



Article

# A Multiscale Land Use Regression Approach for Estimating Intraurban Spatial Variability of PM<sub>2.5</sub> Concentration by Integrating Multisource Datasets

Yuan Shi <sup>1,\*</sup> , Alexis Kai-Hon Lau <sup>2,3,4</sup> , Edward Ng <sup>1,5,6</sup>, Hung-Chak Ho <sup>7</sup> and Muhammad Bilal <sup>8</sup>

<sup>1</sup> Institute of Future Cities (IOFC), The Chinese University of Hong Kong, Hong Kong, China

<sup>2</sup> Division of Environment and Sustainability, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China; alau@ust.hk

<sup>3</sup> Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

<sup>4</sup> Institute for the Environment, The Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong, China

<sup>5</sup> School of Architecture, The Chinese University of Hong Kong, Hong Kong, China; edwardng@cuhk.edu.hk

<sup>6</sup> Institute of Environment, Energy and Sustainability (IEES), The Chinese University of Hong Kong, Hong Kong, China

<sup>7</sup> Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China; hcho21@hku.hk

<sup>8</sup> Lab of Environmental Remote Sensing (LERS), School of Marine Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China; muhammad.bilal@connect.polyu.hk

\* Correspondence: shiyuan@link.cuhk.edu.hk



**Citation:** Shi, Y.; Lau, A.K.-H.; Ng, E.; Ho, H.-C.; Bilal, M. A Multiscale Land Use Regression Approach for Estimating Intraurban Spatial Variability of PM<sub>2.5</sub> Concentration by Integrating Multisource Datasets. *Int. J. Environ. Res. Public Health* **2022**, *19*, 321. <https://doi.org/10.3390/ijerph19010321>

Academic Editor: Paul B. Tchounwou

Received: 18 November 2021

Accepted: 28 December 2021

Published: 29 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Poor air quality has been a major urban environmental issue in large high-density cities all over the world, and particularly in Asia, where the multiscale complex of pollution dispersal creates a high-level spatial variability of exposure level. Investigating such multiscale complexity and fine-scale spatial variability is challenging. In this study, we aim to tackle the challenge by focusing on PM<sub>2.5</sub> (particulate matter with an aerodynamic diameter less than 2.5 μm,) which is one of the most concerning air pollutants. We use the widely adopted land use regression (LUR) modeling technique as the fundamental method to integrate air quality data, satellite data, meteorological data, and spatial data from multiple sources. Unlike most LUR and Aerosol Optical Depth (AOD)-PM<sub>2.5</sub> studies, the modeling process was conducted independently at city and neighborhood scales. Correspondingly, predictor variables at the two scales were treated separately. At the city scale, the model developed in the present study obtains better prediction performance in the AOD-PM<sub>2.5</sub> relationship when compared with previous studies ( $R^2$  from 0.72 to 0.80). At the neighborhood scale, point-based building morphological indices and road network centrality metrics were found to be fit-for-purpose indicators of PM<sub>2.5</sub> spatial estimation. The resultant PM<sub>2.5</sub> map was produced by combining the models from the two scales, which offers a geospatial estimation of small-scale intraurban variability.

**Keywords:** PM<sub>2.5</sub>; spatial variability; geographic information system; multiscale; multi-source datasets

## 1. Introduction

Poor air quality has been a major urban environmental issue in cities, especially those large and compact cities in Asia, for the last several decades [1]. Urbanization alters the local climate, thus contributing to ambient air pollution levels in cities [2]. The interaction between the urban environment and air pollution dispersion is a complex multiscale mechanism [3]. The spatial scales of atmospheric pollution can range from a few hundred meters for urban street canyons to a few hundred kilometers for a whole Megalopolis [4]. At the mesoscale, the spatial layout of artificial urban land cover types changes the near-surface aerodynamic roughness, which interacts with the atmospheric circulation, and

consequently, alters the transportation of air pollutants [5]. At the city scale, urban land use planning changes the spatial distribution of the emission sources of air pollutants. As a result, the air quality situation in large cities is not homogeneously distributed. A spatially heterogeneous urban environment (different combinations of building density and functions, road network, open space, etc.) makes the concentration levels of air pollutants vary from place to place [6]. At the neighborhood scale, the geometrical shapes of building clusters, street canyons, the arrangement of street trees, and vegetation are all influential factors of street-level air quality [7,8]. The above multiscale phenomenon creates a high-level spatial variability of pollution exposure levels in the urban environment. Therefore, investigating the multiscale complexity and fine-scale spatial variability of air pollution is challenging [9].

There are two types of air pollution modeling methods—deterministic and statistical [10]. Deterministic methods (numerical and computational modeling) are most commonly used in the modeling of mesoscale or city-scale air pollution. In the last few decades, various numerical and computational models have been developed from the perspective of atmospheric chemistry and physics [11,12]. Most of the Lagrangian and Eulerian grid dispersion models provide hourly-resolved fine temporal resolution but at a coarse spatial resolution of several kilometers [13,14]. Computational fluid dynamics (CFD) enables the accurate simulation of the movement of air pollutants at a very fine spatiotemporal scale. CFD is mostly used for air pollution investigation in street canyons [15] or city blocks [16] as it can only cover a limited spatial extent due to the high demands for computational powers and time costs. However, for air quality management and health risk assessment nowadays, the air pollution data needs to have both a relatively large spatial extent (i.e., an entire city) and a sufficiently high spatial resolution (sometimes, with a grid cell size of fewer than one hundred meters) to adequately capture small-scale intraurban spatial variability [17]. With geographical information systems (GIS), statistical methods offer data with a finer spatial resolution to help cater to the above need.

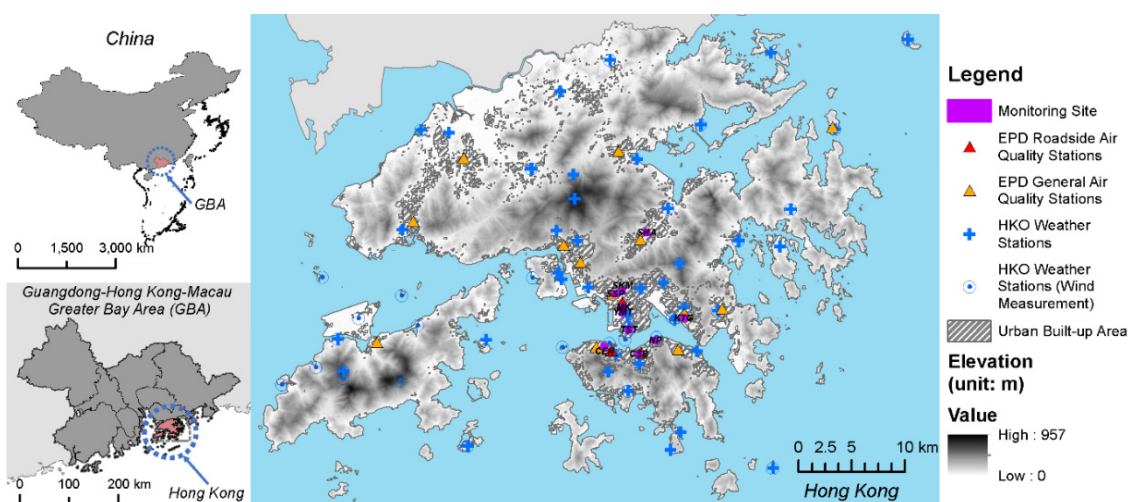
Statistical methods employ and combine in-situ monitored data from various sources, which can provide concise and quantitative information on spatiotemporal variability. Most of the routine air quality monitoring networks have long-term air quality data at an hourly-resolved temporal scale. Mobile monitoring techniques based on portable sensors capture fine-scale spatial variability [18,19]. AOD-ground-level PM<sub>2.5</sub> correlation analysis is also a popular method [20,21] as the Satellite-derived aerosol observations provide extensive spatial coverage and temporal continuity [22]. Land-use regression (LUR) models from exposure science research provide fit-for-purpose predictions on long-term concentration at a fine spatial scale [23–26]. By combining the above data sources and techniques, studies on the development of exposure models have been conducted and show promising results [27,28]. Moreover, developing multiscale exposure models is also found to be a feasible way of capturing the intraurban spatial variability in air pollution distribution [29]. Such models can estimate air pollution concentration levels for large-scale spatial coverage while capturing small-scale spatial variability, which not only provides useful information on spatial exposure but also enables a quantitative understanding of the multiscale influence of urbanization on air pollution dispersion in cities. In addition, it also helps with the identification of influential factors of intraurban spatial variability of air pollution. In this study, aiming at tackling this multiscale complexity in the spatial variability of air pollution affecting pedestrians' exposure levels, we propose a multiscale LUR model and offer a geospatial estimation of small-scale intraurban variability of air pollution. Particularly, we focus on exposure to PM<sub>2.5</sub> (particulate matter with an aerodynamic diameter of less than 2.5 μm). A geospatial approach—multiscale LUR modeling that combines various types of multiscale data sources was adopted to estimate the intraurban spatial variability of PM<sub>2.5</sub> exposure in the compact built environment using the city of Hong Kong as a study area. Unlike most LUR and AOD-PM<sub>2.5</sub> studies, the modeling process was conducted independently at city and neighborhood scales. Correspondingly, predictor variables at the two scales were treated separately. Moreover, point-based building morphological indices

and road network centrality metrics were also examined in the modeling process, which is rare in other existing LUR studies.

