

## **Did Purple-Air Sensors Agree with EPA Regulatory Air Quality Monitors in North Carolina During the Canadian Wildfire Smoke Episode in the Summer of 2023?**

Jackson Coley  
Department of Atmospheric Sciences  
The University of North Carolina Asheville  
One University Heights  
Asheville, North Carolina 28804 USA  
Faculty Mentors: Evan Couzo, Elaine Godfrey

**ABSTRACT:** Particulate matter with a diameter between 1.0–2.5  $\mu\text{m}$  ( $\text{PM}_{2.5}$ ) is an atmospheric pollutant that primarily forms as a product of combustion. Because of the risk to human health, the EPA regulates  $\text{PM}_{2.5}$  using Air Quality Monitoring Stations (AQMS). Due to extreme cost, AQMS are only located in regions of large population. However, low-cost sensors, such as the Purple-Air PA-II, have been developed to provide  $\text{PM}_{2.5}$  data in less dense regions of population. Previous studies have shown that under low-pollution levels ( $<25 \mu\text{g m}^{-3}$ ), the PA-II unit is accurate when compared to the EPA AQMS. In contrast, under moderate to high pollution levels ( $>25 \mu\text{g m}^{-3}$ ) and under high relative humidity, the PA-II overestimates  $\text{PM}_{2.5}$  concentration. In summer 2023, wildfires raged across northern Canada, leading to thick smoke protruding into the southeastern United States. North Carolina was affected by this smoke as there were several days where reported  $\text{PM}_{2.5}$  concentrations exceeded the EPA daily mean limit of  $35 \mu\text{g m}^{-3}$ . This work will examine the performance of the PA-II unit compared to EPA AQMS in North Carolina in June and July 2023, providing the opportunity to identify the effects of wildfire smoke on the PA-II unit in a region known to have high relative humidity in the summer months. Insight provided from this study will help inform residents of North Carolina on the integrity of the PA-II data.

## Introduction

An aerosol is a solid or liquid particle suspended in a gas. Three different classifications of aerosols exist, based upon the size of the particle. Particles with a diameter greater than 2.5  $\mu\text{m}$  are classified as coarse aerosols ( $\text{PM}_{10}$ ), particles with a diameter between 0.1–2.5  $\mu\text{m}$  are classified as fine aerosols ( $\text{PM}_{2.5}$ ), and particles with a diameter less than 0.1  $\mu\text{m}$  are classified as ultrafine aerosols ( $\text{PM}_{1.0}$ ) (Jacobson 2012, p. 101). Fossil fuel combustion, such as power plants, waste incinerators, automobiles, airplanes, etc. are the leading sources of  $\text{PM}_{2.5}$ . When inhaled by humans, the small size of  $\text{PM}_{2.5}$  allows it to travel far into the lungs, and even into the bloodstream (Ling and van Eeden 2009). The abundance of  $\text{PM}_{2.5}$  in the atmosphere results in negative health impacts for many humans across the globe (Cohen et al. 2017). Due to the danger it poses to humans,  $\text{PM}_{2.5}$  is extensively studied.

## EPA Regulations

In the United States, the Environmental Protection Agency (EPA) is tasked with setting the National Ambient Air Quality Standards (NAAQS) for criteria air pollutants. The EPA identifies the criteria pollutants as ozone, nitrogen dioxide, carbon monoxide, sulfur dioxide, lead, and PM (Jacobson 2012, p. 178). These six pollutants are the most dangerous to human health. The NAAQS are in place to establish pollution limits one region (usually a city or county) can allow for a given period. The annual mean NAAQS value for  $\text{PM}_{2.5}$  is 9  $\mu\text{g m}^{-3}$  and the daily mean NAAQS value for  $\text{PM}_{2.5}$  is 35  $\mu\text{g m}^{-3}$  (Feenstra et al. 2019). The EPA regularly reviews the NAAQS to determine if they need to be lowered to continuously protect humans and the environment. A known toxin to humans,  $\text{PM}_{2.5}$  causes millions of deaths per year worldwide (Cohen et al. 2017). Without the NAAQS, people and businesses could pollute freely, which would cause human harm (Cohen et al. 2017). The EPA uses highly advanced air quality measuring devices to record  $\text{PM}_{2.5}$ , which provide the data used in the NAAQS calculations. These devices are known as Air Quality Monitoring Stations (AQMS) and are deployed across the United States. They calculate  $\text{PM}_{2.5}$  concentrations using the Federal Equivalent Method (FEM), which uses beta ray attenuation to measure particles (Ardon-Dryer et al. 2020). The EPA only installs AQMS in cities or counties with large populations due to the cost of initial manufacturing and frequent maintenance. Although AQMS cost tens of thousands of dollars, their frequent calibration and utilization of the FEM makes their data trustworthy. But due to the high spatial resolution of  $\text{PM}_{2.5}$  observations, a region may have nonuniform  $\text{PM}_{2.5}$  concentrations if one specific area within that region has a point source of pollution. This will cause conflicting statuses of attainment, meaning that a region can meet the NAAQS requirement, despite polluting lots of  $\text{PM}_{2.5}$ . As the NAAQS continue to grow more stringent over time (Jacobson 2012, pp. 178–187), questions arise regarding whether enough AQMS exist to assess a region's attainment status (i.e., the ability to meet the defined NAAQS requirements).

## Low-cost sensors

A practical, alternative method of measuring air quality may come with the adaption of low-cost sensors. Low-cost air quality monitors can measure the composition of ambient

air and are easy for the public to use. These devices range from a few hundred to a few thousand dollars and continue to grow in popularity as a means of measuring air quality (Feenstra et al. 2019). Due to their price, low-cost sensors can be placed in regions that do not have an AQMS, allowing for the collection of air quality data in less populated regions of the country. One of the most popular versions of the low-cost sensor is the PurpleAir PA-II unit. This device is designed to measure  $PM_{1.0}$ ,  $PM_{2.5}$ , and  $PM_{10}$  in real-time and report it to the PurpleAir website to display live (Ardon-Dryer et al. 2020). This study will focus on the use of the PA-II unit. Due to their ease-of-access and ability to provide data to consumers, there are thousands of PA-II sensors deployed worldwide, resulting in a large amount of available air quality data. However, experts have raised questions regarding the value of PA-II data because of the lack of research done to verify its correctness (Ardon-Dryer et al. 2020).

Many studies have aimed to determine if the PA-II unit reports data that is in agreement with reference monitors (e.g., Feenstra et al. 2018; Ardon-Dryer et al. 2020; Magi et al. 2020; Wallace et al. 2021; Barkjohn et al. 2022). Clements et al. (2017) showed that changes in atmospheric variables such as temperature, air pressure, and relative humidity (RH) can cause the PA-II to provide inaccurate data. High RH is a recurring issue that affects the monitors' ability to report accurate  $PM_{2.5}$  concentration. Under humid conditions,  $PM_{2.5}$  can exhibit hygroscopic growth if water vapor were to condense upon it, causing an inaccurate overestimation of  $PM_{2.5}$  concentration (Ardon-Dryer et al. 2020).

## Corrections of PA-II data

Due to the limitations of the PA-II, researchers often apply corrections to the  $PM_{2.5}$  data to bring the unit in agreement with a reference sensor (usually an AQMS). PurpleAir has developed their own method of correction based on advice from Plantower, the manufacturer of the particulate matter sensors within the PA-II unit. Wallace et al. (2021) compared PA-II sensors with AQMS in California and derived a method of correction that they deemed better than the Plantower method. Their method relied on calculating the volume of the particles based on the quantity recorded by the sensor and then used a constant density to calculate mass. Magi et al. (2020) and Ardon-Dryer et al. (2020) used a multivariate linear regression to correct the PA-II  $PM_{2.5}$  values using parameters of temperature, RH, and the recorded AQMS  $PM_{2.5}$  values. Despite variations in methods for calculating correction factors, those working with the PA-II unit generally reached a consensus that some kind of correction becomes necessary for the data to suit scientific use (e.g., Ardon-Dryer et al. 2020; Magi et al. 2020; Wallace et al. 2021; Barkjohn et al. 2022). Most research correcting PA-II data has been done during periods of low air pollution, which leaves a gap in knowledge about the PA-II's performance during high air pollution periods.

## PA-II performance during wildfire/high pollution events

Exceptionally bad air pollution events provide unique opportunities to study the PA-II unit due to elevated pollution levels. Wildfires are a large source of atmospheric  $PM_{2.5}$ . Among many gases, combustion releases carbon dioxide, nitrogen dioxide, nitrogen monoxide, volatile organic compounds, and  $PM_{2.5}$  into the atmosphere (Jaffe et al. 2020). Due to vast amounts of smoke during wildfires,  $PM_{2.5}$  concentrations can rise above 250–

500  $\mu\text{g m}^{-3}$ , which are incredibly unhealthy (Barkjohn et al. 2022). Typically, low-cost sensors measure the air under healthy conditions (0–50  $\mu\text{g m}^{-3}$ ), but during wildfires,  $\text{PM}_{2.5}$  concentrations increase, allowing for analysis of low-cost sensor performance against AQMS under these conditions.

