

# Mapping Microscale PM<sub>2.5</sub> Distribution on Walkable Roads in a High-Density City

Chengzhuo Tong , Zhicheng Shi , Wenzhong Shi , Pengxiang Zhao, and Anshu Zhang

**Abstract**—Monitoring pollution of PM<sub>2.5</sub> on walkable roads is important for resident health in high-density cities. Due to the spatiotemporal resolution limitations of aerosol optical depth observation, fixed-point monitoring, or traditional mobile measurement instruments, the microscale PM<sub>2.5</sub> distribution in the walking environment cannot be fully estimated at the fine scale. In this article, by the integration of mobile measurement data, OpenStreetMap (OSM) data, Landsat images, and other multisource data in land-use regression (LUR) models, a novel framework is proposed to estimate and map PM<sub>2.5</sub> distribution in a typical microscale walkable environment of the high-density city Hong Kong. First, the PM<sub>2.5</sub> data on the typical walking paths were collected by the handheld mobile measuring instruments, to be selected as the dependent variables. Second, geographic prediction factors calculated by Google Street View, OSM data, Landsat images, and other multisource data were further selected as independent variables. Then, these dependent and independent variables were put into the LUR models to estimate the PM<sub>2.5</sub> concentration on sidewalks, footbridges, and footpaths in the microscale walkable environment. The proposed models showed high performance relative to those in similar studies (adj  $R^2$ , 0.593 to 0.615 [sidewalks]; 0.641 to 0.682 [footpaths]; 0.783 to 0.797 [footbridges]). This article is beneficial for mapping PM<sub>2.5</sub> concentration in the microscale walking environment and the identification of hot spots of air pollution, thereby helping people avoid the PM<sub>2.5</sub> hotspots and indicating a healthier walking path.

**Index Terms**—Air pollution, pollution measurement, urban areas.

## I. INTRODUCTION

**I**N RECENT years, exposure to particulate matter of 2.5  $\mu\text{m}$  or smaller (PM<sub>2.5</sub>) has attracted notable attention. PM<sub>2.5</sub> tends to adhere to toxic and harmful substances (such as heavy metals and microorganisms) better than coarser atmospheric

Manuscript received January 10, 2021; revised March 16, 2021 and April 4, 2021; accepted April 11, 2021. Date of publication April 26, 2021; date of current version July 20, 2021. This work was supported in part by The Hong Kong Polytechnic University under Grant 1-ZVN6, in part by National Key R&D Program of China under Grant 2019YFB2103102, and in part by National Natural Science Foundation of China under Grant 41861052. (Corresponding author: Zhicheng Shi.)

Chengzhuo Tong, Wenzhong Shi, and Anshu Zhang are with the Smart Cities Research Institute and the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong (e-mail: chengzhuo.tong@connect.polyu.hk; lswzshi@polyu.edu.hk; april-anshu.zhang@polyu.edu.hk).

Zhicheng Shi is with the Research Institute for Smart Cities, School of Architecture and Urban Planning, Shenzhen University, Shenzhen 518060, China (e-mail: shizhic@hotmail.com).

Pengxiang Zhao is with the GIS Center, Department of Physical Geography and Ecosystem Science, Lund University, 22100 Lund, Sweden (e-mail: pengxiang.zhao@nateko.lu.se).

Digital Object Identifier 10.1109/JSTARS.2021.3075442

particles. Some studies have found that PM<sub>2.5</sub> can cause a series of health problems including respiratory tract and cardiovascular diseases [1]–[4]. A recent study in Singapore further reported that walking was the worst mode of commuting for particle exposure [5]. Furthermore, walkable roads in Asian cities usually have flat sidewalks alongside the roads, which maximizes pedestrian exposure to PM<sub>2.5</sub> [6]. Therefore, a method that can accurately evaluate the spatiotemporal variations in the PM<sub>2.5</sub> distribution at the walkable road level has become a focal point for researchers and policymakers.

The mainstream approach to air quality monitoring is the use of completely fixed air quality monitoring networks, which have been used in most cities [7]. These ground monitoring systems can accurately measure temporal variations in the levels of particulate matter to determine the physical and optical properties [8]. However, the traditional ground monitoring method performs poorly in measuring the spatial variability of particulate matter throughout an area [9]. Furthermore, retrieving aerosol optical depth by remote sensing images has been widely applied to estimate ground-level PM<sub>2.5</sub> concentration [10], [11]. However, most of these studies have focused on estimating PM<sub>2.5</sub> at large spatial scales [12], [13], while studies on microscale are relatively limited [14]. In some high-density cities, the spread of PM<sub>2.5</sub> is hampered by the influence of urban morphology through effects such as the street canyon effect [15], [16]. The traditional methods cannot accurately reflect the spatial differences in the PM<sub>2.5</sub> distribution in cities, especially at the microscale, which may lead to further underestimation of its health effects on humans. Ever since the land-use regression (LUR) model was proposed by Briggs *et al.* [17], it has remained a common and cost-effective method to obtain a more accurate spatial distribution of urban-scale air pollutants due to its integration of land use, meteorological factors and other urban geographic factors [18], [19]. However, the LUR models applied to PM<sub>2.5</sub> require very dense monitoring networks and a large amount of training data [20].

As mobile sensing and information communication technology has advanced, the mobile monitoring method has been increasingly used to collect air quality data. Compared with station-fixed professional instruments, mobile monitoring can increase the spatial resolution of PM<sub>2.5</sub> samples with monitoring instruments [21]. A combination of the LUR model and portable PM<sub>2.5</sub> monitoring sensors has thus been widely applied to related studies of PM<sub>2.5</sub> prediction [22], [23]. However, in most of these studies, the LUR models were developed to obtain the PM<sub>2.5</sub> distribution of the entire urban scale based

on the vehicle or bicycle mobile monitoring platform used to sample PM<sub>2.5</sub> data [24]–[27]. In addition, most PM<sub>2.5</sub> maps are confined by the limitations of these sampled data and the spatial resolution, for example, the number of roadside sampling locations is always limited; the traditional vehicle or bicycle mobile monitoring platform can only follow the vehicle or bicycle route to measure PM<sub>2.5</sub>. As a result, most PM<sub>2.5</sub> maps do not accurately reflect the PM<sub>2.5</sub> distribution in the urban micro-environment, especially in the walking environment of a high-density city [27], [28].

Hong Kong is among the most famous high-density cities in the world, and it has a complex network of pedestrian passages connected by transport hubs, markets, shopping centres, and residential areas. Walking is among the most common travel modes in the city. A recent study reported that Hong Kong residents walk around 6 880 steps per day [28]. However, a study on premature death in Hong Kong people affected by PM<sub>2.5</sub> showed that 32% of patients were affected by PM<sub>2.5</sub> pollution while walking [29]. Due to the high cost of ground monitoring sites, only three roadside PM<sub>2.5</sub> monitoring stations have been established on the sidewalks in the city center to provide hourly average monitoring results. These stations do not provide a comprehensive and intuitive road-level PM<sub>2.5</sub> distribution. At the same time, the existing LUR model-based PM<sub>2.5</sub> studies in Hong Kong also failed to reflect the high-spatial-resolution PM<sub>2.5</sub> distribution in a microscale walking environment [14], [30], [31]. Furthermore, most studies had insufficient extraction and analysis of urban environmental factors that affect changes in PM<sub>2.5</sub> concentrations [6]. Overall, few studies have adopted the LUR model based on data sampled by handheld portable motion sensors. However, it may provide a cost-effective method to model and map air pollution in high-density urban microscale environments.

The purpose of this article is to develop the framework of LUR models for PM<sub>2.5</sub> assessment in the microscale walkable environment of Hong Kong to protect the health of residents while travelling. First, mobile measurement sensors (AirBeam2) were deployed at the same locations as the high-precision instruments to determine their accuracy, and multiple cross-measurements of multiple AirBeam2 sensors were then arranged along a typical walking route to verify their precision. Second, movement sampling of PM<sub>2.5</sub> was performed by repeatedly walking along three typical walking routes over a period of about 12 weeks, followed by data cleaning and screening. Third, combining the predictive factors generated by microscale geographic data like mobile measurement data, OpenStreetMap (OSM) data, Landsat images, and Google Street View images, microscale LUR models were used to determine the PM<sub>2.5</sub> distribution at a higher spatial resolution on a multitype walkable road so that the public could more intuitively obtain PM<sub>2.5</sub> distribution results and plan healthier walking paths with less PM<sub>2.5</sub>, while supporting pedestrian city planning and decision-making. This work explores the potential use of handheld portable mobile sampling sensors, LUR models, and new urban environmental factors to characterize PM<sub>2.5</sub> pollution in the urban microscale walking environment with good spatial resolution.

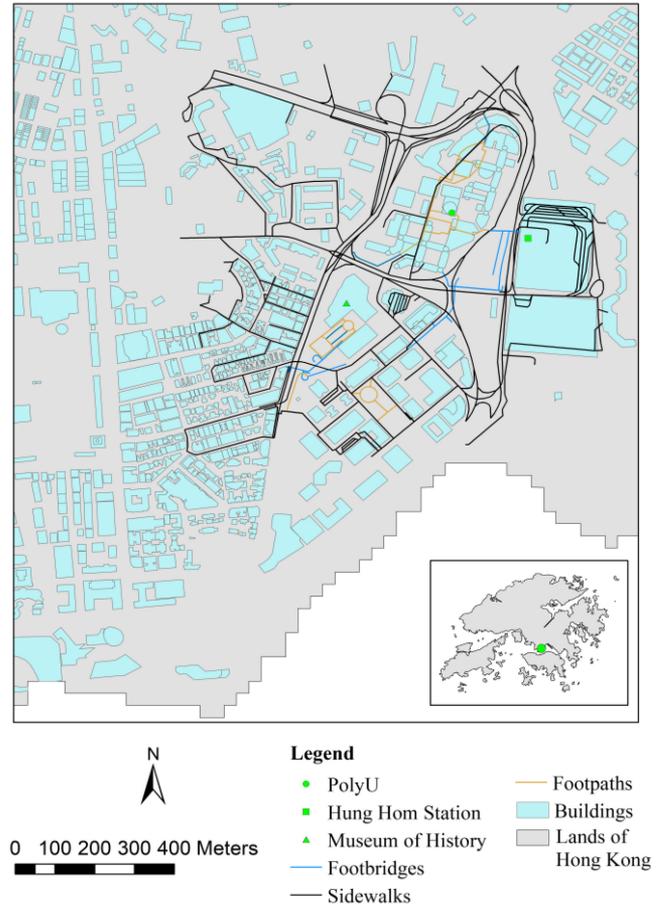


Fig. 1. Study area.

## II. STUDY AREA AND DATASET

### A. Study Area

To combine the abovementioned urban construction characteristics in Hong Kong, the area within 500 m of The Hong Kong Polytechnic University (POLYU) was selected as the center (as shown in Fig. 1).

This area includes schools, commercial areas, residential areas, tourist attractions, bus stations, resting green areas, and so forth, while the walkable road types with different backgrounds mainly include footbridges (car-free environment), sidewalks, and footpaths (as shown in Fig. 2). Sidewalks include not only those in open areas of green vegetation, but also those in narrow areas with almost no vegetation. The areas in which the footpaths are located include schools, tourist attractions, leisure parks, and residential areas. Furthermore, the footbridges connect to a transport hub and to residential buildings and commercial office buildings. This study area is a typical representative example of Hong Kong's pedestrian network.

