

Mapping Air Pollution with Google Street View Cars: Efficient Approaches with Mobile Monitoring and Land Use Regression

Kyle P. Messier,^{†,‡} Sarah E. Chambliss,[†] Shahzad Gani,[†] Ramon Alvarez,[‡] Michael Brauer,[§] Jonathan J. Choi,[‡] Steven P. Hamburg,[‡] Jules Kerckhoffs,^{||} Brian LaFranchi,[⊥] Melissa M. Lunden,[⊥] Julian D. Marshall,[#] Christopher J. Portier,[‡] Ananya Roy,[‡] Adam A. Szpiro,[○] Roel C. H. Vermeulen,^{||} and Joshua S. Apte^{*,†}

[†]Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, Austin, Texas 78712, United States

[‡]Environmental Defense Fund, New York, New York 10010, United States

[§]School of Population and Public Health, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada

^{||}Institute for Risk Assessment Science, Utrecht University, Utrecht 3584 CM, Netherlands

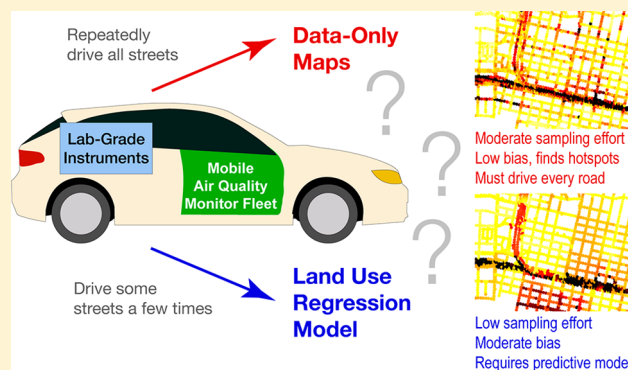
[⊥]Aclima, Inc., 10 Lombard Street, San Francisco, California 94111, United States

[#]Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington 98195, United States

[○]Department of Biostatistics, University of Washington, Seattle, Washington 98195, United States

Supporting Information

ABSTRACT: Air pollution measurements collected through systematic mobile monitoring campaigns can provide outdoor concentration data at high spatial resolution. We explore approaches to minimize data requirements for mapping a city's air quality using mobile monitors with "data-only" versus predictive modeling approaches. We equipped two Google Street View cars with 1-Hz instruments to collect nitric oxide (NO) and black carbon (BC) measurements in Oakland, CA. We explore two strategies for efficiently mapping spatial air quality patterns through Monte Carlo analyses. First, we explore a "data-only" approach where we attempt to minimize the number of repeated visits needed to reliably estimate concentrations for all roads. Second, we combine our data with a land use regression-kriging (LUR-K) model to predict at unobserved locations; here, measurements from only a subset of roads or repeat visits are considered. Although LUR-K models did not capture the full variability of on-road concentrations, models trained with minimal data consistently captured important covariates and general spatial air pollution trends, with cross-validation R^2 for log-transformed NO and BC of 0.65 and 0.43. Data-only mapping performed poorly with few (1–2) repeated drives but obtained better cross-validation R^2 than the LUR-K approach within 4 to 8 repeated drive days per road segment.



1. INTRODUCTION

Exposure to air pollution is a major risk factor for adverse health effects and premature death.^{1–6} Urban air pollutant concentrations can vary sharply over short spatial scales,^{7–9} with important consequences for population exposures and environmental health.¹⁰ In the context of epidemiological studies, highly spatially resolved exposure estimates can reduce exposure measurement error¹¹ and enable researchers to study populations within small study areas.¹² Other applications of high-resolution air quality maps may include assessment of environmental equity, studies of pollutant sources and dynamics, and the basis for air quality management and public awareness.

Numerous approaches exist for quantifying intraurban variation in air pollutant concentrations.^{9,13} Land use regression

(LUR) has been a popular approach¹⁴ because of its simplicity, interpretability, and ability to predict fine-scale variations in pollution. LUR models are traditionally developed with central-site air quality monitors^{15–18} or networks of fixed-location air quality instruments designed specifically for creating a LUR within a given city or study area.^{19–24} More recently mobile monitoring campaigns, including both the approaches of collecting air quality data in a vehicle while in motion^{25–36} and short-term stationary measurements (e.g., 30 min) at

Received: June 20, 2018

Revised: October 2, 2018

Accepted: October 3, 2018

Published: October 24, 2018

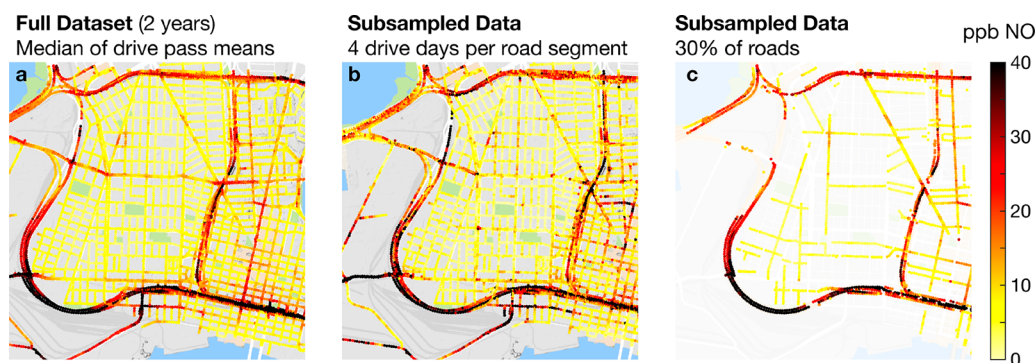


Figure 1. (a) Median of drive pass mean NO concentrations for the full 2-year data set. This data set is used to validate model predictions. (b) Representative example of subsampled training data set with 4 days of driving at each road segment. This Monte Carlo subsampled data set results in the LUR-K median predictive performance out of the 100 subsampled data sets with 4 drive days. (c) Representative example (also median predictive performance) of a subsampled training data set incorporating 30% of the roads in the domain. Map data © 2018 Google.

predetermined stops,^{29,31} have been used to successfully develop LUR models. A strength of mobile air quality monitoring is its ability to increase the spatial resolution of measurements to a substreet level, however at the expense of temporal sparsity at any given location. As mobile monitoring is increasingly contemplated as an approach for developing high-resolution air quality maps over large spatial domains, a better understanding of the data requirements to efficiently produce reliable estimates is needed. Here, we systematically assess how key study design parameters such as repetition frequency and road coverage affect the performance of alternative mobile monitoring study designs. By exploring the trade-off between sampling effort and mapping fidelity, our results illuminate efficient approaches for large-scale mobile data collection.

The goal of this study is to systematically assess the data requirements for accurately and efficiently mapping long-term average air quality at a high-spatial resolution using mobile monitoring. Using an unusually rich data set of repeated mobile air quality measurements collected with specially equipped Google Street View cars, we explore and evaluate two alternative strategies for efficiently mapping spatial air quality patterns. First, following Apte et al. (2017),⁷ we use a “data-only” approach where we map concentrations solely on the basis of repeated observations while attempting to minimize the number of repeated visits to each road. Second, we combine our data with a land use regression-Kriging (LUR-K) model to predict at unobserved locations and consider sampling schemes where only a subset of a city’s roads or repeat visits are measured. The LUR-K approach can make spatial predictions at a large number of locations even with a small number of repeated measurements on a subset of city roads. Potential advantages of the model-free “data-only” approach include avoiding the extra effort of training and evaluating predictive models, avoiding potential biases that may result from the model structure, and avoiding assumptions regarding the relationship between air pollution and land-use. To evaluate these approaches, we use a Monte Carlo scheme to systematically subsample the full data set and then develop data-only maps or LUR-K models and compare the reduced-data results to our full data set of long-term maps.