## 2. Materials and Methods

### 2.1. Study Area

Hong Kong (Figure 1), a highly urbanized large city with serious air pollution, was selected as the testbed city for developing multiscale LUR models of  $PM_{2.5}$  exposure. Hong Kong is one of the world's most compact cities with a large population of over seven million people living in a total land area of 1100 km<sup>2</sup>. Its population density is around 6700 persons/km<sup>2</sup>. Like most other large cities in the world, Hong Kong also has air pollution issues. The highly complex and diverse urban environment makes Hong Kong an ideal study area for the development of multiscale LUR models.



**Figure 1.** The air quality monitoring network, weather stations for wind measurement, and site selection in this study.

### 2.2. $PM_{2.5}$ and AOD Data

#### 2.2.1. Satellite-Derived AOD Data at City Scale

In this study, MODIS Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD (MCD19A2) was used to represent the spatial distribution of surface  $PM_{2.5}$  concentration over the study area corresponding to the period of ground monitoring and sampling (described in Sections 2.2.2 and 2.2.3). Compared with the 3-km MODIS product [30], MAIAC has a satisfying accuracy at a finer spatial resolution (1-km) and higher spatial coverage of retrieval [31]. Therefore, MAIAC AOD is more suitable for air quality studies, notably, it is better at providing details of fine-scale aerosol characteristics over study areas with heterogeneous geographic context. A QA filter should be applied to select the best-quality MAIAC AOD data [31]. Most applications use the filter of QA.CloudMask = "Clear" in order to ensure a high quality of the retrieved AOD. However, it has been identified that in some locations with high spatial aerosol variability, this filter may systematically erase AOD retrievals in cloud-free conditions over certain urban areas [32]. In such cases, the following filters to define high-quality AOD retrievals: QA.CloudMask = "Clear" or "Possibly cloudy". Furthermore, pixels with the filter of QA for AOD = "AOD within  $\pm 2$  km from the coastline" were also included to extend the spatial coverage of prediction, as there are major urban built-up areas in Hong Kong within 2 km from the coastline (can be seen in Figure 1). MAIAC jointly processes MODIS Terra and Aqua sensors as a single sensor due to the Terra-to-Aqua cross-calibration of the entire C6 MODIS data collection [33]. In the data collection, most daily files contain multiple orbit overpasses for combined Terra and Aqua. Data from Aqua and Terra were combined

and averaged to provide an estimation of the daily average AOD value [34]. Chemical transport models, AOD validation, and inter-comparison commonly use the input which is standardized to 0.55  $\mu\text{m}$  [35]. Therefore, to keep the consistency, AOD retrieved at 0.55  $\mu\text{m}$  was selected by this study.

### 2.2.2. Long-term PM<sub>2.5</sub> Monitoring Data at City Scale

During the study period (years 2018 and 2019), data from a total of 16 stations were available in the local air quality monitoring network in the study area (Figure 1). The stations in the network are located at representative places that cover various combinations of land use, road traffic network, natural topography (information shown in Table S1 of the Supplementary Materials). These stations are operated and regularly maintained by the local authority—Hong Kong Environmental Protection Department (HKEPD). The data archive of the hourly concentration level of PM<sub>2.5</sub> is openly available to the public. The monitoring data from the network are readily comparable to air quality data from other cities, as the network has been recognized by the United Nations Environmental Programme (UNEP) and complies with international standards [36]. Therefore, we selected this data source for investigating the city-scale spatial variability of PM<sub>2.5</sub>.

### 2.2.3. In-Situ PM<sub>2.5</sub> Sampling at Neighborhood Scale

**Sampling site selection:** The city of this study is a large city with a compact and complex urban environment. The spatial variability in air pollution affecting pedestrians' exposure levels in such a spatially heterogeneous city context cannot be fully represented by the sparsely located 16 air quality stations operated in the local air quality monitoring network. Therefore, this study selected ten study sites for air pollution sampling based on not only land use and natural topography but also district block function, building density and morphology, and road network layout. The ten sampling sites selected are representative of the diverse urban contexts of Hong Kong. The details of the ten selected sites are shown in Table S2 of the Supplementary Materials.

**Sampling routes:** The present study adopted a PM<sub>2.5</sub> sampling method which was tested in a previous pilot study on small-scale spatial variability of pedestrian level particulate matter in three downtown commercial districts [37]. To investigate the spatial variability of PM<sub>2.5</sub> exposure levels, one person kept strolling and rambling in a selected site area at a common pedestrian walking speed of 3 km/h (0.8 m/s) [38] with a backpack sampling unit (see the next paragraph—Sampling instrumentation). The sampling routes are shown in Figure S1 of the Supplementary Materials.

**Sampling instrumentation:** The instrument used in the previous study is an all-in-one backpack sampling unit assembled by combining various monitors and sensors. In the present study, the instrument was further improved, becoming smaller in size and more lightweight. The concentration levels of PM<sub>2.5</sub> were continuously monitored using the TSI SidePak<sup>TM</sup> personal aerosol monitor AM520 (a more portable laser-scattering device that is dedicatedly designed for assessing personal exposure to PM) instead of the DUSTTRAK<sup>TM</sup> model 8534 monitor. The AM520 was set to sample and log PM<sub>2.5</sub> concentration data with the time interval of 1 s (with this sampling interval, the spatial interval of sampling data points was approximately 1 m which is a fine resolution for spatial mapping and exposure assessment). A sample extension tube was connected to the inlet of AM520 to extend the inlet to the height of 1.6 m above the ground. During the measurement, the person who carried the backpack sampling unit stayed away from any random pollution sources nearby (e.g., smoking people, roadside vendors) in order to minimize random noise introduced in the data. Ambient air temperature ( $T_a$ , °C) and relative humidity ( $RH$ , %) were synchronously sampled by a set of Testo 480 data logger and humidity/temperature probe. The  $RH$  data were used for the sampled PM<sub>2.5</sub> data calibration. The AM520 sampled data were corrected using synchronously sampled  $RH$  based on the following equation [39]:

$$\text{Correction Factor} = 1 + 0.25 \frac{RH^2}{(1 - RH)} \quad (1)$$

The Global Positioning System (GPS) receiver was used to label the geolocation of each sample in the WGS 1984 coordinate system. The data logging time stamps were synchronized to Coordinated Universal Time (UTC).

*Sampling periods:* During the warm season (May to September) of Hong Kong, local emission sources dominate the local condition of PM<sub>2.5</sub> [40], which minimizes the overwhelming effects of high background concentration levels caused by the long-distance transport of non-local emission sources. Therefore, in-situ measurements were conducted during the warm season (May to September) of the year 2018 and 2019 (the measurement date are shown in Table S2 of the Supplementary Materials). This experimental design simulated a scenario of typical city pedestrians' outdoor activities. The sampling campaign was conducted in three different time slots (9 a.m.–11 a.m., 2 p.m.–4 p.m., and 7 p.m.–9 p.m.) each day to cover different periods and situations of traffic and human activities in a day. To investigate the spatial variability, day-to-day and hour-to-hour temporal variation in the measurement, data were adjusted based on the background air quality monitoring station of Hong Kong using the adjustment method that has been tested and adopted in a previous study of vehicle-based mobile measurement [41].

### 2.3. Spatial and Temporal Predictor Data

In this study, multiple data sources were collated to a predictor variables dataset for the multiscale LUR modeling process. Table 1 summarizes the predictor variables and the corresponding data sources used in the multiscale LUR modeling of this study.

**Table 1.** Summary of the predictor variables and the corresponding data sources used in the multiscale LUR modeling in the present study.

