Environmental Science & Technology

pubs.acs.org/est

Article

National Empirical Models of Air Pollution Using Microscale Measures of the Urban Environment

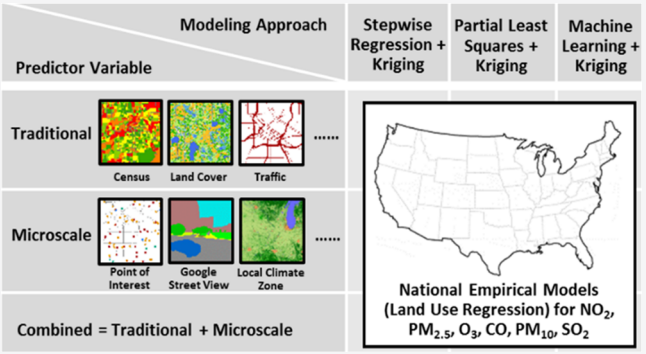
Tianjun Lu,* Julian D. Marshall, Wenwen Zhang, Perry Hystad, Sun-Young Kim, Matthew J. Bechle, Matthias Demuzere, and Steve Hankey

Cite This: *Environ. Sci. Technol.* 2021, 55, 15519−15530

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: National-scale empirical models of air pollution (e.g., Land Use Regression) rely on predictor variables (e.g., population density, land cover) at different geographic scales. These models typically lack microscale variables (e.g., street level), which may improve prediction with fine-spatial gradients. We developed microscale variables of the urban environment including Point of Interest (POI) data, Google Street View (GSV) imagery, and satellite-based measures of urban form. We developed United States national models for six criteria pollutants (NO_2, $PM_{2.5}$, O_3, CO, PM_{10}, SO_2) using various modeling approaches: Stepwise Regression + kriging (SW-K), Partial Least Squares + kriging (PLS-K), and Machine Learning + kriging (ML-K). We compared predictor variables (e.g., traditional vs microscale) and emerging modeling approaches (ML-K) to well-established approaches (i.e., traditional variables in a PLS-K or SW-K framework). We found that combined predictor variables (traditional + microscale) in the ML-K models outperformed the well-established approaches (10-fold spatial cross-validation (CV) R^2 increased 0.02−0.42 [average: 0.19] among six criteria pollutants). Comparing all model types using microscale variables to models with traditional variables, the performance is similar (average difference of 10-fold spatial CV R^2 = 0.05) suggesting microscale variables are a suitable substitute for traditional variables. ML-K and microscale variables show promise for improving national empirical models.



KEYWORDS: *Empirical models, street-level features, urban form, exposure assessment, machine learning*

1. INTRODUCTION

Ambient air pollution contains a mixture of particles and gases, many of which have adverse effects on human health.[1,2] To assess air-quality patterns at unmonitored locations, empirical modeling (earlier name: land use regression [LUR]) is a well-established method that is based on the correlation of monitoring data with predictor variables (e.g., land use and geographic factors).[3,4] Developing LUR models with an improved efficiency and spatial resolution is an important goal for improving the exposure assessment and understanding issues of health impacts,[5] health disparities,[6] environmental justice,[7] and urban planning.[8]

While there is a growing number of studies that develop empirical models for large geographies (e.g., national), traditional predictor variables (e.g., traffic intensity, land use types, and population dynamics) have limitations. For example, variables tabulated within Census geographies[9−11] may suffer from issues like the modifiable areal unit problem.[12] Variables assembled using data developed by administrative sources may miss local information,[13] making it difficult to generalize LUR models across regions.[14] Standardized land use (e.g., grid-based land cover data) and traffic data developed by federal agencies may capture regional air pollution emission sources but likely do not capture many local emission sources or features that modify air pollution dispersion.[15] This is partly because the process by which they are developed is slow, and they serve administrative needs that do not necessarily align with goals of air-quality modeling.[12] Data available at the microscale (e.g., street level) through emerging data science (e.g., data mining) and open-access data platforms may help address these issues.[16,17]

Google Point of Interest (POI) data have the potential to improve an empirical-model performance by capturing street-level attributes (e.g., restaurants, gas stations) related to air quality.[18] Similarly, Google Street View (GSV) imagery could provide information on the street-level built environment (e.g., greenness, infrastructure, sky view factor)[19−21] for modeling air

Received: June 18, 2021
Revised: October 22, 2021
Accepted: October 25, 2021
Published: November 5, 2021

https://doi.org/10.1021/acs.est.1c04047
Environ. Sci. Technol. 2021, 55, 15519−15530

Table 1. Candidate Predictor Variables Assembled for Developing LUR Models

variable scenario			variable category	variable name	variable type	spatial resolution	description	data source[b]
combined	microscale	traditional						
X		X	Geographic[a]	Traffic	Length/density in buffer (km)	0.05−15 km	Any road, truck route, intersections, etc.	TeleAtlas
X		X		Population	Count in buffer	Block group	Population in block groups (0.5−3 km)	US Census
X		X		Land use/ land cover	Area in buffer (%)	30 m	Built land, open space, agricultural land, etc. (0.05−15 km)	NLCD
X		X		Sources	Length in buffer (m)	Point	Distance to the nearest source (e.g., railroad, airport)	NEI
X		X		Emissions	Point in buffer (ton)	Point	Sum of site-specific facility emissions (3−30 km)	NEI
X		X		Vegetation	Area in buffer	30 m	Normalized difference vegetation index (0.5−10 km)	University of Maryland
X		X		Impervious	Area in buffer (%)	30 m	Impervious surface value (0.05−5 km)	NLCD
X		X		Elevation	Value	30 m	Elevation above sea levels; counts of points above or below a threshold (1−5 km)	USGS NED
X	X	X	Satellite[b]	Air pollution estimates	Column abundance or surface (μg/m3 or ppb)	10−25 km	Satellite-based air pollution estimates (NO_2, SO_2, CO, HCHO, $PM_{2.5}$)	Multiple sources[b]
X	X		POI[b]	Point of interest	Count in buffer	Point	90 categories of POI (e.g., gas station, restaurant)	Google
X	X		GSV[b]	Google street view	Object pixel (%)	Point	57 categories of ambient GSV-related features (e.g., tree, grass, person, building)	Google
X	X		LCZ[b]	Local climate zones	Counts in buffer	100 m	15 categories of LCZ (e.g., compact high-rise, dense trees)	Demuzere et al., 2020

[a]A detailed description can be found in the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air) Database (Kim et al., 2020). [b]A detailed description is in the Supporting Information.

pollution.[22,23] A newly developed satellite-based urban form database (Local Climate Zones [LCZ]) that classifies built and natural environments based on urban morphology (e.g., street configuration, building heights) and climate-related properties[24,25] may also be helpful in air-quality modeling.[26] While these microscale variables provide a consistent way of capturing local features across various regions, questions remain as to whether they can be used to improve empirical models for large geographies.

The inclusion of microscale predictor variables and traditional LUR predictor variables at multiple spatial scales may require more flexible modeling approaches. Traditional LUR models employ a stepwise regression approach[22,27] or hybrid modeling framework (e.g., partial least-squares in a kriging framework [hereafter PLS-K])[15] as well as many others (e.g., generalized additive model).[28,29] Emerging air-quality modeling studies report that machine learning (ML)-based models (e.g., random forest, neural network) allow for processing large data inputs with high predictive power and faster processing time.[30−33] An understudied topic is how ML-based approaches could be leveraged along with the microscale predictor variables to develop national LUR models.

In this study, we develop national LUR models to predict annual average concentrations of six criteria pollutants (i.e., NO_2, O_3, $PM_{2.5}$, CO, PM_{10}, and SO_2) in the contiguous United States based on regulatory monitor measurements. We compare LUR model performance using (1) different sets of predictor variables (i.e., traditional, microscale, and combined) and (2) different modeling approaches (i.e., traditional vs ML-based). We define traditional variables as including institutional data sets (e.g., Census-based geographic variables, satellite-based land cover data) and satellite-based air pollution estimates used in previous national models.[15] We develop a set of microscale variables that includes destinations derived from Google POI data, street-level built environment features for GSV imagery, and measures of urban form and morphology based on estimates of LCZs.[24] We compare combinations of predictor variables (e.g., traditional vs microscale vs a combination of all variables) and various established (Stepwise Regression + kriging [hereafter SW-K], PLS-K) and emerging (Machine Learning + kriging [hereafter ML-K]) modeling approaches. We focus our comparison on how emerging modeling approaches (ML-K) coupled with new microscale variables compare to well-established modeling approaches (e.g., SW-K or PLS-K with traditional variables). Our overarching goal is to evaluate whether microscale variables and ML approaches can be leveraged to improve empirical models.