2. MATERIAL AND METHODS

2.1. Air Quality Data Collection and Processing.

We used two Google Street View cars equipped with the Aclima mobile platform (Aclima Inc., San Francisco, CA) to measure

air quality on city streets in Oakland, CA on 326 days from May 28, 2015 to May 19, 2017. This data set is based on the instrumentation platform and mobile sampling approach described in detail in Apte et al. (2017).⁷ We consider two species that are markers of traffic air pollution: nitric oxide (NO, measured by chemiluminescence) and black carbon (BC, measured by photoacoustic spectroscopy). Further details on the measurement platform and summary driving statistics are available elsewhere⁷ and as [Supporting Information \(SI\)](#). We repeatedly measured air quality during weekday daytime hours on every road in a 30 km² domain in Oakland, CA. Described in detail in the [SI](#), this domain incorporates residential, commercial, and industrial areas. We display results for two neighborhoods to emphasize the high spatial resolution data, West Oakland and Downtown, with additional results for East Oakland presented as [SI](#). Overall, we collected approximately 3.5 million (NO) and 3.7 million (BC) 1-Hz observations. This data set is unique for nearly complete coverage of all city streets within the domain and for the many repeated measurements at each location (minimum and median of 10 and 41 days of repetition at each location, see [Figure S1](#)).

Our data processing steps followed those of Apte et al.⁷ We first applied a multiplicative time-of-day factor based on central-site monitoring to adjust for diurnal variation in ambient air quality, which had only a minor ($\pm 10\%$) effect on long-term average spatial patterns.⁷ We then divided the ~ 640 km of roads in our domain into 19,149 road segments, each 30 m in length, and then “snapped” each pollution observation to the nearest segment. To ensure that each repeated drive through a given road segment (drive pass), which had varying numbers of highly correlated 1-Hz measurements, was represented equally in our analysis, we updated our data reduction scheme here as follows. First, we reduced the measurements for each “drive pass” through a 30-m road segment (typically ~ 3 – 10 s) into a single drive pass mean concentration. We then computed the median of repeated drive pass mean concentrations as our core metric for analysis, as shown [Figure 1a](#) for NO (see [Figure S2](#) for BC). Because this “median of drive pass means” approach incorporates information from numerous repeated drive passes, it is robust to anomalously or idiosyncratically polluted drive passes, and it produces a concentration map that is highly correlated ($R^2 > 0.9$) with the data reduction approach used in Apte et al.⁷

2.2. Monte Carlo Simulation of Mobile Monitoring Scenarios.

We outline here our approach to evaluating

alternative schemes for developing high-resolution air quality maps with mobile monitoring. We consider high spatial resolution as the ability to precisely distinguish between and within street-to-street variability. As a starting point, we take our full 2-year data set of average (i.e., median of drive pass means) concentrations on every 30-m road segment to be our reference data set, representing our best estimate of the true spatial variability of weekday daytime concentrations (Figure 1a). The level of monitoring effort (~1,400 h, 40,000 km of driving) required to generate this data set could be difficult to replicate elsewhere. We then evaluate the following suite of alternative mapping approaches that could potentially reduce the data collection effort:

- **Data-Only Map:** Starting with the full 2 years of observations, we develop a subsampled data set with N driving days at each 30-m road segment from the full 2 years of observations. We estimate the long-term concentrations at each 30-m road segment as the median of drive pass means for this subsample.
- **LUR-Kriging Model:** We train LUR-Kriging (LUR-K) models using a subset of the full 2-year data set. We consider three alternative approaches to subsampling data to train LUR-K models, described briefly here and in detail below. First, we consider a “drive day” sampling scheme: mobile monitors collect N days of data for all 30-m road segments in the domain, and then a LUR-K prediction for all road segments is trained on this temporal subsample of measurements. Second, we consider a “road coverage” sampling scheme, where all 2 years of data for only a portion of the roads in the sampling domain are included for training a LUR-K model. Third, we consider “joint” scenarios in which LUR-K predictions are developed on the basis of a subsampled data set where a limited number of repeated observations are collected on a limited number of roads. In each case, the prediction performance of LUR-K models is evaluated relative to our long-term reference data set using a 10-fold cross-validation scheme described in Section 2.5.

We employed the following Monte Carlo (MC) subsampling schemes to implement each of the above scenarios in 100 independent draws.

2.2.1. Sampling by Number of Days of Driving Per Road Segment. We develop subsampled data sets where every road segment is sampled on N unique days. Here, we summarize the process with a more detailed description in the SI. First, we developed 16 scenarios where we randomly selected without replacement $N = \{1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45\}$ days with valid measurements within the Oakland sampling domain from our full set of 2 years of repeated observations, preserving at least 95% of all road segments in the domain to ensure our domain does not change substantially from one subsample to the next. Our sampling routes differed daily, and the domain took several days to fully drive, thus the selection procedure resulted in fewer than the target N days of repeated measurement at each individual 30-m road segment. On average, ~34 unique calendar days of sampling were required to construct a map with 1 complete drive day for the full domain, and ~190 unique calendar days of driving were needed to provide 10 days of measurements everywhere (see Figures S3 and S4). The number of calendar days required to completely sample our domain appears not to meaningfully affect our findings (see sensitivity analysis in

Figure S5). For large values of N , the desired number of days occasionally exceeds the number of repeated observations for a road segment, resulting in the full data set being sampled (~75% of road segments have ≥ 25 days of observations; see the SI).

We then computed the median-of-drive-pass-mean concentrations for every road segment subsample (see example map in Figure 1b). We applied these reduced data sets to the two schemes described above: data-only (wherein a map is made simply by reducing repeated measurements) and LUR-K (as training inputs to a subsequent LUR-K modeling process).

2.2.2. Sampling by Road Coverage. We developed scenarios to simulate a mapping approach where predictive models are trained to estimate concentrations for an entire domain on the basis of measurements collected on a small subset of roads. As a starting point, we decided for two reasons that subsampled maps would always include the full highway network. First, given the low overall contribution of highways to the total road length in our sample, we believed that further restricting the number of highways in a subsample would lead to unstable predictions at the high end of our concentration range. Second, these roads were routinely driven here out of operational necessity, as would likely be the case many mobile sampling efforts in other urban settings. We then randomly selected without replacement nonhighway roads to construct maps where between 10% and 90% of the nonhighway roads were included, increasing in 10% increments. Subsampling was performed by street name to ensure contiguous, connected road segments. For a scenario targeting $M\%$ road coverage, we used an iterative approach to repeatedly and randomly select road segments by street name until a map containing $M \pm 1\%$ of the road network was completed. Note that the subsampling by road coverage occurs in the training set for each iteration in the 10-fold cross-validation process. The test set is always the median of drive pass means based on the full 2-year data set. For each road segment, we used the median of drive pass means from the full 2-year data set as the basis for model training.

