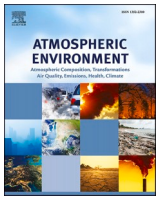
Atmospheric Environment 326 (2024) 120486

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv



High spatio-temporal resolution predictions of $PM_{2.5}$ using low-cost sensor data

Armita Kar [a], Mohammed Ahmed [b], Andrew A. May [b,**], Huyen T.K. Le [a,*]

[a] Department of Geography, The Ohio State University, Columbus, OH, 43210, USA
[b] Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, 43210, USA

HIGHLIGHTS

- Using low-cost sensor data, we predicted hour-specific daily $PM_{2.5}$ at a 100 m resolution in three U.S. cities.
- Spatial-temporal kriging models show better performance than non-spatio-temporal machine learning models.
- Our high-resolution predictions could also facilitate studies on short-term, traffic-based exposure assessment.

ARTICLE INFO

Keywords:
Space-time modeling
Machine learning
Kriging
Air pollution exposure
Multi-city analysis

ABSTRACT

We generated $PM_{2.5}$ predictions at a high spatio-temporal resolution in the Columbus, OH, Denver, CO, and Pittsburgh, PA metropolitan areas using low-cost PurpleAir sensor data. We used multiple modeling approaches, namely random forest (RF), random forest spatial interpolation (RFSI), space-time regression kriging (STRK), and random forest kriging (RFK). We trained separate models for each combination of hour, month, and city to predict $PM_{2.5}$ concentrations at 8 a.m. and 6 p.m. on any specific day at a spatial resolution of 100m. In most cases, models that account for the spatio-temporal relationships (e.g., STRK, RFK, RFSI) show better performance than non-spatio-temporal machine learning models (e.g., RF). On average, considering all models of all cities, RFSI (mean MAE = 1.75, R^2 = 0.67) and STRK (mean MAE = 1.74, R^2 = 0.63) models perform better than RFK models (mean MAE = 2.11, R^2 = 0.59), and STRK has clearest spatial patterns. We found that kriging models, especially STRK, are superior in capturing the spatio-temporal relationships and resemble the generic land use pattern of the city, while RFSI models are effective when dealing with very large datasets with missing cases. Our study demonstrates a multi-model approach that could inform low-cost sensor deployment to facilitate air quality modeling. Our high-resolution predictions could also facilitate studies on short-term, traffic-based exposure assessment.

1. Introduction

Fine particulate matter with an aerodynamic diameter of 2.5 μm or less ($PM_{2.5}$) is associated with high mortality rates and a wide range of health issues globally (Apte et al., 2015; Brook et al., 2010; Hu et al., 2017; Li et al., 2022; McDuffie et al., 2021; Pope and Dockery, 2006). Quantifying the health impacts of $PM_{2.5}$ relies on accurate measurement and assessment of $PM_{2.5}$ exposure (Bi et al., 2020a; Burnett et al., 2014; O'Lenick et al., 2017; Sarnat et al., 2015). In the US, $PM_{2.5}$ concentrations are strongly associated with fuel combustion and emissions from the residential, transportation, and energy sectors, which can vary within a city (Askariyeh et al., 2020; McDuffie et al., 2021; Weagle et al., 2018). Hence, fine-grained measurement and estimation of $PM_{2.5}$ would aid in both exposure assessment and the quantification of the short-and long-term health impacts of $PM_{2.5}$ (Liu et al., 2015).

Air pollution exposures are traditionally measured at home and work locations. For this purpose, low-resolution data can serve as a useful source to understand the long-term health impacts. However, as more efforts transition to dynamic exposure, i.e., understanding the exposure level (and disparity) during daily travel as people quickly move across space, we observe a rising need for higher-resolution data, both spatially and temporally. Without considering this spatial granularity, research

* Corresponding author.
** Corresponding author.
E-mail addresses: may.561@osu.edu (A.A. May), le.253@osu.edu (H.T.K. Le).

https://doi.org/10.1016/j.atmosenv.2024.120486
Received 15 November 2023; Received in revised form 14 March 2024; Accepted 25 March 2024
Available online 26 March 2024

results may be subject to biases that over- or underestimate the true exposure level of individuals living, working, and traveling in different environments with various levels of pollution concentrations (Kim and Kwan, 2019; Ma et al., 2020; Park and Kwan, 2017).

One challenge for developing high-resolution predictions of $PM_{2.5}$ is the lack of granular $PM_{2.5}$ measurements. In most cases, $PM_{2.5}$ measurements are obtained from regulatory air quality monitoring stations (e.g., the U.S. Environmental Protection Agency (EPA) air quality system (AQS) stations) (Heo et al., 2016; Hu et al., 2019a), which are often very sparse in space due to the cost and labor (Bi et al., 2020a; Clements and Vanderpool, 2019). Low-cost particle sensors, on the other hand, are much more affordable than EPA AQS stations (Clements and Vanderpool, 2019). However, these sensors are often criticized for their reliance on a light-scattering technique, which both reduces the accuracy and generates more uncertainty in estimated particle mass concentrations, compared to reference-grade monitors (Bi et al., 2020a; Hagan and Kroll, 2020; He et al., 2020; Morawska et al., 2018; Tryner et al., 2020; Zou et al., 2021). Nevertheless, low-cost sensors are growing in popularity. For example, the PurpleAir low-cost sensor network provides both outdoor and indoor estimates of different PM concentrations (PM_1, $PM_{2.5}$, and PM_{10}), in addition to other environmental parameters such as air temperature, relative humidity, and barometric pressure, at a 2-min temporal-scale (Bi et al., 2022). The PurpleAir website hosts the recorded datasets and allows data downloading at high spatio-temporal resolution (Bi et al., 2020a; Yang et al., 2022).

Prior studies have developed calibration methods to improve the accuracy of low-cost sensor data. Malings et al. (2020) developed two types of long-term sensor corrections using 9 PurpleAir sensors (PA-II) and 25 Met One Neighborhood Particulate Monitors (NPM). The first correction is a physics-based approach, which uses particle composition to predict aerosol hygroscopic growth, then adjusts for particle mass below the optical sensor size cut-point with respect to the Beta Attenuation Monitor. The other correction is based on a statistical approach using environmental variables such as temperature (T) and relative humidity (RH) (Malings et al., 2020). Similarly, U.S. EPA researchers developed a statistical calibration equation to correct PurpleAir sensors, using comparisons to AQS monitors and including the PurpleAir's reported RH values (Barkjohn et al., 2021). PurpleAir sensors generally exhibit high inter-unit consistency ($r > 0.9$) for $PM_{2.5}$ observations (Kelly et al., 2017; Malings et al., 2020; South Coast Air Quality Management District, 2024). Moreover, while PurpleAir sensors experience lower accuracy compared to governmental instruments, two studies in the U.S. show a very good association between PurpleAir sensors and the AQS monitors ($R^2 = 0.88$ and > 0.90, respectively) (Kelly et al., 2017; South Coast Air Quality Management District, 2024).

In recent years, researchers have demonstrated the use of these calibrated low-cost sensor measurements, combined with the appropriate spatial and temporal predictors, to generate more accurate $PM_{2.5}$ prediction models (Bi et al., 2020b; Reis et al., 2015; Saltelli et al., 2010). Low-cost sensors deployed by scientists and non-scientist entities form a network that, as a whole, is larger in number than sparsely distributed regulatory monitors alone. Thus, prediction models designed through high volume low-cost sensor datasets can capture more spatio-temporal variations, especially among residential locations that may be far away from regulatory stations, (Bi et al., 2022) and can yield more accurate $PM_{2.5}$ predictions (Bi et al., 2020a, 2022; Li et al., 2020; Lin et al., 2020; Lu et al., 2021a; Schulte et al., 2020) than the models generated from regulatory monitors (Kloog et al., 2012, 2014; Lin et al., 2020). Such accurate predictions are particularly crucial in investigating short-term, outdoor, or in-traffic exposure to environmental pollution (Reis et al., 2015).

While low-cost sensors offer valuable high-resolution data, their utilization in developing $PM_{2.5}$ prediction models remains limited to only a few U.S.-based studies (Bi et al., 2020a, 2020b, 2022; Jain et al., 2021; Lu et al., 2021a, 2022; Vu et al., 2022). For example, Lu et al. predicted hourly $PM_{2.5}$ for Los Angeles County with a spatial resolution of 500 × 500 m (Lu et al., 2021b). Jain et al. also predicted $PM_{2.5}$ in Pittsburgh, PA, using a 50 × 50 m spatial resolution with an output of daily $PM_{2.5}$ concentrations (Jain et al., 2021). However, these models can further be advanced to improve their reliability and granularity to enable the prediction of $PM_{2.5}$ concentrations at any place and time (Liu et al., 2022; van Donkelaar et al., 2015; Xie et al., 2015).

Past studies designed air quality models using either data-driven machine learning (ML) or geostatistical modeling approaches. Random forest (RF) regression, a tree-based ensemble model, is the most popular ML algorithm adopted for $PM_{2.5}$ predictions (Bi et al., 2020b; Lu et al., 2021b; Vu et al., 2022), while one study applied neural networks and gradient boosting (Di et al., 2019). Prior studies also used various statistical and geospatial modeling approaches for air quality predictions, such as land use regression (LUR) (Kloog et al., 2012; Lee, 2019), space-time regression kriging (Hu et al., 2019b), and spatio-temporal mixed-effect models (Bi et al., 2022; Kloog et al., 2014; Xie et al., 2015). Some studies adopted a combined approach of using ML models for data processing and imputation and statistical models, e.g., ordinary kriging for prediction and visualization (Chang et al., 2020; Li et al., 2020). However, few studies compared the performances of (geo)statistical and ML models within the same study. For example, a comparison between LUR and RF showed that the latter model outperforms the former (Jain et al., 2021). Another study compared the regression-based and ML-based kriging models and found that the ML-geospatial combination performs better than the traditional kriging (Lu et al., 2021a). Thus, while these studies suggested that ML approaches are promising for spatio-temporal predictions, further efforts are needed to fully explore their performance and potential limitations.

