Evaluating the Meteorological Effects on the Urban Form–Air Quality Relationship Using Mobile Monitoring

Environmental Science & Technology

pubs.acs.org/est

Article

Evaluating the Meteorological Effects on the Urban Form−Air Quality Relationship Using Mobile Monitoring

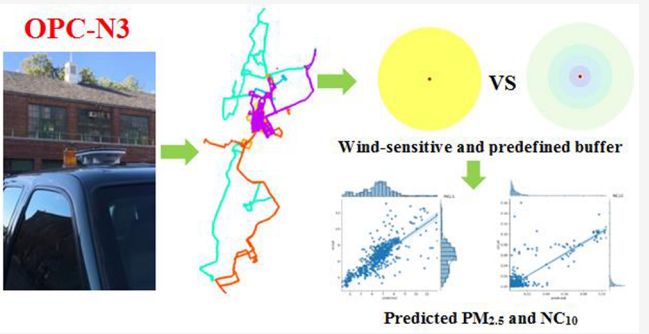
Ye Tian, Priyanka deSouza, Simone Mora,* Xiaobai Yao, Fabio Duarte, Leslie K Norford, Hui Lin, and Carlo Ratti

Cite This: https://doi.org/10.1021/acs.est.1c04854

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Predictive models based on mobile measurements have been increasingly used to understand the spatiotemporal variations of intraurban air quality. However, the effects of meteorological factors, which significantly affect the dispersion of air pollution, on the urban-form−air-quality relationship have not been understood on a granular level. We attempt to fill this gap by developing predictive models of particulate matter (PM) in the Bronx (New York City) using meteorological and urban form parameters. The granular PM data was collected by mobile low-cost sensors as the ground truth. To evaluate the effects of meteorological factors, we compared the performance of models using the urban form within fixed and wind-sensitive buffers, respectively. We find better predictive power in the wind-sensitive group (R = 0.85) for NC_{10} (number concentration for particles with diameters of 1 μm-10 μm) than the control group (R = 0.01), and modest improvements for $PM_{2.5}$ (R = 0.84 for the wind sensitive group, R = 0.77 for the control group), indicating that incorporating meteorological factors improved the predictive power of our models. We also found that urban form factors account for 62.95% of feature importance for NC_{10} and 14.90% for $PM_{2.5}$ (9.99% and 4.91% for 3-D and 2-D urban form factors, respectively) in our Random Forest models. It suggests the importance of incorporating urban form factors, especially for the uncommonly used 3-D characteristics, in estimating intraurban PM. Our method can be applied in other cities to better capture the influence of urban context on PM levels.



KEYWORDS: *meteorological factors, mobile monitoring, urban-form−air-quality relationship, 2-D and 3-D urban form, Random Forest*

1. INTRODUCTION

Air pollution is the single most significant environmental health risk worldwide and is associated with increased mortality and a range of serious diseases.[1−3] Air quality data is critical for the development of effective pollution management plans. However, regulatory air pollution monitors have high operating and capital costs. Even in the United States, only 60% of the U.S. census urban areas have a regulatory monitor.[4] The sparse nature of the stationary regulatory monitoring network makes it challenging to quantify intraurban air pollution that has a high level of spatiotemporal heterogeneity.[4−6]

With the advent of *IoT* (Internet of Things) technology, mobile monitoring has proved to be a suitable approach to characterize intraurban air pollution concentrations at a finer scale.[7] Dedicated mobile platforms have been used to collect air quality data, using mobile vans,[8] Google street-view cars,[4] and *UAV* (Unmanned Aerial Vehicles).[9] More recently, routine fleets of vehicles, such as trash trucks,[10] public transport vehicles,[11] and bicycles[12] have been used to measure air pollutants in cities on a real-time basis.

In conjunction, Land Use Regression (*LUR*) models have become increasingly more sophisticated.[13] Recently, LUR models relying on machine learning techniques have been developed to better capture the spatiotemporal distribution of air pollutants using mobile measurements.[6] Also, incorporating urban form factors has been shown to improve the performance of these models.

The urban form of a city captures the spatial configurations and arrangements of different urban elements, such as land use, population distribution, transportation networks, and urban infrastructure,[14−16] which can impact the spatiotemporal dynamics of air pollution. For example, industrial and traffic-heavy areas tend to have a high level of PM concentration.[17] Proximity to waterfronts or vegetation is associated with greater deposition velocity and can impact PM concentra-

Special Issue: Urban Air Pollution and Human Health

Received: July 23, 2021
Revised: January 10, 2022
Accepted: January 11, 2022

tions.[18] Open spaces contribute to better air movement and hence facilitate pollution dispersion, leading to lower pollution levels.[19] Urban form metrics used in previous predictive models, such as area, land use mix, population density, aggregation index, and fractal dimension, tend to be 2-D.[20−22] Recently, 3-D urban form characteristics, such as *FAI* (Frontal Area Index) and *SVF* (Sky View Factor), have been shown to improve the predictive performance of models of PM concentrations in fields such as urban ventilation,[23] urban climate modeling,[24] and urban relief visualization.[25] For example, Ghassoun et al. (2017)[26] improved the performance of predicting the total PM concentration by adding FAI into the 3-D LUR model. Shi et al. (2016)[27] found FAI, coupled with road network density and traffic volume, were the most influential factors in predicting street-level PM in Hong Kong. Tang et al. (2013)[28] used building heights and geometry to enhance the estimation of land use related variables and the pollution dispersion fields for long-term air pollutants. Edussuriya et al. (2011)[29] found that air pollution concentrations were impacted by the aspect ratio, building volume, and standard deviation of building height.

Meteorological factors also significantly influence ambient particulate matter levels by impacting ventilation rates, dry deposition, chemical reactions, natural emissions, and background aggregation.[30] Wind speed and direction, in particular, are key meteorological variables that have significant impacts on horizontal transport and distribution of pollution concentrations, as well as vertical mixing and dispersion in a region.[31] Few studies, however, have evaluated the effects of wind on the urban-form−air-quality relationship. Arain et al. (2007)[32] were among the first to show that including wind fields in LUR models could improve air pollution prediction. Vienneau et al. (2009)[33] used eight directional grids and wind rose diagrams to capture the effects of speed and direction on the wind-air pollution relationship. Naughton et al. (2018)[34] also built on the traditional LUR models by adding interactions between wind speed and direction to improve model performance. Contreras et al. (2016)[35] proposed using a wind-sensitive local *IDW* (Inverse Distance Weighted) model to incorporate the relationship between wind and urban form to analyze the spatiotemporal distribution of $PM_{2.5}$.

