Environmental Science & Technology

pubs.acs.org/est

Article

National Exposure Models for Source-Specific Primary Particulate Matter Concentrations Using Aerosol Mass Spectrometry Data

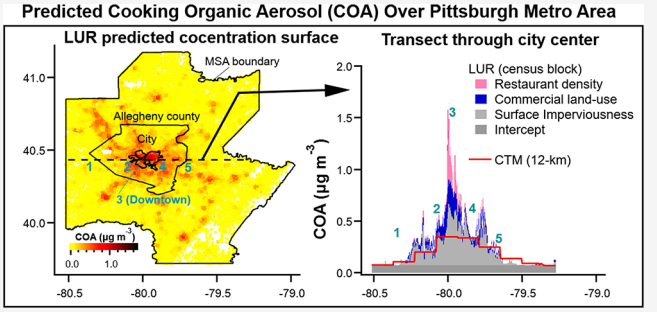
Provat K. Saha, Albert A. Presto,* Steve Hankey, Benjamin N. Murphy, Chris Allen, Wenwen Zhang, Julian D. Marshall, and Allen L. Robinson*

Cite This: *Environ. Sci. Technol.* 2022, 56, 14284−14295

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: This paper investigates the feasibility of developing national empirical models to predict ambient concentrations of sparsely monitored air pollutants at high spatial resolution. We used a data set of cooking organic aerosol (COA) and hydrocarbon-like organic aerosol (HOA; traffic primary organic PM) measured using aerosol mass spectrometry across the continental United States. The monitoring locations were selected to span the national distribution of land-use and source-activity variables commonly used for land-use regression modeling (e.g., road length, restaurant count, etc.). The models explain about 60% of the spatial variability of the measured data (R^2 0.63 for the COA model and 0.62 for the HOA model). Extensive cross-validation suggests that the models are robust with reasonable transferability. The models predict large urban−rural and intra-urban variability with hotspots in urban areas and along the road corridors. The predicted national concentration surfaces show reasonable spatial correlation with source-specific national chemical transport model (CTM) simulations (R^2: 0.45 for COA, 0.4 for HOA). Our measured data, empirical models, and CTM predictions all show that COA concentrations are about two times higher than HOA. Since COA and HOA are important contributors to the intra-urban spatial variability of the total $PM_{2.5}$, our results highlight the potential importance of controlling commercial cooking emissions for air quality management in the United States.



KEYWORDS: *fine particulate matter, spatial modeling, aerosol mass spectrometry*

1. INTRODUCTION

Over the past two decades, researchers have used land-use regression and other empirical models to predict air pollution concentrations at high (census block or zip code) spatial resolution at the country or continental scale.[1−6] These models use a variety of approaches that combine satellite measurements, chemical transport model (CTM) simulations, and/or land-use measures to obtain spatially and temporally resolved estimates of ground-level concentrations.[7−14] These models have enabled investigation of air pollution health impacts[15−17] and exposure disparities[18−20] in very large populations. These studies have provided substantial insight into the health impacts of pollutant concentrations among different populations.

Deriving empirical exposure models requires large amounts of air pollution monitoring data. Therefore, national-scale empirical exposure models were first developed for criteria pollutants ($PM_{2.5}$, NO_2, etc.), which have large, well-developed regulatory monitoring networks. Few national-scale empirical models have been developed for pollutants that are costly to monitor, and thus, national networks do not exist (e.g., ultrafine particles, air toxics, and particulate matter components). There is substantial interest in the potential health impacts of these types of air pollutants. Of particular interest are source and chemically specific subcomponents of fine particulate matter ($PM_{2.5}$),[21−30] which could disproportionally contribute to the adverse health associated with $PM_{2.5}$ mass.

The lack of monitoring data is a major challenge for developing large-spatial scale empirical models for source- and component-specific $PM_{2.5}$. National empirical model predictions exist for the total $PM_{2.5}$ concentrations, but they typically are not speciated. A few studies have developed empirical models for $PM_{2.5}$ components (e.g., OC, EC, SO_4, and NO_3) using chemically resolved measurements, land-use regression, and chemical transport modeling,[31−33] but they are not source resolved. CTMs have been used to estimate source-resolved $PM_{2.5}$ concentrations,[34−36] but these simulations typically have relatively coarse spatial resolution for a national-scale simulation. There is a paucity of high-spatial-resolution

Received: May 12, 2022
Revised: September 12, 2022
Accepted: September 12, 2022
Published: September 26, 2022

exposure estimates for source-resolved or novel markers of PM, especially at the national scale.

High-resolution aerosol mass spectrometry (HR-AMS) provides nearly real-time chemical characterization of major organic and inorganic species in $PM_{2.5}$.[37,38] Source apportionment of the HR-AMS mass spectra using positive matrix factorization (PMF) provides concentrations of source-specific organic aerosols (e.g., traffic, cooking, biomass burning, and secondary organic aerosols).[39] Over the past decades, this measurement and source apportionment technique has been widely applied in atmospheric chemistry research.[39−41] The HR-AMS is a complex and expensive instrument; its application in air pollution spatial modeling and exposure assessment is rare.

The goal of this paper is to investigate the feasibility of developing national-scale models for source-specific primary components of $PM_{2.5}$ using a relatively sparse (compared to regulatory networks) HR-AMS data set. We have developed national models for primary organic aerosols from two important urban sources: traffic and cooking. Recent measurements in several U.S. cities[42−44] showed that these two primary $PM_{2.5}$ components contribute 5−25% of the overall $PM_{2.5}$ mass and are major contributors (more than 50%) to intra-urban spatial variability of the total $PM_{2.5}$. It is noted that we develop models for two primary PM components (HOA and COA). Other important primary $PM_{2.5}$ components include biomass-burning organic aerosol (BBOA) and black carbon (BC), which are not modeled here.

In this paper, we used a nationally representative HR-AMS data set and a supervised linear regression technique to develop national models to predict traffic and cooking primary organic aerosol concentrations at high spatial resolution. The specific objectives of this paper are to (a) examine the representativeness of the measured HR-AMS data set for developing national models, (b) develop models for traffic and cooking primary organic aerosols, (c) investigate the robustness and transferability of the models, (d) apply cross-validated models for predicting national concentration surfaces, and (e) compare empirical model predicted concentration surfaces with CTM predictions.

2. MATERIALS AND METHODS

2.1. Air Pollution Data Set. Empirical models that predict the spatial variability of pollution concentrations are derived by fitting measured data that capture the spatial variability of both the dependent variables (concentrations) and the independent (predictor) variables.[45] Previous studies[45−47] demonstrate that beyond a certain number of locations, additional sites with similar land-use characteristics do not add much additional value for empirical land-use regression model development. Capturing the spatial variability is more important than gross data coverage (i.e., the total number of points used for model training). For national model development, monitoring locations need to capture intra-urban, inter-urban, and urban−rural spatial variability. Using the above principle, we previously developed a national model for ultrafine particle number concentrations in the United States using a relatively sparse data set.[48] In this paper, we applied the same approach to develop national models for source-specific $PM_{2.5}$ components using targeted HR-AMS measurements.

The HR-AMS data set includes intra-urban, urban, and rural background measurements, collected using a hybrid (mobile and stationary) sampling approach. We compiled this data set from our recent mobile measurements as well as data reported in the literature. The data set includes urban and rural concentration measurements in 12 states (AL, CA, CO, GA, IL, MD, MI, NY, OK, PA, TN, and TX). It includes data from 11 cities (Atlanta, Baltimore, Conroe, Fort Worth, Fresno, Houston, New York City, Oakland, Pasadena, Pittsburgh, and St. Louis) and 11 suburban/remote locations across the continental United States. We briefly describe the data set below.

To characterize intra-urban spatial variabilities, we performed mobile measurements in three cities: Pittsburgh, PA; Oakland, CA; and Baltimore, MD. A mobile laboratory with an Aerodyne HR-AMS was slowly and repeatedly driven on predefined sampling routes along many streets in each city over multiple days: Oakland (20 days, 2017), Pittsburgh (33 days, 2016−2017), and Baltimore (9 days, 2019). The driving routes in each city were selected to span a range of land-use and source-activity variables commonly used for land-use regression models. To characterize a representative stable mean, mobile monitoring data were collected covering different times of the day (i.e., the same street was repeatedly visited covering the morning, midday, and evening). We derived organic PM concentrations of traffic, cooking, and secondary organic aerosol by performing PMF analysis of the HR-AMS data. In the AMS source apportionment literature,[39] the traffic factor is commonly referred to as hydrocarbon-like organic aerosol (HOA) and the cooking factor as cooking organic aerosol (COA). We use the same terminology in this paper. The HR-AMS results for the Pittsburgh[49] and Oakland[43,50] data sets are described in previous publications. Similar methods were used for the collection and analysis of the Baltimore data.

We averaged the mobile monitoring data over space and time to characterize spatial patterns of long-term concentrations. This analysis was done using two different grids: 200 m and 1 km. To determine the average concentration in each grid cell, we first calculated the median concentration measured for each sampling day and then calculated the average of the daily medians across all sampling days. This procedure was done separately for each grid cell.

Past mobile monitoring studies indicate that between 7 and 15 days of repeated measurements are needed to characterize representative concentration distributions.[51,52] Therefore, we only considered grid cells with data collected on seven or more days (range 7−33 days). Application of this selection criterion resulted in average concentration estimates for 37 1 km grid cells in Pittsburgh, 31 in Oakland, and 15 in Baltimore. For 200 m grid cells, average concentration values were available for 243 locations in Pittsburgh, 378 in Oakland, and 108 in Baltimore.

