THESIS

UNDERSTANDING METEOROLOGICAL IMPACTS ON AMBIENT PM2.5 CONCENTRATIONS USING RANDOM FOREST MODELS IN BEIJING

Submitted by

Collin Brehmer

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2022

Master's Committee:

Advisor: Ellison Carter

Tami Bond
Kenneth Carlson
Jeffery Pierce

ABSTRACT

UNDERSTANDING METEOROLOGICAL IMPACTS ON AMBIENT PM2.5 CONCENTRATIONS USING RANDOM FOREST MODELS IN BEIJING

Policymakers and non-governmental organizations have been implementing policies and interventions designed to reduce exposure to hazardous air pollution. Having knowledge of how non-policy related factors (i.e., meteorology) impact air pollution concentrations in a given study area can better inform longitudinal studies of the effects of the policy on air pollution and health. In this study, we apply a random forest machine learning approach to evaluate how meteorological factors including temperature, relative humidity, wind speed, wind direction, and boundary layer height influence daily $PM_{2.5}$ concentrations in rural Beijing villages during heating months (January and February of 2019 and 2020). Ten-fold cross validation indicated good model performance with an overall r^2 of 0.85 for season 1, and 0.93 for season 2. The models were able to identify variables that were the most important for predicting $PM_{2.5}$ concentrations both field seasons (relative humidity) and variables that had changes in relative importance between seasons (temperature and boundary layer height). Additionally, examination of one and two-way partial dependence plots as well as interactions through Friedman's H-statistic granted insight into how meteorology variables influence $PM_{2.5}$ concentrations. Findings from this work provide a basis for adjusting for meteorological variability in important indicators of air quality like $PM_{2.5}$ concentrations in an ongoing real-world policy evaluation of a province-wide ban on household use of coal for space heating in Beijing, which is critical for isolating (to the extent possible) changes in measured pollutant concentrations attributable to the policy.

ACKNOWLEDGEMENTS

This work would not have been possible without the expert field management of Dr. Xiaoying Li. I want to thank Talia Sternbach, Xiang Zhang, Xinwei Liu, and the field staff for their assistance with data collection and management. I would also like to extend my gratitude to the village leaders, and study participants in China for letting us take measurements over many seasons. Additionally, Christian L'Orange and Kylie Rasmussen from Colorado State University for their assistance with sample and data processing. This work was funded by the Canadian Institute of Heath Research (project grants #148697 and #159477) and the Health Effects Institute (assistance award no. R-82811201).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Fine particulate matter air pollution ($PM_{2.5}$) is a heterogenous, particle phase air pollutant that is emitted from combustion, abrasion, and industrial processes. Evidence from epidemiological studies have identified associations between exposure to $PM_{2.5}$ and cardiovascular disease.[1–7] Emerging research has also identified $PM_{2.5}$ exposure as a risk factor for obesity, diabetes, and neurogenerative diseases.[8–11] These studies and others have established exposure to $PM_{2.5}$ ass a leading environmental health risk factor globally.[2] In addition to the well-studied adverse health outcomes associated with $PM_{2.5}$ exposure, certain chemical components of $PM_{2.5}$, like black carbon, can absorb heat and contribute to atmospheric warming.[12,13]

In response to the health and climate impacts of $PM_{2.5}$ pollution, governments, non-profits, and non-governmental organizations have enacted polices and interventions designed to improve air quality.[14] Such policies and interventions include those intended for short term (e.g. temporary restriction on traffic) and long term (e.g. improved vehicle standards) impact. A recent focus of air pollution interventions has been on reducing exposure to household air pollution generated from solid fuel-based cooking and heating. Approximately three billion globally cook and or heat their homes with solid fuel stoves, which generates high amounts of hazardous air pollution indoors as well as contributing to poor outdoor air quality.[15–17] Household air pollution is a persistent source of harmful exposures in China, where estimates through 2012 indicate that nearly 85% of rural households burn solid fuels to heat their homes and 40% of rural households still use some solid fuels to cook.[18] Interventions for settings with household solid fuel combustion include a range of sociological and technological efforts to introduce cleaner burning solid fuel stoves and/or fuels to promote transitions to electric or natural gas powered heating/cooking.[19,20]

A thorough evaluation of how air quality interventions work in practice (i.e. “accountability studies”) is critical for understanding how and to what extent they benefit public health.[21] In addition to pre- and post-intervention air quality measurements, rigorous accountability studies should account for confounding variables that can also influence air quality, including other regulations that may directly or indirectly influence air quality, adoption of policies, technologies and or strategies, and atmospheric conditions. One study found that wintertime ambient $PM_{2.5}$ concentrations in north and south east Asia could vary by up to 25% between years due to different meteorological conditions.[22] Given that China has enacted numerous policies designed to improve ambient air quality, knowledge of how meteorology impacts $PM_{2.5}$ concentrations over space (i.e. area where policy is enacted) and time (i.e. period when policy is expected to have an impact) is important in the broader context of efforts to evaluate the effectiveness an impacts of individual policies on air pollution.

In the past, accountability studies of air quality interventions have most commonly adjusted for meteorology in a few ways. For example, in regression models that are constructed to evaluate the effects of an air pollution intervention policy on one or several downstream health outcomes, researchers may include terms for common meteorological variables (e.g., temperature and relative humidity). In other examples, researchers run several emissions model scenarios with different air pollution emissions inputs that reflect the potential outcome scenarios of the air pollution intervention policy. The relationship between meteorology and air pollution concentrations is complex and non-linear, so having to correctly adjust each meteorological parameter to meet regression model assumptions can be challenging and make model interpretation difficult.[23] Additionally, the ability of emissions modeling to properly characterize and quantify relationships between air pollution and meteorological variables relies on the robustness of the underlying model assumptions and the accuracy of the emissions inventories which may not be available for all sources in every region. [24,25] The challenges encountered using parametric methods to model the impact of meteorology on $PM_{2.5}$ concentrations may be more easily handled by robust, non-parametric machine learning algorithms such as random forest models.[26] These models do not require pre-determined information about the structure (i.e., linear, spline, cubic) of the relationship between variables. Additionally, assumptions about the distributions of data are more relaxed than for parametric models, which is beneficial for environmental data that often have skewed distributions.

2. OBJECTIVES, OUTPUTS, AND OUTCOMES

This thesis contributes to a larger study that seeks to evaluate the impact of transitioning households from coal-based space heating to electric or natural-gas based heating on air quality and health. The context of this study is a natural experiment (i.e., observational study) where the coal-to-electricity energy transition occurs at the village level throughout the Beijing province in China. To successfully conduct the analysis for the larger study, variables that have the potential to confound the impacts of the policy on air pollution and health must be characterized and quantified. This thesis pursues the following objectives: (1) quantify the joint and independent effects of meteorological variables on ambient $PM_{2.5}$ during two heating seasons; (2) determine whether and how meteorological variables should be represented in the health-based analyses (i.e., mediation analysis of the impact of changes in $PM_{2.5}$ due to the policy on health) as confounding variables; and (3) propose approaches to representing meteorology in the mediation analysis. I hypothesize that the influence of meteorological variability on local, outdoor $PM_{2.5}$ where we are evaluating effects of a coal-ban policy may vary over time and should thus be accounted for separately in each season over the entire period of the analysis (four years). I propose random forest models as an effective model type for the above-mentioned objectives. An effective model in this study is one that meets objective 1 by correctly modeling the relationships between meteorological variables and $PM_{2.5}$, as well as providing results that are easily interpretable and useful for objectives 2 and 3.

With this hypothesis and rationale in mind, the work conducted as part of this thesis was designed to yield the following outputs: a measure of how strongly meteorological variables predict outdoor $PM_{2.5}$ in the study setting (i.e., predictability); a measure of how much including meteorological variables increases (or decreases) model accuracy (i.e., variable importance); the marginal effects of individual meteorological variables on prediction of outdoor PM in the study setting (i.e., partial dependence); and the simultaneous effect of two or more meteorological variables on prediction of outdoor $PM_{2.5}$ in the study setting partial dependence, and variable interactions. For objective 1, I will quantify the impact of meteorological variables on ambient $PM_{2.5}$ in our study by evaluating the predictability of random forest models that include meteorological variables. Differences in the predictability between models developed for two heating seasons will be used to assess the differences in impacts of meteorology on PM between heating seasons to determine if meteorology should be modeled separately for each season. Identification of meteorological variables that should be included in the mediation analysis (objective 2) will be assessed by comparing the relative variable importance and two-way variable interaction strength among meteorological variables. We will also use the shape of the partial dependence plots to inform how to model meteorological variables in the mediation analysis for objective 2. Further, our findings about the relationships and interactions between meteorological variables and $PM_{2.5}$ will be compared to known physical relationships in our study region. For objective 3, we synthesize our results of predictability, variable importance, partial dependence, and variable interactions to make recommendations about how to model meteorology in our study setting and, ultimately, correct for meteorological variability in our study setting so that we can evaluate the effectiveness of the emissions reduction policy in a real-world context.

3. REVIEW OF LITERATURE

3.1. $PM_{2.5}$ exposure and health

3.1.1. Sources of $PM_{2.5}$

Fine particulate matter ($PM_{2.5}$) air pollution is a mixture of airborne solid particles and liquid droplets with a diameter of 2.5 microns or less. Dominant sources of ambient $PM_{2.5}$ vary by region based on proximity to different sources. Urban areas across the world share three common sources of $PM_{2.5}$: resuspended crustal material (dust), vehicle emissions, and secondary aerosols. [27–31] Most urban areas also contain sources related to industrial processes in the region that are usually characterized by concentrations of specific metals related to the industry.[32–35] In regions that burn solid fuel for cooking and or heating, $PM_{2.5}$ resulting from the combustion of coal and or biomass may also be a present source.[36,37] Other sources in urban areas are more specific to the region and can include ship emissions in coastal cities, solid waste burning, construction dust, or oil refining.[31,32,36,38] Rural areas usually have fewer sources of ambient $PM_{2.5}$ than urban areas and are less commonly analyzed for sources. Rural areas that burn solid fuel for household energy needs can see large contributions from solid fuel combustion.[39,40] Depending on the proximity to a major city, rural areas can be subject to pollution transported from a nearby city. Secondary sources and dust have also been found to contribute to ambient $PM_{2.5}$ in rural areas.[39]

In addition to emissions from a range of sources and source activities, ambient $PM_{2.5}$ concentrations can be impacted by meteorological conditions.[23,41,42] Variables like temperature, relative humidity, wind speed, wind direction, and boundary layer height can impact the dispersion, transport, deposition, and rate of particle formation for $PM_{2.5}$. For example, increased temperature and relative humidity can increase secondary aerosol formation, which is why it can be more common to observe higher contributions to ambient $PM_{2.5}$ from secondary sources in the summertime.[43–45] Higher wind speed and boundary layer height can increase horizontal and vertical dispersion respectively, reducing ground level $PM_{2.5}$ concentrations.[46,47] Wind direction can positively or negatively influence $PM_{2.5}$ concentrations depending on the proximity and direction relative to sources.[48]

2.1.2. Health effects of $PM_{2.5}$

Exposure to $PM_{2.5}$ can be more hazardous than exposure to larger particles (e.g., PM_{10}, total suspended particulates (TSP)) because $PM_{2.5}$ is small enough to penetrate deep into the lungs and be absorbed into the blood stream.[3] The mechanisms by which exposure to $PM_{2.5}$ cause disease is still an area of active study, but evidence has been presented for several biochemical mechanisms. For example, $PM_{2.5}$ may cause disease through the generation of cellular reactive oxygen species (ROS).[49,50] Redox-active components of $PM_{2.5}$ like transition metals and polycyclic aromatic hydrocarbons (PAHs) can react in cells and generate ROS which can lead to cellular damage and disease.[49,51–54] Exposure to $PM_{2.5}$ has also been found to be associated with the upregulation of inflammation-related genes.[55] Further, chronic inflammation has been associated with several chronic diseases including hypertension and cardiovascular disease.[56,57]

Decades of epidemiological evidence provide, arguably conclusive, evidence that human exposure to $PM_{2.5}$ is associated with increased risk of cardiovascular disease.[1,3,7,58] In the global burden of disease studies, preventable exposure to air pollution is a leading cause of death globally and was estimated to have contributed to 4.58 million deaths in 2017.[59] More recent evidence has linked $PM_{2,5}$ exposure to increased risk of diabetes, obesity, and neurological disease.[8,10,11] Not all areas of the world are impacted by $PM_{2.5}$ exposure equally. People living in low-income countries, the elderly, and small children are those who may experience elevated exposures, as well as elevated susceptibility to health effects of $PM_{2.5}$ exposure[2,59] Leading environmental risk factors include exposure to household air pollution from burning solid fuels indoors and at home, which is more common in low-income countries.[15,60]

3.2. Accountability studies of air quality interventions

3.2.1. Accountability study frameworks

An accountability study is one that aims to assess the effectiveness of a given policy or intervention. In 2003, the Health Effects Institute (HEI) published a conceptual framework outlining the "accountability chain" which identifies the relationship between a regulatory action and a human health outcome by evaluating relevant "links" in the chain.[61] In order, the links are regulatory action, emission, ambient air quality, exposure/dose, and human health response. This framework also outlines potential confounding variables that can affect the relationship between any two sequential links in the chain. For example, how much a group targeted by a policy complies with said policy will impact how much the regulatory action impacts emissions. The purpose of each link past the "regulatory action" link is to provide researchers with specific outcomes that can be used to evaluate the impacts of a regulatory action. Under this framework, each link should be evaluated in full to understand the regulation at a mechanistic level.

In contrast to the HEI suggested accountability chain, Zigler and Dominici 2014 propose a "direct accountability" approach for evaluating the effectiveness of a regulatory action.[62] In the direct accountability framework, relationships between regulations, air quality, and health outcomes are evaluated using statistical methods. Zigler and Dominici argue that using a potential outcomes study design, some of the links in the accountability chain do not need to be examined directly. "Potential outcomes" refers to the issue of only being able to observe the air quality under one of following scenarios: 1. The policy is enacted; 2. The policy is not enacted. In an ideal scenario from the research design perspective, we would be able to measure the air quality under both scenarios and compare the difference to evaluate the effectiveness of a given regulation. However, this is infeasible. Instead, to replicate the "counterfactual" (i.e., observing air quality under the conditions without regulatory action), we can select a control set of conditions that is similar in ways to the treated set of conditions in ways that are related to air quality and health. Selecting an adequate control allows for the application of statistical methods that do not require evaluation of every link in accountability chain chain.

3.2.2. Examples of previous accountability studies

Extensive reviews that synthesize the results of accountability studies related to air quality have been published elsewhere.[24,63,64] The focus of this section will be on the examination of two accountability settings where initial evaluations of the regulations found them effective at improving air quality and health endpoints, but later studies called the initial findings into question.

The first example we turn to is the the 1996 Summer Olympic Games, which were hosted in Atlanta, Georgia. To reduce traffic related air pollution during the games, several control measures were put into place. Examples of control measure include an increase in public transport available 24-hrs a day during the Games, restricting traffic in the downtown area near the games, and efforts to reduce vehicle travel during commuting hours. A 2001 analysis by Friedman et al., was one of the first to evaluate the impact of short-term traffic-related air pollution interventions and is frequently cited as evidence that they are effective.[65] Their analysis found a reduction in hospital visits associated with pediatric asthma, as well as reductions in outdoor ozone during the Games. Additionally, they noted that meteorological conditions during the Games did not differ by a large amount compared to the baseline measurements. In 2010, Peel et al. noted the limitations of the Friedman and colleagues 2010 study and performed a follow-up analysis using additional meteorological and health data.[66] While Peel et al. (2010) found evidence of reduced ozone concentrations, they attributed them to meteorological conditions that were not conducive to ozone production. Additionally, they did not identify changes in respiratory- or cardiovascular-related emergency department visits for adults or children during the Games after adjusting for seasonal trends in air pollution concentrations and health outcomes during the years before and after the Games.

In a second example, we consider 29th Olympic Games held in August of 2008 in the capital city of Beijing by host country China. As a stipulation of hosting the Olympic and Paralympic events, the Chinese government agreed to reduce emissions and improve air quality for the duration of the Games. Starting in late July of 2008, local control measures including the closure of heavily polluting factories, replacement of coal-burning furnaces with natural gas replacements, and restriction of on-road vehicle traffic patterns. Similar measures were also taken in nearby regions of China to reduce the impact of regional transport on air quality in Beijing. The planned nature of these restrictions let researchers use this as an opportunity for a "natural experiment" or observational study (discussed in section 2.2.3).

Several studies in Beijing recruited participants and measured baseline health and air quality prior to the restrictions being implemented. X. Wang 2009 and M. Wang 2009 attributed reductions in measured outdoor black carbon (a chemical component of PM), carbon monoxide, nitrogen oxides, and traffic-related volatile organic compounds to traffic-control measures.[67,68] Ambient concentrations of ozone, carbon monoxide, nitrogen oxides, and sulfur dioxide concentrations were also found to be lower during the Games compared to the previous two years.[69] Additionally, improvements in health endpoints were measured for biomarkers of inflammation, birthweight, peak expiratory flow, and asthma related hospital visits.[70–73] Most of the aforementioned studies related to the Beijing Olympic Games air emissions interventions account for meteorology in some way, most commonly by including variables for temperature and relative humidity in their statistical models. In 2009, W Wang and colleagues preformed an in-depth meteorological analysis and found that meteorological conditions accounted for 40% of the variability in PM_{10} during and post-Games, whereas source control only account for 16% of the variability.[74] Few of the previous studies that attempted to control for meteorology mention the effect of meteorology in these situations. While these studies predominantly concluded that the air pollution emissions restrictions in place during the Games had positive impacts on multiple health and exposure endpoints, the work by W Wang et al. (2009) shows that they failed to elucidate the multiple pathways through which air pollution concentrations and health endpoints may be changing.