Previous studies that captured the impact of meteorology on the urban-form−air-quality relationship tended to (1) rely on linear models that likely do not capture the complex relationships between urban form and PM, (2) rely on measurements from fixed monitors that do not capture hyperlocal variations in urban PM concentrations, and (3) only consider $PM_{2.5}$ as the main pollutant of interest.

Our study attempts to quantify the importance of meteorological factors on the urban-form−air-quality relationships using a mobile monitoring data set with the following special design: (1) We incorporated both 2-D and 3-D urban form characteristics as predictors in a Random Forest (*RF*) model to better capture the nonlinearity between urban form factors and PM concentrations; (2) We used wind-sensitive buffer distances to capture the effect of meteorological factors on the urban-form−air-quality relationship; (3) We used PM concentration measurements collected from mobile low-cost sensors; (4) In addition to developing models to predict $PM_{2.5}$ concentrations, our study also predicted that of large particles, NC_{10} (number concentrations for coarser particles with diameters of 1 μm-10 μm). We did this for two reasons: (1) NC_{10} is important because, as larger particles tend to travel shorter distances than finer particles under similar wind conditions,[36] a high concentration of NC_{10} indicates the presence of significant local sources of pollution. Identifying locations of the high-impact local sources can help policymakers take necessary actions; and (2) NC_{10} can be measured accurately using low-cost sensors under conditions of low humidity. This will be discussed further in Section 2.2.

2. MATERIALS AND METHODS

2.1. Research Area. The Bronx, New York City (NYC) (Figure 1) was chosen to be the site of deployment as it has

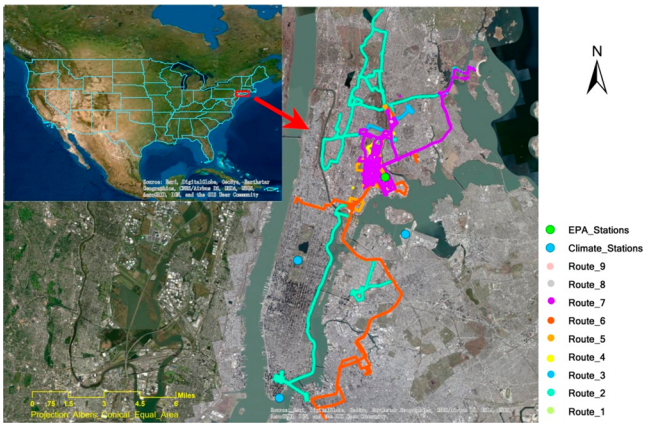


Figure 1. We deployed nine (five for pilot 1 and four for pilot 2) low-cost mobile monitors on vehicles in the Bronx, with a small number of measurements made in Manhattan, Queens, and Brooklyn from 01/20/2020 to 02/17/2020 (pilot 1) and 10/15/2020 to 02/08/2021 (pilot 2). The satellite background image comes from ArcGIS in ESRI.

the highest asthma hospitalization rate, the highest population percentage of minorities, the lowest mean household income, and the lowest average educational attainment level of any borough in New York City. The Bronx is also home to a dense network of highways and truck routes and other noxious land uses (e.g., Toxic Release Inventory facilities and other stationary pollution sources), which are important sources of pollution.[37−39]

2.2. Materials. The mobile monitoring platforms used in this study were developed as part of the City Scanner project (http://senseable.mit.edu/cityscanner) at Senseable City Lab, Massachusetts Institute of Technology, and were installed on municipal vehicles operated by the New York City Department of Parks and Recreation and the Department of Sanitation. Each sensing node included an Alphasense OPC-N3 (Optical Particle Counter) that reports fine particulate matter ($PM_{2.5}$) concentrations.[40] More details about Alphasense OPC-N3 can be found in Supporting Information (SI) Section S1: Alphasense OPC-N3. The nodes recorded measurements every two seconds. A total of 238,603 (129,348 in pilot 1 and 109,255 in pilot 2) measurements over 144 days (28 in pilot 1 and 116 in pilot 2) were made.

In addition to reporting $PM_{2.5}$ measurements, the OPC-N3 also reports concentrations of ambient particles in different size bins. The measured particle counts agree well with reference instrument measurements for coarser particles (>0.78 μm) under conditions of low relative humidity (RH< 80%)[41,42] (SI Figure S2: Dependence on the Relationship Between $PM_{2.5}$ from the OPC-N3 and Reference Monitor on Humidity). We

https://doi.org/10.1021/acs.est.1c04854
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

thus used the raw concentration metrics of coarser particles for measurements made in both pilots when the RH < 80% as a key outcome of interest: NC_{10} in this study. Other data sources are summarized in Table 1. More details can be found in SI Section S2: Covariates Description.

Table 1. Summary of Independent Variables[a]

	data type	units/ resolution	sources
meteorological factors	temperature	Fahrenheit	*IEM* stations in New York State (i.e., *HPN*, *JFK*, *JRB*, and *LGA*)
	feel temperature		
	wind speed	mile per hour	
	wind direction	degree	
	relative humidity	percentage	
	precipitation	millimeter	
Land use	tree canopy		NYC Open Data
	grass/shrub		
	bare soil		
	water		
	roads and railroads		
	other impervious surfaces		
	buildings with height information	meter	Microsoft Building Footprint Data
	landfill facilities		*HIFLD*
elevation	*DEM* (Digital Elevation Model)	30 m	*USGS*
traffic	*AADT*	daily	Department of Transportation in NYC
	bus stops		NYC Open data
	track routes		
socioeconomic factors	population	count	U.S. Census Bureau
	household income	dollar	

[a]*IEM*: Iowa Environmental Mesonet, *OSM*: Open Street Map, *HIFLD*: Homeland Infrastructure Foundation-Level Data, *USGS*: United States Geological Survey, AADT: Annual Average Daily Traffic.

2.3. Methods. The research design is summarized as a flowchart illustrated in SI Figure S5: Research Flowchart. Briefly, to predict $PM_{2.5}$ and NC_{10} data, we developed two models. In Model 1, we accounted for the impact of meteorology on the urban-form−PM relationship by using a wind-sensitive buffer radius for calculating the 2-D and 3-D urban form factors. More details about urban form are explained in Section 2.3.2. In Model 2, all urban form factors were calculated within fixed buffer radii (i.e., from 100 to 5000 m) of each measurement. All other variables depicted in Table 1 were included as covariates in both models.

