Landscape and Urban Planning 189 (2019) 15–26

Contents lists available at ScienceDirect

Landscape and Urban Planning

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

Research Paper

Investigating the influence of urban land use and landscape pattern on $PM_{2.5}$ spatial variation using mobile monitoring and WUDAPT

Yuan Shi[a,*], Chao Ren[c,e], Kevin Ka-Lun Lau[b,c,d], Edward Ng[a,b,c]

[a] School of Architecture, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong Special Administrative Region
[b] The Institute of Environment, Energy and Sustainability (IEES), The Chinese University of Hong Kong, Shatin, N.T., Hong Kong Special Administrative Region
[c] Institute of Future Cities (IOFC), The Chinese University of Hong Kong, Shatin, N.T., Hong Kong Special Administrative Region
[d] CUHK Jockey Club Institute of Ageing, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong Special Administrative Region
[e] Faculty of Architecture, The University of Hong Kong, Hong Kong Special Administrative Region

A R T I C L E I N F O

Keywords:
Air pollution
Land use
Landscape pattern
Mobile monitoring
WUDAPT

A B S T R A C T

Particulate matter that < 2.5 μm in aerodynamic diameter ($PM_{2.5}$) has been recognized as one of the principal pollutants that degrades air quality and increases health burdens. In this study, we employ the MLR and GWR modelling method to obtain estimation models for $PM_{2.5}$ with a set of land use/landscape metrics as predictor variables. The study focused on investigating the influence of urban land use and landscape pattern on $PM_{2.5}$ spatial variation, specifically, on identification of influential landscape classes/types that regulate $PM_{2.5}$ concentration levels. The spatial $PM_{2.5}$ concentration in the compact urban scenario of Hong Kong was sampled by conducting a series of mobile monitoring campaigns. The Local Climate Zone (LCZ) Scheme and World Urban Database and Portal Tools (WUDAPT) level 0 database were adopted as the basis of the calculation of land use/landscape metrics. These metrics were then adopted as the predictors to explain the spatial variations in $PM_{2.5}$. 62% of the variance in $PM_{2.5}$ can be explained by the resultant GWR model using only five land use/landscape classes, and without using any traffic-related variables or data from emission inventory. The findings can inform the urban planning strategies for mitigating air pollution and also indicate the usefulness of LCZ and WUDAPT in estimating the spatial variation of urban air quality.

1. Introduction

More than half of people globally live in urban area and even will increase to over two-thirds by 2050 (UN, 2014). Nowadays, unprecedented rate of urbanization results in air pollution in urban areas and subsequent health impacts on urban population. Over 90 percent of the global population are exposing to air pollution that beyond the recommended level confirmed by WHO recently (UN, 2016). The amount of death caused by air pollution reached to 650 million in 2012 which accounts for 11.6% of the annual death toll in the world. Hence, the life risks caused by exposure to air pollution requires global attention (UNEP, 2012). With the rapid urban development in recent years, environmental issues associated with air pollution have become an enormous challenge to most of the large cities in Asia (Schwela, Haq, Huizenga, Han, & Fabian, 2012). As one of the most compact cities in Asia, Hong Kong is experiencing the challenges from severe air pollution (Kim Oanh et al., 2006; Schwela et al., 2012). Air quality monitoring data from local authority indicates that Hong Kong still fails to meet the WHO air quality standards (Brajer, Mead, & Xiao, 2006) despite efforts in the last decade (HKEPD, 2005). Notably, the annual average $PM_{2.5}$ concentration is double of the WHO standard. $PM_{2.5}$ – particulate matter (PM) that < 2.5 μm in aerodynamic diameter, has been recognized as one of the principal pollutants that degrades air quality and is associated with cardiovascular and respiratory mortality and hospitalizations (Lin et al., 2017; Wong, Tam, Yu, & Wong, 2002). According to the World Health Statistics (WHO, 2016), approximate 90% of the population living in cities was exposed to PM concentrations exceeding the WHO air quality guidelines (AQGs) (WHO & UNAIDS, 2006). It heavily influences the liveability of urban areas and the living quality of urban population.

Urban development significantly changes the natural land cover and landscape patterns (Landsberg, 1981) and such a highly artificial landscape and land cover in urbanized areas considerably altered local climate (Pielke & Avissar, 1990), air quality (Bogucki & Turner, 1987) and biodiversity (Alkemade et al., 2009). As such, it is important to optimize land use allocation/landscape planning for an environmental

* Corresponding author at: Rm 905, YIA Building, The Chinese University of Hong Kong, Shatin, NT, Hong Kong Special Administrative Region.
E-mail address: shiyuan@cuhk.edu.hk (Y. Shi).

https://doi.org/10.1016/j.landurbplan.2019.04.004
Received 14 November 2018; Received in revised form 9 April 2019; Accepted 11 April 2019
Available online 21 April 2019

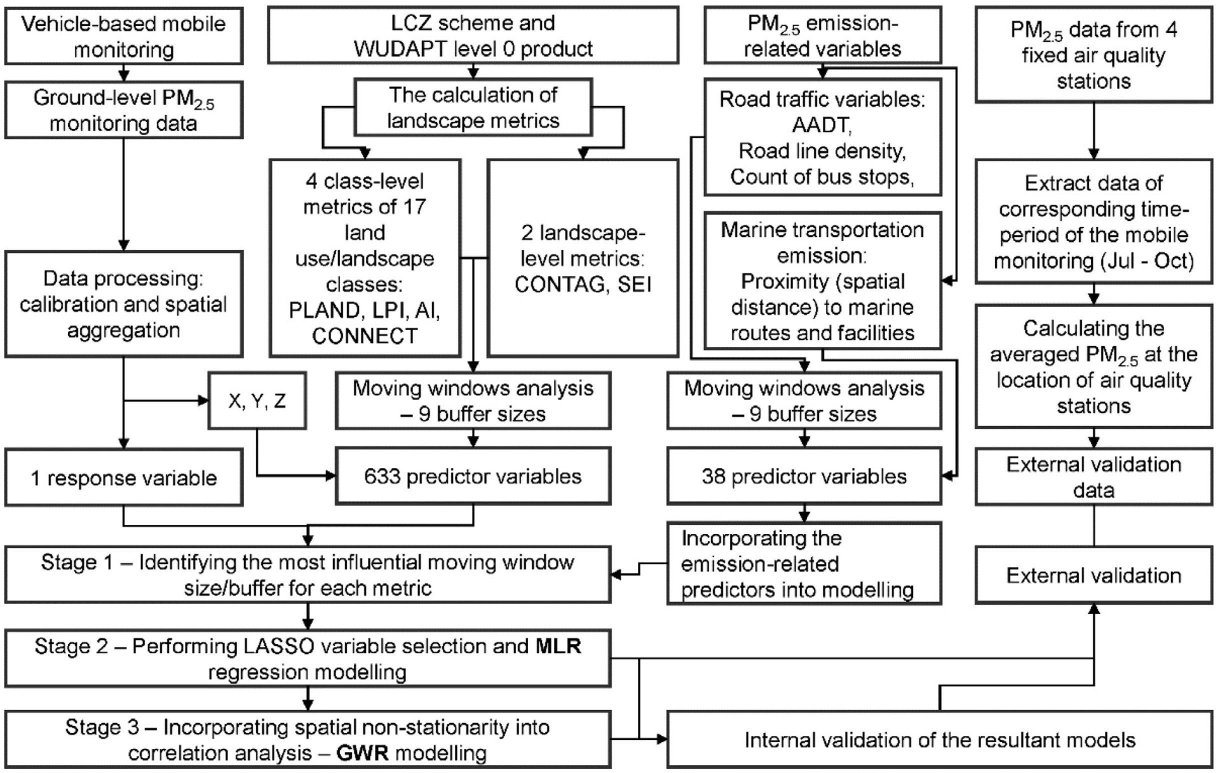

Fig. 1. The workflow chart of the present study.

and sustainable urban development has been emphasized (de Groot, Alkemade, Braat, Hein, & Willemen, 2010). It has been observed that the spatial variation of intraurban air pollution closely relates to land use planning (Foley et al., 2005; Xian, 2007). Different land use types in the city have varied effects on the urban air quality. Industrial areas and heavy traffic usually contribute to a considerably high concentration level of both particulate matters ($PM_{2.5}$, PM_{10}) and gaseous pollutants (CO, NO_X) due to the large emission intensity (de Hoogh et al., 2013; Habermann, Billger, & Haeger-Eugensson, 2015; Ross et al., 2006). In compact urban areas, zones with high level of air pollution spatially correlate with commercial and residential land use because the compact urban form blocks air ventilation and, consequently, impede the dispersion of air pollutants (Shi, Lau, & Ng, 2017). Open space is also influential to pollutant dispersion. Proximity to open urban public space (e.g. public squares, city parks, playgrounds) contributes to a better air movement (Ng, 2009) and hence benefits pollutant dispersion. Differently, proximity to waterfront area has both benefits and inconveniences to local air quality. The specific condition might depend on the climatic characteristics and geographical contexts. Proximity to waterfronts often provides better ventilation for pollutant dispersion. However, a strong radiation condition plus the presence of primary air pollutants react and form troposphere ozone in waterfronts areas (Simpson, 1994).

