Environment International 131 (2019) 105022

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea

Chris C. Lim[a,*], Ho Kim[b], M.J. Ruzmyn Vilcassim[a], George D. Thurston[a], Terry Gordon[a], Lung-Chi Chen[a], Kiyoung Lee[b], Michael Heimbinder[c], Sun-Young Kim[d]

[a] Department of Environmental Medicine, New York School of Medicine, New York, NY, United States of America
[b] Graduate School of Public Health, Seoul National University, Seoul, South Korea
[c] HabitatMap, Brooklyn, NY, United States of America
[d] Graduate School of Cancer Science and Policy, National Cancer Center, Gyeonggi, South Korea

ARTICLE INFO

Handling Editor: Xavier Querol

ABSTRACT

Recent studies have demonstrated that mobile sampling can improve the spatial granularity of land use regression (LUR) models. Mobile sampling campaigns deploying low-cost (< $300) air quality sensors could potentially offer an inexpensive and practical approach to measure and model air pollution concentration levels. In this study, we developed LUR models for street-level fine particulate matter ($PM_{2.5}$) concentration levels in Seoul, South Korea. 169 h of data were collected from an approximately three week long campaign across five routes by ten volunteers sharing seven AirBeams, a low-cost ($250 per unit), smartphone-based particle counter, while geospatial data were extracted from OpenStreetMap, an open-source and crowd-generated geographical dataset. We applied and compared three statistical approaches in constructing the LUR models – linear regression (LR), random forest (RF), and stacked ensemble (SE) combining multiple machine learning algorithms – which resulted in cross-validation R^2 values of 0.63, 0.73, and 0.80, respectively, and identification of several pollution 'hotspots.' The high R^2 values suggest that study designs employing mobile sampling in conjunction with multiple low-cost air quality monitors could be applied to characterize urban street-level air quality with high spatial resolution, and that machine learning models could further improve model performance. Given this study design's cost-effectiveness and ease of implementation, similar approaches may be especially suitable for citizen science and community-based endeavors, or in regions bereft of air quality data and preexisting air monitoring networks, such as developing countries.

1. Introduction

Ambient air pollution is a major global public health concern, with the World Health Organization estimating that 4.2 million premature deaths annually are attributable to fine particulate matter ($PM_{2.5}$) exposure (WHO, 2018). Government and regulatory agencies throughout the world have traditionally relied on networks of fixed-site monitors in order to measure air quality and establish standards. Owing to their prohibitive equipment and operational costs, these monitors tend to be sparsely located even in large metropolitan cities, or may be entirely missing in many locales. However, as concentrations of air pollutants can vary markedly over small distances and short time periods, the urban environment cannot be fully characterized using information from sparse, static networks of air pollution monitors (Kumar et al., 2015). To empirically model and characterize the spatial or spatiotemporal variability of $PM_{2.5}$ concentrations, land use regression (LUR) models based on data from monitoring networks have been employed. Recently, LUR models based on data collected from mobile sampling designs – where predetermined locations or routes are repeatedly sampled on modes of transport – have gained traction, offering improved spatial resolution at a lower cost (e.g., Hankey and Marshall, 2015; Shi et al., 2016; Deville Cavellin et al., 2016).

Recent technological advancements and proliferation of air quality sensors offer additional avenues to refine the spatiotemporal characterization of air pollution levels (Morawska et al., 2018. Numerous instruments from commercial entities, non-profits, and startups have entered the market to date (Borghi et al., 2017; McKercher et al., 2017), although the performance of these sensors can differ substantially

* Corresponding author at: Department of Environmental Medicine, New York University School of Medicine, 341 East 25th Street, New York, NY 10010, United States of America.
E-mail address: ccl414@nyu.edu (C.C. Lim).

https://doi.org/10.1016/j.envint.2019.105022
Received 1 March 2019; Received in revised form 26 June 2019; Accepted 15 July 2019
Available online 27 July 2019

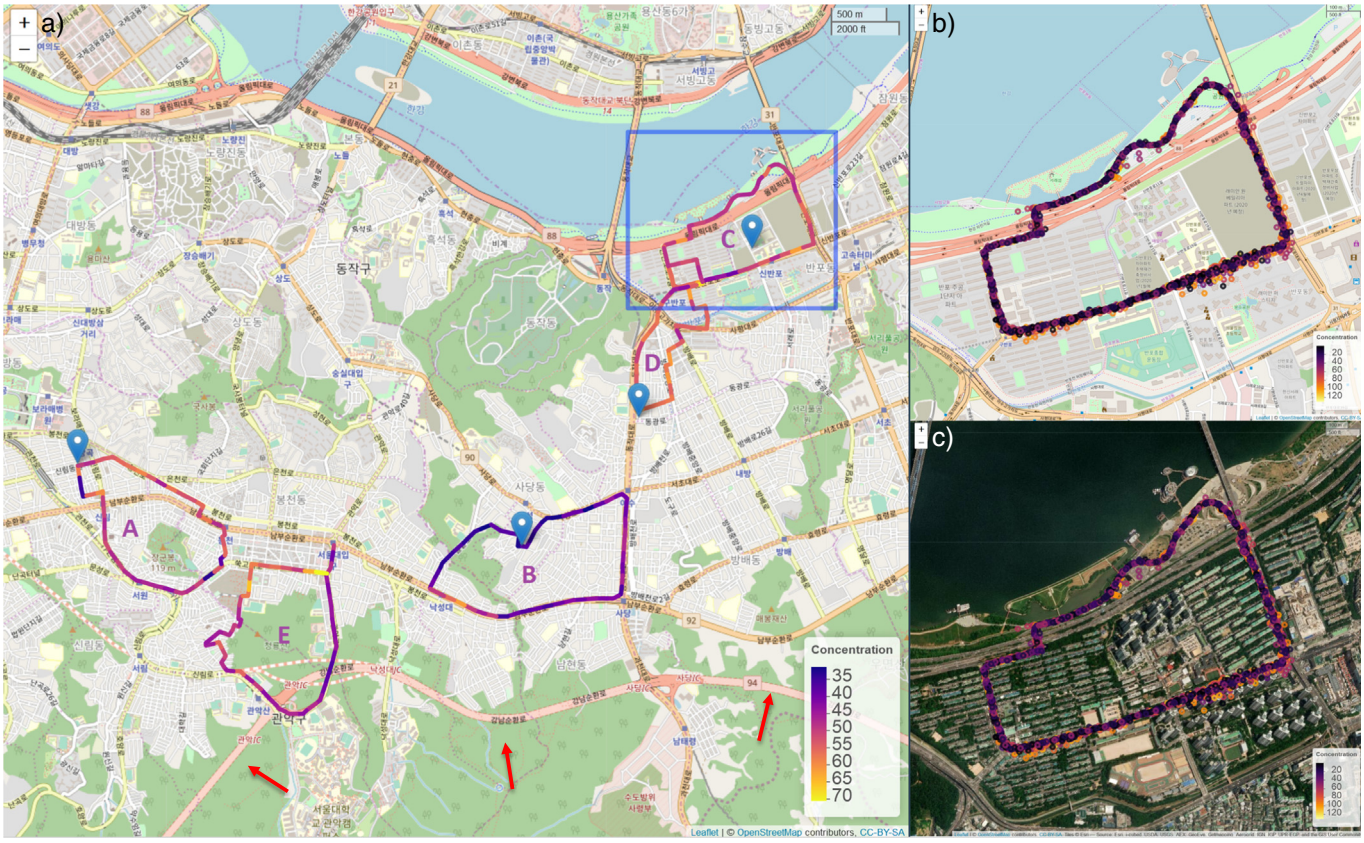

Fig. 1. (a) Locations of the five sampling routes in Seoul and government-run, fixed-site monitors (blue markers). Mean $PM_{2.5}$ concentration levels (μg/m^3) during the sampling period at each of the 100 m segments are also depicted. The red arrows point to an underground roadway, which was not included in analyses. We also present close-up views of route C as an example to depict sampled data points, with (b) OpenStreetMap and (c) satellite backgrounds. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

between the different models as well as between individual units, as noted by evaluations in field and laboratory settings (Jiao et al., 2016; Jerrett et al., 2017; Castell et al., 2017; Kelly et al., 2017; Feinberg et al., 2018; Levy Zamora et al., 2019). Offering the capability to inexpensively generate a large volume of data, distributed networks of low-cost air quality sensors are beginning to be established to augment existing monitoring networks or provide novel real-time data streams (Gao et al., 2015; Schneider et al., 2017; Zikova et al., 2017). Noteworthy examples of collaborative endeavors between government agencies, research organizations, and communities include: 'OpenSense' in Geneva, Switzerland (Hasenfratz et al., 2015), 'Array of Things' in Chicago, U.S (Catlett et al., 2017), and the Imperial County Community Air Monitoring Network (English et al., 2017) in California, U.S.