3.2.3. The role of observational studies for accountability

The gold standard for evaluating the effect of any treatment on any outcome is a randomized controlled trial (RCT).[75,76] RCTs are appealing to scientists because through the process of random treatment assignment, the relationship between confounding variables and the outcome is severed. This allows for simple comparisons of the outcome between the treatment and control group to make causal statements about treatment effectiveness. A RCT approach would not work for air pollution studies for several reasons. First, it would be unethical to forcibly expose humans to air pollution given the well documented health effects. Second, regulations are large in scope, expensive, and logistically difficult to implement without having to worry about doing so in a RCT framework. Third, it is difficult to blind participants to treatment status, which may impact behaviors associated with the health or air pollution outcome of interest.

Observational studies are perhaps the most common study design in environmental health because they do not suffer from the ethical dilemma that a RCT would.[76,77] In an observational study, researchers observe an outcome of interest without having a direct say in what the treatment is, who receives the treatment, and how/when the treatment is administered. This lack of randomization and control of the study design means that observational studies are subject to bias by confounding variables. Around the turn of the 21st century, researchers proposed a potential outcomes framework for addressing causality in observational studies. The potential outcomes framework states that casual effects of an intervention are the difference between the outcomes of the person's current treatment status (e.g., treated) to what it would be if they had the opposite treatment status (e.g., untreated). Therefore, causal statements about the impact of interventions measured through observational studies can be made with sufficient understanding of the potential outcomes under opposite treatment status (i.e., "counterfactual"). Details of adjusting for confounding and constructing a counterfactual scenario are discussed in the following section.

3.3. Tools for addressing meteorology in accountability studies

3.3.1. Summary of current approaches for addressing meteorology in accountability studies

The current approach to accountability studies can be divided into two main categories based on their outcome of interest. The first major category of air pollution accountability studies evaluates the impact of a regulation on air quality through two major analysis routes. The first analysis route consists of applying statistical tools for detrending meteorological effects like

Autoregressive Integrated Moving Average models or Kolmogorov-Zurbenko (KZ) filters to time series of air pollution concentrations.[78,79] These methods use multiple moving averages taken across several different time periods of interest. For example, passing a time series of pollution concentrations through the developed KZ filter can remove the impacts of long term and seasonal meteorological effects, leaving behind a residual signal that can be interpreted as short term meteorological impacts. Regression methods (i.e., multiple linear, general additive, generalized estimating equations) with the residual signal as the response and meteorological variables as the predictors can then be applied to determine the associations between meteorological variables and air pollution concentrations. Statistical detrending methods require long-term measurements of pollutions over multiple years to get accurate estimates of meteorological impacts. Some interventions are implemented for only short periods of time (i.e., emissions restrictions for the Olympic Games), in which case, detrending may not be feasible. Additionally, studies of interventions administered in areas of the world that lack extensive monitoring networks (i.e., rural China) are usually too resource-constrained to conduct multi-year, real-time measurements of pollutants. However, statistical detrending methods can still prove useful in areas where sufficient data are available.

The second most common way to understand how a regulation impacts air quality is through the use of chemical transport models (CTMs).[24,80–82] Commonly used CTMs are the Community Multiscale Air Quality model (CMAQ) and the Goddard Earth Observing System chemistry model (GEOS-Chem). Both the CMAQ and GEOS-Chem models use emissions inventories, meteorological variables, and atmospheric chemistry to simulate air pollution concentrations. To evaluate the effect of a particular air pollution regulation or policy, changes to an emissions inventory based on the expected impacts of the regulation can be made and compared to a model with the non-altered emissions inventories as an input. CTMs are very comprehensive and incorporate decades of knowledge on metrological and atmospheric chemistry impacts on air pollution. They can be useful for predicting the impacts of a regulation over large spatial scales. However, CTMs are limited by the quality of the emissions inventories supplied to them. Up to date emissions inventories are not available for every region of the world. Additionally, these models do not take into consideration human behavior which can impact policy adoption, and therefore emissions, especially if a policy is implemented in a heterogenous manner over time and space (e.g., over multiple years across several or many cities or villages).

The second common category of accountability studies are epidemiological studies that evaluate the health impact of air pollution regulations.[65,66,70,71] These studies use regression-based methods such as mixed effects models, general estimating equations, or general additive models to quantify the relationship between $PM_{2.5}$, a regulation, and continuous health outcomes like blood pressure or binary health outcomes like hospitalizations. A common statistical framework within in a cohort study is the difference-in-difference (DiD) approach, which compares the difference of a response between the treatment and control group pre- and post-treatment.[83,84] As mentioned in section 2.2.1, selecting a control group with similar characteristics for variables that are relevant to the outcome can strengthen the statistical power of an observational study by eliminating, or at least accurately account for, confounding variables. Most studies account for meteorology by including, most commonly, terms for temperature and relative humidity in the regression model statements. While temperature and relative humidity are important meteorological variables to consider, there are other variables relevant to consider such as wind speed and boundary layer height. Additionally, the relationship between $PM_{2.5}$ and meteorological variables can be non-linear and include interaction terms that may be hard to address in regression models. Not including important confounding variables or mis-specifying the outcome-confounder relationship in a regression model can lead to inaccurate estimations of the treatment effect.

Air quality interventions (i.e., coal ban, clean cookstove programs) impact several variables ("mediators") that are associated with health outcomes including exposure to multiple pollutants, household temperature, and subjective wellbeing. Mediation analysis can be applied to disentangle the "indirect effects" of the policy on health caused by one mediator variable (e.g., $PM_{2.5}$ exposure or outdoor $PM_{2.5}$) compared to other mediator variables (e.g., indoor temperature) as well as the "direct effect" of the policy itself on health. This type of analysis is useful because it enables researchers to make causal statements about how much of the impact of a given air quality intervention on health is due to changes for a given mediator (e.g., $PM_{2.5}$) which gives insight into how the intervention works. However, this type of analysis is challenging because it requires accounting for as many confounding variables, including meteorology, as possible.

Mediation analysis is sensitive to mediator-outcome confounding, among other confounding biases. Within the context of the worked presented herein, meteorological variables like temperature and relative humidity can impact both the health outcome (e.g., blood pressure) and $PM_{2.5}$ concentrations. Because mediation analysis is sensitive to confounding in this way, methods to control for confounders during statistical analysis stage are required. The two most common ways of controlling for mediator-outcome confounding in mediation analysis are sequential g-estimation and inverse probability weighting with propensity score matching. Sequential g-estimation estimates the controlled direct effect by first modeling the outcome as a function of the mediator (e.g., $PM_{2.5}$), policy (e.g., coal ban), and confounders (e.g., meteorological variables). Next, a new outcome variable is calculated by subtracting the effect of the mediator from the outcome variable. Finally, the new controlled outcome variable is modeled as a function of the mediator and confounders.

Inverse probability weighting takes a different approach by first modeling the mediator as a function of confounding variables. Each observation is then assigned a weight based on how much the mediator can be explained by the confounders. Observations receive a lower weight the more the mediator can be explained by the confounding variables. The weights are then included in a standard regression where the outcome is modeled as a function of the outcome, mediator, and policy. Inverse probability weighting has been shown in a comparative simulation study to be sensitive to properly modeling the mediator-confounding relationship, which in our study would include the $PM_{2.5}$-meteorology relationship.[85]

3.3.2. Machine learning approach

The application of machine learning (ML) in academic research has risen in popularity in recent years, in part, due to the development of user-friendly packages for commonly used statistical programing languages like R and Python.[86–88] Examples of commonly used ML methods include random forest, XGBoost, Bayesian regression trees, and neural networks. Most regression-based ML methods employ a decision tree-based method to develop a predictive model. Specific details of the method used in this study will be discussed in the methods section and the basics of a decision tree model will be described here. Decision trees are constructed by testing which value for a particular predictor variable does the best job at sorting the response variable into two separate bins.[89] For example, predicted $PM_{2.5}$ could be sorted into two bins based on whether the associated relative humidity was above or below a certain value. The response then continues to be sorted by predictor values until some criteria has been met (e.g., some function is optimized. or the new way of sorting is not better than the previous way of sorting).

Some ways that tree-based ML methods differ from one another are how large a decision tree grows, how much influence each variable has, and how a stopping point is determined. Decision trees in random forest models are weighted equally and continue to grow until the sorting of a new node no longer improves on the sorting of the previous node. In contrast, decision trees in XGBoost consist of only one sorting variable per tree and each tree assigned a weight based on a selected optimization function. The random forest approach was chosen over other decision tree models because its relatively low barrier to understand and implement.

ML methods can be useful tools for analysis because they are not bound by the same assumptions as standard regression techniques. First, ML does not require prior knowledge of the type of relationship between a predictor and response. In comparison, regression methods usually require transformation or other adjustments (e.g., polynomial terms, splines) to the model statement to meet the assumption of a linear relationship between a response and predictor. Additionally, due to the sequential and random nature of variable evaluation at each step of constructing a decision tree, ML models fit interactions between predictor variables. For example, let us define hypothetical scenario (equation 1) with response Y as a function of predictors x_1, x_2, x_3, and x, with two terms, in this case, representing interactions between x_1 and x_4, and between x_2 and x_3.

$$(1)\ Y = x_1 + x_2 + x_3 + x_4 + (x_1 \times x_4) + (x_2 \times x_3)$$

An example model statement for a ML method would be (equation 2):

$$(2)\ Y = x_1 + x_2 + x_3 + x_4$$

At some point in the construction of the decision tree, x_1 will be selected and determined to be the best variable for sorting the response at that point of the tree. At the next sorting step, if x_4 is the best variable to sort by it would be able to reflect the interaction between itself and x_1, because the variable had previously been sorted by x_1. The variable x_4 has a higher likelihood to be chosen as the next variable to sort the predictor by because it should minimize any sorting error by modeling the interaction term.

The most common criticism of ML models is their "black-box" nature, meaning researchers are not able to observe or determine how the model is making decisions or interpret the implications of those decisions. Previously, the lack of ML model interpretability has limited their use to predictive tasks, where an understanding of how the model is reaching its' decisions is less important than in descriptive models. Recent research has led to the development of explanatory tools to increase the transparency of ML models. The explanatory tools extend the use of ML models past prediction and enable descriptive information to be extracted from the model. Being able to identify descriptive information is advantageous given that ML models are useful for modeling data with complex underlying relationships. While the descriptive information may not be causal in all cases, it can still provide insight into relationships that are important to consider in explanatory models.[90]

Here, I present a set of common explanatory tools and metrics for ML models that can be used for a range of descriptive purposes. To begin, to determine which variables are contributing to the prediction of a given response and their contributions relative to one another, variable importance metrics can be calculated for predictor variables to determine which variables are contributing to the prediction of the response and their contributions relative to one another. If, in addition to this, we would like to assess the marginal effect of a predictor on the response variable (similar to a regression coefficient), partial dependence plots (PDPs) are appropriate.[89] PDPs are calculated using the same method as Pearl's back-door adjustment (i.e. a generalizable approach to adjusting for the effect of a measured confounder in conventional causal analysis frameworks), so it has been argued that they can also be interpreted casually in certain circumstances.[91,92] To instead to understand the conditional effect of a predictor on the response, individual conditional expectance (ICE) can be constructed.[94] Further, Friedman's H-statistic can be calculated to quantify how much of the main effect (i.e., marginal effect from the partial dependence plot) of a predictor variable is explained by interactions. Information about interactions is useful for understanding how combinations of values of two or more variables may alter the predictor-response relationship shown in the partial dependence plots. In the context of this work, collectively, these explanatory tools are useful for determining whether and how meteorological variables should be included in the mediation analysis (objective 2) as well as showing how meteorology could be represented in the mediation analysis (objective 3).

ML models have been applied to predict $PM_{2.5}$ concentrations in several studies with high predictive power ($r^2 > 0.8$) using satellite measurements, meteorology, and land use variables. Most studies use satellite-based aerosol optical depth (AOD) values as surrogate measure of $PM_{2.5}$ concentrations over the study area.[95–98] Robust relationships between AOD and $PM_{2.5}$ have been established, but there are limitations to using AOD over measured PM concentrations including missing values due to cloud cover and low spatial resolution.[99] As a result of using satellite measurements in these ML modeling efforts, most studies tend to predict $PM_{2.5}$ concentrations over a large spatial area with low spatial resolution (e.g. country, county, state).[100,101] Doing so requires interpolation of $PM_{2.5}$ concentrations from AOD or ground monitors which may not accurately reflect the spatial heterogeneity within a particular grid cell.

To our knowledge, very few studies have applied ML methods and explanatory tools to model environmental data collected on the ground in the field as part of an observational study. At present, it is unclear how capable ML methods are for modeling known physical relationships from field-based observational data. Given the widespread use of observational studies in environmental health settings, and the need for tools to account for confounding variables like meteorology, especially in field-based observational studies of air quality policy interventions, more evaluations of ML methods and explanatory tools in the context of observational studies, such as the work presented here, is warranted.

4. METHODS

4.1. Study Setting

This analysis uses data that were collected from November to February in 2018/2019 and 2019/2020 as part of a larger longitudinal study designed to assess the air pollution and health impacts associated with the transition from coal-based residential space heating to the use of electric and natural gas-powered heat pumps for household heating (Figure 1). Specifically, outdoor air quality data were gathered from 49 villages in the Beijing region, situated in the rural, mountainous regions of the Miyun, Huairou, Fangshan and Mentougou districts. These villages were located far (> 40 km) from the Beijing city center, where outdoor air pollution concentrations were most likely to be most strongly influenced by local sources. At baseline (i.e., winter 2018-2019), households in these villages reported using coal and biomass as their primary sources of fuel for household space heating.

4.2. PM2.5 Measurements and Calibration

$PM_{2.5}$ concentrations were collected at 1-minute resolution using ZeFan sensors (www.zfznkj.com), which measure PM via light-scattering using the Plantower PMS7003 sensor. These sensors have shown high agreement with reference instruments in several studies.[102,103] One to three sensors were placed in each village during the heating season (Table S1). The Spearman correlation coefficients between sensors in the same village were generally above 0.80 (Figure S1), so reported values and values used in this analysis represent the average of $PM_{2.5}$ concentrations measured by all sensors in a given village. All sensors were calibrated using two methods. First, before and after sensor deployment for each season, they were co-located and calibrated against two standard reference instruments (TEOM 1400A and BAM model 5030) over the course of approximately 10 days. Additionally, a correction factor was developed by comparing time-integrated sensor concentrations to filter-based $PM_{2.5}$ measurements collected in each village throughout the duration of each field season. These filters were collected from each village approximately every 7 days during the heating season (November-February) on Zefluor filter media (PTFE 37 mm with 2 μm pore size, Pall Labs) using ultrasonic personal air samplers (Access Sensor Technologies, CO, USA).[104] Briefly, air was actively sampled continuously at a rate of 1.0 L min^{-1} over filter sampling media (Zefluor PTFE 37 mm with 2 μm pore size, Pall Labs). Sampling flow rates were calibrated and checked before and following each UPAS deployment using a mass flow meter (Alicat Scientific, AZ, USA). Additional information about sensor calibration, the development of correction factors, and gravimetric analysis is available in the appendix (Text S1 and Figures S2-S6).

4.3. Meteorological Data

Hourly boundary layer height, 2-m temperature, 2-m dew point temperature, and 2-m horizontal wind speed components (u, v) were obtained from the European Center for Midrange Weather Forecasting ERA5 reanalysis dataset (0.25 x 0.25 resolution).[105] Village-level meteorology values were found by identifying the four surrounding grid points with values available from the ERA5 reanalysis, and then applying inverse distance weighted interpolation from those four grid points to the village. Percent relative humidity was calculated from the 2-m dew point temperature using the "weathermetrics" package (version 1.2.2) in R.[106] Total hourly wind speed and wind direction were calculated from the horizontal wind speed components.

4.4. Random Forest Modeling

Random forest (RF) models are a machine learning approach that generates an ensemble (forest) of independently constructed decision trees that are used for prediction.[107,108] Each tree is built by selecting a set number of predictor variables at each node, sorting the values of the outcome variable by the values of the predictor variables, and selecting the predictor variable that has the lowest sorting error for each node. Sorting error is determined at each node by evaluating residual sum of squares for all possible binary classifications of each variable selected to be tested. The tree is fully constructed once the sorting error of a new node does not improve upon that of any previous nodes. The models have two main parameters: the number of trees in each forest and the number of parameters tested at each node. The value for these parameters that minimize the prediction error can be found by generating forests parameter values and selecting the value that results in the lowest prediction error. For this study, the number of trees grown per forest was 500 and 2 variables were tested at each node.

Models of daily $PM_{2.5}$ were constructed using the meteorological variables mentioned previously, as well as dummy variables for village, district, and day of year. Due to differences in sensor deployment between the first season and second season, the number of observations in November and December in season one was considerably lower than in season two. Therefore, we restricted our analysis to only data collected in January and February in season one and December, January, and February in season two to achieve as balanced a dataset as possible, with respect to the number of observations per month. Unbalanced data can lead to the model being better at predicting outcomes in the level of a variable with more observations because it has been trained on more observations of that type.[109,110] The models in both seasons include values from January and February of the respective field season and the season one model was built using data from 38 (76% of total study villages) villages, whereas the season two model was built with data from 49 (98% of total study villages) villages. The capacity for the season one data to predict the season two data was evaluated only for the villages and days of year included in the season one model.

