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:

Nancy L. Murray	Date

Ambient Air Pollution Estimation Using Bayesian Hierarchical Models

By

Nancy L. Murray

Doctor of Philosophy

Biostatistics

Howard H. Chang, PhD
Advisor

Stefanie Ebelt, ScD
Committee Member

Yang Liu, PhD
Committee Member

Robert Lyles, PhD
Committee Member

Lance Waller, PhD
Committee Member

Accepted:

Kimberly Jacob Arriola, PhD, MPH
Dean of the James T. Laney School of Graduate Studies

Date

Ambient Air Pollution Estimation Using Bayesian Hierarchical Models

By

Nancy L. Murray
BS, University of Tennessee, Knoxville
MS, Emory University

Advisor: Howard H. Chang, 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 Biostatistics
2021

Abstract

Ambient Air Pollution Estimation using Bayesian Hierarchical Models

By

Nancy L. Murray

Ambient fine particulate matter less than 2.5 μm in aerodynamic diameter ($PM_{2.5}$) negatively affects various health outcomes. However, the sparsity of existing air quality monitors greatly restricts the spatial-temporal coverage of $PM_{2.5}$ data, potentially limiting the accuracy of $PM_{2.5}$-related health studies. Various methods exist to address these limitations by supplementing air quality monitoring measurements with additional data. We aim to contribute to these methods with ambient air pollution estimation using Bayesian models. First, we develop a method to combine $PM_{2.5}$ estimated from satellite-retrieved aerosol optical depth (AOD) and chemical transport model (CTM) simulations using statistical models. In an application of estimating daily $PM_{2.5}$ in the Southeastern US, the ensemble approach outperforms previously developed spatial-temporal statistical models that use either AOD or bias-corrected CTM simulations in cross-validation (CV) analyses. Second, we evaluate the potential impact of differential exposure measurement error in $PM_{2.5}$ when examining differences in associations among subpopulations defined by spatial regions. In a simulation study, we observe bias when performing stratified analyses by neighborhood-level socioeconomic status measures when exposure granularity is ignored. Finally, we further develop the ensemble approach for $PM_{2.5}$ using multiple models and improve accuracy of methods by incorporating covariates into the weights. Bayesian estimation is accomplished through data augmentation with parameter expansion. The resulting weights are then used in a Bayesian ensemble averaging framework to combine estimates across data integration techniques.

Ambient Air Pollution Estimation Using Bayesian Hierarchical Models

By

Nancy L. Murray
BS, University of Tennessee, Knoxville
MS, Emory University

Advisor: Howard H. Chang, PhD

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 Biostatistics
2021

Acknowledgements

It is impossible to thank everyone who has shared this PhD journey with me by name. I feel lucky to have met and been supported by so many people at Emory and beyond.

I cannot begin to express my thanks to my advisor, Dr. Howard Chang, who exhibits infinite patience. This dissertation would not have been possible without you and your insights. I extend sincere gratitude to my committee members, whose thoughtful comments on this dissertation taught me valuable lessons until the very end of my time at Emory. Specifically, Dr. Bob Lyles offered me endless support and encouragement throughout the PhD process; Dr. Stefanie Ebelt and Dr. Yang Liu served as the content experts who constantly elevated my work; and Dr. Lance Waller supported me, first, as department chair, then committee member and offered mentorship and accountability along the way.

Many thanks to everyone who has taken a chance on me over the last seven years. Dr. Amita Manatunga and Dr. Reneé Moore encouraged my growth and were trusted supporters of not only my research but my overall being. Dr. John Hanfelt always supported me throughout my PhD process. I would also like to acknowledge Dr. Manish Patel, who hired a PhD student with an unfinished dissertation during the COVID-19 pandemic.

I want to send thanks and gratitude to all faculty and staff in the Department of Biostatisics and Bioinformatics, for making time and space for me to grow. I cannot leave Emory without thanking Mary Abosi for her friendliness and compassion. Additionally, I thank Dr. Steve Qin and Angela Guinyard for their commitment to the success of our students. I must thank all of my student colleagues in the department for resources, support, and inspiration to aspire to your level of greatness. I had the great pleasure of getting to know and learn from so many of you.

Many thanks to my friends, who constantly cheered me on and showed me compassion when I needed it most.

Finally, I am extremely grateful for my family, who have always promised to love me, regardless of what letters come after my name. Special thanks to Tyler Rowe, for everything.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Particulate matter less than 2.5 μm in diameter

Ambient air pollution, consisting of particulate matter (PM), nitrogen dioxide (NO_2), ozone, and other components, has been proven to be harmful to human health. (Brunekreef and Holgate, 2002; Hoek et al., 2013; Liu et al., 2013; Clark et al., 2014*a*; Evans et al., 2014; Brook et al., 2017) Due to monitoring of these dangerous pollutants through monitoring networks and through satellite data, researchers have been able to quantify the amount of pollution on a given day and, therefore, can relate adverse health outcomes to these specific pollutants in epidemiologic studies. (Bowatte et al., 2018; Jerrett et al., 2008; Miller and Marty, 2010; Strickland et al., 2010; Sarnat et al., 2010) With these important relationships in mind, air pollution can be curtailed through lowering emissions and modification of human activity through air quality regulations and, as a result, related adverse health outcomes can be reduced.

This dissertation focuses on $PM_{2.5}$, or particulate matter (PM) less than 2.5 μm in aerodynamic diameter. $PM_{2.5}$ consists of small particles and, subsequently, human exposure is virtually unavoidable in ambient air. $PM_{2.5}$ is composed of primary and secondary PM, with primary PM emitted directly from sources like wildfires, erosion, and pollen and secondary PM resulting from chemical reactions in the atmosphere. (United States Environmental Protection Agency, 2009) The availability of accurate $PM_{2.5}$ estimates is crucial in understanding the harmful effects of $PM_{2.5}$. Throughout this work, we aim to combine strengths of ground monitoring data, chemical transport models, and satellite data.

Population-based studies of air pollution and health contribute significantly to setting air quality standards worldwide. However, these studies draw criticism due to the routine use of regulatory monitoring networks to estimate exposures. Monitors in these networks are preferentially located in specific geographic areas, often in areas with high pollution levels and large populations. Due to the high cost of maintenance, $PM_{2.5}$ monitor measurements are spatially sparse, such that using these measurements over a large spatial domain may be inappropriate, and are sometimes temporally available only in 1-in-3 or 1-in-6 days time

periods.

Chemical transport models (CTM) are a type of numerical model used in air pollution research to estimate air pollution levels. CTMs are 3-dimensional deterministic models that simulate gridded air pollution concentrations based on state-of-the-art knowledge on drivers of air quality. (Chipperfield, 1999) Advantages of CTMs include their complete spatial-temporal coverage and the ability to incorporate chemical and physical processes associated with air pollution. However, CTMs are computationally expensive and often are only available at crude spatial resolutions.

For $PM_{2.5}$, remotely-sensed aerosol optical depth (AOD) has been examined extensively in its ability to predict $PM_{2.5}$ in combination with other meteorological and land use variables. (Liu et al., 2005, 2009) AOD measures the degree to which aerosols prevent light from penetrating the atmosphere. Some main advantages of satellite-based AOD are its fine spatial resolution, global coverage, and public accessibility. However, remotely-sensed data can suffer from missing data due to retrieval error and cloud cover.

CTMs and AOD values cannot be used directly in health analyses because complex spatial-temporal bias exists when compared to ground-level monitoring data. (Marmur et al., 2006; Friberg et al., 2017, 2018; Loría-Salazar et al., 2017) Therefore, statistical data fusion models that calibrate CTM and AOD data against observed measurements are needed. (Berrocal et al., 2010; Chang et al., 2014)

1.2 Bayesian Spatial Hierarchical Models

Bayesian models provide a flexible framework to deal with the complexities of models with space-time measurements. (Wikle et al., 1998) Due to the advances in computation over the years, Bayesian methods can now be efficiently implemented for spatial data analysis. (Hepple, 1995) Likelihood-based methods from a frequentist approach do exist. (Gelfand et al., 2010) However, the main advantage of using Bayesian inference is the ability to bet-

ter quantify estimation uncertainties, especially for the unobserved spatial-temporal trends. In analyzing air pollution data, while an abundance of data is available from monitoring networks, much of the geographical region is unmonitored.

While monitoring air pollution through satellite data has advanced over the years, it is economically infeasible and improbable that we will have complete spatial-temporal coverage in the near future. Various spatial models for air pollution exist, including Bayesian spatial hierarchical models. These hierarchical models have multiple stages, with each stage building upon the next as described below,

$$\text{Data Model: } Y|\beta, \eta, \sigma_\epsilon^2 \sim Gaussian(X\beta + H\eta, \sigma_\epsilon^2 I) \tag{1.1}$$

$$\text{Process Model: } \eta|\theta \sim Gaussian(0, \Sigma(\theta)) \tag{1.2}$$

$$\text{Parameter Model: } [\beta, \sigma_\epsilon^2, \theta] \tag{1.3}$$

Notice here the inclusion of the latent process $\boldsymbol{\eta}$, where $\eta = (\eta(s_1, ..., \eta(s_n))'$ and $\eta(s_i)$ is a Gaussian spatial process with spatial location s_i. Additionally, X is a matrix of covariates; β is some vector of parameters; H connects the observations Y to η; σ_ϵ^2 is a spatial effect with measurement error process ϵ; and θ contains parameters utilized in the spatial covariance function. (Gelfand et al., 2010)

1.3 Data Fusion

More recently, an important research area in environmental engineering and epidemiology involves developing data fusion products that supplement monitoring measurements with numerical model simulations and remotely-sensed observations from satellites. These data fusion models typically involve hindcasting, or using estimates and observed data from a previous time period to assess model performance, as a means of supplementing health analyses. The overarching goal of data fusion is to increase the spatial-temporal coverage

of air quality data to support health analyses and health impacts assessments, as seen in existing data fusion literature that supports epidemiological studies of birth outcomes. (Gray et al., 2014; Warren et al., 2016; Chang et al., 2011)

Previous studies have used available data from satellite imaging at the available spatial resolution of the data. However, technology is expanding and available resolutions are narrowing. For example, the aerosol remote sensor Moderate Resolution Imaging Spectroradiometer (MODIS) now has spatial resolution as low as 0.25 km. (National Aeronautics and Space Administration, 2018). With this evolving technology, it is imperative that our estimation techniques allow the use of these spatially refined data. On the other hand, these advanced data take time to process, with processing times increasing as the spatial resolution becomes smaller. Additionally, different data layers may not have the same resolutions. Our proposed methods aim to accommodate these differences in spatial resolution to obtain improved estimates as the technology in the air pollution monitoring field evolves.

Existing statistical methods that combine data, such as Bayesian melding, require modeling of the entire unknown pollution surface. Additionally, the bias between predictors and observations is usually assumed to take simple parametric forms (e.g., linear or quadratic). (Fuentes and Raftery, 2005; Crooks and Isakov, 2013; Zidek et al., 2012) Our proposed methods have the computational advantage of connecting point measurements and gridded predictors.

1.4 Specific Aims

We use Bayesian hierarchical modeling methods and data fusion methods throughout this dissertation in order to meet the following aims:

Aim 1. To combine estimates of $PM_{2.5}$ from statistical downscalers based on satellite imagery and numerical model simulation. We utilize the Bayesian Model Averaging framework to create a method that combines estimates and their uncertainties

with spatially-varying weights. Our goal is to improve prediction performance by borrowing strengths from the existing methods and data inputs.

Aim 2. To evaluate the potential impact of differential exposure measurement error when examining differences in associations among subpopulations defined by spatial regions. We utilize the 20-county metropolitan Atlanta area emergency department data and ZIP-code level socioeconomic status data. Using estimated $PM_{2.5}$ at 1km $\times$ 1km spatial resolution building on Aim 1, we conduct a simulation study and real data analysis to assess the impact of exposure measurement error when cruder exposure metrics are used to estimate effect modification.

Aim 3. To combine estimates of $PM_{2.5}$ from multiple models and improve accuracy of methods by incorporating covariates into the weights. We develop a modeling framework for incorporating potentially important information into the weights from Aim 1. We adapt existing methods for spatial classification such as multinomial probit models.

Chapter 2

A Bayesian Ensemble Approach to Combine $PM_{2.5}$ Estimates from Statistical Models Using Satellite Imagery and Numerical Model Simulation

2.1 Introduction

Air pollution negatively impacts human health, as supported by various studies around the world. (Brunekreef and Holgate, 2002; Hoek et al., 2013; Liu et al., 2013; Clark et al., 2014*a*; Evans et al., 2014; Brook et al., 2017) While air pollution represents a complex mixture of chemicals, particulate matter (PM) less than 2.5 μm in aerodynamic diameter ($PM_{2.5}$), in particular, has received increasing interest in public health. (Pui et al., 2014; Hart et al., 2015*a*; Maji et al., 2017) $PM_{2.5}$ is a mixture of solids and liquids that can penetrate deep into the lower respiratory system to affect the lungs and circulatory system. (Brook et al., 2002; Maté et al., 2010; Adam et al., 2015) $PM_{2.5}$ is composed of primary and secondary PM, with primary PM coming from sources like wildfires, erosion, and pollen; and secondary PM resulting from chemical reactions in the atmosphere. (United States Environmental Protection Agency, 2009) Also, sources of $PM_{2.5}$ include power generation, industrial operations, and automobiles. These anthropogenic emissions and the changing climate can have notable impacts on $PM_{2.5}$ concentrations and, subsequently, on human health. As a result, the United States Environmental Protection Agency (USEPA) regulates $PM_{2.5}$ as one of its criteria pollutants to protect public health. (Hubbell et al., 2009)

Population-based studies of air pollution and health have contributed significantly to setting regulatory standards worldwide. However, these studies often suffer from routine estimation of exposures using regulatory monitoring networks. Monitors in these networks are preferentially located in specific geographic areas, often in areas with high pollution levels and large populations. Due to high cost of maintenance, $PM_{2.5}$ monitor measurements are spatially sparse, such that extrapolating these measurements over a large spatial domain may be inappropriate, and are sometimes temporally available only in 1-in-3 or 1-in-6 days time periods.

More recently, an important research area involves developing data fusion products that supplement monitoring measurements with numerical model simulations and remotely sensed observations. These data fusion models typically involve retrospectively estimating expo-

sures. The overarching goal of data fusion is to increase the spatial-temporal coverage of air quality data to support health analyses and health impacts assessments, as seen in existing applications of data fusion products for epidemiological studies involving exposure estimation. (Gray et al., 2014; Warren et al., 2016; Chang et al., 2011)

Numerical model simulations used in air pollution research are known as chemical transport models (CTM). CTMs are 3-dimensional deterministic models that simulate gridded air pollution concentrations based on state-of-the-art knowledge on drivers of air quality. (Chipperfield, 1999) Advantages of CTM include its complete spatial-temporal coverage and the ability to incorporate chemical and physical processes associated with air pollution. However, CTM is computationally expensive and, as a result, is often only available at crude spatial resolutions. Because CTM are often archived and shared, CTMs continue to be used in estimating $PM_{2.5}$.

Remotely sensed aerosol optical depth (AOD) has been examined extensively in its ability to predict $PM_{2.5}$ in combination with other meteorological and land use variables. (Liu et al., 2005, 2009) AOD measures the degree to which aerosols prevent light from penetrating the atmosphere. AOD measurements can come from both polar orbiting or geostationary satellites. (Levy et al., 2007; Zhou et al., 2018) We focus on polar orbiting satellites in this work. Some main advantages of satellite-based AOD are its fine spatial resolution, global coverage, and public accessibility. However, remotely sensed data represents columnular measurements and can suffer from missing data due to retrieval error and cloud cover.

CTM simulations and AOD values cannot be used directly in health analyses because complex spatial-temporal bias exists when compared to ground-level monitoring data. (Marmur et al., 2006; Friberg et al., 2017, 2018; Loría-Salazar et al., 2017) For example, the Community Multiscale Air Quality (CMAQ) model, a type of CTM, may suffer from underprediction or overprediction due to error in inputs and discretization over space and time. (Mebust et al., 2003; Lim et al., 2010) AOD measures aerosol over the entire atmospheric column and its relationship with ground-level $PM_{2.5}$ can depend on various factors. There-

fore, statistical data fusion models that calibrate CTM and AOD data against observed measurements are needed. (Berrocal et al., 2010; Chang et al., 2014)

Most existing data fusion models that incorporate uncertainty quantification have been developed to utilize only one data source: CTM or satellite AOD. Concurrent utilization of both data sources in the fusion process may provide more accurate $PM_{2.5}$ estimates. Specifically, CTM simulation can address the missing data problem in satellite AOD, while satellite AOD can provide additional fine-scale spatial information to CTM simulation. Current approaches center around using CTM simulations to impute missing AOD values, followed by using the gap-filled AOD field as a predictor of $PM_{2.5}$ in multi-stage regression models (de Hoogh et al., 2018; Xiao et al., 2017) or machine learning algorithms. (Di et al., 2016; Hu et al., 2017; Reid et al., 2015) Kloog et al. provides full coverage by first fitting a model with available data, then smoothing predictions from this model to achieve complete spatial-temporal coverage. (Kloog et al., 2015) Because of this multi-stage approach, obtaining prediction standard error is challenging. Similarly, in the Global Burden of Disease project of Van Donkelaar et al., annual $PM_{2.5}$ averages are obtained by using satellite AOD values that are informed by CTM simulations to account for the vertical aerosol profile.(Van Donkelaar et al., 2016) However, one cannot conduct an epidemiological study that examines short-term effects of air pollution based on results from the Global Burden of Disease project. (Brauer et al., 2012) Most of the aforementioned studies rely on long-term averages, which do not capture the complex daily missingness of AOD data. Our method uses a sophisticated statistical model to estimate daily $PM_{2.5}$ exposure while also propagating uncertainty.

In this article, we describe a way to combine estimates of $PM_{2.5}$ from spatial-temporal statistical models using Bayesian ensemble averaging. Specifically, predictions from statistical data fusion models using either CTM simulation or satellite AOD are combined with spatially varying weights. The focus on statistical models is motivated by the need to provide uncertainties in $PM_{2.5}$ estimates, in terms of prediction standard error, that can be used

in subsequent health effect and health impact analyses. Our model-based approach offers several advantages compared to previous methods, namely the ability to incorporate various sources of uncertainty in predictions and to characterize the relative prediction performance of CTM versus satellite AOD. Model-based approaches also provide data-driven information on relationships between $PM_{2.5}$ and predictors that is often not available from algorithmic, or machine learning, methods. In an application, we evaluate the proposed method for predicting daily $PM_{2.5}$ in the Southeastern United States (Southeastern US) using 12km CTM simulations and 1km satellite-derived AOD.

2.2 Methods

2.2.1 Data

We obtained daily ground-level 24-hour average measurements of $PM_{2.5}$ from 63 monitors in the Southeastern US over the period 2003 to 2005 via the USEPA's Air Quality System (AQS). We strategically use this period of time in order to perform subsequent health analyses with data from the same time period. CTM simulations were obtained from the USEPA Models-3/Community Multiscale Air Quality (CMAQ) model version 4.5 at a 12 km $\times$ 12 km horizontal spatial resolution. (Byun and Schere, 2006) We acquired satellite-retrieved AOD measurements by the aerosol remote sensor Moderate Resolution Imaging Spectroradiometer (MODIS), which orbits the Earth on the National Aeronautics and Space Administration's Aqua and Terra satellites. We utilized a new multiangle implementation of atmospheric correction (MAIAC) algorithm that provides AOD values at a 1km $\times$ 1km spatial resolution. (Lyapustin, Martonchik, Wang, Laszlo and Korkin, 2011; Lyapustin, Wang, Laszlo, Kahn, Korkin, Remer, Levy and Reid, 2011) For each AOD grid cell, we also compiled variables including: elevation from the US Geological Survey, forest cover and road lengths from the 2001 National Land Cover data, meteorology (e.g. wind speed) from the North American Land Data Assimilation Systems, and $PM_{2.5}$ primary emission point sources from the 2002

USEPA National Emissions Inventory. As in Hu et al., forest cover and elevation were averaged from their original resolutions of about 1 km and about 30 m, respectively, to the 1 km $\times$ 1 km MAIAC grid cell level.(Hu et al., 2013) Additionally, road lengths and point emissions were summed over the 1 km $\times$ 1 km MAIAC grid cell level.

Figure 2.1 shows the locations of the 63 AQS monitors in our study region and gridded $PM_{2.5}$ simulations from CMAQ for an example day of March 17, 2005. Similarly, Figure 2.2, with an overlay of the same AQS monitor locations, shows the 1km-level satellite MAIAC AOD values on the same day with a considerable amount of missing data. Overall, the MAIAC AOD is missing for approximately 57% of the days and grid cells in our study. The differences in spatial resolution are also apparent between CMAQ and MAIAC AOD in Figures 2.1 and 2.2.

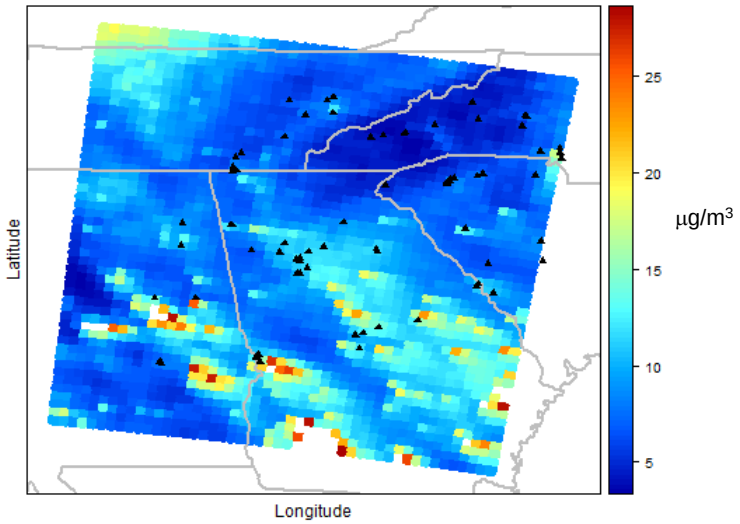

Figure 2.1: Simulation of $PM_{2.5}$ from the Community Multiscale Air Quality (CMAQ) model at 12 km resolution on March 17, 2005. Black triangles indicate AQS monitoring locations.

2.2.2 Statistical Modeling

2.2.2.1 Bayesian Hierarchical Modeling for Daily $PM_{2.5}$

We first describe the model for combining monitoring data with CMAQ outputs or AOD retrievals as predictors for point-referenced AQS monitoring measurements in a Bayesian spatial-temporal hierarchical model (BHM). Predictions of $PM_{2.5}$ from the $PM_{2.5}$-CMAQ BHM and $PM_{2.5}$-AOD BHM are subsequently used as inputs to the ensemble model.

Let $Y(\boldsymbol{s}, t)$ represent the observed $PM_{2.5}$ concentration on day t at locations $\boldsymbol{s}$. Following Berrocal et al. and Chang et al., our statistical model has the form of a BHM:

$$Y(\boldsymbol{s}, t) = \alpha_1(\boldsymbol{s}) + \alpha_2(\boldsymbol{s})X(\boldsymbol{s}, t) + \beta_1(t) + \beta_2(t)X(\boldsymbol{s}, t) + \boldsymbol{Z}(\boldsymbol{s}, t)\boldsymbol{\gamma} + \epsilon(\boldsymbol{s}, t), \tag{2.1}$$

where $X(\boldsymbol{s}, t)$ is the linked AOD or CMAQ values in the grid cells containing the monitor at locations $\boldsymbol{s}$, and $\boldsymbol{Z}(\boldsymbol{s}, t)$ is a vector of additional predictors with coefficient $\boldsymbol{\gamma}$.(Berrocal et al., 2010; Chang et al., 2014) For the AOD model, $\boldsymbol{Z}(\boldsymbol{s}, t)$ includes the following land use and meteorology variables: elevation, forest cover, road length, primary emission source, wind speed, and temperature. Because CMAQ uses information on emissions and meteorology to

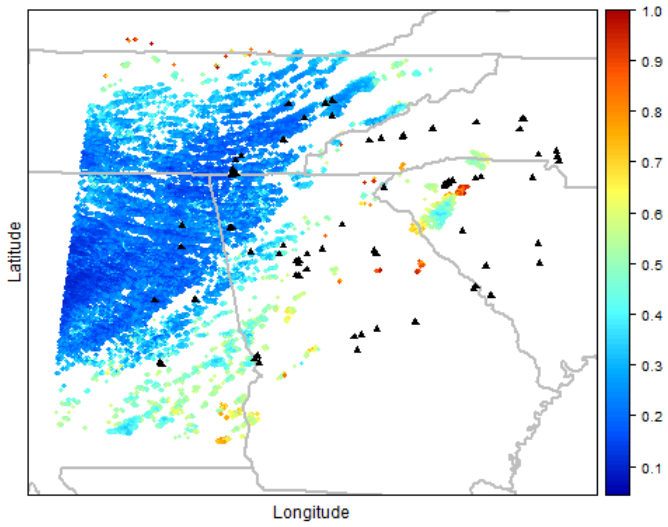

Figure 2.2: Satellite-derived aerosol optical depth (AOD) at 1 km $\times$ 1 km gridded resolution on March 17, 2005. Black triangles indicate AQS monitoring locations.

perform simulations, $\boldsymbol{Z}(\boldsymbol{s}, t)$ is not included in the $PM_{2.5}$-CMAQ BHM. Preliminary analysis also showed that including additional covariates does not improve prediction performance for the CMAQ model. Finally, the residual error term, $\epsilon(\boldsymbol{s}, t)$, is independent normally distributed with mean zero and variance σ_y^2.