A modification in landscape patterns also affects the spatial variation of air pollution by interfering with critical atmospheric processes that are decisive to the transport, deposition, and dispersion of the air pollutants (Pielke et al., 2002; Weaver & Avissar, 2001). For example, there have been many studies emphasizing the importance of urban greening and forests to the improvement of urban air quality (Escobedo, Kroeger, & Wagner, 2011; Nowak, Crane, & Stevens, 2006). Vegetation has the capacity to separating aerosols and chemicals from the atmosphere. Generally speaking, the concentrations of particulate air pollutants can be significantly reduced due to the influence of vegetation on the deposition velocity, particularly, when the vegetation is close to the emission sources (Janhäll, 2015). However, the specific situation depends on the types of pollutants (e.g. $PM_{2.5}$ or VOC), the types of vegetations (e.g. tall tree or low bush), and the geometrical characteristics of street canyons (Vos, Maiheu, Vankerkom, & Janssen, 2013).

Despite that the land use is one of the most important determinants of urban air quality, most of the current studies only adopted the areal composition (the total area of each type of land use in a certain spatial extent) as the indicator to quantify the land use (Hoek et al., 2008). The spatial pattern (e.g. the allocation, layout, evenness, fragmentation, etc.) of different land use types have been rarely considered in the investigation of the spatial variation of intraurban air quality. This is an obvious research gap because the spatial variation of intraurban air quality associates with the land use planning via many different pathways (Frank et al., 2006). Facilitated by the rapid development in geographic information system (GIS) technologies, hundreds of indicators/metrics have been developed to quantify land use allocation and landscape patterns (Gustafson, 1998), which has been considered as the prerequisite to the studies in urban ecological research (McGarigal, 2006). However, there are only a limited amount of studies that focuses on the relationship between urban land use/landscape patterns and urban air pollution (Wu, Xie, Li, & Li, 2015). Therefore, the present study aims to achieve a comprehensive understanding on the influence of land use and landscape planning on the spatial pattern of $PM_{2.5}$ in Hong Kong.

Mobile monitoring, as an efficient method of the spatial investigation, has been increasingly used in the intraurban air quality research and pollution exposure studies (Adams & Kanaroglou, 2016; Hagler, Thoma, & Baldauf, 2010; Isakov, Touma, & Khlystov, 2007; Westerdahl, Fruin, Sax, Fine, & Sioutas, 2005; Xu et al., 2017) due to its advantages of spatial coverage over the limited amount of the sparsely distributed air quality monitoring stations. In this study, by conducting a series of vehicular-based mobile monitoring campaigns, the ground-level $PM_{2.5}$ concentrations were sampled in different parts of Hong Kong with varied land use/landscape. After the mobile monitoring campaigns, the mobile monitored spatial $PM_{2.5}$ data were collated in GIS. Meanwhile, the land use and landscape pattern of Hong Kong was quantified by calculating a set of well-established landscape pattern metrics based on the globally standardized Local Climate Zone (LCZ) scheme (Stewart & Oke, 2012). Multivariate statistical correlation analysis was then performed to correlate the spatial $PM_{2.5}$ data with the landscape pattern metrics. As the results, the correlation models were developed for the spatial estimation of intraurban air pollution. The resultant models were validated by the monitoring data from fixed air quality stations operated by the local authority. On top of the models, the influence of urban land use and landscape planning on the spatial patterns of $PM_{2.5}$ was investigated by identifying the critical metrics of land use/landscape patterns. Fig. 1 demonstrates the workflow of the present study.

2. Methods

2.1. Quantifying the land use and landscape spatial pattern

2.1.1. The application of LCZ scheme and WUDAPT level 0 product

Most of the current studies focused on the influence of land use on urban air quality only use the area and distance as the indicator/predictor to measure the land use and proximity to open space (Hoek et al., 2008). The detailed spatial pattern (e.g. the configuration, allocation, evenness, fragmentation, clustering, edge effects, etc.) of different land use/landscape types have rarely been quantified in the investigation of the spatial variation of intraurban air pollution. Previous land use/ landscape studies usually use the land use data provided by the local governmental authorities of the study area, which makes the cross-cities comparison becomes difficult due to the varied land use classification schemes. Currently, the USGS land-use and land-cover category is an internationally idiomatic standard classification of land use (Anderson, 1976). For example, the most popularly used model in the field of atmospheric pollution modelling – Weather Research and Forecasting model coupled with Chemistry (WRF-Chem) adopts the USGS 24-category land-use data as the default built-in land use data (Grell et al., 2005). In Hong Kong, the lack of land resources and high degree of population aggregation jointly form a compact and vertical mode of urban development. In the high-density urban built-up areas, there are a considerable number of high-rise buildings/skyscrapers with varied functions at different floors (Lau, Giridharan, & Ganesan, 2005). In contrast to the high-density urban core, there are > 70% of the total land area (approximately 1100 km^2) are vegetated mountainous areas and urban forests (Taylor, 1986). The mixing of highly diverse land use shapes an extremely heterogeneous landscape of Hong Kong so such a unique urban context cannot be well depicted by the USGS 24-category. Based on the widely used land surface classification scheme – local climate zone (LCZ) (Bechtel et al., 2015; Stewart & Oke, 2012), both built-up areas and natural land cover can be classified into 17 distinct types for the depiction of the land use diversity and variability of the context of Hong Kong, especially for the densely built-up areas (Table 1).

WUDAPT is a global initiative project volunteered by local urban experts. It aims to establish a global urban database based on the LCZ scheme (Mills, Ching, See, Bechtel, & Foley, 2015). The level 0 data provides a 17-type land classification map at fine spatial scale and has sufficient quality for environmental research application (Bechtel et al., 2019). Now there are over 150 cities' standardized LCZ data available on the WUDPAT data platform and this initiative has attracted a growing multi-disciplinary research community's interest. In Hong Kong, a WUDAPT level 0 database has been developed at a fine spatial resolution of 100 m in a series of previous studies based on satellite images (Ren et al., 2016; Wang, Ren, Xu, Lau, & Shi, 2017). The results of the accuracy assessment indicate that the resultant WUDAPT classification of Hong Kong is suitable for depicting of the diversity and variability of the landscape of Hong Kong. Therefore, it was adopted by the present study as the basis of land use and landscape analysis.

Table 1
The land use categories of Hong Kong – a comparison between WUDAPT and USGS 24-category land use classification.

WUDAPT Classification based on LCZ		USGS 24-category Land Use Classification	
LCZ Category	Land Use/Landscape Description	Land Use Category	Land Use/Landscape Description
LCZ 1	Compact High-rise	20	Urban (High-rise)
LCZ 4	Open High-rise		
LCZ 2	Compact Mid-rise	1	Urban (Mid-rise)
LCZ 5	Open Mid-rise		
LCZ 3	Compact Low-rise	23	Urban (Low-rise)
LCZ 6	Open Low-rise		
LCZ 7	Lightweight Low-rise		
LCZ 8	Large Low-rise		
LCZ 9	Sparsely Built		
LCZ 10	Heavy Industry		
LCZ A	Dense Trees	15	Mixed Forest
LCZ B	Scattered Trees		
LCZ C	Bush, Scrub	8	Shrubland
LCZ D	Low Plant	5	Cropland/Grassland Mosaic
LCZ E	Bare Rock or Paved	19	Barren or Sparsely Vegetated
LCZ F	Bare Soil or Sand		
LCZ G	Water	16	Water Bodies

2.1.2. The calculation of landscape metrics

As highly quantifiable measures, the calculation of landscape metrics have been incorporated into the satellite image-based land use/ land cover analysis (Southworth, Nagendra, & Tucker, 2002). Most of the landscape metrics are developed based on the classic "patch-corridor-matrix" theory in landscape ecology (Forman, 1995). In the present study, six landscape metrics were selected based on literature (Neel, McGarigal, & Cushman, 2004; Roy & Mark, 1996) to quantify the detailed spatial pattern of different land use/landscape types by utilizing knowledge of landscape ecology. They are separated into two different groups because they belong to different landscape levels: the class-level and the landscape-level. Briefly speaking, four class-level metrics represent the quantity and the spatial pattern of one particular type of land use/cover within the unit area. Two landscape-level metrics evaluate the combination, arrangement, and mixing of all different types of land use/cover within the unit area. Four class-level landscape metrics were selected to represents the spatial pattern of each type of land use/landscape classes – percentage of landscape types (PLAND), Largest Patch Index (LPI), Aggregation Index (AI), Connectance Index (CONNECT). Two widely used landscape-level metric – contagion index (CONTAG) and Shannon's Evenness Index (SEI) was adopted to quantify the diversity of the land use. In this study, Fragstats (version 4) – a widely used program for spatial pattern analysis of categorical maps, was used to calculate all landscape metrics (McGarigal, Cushman, & Ene, 2012).