Gupta et al. (2018) and Barkjohn et al. (2022) have tested the PA-II's performance during wildfires. Gupta et al. (2018) studied the PA-II unit for wildfires in California. They did not use a correction factor for the raw  $\text{PM}_{2.5}$  concentrations reported by the PA-II and found that the unit tended to overestimate  $\text{PM}_{2.5}$  values when compared to an AQMS (Gupta et al. 2018). Barkjohn et al. (2022) studied the PA-II unit for wildfires across the continental United States and found that if smoke concentration was less than 200  $\mu\text{g m}^{-3}$ , the PA-II unit usually responded in a linear fashion; this means that a linear regression could be used to correct the data (e.g., Magi et al. 2020 and Ardon-Dryer et al. 2020). But once smoke concentration became greater than 200  $\mu\text{g m}^{-3}$ , the PA-II unit responded in a non-linear fashion (Barkjohn et al. 2022). They discovered that the best method for correcting PA-II data at very high smoke concentrations was to apply a quadratic regression (Barkjohn et al. 2022). The results from both Gupta et al. (2018) and Barkjohn et al. (2022) indicate that it may be necessary to apply a correction factor to the PA-II data to compare it to data recorded by an AQMS for wildfire events.

Another consideration with severe wildfire smoke is the possibility of increased cloud formation due to lower temperatures. If smoke is thick enough to significantly prevent sunlight from reaching the surface, daytime temperatures will decrease, which may cause clouds to form (Conrick et al. 2021). Cloud formation is also induced by the  $\text{PM}_{2.5}$  acting as cloud condensation nuclei. If the high smoke concentrations cause clouds to form, the PA-II unit may begin to overestimate actual  $\text{PM}_{2.5}$  concentration due to the increased humidity. This is potentially the reason behind Barkjohn et al. (2022) figuring out that a quadratic regression was necessary to correct PA-II  $\text{PM}_{2.5}$  values at high smoke concentrations.

Numerous studies have investigated trends in PA-II data during wildfires, with a significant portion of them conducted in the western region of the United States (e.g., Gupta et al. 2018 and Barkjohn et al. 2022). As previously stated, under high RH conditions, the PA-II tends to overestimate the actual  $\text{PM}_{2.5}$  concentration. Given that the western United States is drier than the east (i.e. deserts in the west; temperate forests in the east), there is the possibility that a moister climate could cause the PA-II to behave differently depending on its location. Little research exists on the PA-II during wildfires in the eastern United States, leaving a gap in known vulnerabilities of the sensor. In the summer of 2023, wildfires raged across northern Canada, leading to thick smoke protruding into the southeastern United States, causing dreadful air quality (North Carolina Department of Environmental Quality). North Carolina (NC) was especially affected by this smoke as there were several days where reported  $\text{PM}_{2.5}$  concentrations exceeded the daily mean NAAQS value of 35  $\mu\text{g m}^{-3}$ . This work will examine the performance of the PA-II unit compared to EPA AQMS in NC in June and July 2023, providing the opportunity to identify the effects of wildfire smoke on the PA-II unit in a region known to have high relative humidity in the summer months.

# Data and Methods

## Acquisition of Data

The NASA Worldview satellite imagery tool (NASA 2023) was used to identify the date range when NC was affected by Canadian wildfire smoke. The corrected reflectance color band makes wildfire smoke easily identifiable because it has a grey tint, which is distinguishable from clouds. This color band is derived from the Visible Infrared Imaging Radiometer Suite aboard the NASA JPSS-1 satellite, with an image resolution of 250 m (NASA 2023). Daily imagery from 1 April 2023 through 31 August 2023 showed Canadian wildfire smoke covering some or all of NC for most of June and July. To grasp a full picture of how smoke affected PM<sub>2.5</sub> concentrations at the air quality monitors, it was determined that the date range for the study would be 29 May 2023 through 29 July 2023. Wildfire smoke began its intrusion into NC during the first week of June, so beginning the date range on 29 May allowed for the inclusion of a few days of clean air, providing the comparison of PA-II to AQMS under low PM<sub>2.5</sub> concentrations. The beginning of August showed that the Canadian wildfire smoke was no longer being forced into NC, so the date range was cut off on 29 July. Hourly observations were used, resulting in approximately 1463 hours of PM<sub>2.5</sub> data for each PA-II and AQMS.

The EPA AirData Map (EPA 2023a) was used to identify all active PM<sub>2.5</sub> AQMS in NC. Hourly raw PM<sub>2.5</sub> concentrations were retrieved from each AQMS for the two-month period. To provide an accurate comparison between the AQMS and PA-II, there needed to be one PA-II co-located with each AQMS. A radius of 5 miles was determined as the maximum allowable distance between the PA-II and AQMS for co-location. This regulation helped eliminate co-located pairs that were very far apart, while keeping enough pairs for a robust comparison. The EPA Fire and Smoke map (EPA 2023b) was used to identify PA-II units that were close to AQMS sites. The Fire and Smoke map is a tool developed by the EPA that displays all AQMS and PA-II units on an interactive map, making air quality data accessible. Once the nearest PA-II unit was identified, its sensor ID was recorded. The PurpleAir data download tool was used to acquire hourly averaged PA-II PM<sub>2.5</sub> concentrations for each PA-II unit. After a co-located PA-II unit was identified for each AQMS, there were a total of eight sites across NC. The co-located pairs were named according to the assigned AQMS site name from the EPA. There are co-located pairs in Asheville, Charlotte (Friendship Park), Winston Salem (Hattie Avenue), Greensboro, Durham, Raleigh, Triple Oak, and Fayetteville.

In alignment with previous research, moisture variables were collected to analyze the performance of the PA-II during periods of high humidity. The Cardinal Data Retrieval System (State Climate Office of North Carolina) was used to pull hourly RH and specific humidity data for each co-located pair of PA-II and AQMS from 29 May 2023 to 29 July 2023. The RH and specific humidity were then time paired to the PA-II and AQMS PM<sub>2.5</sub> data. The weather stations used were a combination of NC ECONet and ASOS sites. The Asheville, Greensboro, Durham, and Raleigh co-located pairs utilized the ECONet sites of the UNC Asheville Weather Tower, the NC A&T Research Farm, the North Durham Water Reclamation Facility, and the Reedy Creek Field Laboratory, respectively. The Charlotte, Winston-Salem, Fayetteville, and Triple Oak co-located pairs utilized the ASOS

sites of Charlotte Douglass International Airport, Smith Reynolds Airport, Fayetteville Regional Airport, and Raleigh-Durham International Airport, respectively.

Previous literature has stated that high RH can cause overestimation by the PA-II units under the justification that the  $PM_{2.5}$  grows hygroscopically (Clements et al. 2017 and Ardon-Dryer et al. 2020). However, in a meteorological sense, RH is not the best moisture variable to measure the vapor content of the atmosphere. RH is dependent on the saturation vapor pressure of the air, which is a function of temperature (Petty 2008, p. 174). Due to this dependence, high RH can be observed during times when it is not particularly humid. For these reasons, the decision was made to also investigate specific humidity, which is not dependent on temperature. Specific humidity is the ratio of mass of water vapor to the total mass of the air and has units of  $g\ kg^{-1}$  (Petty 2008, p. 73). It is a better moisture variable to use when we are concerned with the absolute vapor content of the atmosphere. Utilizing both RH and specific humidity will provide the opportunity to investigate how the PA-II performs against the AQMS under various humidity levels.

## Comparison of data from co-located pairs

Two sample t-tests were calculated using the PA-II and AQMS for each co-located pair to compare the sample means of the  $PM_{2.5}$  concentrations. Results from the tests indicated that all co-located pairs had  $PM_{2.5}$  concentrations that had statistically different sample means.

Scatterplots were then used to gain a visual understanding of the data. To mask some of the PA-II outliers, the axis bounds for each scatterplot were set at  $100\ \mu g\ m^{-3}$ . Two sets of scatterplots were made, one using the RH data, and the other using the specific humidity data. For each co-located pair, the PA-II data was plotted against the AQMS data, with the humidity variable depicted by the color of the marker. It was immediately clear that there were too many overlapping data points to gain a grasp of the role humidity played in overestimation. To rectify this, the  $PM_{2.5}$  values were separated into three bins corresponding to the value of the humidity variable at that hour. Ardon-Dryer et al. (2020) defined high RH as  $90\% < RH$ , which served as a baseline in this study. After some adjustments, and making an effort to minimize skew, the low RH bin was set at  $RH \leq 60\%$ , the medium RH bin was set at  $60\% < RH \leq 85\%$ , and the high bin was set at  $85\% < RH$ . For specific humidity, the absence of previous research made it more difficult to define bin widths. Ultimately, equal bin widths were selected, and the low bin was set at  $q \leq 10\ g\ kg^{-1}$ , the medium bin was set at  $10\ g\ kg^{-1} < q \leq 16\ g\ kg^{-1}$ , and the high bin was set at  $16\ g\ kg^{-1} < q$ . These bins were essentially bounded on either end by the minimum specific humidity recorded over the period being approximately  $4\ g\ kg^{-1}$  and the maximum specific humidity recorded being approximately  $22\ g\ kg^{-1}$ . Additionally, these binned RH and specific humidity plots were made combining all eight co-located sites  $PM_{2.5}$  data, to provide the opportunity to identify disparities between a specific co-located site and the statewide average.

