



A MAIA-like modeling framework to estimate PM_{2.5} mass and speciation concentrations with uncertainty

Zhihao Jin^a, Qiang Pu^a, Nathan Janecek^b, Huanxin Zhang^b, Jun Wang^b, Howard Chang^c, Yang Liu^{a,*}

^a Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA

^b Department of Chemical and Biochemical Engineering, Iowa Technology Institute, Center for Global and Regional Environmental Research, University of Iowa, Iowa City, IA 52242, USA

^c Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA

ARTICLE INFO

Edited by Menghua Wang

Keywords:

PM_{2.5}
MAIA
Bayesian inference
GOES-16 AOD
WRF-Chem
Exposure uncertainty

ABSTRACT

Ambient fine particulate matter (PM_{2.5}) is strongly associated with various adverse health outcomes. However, the lack of extensive PM_{2.5} measurements, and especially of its components, hinders the assessment of negative health effects caused by PM_{2.5} in many parts of the world. To address this issue, a new satellite instrument, the Multi-angle Imager for Aerosols (MAIA), with improved design for providing aerosol optical depth (AOD) of high quality, will be helpful in determining concentrations of total and speciated PM_{2.5}. According to the retrieval algorithm of MAIA particulate matter (PM) products, level 2 (L2) PM products are generated based on MAIA AOD on days of observation. Bias-corrected chemical transport model (CTM) outputs are then merged with the L2 PM products to fill their gaps using a Bayesian Model Averaging (BMA) ensemble framework. This process creates the MAIA Level 4 (L4) gap-filled PM products. In this study, we aim to implement the MAIA framework and validate its feasibility after the launch of the MAIA satellite instrument. We used both Bayesian hierarchical model (BHM) and a Bayesian additive regression tree (BART) to predict L2 and CTM-based daily $1 \times 1 \text{ km}^2$ PM_{2.5} mass and speciation concentrations, along with prediction uncertainties, over the MAIA Primary Target Area in the Northeastern US in 2018. We then employed the BMA ensemble model to combine the L2 and CTM-based PM_{2.5} mass predictions to fill gaps in L2 PM_{2.5} mass and produce Level 4 (L4) gap-filled PM_{2.5} mass. Our cross-validation experiments showed that both the BHM (R^2 ranging from 0.60 to 0.82) and BART (R^2 ranging from 0.59 to 0.79) models performed well in predicting CTM-based PM_{2.5} speciation, with better results for sulfate, organic carbon, and elemental carbon. At the stage of L4 PM_{2.5} mass predictions, both BHM-based and BART-based BMA ensemble models demonstrated improved performance with their traditional R^2 of 0.81 and 0.73, surpassing the input L2 and CTM-based PM_{2.5} mass. Additionally, our models showed excellent prediction uncertainty control with the coverage rates of 95% posterior prediction confidence interval associated of concentration estimates to be 95% for BHM and 75% for BART across PM_{2.5} mass and speciation. Results from the proposed modeling techniques contribute to a deeper understanding of the health effects of PM_{2.5} for future epidemiological studies and provide insights into the MAIA mission for producing improved PM products for health research.

1. Introduction

Previous studies have documented the adverse health effects of air pollution, accounting for 6.4 million premature deaths and 209 million disability adjusted life years worldwide (Chowdhury et al., 2023). Among different types of air pollution, many studies have focused on airborne particulate matter, particularly the ambient fine particulate

matter (PM_{2.5}, particles with a diameter of $<2.5 \mu\text{m}$). PM_{2.5} has been linked to premature death (Orellano et al., 2020), cardiovascular and respiratory diseases (Yang et al., 2022; Zhang et al., 2022b), lung cancer (Pun et al., 2017), and adverse birth outcomes (Li et al., 2019). It is also increasingly recognized that different PM_{2.5} chemical components show distinct health impacts. For example, decreased sulfate levels were found to be associated with decreased cases of non-accidental and

* Corresponding author.

E-mail address: yang.liu@emory.edu (Y. Liu).

<https://doi.org/10.1016/j.rse.2024.113995>

Received 24 April 2023; Received in revised form 8 November 2023; Accepted 10 January 2024

Available online 19 January 2024

0034-4257/© 2024 Elsevier Inc. All rights reserved.

cardiopulmonary deaths (Meng et al., 2023). Sulfate and nitrate were suggested to play important roles in the effects of PM_{2.5} exposure on being lower or higher for gestational age (Shen et al., 2022). However, these studies are often limited due to sparsely distributed monitoring networks that provide PM_{2.5} observations (Zhang et al., 2020c) or PM_{2.5} simulations with coarse spatial-temporal resolutions (Liu et al., 2017). Additionally, monitors capable of measuring PM_{2.5} speciation are especially rare due to the high operating and analysis cost, severely constraining our ability to examine the toxicity of different chemical components of PM_{2.5} in population-based epidemiological studies (Diner et al., 2018; Geng et al., 2020; Meng et al., 2018). To increase the spatial and temporal coverage of both PM_{2.5} mass and speciation exposure estimates, data fusion models that incorporate information from monitoring measurements with complementary data sources are needed.

Aerosol optical depth (AOD) retrieved by satellite instruments can provide insights into the abundance of airborne particles. Since the launch of NASA's *Terra* satellite in late 1999, multiple AOD products have been developed based on measurements from advanced spaceborne instruments such as the Multi-angle Imaging SpectroRadiometer (MISR) (Garay et al., 2020), Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) (Winker et al., 2009), Moderate resolution Imaging Spectroradiometer (MODIS) (Lyapustin et al., 2018), and Visible Infrared Imaging Radiometer Suite (VIIRS) (Murphy et al., 2006). Because of the high spatiotemporal resolution and large geographical coverage, satellite-retrieved AOD has been widely used as a predictor of near-surface PM₁₀ and PM_{2.5} concentrations in various statistical and machine learning models (Chen et al., 2018; He et al., 2021; Pu and Yoo, 2022; She et al., 2020). For example, Zheng et al. (Zheng et al., 2017) explored the factors that influence the relationship between PM_{2.5} and AOD in Beijing. To overcome the issue of AOD missingness due to cloud cover or bright surfaces, chemical transport model (CTM) simulations are increasingly being adopted in AOD gap filling (Xiao et al., 2017; Xue et al., 2019). Satellite-driven PM_{2.5} mass concentrations have been extensively applied to investigate the associations between PM_{2.5} exposure and various adverse health outcomes (Cohen et al., 2017; Crouse et al., 2015). On the other hand, only a few satellite aerosol data products possess the capability to characterize particle types, which is essential for filling the information gap created by the sparsely distributed speciation monitors. Since MISR fractional AOD values contain information on particle size distribution, shape, and light absorption, they have been used as important predictors of PM_{2.5} speciation (Franklin et al., 2017; Geng et al., 2020; Hang et al., 2022; Liu et al., 2007a; Liu et al., 2007b; Meng et al., 2018; Meng et al., 2023; Zhang et al., 2020b). This level of information is not currently available from MODIS and VIIRS. However, MISR's relatively low spatial resolution (4.4 km for the aerosol product), narrow swath (~400 km), and low revisit frequency (~ once in 9 days), often limit its ability to generate PM_{2.5} speciation estimates to support large-scale air pollution epidemiological research. CALIPSO is also limited by its narrow swath (~ 60 km) and low revisit frequency (~ once in 16 days) (Yang et al., 2021).

In addition to the sensors' limitation, another challenge in air pollution exposure assessment is exposure error, which tends to attenuate the estimated health risks towards the null (Goddard et al., 2020; Keller et al., 2017; Levy et al., 2019). Current exposure models focusing on fusing CTM simulations with AOD values or imputing missing AOD values primarily use multi-stage regression models or machine learning algorithms (Amini et al., 2022; Ma et al., 2022; Zhang et al., 2020a). However, most of these models are unable to estimate, or effectively estimate, prediction uncertainty with the scarce ground-based observations, CTMs, and satellite aerosol products. To address this issue, Bayesian inference methods have been proposed, offering more realistic and model-based uncertainty estimations by leveraging Markov chain Monte Carlo (MCMC) techniques and posterior probability distributions (Box and Tiao, 2011). For example, Chang et al. (Chang et al., 2014) introduced a Bayesian hierarchical model (BHM) that calibrates MODIS

AOD to predict daily PM_{2.5} concentrations in the southeastern US, achieving an R² of 0.78 in cross validation (CV) experiments. Murray et al. (Murray et al., 2019) used the Bayesian Model Averaging (BMA) ensemble framework to integrate PM_{2.5} estimates from both the CMAQ-based BHM and AOD-based BHM. The BMA ensemble model outperformed its input BHMs in 10-fold CV and yielded with an R² of 0.83 and a 97.15% coverage rate of the 95% prediction confidence interval (CI).

Recognizing the limitations of current aerosol sensors such as MISR, and the challenges in exposure error quantification, an advanced satellite instrument, the Multi-Angle Imager for Aerosols (MAIA), has been developed to address these shortcomings. The MAIA mission is a collaborative endeavor between NASA and the Italian Space Agency (Agenzia Spaziale Italiana or ASI). This collaboration encompasses a joint scientific program, an integrated spaceborne observatory comprised of NASA's MAIA satellite instrument and ASI's PLATI^{NO}-2 spacecraft, networks of surface-based pollution sensors, and systems to launch and manage the observatory. MAIA aims to provide high-resolution ground-level PM composition information in a select set of large population centers around the world (Primary Target Areas or PTAs in MAIA terminology), and examine their associations with various adverse health outcomes (Liu and Diner, 2017). The MAIA instrument integrates multispectral, polarimetric, and multiangular capabilities for detailed mapping of total and speciated PM at the neighborhood level. A two-axis gimbal is used to mount the MAIA camera, which allows for more frequent sampling of selected target areas. Unlike MISR, which uses 9 fixed cameras, MAIA employs a pointable single camera, providing enhanced flexibility in its observations. While MISR is equipped with 4 spectral bands, MAIA has a detailed set of 14 spectral bands, spanning ultraviolet to shortwave-infrared. Notably, MAIA introduces 3 polarimetric bands, a feature unavailable in MISR, which amplifies its capability to measure additional aerosol microphysical properties. The MAIA observatory will be deployed in a low-Earth polar orbit, 740 km above the surface. The anticipated launch is set for 2025.

Built upon MISR's legacy with improved instrumental design, MAIA employs a sophisticated PM modeling framework that integrates MAIA aerosol retrievals, CTM simulations from the Unified Inputs for WRF-Chem (UI-WRF-Chem), and ground observations. The framework produces different level of PM products reflecting daily average PM_{2.5} components concentration at 1 × 1 km² resolution. In MAIA's terminology, the Level 2 (L2) products utilize satellite aerosol retrieval for input. The CTM-based products are developed by calibrating CTM simulations. The final Level 4 gap-filled products (L4) are the ensemble merging of the L2 and the CTM-based products (Diner et al., 2018). To quantify the prediction uncertainty, MAIA adopts the BHM and BMA models as the basis of the modeling framework, as these models also suggest potential benefits of downscaling spatial resolution (Geng et al., 2018; Wang et al., 2013). Specifically, BHM is used to produce L2 products and CTM-based PM estimates, and they are subsequently merged in a BMA ensemble model to derive L4 gap-filled PM_{2.5} products.

In this study, we assessed the ability of the MAIA PM_{2.5} modeling framework in producing L2 products, CTM-based estimates and L4 PM_{2.5} products in one of MAIA's domestic PTAs. UI-WRF-Chem PM_{2.5} simulations were used as operational CTM in the MAIA framework and AOD from the Geostationary Operational Environmental Satellite (GOES) 16 satellite (GOES-16) was used as the proxy of MAIA AOD. We closely followed the Algorithm Theoretical Basis Document (Diner et al., 2019) of the MAIA PM_{2.5} products framework to design a pre-launch version of the BHM and BMA models that estimate MAIA-like total PM_{2.5} mass and PM_{2.5} speciation concentrations with prediction uncertainties. As an alternative to BHM, we also explored the performance of a novel machine learning algorithm called the Bayesian Additive Regression Trees (BART) and tested the BMA ensemble model's extensibility. Taking advantage of the relatively abundant ground PM_{2.5} speciation measurements in our chosen target area, we conducted sensitivity analyses using reduced model training datasets in order to evaluate the reliability

of our proposed models in data-poor regions (e.g., the PTAs located in low- or middle-income countries with sparser ground observations). Similar to many satellite-derived products, the generation of the MAIA PM product is iterative. This iterative process evolves as more training samples are incorporated over time. As the launch of MAIA is approaching, our study establishes a foundational performance benchmark and provides valuable insights into the MAIA-based PM products that aim to support future air pollution health effect research.

