DISSERTATION

INVESTIGATING THE ENHANCEMENT OF AIR POLLUTANT PREDICTIONS AND UNDERSTANDING AIR QUALITY DISPARITIES ACROSS RACIAL, ETHNIC, AND ECONOMIC LINES AT US PUBLIC SCHOOLS

Submitted by

Michael J. Cheeseman

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2022

Doctoral Committee:

Advisor: Jeffrey R. Pierce

Elizabeth Barnes
Emily Fischer
Bonne Ford
John Volckens

ABSTRACT

INVESTIGATING THE ENHANCEMENT OF AIR POLLUTANT PREDICTIONS AND UNDERSTANDING AIR QUALITY DISPARITIES ACROSS RACIAL, ETHNIC, AND ECONOMIC LINES AT US PUBLIC SCHOOLS

Ambient air pollution has significant health and economic impacts worldwide. Even in the most developed countries, monitoring networks often lack the spatiotemporal density to resolve air pollution gradients. Though air pollution affects the entire population, it can disproportionately affect the disadvantaged and vulnerable communities in society. Pollutants such as fine particulate matter ($PM_{2.5}$), nitrogen oxides (NO and NO_2), and ozone, which have a variety of anthropogenic and natural sources, have garnered substantial research attention over the last few decades. Over half the world and over 80% of Americans live in urban areas, and yet many cities only have one or several air quality monitors, which limits our ability to capture differences in exposure within cities and estimate the resulting health impacts. Improving sub-city air pollution estimates could improve epidemiological and health-impact studies in cities with heterogeneous distributions of $PM_{2.5}$, providing a better understanding of communities at-risk to urban air pollution. Biomass burning is a source of $PM_{2.5}$ air pollution that can impact both urban and rural areas, but quantifying the health impacts of $PM_{2.5}$ from biomass burning can be even more difficult than from urban sources. Monitoring networks generally lack the spatial density needed to capture the heterogeneity of biomass burning smoke, especially near the source fires. Due to limitations of both urban and rural monitoring networks several techniques have been developed to supplement and enhance air pollution estimates. For example, satellite aerosol optical depth (AOD) can be used to fill spatial gaps in PM monitoring networks, but AOD can be disconnected from time-resolved surface-level PM in a multitude of ways, including the limited overpass times of most satellites, daytime-only measurements, cloud cover, surface reflectivity, and lack of vertical-profile information. Observa-

tions of smoke plume height (PH) may provide constraints on the vertical distribution of smoke and its impact on surface concentrations. Low-cost sensor networks have been rapidly expanding to provide higher density air pollution monitoring. Finally, both geophysical modeling, statistical techniques such as machine learning and data mining, and combinations of all of the aforementioned datasets have been increasingly used to enhance surface observations. In this dissertation, we explore several of these different data sources and techniques for estimating air pollution and determining community exposure concentrations.

In the first chapter of this dissertation, we assess PH characteristics from the Multi-Angle Implementation of Atmospheric Correction (MAIAC) and evaluate its correlation with co-located $PM_{2.5}$ and AOD measurements. PH is generally highest over the western US. The ratio $PM_{2.5}$:AOD generally decreases with increasing PH:PBLH (planetary boundary layer height), showing that PH has the potential to refine surface $PM_{2.5}$ estimates for collections of smoke events.

Next, to estimate spatiotemporal variability in $PM_{2.5}$, we use machine learning (Random Forests; RFs) and concurrent $PM_{2.5}$ and AOD measurements from the Citizen Enabled Aerosol Measurements for Satellites (CEAMS) low-cost sensor network as well as $PM_{2.5}$ measurements from the Environmental Protection Agency's (EPA) reference monitors during wintertime in Denver, CO, USA. The RFs predicted $PM_{2.5}$ in a 5-fold cross-validation (CV) with relatively high skill (95% confidence interval R^2=0.74-0.84 for CEAMS; R^2=0.68-0.75 for EPA) though the models were aided by the spatiotemporal autocorrelation of the $PM_{2.5}$ measurements. We find that the most important predictors of $PM_{2.5}$ are factors associated with pooling of pollution in wintertime, such as low planetary boundary layer heights (PBLH), stagnant wind conditions, and, to a lesser degree, elevation. In general, spatial predictors are less important than spatiotemporal predictors because temporal variability exceeds spatial variability in our dataset. Finally, although concurrent AOD is an important predictor in our RF model for hourly $PM_{2.5}$, it does not improve model performance during our campaign period in Denver. Regardless, we find that low-cost $PM_{2.5}$ measurements incorporated into an RF model were useful in interpreting meteorological and geographic drivers of

$PM_{2.5}$ over wintertime Denver. We also explore how the RF model performance and interpretation changes based on different model configurations and data processing.

Finally, we use high resolution $PM_{2.5}$ and nitrogen dioxide (NO_2) estimates to investigate socioeconomic disparities in air quality at public schools in the contiguous US. We find that a higher proportion of Black and African American, Hispanic, and Asian or Pacific Islander students attend schools in locations where the ambient concentrations of NO_2 and $PM_{2.5}$ are above the World Health Organization's (WHO) guidelines for annual-average air quality. Specifically, we find that ~95% of students that identified as Asian or Pacific Islander, 94% of students that identified as Hispanic, and 89% of students that identified as Black or African American, attended schools in locations where the 2019 ambient concentrations were above the WHO guideline for NO_2 (10 μg m^{-3} or ~5.2 ppbv). Conversely, only 83% of students that identified as white and 82% of those that identified as Native American attended schools in 2019 where the ambient NO_2 concentrations were above the WHO guideline. Similar disparities are found in annually averaged ambient $PM_{2.5}$ across racial and ethnic groups, where students that identified as White (95%) and Native American (83%) had the smallest percentage of students above the WHO guideline (5 μg m^{-3}), compared to students that identified with other minoritized groups (98-99%). Furthermore, the disparity between white students and other minoritized groups, other than Native Americans, is larger at higher $PM_{2.5}$ concentrations. Schools with a higher percentage of students eligible for free or reduced meals, which we use as a proxy for poverty, are also associated with ambient air pollutant concentrations that exceed WHO guidelines. These disparities also tend to increase in magnitude at higher concentrations of NO_2 and $PM_{2.5}$. We investigate the intersectionality of disparities across racial/ethnic and poverty lines by quantifying the mean difference between the lowest and highest poverty schools, and the most and least white schools in each state, finding that most states have disparities above 1 ppbv of NO_2 and 0.5 μg m^{-3} of $PM_{2.5}$ across both. We also identify distinct regional patterns of disparities, highlighting differences between California, New York, and Florida. Finally, we also highlight that disparities do not only exist across an urban and non-urban divide, but also within urban areas.

ACKNOWLEDGEMENTS

This dissertation is the culmination of the most formative years of my life. I cannot express enough gratitude for my advisor Jeffrey Pierce and research scientist Bonne Ford. They have shown more patience, trust, and kindness than any graduate student could ask for. They have guided me to become an independent thinker. I would also like to thank my committee members, John Volckens, Elizabeth Barnes, and Emily Fischer. Emily taught me how to present research like a story. Elizabeth showed me that statistics could be fun. John pushed me to think about the bigger picture. Furthermore, Scott Denning, my Masters advisor, taught me the importance of engaging every person, of all varieties, in the scientific process and our understanding of the Earth. I would not have made it this far without this incredible group of scientists and people.

Before I began graduate school, I received encouragement, guidance, and camaraderie from many of my professors at Appalachian State University. Dr. Barkley Sive was the first person to invite me to conduct research during my free time in college. If it were not for him, I would likely have never come to CSU. He, more than any other, was always willing to sit down and discuss anything, from questions about research to those about life. I will always appreciate his guidance and support. I would also thank Dr. Scott Marshall, who taught me how to code and awakened the puzzle solver in my brain. Finally, Dr.'s Carla Penders and Chris Thaxton two of Appalachian States top professors, were unceasing in their support and kindness towards me. I would have burned out on school long ago were it not for them. It was Carla who taught me that I am not my grades, my research, or my list of accolades, and that I must always balance these pursuits with other things that bring me joy.

I also cannot express enough gratitude for the CSU Atmospheric Science community. The front office staff here at CSU-Atmos is especially impactful. I can recall at least three occasions where I was freaking out until Sarah Tisdale quickly dispatched with my concerns and helped me move forward. The Pierce group, past and present, helped me become a better scientist and communicator. Ali always laughed with me, often times at nonsense. The greater Atmos community

such as Kyle, Kate, Jack, Steve, Marie, Bryn, Drea, Tim, Rachel, Brad, and many, many others, are always supporting each other, planning trips, and drinking good Colorado beers. Clarissa put up with my stress and loved me anyway. My climbing, church, and soccer friends made all the good times better and the stressful times bearable. Finally, I would like to thank my family, who's support provided me with distractions, strength, and love in equal measure. I could not have finished graduate school without the support, patience, and kindness of every person I have listed here and more. Thank you.

DEDICATION

This dissertation is dedicated to my grandmother Audrey Ross, or "Gragra", whose unconditional love, kindness, and selflessness has inspired me every day of my adult life. Thank you Gragra, for everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 An Overview of Air Pollution in Society

Air pollution has long been a plague on the World's societies, and one exponentially heightened as an unforeseen consequence of industrialization and urbanization. The impacts of air pollution have been noted by many over the course of human history: smoke plumes from metallurgists, pre-industrial workshops, and homes in ancient Rome were described as "heavy heavens" and "infamous air" (Hughes, 1996). The problems became more apparent and ubiquitous when societies started burning fossil fuels instead of wood. Lamenting the state of London's polluted atmosphere, due to the increasing use of coal for residential heating, poet John Evelyn wrote in 1661:

> "... that Hellish and dismall Cloud of SEA-COAL ... perpetually imminent over her head ... mixed with the otherwise wholesome and excellent Aer, that her Inhabitants breathe nothing but an impure and thick Mist accompanied with a fuliginous and filthy vapour, which renders them obnoxious to a thousand inconveniences, corrupting the Lungs, and disordering the entire habits of their Bodies; so that Catharrs, Phthisicks, Coughs and Consumptions rage more in this one City than in the whole Earth besides." (Evelyn, 1976)

Despite the long history of intersection between human health and air pollution, it remains the greatest environmental hazard for human health globally (Cohen et al., 2017; World Health Organization, 2013). In 2019, ambient fine particulate matter ($PM_{2.5}$) caused an estimated 4.2 million premature deaths worldwide (Murray et al., 2020). Exposure to $PM_{2.5}$, as well as to other toxins such as nitrogen oxides (NO_x) and ozone (O_3) are linked to a host of cardiovascular and respiratory diseases (Atkinson et al., 2018; Cohen et al., 2017; Faustini et al., 2014; Hoek et al., 2013; World Health Organization, 2013). Air pollution has also been linked to decreased cognitive function, lower intelligence quotient (IQ) scores, and, most recently, to increased rates of depression and anxiety (Myhre et al., 2018; Roberts et al., 2019).

The health impacts of air pollution also result in huge economic losses worldwide. A study over Europe found that a 1 μg m^{-3} increase in $PM_{2.5}$ concentrations causes a 0.8% reduction in real gross domestic product (GDP) (Dechezleprêtre et al., 2019), mostly due to increased worker absenteeism and work productivity. Similarly, the World Bank estimated that North America loses an estimated 2.4% GDP equivalent to welfare losses due to ambient $PM_{2.5}$ (World Bank, 2016). In 2013, air pollution cost the world economy an estimated $5.11 trillion in welfare losses and an additional $225 billion in labor income losses (World Bank, 2016).

While air pollution impacts our entire society, the greatest burden of air pollution in the US often disproportionately impacts groups that are already most at risk. Studies investigating the intersection between air pollution and race/ethnicity as well as socioeconomic status, find that racial/ethnically minoritized and impoverished people are more likely to be near emission sources (Chakraborty et al., 2011; Maantay, 2007; Maantay et al., 2009; Mohai & Saha, 2007; Pastor et al., 2004). Past studies have found that racially/ethnically minoritized groups in the US had disproportionate exposures to atmospheric neurotoxins (Grineski & Collins, 2018), $PM_{2.5}$ (Bell & Ebisu, 2012; O'Neill et al., 2003; Woo et al., 2019), and NO_2 (Woo et al., 2019). Residential segregation is a crucial factor leading to disparities in air pollutant exposure: Woo et al. (2019) found that segregating neighborhoods not only serves to keep minoritized people in separate residential spaces but also keeps them in closer proximity to environmental pollution sources. Studies have also found disproportionate health impacts in communities with larger proportions of people of color such as increased $PM_{2.5}$-attributable disease burdens (Castillo et al., 2021; Hajat et al., 2015; Tessum et al., 2019). Populations may also have different health responses to the same levels of exposure to various air pollutants (Bell & Dominici, 2008; Bell & Ebisu, 2012; Grineski et al., 2010; Zanobetti & Schwartz, 2000), which is an open area of research. Finally, some communities have less access to healthcare and are less able to mitigate pollution impacts.

Fortunately, due to regulations in the US, anthropogenic emissions of most primary air pollutants have been declining for decades (e.g., Anenberg et al., 2021; Hammer et al., 2020), which may serve to lessen the burden of air pollution on all communities, including those most at-risk.

However, anthropogenic emissions are not the only important emission sources. Landscape fires are a major source of both $PM_{2.5}$ (~40% of primary emissions in the US) (EPA, 2017), and volatile organic compounds (VOCs), including many hazardous air pollutants (HAPs) (Akagi et al., 2011). Furthermore, due to a warming climate, the western US has become hotter and drier, and wildfires have been increasing in frequency and intensity in much of the western US (Westerling, 2006; Westerling et al., 2003). As a result, landscape fires and their contribution to US $PM_{2.5}$ has been increasing in the western US and are likely to continue increasing throughout the Contiguous US (CONUS) as the climate changes (Brey et al., 2021; Ford et al., 2018; J. C. Liu et al., 2016; O'Dell et al., 2019). It is unclear, however, whether there are racial/ethnic and socioeconomic disparities in the impact of landscape fires, and if so, how these disparities may change as the climate continues to warm.

1.2 The Current State of Air Quality Monitoring

Measuring people's exposure to air pollution from anthropogenic/natural sources, understanding their impacts, and assessing the resulting social disparities, are not possible without sufficient monitoring of air pollutants. Air quality monitoring, however, is a relatively recent phenomena. After tragic events such as the lethal haze that enveloped Donora, PA (1948) and the frequent episodes of smog that hit heavily trafficked Los Angeles, CA (beginning in the 1940s), the US federal and state governments were pushed into action, culminating in the establishment of the environmental protection agency (EPA) and passing of the Clean Air Act (1970). The Clean Air Act and its amendments require that the EPA establish regulatory monitoring networks to capture concentrations of six criteria pollutants, including $PM_{2.5}$ and NO_2, across the US. Though these regulatory networks have provided invaluable information on air pollution over several decades, they often have large spatial gaps, especially in sparsely populated areas of the US. These spatial gaps make capturing the complex and heterogeneous nature of air pollution difficult and can limit our ability to regulate emission sources, evaluate regulatory models, and accurately estimate the health impacts of air pollutants (Bi et al., 2019; Gao et al., 2015; Just et al., 2015; Wang & Oliver

Gao, 2011). Several methods have been developed to fill in the gaps of regulatory networks and increase our understanding of air pollution including: (1) satellite observations; (2) low-cost sensor networks; (3) chemical transport models; and (4) enhancing observations using data mining, data fusion, machine learning, and other statistical techniques.

Satellite observations of aerosol optical depth (AOD), which are estimates of column-integrated light extinction due to aerosols, are commonly used to estimate $PM_{2.5}$ concentrations (e.g. Hu et al., 2014; van Donkelaar et al., 2006, 2010, 2011). AOD observations, however, have a complex relationship with surface $PM_{2.5}$ measurements that depend on the aerosol size-distribution, hygroscopicity, vertical profile, and optical properties. These factors, especially the vertical profile of aerosols can be especially important when interpreting AOD observations from landscape fire smoke plumes. The height of smoke plumes can vary dramatically due to fire intensity and local meteorology (Freitas et al., 2006; Vernon et al., 2018), which can result in very different surface concentrations of aerosols. Furthermore, AOD measurements have often been at resolutions too coarse (e.g. > 10 km) to properly capture the heterogeneity of air pollutant sources such as landscape fires (Lyapustin et al., 2019). Although recent satellite products such as high resolution (1 km) AOD and smoke plume height observations from the Multi-Angle Implementation of Atmospheric Correction (MAIAC) data product may help remedy these issues (Lyapustin et al., 2018, 2019).

Due to the complex nature of the AOD:$PM_{2.5}$ relationship, AOD measurements are often translated to $PM_{2.5}$ estimates by combining satellite observations with a variety of modeling and statistical approaches. One common method is to use a global chemical transport model to simulate the $PM_{2.5}$:AOD relationship globally (or regionally) and then multiply satellite observations of AOD by the modeled ratio, which results in satellite-derived $PM_{2.5}$ estimates (Y. Liu et al., 2004; van Donkelaar et al., 2010, 2011, 2015). However, the $PM_{2.5}$:AOD relationship is still poorly understood and requires further validation from ground based sensors. For that reason, the Surface PARTiculate mAtter Network (SPARTAN) was developed to provide co-located, high-quality AOD and $PM_{2.5}$ measurements in different regions across the globe (Snider et al., 2015, 2016).

Unfortunately, networks such as SPARTAN, traditional regulatory networks, and satellites all have high costs associated with their development, implementation, and maintenance.

Both researchers and private businesses have investigated cheaper alternatives to the traditional, high-cost air pollution measurement methods. Thus, low-cost sensor (LCS) networks of air quality monitors have become increasingly used in conjunction with regulatory measurements at cheaper costs. For example, the PurpleAir network, which uses low-cost light-scattering sensors to capture sub-hourly $PM_{2.5}$ measurements, has been growing rapidly over the past few years (Delp & Singer, 2020; Krebs et al., 2021). LCS networks have successfully been used in identifying pollution hotspots and sources (Gao et al., 2015; Rickenbacker et al., 2019; Zikova et al., 2017) and targeting pollution reduction targets (Gillooly et al., 2019). These low-cost networks also have the added benefit of being easier to deploy and maintain than regulatory networks, which allows citizen scientists to get involved, take measurements, and potentially learn more about local air quality issues (e.g., Ford et al., 2019; Gupta et al., 2018). Despite the benefits of LCS networks, they also have moderate-to-high uncertainties due to the lower quality of their sensors (Gupta et al., 2018; Snyder et al., 2013). For example, many LCS networks like PurpleAir that measure $PM_{2.5}$ use light-scattering sensors, which are sensitive to variations in particle size distributions, relative humidity (RH), and aerosol refractive index (Austin et al., 2015; Levy Zamora et al., 2019; Singer & Delp, 2018; Sousan et al., 2017; Tryner et al., 2020; Y. Wang et al., 2015). In particular, variations in RH have been shown to result in large biases in measurements of $PM_{2.5}$ from these low-cost sensors (Jayaratne et al., 2018; Levy Zamora et al., 2019; Magi et al., 2020; Malings et al., 2020; Tryner et al., 2020).

Another burgeoning field in air quality research is the use of data mining and machine learning techniques (ML) to supplement regulatory monitoring networks, often through the use of multiple additional data sources such as satellite observations, meteorological variables, and geographic information (e.g., Bellinger et al., 2017; Di et al., 2016; Lightstone et al., 2017; Y. Liu et al., 2018; Reid et al., 2015; Suleiman et al., 2019; Xi et al., 2015). Data mining includes several distinct statistical tools that are often used to discern patterns in large datasets, sometimes in an unsuper-

vised manner, and then use those patterns to create predictions of new or unknown events. ML, a subgroup of data mining, can create predictor functions from high-dimensional datasets, oftentimes creating accurate predictions where traditional statistical models may fail due to underlying assumptions (Bellinger et al., 2017). Since ML methods can recognize patterns in complex and large datasets efficiently without making assumptions about the data, they have the potential to aid interpretations of the drivers of air pollutants and develop highly skilled predictive models (e.g. Di et al., 2016; Lightstone et al., 2017; Liu et al., 2018; Reid et al., 2015; Suleiman et al., 2019; Xi et al., 2015). ML methods have already been shown to be more successful at predicting $PM_{2.5}$ than traditional chemical transport models in some cases (Lightstone et al., 2017; Xi et al., 2015) and are generally easier to implement. Techniques such as data mining and ML will become increasingly useful as future satellite missions, LCS networks, and even, perhaps, regulatory monitors continue to produce ever-larger and more complex datasets.

1.3 Scope of Dissertation

In this dissertation, we conduct quantitative assessments of $PM_{2.5}$ and its relationship with: 1) smoke-plume heights and satellite AOD over the Western US; 2) low-cost sensor AOD and meteorological/geographical factors over wintertime Denver; and 3) poverty and racial/ethnic demographics at all US public schools (in conjunction with NO_2 concentrations). We do these assessments with the goals of: 1) preparing for future investigations in landscape fires and their impacts on human health; 2) providing guidance for future use of ML, especially random forests (RF), in air quality prediction over urban areas; and 3) understanding how the legacy of racial/ethnic and economic bias has impacted the air quality of our nation's children. The following paragraphs describe the direction of each chapter.

Chapter 2 is a research article that was published in Geophysical Research Letters (Cheeseman et al., 2020). For this paper, we co-located high resolution (1 km) MAIAC smoke plume heights and AOD with surface $PM_{2.5}$ observations and planetary boundary layer heights. We quantified the relationship between AOD and plume height to test the hypothesis that optically thicker smoke

plumes may be more buoyant. We displayed regional patterns of plume height observations and their relationship with co-located boundary layer heights. Finally, we quantified the $PM_{2.5}$:AOD relationship when smoke plumes either remain in the boundary layer or escape into the free troposphere.

Chapter 3 is a research article under review at Atmospheric Chemistry and Physics (Cheeseman et al., 2021). We analyzed measurements taken from a citizen-science field campaign (Citizen Enabled Aerosol Measurements for Satellites [CEAMS]) that used low-cost monitors deployed from Nov 14th, 2019 to January 20th, 2020 in Denver, Colorado. We quantified the relationship between coincident low-cost sub-hourly measurements of $PM_{2.5}$ and AOD, as well as meteorological and geographical variables using a RF ML framework. We assessed the added predictability of intra-city $PM_{2.5}$ based on whether AOD is used as a predictor. Finally, we describe our ML methods in detail and discuss how they may be misused in future studies.

Chapter 4 is a letter-length manuscript in preparation for a journal yet to be determined. We used high-resolution annually averaged $PM_{2.5}$ (1 km) and NO_2 (~2.8 km) estimates to quantify disparities in air pollutants across residential categories (urban, suburban, town, rural), racial/ethnic demographics, and poverty levels at US public school locations. We quantified disparities for each state across poverty and racial/ethnic lines. Additionally we evaluated distinct regional patterns of disparities and investigated the intersectional nature of residential classification, poverty, and racial/ethnic demographics as they pertain to air quality at school. We also highlight intra-urban disparities in the Bay Area of California. Finally, in Chapter 5, we provide a summary of the conclusions from the aforementioned chapters. We also discuss the implications of this dissertation and how it may influence future work.

Chapter 2

The Relationship Between MAIAC Smoke Plume Heights and Surface PM [1]

2.1 Introduction

The Global Burden of Disease (GBD) study estimates that human exposure to $PM_{2.5}$ (particles with diameters smaller than 2.5 μm) is the fifth leading risk factor for premature mortality worldwide (healthdata.org). $PM_{2.5}$ consists of particles small enough to penetrate deeply into the lungs, with prolonged exposure resulting in respiratory and cardiovascular diseases (Barregard et al., 2006; Gan et al., 2017; J. C. Liu et al., 2015). Biomass burning (defined here as wild, prescribed, and agricultural fires) is a major source of $PM_{2.5}$ (e.g., Akagi et al., 2011; Koss et al., 2018). Wildfires have been increasing in severity and frequency in the western United States (Dennison et al., 2014; Westerling, 2006) impacting $PM_{2.5}$ (McClure & Jaffe, 2018; O'Dell et al., 2019), and this trend is predicted to continue (Ford et al., 2018; Spracklen et al., 2009; Yue et al., 2013). Therefore, the health impacts of biomass burning smoke are important and may become increasingly so.

Exposure to smoke $PM_{2.5}$ has been linked to respiratory health impacts (Gan et al., 2017; J. C. Liu et al., 2015; Reid et al., 2016), higher all-cause mortality rates (J. C. Liu et al., 2015; Reid et al., 2016), higher hospitalization rates (Gan et al., 2017; J. C. Liu et al., 2017), emergency room admissions (Wettstein et al., 2018), and low birth weights (Holstius et al., 2012). However, quantifying the extent and variety of health impacts due to biomass burning smoke is challenging due to the episodic nature of these events and our inability to estimate exposure accurately (J. C. Liu et al., 2015).

[1] Cheeseman, M., Ford, B., Volckens, J., Lyapustin, A., & Pierce, J. R. (2020). The Relationship Between MAIAC Smoke Plume Heights and Surface PM. *Geophysical Research Letters*, 47(17), e2020GL088949. https://doi.org/10.1029/2020GL088949

Previous estimates of smoke $PM_{2.5}$ exposure in the United States tend to rely on ground-based monitoring networks (J. C. Liu et al., 2015). Unfortunately, surface monitors, though accurate, are sparsely distributed in many regions of the country, making it difficult to capture the heterogeneous nature of smoke plumes, especially in rural/remote areas (Lassman et al., 2017). Furthermore, monitoring data often cannot directly distinguish biomass burning smoke from other $PM_{2.5}$ sources. Alternatively, chemical transport models (CTMs) can produce spatially continuous $PM_{2.5}$ estimates that can be attributed to specific sources. However, CTMs struggle simulating biomass burning events, particularly smoke heights (Baker et al., 2016; Lassman et al., 2017; Paugam et al., 2016; Val Martin et al., 2012). Finally, satellite retrievals of aerosol optical depth (AOD), which measures light extinction due to aerosols in an atmospheric column, have become increasingly important for $PM_{2.5}$ exposure estimates (Cohen et al., 2017; Jerrett et al., 2017). By providing near-global coverage of clear-sky regions every day, satellite retrievals of AOD can fill the spatial gaps within monitoring networks. However, most satellite AOD products are from polar-orbiting satellites, which only provide retrievals during specific daytime overpass times. Furthermore, as a column-integrated value, AOD products do not provide information on aerosol loading at the surface.

A common technique for translating AOD to $PM_{2.5}$ involves establishing a relationship between co-located $PM_{2.5}$ and AOD (i.e., the ratio $PM_{2.5}$:AOD) using either ground-based monitoring networks (Engel-Cox et al., 2004; Y. Liu et al., 2007; Wang, 2003) or chemical transport models (van Donkelaar et al., 2006, 2010). The $PM_{2.5}$:AOD values are then applied to satellite-retrieved AOD as a scaling factor to estimate surface $PM_{2.5}$ (van Donkelaar et al., 2010). However, a crucial source of uncertainty in estimating $PM_{2.5}$ from AOD is the vertical profile of aerosol loading in the atmosphere (Snider et al., 2015; van Donkelaar et al., 2010). This profile is important during biomass burning events because the vertical profile of smoke depends on the altitude reached by the plume (i.e., the plume height, PH) (Vernon et al., 2018). For example, if a smoke plume entirely escapes the boundary layer into the free troposphere (i.e., the PH is higher than the PBLH) the $PM_{2.5}$:AOD will be small because the surface concentration is low relative to the column total.

Thus, the common techniques for estimating $PM_{2.5}$:AOD may fail during smoke events because regionally averaged $PM_{2.5}$:AOD from monitors may not be representative of the smoke event, and chemical transport models often struggle with predicting smoke PHs (Freitas et al., 2006). In fact, many models prescribe PHs (e.g., all emissions at the surface or within the boundary layer such as the base model described by Zhu et al., 2018). Therefore, accurate estimates of smoke PH could improve predictions of $PM_{2.5}$ concentrations by constraining $PM_{2.5}$:AOD estimates during smoke events.

Space-borne PH retrievals are currently available from three sources: (1) the Cloud-Aerosol LiDAR with Orthogonal Polarization (CALIOP) (Winker et al., 2009); (2) the Multi-angle Imaging SpectroRadiometer (MISR) (Kahn et al., 2007; Nelson et al., 2013); and, more recently, (3) the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm using the Moderate Resolution Imaging Spectroradiometer (MODIS) instruments aboard both Terra and Aqua satellites (Lyapustin et al., 2019). Space-borne lidars, such as CALIOP, can capture aerosol layer heights with high precision but have poor spatial sampling. Space-borne stereo imaging from MISR has greater horizontal sampling than CALIOP but has global coverage only every 9 days. Therefore, we use PH from the MAIAC algorithm to investigate the relationship between PH and $PM_{2.5}$:AOD ratios. Since MODIS is on both Terra and Aqua satellites, and swaths from different overpasses can overlap, MODIS can observe a location up to four times a day in the midlatitudes. Lyapustin et al. (2019) found that MAIAC and MISR PH products were in good agreement: 60% of MAIAC PH were within 500 m of the MISR PH but ~450 m lower on average. Similar statistics were found in the comparison of MAIAC and CALIOP PH retrievals, namely a 216 m low bias in MAIAC (Lyapustin et al., 2019). The low bias compared to MISR and CALIOP is to be expected because MISR and CALIOP PH retrievals tend to be associated with plume tops, while MAIAC estimates the effective PH using a thermal technique that is reliant on 11 μm absorption of emitted gases entrained in the plume. Therefore, MAIAC retrievals may be associated with altitudes closer to plume tops only for plumes that are very optically thick at 11 μm. Due to MAIAC's use of 11 μm absorption optical thickness, we hypothesize that there may be a relationship between MAIAC

PH and the extinction AOD (if AOD is correlated with the absorption optical depth at 11 μm). However, PH and AOD may also be positively correlated due to a physical mechanism: Smoke plumes from extreme wildfire events generally have higher emissions and greater buoyancy than smaller, optically thinner, plumes, which may cause a positive correlation between PHs and AOD values.

In this study, we present regional characteristics of PH from MAIAC over the United States, and we investigate (1) the spatial relationship between PH and PBLH to estimate the frequency that smoke may be lofted above the boundary layer, (2) the correlation between PH and AOD, and (3) the relationship between $PM_{2.5}$:AOD and PH:PBLH over the western United States.

2.2 Data and Methods

For all data sources described below, we gathered data over the contiguous United States during the months of July to September over a 9 year period (2010-2018). We limited our analysis to these 3 months each year because they are generally the most active wildfire months in the western United States (Brey et al., 2018). When analyzing the relationship between $PM_{2.5}$:AOD and PH:PBLH, we limit our analysis to the western United States, which we define as the contiguous United States west of -103° longitude.

We used AOD and PH data from the daily MAIAC atmospheric product MCD19A2, which is part of the MODIS Collection 6 record (Lyapustin et al., 2018, 2019; Lyapustin & Wang, 2018). It should be noted that the PH data in MAIAC are labeled as "plume injection height." However, as the PH may change after injection but still be measured by MAIAC, we refer to the MAIAC smoke height product as "plume height." It should also be noted that each MAIAC retrieval only provides a single PH altitude, regardless of the plume vertical depth. MODIS has global coverage every 1–2 days at ~10:30 local time (Terra) and ~13:30 local time (Aqua). We use both Terra and Aqua orbits separately to test how the different overpass times (morning and afternoon, respectively) affect our analysis. Smoke plumes have high concentrations of gases that absorb 11 μm radiation such as ammonia, nitric acid, and some VOCs. For plumes with high enough gas concentration, the smoke

pixels have a colder brightness temperature (T_b) at 11 μm relative to nearby smoke-free pixels. MAIAC uses the thermal contrast between detected smoke and neighboring smoke-free pixels (ΔT_b) to estimate PH assuming a lapse rate. Lyapustin et al. (2019) found that three conditions are necessary to have confidence in PH retrievals: (1) MAIAC detected smoke (Lyapustin et al., 2012a, b) and the total AOD in the 0.47 μm channel is high enough ($\geq$0.8) to ensure sufficient gaseous absorption at 11 μm; (2) there must be enough nearby smoke/cloud-free land surface pixels available to determine the thermal contrast (ΔT_b); and (3) the contrast is negative ($\Delta T_b < 0$), that is, the smoke pixel is "colder" than the background. The current version of MAIAC assumes a fixed 6.5° km^{-1} lapse rate to convert from brightness temperature to PH, which could introduce errors. The next public version of MAIAC will use reanalysis meteorology to achieve more realistic spatiotemporal variability in lapse rates. We used the 0.55 μm total AOD in the MCD19A2 product. Finally, Lyapustin et al. (2019) recommends that MAIAC PH retrievals only be used within a buffer distance of 75-150 km away from the detected fire hotspots (reported in MAIAC QA) because of poor observed agreement between MISR and MAIAC over transported smoke. This poor agreement is likely due, in part, to the dilution and short chemical lifetimes of gases that absorb 11 μm. Thus, we repeat our analysis using MAIAC retrievals only within 150 km of detected fires and find similar qualitative results, but a 67% reduction in data.

We used *in situ* $PM_{2.5}$ measurements from the Environmental Protection Agency's (EPA) Air Quality System (AQS) monitoring network (https://aqs.epa.gov/aqsweb/airdata.html). The AQS network includes federal, state, and tribal monitors in both rural and urban areas. Measurements were taken from sites using the federal reference method (FRM) (EPA parameter code 88101) and acceptable non-FRM methods (EPA parameter code 88502) in the western United States (as defined above), and we used 24 hr averaged data.

Planetary boundary layer heights (PBLH) from the National Center for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR) product (Mesinger et al., 2006) were used in our analysis. The NCEP-NARR is an extension of NCEP's Global Reanalysis and assimilates precipitation and weather observations, land surface models, and other supporting in-

formation in order to predict a suite of atmospheric variables including PBLH. The PBLH is output eight times daily (3-hourly) as instantaneous values with a 32 km grid resolution.

In order to observe the relationship between PH:PBLH and $PM_{2.5}$:AOD, we matched the PH, AOD, $PM_{2.5}$, and PBLH data sets in space and time. The PBLH was linearly interpolated spatially from the NCEP grid (32 km) to the MAIAC grid resolution (1 km) and temporally interpolated to satellite overpasses. Next, we colocated the MAIAC and NCEP data to the EPA-AQS sites by averaging all PH, AOD, and PBLH data within a 10 km radius from each monitor (results were similar when using a 5 km radius but with less data included). We used the averages of the AOD, PH, and PBLH data within 10 km of each EPA site independently, even if there is overlap between data selected for each EPA monitor (i.e., if monitors are close together). Finally, we found the ratio of colocated PH:PBLH and $PM_{2.5}$:AOD around each EPA site. It should be noted that the variability in the AOD during smoke events was low relative to the variability in $PM_{2.5}$, and thus $PM_{2.5}$ described most of the variability in the $PM_{2.5}$:AOD values used in our analysis.

2.3 Results

We show an example of the different data sets used in this study in Figure 2.1, which displays PH and AOD retrievals, $PM_{2.5}$ concentrations, and PBLH for 18 August 2016 in central California. There were three major fires in the region on this day: the Cedar, Soberanes, and Chimney fires (https://fire.ca.gov/) (Figure 2.1a). There is a strong gradient in AOD between the plumes and background air (Figure 2.1b). This scene displays the potential usefulness of satellite-retrieved PH. For example, the Cedar fire shows a clear smoke plume with high AOD (Figure 2.1b) but relatively low PH (Figure 2.1c) and PH:PBLH (Figure 2.1d) values. This indicates that the smoke plume is remaining within the boundary layer; thus, we would expect high $PM_{2.5}$ concentrations downwind of the fire as confirmed by enhancements in $PM_{2.5}$ measured at the downwind monitor. The Soberanes fire, on the other hand, shows relatively high PH and PH:PBLH and the downwind monitors of this fire, though clearly below a layer of smoke indicated by the AOD, do not show large $PM_{2.5}$ enhancements. These examples indicate that satellite-retrieved PH may be indicative

of the $PM_{2.5}$:AOD relationship for smoke plumes. It should be emphasized that the $PM_{2.5}$ measurements shown in Figure 2.1 are 24 hr averages, which means the reported $PM_{2.5}$ may not be representative of the aerosol concentrations during the satellite overpass time. The 24 hr $PM_{2.5}$ averaging could explain why some monitors indicate high $PM_{2.5}$ concentrations (~15-20 μg m^{-3}) downwind of the Chimney fire even though the AOD is relatively low (~0.2) (Figure 2.1b).

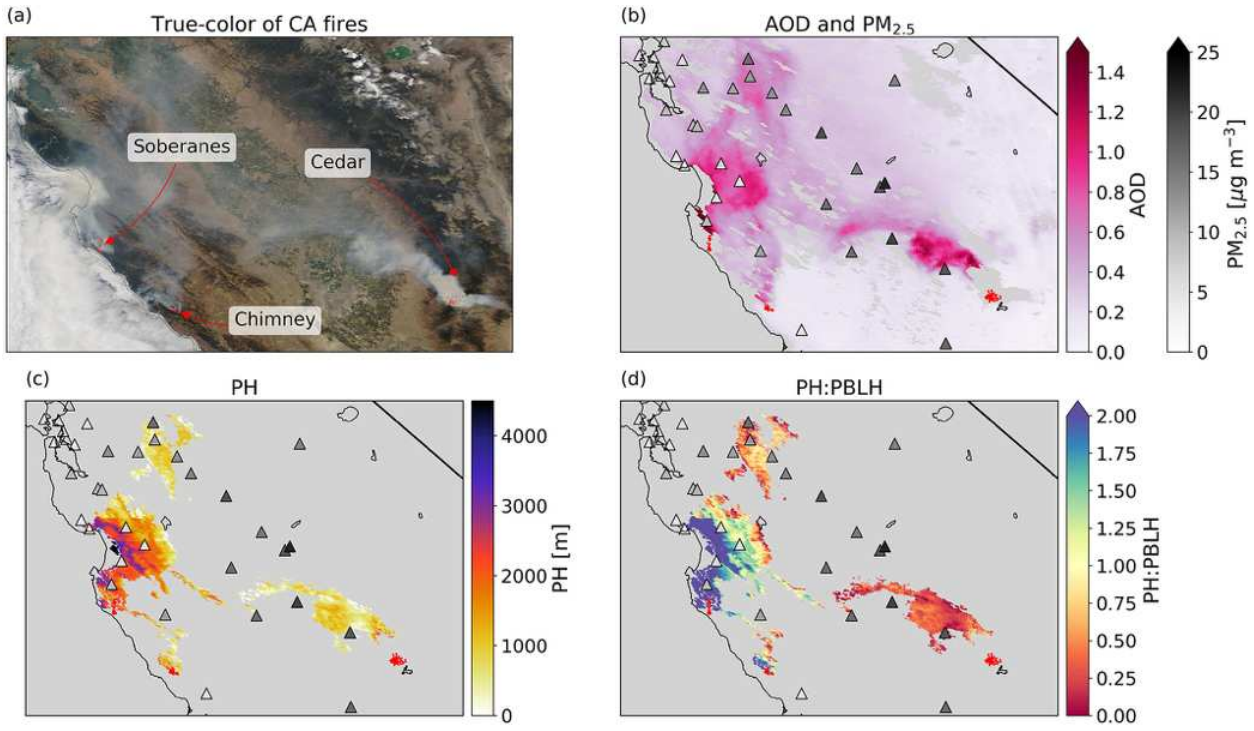

Figure 2.1: (a) MODIS Terra true-color image of multiple fires in California on 18 August 2016 from NASA Worldview (https://worldview.earthdata.nasa.gov/). The red dots indicate MODIS thermal anomalies from worldview. (b) Terra MODIS-MAIAC AOD measurements, 24 hr mean $PM_{2.5}$ from EPA-AQS monitors (filled triangles), and Terra MODIS thermal anomalies (red dots). (c) Terra MODIS-MAIAC PH retrievals and repeated 24 hr mean $PM_{2.5}$ and thermal anomalies. (d) The ratio of MODIS-MAIAC PH and colocated PBLH and repeated 24 hr mean $PM_{2.5}$ and thermal anomalies.

Figure 2.2 shows large regional differences in the number of days with PH retrievals and their relationship to colocated PBLH for July to September across 2010–2018. For this time period, the MAIAC algorithm identified an order of magnitude more days with PH retrievals on the West Coast of the United States and northern Rockies than the East Coast for both Aqua and Terra retrievals (Figures 2.2a and 2.2b). The West Coast and a few states in the eastern United States generally have

the highest percentages of PH:PBLH that are greater than 1.2 (i.e., PH $\geq$ PBLH * 1.2), which we use as a proxy here to indicate plumes that may have escaped the boundary layer (the sensitivity of this ratio for determining $PM_{2.5}$:AOD is discussed later). We repeated this analysis while including PH retrievals greater than 150 km away from MAIAC thermal hotspots (Figures A.3–A.5 in the supporting information) and found similar regional patterns. We suspect that the high percentage of smoke plumes exceeding our PH:PBLH $>$ 1.2 criteria in some eastern states may be due to free-tropospheric long-range transport of smoke from the western United States or Canada. See the supporting information for examples of long-range transport of smoke with PH:PBLH $>$ 1.2 (Figures A.6 and A.7). These states have significantly fewer PH retrievals (Figure A.2); thus, a small number of long-range transport events could bias the average PH high in these states.

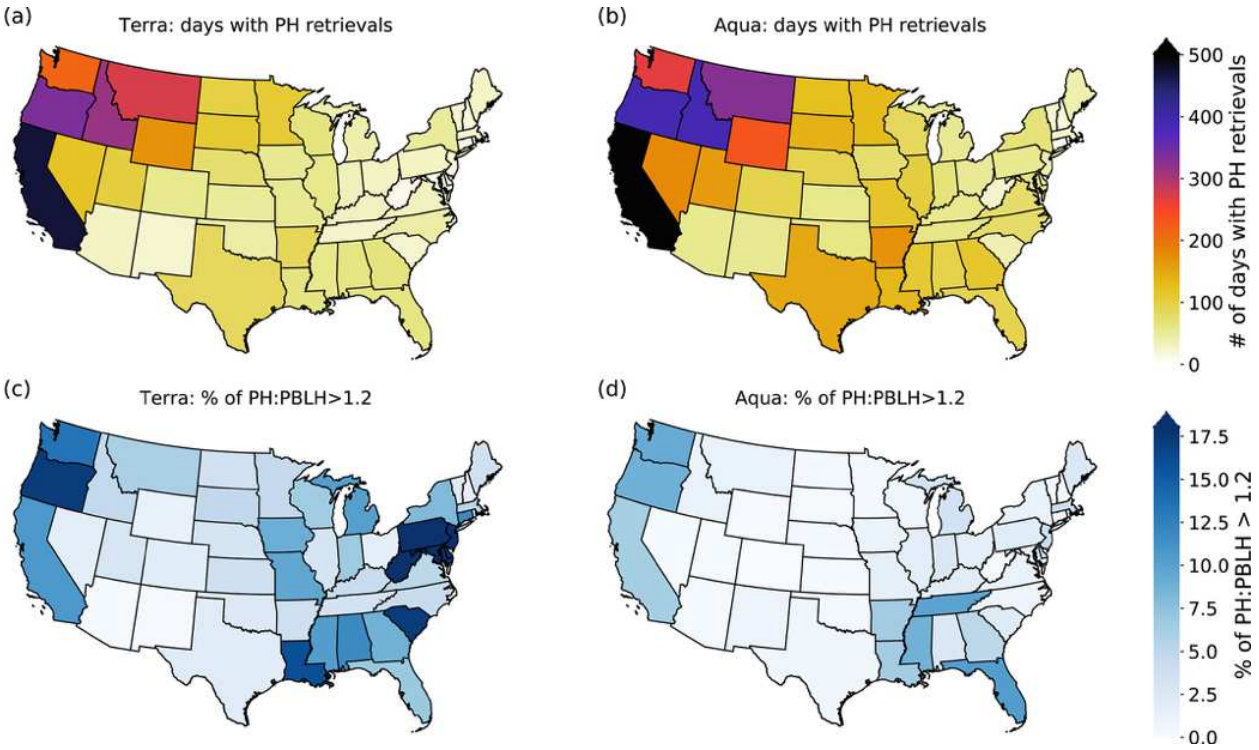

Figure 2.2: The number of days with MAIAC PH retrievals during the study period in each state for Terra (a) and Aqua (b), and the percentage of PH retrievals that are greater than PBLH by a factor of 1.2 from Terra (c) and Aqua (d).

In addition to analyzing regional characteristics of PH, we investigated the relationship between PH and AOD. We hypothesized there may be a relationship between MAIAC PH and extinction

AOD for two reasons: (1) larger, more optically thick plumes, also tend to be more buoyant, and thus optical thickness may be correlated with PH; and (2) the MAIAC thermal technique's reliance on absorption optical thickness may lead to a positive correlation between PH and the AOD of the plume. We examine this hypothesis by segregating PH values into AOD bins (Figure 2.3) and calculating a 25th, 50th, and 75th quartile for each AOD bin. All of the 25th, 50th, and 75th quartiles of PH distributions (calculated for each AOD bin) increase with increasing AOD (Figure 2.3), and the correlation between mean AOD and mean PH for each bin is high (R^2 = 0.97, slope = 527). However, in the absence of binning and averaging, the correlation of colocated PH and AOD is low (R^2 = 0.09, slope = 546 m) (Figure A.8). Though there is clearly a relationship between PH and AOD, the relative contributions of (1) the physical "buoyancy" mechanism and (2) the relationship between absorption optical depth at 11 μm and AOD are still unknown. However, we suspect that the physical buoyancy mechanism may be the leading contributor because we repeated this analysis using data only within 150 km of detected fires (Figure A.9), which is the distance within which MAIAC PH showed good agreement with MISR and CALIOP (Lyapustin et al., 2019), and found similar results with and without the buffer.

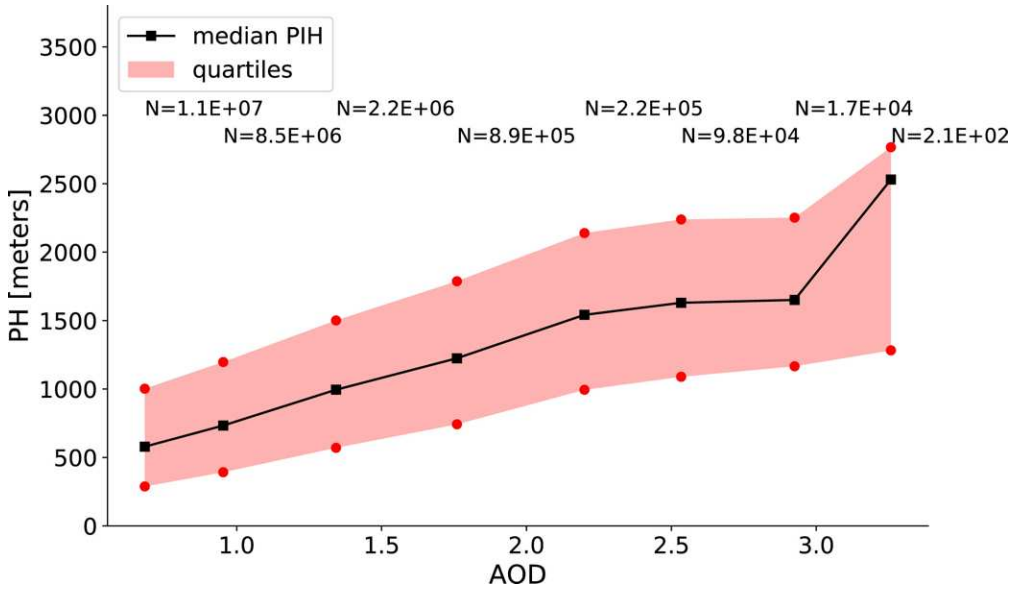

Figure 2.3: The 25th, 50th, and 75th quartiles of PH of colocated PH and AOD retrievals for binned AOD values. The number of PH retrievals for each AOD bin are shown at the top.

As discussed in section 1, $PM_{2.5}$:AOD is used to translate satellite AOD into $PM_{2.5}$ estimates. However, the $PM_{2.5}$:AOD value is especially uncertain during smoke events because smoke may be injected into the boundary layer or the free troposphere (Lassman et al., 2017). Thus, we hypothesized that PH could be used to constrain $PM_{2.5}$:AOD during smoke events because smoke plumes injected entirely into the free troposphere would cause low $PM_{2.5}$:AOD ratios downwind of the fire because the smoke would lead to high AOD values but smaller $PM_{2.5}$ enhancements. Alternatively, if smoke remains trapped in the boundary layer, $PM_{2.5}$ would be enhanced, leading to relatively higher $PM_{2.5}$:AOD values. To test this hypothesis, we assess the relationship between PH:PBLH and $PM_{2.5}$:AOD using MODIS-MAIAC Terra (Figure 2.4a) and Aqua (Figure 2.4b). We find that the 25th, 50th, and 75th quartiles of $PM_{2.5}$:AOD decrease monotonically (other than the highest PH:PBLH bin for Aqua and the central bin for Terra) with increasing PH:PBLH bins. Note that the bins are spaced unevenly due to the lower number of high PH:PBLH values colocated with PM monitors. The variability of $PM_{2.5}$:AOD is large within most PH:PBLH bins less than 2. The high variability in $PM_{2.5}$:AOD is likely because (1) high PH:PBLH does not preclude high $PM_{2.5}$ concentrations and (2) monitor 24 hr averages may not be representative of smoke concentrations at satellite overpass times. Linear-regression models were fit to all colocated PH:PBLH versus $PM_{2.5}$:AOD data points (Figure A.10) and binned means of PH:PBLH and $PM_{2.5}$:AOD (Figure 2.4) for the combined Aqua and Terra record. The linear models were created as follows:

$$\mathbf{PH:PBLH} = \mathbf{m} \times [\mathbf{PM_{2.5}:AOD}] + \mathbf{b}, \tag{2.1}$$

The slopes ($\mathbf{m}$) were -9 μg m^{-3} and the intercepts ($\mathbf{b}$) were 48 μg m^{-3} both when using all individual data points (no binning and averaging) and when using the mean $PM_{2.5}$:AOD values in the PH:PBLH bins. The correlation coefficients were -0.12 when using all individual data points and -0.97 when using the mean $PM_{2.5}$:AOD values in the PH:PBLH bins. The low correlation when fitting all individual data points is due to the spread of $PM_{2.5}$:AOD when PH:PBLH is low. On the other hand, the strong anticorrelation when using the mean $PM_{2.5}$:AOD values in the PH:PBLH bins shows that PH:PBLH has skill for predicting $PM_{2.5}$:AOD for large collections of data. The

overall relationship displayed in Figure 2.4 and the results of the linear model are consistent with our hypothesis that a smoke plume injected higher into the atmosphere (relative to the PBLH) will result in enhanced AOD but a relatively low $PM_{2.5}$. We found that removing data outside of the 150 km buffer distance (PH retrievals >150 km away from a thermal hot spot) did not change the main result, (Figure A.11), though the number of data points decreased by over half.