LUR models based on data collected from mobile sampling with low-cost (< $300) consumer-based sensors are very limited thus far, which could potentially offer a highly cost-effective approach to model and map air pollution concentration levels. The main aim of this study was to deploy multiple units of the smartphone-based particle counter 'AirBeam' to measure and model street-level urban air quality in Seoul, South Korea, a location with limited fixed regulatory monitoring sites relative to the high population and diverse urban environments. The individual AirBeam units were first collocated with a pDR-1500 within a laboratory setting to adjust for intra-instrument variability and equate particle counts to mass equivalents, and a mobile sampling campaign was conducted by repeatedly walking across five routes during an approximately three-week period. The collected air pollution data, together with an openly available and crowd-sourced geographical data source OpenStreetMap (OSM), were then used to construct LUR models with both linear regression and machine learning methods. This work explores the potential of mobile sampling with low-cost air quality sensors, machine learning models, and 'open data' sources to characterize street-level air quality in urban locations with fine spatial resolution.

2. Materials and methods

2.1. Equipment description and intra-instrument variability adjustment

The internal optical particle sensor of the AirBeam (dimensions: 105 × 95 × 43.5 mm; weight: 198 g) is the PPD60PV-T2 (detectable particle range: 0 to 400 μg/m^3; detective particle size 0.5–2.5 μm) from Shinyei Technology Co. LTD. (Kyoto, Japan), connected to an Android OX smartphone running the AirCasting application (aircasting.org). Supplemental Fig. 1 depicts the AirBeam, its specifications, and the Android AirCasting app. This mobile system is capable of continuous measurement (programmable intervals as little as per 1 s) and mapping (by GPS and Google Maps). The platform code is open-source, and collected data can be shared and mapped via an online platform, 'Aircasting' (www.aircasting.org/map).

To adjust for potential intra-instrument variability and to convert particle counts to $PM_{2.5}$ mass equivalents, the AirBeam units were collocated with a DataRAM pDR-1500 (Thermo Scientific, Franklin, MA) within a concentrated air particle (CAP) system in Sterling Forest, New York. The system draws in and concentrates ambient air through a cyclone inlet that first removes most of the particles larger than 2.5 μm in aerodynamic diameter. The cyclone outflow is passed over the warm bath of water and is then rapidly cooled in the condenser, resulting in supersaturation and particle growth (Maciejczyk et al., 2005). The pDR-1500 was initially calibrated with ambient particles via its internal

gravimetric filter and pump system at a flow rate of 1.5 L/min in the CAP chamber. The individual AirBeam units were then calibrated with the pDR-1500; first, the individual AirBeam units were placed within the CAPS chamber together with the pDR-1500 and tested for approximately 3 to 4 h periods per day and between 2 and 3 days per unit, and separate linear regression models were then fit for each unit.

2.2. Sampling location and protocol

Seoul, the capital of South Korea and the 5th most populous metropolitan area in the world, experiences one of the highest air pollution concentration levels among cities in developed countries. The city is characterized by extremely high urban density, abundance of high-rise buildings and apartments, and a mountainous terrain. This study was carried out in the southern part of Seoul, south of the Han River, in three districts: Dongjak-gu (area = 16.35 km^2; population density = 24,000/km^2), Seocho-gu (area = 47.14 km^2; density = 8300/km^2), and Gwanak-gu (area = 29.57 km^2; density = 18,000/km^2). The sampling campaign was conducted during an approximately three-week period (July 23rd to August 11th) in the summer of 2015, on weekdays only (12 days total, on non-rainy days) during three different time periods: morning (8–10 am), evening (6–8 pm), and night (9–11 pm). Ten volunteers sharing 7 AirBeam units were instructed to repeatedly sample the five routes without predetermined beginning/ending locations and times.

The five routes (Fig. 1), four of which were based near or around government-run regulatory monitors, were designed to span various neighborhoods and to obtain spatial coverage of a wide range of types of geographical variables, such as major roads and highways, green spaces, and both low and high density residential areas. Route A is located in Sillim; the neighborhood is largely residential with low-rise buildings and houses. Route B is in Sadang, which is also mainly residential with a large park and three major roads that surround the neighborhood. Route C is in Seocho, where the central bus transport terminal for Seoul is located, as well as the main city highway, a riverside park, and high-rise apartment buildings. Route D is located at Isu, where major highways and high-density residential areas are present. Route E is located near Seoul National University, a large university campus located at the base of a mountain; the area is hilly and tree-covered, and has a relatively low volume of traffic, mainly consisting of buses used for student transport. The lengths of the routes ranged from 3.9 km to 4.9 km, and the total sum length of all the routes was 21.5 km.

2.3. Data source for land use predictors

Geospatial data for the city of Seoul, South Korea were downloaded from OpenStreetMap (OSM), a freely available, crowd-sourced and user-generated online mapping system. The dataset included > 60 variables, grouped by the following categories: roads (cycleway, footway, living, path, pedestrian, residential, primary, secondary, road, secondary link, service, steps, subway, tertiary, trunk, trunk link, unclassified); land use (cemetery, farm, footway, forest, garden, golf, grass, hospital, island, park, parking, pitch, place of worship, playground, residential, school, sports center, substation, university, wood); buildings (apartments, cathedral, church, commercial, hospital, hotel, house, public, residential, retail, school, university, identified/unidentified), public amenities (fire station, fuel station, hospital, library, police, school, town hall); transportation points (bus stop, motorway junction, station, subway entrance); and water areas and waterways (stream, river, riverbank, water). Several variables in different categories that repeatedly describe the same land use morphology – e.g. "university", which is counted as land use, buildings, and public amenities – were all initially included in the analysis. After removing the subway variable (as it describes underground paths), there were 67 predictor variables available for analysis (Supplemental Table 1).

2.4. Data reduction

As the frequency of data collection was in 1-second intervals, the data points were first aggregated into 1-min averages to match the pDR-1500 sampling frequency and to reduce data noise. Measurement points with obvious GPS (e.g. located in middle of rivers) and sampling errors (e.g. volunteer did not follow sampling route properly) were removed by restricting data points to < 50 M away from the routes and also by manually after visual inspection. We then employed a "snapping" procedure to assign the collected data points to the nearest route segment on the basis of measured GPS coordinates to allow measurements along the same segment to be analyzed as a group, as per previous mobile LUR studies (Hankey and Marshall, 2015). Segments were first defined by length from a starting point along a route, and buffers with different radiuses were drawn around centroids of the route segments, with geospatial data from OSM within the buffers then extracted. Each road segment was thereby associated with land use, built, and natural environment variables, calculated as different OSM variables within the buffers of different sizes. We calculated road segments at 5 different lengths (25 M, 50 M, 100 M, 150 M, 250 M) and 5 buffer radiuses (50 M, 100 M, 150 M, 350 M, 500 M) in order to build the LUR models as well as to assess how these parameters influence the LUR model performance.