Data Type	Predictor Variables	Unit	Abbr.	Raw Data Source	Spatial Scale
Weather data	Air temperature	°C	TEMP	Historical records are publicly accessible from local authorities of weather monitoring—Hong Kong Observatory (HKO)	Temporal-resolved variable with city-scale spatial variability
	Relative humidity	%	RH		
	Wind speed	m/s	WSPD		
	Rainfall	mm	RF		
	Mean sea level pressure	hPa	MSLP		
Atmospheric soundings	Sounding indices examined in this study are listed in Table S3 of the Supplementary Material			Wyoming Weather Web	Temporal variable
Land use (Areal fraction of the land use type within certain circular buffer range <sup>1</sup> )	Residential land	%	RES	Derived from the open data from the Planning Department of Hong Kong (PlanD)	City scale
	Commercial land	%	COM		
	Industrial land	%	IND		
	Government land	%	GOV		
	Open space land	%	OPN		
	Greening cover ratio	%	GCR	Calculated based on above data	City scale
Geolocation of air quality monitoring stations and weather stations	Longitude	degree	LONG	The GeoInfo Map of Hong Kong	City scale
	Latitude	degree	LAT		
	Elevation above sea level	m	ELEV		
Population	Population density <sup>1</sup>	People/km <sup>2</sup>	POP	WorldPop Global Project Population Data	City scale

**Table 1.** *Cont.*

Data Type	Predictor Variables	Unit	Abbr.	Raw Data Source	Spatial Scale
Road network density (For road: Line density of the land use type within certain circular buffer range <sup>1</sup> ; For bus station: the total number of stations within buffer)	Trunk road/expressways	km/km <sup>2</sup>	RD0	Spatial data layers extracted from Open Street Map (OSM)	City scale
	Primary road	km/km <sup>2</sup>	RD1		
	Secondary road	km/km <sup>2</sup>	RD2		
	Tertiary road	km/km <sup>2</sup>	RD3		
	Ordinary road	km/km <sup>2</sup>	RD4		
	Bus stations	–	BUS		
Road segment attributes (a set of metrics commonly used as a proxy of the road usage and traffic distribution)	Normalized straightness	–	STRAIGHT	Calculation based on network centrality analysis	Neighborhood scale (each road segment corresponds to a value)
	Normalized betweenness	–	BETWEEN		
	Normalized closeness	–	CLOSE		
	Connectivity	–	CONNECT		
	Control value	–	CONTROL	Calculation based on Spatial Syntax	
	Mean depth	–	MDEPTH		
	Global integration	–	GINTEG		
	Local integration	–	LINTEG		
Building morphological data	Frontal area index <sup>1</sup>	–	FAI	Calculated from the building dataset produced by Ren, et al. [42]	City scale
	Point-based FAI	–	FAI <sub>Point</sub>		Neighborhood scale
	Sky view factor	[0–1]	SVF		Neighborhood scale
	Surface roughness length	m	ROUGHNESS		City scale

<sup>1</sup> For this spatial predictor variable, multiple values were calculated using a series of circular buffer radius: 50 m, 100 m, 200 m, 300 m, 400 m, 500 m, 750 m, 1000 m, 1500 m, 2000 m.

### 2.3.1. Weather Data and Sounding Data at City Scale

Incorporating meteorological information for exposure analysis helps with explaining the small-scale spatial variability in air pollution [43]. Particularly, incorporating interpolated observed wind information from a weather station network improves the model prediction performance [44,45]. In this study, near-surface meteorological data (i.e., air temperature, relative humidity, wind speed, rainfall, mean sea level pressure) were retrieved from the Hong Kong Observatory (HKO) weather monitoring network (Figure 1). Kriging, as one of the mostly-used geostatistical methods for meteorological applications [46], was performed to generate spatially interpolated meteorological data layers that cover the entire study area using a fine spatial resolution of 10 m, which is consistent with the land use information of Hong Kong introduced in the next section.

### 2.3.2. Land-Use, Population, Road Network at City Scale

At the city scale, the distribution of residential zones, population density, the location of industrial and commercial point sources, traffic-related line sources on road networks are all factors that influence the PM<sub>2.5</sub> spatial variability. The circular buffering method was used for quantifying the above influential factors and creating predictor variables (e.g., the total area of industrial zones, road network density, or population density within the circular buffer, etc.). The predictor variables were calculated in a series of circular buffers around the location of each long-term monitoring station. The land-use zoning information is publicly available from the local city planning authority—Planning Department (in the form of a land use map at 10 m resolution). The original complicated land-use zoning was reclassified to a five-class scheme: residential zone, commercial zone, industrial zone, governmental and facilities, greening, and open space, which makes the present study

consistent and comparable to existing studies [24]. Spatial data of urban road networks, location of car parks, bus stops, and other transport facilities were extracted from Open Street Map (OSM). The population density map at 100 m resolution was derived from the WorldPop dataset [47].

### 2.3.3. Road Segment Centrality and Accessibility as Proxy of Traffic Distribution

Traffic distribution is usually measured as the volume of vehicles passing through a road segment. A higher traffic volume leads to more traffic-related pollutant emissions. However, obtaining a full record of traffic volume that covers all road segments of a large network usually needs intensive resources or technical input. Regression modeling has been enlisted as a more economically feasible way of estimating traffic distribution [48,49], in which topological measures are an essential set of predictors. The network centrality, as a key factor of network analysis, has been used to investigate the characteristics of the urban roadways [50] and employed in the estimation of traffic distribution [51,52]. The Space Syntax approach is another road network topology analysis method based on the accessibility of road segments that has been used in understanding the traffic as well [53–55]. Based on the literature, in this study, we selected and calculated three centrality metrics and five spatial syntax metrics (as shown in Table 1) for all road segments in the study area using the Urban Network Analysis Tool [56] and Axwoman [57], respectively. These metrics were used as predictor variables in the neighborhood-scale modeling.

### 2.3.4. Building Morphological Data at Neighborhood Scale

In the highly urbanized context, urban ventilation depends largely on the geometrical characteristics of buildings, which have a greater influence on the spatial variability of urban pollution [58]. A previous local study found that the pollutant concentration in street canyons had a strong and significant correlation with specific building morphological parameters [41]. In this study, building morphological parameters at both city-scale and neighborhood-scale were adopted for building multiscale models. At the city scale, the frontal area index (*FAI*), as a commonly-used wind direction-dependent measure of evaluating urban ventilation and a widely recognized influencing factor of air quality [59], was calculated using the following equation:

$$FAI = \sum_{\theta=1}^8 \left[ \frac{A_{F(\theta)}}{A_T} \right] \cdot P(\theta) \quad (2)$$

where,  $A_T$  is the lot area of the land parcel.  $A_{F(\theta)}$  is the total projected area along a particular wind direction ( $\theta$ ) of all buildings in the land parcel.  $P(\theta)$  is the wind direction probability of the eight corresponding principal wind directions ( $\theta$ ). Similar to other city-scale predictor variables, the averaged *FAI* value within circular buffers around each long-term monitoring station was calculated. At the neighborhood scale, two point-based building morphological parameters, which are sky view factor (*SVF*) [60] and point-based *FAI* [61] were calculated for all recorded locations of the in-situ  $PM_{2.5}$  sampling data. *SVF* has been found associated with air quality [62] and calculated in the present study by using the method proposed by Dozier and Frew [63] which is shown in the equation below:

$$SVF = \frac{1}{2\pi} \int_0^{2\pi} \left[ \cos \beta \cos^2 \varphi + \sin \beta \cdot \cos(\Phi - \alpha) \cdot (90 - \varphi - \sin \varphi \cos \varphi) \right] d\Phi \quad (3)$$

where *SVF* is calculated for each location of the digital surface model (DSM, a raster layer contains elevation value of buildings and ground surface at each pixel) of sampling sites with slope aspect  $\alpha$  and angle  $\beta$  based on the horizon angles  $\varphi$  in azimuth directions  $\Phi$  of the hemisphere circle with the radius  $d$ . Point-based *FAI* ( $FAI_{point}$ ) is a fine-scale building morphological parameter developed for the fit-for-purpose estimation of the pedestrian-level wind speed at a high spatial resolution within urban areas [61]. Compared with conventional computational fluid dynamics (CFD) simulation,  $FAI_{point}$  is much more cost-effective as it can estimate pedestrian-level wind speed for a large spatial extent (i.e., the

entire city) in almost real-time. Wind speed has been commonly regarded as a proxy of pollutants concentration level in the investigation of spatial variability in exposure [64] as stagnation of air in compact building clusters is often associated with elevated concentration of air pollutants and increased exposure to air pollution [65,66].  $FAI_{Point}$  is an extension of  $FAI$ . Therefore, their calculation shares common ground. It is calculated using the equation below:

$$FAI_{Point} = \iint w_l \left[ \sum_{\theta=1}^8 \frac{A_{F(\theta,x,y)}}{A_{T,circle}} \cdot P(\theta) \right] dl / A_{T,circle} \quad (4)$$

$$w_l = [0, 1]$$

$$w_l = \left( \frac{R-l}{R} \right)^c \quad (5)$$

$$c = 2, l = [0, 200]$$

where  $A_{T,circle}$  is the circular area using a radius of  $R = 200$  m and the location of the test point  $(x, y)$  as the center of the circle.  $A_{F(\theta,x,y)}$  is calculated for each location in the circular area. An exponential ( $c = 2$ ) decay weighting factor (Equation (5)) is applied to represent the decreasing effect of the roughness elements along with the increase in the distance  $l$  between roughness elements and the test point [67].