2. Material and methods

2.1. Study domain

Our study domain is located in the northeastern US (Fig. 1), and covers part of Pennsylvania, New Jersey, New York, Vermont, New Hampshire, Maine, Massachusetts, and the entirety of Connecticut and Rhode Island, which corresponds to the MAIA Northeastern US PTA (Diner et al., 2018). After excluding the grid cells off the coast, there are a total of 102,896 $1 \times 1 \text{ km}^2$ grid cells remaining.

2.2. Data

2.2.1. Air quality monitoring data

The daily (24-h average) concentrations of total $\text{PM}_{2.5}$ mass and $\text{PM}_{2.5}$ species (sulfate, nitrate, organic carbon [OC], elemental carbon [EC], and dust) across the study domain in 2018 were obtained from the US Environmental Protection Agency's (EPA) Air Quality System (AQS) and the Interagency Monitoring of Protected Visual Environments (IMPROVE) Network. The data was collected from 49 total $\text{PM}_{2.5}$ mass monitoring stations and 16 $\text{PM}_{2.5}$ species monitoring stations. To maximize our sample size, we included one station located slightly

outside our study domain. The mean concentration and component percentages of $\text{PM}_{2.5}$ mass and speciation are summarized in Table S1, with an average $\text{PM}_{2.5}$ concentration of $6.88 \mu\text{g}/\text{m}^3$. OC contributes 26.76%, representing the largest proportion, whereas dust constitutes the smallest proportion at 5.98% of the total $\text{PM}_{2.5}$.

2.2.2. GOES-16 ABI AOD

The Geostationary Operational Environmental Satellite (GOES) 16 satellite (launched in November 2016) carrying the Advanced Baseline Imager (ABI) is the first of the new GOES-R series operated by NASA and the National Oceanic and Atmospheric Administration (NOAA) (Laszlo and Liu, 2016). To simulate future MAIA AOD products, we utilized GOES-16 ABI 5-min Level 2 AOD retrievals at 550 nm at a $2 \times 2 \text{ km}^2$ resolution as a proxy. Given our limited model training period (i.e., 12 months), we chose GOES-16 because it provides high temporal resolution (5-min) AOD over North America, ensuring the largest dataset for training when matched with ground-based observations (Laszlo and Liu, 2016). Among the widely used AOD products, MAIAC has a superior spatial resolution of 1 km. However, its daily revisit rate, even operating in both Terra and Aqua MODIS sensors, does not provide sufficient data for our 12-month modeling period. While MISR provides fractional AOD retrievals, its revisit frequency of 1–2 times weekly is insufficient to compile our model training dataset. Finally, VIIRS AOD has a coarse spatial resolution of 6 km along with a daily revisit schedule.

To ensure the consistent temporal resolution with the MAIA AOD and PM products and reduce missingness, we aggregated the 5-min level AOD values into hourly AOD, then further averaged the hourly data into daily means. This process also reduced the average AOD missing rate in the PTA from 91.10% at the 5-min level to 37.85% at the daily level (Fig. S1).

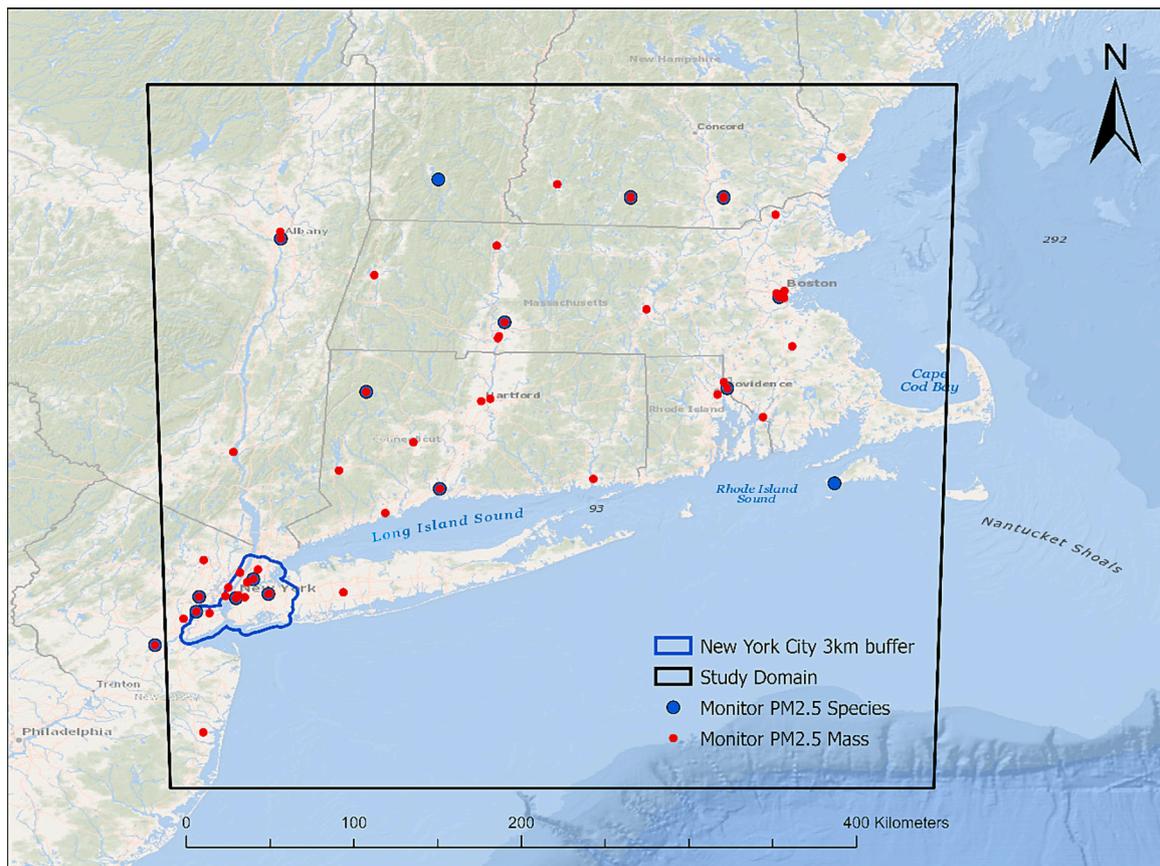


Fig. 1. Study region showing all available $\text{PM}_{2.5}$ mass (red color) and species (blue color) monitors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2.3. UI-WRF-Chem

In this study, the Unified Inputs (initial and boundary conditions) for WRF-Chem (Sha et al., 2021; Zhang et al., 2022a) was used to simulate meteorological fields and total $PM_{2.5}$ and $PM_{2.5}$ speciation. UI-WRF-Chem is specialized by its application of NASA GEOS-5 data to provide both meteorological and chemical initial and boundary conditions for performing standard WRF-Chem forecast and was developed as the CTM to support the MAIA satellite mission (Fast et al., 2006; Grell et al., 2005). We set up the UI-WRF-Chem simulation as a nested domain with a $12 \times 12 \text{ km}^2$ horizontal spatial resolution for the outer domain and a $4 \times 4 \text{ km}^2$ horizontal spatial resolution for the inner domain focusing on the MAIA target areas. Both outer and inner domains have 47 vertical levels from surface to 50 hpa. Here, we conducted UI-WRF-Chem simulation from January 1 through December 31, 2018. The model was initiated on December 25, 2017 and allowed to run 7 days as a spin-up period to allow the model results to equilibrate. We used the hourly UI-WRF-Chem outputs from the inner domain ($4 \times 4 \text{ km}^2$) for our study PTA, including total $PM_{2.5}$ mass, sulfate, nitrate, OC, EC, and dust. We adopted the simulations at the lowest model layer to better capture the near-surface air pollution, and we calculated the 24-h averages as daily means, which were further assigned to the $1 \times 1 \text{ km}^2$ grid using nearest neighbor method. Table S1 illustrates significant variability in the correlation between observed $PM_{2.5}$ and WRF-Chem simulations. The strongest correlation is observed with OC, while the weakest is with dust. More details about the UI-WRF-Chem can be found in the Supplementary S1.

2.2.4. Other predictors

To account for the influences of meteorological conditions and human activities, a number of spatially and temporally varying characteristics of the surface and atmosphere were considered in the MAIA models, including meteorological variables, Normalized Difference Vegetation Index (NDVI), population density, elevation, and road density. Hourly meteorological variables such as air temperature (in Kelvin), planetary boundary layer height (in meters), relative humidity (in percentage), and surface wind speed (in meters/s) at $4 \times 4 \text{ km}^2$ resolution were obtained from the UI-WRF-Chem simulation. They were similarly processed for each $1 \times 1 \text{ km}^2$ grid as UI-WRF-Chem $PM_{2.5}$ simulations. Monthly NDVI data was obtained from the MAIA Ancillary Geographic Product (AGP), derived from multi-year averages of the MODIS/Terra 16-Day L3 global 250 m vegetation index product (MOD13Q1 V006). MOD13Q1 has a spatial resolution of 250 m, so 16 raster grids were aggregated to calculate the NDVI for each $1 \times 1 \text{ km}^2$ grid. We acquired the LandScan Global population data (people/ km^2) from 2018 at $1 \times 1 \text{ km}^2$ resolution (Rose et al., 2019). We used elevation data at 1 arc sec resolution (approximately 30 m) from ASTER Global Digital Elevation (DEM) Model Version 3. The original elevation data was aggregated at each $1 \times 1 \text{ km}^2$ grid and averaged. The roadway density (meters/ km^2) of primary and secondary roadways for each $1 \times 1 \text{ km}^2$ grid was calculated based on the 2018 TIGER/Line Shapefiles produced by the US Census Bureau. Several geospatial predictors are not currently generated by MAIA AGP in our PTA during 2018, including population, elevation, and roadway density, so we used alternative data sources instead.

2.3. Statistical methods

The MAIA L2 PM products are generated using a set of Bayesian hierarchical models (BHM), specifically one separate BHM per particle type (total $PM_{2.5}$ mass concentration, PM_{10} mass concentration, and the concentration of major $PM_{2.5}$ constituents including sulfate, nitrate, elemental carbon, organic carbon and dust) per target area. The L2 $PM_{2.5}$ retrievals depend on valid aerosol retrievals; therefore, gaps still exist due to the missing AOD values. In contrast, the CTM-based $PM_{2.5}$ retrievals rely on CTM simulations that have complete coverage. The L4 $PM_{2.5}$ product then takes a statistical data fusion approach called

Bayesian Model Averaging (BMA) using L2 $PM_{2.5}$ retrievals and CTM-based $PM_{2.5}$ retrievals as inputs to achieve full coverage in space and time (Diner et al., 2018). In this study, we followed the MAIA operational retrieval algorithm by building a BHM to generate the MAIA-like L2 $PM_{2.5}$ product (Fig. 2). Here, GOES-16 AOD retrievals and UI-WRF-Chem simulations were used to represent future MAIA AOD retrievals and CTM simulations. In this step, we also included a Bayesian statistical model called the Bayesian additive regression trees (BART) as a possible alternative to the BHM. BHM, BART, and BMA are all Bayesian statistical models that enable us to obtain the mean prediction and SD from posterior predictions. For the gap-filled $PM_{2.5}$ speciation products, we only produced the CTM-based results since GOES does not provide multi-angle images, which makes GOES unsuitable for speciated PM retrievals. Moreover, the ground-based $PM_{2.5}$ speciation observations are insufficient to generate the necessary model training dataset during our study period. The CTM-based $PM_{2.5}$ speciation estimates may serve as the lower bound of performance for the future MAIA L4 $PM_{2.5}$ speciation product. The details of our model development are provided below.

