

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Jianzhao Bi

Date

Assessment of High-Resolution PM_{2.5} Exposures and Changes in PM_{2.5} Cardiorespiratory
Disease Associations Over Time

By

Jianzhao Bi
Doctor of Philosophy

Environmental Health Sciences

Yang Liu, Ph.D.
Advisor

Stefanie E. Sarnat, Sc.D.
Committee Member

Howard H. Chang, Ph.D.
Committee Member

Avani Wildani, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Assessment of High-Resolution PM_{2.5} Exposures and Changes in PM_{2.5} Cardiorespiratory
Disease Associations Over Time

By

Jianzhao Bi
M.Sc., Tsinghua University, 2016
B.Eng., Wuhan University, 2014

Advisor: Yang Liu, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Environmental Health Sciences
2020

Abstract

Assessment of High-Resolution PM_{2.5} Exposures and Changes in PM_{2.5} Cardiorespiratory Disease Associations Over Time

By Jianzhao Bi

Fine particulate matter with aerodynamic diameters less than 2.5 micrometers (PM_{2.5}) is one of the six criteria air pollutants defined by the National Ambient Air Quality Standards. Numerous epidemiological studies have shown the associations between long-term and short-term exposure to PM_{2.5} and increased risks of cardiovascular and respiratory diseases. Understanding the accurate distribution of ground PM_{2.5} concentrations is of growing importance for studying the acute and chronic health effects of PM_{2.5}. Measurements from satellites (aerosol optical depth, AOD) and low-cost air pollution sensors have been increasingly utilized in improving the estimation of ground PM_{2.5} concentrations due to their extensive spatiotemporal coverage. However, an important issue influencing the effective use of satellite AOD retrievals is the large proportion of non-random missing data caused by snow and cloud cover. This study examined the impacts of snow and cloud cover on AOD and PM_{2.5} and made full-coverage PM_{2.5} predictions with the consideration of these impacts. In addition, little has been done to incorporate low-cost sensor PM_{2.5} measurements in large-scale PM_{2.5} exposure modeling. This study conducted spatially varying calibration and developed a down-weighting strategy to optimize the use of low-cost sensor data in PM_{2.5} estimation. Finally, although PM_{2.5} is a complex mixture composed of different chemical components, it is commonly treated as a single pollutant to assess its health effects given that ambient air regulations focus on PM_{2.5} (and not its components). This study examined temporal changes in the risk of emergency department visits for cardiovascular diseases and asthma associated with short-term increases in ambient PM_{2.5} concentrations. By generating the improved PM_{2.5} exposures and examining the contribution of PM_{2.5} components to its overall toxicity, this study sought to broaden the application of satellite and low-cost sensor observations to PM_{2.5} exposure assessment and to provide new information for the health effects of PM_{2.5} mixture.

Assessment of High-Resolution PM_{2.5} Exposures and Changes in PM_{2.5} Cardiorespiratory
Disease Associations Over Time

By

Jianzhao Bi
M.Sc., Tsinghua University, 2016
B.Eng., Wuhan University, 2014

Advisor: Yang Liu, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Environmental Health Sciences
2020

Acknowledgments

Foremost, I would like to thank the wonderful members of my dissertation committee. The completion of this work would not have been possible without their mentorship and guidance. I would especially like to extend my deepest appreciation to Yang Liu, my principal advisor. Throughout the span of my tenure as a student, he exhibited unwavering patience, encouragement, and guidance, which have shaped me into the environmental health researcher that I am today. I would also like to express my gratitude to Stefanie Ebelt, Howard Chang and Vaughn Barry, who lent tremendous support in helping me shape the epidemiological methods and interpretations of my work. I am truly grateful to Avani Wildani, who was instrumental in educating and guiding me in the machine learning methods. I would also like to extend a special thanks to all who made this work possible, including Xuefei Hu, Qingyang Xiao, Jessica Belle, Jennifer Stowell, Xia Meng, Guannan Geng, Fengchao Liang, Bryan Vu, Keyong Huang, Qiannan She, Lin Wang, Rohan D'Souza, Shannon Moss, and our collaborators at the University of Rochester, Clarkson University, Georgia Institute of Technology, and NASA.

I would like to thank the many faculty members in the Rollins School of Public Health who lent educational support and guidance during my sojourn as a doctoral student. I would be remiss not to thank my fellow doctoral students as well, who were always available to listen and lend advice. So, thanks to Brittney Baumert, Danielle Clarkson-Townsend, Frederick Goddard, Tyiesha Johnson, Jiawen Liao and many others for their genuine friendship and assistance.

Finally, I would like to thank and dedicate this dissertation to my loving family. I am tremendously grateful for their encouragement and support throughout this journey. I would especially like to thank my wife, Yan Luo, for her love and all the time spent supporting me when times were especially busy or intense. She is truly my best friend – someone that I can always count on and who is always there with kind and reassuring words. I am also truly grateful for my parents, Mei Li and Jinfeng Bi, and my grandparents, Xiezhao Song and Shuzhen Bi. Thank you for instilling in me a tough work ethic and just enough stubbornness to keep going – even when things were difficult.

Table of Contents

1	Background and Significance	1
2	Description of Aims	6
3	Manuscript I: Impacts of Snow and Cloud Covers on Satellite-Derived PM_{2.5} Levels	8
3.1	Abstract	8
3.2	Introduction	9
3.3	Data and Methods	11
3.3.1	Study Areas	11
3.3.2	PM _{2.5} Measurements	11
3.3.3	MAIAC AOD Data	12
3.3.4	MODIS Cloud and Snow Fractions	13
3.3.5	Meteorological Data	13
3.3.6	Land-Use Variables	14
3.3.7	Data Matching	14
3.3.8	AOD Gap-Filling Model	14
3.3.9	PM _{2.5} Prediction Model	15
3.3.10	Comparison With Another Gap-Filling Method	17
3.4	Results	18
3.4.1	Descriptive Statistics for MAIAC AOD Missingness	18
3.4.2	AOD Gap-Filling by Random Forests	19
3.4.3	PM _{2.5} Prediction by Random Forests	20
3.4.4	The Importance of Gap-Filled AOD With Snow Cover Parameter	24
3.5	Discussion	25
3.6	Conclusions	28
4	Manuscript II: Incorporating Low-Cost Sensor Measurements Into High-Resolution PM_{2.5} Modeling at a Large Spatial Scale	29

4.1	Abstract	29
4.2	Introduction	29
4.3	Data and Methods	32
4.3.1	Study Domain and Modeling Strategy	32
4.3.2	Data	33
4.3.3	PurpleAir PM _{2.5} Calibration and Weighted PM _{2.5} Modeling	35
4.4	Results	41
4.4.1	PurpleAir PM _{2.5} Calibration	41
4.4.2	Weighted PM _{2.5} Modeling	45
4.5	Discussion	48
4.6	Conclusions	52
5	Manuscript III: Temporal Changes in Short-Term Associations Between Cardiorespiratory Emergency Department Visits and PM_{2.5} in Los Ange- les, 2005 to 2016	53
5.1	Abstract	53
5.2	Introduction	54
5.3	Data and Methods	56
5.3.1	Study Population	56
5.3.2	Air Pollution and Weather	56
5.3.3	Emissions and Time Periods	57
5.3.4	Statistical Analysis	58
5.3.5	Sensitivity Analysis	61
5.4	Results	62
5.4.1	PM _{2.5} Concentrations	62
5.4.2	Emergency Department Visits Data	64
5.4.3	Relative Risks Associated With PM _{2.5} Total Mass	65
5.4.4	Relative Risks Associated With PM _{2.5} Components	67
5.4.5	Relative Risks Associated With the Remaining PM _{2.5} Mass	69
5.4.6	Sensitivity Analysis	70

5.5	Discussion	70
5.6	Conclusions	74
6	Conclusions and Future Directions	76
7	Appendices	78
7.1	Manuscript I Supplemental	78
7.2	Manuscript II Supplemental	83
7.2.1	Quality Control for PurpleAir PM _{2.5} Measurements	83
7.2.2	Evaluation of PurpleAir PM _{2.5} Measurements	83
7.2.3	Nonlinearity of PurpleAir Systematic Bias	84
7.2.4	Validation of Scale Factor	84
7.3	Manuscript III Supplemental	97
	References	107

List of Figures

1	New York State Study Domain	12
2	New York State AOD Distributions	19
3	New York State Cross-Validation	20
4	New York State PM _{2.5} Distributions	21
5	PM _{2.5} Accumulation Effect	23
6	California Study Domain	33
7	Time Series and Box Plots	42
8	Generalized Additive Model Relationships	44
9	California PM _{2.5} Distributions	48
10	Los Angeles PM _{2.5} Box Plots	62
11	Cardiovascular Relative Risk	66
12	Asthma Relative Risk	67
13	Cardiovascular Relative Risk (Components)	68
14	Asthma Relative Risk (Components)	69
S1	New York State PM _{2.5} Seasonal Distributions	81
S2	New York State PM _{2.5} Differences	82
S3	Geographically Weighted Regression Spatial Distribution	89
S4	Three Clustered Sub-Domains	90
S5	California Cross-Validation	91
S6	Ferguson Fire	92
S7	California PM _{2.5} Differences	93
S8	PurpleAir Dual-Channel	94
S9	Locally Weighted Polynomial Regression	95
S10	Scale Factor	96
S11	Los Angeles Study Domain	97
S12	Components Box Plots	98
S13	Components Box Plots (%)	99
S14	Cardiovascular Relative Risk (Remainder)	100

S15	Asthma Relative Risk (Remainder)	101
S16	Temperature Splines	102
S17	Minimum Temperature	103
S18	Annual Knots	104
S19	Redefined Periods	105
S20	Ozone Confounding	106

List of Tables

1	New York State AOD Missing Rates	18
2	California Independent Variables	39
3	PurpleAir Weights	46
4	California PM _{2.5} Model Performance	47
5	Los Angeles Summary Statistics	64
6	Los Angeles Rate Ratios	65
S1	New York State AOD Selection Criteria	78
S2	New York State PM _{2.5} Model Performance (Annual)	79
S3	New York State PM _{2.5} Model Performance (AOD)	80
S4	Subsets of AQS Sites	86
S5	Hierarchical Clustering Variables	87
S6	AQS Density in the U.S.	88

1 Background and Significance

Fine particulate matter with aerodynamic diameters less than 2.5 micrometers ($\text{PM}_{2.5}$) is one of the six criteria air pollutants defined by the National Ambient Air Quality Standards (NAAQS). Numerous epidemiological studies have shown the associations between long-term exposure to $\text{PM}_{2.5}$ and increased risks of cardiovascular and respiratory diseases [1]. Growing evidence also shows the adverse effects of short-term exposure to $\text{PM}_{2.5}$ on cardiorespiratory diseases [2, 3]. Biological hypotheses suggested that short-term exposure to $\text{PM}_{2.5}$ may lead to or exacerbate cardiovascular diseases through neurogenic and inflammatory processes [4] and the acceleration of the development of atherosclerosis [5]. The contribution of $\text{PM}_{2.5}$ to oxidative stress and allergic inflammation may also lead to more immediate exacerbation of respiratory diseases, especially asthma [6–9].

Understanding the accurate distribution of ground $\text{PM}_{2.5}$ concentrations is of growing importance for pollution control and public health. While regulatory air quality stations have been employed for the ground-level $\text{PM}_{2.5}$ measurement, they lack spatiotemporal coverage to fully capture population exposures for the research of adverse impacts of $\text{PM}_{2.5}$. Recently, satellite-based aerosol optical depth (AOD) has been increasingly utilized in $\text{PM}_{2.5}$ exposure assessment due to its extensive spatiotemporal coverage. For example, with a 1-km resolution, the Multi-Angle Implementation of Atmospheric Correction (MAIAC) dataset is able to reflect detailed AOD pollution patterns and a more precise link between $\text{PM}_{2.5}$ and microenvironment [10–12]. AOD has a non-linear relationship with ground-level $\text{PM}_{2.5}$, which varies spatially and temporally [13]. Statistical models have been developed to account for this non-linear and variant AOD- $\text{PM}_{2.5}$ relationship. For example, previous studies adopted multi-stage regression models to generate accurate $\text{PM}_{2.5}$ predictions based on satellite AOD data [14–18]. Recently, non-parametric machine learning models, especially Artificial Neural Networks (ANN) [19, 20] and Random Forests (RF) [21, 22], have been increasingly applied to improve the modeling accuracy and precision.

An important issue influencing the effective use of satellite AOD data is the large proportion of non-random missing data [23], the majority of which is caused by cloud cover and

high surface brightness (*e.g.*, snow/ice and desert) [24]. This data gap issue is especially severe in high-latitude areas with large areal extents of snow cover in winter. Furthermore, previous studies showed that AOD/PM_{2.5} levels changed when there was cloud cover due to changes in AOD/PM_{2.5} physical characteristics under different meteorological conditions [25–29]. For example, Belle et al. [26] found that changes in cloud properties such as effective radius, optical depth, and emissivity were associated with changes in PM_{2.5} concentrations and composition. Several strategies have been developed to deal with non-random AOD missingness with meteorological interactions [16, 17, 30, 31]. Kloog et al. [17] proposed a spatial smoothing model based on universal kriging with daily mean PM_{2.5} concentrations and random cell-specific slopes to generate fully covered PM_{2.5} predictions across New England of the United States. Kloog et al. [16] incorporated inverse probability weighting (IPW) in PM_{2.5} regression to address the selection bias caused by AOD missingness and generated PM_{2.5} predictions with high reliability. Xiao et al. [31] proved that cloud-AOD interactions can be partially explained by incorporating cloud features into AOD gap-filling models. In general, when estimating missing AOD data, the significant influence of cloud cover has to be considered in order to minimize estimation biases. Similar to cloud cover, snow cover is also associated with the AOD/PM_{2.5} relationship. Several studies reported that PM_{2.5} levels were higher on snowy days in mountainous terrains because of colder, more humid, and more stagnant atmospheric conditions [32–34]. To our knowledge, however, limited studies have been conducted to explicitly address the AOD missingness due to snow cover. Incorporating snow features into the AOD gap-filling model is beneficial to obtaining more accurate AOD estimations in the regions with extensive snow cover.

Another limitation regarding PM_{2.5} exposure assessment is sparse regulatory PM_{2.5} stations. Due to the high installation and maintenance cost of regulatory stations, even in the United States with the most extensive monitoring network, a large number of less populated counties have still not been covered by regulatory PM_{2.5} monitoring [14, 35, 36]. The sparse monitoring network hinders a comprehensive picture of intra-urban PM_{2.5} pollution details, causing difficulties in accurately reflecting PM_{2.5} exposure in the areas without monitoring stations. With attractive features such as easy installation and low maintenance, low-cost PM_{2.5} sensors have recently emerged as a means to build more extensive networks of mea-

surement sites, though few sensor brands have undergone the rigorous quality examinations that Federal Reference and Equivalent Methods (FRM/FEM) need to pass [37]. For example, a worldwide low-cost air quality monitoring network, PurpleAir, has more than 2,000 sensors in California, compared to ~ 150 regulatory air quality stations in the state. However, in contrast to the well-maintained regulatory air quality stations, low-cost sensors have a significantly lower data quality. Many efforts have been made to calibrate low-cost sensors against collocated high-accurate reference instruments [38–43]. For example, Holstius et al. [41] calibrated self-designed low-cost $\text{PM}_{2.5}$ instruments using FEM measurements in Oakland, California with linear modeling R^2 values of 0.6 at a 1-h scale and 0.72 at a 24-h scale. Kelly et al. [42] conducted ambient tests for low-cost $\text{PM}_{2.5}$ sensors and reported a non-linear response between the sensor measurements and FRM observations when the $\text{PM}_{2.5}$ concentration exceeded $40 \mu\text{g}/\text{m}^3$. By evaluating low-cost air quality monitors in both laboratory and ambient settings, Castell et al. [40] found their performance varied spatially and temporally, depending on $\text{PM}_{2.5}$ composition and weather conditions. Broday and Citi-Sense Project [38] validated the quality of low-cost air quality monitors by a range of criteria, indicating the necessity of frequent calibrations to address quality degradation. Current work has shown the potential of low-cost sensors to shed light on fine spatiotemporal variations of $\text{PM}_{2.5}$ pollution. At a larger spatial scale, however, no study has been devoted to evaluate and calibrate a regional low-cost sensor network, especially for the sensors far away from regulatory stations. The lack of large-scale quality control has limited the use of low-cost sensor data in high-resolution $\text{PM}_{2.5}$ exposure assessment.

Although $\text{PM}_{2.5}$ is a complex mixture composed of different chemical components, it is commonly treated as a single pollutant to assess its health effects given that ambient air regulations focus on $\text{PM}_{2.5}$ (and not its components) and because $\text{PM}_{2.5}$ measurements are generally more readily available. Previous toxicological studies reported that certain $\text{PM}_{2.5}$ components may have higher toxicity than others for certain health outcomes [44, 45]. This toxicological evidence is echoed by the epidemiological hypothesis that the temporal and regional heterogeneity in $\text{PM}_{2.5}$ health effects may be explained by the variation of $\text{PM}_{2.5}$ composition [46–49] as well as the size and solubility [50] of the particles. National-scale epidemiological studies have indicated that the estimated acute effects of $\text{PM}_{2.5}$ may vary

by region [47, 51, 52], and differences in $\text{PM}_{2.5}$ composition across regions may in part explain this variability. However, the regional variation in acute effects of $\text{PM}_{2.5}$ may also be explained by other geographical factors, such as the spatially heterogeneous population susceptibility and exposure misclassification. In detail, it is likely that some regions have higher proportions of population at a greater risk of $\text{PM}_{2.5}$ -related health effects. It is also possible that the estimated $\text{PM}_{2.5}$ exposure levels are more representative to the true pollution levels at certain regions than others, especially when the exposure assessment relies on central air quality stations. In comparison, assessing the temporal variation in acute effects of $\text{PM}_{2.5}$ at the same location would be a more ideal way to mitigate the influence of these unmeasured geographical factors while assessing the influence of $\text{PM}_{2.5}$ composition changes on $\text{PM}_{2.5}$ -related health effects. Currently, very few studies have systematically assessed the temporal variation in the health effects of $\text{PM}_{2.5}$ respect to changes in $\text{PM}_{2.5}$ composition. Recent work evaluated the health effects of short-term exposure to $\text{PM}_{2.5}$ in New York State before, during, after a period between 2005 and 2016 when major emission regulations went into effect and significant emission changes occurred [53–55]. This series of studies found that even with decreasing $\text{PM}_{2.5}$ concentrations, its health effects on cardiovascular [55] and respiratory diseases [53, 54] were elevated with each interquartile-range increase in concentration after the implementation of emission policies, suggesting a possible association between the overall toxicity of $\text{PM}_{2.5}$ mixture and its composition [56]. Changes in the acute response to $\text{PM}_{2.5}$ over time have also been observed in other regions. For example, Abrams et al. [57] found weaker associations between $\text{PM}_{2.5}$ and cardiorespiratory emergency department (ED) visits after the emission control policies were fully realized in Atlanta, Georgia. Outside of the United States, Kim et al. [58] reported temporal variation of estimated risks of $\text{PM}_{2.5}$ on asthma hospitalization in Seoul, South Korea from 2003 to 2011 when the Korean air quality standards had been strengthened. However, to the best of our knowledge, no $\text{PM}_{2.5}$ component-specific epidemiological analysis has been conducted to examine if the changes in the health effects of $\text{PM}_{2.5}$ can be attributed to specific $\text{PM}_{2.5}$ components.

Given the above-mentioned limitations regarding $\text{PM}_{2.5}$ exposure assessment and health analysis, this study aimed to develop an improved $\text{PM}_{2.5}$ exposure dataset with spatiotemporally high resolutions and complete coverage by incorporating satellite, low-cost sensor, and regu-

latory monitoring data, and to further explore variations in overall acute effects of $PM_{2.5}$ on cardiorespiratory disease outcomes attributed to specific $PM_{2.5}$ components. By generating the improved $PM_{2.5}$ exposures and examining the contribution of $PM_{2.5}$ components to its overall toxicity, this study sought to broaden the application of satellite and low-cost sensor observations to $PM_{2.5}$ exposure assessment and to provide new information for the health effects of $PM_{2.5}$ mixture. Although the study regions were limited to the United States due to the accessibility of air quality and epidemiological data, the methodology is expected to be generalizable to other countries with limited regulatory air quality measurements.

2 Description of Aims

PM_{2.5} pollution has contributed to a growing health burden worldwide, causing a variety of mortality and morbidity. In this study, data from novel PM_{2.5} measurement platforms were utilized to better reveal PM_{2.5} pollution features, including satellite AOD data which has been increasingly applied to the large-scale prediction of PM_{2.5} pollution and become an important supplement to ground-based regulatory PM_{2.5} monitoring, and low-cost sensor data which have provided a valuable opportunity to improve the spatiotemporal coverage of PM_{2.5} exposure assessment. By means of the appropriate approaches for satellite AOD gap-filling and low-cost sensor calibration, reliable PM_{2.5} exposures with fine spatiotemporal resolutions and broad coverage were developed. An epidemiological analysis was also conducted by looking further at not only the overall health effects of PM_{2.5} as a single pollutant, but also the influence of PM_{2.5} composition on the health effects of PM_{2.5} as a mixture of pollutants. Specifically,

- (i) Aim 1 built a gap-filling model incorporating satellite snow/cloud fractions to estimate missing satellite AOD data in the regions with extensive snow and cloud cover. A PM_{2.5} prediction model based on the gap-filled AOD was developed to validate the quality of the gap-filling dataset. The importance of the satellite-observed snow fraction in the gap-filling process was evaluated by comparing the full-model predictions to the predictions from a reduced model without the snow parameter. New York State with extensive snow and cloud cover was selected as the study domain to examine the validity of the proposed gap-filling process.
- (ii) Aim 2 developed a spatially-varying calibration model dealing with the measurement errors of low-cost PM_{2.5} sensors. A weighted prediction model with the regulatory PM_{2.5} observations, calibrated sensor PM_{2.5} measurements, gap-filled satellite AOD, and other variables were built to predict PM_{2.5} exposure levels with spatial details. California with dense low-cost PM_{2.5} sensors was selected as the study domain to validate the proposed calibration and exposure prediction methods.
- (iii) Aim 3 assessed the temporal variation in the associations between the emergency de-

partment visits for cardiovascular and asthma outcomes and short-term exposure to $\text{PM}_{2.5}$ in Los Angeles, California during 2005 – 2016, which experienced significant changes in $\text{PM}_{2.5}$ concentrations and composition due to the implementation of emissions reduction policies during this period. The health effects of specific $\text{PM}_{2.5}$ components were further analyzed to explore whether the changes in the health effects of $\text{PM}_{2.5}$ over time could be attributed to specific $\text{PM}_{2.5}$ components.

3 Manuscript I: Impacts of Snow and Cloud Covers on Satellite-Derived PM_{2.5} Levels

Jianzhao Bi, Jessica H. Belle, Yujie Wang, Alexei I. Lyapustin, Avani Wildani, and Yang Liu
Remote Sensing of Environment **2019**, 221, 665–674; doi: 10.1016/j.rse.2018.12.002

3.1 Abstract

Satellite aerosol optical depth (AOD) has been widely employed to evaluate ground fine particle (PM_{2.5}) levels, whereas snow/cloud covers often lead to a large proportion of non-random missing AOD values. As a result, the fully covered and unbiased PM_{2.5} estimates will be hard to generate. Among the current approaches to deal with the data gap issue, few have considered the cloud-AOD relationship and none of them have considered the snow-AOD relationship. This study examined the impacts of snow and cloud covers on AOD and PM_{2.5} and made full-coverage PM_{2.5} predictions with the consideration of these impacts. To estimate missing AOD values, daily gap-filling models with snow/cloud fractions and meteorological covariates were developed using the random forest algorithm. By using these models in New York State, a daily AOD data set with a 1-km resolution was generated with a complete coverage. The “out-of-bag” R² of the gap-filling models averaged 0.93 with an interquartile range from 0.90 to 0.95. Subsequently, a random forest-based PM_{2.5} prediction model with the gap-filled AOD and covariates was built to predict fully covered PM_{2.5} estimates. A ten-fold cross-validation for the prediction model showed a good performance with an R² of 0.82. In the gap-filling models, the snow fraction was of higher significance to the snow season compared with the rest of the year. The prediction models fitted with/without the snow fraction also suggested the discernible changes in PM_{2.5} patterns, further confirming the significance of this parameter. Compared with the methods without considering snow and cloud covers, our PM_{2.5} prediction surfaces showed more spatial details and reflected small-scale terrain-driven PM_{2.5} patterns. The proposed methods can be generalized to the areas with extensive snow/cloud covers and large proportions of missing satellite AOD data for predicting PM_{2.5} levels with high resolutions and complete coverage.

3.2 Introduction

Fine particulate matter ($\text{PM}_{2.5}$) may be inhaled and deposited in alveoli, increasing the risk of cardiorespiratory disease [59–62]. Though air quality stations have been employed to measure $\text{PM}_{2.5}$ and its composition, their spatial coverage is insufficient to cover larger populations to study the impacts of $\text{PM}_{2.5}$ exposure on human health. Recently, satellite aerosol optical depth (AOD) with broad spatiotemporal availability has been widely applied in $\text{PM}_{2.5}$ exposure modeling [16, 22, 63]. With a 1-km resolution, Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD has been able to reveal pollution patterns with more detail and estimate the link between $\text{PM}_{2.5}$ and microenvironment more precisely [12, 14]. High-resolution AOD is critical for reflecting $\text{PM}_{2.5}$ pollution patterns in the regions dominated by local sources, *e.g.*, New York State, U.S., in which residential wood combustion and on-road emissions are the primary $\text{PM}_{2.5}$ sources [64–66].

Though satellite AOD can serve as a surrogate of ground $\text{PM}_{2.5}$, a non-linear relationship is shown between the two, which varies spatially and temporally [13]. To reflect this relationship, various statistical models have been developed [15, 31, 67]. For instance, Kloog et al. [15] employed mixed models to estimate high-resolution $\text{PM}_{2.5}$ with an out-of-sample R^2 (coefficient of determination) of 0.88 in the Northeastern U.S. Xiao et al. [31] built two-stage models to produce 1-km $\text{PM}_{2.5}$ estimates with complete coverage over the Yangtze River Delta (YRD) in China with cross-validation (CV) R^2 values of ~ 0.8 . In these parametric models, however, restrictive assumptions of independence and population distributions are needed, which can be a challenge for a complex mixture, *e.g.*, $\text{PM}_{2.5}$. In contrast, non-parametric models (*e.g.*, machine learning models) can capture the non-linear relationships and interactions between the variables with fewer *a priori* assumptions. Popular machine learning models for $\text{PM}_{2.5}$ prediction consist of Artificial Neural Networks (ANN) [19, 20, 68] and Random Forests (RF) [22, 69]. To be specific, random forests are advantaged in the interpretability of modeling outcomes by their measures of variable importance [70]. Hu et al. [22] developed a random forest model with satellite AOD and covariates, generating 12-km $\text{PM}_{2.5}$ predictions over the Contiguous U.S. with a CV R^2 of 0.80 and a root-mean-square error (RMSE) of $1.78 \mu\text{g}/\text{m}^3$.

The large proportion of non-random missing data is considered as a critical problem associated with the effective use of satellite AOD, which primarily results from cloud cover and high surface brightness from factors such as snow and ice [24, 71, 72]. This non-random missingness has significantly affected data availability and generated biases in the stages of data processing [31]. Several strategies have been developed to cope with this data gap problem. Donkelaar et al. [73] adopted adjusted cloud filtering standards in the Dark Target algorithm and achieved a 21% increase of available AOD during Moscow wildfire events in 2010. Kloog et al. [17], based on universal kriging with daily mean $PM_{2.5}$ levels and random slopes, proposed a spatial smoothing model and generated full-coverage $PM_{2.5}$ predictions across New England within the U.S. Kloog et al. [16] incorporated inverse probability weighting (IPW) with $PM_{2.5}$ regression to reduce the selection bias caused by AOD missingness. They achieved $PM_{2.5}$ predictions with higher reliability. However, it has been found that AOD and $PM_{2.5}$ levels were changed in the presence of cloud cover due to the shifted AOD/ $PM_{2.5}$ physical characteristics under various meteorological conditions [25–28]. For instance, Belle et al. [26] found that variations in cloud properties (e.g., cloud optical depth and emissivity) were correlated with the shifted levels of $PM_{2.5}$ and its components. Accordingly, the gap-filling process without considering cloud interactions may cause estimation biases. Xiao et al. [31] showed that the cloud-AOD interactions can be partially explained by incorporating cloud features in AOD gap-filling processes. By using a multiple imputation model coupled with cloud fractions, they found a higher level of gap-filled AOD on cloudy days in the YRD of China. Similar to cloud, snow cover, with a large areal extent in high-latitudes of Northern Hemisphere [74], has also resulted in a large proportion of missing AOD [12, 23] and the changes of AOD/ $PM_{2.5}$ levels [10, 32–34]. Emili et al. [10] found that due to the contamination of cloud and snow pixels, satellite AOD would be occasionally overestimated. Studies suggested that in mountainous terrains, $PM_{2.5}$ levels were higher on snowy days because of more stagnant atmospheric conditions [32, 34]. To our knowledge, no study has been conducted to deal with the snow-related AOD missingness in $PM_{2.5}$ modeling. Thus, incorporating snow features in AOD gap-filling processes may help to obtain more reliable $PM_{2.5}$ estimates in the regions with extensive snow cover.

In this case study, we incorporated satellite-retrieved snow and cloud fractions in an AOD

gap-filling model to generate fully covered AOD data in New York State where there were extensive snow/cloud covers, especially in winter. Based on the AOD data, we generated high-resolution and fully covered $\text{PM}_{2.5}$ predictions. In the meantime, we examined the significance of the snow fraction in the AOD gap-filling and $\text{PM}_{2.5}$ prediction processing. We also identified the patterns of local $\text{PM}_{2.5}$ pollution in New York State using our high-resolution $\text{PM}_{2.5}$ predictions.

3.3 Data and Methods

3.3.1 Study Areas

New York State and its surrounding areas were selected as our study areas (Figure 1). The surrounding areas served as the buffer to minimize the quality degradation on the edge of the areas and to ensure sufficient ground $\text{PM}_{2.5}$ stations in the study areas. These areas were suitable for examining the performance of our models since 1) the cloud cover and heavy snowfall led to a considerable amount of missing satellite AOD data, particularly in winter [26]; 2) local $\text{PM}_{2.5}$ emission sources (e.g., residential wood combustion and on-road emissions) were dominated in this state so that reliable predictions with a high spatial resolution were needed to reveal the spatial details of $\text{PM}_{2.5}$ pollution [64, 66]. All analyses were based on the 1-km grid of MAIAC AOD, covering 474,392 grid cells in the entire areas.

3.3.2 $\text{PM}_{2.5}$ Measurements

$\text{PM}_{2.5}$ measurements inside the U.S. were provided by the United States Environmental Protection Agency (EPA) (<https://www.epa.gov/>). The measurements inside Canada were provided by the National Air Pollution Surveillance (NAPS) (<http://maps-cartes.ec.gc.ca/>). NAPS has three types of samplers (Dichotomous, Partisol and Speciation) for $\text{PM}_{2.5}$ measurements. Only Dichotomous and Partisol measurements were employed to ensure the data quality. For each ground station from AQS or NAPS, daily mean $\text{PM}_{2.5}$ levels were calculated and matched with AOD and other variables in the MAIAC grid cells. The entire areas covered 137 $\text{PM}_{2.5}$ ground stations (with 127 inside the U.S. and 10 inside Canada;

Figure 1).

3.3.3 MAIAC AOD Data

MAIAC is an advanced AOD data set based on Moderate Resolution Imaging Spectroradiometer (MODIS) Collection 6 Level 1B data. Using a time-series algorithm, the MAIAC data set has increased the accuracy of aerosol detection and optimized the spatial resolution to 1-km [12]. Since MAIAC is a novel algorithm for MODIS AOD, the quality assessment of this product is insufficient. A current validation over South America suggested that $\sim 66\%$ MAIAC AOD retrievals were within the expected error ($\pm(0.05+0.05\times\text{AOD})$) [75]. Terra/Aqua satellites provide two MODIS data sets with different crossing times. In our $\text{PM}_{2.5}$ prediction models, Terra (descending node at 10:30 A.M. local time) and Aqua (ascending node at 1:30 P.M. local time) AOD served as two separate predictors to reflect diurnal changes of AOD.

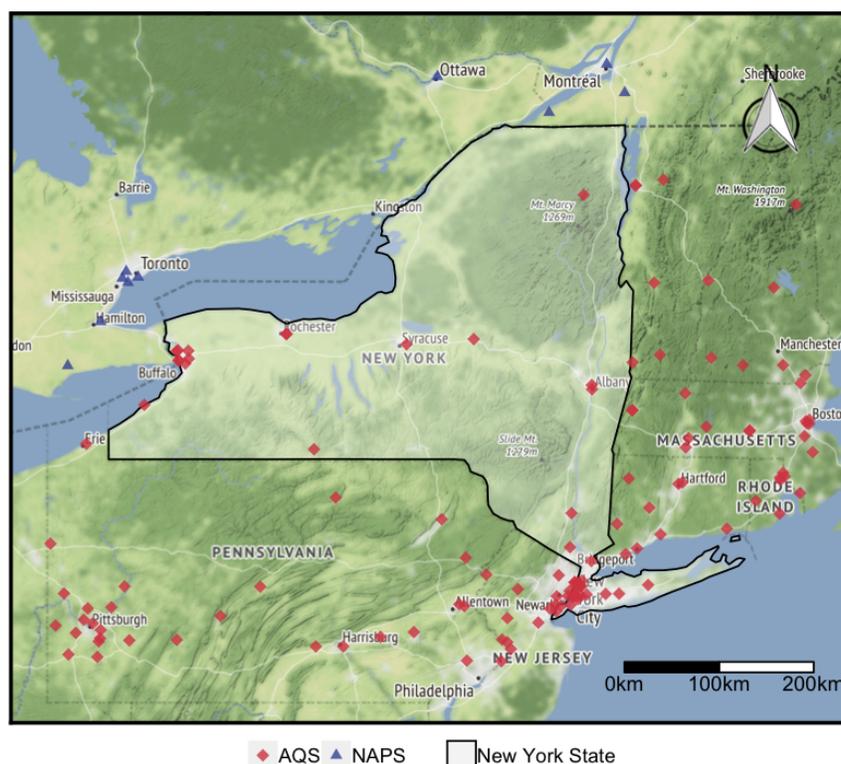


Figure 1: Study areas. Latitude: $[40.1^\circ\text{N}, 45.6^\circ\text{N}]$; Longitude: $[80.5^\circ\text{W}, 71^\circ\text{W}]$. The area outside New York State served as the buffer. Red diamonds are EPA AQS stations. Blue triangles are NAPS stations.

3.3.4 MODIS Cloud and Snow Fractions

Cloud and snow fractions as the percentages of cloud and snow in a grid cell are critical for the quantitative estimation of cloud- and snow-AOD interactions in the gap-filling processing. Cloud fractions were generated from MODIS Level-2 Cloud product (MOD06_L2/MYD06_L2, <https://modis.gsfc.nasa.gov/>). In this product, the cloud fraction has been retrieved from infrared and visible wavelengths along with other physical and radiative cloud properties [76]. A global evaluation by Ackerman et al. [77] suggested that MODIS cloud detection reached an agreement of $\sim 85\%$ with the ground lidar observations. Snow fractions were generated from MODIS Snow Cover product (MOD10/MYD10, <https://modis.gsfc.nasa.gov/>). In this product, the snow fraction has been generated from the Normalized Difference Snow Index (NDSI) [78]. MODIS Snow Cover product was found with a high correlation ($r = 0.9$) with ground-truth observations [79]. Terra and Aqua data sets were employed respectively and matched with corresponding AOD data sets. Only daytime cloud/snow observations were adopted.

3.3.5 Meteorological Data

Meteorological data were obtained from the North American Regional Reanalysis (NARR) (<http://www.emc.ncep.noaa.gov/>) [80] and the North American Land Data Assimilation System (NLDAS) (<http://ldas.gsfc.nasa.gov/nldas>) [81]. NARR had a resolution of 0.25° and a 3-h temporal interval. NLDAS had a resolution of 0.125° and a 1-h temporal interval. Since NLDAS has higher spatiotemporal resolutions, the parameters from this data set were used first. The parameters not provided by NLDAS were extracted from NARR. The meteorological parameters from NLDAS consisted of air temperature, humidity, surface pressure, precipitation, wind speed, potential evaporation, downward shortwave radiation as well as convective available potential energy (CAPE). The parameters from NARR covered planetary boundary layer height (HPBL) and visibility. To match the observation time of MAIAC AOD data, we calculated the daily values of meteorological parameters by averaging the simulations from 9 A.M. to 3 P.M. (local time, *i.e.*, GMT 1400 – 2000 for NLDAS and GMT 1500, 1800, 2100 for NARR).

3.3.6 Land-Use Variables

The land-use parameters covered (1) the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation at a 1 arc-second (~ 30 m) resolution (<https://asterweb.jpl.nasa.gov/gdem.asp>); (2) LandScan ambient population in 2015 at a 900-m resolution (<https://web.ornl.gov/sci/landscan/>); (3) Normalized Difference Vegetation Index (NDVI) from MODIS vegetation indices at a 500-m resolution (MOD13/MYD13, <https://modis.gsfc.nasa.gov/>); (4) distances to highways and major roads computed from ESRI StreetMap (Environmental System Research Institute, Inc., Redland, California).

3.3.7 Data Matching

All data sets with various spatial resolutions were matched and fitted into the MAIAC AOD 1-km grid. For the data sets with spatial resolutions coarser than 1 km (e.g., meteorological fields, cloud and snow fractions, and NDVI), the inverse distance weighting (IDW) interpolation was employed [82]. For the data sets with finer resolutions (e.g., ASTER GDEM), the down-sampling was conducted, *i.e.*, a process of averaging neighboring pixels into one value. For the LandScan population, each data point was assigned to its nearest MAIAC grid cell. Subsequently, the population in that grid cell was summarized.