In this study, we demonstrate applications of low-cost sensor datasets in predicting $PM_{2.5}$ concentrations at a high spatio-temporal resolution. Our study overcomes the spatio-temporal limitations of past studies by creating prediction surfaces of hour-specific daily $PM_{2.5}$ concentration at 100 × 100m resolution using data from 474 PurpleAir sensors across the three different U.S. cities, namely the Columbus, OH, Denver, CO, and Pittsburgh, PA metropolitan areas. We developed the prediction models using both ML and geostatistical methods, namely RF (non-spatial model), random forest spatial interpolation (RFSI), random forest space-time kriging (RFK), and space-time regression kriging (STRK). This application of multiple models using datasets from multiple cities will help us understand the differences in model performance for high spatio-temporal predictions based on the sensor number, locations, and observed datasets from various urban contexts. Moreover, by focusing on three different geographic locations, we can establish a model that has greater potential to be generalized to most US urban areas with similar land use and urban characteristics.

2. Methods

2.1. Study area and data

We obtained data on $PM_{2.5}$ concentrations from outdoor PurpleAir sensors (PurpleAir, 2023) installed in three metropolitan areas in the US: Columbus, OH, Denver, CO, and Pittsburgh, PA. The data used in this study include hourly $PM_{2.5}$, T, and RH measures recorded from 48 sensors in Columbus, 172 sensors in Denver, and 254 sensors in Pittsburgh from July 1, 2021 to April 30, 2022. We deployed some of the sensors in Columbus for our study, but the remainder were installed by other users across all three cities. However, not all these sensors have data captured for all date-time stamps used in this study (e.g., some of our sensors were not installed until September 2021). We adjusted the raw $PM_{2.5}$ values using the calibration equation established by Barkjohn et al. (2021). Prior to deploying our sensors, we conducted a simple outdoor comparison study to investigate inter-device agreement (see Figures S1 – S2 and Tables S1 – S4). Additionally, we checked the percentile distribution of $PM_{2.5}$ (Table S5) and excluded any data point that is greater than 50 $\mu g/m^3$, suspecting that these are an outcome of

instrumental errors or extreme weather events. From a practical standpoint, these extreme values would be difficult to capture in any predictive model because their occurrences are likely due to processes that a statistical model could not capture (e.g., transport of wildland fire smoke). After excluding the outliers above 50 μg/m^3, we retained 97–99% of observations from the original datasets (Table S6).

Table 1 provides a list of the predictor variables used for modeling $PM_{2.5}$, data sources, and spatial and temporal resolution. We used T and RH, two common predictors used in modeling $PM_{2.5}$ in past studies (Bi et al., 2020b; Di et al., 2019; Kloog et al., 2012, 2014), and obtained these data directly from the PurpleAir sensors for the locations where they were deployed. For the rest of the study area, we used datasets from the NOAA High-Resolution Rapid Refresh (HRRR) data archive (Benjamin et al., 2016; The University of Utah Mesowest Group, 2022) to obtain complete spatial coverage in our modeling domains. Other predictors in our models are commonly used in previous studies, such as population, housing density, and land use classifications (Bi et al., 2020b, 2022; Di et al., 2019; Jain et al., 2021; Kloog et al., 2012, 2014), as well as road functional classes, road length, distance to roads, and traffic volume to account for fuel combustion from the transportation sector (Bi et al., 2020b, 2022; Jain et al., 2021; McDuffie et al., 2021; Weagle et al., 2018). We collected the land cover data from the National Land Cover Database (NLCD) (Earth Resources Observation and Science (EROS) Center, 2019), containing 20 land cover classes that we further combined into eight groups (see SI Section 2 and Figures S3-S5 for details).

To capture the temporal variation, we included other predictors that are more temporally-variant such as hourly-monthly visitors at accommodation and food service locations and art, entertainment, and recreational locations as proxies for cooking activities in commercial areas, as emission from restaurant cooking is another major source of air pollution in urban settings (Sinaga et al., 2020; Song et al., 2021). Although emissions from retail activities, industries, and energy sectors also contribute toward increased $PM_{2.5}$ concentrations, the locations of these activities are not within close proximity to the PurpleAir sensors (i.e., 100 m) in the study areas; therefore, we did not enter these variables into the model, and consequently, their impacts on $PM_{2.5}$ cannot be captured in the prediction models.

Table 1
Data on land use and human activity patterns collected from secondary sources.

Data	Unit	Source	Spatial resolution	Temporal resolution
Temperature Relative humidity	°F %	PurpleAir/ NOAA High-Resolution Rapid Refresh (HRRR)	PurpleAir: Sensor-specific NOAA HRRR: 3 km grids	Hourly data
Population density	People per m^2	5-year estimates American Community Surveys (2018)	Census block group	1y
Road length	m	TIGER/Lines shapefiles	Total length in meters by census block group	1y
Land cover data: high, medium, and low-intensity development, open space, grassland, forest, wetland, water	m^2	National Land Cover Database (NLCD), 2019	30-m resolution	5y
Visitor counts at accommodation and food service locations	People	SafeGraph (cell phone data)	At each point of interest location	Hourly visitors in a month
Visitor counts at arts, entertainment, and recreation locations	People	SafeGraph (cell phone data)	At each point of interest location	Hourly visitors in a month

We compiled the hourly $PM_{2.5}$ data at 8 a.m. and 6 p.m. each day, created a 100 × 100 m grid for the study areas, and assigned hourly $PM_{2.5}$ values to each cell in this grid. We chose a grid cell size of 100 × 100 m to allow for good spatial resolution while ensuring computational feasibility during the modeling procedure. We modeled data at 8 a.m. and 6 p.m. to capture the variations in air quality during morning and evening peak hours of traffic. We adjusted the spatial resolution of all variables to the grid cell size of 100 m and estimated their aggregated values for each grid (see Section 3 in the SI and Figures S6 – S8 for details).

2.2. Modeling approaches

We modeled air quality as a function of temperature, relative humidity, the transportation network, land use, and visitor counts at food and recreational establishments (Table 1) using the gridded datasets with PurpleAir sensor locations, as well as a binary variable to account for weekday/weekend differences. We used four modeling approaches, namely RF, RFSI, STRK, and RFK. The RF models predict $PM_{2.5}$ disregarding the underlying spatio-temporal variabilities in the dataset. RFSI models advance RF by incorporating additional covariates that capture spatial dependencies among the observed locations. Kriging models such as STRK and RFK measure $PM_{2.5}$ at an unobserved location and time by fitting a linear regression and RF-based trend function, followed by spatial-temporal ordinary kriging for residuals interpolation (see Section 4 in the SI for descriptions of these models).

We generated separate models to measure $PM_{2.5}$ at 8 a.m. and 6 p.m. for each month and each city using various statistical and machine learning methods. Specifically, we designed 20 RF models (10 months × 2 h), 20 RFSI models, 20 RFK models, and 20 STRK models for each city. Each model was trained using hour-specific data observations for all days within a particular month and city. Consequently, each model can predict hour-specific $PM_{2.5}$ concentrations for any day within that same month and city. For instance, an STRK model, specific to 8 a.m. of October in Denver, can predict $PM_{2.5}$ at all 100 × 100m grid locations in Denver for 8 a.m. on any day in October, say October 12th.

We determined this hour- and month-specific modeling approach based on the data requirements of kriging-based approaches, which necessitate a dataset without missing values for any observed location and timestamp. Across all study areas, very few sensors were operational throughout the entire 10-month study period, meeting the requirement for a complete dataset. Consequently, relying on a single model to predict hourly $PM_{2.5}$ would depend on this limited number of sensors. Instead, we opted for models tailored to specific hours and months, allowing us to benefit from broader sensor coverage and a more extensive dataset for each model.

The hour-specific monthly models, therefore, have different numbers of sensors and only include the sensors that do not contain any missing data for any days within that particular hour and month of interest. The availability of sensors and their data completeness improved over the months for all cities. In Columbus, the first five months had fewer sensors with 100% data completeness than the rest of the study months. In Denver and Pittsburgh, datasets from about 47 to 78% and 51 to 65% of total available sensors were used for the $PM_{2.5}$ models (Table S6). Additionally, there were some cases where the corresponding sensor grids did not contain any waterbodies, forests, and certain activity locations (accommodation, food services, and recreational activities) and generated zero-values in the input dataset for the respective land use variables and null model coefficients in the STRK models. We also omitted the corresponding variables with zero values from that hour-specific monthly model in these cases. For instance, Columbus prediction models only from October to April included hourly visitor counts to food and recreational establishments (Figure S6). For Denver, the

models only included visitor counts to food services (Figure S8).

2.3. Model validation

We validated the model results in two ways: (1) performing *k*-fold cross-validation using PurpleAir sensor data, and (2) comparing model predictions with the $PM_{2.5}$ measurements from the EPA monitors.

Cross-validation during the modeling process. We applied a *k*-fold leave-location-out cross-validation technique to evaluate model accuracy. In this method, data from all sensor locations were divided into *k* parts to ensure that all daily observations at a given location are in the same fold. At each step of the validation, one-fold was set aside for testing and measuring prediction accuracy, whereas the rests (*k*-1 folds) were used for model training. We used mean absolute error (MAE) and coefficients of determination (R^2) as indicators of model performance.