2.3.1. Bayesian spatial-temporal hierarchical model (BHM)

The BHM is a hierarchical model that considers the dependencies between observations across different spatial locations and scales and its detailed description has been published elsewhere (Chang et al., 2014; Geng et al., 2018; Murray et al., 2019). This approach allows for a smooth and stable representation of the spatial patterns in the data, as well as the incorporation of prior knowledge and information from other sources. As the operational MAIA PM product algorithm, the BHM model has several technical advantages. Firstly, it provides uncertainty quantifications for PM predictions. Secondly, it allows for flexibility in accounting for different data availability across targets (Banerjee et al., 2014). The data for certain monitors may be limited or incomplete, making it difficult to obtain accurate estimates for those monitors. BHM can address this issue by allowing for different levels of data availability across different monitors, but this flexibility may also cause the model difficulties in estimating the model parameters accurately (Banerjee et al., 2014). In this study, we designed two BHM models to allow the integration of ground-based observations with either the UI-WRF-Chem simulations or GOES AOD retrievals as the main predictor. Our BHM model can be expressed as follows:

$$Y_{st} = \alpha_{st} + \beta_{st}X_{st} + \varepsilon_{st} \quad (1)$$

where Y_{st} is the measured concentrations of $PM_{2.5}$ mass or speciation at site s on day t . X_{st} is the main predictor value at site s on day t , which can be either GOES AOD value or UI-WRF-Chem simulations for estimating the L2 $PM_{2.5}$, Y_{st}^{L2} , and the CTM-based $PM_{2.5}$, $Y_{st}^{CTM-based}$, respectively. α_{st} and β_{st} represent the spatial-temporal random intercepts (additive bias) and random slopes (multiplicative bias) that are assumed to be day-specific and site-specific, respectively. ε_{st} is the residual error term that is assumed to be independent and normally distributed with a mean of zero and variance σ^2 .

The spatial-temporal regression coefficients, α_{st} and β_{st} , are obtained from two second-level regression models as follows:

$$\alpha_{st} = \alpha'_s + \alpha'_t + \gamma Z_{st} \quad (2)$$

$$\beta_{st} = \beta'_s + \beta'_t \quad (3)$$

where α' and β' are the unobserved correlation random effects accounting for the solely spatial and solely temporal trends in the intercepts and the slopes. The fixed-effect regression coefficient, γ , is associated with the Z_{st} vector, which contains additional spatial and spatiotemporal predictors including meteorological variables, NDVI, population density, elevation, and roadway density.

2.3.2. Bayesian additive regression trees (BART)

MAIA offers a range of products besides the final Level 4 PM product,

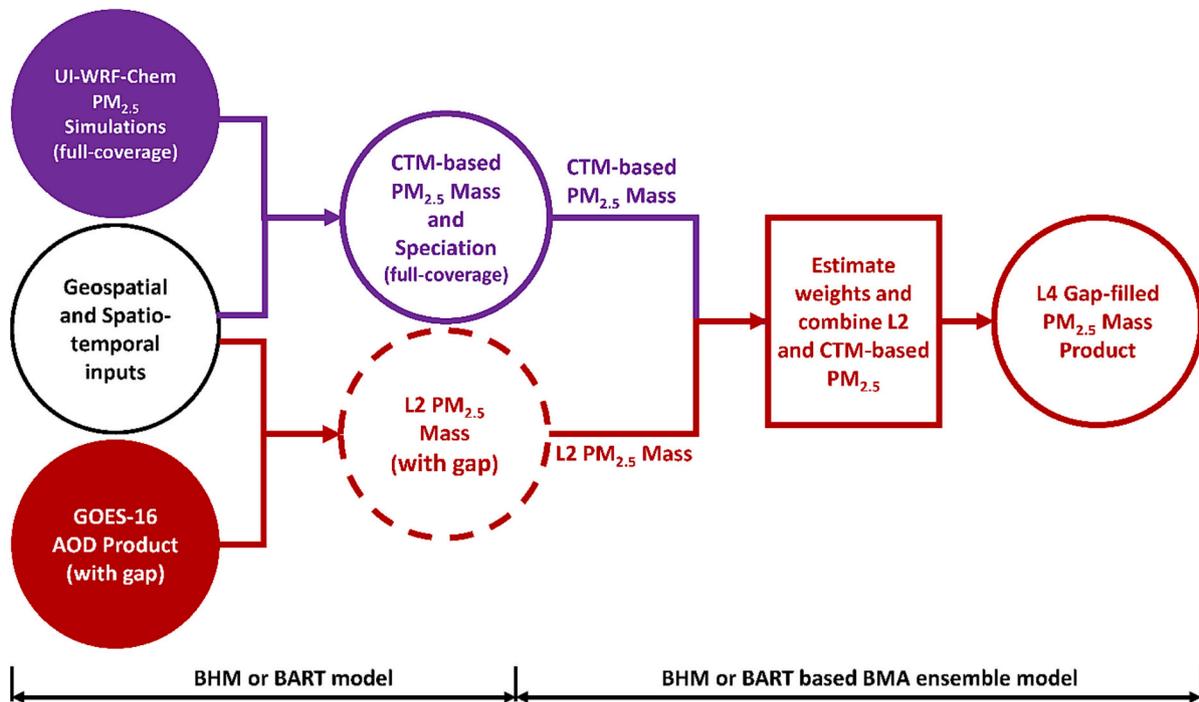


Fig. 2. Framework of the MAIA-like Level 4 gap-filled PM products.

including AOD retrievals products, Level 2 and CTM-based PM products. Consequently, many studies may not solely use the level 4 product but leverage other MAIA products to develop their own PM products. Furthermore, the BMA ensemble framework enables MAIA users to combine any number of Bayesian statistical models and models that are unable to generate prediction uncertainties. In this context, we used another Bayesian inference method, BART, as an alternative Bayesian model to test the model extensibility of BMA as well as the predictive ability of BART. BART combines the notion of sum-of-trees ensemble framework and Bayesian inference (Chipman et al., 2010). These concepts provide BART with the ability to capture the non-linear relationship and complicated interactions among predictors with high-dimensional data. Specifically, let Y represent the output corresponding to the input p -dimensional vector input ($X = x_1, x_2, \dots, x_p$). A BART model with m trees can be expressed as:

$$Y = \sum_{j=1}^m T_j(M_j; X) + \varepsilon \quad (4)$$

where $T_j(M_j; X)$ denotes a single decision tree of BART that is characterized by the tree structure (T_j) and a set of terminal nodes (M_j , also known as leaves) that are determined by the predictor vector X . Each decision tree includes a root and two node sets (internal nodes and terminal nodes). Internal nodes are formed by splitting decision rules based on a single predictor, $x_i \leq c$ or $x_i > c$, where c is the threshold and x_i is the splitting variable. The splitting process continues until reaching a terminal node and an observation value $\mu_{ij} \in M_j = \{\mu_{1j}, \mu_{2j}, \dots, \mu_{bj}\}$ is assigned to the terminal of tree j with b terminal nodes. ε is the error term following a normal distribution with variance σ^2 .

The BART algorithm, which is based on the Markov Chain Monte Carlo (MCMC), can be fitted and generate all the possible predictions from its corresponding posterior probability distribution. In this study, the hyperparameters $m, \alpha, \beta, k, v, q$ were optimized through grid search in spatial cross-validation to strike the best balance between accuracy and uncertainty. The R package *BART* was used to fit the BART model. More details about the BART model's structure and hyperparameters can be found in the Supplementary S2.

2.3.3. Bayesian model averaging (BMA) ensemble model

The total $PM_{2.5}$ mass concentrations of AOD-based L2 products and CTM-based estimates are combined using a Bayesian Model Averaging (BMA) ensemble framework, which uses the Markov Chain Monte Carlo (MCMC) approach to obtain the weights at each monitoring location (Raftery et al., 2005). The BMA model fills the gaps in L2 $PM_{2.5}$ mass and incorporates the predictive power of base models (e.g., BHM, BART) to improve the final predictions. As shown in Eq. (5), a Beta (1,1) prior is assumed on each monitoring station's weight, w_s , which is then updated using a random variable with Bernoulli distribution at each iteration (Raftery et al., 2005). The median value of the chain of values obtained after all iterations is used as the final weight. We similarly applied the framework (Fig. 2) to fuse the L2 and CTM-based $PM_{2.5}$ mass predictions from Bayesian models (BHM or BART) developed with GOES-16 AOD or UI-WRF-Chem simulations as the main predictor. The weights for L2 and CTM-based $PM_{2.5}$ mass estimates are calculated using the BMA ensemble framework for each monitoring stations and interpolated to grid cells without available monitors using a simple inverse distance weight (IDW).

We built the following model for final prediction of L4 $PM_{2.5}$ mass:

$$Y_{st}^{L4} = w_s Y_{st}^{CTM-based} + (1 - w_s) Y_{st}^{L2} \quad (5)$$

Where $Y_{st}^{CTM-based}$ and Y_{st}^{L2} are posterior means of CTM-based and L2 total $PM_{2.5}$ mass estimates obtained from the UI-WRF-Chem simulation and the GOES-16 AOD at site s on day t , respectively. w_s is the optimized weight for the UI-WRF-Chem downscaler at site s . Separate BMA ensemble models were built for either BHM modeled or BART modeled pairs of CTM-based and L2 $PM_{2.5}$ mass.

Similarly, the SD for $Y_{st}^{ensemble}$ is defined as

$$\hat{\sigma}_{st}^{ensemble} = \sqrt{(w_s \sigma_{st}^{CTM-based})^2 + ((1 - w_s) \sigma_{st}^{L2})^2} \quad (6)$$

which enables the calculation of uncertainties and inferences through the Bayesian ensemble model.

2.4. Model performance evaluation

We conducted a two-stage cross-validation (CV) to evaluate the model performance in predicting PM_{2.5} concentration at grid cells where both monitoring observations and UI-WRF-Chem/GOES-16 AOD were available. The CV was conducted in two types: traditional CV and spatial CV. The spatial CV enabled us to assess model's ability to predict PM_{2.5} in regions without monitoring networks.

In the first stage, the entire dataset was divided into a specific number of groups according to different CV types. At each iteration, one group was randomly selected as the testing dataset, while the rest served as the training dataset for building the BHM or BART model. The model was subsequently used in generating the fitted value from the testing dataset. This process was repeated after all iterations. In the second stage, because large data gaps existed in the GOES-16 AOD, we filtered the CV results from the first stage to include only grid-day records with

both UI-WRF-Chem and GOES-16 AOD data available. This process enables us to better reflect the model's predictive performance in real-world situations. The filtered pairs of L2 and CTM-based PM_{2.5} mass estimations were then divided into testing datasets and training datasets and used to build and test the BMA ensemble model. The ensemble weights are also computed and interpolated at each iteration. The traditional CV was conducted by randomly dividing the dataset into 10 folds, while leave-one-station-out is used for the spatial CV in evaluating PM_{2.5} speciation prediction due to fewer stations, and 10-fold CV divided by stations was used for PM_{2.5} mass. The CV results were evaluated in terms of accuracy and uncertainty using multiple performance metrics, including coefficient of determination (R²), the root-mean-square error (RMSE), slopes, the coverage rate of 95% prediction CI, and prediction SD. The 95% CI of predictions was derived by the 2.5th and the 97.5th quantiles of posterior samples, which are based on the mean values and the SD measure of posterior predictions. Therefore, the