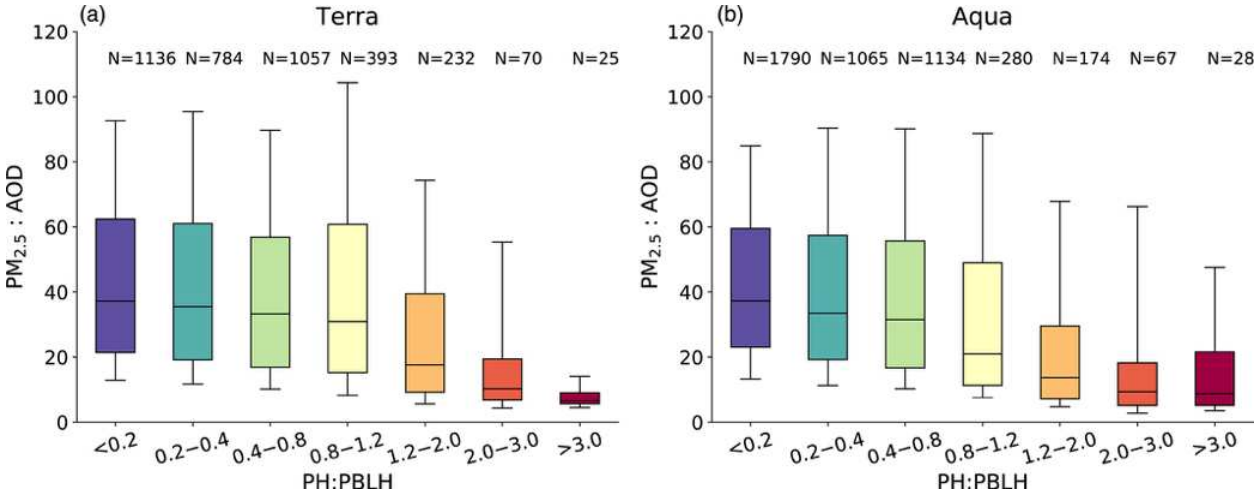

Figure 2.4: The $PM_{2.5}$:AOD as a function of PH:PBLH using PH and AOD estimates from Terra (a) and Aqua (b), PBLH, and $PM_{2.5}$ measurements for all colocated data (within 10 km of a monitor) over the western United States for July to September between 2010 and 2018. The number of observations in each PH:PBLH bin is listed above each box plot. The horizontal lines on each box indicate the 25th, 50th, and 75th percentiles of each $PM_{2.5}$:AOD distribution. The whiskers indicate the 10th and 90th percentiles of each distribution. The box-and-whisker color ranges from blue to red indicating an increase in PH:PBLH.

2.4 Conclusions

This work adds to the analysis of the MAIAC "plume injection height" product (Lyapustin et al., 2019), and our results suggest that incorporating PH information could improve $PM_{2.5}$ estimates in smoke plumes when aggregating across many smoke events. However, more work is needed to assess the value of PH for predicting $PM_{2.5}$ from individual AOD retrievals. We found that the western United States and northern Rockies had the highest number of days with PH retrievals (Figure 2.2) and relatively high average PH (Figures A.1a and A.1b in the supporting information). We also found a correlation between the mean PH values within AOD bins from

MAIAC, which may indicate that optically thicker smoke plumes are associated with higher PHs, but possibly the correlation may be due to the MAIAC PH retrieval algorithm, which can have a low bias for thin plumes. Finally, our analysis found that, on average, increasing PH:PBLH values are associated with decreasing $PM_{2.5}$:AOD, which indicates that smoke plumes escaping the boundary layer are associated with lower $PM_{2.5}$ concentrations. Therefore, this relationship could improve satellite-derived estimates of $PM_{2.5}$ during smoke events by providing a constraint on $PM_{2.5}$:AOD values if many AOD and PH retrievals are used in aggregate. This relationship could be especially impactful in regions where surface monitors do not exist. However, there is variability for all PH:PBLH bins, where high PH:PBLH cases may have smoke that extends below into the boundary layer and low PH:PBLH cases may be free-tropospheric plumes that are not opaque at 11 μm.

There are several additional aspects of the MAIAC PH retrievals that should be investigated in order to improve their ability to constrain $PM_{2.5}$ and provide more accurate PH constraints for CTMs, such as (1) understanding the accuracy of MAIAC PH as a function of AOD; (2) quantifying the chemical components of the plume that dominate the optical absorption at 11 μm and how the relative concentrations of these components change with dilution, chemistry, fire intensity, and fuel type; (3) studying the effect of regional smoke mixing with fresh smoke; and (4) testing MAIAC PIH against recent field airborne campaigns (e.g., WE-CAN, FIREX-AQ) with lidar measurements of vertical plume extent.

2.5 Funding and Data Availability

This work was funded by NASA Grant 80NSSC18M0120. Data sets for this research are available in these in-text citations: Lyapustin et al. (2019), Lyapustin and Wang (2018), Mesinger et al. (2006), and the EPA-AQS website (www.epa.gov/outdoor-air-quality-data). NARR data were provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, from their website (psl.noaa.gov/data/gridded/dat
MAIAC data can be found at the USGS website (https://lpdaac.usgs.gov/products/mcd19a2v006/).
We acknowledge the use of imagery from the NASA Worldview application (https://worldview.earthdata.nasa.gov)

Chapter 3

Investigating Sub-City Gradients of Air Quality: Lessons Learned with Low-Cost $PM_{2.5}$ and AOD Monitors and Machine Learning[2]

3.1 Introduction

Exposure to high concentrations of airborne particulate matter, especially particles with aerodynamic diameters smaller than 2.5 μm ($PM_{2.5}$), has adverse effects on public health (Forouzanfar et al., 2016; Hennig et al., 2018; Lelieveld et al., 2019; Pope et al., 2002; Schwartz et al., 1996). Increased exposure to $PM_{2.5}$ also imposes large economic burdens due to medical costs, welfare loss, disruptions to work productivity, and elevated crime rates (Burkhardt et al., 2019; Dechezleprêtre et al., 2019). As $PM_{2.5}$ concentrations are generally higher in urban areas, this burden can be especially large in major cities (Anenberg et al., 2019; Marlier et al., 2016). Research has shown that urban concentrations of $PM_{2.5}$ can be uniform with relatively larger heterogeneity in black carbon, organic aerosol, and particle number concentrations (Gu et al., 2018; Saha et al., 2021), though this likely varies by city. In addition, there can still be sub-city spatial and sub-daily temporal gradients in $PM_{2.5}$ that are difficult to measure due to the low spatial density of reference monitoring networks (Just et al., 2015; Bi et al., 2019; Gao et al., 2015; Wang and Oliver Gao, 2011). Improving predictions of $PM_{2.5}$ across cities could aid epidemiological investigations into the public health impacts of poor air quality (Southerland et al., 2021).

Low-cost sensor networks have been increasingly used to supplement reference networks and increase the spatiotemporal density of $PM_{2.5}$ measurements (Bi et al., 2020; Gao et al., 2015;

[2]Cheeseman, M., Ford, B., Rosen, Z., Wendt, E., DesRosiers, A., Hill, A. J., L'Orange, C., Quinn, C., Long, M., Jathar, S. H., Volckens, J., & Pierce, J. R. (2021). Technical note: Investigating sub-city gradients of air quality: lessons learned with low-cost $PM_{2.5}$ and AOD monitors and machine learning. *Atmospheric Chemistry and Physics Discussions*, 1–30. https://doi.org/10.5194/acp-2021-751 (In review)

Gupta et al., 2018; Snyder et al., 2013). These networks can be deployed by citizen scientists, thus simultaneously contributing to our understanding of air pollution and increasing public awareness of air quality issues (Ford et al., 2019; Gupta et al., 2018). For example, the PurpleAir network (https://www.purpleair.com), which uses light-scattering sensors to estimate $PM_{2.5}$ at sub-hourly timescales, has thousands of citizen-deployed monitors across the US and has been growing rapidly over recent years (Delp and Singer, 2020; Krebs et al., 2021). Despite the usefulness of low-cost sensor networks, they are often limited by their lower quality monitors, which can result in moderate to large uncertainties in their measurements (Gupta et al., 2018; Snyder et al., 2013). Furthermore, many regions in the US lack both low-cost and reference measurements of $PM_{2.5}$, which limits our understanding of public exposure to air pollutants.

Satellite observations of aerosol optical depth (AOD), an estimate of light extinction due to aerosols in an atmospheric column, can provide near-global coverage of clear-sky regions every 1-2 days; these observations are useful for filling in the gaps of $PM_{2.5}$ monitoring networks. Since satellite-retrieved AOD does not provide information about surface $PM_{2.5}$ directly, various techniques have been developed to leverage AOD measurements to inform surface $PM_{2.5}$ predictions. These techniques can generally be grouped into two categories: geophysical and statistical approaches. The geophysical approach to translate satellite AOD into $PM_{2.5}$ uses chemical transport models (CTMs) to simulate the relationship between $PM_{2.5}$ and AOD (Hammer et al., 2020; Liu et al., 2004, 2005; van Donkelaar et al., 2006, 2013, 2011) on global to local scales. The modeled $PM_{2.5}$:AOD ratios are then multiplied by the satellite AOD to derive an estimate of $PM_{2.5}$. While this approach is useful for annual-average concentrations (Hammer et al., 2020; van Donkelaar et al., 2010) and on shorter timescales for some locations and seasons (van Donkelaar et al., 2012), there are many limitations to this approach. For example, most satellites that capture AOD are polar-orbiting satellites, which only provide coverage during specific daytime-only (and cloud-free) overpass times, and hence fully rely on the model's predicted diurnal cycles for daily mean $PM_{2.5}$ estimates. Modeled $PM_{2.5}$:AOD relationships have also been found to be a large source of uncertainty in satellite-derived $PM_{2.5}$ (Ford and Heald, 2016; Jin et al., 2019), and a lack of

reference measurements of $PM_{2.5}$:AOD means they are difficult to validate. Monitoring networks such as SPARTAN have been developed to provide high fidelity $PM_{2.5}$:AOD observations but the monitoring sites are expensive and there are few worldwide (Snider et al., 2015, 2016). Finally, the resolution of satellite AOD measurements and CTM grid cells tends to be too coarse (e.g., >10 km) to study the fine-scale spatiotemporal resolutions necessary to capture the heterogeneity of $PM_{2.5}$ concentrations in urban areas, although recent satellite AOD products (Lyapustin et al., 2018) and high-resolution simulations (Jena et al., 2021; Kirwa et al., 2021) may remedy these issues.

Alternatively, satellite AOD retrievals can be incorporated into a statistical model to estimate surface $PM_{2.5}$. The simplest of these approaches uses co-located satellite AOD and surface $PM_{2.5}$ measurements in a linear regression model (Engel-Cox et al., 2004; Koelemeijer et al., 2006). However the relationship between AOD and $PM_{2.5}$ is complex and can vary due to changes in the aerosols' vertical distribution, water content, speciation, optical properties, and size distribution (Ford and Heald, 2016; Snider et al., 2015; van Donkelaar et al., 2010, 2006, 2013). Thus, many techniques for $PM_{2.5}$ estimation have been developed to incorporate information from many data sources including but not limited to AOD, meteorology, and geographic information such as land-use regression (Hoogh et al., 2016; Song et al., 2014) and geographically weighted regression (e.g. Lassman et al., 2017), not all of which are inherently suited to estimate both spatial and temporal variability in $PM_{2.5}$ concentrations. Even more complex computational and machine learning (ML) methods are also becoming increasingly common in estimating $PM_{2.5}$ (e.g. Di et al., 2016; Lightstone et al., 2017; Liu et al., 2018; Reid et al., 2015; Suleiman et al., 2019; Xi et al., 2015). Already, ML methods have been shown to be more accurate at predicting $PM_{2.5}$ than traditional CTM methods under certain conditions (Lightstone et al., 2017; Xi et al., 2015), and they can require less expertise to operationalize than CTMs.

ML represents a range of computational methods that build predictive models without explicit programming and with limited human intervention. One of the benefits of ML methods is that most can capture complex, non-linear relationships between many predictors (e.g., wind speed, AOD, land use) and a target variable (in this case, $PM_{2.5}$) in order to produce explicit predictions

of the target variable. Generally, ML models find relationships between predictors and the target using a training dataset, which is then validated using an independent testing dataset. Although the training and testing process is done with little human interference, the complexity and flexibility of ML models must be decided beforehand, and it is difficult to know what model configurations will result in the highest prediction skill. Thus, models must be tuned to find optimal configurations that reduce the risk of overfitting or underfitting the training dataset. As ML methods become more widely used, transparency in the execution of these methods and how they are validated will be key for the research community to ensure the quality of results obtained.

In this work, we use ML methods to investigate spatiotemporal variability in wintertime Denver. This work uses low-cost sensor measurements from the Citizen Enabled Aerosol Measurements for Satellites (CEAMS) project in addition to regulatory $PM_{2.5}$ measurements. The CEAMS project has (1) developed a low-cost monitor that can capture sub-hourly coincident $PM_{2.5}$ and AOD measurements and (2) trained citizen scientists to deploy them to study fine-scale spatiotemporal variability in the relationship between $PM_{2.5}$ and AOD. The CEAMS team conducted a deployment of these monitors during the winter of 2019-2020 in Denver, Colorado, United States (hereafter just "Denver"). To our knowledge, this was the first high-density network of low-cost, coincident sub-hourly AOD and $PM_{2.5}$ sensors deployed in a single city. We investigate the potential drivers of fine-scale $PM_{2.5}$ spatiotemporal variability in wintertime Denver by incorporating meteorological and geographical variables into a random forest (RF) ML regression framework (Breiman, 2001). We use a permutation metric to assess the relative importance of different predictor variables. We test whether co-located AOD measurements are identified as an important predictor of $PM_{2.5}$ and whether they increase the overall RF prediction skill compared to RFs that only used geographic and meteorological variables. The RF method was used here because it has been used to skillfully estimate $PM_{2.5}$ in past studies (Considine et al., 2021; Reid et al., 2015). We also compare our analysis of CEAMS $PM_{2.5}$ with results using reference $PM_{2.5}$ measurements from the Environmental Protection Agency (EPA). Finally, we discuss our RF methods in detail

and discuss how decisions made during data processing and model configuration may have influenced our results and the subsequent interpretation.

3.2 Methods

3.2.1 Data Sources

The CEAMS team developed two generations of low-cost monitors called the Aerosol Mass and Optical Depth (AMOD) monitors (Wendt et al., 2021). The AMODs used in this study are second-generation instruments (i.e., AMOD-v2) but, as the first version is no longer in use, we simply refer to the devices as AMODs herein. The CEAMS team trained citizen scientists to deploy AMODs in several different campaigns in northern Colorado (e.g. Ford et al., 2019). Here we analyze data from the CEAMS deployment during the winter of 2019-2020 in Denver (Figure 3.1). Thirty-two participants were recruited from across Denver through collaboration with the Community Collaborative Rain, Hail and Snow (CoCoRaHS) citizen scientist network (Cifelli et al., 2005) and other media outreach. Participants were trained by CEAMS researchers to set up devices using a mobile application (Quinn et al., 2019) and replace aerosol filters once a week. Measurements were taken from 14 November 2019 to 20 January 2020.

The AMOD is a low-cost ($1,200 manufacturing cost) $PM_{2.5}$ and AOD monitor that measures $PM_{2.5}$ in two ways: (1) real-time measurements using an inexpensive light-scattering sensor and (2) time-integrated measurements by collecting particles onto a filter using a size-selective cyclone separator and an ultrasonic pumping system (Volckens et al., 2017; Wendt et al., 2019, 2021). The real-time $PM_{2.5}$ sensor is the Plantower PMS5003, which has been widely deployed by networks such as PurpleAir, and validated in past work (Bulot et al., 2019; Sayahi et al., 2019). The AMOD also measures AOD at four discrete wavelengths (440, 500, 675, and 870 nm) using optically filtered photodiodes. The AMOD uses a solar tracking procedure that allows for automated AOD measurements throughout the day (Wendt et al., 2021). When AMODs were co-located with Aerosol Robotics Network (AERONET) AOD monitors in a series of validation experiments, the mean absolute error was 0.057 over AOD values ranging from 0.030 to 1.51 (Wendt et al., 2021).

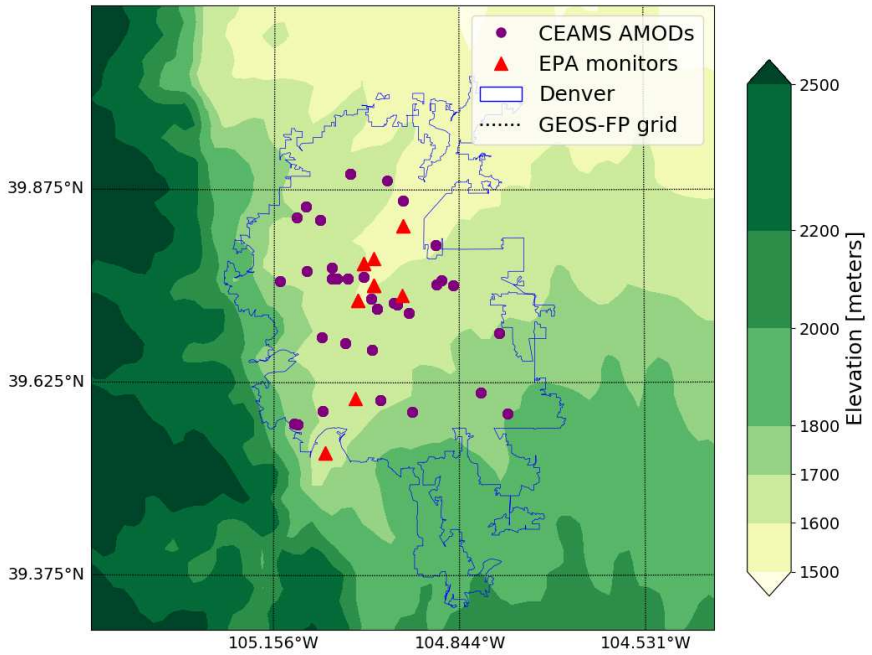

Figure 3.1: Map of elevation (Amante and Eakins, 2009) over Denver, CO, with locations of CEAMS Aerosol and Mass Optical Depth (AMOD) monitors (purple), EPA reference $PM_{2.5}$ monitors (red), outlines of the GEOS-FP grid-boxes (black), and outlines of the greater Denver-Aurora area (blue; based on cartographic files from 2015 TIGER/Line Shapefiles).

Real-time $PM_{2.5}$ and AOD can be sent to a central server by the AMOD over Wi-Fi every 20-minutes. The real-time $PM_{2.5}$ values used in this study were an average of instantaneous 1s values reported every 15-seconds over a period of 2.5 minutes (after a 30 second warm up period), taken every 20 minutes.

The real-time $PM_{2.5}$ data were quality controlled for possible sources of error and bias. First, any real-time $PM_{2.5}$ measurement reported over 500 μg m^{-3} was removed based on the manufacturer's guidance, similar to (Lu et al., 2021). Second, the AMOD $PM_{2.5}$ measurements were aggregated in two different temporal resolutions: 1-hour and 24-hour averages. The 24-hour averaged $PM_{2.5}$ measurements were only used if there were measurements for at least ¾ of the day to ensure it was representative of the entire day. Third, to correct for known biases of Plantower data tied to relative humidity (RH), we applied the following simple additive model in Eq. (1) tested by the US EPA (Barkjohn and Clements, 2020):

$$\mathbf{PM_{2.5}} = 0.524 \times \mathbf{AMOD_{CF1}} - 0.0862 \times \mathbf{RH} + 575, \tag{3.1}$$

The Plantower reports multiple $PM_{2.5}$ values based on different corrections, but $AMOD_{CF1}$ refers to $PM_{2.5}$ reported by the Plantower that was not corrected by the manufacturer's built-in atmospheric correction. The RH values used to correct each $PM_{2.5}$ data point were taken from the Plantower sensor as well. The Barkjohn and Clements (2020) model was developed specifically for the low-cost PurpleAir PM monitoring network, which uses either the PMS5003 or PMS7001 Plantower sensors. In this study, the Plantower $PM_{2.5}$ data were not corrected using the time-integrated filter measurements of $PM_{2.5}$ taken by the AMODs as in Ford et al. (2019).

The AMOD 500 nm AOD data were quality controlled based on a procedure previously described by Ford et al. (2019) that is based on similar methods used by AERONET. As long as the sun is greater than 10 degrees above the horizon (estimated by the solar tracking algorithm), the device will attempt to take 3 AOD measurements, or a triplet, within a 1-minute period at the start of each 20-minute interval. We did not require that AOD was measured over ¾ of the day, as we did with $PM_{2.5}$, since successful AOD measurements were less frequent and similar measurements from satellites only capture 1-2 times per day. Quality control and cloud screening were then applied in post-processing on each triplet at each wavelength. If less than 2 measurements per triplet attempt were taken or the range of AOD values was too large (>0.02) at any wavelength, then no measurements from that interval were used in this analysis. AOD was also filtered to remove measurements with air mass factors > 5 or an Ångström exponent < 0. The Angstrom exponent was measured between the 440 nm and 875 nm wavelengths. Finally, we assumed that 500 nm AOD values that were outside of the range 0-1 were likely the result of measurement errors, such as cloud contamination, though we acknowledge this may be wrong for dust or smoke-impacted scenes (which are uncommon in wintertime Denver).

We used 24-hour averaged $PM_{2.5}$ measurements from the Environmental Protection Agency's (EPA) Air Quality System (AQS) network from eight sites based in Denver (https://aqs.epa.gov/aqsweb/airdata.html) as shown in Figure 3.1. We limited our analysis to the

eight sites that use federal reference methods or federal equivalent methods for $PM_{2.5}$ and report local conditions (EPA parameter code 88101). We show in Table B1 the characteristics of each EPA monitoring site used in this study. Since EPA AQS $PM_{2.5}$ data is available for multiple years, we analyze data from November 1st - January 31st for the winters of 2017-2018, 2018-2019, 2019-2020. In Section 3.3, we choose to show results using the three years of data rather than limiting to the time period of the CEAMS deployment. However, we did the analysis for both time periods, and as we will discuss, the EPA RF results are similar, though noisier, if we limit EPA data to the time period of the CEAMS deployment.

We used meteorological data (Table 3.1) from the Goddard Earth Observing System forward-processing dataset (GEOS-FP) provided by the Global Modeling and Assimilation Office. GEOS-FP is produced with a native resolution of 0.25° (longitude) x 0.3125° (latitude) (~25 km horizontal resolution) with 72 hybrid vertical layers. The GEOS-FP data used were hourly or 3-hourly time averaged, depending on the variable. The 3-hourly data were linearly interpolated to hourly time resolution. Finally, all of the GEOS-FP variables were linearly interpolated spatially to the CEAMS and EPA monitor locations to better relate $PM_{2.5}$ observations with the environment for RF predictions.

We used multiple spatially varying datasets to describe each CEAMS and EPA monitoring location. Elevation information at each monitor location was extracted from the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) provided by the US Geological Survey (USGS) and the National Geospatial-Intelligence Agency (NGA) (Danielson and Gesch, 2011). The GMTED2010 data used in this analysis were at a 15-arc second (450 meters) horizontal resolution, which provides a unique elevation value for each CEAMS AMOD and EPA site. The slope of the terrain at the dataset resolution was calculated from GMTED's elevation data using QGIS. Additionally, we considered two predictors that have been found to be significant in land use regression (LUR) modeling: population density and travelled miles. We used census tract population density estimates from the Colorado Department of Public Health and Environment (CDPHE) (https://data-cdphe.opendata.arcgis.com) as a predictor in our RF models. Population

density estimates on the CDPHE site, measured in population density per square mile of land area within the given census tract, are directly from the 2013-2017 American Community Survey. Finally, we also incorporated traffic and road density data into our RF $PM_{2.5}$ predictions. We used model-assigned 2020 All-day Traffic Volumes from the Denver travel model, "Focus" (Model Cycle: RTP-2020 and Focus 2.3; https://drcog.org/services-and-resources/data-maps-and-modeling/travel-modeling) developed by the Denver Regional Council of Governments. This data includes shapefiles of large, medium, and small (i.e., arterial, collector, and local) road segments and a model estimate of annual average traffic volumes on each segment, which is measured in vehicles that traveled on each segment per year. We determined the length of each road segment within a 500 m buffer (i.e., intersection length) around each CEAMS and EPA monitor and then multiplied the intersection lengths by the traffic volumes of each segment, thereby producing an estimate of miles traveled by vehicles per year within 500 m of each monitor. As will be shown later, both population density and traveled miles were not found to be significant in estimating variability in $PM_{2.5}$ and hence we decided not to include any more LUR predictors in our model. A more comprehensive dataset of LUR predictors could be used in conjunction with geographical and meteorological predictors in future RF modeling.

3.2.2 Random Forest Models

To investigate the complex relationships between meteorology, geography, and air quality as well as the value that AOD can add to predicting air quality, we used RF ML regression models. RF models are made up of a group of unique and weakly correlated decision trees that are leveraged together to make a prediction. A decision tree begins with a random subset of the training data at the first node (i.e., the root node) and successively splits the data into branch nodes (Figure 3.2). Here, a mean value of $PM_{2.5}$ is predicted to represent all the samples in each node and the mean squared error is found. The data samples are then split at each branch node using a true or false question about one of the predictors (e.g., "is the temperature > 290 K?"), which is chosen to reduce the mean squared error of the samples in the following nodes. This process continues

until the decision tree reaches its termination criterion, such as when there are not enough samples to form a new branch node or the tree depth (i.e., number of branches) reaches some maximum set beforehand. At this point leaf nodes are formed with final predictions. These trees are built during the training process and then the testing data will follow the split nodes until they arrive at leaf nodes, which provide predictions for each value. This process is repeated using each tree in the forest, and the final prediction for each testing sample is given as an average of the predictions across all of the trees. The strength of RF models is that they leverage predictions from many weakly correlated decision trees, which helps protect the model against biases. The RF ensures that decision trees are weakly correlated and unique by giving a random subset of the samples and predictors to each decision tree, and another random subset of the predictors to choose from at each branch node during the training process. In this study, we created RF models with the scikit-learn Python package (Pedregosa et al., 2011) to predict the spatial and temporal variability of $PM_{2.5}$ using the predictors in Table 3.1.

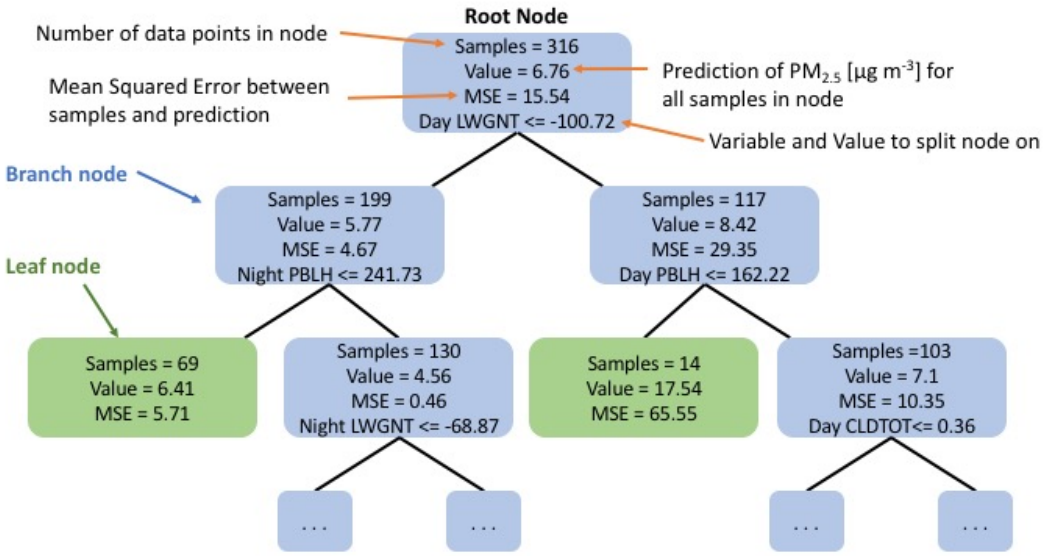

Figure 3.2: Example schematic of a random forest decision tree. The root and branch nodes are in blue, while the leaf nodes, which hold the final predictions, are in green. The black lines represent a splitting of a branch node into two more nodes.

Table 3.1: Predictor variables used in our RF models. The † symbol means that daytime (11am-3pm) and nighttime (11pm-3am) averages of these variables were used as separate predictors when using the 24-hour averaged $PM_{2.5}$ from the CEAMS and EPA dataset. For hourly data, these predictors were taken from the same location and hour of the $PM_{2.5}$ data.

Predictors	**Description**	**Units**	**Data Source**	**Models Used In**
Temp†	Surface skin temperature	Kelvin	GEOS-FP	All CEAMS + EPA
RH†	Relative humidity	%	GEOS-FP	All CEAMS + EPA
Wind Speed †	Wind speed at 10m height	m s^{-1}	GEOS-FP	All CEAMS + EPA
U*†	Friction velocity	m s^{-1}	GEOS-FP	All CEAMS + EPA
Precip †	Precipitation total	inches	GEOS-FP	All CEAMS + EPA
Cloud Frac†	Cloud total fraction	N/A	GEOS-FP	All CEAMS + EPA
LWGNT †	Longwave net radiation	W m^{-2}	GEOS-FP	All CEAMS + EPA
SWGDN †	Shortwave downwelling radiation	W m^{-2}	GEOS-FP	All CEAMS + EPA
PBLH †	Planetary boundary layer height	meters	GEOS-FP	All CEAMS + EPA
Elevation	Elevation above sea level	meters	GMTED 1km grid	All CEAMS + EPA
Slope	Terrain slope	degrees	calculated from GMTED	All CEAMS + EPA
Pop Density	Population density	people per mi^2	US Census	All CEAMS + EPA
Traveled miles	Annual mean miles traveled within 500m of monitor	mi yr^{-1}	Focus 2.3	All CEAMS + EPA
AOD	Daily mean 500nm AOD	dimen -sionless	CEAMS AMODs	CEAMS Test - AOD

In addition to investigating the relative importance of all of the different variables to predicting wintertime Denver $PM_{2.5}$, we also wanted to specifically test the skill that AOD adds to the prediction of $PM_{2.5}$. Thus, we created 3 subsets of both hourly and 24-hour CEAMS data, resulting in 6 datasets. The first subset, called the "Full Dataset," used all hourly or 24-hour CEAMS $PM_{2.5}$ data points and the co-located meteorological and geographic variables; however, it did not use the CEAMS AOD as a predictor. The second subset, called "Test - AOD," only used hourly or 24-hour CEAMS $PM_{2.5}$ at times and sites where CEAMS AOD was also available. However, similar to the first subset, the second subset did not use AOD as a predictor. Finally, the third subset, called "Test + AOD," is the same as the second, but CEAMS AOD was used as an additional predictor. Using these 3 subsets allowed for the investigation of three questions: 1) What is the change in prediction skill of our models if we limit the data to locations and days where AOD is available but we do not use AOD as a predictor? 2) When we use AOD as an additional predictor, how important is it for predicting $PM_{2.5}$ over wintertime Denver using the permutation metric? 3) How does the overall RF model skill change for predicting $PM_{2.5}$ after AOD is included as a predictor? Using models with both hourly and 24-hour data allowed us to analyze the relationships among air quality, meteorological and geographical factors, and the prediction skill of AOD measurements at different timescales.

To compare the CEAMS RF results to reference measurements, an additional RF model was created to predict 24-hour averaged EPA $PM_{2.5}$. We used the same predictors in our EPA RF model as we did with the CEAMS data, except for AOD as the EPA monitors do not have co-located AOD monitors. While there are fewer EPA monitors, they provide three full winters of data, allowing us to test whether our conclusions are robust when applied to a longer time period. The length of the datasets used in each CEAMS and EPA model can be found in Table 3.2.

When implementing a ML technique, such as RFs, models must be appropriately tuned. Tuning is the process for configuring the structure and assumptions of the model. For RF models specifically, the tuning process generally controls the number and complexity of the decision trees that make up the random forest, the way in which data are sampled, and the number of predictors

Table 3.2: The number of data points used for the training and testing of each RF model is listed. A dash is used to indicate that a column is not applicable.

	CEAMS Full Dataset	**CEAMS Test +AOD**	**CEAMS Test -AOD**	**EPA model**
Number of data points in 24-hour models:	634	307	307	2411
Number of data points in hourly models:	18969	1043	1043	-

that should be considered at each split in each tree (i.e., hyperparameters). Tuning is necessary to ensure that the model does not underfit or overfit the training data. For an RF model, overfitting the training data occurs when the decision trees begin fitting onto the noise that exists in the training dataset. As a consequence, the RF could skillfully re-create the $PM_{2.5}$ values of the training dataset if fed the same predictor values associated with those training values (i.e., the same combination of meteorology, geography, etc.) because it learned to predict even the noise of the training data. However, if this same RF model was given an unseen set of $PM_{2.5}$ values and their associated predictors, such as $PM_{2.5}$ from a different time period or a different monitor, it could perform poorly. Overfitting can go unnoticed if there is data leakage between the testing and training data, which could occur if the data in both sets are autocorrelated. Underfitting, on the other hand, is more straightforward; it occurs when the decision trees are too simple and fail to capture the relationships that exist between the predictor and target variables. Typically, to ensure against over or underfitting, the data are split into separate tuning, training, and testing datasets. However, since our dataset spans only a couple months, we tuned, trained, and tested our RF models using a cross-validation (CV) method.

As ML methods become popular in air-quality research, we hope that transparency about our tuning process allows for reproducibility and serves as a guide for future work. We used a *k*-fold cross-validation method to tune each model over a selection of hyperparameters (Table 3.3) using the scikit-learn package GridSearchCV (Pedregosa et al., 2011). GridSearchCV automatically trains and validates an RF model for every combination of hyperparameters given to it (Table 3.3) over *k* number of folds (in our case, 5 folds for each hyperparameter combination). A *k*-fold CV

entails chunking the data into k number of equally sized groups, using k-1 number of folds for training the RF model, validating that RF model using the remaining fold, and then repeating that process until every fold has been used for validation. We chose the final hyperparameters for our 6 CEAMS RF models and our EPA RF model based on the best MSE for each combination of hyperparameters in each model. However, if a similar MSE was found for a hyperparameter selection that allowed for simpler tree structures (shallower depth, fewer trees, etc.), the simpler model was chosen instead of the more complex model. For example, Figure B.1a shows that RF model skill when using 120 trees or 90 trees result in very similar distributions of model skill. Thus, while tuning the RF model that used the 24-hour CEAMS Full Dataset, we chose 90 trees to limit the complexity of the model without losing performance. Similarly, we also chose a maximum depth of 15, 2 samples needed to form a leaf, and 5 samples needed to split a branch node. More information about the hyperparameters chosen for each RF model is in the Supplement (Figure B.1-B.7).

Table 3.3: Hyperparameters tested during the tuning of each random forest model.

Hyperparameters in scikit-learn	**Description**	**Values tested**
n_estimators	Number of trees	20, 30, 40, 50, 60, 70, 90, 120
max_depth	Maximum depth of each tree	Varies with different models
min_samples_split	Minimum samples needed to split an internal node	1, 2, 5
min_samples_leaf	Minimum samples needed to split at a leaf node	1, 2, 4
max_features	Maximum number of predictors to consider for each split	'Sqrt', 'Auto'
bootstrap	Each decision tree will be built with a bootstrapped sample of the dataset	True, False

Once the final model hyperparameters were chosen, the models were trained and tested over another 5-fold CV. Although the CV method was used in both our tuning and testing methodology, each was done using a different random shuffling of the data. See Figure B.8 for an example

of a comparison between CEAMS 24-hour $PM_{2.5}$ and the RF prediction after validating against 1 testing fold. We estimated the uncertainty in our RF model predictions by calculating 95% confidence intervals for the performance metrics of each RF model using a bootstrapping method. Bootstrapping entailed taking random samples of the model predictions and the associated $PM_{2.5}$ measurements, with replacement, and finding the errors statistics (e.g. the root mean squared error [RMSE] and the coefficient of determination [R^2]) of each random sample. This process was repeated until a distribution of each error statistic was created. Then the error statistics were sorted into ascending order and the values at the 2.5% and 97.5% percentiles represented the 95% confidence interval. Finally, to investigate the relative importance of each predictor for the RF predictions, a permutation importance metric was used, which tests the change in model prediction skill after randomly shuffling one predictor of the validation data at a time. Thus, the higher the permutation importance, the greater loss of prediction skill if that predictor was randomized. To test the robustness of each permutation importance score, the metric was calculated 100 times for each predictor for each of the 5 iterations of the 5-fold CV, resulting in a distribution of 500 permutation importance scores per predictor.

The CEAMS and EPA $PM_{2.5}$ measurements were autocorrelated at both hourly and 24-hour timescales. This lack of independence can result in information being shared between the training and testing datasets. This information sharing makes it much easier for the RF models to predict $PM_{2.5}$ from the testing dataset because it looks very similar to the data the models were trained on. Thus, the RF models' prediction skills can be inflated. In our models, this information sharing occurred when the $PM_{2.5}$ and predictor variables were randomly shuffled before we chunked the data into *k*-folds (i.e., "shuffled *k*-folds"), which is the default behavior of many *k*-fold CV procedures because they assume each value is an independent sample. Alternatively, data can be chunked into consecutive *k*-folds, which reduces information sharing between testing and training datasets (Figure 3.3).

To understand the impact of autocorrelation on our models, we present results from training and validation with shuffled and consecutive *k*-folds in the Results and SI, respectively. Our analysis of

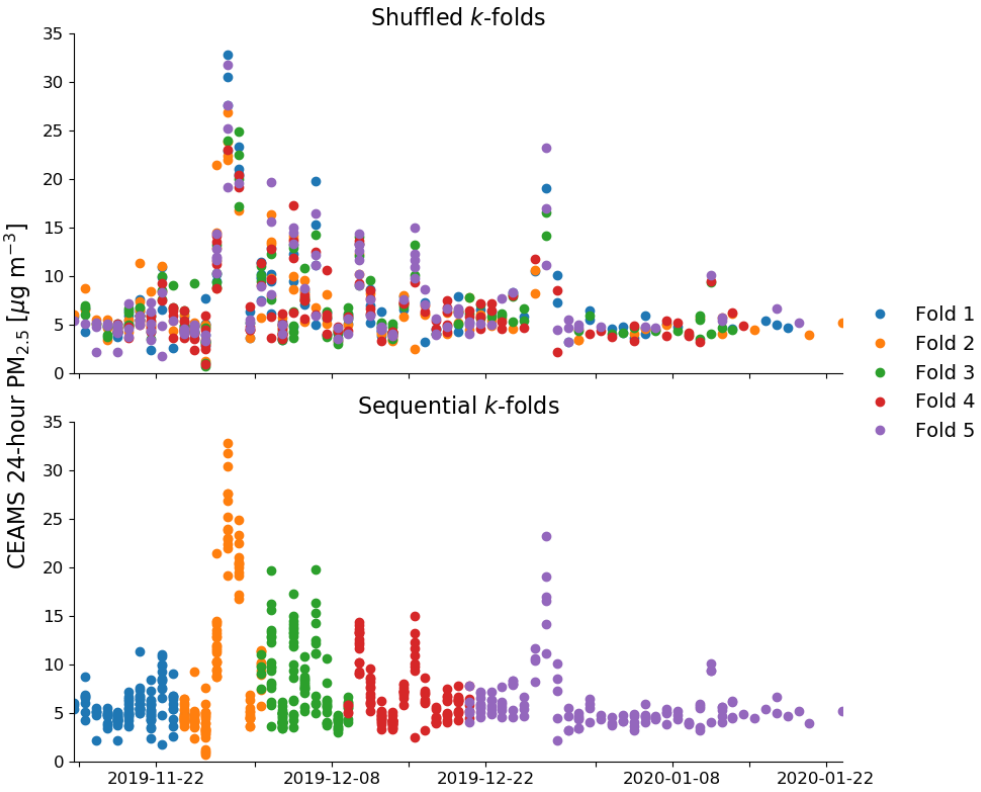

Figure 3.3: Example of CEAMS 24-hour data split into 5 randomly shuffled folds (a) and 5 consecutive (i.e., chunked in time) folds (b) used in the RF model training and testing process. During a 5-fold CV, 4 of these folds are used to train a model and the remaining fold is used to validate the model, which is repeated iteratively until each fold has been used for validation.

the CEAMS data, which was limited to several weeks of measurements, showed larger differences between results using shuffled data and results using consecutive k-folds compared to our EPA analysis, as discussed in Section 3.4. This is likely due to the greater noise, residual RH bias of the Plantower sensors, and inconsistent sampling pattern of the citizen science deployment (Figure B.9), which limited the predictability of CEAMS $PM_{2.5}$. Thus, the CEAMS RF models using consecutive folds often performed poorly and our confidence in the chosen predictors was low. Hence, while the consecutive method is preferable for long, comprehensive data sets, we present here the CEAMS results of our RF models using shuffled data because we found that, even though their predictive ability appears inflated for unseen data due to the autocorrelation, they still allow for useful interpretations of meteorological, geographical, and other predictors of $PM_{2.5}$.

3.3 Results

3.3.1 CEAMS Denver Deployment Data

During the CEAMS pilot deployment in wintertime Denver our AMODs retrieved over 18,000 hourly averaged quality-controlled $PM_{2.5}$ measurements ($\boldsymbol{\mu}$ = 8.2 μg m^{-3}; σ = 12.6 μg m^{-3}) and over 1000 hourly averaged quality-controlled AOD measurements ($\boldsymbol{\mu}$ = 0.06; σ = 0.05) (Table 3.2). There were only a few periods of significantly elevated $PM_{2.5}$ (24-hour means > 10 μg m^{-3}) during the deployment (Figure 3.4a), and they did not often coincide with a proportional increase in daytime AOD (Figure 3.4b). Thus, there was a low correlation between $PM_{2.5}$ and AOD (Figure 3.4c). The days with elevated 24-hour averaged $PM_{2.5}$ tended to be driven by late afternoon and overnight buildup of air pollution potentially caused by automobile emissions and residential heating during stable winter nighttime conditions over Denver (Figure B.10). There were also strong sub-city gradients of concentrations during some periods as shown in Figure B.11.

To enhance our understanding of the potential drivers of $PM_{2.5}$ over wintertime Denver, prior to creating ML models, we investigated the relationship between our CEAMS 24-hour $PM_{2.5}$ measurements and different spatial and spatiotemporal predictors (Figure 3.5). This analysis helps set

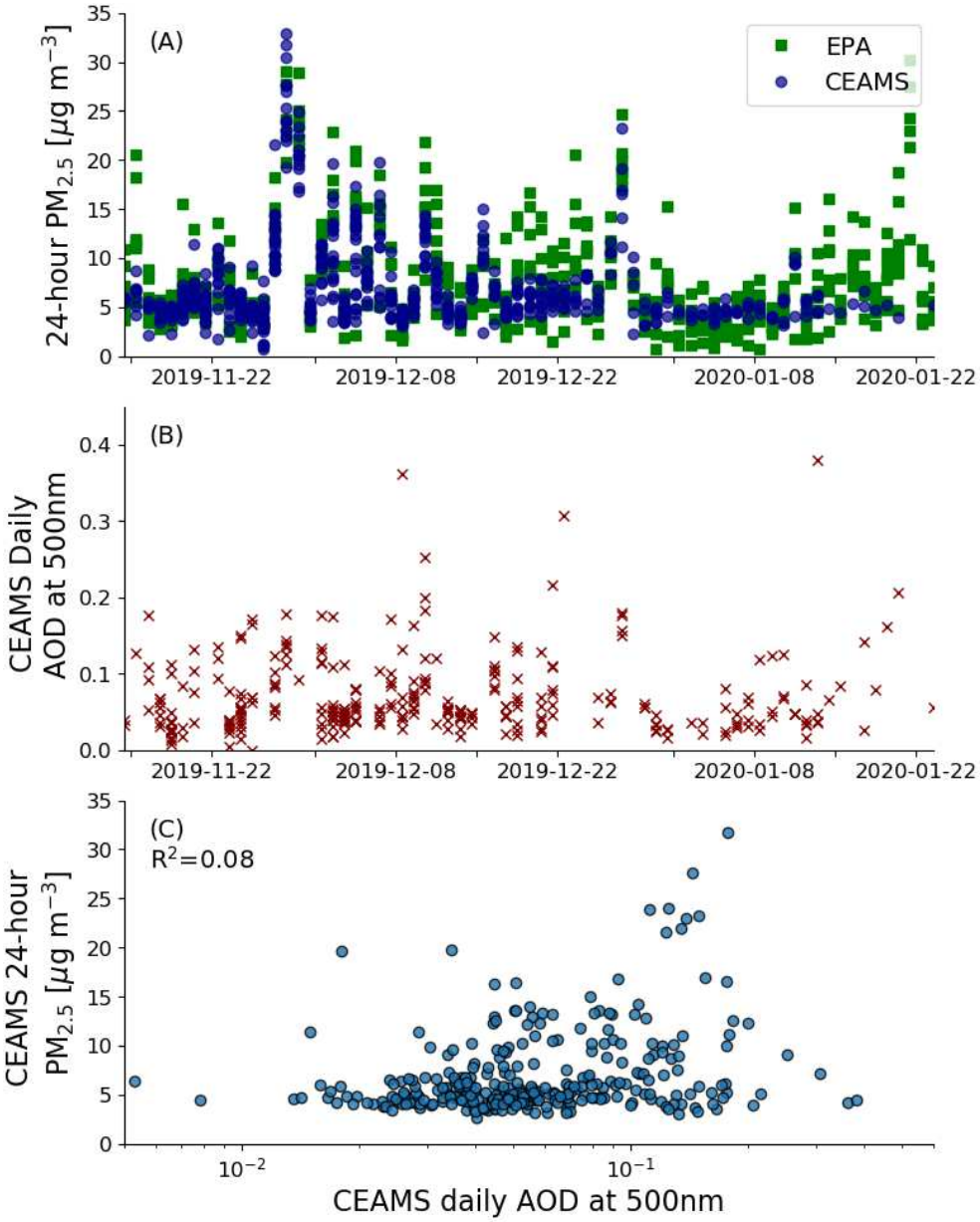

Figure 3.4: (a) CEAMS AMODs' 24-hour averaged $PM_{2.5}$ measurements taken in Denver, CO by 32 citizen scientists between November 14th, 2019, and January 20th, 2020. EPA reference measurements of 24-hour $PM_{2.5}$ from 8 sites are also shown for the same period. (b) CEAMS AMODs daily averaged AOD measurements taken by the same devices shown in panel a. (c) The relationship between 24-hour time averaged $PM_{2.5}$ and daily averaged AOD taken by the same CEAMS AMODs.

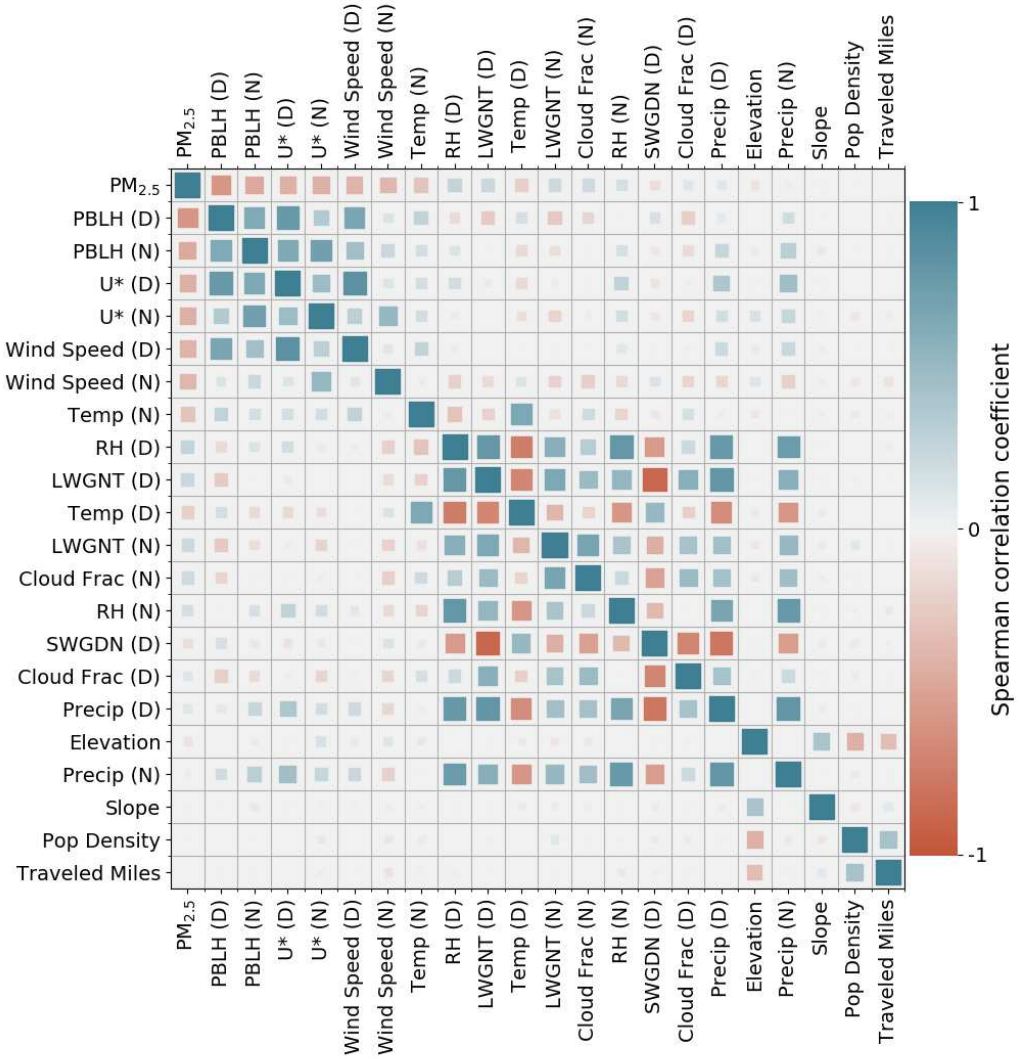

Figure 3.5: Correlation matrix of CEAMS 24-hour $PM_{2.5}$ data and all RF model predictors used in each 24-hour model. The "(D)" and "(N)" represent daytime (11am-3pm) and nighttime (11pm-3am) averages, respectively, of each meteorological predictor. The predictors are in order from greatest to least corresponding to the absolute value of the spearman rank correlation. The size of each box also corresponds to the absolute value of the spearman rank correlation between each variable so that the least important predictors have the smallest boxes. The same predictors are used in the RF models that predict hourly $PM_{2.5}$ but hourly averaged meteorological factors were used instead of daytime and nighttime averages.

expectations for potentially important predictors in the ML models. We found that 24-hour $PM_{2.5}$ was negatively correlated with daytime and nighttime planetary boundary layer heights (PBLH), friction velocity (U*), wind speed, and nighttime temperature (all of which are positively correlated with each other). The correlation between our $PM_{2.5}$ measurements and these meteorological predictors is likely due to wintry conditions in Denver that lead to stagnant air, thermal inversions, and low boundary layers, which can all serve to slow the ventilation and downwind transport of urban air pollution. We also hypothesize that wintry conditions also may have led to increased wood burning for residential heat, which would enhance $PM_{2.5}$ build up, especially overnight. However, this temperature-emission connection is a hypothesis that we do not test here. $PM_{2.5}$ tended to be elevated during higher RH conditions as well, which may be due to a combination of the physical connection between $PM_{2.5}$ and meteorological conditions, as well as remaining RH bias in the measurements that was not removed using the Barkjohn and Clements (2020) correction. We explore this RH connection more in our discussion of variable importance in our CEAMS and EPA RF models. The spatial-only predictors (elevation, slope, population density, and vehicle travelled miles) were only weakly correlated with $PM_{2.5}$ because temporal variability dominated over spatial variability in our dataset; however, these spatial predictors may still provide information to refine the ML estimates. Figure 3.5 shows that $PM_{2.5}$ likely has a complicated and nonlinear relationship with local meteorology during our deployment. However, it is difficult to interpret which variables or combinations of variables are more useful for predicting $PM_{2.5}$ which is why we chose to use RF models to quantitatively determine predictors of spatiotemporal variability of $PM_{2.5}$ in wintertime Denver.