The prediction results from the models generated in this study were evaluated using 10-fold cross validation. In this approach, the data is randomly divided into 10 equally sized bins ("folds") that each contain 10% of the data. One of these folds is held out ("out-of-bag"), and a forest is generated using the other 90% of the data. This is repeated 9 times for a total of 10 forests, with each iteration holding out a different fold. The overall model root mean square error (RMSE) and r^2 is calculated by a simple linear regression using the observed values of the held-out folds and the predicted values from the forest where that fold was held out. These models were created using the "caret" package (version 6.0-86) in R.[88]

Variable importance for each variable is calculated by permuting each predictor variable and comparing the out-of-bag prediction error to the unpermuted case.[111] The difference is averaged across all trees and normalized by the standard deviation. Variable importance is given as percent increase in MSE (when the variable is permuted versus not). A larger percent increase in MSE indicates a higher degree of importance relative to variables with a lower percent increase in MSE. Partial dependence curves were calculated for each meteorological variable included in the model. For each observation, all predictor variables were held constant except the predictor variable of interest (x_i), and predictions for each observation at each value of x_i in the dataset were made. The predictions for each observation at each value of x_i were then averaged to produce the final partial dependence plot. The derivation and additional explanation of the partial dependence plots can be found elsewhere.[112] For the purpose of this thesis, one of the main objectives was to determine whether and how meteorological variables should be represented in a health-based mediation analysis. The partial dependence plots contribute to this objective by showing the shape of the relationship between a predictor and response variable, which can then be used to transform the meteorological variable in a mediation analysis that assumes linearity.

Two-way interactions between select meteorological variables were quantified by Friedman's H-statistic and given as the percent of the effect of a given predictor variable on the response due to interactions with other predictor variables. Briefly, the H-statistic is calculated by comparing the variance of partial dependence of a selected predictor variable to the variance of the partial dependence of all other variables (or a second variable of interest for two-way interactions). Detailed mathematical explanations of the H-statistic can be found elsewhere in the literature.[113] In the context of this work, understanding how different combinations of different predictors (i.e., meteorological variables) impact the response (i.e., ambient $PM_{2.5}$) is critical to determining how to represent these variables, which are confounders, in the larger study context (i.e., mediation analysis to understand the impact of an air quality policy intervention on pollution levels). The evaluation of two-way interactions is valuable for further understanding how combinations of different predictors impact the response (objective 2). For this thesis, the following two-way interactions are presented as a demonstration for the season two data only: relative humidity-boundary layer height, wind speed-wind direction, and wind speed-boundary layer height. These interactions were investigated because they were the strongest interaction (relative humidity-boundary layer height) for the variable with the highest importance (relative humidity) or have a known interaction in the literature (wind speed-wind direction) or are related to air stagnation (wind speed-boundary layer height).[114,115]

Random forest models have been shown to be sensitive to spatial and temporal autocorrelation due to the potential for the formation of non-spatially and non-temporally independent folds.[116] To account for this, we also preformed separate spatial and temporal 10-fold cross validation. The folds in these models were not randomly generated, but rather split by the spatial and temporal variable of interest. For this study, the spatially-balanced cross validation models were generated on folds that each contained data for 10% of the total villages in the study. The folds for the temporally-balanced cross validation were split so that each had data for 10% of the days of the year. The results from these models give insight into how well the overall model, where the folds were split randomly, preform over space and time. The folds for the spatial and temporal cross validation models were generated using the "CAST" package (version 0.5.1).[117]

We conducted several sensitivity analyses to determine the impacts of our data and variable choices on the predictability of the season one model for the season two data. First, we developed a model that did not contain the day of year variable, which could absorb some unexplained variation each year without having it be useful for predicting the next year. We also limited the observations in season two that were predicted by the season one model to those that had meteorological values within the range measured in the season one to prevent extrapolation by the season one model for the season two data. Finally, villages that had transitioned from coal to electricity-based heating between seasons one and two were excluded from the season two data being predicted by the season one model.

5. RESULTS AND DISCUSSION

5.1. Descriptive statistics

Boxplots of daily average $PM_{2.5}$ and meteorological parameters by district and month are provided in the appendix (Figure S7). Median (inter quartile range) daily $PM_{2.5}$ concentrations during the study durations in seasons one and two respectively were: Miyun: 38.1 (39.5) µg m^{-3}, 52.9 (50.9) µg m^{-3}; Huairou: 33.6 (33.4) µg m^{-3}, 31.4 (35.5) µg m^{-3}; Fangshan: 67.3 (51.4) µg m^{-3}, 66.1 (57) µg m^{-3}; and Mentougou: 34.3 (41.3) µg m^{-3}, 42.1 (41.4) µg m^{-3}. Outdoor $PM_{2.5}$ concentrations in these districts were lower than in rural regions of the nearby Northern China Plain, where average wintertime concentrations in recent years have ranged from 75 µg m^{-3} to over 100 µg m^{-3}.[118–120] Pollution concentrations were also lower than the typical wintertime average of at least 80 µg m^{-3} in the Beijing city center.[121,122] Notably, the average daily concentrations in all districts were much higher than the newly released World Health Organization guideline of 5 µg m^{-3} annual average and 15 µg m^{-3} 24-h average.[123]

5.2. Random forest models

Results for the overall, the spatially-balanced, and the temporally-balanced 10-fold cross validation models the first winter season are shown in Figure 2. The overall model had a relatively large r^2 value of 0.85 and a RMSE of 13.1 µg m^{-3}. For the spatially-balanced and temporally-balanced cross validation models, the r^2 decreased to 0.76 and 0.54, respectively. The RMSE for these models also increased to 16.6 µg m^{-3} for the spatially-balanced model, and 22.8 µg m^{-3} for the temporally-balanced model. Figure 3 shows the predicted daily $PM_{2.5}$ concentrations for the season two data using the season one model. The season one model in general overpredicted the season two data and performed relatively poorly with large error (r^2 =

0.45, RMSE = 33.5 μg m^{-3}). Figure 4 presents the results for the overall, spatial, and temporal cross-validation models developed with the data from the second winter season. The models developed using the season two data predict the season two data much better than the season one model (overall $r^2 = 0.93$, RMSE = 11.1). The spatially-balanced and temporally-balanced season two models also see decreases in the r^2 (spatial = 0.85, temporal: 0.73) and increases in the RMSE (spatial: 16.3 μg m^{-3}, temporal: 22.8 μg m^{-3}).

The high within-season predictability of our models shows that meteorological variables are important predictors of ambient $PM_{2.5}$ in our study and should be controlled for in a mediation analysis. These findings contribute to objective 1 of determining the joint impacts of meteorology on $PM_{2.5}$ in our study. Random forest models that include meteorological and land use variables tend to have the highest predictability ($r^2 > 0.80$) among random forest models that predict ambient $PM_{2.5}$ because they include information that accounts for most of the variability in $PM_{2.5}$ concentrations.[100,101,124] However, our models that included only meteorological variables approach the same predictability of those that have included additional variables (i.e., land use). For example, the random forest model constructed in Huang et al. 2018 to predict $PM_{2.5}$ concentrations in the Northern China Plain, including both rural and urban areas, ($r^2 =$ 0.88; RMSE = 14.89 μg m^{-3}) included 34 variables such as fraction of various surface cover, population density, and cloud cover. Our findings that fewer variables were needed to achieve similar levels of predictability may be explained, in part, by the potentially smaller number of sources that contribute to $PM_{2.5}$ in rural relative to urban areas. For example, a land use variable, like length of roadway in a given area, that could be used as a proxy for roadway emissions would probably not have been beneficial in our rural study setting where the contributions of traffic to overall $PM_{2.5}$ are much smaller compared to urban settings.

The worse performance of the spatially-balanced and temporally-balanced cross validation models indicates that the overall model does not predict as well across villages or across days of the year. In the spatially-balanced cross validation model, the data for 10% of villages was contained in each fold (~5 villages in each fold), eliminating the use of the "village" variable for prediction because the model did not include any data from villages assigned to the held-out folds. This would be similar to removing the village random effect in a mixed effects model or the village specific intercept in a regular linear model. These village level adjustments are important in the overall model because, without them, our data trends ("shrinks") more towards the population mean. We don't observe a similar shrinkage towards the mean in the temporally-balanced cross validation models, where each fold contained data for 10% of the days of the year measured in our study. However, we do observe a more general, balanced, spreading of the data around the 1:1 line compared to the overall model. This suggests that the predictability of the season one and season two overall model is inflated by temporal autocorrelation.

Our findings highlight the importance of spatially-balanced and temporally-balanced cross validation in air pollution studies. These cross validations improve interpretability of the model results by showing how the spatial and temporal effects, if not accounted for explicitly as was done here, may cause us to overestimate the prediction power of the other variables included in the model (i.e., meteorological variables). For example, one study designed to predict forest biomass variation found that after spatially-balanced cross validation, their overall model performance ($r^2 = 0.53$) was heavily influenced by the spatial variable included (spatially-balanced cross validation model r^2: 0.14). We recommend that future studies using time series data in random forest models apply similar spatially-balanced and temporally-balanced cross validations to contextualize the model's predictive power.

The poor predictability of the season one model for the season two data was robust to sensitivity analyses that included restricting the season two data to observations with meteorological values in the range of the season one values, excluding villages that had transitioned from coal to electricity-based heating between season one and two, and removing the day of year variable (Table S2). These sensitivity analyses were conducted to evaluate if the poor predictability of the season one model for the season two data was due to extrapolation, changes in energy use, or use of a variable that may not be useful for predicting between years (DOY) . The individual and combined sensitivity analyses had small impacts on predictability (change in r^2: -0.03 - +0.04). Additionally, the season two model outperformed the predictive power of the season one model for the spatially and temporally-balanced cross validation. Together, these findings provide compelling evidence that the variables used in our model, including meteorological variables, were not as good of predictors in the first season compared to the second season. The poor predictability of the season one model for the season two data and the difference in predictability between the two models demonstrates the limitations of using the understanding of meteorological impacts from only one year to the next year. Local meteorology is shaped by larger, synoptic patterns (like the East Asian Winter Monsoon in our study setting) that can vary considerably year to year. The chemical components of $PM_{2.5}$ also interact with meteorological variables differently (discussed in detail in the section 5.3.), so variations in the $PM_{2.5}$ chemistry between season one and two may be contributing to the differences in predictability of meteorological variables in season one compared to season two. Our findings indicate that in short-term observational studies, such as the Beijing household coal ban policy under consideration here, in which we have measurements spanning multiple years, the impacts of meteorology on ambient $PM_{2.5}$ should be modeled on a year-to-year basis.

While we used daily averaged $PM_{2.5}$ and meteorology values, our models still also have some spatial and temporal autocorrelation that can be difficult to account for in observational study settings where researchers can be limited by when and where samples are collected. Future studies using observational data that are focused on model predictability may consider using convolution layers as an additional variable in the random forest model(s) to reduce the effect of autocorrelation on model output.[125] These layers group data that is highly correlated in space and time to improve model performance by reducing spatial and temporal autocorrelation. However, previous examples of applications of this approach have found only small improvements in model predictability (r^2 increases of ~0.02-0.06).[100,125]

5.3. Variable importance and partial dependence

Variable importance is given in Figure 5 as percent increase in mean squared error (MSE) from the base model compared to a model where the variable had been randomly permuted. Across both seasons, relative humidity was the most important variable by a large margin. Boundary layer height and temperature were more important in season two compared to season one, whereas wind speed was more important in season one compared to season two. While the importance of wind direction, village, and district are low relative to other variables, the increase in MSE when these variables was randomly permuted greater than 250%, which indicates that they are still important to include in the models.

Partial dependence plots for the meteorological variables included in the models are given in Figure 6. The partial dependence curves for each variable follow a similar trend across seasons. Relative humidity (Figure 6A) and temperature (Figure 6B), generally, have a constant relationship with $PM_{2.5}$ at the tail ends of the values and a positive, monotonically increasing relationship in in the middle range of values. Boundary layer height (Figure 6C) and wind speed (Figure 6D) were negatively associated with daily $PM_{2.5}$ at lower values and had constant relationships at higher values.

The general trend in the relationships observed between boundary layer height, wind speed, and wind direction to predicted ambient $PM_{2.5}$ concentrations observed in the partial dependence plots are consistent with the current literature. Boundary layer height and wind speed are meteorological variables that influence how well local air pollution can disperse in the vertical and horizontal directions respectively, so higher values generally act to lower pollution.[46,47] One study conducted in the Beijing region has indicated that northeasterly winds carry pollution from the Beijing city center and other polluted regions up mountain, increasing predicted $PM_{2.5}$ concentrations. Southwesterly winds bring in air from cleaner regions which is associated with lower predicted $PM_{2.5}$ levels.[48] The impacts of wind direction on $PM_{2.5}$ are consistent with our findings (Figure 6E) where higher predicted $PM_{2.5}$ concentrations occurred when the air mass originated from the southwest (approximately 0-200 degrees) and lower concentrations occurred when air masses originated from the northeast (approximately 200-300 degrees).

The effects of temperature and relative humidity (Figure 6A) on ambient $PM_{2.5}$ are less generalizable and can vary based on season or the physical and chemical properties of air pollutants present.[126] Some studies have found negative associations between temperature and $PM_{2.5}$ concentrations, citing more dispersion from increased convection as a potential mechanism.[127] A positive association between temperature and $PM_{2.5}$, as was observed in our study (Figure 6), may be indicative of increased particle formation, or temperature inversions, which limit or prevent the vertical dispersion of the pollution. Associations between $PM_{2.5}$ and relative humidity have also been found to vary in proportionality and linearity, with some studies citing negative relationships due to increased particle growth and deposition.[44,128] Our study observed a positive association between increased relative humidity and predicted $PM_{2.5}$. Similar positive relationships between relative humidity and $PM_{2.5}$ have been attributed to increased gas-particle phase partitioning of the organic fraction of coal combustion aerosol as well as increased sulfate formation at higher relative humidity levels.[45] Since coal combustion is a large source of air pollution emissions in our study setting, the increased rate of aerosol chemistry under conditions of increasing relative humidity is the most likely explanation for the relationship we observed between $PM_{2.5}$ and relative humidity in this study.

The utility of the variable importance and partial dependence plots is three-fold. First, the plots contribute to the goal of understanding the individual impacts of meteorology on $PM_{2.5}$ in our study by contextualizing our findings from the overall models that show high predictability for meteorology in both heating seasons. In addition to understanding the predictability of meteorology in our study setting, these plots provide insight into what variables are driving prediction, which can directly inform what meteorological variables to incorporate into air pollution accountability studies. Because accountability studies aim to understand how a given policy is impacting an outcome in addition to determining if it was effective or not, it is vital that the confounding effect of factors such as meteorology be isolated from the factors of concern. Efforts to more robustly account for meteorology in air pollution intervention studies have only recently intensified, and this study makes a meaningful contribution to that emerging literature. Second, the variable importance plots inform what variables should be included in the mediation analysis and the partial dependence plots show how they should be modeled (objective 2). The variable importance of greater than 250% increase in MSE for our variables indicates they are all important predictors of $PM_{2.5}$ and should be considered for inclusion in a mediation analysis. Future studies may consider developing a threshold % increase in MSE value of variable importance in the case that it is not immediately apparent which variables are important or not. Additionally, the partial dependence plots show that, in a linear model, boundary layer height and wind speed could be transformed logarithmically whereas relative humidity, temperature, and wind direction could be modeled using splines with various degrees of freedom. Finally, comparing the partial dependence plots to air pollution literature in our study region supports our hypothesis that the random forest model can accurately model the physical relationships between ambient $PM_{2.5}$ and the meteorological variables we included. In future studies, partial dependence plots could be used as measures of model quality assurance and control. For example, if the partial dependence plots in future studies do not align with our current understanding of the underlying physical processes, additional investigation should be undertaken to understand whether the random forest model were producing spurious results.

5.5. Variable interactions

The Friedman's H-statistic for cumulative interaction strength (percent of main effect of a single variable explained by interactions of that variable with other variables) and select two-way interactions are shown in Figure 7 and Figure 8 respectively. Values of the H-statistic can range from 0 to 1, with a value of 0 indicating that none of the main effect of a single variable is explained by interactions of that variable with other variables in the model, while a value of 1 would indicate that all of the main effect of the variable is explained by interactions of that variable with all other variables in the model. The interaction strength for all variables included in the models except district in season two was greater than 0.15, which indicates there are interactions between spatial (village and district), temporal (day of year) and meteorological variables. Relative humidity, wind speed, and boundary layer height were three variables selected to evaluate two-way interactions (Figure 8) because of their relative importance in the model, as well as their known interactions with other variables. Figure 8A shows a relatively strong interaction between relative humidity and boundary layer height in season two (H-statistic = 0.21). Other relatively strong interactions include wind speed and day of year in season two (H-statistic = 0.18) as well as wind speed and boundary layer height in season one (H-statistic = 0.17).

The H-statistic as a measure of interaction strength is useful for understanding what interactions, if any, are contributing to the impacts of meteorology on $PM_{2.5}$ in our study setting. Our findings show that 20-30% of the main effect of the village variable, relative humidity, boundary layer height, and wind speed in both the season one and two models were due to interactions. As previously mentioned, causal claims about the impact and effectiveness of a policy requires a robust adjustment for confounding variables. A mediation analyses for this study that adjust for the impacts of meteorology on $PM_{2.5}$ and did not include any interaction terms between meteorological variables would be failing to capture a part of the impact of meteorology on $PM_{2.5}$, which could result in inaccurate estimations or conclusions of a treatment effect. The two-way interaction strength gives insight into what these specific interactions between two variables are, which could potentially be included as interaction terms in linear models. Specifically, our results show that we should consider interaction terms among boundary layer height, wind speed, and relative humidity. The method we used to evaluate interaction strength can be further applied to three-way, four-way, or higher order interactions that could be considered for including in linear models with sufficient physical justification.