2.3.1. Air Quality Data Preprocessing. To calibrate the error-prone $PM_{2.5}$ concentrations from the mobile OPC-N3, we developed a calibration function from our colocation experiment. We colocated an additional OPC-N3 from the same manufacturing batch with a reference monitor in the same general sampling area in the Bronx during pilot 2 (SI Figure S1: Mobile Monitoring Platform and EPA Fixed Station in the Bronx; SI Figure S3: Time-Series Analysis of $PM_{2.5}$ Measurements from the Mobile Sensor and the EPA Reference Monitor). We found that the RF model with time-varying covariates was effective (R^2 > 90%) at correcting the OPC-N3 measurements. For more information on the calibration procedure, see SI Section S3: Data Calibration. We applied this calibration function to the raw $PM_{2.5}$ measurements made by the mobile sensors in pilot 2. As no colocation experiment was performed in pilot 1, we only use $PM_{2.5}$ measurements from pilot 2. Henceforth, $PM_{2.5}$ refers to the calibrated $PM_{2.5}$ measurements.

To standardize measurements made at the same location in NYC on different days and at different times, we needed to correct them for background variation in $PM_{2.5}$ and NC_{10}, during a day and over different days that could obscure changes in local PM concentrations. To do this, we assumed that background levels vary temporally but not spatially over the region of interest. Essentially, we assumed that the background level of each pollutant of interest is due to transported aerosol from sources outside the study region. We used a time-series, spline-of-minimums approach to estimate the background concentrations for $PM_{2.5}$ and NC_{10}.[43]

After estimating the background $PM_{2.5}$ and NC_{10} values, we performed a background time-of-day correction. For more details on this approach as well as our assessment of the robustness of the background correction, refer to SI Section S4: Background Correction. Henceforth, NC_{10} and $PM_{2.5}$ in this article refer to the background-corrected concentrations.

2.3.2. Urban Form Metrics. Estimation of Wind-Sensitive Urban Form Covariates. To explore the effects of wind and urban form on air pollution concentrations, we devised a wind-sensitive buffer around each air pollution measurement, where the radius was determined by the wind speed and temporal granularity of the wind (SI Figure S6: Two Scenarios to Calculate Urban form). In this study, the finest temporal granularity for wind factor is 5 min according to the IEM station. The radii we chose describe how far air masses have moved during these 5 min because of wind. This radius represents the maximum potential influencing area by the wind for each measurement. We then estimated model covariates of all urban form metrics (described below) within this wind-sensitive buffer. Such a variable buffer allows us to factor in how different covariates impact local air pollution differently at different wind speeds.

To test if the inclusion of covariates calculated in a wind-sensitive buffer region as opposed to a fixed buffer area improved the model, we compared the predictive accuracy of models with wind-sensitive buffers with models that only included urban form metrics estimated in fixed buffer radii (i.e., 50 m, 100 m, 300 m, 500 m, 1000 m, 2000 m, 3000 m, and 5000 m) used in previous studies, around each data point.[34,44,45] Note all other covariates that were not related to urban form (Table 1) were constructed using fixed buffers in both models.

2-D Urban Form Metrics. We calculated the 2-D urban form in the study area through the simultaneous utilization of land use and landscape metrics. Briefly speaking, we calculated six such metrics, including total area (CA), edge density (ED), patch density (PD), largest patch index (LPI), landscape shape index (LSI), and aggregation index (AI), for each land use type

https://doi.org/10.1021/acs.est.1c04854
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

in buffer zones constructed around each measurement. These metrics capture the spatial coverage, fragmentation, patch dominance, and shape complexity of land use patterns[46−48] that likely shape air pollution patterns.[49−53] SI Table S1 (Summary of Selected Urban Form Metrics)[54] describes each metric in detail. The interaction between the land use and landscape metrics also likely impacts PM concentrations. Therefore, in addition to using each metric as an independent covariate, we also added interactions of them as predictors. This gives us 48 urban form metrics in total (eight land uses times six landscape metrics). Details can be seen in SI Table S2 (Summary Statistics for 2-D Urban Form Metrics for the Wind-Sensitive Group). We will now describe the unconventional 3-D urban form metrics proposed in this study:

3-D Urban Form. We used Frontal Area Index (*FAI*), Sky View Factor (*SVF*), and height variation (*H_Var*) to capture 3-D urban form factors.

FAI (Frontal Area Index). FAI is a wind direction-dependent parameter for estimating aerodynamic resistance of the urban surface as a prediction of wind ventilation that has been widely used to evaluate horizontal permeability of the wind from a specific direction within an urban plot.[27] In this study, we further improved the estimation of FAI in the following manner: (1). We used real-time direction instead of a fixed[27] or dominant[26] wind direction during a certain period; (2) We not only considered the first windward area but the remaining windward area. These buildings in the remaining windward area need to be counted when they are higher or wider than the first windward buildings. More details can be found in SI Section S5: *FAI*.

We first resampled the measurements every 5 min in accordance with the finest temporal granularity of the IEM station and calculated the FAI values for 2166 resampled data points based on the real-time wind direction. The FAI surface was then smoothed by Kriging (SI Figure S8: *FAI* Value in the Bronx Area). The smooth results cannot cover the entire Bronx due to the uneven distribution of collected data points. However, it does not affect the model construction because the prediction is point-based. According to the results, most FAI values in the study range between 0.01 and 0.09, suggesting that the buildings in the study area did not seriously block the wind flow. The FAI is lower in the northern Bronx and higher in the southern Bronx, where wider and taller buildings are seen and most commercial and residential land uses are located, which accords with previous findings.[55] SI Table S3 (Sum FAI Values for Different Land Uses) shows the sum value of FAI for different land uses. As can be seen, the artificial land uses (e.g., buildings, roads, and other impervious surfaces) have overall higher FAI values than urban vegetation (grass/shrub), suggesting that wind is more blocked in building areas where air pollutants are prone to be retarded leading to poor air ventilation, and vice versa for the water and bare soil areas.

SVF (Sky View Factor). SVF is a measure of how much sky is visible at a given location. We include it as a 3D form metric of the built environment, as it quantifies the openness to the sky of a given location. More details can be found in SI Section S6: *SVF*. Various data sets have been used to estimate SVF, such as *DEM* (Digital Elevation Model),[56] fisheye photographs,[24] and terrestrial LiDAR.[56] In this study, due to data availability, we used the DEM as the background elevation and building height to calculate SVF. SVF ranges between 0 and 1 where close to 1 indicates that almost the entire hemisphere is visible (e.g., planes and peaks) and close to 0 means almost no sky is visible (e.g., deep sinks and valleys).[25]

We calculated SVF through 136 406 buildings in the Bronx and smoothed the results by the Kriging method (SI Figure S9: *SVF* Value in the Bronx Area). As can be seen, most areas have relatively low SVF values (close to 0), indicating that most regions have an almost completely obstructed sky. This is because many regions in this zone are highly urbanized areas with densely distributed tall buildings and relatively low sky visibility for each observation point. Such dense streets could trap particulates within their boundaries, leading to high PM levels within the immediate areas around the roadway.[57] Higher SVF values spread over parks, rivers, and open spaces in the research area in which air pollutants are more easily transported and diluted.

