

Integrating Mobile and Fixed-Site Black Carbon Measurements to Bridge Spatiotemporal Gaps in Urban Air Quality

Chirag Manchanda, Robert A. Harley, Julian D. Marshall, Alexander J. Turner, and Joshua S. Apte*



Cite This: <https://doi.org/10.1021/acs.est.3c10829>



Read Online

ACCESS |



Metrics & More



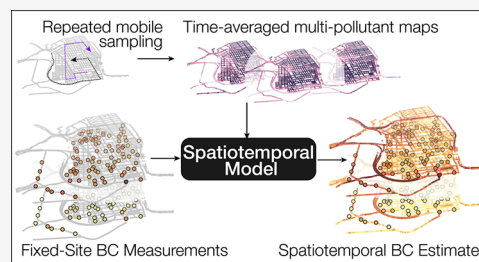
Article Recommendations



Supporting Information

ABSTRACT: Urban air pollution can vary sharply in space and time. However, few monitoring strategies can concurrently resolve spatial and temporal variation at fine scales. Here, we present a new measurement-driven spatiotemporal modeling approach that transcends the individual limitations of two complementary sampling paradigms: mobile monitoring and fixed-site sensor networks. We develop, validate, and apply this model to predict black carbon (BC) using data from an intensive, 100-day field study in West Oakland, CA. Our spatiotemporal model exploits coherent spatial patterns derived from a multipollutant mobile monitoring campaign to fill spatial gaps in time-complete BC data from a low-cost sensor network. Our model performs well in reconstructing patterns at fine spatial and temporal resolution (30 m, 15 min), demonstrating strong out-of-sample correlations for both mobile (Pearson's $R \sim 0.77$) and fixed-site measurements ($R \sim 0.95$) while revealing features that are not effectively captured by a single monitoring approach in isolation. The model reveals sharp concentration gradients near major emission sources while capturing their temporal variability, offering valuable insights into pollution sources and dynamics.

KEYWORDS: black carbon, spatiotemporal modeling, mobile monitoring, low-cost sensors, hyperlocal, urban air quality



1. INTRODUCTION

Air pollution adversely impacts public health and the environment.^{1–3} Owing to the interplay of atmospheric dynamics and unevenly distributed sources, air pollution can vary sharply in space and time.^{4–8} While conventional air pollution measurements are sparse in space and/or time, recognizing the historically high cost of acquiring measurements, spatiotemporally resolved air pollution data are increasingly in demand. This need arises from the desire to gain detailed insights into the sources and processes in urban environments, address societal impacts such as exposures and inequalities, and facilitate effective management strategies.^{9–11}

There has been a surge of interest in hyperlocal air pollution monitoring, including mobile monitoring^{7,8,12–27} and fixed-site low-cost sensor (LCS) networks.^{7,28} These methods can overcome the limitations of traditional monitoring to better capture the fine-scale structure of air pollution in urban settings but have their own advantages and drawbacks.⁷ Mobile measurements can provide high spatial resolution to quantify fine-scale concentration gradients, thereby identifying previously undiscovered pollutant sources and hotspots. However, they have intermittent temporal coverage, resulting in data that may miss critical pollution events. In contrast, like their regulatory monitoring counterparts, fixed-site LCS networks typically provide temporally complete data. However, despite considerably higher spatial density in comparison to regulatory monitoring, LCS networks often still have large spatial gaps compared to the street-level resolution attainable through

mobile monitoring. Here, we present and explore a new method for capitalizing on the complementary strengths of these two hyperlocal monitoring paradigms.

Previous studies have explored methods to address the limitations of these monitoring strategies.²⁹ These approaches include land-use regression (LUR),^{10,18,30–33} kriging methods,^{10,34} and a combination of other statistical learning techniques.^{4,13,21,30,34} However, few studies have explored combining fixed-site and mobile measurements into a single spatiotemporal data product. Adams and Kanaroglou³⁶ utilized a neural network-LUR framework to combine stationary and mobile PM_{2.5} measurements with land-use covariates and meteorology to develop spatially dense hourly estimates of PM_{2.5}. Following a similar approach, Simon et al.³⁷ employed mobile measurements to calculate spatial enhancements in UFP concentration in relation to a stationary monitor, subsequently integrating these enhancements into a regression model. However, very few studies have attempted to integrate stationary and mobile pollutant measurements into a fully measurement-driven spatiotemporal model, that is, a model

Received: December 22, 2023

Revised: June 7, 2024

Accepted: June 10, 2024

relying exclusively on observations as inputs without additional predictor variables.

Our objective here is to describe, validate, and apply a new spatiotemporal modeling methodology for capturing fine-scale spatiotemporal variation in black carbon (BC) by fusing data from a LCS network and mobile monitoring. By leveraging information from coherent multipollutant spatial patterns measured by mobile monitoring, we effectively reveal nuanced spatiotemporal patterns in BC that are not readily apparent in either underlying data set.

2. MATERIALS AND METHODS

2.1. Measured Data and Study Area. Our modeling approach entails fusing *time-resolved* but spatially sparse fixed-site data from low-cost sensors with *time-averaged* but spatially dense maps from mobile monitoring. The approach allows us to develop a BC model at high spatial (30 m) and temporal (15 min) resolution. To do so, we exploit an unusually rich data set of mobile- and fixed-site measurements collected in West Oakland (WO), California.

During a 100 day period from 19th May to 27th August 2017, Caubel et al.²⁸ deployed a BC measurement network consisting of 100 sensors (“100 × 100 BC network”) distributed over an area of 15 km² in WO. This network relied on a custom low-cost sensor (LCS), the Aerosol Black Carbon Detector (ABCD), which functions similarly to an aethalometer.^{38,39} The 100 LCS sites were distributed across residential, industrial, and high-traffic microenvironments. These data were recently used by Wai et al.⁴⁰ in a separate effort to develop a spatiotemporal BC model. We focus here on 97 sites located within 30 m of the road network covered by mobile monitoring. The LCS sensors natively report BC at a resolution of 0.5 Hz.²⁸ We systematically experimented with multiple LCS averaging times between 1 and 120 min, ultimately selecting a time resolution of 15 min to balance between preserving spatial heterogeneity and reducing instrument noise (see Supporting Information (SI) Section S1.1 and Figure S1 for details).⁷

In addition, the LCS campaign coincided with an extensive ongoing mobile monitoring effort that used two custom-equipped Google Street View cars, which sampled repeatedly on every city block of WO during 2015–2018. As described by Chambliss et al.⁴¹ and references therein, the vehicles recorded instantaneous (1 Hz) measurements of GPS location and concentrations of BC and other pollutants (NO, NO₂, ultrafine particles [UFP], and 6 size-resolved particle concentrations bins from 0.3 to 10 μm).^{7,12,18} BC was measured using photoacoustic extinctions (PAX, Droplet Measurement Technologies, Longmont, CO);⁴² see SI Section S1.2 for a description of the full measurement suite for other species. Chambliss et al.⁴¹ found strong instrument–instrument agreement between the two mobile PAX instruments and between the PAX and the fixed-site ABCD sensors (PAX–PAX comparison: Pearson $R^2 = 0.97$ and NRMSE = 0.15; PAX–ABCD comparison: Pearson $R^2 = 0.90$ and NRMSE = 0.33).

For our core analysis, we use mobile monitoring data collected between 6 AM and 8 PM on 49 days within the 100 day period that the LCS sensor network operated. This data set emphasizes weekday, daytime conditions (70% on weekdays; 71% between 9 am and 4 pm). In previous work, Apte et al.¹² reported that 10–20 repeat drive passes were sufficient to reproduce key spatial patterns with good precision

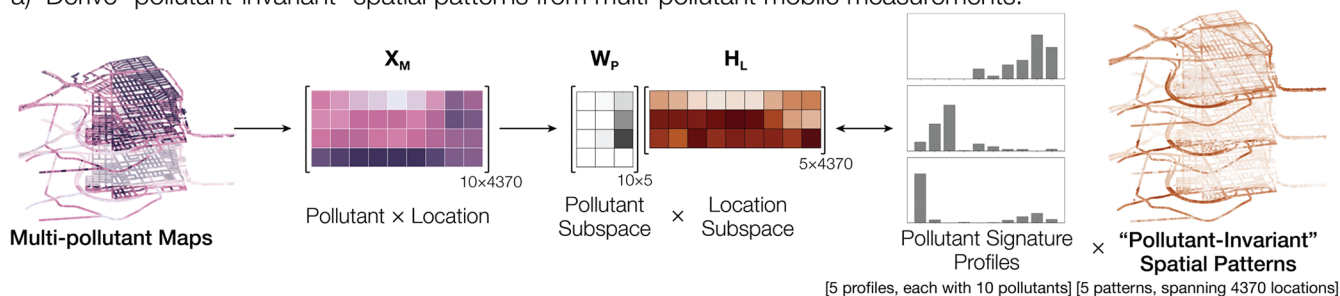
and minimal bias. Here, we restrict the spatial domain to those roads with a minimum of 15 repeated drive visits (median visits = 31). The ~150k 1-Hz time-resolved measurements were aggregated to “median-of-drive-pass-mean” concentrations for ~4300 30-m-long road segments following the approach of Messier et al.¹⁸ That step resulted in 10 time-integrated maps for BC and the 9 other measured pollutants. This suite of multipollutant measurements played an essential role in developing the spatiotemporal model for BC.