Note that the experimental area includes one of Hong Kong's busiest transport hubs, the Hung Hom Station, and an open area with lush green vegetation, the History Museum and the Science and Technology Museum. In addition, one of the three roadside air quality monitoring stations of the Hong Kong Government, Mong Kok Station, is near the study area. Therefore,



Fig. 2. Typical walking roads include (a) sidewalks, (b) footpaths, and (c) footbridges.

this research area covers the basic types of walkable roads in Hong Kong, and the background also considers the possible environment of Hong Kong's walkable roads. It thus provides a considerable sample of our research and makes our results scalable.

### B. Mobile Monitoring Data

1) *Mobile Monitoring Instrument of PM<sub>2.5</sub>*: A handheld mobile air quality monitoring instrument (as shown in AirBeam2) was used to measure the PM<sub>2.5</sub> concentration (Fig. 3). The results of field tests to evaluate the performance of the AirBeam2 in measuring the PM<sub>2.5</sub> concentration (South Coast Air Quality Management District, USA:<sup>1</sup> showed that it can detect

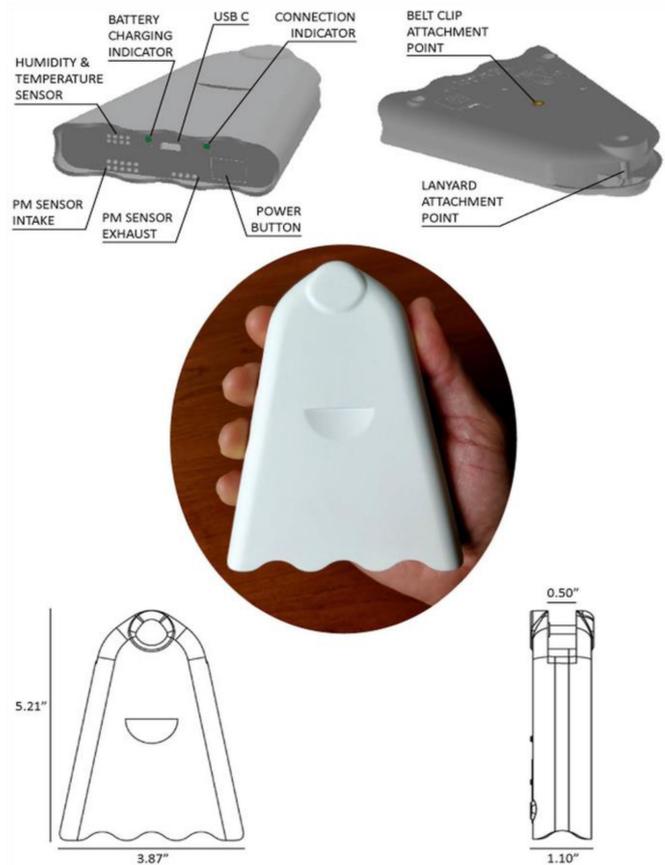


Fig. 3. Internal and external structures of AirBeam2.<sup>2</sup>

microscale variations in the PM<sub>2.5</sub> concentration. A recent study proved its effectiveness for mobile monitoring and indicated that the AirBeam2 is suitable for the purposes of this article [27], which indicates that AirBeam2 is suitable for this article. It should be noted that the PM<sub>2.5</sub> concentration values obtained from the AirBeam2 are not considered as absolute concentration values but are used as estimates to investigate PM<sub>2.5</sub> exposure on walkable roads. Therefore, the PM<sub>2.5</sub> data obtained by the AirBeam2 and inferred in this article are only used as estimates to analyze the variation in PM<sub>2.5</sub> exposure on walkable roads, not as absolute values.

2) *Mobile Data Collection*: The AirBeam2 was used to sample mobile data in the following three situations. 1) LUR models geomapping: Considering the environment of the study area and a sufficient sampling amount of each type of road section, three routes were selected. These routes are shown in Fig. 4(a) and described in Table I. 2) LUR models validation: This route [Fig.4 (b)] began on Mira Place 1 [point A] to InterContinental Grand Stanford Hong Kong [point B]. 3) The precision test of the Airbeam2: The walking route from the Z core of POLYU (point A) to the student dormitory (point B) in the study area was selected to monitor the precision of the instruments during the morning and evening peaks [as shown in Fig. 4(c)].

<sup>1</sup>[Online]. Available: <http://www.aqmd.gov/aqspect/evaluations/field>

<sup>2</sup>[Online]. Available: <https://www.habitatmap.org/blog/airbeam2-technical-specifications-operation-performance>

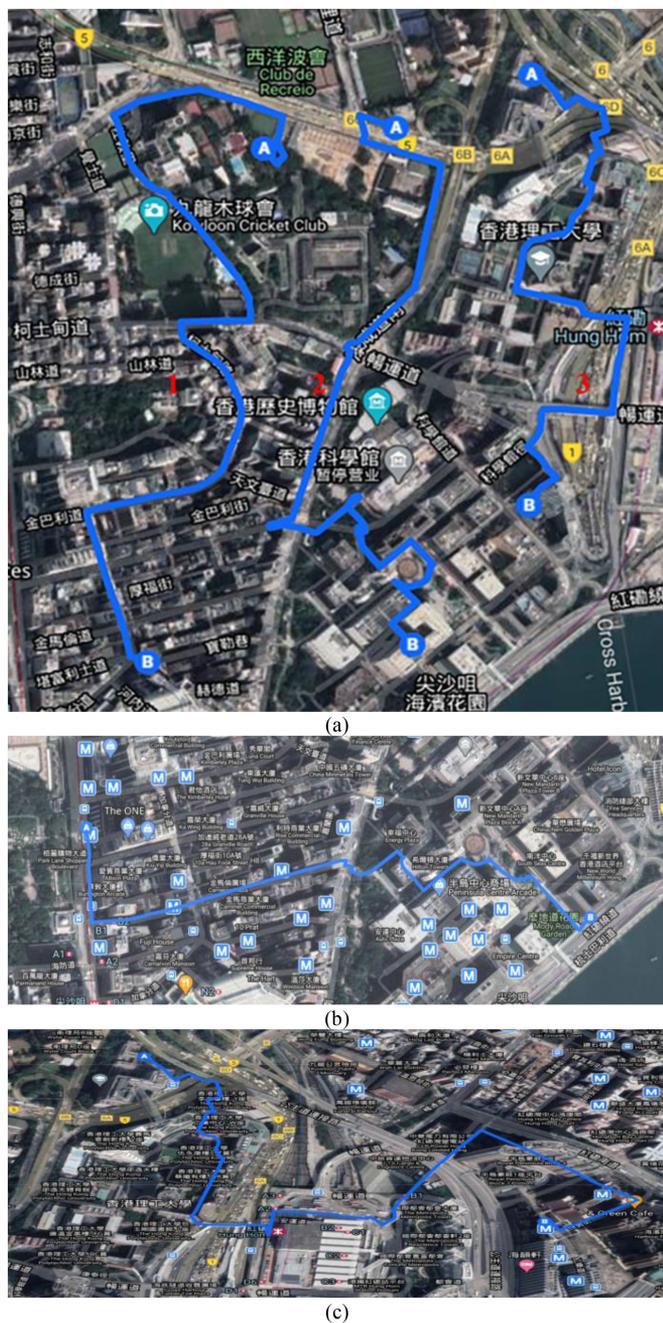


Fig. 4. The sampling roads for (a) LUR models geomapping, (b) LUR models validation, and (c) the precision test of the Airbeam2.

To avoid the region-dominant influence of PM<sub>2.5</sub> pollution from the Pearl River Delta region of mainland China in other seasons, measurements were taken from June to September, 2018 because the PM<sub>2.5</sub> concentration is determined mainly by local emission during these times. This article mainly considers the PM<sub>2.5</sub> concentration in the walking environment during the morning peak (08:00 to 09:30) and the evening peak (18:00 to 19:30). During these periods, data were collected from 168 measurements over 12 weeks.

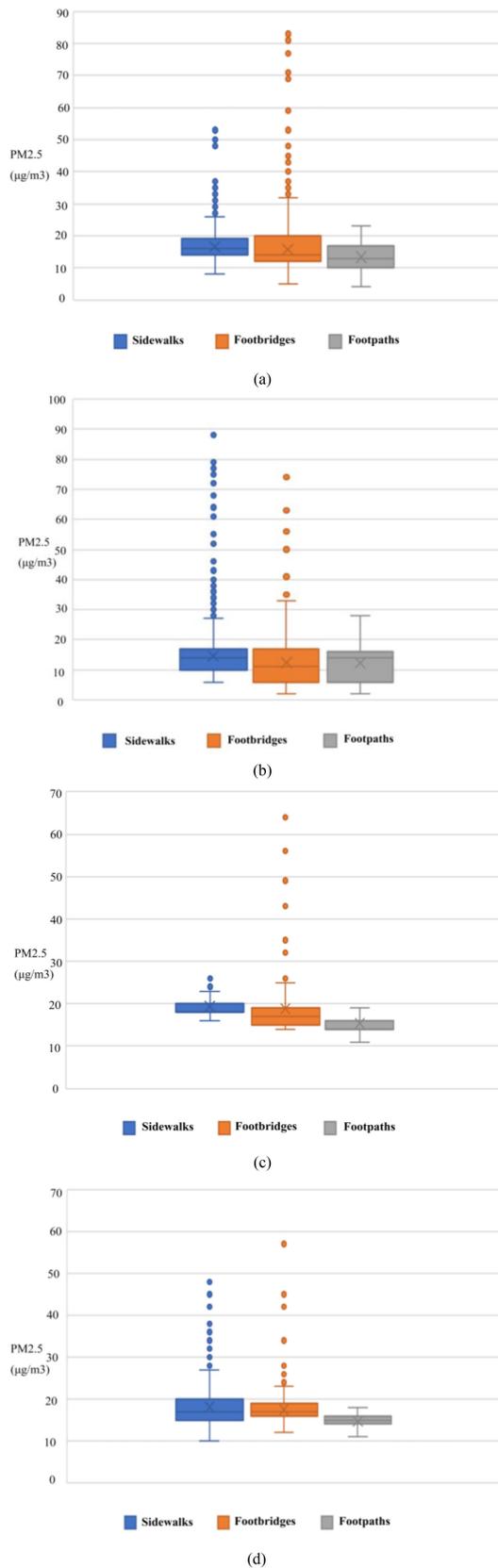


Fig. 5. Boxplots of data in (a) morning peaks on weekdays, (b) evening peaks on weekdays, and (c) morning peaks on weekends, and (d) evening peaks on weekends.