Comparison of model predictions with EPA measurements. We retrieved $PM_{2.5}$ observations from EPA regulatory monitors in Columbus (n = 1), Denver (n = 7), and Pittsburgh (n = 13) to serve as the ground truth data. Instead of using specific grid cells containing EPA monitors, we used neighborhoods of cells for validation purposes because the prediction within any grid cell may deviate from the ground truth data. Each validation neighborhood contains 9 grid cells, including the cell containing the EPA monitor and 8 neighboring cells (i.e., those that share an edge or vertex with the EPA-contained cell).

We predicted hour-specific $PM_{2.5}$ concentrations for six days per month for the cell neighborhoods in each city using the hour-specific monthly RFSI, STRK, and RFK models. The dates were chosen five days apart, starting from the 3rd day of the month. Then, for each validation neighborhood, we calculated the mean values of our modeled data for each day (8 a.m. or 6 p.m.) and compared them with the EPA measurements. We used errors and biases in both absolute and relative terms as the evaluation metrics. The absolute error and bias represent the absolute (non-negative) and raw differences between EPA measurements and $PM_{2.5}$ predictions. The relative error and relative bias represent absolute error and absolute bias divided by the EPA measurement.

3. Results

3.1. Descriptive statistics

Fig. 1 illustrates the monthly variation in $PM_{2.5}$ concentrations from the PurpleAir sensors between July 2021 and April 2022 for the three study areas at 8 a.m. (left) and 6 p.m. (right). Columbus, Pittsburgh, and Denver experience similar concentrations of $PM_{2.5}$ throughout the year (i.e., the mean of the means is roughly 9 ± 4 μg/m^3 for all three cities). The sensors located in Pittsburgh and Denver yield more outliers (red dots in Fig. 1) compared to Columbus, which are likely attributable to a combination of the number of sensors in Columbus (n = 48 vs. 172 in Denver and 254 in Pittsburgh) and the geographic distribution of these sensors (e.g., the sensors are sparser and more clustered in Columbus than the other two cities). In general, $PM_{2.5}$ concentrations are higher in both summer (July–August) and winter (December and January) for all three cities (Figure S9). Moreover, Columbus and Pittsburgh are in a

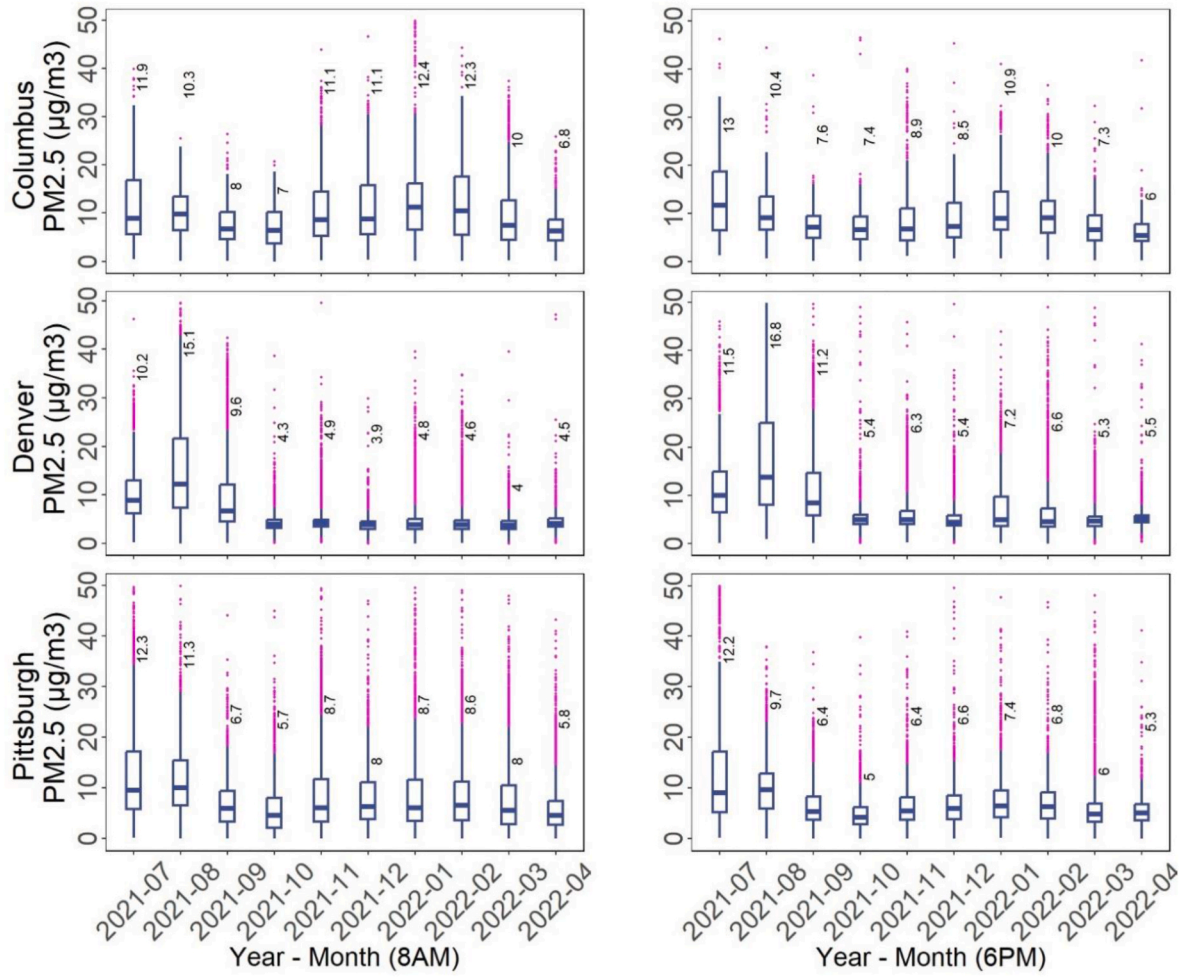


Fig. 1. Boxplots showing the monthly variation in $PM_{2.5}$ concentration for 8 a.m. (left) and 6 p.m. (right), based on PurpleAir data. The values along the boxplots represent their mean values.

cool, humid climate zone, while Denver is in a cold, dry climate zone (International Code Council (ICC), 2021). Due to their close spatial proximity (~260 km), Columbus and Pittsburgh often experience similar temporal variations in weather (Figure S9 – S11).

3.2. Comparison across models

We developed 80 models (10 months × 2 h x 4 modeling approaches) for each city using the PurpleAir sensor datasets. Considering both 8 a.m. and 6 p.m. models, the average mean absolute error (MAE) scores respectively for RF, RFSI, STRK, and RFK models are 3.16, 2.03, 1.50, 1.85 in Columbus, 2.67, 1.35, 1.67, 2.00 in Denver and 3.19, 1.89, 1.96, 2.46 in Pittsburgh. Moreover, the average R^2 for RF, RFSI, STRK, and RFK models are 0.38, 0.78, 0.81, 0.84 in Columbus; 0.25, 0.62, 0.53, 0.44 in Denver; and 0.34, 0.63, 0.62, 0.52 in Pittsburgh.

Figs. 2 and 3 represent the MAE and R^2, respectively, of these models for each hour, month and city. For example, the first orange-colored box on the left-most panel in Fig. 2 represents the distribution of MAE estimated from 10 RF monthly models for 8 a.m. in Columbus. The RF models yield higher MAE, indicating worse prediction performance than the other modeling approaches, especially for Columbus and Pittsburgh (Fig. 2). For Denver, RF and RFK show similar model performance, whereas RFSI and STRK demonstrate slightly better performance with lower ranges of MAE. Also, we observe differences in model performance between morning and evening hour models in all cities. For instance, the 6 p.m. models show relatively better performance for Columbus and Pittsburgh, whereas it is the opposite for Denver.

Similarly, R^2 values are notably higher for RFSI, STRK, and RFK models than the RF models, indicating that they can better explain the variabilities in the observed datasets (Fig. 3). However, the ranges of R^2 values differ substantially between 8 a.m. and 6 p.m. models for Columbus and Pittsburgh, and the 6 p.m. models yield a relatively wider range of R^2. Similarly, Denver models have a wide range of R^2 for both 8 a.m. and 6 p.m. This wide range of R^2 indicates that the current predictors may not adequately explain the variabilities in $PM_{2.5}$ concentrations for all monthly models – while this set of predictors satisfactorily explains the variabilities in $PM_{2.5}$ concentrations for some models of some months, they fail to do so for other models, resulting in the high variations in R^2 values. The distribution of sensors and the resultant land coverage data availability may have contributed to this fluctuation in R^2. In all cities, the sensors are mainly concentrated in the core urban areas, and the corresponding grid cells only contain information on land use and activity patterns around the sensors. Therefore, our models do not capture the land use patterns more likely to be available in suburban areas (such as grassland, forest, and wetland) and their effects on $PM_{2.5}$ concentrations, especially for Denver and Pittsburgh. Due to the relatively poor performance of RF in both Figs. 2 and 3, we exclude this model from further discussion.

Fig. 4 illustrates the predicted $PM_{2.5}$ concentrations from the STRK, RFSI, and RFK models for the core urban areas of each metropolitan area on February 18, 2022, at 8 a.m., a time and date chosen at random and irrespective of the model performance indicators. Additional prediction maps are provided in SI Section 6, including maps of the entire study areas (for the same date) and core urban areas (on different dates) (Figures S12-S14). We also provide a temporally-averaged prediction map (across days in February 2022) in Figure S15 to investigate underlying spatial patterns. The color bar scaling within each map differs to highlight spatial differences.