H_Var. Building height variation describes the smoothness of the building, which impacts the distribution of wind flow. For example, high and dense constructions can trap pollutants between them and impede their dilution.[58] Likewise, the buildings are calculated for both wind-sensitive and fixed buffer distances. More details can be found in SI Section S7: *H_Var*.

2.3.3. Covariate Selection. Considering possible collinearity between the covariates, we used a Greedy Stepwise algorithm to filter correlated covariates and select the subset of the most significant covariates as candidates in the RF model. The preferred subsets of features are highly correlated with the dependent variable while having low intercorrelation.[59]

2.3.4. Prediction Model. Machine learning models have been widely used to model PM concentrations as they can account for nonlinear relationships, deal with multidimensional independent variables, and allow for complex interactions between the various predictors.[60] Previous researchers have found that RF models developed to predict air pollution have been at least as accurate and, in many cases, more accurate than linear models.[61] Robust cross-validation can prevent overfitting,[62] which is crucial given the nonlinear characteristics of air pollution.[63] Compared to other nonparametric techniques, such as neural networks, RF also provides metrics that capture the relative importance of the different independent covariates in predicting the outcome. Feature importance is computed as the decrease in node impurity weighted by the probability of reaching that node, which is also called *MDI* (Mean Decrease in Impurity). It counts the times a feature is used to split a node weighted by the number of samples it splits.[64] This study used the various covariates in an RF model to predict $PM_{2.5}$ and NC_{10} concentrations. To avoid overfitting, we used a 10-fold cross-validation technique to tune the hyper-parameters of our model. To illustrate, the original sample is randomly partitioned into 10 equal-size subsamples. A single subsample is retained as the validation data for testing the model, and the remaining nine subsamples are used as training data. We built 500 trees in RF in the software Weka for $PM_{2.5}$ and NC_{10}, respectively, and calculated the prediction interval by considering the range of prediction values returned by the individual trees in the forest, which indicates the range in which the true value is expected to fall. SI Figure S10 describes RF construction. We also compared the performance of using an RF model with other commonly used models (i.e., Artificial Neural Network, Multiple Linear Regression) to ensure that it is the best model for our data. In addition, we predicted NC_{10} and $PM_{2.5}$ based on the grid centroids with a resolution of 600 m (same as the urban plot for which we calculated FAI) in the Bronx. To qualitatively test

https://doi.org/10.1021/acs.est.1c04854
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

the robustness of the trends of $PM_{2.5}$ observed from our model, we compared the spatial variations of $PM_{2.5}$ produced by the *NYCCAS* (The New York City Community Air Survey) results, a project run by the NYC Department of Health and Mental Hygiene. A direct comparison between the two results is not possible, as the NYCCAS data are three-month average concentrations made in 2018, whereas our data was generated from 2020 to 2021 over a much shorter period.

3. RESULTS AND DISCUSSIONS

3.1. Model Performance. A key aim of this study is to evaluate whether using a wind-sensitive buffer improves the predictive power of our model. Figure 2 shows the results from

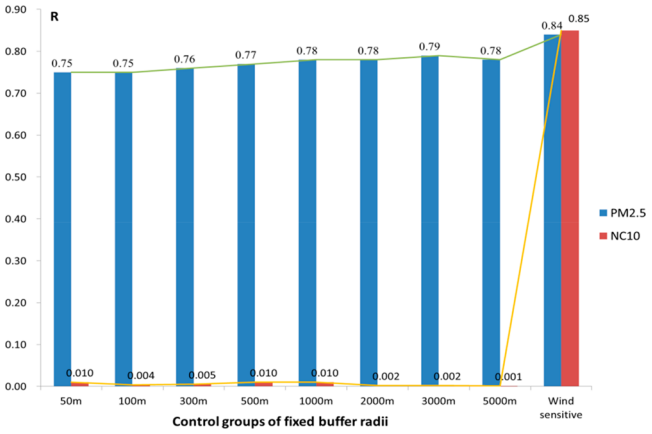
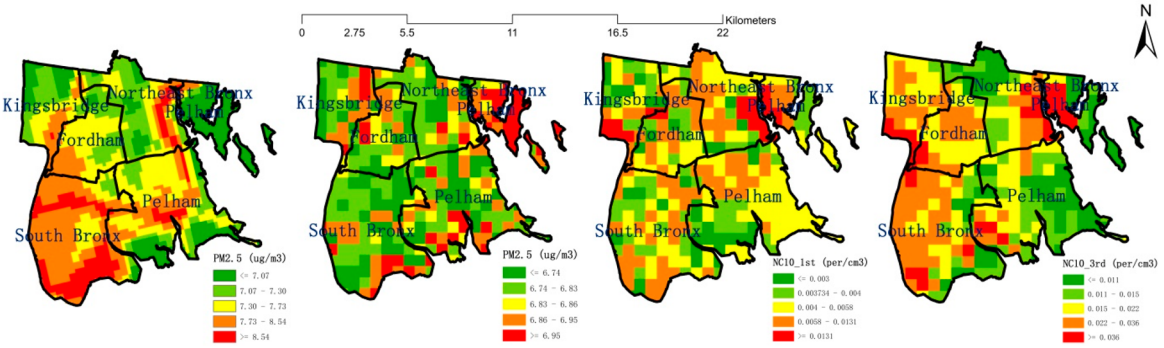


Figure 2. Model performances for the control groups (fixed buffer radii) and the wind-sensitive group.

10-fold cross-validation for Model 1 (predefined distances) and Model 2 (wind-sensitive). The wind-sensitive model outperforms the models using fixed buffer radii (from 50m to 5000 m), which represent the control groups, especially for NC_{10} (R increases from 0.01 to 0.85). Predictions of $PM_{2.5}$ are not as sensitive to buffer distance with modest improvement from Model 1 to Model 2 (R increases from 0.77 to 0.84). NC_{10} is likely more sensitive to the wind effect because it comprises coarser particles from nearby sources that tend to be transported predominantly by the wind. As a result, incorporating meteorological factors into the urban-form−air-quality model better captures the spatiotemporal heterogeneity of the concentration of larger particles. SI Table S4 (Model Performance for Different Buffer Distances) displays detailed model performances. Overall, the high predictive power of the wind-sensitive model for both NC_{10} (R = 0.85) and $PM_{2.5}$ (R = 0.84) is encouraging.