3.3.8 AOD Gap-Filling Model

Random Forest (RF) is an “ensemble learning” method creating numerous decision trees and aggregating the regressing results from these trees [83, 84]. It adopts two types of bootstrap aggregating strategies (*i.e.*, a bootstrap training sample and a group of randomly selected independent variables) for each decision tree. These strategies allow the random forest to be a robust algorithm against overfitting [83]. A random forest model has two major parameters, *i.e.*, the number of decision trees in the forest (n_{tree}) and the number of independent variables in each tree (m_{try}). It also provides variable importance measures that inform variable weights and contribute to the interpretation of the model [84]. Our AOD

gap-filling model was based on the random forest algorithm, which is expressed as

$$\text{AOD}_{st} = f(\text{cloud fraction}_{st}, \text{snow fraction}_{st}, \text{air temperature}_{st}, \text{specific humidity}_{st}, \text{relative humidity}_{st}, \text{previous day precipitation}_{st}, \text{elevation}_s, X_s, Y_s) \quad (1)$$

where s denotes the location of a grid cell, and t is the time of an observation. The dependent variable was the satellite AOD. Independent variables covered cloud and snow fractions, meteorological parameters (*i.e.*, air temperature, specific humidity, relative humidity and precipitation on the previous day), elevation, and spatial coordinates. The spatial coordinates were the real distances in kilometers to the central point of the study areas. By comparing the results with different settings, n_{tree} and m_{try} were set as 200 and 3 respectively to balance the prediction accuracy and computational efficiency.

The AOD gap-filling model (Equation (1)) was fitted daily for both Terra and Aqua AOD data sets. Equation (1) without the snow fraction (a.k.a. cloud-only gap-filling model) was also fitted to examine the impact of this variable on the gap-filled AOD. Since MAIAC AOD had a large proportion of missing data, we employed three rolling-day samples for the middle day’s model. Out-of-bag (OOB) R^2 and RMSE were employed to assess the modeling performance. OOB R^2 was calculated from the predictions not in the bootstrap sample (a.k.a. “out-of-bag” sample) [83]. Since the mechanism of OOB is similar to cross-validation (CV), OOB R^2 is nearly equal to CV R^2 under large sample size [70]. RMSE was calculated by aggregating the errors of OOB predictions. Model-estimated “permutation accuracy importance” [85] was used to suggest the variable importance. This importance measure was estimated in line with the fall of prediction accuracy after randomly permuting the OOB sample of the targeting variable [84].

3.3.9 PM_{2.5} Prediction Model

The PM_{2.5} prediction model also followed the random forest algorithm (Equation (2)). The dependent variable was the PM_{2.5} measurements from AQS and NAPS. Independent variables covered gap-filled Terra and Aqua AOD, meteorological parameters (air temperature,

dew-point temperature, surface pressure, specific humidity, wind speed, visibility, planetary boundary layer height, potential evaporation, downward shortwave radiation, and CAPE), land-use parameters (population, distances to highways and major roads, elevation, and NDVI), and dummy variables for months and Julian days. Moreover, a convolutional layer of $PM_{2.5}$ levels was applied, representing the weighted averages of nearby $PM_{2.5}$ levels. This parameter helps to fully exploit the spatial autocorrelation of $PM_{2.5}$ and can significantly increase the accuracy of $PM_{2.5}$ prediction [19, 22]. By comparing the results with different settings, n_{tree} and m_{try} were set as 500 and 7 respectively to achieve the best prediction accuracy.

$$\begin{aligned}
 PM_{2.5st} = f(& \text{Terra AOD}_{st}, \text{Aqua AOD}_{st}, \text{air temperature}_{st}, \text{dew point temperature}_{st}, \\
 & \text{surface pressure}_{st}, \text{specific humidity}_{st}, \text{wind speed}_{st}, \text{visibility}_{st}, \text{HPBL}_{st}, \\
 & \text{potential evaporation}_{st}, \text{downward shortwave radiation}_{st}, \text{CAPE}_{st}, \text{population}_s, \\
 & \text{NDVI}_{st}, \text{highway distance}_s, \text{major distance}_s, \text{elevation}_s, \text{convolutional layer}_{st}, \\
 & \text{month}_t, \text{day}_t)
 \end{aligned} \tag{2}$$

where s denotes the location of a grid cell and t denotes the time of an observation. The variable selection strategy used in Hu et al. [22] was adopted here, in which the variables with low importance values were discarded (e.g., precipitation and relative humidity in this analysis). 10-fold CVs was employed to assess the model performance, including overall, spatial, and temporal CVs [31]. The spatial CV generated validation samples in accordance with the locations of the $PM_{2.5}$ measurements. The temporal CV generated validation samples in line with the Julian days of the measurements. Besides the original model, Equation (2) without both Terra and Aqua gap-filled AOD parameters (a.k.a. no-AOD prediction model) was fitted to examine the impact of the gap-filled AOD on the $PM_{2.5}$ predictions. Equation (2) with AOD gap-filled by cloud fractions (a.k.a. cloud-only prediction model) was also fitted to verify the importance of the snow parameter in the $PM_{2.5}$ prediction.

3.3.10 Comparison With Another Gap-Filling Method

To assess our gap-filling model, our full-coverage PM_{2.5} predictions were compared to the gap-filled PM_{2.5} generated by a previous gap-filling method proposed by Kloog et al. [17] and Just et al. [86] (Equation (3)).

$$\sqrt{\text{PredPM}_{st}} = \alpha_0 + \alpha_1 \times \sqrt{\text{MeanPM}_t} + s(X, Y)_s + \epsilon_{st} \quad (3)$$

where s denotes the location of a grid cell, and t is the time of an observation. The dependent variable $\sqrt{\text{PredPM}_{st}}$ denotes the existing PM_{2.5} predictions at location s at time t . $\sqrt{\text{MeanPM}_t}$ denotes an average PM_{2.5} measurement on day t . $s(X, Y)_s$ is a spline surface of the coordinates of the grid cells. The reason for applying square root was to ensure the positive values of PM_{2.5} [86]. We had also tried calculating $\sqrt{\text{MeanPM}_t}$ within a 100-km buffer [15]. Yet the PM_{2.5} distribution showed significant artificial patterns due to the sparse monitoring network so that the buffer was not adopted. The gap-filling model was fitted monthly to better reveal the temporal variations of PM_{2.5} [15]. A 10-fold CV was employed to examine the modeling performance.

3.4 Results

3.4.1 Descriptive Statistics for MAIAC AOD Missingness

Table 1: Average daily missing rates of MAIAC AOD caused by cloud and snow covers in NYS in 2015. The first 15 weeks of 2015 are defined as the “snow season”. The $\sim 10\%$ gap between the overall missing rate and the total rate caused by cloud and snow resulted from the MAIAC pixels outside the sensor scanning range (*i.e.*, no measurement). The contribution of water/ice to the missing AOD was negligible in NYS.

AOD	Missing Type	Mean	Median	25 th – 75 th Quantiles
Aqua	Overall	90.27%	96.54%	87.62% – 99.84%
	Cloud-related	75.58%	79.47%	64.74% – 91.48%
	Snow-related	6.14%	0.00%	0% – 2.48%
	Snow-related (first 15 weeks)	21.15%	14.03%	5.23% – 31.18%
Terra	Overall	89.50%	96.54%	84.88% – 99.75%
	Cloud-related	76.66%	81.32%	66.29% – 92.12%
	Snow-related	5.55%	0.00%	0% – 2.88%
	Snow-related (first 15 weeks)	19.03%	10.78%	5.00% – 29.84%

The Quality Assessment (QA) flags of MAIAC AOD were employed to infer the rates of missing AOD caused by cloud cover, snow cover and water/ice (Table S1). The average daily missing rates associated with cloud and snow are listed in Table 1 (the contribution of water/ice-related missing AOD was negligible in NYS). Since the missing AOD caused by snow cover was primarily in the first 15 weeks (105 days) of 2015, the average missing rates in this period are separately summarized in the table. This period is also referred to as the “snow season” hereinafter. The overall AOD missing rate was $\sim 90\%$ in 2015, and the cloud-related missing AOD was $\sim 76\%$ of the total data. Though the snow-related missingness only took up $\sim 6\%$ in this year, the percentage increased to $\sim 20\%$ in the snow season. There was a $\sim 10\%$ gap between the overall missing rate and the total rate caused by cloud and snow. This gap resulted from the MAIAC pixels located in the areas outside the sensor scanning range (*i.e.*, no measurement). In general, though the missing AOD was primarily caused by cloud cover, snow-related AOD missingness remained a severe issue in the snow season.

3.4.2 AOD Gap-Filling by Random Forests

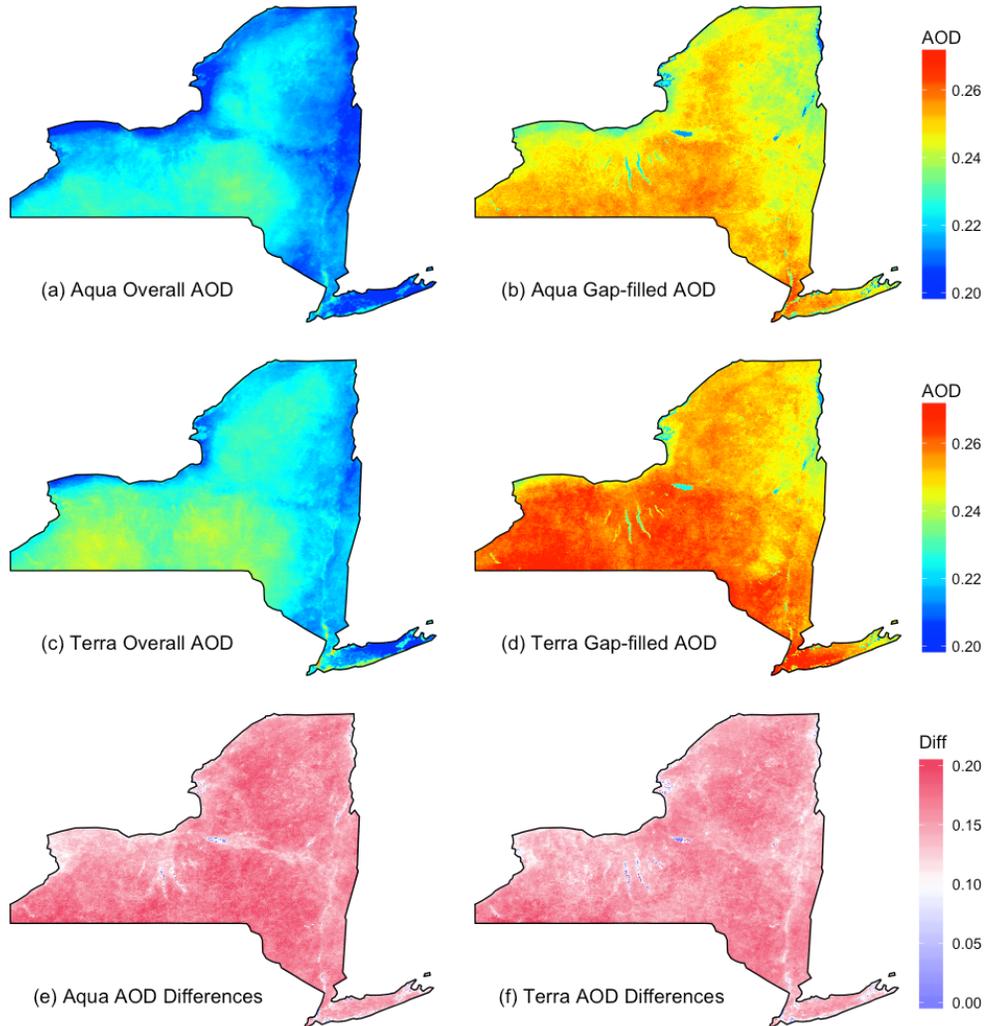


Figure 2: AOD spatial distributions in 2015: (a) annual distribution of Aqua AOD (original and gap-filled AOD); (b) annual distribution of gap-filled Aqua AOD; (c) annual distribution of Terra AOD (original and gap-filled AOD); (d) annual distribution of gap-filled Terra AOD; (e) differences between gap-filled and original Aqua AOD (gap-filled minus original AOD); (f) differences between gap-filled and original Terra AOD (gap-filled minus original AOD).

By Equation (1), full-coverage AOD was generated for both Terra and Aqua data sets. Daily RF models had a mean OOB R^2 of 0.93 (for both Terra and Aqua data sets) with interquartile ranges (IQR) from 0.91 to 0.95 for Terra and 0.90 to 0.95 for Aqua. The annual AOD distributions and the spatial differences between gap-filled and original AOD are shown in Figure 2. The overall and gap-filled AOD showed the similar spatial patterns. The peak

AOD levels appeared in the Allegheny Plateau of Southern NYS. Some populated areas in NYS (e.g., New York City) also had high-level AOD. For Terra, original and gap-filled AOD had the annual means of 0.11 (IQR: [0.10, 0.12]) and 0.26 (IQR: [0.25, 0.26]), respectively. For Aqua, original and gap-filled AOD had the annual means of 0.09 (IQR: [0.08, 0.10]) and 0.25 (IQR: [0.24, 0.25]), respectively. As Figure 2(e) and (f) suggest, the gap-filled AOD was higher than the original AOD throughout the state. This outcome is consistent with that of Belle et al. [26] and Xiao et al. [31] who found clear evidences that the hygroscopic growth of aerosol droplets under cloudy and humid weather resulted in higher levels of AOD estimates.

3.4.3 PM_{2.5} Prediction by Random Forests

3.4.3.1 Modeling Performance

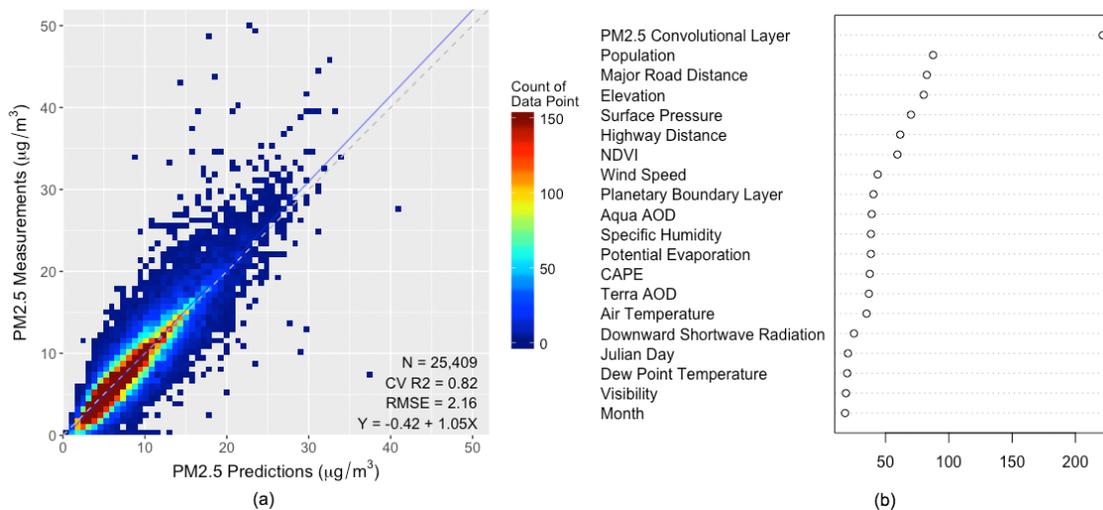


Figure 3: Random forest modeling performance: (a) 10-fold cross-validation (CV) scatters with an R^2 of 0.82 and an RMSE of $2.16 \mu\text{g}/\text{m}^3$; (b) variable importance ranking. The PM_{2.5} convolutional layer had the highest value and five of the top-seven important variables were land-use terms.

By Equation (2), 1-km PM_{2.5} predictions with full coverage were generated. The prediction model showed an overall CV R^2 of 0.82 (Figure 3(a)) with spatial and temporal CV R^2 values of 0.74 and 0.81, respectively. The model had an RMSE of $2.16 \mu\text{g}/\text{m}^3$, suggesting a good prediction accuracy. Figure 3(b) shows the ranking of the variable importance. The PM_{2.5}

convolutional layer had the highest importance value, which is consistent with that of Hu et al. [22] who confirmed an improvement of model accuracy by the $PM_{2.5}$ convolutional layer. Five of the top-seven important variables were land-use terms (e.g., population, distances to major roads, elevation, distances to highways, and NDVI). This outcome could suggest that the $PM_{2.5}$ sources in New York State were primarily local sources, and the transportation of regional pollution was relatively weak, which is consistent with the $PM_{2.5}$ emission inventory of New York State in 2014. Though the gap-filled AOD parameters did not have high importance values, they still contributed to a changed spatial pattern of $PM_{2.5}$ estimates. We also performed the $PM_{2.5}$ prediction for 2002 – 2012, finding that the modeling performance was stable in terms of CV R^2 and variable importance ranking (Table S2).

3.4.3.2 $PM_{2.5}$ Predictions

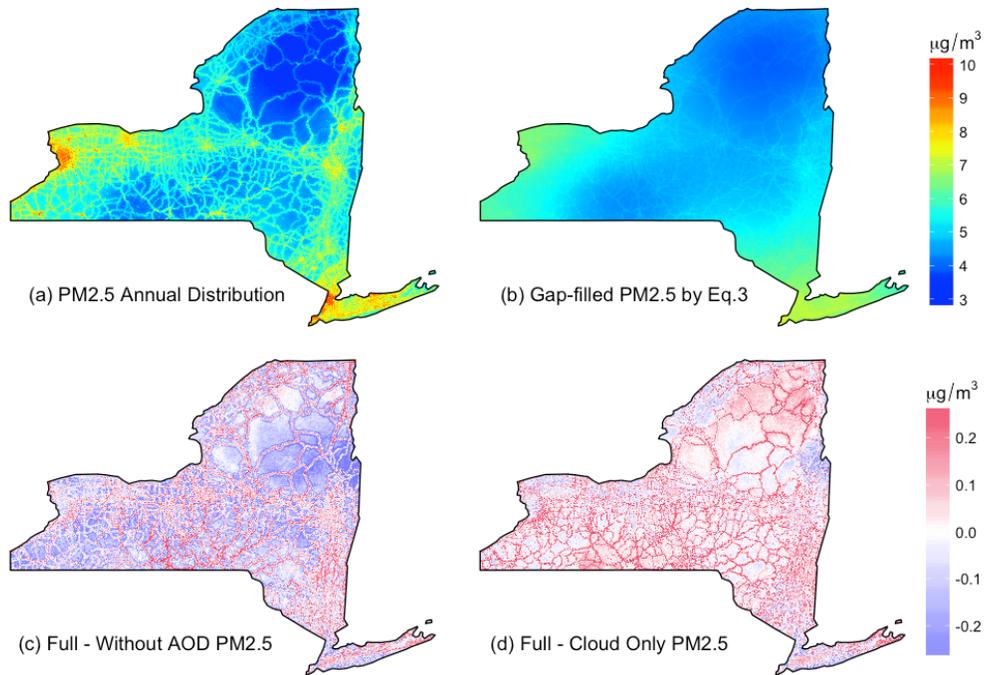


Figure 4: $PM_{2.5}$ spatial distributions in 2015: (a) annual distribution of $PM_{2.5}$ with a 1-km resolution by Equation (2); (b) gap-filled $PM_{2.5}$ by Equation (3); (c) differences between full-model and no-AOD $PM_{2.5}$ in the snow season (full-model minus no-AOD $PM_{2.5}$); (d) differences between full-model and cloud-only $PM_{2.5}$ in the snow season (full-model minus cloud-only $PM_{2.5}$).

Figure 4(a) shows the annual $\text{PM}_{2.5}$ distribution in NYS in 2015 (the $\text{PM}_{2.5}$ distributions in different seasons shown in Figure S1). The $\text{PM}_{2.5}$ levels had a mean of $5.30 \mu\text{g}/\text{m}^3$ with an IQR from 4.43 to $6.08 \mu\text{g}/\text{m}^3$. The distribution showed clear patterns of higher $\text{PM}_{2.5}$ along with the roads and in the populated areas. These patterns are consistent with the $\text{PM}_{2.5}$ emission inventory of NYS, suggesting that the $\text{PM}_{2.5}$ sources were primarily residential wood combustion and on-road emissions. The highest $\text{PM}_{2.5}$ levels appeared in the large cities, e.g., New York City, Long Island, and Buffalo. Other major cities, e.g., Albany, Rochester, Yonkers, and Syracuse also had relatively high $\text{PM}_{2.5}$ levels. In contrast, the Adirondack Mountains in the Northeastern NYS with the lowest population density showed the lowest $\text{PM}_{2.5}$ levels.

Since New York State lies upon the portion of Northeast Appalachians, most of its areas, in particular Upstate New York, are dominated by the mountainous terrain. Our high-resolution $\text{PM}_{2.5}$ predictions showed obvious evidence of the correlation between elevation and $\text{PM}_{2.5}$ in NYS. This correlation can be partially reflected by the variable importance since the elevation was the 4th important variable in the prediction model. Besides, we found obvious patterns of $\text{PM}_{2.5}$ accumulation in the valleys of the Allegheny Plateau in winter. An area in the Allegheny Plateau was selected to highlight the valley accumulation of $\text{PM}_{2.5}$. Figure 5 shows the $\text{PM}_{2.5}$ distribution with contours (140-m intervals) in this area in winter (January, February, and December in 2015). Elevated $\text{PM}_{2.5}$ levels in the valleys were clearly shown. In these valleys, $\text{PM}_{2.5}$ levels were $1 - 2 \mu\text{g}/\text{m}^3$ higher than those in the surroundings. These results suggested that our 1-km $\text{PM}_{2.5}$ predictions could reflect small-scale $\text{PM}_{2.5}$ features driven by local geographical factors, which would be smeared with a coarser resolution.

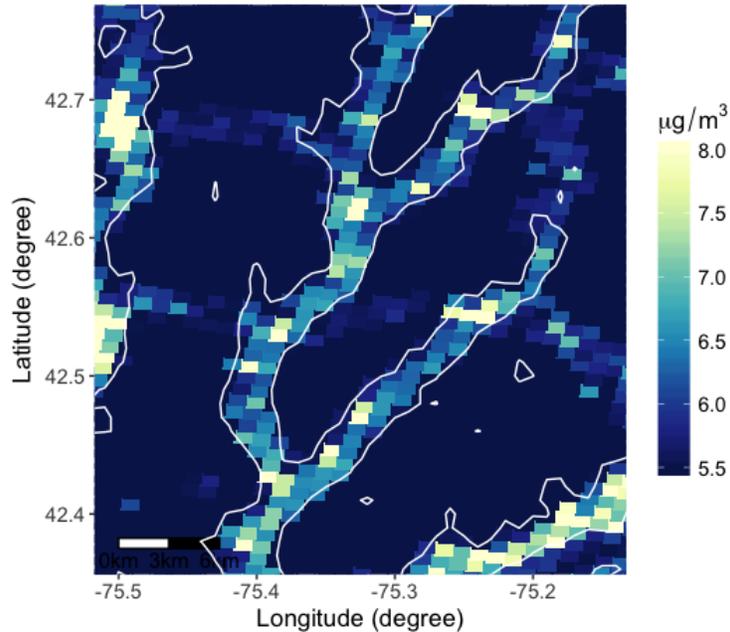


Figure 5: $PM_{2.5}$ accumulation effect in the valleys of Upstate New York in winter. The areas are on the border of Chenango County and Otsego County (Latitude: $[42.375^\circ N, 42.75^\circ N]$; Longitude: $[75.5^\circ W, 75.15^\circ W]$). The widths of the valleys are ~ 3 km.

3.4.3.3 Comparison With Another Gap-filling Approach

To verify the effectiveness of our gap-filling model, we compared our $PM_{2.5}$ predictions to the $PM_{2.5}$ generated using a reported gap-filling method (Equation (3)). For a fair comparison, Equation (3) followed a subset of our $PM_{2.5}$ predictions derived from the original MAIAC AOD (*i.e.*, removing the $PM_{2.5}$ predictions generated from the gap-filled AOD). The ten-fold cross-validation R^2 values of the monthly models had a mean of 0.69 ranged from 0.60 to 0.78, lower than the OOB R^2 values of our gap-filling model. Figure 4(b) shows the annual $PM_{2.5}$ distribution generated by Equation (3). Even though the model had full-coverage predictions, the $PM_{2.5}$ were spatially over-smoothed. In contrast, since the major $PM_{2.5}$ sources were local sources in NYS (*i.e.*, residential wood combustion and on-road emissions) [66], our $PM_{2.5}$ spatial patterns well captured these well-defined sources.

3.4.4 The Importance of Gap-Filled AOD With Snow Cover Parameter

3.4.4.1 Contribution of Gap-filled AOD to PM_{2.5} Predictions

Since the gap-filled AOD parameters were not among the top-important variables (Figure 3(b)), their contribution to our PM_{2.5} predictions was examined to verify their validity in the prediction model. A prediction model without gap-filled AOD parameters was fitted (a.k.a. no-AOD prediction model), and the analysis of the PM_{2.5} estimates emphasized on the snow season (first 105 days in 2015). Figure 4(c) shows the spatial differences between full-model and no-AOD PM_{2.5} in the snow season. Compared with the no-AOD PM_{2.5}, the full-model PM_{2.5} had higher levels along with the roads and lower levels in the background areas (e.g., the Adirondack Mountains). Hence, after the addition of the AOD parameters in the prediction model, the pattern of PM_{2.5} sources was intensified, in particular on-road emissions. Quantitatively, the mean absolute difference between two PM_{2.5} data sets was 0.13 $\mu\text{g}/\text{m}^3$ with an IQR from 0.06 to 0.18 $\mu\text{g}/\text{m}^3$. The maximum absolute difference reached 0.99 $\mu\text{g}/\text{m}^3$, suggesting a significant change in terms of a 105-day average. Also, the full-model PM_{2.5} was compared with the PM_{2.5} predictions derived from the original MAIAC AOD (Figure S2). The significant differences between two data sets suggested the impact of not only the systematic changes of gap-filled AOD values, but also the significant increase of the sample size. These comparisons suggested that the pollution information incorporated in the gap-filled AOD significantly impacted the PM_{2.5} prediction, leading to discernible changes in PM_{2.5} patterns. The reason for the intensified pattern of on-road PM_{2.5} pollution after the addition of the gap-filled AOD needs further studies.

3.4.4.2 Influences of Snow Cover Parameter on AOD and PM_{2.5}

During and after the snow season, the rank change of the snow fraction's importance in the AOD gap-filling model indicated the importance of this variable in the gap-filling process. Among the 9 variables in the AOD gap-filling model, the average rank of snow fraction's importance was valued as 5 during the snow season and dropped to 8 after the snow season. In contrast, the rank of cloud fraction's importance was stable with an average of 2 throughout the year. As expected, this outcome indicated that the snow fraction was more important

when it was snowy. Further, we fitted a cloud-only AOD gap-filling model (*i.e.*, the snow fraction was removed) to examine the impact of the snow fraction on the gap-filled AOD (the cross-validation performance listed in Table S3). We found that the mean absolute differences between the full-model AOD and the cloud-only AOD were 0.002 (IQR: [0.001, 0.003]) and 0.001 (IQR: [0.0004, 0.001]) for Terra and Aqua, respectively. During the snow season, the maximum absolute difference was 0.007 in terms of a 105-day average. Moreover, we also examined the impact of the snow fraction on the PM_{2.5} predictions. By using the full-model and cloud-only AOD in the PM_{2.5} prediction, two PM_{2.5} data sets were generated: 1) full-model PM_{2.5} and 2) cloud-only PM_{2.5}. Figure 4(d) shows the differences between two PM_{2.5} data sets in the snow season. Compared with the cloud-only PM_{2.5}, the full-model PM_{2.5} tended to be higher in the regions with high PM_{2.5} levels (*e.g.*, populated areas and major roads) and lower in backgrounds. Hence, the snow fraction improved the source patterns of PM_{2.5}. In the snow season, the mean absolute difference between two PM_{2.5} data sets reached 0.09 $\mu\text{g}/\text{m}^3$ (IQR: 0.02 to 0.12 $\mu\text{g}/\text{m}^3$), and the maximum absolute difference reached 0.99 $\mu\text{g}/\text{m}^3$. The mean difference was at a similar scale as the mean difference caused by the gap-filled AOD, which was 0.13 $\mu\text{g}/\text{m}^3$ in the snow season. This suggested that the snow cover might take up the majority of gap-filled AOD’s impact on the PM_{2.5} predictions. In brief, by influencing the gap-filled AOD, the impact of the snow cover was transferred to the PM_{2.5} predictions, resulting in the discernible changes in PM_{2.5} patterns. The reason for the intensified source patterns after the addition of this parameter needs further studies.

3.5 Discussion

In this study, fully covered and high-resolution PM_{2.5} levels were estimated based on the gap-filled MAIAC AOD in New York State in 2015. To the best of our knowledge, this is the first study applying both snow and cloud cover parameters in the gap-filling process to increase the spatiotemporal availability of satellite AOD and the accuracy of PM_{2.5} estimation. Despite the $\sim 90\%$ AOD missingness in the region, our gap-filling models still had excellent performance with a mean OOB R² of 0.93. Xiao et al. [31] adopted a multiple imputation model with similar predictors (except for the snow fraction) to conduct AOD gap-filling in

the YRD of China, in which there was nearly no snow cover, and the AOD missing rate was $\sim 60\%$. They had a mean modeling R^2 of 0.77 with an IQR from 0.71 to 0.82. The improvement in our modeling performance reflected the advantages over the model of Xiao et al. [31] and the potential of machine learning models dealing with the AOD gap-filling with complex snow/cloud-AOD interactions. Our gap-filled AOD was significantly higher than the original MAIAC AOD by an average of 0.15. This outcome is consistent with the findings of previous studies suggesting that the increased humidity caused by cloud could lead to the aerosol hygroscopic growth [26, 27, 31].

Our $PM_{2.5}$ prediction model showed a good performance with a CV R^2 of 0.82 and an RMSE of $2.16 \mu\text{g}/\text{m}^3$. Hu et al. [22], also applying the random forest algorithm to the $PM_{2.5}$ prediction, had a CV R^2 of 0.79 and an RMSE of $2.84 \mu\text{g}/\text{m}^3$ in the Northeastern U.S. (including New York and New England states). Compared to the model of Hu et al. [22], our model achieved the major improvements of 1) conducting satellite AOD gap-filling and making full-coverage $PM_{2.5}$ predictions, 2) conducting $PM_{2.5}$ predictions with a higher 1-km spatial resolution, and 3) achieving similar modeling performance without the convolutional layers of land-use variables. Kloog et al. [15] estimated high-resolution $PM_{2.5}$ with an “out-of-sample” R^2 of 0.88 and an RMSE of $2.33 \mu\text{g}/\text{m}^3$ in the Northeastern U.S. using a linear mixed effects model. However, their $PM_{2.5}$ gap-filling model, virtually a generalized additive model (GAM) with a spline surface (Equation (3)), tended to over-smooth the $PM_{2.5}$ spatial details. In contrast, our $PM_{2.5}$ predictions showed strong local sources in the populated areas and major roads. 2014 National Emissions Inventory suggested that the largest $PM_{2.5}$ emission source in New York State was the residential wood combustion, resulting in 17,916 tons of $PM_{2.5}$ emissions. This emission source was primarily gained around the large and densely populated cities (e.g., New York metropolitan area, Buffalo, Rochester, Syracuse and Albany) [65]. Thus, the correlation between high $PM_{2.5}$ levels and high populations in our areas was reasonable. Besides, the emission inventory suggested that among the top-twelve $PM_{2.5}$ emission sources in New York State (out of 60 Emissions Inventory System emission sectors), at least four of them had the direct correlation with on-road emissions. These sources (paved road dust, on-road diesel heavy-duty vehicles, unpaved road dust, and on-road non-diesel light-duty vehicles) led to 19,753 tons of $PM_{2.5}$ emissions in 2014.

These high-level $\text{PM}_{2.5}$ emissions could interpret the strong $\text{PM}_{2.5}$ signals along with the roads in our areas. Due to the 1-km resolution, our predictions could reflect small-scale $\text{PM}_{2.5}$ features driven by local topographic factors. In some mountainous areas of Upstate New York, $1 - 2 \mu\text{g}/\text{m}^3$ higher $\text{PM}_{2.5}$ accumulated in valleys in winter was found. This phenomenon is consistent with the findings of previous studies suggesting that the stagnant weather conditions in valleys caused by strong temperature inversions could lead to the increase in $\text{PM}_{2.5}$ levels [87–89].

This study considered snow cover in AOD gap-filling process by introducing the snow fraction in the gap-filling model. The significance of the snow fraction was partially reflected by its importance ranking in the gap-filling models. In the snow season (first 15 weeks of 2015), when the heaviest snowfalls appeared, the snow fraction would be the 5th/6th important variable, compared with the rest of the year when its importance only ranked 8th/9th. To further examine the impact of the snow fraction on the $\text{PM}_{2.5}$ predictions, two prediction models without the gap-filled AOD or snow fraction were built. In the snow season, the gap-filled AOD alone caused the absolute changes of $\text{PM}_{2.5}$ by an average of $0.13 \mu\text{g}/\text{m}^3$, and the snow fraction resulted in an average absolute change of $0.09 \mu\text{g}/\text{m}^3$. Accordingly, the impact of the snow fraction might be responsible for the primary influence of gap-filled AOD on the $\text{PM}_{2.5}$ predictions in the snow season. For the spatial pattern, the snow fraction led to the increase of $\text{PM}_{2.5}$ estimates in the areas with large $\text{PM}_{2.5}$ sources (e.g., populated areas and major roads). Previous studies found that the elevated $\text{PM}_{2.5}$ levels appeared to occur on snowy days due to more stagnant atmospheric conditions [32, 34]. The impact of snow on $\text{PM}_{2.5}$ patterns in this analysis could partially reflect this phenomenon. On the whole, by affecting the gap-filled AOD, the impact of the snow fraction was transferred to the $\text{PM}_{2.5}$ predictions, leading to the discernible changes in $\text{PM}_{2.5}$ patterns. These analyses also suggested the advantages of estimating missing satellite AOD data for $\text{PM}_{2.5}$ prediction, instead of directly estimating the missing $\text{PM}_{2.5}$ levels. The former approach can, to a greater extent, draw upon the pollution characteristics incorporated in the satellite AOD and the additional information provided by AOD-related meteorological features. Though this study was limited to New York State, the methodology relative to the AOD gap-filling and $\text{PM}_{2.5}$ prediction is generalizable to other areas with extensive snow/cloud covers and

large proportions of missing satellite AOD data.

The major limitation of this study is that the physical characteristics of snow and cloud are insufficiently considered. The parameters used in this study only reflected the coverage of snow and cloud. Their different physical features, however, may cause the changes in the AOD and $PM_{2.5}$ levels [26]. Accordingly, more snow/cloud characteristics (e.g., cloud optical depth, cloud emissivity, surface albedo, *etc.*) can be applied to better interpret the interactions between snow/cloud and AOD/ $PM_{2.5}$. With an increased number of snow/cloud parameters, a more suitable strategy to incorporate them, instead of additively applying them to the model, also deserves to be considered. Furthermore, the reason for snow/cloud-related AOD/ $PM_{2.5}$ pattern changes needs further studies.

3.6 Conclusions

In this study, an AOD gap-filling model and a $PM_{2.5}$ prediction model based on the random forest algorithm were developed to estimate fully covered and high-resolution ground $PM_{2.5}$ in New York State in 2015. By introducing the MODIS snow/cloud fractions into the gap-filling process, a 100% gap-filled AOD data set was produced with an excellent modeling performance. The 1-km $PM_{2.5}$ predictions derived from the gap-filled AOD could reflect the detailed emission patterns and small-scale terrain-driven features. It is the first attempt where both snow and cloud parameters are introduced into the AOD gap-filling process. Though we only applied fraction measures of snow and cloud, the importance of these parameters was still reflected, and the discernible interactions between snow/cloud and AOD/ $PM_{2.5}$ were observed. It is necessary for future applications to adopt more physical characteristics of snow and cloud and to explore more suitable strategies to introduce these parameters into the gap-filling process. The methodology of this study can be generalized to other areas with extensive snow/cloud covers and large proportions of missing satellite AOD data to estimate $PM_{2.5}$ exposures that previously could not be obtained. The improved $PM_{2.5}$ exposures with an increased sample size and good data quality are expected to contribute to downward epidemiological studies.

4 Manuscript II: Incorporating Low-Cost Sensor Measurements Into High-Resolution PM_{2.5} Modeling at a Large Spatial Scale

Jianzhao Bi, Avani Wildani, Howard H. Chang, and Yang Liu

Environmental Science & Technology **2020**, 54(4), 2152–2162; doi: 10.1021/acs.est.9b06046

4.1 Abstract

Low-cost air quality sensors are promising supplements to regulatory monitors for PM_{2.5} exposure assessment. However, little has been done to incorporate the low-cost sensor measurements in large-scale PM_{2.5} exposure modeling. We conducted spatially varying calibration and developed a down-weighting strategy to optimize the use of low-cost sensor data in PM_{2.5} estimation. In California, PurpleAir low-cost sensors were paired with Air Quality System (AQS) regulatory stations and calibration of the sensors was performed by Geographically Weighted Regression. The calibrated PurpleAir measurements were then given lower weights according to their residual errors and fused with AQS measurements into a Random Forest model to generate 1-km daily PM_{2.5} estimates. The calibration reduced PurpleAir’s systematic bias to $\sim 0 \mu\text{g}/\text{m}^3$ and residual errors by 36%. Increased sensor bias was found to be associated with higher temperature and humidity as well as a longer operating time. The weighted prediction model outperformed the AQS-based prediction model with an improved random cross-validation (CV) R^2 of 0.86, an improved spatial CV R^2 of 0.81, and a lower prediction error. The temporal CV R^2 did not improve due to the temporal discontinuity of PurpleAir. The inclusion of PurpleAir data allowed the predictions to better reflect PM_{2.5} spatial details and hot-spots.

4.2 Introduction

Particulate matter with an aerodynamic diameter $\leq 2.5 \mu\text{m}$ (PM_{2.5}) is associated with a broad range of adverse health outcomes [90, 91] and is a major contributor to the global

burden of disease [92]. Precise and detailed ambient $\text{PM}_{2.5}$ exposure assessment is fundamental to reliably describing $\text{PM}_{2.5}$ -disease relationships [60, 93, 94] and developing $\text{PM}_{2.5}$ pollution control policies [63, 95]. Ambient $\text{PM}_{2.5}$ exposure assessment has traditionally relied on regulatory air quality monitoring stations such as the U.S. Environmental Protection Agency (EPA) Air Quality System (AQS) stations. Due to high instrumentation and maintenance cost, regulatory monitoring is only performed at limited locations for examining the compliance of air quality standards. Given the spatial variability of $\text{PM}_{2.5}$ at the kilometer scale [96], sparse and uneven regulatory monitoring has a limited ability to reflect $\text{PM}_{2.5}$ pollution details [97], especially at remote communities or when impacted by episodic events such as wildfires [98, 99]. This paradigm is shifting with the development of citizen science where many individuals voluntarily collect large amounts of air quality data through low-cost air quality sensors. These low-cost sensors typically cost $< \$2,500$ and have desirable features such as flexibility of deployment and ease of maintenance. Due to the lower costs, they can be deployed more densely than government-operated regulatory stations. Low-cost sensor data have the potential to provide meaningful air quality information in a spatiotemporally more frequent manner.