TABLE I  
INTRODUCTION OF SAMPLING ROADS

Id	Starting point	Endpoint	Description
1	United Services Recreation Club	7-Eleven (33-35 Carnarvon Rd)	Sidewalks Majority of roads in high-density environments, some in an open environment Some roads with many green plants, some with no greenery
2	The Hong Kong Girl Guide Association	Regal Kowloon Hotel	Sidewalks, footpaths, bridges Sidewalks: an open environment Footpaths: garden, square
3	HK POLYU (Block Z)	Hotel ICON	Footpaths, bridges, sidewalks Footpaths: POLYU campus Bridges: Crossing complex environments

Fig. 5 shows the variability of the measured PM<sub>2.5</sub> data on footpaths, sidewalks, and footbridges in Morning and evening peaks on weekdays and weekends were shown in Fig. 4. As can be seen from the comparison of the length and shape of the box plots in groups of different periods, the median PM<sub>2.5</sub> corresponding to different road types is relatively close, but the distribution difference is self-evident. Among them, the distribution of PM<sub>2.5</sub> sampling data on footpaths is relatively concentrated, while the distribution of PM<sub>2.5</sub> data on sidewalks and footbridges is more dispersed. At the same time, the PM<sub>2.5</sub> box plots that correspond to sidewalks and footbridges have outliers concentrated on one side of the larger value, whilst the PM<sub>2.5</sub> box plots that correspond to footpaths have no outliers. This suggests that because the background environment is relatively stable, the PM<sub>2.5</sub> distribution for footpaths in a car-free walking environment tends to be stable. Due to the influence of traffic flow, especially near the Hung Hom transport hub, the PM<sub>2.5</sub> concentrations on sidewalks and footbridges tend to fluctuate, and high outliers are numerous.

At the same time, to verify the instrument's accuracy, we placed an AirBeam2 at approximately the same location as the roadside monitoring station near Mong Kok Station for comparison. The continuous sampler used to monitor PM<sub>2.5</sub> at the Mong Kok Station is a Thermo Scientific environmental particle monitor and tapered element oscillating microbalance that has high data quality, precision, reliability, and unparalleled support.

Moreover, to evaluate the precision of the AirBeam2 in the measurement process, the following verification was arranged during a portion of the study period. Three AirBeam2 instruments, labeled AirBeam2-A, AirBeam2-B, and AirBeam2-C, were used on the same sampling road. AirBeam2-A and AirBeam2-B were used to sample in the same direction, and AirBeam2-A and AirBeam2-C were used to sample in different directions.

a) *Data Processing*: As mentioned above, the AirBeam2 measures the PM<sub>2.5</sub> concentration by light scattering. When aerosols are monitored using light-scattering techniques, an excessively high humidity level will cause water to condense and lead to an increase in the measured concentration of PM<sub>2.5</sub>. To

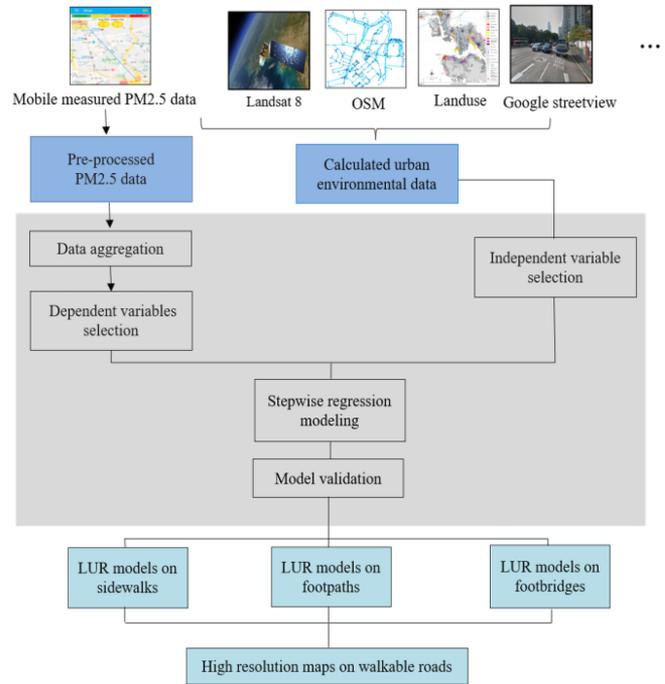


Fig. 6. Methodology of LUR model development on walk roads.

correct the possible error of AirBeam2 measurements due to the relative humidity, this article simultaneously used an AirBeam2 to monitor the obtained relative humidity data to correct the instrument [32]. The corrected PM<sub>2.5</sub> concentration is called PM<sub>2.5corrected</sub>. The formula is as follows [32]:

$$PM_{2.5corrected} = \frac{PM_{2.5AirBeam2}}{1 + 0.25 \frac{RH^2}{1-RH}} \quad (1)$$

where PM<sub>2.5AirBeam2</sub> is the PM<sub>2.5</sub> concentration measured by the AirBeam2; RH represents the relative humidity measured by the AirBeam2.

To determine the typical environmental impact during mobile measurements, we took photographs of the scene and recorded the time. The data affected by extraneous factors were then deleted during data preprocessing. For example, while a person waited to cross the road, the PM<sub>2.5</sub> concentration data measured when a diesel vehicle released a large amount of exhaust gas were removed because the data were significantly affected by the polluted exhaust gas. The data collected at a location near a construction site were also deleted. The noise of the measured data was filtered. It should be noted that none of the people who measured the data for this article were smokers.

### III. METHODS

Fig. 6 shows the overall methodological framework to estimate the PM<sub>2.5</sub> distribution on walkable roads. First, the collected PM<sub>2.5</sub> data and multiple-source urban environmental factors were preprocessed to obtain the dependent variables and independent variables required to establish the model, and the independent variables and dependent variables were then selected respectively based on the sidewalks, footpaths, and

footbridges. The related stepwise linear regression models of PM2.5 estimation were further established and validated, and high-resolution PM2.5 distribution maps for walkable roads were eventually obtained.

### A. Dependent Variables of LUR Models

In this article, the PM2.5 concentration data obtained from the mobile measurement were regarded as the dependent variable in the model. To represent the spatial variability of the PM2.5 on walking roads, a spatially balanced points network from the continuously mobile measured PM2.5 needs to be established. So it is important to choose the optimal spatial scale of mobile measured PM2.5 data aggregation. If the spatial scale of the grid used for spatial aggregation is too small, there will be too few samples in many grids to obtain an actual average to represent the true PM2.5 of the corresponding pedestrian zone. Conversely, if the spatial scale of the grid is too large, many local subtle changes in the original data will be masked. Several studies have found that in the field of environmental and epidemiological studies, misleading conclusions can be obtained when the spatial resolution of spatial aggregation is inappropriate.

In this article, the PM2.5 concentration data obtained from the mobile measurements were regarded as the dependent variable in the model. To represent the spatial variability of the PM2.5 concentration on walkable roads, a spatially balanced point network must be established from the continuous mobile measurements of PM2.5. It is therefore important to choose the optimal spatial scale of mobile measurements of PM2.5 data aggregation. If the spatial scale of the grid used for spatial aggregation is too small, many grids will have too few samples to obtain an actual average to represent the true PM2.5 concentration of the corresponding pedestrian zone. Conversely, if the spatial scale of the grid is too large, many subtle local changes in the original data will be masked. Several studies have shown that in environmental and epidemiological studies, misleading conclusions can be obtained when the spatial resolution of spatial aggregation is inappropriate.

This article specifically analyzes the effects of changes in spatial resolution on the sample size (i.e., the number of airborne moving sample points per aggregation point) to select the best spatial resolution for data aggregation [29]. The average sample size for various resolutions was initially tested using five spatial resolutions (5, 10, 20, 30, and 40 m). Fig. 7(a) and (b) shows the statistics of the PM2.5 sampling points for each aggregation point and the total number of aggregation points for each aggregation resolution of PM2.5. The results of the initial tests indicated that spatial resolutions of 5 and 10 m introduce a very small sample size per aggregation point which does not provide statistical reliability for the typical geographical studies performed herein. The minimum sample size introduced by the 5-m and 10-m spatial resolutions cannot even be assumed to have a normal distribution to support the basic assumptions required for many statistical analyses. Therefore, spatial resolutions of 5 and 10 m should not be used for data aggregation. At the same time, when a spatial resolution of 30 or 40 m is adopted, the total number of aggregation points is smaller. This is not suitable to

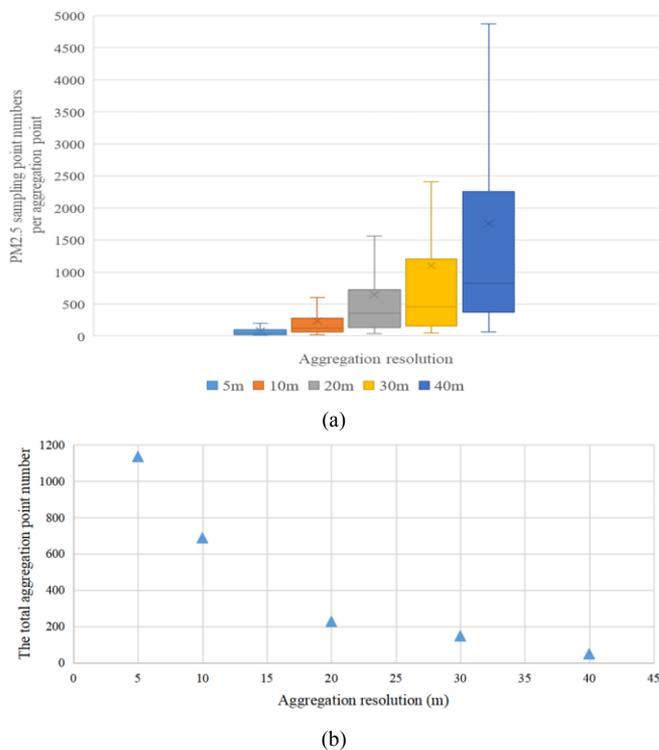


Fig. 7. Boxplot and total aggregation point number of the mobile sampling point number per aggregation point for each aggregation resolution for PM2.5.

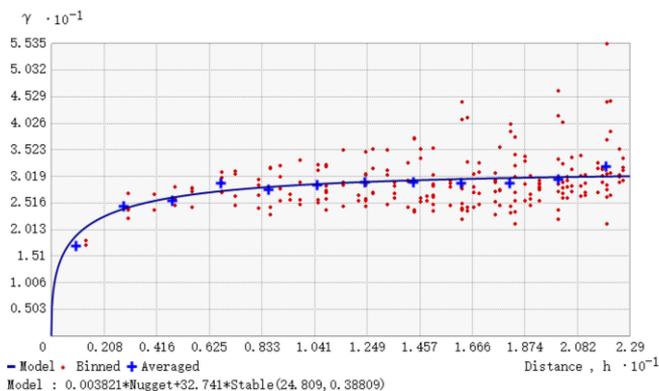


Fig. 8. Resulting empirical semivariogram model for testing the appropriate spatial aggregation resolution.

perform PM2.5 refinement inversion in this study area under microscale to determine local subtle feature changes.