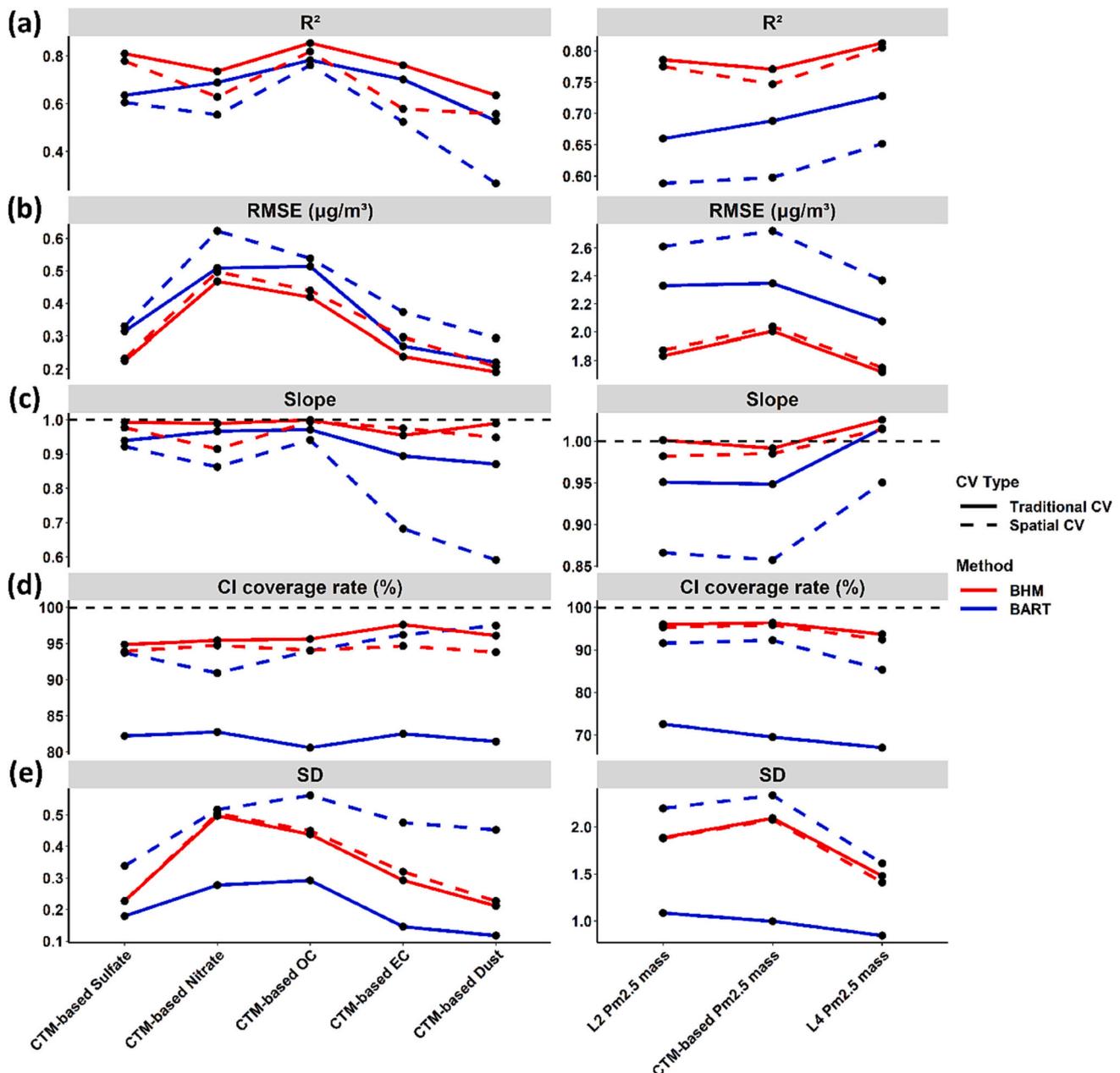


Fig. 3. Traditional and spatial Cross-validation (CV) results for BHM and BART models. (a), R² (b), Root-mean-square error (RMSE, µg/m³) (c), Slope (d), 95% Confidence interval (CI) coverage rate (%) (e), Standard deviation (SD) of posterior predictions. Note: this figure is subtitled by rows.

coverage rate of 95% prediction CI measures both prediction accuracy and uncertainties. For example, a coverage rate of 90% means 90% of the 95% CIs of posterior predictions captured the ground-based observations.

3. Results

3.1. Model performance evaluation

Fig. 3 displays the traditional and spatial CV results from BHM and BART models. The BHM generally showed better accuracy compared to

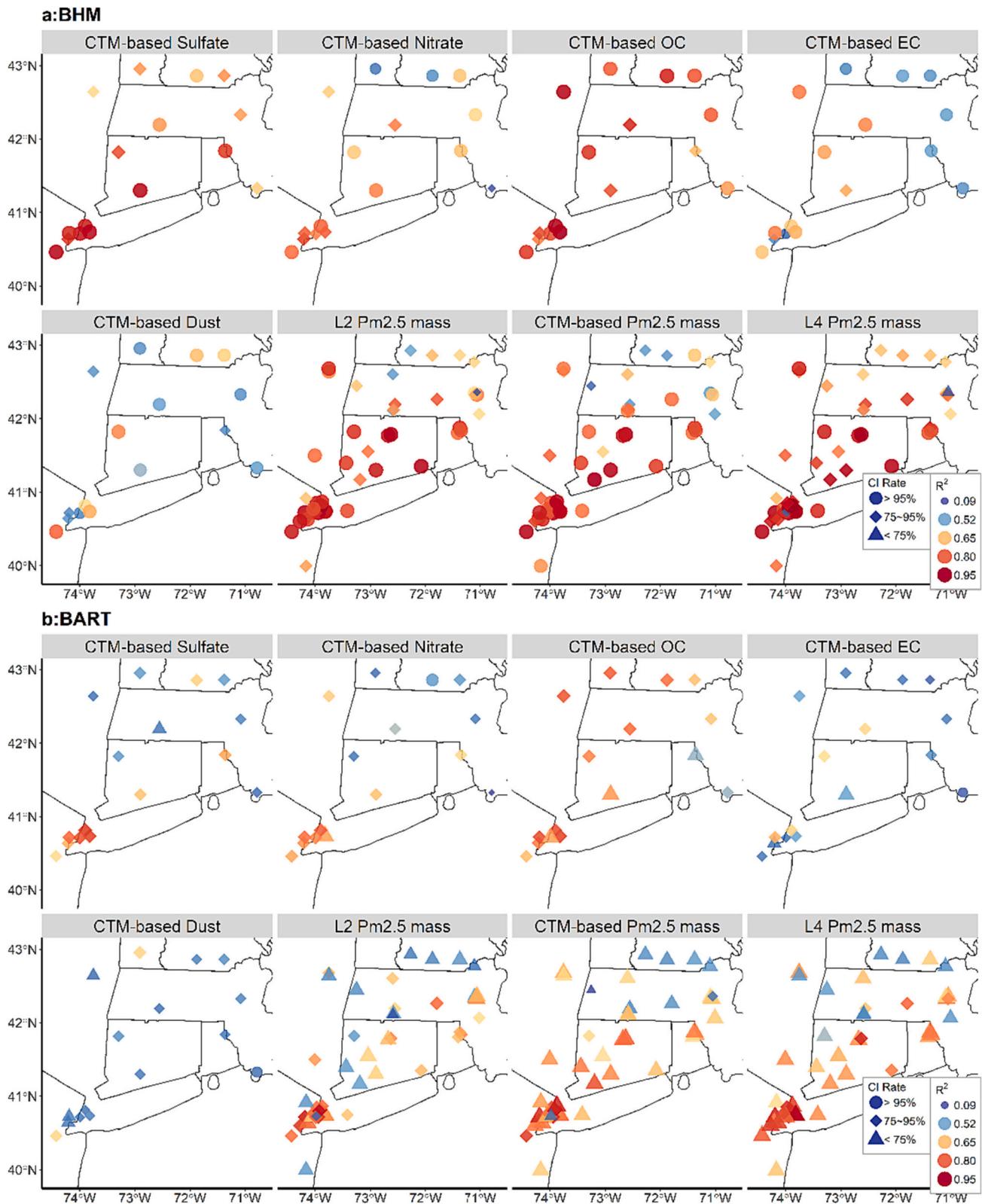


Fig. 4. Maps of traditional CV R^2 and confidence interval (CI) coverage rate at $PM_{2.5}$ speciation and mass stations. (a), BHM (b), BART.

the BART model, with higher R^2 (Fig. 3a) and slightly lower RMSE (Fig. 3b). The slope plot (Fig. 3c) illustrates that the BART models tend to underestimate the concentrations. The spatial CV results also indicate that the BHM model has a better predictive ability at locations without monitoring data. Both BHM and BART performed differently for different $PM_{2.5}$ species, with both exhibiting the best performance for OC (BHM [traditional CV $R^2 = 0.85$]; BART [traditional CV $R^2 = 0.78$]) and the worst performance for dust (BHM [traditional CV $R^2 = 0.64$]; BART [traditional CV $R^2 = 0.53$]). The BMA ensemble model demonstrated improved performance in predicting L4 $PM_{2.5}$ mass, with a traditional CV R^2 of 0.81 and 0.73, which outperformed their input modeled $PM_{2.5}$ mass by BHM (L2 [$R^2 = 0.79$]; CTM-based [$R^2 = 0.77$]) and BART (L2 [$R^2 = 0.66$]; CTM-based [$R^2 = 0.69$]). The BART model consistently showed a lower coverage rate of 95% prediction CI (Fig. 3d), which means that prediction CIs from BART are less likely to contain the true values. The higher coverage rates of BHM models also imply wider prediction CIs caused by higher SD (Fig. 3e), suggesting that CI coverage rate alone does not provide a complete picture of the model's performance. Comparing the results between traditional CV and spatial CV, BHM performed more robustly among the two CV experiments, while BART models showed worse spatial CV performance as measured by R^2 , RMSE, slope, and SD. The higher SD, on the contrary, significantly improve the CI coverage rates of BART in spatial CV. For details of CV results, please refer to Table S2.

To evaluate the spatial variation of model performance, we further analyzed the CV results at specific monitor locations. Both BHM and BART models' results for L2, CTM-based, and L4 products had higher R^2 in the southern part of the study domain (Fig. 4), especially in New York City (NYC) with a denser monitor network. Among different $PM_{2.5}$

speciation, EC and dust showed a generally weaker accuracy than the other speciation, especially in the southern urban centers. However, the difference of CI coverage rates across space was not significant for different $PM_{2.5}$ products. BHM had higher R^2 (mean $R^2 = 0.72$) and CI coverage rates (mean $R^2 = 0.95$) than BART (mean $R^2 = 0.61$; mean CI coverage rate = 0.74) at the station level (Fig. 4), as seen in Fig. 3. For stations that are sparsely located outside the NYC area, their performance is acceptable in both traditional and spatial CV experiments (Fig. 4 and Fig. S2), but not as robust as that of the stations densely situated within the NYC region. This suggests that stations which are densely distributed offer a more precise reflection of $PM_{2.5}$ levels in a particular area compared to those spaced further apart.

The performance of the models varied by season (Fig. 5), with generally better prediction accuracy (lower RMSE) during autumn, when PM concentrations were lower. There was no obvious seasonal pattern for the BHM in terms of the CI coverage rate, but the rates were lower during summer and winter for BART. We further investigated model performance at different concentrations intervals (Fig. S3), and the results consistently indicated the worst performance (highest RMSE and lowest coverage rate) in predicting the highest 10% values of each $PM_{2.5}$ product, which may explain the unsatisfactory results during seasons with higher pollutants concentration.