A linear regression was calculated for each co-located pair in each humidity bin for each humidity variable for PA-II vs AQMS  $PM_{2.5}$  concentrations. The slope and y-intercept were used from the regression to plot a line of best fit on each scatterplot corresponding

to the PM<sub>2.5</sub> data. The linear regression also returned the Pearson correlation coefficient (R), which was used to calculate the coefficient of determination (R<sup>2</sup>).

## Results

### Scatterplots of PA-II and AQMS PM<sub>2.5</sub> as a function of Relative Humidity

Figure 1 shows an aggregated scatterplot of PA-II vs AQMS PM<sub>2.5</sub> as a function of RH for all co-located sites. Figures 2–9 show the resultant scatterplots of PA-II vs AQMS PM<sub>2.5</sub> as a function of RH at each co-located site.

The aggregated scatterplot shows that the PA-II overestimated actual PM<sub>2.5</sub> concentrations for all three RH bins (Fig. 1). This is easily identifiable because most of the data resides above the 1:1 line. Data points lying above the 1:1 line indicate overestimation because the PA-II reports a higher PM<sub>2.5</sub> concentration than its co-located AQMS. The 1:1 line is in place to depict the perfect comparison. This line has a slope of 1.0 and data points falling on this line indicate the PA-II unit and AQMS reported the same PM<sub>2.5</sub> concentration at that given time, which is the most ideal scenario. And as we see with all three RH bins, slopes greater than one indicates that as the AQMS PM<sub>2.5</sub> concentration increased, the PA-II performance decreases (i.e., the PA-II overestimation increased with increasing PM<sub>2.5</sub> concentration). The high RH bin showed the greatest overestimation, with a slope of 1.32. The R<sup>2</sup> value was 0.82, which indicates that the linear regression line did a decent job of fitting the data. The low RH bin showed the second most overestimation, with a slope of 1.16 and R<sup>2</sup> value of 0.74. The medium RH bin had the lowest overestimation, with a slope of 1.10 and R<sup>2</sup> value of 0.62. This R<sup>2</sup> value shows that the linear regression line did not do as good of a job representing the data, meaning that there was more variability.

When looking at the comparisons of the PA-II vs AQMS units as a function of RH for an individual co-located pair, they are like the aggregated scatterplot. Asheville (Fig. 2), Charlotte (Fig. 3), Durham (Fig. 4), Fayetteville (Fig. 5), Raleigh (Fig. 7), and Triple Oak (Fig. 8) all see positive slopes for all three RH bins. These slopes range from 1.10 – 1.50. Across all eight co-located pairs, the low RH bin had an average slope of 1.23, the medium RH bin had an average slope of 1.17, and the high RH bin had an average slope of 1.33. These average slopes show that the high RH bin saw the most PA-II overestimation, while the medium bin saw the least overestimation. The high RH bin had the highest mean R<sup>2</sup> value of 0.86, while the medium RH bin had the lowest mean R<sup>2</sup> of 0.76.

Two plots that did not show overestimation were the middle RH bin for Greensboro (Fig. 6) and the low and middle RH bins for Winston-Salem (Fig. 9). The middle RH bin for Greensboro has a slope of 0.79, which is undoubtedly influenced by many underreported PA-II outlier values beyond 100 µg m<sup>-3</sup>. This could have been caused by a piece of trash or larger natural material such as a blade of grass blocking the inlet to the PA-II particulate matter sensor. The low and medium RH bins for Winston-Salem (Fig. 9) have slopes of 0.99, and 1.00 respectively. This indicates that the PA-II did not show increasing overestimation with increasing humidity, which is a good sign.

## Scatterplots of PA-II and AQMS PM<sub>2.5</sub> as a function of Specific Humidity

Figure 10 shows an aggregated scatterplot of PA-II vs AQMS PM<sub>2.5</sub> as a function of specific humidity for all co-located sites. Figures 11–18 show the resultant scatterplots of PA-II vs AQMS PM<sub>2.5</sub> as a function of specific humidity at each co-located site.

Once again, the aggregated scatterplot shows that the PA-II generally overestimated actual PM<sub>2.5</sub> concentrations for all three specific humidity bins (Fig. 10). We see a large percentage of the data residing above the 1:1 line along with positive slopes in each bin. The high specific humidity bin showed the greatest overestimation, with a slope of 1.24, however, the R<sup>2</sup> value was only 0.73, which indicated that the linear regression line did not do a great job of fitting the data. The low specific humidity bin showed the second most overestimation, with a slope of 1.18. The medium specific humidity bin had the lowest overestimation, with a slope of 1.16. There was less variability between the slopes of the aggregated plot, but overall, more variability between the data within each bin, as the R<sup>2</sup> values were not as large as for the RH aggregated plot (Fig. 1).

When looking at the comparisons of the PA-II vs AQMS units as a function of specific humidity for an individual co-located pair, they are somewhat like the aggregated scatterplot. Durham (Fig. 13), Fayetteville (Fig. 14), Raleigh (Fig. 16), and Triple Oak (Fig. 17) all see positive slopes for all three specific humidity bins. These slopes range from 1.15–1.51. Asheville (Fig. 11), Charlotte (Fig. 12), Greensboro (Fig. 15), and Winston-Salem (Fig. 18) all have at least one specific humidity bin with a slope at or less than 1.0. The high specific humidity bin for Asheville has a slope of 0.82, the low specific humidity bin for Charlotte has a slope of 0.93, the middle specific humidity bin for Greensboro has a slope of 0.97, and the low specific humidity bin for Winston-Salem has a slope of 1.00. Across all eight co-located pairs, the low specific humidity bin had an average slope of 1.26, the medium specific humidity bin had an average slope of 1.19, and the high specific humidity bin had an average slope of 1.23. These average slopes show that the low specific humidity bin saw the most PA-II overestimation, while the medium bin saw the least overestimation. The low specific humidity bin saw the highest mean R<sup>2</sup> value of 0.81, while the high specific humidity bin saw the lowest mean R<sup>2</sup> of 0.74.

## Discussion

For all eight co-located sites, the PA-II units showed overestimation at high RH, which was anticipated. Something that was unexpected was the continued overestimation at the middle and low RH bins. In previous research, there were not many claims regarding the PA-II performance at middle and low RH levels. Because of this, it was assumed that the PA-II generally showed less overestimation and tended to agree more with the reference monitors at lower humidity. This leads to the question regarding RH bin sizes. Ardon-Dryer et al. (2020) defined high RH as RH > 90% and low RH as RH < 40%, while this study used high RH as RH > 85% and low RH as RH ≤ 60%. In this study, the bins were chosen to minimize skew, which caused the high RH bin to be reduced to 85% and the low RH bin to increase to 60%. While perusing the data in the preliminary stages of this project, it was clear that there were few hours of data with low RH, meaning a cut-off of 40% would cause the low RH bin to have little population. Evidently, this could end up

being a major discrepancy because with these results, we cannot definitively say that only high RH caused PA-II overestimation. Another question to consider is climate differences. Ardon-Dryer et al. (2020) used PA-II units in Salt Lake City, Utah; Denver, Colorado; San Francisco, California; and Vallejo, California. These four climates are in the western United States, which is known to be drier than the east. A low RH in NC during the summer is more than likely going to be at least a mid-range RH for the four cities used in Ardon-Dryer et al. (2020). Since a large percentage of hourly data in NC is going to be during times of at least moderate humidity (middle or high RH bin), it may be necessary to define a lower threshold for humidity at which the PA-II begins to significantly overestimate PM<sub>2.5</sub> concentrations.

The results for PA-II vs AQMS performance as a function of specific humidity were not as consistent as they were for RH. Unlike RH, the high specific humidity bin saw an average R<sup>2</sup> value that was less than the middle or low bin. This is likely attributed to the high specific humidity bin not being set at an appropriate value. As previously stated, the lack of previous literature investigating the PA-II at different levels of specific humidity left severe uncertainty regarding the correct bin widths. The results suggest that the linear regression line on average did not do as good of a job for the high specific humidity bin as it did with the low specific humidity bin. This means that the data was more variable and not as organized for higher specific humidity, which further leads to the concern that the binning methodology needs refinement. Additionally, the middle specific humidity bin saw much higher data representation, with 63% of all hours residing in this bin. Based on the average slopes, the low specific humidity bin saw the most overestimation, which does not align with the trend for RH. However, the average slopes for each three bins were all very similar.

## Conclusion

The Canadian wildfire smoke intrusion of June and July 2023 provided an opportunity to study how low-cost PA-II units performed against expensive EPA AQMS in NC. Incorporating RH analysis into PA-II performance allowed for the comparison of this study to previous literature. Overall, the PA-II overestimated at high RH, which agreed with previous research. There was a disparity though at lower RH values. Previous research did not find significant overestimation at lower RH values, but this research did. This discrepancy could be attributed to NC being in a more humid climate, meaning that high RH persists most of the season. Among PA-II units located in the southeastern United States, there is likely always going to be overestimation, at least during the summer months when humidity is high.

The specific humidity analysis was introduced to potentially find an alternative moisture variable that affects reported PA-II PM<sub>2.5</sub> concentrations. Specific humidity is independent of temperature, meaning it is a better measure of the true vapor content of the atmosphere. Results from this study are inconclusive based on the uncertainty regarding proper bin size for specific humidity values. The results were not the same as RH (i.e., higher slopes for low bin, indicating more overestimation), meaning that for a contradictory claim to be made (i.e. more overestimation for low humidity values), more research is needed.