PLAND is the most basic class-level metric of landscape composition. It calculates the areal proportion of a certain type of landscape in the focused area (refers to each moving window/round buffer in this study, see Section 2.3) as a percentage value, which can be calculated as follows:

$$PLAND = P_i = \frac{\sum_{j=1}^{n} a_{ij}}{A} * 100 \tag{1}$$

where, P_i is the PLAND of landscape type (LCZ class, in this study) i. n is the total number of patches of the specified landscape type in the study area. a_{ij} is the area of landscape patch j of the landscape type i. LPI is also a measure of the areal proportion of specific landscapes, which is similar with PLAND. The only difference is that LPI is only calculating the percentage of the largest single patch instead of accounting all patches of the specified landscape type (Eq. (2)). Therefore, it is a measure of the dominance of each landscape type in the study area.

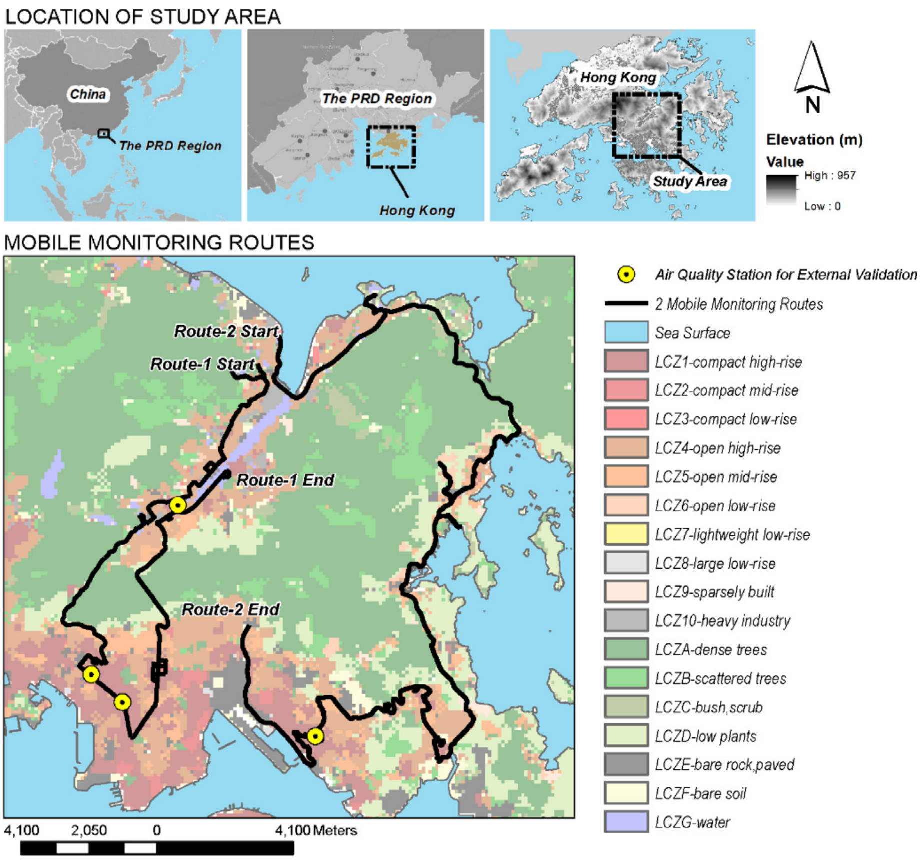

Fig. 2. The location of Hong Kong, and the two mobile monitoring routes used in the present study in sampling the ground-level $PM_{2.5}$ in different types of the land use/landscapes of Hong Kong (based on the LCZ classification scheme). $PM_{2.5}$ data of corresponding time period of mobile monitoring campaign from the four labelled air quality stations will be used as the external validation dataset. Modified from: Shi et al. (2018).

$$LPI = \frac{\max_j a_{ij}}{A} * 100 \tag{2}$$

AI has been developed to measure the spatial aggregation levels of a specific landscape type in the study area (He, DeZonia, & Mladenoff, 2000). Some earlier developed landscape metrics are scale-dependent which means that the calculation results will be to a certain extent sensitive to the map resolution (Turner & Gardner, 2015). AI overcomes the above limitation of those previous metrics, therefore, was selected by this study to evaluate the aggregation level of patches of each landscape type. AI is a percentage value of the frequency of the spatial adjacencies between the patches of a specified landscape type. AI = 0, when all patches of the specified landscape type are entirely dispersed. The details of calculation have been demonstrated by He et al. (2000) in their study. CONNECT quantitatively evaluates the functional connectivity between patches of each built-up or landscape type (in this study, the LCZ sites). It is an important concept in landscape ecology (Tischendorf & Fahrig, 2000). All patches of the same type in the study area is firstly paired. Based on a threshold of distance, each pair of patches is defined to be either connected or unconnected in terms of their landscape function. As an indicator of the functional connectivity, CONNECT calculates the percentage of connected pairs (Eq. (3)):

$$CONNECT = \frac{\sum_{j \neq k}^{n} c_{ijk}}{\frac{n_i(n_i - 1)}{2}} * 100 \tag{3}$$

where n_i is the total number of patches of the specified landscape type in this study area (there are a total of $n_i(n_i - 1)/2$ pairs). j and k are the two patches of a pair. $c_{ijk} = 1$ if the two patches are connected, otherwise, $c_{ijk} = 0$. At the landscape-level, the evaluation of landscape contagion in this study is based on an improved metric – CONTAG which is developed by Li and Reynolds (1993) (the detailed algorithm has been demonstrated by their study). CONTAG evaluates the landscape aggregation in a certain study area by taking all landscape types into consideration. It is a percentage value ranges from 0 (all landscape types in the study area are maximally disaggregated) to 100 (all landscape types in the study area are maximally aggregated). SEI is another widely-used landscape-level metric of measuring the diversity of land use/landscape composition in a certain area, which ranges from 0 (no diversity, one single type of landscape dominates the entire study area) to 1 (high diversity without dominance effect, the proportion of all types of landscape are perfectly the same in the study area). SEI was calculated based on the following equation:

$$SEI = \frac{-\sum_{i=1}^{m} (P_i * \ln P_i)}{\ln m} \tag{4}$$

where, P_i is the PLAND of landscape type (LCZ class) i. m is the total number of landscape types in the study area. All landscape metrics were normalized to [0,1] and used as the predictors of $PM_{2.5}$ concentration in later analysis (Section 2.3.1).

2.2. $PM_{2.5}$ mobile monitoring campaigns

2.2.1. Monitoring plan

Mobile monitoring method has been increasingly adopted to investigate the spatial variation of urban air quality (Adams & Kanaroglou, 2016; Xu et al., 2017). In this present study, the spatial variation of ground-level $PM_{2.5}$ in Hong Kong were investigated by a series of vehicular-based mobile monitoring. The mobile monitoring method has been successfully adopted in a preliminary study of Hong Kong to investigate the street-level particulate air pollution in the downtown area of Hong Kong within a relatively small spatial extent (Shi, Lau, & Ng, 2016). However, the design of monitoring route is largely determined by the study objective. It should be noticed that Hong Kong is a mountainous city with a highly heterogeneous landscape pattern and a compact urban scenario in its built-up area. To serve the objective of the present study, the mobile monitoring route has to be entirely redesigned in order to cover a broad range of various types of land use and landscapes. As the results, two monitoring routes with a total length of about 90 km were designed by the present study (Fig. 2). The first route has a length of 35 km and mainly passes through more built-up areas with artificial land covers and landscapes. The second route has a length of 55 km mainly covers the natural landscapes/land cover.

Mobile monitoring campaigns with repeated monitoring runs on the same route at properly-selected time slots are required for reliable observations (Elen et al., 2013). On top of that, the spatiotemporal data can be spatially aggregated for each monitored location on the route to obtain a robust estimation on spatial pattern of the air quality (Hatzopoulou et al., 2017). Three particular time slots of each day were selected by this study for monitoring the $PM_{2.5}$ spatial variation in a diurnal cycle, which are 09:00 am to 11:00 am, 2:00 pm to 4:00 pm, and 7:00 pm to 9:00 pm. Considering the regional transportation of the $PM_{2.5}$ from the Pearl River Delta region (PRD) of Mainland China affects Hong Kong only one-third of time in the year mainly during the winter time (Lau, Lo, Gray, Yuan, & Loh, 2007; Yuan et al., 2006), all monitoring campaigns were conducted between July and October to avoid the dominance effect of regional air pollution. The mobile monitoring campaigns were shared by the present study and another previous study on the spatial investigation of air temperature (Shi, Lau, Ren, & Ng, 2018). Therefore, details information has been provided by the above previous study.