3.2. $PM_{2.5}$ predictions

Fig. 6a and b depict the annual L4 $PM_{2.5}$ mass predictions from the BHM-based and BART-based BMA ensemble models in the study domain, respectively. The results of both models show higher concentrations in NYC, Boston, and coastal cities. The BART-based BMA tends

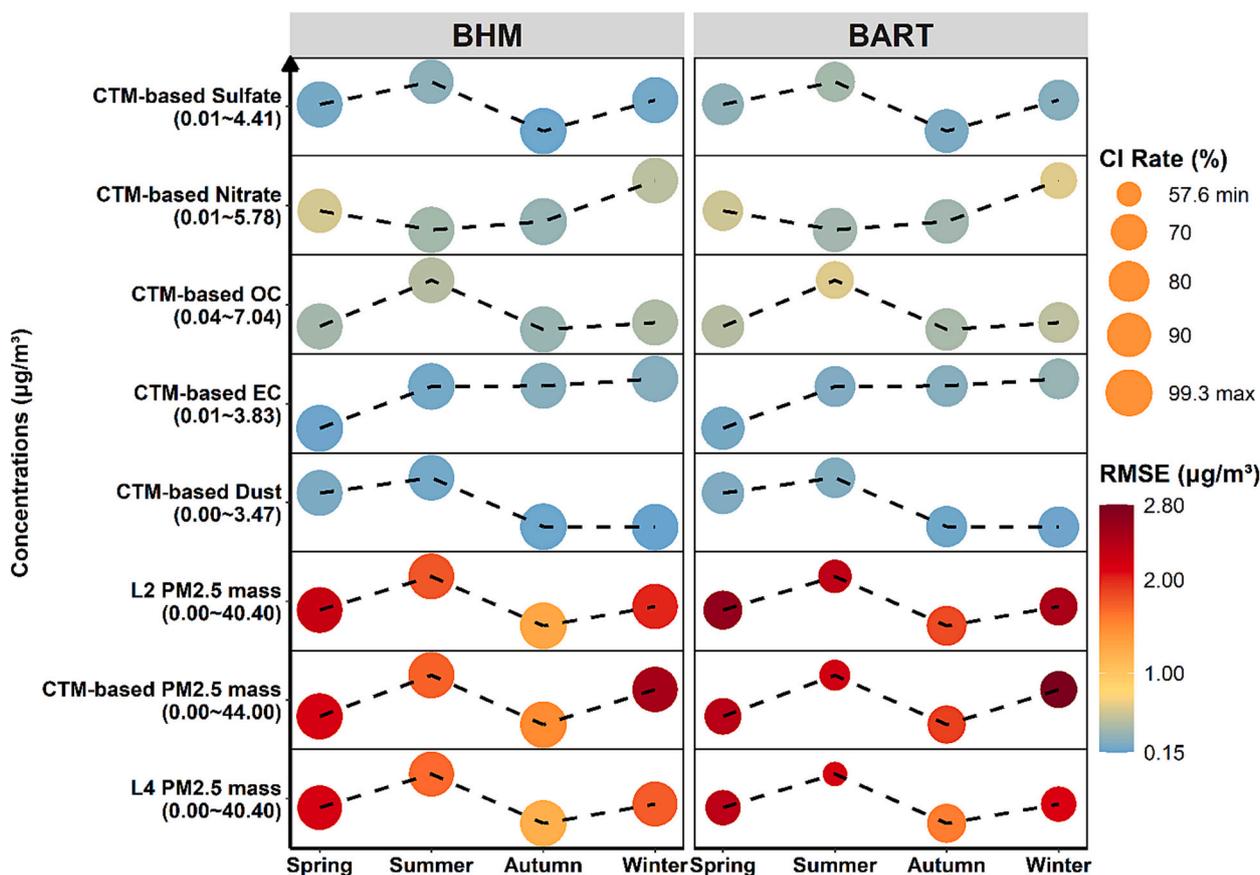


Fig. 5. CV RMSE and CI coverage rate by seasons. Note: The relative heights of the four seasons in each separated box of the plot refer to the concentration difference for each $PM_{2.5}$ product. The overall range of concentrations is labeled below each pollutant's name, corresponding to the seasons with the lowest and highest concentrations in each box. The seasons were defined as follows: spring (March–May), summer (June–August), autumn (September–November), and winter (December–February).

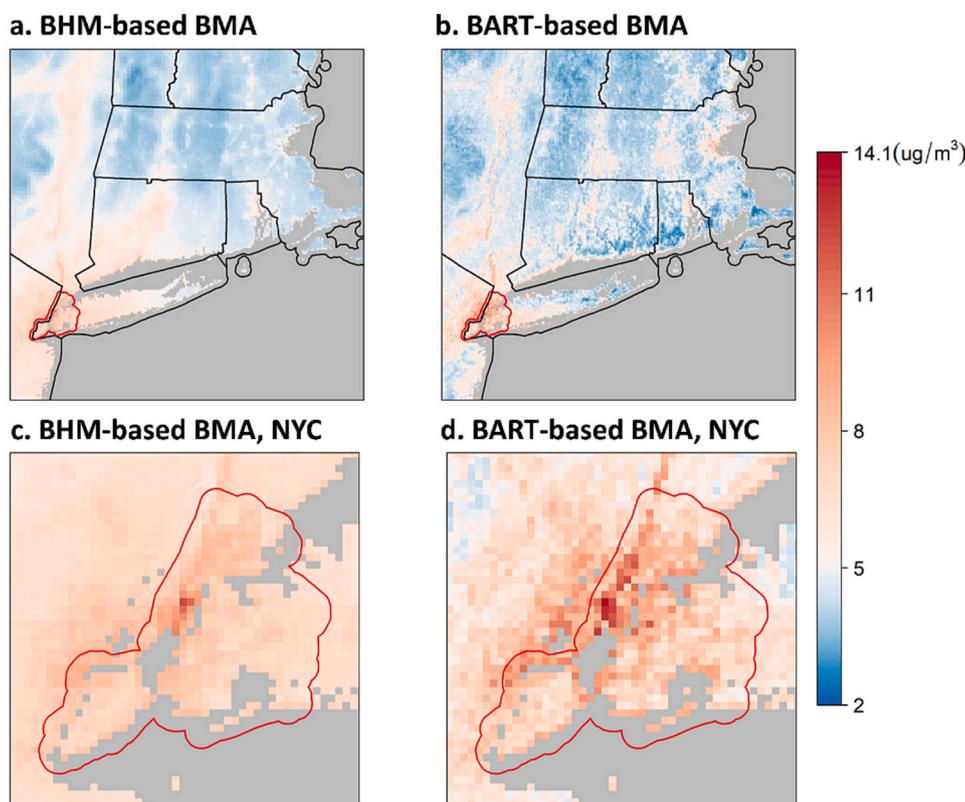


Fig. 6. Mean prediction results of L4 $PM_{2.5}$ mass in 2018. (a), Mean $PM_{2.5}$ mass concentrations predicted by BHM-based BMA model (b), Mean $PM_{2.5}$ mass concentrations predicted by BART-based BMA model (c), Mean $PM_{2.5}$ mass concentrations in NYC predicted by BHM-based BMA model (d), Mean $PM_{2.5}$ mass concentrations in NYC predicted by BART-based BMA model.

to be less spatially smooth compared to the BHM-based BMA, which was also illustrated in terms of the spatial variation of prediction uncertainties (Fig. S4). The SD of predictions was clearly higher along urban center, highways and Hudson River for BART model (Fig. S4b and S4d), while the BHM-based BMA displayed more consistent prediction uncertainties across the study domain (Fig. S4a), despite a slight increase in Manhattan (Fig. S4c). The same spatial patterns of concentrations and differences between the two models are also evident in $PM_{2.5}$ speciation predictions (Fig. S5). Higher pollution levels of all CTM-based $PM_{2.5}$ speciation were estimated in NYC, Albany, Boston, Providence, coastal areas in New Jersey and Connecticut, areas along the Hudson River, and along the city lines of New Haven, Hartford, and Springfield, while Vermont and New Hampshire exhibit the lowest concentration level. Prediction SD also showed larger uncertainties in regions with high estimated concentrations. Among different $PM_{2.5}$ components, nitrate and dust disseminated further northward within New York State. The BART model had larger spatial variation for predictions and SD, and showed drastic higher uncertainties at highways and water bodies than peripheries. In contrast, BHM model showed a broader dissemination of the aerosol with high concentrations covering more rural areas than the BART models. OC showed the least spatial difference between BHM and BART models.

As an independent validation, we compared our predictions with $PM_{2.5}$ measurements of the New York City Community Air Survey (NYCCAS) network. The NYCCAS includes 93 stations (Fig. S6) and provided bi-weekly mean $PM_{2.5}$ mass and sulfur measurements. Both BHM and BART models accurately depicted the substantial differences in $PM_{2.5}$ mass concentrations between urban and rural areas in NYC (Fig. 6c and d). Furthermore, they captured the temporal variation within 2018 (Fig. 7a) and aligned with the real-world observations from NYCCAS (Fig. S7). The 95% CI of the BART model was slightly wider than the BHM, yet the NYCCAS observations were covered within the

95% CIs of both models. We estimated the sulfate (SO_4^{2-} , molecular mass: 96 g/mol) concentration by multiplying NYCCAS sulfur (S, molecular mass: 32 g/mol) by three, based on the assumption that all sulfur is fully oxidized in the form of sulfate (Brown et al., 2002). Despite underestimation due to the assumption, both BHM and BART models the BHM and BART models adequately capture the temporal variation (Fig. 7b), and the 95% CI of both models covered the transformed sulfate of NYCCAS.

4. Discussion

In our study, we followed the MAIA modeling framework and applied the BHM and BART models to estimate the L2 and CTM-based daily average $PM_{2.5}$ mass and speciation at a $1 \times 1 \text{ km}^2$ spatial resolution for the MAIA Northeastern US PTA, which encompasses New York City and Boston in 2018. To fill the gaps in the L2 $PM_{2.5}$ mass products, we used the BMA ensemble models to integrate the L2 and CTM-based $PM_{2.5}$ mass predictions. The proposed modeling framework effectively captures the spatiotemporal variation of $PM_{2.5}$ and quantifies the prediction uncertainties, thereby reducing potential exposure misclassification in future health studies. Previous studies have primarily focused on estimating total $PM_{2.5}$ mass using various statistical and machine learning models such as extreme gradient boosting ($CV R^2 = 0.87$) (Just et al., 2020), Gaussian Markov Random Field ($CV R^2 = 0.83$) (Sarafian et al., 2019), and mixed effects model ($CV R^2 = 0.88$) (Kloog et al., 2014). In our study, we employed a BHM-based BMA ensemble model, which achieved slightly lower prediction accuracy ($CV R^2 = 0.81$), potentially due to fewer predictors in the operational MAIA algorithm and fewer monitors for training dataset in our study domain. Less attention has been paid to estimating $PM_{2.5}$ composition and most published studies focused on California and the Northeastern US, which have denser $PM_{2.5}$ speciation monitoring networks. For example, Geng