Geostatistical analysis is then used to further quantify and determine the optimal spatial resolution for spatial aggregation. A good grid should have sufficiently large cells to provide a sample size appropriate to ensure statistical reliability, but not such large cells that the spatial characteristics of locally varying data cannot be well preserved. Based on the experience of PM2.5 spatial modeling studies, semivariogram modeling is used to determine the optimal spatial resolution of the dependent variable data aggregation. The semivariogram is a function of distance that increases with distance over a certain range until it reaches stability. Within the range of stable values, the sampling points are spatially related to each other, but outside this range,

TABLE II  
DATA INTRODUCTION

Data source	Variables	Buffer/ distance/point	Unit
Landsat 8	NDVI	Point	-1~1
	IBI	Point	-1~1
Google Street View Image	Mean of GVI ( <i>me-GVI</i> )	Buffer	%
OSM	Primary, secondary, trunk road line density ( <i>rd-main</i> )	Buffer	m/m <sup>2</sup>
Road network	Tertiary road line density ( <i>rd-ter</i> )	Buffer	m/m <sup>2</sup>
	Service road line density ( <i>rd-ser</i> )	Buffer	m/m <sup>2</sup>
	All motor vehicle roads line density ( <i>rd-all</i> )	Buffer	m/m <sup>2</sup>
Land use data	Commercial area ( <i>lu-com</i> )	Buffer	m <sup>2</sup>
	Government, Institution & Community area ( <i>lu-gov</i> )	Buffer	m <sup>2</sup>
	Open space area ( <i>lu-ops</i> )	Buffer	m <sup>2</sup>
	Residential area ( <i>lu-res</i> )	Buffer	m <sup>2</sup>
DSM data	SVF	Point	%
Building height data	Mean of building height ( <i>me-bh</i> )	Buffer	m
Bus stop data	The closest distance to bus stop ( <i>CDS</i> )	Distance	m
	Elevation	Point	m
Google	Terrain slope	Point	Degree
Elevation	Terrain aspect	Point	Degree
Meteorological data	Wind speed (2m, 10m) ( <i>WS2, WS10</i> )	Point	m/s
	Atmospheric pressure ( <i>PS</i> )	Point	Pa
	Temperature near the ground ( <i>T2m</i> )	Point	F
	Planetary boundary layer ( <i>PBLH</i> )	Point	km

buffers (meters): 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 400, 500, 1000

the sampling points are spatially independent. Semivariograms can be used to describe the spatial correlation and autocorrelation of sample points. The spatial dependence between sampling points is a good indicator of the spatial aggregation scale. The semivariogram equation is as follows:

$$\gamma = \frac{1}{2N(d)} \sum_{X_i - X_j} = d(V_i - V_j)^2 \quad (2)$$

where  $\gamma$  is a semivariogram. The sample points  $X_i$  and  $X_j$  are paired by creating a semivariogram.  $V_i$  and  $V_j$  are the measured values of the sampling points  $X_i$  and  $X_j$ .  $d$  is the distance between  $X_i$  and  $X_j$ .  $N(d)$  is the total number of all pairs of sample points.

The best stable-type semivariogram model finally established shows that the optimal spatial resolution of PM2.5 data aggregation is approximately 24.81 m (as shown in Fig. 8), which can reveal the local characteristics of PM2.5 variation on walkable roads. To be consistent with the spatial grid system constructed by the existing environmental research, the 20-m spatial resolution was finally selected for spatial aggregation of the PM2.5 mobile measurement data.

The PM2.5 measurement data on the sampling routes were thus covered by 228 grids with a spatial resolution of 20 m. For each grid cell, the PM2.5 concentration data for the sampling points within the grid were used to calculate the average concentration in each grid. On this basis, we counted the number of data points in each grid and selected the grids with a higher number of data points to build the model. This ensures that the calculated PM2.5 means in these grids are reliable and provides a good tradeoff between the spatial input distribution and model performance. The grid units selected included 30 grid units in the footbridge section, 43 on footpaths and 100 in the sidewalk section. This ensured the reliability of the calculated average concentration of PM2.5.

### B. Independent Variables of LUR Models

After considering the high-density building environment in Hong Kong and the actual environment of the study area, the following kinds of data were used to calculate the potential independent variables in this article.

- 1) Meteorological data (mainly from resampled from the GEOS 5-FP meteorological data).

TABLE III  
OPTIMAL BUFFER-BASED VARIABLES OF SIDEWALKS, FOOTPATHS, AND FOOTBRIDGES

Buffer-based variables	Optimal distances (sidewalks)	Optimal distances (footpaths)	Optimal distances (footbridges)
<i>rd-main</i>	200m	100m	30m
<i>rd-ter</i>	300m	180m	40m
<i>rd-ser</i>	180m	60m	60m
<i>rd-all</i>	100m	40m	20m
<i>lu-com</i>	180m	250m	300m
<i>lu-gov</i>	250m	180m	140m
<i>lu-ops</i>	80m	60m	60m
<i>lu-res</i>	300m	300m	250m
<i>me-bh</i>	40m	140m	100m
<i>me-GVI</i>	60m	/	/

TABLE IV  
ACCURACY TEST OF AIRBEAM2 IN HONG KONG

Time Period	Duration	Weekdays		Weekends	
		Average PM2.5 ( $\mu\text{g}/\text{m}^3$ ) AirBbeam2	Average PM2.5 ( $\mu\text{g}/\text{m}^3$ ) TEOM	Average PM2.5 ( $\mu\text{g}/\text{m}^3$ ) AirBeam2	Average PM2.5 ( $\mu\text{g}/\text{m}^3$ ) TEOM
8:00-10:00	1-hr	10.276	12	12.105	15.5
18:00-20:00	1-hr	10.682	14	10.062	13

TABLE V  
PRECISION TEST OF AIRBEAM2 IN HONG KONG

Time period	PM2.5( $\mu\text{g}/\text{m}^3$ ) X-Axis	PM2.5( $\mu\text{g}/\text{m}^3$ ) Y-Axis	Linear regression	R <sup>2</sup>
08:33-08:53	AirBeam2-A	AirBeam2-B	$y=0.9855x+0.2304$	0.8873
08:33-08:53	AirBeam2-A	AirBeam2-C	$Y=0.993x+0.151$	0.8731
17:51-18:11	AirBeam2-A	AirBeam2-B	$y=0.9909x+0.1034$	0.8816
17:51-18:11	AirBeam2-A	AirBeam2-C	$y=0.9906x+0.1159$	0.8663

TABLE VI  
SUMMARY OF LUR REGRESSION MODELS OF SIDEWALKS

Time	Period	Adjusted R <sup>2</sup>	RMSE	P-value	R <sup>2</sup>
Weekdays	Morning peak	<i>Sidewalks</i>	<i>Sidewalks</i>	<i>Sidewalks</i>	<i>Sidewalks</i>
		0.645	6.085	<0.0001*	0.615
		<i>Footpaths</i>	<i>Footpaths</i>	<i>Footpaths</i>	<i>Footpaths</i>
		0.743	6.112	<0.0001*	0.676
		<i>Footbridges</i>	<i>Footbridges</i>	<i>Footbridges</i>	<i>Footbridges</i>
		0.815	7.305	<0.0001*	0.796
	Evening peak	<i>Sidewalks</i>	<i>Sidewalks</i>	<i>Sidewalks</i>	<i>Sidewalks</i>
		0.637	6.067	<0.0001*	0.608
		<i>Footpaths</i>	<i>Footpaths</i>	<i>Footpaths</i>	<i>Footpaths</i>
		0.751	6.156	<0.0001*	0.682
		<i>Footbridges</i>	<i>Footbridges</i>	<i>Footbridges</i>	<i>Footbridges</i>
		0.791	7.158	<0.0001*	0.783
Weekends	Morning peak	<i>Sidewalks</i>	<i>Sidewalks</i>	<i>Sidewalks</i>	<i>Sidewalks</i>
		0.614	5.876	<0.0001*	0.593
		<i>Footpaths</i>	<i>Footpaths</i>	<i>Footpaths</i>	<i>Footpaths</i>
		0.735	6.025	<0.0001*	0.655
		<i>Footbridges</i>	<i>Footbridges</i>	<i>Footbridges</i>	<i>Footbridges</i>
		0.796	7.171	<0.0001*	0.785
	Evening peak	<i>Sidewalks</i>	<i>Sidewalks</i>	<i>Sidewalks</i>	<i>Sidewalks</i>
		0.629	5.996	<0.0001*	0.603
		<i>Footpaths</i>	<i>Footpaths</i>	<i>Footpaths</i>	<i>Footpaths</i>
		0.727	5.592	<0.0001*	0.641
		<i>Footbridges</i>	<i>Footbridges</i>	<i>Footbridges</i>	<i>Footbridges</i>
		0.803	7.202	<0.0001*	0.797

TABLE VII  
THE SIGNIFICANCE AND COLLINEARITY OF THE VARIABLES

Independent variable		Estimate	Prob >  t	VIF
Sidewalks	GVI	-4.0824	0.0403*	1.646
	Elevation	-0.6139	<0.0001*	1.245
	PBLH	-0.6139	0.0313*	1.487
	SVF	22.1943	<0.0001*	1.168
	Rd-all	4.0435	<0.0001*	1.503
Footpaths	Intercept	-6.4824	<0.0001*	/
	IBI	82.724	<0.0001*	1.048
	NDVI	-0.6139	0.0317*	1.733
	WS2	-2.793	<0.0001*	1.176
	T2M	23.7349	0.0274*	1.359
	PBLH	-1.864	0.0435*	1.804
	Intercept	-70.9862	<0.0001*	/
Footbridges	IBI	33.567	0.00321*	1.596
	RD-all	5.700	<0.0001*	1.158
	WS2	-2.669	<0.0001*	1.253
	T2M	42.534	0.0314*	1.702
	STOP	-2.244	<0.0001*	1.057
	Intercept	-21.249	0.000246*	/

TABLE VIII  
VALIDATION OF LUR MODELS

Time	Period	Adjusted R <sup>2</sup> (footways)	Adjusted R <sup>2</sup> (sidewalks)	Adjusted R <sup>2</sup> (footbridges)
Weekdays	Morning peak	0.646	0.583	0.767
	Evening peak	0.652	0.565	0.743
Weekends	Morning peak	0.621	0.541	0.752
	Evening peak	0.603	0.552	0.773

- 2) Landsat 8 data.
- 3) Google Street View and Google Elevation.
- 4) Digital Surface Model (DSM) data.
- 5) OSM road network data;
- 6) Bus stop data.
- 7) Land use data.

Therefore, this article selected several indicators as shown in Table II, including 10 buffer variables describing road lines, land use, building height (at 25 buffer lengths each), one distance variable describing the closest distance to a bus stop, and 11 point variables describing normalized difference vegetation index (NDVI), the index-based built-up index (IBI), DSM, elevation factors, and meteorological factors, resulting in a total of 262 (i.e.,  $10 \times 25 + 12$ ) variables available for selection as the independent variables in each model.

Compared with similar research, this article introduced several new urban morphology factors and focused on the walkable environment. Therefore, the sky view factor (SVF) related to the three-dimensional environment of urban buildings [33], [34], the green view index (GVI) factor of green vegetation coverage of the urban canopy [35], [36], and the IBI, related to urban impervious surfaces, were all used in the modeling process [37], [38].

### C. LUR Modeling and Validation

Significant differences in the PM<sub>2.5</sub> concentrations were found among sidewalks, footpaths, and footbridges, between

the morning and evening peaks, and on weekdays and weekends due to differences in environmental background and traffic conditions. Therefore, different prediction models for PM<sub>2.5</sub> on sidewalks, footpaths, and footbridges were established for various time periods.