2.2. Spatiotemporal Modeling Framework. **2.2.1. Conceptual Framework.** The multifaceted interplay governing air pollution unfolds across three fundamental dimensions: location, time, and pollutant composition. However, most air quality monitoring strategies cannot capture the three dimensions simultaneously. Rather, dominant monitoring paradigms tend to emphasize at most two dimensions. For example, regulatory monitoring systems generally provide continuous measurements of multiple pollutants but only at a small number of distinct locations, thus constituting a *pollutant-time* system. Maps derived from repeated mobile measurements frequently yield time-integrated data for many pollutants at high spatial density, thus representing a *pollutant-location* system. LCS networks excel at providing spatial coverage at high time resolution, constituting a *location-time* system, but generally have a more limited capacity to measure detailed chemical speciation. Similar to a few other recent data-driven air pollution studies,^{43,44} our modeling framework is inspired by a common signal processing technique called compressive sensing (SI Section S1.3).^{45–47} Unlike more conventional spatiotemporal air pollution modeling methods, this approach relies exclusively on observations as inputs and does not require spatial or temporal predictor variables (e.g., land-use data, meteorology).^{30–32,48}

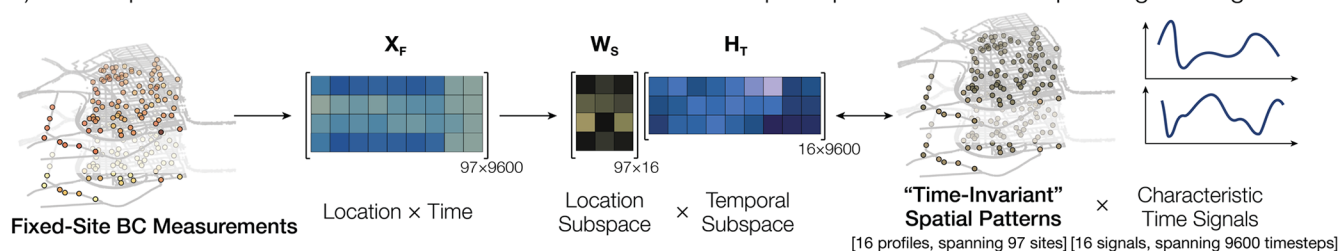
Matrix factorization techniques (e.g., principal component analysis [PCA], non-negative matrix factorization [NMF]) are widely used in air quality research.^{49–52} These techniques exploit coherence across spatial, temporal, or chemical variability to uncover common pollution sources or meteorological conditions concurrently impacting these dimensions. Matrix factorizations are routinely applied to *pollutant-time* systems to leverage covariance among multiple pollutant time series to apportion each pollutant to one or more underlying contributing factors or sources represented by a set of covarying pollutants.⁴⁹ The same techniques can also be extended to *pollutant-location*⁵³ and *location-time*⁵⁰ systems.

Our approach is grounded in the assumption that if common sources or meteorological conditions influence the spatial distribution of multiple pollutants across several time scales, information captured along one dimension can be leveraged to bridge gaps along another dimension. Here, we use the spatial variability captured by time-averaged multipollutant mobile measurements to fill spatial gaps between fixed-site BC measurements. At its core, the model decomposes a *pollutant-location* system, represented by mobile measurement-derived maps for multiple pollutants, and a *location-time* system, represented by continuous fixed-site BC measurements. This decomposition allows for the extraction of spatial patterns that repeat across multiple pollutants and across different time scales, respectively. By decomposing the input multipollutant maps (*pollutant-location*) into a smaller set of repetitive patterns, we are able to express each map as a weighted combination of these patterns (see SI Section S1.4 and Figure S2). Likewise, we decompose the 97 unique LCS

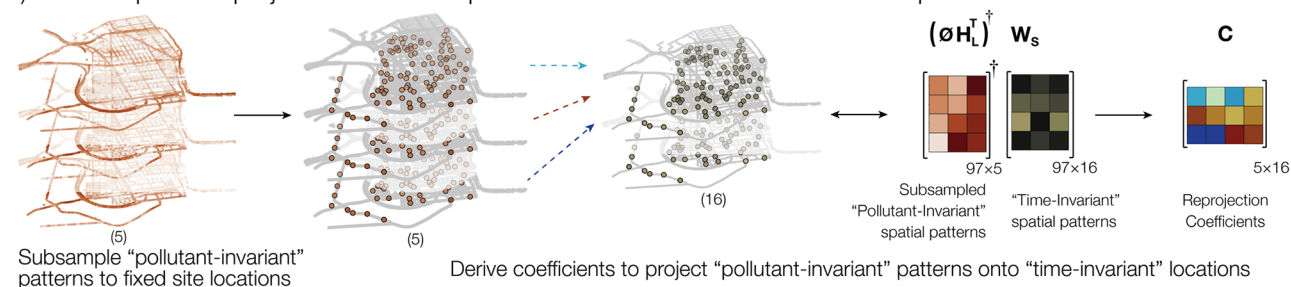
a) Derive “pollutant-invariant” spatial patterns from multi-pollutant mobile measurements.



b) Decompose fixed-site BC measurements into “time-invariant” spatial patterns and corresponding time signals.



c) Derive optimal reprojection between “pollutant-invariant” and “time-invariant” patterns.



d) Combine ‘pollutant invariant’ spatial patterns with characteristic time signals to generate BC estimates.

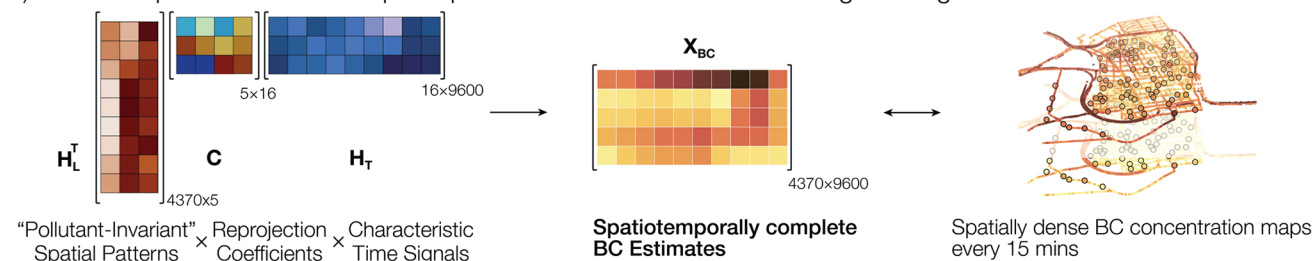


Figure 1. Schematic of a spatiotemporally complete model of BC based on the fusion of mobile and fixed-site data sets collected contemporaneously over 100 days in the Summer of 2017. Conceptually, the approach uses multipollutant information from (a) mobile monitoring, which produces time-averaged but spatially complete multipollutant maps, to fill in spatial gaps in a (b) sparser network of 97 fixed-site BC monitors that provide temporally complete data. To do so, we use matrix factorization (a) to represent time-averaged mobile monitoring maps for 10 pollutants as a linear combination of 5 factors with each factor having its own richly detailed spatial pattern (“pollutant-invariant spatial patterns”). Likewise, we decompose (b) the 100-day temporal variation of 15 min data (9600 observations) at fixed sites into 16 characteristic time signals. The relative weighting of these signals over all sites forms 16 “time-invariant spatial patterns”. In part (c), we derive an optimal reprojection matrix that links the 5 spatially dense multipollutant spatial patterns from part (a) to the 16 spatially sparser temporal patterns from part (b). Finally, these reprojection coefficients are employed to develop a spatially dense temporally continuous model for BC (d) that provides 9600 time estimates (100 days with 15 min data) over a map of 4370 locations in our measurement domain.

time series (*pollutant-time*) into a smaller set of temporal patterns, with each site’s time series expressed as a weighted combination of these patterns. Given that the same meteorological conditions and emission sources influence both sets of patterns, the model establishes a best-fit relationship between them. Consequently, the model utilizes the more comprehensive multipollutant spatial patterns from mobile measurements to fill in the spatial gaps in BC observed in the sparser patterns from fixed-site measurements.