2.2.3. Sampling Jointly by Road Coverage and Drive Days. We simulated a mobile measurement approach where a fraction of roads in the measurement domain is sampled on a restricted number of repeated occasions. Here, we combined elements of the algorithms devised in the two preceding sections. We first constructed an air quality map via subsampling for the complete measurement domain on a restricted number of days. Then, using the road subsampling algorithm described in section 2.2.2, we randomly selected a subset of all roads in the domain for inclusion in model training. We considered scenarios with combinations of 1, 4, and 14 days of repetition per road segment on $\{10, 20, \dots, 80, 90\}$ percent of the roads in the modeling domain.

2.3. Model Development: Geographic Covariates. Geographic covariates were constructed prior to model development. The candidate set of variables is summarized below, with details available in the SI (Table S2). Briefly, the 121 variable candidate set of covariates included binary road classifications, binary local truck routes, local zoning classifications, normalized difference vegetative index, percent landcover, road length, population density, and continuous point source variables such as National Priority Listing sites, airports, and ports (see the SI). Continuous variables had a distance hyperparameter such as exponential decay distance³⁷ or buffer size, with a minimum buffer size of 50 m. Owing to the limited availability of fine-scale predictor data sets, our set of candidate

model covariates most strongly represents features at scales of ≥ 50 m.

2.4. Land Use Regression with Kriging (LUR-K). We used a LUR-K approach to model the observed 30-m median of drive-pass mean pollutant concentrations. Prior to model development the distributions for NO and BC were examined graphically for normality. Statistical analyses used log-transformed data for NO and untransformed data for BC. LUR-Kriging models were selected following a similar approach developed for the European Study of Cohorts for Air Pollution Effects (ESCAPE) studies.³⁸ Briefly, an ordinary least-squares (OLS) LUR was fit using a modified stepwise procedure. Variables were added based on an increase in model R^2 , variables were required to be statistically significant to enter the model, variables were constrained to a priori assumption of physical interpretations (i.e., sources are expected to increase pollution therefore their coefficients are positive), and variance inflation (VIF) was maintained below 3.

Once the geographic covariates are selected and the model is trained, the LUR was integrated into a Kriging framework to produce a LUR-Kriging model, sometimes known as Kriging with an external drift.^{39–41} LUR-Kriging combines an ordinary linear LUR plus a spatially explicit error term that accounts for the spatial autocorrelation between points. Here, the error covariance was modeled as an exponential model with a nugget effect.⁴² The full set of covariance model parameters and explanatory model coefficients was estimated using an iterative algorithm in which the covariance was estimated from the LUR residuals followed by a generalized least-squares estimate of all parameters with the updated covariance matrix.⁴³

2.5. Performance Evaluation. For the model-based approaches in the subsampling analyses, we implemented K -fold cross-validation to assess the ability of LUR-K models to select consistent, parsimonious models and make independent, out-of-sample predictions of our long-term average measurements. Using an algorithm described in the SI, we defined $K = 10$ contiguous, similarly sized spatially clustered cross-validation groups¹⁶ to minimize the spatial autocorrelation of near-neighbors in the cross-validation (Figure S6). Owing to the high spatial density of mobile monitoring samples, this cluster approach to cross-validation reduces the effect of the extremely close neighbors and more rigorously approximates out-of-sample prediction performance. In 10-fold cross-validation, the subsampled road segments selected for model training are divided into $K = 10$ spatially clustered folds. We then cycle through the 10 possible permutations of $K-1 = 9$ folds, each time training a LUR-K model on 9 out of the 10 folds, while reserving data from the 10th fold for independent model evaluation.

For each set of 10 folds, we apply the fitted LUR-K model to make predictions in the single held-out spatial cluster. In conventional LUR modeling practice, model predictions would be compared against the data withheld from the training data set for each of the K folds. In contrast, our core analyses of model performance compared model predictions for each road segment within the held-out cluster with the long-term median-of-drive-pass-mean concentrations at those locations. Whereas the former analysis approach provides information on how well the model reproduces the training data set, the latter approach summarizes how well the model predicts long-term average concentrations.

For each set of 10 folds, we summarize model performance in terms of two statistics: R^2 (i.e., squared Pearson correlation

coefficient) and the normalized root mean squared error (NRMSE), root mean squared error normalized to the observed mean). The R^2 and NRMSE for the spatially clustered K -folds were calculated as follows:

$$R^2 = R^2(Z, Z^*) \quad (1)$$

$$\text{NRMSE} = \frac{1}{K} \sum_{k=1}^K \frac{\sqrt{\frac{1}{n_k} \sum_{i=1}^{n_k} (Z_i - Z_i^*)^2}}{\frac{K}{n_k} \sum_{i=1}^{n_k} Z_i}, \quad K = 10 \quad (2)$$

Here, Z_i are the observed medians of drive-pass means based on the entire 2-year data set at K -fold i , Z_i^* are the predicted medians of drive-pass means at K -fold i , and n_k is the sample size for the k -fold. For NO, R^2 is based on the log-transformed NO data, and the NRMSE is calculated in the normal/untransformed space. The best value of NRMSE is 0, while favorable values of NRMSE are considered to be less than 1, which indicate the mean error is less than the sample mean.

10-fold cross-validation is not feasible in data-only approaches since there is no basis for prediction at unsampled locations. We assessed model performance for the data-only approach by comparing the simulated median of drive-pass means based on a given MC iteration to the median of drive-pass means for the full 2-year data set.

For each scenario, we repeatedly developed models on 100 independent MC subsamples of the full data set. For each subsample, we evaluate R^2 and NRMSE and then compute the distribution of these performance metrics for repeated subsamples.

3. RESULTS AND DISCUSSION

In this study, we examine strategies to efficiently develop air quality maps from mobile monitoring data, either via a “data-only” scheme that averages repeated measurements or via LUR-K models trained on repeated measurements. As the following sections discuss in detail, two key results of our analysis are that (i) robust LUR-K models can be effectively developed even with very sparse mobile monitoring data, but (ii) the data-only approach outperforms LUR-K in precision (R^2) after a small number of drive days (cf. Figure 2a-b).

Figure 2 illustrates representative results and residuals for these two approaches (left column: maps of daytime NO, right column: residuals in comparison to Figure 1a). Visual inspection suggests that each approach recreates some key features of the long-term observed concentrations. NO concentrations are elevated strongly on highways (and modestly on arterials) relative to residential streets. Elevated NO levels in Downtown Oakland are evident in each of the maps. However, the full data set (Figure 1a) contains numerous localized pollution hotspots at road intersections, industries, and other emissions sources, only some of which are reproduced in the Figure 2 maps (compare Figures S2 and S7 for BC). Figure S9 illustrates similar results for the East Oakland domain.