2.5. Adjustment for background temporal trends

Previous mobile sampling investigations adjusted for potential temporal bias through several approaches; for example, Tessum et al. (2018) adjusted for between-day temporal trends by subtracting the daily fifth percentile from all measured concentration values on a given day. Deville Cavellin et al. (2016) used linear and quadratic terms for temperature as independent variables in the model as adjustment for potential temporal variability. We modified an approach applied by multiple studies (Larson et al., 2009; Dons et al., 2012; Clougherty et al., 2013; Van den Bossche et al., 2015; Apte et al., 2017) that used background concentration levels from a nearby regulatory monitor to adjust for temporal trends and normalize measured values. Leveraging the available information on background $PM_{2.5}$ concentrations from multiple fixed-site regulatory monitors nearby the sampling routes, we adjusted each 1-min averaged measurements from AirBeams for each day by applying a multiplicative hourly factor (defined as the ratio of mean concentration level during the entire sampling period to corresponding hour in which that measurement is taken) derived from the nearby regulatory monitor. For route E, which was not designed around a regulatory monitor, we used averaged values from the two nearby monitors (approx. 2–4 km away) located by routes A and B. This resulted in 6 factors per each sampling day for each of the 5 routes. Using multiple nearby monitors, instead of a single monitor as done in past studies, allowed for variable temporal adjustments across several locations. This approach minimizes the effect of day-to-day variations in background air quality on the measurements, thereby decreasing the amount of required sampling data (Van den Bossche et al., 2015). Hourly measurements from regulatory monitors in Seoul revealed considerable temporal variability during the study period, with hourly $PM_{2.5}$ levels as low as 5 μg/m^3 and reaching 67 μg/m^3 during pollution episodes (Fig. 2).

2.6. LUR model building

We first tested the potential effects of spatial aggregation by different route segment lengths and buffer sizes in the linear regression model by including all available 67 variables into a linear regression model, and we selected 100 m route segments to spatially aggregate the collected data points based on the high adj-R^2, resulting in 215 available segments for subsequent analyses. We then applied and compared three statistical approaches for building the LUR model: linear

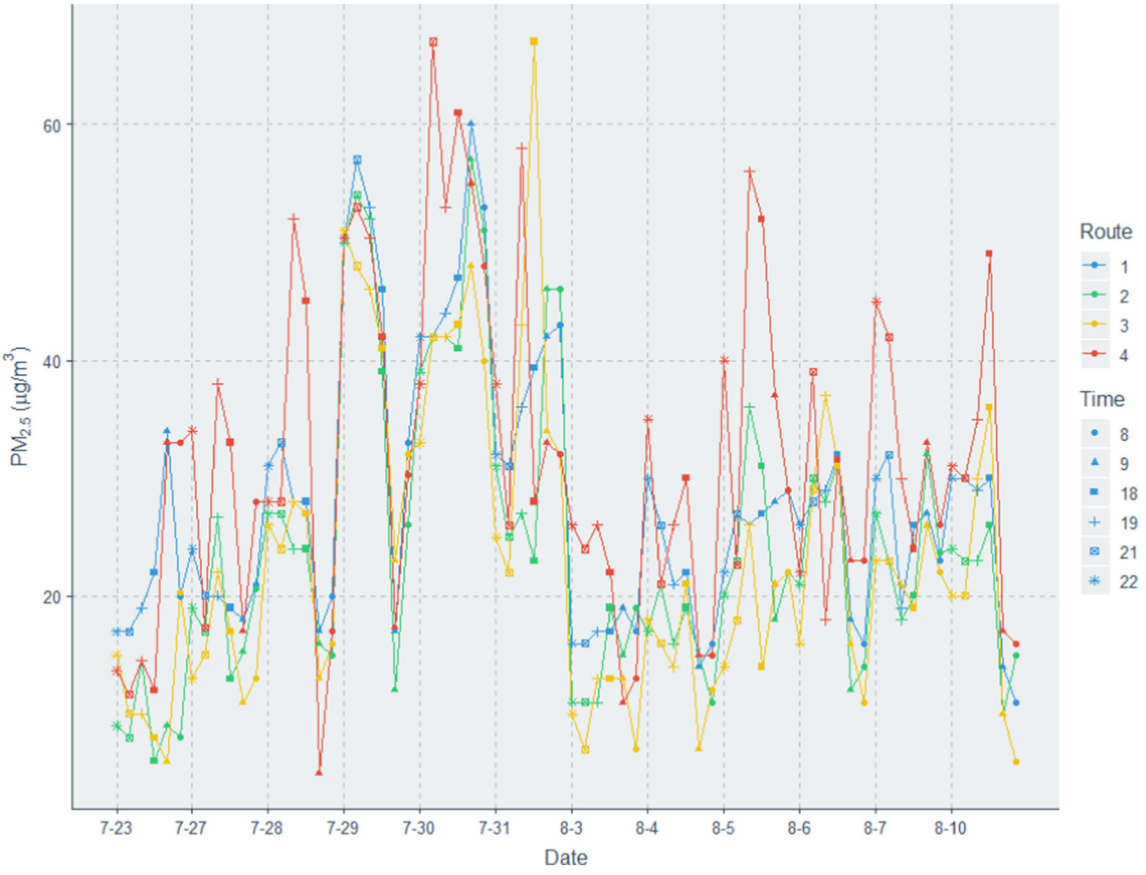

Fig. 2. Hourly (at 8 am, 9 am, 6 pm, 7 pm, 9 pm, 10 pm) $PM_{2.5}$ concentration levels during the sampling period (7/23/15 to 8/10/15) at the four regulatory background monitors.

regression (LR), random forest (RF), and stacked ensemble (SE).

In the linear regression model, the GIS variables were retained for multivariable models based on a distance-decay regression selection strategy (ADDRESS) to screen and select informative candidate variables and corresponding buffer size from all of the available potential variables (Su et al., 2015). We then applied a supervised forward search approach, adding the variables one at a time in the LR model and keeping the variable only if it increased the R^2 of the model by 1.0% and if all predictor variables have statistically significant coefficients ($p < 0.05$) (Van den Bossche et al., 2018). We also applied the random forest (RF) model after first removing highly correlated variables (absolute correlation > 0.8). Random forests, in brief, are an ensemble of decision trees and each tree is constructed using the best split for each node among a subset of predictors randomly chosen. Random search, which randomly chooses combination of hyperparameters at every iteration, was used to tune and optimize the model (Bergstra and Bengio, 2012). Finally, we employed the stacked ensemble (SE) model, a machine learning ensemble approach that involves training a learning algorithm to combine the predictions of several other learning algorithms; first, all of the other algorithms are trained using the available data, then a 'meta-classifier' algorithm (chosen from the list of algorithms) is trained to make a final prediction combine all the predictions of the other algorithms as additional inputs. We evaluated and selected a diverse group of machine learning algorithms, including random forest ('rf'), Bayesian generalized linear model ('bayesglm'), k-nearest neighbors ('knn'), recursive partitioning and regression trees ('rpart'), and partitioning using deletion, substitution, and addition moves ('partDSA').

We applied 10-fold cross validation (with 500 repeats) to calculate mean CV-R^2 (cross-validation R^2; 1-(mean square error/variance)) and root mean square errors (RMSE; a measure of the differences between values predicted by a model and the values observed) for the three methods to quantify their accuracy. We used packages 'ggplot2' and 'leaflet' for visualization and 'caret' for statistical analyses in R (version 3.4.4).

3. Results

3.1. Adjustment for intra-instrument variability

We fit univariate linear regression models for each of the deployed Airbeam unit in order to adjust for intra-unit variability and to convert particle counts to $PM_{2.5}$ mass concentrations. During the collocated sessions with the DataRAM pDR-1500 in the CAP chamber, the $PM_{2.5}$ concentration (as measured by pDR-1500) ranged from 0 to 81 µg/m^3. The AirBeams revealed strong agreements with the pDR-1500 (adj-R^2 = 0.95–0.98) and noticeable differences in responses between the individual units (Fig. 3). The regression models' intercepts, slopes, and RMSE values varied across the units; detailed statistical summaries of the models are presented in Table 1.