2. MATERIALS AND METHODS

2.1. Regulatory Monitoring Data. We followed a previously published national air-quality modeling study and calculated annual average concentrations of criteria pollutants (i.e., NO_2, O_3, $PM_{2.5}$, CO, PM_{10}, SO_2) in year 2015 based on the U.S. Environmental Protection Agency (EPA) Air Quality System (AQS) monitoring locations.[15] All concentrations (except O_3) were annualized based on monitors with (1) at least 18 h of valid measurements per day, (2) at least 244 d per year, at least 61 d if reporting measurements each 3 d, or at least 41 d if reporting measurements each 6 d, and (3) no more than 45 consecutive days with no measurements. The concentrations of O_3 were annualized using the daily maximum

https://doi.org/10.1021/acs.est.1c04047
Environ. Sci. Technol. 2021, 55, 15519−15530

of the 8 h moving average at monitors with at least 18 h/d during the ozone season (i.e., May to September). All the other pollutants were based on a 24 h daily mean. To meet the normality assumption for linear models, concentrations were transformed by taking the square root following a previous study.[15]

2.2. Predictor Variables for LUR Modeling. We assembled multiple categories of candidate predictor variables for developing LUR models (Table 1) with the goal of exploring how different variable inputs impact the model performance. Generally, we used three sets of variable inputs: traditional only (i.e., geographic and satellite estimates of air quality), microscale only (i.e., microscale variables [POI, GSV, and LCZ] and satellite estimates of air quality), and combined (all candidate variables). We included satellite estimates of air quality in all models, since these variables have been shown to be important in several studies[13,15,34] and are available consistently across the U.S. While the traditional variables used in this study may include geographic and land use information at small spatial scales (e.g., road lengths in small buffers), we categorized all commonly used covariates as traditional variables.[15] We categorized POI, GSV, and LCZ as microscale variables, since these new variables were designed to capture street-level and neighborhood information.

2.2.1. Geographic Variables. We used the same geographic variables used in a previous study including eight categories of covariates (e.g., traffic, population, land use/land cover, and vegetation).[15] The variables were calculated and tabulated as count, length, and area within different buffers according to different data types, since it was unclear which size of buffers best represents each variable. This process resulted in ∼360 geographic variables for an LUR model development (Tables 1 and S1). Detailed data processing and variable calculation procedures are available elsewhere.[15]

2.2.2. Satellite-Based Estimates of Air Pollution. The annual average estimates of satellite-based air pollution concentrations (i.e., column abundance [atmospheric trace gas in the vertical column] or surface [ground-level estimates]) for NO_2, $PM_{2.5}$, SO_2, and CO were obtained from different satellite products and data sets (e.g., aerosol optical depth [AOD], ozone monitoring instrument [OMI]; some estimates were processed using chemical transport models).[35−37] The data for HCHO (formaldehyde) were a 12-year average (2005−2016) rather than annual average (Table S2). The resolution of these gridded satellite-based estimates varied (NO_2, $PM_{2.5}$, and HCHO: 0.1° × 0.1°; SO_2 and CO: 0.25° × 0.25°). The satellite-based estimates were assigned to the grid where each regulatory monitor was located.

2.2.3. Google Point of Interest (POI). To explore new/alternative land use data sets, we web-scraped Google POI from the Google Places application programming interface (API) that returns all point-based locations of interest within a specified buffer. There are 90 POI categories, and we collected POI data at 5 buffers (100m, 250m, 500m, 750m, 1000m) resulting in 450 total variables (Table S3). In general, the POI categories represent locations that might not be included in other data sets, for example, specific local businesses (e.g., gas stations, restaurants) that might have direct emissions. More importantly, the POI data may serve as a uniform and localized land use proxy for tracking changes over time and assessing air-quality impacts uniformly across administrative boundaries, allowing for more generalizable air-quality models across large geographies.

2.2.4. Google Street View (GSV). Another promising data source capable of capturing microscale features is GSV imagery, which provides georeferenced images along major road networks.[21] We processed available GSV images (640 × 640 pixels) nearest to year 2015 at locations near all EPA regulatory monitors. Specifically, at least five random locations were sampled within a 100m buffer of each monitoring location; each image location was required to be 20m apart to properly capture the surrounding environment. We extracted four images in four directions (0°, 90°, 180°, 270°) at each sampling location. This step resulted in at least 20 GSV images per monitoring location and a total of 5470 images around all monitoring locations. Then, we used a deep learning image segmentation algorithm (i.e., pyramid scene parsing network [PSPNet]) to classify each pixel in the image into 150 feature categories, including natural and built environment features (e.g., trees, buildings, and cars).[38] We summarized each image to give a percentage of all feature categories (Figures S1 and S2). While the original algorithm was developed to characterize both indoor and outdoor features, we only tabulated 57 categories of outdoor-related GSV features (Table S4) for the purpose of ambient air-quality modeling. For model development, the results among all images at the GSV sampling locations within 100m of monitoring locations (at least 20 GSV images per monitor) were averaged to obtain the final GSV-based variables.

2.2.5. Local Climate Zones (LCZ). An LCZ characterizes the urban form and morphology (e.g., building heights, layout) over time with a potential improvement over existing administrative data sets. We derived LCZ-based variables from the first U.S.-wide LCZ map.[24] LCZ captures a building configuration including a compact versus open arrangement, building height, and natural features. The LCZ map has a spatial resolution of 100m (with a lower thematic resolution due to the Gaussian filter applied) and covers 15 LCZ classes. It was produced in Google's Earth Engine based on a pixel-based random forest process using a combination of expert and crowd-sourced training data and satellite images.[24,39] While the LCZ reflects neighborhood-level information in a design for urban climatology (∼400 m radius), for simplicity, LCZs were grouped with the new microscale variables in our study. The full LCZ classifications (Figure S3) include built environment variables (LCZ 1−10) and land cover variables (LCZ A-G).[25] Figure S4 shows the LCZ surface for the continental U.S.; Figure S5 shows two selected urban areas. We summarized the counts of each LCZ category within five buffer sizes (i.e., 500m, 1000m, 1500m, 2000, 2500m) at all air-quality monitors. All calculations were conducted in ESRI ArcGIS (ver. 10.6).

2.3. Modeling Approach. We compared three modeling approaches (i.e., two traditional and one machine learning) using the three sets of candidate predictor variables (traditional; microscale; combined), resulting in nine combinations of modeling approaches and predictor variables.

2.3.1. Stepwise Regression + Kriging (SW-K). We added a kriging process (based on a minimum mean squared error interpolation) after a stepwise regression following a similar approach to that of Mercer et al.[40] For the stepwise regression, we used the same approach commonly used in LUR models.[22,41] The variable selection process included two steps: (1) selecting the most correlated variable with the dependent variable and (2) adding the variable that was most correlated with the model residuals among the remaining

https://doi.org/10.1021/acs.est.1c04047
Environ. Sci. Technol. 2021, 55, 15519−15530

		Random 10-fold CV						Spatial 10-fold CV					
		R^2			RMSE			R^2			RMSE		
		SW-K	PLS-K	ML-K	SW-K	PLS-K	ML-K	SW-K	PLS-K	ML-K	SW-K	PLS-K	ML-K
NO_2	traditional	0.88	0.86	0.90	0.21	0.23	0.19	0.83	0.73	0.79	0.26	0.33	0.28
	microscale	0.83	0.78	0.88	0.26	0.29	0.21	0.71	0.67	0.76	0.35	0.37	0.30
	combined	0.86	0.87	0.91	0.24	0.22	0.18	0.72	0.77	0.79	0.33	0.30	0.29
$PM_{2.5}$	traditional	0.73	0.83	0.88	0.16	0.13	0.11	0.65	0.64	0.67	0.19	0.19	0.18
	microscale	0.74	0.77	0.83	0.16	0.15	0.12	0.66	0.51	0.65	0.18	0.22	0.19
	combined	0.76	0.82	0.89	0.15	0.13	0.10	0.68	0.67	0.66	0.18	0.17	0.18
O_3	traditional	0.71	0.75	0.82	0.06	0.06	0.05	0.00	0.39	0.55	0.12	0.09	0.08
	microscale	0.36	0.76	0.72	0.09	0.06	0.06	0.00	0.49	0.31	0.12	0.08	0.10
	combined	0.71	0.75	0.82	0.06	0.06	0.05	0.45	0.45	0.55	0.09	0.09	0.08
CO	traditional	0.74	0.33	0.74	0.15	0.25	0.15	0.52	0.30	0.66	0.21	0.25	0.18
	microscale	0.71	0.35	0.66	0.16	0.24	0.16	0.49	0.33	0.60	0.22	0.25	0.19
	combined	0.67	0.50	0.80	0.17	0.20	0.14	0.53	0.38	0.71	0.21	0.24	0.16
PM_{10}	traditional	0.49	0.70	0.80	0.35	0.27	0.21	0.34	0.52	0.55	0.40	0.34	0.08
	microscale	0.64	0.64	0.75	0.29	0.30	0.24	0.46	0.46	0.62	0.36	0.36	0.31
	combined	0.71	0.72	0.82	0.27	0.26	0.21	0.51	0.54	0.57	0.35	0.33	0.32
SO_2	traditional	0.64	0.22	0.68	0.45	0.65	0.39	0.48	0.15	0.51	0.53	0.68	0.52
	microscale	0.54	0.11	0.61	0.50	0.70	0.46	0.38	0.08	0.24	0.58	0.71	0.64
	combined	0.53	0.20	0.59	0.51	0.66	0.42	0.46	0.18	0.57	0.54	0.67	0.48