3.2. Model Comparison. We compared the performance of the RF model using the wind-sensitive covariates with other techniques, *ANN* (Artificial Neural Network) and *MLR* (Multiple Linear Regression), for NC_{10} and $PM_{2.5}$ within Model 2. It was found that the RF model outperforms the other models with higher R and lower RMSE (SI Table S5: Performance Comparison Between *RF*, *ANN*, and *MLR*), which is in line with previous findings.[61]

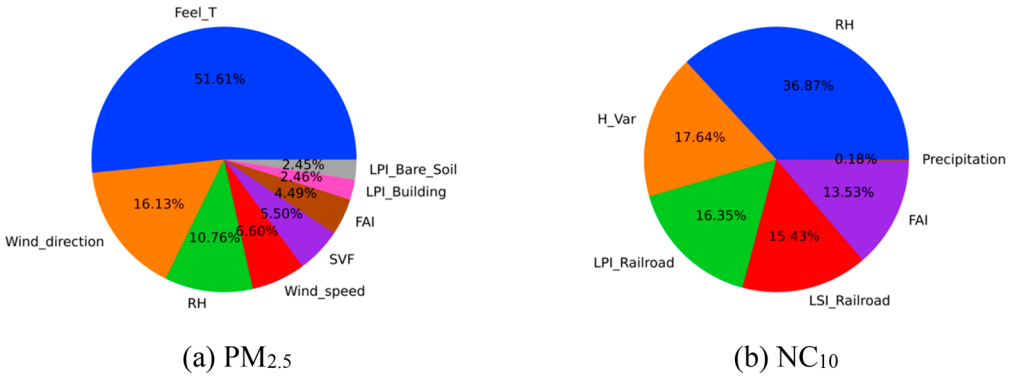
3.3. Prediction Results. We explored the performance of the prediction results of Model 2 using the RF. It appears that the distribution of NC_{10} concentrations is narrower than $PM_{2.5}$ concentrations, see SI Figure S11: Prediction Results for Different Air Pollutants. Figure 3 displays the predicted $PM_{2.5}$ and NC_{10} in the Bronx using the wind-sensitive model. As NC_{10} is sensitive to wind factors, we predict NC_{10} using the first (5.75 mph, 50 degree) and third quartiles (11.50 mph, 300 degree) during the research period (144 days) to represent the median of the lower and upper half of the wind information, respectively. $PM_{2.5}$ is not as sensitive to wind factors, so we used the average wind speed (8.02 mph, 174 degree) for prediction. We compared the predicted $PM_{2.5}$ concentrations with three-month winter averages derived from the NYCCAS program, which was conducted by the Department of Health and Mental Hygiene and Queens College to evaluate air quality distribution across NYC. Although the NYCCAS data is for a different year and is averaged over three months in 2018, it still provides us with ground truth to compare with the general trends of our predicted results. We found that there are several common hotspots (relatively higher level of $PM_{2.5}$ concentration) and cold spots in Figure 3a,b. For example, both sets of results show higher $PM_{2.5}$ concentrations in the border between northern Pelham and Northeast Bronx, the southern part of South Bronx, and southern Pelham. Likewise, some cold spots are found in northwestern Kingsbridge, which qualitatively demonstrates the credibility of Model 2 (wind-sensitive model). For NC_{10}, the spatial trends between Figure 3c,d are more similar. For example, the heavily polluted areas located in the border between northern Pelham and Northeast Bronx, western Kingsbridge, and the southern part of South Bronx. In the western Pelham, the level of NC_{10} concentration is relatively lower for both the first and third quartile.



(a).$PM_{2.5}$ (b).$PM_{2.5}$ (*NYCCAS*) (c).NC_{10} (1st quartile) (d).NC_{10} (3rd quartile)

Figure 3. Comparison between predicted $PM_{2.5}$ (ug/m^3) and NYCCAS in 2018 and NC_{10} (per/cm^3) results based on 1st and 3rd wind quartile.

https://doi.org/10.1021/acs.est.1c04854
Environ. Sci. Technol. XXXX, XXX, XXX−XXX



(a) $PM_{2.5}$ (b) NC_{10}

Figure 4. Feature importance (%) for different air pollutants within the wind-sensitive group.

We also find that Model 2 underestimates NC_{10} concentrations while the reverse is true for $PM_{2.5}$ (SI Figure S12: Prediction Interval for Different Air Pollutants). This suggests that the current model is more stable in predicting air pollutants with relatively lower concentration levels. It accords with SI Figure S11, which shows that values with high concentration levels diverge more from the fitted estimate.

3.4. Feature Importance. Figure 4 shows the distribution of feature importance for Model 2. Higher feature importance indicates the higher frequency of being chosen in the model and more influence on the air pollution concentration. After feature filtration, no selected features have shown obvious collinearity (SI Figure S13: Scatter Plot Matrix for the Selected Features).

For NC_{10}, the urban form factors account for 62.95% of feature importance, 31.17% for 3-D urban form (17.64% for H_Var and 13.53% for FAI), and 31.78% for 2-D urban form (16.35% for LPI-Railroad and 15.43% for LSI-Railroad), which is much higher than the remaining meteorological factors (36.31%). For $PM_{2.5}$, the feature importance of 3-D urban form factors (9.99%: 5.50% for SVF, 4.49% for FAI) almost doubles the 2-D urban form counterparts (4.91%: 2.46% for LPI-Building, 2.45% for LPI-Bare_Soil). According to SI Figure S14 (Feature Importance of NC_{10} for Groups with Different Buffer Radius), we also find that 3-D urban form factors, such as FAI, SVF, and H_Var, take essential proportions of feature importance for most groups using fixed buffer distances when estimating NC_{10}. It further proves that incorporating 3-D urban form is essential for estimating intraurban air pollutants, which accords with previous findings that urban morphology is more decisive to the street-level air quality in high-density cities.[27] For the 2-D urban form in Figure 4, the shape complexity (LSI) and path dominance (LPI) of the railroad, buildings, and bare soil are the final selected features. To illustrate, on the one hand, the aggregated and dominated distribution of railroad and buildings indicate intensive emission hotspots. On the other hand, air pollutants in areas with bare soil and smaller FAI value (SI Table S3), are easier to be transported and diluted. SI Table S6 (Feature Importance for Different Air Pollutants) shows the feature importance for all 10 folds. As for meteorological factors, RH is the only selected feature for all air pollutants and is also highly correlated with both $PM_{2.5}$ and NC_{10} (SI Figure S15: Correlation Matrix of Selected Features and Air Pollutants).