3.3.2 Random Forest Model Skill

In Figure 3.6, we present the 95% confidence intervals of the performance metrics for each RF model using shuffled *k*-folds. We found that, of the RF models predicting 24-hour $PM_{2.5}$ measurements, the model using the CEAMS Full Dataset showed the highest coefficient of determination, lowest RMSE, and a slope nearest to 1 between predictions and $PM_{2.5}$ measurements. The 95%

confidence intervals of the Full Dataset model overlapped with the Test - AOD and Test + AOD RF models, which implies that limiting the CEAMS data to locations and days where AOD was available did not result in a significant reduction in model skill for 24-hour $PM_{2.5}$ prediction. We did see a small reduction in skill for hourly $PM_{2.5}$ predictions (Figure 3.6e-g) after limiting the dataset to only locations and hours where AOD was also taken. The 24-hour CEAMS Full Dataset model also showed similar skill to the EPA model for all metrics (Figure 3.6a-d). However, results for the RF models using consecutive *k*-folds showed a significant decrease in prediction skill, especially for the CEAMS Full Dataset, while the EPA model results showed a less substantial decrease in skill (Figure B.12). We also found that the RF models were better at capturing temporal variability than spatial variability during the CEAMS deployment. The hourly $PM_{2.5}$ observations showed an average spatial standard deviation of ~2.5 μg m^{-3} while the RF model predictions showed an average spatial standard deviation of ~1.5 μg m^{-3} for shuffled *k*-folds (Figure B.13) and only ~0.6 for consecutive *k*-folds (Figure B.14).

By comparing the CEAMS Test - AOD and the CEAMS Test + AOD model performance metrics, we investigated the change in model performance if AOD was used as an additional predictor of $PM_{2.5}$. We found that the confidence intervals of the Test - AOD and Test + AOD models almost entirely overlapped for 24-hour $PM_{2.5}$ predictions (Figure 3.6a-d), which shows that the daily averaged AOD did not add to the overall prediction skill of the RF models. We found a small increase in mean model skill when comparing hourly $PM_{2.5}$ predictions between the Test - AOD and Test + AOD, indicated by the increased R^2, decreased RMSE, and a slope nearer to 1 for the Test + AOD model, but the confidence intervals overlap, which indicates that the difference between models had low statistical significance. This finding may be because AOD can be disconnected from $PM_{2.5}$ in a variety of ways. For example, daytime-only measurements such as AOD would be unable to capture evening buildup of $PM_{2.5}$ that we often saw in the Denver pilot deployment. Furthermore, $PM_{2.5}$ and AOD share a nonlinear relationship that can be altered by the aerosol hygroscopicity, aerosol vertical profile, size distribution, and chemical composition. However, AOD would likely provide greater predictive skill in the spatial variability of long-term averages in $PM_{2.5}$ (Hu et al.,

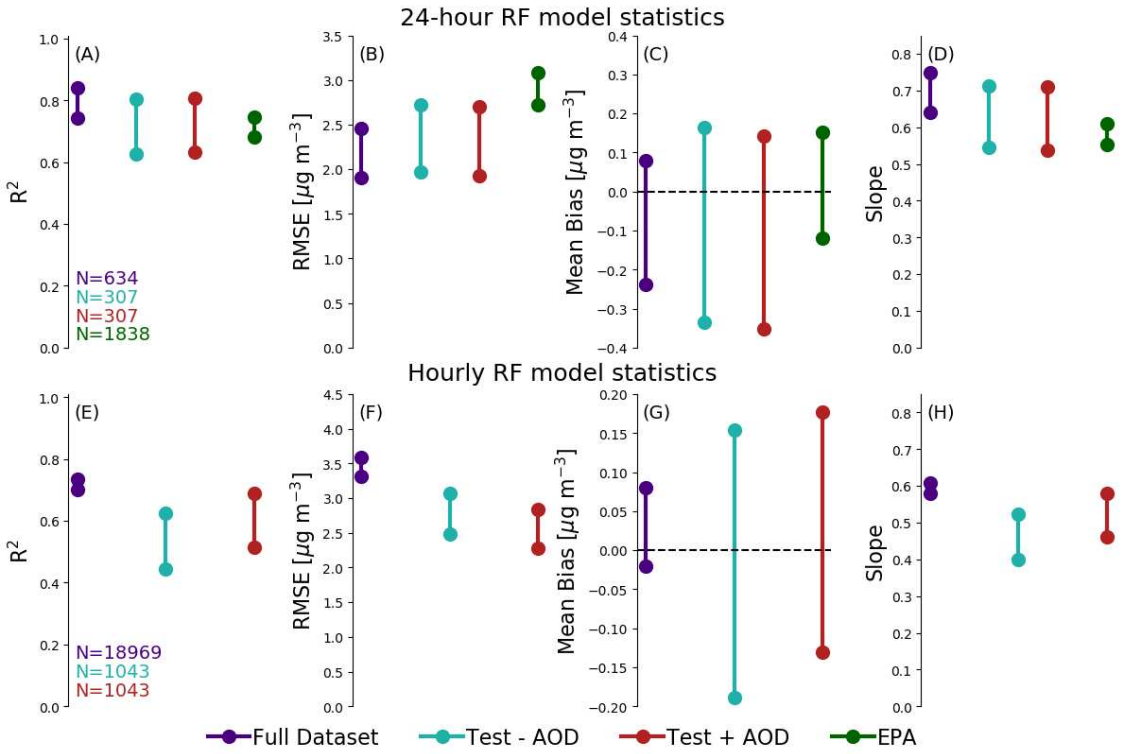

Figure 3.6: The 95% confidence interval of the error metrics for all of the CEAMS RF models (Full Dataset, Test - AOD, and Test + AOD) in predicting both 24-hour and hourly $PM_{2.5}$ and the error metrics for the 24-hour EPA model. The 95% confidence intervals show an estimate of the uncertainty range and, thus, if the intervals of two different models overlap, any difference in their error metrics are likely not statistically significant. The error metrics for each 24-hour $PM_{2.5}$ RF model includes (a) the coefficient of determination (R^2) (b) root mean squared error (RMSE), (c) mean bias, (d) and slope of the linear regression. Plots (e), (f), (g), and (h) show analogous results but for the hourly $PM_{2.5}$ predictions, which we did not predict for the EPA dataset. The size of each 24-hour and hourly dataset, before being split into k-folds, is shown in the bottom left corner of plot (a) and (e).

2014; Liu et al., 2005; van Donkelaar et al., 2010) and locations where $PM_{2.5}$ is driven by daytime variability (van Donkelaar et al., 2011).

3.3.3 Variable Importance for Spatiotemporal $PM_{2.5}$ Predictions

We use our RF models not only to estimate $PM_{2.5}$ concentrations but also to investigate the variables importance in predicting $PM_{2.5}$ for wintertime in Denver. We show distributions of permutation importance for the top 10 predictors, ranked by their median permutation importance, of each RF model that predicted CEAMS and EPA $PM_{2.5}$ concentrations (Figure 3.7). We found that the meteorological predictors vary largely between 24-hour (Figure 3.7a-d) and hourly (Figure 3.7e-g) models of $PM_{2.5}$. The daytime (11am-3pm) averaged PBLH and RH were consistently strong predictors in each CEAMS 24-hour RF model (Figure 3.7a-c) and daytime PBLH was the strongest predictor in the EPA model (Figure 3.7d). It was not surprising that PBLH was a strong predictor of $PM_{2.5}$, though we expected nighttime PBLH to be a stronger predictor than daytime PBLH because high $PM_{2.5}$ usually occurs during the late evening to early morning hours (Figure B.10). However, it may be that low daytime PBLH values were better correlated with periods where $PM_{2.5}$ was elevated for extended periods of time, because ventilation of air pollution was hampered by stagnant air masses. Additionally, day and night PBLH are correlated so day PBLH may act to predict nighttime $PM_{2.5}$ buildup (Figure 3.5). The strength of daytime-averaged RH in our CEAMS RF models may be due to physical connections between $PM_{2.5}$ and RH, because high RH is tied to colder conditions (Figure 3.5), which is subsequently correlated with boundary layer heights (Figure 3.5). However, this may also be due to residual bias of the Plantower measurements for which the Barkjohn and Clements (2020) correction was unable to account. The nighttime (11pm-3am) averaged cloud fraction was the third most important predictor in the CEAMS Full Dataset while daytime longwave net radiation was the third most important for the CEAMS Test - AOD and Test + AOD models. As expected, since we saw no change in prediction skill between the CEAMS Test - AOD and Test + AOD 24-hour predictions (Figure 3.6a-d), AOD also did not have high permutation importance in the 24-hour CEAMS Test + AOD model.

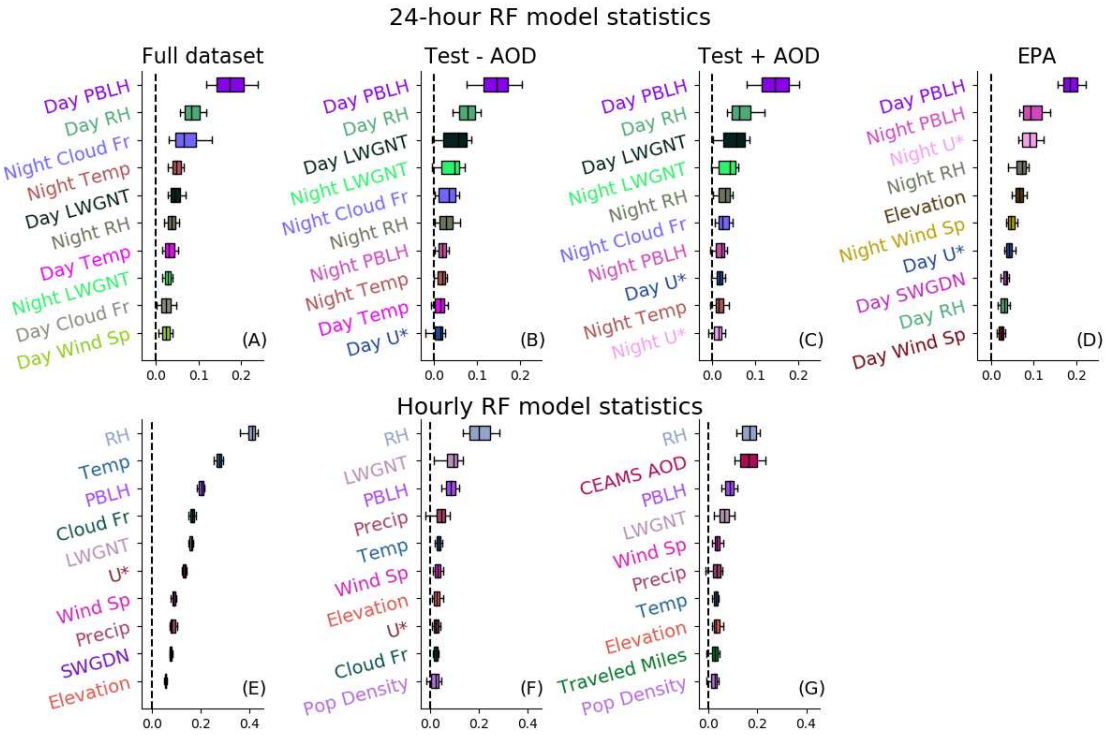

Figure 3.7: Box-and-whisker plots of the distribution of 500 permutation importance measurements from the top 10 predictors of each model. The 500 permutation importance values are taken from 100 repeats of permutation importance from each of the 5 testing folds. The whiskers of each box are the 10th and 90th percentile of the permutation importance distribution. The edges of each box represent the 25th and 75th percentile and, finally, the centerline of each box represents the median of the permutation importance distribution. (a) The 24-hour $PM_{2.5}$ predictions of the CEAMS Full Dataset, (b) The 24-hour $PM_{2.5}$ predictions of the CEAMS Test - AOD dataset, (c) The 24-hour $PM_{2.5}$ predictions of the CEAMS Test + AOD dataset. Plots (d), (e), and (f) are analogous to the CEAMS 24-hour $PM_{2.5}$ RF models in plots (a), (b), and (c) but for hourly averaged $PM_{2.5}$ and associated predictors. The permutation importance of every predictor used for each model is in the Supplement (Figure B.15-B.16).

The hourly RF models relied more on different meteorological variables than the 24-hour models (Figure 3.7e-g). The most important predictor for all three hourly models was hourly averaged RH. We hypothesize that RH is the strongest predictor because the RH correction factor that we applied to our hourly-averaged $PM_{2.5}$ data is based on Barkjohn and Clements (2020), which used 24-hour averaged $PM_{2.5}$. Thus, the importance of RH in our model may be more reflective of the RH bias in the sensor measurement than the physical connection between $PM_{2.5}$ and RH. Unlike the 24-hour CEAMS models, hourly-averaged AOD was the second strongest predictor in our Test + AOD model, and we saw improvement in mean prediction skill of this model compared to the hourly Test - AOD model (Figure 3.7e-h). This result implies that hourly AOD added some skill in predicting hourly $PM_{2.5}$ data. The increased importance of AOD in the hourly models relative to the 24-hour models is likely because the AOD is co-located in time (within the hour) with the $PM_{2.5}$ measurement but not with most of the 24-hour period (as AOD is only available during daylight hours). Finally, we expected spatial predictors such as elevation, vehicle miles traveled, and population density to be more important for all of the RF models, because we hypothesized that air pollution would pool at low elevations during the winter in Denver, as late evening traffic and residential wood burning emissions were trapped over Denver by stagnant air. Instead, we found that their permutation importance was near zero for all of these variables in the CEAMS models (Figure B.15-B.18). However, our EPA model results indicated that elevation was a moderately important predictor when we used shuffled *k*-folds (Figure 3.7d) and the 2nd most important when we used consecutive *k*-folds (Figure B.19). Vehicle miles traveled and population density are generally not important predictors in our RF models, which may be due to temporal variability being larger than spatial variability in our dataset as well as these predictors having no correlation with $PM_{2.5}$ (Figure 3.5).

3.3.4 Sensitivity of Results to Data Processing and RF Setup Decisions

One critical caveat to our CEAMS data analysis is that there is strong autocorrelation in daily, and especially hourly, $PM_{2.5}$. Thus, when we shuffle the $PM_{2.5}$ data and their associated predictors

before splitting the data into *k*-folds for training and testing, information will be shared between the training and testing datasets. We tested the potential impact of autocorrelation on our model skill by repeating the CEAMS and EPA analyses without shuffling the data before splitting it into *k*-folds. We saw a significant decrease in skill for the CEAMS RF models, especially hourly (Figure B.12), and a decrease in the consistency of predictor ranking (Figure B.17-B.18). Our analysis of the EPA RFs, however, showed a smaller decrease in predictive skill when we compared the results from models trained and tested using shuffled vs. consecutive *k*-folds. For example, the upper bound of the 95% confidence intervals decreased by 0.5 and 0.3 for the CEAMS Full Dataset and EPA 24-hour models, respectively, when we used consecutive *k*-folds. Furthermore, the meteorological and geographical predictors remained more consistent in the EPA model when we used consecutive *k*-folds (Figure B.19). To test whether these results were due to the increased length of the EPA dataset, we repeated the analysis only using EPA measurements from 15 November 2019 - 15 January 2020, the same time period as our CEAMS deployment, and found similar results for both shuffled (Figure B.20) and consecutive (Figure B.21) *k*-folds. Thus, the sharp decrease in CEAMS RF skill may be due to the quality of the Plantower sensor measurements and/or the inconsistent sampling patterns of the CEAMS AMOD citizen science deployment, which means that the model would not be able to train itself appropriately to compare well to unseen data. Furthermore, as we mentioned in our discussion of Figure 3.2, there are only a few short periods of significantly elevated $PM_{2.5}$ during the CEAMS deployment, which led to consistent under-prediction of high $PM_{2.5}$ in the RF models (Figure 3.7d and 3.7h), especially in those that used consecutive *k*-folds (Figure B.12). However, even though we do not have confidence that our CEAMS model would have predictive skill for new time periods, we do have more confidence that our interpretation of the top meteorological and geographical relationships is valid under the conditions of the CEAMS campaign. In addition to the impact of autocorrelation and shuffling on our results, we found that various decisions made in processing our data could lead to variations in the predictive skill of our models and the order of variable importance's. For example, we found that using a linear interpolation method instead of nearest neighbor for co-locating the GEOS-FP meteorological data

to the CEAMS monitors affected which predictors were considered most important (not shown here), likely because the linear interpolation method introduced greater spatial variability among predictors when comparing $PM_{2.5}$ from monitors in the same grid-box. We also saw that our results were sensitive to the use of an RH correction for the CEAMS $PM_{2.5}$ because the relative importance of RH variables decreased after the RH correction was applied in the 24-hour CEAMS RF models. Finally, we found it useful to tune our models on a greater selection of hyperparameters than the maximum depth and the number of trees. We recommend that future investigations of $PM_{2.5}$ with machine learning (RF in particular) carefully consider the decisions described above.

3.4 Conclusions

The CEAMS pilot campaign provided a novel high-spatial-density, low-cost network of citizen-scientist-deployed monitors that captured coincident sub-hourly $PM_{2.5}$ and AOD measurements in Denver. For the measurements gathered in this work in Denver over wintertime, $PM_{2.5}$ concentrations varied much more with time than in space. This finding, that $PM_{2.5}$ varies less with space than time within an urban environment, is generally consistent with recent $PM_{2.5}$ measurements made in other US cities including Oakland, CA (e.g. Shah et al., 2018) and Pittsburgh, PA (Gu et al., 2018).

To understand potential drivers of $PM_{2.5}$ over wintertime Denver, we analyzed the importance of various meteorological and geographical features in predicting spatiotemporal variability of $PM_{2.5}$ from both the CEAMS low-cost and EPA reference networks. We found that daytime-averaged (11am-3pm) PBLH was the strongest predictor of intra-city spatiotemporal variability for both low-cost and reference measurements of 24-hour averaged $PM_{2.5}$. The ranking of less important predictors in our CEAMS and EPA RF models differed, however. For example, nighttime-averaged PBLH and friction velocity were strong predictors of EPA 24-hour averaged $PM_{2.5}$, while daytime-averaged RH was a strong predictor of CEAMS 24-hour averaged $PM_{2.5}$. We also found that hourly averaged RH was the strongest predictor of CEAMS hourly-averaged $PM_{2.5}$. However, we expect that the RFs' reliance on RH for CEAMS 24-hour and hourly $PM_{2.5}$ prediction was

likely due, in part, to residual RH bias in the Plantower measurements of $PM_{2.5}$, especially since RH was not one of the top 3 predictors in our EPA RF model. Spatial variables such as population density and number of vehicle miles traveled were consistently unimportant predictors in our RF models, although elevation was important in our multi-year EPA model. Perhaps due to the lack of importance placed on spatial variables, our RF models were unable to fully capture the extent of the spatial variability of $PM_{2.5}$ seen over Denver (Figure B.12-B.13). Historically, most LUR modeling has relied on spatial variables to explain differences in $PM_{2.5}$ concentrations and discounted temporal variability since the objective is usually to quantify the average $PM_{2.5}$ exposure over a time period of interest (e.g., seasonal, annual). In cases where studies have developed spatiotemporal LUR models because there is an interest in quantifying the time-resolved $PM_{2.5}$ exposure (Martenies et al., 2021), they do not appear to use meteorological variables directly. This work suggests that LUR modeling can benefit from using meteorological variables (e.g., PBLH) in addition to spatial and geographical variables in estimating $PM_{2.5}$.

Finally, we tested whether coincident AOD measurements added predictive skill to hourly and 24-hour averaged $PM_{2.5}$ predictions beyond what was achievable using only geographical and meteorological information in wintertime Denver, as may be possible with satellite AOD retrievals. We found that daily-averaged AOD measurements did not improve RF model predictions of CEAMS 24-hour $PM_{2.5}$, nor was AOD identified as a strong predictor of 24-hour $PM_{2.5}$ based on the permutation metric. The lack of skill added by AOD to 24-hour $PM_{2.5}$ prediction is likely because 24-hour $PM_{2.5}$ in wintertime Denver is largely driven by evening and overnight build-up of air pollution, which daytime-only measurements such as AOD cannot capture. However, when incorporating CEAMS AOD as a predictor in our RF model of hourly-averaged $PM_{2.5}$, we found an increase in average prediction skill, and the hourly averaged AOD was the second strongest predictor based on a permutation importance metric (although the 95% confidence intervals overlapped, which implies that the increase in model skill had low statistical significance). This implies that AOD retrieved from geostationary satellites may be a better predictor of $PM_{2.5}$ than AOD from polar-orbiting satellites, because they may help capture more of the diurnal cycle of aerosols. We

also expect AMOD AOD to be a better predictor of daily and hourly averaged $PM_{2.5}$ in other seasons or locations where enhanced $PM_{2.5}$ is not driven as strongly by nighttime conditions.

The CEAMS deployment in Denver for the winter of 2019-2020 was hampered by inconsistencies in sampling locations, sampling times, and machine errors, which resulted in a limited dataset. Despite these setbacks, this deployment provided a novel dataset that informed us about possible interactions between meteorological and geographical variables, as well as the potential for low-cost AOD measurements to aid in the prediction of high-resolution spatiotemporal variability in $PM_{2.5}$. We recommend that future work mostly concerned about predicting high $PM_{2.5}$ days in cities consider using classification RF models that only work to predict "low" and "high" $PM_{2.5}$ days. This may provide more insight into how different spatiotemporal predictors play a role in elevated $PM_{2.5}$ events. We also recommend that future work incorporate a more thorough list of spatial predictors or use a hybrid approach that combines traditional LUR techniques and ML, such as Considine et al. (2021), to improve predictions of spatial variability for sub-city $PM_{2.5}$.

3.5 Funding and Data Availability

This research was supported by the National Aeronautics and Space Administration (NASA) grant number 80NSSC18M0120 and NASA Health and Air Quality Applied Sciences Team grant number 80NSSC21K0429. We are grateful for the CEAMS citizen scientists who participated in gathering the data used in this study as well as to Katelyn O'Dell, John Mehaffy, and the CSU Stats Helpdesk for their assistance in preparing data for Random Forest model analysis. The GEOS-FP data used in this study/project have been provided by the Global Modeling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center. All air quality data collected and used for this study during the CEAMS Denver, CO deployment are available at the following URL: https://hdl.handle.net/10217/233884.

Chapter 4

Investigating Disparities of $PM_{2.5}$ and NO_2 at US Public Schools[3]

4.1 Introduction

Air pollution poses significant health and economic burdens worldwide. Certain pollutants have garnered significant research attention including, but not limited to, fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), and ozone (O_3). $PM_{2.5}$, which has a variety of natural and anthropogenic sources, is considered the greatest environmental risk factor for premature death worldwide (Cohen et al., 2017; World Health Organization, 2013). NO_2 is toxic gas and a component of nitrogen oxides (NO_x), which are primarily produced through fuel combustion. Strong causal relationships have been found between $PM_{2.5}$/NO_2 pollution and cardiovascular/respiratory diseases, lung cancer, and all-cause mortality (Atkinson et al., 2018; Cohen et al., 2017; Faustini et al., 2014; Hoek et al., 2013; World Health Organization, 2013).

Although air pollution impacts people of all ages, it can be especially hazardous to children. Children are at greater risk due to still developing organs (Brockmeyer & D'Angiulli, 2016; Gehring et al., 2013; Kulkarni & Grigg, 2008); increased pollutant intake to body mass ratios (Brockmeyer & D'Angiulli, 2016); and generally spending more time outside (Bateson and Schwartz, 2008). Exposure to high concentrations of air pollutants during childhood has been linked to increased risk for developing or exacerbating respiratory diseases such as asthma, bronchitis, and pneumonia (Chatkin et al., 2021; Han et al., 2021; Madureira et al., 2015; van Zoest et al., 2020), as well as reduced lung function (Gehring et al., 2013). Studies have also found links between children's exposure to air pollution and diminished cognitive function, lower intelligence

[3]This work is in preparation for *GeoHealth*: Cheeseman, M., Ford, B., Anenberg, S., C., Cooper, M. J., Fischer, E., V., Hammer, M., S., Magzamen, S., Martin, R., V., van Donkelaar, A., Volckens, J., & Pierce, J. R. (2022). Strong disparities of air pollutants across racial/ethnic and poverty groups at US public schools.

quotient (IQ) scores (Mohai et al., 2011; Sunyer, 2008), and emerging mental health problems such as attention deficit/hyperactivity disorder, anxiety, and depression (Myhre et al., 2018; Roberts et al., 2019).

There are many important factors that dictate which children in the US are exposed to high levels of air pollution but the location of school that each child attends is important (e.g. Mohai et al. (2011)). There are over 54 million children at US public and private schools (National Center for Education Statistics, 2021) and they spend an average of 6.64 hours a day in school for 180 days a year (i.e. 1200 hours a year) (NCES Schools and Staffing Survey, 2008). Schools are often in proximity to heavily trafficked roads and commercial areas (Shaori et al., 2019) and students generally have little choice in what school they attend (Kweon et al., 2018). While at school, students spend time in outdoor environments during activities like recess, physical education class, free time between class, and extracurricular activities. Outdoor air pollutants can also infiltrate indoor environments (e.g. Habre et al., 2014; Reche et al., 2015). Furthermore, depending on the age and building quality of the school, children may also attend class in rooms with poor ventilation. Students also spend time commuting to and from school, oftentimes aboard heavy polluters such as diesel-powered buses.

As a result, studies have found that schools can be a crucial site of exposure for children. One study of personal exposure for 250 students in London found that children are exposed to high levels of $PM_{2.5}$ and NO_2 at schools and even more when commuting to and from school (Varaden et al., 2018). Proximity to major roadways and high road density generally have positive correlations with air pollutant concentrations or related health outcomes at schools, while proximity to greenspaces have negative correlations, by acting as a buffer against air pollutants (Amram et al., 2011; Dadvand et al., 2015; Hauptman et al., 2020; Kweon et al., 2018; Requia et al., 2021; Zeng et al., 2020). Air pollution levels near schools have been linked to lower student test scores, grade point averages, and attention retention (Mohai et al., 2011; Sunyer, 2008). Poorer air quality at schools has also been associated with chronic absenteeism, which is a critical health performance indicator, since absenteeism is often due to health reasons (such as asthma), as opposed to defiant

behavior (Ready, 2010). Additionally, chronic absenteeism is associated with poorer academic performance in kindergarten to 12th grade (Ready, 2010). Finally, a longitudinal study of 16,000 US primary school students revealed that higher concentrations of hazardous air pollutants (HAPs) were associated with lower scores in reading, science, and math (Grineski et al., 2020).

The burden of air pollution often falls disproportionately on the most under-priveleged communities in our society, both in terms of wealth and race/ethnicity. There is evidence that lower economic positions are associated with higher levels of exposure to both NO_2 and $PM_{2.5}$ (Amram et al., 2011; Clark et al., 2014; Fairburn et al., 2019; Fecht et al., 2015). Fecht et al., 2015 found that the poorest decile of people in England had NO_2 levels that were 7.9 μg m^{-3} higher, on average, than the wealthiest decile of citizens. The same study found that the poorest decile had $PM_{2.5}$ levels elevated by 2.6 μg m^{-3}, on average, compared to the wealthiest decile (Fecht et al., 2015). A meta-analysis over Western Europe found lower social dimensions (e.g., economic position, gender, ethnicity, and occupation) were associated with higher air pollution levels. However, when they investigated specific ethnicities, they found mixed results because some minoritized groups had higher exposure to air pollutants while others had lower relative to the majority population (Fairburn et al., 2019). Fairburn et al. (2019) also found that the most privileged and least privileged societal groups in Europe can be exposed to similar pollution levels compared to other social groups because they tend to concentrate in metropolitan areas. This can result in a U-shaped exposure curve across socioeconomic levels (Fairburn et al., 2019; Walker et al., 2003). However, even within city and regional studies, the lowest income groups still were associated with the highest levels of pollution. Studies have found similar results in the US and Canada. Gray et al. (2013) found that lower socioeconomic status and a greater proportion of minoritized groups were associated with higher levels of $PM_{2.5}$ for census tracts in North Carolina. Gray et al. (2013) found that the opposite was true for O_3, however. Pinault et al. (2016) found correlations between social/material deprivation and higher NO_2 exposure in the three largest cities in Canada. Recent studies have shown that air pollution contributes to the disparities in race of preterm births (Ben-

marhnia et al., 2017, Heo et al, 2019) and low birth weight of infants (Heo et al., 2019) in the US.

Understanding the disparities in air pollution that children of various racial/ethic and socioeconomic backgrounds are exposed to at schools in the US is crucial to future school siting plans and adequately devising policies to combat inequalities that currently exist. Past studies of disparities in air pollutants at public schools have focused on school districts in California (Gaffron & Niemeier, 2015; Green et al., 2004; Pastor et al., 2004, 2006, 2002) and Utah (Collins and Grineski, 2020; Mullen et al., 2020). These studies found that modeled average pollutant concentrations were higher (Morello-Frosch et al., 2002) or the proximity to polluters was more likely at schools that had higher percentages of Black and African American, Hispanic, and multi-ethnic students (Chakraborty & Zandbergen, 2007; Green et al., 2004; Maantay, 2002). Gaffron and Niemeier (2015) found that $PM_{2.5}$ emissions from road traffic were positively correlated with the percentage of students eligible for subsidized meals, which is often used as a proxy for poverty in school (and a measure we use in our analysis). Another study co-located the schools and homes of individual students in Orange County, FL to nearby major point sources, small area emitters, and major roadways, and found that Black and Hispanic children were significantly more likely to be in proximity to major polluters (Chakraborty and Zandbergen, 2007). Similar disparities have been found at schools outside the US such as in Canada (Amram et al., 2011; Batisse et al., 2017) and the United Kingdom (Shaori et al., 2019). Air pollution inequalities at schools have also been found to be dynamic due to changing weather conditions. For example, Mullen et al. (2020) found that public schools in Salt Lake County, Utah that served a higher proportion of low-income students had disproportionately higher concentrations of $PM_{2.5}$ under clean air conditions and during moderate strength persistent cold air pool events, which consist of cold air and high pressure systems trapping air pollutants near the surface. However, during strong cold air pool events when $PM_{2.5}$ was especially high over the region, disparities in $PM_{2.5}$ exposure disappeared.

To our knowledge, few studies have investigated pollution disparities at schools across the US nationally. Recently, the first nationwide study on air pollution disparities between schools focused

on airborne neurotoxins (Grineski and Collins, 2018). This study used the Environmental Protection Agency's (EPA) National Air Toxics Assessment (NATA), which is based on state and federal emissions data, a chemical transport model, and a dispersion model that produces census-tract level concentrations of HAPs. They found that students attending schools considered to be "high risk" for neurotoxin concentrations were significantly more likely to be Hispanic, Black, or Asian/Pacific Islander, and eligible for subsidized meals. As a result, they found that students attending high risk schools were much less likely to be white or races other than those in the last sentence. An analogous work to Grineski and Collins (2018) has not been done for two crucial pollutants: $PM_{2.5}$ and NO_2. Furthermore, census-tract modeled concentrations may lack the spatial resolution to capture the heterogeneity of air pollution gradients that could cause disparities between nearby schools. The recent development of high-resolution $PM_{2.5}$ (Hammer et al., 2020) and NO_2 (Anenberg et al., 2021; Cooper et al., 2020) products allow us to better understand the intersection of race, poverty, and air pollution. We hope our analysis provides motivation and guidance for future public school siting priorities and increased awareness of the air pollutant disparities that currently exist at US public schools.

4.2 Methods

We retrieved general information, demographics, and financial information for 85150 public schools for the continental United States (CONUS) from the National Center for Education Statistics (NCES). The NCES dataset can be found and indexed using the Elementary/Secondary Information System (ElSi) table generator tool (https://nces.ed.gov/ccd/elsi/tableGenerator.aspx). The NCES data has been used for several past studies investigating disparities in various pollutants near public schools (e.g. Collins et al., 2019; Grineski and Collins, 2018; Mullen et al., 2020). The NCES data is based on data retrieved by the Common Core of Data, which is run by the US Department of Education and retrieves financial and non-financial information on US public schools, school districts, and state education agencies every year. Private School information was also collected, but as no geographic coordinates were provided, they could not be integrated into

our analysis. We used data collected for the 2019-2020 school year. We found the fraction of white and non-white students at each school and the fraction of students eligible for free and reduced lunch. We limit our analysis to only schools with at least 100 students. We use the fraction of students eligible for free or reduced lunch as a proxy for poverty levels in schools, which has been commonly used in previous research (Grinesk and Collins, 2018, Mohai et al., 2011, Morello-Frosch et al., 2002; Pastor et al., 2006). This proxy is not a perfect measure of poverty (National Center for Education Statistics, 2015) for multiple reasons. For example, eligibility for the free and reduced lunch program in the US is based on poverty levels set at the federal level, which means that eligibility for this program relative to regional cost of living and standard of living can vary across the US. The NCES also categorizes school urbanity as City (though we define here as Urban), Suburban, Town, or Rural locations, and each of these categories have subcategories of large, mid-size, and small (Urban and Suburban) and fringe, distant, and remote (Town and Rural). We use NCES categories of urbanity as well as the demographics and poverty levels of students at each school to investigate disparities in air pollutant concentrations.

To study disparities in particulate matter across US public schools, we use annually averaged $PM_{2.5}$ concentrations from Hammer et al. (2020), which combined AOD measurements from NASA MODIS, MISR, and SeaWIFs with the GEOS-Chem chemical transport model (Figure C.1). The $PM_{2.5}$ estimates were then calibrated regionally to surface observations using a geographically weighted regression model. Data is provided at the finest resolution possible (0.01° × 0.01°), though particulate matter gradients at this resolution are not truly resolved due to data inputs at coarser resolution (Hammer et al., 2020). We analyze 2018 for our study since this was the last available year. As these were gridded hybrid-model datasets, we co-located each school to the nearest gridcell.

We used a variety of national, annually averaged NO_2 products over CONUS for our analysis including datasets developed by Anenberg et al. (2021) and Cooper et al. (2020). We show results from the Anenberg et al. (2021) dataset in Section 4.3 since this dataset was at a higher resolution than Cooper et al. (2020). Also, the use of a land use regression (LUR) model allowed

Anenberg et al. (2021) to explicitly incorporate traffic data when developing their NO_2 estimates. The reliance on traffic data is important in our analysis since previous studies have shown that schools with higher poverty levels also tended to be closer to major roadways (Amram et al., 2011; Green et al., 2004). We present results using the Cooper et al. (2020) dataset as a sensitivity analysis in the Supplement. The Anenberg et al. (2021) NO_2 dataset contains annually averaged NO_2 estimates at a 1km resolution from 1990-2019 (Figure C.2). The NO_2 estimates were created using a combination of LUR model predictions and column density NO_2 observations from the Ozone Monitoring Instrument (OMI) satellite sensor. The Cooper et al. (2020) dataset used a chemical transport model in combination with satellite observations from the TROPOMI satellite (Figure C.3). We use data from 2019, the last year of data available, for our analysis and co-locate each school to the nearest gridcell.

4.3 Results

4.3.1 Nationwide Disparities in $PM_{2.5}$ and NO_2

Figure 4.1a shows the complementary cumulative distribution function (CDF) of NO_2 for students in CONUS public schools from each NCES racial/ethnic demographic category (except those who identify as 2 or more races). White and Native American students were the least likely to attend schools with ambient air quality levels that exceed the recently updated WHO guideline for annual average NO_2 concentrations (83% and 82% above the WHO guideline). Conversely, Black and African American (89% above the WHO guideline), Asian/Pacific Islander (95% above the WHO guideline), and Hispanic (94% above the WHO guideline) students were the most likely to attend schools with ambient NO_2 levels that exceed the WHO guideline.

Figure 4.1b shows the complementary CDF of $PM_{2.5}$ for students in CONUS public schools. $PM_{2.5}$ had a much steeper curve than NO_2 and shows that the vast majority of students in the US went to schools with annually averaged $PM_{2.5}$ concentrations below 8 μg m^{-3}. The WHO recently updated the guideline for annually averaged $PM_{2.5}$ to 5 μg m^{-3}. Thus, over 95% of every racial/ethnic demographic, except Native Americans, attended schools in locations where

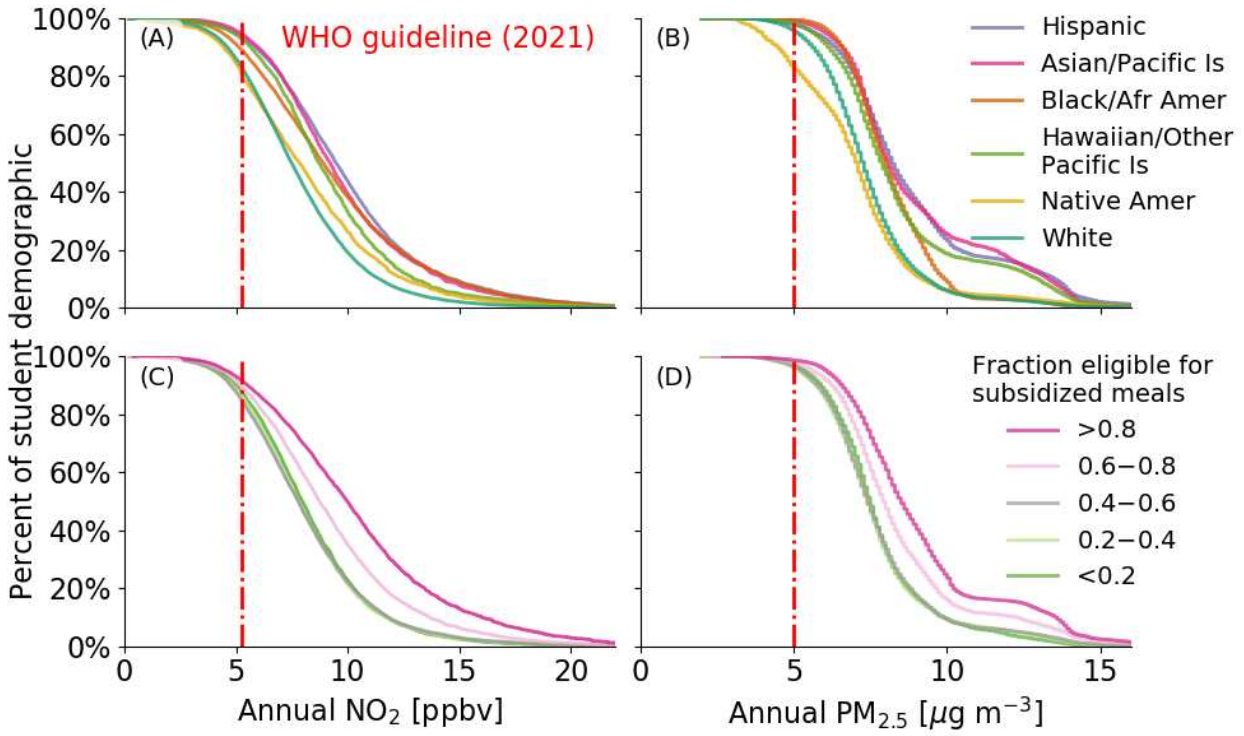

Figure 4.1: Complementary cumulative distribution functions (CDF) of annually averaged NO_2 (a) and $PM_{2.5}$ (b) at schools for students that belong to specific racial/ethnic demographics. Complementary CDFs of annually averaged NO_2 (c) and $PM_{2.5}$ (d) at schools of differing poverty levels, measured by the fraction of students eligible for subsidized meals at each school. Each plot shows the percentage of the students that attend schools where the co-located annually averaged mean of each pollutant is above a given concentration. The current WHO guidelines for annually averaged NO_2 and $PM_{2.5}$ are shown in red dashed lines.

the ambient concentrations of $PM_{2.5}$ were above the WHO guideline. If we repeat this analysis relative to the previous WHO guideline for $PM_{2.5}$ (i.e., 10 μg m^{-3}), we find that Asian/Pacific Islander (25% above the WHO guideline), Hispanic (23% above the WHO guideline), and Native Hawaiian/Other Pacific Islander (18% above the WHO guideline) students were the most likely to attend schools that exceeded the previous WHO guideline. Black and African American (9%), Native American (6%), and white students (4%) were the least likely to attend schools with ambient $PM_{2.5}$ concentrations above the previous WHO guideline. However, Black and African American students attend schools with $PM_{2.5}$ concentrations above Native American and white students for concentrations below ~8 μg m^{-3}. The steep drop-off in concentrations for Black and African American students likely occurs because there are more Black and African American students in the rural and suburban South and they are less concentrated in the Central Valley of California (Figure C.4-C.8), which is the region with the highest concentrations of annually averaged $PM_{2.5}$ in the Hammer et al. (2020) dataset (Figure C.1).

We also calculated the complementary CDF of students' co-located NO_2 (Figure 4.1c) and $PM_{2.5}$ (Figure 4.1d), where students were grouped by the fraction of students eligible for subsidized meals at the schools that each student attends. We found that 92% of students that attended schools with the highest poverty levels (i.e., the percentage of students eligible for subsidized meals was greater than 80%) attended schools with ambient NO_2 concentrations above the WHO guideline (Figure 4.1c). In contrast, only ~85-88% of students that attended schools with moderate-to-low levels of poverty attended schools with NO_2 concentration above the WHO guideline. Similarly, 99% of students that attended the highest poverty schools were at schools with $PM_{2.5}$ concentrations above the 2021 WHO guideline, while 97% of the students that attended the lowest poverty schools were above the WHO guideline. This disparity may seem relatively insignificant, but it is a large number of actual students. Additionally, if we compare the levels of $PM_{2.5}$ between the richest and poorest schools relative to the previous WHO guideline for $PM_{2.5}$ (i.e., 10 μg m^{-3}), over 23% of the poorest schools were above the previous guideline, compared to only 9% of the richest

schools. Figure 4.1 demonstrates that there were differences in the air quality levels experienced by children attending public schools in the US across both on racial and economic lines.

We further investigated the intersection of poverty levels, locale categories, race, and pollutant concentrations for NO_2 and $PM_{2.5}$ around CONUS public schools in Figure 4.2. Figure 4.2a shows that the 25th, 50th, and 75th quartiles of NO_2 concentrations increased monotonically from the least to the most impoverished schools at urban locations. In contrast, quartiles of NO_2 at suburban, town, and rural schools overlapped considerably from the least to most impoverished schools. Schools that were the most white (greater than 65% white students, based on the 60th percentile nationwide) were more likely to be wealthier (depicted by box widths), and concentrated in suburbs, town, and rural areas, especially in the northeastern and midwestern US (Figure 4.2c). Schools that were the most non-white (fewer than 39.3% white students, based on the 40th percentile nationwide) were more likely to be poorer and disbursed in urban areas nationwide, as well as in suburban, town, and rural areas in the southeastern US, Texas, and California (Figure 4.2c). When we separate suburban schools from town and rural schools, we find that NO_2 concentrations at suburban schools generally increase with increasing poverty levels for mostly non-white schools and stay the same for mostly white schools (Figure C.9a). Additionally, NO_2 concentrations around schools in town and rural areas generally decrease with increasing poverty level (Figure C.10a). Thus, generally, the poorest urban schools had the highest distribution of NO_2 concentrations, while the poorest rural schools had the lowest distribution of NO_2 concentrations. At each poverty level and locale category, non-white schools were more likely to have higher concentrations of NO_2 than their white counter-part (Figure 4.2a). For example, around the poorest urban schools, the median NO_2 at mostly non-white schools was approximately 1.5 ppbv higher than mostly white schools. Similarly at suburban, town, and rural schools, the median NO_2 concentrations were approximately 2 ppbv higher at mostly non-white schools than at mostly white schools. As NO_2 concentrations are higher in urban areas, these locale categories exacerbate the disparities in ambient NO_2 concentrations across poverty levels. Thus, when we compared the median NO_2 concentration of the mostly non-white urban high-poverty schools against mostly white

suburban, town, and rural moderate-poverty schools, we found a disparity of approximately 4 ppbv (urban high-poverty schools ~50% higher) of annually averaged NO_2 and even larger disparities in the tails of the distributions.

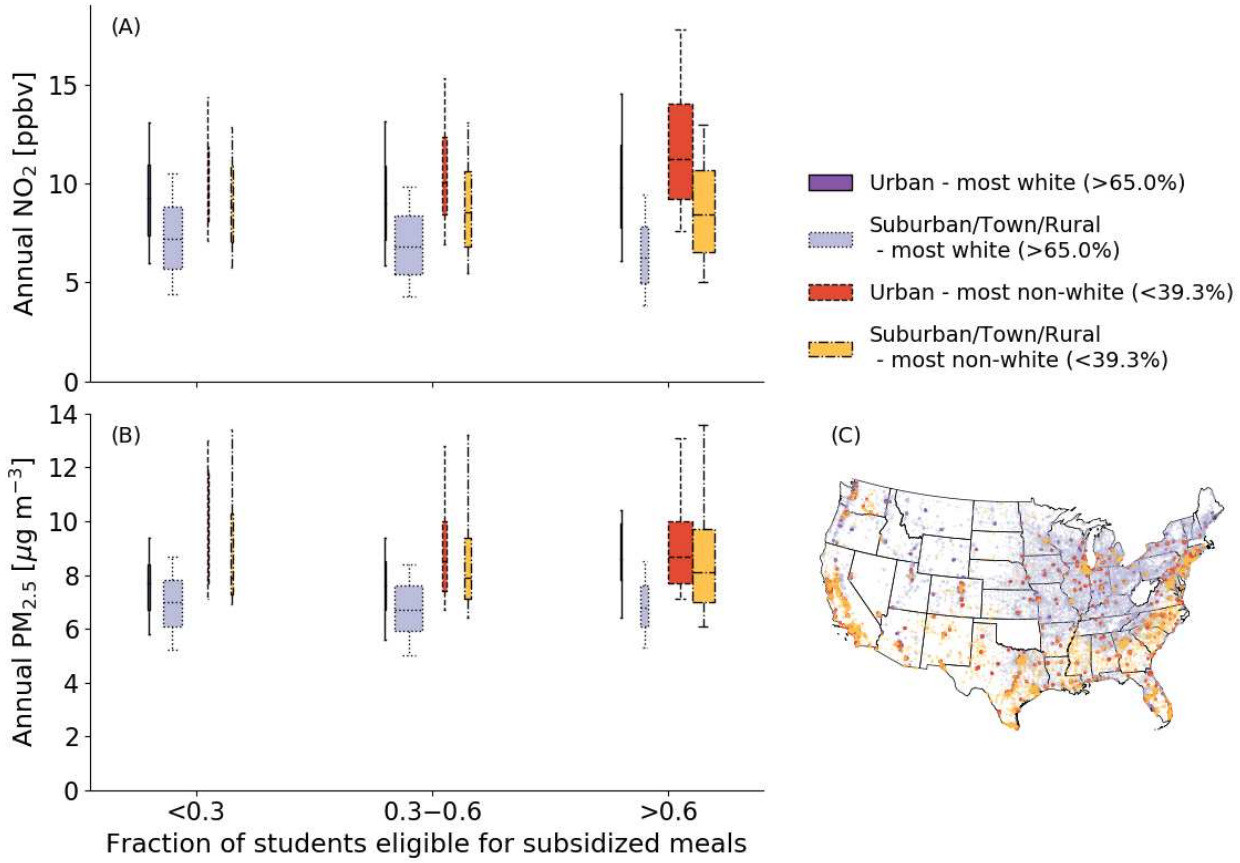

Figure 4.2: Boxplots of annually averaged (a) NO_2 and (b) $PM_{2.5}$ surface mixing ratios split into categories of poverty level, which is defined by the fraction of students eligible for free or reduced lunch. A greater fraction of free or reduced lunch indicates a higher level of poverty. Within each poverty-level bin, separate boxplots are shown for schools in urban and combined suburban, town, and rural locations for mostly white schools (>65% white students) and mostly non-white schools (<39.3% white students). The thresholds for mostly white and mostly non-white schools are based on the 60th and 40th percentile of the percentage of white students in all public schools across CONUS. The width of each boxplot is proportional to the number of schools in each distribution. (c) A map of schools that belong in each race and locale category.

Relative to NO_2, there was more overlap in $PM_{2.5}$ concentrations across poverty levels and between urban and suburban, town, and rural schools (Figure 4.2b). This is unsurprising since the chemical lifetime of NO_2 is much shorter than for $PM_{2.5}$, which allows $PM_{2.5}$ to become more regionally uniform. In Figure 4.2b, the median $PM_{2.5}$ at mostly white schools increased from

lowest to highest poverty level (0.9 μg m^{-3} increase). Contrastingly, the median $PM_{2.5}$ decreased at mostly non-white schools from lowest to highest poverty level (0.8 μg m^{-3} decrease), though there are two orders of magnitude more non-white schools in the highest poverty bin than the lowest poverty bin (Table C.1), and the tails of the distribution are similar. For both mostly white and mostly non-white schools, the concentrations of $PM_{2.5}$ at urban schools largely overlapped with those at suburban, town, and rural schools, but the mostly non-white schools had higher distributions of $PM_{2.5}$ than their mostly white counterparts within each poverty bin. Furthermore, similarly to NO_2, the largest disparities were the result of mostly white schools being concentrated in wealthier and more suburban, town, and rural areas, while mostly non-white schools were concentrated in the poorest and mostly urban and suburban areas (Figure 4.2c). Thus, the median $PM_{2.5}$ concentrations for mostly white schools in moderately impoverished schools were 2 μg m^{-3} lower (~25% decrease) than mostly non-white, urban schools.