2.2.2. Mathematical Model. Figure 1 provides a schematic illustration outlining our approach. Figure 1a highlights how the multipollutant maps can be represented as a *pollutant \times location* matrix, say $X_M \in R^{p \times l}$, with each matrix element representing time-averaged concentrations of the p^{th} pollutant and l^{th} location. In this study, $p = 10$ pollutants and $l = 4370$ locations (i.e., 4273 road segments plus 97 LCS network sites; we extended the pollutant measurements from the road segments to the nearby LCS sites using ordinary kriging).⁵³

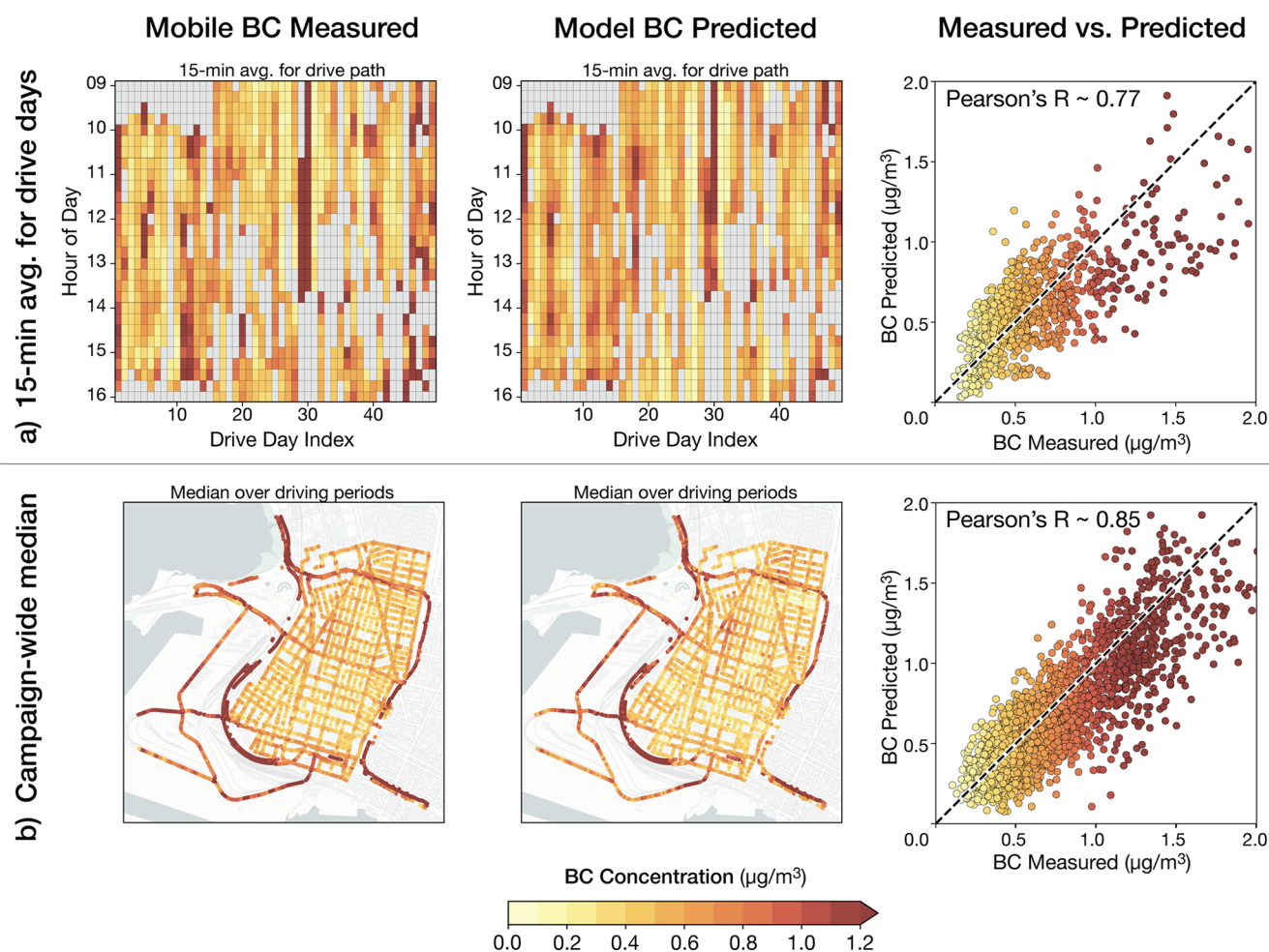


Figure 2. Evaluation of model performance across multiple spatiotemporal scales. (a) Comparison of the spatiotemporal averages along the sampled drive path for mobile measurements (*left*) and model predictions (*center*) for all days when mobile sampling occurred. Each data point and cell represents a 15 min “Lagrangian” spatiotemporal average along the vehicle drive path. (b) Temporal aggregation of time-resolved mobile measurements (*left*) and model predictions (*center*) to produce spatial maps of the campaign-integrated median BC concentrations. Note the high coherence between the spatial patterns of mobile measurements and model predictions.

This matrix \mathbf{X}_M can now be factorized into a pollutant subspace \mathbf{W}_P and a location subspace \mathbf{H}_L , using NMF as follows

$$\mathbf{X}_M = \mathbf{W}_P \mathbf{H}_L \quad \mathbf{W}_P \in \mathbb{R}^{p \times k}; \mathbf{H}_L \in \mathbb{R}^{k \times l} \quad (1)$$

Drawing on analogy to the concept of *pollutant-time* source apportionment, NMF applied to this *pollutant-location* matrix apportions the spatial variation in concentration to k -source profiles. In the present case, using the knee-point method, we found a five-factor solution (i.e., $k = 5$) to be optimal for describing the input data (See SI Section S1.5 and Figure S1 for details). \mathbf{W}_P represents the fractional abundance of the p^{th} pollutant in the k^{th} source, and thus, each column vector in \mathbf{W}_P represents a source profile or source signature (Figure 1a). \mathbf{H}_L represents the normalized concentration attributable to the k^{th} source at the l^{th} location, and thus, each row in \mathbf{H}_L represents a *pollutant-invariant* spatial pattern. Analogous to how *pollutant-time* source apportionment yields source profiles and corresponding time signals, each *pollutant-invariant* spatial pattern describes the impact corresponding to each derived source profile. While these source profiles may contain interpretable information, here, our focus is on using their spatial signatures to translate coherent patterns from one dimension to another.

The second component of the spatiotemporal model consists of the BC measurements obtained from the LCS network. Figure 1b represents the fixed-site BC measurements as a *location* \times *time* matrix, say $\mathbf{X}_F \in \mathbb{R}^{s \times t}$. For the current study, we have 97 LCS sites (s). We averaged measurements at 15 min time resolution over the 100 day study duration, resulting in 9600 timesteps (t). \mathbf{X}_F can similarly be decomposed into a spatial (\mathbf{W}_S) and temporal (\mathbf{H}_T) subspace using NMF as follows:

$$\mathbf{X}_F = \mathbf{W}_S \mathbf{H}_T \quad \mathbf{W}_S \in \mathbb{R}^{s \times q}; \mathbf{H}_T \in \mathbb{R}^{q \times t} \quad (2)$$

Through systematic exploration, we determined that the most suitable value for q was 16 factors (See SI Section S1.5 and Figure S1 for details.) Here, each of the $q = 16$ rows of \mathbf{H}_T represents one of the 16 characteristic time signals, and each of the columns \mathbf{W}_S represents a spatial pattern that remains invariant for that corresponding signal, thus a *time-invariant* BC concentration field or *time-invariant spatial pattern* (Figure 1b). These patterns denote sets of spatial points that covary according to the same time signal. While we do not focus here on the interpretability of these patterns, they may reflect diurnal factors such as traffic or industries or meteorological conditions like wind patterns or urban infrastructure impacts,

affecting similar locations in a similar manner. The NMF-based decomposition of the mobile and fixed-site data yields two matrices, \mathbf{H}_L and \mathbf{W}_S , corresponding to a set of *pollutant-invariant* and *time-invariant* spatial patterns, respectively. Revisiting our initial assumption, if we posit that common pollution sources and meteorological conditions influence the spatial distribution of pollutant concentrations consistently over time, then these *pollutant-invariant* and *time-invariant* spatial patterns essentially portray different perspectives of the same underlying reality. Consequently, they can be reciprocally reprojected, effectively expressing one in terms of the other. In other words, each pattern in one of the sets can be expressed as a combination of patterns in the other set.