The two-way partial dependence plots contribute to our understanding on the joint impacts of meteorological variables on $PM_{2.5}$ (objective 1). Figure 9 shows the two-way partial dependence plots for relative humidity-boundary layer height, wind speed-wind direction, and wind speed-boundary layer height from the season two model. The combination of decreased vertical dispersion at lower boundary layer height, and increased particle formation at higher relative humidity could explain the relatively high interaction strength between these two variables (Figure 9A). As previously mentioned, difference in wind direction had an impact on predicted $PM_{2.5}$ concentrations in our study region (Figure 6E). Figure 9B shows how the wind direction was most heavily modulated by windspeed at values about ~ 2 m/s in season two. The interaction strength between wind speed and boundary layer height in season two was low (H-statistic = 0.064), so there are fewer extreme effects of wind speed on boundary layer height relative to relative humidity-boundary layer height and wind speed-wind direction interactions (H-statistic > 0.10).

The trends observed in the two-way partial dependence plots depicting the relationship between predicted daily $PM_{2.5}$ concentrations and meteorological variables (Figure 9) are consistent with the individual trends in the single variable partial dependence plots. The plots are additionally useful for understanding how different combinations of variables may contribute to conditions favorable, or unfavorable, to ambient $PM_{2.5}$ concentrations, which directly addresses our third objective by providing evidence for how meteorological variables could be represented in a mediation analysis. For example, previous work has focused on developing meteorological composite indices that can be used to describe the general state of the atmospheric conditions at a given time.[114,129] These methods involve the construction of similar bi-variate plots, though not from machine learning models, to develop cutoff points to classify the atmosphere by sets of conditions that are related to $PM_{2.5}$ concentrations. The two-way partial dependence plots (and their three-way extension) could be used to develop similar indices on a study-by-study basis. For example, we could develop an index with four levels for this study that captures meteorological effects related to dispersion based on the most predominant colors in Figure 9B and 9C (yellow, light blue/green, blue, and dark purple). These indices could then be used in health or economic models to adjust for the dispersive effects of meteorology on $PM_{2.5}$ instead of adjusting for wind speed, wind direction, and boundary layer height as individual variables.

This study was the first to apply random forest models to $PM_{2.5}$ measurements collected in the field as part of an observational study. Features of the study setting that make it unique and interesting from an air pollution standpoint are the use of solid fuel for household heating, proximity to a large city center, and the mountainous region where the villages were located. Given that household solid fuel combustion often occurs in rural regions with complex geography and is a frequent target of policy intervention, it is valuable to verify that the tools used to evaluate these interventions provide useful information in theses settings. While some studies have evaluated outdoor $PM_{2.5}$ in the rural, mountainous region of Beijing, they tend to be limited to only a few villages or rural sampling stations per study.[48,130,131] Our study, which includes sampling from a ~50 villages across four administrative districts at a high temporal and spatial resolution, provides additional data about this region. Further, the non-parametric nature of the random forest method allowed us to model the relationships between meteorology variables and PM without assuming degree or nature of the relationship.

There are several limitations to our study. First, we are using a predictive model for descriptive purposes. The random forest method, like other machine learning approaches, optimizes for out of sample fit instead of in sample fit like descriptive models. However, as previously stated, an advantage of the random forest approach over other machine learning methods is that it has outputs (i.e., variable importance and partial dependence) that have similar function to descriptive models (i.e., regression coefficients). The meteorology data was gridded at ~30 km^2 resolution, so villages in the same district tended to have values interpolated from the same 4 grid points. Our interpolations also did not take terrain into consideration. Windspeed and boundary layer height decrease in areas with more obstacles (i.e., buildings, mountains), so our estimation of these values may be different than expected in some villages. The accuracy of the interpolated meteorological values could be improved by including land use variables for terrain types in the model, using terrain information to adjust the meteorological variables at the interpolation stage, or developing correction factors by placing low-cost weather stations and comparing the measured to those interpolated ones. While these limitations likely weakened the predictability of our models, it is unlikely that they would have changed the fundamental physical relationship between meteorological variables and $PM_{2.5}$ concentrations. For example, it is unlikely that addressing all the limitations of our study would lead to increased wind speed being associated with an increase in predicted $PM_{2.5}$ concentrations. However, more accurate meteorological values may change the magnitude of the relationship with $PM_{2.5}$, similar to how an effect modification would change the magnitude of a coefficient in a regression model. This potential limitation does not greatly impact the outcome of this study, given that two primary objectives were to identify which meteorological variables should be included in future mediation analysis and to make recommendations on how to model meteorology in our study setting, rather than to specify the magnitude of those relationships.

6. CONCLUSIONS

Our findings demonstrated that random forest models can be applied in settings where observational studies are underway to gain insight into how meteorology variables influence local daily $PM_{2.5}$ concentrations. By building separate models for each field season, we were able to quantify the impacts of meteorology on ambient $PM_{2.5}$ and identify differences in the predictability of meteorological variables between the two seasons. The variable importance plots showed that all the meteorological variables included in our analysis (temperature, relative humidity, boundary layer height, wind speed, wind direction) were important and should be controlled for in a mediation analysis. The partial dependence plots and interaction strength further inform how these variables should be treated if included as individual terms in a linear model. Here, we found that boundary layer height and wind speed should be transformed logarithmically whereas relative humidity, temperature, and wind direction should be modeled using splines with various degrees of freedom. We also found that interaction terms among boundary layer height, relative humidity, and wind speed should be considered in a linear model.

The two-way partial dependence plots showed potential for the meteorological impacts on $PM_{2.5}$ to be represented in composite meteorological indices. Meteorological variables (i.e., wind speed, boundary layer height) with similar impacts on meteorology (i.e., dispersion) could be grouped together and ranges of common values computed for similar $PM_{2.5}$ concentrations. From our plots (Figure 9), we could develop an index with four categories based on the most prominent visible colors. A composite meteorological index may be appealing to health researchers because it simplifies many meteorological terms into a few index variables. However, transforming many variables into an index with a finite number of levels causes some loss of information because levels are often not perfectly discrete. Appropriate caution and incorporation of uncertainty surrounding the index levels should be taken when using a composite meteorological index for mediation analysis.

A third option to translate our findings about meteorological impacts on $PM_{2.5}$ concentrations is to develop meteorologically-normalized $PM_{2.5}$ concenrations.[132,133] The most applied approach involves randomly sampling values for meteorological variables for each observation, using the developed model to predict a $PM_{2.5}$ concentration, and averaging predicted values over many runs (~1000 runs). The advantage of this approach is that it requires very little additional work after the initial random forest model has been developed and provides an output that is ready to be used for inference (meteorologically-normalized $PM_{2.5}$ concentrations for each observation in the model). The work presented in this thesis provides the random forest models for the first two seasons of the overall accountability study. The same random forest modeling approach could be applied to the remaining two seasons of data collection and subsequently used to develop meteorologically-normalized $PM_{2.5}$ concentrations for each season.

It is unclear at this time how adjusting for meteorology in each of the three ways previously mentioned (terms for each variable, composite meteorological index, meteorological normalization) would impact the results of a mediation analysis. Future work for this project includes developing a composite meteorological index and the meteorologically-normalized values and evaluating how they influence the mediation analysis. Other related work on methods to translate modeled information about the metrological impacts on $PM_{2.5}$ into ready-to-use information for epidemiologists, social scientists, and other non-physical scientists is needed.

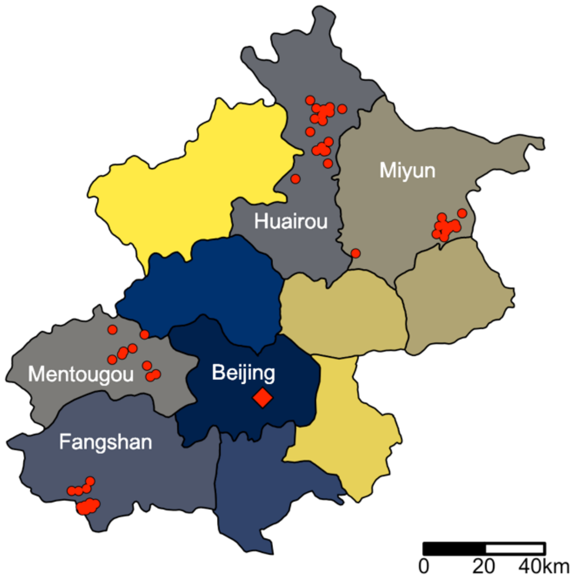

Figure 1. Map of study area. Villages are red circles and administrative districts are labeled.

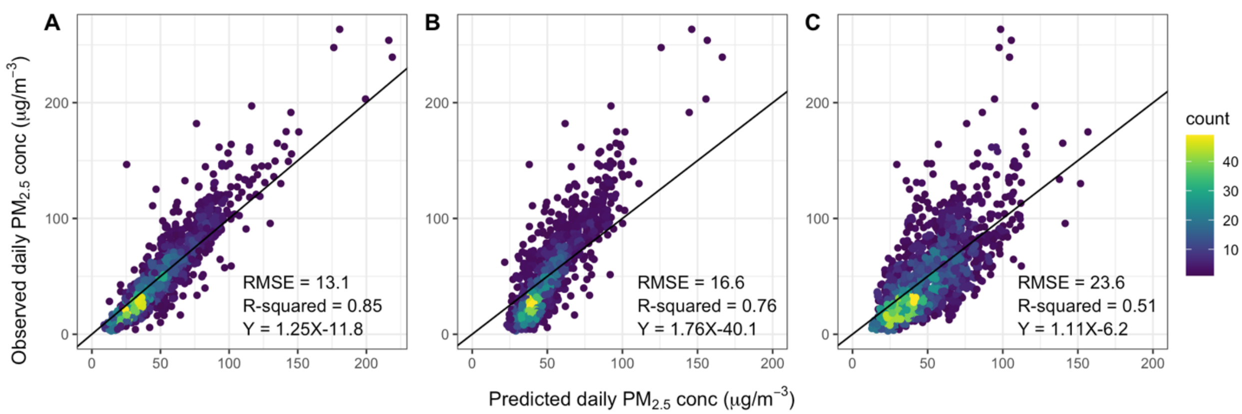

Figure 2. Observed vs predicted daily $PM_{2.5}$ concentrations from the season one models. A: overall (random folds); B: spatial (village split equally among folds); C: temporal (day of year split equally among folds) 10-fold cross validation models.

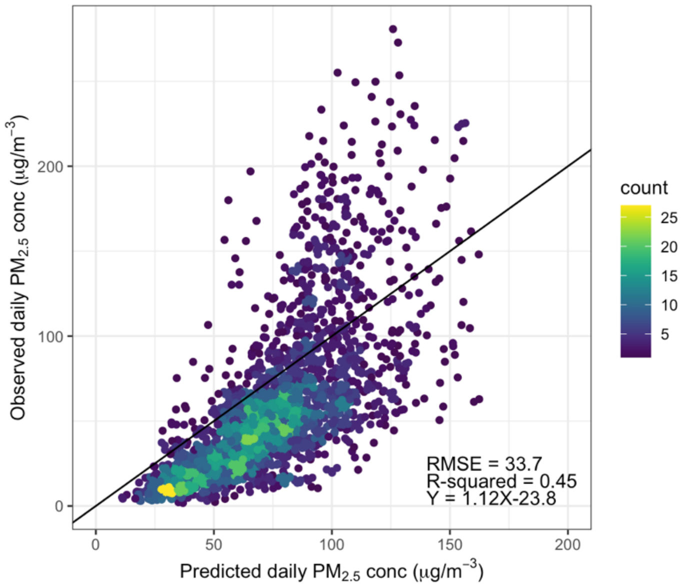

Figure 3. Observed daily $PM_{2.5}$ concentrations for season two random forest model vs daily PM2.5 concentrations predicted by the season one random forest model.

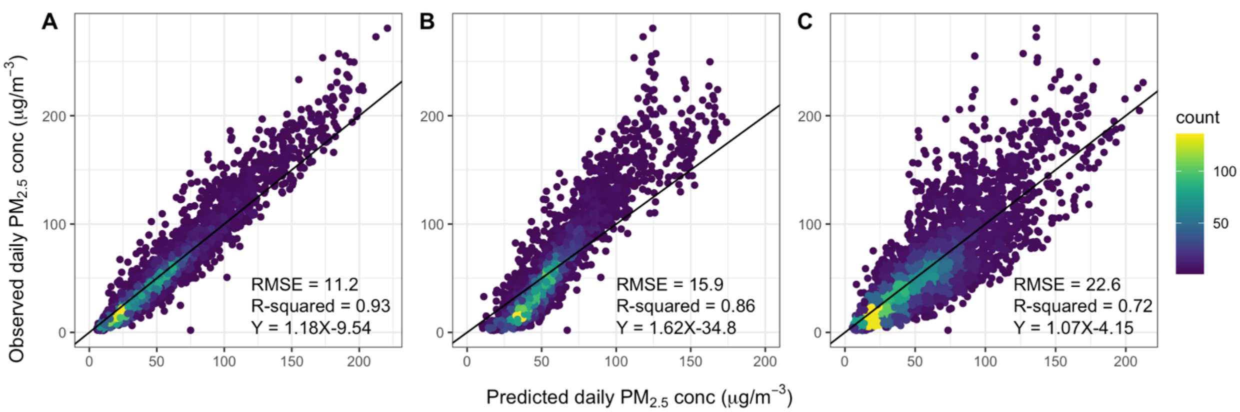

Figure 4. Observed vs predicted daily $PM_{2.5}$ concentrations from the season two models. A: overall (random folds); B: spatial (village split equally among folds); C: temporal (day of year split equally among folds) 10-fold cross validation models.

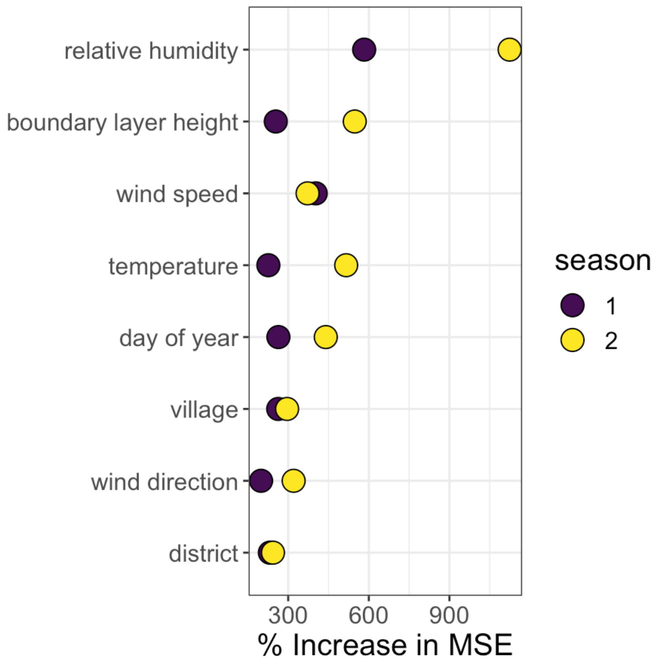

Figure 5. Variable importance given as percent increase in the mean squared error (MSE) of the model when the variable was randomly permuted compared to the base model. Larger value indicates higher variable importance.

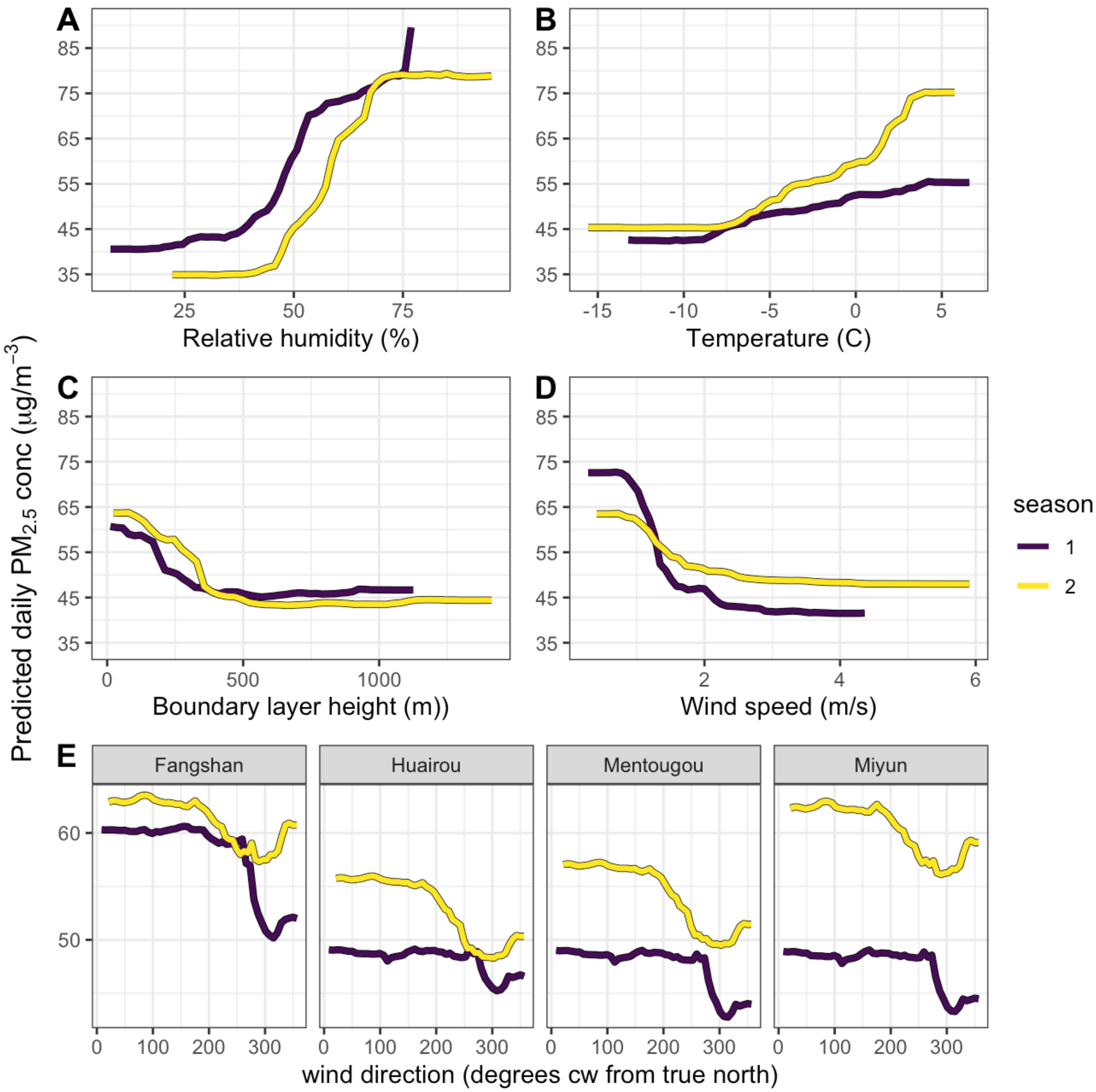

Figure 6. Partial dependence plot for the season one and season two models predicting daily PM2.5 concentrations for a given variable holding the value of all other variables constant.