Figure 2a depicts an example data-only map for NO concentrations in our Oakland domain (excluding East Oakland), which incorporates 4 days of sampling at each location (85 total hours of measurement, approximately 6% of the full measurement data set). Even with only 4 days of measurement data per road segment, the key spatial patterns of NO are evident (median $R^2 = 0.65$, median NRMSE = 1.18), including pollution hotspots near some industries and intersections. However, there is evident “noise” – errors that appear to be

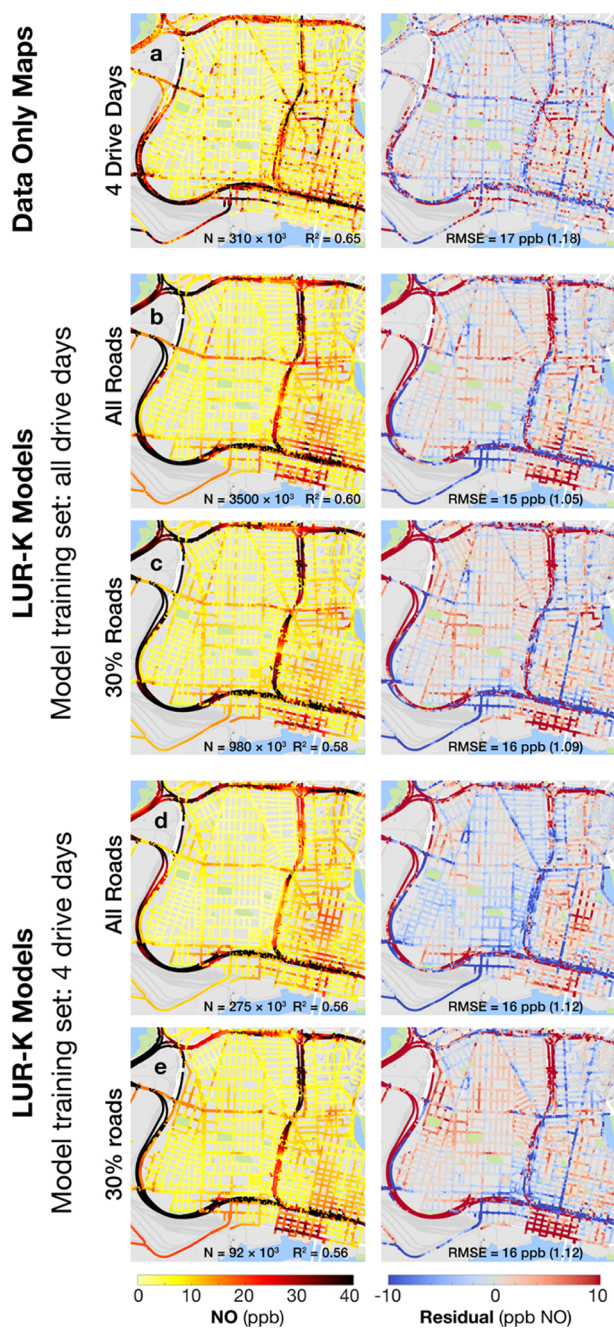


Figure 2. Example air quality maps constructed using sampling and/or modeling approaches. (a) Data only map drawn from Monte Carlo subsample with 4 days at each road segment. Residuals are computed as the difference between subsampled 4 day map and the long-term concentrations shown in Figure 1a. (b) 10-fold cross-validation LUR-K prediction (and residual) surface trained on all the road segments and the entire 2-year data set. (c) 10-fold cross-validation LUR-K prediction (and residual) surface trained on the full 2-year data set and 30% subsample of road segments. Note the similarity in predictions between (b) and (c). (d) 10-fold cross-validation LUR-K predictions trained on a 4-day subsample for the full domain of road segments. (e) 10-fold cross-validation LUR-K predictions trained on a 4-day subsample and 30% of the road segments. The number of training data is given as N and span a range from 92,000 independent points (~ 25 h of sampling to cover 30% of roads in domain 4 times each) up to the size of the full data set. For all panels, the R^2 is based on the log-transformed NO data, and the RMSE is calculated in untransformed space. Normalized RMSE values are provided in parentheses. Map data © 2018 Google.

approximately randomly distributed in space – that arises because of the limited number of samples at each location.

Figures 2b–e illustrate four alternative approaches to training a LUR-K model to predict concentrations at every 30-m road segment. Figure 2b represents a scenario where the LUR-K model incorporates the full 2 years of measurement data for training and is described in greater detail in the following section. By incorporating the full data set, this model achieves the best performance ($R^2 = 0.60$, NRMSE = 1.05) of all LUR-K models in our analysis. Of course, it would generally be both impractical and illogical to collect extensive repeated measurements at every location for which one wishes to make model predictions. Figures 2c–e illustrate that the LUR model performance remains similar even when the amount of model training data is substantially restricted. In Figure 2c, the training data set is restricted to a subset of roads accounting for all highways and a random set of 30% of the nonhighway road network (20% of the full data set hours), resulting in only a negligible change in model predictions and performance (median $R^2 = 0.58$, median NRMSE = 1.09). In Figure 2d, the training data set is restricted to only 4 days of observation, resulting in a different model with a slight decrement in performance (median $R^2 = 0.56$, median NRMSE = 1.12), but with a large drop in training data requirements (6% of full data set; ~ 80 h). Figure 2e illustrates an example model trained on a highly restricted data set (30% road coverage, 4 days of observation) with a dramatic reduction in data requirements (2% of full data set; ~ 25 h), also accompanied by a slight reduction in model performance (median $R^2 = 0.56$, median NRMSE = 1.12).

3.1. Full Data Set Model. LUR-K predictions incorporating the full 2 years of driving for NO are shown in Figure 2b. The predictions capture regional and local variability but often fail to correctly predict fine-scale hotspots. GIS covariates selected in each of the 10-folds included road type indicators (highway roads, residential roads, etc.), local truck route indicator, NDVI within a 50 m buffer, distance to the port, and elevation. Table 1 shows covariates selected for log-NO and

Table 1. Log-NO Models for the 10-Fold Cross-Validation of the Full 2-Year Data Set^a

variable name	units	average coefficient (min, max)	no. out of 10-folds
highway roads	binary	0.90 (0.73, 1.78)	10
major roads	binary	0.66(–)	1
residential roads	binary	–0.70 (–0.71, –0.57)	9
local trucks	binary	0.49 (0.40, 0.58)	10
NDVI 50 m	unitless	–4.2 (–4.7, –2.4)	7
elevation 1000 m	meters	0.007 (0.006, 0.007)	6
port 5000 m	unitless	0.10 (0.075, 0.15)	3
developed high 250 m	percent	0.008 (0.008, 0.008)	2
develop low 2500 m	percent	0.05 (–)	1
inverse distance to airport	meters ^{–1}	3.3 (–)	1

^aColumns include the variable name and buffer size if applicable, variable units, the average coefficient over 10-folds, and the number of folds it appeared in.

BC models including the average (min, max) coefficients across each fold and the number of occurrences. The LUR-K 10-fold cross-validation has a NRMSE of 1.05 and R^2 of 0.60.

Black carbon LUR-K predictions based on the full 2 years of monitoring are available in the [Supporting Information](#)

(Figure S7). Observed BC has a similar spatial pattern as NO, with an R^2 between the full data set NO and BC LUR-K predictions of 0.73. Fine-scale variability in observed BC is more pronounced owing to 1 Hz noise in the raw data from the photoacoustic BC instrument. The final LUR-K BC model resulted in a 10-fold cross-validation R^2 of 0.43 and NRMSE of 0.60 (Table S1).