There is also the possibility that the concentrated wildfire smoke caused unknown variability in how the PA-II reports its PM<sub>2.5</sub> concentrations. It is known that the PA-II overestimates during high RH, but what about when there is a high concentration of wildfire smoke as well? There could be chemistry taking place between water vapor and smoke particles in the air, causing changes in the reported PM<sub>2.5</sub> concentration from the PA-II particulate sensor.

Residents of NC should be cautious about PA-II data. The best source for accurate air quality data is still AQMS – their frequent maintenance and calibration ensure that accurate air quality is measured around the clock. The EPA Fire and Smoke map is a fantastic tool to use to obtain real time air quality because they preprocess the PA-II data and apply a correction before displaying it on the map. Users of the EPA Fire and Smoke map can safely assume that the PA-II data displayed on the site is trustworthy.

## Acknowledgment

I acknowledge the use of imagery from the NASA Worldview application (<https://worldview.earthdata.nasa.gov>), part of the NASA Earth Observing System Data and Information System (EOSDIS).

I would like to thank Dr. Elaine Godfrey and Dr. Evan Couzo for their guidance as I progressed on this project throughout the 2023-2024 academic year.

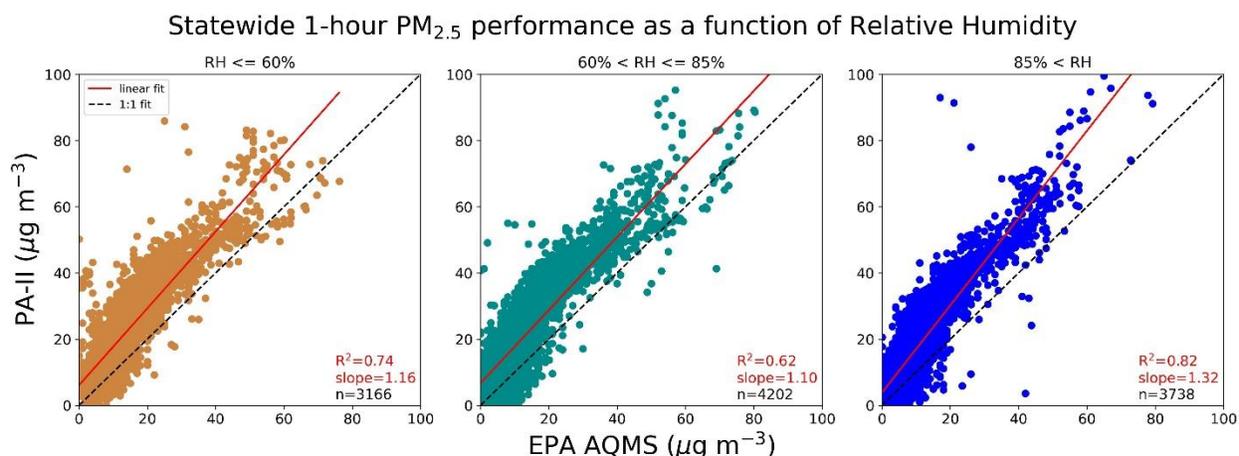


Figure 1. Aggregated comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for all co-located pairs as a function of Relative Humidity. The left, middle, and right panels represent PM<sub>2.5</sub> for RH  $\leq$  60%, 60% < RH  $\leq$  85%, and 85% < RH. The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and R<sup>2</sup> value. The number of data points in each panel is represented by n.

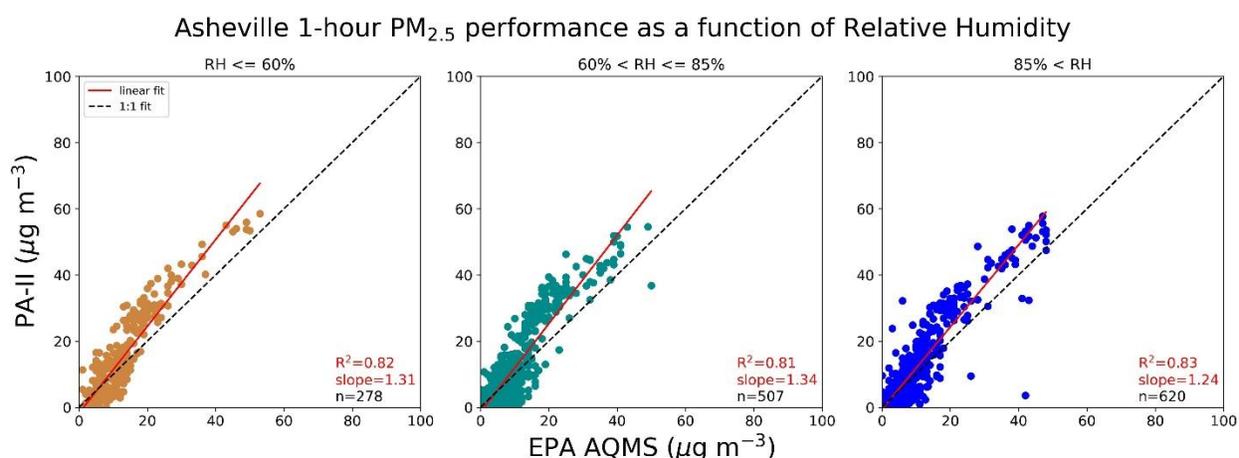


Figure 2. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Asheville co-located pair as a function of Relative Humidity. The left, middle, and right panels represent PM<sub>2.5</sub> for RH  $\leq$  60%, 60% < RH  $\leq$  85%, and 85% < RH. The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and R<sup>2</sup> value. The number of data points in each panel is represented by n.

### Charlotte 1-hour PM<sub>2.5</sub> performance as a function of Relative Humidity

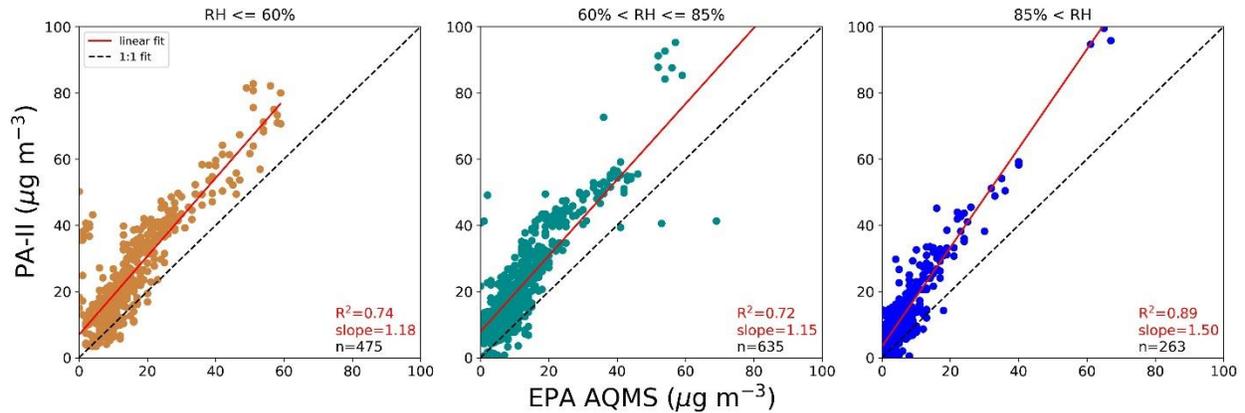


Figure 3. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Charlotte co-located pair as a function of Relative Humidity. The left, middle, and right panels represent PM<sub>2.5</sub> for RH  $\leq$  60%, 60% < RH  $\leq$  85%, and 85% < RH. The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and  $R^2$  value. The number of data points in each panel is represented by n.

### Durham 1-hour PM<sub>2.5</sub> performance as a function of Relative Humidity

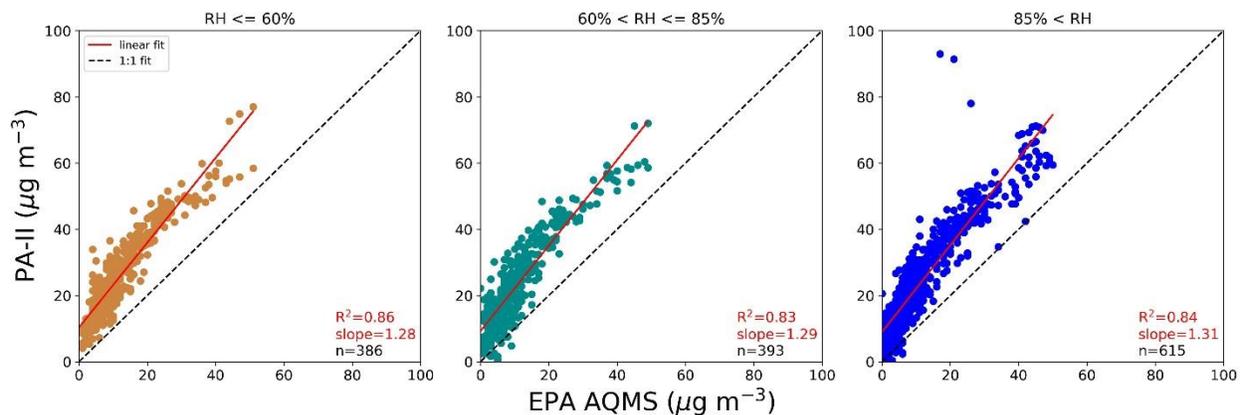


Figure 4. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Durham co-located pair as a function of Relative Humidity. The left, middle, and right panels represent PM<sub>2.5</sub> for RH  $\leq$  60%, 60% < RH  $\leq$  85%, and 85% < RH. The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and  $R^2$  value. The number of data points in each panel is represented by n.