In addition to the detailed intra-urban mobile sampling data in three cities, we compiled HR-AMS measured PMF factors from various field campaigns in the United States. These are stationary measurements made at both urban and rural background locations for a period of 1 month or more. They provide information on the inter-urban and urban−rural concentration differences. Table S1 provides details on these field campaigns. PMF analysis of the HR-AMS spectra identified HOA and COA factors at all urban sites. However, HOA and COA factors were not identified at most remote background/forested locations. This is not surprising because HOA/COA concentrations are expected to be very low at remote locations (below detection limit). For the remote sites,

https://doi.org/10.1021/acs.est.2c03398
Environ. Sci. Technol. 2022, 56, 14284−14295

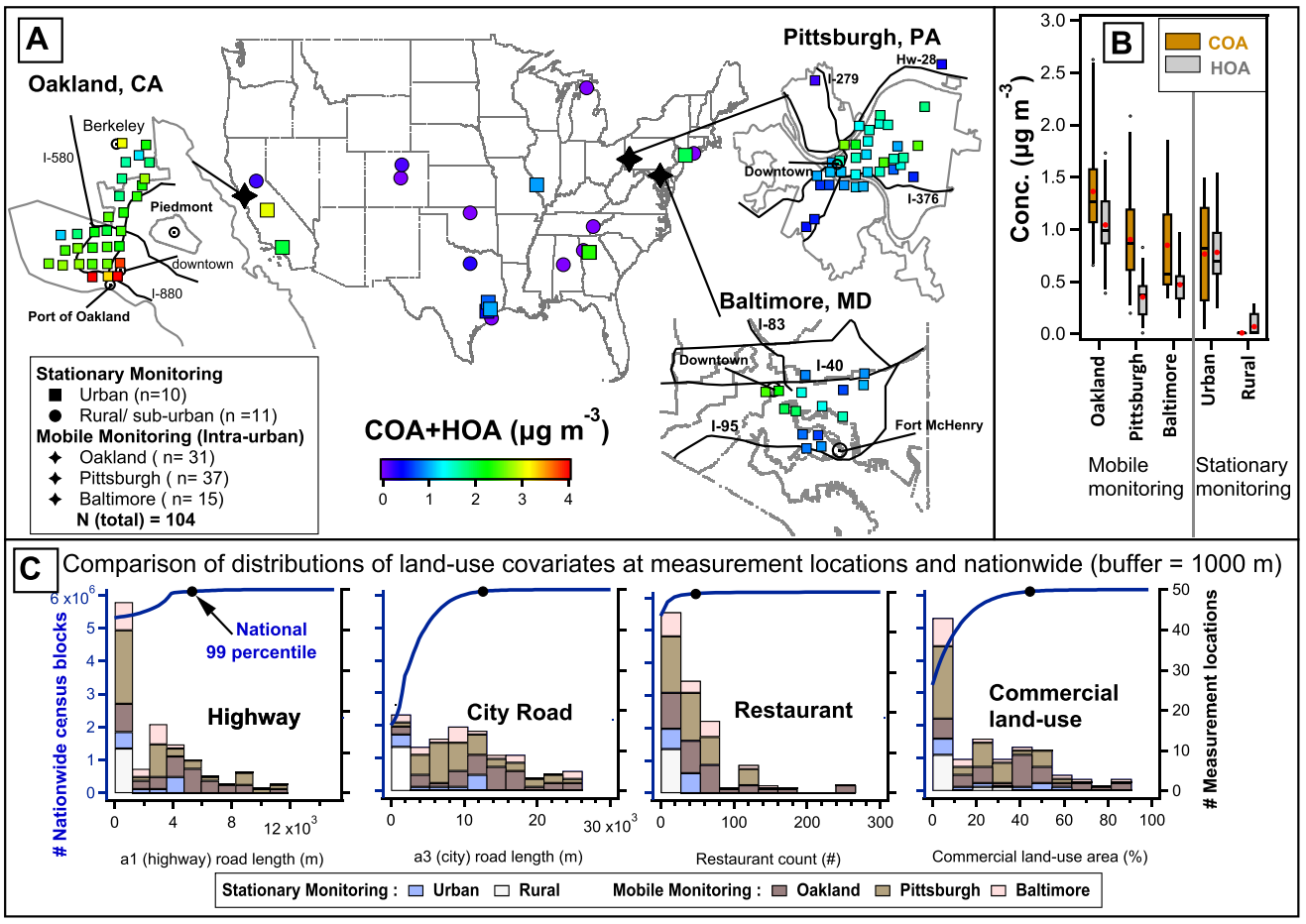


Figure 1. Measured cooking (COA) and traffic (HOA) organic $PM_{2.5}$ concentrations. (A) Measurement locations colored by COA + HOA concentrations. The inset maps in (A) show the mobile monitoring data in three cities (Oakland, CA; Pittsburgh, PA; and Baltimore, MD) on a 1 km grid. (B) Box−whisker plots of distribution of COA and HOA concentrations measured at different geographical location groups (intra-urban: Oakland, Pittsburgh, Baltimore; point locations in other urban and rural areas). (C) Comparison of potential predictor land-use variables at measurement locations and nationwide census-block-level distribution. Each plot shows the histograms (bar) of variable values across measurement locations in 10 equally spaced bins and a cumulative frequency distribution (CDF; blue curve) of that variable across all national census blocks (n = 6,174,588). The black circle on the blue curve is the 99th percentile value of the national distribution. Figure S1 shows a similar figure with the mobile monitoring data aggregated on a 200 m grid.

if COA and HOA are not reported, we assigned a near-zero concentration of 0.01 $\mu g\ m^{-3}$ at these sites for model development.

2.2. Quality Assurance. Since the data set includes measurements made by different groups and at different times, we applied various selection criteria and corrections to ensure temporal and spatial representativeness of the data. We briefly describe these below. Details are given in the SI (Figures S2−S5).

While most of the data used for our model building is from 2016 to 2017, a subset is from different periods and/or collected using different sampling platforms. We applied temporal correction factors to data not measured in 2016−2017 (see Figure S2 for details). These correction factors are based on national trends in traffic- and cooking-related pollutant concentrations and emissions relative to 2017. These corrections are needed for a small subset (∼15%) of total data set. The applied correction factors were relatively small. For HOA, they varied between 0.87 and 1.04 (i.e., between −13% and +4%), and for COA, 0.93 and 1.3 (i.e., between −7% and +30%).

The data were collected in different seasons, and the mobile monitoring was performed during the daytime (between 7 AM and 7 PM). We analyzed a comprehensive data set from Pittsburgh to investigate the potential uncertainty created by day−night and seasonal patterns. Pittsburgh mobile measurements were collected over 5 months in two seasons; continuous data at an urban background location was collected for a subset of these 5 months (for those days, mobile monitoring was not conducted). Pittsburgh continuous data (see Figure S3) indicates a substantial diurnal variation in measured COA and HOA concentrations, e.g., elevated concentrations during morning and evening rush hours. While comparing the 24 h versus daytime (7 AM to 7 PM) average, the daytime is 5−10 % higher. In Pittsburgh mobile monitoring, the difference between summer and winter in HOA and COA concentrations is 5−15% (Figure S4). We also utilized CTM-simulated concentrations (discussed in Section 2.5) to investigate the seasonal variation of COA and HOA. For COA national mean, seasonal averages vary between −10% and +20% of the annual average (for HOA, −17% and +45%) (Figure S5). These comparisons suggest that the effects of diurnal and seasonal variation are relatively small compared

https://doi.org/10.1021/acs.est.2c03398
Environ. Sci. Technol. 2022, 56, 14284−14295

to a factor of 10 spatial variations in measured concentration data (Figure 1).

2.3. Land-Use Regression Modeling. We compiled a large data set of land-use and source-activity variables at high spatial resolution to investigate the spatial patterns in HR-AMS-derived COA and HOA concentrations. A complete list of the variables is shown in Table S2. The majority of the variables are from Kim et al.[6] who compiled a large database of traffic-specific, land-use type, imperviousness of land surface, criteria pollutant emissions, satellite-based air pollution estimates, surface elevation, and other data at the centroid of each census block across the contiguous United States for different buffer sizes (100 m to 15 km). For this work, we created a new national data set of restaurant activity (i.e., restaurant count within various buffer sizes) at the centroid of each census block using publicly available point of interest (POI) data from Yelp. Since our mobile data from three cities were aggregated over two different grids, we averaged the covariate data over each grid cell and then assigned that value to the grid-cell center. For the fixed sites, we assigned the nearest census block's covariate data to the measurement location.

To quantify the relationship between the measured concentrations and the source-activity/land-use variables, we used a supervised stepwise linear regression (commonly referred to as a land-use regression (LUR)) similar to the ESCAPE protocol.[47,53,54] The approach systematically identifies which variables are candidates for inclusion in the model (i.e., correlated with the measured spatial patterns in pollutant concentrations) and the fraction of the measured concentration variability explained by each of these variables. The output of the analysis is a multi-linear regression model that describes the spatial patterns in the measured pollutant concentrations as a function of the source-activity and land-use covariates.