Although, Figure 4.2 shows that nationwide disparities exist, in part, due to regional differences in the location of different racial/ethnic groups, we also found distinct patterns of disparities in different states across the US. To investigate how the disparities across race and poverty compare between states, we calculated the differences in mean NO_2 and $PM_{2.5}$ between the most white and least white schools (Figure 4.3a-b) and most impoverished and least impoverished schools (Figure 4.3c-d) in each state. We combined schools from urban, suburban, town, and rural locations in Figure 4.3. Again, we defined the "most white" and "most non-white" schools as those with greater than 65% white students and fewer than 39.3% white students, based on the 60th and 40th percentiles for the percentage of white students at schools nationwide. States in the Northeast and Midwest show the largest disparities in NO_2 across racial/ethnic lines (Figure 4.3a-b). For example, there were large disparities in ambient NO_2 concentrations between the most white and non-white schools in New York (8.5 ppbv), Illinois (5.9 ppbv), and Michigan (4.5 ppbv). Figure 4.3c displays the differences in mean NO_2 between "high poverty" and "low poverty" schools, which are schools that have a percentage of students eligible for subsidized meals above 75% or below 25%, respectively. Figure 4.3b shows similarly that Northeastern and Midwestern states

such as New York (7.7 ppbv), Illinois (4.4 ppbv), and Michigan (2.8 ppbv) have the strongest disparities between high and low poverty schools. On average, racial/ethnic disparities appear to be larger than just poverty disparities for most states, though these problems are intersectional and not independent of each other. Much of the disparities displayed for NO_2 are driven by racial divides across urban, suburban, town, and rural areas. We investigate regional disparities in NO_2 in more detail in Section 4.3.2.

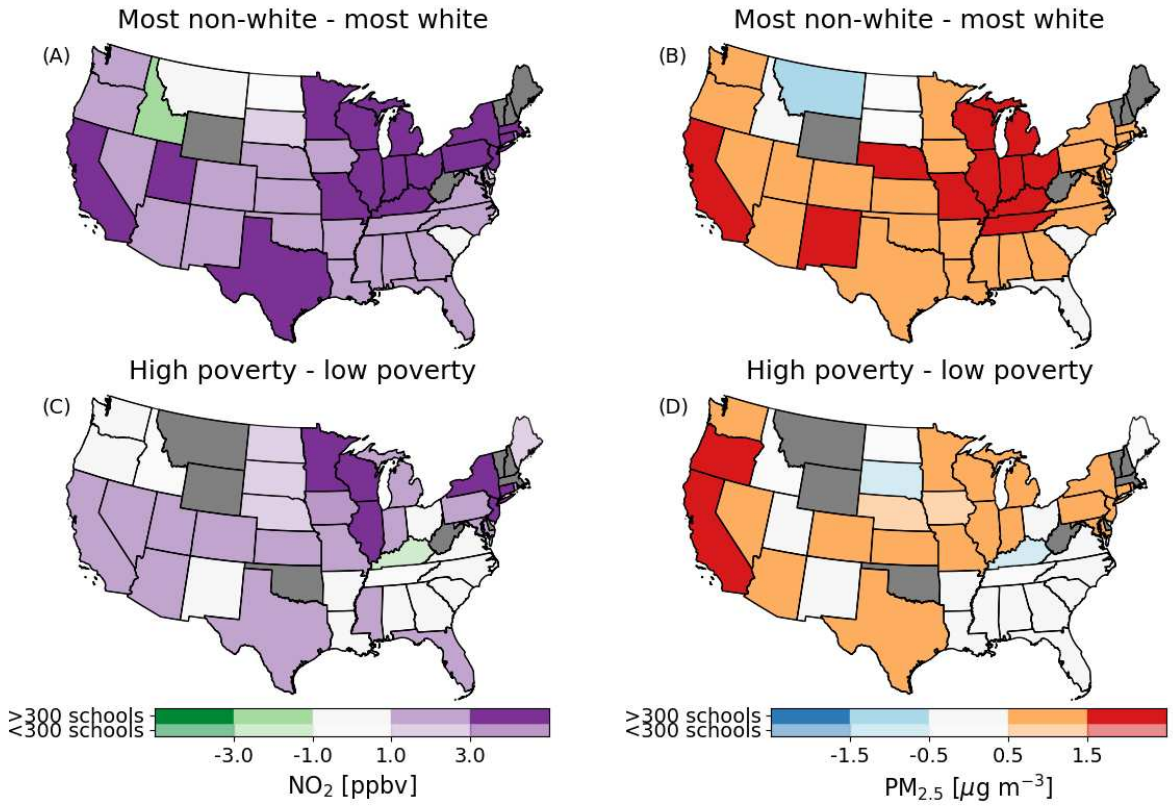

Figure 4.3: Difference in mean (a) NO_2 and (b) $PM_{2.5}$ between mostly non-white schools (fewer than 39.3% white students based on the 40th percentile nationwide) and mostly white schools (greater than 65% white students based on the 60th percentile nationwide). Difference in mean (c) NO_2 and (d) $PM_{2.5}$ between high poverty (greater than 75% students eligible for free or reduced lunch) and low poverty schools (fewer than 25% students eligible for free or reduced lunch). The colors are lighter shades if fewer than 300 total schools are used to find the difference between each category in each state. The colors are grey if there is no data for that state or there are fewer than 10 schools in either poverty or racial/ethnic categories.

$PM_{2.5}$ shows different patterns of disparities across race and poverty categories (Figure 4.3b and d) but most states still show increased $PM_{2.5}$ at schools with higher poverty levels and with

more non-white students. California had the largest difference in annually averaged $PM_{2.5}$ between mostly white and mostly non-white schools (2.1 μg m^{-3}) and the third largest across poverty levels (1.6 μg m^{-3}). Oregon and Rhode Island show the largest disparities across poverty levels (1.8 μg m^{-3} for both), though fewer than 300 schools were used to calculate the difference in Rhode Island. Michigan shows large disparities across racial/ethnic categories (1.9 μg m^{-3}) but a smaller difference across poverty levels (1.1 μg m^{-3}). The magnitudes of these disparities may be dependent on the thresholds used to divide racial/ethnic categories and poverty levels, but the patterns of disparity remain consistent.

4.3.2 Regional Disparities in NO_2

Concentrations of NO_2 are higher in cities, especially near sources such as major roadways and industrial areas (e.g. Anenberg et al., 2021). Thus, unsurprisingly, there were clear disparities in NO_2 concentrations at public schools across CONUS based on whether schools were in urban, suburban, town, or rural areas (Figure 4.2a). However, in certain regions, this urban/rural divide strongly intersects with poverty and race divisions (Figure 4.4). For example, in New York (Figure 4.4a), wealthier white public school students tended to be more dispersed across the state in suburban, town, and rural areas, while students from racially/ethnically minoritized and economically disadvantaged groups were heavily concentrated in urban areas, leading to disparities of up to ~11 ppbv NO_2 between mostly white and non-white schools. Note that in Figures 4.4-4.5, we define "most white" and "most non-white" based on statewide percentiles of the percentage of white students in each school, instead of the nationwide percentiles used in Figures 4.2-4.3.

In other regions, segregation in residential classification did not strictly coincide with economic status. In California, most schools were classified as urban or suburban, but NO_2 and $PM_{2.5}$ distributions still increased with increasing poverty levels and were generally higher for non-white schools regardless of locale category. Thus, in California, we saw large disparities across poverty levels that were not captured by segregation in residential classifications (Figure 4.4b). Finally, in some states, such as Florida (Figure 4.4c), there is a mix of white and non-white students in urban

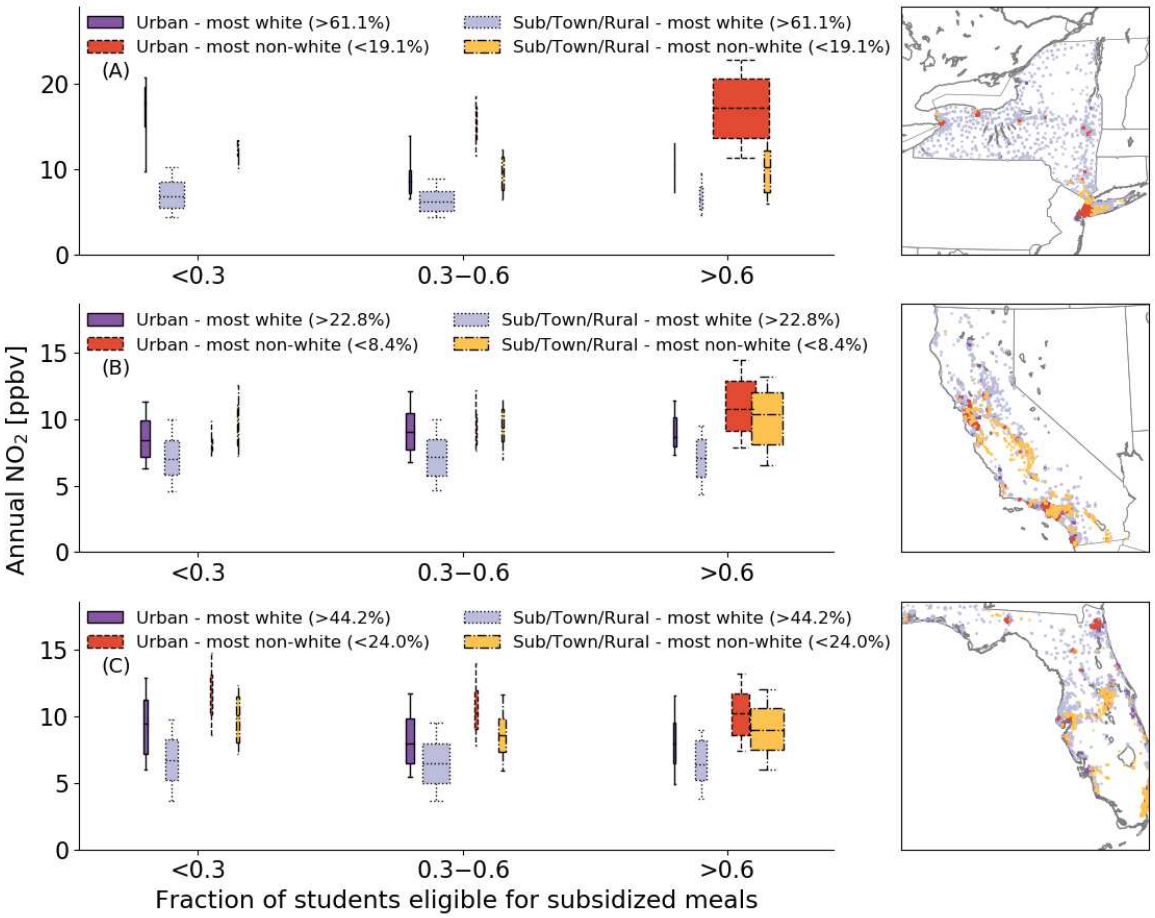

Figure 4.4: Boxplots of annually averaged NO_2 at mostly white schools (based on the 60th percentile of the percentage of white students in the state) and mostly non-white (based on the 40th percentile of the percentage of white students in the state) split into bins of poverty level in New York (a), California (b), and Florida (c).

areas around the state, especially in Tampa and Jacksonville, and yet, there are clear disparities across racial lines that do not also correspond to poverty lines. Thus, we find that there are three major types of disparities that appear to exist in NO_2 concentrations at US public schools in different regions: 1) wealthy white students in suburban, town, and rural environments have lower concentrations than poorer, racially/ethnically minoritized students concentrated in urban areas; 2) within many suburban/urban environments, wealthier, whiter schools have lower concentrations than corresponding poorer, less white schools; or 3) less white schools have higher concentrations regardless of poverty level. We further investigated these second and third points, by repeating our analysis only looking at schools in urban regions. This analysis shows that there are still air quality disparities across race and poverty categories. For example, in the Bay Area of California, we found higher concentrations of NO_2 at mostly non-white schools located in Chinatown, the Mission District, and Oakland, while more white schools had lower NO_2 in the Richmond district, which is near the greenspace of Golden Gate Park, and other surrounding areas (Figure C.11). The analyses of each state separately shows that disparities across racial/ethnic and poverty lines are complex and generally ubiquitous in the US.

4.3.3 Regional Disparities in $PM_{2.5}$

Consistent with the national analysis, there were similar but smaller disparities in $PM_{2.5}$ concentrations, relative to NO_2, in New York (Figure 4.5a), California (Figure 4.5b), and Florida (Figure 4.5c), likely due to $PM_{2.5}$'s longer chemical lifetime. In New York, the largest disparities in $PM_{2.5}$ were a result of racial/ethnic stratification across urban and suburban, town, and rural areas, which led to disproportionately high concentrations of $PM_{2.5}$ at mostly non-white urban schools (1.0-1.5 μg m^{-3} higher median $PM_{2.5}$ than mostly white suburban, town, and rural schools). In California, the most white and non-white urban schools had similar median concentrations of $PM_{2.5}$ and both had higher median $PM_{2.5}$ with increasing levels of poverty (~1.3-1.5 μg m^{-3} increase from lowest to highest poverty levels). However, the most non-white schools in suburban, town, and rural areas were consistently higher (~13.6 μg m^{-3}) than their white coun-

terparts (~11 μg m^{-3}). We found very little disparities of $PM_{2.5}$ concentrations in Florida (Figure 4.5c) across poverty levels. The largest difference across poverty levels occurred in mostly white, urban schools, where the wealthiest schools had slightly higher median concentration of $PM_{2.5}$ than the poorest schools (0.8 μg m^{-3} decrease in median $PM_{2.5}$). There were small disparities across racial/ethnic categories in suburban, town, and rural schools, resulting in a modest increase in median $PM_{2.5}$ at the poorest non-white schools compared to their white counterparts (0.8 μg m^{-3} increase).

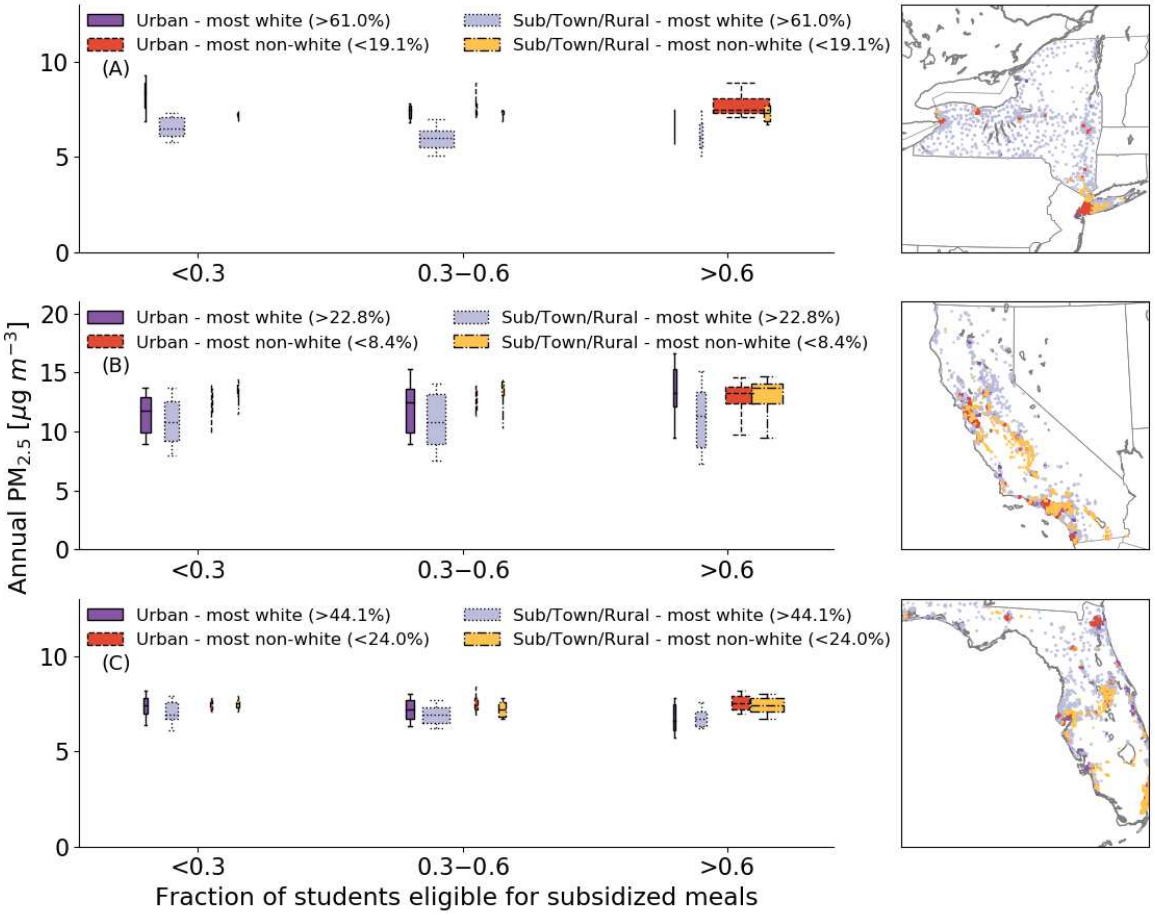

Figure 4.5: Boxplots of annually averaged $PM_{2.5}$ at mostly white schools (based on the 60th percentile of the percentage of white students in the state) and mostly non-white (based on the 40th percentile of the percentage of white students in the state) split into bins of poverty level in New York (a), California (b), and Florida (c).

4.4 Conclusions

Children spend a considerable amount of early life at schools in the US (NCES Schools and Staffing Survey, 2008), and while there, they are potentially being exposed to high levels of air pollutants. Greater than 95% of all US public school students are attending locations above the 2021 WHO guideline for ambient annually averaged $PM_{2.5}$ concentrations (~10% students above the 2005 WHO guideline) and greater than 87% are above the 2021 WHO guideline for annually averaged NO_2. Exposure to air pollutants such as $PM_{2.5}$ and NO_2 can have significant impacts on respiratory and cardiovascular health (Atkinson et al., 2018; Cohen et al., 2017; Faustini et al., 2014; Hoek et al., 2013; World Health Organization, 2013), especially in children (Brockmeyer & D'Angiulli, 2016; Gehring et al., 2013; Kulkarni & Grigg, 2008), and has been linked to higher levels of depression, anxiety, and ADHD (Myhre et al., 2018; Roberts et al., 2019). Thus, it is unsurprising that air pollution at schools has been linked to lower cognitive functioning, test and IQ scores, performance in math, science, reading, and overall performance. These impacts do not necessarily stop with impacts on school performance: instead, the lifetime work and productivity of students may be degraded by these higher exposures to air pollutants, which could have significant impacts on future generations of workers in the US.

Furthermore, the brunt of this pollution is not being shared equally across racial/ethnic and poverty lines. Instead, in most regions of the US, school children that attend schools with higher percentages of racially/ethnically minoritized students and higher levels of poverty are significantly more likely to be near higher concentrations of both $PM_{2.5}$ and NO_2. Much of these disparities appear to be due to the increased likelihood of economically disadvantaged and racially/ethnically minoritized children to live in urban areas, but we also find that even within urban environments, disparities still exist across racial/ethnic and poverty lines. This is likely because poorer and more non-white schools are more likely to be placed near major pollution sites (Chakraborty & Zandbergen, 2007; Green et al., 2004; Maantay, 2002). We did note that, nationally, the poorest town and rural schools had the lowest distribution of NO_2 but the poorest schools were more likely to be suburban or urban, which had the highest distribution of NO_2 (Figure C.10). Disparities nationally

do not occur only because certain demographics tend to live in different regions of the US (Figure 4.2c). Instead, within specific regions of the US, we found that schools with even slightly more white students were more likely to have better air quality (Figure 4.4-4.5). Finally, disparities in air pollutants at schools will likely only exacerbate already existing disparities in our society because poorer and more impoverished students will also suffer disproportionate impacts due to air quality on school and future job performance, as well as mental and physical health, which can in itself result in higher medical costs.

To our knowledge, this is the first investigation of NO_2 and $PM_{2.5}$ disparities across race and poverty in CONUS public schools though it builds on and is in agreement with work by Grineski and Collins (2018) and Mullen et al. (2020) that focused on disparities in atmospheric neurotoxins and $PM_{2.5}$ over Salt Lake City schools, respectively. Currently, there are no federal or specific guidelines to protect children from pollutants while attending schools (Sampson, 2012). Instead, the EPA releases school siting guidelines (EPA, 2011) that others can use voluntarily to inform decision making. Our work suggests that specific policies are needed to avoid future disparities in air pollution at schools, and certain states will have to face distinct challenges. For example, the stark racial segregation of students in New York state across rural/urban lines will likely be difficult to overcome with policy changes in future school sitings. More research is needed to devise sound policies that can lessen current disparities and avoid future disparities, though past research has shown that proximity to major roadways and industrial areas (Chakraborty and Zandbergen, 2007; Green et al., 2004; Kweon et al., 2018; Maantay 2002) are detrimental while proximity to green spaces (Amram et al., 2011) is beneficial. Future school siting guidelines should also take into account factors in addition to the pollutant levels of the possible school site. For example, children's exposure to pollution during long commutes or commutes through densely trafficked areas may diminish the benefits of siting schools in locations with better air quality (Wolfe et al., 2020).

There are several caveats to take into account for our study: 1) It is important to note that ambient concentrations of pollutants are different from personal exposure estimates, especially

when children do not spend all of their time at schools and spend more time indoors than out while at school. Thus, we did not measure US children's personal exposure to $PM_{2.5}$ and NO_2 at and within schools; instead, we used spatially gridded estimates of annual-mean concentrations of ambient $PM_{2.5}$ and NO_2 around current school locations. 2) The high resolution datasets of $PM_{2.5}$ and NO_2 have moderate uncertainties and spatial gaps in reference monitoring networks make consistent calibration and evaluation throughout CONUS difficult, especially in remote areas (Anenberg et al., 2021). 3) Fine scale gradients of $PM_{2.5}$ used in this study are not fully resolved because the model inputs were coarser than the output $PM_{2.5}$ resolution (Hammer et al., 2020). 4) Our sensitivity analysis using the Cooper et al. (2020) dataset revealed similar qualitative results but much lower concentrations of NO_2 on average across most of the US and, likely due to the coarser resolution, we saw smaller differences between urban and suburban, town, and rural NO_2 concentrations in California (Figure C.12-C.15). 5) We were unable to include private schools in our analysis, which account for roughly 10% of students from kindergarten to grade 12 (National Center for Education Statistics, 2021). It is unclear if including private schools would reinforce the patterns we have seen here or would increase the number of mostly white urban schools that are potentially exposed to high concentrations of pollution. 6) Finally, the fraction of students eligible for subsidized meals is an imperfect proxy for poverty and should not be confused with a measure for socioeconomic status, which requires knowledge of several financial and personal factors for each student.

Our study highlights the importance of investigating current and potential future disparities in environmental pollutants in the US. Future investigations of disparities in air pollution at US schools should consider the following points: 1) It is crucial that we have estimates of indoor air pollution and personal exposure estimates for children in US schools. 2) Investigation is needed into how age and socioeconomic factors of schools impact the infiltration of outdoor air into the indoor environment. 3) We need estimates of air pollutant concentrations and exposure during school commutes, especially on diesel powered buses. 4) In-depth analyses are needed to understand the many social, environmental, geographical, and financial factors that lead to disparities in

schools, so that they may be avoided in the future. Work such as Shaori et al. (2020) serves as an example. This could also be done using a machine learning approach that predicts air quality at schools using these measurable factors. 5) As anthropogenic air pollutants continue to decrease in upcoming decades, the importance of landscape fires for pollutant exposure at US schools will increase. Thus, we need to understand students' exposure to smoke at schools and whether smoke disproportionately affects various specific racial/ethnic and socioeconomic groups.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

Ambient $PM_{2.5}$ is the greatest environmental risk factor for global health, and yet, much of the world lacks sufficient regulatory networks to provide accurate, high-spatiotemporal resolution estimates of $PM_{2.5}$, which is crucial for population exposure estimates. Thus, this dissertation provides quantitative assessments of $PM_{2.5}$ and its relationship to environmental, geographical, meteorological, and socioeconomic variables with the goals of improving future predictions of $PM_{2.5}$ and contributing to our understanding of environmental injustices at US public schools.

In Chapter 2, we found that the number of smoke plume height observations and the number of observed smoke plume heights above co-located reanalysis boundary layer heights were greater by orders of magnitude in California, the Pacific Northwest, and the Northern Rockies. The disparities in the number of smoke plume observations may also suggest that small agricultural fires are being missed by the thermal observation technique in MAIAC, perhaps due to the low optical thickness of small plumes at 11 μm and cloud screening. Smoke plumes with higher AOD were also associated with higher smoke plume heights, which may be due to a physical tendency for more optically thick plumes being more bouyant, or due to a bias in the plume height observations for optically thick plumes. Finally, we found that $PM_{2.5}$ and the ratio of $PM_{2.5}$:AOD tended to be lower near smoke plumes that were higher than co-located boundary layer heights. This work suggests that plume height observations could be incorporated into future predictive models of $PM_{2.5}$, especially when smoke plume heights are determined near source fires.

In Chapter 3, we developed RF machine learning models to predict sub-city gradients of $PM_{2.5}$, finding that meteorological variables including boundary layer heights, wind speeds, and RH provided most of the prediction skill for $PM_{2.5}$. Contrastingly, AOD provided little additional skill for $PM_{2.5}$ prediction in wintertime Denver, especially for 24-hour $PM_{2.5}$ predictions. AOD was iden-

tified as a stronger predictor for hourly averaged $PM_{2.5}$, though it did not provide additional model skill overall when used as an additional predictor. Thus, for wintertime Denver, daytime-only AOD was disconnected from diurnal drivers of $PM_{2.5}$. To aid future investigations into fine-scale $PM_{2.5}$ predictions, we provide in-depth instructions of the RF technique, model tuning, and how choices in our methodology impacted our results.

In Chapter 4, we found that Hispanic, Asian/Pacific Islander, and Black or African American students were significantly more likely to attend schools above the WHO guidelines for ambient $PM_{2.5}$ and NO_2 concentrations than students that identified with any other racial or ethnic group. There were disparities in pollutant concentrations (especially NO_2) in most of CONUS and strong disparities clustered in the Northeast and Midwest regions. We identified three distinct patterns of disparities that occur in different regions of the US: 1) disparities driven mostly by residential segregation, where wealthier, more white schools were dispersed outside of urban areas and poorer, less white schools were clustered in urban areas; 2) disparities existing across poverty and racial/ethnic lines occurring even within urban and suburban environments; and 3) little disparities across poverty levels but disparities persisted across racial/ethnic demographics. This work contributes to a growing body of environmental justice literature that assesses the legacy of racial, ethnic, and economic bias in the US and the resulting disproportionate impacts that this legacy may be having on the health of at-risk communities.

Our work in Chapters 2-3 suggests that the incorporation of multiple aerosol and geophysical data products into a flexible ML framework could allow $PM_{2.5}$ predictions to account for various weaknesses of each individual dataset. For example, in Chapter 2, we found that there are likely circumstances where plume height, PBLH, and AOD could potentially be used in conjunction to predict $PM_{2.5}$. Furthermore, we found that our RF models in Chapter 3 were likely able to leverage the physical relationships between $PM_{2.5}$ and Denver meteorology and geography. These methods have already been used to fill-in spatial and temporal gaps in $PM_{2.5}$ observation networks (e.g. Di et al., 2016; Lightstone et al., 2017; Liu et al., 2018; Reid et al., 2015; Suleiman et al., 2019; Xi et al., 2015), but we will likely see great improvements in predictive models in the

near future. However, our work emphasizes the importance of understanding the autocorrelation present in machine learning datasets. It is also crucial to understand the biases in observations used in machine learning, as they may impact the interpretability of the results.

Our work also suggests that AOD likely should not be used alone for predicting $PM_{2.5}$ in many regions and circumstances. For example, during landscape fire events, we found large variabilities in $PM_{2.5}$:AOD, especially when the plume height was low. We also found that satellite AOD and co-located *in situ* AOD was poorly correlated with $PM_{2.5}$ over wintertime Denver, though this is not surprising since AOD and $PM_{2.5}$ are generally less correlated during wintertime conditions and when daily mean $PM_{2.5}$ variability is driven by nighttime conditions. Finally, our work in Chapter 4 only emphasizes the importance of improving our ability to estimate $PM_{2.5}$ and other air pollutants at a high spatiotemporal resolution. Without future improvements to air pollutant concentration predictions, we will be limited in our ability to: 1) identify various emission sources; 2) accurately estimate health and economic costs of air pollutants, especially in a changing climate; 3) identify current injustices in air pollutant exposure and their causes; 4) devise policies to reduce future pollution exposures and subsequent injustices in our society.

5.2 Recommendations for future work

5.2.1 Concerning predictions of the $PM_{2.5}$ during landscape fires

As discussed in Chapter 2, smoke plume heights have the potential to aid in $PM_{2.5}$ predictions during landscape fire events, especially near the fire. However, most current satellite products are polar-orbiting satellites, and thus limited to 1-2 overpasses per day in most cloud-free regions of the world. This limits our ability to capture smoke plume characteristics at a sub-daily temporal resolution needed for reliable $PM_{2.5}$ prediction near fires, since the direction and height of smoke plumes can change rapidly due to changes in prevailing winds, fire intensity, and boundary layer dynamics; in particular, fire intensity has a strong diurnal cycle. With the implementation of thermal smoke-height retrievals from current geostationary satellites (GOES-West and East) as well as future retrievals from upcoming geostationary platforms, there is potential to improve our

understanding of $PM_{2.5}$:AOD. Geostationary satellites can retrieve sub-hourly estimates of AOD and plume height, which could be incorporated into a ML framework to predict surface $PM_{2.5}$ or $PM_{2.5}$:AOD under smoke plumes. Improving prediction accuracy of $PM_{2.5}$:AOD during smoke episodes could improve worldwide estimates of exposure in landscape fire smoke and the resulting health impacts.

5.2.2 Intersectionality in poverty, race, ethnicity, and air pollution

Our results in Chapter 4 suggest large disparities in ambient air pollution along poverty and racial/ethnic lines at US public schools, yet there is more research needed to better understand the nature, scale, and impact of these disparities, and, crucially, how to reduce them. Given the potentially large impacts that air pollution may be having on US schoolchildren, future research should focus on capturing student exposure to indoor air pollution at school. Furthermore, since the US lacks mandatory federal guidelines for future US public school sitings and for reducing current air pollution exposures, research should be directed toward providing solutions that can be adopted on smaller scales than the federal level. Additionally, future studies of pollution exposure at schools should consider investigating: 1) how much ambient air pollution dictates the indoor air of schools and how that changes based on physical and socioeconomic characteristics of each school (e.g. age of the school, demographics, poverty levels, HVAC system); 2) student exposure to pollutants during commutes to/from school, especially in heavily trafficked areas or on diesel-powered school buses, which could be accomplished through a mobile monitoring campaign; and 3) differences in disparities based on student age, giving special attention to the youngest schoolchildren (middle school age and below) since they are more susceptible to health impacts from air pollution.

In addition to investigating personal exposure to air pollutants at schools, insight may be gained by quantifying how disparities have developed or changed over time. The $PM_{2.5}$ dataset (Hammer et al., 2019) used in Chapter 4, for example, is available from 1991 to the present. Thus, our method in Chapter 4 could be applied to quantify trends in air pollution disparities over CONUS for the last two decades, and we may consider this before submitting for peer review. Furthemore, special

attention could be given to school districts that have adopted the EPA's voluntary school siting guidelines (EPA, 2011) by comparing disparities before and after these policies were adopted. Investigations should also consider using a combination of regulatory and low-cost sensor networks to quantify community-level pollution disparities across racial/ethnic and poverty lines. Using a high density network of monitors taking sub-hourly measurements could provide crucial insights into the nature and patterns of pollution disparities, especially since regulatory networks routinely miss pollution (Reuters, 2020). For example, a future study could use PurpleAir monitors in Los Angeles county to investigate diurnal patterns of pollution hotspots, identify potential sources, and quantify how they relate to the socioeconomic characteristics of various schools or, given data availability, neighborhoods. However, it should be noted that PurpleAir monitors are less likely to be located in low-income and vulnerable communities. Therefore, it will also be essential for future research to prioritize increasing the monitoring capabilities in these communities so that we can more accurately assess their pollution patterns.

The influence of smoke on air pollution and health disparities is an open area of research and has yet to be investigated at schools. There is no clear mechanism for landscape fire smoke to disproportionately impact minoritized or socioeconomically disadvantaged groups, but regional patterns in where various groups live and attend school may overlap with regional patterns of smoke abundance. This is especially true in California, where wildfires are common and there is a large proportion of minoritized people. However, smoke has large impacts on people's health in the eastern US as well as the West (O'Dell et al., 2021), which may negate any disproportionate impacts that smoke may have on minoritized groups in California. Even if smoke does not impact regions with high proportions of minoritized groups more than others, differences in the quality of building materials and filtration systems of homes and schools that belong to minoritized groups compared to the majority population may create disparities in health outcomes, as indoor air may be worse during smoke events in lower-quality buildings.

Future studies should also consider using interpretable ML techniques (such as RFs) to derive quantitative estimates of driving factors in environmental disparities at US schools and within

communities in general. For example, predictive ML models could be built using socioeconomic, geographic, and demographic information of each school (or each census block) to predict pollution levels and discern prevailing patterns of disparities in $PM_{2.5}$, NO_2, and O_3. Our work in Chapter 3 and 4 also suggests that there is great opportunity to leverage citizen science and low-cost networks to address environmental injustices. Our results in Chapter 2 show that low-cost sensors, despite having greater noise and biases than traditional reference measurements, are able to capture temporal variability in surface $PM_{2.5}$ that resulted in similar interpretations of meteorological drivers over wintertime Denver. There are many non-profit and community organizations that are working to improve the lives of low-income and minoritized people. By developing partnerships between academics, nonprofit organizations, and community leaders, there are critical opportunities for improving air quality monitoring in historically under-monitored areas of the US and developing practices and policies that help mitigate the health impacts of air pollution disparities.

Finally, to understand the full breadth of potential health impacts that air pollution has on US children, future work should also quantify the combined exposure of children to multiple pollutants including various HAPs and the EPA's 6 criteria pollutants (NO_2, ground-level O_3, particulate matter, carbon monoxide, sulfur dioxide) both at home and school. Quantifying the personal exposure of children around the US to multiple pollutants and their combined health impact may portray varying patterns of disparities across racial, ethnic, and poverty groups due to regional, local, and seasonal differences in sources of the air pollutants and the resulting exposures. Considering multiple pollutants may also show that the combined health impacts due to pollutants from different sources either exacerbate pollution disparities, similar to those we investigate here, or diminish disparities due to different populations being exposed to different pollutants.

References

Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., ... Wennberg, P. O. (2011). Emission factors for open and domestic biomass burning for use in atmospheric models. *Atmospheric Chemistry and Physics*, *11*(9), 4039–4072. URL https://www.atmos-chem-phys.net/11/4039/2011/ doi: 10.5194/acp-11-4039-2011

Amante, C., & Eakins, B. W. (2009). ETOPO1 Global Relief Model converted to PanMap layer format. *NOAA-National Geophysical Data Center*. URL https://doi.pangaea.de/10.1594/PANGAEA.769615 doi: 10.1594/PANGAEA.769615

Amram, O., Abernethy, R., Brauer, M., Davies, H., & Allen, R. W. (2011). Proximity of public elementary schools to major roads in Canadian urban areas. *International Journal of Health Geographics*, *10*(1), 68. URL https://doi.org/10.1186/1476-072X-10-68 doi: 10.1186/1476-072X-10-68

Anenberg, S., Mohegh, A., Goldberg, D. L., Brauer, M., Burkart, K., Hystad, P., ... Wozniak, S. (2021). *Long-term trends in urban NO2 concentrations and associated pediatric asthma cases: estimates from global datasets* [preprint]. URL http://www.essoar.org/doi/10.1002/essoar.10506660.1 doi: 10.1002/essoar.10506660.1

Anenberg, S. C., Achakulwisut, P., Brauer, M., Moran, D., Apte, J. S., & Henze, D. K. (2019). Particulate matter-attributable mortality and relationships with carbon dioxide in 250 urban areas worldwide. *Scientific Reports*, *9*(1), 11552. URL https://www.nature.com/articles/s41598-019-48057-9 doi: 10.1038/s41598-019-48057-9

Atkinson, R. W., Butland, B. K., Anderson, H. R., & Maynard, R. L. (2018). Long-term Concentrations of Nitrogen Dioxide and Mortality. *Epidemiology (Cambridge, Mass.)*, *29*(4), 460–472. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5991178/ doi: 10.1097/EDE.0000000000000847

Austin, E., Novosselov, I., Seto, E., & Yost, M. G. (2015). Laboratory Evaluation of the Shinyei PPD42NS Low-Cost Particulate Matter Sensor. *PLoS One*, *10*(9), e0137789. URL

https://www.proquest.com/docview/1719284650/abstract/3F95DF72A41A4C27PQ/1 doi: http://dx.doi.org/10.1371/journal.pone.0137789

Baker, K., Woody, M., Tonnesen, G., Hutzell, W., Pye, H., Beaver, M., ... Pierce, T. (2016). Contribution of regional-scale fire events to ozone and PM2.5 air quality estimated by photochemical modeling approaches. *Atmospheric Environment*, *140*, 539–554. URL https://linkinghub.elsevier.com/retrieve/pii/S1352231016304654 doi: 10.1016/j.atmosenv.2016.06.032

Bank, W. (2016). *The Cost of Air Pollution: Strengthening the Economic Case for Action.*

Barkjohn, K. K., & Clements, A. L. (2020). *Development and Application of a United States wide correction for PM2.5; data collected with the PurpleAir sensor* (preprint). Aerosols/In Situ Measurement/Validation and Intercomparisons. URL https://amt.copernicus.org/preprints/amt-2020-413/ doi: 10.5194/amt-2020-413

Barregard, L., Sällsten, G., Gustafson, P., Andersson, L., Johansson, L., Basu, S., & Stigendal, L. (2006). Experimental Exposure to Wood-Smoke Particles in Healthy Humans: Effects on Markers of Inflammation, Coagulation, and Lipid Peroxidation. *Inhalation Toxicology*, *18*(11), 845–853. URL https://doi.org/10.1080/08958370600685798 doi: 10.1080/08958370600685798

Bateson, T. F., & Schwartz, J. (2008). Children's response to air pollutants. *Journal of Toxicology and Environmental Health. Part A*, *71*(3), 238–243. doi: 10.1080/15287390701598234

Batisse, E., Goudreau, S., Baumgartner, J., & Smargiassi, A. (2017). Socio-economic inequalities in exposure to industrial air pollution emissions in Quebec public schools. *Canadian Journal of Public Health*, *108*(5/6), E503–E509. URL http://www.proquest.com/docview/1991566858/abstract/70BC1B39B454D13PQ/1 doi: http://dx.doi.org.ezproxy2.library.colostate.edu/10.17269/CJPH.108.6166

Bell, M. L., & Dominici, F. (2008). Effect Modification by Community Characteristics on the Short-term Effects of Ozone Exposure and Mortality in 98 US Communities. *American journal of epidemiology*, *167*(8), 986–997. URL

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2430754/ doi: 10.1093/aje/kwm396

Bell, M. L., & Ebisu, K. (2012). Environmental Inequality in Exposures to Airborne Particulate Matter Components in the United States. *Environmental Health Perspectives*, *120*(12), 1699–1704. URL https://ehp.niehs.nih.gov/doi/10.1289/ehp.1205201 doi: 10.1289/ehp.1205201

Bellinger, C., Mohomed Jabbar, M. S., Zaïane, O., & Osornio-Vargas, A. (2017). A systematic review of data mining and machine learning for air pollution epidemiology. *BMC Public Health*, *17*(1), 907. URL https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-017-4914-3 doi: 10.1186/s12889-017-4914-3

Benmarhnia, T., Huang, J., Basu, R., Wu, J., & Bruckner, T. A. (2017). Decomposition Analysis of Black–White Disparities in Birth Outcomes: The Relative Contribution of Air Pollution and Social Factors in California. *Environmental Health Perspectives (Online)*, *125*(10). URL http://www.proquest.com/docview/2042708123/abstract/4354E07CF6544CEFPQ/1 doi: http://dx.doi.org.ezproxy2.library.colostate.edu/10.1289/EHP490

Bi, J., Belle, J. H., Wang, Y., Lyapustin, A. I., Wildani, A., & Liu, Y. (2019). Impacts of snow and cloud covers on satellite-derived PM2.5 levels. *Remote Sensing of Environment*, *221*, 665–674. URL https://linkinghub.elsevier.com/retrieve/pii/S003442571830556X doi: 10.1016/j.rse.2018.12.002

Bi, J., Wildani, A., Chang, H. H., & Liu, Y. (2020). Incorporating Low-Cost Sensor Measurements into High-Resolution PM $_{2.5}$ Modeling at a Large Spatial Scale. *Environmental Science & Technology*, *54*(4), 2152–2162. URL https://pubs.acs.org/doi/10.1021/acs.est.9b06046 doi: 10.1021/acs.est.9b06046

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. URL https://doi.org/10.1023/A:1010933404324 doi: 10.1023/A:1010933404324

Brey, S. J., Barnes, E. A., Pierce, J. R., Swann, A. L. S., & Fischer, E. V. (2021). Past Variance and Future Projections of the Environmental Conditions Driving West-

ern U.S. Summertime Wildfire Burn Area. *Earth's Future*, *9*(2), e2020EF001645. URL http://onlinelibrary.wiley.com/doi/abs/10.1029/2020EF001645 doi: 10.1029/2020EF001645

Brey, S. J., Barnes, E. A., Pierce, J. R., Wiedinmyer, C., & Fischer, E. V. (2018). Environmental Conditions, Ignition Type, and Air Quality Impacts of Wildfires in the Southeastern and Western United States. *Earth's Future*, *6*(10), 1442–1456. URL http://doi.wiley.com/10.1029/2018EF000972 doi: 10.1029/2018EF000972

Brockmeyer, S., & D'Angiulli, A. (2016). How air pollution alters brain development: the role of neuroinflammation. *Translational Neuroscience*, *7*(1), 24–30. URL http://www.degruyter.com/document/doi/10.1515/tnsci-2016-0005/html doi: 10.1515/tnsci-2016-0005

Bulot, F. M. J., Johnston, S. J., Basford, P. J., Easton, N. H. C., Apetroaie-Cristea, M., Foster, G. L., ... Loxham, M. (2019). Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment. *Scientific Reports*, *9*(1), 7497. URL https://www.nature.com/articles/s41598-019-43716-3 doi: 10.1038/s41598-019-43716-3

Burkhardt, J., Bayham, J., Wilson, A., Carter, E., Berman, J. D., O'Dell, K., ... Pierce, J. R. (2019). The effect of pollution on crime: Evidence from data on particulate matter and ozone. *Journal of Environmental Economics and Management*, *98*, 102267. URL https://www.sciencedirect.com/science/article/pii/S0095069619301901 doi: 10.1016/j.jeem.2019.102267

Castillo, M. D., Kinney, P. L., Southerland, V., Arno, C. A., Crawford, K., van Donkelaar, A., ... Anenberg, S. C. (2021). Estimating Intra-Urban Inequities in PM2.5-Attributable Health Impacts: A Case Study for Washington, DC. *GeoHealth*, *5*(11), e2021GH000431. URL http://onlinelibrary.wiley.com/doi/abs/10.1029/2021GH000431 doi: 10.1029/2021GH000431

Chakraborty, J., Maantay, J. A., & Brender, J. D. (2011). Disproportionate Proximity to Environmental Health Hazards: Methods, Models, and Mea-

surement. *American Journal of Public Health*, *101*(S1), S27–36. URL https://www.proquest.com/docview/906290163/abstract/E1C4049AD90A424BPQ/1

Chakraborty, J., & Zandbergen, P. A. (2007). Children at risk: measuring racial/ethnic disparities in potential exposure to air pollution at school and home. *Journal of Epidemiology and Community Health*, *61*(12), 1074–1079. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2465656/ doi: 10.1136/jech.2006.054130

Chatkin, J., Correa, L., & Santos, U. (2021). External Environmental Pollution as a Risk Factor for Asthma. *Clinical Reviews in Allergy & Immunology*. URL https://doi.org/10.1007/s12016-020-08830-5 doi: 10.1007/s12016-020-08830-5

Cheeseman, M., Ford, B., Rosen, Z., Wendt, E., DesRosiers, A., Hill, A. J., ... Pierce, J. R. (2021). Technical note: Investigating sub-city gradients of air quality: lessons learned with low-cost $PM_{2.5}$ and AOD monitors and machine learning. *Atmospheric Chemistry and Physics Discussions*, 1–30. URL https://acp.copernicus.org/preprints/acp-2021-751/ doi: 10.5194/acp-2021-751

Cheeseman, M., Ford, B., Volckens, J., Lyapustin, A., & Pierce, J. R. (2020). The Relationship Between MAIAC Smoke Plume Heights and Surface PM. *Geophysical Research Letters*, *47*(17), e2020GL088949. URL http://onlinelibrary.wiley.com/doi/abs/10.1029/2020GL088949 doi: 10.1029/2020GL088949

Cifelli, R., Doesken, N., Kennedy, P., Carey, L. D., Rutledge, S. A., Gimmestad, C., & Depue, T. (2005). The Community Collaborative Rain, Hail, and Snow Network: Informal Education for Scientists and Citizens. *Bulletin of the American Meteorological Society*, *86*(8), 1069–1077. URL https://www.jstor.org/stable/26221344

Clark, L. P., Millet, D. B., & Marshall, J. D. (2014). National Patterns in Environmental Injustice and Inequality: Outdoor NO2 Air Pollution in the United States. *PLOS ONE*, *9*(4), e94431. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0094431 doi: 10.1371/journal.pone.0094431

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., ... Forouzanfar, M. H. (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. *The Lancet*, *389*(10082), 1907–1918. URL https://linkinghub.elsevier.com/retrieve/pii/S0140673617305056 doi: 10.1016/S0140-6736(17)30505-6

Collins, T. W., & Grineski, S. E. (2019). Environmental Injustice and Religion: Outdoor Air Pollution Disparities in Metropolitan Salt Lake City, Utah. *Annals of the American Association of Geographers*, *109*(5), 1597–1617. doi: 10.1080/24694452.2018.1546568

Considine, E. M., Reid, C. E., Ogletree, M. R., & Dye, T. (2021). Improving accuracy of air pollution exposure measurements: Statistical correction of a municipal low-cost airborne particulate matter sensor network. *Environmental Pollution*, *268*, 115833. doi: 10.1016/j.envpol.2020.115833

Cooper, M. J., Martin, R. V., McLinden, C. A., & Brook, J. R. (2020). Inferring ground-level nitrogen dioxide concentrations at fine spatial resolution applied to the TROPOMI satellite instrument. *Environmental Research Letters*, *15*(10), 104013. URL https://doi.org/10.1088/1748-9326/aba3a5 doi: 10.1088/1748-9326/aba3a5

Dadvand, P., Rivas, I., Basagaña, X., Alvarez-Pedrerol, M., Su, J., De Castro Pascual, M., ... Nieuwenhuijsen, M. J. (2015). The association between greenness and traffic-related air pollution at schools. *Science of The Total Environment*, *523*, 59–63. URL https://linkinghub.elsevier.com/retrieve/pii/S0048969715003782 doi: 10.1016/j.scitotenv.2015.03.103

Danielson, J. J., & Gesch, D. B. (2011). *Global multi-resolution terrain elevation data 2010 (GMTED2010)* (Report No. 2011-1073). URL http://pubs.er.usgs.gov/publication/ofr20111073 doi: 10.3133/ofr20111073

Dechezleprêtre, A., Rivers, N., & Stadler, B. (2019). *The economic cost of air pollution: Evidence from Europe* (Tech. Rep.). Paris: OECD. doi: 10.1787/56119490-en

Delp, W. W., & Singer, B. C. (2020). Wildfire Smoke Adjustment Factors for Low-Cost and Professional PM2.5 Monitors with Optical Sensors. *Sensors*, *20*(13), 3683. URL https://www.mdpi.com/1424-8220/20/13/3683 doi: 10.3390/s20133683

Dennison, P. E., Brewer, S. C., Arnold, J. D., & Moritz, M. A. (2014). Large wildfire trends in the western United States, 1984–2011. *Geophysical Research Letters*, *41*(8), 2928–2933. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL059576 doi: 10.1002/2014GL059576

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., & Schwartz, J. (2016). Assessing PM $_{2.5}$ Exposures with High Spatiotemporal Resolution across the Continental United States. *Environmental Science & Technology*, *50*(9), 4712–4721. URL https://pubs.acs.org/doi/10.1021/acs.est.5b06121 doi: 10.1021/acs.est.5b06121

Engel-Cox, J. A., Hoff, R. M., & Haymet, A. (2004). Recommendations on the Use of Satellite Remote-Sensing Data for Urban Air Quality. *Journal of the Air & Waste Management Association*, *54*(11), 1360–1371. URL https://www.tandfonline.com/doi/full/10.1080/10473289.2004.10471005 doi: 10.1080/10473289.2004.10471005

EPA. (2011). *School Siting Guidelines.* URL https://www.epa.gov/sites/default/files/2015-06/documents/school_siting_guidelines-2.pdf

EPA. (2017). *2017 National Emissions Inventory (NEI) Data.* URL https: //www.epa.gov/air-emissions-inventories/2017-national-emissions-inventory-nei-data.