Parameters $\alpha_1(\boldsymbol{s})$ and $\alpha_2(\boldsymbol{s})$ in Equation (2.1) are the spatial random intercept and spatial random slope, respectively; $\beta_1(t)$ and $\beta_2(t)$ in Equation (2.1) are the temporal random intercept and temporal random slope, respectively. $\alpha_1(\boldsymbol{s})$, $\alpha_2(\boldsymbol{s})$, $\beta_1(t)$, and $\beta_2(t)$ are sometimes referred to as calibration parameters because they correct for the additive and multiplicative bias associated with CMAQ or AOD. Additional details about the modeling assumptions for BHM can be found in Appendix A.

2.2.2.2 Combining Estimates from Statistical Models

Our proposed method to combine $PM_{2.5}$ estimates from the CMAQ-only and AOD-only model is based on the Bayesian Model Averaging (BMA) framework. BMA has been applied to probabilistic weather forecasting in order to combine forecasts from different numerical weather models.(Raftery et al., 2005) Here, we extend the approach for estimating spatial-

temporal air pollution concentrations when predictions from multiple statistical models are available. To our knowledge, this framework has not previously been used in modeling spatial-temporal air pollution.

We consider the following model:

$$p(y_{st} \mid M_1, M_2) = w_s f_1(y_{st} \mid M_1) + (1 - w_s) f_2(y_{st} \mid M_2), \tag{2.2}$$

where y_{st} is the $PM_{2.5}$ value; $f_k(y_{st} \mid M_k)$ is the posterior *predictive* distribution of y_{st} from model M_k, and w_s is the weight for the $PM_{2.5}$-CMAQ BHM at location s. w_s ranges from 0 to 1, with a default value of 1 on days where AOD is missing.

Equation (2.2) can be viewed as a predictive model, where w_s is the posterior probability (ensemble weight) that the $PM_{2.5}$-CMAQ BHM is the better estimate of $PM_{2.5}$ at monitor s. Here we assume $f_k(y_{st} \mid M_k) \equiv \phi\left(y_{st} \mid \mu_{st}^{(k)}, \sigma_{st}^{2,(k)}\right)$, i.e., a Normal posterior predictive distribution of y_{st} with mean $\mu_{st}^{(k)}$ and variance $\sigma_{st}^{2,(k)}$ using either the $PM_{2.5}$-CMAQ BHM ($k = 1$) or the $PM_{2.5}$-AOD BHM ($k = 2$). Hence, the point prediction of y_{st} can be defined by its posterior mean

$$\hat{y}_{st} = w_s \mu_{st}^{(1)} + (1 - w_s)\mu_{st}^{(2)}, \tag{2.3}$$

which is a weighted average of predictions from the $PM_{2.5}$-CMAQ BHM and the $PM_{2.5}$-AOD BHM. Similarly, the error for y_{st} is defined as

$$\mathrm{Var}(y_{st}|\mu_{st}^{(1)}, \mu_{st}^{(2)}) = w_s\sigma_{st}^{2,(1)} + (1-w_s)\sigma_{st}^{2,(2)} + w_s\mu_{st}^{2,(1)} + (1-w_s)\mu_{st}^{2,(2)} - \left(w_s\mu_{st}^{(1)} + (1 - w_s)\mu_{st}^{(2)}\right)^2 \tag{2.4}$$

which allows us to quantitatively define uncertainties and make inferences. Bayesian inference also allows us to capture the uncertainty in the weight estimation procedure. Additionally, the posterior interval can be defined as the 2.5% and the 97.5% interval of the conditional distribution. To allow for spatial interpolation of the ensemble weight to locations without monitors, we further assume that $q_s = \mathrm{logit}(w_s)$ is a Gaussian process with

an exponential covariance function, i.e., $\text{Cov}(q_s, q_{s'}) = \tau^2 e^{-||s-s'||/\rho}$, where τ^2 controls the spatial variability, and ρ controls the rate of spatial decay in dependence.

2.2.2.3 Estimation and Prediction

Estimation and prediction are accomplished in seven stages, which we describe in the enumerated steps below.

1. Fit the $PM_{2.5}$-CMAQ BHM using Equation (2.1) to obtain posterior predictive means, $\mu_{st}^{(1)}$, and variances, $\sigma_{st}^{2,(1)}$, for each day and location.

2. Fit the $PM_{2.5}$-AOD BHM using Equation (2.1) to obtain posterior predictive means, $\mu_{st}^{(2)}$, and variances, $\sigma_{st}^{2,(2)}$, for each day and location where we have observed AOD values.

3. Create out-of-sample CMAQ-based predictions. Randomly leave 10% of the $PM_{2.5}$ observations out then obtain prediction means and prediction variances using the remaining 90% of the data. Repeat this ten times. Stack the predictions to create a dataset.

4. Create out-of-sample AOD-based predictions. Randomly leave 10% of the $PM_{2.5}$ observations out then obtain prediction means and prediction variances using the remaining 90% of the data. Repeat this ten times. Stack the predictions to create a dataset. Note: the training folds and validation folds are the same for CMAQ and AOD.

5. Estimate spatially varying weights based on $PM_{2.5}$ measurements and out-of-sample prediction datasets from Steps 3 and 4 using Equation (2.2).

6. Interpolate the weights to 1 km $\times$ 1 km grid cells using kriging.

7. Combine the estimates from Steps 1 and 2 using weights from Step 5 in the same fashion as Equation (2.3) to obtain the *ensemble estimate*.

Notice in Steps 3 and 4, to avoid overfitting while estimating ensemble weights, we fit the BHMs repeatedly, but we leave-out and back-predict observations in a *cross-validation* experiment, similar to approaches employed in stack regression and SuperLearner techniques. (LeBlanc and Tibshirani, 1996; Polley and van der Laan, 2010)

Estimation and inference are carried out in a Bayesian framework by specifying priors for all model parameters. Markov chain Monte Carlo (MCMC) methods are used to obtain samples from posterior distributions; we use Gibbs sampler when the full conditional distributions are in closed-form and the random-walk Metropolis-Hasting algorithm otherwise. MCMC computations are standard for Bayesian hierarchical modeling and are provided elsewhere. (Chang et al., 2014) MCMC details for fitting the BHM (Step 1 and Step 2) and the ensemble weights (Step 5) are provided in Appendix A.

We also investigate alternative approaches to estimate the ensemble weights. In addition to the 10-fold cross-validation (CV) predictions to estimate the weights, we also use a spatial (leave-one-monitor-out) CV approach. We also consider estimating the weights by using a two-stage approach, which first estimates the optimal weight at each model separately, then performs spatial interpolation in a second stage. This method differs from Step 5 in that the uncertainty in the monitor-specific weight is not accounted for in the spatial interpolation.

We use R version 3.5.1 for all estimation and prediction. (R Core Team, 2018) The MCMC algorithm is available, coded in R, through the corresponding author's Github site. Sample data is also posted.

2.2.2.4 Assessing Model Performance

We evaluated the prediction performance of the proposed ensemble approach using three out-of-sample cross-validation (CV) experiments. First, in a 10-fold CV, we randomly divided the dataset into 10 subsets. Repeatedly, we left out each subset (10% of the data) and used the other 90% of the data to fit the prediction model. Because data are available at each monitor in each CV fold, this 10-fold CV experiment allowed us to evaluate the

model's ability to perform *temporal* interpolation when daily $PM_{2.5}$ is missing at monitoring locations.

We also performed a *spatial* CV experiment where all observations at each monitor were left out one-monitor-at-a-time. This allowed us to evaluate the model's ability to perform *spatial* interpolation to estimate $PM_{2.5}$ at locations without monitors.

Finally, we performed *spatially clustered* CV, where 20 clusters formed through hierarchical clustering by proximity of monitoring locations (the *hclust* function in the *stats* package of R) were dropped, and the remaining data were used to estimate $PM_{2.5}$ at multiple locations without monitors (leave one-cluster-at-a-time out). (Johnson, 1967) The twenty clusters, as well as more details about their formation, are given in Figure A.1 in Appendix A. The spatially clustered CV simulates a more realistic scenario where the modeling approaches are tasked with spatially interpolating a larger group of spatially missing data rather than a single missing location. (Young et al., 2016)

We quantified the performance of different methods using the following statistics: prediction root-mean-square error (RMSE), 95% coverage probability of the posterior intervals (PI), average posterior standard deviation (SD), and R^2. R^2 and RMSE were calculated based on posterior predictive means of the left-out observed $PM_{2.5}$ concentrations. Posterior prediction intervals were based on the 2.5^{th} and the 97.5^{th} percentiles of the posterior distribution of the two-component predictive model distribution in Equation (2.2).

2.3 Results

Due to the deterministic construction of CMAQ simulations, we have full spatial-temporal coverage for CMAQ across the Southeastern US during the study period of 2003-2005. AOD, on the other hand, is only available at 57.4% of locations and days. $PM_{2.5}$ observations from AQS monitoring sites are available at 75.8% of all days over the three-year study period. Observed $PM_{2.5}$ has a mean of 14.54 $\mu g/m^3$ and a standard deviation of 7.02 $\mu g/m^3$.

The mean value of $PM_{2.5}$ as determined by the CMAQ simulation is 12% lower at 12.78 $\mu g/m^3$. Mean AOD is 0.24. Pearson correlations show moderate linear relationships between observed $PM_{2.5}$ and CMAQ at 0.57 and observed $PM_{2.5}$ and AOD at 0.54. CMAQ and AOD are weakly correlated with a Pearson correlation coefficient of 0.13.

Table 2.1 gives model performance results for the i) 10-fold CV experiment, ii) spatial CV experiment, and iii) spatially clustered CV experiment. Overall, the ensemble approach resulted in improved out-of-sample predictions. Specifically, using inputs derived from the 10-fold CV, the ensemble model achieved the lowest RMSE and highest R^2 in all three evaluations, with the RSME of the ordinary CV being 43% of the standard deviation of $PM_{2.5}$ measurements. The posterior prediction SD for the 10-fold CV increases, but this increase allows us to maintain the proper coverage of at least 95%. While results of the spatial CV and spatially clustered CV show similar trends as in the 10-fold CV experiment, we find the improvement of the ensemble approach over separate models tends to be better, suggesting the ensemble approach is particularly beneficial for spatial interpolation compared to using only CMAQ or only AOD. Among the three CV experiments, prediction performance decreases from 10-fold CV to spatial CV due to the need to spatially interpolate to locations without monitors; prediction performance also decreases from spatial CV to spatially clustered CV because the number of nearby monitors to aid in interpolation is limited. Despite these differences in CV experiment results, the ensemble approach continues to outperform separate statistical models in all three types of CV experiments. Using a two-stage estimation approach resulted in a negligible reduction in prediction performance compared to the joint estimation method utilized above; these results can be found in Appendix A in Table A.1 and Table A.2. The spatial CV prediction inputs result in weaker results but still justify use of our method compared to the individual models, as seen in Appendix A in Table A.2.

Figure 2.3 clearly demonstrates the need for spatially varying weights due to the $PM_{2.5}-$CMAQ BHM receiving a higher assigned weight value for the predictive model in certain areas, whereas the $PM_{2.5}-$AOD BHM receives higher weights in more rural areas but also close to

Evaluation Method	Statistical Model	RMSE	Coverage of 95% PI	Average Posterior SD	R^2
Ordinary (10-fold) CV	$PM_{2.5}$-AOD BHM	3.40	94.07	3.30	0.78
	$PM_{2.5}$-CMAQ BHM	3.14	95.05	3.28	0.81
	Ensemble	3.00	97.15	3.52	0.83
Spatial (Leave-one-monitor-out) CV	$PM_{2.5}$-AOD BHM	3.45	94.25	3.39	0.77
	$PM_{2.5}$-CMAQ BHM	3.33	95.32	3.45	0.78
	Ensemble	2.99	96.81	3.52	0.83
Spatially clustered (Leave-one-cluster-out) CV	$PM_{2.5}$-AOD BHM	3.62	94.43	3.59	0.74
	$PM_{2.5}$-CMAQ BHM	3.93	93.34	3.58	0.69
	Ensemble	3.13	95.73	3.43	0.81

RMSE: root mean squared error (in $\mu g/m^3$); PI: prediction interval; SD: standard deviation (in $\mu g/m^3$); CV: cross-validation; $PM_{2.5}$: particulate matter less than 2.5 μm; AOD: aerosol optical depth; BHM: Bayesian hierarchical model; CMAQ: Community Multiscale Air Quality

Table 2.1: Prediction performance for daily $PM_{2.5}$ concentrations in 10-fold cross-validation (CV) comparing ensemble averaging with a Bayesian hierarchical model (BHM) using satellite-derived aerosol optical depth (AOD) or a BHM using a numerical model (CMAQ) simulation. Ensemble weights were derived from first performing 10-fold CV.

some urban centers across the study time period. We can also illustrate spatially kriging the weight estimates from the 10-fold CV experiment to areas without monitoring locations at a finer spatial resolution of 1 km $\times$ 1 km across the Southeastern US (Figure A.2 in Appendix A).

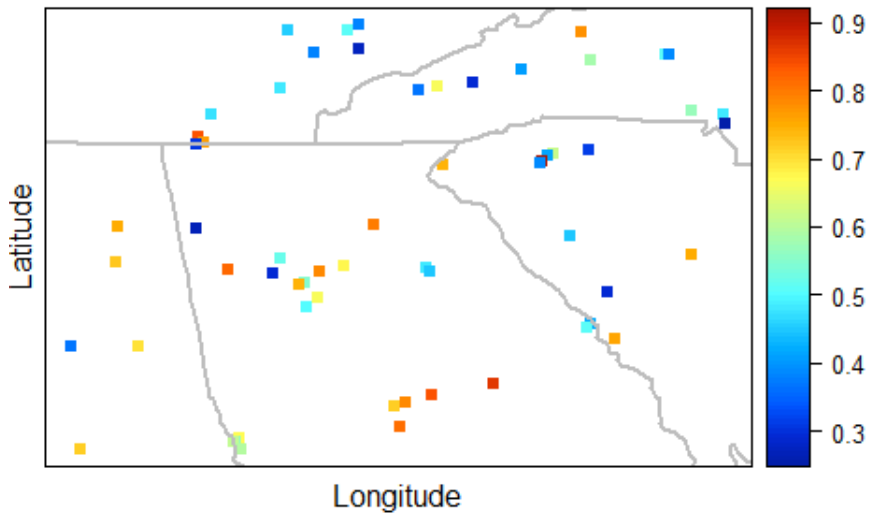

Figure 2.3: Ensemble weights for predictions from the $PM_{2.5}$-CMAQ Bayesian hierarchical model at AQS monitoring locations.

While the derived weights spatially vary across the Southeastern US, the metropolitan Atlanta, GA area is a clear example of the varying weights the $PM_{2.5}-$CMAQ BHM receives within a relatively small geographical area. $PM_{2.5}$'s environmental health effects are well-documented in Atlanta, GA. (Alhanti et al., 2016; Gass et al., 2015) To that end, we illustrate the use of ensemble estimates of $PM_{2.5}$ within the 20-county metropolitan Atlanta, GA area. We aim to contrast results from the two individual AOD and CMAQ models with our results from the combined, ensemble method. This Atlanta region contains 16,063 AOD grid cells and 143 CMAQ grid cells.

Figure 2.4 demonstrates the applicability of the ensemble approach for a single day. The 20-county metropolitan Atlanta area has 9 AQS monitors, but the ensemble approach, combined with spatial kriging and interpolation, allows us to extend the use of weights beyond

areas with monitors to obtain posterior predictive mean $PM_{2.5}$ concentrations across a wider swath of land. Figure 2.4(AOD model) shows the spatially refined $PM_{2.5}$ estimates from the $PM_{2.5}$-AOD BHM. Figure 2.4(CMAQ model), the $PM_{2.5}$-CMAQ BHM results, starkly differs from Figure 2.4(Ensemble model), the ensemble averaged results, in terms of smoothness. On this particular day, the $PM_{2.5}$-AOD BHM predicts lower $PM_{2.5}$ concentrations over Atlanta than the $PM_{2.5}$-CMAQ BHM (Figure 2.4(AOD model) and Figure 2.4(CMAQ model); also seen in Appendix A Figure A.3 and A.5. The standard error of the $PM_{2.5}$-AOD BHM is also lower than that of the $PM_{2.5}$-CMAQ BHM (Figure A.4 and A.6). The ensemble approach leads to an average of the $PM_{2.5}$-AOD BHM and the $PM_{2.5}$-CMAQ BHM predictions and, thereby, allows for depictions of seamless $PM_{2.5}$ estimates between neighboring spatial fields for which CMAQ alone does not have the complexity.

Figure 2.5 displays the long-term 3-year $PM_{2.5}$ concentration estimates over Atlanta from the $PM_{2.5}$-AOD BHM (Figure 2.5(AOD model)), the $PM_{2.5}$-CMAQ BHM (Figure 2.5(CMAQ model), and ensemble averages restricted to days when AOD was observed (Figure 2.5(Ensemble model (AOD observed))) or across all days (Figure 2.5(Ensemble model (all days))). The combination of information from the $PM_{2.5}$-AOD BHM and $PM_{2.5}$-CMAQ BHM permits more granularity in the maps on both a daily level (Figure 2.4(Ensemble model)) and when averaging across days where AOD is observed (Figure 2.5(Ensemble model (AOD observed))). This finer resolution on a daily level or on days with observed AOD will aid in acute environmental health effect analyses. However, in Figure 2.5(Ensemble model (all days)), the predictions from the $PM_{2.5}$-CMAQ BHM dominate, likely due to the large amount of temporally missing AOD in this region over the study time period (about 57%).

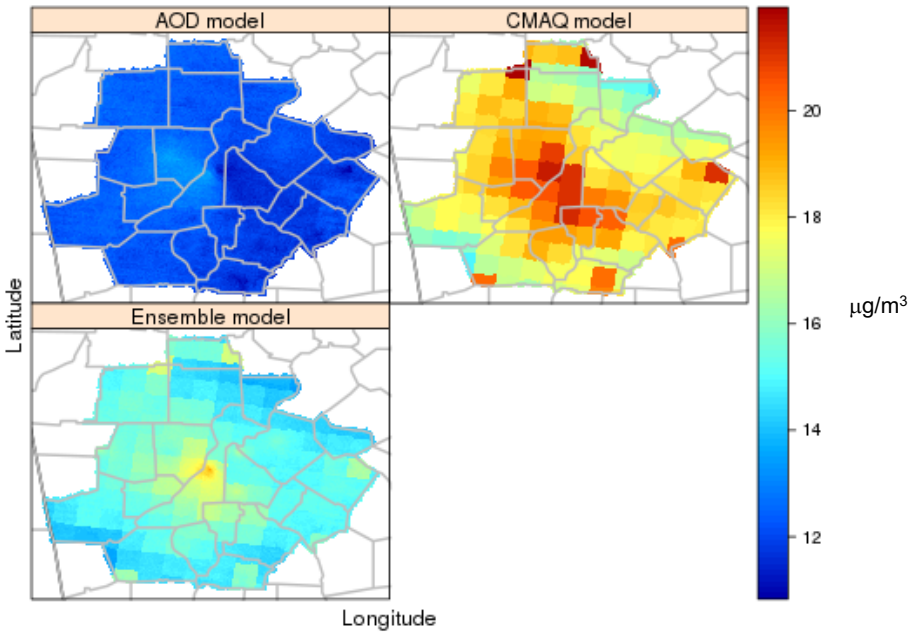

Figure 2.4: Daily estimates of $PM_{2.5}$ concentrations on March 26, 2005 in the 20-county metropolitan Atlanta, GA area using estimates from (top left) the $PM_{2.5}$-AOD Bayesian hierarchical model (BHM) , (top right) the $PM_{2.5}$-CMAQ BHM , and (bottom left) the ensemble method.

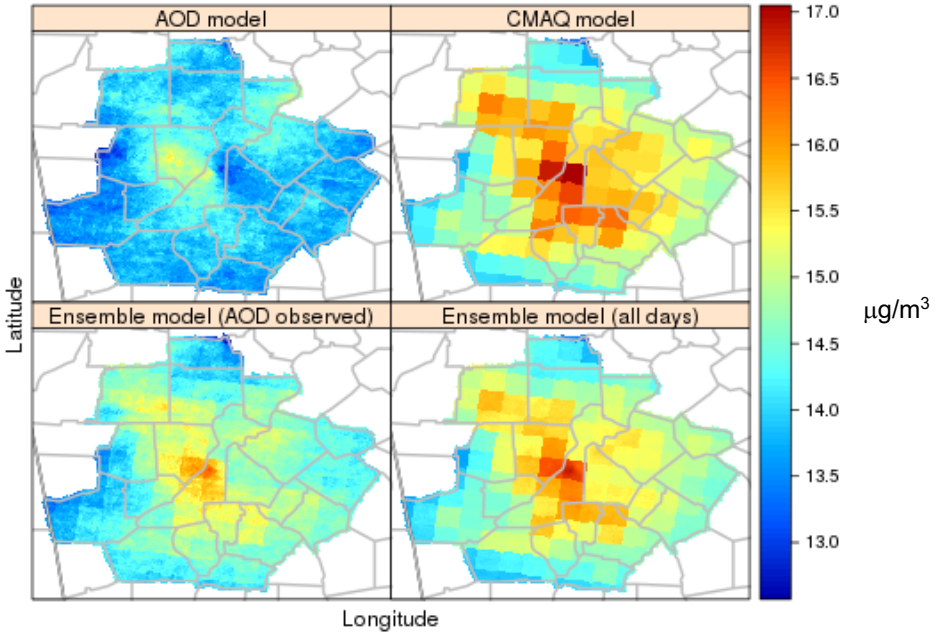

Figure 2.5: Posterior averages of $PM_{2.5}$ concentrations across 2003-2005 in the 20-county metropolitan Atlanta, GA area based on (top left) the $PM_{2.5}$-AOD Bayesian hierarchical model (BHM), (top right) the $PM_{2.5}$-CMAQ BHM, (bottom left) the ensemble method for days in the three-year time period where AOD is observed, and (bottom right) the ensemble method for all days in the three-year time period.

2.4 Discussion

Instead of relying solely upon numerical CTM simulations or satellite data to perform data fusion, the proposed combined statistical model framework allows us to incorporate both sources of information and harness their collective predictive power. Existing statistical methods that combine data, such as Bayesian melding, require modeling of the entire unknown pollution surface, which is computationally intensive, and the data source with the largest sample size (i.e. CTM) will dominate over monitoring measurements and satellite imagery. The proposed ensemble method does not incur significant additional computational burden because it estimates ensemble weights from 10-fold CV predictions, which are routinely performed by researchers when comparing prediction performance of different models.

Another advantage of the ensemble approach entails accounting for differences in spatial resolution between different gridded data because CTM and satellite data are first calibrated to the point-level using monitoring data via Bayesian hierarchical modeling. Finally, in our $PM_{2.5}$ application, the ensemble approach also naturally accounts for the missing values in satellite retrievals, providing $PM_{2.5}$ estimates with complete spatial-temporal coverage. Specifically, in settings with more than two inputs, when satellite AOD is missing, ensemble weights for different inputs can be reweighted among available inputs. In the current version of the ensemble method, missing AOD results in assigning the other input (CTM) a weight of one and proceeding with the estimation. This differs from existing approaches where AOD needs to be imputed before being used as a predictor for $PM_{2.5}$, increasing computational burden and introducing another source of prediction uncertainty.

The computation time of estimating weights for the ensemble method itself is not limiting because it is based on the number of monitors in the area. However, the computation time for predictions do take some time at the fine-scale resolution of 1 km $\times$ 1 km. For this reason, we displayed results in Atlanta, GA instead of the entire Southeastern region. Specifically, we presented predictions that incorporated spatial correlation between grid cells (i.e. predicting

"maps" of $PM_{2.5}$ concentrations jointly), but if that is not needed, weights can be spatially interpolated one grid cell at a time, which is highly parallelizable.