3.2. Mobile sampling summary statistics

The mobile sampling campaign yielded a total of 10,871 min of data, of which after removing GPS and sampling errors, 10,177 min (93.6%) of data remained, equaling > 169 h of total data across the 5 sampled routes (Table 2, Supplemental Tables 2 & 3). 1992 min (33.2 h) of sampling data were collected at Route A; 2449 min (40.8 h) at Route B; 2313 min (38.6 h) at Route C; 1970 min (32.8 h) at Route D; and 1453 min (24.2 h) at Route E. Route D, which is located near major roads and highways, had the highest concentration levels (55.5 ± 27.7 µg/m^3), while Route B (42.0 ± 24.2 µg/m^3) and Route E

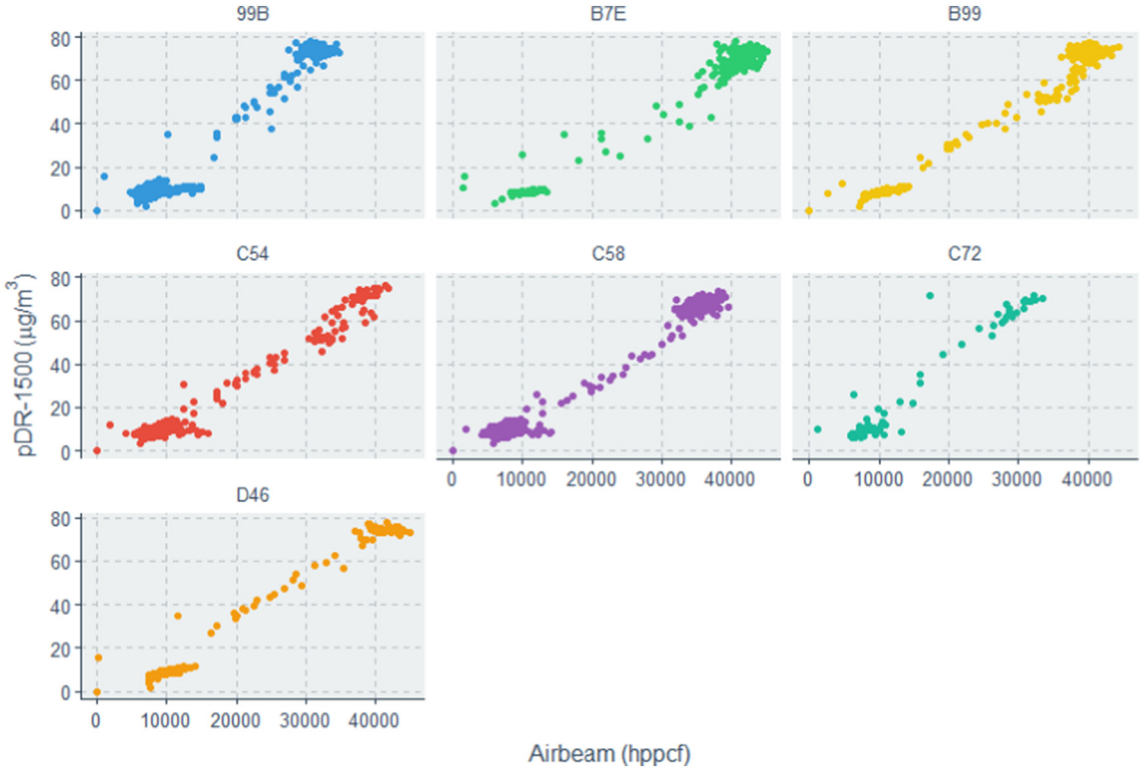

Fig. 3. DataRam pDR-1500 (mass; µg/m^3) vs. 1-minute averaged AirBeam (hundreds of particles per cubic feet; hppcf) measurements in the concentrated air particle chamber (CAP).

(48.4 ± 31.3 µg/m^3) had the lowest concentration levels. Notable differences between morning, evening, and night were also observed across the five routes, especially for Route D, which had elevated levels during morning (70.7 ± 25.5 µg/m^3) compared to evening (46.6 ± 28.3 µg/m^3) and night (54.8 ± 24.1 µg/m^3). The amount of sampling data varied across the 215 segments, with a median of 44 min per segment (minimum = 5; 25% percentile = 34; 75% percentile = 55; maximum = 179). Summary statistics for minutes of sampling per 100 m segment for each of the five routes are visualized as boxplots in Fig. 4.

3.3. Model results

The LUR models were sensitive to different segment lengths and buffer radiuses, with adj-R^2 generally increasing with larger buffer radiuses (Fig. 5), while 100 m to 150 m segments for spatial aggregation performed the best. Fitting individual equations to account for intra-instrument variability for each AirBeam unit generally improved the accuracy of the constructed LUR models, with an increase in CV-R^2 values by ~0.10–0.15.

In constructing the LR model, we screened and removed several point variables (e.g. fire stations) that were not frequently present across the sampling space but clustered near the pollution hotspots, as these variables ended up having very strong influences on the models. The final LR LUR model showed high goodness-of-fit with a CV-R^2 of 0.63 and RMSE of 7.01, and the following variables were included in the model: wood, secondary link, residential road, cathedral, station, pitch, and apartments (Table 3). The machine learning approaches explained a greater proportion of the variance of $PM_{2.5}$ concentrations than the LR model. The random forest model identified mostly different variables as important (wood, residential road, living street, school, park, apartments, residential, building, tertiary, and service) and also revealed better performance metrics compared to the LR model, with higher mean CV-R^2 (0.73) and lower RMSE (6.20). The stacked ensemble model with random forest as the meta-predictor algorithm performed the best, and the SE model outperformed both LR and RF models, with higher CV-R^2 (0.80) and lower RMSE (5.22). Individual R^2 values for the algorithms in the ensemble were 0.74 for random forest, 0.45 for partDSA, 0.50 for rpart, 0.70 for bayesglm, and 0.69 for knn.

Adjusting for background temporal trends changed the overall morning average concentration levels from 49.4 to 59.2 µg/m^3; evening from 46.4 to 45.7 µg/m^3; and night from 51.5 to 47.3 µg/m^3. The changes in concentration levels after temporal adjustment during the three sampling periods differed significantly across the routes (Supplemental Table 4). This adjustment also improved the CV-R^2 for the three approaches, as not doing so resulted in lower CV-R^2 values of

Table 1
Linear regression equations to convert particle counts to mass for each of the AirBeam unit.

Unit name	Intercepts (standard error)	Slope (standard error)	RMSE	Adj-R^2
99B	−10.72 (0.29)	0.002616 (1.77 × 10–5)	23.48	0.95
B7E	−11.69 (0.44)	0.001974 (1.41 × 10–5)	14.84	0.98
B99	−13.16 (0.42)	0.002102 (1.47 × 10–5)	15.50	0.98
C54	−6.68 (0.18)	0.001905 (1.35 × 10–5)	8.92	0.96
C58	−4.91 (0.16)	0.002000 (1.05 × 10–5)	9.16	0.98
C72	−9.94 (0.34)	0.002537 (3.17 × 10–5)	10.50	0.95
D46	−11.26 (0.37)	0.002049 (1.64 × 10–5)	11.39	0.98

Table 2
Summary statistics for measurements across the five routes.