Since the majority of low-cost $\text{PM}_{2.5}$ sensors are based on the light-scattering principle [100], they tend to have a higher uncertainty than reference-grade monitors. The uncertainty may be caused by the measurement principle itself such as the uncertainty in measured particle counts and the conversion from particle counts to mass concentrations [40, 101]. The manufacturing calibration which uses manufactured aerosols with different composition and properties than those in the ambient environment is another source of uncertainty [100, 102]. The sensors may also experience quality degradation over time [103] and other logistical issues during deployment and maintenance. Therefore, the data quality of low-cost sensors varies with sampling locations and conditions [40, 100, 104]. Previous studies suggested that the pre-test and calibration of low-cost sensors should be conducted where the sensors are intended to be deployed [40, 100, 104]. Current laboratory and field calibration of low-cost sensors mainly focuses on reducing their systematic bias [105–107]. Humidity and temperature were found to be two important factors affecting the systematic bias [102, 108], especially when humidity is high [101, 105, 109]. Multivariate regression models with

these factors as covariates have been widely used to calibrate the sensors against collocated reference-grade monitors, which are able to significantly improve the accuracy of the data but have a limited ability to reduce their residual errors [41, 105, 110].

Currently, there are two primary uses for low-cost $\text{PM}_{2.5}$ sensors. First, they improve monitoring coverage in areas where there is insufficient regulatory monitoring. For example, Pope et al. [111] identified spatial features and diurnal behavior of PM pollution based on low-cost sensor data deployed at three locations in Nairobi, Kenya, a city without long-term reference-grade PM measurements. The applications of low-cost sensors in regions with limited access to regulatory monitoring have advanced local communities' awareness and understanding of air pollution [99, 100, 112]. The second major use of low-cost sensors is to assist with revealing the fine-scale variability of $\text{PM}_{2.5}$, especially in developed countries. In this case, low-cost sensor data are used as a supplement to regulatory measurements in physical or statistical models to fill in the spatiotemporal gaps of $\text{PM}_{2.5}$ concentrations. For example, Masiol et al. [113] incorporated continuous PM concentrations from commercially available low-cost sensors in land-use regression models to derive hourly-resolved PM predictions in Monroe County of New York State. With the addition of calibrated low-cost sensor data to the few regulatory measurements from existing air quality stations in Imperial County of California, Bi et al. [114] found that low-cost sensor data could improve the accuracy of $\text{PM}_{2.5}$ predictions with more reasonable spatial details.

Though it holds promise, there are two major limitations with regard to using a low-cost sensor network to improve $\text{PM}_{2.5}$ pollution mapping and exposure assessment. First, due to the significant cost of extensive field testing by trained scientists [111], the side-by-side low-cost sensor calibration against reference-grade monitors has mostly been confined in a small region, *e.g.*, at a city or county level. In other words, even though low-cost sensors are individually cheap, the high cost of field calibration makes their use at large spatial scales expensive. Field calibration is more difficult for the low-cost sensor networks established by third parties for other purposes. Secondly, even though low-cost sensor data attain a relatively low systematic bias after calibration, their precision is still not comparable to that of reference-grade measurements. The residual measurement errors of sensor data are difficult

to be reduced by current multivariate calibration models [114]. When the calibrated sensor data are treated as ground truth, their residual errors may still significantly influence the reliability of their downstream applications such as hot-spot detection, source identification, and epidemiological analysis [113]. These limitations also apply to other citizen science programs with large amounts of low-quality volunteer-generated data, such as the personal weather data collected by citizens across the U.S. for the Citizen Weather Observer Program [97].

In this study, we proposed a two-step approach to address the aforementioned limitations and optimize the use of low-cost sensor measurements in a spatially extensive, high-resolution $\text{PM}_{2.5}$ exposure assessment. Using a commercial low-cost sensor network as an example, we first conducted a large-scale spatially varying calibration for low-cost $\text{PM}_{2.5}$ data against existing reference-grade measurements. A down-weighting process was then conducted in the prediction stage to reduce the negative impacts of the residual errors of the calibrated sensor data. Our framework is designed to integrate low-cost sensor data with regulatory monitoring data and other sources of information such as satellite, meteorological, and land-use data to improve the high-resolution $\text{PM}_{2.5}$ exposure assessment. This framework could also be informative to other citizen science programs to improve the accuracy of volunteer-generated data.

4.3 Data and Methods

4.3.1 Study Domain and Modeling Strategy

California is the most populous U.S. state with over 39 million residents and the one with the most severe PM pollution, especially in metropolitan areas and the Central Valley [115]. California has a relatively dense regulatory air quality monitoring network and dense low-cost sensors for tracking local air quality. By the end of 2018, there were 157 AQS stations providing $\text{PM}_{2.5}$ measurements and 2,090 outdoor sensors from PurpleAir, a commercial low-cost sensor network, providing sub-hourly $\text{PM}_{2.5}$ measurements within the state. Figure 6(a) shows our study domain with the locations of AQS and PurpleAir monitors. To take advantage of the dense ground monitors and high-resolution satellite aerosol data, we defined

a grid at a 1-km resolution for $\text{PM}_{2.5}$ modeling. The entire study domain consists of 493,561 grid cells. A brief workflow of our two-step modeling approach is shown in Figure 6(b).

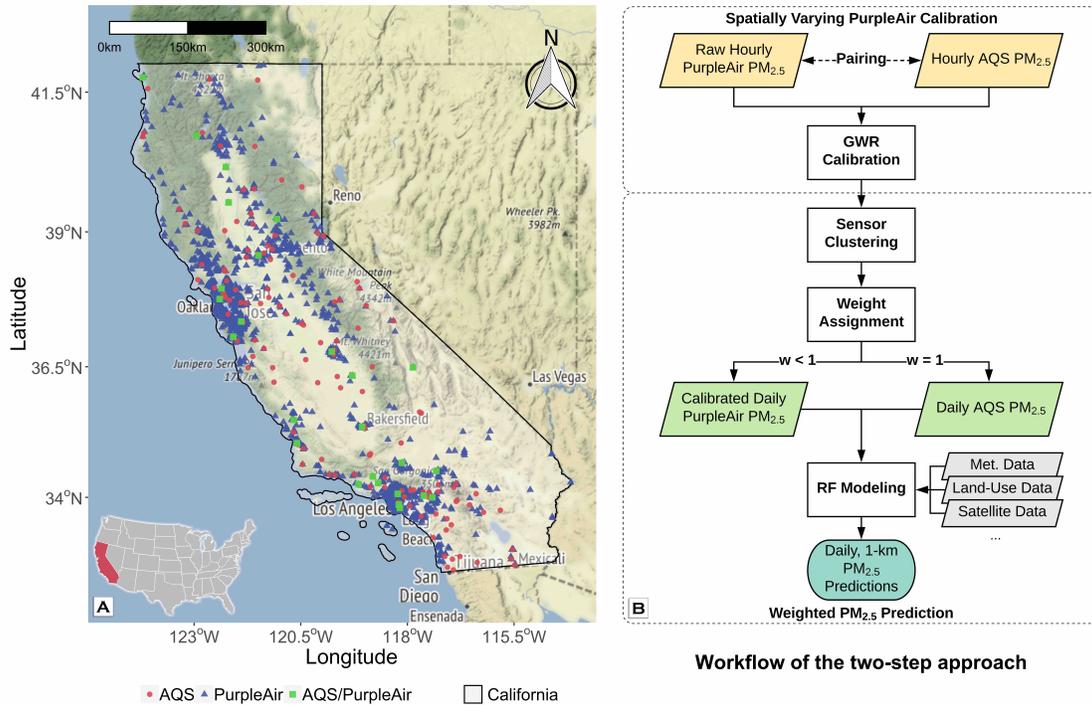


Figure 6: (a): The study domain, California, with the locations of AQS (red), PurpleAir (blue), and paired AQS/PurpleAir monitors (green). The latitude/longitude range of California is $[32^{\circ}30'N, 42^{\circ}N]$ and $[114^{\circ}8'W, 124^{\circ}24'W]$. (b): The workflow of the two-step approach including spatially varying PurpleAir calibration and weighted $\text{PM}_{2.5}$ prediction.

4.3.2 Data

4.3.2.1 $\text{PM}_{2.5}$ Measurements

PurpleAir is a citizen-based, real-time low-cost PM sensor network started in 2015 (<https://www.purpleair.com/>). By the end of 2018, PurpleAir had almost 7,000 sensors worldwide with a growing rate of ~ 30 sensors per day. PurpleAir provides minute-level indoor/outdoor measurement for $\text{PM}_{2.5}$ and other environmental parameters (humidity, barometric pressure, and temperature). We obtained hourly $\text{PM}_{2.5}$ measurements from 2,090 outdoor PurpleAir sensors in 2018 in California ($N = 5,842,404$). Quality control was conducted for these measurements to minimize the outliers (Section 7.2.1, Supporting Information). The raw PurpleAir $\text{PM}_{2.5}$ measurements appeared to bias substantially high against reference-grade

measurements (Section 7.2.2, Supporting Information).

Reference-grade PM measurements were obtained from the EPA AQS regulatory monitoring network (<https://www.epa.gov/aqs>). In 2018, 157 AQS stations provided 50,870 daily $\text{PM}_{2.5}$ measurements in California, and 109 of them provided 499,940 hourly $\text{PM}_{2.5}$ measurements. The hourly $\text{PM}_{2.5}$ measurements from the AQS stations near the PurpleAir monitors were used for PurpleAir evaluation and calibration. Daily PM_{10} measurements were also obtained, which were utilized to generate an ancillary predictor, the $\text{PM}_{2.5}/\text{PM}_{10}$ ratio. This predictor is a continuous surface interpolated from the $\text{PM}_{2.5}/\text{PM}_{10}$ ratio scatters at the locations of AQS stations, representing the distribution of the percentages of $\text{PM}_{2.5}$ in PM_{10} in the study domain. The interpolation was performed by ordinary kriging with month-specific variograms fitted in a spherical model. The $\text{PM}_{2.5}/\text{PM}_{10}$ ratio was shown to be an important predictor of ground-level $\text{PM}_{2.5}$ in California due to relatively high coarse-particle loadings [114].

4.3.2.2 Ancillary Data

Aerosol Optical Depth (AOD) is the integral of aerosol extinction of the solar beam along the entire vertical atmospheric column, which is an important predictor of ground-level $\text{PM}_{2.5}$ [96, 116]. We adopted the satellite AOD retrievals from the Moderate Resolution Imaging Spectroradiometer (MODIS) Multi-Angle Implementation of Atmospheric Correction (MAIAC) product (MCD19A2, <https://1pdaac.usgs.gov/products/mcd19a2v006/>) [117]. MODIS aerosol retrievals have 40 – 50% missing values on average in California due to cloud cover [118]. Therefore, AOD gap-filling was performed by following Bi et al. [96] in which daily-level AOD prediction models were built with satellite-observed cloud fractions and AOD-related meteorological parameters (humidity, visibility, downward shortwave radiation, and wind speed and direction) to derive complete daily AOD surfaces.

Meteorological parameters were obtained from the North American Regional Reanalysis (NARR) (<http://www.emc.ncep.noaa.gov/>) at a 32-kilometer ($\sim 0.3^\circ$) resolution [80] and the North American Land Data Assimilation System (NLDA) (<https://ldas.gsfc.nasa.gov/>) at a 0.125° resolution [81]. The meteorological parameters used in $\text{PM}_{2.5}$ modeling

include visibility, 2-meter air temperature and specific humidity, planetary boundary layer height, 10-meter zonal and meridional wind speeds, shortwave/longwave radiation flux downwards, aerodynamic conductance, convective available potential energy, convective precipitation, and total precipitation. These reanalysis data were aggregated from sub-daily to daily to match the $PM_{2.5}$ data.

The land-use and demographic parameters were obtained from 1) the 2011 National Land Cover Database (NLCD) at a 30-meter resolution (<https://www.mrlc.gov/>), 2) the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation at a 1 arc-second resolution (<https://asterweb.jpl.nasa.gov/>), 3) the Land-Scan ambient population in 2017 at a 900-meter resolution (<https://landscan.ornl.gov/>), 4) the Normalized Difference Vegetation Index (NDVI) from MODIS vegetation indices at a 500-meter resolution (<https://modis.gsfc.nasa.gov/>), 5) the distances to the nearest primary and secondary roads computed from the U.S. Census TIGER/Line[®] Shapefiles (<https://www.census.gov/>), and 6) active fire distributions computed from satellite-retrieved active fire spots (<https://firms.modaps.eosdis.nasa.gov/>).

4.3.3 PurpleAir $PM_{2.5}$ Calibration and Weighted $PM_{2.5}$ Modeling

4.3.3.1 Spatially Varying PurpleAir $PM_{2.5}$ Calibration

The PurpleAir measurements were calibrated against the “gold-standard” AQS measurements. Since both AQS and PurpleAir were existing networks, there were very few strictly collocated AQS/PurpleAir sites in California during the time this analysis was conducted. Instead, we matched a PurpleAir sensor to its nearest AQS station within a 500-m radius so that each AQS/PurpleAir pair was within a 1-km modeling grid cell. The calibration was conducted at the level of single PurpleAir sensors, *i.e.*, the measurements from multiple PurpleAir sensors around the same AQS station were treated separately rather than aggregated together in calibration. A sensitivity analysis indicated that the selected AQS/PurpleAir pairs were robust within a range between 100 to 1000 m without a significant change of the number of pairs. During the study period, 54 PurpleAir sensors were matched to 26 AQS stations, providing 128,777 paired hourly $PM_{2.5}$ measurements.

Given the spatially varying agreement between paired AQS and PurpleAir measurements, Geographically Weighted Regression (GWR) was conducted for the PurpleAir calibration. GWR allows smoothed local relationships between AQS and PurpleAir measurements. Temperature and relative humidity (RH) were used as covariates of the GWR calibration model because these parameters are associated with the data quality of low-cost sensors [40, 105]. In addition, low-cost sensors may experience quality degradation over time [103, 119, 120], thereby the total operating time of a sensor (the duration between the measurement time and the installation time) were used to adjust the effect of sensor aging. Finally, the sensor uptime (the time during which a sensor is in consecutive operation from the last boot time) were used to adjust the potential impact of sensor’s operational stability on data quality. A linear specification was used to describe the relationship between the bias of PurpleAir measurements and four covariates (temperature, RH, operating time, and uptime) (see Section 7.2.3 of Supporting Information for the nonlinearity analysis of PurpleAir bias). The GWR model can be expressed as:

$$\begin{aligned} \text{AQS PM}_{2.5i} = & \beta_0(\mathbf{u}_i, \mathbf{v}_i) + \beta_1(\mathbf{u}_i, \mathbf{v}_i) \times \text{PurpleAir PM}_{2.5i} + \beta_2(\mathbf{u}_i, \mathbf{v}_i) \times T_i + \\ & \beta_3(\mathbf{u}_i, \mathbf{v}_i) \times \text{RH}_i + \beta_4(\mathbf{u}_i, \mathbf{v}_i) \times \text{Optime}_i + \beta_5(\mathbf{u}_i, \mathbf{v}_i) \times \text{Uptime}_i + \epsilon_i \quad \epsilon_i \sim N(0, \tau^2) \end{aligned} \quad (4)$$

where $\beta(\mathbf{u}_i, \mathbf{v}_i)$ indicates the vector of the location-specific parameter estimates and $(\mathbf{u}_i, \mathbf{v}_i)$ represents the geographic coordinates of location i . AQS $\text{PM}_{2.5i}$ and PurpleAir $\text{PM}_{2.5i}$ are paired hourly $\text{PM}_{2.5}$ measurements at location i . T_i , RH_i , Optime_i , and Uptime_i represent temperature, relative humidity, operating time, and uptime of the PurpleAir sensor at location i . The error term ϵ_i is normally distributed with a mean of zero and an overall error variance τ^2 . The optimal hyperparameters of GWR, *i.e.*, the kernel and bandwidth, were chosen based on the corrected Akaike Information Criterion (AICc). In this analysis, the optimal kernel was a Gaussian kernel and the optimal bandwidth was 5,401 nearest neighboring points. All covariates were statistically significant at an alpha level of 0.05 in the GWR calibration model. The GWR was fitted using the R package “GWmodel” version 2.0-7 [121]. In order to examine the impact of the number of collocated AQS stations for

PurpleAir calibration, a sensitivity analysis was conducted with subsets of randomly selected collocated stations (Table S4).

Besides the calibration, another generalized additive model (GAM) was built to quantify the impacts of temperature, RH, operating time, and uptime on the bias of PurpleAir measurements (Equation (5)). This model describes the relationships of the absolute bias of PurpleAir measurements (against paired AQS measurements) and the four covariates. The model can be expressed as:

$$|\text{AQS PM}_{2.5i} - \text{PurpleAir PM}_{2.5i}| = \alpha_0 + s(T_i) + s(\text{RH}_i) + s(\text{Optime}_i) + s(\text{Uptime}_i) + \epsilon_i \quad (5)$$

where i represents a specific paired record and $s(\cdot)$ indicates the smooth function with degrees of freedom of 2 to minimize the random fluctuation in the estimated relationships.

4.3.3.2 Weighted PM_{2.5} Modeling With AQS and PurpleAir Data

After calibration, AQS and PurpleAir measurements were aggregated to daily, 1-km averages for PM_{2.5} modeling. For the 1-km grid cells containing both AQS and PurpleAir measurements, only the AQS measurements were selected to better represent the pollution levels. A weighted Random Forest (RF) model was adopted to generate daily, 1-km PM_{2.5} predictions based on the aggregated daily measurements. Random forests are an ensemble learning method combining the predictions from a multitude of decision trees [83]. RF provides variable importance measures to explain the relative importance and contribution of each predictor. The RF algorithm has been increasingly applied to predicting ground PM_{2.5} levels [22, 96]. An advantage of using RF in this analysis is that it can assign an individual weight to each dependent observation so that the high-quality AQS measurements could have a higher weight than the PurpleAir measurements [122]. An observation with a higher weight will be selected with a higher probability in the samples for building decision trees, therefore having a greater influence on the predictions.

We followed Hu et al. [22] and Bi et al. [96] to perform variable selection and model evalua-

tion based on RF variable importance and random cross-validation (CV). The independent variables used in the prediction models are shown in Table 2. Two major RF hyperparameters, the number of decision trees (n_{tree}) and the number of predictors randomly tried at each split (m_{try}), were tuned based on CV performance. In this analysis, the optimal values of n_{tree} and m_{try} were 500 and 5, respectively. Apart from the RF model with individual weights (refer to as “the weighted model” hereinafter), two reference models were built: one based solely on the AQS measurements (a.k.a. the AQS-based model) and another based on the AQS and PurpleAir measurements without weighting (a.k.a. the non-weighted model). We used 10-fold random, spatial, and temporal CV to evaluate these models. The 10-fold spatial CV procedure creates validation sets according to the locations of the measurements (*i.e.*, dropping 10% of all locations) and the temporal CV creates validation sets according to Julian days. R^2 and root-mean-square prediction error (RMSPE) were the major gauging metrics of CV. It is worth noting that CV was only performed on AQS measurements not used in calibrating PurpleAir to ensure the CV only evaluates out-of-sample model prediction performance. This avoids the issue that calibrated PurpleAir measurements will likely share similar features of matched AQS monitors.

Table 2: Independent variables used in the $PM_{2.5}$ prediction models (s : spatially varying; t : temporally varying).

Prediction Variables	
MAIAC AOD	Ancillary variables
- Gap-filled Terra AOD _(s,t)	- PM _{2.5} /PM ₁₀ ratio _(s,t)
- Gap-filled Aqua AOD _(s,t)	- Day of year _(t)
Land-use variables	- Month _(t)
- Elevation _(s)	Meteorological variables
- Population _(s)	- Visibility _(s,t)
- NDVI _(s,t)	- 2-meter air temperature _(s,t)
- Nearest distance to roads _(s)	- 2-meter specific humidity _(s,t)
- Percentage of shrublands _(s)	- Planetary boundary layer height _(s,t)
- Percentage of herbaceous areas _(s)	- Longwave radiation flux downwards _(s,t)
- Percentage of developed areas _(s)	- Shortwave radiation flux downwards _(s,t)
- Percentage of cultivated areas _(s)	- 10-meter zonal wind speed _(s,t)
- Percentage of forests _(s)	- 10-meter meridional wind speed _(s,t)
- Percentage of water bodies _(s)	- Aerodynamic conductance _(s,t)
- Percentage of wetlands _(s)	- Convective available potential energy _(s,t)
- Percentage of barren lands _(s)	- Convective precipitation _(s,t)
- Active fire distribution _(s,t)	- Total precipitation _(s,t)

Although the systematic bias of PurpleAir data could be reduced by calibration, these measurements still had substantial residual errors which might adversely impact the accuracy of $PM_{2.5}$ predictions. We assigned lower weights to PurpleAir measurements in the prediction process according to their estimated residual errors to mitigate such influences. Similar to the bias, we assumed that the residual errors in calibrated PurpleAir measurements would vary under different environmental conditions. Accordingly, the study domain was partitioned into several sub-domains based on selected variables using Hierarchical Agglomerative Clustering (HAC) [123]. The domain partitioning aimed to obtain distinct $PM_{2.5}$ pollution conditions under which the PurpleAir residual errors would vary. The selected variables were the top-10 predictors with the highest importance values in the AQS-based prediction model (Table S5). HAC performs “bottom-up” clustering, *i.e.*, each unclassified item starts in its own cluster and the two most similar items are merged into a new cluster. This step is iterated until all items are aggregated into a single cluster to form a hierarchical structure. Hierarchical clustering has no hidden assumptions about the distribution of underlying data,

which was suitable in our case as we had little *a priori* understanding of the 10-variable feature space. The number of clusters (K) was determined with the R package “NbClust” [124]. “NbClust” provides 30 indices for determining K and the optimal K can be decided through the majority vote of these indices. The optimal K of our feature space was determined to be 3. Month-specific clustering was conducted as a sensitivity test which showed that our three-cluster partitioning was robust over time.

Weights assigned to calibrated PurpleAir measurements reflected their relative importance to AQS measurements in the $PM_{2.5}$ prediction process. All calibrated PurpleAir measurements in each sub-domain were given the same weight determined using a three-parameter formula (Equation (6)). First, the mean PurpleAir residual variance in each sub-domain (τ_j^2), measured by the variance of the differences between paired PurpleAir and AQS measurements, represents the overall PurpleAir residual error in the sub-domain. Secondly, the error associated with the prediction model structure (σ^2) was estimated as the CV mean squared prediction error (MSPE) of the AQS-based model. σ^2 was the same across different sub-domains. The proportion of the model structure error (σ^2 in the total possible variance ($\sigma^2 + \tau_j^2$)) served as the upper bound of the weight. Finally, in addition to the uncertainty related to the low-cost sensing technology, other factors such as the lack of a consistent siting plan for spatial representativeness can also influence the quality of PurpleAir measurements. To summarize the impact of these unquantifiable circumstances, we included a data-driven scale factor (ρ) with a range of (0, 1) in the weighting formula. Its value was tuned based on CV RMSPE and was determined to be 0.23 in this analysis (Section 7.2.4, Supporting Information). Intuitively, as the overall residual error and unquantifiable uncertainty in calibrated PurpleAir measurements decrease, the weight of the PurpleAir measurements increases within (0, 1) in the prediction model. As a reference, the weight of the “gold-standard” AQS measurements was fixed to 1.

$$w_j = \rho \cdot \frac{\sigma^2}{\sigma^2 + \tau_j^2} \quad \sigma^2 > 0 \quad \tau_j^2 > 0 \quad j = 1, 2, 3 \quad (6)$$

4.4 Results

4.4.1 PurpleAir PM_{2.5} Calibration

4.4.1.1 Evaluation of Uncalibrated PurpleAir Measurements

A linear regression of uncalibrated PurpleAir measurements against AQS had an R^2 of 0.74 and a slope of 0.61. This R^2 was slightly lower than the R^2 values reported by previous studies [42, 125]. Our relaxed pairing strategy between AQS and PurpleAir might lead to this lower agreement. Figure 7(a) shows that uncalibrated PurpleAir PM_{2.5} measurements tracked well with AQS in time but detected more spikes and biased high against AQS by 1.9 $\mu\text{g}/\text{m}^3$. These spikes might be caused by high-level local pollution in the microenvironment near the PurpleAir sensors such as cigarette smoke, barbecues, fireplaces, and idling trucks [110] as most PurpleAir sensors have been installed in residential areas by citizens.

When evaluating PurpleAir sensors at the locations of 21 AQS stations with more than 70 paired PM_{2.5} measurements, a clear variation of the AQS/PurpleAir agreement was observed. Figure 7(b) shows the box plots of the differences between PurpleAir and AQS data at each AQS site. The site-specific R^2 between PurpleAir and AQS ranged from 0.03 to 0.93. The site-specific slope also had a large variation from 0.06 to 1.23. These substantial variations emphasize the necessity of calibrating and assigning lower weights to PurpleAir measurements in a spatially varying manner. Figure 7(b) also shows that the variation of the AQS/PurpleAir agreement was not correlated with their actual distance (correlation coefficient < 0.001), suggesting that 500 m was a reasonable distance for pairing AQS and PurpleAir. As shown in Figure 7(b), some paired AQS/PurpleAir measurements had large differences ($> 50 \mu\text{g}/\text{m}^3$). No temporal patterns among these large differences were found. High-level local pollution in the microenvironment of PurpleAir sensors is believed to be a potential reason for these occasional inconsistencies.

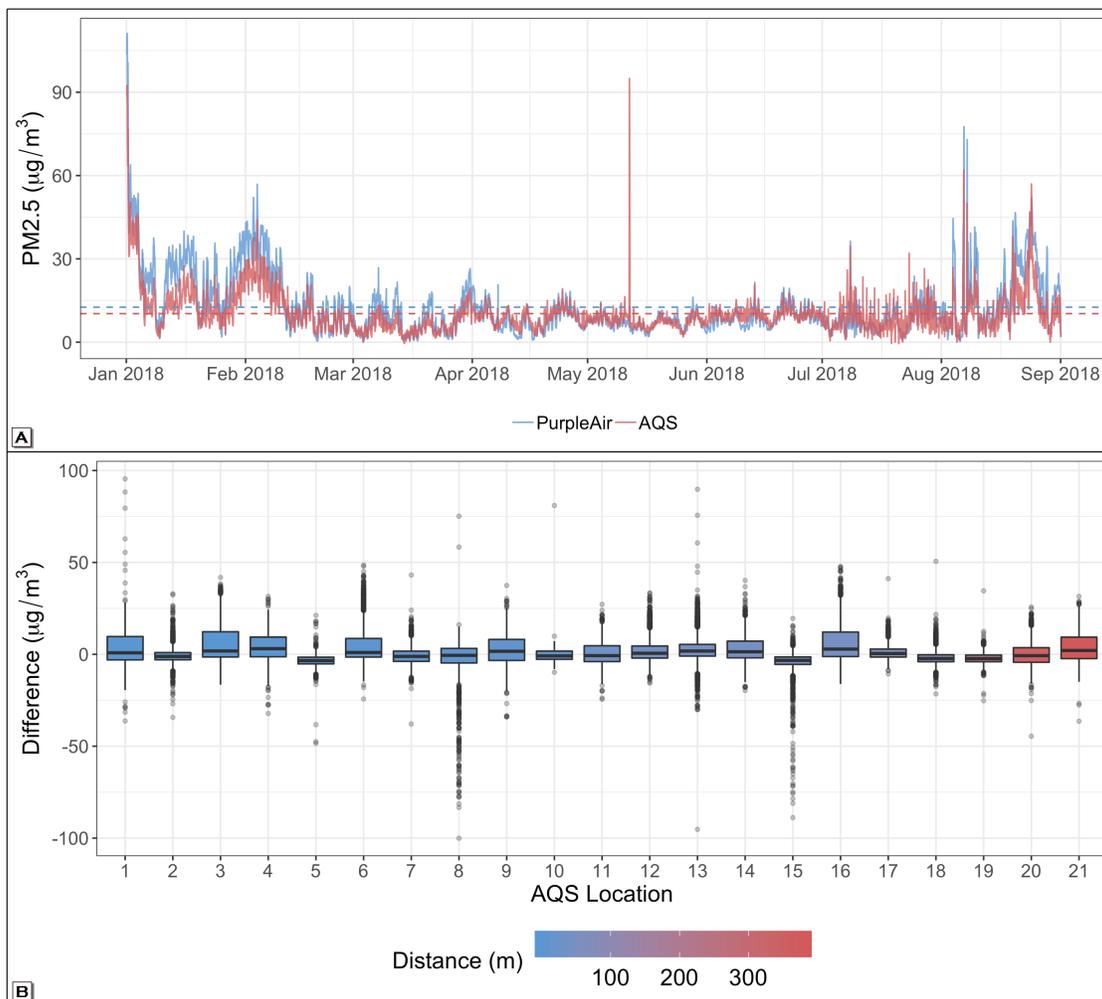


Figure 7: (a): Time series of paired AQS (red) and PurpleAir (blue) hourly measurements with their mean values (dashed lines). The paired measurements were only available from January to August of 2018. (b): Box plots of the differences between paired AQS and PurpleAir hourly measurements (PurpleAir minus AQS) at the locations of 21 AQS stations with more than 70 paired hourly $PM_{2.5}$ measurements (in ascending order of mean distance).

4.4.1.2 Spatially Varying PurpleAir $PM_{2.5}$ Calibration

The GWR slopes of PurpleAir (β_1 in Equation (4)) averaged 0.64 with an interquartile range (IQR) of 0.02. The largest slope was 0.67 near the U.S.-Mexico border in Southern California and the smallest was 0.62 near the coast of Northern California (Figure S3). Although the individual slopes of the AQS/PurpleAir pairs varied significantly across the domain, the calibration slopes had a narrower range because the GWR model fitted the paired measurements in a wider area at each location and other covariates also worked to remove much

of the variation. This is a conservative strategy for mitigating the influence of few paired measurements with extreme coefficients on the calibration model. After calibration, the overall systematic bias of PurpleAir decreased from $1.9 \mu\text{g}/\text{m}^3$ to $\sim 0 \mu\text{g}/\text{m}^3$. The overall PurpleAir residual error was also reduced to some degree, reflected in a decreased standard deviation of the AQS/PurpleAir differences from $8.18 \mu\text{g}/\text{m}^3$ to $5.20 \mu\text{g}/\text{m}^3$ (i.e., a 36% decrease). The calibration model had a 10-fold CV R^2 of 0.78 which is higher than the R^2 of 0.74 between AQS and uncalibrated PurpleAir data, again indicating the improvement of the overall precision of PurpleAir data. Table S4 shows the results of the sensitivity analysis based on randomly selected subsets of collocated AQS stations. When keeping 90% of the collocated AQS stations (23 stations), the calibrated PurpleAir data only had negligible changes. However, when keeping $\sim 80\%$ of the collocated stations (20 stations), although the hourly-level mean absolute difference between the fully calibrated data and the calibrated data based on the subset of collocated stations was still minor ($0.35 \mu\text{g}/\text{m}^3$), the maximum absolute difference started becoming significant ($> 10 \mu\text{g}/\text{m}^3$).

4.4.1.3 PurpleAir Sensor Bias and Influential Factors

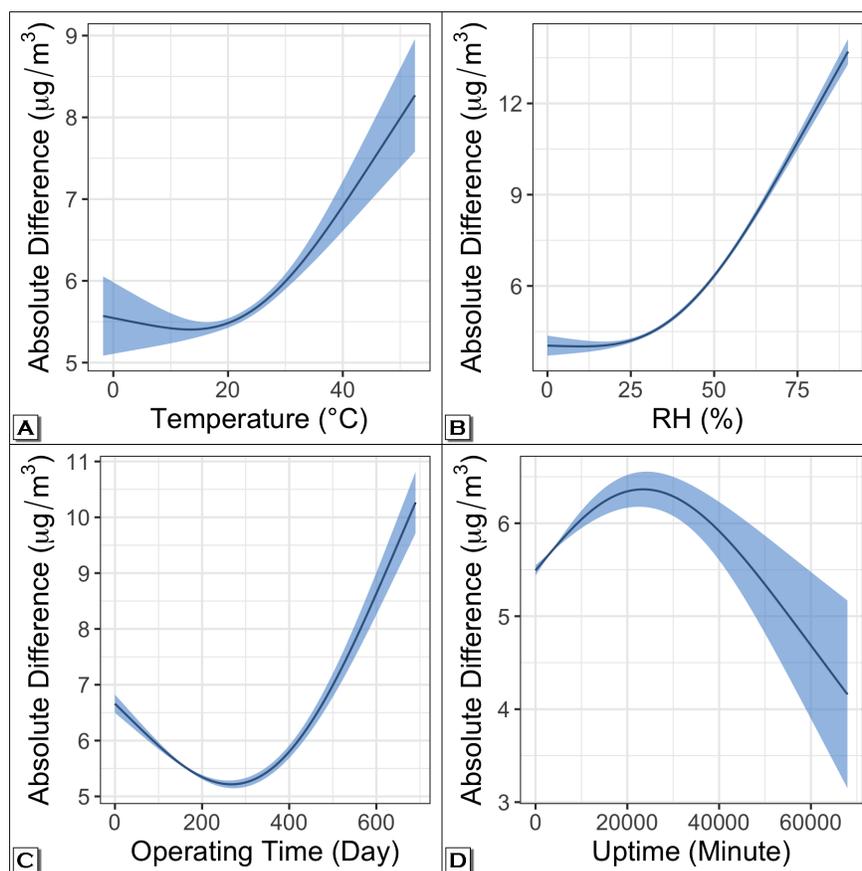


Figure 8: The GAM-fitted relationships with 95% confidence intervals between the absolute differences of paired AQS/PurpleAir hourly measurements and (a) temperature, (b) RH, (c) sensor operating time, and (d) sensor uptime after controlling for other three factors.

Figure 8 shows the GAM-fitted relationships of the AQS/PurpleAir absolute differences and temperature, RH, operating time, and uptime, respectively. The 95% confidence intervals (CIs) of the relationships are shown as the shaded area. In the paired data, temperatures ranged from -1.8°C to 52.6°C with an average of 22.4°C . Temperature was associated with the smallest absolute bias at $\sim 20^{\circ}\text{C}$ after adjusting for other covariates (Figure 8(a)). The bias significantly increased when temperature became higher. At 50°C , the absolute bias was ~ 1.5 times ($\sim 2.5 \mu\text{g}/\text{m}^3$) higher than at 20°C . In contrast, a lower temperature was only associated with a minor increase of bias. RH measures ranged from 0% to 90.1% with an average of 38.8%. RH was positively associated with the absolute bias after adjusting for other covariates (Figure 8(b)). Specifically, the absolute bias was relatively stable at RH <

25% but increased exponentially at $\text{RH} > 25\%$. At 90%, the absolute bias was ~ 3 times ($\sim 9 \mu\text{g}/\text{m}^3$) higher than at 25%. Operating times ranged from 1 to 690 days with an average of 198 days. After controlling for temperature, RH, and uptime, the absolute bias showed a U-curve with a minimum value at ~ 280 days (~ 9 months) (Figure 8(c)). As the operating time became shorter or longer, the bias increased, and the rate of increase was faster for a longer operating time. A sensor with an operating time of 700 days (~ 23 months) had a ~ 2 times ($\sim 5 \mu\text{g}/\text{m}^3$) higher absolute bias than a sensor with an operating time of 280 days. Sensor uptimes ranged from 1 to 67,964 minutes (~ 47 days) with an average of 4,881 minutes (~ 3.5 days). After adjusting for temperature, RH, and operating time, the absolute bias peaked at $\sim 23,000$ minutes (~ 16 days) and became smaller when the uptime was shorter or longer (Figure 8(d)). However, the derived relationship about the uptime had a large uncertainty.

4.4.2 Weighted $\text{PM}_{2.5}$ Modeling

4.4.2.1 Residual Errors and Weights

The clustered sub-domains correspond well with the topographic, meteorological, and land-cover features in California (Figure S4): 1) the first sub-domain consisting of agricultural, humid, and developed areas where most of the population resides, 2) the second sub-domain consisting of mountainous areas such as the Sierra Nevada, and 3) the third sub-domain mainly consisting of the arid areas in the state. The estimated mean residual variances (τ_j^2) and the corresponding weights of PurpleAir (w_j) in the sub-domains are summarized in Table 3. The residual variances were distinct in different sub-domains, varying from 11.2 to 50.0. The variance was smallest in arid areas and largest in mountainous areas, and was modest in agricultural/developed areas. The domain-specific weights of PurpleAir ranged from 0.10 to 0.17.

Table 3: Numbers of paired AQS/PurpleAir hourly measurements, mean PurpleAir residual variances, and PurpleAir weights in three clustered sub-domains. The weights were calculated based on an AQS-based CV MSPE (σ^2) of 33.4 and a scale factor (ρ) of 0.23.