Our approach deviates from the existing methods in our lack of imputed AOD because of our focus on uncertainty quantification, an important assessment other methods, such as machine learning ensemble methods, cannot provide. Specifically, our Bayesian modeling framework provides prediction standard error, which can then be used for uncertainty quantification in subsequent health effect analyses. Although here we focus on ambient air pollution for the application of this method, the approach is also highly relevant to the estimation of other environmental exposures (e.g. temperature, precipitation) that utilize information from both satellite imagery and numerical model simulations.

This current case study presents a relatively small geographic area; however, the method can be extended to other regions with different meteorological and land characteristics as well. A previous analysis using a non-Bayesian ensemble approach, which does not provide prediction error quantification, to combine estimates from $PM_{2.5}$-AOD BHM and $PM_{2.5}$-CMAQ BHM in Colorado has shown similar improvements. (Geng et al., 2018)

While this is not the first use of Bayesian Model Averaging (BMA) to perform ensemble modeling in a spatial setting, we focus on combining statistical models rather than deterministic outputs from climate model simulations. (Berrocal et al., 2007) Bhat, et al. also use spatial-temporal BMA but to combine global climate projections. (Bhat et al., 2011) In contrast, our method interpolates the ensemble weights and uses the model on a much more localized level. Specifically, although the monitoring stations are not randomly placed, we are still able to obtain fine-scale spatially smoothed estimates due to the use of CMAQ and AOD. We also reconcile the spatial resolution differences between our statistical model estimates. Finally, as previously mentioned, we can handle data with high spatial sparsity as demonstrated by the performance of the spatial and spatially clustered CVs.

Several extensions of the proposed method warrant additional investigations. First, ensemble modeling can be generalized to consider multiple sources of information. For ex-

ample, one can consider a model only driven by fine-scale land use variables with AOD missing. Specifically, the two-component predictive model utilized here can be extended to have multiple weights (i.e. more than two) that are estimated with a multinomial latent variable. In the air pollution application, this may include (1) CTM simulations driven by different assumptions on emission levels and pollution composition for each emission source, (2) multiple satellite parameters that may inform different characteristics of aerosol, and (3) AOD retrievals from different satellites. Our model inputs are also not limited to BHM; we can adapt this method to obtain spatial weights based on the performance of other popular techniques that provide prediction standard error such as kriging or machine learning techniques such as random forests. We modeled spatially varying weights largely due to the ability of satellite-retrieved AOD to predict $PM_{2.5}$ over large areas at a fine scale in certain areas and the error in CMAQ simulation being likely to exhibit spatial variation. Another extension of the ensemble method is to allow weights to depend on spatial and temporal covariates (e.g. land use and meteorology). This may further improve $PM_{2.5}$ prediction and provide insights into which factors are associated with the relative underperformance of the $PM_{2.5}$-CMAQ BHM and $PM_{2.5}$-AOD BHM.

Chapter 3

Impacts of $PM_{2.5}$ Exposure Spatial Resolutions on Estimating Neighborhood-Level Socioeconomic Status as an Effect Modifier

3.1 Introduction

Exposure to outdoor air pollution has been linked to increased asthma morbidity, especially among children. (Wardlaw, 1993; Guarnieri and Balmes, 2014; Bowatte et al., 2018; Jerrett et al., 2008; Miller and Marty, 2010) Some populations are disproportionately exposed to higher levels of outdoor air pollution due to their residential proximity to emission sources. (Bell and Ebisu, 2012; Clark et al., 2014*b*; Mikati et al., 2018) One active area of research in environmental epidemiology focuses on examining how health effects of air pollution may vary across sub-populations defined by measures of socioeconomic status (SES). Effect modification of air pollution by SES may be attributed to differences in exposure levels, disease severity, and prevalence of comorbid conditions. (Neidell, 2004; Forastiere et al., 2007; Munoz-Pizza et al., 2020)

Particulate matter less than 2.5 μm in diameter ($PM_{2.5}$) lends itself to exacerbating existing respiratory health issues due to its chemically diverse composition and the small particles' ability to enter the lungs and become irritants. (Xing et al., 2016) In population-based epidemiologic studies, $PM_{2.5}$ exposures are typically estimated from ground monitors, as well as using a host of different data inputs and techniques. (United States Environmental Protection Agency, 2009; Xiao et al., 2017; Di et al., 2016; Van Donkelaar et al., 2016) The process of linking $PM_{2.5}$ exposure estimates to health outcomes may lead to complex exposure measurement errors due to insufficient characterization of the spatial-temporal variation in pollution concentration. (Gryparis et al., 2009; Alexeeff et al., 2016; Dionisio et al., 2014; Szpiro et al., 2011)

Neighborhood-level measures of SES are frequently used in times-series and case-crossover analyses for examining modifications of short-term air pollution health effects. However, the use of spatially-varying effect modifiers may result in exposure measurement errors that are *differential* across sub-populations. For example, low SES neighborhoods located in urban centers may have less exposure measurement error due to availability of nearby monitoring stations compared to high SES neighborhoods in the suburbs. Hence spatial patterns

of monitor locations and population demographics may result in spatially-varying exposure measurement errors that are dependent on the outcome, resulting in biased health effect estimates that are not necessarily towards the null. (Carrothers and Evans, 2000; VanderWeele and Hernán, 2012; van Smeden et al., 2020)

The goal of this study is to examine how exposure measurement error may impact our ability to estimate effect modifications of short-term $PM_{2.5}$ effects by neighborhood-level SES measures. First, we constructed estimates of daily $PM_{2.5}$ concentrations at 1km spatial resolution in southeastern United States using an ensemble approach to combine predictions from two random forest models. (Breiman, 2001) Then, we compared estimated short-term $PM_{2.5}$ associations with pediatric asthma-related emergency department visits in Atlanta using either the spatially-resolved $PM_{2.5}$ estimates across the study area or a non-spatially-varying exposure estimate derived from monitoring data. Finally, we conducted a simulation study to quantify bias due to exposure measurement error when performing stratified case-crossover analyses by neighborhood-level SES measures.

3.2 Methods

3.2.1 Data

3.2.1.1 Emergency Department Visit Data

The Study of Particulates and Health in Atlanta (SOPHIA) is a long-standing study on the short-term health effects of air pollution and the study has established clear connections between air pollution and emergency department (ED) visits for issues ranging from respiratory problems to circulatory disease in the metropolitan Atlanta area. (Peel et al., 2005; Metzger et al., 2004; Strickland et al., 2010; Sarnat et al., 2010) We acquire SOPHIA ED visits data originally analyzed in a case-crossover study by O'Lenick et al. that investigated ZIP code-level poverty measures as an effect modifier for short-term health effects of $PM_{2.5}$

and other pollutants on asthma-related pediatric ED visits.(O'Lenick et al., 2017) Specifically, patient-level ED visits are obtained from individual hospitals and the Georgia Hospital Association from the Atlanta area for children, 5 to 18 years old, between January 1, 2002 and December 31, 2008. ED visits for asthma or wheeze are based on primary or secondary ICD9 diagnosis codes. Daily ED visits are then aggregated at the residential ZIP code-level and matched to one of 173 ZIP code tabulation areas (ZCTAs, using 2010 Census Bureau boundaries, created from census blocks to approximate ZIP codes) in order to inspect effect modification by ZCTA-level SES measures in the air pollution health analysis.

3.2.1.2 Socioeconomic Status Data

Our socioeconomic status measures (SES) are derived from the 2000 US Census long-form and the American Community Survey (ACS) 5-year (2007-2011) summary files. We use linear interpolation between the 2000 Census and the ACS data to capture yearly changes in ZCTA-level SES over our 2002-2008 study period. (O'Lenick et al., 2017) We examine the difference in air pollution effects between high and low levels of SES by classifying the study areas into four sub-regions based on quartiles of percent of households in a ZCTA living below the federal poverty line. Quartile values of percent below poverty are calculated across all years and areas as: Q1 = <7.6%; Q2 = $\geq$ 7.6% to <11.4%; Q3 = $\geq$11.4% to <16.2%; Q4 = $\geq$16.2%.

3.2.1.3 $PM_{2.5}$ Monitoring Data and Predictors

We obtain daily ground-level 24-hour average measurements of $PM_{2.5}$ from 115 United States Environmental Protection Agency (USEPA) Air Quality System (AQS) monitors spanning six states in the Southeastern United States over the period 2002 to 2008 in order to train our exposure models. We use satellite-retrieved aerosol optical depth (AOD) measurements from the Earth-orbiting National Aeronautics and Space Administration's Aqua and Terra satellites' aerosol remote sensor Moderate Resolution Imaging Spectroradiometer (MODIS).

We estimate AOD using the multiangle implementation of atmospheric correction (MAIAC) algorithm that provides AOD values at a fine spatial resolution of 1km $\times$ 1km. (Lyapustin, Martonchik, Wang, Laszlo and Korkin, 2011; Lyapustin, Wang, Laszlo, Kahn, Korkin, Remer, Levy and Reid, 2011) For each 1 km-resolution AOD grid cell, in addition to spatial information from the grid cell centroid, we also obtain predictor variables including: percent forest cover and lengths of major road types from the 2001 National Land Cover data, daily meteorology (temperature, wind speed, and relative humidity) from the North American Land Data Assimilation Systems, and $PM_{2.5}$ primary emission point sources from the 2002 USEPA National Emissions Inventory. We also include planetary boundary layer height from the North American Regional Reanalysis. Figure 3.1 displays the 20-county metropolitan Atlanta area, along with ground monitors, primary roads, and quartiles of SES for a single year (2005).

3.2.2 Exposure Modeling via an Ensemble of Random Forests

While satellite-derived AOD has been shown to improve spatial-temporal $PM_{2.5}$ predictions, AOD can be informatively missing due to cloud cover or high surface reflectance. Imputation of missing AOD data may be an appealing approach for full spatial-temporal coverage, but for random forest-based $PM_{2.5}$ predictions, imputation has been shown to not necessarily yield greatly improved results. (Kianian et al., 2021) Instead, to obtain $PM_{2.5}$ estimates with complete spatial and temporal coverage, we combine estimates from two random forest models through an ensemble approach that allows us to use the rich set of covariate information for $PM_{2.5}$ predictions, as well as satellite-derived AOD when available. (Murray et al., 2019) Random forest is an approach increasingly used by air pollution researchers due to its accuracy and ease of use. (Hu et al., 2017; Brokamp et al., 2018; Huang et al., 2018) Random forest is a decision tree-based approach in which many small decision trees are grown and then aggregated to provide a single prediction averaged across trees. Random forest can be preferable to regression-based approaches due to its ability to better accommodate non-

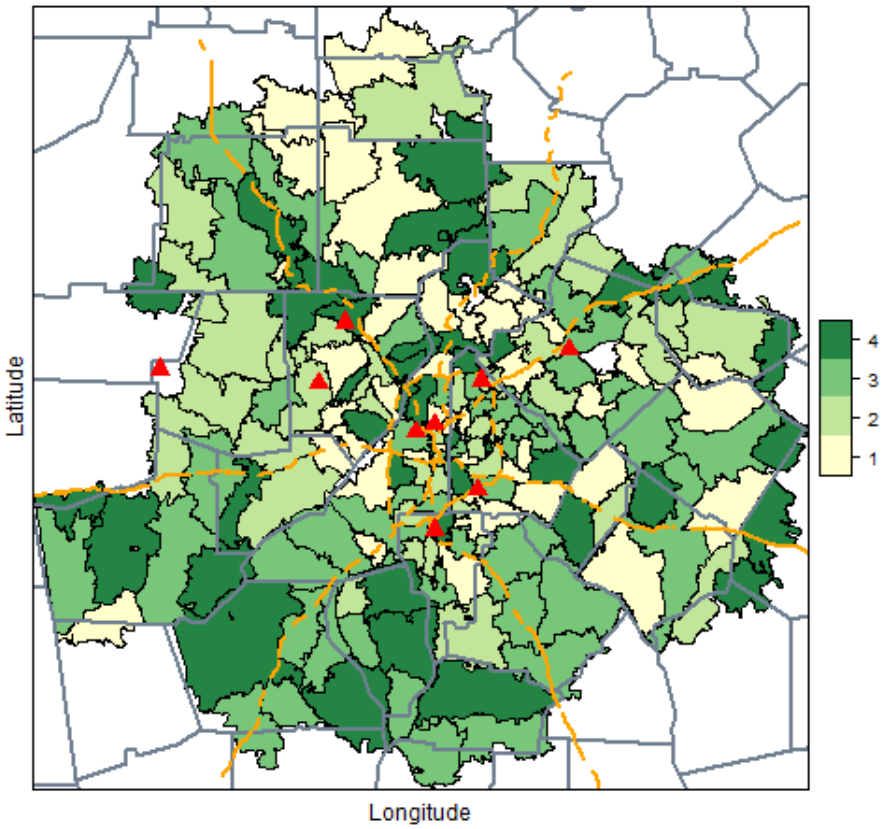

Figure 3.1: 20-county Atlanta study area with ZIP code tabulation area (ZCTA)-level socioeconomic status (SES) indicators in green with each shade of green indicating quartiles of percent below poverty from Q1(1) to Q4(4). Major highways are indicated in orange. USEPA AQS monitors based on 2005 data are shown as red triangles, and counties are outlined in gray.

linearity and general flexibility in specifying model inputs. More importantly, in contrast to other machine learning techniques, prediction uncertainties can be derived for random forest predictions useful for the ensemble approach.

We train two random forest models separately for each year. The first model uses land-use variables, meteorological data, and time covariates, including day of week, month, and day of year. We also include the projected x-y coordinates. We refer to this model as the "Base" model. In the second model, we additionally include AOD as a covariate where available. We refer to the second model as the "Base+AOD" model.

Each random forest model is evaluated by conducting an overall ten-fold cross-validation experiment and calculating the out-of-sample root mean squared error (RMSE), R^2, average prediction standard error, and 95% coverage probability of the prediction intervals. First, using data from years 2002, 2005, and 2008, we explore different hyperparameters to identify the combination that results in the best prediction performance. The chosen hyperparameter combinations consider all possible partitions of factor variables and are 1000, 7, 3 (number of trees, number of variables to consider at each split in a node, minimal node size, respectively) for the "AOD" model and 1000, 14, 3 for the "Base" model. (R Core Team, 2019; Wright and Ziegler, 2017) We find that within a certain year, the hyperparameter combinations result in similar predictive summaries.

In our previous study (Murray et al., 2019), we assign ensemble weights to estimates from models across multiple years. Here, we design the model to allow ensemble weights to vary for different years. This approach allows for more flexibility for the weighting across years. We combine predictions from the two random forest models using a Bayesian ensemble approach as follows. First, we assume the relationship between true $PM_{2.5}$ concentrations and predictions from random forest models follow a Normal distribution with density $f_r(y_{st} \mid M_r) \equiv \phi\left(y_{st} \mid \mu_{st}^{(r)}, \sigma_{st}^{2,(r)}\right)$, where y_{st} is the $PM_{2.5}$ value at location s and time t; ϕ denotes the Normal density function; $\mu_{st}^{(r)}$ is the corresponding prediction from the r^{th} random forest; and $\sigma_{st}^{2,(r)}$ is the variance of prediction. (Biau and Scornet, 2016)

Given predictions from two random forest models, the probabilistic predictions for $PM_{2.5}$ concentration are given by:

$$p(y_{st} \mid M_1, M_2) = w_s f_1(y_{st} \mid M_1) + (1 - w_s) f_2(y_{st} \mid M_2), \tag{3.1}$$

where $f_r(y_{st} \mid M_r)$ is the *predictive* distribution of y_{st} from model M_r, and w_s is the weight for the random forest at location s, where ensemble weights are allowed to be spatially-varying and range from 0 to 1. Let M_1 be the "Base" model and M_2 be the "Base+AOD" model. We set $w_s = 1$ on days where AOD is missing. We run 10,000 MCMC iterations with a burn-in of 2,500. Estimation of the spatially-varying ensemble weights and prediction are accomplished in multiple stages, which we omit here and refer the reader to the Supplementary Materials.

Once we obtain predictions of daily $PM_{2.5}$ from the ensemble model, we average the 1 km $\times$ 1 km estimates over each ZCTA area for use as exposure estimates in the health model.

3.2.3 Case-Crossover Analysis of Emergency Department Visit Data

We use case-crossover analysis to assess the short-term effect of $PM_{2.5}$ on ED visits for pediatric asthma. Each unique, independent ED visit (case) is assigned control days matched on year, month, and day of week. Specifically, each case serves as its own control for the other days of week during that month, thereby limiting potential confounding. Frequently, this results in three or four controls to one case. We implement a conditional logistic regression model with time-stratification by year, month, and day of the week. This approach to selecting control days minimizes overlap bias and time trend bias. (Janes et al., 2005) We also prefer this approach due to its ability to minimize temporal autocorrelation and remove for the effect of day of the week, which in the case of ED visits, is a known confounder. (Carracedo-Martínez et al., 2010)

We investigate the role of SES in the association between asthma ED visits and $PM_{2.5}$

exposure through stratified analyses by SES quartile. Let $Y_{kt}^{(q)}$ denote the binary variable for an ED visit for patient k on day t living in the qth quartile of SES (q=1,...,4 with q=1 representing the highest level of SES). The corresponding regression models are:

$$\text{logit}\{Pr(Y_{kt}^{(q)} = 1)\} = \sum_{k=1}^{N_p} \alpha_k V_k + \beta^{(q)} PM_{t-1}^{(q)} + \boldsymbol{\gamma} \boldsymbol{X}_t, \tag{3.2}$$

where V_k is the indicator variable for patient k and α_k is the patient-specific baseline risk coefficient for V_k; N_p is the number of case-control matched sets; and $\beta^{(q)}$ is the log odds ratio for 1-day lagged exposure $PM_{t-1}^{(q)}$. The set of temporally-varying covariates $\boldsymbol{X}_t$ includes: indicator variables for four seasons, for hospital participation on ED case days, and for holidays during the study period. We also include a cubic spline on day of year (5 degrees of freedom) to smoothly control for within-window seasonal trends. Meteorology is controlled with cubic polynomials for 3-day moving average (lags 0-2) of maximum temperature and mean dew point. We also include interaction terms between season and maximum temperature.

We examine two methods to define $PM_{2.5}$ exposures in order to evaluate the benefit of having spatially-varying exposures in estimating SES-specific odds ratios. The first method uses ZCTA-level averaged estimates (spatially-varying) derived from ensembled 1 km $\times$ 1 km exposures, as described in Section 3.2.2. The second method uses a spatially-homogeneous exposure ZCTA average located centrally in the study region.

3.3 Application to Asthma Emergency Department Data

The mean daily $PM_{2.5}$ for the "Base" model is 14.58 μg/m^3 whereas the mean daily $PM_{2.5}$ for the "Base+AOD" model is 15.26 μg/m^3 across the southeastern US during our study period of 2002-2008. AOD is missing in 54% of the total space-time units for our 2002-2008 data.

The overall population of the 173 ZCTAs over the 20-county metropolitan Atlanta area counties is 5,079,436 according to the 2010 Census data. There were a total of 127,396

pediatric asthma ED visit across the study period, and 22.01% of all space-time combinations in the study period had at least one pediatric asthma ED visit.

3.3.1 Random Forest Ensemble Exposure Modeling Results

The ensemble model and its accompanying weights are formed by relying on information from the entire southeastern region for an individual year, and because the "Base" model has better predictive performance across the study area with RMSE of 3.46 and R^2 of 76% averaged across study years (Supplementary Table 1), the ensemble model favors the "Base" model estimates in a 2005 example we highlight throughout.

Figure 3.2 shows the fine-scale, 1 km $\times$ 1 km predictions we are able to obtain from the ensembled RF model for a single day in Atlanta. We see the 20-county metropolitan area and demonstrate that even for a relatively small region, $PM_{2.5}$ varies not only from county to county but within each county.

Figure 3.3 illustrates the effect of averaging the fine scale ensemble predictions over each ZCTA for use in the health effects analysis. We see the same patterns in Figure 3.2 and Figure 3.3, where the more densely populated and trafficked areas have higher levels of $PM_{2.5.}$ and the more suburban areas have lower levels.

Figure 3.4 displays the year-long average for $PM_{2.5}$ concentration estimates averaged over the ZCTA areas. On average throughout the year, the ensemble is able to differentiate between the higher $PM_{2.5}$ levels in more urban areas versus the suburban areas.

3.3.2 Associations between $PM_{2.5}$ and Asthma ED Visits

In an unstratified analysis, the estimated odds ratio between yesterday's $PM_{2.5}$ exposure and asthma ED visits was 1.014 (95% CI: (0.998, 1.031)) per 10 $\mu g/m^3$ increase $PM_{2.5}$ concentration when using the ZCTA-level spatially-varying exposure and 1.010 (95% CI: (0.996, 1.024)) when using the spatially-homogeneous exposure.

Figure 3.5 gives estimated associations when the analysis is stratified by quartiles of

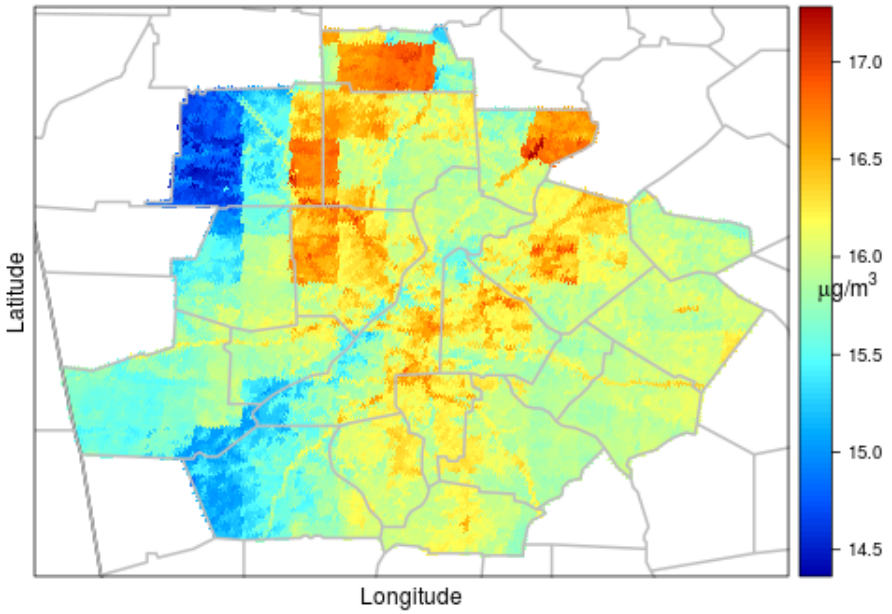

Figure 3.2: Ensemble $PM_{2.5}$ from the 20-county metro Atlanta area at 1 km $\times$ 1 km resolution on June 16, 2005.

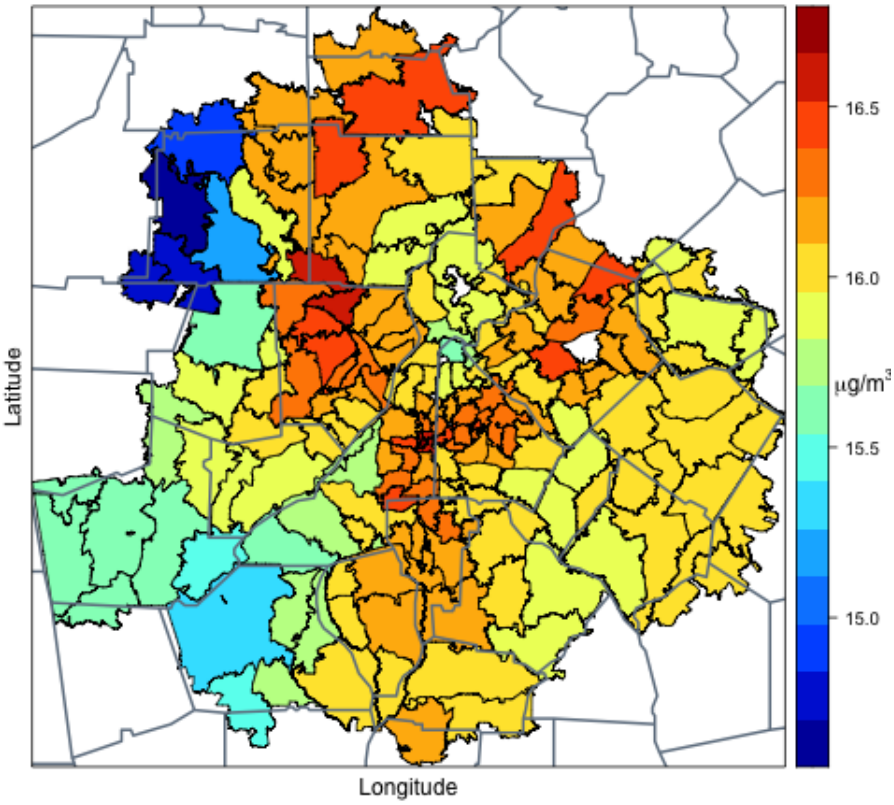

Figure 3.3: Ensemble $PM_{2.5}$ from the 20-county metro Atlanta area averaged over ZCTAs on June 16, 2005.