Route ID	Route name	AirBeam units deployed	Total			Morning			Evening			Night		
			Minutes sampled	Average (Std. Dev), μg/m^3	IQR	Minutes sampled	Average (Std. Dev), μg/m^3	IQR	Minutes sampled	Average (Std. Dev), μg/m^3	IQR	Minutes sampled	Average (Std. Dev), μg/m^3	IQR
A	Sillim	B7E, B99, C54, D46	1992	51.3 (32.6)	56.3	901	43.3 (32.8)	40.4	462	58.6 (27.9)	49.6	629	57.4 (33.0)	47.4
B	Sadang	99B, B7E, C58, D46	2449	42.0 (24.2)	36.5	744	40.5 (24.8)	34.2	858	38.6 (21.2)	38.1	847	46.7 (25.8)	30.3
C	Seocho	99B, C58, C72	2313	49.9 (31.4)	50.3	574	47.9 (35.2)	69.2	892	46.7 (29.4)	43.3	847	54.5 (30.1)	38.1
D	Isu	99B, C72, D46	1970	55.5 (27.7)	33.9	477	70.7 (25.5)	39.6	755	46.6 (28.3)	43.3	738	54.8 (24.1)	26.1
E	Seoul National University	B7E, B99, C54, D46	1453	48.4 (31.3)	48.4	396	56.8 (39.4)	74.3	528	47.4 (26.5)	23.2	529	43.0 (27.4)	54.0

0.54, 0.65, and 0.71 for the LR, RF, and SE models, respectively. The constructed LUR models were used to create prediction maps of street-level $PM_{2.5}$ concentration levels in Seoul nearby the sampled locations, which revealed several 'hotspots' with elevated $PM_{2.5}$ levels (Fig. 6). The prediction maps revealed similar spatial patterns between the three modeling approaches with emphasis on similar locations as hotspots, especially at locations with major roads/highways and high population density. Conversely, the lowest concentrations were predicted at greenspace locations, such as parks and mountains. The three approaches resulted in relatively similar mean predicted values across the exposure surface, at 47.31, 48.86, and 49.43 μg/m^3, for LR, RF, and SE, respectively. However, the LR prediction map predicted lower values than machine learning approaches at the extremes (range: 26.36–68.96 μg/m^3), while maps for RF (34.97–71.43 μg/m^3) and especially SE (33.50–83.19 μg/m^3) models resulted in higher predicted values.

4. Discussion

In this study, we conducted a mobile sampling campaign in Seoul, South Korea deploying low-cost smartphone-based air quality sensors and utilized the collected data to construct LUR models employing three statistical approaches. The strengths of the resulting R^2 values were comparable to recent, similar studies across multiple locations around the world that utilized more advanced equipment. Our study is unique for developing LUR models using multiple low-cost (< $300), mobile sensors; priced at $250 per unit, AirBeams are order(s) of magnitude less expensive than the commercially available portable (in the thousands; the pDR-1500 used in this study cost ~$5700) and federal standard (in the tens of thousands) instruments. AirBeam and its operating platform, Aircasting, is also notable for being primarily developed for citizen science whereby users can upload their measurements to share with the public, as well as for being open-sourced, allowing developers and researchers to program and customize the instruments and the smartphone app according to their needs and requirements. Many similarly priced ($200–$300) sensors have entered the market since the present study was conducted, underlining the public's increasing interest in the capability to measure personalized real-time exposure data (Caplin et al., 2019). Through deployment of such low-cost sensors, we were able to characterize the spatial variability of street-level $PM_{2.5}$ in Seoul, the main source of which is likely to be from traffic given the near-road sampling approach applied in this study. Past source apportionment studies also identified the primary source of $PM_{2.5}$ in Seoul as motor vehicle emissions and road dust (Heo et al., 2008; Ryou et al., 2018).

Recent mobile sampling approaches for LUR model building have employed a variety of study designs and instruments. For example, Hankey and Marshall (2015) collected over 85 h of data on a bicycle-based sampling platform in Minneapolis, MN and constructed LUR models for particle size, black carbon, and $PM_{2.5}$ with modest goodness-of-fit (adj-R^2 of ~0.5 for particle number and ~0.4 for $PM_{2.5}$). Apte et al. (2017) analyzed data collected from a Google Street View mapping vehicle equipped with air quality sensors that repeatedly sampled every street in a 30-km^2 area of Oakland, CA, to model and reveal urban air pollution patterns at 4–5 orders of magnitude greater spatial precision than possible with current central-site ambient monitoring. The 'OpenSense' project in Zurich, Switzerland (Hasenfratz et al., 2015) utilized mobile sensor nodes installed on top of public transport tram vehicles in the city to create high-resolution pollution prediction maps for ultrafine particles and particle counts. Vehicle-based mobile measurements were also applied to create LUR models to estimate the spatial variation of street-level $PM_{2.5}$ and PM_{10} in the downtown area of Hong Kong (Shi et al., 2016), and integration of urban/building morphology as independent variables increased the adj-R^2 of the LUR model, suggesting that incorporating detailed 3D characteristics of the land use can improve the predictive power of such models.

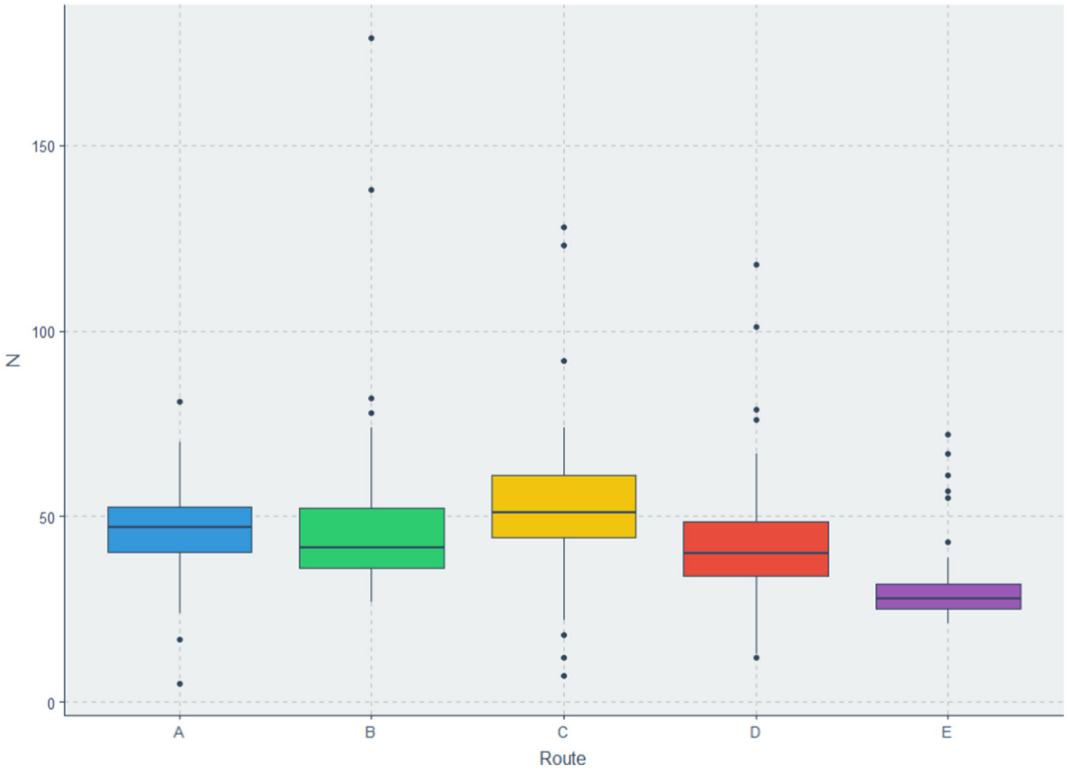

Fig. 4. Boxplot demonstrating distribution of minutes of sampling per 100 m segment for each sampling route.

Our study and sampling design highlight the potential advantages of mobile sampling with low-cost and portable air quality sensors in constructing the LUR models. The aforementioned studies were largely based on sampling campaigns conducted on modes of transport (e.g. cars) visiting a single location at a given time, which may potentially result in a low number of visits per location. The results from this and past studies found that mobile LUR models are highly sensitive to parameters such as the number of route segments, radiuses of buffers, and number of measurements per segment (Minet et al., 2017). Hatzopoulou et al. (2017) evaluated the influence of the number of sampling locations and durations of sampling on LUR model performance, noting that mobile sampling campaigns can be inefficient due to low sampling frequency at a large number of locations, and that spatial variability may be more important than the numbers of locations when designing the sampling routes. The authors also found that the LUR models became relatively robust after 150–200 segments and 10–12 visits per segment. In the present study, walking at a slow speed, instead of sampling on mechanical modes of transportation, resulted in each route generally having a high number of data points (median = 44) per segment. This approach also allows for assessing personal-level exposure in urban areas where there are a larger number of people on the streets than in cars. The disadvantage of shorter distances being covered when sampling on foot was offset by the low cost and portability of AirBeams, which allowed for several units that could be deployed simultaneously across multiple locations at a given time and thereby maximize spatial coverage, as opposed to the majority of past mobile sampling studies that were carried out on a single platform. Simultaneous measurements within a structured sampling design could