3.5. Meteorological Effects. To further explore the meteorological effects on the urban-form−air-quality relationship, we find that FAI, the wind-dependent variable, is the only factor that has been selected in all NC_{10} models and it also accounts for a large proportion of feature importance in most groups using different buffer distances. Besides, the feature importance of the same urban form factors also varies in different groups (SI Figure S14). We also find that using the fixed radius may be unreliable to make accurate predictions compared to the wind-sensitive group (SI Figure S16: Predicted NC_{10} for First and Third Quantile Meteorological Factors for Different Groups). These findings reconfirm the significance of meteorological effects on the urban-form−air-quality relationship.

4. IMPLICATIONS

Our results revealed that (1) including the effects of meteorology on the relationship between urban form and PM concentrations improved our prediction accuracy of NC_{10} and to some extent $PM_{2.5}$; (2) Using the RF model that allowed for nonlinear relationships between the various covariates and PM concentrations led to greater predictive capacity, and it outperformed ANN and MLR; (3) 2-D and 3-D urban form factors had considerable feature importance in our RF model. Thus, using mobile measurements from low-cost sensors has the potential to develop fine-scale granular maps of PM concentrations in cities around the world. NC_{10} is an important parameter for officials to understand as it can aid in the development of plans to control sources within their local jurisdiction. However, our study has several limitations:

1 More work needs to be done to validate the performance of low-cost sensors in a mobile setting and the results should be compared with advanced air dispersion models to provide a more accurate assessment of pollution dispersion.[65,66]
2 In our construction of urban form factors, we regarded buildings as regular cuboid shapes. However, real-world building shapes could be more complicated than the assumed shapes. Therefore, more data sets, such as LiDAR,[67] high-resolution DEM,[56] and the existing building database,[26] can be used to calculate 3-D urban form. It is necessary to construct the 3-D model on finer scales and analyze how it will affect pollution dilution and transportation.
3 We used a circular buffer to calculate urban form. However, it is necessary to explore other buffer shapes, such as plume buffers[68] or semicircular regions.[61]
4 The spatiotemporal coverage of the air quality data is limited in this study (01/20/2020 to 02/17/2020 for pilot 1 and 10/15/2020 to 02/08/2021 for pilot 2), so

https://doi.org/10.1021/acs.est.1c04854
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

we cannot explore the effects of seasonality and heterogeneity of air pollution on a broad scale. Future pilots may deploy mobile sensors within multiple cities combined with satellite-derived *AOD* (Aerosol Optical Depth) to further explore the spatiotemporal heterogeneity of urban air pollutants.

5 Although we analyzed the feature importance for different urban form factors, the marginal effects and the monotonicity of these factors on the prediction results remain unclear. Therefore, it will be helpful to add partial dependence plots[69] and SHAP value[70] distribution of urban form factors to further explore them in future studies.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.1c04854.

(PDF)

■ AUTHOR INFORMATION

Corresponding Author

Simone Mora − *Senseable City Laboratory, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States*; Phone: 1-617-902-8703; Email: moras@mit.edu

Authors

Ye Tian − *School of Geography and Environment, Jiangxi Normal University, Nanchang 330022, China; Senseable City Laboratory, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States; Department of Geography, University of Georgia, Athens, Georgia 30602, United States*; orcid.org/0000-0001-9532-6086

Priyanka deSouza − *Department of Urban Studies and Planning, University of Colorado Denver, Denver, Colorado 80202, United States*

Xiaobai Yao − *Department of Geography, University of Georgia, Athens, Georgia 30602, United States*

Fabio Duarte − *Senseable City Laboratory, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States; Pontifícia Universidade Católica do Paraná, Curitiba 80215, Brazil*

Leslie K Norford − *Department of Architecture, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States*

Hui Lin − *School of Geography and Environment, Jiangxi Normal University, Nanchang 330022, China*

Carlo Ratti − *Senseable City Laboratory, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.1c04854

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank all other members of the MIT Senseable City Lab Consortium (Ford Motor Company, RATP, Dover Corporation, Teck Resources, Lab Campus, Anas S.p.A., ENEL Foundation, EKTU, Università di Pisa, KTH (Sweden), ITB (Indonesia), UTEC (Peru), Politecnico di Torino, SMART (Singapore), AMS Institute, and the cities of Laval, Curitiba, Stockholm, Amsterdam, and Helsingborg) for funding this research. We thank the New York City Mayor's Office of the Chief Technology Officer, the Department of Health and Mental Hygiene, the Department of Citywide Administrative Services, as well as the EDF - Environmental Defense Fund, for providing the sensing fleet and providing feedback on the research work.

■ REFERENCES

(1) Haley, V. B.; Talbot, T. O.; Felton, H. D. Surveillance of the short-term impact of fine particle air pollution on cardiovascular disease hospitalizations in New York State. *Environmental Health* **2009**, *8* (1), 42.

(2) Eilstein, D. Exposition prolongée à la pollution atmosphérique et mortalité par pathologies respiratoires. *Revue Française d'Allergologie* **2010**, *50* (2), 51−61.

(3) Kheirbek, I.; Haney, J.; Douglas, S.; Ito, K.; Matte, T. The contribution of motor vehicle emissions to ambient fine particulate matter public health impacts in New York City: a health burden assessment. *Environmental Health* **2016**, *15* (1), 89.

(4) 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.; Hamburg, S. P. High-resolution air pollution mapping with Google street view cars: exploiting big data. *Environ. Sci. Technol.* **2017**, *51* (12), 6999−7008.

(5) Hofman, J.; Samson, R.; Joosen, S.; Blust, R.; Lenaerts, S. Cyclist exposure to black carbon, ultrafine particles and heavy metals: An experimental study along two commuting routes near Antwerp, Belgium. *Environmental research* **2018**, *164*, 530−538.

(6) Do, T. H.; Tsiligianni, E.; Qin, X.; Hofman, J.; La Manna, V. P.; Philips, W.; Deligiannis, N. Graph-Deep-Learning-Based Inference of Fine-Grained Air Quality from Mobile IoT Sensors. *IEEE Internet of Things Journal* **2020**, *7* (9), 8943−8955.

(7) Van den Bossche, J.; Peters, J.; Verwaeren, J.; Botteldooren, D.; Theunis, J.; De Baets, B. Mobile monitoring for mapping spatial variation in urban air quality: Development and validation of a methodology based on an extensive dataset. *Atmos. Environ.* **2015**, *105*, 148−161.

(8) Restrepo, C.; Zimmerman, R.; Thurston, G.; Clemente, J.; Gorczynski, J.; Zhong, M.; Blaustein, M.; Chen, L. C. A comparison of ground-level air quality data with New York State Department of Environmental Conservation monitoring stations data in South Bronx, New York. *Atmos. Environ.* **2004**, *38* (31), 5295−5304.