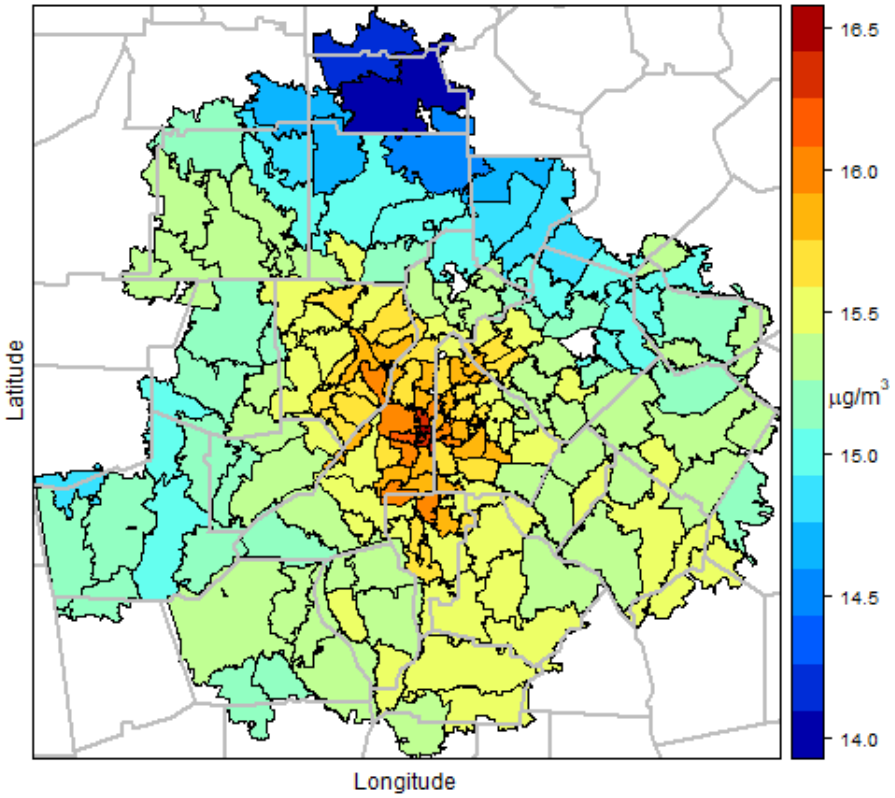

Figure 3.4: Ensemble $PM_{2.5}$ from 20-county metro Atlanta area averaged over 173 ZIP code tabulation areas (ZCTAs) daily over 2005.

SES. Associations between $PM_{2.5}$ and asthma ED visits across SES strata show the same pattern for both exposure assessment methods: the highest levels of SES (Q1) had the largest magnitude of association (1.033 and 1.025) and the 3rd quartile (Q3) had the smallest magnitude (0.996 and 0.992). The differences in estimates across quartiles are more pronounced with the spatially-varying exposure due to the Q1 estimate being slightly higher when a spatially-varying exposure is used for estimation.

3.4 Simulation for Emergency Department Visit Modeling

3.4.1 Simulation Settings

Population-based studies of $PM_{2.5}$ exposure and health often assess "exposure" differently. While exposures derived from monitoring networks are common, the method of assigning exposures to at-risk populations varies. For example, some analyses assign monitor values based on proximity to a patient's residence (Baxter et al. (2013); Boutin-Forzano et al. (2004) or ZIP code (Bell and Ebisu (2012)); others assign the monitoring values to a broader area (Glad et al. (2012); Grineski et al. (2015)).

For our analyses, we develop $PM_{2.5}$ exposure estimates at the ZCTA-level based on averaging over all 1 km $\times$ 1 km estimates within a certain ZCTA. We hypothesize that these spatially-varying exposure estimates will result in more accurate estimates of effect modification by SES. In this section, we conduct a simulation study to quantify the difference between the effect sizes estimated using a spatially coarser exposure estimate, when the true exposure is more spatially heterogeneous. Specifically, we first simulate multiple datasets of daily ED visits using "true" ZCTA-level exposures obtained from ensemble modeling (Section 3.2.2). Then, we estimate $PM_{2.5}$ health effects using a coarser exposure defined as the central ZCTA-level estimate that is assumed to be common across all ZCTAs.

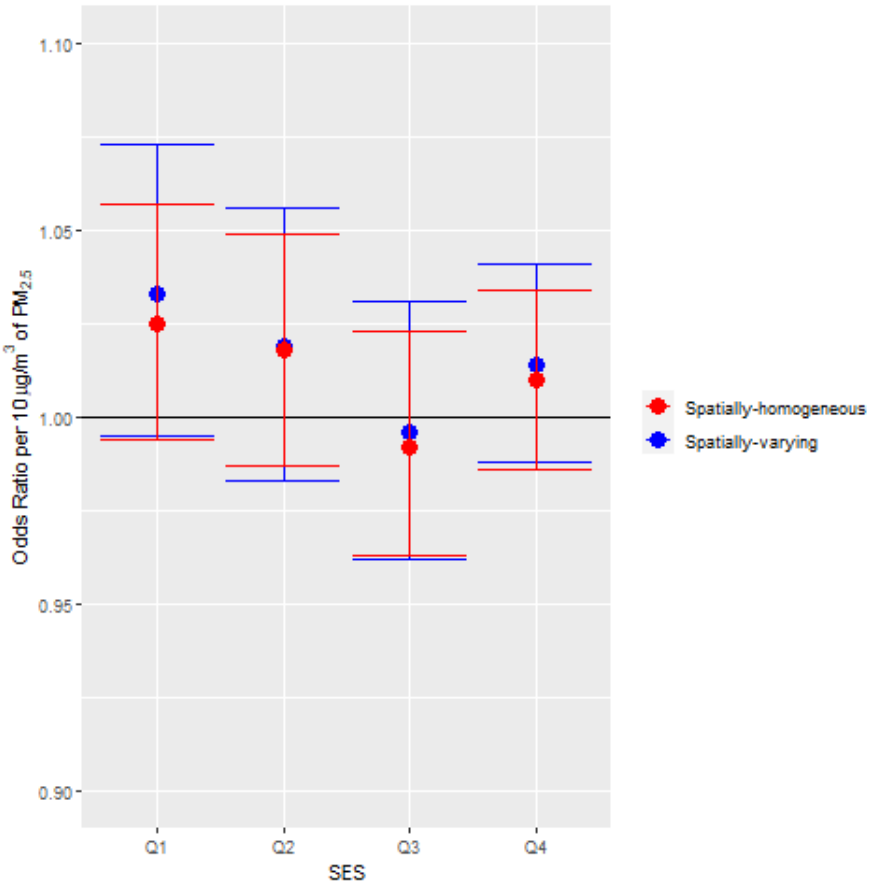

Figure 3.5: Odds ratios (per 10 μg/m^3 of $PM_{2.5}$) of asthma emergency department visits based on a spatially-varying ZIP code tabulation area (ZCTA)-level $PM_{2.5}$ exposure and spatially-homogeneous ZCTA-level $PM_{2.5}$ exposure, stratified by quartiles of socioeconomic status (SES) defined by percent of ZCTA living in poverty. Quartile values of percent below poverty were defined as: Q1 = <7.6%; Q2 = $\geq$ 7.6% to <11.4%; Q3 = $\geq$11.4% to <16.2%; Q4 = $\geq$16.2%. spatially-homogeneous- $PM_{2.5}$ exposure ZCTA average located centrally in the study region. spatially-varying- $PM_{2.5}$ exposure ZCTA average from each ZCTA in the study region.

Our main objective is to examine the difference in $PM_{2.5}$ short-term effects between high SES and low SES areas. In the simulation study, we define ZCTAs with the lowest quartile of poverty as high SES and all other quartiles as low SES. We use the binomial model in Equation (3.3) to simulate the number of cases for each ZCTA on each day for a one-year period.

$$\text{Binomial}(n_s, p_{st}) \text{ where } p_{st} = \frac{\exp(\beta_0 + PM_{st}\beta)}{1 + \exp(\beta_0 + PM_{st}\beta)}, \tag{3.3}$$

where n_s is the ZCTA-level (s) population in 2010, β_0 is chosen to provide similar incidence as the observed ED counts in the year 2005, PM_{st} is the true ZCTA-level (s) $PM_{2.5}$ concentration on day t, and β is the true health effect assigned based on SES level. We assign β in two different ways in order to investigate the bias when either high or low SES area has the higher effect from $PM_{2.5}$. When one SES level is assigned no effect ($\beta = 0$), the other SES level is assigned an effect invoking approximately a 5% difference per 10 $\mu g/m^3$ of $PM_{2.5}$ ($\beta = 0.0048$). (Kuo et al., 2018)

For each simulated dataset, we fit the stratified conditional logistic regression (Equation (3.2)), similar to those in Section 3.2.3 but without the additional confounders by SES level, i.e., no $\boldsymbol{X}_t$ term.

We compare health effect estimates from using three exposure assessment methods: (a) with spatially-varying ZCTA-level averaged estimates (true exposure), (b) a spatially-homogeneous exposure, derived from a single ZCTA-level average located in the center of the study area (spatially-homogeneous-single-site), and (c) a spatially-homogeneous exposure, derived by averaging ZCTA-level exposures from multiple locations that correspond to locations with US EPA monitors (spatially-homogeneous-average-of-sites). This allows us to examine differences in estimated effect modification for exposures either assigned close to a residence versus a more central monitor or an average of monitoring locations. We perform this simulation across 1000 times and assess this potential bias on average, as well as RMSE, 95% and 90% coverage probabilities, average standard error, Monte Carlo error, and power

or Type I error. (Alexeeff et al., 2015)

3.4.2 Simulation Results

Table 3.1 shows results where both SES levels (high and low) had null $PM_{2.5}$ effect. Within each SES group, estimation performance is similar across different exposure assessment methods in terms of average bias, RMSE, and coverage. The use of different exposures results in similar Type I error rates that are close to the nominal 5% level, with the low SES population being more conservative. Therefore, when the true effect is null, the impact of potential differential exposure measurement is minimal, and we did not detect biases away from the null.

Table 3.2 shows results when each SES group has a log odds ratio of 0.0048 (i.e., an effect size resulting in about a 5% increase in odds ratios per 10 $\mu g/m^3$). Among both high and low SES ZCTAs, the spatially-homogeneous-single-site exposure resulted in larger, negative biases compared to the gold standard (spatially-varying), around 10% of the true effect size. Biases associated with a spatially-homogeneous-average-of-sites exposures are not as pronounced, but they are still negative. Therefore, when the true exposures are spatially-varying, the use of spatially-homogeneous exposures may result in effect attenuation when examining SES-specific associations. We note that the use of spatially-homogeneous exposures had lower RMSE, likely due to smaller average standard error.

Table 3.1: Simulation with no effect: True $\beta_1 = 0$ Results of 1,000 case-crossover dataset results of stratified quartile analyses for 1) all days (lag 1) between 01/01/2005 and 12/31/2005 with assigned ensemble exposure averaged over each ZCTA (spatially-varying; gold standard); 2) all days (lag 1) between 01/01/2005 and 12/31/2005 with assigned ensemble exposure averaged over the ZCTA containing the central monitor and assigned to all ZCTAs on that day (spatially-homogeneous-single-site) 3) all days (lag 1) between 01/01/2005 and 12/31/2005 with assigned ensemble exposure averaged over the ZCTAs containing the central monitors in Atlanta and assigned to all ZCTAs on that day (spatially-homogeneous-average-of-sites)

SES	Data	Case Count	Average Bias*	RMSE*	95% Coverage	90% Coverage	Average SE*	MC Error*	Type I Error
High	spatially-varying (Gold standard)	5137	0.103	3.310	94.4	88.4	3.19	0.1	5.6
	spatially-homogeneous (average of sites)	5137	0.11	3.17	94.9	88.8	3.07	0.1	5.1
	spatially-homogeneous (single site)	5137	0.11	3.0	94.8	89.5	2.92	0.09	5.2
Low	spatially-varying (Gold standard)	12779	-0.02	1.95	95.9	91.0	1.97	0.06	4.1
	spatially-homogeneous (average of sites)	12779	0.017	1.95	95.7	90.0	1.95	0.062	4.3
	spatially-homogeneous (single site)	12779	-0.002	1.85	95.6	90.7	1.85	0.06	4.4

Footnotes: SES- socioeconomic status; High SES: Q1; Low SES: Q2-Q4; SE- standard error; RMSE- root mean squared error; MC- Monte Carlo; spatially-homogeneous (single site) - for the same day, every ZCTA receives the same exposure assignment from the central monitor containing ZCTA; spatially-homogeneous (average of sites) - for the same day, every ZCTA receives the same exposure assignment from the average of central monitors containing ZCTAs in Atlanta

* - estimates times 1,000

Table 3.2: Simulation with effect: True $\beta_1 = 0.0048$ Results of 1,000 case-crossover dataset results of stratified quartile analyses for 1) all days (lag 1) between 01/01/2005 and 12/31/2005 with assigned ensemble exposure averaged over each ZCTA (spatially-varying; gold standard); 2) all days (lag 1) between 01/01/2005 and 12/31/2005 with assigned ensemble exposure averaged over the ZCTA containing the central monitor and assigned to all ZCTAs on that day (spatially-homogeneous-single-site) 3) all days (lag 1) between 01/01/2005 and 12/31/2005 with assigned ensemble exposure averaged over the ZCTAs containing the central monitors in Atlanta and assigned to all ZCTAs on that day (spatially-homogeneous-average-of-sites)

SES	Data	Case Count	Average Bias*	RMSE*	95% Coverage	90% Coverage	Average SE*	MC Error*	Power
High	spatially-varying (Gold standard)	5538	0.11	3.18	93.8	88.5	3.05	0.1	37.6
	spatially-homogeneous (average of sites)	5538	-0.17	3.07	94.7	89.0	2.938	0.097	36.2
	spatially-homogeneous (single site)	5538	-0.47	2.94	94.3	88.8	2.79	0.09	35.2
Low	spatially-varying (Gold standard)	13793	-0.045	1.86	96.3	89.7	1.88	0.06	70.4
	spatially-homogeneous (average of sites)	13793	-0.18	1.86	95.4	89.4	1.86	0.059	70.1
	spatially-homogeneous (single site)	13793	-0.45	1.82	93.6	88.9	1.77	0.06	69

Footnotes: SES- socioeconomic status; High SES: Q1; Low SES: Q2-Q4; SE- standard error; RMSE- root mean squared error; MC- Monte Carlo; spatially-homogeneous (single site) - for the same day, every ZCTA receives the same exposure assignment from the central monitor containing ZCTA; spatially-homogeneous (average of sites) - for the same day, every ZCTA receives the same exposure assignment from the average of central monitors containing ZCTAs in Atlanta

* - estimates times 1,000

3.5 Discussion

This study quantifies the impacts of using different spatial resolutions of $PM_{2.5}$ exposure in a short-term health effect study of pediatric asthma emergency department visits in Atlanta.

Using ED visit data from 2002 to 2008, we do not find a statistically significant difference in effect estimates between ZCTAs classified by different poverty levels. We do note a stronger association in the highest and lowest SES ZCTAs, regardless of the exposure granularity (Figure 3.5), similar to results reported in O'Lenick et al. (2017). The observed differences may also be explained by varying baseline risk between the lowest and highest SES level groups, where similar additive risks will result in apparent differences in multiplicative risk measures.

One concern from previous studies of effect modification by SES is that observed differences in associations may be due to differential exposure measurement error. In our simulation study, we observe similar magnitude of attenuation for both the high and low SES neighborhoods when error-prone exposures are used; we also do not observe spurious associations when the effect is null. Hence, given Atlanta's spatial distribution of SES as measured by percent population below poverty and $PM_{2.5}$ monitor placements, we do not find evidence that observed effect modifications by SES may be due to exposure spatial misalignment.

When comparing spatially-varying and spatially-homogeneous exposures, odds ratios in ZCTAs with the highest levels of SES differ the most. This pattern is in agreement with bias we see in the high SES group from our simulation study. We also observe higher upper confidence limits for the spatially-varying exposure, consistent with simulation results.

This study utilizes $PM_{2.5}$ concentrations estimated through random forests and ensemble techniques at a fine spatial resolution that can reflect local $PM_{2.5}$ exposures. This exposure assessment approach is increasingly being used by environmental health researchers to investigate potential links between air pollution and adverse health events.

In our study, the optimal random forest hyperparameter combinations for the models

with and without AOD differ. This is likely due to AOD being a useful predictor of $PM_{2.5}$; when AOD is not a potential covariate, the model needs to rely on many more covariates to optimize prediction.

Combining predictions from both random forest models with and without AOD allows for more modeling flexibility based on data availability, as well as provide complete spatial and temporal coverage for exposure assessment. Table B.1 shows a slightly better performing model when AOD is not included, but this is may not be the case for other areas and years outside the modeling domain.

Methodology to characterize the uncertainty of predictions from random forests is a recent research focus. There is no "gold standard" method for random forest prediction standard error. We use naive estimates of uncertainty based on variability across trees; other options include more advanced Monte Carlo methods (Coulston et al.) or quantile regression methods (Meinshausen and Ridgeway) (2016; 2006). A naive estimate provided better coverage for our predictions. A more detailed exploration of the different options for random forest uncertainty in our analysis can be found in the Supplementary Materials.

Our results are limited to short-term exposures and pediatric asthma morbidity outcomes as measured by emergency department visits, and the results from this study may not be applicable to long-term exposures or other asthma-related outcomes in adults. We also only explore exposure measurement error due to spatial-misalignment, and we assume the health outcome is ascertained without error. Other simpler methods of assigning $PM_{2.5}$ to a spatial area exist that are not evaluated in this study. For example, Hart et al. (2015*b*) and Sarnat et al. (2010) assign monitor measurements to the nearest ZCTA and assess potential measurement error. Additionally, we are limited in our conclusions by the ZIP code-level ED data; we know patients' SES levels are heterogeneous within a ZIP code, and smaller scale spatial data would allow us to make more use of our spatially refined exposure data. Finally, our simulation and data analyses rely on having complete daily time-series of exposure data. Systematic missing data may bias effect estimates in the case-crossover framework when we

do not have complete time stratified controls. Janes et al. (2005) discuss these issues and explain the localizable, nonignorable nature of this referent selection.

Our simulation study is based on the motivating Atlanta case-crossover analysis, and results are dependent on the spatial distributions in $PM_{2.5}$ levels and ZCTA-level SES measures. Other simulation settings warrant further considerations. Particularly, the data from the simulation is created with a slightly different model (unconditional logistic regression) than the final analysis model (conditional logistic regression). While this approach allows us to simulate case data, there may be some residual bias. This could explain why our spatially-varying exposures are not always the top performing result across *all* performance statistics despite being the "true exposure."

We use a time-stratified approach for our case-crossover modeling, but other designs such as the symmetric bidirectional designs may also be explored for short-term health outcomes. (Carracedo-Martínez et al., 2010) Additionally, the conditional logistic regression and referent strategy we used could be replaced with another modeling strategy such as one that relaxes some distributional assumptions or expands the number or creation of control days.

In summary, this study assesses the effect of spatial misalignment bias on associations of $PM_{2.5}$ and asthma ED visits with effect modification by area-level SES through the use of real and simulated data. We recommend using spatially-varying $PM_{2.5}$ exposures when examining health effects by a spatially-varying effect modifier to minimize potential additional bias based on our simulation results. With finer resolution air pollution data becoming more widely accessible, future studies should take advantage of these data. Differences in the effects of $PM_{2.5}$ on asthma ED visits in subpopulations should continue to be investigated as the results have potential policy implications. Limiting harmful exposures to all populations should be a priority to reduce health inequities. (Adler and Newman, 2002)

Chapter 4

Combining Air Pollution Estimates from Multiple Models Using Bayesian Ensemble Averaging

4.1 Introduction

Particulate matter less than 2.5 μm in aerodynamic diameter, $PM_{2.5}$, is a type of air pollutant that globally harms human health. (Pui et al., 2014; Hart et al., 2015*a*; Maji et al., 2017) Availability of $PM_{2.5}$ data is critical to understand the impact of air pollution on human health. However, high-quality $PM_{2.5}$ levels are typically measured at regulatory monitoring stations, and due to the sparsity and preferential locations of monitoring stations, estimated $PM_{2.5}$ derived from observed measurements are often used in health analyses. (United States Environmental Protection Agency, 2009) Estimates can be obtained through averaging of monitoring data, e.g., Zeger et al. (2008), or performing spatial kriging, e.g., Wang et al. (2017). Other popular estimation techniques involve data fusion, where $PM_{2.5}$ estimates come from combining monitor measurements with other data sources. (Berrocal et al., 2010) Another estimation method gaining popularity among environmental researchers is random forests. (Hu et al., 2017; Brokamp et al., 2018; Huang et al., 2018) This method is appealing to researchers due to its ease of use and its performance compared to traditional methods. (Breiman, 2001) However, it has been demonstrated to underestimate more extreme values and potentially misrepresent the variability. (Berrocal et al., 2019)

Various methods exist to estimate air pollution levels across large spatial areas without monitors, but many methods lack the ability to quantify the uncertainty of those air pollution estimates. Statistical modeling makes uncertainty quantification possible. (Hoeting et al., 1999; Clyde, 2000)

In Murray et al. (2019), we used Bayesian model averaging to combine estimates of $PM_{2.5}$ from Bayesian hierarchical models with satellite derived aerosol optical depth and numerical model simulation as members to the ensemble models. (Murray et al., 2019) Combining these estimates allowed us to use all available information, while accounting for informative missingness associated with AOD. The ensemble estimates have better predictive performance than individual models in cross-validation experiments. By allowing spatially varying weights, we also found that Bayesian hierarchical models with numerical model

simulation (i.e., CMAQ) or remote sensing (i.e., AOD) as the main predictor may be favored more over the other in certain locations. However, in the two-member method, we did not examine how spatial or temporal covariates affect the ensemble weights.

The main contribution of this paper is to 1) extend the ensemble to more than two members and 2) include covariates into the weights. The weight estimation is accomplished through assigning a latent variable structure to the weights so that the weights given to each model member vary spatially. Estimation of weights is accomplished through the use of data augmentation with parameter expansion, a computational approach for efficient Bayesian analysis. This method of estimation allows us to incorporate covariates into the weight estimates. In the application, we combine the estimates from the two Bayesian spatial hierarchical models with either AOD or CMAQ as the main predictors and a third model member, estimates from random forests. Random forests predictively perform differently than the traditionally used models. Spatial statistical models have been shown to perform better for spatial interpolation, while random forest is better at explaining temporal variation. (Berrocal et al., 2019; Hu et al., 2017) Being able to combine estimates from approaches with different performance in space and time may result in better performing overall estimates.

4.2 Modeling

In our past work, we have detailed the usefulness of combining Bayesian hierarchical models in order to estimate $PM_{2.5}$. (Murray et al., 2019) In the current work, we achieve covariate incorporation through the use of multinomial probit modeling. While the goal of multinomial probit modeling is to model a categorical outcome, each outcome is assigned a probabilistic value of the model choosing that outcome. We are able to use these probabilities as our "weights" for a multi-member ensemble. We show the multinomial probit approach herein, but it can be easily reduced to binomial probit for a two statistical model member.

We first describe the ensemble model in a generalized form with J members:

$$p(y_{st} \mid M_1, M_2, ..., M_J) = w_s^{(1)} f_1(y_{st} \mid M_1) + w_s^{(2)} f_2(y_{st} \mid M_2) + ... + w_s^{(J)} f_J(y_{st} \mid M_J), \quad (4.1)$$

where y_{st} is the $PM_{2.5}$ concentration at location s on day t; $f_j(y_{st} \mid M_j)$ is the prediction distribution of the j^{th} member, which may correspond to the posterior predictive distribution of y_{st} from model M_j for Bayesian models or other ad-hoc distributions. For example, for random forest, one can define $f_j(y_{st} \mid M_j)$ as a normal distribution with mean $\mu_{st}^{(k)}$ and variance $\sigma_{st}^{2,(k)}$. Parameter $w_s^{(j)}$ is the non-negative weight for the jth model member at location s, with vector $\boldsymbol{w}_s = (w_s^{(1)}, w_s^{(2)}, ..., w_s^{(J)})$ and $\sum_{j=1}^{J} w_s^{(j)} = 1$. $\boldsymbol{w}_s$ is a probability vector from a multinomial probit model. $\boldsymbol{U}_s = (U_{s,1}, U_{s,2}, ..., U_{s,J-1})$ with $\boldsymbol{U}_s$ being multivariate normally distributed with mean $\boldsymbol{\mu}$ and variance Σ, then specifically define $w_s^{(J)} = \text{Prob}(\max(\boldsymbol{U}_s) < 0)$, and $w_s^{(j)} = \text{Prob}(\max(\boldsymbol{U}_s) = U_{s,j} > 0)$. Weights have a default value of $w_s^{(CMAQ)} = 1$ on locations where AOD is missing.