Model covariates are selected based on adjusted R^2 (coefficient of determination) of univariate linear regressions, starting with the covariate that provides the highest adjusted R^2. The next covariate selected is the one that provides the highest adjusted R^2 with the model residuals. We repeat this procedure until including the best of the remaining variables improves the overall model adjusted R^2 by less than 1%. Then, among the selected covariates, we removed the ones with p-values (predictor significance) greater than 0.05. We characterized the model performance in terms of adjusted R^2 and root-mean-square error (RMSE). We evaluated the performance of LUR models using 10-fold cross-validation (CV).

We created separate LUR models using the 1 km and 200 m grid averaged mobile monitoring data. The model derived from the 1 km grid data is the base model. It is fit to 104 data points (Pittsburgh, 37; Oakland, 31; and Baltimore, 15; urban fixed sites, 10; rural fixed sites, 11). We used the 200 m data to investigate the sensitivity of grid resolution. However, models based on the entire 200 m data set had intercepts that were much larger than the measured rural concentrations. This occurred because urban concentrations were overrepresented in the 200 m data set (11 rural locations versus more than 700 urban locations). We used two approaches to overcome this problem. First, we randomly divided 200 m grid data from each city into nine folds and developed nine models using each fold from each city and all of the fixed-site data. Each of these 200 m models is fit to 102 data points. Second, we fit a model using all 200 m grid averaged data along with fixed-site data (it is fit to 751 data points) and forced the intercept of the regression model through zero. This is a reasonable constraint because data from rural locations indicate that the intercept (approximately the rural background level) for HOA/COA national models should be close to zero. We compared the predicted concentration surfaces for the three sets of models.

2.4. Assessment of Robustness and Transferability of LUR Models. We performed extensive analysis to evaluate the robustness and transferability of the LUR models. Specifically, we systematically developed a series of LUR models using different subsets of the air pollution concentration data and then evaluated the performance of these models against the entire data set. We use this approach to examine the following questions: (1) Are the models highly sensitive to specific subsets of the input data? (2) How sensitive are model predictions to the number of locations used for model building? (3) How robust are model predictions in locations without any training data?

To examine the transferability, we systematically developed models by removing data from a particular city and then applied that model to predict the measured concentrations in the holdout city. Our assessment criteria assume that if the underlying model-building data set has the power to predict concentrations in a city/area without any training data from that city/area, that indicates good model transferability.

2.5. Chemical Transport Model Simulations. We compare the results of the national-scale model to bottom-up predictions of COA and HOA using the regional-scale CTM CMAQv5.3.3 (Community Multiscale Air Quality model; U.S. EPA)[55] for the continental United States at 12 km horizontal resolution for the full year 2016. Details on CTM simulation are given in the SI (Section S1) and briefly described here.

CMAQ used input data developed for the EPA Air QUAlity TimE Series (EQUATES) Project,[56] including meteorology from the Weather Research and Forecasting (WRF) model v4.1.1, anthropogenic emissions developed for 2016, and land-use parameters needed for modeling pollutant deposition as well as bidirectional volatilization of NH_3 and energy transfer between the atmosphere and the underlying land surface. Boundary conditions were constrained by a coarse-resolution (108 km) CMAQ simulation over the entire northern hemisphere, and the model was initialized on December 22, 2015 (i.e., a 10-day model spin-up). CMAQ was run with all default parameters (base case) and once each with organic particulate emissions from a key source neglected. These sources included cooking sources, on-road vehicles, and all on-road and non-road mobile sources. With these four cases, source-based POA concentrations were calculated as the difference between each run with neglected emissions and the base simulation.

3. RESULTS AND DISCUSSION

3.1. Spatial Variability in Measured Source-Specific PM Components (COA and HOA). The HR-AMS measured COA and HOA concentration data set used for LUR model development is shown in Figure 1 (mobile monitoring data averaged on a 1 km grid). Figure S1 shows the 200 m grid averaged mobile data. There is substantial spatial variability in COA and HOA concentrations. Urban concentrations are a factor of 4−10 higher than rural concentrations. The measured urban COA and HOA levels constitute a significant fraction of

https://doi.org/10.1021/acs.est.2c03398
Environ. Sci. Technol. 2022, 56, 14284−14295

the total $PM_{2.5}$ mass relative to both the new WHO guideline and the current U.S. EPA NAAQS.

Spatially dense mobile sampling data indicate a factor of 3−12 variability within urban areas, estimated as the ratio of 95th and 5th percentiles values of measured COA and HOA concentrations in different cities. Within a city, concentrations are higher in downtown and business districts highlighting the strong influence of local sources on intra-urban spatial variability of primary $PM_{2.5}$ components. The within-urban area spatial variability is substantially larger than between-city variability (a factor of 1.5−3) for both HOA and COA. As expected, the intra-city distribution of concentrations is wider for the 200 m grid averaged data (Figure S1). COA concentrations are higher than HOA in each sampled city by a factor of 1.3, 1.8, and 2.6 in Oakland, Baltimore, and Pittsburgh, respectively. This means that the primary organic $PM_{2.5}$ concentrations from cooking are 30−160% higher than traffic in these sampled cities.

3.2. Association between Source-Specific PM Concentrations and Relevant Land-Use Variables. Figure 2

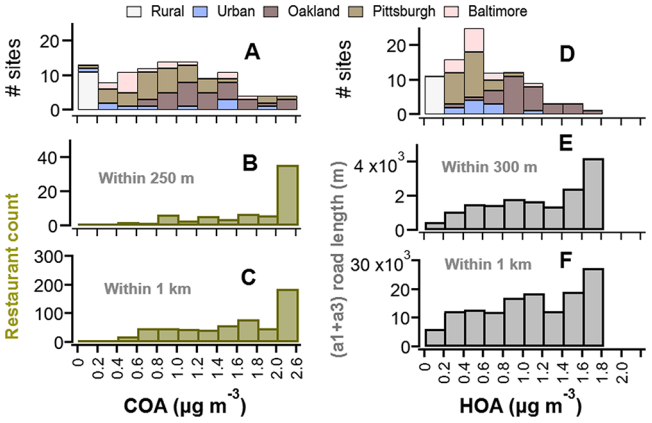


Figure 2. Comparison of measured source-specific $PM_{2.5}$ concentrations and relevant land-use variables in measurement locations. (A) Histogram of measured COA concentrations. (B and C) average restaurant density with (B) 250 m and (C) 1 km of measurement locations for each concentration bin. (D) Histogram of measured HOA concentrations. (E and F) Same as (B and C) but for road density (a1 and a3 road length within 300 m and 1 km buffer radius from measurement locations). a1 road is the highway, and a3 is the city arterial road.

shows that the concentrations of the HR-AMS-derived source-resolved $PM_{2.5}$ components are correlated with source-specific land-use variables. For example, COA concentrations increase with restaurant density. The univariate correlation coefficient (Pearson R) between COA and restaurant count within 1 km is 0.65; the correlation with restaurant count within 250 m is 0.45. HOA concentrations increase with road density. The univariate correlation coefficient between HOA and road length (highway + city road) within 1 km and 300 m is 0.47 and 0.50, respectively. The relationships using the 200 m grid averaged mobile data and source-specific land-use variables are similar (Figure S7).

Although the absolute concentrations of the source-resolved components vary by city, the measured COA and HOA mass spectra are very similar.[43,49] This suggests that the composition of cooking- and traffic-related organic aerosols is similar across American cities. However, the level of activity can lead to different concentration levels. This suggests that if one can collect source-resolved monitoring data covering the possible range of variability across the national domain, it may be possible to build a robust national model. Figure 2 highlights that spatially dense sampling is required to capture intra-city variability in both primary PM concentrations and land-use covariates. The data from the 10 urban fixed sites do not capture the wide range of intra-urban variability (e.g., the blue bars in Figure 2 do not capture the high end of measured concentrations and land-use covariates). Therefore, the mobile data are critical for capturing the intra-urban trends and the most source-impacted areas.

3.3. Representativeness of Concentration Data Set for Developing National LUR Models. To develop an LUR model, it is important to collect the monitoring data across the entire range of pollutant exposure and land-use/source-activity variable values (e.g., national distribution in our case). Figure 1C compares the traffic and cooking variables at our monitoring locations to the national distribution. The measurement locations span the 0−99th percentile of the national data. The basic trends and conclusions are similar for all variables (Figures S8 and S9); our monitoring locations span the range of national distributions of relevant source-activity variables. These comparisons demonstrate that our measurement locations span the entire national range of land-use and source-activity variables.

The concentration data set also has reasonable geographical coverage. Measurements were made in 12 states, 11 cities, and 11 suburban and remote locations across the continental United States. For model development, we used 104 (using 1 km grid averaged mobile data) to 751 (using 200 km grid averaged mobile data) data points. This amount of data is similar to data sets used to develop national models for criteria air pollutants. For example, Kim et al.[6] developed national models for criteria air pollutants in the United States using regulatory monitoring data from a few hundred locations (e.g., CO, 218; NO_2, 327; SO_2, 370; O_3, 850, and $PM_{2.5}$, 934, for the modeling year 2010) along with satellite remote sensing data. A key aspect of our strategy is to utilize mobile monitoring data to characterize the large intra-urban spatial gradients in COA and HOA concentrations. In comparison, regulatory monitors are subject to detailed siting criteria (typically located in urban background and rural locations) and therefore may span less of the covariate space than our modestly smaller data set.