(9) Hu, Z.; Bai, Z.; Yang, Y.; Zheng, Z.; Bian, K.; Song, L. UAV aided aerial-ground IoT for air quality sensing in smart city: Architecture, technologies, and implementation. *IEEE Network* **2019**, *33* (2), 14−22.

(10) DeSouza, P.; Anjomshoaa, A.; Duarte, F.; Kahn, R.; Kumar, P.; Ratti, C. Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: Methods development and lessons learned. *Sustainable Cities and Society* **2020**, *60*, 102239.

(11) Hasenfratz, D.; Saukh, O.; Walser, C.; Hueglin, C.; Fierz, M.; Thiele, L. Pushing the Spatio-Temporal Resolution Limit of Urban Air Pollution Maps. In *IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE*, **2014**; pp 69−77.

(12) Hankey, S.; Marshall, J. D. Land use regression models of on-road particulate air pollution (particle number, black carbon, PM2. 5, particle size) using mobile monitoring. *Environ. Sci. Technol.* **2015**, *49* (15), 9194−9202.

(13) 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* (7), 3938−3947.

(14) Dulal, H. B.; Brodnig, G.; Onoriose, C. G. Climate change mitigation in the transport sector through urban planning: A review. *Habitat International* **2011**, *35* (3), 494−500.
(15) Guidance, E., Improving Air Quality Through Land Use Activities. In *Transportation and Regional Programs Division, Office of Transportation and Air Quality, US Environmental Protection Agency*, **2005**.
(16) Tsai, Y.-H. Quantifying urban form: compactness versus' sprawl'. *Urban studies* **2005**, *42* (1), 141−161.
(17) Habermann, M.; Billger, M.; Haeger-Eugensson, M. Land use regression as method to model air pollution. Previous results for Gothenburg/Sweden. *Procedia Engineering* **2015**, *115*, 21−28.
(18) Janhäll, S. Review on urban vegetation and particle air pollution−Deposition and dispersion. *Atmospheric environment* **2015**, *105*, 130−137.
(19) Simpson, J. E. *Sea breeze and local winds*. Cambridge University Press, **1994**.
(20) Kang, J. E.; Yoon, D. K.; Bae, H.-J. Evaluating the effect of compact urban form on air quality in Korea. *Environment and Planning B: Urban Analytics and City Science* **2019**, *46* (1), 179−200.
(21) Lee, C. Impacts of urban form on air quality in metropolitan areas in the United States. *Computers, Environment and Urban Systems* **2019**, *77*, 101362.
(22) Liang, Z.; Wei, F.; Wang, Y.; Huang, J.; Jiang, H.; Sun, F.; Li, S. The Context-Dependent Effect of Urban Form on Air Pollution: A Panel Data Analysis. *Remote Sensing* **2020**, *12* (11), 1793.
(23) Wong, M.; Nochol, J.; Ng, E.; Guilbert, E.; Kwok, K.; To, P.; Wang, J. GIS Techniques for mapping urban ventilation, using frontal area index and least cost path analysis. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2010**, *38*, 586−591.
(24) Middel, A.; Lukasczyk, J.; Maciejewski, R.; Demuzere, M.; Roth, M. Sky View Factor footprints for urban climate modeling. *Urban climate* **2018**, *25*, 120−134.
(25) Zakšek, K.; Oštir, K.; Kokalj, Ž. Sky-view factor as a relief visualization technique. *Remote sensing* **2011**, *3* (2), 398−415.
(26) Ghassoun, Y.; Löwner, M.-O. Land use regression models for total particle number concentrations using 2D, 3D and semantic parameters. *Atmos. Environ.* **2017**, *166*, 362−373.
(27) Shi, Y.; Lau, K. K.-L.; Ng, E. Developing street-level PM2. 5 and PM10 land use regression models in high-density Hong Kong with urban morphological factors. *Environ. Sci. Technol.* **2016**, *50* (15), 8178−8187.
(28) Tang, R.; Blangiardo, M.; Gulliver, J. Using building heights and street configuration to enhance intraurban PM10, NOx, and NO2 Land use regression models. *Environ. Sci. Technol.* **2013**, *47* (20), 11643−11650.
(29) Edussuriya, P.; Chan, A.; Ye, A. Urban morphology and air quality in dense residential environments in Hong Kong. Part I: District-level analysis. *Atmos. Environ.* **2011**, *45* (27), 4789−4803.
(30) Jacob, D. J.; Winner, D. A. Effect of climate change on air quality. *Atmospheric environment* **2009**, *43* (1), 51−63.
(31) Seaman, N. L. Meteorological modeling for air-quality assessments. *Atmospheric environment* **2000**, *34* (12−14), 2231−2259.
(32) Arain, M.; Blair, R.; Finkelstein, N.; Brook, J.; Sahsuvaroglu, T.; Beckerman, B.; Zhang, L.; Jerrett, M. The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies. *Atmos. Environ.* **2007**, *41* (16), 3453−3464.
(33) Vienneau, D.; De Hoogh, K.; Briggs, D. A GIS-based method for modelling air pollution exposures across Europe. *Sci. Total Environ.* **2009**, *408* (2), 255−266.
(34) Naughton, O.; Donnelly, A.; Nolan, P.; Pilla, F.; Misstear, B.; Broderick, B. A land use regression model for explaining spatial variation in air pollution levels using a wind sector based approach. *Sci. Total Environ.* **2018**, *630*, 1324−1334.
(35) Contreras, L.; Ferri, C. Wind-sensitive interpolation of urban air pollution forecasts. *Procedia Computer Science* **2016**, *80*, 313−323.
(36) Wilson, J. G.; Kingham, S.; Pearce, J.; Sturman, A. P. A review of intraurban variations in particulate air pollution: Implications for epidemiological research. *Atmos. Environ.* **2005**, *39* (34), 6444−6462.
(37) Luttinger, D.; Wilson, L. A study of air pollutants and acute asthma exacerbations in urban areas: status report. *Environ. Pollut.* **2003**, *123* (3), 399−402.
(38) Maciejczyk, P. B.; Offenberg, J. H.; Clemente, J.; Blaustein, M.; Thurston, G. D.; Chen, L. C. Ambient pollutant concentrations measured by a mobile laboratory in South Bronx, NY. *Atmospheric environment* **2004**, *38* (31), 5283−5294.
(39) Corburn, J.; Osleeb, J.; Porter, M. Urban asthma and the neighbourhood environment in New York City. *Health & place* **2006**, *12* (2), 167−179.
(40) Alphasense OPC-N3 Particle Monitor. For use in high polution urban environments. https://www.alphasense.com/index.php/air/downloads/OPC-N3.pdf.
(41) Crilley, L. R.; Shaw, M.; Pound, R.; Kramer, L. J.; Price, R.; Young, S.; Lewis, A. C.; Pope, F. D. Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring. *Atmospheric Measurement Techniques* **2018**, *11* (2), 709−720.
(42) Sousan, S.; Koehler, K.; Hallett, L.; Peters, T. M. Evaluation of the Alphasense optical particle counter (OPC-N2) and the Grimm portable aerosol spectrometer (PAS-1.108). *Aerosol Sci. Technol.* **2016**, *50* (12), 1352−1365.
(43) Brantley, H.; Hagler, G.; Kimbrough, E.; Williams, R.; Mukerjee, S.; Neas, L. Mobile air monitoring data-processing strategies and effects on spatial air pollution trends. *Atmospheric measurement techniques* **2014**, *7* (7), 2169−2183.
(44) Meng, X.; Chen, L.; Cai, J.; Zou, B.; Wu, C.-F.; Fu, Q.; Zhang, Y.; Liu, Y.; Kan, H. A land use regression model for estimating the NO2 concentration in Shanghai, China. *Environmental research* **2015**, *137*, 308−315.
(45) Tian, Y.; Yao, X.; Chen, L. Analysis of spatial and seasonal distributions of air pollutants by incorporating urban morphological characteristics. *Computers, Environment and Urban Systems* **2019**, *75*, 35−48.
(46) Liu, Y.; Wu, J.; Yu, D. Characterizing spatiotemporal patterns of air pollution in China: A multiscale landscape approach. *Ecological Indicators* **2017**, *76*, 344−356.
(47) Liu, Y.; Wu, J.; Yu, D.; Ma, Q. The relationship between urban form and air pollution depends on seasonality and city size. *Environmental Science and Pollution Research* **2018**, *25*, 1−14.
(48) Buyantuyev, A.; Wu, J.; Gries, C. Multiscale analysis of the urbanization pattern of the Phoenix metropolitan landscape of USA: time, space and thematic resolution. *Landscape and Urban Planning* **2010**, *94* (3−4), 206−217.
(49) Borrego, C.; Martins, H.; Tchepel, O.; Salmim, L.; Monteiro, A.; Miranda, A. I. How urban structure can affect city sustainability from an air quality perspective. *Environmental modelling & software* **2006**, *21* (4), 461−467.
(50) Bechle, M. J.; Millet, D. B.; Marshall, J. D. Effects of income and urban form on urban NO2: Global evidence from satellites. *Environ. Sci. Technol.* **2011**, *45* (11), 4914−4919.
(51) Martins, H. Urban compaction or dispersion? An air quality modelling study. *Atmospheric environment* **2012**, *54*, 60−72.
(52) Cho, H.-S.; Choi, M. Effects of compact urban development on air pollution: Empirical evidence from Korea. *Sustainability* **2014**, *6* (9), 5968−5982.
(53) Rodríguez, M. C.; Dupont-Courtade, L.; Oueslati, W. Air pollution and urban structure linkages: Evidence from European cities. *Renewable and Sustainable Energy Reviews* **2016**, *53*, 1−9.
(54) McGarigal, K. *FRAGSTATS help. University of Massachusetts*; Amherst, M., Ed., 2015.
(55) Burian, S.; Brown, M.; Linger, S. *Morphological Analyses using 3D Building Databases*, LAUR020781; Los Alamos National Laboratory: Los Angeles, CA, 2002; pp 36−42.
(56) Lindberg, F.; Grimmond, C. Continuous sky view factor maps from high resolution urban digital elevation models. *Climate Research* **2010**, *42* (3), 177−183.
(57) Britter, R.; Hanna, S. Flow and dispersion in urban areas. *Annu. Rev. Fluid Mech.* **2003**, *35* (1), 469−496.