#### 2.4. Multiscale Land Use Regression Modeling

As mentioned in the introduction of the study, we aimed to develop a multiscale LUR model which enables a cost-efficient geospatial estimation of small-scale intraurban variability of intraurban air pollution for the study area. Therefore, the LUR modeling was conducted at two different scales. First, we performed city-scale LUR modeling based on long-term monitoring and city-scale predictors. This first stage LUR model provides an overview of the seasonal average concentration level of  $PM_{2.5}$  over the spatial extent. Then, we performed the second-stage LUR at a finer spatial scale using the in-situ  $PM_{2.5}$  sampling data neighborhood-scale predictors. The second-stage LUR model explains the fine-scale spatial variability which was used to downscale the city-scale spatial  $PM_{2.5}$  map from the first stage LUR modeling.

##### 2.4.1. City-Scale LUR Modeling and Mapping

The first stage LUR modeling provided an overview of the spatial distribution of the four seasonal average concentration levels of  $PM_{2.5}$  for the city. The development of city-scale LUR models consisted of three steps, which were (i) variable selection, (ii) regression modeling, and (iii) spatial mapping. At this step, instead of conventional multiple linear regression (MLR), we adopted geographically and temporally weighted regression (GTWR) modeling by using the predictor variable subset derived above. GTWR considers spatial and temporal variability, as it has been found that geographically weighted models perform better than MLR in the investigation of the effects of land use on urban air pollution variations [68–70]. The general model structure can be represented as follows:

$$PM_{2.5ij} \sim (\alpha_0 + \beta_{ij}) + (\alpha_1 + \beta_{ij}) \times AOD_{ij} + \alpha_{2ij}P_{0ij} + \alpha_{3ij}P_{1ij} + \dots + \alpha_{k+2ij}P_{kij} + \varepsilon \quad (6)$$

where  $AOD_{ij}$  is the observed AOD at location  $i$  on time  $j$ , which is forced to be included in the model.  $(\alpha_0 + \beta_{ij})$  is the intercept of the model.  $(\alpha_1 + \beta_{ij})$  is the slope of  $AOD_{ij}$ .  $\alpha_{2ij}, \alpha_{3ij}, \dots, \alpha_{k+2ij}$  are the slopes for city-scale predictors  $P_{0ij}, P_{1ij}, \dots, P_{kij}$  (that are derived from weather data, land use, population, and road network density) at location  $i$  on time  $j$ , which are all geographically and temporally varied.  $\varepsilon$  is residuals. First, all city-scale spatial predictor variables described in Table 1 were examined to screen out a subset of variables that could result in the best-performing model. Generally, the selection of buffer radius is based upon the decay correlation with the modeled pollutant [24]. The distance-decay curve method [71] that has always been widely used so far [72,73] was adopted by the present study to identify appropriate spatial predictor variables and the corresponding associated buffer distance. Specifically, the buffer radius with the highest correlation coefficient between the  $PM_{2.5}$  level and spatial predictor variables was identified, and



the variables at their corresponding buffer were selected for further stepwise regression analysis. Then, the spatial predictor variables at their optimal buffer and all temporal variables were put together as the input for a stepwise regression analysis to derive a subset of variables that could result in the best-performing model. In this study, the stepwise regression yielded the following rules: (1)  $p$ -value  $< 0.05$  for all variables included in the model; (2) variance inflation factor (VIF)  $< 3$  for all variables to minimize the collinearity and overfitting issues. The predictor subset derived from the above stepwise regression was used to build seasonal GTWR models. Finally, based on the spatiotemporal GTWR models, we generated the seasonal averaged  $PM_{2.5}$  spatial map. AOD data (with an original resolution of 1 km) was smoothed with the bilinear interpolation to keep the resolution consistent with the land use data layers and interpolated meteorological data layers. The modeling and mapping results are shown in Section 3.1.

#### 2.4.2. Neighborhood-Scale LUR Modeling

To incorporate the neighborhood-scale spatial variability of  $PM_{2.5}$  concentration, we further developed the second stage LUR model using the data from in-situ sampling at the ten sites (refers to Section 2.2.3). All neighborhood-scale spatial variables (mentioned in Table 1) were used as predictors. The same variable selection method and criteria used in the first stage modeling were also adopted here. Instead of predicting the absolute value of in-situ sampling  $PM_{2.5}$  data, the neighborhood-scale LUR modeling focused on investigating the variation among different points in a site. Therefore, the variation in  $PM_{2.5}$  ( $\Delta PM_{2.5}$ ) was used as the response variable.  $\Delta PM_{2.5}$  is the difference between the measured  $PM_{2.5}$  concentration at a point and the site averaged  $PM_{2.5}$  (which is the mean of all measurement points in the entire site). The  $\Delta PM_{2.5}$  value will be positive if the measured concentration at a point is higher than the site average, vice versa. The model structure is as follows:

$$\Delta PM_{2.5} \sim \beta + \alpha_1 P_1 + \alpha_2 P_2 + \alpha_3 P_3 + \dots + \alpha_n P_n + \varepsilon \quad (7)$$

where  $\alpha_1, \alpha_2, \dots, \alpha_n$  are the slopes for city-scale predictors  $P_1, P_2, \dots, P_n$ . As to the regression modeling, we used MLR instead of GTWR for the development of the neighborhood-scale model, as the in-situ  $PM_{2.5}$  sampling data mainly reflects the microscale effect. The microscale effect that happens in street canyons depends on the environmental condition of the specific road segment, which is supposed to be geolocation-independent. The  $\Delta PM_{2.5}$  was calculated for all road segments. A road segment was treated as a line segment for the calculation of centrality and accessibility (described in Section 2.3.3), and the calculated values were spatially assigned to all locations within the street canyon corresponding to the road segment in the process of spatial mapping. As the model estimates the variation in  $PM_{2.5}$  instead of the absolute concentration value, the resultant map had a data distribution across positive and negative. It is important to note that the regression model is only valid within the spatial range of road and street area (the centrality and accessibility variables representing traffic were only calculated at the road network) and the numerical range of input predictor variables (no “extrapolation” [74] was done for the model to reduce the uncertainty). The resultant map reflects the variation at a finer spatial scale, but it does not have full spatial coverage of the city. The valid part of  $PM_{2.5}$  variability estimation was overlaid together with the city-scale  $PM_{2.5}$  map from Section 2.4.1 in order to provide the final  $PM_{2.5}$  concentration map.

### 3. Results

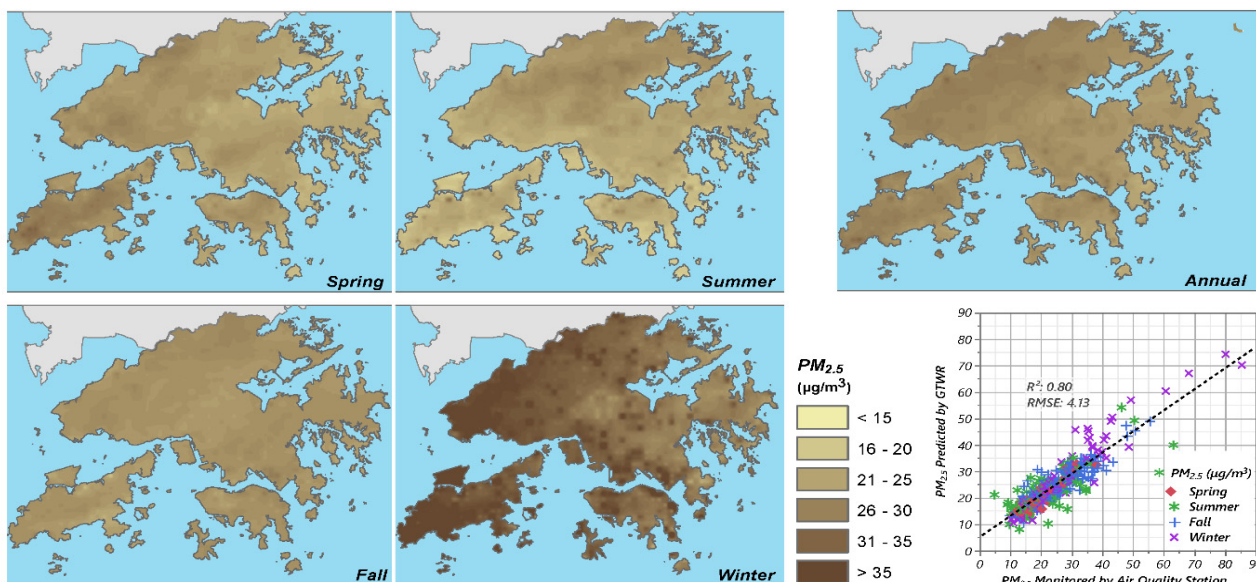
#### 3.1. City-Scale LUR Modeling and Mapping Results

The section reports the results from the first stage city-scale LUR. The LUR results include the resultant GTWR models and the seasonal maps of  $PM_{2.5}$ . The structures of the four seasonal GTWR models are shown in Table 2, respectively. The estimation maps and corresponding regression plots are shown in Figure 2. The predictor variables selected by stepwise regression mainly reflect features that affect the  $PM_{2.5}$  level from the following three aspects: spatial location, built environment density, meteorological condition. For

GTWR models, the coefficients were geographically and temporally varied (please refer to Table S4 of the supplementary material). AICc was calculated, which is the corrected version of the Akaike information criterion (AIC) [75]. Leave-one-out cross-validation (CV) which is commonly used for GWR-based model validation [76], was also performed to validate the model performance.