For each city, some spatial similarities exist across the $PM_{2.5}$ predictions derived from each model. For instance, the central urban areas consistently have higher concentrations of $PM_{2.5}$, marked by red-colored grids in both Columbus and Denver. The fringe areas of these two study areas have lower $PM_{2.5}$ values in the suburbs than in the main city. In Pittsburgh, high $PM_{2.5}$ values are more concentrated near the waterbodies and their surroundings than the rest of the core city. The predicted patterns of $PM_{2.5}$ observed in Pittsburgh are consistent with the geographical features, i.e., multi-lane highways and some polluted industrial sources are located adjacent to the river valleys. Moreover, this is confirmed by prior research with mobile observations of air pollution within Pittsburgh (Gu et al., 2018; Li et al., 2018; Tan et al., 2014).

The prediction differences and spatial dissimilarities in the RFSI, STRK, and RFK prediction maps for each city are perhaps an outcome of their methodological differences in capturing spatial relationships. Using a spatial interpolation approach, STRK maps reflect the generic land use pattern for each city and include very spatially smooth prediction surfaces. The prediction values from STRK follow a wider range than the other two models and often contain outliers due to their dependency on linear regression as the base model algorithm. Even though we used a global threshold value to exclude extreme values from all models, the data distribution of $PM_{2.5}$ was non-normal for some months and study areas, indicating many outliers in the training set. Applying a linear regression model on such skewed datasets tends to produce higher

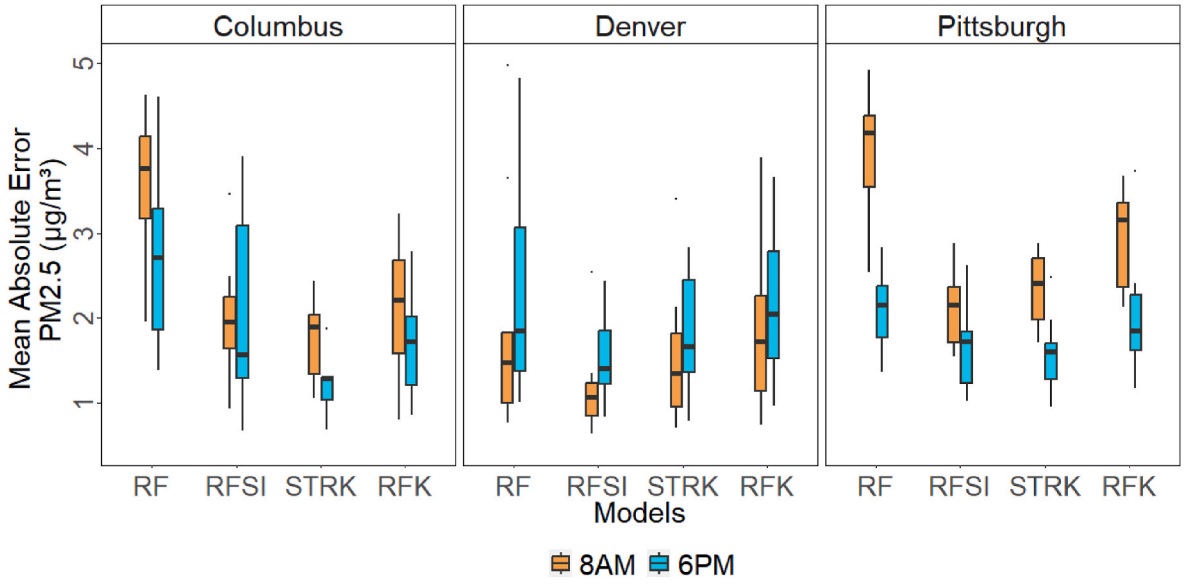


Fig. 2. Mean absolute errors (MAE) of the RF, RFSI, STRK, and RFK models, estimated for 8 a.m. and 6 p.m. for Columbus, Denver, and Pittsburgh. Each box plot represents a distribution of the MAE for the 10 models specific to an hour and month for the respective study areas.

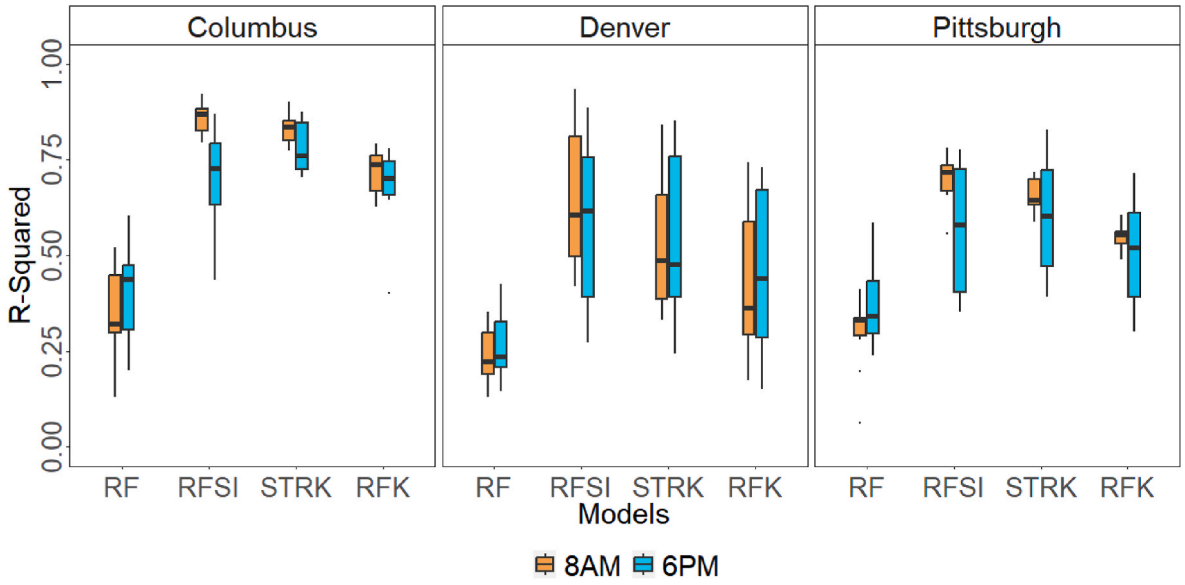


Fig. 3. R^2 of the RF, RFSI, STRK, and RFK models, estimated for 8 a.m. and 6 p.m. for Columbus, Denver, and Pittsburgh. Each box plot represents a distribution of R^2 of the 10 models specific to an hour and month for the respective study areas.

values and outliers, reflected in the STRK predictions.

Similarly, the spatial patterns in the RFK prediction maps somewhat align with the generic land use pattern due to its use of kriging to embed spatio-temporal variabilities within the RF algorithm. However, compared to STRK, the resemblance of RFK maps to the land use pattern is less pronounced, and predictions also follow a smaller range than STRK predictions. Unlike linear regression models, machine learning models such as RF generate predictions that are close to the observed values in the training set; resulting in low values of model residuals (difference between the observed data and RF predictions). Therefore, the RFK predictions are mainly driven by the RF component rather than the kriging component of the model.

The RFSI maps are more abstract since they only include observations and distances from neighbors in the predictor list to indicate the spatial component. When training the models, RFSI uses a fixed bandwidth of neighborhood distances based on the spatial distribution of sensors. RFSI assumes distances between the prediction grids and their neighbors are also within this bandwidth and uses the same prediction mechanism even if they are not (Sekulić et al., 2020). This could be problematic for our study areas. In all cities, the sensors were clustered in central urban areas, resulting in a small distance bandwidth (Figure S3-S5). The fringe area grids were at a distance shorter or longer than the threshold, yet received similar predictions assuming that the neighbors were at the minimum or maximum neighborhood distance of the trained model.

3.3. Model evaluation and comparison with EPA data

Fig. 5 illustrates the distribution of errors and biases (absolute and relative) estimated for all cities and models by comparing the predicted values and the ground truth data from EPA. Overall, the mean absolute errors are slightly lower for predictions from the RFSI models than those from the STRK and RFK models. In terms of mean absolute biases, STRK models perform better or similar to the other two models. The mean absolute biases are mostly negative for all cities and models, indicating over-predictions relative to EPA observations. For context, Murphy et al. report an average root-mean square error of roughly 1.8 $\mu g/m^3$ and an average mean bias of roughly -0.3 $\mu g/m^3$ for a state-of-the-science regional chemical transport model (Murphy et al., 2023).

We use the relative error and bias estimates to enable a cross-comparison across cities that we cannot infer from absolute errors alone. The relative errors exhibit similar patterns across each city and model as in Fig. 5; on average, these relative errors range from roughly 20% to roughly 40%. Similarly, the relative biases, on average, range from roughly -30% to -10%. Thus, there are arguably minor differences in the prediction performance across all cities and models.

3.3.1. Influence of wildfire events

Denver was impacted by transported wildfire smoke in 2021 (and many other years). Langford et al. (2023) highlight enhancements to $PM_{2.5}$ in Colorado due to so-called "megafires" in California and Arizona during 2021 (Langford et al., 2023). In general, the daily-averaged $PM_{2.5}$ concentrations in 2021 were roughly double those in 2019 and 2022. However, we have no strong evidence that wildfire smoke has influenced our model results. We screened data to remove extreme events (>50 $\mu g/m^3$), and both the errors and the biases – both absolute and relative – compared to the EPA AQS monitors for Denver are similar to those for Columbus and Pittsburgh (Fig. 5).