To link the ensemble model to the multinomial probit model, define Z_s as a multinomial outcome where each level of Z_s represents the best model to estimate $PM_{2.5}$ among three different prediction models and s = 1,...,N_S, where N_S is the total number of spatial locations. Each Z_s has an associated choice probability of $\boldsymbol{w}_s$ with $w_s^{(1)} + w_s^{(2)} + w_s^{(3)} = 1$. In order to estimate Z_s and, subsequently, $\boldsymbol{w}_s$, we use established data augmentation techniques where Z_s is modeled using a latent variable $\boldsymbol{U}_s$. (Tanner and Wong, 1987) We define $\boldsymbol{U}_s = \boldsymbol{X}_s\boldsymbol{\beta} + \boldsymbol{\epsilon}_s$, where $\boldsymbol{X}_s$ is constructed as a combination of choice-specific and individual-specific covariates; $\boldsymbol{\beta}$ are the coefficients corresponding to the $\boldsymbol{X}_s$ matrix, and $\boldsymbol{\epsilon}_s$ is the error term, distributed as $N(\boldsymbol{0}, \boldsymbol{\Sigma})$. $\boldsymbol{\Sigma}_{11}$ is set as 1 to avoid identifiability issues. (Imai and Van Dyk, 2005) The ordering of the choice outcomes in Z_s is irrelevant so long as the indexing is consistent, e.g., all indices of 1 correspond to the the first member model, 2 corresponds to the second member model, etc.

We assume $\boldsymbol{\beta}$ is distributed as $N(\boldsymbol{0}, A^{-1})$, where $\boldsymbol{0}$ is the prior mean, and A^{-1} is the prior

variance of $\boldsymbol{\beta}$. $\boldsymbol{\Sigma}$ is a positive definite 2×2 matrix with $\boldsymbol{\Sigma}_{11} = 1$. The prior for $\boldsymbol{\Sigma}$ is defined as

$$p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(\nu+3)/2}[\text{trace}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})]^{-\nu(2)/2}, \tag{4.2}$$

where ν is the prior degrees of freedom for $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$ is the prior scale of $\boldsymbol{\Sigma}$ with $S_{11} = 1$. The prior for Σ is proper if $\nu \geq p - 1 = 2$.

Typically, Z_s is observed in multinomial probit models. In our ensemble model, *both* Z_s and $\boldsymbol{U}_s$ are latent variables. The main motivation of estimating Z_s is to subsequently obtain an estimate of $\boldsymbol{w}_s$ to aid in combining estimates from ensemble members.

4.3 Estimation and Inference

We describe model estimation techniques of a three-member model given the motivating application. Similar approaches can be extended to an ensemble model with more than three members.

We use parameter-expanded data augmentation for our Markov chain Monte Carlo (MCMC) estimation. (Imai and Van Dyk, 2005; Schliep and Hoeting, 2015) The parameter expansion helps with algorithm efficiency using an unidentified parameter, α. (Meng and Van Dyk, 1999) Details on sampling with the use of parameter expansion are below.

4.3.1 Covariate Matrix Construction and Dimension-Reduced Parameters

We introduce two different types of covariates to consider for these models: individual-specific and choice-specific. Individual-specific covariates do not vary across choices. They are the same for each choice but differ among individuals (or in our application, spatial locations). The intercept is also treated as "individual-specific" due to choices having shared $\boldsymbol{\beta}$'s.

The individual-specific matrix here will be $N_S \times m^*$, where m^* is the number of individual-specific covariates. Choice-specific covariates differ among individuals (i.e., locations) and choice (i.e., models). The choice-specific matrix here will be $(N_S \times p) \times m$, where m is the number of choice-specific covariates, and p is the number of total choices (before reducing to $p-1$). Keane (1992) makes the case for at least one choice-specific covariate in a covariate matrix in addition to individual-specific covariates. Quinn et al. (1999) offers an explanation on construction of $\boldsymbol{X}$ with both choice-specific and individual-specific covariates as a stacked $\boldsymbol{X}$ matrix; we follow this construction herein. We consider three different covariate combinations: (1) intercept + choice-specific covariate, (2) intercept + choice-specific covariate + individual-specific covariate, and (3) intercept + choice-specific covariate + individual-specific covariate + latitude + longitude.

Both $\boldsymbol{U}$ and $\boldsymbol{X}$, and by relation $\boldsymbol{U}_s$ and $\boldsymbol{X}_s$, are dimension reduced parameters that are created by using the third model member's values as a "reference" group. Specifically, if we let $\boldsymbol{U}^+$ and $\boldsymbol{X}^+$ be the non-dimension reduced values, we can write $\boldsymbol{U}_s$ as
$U_{s,j} = U^+_{s,j} - U^+_{s,3}$, j=1,2, where $\boldsymbol{U}$ is a $(N_S \times 2)$ matrix, $\boldsymbol{U}^+$ is a $(N_S \times 3)$ matrix, and $\boldsymbol{U}_s$ is a (1×2) matrix. The construction of $\boldsymbol{X}^+$ and, subsequently, $\boldsymbol{X}$ follows. The individual-specific covariate matrix will have dimension $N_S \times$ m* before expanding it to combine with the choice-specific covariate matrix, where it will become $(N_S \times 3{=}183) \times (m^* \times 3)$ based on cross multiplying the original individual-specific covariate matrix by I_p, the identity matrix. Then, deleting every p^{th} row and column, as in Quinn et al. (1999), will result in dimension $(N_S \times 2{=}122) \times (m^* \times 2)$.

Any choice-specific covariate matrix will be of dimension $(N_S \times 3 = 183) \times m$ before reducing the dimension to $(N_S \times (p-1)) \times m$, which will result in dimension $(N_S \times 2 = 122) \times m$.

When we stack the individual-specific covariate matrix and the choice-specific covariate matrix, we ultimately will have an $\boldsymbol{X}$ matrix that is of dimension $(N_S \times 2 = 122) \times l$, where l is the result of combining columns, $m^* + m$. Individually, when we see $\boldsymbol{X}_s$ referenced in

the model, $\boldsymbol{X}_s$ is of dimension $(2 \times l)$.

$\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are also affected by the dimension reduction where $\boldsymbol{\beta}$ becomes $(p-1) \times 1$, and $\boldsymbol{\Sigma}$ becomes $(p-1) \times (p-1)$.

4.3.2 Estimation Algorithm with Parameter Expansion

Before beginning the algorithm, we first offer some notational clarity. Much of the notation used mirrors that of Imai and Van Dyk (2005).

Asterisk (*) indicates an intermediate step and that sampled value is not saved for posterior inference. Tilde ($\tilde{}$) indicates an unidentified parameter, i.e., a parameter altered by the parameter expansion, α^2. Overall, our MCMC involves sequential sampling of: $\boldsymbol{U}_s^*$, α^{2*} $\big($both $(\alpha^2)^{*(t-1)}$ and $(\alpha^2)^{*(t)}\big)$, $\tilde{U}_s^*$, $\tilde{\boldsymbol{\beta}}^*$, $\tilde{\boldsymbol{\Sigma}}^*$, and $\tilde{\boldsymbol{S}}$.

Data augmentation with parameter expansion allows us to more efficiently sample our parameters. When we introduce the working parameter, α, following Imai and Van Dyk, we can use any working prior distribution. α_0^2, a positive constant, is the prior for α.

Under parameter expansion, set $\alpha^2 = \tilde{\boldsymbol{\Sigma}}_{11}$ and $\boldsymbol{\Sigma} = \tilde{\boldsymbol{\Sigma}}/\tilde{\boldsymbol{\Sigma}}_{11}$ to obtain
$p(\boldsymbol{\Sigma}, \alpha^2) \propto |\boldsymbol{\Sigma}|^{-(\nu+3)/2} \text{exp}[-\frac{\alpha_0^2}{2\alpha^2} \text{trace}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})](\alpha^2)^{-[\nu(2)/2+1]}$, where
$\alpha^2|\boldsymbol{\Sigma} \sim \alpha_0^2 \text{trace}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})/\chi^2_{\nu(p-1=2)}$.
Finally, by combining the priors for $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, and the joint distribution of $(\boldsymbol{\Sigma}, \alpha^2)$, then transforming to $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Sigma}})$, we have $\tilde{\boldsymbol{\beta}}|\tilde{\boldsymbol{\Sigma}} \sim N(\boldsymbol{0}, \tilde{\boldsymbol{\Sigma}}_{11} A^{-1})$ with $\tilde{\boldsymbol{\Sigma}} \sim$ inv Wishart$(\nu, \tilde{\boldsymbol{S}})$ with $\tilde{\boldsymbol{S}} = \alpha_0^2 S$.
Begin algorithm:

1. Draw $\left((\alpha^2)^*, \tilde{\boldsymbol{U}}_s^*\right)$ via $p(\alpha^2, \boldsymbol{U}_s|\boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1}, Z_s)$

   (a) Draw $(\alpha^2)^*$ from $p(\alpha^2|\boldsymbol{\Sigma}) \sim \alpha_0^2 \text{trace}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})/\chi^2_{\nu(p-1=2)}$.

   (b) Draw $N_S \times 2$ conditionals of $\boldsymbol{U}^*_{s,j}|\boldsymbol{U}_{s,-j}, \boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1}, Z_s$ for $s = 1, ..., N_S$ and $j = 1, 2$. Specifically, draw

      i. $\boldsymbol{U}^*_{s,1}|\boldsymbol{U}_{s,2}, \boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1}, Z_s$

ii. $\boldsymbol{U}^*_{s,2}|\boldsymbol{U}_{s,1}, \boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1}, Z_s$

where the conditionals are truncated univariate normal (details in Appendix C.)

Set $\tilde{\boldsymbol{U}}^*_{s,j} = \alpha^* \boldsymbol{U}^*_{s,j}$, where $\alpha^* = \sqrt{(\alpha^2)^*}$.

2. Draw $\left((\alpha^2)^*, \tilde{\boldsymbol{\beta}}^*\right)$ via $p(\alpha^2, \boldsymbol{\beta}|\tilde{\boldsymbol{U}}^*_s, \boldsymbol{\Sigma}^{-1}, Z_s)$

   (a) Draw $\alpha^{2*}|\tilde{\boldsymbol{U}}^*, \boldsymbol{\Sigma}^{-1}$

   $$\alpha^{2*}|\tilde{\boldsymbol{U}}^*, \boldsymbol{\Sigma}^{-1} \sim \frac{\sum_{s=1}^{N_S}(\tilde{\boldsymbol{U}}^*_s - \boldsymbol{X}_s\hat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{U}}^*_s - \boldsymbol{X}_s\hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\beta}}^T\boldsymbol{A}\hat{\boldsymbol{\beta}} + \text{trace}[\tilde{\boldsymbol{S}}\boldsymbol{\Sigma}^{-1}]}{\chi^2_{(N_S+\nu)(p-1=2)}}$$

   where

   $$\hat{\boldsymbol{\beta}} = \left[\sum_{s=1}^{N_S} \boldsymbol{X}_s^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_s + \boldsymbol{A}\right]^{-1}\left[\sum_{s=1}^{N_S} \boldsymbol{X}_s^T\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{U}}^*_s\right]$$

   (b) Draw $\tilde{\boldsymbol{\beta}}^*|\tilde{\boldsymbol{U}}^*, \boldsymbol{\Sigma}^{-1} \sim \text{Normal}\left[\hat{\boldsymbol{\beta}}, (\alpha^2)^*\left(\sum_{s=1}^{N_S} \boldsymbol{X}_s^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_s + \boldsymbol{A}\right)^{-1}\right]$

   Then, set $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}^*/\alpha^*$

3. Draw $\left((\alpha^2)^*, (\tilde{\boldsymbol{\Sigma}})^{-1*}\right)$ via $p(\alpha^2, \boldsymbol{\Sigma}^{-1}|\tilde{\boldsymbol{U}}^*_s, \boldsymbol{\beta}, Z_s)$

   $(\tilde{\boldsymbol{\Sigma}})^{-1*}|\boldsymbol{\beta}, (\tilde{\boldsymbol{U}}^*_s - \boldsymbol{X}_s\tilde{\boldsymbol{\beta}}^*) \sim \text{Wishart}\left[N_S + \nu, \{\tilde{\boldsymbol{S}} + \sum_{s=1}^{N_S}(\tilde{\boldsymbol{U}}^*_s - \boldsymbol{X}_s\tilde{\boldsymbol{\beta}}^*)(\tilde{\boldsymbol{U}}^*_s - \boldsymbol{X}_s\tilde{\boldsymbol{\beta}}^*)^T\}^{-1}\right]$

   For $\tilde{\boldsymbol{\Sigma}}^*_{11}$ to meet the constraints of the $\boldsymbol{\Sigma}$ prior, $\tilde{e}_s(\tilde{\boldsymbol{\Sigma}}^*_{11}) = e_s + (\tilde{\boldsymbol{\Sigma}}^*_{11})\boldsymbol{X}_s\boldsymbol{\beta}$ must satisfy

   $$\begin{cases} \max\{\tilde{e}_{s1}(\tilde{\boldsymbol{\Sigma}}^*_{11}), ..., \tilde{e}_{sj}(\tilde{\boldsymbol{\Sigma}}^*_{11})\} < 0 \text{ if } Z_s = 0 \\ \max\{0, \tilde{e}_{s1}(\tilde{\boldsymbol{\Sigma}}^*_{11}), ..., \tilde{e}_{sj}(\tilde{\boldsymbol{\Sigma}}^*_{11})\} = \tilde{e}_{sk}(\tilde{\boldsymbol{\Sigma}}^*_{11}) \text{ if } Z_s = k \end{cases}$$

   Jiao and van Dyk (2015) recommend simple rejection sampling, for which details can be found in Appendix C.

   Set $\alpha^2 = \tilde{\boldsymbol{\Sigma}}^{2*}_{11}$ (Note: $\tilde{\boldsymbol{\Sigma}}^{2*}_{11}$ comes from the first diagonal element of $\tilde{\boldsymbol{\Sigma}}^*$ NOT $\tilde{\boldsymbol{\Sigma}}^{-1*}$.)

   Set $\boldsymbol{\Sigma}^{-1} = \tilde{\boldsymbol{\Sigma}}^{-1*} \times \alpha^2$

4. Draw Z_s via $\text{p}(Z_s|\boldsymbol{w}_s)$ for $s = 1, \ldots, N_S$.

Calculate $\boldsymbol{w_s}$ via $\mathrm{p}(\boldsymbol{w_s}|\boldsymbol{\beta}, \boldsymbol{\Sigma})$

$\boldsymbol{U}_s \sim \mathrm{Normal}(\boldsymbol{X}_s\boldsymbol{\beta}, \Sigma)$

$w_s^{(1)} = Pr(\boldsymbol{U}_{s,1} > \boldsymbol{U}_{s,2}, \boldsymbol{U}_{s,1} > 0)$

$w_s^{(2)} = Pr(\boldsymbol{U}_{s,1} < \boldsymbol{U}_{s,2}, \boldsymbol{U}_{s,2} > 0)$

$w_s^{(3)} = Pr(max(\boldsymbol{U}_{s,1}, \boldsymbol{U}_{s,2}) < 0)$

$\mathrm{p}(Z_s|\boldsymbol{w}_s) \propto p_{s,1} * p_{s,2} * p_{s,3}$, where

$$p_{s,j} = \mathrm{Prob}(Z_s = j|\cdot) = \frac{w_s^{(j)} \prod_{t=1}^{T} \phi(y_{st}; \mu^{(j)}, \sigma^{2,(j)})}{\sum_{j=1}^{3} w_s^{(j)} \prod_{t=1}^{T} \phi(y_{st}; \mu^{(j)}, \sigma^{2,(j)})}$$

where $\mu^{(j)}$ and $\sigma^{2,(j)}$ correspond to the predictive probability distribution mean and variance, respectively, and $\phi \equiv$ a Normal distribution.

End algorithm.

4.3.3 Inference

The point prediction of y_{st} can be defined by its posterior mean

$$\hat{y}_{st} = w_s^{(1)}\mu_{st}^{(1)} + w_s^{(2)}\mu_{st}^{(2)} + w_s^{(3)}\mu_{st}^{(3)}, \tag{4.3}$$

which is a weighted average of predictions from the three members. The error for y_{st} is defined as

$$\mathrm{Var}(y_{st}|\mu_{st}^{(1)}, \mu_{st}^{(2)}, \mu_{st}^{(3)}) = w_s^{(1)}\sigma_{st}^{2,(1)} + w_s^{(2)}\sigma_{st}^{2,(2)} + w_s^{(3)}\sigma_{st}^{2,(3)} + w_s^{(1)}\mu_{st}^{2,(1)} + w_s^{(2)}\mu_{st}^{2,(2)} + w_s^{(3)}\mu_{st}^{2,(3)} - \left(w_s^{(1)}\mu_{st}^{(1)} + w_s^{(2)}\mu_{st}^{(2)} + w_s^{(3)}\mu_{st}^{(3)}\right)^2$$

Bayesian inference also allows us to capture the uncertainty in the weight estimation procedure. Using Bayesian techniques, one can report any appropriate summary measures to describe the centrality and variation in the weight estimation.

4.4 Application

We obtained daily ground-level 24-hour average measurements of $PM_{2.5}$ from 63 monitors in the Southeastern US over the period 2003 to 2005 via the United States Environmental Protection Agency's Air Quality System (AQS). To supplement the monitoring data, chemical transport model (CTM) simulations were obtained from the USEPA Models-3/Community Multiscale Air Quality (CMAQ) model version 4.5 at a 12 km $\times$ 12 km horizontal spatial resolution. (Byun and Schere, 2006) Satellite-retrieved aerosol optical depth (AOD) measurements by the aerosol remote sensor Moderate Resolution Imaging Spectroradiometer (MODIS), which orbits the Earth on the National Aeronautics and Space Administration's Aqua and Terra satellites, are also available. We utilized a new multiangle implementation of atmospheric correction (MAIAC) algorithm that provides AOD values at a 1km $\times$ 1km spatial resolution. (Lyapustin, Martonchik, Wang, Laszlo and Korkin, 2011; Lyapustin, Wang, Laszlo, Kahn, Korkin, Remer, Levy and Reid, 2011) For each AOD grid cell, we also compiled variables including: elevation from the US Geological Survey, forest cover and road lengths from the 2001 National Land Cover data, meteorology (e.g. wind speed) from the North American Land Data Assimilation Systems, and $PM_{2.5}$ primary emission point sources from the 2002 USEPA National Emissions Inventory. As in Hu et al., forest cover and elevation were averaged from their original resolutions of about 1 km and about 30 m, respectively, to the 1 km $\times$ 1 km MAIAC grid cell level. (Hu et al., 2013) Additionally, road lengths and point emissions were summed over the 1 km $\times$ 1 km MAIAC grid cell level.

4.4.1 Random Forest

Random forest is a decision tree-based approach in which many small decision trees are grown and then aggregated to provide a single prediction averaged across trees. Random forest can be preferable to regression-based approaches due to its ability to better accommodate non-linearity and general flexibility in specifying model inputs. Additionally, integral to our

ensemble approach, prediction uncertainties can be derived for random forest predictions.

Our random forest incorporates 1 km-resolution spatial information such as latitude and longitude, percent forest cover and lengths of major and local road types from the 2001 National Land Cover data, daily meteorology (temperature, wind speed, and relative humidity) from the North American Land Data Assimilation Systems, and $PM_{2.5}$ primary emission point sources and distance from the emission point source from the 2002 USEPA National Emissions Inventory. We also include planetary boundary layer height from the North American Regional Reanalysis. Finally, we input CMAQ, AOD, and time ID values.

The function randomForest in R has parameters that the user can specify, e.g., number of trees, number of variables to consider at each split in a node, and minimal node size. We ultimately use random forests with 500 trees, 6 variables considered at each split, and a minimum of 5 terminal nodes, all of which is the default. The point predictions and standard deviations for the random forests result from taking the mean and standard deviation, respectively, of the 500 trees at each observation. (Breiman, 2001)

4.4.2 Bayesian Hierarchical Modeling for Daily $PM_{2.5}$

We have previously described the model for combining monitoring data with CMAQ outputs or AOD retrievals as predictors for point-referenced AQS monitoring measurements in a Bayesian spatial-temporal hierarchical model (BHM). (Murray et al., 2019)

Let $Y(\boldsymbol{s}, t)$ represent the observed $PM_{2.5}$ concentration on day t at locations $\boldsymbol{s}$. Following Berrocal et al. and Chang et al., our statistical model has the form of a BHM:

$$Y(\boldsymbol{s}, t) = \alpha_1(\boldsymbol{s}) + \alpha_2(\boldsymbol{s})X(\boldsymbol{s}, t) + \beta_1(t) + \beta_2(t)X(\boldsymbol{s}, t) + \boldsymbol{Z}(\boldsymbol{s}, t)\boldsymbol{\gamma} + \epsilon(\boldsymbol{s}, t), \tag{4.4}$$

where $X(\boldsymbol{s}, t)$ is the linked AOD or CMAQ values in the grid cells containing the monitor at locations $\boldsymbol{s}$, and $\boldsymbol{Z}(\boldsymbol{s}, t)$ is a vector of additional predictors with coefficient $\boldsymbol{\gamma}$. (Berrocal et al., 2010; Chang et al., 2014) For the AOD model, $\boldsymbol{Z}(\boldsymbol{s}, t)$ includes the following land use and

meteorology variables: elevation, forest cover, road length, primary emission source, wind speed, and temperature. Because CMAQ uses information on emissions and meteorology to perform simulations, $\boldsymbol{Z}(\boldsymbol{s}, t)$ is not included in the $PM_{2.5}$-CMAQ BHM. Preliminary analysis also showed that including additional covariates does not improve prediction performance for the CMAQ model. Finally, the residual error term, $\epsilon(\boldsymbol{s}, t)$, is independent normally distributed with mean zero and variance σ_y^2.

Parameters $\alpha_1(s)$, $\alpha_2(s)$ and $\beta_1(t)$, $\beta_2(t)$ in Equation (4.4) are spatial and temporal random effects, sometimes referred to as calibration parameters because they correct for the additive and multiplicative bias associated with CMAQ or AOD. Additional details about the modeling assumptions for BHM can be found in Murray et al. (2019).

4.4.3 Estimation and Prediction

Estimation and prediction are accomplished in three stages. In Stage I, we fit the CMAQ and the AOD Bayesian hierarchical models and the random forest to obtain posterior predictive means, $\mu_{st}^{(j)}$, and variances for each observed $PM_{2.5}$ value. To avoid overfitting, each observation was left-out and back-predicted in a *cross-validation* experiment, similar to approaches employed in stack regression. (LeBlanc and Tibshirani, 1996) In Stage II, we fit the proposed ensemble model to estimate covariate-dependent weights using the posterior predictive distributions from Stage I. Finally, in Stage III, the CMAQ and the AOD Bayesian hierarchical models and the random forest are fitted again with all $PM_{2.5}$ observations. Predictions are made at all locations and then combined using ensemble weights obtained based on parameter estimates from Stage II.

We consider two types of cross-validation experiments to assess model performance. First, we use a traditional ten-fold cross-validation technique to assess prediction. This requires us to separate the data into 10 different subsets with 90% of the data used as training data and the other 10% of the data used as validation data within each subset. This allows us to evaluate prediction performance in scenarios with temporally missing observations

of $PM_{2.5}$. Then, in order to evaluate the prediction performance of the proposed method at locations absent of ground monitors, especially over large geographical areas, we use a spatially clustered cross-validation technique. This cross-validation most closely mimics the real world scenario of needing to estimate ambient air pollution in areas without monitoring data. (Young et al., 2016) Clusters are formed through hierarchical clustering based on distance. (Johnson, 1967) We leave out one spatial cluster at a time (validation data) and use the other nineteen clusters as training data.

We run 2500 MCMC iterations with a burn-in of 1000 for cross-validation results. We use R version 3.5.1 for all estimation and prediction. (R Core Team, 2018)

Out-of-sample prediction performance is based on four statistics: prediction root-mean-square error (RMSE), 95% coverage probability of the posterior intervals (PI), average posterior standard deviation (SD), and R^2. R^2 and RMSE were calculated based on posterior predictive means of the left-out observed $PM_{2.5}$ concentrations. Posterior prediction intervals were based on the 2.5^{th} and the 97.5^{th} percentiles of the posterior distribution of the two-component predictive model distribution in Equation (4.1).