Sub-Domain (j)	N	Mean Residual Variance (τ_j^2)	Weight (w_j)
Agricultural/Developed	118,912	27.22	0.13
Mountainous	3,531	50.00	0.10
Arid	6,334	11.21	0.17

4.4.2.2 Modeling Performance and PM_{2.5} Predictions

Table 4 shows the CV performance of the prediction models. Figure S5 shows the CV scatter plots of the models, indicating that the predictions from all models were slightly underestimated against AQS measurements with slopes of ~ 1.1 . The underestimation is mainly because the RF algorithm is conservative for extreme values and in this analysis, extremely high PM_{2.5} pollution levels tended to be predicted as lower values. The spatial and temporal CV of the AQS-based model had baseline R² values of 0.75 and 0.77, respectively, which were lower than its random CV R² of 0.83. The lower spatial/temporal R² values reflect slightly decreased abilities to extrapolate the PM_{2.5} estimates from the spatial/temporal ranges of the training data to the entire domain/time span. The contribution of PurpleAir data is shown by the higher random CV R² values of both non-weighted and weighted models than that of the AQS-based model. The spatial CV R² values of both non-weighted and weighted models also increased from 0.75 to 0.81. This change indicates that PurpleAir measurements captured PM_{2.5} pollution in more microenvironments despite the network’s lack of a coordinated siting strategy. The temporal CV R² of the non-weighted model decreased from the baseline value of 0.77 to 0.75 possibly due to the lack of sampling continuity of PurpleAir. Unlike AQS, most of the PurpleAir sensors were newly installed during the study period and maintained by untrained citizens so the operations were often intermittent. This lack of sampling continuity could render the measurements less representative in time. The weighted model had a spatial CV R² higher than the baseline value and a temporal CV R² comparable to the baseline value. The model also had the best random CV R² of 0.86 and the lowest RMSPE of 5.62 $\mu\text{g}/\text{m}^3$. These results indicate that the weighting strategy could not only result in higher spatial predictability provided by dense PurpleAir sensors but also

maintain high temporal predictability provided by continuous AQS monitors.

Table 4: Cross-validation performance of three prediction models. CV was only performed on AQS measurements not used in calibrating PurpleAir ($N = 32,981$).

Model	Random CV R^2	Spatial CV R^2	Temporal CV R^2	CV RMSPE
AQS-based	0.83	0.75	0.77	6.04
Non-weighted	0.85	0.81	0.75	5.95
Weighted	0.86	0.81	0.77	5.62

The $PM_{2.5}$ prediction surfaces illustrate the contribution of PurpleAir and the weighting strategy from a different angle. Figure 9 shows the annual mean $PM_{2.5}$ distributions generated from the AQS-based and weighted models as well as their differences. The AQS-based model had an averaged $PM_{2.5}$ prediction of $9.4 \mu\text{g}/\text{m}^3$ and the weighted model had an average of $10.0 \mu\text{g}/\text{m}^3$. The weighted predictions were higher than the AQS-based predictions in almost all areas except for the San Francisco Bay Area, Imperial Valley, and the desert mountain ranges in Southeastern California. The higher predictions were mainly caused by higher calibrated PurpleAir measurements. During the study period, the daily calibrated PurpleAir $PM_{2.5}$ measurements had an average of $12.1 \mu\text{g}/\text{m}^3$, higher than the daily AQS measurements averaged $11.5 \mu\text{g}/\text{m}^3$. Figure 9(c) shows some hot-spots where the weighted predictions were considerably higher than the AQS-based predictions. These hot-spots appear to spatiotemporally coincide with the California wildfires. The black points in Figure 9(c) label the locations of the four most destructive California wildfires in 2018 (Carr Fire, Camp Fire, Mendocino Complex Fire, and Ferguson Fire). The extreme weather conditions during the wildfire events, especially high air temperatures, might influence the quality of PurpleAir sensors. However, when checking the temperature measurements from the PurpleAir sensors near the wildfires, we found that the maximum measurement, 50.3°C , was still within the temperature range of the calibration model (Figure 8). This finding indicates that the PurpleAir measurements near these wildfires were well calibrated and the observed $PM_{2.5}$ hot-spots are not likely caused by highly biased PurpleAir measurements due to high temperatures. As shown in Figure S6, we infer that the density of PurpleAir sensors in the study domain, allowing them to measure such episodic and high-level pollution events, could be one of the reasons for their ability to better reflect the hot-spots. The utility of

low-cost sensors under more extreme conditions still warrants future research. Figure S7 shows that the non-weighted predictions were higher in most of the study domain than the weighted predictions. However, the larger impact of PurpleAir residual errors on the non-weighted model reduced the credibility of its predictions. As PurpleAir measurements were aggregated to daily-level data in this analysis due to the difficulty of current $PM_{2.5}$ models generating predictions at a finer temporal scale, the contribution of high temporal frequency of PurpleAir data on $PM_{2.5}$ predictions warrants further research with improved prediction models.

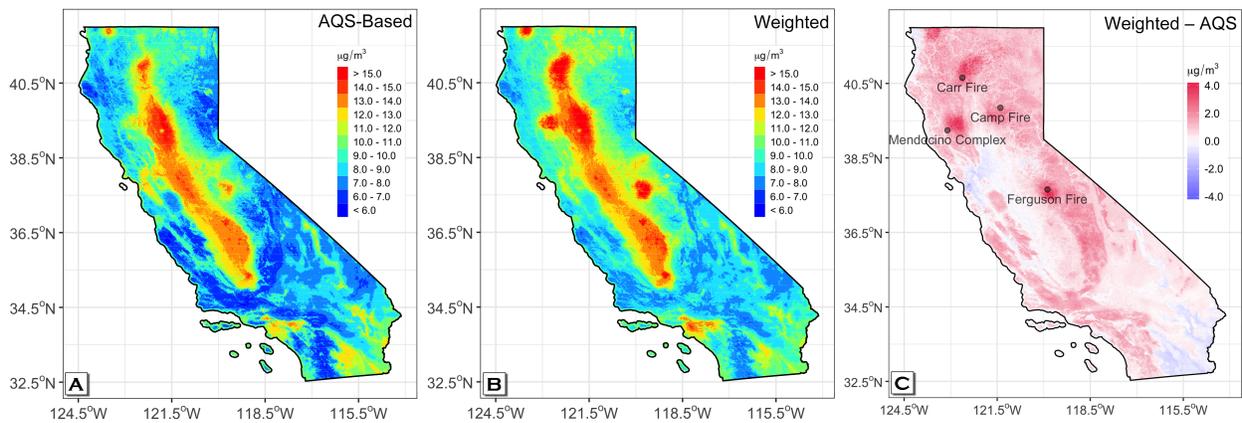


Figure 9: (a) – (b): Annual mean $PM_{2.5}$ distributions for the year of 2018 derived by (a) the AQS-based model and (b) the weighted model. (c): Annual mean $PM_{2.5}$ differences between the weighted and AQS-based models (weighted minus AQS-based) with the locations of the four most destructive wildfires in California in 2018.

4.5 Discussion

In this study, we conducted a spatially varying calibration and developed a down-weighting strategy to integrate low-cost sensor data into high-resolution $PM_{2.5}$ modeling in California. To the best of our knowledge, this is the first time such a framework has been proposed to enable $PM_{2.5}$ prediction models to take advantage of a large volume of low-cost sensor measurements while minimizing the adverse influence of their uncertainties.

Strict side-by-side collocation against reference-grade monitors has been reported in many field calibration studies of low-cost sensors [101, 105, 107, 110]. Although it ensures robustness of calibration, a number of limitations prevent larger-scale implementation of this

method. Because side-by-side collocation in a field measurement campaign is costly and time-consuming, it is usually restricted to a relatively small area such as a city [106, 107] or a county [105]. Calibration coefficients fitted in a small area are difficult to apply in other regions as the low-cost sensor bias may vary under different environmental conditions [40, 100, 104]. More importantly, strict collocation is difficult when both low-cost sensor and regulatory networks are already established. Instead, we tested a less stringent collocation strategy by matching a PurpleAir sensor to its nearest AQS station within a radius of 500 m so that each AQS/PurpleAir pair was within a 1-km modeling grid cell. The reasonability of our collocation strategy was bolstered by the fact that agreement between AQS and PurpleAir data was not related to the actual distance between monitors within 500 m. Furthermore, it allowed for sufficient collocated samples to conduct the calibration. As the AQS/PurpleAir agreement was heterogeneous across the domain where the bias of PurpleAir data was lower near the coastal area of Northern California and higher near the U.S.-Mexico border, the PurpleAir calibration was performed with a GWR model in a spatially varying manner. The calibration reduced the overall systematic bias of PurpleAir data from $1.9 \mu\text{g}/\text{m}^3$ to $\sim 0 \mu\text{g}/\text{m}^3$. The overall residual error of the measurements was also reduced by 36%. Results from the sensitivity analysis examining a reduced set of collocated AQS stations (Table S4) suggest that for a region the size of California, at least ~ 20 well-distributed, continuous reference-grade monitors (capable of providing hourly $\text{PM}_{2.5}$ measurements) are needed to effectively calibrate hourly-level PurpleAir measurements. Thus, our recommended reference-grade monitor density is ~ 5 stations per $100,000 \text{ km}^2$. By the end of 2018, 37 states in the Contiguous United States (CONUS) other than California had a density of continuous AQS stations greater than 5 per $100,000 \text{ km}^2$ (Table S6). Accordingly, the proposed PurpleAir calibration framework may potentially be generalized to the majority of CONUS states without deploying new reference-grade air quality stations. In fact, for the states with a lower network density, an effective calibration could still be conducted by grouping with surrounding states.

Due to the lack of detailed operational conditions of the PurpleAir sensors, assigning site-specific weights was unrealistic. Instead, we clustered these sensors into groups and assigned each group a single weight. We demonstrated a down-weighting approach to minimize the

influences of residual errors and other factors affecting low-cost sensor measurements. The population-level down-weighting formula consists of two parts: an explicit error-variance ratio and a data-driven scale factor (ρ). The error-variance ratio accounts for the proportion of the mean PurpleAir residual variance in the total possible error variance of the model. As the residual error becomes smaller, this ratio becomes larger, and so does the weight. The scale factor ρ was used as a proxy of the negative impacts of implicit factors of low-cost sensors on modeling performance, such as sampling discontinuity and less representative sensor siting [126]. Since the impacts of these implicit factors were unquantifiable, the optimal ρ was determined by our model-fitting data within a range from 0 to 1. It is worth noting that for a set of measurements with an overall quality close to the “gold-standard” measurements, ρ may approach 1 (Section 7.2.4, Supporting Information). In this analysis, the population-level weights were between 0.10 to 0.17, indicating that even though the bias of PurpleAir measurements could be eliminated by a statistical calibration, the contribution of PurpleAir measurements was still no more than 20% of that of AQS measurements in achieving the best model-predicting performance of daily $\text{PM}_{2.5}$ concentrations. Compared to the non-weighted prediction model, the weighted prediction model had a higher random CV R^2 of 0.86, a higher temporal CV R^2 of 0.81, and a lower RMSPE, indicating that the weighting strategy was able to compensate the loss of predictability caused by the PurpleAir residual errors. Through this study and its pilot study conducted in a region with insufficient AQS stations [114], we found that the improvement of statistical metrics, such as the increase of CV R^2 values, can only partly reflect the contribution of low-cost sensors to the quality of $\text{PM}_{2.5}$ prediction, especially when CV is based solely on reference-grade measurements as in this analysis. In the absence of high-quality and high-coverage measurements, we believe that examining the spatial distribution of $\text{PM}_{2.5}$ predictions (Figure 9 and Figure S7) is an important complement to CV metrics to evaluate low-cost sensor data. As shown in this analysis, despite the small improvement on CV metrics, small-scale pollution features are able to be well captured by low-cost sensors. Another possible way to mitigate the influence of low-cost sensor measurement errors on $\text{PM}_{2.5}$ modeling is interpolating these discrete measurements into continuous surfaces and treating the surfaces as an independent variable in the model. We think our weighting strategy is advantageous because incorporating low-

cost sensor measurements into the dependent variable will lead to a significantly larger training sample capable of providing considerably more detailed spatiotemporal information about the pollutant. In this analysis, PurpleAir provided ~ 5 times more training samples than AQS and these samples could help improve the model predictability and better measure pollution hot-spots.

The dense and spatially extensive low-cost sensor measurements allowed us to analyze the potential factors related to the bias and residual error of low-cost sensor measurements. Increased temperature and RH were related to a near-exponentially increased PurpleAir data bias. The observed influences of high temperature/humidity on low-cost sensor bias may be related to the issues in electronic circuits and the hygroscopic growth of fine particulates [109, 127, 128]. The sensor operating time was an influential factor of the bias as well, where a PurpleAir sensor with an operating time of 2 years tended to have a ~ 2 times higher bias than a sensor with an operating time of 9 months. The increased bias over time may reflect the aging effect of sensors [103, 119, 120]. A shorter operating time than 9 months was also associated with a slightly increased sensor bias, suggesting a “break-in” or “warm-up” period of the sensor. The mechanism of the “break-in” warrants further investigation. A longer sensor uptime was in general associated with a lower sensor bias, indicating that stable operation would generally result in better data quality. However, this relationship had a large degree of uncertainty, probably because the sensor’s operational stability is associated with many factors other than the sensor itself, such as the reliability of power supply. In terms of the residual error, the mountainous areas had the highest estimated PurpleAir residual error while the arid areas had the lowest. This difference indicates that 1) humidity may still play a role in the residual errors of low-cost sensor data even after controlling it in the calibration stage and 2) the change of $\text{PM}_{2.5}$ composition in different land-use types may differentially affect the accuracy of the formula the manufacturer of PurpleAir used to convert light scatter to mass concentration. Given the limited information about the sensors we were able to collect, the factors other than temperature, humidity, sensor operating time, and uptime could not be analyzed. More in-depth analyses on the influential factors of sensor data quality are needed but they are beyond the scope of this study.

4.6 Conclusions

In this study, a two-step approach, *i.e.*, spatially varying calibration and down-weighting modeling, was developed to combine low-cost sensor data with regulatory measurements to improve the quality of high-resolution spatiotemporal $\text{PM}_{2.5}$ modeling. The proposed approach was able to mitigate the negative impact of the high noises in low-cost sensor measurements on $\text{PM}_{2.5}$ prediction accuracy. Dense low-cost sensor measurements in the study domain also showed their potential to help the prediction model better reflect $\text{PM}_{2.5}$ hot-spots such as wildfires. This study demonstrated that the integration of low-cost sensors with regulatory monitoring and other sources of information such as satellite remote sensing can provide new insights into $\text{PM}_{2.5}$ pollution. PurpleAir is a global monitoring network with a rapid growth rate. All other supporting data in this analysis, including satellite, meteorological, land-use, and demographic data, are not limited to our study domain. Therefore, our two-step approach can be generalized to other regions to derive high-resolution $\text{PM}_{2.5}$ exposure estimates. The proposed approach is also informative to other meteorological, geographical, and ecological citizen science applications to calibrate large volumes of low-quality volunteer-generated data.

5 Manuscript III: Temporal Changes in Short-Term Associations Between Cardiorespiratory Emergency Department Visits and PM_{2.5} in Los Angeles, 2005 to 2016

Jianzhao Bi, Rohan R. D'Souza, David Q. Rich, Philip Hopke, Armistead G. Russell, Yang Liu, Howard H. Chang, and Stefanie Ebel

5.1 Abstract

Background: Emissions control programs targeting certain air pollution sources may alter PM_{2.5} composition, as well as the risk of adverse health outcomes associated with PM_{2.5}. **Objectives:** We examined temporal changes in the risk of emergency department (ED) visits for cardiovascular diseases (CVDs) and asthma associated with short-term increases in ambient PM_{2.5} concentrations in Los Angeles, California. **Methods:** Poisson log-linear models with unconstrained distributed exposure lags were used to estimate the risk of CVD and asthma ED visits associated with short-term increases in daily PM_{2.5} concentrations, controlling for temporal and meteorological confounders. The models were run separately for three predefined time periods, selected based on the implementation of multiple emissions control programs (EARLY: 2005 – 2008; MIDDLE: 2009 – 2012; LATE: 2013 – 2016). Two-pollutant models with individual PM_{2.5} components and the remaining PM_{2.5} mass were also considered to assess the influence of changes in PM_{2.5} composition on changes in the risk of CVD and asthma ED visits associated with PM_{2.5} over time. **Results:** The relative risk of CVD ED visits associated with a 10 $\mu\text{g}/\text{m}^3$ increase in 4-day PM_{2.5} concentration (lag 0-3) was higher in the LATE period (rate ratio = 1.020, 95% confidence interval = [1.010, 1.030]) compared to the EARLY period (1.003, [0.996, 1.010]). In contrast, for asthma, relative risk estimates were largest in the EARLY period (1.018, [1.006, 1.029]), but smaller in the following periods. Similar temporal differences in relative risk estimates for CVD and asthma were observed among different age groups. No single component was identified as an

obvious contributor to the changing risk estimates over time, and some components exhibited different temporal patterns in risk estimates from $\text{PM}_{2.5}$ total mass, such as a decreased risk of CVD ED visits associated with sulfate over time. **Conclusions:** Temporal changes in the risk of CVD and asthma ED visits associated with short-term increases in ambient $\text{PM}_{2.5}$ concentrations were observed. These changes could be due to changes in $\text{PM}_{2.5}$ composition (e.g., an increasing fraction of organic carbon and a decreasing fraction of sulfate in $\text{PM}_{2.5}$). Other factors such as improvements in healthcare and differential exposure misclassification might also contribute to the changes.

5.2 Introduction

Fine particulate matter ($\text{PM}_{2.5}$) is a well-established environmental health risk factor. Numerous epidemiological studies have shown associations between long-term exposure to $\text{PM}_{2.5}$ and the increased risk of cardiorespiratory diseases [1]. Growing evidence also shows the adverse effects of short-term exposure to $\text{PM}_{2.5}$ on cardiorespiratory diseases [2, 3]. Biological hypotheses suggest that short-term $\text{PM}_{2.5}$ exposure may lead to or exacerbate cardiovascular diseases (CVDs) through neurogenic and inflammatory processes [4] and the acceleration of the development of atherosclerosis [5]. The contribution of $\text{PM}_{2.5}$ to oxidative stress and allergic inflammation may lead to more immediate exacerbations of respiratory diseases, especially asthma [6–9].

As a mixture of many chemical components, certain $\text{PM}_{2.5}$ components may have higher toxicity than others for certain health outcomes [44, 45]. National-scale epidemiological studies have indicated that the risk of adverse health outcomes associated with short-term increases in $\text{PM}_{2.5}$ concentrations varied by region and sub-populations, leading to the hypothesis that the observed heterogeneity may be related to regional differences in $\text{PM}_{2.5}$ composition [46–49]. However, factors other than differences in $\text{PM}_{2.5}$ composition such as different levels of population susceptibility and differential exposure misclassification may also contribute to the observed regional variation. In contrast, estimating temporal changes in $\text{PM}_{2.5}$ health associations in the same region is an approach to mitigate the influence of these other factors. Few epidemiological studies have assessed temporal variation in the risk of cardiorespiratory

disease outcomes associated with short-term increases in $PM_{2.5}$ concentrations. For example, recent work evaluated health effects of short-term exposure to $PM_{2.5}$ in New York State before, during, and after a period between 2005 and 2016 when major emission regulations went into effect and significant emission changes occurred [53–55]. This series of studies found that even with decreasing $PM_{2.5}$ concentrations, the risk of cardiovascular [55] and respiratory diseases [53, 54] was elevated after the implementation of emission policies and an economic recession, which could be driven by temporal changes in $PM_{2.5}$ composition and increased toxicity of the $PM_{2.5}$ mixture [56]. Changes in the acute response to $PM_{2.5}$ over time have also been observed in other regions. Abrams et al. [57] found a smaller risk of cardiorespiratory emergency department (ED) visits associated with short-term increases in $PM_{2.5}$ concentrations after emissions control programs implemented during 1999 – 2013 were fully realized in Atlanta, Georgia. Outside of the United States, Kim et al. [58] reported an increased risk of asthma hospitalizations associated with short-term increases in $PM_{2.5}$ concentrations in Seoul, South Korea from 2003 to 2011 when the Korean air quality standards had been strengthened. In summary, the observed temporal changes in $PM_{2.5}$ health associations reported by previous studies were inconsistent, and few studies also examined temporal changes in associations between individual $PM_{2.5}$ components and adverse health outcomes.

Southern California has some of the highest $PM_{2.5}$ levels in the United States, and the area has implemented stringent control programs. These programs cover almost all controllable emission sources, including on-road and off-road mobile emissions, stationary sources such as fuel combustion, waste disposal, and industrial processes, and area-wide sources such as solvent evaporation, to achieve the compliance of the National Ambient Air Quality Standards (NAAQS) reducing $PM_{2.5}$ and its major precursors (e.g., nitrogen oxides, sulfur oxides, and volatile organic compounds) [129]. In addition, the great recession in the late 2000s may have also accelerated the emission reductions in southern California [130]. In response, the air quality in southern California has significantly improved. The changes in $PM_{2.5}$ concentrations and composition in southern California provide a unique opportunity to investigate whether the risk of acute cardiorespiratory health events associated with each unit change in $PM_{2.5}$ concentration, an indicator of its toxicity, has changed over time due

to different source emissions and resulting mixtures. Therefore, we examined the temporal variation in the risk of CVD and asthma ED visits associated with short-term increases in $PM_{2.5}$ concentrations over the period of 2005 to 2016 in Los Angeles, California. We also similarly examined the temporal variation in the risk of CVD and asthma ED visits associated with individual $PM_{2.5}$ components.

5.3 Data and Methods

5.3.1 Study Population

ED visits data were provided by the California Office of Statewide Health Planning and Development (OSHPD). The study population was restricted to ED patients who lived in any ZIP code area located within 15 miles and visited an ED within 30 miles of the $PM_{2.5}$ monitoring sites in downtown Los Angeles and a nearby community Rubidoux (a total of 147 hospitals) from January 1, 2005 to December 31, 2016. Figure S11 shows the study domains. We included patients with a primary diagnosis (at time of ED visit) of CVD, including ischemic heart disease, cardiac dysrhythmia, congestive heart failure, peripheral vascular disease, and stroke (International Classification of Disease, ICD 9 = 410, 411, 412, 413, 414, 427, 428, 433, 434, 435, 436, 437, 440, 443, 444, 445, 447; ICD 10 = G45, I20, I21, I22, I24, I25, I46, I47, I48, I49, I50, I63, I64, I65, I66, I67, I70, I73, I74, I75, I77, I79), and asthma (ICD 9 = 493; ICD 10 = J45, J46). Multiple ED visits by the same patient on the same day for the same outcome were only counted once. Overall, there were 1,172,516 ED visits for CVD and 522,379 ED visits for asthma over the study period. ED visits were aggregated by day to obtain a daily count time-series for each outcome around each monitor-buffer. The Institutional Review Board (IRB) at Emory University approved this study and granted an exemption from informed consent requirements, given the minimal risk nature of the study and the infeasibility of obtaining informed consent from individual patients.

5.3.2 Air Pollution and Weather

Daily average $PM_{2.5}$ mass and component concentrations were retrieved from two air quality monitoring stations in downtown Los Angeles (Air Quality System, AQS site ID: 06-037-

1103) and Rubidoux (AQS site ID: 06-065-8001). These stations are both NCore Multipollutant Monitoring Network sites (Figure S11). Major $\text{PM}_{2.5}$ components that were monitored included elemental carbon (EC), organic carbon (OC), nitrate, and sulfate. Trace components were also monitored, and components with less than 20% of observations below the minimum detection level (MDL) were selected for inclusion in this analysis, including iron (Fe), sulfur (S), calcium (Ca), potassium (K), silicon (Si), zinc (Zn), bromine (Br), and copper (Cu). $\text{PM}_{2.5}$ mass concentrations were primarily measured using the Federal Reference Methods (FRM) and Federal Equivalent Methods (FEM). Acceptable $\text{PM}_{2.5}$ air quality index & speciation mass (non-FRM/FEM) measurements were also used to increase data coverage. Non-FRM/FEM measurements were linearly calibrated with the FRM/FEM measurements. For EC and OC, linear adjustments were conducted to merge measurements from different samplers and analytical methods [131]. At the two monitoring sites, $\text{PM}_{2.5}$ mass concentrations were generally sampled daily while component data were collected at 1-in-3 or 1-in-6 day intervals. Daily maximum 8-hour ozone concentrations (unit: parts-per-million, ppm, 10^{-6}) at the two NCore stations were also acquired.

Meteorological data were retrieved from the California Irrigation Management Information System (CIMIS) weather stations managed by the California Department of Water Resources (<https://cimis.water.ca.gov/>). Meteorological variables included daily maximum air temperature and mean dew-point temperature. These variables were observed at the daily level at two CIMIS stations inside the two monitor-buffers.

5.3.3 Emissions and Time Periods

Emissions control programs implemented in southern California from 1990s onwards covered virtually all controllable air pollution sources. Key programs targeted on-road vehicle emissions, such as Low Emission Vehicle (LEV I and II) starting in the early 1990s and the heavy-duty diesel vehicle emission standard reductions and fuel reformulation programs in the 2000s. Important programs also targeted off-road emissions from oceangoing vessels, harbor craft, trains, and agricultural equipment in the 2000s and 2010s. Emissions from stationary point sources and area sources were controlled during these periods as well. These

programs directly led to the reduction of primary $\text{PM}_{2.5}$ emissions and indirectly contained the formation of secondary $\text{PM}_{2.5}$ by controlling its precursors such as sulfur oxides (SO_x), nitrogen oxides (NO_x) and volatile organic compounds (VOCs) [129].

Since these broad emissions control programs were often modified during implementation, we relied on annualized emissions inventories to evaluate their cumulative effects and define time periods for our epidemiological analysis corresponding to different periods of implementation. Annualized emission data were retrieved from the California Emissions Projection Analysis Model (CEPAM) 2016 SIP – Standard Emission Tool based on the California Air Resources Board (CARB) 2012 inventory. The geographic area of the inventory is the South Coast Air Basin (SoCAB) which captures almost all of the two monitor-buffers. During the period of 2005 – 2016, primary $\text{PM}_{2.5}$ emission decreased steadily from ~ 35 to ~ 18 tons per day. The source category with the most significant reduction was mobile sources. This emission trend was in tandem with restrictive control programs targeting mobile sources in this period. In comparison, the $\text{PM}_{2.5}$ emission from stationary and area-wide sources remained almost constant at ~ 15 and ~ 30 tons per day, respectively. Since the emissions control programs, especially mobile sources-related programs, took effect over long time periods and their effects were not always immediately fully realized, there were no clear-cut reference or intervention intervals during our 2005 – 2016 study period. Therefore, we defined three equally separated time periods, the EARLY (2005 – 2008), MIDDLE (2009 – 2012), and LATE (2013 – 2016) periods, for our epidemiological analysis.

5.3.4 Statistical Analysis

5.3.4.1 Model Specification

Quasi-Poisson log-linear models were used to estimate the risk of CVD and asthma ED visits associated with short-term increases in $\text{PM}_{2.5}$ concentrations separately for the EARLY, MIDDLE, and LATE periods. Rate ratios (RRs) with 95% confidence intervals (CIs) were calculated based on per $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ concentration to enable the comparison of relative risk estimates between time periods.

Poisson models were specified with distributed lags to reflect cumulative effects of $\text{PM}_{2.5}$

exposure over four days (lag 0-3) and over eight days (lag 0-7) (lag 0 is the same day, and lag 1 is the previous day, *etc.*), which was motivated by previous studies suggesting that the effect of PM_{2.5} may occur over multiple days [53–55]. Models were also controlled for meteorology, via moving averages (MAs) of daily maximum air temperature and mean dew-point temperature using cubic splines with 4 degrees of freedom. The MAs of air temperature and dew-point temperature corresponded to the distributed lags of exposure (*i.e.*, MA of lag days 0-3 or lag days 0-7). Cubic splines for calendar date using 6 (for CVD) or 12 (for asthma) knots per year were included to control for long-term time trends and seasonality. The degrees of freedom of temperature and time splines were determined based on model specification in previous studies on short-term associations between PM_{2.5} and cardiorespiratory disease outcomes [132, 133]. Indicator variables for day-of-week (Monday through Sunday) and holidays (0 = non-holidays, 1 = federal and Federal Reserve Board holidays) were also included. As data were not available for some hospitals over the entire study period (N = 31), an indicator variable was included for these hospitals to account for any changes to ED visit totals attributable to hospital data availability. For asthma, the ED visit counts for influenza were used as an additional confounder to control for viral-induced asthma in flu seasons [134]. The distributed lag model can be expressed as:

$$\log(E(Y_t)) = \alpha + \sum_{q=0}^D \beta_{t-q} \text{PM}_{2.5(t-q)} + [\text{confounders}] \quad D = 3 \text{ or } 7 \quad (7)$$

where $E(Y_t)$ is the expectation of the ED visit counts on day t , $\text{PM}_{2.5(t-q)}$ is the PM_{2.5} concentration q days before day t , and the sum of β_{t-q} is the main parameter of interest for distributed lagged associations.

The relative risk estimates of each outcome (CVD and asthma), lag structure (lags 0-3 and 0-7), and time period (EARLY, MIDDLE, LATE) were estimated separately for Los Angeles and Rubidoux. The effects for each model specification were then pooled together across sub-domains by inverse-variance weighting. Statistical significance of the difference between the RR estimates in two time periods was estimated with the assumption that the logarithms of two RRs were independent and normally distributed.

5.3.4.2 Factors Influencing PM_{2.5} Health Associations Over Time

In order to assess and account for changes in population age structure over time as a potential factor contributing to observed differences in PM_{2.5} health associations between time periods, we conducted age-stratified analyses, in which Poisson models were run by age group (ages 1-18, 19-64, and 65+). We anticipated that within these groupings, the relative risk estimates of CVD and asthma ED visits associated with PM_{2.5} should remain similar across time periods. Any remaining differences in relative risk estimates between time periods should be due to external factors (such as changes in PM_{2.5} composition).

In addition, to examine whether changes in PM_{2.5} composition contributed to the observed differences in PM_{2.5} health associations between time periods, we considered two-pollutant models that included (one by one) individual PM_{2.5} components and the remaining proportion of PM_{2.5} mass (calculated as PM_{2.5} – a specific PM_{2.5} component). In this manner, the health associations for individual PM_{2.5} components controlling for the remaining PM_{2.5} mass were estimated. Since the component data were observed every 3 or 6 days, moving averages with available observations over four days (*i.e.*, MA 0-3) and over eight days (*i.e.*, MA 0-7), instead of distributed lags, were calculated to reflect multiple-day cumulative effects of exposure. Although the moving average method might potentially lead to higher exposure misclassification, this approach considerably increased data coverage and statistical power. The two-pollutant model can be expressed as:

$$\log(E(Y_t)) = \alpha + \beta_1 \times \text{component}_t + \beta_2 \times (\text{PM}_{2.5} - \text{component})_t + [\text{confounders}] \quad (8)$$

where $E(Y_t)$ is the expectation of ED visit counts, and component_t and $(\text{PM}_{2.5} - \text{component})_t$ denote the moving averages of a specific component and the remaining PM_{2.5} mass from day t to previous 3 or 7 days, respectively. The confounders here are the same as which in Equation (7). In this two-pollutant model, β_1 reflects the log-ratio estimate of a PM_{2.5} component controlling for the remaining PM_{2.5} mass, and β_2 reflects the log-ratio estimate of the remaining PM_{2.5} mass controlling for the specific component. We anticipated that the

health associations for specific $\text{PM}_{2.5}$ components (controlling for the remaining $\text{PM}_{2.5}$ mass) should remain similar between time periods (assuming no other external factors contributing to changes in associations), given that individual $\text{PM}_{2.5}$ components are less complex in terms of composition than the $\text{PM}_{2.5}$ mixture and as such their toxicity should be less variant over time. We also anticipated that if the health associations for the remaining $\text{PM}_{2.5}$ mass (controlling for the specific component) corresponded to the temporal pattern displayed by $\text{PM}_{2.5}$ total mass, then this specific component may not influence the changing $\text{PM}_{2.5}$ toxicity over time.

5.3.5 Sensitivity Analysis

Sensitivity analysis was conducted to examine the robustness of the relative risk estimates across time periods. For meteorological confounding, we tested different degrees of freedom of the cubic splines of daily maximum air temperature and mean dew-point temperature ($df = 2 - 6$). We also examined the addition of cubic splines of daily minimum air temperature ($df = 4$). For long-term time trends, different annual knots were tested ($df = 4 - 8$ for CVD and $10 - 14$ for asthma). Additionally, we evaluated associations between tomorrow's pollutant levels (lag -1) and today's ED visits controlling for today's pollutant (lag 0). Tomorrow's pollutant levels should not be associated with today's ED visits as cause must precede effect [135]. Furthermore, we redefined three time periods by 1) moving the year of 2008 to the MIDDLE period and 2) moving the year of 2012 to the LATE period to test the sensitivity of the risk estimates on interval separation. Finally, we evaluated the potential for confounding by exposure to ozone collocated with $\text{PM}_{2.5}$ by adding daily maximum 8-hour ozone concentrations as an additional confounder, as there is a large body of research showing the risk of cardiorespiratory disease outcomes associated with short-term increases in ozone concentrations [136, 137].

5.4 Results

5.4.1 PM_{2.5} Concentrations

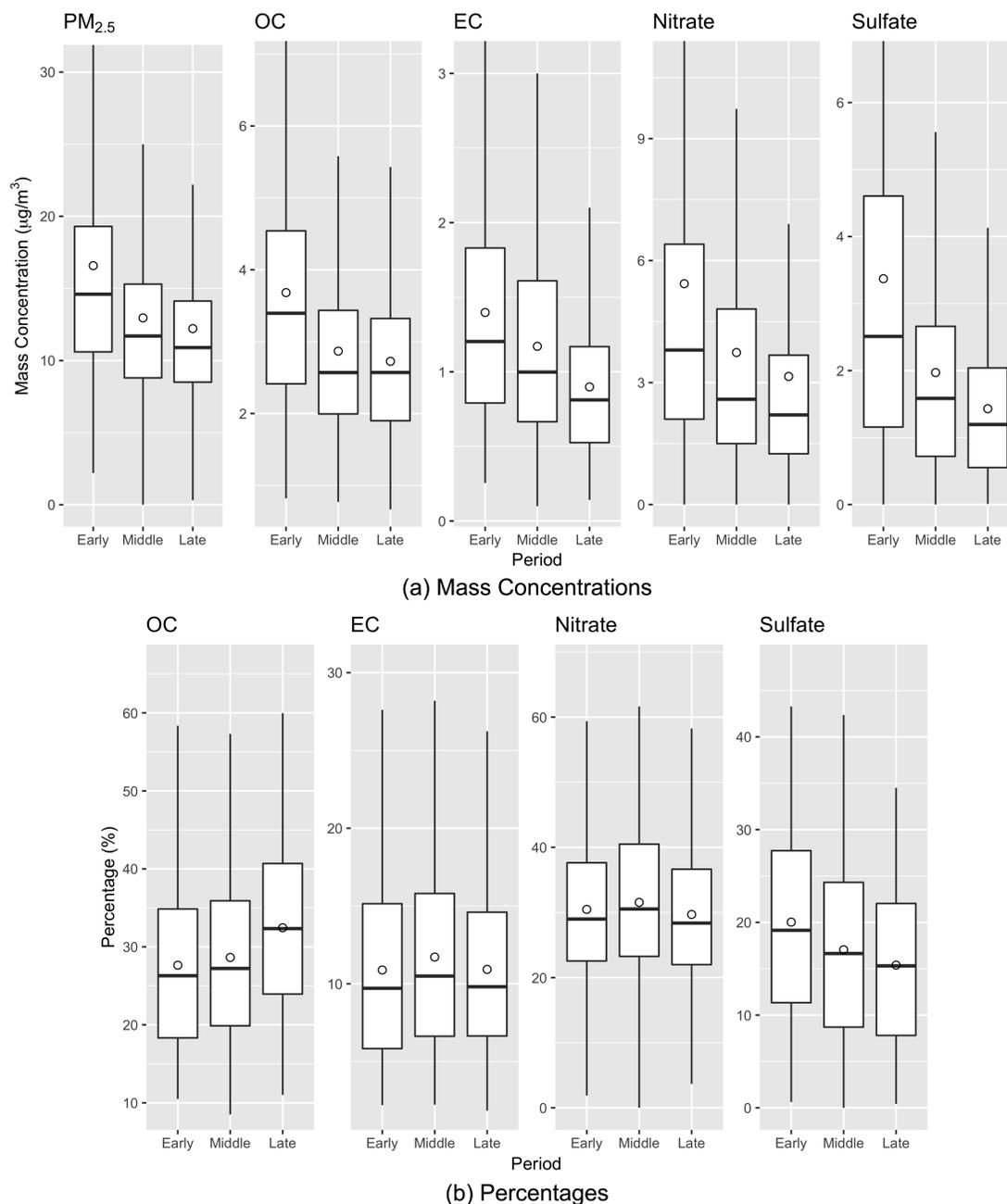


Figure 10: Box plots of (a) mass concentrations of PM_{2.5} and four major components and (b) percentages of four major components in PM_{2.5} total mass during the three time periods (EARLY: 2005 – 2008; MIDDLE: 2009 – 2012; LATE: 2013 – 2016). The measurements are the averages of two sub-domains. The box shows the 25th, 50th, and 75th percentiles and the circle shows the mean value.

Figure 10(a) shows the means, medians, and 25th/75th percentiles of the concentrations of PM_{2.5} total mass and four major components (OC, EC, sulfate, and nitrate) during the three time periods (averages of two sub-domains). The concentrations of PM_{2.5} and four major components decreased over time. The mean PM_{2.5} concentration dropped by 26% from 16.6 $\mu\text{g}/\text{m}^3$ in the EARLY period to 12.2 $\mu\text{g}/\text{m}^3$ in the LATE period. The mean concentrations of OC, EC, sulfate, and nitrate dropped by 26%, 36%, 42%, and 58%, respectively. For most components, the largest drop in concentration occurred between the EARLY and the MIDDLE period, indicating that in addition to emissions control programs, the 2008 economic recession may have also played an important role in the decreased pollution levels. The mean concentrations of other trace PM_{2.5} components across time periods are shown in Figure S12. The trace components with a decreasing trend in concentration over time included Ca, Cu, Fe, Si, S, and Zn. The concentration of Br remained about the same while the concentration of K increased slightly over time.

Figure 10(b) shows the percentage changes of four major PM_{2.5} components over the time periods. The percentage changes of PM_{2.5} components in PM_{2.5} total mass over time were analyzed with the reconstructed PM_{2.5} mass, *i.e.*, the sum of the masses of 12 components. The percentages of EC and nitrate in PM_{2.5} total mass were similar over time with means of $\sim 11\%$ and $\sim 30\%$, respectively. The percentage of sulfate in PM_{2.5} decreased from 20% in the EARLY period to 15% in the LATE period. A decreased percentage of sulfate over time may reflect the effectiveness of emissions control programs on on-road mobile sources, stationary sources, and changes at the Ports of Long Beach and Los Angeles in fuels and in-port electrification, in which the emission of a major precursor of sulfate, SO_x, had a significant reduction [129]. In contrast, the percentage of OC in PM_{2.5} increased from 28% in the EARLY period to 33% in the LATE period. The increased percentage of OC over time echoes a previous finding that there was a slight increase in secondary organic aerosols (SOAs) as NO_x emissions decreased in Los Angeles because non-methane organic gas (NMOG) that previously reacted with NO_x was now available to form more SOAs [138]. Figure S13 shows the percentage changes of 8 trace components over the time periods. The components contributing to an increasing fraction of PM_{2.5} mass over time included Br, Ca, Cu, Fe, K, and Si as the total PM_{2.5} mass concentration declined. The percentage of Zn in PM_{2.5}

remained about the same over time.