Figure 1c illustrates this reprojection process. It starts by taking a transpose of *pollutant-invariant* location subspace $\mathbf{H}_L^T \in \mathbb{R}^{l \times k}$ and subsampling it to the locations where fixed-site measurements are available, using a binary sampling matrix $\emptyset \in \mathbb{R}^{s \times l}$, resulting in $\emptyset \mathbf{H}_L^T \in \mathbb{R}^{s \times k}$, which represent the 5 *pollutant-invariant* patterns at the 97 LCS sites, i.e., the overlap between the *pollutant-invariant* and *time-invariant* patterns. This subsampled matrix subsequently undergoes a Moore–Penrose Inverse with respect to the *time-invariant* location subspace $\mathbf{W}_S \in \mathbb{R}^{s \times q}$, yielding a coefficient matrix $\mathbf{C} \in \mathbb{R}^{k \times q}$. Each column of this coefficient matrix signifies the best-fit coefficients for representing each of the 16 *time-invariant* patterns (derived from eq 2) as a combination of the 5 *pollutant-invariant* patterns such that

$$\underset{\mathbf{C}}{\operatorname{argmin}} \|\mathbf{W}_S - \emptyset \mathbf{H}_L^T \mathbf{C}\|_2 \quad (3)$$

The matrices $\emptyset \mathbf{H}_L^T \in \mathbb{R}^{97 \times 5}$ and $\mathbf{W}_S \in \mathbb{R}^{97 \times 16}$ represent the *pollutant-invariant* patterns and the *time-invariant* patterns at the set of spatial points where both patterns overlap: the LCS network sites (Figure 1c). The coefficient matrix $\mathbf{C} \in \mathbb{R}^{5 \times 16}$ effectively maps one set onto the other. While we derived \mathbf{C} using an overlapping subset between the *time-invariant* and *pollutant-invariant* patterns, the *pollutant-invariant* set $\mathbf{H}_L^T \in \mathbb{R}^{4370 \times 5}$ possesses denser spatial information. Leveraging \mathbf{C} , we can now “fill in the gaps” by combining the *pollutant-invariant* patterns using \mathbf{C} to generate spatially augmented versions of the *time-invariant* patterns, expressed as $\mathbf{H}_L^T \mathbf{C} \in \mathbb{R}^{4370 \times 16}$. These augmented patterns are finally multiplied with the corresponding characteristic time signals, $\mathbf{H}_T \in \mathbb{R}^{16 \times 9600}$ (eq 4), as shown in Figure 1d, to generate the desired model output: spatiotemporally complete BC estimates.

$$\mathbf{X}_{BC} = \mathbf{H}_L^T \mathbf{C} \mathbf{H}_T \quad \mathbf{H}_L^T \in \mathbb{R}^{k \times l}; \mathbf{C} \in \mathbb{R}^{k \times q}; \mathbf{H}_T \in \mathbb{R}^{q \times t} \quad (4)$$

The matrix \mathbf{X}_{BC} represents the resultant model estimates of BC integrating the spatial density of mobile measurements and temporal completeness of the fixed-site LCS measurements. In essence, the model leverages dense spatial patterns observed across multiple pollutants to fill gaps in sparser spatial patterns repeating over time to estimate a spatiotemporally complete BC surface.

2.3. Core Model and Sensitivity Cases. Here, we briefly describe the core model, which we present in Figures 2–4, as well as a set of five alternative models (see Table S1) developed to interrogate the robustness of our overall approach. The core model is developed using data from 97 fixed-site BC sensors, operating over the 100 days from May 19th to Aug 27th, 2017, and time-averaged multipollutant maps developed based on 49 days of available data from the

two Google Street View cars operating in this time period. The factor analyses, core model, and all sensitivity test models were primarily developed using Python, leveraging multiple open-source libraries and packages (see SI Section S1.6 for details).

Sensitivity cases A and B aid in evaluating the out-of-sample prediction fidelity of our model. In sensitivity case A, we iteratively trained the model twice, each time using data from only one of the two Google Street View cars (and all of the data from the LCS sites), using the time-resolved and time-averaged data from the held-out car exclusively for model evaluation.¹² In sensitivity case B, we trained the model on a randomly selected subset of 70% of the LCS sites (68) and iteratively refit the model 1000 times using different permutations of sites for the training. We then use the remaining 30% of fixed sites to test the out-of-sample prediction performance of the spatiotemporal model. In sensitivity case C, we explored how model performance depends on the number of overall LCS fixed sites by parametrically sampling and refitting the model 1000 times each for a random subset of 10 to 80 LCS sites. Since the model relies on “representative” time-averaged mobile maps¹² rather than time-varying mobile measurements, in sensitivity case D, we explored the sensitivity of our results to the period of mobile monitoring data. To do so, we incorporated mobile monitoring from a random selection (repeated with 100 Monte Carlo iterations) of 49 drive days from other years and/or seasons collected during the time window between May 2015 and December 2017 but excluding the 100 day window of the LCS sampling campaign. In sensitivity case E, we systematically assessed how model performance is impacted by the number of pollutants (in addition to BC) measured using mobile monitoring (see Table S1).

3. RESULTS AND DISCUSSION

Our spatiotemporal model estimates BC concentrations during 100 days at 15 min resolution for all ~4300 30-m road segments in the ~15 km² West Oakland domain. We first describe our process for evaluating model performance and then discuss the key insights from the outputs.

3.1. Model Performance Evaluation. Because our model makes spatiotemporally complete estimates, whereas available observation data sets are sparser in space and/or time, careful consideration is needed to evaluate our predictions at an appropriate “apples-to-apples” spatiotemporal resolution. In addition to directly validating our core model against mobile and LCS data at multiple spatial, temporal, and spatiotemporal scales, we performed cross-validation analyses in sensitivity cases A and B to use fully independent data sets for model validation.

3.1.1. Evaluation Using Mobile Monitoring Data. Because our model uses mobile monitoring data that has been extensively time-averaged, *time-varying* mobile BC measurements are useful for validation, especially for temporal performance (Figure 2). We compare our model output against mobile measurements in 15 min model timesteps using a “Lagrangian” approach, where we follow the vehicle’s sampling path. For each 15 min time window with available mobile data, we compare (i) the time average of the measured 1 Hz BC concentrations along the sampling route during this time interval and (ii) the spatial average of the 15 min-average model predictions for all road segments traversed by the vehicle (see SI Section S1.7 for details of our Lagrangian comparison method).