2.2.2. Instrumentation and data calibration

A compact multi-purpose vehicle with a $PM_{2.5}$ monitor and micro-climate probes equipped was used for the mobile measurement campaigns in the present study. The concentration level of $PM_{2.5}$ was monitored by a DustTrak DRX aerosol monitor (hereinafter the DustTrak) with a temporal frequency of 1 Hz. The air temperature (T_a, °C) and relative humidity (*RH*, %) were synchronously monitored by a set of TESTO™ 480 Thermometers. A GPS locator and a video camera were also installed to record the geographical position and the surrounding conditions. Before being installed on the measurement vehicle, the DustTrak monitor was collocated with a roadside air quality monitoring station (Mong Kok Station) of the Hong Kong Environmental Protection Department (HKEPD) (HKEPD, 2013) for calibration. The annual average *RH* of Hong Kong is approximate 80% which is a relatively high level. All main chemical components of the aerosol are measured by the DustTrak (a light scattering instrument), which account for about 70% or more of $PM_{2.5}$ mass. Consequently, the reading will increase with high relative humidity due to the increase in the average particle size associated with condensational growth of hygroscopic components of the aerosol (Swietlicki, 2004; Zhang, 1996). Therefore, the DustTrak readings were firstly corrected to remove the influence of the particle-bound water using the synchronously monitored *RH* (Eq. (5)) based on Ramachandran, Adgate, Pratt, and Sexton (2003):

$$PM_{2.5DRXRH} = PM_{2.5DRX} / \left[1 + 0.25\frac{RH^2}{(1 - RH)} \right] \tag{5}$$

where the $PM_{2.5DRX}$ is the uncorrected DustTrak readings. The $PM_{2.5DRXRH}$ is the corrected readings in which the remove the influence of the particle-bound water has been removed. The collocation comparison method with a linear relationship-based calibration is commonly used for the calibration of DustTrak DRX and has been used in previous studies in Hong Kong (Che, Frey, & Lau, 2016; Li, Che, Frey, & Lau, 2018; Li, Che, Frey, Lau, & Lin, 2017). Similarly, in the present study, the DustTrak used for mobile monitoring was collocated with the aforementioned monitoring station for a 12-hour collocation campaign. A linear regression was then performed to derive the relationship between hourly averaged readings from the DustTrak and the hourly monitoring data of the reference station. The resultant $r^2 = 0.898$ indicates a good relationship. Therefore, the slope of the linear regression (which is 1.69) was used as the calibration factor ($CF_{PM2.5DRX}$) or the photometric calibration of the DustTrak monitor (Eq. (6)).

$$PM_{2.5DRXRHCF} = \frac{PM_{2.5DRXRH}}{CF_{PM2.5DRX}} \tag{6}$$

where the $PM_{2.5DRXRH}$ is the hourly averaged readings from the DustTrak after humidity correction. $PM_{2.5DRXRHCF}$ is the resultant value after both the humidity correction and photometric calibration. The above humidity correction and photometric calibration were performed for all $PM_{2.5}$ data from the mobile monitoring campaigns before further data processing.

2.2.3. Data processing

The complex roadside environment of Hong Kong is usually being influenced by intense traffic flows and other roadside anthropogenic activities. Some abnormal data samples (show as the spikes and outliers in the dataset) that influenced by anomalous pollution sources have been observed in our measurement data. The sampled $PM_{2.5}$ values could be much higher than the typical ambient concentration level, when driving closely behind heavy-duty diesel vehicles or driving near building construction sites/roadside food restaurants. In this study, a 4-order polynomial Savitzky–Golay (S-G) filter was used to deal with the abnormal data spikes. S-G filter is a moving average filter developed for eliminating the data noise without significant distortion of the data (Orfanidis, 1995). A data span of 11 (the mean of the sampling point numbers in per HK LCZ map cell) was used as the data span for performing the data filter.

Temporal effects in background $PM_{2.5}$ concentration level need to be removed from the spatial dataset. Temporal adjustments were made for each mobile monitoring dataset to eliminate the impacts of hour-to-hour difference. Hourly $PM_{2.5}$ monitoring data from the nearest HKEPD general air quality monitoring station were used as the reference for the temporal adjustment of each mobile monitoring data point based on a linear assumption of temporal changes in background $PM_{2.5}$ concentration level. The reference air quality monitoring stations in the study area were shown in Fig. 2. The spatial estimation of air quality trend is also sensitive to the data processing strategies (Brantley et al., 2014). An appropriate spatial scale is also essential to the spatial investigation of air quality (Lightowlers, Nelson, Setton, & Keller, 2008). In this study, the spatial scale of data aggregation is determined to be in conformity with the spatial resolution of the Hong Kong LCZ map. The cell size of Hong Kong LCZ map as WUDAPT level 0 product is 100 m × 100 m. Therefore, a distance of 100 m was used as the spatial interval to create a groups of equally spaced aggregation points along the two mobile monitoring routes (a total of 826 aggregation points was generated). All measured $PM_{2.5}$ data were then aggregated to these aggregation points by mean and used as the response variables in the statistical modelling later.

2.3. Correlating the land use/landscape pattern with $PM_{2.5}$ observations

2.3.1. Predictor variables and response variable

The spatial gradient of all landscape metrics mentioned above was analyzed over entire Hong Kong by using a moving windows method. A round-shaped buffer was used as the shape of the moving window. It was created for each cell of the LCZ classification map (mentioned in Section 2.1) so that the six metrics can be calculated for each land use type at each location of Hong Kong. A series of buffer radius (R_{Buffer}) were adopted to investigate the landscape pattern at different spatial scales – 100, 200, 300, 400, 500, 750, 1000, 1500, 2000 m. The R_{Buffer} ranges from a small spatial scale of a small street block (100 m) to a large spatial size that similar to a common Tertiary Planning Unit (TPU) of Hong Kong (2000 m). The calculated values of all above metrics at all $PM_{2.5}$ data aggregation points (results from Section 2.2) were extracted and used as the predictor datasets. The longitude (X-coordinates), latitude (Y-coordinates), and altitude (Z) of each point (based on HK1980 Transverse Mercator project coordinate system) were also used as the candidate predictor variables of the correlation model. The corresponding aggregated $PM_{2.5}$ data were used as the response variables.

2.3.2. Developing the correlation model

As introduced in Section 2.1.2, the four class-level metrics among the six metrics are designed to represent the spatial pattern of each land use/landscape type, which means that these metrics need to be calculated for each land use/landscape type listed in Table 1 (17 times in total). The same metric calculated using two different R_{Buffer} are used as two separate predictor variables in the development of correlation model. For example, the PLAND of the type LCZ 1 calculated using 100 m and 200 m buffers will be regarded as two different metrics in this study, such that there are 70 metrics need to be calculated using nine different buffers. With the geo-coordinates (X, Y, and Z), as the results, a total of 633 predictor variables need to be examined during the modelling process which possibly leads to multicollinearity issues due to this large number of predictors (Franke, 2010). The multicollinearity in predictor variable data causes unreliable regression modelling results in environmental and ecological research, which should be minimized (Abdul-Wahab, Bakheit, & Al-Alawi, 2005; Graham, 2003). To serve as a reference for urban land use planning and landscape management, our regression modelling process aims to include those most significant predictors that would explain as much as of the influence of land use and landscape in the spatial variation of the response variables – $PM_{2.5}$ concentration. Therefore, the following stages of works were performed to screen all candidate variables and retain only a subset of significant variables. Only a limited number of variables will be finally included in the resultant model.

Stage 1 – Identifying the most influential moving window size/buffer for each metric. The impact range of different land use/landscape types may vary due to the differences in the emission, deposition rate as well as the complex physical or chemical basis of the particulate air pollutant diffusion and dispersion. Geographically, land use/landscape pattern quantified by a specific metrics within its most influential buffers explains the variation of $PM_{2.5}$ concentration to the greatest extent. Above is the reason behind performing the moving windows analysis based on a series of different R_{Buffer}. For example, the heavy industrial land use (LCZ 10 in Table 1) could affect the $PM_{2.5}$ concentration level within a geographical extent of several kilometers, while an isolated small piece of vegetated area (e.g. a small urban park in LCZ B) could only improve the ambient air quality within a couple of hundred meters. Therefore, the most influential size of moving windows – R_{Buffer} will not be identical for those variables included in the resultant model. The correlation coefficient (r) between the response variable – $PM_{2.5}$ concentration and each metric calculated within the nine R_{Buffer} were calculated based on simple linear regression. Only the R_{Buffer}-based metric which has the highest $|r|$ (considering that r could be either positive or negative, absolute value was used for the correlation comparison) were selected as the predictor variables and included in the next stage of the correlation analysis.