3.2. Monte Carlo Subsampling Analysis. **3.2.1. Reduced Drive Days for LUR-K Models.** Figure 2 shows example maps and residuals for reduced data approaches. The residuals of the data-only approach appear more randomly distributed than the LUR-K model residuals. We observe similar predictions and residuals for the LUR-K models in Figure 2b-c, an illustration of a key result: the amount of road coverage used for model training does not strongly affect modeled predictions. The LUR-K models based on 4 days of training data at all road segments (Figure 2d, all roads and Figure 2e, 30% of roads) produce comparable predictions to the full data set models, although with a slightly larger error and lower R^2 .

Figure 3 shows the R^2 and NRMSE for the MC analyses as a function of reduced drive days and road coverage for log-NO

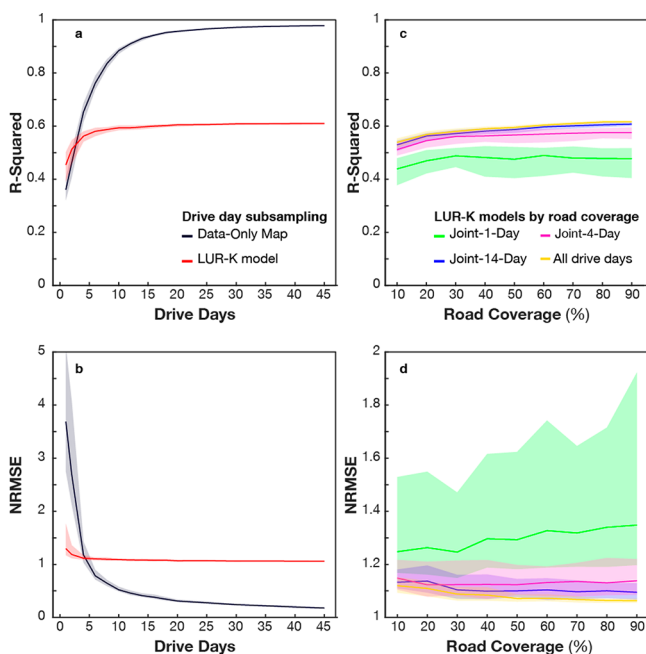


Figure 3. Performance evaluation for subsampled maps in terms of R^2 (a,c) and NRMSE (b,d). (a,b) Subsampling schemes based on the number of drive days per road segment. The black trace indicates the data-only mapping scheme, and the red trace indicates a LUR-K model trained with an equal number of measurements days on all road segments in the domain. (c,d) LUR-K models trained on a specified fraction of road segments on the domain for four alternative levels of repetition frequency on each road segment. “Joint” models refer to cases with jointly reduced road coverage and drive days per road segment. Shaded area and solid lines represent, respectively, the interquartile range and median model performance for 100 Monte Carlo permutations of our full data set.

(see Figure S8 for scatter plots and S10 for BC). Figure 3a-b shows the direct comparison of a LUR-K approach versus a data-only approach as drive days are reduced. The R^2 of the LUR-K models for log-NO was remarkably consistent as drive days were reduced toward a single drive day. For example, the median R^2 was above 0.50 for cases with four or more drives

per road segment. Even with 2 drive days per road segment, the median R^2 (0.52) was within 15% of models developed on 45+ drive days. The prediction bias (NRMSE) for log-NO was also generally consistent as the number of drive days per road segment was reduced, although a small increase in bias was observed below 6 days. In general, BC models followed similar trends, although the upper-bound prediction performance for BC was somewhat inferior to NO. Results were stable among repeated MC iterations, as evidenced by the small interquartile range in predictions. Overall, these results suggest that the sampling error introduced by a low repetition frequency at each road segment does not strongly degrade the ability of a LUR model to make out-of-sample predictions of long-term average concentrations. There are many training road segments representing each part of our prediction variable space, which perhaps allows for an “averaging out” of the sampling error or temporal variability at each location.

3.2.2. Reduced Drive Days for Data-Only Maps. We find that data-only mapping is extremely variable and poorly replicates long-term concentrations with ≤ 2 drive days per road segment, but each subsequent repeat visit results in significant improvements of prediction accuracy and bias. With only a small number of drive days (log-NO: 4–6 days, BC: 6–8 days), the data-only approach consistently outperforms the best LUR-K models, with $R^2 > 0.7$. Moreover, the data-only approach continues to improve until about 25 drive days and approaches the full 2-year data set results ($R^2 > 0.9$), whereas the LUR-K models rapidly approach a ceiling of moderate performance ($R^2 \sim 0.6$ for log-NO).

3.2.3. Reduced Road Coverage for LUR-K Models. A key efficiency advantage of predictive modeling approaches is the ability to make wide-area exposure estimates on the basis of spatially limited training measurements. Figures 3c-d show the log-NO R^2 and NRMSE for the MC analyses as a function of reduced road coverage. An important finding here is that LUR-K prediction performance is only weakly dependent on the amount of road coverage, especially for $>10\%$ road coverage. For example, NO models trained on a random 10% and 20% sample of the roads in our domain performed nearly as well ($R^2 = 0.51, 0.56$) as a model trained on 90% of the domain roads ($R^2 = 0.61$), despite the nearly order-of-magnitude difference in the size of the training data set. Our assessment of the joint subsampling of road coverage and repetition frequency suggests that there is little interaction between these two variables. In general, the LUR-K models based on 1-drive day performed worse than the other LUR-K models (mean $R^2 = 0.43$ – 0.47 for 10–90% road coverage), while our road coverage subsampling for the 4, 14, and all-day models had similar results.

Our results were largely consistent with Hatzopoulou et al.⁴⁴ and Minet et al.⁴⁵ in which they found a consistent mean with increasing variability in R^2 as the number of road segments in the domain decreased. This finding also implies that the common practice of training mobile monitoring LUR models on a subset of a domain’s roads does not impose a meaningful prediction bias. An important qualification is that for out-of-sample predictions to be successful, the training data set must be representative of the full range of conditions for the prediction domain. To test this point, we conducted a sensitivity analysis where we restricted the road subsampling to include only arterial roads and highways and found substantially diminished model performance ($R^2 \sim 0.35$ – 0.5).

3.3. Consistency of Results. Correlations between the reduced data set LUR-K predictions and the full model were

generally high ($R^2 > 0.8$ – 0.9). To explain why, we examined the sensitivity of the LUR-K model structure (i.e., the selected covariates) to reducing the size of training data sets. We found that variables selected for the full 2-year data set model were often selected in the cases with reduced data sets. In other instances, highly correlated explanatory variables were selected instead (e.g., mutually exclusive road classes), resulting in similar predictions as the full model variables. We observed that GIS covariates explaining the largest proportion of variance were selected frequently and tended to be ubiquitous city-scale features such as road class. In contrast, fine-scale features (e.g., NDVI within 50 m) were less reliably selected as the number of the samples decreased. Much of the unexplained variance in our best-case model appears attributable to very fine-scale spatial variation (e.g., local hotspots) for which good predictor variables are unavailable. We speculate that if future models were to include more effective covariates for predicting the residual fine-scale variation, LUR-K results might become more sensitive to the amount of training data even as the best-case performance of a model improves.