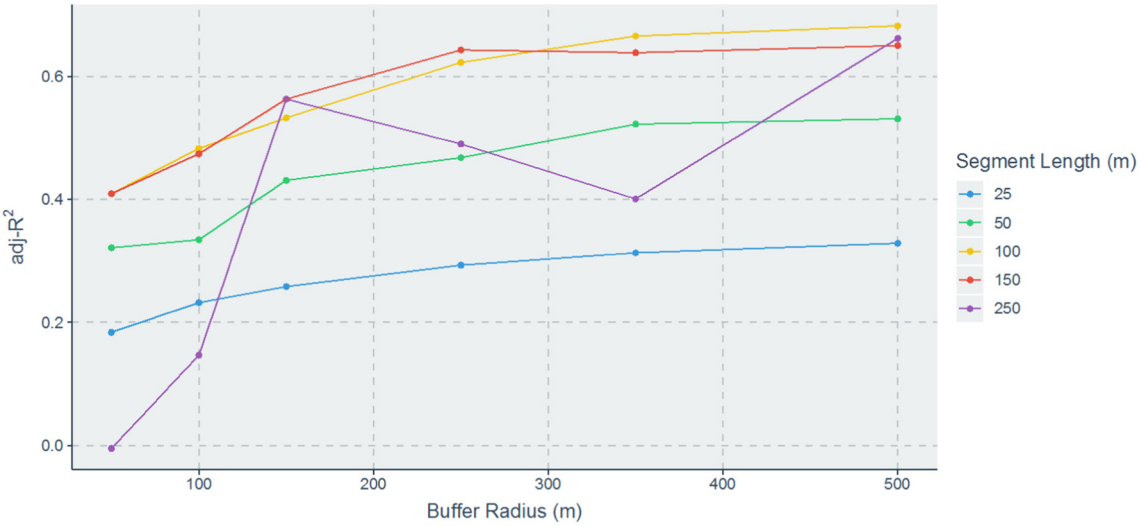

Fig. 5. Adjusted R^2 of LR LUR models (including all available 67 predictor variables) for mass, by segment radius and buffer sizes.

Table 3
Selected LUR model predictor variables in the LR and RF models and associated statistics.

Variable name	Variable type	Buffer length	Linear regression			Random forest[a]
			B	Std. error	*P*-value	Importance
Intercept			50.02	1.21	< 0.001	
Wood	Area	500 m	-3.80×10^{-5}	4.25×10^{-6}	< 0.001	14.45
Residential Road	Line	500 m	2.59×10^{-5}	5.88×10^{-6}	< 0.001	13.10
Secondary Link	Line	500 m	6.88×10^{-3}	1.00×10^{-3}	< 0.001	
Cathedral	Point	500 m	-2.47×10^{-3}	1.03×10^{-3}	0.02	
Station	Point	500 m	−3.75	1.02	< 0.001	
Pitch	Area	350 m	-1.88×10^{-4}	4.28×10^{-5}	< 0.001	
Apartments	Point	500 m	7.70×10^{-5}	4.02×10^{-5}	0.05	10.21
School	Area	500 m				10.73
Living Street	Line	500 m				10.85
Park	Area	500 m				10.31
Residential	Area	500 m				9.93
Building (Unclassified)	Area	500 m				9.72
Tertiary	Line	350 m				9.70
Service	Line	350 m				8.97

[a] Top ten variables by variable importance are shown in the table.

decrease the amount of collected data (and manpower) required to construct robust models, whereas participatory sensing where sampling is done 'opportunistically' could lead to unstructured data that is more difficult to interpret (Van den Bossche et al., 2016). Furthermore, AirBeam's ease of operation meant that minimal training (a few minutes at most) was required prior to field deployment, resulting in a relatively large volume of data being generated within the short sampling campaign period during this study.

This study leveraged OpenStreetMap (OSM), an openly available and crowd-sourced GIS dataset, which provided a rich and comprehensive source of geospatial data for a wide range of LUR variables. OSM and other 'open data' sources offer underexplored but valuable information for data-driven methods to predict air pollution levels (VoPham et al., 2018). Notably, the OSM GIS variables were highly developed for Seoul and provided detailed and differentiated data for the numerous types of roads and buildings, which are the land use categories that usually provide the highest predictive power for air pollution LUR models. Another advantageous aspect of crowd-sourced data is that it is continually updated; for example, using an earlier download of OSM from September 2015 (versus January 2018 in this analysis) with less developed characterization of Seoul resulted inLUR models with lower CV-R^2, suggesting that in locations with lacking geospatial data, crowd-sourced efforts to generate the relevant GIS variables could be carried out in concert with the air pollution sampling campaign to strengthen the predictive capability of LUR models. Despite recent endeavors to democratize data by agencies and organizations throughout the world as part of the 'open data' movement, many detailed GIS files remain proprietary and thereby cost-prohibitive, and freely available data like OSM offer an alternative and important source of detailed spatial data for researchers and communities.

Machine learning methods offered improved goodness-of-fit compared to traditional stepwise linear regression in constructing the LUR models. Prior work on machine learning applications in both national (Hu et al., 2017; Di et al., 2016) and local-level (Adams and Kanaroglou, 2016; Weichenthal et al., 2016; Brokamp et al., 2017) predictions of air pollution concentration levels highlight the advantages associated with the approach, including higher accuracy and identification of important variables. A recent example further underlines additional potential benefits; a study in Los Angeles, USA used a multi-step and flexible spatial data mining approach using machine learning to select for most important OSM geographic features and predict $PM_{2.5}$ concentrations, removing the need for a priori selection of predictors for exposure modeling (Lin et al., 2017). Similarly in our analysis, applying the traditional step-wise linear regression LUR approach with the highly correlated OSM dataset, which also contained several highly influential variables, required manual screening and removal of predictor variables prior to input and during the model building process. Notably, the stacked ensemble model combining multiple machine learning algorithms outperformed both LR and RF in this study. In recent years ensemble machine learning methods have emerged as an important tool for modeling complex relationships and have been applied successfully in various research areas (Yang et al., 2010). Application of ensembles have been generally limited in air pollution exposure assessment and modeling efforts to date, and the results here suggest that ensemble-based approaches could further enhance the predictive performance of LUR models.

We note several potential weaknesses that are present in this study. As we evaluated the AirBeam units in a carefully controlled experimental chamber drawing in air from a forested and rural area (Tuxedo, New York), the particle composition and the environmental conditions (e.g., humidity and temperature) encountered during the experiment are likely to be significantly different from the heavily urban location where this study was carried out. Although the potential impacts of these factors were not assessed in this study, previous performance evaluations of AirBeams in various laboratory and field settings offer insight. The initial manufacturer calibration was conducted in a similarly urban setting (New York City), which revealed high correlations with both gravimetric sampling and pDR-1500 (takingspace.org, 2014). Comparison against federal equivalent method monitors showed high agreements with GRIMM (R^2~0.6–0.8) (Mukherjee et al., 2017; SCAQMD, 2017; Feinberg et al., 2018), but mixed results were observed with BAM (R^2~0.2–0.7) (Jiao et al., 2016; SCAQMD, 2017). A study of sensor responses to Arizona road dust, salt, and welding fumes (Sousan et al., 2017) demonstrated that particle types had significant impacts on AirBeam (and other low-cost sensors) measurements. Relative humidity (RH) levels also influenced the measurements; a laboratory evaluation found that bias was observed when both RH (> 65%) levels and concentration levels (> 100 μg/m^3) were elevated (SCAQMD, 2017), while another study (Feinberg et al., 2018) found that the particle counts measurements were affected by higher humidity levels in a field setting. Highly humid summers in Korea would likely influence the absolute measurement values, but the potential impact on prediction model performance is likely to be minimal as the spatial variability of humidity levels is relatively uniform across a city. Nevertheless, these findings emphasize the need to consider the potential influence of environmental factors in sensor deployments, and performance evaluations at the study location is suggested for similar studies applying low-cost sensors. In addition, the particle concentration levels encountered during sampling in Seoul were higher than the range used for constructing calibration equations for the AirBeam units, which may ignore