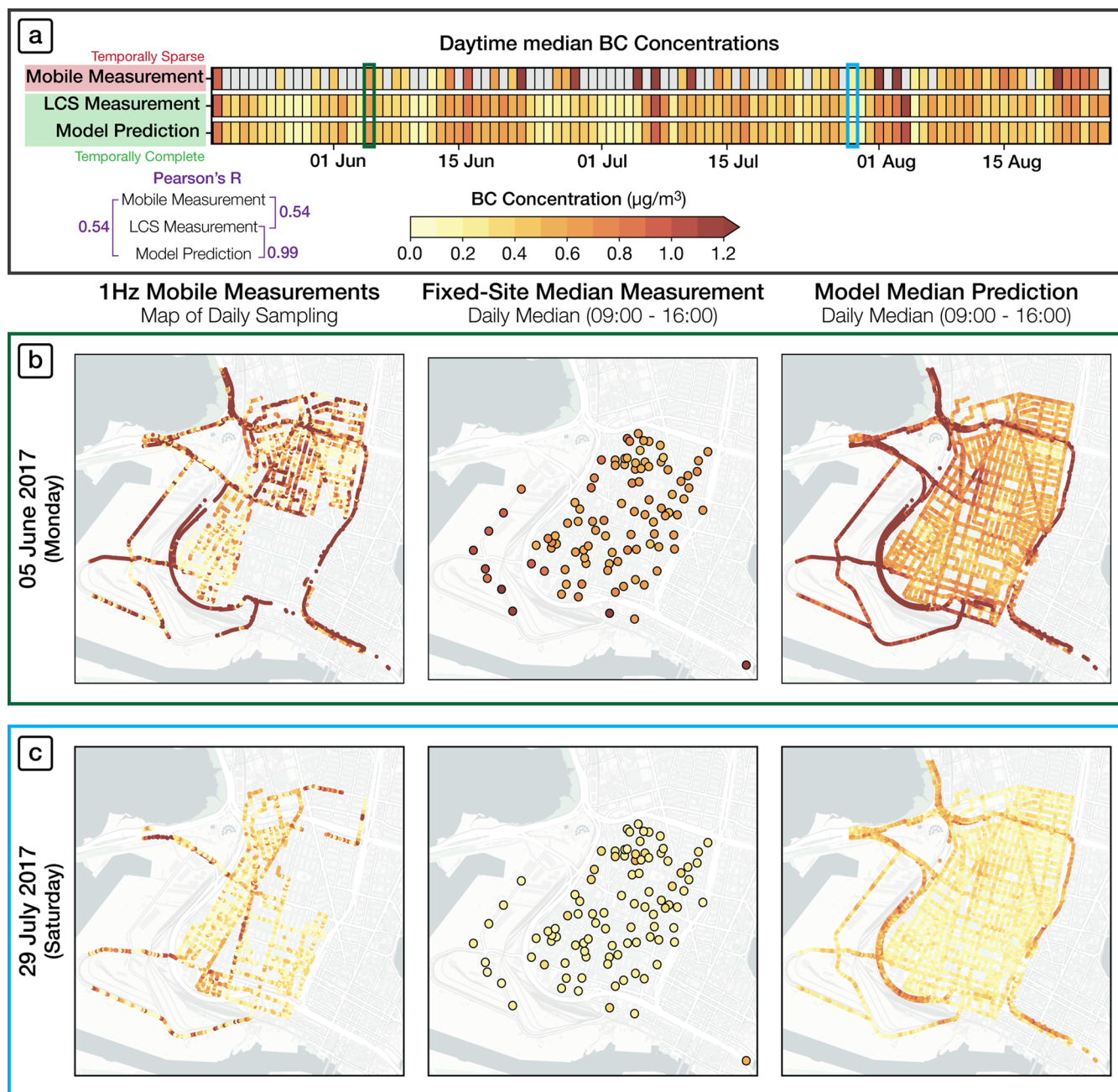


Figure 3. Temporal gap-filling. (a) Time-series heatmap shows median BC concentrations for daytime hours, comparing mobile measurements, fixed-site measurements, and model predictions. Note how the poor spatial and temporal representativeness of mobile measurements on any given day leads to a low correlation with measurements. Panels (b) and (c) show results for Monday June 5 and Saturday July 29. Maps from a single day of driving (*left*) have large spatial gaps and are not temporally representative. Fixed-site measurements (*center*) are temporally complete but spatially sparse. In contrast, predicted BC maps from the model (*right*) provide spatial completeness at any temporal scale. Here, we show daytime median BC predictions for two different days, revealing distinct spatial patterns that likely arise from differences in activity, emissions, and meteorology.

3.1.1.1. 15 min Spatiotemporal Averages. In Figure 2a, we compare 15 min spatiotemporal averages for mobile measurements (Figure 2a, left) and model predictions (Figure 2a, center). This evaluation spans all 49 days when mobile sampling took place during the 100 day observational period between 9 AM and 4 PM. We find a high correlation $R = 0.77$ between measured and predicted BC concentrations, even when assessed at a 15 min temporal resolution (Figure 2a, right). (Intriguingly, Figure S3 shows that that our model is also capable of reproducing spatial patterns *within* individual 15 min time intervals with good fidelity: median $R = 0.52$.)

The model performs well in capturing 15 min-average concentrations with low to moderate BC concentrations ($0.2\text{--}1.5 \mu\text{g}/\text{m}^3$) and tends to slightly underpredict the highest concentrations, with the peak absolute error typically falling within the order of approximately 25% (Figure S4). We attribute this in part to the model's inherent limitations in accurately resolving sudden sharp peaks, a challenge shared by other spatiotemporal models. Figure S4a depicts residuals between measured and predicted spatiotemporal averages for all 15 min intervals across all drive days. The distribution of

errors appears random, with no evident temporal bias (see Figure S4a).

We next investigate whether our model is capturing unique *spatiotemporal* features, rather than simply tracking the general temporal evolution of urban background concentrations, following the approach of Wai et al.⁴⁰ We find that our core model has a substantially superior performance against mobile measurements than a time series of area-wide-average model predictions ($R = 0.77$ vs 0.54 , see Figure S5), indicating that our model is capturing intricate spatiotemporal features, not just general temporal patterns that hold across the domain.

Finally, we iteratively held out one of the two cars' monitoring from model building and then used those data for model testing (sensitivity case A, see Section 2.3 and Table S1). Our results show similar performance using one car exclusively for model validation, with strong spatiotemporal correlations of $R = 0.73$ and 0.80 (see Figure S6).

3.1.1.2. Campaign-Integrated Spatial Patterns. Next, we assess the model's ability to replicate observed spatial patterns for the 100 day average (Figure 2b). We compare the time-integrated spatial patterns (median-of-drive-pass-mean concentrations) that emerge from the full mobile monitoring campaign, comparing the aggregation of mobile measurements (Figure 2b, left) against the aggregation of time-resolved model predictions made along the Lagrangian trajectory of the mobile laboratories' repeated visits to each 30 m road segment. This comparison shows an even stronger correlation, $R = 0.85$ (Figure 2b, right), with no systematic spatial pattern of residuals (Figure S4b). This investigation highlights that the spatial performance of the model improves somewhat when averaging over a longer period.

3.1.2. Evaluation Using Fixed-Site LCS Data. To assess the skill of our model at predicting time-series data at fixed sites, we repeatedly used the 30% holdout scheme we described in sensitivity case B to make out-of-sample predictions at LCS sites excluded from model development (see Section 2.3, Table S1, and Figure S7a). The results indicate a very high correlation of predicted and measured BC time series (for the median out-of-sample site, the temporal $R = 0.95$, 10th to 90th percentile range of 0.80 – 0.97) at the LCS network sites as shown in Figures S7 and S8. The normalized mean bias (NMB) is approximately -5% , while the normalized mean absolute error (NMAE) and normalized root-mean-squared error (NRMSE) are 20 and 25% , respectively. Thus, on average, the model exhibits a slight tendency to underpredict the observed concentrations. It is interesting to contrast our results with those from the spatiotemporal model of Wai et al.,⁴⁰ who report a lower mean temporal R^2 value of 0.60 (compared to a Pearson R^2 of ~ 0.90 here) when predicting LCS measurements in a similar leave-site-out evaluation. We attribute the improvement in temporal performance in part to the additional spatial information that mobile monitoring data contributes to our spatiotemporal model.

Considering that mobile monitoring was predominantly conducted on weekdays from 9 AM to 4 PM, we also assessed whether this limited temporal coverage impacts model performance during time periods when only LCS measurements are available. To do so, we segmented the model output by weekday/weekend and daytime/nighttime categories for each of the leave-site-out cross-validation trials. Figure S7b provides an overview of these results, demonstrating consistent model performance across all temporal subset conditions. This result is noteworthy as it underscores that the model's

performance remains relatively unaffected by the absence of mobile measurements during specific days or times, suggesting that the continuous temporal coverage provided by the LCS enables the model to dynamically adjust the weighting of the spatially coherent patterns derived from mobile monitoring data. As a result, the model can identify spatial patterns that might be missed by LCS sites.

3.2. Model Application: Filling in Spatiotemporal Gaps. Next, we explore how the spatiotemporal model can provide new insights by filling monitoring gaps in space and time.

3.2.1. Temporal Gap-Filling. We start with an illustration (Figure 3), which depicts how temporally complete model predictions can offer new insights into the complete spatial patterns of BC at finer time resolution than would be possible with the temporally incomplete mobile monitoring data. Figure 3a depicts the daily time series of spatial averages of daytime (9:00–16:00) median BC concentrations as observed by mobile measurements, LCS measurements, and model predictions over the 100 day study period. We find that the time series of daily spatial averages of our spatiotemporal model closely reproduces that of the LCS sensors ($R = 0.99$). In contrast, we find a lower temporal correlation ($R = 0.54$) between the spatial averages of the daily mobile measurements and our spatiotemporal model. This result makes sense. Within just 1 day of driving, mobile monitoring does not generally capture spatially representative concentration patterns, whereas a spatiotemporally complete model can represent fine-scale spatial patterns for a single day. Moreover, temporally complete model predictions can help address notable limitations of mobile monitoring, including subtle but irreducible temporal sampling biases and the emphasis on daytime sampling.

Next, we contrast daily median spatial patterns for two illustrative days during the campaign: Monday, June 5th, and Saturday, July 29th, 2017 (Figure 3b–c). For mobile measurements, maps derived from a single drive pass are inherently spatially complete and temporally unrepresentative (Figure 3b–c, left). In contrast, fixed-site measurements (Figure 3b–c, center) and model predictions (Figure 3b–c, right) are temporally complete and thus are representative of spatial variability for a given day. Our daily model predictions are well correlated with LCS observations for these 2 days (spatial $R = 0.8$ and 0.75 , respectively; see Figure S9).

Relative to the LCS network, a key advantage of the model is that it can reveal distinct spatial patterns linked to activity that differ from day to day, including sharp concentration gradients near highways, industrial sites, or truck routes converging to the port. Many of these spatial features are in areas overlooked by the sparser LCS network. The time-resolved (i.e., daily) spatial patterns from the model output (Figure 3b–c, right) can deviate substantially from the campaign-averaged map of mobile measurement (cf. Figure 2b, left), revealing intermittent spatial features that are the signature of many of the BC sources in the study domain.

3.2.2. Spatial Gap-filling of Time-Series Data. We next consider how the spatiotemporal model can fill spatial monitoring gaps. We demonstrate that our spatiotemporal model reveals how the temporal variation in BC levels can be quite distinct over even short spatial distances.