We apply the ensemble model with covariate-dependent weights to the $PM_{2.5}$ data in the southeastern United States. The $PM_{2.5}$ measurements are present at 63 monitoring stations. (Murray et al., 2019) When choosing covariates that inform the weights, we only consider covariates that are present at all spatial locations where we estimate $PM_{2.5}$ to avoid unnecessary computation, (i.e., interpolation). The covariates we examine as potential weight predictors are: i. (standardized) elevation, ii. (standardized) percent forest coverage, iii. average estimated $PM_{2.5}$, defined as the three-year average estimated PM per model, and iv. above average estimated $PM_{2.5}$ per model, defined as whether the three-year average estimated $PM_{2.5}$ is higher than the USEPA's National Ambient Air Quality Standards (NAAQS) from the 2006 annual threshold of 15 $\mu g/m^3$.

4.4.4 Cross-Validation Results

Table 4.1shows the relative performance of the ensemble model compared to individual models for spatially clustered cross-validation. Similarly, Table 4.2 shows the same comparisons but for ten-fold cross-validation. We see similar statistics for spatially clustered CV ensemble results (Table 4.1) compared to ten-fold CV ensemble results (Table 4.2).

For the spatially clustered cross-validation, the proposed three-member ensemble has an improved RMSE of 2.99 compared to the RMSE of 3.13 of the two-member ensemble model. The three-member ensemble also has a slightly higher R^2 of 0.83 compared to the two-member ensemble's R^2=0.81. We see little improvement in the RMSE (0.7% decrease) in the ordinary ten-fold cross-validation experiment compared to the two-member ensemble while the R^2 remains the same for the two ensembles. The three-member ensemble model has a higher overall standard deviation than the two-member ensemble model (3.91 vs. 3.43). The best ensemble model has weights dependent on elevation and estimated three-year $PM_{2.5}$ average at each location.

Spatially clustered CV results informed which models we compared for ten-fold CV results. We see little, if any, improvement in the RMSE (2.98 vs. 3.00) in the ten-fold CV experiment. Similar to the spatially clustered CV experiment, the overall standard deviation is higher than that of the two-member ensemble model.

4.4.5 Spatial Results

In Figure 4.1, we show an example of the spatially-varying weights based on using percent forest cover and average estimate over EPA-standard as covariates. This figure illustrates the spatially-varying performance of the three models, as evidenced by the estimated weights at monitoring locations. Random forest receives the majority of the weight in most areas, but some areas may slightly favor (> 0.33) the $PM_{2.5}$-AOD BHM.

Table 4.1: Prediction performance for daily $PM_{2.5}$ concentrations in spatial clustering cross-validation (CV) comparing ensemble averaging with a Bayesian hierarchical model (BHM) using satellite-derived aerosol optical depth (AOD), a BHM using a numerical model (CMAQ) simulation, or random forest. Ensemble inputs were derived from first performing 10-fold CV. Ensembles are based on 2500 runs with 1000 burn in using the mean weight from the runs.

Method	RMSE	Coverage of 95% PI	Average Posterior SD	R^2
$PM_{2.5}$-AOD BHM	3.62	94.43	3.59	0.74
$PM_{2.5}$-CMAQ BHM	3.93	93.34	3.58	0.69
Random Forest	3.83	96.86	4.17	0.71
Ensemble (Hi/low PM)	2.98	98.14	3.93	0.83
Ensemble (Avg PM from Model)	2.98	98.13	3.93	0.83
Ensemble (Forest cover + Hi/low PM)	3.00	98.16	3.91	0.82
Ensemble (Forest cover + Avg PM from Model)	3.01	98.18	3.91	0.82
Ensemble (Elevation + Hi/low PM)	2.99	98.13	3.92	0.83
Ensemble (Elevation + Avg PM from Model)	2.99	98.12	3.92	0.83
Ensemble (Lat/Long + Hi/low PM)	3.01	98.04	3.91	0.82
Ensemble (Lat/Long + Avg PM from Model)	3.01	98.12	3.91	0.82
Ensemble (Lat/Long + Elevation + Avg PM from Model)	3.03	98.09	3.90	0.82

RMSE: root mean squared error; PI: prediction interval; SD: standard deviation; CV: cross-validation; $PM_{2.5}$: particulate matter less than 2.5 μm; AOD: aerosol optical depth; BHM: Bayesian hierarchical model; CMAQ: Community Multiscale Air Quality; hi/low PM: indicator variable for estimated three-year average $PM_{2.5}$ value above the 2006 NAAQS annual threshold from each of the models (choice-specific variable); Avg PM from Model: estimated three-year average $PM_{2.5}$ value from each of the models (choice-specific variable).

Table 4.2: Prediction performance for daily $PM_{2.5}$ concentrations in ordinary (ten-fold) cross-validation (CV) comparing ensemble averaging with a Bayesian hierarchical model (BHM) using satellite-derived aerosol optical depth (AOD), a BHM using a numerical model (CMAQ) simulation, or random forest. Ensemble inputs were derived from first performing 10-fold CV. Ensembles are based on 2500 runs with 1000 burn in using the mean weight from the runs.

Method	RMSE	Coverage of 95% PI	Average Posterior SD	R^2
$PM_{2.5}$-AOD BHM	3.40	94.07	3.30	0.78
$PM_{2.5}$-CMAQ BHM	3.14	95.05	3.28	0.81
Random Forest	3.30	97.79	3.89	0.79
Ensemble (Hi/low PM)	2.98	98.19	3.93	0.83
Ensemble (Avg PM from Model)	2.97	98.19	3.92	0.83
Ensemble (Forest cover + Avg PM from Model)	2.99	98.20	3.91	0.83
Ensemble (Elevation + Avg PM from Model)	2.98	98.13	3.92	0.83
Ensemble (Lat/Long + Avg PM from Model)	3.00	98.12	3.91	0.83
Ensemble (Lat/Long + Elevation + Avg PM from Model)	3.01	98.10	3.90	0.82

RMSE: root mean squared error; PI: prediction interval; SD: standard deviation; CV: cross-validation; $PM_{2.5}$: particulate matter less than 2.5 μm; AOD: aerosol optical depth; BHM: Bayesian hierarchical model; CMAQ: Community Multiscale Air Quality; hi/low PM: indicator variable for estimated three-year average $PM_{2.5}$ value above the 2006 NAAQS annual threshold from each of the models (choice-specific variable); Avg PM from Model: estimated three-year average $PM_{2.5}$ value from each of the models (choice-specific variable).

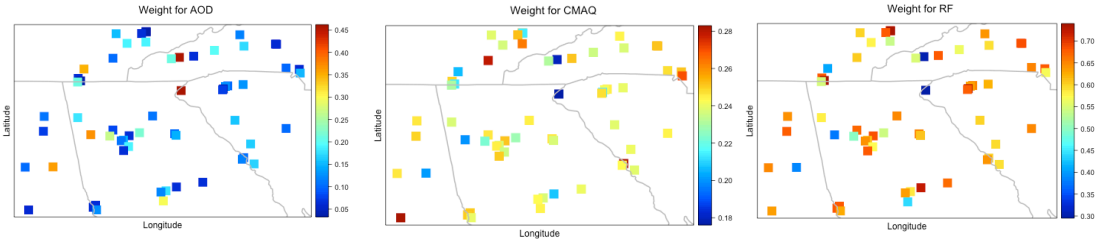

Figure 4.1: Estimated weights using percent forest cover and average estimate over EPA standard as covariates. Left to right: estimated weights at monitoring locations for $PM_{2.5}$-AOD Bayesian Hierarchical Models (AOD), estimated weights at monitoring locations for $PM_{2.5}$-CMAQ Bayesian Hierarchical Models (CMAQ), and estimated weights at monitoring locations for random forest (RF).

4.5 Discussion

The new Bayesian ensemble approach allows for the integration of multiple $PM_{2.5}$ estimates obtained from different prediction models. Unlike previous ensemble approaches, our method provides estimates of prediction uncertainty, as well as identifies the relative spatially-varying predictive ability of each ensemble member based on the estimated ensemble weights. Ultimately, the ensemble approach may improve health effect estimation and health impact assessment by improving exposure estimation with complete spatial-temporal coverage.

When combining estimates from different models, especially those without ground monitoring measurements, it can be useful to incorporate covariates into ensemble weight estimates. Previous data fusion work has successfully combined multiple numerical models but not different types of estimation inputs such as satellite data and numerical model outputs. (Crooks and Isakov, 2013) The ensemble approach helps to identify which model is performing the strongest at a certain location through the weight estimation.

The parameter estimates obtained from the post-burn-in MCMC runs to derive the weights can be used to extrapolate the weights to spatial locations without monitors. Note that the β values do *not* correspond to subsequent increases or decreases in ensemble weights. The βs do, however, indirectly relate to the weights through U_s.

Although the method demonstrated here shows small improvements, these relative performances are limited to the southeastern United States. The models that make up the ensemble method typically perform well in this region of the United States. Other areas of the United States where certain member models perform relatively worse than other member models should be studied. (Geng et al., 2018)

Note that our algorithm has a prior mean of $\boldsymbol{\beta}$ set to $\mathbf{0}$. Algorithm 2 in Imai and Van Dyk (2005) allows for a non-zero prior mean, and the rest of our algorithm can be adjusted accordingly.

While we choose to use the $\boldsymbol{\Sigma}_{11}$ restriction, there are other options to maintain identifiability; another popular restriction is from Burgette and Nordheim where the trace of the $\boldsymbol{\Sigma}$

matrix is restricted to be a constant. (Burgette and Nordheim, 2012)

For critics of the random forest and its lack of spatial dependency compared to the spatial Bayesian hierarchical models, the results may be unsatisfactory. While random forests do not have a spatial modeling component, they do have the ability to include spatial covariates such as latitude, longitude, monitor ID, and land-use covariates.

Some readers may question the multiple uses of covariates in both the ensemble modeling stage and the weighting estimation stages. However, the covariate usage serves two different purposes. The covariates are intended to help predict PM in the modeling phase while covariates are helping to choose a best model in the weighting stage. Therefore, the same covariate can be comfortably used in both stages if desired.

While spatial extensions of the multinomial probit model exist, we chose to evaluate the usefulness of the non-spatial multinomial probit method; however, the covariates we used spatially vary. The algorithm can be adjusted to incorporate a spatial structure as well.

When looking beyond three model members, even more efficient algorithms beyond the use of multinomial probit may used in the same way, e.g., the Diagonal Orthant Multinomial Probit Models. (Johndrow et al., 2013)

Bayesian hierarchical models allow us to use data such as AOD or CMAQ as covariates in a model to predict $PM_{2.5}$ while also incorporating structures for spatial and temporal dependence. New to the proposed method herein is the inclusion of the machine learning approach, random forests. Unlike Bayesian hierarchical models, random forests do not have spatial components built in, but it is still able to perform well based on its ability to build numerous decision trees to identify which model has the best prediction performance. (Breiman, 2001)

Incorporating covariate information into the weights provided more information about what characteristics influence a higher weight for each predictive member and, therefore, provided more information to easily interpolate the weights to locations without monitors in future analyses.

Appendix A

Appendix for Chapter 2

A.1 Bayesian Hierarchical Modeling Details

$$Y(\boldsymbol{s},t) = \alpha_1(\boldsymbol{s}) + \alpha_2(\boldsymbol{s})X(\boldsymbol{s},t) + \beta_1(t) + \beta_2(t)X(\boldsymbol{s},t) + \boldsymbol{Z}(\boldsymbol{s},t)\boldsymbol{\gamma} + \epsilon(\boldsymbol{s},t), \tag{A.1}$$

Temporal dependence in $\beta_1(t)$ and $\beta_2(t)$ is modeled using a first-order conditional autoregressive model (CAR). The CAR model is defined via temporal adjacencies. Let $t \sim t'$ indicate that days t and t' are 1 day apart. The full conditional distribution of $\beta_1(t)$ is Gaussian with $E[\beta_1(t)] = \eta_{\beta_1} \sum_{t' \sim t} \beta_{1,t'}/n_t$ and $\text{Var}[\beta_1(t)] = \sigma^2_{\beta_1}/n_t$, where n_t is the number of temporal neighbors and $\eta_{\beta_1} \in [0,1]$ controls the degree of temporal dependence. Temporal random slopes $\beta_2(t)$ for the Community Multiscale Air Quality modeling system (CMAQ) or aerosol optical depth (AOD) are defined similarly. Spatial dependence in $\alpha_1(\boldsymbol{s})$ and $\alpha_2(\boldsymbol{s})$ is modeled jointly using a linear coregionalization model. Specifically, we assume $(\alpha_1(\boldsymbol{s}), \alpha_2(\boldsymbol{s}))^T = \mathbf{A}\mathbf{v}(\boldsymbol{s})$, where $\mathbf{A}$ is a 2×2 lower triangular matrix, and $\mathbf{v}(\boldsymbol{s})$ is a 2×1 vector $(v_1(\boldsymbol{s}), v_2(\boldsymbol{s}))^T$, where $v_1(\boldsymbol{s})$ and $v_2(\boldsymbol{s})$ represent two latent independent Gaussian processes with marginal variances of 1 and exponential covariance functions with range parameters θ_j, i.e. $\text{Cov}(v_i(\boldsymbol{s}), v_i(\boldsymbol{s'})) = e^{-||\boldsymbol{s}-\boldsymbol{s'}||/\theta_j}$ for $j = 1, 2$.

A.2 Inference Details

A.2.1 Bayesian Hierarchical Model Inference

For the Bayesian Hierarchical Models (BHM), each component of the fixed effect $\boldsymbol{\gamma}$ is assigned a flat prior ($\propto 1$), and each element of $\mathbf{A}$ is assigned $N(0, 1 \times 10^3)$. The BHM's temporal CAR parameters η_{β_1} and η_{β_2} are discretized into 1,000 intervals in $[0, 1]$. Variance components (marginal variances for the Gaussian process, τ^2; BHM's residual error variance, σ_y^2; conditional variance of the temporal CAR model, $\sigma_{\beta_1}^2$ and $\sigma_{\beta_2}^2$) are assigned Inverse-Gamma (a, b), with a and b chosen to be small and non-informative. Range parameters for Gaussian processes were assigned Gamma (0.5, 0.05).

A.2.2 Ensemble Model Inference

Here we only present the Markov Chain Monte Carlo (MCMC) algorithm for the ensemble model.

We first introduce a latent variable z_{st}, where $z_{st} = 1$ if the prediction from CMAQ performs superiorly to AOD, and $z_{st} = 0$ otherwise, since z_{st} is Bernoulli. After initialization, we update z_{st}, w_s, τ^2, and ρ as follows.

1. Update z_{st} for $s = 1, \ldots, S$ and $t = 1, \ldots, T_s$. The full conditional distribution of $z_{st} \sim$ Bernoulli with probability

$$\frac{w_s * \phi\left(y_{st} \mid \mu_{st}^{(1)}, \sigma_{st}^{2,(1)}\right)}{w_s * \phi\left(y_{st} \mid \mu_{st}^{(1)}, \sigma_{st}^{2,(1)}\right) + (1 - w_s) * \phi\left(y_{st} \mid \mu_{st}^{(2)}, \sigma_{st}^{2,(2)}\right)}.$$

2. Update w_s for $s = 1, \ldots, S$. At the r^{th} iteration, generate a proposal $q_s^{(r)} = \text{logit}(w_s^{(r)})$ from a Normal distribution with mean $q_s^{(r-1)}$ and variance κ_w. Accept $q_s^{(r)}$ with prob-

ability

$$\frac{\exp\left[\sum_{t=1}^{T_s}(z_{st} * q_s^{(r)}) - T_s * \log(1+\exp(q_s^{(r)})\right] h_1(q_s^{(r)} \mid \tau^2, \rho, \mathbf{q}_{-s})}{\exp\left[\sum_{t=1}^{T_s}(z_{st} * q_s^{(r-1)}) - T_s * \log(1+\exp(q_s^{(r-1)})\right] h_1((q_s^{(r-1)} \mid \tau^2, \rho, \mathbf{q}_{-s})},$$

where $\exp\left[\sum_{t=1}^{T_s}(z_{st} * q_s^{(r)}) - T_s * \log(1+\exp(q_s^{(r)})\right]$ is the likelihood of the Bernoulli distribution given q_s, and $h_1(\cdot)$ is the univariate conditional Normal distribution given $\mathbf{q}_{-s}$, the vector of logit weights for all locations except location s.

3. Update τ^2. The full conditional distribution of $\tau^2 \sim$ Inverse-Gamma$(a + S/2, b + 1/2\mathbf{W}^T\Sigma^{-1}\mathbf{W})$, where $\mathbf{W} = (q_1, \ldots, q_S)^T$, and Σ is the spatial covariance matrix of $\mathbf{W}$.

4. Update ρ. Generate proposal $\rho^{(r)}$ from a log-normal distribution with mean $\rho^{(r-1)}$ and variance κ_ρ. Accept $\rho^{(r)}$ with probability

$$\frac{l_1(\mathbf{W} \mid \tau^2, \rho^{(r)})\, l_2(\rho^{(r)})\rho^{(r)}}{l_1(\mathbf{W} \mid \tau^2, \rho^{(r-1)})\, l_2(\rho^{(r-1)})\rho^{(r-1)}},$$

where l_1 is the multivariate Normal distribution for the Gaussian process, and l_2 is the Gamma prior distribution.

Updating q_s, and thereby w_s, individually when the number of monitoring locations is large can be computationally demanding. Hence, we also consider a *two-stage* approach. First, the ensemble weight w_s is estimated separately at each location by assuming a prior distribution $w_s \sim$ Beta$(1, 1)$. Bayesian kriging is then applied to the posterior medians of w_s across locations, assuming a similar Gaussian process model as above. Compared to the original *joint* estimation approach, the two-stage approach assumes q_s to be known when performing spatial interpolation.

Equations for evaluation statistics follow. R^2 refers to

$$R^2 = 1 - \frac{\sum_{s=1}^{S}\sum_{t=1}^{T_s}(y_{st} - \hat{y}_{st})^2}{\sum_{s=1}^{S}\sum_{t=1}^{T_s}(y_{st} - \bar{y}_{st})^2}, \tag{A.2}$$

while root mean squared error (RMSE) is defined as

$$RMSE = \sqrt{\frac{\sum_{s=1}^{S}\sum_{t=1}^{T_s}(y_{st} - \hat{y}_{st})^2}{\sum_{s=1}^{S} s \times T_s}} \tag{A.3}$$

A.3 Spatially Clustered Cross-Validation

For the spatially clustered cross-validation (CV) approach, we use hierarchical clustering to group the United States Environmental Protection Agency's Air Quality System monitoring locations based on proximity. Complete, agglomerative hierarchical clustering starts with each point belonging to its own "cluster," then finds the best match based on distance, and continues to do so until all points belong to a single cluster. With this approach, we do not have to specify the number of clusters ahead of time, allowing for efficient comparison of results with different numbers of clusters. The distance was determined by Euclidean distance between AQS monitors. To best illustrate real-life scenarios, we use 20 clusters in the spatially-clustered CV experiment.

A.4 Non-Bayesian Mixed Models

In order to show the advantages of using individual Bayesian hierarchical models (BHM) and the ensemble method with BHMs as its inputs, we provide a non-Bayesian mixed model comparison in Table S3. A single-input mixed model mirrors that of the BHM: $PM_{2.5}$ is the outcome, and AOD and covariates are used as a fixed effects with space and time as independent random effects ($PM_{2.5}$-AOD MM). The single-input CMAQ mixed model ($PM_{2.5}$-CMAQ MM) has the same construction without the covariates. The multi-input

mixed model is the non-Bayesian comparison to the ensemble model. The multi-input mixed model contains fixed effects for AOD, CMAQ, and covariates with independent, random space and time effects. Table 1 from the manuscript can be compared to Table S3 herein, and clear performance advantages from using the Bayesian methods are demonstrated through each performance metric, i.e. RMSE, coverage, Average SD, and R^2.

A.5 Figures

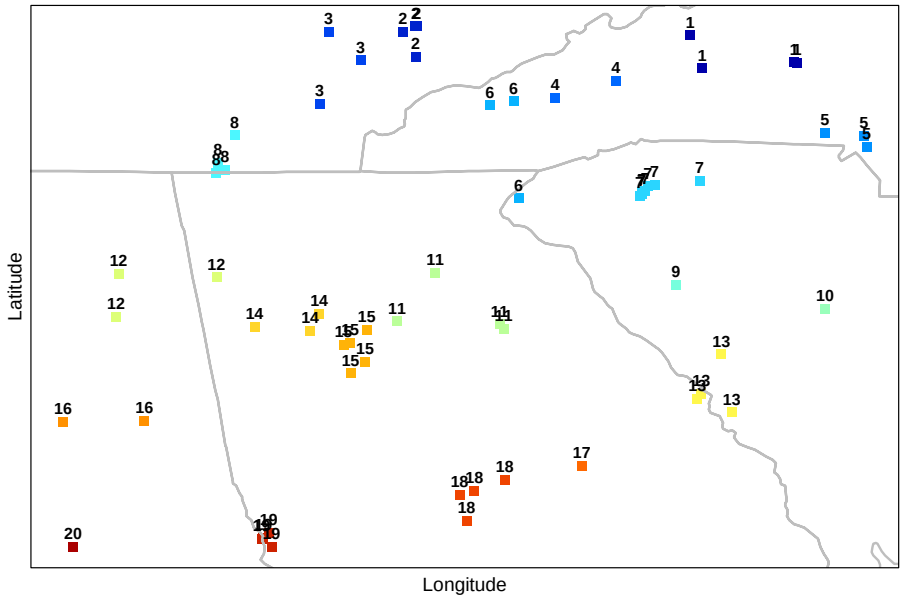

Figure A.1: United States Environmental Protection Agency Air Quality System monitoring sites grouped through hierarchical clustering. Each number and color corresponds to which cluster that monitoring site belongs.

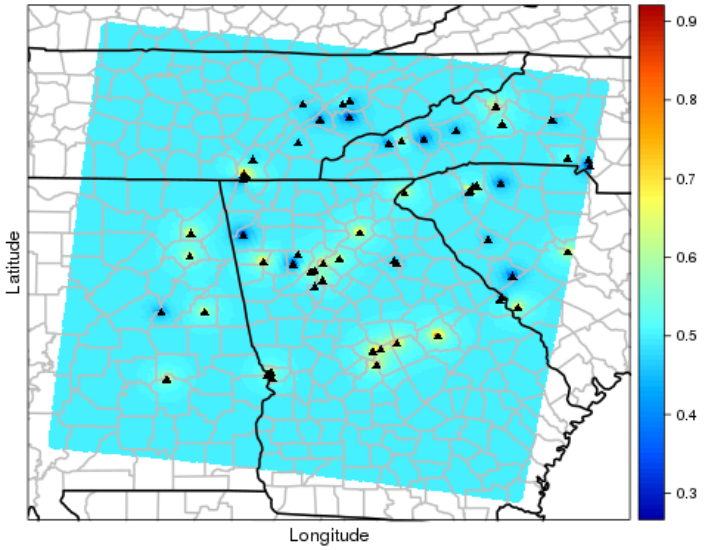

Figure A.2: Spatially interpolated ensemble weights for predictions from the $PM_{2.5}$-Community Multiscale Air Quality (CMAQ) Bayesian hierarchical model at 1 km $\times$ 1 km resolution. Black triangles indicate AQS monitoring locations.

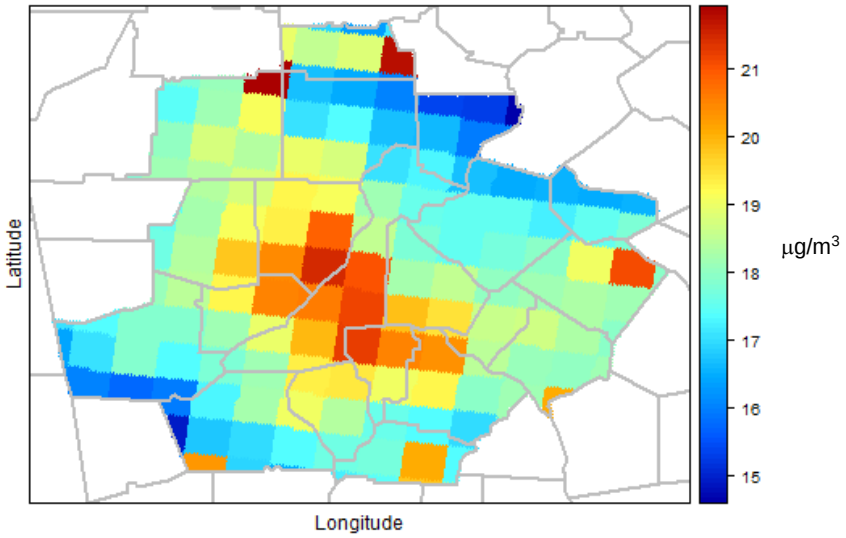

Figure A.3: Posterior means of $PM_{2.5}$ from a Bayesian hierarchical model with simulations from the Community Multiscale Air Quality (CMAQ) model on March 26, 2005.