### Fayetteville 1-hour PM<sub>2.5</sub> performance as a function of Relative Humidity

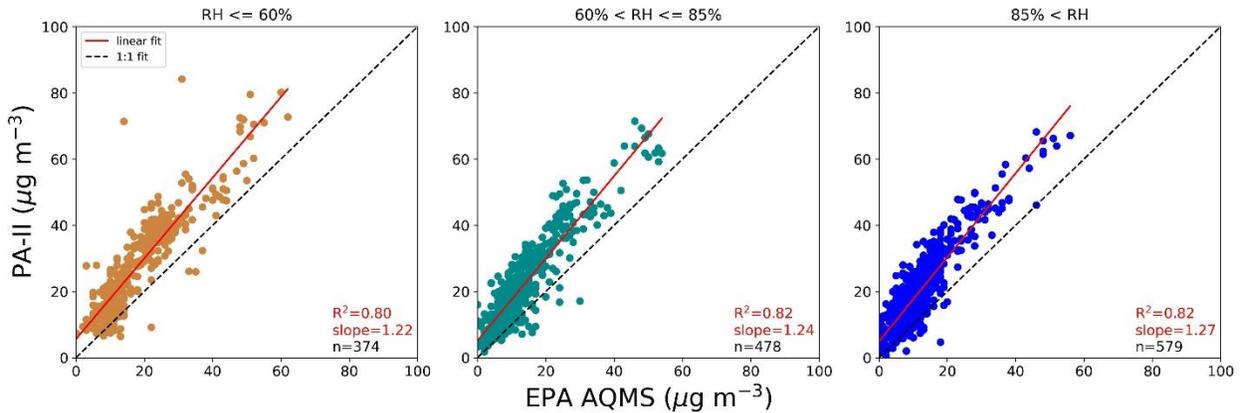


Figure 5. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Fayetteville co-located pair as a function of Relative Humidity. The left, middle, and right panels represent PM<sub>2.5</sub> for RH  $\leq$  60%, 60% < RH  $\leq$  85%, and 85% < RH. The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and  $R^2$  value. The number of data points in each panel is represented by n.

### Greensboro 1-hour PM<sub>2.5</sub> performance as a function of Relative Humidity

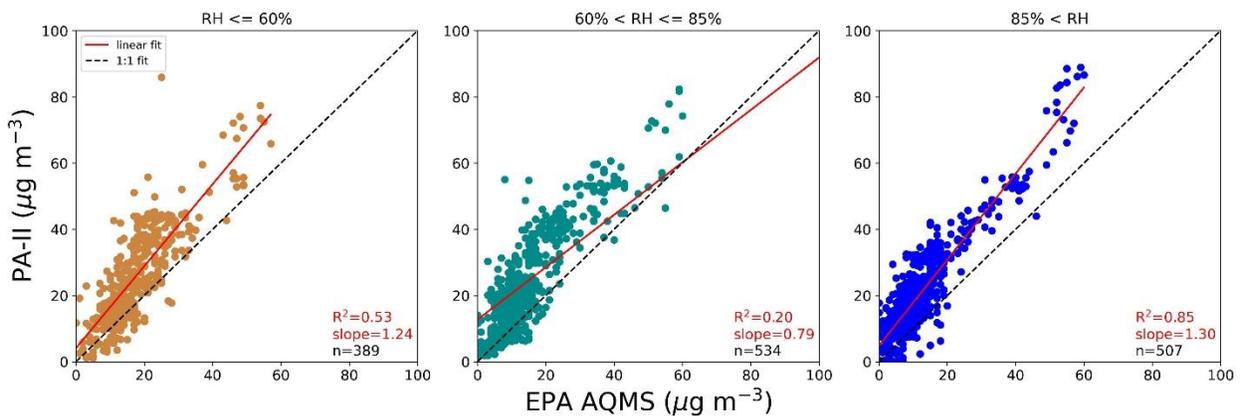


Figure 6. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Greensboro co-located pair as a function of Relative Humidity. The left, middle, and right panels represent PM<sub>2.5</sub> for RH  $\leq$  60%, 60% < RH  $\leq$  85%, and 85% < RH. The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and  $R^2$  value. The number of data points in each panel is represented by n.

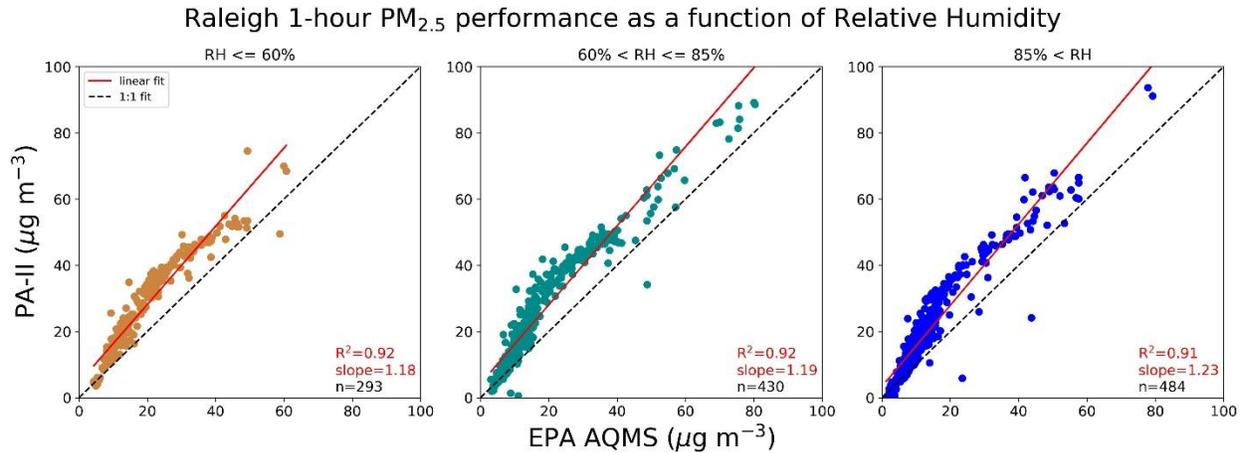


Figure 7. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Raleigh co-located pair as a function of Relative Humidity. The left, middle, and right panels represent PM<sub>2.5</sub> for RH  $\leq$  60%, 60% < RH  $\leq$  85%, and 85% < RH. The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and R<sup>2</sup> value. The number of data points in each panel is represented by n.

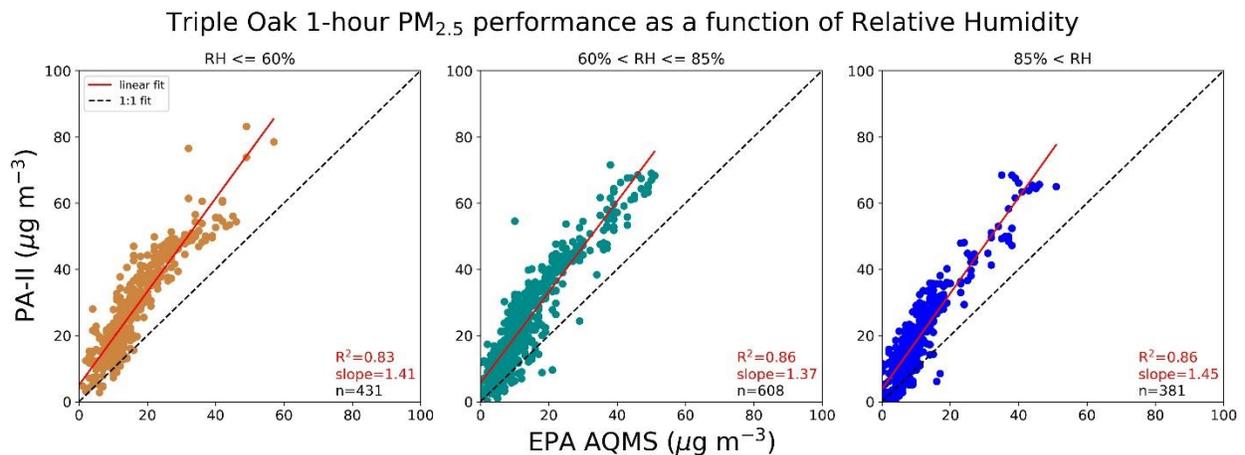


Figure 8. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Triple Oak co-located pair as a function of Relative Humidity. The left, middle, and right panels represent PM<sub>2.5</sub> for RH  $\leq$  60%, 60% < RH  $\leq$  85%, and 85% < RH. The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and R<sup>2</sup> value. The number of data points in each panel is represented by n.