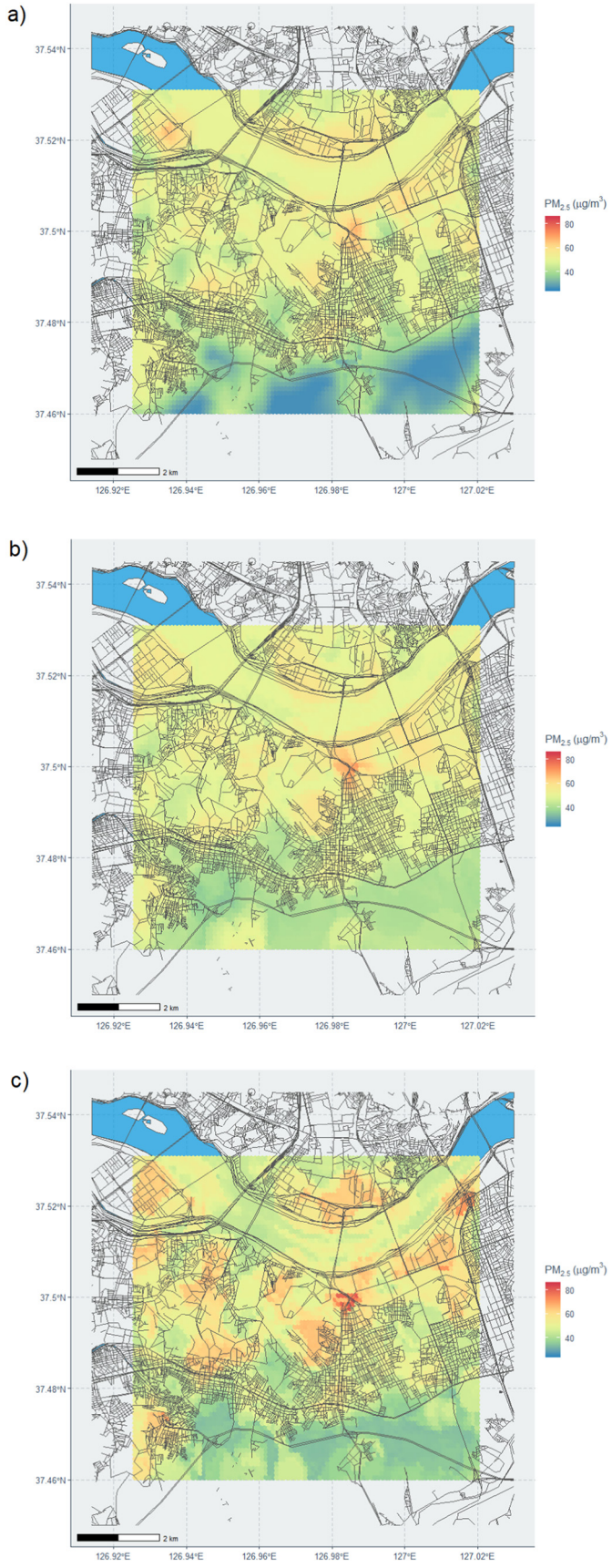

Fig. 6. $PM_{2.5}$ prediction maps nearby sampled areas constructed applying (a) linear regression, (b) random forest, and (c) stacked ensemble approaches.

the potential nonlinearity of sensor responses. We also did not check for potential sensor drift – a common issue for low-cost air quality sensors – during and after the mobile sampling, although this is unlikely due to the relatively short sampling period. These issues may have contributed to predicted values that were significantly higher than observed values from nearby fixed-site monitors, although it is also possible that such differences are due to the fact that fixed-site monitors are often located well above ground and tend to underestimate personal exposures when walking near traffic (Deville Cavellin et al., 2016). Another potential weakness is that the OSM data quality and density could be potentially uneven across locations, as some areas could be characterized in more detail than others. For example, in some of the sampled areas in this study, several of the houses in residential areas were not captured in the OSM file and thereby could have influenced model quality; however, as OSM data coverage and quality continues to improve this should become less of an issue over time.

5. Conclusions

Low-cost sensors represent an opportunity to bridge the data gap, thereby promoting public discourse, influencing air pollution regulations, and protecting public health (Amegah, 2018). This study highlights the advantages and potential of applying data collected from mobile sampling with multiple low-cost sensors to model and map street-level air pollution levels in urban locations, especially the capability to generate a large volume of sampling data with ease. The predictive power of models developed here, despite deploying only a limited number of significantly less expensive, consumer-based air quality sensors, were comparable to the past mobile sampling LUR studies, especially after adjusting for intra-instrument variability and temporal trends. To minimize the potential influence of local particle characteristics and environmental conditions, calibration with collocated reference monitors at the sampling location is suggested for future projects using similar low-cost sensors, as well as to convert particle counts to mass concentration, a unit of measurement that is more readily transferable for policy-relevant metrics. Initial calibrations should also carefully evaluate and adjust for the potential effects of relative humidity levels, which can have significant influences on the low-cost sensors. Overall, the findings here suggest that similar mobile sampling designs using low-cost sensors and 'open data' sources could be applied to generate a large volume of data and construct LUR models and maps with fine spatial granularity, and that machine learning methods could further improve model performance. Our study design and approach may be especially suitable for citizen science and community-based endeavors, or in locations without preexisting air monitoring networks, such as developing countries.

Acknowledgement

This study was funded by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the South Korea Ministry of Education (2018R1A2B6004608), the NSF East Asia and Pacific Summer Institute (EAPSI) Fellowship, the Air & Waste Management Air Pollution Education and Research Grant (APERG), the EPA STAR Graduate Fellowship, and by a grant from the National Institutes of Environmental Health Sciences Center (ES00260). This publication was developed under Assistance Agreement No. FP917825 awarded by the U.S. Environmental Protection Agency to Chris C. Lim. It has not been formally reviewed by EPA. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://

doi.org/10.1016/j.envint.2019.105022.

References

Adams, M.D., Kanaroglou, P.S., 2016. Mapping real-time air pollution health risk for environmental management: combining mobile and stationary air pollution monitoring with neural network models. J. Environ. Manag. 168, 133–141.

Amegah, A.K., 2018. Proliferation of Low-Cost Sensors. What Prospects for Air Pollution Epidemiologic Research in Sub-Saharan Africa? vol. 241. pp. 1132–1137.

Apte, J.S., Messier, K.P., Gani, S., Brauer, M., Kirchstetter, T.W., Lunden, M.M., ... Hamburg, S.P., 2017. High-resolution air pollution mapping with Google street view cars: exploiting big data. Environmental Science & Technology 6999–7008.

Bergstra, James, Bengio, Yoshua, 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13.Feb, 281–305.

Borghi, F., Spinazz, A., Rovelli, S., Campagnolo, D., Buono, L. Del, Cattaneo, A., & Cavallo, D. M. (2017). Miniaturized monitors for assessment of exposure to air pollutants: a review.

Brokamp, C., Jandarov, R., Rao, M.B., LeMasters, G., Ryan, P., 2017. Exposure assessment models for elemental components of particulate matter in an urban environment: a comparison of regression and random forest approaches. Atmos. Environ. 151, 1–11.

Caplin, A., Ghandehari, M., Lim, C., Glimcher, P., Thurston, G., 2019. Advancing environmental exposure assessment science to benefit society. Nat. Commun. 10 (1), 1236.

Castell, N., Dauge, F.R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Bartonova, A., 2017. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? Environ. Int. 99, 293–302.

Catlett, C.E., Beckman, P.H., Sankaran, R., Galvin, K.K., 2017. Array of things: a scientific research instrument in the public way: Platform design and early lessons learned. In: Proceedings of the 2nd International Workshop on Science of Smart City Operations and Platforms Engineering. ACM, pp. 26–33 April.

Clougherty, J.E., Kheirbek, I., Eisl, H.M., Ross, Z., Pezeshki, G., Gorczynski, J.E., Johnson, S., Markowitz, S., Kass, D., Matte, T., 2013. Intra-urban spatial variability in wintertime street-level concentrations of multiple combustion- related air pollutants: the New York City Community Air Survey (NYCCAS). J. Expo. Sci. Environ. Epidemiol. 23, 232e240.