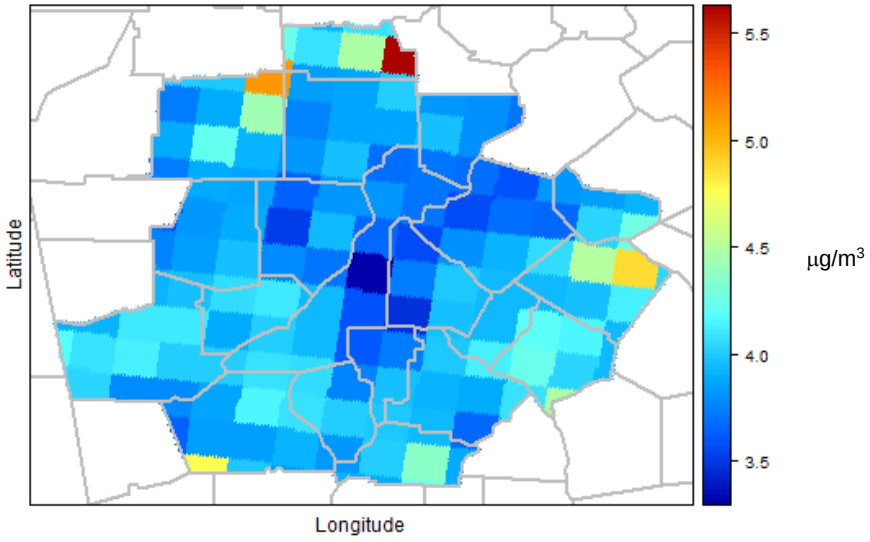

Figure A.4: Posterior standard errors of $PM_{2.5}$ from a Bayesian hierarchical model with simulations from the Community Multiscale Air Quality (CMAQ) model on March 26, 2005.

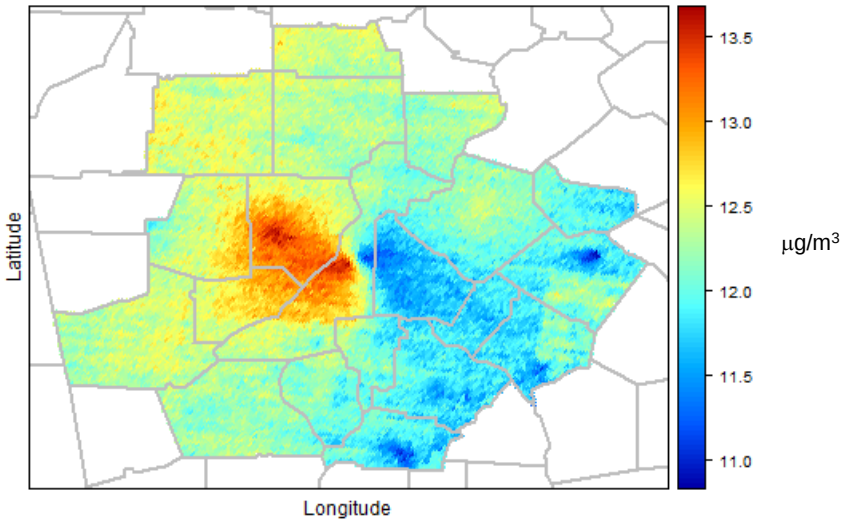

Figure A.5: Posterior means of $PM_{2.5}$ from a Bayesian hierarchical model with satellite-derived aerosol optical depth (AOD) on March 26, 2005.

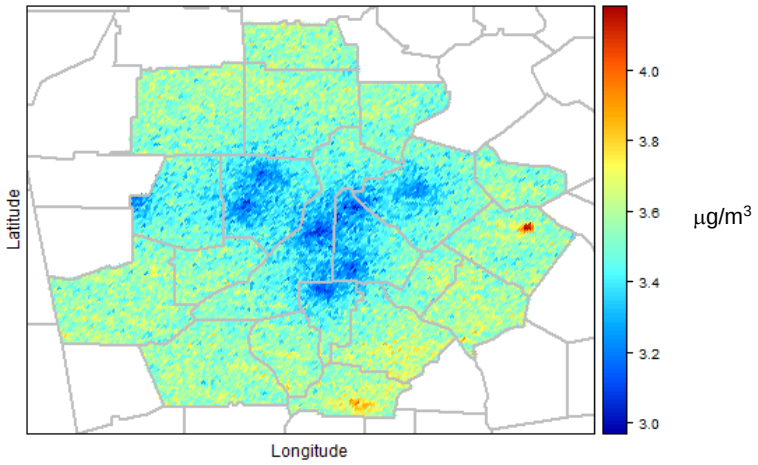

Figure A.6: Posterior standard errors of $PM_{2.5}$ from a Bayesian hierarchical model with satellite-derived aerosol optical depth (AOD) on March 26, 2005.

A.6 Tables

Table A.1: Prediction performance for daily $PM_{2.5}$ concentrations in 10-fold cross-validation (CV) comparing ensemble averaging with a Bayesian hierarchical model (BHM) using satellite-derived aerosol optical depth (AOD) or a BHM using a numerical model (CMAQ) simulation. Ensemble inputs were derived from first performing either 10-fold or leave-one-monitor-out (spatial) CV.

Method	Estimation	RMSE	Coverage of 95% PI	Average Posterior SD	R^2
$PM_{2.5}$-AOD BHM	–	3.40	94.07	3.30	0.78
$PM_{2.5}$-CMAQ BHM	–	3.14	95.05	3.28	0.81
Ensemble (10-fold CV input)	Joint	3.00	97.15	3.52	0.83
	two-stage	3.01	97.09	3.50	0.82
Ensemble (spatial CV input)	Joint	3.13	97.39	3.62	0.81
	two-stage	3.14	97.33	3.60	0.81

RMSE: root mean squared error (in $\mu g/m^3$); PI: prediction interval; SD: standard deviation (in $\mu g/m^3$); CV: cross-validation; $PM_{2.5}$: particulate matter less than 2.5 μm; AOD: aerosol optical depth; BHM: Bayesian hierarchical model; CMAQ: Community Multiscale Air Quality

Table A.2: Prediction performance for daily $PM_{2.5}$ concentrations in leave-one-monitor-out (spatial) cross-validation (CV) comparing ensemble averaging with a Bayesian hierarchical model (BHM) using satellite-derived aerosol optical depth (AOD) or a BHM using a numerical model (CMAQ) simulation. Ensemble inputs were derived from first performing either 10-fold or leave-one-monitor-out (spatial) CV.

Method	Estimation	RMSE	Coverage of 95% PI	Average Posterior SD	R^2
$PM_{2.5}$-AOD BHM	–	3.45	94.25	3.39	0.77
$PM_{2.5}$-CMAQ BHM	–	3.33	95.32	3.45	0.78
Ensemble (10-fold CV input)	Joint	2.99	96.81	3.52	0.83
	two-stage	3.01	96.68	3.53	0.82
Ensemble (spatial CV input)	Joint	3.14	97.23	3.65	0.81
	two-stage	3.14	97.23	3.66	0.81

RMSE: root mean squared error (in $\mu g/m^3$); PI: prediction interval; SD: standard deviation (in $\mu g/m^3$); CV: cross-validation; $PM_{2.5}$: particulate matter less than 2.5 μm; AOD: aerosol optical depth; BHM: Bayesian hierarchical model; CMAQ: Community Multiscale Air Quality

Appendix B

Appendix for Chapter 3

B.1 Random Forest Tuning

Rather than simply using the defaults from the randomForest or ranger packages in R, we explore different combinations of hyperparameters in the random forest framework to see which combination results in the best predictive statistics for our data. Instead of tuning all seven years of data, we look at 2002, 2005, and 2008 as selected representations of the seven years.

The hyperparameters we vary are: 1) the number of trees (ntree) to build the random forest. The default of most random forest algorithms is 500, but research shows that as few as 250 may suffice. (Probst and Boulesteix, 2017) We try 250, 500, and 1000. 2) the number of variables to randomly sample as candidate covariates at each split of the decision tree (mtry). The default for regression is p/3. We start at the default value and increase by equal size until we reach the maximum number of covariates; specifically, we try (7, 11, 15, 19) for the AOD only dataset and (6, 10, 14, 18) for the all dataset. 3) the minimum number of samples within the terminal nodes (nodesize). This controls the tree depth. The default for regression is 5. We try 3, 5, 7, 10.

Because we use a 10 fold cross-validation approach to evaluate the predictive ability

of each random forest, we report the summary statistics *on average* across the 10 folds. The best performing hyperparameter combination is subjectively defined as the combination having the lowest RMSE, on average, and the highest R^2, on average, while maintaining at least 95% coverage, on average.

We also compare different options to code the categorical covariates such as day of week and month. "Order" is the computationally efficient option and results do not considerably differ from partition, so all random forests considered here use the order option opposed to partition, which is computationally expensive with little improvement. This is because partition allows "all possible 2-partitions" consideration for splitting (Wright and Ziegler, 2017). Order for multiclass classification has the factor levels ordered by the first principal component of the weighted covariance matrix of the contingency table (Wright and Ziegler, 2017). Finally, we consider the option of manually coding the multicategory covariates as binary covariates.

The chosen hyperparameter combinations consider all possible partitions of factor variables and are 1000, 7, 3 (ntrees, mtry, nodesize, respectively) for the AOD only dataset and 1000, 14, 3 for the all dataset. Overall, within a certain tuning year, the hyperparameter combinations result in similar predictive summaries, and while tuning is recommended, it may not be necessary to achieve favorable results (Probst and Boulesteix, 2017; Friedman et al., 2001). As for our method, one can still expect this ensemble method to work regardless of tuning practices.

B.2 Random Forest Estimation Methods

B.2.1 Discussion of random forest standard errors

Although random forest methods have been used for many years (Breiman), a recent topic of exploration is the methodology used to characterize the uncertainty of the predictions from random forests. (Breiman, 2001) We use naive estimates of uncertainty, but other

options include more advanced Monte Carlo methods (Coulston et al. (2016)) or quantile regression methods (Meinshausen and Ridgeway (2006)). Neither of these methods are currently considered "gold standard" estimates for random forest prediction standard error.

Further, it is important to differentiate between methods that produce prediction intervals and confidence intervals. The ranger package has the infinitesimal jackknife method of Wager et al. as the default prediction standard error method. (Wager et al., 2014) However, the infinitesimal jackknife is expressly meant for confidence intervals and provides "estimates for standard errors of the expected values of prediction" (Hengl et al., 2018). There are other confidence interval methods as well, but they are not as readily implementable in R, and because random forest standard error was not the focus of this project, we did not implement these methods (Sexton and Laake, 2009; Mentch and Hooker, 2016).

We use a naive standard deviation estimate for the random forest estimates where we take the standard deviation of the 1000 tree estimates. Other uncertainty measures exist for the random forest variance estimates, with the infinitesimal jackknife method of Wager et al. being the default in the ranger package (Wager et al., 2014). However, for our application, the infinitesimal jackknife approach results in inadequate coverage results.

Bivand et al. offers an approach to assess how the prediction standard error performs by looking at z-scores for each location (Bivand et al., 2008). If the score is lower than 1, that indicates that the estimate overestimates the error. If the score is higher than 1, the estimate is underestimating the error. Especially since we planned to use the estimates and their related uncertainties in a health analysis, we would rather the estimates of uncertainty be conservative and, therefore, be under 1. We found that our naive approach provides conservative standard error estimates while the infinitesimal jackknife approach provides too small error estimates for prediction purposes.

B.2.2 Estimation details

Estimation and prediction are accomplished in seven stages, which we describe in the enumerated steps below.

1. Fit the "Base" random forest to obtain predictive means, $\mu_{st}^{(1)}$, and variances, $\sigma_{st}^{2,(1)}$, for each day and location where we have observed AOD values.

2. Fit the "Base+AOD" random forest to to obtain predictive means, $\mu_{st}^{(2)}$, and variances, $\sigma_{st}^{2,(2)}$, for each day and location regardless of AOD values present.

3. Create out-of-sample "Base" model predictions. Randomly leave 10% of the $PM_{2.5}$ observations out then obtain prediction means and prediction variances using the remaining 90% of the data. Repeat this ten times. Stack the predictions to create a dataset.

4. Create out-of-sample "Base+AOD" predictions. Randomly leave 10% of the $PM_{2.5}$ observations out then obtain prediction means and prediction variances using the remaining 90% of the data. Repeat this ten times. Stack the predictions to create a dataset.

5. Estimate spatially varying weights based on $PM_{2.5}$ measurements and out-of-sample prediction datasets from Steps 3 and 4 using Equation (B.1).

$$p(y_{st} \mid M_1, M_2) = w_s f_1(y_{st} \mid M_1) + (1 - w_s) f_2(y_{st} \mid M_2), \tag{B.1}$$

6. Interpolate the weights to 1 km $\times$ 1 km grid cells using kriging.

7. Combine the estimates from Steps 1 and 2 using weights from Step 5 in the same fashion as Equation (B.2) to obtain the *ensemble estimate*.

$$\hat{y}_{st} = w_s \mu_{st}^{(1)} + (1 - w_s) \mu_{st}^{(2)}, \tag{B.2}$$

Notice in Steps 3 and 4, to avoid overfitting while estimating ensemble weights, we fit the RF models repeatedly, but we leave-out and back-predict observations in a *cross-validation* experiment, similar to approaches employed in stack regression and SuperLearner techniques. (LeBlanc and Tibshirani, 1996; Polley and van der Laan, 2010)

B.3 Tables

Table B.1: $PM_{2.5}$ predictions summary statistics by year, AOD only subset.

Year	Base Model				Base+AOD Model			
	RMSE	Coverage of 95% PI	Avg SD	R^2	RMSE	Coverage of 95% PI	Avg SD	R^2
2002	3.52	0.98	4.57	0.76	3.78	0.98	4.64	0.73
2003	3.25	0.98	4.29	0.77	3.48	0.98	4.37	0.74
2004	3.15	0.98	4.19	0.79	3.39	0.98	4.30	0.76
2005	3.76	0.98	4.81	0.77	3.77	0.98	4.73	0.77
2006	3.32	0.98	4.28	0.78	3.54	0.98	4.44	0.75
2007	4.07	0.98	4.97	0.74	4.12	0.98	4.96	0.73
2008	3.18	0.98	3.86	0.72	3.37	0.98	4.06	0.68

Table B.2: Case-crossover results of overall analyses for 1) all days (lag 1) between 01/02/2002 and 12/31/2008 with assigned ensemble exposure averaged over each ZCTA (spatially-varying); 2) all days (lag 1) between 01/02/2002 and 12/31/2008 with assigned ensemble exposure averaged over the ZCTA containing the central monitor and assigned to all ZCTAs on that day (spatially-homogeneous)

Data	OR	95% CI	β	Standard Error	p-value for $H_0 : \beta = 0$
All days (spatially-varying)	1.014	0.998, 1.031	0.0014	0.0008	0.09
All days (spatially-homogeneous)	1.010	0.996, 1.024	0.0010	0.0007	0.16

Footnotes: β is the log odds ratio for that model. Spatially-homogeneous: for the same day, every ZCTA receives the same exposure assignment from the central monitor containing ZCTA. OR are in terms of every 10 $\mu g/m^3$.

Table B.3: Case-crossover results of stratified quartile analyses for 1) all days (lag 1) between 01/02/2002 and 12/31/2008 with assigned ensemble exposure averaged over each ZCTA (spatially-varying); 2) all days (lag 1) between 01/02/2002 and 12/31/2008 with assigned ensemble exposure averaged over the ZCTA containing the central monitor and assigned to all ZCTAs on that day (spatially-homogeneous)

Data	Quartile	OR	95% CI	β	Standard Error	p-value for $H_0 : \beta = 0$	p-value for $H_0 : \mu_1 = \mu_{other}$
Spatially-varying	Q1	1.033	0.995, 1.073	0.0033	0.0019	0.09	-
	Q2	1.019	0.983, 1.056	0.0019	0.0018	0.31	0.60
	Q3	0.996	0.962, 1.031	-0.0004	0.0018	0.82	0.16
	Q4	1.014	0.988, 1.041	0.0014	0.0013	0.29	0.43
Spatially-homogeneous	Q1	1.025	0.994, 1.057	0.0025	0.0016	0.12	-
	Q2	1.018	0.987, 1.049	0.0018	0.0016	0.26	0.75
	Q3	0.992	0.963, 1.023	-0.0008	0.0015	0.62	0.14
	Q4	1.010	0.986, 1.034	0.0010	0.0012	0.42	0.45

Footnotes: β is the log odds ratio for that stratum's model. μ_1 is the log odds ratio for Q1. μ_{other} is the log odds ratio for Q2, Q3, or Q4. Quartile values of percent below poverty were defined as: Q1 = <7.6%; Q2 = $\geq$ 7.6% to <11.4%; Q3 = $\geq$11.4% to <16.2%; Q4 = >16.2%. OR are in terms of every 10 $\mu g/m^3$. Spatially-homogeneous: for the same day, every ZCTA receives the same exposure assignment from a central monitor containing ZCTA.

Appendix C

Appendix for Chapter 4

C.1 Truncated Normal Mean and Variance

Truncated univariate normal details from Step 1 of estimation algorithm has truncation points from R^2 region moments from

$\boldsymbol{U}^*_{s,j}|\boldsymbol{U}_{s,-j}, \boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1}, Z_s \sim N(m_{s,j}, \tau^2_{s,j})$, where $m_{s,j}$ and $\tau^2_{s,j}$ are formed from partitioning the $\boldsymbol{\Sigma}^{-1}$ matrix.

Permute Σ^{-1} so the j^{th} row and column are last.

e.g., for $j = 1$,

$$\Sigma^{-1} = \begin{bmatrix} \sigma_{22} & \sigma_{21} \\ \sigma_{12} & \sigma_{11} \end{bmatrix}^{-1} = \begin{bmatrix} \sigma_{22}^{-1} + FE^{-1}F & -FE^{-1} \\ -EF' & E^{-1} \end{bmatrix},$$

where $E = \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21} = \tau^2_{s,1}$

$F = \sigma_{22}^{-1}\sigma_{21}$

$m_{s,1} = \boldsymbol{X}_s[1,]\beta + F'(U_{s,2} - \boldsymbol{X}_s[2,]\beta)$

C.2 Simple Rejection Sampling

Let $e_s = \tilde{\boldsymbol{U}}^*_s - \alpha^* \boldsymbol{X}_s \boldsymbol{\beta}$, where $\boldsymbol{\beta} = \dfrac{\tilde{\boldsymbol{\beta}}^*}{\alpha^*}$. Also, $\alpha = \tilde{\boldsymbol{\Sigma}}^*_{11}$ and α^* is the proposed α for that iteration.

In order for the transformation from $(e_s, \boldsymbol{\beta}, \alpha, \tilde{\boldsymbol{\Sigma}}^*)$ to $(\boldsymbol{U}, \boldsymbol{\beta}, \alpha, \boldsymbol{\Sigma})$ to be correct, we must set

$$\boldsymbol{U}_s = \frac{e_s + \alpha \boldsymbol{X}_s \boldsymbol{\beta}}{\alpha} = \frac{\tilde{\boldsymbol{U}}_s^* - \alpha^* \boldsymbol{X}_s \dfrac{\tilde{\boldsymbol{\beta}}^*}{\alpha^*} + \alpha \boldsymbol{X}_s \dfrac{\tilde{\boldsymbol{\beta}}^*}{\alpha^*}}{\alpha}$$

$$= \tilde{\boldsymbol{U}}_s^*/\alpha - \boldsymbol{X}_s \tilde{\boldsymbol{\beta}}^*/\alpha + \boldsymbol{X}_s \tilde{\boldsymbol{\beta}}^*/\alpha^*$$

(The above helps the stationarity of $\boldsymbol{U}$.)

Therefore, set $\boldsymbol{U}_s = \dfrac{e_s + \alpha \boldsymbol{X}_s \boldsymbol{\beta}}{\alpha}$

Bibliography

Adam, M., Schikowski, T., Carsin, A. E., Cai, Y., Jacquemin, B., Sanchez, M., Vierkötter, A., Marcon, A., Keidel, D., Sugiri, D. and Al Kanani, Z. (2015), 'Adult lung function and long-term air pollution exposure. ESCAPE: A multicentre cohort study and meta-analysis', *European Respiratory Journal* **45**(1), 38–50.

Adler, N. E. and Newman, K. (2002), 'Socioeconomic disparities in health: Pathways and policies', *Health Affairs* **21**(2), 60–76.

Alexeeff, S. E., Carroll, R. J. and Coull, B. (2016), 'Spatial measurement error and correction by spatial SIMEX in linear regression models when using predicted air pollution exposures', *Biostatistics* **17**(2), 377–389.

Alexeeff, S. E., Schwartz, J., Kloog, I., Chudnovsky, A., Koutrakis, P. and Coull, B. A. (2015), 'Consequences of kriging and land use regression for $PM_{2.5}$ predictions in epidemiologic analyses: Insights into spatial variability using high-resolution satellite data', *Journal of Exposure Science and Environmental Epidemiology* **25**(2), 138–144.

Alhanti, B. A., Chang, H. H., Winquist, A., Mulholland, J. A., Darrow, L. A. and Sarnat, S. E. (2016), 'Ambient air pollution and emergency department visits for asthma: A multi-city assessment of effect modification by age', *Journal of Exposure Science and Environmental Epidemiology* **26**(2), 180–188.

Baxter, L. K., Burke, J., Lunden, M., Turpin, B. J., Rich, D. Q., Thevenet-Morrison, K., Hodas, N. and Özkaynak, H. (2013), 'Influence of human activity patterns, particle com-

position, and residential air exchange rates on modeled distributions of $PM_{2.5}$ exposure compared with central-site monitoring data', *Journal of Exposure Science and Environmental Epidemiology* **23**(3), 241–247.

Bell, M. L. and Ebisu, K. (2012), 'Environmental inequality in exposures to airborne particulate matter components in the United States', *Environmental Health Perspectives* **120**(12), 1699–1704.

Berrett, C. and Calder, C. A. (2012), 'Data augmentation strategies for the Bayesian spatial probit regression model', *Computational Statistics & Data Analysis* **56**(3), 478–490.

Berrocal, V. J., Gelfand, A. E. and Holland, D. M. (2010), 'A spatio-temporal downscaler for output from numerical models', *Journal of Agricultural, Biological, and Environmental Statistics* **15**(2), 176–197.

Berrocal, V. J., Guan, Y., Muyskens, A., Wang, H., Reich, B. J., Mulholland, J. A. and Chang, H. H. (2019), 'A comparison of statistical and machine learning methods for creating national daily maps of ambient $PM_{2.5}$ concentration', *Atmospheric Environment* p. 117130.

Berrocal, V. J., Raftery, A. E. and Gneiting, T. (2007), 'Combining spatial statistical and ensemble information in probabilistic weather forecasts', *Monthly Weather Review* **135**(4), 1386–1402.

Bhat, K. S., Haran, M., Terando, A. and Keller, K. (2011), 'Climate projections using Bayesian model averaging and space–time dependence', *Journal of Agricultural, Biological, and Environmental Statistics* **16**(4), 606–628.

Biau, G. and Scornet, E. (2016), 'A random forest guided tour', *TEST* **25**(2), 197–227.

Bivand, R. S., Pebesma, E. J., Gomez-Rubio, V. and Pebesma, E. J. (2008), *Applied Spatial Data analysis with R*, Vol. 747248717, Springer.

Boutin-Forzano, S., Adel, N., Gratecos, L., Jullian, H., Garnier, J., Ramadour, M., Lanteaume, A., Hamon, M., Lafay, V. and Charpin, D. (2004), 'Visits to the emergency room for asthma attacks and short-term variations in air pollution', *Respiration* **71**(2), 134–137.

Bowatte, G., Lodge, C. J., Knibbs, L. D., Erbas, B., Perret, J. L., Jalaludin, B., Morgan, G. G., Bui, D. S., Giles, G. G., Hamilton, G. S. and Wood-Baker, R. (2018), 'Traffic related air pollution and development and persistence of asthma and low lung function', *Environment International* **113**, 170–176.