To motivate this discussion, we first compare domain-wide maps of the campaign medians of the fixed-site LCS measurements (Figure 4a), and the spatiotemporal model

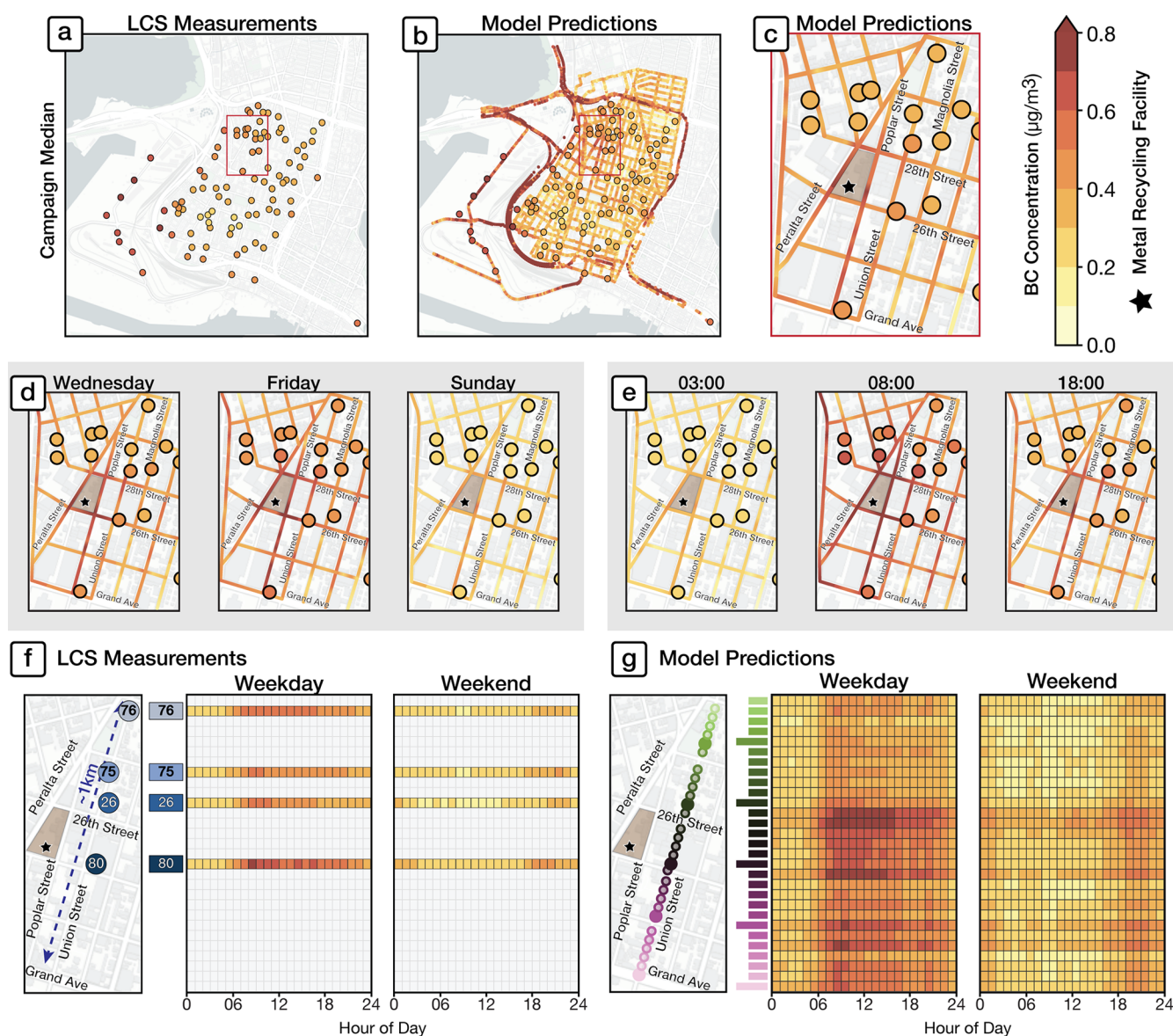


Figure 4. Spatial gap-filling of time-series data. (a) Campaign median of fixed-site BC measurements. (b) Median model predictions overlaid on the sensor network, revealing within-neighborhood variation. (c) Close-up view in a neighborhood surrounding a metal recycling cluster. This spatial gap-filling can be extended to distinct days of the week (d) or times of day (e), revealing localized pollution hotspots near emission sources that are not captured by the spatially sparser monitoring network. We contrast weekday and weekend diurnal time series along a single transect (Union Street) within the zoomed-in region. (f) Diurnal cycles for four closely spaced monitors within three city blocks (site IDs from Caubel et al.²⁵), illustrating the presence of localized pollution sources. (g) Model predictions over the same ~ 1 km transect, illustrating how the magnitude of BC and the corresponding diurnal signatures vary markedly among unobserved locations.

overlaid on the LCS measurements (Figure 4b). Figure 4c provides a zoomed-in view of a single neighborhood containing both residences and multiple industrial facilities, where model predictions exhibit fine-scale spatial variability not captured by the sparser fixed-site sensors. Considering distinct days of week and times of day, Figures 4d-e illustrate that while the LCS sites capture the overall contrast, the finer-scale spatial variation of the model reveals more about the locations of specific activities. Model predictions are overlaid onto LCS sites, revealing the spatial variability captured versus missed across different days of the week or times of the day. The pronounced contrast in fine-scale spatial features, particularly comparing Friday to Sunday (weekday/weekend) and 3:00 am to 8:00 am (night/day), illustrates how spatial features

correspond with the operating hours of major emissions sources in the neighborhood.

To further illustrate the model's proficiency in capturing fine-scale spatial patterns over time, we investigate the fine-scale spatial variability of the time series predicted by the model. We examine how the diurnal variation of BC varies along a ~ 1 km spatial transect defined by a specific road, Union Street (Figure S10). Figure 4f shows heatmaps of the measured diurnal cycles for four closely spaced LCS sites within three city blocks along Union Street. The diurnal patterns vary across the four sites, revealing localized pollution hotspots near facilities such as a metal recycling plant and a moving company.²⁸ We find that the model is capable not only of reproducing similar diurnal profiles (Figure S11) at these four locations (Figure 4g) but also reveals sharp variation in the

diurnal profiles at *unobserved* locations along the Union Street transect. These nuanced diurnal signatures over short distances correspond to localized differences in the microenvironment in this mixed neighborhood, including major roadways, areas near industrial sites, and residential zones. Imagery in Figure S10 provides further context to these findings. This vignette within one heterogeneous neighborhood shows how a spatiotemporally complete model can fill measurement gaps to identify intricate spatial patterns in BC concentrations.

3.3. Limitations and Design Considerations for Future Campaigns. In the present study, we were able to leverage an unusually rich data set of mobile and fixed-site monitoring. We conducted a series of data experiments to explore the potential of our method to work with monitoring data that is more limited along multiple distinct axes: fewer fixed sensors, less temporally representative mobile monitoring, or fewer pollutants.

In sensitivity case C (see Section 2.3 and Table S1), we explore how model performance varies as a function of fixed sites used to construct the model. We find sharply diminishing returns to model performance after including more than 40 randomly selected sites out of the existing 100 sites (Figure S12). If we had sufficient information (e.g., from an initial deployment of many LCS) to place sensors *optimally* rather than randomly, fewer sensors would be required (see SI Section S1.8).⁴³ In Figure S13, we illustrate results from an optimal sensor placement algorithm,⁴⁵ which suggests that as few as 20 well-placed sensors would enable us to reproduce the performance of our core model. As detailed in Section S1.8, the optimal sensor placement algorithm is designed to maximize the spatiotemporal variability captured for a given number of sites. Although the algorithm is purely data-driven, the 20 sites identified as optimal in Figure S13e are located near industrial facilities, designated truck routes, freeways, the port, and railyards—areas where high spatiotemporal variability in BC emissions is expected.

Likewise, it may not always be feasible to simultaneously collect mobile and fixed-site measurements. However, since our model only requires time-averaged maps from mobile measurements, we used sensitivity case D to assess whether mobile measurement maps derived from a period outside the fixed-site measurement's time frame could still yield viable results (see Section 2.3 and Table S1). As shown in Figure S14, we found our model's spatiotemporal performance declined modestly (median $R = 0.67$ vs 0.77) when trained on mobile monitoring data outside of the 100 day study period.

Another consideration is the number of unique pollutants measured by multipollutant mobile monitoring. In sensitivity case E, we systematically considered models developed with combinations of BC and fewer than the 9 total species measured by our mobile monitoring. As we increased the number of species incorporated from 4 to 10 (Figure S15), we found meaningful improvement in R (from ~ 0.6 to 0.77) and other measures of model performance. We attribute this improvement to the broader range of pollutants, along with their shared variance, which enables the model to better identify pollutant-invariant patterns and thereby more effectively fill spatiotemporal gaps. However, with the available evidence, it is difficult to reliably isolate the unique influence of a specific pollutant on model performance since pollutants have correlated spatial patterns (Figure S2) that arise because of common sources and meteorology. Rather, Figure S15 suggests that the model is more sensitive to the number of

pollutants included than to which pollutants are included, implying that our factor analysis benefits principally from having a large set of pollutants, each with somewhat unique spatial patterns. We do not see clear evidence of diminishing returns to adding additional pollutants to the mobile suite from $n = 4$ to 10.