R^2 Low High RMSE Low High

Figure 1. Random and spatial 10-fold CV results of models for six criteria pollutants. SW-K is Stepwise regression + kriging; PLS-K is Partial Least Squares + kriging; ML-K is Machine Learning + kriging. RMSE is standardized (i.e., RMSE/mean concentrations of all monitors).

variables. This process was repeated until either a variable was not significant ($p > 0.05$) or the multicollinearity indicator (variance inflation factor [VIF]) was greater than 5. We allowed variables to be selected with multiple buffers, since the main purpose of this study is to make predictions, and we were not sure which buffer would be the best. For the second-stage kriging process, we incorporated spatial smoothing by kriging the residuals after estimating a trend in the stepwise regression.

2.3.2. Partial Least Squares + Kriging (PLS-K). We employed the same PLS-K modeling approach used in our previous LUR study.[15] The modeling process involved two major components: variance and mean. Specifically, the universal kriging (using exponential covariance function for variogram) accounts for the variance component, and PLS accounts for the mean component by reducing the dimensions of independent variables used in a linear regression process. All kriging covariance parameters and PLS summary variables were based on a maximum likelihood approach. Following a previous study,[15] which highlighted that parsimonious models with 3−30 variables outperformed models with all variables, we first selected the top 30 variables by a forward selection, then used those variables in a PLS reduction and regression modeling. Specifically, we used a dimension reduction approach instead of a stepwise selection; the dimension reduction approach estimates a few summary predictors ($n = 2$ or 3 components depending on the cross-validation (CV) process in Section 2.4). Table S7 shows the top 30 variables selected as predictor variables for PLS-K models. Details are in a report by Kim et al.[15]

2.3.3. Machine Learning + Kriging (ML-K). Finally, to select our best ML algorithm, we first conducted experiments with eight common ML algorithms (i.e., ridge regression, elasticnet, lasso, adaboost, bagging, random forest, gradient boosting, and extratrees) integrated in Python scikit-learn packages (Python ver. 3.6.10; Scikit-learn ver. 0.22.1[42]) to develop LUR models for all three sets of variables (traditional; microscale; combined). We fine-tuned the parameters using various combinations of hyperparameters (Table S5) and selected the ML algorithm with the best performance (highest predictive power and lowest error; metrics are shown in Section 2.4). Then, similar to SW-K, we also added a second-stage kriging step after estimating a trend using the ML algorithms. We kriged the residuals of the ML algorithm with the best performance (using the exponential covariance function for variogram).

2.4. Modeling Evaluation. We conducted two types of 10-fold cross-validation to evaluate the model performance: a conventional CV (hereafter random CV), which divided the monitoring locations into 10 groups randomly, and a spatially clustered CV (hereafter spatial CV), which divided the 10 groups spatially using k-means clustering for training and testing.[15,43] Each CV separately involves 10 iterations of the following processes: (1) selecting one group out of the 10 groups as the hold-out group and (2) developing models using the remaining nine groups to predict concentrations at the hold-out group. Random CV reflects the model performance at random monitoring locations; spatial CV reflects model performance for locations distant from a monitor.

We used standardized root-mean-square error (RMSE) and mean square error (MSE)-based R^2 to evaluate the CV performance. Standardized RMSE (i.e., RMSE/mean concentrations of all monitors; hereafter RMSE) allows for a comparison across the criteria pollutants. The MSE-R^2 (i.e., one minus the sum of squared error between the observations and the predictions divided by the sum of squared error between the observations and the mean of the observations; hereafter R^2) assesses the agreement between observations and predictions on the 1:1 line instead of the regression line.[15]

2.5. Comparison of Variable Inputs and Modeling Approaches. To investigate how the LUR models performed among different scenarios, we focused on several comparisons below. We compared the model performance across the three sets of variables (traditional; microscale; combined) with the focus on whether (1) the microscale variables could be alternative choices for LUR models when traditional data sources are not available and (2) whether adding the

https://doi.org/10.1021/acs.est.1c04047
Environ. Sci. Technol. 2021, 55, 15519−15530

microscale variables to the traditional variables could improve model performance. Our ultimate goal was to compare combinations of predictor variables and modeling types to well-established LUR models (e.g., SW-K or PLS-K with traditional variables). We compared results primarily using the two types of CV. Finally, we compared the variable importance for different models and pollutants.

To further compare the model performance, we also evaluated the model performance among quantiles of air pollution concentrations and regions of monitoring locations. We used the mean normalized bias (MNB) to evaluate the model performance: MNB = $\sum_{i=1}^{n}$(concentration estimates, i −raw concentrations, i)/ $\sum_{i=1}^{n}$(raw concentrations, i). First, we compared the model performance by a variable input and modeling approach based on 10 quantiles of air pollution concentrations. Second, we compared the model performance based on monitoring locations according to the four Census Bureau-designated regions (i.e., West, Midwest, Northeast, and South).[44] Third, we compared the model performance between monitors in urban areas and rural areas according to the U.S. Census classification.[45] Lastly, we used the Environmental Justice Screening and Mapping Tool (EJSCREEN)[46] to compare the model performance between monitors in low-income minority areas (with a demographic index over 75% percentile) to all other areas. We used the MNB to evaluate the performance of all analyses.

3. RESULTS AND DISCUSSION

3.1. Summary of Monitored Air Pollution Concentrations. The number of valid regulatory monitors in 2015 ranges from 196 (CO) to 821 (O_3). Annual mean (median) concentrations were 7.8 (7.1) ppb [NO_2], 7.7 (8.0) $\mu g/m^3$ [$PM_{2.5}$], 44.2 (44.5) ppb [O_3], 0.3 (0.3) ppm [CO], 16.0 (15.6) $\mu g/m^3$ [PM_{10}], and 1.2 (1.0) ppb [SO_2]. Detailed descriptive statistics are reported in Table S6.

3.2. LUR Model Performance by Variable and Modeling Approach. Among the eight ML algorithms, gradient boosting and random forest generally showed the best performance (Figures S6−S9). Gradient boosting optimizes the model prediction in an iterative fashion by fitting on the negative gradients. Random forest has a set of decision trees (constructed by the best splits randomly chosen through subset predictors) averaging for regression results in the final prediction. Since gradient boosting performed slightly better for the most widely modeled criteria pollutants (i.e., NO_2 and $PM_{2.5}$), we used the gradient boosting algorithm to represent the ML approach.

Figure 1 shows the random and spatial 10-fold CV results of the six criteria pollutants. When comparing the various combinations of predictor variables alone, we found that the microscale variables may be a useful substitute for traditional variables. For example, LUR models using the microscale variables alone performed similarly to models using the traditional variables (average difference in the model random [spatial] CV $R^2 \approx 0.06$ [0.05] among the three modeling approaches; CV RMSE ≈ 0.02 [0.03]). We found similar results when we compared models using combined variables to those developed with the traditional variables alone (average difference in the model random [spatial] CV $R^2 \approx 0.01$ [0.05] among the three modeling approaches; CV RMSE ≈ 0 [0.01]). While the model performance did not always improve using the microscale variables for each pollutant, our findings suggest that microscale variables are a potential substitute for traditional variables in LUR models.

When comparing the modeling approaches using the same predictor variables, we found that ML-K models showed the best performance. For example, ML-K models among the three variable inputs generally outperformed SW-K (average difference in model random [spatial] CV $R^2 \approx 0.10$ [0.11]; CV RMSE ≈ 0.04 [0.04]) and PLS-K (average difference in model random [spatial] CV $R^2 \approx 0.17$ [0.14]; CV RMSE ≈ 0.08 [0.06]).