1) *Independent Variable Selection*: Combined with the environment and the characteristics of the data related to the independent variables, all 22 variables mentioned above were applied to the model. First, the correlations of independent variables were analyzed and the highly correlated independent variables were deleted, which helped to better distinguish the individual contributions of independent variables and prevent redundancy.

Moreover, we used a distance decay regression selection strategy to select the variables obtained via buffer analysis using different distances [39]. First, the correlation coefficient between the PM<sub>2.5</sub> concentration of the sidewalk points and all variables was selected to establish a series of distance attenuation curves. Each curve then represented the trend of the correlation coefficient corresponding to a series of independent variables of different buffer sizes. Finally, these selected variables were used in the further modeling process (Table III). The variance inflation factor (VIF) was checked to ensure that there is no significant collinearity among these selected independent variables in resultant models. The selected independent variables with VIF less than 2 were used to select independent variables with significant collinearity.

TABLE IX  
COMPARISON WITH EXISTING STUDIES USING MOBILE MEASUREMENTS TO DEVELOP LUR MODELS OF PM<sub>2.5</sub> IN URBAN AREA

Study	Location	Mode	Regression	Model R <sup>2</sup>
Su et al. (2011)	Upstate New York	Motor vehicle	Linear	0.58
Zwack et al. (2011)	Brooklyn, New York	Walking	Generalized Additive Model (GAM)	0.32
Li et al. (2013)	Los Angeles, CA	Hybrid vehicle	Linear GAM	0.51
Hankey et al. (2015)	Minneapolis, MN	Bicycle	Linear	Adj-R <sup>2</sup> : Morning, 0.30 Afternoon, 0.49
Lim et al. (2019)	Seoul	Walking	Linear	0.63
This study	Hong Kong	Walking	Liner	Adjusted R <sup>2</sup> : 0.541 - 0.773

In particular, for footpath models, because the Google Street View images were taken by a Google Street View car (for which the coverage is mainly restricted to areas where sidewalks and footbridges are located), the car-free area in which the footpaths were located was not covered. Therefore, for footpaths, we excluded the GVI factor, while including the remaining 21 factors. For the footbridge models, meanwhile, 21 of the variables mentioned above were again applied to the models, but in this case excluding the wind speed at a height of 2 m.

2) *Stepwise Regression Modeling*: After the independent variables were selected, a stepwise multiple linear regression method was used to measure the goodness of the statistical model by the Akaike information criterion and to determine the LUR regression models of PM<sub>2.5</sub> in different road sections in different periods [40]. On this basis, the t-test ( $\text{prob} > |t|$ ) was used to determine the significance level of the variables, and the VIF value was obtained to determine whether the regression models had a collinearity problem. At the same time, the adjusted  $R^2$  value for each model was also checked to test the model performance. Moreover, in the validation process, this article used the 10-fold cross-validation method to test the performance of these models.

3) *LUR Models Validation*: To better validate the models, we used the mobile measurements on a typical section of Tsim Sha Tsui in our study area on weekends (21 and 22 July) and weekdays (23 to 28 July) to further analyze the performance of the LUR model. Surrounded by a densely constructed area, this route consists of sections such as footpaths, sidewalks, and flyovers [as shown in Fig. 4(b)].

## IV. RESULTS

### A. Instrument Performance

1) *Accuracy Test of AirBeam2*: We measured the PM<sub>2.5</sub> concentration with an AirBeam2 at a location similar to that of the Mong Kok Monitoring Station. We calculated the average PM<sub>2.5</sub> concentrations per hour measured with the AirBeam2 and compared it to the average PM<sub>2.5</sub> concentrations per hour

measured with a tapered element oscillating microbalance. The results showed measurement accuracy values of about 0.734 and 0.763, and in the evening peaks on workdays and weekends, the corresponding measurement accuracy values were about 0.781 and 0.774, respectively (as shown in Table IV). The PM<sub>2.5</sub> concentration measurements measured with the AirBeam2 showed good correlations with the corresponding tapered element oscillating microbalance measurements.

2) *Precision Test of AirBeam2*: To evaluate the precision of the AirBeam2 in a more systematic and comprehensive manner, the walking route from the Z core of POLYU (point A) to the student dormitory (point B) in the study area was selected to monitor the precision of the instruments during the morning and evening peaks [as shown in Fig. 4(c)]. This route has several road types, such as flyovers, sidewalks, and footpaths. The surrounding environment includes bus stops, campuses, and residential areas. As a result, the precision of these instruments can be evaluated more rigorously. The precision of the instruments was evaluated by computing the linear regression and the correlation of two of the same instruments.

We arranged a route from the student residence to the Z core for the morning peak and a route from the Z core to the student residence for the evening peak. Three AirBeam2 instruments (AirBeam2-A, AirBeam2-B, and AirBeam2-C) were used in this experiment. AirBeam2-A and AirBeam2-B were used to sample in the same direction throughout the study, and AirBeam2-A and AirBeam2-C were used to sample in different directions. The results of these experiments are shown in Table V.

### B. LUR Models Results

Table VI is a result summary of the LUR models for sidewalks, footpaths, and footbridges after the previous steps. First, the correlation coefficients of the sidewalks, footpaths, and footbridge models for different periods were all greater than 0.61, 0.72, and 0.79, respectively. After ten-fold cross-validation, the three kinds of LUR models also displayed good performance. Moreover, it was observed that all selected independent variables passed the significance test (as shown in Table VII).

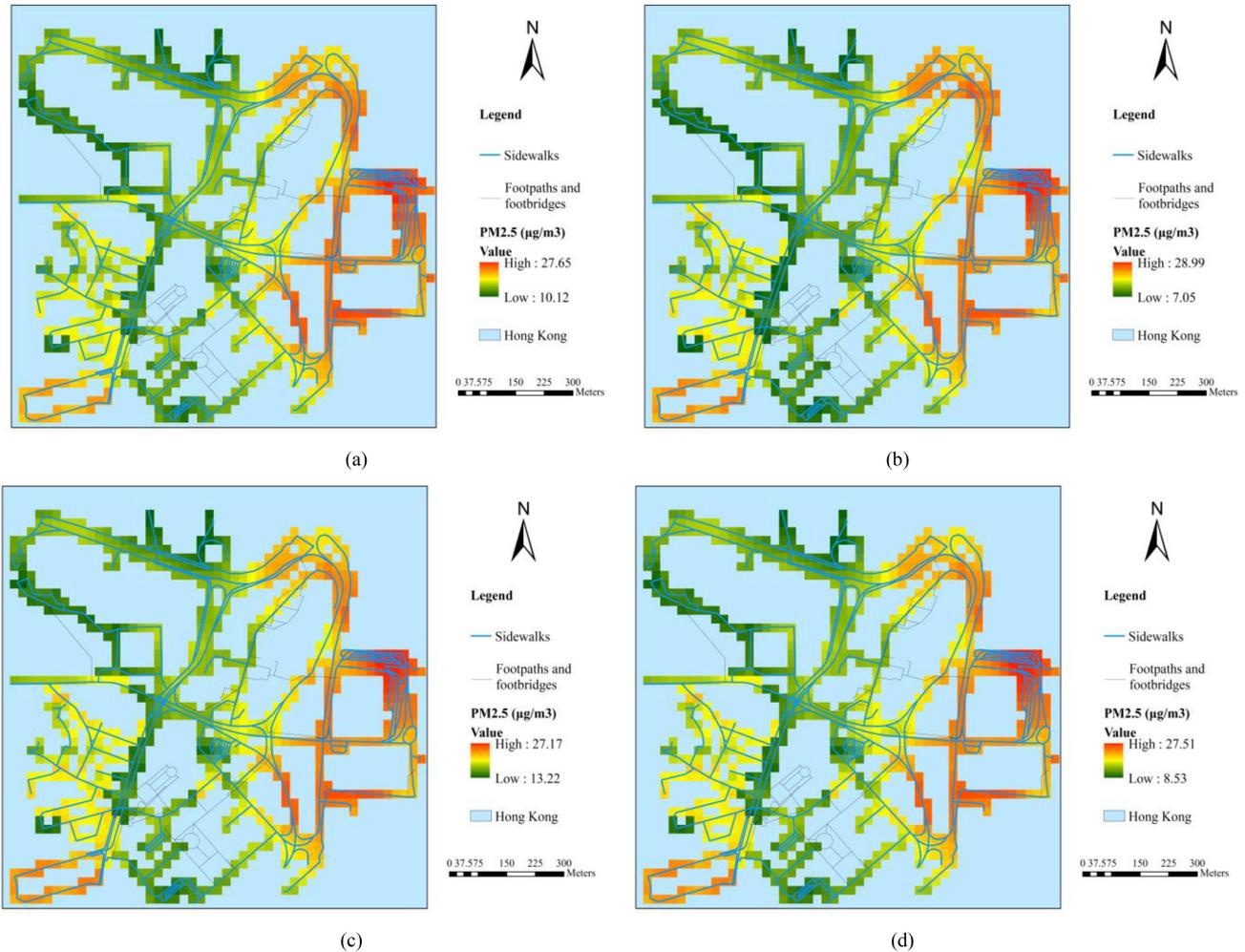


Fig. 9. Thematic maps of PM<sub>2.5</sub> on sidewalks during (a) the morning peak on the weekday, (b) the evening peak on the weekday, (c) the morning peak on the weekend, and (d) the evening peak on the weekend.

It can be seen that when obtaining micro-scale PM<sub>2.5</sub> pollution information in the special walking environment of a high-density city such as Hong Kong, attention should be paid to the difference in the PM<sub>2.5</sub> statistics caused by the background environment differences among sidewalks, footpaths, and footbridges, and urban three-dimensional, green vegetation coverage and building indices should be added.

### C. LUR Models Validation

To better validate the models, we used the mobile measurements for a typical section of Tsim Sha Tsui in our study area on weekends (21 and 22 July) and weekdays (23 to 28 July) to further analyze the performance of the LUR models.

This route is surrounded by a densely constructed area and consists of sections such as footpaths, sidewalks and flyovers. The results are shown in Table VIII. A comparison with similar studies [24]–[26], [41], [42] (as shown in Table IX) showed that the LUR models in this article could accurately estimate the spatial variation in the PM<sub>2.5</sub> concentration on the walking paths of the study area.

### D. LUR Models Geomapping

Using the PM<sub>2.5</sub> LUR models of sidewalks, footpaths, and footbridges to process the data, we obtained PM<sub>2.5</sub> distribution maps at a spatial resolution of 20 m, as shown in Figs. 9, 10, and 11. The PM<sub>2.5</sub> distribution maps for weekdays were obtained by processing the data for the morning and evening peaks on 13 August, 2018, whilst the PM<sub>2.5</sub> distribution maps for weekends were obtained by processing the data for 17 August, 2018.

First, the PM<sub>2.5</sub> maps of the sidewalks on weekdays and weekends show higher PM<sub>2.5</sub> concentrations around the Hung Hom Transport Hub and Tsim Sha Tsui Block. In contrast, the PM<sub>2.5</sub> concentrations on Jordan Road and in the Urban Park's Centennial Garden were relatively low.