5.4.2 Emergency Department Visits Data

Table 5: Summary statistics for two health outcomes (CVD and asthma), $PM_{2.5}$ composition (total mass, four major components, and eight trace components), and weather parameters (daily maximum air temperature and mean dew-point temperature) in Los Angeles and Rubidoux during the three time periods.

	Los Angeles			Rubidoux		
	EARLY (2005 – 2008)	MIDDLE (2009 – 2012)	LATE (2013 – 2016)	EARLY (2005 – 2008)	MIDDLE (2009 – 2012)	LATE (2013 – 2016)
CVD Cases						
N	277,710	292,657	308,303	77,723	104,375	111,748
Age (years)	66.2 (55.0, 80.0)*	66.4 (55.0, 80.0)	66.1 (55.0, 79.0)	61.9 (50.0, 76.0)	61.9 (51.0, 75.0)	62.7 (52.0, 75.0)
Gender (% male)	49.1	49.4	51.3	49.0	50.8	52.0
Race (% white)	51.0	47.8	49.2	65.7	61.4	57.3
Asthma Cases						
N	124,577	130,657	133,717	36,085	48,899	48,444
Age (years)	28.8 (7.0, 47.0)*	29.0 (7.0, 48.0)	29.2 (8.0, 47.0)	25.8 (7.0, 42.0)	26.1 (7.0, 43.0)	27.1 (8.0, 44.0)
Gender (% male)	47.1	47.8	49.4	48.3	48.8	49.5
Race (% white)	38.1	36.8	39.8	54.8	53.5	49.0
Pollutants ($\mu\text{g}/\text{m}^3$)						
Total $PM_{2.5}$	16.58 (9.55)**	12.96 (6.73)	12.21 (6.47)	18.76 (11.90)	13.88 (8.70)	12.48 (7.80)
EC	1.40 (0.79)	1.17 (0.68)	0.90 (0.48)	1.22 (0.80)	1.01 (0.70)	0.80 (0.52)
OC	3.68 (1.69)	2.87 (1.34)	2.73 (1.18)	3.51 (1.84)	2.62 (1.34)	2.56 (1.22)
Nitrate	5.43 (5.73)	3.74 (3.40)	3.15 (3.31)	7.02 (6.34)	5.09 (4.58)	3.74 (3.85)
Sulfate	3.37 (2.91)	1.97 (1.57)	1.43 (1.09)	2.52 (1.94)	1.63 (1.23)	1.26 (1.08)
Fe	0.21 (0.17)	0.20 (0.14)	0.18 (0.11)	0.17 (0.13)	0.14 (0.10)	0.15 (0.09)
S	1.08 (1.01)	0.67 (0.52)	0.52 (0.38)	0.83 (0.64)	0.55 (0.41)	0.44 (0.33)
Ca	0.08 (0.05)	0.07 (0.05)	0.06 (0.04)	0.14 (0.15)	0.08 (0.05)	0.08 (0.06)
K	0.12 (0.61)	0.08 (0.28)	0.09 (0.32)	0.12 (0.33)	0.09 (0.22)	0.11 (0.45)
Si	0.12 (0.14)	0.12 (0.11)	0.11 (0.08)	0.18 (0.19)	0.14 (0.11)	0.16 (0.15)
Zn	0.02 (0.02)	0.01 (0.01)	0.01 (0.01)	0.02 (0.02)	0.01 (0.01)	0.01 (0.01)
Br	0.01 (0.00)	0.00 (0.00)	0.01 (0.00)	0.00 (0.00)	0.01 (0.00)	0.01 (0.00)
Cu	0.02 (0.02)	0.01 (0.01)	0.01 (0.01)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)
Weather ($^{\circ}\text{C}$)						
Maximum Temp	25.0 (19.7, 30.2)*	24.9 (19.9, 29.4)	26.4 (21.8, 30.9)	25.5 (19.6, 31.6)	25.2 (19.7, 30.8)	26.2 (21.3, 31.4)
Dew-Point Temp	8.7 (5.2, 13.3)	8.7 (5.1, 13.3)	9.2 (5.2, 14.1)	6.0 (2.2, 11.1)	5.9 (2.4, 11.0)	6.3 (2.0, 11.6)

*Mean (25th, 75th Percentiles)

**Mean (Standard Deviation)

Table 5 summarizes the characteristics of ED patients and numbers of ED visits by health outcome (CVD and asthma), sub-domain (Los Angeles and Rubidoux), and period (EARLY, MIDDLE, and LATE). ED visit totals for CVD and asthma increased by 15% and 12%, respectively, from the EARLY to LATE period in the study domains. Three time periods had similar age structures for two health outcomes. Patients with CVD had a mean age of 65 years (25th, 75th percentiles = [54 years, 79 years]) and those with asthma had a mean age of 28 years (25th, 75th percentiles = [7 years, 46 years]). There was a relatively even split between genders in the CVD and asthma patient populations, with a slight increase in percentage of male patients over time. Patients with CVD and asthma were predominantly white. The structure of race/ethnicity was stable over time. Overall, the characteristics of the ED patients remained relatively consistent over the study period and the increased

ED visit numbers might be related to the increased population in Los Angeles (<https://data.census.gov/>).

5.4.3 Relative Risks Associated With $PM_{2.5}$ Total Mass

Table 6: Rate ratios (95% confidence intervals) of $PM_{2.5}$ and CVD and asthma ED visits (per $10 \mu g/m^3$ increase in $PM_{2.5}$ concentration).

Outcome	Lag	Period	Rate Ratio	Difference of Rate Ratios Between Two Periods	
			(95% Confidence Interval)	(95% Confidence Interval)	
CVD	0-3	EARLY	1.003 (0.996, 1.010)	MIDDLE – EARLY	0.0058 (-0.0054, 0.0170)
		MIDDLE	1.009 (1.000, 1.018)*	LATE – MIDDLE	0.0112 (-0.0021, 0.0246)
		LATE	1.020 (1.010, 1.030)*	LATE – EARLY	0.0170 (0.0049, 0.0292)*
	0-7	EARLY	0.991 (0.981, 1.002)	MIDDLE – EARLY	0.0190 (0.0027, 0.0354)*
		MIDDLE	1.010 (0.998, 1.023)	LATE – MIDDLE	0.0223 (0.0023, 0.0423)*
		LATE	1.033 (1.017, 1.049)*	LATE – EARLY	0.0413 (0.0227, 0.0599)*
Asthma	0-3	EARLY	1.018 (1.006, 1.029)*	MIDDLE – EARLY	-0.0311 (-0.0497, -0.0125)*
		MIDDLE	0.986 (0.972, 1.001)	LATE – MIDDLE	0.0171 (-0.0046, 0.0388)
		LATE	1.003 (0.988, 1.020)	LATE – EARLY	-0.0140 (-0.0336, 0.0056)
	0-7	EARLY	1.036 (1.018, 1.056)*	MIDDLE – EARLY	-0.0553 (-0.0846, -0.0260)*
		MIDDLE	0.981 (0.959, 1.003)	LATE – MIDDLE	0.0173 (-0.0179, 0.0525)
		LATE	0.998 (0.971, 1.025)	LATE – EARLY	-0.0380 (-0.0704, -0.0056)*

*Statistically significant at an alpha level of 0.05

Table 6 and Figure 11(a) show the rate ratios (RRs) and 95% confidence intervals (CIs) for the associations of CVD ED visits and $PM_{2.5}$ total mass during the three time periods. Both lags 0-3 and 0-7 showed a similar increasing trend over time, where the risk estimates of CVD ED visits associated $PM_{2.5}$ were smaller in the EARLY period and larger in the later periods. For lag 0-3, the RR in the LATE period (1.020, 95% CI = [1.010, 1.030]) was significantly larger than that in the EARLY period (1.003, [0.996, 1.010]) (Table 6). For lag 0-7, all three period-specific RRs were significantly different. Figure 11(b)–(c) show the age-specific risk estimates of CVD ED visits associated with $PM_{2.5}$ during the three time periods. The age groups only included ages of 19-64 and 65+ because there were very few ED visits for CVD under the age of 18 (< 2 visits per day on average). The two age groups showed a similar trend in RR across time periods as the all-age analysis where the RRs increased significantly over time.

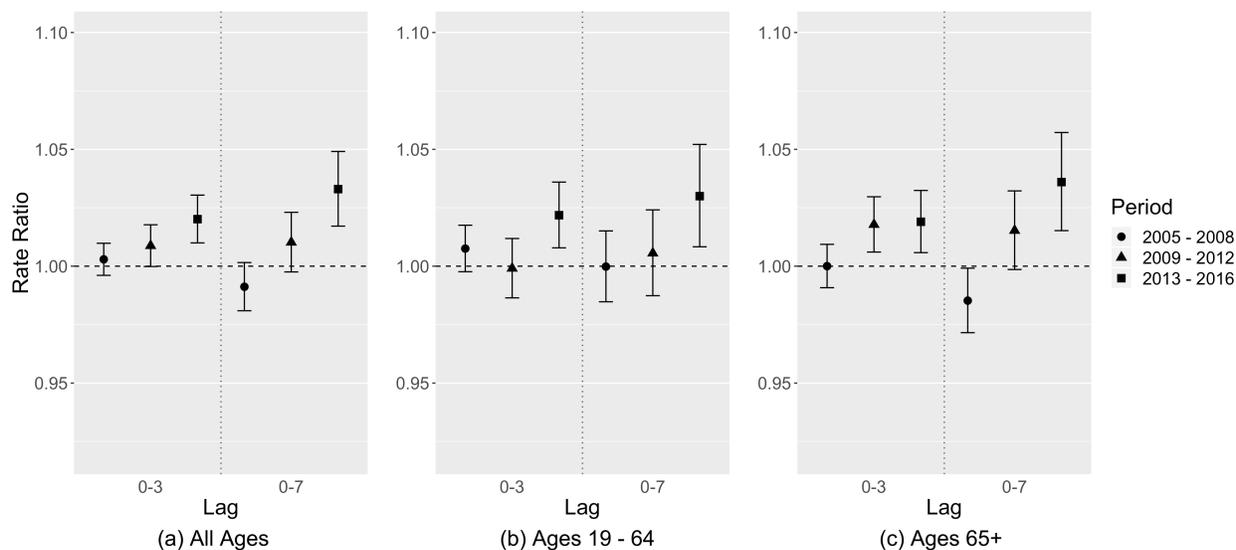


Figure 11: Relative risk estimates of CVD ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ mass concentration of (a) all patients, (b) patients of ages 19-64, and (c) patients of ages 65+, shown as rate ratios (dots) and 95% confidence intervals (whiskers).

Table 6 and Figure 12(a) show the RRs with 95% CIs for the associations of asthma ED visits and $\text{PM}_{2.5}$ total mass during the three time periods. Both lag structures showed a similar pattern over time, where in contrast to the results for CVD the EARLY period had the largest and significant RRs (lag 0-3: 1.018, [1.006, 1.029]; lag 0-7: 1.036, [1.018, 1.056]) compared to the following two time periods. Figure 12(b)–(c) show the age-specific risk estimates during the three time periods for ages of 1-18 and 19-64, respectively. The RRs in the elderly group (ages 65+) had large 95% CIs because of a small sample size (~ 10 visits per day on average) and are not shown in the figure. Adult groups (ages 1-18 and 19-64) had a similar trend in RR to the all-age analysis where the RRs were largest in the EARLY period. For children (ages 1-18), the risk estimates of asthma ED visits associated with $\text{PM}_{2.5}$ were similar in the EARLY and LATE periods, and smaller in the MIDDLE period, while the 95% CIs were large.

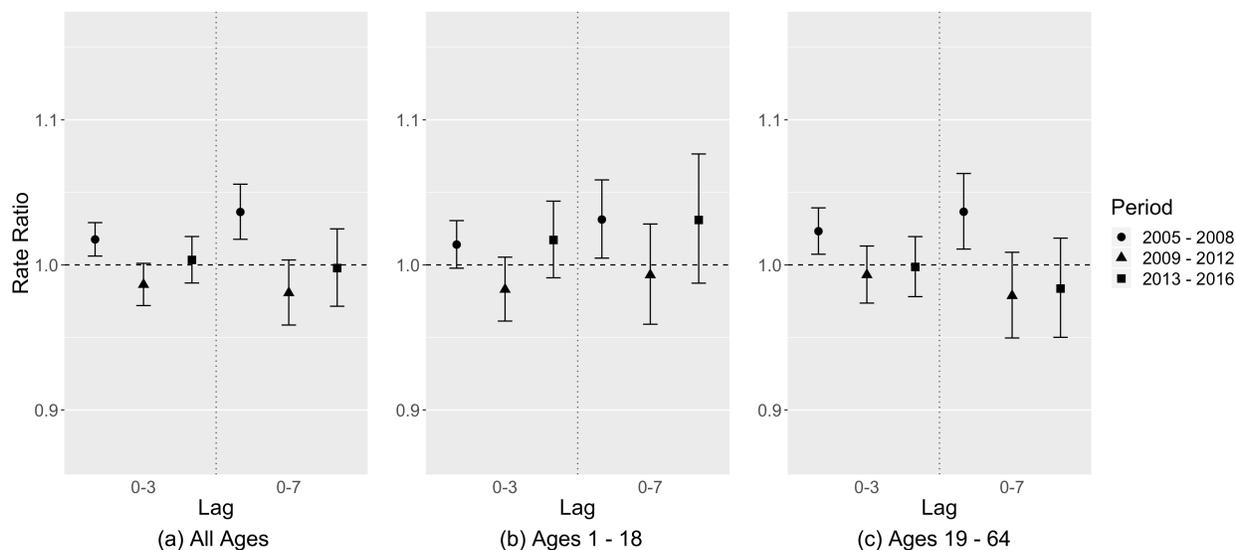


Figure 12: Relative risk estimates of asthma ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ mass concentration of (a) all patients, (b) patients of ages 1-18, and (c) patients of ages 19-64, shown as rate ratios (dots) and 95% confidence intervals (whiskers).

5.4.4 Relative Risks Associated With $\text{PM}_{2.5}$ Components

For each $\text{PM}_{2.5}$ component, we ran a two-pollutant model that included the $\text{PM}_{2.5}$ component of interest and the remaining $\text{PM}_{2.5}$ mass (*i.e.*, $\text{PM}_{2.5}$ – that specific $\text{PM}_{2.5}$ component). Due to sparser observations and the use of moving averages, $\text{PM}_{2.5}$ component concentrations had less temporal variation, resulting in larger uncertainties than that of $\text{PM}_{2.5}$ total mass.

Figure 13 shows the relative risk estimates of CVD ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ component concentration, controlling for the remaining $\text{PM}_{2.5}$ mass. OC and nitrate showed a similar increasing trend in RR over time to $\text{PM}_{2.5}$ total mass (Figure 11(a)). EC, Fe, and Ca showed small RRs which tended to be less than 1.0 in the EARLY period and close to 1.0 in the following periods. Sulfate and S had a high correlation ($r = 0.95$) during the study period so they had a similar pattern, where the RRs were largest and significant in the EARLY period and close to 1.0 in the later periods. The relative risk associated with Si was less than 1.0 and not significant. K had a unique pattern where the MIDDLE period tended to have the largest risk though with a large uncertainty. Figure 14 shows the relative risk estimates of asthma ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ component concentration, controlling for the remaining $\text{PM}_{2.5}$ mass. Nitrate, OC,

and EC showed a similar pattern to $PM_{2.5}$ total mass where the EARLY period had the largest RRs. For sulfate, S, and K, the MIDDLE-period RRs tended to be largest and significant. Associations of Zn, Br, and Cu with CVD and associations of Fe, Ca, Si, Zn, Br, and Cu with asthma were consistent with the null ($RR = 1.0$), with large uncertainties in risk estimates, which are not shown in the results.

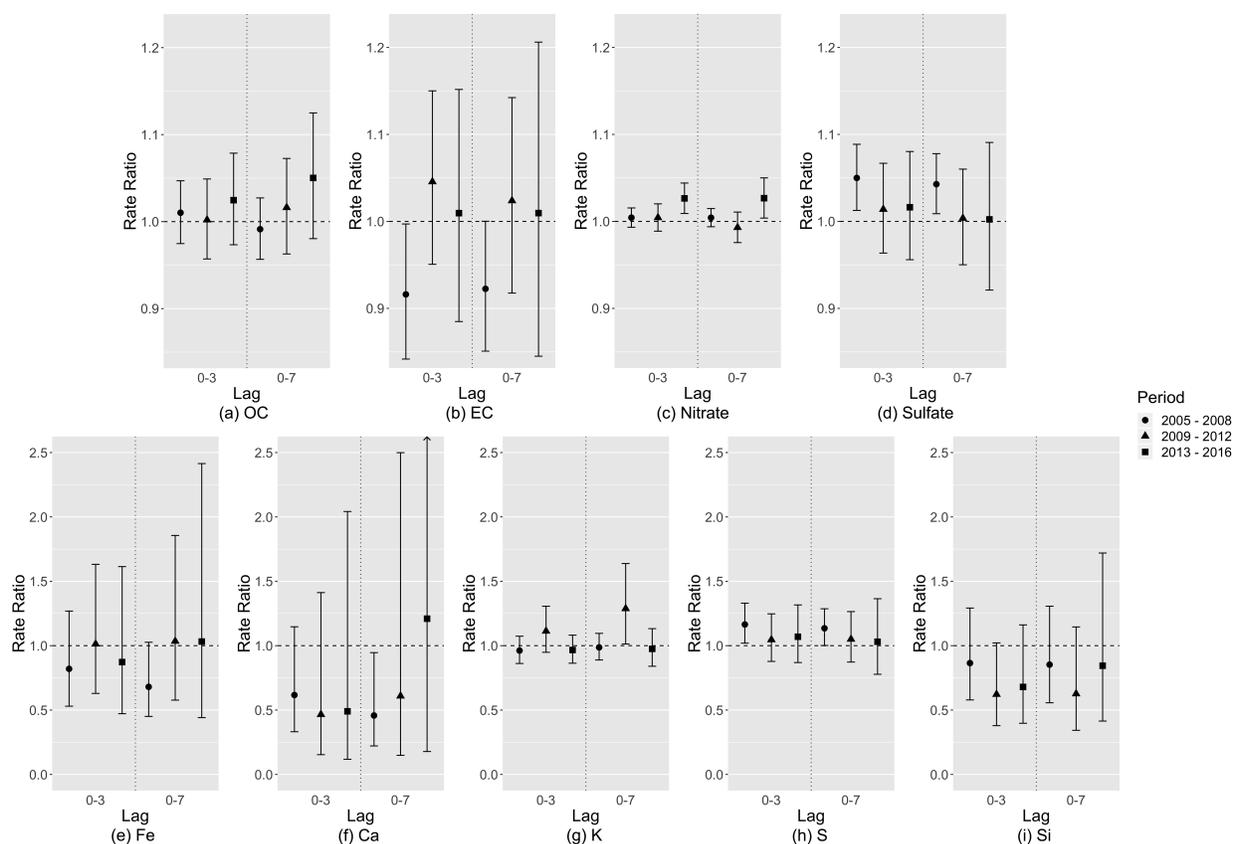


Figure 13: Relative risk estimates of CVD ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in $PM_{2.5}$ component concentration, controlling for the remaining $PM_{2.5}$ mass, shown as rate ratios (dots) and 95% confidence intervals (whiskers).

Overall, this analysis demonstrated temporal variation in the risk of CVD and asthma ED visits associated with short-term increases in $PM_{2.5}$ component concentrations. Given that these $PM_{2.5}$ components are less complex in terms of composition than the $PM_{2.5}$ mixture and that their toxicities should be less variant over time, the observed temporal variation in relative risk suggest that these components might still be proxies of some complex mixtures.

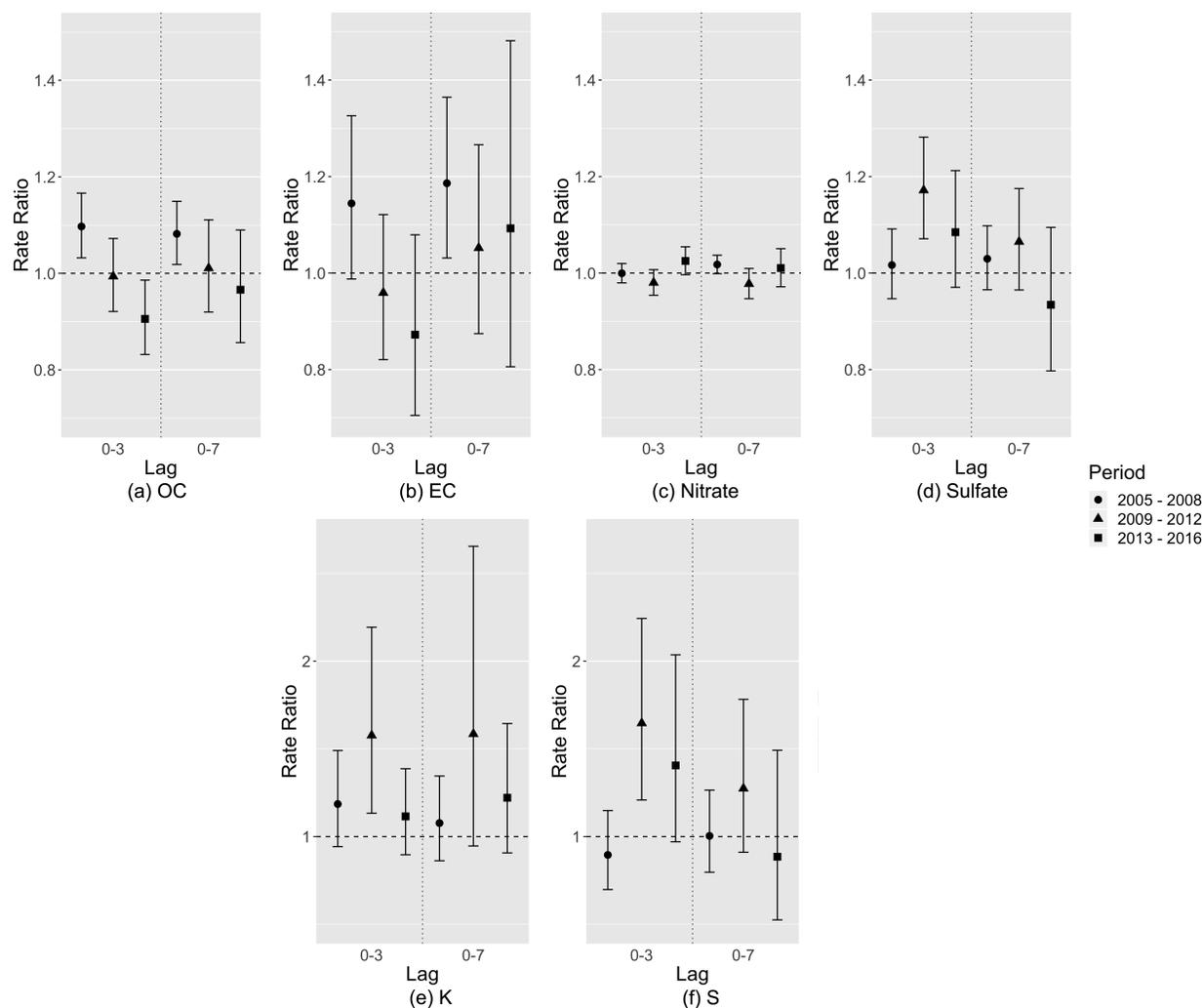


Figure 14: Relative risk estimates of asthma ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ component concentration, controlling for the remaining $\text{PM}_{2.5}$ mass, shown as rate ratios (dots) and 95% confidence intervals (whiskers).

5.4.5 Relative Risks Associated With the Remaining $\text{PM}_{2.5}$ Mass

Figure S14 shows the relative risk estimates of CVD ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in the remaining $\text{PM}_{2.5}$ mass concentration (*i.e.*, β_2 in Equation (8)). Apart from nitrate, the remaining $\text{PM}_{2.5}$ mass for all other components showed a similar pattern to $\text{PM}_{2.5}$ total mass. For nitrate, all RRs were close to 1.0 and not statistically significant, which might be caused by unstable estimated coefficients due to a high correlation between $\text{PM}_{2.5}$ total mass and nitrate ($r = 0.88$). Figure S15 shows the relative risk estimates of asthma ED visits associated with increased remaining $\text{PM}_{2.5}$ mass concentrations. Similarly,

the remaining $\text{PM}_{2.5}$ mass was similar to $\text{PM}_{2.5}$ total mass in terms of the temporal trend of RRs, and some variation might happen by chance (e.g., OC) or due to a high correlation with $\text{PM}_{2.5}$ total mass (e.g., nitrate). In general, for all formulations of remaining $\text{PM}_{2.5}$ mass, the temporal trends in relative risk were similar to that of $\text{PM}_{2.5}$ total mass, thus indicating that no single $\text{PM}_{2.5}$ component (when removed from the $\text{PM}_{2.5}$ mixture) was an obvious contributor to those trends.

5.4.6 Sensitivity Analysis

With different degrees of freedom ($df = 2 - 6$) of cubic splines of daily maximum air temperature and mean dew-point temperature, the relative risk estimates of CVD and asthma ED visits associated with $\text{PM}_{2.5}$ were consistent (Figure S16). When adding daily minimum air temperature ($df = 4$) as another confounder, most of the risk estimates were stable apart from the risk of CVD ED visits in the MIDDLE period (Figure S17). However, the RRs were still within the original 95% CIs, indicating that the change might occur by chance. Figures S18 and S19 show that the risk estimates were consistent with different annual knots in time splines and in redefined time periods, respectively. After controlling for ozone, the risk estimates remained consistent (Figure S20). Finally, there were no significant associations between tomorrow's pollutant levels (lag -1) and today's ED visits when controlling for today's pollutant levels (lag 0), and the lag 0 RRs and 95% CIs remained about the same before and after adding tomorrow's pollutant levels.

5.5 Discussion

In this study, we analyzed temporal changes in the risk of CVD and asthma ED visits associated with short-term increases in $\text{PM}_{2.5}$ concentrations in Los Angeles, California. This study focused on the period of 2005 – 2016 during which comprehensive emissions control programs and economic drivers influenced air quality in the region. Similar to previous studies on short-term associations between $\text{PM}_{2.5}$ and CVD [139] and asthma [140, 141] health events, a significantly increased risk of CVD and asthma ED visits associated with increases in $\text{PM}_{2.5}$ concentrations were observed. More importantly, we also observed temporal varia-

tion in the relative risk with changes in $\text{PM}_{2.5}$ concentrations and composition.

For CVD ED visits, the relative risk estimates were significantly larger in the LATE (2013 – 2016) compared to the EARLY (2005 – 2008) period. The estimated 4-day exposure (lag 0-3) RR increased from 1.003 to 1.020 and 8-day exposure (lag 0-7) RR increased from 0.991 to 1.033 per $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ concentration between the EARLY and LATE periods. For asthma ED visits, the largest RRs were found in the EARLY period (lag 0-3 RR = 1.018; lag 0-7 RR = 1.036) while the RRs were smaller in the following periods. In general, there were significant temporal trends for the risk of CVD and asthma ED visits associated with short-term increases in $\text{PM}_{2.5}$ concentrations, and the trends were similar for different lag days and age groups.

The temporal variation in $\text{PM}_{2.5}$ relative risks could be a result from a number of factors, among which changes in $\text{PM}_{2.5}$ composition could be an important one. With two-pollutant models, we investigated the hypothesis that changes in $\text{PM}_{2.5}$ relative risks were associated with changing fractions of individual $\text{PM}_{2.5}$ components in $\text{PM}_{2.5}$ total mass over time. We hypothesized that if the relative risk of CVD and asthma ED visits associated with a $\text{PM}_{2.5}$ mixture without a certain $\text{PM}_{2.5}$ component was similar across time periods, then this would be support to show that the component in question was a contributor to the observed temporal variation in relative risk. However, we found that the remaining $\text{PM}_{2.5}$ mass without individual components had similar temporal variation in relative risk to $\text{PM}_{2.5}$ total mass, indicating that the observed temporal variation might not be caused by any single component. Another possibility is that the observed temporal variation could result from changing fractions of a group of $\text{PM}_{2.5}$ components acting on different physiological mechanisms. To examine this hypothesis, a multi-pollutant model should be used to analyze the overall association between multiple $\text{PM}_{2.5}$ components and a health outcome. However, due to high correlations between different $\text{PM}_{2.5}$ components, multicollinearity could inflate the uncertainty of risk estimates [142]. Subtracting multiple components from the total $\text{PM}_{2.5}$ could be another way to assess the hypothesis, but as these components had distinct measurement uncertainties, the cumulative error in the generated remaining $\text{PM}_{2.5}$ mass concentrations could make the risk estimates unreliable. Given these limitations, novel and

robust statistical approaches for mixtures are needed to further analyze the combinations of $PM_{2.5}$ components affecting $PM_{2.5}$ relative risks [143].

Despite the fact that no single component was identified as an obvious contributor to the temporal variation in the risk of CVD and asthma ED visits associated with short-term increases in $PM_{2.5}$ concentrations, the component-specific risk estimates still exhibited unique temporal patterns, some of which were different from the pattern of $PM_{2.5}$ total mass. For example, the risk of CVD ED visits associated with increased OC concentrations had an increasing trend over time, which coincided with the increasing percentage of OC in $PM_{2.5}$ total mass. An increasing percentage of OC might include both reactive oxygen species and species with oxidative potential, which could potentially lead to increased oxidative stress and exacerbate cardiorespiratory diseases [144]. Previous studies conducted in New York State also suggested that secondary OC could be a key component leading to temporal changes in the risk of adverse health outcomes associated with $PM_{2.5}$ as oxidative stress is associated more with secondary organic aerosols [53–55]. However, even though increased oxidative stress could be a plausible explanation for the larger risk of CVD ED visits associated with $PM_{2.5}$ over time, it is difficult to explain the smaller risk of asthma ED visits over time. The inconsistent patterns in different health outcomes indicate that 1) it is possible that secondary OC may have different effects on CVD and asthma, 2) there could be other unmeasured factors leading to different risks of different health outcomes associated with OC, or 3) confounding from co-exposure to other pollutants that impact CVD and asthma not fully captured by the model. Besides, due to the lack of source-specific measurements or predictions for secondary OC levels in Los Angeles, we were not able to fully examine its role on the observed temporal variation in relative risk in this study. Additional source apportionment research is needed to further analyze secondary OC and its health effects.

In southern California, sulfate could be a proxy of pollution mixtures related to the combustion of sulfur-containing fuel from motor vehicles, locomotives, ships, and off-road diesel equipment. Residual oil would be a prevalent sulfate source in this region especially at the early end of the study period when the ships started to be forced to switch to lower

sulfur-containing fuel. Rich et al. [145] found that residual oil particles and exhaust gas from spark-ignition and diesel vehicles were associated with increased rates of CVD hospitalizations over the next day. In this analysis, the risk of CVD ED visits associated with increased sulfate concentrations was large and significant in the EARLY period and decreased over time. Since these fuel combustion sources had been well-controlled due to the emissions control programs during the study period [129], it is expected that their adverse effects on CVD could be mitigated. However, sulfate was associated with asthma ED visits in a different manner, where the MIDDLE period (2009 – 2012) seemed to have the largest risk. Again, this inconsistency may be caused by different effects of sulfate on different health outcomes, other unmeasured time-variant factors, or confounding from co-exposure to uncaptured pollutants.

The temporal trends in the risk estimates of both CVD and asthma ED visits associated with increased K concentrations were similar, where the largest risk was in the MIDDLE period. K has been extensively used as an indicator of biomass burning [146], and the MIDDLE period had the lowest mean K concentration associated with fewer wildfire events in southern California. This pattern indicates that the risk associated with emission sources containing K was largest when there were fewer wildfire events and lower K concentrations, which was not expected. Therefore, further research is needed to confirm this health association.

Based on the observed evidence, we infer that other measured or unmeasured time-variant factors, in addition to changes in $PM_{2.5}$ composition, may also play an important role in the change in $PM_{2.5}$ relative risks over time. Asthma may be exacerbated by respiratory infections such as influenza, which can cause inflammation of the airways [134]. Although we controlled for the ED visit counts for influenza in the asthma models, it was still possible that the control was insufficient due to potential under-detection and under-diagnosis of influenza [147, 148]. Exposure misclassification could be another potential factor. If the $PM_{2.5}$ measurements at the monitoring stations were a more representative of population exposures within the respective monitor-buffers in some periods than others, there would be a smaller estimation bias in these periods. However, we would expect for such differential exposure misclassification to affect the risk of both CVD and asthma ED visits and thus

should not be a major factor influencing the observed differences in temporal patterns of two health outcomes. Other time-variant factors such as changes in population vulnerability (e.g., socioeconomic conditions, underlying diseases) and health care accessibility could also be potential effect modifiers for short-term $\text{PM}_{2.5}$ -cardiorespiratory associations. A full investigation for these factors needs detailed community-level information, which warrants further research.

While previous studies reported that $\text{PM}_{2.5}$ -cardiorespiratory disease associations may vary by region and sub-populations [51, 149], this study provides further evidence that these associations may also vary by time, and changes in $\text{PM}_{2.5}$ composition related to emissions control programs and economic changes could be an important driving factor. Apart from those already mentioned, there were several additional limitations in this study. First, the remaining $\text{PM}_{2.5}$ mass concentrations generated by subtracting the mass of individual components from $\text{PM}_{2.5}$ total mass might have some uncertainty due to different measurement errors in different $\text{PM}_{2.5}$ components. This uncertainty might bias the estimated health associations of both the component and remaining $\text{PM}_{2.5}$ parts. In addition, our exposure assignment relying on central air quality stations might have resulted in exposure misclassification, which was a combination of Berkson and classical error, leading to underestimates of risk. The moving average method dealing with temporal sparsity in the component measurements was another possible source of exposure misclassification. Exposure misclassification could result in a bias toward the null and underestimated health associations [150]. Finally, the diagnosis classification codes changed on October 1, 2015, from ICD-9 to ICD-10, might be additional potential concern. However, all ICD-9 and ICD-10 codes were carefully reviewed to ensure consistency of disease groups, and any outcome misclassification should be minimal.

5.6 Conclusions

In this study, we observed temporal changes in the risk of CVD and asthma ED visits associated with short-term increases in $\text{PM}_{2.5}$ mass and component concentrations. These temporal changes could be due to changes in the $\text{PM}_{2.5}$ mixtures such as the increasing

fraction of OC and decreasing fraction of sulfate in $PM_{2.5}$ total mass resulted from comprehensive emissions control programs and economic changes. However, the evidence at the single-component level was not clear. Other factors such as improvements in healthcare and differential exposure misclassification might also contribute to the temporal changes. The complex relationship between changes in the $PM_{2.5}$ mixture and different health outcomes warrants further validations in other geographical regions.

6 Conclusions and Future Directions

In this work, advanced exposure prediction methods were proposed to improve high-resolution $\text{PM}_{2.5}$ exposure assessment based on novel measurement data. Even though we focused more on $\text{PM}_{2.5}$ total mass in this work, these advanced methods can be further applied to the assessment of high-resolution $\text{PM}_{2.5}$ component exposure with the next generation of satellite monitors and low-cost sensors capable of providing more reliable $\text{PM}_{2.5}$ component measurements. The improved $\text{PM}_{2.5}$ exposure data are promising to be utilized in the future health analysis to further improve our understanding of the influence of $\text{PM}_{2.5}$ composition on its overall toxicity.

In Aim 1, an AOD gap-filling model and a $\text{PM}_{2.5}$ prediction model based on the random forest algorithm were developed to estimate fully covered and high-resolution ground $\text{PM}_{2.5}$ concentrations in New York State in 2015. By introducing the MODIS snow/cloud fractions into the gap-filling process, a 100% gap-filled AOD data set was generated with an excellent modeling performance. The 1-km $\text{PM}_{2.5}$ predictions derived from the gap-filled AOD could reflect detailed emission patterns and small-scale terrain-driven features. This study was the first attempt where both snow and cloud parameters had been introduced into the AOD gap-filling process. It is necessary for future applications to adopt more physical characteristics of snow and cloud and to explore more suitable strategies to introduce these parameters into the gap-filling process. The methodology of this study can be generalized to other regions with extensive snow/cloud cover and large proportions of missing satellite AOD data to estimate $\text{PM}_{2.5}$ exposures that previously could not be obtained.

In Aim 2, a two-step approach, *i.e.*, spatially varying calibration and down-weighting modeling, was developed to combine low-cost sensor data with regulatory measurements to improve the quality of high-resolution spatiotemporal $\text{PM}_{2.5}$ modeling. The proposed approach was able to mitigate the negative impact of the high noises in low-cost sensor measurements on $\text{PM}_{2.5}$ prediction accuracy. Dense low-cost sensor measurements in the study domain also showed their potential to help the prediction model better reflect $\text{PM}_{2.5}$ hot-spots such as wildfires. This study demonstrated that the integration of low-cost sensors with regulatory

monitoring and other sources of information such as satellite remote sensing can provide new insights into $PM_{2.5}$ pollution. The two-step approach can be generalized to other regions to derive high-resolution $PM_{2.5}$ exposure estimates. The proposed approach is also informative to other meteorological, geographical, and ecological citizen science applications to calibrate large volumes of low-quality volunteer-generated data.

In Aim 3, significant temporal changes in the rates of ED visits for CVD and asthma associated with changes in $PM_{2.5}$ concentrations and composition were observed. These temporal changes could be due to changes in $PM_{2.5}$ mixtures, such as increasing OC and decreasing sulfate, resulted from comprehensive emissions control policies and economic changes, although the evidence at the single-component level was not clear. Other factors, such as improvements in healthcare facilities and urban environments, exposure misclassification, and residual confounding, might also contribute to the changes. The complex relationship between changing $PM_{2.5}$ mixture and different health outcomes warrants further research in other geographical regions.

7 Appendices

7.1 Manuscript I Supplemental

Table S1: Selection criteria for cloud-, snow-, and water/ice-related AOD missingness by MAIAC AOD Quality Assessment (QA) flags. The digits between the parentheses are the QA flags of the corresponding masks.