We also compared the model type and predictor variable combinations to the well-established modeling approaches (e.g., traditional variables combined with PLS-K). We found that, when including microscale variables in combination with machine learning approaches, it was possible to improve the model performance, but improvements varied by pollutant. For example, using combined (traditional + microscale) predictor variables in the ML-K models achieved a better performance for all pollutants based on the 10-fold random and spatial CVs (average random [spatial] CV R^2 increased 0.19 [0.19] among six criteria pollutants; CV RMSE decreased 0.08 [0.06]). The models improved the most for CO and SO_2 (average random [spatial] CV R^2 increased 0.42 [0.42]; CV RMSE decreased 0.17 [0.15]). However, when using PLS-K, there were minimal changes between models that used traditional alone versus microscale alone variables (average difference of random [spatial] CV $R^2 \approx 0.05$ [0.03]; CV RMSE ≈ 0.03 [0.02]) or traditional versus combined variables (average difference of random [spatial] CV $R^2 \approx 0.03$ [0.04]; CV RMSE ≈ 0.01 [0.01]). When comparing ML-K models using combined variables to models using traditional variables, the difference in the model performance was small (average 10-fold random [spatial] CV R^2 increased by 0.01 [0.02]; CV RMSE increased by 0 [0.03]). These findings suggest that the combination of ML-K models with the addition of microscale variables have the potential to improve the LUR model performance and that much of the increase in the model performance may be attributable to the use of an ML-based modeling approach.

An important finding of our work is that ML-K models using the microscale variables may be more helpful for pollutants that are less commonly included in LUR studies at the national scale. For example, when using the microscale variables for CO, PM_{10}, and SO_2, ML-K outperformed PLS-K (average random [spatial] CV R^2 increased 0.31 [0.20]; CV RMSE decreased 0.13 [0.06]) and SW-K (average random [spatial] CV R^2 increased 0.04 [0.04] among CO, PM_{10}, and SO_2; CV RMSE decreased 0.03 [0.01]). These results suggest that the microscale variables improve the model performance to some degree but likely require the use of ML-based modeling tools.

An additional benefit of using the ML-K approach is that it runs faster than other modeling approaches. For example, the time needed on a desktop computer (random access memory: 3.00 GB) to develop models for $PM_{2.5}$ (with 757 monitoring locations) was ∼135 min when using PLS-K and less than 30 min for ML-K. This difference could be helpful when running many different modeling scenarios. ML-K in conjunction with the microscale variables could be a more practical and efficient modeling framework while ensuring reliable performance for the criteria pollutants; however, further studies may be needed to confirm the importance of microscale variables, especially across administrative boundaries (e.g., multiple countries).

3.3. LUR Model Performance by CV Type. As expected, and consistent with the literature, random CV consistently

https://doi.org/10.1021/acs.est.1c04047
Environ. Sci. Technol. 2021, 55, 15519−15530

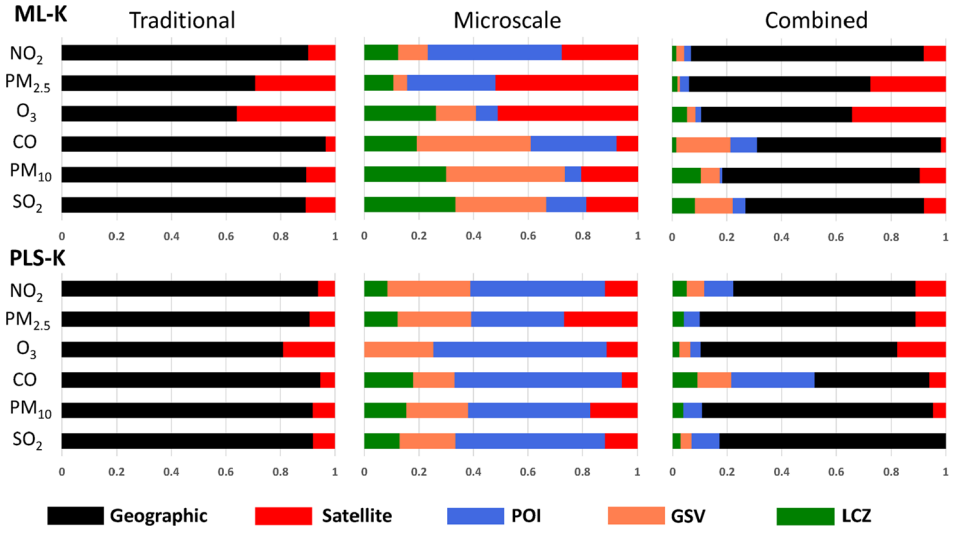


Figure 2. Sum of variable importance of all variables by variable type (traditional vs microscale vs combined) for the ML-K models (upper panels: relative importance scores) and PLS-K models (bottom panels: weighted VIP score).

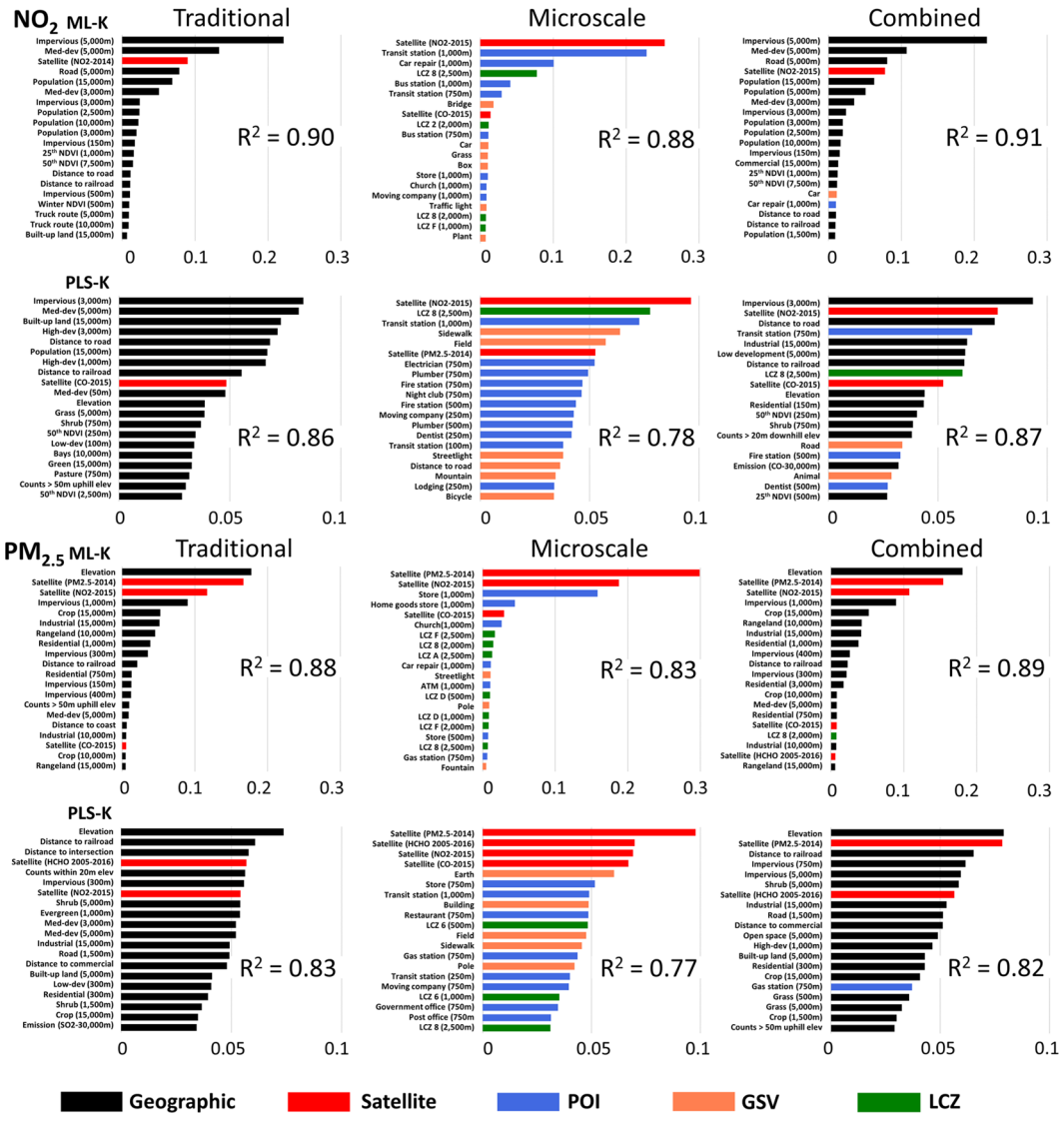


Figure 3. Variable importance of the top 20 variables by variable type (traditional vs microscale vs combined) for the ML-K models (upper panels: relative importance scores) and PLS-K models (bottom panels: weighted VIP score) of NO_2 and $PM_{2.5}$.

https://doi.org/10.1021/acs.est.1c04047
Environ. Sci. Technol. 2021, 55, 15519−15530

outperformed spatial CV (Figure S10). Among all pollutants and variable scenarios, the random CV model performance was better than the spatial CV using ML-K (average R^2 increased 0.18; RMSE decreased 0.06), PLS-K (R^2 increased 0.15; RMSE decreased 0.05), and SW-K (average R^2 increased 0.19; RMSE decreased 0.05), respectively. Among all pollutants and modeling approaches, the random CV model performance was better than the spatial CV with traditional variables (R^2 increased 0.19; RMSE decreased 0.05), microscale variables (R^2 increased 0.18; RMSE decreased 0.06), and combined variables (R^2 increased 0.15; RMSE decreased 0.05), respectively. A possible explanation of the relatively worse performance of the spatial CV could be that EPA monitoring locations share common attributes (e.g., land use types). These findings indicate that our model performance is likely less reliable where few or no EPA regulatory monitors are in close proximity to a prediction location.