https://doi.org/10.1021/acs.est.1c04854
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

(58) Shen, J.; Gao, Z.; Ding, W.; Yu, Y. An investigation on the effect of street morphology to ambient air quality using six real-world cases. *Atmos. Environ.* **2017**, *164*, 85−101.
(59) Hall, M. A., Correlation-based feature selection for machine learning **1999**.
(60) 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*, 105827.
(61) Tian, Y.; Yao, X. A.; Mu, L.; Fan, Q.; Liu, Y. Integrating meteorological factors for better understanding of the urban form-air quality relationship. *Landscape Ecology* **2020**, *35* (10), 2357−2373.
(62) Breiman, L. Random forests. *Machine learning* **2001**, *45* (1), 5−32.
(63) Gardner, M. W.; Dorling, S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment* **1998**, *32* (14−15), 2627−2636.
(64) Perrier, A. Feature Importance in Random Forests. https://alexisperrier.com/datascience/2015/08/27/feature-importance-random-forests-gini-accuracy.html.
(65) Hodgson, S.; Nieuwenhuijsen, M. J.; Colvile, R.; Jarup, L. Assessment of exposure to mercury from industrial emissions: comparing "distance as a proxy" and dispersion modelling approaches. *Occupational and environmental medicine* **2006**, *64* (6), 380−388.
(66) Maantay, J. A.; Tu, J.; Maroko, A. R. Loose-coupling an air dispersion model and a geographic information system (GIS) for studying air pollution and asthma in the Bronx, New York City. *International Journal of Environmental Health Research* **2009**, *19* (1), 59−79.
(67) Heo, H. K.; Lee, D. K.; Park, C. Y.; Kim, H. G. Sky view factor calculation in complex urban geometry with terrestrial LiDAR. *Physical Geography* **2021**, *42*, 1−21.
(68) Bellander, T.; Berglind, N.; Gustavsson, P.; Jonson, T.; Nyberg, F.; Pershagen, G.; Järup, L. Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. *Environ. Health Perspect.* **2001**, *109* (6), 633−639.
(69) Liu, M.; Chen, H.; Wei, D.; Wu, Y.; Li, C. Nonlinear relationship between urban form and street-level PM2. 5 and CO based on mobile measurements and gradient boosting decision tree models. *Building and Environment* **2021**, *205*, 108265.
(70) Wang, A.; Xu, J.; Tu, R.; Saleh, M.; Hatzopoulou, M. Potential of machine learning for prediction of traffic related air pollution. *Transportation Research Part D: Transport and Environment* **2020**, *88*, 102599.

https://doi.org/10.1021/acs.est.1c04854
Environ. Sci. Technol. XXXX, XXX, XXX−XXX