3.4. LUR Models for Source-Specific PM Concentrations. The combination of the correlation between the measured concentrations and source-specific land-use variables (Figure 2) and the fact that our measurement locations span the national land-use/source-activity variable space (Figure 1C) provides confidence that we can derive a reasonable national model for predicting COA and HOA spatial patterns across the continental United States. We developed multi-parameter regression models (LUR models) to describe the spatial patterns in the measured COA and HOA concentrations as a function of the source-activity and land-use covariates. Separate models were developed using 1 km and 200 m grid averaged mobile monitoring data. Figure 3 and Figure S10 show the measured versus predicted concentrations for models developed using 1 km and 200 m grid averaged mobile monitoring data, respectively. Tables S3 and S4 summarize the predictor covariates and performance metrics of all these models.

https://doi.org/10.1021/acs.est.2c03398
Environ. Sci. Technol. 2022, 56, 14284−14295

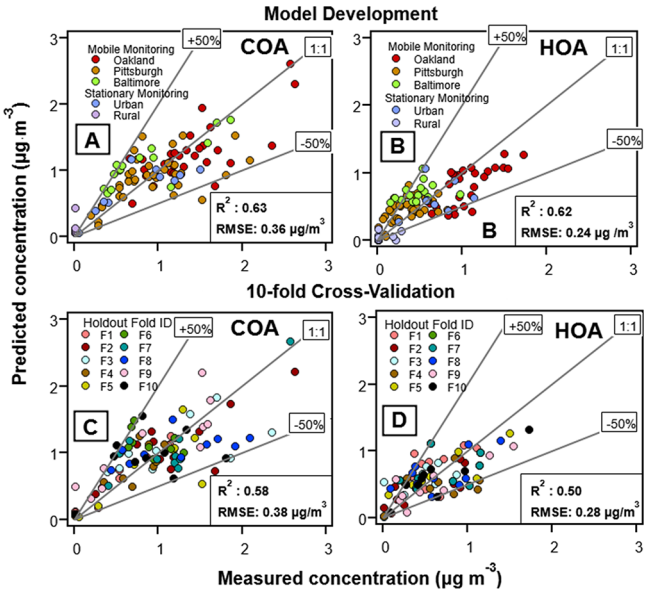


Figure 3. Measured versus predicted concentrations from LUR models. (A and B) Models based on all data. (C and D) Predictions at holdout locations from random 10-fold CV models.

The models selected four to five predictor variables, indicating that they are not overfitted. The selected variables are indicators for relevant sources and land uses (Tables S3 and S4). They are therefore physically interpretable, which highlights the value of fitting source-resolved $PM_{2.5}$ components. For example, the model selects restaurant density, commercial land-use, and urbanicity-related variables to explain the spatial variability of COA. The model selects traffic, transportation land-use, and urbanicity-related variables to explain the spatial variability of HOA. The variable "impervious land surface" is present in most models, which likely captures the urban−rural gradients. Some of these variables are well correlated with population density (e.g., impervious land surface and restaurant density). All models select variables with a range of buffer sizes. The variables with small buffer distances (e.g., restaurant count within 100 m and highway length within 150 m) describe the near-source variation, whereas variables with larger buffer distances (e.g., restaurant count within 1000 m and the commercial land-use area within 1500 m) likely reflect the neighborhood/city-scale and urban/rural variations.

The R^2 values of the COA and HOA LUR base models (derived using the 1 km grid averaged mobile monitoring data) are 0.63 and 0.62, respectively; RMSE are 0.36 and 0.24 μg m^{-3}, respectively (Figure 3). The R^2 values of the models developed for random 10-fold CV are 7−15% lower than the model fits of the entire data set (Figure 3).

The R^2 values of the COA and HOA models using all 200 m mobile data (n = 751) are 0.52 and 0.51, respectively, with RMSE of 0.39 and 0.34 μg m^{-3}, respectively. The random 10-fold CV R^2 values are almost same as the models fit using the entire data set (Figure S10). The R^2 values of the models using a different subset of the 200 m grid data range from 0.6 to 0.76 for COA and 0.5 to 0.67 for HOA (see Tables S3 and S4); the 10-fold CV R^2 values of the models are 5−10% lower than fit R^2.

There were only slight differences in the performance metrics (e.g., R^2 values) between models fit to the entire data set and the 10-fold CV (Figure 3 and Figure S10). This indicates that the models are not overly sensitive to random subsets of the data and fits are statistically robust. In addition, across the base and range of sensitivity cases, available predictor variables consistently predict more than 50% of the measured variability. This level of model performance is typically considered good in the LUR literature.[45]

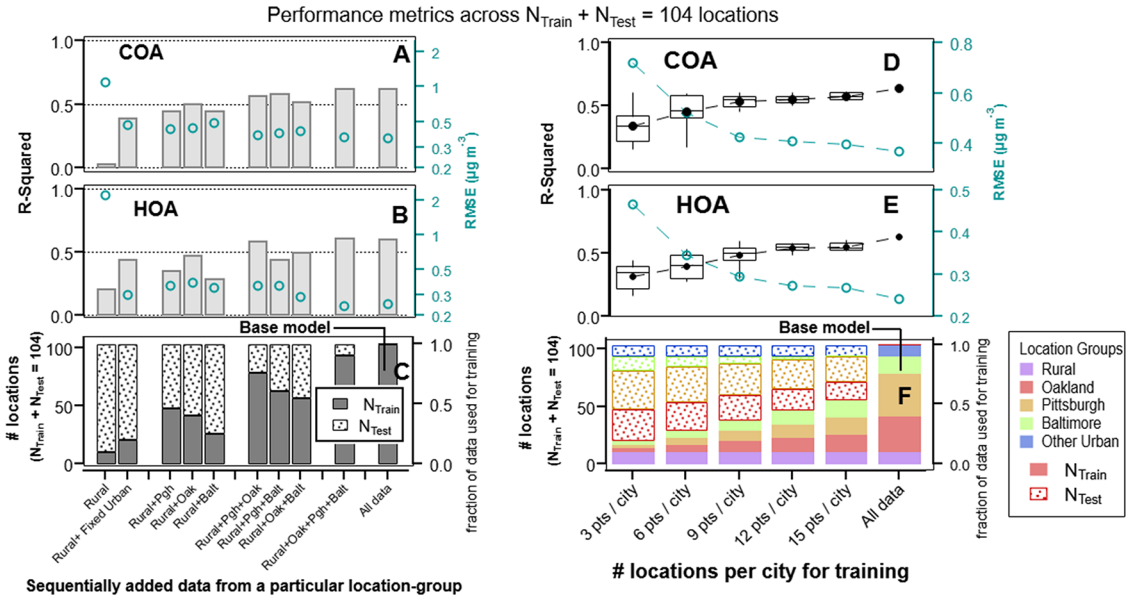


Figure 4. Predictive performance of LUR models developed using different subsets of the data. (A−C) Models fit using different geographical location-based subgroups (rural fixed sites, urban fixed sites, Pittsburgh, Oakland, and Baltimore). (D−F) Models fit using a subset of data from each mobile sampling city along with rural fixed sites. The box−whisker plots in panels (D and E) show the distribution of R^2 from 10 random selections, and the circles are the mean; only means are shown for RMSE. The 1 km grid averaged mobile data are used in this analysis. Panels (C and F) show the number of locations used for training (N_{Train}) and testing (N_{Test}) for different cases. The R^2 and RMSE are calculated using all of the measured data (fitting and holdout).

https://doi.org/10.1021/acs.est.2c03398
Environ. Sci. Technol. 2022, 56, 14284−14295