Evelyn, J. (1976). *Fumifugium.* URL http://books.google.com/books?id=jKY_AQAAMAAJ

Fairburn, J., Schüle, S. A., Dreger, S., Karla Hilz, L., & Bolte, G. (2019). Social Inequalities in Exposure to Ambient Air Pollution: A Systematic Review in the WHO European Region. *International Journal of Environmental Research and Public Health*, *16*(17), 3127. URL https://www.mdpi.com/1660-4601/16/17/3127 doi: 10.3390/ijerph16173127

Faustini, A., Rapp, R., & Forastiere, F. (2014). Nitrogen dioxide and mortality: review and meta-analysis of long-term studies. *The European Respiratory Journal*, *44*(3), 744–753. doi:

10.1183/09031936.00114713

Fecht, D., Fischer, P., Fortunato, L., Hoek, G., de Hoogh, K., Marra, M., ... Hansell, A. (2015). Associations between air pollution and socioeconomic characteristics, ethnicity and age profile of neighbourhoods in England and the Netherlands. *Environmental Pollution*, *198*, 201–210. URL https://www.sciencedirect.com/science/article/pii/S0269749114005144 doi: 10.1016/j.envpol.2014.12.014

Ford, B., & Heald, C. L. (2016). Exploring the uncertainty associated with satellite-based estimates of premature mortality due to exposure to fine particulate matter. *Atmospheric Chemistry and Physics*, *16*(5), 3499–3523. URL https://acp.copernicus.org/articles/16/3499/2016/ doi: 10.5194/acp-16-3499-2016

Ford, B., Pierce, J. R., Wendt, E., Long, M., Jathar, S., Mehaffy, J., ... Volckens, J. (2019). A low-cost monitor for measurement of fine particulate matter and aerosol optical depth – Part 2: Citizen-science pilot campaign in northern Colorado. *Atmospheric Measurement Techniques*, *12*(12), 6385–6399. URL https://amt.copernicus.org/articles/12/6385/2019/ doi: 10.5194/amt-12-6385-2019

Ford, B., Val Martin, M., Zelasky, S. E., Fischer, E. V., Anenberg, S. C., Heald, C. L., & Pierce, J. R. (2018). Future Fire Impacts on Smoke Concentrations, Visibility, and Health in the Contiguous United States. *GeoHealth*, *2*(8), 229–247. URL http://onlinelibrary.wiley.com/doi/abs/10.1029/2018GH000144 doi: 10.1029/2018GH000144

Forouzanfar, M. H., Afshin, A., Alexander, L. T., Anderson, H. R., Bhutta, Z. A., Biryukov, S., ... Murray, C. J. L. (2016). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, *388*(10053), 1659–1724. URL https://linkinghub.elsevier.com/retrieve/pii/S0140673616316798 doi: 10.1016/S0140-6736(16)31679-8

Freitas, S. R., Longo, K. M., & Andreae, M. O. (2006). Impact of including the plume rise of vegetation fires in numerical simulations of associated atmospheric pollutants. *Geophysical Research Letters*, *33*(17). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL026608 doi: 10.1029/2006GL026608

Gaffron, P., & Niemeier, D. (2015). School Locations and Traffic Emissions — Environmental (In)Justice Findings Using a New Screening Method. *International Journal of Environmental Research and Public Health*, *12*(2), 2009–2025. URL https://www.mdpi.com/1660-4601/12/2/2009 doi: 10.3390/ijerph120202009

Gan, R. W., Ford, B., Lassman, W., Pfister, G., Vaidyanathan, A., Fischer, E., ... Magzamen, S. (2017). Comparison of wildfire smoke estimation methods and associations with cardiopulmonary-related hospital admissions. *GeoHealth*, *1*(3), 122–136. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GH000073 doi: 10.1002/2017GH000073

Gao, M., Cao, J., & Seto, E. (2015). A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China. *Environmental Pollution*, *199*, 56–65. URL https://www.sciencedirect.com/science/article/pii/S0269749115000160 doi: 10.1016/j.envpol.2015.01.013

Gehring, U., Gruzieva, O., Agius, R. M., Beelen, R., Custovic, A., Cyrys, J., ... Brunekreef, B. (2013). Air Pollution Exposure and Lung Function in Children: The ESCAPE Project. *Environmental Health Perspectives*, *121*(11-12), 1357–1364. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3855518/ doi: 10.1289/ehp.1306770

Gillooly, S. E., Zhou, Y., Vallarino, J., Chu, M. T., Michanowicz, D. R., Levy, J. I., & Adamkiewicz, G. (2019). Development of an in-home, real-time air pollutant sensor platform and implications for community use. *Environmental Pollution*, *244*, 440–450. URL https://www.sciencedirect.com/science/article/pii/S0269749118319584 doi: 10.1016/j.envpol.2018.10.064

Gray, S. C., Edwards, S. E., & Miranda, M. L. (2013). Race, socioeconomic status, and air pollution exposure in North Carolina. *Environmental Research*, *126*, 152–158. URL https://www.sciencedirect.com/science/article/pii/S0013935113001138 doi: 10.1016/j.envres.2013.06.005

Green, R. S., Smorodinsky, S., Kim, J. J., McLaughlin, R., & Ostro, B. (2004). Proximity of California public schools to busy roads. *Environmental Health Perspectives*, *112*(1), 61–66. URL https://ehp.niehs.nih.gov/doi/10.1289/ehp.6566 doi: 10.1289/ehp.6566

Grineski, S. E., & Collins, T. W. (2018). Geographic and social disparities in exposure to air neurotoxicants at U.S. public schools. *Environmental Research*, *161*, 580–587. URL https://www.sciencedirect.com/science/article/pii/S0013935117317188 doi: 10.1016/j.envres.2017.11.047

Grineski, S. E., Collins, T. W., & Adkins, D. E. (2020). Hazardous air pollutants are associated with worse performance in reading, math, and science among US primary schoolchildren. *Environmental Research*, *181*, 108925. URL https://linkinghub.elsevier.com/retrieve/pii/S0013935119307224 doi: 10.1016/j.envres.2019.108925

Grineski, S. E., Staniswalis, J. G., Peng, Y., & Atkinson-Palombo, C. (2010). Children's asthma hospitalizations and relative risk due to nitrogen dioxide (NO2): Effect modification by race, ethnicity and insurance status. *Environmental research*, *110*(2), 178. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2819647/ doi: 10.1016/j.envres.2009.10.012

Gu, P., Li, H. Z., Ye, Q., Robinson, E. S., Apte, J. S., Robinson, A. L., & Presto, A. A. (2018). Intracity Variability of Particulate Matter Exposure Is Driven by Carbonaceous Sources and Correlated with Land-Use Variables. *Environmental Science & Technology*, *52*(20), 11545–11554. URL https://doi.org/10.1021/acs.est.8b03833 doi: 10.1021/acs.est.8b03833

Gupta, P., Doraiswamy, P., Levy, R., Pikelnaya, O., Maibach, J., Feenstra, B., ... Mills, K. C. (2018). Impact of California Fires on Local and Regional Air Quality: The

Role of a Low-Cost Sensor Network and Satellite Observations. *GeoHealth*, *2*(6), 172–181. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GH000136 doi: https://doi.org/10.1029/2018GH000136

Habre, R., Coull, B., Moshier, E., Godbold, J., Grunin, A., Nath, A., ... Koutrakis, P. (2014). Sources of indoor air pollution in New York City residences of asthmatic children. *Journal of Exposure Science & Environmental Epidemiology*, *24*(3), 269–278. doi: 10.1038/jes.2013.74

Hajat, A., Hsia, C., & O'Neill, M. S. (2015). Socioeconomic Disparities and Air Pollution Exposure: a Global Review. *Current Environmental Health Reports*, *2*(4), 440–450. doi: 10.1007/s40572-015-0069-5

Hammer, M. S., van Donkelaar, A., Li, C., Lyapustin, A., Sayer, A. M., Hsu, N. C., ... Martin, R. V. (2020). Global Estimates and Long-Term Trends of Fine Particulate Matter Concentrations (1998–2018). *Environmental Science & Technology*, *54*(13), 7879–7890. URL https://doi.org/10.1021/acs.est.0c01764 doi: 10.1021/acs.est.0c01764

Han, K., Ran, Z., Wang, X., Wu, Q., Zhan, N., Yi, Z., & Jin, T. (2021). Traffic-related organic and inorganic air pollution and risk of development of childhood asthma: A meta-analysis. *Environmental Research*, *194*, 110493. URL https://www.sciencedirect.com/science/article/pii/S0013935120313906 doi: 10.1016/j.envres.2020.110493

Hauptman, M., Gaffin, J. M., Petty, C. R., Sheehan, W. J., Lai, P. S., Coull, B., ... Phipatanakul, W. (2020). Proximity to major roadways and asthma symptoms in the School Inner-City Asthma Study. *The Journal of Allergy and Clinical Immunology*, *145*(1), 119–126.e4. doi: 10.1016/j.jaci.2019.08.038

Hennig, F., Quass, U., Hellack, B., Küpper, M., Kuhlbusch, T. A. J., Stafoggia, M., & Hoffmann, B. (2018). Ultrafine and Fine Particle Number and Surface Area Concentrations and Daily Cause-Specific Mortality in the Ruhr Area, Germany, 2009–2014. *Environmental Health Perspectives*, *126*(2), 027008. URL https://ehp.niehs.nih.gov/doi/10.1289/EHP2054 doi:

10.1289/EHP2054

Heo, S., Fong, K. C., & Bell, M. L. (2019). Risk of particulate matter on birth outcomes in relation to maternal socio-economic factors: a systematic review. *Environmental Research Letters*, *14*(12), 123004. URL https://doi.org/10.1088/1748-9326/ab4cd0 doi: 10.1088/1748-9326/ab4cd0

Hoek, G., Krishnan, R. M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., & Kaufman, J. D. (2013). Long-term air pollution exposure and cardio- respiratory mortality: a review. *Environmental Health*, *12*(1), 43. URL https://doi.org/10.1186/1476-069X-12-43 doi: 10.1186/1476-069X-12-43

Holstius, D. M., Reid, C. E., Jesdale, B. M., & Morello-Frosch, R. (2012). Birth Weight following Pregnancy during the 2003 Southern California Wildfires. *Environmental Health Perspectives*, *120*(9), 1340–1345. URL https://ehp.niehs.nih.gov/doi/10.1289/ehp.1104515 doi: 10.1289/ehp.1104515

Hoogh, K. d., Gulliver, J., Donkelaar, A. v., Martin, R. V., Marshall, J. D., Bechle, M. J., … Hoek, G. (2016). Development of West-European PM2.5 and NO2 land use regression models incorporating satellite-derived and chemical transport modelling data. *Environmental Research*, *151*, 1–10. URL https://www.sciencedirect.com/science/article/pii/S0013935116302894 doi: 10.1016/j.envres.2016.07.005

Hu, X., Waller, L. A., Lyapustin, A., Wang, Y., & Liu, Y. (2014). 10-year spatial and temporal trends of $PM_{2.5}$ concentrations in the southeastern US estimated using high-resolution satellite data. *Atmospheric Chemistry and Physics*, *14*(12), 6301–6314. URL https://www.atmos-chem-phys.net/14/6301/2014/ doi: 10.5194/acp-14-6301-2014

Hughes. (1996). *Pan's travail : environmental problems of the ancient Greeks and Romans*. Baltimore: ohns Hopkins University Press.

Jayaratne, R., Liu, X., Thai, P., Dunbabin, M., & Morawska, L. (2018). The influence of humidity on the performance of a low-cost air particle mass sensor and the effect

of atmospheric fog. *Atmospheric Measurement Techniques*, *11*(8), 4883–4890. URL https://amt.copernicus.org/articles/11/4883/2018/ doi: 10.5194/amt-11-4883-2018

Jena, C., Ghude, S. D., Kumar, R., Debnath, S., Govardhan, G., Soni, V. K., ... Rajeevan, M. (2021). Performance of high resolution (400 m) PM 2.5 forecast over Delhi. *Scientific Reports*, *11*(1), 4104. URL https://www.nature.com/articles/s41598-021-83467-8 doi: 10.1038/s41598-021-83467-8

Jerrett, M., Turner, M. C., Beckerman, B. S., Pope, C. A., van Donkelaar, A., Martin, R. V., ... Burnett, R. T. (2017). Comparing the Health Effects of Ambient Particulate Matter Estimated Using Ground-Based versus Remote Sensing Exposure Estimates. *Environmental Health Perspectives*, *125*(4), 552–559. URL https://ehp.niehs.nih.gov/doi/10.1289/EHP575 doi: 10.1289/EHP575

Jin, X., Fiore, A. M., Curci, G., Lyapustin, A., Civerolo, K., Ku, M., ... Martin, R. V. (2019). Assessing uncertainties of a geophysical approach to estimate surface fine particulate matter distributions from satellite-observed aerosol optical depth. *Atmospheric Chemistry and Physics*, *19*(1), 295–313. URL https://www.atmos-chem-phys.net/19/295/2019/ doi: 10.5194/acp-19-295-2019

Just, A. C., Wright, R. O., Schwartz, J., Coull, B. A., Baccarelli, A. A., Tellez-Rojo, M. M., ... Kloog, I. (2015). Using High-Resolution Satellite Aerosol Optical Depth To Estimate Daily PM2.5 Geographical Distribution in Mexico City. *Environmental Science & Technology*, *49*(14), 8576–8584. doi: 10.1021/acs.est.5b00859

Kahn, R. A., Li, W.-H., Moroney, C., Diner, D. J., Martonchik, J. V., & Fishbein, E. (2007). Aerosol source plume physical characteristics from space-based multi-angle imaging. *Journal of Geophysical Research: Atmospheres*, *112*(D11). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006JD007647 doi: 10.1029/2006JD007647

Kirwa, K., Szpiro, A. A., Sheppard, L., Sampson, P. D., Wang, M., Keller, J. P., ... Kaufman, J. D. (2021). Fine-Scale Air Pollution Models for Epidemiologic Research: Insights From Ap-

proaches Developed in the Multi-ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Current Environmental Health Reports*. doi: 10.1007/s40572-021-00310-y

Koelemeijer, R. B. A., Homan, C. D., & Matthijsen, J. (2006). Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. *Atmospheric Environment*, *40*(27), 5304–5315. doi: 10.1016/j.atmosenv.2006.04.044

Koss, A. R., Sekimoto, K., Gilman, J. B., Selimovic, V., Coggon, M. M., Zarzana, K. J., ... de Gouw, J. (2018). Non-methane organic gas emissions from biomass burning: identification, quantification, and emission factors from PTR-ToF during the FIREX 2016 laboratory experiment. *Atmospheric Chemistry and Physics*, *18*(5), 3299–3319. URL https://www.atmos-chem-phys.net/18/3299/2018/ doi: 10.5194/acp-18-3299-2018

Krebs, B., Burney, J., Zivin, J. G., & Neidell, M. (2021). Using Crowd-Sourced Data to Assess the Temporal and Spatial Relationship between Indoor and Outdoor Particulate Matter. *Environmental Science & Technology*, *55*(9), 6107–6115. URL https://doi.org/10.1021/acs.est.0c08469 doi: 10.1021/acs.est.0c08469

Kulkarni, N., & Grigg, J. (2008). Effect of air pollution on children. *Paediatrics and Child Health*, *18*(5), 238–243. URL https://www.sciencedirect.com/science/article/pii/S1751722208000280 doi: 10.1016/j.paed.2008.02.007

Kweon, B.-S., Mohai, P., Lee, S., & Sametshaw, A. M. (2018). Proximity of public schools to major highways and industrial facilities, and students' school performance and health hazards. *Environment and Planning B: Urban Analytics and City Science*, *45*(2), 312–329. URL https://doi.org/10.1177/0265813516673060 doi: 10.1177/0265813516673060

Lassman, W., Ford, B., Gan, R. W., Pfister, G., Magzamen, S., Fischer, E. V., & Pierce, J. R. (2017). Spatial and temporal estimates of population exposure to wildfire smoke during the Washington state 2012 wildfire season using blended model, satellite, and in situ data. *GeoHealth*, *1*(3), 106–121. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GH000049 doi: 10.1002/

2017GH000049

Lelieveld, J., Klingmüller, K., Pozzer, A., Pöschl, U., Fnais, M., Daiber, A., & Münzel, T. (2019). Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions. *European Heart Journal*, *40*(20), 1590–1596. URL https://academic.oup.com/eurheartj/article/40/20/1590/5372326 doi: 10.1093/eurheartj/ehz135

Levy Zamora, M., Xiong, F., Gentner, D., Kerkez, B., Kohrman-Glaser, J., & Koehler, K. (2019). Field and Laboratory Evaluations of the Low-Cost Plantower Particulate Matter Sensor. *Environmental Science & Technology*, *53*(2), 838–849. URL https://doi.org/10.1021/acs.est.8b05174 doi: 10.1021/acs.est.8b05174

Lightstone, S. D., Moshary, F., & Gross, B. (2017). Comparing CMAQ Forecasts with a Neural Network Forecast Model for PM2.5 in New York. *Atmosphere*, *8*(9), 161. URL https://www.mdpi.com/2073-4433/8/9/161 doi: 10.3390/atmos8090161

Liu, J. C., Mickley, L. J., Sulprizio, M. P., Dominici, F., Yue, X., Ebisu, K., ... Bell, M. L. (2016). Particulate air pollution from wildfires in the Western US under climate change. *Climatic Change*, *138*(3-4), 655–666. URL http://link.springer.com/10.1007/s10584-016-1762-6 doi: 10.1007/s10584-016-1762-6

Liu, J. C., Pereira, G., Uhl, S. A., Bravo, M. A., & Bell, M. L. (2015). A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke. *Environmental Research*, *136*, 120–132. URL https://linkinghub.elsevier.com/retrieve/pii/S0013935114003788 doi: 10.1016/j.envres.2014.10.015

Liu, Y., Cao, G., Zhao, N., Mulligan, K., & Ye, X. (2018). Improve ground-level PM2.5 concentration mapping using a random forests-based geostatistical approach. *Environmental Pollution*, *235*, 272–282. URL https://www.sciencedirect.com/science/article/pii/S0269749117316469 doi: 10.1016/j.envpol.2017.12.070

Liu, Y., Franklin, M., Kahn, R., & Koutrakis, P. (2007). Using aerosol optical thickness

to predict ground-level PM2.5 concentrations in the St. Louis area: A comparison between MISR and MODIS. *Remote Sensing of Environment*, *107*(1-2), 33–44. URL https://linkinghub.elsevier.com/retrieve/pii/S0034425706004317 doi: 10.1016/j.rse.2006.05.022

Liu, Y., Park, R. J., Jacob, D. J., Li, Q., Kilaru, V., & Sarnat, J. A. (2004). Mapping annual mean ground-level PM2.5 concentrations using Multiangle Imaging Spectroradiometer aerosol optical thickness over the contiguous United States. *Journal of Geophysical Research: Atmospheres*, *109*(D22). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JD005025 doi: 10.1029/2004JD005025

Liu, Y., Sarnat, J. A., Kilaru, V., Jacob, D. J., & Koutrakis, P. (2005). Estimating Ground-Level PM2.5 in the Eastern United States Using Satellite Remote Sensing. *Environmental Science & Technology*, *39*(9), 3269–3278. URL https://doi.org/10.1021/es049352m doi: 10.1021/es049352m

Lu, Y., Giuliano, G., & Habre, R. (2021). Estimating hourly PM2.5 concentrations at the neighborhood scale using a low-cost air sensor network: A Los Angeles Case Study. *Environmental Research*, 110653. URL https://linkinghub.elsevier.com/retrieve/pii/S0013935120315504 doi: 10.1016/j.envres.2020.110653

Lyapustin, A., Korkin, S., Wang, Y., Quayle, B., & Laszlo, I. (2012a). Corrigendum to "Discrimination of biomass burning smoke and clouds in MAIAC algorithm" published in Atmos. Chem. Phys., 12, 9679–9686, 2012. *Atmospheric Chemistry and Physics*, *12*(21), 10631–10631. URL http://www.atmos-chem-phys.net/12/10631/2012/ doi: 10.5194/acp-12-10631-2012

Lyapustin, A., Korkin, S., Wang, Y., Quayle, B., & Laszlo, I. (2012b). Discrimination of biomass burning smoke and clouds in MAIAC algorithm. *Atmospheric Chemistry and Physics*, *12*(20), 9679–9686. URL https://www.atmos-chem-phys.net/12/9679/2012/ doi: 10.5194/acp-12-9679-2012

Lyapustin, A., Wang, Y., Korkin, S., & Huang, D. (2018). MODIS Collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques*, *11*(10), 5741–5765. URL https://www.atmos-meas-tech.net/11/5741/2018/ doi: 10.5194/amt-11-5741-2018

Lyapustin, A., Wang, Y., Korkin, S., Kahn, R., & Winker, D. (2019). MAIAC Thermal Technique for Smoke Injection Height From MODIS. *IEEE Geoscience and Remote Sensing Letters*, 1–5;. URL https://ieeexplore.ieee.org/document/8834856/ doi: 10.1109/LGRS.2019.2936332

Maantay, J. (2002). Mapping environmental injustices: pitfalls and potential of geographic information systems in assessing environmental health and equity. *Environmental Health Perspectives*, *110*(Suppl 2), 161–171. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1241160/

Maantay, J. (2007). Asthma and air pollution in the Bronx: Methodological and data considerations in using GIS for environmental justice and health research. *Health & Place*, *13*(1), 32–56. URL https://linkinghub.elsevier.com/retrieve/pii/S1353829205000675 doi: 10.1016/j.healthplace.2005.09.009

Maantay, J., Tu, J., & Maroko, A. R. (2009). Loose-coupling an air dispersion model and a geographic information system (GIS) for studying air pollution and asthma in the Bronx, New York City. *International Journal of Environmental Health Research*, *19*(1), 59–79. URL https://doi.org/10.1080/09603120802392868 doi: 10.1080/09603120802392868

Madureira, J., Paciência, I., Rufo, J., Ramos, E., Barros, H., Teixeira, J. P., & de Oliveira Fernandes, E. (2015). Indoor air quality in schools and its relationship with children's respiratory symptoms. *Atmospheric Environment*, *118*, 145–156. URL https://www.sciencedirect.com/science/article/pii/S1352231015302272 doi: 10.1016/j.atmosenv.2015.07.028

Magi, B. I., Cupini, C., Francis, J., Green, M., & Hauser, C. (2020). Evaluation of PM2.5 measured in an urban setting using a low-cost optical particle counter and a Federal Equivalent Method Beta Attenuation Monitor. *Aerosol Science and Technology*, *54*(2), 147–159. URL

https://doi.org/10.1080/02786826.2019.1619915 doi: 10.1080/02786826.2019.1619915

Malings, C., Tanzer, R., Hauryliuk, A., Saha, P. K., Robinson, A. L., Presto, A. A., & Subramanian, R. (2020). Fine particle mass monitoring with low-cost sensors: Corrections and long-term performance evaluation. *Aerosol Science and Technology*, *54*(2), 160–174. URL https://doi.org/10.1080/02786826.2019.1623863 doi: 10.1080/02786826.2019.1623863

Marlier, M. E., Jina, A. S., Kinney, P. L., & DeFries, R. S. (2016). Extreme Air Pollution in Global Megacities. *Current Climate Change Reports*, *2*(1), 15–27. URL https://doi.org/10.1007/s40641-016-0032-z doi: 10.1007/s40641-016-0032-z

McClure, C. D., & Jaffe, D. A. (2018). US particulate matter air quality improves except in wildfire-prone areas. *Proceedings of the National Academy of Sciences*, *115*(31), 7901–7906. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1804353115 doi: 10.1073/pnas.1804353115

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., ... Shi, W. (2006). North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, *87*(3), 343–360. URL http://journals.ametsoc.org/doi/10.1175/BAMS-87-3-343 doi: 10.1175/BAMS-87-3-343

Mohai, P., Kweon, B.-S., Lee, S., & Ard, K. (2011). Air Pollution Around Schools Is Linked To Poorer Student Health And Academic Performance. *Health Affairs*, *30*(5), 852–62. URL http://www.proquest.com/docview/868915495/abstract/7DE3C17CC1554099PQ/1

Mohai, P., & Saha, R. (2007). Racial Inequality in the Distribution of Hazardous Waste: A National-Level Reassessment. *Social Problems*, *54*(3), 343–370. URL https://www.jstor.org/stable/10.1525/sp.2007.54.3.343 doi: 10.1525/sp.2007.54.3.343

Morello-Frosch, R., Pastor, M., & Sadd, J. (2002). Integrating Environmental Justice and the Precautionary Principle in Research and Policy Making: The Case of Ambient Air Toxics Exposures and Health Risks among Schoolchildren in Los Angeles. *The ANNALS of the American Academy of Political and Social Science*, *584*(1), 47–68. URL https://doi.org/10.1177/000271620258400104 doi: 10.1177/000271620258400104

Mosley, S. (2014). Environmental History of Air Pollution and Protection. In M. Agnoletti & S. Neri Serneri (Eds.), *The Basic Environmental History* (pp. 143–169). Cham: Springer International Publishing. URL https://doi.org/10.1007/978-3-319-09180-8_5 doi: 10.1007/978-3-319-09180-8_5

Mullen, C., Grineski, S., Collins, T., Xing, W., Whitaker, R., Sayahi, T., ... Kelly, K. (2020). Patterns of distributive environmental inequity under different PM2.5 air pollution scenarios for Salt Lake County public schools. *Environmental Research*, *186*, 109543. URL https://www.sciencedirect.com/science/article/pii/S0013935120304369 doi: 10.1016/j.envres.2020.109543

Murray, C. J. L., Aravkin, A. Y., Zheng, P., Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., ... Lim, S. S. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, *396*(10258), 1223–1249. URL https://www.sciencedirect.com/science/article/pii/S0140673620307522 doi: 10.1016/S0140-6736(20)30752-2

Myhre, O., Låg, M., Villanger, G. D., Oftedal, B., Øvrevik, J., Holme, J. A., ... Dirven, H. (2018). Early life exposure to air pollution particulate matter (PM) as risk factor for attention deficit/hyperactivity disorder (ADHD): Need for novel strategies for mechanisms and causalities. *Toxicology and Applied Pharmacology*, *354*, 196–214. URL https://www.sciencedirect.com/science/article/pii/S0041008X18300991 doi: 10.1016/j.taap.2018.03.015

National Center for Education Statistics. (2008). *National Center For Education Statistics (NCES). Schools and Staffing Survey (SASS).* URL https://nces.ed.gov/surveys/sass/tables/sass0708_035_s1s.asp

National Center for Education Statistics. (2015). *NCES Blog | Free or reduced price lunch: A proxy for poverty?* URL https://nces.ed.gov/blogs/nces/post/free-or-reduced-price-lunch-a-proxy-for-poverty

National Center for Education Statistics. (2021). *The NCES Fast Facts Tool provides quick answers to many education questions (National Center for Education Statistics).* URL https://nces.ed.gov/fastfacts/display.asp?id=372

Nelson, D., Garay, M., Kahn, R., & Dunst, B. (2013). Stereoscopic Height and Wind Retrievals for Aerosol Plumes with the MISR INteractive eXplorer (MINX). *Remote Sensing*, *5*(9), 4593–4628. URL http://www.mdpi.com/2072-4292/5/9/4593 doi: 10.3390/rs5094593

O'Neill, M. S., Jerrett, M., Kawachi, I., Levy, J. I., Cohen, A. J., Gouveia, N., ... null, n. (2003). Health, wealth, and air pollution: advancing theory and methods. *Environmental Health Perspectives*, *111*(16), 1861–1870. URL https://ehp.niehs.nih.gov/doi/10.1289/ehp.6334 doi: 10.1289/ehp.6334

O'Dell, K., Bilsback, K., Ford, B., Martenies, S. E., Magzamen, S., Fischer, E. V., & Pierce, J. R. (2021). Estimated Mortality and Morbidity Attributable to Smoke Plumes in the United States: Not Just a Western US Problem. *GeoHealth*, *5*(9), e2021GH000457. URL http://onlinelibrary.wiley.com/doi/abs/10.1029/2021GH000457 doi: 10.1029/2021GH000457

O'Dell, K., Ford, B., Fischer, E. V., & Pierce, J. R. (2019). Contribution of Wildland-Fire Smoke to US PM $_{2.5}$ and Its Influence on Recent Trends. *Environmental Science & Technology*, *53*(4), 1797–1804. URL https://pubs.acs.org/doi/10.1021/acs.est.8b05430 doi: 10.1021/acs.est.8b05430

Pastor, M., Morello-Frosch, R., & Sadd, J. L. (2006). Breathless: Schools, Air Toxics, and Environmental Justice in California. *Policy Studies Journal*, *34*(3), 337–362. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0072.2006.00176.x doi: 10.1111/j.1541-0072.2006.00176.x

Pastor, M., Sadd, J. L., & Morello-Frosch, R. (2004). Waiting to Inhale: The Demographics of Toxic Air Release Facilities in 21st-Century California*. *Social Science Quarterly*, *85*(2), 420–440. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/j.0038-4941.2004.08502010.x doi: 10.1111/j.0038-4941.2004.08502010.x

Pastor, M., Jr., Sadd, J. L., & Morello-Frosch, R. (2002). Who's Minding the Kids? Pollucion, Public Schools, and Environmental Justice in Los Angeles. *Social Science Quarterly*, *83*(1), 263–280. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1540-6237.00082 doi: 10.1111/1540-6237.00082

Paugam, R., Wooster, M., Freitas, S., & Val Martin, M. (2016). A review of approaches to estimate wildfire plume injection height within large-scale atmospheric chemical transport models. *Atmospheric Chemistry and Physics*, *16*(2), 907–925. URL https://www.atmos-chem-phys.net/16/907/2016/ doi: 10.5194/acp-16-907-2016

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 6.

Pinault, L., Crouse, D., Jerrett, M., Brauer, M., & Tjepkema, M. (2016). Spatial associations between socioeconomic groups and NO2 air pollution exposure within three large Canadian cities. *Environmental Research*, *147*, 373–382. URL https://www.sciencedirect.com/science/article/pii/S0013935116300755 doi: 10.1016/j.envres.2016.02.033

Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., & Thurston, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA*, *287*(9), 1132–1141. doi: 10.1001/jama.287.9.1132

Quinn, C., Ford, B., & Volckens, J. (2019). Mobilizing the Low-Cost Sensor Revolution with Smartphones and Citizen Science. *The Magazine for Environmental Managers*, 5.

Ready, D. D. (2010). Socioeconomic Disadvantage, School Attendance, and Early Cognitive Development: The Differential Effects of School Exposure. *Sociology of Education*, *83*(4), 271–286. URL https://doi.org/10.1177/0038040710383520 doi: 10.1177/0038040710383520

Reche, C., Rivas, I., Pandolfi, M., Viana, M., Bouso, L., Àlvarez Pedrerol, M., ... Querol, X. (2015). Real-time indoor and outdoor measurements of black

carbon at primary schools. *Atmospheric Environment*, *120*, 417–426. URL https://www.sciencedirect.com/science/article/pii/S1352231015302892 doi: 10.1016/j.atmosenv.2015.08.044

Reid, C. E., Brauer, M., Johnston, F. H., Jerrett, M., Balmes, J. R., & Elliott, C. T. (2016). Critical Review of Health Impacts of Wildfire Smoke Exposure. *Environmental Health Perspectives*, *124*(9), 1334–1343. URL https://ehp.niehs.nih.gov/doi/10.1289/ehp.1409277 doi: 10.1289/ehp.1409277

Reid, C. E., Jerrett, M., Petersen, M. L., Pfister, G. G., Morefield, P. E., Tager, I. B., ... Balmes, J. R. (2015). Spatiotemporal Prediction of Fine Particulate Matter During the 2008 Northern California Wildfires Using Machine Learning. *Environmental Science & Technology*, *49*(6), 3887–3896. URL https://pubs.acs.org/doi/10.1021/es505846r doi: 10.1021/es505846r

Requia, W. J., Roig, H. L., & Schwartz, J. D. (2021). Schools exposure to air pollution sources in Brazil: A nationwide assessment of more than 180 thousand schools. *Science of The Total Environment*, *763*, 143027. URL https://www.sciencedirect.com/science/article/pii/S0048969720365578 doi: 10.1016/j.scitotenv.2020.143027

Reuters. (2020). *Special Report: U.S. air monitors routinely miss pollution - even refinery explosions.* URL https://www.reuters.com/article/usa-pollution-airmonitors-specialreport-idUSKBN28B4RT

Rickenbacker, H., Brown, F., & Bilec, M. (2019). Creating environmental consciousness in underserved communities: Implementation and outcomes of community-based environmental justice and air pollution research. *Sustainable Cities and Society*, *47*, 101473. URL https://linkinghub.elsevier.com/retrieve/pii/S2210670718321449 doi: 10.1016/j.scs.2019.101473

Roberts, S., Arseneault, L., Barratt, B., Beevers, S., Danese, A., Odgers, C. L., ... Fisher, H. L. (2019). Exploration of NO2 and PM2.5 air pollution and mental health problems using high-resolution data in London-based children

from a UK longitudinal cohort study. *Psychiatry Research*, *272*, 8–17. URL https://www.sciencedirect.com/science/article/pii/S016517811830800X doi: 10.1016/j.psychres.2018.12.050

Saha, P. K., Hankey, S., Marshall, J. D., Robinson, A. L., & Presto, A. A. (2021). High-Spatial-Resolution Estimates of Ultrafine Particle Concentrations across the Continental United States. *Environmental Science & Technology*, *55*(15), 10320–10331. URL https://doi.org/10.1021/acs.est.1c03237 doi: 10.1021/acs.est.1c03237

Sampson, N. (2012). Environmental Justice at School: Understanding Research, Policy, and Practice to Improve Our Children's Health. *Journal of School Health*, *82*(5), 246–252. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1746-1561.2012.00694.x doi: 10.1111/j.1746-1561.2012.00694.x

Sayahi, T., Butterfield, A., & Kelly, K. E. (2019). Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors. *Environmental Pollution*, *245*, 932–940. URL https://www.sciencedirect.com/science/article/pii/S0269749118316129 doi: 10.1016/j.envpol.2018.11.065

Schwartz, J., Dockery, D. W., & Neas, L. M. (1996). Is Daily Mortality Associated Specifically with Fine Particles? *Journal of the Air & Waste Management Association*, *46*(10), 927–939. URL https://doi.org/10.1080/10473289.1996.10467528 doi: 10.1080/10473289.1996.10467528

Shah, R. U., Robinson, E. S., Gu, P., Robinson, A. L., Apte, J. S., & Presto, A. A. (2018). High-spatial-resolution mapping and source apportionment of aerosol composition in Oakland, California, using mobile aerosol mass spectrometry. *Atmospheric Chemistry and Physics*, *18*(22), 16325–16344. URL https://www.atmos-chem-phys.net/18/16325/2018/ doi: 10.5194/acp-18-16325-2018

Shoari, N., Heydari, S., & Blangiardo, M. (2022). School neighbourhood and compliance with WHO-recommended annual NO2 guideline: A case study of Greater London. *Science of The Total Environment*, *803*, 150038. URL

https://www.sciencedirect.com/science/article/pii/S0048969721051135 doi: 10.1016/j.scitotenv.2021.150038

Singer, B. C., & Delp, W. W. (2018). Response of consumer and research grade indoor air quality monitors to residential sources of fine particles. *Indoor Air*, *28*(4), 624–639. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/ina.12463 doi: 10.1111/ina.12463

Snider, G., Weagle, C. L., Martin, R. V., van Donkelaar, A., Conrad, K., Cunningham, D., ... Liu, Y. (2015). SPARTAN: a global network to evaluate and enhance satellite-based estimates of ground-level particulate matter for global health applications. *Atmospheric Measurement Techniques*, *8*(1), 505–521. URL https://www.atmos-meas-tech.net/8/505/2015/ doi: 10.5194/amt-8-505-2015

Snider, G., Weagle, C. L., Murdymootoo, K. K., Ring, A., Ritchie, Y., Stone, E., ... Martin, R. V. (2016). Variation in global chemical composition of PM $_{2.5}$:emerging results from SPARTAN. *Atmospheric Chemistry and Physics*, *16*(15), 9629–9653. URL https://www.atmos-chem-phys.net/16/9629/2016/ doi: 10.5194/acp-16-9629-2016

Snyder, E. G., Watkins, T. H., Solomon, P. A., Thoma, E. D., Williams, R. W., Hagler, G. S. W., ... Preuss, P. W. (2013). The Changing Paradigm of Air Pollution Monitoring. *Environmental Science & Technology*, *47*(20), 11369–11377. URL https://pubs.acs.org/doi/10.1021/es4022602 doi: 10.1021/es4022602

Song, W., Jia, H., Huang, J., & Zhang, Y. (2014). A satellite-based geographically weighted regression model for regional PM2.5 estimation over the Pearl River Delta region in China. *Remote Sensing of Environment*, *154*, 1–7. URL https://www.sciencedirect.com/science/article/pii/S0034425714003058 doi: 10.1016/j.rse.2014.08.008

Sousan, S., Koehler, K., Hallett, L., & Peters, T. M. (2017). Evaluation of consumer monitors to measure particulate matter. *Journal of Aerosol Science*, *107*, 123–133. URL https://www.sciencedirect.com/science/article/pii/S002185021630338X doi: 10.1016/j.jaerosci.2017.02.013

Southerland, V. A., C., A. S., Harris Maria, Apte Joshua, Hystad Perry, van Donkelaar Aaron, ... Roy Ananya (2021). Assessing the Distribution of Air Pollution Health Risks within Cities: A Neighborhood-Scale Analysis Leveraging High-Resolution Data Sets in the Bay Area, California. *Environmental Health Perspectives*, *129*(3), 037006. URL https://ehp.niehs.nih.gov/doi/10.1289/EHP7679 doi: 10.1289/EHP7679

Spracklen, D. V., Mickley, L. J., Logan, J. A., Hudman, R. C., Yevich, R., Flannigan, M. D., & Westerling, A. L. (2009). Impacts of climate change from 2000 to 2050 on wildfire activity and carbonaceous aerosol concentrations in the western United States. *Journal of Geophysical Research*, *114*(D20), D20301. URL http://doi.wiley.com/10.1029/2008JD010966 doi: 10.1029/2008JD010966

Suleiman, A., Tight, M. R., & Quinn, A. D. (2019). Applying machine learning methods in managing urban concentrations of traffic-related particulate matter (PM10 and PM2.5). *Atmospheric Pollution Research*, *10*(1), 134–144. URL https://www.sciencedirect.com/science/article/pii/S1309104218301272 doi: 10.1016/j.apr.2018.07.001

Sunyer, J. (2008). The neurological effects of air pollution in children. *European Respiratory Journal*, *32*(3), 535–537. URL https://erj.ersjournals.com/content/32/3/535 doi: 10.1183/09031936.00073708

Tessum, C. W., Apte, J. S., Goodkind, A. L., Muller, N. Z., Mullins, K. A., Paolella, D. A., ... Hill, J. D. (2019). Inequity in consumption of goods and services adds to racial–ethnic disparities in air pollution exposure. *Proceedings of the National Academy of Sciences*, *116*(13), 6001–6006. URL https://www.pnas.org/content/116/13/6001 doi: 10.1073/pnas.1818859116

Tryner, J., L'Orange, C., Mehaffy, J., Miller-Lionberg, D., Hofstetter, J. C., Wilson, A., & Volckens, J. (2020). Laboratory evaluation of low-cost PurpleAir PM monitors and in-field correction using co-located portable filter samplers. *Atmospheric Environment*, *220*, 117067. URL https://linkinghub.elsevier.com/retrieve/pii/S135223101930706X doi: 10.1016/j.atmosenv.2019.117067

Val Martin, M., Kahn, R. A., Logan, J. A., Paugam, R., Wooster, M., & Ichoku, C. (2012). Space-based observational constraints for 1-D fire smoke plume-rise models. *Journal of Geophysical Research: Atmospheres*, *117*(D22). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012JD018370 doi: 10.1029/2012JD018370

van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010). Global Estimates of Ambient Fine Particulate Matter Concentrations from Satellite-Based Aerosol Optical Depth: Development and Application. *Environmental Health Perspectives*, *118*(6), 847–855. URL https://ehp.niehs.nih.gov/doi/10.1289/ehp.0901623 doi: 10.1289/ehp.0901623

van Donkelaar, A., Martin, R. V., Levy, R. C., da Silva, A. M., Krzyzanowski, M., Chubarova, N. E., ... Cohen, A. J. (2011). Satellite-based estimates of ground-level fine particulate matter during extreme events: A case study of the Moscow fires in 2010. *Atmospheric Environment*, *45*(34), 6225–6232. URL https://linkinghub.elsevier.com/retrieve/pii/S135223101100851X doi: 10.1016/j.atmosenv.2011.07.068

van Donkelaar, A., Martin, R. V., & Park, R. J. (2006). Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing. *Journal of Geophysical Research: Atmospheres*, *111*(D21). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006996 doi: 10.1029/2005JD006996

van Donkelaar, A., Martin, R. V., Pasch, A. N., Szykman, J. J., Zhang, L., Wang, Y. X., & Chen, D. (2012). Improving the Accuracy of Daily Satellite-Derived Ground-Level Fine Aerosol Concentration Estimates for North America. *Environmental Science & Technology*, *46*(21), 11971–11978. URL https://pubs.acs.org/doi/10.1021/es3025319 doi: 10.1021/es3025319

van Donkelaar, A., Martin, R. V., Spurr, R. J. D., & Burnett, R. T. (2015). High-Resolution Satellite-Derived PM $_{2.5}$ from Optimal Estimation and Geographically Weighted Regression

over North America. *Environmental Science & Technology*, *49*(17), 10482–10491. URL https://pubs.acs.org/doi/10.1021/acs.est.5b02076 doi: 10.1021/acs.est.5b02076

van Donkelaar, A., Martin, R. V., Spurr, R. J. D., Drury, E., Remer, L. A., Levy, R. C., & Wang, J. (2013). Optimal estimation for global ground-level fine particulate matter concentrations. *Journal of Geophysical Research: Atmospheres*, *118*(11), 5621–5636. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50479 doi: 10.1002/jgrd.50479

van Zoest, V., Hoek, G., Osei, F., & Stein, A. (2020). Bayesian analysis of the short-term association of NO2 exposure with local burden of asthmatic symptoms in children. *Science of The Total Environment*, *720*, 137544. URL https://www.sciencedirect.com/science/article/pii/S004896972031055X doi: 10.1016/j.scitotenv.2020.137544

Varaden, D., McKevitt, C., & Barratt, B. (2018). Making the invisible visible: Engaging school children in monitoring air pollution in London. *Research for All*, *2*(2), 267–288. URL https://scienceopen.com/document?vid=7c2fc37f-5854-4e77-8a9f-af29ec030f07 doi: 10.18546/RFA.02.2.06

Vernon, C. J., Bolt, R., Canty, T., & Kahn, R. A. (2018). The impact of MISR-derived injection height initialization on wildfire and volcanic plume dispersion in the HYSPLIT model. *Atmospheric Measurement Techniques*, *11*(11), 6289–6307. URL https://www.atmos-meas-tech.net/11/6289/2018/ doi: 10.5194/amt-11-6289-2018

Volckens, J., Quinn, C., Leith, D., Mehaffy, J., Henry, C. S., & Miller-Lionberg, D. (2017). Development and evaluation of an ultrasonic personal aerosol sampler. *Indoor Air*, *27*(2), 409–416. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ina.12318 doi: 10.1111/ina.12318

Walker, G., Fairburn, John, Smith, Graham, & Mitchell, Gordan. (2003). *Environmental quality and social deprivation: phase II : national analysis of flood hazard, IPC Industries and air quality*.

Wang, J. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM $_{2.5}$ mass: Implications for air quality studies. *Geophysical Research Letters*, *30*(21), 2095. URL http://doi.wiley.com/10.1029/2003GL018174 doi: 10.1029/2003GL018174

Wang, X., & Oliver Gao, H. (2011). Exposure to fine particle mass and number concentrations in urban transportation environments of New York City. *Transportation Research Part D: Transport and Environment*, *16*(5), 384–391. URL https://linkinghub.elsevier.com/retrieve/pii/S136192091100040X doi: 10.1016/j.trd.2011.03.001

Wang, Y., Li, J., Jing, H., Zhang, Q., Jiang, J., & Biswas, P. (2015). Laboratory Evaluation and Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement. *Aerosol Science and Technology*, *49*(11), 1063–1077. URL https://doi.org/10.1080/02786826.2015.1100710 doi: 10.1080/02786826.2015.1100710

Wendt, E. A., Quinn, C., L'Orange, C., Miller-Lionberg, D. D., Ford, B., Pierce, J. R., ... Volckens, J. (2021). A low-cost monitor for simultaneous measurement of fine particulate matter and aerosol optical depth - Part 3: Automation and design improvements. *Atmospheric Measurement Techniques Discussions*, 1–25. URL https://amt.copernicus.org/preprints/amt-2021-73/ doi: 10.5194/amt-2021-73

Wendt, E. A., Quinn, C. W., Miller-Lionberg, D. D., Tryner, J., L'Orange, C., Ford, B., ... Volckens, J. (2019). A low-cost monitor for simultaneous measurement of fine particulate matter and aerosol optical depth - Part 1: Specifications and testing. *Atmospheric Measurement Techniques*, *12*(10), 5431–5441. URL https://www.atmos-meas-tech.net/12/5431/2019/ doi: 10.5194/amt-12-5431-2019

Westerling, A. L. (2006). Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity. *Science*, *313*(5789), 940–943. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1128834 doi: 10.1126/science.1128834

Westerling, A. L., Gershunov, A., Brown, T. J., Cayan, D. R., & Dettinger, M. D. (2003). Climate

and Wildfire in the Western United States. *Bulletin of the American Meteorological Society*, *84*(5), 595–604. URL http://journals.ametsoc.org/doi/10.1175/BAMS-84-5-595 doi: 10.1175/BAMS-84-5-595

Wettstein, Z. S., Hoshiko, S., Fahimi, J., Harrison, R. J., Cascio, W. E., & Rappold, A. G. (2018). Cardiovascular and Cerebrovascular Emergency Department Visits Associated With Wildfire Smoke Exposure in California in 2015. *Journal of the American Heart Association*, *7*(8). URL https://www.ahajournals.org/doi/10.1161/JAHA.117.007492 doi: 10.1161/JAHA.117.007492

Winker, D. M., Vaughan, M. A., Omar, A., Hu, Y., Powell, K. A., Liu, Z., ... Young, S. A. (2009). Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms. *Journal of Atmospheric and Oceanic Technology*, *26*(11), 2310–2323. URL http://journals.ametsoc.org/doi/10.1175/2009JTECHA1281.1 doi: 10.1175/2009JTECHA1281.1

Wolfe, M. K., McDonald, N. C., Arunachalam, S., Baldauf, R., & Valencia, A. (2020). Impact of school location on children's air pollution exposure. *Journal of Urban Affairs*, *0*(0), 1–17. URL https://doi.org/10.1080/07352166.2020.1734013 doi: 10.1080/07352166.2020.1734013

Woo, B., Kravitz-Wirtz, N., Sass, V., Crowder, K., Teixeira, S., & Takeuchi, D. T. (2019). Residential Segregation and Racial/Ethnic Disparities in Ambient Air Pollution. *Race and social problems*, *11*(1), 60–67. URL http://www.proquest.com/docview/2280546241?pq-origsite=primo doi: http://dx.doi.org.ezproxy2.library.colostate.edu/10.1007/s12552-018-9254-0

World Health Organization. (2013). *Review of evidence on health aspects of air pollution – REVIHAAP project: final technical report.* URL https://www.euro.who.int/en/health-topics/environment-and-health/air-quality/publications/2013/review-of-evidence-on-health-aspects-of-air-pollution-revihaap-project-final-technical-report

Xi, X., Wei, Z., Xiaoguang, R., Yijie, W., Xinxin, B., Wenjun, Y., & Jin, D. (2015). A comprehen-

sive evaluation of air pollution prediction improvement by a machine learning method. In *2015 IEEE International Conference on Service Operations And Logistics, And Informatics (SOLI)* (pp. 176–181). doi: 10.1109/SOLI.2015.7367615

Yue, X., Mickley, L. J., Logan, J. A., & Kaplan, J. O. (2013). Ensemble projections of wildfire activity and carbonaceous aerosol concentrations over the western United States in the mid-21st century. *Atmospheric Environment*, *77*, 767–780. URL https://linkinghub.elsevier.com/retrieve/pii/S1352231013004573 doi: 10.1016/j.atmosenv.2013.06.003

Zanobetti, A., & Schwartz, J. (2000). Race, Gender, and Social Status as Modifiers of the Effects of PM10 on Mortality. *Journal of Occupational and Environmental Medicine*, *42*(5), 469–474.

Zeng, X.-W., Lowe, A. J., Lodge, C. J., Heinrich, J., Roponen, M., Jalava, P., ... Dong, G.-H. (2020). Greenness surrounding schools is associated with lower risk of asthma in schoolchildren. *Environment International*, *143*, 105967. URL https://www.sciencedirect.com/science/article/pii/S016041202031922X doi: 10.1016/j.envint.2020.105967

Zhu, L., Val Martin, M., Gatti, L. V., Kahn, R., Hecobian, A., & Fischer, E. V. (2018). Development and implementation of a new biomass burning emissions injection height scheme (BBEIH v1.0) for the GEOS-Chem model (v9-01-01). *Geoscientific Model Development*, *11*(10), 4103–4116. URL https://www.geosci-model-dev.net/11/4103/2018/ doi: 10.5194/gmd-11-4103-2018

Zikova, N., Masiol, M., Chalupa, D. C., Rich, D. Q., Ferro, A. R., & Hopke, P. K. (2017). Estimating Hourly Concentrations of PM2.5 across a Metropolitan Area Using Low-Cost Particle Monitors. *Sensors*, *17*(8), 1922. URL https://www.mdpi.com/1424-8220/17/8/1922 doi: 10.3390/s17081922

Appendix A

Supplementary information for Chapter 2

A.1 Supplementary Figures

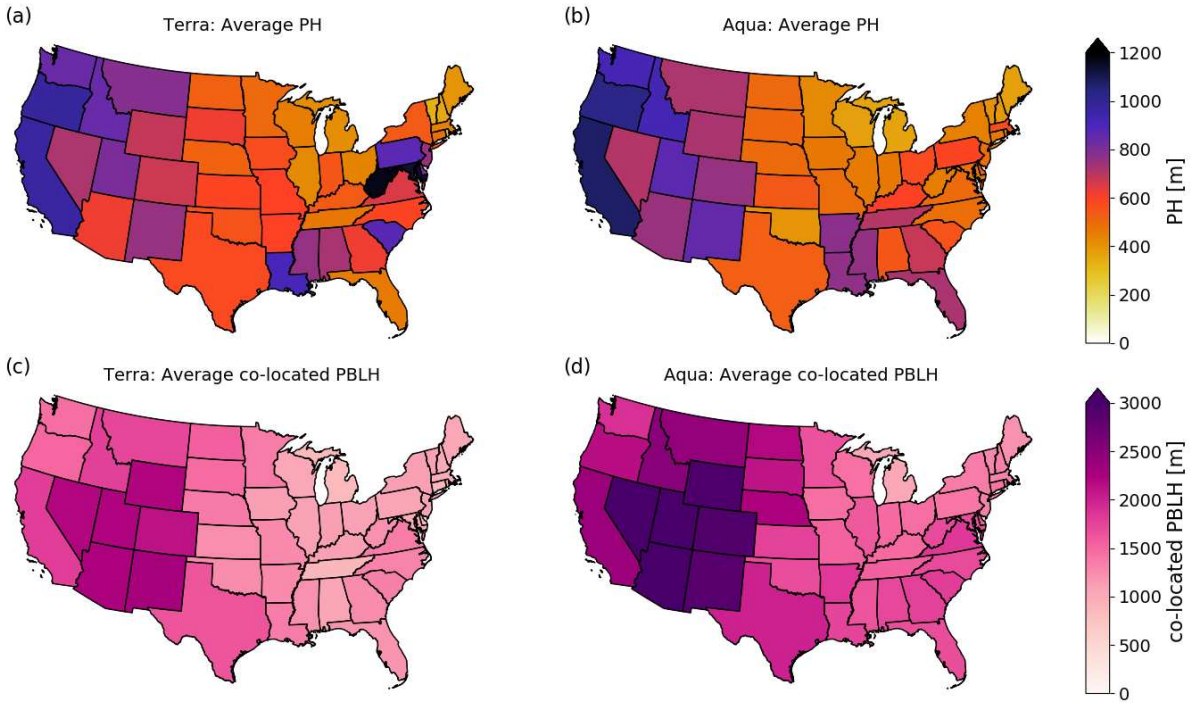

Figure A.1: Average smoke PH values from MAIAC Terra (a) and Aqua (b). Average reanalysis PBLH values that were co-located to PH retrievals for Terra (c) and Aqua (d).