### Winston-Salem 1-hour PM<sub>2.5</sub> performance as a function of Relative Humidity

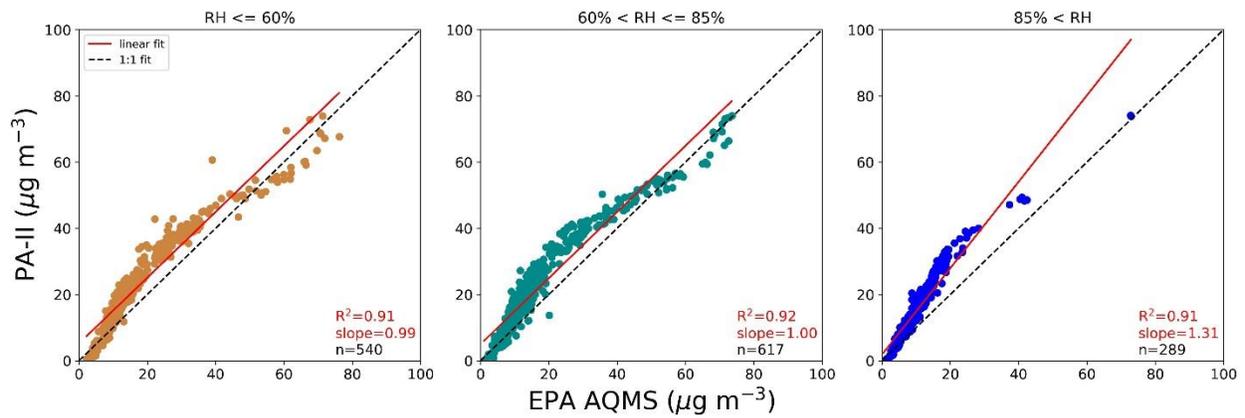


Figure 9. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Winston-Salem co-located pair as a function of Relative Humidity. The left, middle, and right panels represent PM<sub>2.5</sub> for RH  $\leq 60\%$ ,  $60\% < \text{RH} \leq 85\%$ , and  $85\% < \text{RH}$ . The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and  $R^2$  value. The number of data points in each panel is represented by n.

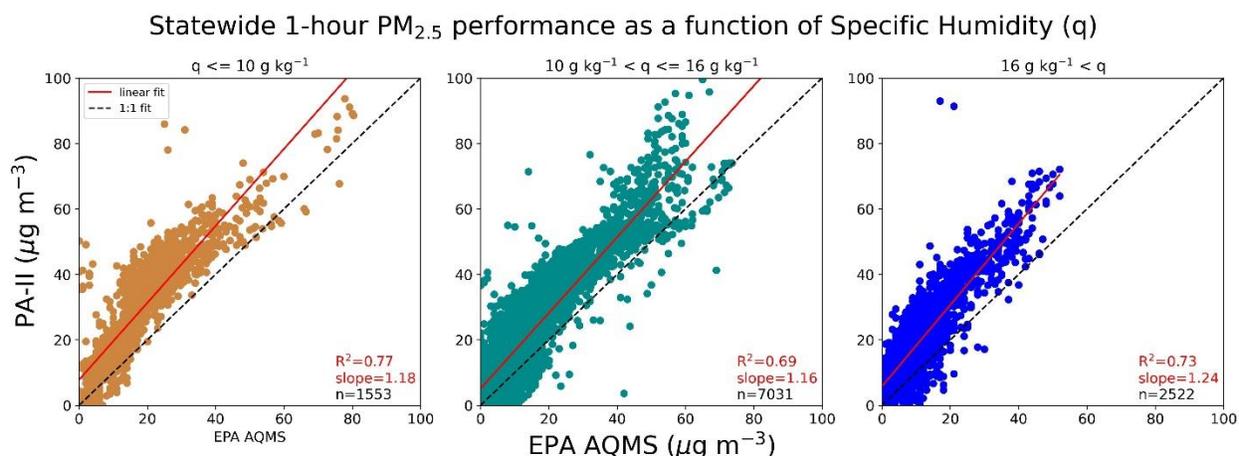


Figure 10. Aggregated comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for all co-located pairs as a function of Specific Humidity (q). The left, middle, and right panels represent PM<sub>2.5</sub> for  $q \leq 10 \text{ g kg}^{-1}$ ,  $10 \text{ g kg}^{-1} < q \leq 16 \text{ g kg}^{-1}$  and  $16 \text{ g kg}^{-1} < q$ . The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and R<sup>2</sup> value. The number of data points in each panel is represented by n.

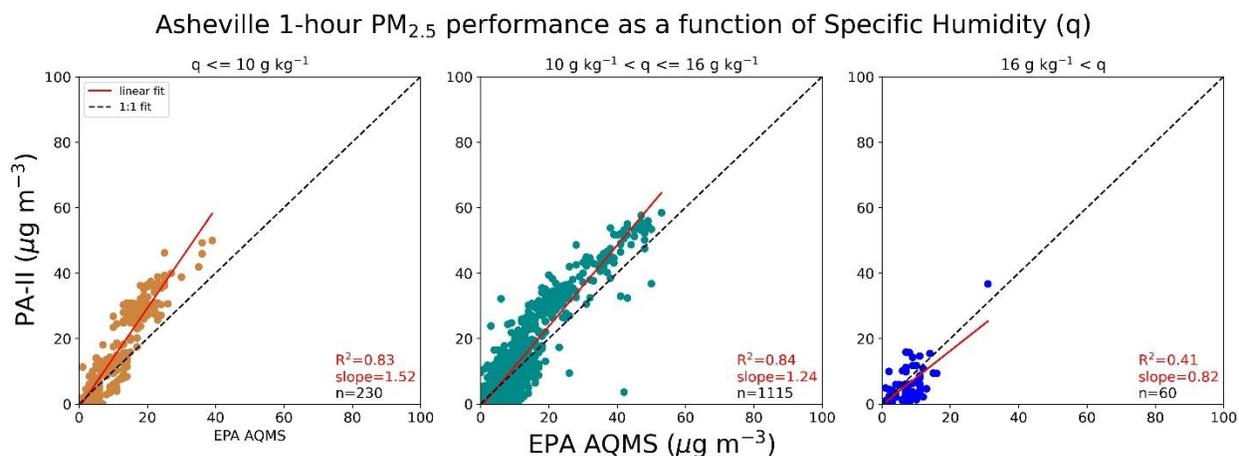


Figure 11. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Asheville co-located pair as a function of Specific Humidity (q). The left, middle, and right panels represent PM<sub>2.5</sub> for  $q \leq 10 \text{ g kg}^{-1}$ ,  $10 \text{ g kg}^{-1} < q \leq 16 \text{ g kg}^{-1}$  and  $16 \text{ g kg}^{-1} < q$ . The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and R<sup>2</sup> value. The number of data points in each panel is represented by n.

### Charlotte 1-hour PM<sub>2.5</sub> performance as a function of Specific Humidity (q)

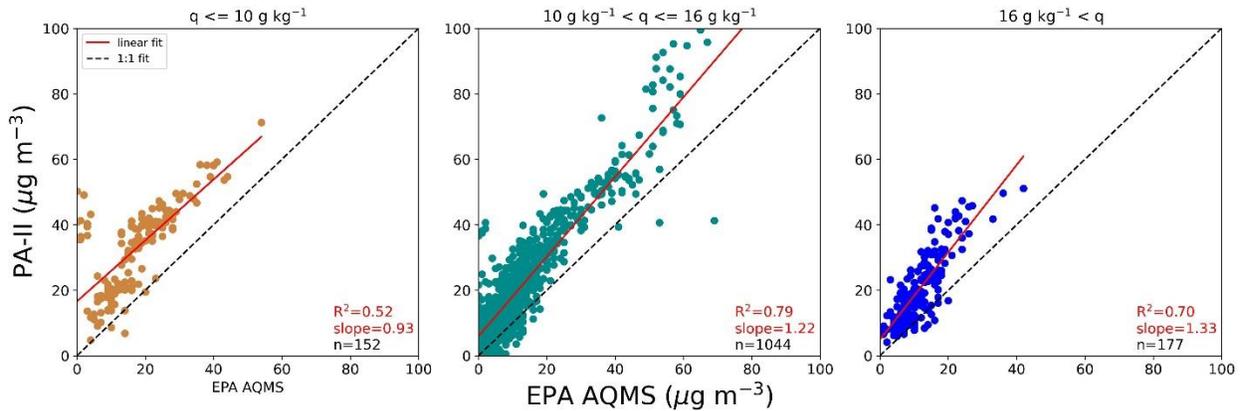


Figure 12. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Charlotte co-located pair as a function of Specific Humidity (q). The left, middle, and right panels represent PM<sub>2.5</sub> for  $q \leq 10 \text{ g kg}^{-1}$ ,  $10 \text{ g kg}^{-1} < q \leq 16 \text{ g kg}^{-1}$  and  $16 \text{ g kg}^{-1} < q$ . The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and R<sup>2</sup> value. The number of data points in each panel is represented by n.