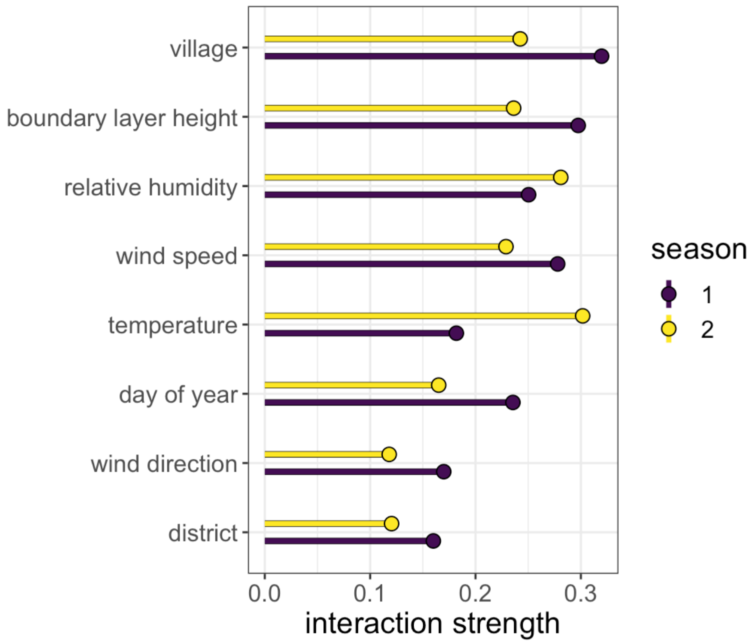

Figure 7. Interaction strength (Friedman's H-statistic) for each variable with all other variables the season one and two random forest models predicting daily PM2.5 concentrations.

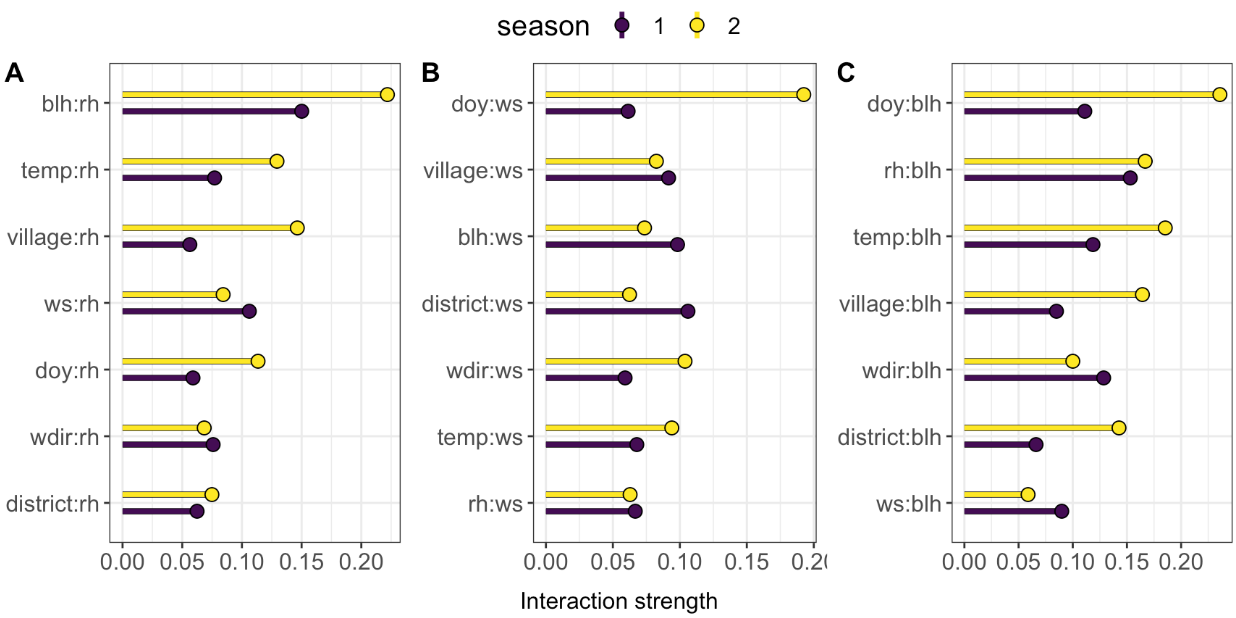

Figure 8. Two-way interaction strengths (Friedman's H-statistic) for (A) relative humidity, (B) wind speed, and (C) boundary layer height from the season one and two random forest models predicting daily PM2.5 concentrations.

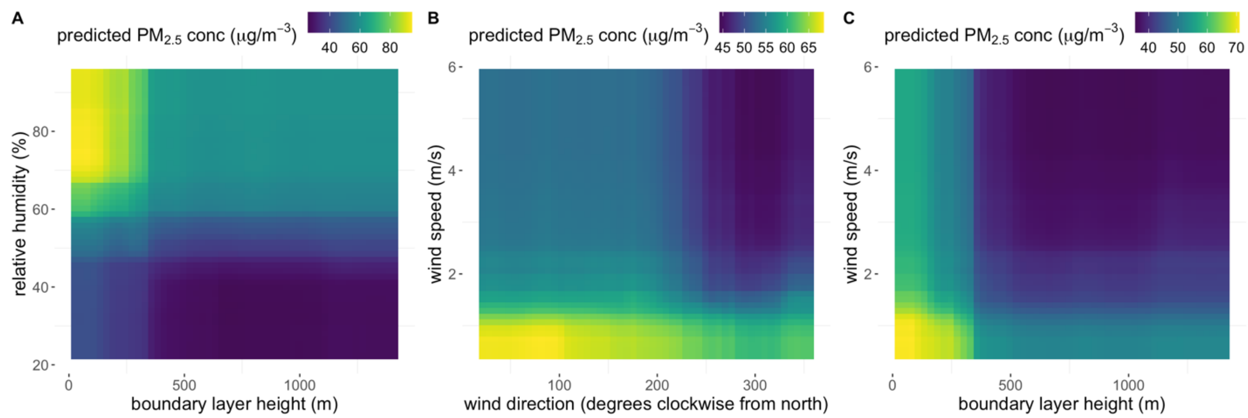

Figure 9. Two-way partial dependence plots depicting the relationship between predicted daily PM2.5 concentrations and (A) relative humidity-boundary layer height, (B) wind speed-wind direction, and (C) boundary layer height-wind speed from the season two random forest model predicting daily PM2.5 concentrations.

REFERENCES

(1) Landrigan, P. J.; Fuller, R.; Acosta, N. J. R.; Adeyi, O.; Arnold, R.; Basu, N. (Nil); Baldé, A. B.; Bertollini, R.; Bose-O'Reilly, S.; Boufford, J. I.; Breysse, P. N.; Chiles, T.; Mahidol, C.; Coll-Seck, A. M.; Cropper, M. L.; Fobil, J.; Fuster, V.; Greenstone, M.; Haines, A.; Hanrahan, D.; Hunter, D.; Khare, M.; Krupnick, A.; Lanphear, B.; Lohani, B.; Martin, K.; Mathiasen, K. V.; McTeer, M. A.; Murray, C. J. L.; Ndahimananjara, J. D.; Perera, F.; Potočnik, J.; Preker, A. S.; Ramesh, J.; Rockström, J.; Salinas, C.; Samson, L. D.; Sandilya, K.; Sly, P. D.; Smith, K. R.; Steiner, A.; Stewart, R. B.; Suk, W. A.; Schayck, O. C. P. van; Yadama, G. N.; Yumkella, K.; Zhong, M. The Lancet Commission on Pollution and Health. The Lancet 2018, 391 (10119), 462–512. https://doi.org/10.1016/S0140-6736(17)32345-0.

(2) Cohen, A. J.; Brauer, M.; Burnett, R.; Anderson, H. R.; Frostad, J.; Estep, K.; Balakrishnan, K.; Brunekreef, B.; Dandona, L.; Dandona, R.; Feigin, V.; Freedman, G.; Hubbell, B.; Jobling, A.; Kan, H.; Knibbs, L.; Liu, Y.; Martin, R.; Morawska, L.; Pope, C. A.; Shin, H.; Straif, K.; Shaddick, G.; Thomas, M.; van Dingenen, R.; van Donkelaar, A.; Vos, T.; Murray, C. J. L.; Forouzanfar, M. H. Estimates and 25-Year Trends of the Global Burden of Disease Attributable to Ambient Air Pollution: An Analysis of Data from the Global Burden of Diseases Study 2015. The Lancet 2017, 389 (10082), 1907–1918. https://doi.org/10.1016/S0140-6736(17)30505-6.

(3) Davidson, C. I.; Phalen, R. F.; Solomon, P. A. Airborne Particulate Matter and Human Health: A Review. Aerosol Science and Technology 2005, 39 (8), 737–749. https://doi.org/10.1080/02786820500191348.

(4) Baumgartner, J.; Zhang, Y.; Schauer, J. J.; Ezzati, M.; Patz, J. A.; Bautista, L. E. Household Air Pollution and Children's Blood Pressure. Epidemiology 2012, 23 (4), 641–642. https://doi.org/10.1097/EDE.0b013e3182593fa9.

(5) Zeng, X.; Xu, X.; Zheng, X.; Reponen, T.; Chen, A.; Huo, X. Heavy Metals in PM2.5 and in Blood, and Children's Respiratory Symptoms and Asthma from an e-Waste Recycling Area. Environmental Pollution 2016, 210, 346–353. https://doi.org/10.1016/j.envpol.2016.01.025.

(6) Yang, A.; Janssen, N. A.; Brunekreef, B.; Cassee, F. R.; Hoek, G.; Gehring, U. Children's Respiratory Health and Oxidative Potential of PM2. 5: The PIAMA Birth Cohort Study. Occup Environ Med 2016, 73 (3), 154–160.

(7) Kim, K.-H.; Kabir, E.; Kabir, S. A Review on the Human Health Impact of Airborne Particulate Matter. Environment International 2015, 74, 136–143. https://doi.org/10.1016/j.envint.2014.10.005.

(8) Mazidi, M.; Speakman, J. R. Ambient Particulate Air Pollution (PM2.5) Is Associated with the Ratio of Type 2 Diabetes to Obesity. Sci Rep 2017, 7 (1), 1–8. https://doi.org/10.1038/s41598-017-08287-1.

(9) Bowe, B.; Xie, Y.; Li, T.; Yan, Y.; Xian, H.; Al-Aly, Z. The 2016 Global and National Burden of Diabetes Mellitus Attributable to PM2·5 Air Pollution. The Lancet Planetary Health 2018, 2 (7), e301–e312. https://doi.org/10.1016/S2542-5196(18)30140-2.

(10) Meo, S. A.; Memon, A. N.; Sheikh, S. A.; Rouq, F. A.; Usmani, A. M.; Hassan, A.; Arian, S. A. Effect of Environmental Air Pollution on Type 2 Diabetes Mellitus. Eur Rev Med Pharmacol Sci 2015, 19 (1), 123–128.
(11) Li, H.; Zhou, D.; Zhang, Q.; Feng, C.; Zheng, W.; He, K.; Lan, Y. Vanadium Exposure-Induced Neurobehavioral Alterations among Chinese Workers. NeuroToxicology 2013, 36, 49–54. https://doi.org/10.1016/j.neuro.2013.02.008.
(12) Jacobson, M. Z. Strong Radiative Heating Due to the Mixing State of Black Carbon in Atmospheric Aerosols. Nature 2001, 409 (6821), 695–697. https://doi.org/10.1038/35055518.
(13) Bond, T. C.; Doherty, S. J.; Fahey, D. W.; Forster, P. M.; Berntsen, T.; DeAngelo, B. J.; Flanner, M. G.; Ghan, S.; Kärcher, B.; Koch, D.; Kinne, S.; Kondo, Y.; Quinn, P. K.; Sarofim, M. C.; Schultz, M. G.; Schulz, M.; Venkataraman, C.; Zhang, H.; Zhang, S.; Bellouin, N.; Guttikunda, S. K.; Hopke, P. K.; Jacobson, M. Z.; Kaiser, J. W.; Klimont, Z.; Lohmann, U.; Schwarz, J. P.; Shindell, D.; Storelvmo, T.; Warren, S. G.; Zender, C. S. Bounding the Role of Black Carbon in the Climate System: A Scientific Assessment. Journal of Geophysical Research: Atmospheres 2013, 118 (11), 5380–5552. https://doi.org/10.1002/jgrd.50171.
(14) Burns, J.; Boogaard, H.; Polus, S.; Pfadenhauer, L. M.; Rohwer, A. C.; van Erp, A. M.; Turley, R.; Rehfuess, E. A. Interventions to Reduce Ambient Air Pollution and Their Effects on Health: An Abridged Cochrane Systematic Review. Environment International 2020, 135, 105400. https://doi.org/10.1016/j.envint.2019.105400.
(15) Bonjour, S.; Adair-Rohani, H.; Wolf, J.; Bruce, N. G.; Mehta, S.; Prüss-Ustün, A.; Lahiff, M.; Rehfuess, E. A.; Mishra, V.; Smith, K. R. Solid Fuel Use for Household Cooking: Country and Regional Estimates for 1980–2010. Environ Health Perspect 2013, 121 (7), 784–790. https://doi.org/10.1289/ehp.1205987.
(16) Chafe, Z. A.; Brauer, M.; Klimont, Z.; Van, D. R.; Mehta, S.; Rao, S.; Riahi, K.; Dentener, F.; Smith, K. R. Household Cooking with Solid Fuels Contributes to Ambient PM2.5 Air Pollution and the Burden of Disease. Environmental Health Perspectives 2014, 122 (12), 1314–1320. https://doi.org/10.1289/ehp.1206340.
(17) Liao, J.; Zimmermann Jin, A.; Chafe, Z. A.; Pillarisetti, A.; Yu, T.; Shan, M.; Yang, X.; Li, H.; Liu, G.; Smith, K. R. The Impact of Household Cooking and Heating with Solid Fuels on Ambient PM2.5 in Peri-Urban Beijing. Atmospheric Environment 2017, 165, 62–72. https://doi.org/10.1016/j.atmosenv.2017.05.053.
(18) Tao, S.; Ru, M. Y.; Du, W.; Zhu, X.; Zhong, Q. R.; Li, B. G.; Shen, G. F.; Pan, X. L.; Meng, W. J.; Chen, Y. L. Quantifying the Rural Residential Energy Transition in China from 1992 to 2012 through a Representative National Survey. Nature Energy 2018, 3 (7), 567–573.
(19) Barrington-Leigh, C.; Baumgartner, J.; Carter, E.; Robinson, B. E.; Tao, S.; Zhang, Y. An Evaluation of Air Quality, Home Heating and Well-Being under Beijing's Programme to Eliminate Household Coal Use. Nat Energy 2019, 4 (5), 416–423. https://doi.org/10.1038/s41560-019-0386-2.
(20) Baumgartner, J.; Clark, S.; Carter, E.; Lai, A.; Zhang, Y.; Shan, M.; Schauer, J. J.; Yang, X. Effectiveness of a Household Energy Package in Improving Indoor Air Quality and Reducing Personal Exposures in Rural China. Environ. Sci. Technol. 2019, 53 (15), 9306–9316. https://doi.org/10.1021/acs.est.9b02061.