The PM<sub>2.5</sub> maps of the footpaths on weekdays and weekends also show that the PM<sub>2.5</sub> concentrations were relatively low for most footpaths on the POLYU campus, likely due to its completely car-free environment. However, due to the traffic conditions of Hung Hom Station, the PM<sub>2.5</sub> concentrations of the campus footpaths near Hung Hom Station were relatively high. Due to the smoke from nearby restaurants, higher PM<sub>2.5</sub> concentrations were found on a narrow sidewalk between the

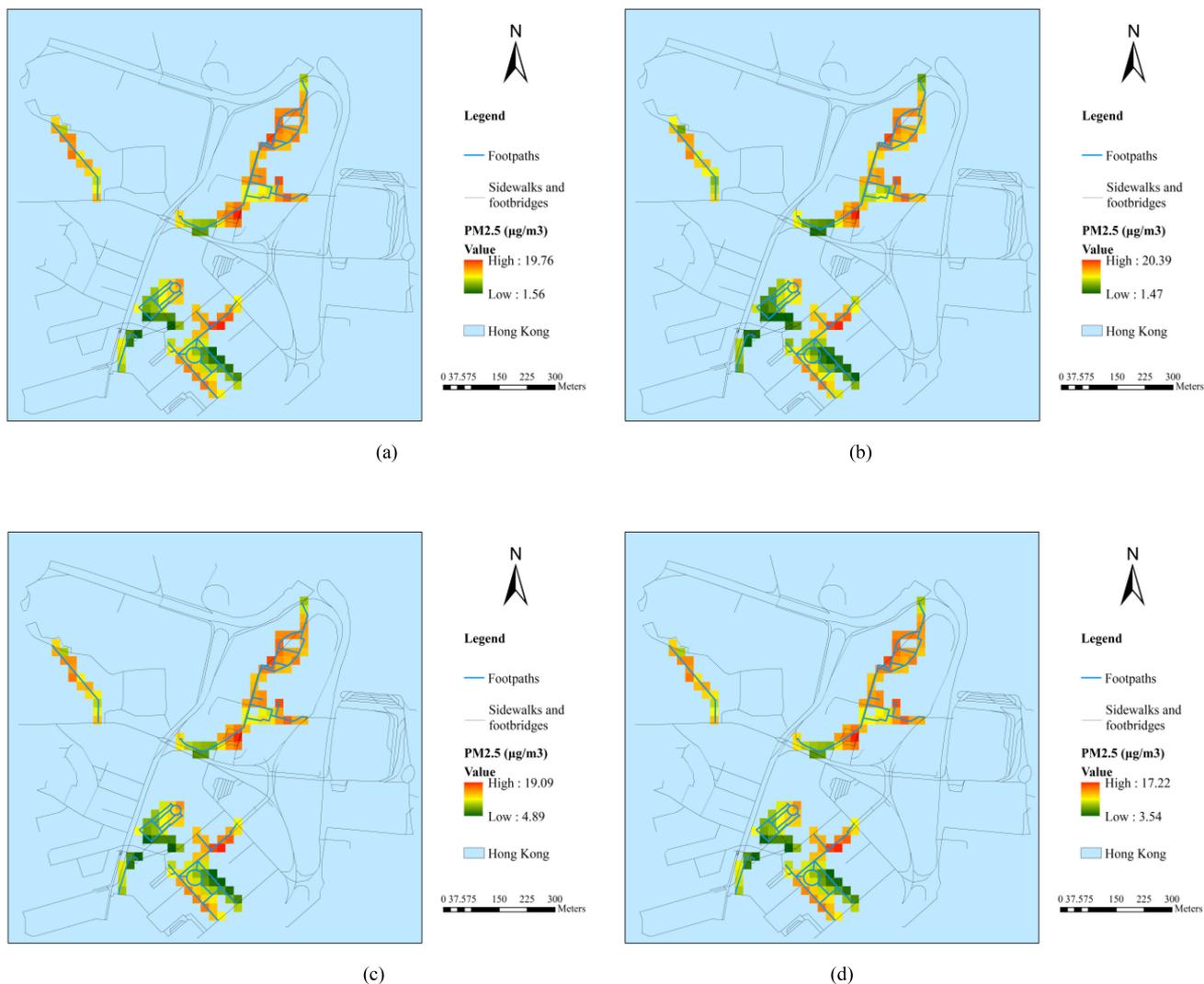


Fig. 10. Thematic maps of PM<sub>2.5</sub> on footpaths during (a) the morning peak on the weekday, (b) the evening peak on the weekday, (c) the morning peak on the weekend, and (d) the evening peak on the weekend.

Science Museum road and the Centennial Garden of the Municipal Council. In contrast, the PM<sub>2.5</sub> concentrations in the open environment in front of the Hong Kong Science Museum and the Hong Kong Museum of History were relatively low.

The PM<sub>2.5</sub> maps of the footbridges on weekdays and weekends show very high PM<sub>2.5</sub> concentrations (up to 100  $\mu\text{g}/\text{m}^3$ ) on the footbridge between Hung Hom Station and POLYU. The PM<sub>2.5</sub> concentration on the footbridge surrounded by the highways of Granville and Chatham Road South was also very high. In contrast, a lower PM<sub>2.5</sub> concentration was found on the footbridge above the Science Museum path.

These results show that the distribution of PM<sub>2.5</sub> on the walkways in this high-density area was greatly affected by variables such as the traffic situation of the Hung Hom transport hub and cooking fumes. The three-dimensional shape and green vegetation distribution in the area also influence the distribution of PM<sub>2.5</sub>. Furthermore, the PM<sub>2.5</sub> concentrations on sidewalks, footpaths, and footbridges on the weekends were higher than those on weekdays, and the PM<sub>2.5</sub> concentrations on sidewalks,

footpaths, and footbridges during the morning peaks were higher than those during the evening peaks.

#### E. Developing the Healthiest Path Planning App

Based on the mapping results of PM<sub>2.5</sub> on walkable roads, a healthiest path planning app was developed, to provide the healthiest walking paths for PM<sub>2.5</sub> in the 500m area near Hong Kong Polytechnic University. In the process, it assumed that walking time between the edges was proportional to the distance. And the intensity of PM<sub>2.5</sub> pollution was proportional to the exposure time. Therefore, we multiplied the PM<sub>2.5</sub> concentration of the pixels along the edge by the length of the corresponding road segment, to directly get the PM<sub>2.5</sub> weighted value of each edge (Fig. 12). Then the healthiest path between the starting point and the endpoint was supported. Furthermore, the app calculated the distance traveled by the health path and the percentage of PM<sub>2.5</sub> pollution that can be reduced relative to the shortest path. Three typical examples were shown in

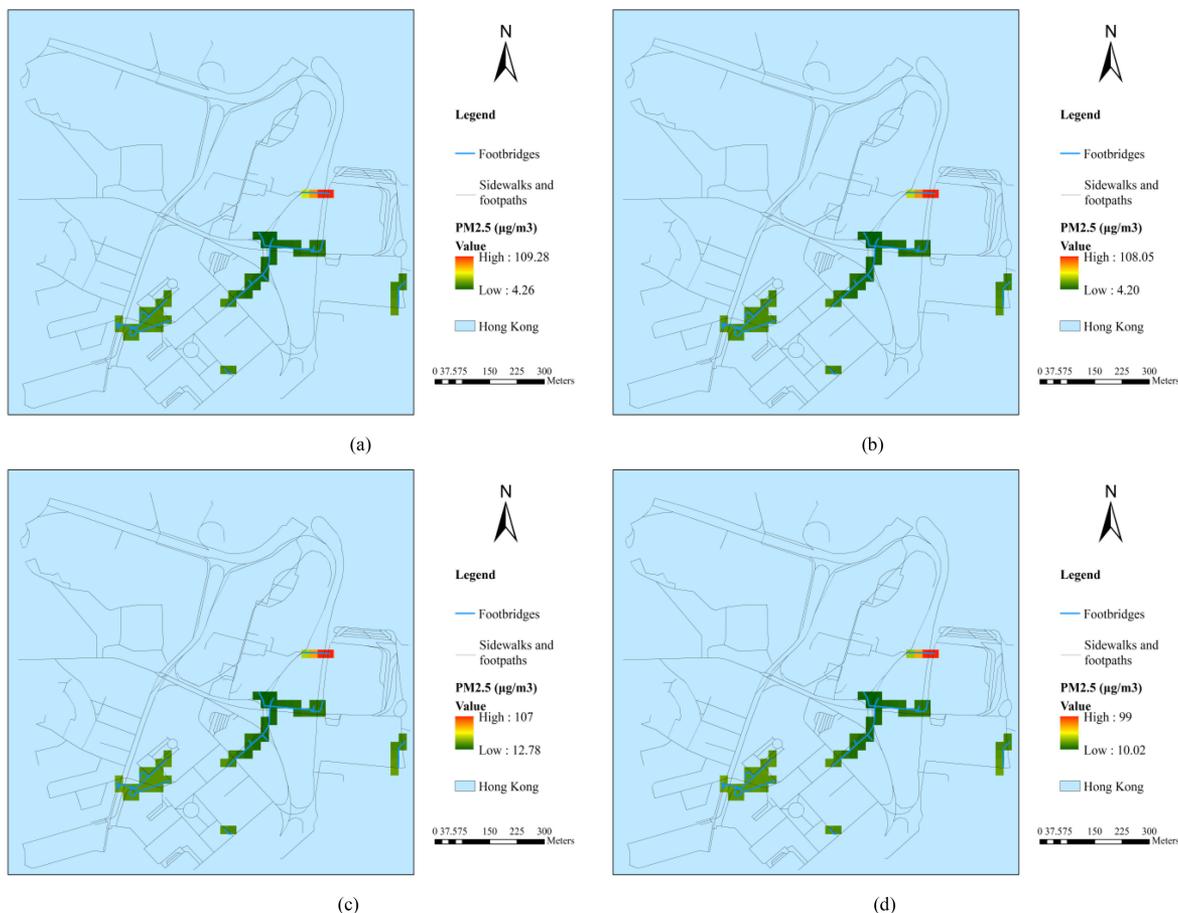


Fig. 11. Thematic maps of PM2.5 on footbridges during (a) the morning peak on the weekday, (b) the evening peak on the weekday, (c) the morning peak on the weekend, and (d) the evening peak on the weekend.

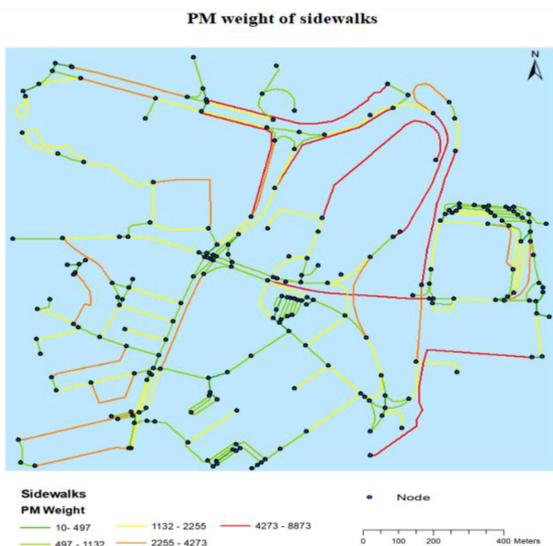


Fig. 12. PM2.5 weighted value of each sidewalk edge.