Brauer, M., Amann, M., Burnett, R. T., Cohen, A., Dentener, F., Ezzati, M., Henderson, S. B., Krzyzanowski, M., Martin, R. V., Van Dingenen, R. et al. (2012), 'Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution', *Environmental Science & Technology* **46**(2), 652–660.

Breiman, L. (2001), 'Random forests', *Machine Learning* **45**(1), 5–32.

Brokamp, C., Jandarov, R., Hossain, M. and Ryan, P. (2018), 'Predicting daily urban fine particulate matter concentrations using a random forest model', *Environmental Science & Technology* **52**(7), 4173–4179.

Brook, R. D., Brook, J. R., Urch, B., Vincent, R., Rajagopalan, S. and Silverman, F. (2002), 'Inhalation of fine particulate air pollution and ozone causes acute arterial vasoconstriction in healthy adults', *Circulation* **105**(13), 1534–1536.

Brook, R. D., Newby, D. E. and Rajagopalan, S. (2017), 'The global threat of outdoor ambient air pollution to cardiovascular health: Time for intervention', *JAMA Cardiology* **2**(4), 353–354.

Brunekreef, B. and Holgate, S. T. (2002), 'Air pollution and health', *The Lancet* **360**(9341), 1233–1242.

Burgette, L. F. and Nordheim, E. V. (2012), 'The trace restriction: An alternative identification strategy for the Bayesian multinomial probit model', *Journal of Business & Economic Statistics* **30**(3), 404–410.

Byun, D. and Schere, K. L. (2006), 'Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system', *Applied Mechanics Reviews* **59**(2), 51–77.

Carracedo-Martínez, E., Taracido, M., Tobias, A., Saez, M. and Figueiras, A. (2010), 'Case-crossover analysis of air pollution health effects: A systematic review of methodology and application', *Environmental Health Perspectives* **118**(8), 1173–1182.

Carrothers, T. J. and Evans, J. S. (2000), 'Assessing the impact of differential measurement error on estimates of fine particle mortality', *Journal of the Air & Waste Management Association* **50**(1), 65–74.

Chang, H. H., Hu, X. and Liu, Y. (2014), 'Calibrating MODIS aerosol optical depth for predicting daily $PM_{2.5}$ concentrations via statistical downscaling', *Journal of Exposure Science and Environmental Epidemiology* **24**(4), 398–404.

Chang, H. H., Reich, B. J. and Miranda, M. L. (2011), 'Time-to-event analysis of fine particle air pollution and preterm birth: Results from North Carolina, 2001–2005', *American Journal of Epidemiology* **175**(2), 91–98.

Chipperfield, M. (1999), 'Multiannual simulations with a three-dimensional chemical transport model', *Journal of Geophysical Research: Atmospheres* **104**(D1), 1781–1805.

Clark, L. P., Millet, D. B. and Marshall, J. D. (2014*a*), 'National patterns in environmental injustice and inequality: outdoor NO_2 air pollution in the United States', *PloS one* **9**(4), 1–8.

Clark, L. P., Millet, D. B. and Marshall, J. D. (2014*b*), 'National patterns in environmental injustice and inequality: Outdoor NO_2 air pollution in the united states', *PloS One* **9**(4), e94431.

Clyde, M. (2000), 'Model uncertainty and health effect studies for particulate matter', *Environmetrics* **11**(6), 745–763.

Coulston, J. W., Blinn, C. E., Thomas, V. A. and Wynne, R. H. (2016), 'Approximating prediction uncertainty for random forest regression models', *Photogrammetric Engineering & Remote Sensing* **82**(3), 189–197.

Crooks, J. and Isakov, V. (2013), 'A wavelet-based approach to blending observations with deterministic computer models to resolve the intraurban air pollution field', *Journal of the Air & Waste Management Association* **63**(12), 1369–1385.

de Hoogh, K., Héritier, H., Stafoggia, M., Künzli, N. and Kloog, I. (2018), 'Modelling daily $PM_{2.5}$ concentrations at high spatio-temporal resolution across Switzerland', *Environmental Pollution* **233**, 1147–1154.

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y. and Schwartz, J. (2016), 'Assessing $PM_{2.5}$ exposures with high spatiotemporal resolution across the continental United States', *Environmental Science & Technology* **50**(9), 4712–4721.

Dionisio, K. L., Baxter, L. K. and Chang, H. H. (2014), 'An empirical assessment of exposure measurement error and effect attenuation in bipollutant epidemiologic models', *Environmental Health Perspectives* **122**(11), 1216–1224.

Evans, K. A., Halterman, J. S., Hopke, P. K., Fagnano, M. and Rich, D. Q. (2014), 'Increased ultrafine particles and carbon monoxide concentrations are associated with asthma exacerbation among urban children', *Environmental Research* **129**, 11–19.

Forastiere, F., Stafoggia, M., Tasco, C., Picciotto, S., Agabiti, N., Cesaroni, G. and Perucci, C. A. (2007), 'Socioeconomic status, particulate air pollution, and daily mortality: Differential exposure or differential susceptibility', *American Journal of Industrial Medicine* **50**(3), 208–216.

Friberg, M. D., Kahn, R. A., Holmes, H. A., Chang, H. H., Sarnat, S. E., Tolbert, P. E., Russell, A. G. and Mulholland, J. A. (2017), 'Daily ambient air pollution metrics for five cities: Evaluation of data-fusion-based estimates and uncertainties', *Atmospheric Environment* **158**, 36–50.

Friberg, M. D., Kahn, R. A., Limbacher, J. A., Appel, K. W. and Mulholland, J. A. (2018), 'Constraining chemical transport $PM_{2.5}$ modeling outputs using surface monitor measurements and satellite retrievals: Application over the San Joaquin Valley', *Atmospheric Chemistry & Physics* **18**(17).

Friedman, J., Hastie, T. and Tibshirani, R. (2001), *The Elements of Statistical Learning*, Vol. 1, Springer Series in Statistics New York.

Fuentes, M. and Raftery, A. E. (2005), 'Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models', *Biometrics* **61**(1), 36–45.

Gass, K., Klein, M., Sarnat, S. E., Winquist, A., Darrow, L. A., Flanders, W. D., Chang, H. H., Mulholland, J. A., Tolbert, P. E. and Strickland, M. J. (2015), 'Associations between ambient air pollutant mixtures and pediatric asthma emergency department visits in three cities: A classification and regression tree approach', *Environmental Health* **14**(1), 1–14.

Gelfand, A. E., Diggle, P., Guttorp, P. and Fuentes, M., eds (2010), *Handbook of Spatial Statistics*, CRC Press.

Geng, G., Murray, N. L., Tong, D., Fu, J. S., Hu, X., Lee, P., Meng, X., Chang, H. H. and

Liu, Y. (2018), 'Satellite-Based Daily $PM_{2.5}$ Estimates During Fire Seasons in Colorado', *Journal of Geophysical Research: Atmospheres* **123**(15), 8159–8171.

Glad, J. A., Brink, L. L., Talbott, E. O., Lee, P. C., Xu, X., Saul, M. and Rager, J. (2012), 'The relationship of ambient ozone and $PM_{2.5}$ levels and asthma emergency department visits: possible influence of gender and ethnicity', *Archives of environmental & occupational health* **67**(2), 103–108.

Gray, S. C., Edwards, S. E., Schultz, B. D. and Miranda, M. L. (2014), 'Assessing the impact of race, social factors and air pollution on birth outcomes: A population-based study', *Environmental Health* **13**(1), 1–8.

Grineski, S. E., Herrera, J. M., Bulathsinhala, P. and Staniswalis, J. G. (2015), 'Is there a Hispanic Health Paradox in sensitivity to air pollution? Hospital admissions for asthma, chronic obstructive pulmonary disease and congestive heart failure associated with NO_2 and $PM_{2.5}$ in el paso, tx, 2005–2010', *Atmospheric Environment* **119**, 314–321.

Gryparis, A., Paciorek, C. J., Zeka, A., Schwartz, J. and Coull, B. A. (2009), 'Measurement error caused by spatial misalignment in environmental epidemiology', *Biostatistics* **10**(2), 258–274.

Guarnieri, M. and Balmes, J. R. (2014), 'Outdoor air pollution and asthma', *The Lancet* **383**(9928), 1581–1592.

Hart, J. E., Liao, X., Hong, B., Puett, R. C., Yanosky, J. D., Suh, H., Kioumourtzoglou, M.-A., Spiegelman, D. and Laden, F. (2015*a*), 'The association of long-term exposure to $PM_{2.5}$ on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction', *Environmental Health* **14**(1), 1–9.

Hart, J. E., Liao, X., Hong, B., Puett, R. C., Yanosky, J. D., Suh, H., Kioumourtzoglou, M.-A., Spiegelman, D. and Laden, F. (2015*b*), 'The association of long-term exposure to $PM_{2.5}$

on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction', *Environmental Health* **14**(1), 38.

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. and Gräler, B. (2018), 'Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables', *PeerJ* **6**, e5518.

Hepple, L. W. (1995), 'Bayesian techniques in spatial and network econometrics: 2. Computational methods and algorithms', *Environment and Planning A* **27**(4), 615–644.

Hoek, G., Krishnan, R. M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B. and Kaufman, J. D. (2013), 'Long-term air pollution exposure and cardio-respiratory mortality: A review', *Environmental Health* **12**(1), 1–15.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999), 'Bayesian model averaging: A tutorial', *Statistical Science* **14**(4), 382–401.

Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., Strickland, M. J. and Liu, Y. (2017), 'Estimating $PM_{2.5}$ concentrations in the conterminous United States using the random forest approach', *Environmental Science & Technology* **51**(12), 6936–6944.

Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes, M. G., Estes, S. M., Quattrochi, D. A., Sarnat, J. A. and Liu, Y. (2013), 'Estimating ground-level $PM_{2.5}$ concentrations in the southeastern US using geographically weighted regression', *Environmental Research* **121**, 1–10.

Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., Gu, D. and Liu, Y. (2018), 'Predicting monthly high-resolution $PM_{2.5}$ concentrations with random forest model in the North China Plain', *Environmental Pollution* **242**, 675–683.

Hubbell, B. J., Crume, R. V., Evarts, D. M. and Cohen, J. M. (2009), 'Policy monitor: Regu-

lation and progress under the 1990 Clean Air Act amendments', *Review of Environmental Economics and Policy* **4**(1), 122–138.

Imai, K. and Van Dyk, D. A. (2005), 'A Bayesian analysis of the multinomial probit model using marginal data augmentation', *Journal of Econometrics* **124**(2), 311–334.

Janes, H., Sheppard, L. and Lumley, T. (2005), 'Case-crossover analyses of air pollution exposure data: Referent selection strategies and their implications for bias', *Epidemiology* **16**(6), 717–726.

Jerrett, M., Shankardass, K., Berhane, K., Gauderman, W. J., Künzli, N., Avol, E., Gilliland, F., Lurmann, F., Molitor, J. N., Molitor, J. T. and Thomas, D. C. (2008), 'Traffic-related air pollution and asthma onset in children: A prospective cohort study with individual exposure measurement', *Environmental Health Perspectives* **116**(10), 1433.

Johndrow, J., Dunson, D. and Lum, K. (2013), Diagonal orthant multinomial probit models, *in* 'Artificial Intelligence and Statistics', pp. 29–38.

Johnson, S. C. (1967), 'Hierarchical clustering schemes', *Psychometrika* **32**(3), 241–254.

Keane, M. P. (1992), 'A note on identification in the multinomial probit model', *Journal of Business & Economic Statistics* **10**(2), 193–200.

Kianian, B., Liu, Y. and Chang, H. H. (2021), 'Imputing satellite-derived aerosol optical depth using a multi-resolution spatial model and random forest for $PM_{2.5}$ prediction', *Remote Sensing* **13**(1), 126.

Kloog, I., Sorek-Hamer, M., Lyapustin, A., Coull, B., Wang, Y., Just, A. C., Schwartz, J. and Broday, D. M. (2015), 'Estimating daily $PM_{2.5}$ and PM_{10} across the complex geo-climate region of israel using MAIAC satellite-based AOD data', *Atmospheric Environment* **122**, 409–416.

Kuo, C.-L., Duan, Y. and Grady, J. (2018), 'Unconditional or conditional logistic regression model for age-matched case–control data?', *Frontiers in Public Health* **6**, 57.

LeBlanc, M. and Tibshirani, R. (1996), 'Combining estimates in regression and classification', *Journal of the American Statistical Association* **91**(436), 1641–1650.

Levy, R. C., Remer, L. A. and Dubovik, O. (2007), 'Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land', *Journal of Geophysical Research: Atmospheres* **112**(D13).

Lim, C. Y., Stein, M., Ching, J. and Tang, R. (2010), 'Statistical properties of differences between low and high resolution CMAQ runs with matched initial and boundary conditions', *Environmental Modelling & Software* **25**(1), 158–169.

Liu, T., Li, T. T., Zhang, Y. H., Xu, Y. J., Lao, X. Q., Rutherford, S., Chu, C., Luo, Y., Zhu, Q., Xu, X. J. and Xie, H. Y. (2013), 'The short-term effect of ambient ozone on mortality is modified by temperature in Guangzhou, China', *Atmospheric Environment* **76**, 59–67.

Liu, Y., Paciorek, C. J. and Koutrakis, P. (2009), 'Estimating regional spatial and temporal variability of $PM_{2.5}$ concentrations using satellite data, meteorology, and land use information', *Environmental Health Perspectives* **117**(6), 886–892.

Liu, Y., Sarnat, J. A., Kilaru, V., Jacob, D. J. and Koutrakis, P. (2005), 'Estimating ground-level $PM_{2.5}$ in the eastern United States using satellite remote sensing', *Environmental Science & Technology* **39**(9), 3269–3278.

Loría-Salazar, S. M., Panorska, A., Arnott, W. P., Barnard, J. C., Boehmler, J. M. and Holmes, H. A. (2017), 'Toward understanding atmospheric physics impacting the relationship between columnar aerosol optical depth and near-surface $PM_{2.5}$ mass concentrations in Nevada and California, USA, during 2013', *Atmospheric Environment* **171**, 289–300.

Lyapustin, A., Martonchik, J., Wang, Y., Laszlo, I. and Korkin, S. (2011), 'Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables', *Journal of Geophysical Research: Atmospheres* **116**, 1–9.

Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., Levy, R. and Reid, J. (2011), 'Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm', *Journal of Geophysical Research: Atmospheres* **116**, 1–15.

Maji, K. J., Dikshit, A. K. and Deshpande, A. (2017), 'Disability-adjusted life years and economic cost assessment of the health effects related to $PM_{2.5}$ and PM_{10} pollution in Mumbai and Delhi, in India from 1991 to 2015', *Environmental Science and Pollution Research* **24**(5), 4709–4730.

Marmur, A., Park, S.-K., Mulholland, J. A., Tolbert, P. E. and Russell, A. G. (2006), 'Source apportionment of $PM_{2.5}$ in the southeastern United States using receptor and emissions-based models: Conceptual differences and implications for time-series health studies', *Atmospheric Environment* **40**(14), 2533–2551.

Maté, T., Guaita, R., Pichiule, M., Linares, C. and Díaz, J. (2010), 'Short-term effect of fine particulate matter ($PM_{2.5}$) on daily mortality due to diseases of the circulatory system in Madrid (Spain)', *Science of the Total Environment* **408**(23), 5750–5757.

Mebust, M. R., Eder, B. K., Binkowski, F. S. and Roselle, S. J. (2003), 'Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component 2. Model evaluation', *Journal of Geophysical Research: Atmospheres* **108**, 1–18.

Meinshausen, N. and Ridgeway, G. (2006), 'Quantile regression forests', *Journal of Machine Learning Research* **7**(6), 983–999.

Meng, X.-L. and Van Dyk, D. A. (1999), 'Seeking efficient data augmentation schemes via conditional and marginal augmentation', *Biometrika* **86**(2), 301–320.

Mentch, L. and Hooker, G. (2016), 'Quantifying uncertainty in random forests via confidence intervals and hypothesis tests', *The Journal of Machine Learning Research* **17**(1), 841–881.

Metzger, K. B., Tolbert, P. E., Klein, M., Peel, J. L., Flanders, W. D., Todd, K., Mulholland, J. A., Ryan, P. B. and Frumkin, H. (2004), 'Ambient air pollution and cardiovascular emergency department visits', *Epidemiology* **15**(1), 46–56.

Mikati, I., Benson, A. F., Luben, T. J., Sacks, J. D. and Richmond-Bryant, J. (2018), 'Disparities in distribution of particulate matter emission sources by race and poverty status', *American Journal of Public Health* **108**(4), 480–485.

Miller, M. D. and Marty, M. A. (2010), 'Impact of environmental chemicals on lung development', *Environmental Health Perspectives* **118**(8), 1155.

Munoz-Pizza, D. M., Villada-Canela, M., Reyna, M., Texcalac-Sangrador, J. L. and Osornio-Vargas, Á. R. (2020), 'Air pollution and children's respiratory health: A scoping review of socioeconomic status as an effect modifier', *International Journal of Public Health* **65**, 649–660.

Murray, N. L., Holmes, H. A., Liu, Y. and Chang, H. H. (2019), 'A Bayesian ensemble approach to combine $PM_{2.5}$ estimates from statistical models using satellite imagery and numerical model simulation', *Environmental Research* **178**, 108601.

National Aeronautics and Space Administration (2018), 'Moderate Resolution Imaging Spectroradiometer Data', `https://modis.gsfc.nasa.gov/data/`. [Online; accessed June 25, 2018].

Neidell, M. J. (2004), 'Air pollution, health, and socio-economic status: The effect of outdoor air quality on childhood asthma', *Journal of Health Economics* **23**(6), 1209–1236.

O'Brien, G. M., Stein, M. D., Zierler, S., Shapiro, M., O'Sullivan, P. and Woolard, R.

(1997), 'Use of the ED as a regular source of care: Associated factors beyond lack of health insurance', *Annals of Emergency Medicine* **30**(3), 286–291.

O'Lenick, C. R., Winquist, A., Mulholland, J. A., Friberg, M. D., Chang, H. H., Kramer, M. R., Darrow, L. A. and Sarnat, S. E. (2017), 'Assessment of neighbourhood-level socioeconomic status as a modifier of air pollution–asthma associations among children in Atlanta', *Journal of Epidemiology and Community Health* **71**(2), 129–136.

Peel, J. L., Tolbert, P. E., Klein, M., Metzger, K. B., Flanders, W. D., Todd, K., Mulholland, J. A., Ryan, P. B. and Frumkin, H. (2005), 'Ambient air pollution and respiratory emergency department visits', *Epidemiology* **16**(2), 164–174.

Polley, E. C. and van der Laan, M. J. (2010), 'Super Learner in Prediction', *U.C. Berkeley Division of Biostatistics Working Paper* (266), 1–19.

Probst, P. and Boulesteix, A.-L. (2017), 'To tune or not to tune the number of trees in random forest', *The Journal of Machine Learning Research* **18**(1), 6673–6690.

Pui, D. Y., Chen, S.-C. and Zuo, Z. (2014), '$PM_{2.5}$ in China: Measurements, sources, visibility and health effects, and mitigation', *Particuology* **13**, 1–26.

Quinn, K. M., Martin, A. D. and Whitford, A. B. (1999), 'Voter choice in multi-party democracies: A test of competing theories and models', *American Journal of Political Science* pp. 1231–1247.

R Core Team (2018), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M. (2005), 'Using Bayesian

model averaging to calibrate forecast ensembles', *Monthly Weather Review* **133**(5), 1155–1174.

Reid, C. E., Jerrett, M., Petersen, M. L., Pfister, G. G., Morefield, P. E., Tager, I. B., Raffuse, S. M. and Balmes, J. R. (2015), 'Spatiotemporal prediction of fine particulate matter during the 2008 Northern California wildfires using machine learning', *Environmental Science & Technology* **49**(6), 3887–3896.

Sarnat, S. E., Klein, M., Sarnat, J. A., Flanders, W. D., Waller, L. A., Mulholland, J. A., Russell, A. G. and Tolbert, P. E. (2010), 'An examination of exposure measurement error from air pollutant spatial variability in time-series studies', *Journal of Exposure Science and Environmental Epidemiology* **20**(2), 135–146.

Schliep, E. M. and Hoeting, J. A. (2015), 'Data augmentation and parameter expansion for independent or spatially correlated ordinal data', *Computational Statistics and Data Analysis* **90**, 1–14.

Sexton, J. and Laake, P. (2009), 'Standard errors for bagged and random forest estimators', *Computational Statistics & Data Analysis* **53**(3), 801–811.

Strickland, M. J., Darrow, L. A., Klein, M., Flanders, W. D., Sarnat, J. A., Waller, L. A., Sarnat, S. E., Mulholland, J. A. and Tolbert, P. E. (2010), 'Short-term associations between ambient air pollutants and pediatric asthma emergency department visits', *American Journal of Respiratory and Critical Care Medicine* **182**(3), 307–316.

Szpiro, A. A., Sheppard, L. and Lumley, T. (2011), 'Efficient measurement error correction with spatially misaligned data', *Biostatistics* **12**(4), 610–623.

Tanner, M. A. and Wong, W. H. (1987), 'The calculation of posterior distributions by data augmentation', *Journal of the American statistical Association* **82**(398), 528–540.

United States Environmental Protection Agency (2009), 'Integrated Science Assessment (ISA) for Particulate Matter (Final Report).'.

Van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., Lyapustin, A., Sayer, A. M. and Winker, D. M. (2016), 'Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors', *Environmental Science & Technology* **50**(7), 3762–3772.

van Smeden, M., Lash, T. L. and Groenwold, R. H. (2020), 'Reflection on modern methods: Five myths about measurement error in epidemiological research', *International Journal of Epidemiology* **49**(1), 338–347.

VanderWeele, T. J. and Hernán, M. A. (2012), 'Results on differential and dependent measurement error of the exposure and the outcome using signed directed acyclic graphs', *American Journal of Epidemiology* **175**(12), 1303–1310.

Wager, S., Hastie, T. and Efron, B. (2014), 'Confidence intervals for random forests: The jackknife and the infinitesimal jackknife', *The Journal of Machine Learning Research* **15**(1), 1625–1651.

Wang, Y., Shi, L., Lee, M., Liu, P., Di, Q., Zanobetti, A. and Schwartz, J. D. (2017), 'Long-term exposure to $PM_{2.5}$ and mortality among older adults in the southeastern US', *Epidemiology* **28**(2), 207.

Wardlaw, A. (1993), 'The role of air pollution in asthma', *Clinical & Experimental Allergy* **23**(2), 81–96.

Warren, J. L., Stingone, J. A., Herring, A. H., Luben, T. J., Fuentes, M., Aylsworth, A. S., Langlois, P. H., Botto, L. D., Correa, A., Olshan, A. F. and National Birth Defects Prevention Study (2016), 'Bayesian multinomial probit modeling of daily windows of susceptibility for maternal $PM_{2.5}$ exposure and congenital heart defects', *Statistics in Medicine* **35**(16), 2786–2801.

Wikle, C. K., Berliner, L. M. and Cressie, N. (1998), 'Hierarchical Bayesian space-time models', *Environmental and Ecological Statistics* **5**(2), 117–154.

Wright, M. N. and Ziegler, A. (2017), 'ranger: A fast implementation of random forests for high dimensional data in C++ and R', *Journal of Statistical Software* **77**(1), 1–17.

Xiao, Q., Wang, Y., Chang, H. H., Meng, X., Geng, G., Lyapustin, A. and Liu, Y. (2017), 'Full-coverage high-resolution daily $PM_{2.5}$ estimation using MAIAC AOD in the Yangtze River Delta of China', *Remote Sensing of Environment* **199**, 437–446.

Xing, Y.-F., Xu, Y.-H., Shi, M.-H. and Lian, Y.-X. (2016), 'The impact of $PM_{2.5}$ on the human respiratory system', *Journal of Thoracic Disease* **8**(1), E69.

Young, M. T., Bechle, M. J., Sampson, P. D., Szpiro, A. A., Marshall, J. D., Sheppard, L. and Kaufman, J. D. (2016), 'Satellite-based NO_2 and model validation in a national prediction model based on universal kriging and land-use regression', *Environmental Science & Technology* **50**(7), 3686–3694.

Zeger, S. L., Dominici, F., McDermott, A. and Samet, J. M. (2008), 'Mortality in the Medicare population and chronic exposure to fine particulate air pollution in urban centers (2000–2005)', *Environmental Health Perspectives* **116**(12), 1614–1619.

Zhou, M., Laszlo, I. and Liu, H. (2018), Preliminary evaluation of GOES-16 ABI aerosol optical depth product, *in* 'AGU Fall Meeting Abstracts'.

Zidek, J. V., Le, N. D. and Liu, Z. (2012), 'Combining data and simulated data for space–time fields: Application to ozone', *Environmental and Ecological Statistics* **19**(1), 37–56.