Deville Cavellin, L., Weichenthal, S., Tack, R., Ragettli, M.S., Smargiassi, A., Hatzopoulou, M., 2016. Investigating the use of portable air pollution sensors to capture the spatial variability of traffic-related air pollution. Environmental Science & Technology 50 (1), 313–320.

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., Schwartz, J., 2016. Assessing PM 2.5 Exposures with High Spatiotemporal Resolution across the Continental United States. Environ Sci Technol. 50 (9), 21–4712.

Dons, E., Int Panis, L., Van Poppel, M., Theunis, J., Wets, G., 2012. Personal exposure to Black Carbon in transport microenvironments. Atmos. Environ. 55, 392–398.

English, P.B., Olmedo, L., Bejarano, E., Lugo, H., Murillo, E., Seto, E., Wong, M., King, G., Wilkie, A., Meltzer, D., Carvlin, G., 2017. The Imperial County Community Air Monitoring Network: a model for community-based environmental monitoring for public health action. Environ. Health Perspect. 125 (7).

Feinberg, S., et al., 2018. Long-term evaluation of air sensor technology under ambient conditions in Denver, Colorado. Atmos. Meas. Tech. 11, 4605–4615.

Gao, M., Cao, J., Seto, E., 2015. A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of $PM_{2.5}$ in Xi'an, China. Environ. Pollut. 199, 56–65.

Hankey, S., Marshall, JD., 2015. 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. 49 (15), 202–9194.

Hasenfratz, D., et al., 2015. Deriving high-resolution urban air pollution maps using mobile sensor nodes. Pervasive Mob. Comput. 16, 268–285.

Hatzopoulou, M., Valois, M.F., Levy, I., Mihele, C., Lu, G., Bagg, S., ... Brook, J., 2017. Robustness of land-use regression models developed from mobile air pollutant measurements. Environ. Sci. Technol. 51 (7), 3938–3947.

Heo, J.B., Hopke, P.K., Yi, S.M., 2008. Source apportionment of $PM_{2.5}$ in Seoul. Korea.Atmos. Chem. Phys. Discuss. 8, 20427e2046.

Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., & Strickland, M. J. (2017). Estimating $PM_{2.5}$ concentrations in the conterminous United States using the random forest approach.

Jerrett, M., et al., 2017. Validating novel air pollution sensors to improve exposure estimates for epidemiological analyses and citizen science. Environ. Res. 158, 286–294.

Jiao, W., et al., 2016. Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. Atmos. Meas. Tech. Discuss. 1–24.

Kelly, K.E., et al., 2017. Ambient and laboratory evaluation of a low-cost particulate matter sensor. Environ. Pollut. 221, 491–500. https://doi.org/10.1016/j.envpol.2016.12.039. Feb.

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L., Britter, R., 2015. The rise of low-cost sensing for managing air pollution in cities. Environ. Int. 75, 199–205.

Larson, T., Henderson, S.B., Brauer, M., 2009. Mobile Monitoring of Particle Light Absorption Coefficient in an Urban Area as a Basis for Land Use Regression. Environ Sci Technol. 43 (13), 8–4672.

Levy Zamora, Misti, et al., 2019. Field and laboratory evaluations of the low-cost plantower particulate matter sensor. Environ. Sci. Technol. 53 (2), 838–849 American Chemical Society.

Lin, Y., Chiang, Y.-Y., Pan, F., Stripelis, D., Ambite, J.L., Eckel, S.P., Habre, R., 2017. Mining public datasets for modeling intra-city $PM_{2.5}$ concentrations at a fine spatial resolution. In: Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, Los Angeles area, CA, pp. 1–10.

Maciejczyk, P., Zhong, M., Li, Q., Xiong, J., Nadziejko, C., Chen, L.C., 2005. Effects of subchronic exposures to concentrated ambient particles (CAPs) in mice: II. The design of a CAPs exposure system for biometric telemetry monitoring. Inhal. Toxicol. 17 (4–5), 189–197.

McKercher, G.R., Salmond, J.A., Vanos, J.K., 2017. Characteristics and applications of small, portable gaseous air pollution monitors. Environ. Pollut. https://doi.org/10.1016/j.envpol.2016.12.045.

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

Morawska, L., et al., 2018. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: how far have they gone? Environ. Int. 116, 286–299.

Mukherjee, Anondo, Stanton, Levi, Graham, Ashley, Roberts, Paul, August 5, 2017. Assessing the utility of low-cost particulate matter sensors over a 12-week period in the Cuyama Valley of California. Sensors 17 (8), 1805.

Ryou, H., Heo, J., Kim, S.Y., 2018. Source apportionment of PM10 and $PM_{2.5}$ air pollution, and possible impacts of study characteristics in South Korea. Environ. Pollut. 240, 963–972.

Schneider, P., Castell, N., Vogt, M., Dauge, F.R., Lahoz, W.A., Bartonova, A., 2017. Mapping urban air quality in near real-time using observations from low-cost sensors and model information. Environ. Int. 106 (December 2016), 234–247.

Shi, Y., Lau, K.K.L., Ng, E., 2016. Developing street-level $PM_{2.5}$ and PM_{10} land use regression models in high-density Hong Kong with urban morphological factors. Environ. Sci. Technol. 50 (15), 8178–8187.

Sousan, S., Koehler, K., Hallett, L., Peters, T.M., 2017. Evaluation of consumer monitors to measure particulate matter. J. Aerosol Sci. 107, 123–133.

South Coast AQMD, 2017. AirBeam summary report [online]. Available at. http://www.aqmd.gov/aq-spec/sensordetail/airbeam, Accessed date: June 2019.

Su, J.G., Hopke, P.K., Tian, Y., Baldwin, N., Thurston, S.W., Evans, K., Rich, D.Q., 2015. Modeling particulate matter concentrations measured through mobile monitoring in a deletion/substitution/addition approach. Atmos. Environ. 122, 477–483.

Takingspace.org, 2014. AirBeam Technical Specifications, Operation & Performance: Taking Space. [online] Available at. http://www.takingspace.org/airbeam-technical-specifications-operation-performance/, Accessed date: June 2019.

Tessum, M.W., et al., 2018. Mobile and fixed-site measurements to identify spatial distributions of traffic-related pollution sources in Los Angeles. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.7b04889. acs.est.7b04889.

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

Van den Bossche, Joris, Theunis, Jan, Elen, Bart, Peters, Jan, Botteldooren, Dick, De Baets, Bernard, 2016. Opportunistic mobile air pollution monitoring: a case study with city wardens in Antwerp. Atmos. Environ. 141, 408–421.

Van den Bossche, J., De Baets, B., Verwaeren, J., Botteldooren, D., Theunis, J., 2018. Development and evaluation of land use regression models for black carbon based on bicycle and pedestrian measurements in the urban environment. Environ. Model. Softw. 99, 58–69.

VoPham, T., Hart, J.E., Laden, F., Chiang, Y.-Y., 2018. Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. Environmental Health: A Global Access Science Source 17 (1), 1–6.

Weichenthal, S., Ryswyk, K. Van, Goldstein, A., Bagg, S., Shekkarizfard, M., Hatzopoulou, M., 2016. A land use regression model for ambient ultrafine particles in Montreal, Canada: a comparison of linear regression and a machine learning approach. Environ. Res. 146, 65–72.

World Health Organization, May 2, 2018. 9 out of 10 People Worldwide Breathe Polluted Air, but More Countries Are Taking Action. Available at. https://www.who.int/news-room/detail/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action, Accessed date: 16 September 2018.

Yang, P., Yang, Y.H., Zhou, B.B., Zomaya, A.Y., 2010. A review of ensemble methods in bioinformatics. Curr. Bioinforma. 5, 296–308.

Zikova, N., et al., 2017. Estimating hourly concentrations of $PM_{2.5}$ across a metropolitan area using low-cost particle monitors. Sensors 17, 1922.