(21) Institute, H. E. Proceedings of an HEI Workshop on Further Research to Assess the Health Impacts of Actions Taken to Improve Air Quality; Health Effects Institute Boston^ eMA MA, 2010.
(22) Jeong, J. I.; Park, R. J. Winter Monsoon Variability and Its Impact on Aerosol Concentrations in East Asia. Environmental Pollution 2017, 221, 285–292. https://doi.org/10.1016/j.envpol.2016.11.075.
(23) Barmpadimos, I.; Hueglin, C.; Keller, J.; Henne, S.; Prévôt, A. S. H. Influence of Meteorology on PM_{10} Trends and Variability in Switzerland from 1991 to 2008. Atmospheric Chemistry and Physics 2011, 11 (4), 1813–1835. https://doi.org/10.5194/acp-11-1813-2011.
(24) Henneman, L. R. F.; Liu, C.; Mulholland, J. A.; Russell, A. G. Evaluating the Effectiveness of Air Quality Regulations: A Review of Accountability Studies and Frameworks. Journal of the Air & Waste Management Association 2017, 67 (2), 144–172. https://doi.org/10.1080/10962247.2016.1242518.
(25) Henneman, L. R. F.; Liu, C.; Hu, Y.; Mulholland, J. A.; Russell, A. G. Air Quality Modeling for Accountability Research: Operational, Dynamic, and Diagnostic Evaluation. Atmospheric Environment 2017, 166, 551–565. https://doi.org/10.1016/j.atmosenv.2017.07.049.
(26) Qiu, M.; Zigler, C. M.; Selin, N. E. Statistical and Machine Learning Methods for Evaluating Trends in Air Quality under Changing Meteorological Conditions. Atmospheric Chemistry and Physics Discussions 2022, 1–45. https://doi.org/10.5194/acp-2022-232.
(27) Squizzato, S.; Masiol, M.; Rich, D. Q.; Hopke, P. K. A Long-Term Source Apportionment of PM2.5 in New York State during 2005–2016. Atmospheric Environment 2018, 192, 35–47. https://doi.org/10.1016/j.atmosenv.2018.08.044.
(28) Goyal, P.; Gulia, S.; Goyal, S. K. Review of Land Use Specific Source Contributions in PM2.5 Concentration in Urban Areas in India. Air Qual Atmos Health 2021, 14 (5), 691–704. https://doi.org/10.1007/s11869-020-00972-x.
(29) Zhu, Y.; Huang, L.; Li, J.; Ying, Q.; Zhang, H.; Liu, X.; Liao, H.; Li, N.; Liu, Z.; Mao, Y.; Fang, H.; Hu, J. Sources of Particulate Matter in China: Insights from Source Apportionment Studies Published in 1987–2017. Environment International 2018, 115, 343–357. https://doi.org/10.1016/j.envint.2018.03.037.
(30) Hasheminassab, S.; Daher, N.; Shafer, M. M.; Schauer, J. J.; Delfino, R. J.; Sioutas, C. Chemical Characterization and Source Apportionment of Indoor and Outdoor Fine Particulate Matter (PM2.5) in Retirement Communities of the Los Angeles Basin. Science of The Total Environment 2014, 490, 528–537. https://doi.org/10.1016/j.scitotenv.2014.05.044.
(31) Larson, T.; Gould, T.; Simpson, C.; Liu, L.-J. S.; Claiborn, C.; Lewtas, J. Source Apportionment of Indoor, Outdoor, and Personal PM2.5 in Seattle, Washington, Using Positive Matrix Factorization. Journal of the Air & Waste Management Association 2004, 54 (9), 1175–1187. https://doi.org/10.1080/10473289.2004.10470976.
(32) Brehmer, C.; Norris, C.; Barkjohn, K. K.; Bergin, M. H.; Zhang, J.; Cui, X.; Zhang, Y.; Black, M.; Li, Z.; Shafer, M.; Schauer, J. J. The Impact of Household Air Cleaners on the Chemical Composition and Children's Exposure to PM2.5 Metal Sources in Suburban Shanghai. Environmental Pollution 2019, 253, 190–198. https://doi.org/10.1016/j.envpol.2019.07.003.

(33) Yu, L. Characterization and Source Apportionment of PM2.5 in an Urban Environment in Beijing. Aerosol and Air Quality Research 2013. https://doi.org/10.4209/aaqr.2012.07.0192.
(34) Chen, X.-C.; Jahn, H. J.; Engling, G.; Ward, T. J.; Kraemer, A.; Ho, K.-F.; Yim, S. H. L.; Chan, C.-Y. Chemical Characterization and Sources of Personal Exposure to Fine Particulate Matter (PM2.5) in the Megacity of Guangzhou, China. Environmental Pollution 2017, 231, 871–881. https://doi.org/10.1016/j.envpol.2017.08.062.
(35) Liu, B.; Song, N.; Dai, Q.; Mei, R.; Sui, B.; Bi, X.; Feng, Y. Chemical Composition and Source Apportionment of Ambient PM2.5 during the Non-Heating Period in Taian, China. Atmospheric Research 2016, 170, 23–33. https://doi.org/10.1016/j.atmosres.2015.11.002.
(36) Liu, W.; Xu, Y.; Liu, W.; Liu, Q.; Yu, S.; Liu, Y.; Wang, X.; Tao, S. Oxidative Potential of Ambient PM2.5 in the Coastal Cities of the Bohai Sea, Northern China: Seasonal Variation and Source Apportionment. Environmental Pollution 2018, 236, 514–528. https://doi.org/10.1016/j.envpol.2018.01.116.
(37) Shang, J.; Khuzestani, R. B.; Tian, J.; Schauer, J. J.; Hua, J.; Zhang, Y.; Cai, T.; Fang, D.; An, J.; Zhang, Y. Chemical Characterization and Source Apportionment of PM2.5 Personal Exposure of Two Cohorts Living in Urban and Suburban Beijing. Environmental Pollution 2019, 246, 225–236. https://doi.org/10.1016/j.envpol.2018.11.076.
(38) Chuang, M.-T.; Chen, Y.-C.; Lee, C.-T.; Cheng, C.-H.; Tsai, Y.-J.; Chang, S.-Y.; Su, Z.-S. Apportionment of the Sources of High Fine Particulate Matter Concentration Events in a Developing Aerotropolis in Taoyuan, Taiwan. Environmental Pollution 2016, 214, 273–281. https://doi.org/10.1016/j.envpol.2016.04.045.
(39) Lai, A. M.; Carter, E.; Shan, M.; Ni, K.; Clark, S.; Ezzati, M.; Wiedinmyer, C.; Yang, X.; Baumgartner, J.; Schauer, J. J. Chemical Composition and Source Apportionment of Ambient, Household, and Personal Exposures to PM2.5 in Communities Using Biomass Stoves in Rural China. Science of The Total Environment 2019, 646, 309–319. https://doi.org/10.1016/j.scitotenv.2018.07.322.
(40) Secrest, M. H.; Schauer, J. J.; Carter, E. M.; Lai, A. M.; Wang, Y.; Shan, M.; Yang, X.; Zhang, Y.; Baumgartner, J. The Oxidative Potential of PM 2.5 Exposures from Indoor and Outdoor Sources in Rural China. Science of The Total Environment 2016, 571, 1477–1489. https://doi.org/10.1016/j.scitotenv.2016.06.231.
(41) van Erp, A. M.; Kelly, F. J.; Demerjian, K. L.; Pope, C. A.; Cohen, A. J. Progress in Research to Assess the Effectiveness of Air Quality Interventions towards Improving Public Health. Air Qual Atmos Health 2012, 5 (2), 217–230. https://doi.org/10.1007/s11869-010-0127-y.
(42) Westervelt, D. M.; Horowitz, L. W.; Naik, V.; Tai, A. P. K.; Fiore, A. M.; Mauzerall, D. L. Quantifying PM2.5-Meteorology Sensitivities in a Global Climate Model. Atmospheric Environment 2016, 142, 43–56. https://doi.org/10.1016/j.atmosenv.2016.07.040.
(43) Xie, Y.; Liu, Z.; Wen, T.; Huang, X.; Liu, J.; Tang, G.; Yang, Y.; Li, X.; Shen, R.; Hu, B.; Wang, Y. Characteristics of Chemical Composition and Seasonal Variations of PM2.5 in Shijiazhuang, China: Impact of Primary Emissions and Secondary Formation. Science of The Total Environment 2019, 677, 215–229. https://doi.org/10.1016/j.scitotenv.2019.04.300.
(44) Lou, C.; Liu, H.; Li, Y.; Peng, Y.; Wang, J.; Dai, L. Relationships of Relative Humidity with PM2.5 and PM10 in the Yangtze River Delta, China. Environ Monit Assess 2017, 189 (11), 582. https://doi.org/10.1007/s10661-017-6281-z.

(45) Sun, Y.; Wang, Z.; Fu, P.; Jiang, Q.; Yang, T.; Li, J.; Ge, X. The Impact of Relative Humidity on Aerosol Composition and Evolution Processes during Wintertime in Beijing, China. Atmospheric Environment 2013, 77, 927–934. https://doi.org/10.1016/j.atmosenv.2013.06.019.

(46) Ly, B.-T.; Matsumi, Y.; Vu, T. V.; Sekiguchi, K.; Nguyen, T.-T.; Pham, C.-T.; Nghiem, T.-D.; Ngo, I.-H.; Kurotsuchi, Y.; Nguyen, T.-H.; Nakayama, T. The Effects of Meteorological Conditions and Long-Range Transport on PM2.5 Levels in Hanoi Revealed from Multi-Site Measurement Using Compact Sensors and Machine Learning Approach. Journal of Aerosol Science 2021, 152, 105716. https://doi.org/10.1016/j.jaerosci.2020.105716.

(47) Wang, H. L.; Qiao, L. P.; Lou, S. R.; Zhou, M.; Ding, A. J.; Huang, H. Y.; Chen, J. M.; Wang, Q.; Tao, S. K.; Chen, C. H.; Li, L.; Huang, C. Chemical Composition of PM2.5 and Meteorological Impact among Three Years in Urban Shanghai, China. Journal of Cleaner Production 2016, 112, 1302–1311. https://doi.org/10.1016/j.jclepro.2015.04.099.

(48) Zhao, X.; Zhang, X.; Xu, X.; Xu, J.; Meng, W.; Pu, W. Seasonal and Diurnal Variations of Ambient PM2.5 Concentration in Urban and Rural Environments in Beijing. Atmospheric Environment 2009, 43 (18), 2893–2900. https://doi.org/10.1016/j.atmosenv.2009.03.009.

(49) Lushchak, V. I. Free Radicals, Reactive Oxygen Species, Oxidative Stress and Its Classification. Chemico-Biological Interactions 2014, 224, 164–175. https://doi.org/10.1016/j.cbi.2014.10.016.

(50) Bates, J. T.; Weber, R. J.; Abrams, J.; Verma, V.; Fang, T.; Klein, M.; Strickland, M. J.; Sarnat, S. E.; Chang, H. H.; Mulholland, J. A.; Tolbert, P. E.; Russell, A. G. Reactive Oxygen Species Generation Linked to Sources of Atmospheric Particulate Matter and Cardiorespiratory Effects. Environ. Sci. Technol. 2015, 49 (22), 13605–13612. https://doi.org/10.1021/acs.est.5b02967.

(51) Brehmer, C.; Lai, A. M.; Clark, S.; Shan, M.; Ni, K.; Ezzati, M.; Yang, X.; Baumgartner, J.; Schauer, J. J.; Carter, E. M. The Oxidative Potential of Personal and Household PM2.5 in a Rural Setting in Southwestern China. Environ. Sci. Technol. 2019. https://doi.org/10.1021/acs.est.8b05120.

(52) Al Hanai, A. H.; Antkiewicz, D. S.; Hemming, J. D. C.; Shafer, M. M.; Lai, A. M.; Arhami, M.; Hosseini, V.; Schauer, J. J. Seasonal Variations in the Oxidative Stress and Inflammatory Potential of PM2.5 in Tehran Using an Alveolar Macrophage Model; The Role of Chemical Composition and Sources. Environment International 2019, 123, 417–427. https://doi.org/10.1016/j.envint.2018.12.023.

(53) Saffari, A.; Daher, N.; Shafer, M. M.; Schauer, J. J.; Sioutas, C. Global Perspective on the Oxidative Potential of Airborne Particulate Matter: A Synthesis of Research Findings. Environmental Science & Technology 2014, 48 (13), 7576–7583. https://doi.org/10.1021/es500937x.

(54) Tuet, W. Y.; Fok, S.; Verma, V.; Tagle Rodriguez, M. S.; Grosberg, A.; Champion, J. A.; Ng, N. L. Dose-Dependent Intracellular Reactive Oxygen and Nitrogen Species (ROS/RNS) Production from Particulate Matter Exposure: Comparison to Oxidative Potential and Chemical Composition. Atmospheric Environment 2016, 144, 335–344. https://doi.org/10.1016/j.atmosenv.2016.09.005.

(55) Lai, A.; Baumgartner, J.; J. Schauer, J.; Rudich, Y.; Pardo, M. Cytotoxicity and Chemical Composition of Women's Personal PM 2.5 Exposures from Rural China. Environmental Science: Atmospheres 2021, 1 (6), 359–371. https://doi.org/10.1039/D1EA00022E.

(56) Furman, D.; Campisi, J.; Verdin, E.; Carrera-Bastos, P.; Targ, S.; Franceschi, C.; Ferrucci, L.; Gilroy, D. W.; Fasano, A.; Miller, G. W.; Miller, A. H.; Mantovani, A.; Weyand, C. M.; Barzilai, N.; Goronzy, J. J.; Rando, T. A.; Effros, R. B.; Lucia, A.; Kleinstreuer, N.; Slavich, G. M. Chronic Inflammation in the Etiology of Disease across the Life Span. Nat Med 2019, 25 (12), 1822–1832. https://doi.org/10.1038/s41591-019-0675-0.
(57) The Inflammation Theory of Disease. EMBO reports 2012, 13 (11), 968–970. https://doi.org/10.1038/embor.2012.142.
(58) Kelly, F. J.; Fussell, J. C. Improving Indoor Air Quality, Health and Performance within Environments Where People Live, Travel, Learn and Work. Atmospheric Environment 2019, 200, 90–109. https://doi.org/10.1016/j.atmosenv.2018.11.058.
(59) Bu, X.; Xie, Z.; Liu, J.; Wei, L.; Wang, X.; Chen, M.; Ren, H. Global PM2.5-Attributable Health Burden from 1990 to 2017: Estimates from the Global Burden of Disease Study 2017. Environmental Research 2021, 197, 111123. https://doi.org/10.1016/j.envres.2021.111123.
(60) Gakidou, E.; Afshin, A.; Abajobir, A. A.; Abate, K. H.; Abbafati, C.; Abbas, K. M.; Abd-Allah, F.; Abdulle, A. M.; Abera, S. F.; Aboyans, V. Global, Regional, and National Comparative Risk Assessment of 84 Behavioural, Environmental and Occupational, and Metabolic Risks or Clusters of Risks, 1990–2016: A Systematic Analysis for the Global Burden of Disease Study 2016. The Lancet 2017, 390 (10100), 1345–1422.
(61) Group, H. A. W. Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research. HEI communication 2003, 11.
(62) Zigler, C. M.; Dominici, F. Point: Clarifying Policy Evidence With Potential-Outcomes Thinking—Beyond Exposure-Response Estimation in Air Pollution Epidemiology. American Journal of Epidemiology 2014, 180 (12), 1133–1140. https://doi.org/10.1093/aje/kwu263.
(63) Rich, D. Q. Accountability Studies of Air Pollution and Health Effects: Lessons Learned and Recommendations for Future Natural Experiment Opportunities. Environment International 2017, 100, 62–78. https://doi.org/10.1016/j.envint.2016.12.019.
(64) Fisk, W. J. Health Benefits of Particle Filtration. Indoor Air 2013, 23 (5), 357–368. https://doi.org/10.1111/ina.12036.
(65) Friedman, M. S.; Powell, K. E.; Hutwagner, L.; Graham, L. M.; Teague, W. G. Impact of Changes in Transportation and Commuting Behaviors during the 1996 Summer Olympic Games in Atlanta on Air Quality and Childhood Asthma. Jama 2001, 285 (7), 897–905.
(66) Peel, J. L.; Klein, M.; Flanders, W. D.; Mulholland, J. A.; Tolbert, P. E.; Committee, H. H. R. Impact of Improved Air Quality during the 1996 Summer Olympic Games in Atlanta on Multiple Cardiovascular and Respiratory Outcomes. 2010.
(67) Wang, M.; Zhu, T.; Zheng, J.; Zhang, R. Y.; Zhang, S. Q.; Xie, X. X.; Han, Y. Q.; Li, Y. Use of a Mobile Laboratory to Evaluate Changes in On-Road Air Pollutants during the Beijing 2008 Summer Olympics. Atmospheric Chemistry and Physics 2009, 9 (21), 8247–8263.
(68) Wang, X.; Westerdahl, D.; Chen, L. C.; Wu, Y.; Hao, J.; Pan, X.; Guo, X.; Zhang, K. M. Evaluating the Air Quality Impacts of the 2008 Beijing Olympic Games: On-Road Emission Factors and Black Carbon Profiles. Atmospheric Environment 2009, 43 (30), 4535–4543.
(69) Wang, T.; Nie, W.; Gao, J.; Xue, L. K.; Gao, X. M.; Wang, X.; Qiu, J.; Poon, C. N.; Meinardi, S.; Blake, D. Air Quality during the 2008 Beijing Olympics: Secondary