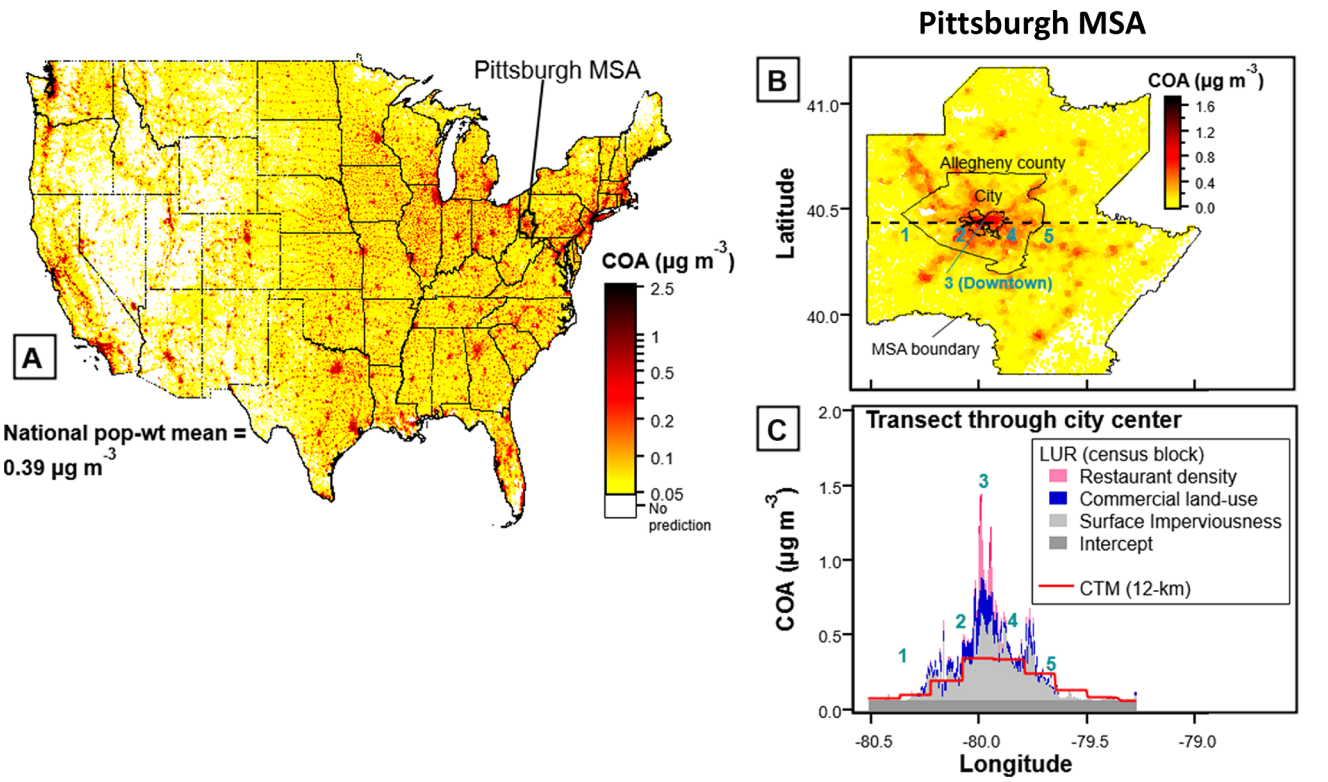


Figure 5. LUR-predicted COA concentrations and comparison with CTM. (A) Predictions across the contiguous United States at the census block. (B) Zoom-in over the Pittsburgh MSA. (C) A concentration profile along a transect across the Pittsburgh MSA that passes through the city center (downtown Pittsburgh). The 12 km CTM-simulated concentrations along the transect line are also shown in (C). Figure S13 shows the same plots for HOA.

3.5. Robustness and Transferability of LUR Models.

This section rigorously evaluates the predictive power of LUR models by systematically building many models with different subsets of the data. The models are then applied to predict concentrations at the remaining (holdout) locations. The models are built by incrementally including additional data from various subgroups. Subgroups are created using different criteria: geographical location-based groups (rural fixed sties, urban fixed sites, Pittsburgh, Oakland, and Baltimore), land-use-based groups (e.g., low, medium, and high restaurant density locations), etc. In all cases, the performance metrics (R^2 and RMSE) of the subsampled models are calculated by comparing predictions to all measured concentrations. The results are compared with the base model, which fits the entire data set (Figure 3A,B). The results from these analyses are summarized in Figure 4 and Figure S11.

Figure 4A–C shows that, as expected, a model only fit to the rural data poorly predicts concentrations in urban holdout sites. Performance improves with additional training data from various subgroups: fixed-urban, intra-city data from one or two cities, etc.

Figure 4D–F shows the predictive performance of various LUR models developed by incrementally increasing the amount of mobile monitoring data used to fit the model. The model performance gradually improves as additional data are included. However, after including 10–15 data points from each mobile monitored in the city (about 40–60% of overall data from the entire data set), the predictive performance of a model becomes close to the model fit to the whole data set. Figure S11 shows a similar analysis by incrementally including data from different land-use variables. The basic trends and conclusions are similar across analyses using multiple grouping approaches; a model using a subset of locations covering rural, intra-, and inter-urban (40–60% of the entire data) data can reasonably predict concentrations at holdout locations.

The results in Figure 4 also highlight the importance of the mobile monitoring. For example, a model fit with the mobile monitoring data from three cities reasonably predicts fixed-site urban background measurements, but a model based on the urban background data poorly predicts intra-urban variability characterized by the mobile monitoring. This is because the mobile monitoring locations overlap with the land-use and source-activity covariate space of urban background locations, but not vice versa (Figure 2).

We also examined the transferability of LUR models through spatial (city) holdouts. We have conducted these holdout experiments for different subgroups of the entire data set (e.g., holding out all Oakland data points, all Pittsburgh, all Baltimore, all urban fixed sites, etc.) (see Figure S12). In most cases, the model predicts spatial variability in a holdout city with a moderate R^2. In one case (HOA prediction in Oakland when holding out all of the Oakland data), there are systematic biases with the model predicting intra-urban variability but not the right absolute level. Oakland had the highest measured HOA concentrations (Figure 2). Holding out Oakland data from model fitting yields lower performance measures and subsequent underpredictions when applying that model to predict the measured Oakland concentrations (Figure S12).

The HOA concentrations in Oakland measured were substantially higher than other cities, but the values of the available traffic variables (e.g., road length) were across all

https://doi.org/10.1021/acs.est.2c03398
Environ. Sci. Technol. 2022, 56, 14284–14295

cities (Figure 2). This suggests that the nationally available traffic-related variable data (e.g., road length) may not capture measured hotspots well. Improved traffic composition and activity-specific variables are required to better capture the near-source and intra-city hotspots.

In summary, the predictive performance of the various models developed using different subsets of the entire data set (discussed in Figure 4 and Figures S12 and S13) suggests that our monitoring data set and available land-use and source-activity variables capture the broad spatial patterns in COA and HOA across the United States. However, improved covariates (activity-specific) and additional data from other cities may further improve the predictions.

3.6. Predicted National Concentration Surfaces. The national representativeness of our model-building data (Figure 1C) and extensive sensitivity analysis (Figure 4 and Figures S11 and S12) suggest that our LUR models can provide reasonable prediction surfaces at high spatial resolution. Figure 5A shows the base-model predicted COA concentrations at census-block resolution across the contiguous United States. The model equation is COA = 0.07 + 0.008 × impervious surface area (1000 m) + 0.004 × restaurant count (1000 m) + 0.011 × commercial landuse area (3000 m) + 0.06 × restaurant count (100 m). Figure S13 shows the HOA predictions at the census block using the base model: HOA = 0.003 + 0.005 × commercial landuse area (1500 m) + 0.0003 × $a1$ (highway) road length (150 m) + 0.005 × impervious surface area (750 m) + 0.02 × transportation landuse area (3000 m).

Figure S14 compares the predicted concentrations for different models (1 km grid, 200 m grid, and subset of 200 m grid). The concentration surfaces of these different models are highly correlated (R^2 > 0.8), and the absolute concentrations agree within ±30%.

To ensure that we are not extrapolating in the covariate space, Figure 5A only shows the model predictions at locations whose land-use covariate values fall within the 1st and 99th percentile range of the data set used for model building. This yields concentration predictions at 6,026,961 residential census-block centroids in the contiguous United States with nonzero population, which covers 97.6% of census blocks and 97.4% of the total population according to the 2010 U.S. census. The remaining blocks are generally in extreme urban or rural locations. The white areas in Figure 5A fall outside the prediction criteria.

The models predict relatively uniform regional background levels for HOA and COA with hotspots corresponding to urban areas and along highway corridors (for HOA; see Figure S13). There is substantial spatial variability between rural and urban and within and between urban areas. The population-weighted national averages and interquartile range of predicted concentrations are 0.39 (0.09−0.44) μg m^{-3} for COA and 0.23 (0.02−0.26) μg m^{-3} for HOA. The predicted HOA and COA concentration surfaces are strongly correlated. R^2 between block-level national surfaces: 0.87; within-MSA R^2: mean 0.86, 5th−95th range 0.76−0.96 (n = 363 MSAs). A high correlation is expected since urban sources drive the concentrations of both.

Figure 5B,C shows the predictions for the Pittsburgh Metropolitan Statistical Area (MSA) to illustrate the spatial patterns for a representative urban region. Within this MSA, the predicted concentration hotspots are in the city center (highest in the downtown area) and densely populated areas. The transect through the center of Pittsburgh shown in Figure 5C illustrates that the predicted concentrations can vary widely within the city and gradually decrease as one moves away from the city center. Figure 5C also shows the contributions of the different model covariates to the predicted concentrations. There is a near-zero intercept, which represents the background concentrations. The impervious land surface and commercial land use likely describe the urban background levels. Finally, source-specific covariates (e.g., restaurant density within 100 m and restaurant density within 1000 m for the COA model) drive the large intra-urban variability in model predictions.

3.7. Comparison with CTM Simulations. The CTM-simulated 2016 annual average concentrations of primary organic aerosols from cooking and mobile sources are shown in Figures S15 and S16, respectively. We aggregated the LUR predictions to the 12 km CTM grid to make quantitative comparisons between the different models. The CTM predicts lower HOA and COA concentrations than the LUR. The predicted national population-weighted mean COA from the LUR and the CTM is 0.39 and 0.22 μg m^{-3}, respectively. For HOA, it is 0.23 and 0.09 μg m^{-3}, respectively. It should be noted that the CTM-predicted HOA includes both on-road and non-road mobile sources. If one only accounts the on-road mobile source, the CTM-predicted national population-weighted mean HOA is 0.05 μg m^{-3} (see Figure S17 for additional details).

Figure S18 directly compares the CTM predictions against the monitoring data. There is reasonable linearity (R^2: 0.56 for COA and 0.61 for HOA), but the measured concentrations are substantially higher than CTM predictions. The slope of the measured versus CTM is 1.5 for COA and 3.3 for HOA (Figure S18). The bias in the CTM predictions could be due to various reasons, including underrepresentation in the emission inventory, coarse spatial resolution of the CTM, etc. The spatial allocation below the county level of the emission inventory used for the CTM simulations is expected to be very uncertain and is difficult to constrain. Our analysis suggests that a resolution smaller than 12 km is needed to capture the spatial patterns of primary PM concentrations over urban areas (Figure 5C and Figure S13C).