4. Discussion

4.1. Summary of the results

In this study, we generated high spatio-temporal resolution prediction surfaces of $PM_{2.5}$ using data from low-cost sensors in the Columbus, OH, Denver, CO, and Pittsburgh, PA metropolitan areas. We applied ML and geostatistical approaches, namely RF, RFSI, STRK, and RFK, to predict daily $PM_{2.5}$ concentrations at a specific hour at 100×100 m resolution. This spatio-temporal granularity is an improvement over the existing models from past studies that tend to be aggregated more coarsely over time and/or space. We chose the predictions of STRK model as the final predictions because of the similar or better model fits and the spatial realism in the prediction surfaces as compared to those of other modeling approaches.

Although our results are not directly comparable with prior efforts given different datasets and contexts, we found that the model performance indicators are comparable with those of the previous studies. For instance, considering both 8 a.m. and 6 p.m. models, the mean R^2 of all 8 a.m. and 6 p.m. STRK models for Pittsburgh is 0.62, which is lower than the R^2 of 0.78 found in another model evaluation in Pittsburgh (Jain et al., 2021). It is worth noting that our models use much more granular data with high spatio-temporal resolutions, which often involve more

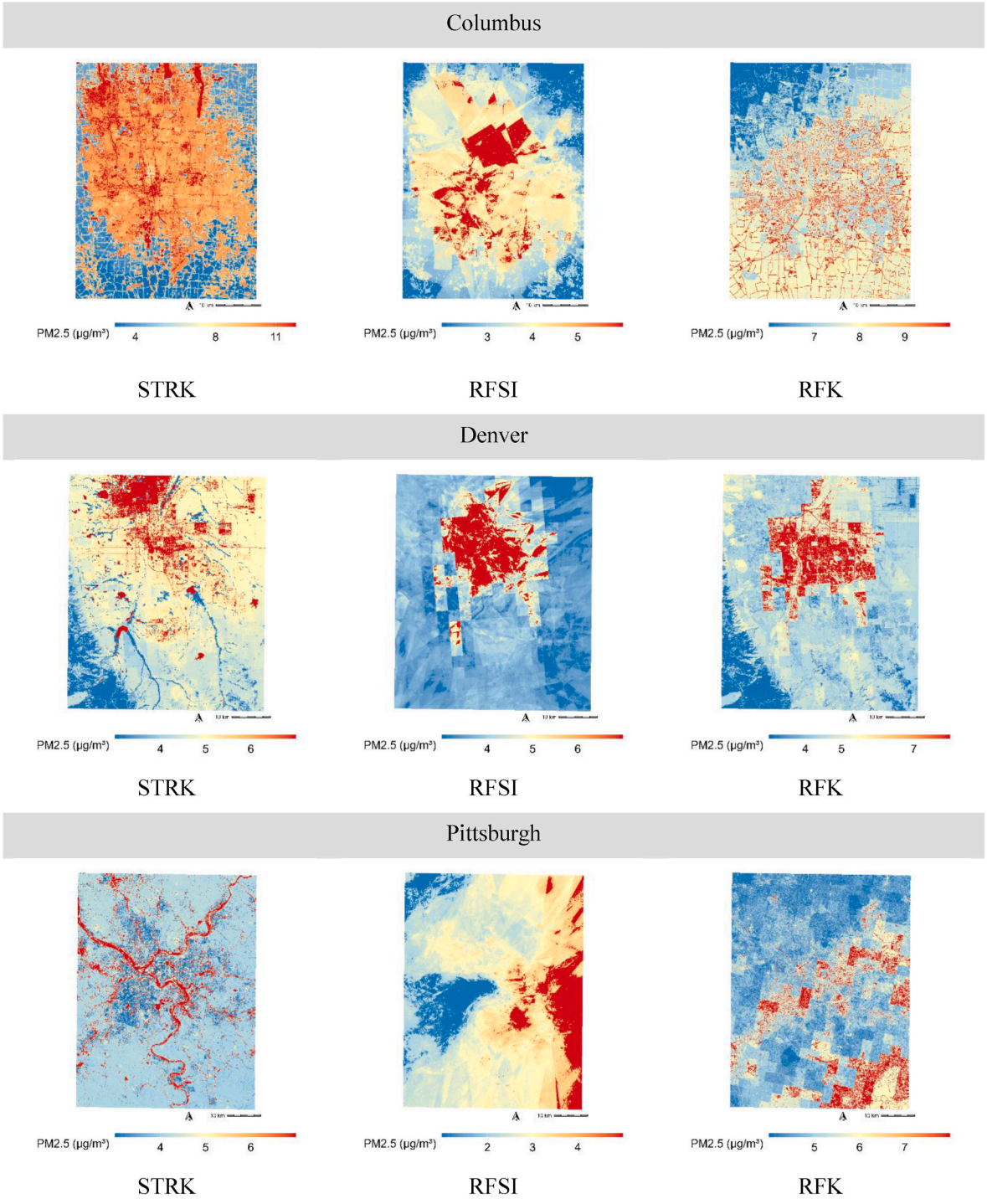


Fig. 4. STRK, RFSI, and RFK prediction maps for three cities on February 18, 2022, at 8 a.m. (month and date are arbitrarily selected). The numeric values in each color scale represent the range of predicted values between 10th to 90th percentile. We selected different color scales for each city to highlight spatial differences in the prediction surfaces among the maps.

variations in the data, more complicated modeling techniques, and thus can have lower R^2 compared to the same methods applied to data with lower temporal or spatial resolutions. For example, the previous study on Pittsburgh (Jain et al., 2021) used daily estimates of $PM_{2.5}$ that may have contributed to producing a more stable training dataset and slightly better model performance than ours. Similarly, other studies model $PM_{2.5}$ at larger spatial resolutions (≥500 m) and/or at longer temporal resolutions (e.g., daily estimates) (Bi et al., 2020a, 2022) with low-cost sensor data for the state of California (Bi et al., 2020a, 2022; Vu et al., 2022), Los Angeles County (Lu et al., 2021b), and other cities in the United States, all achieving higher R^2 (i.e., above 0.85) (Bi et al., 2020a, 2022; Lu et al., 2021b, 2022; Vu et al., 2022). However, as our goal is to improve prediction accuracy, MAE is a better indicator of model predictive power because models with a high R^2 may be subject to overfitting and limit their generalizability. Overall, our models have lower error rates (our mean MAE = 1.74 µg/m^3) as compared to models in past studies using low-cost sensor data, such as Jain et al. (2021) with MAE = 1.81 µg/m^3 (Jain et al., 2021), Lu et al. (2021) with RMSE = 3.23 µg/m^3 (Lu et al., 2021b), Bi et al. (2020) with RMSE = 3.71 µg/m^3 (Bi et al., 2020b), and Lu et al. (2022) with MAE = 2.01 µg/m3 and R^2 =

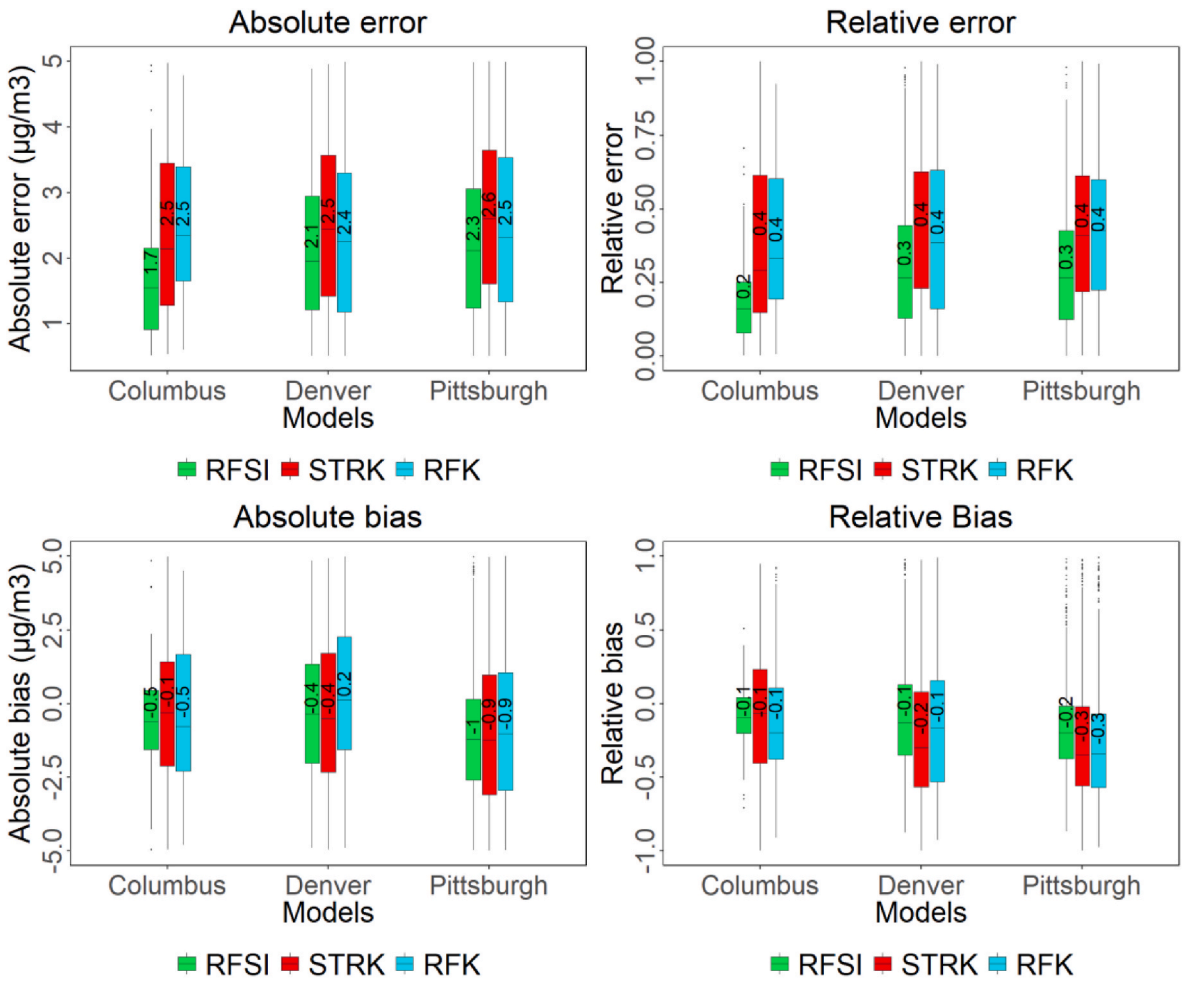


Fig. 5. Distribution of errors and biases (absolute and relative) between the predicted $PM_{2.5}$ and the estimates from EPA monitors of all cities and models. The values along the boxplots indicate their mean.