3.4. Rank of Variable Importance. We compared the variable importance among pollutant, variable type, and modeling approach. We focused on the variable importance for the ML-K models and PLS-K models. For ML-K models, we used the relative importance score to represent importance ranks of each variable, which does not give any information on the direction of impact but does offer insight for a prediction (a trade-off of using ML-based approaches). For PLS-K models, we used the weighted variable importance in projection [VIP] score to measure the contribution of the predictor variables. Figure 2 shows the sum of variable importance of all variables by variable type (traditional vs microscale vs combined) for the ML-based models and PLS-K models.

In general, we found that both traditional and microscale variables were important in LUR models. Particularly, when using the traditional variables alone, the geographic variables were important in both ML-K and PLS-K models for all pollutants. While the importance of the geographic variables was also apparent in models with the combined variables, the microscale variables showed noticeable importance especially for CO, PM_{10}, and SO_2 suggesting that the microscale variables may be more helpful for these pollutants. When using the microscale variables alone, a mix of variable types (i.e., GSV, POI, and LCZ) showed a different magnitude of importance among the pollutant and modeling approaches. For example, GSV variables were more important for CO, PM_{10}, and SO_2, while POI variables were important for NO_2, $PM_{2.5}$, and CO in ML-K models. In comparison, POI variables showed importance for all pollutants in PLS-K models. This might be due to the two-stage ML modeling framework; since the PLS-K approach adjusts the coefficients on the PLS components it will likely give less weight to variables that correspond to regional variations.

Figure 3 shows the variable importance of the top 20 variables in ML-K (based on the relative importance scores) and PLS-K models (based on the weighted VIP score) for NO_2 and $PM_{2.5}$. For the individual variable importance, PLS-K included many of the microscale variables for NO_2 and relatively fewer for $PM_{2.5}$ for models using the combined variables. However, this trend was not clear when using ML-K. This result suggests that the microscale variables may offer information on the building configuration and function (e.g., amount of building in an image, industrial low-rise buildings) that might better capture factors that influence air pollution. However, components of some pollutants (e.g., $PM_{2.5}$ formed in the atmosphere) might make it difficult to model the association between the measurement and local land uses. Figures S11−S17 show the variable importance of the rest of the pollutants in the PLS-K, SW-K, and ML-K models.

3.5. LUR Model Performance by Low Versus High Concentration. The 10-quantile analyses showed that, for all pollutants studied, models tended to overestimate areas with the lowest concentrations and underestimate areas with the highest concentrations, regardless of variable type (Figures S18−S20). This finding is consistent with previous LUR studies.[47−49] Among the three modeling approaches, ML-K generally outperformed (e.g., lower deviations in MNB) SW-K and PLS-K. Second, models were more likely to underestimate in the West (e.g., for NO_2, $PM_{2.5}$, and O_3) and overestimate in the South (e.g., for NO_2 and PM_{10}) as well as the Northeast region in the U.S. (e.g., for $PM_{2.5}$, PM_{10}, and CO; Figures S21−S23). Similar to the stratification by concentration level, ML-K generally performed the best across the regional stratification of results. Finally, there was no obvious difference in the model performance between urban and rural areas (Figures S24−S26) or between low-income minority area versus all other areas for most pollutants. For example, for PM_{10}, models tend to underestimate concentrations in low-income minority communities and overestimate in all other categories indicating larger environmental justice (EJ) effects not captured by the models (Figures S27−S29).

3.6. Implications for Developing Future National Air-Quality Models. In this study, we developed national LUR models for estimating the annual average concentrations of six criteria pollutants using various combinations of predictor variables (e.g., microscale vs traditional) and modeling types (e.g., ML-K vs PLS-K). We found that the combined variables (microscale + traditional) in the ML-K models generally outperformed the well-established empirical models (e.g., traditional variables using PLS-K). We also found that models using the microscale variables did not always show a clear improvement for all pollutants. However, the finding that microscale variables are a suitable substitute for traditional variables suggests a promising application when modeling large geographies (e.g., across countries) where land use and other covariates are not harmonized. Additionally, the microscale variables have the potential to better align with ongoing efforts in the air-quality field to create relatively denser measurement networks (e.g., mobile monitoring and low-cost sensor networks). Therefore, the choice of the final "best" model will likely depend on the intended model application and data availability. Leveraging ML-based approaches to use microscale variables can replicate and sometimes improve traditional well-established approaches—a finding that warrants an investigation of these variables more closely with other data sets. As described below, our work could be helpful for selecting predictor variables and modeling approaches for empirical models.

3.6.1. Developing Empirical Models Using Microscale Variables. Our work focuses on the development of enhanced modeling tools capable of estimating air-pollution exposures at the national scale for environmental scientists, urban planners, and public health researchers. A unique aspect of our work is the use of microscale variables that characterize street-level and destination information in a consistent format across administrative boundaries. Unlike traditional variables that are often tabulated using aggregated data (e.g., Census geographies) to serve administrative needs or capture regional

https://doi.org/10.1021/acs.est.1c04047
Environ. Sci. Technol. 2021, 55, 15519−15530

predictor variables, our results suggest that microscale predictor variables could supplement traditional variables to reduce prediction errors while resolving the modifiable areal unit problem for LUR models. For larger geographic areas, microscale variables would stand out by offering a consistent data set while capturing street-level information for model development.

Our work has several implications for improving performance and reliability of empirical air-quality modeling. First, our microscale variables (i.e., GSV, POI, and LCZ) have the advantage of characterizing street-level factors to capture intraurban variations in air quality and help evaluate previously unmeasured emission sources in national models. For example, we found that POI variables (e.g., transit station and car repair) were important for NO_2 in our ML-K models; NO_2 is associated with traffic-related emissions.[50−52] Other POI variables (e.g., stores and gas stations) were found to be significant for estimating $PM_{2.5}$ concentrations.[11,53] Many of the microscale variables were local sources that may not be captured by traditional variables, which may be useful for future empirical modeling efforts. Second, our approach allows for improving the air-pollution exposure for larger geographies (e.g., national) with consistent variables across administrative boundaries, making it more convenient and reliable for conducting ongoing large-scale epidemiological studies.[54] Our models with the microscale variables achieved a similar performance to those with traditional variables, which is particularly significant in global cities or countries where traditional government-based data sources are scarce or less reliable at large spatial extents. Third, our results could offer insights into urban planning and policy. While many traditional predictor variables are indicative of long-term planning efforts on a macroscale (e.g., population density, land use), our microscale variables may inform or suggest interventions for a short-term improvement in microscale infrastructure. For example, we were able to identify that some infrastructure-based GSV variables (e.g., building, sidewalk, tree canopy) were associated with an air quality similar to that of other studies.[20−22,33,52,53] Lastly, our LCZ variables that reflect the urban morphology imply thermal urban characteristics and street canyon effects that may be relevant for the formation of (some of the) key pollutants,[24−26] which could be important for the model improvement. These findings suggest that empirical models with microscale variables could be used to provide air-pollution exposure estimates for epidemiological studies and help to inform strategies for the development of clean and healthy cities through sustainable urban planning strategies (e.g., green design of building, active transportation). Another advantage of using the microscale variables for LUR models is that the retrieval process typically involves only a small number of data sources (e.g., Google API) instead of assembling a variety of variables from multiple local and regional avenues.

3.6.2. Leveraging the Benefit of Microscale Variables Using ML-K Models. Our work is consistent with ongoing efforts to improve exposure assessment by (1) the development of improved models to explain more spatial variations of air pollutants while minimizing estimation biases, (2) harmonizing a consistent set of predictor variables across different areas, and (3) enhancing the modeling efficiency. In terms of model performance, we found that ML-K performed better than PLS-K and SW-K when using the microscale variables. Our work suggests that the use of ML-K is needed to maximize the advantages of using microscale variables for improving LUR models, especially for pollutants that are not as frequently studied in LUR literature (e.g., CO, PM_{10}, and SO_2). The improved performance for these pollutants may be explained by two factors. First, while we added microscale variables to potentially capture more local variations, we also included a range of buffer sizes including some at or above 1000m; the larger buffers may also help capture regional variations. Second, the flexibility of the ML-based approach was helpful when using the microscale variables as compared to the data shrinkage of the PLS approach. For example, the ML-based models are more likely to capture any nonlinear relationship between the predictor variables and the air pollution concentrations that may be missed by the PLS-based models with a focus on linear relationships.