Stage 2 – Constructing multiple linear regression (MLR) model. The statistical correlation analysis starts from a classic multiple linear regression analysis (Eq. (7)):

$$PM_{2.5i} = \alpha_1 Var_1 + \alpha_2 Var_2 + \ldots + \alpha_n Var_n + \gamma + \varepsilon \quad (7)$$

where $PM_{2.5i}$ is the averaged $PM_{2.5}$ concentration value at the aggregation point i. The model includes n land use/landscape metrics as the predictor variables. $\alpha_1 \ldots, \alpha_n$ are the coefficient estimates of the metrics Var_1, Var_n at the aggregation point i. γ is the model intercept, and ε is the residual. For example, Var_1 could be $PLAND_{LCZ1,200m}$ which represent the areal proportion of LCZ 1 calculated within a round-shaped buffer with a radius of 200 m. As the basis of any further correlation analysis, MLR model was firstly constructed based on the variable subset from Stage 1. There will be still dozens of candidate variables were still involved as the potential predictors. Therefore, LASSO (Least Absolute Shrinkage and Selection Operator) is performed to identify a subset of influential predictors which possibly contains the best predictor variables. LASSO is a variable selection method which can be used to automatically screen a subgroup of significant predictor variables of the response variable from a large set of candidate predictors (Tibshirani, 1996), which is particularly useful to the relatively large predictor dataset of the present study where collinearity is potentially a problem. Restrictive VIF rules have been used to ensure that there is no collinearity among final included independent variables in resultant models. For example, the studies by Vienneau et al. (2013) and Shi et al. (2016), etc. The subset of predictor variables was further refined by adopting the following rules: Only variables with a p-value < 0.001 and VIF < 3 in the MLR model will be included. All other variables selected by LASSO will still be excluded.

Stage 3 – Incorporating spatial non-stationarity into correlation analysis. A small number of most influential predictor variables has been selected and used to construct an MLR model at stage 2. However, the MLR model are still constructed based on a fixed effect model structure, in which the effects of predictor variables are presumed to be spatially stationary. However, the influence of some predictors could be spatially variant due to the landscape heterogeneity of Hong Kong. The MLR model developed by performing a stepwise statistical procedure for selecting important independent variables must be further calibrated to deal with the spatial non-stationarity (Leung, Mei, & Zhang, 2000). Therefore, in this study, using the subset of most influential predictor variables that previously identified, geographically weighted regression (GWR) modelling is performed to incorporate the spatial non-stationarity into the correlation model. GWR is a widely-adopted method of dealing with such spatial non-stationarity in $PM_{2.5}$ spatial estimation (van Donkelaar, Martin, Spurr, & Burnett, 2015). GWR deals with the spatial non-stationarity by constructing local correlations for different spatial locations instead of using one global correlation for the entire spatial domain (Brunsdon, Fotheringham, & Charlton, 1998). The coefficient estimates of GWR model variables are spatially variant as well (Eq. (8)):

$$PM_{2.5i} = \sum_n \alpha_n(u_i, v_i) VAR_{n,\ d} + \gamma_i + \varepsilon_i \quad (8)$$

where $PM_{2.5i}$ is the averaged $PM_{2.5}$ concentration value at the aggregation point i. u_i, v_i are the geo-coordinates of the aggregation point i. α_n are the coefficient estimates of the n land use/landscape metrics ($VAR_{n,d}$) calculated within the R_{Buffer} of d. γ_i and ε_i are the intercept and residuals of GWR model.

2.3.3. Model validation

Both internal validation and external validation were conducted to examine the performance of the resultant models. For the internal validation, leave-one-out cross-validation (LOOCV) was adopted. Cross-validation adjusted r^2 (LOOCV r^2) and root-mean-square error (*RMSE*) were calculated. About the external validation, the resultant MLR and GWR model performance were further examined by the monitoring data from four fixed air quality stations operated by the local authority – HKEPD (Fig. 2). The 2016 annual averaged $PM_{2.5}$ data from four air quality stations outside the monitoring route were compared with the estimated $PM_{2.5}$ concentration value based on resultant models.

2.4. Incorporating the Emission-related predictors into models

In the previous section, land use and landscape metrics were used as predictors to estimate spatial $PM_{2.5}$. In this section, based on the same methodology, more predictors directly related to the $PM_{2.5}$ emissions will be examined to further improve the estimation accuracy. Same statistical methods (LASSO, MLR, GWR) and model criteria (p-value < 0.001 and VIF < 3) were adopted to ensure the robustness of resultant model. Road traffic is a major emission source of $PM_{2.5}$ in Hong Kong. Therefore, the annual average daily traffic ($AADT$) values which counted by the local authority to represents the traffic volume and road line density are used as the indicators of traffic-related $PM_{2.5}$ emission. The spatial data of $AADT$ and road line density were analyzed by using the same moving windows method described in Section 2.3.1. The road line density was calculated separately for major roads (RD_{Major}) and minor roads (RD_{Minor}). Additionally, the count of bus stops ($BUSST$) is also calculated using the buffers, since bus as a heavy-duty vehicle is a considerable $PM_{2.5}$ source. The emission from marine transportation is another major $PM_{2.5}$ source in Hong Kong (Lau et al., 2007). To take this into consideration, the proximity (spatial distance) to marine routes and facilities of each $PM_{2.5}$ aggregation points was calculated and used as a predictor variable. In the 1980s, the labor-intensive and high-pollution emission industries of Hong Kong have been relocated to Mainland China. Therefore, the present study does not include any industry pollution-related predictors. As the results, 38 more emission-related predictors were examined for improving the GWR model.

3. Results

The most influential buffers for each metric were identified by only keeping the R_{Buffer}-based metric corresponding to the buffer size which has the highest $|r|$. Additionally, those variables with a weak and/or statistically insignificant correlation with $PM_{2.5}$ ($|r| < 0.1$, p-value > 0.05) were also excluded and not used as the input for LASSO regression modelling. As the results, only 42 variables (include X, Y, and Z) remained to be used for regression modelling (Table 2).

Table 3 and Fig. 3 shows the resultant MLR model from stage 2 (mentioned in Section 2.3.2). Seven predictor variables are included by the MLR model and already explain almost 47% variation in the measured $PM_{2.5}$. The results indicate the significance of land use and landscape pattern in explaining the spatial variation of $PM_{2.5}$. After incorporating spatial non-stationarity into correlation analysis, the model performance was further improved. The adjusted r^2 of GWR model is 0.622 (Table 4 and Fig. 4). The external validation results show that the adjusted r^2 between the modelled $PM_{2.5}$ data and the 2016 annual averaged $PM_{2.5}$ data from the air quality stations are 0.699 and 0.871 for the MLR model and GWR model (without emission-related predictors included) respectively. Fig. 5 shows the $PM_{2.5}$ prediction maps derived from both the MLR and the GWR model.

Table 3
The performance and structure of the resultant MLR model of $PM_{2.5}$ concentration. All variables that meet the criteria of p-value < 0.001 and VIF < 3 in MLR model. Variable name: for example, "LPI_LCZ 1_1500" refers to the Largest Patch Index of land use type-LCZ1 calculated within the buffer of 1500 m.

The resultant MLR model of $PM_{2.5}$ concentration using land use/landscape metrics as predictors	
r^2	474
adjusted r^2	469
LOOCV r^2	464
RMSE	5.093
n	826
AICc	5043.611

Predictor Variables	Coefficient Estimates	95% CI Lower	95% CI Upper	Std Error	t Ratio	VIF
Model Intercept	33.445	30.915	35.975	1.289	25.950	.
$LPI_{LCZ1,1500m}$	0.014	25.333	32.696	1.875	15.470	2.170
$LPI_{LCZ4,1500m}$	794	2.690	10.899	2.091	3.250	1.396
$LPI_{LCZA,500m}$	0.330	−6.479	−2.180	1.095	−3.950	2.235
$LPI_{LCZB,1500m}$	4.052	−60.309	−27.795	8.282	−5.320	1.863
$AI_{LCZ2,400m}$	094	5.182	11.006	1.484	5.460	1.027
$AI_{LCZB,1500m}$	053	2.906	7.199	1.094	4.620	2.383
SEI_{400m}	5.610	−7.953	−3.268	1.193	−4.700	1.617

As described in Section 2.4, using the above MLR model as the basis, 38 more predictors that directly related to the $PM_{2.5}$ emissions were examined ($AADT$, road line density, $BUSST$ and the distance to the marine routes and facilities). The same method (mentioned in Section 2.3.2) was used to identify the most influential emission related predictors. As the results, two influential emission related predictors were identified – $AADT_{100m}$ and $RD_{Major,750m}$. After incorporating these two predictors into the MLR model, the model adjusted r^2 increased from 0.469 to 0.515. Moreover, the predictor $LPI_{LCZ4,1500m}$ becomes statistically insignificant due to the collinearity and therefore being excluded. However, adding these two traffic emission-related predictors doesn't substantially changes the GWR model performance (adjusted $r^2 = 0.599$, AICc = 4852.911). This indicates that the LCZ scheme and WUDAPT level 0 product could indirectly represent the road network organization.

Table 2
Summary of the most influential R_{Buffer} of selected land use/landscape predictor variables for MLR and GWR modelling (unit: m). CONTAG and SEI are landscape-level metrics which are not calculated for each land use/landscape type. Brackets indicate a negative correlation with $PM_{2.5}$ concentration; n.s. – Not significant statistically (p-value > 0.05); n.a. – Not available; Bold font indicates the final subset of variables that meet the criteria of p-value < 0.001 and VIF < 3 in MLR model.