4. IMPLICATION

In this paper, we demonstrate the feasibility of developing national empirical models for sparsely measured air pollutants. We show that the nationally representative data can be generated through a careful and targeted sampling design. The key constraint is that the monitoring data used to fit the model must span the range of variability across the national domain. We show that this can be efficiently achieved using a hybrid approach of combining high-spatial-resolution mobile sampling to capture intra-urban variability with fixed monitoring at rural/urban background locations to capture inter-urban and urban−rural patterns. While additional monitoring data would likely improve the model performance, the measurement locations need to be carefully selected to better span the entire range of national spatial variability. For example, our models are based on a limited number of rural observations; additional rural data are likely needed to predict spatial patterns in rural areas. Data from additional cities might help the model better predict intra- and inter-urban patterns.

A strength of this study was the focus on source-specific components of primary $PM_{2.5}$. This means that the covariates selected by the supervised linear regression are physically

https://doi.org/10.1021/acs.est.2c03398
Environ. Sci. Technol. 2022, 56, 14284−14295

interpretable. For example, traffic source-specific variables included different types of road length, intersections, and distance from roads. This likely helps with robustness and transferability.

We used 1 km and 200 m grid averaged mobile data to develop models, which average out some hyperlocal spatial variability. A higher spatial resolution (100 m or less) is needed to capture the near-source spatial variability, specifically for traffic. Our analysis was limited by the source-specific covariates that are available on a national scale. Improved source and activity-specific variables are likely needed to better predict the measured hyperlocal variability; for example, covariates such as traffic volume, composition of vehicle fleet, vehicle-specific power, cooking data by restaurant type, volume of food cooked, etc. However, development of these types of improved variables is especially challenging for national-scale models. Recently, land-use features extracted from Google Street View (GSV) imagery have been used for street-level air pollution modeling.[57,58] GSV features can better characterize the near-source microenvironment. These are nationally available and may improve the model performance. Finally, different model formulations (e.g., kriging, spatial semi-parametric models, random forest, etc.[6,52,57]) might also improve model's performance.

The results of this paper provide new insights into emerging urban $PM_{2.5}$ sources in the United States. The measured data, LUR predictions, and CTM predictions all indicate that the contribution of cooking organic $PM_{2.5}$ is higher than organic traffic primary $PM_{2.5}$. Traffic emissions have reduced dramatically over the past years due to technology and regulations.[59,60] Currently, cooking emissions are largely unregulated. This suggests that increased regulatory focus on cooking emissions may be a way to reduce exposure hotspots.

NOTES/DATA AVAILABILITY

The predicted census-block-level COA and HOA concentrations in the contiguous United States are publicly and freely available online at https://www.caces.us.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.2c03398.

The description of chemical transport model simulations, temporal correction factors for HOA and COA, analysis of diurnal and seasonal trends of HOA and COA concentrations using the Pittsburgh data set, CTM-simulated seasonal variations in traffic and cooking primary OA concentrations, evaluation of CMAQ simulation, analysis similar to Figure 1 using 200 m aggregated mobile data, analysis similar to panel C of Figure 1 for more covariates and values calculated at two different buffer sizes (500 and 1000 m), analysis similar to Figure 2 using 200 m aggregated mobile data, analysis similar to Figure 3 for models developed with fitting all 200 m grid averaged mobile data, analysis similar to Figure 4 showing incrementally including additional data from land-use-based subgroups, analysis similar to Figure 5 for HOA, assessment of transferability of LUR model-building data set through city holdouts, comparison of LUR-predicted national concentration surfaces from models developed with 1 km and 200 m grid averaged mobile monitoring data, comparison of LUR-predicted COA and HOA concentrations with CTM predictions, comparison of measured COA and HOA concentrations with CTM predictions, description of HR-AMS measured HOA and COA concentrations compiled from various atmospheric field campaigns in the United States, list of land-use covariates used for COA and HOA LUR model development, and list of predictor variables and performance parameters for COA and HOA LUR models (PDF)

AUTHOR INFORMATION

Corresponding Authors

Albert A. Presto − *Center for Atmospheric Particle Studies and Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;* orcid.org/0000-0002-9156-1094; Phone: 412-721-5203; Email: apresto@andrew.cmu.edu

Allen L. Robinson − *Center for Atmospheric Particle Studies and Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;* orcid.org/0000-0002-1819-083X; Phone: 412-268-3657; Email: alr@andrew.cmu.edu

Authors

Provat K. Saha − *Center for Atmospheric Particle Studies and Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;* orcid.org/0000-0002-4044-9350

Steve Hankey − *School of Public and International Affairs, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0002-7530-6077

Benjamin N. Murphy − *Center for Environmental Measurement and Modeling, U.S. Environmental Protection Agency, Durham, North Carolina 27709, United States;* orcid.org/0000-0003-3542-5378

Chris Allen − *General Dynamics Information Technology, Durham, North Carolina 27711, United States*

Wenwen Zhang − *Department of Public Informatics, Rutgers University, New Brunswick, New Jersey 08901, United States*

Julian D. Marshall − *Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington 98195, United States;* orcid.org/0000-0003-4087-1209

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.2c03398

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research is part of the Center for Air, Climate, and Energy Solutions (CACES), which was supported by the Environmental Protection Agency (assistance agreement number RD83587301). The U.S. Environmental Protection Agency (U.S. EPA) through its Office of Research and Development collaborated in the research described here. The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. EPA. Any mention of trade names, manufacturers, or products does not imply an endorsement by the U.S. Government or the U.S. EPA. The EPA and its employees do not endorse any commercial products, services, or enterprises. The authors

would like to thank Havala Pye, Karl Seltzer, Kristen Foley, and the EQUATES team for assistance with CTM simulations.

REFERENCES

(1) Chudnovsky, A. A.; Koutrakis, P.; Kloog, I.; Melly, S.; Nordio, F.; Lyapustin, A.; Wang, Y.; Schwartz, J. Fine Particulate Matter Predictions Using High Resolution Aerosol Optical Depth (AOD) Retrievals. *Atmos. Environ.* **2014**, *89*, 189−198.

(2) Ren, X.; Mi, Z.; Georgopoulos, P. G. Comparison of Machine Learning and Land Use Regression for Fine Scale Spatiotemporal Estimation of Ambient Air Pollution: Modeling Ozone Concentrations across the Contiguous United States. *Environ. Int.* **2020**, *142*, No. 105827.

(3) Kloog, I.; Chudnovsky, A. A.; Just, A. C.; Nordio, F.; Koutrakis, P.; Coull, B. A.; Lyapustin, A.; Wang, Y.; Schwartz, J. 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.* **2014**, *95*, 581−590.

(4) Brokamp, C.; Jandarov, R.; Rao, M. B.; LeMasters, G.; Ryan, P. Exposure Assessment Models for Elemental Components of Particulate Matter in an Urban Environment: A Comparison of Regression and Random Forest Approaches. *Atmos. Environ.* **2017**, *151*, 1−11.

(5) Wang, Y.; Bechle, M. J.; Kim, S.-Y.; Adams, P. J.; Pandis, S. N.; Pope, C. A.; Robinson, A. L.; Sheppard, L.; Szpiro, A. A.; Marshall, J. D. Spatial Decomposition Analysis of NO2 and PM2.5 Air Pollution in the United States. *Atmos. Environ.* **2020**, *241*, No. 117470.

(6) Kim, S.-Y.; Bechle, M.; Hankey, S.; Sheppard, L.; Szpiro, A. A.; Marshall, J. D. Concentrations of Criteria Pollutants in the Contiguous U.S., 1979 − 2015: Role of Prediction Model Parsimony in Integrated Empirical Geographic Regression. *PLoS One* **2020**, *15*, No. e0228535.

(7) Di, Q.; Rowland, S.; Koutrakis, P.; Schwartz, J. A Hybrid Model for Spatially and Temporally Resolved Ozone Exposures in the Continental United States. *J. Air Waste Manage. Assoc.* **2017**, *67*, 39−52.

(8) Chen, J.; de Hoogh, K.; Gulliver, J.; Hoffmann, B.; Hertel, O.; Ketzel, M.; Bauwelinck, M.; van Donkelaar, A.; Hvidtfeldt, U. A.; Katsouyanni, K.; Janssen, N. A. H.; Martin, R. V.; Samoli, E.; Schwartz, P. E.; Stafoggia, M.; Bellander, T.; Strak, M.; Wolf, K.; Vienneau, D.; Vermeulen, R.; Brunekreef, B.; Hoek, G. A Comparison of Linear Regression, Regularization, and Machine Learning Algorithms to Develop Europe-Wide Spatial Models of Fine Particles and Nitrogen Dioxide. *Environ. Int.* **2019**, *130*, No. 104934.

(9) Tong, D. Q.; Lamsal, L.; Pan, L.; Ding, C.; Kim, H.; Lee, P.; Chai, T.; Pickering, K. E.; Stajner, I. Long-Term NOx Trends over Large Cities in the United States during the Great Recession: Comparison of Satellite Retrievals, Ground Observations, and Emission Inventories. *Atmos. Environ.* **2015**, *107*, 70−84.

(10) 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. An Ensemble-Based Model of PM2.5 Concentration across the Contiguous United States with High Spatiotemporal Resolution. *Environ. Int.* **2019**, *130*, No. 104909.

(11) Kelly, J. T.; Jang, C.; Timin, B.; Di, Q.; Schwartz, J.; Liu, Y.; van Donkelaar, A.; Martin, R. V.; Berrocal, V.; Bell, M. L. Examining PM2.5 Concentrations and Exposure Using Multiple Models. *Environ. Res.* **2021**, *196*, No. 110432.