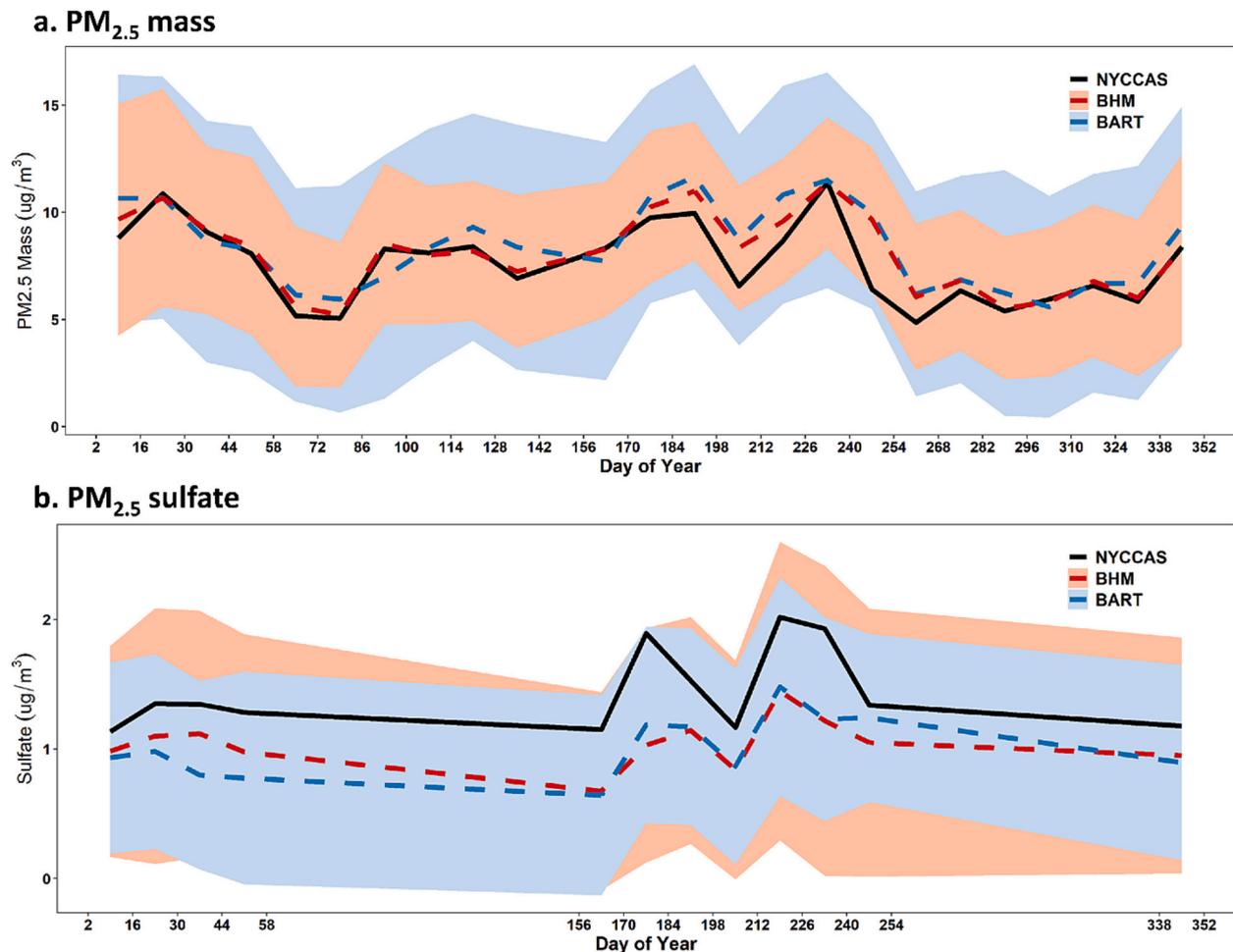


Fig. 7. Comparison between NYCCAS observations and predictions with confidence intervals. (a), L4 total PM_{2.5} mass product (b), CTM-based sulfate estimate. Note: Observations from different sites have been averaged. The x-axis represents the day of year corresponding to the bi-weekly dates of available NYCCAS records.

et al. used MISR AODs and CMAQ simulations in random forest models to estimate sulfate, nitrate, OC and EC levels in California over 2005–2014 (Geng et al., 2020). Di et al. calibrated GEOS-Chem simulations using a backward propagation neural network in the Northeastern US, achieving mean CV R^2 of 0.81, 0.83, 0.69, 0.71, and 0.42 for sulfate, nitrate, OC, EC, and dust (Di et al., 2016), which were lower than our results except for nitrate. The inclusion of spatially and temporally lagged terms in Di et al. may not be practical for most MAIA PTAs, which typically have limited monitor coverage and less frequent measurements of PM_{2.5} composition. Our study also revealed variations in calibration performance among major PM_{2.5} components, with the poorest performance observed in dust. This might be attributed to the fact that the WRF-Chem dust simulation have the poorest agreement with observations (Table S1). Nonetheless, with the utilization of actual MAIA AOD retrievals, we anticipate that the operational L4 PM_{2.5} speciation estimates will have better quality than our simulated results.

There is a scarcity of studies in the air pollution exposure modeling literature that have addressed PM_{2.5} prediction uncertainty, primarily because few reported techniques are capable of quantifying prediction uncertainty. Murray et al. explored the use of BHM and Bayesian ensemble model to estimate total PM_{2.5} mass in the Southeastern US (Murray et al., 2019). Their ensemble model showed improved predictions with an R^2 of 0.83 and a 95% CI coverage rate of 97%. Our BHM-based BMA ensemble model produced similar results, with an R^2 (0.81) and a 95% CI coverage rate of 94%, while the results of our BART-based BMA ensemble model were inferior. We have observed a slight reduction in the coverage rates of prediction CIs in our BMA ensemble

model, despite the improvement in prediction accuracy, which differs from the results reported by Murray et al. This might be attributed to Eq. (4), where it is typical to achieve lower SD and coverage rate of prediction CIs than those of the base models. The complex relationship and interactions between L2 PM products and CTM-based PM estimates with varying degrees of accuracy and uncertainty may influence the calculation of ensemble weight and contribute to this effect as well.

We further investigated our weight results (Fig. S8) and found that the CTM-based PM_{2.5} mass from the BHM-based BMA was assigned a higher weight value, as also found in Murray's paper, while the BART-based BMA favors L2 PM_{2.5} mass. The BART-based weights show less spatial variation than the BHM-based weights, which display higher weights in the northern part of the PTA where monitors are sparsely distributed, corresponding to rural areas. To understand the sources of the weights' difference, we analyzed the associations between weights and prediction accuracy (R^2) and uncertainties (SD) (Fig. S9 and Fig. S10). The results of both BHM and BART implied that the PM product with higher accuracy at site s will receive a higher weight than the other PM_{2.5} mass product (Fig. S9). A larger difference between the accuracy of PM_{2.5} mass products will also result in a larger difference in the weights (Fig. S10). However, the prediction uncertainties play a limited role in determining the weights. The BMA ensemble model favors the PM product with higher accuracy even though it may have larger uncertainties as well.

Regarding the prediction uncertainties of PM_{2.5} components, Zhang et al. used BART to model PM_{2.5} components, including sulfate, nitrate, OC, and EC, in California during 2005–2014 (Zhang et al., 2020b). By

employing variable selection and including CTM $PM_{2.5}$ simulation as input, their BART model demonstrated good predictive ability (R^2 ranging from 0.78 to 0.84) and maintained proper CI coverage rates of at least 95%, outperforming the BART models in our study. BART is constructed with various tuning parameters, such as tree structure tree T_j and terminal nodes M_j , which promote smaller trees and prevent overfitting (Chipman et al., 1998, 2010). Additionally, BART has the advantage of selecting variables that appear most frequently in the fitted sum-of-trees models when the number of trees is small. Both features of BART enable it to handle datasets with high-dimensional predictors. However, we did not incorporate fractional AOD component data in our models' inputs, which reduced our input dimensions and impacted our models' prior hyperparameters. We tried different settings of hyperparameters for each PM products during CV-based tuning, but the tuning results (Table S3) consistently tended to construct models with larger sum-of-trees with deeper depth. This suggests complex relationships in the dataset requires more complex BART model settings. In addition, our study only includes observations from 16 $PM_{2.5}$ speciation monitors in 2018, which is approximately 3 times fewer than Zhang et al.'s study in California from 2005 to 2014. A smaller dataset was used to train the BART models, which may also be responsible for the model complexity. Our findings suggest that the BART model is not able to leverage its advantages but goes against the philosophy of BART under our study design, implying its overfitting and inability to model $PM_{2.5}$ speciation in data-poor PTAs. This also helps explain the sensitivity of BART to different land covers and the worse performance in spatial CV experiments (Fig. 3), which is caused by the lack of generalizability of the overfitted BART and the fact that BART algorithm does not explicitly consider dependencies across different locations.

Our study found that the BHM models outperformed BART models in estimating both total $PM_{2.5}$ mass and speciation. Despite limited

monitoring data for $PM_{2.5}$ species and low-dimensional inputs, the BHM model showed its simplicity and efficiency. It has the potential to be applied in data-poor areas such as the developing countries with sparse $PM_{2.5}$ monitoring networks. Our PTA in northeastern US has few monitors, yet it is already richer than most regions of the world, highlighting that most other MAIA PTAs and Secondary Target Areas suffer from an even greater scarcity of monitors for $PM_{2.5}$ speciation modeling. We further performed sensitivity analyses to evaluate the model's performance under data-poor scenarios by removing monitors from the training dataset. The results (Fig. 8) showed that the BHM model collapses when the number of available monitors is less than four. In addition, it was found that nitrate, EC, and dust require more monitors to achieve an acceptable performance ($R^2 > 0.6$) compared to sulfate and OC. This highlights the challenges that data-poor regions may face when trying to estimate $PM_{2.5}$ speciation, particularly for nitrate, EC, and dust.

Our study has several limitations that are worth noting. Firstly, $PM_{2.5}$ speciation monitors are very sparse within our study domain and there is a considerable amount of missing data in the GOES AOD data. The lack of larger training dataset may limit the model performance, especially for BART models, in estimating $PM_{2.5}$ speciation. This is expected to improve with the use of MAIA, which will have multi-angle capabilities and higher spatial resolution. Secondly, the input covariates were designed in a simple manner with fewer predictors, and the satellite AOD data or CTM simulations of $PM_{2.5}$ mass were not included in the $PM_{2.5}$ speciation modeling. $PM_{2.5}$ -AOD relationships are complicated and have obvious spatiotemporal heterogeneities (Ma et al., 2022). We expect the relationships will be better captured by a large-scale model that incorporates more AOD information provided from multi-angle satellites instruments such as MISR and MAIA. Thirdly, we used a simple IDW function to interpolate and extrapolate the ensemble weights,

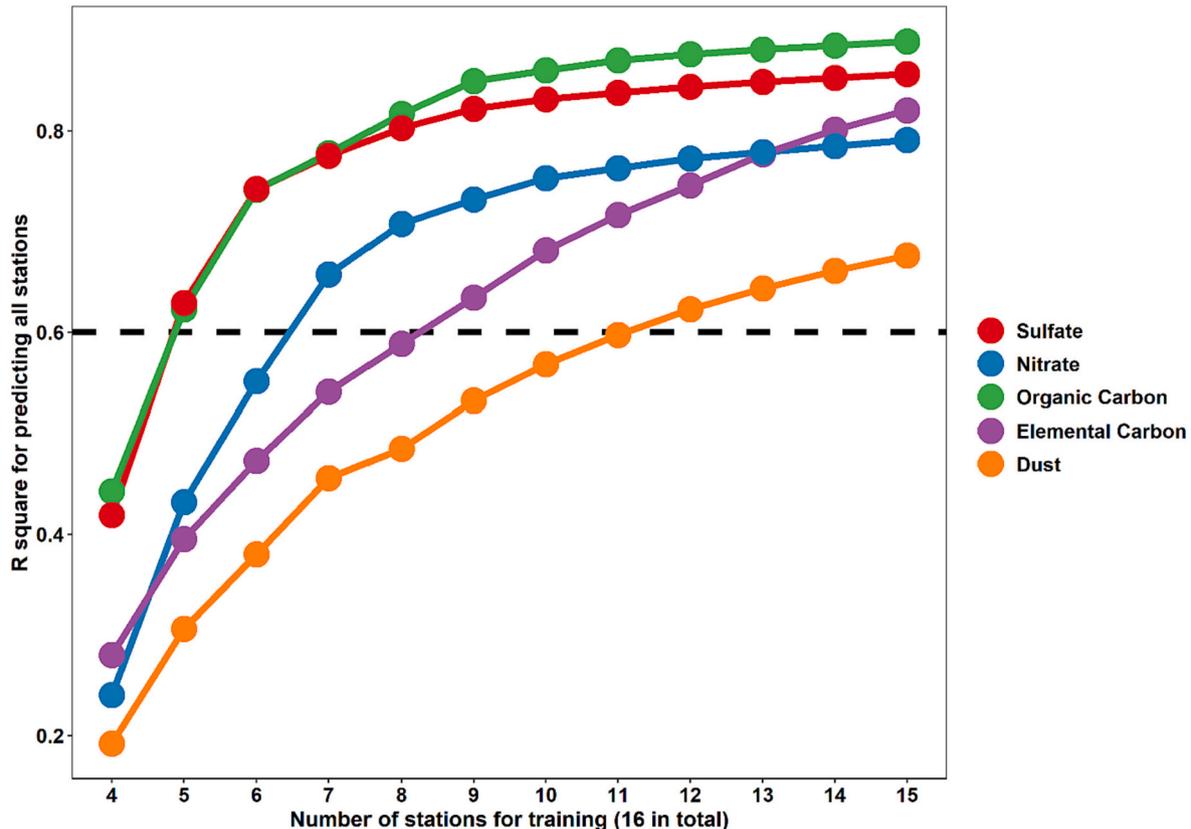


Fig. 8. Sensitivity analyses of BHM models' performance in predicting CTM-based $PM_{2.5}$ speciation with different numbers of stations used in model training. Note: For each number of stations, all possible combinations were listed and up to 500 combinations were randomly selected. The R^2 results from selected combinations were averaged.

which may introduce additional uncertainties. More appropriate interpolation methods, such as kriging and splines interpolation, could be tested and used instead. However, the choice of the best interpolation method will depend on the specific characteristics of the data. Moreover, the contribution of better interpolation methods may be limited in data-poor areas with extremely few monitors. Fourthly, our BHM models collapse when the available monitors for PM_{2.5} speciation are less than four, which limits the applicability of BHM model over data-poor regions and small countries that do not even have four monitors. Therefore, A simpler BHM model that does not consider spatial or temporal random effect is needed. Finally, MAIA will retrieve fractional AODs along with other aerosol properties, which will be used as additional predictors in the PM product retrieval algorithm. Our study utilized GOES-16 AOD at 550 nm as the proxy of MAIA given the data limitation. The GOES-16 AOD retrievals do not capture the information in particle microphysical properties provided by fractional AODs. The current WRF-Chem simulations also show weaker correlations with certain PM_{2.5} species such as nitrate and dust (Table S1). However, our study results establish a performance benchmark for the MAIA mission. Once the actual MAIA AOD retrievals are incorporated, the operational PM_{2.5} products are expected to surpass the quality of the results presented in the current study.