Land use/Landscape Type	Land Use/Landscape Description	PLAND	LPI	AI	CONNECT	CONTAG	SEI
LCZ 1	Compact High-rise	1000	**1500**	750	n.s.	n.a.	n.a.
LCZ 2	Compact Mid-rise	400	400	**400**	n.s.	n.a.	n.a.
LCZ 3	Compact Low-rise	n.s.	n.s.	n.s.	n.s.	n.a.	n.a.
LCZ 4	Open High-rise	2000	**1500**	2000	n.s.	n.a.	n.a.
LCZ 5	Open Mid-rise	500	500	500	n.s.	n.a.	n.a.
LCZ 6	Open Low-rise	(300)	(300)	(300)	n.s.	n.a.	n.a.
LCZ 7	Lightweight Low-rise	n.s.	n.s.	n.s.	n.s.	n.a.	n.a.
LCZ 8	Large Low-rise	n.s.	n.s.	n.s.	n.s.	n.a.	n.a.
LCZ 9	Sparsely Built	(1500)	(1500)	(2000)	n.s.	n.a.	n.a.
LCZ 10	Heavy Industry	n.s.	n.s.	750	n.s.	n.a.	n.a.
LCZ A	Dense Trees	(500)	**(500)**	(500)	n.s.	n.a.	n.a.
LCZ B	Scattered Trees	(1500)	**(1500)**	**(1500)**	n.s.	n.a.	n.a.
LCZ C	Bush, Scrub	(2000)	(2000)	(2000)	n.s.	n.a.	n.a.
LCZ D	Low Plant	(750)	(750)	(750)	n.s.	n.a.	n.a.
LCZ E	Bare Rock or Paved	2000	2000	2000	n.s.	n.a.	n.a.
LCZ F	Bare Soil or Sand	(400)	(400)	(400)	n.s.	n.a.	n.a.
LCZ G	Water	n.s.	n.s.	n.s.	n.s.	n.a.	n.a.
All Types	n.a.	n.a.	n.a.	n.a.	n.s.	400	**(400)**

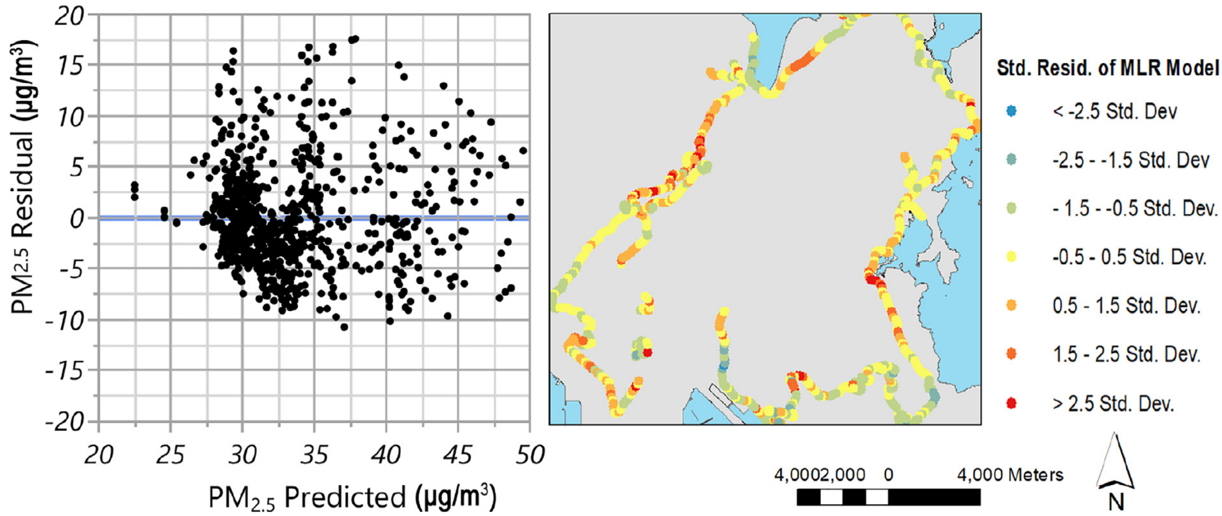

Fig. 3. The residual of the resultant MLR model of $PM_{2.5}$ concentration.

4. Discussion

4.1. Influential moving window sizes/buffers

As mentioned in Section 2.3.2, the most influential moving window size/buffer for each metric has been identified by calculating and keeping the R_{Buffer}-based metric corresponding to the buffer size which has the highest $|r|$. The sensitivity of the correlation between landscape metrics and the response variable to the changes of buffer size was illustrated in Fig. 6.

Findings from the process of identification of influential buffers indicate:

(1) All variables of CONNECT were excluded due to their correlation with $PM_{2.5}$ are weak (all $|r| < 0.1$) and statistically insignificant (p-value > 0.05). Therefore, CONNECT was excluded by the statistical modelling. CONNECT value for each land use/landscape types at most of the aggregation points are 0, which is the cause of the weak and insignificant correlation. This fact affirms the highly heterogeneous and fragmented city landscape pattern of Hong Kong.

(2) For each type of land use/landscape, the influential buffers of PLAND, LPI, and AI are similar, which confirms the differences in the impact range of land use types. Except those excluded variables, the types in urban built-up areas (LCZ 1 to LCZ 6) generally have a smaller influential buffer size than those types in suburban and rural areas (LCZ A to LCZ G). The correlation between LCZ7/LCZ8 and $PM_{2.5}$ are also weak ($|r| < 0.1$) and statistically insignificant (p-value > 0.05) because their areal proportion is quite small in Hong Kong. Similarly, LCZ 9 also has large influential buffer which is possibly because that they are sparsely distributed in different part of Hong Kong and usually has a small area, therefore, may not be included by those smaller buffers (> 1000 m).

(3) LCZ 1 and LCZ 4 have larger influential buffers than other built-up areas, which implies that the higher-rise built environment has a larger impact range. From the viewpoint of urban fluid dynamics, higher-rise buildings usually have larger influential range on the near-surface wind field as such hamper the pollution dispersion in a larger area.

(4) LCZ A represents the densely vegetated trees. The significance of variable $LPI_{LCZA,500m}$ indicates that it is important to have enough amount of high-quality urban greening within a buffer of 500 m. In other words, a large patch of dense trees, for example, a centralized park would be beneficial to the mitigation of air pollution of neighborhoods. The accessibility to urban greenery at the neighborhood scale should be given a high priority in urban planning and design practice. The inclusion of variable – $LPI_{LCZB,1500m}$ indicates that it could also be very useful to sparsely arrange trees at the urban district level. This finding is particularly useful for highly urbanized cities that have only limited land resources can be used

Table 4
The performance and statistical summary of coefficient estimates of the resultant $PM_{2.5}$ concentration GWR model without emission-related variables.

The resultant GWR model of $PM_{2.5}$ concentration by incorporating spatial non-stationarity	
r^2.	622
adjusted r^2.	622
LOOCV r^2.	620
RMSE	3.282
n	826
AICc	4815.135

Predictor Variables	Mean	Std. Dev.	Min	10% Quantiles	25% Quantiles	Median	75% Quantiles	90% Quantiles	Max
Model Intercept	32.036	2.461	27.532	28.527	29.894	32.364	33.906	35.283	36.495
$LPI_{LCZ1,1500m}$	0.007	23.376	10.550	28.464	31.123	33.277	36.068	85.479	119.712
$LPI_{LCZ4,1500m}$	052	15.029	−15.717	−11.428	−7.882	0.986	22.333	26.174	28.570
$LPI_{LCZA,500m}$	0.820	4.454	−6.796	−5.770	−3.481	−2.397	2.837	6.273	10.299
$LPI_{LCZB,1500m}$	9.625	53.490	−90.074	−75.347	−61.330	−38.883	3.012	78.601	119.892
$AI_{LCZ2,400m}$	635	1.316	5.584	5.739	6.260	7.891	8.636	9.412	10.269
$AI_{LCZB,1500m}$	564	3.703	−3.438	−2.083	0.243	4.200	6.438	8.132	9.934
SEI_{400m}	0.694	4.058	−11.542	−10.859	−7.496	−4.772	−1.751	1.302	3.114