Cloud QA Flags	<ol style="list-style-type: none"> 1. MAIAC Cloud Mask <ol style="list-style-type: none"> a. Possible cloudy (010) b. Cloudy (011) c. Cloud shadow (101) d. Fire hot-spot (110) e. Water sediments (111) 2. MAIAC Adjacency Mask <ol style="list-style-type: none"> a. Adjacent to cloud (001) b. Surrounded by more than 8 cloudy pixels (010) c. Single cloudy pixel (011) 3. AOT Quality FLAG <ol style="list-style-type: none"> a. Possible cloud contamination (1)
Snow QA Flags	<ol style="list-style-type: none"> 1. Land Water Snow/Ice Mask <ol style="list-style-type: none"> a. Snow (10) 2. Adjacency Mask <ol style="list-style-type: none"> a. Adjacent to snow (100) b. Snow was previously detected for this pixel (101)
Water/Ice QA Flags	<ol style="list-style-type: none"> 1. Land Water Snow/Ice Mask <ol style="list-style-type: none"> a. Water (01) b. Ice (11)

Table S2: 10-fold overall, spatial, and temporal CV R^2 for the $PM_{2.5}$ predictions from 2002 to 2012. The $PM_{2.5}$ convolutional layer is the most important variable during these years. Land-use terms are also among the top important variables.

Year	Overall CV R^2	Spatial CV R^2	Temporal CV R^2
2002	0.86	0.85	0.80
2003	0.85	0.83	0.82
2004	0.87	0.85	0.85
2005	0.86	0.83	0.83
2006	0.88	0.83	0.85
2007	0.88	0.84	0.86
2008	0.86	0.82	0.84
2009	0.84	0.77	0.82
2010	0.88	0.82	0.86
2011	0.83	0.75	0.81
2012	0.81	0.72	0.80

Table S3: The cross-validation performance of three $PM_{2.5}$ prediction models with different AOD predictors. The full model used gap-filled AOD based on both cloud and snow covers. The cloud-only model used gap-filled AOD based solely on cloud cover. The no-AOD model did not use AOD as predictors. All three models had nearly the same performance. However, cross-validation alone could not fully reflect the contribution of AOD to $PM_{2.5}$ prediction accuracy. The influences of were mainly reflected in the changes of $PM_{2.5}$ distribution patterns.

Model	N	Overall CV R^2	Spatial CV R^2	Temporal CV R^2	RMSE
Full	25,599	0.82	0.74	0.81	2.16 $\mu\text{g}/\text{m}^3$
Cloud-only	25,599	0.82	0.74	0.81	2.15 $\mu\text{g}/\text{m}^3$
No-AOD	25,599	0.83	0.74	0.81	2.13 $\mu\text{g}/\text{m}^3$

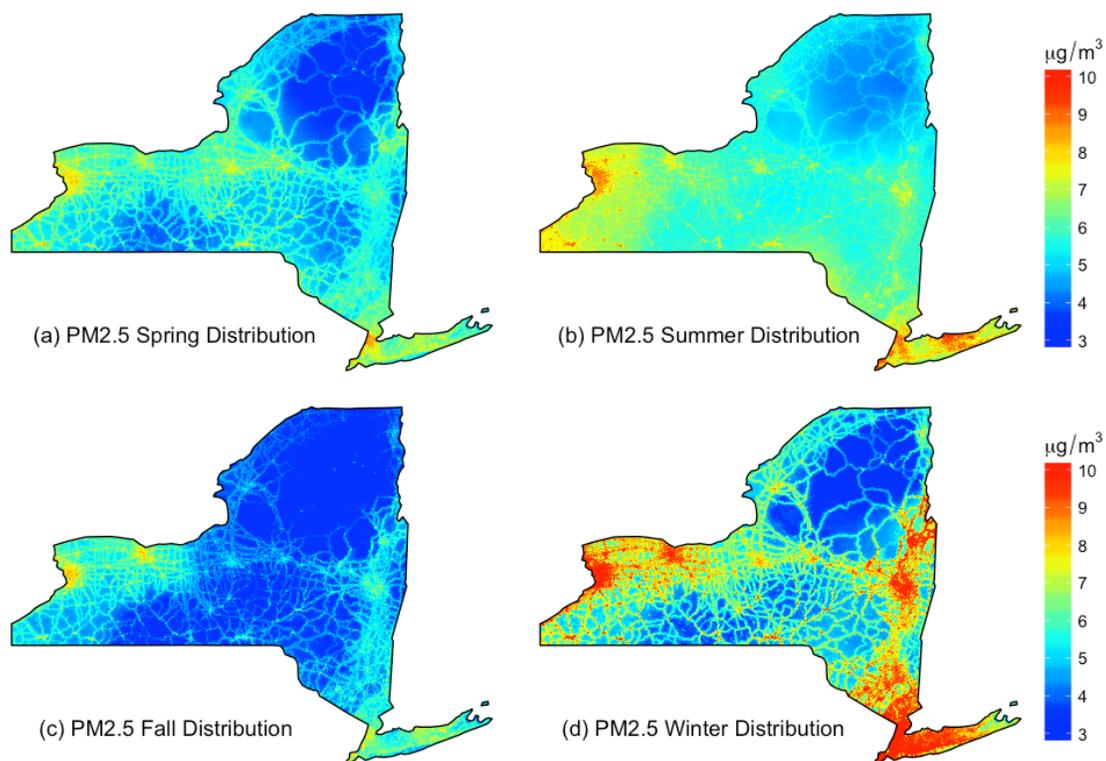


Figure S1: $PM_{2.5}$ spatial distributions in different seasons of 2015: (a) $PM_{2.5}$ distribution in spring; (b) $PM_{2.5}$ distribution in summer; (c) $PM_{2.5}$ distribution in fall; (d) $PM_{2.5}$ distribution in winter.

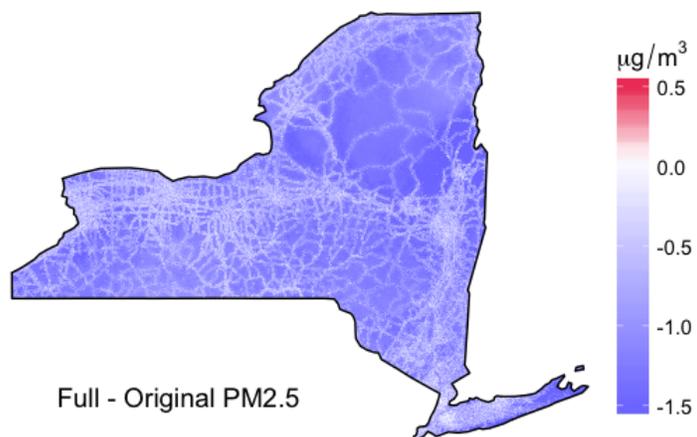


Figure S2: Differences between the full-model and original $\text{PM}_{2.5}$, which had an absolute mean of $0.98 \mu\text{g}/\text{m}^3$ with an IQR from 0.77 to $1.20 \mu\text{g}/\text{m}^3$ and an absolute maximum of $2.29 \mu\text{g}/\text{m}^3$.

7.2 Manuscript II Supplemental

7.2.1 Quality Control for PurpleAir PM_{2.5} Measurements

Each PurpleAir sensor consists of two identical laser particle counters providing two sets of PM readings (Channel A & Channel B). PurpleAir PM_{2.5} data were cleaned based on the dual-channel readings. We first discarded all hourly records with only one channel's reading since the outliers were hard to be identified based on a single channel. Additionally, there were apparent outliers with PM_{2.5} levels greater than 3,000 $\mu\text{g}/\text{m}^3$ in both channels, which were also discarded. The discarded records accounted for $\sim 3\%$ of the total records. The remaining records ($N = 5,658,772$) still had a large dual-channel discrepancy with an R^2 of 0.32 and a slope of 0.60. We then used the absolute percentage bias (APB) computed from dual-channel readings (Equation (S1)) to further filter out the outliers. In Equation (S1), PM_{2.5A} denotes Channel A's reading and PM_{2.5B} denotes Channel B's reading. A percentage threshold of APB was determined according to the improvement of overall dual-channel agreement. When setting the threshold to be 5%, *i.e.*, removing the records with top-5% largest APB values, all apparent outliers disappeared (Figure S8). The remaining data ($N = 5,375,833$) had an excellent dual-channel agreement with an R^2 of 0.98 and a slope of 0.997. The final PurpleAir PM_{2.5} measurements were the average of the dual-channel readings.

$$\text{APB} = \left| \frac{\text{PM}_{2.5\text{B}} - \text{PM}_{2.5\text{A}}}{\text{PM}_{2.5\text{A}}} \right| \times 100\% \quad (\text{S1})$$

7.2.2 Evaluation of PurpleAir PM_{2.5} Measurements

The evaluation of PurpleAir was performed based on the paired hourly PM_{2.5} measurements ($N = 137,068$). The hourly AQS measurements averaged 11.1 $\mu\text{g}/\text{m}^3$ with an interquartile range (IQR) of 9.8 $\mu\text{g}/\text{m}^3$ (25th, 75th percentiles: [5.0 $\mu\text{g}/\text{m}^3$, 14.8 $\mu\text{g}/\text{m}^3$], maximum: 369.0 $\mu\text{g}/\text{m}^3$), while the corresponding PurpleAir measurements averaged 13.0 $\mu\text{g}/\text{m}^3$ with an IQR of 14.8 $\mu\text{g}/\text{m}^3$ (25th, 75th percentiles: [2.8 $\mu\text{g}/\text{m}^3$, 17.6 $\mu\text{g}/\text{m}^3$], maximum: 448.5 $\mu\text{g}/\text{m}^3$). Compared to AQS, PurpleAir measured a higher overall PM_{2.5} level by 1.9 $\mu\text{g}/\text{m}^3$ and

significantly higher peak values. Previous low-cost sensor evaluation studies based on the same sensor (Plantower PMS, Beijing Plantower Co., Ltd) also found that it tended to overestimate $\text{PM}_{2.5}$ compared to reference-grade monitors [42, 151]. For example, Kelly et al. [42] reported that PMS overestimated $\text{PM}_{2.5}$ concentrations when exceeding $10 \mu\text{g}/\text{m}^3$ during several cold-air pools (CAPs) in winter. Badura et al. [151] reported that the raw outputs from PMS overestimated collocated tapered element oscillating microbalance (TEOM) data by a factor of 3.5 during a half-year field campaign.

7.2.3 Nonlinearity of PurpleAir Systematic Bias

The nonlinearity of PurpleAir systematic bias was examined by locally weighted scatter-plot smoothing (LOWESS). LOWESS fits a low-degree polynomial at each point of the data set where the data near the point are given higher weights [152]. LOWESS is a non-parametric strategy for finding a curve of best fit without assuming the distribution of the data. An important hyperparameter of LOWESS is the smoothing span controlling the degree of smoothing. This hyperparameter was tuned with 10-fold cross-validation. A smoothing span of 10% was the optimal value in our case. The LOWESS showed an almost linear curve coinciding with the curve of linear regression (Figure S9), reflecting that linear calibration was satisfactory for PurpleAir.

7.2.4 Validation of Scale Factor

The scale factor ρ was mainly used as a proxy of implicit factors which may impact the prediction quality and further reduce the relative importance of PurpleAir in the model. This factor was set to be a multiplicative term within a range of [0, 1]. Intuitively, for a set of perfect measurements, *i.e.*, the data quality is identical to reference-grade data, this data-driven scale factor should be close to 1. In order to validate this assumption, we pretended the PurpleAir measurements had a perfect quality and used them as ground truth with AQS measurements in the model with 10-fold CV. The trend of CV RMSPE with different ρ values is shown in Figure S10(a). We can see that CV RMSPE reaches its minimum when ρ is closer to 1. This result indicates the reasonability of our assumption, *i.e.*, ρ is a physically meaningful parameter with the value closer to 1 for a perfect data set such as

reference-grade data and closer to 0 for a data set with large uncertainty such as low-cost sensor measurements. The optimal ρ value of the weighted prediction model was tuned with the 10-fold CV (Figure S10(b)). The CV RMSPE shows a U-curve with a minimum at the value of ~ 0.23 . The range of CV RMSPE is as large as $0.2 \mu\text{g}/\text{m}^3$, indicating the large influence of this scale factor on the model performance.

Table S4: Summary statistics of the absolute differences between fully calibrated PurpleAir data and the calibrated data based on subsets of collocated AQS/PurpleAir sites. The total number of collocated AQS/PurpleAir sites was 26 and the subsets were randomly selected from these 26 sites with different proportions from 90% to 10%. This analysis was based on a subset of 10,000 randomly selected PurpleAir measurements.

Proportion*	Mean ($\mu\text{g}/\text{m}^3$)	Q1** ($\mu\text{g}/\text{m}^3$)	Median ($\mu\text{g}/\text{m}^3$)	Q3** ($\mu\text{g}/\text{m}^3$)	Max ($\mu\text{g}/\text{m}^3$)
90%	0.02	0.00	0.01	0.02	0.77
80%	0.35	0.02	0.08	0.44	11.51
70%	0.47	0.07	0.20	0.52	12.58
60%	0.84	0.32	0.65	1.10	13.39
50%	1.06	0.37	0.76	1.34	20.26
40%	1.75	0.72	1.43	2.29	20.70
30%	1.87	0.83	1.62	2.50	20.40
20%	2.26	1.04	2.03	3.11	19.59
10%	2.35	1.01	2.02	3.22	23.49
Raw***	4.59	1.13	2.51	4.89	158.62

* The (gross) proportion of collocated AQS sites being kept

** The 25th and 75th percentiles

*** Uncalibrated PurpleAir data

Table S5: The ten-most important variables of the AQS-based model, based on which the HAC was performed.

PM_{2.5}-Related Variables Used in HAC	
1	PM _{2.5} /PM ₁₀ ratio
2	Elevation
3	Visibility
4	Gap-filled Aqua AOD
5	10-meter meridional wind speed
6	Gap-filled Terra AOD
7	Percentage of shrublands
8	2-meter specific humidity
9	Population
10	Nearest distance to roads

Table S6: Numbers and densities of continuous AQS stations (providing hourly $PM_{2.5}$ measurements) in 47 states of the Contiguous United States (CONUS) (without California). The rows in green are the states with densities of AQS stations > 5 per 100,000 km^2 .

Rank	CONUS State	N of AQS	State Area (km^2)	N of AQS per 100,000 km^2
1	Rhode Island	5	4,001	124.97
2	Delaware	5	6,446	77.57
3	Massachusetts	16	27,336	58.53
4	Connecticut	8	14,357	55.72
5	New Jersey	12	22,591	53.12
6	Maryland	11	32,131	34.24
7	Washington	61	184,661	33.03
8	Pennsylvania	39	119,280	32.70
9	New Hampshire	6	24,214	24.78
10	Florida	39	170,312	22.90
11	Ohio	26	116,098	22.40
12	New York	30	141,297	21.23
13	Indiana	19	94,326	20.14
14	Vermont	4	24,906	16.06
15	Tennessee	16	109,153	14.66
16	North Carolina	20	139,391	14.35
17	Illinois	21	149,995	14.00
18	Kentucky	14	104,656	13.38
19	South Carolina	11	82,933	13.26
20	Wisconsin	18	169,635	10.61
21	Alabama	14	135,767	10.31
22	Idaho	22	216,443	10.16
23	Oklahoma	18	181,037	9.94
24	Minnesota	22	225,163	9.77
25	Georgia	15	153,910	9.75
26	Maine	8	91,633	8.73
27	Virginia	9	110,787	8.12
28	Louisiana	10	135,659	7.37
29	Missouri	13	180,540	7.20
30	Arizona	21	295,234	7.11
31	Iowa	10	145,746	6.86
32	Utah	15	219,882	6.82
33	Michigan	17	250,487	6.79
34	Mississippi	8	125,438	6.38
35	Texas	44	695,662	6.33
36	Colorado	17	269,601	6.31
37	North Dakota	11	183,108	6.01
38	Montana	19	380,831	4.99
39	Wyoming	12	253,335	4.74
40	Nevada	13	286,380	4.54
41	South Dakota	8	199,729	4.01
42	Arkansas	5	137,732	3.63
43	West Virginia	2	62,756	3.19
44	New Mexico	9	314,917	2.86
45	Oregon	6	254,799	2.35
46	Kansas	5	213,100	2.35
47	Nebraska	2	200,330	1.00

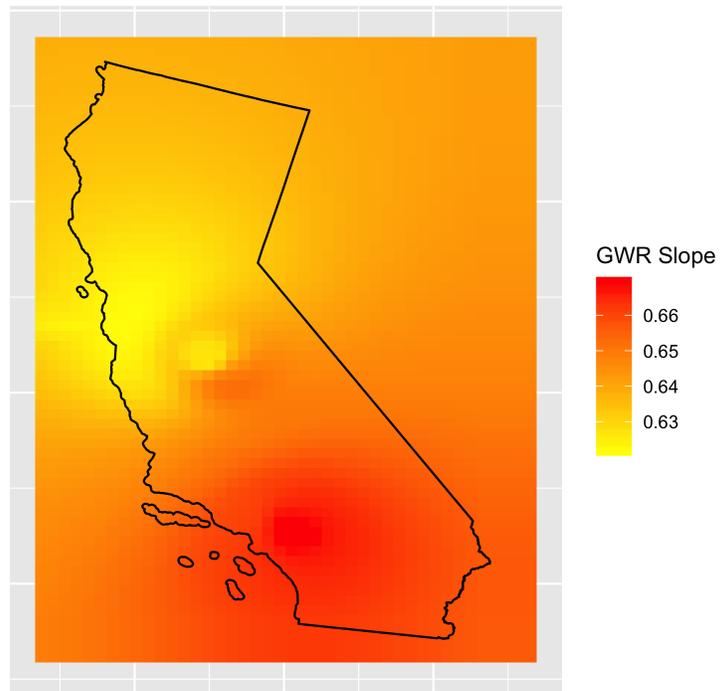


Figure S3: The spatial distribution of GWR slopes.

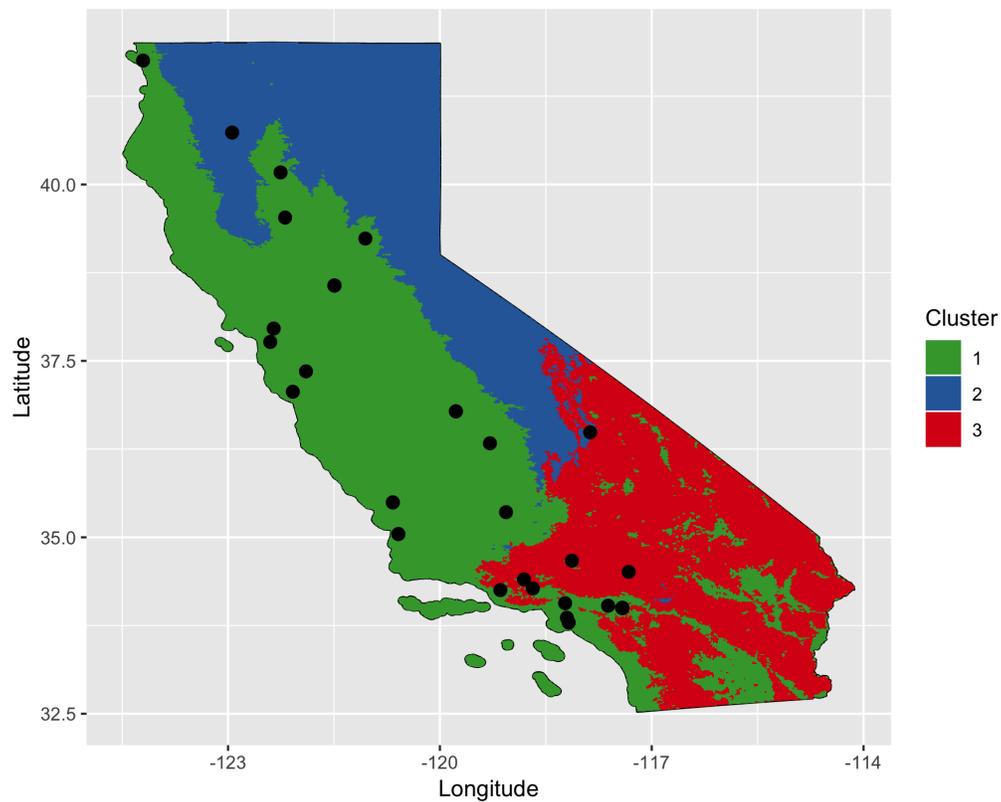
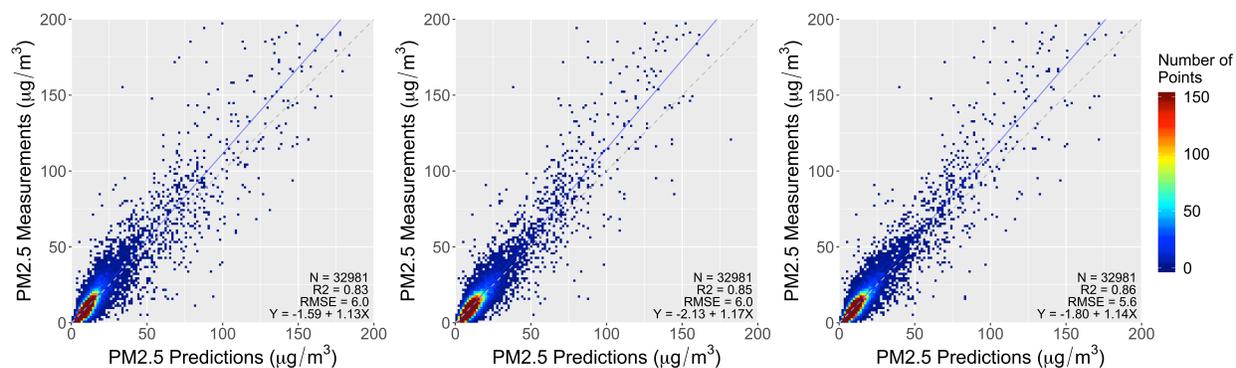


Figure S4: Three clustered sub-domains with the locations of AQS/PurpleAir pairs (black points): 1 – agricultural/developed areas, 2 – mountainous areas, and 3 – arid areas.



(a) CV scatters of the AQS-based model (b) CV scatters of the non-weighted model (c) CV scatters of the weighted model

Figure S5: 10-fold CV scatter plots of (a) the AQS-based model, (b) the non-weighted model, and (c) the weighted model.

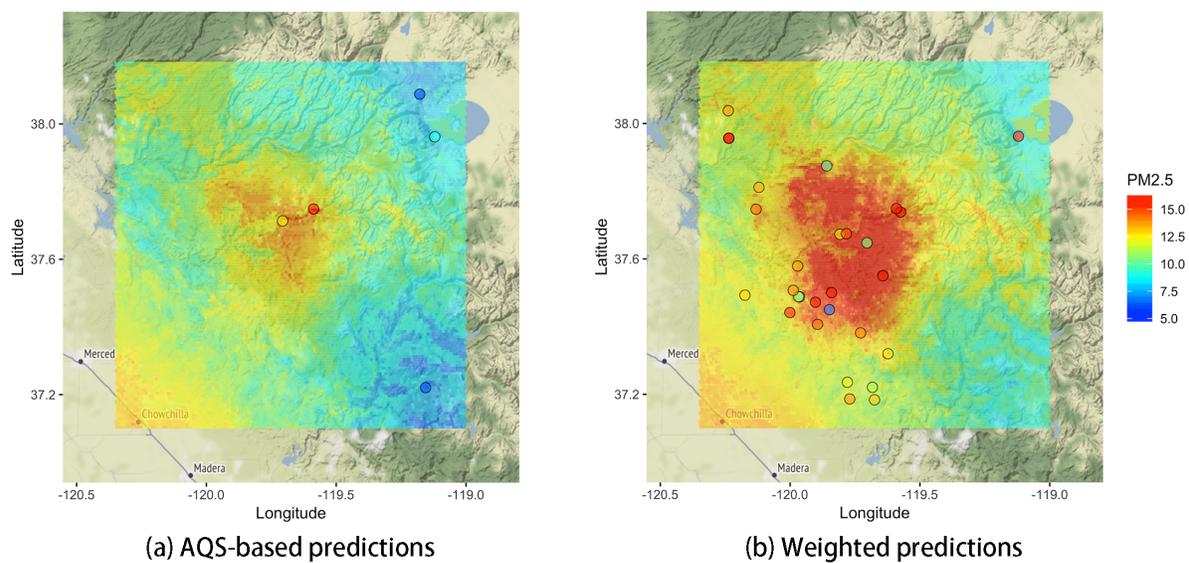


Figure S6: Locations (with annual mean $PM_{2.5}$ levels) of (a) AQS and (b) PurpleAir sites and the annual mean $PM_{2.5}$ distributions derived by the (a) AQS-based and (b) weighted models in the region of Ferguson Fire in 2018.

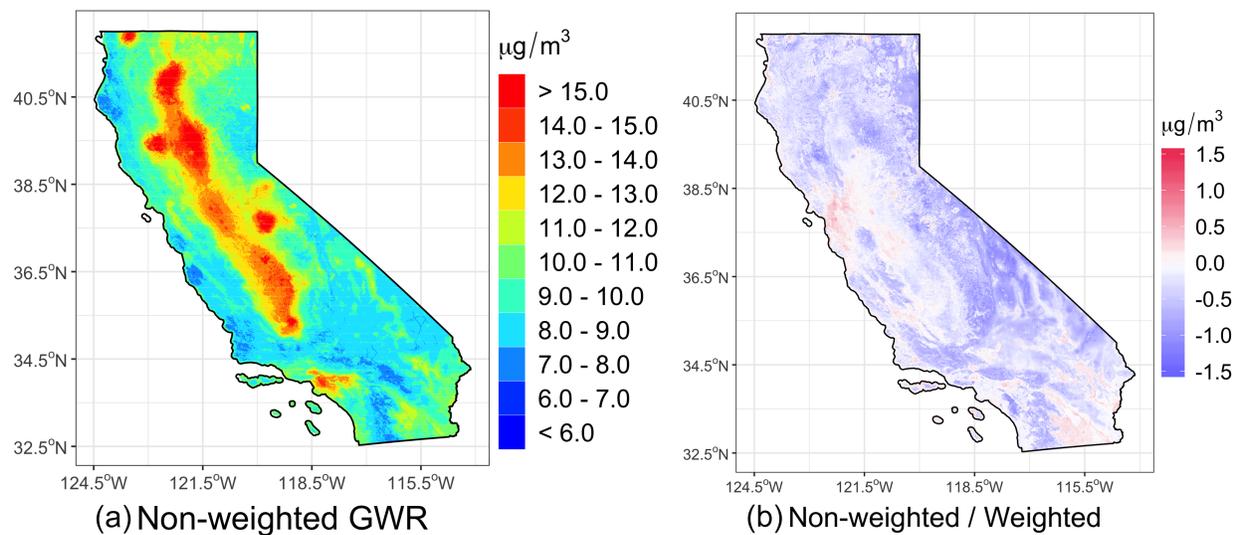


Figure S7: (a): Annual mean $PM_{2.5}$ distribution for 2018 from the non-weighted model. (b): Annual mean $PM_{2.5}$ differences between the non-weighted and weighted predictions (non-weighted minus weighted).

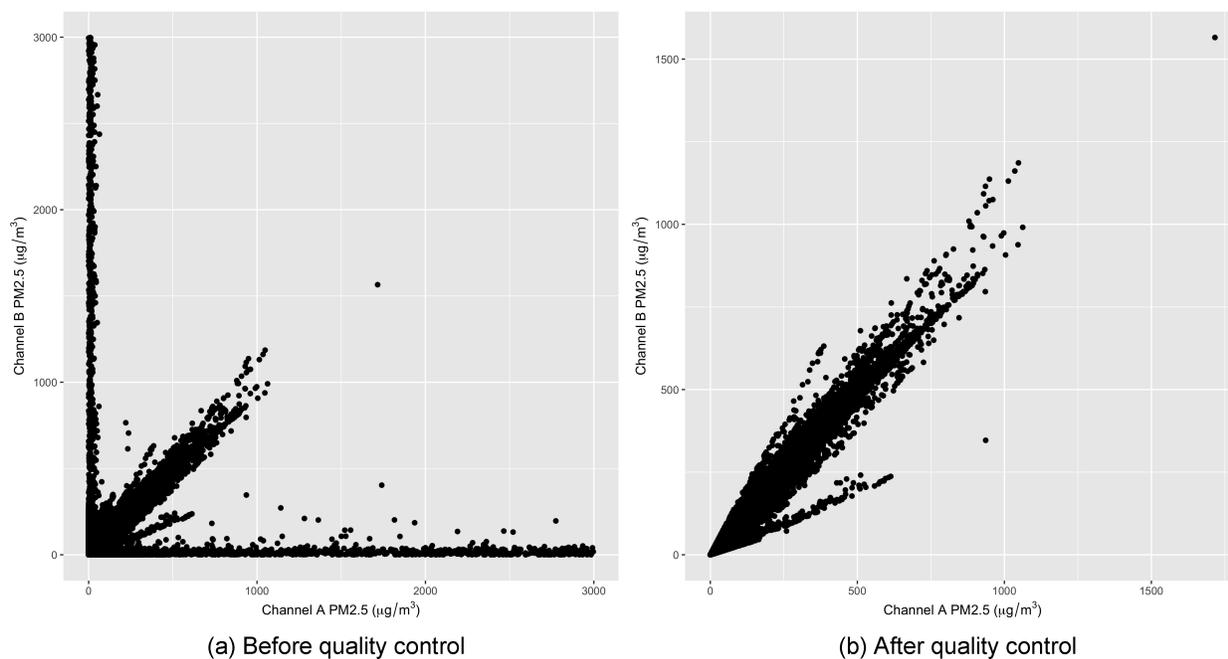


Figure S8: Scatter plots of PurpleAir dual-channel hourly measurements (a) before and (b) after removing the 5% largest absolute percentage biases (APBs).

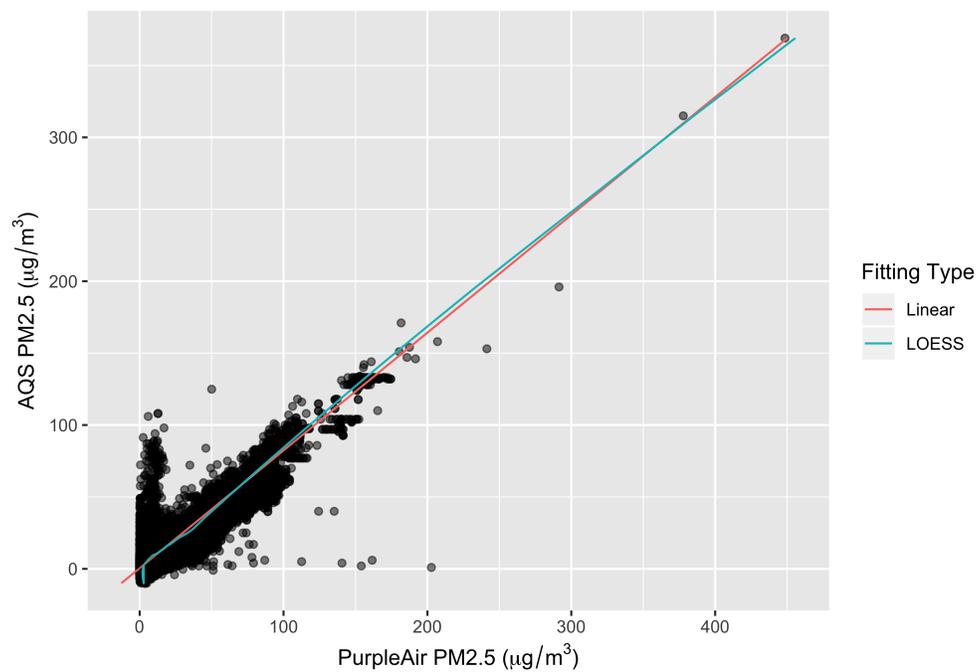


Figure S9: LOWESS (green) and linear (red) fitting curves of the paired AQS/PurpleAir hourly measurements (black scatters).

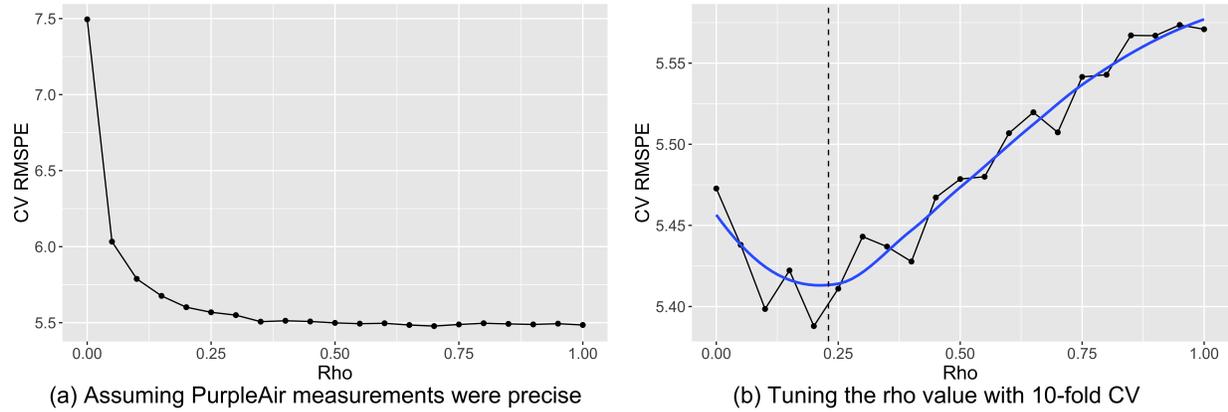


Figure S10: (a): The trend of the 10-fold CV RMSPE with ρ within a range of $[0, 1]$ when assuming PurpleAir measurements were precise. (b): The trend of the 10-fold CV RMSPE with ρ within $[0, 1]$ in the real case. The blue curve is the smoothed fitting curve, showing the minimum of CV RMSPE at a ρ value of ~ 0.23 .

7.3 Manuscript III Supplemental

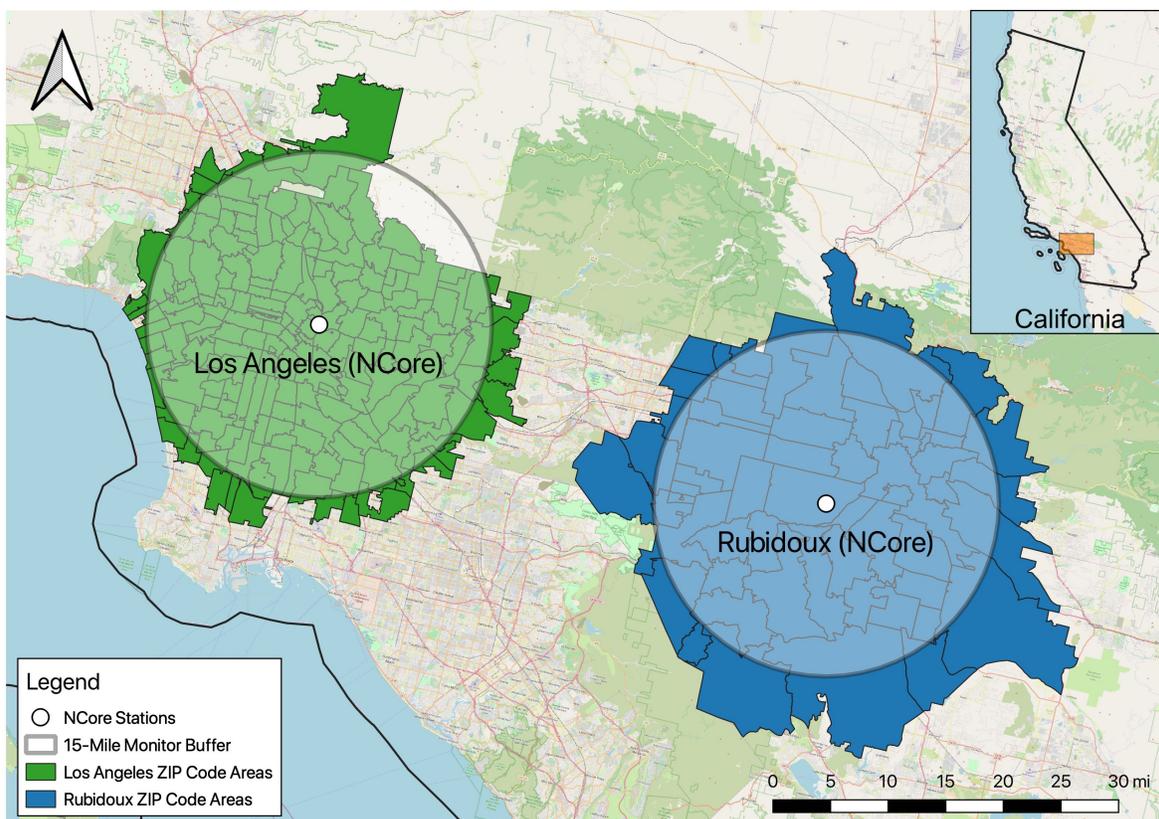


Figure S11: Study domains with Los Angeles and Rubidoux 15-mile monitor-buffers.