(12) Delle Monache, L.; Alessandrini, S.; Djalalova, I.; Wilczak, J.; Knievel, J. C.; Kumar, R. Improving Air Quality Predictions over the United States with an Analog Ensemble. *Weather Forecast.* **2020**, *35*, 2145−2162.

(13) Beckerman, B. S.; Jerrett, M.; Serre, M.; Martin, R. V.; Lee, S.-J.; van Donkelaar, A.; Ross, Z.; Su, J.; Burnett, R. T. A Hybrid Approach to Estimating National Scale Spatiotemporal Variability of PM2.5 in the Contiguous United States. *Environ. Sci. Technol.* **2013**, *47*, 7233−7241.

(14) Lee, S.-J.; Serre, M. L.; van Donkelaar, A.; Martin, R. V.; Burnett, R. T.; Jerrett, M. Comparison of Geostatistical Interpolation and Remote Sensing Techniques for Estimating Long-Term Exposure to Ambient PM2.5 Concentrations across the Continental United States. *Environ. Health Perspect.* **2012**, *120*, 1727−1732.

(15) Apte, J. S.; Marshall, J. D.; Cohen, A. J.; Brauer, M. Addressing Global Mortality from Ambient PM2.5. *Environ. Sci. Technol.* **2015**, *49*, 8057−8066.

(16) Vodonos, A.; Schwartz, J. Estimation of Excess Mortality Due to Long-Term Exposure to PM2.5 in Continental United States Using a High-Spatiotemporal Resolution Model. *Environ. Res.* **2021**, *196*, No. 110904.

(17) Turner, M. C.; Jerrett, M.; Pope, C. A.; Krewski, D.; Gapstur, S. M.; Diver, W. R.; Beckerman, B. S.; Marshall, J. D.; Su, J.; Crouse, D. L.; Burnett, R. T. Long-Term Ozone Exposure and Mortality in a Large Prospective Study. *Am. J. Respir. Crit. Care Med.* **2016**, *193*, 1134−1142.

(18) Clark, L. P.; Millet, D. B.; Marshall, J. D. National Patterns in Environmental Injustice and Inequality: Outdoor NO2 Air Pollution in the United States. *PLoS One* **2014**, *9*, No. e94431.

(19) Tessum, C. W.; Apte, J. S.; Goodkind, A. L.; Muller, N. Z.; Mullins, K. A.; Paolella, D. A.; Polasky, S.; Springer, N. P.; Thakrar, S. K.; Marshall, J. D.; Hill, J. D. Inequity in Consumption of Goods and Services Adds to Racial−Ethnic Disparities in Air Pollution Exposure. *Proc. Natl. Acad. Sci.* **2019**, *116*, 6001−6006.

(20) Liu, J.; Clark, L. P.; Bechle, M. J.; Hajat, A.; Kim, S.-Y.; Robinson, A. L.; Sheppard, L.; Szpiro, A. A.; Marshall, J. D. Disparities in Air Pollution Exposure in the United States by Race/Ethnicity and Income, 1990−2010. *Environ. Health Perspect.* **2021**, *129*, 127005.

(21) Adams, K.; Greenbaum, D. S.; Shaikh, R.; van Erp, A. M.; Russell, A. G. Particulate Matter Components, Sources, and Health: Systematic Approaches to Testing Effects. *J. Air Waste Manage. Assoc.* **2015**, *65*, 544−558.

(22) Laden, F.; Neas, L. M.; Dockery, D. W.; Schwartz, J. Association of Fine Particulate Matter from Different Sources with Daily Mortality in Six U.S. Cities. *Environ. Health Perspect.* **2000**, *108*, 941−947.

(23) Zanobetti, A.; Franklin, M.; Koutrakis, P.; Schwartz, J. Fine Particulate Air Pollution and Its Components in Association with Cause-Specific Emergency Admissions. *Environ. Health* **2009**, *8*, 58.

(24) Fischer, P. H.; Marra, M.; Ameling, C. B.; Velders, G. J. M.; Hoogerbrugge, R.; de Vries, W.; Wesseling, J.; Janssen, N. A. H.; Houthuijs, D. Particulate Air Pollution from Different Sources and Mortality in 7.5 Million Adults — The Dutch Environmental Longitudinal Study (DUELS). *Sci. Total Environ.* **2020**, *705*, No. 135778.

(25) Dai, L.; Bind, M.-A.; Koutrakis, P.; Coull, B. A.; Sparrow, D.; Vokonas, P. S.; Schwartz, J. D. Fine Particles, Genetic Pathways, and Markers of Inflammation and Endothelial Dysfunction: Analysis on Particulate Species and Sources. *J. Exposure Sci. Environ. Epidemiol.* **2016**, *26*, 415−421.

(26) Kioumourtzoglou, M.-A.; Coull, B. A.; Dominici, F.; Koutrakis, P.; Schwartz, J.; Suh, H. The Impact of Source Contribution Uncertainty on the Effects of Source-Specific PM2.5 on Hospital Admissions: A Case Study in Boston, MA. *J. Exposure Sci. Environ. Epidemiol.* **2014**, *24*, 365−371.

(27) Wyzga, R. E.; Rohr, A. C. Long-Term Particulate Matter Exposure: Attributing Health Effects to Individual PM Components. *J. Air Waste Manage. Assoc.* **2015**, *65*, 523−543.

(28) Yang, Y.; Pun, V. C.; Sun, S.; Lin, H.; Mason, T. G.; Qiu, H. Particulate Matter Components and Health: A Literature Review on Exposure Assessment. *J. Public Health Emerg.* **2018**, *2*, 14.

(29) Krall, J. R.; Chang, H. H.; Sarnat, S. E.; Peng, R. D.; Waller, L. A. Current Methods and Challenges for Epidemiological Studies of the Associations Between Chemical Constituents of Particulate Matter and Health. *Curr. Environ. Health Rep.* **2015**, *2*, 388−398.

(30) Chen, H.; Zhang, Z.; van Donkelaar, A.; Bai, L.; Martin, R. V.; Lavigne, E.; Kwong, J. C.; Burnett, R. T. Understanding the Joint Impacts of Fine Particulate Matter Concentration and Composition on the Incidence and Mortality of Cardiovascular Disease: A

https://doi.org/10.1021/acs.est.2c03398
Environ. Sci. Technol. 2022, 56, 14284−14295