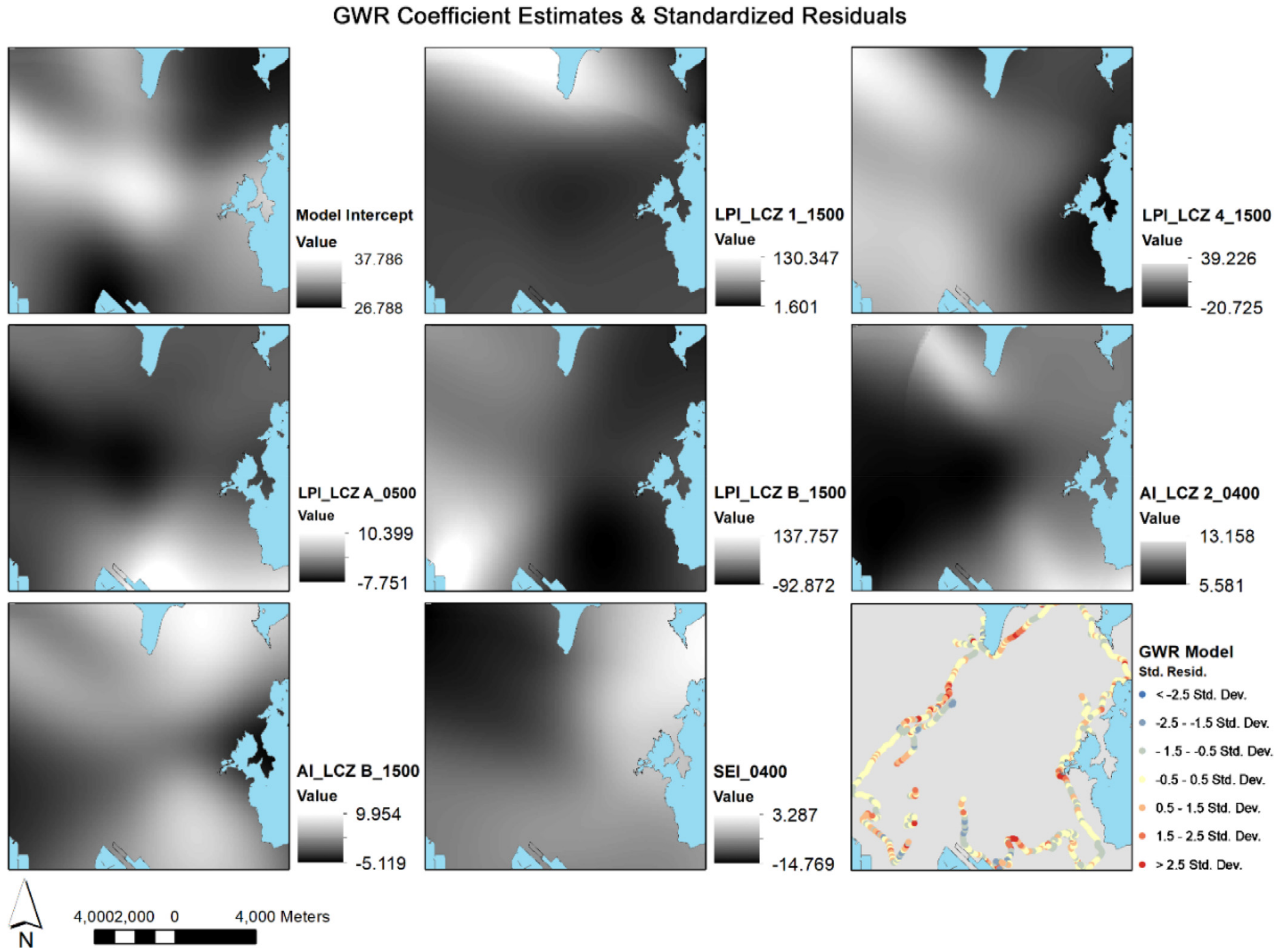

Fig. 4. The spatial non-stationarity in coefficient estimates of predictor variables and model intercept of $PM_{2.5}$ concentration GWR model without emission-related variables.

for greening in their intraurban areas.

(5) CONTAG quantifies the contagion of a certain type of land use (e.g. LCZ 1), while SEI evaluates landscape diversity. CONTAG are positively correlated with $PM_{2.5}$ concentration, while the correlation of SEI is negative. Notably, these two landscape-level metrics share the same influential buffer which is 400 m. Above findings indicate that higher diversity of the neighboring area help with the improvement of air quality. The contagion of those compact land use types in urban built-up areas (LCZ 1, LCZ 2, and LCZ 4), could largely decrease the dynamic potential of pollution dispersion.

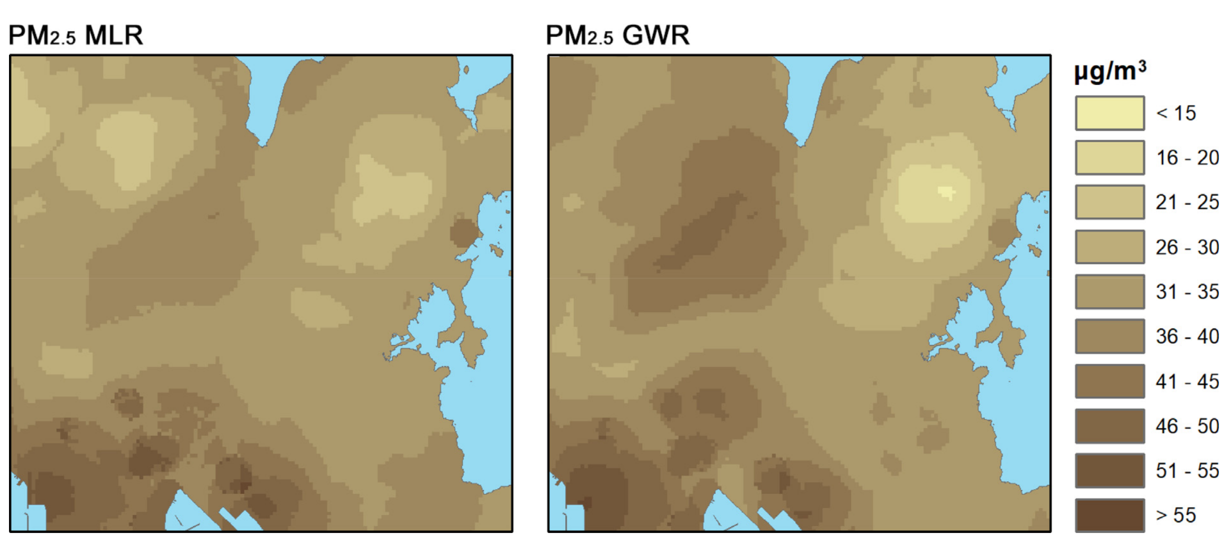

Fig. 5. The $PM_{2.5}$ prediction maps derived from the MLR, and the GWR models.

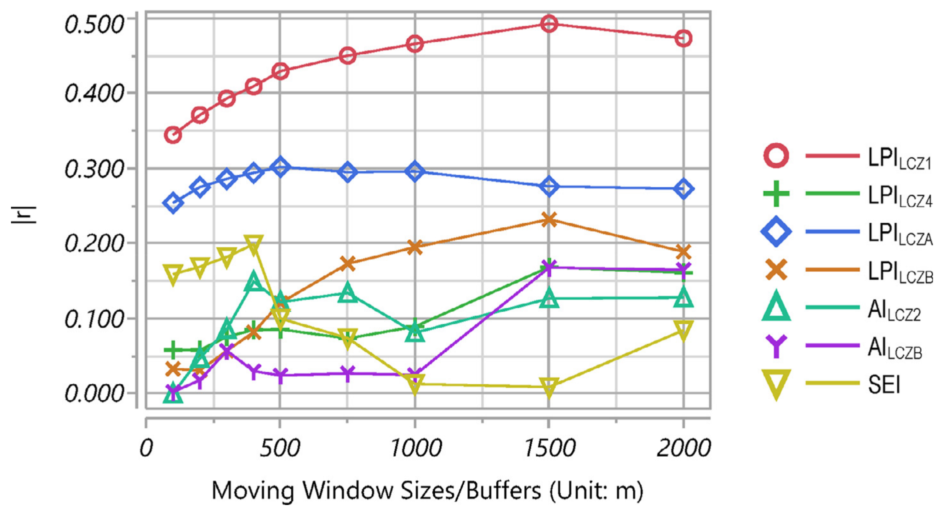

Fig. 6. The plot of the |r| between landscape metric included by the model and $PM_{2.5}$, calculated using various buffer sizes.

4.2. Resultant models and the revelation for urban planning practice

As described in Section 3, two resultant models were developed – MLR and GWR model. The development of the MLR model allows the recognition of the most influential metrics and the identification of their influencing spatial buffers. Seven predictor variables are included by the MLR model and already explain almost one half of the variation in the measured $PM_{2.5}$. In the resultant models, both $LPI_{LCZ1,1500m}$ and $LPI_{LCZ4,1500m}$ is positively related to the $PM_{2.5}$ concentration level indicates that a large area of high-rise building development could hamper the near-ground pollutant emissions, therefore, should be avoided in urban planning process. $AI_{LCZ2,400m}$ also has a positive relationship with $PM_{2.5}$ concentration level. Although LCZ 2 has a lower level of building height than LCZ 1 and LCZ 4, the higher ground coverage ratio still negatively affects the pollution dispersion. Similar findings could also be observed between LCZ 1 and LCZ 4. The coefficient estimates of LCZ 1 is much larger than LCZ 4, which indicates LCZ 1 has a more significant influence on the near-ground air quality due to its higher ground coverage ratio. The aggregation of high-rise and /or high ground coverage ratio building development is not recommended. This recommendation can be also supported by the negative correlation between SEI_{400m} and $PM_{2.5}$ concentration. The inclusion of $LPI_{LCZA,500m}$, $LPI_{LCZB,1500m}$, $AI_{LCZB,1500m}$ affirms that urban greening is an effective way of mitigating urban air pollution.