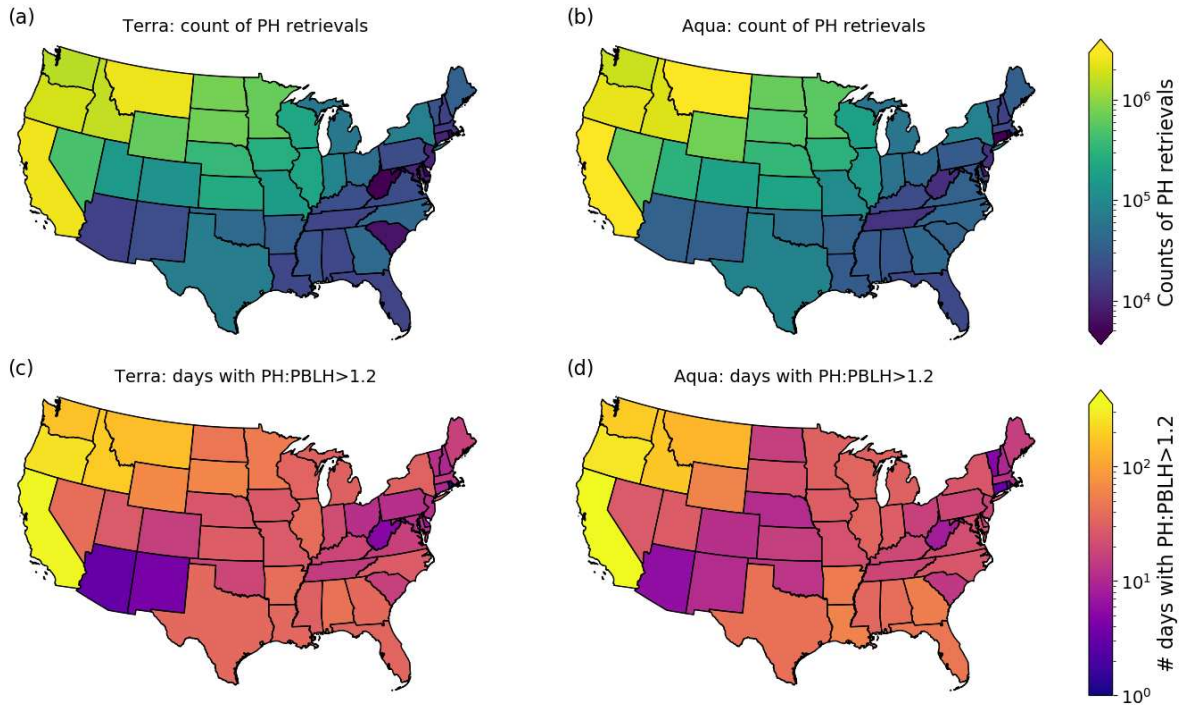

Figure A.2: Count of MAIAC PH retrievals for Terra (a) and Aqua (b) in each state, as well as the number of days that had a PH:PBLH values greater than 1.2 for Terra (c) and Aqua (d).

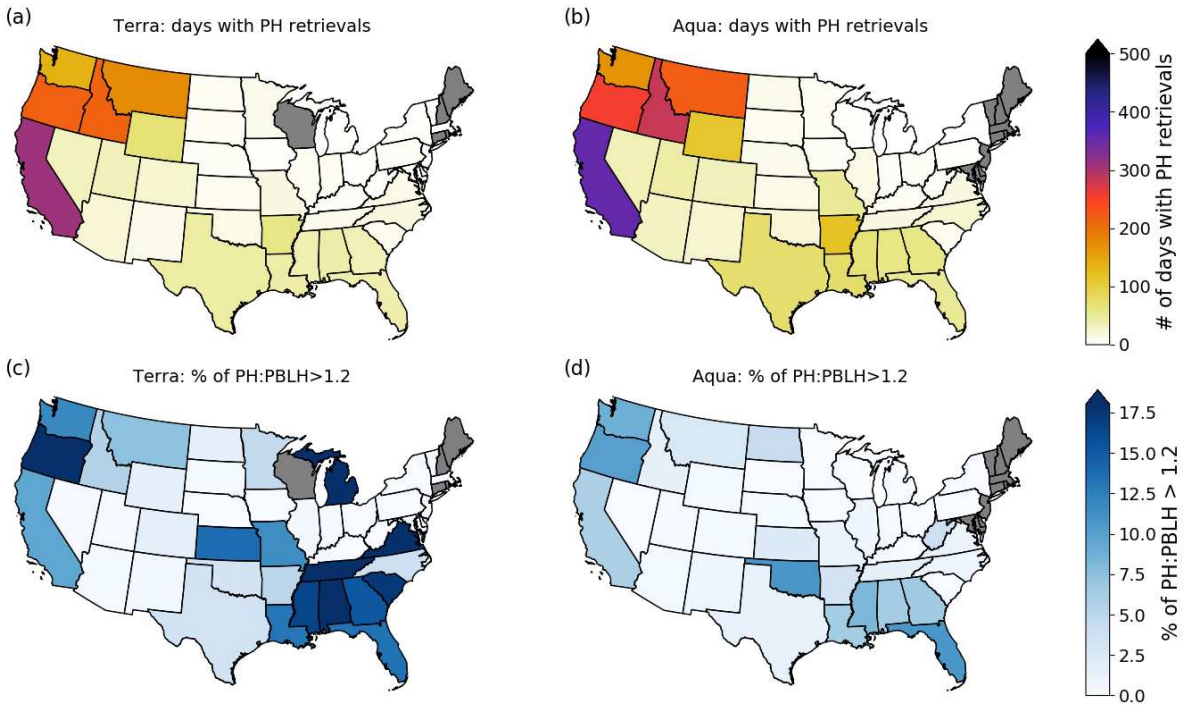

Figure A.3: This figure is the same as Figure 2 (in text) but only uses PH retrievals within 150 km of active fires (determined by the MAIAC thermal hotspots). This figure shows the number of days with PH retrievals during the study period using Terra (a) and Aqua (b). This figure also shows the percent of co-located PH:PBLH values in each state that exceed our 1.2 criteria from Terra (c) and Aqua (d). Gray states indicate that there were no retrievals within 150 km of detected thermal hotspots during the study period.

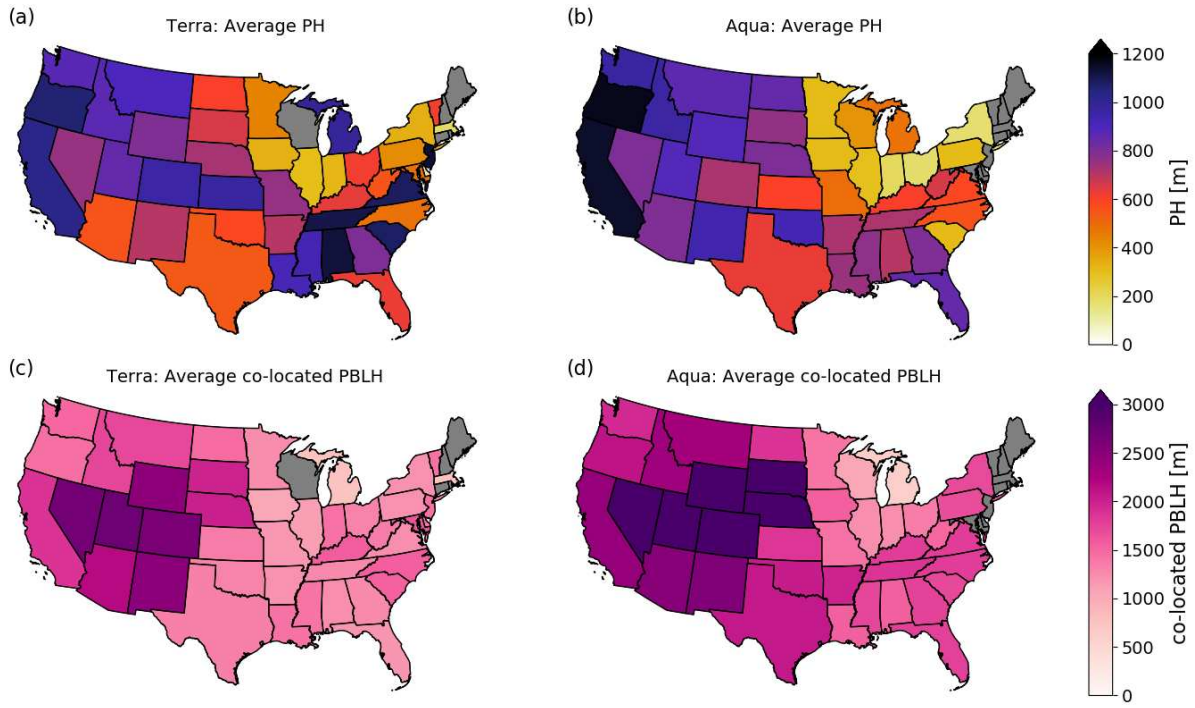

Figure A.4: This figure is the same as Figure A.1 but only uses PH retrievals within 150 km of active fires (determined by the MAIAC thermal hotspots). This figure shows the average smoke PH values from MAIAC Terra (a) and Aqua (b) and average reanalysis PBLH values that were co-located to PH retrievals for Terra (c) and Aqua (d). Gray states indicate that there were no retrievals within 150 km of detected thermal hotspots during the study period.

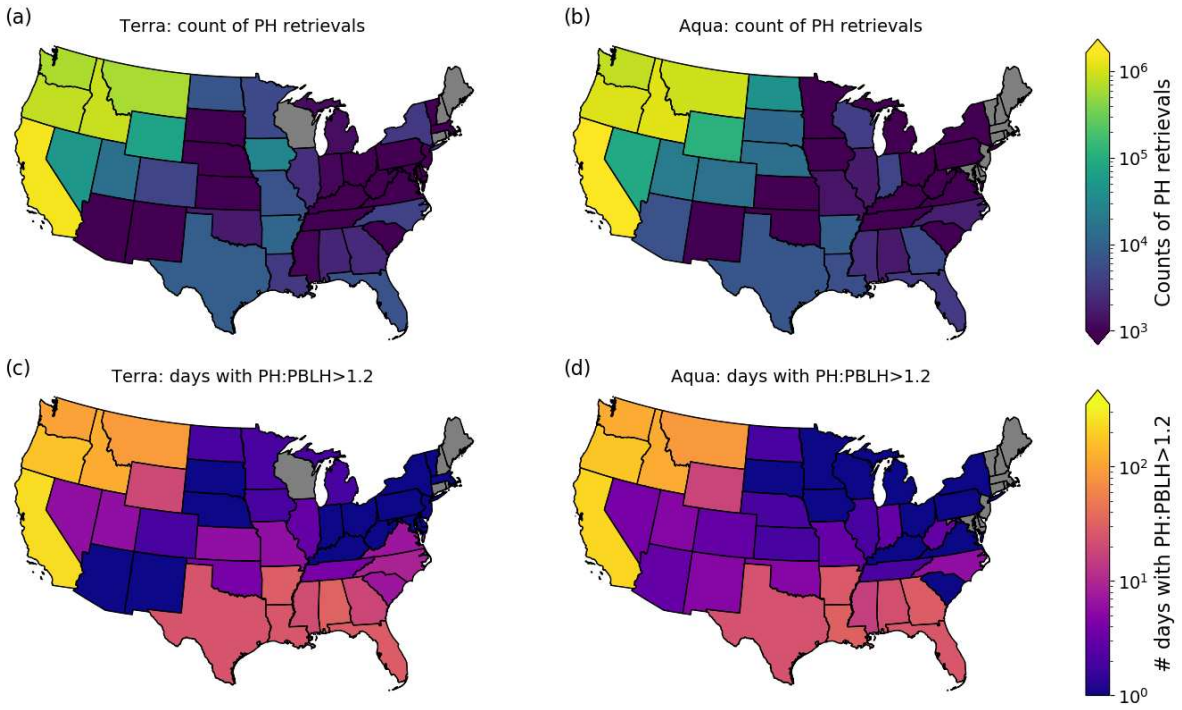

Figure A.5: This figure is the same as Figure A.2 but only uses PH retrievals within 150 km of active fires (determined by the MAIAC thermal hotspots). This figure shows the count of MAIAC PH retrievals in each state for Terra (a) and Aqua (b), as well as number of days that had a PH:PBLH values greater than 1.2 for Terra (c) and Aqua (d). Gray states indicate that there were no retrievals within 150 km of detected thermal hotspots during the study period.

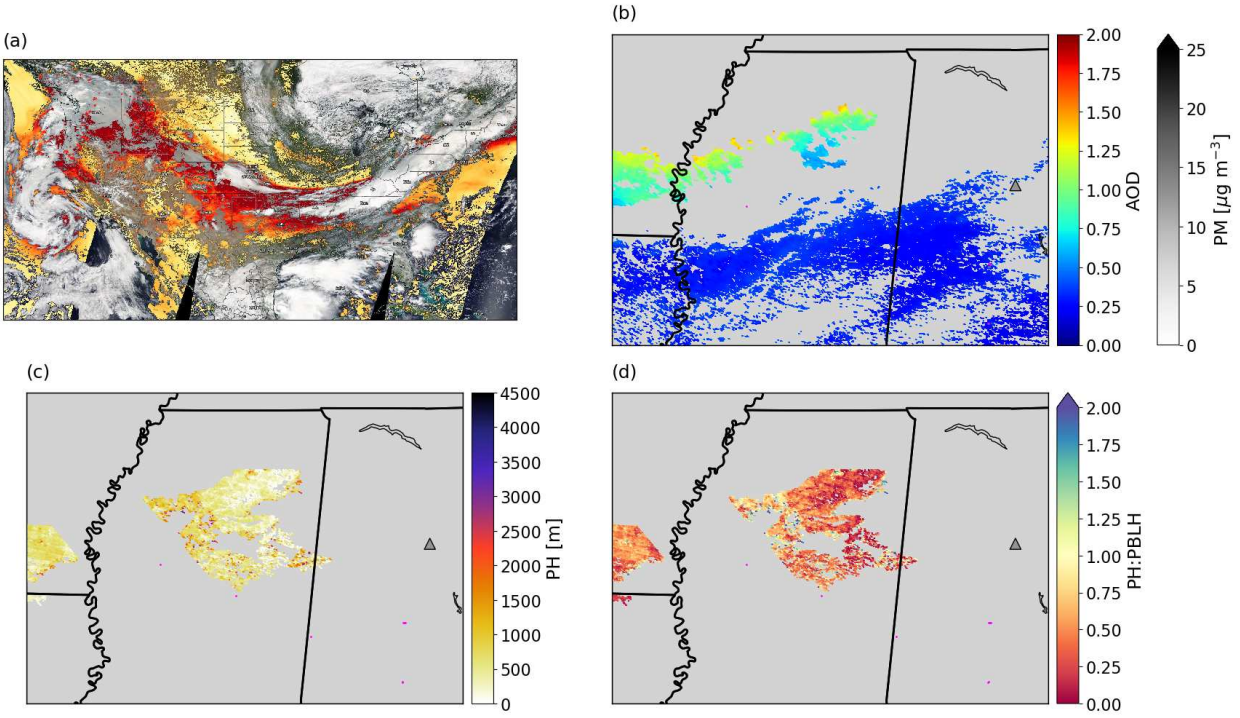

Figure A.6: MODIS Terra true-color image with Terra Dark Target 3km AOD from NASA Worldview (https://worldview.earthdata.nasa.gov/) (a), MAIAC AOD and EPA-AQS $PM_{2.5}$ (b), MAIAC PH (c), and PH:PBLH (d) values using MAIAC PH and NCEP PBLH. This figure displays potential long range transport of smoke from Canada to the eastern US, causing high PH retrievals over Mississippi, despite the lack of any large local fires.

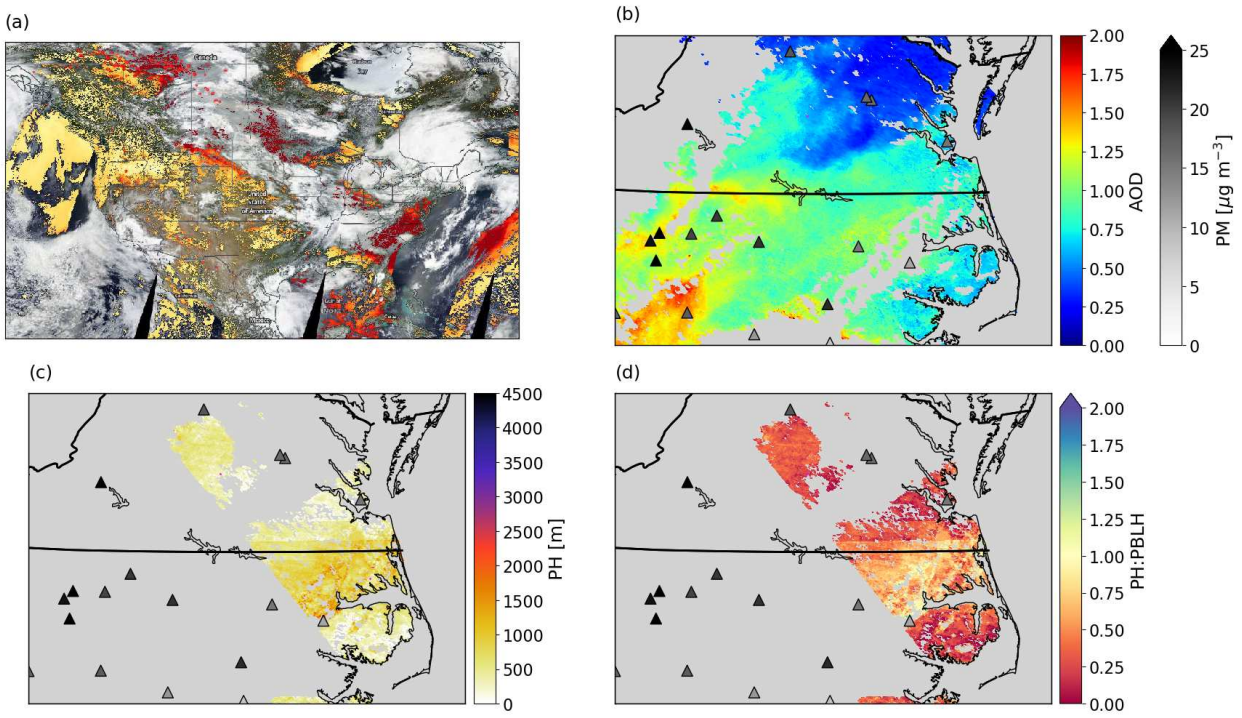

Figure A.7: MODIS Terra true-color image with Terra 3km AOD from NASA Worldview (https://worldview.earthdata.nasa.gov/) (a), MAIAC AOD and EPA-AQS $PM_{2.5}$ (b), MAIAC PH (c), and PH:PBLH (d) values using MAIAC PH and NCEP PBLH. This figure displays potential long range transport of smoke from Canada to the eastern US, causing high PH retrievals over Virginia and North Carolina, despite the lack of any large local fires.

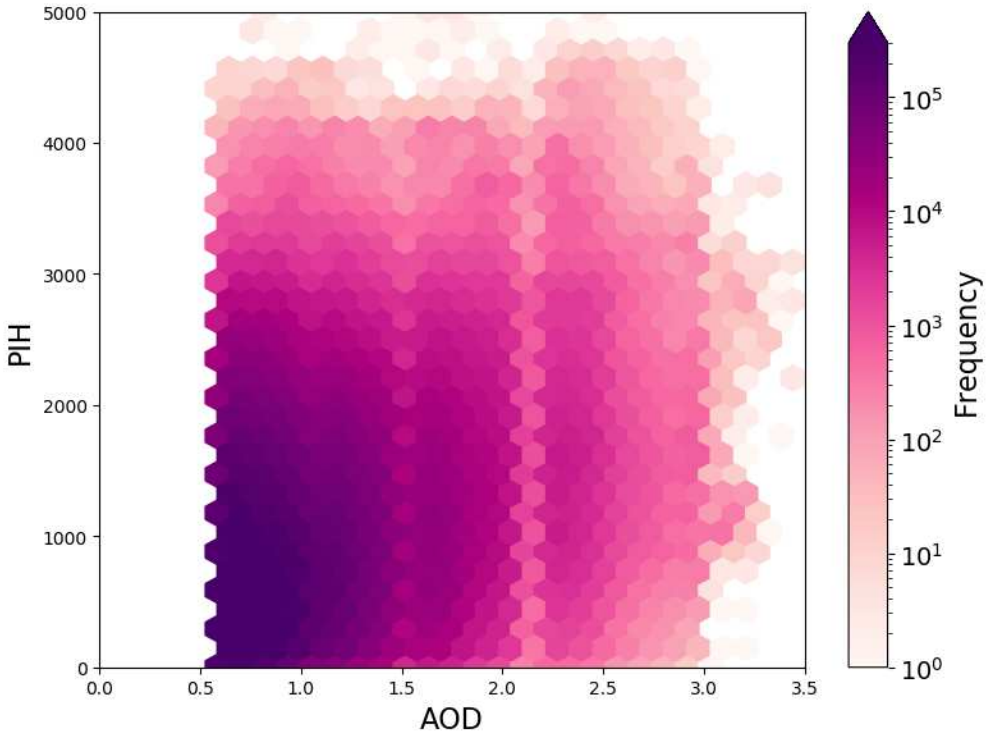

Figure A.8: Hexagonal 2D histogram of co-located PIH and AOD retrievals for the western US from the combined MODIS Terra and Aqua data record.

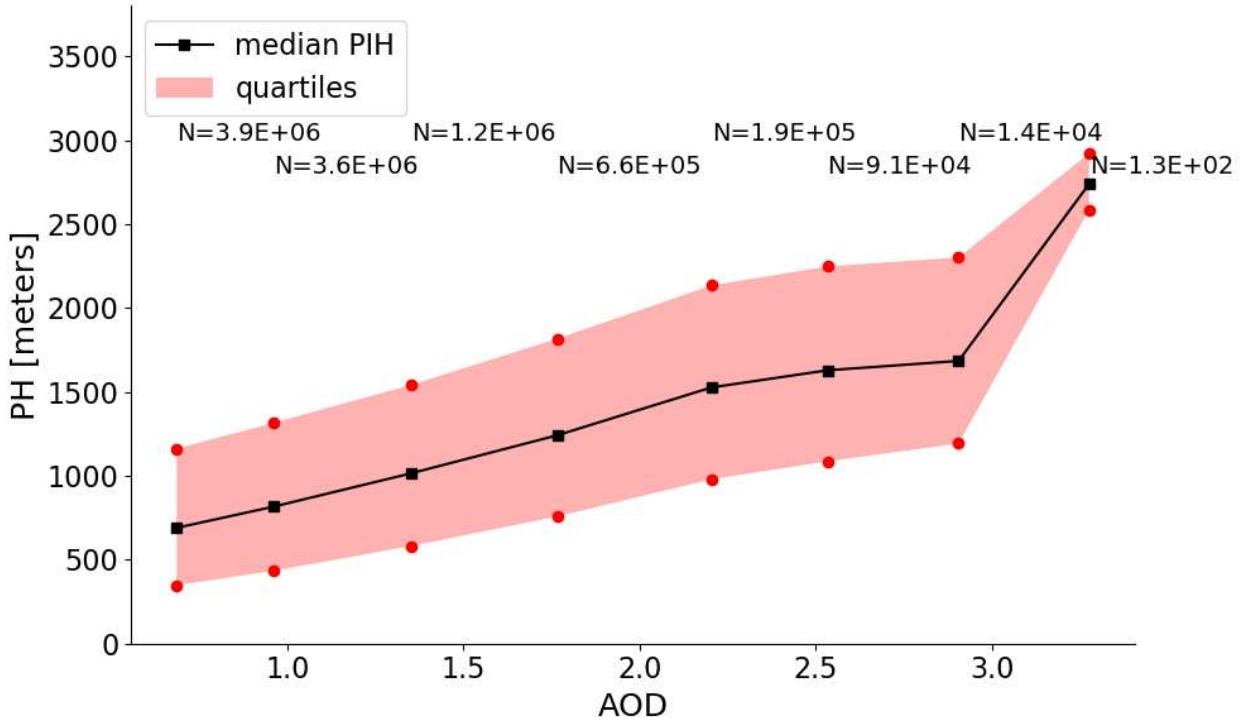

Figure A.9: This is the same as Figure 3 in the text but only uses PH retrievals within 150 km of active fires (determined by the MAIAC thermal hotspots). This figure shows the 25th, 50th, and 75th quartile of PH of co-located PH and AOD retrievals for binned AOD values. The number of PH retrievals for each AOD bin are shown at the top.

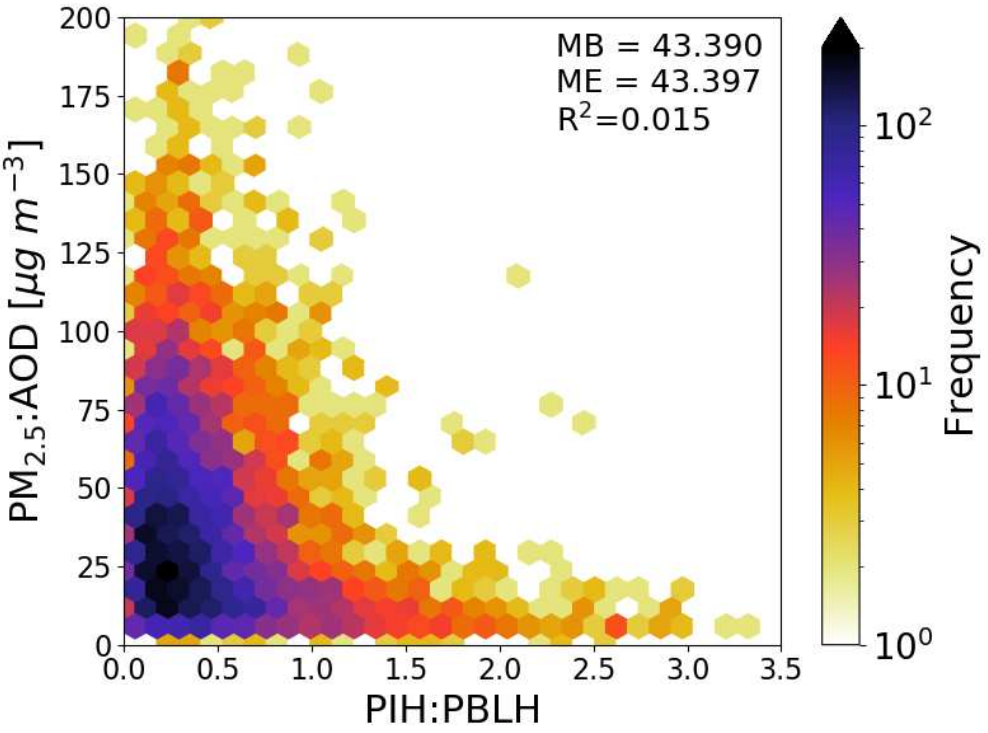

Figure A.10: Hexagonal 2D histogram of co-located PH:PBLH and $PM_{2.5}$:AOD using PH and AOD estimates from MODIS both Terra and Aqua MAIAC datasets, PBLH from NCEP reanalysis, and EPA-AQS 24-hour $PM_{2.5}$ measurements for all co-located data (within 10 km) over the western US for July-September between 2010-2018.

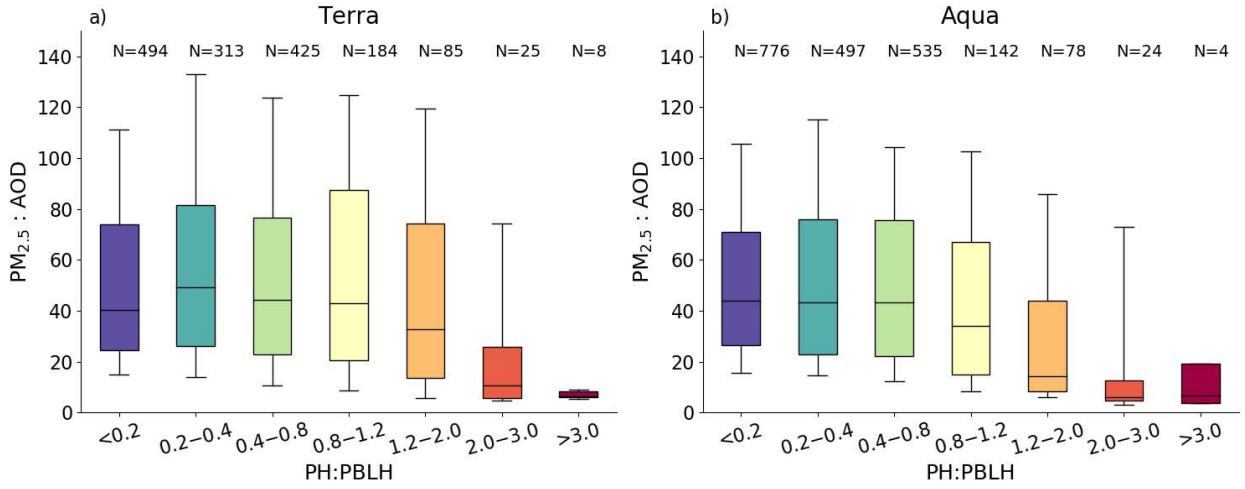

Figure A.11: This figure is the same as Figure 4 in the main text except that the 150 km buffer was applied in this figure. This figure shows the $PM_{2.5}$:AOD as a function of PH:PBLH using PH and AOD estimates from Terra (a) and Aqua (b), PBLH, and $PM_{2.5}$ measurements for all co-located data (within 10 km of a monitor) over the western US for July-September between 2010-2018. The number of observations in each PH:PBLH bin is listed above each box plot. The horizontal lines on each box indicate the 25th, 50th, and 75th percentile of each $PM_{2.5}$:AOD distribution. The whiskers indicate the 10th and 90th percentile of each distribution. The box-and-whisker color ranges from blue to red indicating an increase in PH:PBLH.

Appendix B

Supplementary information for Chapter 3

B.1 Supplementary Tables

Table B.1: Description of EPA-AQS regulatory monitors in Denver, Colorado.

EPA AQS-ID	Latitude, Longitude	Sampling Frequency	Sample Collection Method	Sample Analysis Type
08-001-0010	39.83, -104.94	24 hour	R & P Model 2025 PM-2.5 Sequential Air Sampler w/VSCC	Gravimetric
08-031-0028	39.79, -104.99	1 hour	GRIMM EDM Model 180 with naphion dryer	Laser Light Scattering
08-031-0026	39.78, -105.01	1 hour	Teledyne T640 at 5.0 LPM	Broadband spectroscopy
08-031-0002	39.75, -104.99	24 hour	R & P Model 2025 PM-2.5 Sequential Air Sampler w/VSCC	Gravimetric
08-031-0027	39.73, -105.02	24 hour	R & P Model 2025 PM-2.5 Sequential Air Sampler w/VSCC	Gravimetric
08-031-0013	39.74, -104.94	1 hour	Teledyne T640 at 5.0 LPM	Broadband spectroscopy
08-005-0005	39.60, -105.02	24 hour	R & P Model 2025 PM-2.5 Sequential Air Sampler w/VSCC	Gravimetric
08-035-0004	39.53, -105.07	24 hour	R & P Model 2025 PM-2.5 Sequential Air Sampler w/VSCC	Gravimetric

B.2 Supplementary Figures

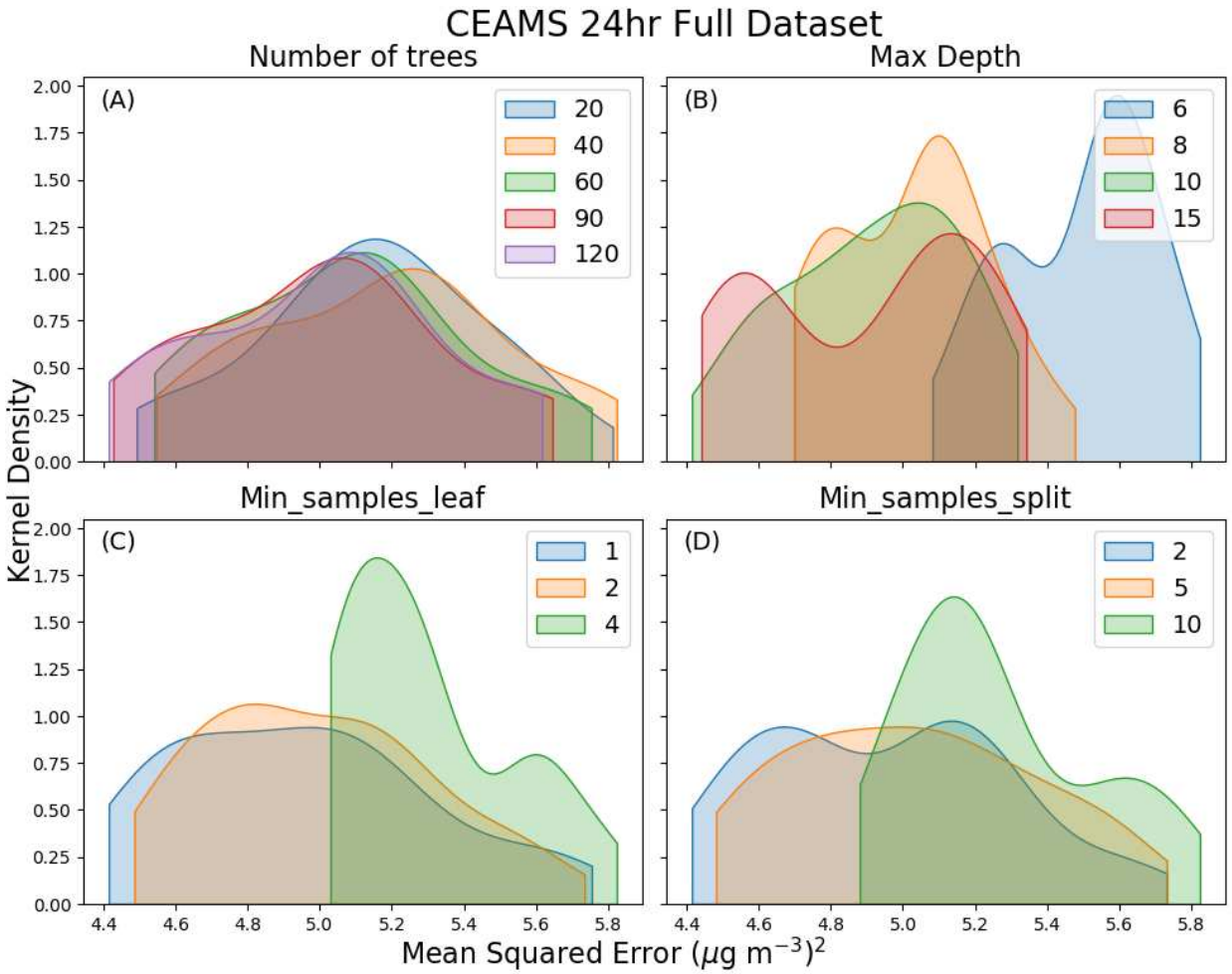

Figure B.1: (a) The probability distribution, smoothed using a kernel function, of mean squared errors of model skill for predicting CEAMS 24-hour $PM_{2.5}$ grouped by the number of trees used in each Random Forest model run. The spread of the distribution is due to how the other hyperparameters are changing (see Table 2 for full list) as well as randomness introduced during the cross-validation. Similarly, the other plots show the kernel density distribution of mean squared errors of model skill grouped by (b) the maximum depth of the trees, (c) minimum samples allowed to form a leaf (end of a branch), and (d) minimum samples needed to split an internal node of a tree.

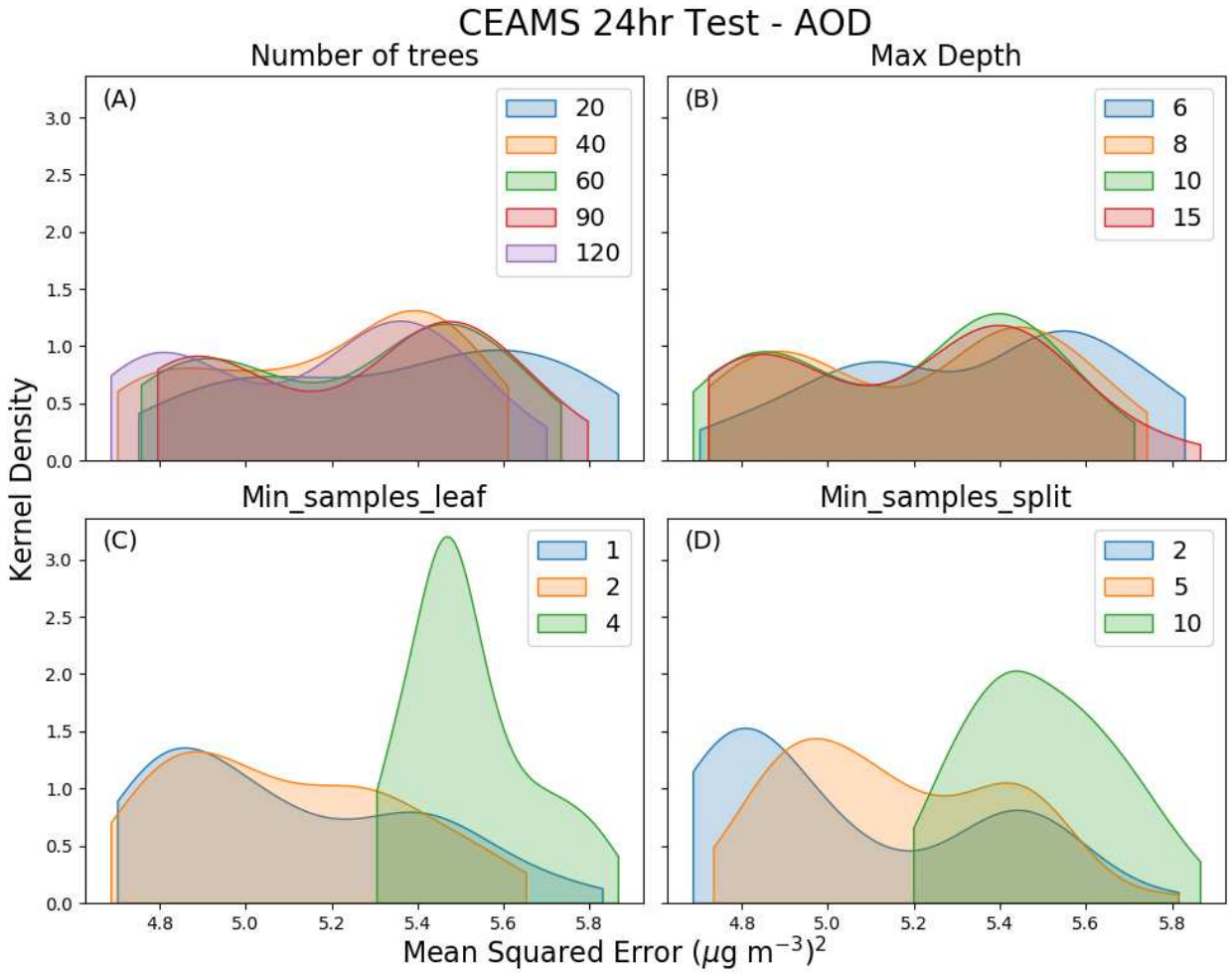

Figure B.2: (a) The probability distribution, smoothed using a kernel function, of mean squared errors of model skill for predicting the "Test - AOD" CEAMS 24-hour $PM_{2.5}$ grouped by the number of trees used in each Random Forest model run. The spread of the distribution is due to how the other hyperparameters are changing (see Table 2 for full list) as well as randomness introduced during the cross-validation. Similarly, the other plots show the kernel density distribution of mean squared errors of model skill grouped by (b) the maximum depth of the trees, (c) minimum samples allowed to form a leaf (end of a branch), and (d) minimum samples needed to split an internal node of a tree.

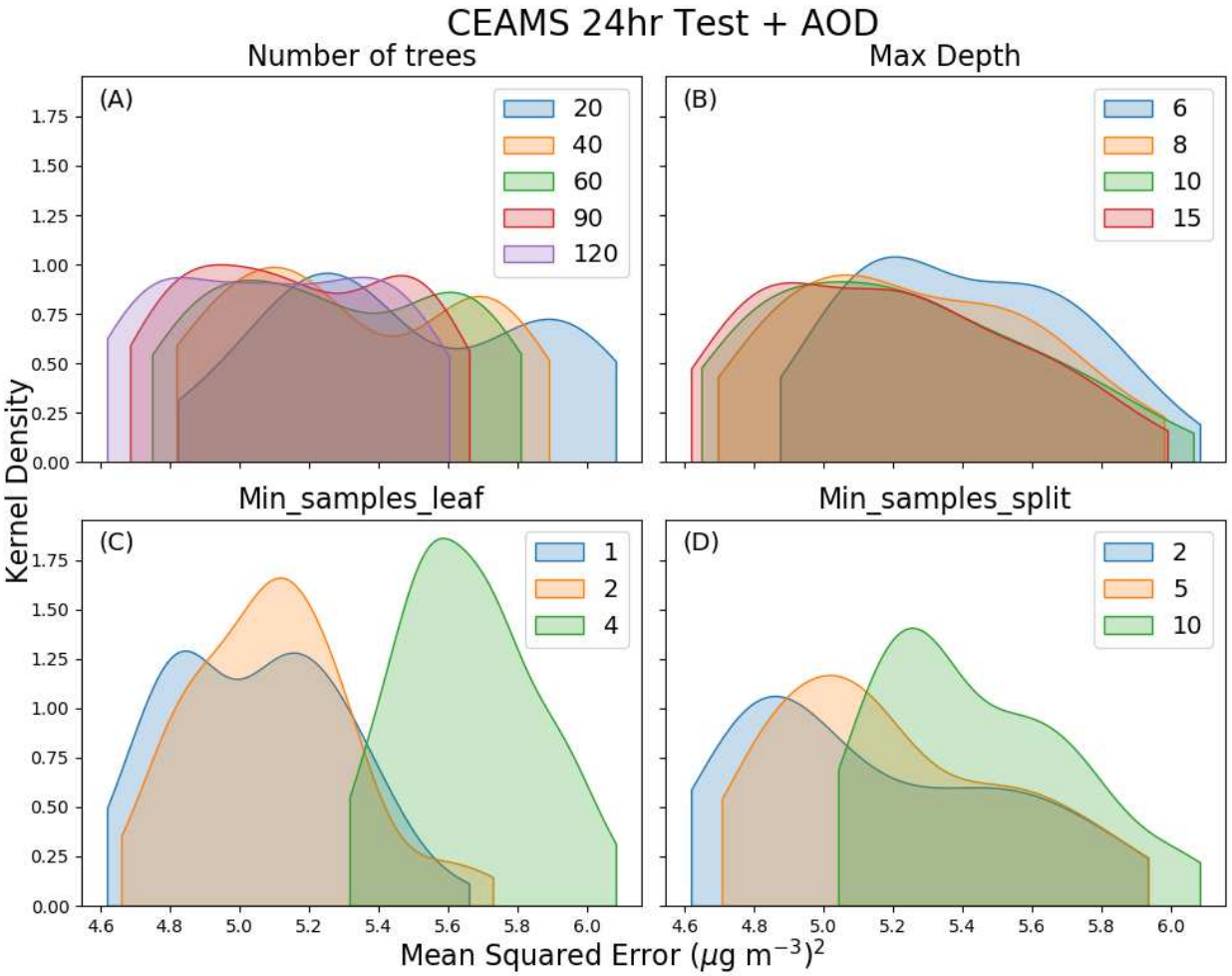

Figure B.3: (a) The probability distribution, smoothed using a kernel function, of mean squared errors of model skill for predicting the "Test + AOD" CEAMS 24-hour $PM_{2.5}$ grouped by the number of trees used in each Random Forest model run. The spread of the distribution is due to how the other hyperparameters are changing (see Table 2 for full list) as well as randomness introduced during the cross-validation. Similarly, the other plots show the kernel density distribution of mean squared errors of model skill grouped by (b) the maximum depth of the trees, (c) minimum samples allowed to form a leaf (end of a branch), and (d) minimum samples needed to split an internal node of a tree.

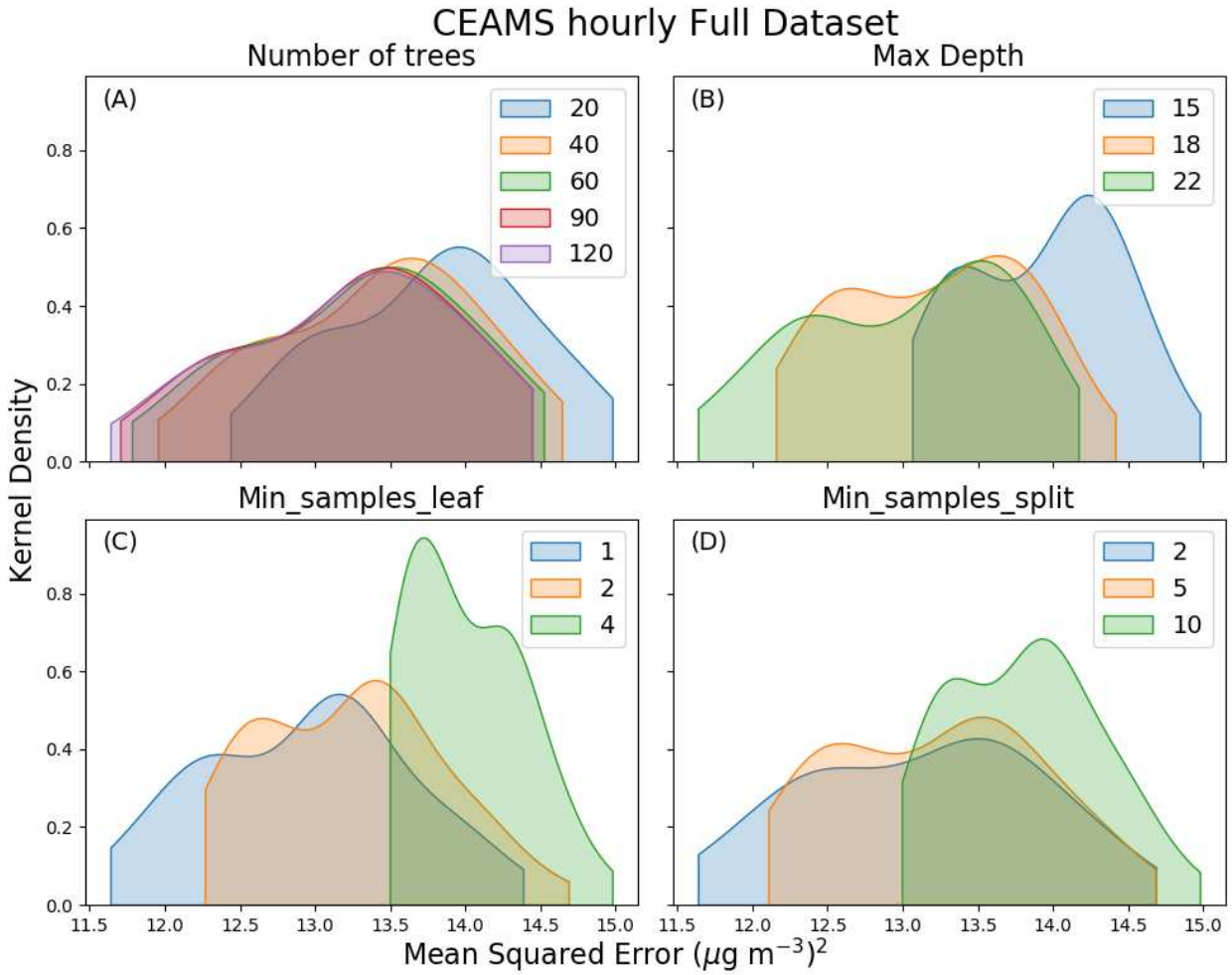

Figure B.4: (a) The probability distribution, smoothed using a kernel function, of mean squared errors of model skill for predicting the "Full dataset" CEAMS hourly $PM_{2.5}$ grouped by the number of trees used in each Random Forest model run. The spread of the distribution is due to how the other hyperparameters are changing (see Table 2 for full list) as well as randomness introduced during the cross-validation. Similarly, the other plots show the kernel density distribution of mean squared errors of model skill grouped by (b) the maximum depth of the trees, (c) minimum samples allowed to form a leaf (end of a branch), and (d) minimum samples needed to split an internal node of a tree.

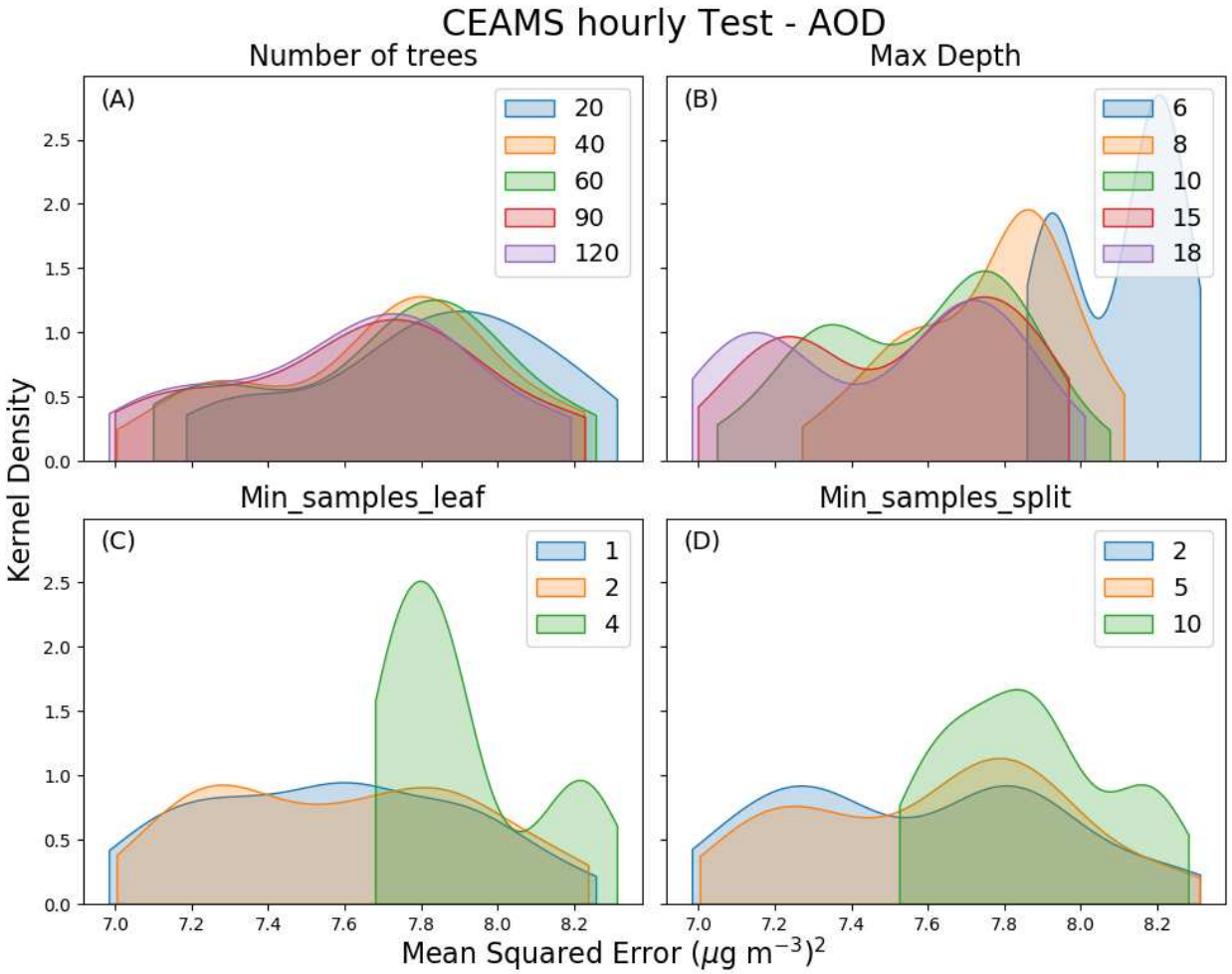

Figure B.5: (a) The probability distribution, smoothed using a kernel function, of mean squared errors of model skill for predicting the "Test - AOD" CEAMS hourly $PM_{2.5}$ grouped by the number of trees used in each Random Forest model run. The spread of the distribution is due to how the other hyperparameters are changing (see Table 2 for full list) as well as randomness introduced during the cross-validation. Similarly, the other plots show the kernel density distribution of mean squared errors of model skill grouped by (b) the maximum depth of the trees, (c) minimum samples allowed to form a leaf (end of a branch), and (d) minimum samples needed to split an internal node of a tree.