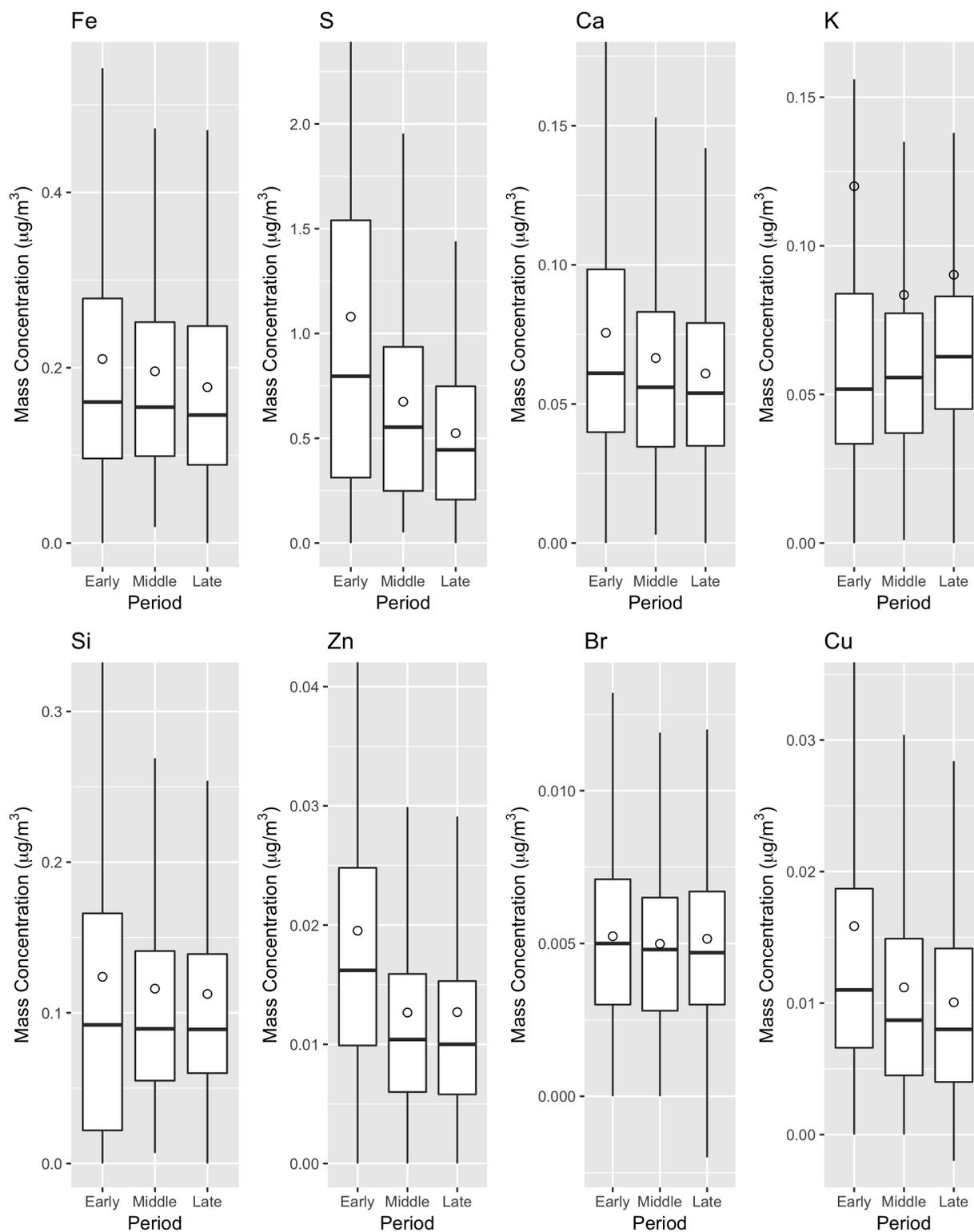


Figure S12: Box plots of mass concentrations of eight $\text{PM}_{2.5}$ trace components during the three time periods (EARLY: 2005 – 2008; MIDDLE: 2009 – 2012; LATE: 2013 – 2016). The measurements are the averages of two sub-domains.

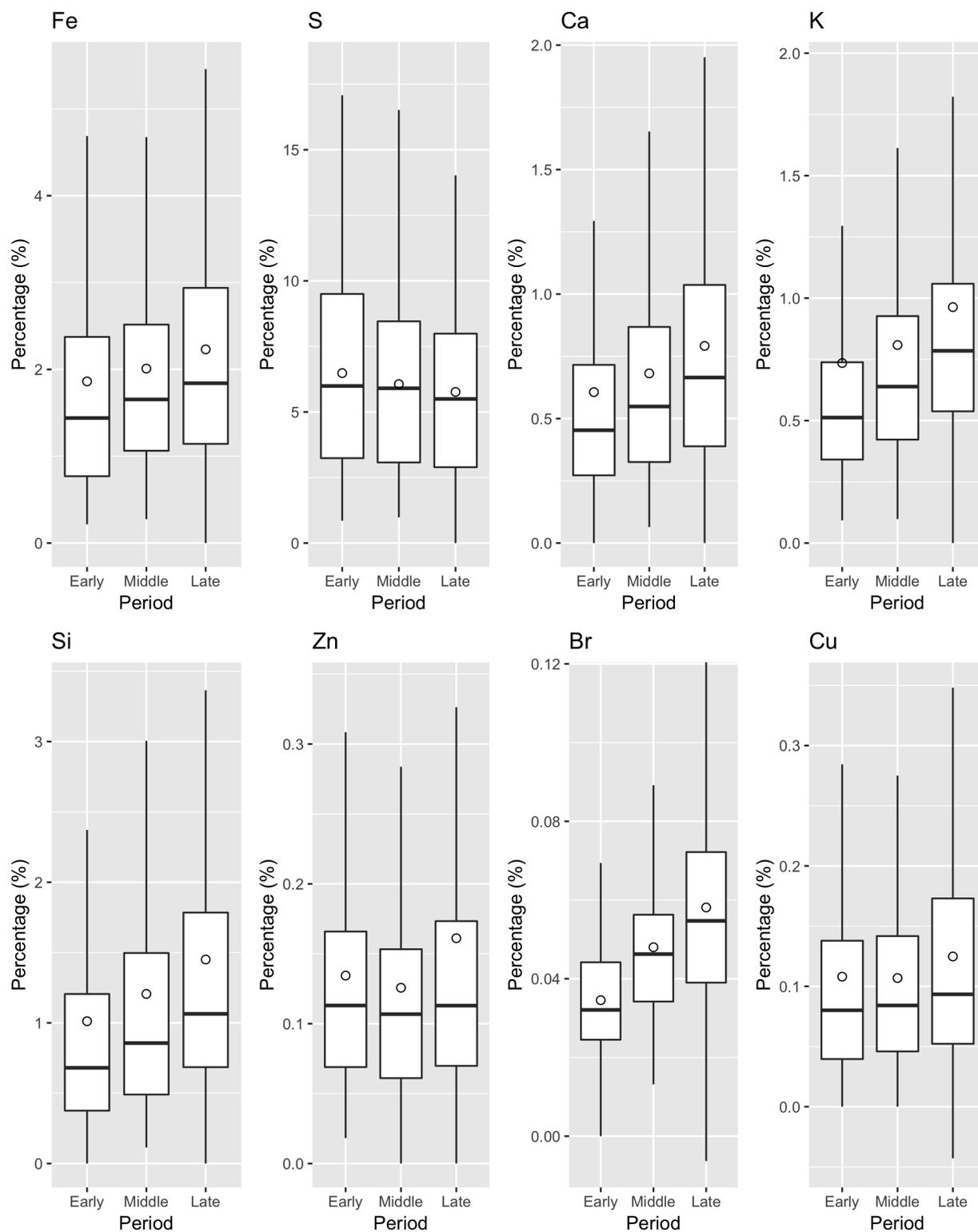


Figure S13: Box plots of percentages of eight $PM_{2.5}$ trace components in $PM_{2.5}$ total mass during the three time periods (EARLY: 2005 – 2008; MIDDLE: 2009 – 2012; LATE: 2013 – 2016). The measurements are the averages of two sub-domains.

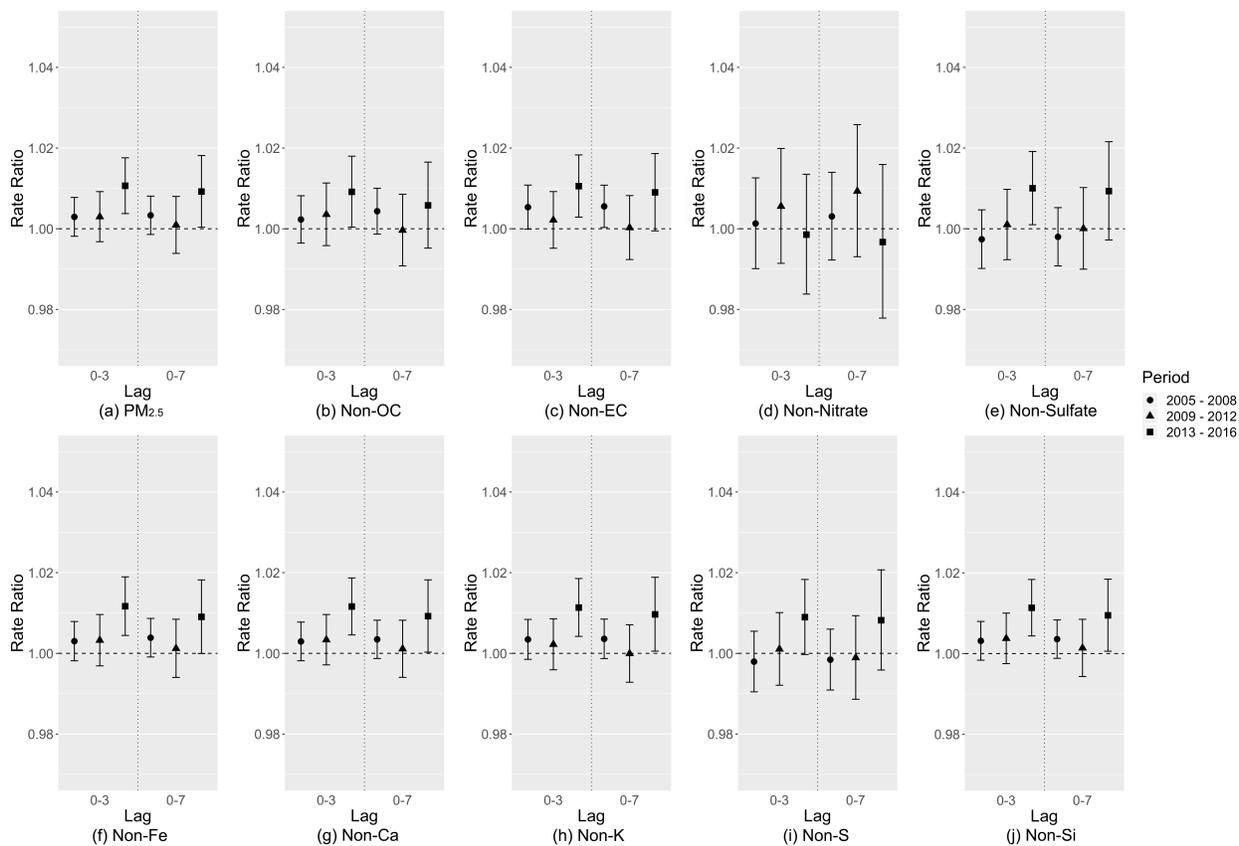


Figure S14: Relative risk estimates of CVD ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in the remaining $\text{PM}_{2.5}$ mass concentration. As a reference, (a) shows the risk estimates associated with $\text{PM}_{2.5}$ based on the moving average of exposure.

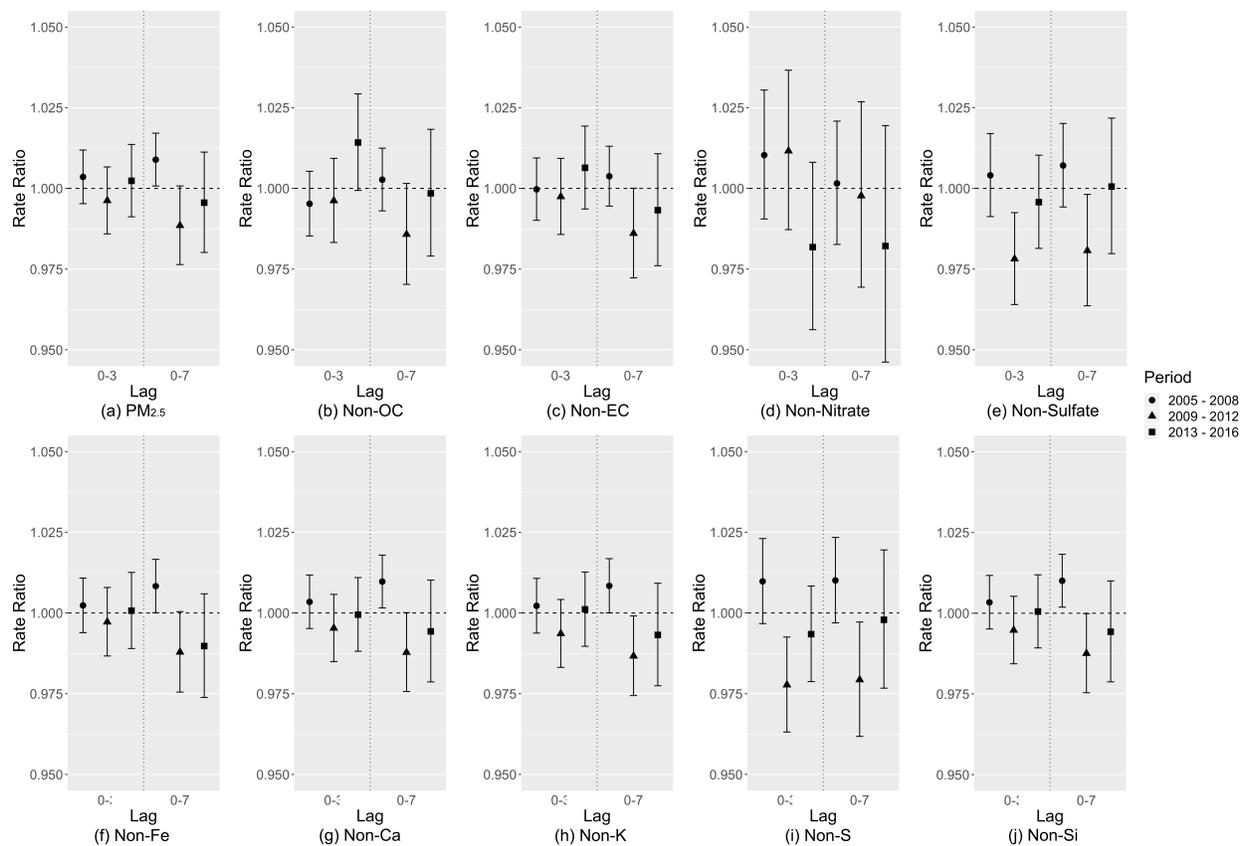


Figure S15: Relative risk estimates of asthma ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in the remaining $\text{PM}_{2.5}$ mass concentration. As a reference, (a) shows the risk estimates associated with $\text{PM}_{2.5}$ based on the moving average of exposure.

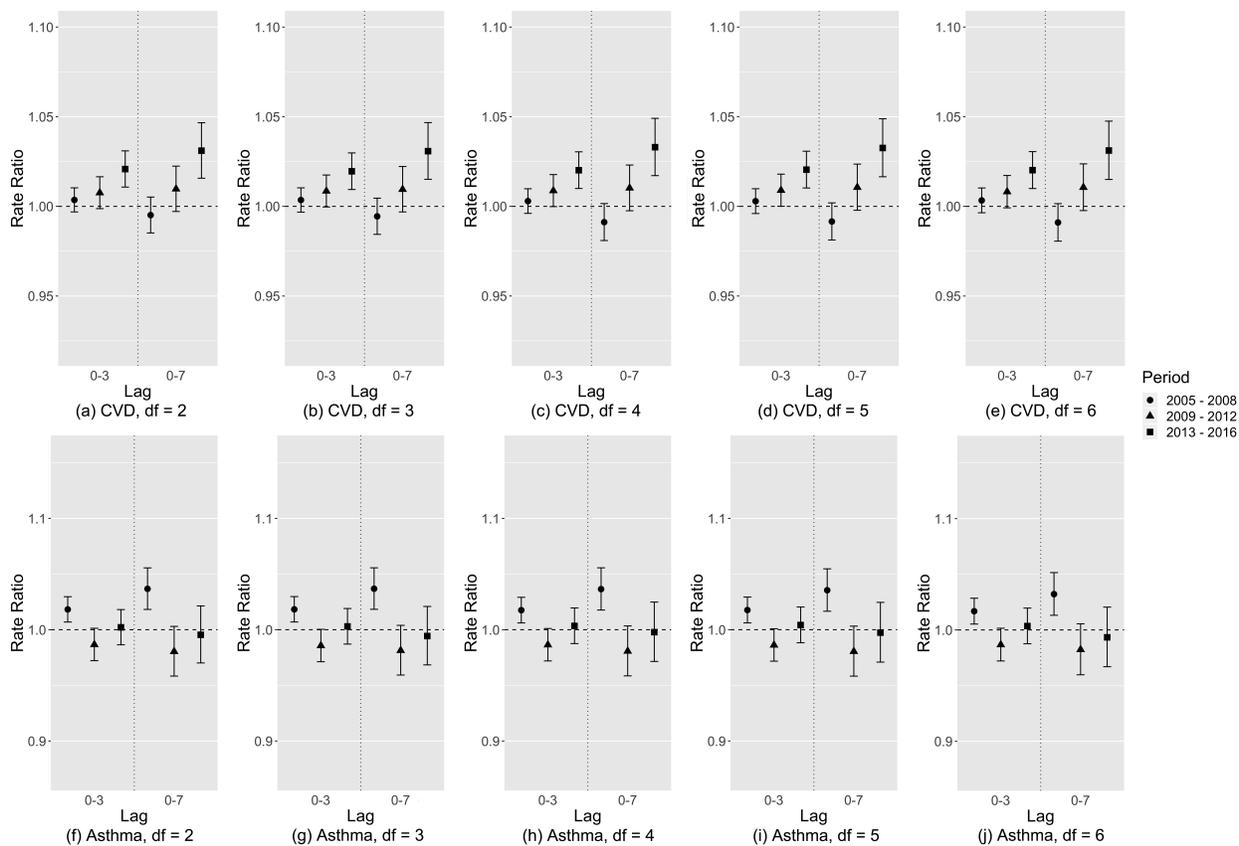


Figure S16: Relative risk estimates of CVD and asthma ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ concentration with different degrees of freedom ($df = 2 - 6$) of the cubic splines of daily maximum air temperature and mean dew-point temperature.

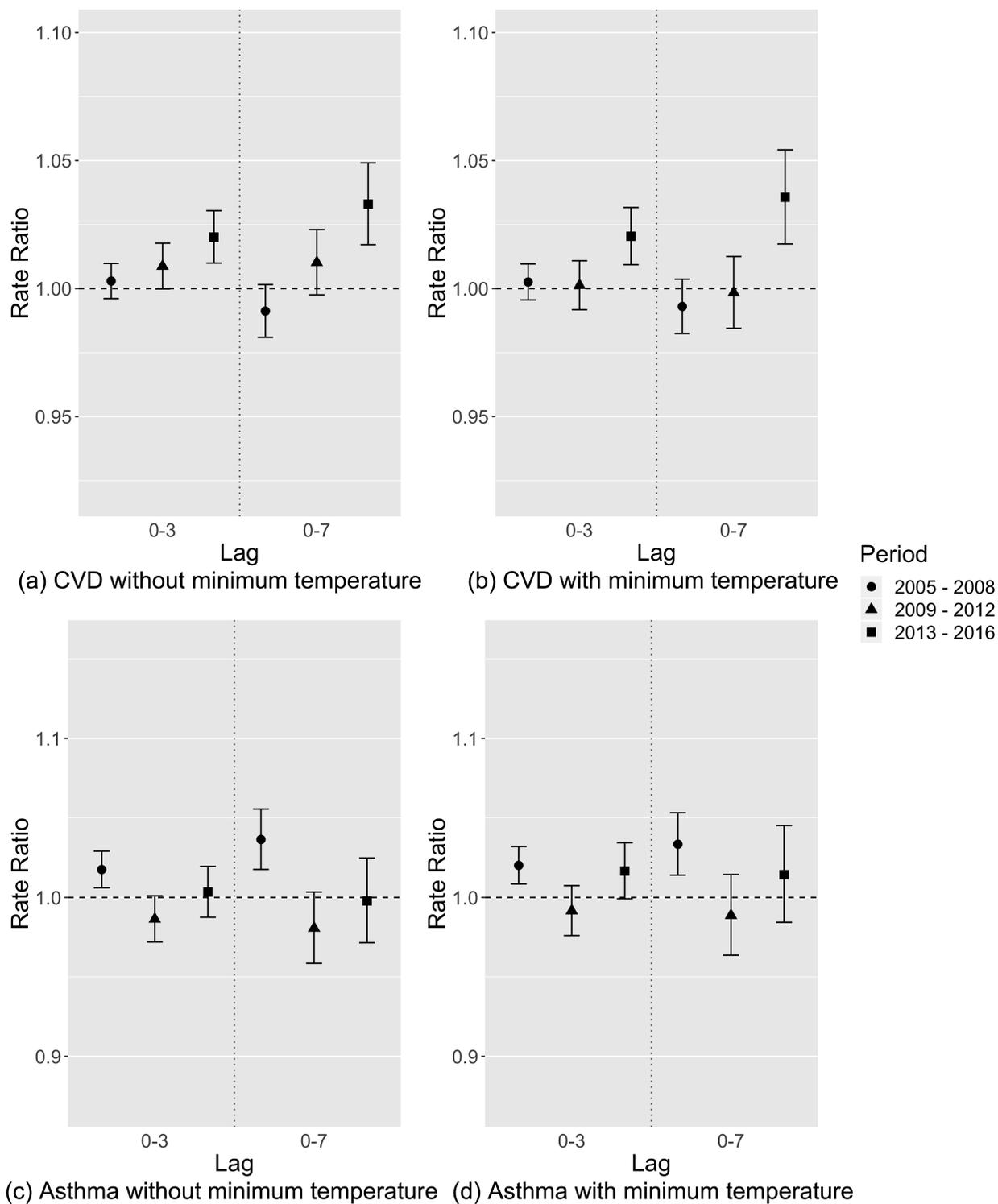


Figure S17: Relative risk estimates of CVD and asthma ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ concentration with and without the control for daily minimum air temperature.

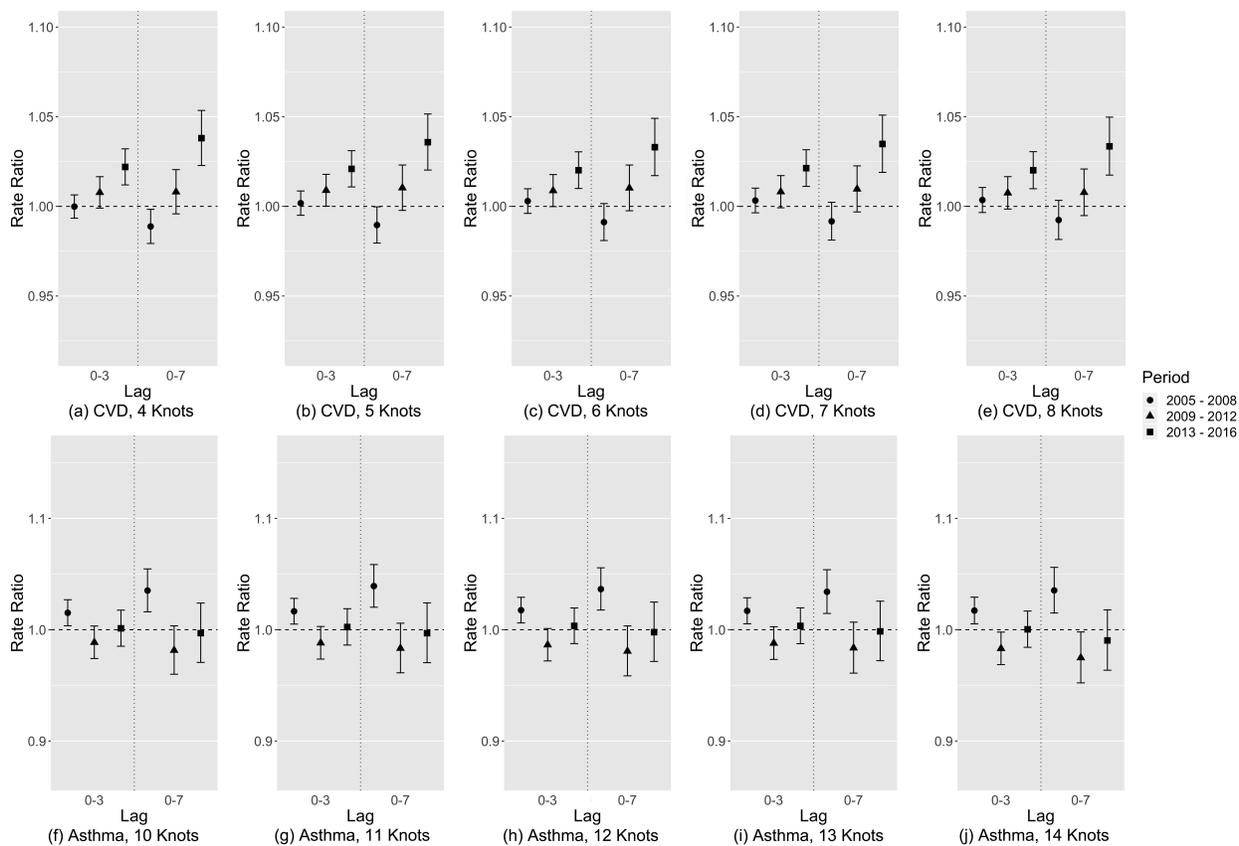


Figure S18: Relative risk estimates of CVD and asthma ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in $PM_{2.5}$ concentration with different annual knots in the time splines.

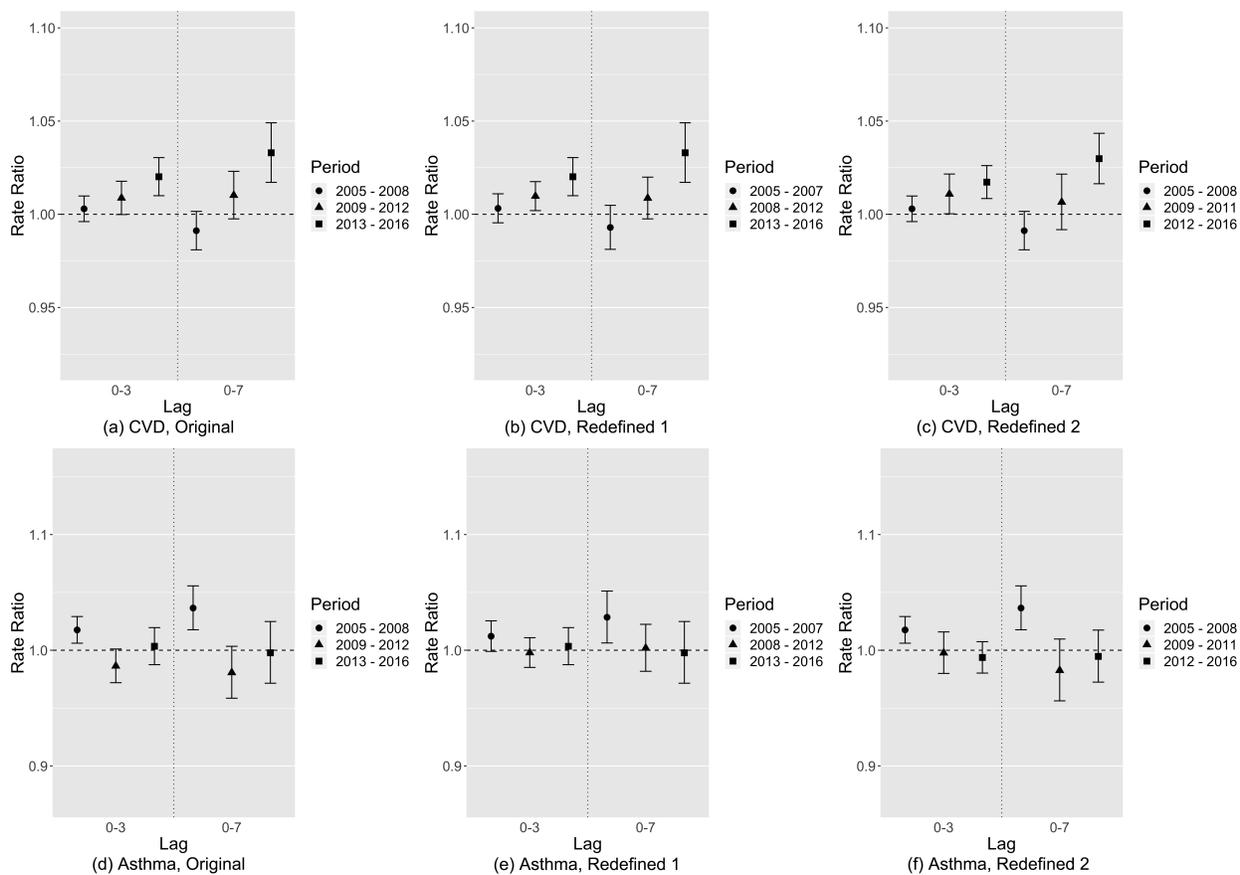


Figure S19: Relative risk estimates of CVD and asthma ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ concentration in different redefined time intervals (Redefined 1: moving the year of 2008 to the MIDDLE period; Redefined 2: moving the year of 2012 to the LATE period).

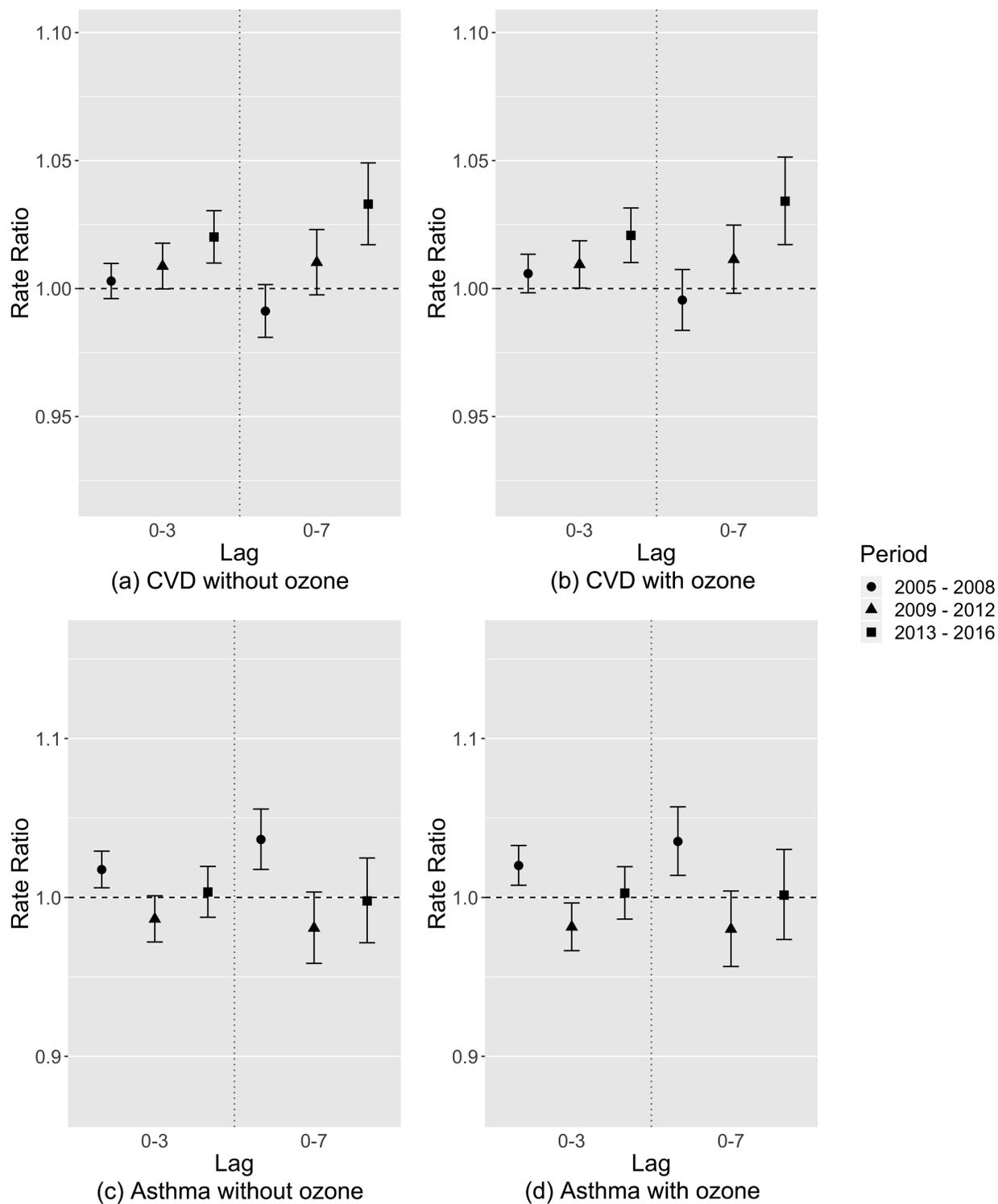


Figure S20: Relative risk estimates of CVD and asthma ED visits associated with each $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ concentration with and without the control for collocated short-term ozone exposure.

References

- [1] F. Lu, D. Q. Xu, Y. B. Cheng, S. X. Dong, C. Guo, X. Jiang, and X. Y. Zheng. “Systematic review and meta-analysis of the adverse health effects of ambient PM_{2.5} and PM₁₀ pollution in the Chinese population”. In: *Environmental Research* 136 (2015), pp. 196–204.
- [2] M. L. Bell, J. M. Samet, and F. Dominici. “Time-series studies of particulate matter”. In: *Annual Review of Public Health* 25 (2004), pp. 247–280.
- [3] M. L. Bell, A. Zanobetti, and F. Dominici. “Evidence on Vulnerability and Susceptibility to Health Risks Associated With Short-Term Exposure to Particulate Matter: A Systematic Review and Meta-Analysis”. In: *American Journal of Epidemiology* 178.6 (2013), pp. 865–876.
- [4] R. D. Brook, B. Franklin, W. Cascio, Y. L. Hong, G. Howard, M. Lipsett, R. Luepker, M. Mittleman, J. Samet, S. C. Smith, and I. Tager. “Air pollution and cardiovascular disease - A statement for healthcare professionals from the expert panel on population and prevention science of the American Heart Association”. In: *Circulation* 109.21 (2004), pp. 2655–2671.
- [5] Q. H. Sun, A. X. Wang, X. M. Jin, A. Natanzon, D. Duquaine, R. D. Brook, J. G. S. Aguinaldo, Z. A. Fayad, V. Fuster, M. Lippmann, L. C. Chen, and S. Rajagopalan. “Long-term air pollution exposure and acceleration of atherosclerosis and vascular inflammation in an animal model”. In: *Jama-Journal of the American Medical Association* 294.23 (2005), pp. 3003–3010.
- [6] A. M. Casillas, T. Hiura, N. Li, and A. E. Nel. “Enhancement of allergic inflammation by diesel exhaust particles: permissive role of reactive oxygen species”. In: *Annals of Allergy Asthma & Immunology* 83.6 (1999), pp. 624–629.
- [7] M. Guarneri and J. R. Balmes. “Outdoor air pollution and asthma”. In: *Lancet* 383.9928 (2014), pp. 1581–1592.
- [8] J. I. Halonen, T. Lanki, T. Yli-Tuomi, M. Kulmala, P. Tiittanen, and J. Pekkanen. “Urban air pollution, and asthma and COPD hospital emergency room visits”. In: *Thorax* 63.7 (2008), pp. 635–641.

- [9] A. Zanobetti, J. Schwartz, and D. Gold. “Are there sensitive subgroups for the effects of airborne particles?” In: *Environmental Health Perspectives* 108.9 (2000), pp. 841–845.
- [10] E. Emili, A. Lyapustin, Y. Wang, C. Popp, S. Korin, M. Zebisch, S. Wunderle, and M. Petitta. “High spatial resolution aerosol retrieval with MAIAC: Application to mountain regions”. In: *Journal of Geophysical Research: Atmospheres* 116.D23 (2011).
- [11] A. I. Lyapustin, Y. Wang, I. Laszlo, T. Hilker, F. G. Hall, P. J. Sellers, C. J. Tucker, and S. V. Korin. “Multi-angle implementation of atmospheric correction for MODIS (MAIAC): 3. Atmospheric correction”. In: *Remote Sensing of Environment* 127 (2012), pp. 385–393.
- [12] A. Lyapustin, J. Martonchik, Y. Wang, I. Laszlo, and S. Korin. “Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables”. In: *Journal of Geophysical Research* 116.D3 (2011).
- [13] C. J. Paciorek and Y. Liu. “Limitations of remotely sensed aerosol as a spatial proxy for fine particulate matter”. In: *Environ Health Perspect* 117.6 (2009), pp. 904–909.
- [14] X. Hu, L. A. Waller, A. Lyapustin, Y. Wang, M. Z. Al-Hamdan, W. L. Crosson, M. G. Estes, S. M. Estes, D. A. Quattrochi, S. J. Puttaswamy, and Y. Liu. “Estimating ground-level PM_{2.5} concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model”. In: *Remote Sensing of Environment* 140 (2014), pp. 220–232.
- [15] I. Kloog, A. A. Chudnovsky, A. C. Just, F. Nordio, P. Koutrakis, B. A. Coull, A. Lyapustin, Y. Wang, and J. Schwartz. “A New Hybrid Spatio-Temporal Model For Estimating Daily Multi-Year PM_{2.5} Concentrations Across Northeastern USA Using High Resolution Aerosol Optical Depth Data”. In: *Atmos Environ (1994)* 95 (2014), pp. 581–590.
- [16] I. Kloog, F. Nordio, B. A. Coull, and J. Schwartz. “Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM_{2.5} exposures in the Mid-Atlantic states”. In: *Environ Sci Technol* 46.21 (2012), pp. 11913–11921.