**Table 2.** Summary of the resultant GTWR model structures. The math operators before each predictor variable indicate its correlation with PM<sub>2.5</sub>. “+” indicates a positive correlation; “−” indicates a negative correlation. Model coefficients are shown in Table S4 of the supplementary material.

Season	Model Structure	R <sup>2</sup>	$\overline{R^2}$	CV R <sup>2</sup>	AICc
Spring	PM <sub>2.5</sub> ~ AOD − LONG + KINX − PWAT	0.881	0.835	0.831	206.368
Summer	PM <sub>2.5</sub> ~ AOD + FAI250 + LAT − KINX − PWAT	0.566	0.504	0.497	445.211
Fall	PM <sub>2.5</sub> ~ AOD + RES500 + BUS400 + TEMP − WSPD + KINX − PWAT	0.772	0.694	0.673	1145.994
Winter	PM <sub>2.5</sub> ~ AOD − LONG + ROUGHNESS50 + CINV + LCLP + LFCV + VTOT	0.898	0.853	0.846	579.051
Annual	A piecewise linear function is the combination of four seasonal models. Only one of the four models will be selected based on the time.	0.798	0.792	N.A.	N.A.



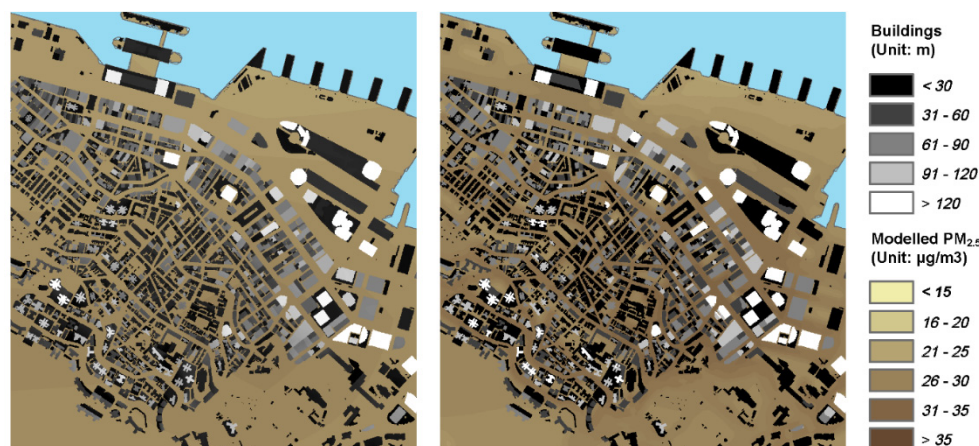
**Figure 2.** The four resulting maps of seasonal average and the map of the annual average of PM<sub>2.5</sub> concentration level. The concentration value of 35 µg/m<sup>3</sup> is the annual limit in Air Quality Objectives (AQOs) set out by the Air Pollution Control Ordinance (Cap. 311) of Hong Kong. The inset picture at the bottom right corner shows the actual predicted plot of the resulting models of PM<sub>2.5</sub> concentration level.

### 3.2. Neighborhood-Scale LUR Modeling Results

The neighborhood-scale regression modeling was conducted by following the steps as stated in Section 2.4.2. The model structure and predictor coefficients are shown in Table 3. The model had a performance of  $\overline{R^2} = 0.508$ , with all predictor variables having VIF < 4. The numerical range and summary statistics of input predictor variables of the neighborhood-scale LUR model are shown in Table S5 of the Supplementary Material. Figure 3 shows the zoom-in city-scale resulting PM<sub>2.5</sub> map of the downtown area with estimated fine-scale variability in PM<sub>2.5</sub> ( $\Delta PM_{2.5}$ ) overlaid. More details of road space PM<sub>2.5</sub> spatial variability are reflected.

**Table 3.** Summary of the resultant neighborhood-scale regression model.

Predictor Variables	Coefficients	Significant Level	VIF
Intercept	$-2.925 \times 10^{-1}$	<0.0001	
Normalized betweenness	$1.431 \times 10^2$	<0.0001	1.009
Normalized closeness	$3.074 \times 10^5$	<0.0001	3.332
Control value	$-1.675 \times 10^{-1}$	<0.0001	2.050
Global integration	3.749	<0.0001	3.657
Sky view factor	$-1.263 \times 10^1$	<0.0001	1.959
Point-based <i>FAI</i>	2.826	<0.0001	1.716

**Figure 3.** The zoom-in city-scale resulting  $PM_{2.5}$  map of the downtown area (left) with estimated fine-scale variability in  $PM_{2.5}$  ( $\Delta PM_{2.5}$ ) overlaid (right).

## 4. Discussion

### 4.1. At the City-Scale

Built environment density determines the pollution emission as a denser area usually has a large population, and also affects the pollutant dispersal as the buildings alter the wind flow in the city. Indeed, a few studies indicate that a higher population density could mediate the balance between the usage of private and public transport [77], foster the implementation of cleaner technology [78], thereby reducing pollution emission. However, in most cases, evidence still shows that high population density is a main factor driving the accumulation of air pollution [79,80], which seems to be more in line with common perceptions. Such perception also appeared in the results of the present study. A statistically significant positive correlation between  $PM_{2.5}$  and residential land use area as well as the number of bus stops was observed in the resultant GTWR models. Spatial location, as another influential aspect, also showed in most models. Longitude was negatively correlated with  $PM_{2.5}$ , while the correlation was opposite for Latitude. Such a pattern indicates a higher level of  $PM_{2.5}$  at northwest territories of Hong Kong, which represents the well-known air quality impact at the regional level [40]. The seasonal spatial maps also clearly showed the seasonal variability (Figure 2) [81]. The *FAI* calculated within a buffer radius of 250 m was identified as a major influential factor during summertime, which indicates the dominant effect of local building density and geometrical factors on the pollution dispersal in summer. Similar findings are also shown in other CFD simulation-based studies on the relationship between urban form and air pollution [82]. The critical buffer radius of 250 m for *FAI* is consistent with the findings from a previous study based on the long-term mobile ground-level air quality monitoring using the public transport vehicle platform [19]. The meteorological condition was also influential. Upper-air sounding indices related to atmospheric stability showed up in all seasonal models and shared the same relationship where higher atmospheric instability corresponded to a lower level of pollution concentration, vice versa. The reasonable explanation of the statistical model

based on theoretical fundamental indicates the interpretability of the models. Compared with the annual model performance ( $\overline{R^2} = 0.72$ ) in our previous study on calibrating the AOD-PM<sub>2.5</sub> relationship [68], the city-scale LUR model achieved better overall prediction performance ( $\overline{R^2} = 0.80$ ). Such a result further supports the conclusions of previous studies that the MODIS collection 6 MAIAC algorithm is more suitable than the 3 km MODIS AOD product for intraurban air quality studies [83]. Comparison between the present study results and the previous study results using 3 km MODIS also showed that the long-term PM<sub>2.5</sub> concentration in Hong Kong during the period of 2018 to 2019 dropped by 46% compared to the years of 2003-2015 (with a reduction in annual average PM<sub>2.5</sub> over the spatial extent of the entire city from 45.8  $\mu\text{g}/\text{m}^3$  to 24.5  $\mu\text{g}/\text{m}^3$ ). This reduction indicates the effectiveness and necessity of regional cooperative actions in air pollution control [84].

#### 4.2. At the Neighborhood-Scale

In our previous study on calibrating the AOD-PM<sub>2.5</sub> relationship [68], we mixed and put all predictor variables at various spatial scales into one single model. Although the previous models achieved moderate prediction performance, the models may still ignore the differences in the influence of predictors at various spatial scales on the dependent variables. The PM<sub>2.5</sub> variations we captured during the in-situ sampling data are mainly determined by the environmental condition of the road segment. Therefore, the neighborhood-scale PM<sub>2.5</sub> variation is supposed to be largely geolocation-independent, and thus should be investigated separately. In the present study, point-based building morphological indices were included in the model. It was found that point-based building morphological indices and road network centrality and accessibility metrics as predictors explained more than 50% of the variability in the in-situ PM<sub>2.5</sub> sampling data. As indicated by the results shown in Table 3, road segments with higher connectivity and accessibility were found to be correlated to a higher level of PM<sub>2.5</sub> concentration, as they usually hold more commercial activities and generally serve a larger traffic volume. The model also indicated that locations with higher openness generally correspond to lower levels of PM<sub>2.5</sub> concentrations.