### Durham 1-hour PM<sub>2.5</sub> performance as a function of Specific Humidity (q)

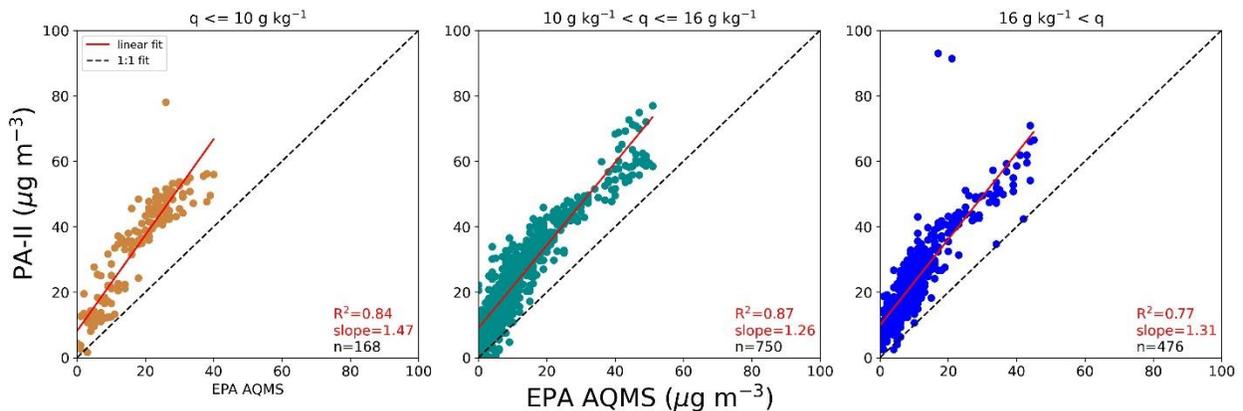


Figure 13. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Durham co-located pair as a function of Specific Humidity (q). The left, middle, and right panels represent PM<sub>2.5</sub> for  $q \leq 10 \text{ g kg}^{-1}$ ,  $10 \text{ g kg}^{-1} < q \leq 16 \text{ g kg}^{-1}$  and  $16 \text{ g kg}^{-1} < q$ . The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and R<sup>2</sup> value. The number of data points in each panel is represented by n.

### Fayetteville 1-hour PM<sub>2.5</sub> performance as a function of Specific Humidity (q)

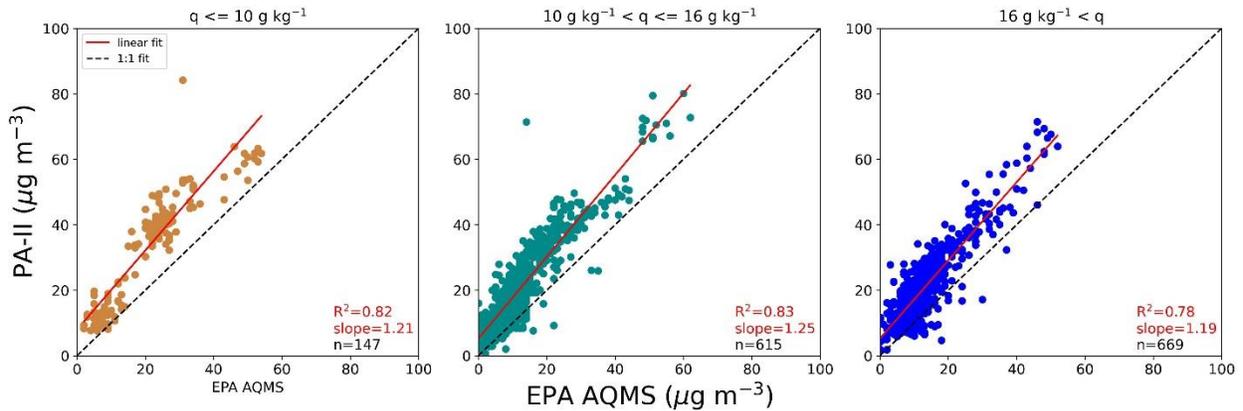


Figure 14. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Fayetteville co-located pair as a function of Specific Humidity (q). The left, middle, and right panels represent PM<sub>2.5</sub> for  $q \leq 10 \text{ g kg}^{-1}$ ,  $10 \text{ g kg}^{-1} < q \leq 16 \text{ g kg}^{-1}$  and  $16 \text{ g kg}^{-1} < q$ . The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and  $R^2$  value. The number of data points in each panel is represented by n.

### Greensboro 1-hour PM<sub>2.5</sub> performance as a function of Specific Humidity (q)

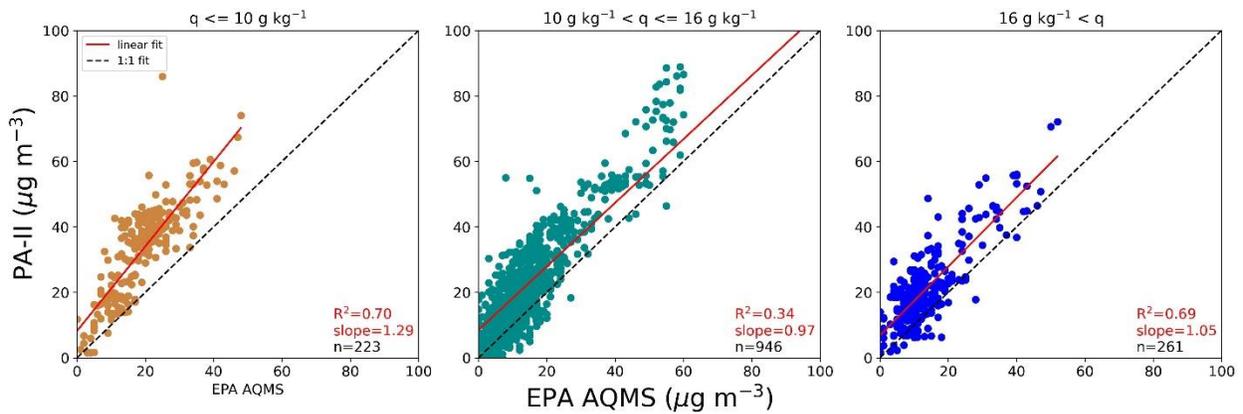


Figure 15. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Greensboro co-located pair as a function of Specific Humidity (q). The left, middle, and right panels represent PM<sub>2.5</sub> for  $q \leq 10 \text{ g kg}^{-1}$ ,  $10 \text{ g kg}^{-1} < q \leq 16 \text{ g kg}^{-1}$  and  $16 \text{ g kg}^{-1} < q$ . The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and  $R^2$  value. The number of data points in each panel is represented by n.

### Raleigh 1-hour PM<sub>2.5</sub> performance as a function of Specific Humidity (q)

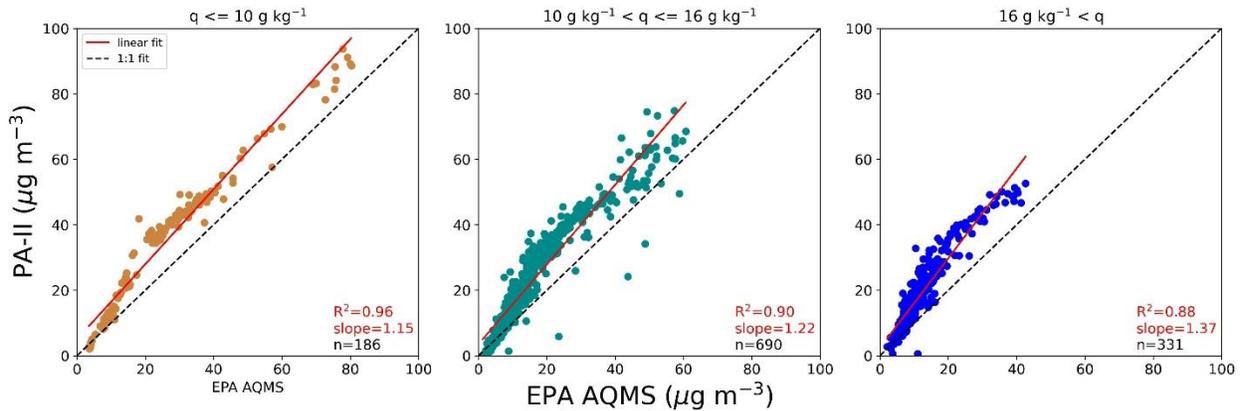


Figure 16. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Raleigh co-located pair as a function of Specific Humidity (q). The left, middle, and right panels represent PM<sub>2.5</sub> for  $q \leq 10 \text{ g kg}^{-1}$ ,  $10 \text{ g kg}^{-1} < q \leq 16 \text{ g kg}^{-1}$  and  $16 \text{ g kg}^{-1} < q$ . The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and R<sup>2</sup> value. The number of data points in each panel is represented by n.