Pollutants and Regional Impact. Atmospheric Chemistry and Physics 2010, 10 (16), 7603–7615.
(70) Li, Y.; Wang, W.; Wang, J.; Zhang, X.; Lin, W.; Yang, Y. Impact of Air Pollution Control Measures and Weather Conditions on Asthma during the 2008 Summer Olympic Games in Beijing. Int J Biometeorol 2011, 55 (4), 547–554. https://doi.org/10.1007/s00484-010-0373-6.
(71) Mu, L.; Deng, F.; Tian, L.; Li, Y.; Swanson, M.; Ying, J.; Browne, R. W.; Rittenhouse-Olson, K.; Zhang, J. (Jim); Zhang, Z.-F.; Bonner, M. R. Peak Expiratory Flow, Breath Rate and Blood Pressure in Adults with Changes in Particulate Matter Air Pollution during the Beijing Olympics: A Panel Study. Environmental Research 2014, 133, 4–11. https://doi.org/10.1016/j.envres.2014.05.006.
(72) Rich, D. Q.; Liu, K.; Zhang, J.; Thurston, S. W.; Stevens, T. P.; Pan, Y.; Kane, C.; Weinberger, B.; Ohman, -Strickland Pamela; Woodruff, T. J.; Duan, X.; Assibey, -Mensah Vanessa; Zhang, J. Differences in Birth Weight Associated with the 2008 Beijing Olympics Air Pollution Reduction: Results from a Natural Experiment. Environmental Health Perspectives 2015, 123 (9), 880–887. https://doi.org/10.1289/ehp.1408795.
(73) Rich, D. Q.; Kipen, H. M.; Huang, W.; Wang, G.; Wang, Y.; Zhu, P.; Ohman-Strickland, P.; Hu, M.; Philipp, C.; Diehl, S. R.; Lu, S.-E.; Tong, J.; Gong, J.; Thomas, D.; Zhu, T.; Zhang, J. (Jim). Association Between Changes in Air Pollution Levels During the Beijing Olympics and Biomarkers of Inflammation and Thrombosis in Healthy Young Adults. JAMA 2012, 307 (19), 2068–2078. https://doi.org/10.1001/jama.2012.3488.
(74) Wang, W.; Primbs, T.; Tao, S.; Simonich, S. L. M. Atmospheric Particulate Matter Pollution during the 2008 Beijing Olympics. Environ. Sci. Technol. 2009, 43 (14), 5314–5320. https://doi.org/10.1021/es9007504.
(75) Rubin, D. B. For Objective Causal Inference, Design Trumps Analysis. The Annals of Applied Statistics 2008, 2 (3), 808–840. https://doi.org/10.1214/08-AOAS187.
(76) Dominici, F.; Bargagli-Stoffi, F. J.; Mealli, F. From Controlled to Undisciplined Data: Estimating Causal Effects in the Era of Data Science Using a Potential Outcome Framework. Harvard Data Science Review 2021. https://doi.org/10.1162/99608f92.8102afed.
(77) Pope, C. A.; Ezzati, M.; Dockery, D. W. Validity of Observational Studies in Accountability Analyses: The Case of Air Pollution and Life Expectancy. Air Qual Atmos Health 2012, 5 (2), 231–235. https://doi.org/10.1007/s11869-010-0130-3.
(78) Henneman, L. R. F.; Holmes, H. A.; Mulholland, J. A.; Russell, A. G. Meteorological Detrending of Primary and Secondary Pollutant Concentrations: Method Application and Evaluation Using Long-Term (2000–2012) Data in Atlanta. Atmospheric Environment 2015, 119, 201–210. https://doi.org/10.1016/j.atmosenv.2015.08.007.
(79) Butler, T. J.; Vermeylen, F. M.; Rury, M.; Likens, G. E.; Lee, B.; Bowker, G. E.; McCluney, L. Response of Ozone and Nitrate to Stationary Source NOx Emission Reductions in the Eastern USA. Atmospheric Environment 2011, 45 (5), 1084–1094. https://doi.org/10.1016/j.atmosenv.2010.11.040.
(80) Henneman, L. R. F.; Dedoussi, I. C.; Casey, J. A.; Choirat, C.; Barrett, S. R. H.; Zigler, C. M. Comparisons of Simple and Complex Methods for Quantifying Exposure to Individual Point Source Air Pollution Emissions. J Expo Sci Environ Epidemiol 2021, 31 (4), 654–663. https://doi.org/10.1038/s41370-020-0219-1.

(81) Henneman, L. R.; Choirat, C.; Zigler, C. M. Accountability Assessment of Health Improvements in the United States Associated with Reduced Coal Emissions Between 2005 and 2012. Epidemiology 2019, 30 (4), 477–485. https://doi.org/10.1097/EDE.0000000000001024.
(82) Bilsback, K. R.; Baumgartner, J.; Cheeseman, M.; Ford, B.; Kodros, J. K.; Li, X.; Ramnarine, E.; Tao, S.; Zhang, Y.; Carter, E. Estimated Aerosol Health and Radiative Effects of the Residential Coal Ban in the Beijing-Tianjin-Hebei Region of China. Aerosol and Air Quality Research 2020, 20 (11), 2332–2346.
(83) Lechner, M. The Estimation of Causal Effects by Difference-in-Difference Methods; Now Hanover, MA, 2011; Vol. 4.
(84) Wing, C.; Simon, K.; Bello-Gomez, R. A. Designing Difference in Difference Studies: Best Practices for Public Health Policy Research. Annual review of public health 2018, 39.
(85) Chatton, A.; Le Borgne, F.; Leyrat, C.; Gillaizeau, F.; Rousseau, C.; Barbin, L.; Laplaud, D.; Léger, M.; Giraudeau, B.; Foucher, Y. G-Computation, Propensity Score-Based Methods, and Targeted Maximum Likelihood Estimator for Causal Inference with Different Covariates Sets: A Comparative Simulation Study. Sci Rep 2020, 10 (1), 9219. https://doi.org/10.1038/s41598-020-65917-x.
(86) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; Chintala, S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Advances in Neural Information Processing Systems; Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F. d', Fox, E., Garnett, R., Eds.; Curran Associates, Inc., 2019; Vol. 32.
(87) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nat Methods 2020, 17 (3), 261–272. https://doi.org/10.1038/s41592-019-0686-2.
(88) Kuhn, M. Caret: Classification and Regression Training; 2020.
(89) Breiman, L. Random Forests. Machine Learning 2001, 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.
(90) Galit Shmueli. To Explain or to Predict? Statistical Science 2010, 25 (3), 289–310. https://doi.org/10.1214/10-STS330.
(91) Zhao, Q.; Hastie, T. Causal Interpretations of Black-Box Models. null 2021, 39 (1), 272–281. https://doi.org/10.1080/07350015.2019.1624293.
(92) Pearl, J. Causality; Cambridge university press, 2009.
(93) Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Advances in Neural Information Processing Systems; Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc., 2017; Vol. 30.
(94) Ribeiro, M. T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International

Conference on Knowledge Discovery and Data Mining; KDD ’16; Association for Computing Machinery: New York, NY, USA, 2016; pp 1135–1144. https://doi.org/10.1145/2939672.2939778.
(95) Streets, D. G.; Canty, T.; Carmichael, G. R.; de Foy, B.; Dickerson, R. R.; Duncan, B. N.; Edwards, D. P.; Haynes, J. A.; Henze, D. K.; Houyoux, M. R.; Jacob, D. J.; Krotkov, N. A.; Lamsal, L. N.; Liu, Y.; Lu, Z.; Martin, R. V.; Pfister, G. G.; Pinder, R. W.; Salawitch, R. J.; Wecht, K. J. Emissions Estimation from Satellite Retrievals: A Review of Current Capability. Atmospheric Environment 2013, 77, 1011–1042. https://doi.org/10.1016/j.atmosenv.2013.05.051.
(96) Jiang, T.; Chen, B.; Nie, Z.; Ren, Z.; Xu, B.; Tang, S. Estimation of Hourly Full-Coverage PM2.5 Concentrations at 1-Km Resolution in China Using a Two-Stage Random Forest Model. Atmospheric Research 2021, 248, 105146. https://doi.org/10.1016/j.atmosres.2020.105146.
(97) Chen, W.; Ran, H.; Cao, X.; Wang, J.; Teng, D.; Chen, J.; Zheng, X. Estimating PM2.5 with High-Resolution 1-Km AOD Data and an Improved Machine Learning Model over Shenzhen, China. Science of The Total Environment 2020, 746, 141093. https://doi.org/10.1016/j.scitotenv.2020.141093.
(98) Goldberg, D. L.; Gupta, P.; Wang, K.; Jena, C.; Zhang, Y.; Lu, Z.; Streets, D. G. Using Gap-Filled MAIAC AOD and WRF-Chem to Estimate Daily PM2.5 Concentrations at 1 km Resolution in the Eastern United States. Atmospheric Environment 2019, 199, 443–452. https://doi.org/10.1016/j.atmosenv.2018.11.049.
(99) Chen, G.; Li, Y.; Zhou, Y.; Shi, C.; Guo, Y.; Liu, Y. The Comparison of AOD-Based and Non-AOD Prediction Models for Daily PM2.5 Estimation in Guangdong Province, China with Poor AOD Coverage. Environmental Research 2021, 195, 110735. https://doi.org/10.1016/j.envres.2021.110735.
(100) Hu, X.; Belle, J. H.; Meng, X.; Wildani, A.; Waller, L. A.; Strickland, M. J.; Liu, Y. Estimating PM $_{2.5}$ Concentrations in the Conterminous United States Using the Random Forest Approach. Environ. Sci. Technol. 2017, 51 (12), 6936–6944. https://doi.org/10.1021/acs.est.7b01210.
(101) Huang, K.; Xiao, Q.; Meng, X.; Geng, G.; Wang, Y.; Lyapustin, A.; Gu, D.; Liu, Y. Predicting Monthly High-Resolution PM2.5 Concentrations with Random Forest Model in the North China Plain. Environmental Pollution 2018, 242, 675–683. https://doi.org/10.1016/j.envpol.2018.07.016.
(102) Badura, M.; Batog, P.; Drzeniecka-Osiadacz, A.; Modzel, P. Evaluation of Low-Cost Sensors for Ambient PM2. 5 Monitoring. Journal of Sensors 2018, 2018.
(103) Bulot, F. M.; Johnston, S. J.; Basford, P. J.; Easton, N. H.; Apetroaie-Cristea, M.; Foster, G. L.; Morris, A. K.; Cox, S. J.; Loxham, M. Long-Term Field Comparison of Multiple Low-Cost Particulate Matter Sensors in an Outdoor Urban Environment. Scientific reports 2019, 9 (1), 1–13.
(104) Volckens, J.; Quinn, C.; Leith, D.; Mehaffy, J.; Henry, C. S.; Miller-Lionberg, D. Development and Evaluation of an Ultrasonic Personal Aerosol Sampler. Indoor Air 2017, 27 (2), 409–416. https://doi.org/10.1111/ina.12318.
(105) The ERA5 global reanalysis - Hersbach - 2020 - Quarterly Journal of the Royal Meteorological Society - Wiley Online Library. https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3803 (accessed 2021-09-09).

(106) Anderson, G. B.; Bell, M. L.; Peng, R. D. Methods to Calculate the Heat Index as an Exposure Metric in Environmental Health Research. Environmental Health Perspectives 2013, 121 (10), 1111–1119. https://doi.org/10.1289/ehp.1206273.
(107) Breiman, L. Random Forests. Machine learning 2001, 45 (1), 5–32.
(108) Breiman, L. Consistency for a Simple Model of Random Forests. 2004.
(109) Kaur, H.; Pannu, H. S.; Malhi, A. K. A Systematic Review on Imbalanced Data Challenges in Machine Learning: Applications and Solutions. ACM Computing Surveys (CSUR) 2019, 52 (4), 1–36.
(110) Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. Learning from Class-Imbalanced Data: Review of Methods and Applications. Expert systems with applications 2017, 73, 220–239.
(111) Biau, G.; Scornet, E. A Random Forest Guided Tour. TEST 2016, 25 (2), 197–227. https://doi.org/10.1007/s11749-016-0481-7.
(112) Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. Annals of statistics 2001, 1189–1232.
(113) Friedman, J. H.; Popescu, B. E. Predictive Learning via Rule Ensembles. The annals of applied statistics 2008, 2 (3), 916–954.
(114) Huang, Q.; Cai, X.; Wang, J.; Song, Y.; Zhu, T. Climatological Study of the Boundary-Layer Air Stagnation Index for China and Its Relationship with Air Pollution. Atmospheric Chemistry and Physics 2018, 18 (10), 7573–7593. https://doi.org/10.5194/acp-18-7573-2018.
(115) Elangasinghe, M. A.; Singhal, N.; Dirks, K. N.; Salmond, J. A.; Samarasinghe, S. Complex Time Series Analysis of PM10 and PM2.5 for a Coastal Site Using Artificial Neural Network Modelling and k-Means Clustering. Atmospheric Environment 2014, 94, 106–116. https://doi.org/10.1016/j.atmosenv.2014.04.051.
(116) Ploton, P.; Mortier, F.; Réjou-Méchain, M.; Barbier, N.; Picard, N.; Rossi, V.; Dormann, C.; Cornu, G.; Viennois, G.; Bayol, N.; Lyapustin, A.; Gourlet-Fleury, S.; Pélissier, R. Spatial Validation Reveals Poor Predictive Performance of Large-Scale Ecological Mapping Models. Nat Commun 2020, 11 (1), 4540. https://doi.org/10.1038/s41467-020-18321-y.
(117) Meyer, H. CAST: “caret” Applications for Spatial-Temporal Models; 2021.
(118) Li, X.; Clark, S.; Floess, E.; Baumgartner, J.; Bond, T.; Carter, E. Personal Exposure to PM2.5 of Indoor and Outdoor Origin in Two Neighboring Chinese Communities with Contrasting Household Fuel Use Patterns. Science of The Total Environment 2021, 800, 149421. https://doi.org/10.1016/j.scitotenv.2021.149421.
(119) Liu, K.; Ren, J. Characteristics, Sources and Health Risks of PM2.5-Bound Potentially Toxic Elements in the Northern Rural China. Atmospheric Pollution Research 2019, 10 (5), 1621–1626. https://doi.org/10.1016/j.apr.2019.06.002.
(120) Zhao, X.; Zhao, X.; Liu, P.; Ye, C.; Xue, C.; Zhang, C.; Zhang, Y.; Liu, C.; Liu, J.; Chen, H.; Chen, J.; Mu, Y. Pollution Levels, Composition Characteristics and Sources of Atmospheric PM2.5 in a Rural Area of the North China Plain during Winter. Journal of Environmental Sciences 2020, 95, 172–182. https://doi.org/10.1016/j.jes.2020.03.053.
(121) Lv, D.; Chen, Y.; Zhu, T.; Li, T.; Shen, F.; Li, X.; Mehmood, T. The Pollution Characteristics of PM10 and PM2.5 during Summer and Winter in Beijing, Suning and Islamabad. Atmospheric Pollution Research 2019, 10 (4), 1159–1164. https://doi.org/10.1016/j.apr.2019.01.021.

(122) Meng, Z.; Wu, L.; Xu, X.; Xu, W.; Zhang, R.; Jia, X.; Liang, L.; Miao, Y.; Cheng, H.; Xie, Y.; He, J.; Zhong, J. Changes in Ammonia and Its Effects on PM2.5 Chemical Property in Three Winter Seasons in Beijing, China. Science of The Total Environment 2020, 749, 142208. https://doi.org/10.1016/j.scitotenv.2020.142208.
(123) Organization, W. H. WHO Global Air Quality Guidelines: Particulate Matter (PM2. 5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide: Executive Summary. 2021.
(124) Chen, G.; Li, S.; Knibbs, L. D.; Hamm, N. A. S.; Cao, W.; Li, T.; Guo, J.; Ren, H.; Abramson, M. J.; Guo, Y. A Machine Learning Method to Estimate PM2.5 Concentrations across China with Remote Sensing, Meteorological and Land Use Information. Science of The Total Environment 2018, 636, 52–60. https://doi.org/10.1016/j.scitotenv.2018.04.251.
(125) Di, Q.; Kloog, I.; Koutrakis, P.; Lyapustin, A.; Wang, Y.; Schwartz, J. Assessing PM2.5 Exposures with High Spatiotemporal Resolution across the Continental United States. Environ. Sci. Technol. 2016, 50 (9), 4712–4721. https://doi.org/10.1021/acs.est.5b06121.
(126) Chen, Z.; Chen, D.; Zhao, C.; Kwan, M.; Cai, J.; Zhuang, Y.; Zhao, B.; Wang, X.; Chen, B.; Yang, J.; Li, R.; He, B.; Gao, B.; Wang, K.; Xu, B. Influence of Meteorological Conditions on PM2.5 Concentrations across China: A Review of Methodology and Mechanism. Environment International 2020, 139, 105558. https://doi.org/10.1016/j.envint.2020.105558.
(127) Li, Y.; Chen, Q.; Zhao, H.; Wang, L.; Tao, R. Variations in PM10, PM2.5 and PM1.0 in an Urban Area of the Sichuan Basin and Their Relation to Meteorological Factors. Atmosphere 2015, 6 (1), 150–163. https://doi.org/10.3390/atmos6010150.
(128) Wang, X.; Zhang, R.; Yu, W. The Effects of PM2.5 Concentrations and Relative Humidity on Atmospheric Visibility in Beijing. Journal of Geophysical Research: Atmospheres 2019, 124 (4), 2235–2259. https://doi.org/10.1029/2018JD029269.
(129) Yuanqin, Y.; Jizhi, W.; Qing, H.; Yaqiang, W. A Plam Index Forecast Method for Air Quality of Beijing in Summer. 应用气象学报 2009, 20 (6), 649–655.
(130) Liao, J.; Zimmermann Jin, A.; Chafe, Z. A.; Pillarisetti, A.; Yu, T.; Shan, M.; Yang, X.; Li, H.; Liu, G.; Smith, K. R. The Impact of Household Cooking and Heating with Solid Fuels on Ambient PM2.5 in Peri-Urban Beijing. Atmospheric Environment 2017, 165, 62–72. https://doi.org/10.1016/j.atmosenv.2017.05.053.
(131) Yu, Q.; Yang, W.; Zhu, M.; Gao, B.; Li, S.; Li, G.; Fang, H.; Zhou, H.; Zhang, H.; Wu, Z.; Song, W.; Tan, J.; Zhang, Y.; Bi, X.; Chen, L.; Wang, X. Ambient PM2.5-Bound Polycyclic Aromatic Hydrocarbons (PAHs) in Rural Beijing: Unabated with Enhanced Temporary Emission Control during the 2014 APEC Summit and Largely Aggravated after the Start of Wintertime Heating. Environmental Pollution 2018, 238, 532–542. https://doi.org/10.1016/j.envpol.2018.03.079.
(132) Grange, S. K.; Carslaw, D. C. Using Meteorological Normalisation to Detect Interventions in Air Quality Time Series. Science of The Total Environment 2019, 653, 578–588. https://doi.org/10.1016/j.scitotenv.2018.10.344.
(133) Qu, L.; Liu, S.; Ma, L.; Zhang, Z.; Du, J.; Zhou, Y.; Meng, F. Evaluating the Meteorological Normalized PM2.5 Trend (2014–2019) in the “2+26” Region of China Using an Ensemble Learning Technique. Environmental Pollution 2020, 266, 115346. https://doi.org/10.1016/j.envpol.2020.115346.