#### 4.3. Limitations and Future Work

There is a limitation on AOD data filtering which should be noticed. As mentioned in Section 2.2.1, the AOD data used in this study is the MAIAC AOD. It has a finer resolution compared with the 3-km MODIS dataset but has more unusable pixels after applying the AOD quality filter. Usually, the QA filter “Best quality”, which combines the two filters: QA.CloudMask = “Clear” and QA.AdjacencyMask = “Clear”, should be used for most cases [31]. In this study, pixels with the filter of QA.CloudMask = “Possibly cloudy” were used to avoid the possible retrieval issue over certain urban areas [32]. In addition, pixels with the filter of QA for AOD = “AOD within  $\pm 2$  km from the coastline” were retained to extend the spatial coverage of prediction. The above filtering scheme possibly introduced unreliable AOD data to the models. In that case, future work should focus on evaluating the uncertainties of the model. Moreover, future work might also be conducted in other testbed cities, which could acquire data outside the present study area and externally validate the present modeling and mapping results. Currently, the GTWR models are developed separately for different seasons. In future work, the model could be developed for various weather types by classifying meteorological conditions of pollutant dispersal. Another limitation of the present study is that the traffic was considered by introducing road network topological measures as the proxy. Although such methods have been employed by existing transport studies, using the traffic volume data from a computational traffic simulation conducted by transport engineers could significantly reduce the uncertainties and increase the neighborhood-scale regression model performance. Moreover, the regression model is only valid within the numerical range of input predictor variables as no model extrapolation was performed. Therefore, the estimation of neighborhood-scale PM<sub>2.5</sub> variability does not have full spatial coverage of the city.

## 5. Conclusions

Investigating the multiscale complexity of spatial variability in air pollution is challenging, especially in compactly built environments. In this study, we tackled such multiscale complexity in the spatial variability of air pollution affecting pedestrian exposure levels by conducting in-situ sampling and multiscale LUR modeling of PM<sub>2.5</sub>, in the testbed city Hong Kong. Using the land-use regression (LUR) modeling technique as the fundamental method, air quality data, satellite data, meteorological data, and spatial data from multiple sources were integrated. We used a different strategy from existing LUR and AOD-PM<sub>2.5</sub> research, which was to conduct the modeling at city- and neighborhood scales independently and examine the predictor variables at the two scales separately. At the city scale, the model developed in the present study obtained better prediction performance in the exploration of the AOD-PM<sub>2.5</sub> relationship when compared with previous studies. At neighborhood scale, the point-based building morphological indices and road network centrality and accessibility metrics were found to be fit-for-purpose indicators of PM<sub>2.5</sub> spatial estimation. The resultant PM<sub>2.5</sub> map was developed by combining the models from the two scales. The map offers an estimation of intraurban variability of air pollution which facilitates local public health research. More importantly, working with practitioners and stakeholders, it is possible to convert the building morphological factors to urban planning and design language which could be incorporated into existing practice guidelines. Therefore, the quantitative information on the relationship between spatial predictors (e.g., building morphological factors) and PM<sub>2.5</sub> concentration level derived from the models are interpretable and transferable to evidence-based strategies for the improvement in the built environment. This will enable knowledge transfer and potentially could inform the strategic planning direction of “planning for a livable high-density city” which is defined in the governmental initiative of planning vision and strategy of “Hong Kong 2030 +” by the Planning Department of Hong Kong [85]. More importantly, the methods and workflows proposed by the research have worldwide applicability, as all the necessary input data are not difficult to acquire in most regions.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ijerph19010321/s1>, Figure S1: The maps and sampling routes in the ten study sites, Table S1: Detailed information of the 16 stations in the air quality monitoring network in Hong Kong, Table S2: The ten study sites selected for the PM<sub>2.5</sub> sampling at neighborhood scale, Table S3: List of all atmospheric soundings indices examined in the stepwise regression modeling of this study, Table S4: Summary of coefficients of the four seasonal GTWR models, Table S5: The numerical range and summary statistics of input predictor variables of the neighborhood scale LUR model.

**Author Contributions:** Conceptualization, Y.S. and E.N.; methodology, Y.S.; formal analysis, Y.S.; investigation, Y.S.; resources, E.N.; writing—original draft preparation, Y.S.; writing—review and editing, E.N., A.K.-H.L., H.-C.H. and M.B.; supervision, E.N.; funding acquisition, E.N. and A.K.-H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is supported by the General Research Fund (GRF No.14610717) from the Research Grants Council (RGC) of Hong Kong.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors appreciate reviewers for their insightful comments and constructive suggestions on our research work. The authors also want to thank editors for their patient and meticulous work on our manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Coria, J.; Köhlin, G.; Xu, J. On the Use of Market-Based Instruments to Reduce Air Pollution in Asia. *Sustainability* **2019**, *11*, 4895. [[CrossRef](#)]
2. Jandaghian, Z.; Akbari, H. The Effect of Increasing Surface Albedo on Urban Climate and Air Quality: A Detailed Study for Sacramento, Houston, and Chicago. *Climate* **2018**, *6*, 19. [[CrossRef](#)]
3. Itahashi, S. Air Pollution Modeling: Local, Regional, and Global-Scale Applications. *Atmosphere* **2021**, *12*, 178. [[CrossRef](#)]
4. Stern, A.C. *Fundamentals of Air Pollution*; Elsevier: Amsterdam, The Netherlands, 2014.
5. Brown, M. *Urban Parameterizations for Mesoscale Meteorological Models*; Wit Press: Southampton, UK, 2000; pp. 193–255.
6. Azarov, V.; Barikaeva, N.; Solovyeva, T. Monitoring of fine particulate air pollution as a factor in urban planning decisions. *Procedia Eng.* **2016**, *150*, 2001–2007. [[CrossRef](#)]
7. Di Sabatino, S.; Buccolieri, R.; Pulvirenti, B.; Britter, R. Flow and pollutant dispersion in street canyons using fluent and adms-urban. *Environ. Model. Assess.* **2008**, *13*, 369–381. [[CrossRef](#)]
8. Janhäll, S. Review on urban vegetation and particle air pollution—Deposition and dispersion. *Atmos. Environ.* **2015**, *105*, 130–137. [[CrossRef](#)]
9. Zlatev, Z.; Dimov, I. *Computational and Numerical Challenges in Environmental Modelling*; Elsevier: Amsterdam, The Netherlands, 2006; Volume 13.
10. Zannetti, P. *Air Pollution Modeling: Theories, Computational Methods, and Available Software*; Computational Mechanics Publications: New York, NY, USA, 1990; Volume 20.
11. Zlatev, Z.; Brandt, J.; Builtjes, P.J.; Carmichael, G.; Dimov, I.; Dongarra, J.; Van Dop, H.; Georgiev, K.; Hass, H.; San José, R. *Large Scale Computations in Air Pollution Modelling*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 57.
12. Holmes, N.S.; Morawska, L. A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available. *Atmos. Environ.* **2006**, *40*, 5902–5928. [[CrossRef](#)]
13. Knowlton, K.; Rosenthal, J.E.; Hogrefe, C.; Lynn, B.; Gaffin, S.; Goldberg, R.; Rosenzweig, C.; Civerolo, K.; Ku, J.-Y.; Kinney, P.L. Assessing ozone-related health impacts under a changing climate. *Environ. Health Perspect.* **2004**, *112*, 1557–1563. [[CrossRef](#)] [[PubMed](#)]
14. Leelőssy, Á.; Molnár, F.; Izsák, F.; Havasi, Á.; Lagzi, I.; Mészáros, R. Dispersion modeling of air pollutants in the atmosphere: A review. *Open Geosci.* **2014**, *6*, 257. [[CrossRef](#)]
15. Tominaga, Y.; Stathopoulos, T. Cfd modeling of pollution dispersion in a street canyon: Comparison between les and rans. *J. Wind Eng. Ind. Aerodyn.* **2011**, *99*, 340–348. [[CrossRef](#)]
16. Kalampounias, A.G.; Stamatis, C.; Kaltsonoudis, C.; Bali, N.; Kalarakis, A.; Pandis, S.N. *Mapping the Air Quality in the City of Patras by Combining Mobile Platform Measurements and Cfd Simulations*; Springer International Publishing: Cham, Switzerland, 2017; pp. 1001–1006.
17. Jones, A.; Thomson, D.; Hort, M.; Devenish, B. *The U.K. Met Office's Next-Generation Atmospheric Dispersion Model, Name III*; Springer: Boston, MA, USA, 2007; pp. 580–589.
18. Hatzopoulou, M.; Valois, M.F.; Levy, I.; Mihele, C.; Lu, G.; Bagg, S.; Minet, L.; Brook, J. Robustness of land-use regression models developed from mobile air pollutant measurements. *Environ. Sci. Technol.* **2017**, *51*, 3938–3947. [[CrossRef](#)]
19. Shi, Y.; Xie, X.; Fung, J.C.-H.; Ng, E. Identifying critical building morphological design factors of street-level air pollution dispersion in high-density built environment using mobile monitoring. *Build Environ.* **2018**, *128*, 248–259. [[CrossRef](#)]
20. Di Nicolantonio, W.; Cacciari, A.; Petritoli, A.; Carnevale, C.; Pisoni, E.; Volta, M.; Stocchi, P.; Curci, G.; Bolzacchini, E.; Ferrero, L.; et al. MODIS and OMI satellite observations supporting air quality monitoring. *Radiat. Prot. Dosim.* **2009**, *137*, 280–287. [[CrossRef](#)] [[PubMed](#)]
21. Fan, Z.; Zhan, Q.; Yang, C.; Liu, H.; Bilal, M. Estimating PM<sub>2.5</sub> Concentrations Using Spatially Local Xgboost Based on Full-Covered SARA AOD at the Urban Scale. *Remote Sens.* **2020**, *12*, 3368. [[CrossRef](#)]
22. Chudnovsky, A.; Lyapustin, A.; Wang, Y.; Tang, C.; Schwartz, J.; Koutrakis, P. High resolution aerosol data from modis satellite for urban air quality studies. *Open Geosci.* **2014**, *6*, 17–26. [[CrossRef](#)]
23. Knibbs, L.D.; Hewson, M.G.; Bechle, M.J.; Marshall, J.D.; Barnett, A.G. A national satellite-based land-use regression model for air pollution exposure assessment in Australia. *Environ. Res.* **2014**, *135*, 204–211. [[CrossRef](#)]
24. Ryan, P.H.; LeMasters, G.K. A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhal. Toxicol.* **2007**, *19*, 127–133. [[CrossRef](#)] [[PubMed](#)]
25. Alvarez-Mendoza, C.I.; Teodoro, A.C.; Torres, N.; Vivanco, V. Assessment of Remote Sensing Data to Model PM<sub>10</sub> Estimation in Cities with a Low Number of Air Quality Stations: A Case of Study in Quito, Ecuador. *Environments* **2019**, *6*, 85. [[CrossRef](#)]
26. Yang, X.; Zheng, Y.; Geng, G.; Liu, H.; Man, H.; Lv, Z.; He, K.; de Hoogh, K. Development of PM<sub>2.5</sub> and NO<sub>2</sub> models in a LUR framework incorporating satellite remote sensing and air quality model data in Pearl River Delta region, China. *Environ. Pollut.* **2017**, *226*, 143–153. [[CrossRef](#)]
27. Lee, H.J.; Chatfield, R.B.; Strawa, A.W. Enhancing the applicability of satellite remote sensing for pm2.5 estimation using modis deep blue aod and land use regression in california, united states. *Environ. Sci. Technol.* **2016**, *50*, 6546–6555. [[CrossRef](#)] [[PubMed](#)]
28. Kloog, I.; Koutrakis, P.; Coull, B.A.; Lee, H.J.; Schwartz, J. Assessing temporally and spatially resolved pm2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos. Environ.* **2011**, *45*, 6267–6275. [[CrossRef](#)]