Although our models overestimate low-concentration areas and underestimate high-concentration areas similar to other studies,[11,15,55] the ML-K generally showed the best performance. In addition, we found that the ML-K approach ran faster than the well-established models. The benefits of using ML-based models are also identified in other studies, including a minimized manual effort for predictor variables,[14] improved prediction accuracy,[9,31,56] and suitability for predicting air pollution for large geographies.[30,57] One could imagine testing our off-the-shelf ML-based approach to facilitate LUR model improvement by including a variety of new microscale variables from open data sources for different geographies.

3.6.3. Limitations and Future Research. Our work could be improved in several ways. First, more information from the microscale variables could potentially be extracted. We pulled the microscale variables at smaller buffers (≤2500 m) than the geographic variables (≤10 km) with the goal of focusing on the addition of information at smaller spatial scales. Future work could explore how larger buffers of the microscale variables could contribute to the model performance when comparing to the traditional variables. For GSV variable tabulation, we simply summarized each image by classifying multiple categories (e.g., tree, awning) based on the feature percentage. Additional work could tailor the need to LUR modeling by combining various measures into more appropriate groups (e.g., vegetation), using metrics to account for the diversity of the microscale features, or specifically classifying important air-pollution sources (e.g., trucks, small industrial sources). A better classification for GSV variables could also be achieved with growing repositories of street-view imagery[58] and emerging image-processing techniques based on computer vision and deep learning algorithms (aside from PSPNet).[59] Similarly, POI variables could benefit from an exhaustive POI data set across different websites and systems (e.g., Facebook Places, Yahoo! Local) to leverage the quantity and heterogeneity of the POI information.[60]

Second, the year of extraction for our microscale variables (∼2017) is unaligned with the year (2015) of our modeled pollutants, which may be an issue for microscale variables that are more likely to change quickly as compared to the traditional geographic variables (e.g., population density). Future work could benefit from collecting data across multiple time periods to characterize changes more accurately. For example, our LCZ variables have the advantage of tracking a street-level urban form in rapidly urbanized areas across many years—a promising prospect for the development of multiyear LUR models. Likewise, the combination of remote-sensed imagery and street-view imagery allows for improved temporal

https://doi.org/10.1021/acs.est.1c04047
Environ. Sci. Technol. 2021, 55, 15519−15530

precision (e.g., daily to yearly updates) to enhance LUR models.[61] Meanwhile, some microscale variables (e.g., GSV imagery) may not be available in rural areas, so models with traditional variables may be a better choice than those using microscale variables.

Third, since our main focus in this study is to explore the utility of microscale variables, we stopped at model building and model evaluation and did not make air-pollution predictions across the U.S. but only for a small downtown area. This is further hindered by a costly service fee for the Google API platform when a large number of GSV images and POI data is retrieved nationally. However, to illustrate how predictions might vary among approaches, we made predictions using both ML-K and PLS-K models of NO_2 and $PM_{2.5}$ for all variable types (i.e., traditional, microscale, and combined) for a single downtown area (with a buffer size of 1500m) in Blacksburg, Virginia—a small town in the U.S. (Figures S30 and S31). We also included a zoning map with major land uses for reference (Figure S32). We found that models using the microscale variables did capture more "hotspots" than models using the traditional and combined variables, which suggests potential benefits of using the microscale variables within an ML framework. As more alternative and freely open-source platforms (e.g., Bing, Open Street Maps, Yelp, and Baidu) become available, our approach could be extended to data sources where predictions might be more easily accomplished. Regardless of the data source, the general approach increases the possibility of the development of multicountry LUR models—an important and understudied topic.

Fourth, our work aims to assess predictor variables that are potentially useful for prediction rather than inference. We only compared a few modeling approaches in this study; a useful future study would be to compare more modeling approaches to assess trade-offs, especially the ones that could capture nonlinear relationships with more interpretable information. As such, while our variable importance may offer insights into a predictor variable selection for LUR models, the sensitivity to different modeling approaches and pollutants cautions the use and interpretation for developing LUR models.

Lastly, similar to other existing national LUR models, our models were developed based on the sparse regulatory monitoring network that was launched for the purpose of regulation compliance. This limitation resulted in a minor model improvement using the microscale variables. With the rapid development in mobile monitoring[41,62] and ubiquitous low-cost sensor networks (e.g., PurpleAir),[63,64] future work could focus on testing our approaches for these monitoring networks with a larger density and better coverage.

Our work suggests that the combination of ML-K and microscale variables hold promise for developing improved national empirical models as compared to the well-established models (i.e., PLS-K or SW-K with traditional variables). Importantly, our proposed modeling framework leverages the value of microscale variables with a finer spatial resolution (i.e., street level) and consistent data for the development of empirical models to track air-pollution exposure for large populations and geographies. More work is needed to determine what aspects of the microscale data are most advantageous or duplicative for empirical models, thus offering more insights into an exposure assessment.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.1c04047.

Candidate geographic predictor variables. Candidate predictor variables of satellite estimates. Categories of POI data. Categories of GSV data. Machine learning hyperparameters. Descriptive statistics of the criteria pollutants. The top 30 predictor variables used in PLS-K models. Example 1 of the extracted GSV features using the deep learning algorithm. Example 2 of the extracted GSV features using the deep learning algorithm. LCZ classification. LCZ surface for the continental United States. LCZ surface for two selected urban areas in the U.S. Random 10-fold CV results (R^2) of the eight ML algorithms. Random 10-fold CV results (RMSE) of the eight ML algorithms. Spatial 10-fold CV results (R^2) of the eight ML algorithms. Spatial 10-fold CV results (RMSE) of the eight ML algorithms. Comparison of random and spatial CV results. Variable importance of the top 20 variables by variable type and modeling approach. Model performance by 10 quantiles. Model performance by four Census Bureau-designated regions. Model performance by urban and rural area. Model performance by demographic index. Predicted NO_2 and $PM_{2.5}$ concentrations for downtown Blacksburg, Virginia. The zoning map of Blacksburg, VA, USA (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Tianjun Lu − *Department of Earth Science & Geography, California State University Dominguez Hills, Carson 90747 California, United States;* orcid.org/0000-0001-8991-4578; Phone: 1-310-243-3815; Email: tilu@csudh.edu

Authors

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

Wenwen Zhang − *Edward J. Bloustein School of Planning and Public Policy, Rutgers University, New Brunswick 08901 New Jersey, United States*

Perry Hystad − *College of Public Health and Human Sciences, Oregon State University, Corvallis 97331 Oregon, United States*

Sun-Young Kim − *Department of Cancer Control and Population Health, Graduate School of Cancer Science and Policy, National Cancer Center, Goyang-si 10408, Korea*

Matthew J. Bechle − *Department of Civil & Environmental Engineering, University of Washington, Seattle 98195 Washington, United States;* orcid.org/0000-0001-8076-5457

Matthias Demuzere − *Urban Climatology Group, Department of Geography, Ruhr-University Bochum, Bochum 44801, Germany*

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

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.1c04047

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This publication was developed as part of the Center for Clean Air Climate Solutions (CACES), which was supported under Assistance Agreement No. R835873 awarded by the U.S. Environmental Protection Agency (EPA). It has not been formally reviewed by the EPA. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication. M.D. is supported by the ENLIGHT project, funded by the German Research Foundation (DFG) under Grant No. 437467569. We thank Dr. A. A. Szpiro from the University of Washington for valuable comments on the model development. We also thank M. Qi from Virginia Tech for providing the Google Street View imagery variables of Blacksburg, Virginia.

■ REFERENCES

(1) Jerrett, M.; Burnett, R. T.; Pope, C. A.; Ito, K.; Thurston, G.; Krewski, D.; Shi, Y.; Calle, E.; Thun, M. Long-Term Ozone Exposure and Mortality. *N. Engl. J. Med.* **2009**, *360* (11), 1085−1095.

(2) Laden, F.; Schwartz, J.; Speizer, F. E.; Dockery, D. W. Reduction in Fine Particulate Air Pollution and Mortality: Extended Follow-up of the Harvard Six Cities Study. *Am. J. Respir. Crit. Care Med.* **2006**, *173* (6), 667−672.

(3) Adam-Poupart, A.; Brand, A.; Fournier, M.; Jerrett, M.; Smargiassi, A. Spatiotemporal Modeling of Ozone Levels in Quebec (Canada): A Comparison of Kriging, Land-Use Regression (LUR), and Combined. *Environ. Health Perspect.* **2014**, *970* (9), 970−976.

(4) Hoek, G.; Beelen, R.; Hoogh, K. De; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A Review of Land-Use Regression Models to Assess Spatial Variation of Outdoor Air Pollution. *Atmos. Environ.* **2008**, *42* (33), 7561−7578.