A unique feature of our analyses is that our subsampled training data sets (e.g., N days per road segment) tend to have relatively imprecise measurements at each location (owing to temporal variability across drive days), while the 2-year median concentrations that are used as a test set are far more precise by virtue of the large number of repeated samples. In most real-world cases, this type of test set is not available, and K -fold cross-validation might instead be used to allow measurements to serve as both training and test data. In a sensitivity analysis we considered how our drive-day models would fare under such an evaluation scheme. An intriguing result arises for cases with a small number of drive days per road segment. Because concentrations at each road segment are “noisy” (error from a small sample size), our LUR-K models do not perform well at predicting their own noisy training data sets, even though they perform quite well in predicting long-term concentrations. This result points to the interesting possibility that a model trained on a small sample adequately predicts the (unknown) true spatial pattern even as it struggles to predict its own training data set. Kerckhoffs et al. found a similar result in The Netherlands: mobile monitoring LUR predictions based on short-term monitoring had poor training R^2 but predicted long-term, independent test set observations with good fidelity.³¹

3.4. Implications for Mobile Monitoring Study Design. **3.4.1. Comparing Results on a Common Scale.** To facilitate comparison among all mapping approaches tested in the MC analyses, Figure 4 rescales the dependent axis (i.e., drive days and road coverage) to be a function of the total number of 1-Hz samples for each of the NO approaches (see Figures S10 and S11 for BC). This rescaling allows a direct comparison of the trends in performance vs sampling effort for the LUR-K drive day, LUR-K road coverage, and data-only approaches. Absolute comparisons between requirements for the various methods require particular care (see section 3.4.2).

Our LUR-K models approach their upper-bound performance quickly and then show little value from increasing sampling. Over 2 orders of magnitude of sample size ($\sim 3 \times 10^4$ – 3.5×10^6 samples, ~ 10 – 1000 h of sampling), LUR-K models consistently have $R^2 \sim 0.5$ – 0.6 . Figure 4 illustrates how for LUR-K modeling, there is little benefit to sampling all road segments: sampling a representative subset of roads produces similar LUR model results with dramatically fewer 1-Hz samples. For a given training sample size, LUR-K performance

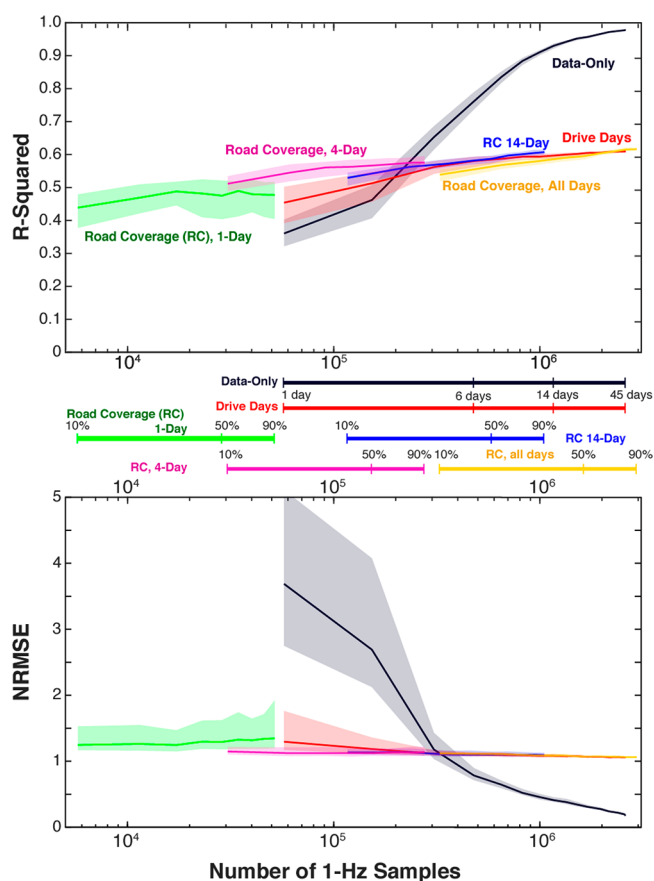


Figure 4. Rescaling data from Figure 3 on a common X-axis representing the number of 1-Hz samples. (Top) R^2 versus the number of 1-Hz samples. Lines represent the median value for all 100 MC subsamples. Shaded area around line represents the observed IQR for all 100 MC subsamples. (Bottom) NRMSE versus the number of 1-Hz samples. The line segments in the middle correspond in color to lines in the figures. The line segments provide context to the number of drive days or percentage of road segments for the corresponding number of 1-Hz samples. The X-axis is on the log-scale.

appears slightly more sensitive to repetition frequency than to road coverage: models perform better with more repetition on a smaller number of road segments. The rescaling also highlights how a moderately performing LUR-K model ($R^2 \sim 0.45$) can be obtained with minimal sampling effort: for example, driving 30% of roads only once would require ~ 5 h of sampling (0.5% of our data set).

Data-only maps considerably outperformed any of the LUR-K models ($R^2 > 0.7$, $\text{NRMSE} < 1$) once $>10\%$ of our measurement data set (~ 100 h, 3.6 M samples) was included in each road segment. Unlike LUR-K models, data-only mapping benefits strongly from additional repeat drives at each location, with substantial improvement up to ~ 15 – 20 repeated drives, and diminishing returns beyond. For a small number of 1-Hz samples, data-only maps generally perform worse than a LUR-K model trained on an equivalent number of 1-Hz samples, because the sampling effort must be dispersed over the full model domain.

3.4.2. Limitations and Generalizability. The findings presented here are particular to our Oakland measurement domain, and so generalizations should be made with care. Notable aspects of our domain include the following: a relatively dense and homogeneous level of urbanization, consistent and steady

westerly winds, an absence of major upwind pollution sources, and an absence of complex topography. Future studies could usefully explore similar questions in a different locale with more pronounced meteorological seasonal variability. Additional considerations for interpreting our results include our sampling design, which required multiple calendar days to completely cover the full domain; the choice of pollutants (NO and BC are dominated by local traffic emissions); and our instrumentation selection (lab-grade instruments, albeit with only fair 1-Hz precision for BC).

Figure 4 in particular should be interpreted with care, since the sampling requirements for each mapping approach may scale differently with respect to the size of a domain. For domains much larger than what we considered in Oakland, diminishing returns for road coverage might arise after sampling far less than 10–20% of the domain, which would make predictive modeling comparatively more efficient.

An additional consideration for interpretation of our results relates to the spatial scale of analysis. Because we were expressly interested in the extent to which LUR-K models could reliably predict fine-scale pollution patterns, our core analyses modeled performance on the basis of predictions at individual 30-m road segments. At coarser analysis scales (e.g., if LUR-K predictions are smoothed over 100+ m), our LUR-K models reproduce the underlying observed data with greater fidelity ($R^2 > 0.7$ – 0.8 , see Figure S12). For epidemiological analyses, a further issue for interpretation is the degree to which on-road concentrations relate to true exposures at residential addresses. Limited evidence from prior mobile monitoring studies suggests that on-road concentrations are correlated with residential address levels and are associated with adverse health effects of subjects at those locations.^{7,12,31}

3.4.3. Implications for Efficient Mobile Monitoring Strategies. Our results indicate that data-only mapping and LUR-K modeling are viable strategies for developing high-resolution air pollution maps that reproduce key patterns in our data set. Relative strengths of the data-only mapping approach include the following: (i) superior performance in estimating long-term mean concentrations when multiple repeated drives are possible, (ii) the simplicity of a model-free approach, and (iii) estimation errors have little remaining spatial structure. Relative strengths of the LUR-K approach include (i) possibly large efficiency gains from very low measurement data requirements and (ii) the ability to develop concentration estimates without making measurements at every location. With limited monitoring funds or time, then the LUR-K approach will provide exposure estimates that capture general spatial trends. If identifying potential air pollution hot-spots is the goal, then the data-only approach has more power to do so. More research is needed to determine whether hyper-local hot-spots are important in population exposure assessment. Therefore, the LUR-K approach, which has less variable estimates than the data-only approach, may be sufficient.