- [17] I. Kloog, P. Koutrakis, B. A. Coull, H. J. Lee, and J. Schwartz. “Assessing temporally and spatially resolved PM_{2.5} exposures for epidemiological studies using satellite aerosol optical depth measurements”. In: *Atmospheric Environment* 45.35 (2011), pp. 6267–6275.
- [18] Z. Ma, Y. Liu, Q. Zhao, M. Liu, Y. Zhou, and J. Bi. “Satellite-derived high resolution PM_{2.5} concentrations in Yangtze River Delta Region of China using improved linear mixed effects model”. In: *Atmospheric Environment* 133 (2016), pp. 156–164.
- [19] Q. Di, I. Kloog, P. Koutrakis, A. Lyapustin, Y. Wang, and J. Schwartz. “Assessing PM_{2.5} Exposures with High Spatiotemporal Resolution across the Continental United States”. In: *Environ Sci Technol* 50.9 (2016), pp. 4712–4721.
- [20] B. Zou, M. Wang, N. Wan, J. G. Wilson, X. Fang, and Y. Tang. “Spatial modeling of PM_{2.5} concentrations with a multifactorial radial basis function neural network”. In: *Environ Sci Pollut Res Int* 22.14 (2015), pp. 10395–10404.
- [21] C. Brokamp, R. Jandarov, M. Hossain, and P. Ryan. “Predicting Daily Urban Fine Particulate Matter Concentrations Using a Random Forest Model”. In: *Environ Sci Technol* 52.7 (2018), pp. 4173–4179.
- [22] X. Hu, J. H. Belle, X. Meng, A. Wildani, L. A. Waller, M. J. Strickland, and Y. Liu. “Estimating PM_{2.5} Concentrations in the Conterminous United States Using the Random Forest Approach”. In: *Environ Sci Technol* 51.12 (2017), pp. 6936–6944.
- [23] Q. Xiao, H. Zhang, M. Choi, S. Li, S. Kondragunta, J. Kim, B. Holben, R. C. Levy, and Y. Liu. “Evaluation of VIIRS, GOCI, and MODIS Collection 6 AOD retrievals against ground sunphotometer observations over East Asia”. In: *Atmospheric Chemistry and Physics* 16.3 (2016), pp. 1255–1269.
- [24] D. A. Chu. “Validation of MODIS aerosol optical depth retrieval over land”. In: *Geophysical Research Letters* 29.12 (2002).
- [25] K. Alam, R. Khan, T. Blaschke, and A. Mukhtiar. “Variability of aerosol optical depth and their impact on cloud properties in Pakistan”. In: *Journal of Atmospheric and Solar-Terrestrial Physics* 107 (2014), pp. 104–112.

- [26] J. H. Belle, H. H. Chang, Y. Wang, X. Hu, A. Lyapustin, and Y. Liu. “The Potential Impact of Satellite-Retrieved Cloud Parameters on Ground-Level PM_{2.5} Mass and Composition”. In: *Int J Environ Res Public Health* 14.10 (2017).
- [27] N. Kang, K. R. Kumar, Y. Yin, Y. Diao, and X. Yu. “Correlation Analysis between AOD and Cloud Parameters to Study Their Relationship over China Using MODIS Data (2003–2013): Impact on Cloud Formation and Climate Change”. In: *Aerosol and Air Quality Research* 15.3 (2015), pp. 958–973.
- [28] G. Myhre, F. Stordal, M. Johnsrud, Y. J. Kaufman, D. Rosenfeld, T. Storelvmo, J. E. Kristjansson, T. K. Berntsen, A. Myhre, and I. S. A. Isaksen. “Aerosol-cloud interaction inferred from MODIS satellite data and global aerosol models”. In: *Atmospheric Chemistry and Physics* 7.12 (2007), pp. 3081–3101.
- [29] C. Yu, L. Di Girolamo, L. Chen, X. Zhang, and Y. Liu. “Statistical evaluation of the feasibility of satellite-retrieved cloud parameters as indicators of PM_{2.5} levels”. In: *J Expo Sci Environ Epidemiol* 25.5 (2015), pp. 457–466.
- [30] S. Li, L. Chen, J. Tao, D. Han, Z. Wang, L. Su, M. Fan, and C. Yu. “Retrieval of aerosol optical depth over bright targets in the urban areas of North China during winter”. In: *Science China Earth Sciences* 55.9 (2012), pp. 1545–1553.
- [31] Q. Xiao, Y. Wang, H. H. Chang, X. Meng, G. Geng, A. Lyapustin, and Y. Liu. “Full-coverage high-resolution daily PM_{2.5} estimation using MAIAC AOD in the Yangtze River Delta of China”. In: *Remote Sensing of Environment* 199 (2017), pp. 437–446.
- [32] L. W. A. Chen, J. G. Watson, J. C. Chow, M. C. Green, D. Inouye, and K. Dick. “Wintertime particulate pollution episodes in an urban valley of the Western US: a case study”. In: *Atmospheric Chemistry and Physics* 12.21 (2012), pp. 10051–10064.
- [33] M. C. Green, J. C. Chow, J. G. Watson, K. Dick, and D. Inouye. “Effects of Snow Cover and Atmospheric Stability on Winter PM_{2.5} Concentrations in Western U.S. Valleys”. In: *Journal of Applied Meteorology and Climatology* 54.6 (2015), pp. 1191–1201.
- [34] C. D. Whiteman, S. W. Hoch, J. D. Horel, and A. Charland. “Relationship between particulate air pollution and meteorological variables in Utah’s Salt Lake Valley”. In: *Atmospheric Environment* 94 (2014), pp. 742–753.

- [35] H. J. Lee, Y. Liu, B. A. Coull, J. Schwartz, and P. Koutrakis. “A novel calibration approach of MODIS AOD data to predict PM_{2.5} concentrations”. In: *Atmospheric Chemistry and Physics* 11.15 (2011), pp. 7991–8002.
- [36] S. C. Wilson, J. Morrow-Tesch, D. C. Straus, J. D. Cooley, W. C. Wong, F. M. Mitlohner, and J. J. McGlone. “Airborne microbial flora in a cattle feedlot”. In: *Appl Environ Microbiol* 68.7 (2002), pp. 3238–3242.
- [37] C. A. Noble, R. W. Vanderpool, T. M. Peters, F. F. McElroy, D. B. Gemmill, and R. W. Wiener. “Federal reference and equivalent methods for measuring fine particulate matter”. In: *Aerosol Science & Technology* 34.5 (2001), pp. 457–464.
- [38] D. M. Broday and C. Citi-Sense Project. “Wireless Distributed Environmental Sensor Networks for Air Pollution Measurement-The Promise and the Current Reality”. In: *Sensors (Basel)* 17.10 (2017), p. 2263.
- [39] T. Cao and J. E. Thompson. “Portable, Ambient PM_{2.5} Sensor for Human and/or Animal Exposure Studies”. In: *Analytical Letters* 50.4 (2017), pp. 712–723.
- [40] N. Castell, F. R. Dauge, P. Schneider, M. Vogt, U. Lerner, B. Fishbain, D. Broday, and A. Bartonova. “Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?” In: *Environment international* 99 (2017), pp. 293–302.
- [41] D. M. Holstius, A. Pillarisetti, K. Smith, and E. Seto. “Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California”. In: *Atmospheric Measurement Techniques* 7.4 (2014), pp. 1121–1131.
- [42] K. E. Kelly, J. Whitaker, A. Petty, C. Widmer, A. Dybwad, D. Sleeth, R. Martin, and A. Butterfield. “Ambient and laboratory evaluation of a low-cost particulate matter sensor”. In: *Environ Pollut* 221 (2017), pp. 491–500.
- [43] Y. Wang, J. Li, H. Jing, Q. Zhang, J. Jiang, and P. Biswas. “Laboratory evaluation and calibration of three low-cost particle sensors for particulate matter measurement”. In: *Aerosol Science and Technology* 49.11 (2015), pp. 1063–1077.
- [44] C. C. Cho, W. Y. Hsieh, C. H. Tsai, C. Y. Chen, H. F. Chang, and C. S. Lin. “In Vitro and In Vivo Experimental Studies of PM_{2.5} on Disease Progression”. In: *International Journal of Environmental Research and Public Health* 15.7 (2018), p. 1380.

- [45] Y. J. Zou, C. Y. Jin, Y. Su, J. R. Li, and B. S. Zhu. “Water soluble and insoluble components of urban PM_{2.5} and their cytotoxic effects on epithelial cells (A549) in vitro”. In: *Environmental Pollution* 212 (2016), pp. 627–635.
- [46] M. L. Bell, F. Dominici, K. Ebisu, S. L. Zeger, and J. M. Samet. “Spatial and temporal variation in PM_{2.5} chemical composition in the United States for health effects studies”. In: *Environmental Health Perspectives* 115.7 (2007), pp. 989–995.
- [47] M. L. Bell, K. Ebisu, R. D. Peng, J. M. Samet, and F. Dominici. “Hospital Admissions and Chemical Composition of Fine Particle Air Pollution”. In: *American Journal of Respiratory and Critical Care Medicine* 179.12 (2009), pp. 1115–1120.
- [48] F. Dominici, R. D. Peng, K. Ebisu, S. L. Zeger, J. M. Samet, and M. L. Bell. “Does the effect of PM₁₀ on mortality depend on PM nickel and vanadium content? A re-analysis of the NMMAPS data”. In: *Environmental Health Perspectives* 115.12 (2007), pp. 1701–1703.
- [49] M. Lippmann, K. Ito, J. S. Hwang, P. Maciejczyk, and L. C. Chen. “Cardiovascular effects of nickel in ambient air”. In: *Environmental Health Perspectives* 114.11 (2006), pp. 1662–1669.
- [50] D. N. Ye, M. Klein, J. A. Mulholland, A. G. Russell, R. Weber, E. S. Edgerton, H. H. Chang, J. A. Sarnat, P. E. Tolbert, and S. E. Sarnat. “Estimating Acute Cardiovascular Effects of Ambient PM_{2.5} Metals”. In: *Environmental Health Perspectives* 126.2 (2018).
- [51] L. K. Baxter, R. M. Duvall, and J. Sacks. “Examining the effects of air pollution composition on within region differences in PM_{2.5} mortality risk estimates”. In: *Journal of Exposure Science & Environmental Epidemiology* 23.5 (2013), pp. 457–465.
- [52] F. Dominici, R. D. Peng, M. L. Bell, L. Pham, A. McDermott, S. L. Zeger, and J. M. Samet. “Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases”. In: *Jama-Journal of the American Medical Association* 295.10 (2006), pp. 1127–1134.
- [53] D. P. Croft, W. J. Zhang, S. Lin, S. W. Thurston, P. K. Hopke, M. Masiol, S. Squizzato, E. van Wijngaarden, M. J. Utell, and D. Q. Rich. “The Association between Respiratory Infection and Air Pollution in the Setting of Air Quality Policy and Eco-

- conomic Change”. In: *Annals of the American Thoracic Society* 16.3 (2019), pp. 321–330.
- [54] P. K. Hopke, D. Croft, W. J. Zhang, S. Lin, M. Masiol, S. Squizzato, S. W. Thurston, E. van Wijngaarden, M. J. Utell, and D. Q. Rich. “Changes in the acute response of respiratory diseases to PM_{2.5} in New York State from 2005 to 2016”. In: *Science of the Total Environment* 677 (2019), pp. 328–339.
- [55] W. J. Zhang, S. Lin, P. K. Hopke, S. W. Thurston, E. van Wijngaarden, D. Croft, S. Squizzato, M. Masiol, and D. Q. Rich. “Triggering of cardiovascular hospital admissions by fine particle concentrations in New York state: Before, during, and after implementation of multiple environmental policies and a recession”. In: *Environmental Pollution* 242 (2018), pp. 1404–1416.
- [56] S. Squizzato, M. Masiol, D. Q. Rich, and P. K. Hopke. “PM_{2.5} and gaseous pollutants in New York State during 2005–2016: Spatial variability, temporal trends, and economic influences”. In: *Atmospheric Environment* 183 (2018), pp. 209–224.
- [57] J. Y. Abrams, M. Klein, L. R. F. Henneman, S. E. Sarnat, H. H. Chang, M. J. Strickland, J. A. Mulholland, A. G. Russell, and P. E. Tolbert. “Impact of air pollution control policies on cardiorespiratory emergency department visits, Atlanta, GA, 1999–2013”. In: *Environment International* 126 (2019), pp. 627–634.
- [58] H. Kim, H. Kim, Y. H. Park, and J. T. Lee. “Assessment of temporal variation for the risk of particulate matters on asthma hospitalization”. In: *Environmental Research* 156 (2017), pp. 542–550.
- [59] S. Bose, N. N. Hansel, E. S. Tonorezos, D. L. Williams, A. Bilderback, P. N. Breysse, G. B. Diette, and M. C. McCormack. “Indoor Particulate Matter Associated with Systemic Inflammation in COPD”. In: *Journal of Environmental Protection* 06.05 (2015), pp. 566–572.
- [60] R. T. Burnett, 3. Pope C. A., M. Ezzati, C. Olives, S. S. Lim, S. Mehta, H. H. Shin, G. Singh, B. Hubbell, M. Brauer, H. R. Anderson, K. R. Smith, J. R. Balmes, N. G. Bruce, H. Kan, F. Laden, A. Pruss-Ustun, M. C. Turner, S. M. Gapstur, W. R. Diver, and A. Cohen. “An integrated risk function for estimating the global burden of

- disease attributable to ambient fine particulate matter exposure”. In: *Environ Health Perspect* 122.4 (2014), pp. 397–403.
- [61] J. Madrigano, I. Kloog, R. Goldberg, B. A. Coull, M. A. Mittleman, and J. Schwartz. “Long-term exposure to PM_{2.5} and incidence of acute myocardial infarction”. In: *Environ Health Perspect* 121.2 (2013), pp. 192–196.
- [62] M. Sorek-Hamer, A. C. Just, and I. Kloog. “Satellite remote sensing in epidemiological studies”. In: *Curr Opin Pediatr* 28.2 (2016), pp. 228–234.
- [63] Z. Ma, X. Hu, A. M. Sayer, R. Levy, Q. Zhang, Y. Xue, S. Tong, J. Bi, L. Huang, and Y. Liu. “Satellite-Based Spatiotemporal Trends in PM_{2.5} Concentrations: China, 2004-2013”. In: *Environ Health Perspect* 124.2 (2016), pp. 184–192.
- [64] P. M. Fine, G. R. Cass, and B. R. Simoneit. “Organic compounds in biomass smoke from residential wood combustion: Emissions characterization at a continental scale”. In: *Journal of Geophysical Research: Atmospheres* 107.D21 (2002).
- [65] A. Reff, P. V. Bhave, H. Simon, T. G. Pace, G. A. Pouliot, J. D. Mobley, and M. Houyoux. “Emissions inventory of PM_{2.5} trace elements across the United States”. In: *Environ Sci Technol* 43.15 (2009), pp. 5790–5796.
- [66] J. G. Su, G. Allen, P. J. Miller, and M. Brauer. “Spatial modeling of residential woodsmoke across a non-urban upstate New York region”. In: *Air Quality, Atmosphere & Health* 6.1 (2013), pp. 85–94.
- [67] Y. Liu, C. J. Paciorek, and P. Koutrakis. “Estimating regional spatial and temporal variability of PM(2.5) concentrations using satellite data, meteorology, and land use information”. In: *Environ Health Perspect* 117.6 (2009), pp. 886–892.
- [68] P. Gupta and S. A. Christopher. “Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach”. In: *Journal of Geophysical Research* 114.D20 (2009).
- [69] C. Brokamp, R. Jandarov, M. Hossain, and P. Ryan. “Predicting Daily Urban Fine Particulate Matter Concentrations Using a Random Forest Model”. In: *Environmental science & technology* (2018).
- [70] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*. Vol. 1. Springer series in statistics New York, 2001.

- [71] N. Hsu, M. Jeong, C. Bettenhausen, A. Sayer, R. Hansell, C. Seftor, J. Huang, and S. Tsay. “Enhanced Deep Blue aerosol retrieval algorithm: The second generation”. In: *Journal of Geophysical Research: Atmospheres* 118.16 (2013), pp. 9296–9315.
- [72] R. C. Levy, L. A. Remer, S. Mattoo, E. F. Vermote, and Y. J. Kaufman. “Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance”. In: *Journal of Geophysical Research: Atmospheres* 112.D13 (2007).
- [73] A. van Donkelaar, R. V. Martin, R. C. Levy, A. M. da Silva, M. Krzyzanowski, N. E. Chubarova, E. Semutnikova, and A. J. Cohen. “Satellite-based estimates of ground-level fine particulate matter during extreme events: A case study of the Moscow fires in 2010”. In: *Atmospheric Environment* 45.34 (2011), pp. 6225–6232.
- [74] D. A. Robinson, K. F. Dewey, and R. R. Heim. “Global Snow Cover Monitoring: An Update”. In: *Bulletin of the American Meteorological Society* 74.9 (1993), pp. 1689–1696.
- [75] V. Martins, A. Lyapustin, L. de Carvalho, C. Barbosa, and E. Novo. “Validation of high-resolution MAIAC aerosol product over South America”. In: *Journal of Geophysical Research: Atmospheres* 122.14 (2017), pp. 7537–7559.
- [76] S. Platnick, S. Ackerman, M. King, K. Meyer, W. Menzel, R. Holz, B. Baum, and P. Yang. “MODIS atmosphere L2 cloud product (06_L2)”. In: *NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA* 10 (2015).
- [77] S. Ackerman, R. Holz, R. Frey, E. Eloranta, B. Maddux, and M. McGill. “Cloud detection with MODIS. Part II: validation”. In: *Journal of Atmospheric and Oceanic Technology* 25.7 (2008), pp. 1073–1086.
- [78] D. K. Hall and G. A. Riggs. “Normalized-difference snow index (NDSI)”. In: *Encyclopedia of snow, ice and glaciers*. Springer, 2011, pp. 779–780.
- [79] V. V. Salomonson and I. Appel. “Development of the Aqua MODIS NDSI fractional snow cover algorithm and validation results”. In: *IEEE Transactions on geoscience and remote sensing* 44.7 (2006), pp. 1747–1756.

- [80] F. Mesinger, G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jović, J. Woollen, E. Rogers, and E. H. Berbery. “North American regional reanalysis”. In: *Bulletin of the American Meteorological Society* 87.3 (2006), pp. 343–360.
- [81] K. E. Mitchell, D. Lohmann, P. R. Houser, E. F. Wood, J. C. Schaake, A. Robock, B. A. Cosgrove, J. Sheffield, Q. Duan, and L. Luo. “The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system”. In: *Journal of Geophysical Research: Atmospheres* 109.D7 (2004).
- [82] P. M. Bartier and C. P. Keller. “Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW)”. In: *Computers & Geosciences* 22.7 (1996), pp. 795–799.
- [83] L. Breiman. “Random forests”. In: *Machine Learning* 45.1 (2001), pp. 5–32.
- [84] A. Liaw and M. Wiener. “Classification and regression by randomForest”. In: *R news* 2.3 (2002), pp. 18–22.
- [85] L. Breiman. “Manual on setting up, using, and understanding random forests v3.1”. In: *Statistics Department University of California Berkeley, CA, USA* 1 (2002).
- [86] A. C. Just, R. O. Wright, J. Schwartz, B. A. Coull, A. A. Baccarelli, M. M. Tellez-Rojo, E. Moody, Y. Wang, A. Lyapustin, and I. Kloog. “Using High-Resolution Satellite Aerosol Optical Depth To Estimate Daily PM_{2.5} Geographical Distribution in Mexico City”. In: *Environ Sci Technol* 49.14 (2015), pp. 8576–8584.
- [87] R. R. Gillies, S.-Y. Wang, and M. R. Booth. “Atmospheric Scale Interaction on Wintertime Intermountain West Low-Level Inversions”. In: *Weather and Forecasting* 25.4 (2010), pp. 1196–1210.
- [88] G. D. Silcox, K. E. Kelly, E. T. Crosman, C. D. Whiteman, and B. L. Allen. “Winter-time PM_{2.5} concentrations during persistent, multi-day cold-air pools in a mountain valley”. In: *Atmospheric Environment* 46 (2012), pp. 17–24.
- [89] S.-Y. Wang, R. R. Gillies, R. Martin, R. E. Davies, and M. R. Booth. “Connecting Subseasonal Movements of the Winter Mean Ridge in Western North America to Inversion Climatology in Cache Valley, Utah”. In: *Journal of Applied Meteorology and Climatology* 51.3 (2012), pp. 617–627.

- [90] R. D. Brook, S. Rajagopalan, 3. Pope C. A., J. R. Brook, A. Bhatnagar, A. V. Diez-Roux, F. Holguin, Y. Hong, R. V. Luepker, M. A. Mittleman, A. Peters, D. Siscovick, J. Smith S. C., L. Whitsel, J. D. Kaufman, E. American Heart Association Council on, C. o. t. K. i. C. D. Prevention, P. A. Council on Nutrition, and Metabolism. “Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association”. In: *Circulation* 121.21 (2010), pp. 2331–2378.
- [91] C. A. Pope and D. W. Dockery. “Health effects of fine particulate air pollution: lines that connect”. In: *J Air Waste Manag Assoc* 56.6 (2006), pp. 709–742.
- [92] A. J. Cohen, M. Brauer, R. Burnett, H. R. Anderson, J. Frostad, K. Estep, K. Balakrishnan, B. Brunekreef, L. Dandona, R. Dandona, V. Feigin, G. Freedman, B. Hubbell, A. Jobling, H. Kan, L. Knibbs, Y. Liu, R. Martin, L. Morawska, 3. Pope C. A., H. Shin, K. Straif, G. Shaddick, M. Thomas, R. van Dingenen, A. van Donkelaar, T. Vos, C. J. L. Murray, and M. H. Forouzanfar. “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”. In: *Lancet* 389.10082 (2017), pp. 1907–1918.
- [93] C. R. O’Lenick, A. Winquist, J. A. Mulholland, M. D. Friberg, H. H. Chang, M. R. Kramer, L. A. Darrow, and S. E. Sarnat. “Assessment of neighbourhood-level socioeconomic status as a modifier of air pollution-asthma associations among children in Atlanta”. In: *J Epidemiol Community Health* 71.2 (2017), pp. 129–136.
- [94] S. E. Sarnat, A. Winquist, J. J. Schauer, J. R. Turner, and J. A. Sarnat. “Fine particulate matter components and emergency department visits for cardiovascular and respiratory diseases in the St. Louis, Missouri-Illinois, metropolitan area”. In: *Environ Health Perspect* 123.5 (2015), pp. 437–444.
- [95] Y. Zhao, S. Wang, L. Duan, Y. Lei, P. Cao, and J. Hao. “Primary air pollutant emissions of coal-fired power plants in China: Current status and future prediction”. In: *Atmospheric Environment* 42.36 (2008), pp. 8442–8452.
- [96] J. Bi, J. H. Belle, Y. Wang, A. I. Lyapustin, A. Wildani, and Y. Liu. “Impacts of snow and cloud covers on satellite-derived PM_{2.5} levels”. In: *Remote Sens Environ* 221 (2019), pp. 665–674.

- [97] E. G. Snyder, T. H. Watkins, P. A. Solomon, E. D. Thoma, R. W. Williams, G. S. Hagler, D. Shelow, D. A. Hindin, V. J. Kilaru, and P. W. Preuss. “The changing paradigm of air pollution monitoring”. In: *Environ Sci Technol* 47.20 (2013), pp. 11369–11377.
- [98] J. Al-Saadi, J. Szykman, R. B. Pierce, C. Kittaka, D. Neil, D. A. Chu, L. Remer, L. Gumley, E. Prins, L. Weinstock, C. MacDonald, R. Wayland, F. Dimmick, and J. Fishman. “Improving National Air Quality Forecasts with Satellite Aerosol Observations”. In: *Bulletin of the American Meteorological Society* 86.9 (2005), pp. 1249–1262.
- [99] P. Gupta, P. Doraiswamy, R. Levy, O. Pikelnaya, J. Maibach, B. Feenstra, A. Polidori, F. Kiros, and K. C. Mills. “Impact of California Fires on Local and Regional Air Quality: The Role of a Low-Cost Sensor Network and Satellite Observations”. In: *Geohealth* 2.6 (2018), pp. 172–181.
- [100] L. Morawska, P. K. Thai, X. Liu, A. Asumadu-Sakyi, G. Ayoko, A. Bartonova, A. Bedini, F. Chai, B. Christensen, and M. Dunbabin. “Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone?” In: *Environment international* 116 (2018), pp. 286–299.
- [101] W. Jiao, G. Hagler, R. Williams, R. Sharpe, R. Brown, D. Garver, R. Judge, M. Caudill, J. Rickard, and M. Davis. “Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States”. In: *Atmospheric Measurement Techniques* 9.11 (2016), pp. 5281–5292.
- [102] A. L. Clements, W. G. Griswold, A. Rs, J. E. Johnston, M. M. Herting, J. Thorson, A. Collier-Oxandale, and M. Hannigan. “Low-Cost Air Quality Monitoring Tools: From Research to Practice (A Workshop Summary)”. In: *Sensors (Basel)* 17.11 (2017), p. 2478.
- [103] N. Masson, R. Piedrahita, and M. Hannigan. “Quantification Method for Electrolytic Sensors in Long-Term Monitoring of Ambient Air Quality”. In: *Sensors (Basel)* 15.10 (2015), pp. 27283–27302.

- [104] E. Austin, I. Novosselov, E. Seto, and M. G. Yost. “Laboratory evaluation of the Shinyei PPD42NS low-cost particulate matter sensor”. In: *PloS one* 10.9 (2015), e0137789.
- [105] G. N. Carvlin, H. Lugo, L. Olmedo, E. Bejarano, A. Wilkie, D. Meltzer, M. Wong, G. King, A. Northcross, M. Jerrett, P. B. English, D. Hammond, and E. Seto. “Development and field validation of a community-engaged particulate matter air quality monitoring network in Imperial, California, USA”. In: *J Air Waste Manag Assoc* 67.12 (2017), pp. 1342–1352.
- [106] M. Gao, J. Cao, and E. Seto. “A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM_{2.5} in Xi’an, China”. In: *Environ Pollut* 199 (2015), pp. 56–65.
- [107] H.-Y. Liu, P. Schneider, R. Haugen, and M. Vogt. “Performance Assessment of a Low-Cost PM_{2.5} Sensor for a near Four-Month Period in Oslo, Norway”. In: *Atmosphere* 10.2 (2019), p. 41.
- [108] A. Ripoll, M. Viana, M. Padrosa, X. Querol, A. Minutolo, K. M. Hou, J. M. Barcelo-Ordinas, and J. Garcia-Vidal. “Testing the performance of sensors for ozone pollution monitoring in a citizen science approach”. In: *Science of the Total Environment* 651 (2019), pp. 1166–1179.
- [109] L. R. Crilley, M. Shaw, R. Pound, L. J. Kramer, R. Price, S. Young, A. C. Lewis, and F. D. Pope. “Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring”. In: *Atmospheric Measurement Techniques* 11.2 (2018), pp. 709–720.
- [110] T. Zheng, M. H. Bergin, K. K. Johnson, S. N. Tripathi, S. Shirodkar, M. S. Landis, R. Sutaria, and D. E. Carlson. “Field evaluation of low-cost particulate matter sensors in high-and low-concentration environments”. In: *Atmospheric Measurement Techniques* 11.8 (2018), pp. 4823–4846.
- [111] F. D. Pope, M. Gatari, D. Ng’ang’a, A. Poynter, and R. Blake. “Airborne particulate matter monitoring in Kenya using calibrated low-cost sensors”. In: *Atmospheric Chemistry Physics* 18.20 (2018), pp. 15403–15418.

- [112] P. B. English, L. Olmedo, E. Bejarano, H. Lugo, E. Murillo, E. Seto, M. Wong, G. King, A. Wilkie, D. Meltzer, G. Carvlin, M. Jerrett, and A. Northcross. “The Imperial County Community Air Monitoring Network: A Model for Community-based Environmental Monitoring for Public Health Action”. In: *Environ Health Perspect* 125.7 (2017), p. 074501.
- [113] M. Masiol, N. Zikova, D. C. Chalupa, D. Q. Rich, A. R. Ferro, and P. K. Hopke. “Hourly land-use regression models based on low-cost PM monitor data”. In: *Environ Res* 167 (2018), pp. 7–14.
- [114] J. Bi, J. Stowell, E. Y. W. Seto, P. B. English, M. Z. Al-Hamdan, P. L. Kinney, F. R. Freedman, and Y. Liu. “Contribution of low-cost sensor measurements to the prediction of PM_{2.5} levels: A case study in Imperial County, California, USA”. In: *Environ Res* 180 (2020), p. 108810.
- [115] A. Mahmud, M. Hixson, J. Hu, Z. Zhao, S. H. Chen, and M. J. Kleeman. “Climate impact on airborne particulate matter concentrations in California using seven year analysis periods”. In: *Atmospheric Chemistry and Physics* 10.22 (2010), pp. 11097–11114.
- [116] R. B. A. Koelemeijer, C. D. Homan, and J. Matthijsen. “Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe”. In: *Atmospheric Environment* 40.27 (2006), pp. 5304–5315.
- [117] A. Lyapustin, Y. Wang, S. Korkin, and D. Huang. “MODIS collection 6 MAIAC algorithm”. In: *Atmos. Meas. Tech. Discuss* 11.10 (2018), pp. 1–50.
- [118] J. Belle and Y. Liu. “Evaluation of Aqua MODIS Collection 6 AOD Parameters for Air Quality Research over the Continental United States”. In: *Remote Sensing* 8.10 (2016), p. 815.
- [119] S. Feinberg, R. Williams, G. S. Hagler, J. Rickard, R. Brown, D. Garver, G. Harshfield, P. Stauffer, E. Mattson, and R. Judge. “Long-term evaluation of air sensor technology under ambient conditions in Denver, Colorado”. In: *Atmospheric Measurement Techniques* 11.8 (2018), pp. 4605–4615.

- [120] A. C. Rai, P. Kumar, F. Pilla, A. N. Skouloudis, S. Di Sabatino, C. Ratti, A. Yasar, and D. Rickerby. “End-user perspective of low-cost sensors for outdoor air pollution monitoring”. In: *Sci Total Environ* 607-608 (2017), pp. 691–705.
- [121] I. Gollini, B. B. Lu, M. Charlton, C. Brunsdon, and P. Harris. “GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models”. In: *Journal of Statistical Software* 63.17 (2015), pp. 1–50.
- [122] M. N. Wright and A. Ziegler. “ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R”. In: *Journal of Statistical Software* 77.1 (2017), pp. 1–17.
- [123] J. H. Ward. “Hierarchical Grouping to Optimize an Objective Function”. In: *Journal of the American Statistical Association* 58.301 (1963), pp. 236–244.
- [124] M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, and M. M. Charrad. “Package ‘nbcust’”. In: *Journal of statistical software* 61 (2014), pp. 1–36.
- [125] M. Levy Zamora, F. Xiong, D. Gentner, B. Kerkez, J. Kohrman-Glaser, and K. Koehler. “Field and Laboratory Evaluations of the Low-Cost Plantower Particulate Matter Sensor”. In: *Environ Sci Technol* 53.2 (2019), pp. 838–849.
- [126] G. Geng, N. L. Murray, H. H. Chang, and Y. Liu. “The sensitivity of satellite-based PM_{2.5} estimates to its inputs: Implications to model development in data-poor regions”. In: *Environ Int* 121.Pt 1 (2018), pp. 550–560.
- [127] G. Allen, C. Sioutas, P. Koutrakis, R. Reiss, F. W. Lurmann, and P. T. Roberts. “Evaluation of the TEOM method for measurement of ambient particulate mass in urban areas”. In: *J Air Waste Manag Assoc* 47.6 (1997), pp. 682–689.
- [128] D. E. Day, W. C. Malm, and S. M. Kreidenweis. “Aerosol light scattering measurements as a function of relative humidity”. In: *J Air Waste Manag Assoc* 50.5 (2000), pp. 710–716.
- [129] F. Lurmann, E. Avol, and F. Gilliland. “Emissions reduction policies and recent trends in Southern California’s ambient air quality”. In: *Journal of the Air & Waste Management Association* 65.3 (2015), pp. 324–335.
- [130] D. Tong, L. Pan, W. W. Chen, L. Lamsal, P. Lee, Y. H. Tang, H. Kim, S. Kondragunta, and I. Stajner. “Impact of the 2008 Global Recession on air quality over the United

- States: Implications for surface ozone levels from changes in NO_x emissions”. In: *Geophysical Research Letters* 43.17 (2016), pp. 9280–9288.
- [131] P. A. Solomon, D. Crumpler, J. B. Flanagan, R. K. M. Jayanty, E. E. Rickman, and C. E. McDade. “U.S. National PM 2.5 Chemical Speciation Monitoring Networks—CSN and IMPROVE: Description of networks”. In: 64.12 (2014), pp. 1410–1438.
- [132] S. Y. Kim, J. L. Peel, M. P. Hannigan, S. J. Dutton, L. Sheppard, M. L. Clark, and S. Vedal. “The Temporal Lag Structure of Short-term Associations of Fine Particulate Matter Chemical Constituents and Cardiovascular and Respiratory Hospitalizations”. In: *Environmental Health Perspectives* 120.8 (2012), pp. 1094–1099.
- [133] Y. H. Tian, X. Xiang, J. Juan, K. X. Sun, J. Song, Y. Y. Cao, and Y. H. Hu. “Fine particulate air pollution and hospital visits for asthma in Beijing, China”. In: *Environmental Pollution* 230 (2017), pp. 227–233.
- [134] W. Glezen. “Asthma, influenza, and vaccination”. In: *Journal of Allergy and Clinical Immunology* 118.6 (2006), pp. 1199–1206.
- [135] W. D. Flanders, M. Klein, L. A. Darrow, M. J. Strickland, S. E. Sarnat, J. A. Sarnat, L. A. Waller, A. Winquist, and P. E. Tolbert. “A method for detection of residual confounding in time-series and other observational studies”. In: *Epidemiology* 22.1 (2011), pp. 59–67.
- [136] R. B. Devlin, K. E. Duncan, M. Jardim, M. T. Schmitt, A. G. Rappold, and D. Diaz-Sanchez. “Controlled Exposure of Healthy Young Volunteers to Ozone Causes Cardiovascular Effects”. In: *Circulation* 126.1 (2012), pp. 104–111.
- [137] M. Ji, D. S. Cohan, and M. L. Bell. “Meta-analysis of the association between short-term exposure to ambient ozone and respiratory hospital admissions”. In: *Environmental Research Letters* 6.2 (2011).
- [138] Y. L. Zhao, R. Saleh, G. Saliba, A. A. Presto, T. D. Gordon, G. T. Drozd, A. H. Goldstein, N. M. Donahue, and A. L. Robinson. “Reducing secondary organic aerosol formation from gasoline vehicle exhaust”. In: *Proceedings of the National Academy of Sciences of the United States of America* 114.27 (2017), pp. 6984–6989.

- [139] E. F. Kirrane, T. J. Luben, A. Benson, E. O. Owens, J. D. Sacks, S. J. Dutton, M. Madden, and J. L. Nichols. “A systematic review of cardiovascular responses associated with ambient black carbon and fine particulate matter”. In: *Environment International* 127 (2019), pp. 305–316.
- [140] J. Fan, S. Li, C. Fan, Z. Bai, and K. Yang. “The impact of PM_{2.5} on asthma emergency department visits: a systematic review and meta-analysis”. In: *Environmental Science and Pollution Research* 23.1 (2016), pp. 843–850.
- [141] X.-Y. Zheng, H. Ding, L.-N. Jiang, S.-W. Chen, J.-P. Zheng, M. Qiu, Y.-X. Zhou, Q. Chen, and W.-J. Guan. “Association between Air Pollutants and Asthma Emergency Room Visits and Hospital Admissions in Time Series Studies: A Systematic Review and Meta-Analysis”. In: *PLOS ONE* 10.9 (2015), e0138146.
- [142] E. A. Gibson, Y. Nunez, A. Abuawad, A. R. Zota, S. Renzetti, K. L. Devick, C. Gennings, J. Goldsmith, B. A. Coull, and M. A. Kioumourtzoglou. “An overview of methods to address distinct research questions on environmental mixtures: an application to persistent organic pollutants and leukocyte telomere length”. In: *Environmental Health* 18.1 (2019).
- [143] K. W. Taylor, B. R. Joubert, J. M. Braun, C. Dilworth, C. Gennings, R. Hauser, J. J. Heindel, C. V. Rider, T. F. Webster, and D. J. Carlin. “Statistical Approaches for Assessing Health Effects of Environmental Chemical Mixtures in Epidemiology: Lessons from an Innovative Workshop”. In: *Environmental Health Perspectives* 124.12 (2016).
- [144] P. K. Hopke, C. Kane, M. J. Utell, D. C. Chalupa, P. Kumar, F. Ling, B. Gardner, and D. Q. Rich. “Triggering of myocardial infarction by increased ambient fine particle concentration: Effect modification by source direction”. In: 142 (2015), pp. 374–379.
- [145] D. Q. Rich, W. Zhang, S. Lin, S. Squizzato, S. W. Thurston, E. Van Wijngaarden, D. Croft, M. Masiol, and P. K. Hopke. “Triggering of cardiovascular hospital admissions by source specific fine particle concentrations in urban centers of New York State”. In: *Environment International* 126 (2019), pp. 387–394.

- [146] J. Li, M. Pósfai, P. V. Hobbs, and P. R. Buseck. “Individual aerosol particles from biomass burning in southern Africa: 2, Compositions and aging of inorganic particles”. In: *Journal of Geophysical Research: Atmospheres* 108.D13 (2003).
- [147] L. Hartman, Y. Zhu, K. M. Edwards, M. R. Griffin, and H. K. Talbot. “Underdiagnosis of Influenza Virus Infection in Hospitalized Older Adults”. In: *Journal of the American Geriatrics Society* (2018).
- [148] M. G. Thompson, M. Z. Levine, S. Bino, D. R. Hunt, T. M. Al-Sanouri, E. A. F. Simoes, R. M. Porter, H. M. Biggs, L. Gresh, A. Simaku, I. Abu Khader, V. L. Tallo, J. K. Meece, M. McMorrow, E. S. Mercado, S. Joshi, N. P. DeGroot, I. Hatibi, F. Sanchez, M. G. Lucero, S. Faouri, S. N. Jefferson, N. Maliqari, A. Balmaseda, D. Sanvictores, C. Holiday, C. Sciuto, Z. Owens, E. Azziz-Baumgartner, and A. Gordon. “Underdetection of laboratory-confirmed influenza-associated hospital admissions among infants: a multicentre, prospective study”. In: *Lancet Child & Adolescent Health* 3.11 (2019), pp. 781–794.
- [149] J. A. Davis, Q. Meng, J. D. Sacks, S. J. Dutton, W. E. Wilson, and J. P. Pinto. “Regional variations in particulate matter composition and the ability of monitoring data to represent population exposures”. In: 409.23 (2011), pp. 5129–5135.
- [150] S. L. Zeger, D. Thomas, F. Dominici, J. M. Samet, J. Schwartz, D. Dockery, and A. Cohen. “Exposure measurement error in time-series studies of air pollution: concepts and consequences”. In: *Environ Health Perspect* 108.5 (2000), pp. 419–426.
- [151] M. Badura, P. Batog, A. Drzeniecka-Osiadacz, and P. Modzel. “Evaluation of Low-Cost Sensors for Ambient PM_{2.5} Monitoring”. In: *Journal of Sensors* 2018 (2018).
- [152] W. S. Cleveland and S. J. Devlin. “Locally weighted regression: an approach to regression analysis by local fitting”. In: *Journal of the American statistical association* 83.403 (1988), pp. 596–610.