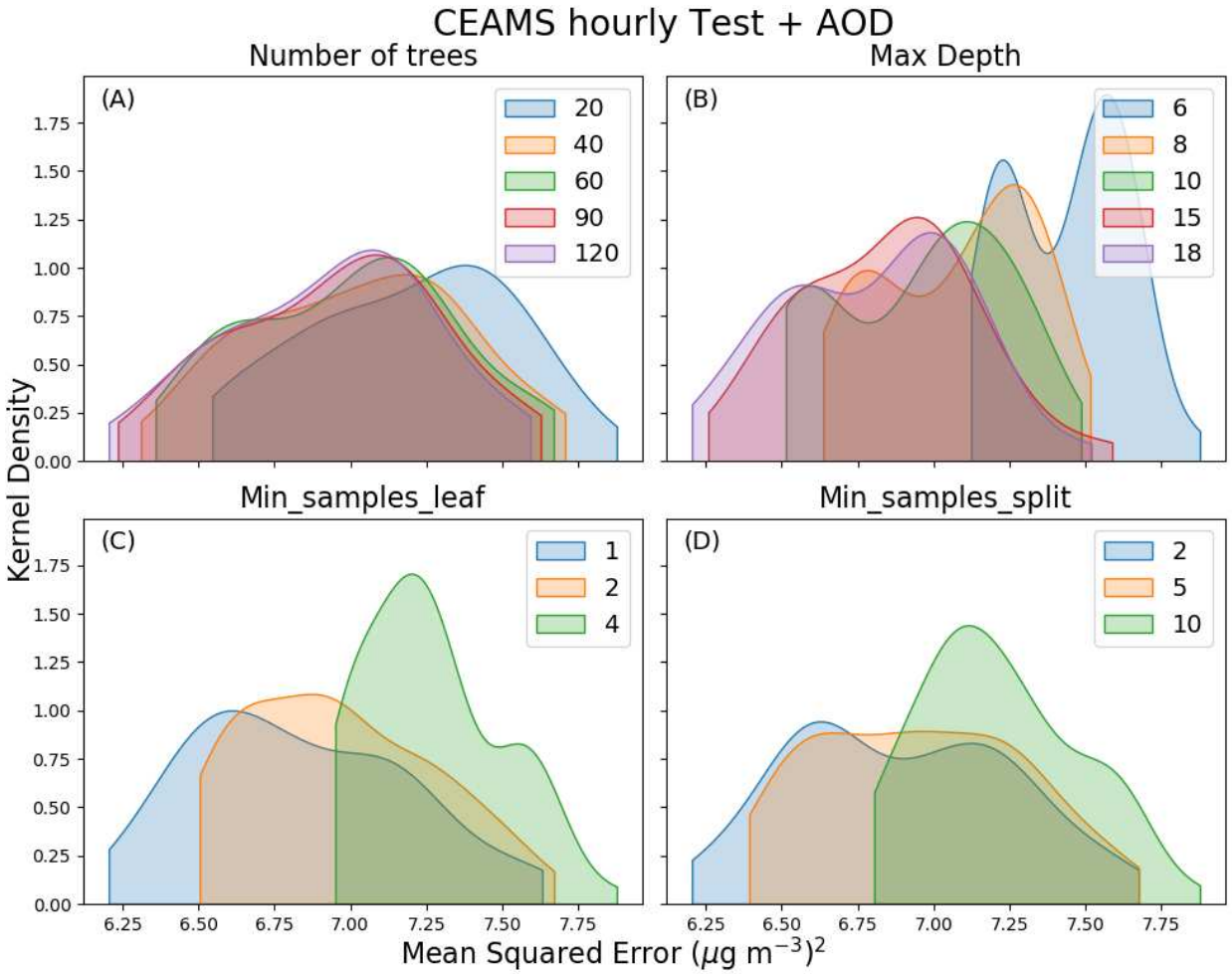

Figure B.6: (a) The probability distribution, smoothed using a kernel function, of mean squared errors of model skill for predicting the "Test + AOD" CEAMS 24-hour $PM_{2.5}$ grouped by the number of trees used in each Random Forest model run. The spread of the distribution is due to how the other hyperparameters are changing (see Table 2 for full list) as well as randomness introduced during the cross-validation. Similarly, the other plots show the kernel density distribution of mean squared errors of model skill grouped by (b) the maximum depth of the trees, (c) minimum samples allowed to form a leaf (end of a branch), and (d) minimum samples needed to split an internal node of a tree.

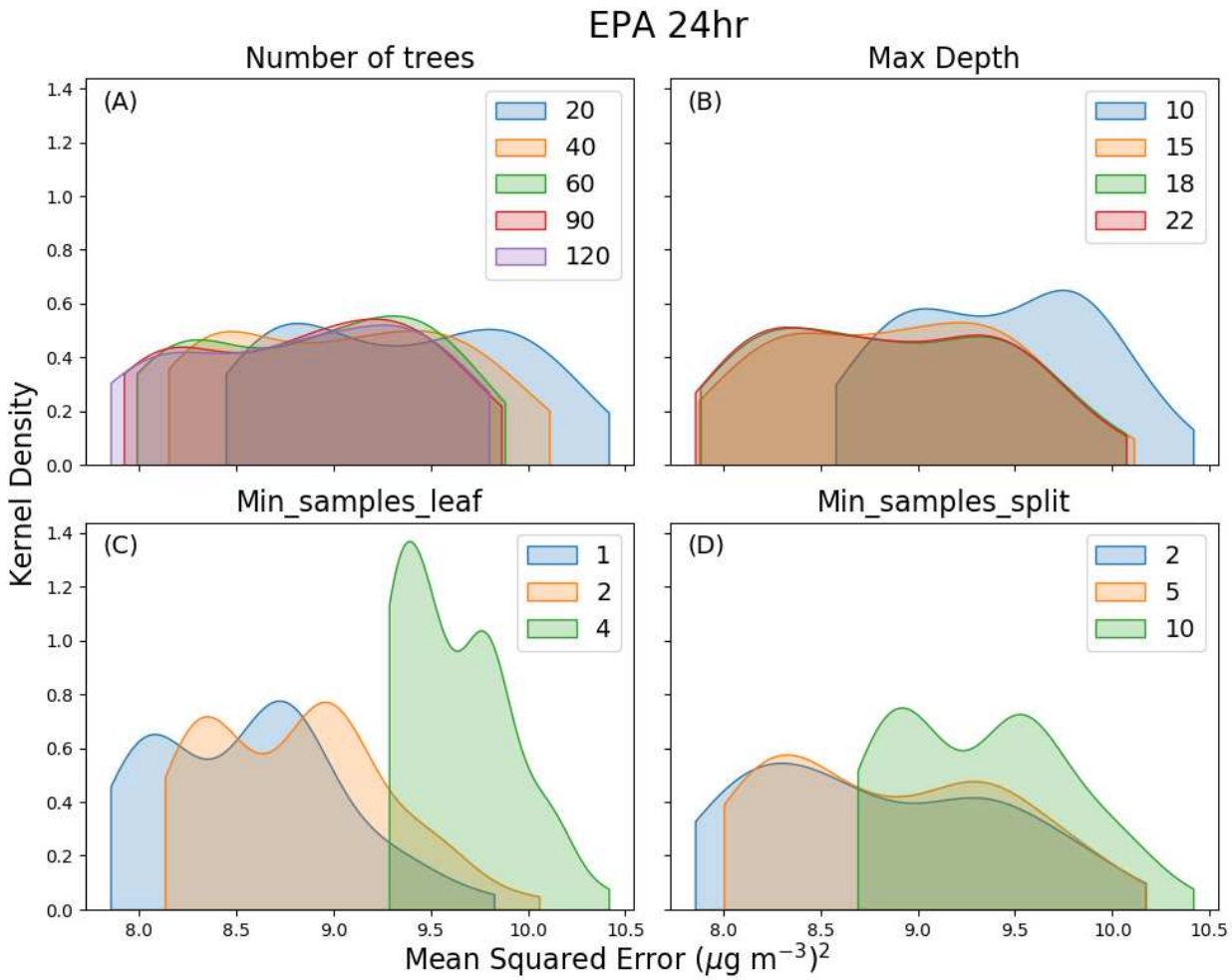

Figure B.7: (a) The probability distribution, smoothed using a kernel function, of mean squared errors of model skill for predicting the EPA 24-hour $PM_{2.5}$ grouped by the number of trees used in each Random Forest model run. The spread of the distribution is due to how the other hyperparameters are changing (see Table 2 for full list) as well as randomness introduced during the cross-validation. Similarly, the other plots show the kernel density distribution of mean squared errors of model skill grouped by (b) the maximum depth of the trees, (c) minimum samples allowed to form a leaf (end of a branch), and (d) minimum samples needed to split an internal node of a tree.

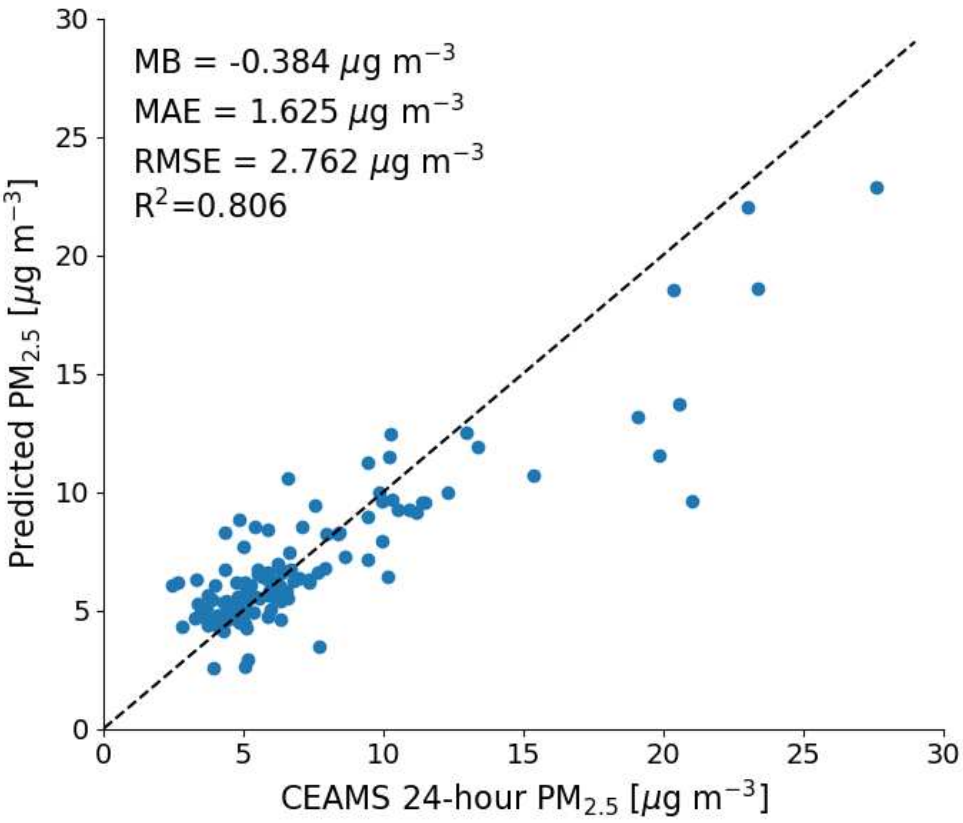

Figure B.8: RF 24-hour $PM_{2.5}$ predictions from 1 testing fold during the 5-fold cross-validation versus the CEAMS 24-hour $PM_{2.5}$ measurements. The mean bias (MB), mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2) are given.

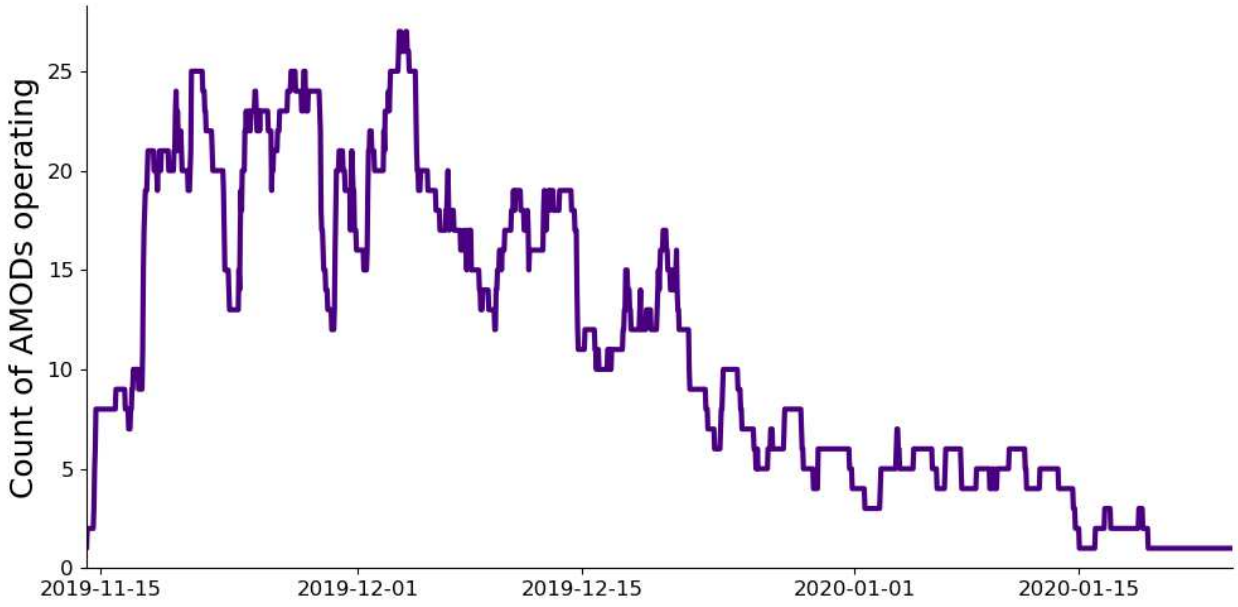

Figure B.9: Time series of the number of AMODs operating per hour of the CEAMS deployment in Denver. CO.

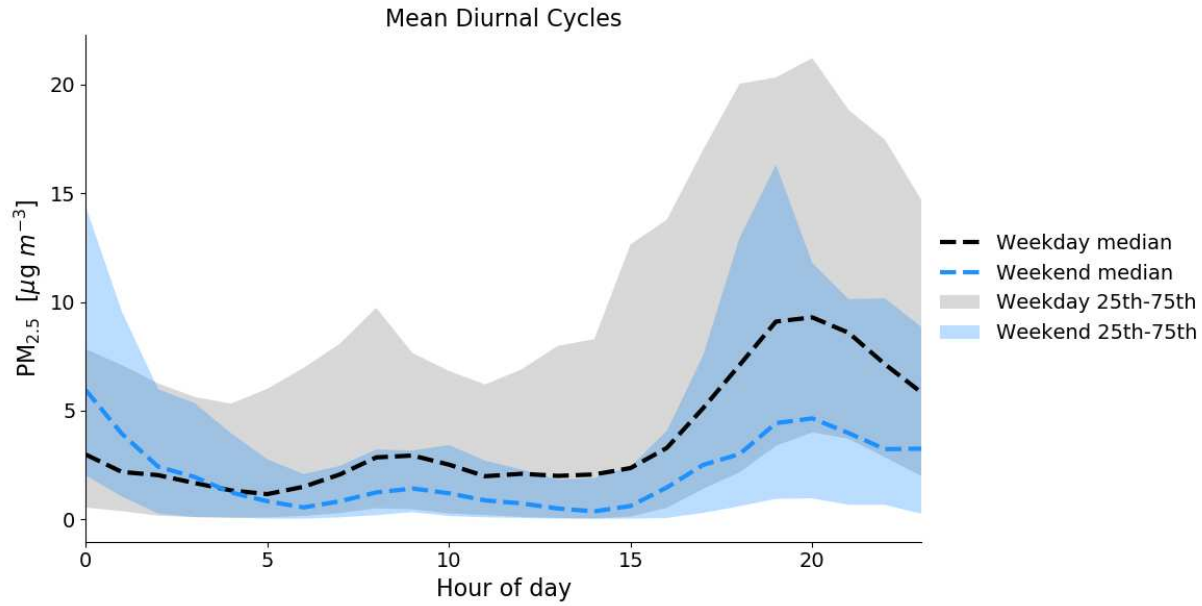

Figure B.10: Median hourly averaged diurnal cycles of weekend (blue) and weekday (black) $PM_{2.5}$ from the Denver, CO wintertime CEAMS deployment. $PM_{2.5}$ measurements from major holidays were removed. The range between the 25th and 75th percentile of each hourly average is shown for the weekend (blue shading) and weekday (grey shading) averages.

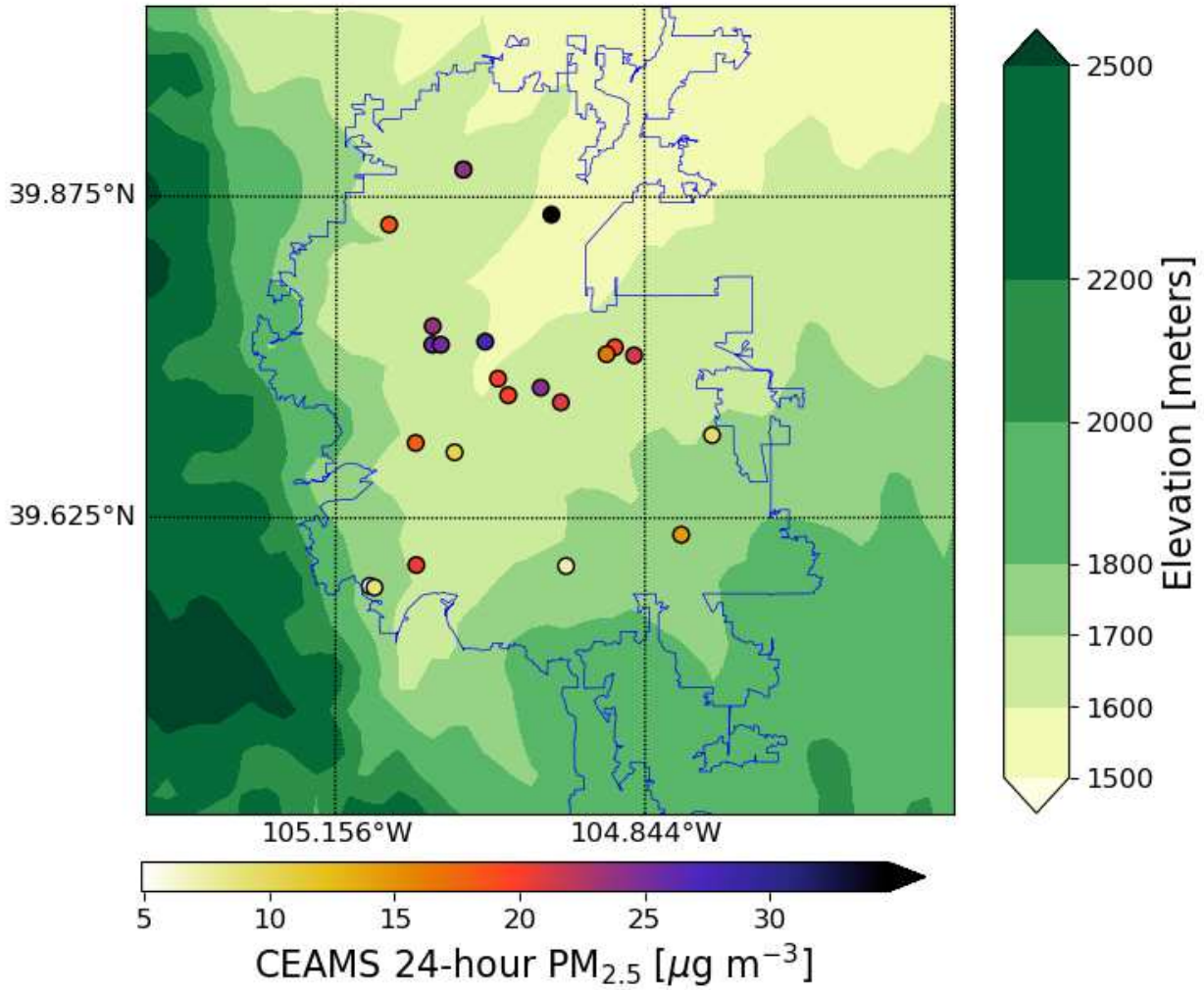

Figure B.11: Map of elevation (Amante and Eakins, 2009) and 24-hour $PM_{2.5}$ averages (points) from the 22 CEAMS low-cost sensors on December 6th, 2019, in Denver, CO. The greater Denver-Aurora area is outlined in blue (based on cartographic files from 2015 TIGER/Line Shapefiles).

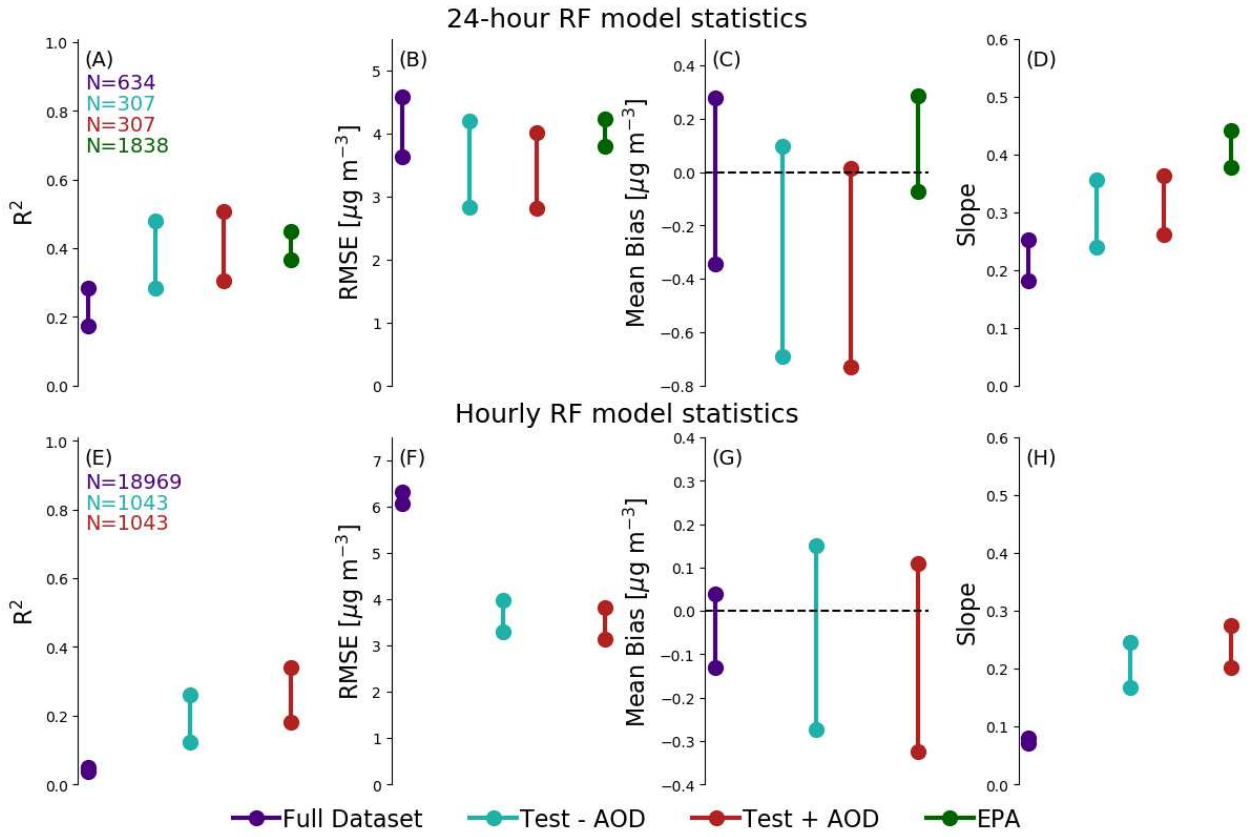

Figure B.12: RF model performance metrics for $PM_{2.5}$ measurements using unshuffled or "consecutive" k-folds (24-hour in the top row and hourly in the bottom row). The 95% confidence interval of the error metrics for all of the CEAMS RF models (Full Dataset, Test - AOD, and Test + AOD) in predicting both 24-hour and hourly $PM_{2.5}$ and the error metrics for the 24-hour EPA model. The 95% confidence intervals show an estimate of the uncertainty range and, thus, if the intervals of two different models overlap, any difference in their error metrics are likely not statistically significant. The error metrics for each 24-hour $PM_{2.5}$ RF model includes (a) the coefficient of determination (R^2) (b) root mean squared error (RMSE), (c) mean bias, (d) and slope of the linear regression. Plots (e), (f), (g), and (h) show analogous results but for the hourly $PM_{2.5}$ predictions, which we did not predict for the EPA dataset. The size of each 24-hour and hourly dataset, before being split into k-folds, is shown in the top left corner of plot (a) and (e).

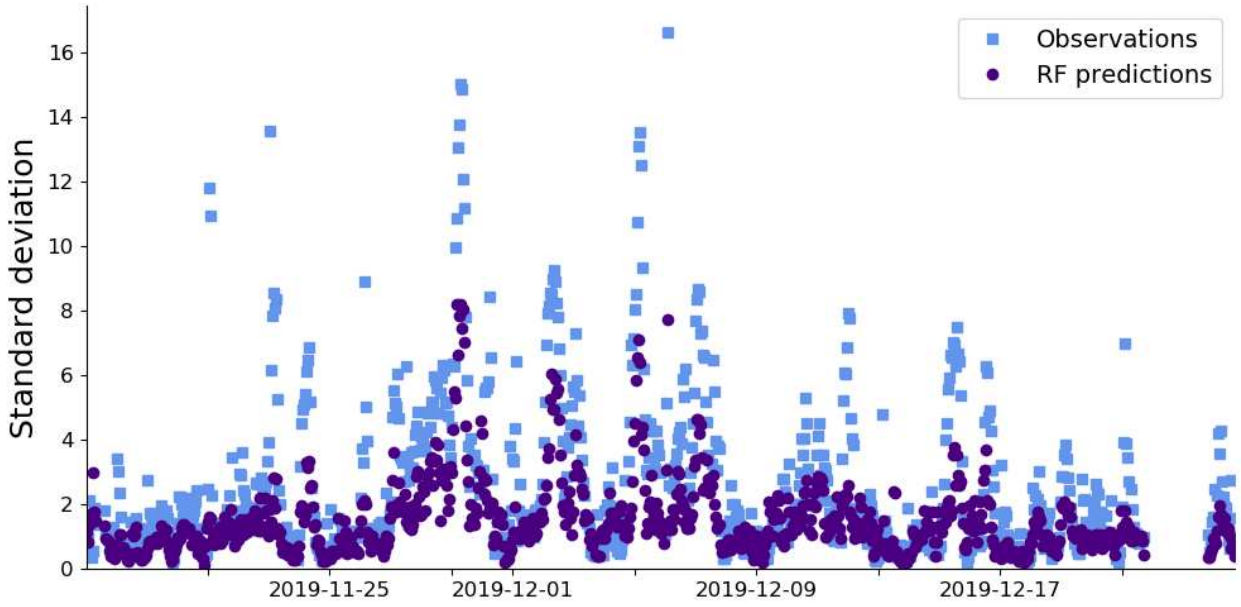

Figure B.13: Time series that shows the standard deviation of the hourly $PM_{2.5}$ for each hour of observations (light blue squares) and CEAMS RF predictions (dark blue circles) for hours that had at least 10 monitors operating at the same time. This plot shows results from RF models that used shuffled k-folds.

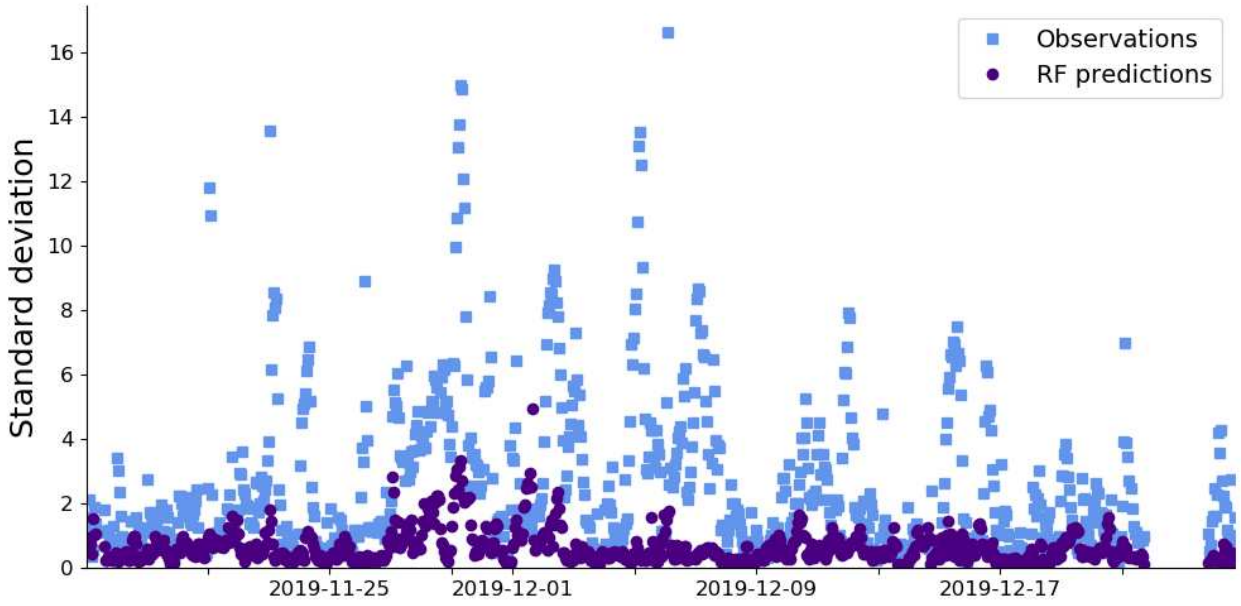

Figure B.14: The same as Figure B.13 but when unshuffled k-folds were used during the training and validation of the RF models predicting CEAMS hourly $PM_{2.5}$.

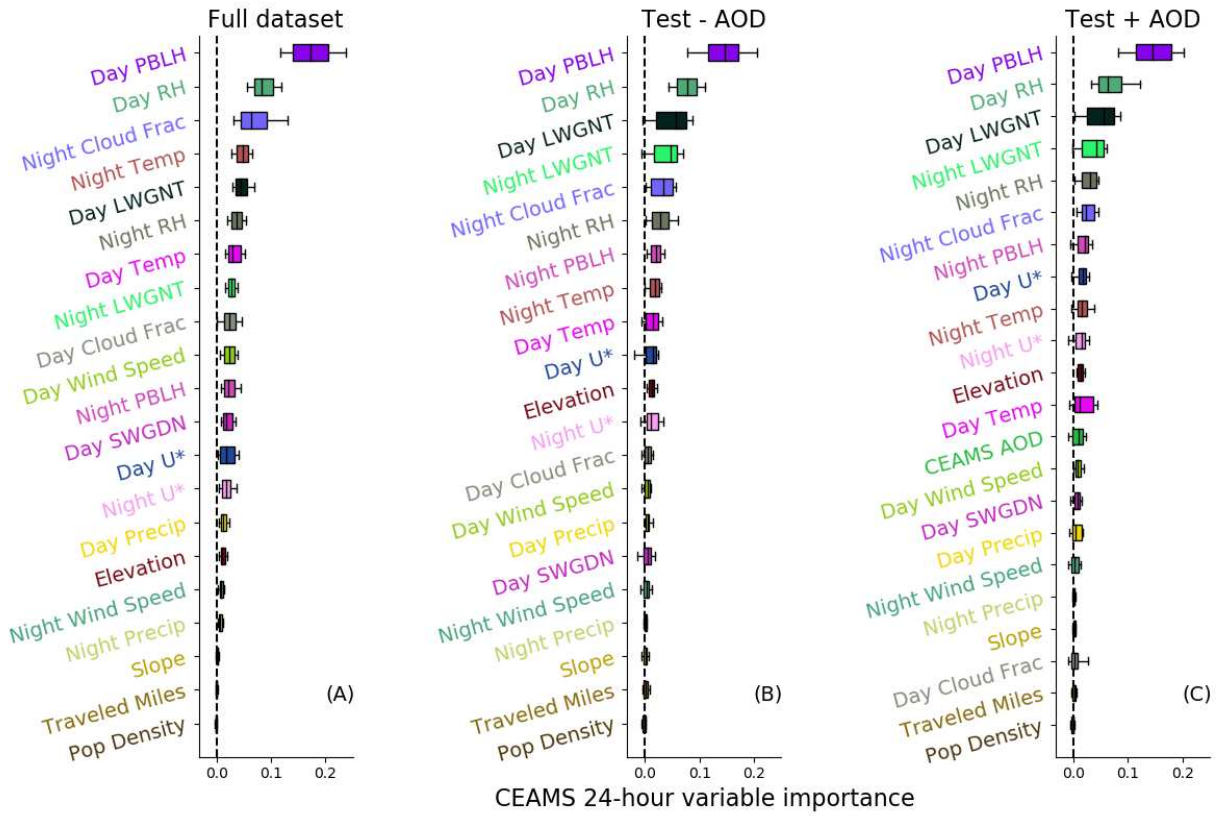

Figure B.15: Box-and-whisker plots of 500 permutation importance values for all of the predictors in the RF models that predict CEAMS 24-hour $PM_{2.5}$. The 500 permutation importance values are taken from 100 repeats of permutation importance from each of the 5 testing folds. The whiskers of each box are the 10th and 90th percentile of the permutation importance distribution. The edges of each box represent the 25th and 75th percentile and, finally, the centerline of each box represents the median (i.e., 50th percentile) of the permutation importance distribution. (a) The 24-hour $PM_{2.5}$ predictions of the CEAMS "Full dataset", which contained all of the available 24-hour $PM_{2.5}$ averages regardless of whether daily AOD was available from each location and day. (b) The 24-hour $PM_{2.5}$ predictions of the CEAMS "Test - AOD" dataset, which only contained 24-hour $PM_{2.5}$ averaged and the associated predictors at locations and days where daily AOD was also available, but we did not use AOD as a predictor for this model. (c) The 24-hour $PM_{2.5}$ predictions of the CEAMS "Test + AOD" dataset, which only contained $PM_{2.5}$ data where AOD was available and we used AOD as an additional predictor to the meteorological and geographical predictors.

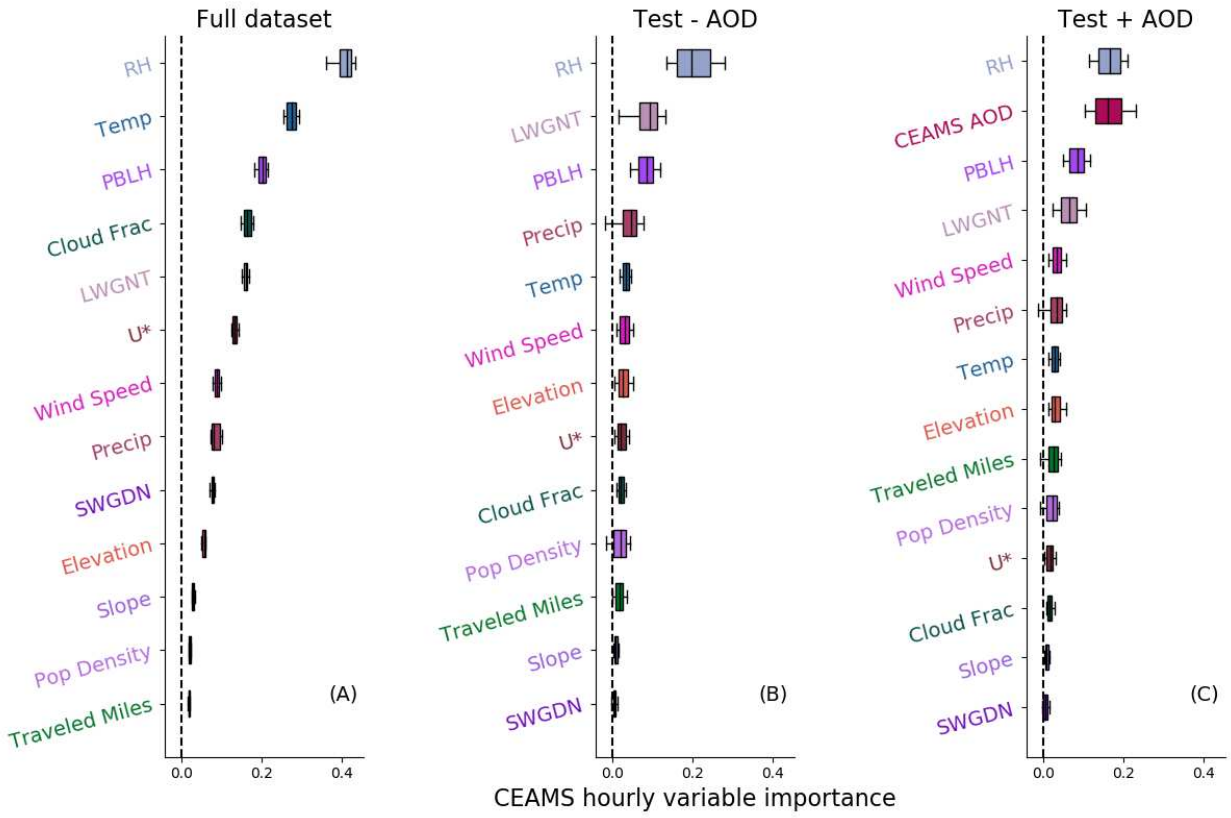

Figure B.16: Box-and-whisker plots of 500 permutation importance values for all of the predictors in the RF models that predict CEAMS hourly $PM_{2.5}$. The 500 permutation importance values are taken from 100 repeats of permutation importance from each of the 5 testing folds. The whiskers of each box are the 10th and 90th percentile of the permutation importance distribution. The edges of each box represent the 25th and 75th percentile and, finally, the centerline of each box represents the median (i.e., 50th percentile) of the permutation importance distribution. (a) The hourly $PM_{2.5}$ predictions of the CEAMS "Full dataset", which contained all of the available 24-hour $PM_{2.5}$ averages regardless of whether daily AOD was available from each location and hour. (b) The hourly $PM_{2.5}$ predictions of the CEAMS "Test - AOD" dataset, which only contained 24-hour $PM_{2.5}$ averaged and the associated predictors at locations and days where daily AOD was also available, but we did not use AOD as a predictor for this model. (c) The hourly $PM_{2.5}$ predictions of the CEAMS "Test + AOD" dataset, which only contained hourly $PM_{2.5}$ data where AOD was available and we used AOD as an additional predictor to the meteorological and geographical predictors.

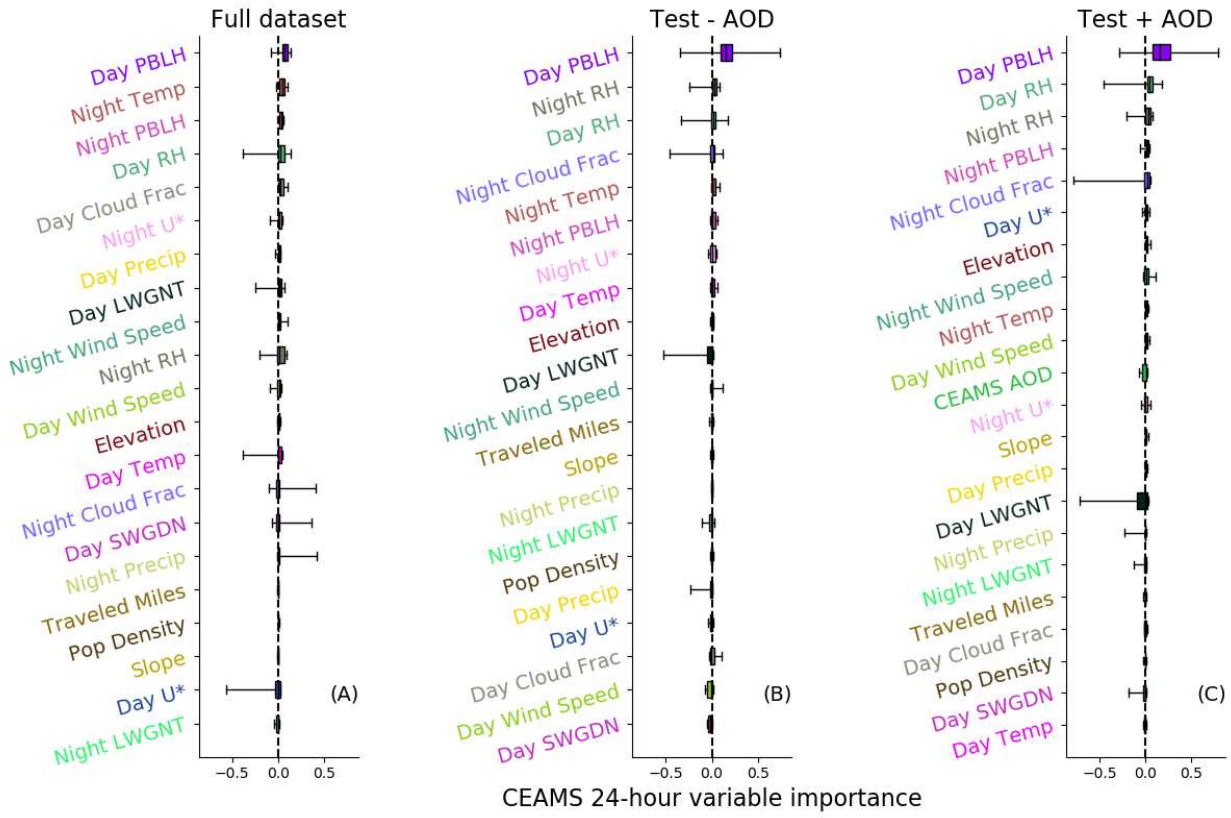

Figure B.17: The same as Figure B.14 but when unshuffled k-folds were used during the training and validation of the RF models predicting CEAMS 24-hour $PM_{2.5}$.

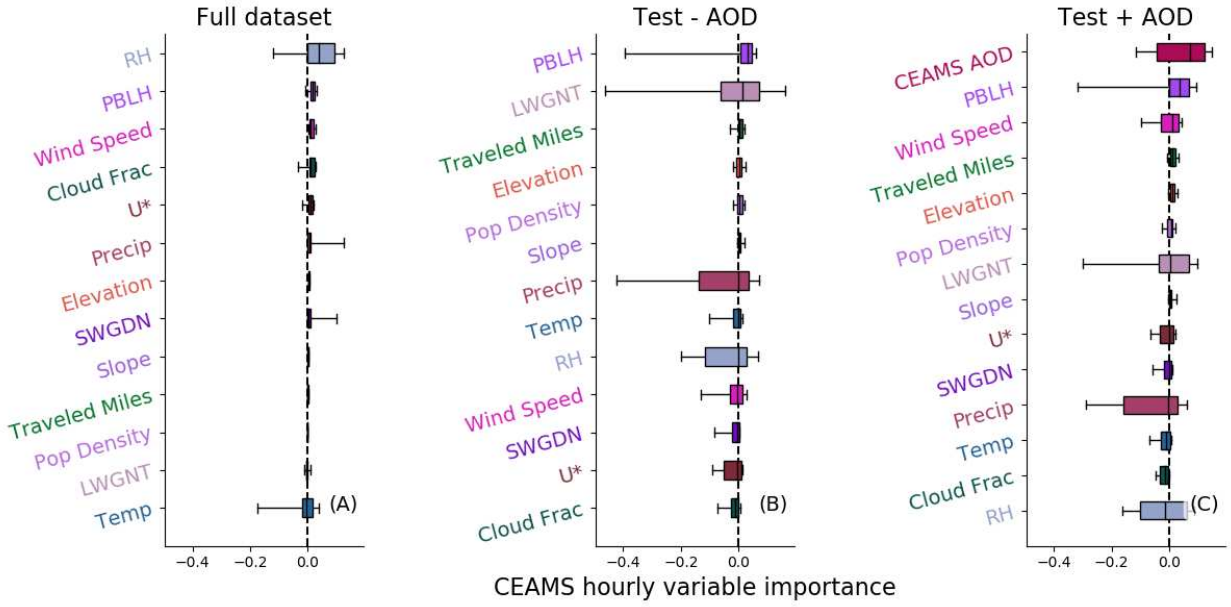

Figure B.18: The same as Figure B.16 but when unshuffled k-folds were used during the training and validation of the RF models predicting CEAMS hourly $PM_{2.5}$.

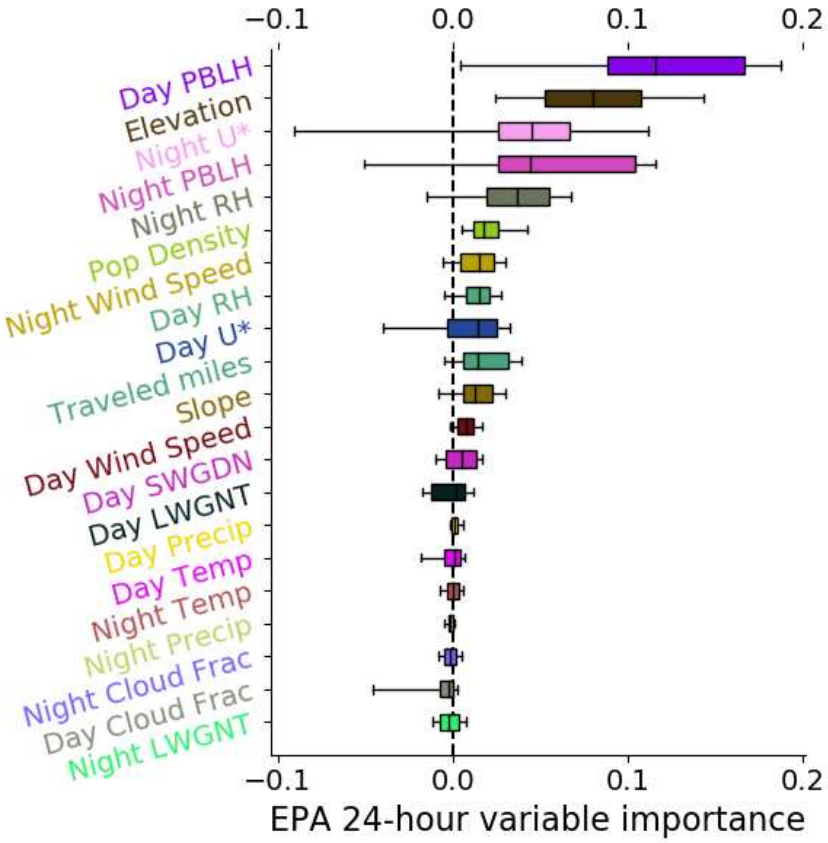

Figure B.19: Box-and-whisker plots of 500 permutation importance values for all of the predictors in the RF models that predict EPA 24-hour $PM_{2.5}$ using consecutive k-folds. The 500 permutation importance values are taken from 100 repeats of permutation importance from each of the 5 testing folds. The whiskers of each box are the 10th and 90th percentile of the permutation importance distribution. The edges of each box represent the 25th and 75th percentile and, finally, the centerline of each box represents the median (i.e., 50th percentile) of the permutation importance distribution.

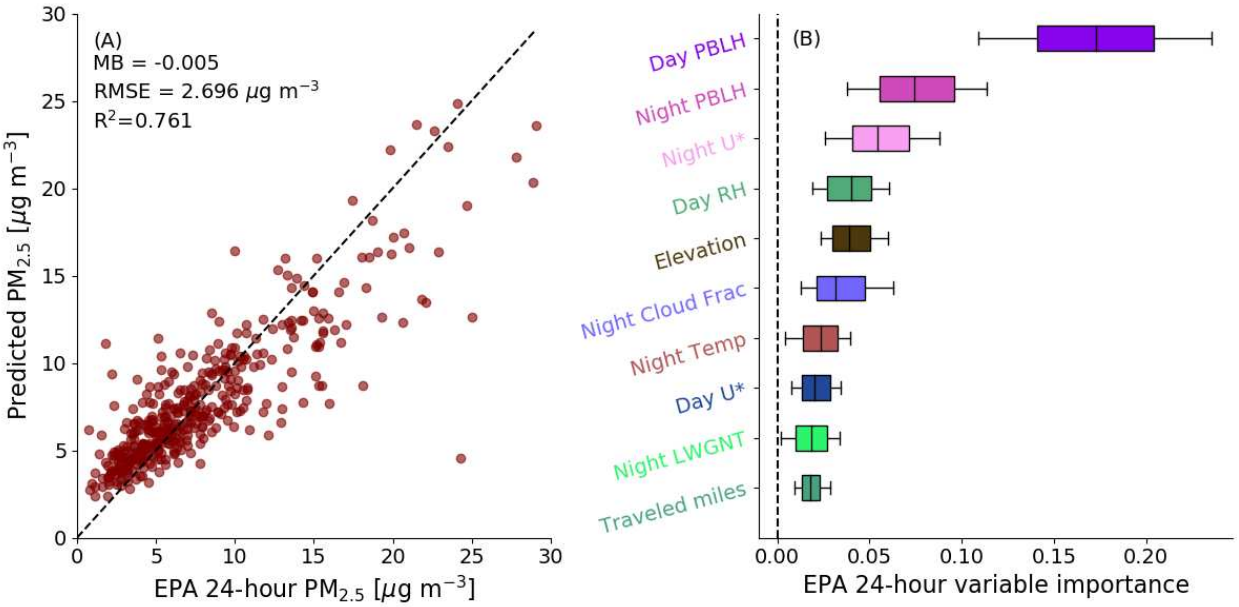

Figure B.20: (a) All points from the testing folds of the 5-fold CV for the EPA 24-hour RF model for only one winter (Dec. 15 - Jan. 15, 2019) and shuffled k-folds. (b) Box-and-whisker plots of the distribution of 100 permutation importance metrics for the top 10 ranked predictors of the 24-hour EPA $PM_{2.5}$ for one winter (Dec. 15 - Jan. 15, 2019) and shuffled k-folds.