In previous work,⁷ we estimated that ~400 mobile monitoring vehicles could make an annual data-only map (20 drives/year) for all roads in the largest 25 US cities (111 M people, 50% of US urban population). Model-based methods such as LUR-K may be used for mapping a large area if every road cannot be driven or budget limitations prevent the deployment of hundreds of mobile monitoring vehicles. If our findings here can be extrapolated, we speculate that ~25 such vehicles might be sufficient to develop moderately performing LUR-K models for the same cities (e.g., by driving 25% of roads 5 times each).

There are numerous strategies for deploying air sensors on vehicle fleets, each of which may lend themselves particularly well to a specific data analysis approach. Here, vehicle scheduling was actively managed for the purpose of collecting air pollution data. Despite using two cars full-time for 2 years, we faced real constraints about how much we could sample, forcing trade-offs between repetition frequency and domain size. In many other applications of mobile monitoring, sensors are deployed “passively” on vehicle fleets that operate for other reasons. Using such fleets presents opportunities (e.g., scalability) while also introducing constraints. In some cases, sampling every road may not be feasible, for example in the case of buses or trams that ply only along fixed routes.^{27,28} In other cases, vehicles may routinely drive every road but only at specific times at each location (e.g., garbage trucks). Our results present a framework for considering the relative benefits of alternative approaches for using mobile monitors to estimate long-term spatial patterns of air pollution.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.8b03395.

Text, figures, and tables with detailed information on experimental methods, calculation of geographic covariates, subsampling and model validation schemes, supplemental results, and sensitivity analyses (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jsapte@utexas.edu.

ORCID

Michael Brauer: 0000-0002-9103-9343

Jonathan J. Choi: 0000-0001-7403-5853

Joshua S. Apte: 0000-0002-2796-3478

Notes

The contents of this article do not necessarily reflect the views or policies of the US EPA, HEI and its sponsors, or any other funders. US EPA and other funders do not endorse any products or commercial services mentioned in this publication.

The authors declare the following competing financial interest(s): J.S.A. was supported by a Google Earth Engine Research Award, and M.M.L. and B.L. are employed by Aclima, Inc.

■ ACKNOWLEDGMENTS

This research was funded in part by a gift to the Environmental Defense Fund from Signe Ostby and Scott Cook. We are grateful for contributions from K. Tuxen-Bettman, A. Raman, R. Moore, D. Herzl, M. Chu Baird, C. Ely, E. Craft, M. Serre, and the Google Street View and Aclima mobile platform teams. This publication was partially developed under Assistance Agreement No. R835873 awarded by the U.S. Environmental Protection Agency (US EPA) and under contract to the Health Effects Institute (HEI), an organization jointly funded by the US EPA (Assistance Award No. R-82811201) and certain motor vehicle and engine manufacturers.

■ REFERENCES

(1) Brook, R. D.; Rajagopalan, S.; Pope, C. A.; Brook, J. R.; Bhatnagar, A.; Diez-Roux, A. V.; Holguin, F.; Hong, Y.; Luepker, R. V.; Mittleman, M. A.; et al. Particulate matter air pollution and

cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation* **2010**, *121* (21), 2331–2378.

(2) Kaufman, J. D.; Adar, S. D.; Barr, R. G.; Budoff, M.; Burke, G. L.; Curl, C. L.; Daviglius, M. L.; Roux, A. V. D.; Gasset, A. J.; Jacobs, D. R.; et al. Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): a longitudinal cohort study. *Lancet* **2016**, *388* (10045), 696–704.

(3) Pope, C. A.; Ezzati, M.; Dockery, D. W. Fine-particulate air pollution and life expectancy in the United States. *N. Engl. J. Med.* **2009**, *360* (4), 376–386.

(4) Apte, J. S.; Marshall, J. D.; Cohen, A. J.; Brauer, M. Addressing global mortality from ambient PM_{2.5}. *Environ. Sci. Technol.* **2015**, *49* (13), 8057–8066.

(5) Lim, S. S.; Vos, T.; Flaxman, A. D.; Danaei, G.; Shibuya, K.; Adair-Rohani, H.; Amann, M.; Anderson, H. R.; Andrews, K. G.; Aryee, M.; Atkinson, C.; Bacchus, L. J.; et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **2012**, *380* (9859), 2224–2260.

(6) Apte, J. S.; Brauer, M.; Cohen, A. J.; Ezzati, M.; Pope, C. A. Ambient PM_{2.5} reduces global and regional life expectancy. *Environ. Sci. Technol. Lett.* **2018**, *5* (9), 546–551.

(7) 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.

(8) Karner, A. A.; Eisinger, D. S.; Niemeier, D. A. Near-roadway air quality: Synthesizing the findings from real-world data. *Environ. Sci. Technol.* **2010**, *44* (14), 5334–5344.

(9) Marshall, J. D.; Nethery, E.; Brauer, M. Within-urban variability in ambient air pollution: Comparison of estimation methods. *Atmos. Environ.* **2008**, *42* (6), 1359–1369.

(10) Health Effects Institute. *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*; Boston, 2010.

(11) Szpiro, A. A.; Sheppard, L.; Lumley, T. Efficient measurement error correction with spatially misaligned data. *Biostatistics* **2011**, *12* (4), 610–623.

(12) Alexeeff, S. E.; Roy, A.; Shan, J.; Liu, X.; Messier, K.; Apte, J. S.; Portier, C.; Sidney, S.; Van Den Eeden, S. K. High-resolution mapping of traffic related air pollution with Google street view cars and incidence of cardiovascular events within neighborhoods in Oakland, CA. *Environ. Health* **2018**, *17* (1), 38.

(13) Jerrett, M.; Arain, A.; Kanaroglou, P.; Beckerman, B.; Potoglou, D.; Sahsuvaroglu, T.; Morrison, J.; Giovis, C. A review and evaluation of intraurban air pollution exposure models. *J. Exposure Sci. Environ. Epidemiol.* **2005**, *15* (2), 185–204.

(14) Hoek, G.; Beelen, R.; de Hoogh, K.; 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.

(15) Reyes, J. M.; Serre, M. L. An LUR/BME framework to estimate PM_{2.5} explained by on road mobile and stationary sources. *Environ. Sci. Technol.* **2014**, *48* (3), 1736–1744.

(16) 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.

(17) Novotny, E. V.; Bechle, M. J.; Millet, D. B.; Marshall, J. D. National satellite-based land-use regression: NO₂ in the United States. *Environ. Sci. Technol.* **2011**, *45* (10), 4407–4414.