5. Conclusions

Our study follows the MAIA modeling framework and provides a reliable estimation of MAIA-like PM_{2.5} mass and speciation with good control over prediction uncertainty. We previewed and validated the feasibility of the MAIA framework and variables in producing L2, CTM-based and L4 PM products. BHM models exhibited greater predictive ability despite a limited number of monitors and low-dimensional input predictors, making it applicable in developing countries with poorly constructed monitoring networks. On the other hand, BART models favor larger datasets with a larger number of input predictors to avoid complex sum-of-trees structures, which may limit its usage in data-poor PTAs. The BMA ensemble models showed improved prediction performance of L4 PM_{2.5} mass compared to their base models of BHM and BART using only UI-WRF-Chem or GOES-16, suggesting it is beneficial in spatial gap-filling by incorporating L2 and CTM-based PM products. The uncertainty measures can be used in cost-benefit assessments and epidemiological studies to account for the bias of effect estimates due to exposure errors (Fann et al., 2018; Gryparis et al., 2009; Johnson and Garcia-Menendez, 2022; Rappold et al., 2014). The study also validated the capability of MAIA framework for fusing satellite AOD data, CTM simulations, meteorological variables, and land-use information to produce gap-filled PM_{2.5} products. After the launch of MAIA aerosol instrument, more comprehensive satellite aerosol-based data will be used in the framework to provide improved PM products for health studies.

Credit author statement

Zhihao Jin: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing. **Qiang Pu:** Conceptualization, Data curation, Writing - original draft. **Nathan Janecek:** Conceptualization, Data curation, Writing - original draft. **Huanxin Zhang:** Conceptualization, Data curation, Writing - original draft. **Jun Wang:** Conceptualization, Data curation, Methodology. **Howard Chang:** Conceptualization, Data curation, Methodology. **Yang Liu:** Conceptualization, Data curation, Funding acquisition, Project administration, Supervision, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The work of Z. Jin and Y. Liu, was partially supported by the NASA (80NSSC19K0191, PI: Y. Liu) and the NASA MAIA investigation led by D. Diner (Subcontract #1558091). It has not been formally reviewed by NASA. We thank Kazuhiko Ito from the New York City Department of Health and Mental Hygiene, and Abigail Nastan from the Jet Propulsion Laboratory for providing the data. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the NASA. NASA does not endorse any products or commercial services mentioned in this publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2024.113995>.

References

- Amini, H., Danesh-Yazdi, M., Di, Q., Requia, W., Wei, Y., Abu-Awad, Y., Shi, L., Franklin, M., Kang, C.-M., Wolfson, J., 2022. Hyperlocal Super-Learned PM_{2.5} Components Across the Contiguous US. <https://doi.org/10.21203/rs.3.rs-1745433/v2>.
- Banerjee, S., Carlin, B.P., Gelfand, A.E., 2014. *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press.
- Box, G.E., Tiao, G.C., 2011. *Bayesian Inference in Statistical Analysis*. John Wiley & Sons.
- Brown, T.L., LeMay, H.E., Bursten, B.E., 2002. *Chemistry: The Central Science*. Pearson Educación.
- Chang, H.H., Hu, X., Liu, Y., 2014. Calibrating MODIS aerosol optical depth for predicting daily PM_{2.5} concentrations via statistical downscaling. *J. Expo. Sci. Environ. Epidemiol.* 24, 398–404. <https://doi.org/10.1038/jes.2013.90>.
- Chen, G., Li, S., Knibbs, L.D., Hamm, N.A.S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M.J., Guo, Y., 2018. A machine learning method to estimate PM_{2.5} concentrations across China with remote sensing, meteorological and land use information. *Sci. Total Environ.* 636, 52–60. <https://doi.org/10.1016/j.scitotenv.2018.04.251>.
- Chipman, H.A., George, E.I., McCulloch, R.E., 1998. Bayesian CART model search. *J. Am. Statist. Ass.* 93, 935–948. <https://doi.org/10.1080/01621459.1998.10473750>.
- Chipman, H.A., George, E.I., McCulloch, R.E., 2010. BART: Bayesian additive regression trees. *Ann. Appl. Stat.* 4, 266–298. <https://doi.org/10.1214/09-AOS285>.
- Chowdhury, S., Pillarisetti, A., Oberholzer, A., Jetter, J., Mitchell, J., Cappuccilli, E., Aamaas, B., Aunan, K., Pozzer, A., Alexander, D., 2023. A global review of the state of the evidence of household air pollution's contribution to ambient fine particulate matter and their related health impacts. *Environ. Int.* 107835 <https://doi.org/10.1016/j.envint.2023.107835>.
- 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 III, 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., 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. *Lancet.* 389, 1907–1918. [https://doi.org/10.1016/S0140-6736\(17\)30505-6](https://doi.org/10.1016/S0140-6736(17)30505-6).
- Crouse, D.L., Peters, P.A., Hystad, P., Brook, J.R., Donkelaar, A.V., Martin, R.V., Villeneuve, P.J., Jerrett, M., Goldberg, M.S., Pope, C.A., Brauer, M., Brook, R.D., Robichaud, A., Menard, R., Burnett, R.T., 2015. Ambient PM_{2.5}, O₃, and NO₂ exposures and associations with mortality over 16 years of follow-up in the Canadian census health and environment cohort (CanCHEC). *Environ. Health Perspect.* 123, 1180–1186. <https://doi.org/10.1289/ehp.1409276>.
- Di, Q., Koutrakis, P., Schwartz, J., 2016. A hybrid prediction model for PM_{2.5} mass and components using a chemical transport model and land use regression. *Atmos. Environ.* 131, 390–399. <https://doi.org/10.1016/j.atmosenv.2016.02.002>.
- Diner, D., Boland, S., Brauer, M., Bruegge, C., Burke, K., Chipman, R., Di Girolamo, L., Garay, M., Hasheminassab, S., Hyer, E., Jerrett, M., Jovanovic, V., Kalashnikova, O., Liu, Y., Lyapustin, A., Martin, R., Nastan, A., Ostro, B., Ritz, B., Schwartz, J., Wang, J., Xu, F., 2018. Advances in multiangle satellite remote sensing of speciated airborne particulate matter and association with adverse health effects: from MISR to