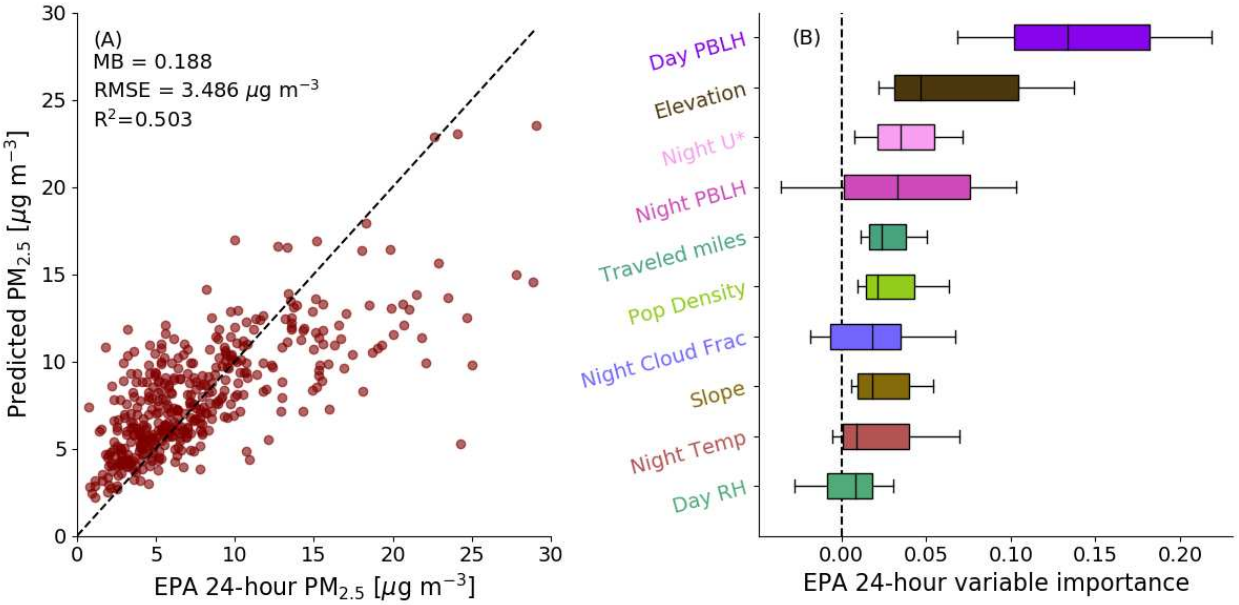

Figure B.21: (a) All points from the testing folds of the 5-fold CV for the EPA 24-hour RF model for one winter (Dec. 15 - Jan. 15, 2019) and consecutive k-folds. (b) Box-and-whisker plots of the distribution of 100 permutation importance metrics for the top 10 ranked predictors of the 24-hour EPA $PM_{2.5}$ for one winter (Dec. 15 - Jan. 15, 2019) and consecutive k-folds.

Appendix C

Supplementary information for Chapter 4

C.1 Supplementary Tables

Table C.1: Count of public schools used for each distribution of Figure 4.2.

Fraction of students eligible for subsidized meals	**<0.3**	**0.3-0.6**	**>0.6**
Urban most white school count	1285	917	281
Urban most non-white school count	637	2230	12408
Suburban, town and rural most white school count	10225	14217	4098
Suburban, town and rural most non-white school count	953	3278	11517

C.2 Supplementary Figures

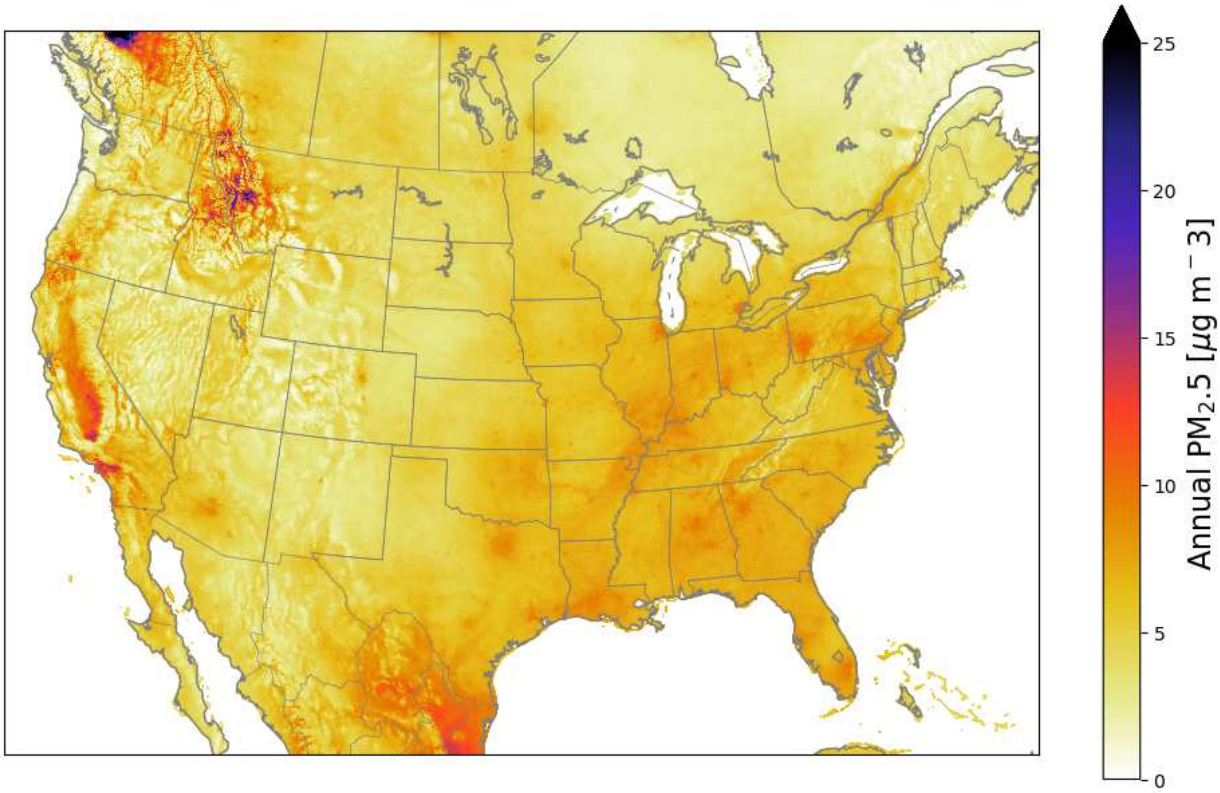

Figure C.1: Hammer et al. (2020) annually averaged $PM_{2.5}$ concentrations over the continental US.

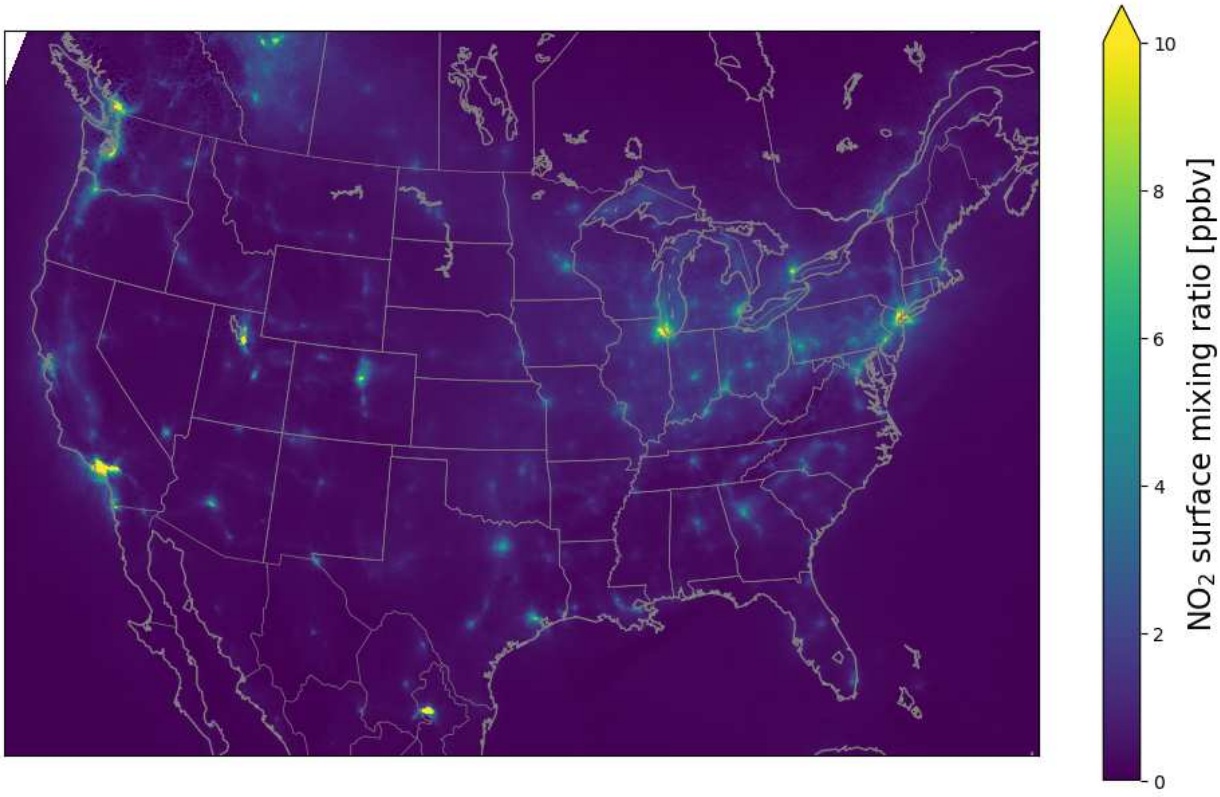

Figure C.2: Cooper et al. (2020) annually averaged NO_2 surface mixing ratios over the continental US.

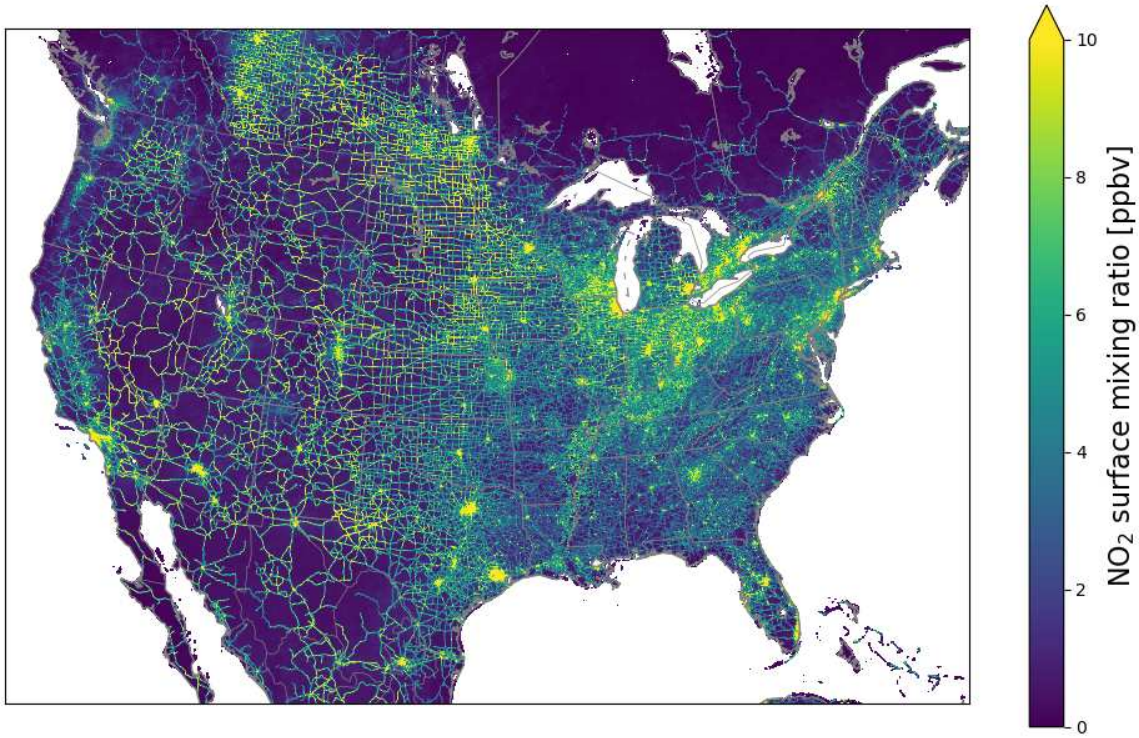

Figure C.3: Anenberg et al. (2021) annually averaged NO_2 surface mixing ratios over the continental US.

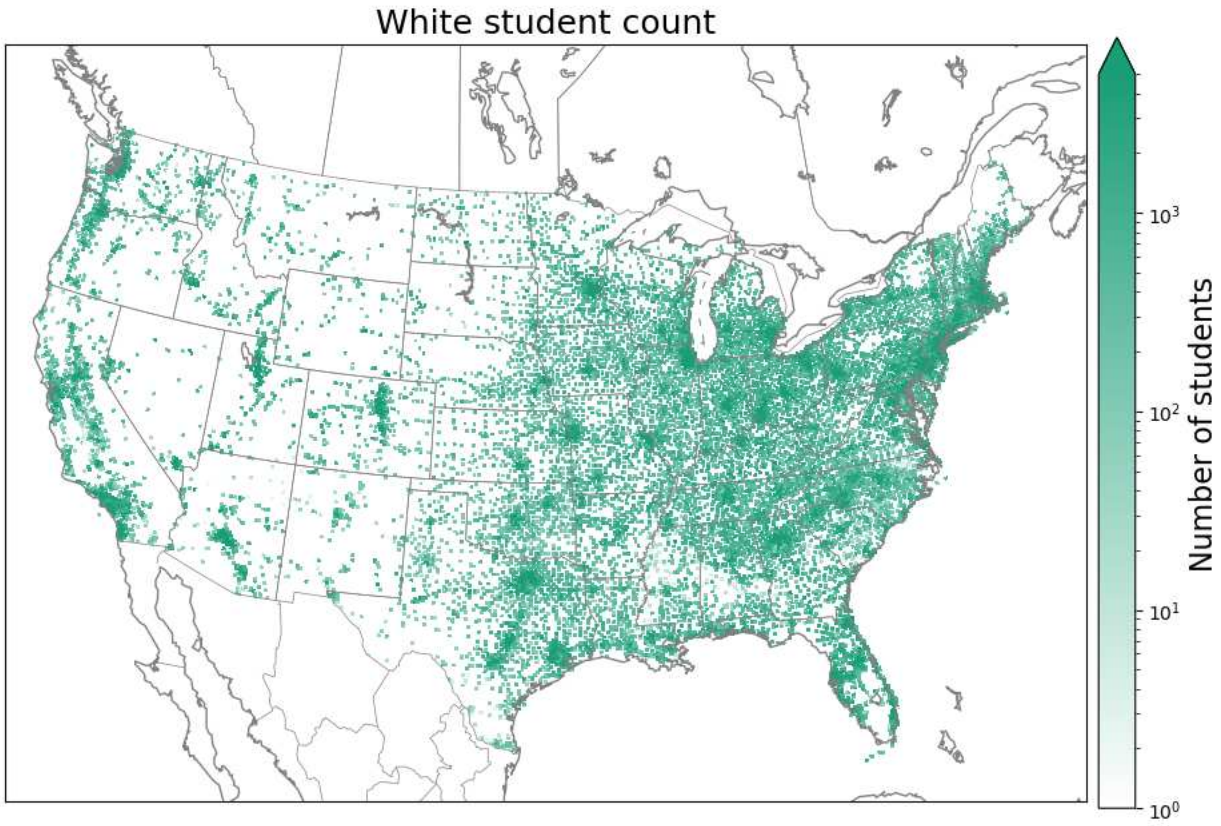

Figure C.4: Gridded counts of students at public schools in the continental US that identified as white based on the National Center for Education Statistics.

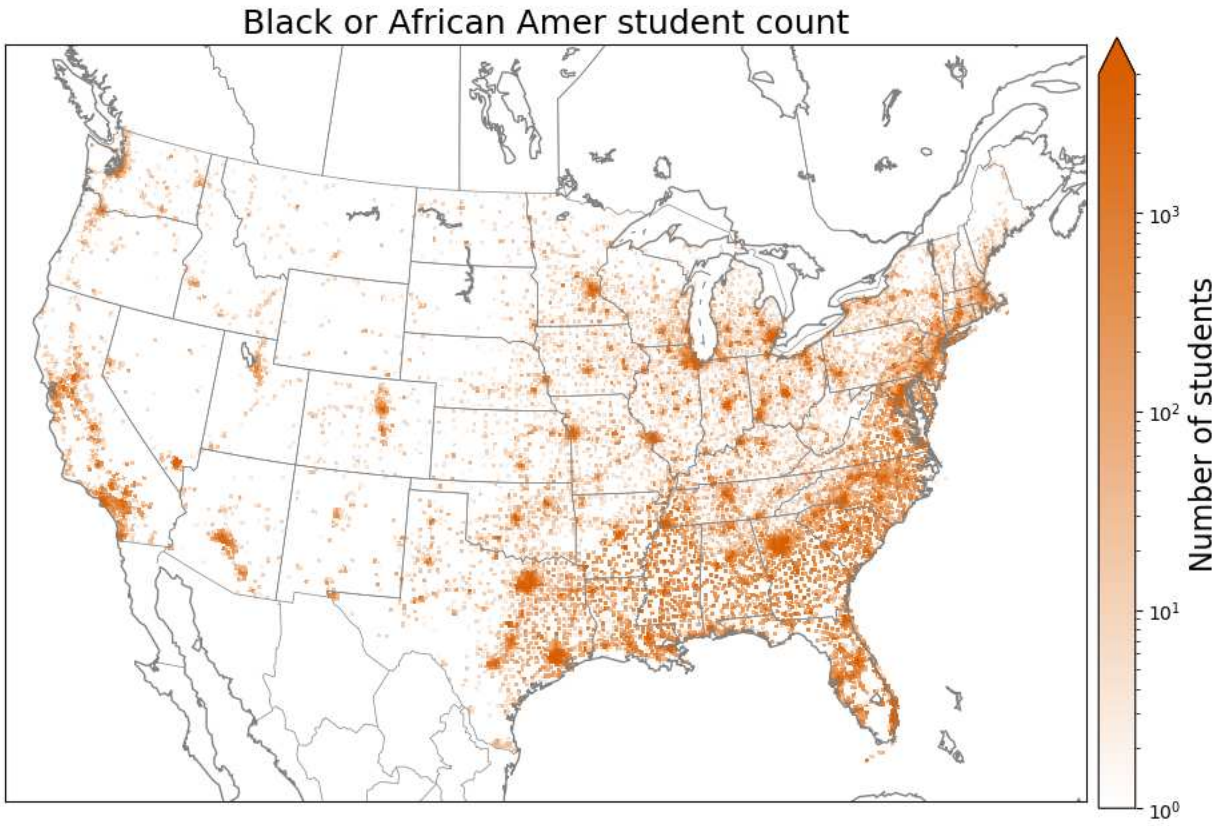

Figure C.5: Gridded counts of students at public schools in the continental US that identified as Black or African American based on the National Center for Education Statistics.

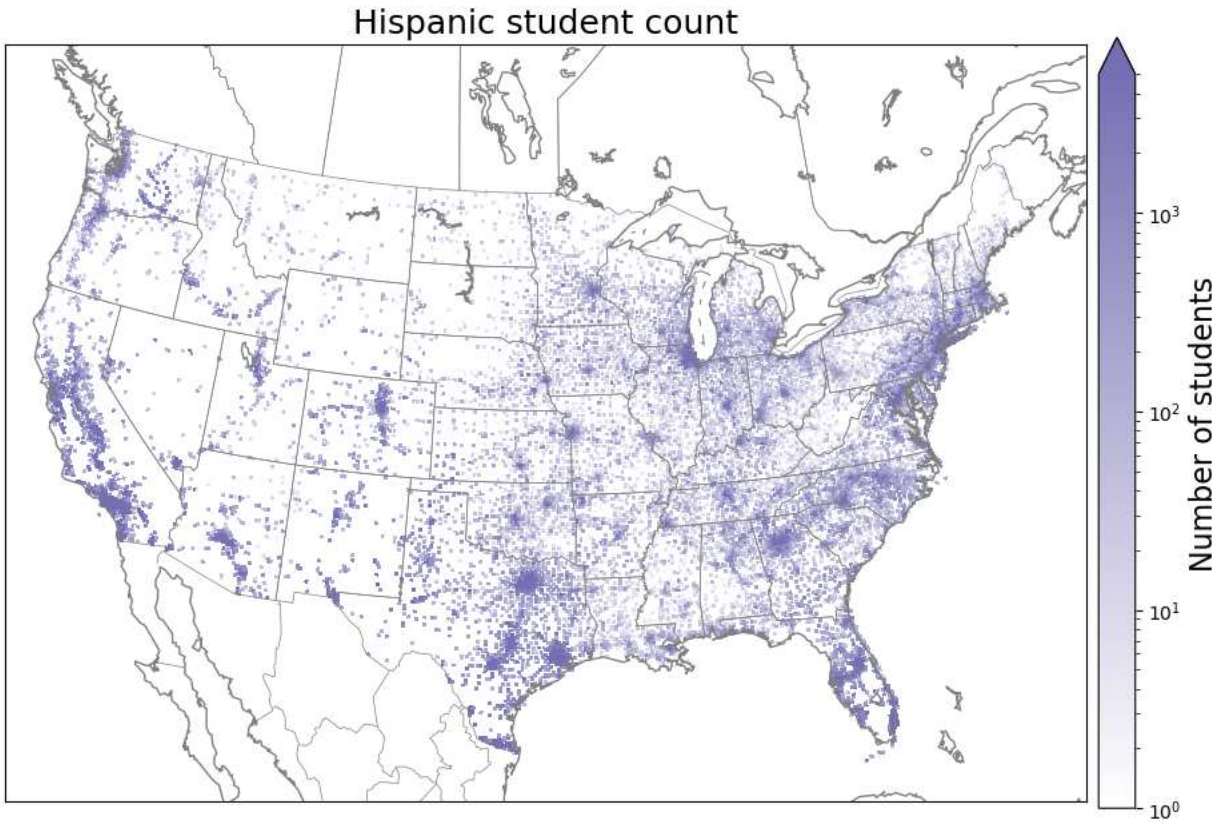

Figure C.6: Gridded counts of students at public schools in the continental US that identified as Hispanic based on the National Center for Education Statistics.

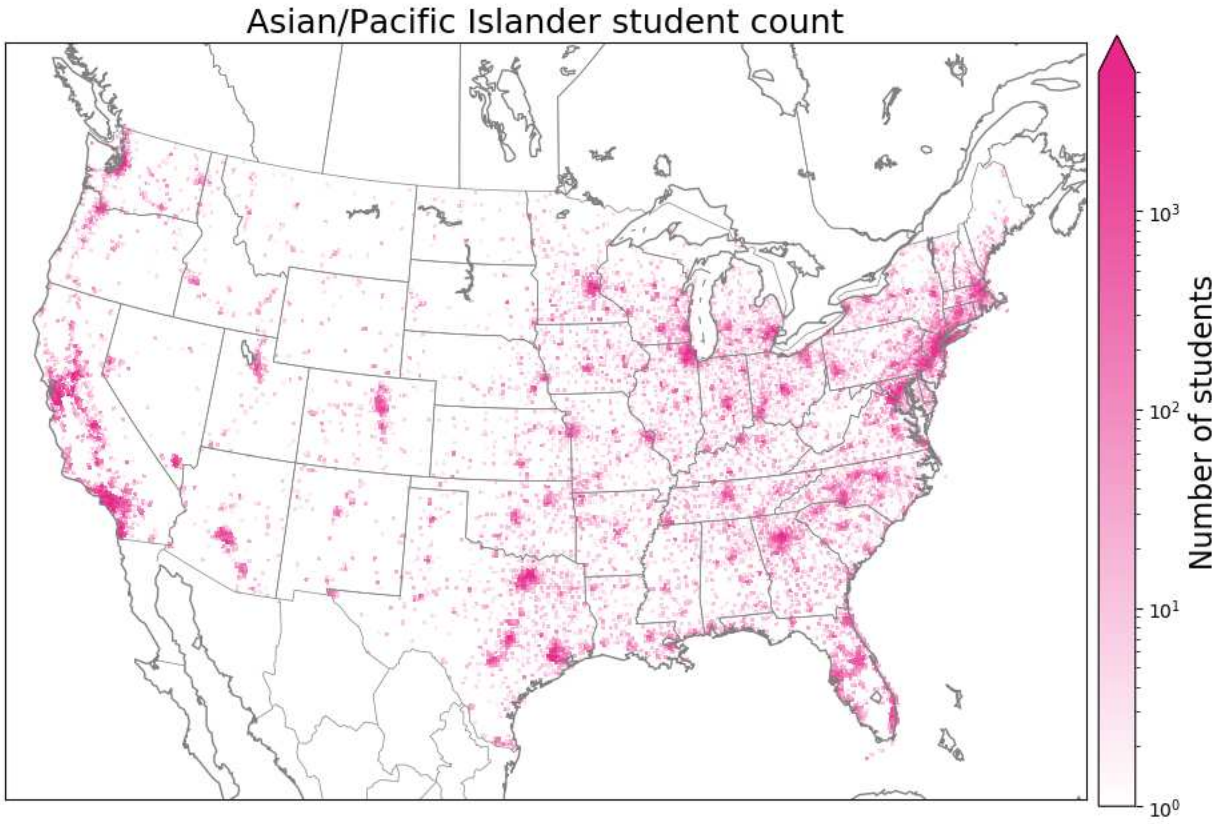

Figure C.7: Gridded counts of students at public schools in the continental US that identified as Asian or Pacific Islander based on the National Center for Education Statistics.

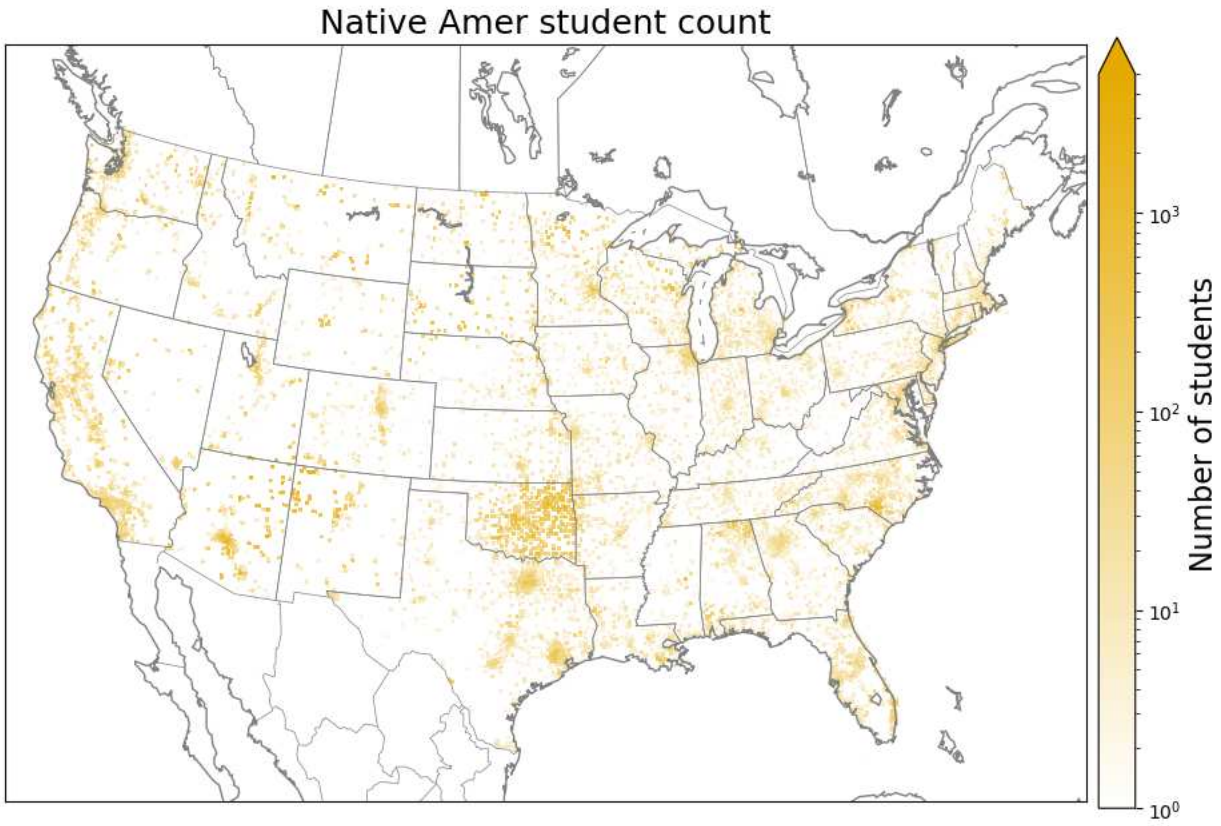

Figure C.8: Gridded counts of students at public schools in the continental US that identified as Native American based on the National Center for Education Statistics.

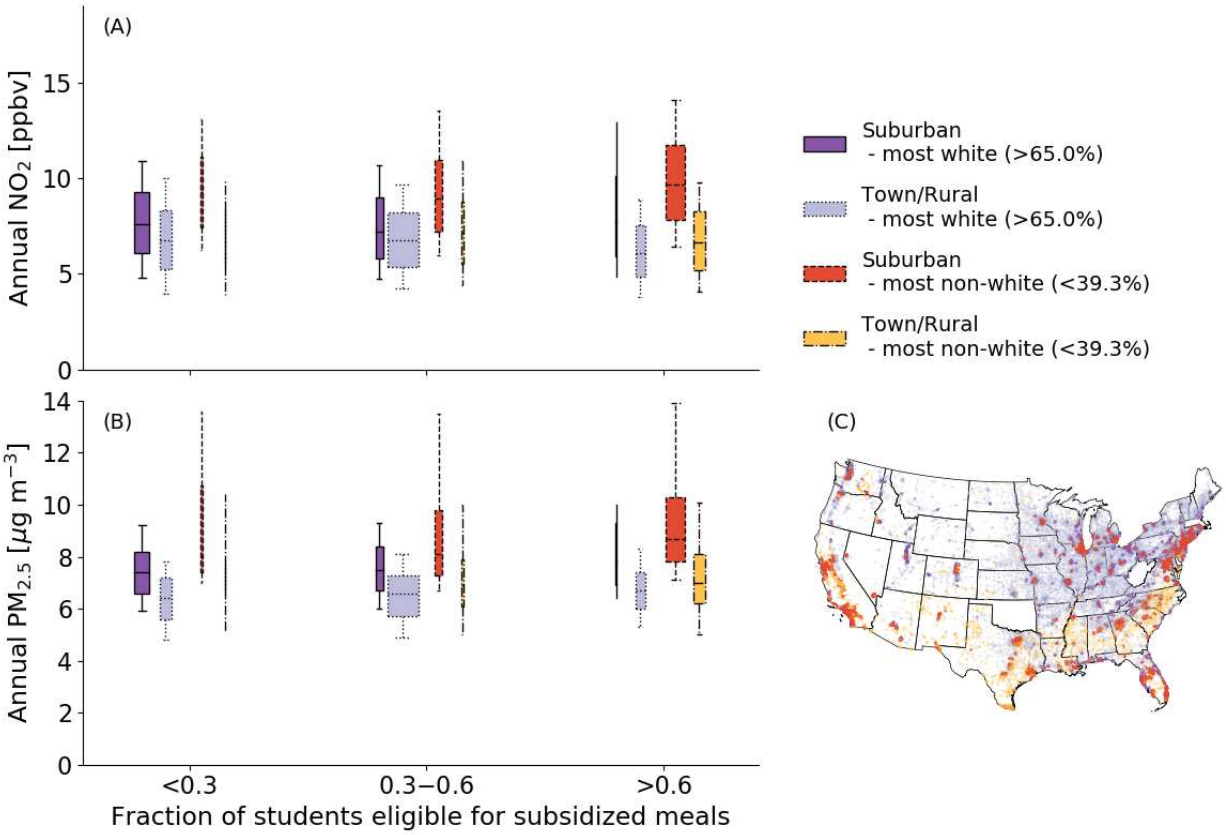

Figure C.9: Boxplots of annually averaged (a) NO_2 and (b) $PM_{2.5}$ surface mixing ratios split into categories of poverty level, which is defined by the fraction of students eligible for free or reduced lunch. A greater fraction of free or reduced lunch indicates a higher level of poverty. Within each poverty-level bin, separate boxplots are shown for schools in suburban and combined town and rural locations for mostly white schools (>65% white students) and mostly non-white schools (<39.3% white students). The thresholds for mostly white and mostly non-white are based on the 60th and 40th percentile of the percentage of white students in all public schools across CONUS. The width of each boxplot is proportional to the number of schools in each distribution. (c) A map of schools that belong in each race and locale category.

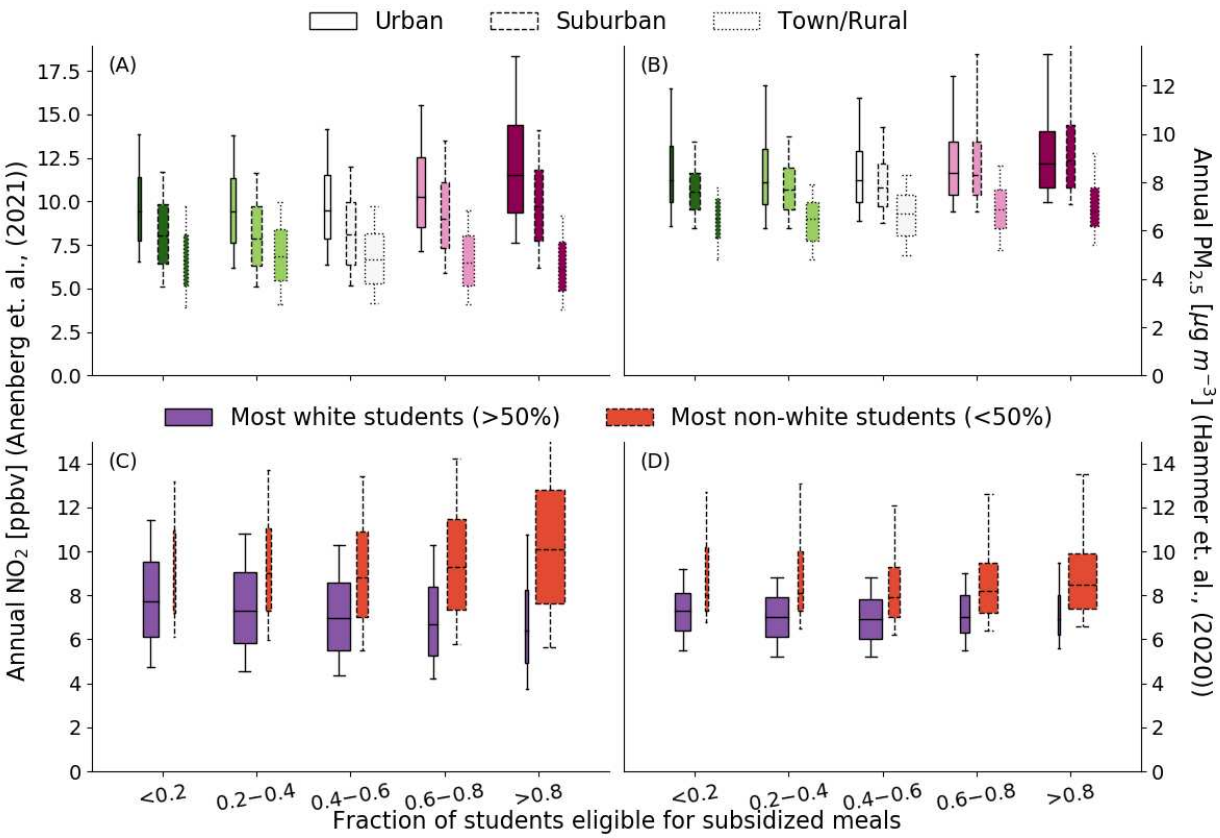

Figure C.10: Boxplots of annually averaged (a) NO_2 and (b) $PM_{2.5}$ split into bins of poverty level, which is defined by the fraction of students eligible for free or reduced lunch. A greater fraction of free or reduced lunch indicates a higher level of poverty. Within each poverty-level bin separate boxplots are shown for schools in urban, suburban, and combined town and rural locations. Boxplots of annually averaged (c) NO_2 and (d) $PM_{2.5}$ at mostly white (>50% white students) and mostly non-white (<50% white students) schools, split into bins of poverty level.

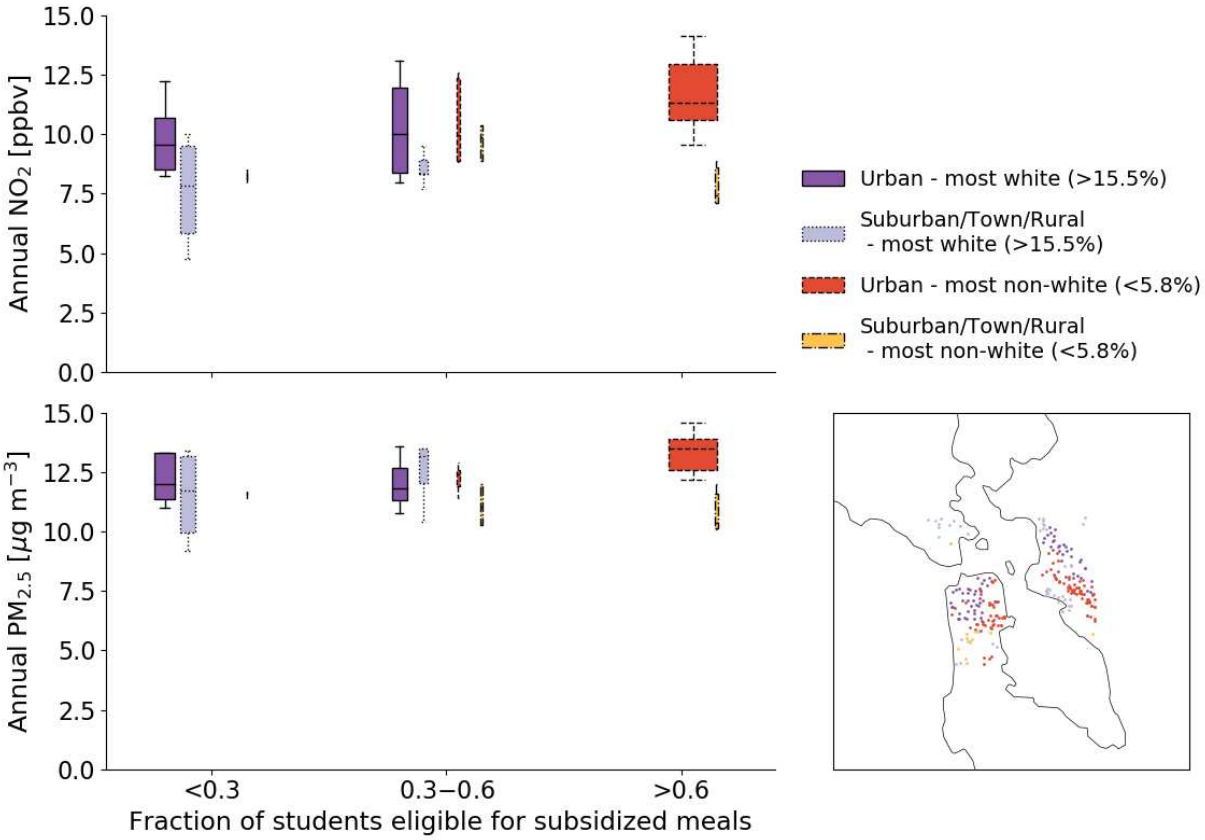

Figure C.11: Boxplots of annually averaged NO_2 at mostly white schools and mostly non-white schools (based on the 60th and 40th percentile of percentage of white students in each school within the region of the map) in the Bay Area of California, United States, split into bins of poverty level.

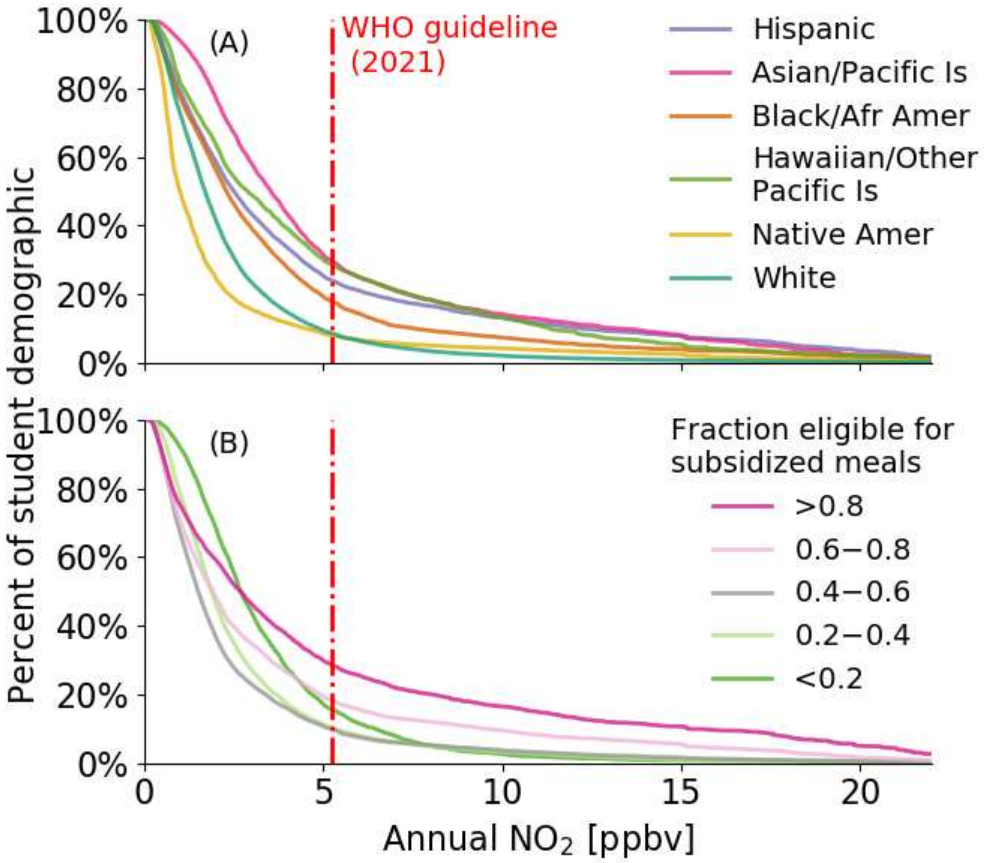

Figure C.12: (a) Complementary cumulative distribution functions (CDF) of annually averaged NO_2 from the Cooper et al. (2020) dataset at schools for students that belong to specific racial/ethnic demographics. (b) Complementary CDFs of annually averaged NO_2 at schools of differing poverty levels, measured by the fraction of students eligible for subsidized meals at each school. Each plot shows the percentage of the students that attend schools where the co-located annually averaged mean of each pollutant is above a given concentration. The current WHO guidelines for annually averaged NO_2 are shown in red dashed lines.

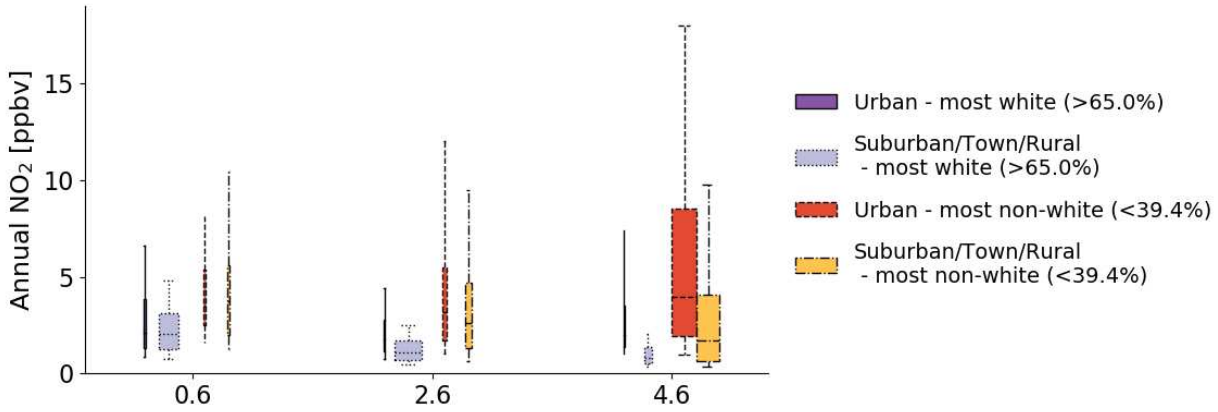

Figure C.13: Boxplots of annually averaged NO_2 surface mixing ratios from the Cooper et al. (2020) dataset split into categories of poverty level, which is defined by the fraction of students eligible for free or reduced lunch. A greater fraction of free or reduced lunch indicates a higher level of poverty. Within each poverty-level bin, separate boxplots are shown for schools in urban and combined suburban, town, and rural locations for mostly white schools (>65% white students) and mostly non-white schools (<39.3% white students). The thresholds for mostly white and mostly non-white are based on the 60th and 40th percentile of the percentage of white students in all public schools across CONUS. The width of each boxplot is proportional to the number of schools in each distribution.

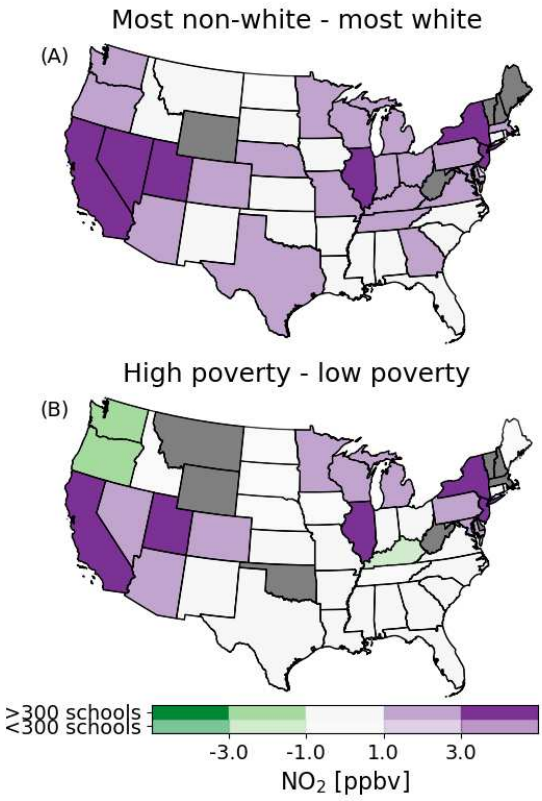

Figure C.14: (a) Difference in mean Cooper et al. (2020) annually averaged NO_2 between mostly non-white schools (fewer than 39.3% white students based on the 40th percentile nationwide) and mostly white schools (greater than 65% white students based on the 60th percentile nationwide). (b) Difference in mean Cooper et al. (2020) NO_2 between high poverty (greater than 75% students eligible for free or reduced lunch) and low poverty schools (fewer than 25% students eligible for free or reduced lunch). The colors are lighter shades if fewer than 300 total schools are used to find the difference between each category in each state. The colors are grey if there is no data for that state or there are fewer than 10 schools in either poverty or racial/ethnic categories.

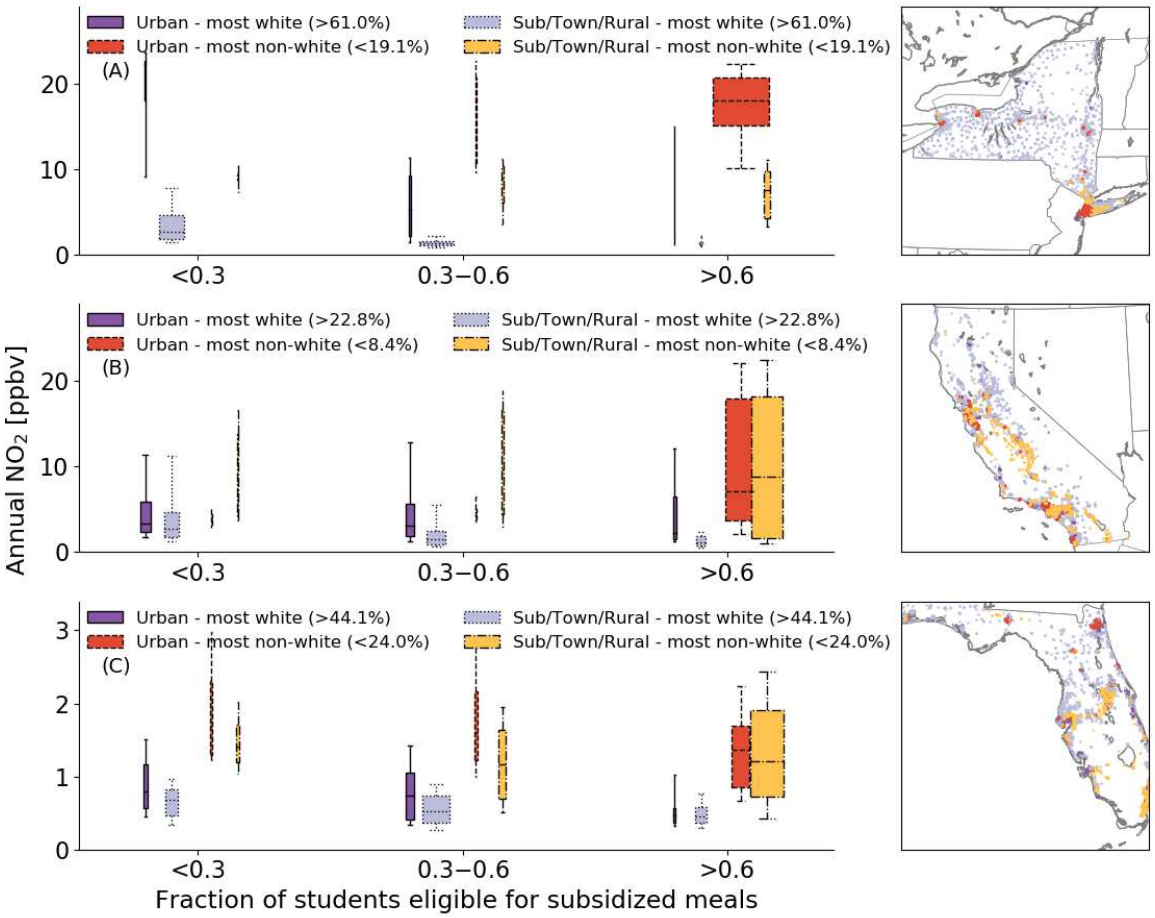

Figure C.15: Boxplots of annually averaged NO_2 from the Cooper et al. (2020) dataset at mostly white schools (based on the 60th percentile of the percentage of white students in the state) and mostly non-white (based on the 40th percentile of the percentage of white students in the state) split into bins of poverty level in New York (a), California (b), and Florida (c).