(18) Bechle, M. J.; Millet, D.; Marshall, J. D. A national spatiotemporal exposure surface for NO₂: monthly scaling of a satellite-derived land-use regression, 2000 - 2010. *Environ. Sci. Technol.* **2015**, *49*, 12297–12305.

(19) Briggs, D. J.; Collins, S.; Elliott, P.; Fischer, P.; Kingham, S.; Lebret, E.; Pryn, K.; Van Reeuwijk, H.; Smallbone, K.; Van Der Veen, A. Mapping urban air pollution using GIS: a regression-based approach. *Int. J. Geogr. Inf. Sci.* **1997**, *11*, 699–718.

(20) Kanaroglou, P. S.; Jerrett, M.; Morrison, J.; Beckerman, B.; Arain, M. A.; Gilbert, N. L.; Brook, J. R. Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach. *Atmos. Environ.* **2005**, *39* (13), 2399–2409.

(21) Su, J. G.; Jerrett, M.; Beckerman, B. A distance-decay variable selection strategy for land use regression modeling of ambient air pollution exposures. *Sci. Total Environ.* **2009**, *407* (12), 3890–3898.

(22) Beelen, R.; Hoek, G.; Vienneau, D.; Eeftens, M.; Dimakopoulou, K.; Pedeli, X.; Tsai, M.-Y.; Künzli, N.; Schikowski, T.; Marcon, A.; et al. Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. *Atmos. Environ.* **2013**, *72*, 10–23.

(23) Saraswat, A.; Apte, J. S.; Kandlikar, M.; Brauer, M.; Henderson, S. B.; Marshall, J. D. Spatiotemporal land use regression models of fine, ultrafine, and black carbon particulate matter in New Delhi, India. *Environ. Sci. Technol.* **2013**, *47* (22), 12903–12911.

(24) Henderson, S. B.; Beckerman, B.; Jerrett, M.; Brauer, M. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ. Sci. Technol.* **2007**, *41* (7), 2422–2428.

(25) Westerdahl, D.; Fruin, S.; Sax, T.; Fine, P. M.; Sioutas, C. Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. *Atmos. Environ.* **2005**, *39* (20), 3597–3610.

(26) Larson, T.; Henderson, S. B.; Brauer, M. Mobile monitoring of particle light absorption coefficient in an urban area as a basis for land use regression. *Environ. Sci. Technol.* **2009**, *43* (13), 4672–4678.

(27) Hagemann, R.; Corsmeier, U.; Kottmeier, C.; Rinke, R.; Wieser, A.; Vogel, B. Spatial variability of particle number concentrations and NO_x in the Karlsruhe (Germany) area obtained with the mobile laboratory “AERO-TRAM”. *Atmos. Environ.* **2014**, *94*, 341–352.

(28) Hasenfratz, D.; Saukh, O.; Walser, C.; Hueglin, C.; Fierz, M.; Arn, T.; Beutel, J.; Thiele, L. Deriving high-resolution urban air pollution maps using mobile sensor nodes. *Pervasive Mob. Comput.* **2015**, *16* (PB), 268–285.

(29) Abernethy, R. C.; Allen, R. W.; McKendry, I. G.; Brauer, M. A land use regression model for ultrafine particles in Vancouver, Canada. *Environ. Sci. Technol.* **2013**, *47* (10), 5217–5225.

(30) Hankey, S.; Marshall, J. D. Land use regression models of on-road particulate air pollution (particle number, black carbon, PM_{2.5}, particle size) using mobile monitoring. *Environ. Sci. Technol.* **2015**, *49* (15), 9194.

(31) Kerckhoffs, J.; Hoek, G.; Messier, K. P.; Brunekreef, B.; Meliefste, K.; Klompmaker, J. O.; Vermeulen, R. Comparison of ultrafine particle and black carbon concentration predictions from a mobile and short-term stationary land-use regression model. *Environ. Sci. Technol.* **2016**, *50* (23), 12894–12902.

(32) Mueller, M. D.; Hasenfratz, D.; Saukh, O.; Fierz, M.; Hueglin, C. Statistical modelling of particle number concentration in Zurich at high spatio-temporal resolution utilizing data from a mobile sensor network. *Atmos. Environ.* **2016**, *126*, 171–181.

(33) Weichenthal, S.; Van Ryswyk, K.; Goldstein, A.; Shekarrizfard, M.; Hatzopoulou, M. Characterizing the spatial distribution of ambient ultrafine particles in Toronto, Canada: A land use regression model. *Environ. Pollut.* **2016**, *208*, 241–248.

(34) Klompmaker, J. O.; Montagne, D. R.; Meliefste, K.; Hoek, G.; Brunekreef, B. Spatial variation of ultrafine particles and black carbon in two cities: Results from a short-term measurement campaign. *Sci. Total Environ.* **2015**, *508*, 266–275.

(35) Sabaliauskas, K.; Jeong, C. H.; Yao, X.; Reali, C.; Sun, T.; Evans, G. J. Development of a land-use regression model for ultrafine particles in Toronto, Canada. *Atmos. Environ.* **2015**, *110*, 84–92.

(36) Rivera, M.; Basagana, X.; Aguilera, I.; Agis, D.; Bouso, L.; Foraster, M.; Medina-Ramon, M.; Pey, J.; Kunzli, N.; Hoek, G. Spatial distribution of ultrafine particles in urban settings: A land use regression model. *Atmos. Environ.* **2012**, *54*, 657–666.

(37) Messier, K. P.; Akita, Y.; Serre, M. L. Integrating address geocoding, land use regression, and spatiotemporal geostatistical estimation for groundwater tetrachloroethylene. *Environ. Sci. Technol.* **2012**, *46* (5), 2772–2780.

(38) Raaschou-Nielsen, O.; Andersen, Z. J.; Beelen, R.; Samoli, E.; Stafoggia, M.; Weinmayr, G.; Hoffmann, B.; Fischer, P.; Nieuwenhuijsen, M. J.; Brunekreef, B.; et al. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *Lancet Oncol.* **2013**, *14* (9), 813–822.

(39) Hengl, T.; Heuvelink, G. B. M.; Stein, A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* **2004**, *120*, 75–93.

(40) de Nazelle, A.; Arunachalam, S.; Serre, M. L. Bayesian maximum entropy integration of ozone observations and model predictions: an application for attainment demonstration in North Carolina. *Environ. Sci. Technol.* **2010**, *44* (15), 5707–5713.

(41) Sampson, P. D.; Szpiro, A. A.; Sheppard, L.; Lindström, J.; Kaufman, J. D. Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data. *Atmos. Environ.* **2011**, *45* (36), 6593–6606.

(42) Irvine, K. M.; Gitelman, A. I.; Hoeting, J. A. Spatial designs and properties of spatial correlation: Effects on covariance estimation. *J. Agric. Biol. Environ. Stat.* **2007**, *12* (4), 450–469.

(43) Cressie, N.; Majure, J. J. Spatio-temporal statistical modeling of livestock waste in streams. *J. Agric. Biol. Environ. Stat.* **1997**, *2* (1), 24–47.

(44) 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.

(45) Minet, L.; Gehr, R.; Hatzopoulou, M. Capturing the sensitivity of land-use regression models to short-term mobile monitoring campaigns using air pollution micro-sensors. *Environ. Pollut.* **2017**, *230*, 280–290.