It remains an open question whether this modeling approach would perform well for reactive or secondary air pollutants. Likewise, it would be useful to confirm our findings for BC in other settings with different sources and meteorology. Replicating this study's resource- and labor-intensive level of monitoring may not be feasible everywhere. However, our sensitivity analyses suggest that this modeling technique is viable even with considerably fewer mobile or fixed-site measurements and thus would provide useful information in filling in spatiotemporal gaps from monitoring. Nonetheless, as a data-driven technique, a key limitation is that our approach is not capable of predicting concentrations for time periods or spatial domains that altogether lack measurements. Moreover, in contrast to physics-based models (e.g., dispersion or chemical-transport models), our model is not capable of simulating the impact of emissions changes or other interventions.

4. IMPLICATIONS AND FUTURE WORK

Finely resolved measurements of air pollution in space and time are increasingly needed. This study demonstrates the benefits of combining two disparate monitoring strategies rather than relying on low-cost sensors or mobile monitoring in isolation for capturing the spatiotemporal variation of a primary pollutant (e.g., BC) within a complex urban environment. The novel spatiotemporal model presented here integrates these measurements effectively, revealing fine-scale features in space and time that would be overlooked if considering one monitoring approach individually. A unique aspect of this approach is in using spatial patterns recurring over multiple pollutants collectively to fill spatial gaps observed for a single pollutant over time. Future endeavors can also explore extending the model's applicability to other pollutants, particularly reactive ones, and to regions with different meteorological regimes.

Our modeling approach holds potential for broader applications. Future investigations could assess the impact of human mobility on pollutant exposures using finely resolved spatiotemporal models.⁵⁴ Additionally, the spatiotemporal completeness provided by our model ensures spatial and temporal representativeness when aggregating the output at different scales. This feature may be particularly valuable for epidemiological studies, which often require processed data products like exposure estimates at residential addresses or census block level, either at a time-resolved or time-averaged scale. The fine resolution and completeness of our model output ensure the representativeness of these aggregates in such scenarios. In the future, this modeling approach may prove useful in expanding the capabilities of emerging hyperlocal air pollution observation systems. If applied over periods of time with changing emissions (e.g., in response to control policies, changes in traffic, or the addition of new sources), this method may aid accountability studies in identifying zones of changing emissions and exposure impact.⁵⁵

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.est.3c10829>.

Additional methodological details, sensitivity analyses, and supporting information figures (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Joshua S. Apte – Department of Civil and Environmental Engineering, University of California, Berkeley, California 94720, United States; School of Public Health, University of California, Berkeley, California 94720, United States; orcid.org/0000-0002-2796-3478; Email: apte@berkeley.edu

Authors

Chirag Manchanda – Department of Civil and Environmental Engineering, University of California, Berkeley, California 94720, United States

Robert A. Harley – Department of Civil and Environmental Engineering, University of California, Berkeley, California 94720, United States

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

Alexander J. Turner – Department of Atmospheric Sciences, University of Washington, Seattle, Washington 98195, United States

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.est.3c10829>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank T. Kirchstetter, C. Preble, R. Cohen, M. Harris, R. Alvarez, A. Franco, and T. Thompson. Portions of this work were supported by Environmental Defense Fund.

■ REFERENCES

- (1) Ou, Y.; West, J. J.; Smith, S. J.; Nolte, C. G.; Loughlin, D. H. Air Pollution Control Strategies Directly Limiting National Health Damages in the US. *Nat. Commun.* **2020**, *11* (1), No. 957.
- (2) 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.
- (3) Pope, C. A.; Dockery, D. W. Health Effects of Fine Particulate Air Pollution: Lines That Connect. *J. Air Waste Manage. Assoc.* **2006**, *56* (6), 709–742.
- (4) Lindström, J.; Szpiro, A. A.; Sampson, P. D.; Oron, A. P.; Richards, M.; Larson, T. V.; Sheppard, L. A Flexible Spatio-Temporal Model for Air Pollution with Spatial and Spatio-Temporal Covariates. *Environ. Ecol. Stat.* **2014**, *21* (3), 411–433.
- (5) 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.
- (6) 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), No. 38, DOI: [10.1186/s12940-018-0382-1](https://doi.org/10.1186/s12940-018-0382-1).
- (7) Chambliss, S. E.; Preble, C. V.; Caubel, J. J.; Cados, T.; Messier, K. P.; Alvarez, R. A.; Lafranchi, B.; Lunden, M.; Marshall, J. D.; Szpiro, A. A.; Kirchstetter, T. W.; Apte, J. S. Comparison of Mobile and Fixed-Site Black Carbon Measurements for High-Resolution Urban Pollution Mapping. *Environ. Sci. Technol.* **2020**, *54* (13), 7848–7857.
- (8) Van den Bossche, J.; Peters, J.; Verwaeren, J.; Botteldooren, D.; Theunis, J.; De Baets, B. Mobile Monitoring for Mapping Spatial Variation in Urban Air Quality: Development and Validation of a Methodology Based on an Extensive Dataset. *Atmos. Environ.* **2015**, *105*, 148–161.
- (9) 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.
- (10) Jerrett, M.; Burnett, R. T.; Ma, R.; Pope, C. A.; Krewski, D.; Newbold, K. B.; Thurston, G.; Shi, Y.; Finkelstein, N.; Calle, E. E.; Thun, M. J. Spatial Analysis of Air Pollution and Mortality in Los Angeles. *Epidemiology* **2005**, *16* (6), 727–736.
- (11) Wang, A.; Paul, S.; deSouza, P.; Machida, Y.; Mora, S.; Duarte, F.; Ratti, C. Key Themes, Trends, and Drivers of Mobile Ambient Air Quality Monitoring: A Systematic Review and Meta-Analysis. *Environ. Sci. Technol.* **2023**, *57* (26), 9427–9444, DOI: [10.1021/acs.est.2c06310](https://doi.org/10.1021/acs.est.2c06310).
- (12) 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* (12), 6999–7008.
- (13) Blanco, M. N.; Bi, J.; Austin, E.; Larson, T. V.; Marshall, J. D.; Sheppard, L. Impact of Mobile Monitoring Network Design on Air Pollution Exposure Assessment Models. *Environ. Sci. Technol.* **2023**, *57* (1), 440–450.
- (14) Brantley, H. L.; Hagler, G. S. W.; Kimbrough, E. S.; Williams, R. W.; Mukerjee, S.; Neas, L. M. Mobile Air Monitoring Data-Processing Strategies and Effects on Spatial Air Pollution Trends. *Atmos. Meas. Tech.* **2014**, *7* (7), 2169–2183.
- (15) 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* (8), 4305–4315.
- (16) Healy, R. M.; Sofowote, U. M.; Wang, J. M.; Chen, Q.; Todd, A. Spatially Resolved Source Apportionment of Industrial VOCs Using a Mobile Monitoring Platform. *Atmosphere* **2022**, *13* (10), No. 1722, DOI: [10.3390/atmos13101722](https://doi.org/10.3390/atmos13101722).
- (17) Larson, T.; Gould, T.; Riley, E. A.; Austin, E.; Fintzi, J.; Sheppard, L.; Yost, M.; Simpson, C. Ambient Air Quality Measurements from a Continuously Moving Mobile Platform: Estimation of Area-Wide, Fuel-Based, Mobile Source Emission Factors Using Absolute Principal Component Scores. *Atmos. Environ.* **2017**, *152*, 201–211.
- (18) Messier, K. P.; Chambliss, S. E.; Gani, S.; Alvarez, R.; Brauer, M.; Choi, J. J.; Hamburg, S. P.; Kerckhoffs, J.; Lafranchi, B.; Lunden, M. M.; Marshall, J. D.; Portier, C. J.; Roy, A.; Szpiro, A. A.; Vermeulen, R. C. H.; Apte, J. S. Mapping Air Pollution with Google Street View Cars: Efficient Approaches with Mobile Monitoring and Land Use Regression. *Environ. Sci. Technol.* **2018**, *52* (21), 12563–12572.
- (19) Miller, D. J.; Actkinson, B.; Padilla, L.; Griffin, R. J.; Moore, K.; Lewis, P. G. T.; Gardner-Frolick, R.; Craft, E.; Portier, C. J.; Hamburg, S. P.; Alvarez, R. A. Characterizing Elevated Urban Air Pollutant Spatial Patterns with Mobile Monitoring in Houston, Texas. *Environ. Sci. Technol.* **2020**, *54* (4), 2133–2142.
- (20) 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.
- (21) Padilla, L. E.; Ma, G. Q.; Peters, D.; Dupuy-Todd, M.; Forsyth, E.; Stidworthy, A.; Mills, J.; Bell, S.; Hayward, I.; Coppin, G.; Moore, K.; Fonseca, E.; Popoola, O. A. M.; Douglas, F.; Slater, G.; Tuxen-