0.66 (Lu et al., 2022), which is equivalent to our study.

4.2. *Monitoring and modeling air quality with low-cost sensors data: limitations and future directions*

While low-cost sensors have lower accuracy and precision than EPA AQS monitors, sensors have the ability to provide useful information. Measurements from PurpleAir sensors and the EPA AQS monitors are highly correlated (Kelly et al., 2017; South Coast Air Quality Management District, 2024). Moreover, PurpleAir sensors tend to have high inter-unit consistency (Malings et al., 2020; South Coast Air Quality Management District, 2024) (also see Figures S1-S2). Aggregation of measurement at the hourly level, as we did, can reduce some of the uncertainty in the sensor measurements (Barkjohn et al., 2021). Together, this means that after careful data processing, we can apply calibration equations, such as the method proposed by Barkjohn et al., to obtain more accurate results (Barkjohn et al., 2021). We acknowledge that uncertainty in the observed sensor concentrations may propagate through to the model results. Nonetheless, we agree with Reis et al. (2015) in that it is still valuable to have more data and results with some uncertainty – our work being one of them – to enrich our current practice of predicting air quality with fewer data yet higher certainty using data from EPA regulatory monitors.

Additionally, regardless of the modeling approach, spatial over-extrapolation may occur when predicting $PM_{2.5}$ at locations far away from the sensor clusters, such as in the urban outskirts. However, this does not affect our evaluation of model performance (Fig. 5) and predictions around the urban centers (Fig. 4), where both EPA monitors and Purple Air sensors are clustered (Figure S3-S5).

Our study informs future air quality research in selecting suitable modeling approaches when working with low-cost sensor data based on sensor network characteristics. This study demonstrates that, for modeling $PM_{2.5}$ at a very granular level, spatio-temporal models such as STRK, RFSI, and RFK are better than non-spatio-temporal models such as RF. For this study, kriging models such as STRK and RFK outperform machine-learning-only models (e.g., RF, RFSI), especially in capturing spatial relationships and producing smoother predictions across geographic areas, owing to their kriging spatial interpolation technique. These kriging models are useful when all sensors work consistently over time (i.e., to estimate the space-time variogram), or with a very large network of sensors such that listwise deletion of missing values does not affect the sample size, which is the case in our study. In contrast, the effectiveness of RFSI models is contingent on the spatial distribution of sensors within the study area. While RFSI inherits the advantages of machine learning models, such as no reliance on statistical distribution assumptions, allowing a large set of input variables, and shorter computational time, these models may produce maps with artifacts if the uneven spatial distribution of sensors across the study area as in our study (and most commonly encountered in the real world). RFSI may be more suitable in future studies with more spatially dispersed sensors, and where missing data or inconsistent sensor data is common.

A good model can only be achieved with good data. Our study underscores the pressing need for a well-coordinated sensor deployment and regular maintenance strategy. While independent deployment is a

strength of low-cost sensor networks, this practice results in an uneven distribution of sensors that leads to inconsistency in the spatio-temporal data. This inconsistency, in turn, leads to many challenges in modeling air quality discussed earlier in this section. A potential remedy is to expand the network in spatial coverage and density, such that when a sensor fails at a certain time, other sensors still provide sufficient and reliable data.

To improve model prediction at high spatio-temporal scales, more fine-grained data are needed, such as hourly (or real-time) traffic on the roads or to certain destinations. To achieve this, sensors need to be strategically placed near major pollution sources, such as commercial cooking activities at accommodation, food service, and recreational locations; as well as retail, industrial facilities, and agricultural activities. Some of these are available in our dataset but were not used in the model due to the fact that no sensors are placed near these establishments. In addition, vehicle traffic volumes, such as those provided by StreetLight, could be useful to generate high resolution predictions of $PM_{2.5}$. Moreover, information about $PM_{2.5}$ concentrations away from core urban areas can aid in the prediction for, e.g., low-intensity development and forested land cover classes (see Figures S3 – S5).

While having a consistent number of sensors over time could help improve predictions of $PM_{2.5}$ concentrations, data completeness is also equally important. This requires regular maintenance of the sensor network. Our current approach handles this data incompleteness by adopting a multiple-modeling approach with increased time complexities where each model represents a specific hour, a month, and a city. Future studies can use existing data imputation techniques to evaluate their performance in predicting $PM_{2.5}$ for the missing time stamps. Moreover, current sensor data quality assessment and calibration do not consider temporal variation. Future research can advance them as dynamic models with time variants to be updated periodically using real-time data. Such dynamic models will enable accurate re-calibration of an existing observation, as well as estimation of missing $PM_{2.5}$ values (Li et al., 2023; Taira et al., 2022; Xiang et al., 2016).

In addition to accounting for the sensor and data challenges, future research may adopt a similar approach and further expand to a single model with high temporal resolution instead of having multiple monthly models. Such models require (1) a consistent number of sensors and data across all months, and (2) input variables that capture spatio-temporal variations, such as hourly vehicular traffic volumes to capture the variations in $PM_{2.5}$ during the peak and off-peak periods, as well as monthly variations in both existing and additional weather variables (e.g., wind speed, precipitation, and atmospheric pressure) to capture the seasonal differences in $PM_{2.5}$.

5. Conclusions

By employing low-cost sensor data and the EPA correction factor, we generated high spatiotemporal-resolution prediction surfaces of $PM_{2.5}$ concentrations in three US urban areas. Our $PM_{2.5}$ prediction surface has the highest spatio-temporal resolution (i.e., hourly-specific daily level, 100m grid), the model performance is comparable with that from past studies, and the model error (MAE) is lower. The results and approach have the potential to generalize to other major US cities with similar urban characteristics. The high spatio-temporal resolution analysis adopted in this study can be useful for assessing dynamic exposure, i.e., exposure to air pollution (especially traffic-related pollutants) during daily travel and its long-term impacts on human health.

The comparison of our model results with EPA measurements demonstrates the ability to model $PM_{2.5}$ using low-cost sensor data with reasonable consistency and accuracy levels compared to regulatory monitors. This signifies the need to monitor $PM_{2.5}$ using both high-quality regulatory monitors and low-cost sensors to complement one another's efforts and support air quality modeling and applications, as each type of sensor has its own limitations and trade-offs.

CRediT authorship contribution statement

Armita Kar: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Mohammed Ahmed:** Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Andrew A. May:** Conceptualization, Data curation, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Huyen T.K. Le:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Huyen T. K. Le, Andew A. May reports financial support was provided by The Ohio State University Sustainability Institute. Huyen T. K. Le, Andew A. May reports financial support was provided by The Ohio State University Translational Data Anlytics Institute. If there are other authors, they 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

The authors do not have permission to share data.

Acknowledgments

We thank our partners, including Kroger, Ohio State University Wexner Medical Center, Columbus Recreation and Parks Department, Mid-Ohio Regional Planning Commission (MORPC), and other individuals, schools, and business owners for supporting our sensor deployment effort, and Dema Alkashkish for her support in setting up PurpleAir sensors. This project was funded by the Translational Data Analytics Institute (TDAI) and Sustainability Institute at The Ohio State University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2024.120486.

References

Apte, J.S., Marshall, J.D., Cohen, A.J., Brauer, M., 2015. Addressing global mortality from ambient PM2.5. Environ. Sci. Technol. 49 (13), 8057–8066. https://doi.org/10.1021/acs.est.5b01236.

Askariyeh, M.H., Zietsman, J., Autenrieth, R., 2020. Traffic contribution to PM2.5 increment in the near-road environment. Atmos. Environ. 224, 117113. https://doi.org/10.1016/j.atmosenv.2019.117113, 117113.

Barkjohn, K.K., Gantt, B., Clements, A.L., 2021. Development and application of a United States-wide correction for $PM_{2.5}$ data collected with the PurpleAir sensor. Atmos. Meas. Tech. 14 (6), 4617–4637. https://doi.org/10.5194/amt-14-4617-2021.

Benjamin, S.G., Weygandt, S.S., Brown, J.M., Hu, M., Alexander, C.R., Smirnova, T.G., Olson, J.B., James, E.P., Dowell, D.C., Grell, G.A., Lin, H., Peckham, S.E., Smith, T.L., Moninger, W.R., Kenyon, J.S., Manikin, G.S., 2016. A North American hourly assimilation and model forecast cycle: the Rapid Refresh. Mon. Weather Rev. 144 (4), 1669–1694. https://doi.org/10.1175/MWR-D-15-0242.1.

Bi, J., Wildani, A., Chang, H.H., Liu, Y., 2020a. Incorporating low-cost sensor measurements into high-resolution PM2.5 modeling at a large spatial scale. Environ. Sci. Technol. 54 (4), 2152–2162. https://doi.org/10.1021/ACS.EST.9B06046/ASSET/IMAGES/LARGE/ES9B06046_0002.JPEG.