(5) Coleman, N. C.; Burnett, R. T.; Ezzati, M.; Marshall, J. D.; Robinson, A. L.; Pope, C. A., III Fine Particulate Matter Exposure and Cancer Incidence: Analysis of SEER Cancer Registry Data from 1992–2016. *Environ. Health Perspect.* **2020**, *128* (10), 107004.

(6) Hystad, P.; Setton, E.; Cervantes, A.; Poplawski, K.; Deschenes, S.; Brauer, M. Creating National Air Pollution Models for Population Exposure Assessment in Canada. *Environ. Health Perspect.* **2011**, *119* (8), 1123−1129.

(7) 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* (4), e94431.

(8) Hankey, S.; Lindsey, G.; Marshall, J. D. Population-Level Exposure to Particulate Air Pollution during Active Travel: Planning for Low-Exposure, Health-Promoting Cities. *Environ. Health Perspect.* **2017**, *125* (4), 527−535.

(9) 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* (13), 7233−7241.

(10) 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*, 9194−9202.

(11) Xu, H.; Bechle, M. J.; Wang, M.; Szpiro, A. A.; Vedal, S.; Bai, Y.; Marshall, J. D. National PM2.5 and NO2 Exposure Models for China Based on Land Use Regression, Satellite Measurements, and Universal Kriging. *Sci. Total Environ.* **2019**, *655* (2), 423−433.

(12) Hankey, S.; Zhang, W.; Le, H. T. K.; Hystad, P.; James, P. Predicting Bicycling and Walking Traffic Using Street View Imagery and Destination Data. *Transp. Res. part D Transp. Environ.* **2021**, *90*, 102651.

(13) Larkin, A.; Geddes, J. A.; Martin, R. V.; Xiao, Q.; Liu, Y.; Marshall, J. D.; Brauer, M.; Hystad, P. Global Land Use Regression Model for Nitrogen Dioxide Air Pollution. *Environ. Sci. Technol.* **2017**, *51* (12), 6957−6964.

(14) Lautenschlager, F.; Becker, M.; Kobs, K.; Steininger, M.; Davidson, P.; Krause, A.; Hotho, A. OpenLUR: Off-the-Shelf Air Pollution Modeling with Open Features and Machine Learning. *Atmos. Environ.* **2020**, *233*, 117535.

(15) Kim, S.; Bechle, M.; Hankey, S.; Sheppard, L.; Szpiro, 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* (2), No. e0228535.

(16) Bellinger, C.; Jabbar, M. S. M.; Zaiane, O.; Osornio-Vargas, A. A Systematic Review of Data Mining and Machine Learning for Air Pollution Epidemiology. *BMC Public Health* **2017**, *17* (1), 907.

(17) Zheng, Y.; Liu, F.; Hsieh, H. U-Air: When Urban Air Quality Inference Meets Big Data. *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery data Min. - KDD '13* **2013**, 1436−1444.

(18) Lu, T.; Lansing, J.; Zhang, W.; Bechle, M. J.; Hankey, S. Land Use Regression Models for 60 Volatile Organic Compounds: Comparing Google Point of Interest (POI) and City Permit Data. *Sci. Total Environ.* **2019**, *677* (2), 131−141.

(19) Middel, A.; Lukasczyk, J.; Maciejewski, R.; Demuzere, M.; Roth, M. Sky View Factor Footprints for Urban Climate Modeling. *Urban Clim.* **2018**, *25*, 120−134.

(20) Larkin, A.; Hystad, P. Evaluating Street View Exposure Measures of Visible Green Space for Health Research. *J. Exposure Sci. Environ. Epidemiol.* **2019**, *29*, 447−456.

(21) Rzotkiewicz, A.; Pearson, A. L.; Dougherty, B. V.; Shortridge, A.; Wilson, N. Systematic Review of the Use of Google Street View in Health Research: Major Themes, Strengths, Weaknesses and Possibilities for Future Research. *Heal. Place* **2018**, *52* (July), 240−246.

(22) Qi, M.; Hankey, S. Using Street View Imagery to Predict Street-Level Particulate Air Pollution. *Environ. Sci. Technol.* **2021**, *55* (4), 2695−2704.

(23) Hong, K. Y.; Pinheiro, P. O.; Weichenthal, S. Predicting Outdoor Ultrafine Particle Number Concentrations, Particle Size, and Noise Using Street-Level Images and Audio Data. *Environ. Int.* **2020**, *144*, 106044.

(24) Demuzere, M.; Hankey, S.; Mills, G.; Zhang, W.; Lu, T.; Bechtel, B. Combining Expert and Crowd-Sourced Training Data to Map Urban Form and Functions for the Continental US. *Sci. data* **2020**, *7* (1), 1−13.

(25) Stewart, I. D.; Oke, T. R. Local Climate Zones for Urban Temperature Studies. *Bull. Am. Meteorol. Soc.* **2012**, *93* (12), 1879−1900.

(26) Steeneveld, G.; Klompmaker, J. O.; Groen, R. J. A.; Holtslag, A. A. M. An Urban Climate Assessment and Management Tool for Combined Heat and Air Quality Judgements at Neighbourhood Scales. *Resour. Conserv. Recycl.* **2018**, *132*, 204−217.

(27) Eeftens, M.; Beelen, R.; De Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dedele, A.; Dons, E.; De Nazelle, A.; Dimakopoulou, K.; Eriksen, K.; Falq, G.; Fischer, P.; Galassi, C.; Grazuleviciene, R.; Heinrich, J.; Hoffmann, B.; Jerrett, M.; Keidel, D.; Korek, M.; Lanki, T.; Lindley, S.; Madsen, C.; Molter, A.; Nador, 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.; Varro, M. J.; Vienneau, D.; von Klot, S.; Wolf, K.; Brunekreef, B.; Hoek, G. Development of Land Use Regression Models for PM2.5, PM 2.5 Absorbance, PM10 and PMcoarse in 20 European Study Areas; Results of the ESCAPE Project. *Environ. Sci. Technol.* **2012**, *46* (20), 11195−11205.

(28) Li, L.; Wu, J.; Wilhelm, M.; Ritz, B. Use of Generalized Additive Models and Cokriging of Spatial Residuals to Improve Land-Use