29. Ma, X.; Longley, I.; Gao, J.; Kachhara, A.; Salmond, J. A site-optimised multi-scale GIS based land use regression model for simulating local scale patterns in air pollution. *Sci. Total Environ.* **2019**, *685*, 134–149. [CrossRef] [PubMed]
30. Remer, L.A.; Mattoo, S.; Levy, R.C.; Munchak, L.A. Modis 3 km aerosol product: Algorithm and global perspective. *Atmos. Meas. Tech.* **2013**, *6*, 1829–1844. [CrossRef]
31. Lyapustin, A.; Wang, Y.; Korkin, S.; Huang, D. Modis collection 6 maiaac algorithm. *Atmos. Meas. Tech.* **2018**, *11*, 5741–5765. [CrossRef]
32. Just, A.C.; Wright, R.O.; Schwartz, J.; Coull, B.A.; Baccarelli, A.A.; Tellez-Rojo, M.M.; Moody, E.; Wang, Y.; Lyapustin, A.; Kloog, I. Using high-resolution satellite aerosol optical depth to estimate daily pm2.5 geographical distribution in Mexico City. *Environ. Sci. Technol.* **2015**, *49*, 8576–8584. [CrossRef] [PubMed]
33. Lyapustin, A.; Wang, Y.; Xiong, X.; Meister, G.; Platnick, S.; Levy, R.; Franz, B.; Korkin, S.; Hilker, T.; Tucker, J. Scientific impact of modis c5 calibration degradation and c6+ improvements. *Atmos. Meas. Tech.* **2014**, *7*, 4353–4365. [CrossRef]
34. Sherman, J.; Gupta, P.; Levy, R.; Sherman, P. An Evaluation of MODIS-Retrieved Aerosol Optical Depth over a Mountainous AERONET Site in the Southeastern US. *Aerosol Air Qual. Res.* **2017**, *16*, 3243–3255. [CrossRef]
35. Lyapustin, A.; Wang, Y. *Modis Multi-Angle Implementation of Atmospheric Correction (Maiaac) Data User's Guide*; NASA: Greenbelt, MD, USA, 2018.
36. HKEPD. Air Quality Monitoring Network of Hong Kong. Available online: <http://www.aqhi.gov.hk/en/monitoring-network/air-quality-monitoring-network.html> (accessed on 9 October 2021).
37. Shi, Y.; Ng, E. Fine-scale spatial variability of pedestrian-level particulate matters in compact urban commercial districts in Hong Kong. *Int. J. Environ. Res. Public Health* **2017**, *14*, 1008. [CrossRef]
38. Penwarden, A.D.; Wise, A.F.E. *Wind Environment around Buildings*; HMSO: Richmond, UK, 1975.
39. Ramachandran, G.; Adgate, J.L.; Pratt, G.C.; Sexton, K. Characterizing indoor and outdoor 15 minute average pm 2.5 concentrations in urban neighborhoods. *Aerosol Sci. Technol.* **2003**, *37*, 33–45. [CrossRef]
40. Lau, A.; Lo, A.; Gray, J.; Yuan, Z.; Loh, C. *Relative Significance of Local vs. Regional Sources: Hong Kong's Air Pollution*; Civic Exchange: Hong Kong, China, 2007.
41. Shi, Y.; Lau, K.K.-L.; Ng, E. Developing street-level pm2.5 and pm10 land use regression models in high-density hong kong with urban morphological factors. *Environ. Sci. Technol.* **2016**, *50*, 8178–8187. [CrossRef]
42. Ren, C.; Cai, M.; Li, X.; Shi, Y.; See, L. Developing a rapid method for 3-dimensional urban morphology extraction using open-source data. *Sustain. Cities Soc.* **2020**, *53*, 101962. [CrossRef]
43. Su, J.G.; Brauer, M.; Ainslie, B.; Steyn, D.; Larson, T.; Buzzelli, M. An innovative land use regression model incorporating meteorology for exposure analysis. *Sci. Total Environ.* **2008**, *390*, 520–529. [CrossRef]
44. Arain, M.A.; Blair, R.; Finkelstein, N.; Brook, J.R.; Sahsuaroglu, T.; Beckerman, B.; Zhang, L.; Jerrett, M. The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies. *Atmos. Environ.* **2007**, *41*, 3453–3464. [CrossRef]
45. Naughton, O.; Donnelly, A.; Nolan, P.; Pilla, F.; Misstear, B.D.; Broderick, B. A land use regression model for explaining spatial variation in air pollution levels using a wind sector based approach. *Sci. Total Environ.* **2018**, *630*, 1324–1334. [CrossRef]
46. Waller, L.; Gotway, C. *Applied Spatial Statistics for Public Health Data*; John Wiley & Sons: Hoboken, NJ, USA, 2004.
47. Tatem, A.J. Worldpop, open data for spatial demography. *Sci. Data* **2017**, *4*, 1–4. [CrossRef] [PubMed]
48. Zhao, F.; Chung, S. Contributing factors of annual average daily traffic in a florida county: Exploration with geographic information system and regression models. *Transp. Res. Rec.* **2001**, *1769*, 113–122. [CrossRef]
49. Selby, B.; Kockelman, K.M. Spatial prediction of traffic levels in unmeasured locations: Applications of universal kriging and geographically weighted regression. *J. Transp. Geogr.* **2013**, *29*, 24–32. [CrossRef]
50. Crucitti, P.; Latora, V.; Porta, S. Centrality measures in spatial networks of urban streets. *Phys. Rev. E* **2006**, *73*, 036125. [CrossRef]
51. Jayasinghe, A.; Sano, K.; Nishiuchi, H. Explaining traffic flow patterns using centrality measures. *Int. J. Traffic Transp. Eng.* **2015**, *5*, 134–149. [CrossRef]
52. Jayasinghe, A.; Sano, K.; Abenayake, C.C.; Mahanama, P.K.S. A novel approach to model traffic on road segments of large-scale urban road networks. *MethodsX* **2019**, *6*, 1147–1163. [CrossRef]
53. Zhang, X.; Ren, A.; Chen, L.; Zheng, X. Measurement and spatial difference analysis on the accessibility of road networks in major cities of china. *Sustainability* **2019**, *11*, 4209. [CrossRef]
54. Giannopoulou, M.; Roukounis, Y.; Stefanis, V. Traffic network and the urban environment: An adapted space syntax approach. *Procedia—Soc. Behav. Sci.* **2012**, *48*, 1887–1896. [CrossRef]
55. Patterson, J.L. Traffic modelling in cities—validation of space syntax at an urban scale. *Indoor Built Environ.* **2016**, *25*, 1163–1178. [CrossRef]
56. Sevtsuk, A.; Mekonnen, M. Urban network analysis. A new toolbox for arcgis. *Rev. Int. Géomatique* **2012**, *22*, 287–305. [CrossRef]
57. Jiang, B. *Axwoman 6.3: An Arcgis Extension for Urban Morphological Analysis*; University of Gävle: Gävle, Sweden, 2015.
58. Yuan, C.; Ng, E.; Norford, L.K. Improving air quality in high-density cities by understanding the relationship between air pollutant dispersion and urban morphologies. *Build. Environ.* **2014**, *71*, 245–258. [CrossRef]
59. Su, J.G.; Brauer, M.; Buzzelli, M. Estimating urban morphometry at the neighborhood scale for improvement in modeling long-term average air pollution concentrations. *Atmos. Environ.* **2008**, *42*, 7884–7893. [CrossRef]