Fig. 13. The first example reflected the healthiest path from the HONG KONG Polytechnic University to the Peninsula Centre (the red path indicates the shortest walking path and

the green path indicates the least polluting path for PM2.5). We can see that the healthiest path has gone 80m more than the shortest path but the PM2.5 pollution on the healthiest path has decreased by 27.7%. The PM2.5 pollution on the healthiest paths of the other two examples has also decreased by 17% and 48.8%.

### V. DISCUSSION

The use of mobile measurement technology and the development of related LUR models have become prevailing trends in air pollution studies. However, previous studies were based on mobile sensors mounted on vehicles or bicycles to collect mobile measurements on main transport roads [1], [23], [43]–[45]. Therefore, this study was the first to provide PM2.5 LUR models combining handheld sensors data and multisource microscale geographic data on walkable roads in a high-density city by conducting a case study in Hong Kong, which complements the coverage and focus of such research. This article highlights the advantages and potential of the use of multiple handheld mobile sensors to easily collect large amounts of sampled data to improve modeling and mapping of microscale air pollution levels in high-density cities. These abilities provide an opportunity to improve the limitations of the existing data source. However,

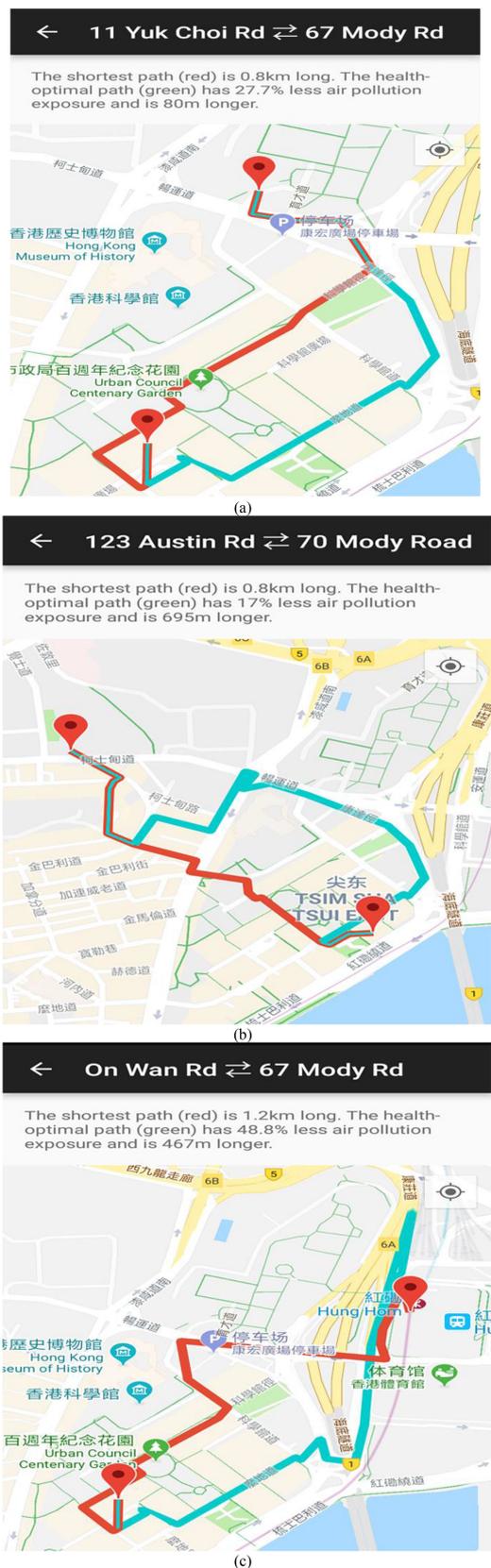


Fig. 13. Three typical examples of the healthiest path planning app.

the handheld sensors are weaker in terms of fineness and stability than the traditional instruments. We were thus required to acquire a large amount of sampling data and eliminate the uncertainty caused by accidental factors when pre-processing the data. Due to the uniqueness and complexity of Hong Kong's pedestrian network, the design of mobile measurement routes and data preprocessing needed to be adjusted according to the specific situation and scene of the selected research area. Furthermore, we had to carefully record the background environment and weather information, especially in the case of key locations and sudden extreme pollution. It is important to note that the mobile measurement data and methods of this study were specifically targeted at PM<sub>2.5</sub> contamination in the pedestrian environment of Hong Kong. If extended to other regions, the route design and data processing method of mobile measurement must be adjusted for the local environmental and weather characteristics. Moreover, nonlinear and nonparametric machine learning algorithms can be further used to improve the accuracy of estimating ground-level PM<sub>2.5</sub> [46]–[48].

Next, compared with traditional LUR models, this article for the first time explored the addition of the SVF factor to reflect the urban three-dimensional structure, the GVI factor to reflect the urban green vegetation distribution of the urban canopy and an urban building normalisation index (i.e., IBI). The results show that these new urban environmental factors from an open data source play a key role in the establishment of the corresponding LUR model. They also show that the three-dimensional morphology, green vegetation distribution and distribution of buildings have a great influence on the diffusion and variation of PM<sub>2.5</sub> in the walking environment of high-density cities.

In addition, the high-spatial-resolution PM<sub>2.5</sub> map obtained shows that the PM<sub>2.5</sub> concentration is higher in some areas of the microscale walking environment due to fumes from transport hubs and various restaurants. The three-dimensional shape of the urban buildings on both sides of the pedestrian road and the distribution of green vegetation also affect the diffusion of PM<sub>2.5</sub>. At the same time, a comparison of the PM<sub>2.5</sub> distribution during the morning and evening peaks and on weekdays and weekends shows relatively high PM<sub>2.5</sub> concentrations on sidewalks, footpaths and footbridges during the morning peak on weekdays. These findings can provide some references for travel and for the government's pollution prevention and control and urban planning.

## VI. CONCLUSION

This article was the first attempt to develop the framework for LUR models of PM<sub>2.5</sub> concentration in a microscale walkable environment of a high-density city. Based on the analysis of PM<sub>2.5</sub> variability for the various types of walking sections and background environments, LUR models were developed for sidewalks, footpaths, and footbridges. The accuracy of the LUR model in estimating PM<sub>2.5</sub> in Hong Kong's complex walking environment was improved by the integration of mobile measurement data, OSM data, Landsat images, and other

multi-source data. This study also produced PM<sub>2.5</sub> distribution maps with fine spatial granularity to provide a good reference for analysis of the PM<sub>2.5</sub> distribution on walkable roads in the microscale urban environment. These maps can be used to analyze the factors and hotspots of PM<sub>2.5</sub> changes in Hong Kong's pedestrian environment, which is a good complement to the Hong Kong government's existing roadside PM<sub>2.5</sub> monitoring network. Therefore, our research design and methods may be particularly well-suited for residential exposure in the walking environment and microscale refinement air pollution analysis, especially in areas in which no air monitoring network has been established. Some key findings are summarized as follows.

- 1) For various kinds of walkable roads in high-density cities, such as footpaths, footbridges, and sidewalks, it is feasible to use handheld mobile sensors and multisource microscale geographic factors for LUR modeling and mapping of microscale PM<sub>2.5</sub> distribution.
- 2) New environmental factors such as SVF, GVI, and IBI generated by open source data that represent three-dimensional morphology, green vegetation distribution, and building distribution can play key roles in LUR modeling of the PM<sub>2.5</sub> distribution in high-density cities. These indicators affect the diffusion of and changes in the PM<sub>2.5</sub> concentration in the walking environment.
- 3) The high spatial resolution maps of the PM<sub>2.5</sub> concentration obtained by LUR modeling can be used to find more pollution hotspots in microscale walking environments and to develop related travel-path search apps to provide healthier walking paths.

This article has three main limitations. First, the design of the mobile measurement route and the data preprocessing during modeling depended on the specific conditions and scenarios of the study area. Second, the precision and stability of the handheld sensors used in this study were relatively weak, which led to the need for significant re-sampling. Therefore, we hope that the transferability and applicable scope of the method adopted in this study will soon be tested in various typical areas of Hong Kong or other high-density cities. Third, a stepwise multiple linear regression method was used in the modeling stage of this study. As data accumulate and more impact factors are gradually taken into consideration over time, machine learning methods can also be considered to model changes in PM<sub>2.5</sub> concentrations on pedestrian roads.