Component-Adjusted Approach. *Environ. Sci. Technol.* **2020**, *54*, 4388−4399.
(31) van Donkelaar, A.; Martin, R. V.; Li, C.; Burnett, R. T. Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. *Environ. Sci. Technol.* **2019**, *53*, 2595−2611.
(32) Di, Q.; Koutrakis, P.; Schwartz, J. A Hybrid Prediction Model for PM2.5 Mass and Components Using a Chemical Transport Model and Land Use Regression. *Atmos. Environ.* **2016**, *131*, 390−399.
(33) Chen, J.; de Hoogh, K.; Gulliver, J.; Hoffmann, B.; Hertel, O.; Ketzel, M.; Weinmayr, G.; Bauwelinck, M.; van Donkelaar, A.; Hvidtfeldt, U. A.; Atkinson, R.; Janssen, N. A. H.; Martin, R. V.; Samoli, E.; Andersen, Z. J.; Oftedal, B. M.; Stafoggia, M.; Bellander, T.; Strak, M.; Wolf, K.; Vienneau, D.; Brunekreef, B.; Hoek, G. Development of Europe-Wide Models for Particle Elemental Composition Using Supervised Linear Regression and Random Forest. *Environ. Sci. Technol.* **2020**, *54*, 15698−15709.
(34) Hu, J.; Ostro, B.; Zhang, H.; Ying, Q.; Kleeman, M. J. Using Chemical Transport Model Predictions To Improve Exposure Assessment of PM2.5 Constituents. *Environ. Sci. Technol. Lett.* **2019**, *6*, 456−461.
(35) Meng, J.; Martin, R. V.; Li, C.; van Donkelaar, A.; Tzompa-Sosa, Z. A.; Yue, X.; Xu, J.-W.; Weagle, C. L.; Burnett, R. T. Source Contributions to Ambient Fine Particulate Matter for Canada. *Environ. Sci. Technol.* **2019**, *53*, 10269−10278.
(36) 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. Source Sector and Fuel Contributions to Ambient PM2.5 and Attributable Mortality across Multiple Spatial Scales. *Nat. Commun.* **2021**, *12*, 3594.
(37) Jimenez, J. L.; Jayne, J. T.; Shi, Q.; Kolb, C. E.; Worsnop, D. R.; Yourshaw, I.; Seinfeld, J. H.; Flagan, R. C.; Zhang, X.; Smith, K. A.; Morris, J. W.; Davidovits, P. Ambient Aerosol Sampling Using the Aerodyne Aerosol Mass Spectrometer. *J. Geophys. Res.: Atmos.* **2003**, *108*, 8425.
(38) DeCarlo, P. F.; Kimmel, J. R.; Trimborn, A.; Northway, M. J.; Jayne, J. T.; Aiken, A. C.; Gonin, M.; Fuhrer, K.; Horvath, T.; Docherty, K. S.; Worsnop, D. R.; Jimenez, J. L. Field-Deployable, High-Resolution, Time-of-Flight Aerosol Mass Spectrometer. *Anal. Chem.* **2006**, *78*, 8281−8289.
(39) Zhang, Q.; Jimenez, J. L.; Canagaratna, M. R.; Ulbrich, I. M.; Ng, N. L.; Worsnop, D. R.; Sun, Y. Understanding Atmospheric Organic Aerosols via Factor Analysis of Aerosol Mass Spectrometry: A Review. *Anal. Bioanal. Chem.* **2011**, *401*, 3045−3067.
(40) Ge, X.; Setyan, A.; Sun, Y.; Zhang, Q. Primary and Secondary Organic Aerosols in Fresno, California during Wintertime: Results from High Resolution Aerosol Mass Spectrometry. *J. Geophys. Res.: Atmos.* **2012**, *117*, D19301.
(41) Ng, N. L.; Canagaratna, M. R.; Zhang, Q.; Jimenez, J. L.; Tian, J.; Ulbrich, I. M.; Kroll, J. H.; Docherty, K. S.; Chhabra, P. S.; Bahreini, R.; Murphy, S. M.; Seinfeld, J. H.; Hildebrandt, L.; Donahue, N. M.; DeCarlo, P. F.; Lanz, V. A.; Prévôt, A. S. H.; Dinar, E.; Rudich, Y.; Worsnop, D. R. Organic Aerosol Components Observed in Northern Hemispheric Datasets from Aerosol Mass Spectrometry. *Atmos. Chem. Phys.* **2010**, *10*, 4625−4641.
(42) Robinson, E. S.; Gu, P.; Ye, Q.; Li, H. Z.; Shah, R. U.; Apte, J. S.; Robinson, A. L.; Presto, A. A. Restaurant Impacts on Outdoor Air Quality: Elevated Organic Aerosol Mass from Restaurant Cooking with Neighborhood-Scale Plume Extents. *Environ. Sci. Technol.* **2018**, *52*, 9285−9294.
(43) Shah, R. U.; Robinson, E. S.; Gu, P.; Robinson, A. L.; Apte, J. S.; Presto, A. A. High-Spatial-Resolution Mapping and Source Apportionment of Aerosol Composition in Oakland, California, Using Mobile Aerosol Mass Spectrometry. *Atmos. Chem. Phys.* **2018**, *18*, 16325−16344.
(44) Sun, Y.-L.; Zhang, Q.; Schwab, J. J.; Demerjian, K. L.; Chen, W.-N.; Bae, M.-S.; Hung, H.-M.; Hogrefe, O.; Frank, B.; Rattigan, O. V.; Lin, Y.-C. Characterization of the Sources and Processes of Organic and Inorganic Aerosols in New York City with a High-Resolution Time-of-Flight Aerosol Mass Apectrometer. *Atmos. Chem. Phys.* **2011**, *11*, 1581−1602.
(45) Hoek, G. Methods for Assessing Long-Term Exposures to Outdoor Air Pollutants. *Curr. Environ. Health Rep.* **2017**, *4*, 450−462.
(46) Hatzopoulou, M.; Valois, M. F.; Levy, I.; Mihele, C.; Lu, G.; Bagg, S.; Minet, L.; Brook, J. Robustness of Land-Use Regression Models Developed from Mobile Air Pollutant Measurements. *Environ. Sci. Technol.* **2017**, *51*, 3938−3947.
(47) Saha, P. K.; Li, H. Z.; Apte, J. S.; Robinson, A. L.; Presto, A. A. Urban Ultrafine Particle Exposure Assessment with Land-Use Regression: Influence of Sampling Strategy. *Environ. Sci. Technol.* **2019**, *53*, 7326−7336.
(48) Saha, P. K.; Hankey, S.; Marshall, J. D.; Robinson, A. L.; Presto, A. A. High-Spatial-Resolution Estimates of Ultrafine Particle Concentrations across the Continental United States. *Environ. Sci. Technol.* **2021**, *55*, 10320−10331.
(49) Gu, P.; Li, H. Z.; Ye, Q.; Robinson, E. S.; Apte, J. S.; Robinson, A. L.; Presto, A. A. Intracity Variability of Particulate Matter Exposure Is Driven by Carbonaceous Sources and Correlated with Land-Use Variables. *Environ. Sci. Technol.* **2018**, *52*, 11545−11554.
(50) Robinson, E. S.; Shah, R. U.; Messier, K.; Gu, P.; Li, H. Z.; Apte, J. S.; Robinson, A. L.; Presto, A. A. Land-Use Regression Modeling of Source-Resolved Fine Particulate Matter Components from Mobile Sampling. *Environ. Sci. Technol.* **2019**, *53*, 8925−8937.
(51) Apte, J. S.; Messier, K. P.; Gani, S.; Brauer, M.; Kirchstetter, T. W.; Lunden, M. M.; Marshall, J. D.; Portier, C. J.; Vermeulen, R. C. H.; Hamburg, S. P. High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data. *Environ. Sci. Technol.* **2017**, *51*, 6999−7008.
(52) Messier, K. P.; Chambliss, S. E.; Gani, S.; Alvarez, R.; Brauer, M.; Choi, J. J.; Hamburg, S. P.; Kerckhoffs, J.; LaFranchi, B.; Lunden, M. M.; Marshall, J. D.; Portier, C. J.; Roy, A.; Szpiro, A. A.; Vermeulen, R. C. H.; Apte, J. S. Mapping Air Pollution with Google Street View Cars: Efficient Approaches with Mobile Monitoring and Land Use Regression. *Environ. Sci. Technol.* **2018**, *52*, 12563−12572.
(53) Eeftens, M.; Beelen, R.; de Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dėdelė, A.; Dons, E.; de Nazelle, A.; Dimakopoulou, K.; Eriksen, K.; Falq, G.; Fischer, P.; Galassi, C.; Gražulevičienė, R.; Heinrich, J.; Hoffmann, B.; Jerrett, M.; Keidel, D.; Korek, M.; Lanki, T.; Lindley, S.; Madsen, C.; Mölter, A.; Nádor, G.; Nieuwenhuijsen, M.; Nonnemacher, M.; Pedeli, X.; Raaschou-Nielsen, O.; Patelarou, E.; Quass, U.; Ranzi, A.; Schindler, C.; Stempfelet, M.; Stephanou, E.; Sugiri, D.; Tsai, M.-Y.; Yli-Tuomi, T.; Varró, M. J.; Vienneau, D.; von Klot, S.; Wolf, K.; Brunekreef, B.; Hoek, G. Development of Land Use Regression Models for PM2.5, PM2.5 Absorbance, PM10 and PMcoarse in 20 European Study Areas; Results of the ESCAPE Project. *Environ. Sci. Technol.* **2012**, *46*, 11195−11205.
(54) Beelen, R.; Hoek, G.; Vienneau, D.; Eeftens, M.; Dimakopoulou, K.; Pedeli, X.; Tsai, M. Y.; Künzli, N.; Schikowski, T.; Marcon, A.; Eriksen, K. T.; Raaschou-Nielsen, O.; Stephanou, E.; Patelarou, E.; Lanki, T.; Yli-Tuomi, T.; Declercq, C.; Falq, G.; Stempfelet, M.; Birk, M.; Cyrys, J.; von Klot, S.; Nádor, G.; Varró, M. J.; Dėdelė, A.; Gražulevičienė, R.; Mölter, A.; Lindley, S.; Madsen, C.; Cesaroni, G.; Ranzi, A.; Badaloni, C.; Hoffmann, B.; Nonnemacher, M.; Krämer, U.; Kuhlbusch, T.; Cirach, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Bellander, T.; Korek, M.; Olsson, D.; Strömgren, M.; Dons, E.; Jerrett, M.; Fischer, P.; Wang, M.; Brunekreef, B.; de Hoogh, K. Development of NO2 and NOx Land Use Regression Models for Estimating Air Pollution Exposure in 36 Study Areas in Europe − The ESCAPE Project. *Atmos. Environ.* **2013**, *72*, 10−23.
(55) US EPA. *CMAQ*; Zenodo: 2021.
(56) US EPA. *EQUATES: EPA's Air QUAlity TimE Series Project*; United States Environmental Protection Agency: 2022.
(57) Lu, T.; Marshall, J. D.; Zhang, W.; Hystad, P.; Kim, S.-Y.; Bechle, M. J.; Demuzere, M.; Hankey, S. National Empirical Models

of Air Pollution Using Microscale Measures of the Urban Environment. *Environ. Sci. Technol.* **2021**, *55*, 15519−15530.
(58) Qi, M.; Hankey, S. Using Street View Imagery to Predict Street-Level Particulate Air Pollution. *Environ. Sci. Technol.* **2021**, *55*, 2695−2704.
(59) Zhao, Y.; Saleh, R.; Saliba, G.; Presto, A. A.; Gordon, T. D.; Drozd, G. T.; Goldstein, A. H.; Donahue, N. M.; Robinson, A. L. Reducing Secondary Organic Aerosol Formation from Gasoline Vehicle Exhaust. *Proc. Natl. Acad. Sci.* **2017**, *114*, 6984−6989.
(60) McDonald, B. C.; de Gouw, J. A.; Gilman, J. B.; Jathar, S. H.; Akherati, A.; Cappa, C. D.; Jimenez, J. L.; Lee-Taylor, J.; Hayes, P. L.; McKeen, S. A.; Cui, Y. Y.; Kim, S.-W.; Gentner, D. R.; Isaacman-VanWertz, G.; Goldstein, A. H.; Harley, R. A.; Frost, G. J.; Roberts, J. M.; Ryerson, T. B.; Trainer, M. Volatile Chemical Products Emerging as Largest Petrochemical Source of Urban Organic Emissions. *Science* **2018**, *359*, 760−764.

https://doi.org/10.1021/acs.est.2c03398
Environ. Sci. Technol. 2022, 56, 14284−14295