APPENDICES

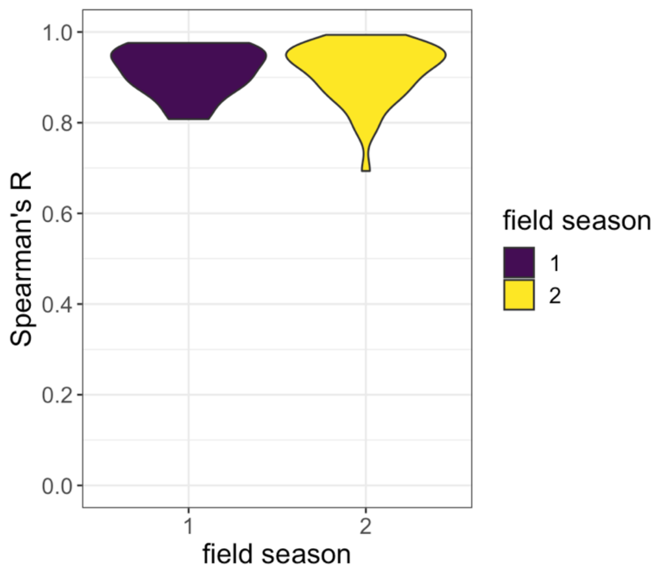

Figure S1. Violin plot of the Spearman's correlation coefficient among sensors in the same village during season 1 and 2. The width of each violin is scaled to the number values observed at a given y.

Table S1. The number of villages with 1-3 real-time $PM_{2.5}$ sensors during each field season

number of sensors	Season 1 (n villages)	Season 2 (n villages)
1	31	6
2	4	44
3	4	0

Text S1. Sensor calibration information

Filter-based $PM_{2.5}$ measurements were transported from the field to Colorado State University for gravimetric analysis. Filters were stored in -20 C freezers while not in use or being analyzed. Gravimetric analysis was conducted using the Automated Air Analysis Facility (AIRLIFT) described in L'Orange et al., 2021. The AIRLIFT system uses a 6-axis articulating robotic arm to weigh samples inside of a temperature and relative humidity-controlled chamber. Filters are weighed in triplicate using a microbalance (Mettler Toledo XS3DU, Columbus, OH, USA) with a 1 μg resolution. Samples were equilibrated in the chamber for 24 hours prior to weighing and statically discharged with a polonium 210 source (2U500, NRD, Grand Island, NY, USA) immediately before weighing.

In the first season (winter 2018-2019), two sets of the same $PM_{2.5}$ sensors purchased at different times (July 2017 and 2018) were used. Prior to deployment, all sensors were co-located with a Thermo Electron Synchronized Hybrid Ambient Real-Time Particulate Monitor (SHARP, model 5030, Thermo Fischer) at Peking University for 7-10 days. A linear regression comparing the time integrated $PM_{2.5}$ measurements to the reference monitor are shown in Figure S2. During the co-location period, an air pollution episode occurred, during which the meteorological conditions rapidly changed and $PM_{2.5}$ concentrations increased considerably in a short period of time. The reference monitor captures $PM_{2.5}$ concentrations at a one-hour time resolution which was not fast enough to capture a continuous range of $PM_{2.5}$ concentrations during the pollution episode. Therefore, we observed a break in $PM_{2.5}$ concentrations from ~100 μg/m^3 to ~160 μg/m^3. $PM_{2.5}$ sensors were also corrected post-deployment by developing a correction factor with filter base measurements. The linear regression comparing filter-based $PM_{2.5}$ measurements to $PM_{2.5}$ sensor measurements integrated over the same time period that the filter was sampled over is shown in Figure S3.

Two sets of the same $PM_{2.5}$ sensor that were purchased at different time (July 2018 and October 2019) were deployed in the second season. $PM_{2.5}$ sensors were co-located with the SHARP monitor at Peking University for 7 days, and a tapered element oscillating balance (TEOM model 1400A; Thermo Fischer) at the University of Chinese Academy of Sciences (UCAS) for 10 days prior to deployment (Figure S4). The quality of the TEOM data was inadequate, so we instead used data collected at the Beijing Environmental Monitoring Station

(Huairou station) over the same time period. After deployment, the sensors were co-located with the SHARP monitor at Peking university for 7 days, and the TEOM at UCAS for 10 days (Figure S5). Additionally, a linear regression between filter-based measurements and time-integrated $PM_{2.5}$ measurements was applied to develop a correction factor for the sensors (Figure S6).

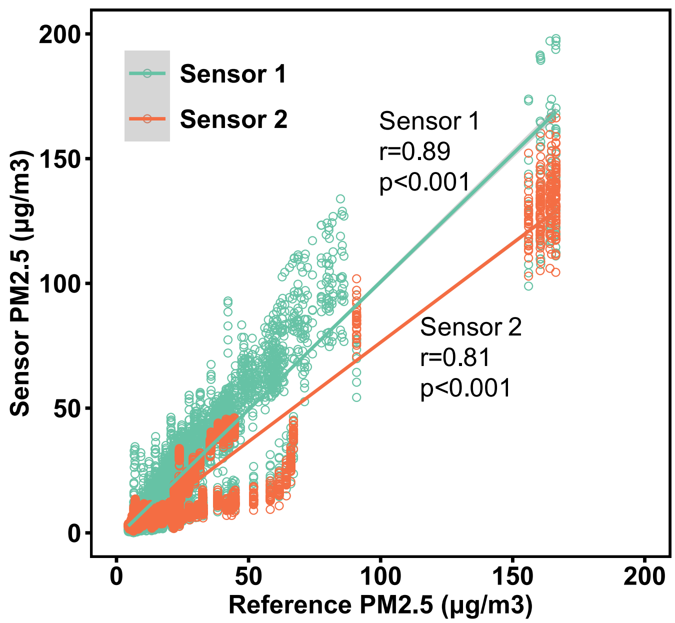

Figure S2. Linear regression between $PM_{2.5}$ measured by a reference instrument (TEOM) and real-time $PM_{2.5}$ sensor in season one prior to deployment. The correlation coefficients (rho) are spearman correlation coefficients.

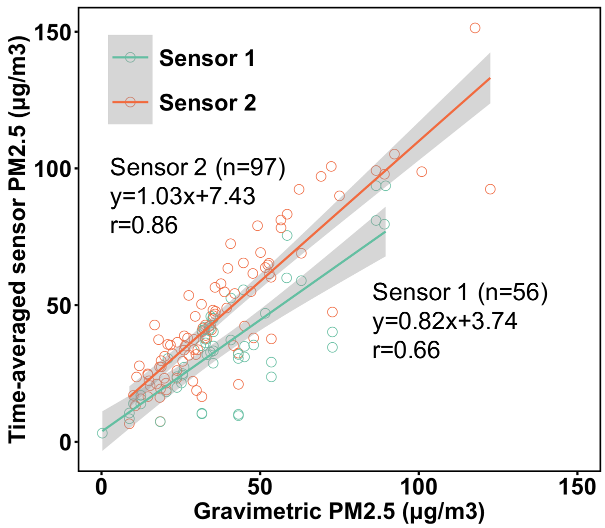

Figure S3. Linear regression between outdoor gravimetric $PM_{2.5}$ and time-averaged sensor-based $PM_{2.5}$ in season one. The correlation coefficients (rho) are spearman correlation coefficients. The grey shade shows the 95% confidence interval.

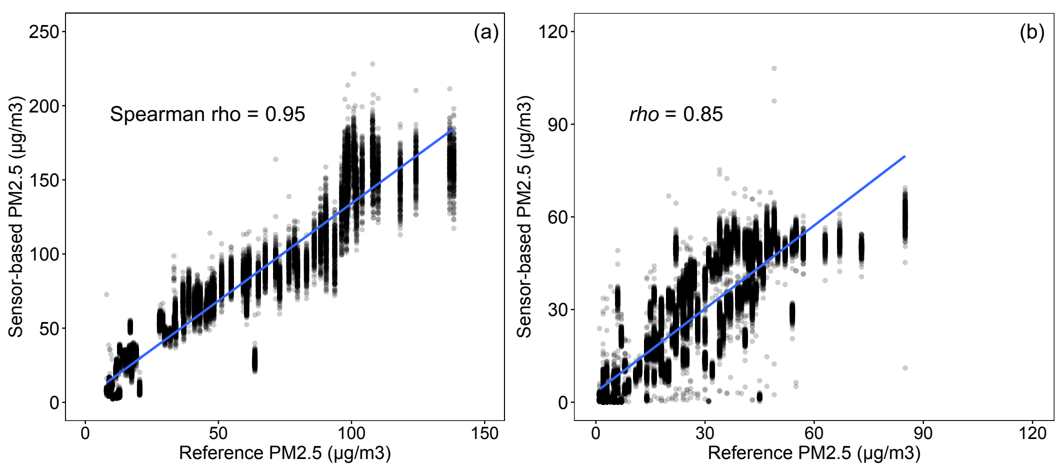

Figure S4. Correlation between sensor and reference instrument measured $PM_{2.5}$ at PKU (a) and UCAS (b) campus before the field campaign in Season 2. Due to the bad data quality of TEOM data from UCAS campus, TEOM data is from the Beijing Environmental Monitoring Station (Huairou station) (Same below).

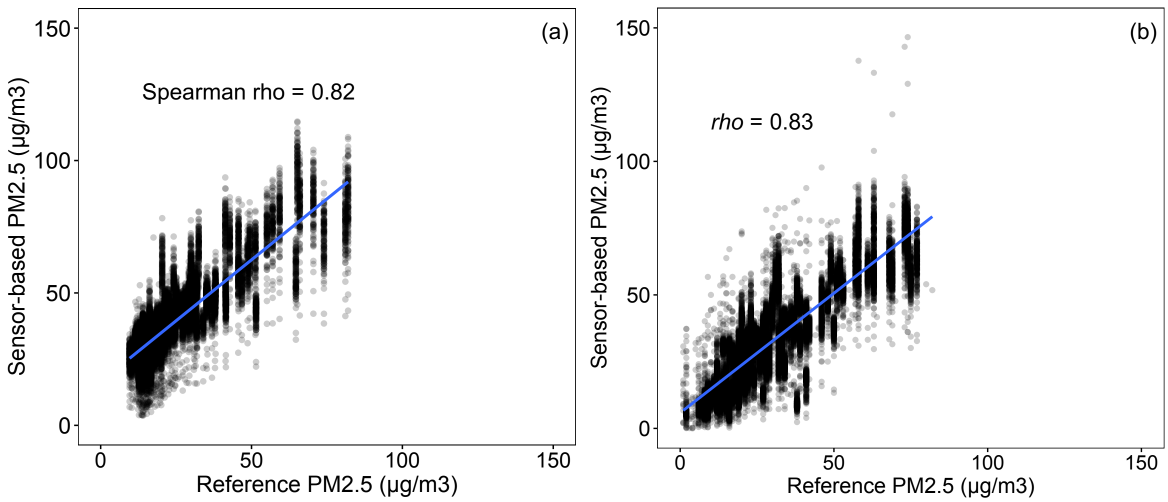

Figure S5. Correlation between sensor and reference instrument measured $PM_{2.5}$ at PKU (a) and UCAS (b) campus after the field campaign in Season 2.

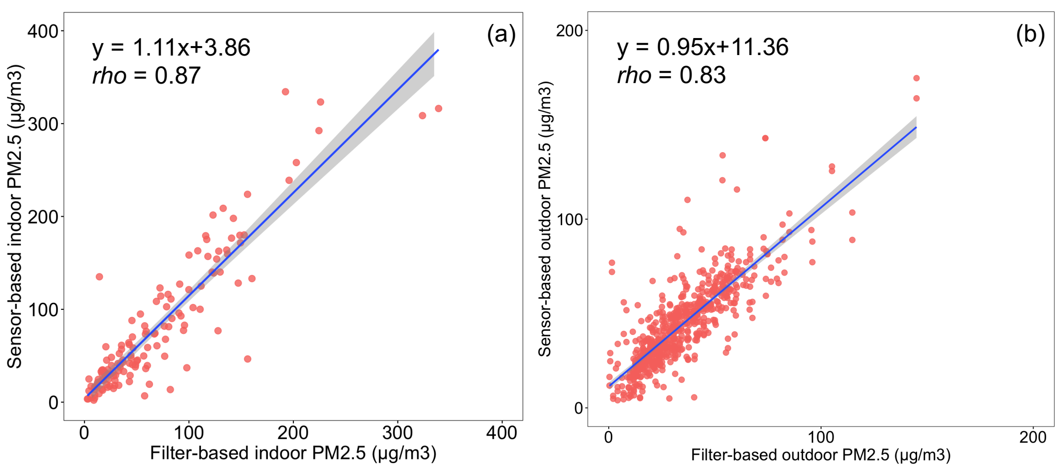

Figure S6. Calibration curves of indoor (a) and outdoor (b) PM sensors by filter-based measurements in season two. The correlation coefficients (rho) are spearman correlation coefficients. The grey shade shows the 95% confidence interval.

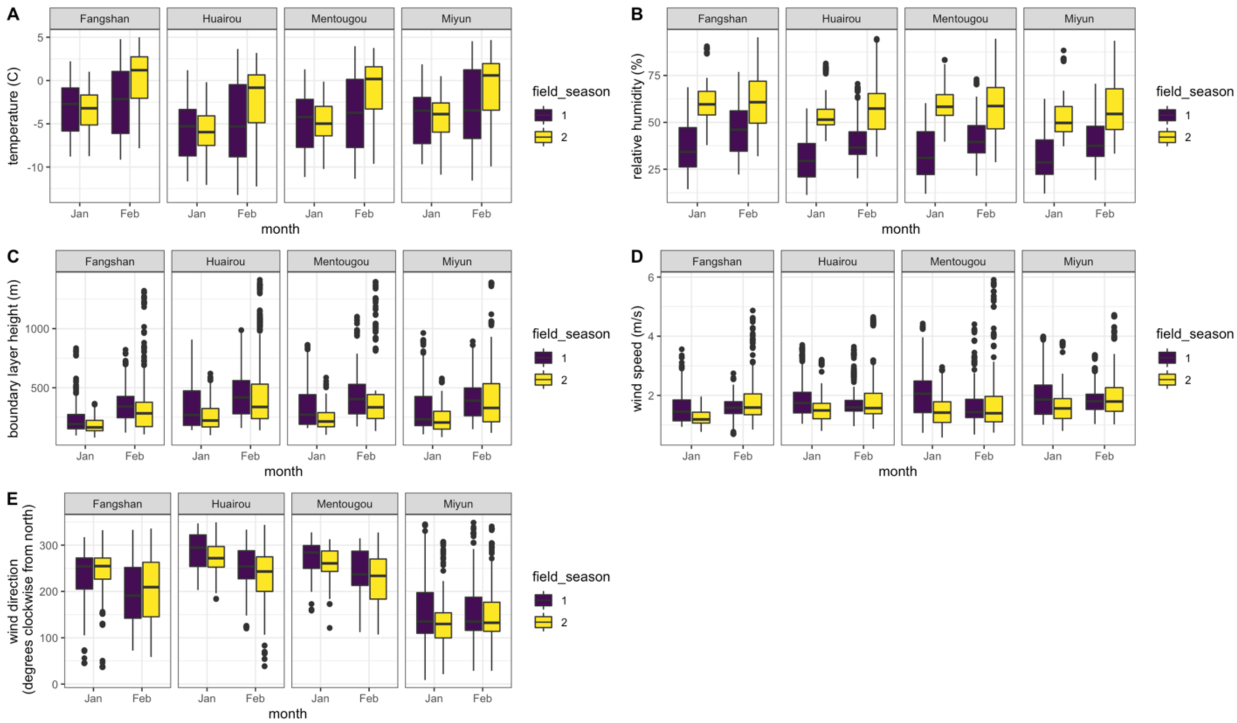

Figure S7. Boxplot of daily average meteorological values by district and season. (A) temperature, (B) relative humidity, (C) boundary layer height, (D) wind speed, and (E) wind direction.

Table S2. Predictive performance for the random forest sensitivity analyses

Sensitivity analysis	original model (r^2, RMSE)	new model (r^2, RMSE)
Removing DOY from model		
Season one model	0.85, 13.1	0.82, 14.4
Season two model	0.93, 11.2	0.90, 13.3
Season one model to predict season two data		
Removing DOY	0.45, 33.8	0.49, 30.7
Restricting the range of season 2 data to include meteorological variables with value in the range of the season one data	0.45, 33.8	0.44, 31.7
Excluding villages that underwent an energy transition between season one and two	0.45, 33.8	0.47, 31.3
Removing DOY, restricting season 2 data range, and removing energy transition villages	0.45, 33.8	0.46, 29.5