Bi, J., Stowell, J., Seto, E.Y.W., English, P.B., Al-Hamdan, M.Z., Kinney, P.L., Freedman, F.R., Liu, Y., 2020b. Contribution of low-cost sensor measurements to the prediction of PM2.5 levels: a case study in Imperial County, California, USA. Environ. Res. 180, 108810 https://doi.org/10.1016/j.envres.2019.108810.

Bi, J., Carmona, N., Blanco, M.N., Gassett, A.J., Seto, E., Szpiro, A.A., Larson, T.V., Sampson, P.D., Kaufman, J.D., Sheppard, L., 2022. Publicly available low-cost sensor measurements for PM2.5 exposure modeling: guidance for monitor deployment and

data selection. Environ. Int. 158, 106897 https://doi.org/10.1016/j.envint.2021.106897.
Brook, R.D., Rajagopalan, S., Pope, C.A., Brook, J.R., Bhatnagar, A., Diez-Roux, A.V., Holguin, F., Hong, Y., Luepker, R.V., Mittleman, M.A., Peters, A., Siscovick, D., Smith, S.C., Whitsel, L., Kaufman, J.D., 2010. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American heart association. Circulation 121 (21), 2331–2378. https://doi.org/10.1161/CIR.0b013e3181dbece1.
Burnett, R.T., Pope, C.A., Ezzati, M., Olives, C., Lim, S.S., Mehta, S., Shin, H.H., Singh, G., Hubbell, B., Brauer, M., Anderson, H.R., Smith, K.R., Balmes, J.R., Bruce, N.G., Kan, H., Laden, F., Prüss-Ustün, A., Turner, M.C., Gapstur, S.M., et al., 2014. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. Environ. Health Perspect. 122 (4), 397–403. https://doi.org/10.1289/ehp.1307049.
Chang, F.-J., Chang, L.-C., Kang, C.-C., Wang, Y.-S., Huang, A., 2020. Explore spatio-temporal PM2.5 features in northern Taiwan using machine learning techniques. Sci. Total Environ. 736, 139656 https://doi.org/10.1016/j.scitotenv.2020.139656.
Clements, A., Vanderpool, R., 2019. FRMs/FEMs and sensors: complementary approaches for determining ambient air quality. EPA Tools and Resources Webinar, 348237.
Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M.B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L.J., Schwartz, J., 2019. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environ. Int. 130, 104909 https://doi.org/10.1016/j.envint.2019.104909.
Earth Resources Observation and Science (EROS) Center, 2019. National land cover Database. U.S. Geological Survey. https://www.usgs.gov/centers/eros/science/national-land-cover-database.
Gu, P., Li, H.Z., Ye, Q., Robinson, E.S., Apte, J.S., Robinson, A.L., Presto, A.A., 2018. Intracity variability of particulate matter exposure is driven by carbonaceous sources and correlated with land-use variables. Environ. Sci. Technol. 52 (20), 11545–11554. https://doi.org/10.1021/acs.est.8b03833.
Hagan, D.H., Kroll, J.H., 2020. Assessing the accuracy of low-cost optical particle sensors using a physics-based approach. Atmos. Meas. Tech. 13 (11), 6343–6355. https://doi.org/10.5194/amt-13-6343-2020.
He, M., Kuerbanjiang, N., Dhaniyala, S., 2020. Performance characteristics of the low-cost Plantower PMS optical sensor. Aerosol. Sci. Technol. 54 (2), 232–241. https://doi.org/10.1080/02786826.2019.1696015.
Heo, J., Adams, P.J., Gao, H.O., 2016. Public health costs of primary PM2.5 and inorganic PM2.5 precursor emissions in the United States. Environ. Sci. Technol. 50 (11), 6061–6070. https://doi.org/10.1021/acs.est.5b06125.
Hu, J., Huang, L., Chen, M., Liao, H., Zhang, H., Wang, S., Zhang, Q., Ying, Q., 2017. Premature mortality attributable to particulate matter in China: source contributions and responses to reductions. Environ. Sci. Technol. 51 (17), 9950–9959. https://doi.org/10.1021/acs.est.7b03193.
Hu, J., Ostro, B., Zhang, H., Ying, Q., Kleeman, M.J., 2019a. Using chemical transport model predictions to improve exposure assessment of PM2.5 constituents. Environ. Sci. Technol. Lett. 6 (8), 456–461. https://doi.org/10.1021/acs.estlett.9b00396.
Hu, H., Hu, Z., Zhong, K., Xu, J., Zhang, F., Zhao, Y., Wu, P., 2019b. Satellite-based high-resolution mapping of ground-level PM2.5 concentrations over East China using a spatiotemporal regression kriging model. Sci. Total Environ. 672, 479–490. https://doi.org/10.1016/j.scitotenv.2019.03.480.
International Code Council (ICC), 2021. International energy conservation Code (IECC). https://codes.iccsafe.org/content/IECC2021P1/chapter-3-ce-general-requirements.
Jain, S., Presto, A.A., Zimmerman, N., 2021. Spatial modeling of daily PM2.5, NO2, and CO concentrations measured by a low-cost sensor network: comparison of linear, machine learning, and hybrid land use models. Environ. Sci. Technol. 55 (13), 8631–8641. https://doi.org/10.1021/ACS.EST.1C02653.
Kelly, K.E., Whitaker, J., Petty, A., Widmer, C., Dybwad, A., Sleeth, D., Martin, R., Butterfield, A., 2017. Ambient and laboratory evaluation of a low-cost particulate matter sensor. Environ. Pollut. 221, 491–500. https://doi.org/10.1016/j.envpol.2016.12.039.
Kim, J., Kwan, M.-P., 2019. Beyond commuting: ignoring individuals' activity-travel patterns may lead to inaccurate assessments of their exposure to traffic congestion. Int. J. Environ. Res. Publ. Health 16 (1). https://doi.org/10.3390/ijerph16010089. Article 1.
Kloog, I., Nordio, F., Coull, B.A., Schwartz, J., 2012. Incorporating local land use regression and satellite aerosol optical depth in A hybrid model of spatiotemporal PM $_{2.5}$ exposures in the mid-atlantic states. Environ. Sci. Technol. 46 (21), 11913–11921. https://doi.org/10.1021/es302673e.
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 PM2.5 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.
Langford, A.O., Senff, C.J., Alvarez II, R.J., Aikin, K.C., Ahmadov, R., Angevine, W.M., Baidar, S., Brewer, W.A., Brown, S.S., James, E.P., McCarty, B.J., Sandberg, S.P., Zucker, M.L., 2023. Were wildfires responsible for the unusually high surface ozone in Colorado during 2021? J. Geophys. Res. Atmos. 128 (12), e2022JD037700 https://doi.org/10.1029/2022JD037700.
Lee, H.J., 2019. Benefits of high resolution PM2.5 prediction using satellite MAIAC AOD and land use regression for exposure assessment: California examples. Environ. Sci. Technol. 53 (21), 12774–12783. https://doi.org/10.1021/acs.est.9b03799.
Li, H.Z., Dallmann, T.R., Li, X., Gu, P., Presto, A.A., 2018. Urban organic aerosol exposure: spatial variations in composition and source impacts. Environ. Sci. Technol. 52 (2), 415–426. https://doi.org/10.1021/acs.est.7b03674.
Li, J., Zhang, H., Chao, C.-Y., Chien, C.-H., Wu, C.-Y., Luo, C.H., Chen, L.-J., Biswas, P., 2020. Integrating low-cost air quality sensor networks with fixed and satellite monitoring systems to study ground-level PM2.5. Atmos. Environ. 223, 117293 https://doi.org/10.1016/j.atmosenv.2020.117293.
Li, H., Yang, Y., Wang, H., Wang, P., Yue, X., Liao, H., 2022. Projected aerosol changes driven by emissions and climate change using a machine learning method. Environ. Sci. Technol. 56 (7), 3884–3893. https://doi.org/10.1021/acs.est.1c04380.
Li, G., Wu, Z., Liu, N., Liu, X., Wang, Y., Zhang, L., 2023. A variational bayesian blind calibration approach for air quality sensor deployments. IEEE Sensor. J. 23 (7), 7129–7141. https://doi.org/10.1109/JSEN.2022.3212009.
Lin, Y.-C., Chi, W.-J., Lin, Y.-Q., 2020. The improvement of spatial-temporal resolution of PM2.5 estimation based on micro-air quality sensors by using data fusion technique. Environ. Int. 134, 105305 https://doi.org/10.1016/j.envint.2019.105305.
Liu, X., Song, Z., Ngai, E., Ma, J., Wang, W., 2015. PM2:5 monitoring using images from smartphones in participatory sensing. 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 630–635. https://doi.org/10.1109/INFCOMW.2015.7179456.
Liu, S., Geng, G., Xiao, Q., Zheng, Y., Liu, X., Cheng, J., Zhang, Q., 2022. Tracking daily concentrations of PM2.5 chemical composition in China since 2000. Environ. Sci. Technol. 56 (22), 16517–16527. https://doi.org/10.1021/acs.est.2c06510.
Lu, T., Marshall, J.D., Zhang, W., Hystad, P., Kim, S.-Y., Bechle, M.J., Demuzere, M., Hankey, S., 2021a. National empirical models of air pollution using microscale measures of the urban environment. Environ. Sci. Technol. 55 (22), 15519–15530. https://doi.org/10.1021/acs.est.1c04047.
Lu, Y., Giuliano, G., Habre, R., 2021b. Estimating hourly PM2.5 concentrations at the neighborhood scale using a low-cost air sensor network: A Los Angeles case study. Environ. Res. 195, 110653 https://doi.org/10.1016/j.envres.2020.110653.
Lu, T., Bechle, M.J., Wan, Y., Presto, A.A., Hankey, S., 2022. Using crowd-sourced low-cost sensors in a land use regression of PM2.5 in 6 US cities. Air Quality, Atmosphere & Health 15 (4), 667–678. https://doi.org/10.1007/s11869-022-01162-7.
Ma, J., Tao, Y., Kwan, M.-P., Chai, Y., 2020. Assessing mobility-based real-time air pollution exposure in space and time using smart sensors and GPS trajectories in beijing. Ann. Assoc. Am. Geogr. 110 (2), 434–448. https://doi.org/10.1080/24694452.2019.1653752.
Malings, C., Tanzer, R., Hauryliuk, A., Saha, P.K., Robinson, A.L., Presto, A.A., Subramanian, R., 2020. Fine particle mass monitoring with low-cost sensors: corrections and long-term performance evaluation. Aerosol. Sci. Technol. 54 (2), 160–174.
McDuffie, E.E., Martin, R.V., Spadaro, J.V., Burnett, R., Smith, S.J., O'Rourke, P., Hammer, M.S., van Donkelaar, A., Bindle, L., Shah, V., Jaeglé, L., Luo, G., Yu, F., Adeniran, J.A., Lin, J., Brauer, M., 2021. Source sector and fuel contributions to ambient PM2.5 and attributable mortality across multiple spatial scales. Nat. Commun. 12 (1), 3594. https://doi.org/10.1038/s41467-021-23853-y.
Morawska, L., Thai, P.K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., Bedini, A., Chai, F., Christensen, B., Dunbabin, M., Gao, J., Hagler, G.S.W., Jayaratne, R., Kumar, P., Lau, A.K.H., Louie, P.K.K., Mazaheri, M., Ning, Z., Motta, N., et al., 2018. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: how far have they gone? Environ. Int. 116, 286–299. https://doi.org/10.1016/j.envint.2018.04.018.
Murphy, B.N., Sonntag, D., Seltzer, K.M., Pye, H.O.T., Allen, C., Murray, E., Toro, C., Gentner, D.R., Huang, C., Jathar, S., Li, L., May, A.A., Robinson, A.L., 2023. Reactive organic carbon air emissions from mobile sources in the United States. Atmos. Chem. Phys. 23 (20), 13469–13483. https://doi.org/10.5194/acp-23-13469-2023.
O'Lenick, C.R., Winquist, A., Mulholland, J.A., Friberg, M.D., Chang, H.H., Kramer, M.R., Darrow, L.A., Sarnat, S.E., 2017. Assessment of neighbourhood-level socioeconomic status as a modifier of air pollution–asthma associations among children in Atlanta. J. Epidemiol. Community 71 (2), 129–136. https://doi.org/10.1136/jech-2015-206530.
Park, Y.M., Kwan, M.-P., 2017. Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. Health Place 43, 85–94. https://doi.org/10.1016/j.healthplace.2016.10.002.
Pope, C.A., Dockery, D.W., 2006. Health effects of fine particulate air pollution: lines that connect. J. Air Waste Manag. Assoc. 56 (6), 709–742. https://doi.org/10.1080/10473289.2006.10464485.
PurpleAir, 2023. Real-time air quality monitoring. PurpleAir, Inc. https://www2.purpleair.com/.
Reis, S., Seto, E., Northcross, A., Quinn, N.W., Convertino, M., Jones, R.L., Maier, H.R., Schlink, U., Steinle, S., Vieno, M., 2015. Integrating modelling and smart sensors for environmental and human health. Environ. Model. Software 74, 238–246.
Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun. 181 (2), 259–270. https://doi.org/10.1016/j.cpc.2009.09.018.
Sarnat, S.E., Winquist, A., Schauer, J.J., Turner, J.R., Sarnat, J.A., 2015. Fine particulate matter components and emergency department visits for cardiovascular and respiratory diseases in the St. Louis, Missouri–Illinois, metropolitan area. Environ. Health Perspect. 123 (5), 437–444. https://doi.org/10.1289/ehp.1307776.
Schulte, N., Li, X., Ghosh, J.K., Fine, P.M., Epstein, S.A., 2020. Responsive high-resolution air quality index mapping using model , regulatory monitor , and sensor data in real-time OPEN ACCESS Responsive high-resolution air quality index mapping using model. Regulatory Monitor , and Sensor Data in Real-Time.
Sekulić, A., Kilibarda, M., Heuvelink, G.B.M., Nikolić, M., Bajat, B., 2020. Random forest spatial interpolation. Rem. Sens. 12 (10), 1687. https://doi.org/10.3390/rs12101687.
Sinaga, D., Setyawati, W., Cheng, F.Y., Lung, S.-C.C., 2020. Investigation on daily exposure to PM2.5 in Bandung city, Indonesia using low-cost sensor. J. Expo. Sci.