60. Watson, I.D.; Johnson, G.T. Graphical estimation of sky view-factors in urban environments. *J. Climatol.* **1987**, *7*, 193–197. [[CrossRef](#)]
61. Yuan, C.; Norford, L.; Britter, R.; Ng, E. A modelling-mapping approach for fine-scale assessment of pedestrian-level wind in high-density cities. *Build. Environ.* **2016**, *97*, 152–165. [[CrossRef](#)]
62. Krüger, E.L.; Minella, F.O.; Rasia, F. Impact of urban geometry on outdoor thermal comfort and air quality from field measurements in Curitiba, Brazil. *Build. Environ.* **2011**, *46*, 621–634. [[CrossRef](#)]
63. Dozier, J.; Frew, J. Rapid calculation of terrain parameters for radiation modeling from digital elevation data. *IEEE Trans. Geosci. Remote Sens.* **1990**, *28*, 963–969. [[CrossRef](#)]
64. Clougherty, J.E.; Wright, R.J.; Baxter, L.K.; Levy, J.I. Land use regression modeling of intra-urban residential variability in multiple traffic-related air pollutants. *Environ. Health* **2008**, *7*, 17. [[CrossRef](#)]
65. Cheng, V.; Ng, E.; Chan, C.; Givoni, B. Outdoor thermal comfort study in a sub-tropical climate: A longitudinal study based in hong kong. *Int. J. Biometeorol.* **2012**, *56*, 43–56. [[CrossRef](#)]
66. Vardoulakis, S.; Fisher, B.E.A.; Pericleous, K.; Gonzalez-Flesca, N. Modelling air quality in street canyons: A review. *Atmos. Environ.* **2003**, *37*, 155–182. [[CrossRef](#)]
67. Belcher, S.; Jerram, N.; Hunt, J. Adjustment of a turbulent boundary layer to a canopy of roughness elements. *J. Fluid Mech.* **2003**, *488*, 369–398. [[CrossRef](#)]
68. Shi, Y.; Ho, H.C.; Xu, Y.; Ng, E. Improving satellite aerosol optical depth-pm2.5 correlations using land use regression with microscale geographic predictors in a high-density urban context. *Atmos. Environ.* **2018**, *190*, 23–34. [[CrossRef](#)]
69. Song, W.; Jia, H.; Li, Z.; Tang, D.; Wang, C. Detecting urban land-use configuration effects on no2 and no variations using geographically weighted land use regression. *Atmos. Environ.* **2019**, *197*, 166–176. [[CrossRef](#)]
70. Widya, L.K.; Hsu, C.-Y.; Lee, H.-Y.; Jaelani, L.M.; Lung, S.-C.C.; Su, H.-J.; Wu, C.-D. Comparison of spatial modelling approaches on pm10 and no2 concentration variations: A case study in surabaya city, indonesia. *Int. J. Environ. Res. Public Health* **2020**, *17*, 8883. [[CrossRef](#)]
71. Su, J.G.; Jerrett, M.; Beckerman, B.; Wilhelm, M.; Ghosh, J.K.; Ritz, B. Predicting traffic-related air pollution in los angeles using a distance decay regression selection strategy. *Environ. Res.* **2009**, *109*, 657–670. [[CrossRef](#)]
72. Basu, B.; Alam, M.S.; Ghosh, B.; Gill, L.; McNabola, A. Augmenting limited background monitoring data for improved performance in land use regression modelling: Using support vector regression and mobile monitoring. *Atmos. Environ.* **2019**, *201*, 310–322. [[CrossRef](#)]
73. Liu, J.; Cai, P.; Dong, J.; Wang, J.; Li, R.; Song, X. Assessment of the dynamic exposure to pm2.5 based on hourly cell phone location and land use regression model in Beijing. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5884. [[CrossRef](#)]
74. Bartley, M.; Hanks, E.; Schliep, E.; Soranno, P.; Wagner, T. Identifying and characterizing extrapolation in multivariate response data. *PLoS ONE* **2019**, *14*, e0225715. [[CrossRef](#)] [[PubMed](#)]
75. Hurvich, C.; Simonoff, J.; Tsai, C. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **1998**, *60*, 271–293. [[CrossRef](#)]
76. Bowman, A. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* **1984**, *71*, 353–360. [[CrossRef](#)]
77. Xie, R.; Wei, D.; Han, F.; Lu, Y.; Fang, J.; Liu, Y.; Wang, J. The effect of traffic density on smog pollution: Evidence from Chinese cities. *Technol. Forecast. Soc. Change* **2019**, *144*, 421–427. [[CrossRef](#)]
78. Chen, J.; Wang, B.; Huang, S.; Song, M. The influence of increased population density in China on air pollution. *Sci. Total Environ.* **2020**, *735*, 139456. [[CrossRef](#)] [[PubMed](#)]
79. Lou, C.-R.; Liu, H.-Y.; Li, Y.-F.; Li, Y.-L. Socioeconomic drivers of pm2.5 in the accumulation phase of air pollution episodes in the Yangtze River delta of China. *Int. J. Environ. Res. Public Health* **2016**, *13*, 928. [[CrossRef](#)] [[PubMed](#)]
80. Ding, Y.; Zhang, M.; Qian, X.; Li, C.; Chen, S.; Wang, W. Using the geographical detector technique to explore the impact of socioeconomic factors on pm2.5 concentrations in China. *J. Clean. Prod.* **2019**, *211*, 1480–1490. [[CrossRef](#)]
81. Fung, W.Y.; Wu, R. Relationship between intraseasonal variations of air pollution and meteorological variables in Hong Kong. *Ann. GIS* **2014**, *20*, 217–226. [[CrossRef](#)]
82. Yang, J.; Shi, B.; Shi, Y.; Marvin, S.; Zheng, Y.; Xia, G. Air pollution dispersal in high density urban areas: Research on the triadic relation of wind, air pollution, and urban form. *Sustain. Cities Soc.* **2020**, *54*, 101941. [[CrossRef](#)]
83. Chudnovsky, A.A.; Koutrakis, P.; Kloog, I.; Melly, S.; Nordio, F.; Lyapustin, A.; Wang, Y.; Schwartz, J. Fine particulate matter predictions using high resolution aerosol optical depth (aod) retrievals. *Atmos. Environ.* **2014**, *89*, 189–198. [[CrossRef](#)]
84. Jiao, Y.; Su, M.; Ji, C.; Yang, S.; Zhang, P. How to design fully cooperative policies to abate transboundary air pollution between two highly asymmetric regions: An abnormal incrementalism analysis. *J. Clean. Prod.* **2021**, *278*, 124042. [[CrossRef](#)]
85. Planning Department of Hong Kong. *Hong Kong 2030+ Planning and Urban Design for a Liveable High-Density City*; Planning Department of Hong Kong: Hong Kong, China, 2016.