Regression Estimates of Nitrogen Oxides in Southern California. *Atmos. Environ.* **2012**, *55*, 220−228.
(29) Hart, J. E.; Yanosky, J. D.; Puett, R. C.; Ryan, L.; Dockery, D. W.; Smith, T. J. Spatial Modeling of PM 10 and NO 2 in the Continental United States, 1985−2000. *Environ. Health Perspect.* **2009**, *1690* (2), 1690−1696.
(30) Di, Q.; Kloog, I.; Koutrakis, P.; Lyapustin, A.; Wang, Y.; Schwartz, J. Assessing PM2.5 Exposures with High Spatiotemporal Resolution across the Continental United States. *Environ. Sci. Technol.* **2016**, *50* (9), 4712−4721.
(31) 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.
(32) Rahman, M. M.; Karunasinghe, J.; Clifford, S.; Knibbs, L. D.; Morawska, L. New Insights into the Spatial Distribution of Particle Number Concentrations by Applying Non-Parametric Land Use Regression Modelling. *Sci. Total Environ.* **2020**, *702*, 134708.
(33) Weichenthal, S.; Hatzopoulou, M.; Brauer, M. A Picture Tells a Thousand··· Exposures: Opportunities and Challenges of Deep Learning Image Analyses in Exposure Science and Environmental Epidemiology. *Environ. Int.* **2019**, *122*, 3−10.
(34) Knibbs, L. D.; Hewson, M. G.; Bechle, M. J.; Marshall, J. D.; Barnett, A. G. A National Satellite-Based Land-Use Regression Model for Air Pollution Exposure Assessment in Australia. *Environ. Res.* **2014**, *135*, 204−211.
(35) Boersma, K. F.; Eskes, H. J.; Dirksen, R. J.; Veefkind, J. P.; Stammes, P.; Huijnen, V. An Improved Tropospheric NO2 Column Retrieval Algorithm for the Ozone Monitoring Instrument. *Atmos. Chem. Phys.* **2011**, *4* (2), 1905−1928.
(36) Deeter, M. N.; Edwards, D. P.; Francis, G. L.; Gille, J. C.; Martínez-alonso, S.; Worden, H. M. A Climate-Scale Satellite Record for Carbon Monoxide: The MOPITT Version 7 Product. *Atmos. Chem. Phys.* **2017**, *10*, 2533−2555.
(37) OMI Science Team. OMI/Aura Level 2 Sulphur Dioxide (SO2) Trace Gas Column Data 1-Orbit subset Swath along CloudSat track 1-Orbit Swath 13 × 24 km, Edited by GES DISC, Greenbelt, MD, USA Goddard Earth Sciences Data and Information Services Center (GES DISC) https://disc.gsfc.nasa.gov/datacollection/OMSO2_CPR_003.html [April 1, 2017].
(38) Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J.; Limited, S. G. Pyramid Scene Parsing Network. *IEEE Conf. Comput. Vis. Pattern Recognit.* **2017**, 6230−6239.
(39) Demuzere, M.; Bechtel, B.; Middel, A.; Mills, G. Mapping Europe into Local Climate Zones. *PLoS One* **2019**, *14* (4), 1−27.
(40) Mercer, L. D.; Szpiro, A. A.; Sheppard, L.; Lindström, J.; Adar, S. D.; Allen, R. W.; Avol, E. L.; Oron, A. P.; Larson, T.; Liu, L. J. S.; Kaufman, J. D. Comparing Universal Kriging and Land-Use Regression for Predicting Concentrations of Gaseous Oxides of Nitrogen (NOx) for the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Atmos. Environ.* **2011**, *45* (26), 4412−4420.
(41) Hankey, S.; Sforza, P.; Pierson, M. Using Mobile Monitoring to Develop Hourly Empirical Models of Particulate Air Pollution in a Rural Appalachian Community. *Environ. Sci. Technol.* **2019**, *53*, 4305−4315.
(42) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(43) Young, M. T.; Bechle, M. J.; Sampson, P. D.; Szpiro, A. A.; Marshall, J. D.; Sheppard, L.; Kaufman, J. D. Satellite-Based NO2 and Model Validation in a National Prediction Model Based on Universal Kriging and Land-Use Regression. *Environ. Sci. Technol.* **2016**, *50* (7), 3686−3694.
(44) US Census Bureau. Census Regions and Divisions of the United States https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf [December 8, 2020].
(45) US Census Bureau. Census Urban Area National https://catalog.data.gov/dataset/tiger-line-shapefile-2017-2010-nation-u-s-2010-census-urban-area-national [March 7, 2020].
(46) US Environmental Protection Agency. EJSCREEN: Environmental Justice Screening and Mapping Tool https://www.epa.gov/ejscreen/download-ejscreen-data [6 August 2021].
(47) Patton, A. P.; Collins, C.; Naumova, E. N.; Zamore, W.; Brugge, D.; Durant, J. L. An Hourly Regression Model for Ultrafine Particles in a Near- Highway Urban Area. *Environ. Sci. Technol.* **2014**, *48*, 3272−3280.
(48) Michanowicz, D. R.; Shmool, J. L. C.; Tunno, B. J.; Tripathy, S.; Gillooly, S.; Kinnee, E.; Clougherty, J. E. A Hybrid Land Use regression/AERMOD Model for Predicting Intra-Urban Variation in PM2.5. *Atmos. Environ.* **2016**, *131*, 307−315.
(49) Hankey, S.; Marshall, J. D. On-Bicycle Exposure to Particulate Air Pollution: Particle Number, Black Carbon, PM2.5, and Particle Size. *Atmos. Environ.* **2015**, *122*, 65−73.
(50) Bechle, M. J.; Millet, D. B.; Marshall, J. D. National Spatiotemporal Exposure Surface for NO2: Monthly Scaling of a Satellite-Derived Land-Use Regression, 2000 − 2010. *Environ. Sci. Technol.* **2015**, No. 2.
(51) de Hoogh, K.; Gulliver, J.; van Donkelaar, A.; Martin, R. V.; Marshall, J. D.; Bechle, M. J.; Cesaroni, G.; Pradas, M. C.; Dedele, A.; Eeftens, M.; Forsberg, B.; Galassi, C.; Heinrich, J.; Hoffmann, B.; Jacquemin, B.; Katsouyanni, K.; Korek, M.; Künzli, N.; Lindley, S. J.; Lepeule, J.; Meleux, F.; de Nazelle, A.; Nieuwenhuijsen, M.; Nystad, W.; Raaschou-Nielsen, O.; Peters, A.; Peuch, V.-H.; Rouil, L.; Udvardy, O.; Slama, R.; Stempfelet, M.; Stephanou, E. G.; Tsai, M. Y.; Yli-Tuomi, T.; Weinmayr, G.; Brunekreef, B.; Vienneau, D.; Hoek, G. Development of West-European PM2.5 and NO2 Land Use Regression Models Incorporating Satellite-Derived and Chemical Transport Modelling Data. *Environ. Res.* **2016**, *151* (2), 1−10.
(52) Hochadel, M.; Heinrich, J.; Gehring, U.; Morgenstern, V.; Kuhlbusch, T.; Link, E.; Wichmann, H. E.; Krämer, U. Predicting Long-Term Average Concentrations of Traffic-Related Air Pollutants Using GIS-Based Information. *Atmos. Environ.* **2006**, *40* (3), 542−553.
(53) Wu, C. Da; Chen, Y. C.; Pan, W. C.; Zeng, Y. T.; Chen, M. J.; Guo, Y. L.; Lung, S. C. C. Land-Use Regression with Long-Term Satellite-Based Greenness Index and Culture-Specific Sources to Model PM2.5 Spatial-Temporal Variability. *Environ. Pollut.* **2017**, *224*, 148−157.
(54) Brousse, O.; Georganos, S.; Demuzere, M.; Dujardin, S.; Lennert, M.; Linard, C.; Snow, R. W.; Thiery, W. Can We Use Local Climate Zones for Predicting Malaria Prevalence across Sub-Saharan African Cities ? OPEN ACCESS Can We Use Local Climate Zones for Predicting Malaria Prevalence across Sub-Saharan African Cities ? *Environ. Res. Lett.* **2020**, *15* (12), 124051.
(55) Weissert, L. F.; Salmond, J. A.; Miskell, G.; Alavi-shoshtari, M.; Williams, D. E. Development of a Microscale Land Use Regression Model for Predicting NO2 Concentrations at a Heavy Trafficked Suburban Area in Auckland, NZ. *Sci. Total Environ.* **2018**, *619−620*, 112−119.
(56) Weichenthal, S.; Ryswyk, K. Van; Goldstein, A.; Bagg, S.; Shekkarizfard, M.; Hatzopoulou, M. A Land Use Regression Model for Ambient Ultrafine Particles in Montreal, Canada: A Comparison of Linear Regression and a Machine Learning Approach. *Environ. Res.* **2016**, *146*, 65−72.
(57) Zhan, Y.; Luo, Y.; Deng, X.; Zhang, K.; Zhang, M.; Grieneisen, M. L.; Di, B. Satellite-Based Estimates of Daily NO2 Exposure in China Using Hybrid Random Forest and Spatiotemporal Kriging Model. *Environ. Sci. Technol.* **2018**, *52* (7), 4180−4189.
(58) Li, X.; Zhang, C.; Li, W.; Ricard, R.; Meng, Q.; Zhang, W. Assessing Street-Level Urban Greenery Using Google Street View and a Modified Green View Index. *Urban For. Urban Green.* **2015**, *14* (3), 675−685.

(59) Guo, Y.; Liu, Y.; Oerlemans, A.; Lao, S.; Wu, S.; Lew, M. S. Deep Learning for Visual Understanding: A Review. *Neurocomputing* **2016**, *187*, 27−48.
(60) Jiang, S.; Alves, A.; Rodrigues, F.; Ferreira, J.; Pereira, F. C. Mining Point-of-Interest Data from Social Networks for Urban Land Use Classification and Disaggregation. *Comput. Environ. Urban Syst.* **2015**, *53*, 36−46.
(61) Lefèvre, S.; Tuia, D.; Wegner, J. D.; Produit, T.; Nassar, A. S. Toward Seamless Multiview Scene Analysis from Satellite to Street Level. *Proc. IEEE* **2017**, *105* (10), 1884−1899.
(62) 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.
(63) Thompson, J. E. Crowd-Sourced Air Quality Studies: A Review of the Literature & Portable Sensors. *Trends Environ. Anal. Chem.* **2016**, *11*, 23−34.
(64) Morawska, L.; Thai, P. K.; Liu, X.; Asumadu-sakyi, A.; Ayoko, G.; Bartonova, A.; Bedini, A.; Chai, F.; Christensen, B.; Dunbabin, M.; Gao, J.; Hagler, G. S. W.; Jayaratne, R.; Kumar, P.; Lau, A. K. H.; Louie, P. K.K; Mazaheri, M.; Ning, Z.; Motta, N.; Mullins, B.; Rahman, M. M.; Ristovski, Z.; Shafiei, M.; Tiondronegoro, D.; Westerdahl, D.; Williams, R. Applications of Low-Cost Sensing Technologies for Air Quality Monitoring and Exposure Assessment: How Far Have They Gone? *Environ. Int.* **2018**, *116* (April), 286−299.

https://doi.org/10.1021/acs.est.1c04047
Environ. Sci. Technol. 2021, 55, 15519−15530