## REFERENCES

- [1] F.K. S.Aammi and M. Petek, "A toxicological and genotoxicological indexing study of ambient aerosols (PM<sub>2.5</sub>-10) using in vitro bioassays," *Chemosphere*, vol. 174, pp. 490–498, 2017.
- [2] Y. Gao and H. Ji, "Microscopic morphology and seasonal variation of health effect arising from heavy metals in PM<sub>2.5</sub> and PM<sub>10</sub>: One-year measurement in a densely populated area of urban Beijing," *Atmos. Res.*, vol. 212, no. 30, pp. 213–226, 2018.
- [3] J. Z. Y. Lin, W. Yang, and C. Q. Li, "A review of recent advances in research on PM<sub>2.5</sub> in China," *Int. J. Environ. Res. Public Health*, vol. 15, no. 3, 2018.
- [4] Y. Yang *et al.*, "Changes in life expectancy of respiratory diseases from attaining daily PM<sub>2.5</sub> standard in China: A nationwide observational study," *Innov.*, vol. 1, no. 3, pp. 100064, 2020.
- [5] S. H. Tan, M. Roth, and E. Velasco, "Particle exposure and inhaled dose during commuting in Singapore," *Atmos. Environ.*, vol. 170, pp. 245–258, 2017.
- [6] A. P. P. Kumar, J. L. Durant, and H. C. Frey, "A review of factors impacting exposure to PM<sub>2.5</sub>, ultrafine particles and black carbon in Asian transport microenvironments," *Atmos. Environ.*, vol. 187, pp. 301–316, 2018.
- [7] S. Z. Sajani, P. L. F. Scotto, F. Galassi, and A. Montanari, "Urban air pollution monitoring and correlation properties between fixed-site stations," *J. Air Waste Manag. Assoc.*, vol. 54, no. 10, pp. 1236–1241, 2004, doi: 10.1080/10473289.2004.10470993.
- [8] V. I. M. Johnson, J. S. Touma, S. Mukerjee, and H. Özkaynak, "Evaluation of land-use regression models used to predict air quality concentrations in an urban area," *Atmos. Environ.*, vol. 44, no. 30, pp. 3660–3668, 2010.
- [9] E. G. Snyder *et al.*, "The changing paradigm of air pollution monitoring," *Environ. Sci. Technol.*, vol. 47, no. 20, pp. 11369–11377, 2013.
- [10] W. S. X. Wang, X. R. K. Zheng, and P. Han, "Estimating hourly PM<sub>2.5</sub> concentrations using MODIS 3 km AOD and an improved spatiotemporal model over Beijing-Tianjin-Hebei, China," *Atmos. Environ.*, vol. 222, 2020, Art. no. 117089.
- [11] W. Chen *et al.*, "Estimating PM<sub>2.5</sub> with high-resolution 1-km AOD data and an improved machine learning model over Shenzhen, China," *Sci. Total Environ.*, vol. 746, 2020, Art. no. 141093.
- [12] H. S. T. Li, Q. Yuan, and L. Zhang, "Geographically and temporally weighted neural networks for satellite-based mapping of ground-level PM<sub>2.5</sub>," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 178–188, 2020.
- [13] J. Wei *et al.*, "Reconstructing 1-km-resolution high-quality PM<sub>2.5</sub> data records from 2000 to 2018 in China: Spatiotemporal variations and policy implications," *Remote Sens. Environ.*, vol. 252, 2021, Art. no. 112136.
- [14] H. C. H. Y. Shi, Y. Xu, and E. Ng, "Improving satellite aerosol optical depth-PM<sub>2.5</sub> correlations using land use regression with microscale geographic predictors in a high-density urban context," *Atmos. Environ.*, vol. 190, pp. 23–34, 2018.
- [15] M. Yuan, Y. Song, Y. Huang, H. Shen, and T. Li, "Exploring the association between the built environment and remotely sensed PM<sub>2.5</sub> concentrations in urban areas," *J. Clean. Prod.*, vol. 220, pp. 1014–1023, 2019.
- [16] B. Hong, B. Lin, and H. Qin, "Numerical investigation on the effect of avenue trees on PM<sub>2.5</sub> dispersion in urban street canyons," *Atmos. (Basel)*, vol. 8, no. 7, 2017, Art. no. 129.
- [17] D. J. Briggs *et al.*, "Mapping urban air pollution using GIS: A regression-based approach," *Int. J. Geogr. Inf. Sci.*, vol. 11, no. 7, pp. 699–718, 1997.
- [18] R. Beelen *et al.*, "Development of NO<sub>2</sub> and NO<sub>x</sub> land use regression models for estimating air pollution exposure in 36 study areas in Europe—The ESCAPE project," *Atmos. Environ.*, vol. 72, no. 2, pp. 10–23, 2013.
- [19] J. D. Marshall, E. Nethery, and M. Brauer, "Within-urban variability in ambient air pollution: Comparison of estimation methods," *Atmos. Environ.*, vol. 42, no. 6, pp. 1359–1369, 2008.
- [20] J. C. M. Gao and E. Seto, "A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM<sub>2.5</sub> in Xi'an, China," *Environ. Pollut.*, vol. 199, pp. 56–65, 2015.
- [21] M. Rogulski, "Low-cost PM monitors as an opportunity to increase the spatiotemporal resolution of measurements of air quality," *Energy Procedia*, vol. 128, pp. 437–444, 2017.
- [22] N. Z. M. Masiol, D. C. Chalupa, D. Q. Rich, A. R. Ferro, and P. K. Hopke, "Hourly land-use regression models based on low-cost PM monitor data," *Environ. Res.*, vol. 167, pp. 7–14, 2018.
- [23] S. S. M. Masiol, D. Q. R. D. Chalupa, and P. K. Hopke, "Spatial-temporal variations of summertime ozone concentrations across a metropolitan area using a network of low-cost monitors to develop 24 hourly land-use regression models," *Sci. Total Environ.*, vol. 654, pp. 1167–1178, 2019.
- [24] J. G. Su, P. J. M. G. Allen, and M. Brauer, "Spatial modeling of residential wood smoke across a non-urban upstate New York region," *Air Qual. Atmos. Heal.*, vol. 6, no. 1, pp. 85–94, 2013.
- [25] S. Hankey and J. D. Marshall, "Land use regression models of on-road particulate air pollution (particle number, black carbon, PM<sub>2.5</sub>, particle size) using mobile monitoring," *Environ. Sci. Technol.*, vol. 49, no. 15, pp. 9194–9202, 2015.
- [26] C. C. Lim *et al.*, "Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea," *Environ. Int.*, vol. 131, pp. 105022, 2019.

- [27] S. Zhou and R. Lin, "Spatial-temporal heterogeneity of air pollution: The relationship between built environment and on-road PM<sub>2.5</sub> at micro scale," *Transp. Res. Part D Transp. Environ.*, vol. 76, pp. 305–322, 2019.
- [28] R. S. T. Althoff, J. L. Hicks, A. C. King, S. L. Delp, and J. Leskovec, "Large-scale physical activity data reveal worldwide activity inequality," *Nature*, vol. 547, no. 7663, pp. 336–339, 2017.
- [29] Y. Gu *et al.*, "Impacts of sectoral emissions in China and the implications: Air quality, public health, crop production, and economic costs," *Environ. Res. Lett.*, vol. 13, no. 8, 2018, Art. no. 084008.
- [30] K. K. L. Lau, Y. Shi, and E. Ng, "Developing street-level PM<sub>2.5</sub> and PM<sub>10</sub> land use regression models in high-density Hong Kong with urban morphological factors," *Environ. Sci. Technol.*, vol. 50, no. 15, pp. 8178–8187, 2016.
- [31] Y. Yang *et al.*, "Long term exposure to air pollution and mortality in an elderly cohort in Hong Kong," *Environ. Int.*, vol. 117, pp. 99–106, 2018.
- [32] N. G. J. Tryner, M. L. C. A. Wilson, J. L. Peel, and J. Volckens, "Variation in gravimetric correction factors for nephelometer-derived estimates of personal exposure to PM<sub>2.5</sub>," *Environ. Pollut.*, vol. 250, pp. 251–261, 2019.
- [33] J. L. A. Middel, M. D. R. Maciejewski, and M. Roth, "Sky view factor footprints for urban climate modeling," *Urban Clim.*, vol. 25, pp. 120–134, 2018.
- [34] J. Yang *et al.*, "Development of an improved urban emissivity model based on sky view factor for retrieving effective emissivity and surface temperature over urban areas," *ISPRS J. Photogramm. Remote Sens.*, vol. 122, pp. 30–40, 2016.
- [35] X. Li, C. Zhang, W. Li, R. Ricard, Q. Meng, and W. Zhang, "Assessing street-level urban greenery using Google street view and a modified green view index," *Urban For. Urban Green.*, vol. 14, no. 3, pp. 675–685, 2015.
- [36] Y. Long and L. Liu, "How green are the streets? An analysis for central areas of Chinese cities using Tencent street view," *PLoS One*, vol. 12, no. 2, pp. 1–18, 2017.
- [37] R. C. Estoque and Y. Murayama, "Classification and change detection of built-up lands from Landsat-7 ETM+ and Landsat-8 OLI/TIRS imageries: A comparative assessment of various spectral indices," *Ecol. Indic.*, vol. 56, pp. 205–217, 2015.
- [38] H. Xu, "A new index for delineating built-up land features in satellite imagery," *Int. J. Remote Sens.*, vol. 29, no. 14, pp. 4269–4276, 2008.
- [39] J. Gulliver, K. deHoogh, G. Hoek, D. Vienneau, D. Fecht, and A. Hansell, "Back-extrapolated and year-specific NO<sub>2</sub> land use regression models for Great Britain—Do they yield different exposure assessment?," *Environ. Int.*, vol. 92–93, no. 2, pp. 202–209, 2016.
- [40] Z. S. I. Poboříková and M. Michalková, "Application of four probability distributions for wind speed modeling," *Procedia Eng.*, vol. 192, pp. 713–718, 2017.
- [41] L. M. Zwack, C. J. Paciorek, J. D. Spengler, and J. I. Levy, "Modeling spatial patterns of traffic-related air pollutants in complex urban terrain," *Environ. Health Perspect.*, vol. 119, no. 6, pp. 852–859, 2011.
- [42] J. W. L. Li, C. S. N. Hudda, S. A. Fruin, and R. J. Delfino, "Modeling the concentrations of on-road air pollutants in Southern California," *Environ. Sci. Technol.*, vol. 47, no. 16, pp. 9291–9299, 2013.
- [43] A. Smargiassi *et al.*, "A spatiotemporal land-use regression model of winter fine particulate levels in residential neighbourhoods," *J. Expo. Sci. Environ. Epidemiol.*, vol. 22, no. 4, pp. 331–338, 2012.
- [44] B. S. Beckerman *et al.*, "A hybrid approach to estimating national scale spatiotemporal variability of PM<sub>2.5</sub> in the contiguous United States," *Environ. Sci. Technol.*, vol. 47, no. 13, pp. 7233–7241, 2013.
- [45] U. S. Y. Su, L. W. J. Debosz, and A. Munoz, "Multi-year continuous PM<sub>2.5</sub> measurements with the federal equivalent method SHARP 5030 and comparisons to filter-based and TEOM measurements in Ontario, Canada," *Atmos. (Basel)*, vol. 9, no. 5, pp. 1–13, 2018.
- [46] J. V. D. Tuia, F. P.-C. L. Alonso, and G. Camps-Valls, "Multioutput support vector regression for remote sensing biophysical parameter estimation," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 804–808, 2011.
- [47] A. K. S. Ahmad, and H. Stephen, "Estimating soil moisture using remote sensing data: A machine learning approach," *Adv. Water Resour.*, vol. 33, no. 1, pp. 69–80, 2010.
- [48] N. S. W. Zhao, H. Lu, and A. Li, "A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression," *J. Hydrol.*, vol. 563, pp. 1009–1024, 2018.



**Chengzhuo Tong** received the M.S. degree in geomatics from The Hong Kong Polytechnic University, Hong Kong, in 2019. He is currently working toward the Ph.D. degree in urban informatics with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University.

His current research interests include urban informatics, and spatial epidemiology.



**Zhicheng Shi** received the Ph.D. degree in geomatics from The Hong Kong Polytechnic University, Hong Kong, China, in 2020.

He is currently a Postdoc with Research Institute for Smart Cities, School of Architecture and Urban Planning, Shenzhen University, Shenzhen, China. His current research interests in spatial big data analysis.



**Wenzhong Shi** received the Ph.D. degree in geomatics from University of Osnabrück, Vechta, Germany, in 1994.

He is an Otto Poon Charitable Foundation Professor in Urban Informatics, Chair Professor in GISci and Remote Sensing, and Director with smart cities Research Institute, The Hong Kong Polytechnic University, Hong Kong. His current research interests include urban informatics and smart cities, GISci and remote sensing, intelligent analytics and quality control for spatial data, artificial-intelligence-based

object extraction and change detection from satellite imagery, and mobile mapping and 3-D modeling based on LiDAR and remote sensing imagery. He has published over 250 academic papers that are indexed by SCI and 15 books.



**Pengxiang Zhao** received the M.S. and Ph.D. degrees in geomatics from Wuhan University, Wuhan, China, in 2012 and 2015, respectively.

He is currently a Researcher with GIS Center, Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden. His research interests include geographic information science, urban mobility, and trajectory data analysis and mining.



**Anshu Zhang** received the B.Sc. degree in geoinformation technology and the Ph.D. degree in geographic information systems from The Hong Kong Polytechnic University, Hong Kong, China, in 2011 and 2017, respectively.

She is currently a Research Assistant Professor with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University. Her research interests include spatial data mining, human mobility modeling, and prediction, with emphasis on improving the robustness and reliability of

data analytics.

Dr. Zhang was the Secretary of Working Group II/1, The International Society for Photogrammetry and Remote Sensing from 2012 to 2016. She has been a Referee of *International Journal of Geographical Information Science* and *ACM Transactions on Data Science*.