### Triple Oak 1-hour PM<sub>2.5</sub> performance as a function of Specific Humidity (q)

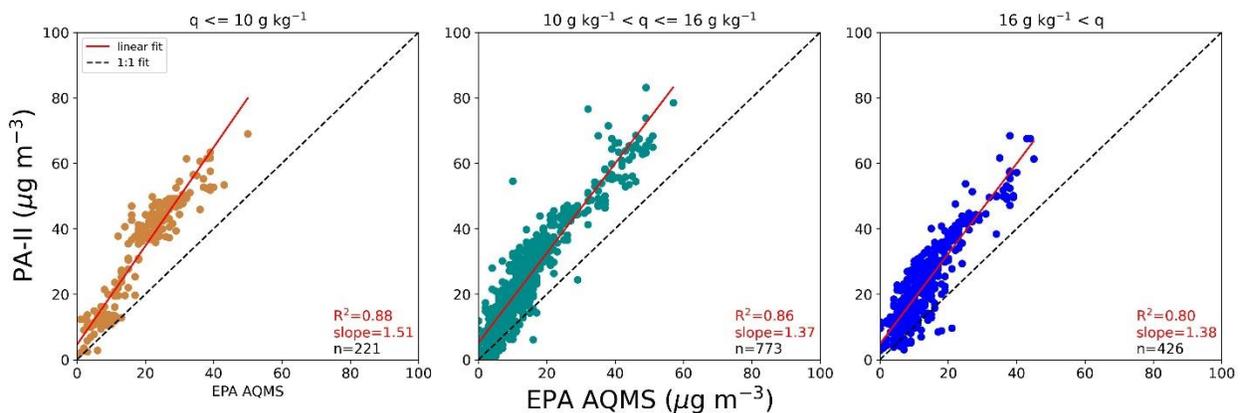


Figure 17. Comparison of PA-II vs AQMS PM<sub>2.5</sub> concentrations ( $\mu\text{g m}^{-3}$ ) for the Triple Oak co-located pair as a function of Specific Humidity (q). The left, middle, and right panels represent PM<sub>2.5</sub> for  $q \leq 10 \text{ g kg}^{-1}$ ,  $10 \text{ g kg}^{-1} < q \leq 16 \text{ g kg}^{-1}$  and  $16 \text{ g kg}^{-1} < q$ . The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and R<sup>2</sup> value. The number of data points in each panel is represented by n.

### Winston-Salem 1-hour $PM_{2.5}$ performance as a function of Specific Humidity ( $q$ )

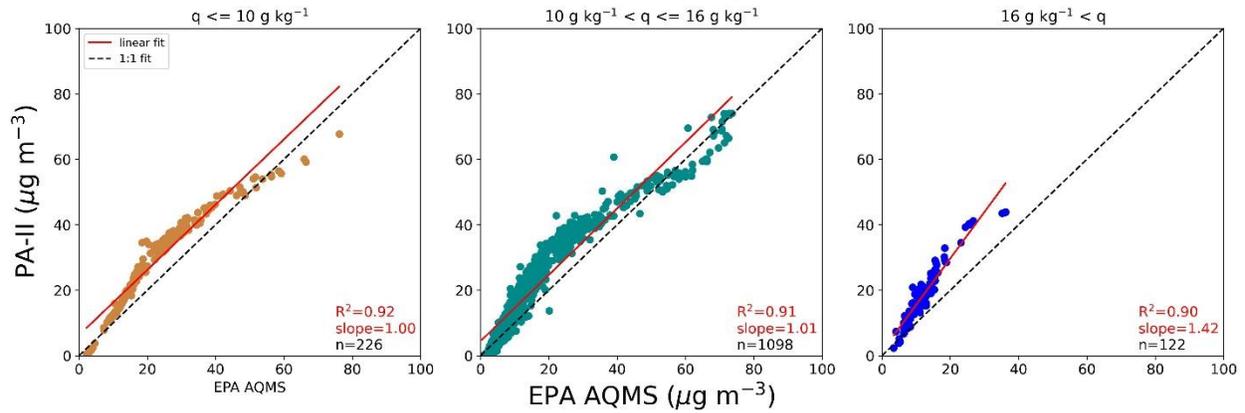


Figure 18. Comparison of PA-II vs AQMS  $PM_{2.5}$  concentrations ( $\mu\text{g m}^{-3}$ ) for the Winston-Salem co-located pair as a function of Specific Humidity ( $q$ ). The left, middle, and right panels represent  $PM_{2.5}$  for  $q \leq 10 \text{ g kg}^{-1}$ ,  $10 \text{ g kg}^{-1} < q \leq 16 \text{ g kg}^{-1}$  and  $16 \text{ g kg}^{-1} < q$ . The dotted black line represents the 1:1 comparison. The solid red line in each panel represents the line of best fit, with the corresponding slope and  $R^2$  value. The number of data points in each panel is represented by n.

## REFERENCES

- Ardon-Dryer, K., Y. Dryer, J. N. Williams, and N. Moghim, 2020: Measurements of PM<sub>2.5</sub> with PurpleAir under atmospheric conditions. *Atmos. Meas. Tech.*, **13**, 5441–5458, <https://doi.org/10.5194/amt-13-5441-2020>.
- Barkjohn, K. K., A. L. Holder, S. G. Frederick, A. L. Clements, and R. Evans, 2021: Sensor Data Cleaning and Correction: Application on the AirNow Fire and Smoke Map. US EPA office of research and development, Accessed 27 November 2023, [https://cfpub.epa.gov/si/si\\_public\\_record\\_report.cfm?dirEntryId=353088&Lab=CEMM](https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=353088&Lab=CEMM).
- , A. L. Holder, S. G. Frederick, and A. L. Clements, 2022: Correction and accuracy of PurpleAir PM<sub>2.5</sub> measurements for extreme wildfire smoke. *Sensors*, **22**, 9669, <https://doi.org/10.3390/s22249669>.
- Clements, A. L., W. G. Griswold, R. S. Abhijit, J. E. Johnston, M. M. Herting, J. Thorson, A. Collier-Oxandale, and M. Hannigan, 2017: Low-cost air quality monitoring tools: From research to practice (a workshop summary). *Sensors*, **17**, 2478, <https://doi.org/10.3390/s17112478>.
- Cohen, A. J., and Coauthors, 2017: Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the Global Burden of Diseases Study 2015. *Lancet*, **389**, 1907–1918, [https://doi.org/10.1016/S0140-6736\(17\)30505-6](https://doi.org/10.1016/S0140-6736(17)30505-6).
- Conrick, R., C. F. Mass, J. P. Boomgard-Zagrodnik, and D. Ovens, 2021: The influence of wildfire smoke on cloud microphysics during the September 2021 Pacific Northwest wildfires. *Wea. Forecasting*, **36**, 1519–1536, <https://doi.org/10.1175/WAF-D-21-0044.1>.
- EPA, 2023a: AirData Map App. Accessed 12 November 2023, <https://www.epa.gov/outdoor-air-quality-data/interactive-map-air-quality-monitors>.
- EPA, 2023b: Fire and Smoke Map. Accessed 12 November 2023, <https://fire.airnow.gov>.
- Feenstra, B., V. Papapostolou, S. Hasheminassab, H. Zhanga, B. D. Boghossiana, D. Cocker, and A. Polidori, 2019: Performance evaluation of twelve low-cost PM<sub>2.5</sub> sensors at an ambient air monitoring site. *Atmos. Environ.*, **216**, 116946, <https://doi.org/10.1016/j.atmosenv.2019.116946>.
- Gupta, P., and Coauthors, 2018: Impact of California fires on local and regional air quality: The role of a low-cost sensor network and satellite observations. *GeoHealth*, **2**, 172–181, <https://doi.org/10.1029/2018GH000136>.
- Jacobson, M. Z., 2012: *Air pollution and global warming*. Cambridge University Press, 375 pp.
- Jaffe, D. A., S. M. O'Neill, N. K. Larkin, A. L. Holder, D. L. Peterson, J. E. Halofsky, and A. G. Rappold, 2020: Wildfire and prescribed burning impacts on air quality in the United States. *J. Air Waste Manage. Assoc.*, **70**, 583–615, <https://doi.org/10.1080/10962247.2020.1749731>.
- Ling, S. H., and S. F. van Eeden, 2009: Particulate matter air pollution exposure: Role in the development and exacerbation of chronic obstructive pulmonary disease. *Int. J. Chronic Obstruct. Pulm. Dis.*, **4**, 233–243, <https://doi.org/10.2147/copd.s5098>.
- Magi, B. I., C. Cupini, J. Francis, M. Green, and C. Hauser, 2020: Evaluation of PM<sub>2.5</sub> measured in an urban setting using a low-cost optical particle counter and a federal equivalent beta attenuation monitor. *Aerosol Sci. Technol.*, **52**, 147–159, <https://doi.org/10.1080/02786826.2019.1619915>.
- NASA, 2023: EOSDIS Worldview. Date accessed 12 November 2023, <https://worldview.earthdata.nasa.gov>.
- North Carolina Department of Environmental Quality, 2023: DEQ Forecasts Code Red or Orange Air Quality for All NC on Wednesday. Accessed 12 November 2023, <https://www.deq.nc.gov/news/press-releases/2023/06/06/deq-forecasts-code-red-or-orange-air-quality-all-nc-wednesday>.
- Petty, G. W., 2008: *A First Course in Atmospheric Thermodynamics*. Sundog Publishing, 337 pp.
- State Climate Office of North Carolina, NC State University. Cardinal [data retrieval interface] available at <https://products.climate.ncsu.edu/cardinal/request>. Accessed March 04, 2024.

Wallace, L., J. Bi, W. R. Ott, J. Sarnat, and Y. Liu, 2021: Calibration of low-cost PurpleAir outdoor monitors using an improved method of calculating PM<sub>2.5</sub>. *Atmos. Environ.*, **256**, 118432, <https://doi.org/10.1016/j.atmosenv.2021.118432>.