- MAIA. *J. Appl. Remote Sens.* 12, 042603 <https://doi.org/10.1117/1.JRS.12.042603>.
- Diner, D.J., Chang, H.H., Liu, Y., Wang, J., Hasheminassab, S., Garay, M.J., Ge, C., Hyer, E.J., Jerrett, M., Kalashnikova, O.V., McDuffie, J.L., Val, S., Wang, Y., Xu, F., 2019. MAIA Level 2 particulate matter and Level 4 Gap-filled particulate matter retrieval algorithm theoretical basis. In: *NASA Earth System Science Pathfinder Program*, JPL D-101518.
- Fann, N., Alman, B., Broome, R.A., Morgan, G.G., Johnston, F.H., Pouliot, G., Rappold, A.G., 2018. The health impacts and economic value of wildland fire episodes in the US: 2008–2012. *Sci. Total Environ.* 610, 802–809. <https://doi.org/10.1016/j.scitotenv.2017.08.024>.
- Fast, J.D., Gustafson Jr., W.I., Easter, R.C., Zaveri, R.A., Barnard, J.C., Chapman, E.G., Grell, G.A., Peckham, S.E., 2006. Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model. *J. Geophys. Res. Atmos.* 111 <https://doi.org/10.1029/2005JD006721>.
- Franklin, M., Kalashnikova, O.V., Garay, M.J., 2017. Size-resolved particulate matter concentrations derived from 4.4km-resolution size-fractionated multi-angle imaging SpectroRadiometer (MISR) aerosol optical depth over Southern California. *Remote Sens. Environ.* 196, 312–323. <https://doi.org/10.1016/j.rse.2017.05.002>.
- Garay, M.J., Witek, M.L., Kahn, R.A., Seidel, F.C., Limbacher, J.A., Bull, M.A., Diner, D.J., Hansen, E.G., Kalashnikova, O.V., Lee, H., Nastan, A.M., Yu, Y., 2020. Introducing the 4.4 km spatial resolution multi-angle imaging SpectroRadiometer (MISR) aerosol product. *Atmos. Meas. Tech.* 13, 593–628. <https://doi.org/10.5194/amt-13-593-2020>.
- Geng, G., Murray, N.L., Tong, D., Fu, J.S., Hu, X., Lee, P., Meng, X., Chang, H.H., Liu, Y., 2018. Satellite-based daily PM_{2.5} estimates during fire seasons in Colorado. *J. Geophys. Res. Atmos.* 123, 8159–8171. <https://doi.org/10.1029/2018JD028573>.
- Geng, G., Meng, X., He, K., Liu, Y., 2020. Random forest models for PM_{2.5} speciation concentrations using MISR fractional AODs. *Environ. Res. Lett.* 15, 034056 <https://doi.org/10.1088/1748-9326/ab76df>.
- Goddard, F.G.B., Chang, H.H., Clasen, T.F., Sarnat, J.A., 2020. Exposure measurement error and the characterization of child exposure to fecal contamination in drinking water. *npj Clean Water* 3, 19. <https://doi.org/10.1038/s41545-020-0063-9>.
- Grell, G.A., Peckham, S.E., Schmitz, R., McKeen, S.A., Frost, G., Skamarock, W.C., Eder, B., 2005. Fully coupled “online” chemistry within the WRF model. *Atmos. Environ.* 39, 6957–6975. <https://doi.org/10.1016/j.atmosenv.2005.04.027>.
- Gryparis, A., Paciorek, C.J., Zeka, A., Schwartz, J., Coull, B.A., 2009. Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics* 10, 258–274. <https://doi.org/10.1093/biostatistics/kxn033>.
- Hang, Y., Meng, X., Li, T., Wang, T., Cao, J., Fu, Q., Dey, S., Li, S., Huang, K., Liang, F., Kan, H., Shi, X., Liu, Y., 2022. Assessment of long-term particulate nitrate air pollution and its health risk in China. *iScience* 25, 104899. <https://doi.org/10.1016/j.isci.2022.104899>.
- He, Q., Gao, K., Zhang, L., Song, Y., Zhang, M., 2021. Satellite-derived 1-km estimates and long-term trends of PM_{2.5} concentrations in China from 2000 to 2018. *Environ. Int.* 156, 106726 <https://doi.org/10.1016/j.envint.2021.106726>.
- Johnson, M.M., Garcia-Menendez, F., 2022. Uncertainty in health impact assessments of smoke from a wildfire event. *GeoHealth*. 6 <https://doi.org/10.1029/2021GH000526>.
- Just, A.C., Arfer, K.B., Rush, J., Dorman, M., Shtein, A., Lyapustin, A., Kloog, I., 2020. Advancing methodologies for applying machine learning and evaluating spatiotemporal models of fine particulate matter (PM_{2.5}) using satellite data over large regions. *Atmos. Environ.* 239, 117649 <https://doi.org/10.1016/j.atmosenv.2020.117649>.
- Keller, J.P., Chang, H.H., Strickland, M.J., Szpiro, A.A., 2017. Measurement error correction for predicted spatiotemporal air pollution exposures. *Epidemiology*. 28, 338. <https://doi.org/10.1097/EDE.0000000000000623>.
- Kloog, I., Chudnovsky, A.A., Just, A.C., Nordio, F., Koutrakis, P., Coull, B.A., Lyapustin, A., Wang, Y., Schwartz, J., 2014. A new hybrid spatio-temporal model for estimating daily multi-year PM_{2.5} concentrations across northeastern USA using high resolution aerosol optical depth data. *Atmos. Environ.* 95, 581–590. <https://doi.org/10.1016/j.atmosenv.2014.07.014>.
- Laszlo, I., Liu, H., 2016. *EPS Aerosol Optical Depth (AOD) Algorithm Theoretical Basis Document*. NOAA NESDIS Center for Satellite Applications and Research.
- Levy, M.C., Collender, P.A., Carlton, E.J., Chang, H.H., Strickland, M.J., Eisenberg, J.N.S., Remais, J.V., 2019. Spatiotemporal error in rainfall data: consequences for epidemiologic analysis of waterborne diseases. *Am. J. Epidemiol.* 188, 950–959. <https://doi.org/10.1093/aje/kwz010>.
- Li, Z., Tang, Y., Song, X., Lazar, L., Li, Z., Zhao, J., 2019. Impact of ambient PM_{2.5} on adverse birth outcome and potential molecular mechanism. *Ecotoxicol. Environ. Saf.* 169, 248–254. <https://doi.org/10.1016/j.ecoenv.2018.10.109>.
- Liu, Y., Diner, D.J., 2017. Multi-angle imager for aerosols: A satellite investigation to benefit public health. *Public Health Rep.* 132, 14–17. <https://doi.org/10.1177/0033354916679983>.
- Liu, Y., Koutrakis, P., Kahn, R., 2007a. Estimating fine particulate matter component concentrations and size distributions using satellite-retrieved fractional aerosol optical depth: part 1—method development. *J. Air Waste Manage. Assoc.* 57, 1351–1359. <https://doi.org/10.3155/1047-3289.57.11.1351>.
- Liu, Y., Koutrakis, P., Kahn, R., Turqueti, S., Yantosca, R.M., 2007b. Estimating fine particulate matter component concentrations and size distributions using satellite-retrieved fractional aerosol optical depth: part 2—A case study. *J. Air Waste Manage. Assoc.* 57, 1360–1369. <https://doi.org/10.3155/1047-3289.57.11.1360>.
- Liu, M., Huang, Y., Ma, Z., Jin, Z., Liu, X., Wang, H., Liu, Y., Wang, J., Jantunen, M., Bi, J., Kinney, P.L., 2017. Spatial and temporal trends in the mortality burden of air pollution in China: 2004–2012. *Environ. Int.* 98, 75–81. <https://doi.org/10.1016/j.envint.2016.10.003>.
- Lyapustin, A., Wang, Y., Korkin, S., Huang, D., 2018. MODIS collection 6 MAIAC algorithm. *Atmos. Meas. Tech.* 11, 5741–5765. <https://doi.org/10.5194/amt-11-5741-2018>.
- Ma, Z., Dey, S., Christopher, S., Liu, R., Bi, J., Balyan, P., Liu, Y., 2022. A review of statistical methods used for developing large-scale and long-term PM_{2.5} models from satellite data. *Remote Sens. Environ.* 269, 112827 <https://doi.org/10.1016/j.rse.2021.112827>.
- Meng, X., Garay, M.J., Diner, D.J., Kalashnikova, O.V., Xu, J., Liu, Y., 2018. Estimating PM_{2.5} speciation concentrations using prototype 4.4 km-resolution MISR aerosol properties over Southern California. *Atmos. Environ.* 181, 70–81. <https://doi.org/10.1016/j.atmosenv.2018.03.019>.
- Meng, X., Hang, Y., Lin, X., Li, T., Wang, T., Cao, J., Fu, Q., Dey, S., Huang, K., Liang, F., Kan, H., Shi, X., Liu, Y., 2023. A satellite-driven model to estimate long-term particulate sulfate levels and attributable mortality burden in China. *Environ. Int.* 171, 107740 <https://doi.org/10.1016/j.envint.2023.107740>.
- Murphy, R., Ardanuy, P., Deluccia, F.J., Clement, J., Schueler, C.F., 2006. The visible infrared imaging radiometer suite. In: *Earth Science Satellite Remote Sensing: Vol. 1: Science and Instruments*, pp. 199–223. https://doi.org/10.1007/978-3-540-37293-6_11.
- Murray, N.L., Holmes, H.A., Liu, Y., Chang, H.H., 2019. A Bayesian ensemble approach to combine PM_{2.5} estimates from statistical models using satellite imagery and numerical model simulation. *Environ. Res.* 178, 108601 <https://doi.org/10.1016/j.envres.2019.108601>.
- Orellano, P., Reynoso, J., Quaranta, N., Bardach, A., Ciapponi, A., 2020. Short-term exposure to particulate matter (PM₁₀ and PM_{2.5}), nitrogen dioxide (NO₂), and ozone (O₃) and all-cause and cause-specific mortality: systematic review and meta-analysis. *Environ. Int.* 142, 105876 <https://doi.org/10.1016/j.envint.2020.105876>.
- Pu, Q., Yoo, E.-H., 2022. A gap-filling hybrid approach for hourly PM_{2.5} prediction at high spatial resolution from multi-sourced AOD data. *Environ. Pollut.* 315, 120419 <https://doi.org/10.1016/j.envpol.2022.120419>.
- Pun, V.C., Kazemparkouhi, F., Manjourides, J., Suh, H.H., 2017. Long-term PM_{2.5} exposure and respiratory, cancer, and cardiovascular mortality in older US adults. *Am. J. Epidemiol.* 186, 961–969. <https://doi.org/10.1093/aje/kwx166>.
- Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* 133, 1155–1174. <https://doi.org/10.1175/MWR2906.1>.
- Rappold, A.G., Fann, N.L., Crooks, J., Huang, J., Cascio, W.E., Devlin, R.B., Diaz-Sanchez, D., 2014. Forecast-based interventions can reduce the health and economic burden of wildfires. *Environ. Sci. Technol.* 48, 10571–10579. <https://doi.org/10.1021/es5012725>.
- Rose, A.N., McKee, J.J., Urban, M.L., Bright, E.A., Sims, K.M., 2019. *LandScan 2018*. In: *In. Oak Ridge National Laboratory, Oak Ridge, TN*.
- Sarafian, R., Kloog, I., Just, A.C., Rosenblatt, J.D., 2019. Gaussian Markov random fields versus linear mixed models for satellite-based PM_{2.5} assessment: evidence from the northeastern USA. *Atmos. Environ.* 205, 30–35. <https://doi.org/10.1016/j.atmosenv.2019.02.025>.
- Sha, T., Ma, X., Zhang, H., Janecek, N., Wang, Y., Wang, Y., Castro García, L., Jenerette, G.D., Wang, J., 2021. Impacts of soil NO_x emission on O₃ air quality in rural California. *Environ. Sci. Technol.* 55, 7113–7122. <https://doi.org/10.1021/acs.est.0c06834>.
- She, Q., Choi, M., Belle, J.H., Xiao, Q., Bi, J., Huang, K., Meng, X., Geng, G., Kim, J., He, K., Liu, M., Liu, Y., 2020. Satellite-based estimation of hourly PM_{2.5} levels during heavy winter pollution episodes in the Yangtze River Delta, China. *Chemosphere* 239, 124678. <https://doi.org/10.1016/j.chemosphere.2019.124678>.
- Shen, Y., Wang, C., Yu, G., Meng, X., Wang, W., Kan, H., Zhang, J., Cai, J., 2022. Associations of ambient fine particulate matter and its chemical constituents with birth weight for gestational age in China: A Nationwide survey. *Environ. Sci. Technol.* 56, 8406–8415 <https://pubs.acs.org/doi/10.1021/acs.est.1c08393>.
- Wang, Y., Jiang, X., Yu, B., Jiang, M., 2013. A hierarchical Bayesian approach for aerosol retrieval using MISR data. *J. Am. Statist. Ass.* 108, 483–493. <https://doi.org/10.48550/arXiv.1107.3351>.
- Winker, D.M., Vaughan, M.A., Omar, A., Hu, Y., Powell, K.A., Liu, Z., Hunt, W.H., Young, S.A., 2009. Overview of the CALIPSO mission and CALIOP data processing algorithms. *J. Atmos. Ocean. Technol.* 26, 2310–2323.
- Xiao, Q., Wang, Y., Chang, H.H., Meng, X., Geng, G., Lyapustin, A., Liu, Y., 2017. Full-coverage high-resolution daily PM_{2.5} estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sens. Environ.* 199, 437–446. <https://doi.org/10.1016/j.rse.2017.07.023>.
- Xue, T., Zheng, Y., Tong, D., Zheng, B., Li, X., Zhu, T., Zhang, Q., 2019. Spatiotemporal continuous estimates of PM_{2.5} concentrations in China, 2000–2016: A machine learning method with inputs from satellites, chemical transport model, and ground observations. *Environ. Int.* 123, 345–357. <https://doi.org/10.1016/j.envint.2018.11.075>.
- Yang, Y., Zhao, C., Wang, Q., Cong, Z., Yang, X., Fan, H., 2021. Aerosol characteristics at the three poles of the Earth as characterized by Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observations. *Atmos. Chem. Phys.* 21, 4849–4868. <https://doi.org/10.5194/acp-21-4849-2021>.
- Yang, X., Wang, Y., Zhao, C., Fan, H., Yang, Y., Chi, Y., Shen, L., Yan, X., 2022. Health risk and disease burden attributable to long-term global fine-mode particles. *Chemosphere* 287, 132435. <https://doi.org/10.1016/j.chemosphere.2021.132435>.
- Zhang, H., Wang, J., García, L.C., Ge, C., Plessel, T., Szykman, J., Murphy, B., Spero, T.L., 2020a. Improving surface PM(2.5) forecasts in the United States using an Ensemble of Chemical Transport Model Outputs: 1. Bias correction with surface observations in

- nonrural areas. *J. Geophys. Res. Atmos.* 125 <https://doi.org/10.1029/2019JD032293>.
- Zhang, T., Geng, G., Liu, Y., Chang, H.H., 2020b. Application of Bayesian additive regression trees for estimating daily concentrations of PM_{2.5} components. *Atmosphere*. 11, 1233. <https://doi.org/10.3390/atmos11111233>.
- Zhang, Y., Ding, Z., Xiang, Q., Wang, W., Huang, L., Mao, F., 2020c. Short-term effects of ambient PM₁ and PM_{2.5} air pollution on hospital admission for respiratory diseases: case-crossover evidence from Shenzhen, China. *Int. J. Hyg. Environ. Health* 224, 113418. <https://doi.org/10.1016/j.ijheh.2019.11.001>.
- Zhang, H., Wang, J., García, L.C., Zhou, M., Ge, C., Plessel, T., Szykman, J., Levy, R.C., Murphy, B., Spero, T.L., 2022a. Improving surface PM_{2.5} forecasts in the United States using an ensemble of chemical transport model outputs: 2. Bias correction with satellite data for rural areas. *J. Geophys. Res. Atmos.* 127 <https://doi.org/10.1029/2021JD035563> e2021JD035563.
- Zhang, L., Wilson, J.P., Zhao, N., Zhang, W., Wu, Y., 2022b. The dynamics of cardiovascular and respiratory deaths attributed to long-term PM_{2.5} exposures in global megacities. *Sci. Total Environ.* 842, 156951 <https://doi.org/10.1016/j.scitotenv.2022.156951>.
- Zheng, C., Zhao, C., Zhu, Y., Wang, Y., Shi, X., Wu, X., Chen, T., Wu, F., Qiu, Y., 2017. Analysis of influential factors for the relationship between PM_{2.5} and AOD in Beijing. *Atmos. Chem. Phys.* 17, 13473–13489. <https://doi.org/10.5194/acp-17-13473-2017>.