- Bettman, K.; Carruthers, D.; Martin, N. A.; Jones, R. L.; Alvarez, R. A. New Methods to Derive Street-Scale Spatial Patterns of Air Pollution from Mobile Monitoring. *Atmos. Environ.* **2022**, *270*, No. 118851.
- (22) Peters, J.; Theunis, J.; Van Poppel, M.; Berghmans, P. Monitoring PM₁₀ and Ultrafine Particles in Urban Environments Using Mobile Measurements. *Aerosol Air Qual. Res.* **2013**, *13* (2), 509–522.
- (23) Riley, E. A.; Schaal, L. N.; Sasakura, M.; Crampton, R.; Gould, T. R.; Hartin, K.; Sheppard, L.; Larson, T.; Simpson, C. D.; Yost, M. G. Correlations between Short-Term Mobile Monitoring and Long-Term Passive Sampler Measurements of Traffic-Related Air Pollution. *Atmos. Environ.* **2016**, *132*, 229–239.
- (24) Solomon, P. A.; Vallano, D.; Lunden, M.; Lafranchi, B.; Blanchard, C. L.; Shaw, S. L. Mobile-Platform Measurement of Air Pollutant Concentrations in California: Performance Assessment, Statistical Methods for Evaluating Spatial Variations, and Spatial Representativeness. *Atmos. Meas. Tech.* **2020**, *13* (6), 3277–3301.
- (25) Sullivan, R. C.; Pryor, S. C. Quantifying Spatiotemporal Variability of Fine Particles in an Urban Environment Using Combined Fixed and Mobile Measurements. *Atmos. Environ.* **2014**, *89*, 664–671.
- (26) Van Poppel, M.; Peters, J.; Bleux, N. Methodology for Setup and Data Processing of Mobile Air Quality Measurements to Assess the Spatial Variability of Concentrations in Urban Environments. *Environ. Pollut.* **2013**, *183*, 224–233.
- (27) Wen, Y.; Wang, H.; Larson, T.; Kelp, M.; Zhang, S.; Wu, Y.; Marshall, J. D. On-Highway Vehicle Emission Factors, and Spatial Patterns, Based on Mobile Monitoring and Absolute Principal Component Score. *Sci. Total Environ.* **2019**, *676*, 242–251.
- (28) Caubel, J. J.; Cados, T. E.; Preble, C. V.; Kirchstetter, T. W. A Distributed Network of 100 Black Carbon Sensors for 100 Days of Air Quality Monitoring in West Oakland, California. *Environ. Sci. Technol.* **2019**, *53* (13), 7564–7573.
- (29) 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.
- (30) 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.
- (31) 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), 9194–9202.
- (32) 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.
- (33) Kim, S.-Y.; Bechle, M.; Hankey, S.; Sheppard, L.; Szpiro, A. A.; Marshall, J. D. Concentrations of Criteria Pollutants in the Contiguous U.S., 1979 – 2015: Role of Prediction Model Parsimony in Integrated Empirical Geographic Regression. *PLoS One* **2020**, *15* (2), No. e0228535.
- (34) 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 (NO_x) for the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Atmos. Environ.* **2011**, *45* (26), 4412–4420.
- (35) Doubleday, A.; Blanco, M. N.; Austin, E.; Marshall, J. D.; Larson, T. V.; Sheppard, L. Characterizing Ultrafine Particle Mobile Monitoring Data for Epidemiology. *Environ. Sci. Technol.* **2023**, *57* (26), 9538–9547.
- (36) Adams, M. D.; Kanaroglou, P. S. Mapping Real-Time Air Pollution Health Risk for Environmental Management: Combining Mobile and Stationary Air Pollution Monitoring with Neural Network Models. *J. Environ. Manage.* **2016**, *168*, 133–141.
- (37) Simon, M. C.; Patton, A. P.; Naumova, E. N.; Levy, J. I.; Kumar, P.; Brugge, D.; Durant, J. L. Combining Measurements from Mobile Monitoring and a Reference Site to Develop Models of Ambient Ultrafine Particle Number Concentration at Residences. *Environ. Sci. Technol.* **2018**, *52* (12), 6985–6995.
- (38) Caubel, J. J.; Cados, T. E.; Kirchstetter, T. W. A New Black Carbon Sensor for Dense Air Quality Monitoring Networks. *Sensors* **2018**, *18* (3), No. 738.
- (39) Hansen, A. D. A.; Rosen, H.; Novakov, T. The Aethalometer — An Instrument for the Real-Time Measurement of Optical Absorption by Aerosol Particles. *Sci. Total Environ.* **1984**, *36*, 191–196.
- (40) Wai, T. H.; Apte, J. S.; Harris, M. H.; Kirchstetter, T. W.; Portier, C. J.; Preble, C. V.; Roy, A.; Szpiro, A. A. Insights from Application of a Hierarchical Spatio-Temporal Model to an Intensive Urban Black Carbon Monitoring Dataset. *Atmos. Environ.* **2022**, *277*, No. 119069, DOI: 10.1016/j.atmosenv.2022.119069.
- (41) Chambliss, S. E.; Pinon, C. P. R.; Messier, K. P.; LaFranchi, B.; Upperman, C. R.; Lunden, M. M.; Robinson, A. L.; Marshall, J. D.; Apte, J. S. Local- and Regional-Scale Racial and Ethnic Disparities in Air Pollution Determined by Long-Term Mobile Monitoring. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118* (37), No. e2109249118, DOI: 10.1073/pnas.2109249118.
- (42) Patrick Arnott, W.; Moosmüller, H.; Fred Rogers, C.; Jin, T.; Bruch, R. Photoacoustic Spectrometer for Measuring Light Absorption by Aerosol: Instrument Description. *Atmos. Environ.* **1999**, *33* (17), 2845–2852.
- (43) Kelp, M. M.; Lin, S.; Kutz, J. N.; Mickley, L. J. A New Approach for Determining Optimal Placement of PM_{2.5} Air Quality Sensors: Case Study for the Contiguous United States. *Environ. Res. Lett.* **2022**, *17* (3), No. 034034.
- (44) Zhou, C.; Gao, M.; Li, J.; Bai, K.; Tang, X.; Lu, X.; Liu, C.; Wang, Z.; Guo, Y. Optimal Planning of Air Quality-Monitoring Sites for Better Depiction of PM_{2.5} Pollution across China. *ACS Environ. Au* **2022**, *2* (4), 314–323.
- (45) Manohar, K.; Brunton, B. W.; Kutz, J. N.; Brunton, S. L. Data-Driven Sparse Sensor Placement for Reconstruction, 2017, DOI: 10.1109/MCS.2018.2810460.
- (46) Baraniuk, R. Compressive Sensing [Lecture Notes]. *IEEE Signal Process. Mag.* **2007**, *24* (4), 118–121.
- (47) Donoho, D. L. Compressed Sensing. *IEEE Trans. Inf. Theory* **2006**, *52* (4), 1289–1306.
- (48) 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.
- (49) Henry, R. C.; Hiiyy, G. M. *Multivariate Analysis of Particulate Sulfate and Other Air Quality Variables by Principal Component Analysis-Part-I*, Annual Data from Los Angeles and New York; Pergamon Press Lki, 1979; Vol. 13.
- (50) Thiem, A.; Schlink, U.; Pan, X.-C.; Hu, M.; Peters, A.; Wiedensohler, A.; Breitner, S.; Cyrys, J.; Wehner, B.; Franck, U. Atmospheric Chemistry and Physics Discussions Using Non-Negative Matrix Factorization for the Identification of Daily Patterns of Particulate Air Pollution in Beijing During. *Atmos. Chem. Phys. Discuss.* **2012**, *12*, 13015–13052.
- (51) Paatero, P.; Tapper, U. Positive Matrix Factorization: A Non-negative Factor Model with Optimal Utilization of Error Estimates of Data Values. *Environmetrics* **1994**, *5* (2), 111–126.
- (52) Hopke, P. K. Review of Receptor Modeling Methods for Source Apportionment. *J. Air Waste Manage. Assoc.* **2016**, *66* (3), 237–259.
- (53) Du, H.; Lu, X. Spatial Distribution and Source Apportionment of Heavy Metal(Loid)s in Urban Topsoil in Mianyang, Southwest China. *Sci. Rep* **2022**, *12* (1), No. 10407, DOI: 10.1038/s41598-022-14695-9.
- (54) Nyhan, M.; Grauw, S.; Britter, R.; Misstear, B.; McNabola, A.; Laden, F.; Barrett, S. R. H.; Ratti, C. Exposure Track”—The Impact of Mobile-Device-Based Mobility Patterns on Quantifying Population Exposure to Air Pollution. *Environ. Sci. Technol.* **2016**, *50* (17), 9671–9681.

(55) Boogaard, H.; van Erp, A. M.; Walker, K. D.; Shaikh, R. Accountability Studies on Air Pollution and Health: The HEI Experience. *Curr. Environ. Health Rep.* **2017**, *4* (4), 514–522.