Environ. Epidemiol. 30 (6), 1001–1012. https://doi.org/10.1038/s41370-020-0256-9.

Song, R., Presto, A.A., Saha, P., Zimmerman, N., Ellis, A., Subramanian, R., 2021. Spatial variations in urban air pollution: impacts of diesel bus traffic and restaurant cooking at small scales. Air Quality, Atmosphere & Health 14 (12), 2059–2072. https://doi.org/10.1007/s11869-021-01078-8.

South Coast Air Quality Management District, 2024. Field evaluation Purple air (PA-II) PM sensor. Retrieved March 14. https://www.aqmd.gov/docs/default-source/aq-spec/field-evaluations/purple-air-pa-ii—field-evaluation.pdf?sfvrsn=2.

Taira, G.R., Leal, A.G., Santos, A.S., Park, S.W., 2022. Bayesian neural network-based calibration for urban air quality sensors. In: Montastruc, L., Negny, S. (Eds.), Computer Aided Chemical Engineering 51, 1549–1554. https://doi.org/10.1016/B978-0-323-95879-0.50259-9. Elsevier.

Tan, Y., Lipsky, E.M., Saleh, R., Robinson, A.L., Presto, A.A., 2014. *Characterizing the spatial Variation of air Pollutants and the Contributions of high emitting Vehicles in Pittsburgh, PA* (world). November 21 ACS Publications; American Chemical Society. https://doi.org/10.1021/es5034074 [Research-article].

The University of Utah Mesowest Group, 2022. NOAA high-resolution Rapid Refresh (HRRR) data archive: AWS open data program. https://mesowest.utah.edu/html/hrrr/.

Tryner, J., L'Orange, C., Mehaffy, J., Miller-Lionberg, D., Hofstetter, J.C., Wilson, A., Volckens, J., 2020. Laboratory evaluation of low-cost PurpleAir PM monitors and in-field correction using co-located portable filter samplers. Atmos. Environ. 220, 117067 https://doi.org/10.1016/j.atmosenv.2019.117067.

van Donkelaar, A., Martin, R.V., Spurr, R.J.D., Burnett, R.T., 2015. High-resolution satellite-derived PM2.5 from optimal estimation and geographically weighted regression over North America. Environ. Sci. Technol. 49 (17), 10482–10491. https://doi.org/10.1021/acs.est.5b02076.

Vu, B.N., Bi, J., Wang, W., Huff, A., Kondragunta, S., Liu, Y., 2022. Application of geostationary satellite and high-resolution meteorology data in estimating hourly PM2.5 levels during the Camp Fire episode in California. Rem. Sens. Environ. 271, 112890 https://doi.org/10.1016/j.rse.2022.112890.

Weagle, C.L., Snider, G., Li, C., van Donkelaar, A., Philip, S., Bissonnette, P., Burke, J., Jackson, J., Latimer, R., Stone, E., Abboud, I., Akoshile, C., Anh, N.X., Brook, J.R., Cohen, A., Dong, J., Gibson, M.D., Griffith, D., He, K.B., et al., 2018. Global sources of fine particulate matter: interpretation of PM2.5 chemical composition observed by SPARTAN using a global chemical transport model. Environ. Sci. Technol. 52 (20), 11670–11681. https://doi.org/10.1021/acs.est.8b01658.

Xiang, Y., Tang, Y., Zhu, W., 2016. Mobile sensor network noise reduction and recalibration using a Bayesian network. Atmos. Meas. Tech. 9 (2), 347–357. https://doi.org/10.5194/amt-9-347-2016.

Xie, Y., Wang, Y., Zhang, K., Dong, W., Lv, B., Bai, Y., 2015. Daily estimation of ground-level PM2.5 concentrations over Beijing using 3 km resolution MODIS AOD. Environ. Sci. Technol. 49 (20), 12280–12288. https://doi.org/10.1021/acs.est.5b01413.

Yang, L.H., Hagan, D.H., Rivera-Rios, J.C., Kelp, M.M., Cross, E.S., Peng, Y., Kaiser, J., Williams, L.R., Croteau, P.L., Jayne, J.T., Ng, N.L., 2022. Investigating the sources of urban air pollution using low-cost air quality sensors at an urban Atlanta site. Environ. Sci. Technol. 56 (11), 7063–7073. https://doi.org/10.1021/ACS.EST.1C07005/ASSET/IMAGES/LARGE/ES1C07005_0003.JPEG.

Zou, Y., Clark, J.D., May, A.A., 2021. Laboratory evaluation of the effects of particle size and composition on the performance of integrated devices containing Plantower particle sensors. Aerosol. Sci. Technol. 55 (7), 848–858. https://doi.org/10.1080/02786826.2021.1905148.