The development of the GWR model allows further incorporation of the spatial non-stationarity. As shown in the resultant GWR model, by incorporating spatial non-stationarity into the spatial analysis, only five land use/landscape classes can already explain > 60% of the spatial variation in $PM_{2.5}$, without using any traffic-related variables or data from emission inventory. Above indicates the considerable influence of urban land use/landscape pattern on the spatial air quality as well as the usefulness of WUDAPT in explaining the spatial variation of urban air quality.

4.3. Limitations

Although the aforementioned findings are informative and useful, there are still several limitations currently did not overcome by the present study. These limitations need to be considered very carefully and should be further investigated by follow-up studies. First, the relatively high cross-validation adjusted r^2 value might be because of the limited number of locations for validation and the stations are relatively close to the measurement routes. Therefore, additional measurements should be conducted to acquire more external validation data that further away from the current mobile monitoring routes). Moreover, the current study is a test case only based on the one case city. The transferability and applicability of the current LCZ-based research methodology for other cities and regions need to be further investigated. To be more specific, for example, the present study does not include any industry pollution-related predictors, which means that the effect of industry type of land use was not investigated. This is reasonable for Hong Kong because the high-pollution emission industries have been relocated outside Hong Kong. This limitation could introduce uncertainties because the effect of industry type of land use is influential to the air quality of other study areas, especially for those industrial-oriented cities. Another limitation is that although external validation has been conducted using another dataset, the dataset is still measured in the same city. Therefore, future work should focus on the external validation and the feasibility test of the current methods for other cities and areas. Considering the collinearity between LCZ related land use/landscape metrics and traffic-related predictors has been found, future work should also focus on investigating the representative of LCZ and WUDAPT level 0 product on the spatial emissions. Last but not least, the possible nonlinear relationship for landscape metrics has not been explored yet as there are a very large number of predictors. The interaction and polynomial terms in the correlation between landscape metrics and $PM_{2.5}$ should be explored by follow-up studies.

5. Conclusions

The present study is one of the first applications of LCZ scheme and WUDAPT level 0 product in the spatial estimation of intraurban air quality. The spatial $PM_{2.5}$ concentration in the compact urban scenario of Hong Kong was sampled by conducting a series of mobile monitoring campaigns. The WUDAPT level 0 database was adopted as the basis of the calculation of land use/landscape metrics which were used as the predictor variables to explain the spatial variations in $PM_{2.5}$ concentration. By utilizing the WUDAPT and combing the knowledge of urban landscape planning, this study investigates the influence of urban land use/landscape patterns on $PM_{2.5}$ concentration and develops spatial models that could explain the $PM_{2.5}$ spatial variation. By providing straightforward quantitative correlation between land use/landscape pattern and $PM_{2.5}$ concentration level, the study outputs could inform the urban planning strategies for mitigating air pollution. The resultant GWR model shows that only five land use/landscape classes can already explain 62% of the spatial variation in $PM_{2.5}$, without using any traffic-related variables or data from emission inventory, which shows the usefulness of LCZ scheme in estimating the spatial variation in urban air quality. This could also be particularly useful to the urban air quality assessment in those cities and areas where the long-term monitoring data, fine-grained traffic data, and detailed emission inventory are not available. More importantly, for the application of the globally standardized WUDAPT level 0 database, this study method can provide opportunities for standardizing $PM_{2.5}$ spatial mapping method and contributing to the global estimation of $PM_{2.5}$. This would greatly help researchers and scientists to quickly estimate the spatial pattern of urban air pollution by using free satellite images and other open resources, such as WUDAPT products.

Acknowledgements

This research is supported by the General Research Fund (GRF Project No.: 14610717 – "Developing urban planning optimization strategies for improving air quality in compact cities using geo-spatial modelling based on in-situ data") from the Research Grants Council (RGC) of Hong Kong Special Administrative Region. The authors deeply thank the reviewers for their insightful comments, feedbacks and constructive suggestions, recommendations on our research work. The authors also want to appreciate editors for their patient and meticulous work for our manuscript.

References

Abdul-Wahab, S. A., Bakheit, C. S., & Al-Alawi, S. M. (2005). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software, 20*(10), 1263–1271.

Adams, M. D., & Kanaroglou, P. S. (2016). Mapping real-time air pollution health risk for environmental management: Combining mobile and stationary air pollution monitoring with neural network models. *Journal of Environmental Management, 168*, 133–141.

Alkemade, R., van Oorschot, M., Miles, L., Nellemann, C., Bakkenes, M., & ten Brink, B. (2009). GLOBIO3: A framework to investigate options for reducing global terrestrial biodiversity loss. *Ecosystems, 12*(3), 374–390.

Anderson, J. R. (1976). *A Land Use and Land Cover Classification System for Use With Remote Sensor Data.* Washington: US Government Printing Office.

Bechtel, B., Alexander, P., Böhner, J., Ching, J., Conrad, O., Feddema, J., ... Stewart, I. (2015). Mapping local climate zones for a worldwide database of the form and function of cities. *ISPRS International Journal of Geo-Information, 4*(1), 199.

Bechtel, B., Alexander, P. J., Beck, C., Böhner, J., Brousse, O., Ching, J., ... Xu, Y. (2019). Generating WUDAPT level 0 data – Current status of production and evaluation. *Urban Climate, 27*, 24–45.

Bogucki, D. J., & Turner, M. G. (1987). *Landscape heterogeneity and disturbance.* Springer.

Brajer, V., Mead, R. W., & Xiao, F. (2006). Valuing the health impacts of air pollution in Hong Kong. *Journal of Asian Economics, 17*(1), 85–102.

Brantley, H. L., Hagler, G. S. W., Kimbrough, E. S., Williams, R. W., Mukerjee, S., & Neas, L. M. (2014). Mobile air monitoring data-processing strategies and effects on spatial air pollution trends. *Atmospheric Measurement Techniques, 7*(7), 2169–2183.

Brunsdon, C., Fotheringham, S., & Charlton, M. (1998). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician), 47*(3), 431–443.

Che, W. W., Frey, H. C., & Lau, A. K. H. (2016). Sequential measurement of intermodal variability in public transportation PM2.5 and CO exposure concentrations. *Environmental Science & Technology*.

de Groot, R. S., Alkemade, R., Braat, L., Hein, L., & Willemen, L. (2010). Challenges in integrating the concept of ecosystem services and values in landscape planning, management and decision making. *Ecological Complexity, 7*(3), 260–272.

de Hoogh, K., Wang, M., Adam, M., Badaloni, C., Beelen, R., Birk, M., ... Dedele, A. (2013). Development of land use regression models for particle composition in twenty study areas in Europe. *Environmental Science & Technology, 47*(11), 5778–5786.

Elen, B., Peters, J., Van Poppel, M., Bleux, N., Theunis, J., Reggente, M., & Standaert, A. (2013). The Aeroflex: A bicycle for mobile air quality measurements. *Sensors (Basel, Switzerland), 13*(1), 221–240.

Escobedo, F. J., Kroeger, T., & Wagner, J. E. (2011). Urban forests and pollution mitigation: Analyzing ecosystem services and disservices. *Environmental Pollution, 159*(8–9), 2078–2087.

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., ... Snyder, P. K. (2005). Global consequences of land use. *Science, 309*(5734), 570–574.

Forman, R. T. T. (1995). *Land mosaics: The ecology of landscapes and regions.* Cambridge University Press.

Frank, L. D., Sallis, J. F., Conway, T. L., Chapman, J. E., Saelens, B. E., & Bachman, W. (2006). Many pathways from land use to health: associations between neighborhood walkability and active transportation, body mass index, and air quality. *Journal of the American Planning Association, 72*(1), 75–87.

Franke, G. R. (2010). *Multicollinearity. Wiley international encyclopedia of marketing*. John Wiley Sons, Ltd.

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology, 84*(11), 2809–2815.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled "online" chemistry within the WRF model. *Atmospheric Environment, 39*(37), 6957–6975.

Gustafson, E. J. (1998). Quantifying landscape spatial pattern: What is the state of the art? *Ecosystems, 1*(2), 143–156.

Habermann, M., Billger, M., & Haeger-Eugensson, M. (2015). Land use regression as method to model air pollution. Previous results for Gothenburg/Sweden. *Procedia Engineering, 115*, 21–28.

Hagler, G. S. W., Thoma, E. D., & Baldauf, R. W. (2010). High-resolution mobile monitoring of carbon monoxide and ultrafine particle concentrations in a near-road environment. *Journal of the Air & Waste Management Association, 60*(3), 328–336.

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

He, H. S., DeZonia, B. E., & Mladenoff, D. J. (2000). An aggregation index (AI) to quantify spatial patterns of landscapes. *Landscape Ecology, 15*(7), 591–601.

HKEPD (2005). *An overview on air quality and air pollution control in Hong Kong.* Hong Kong: The Environment Protection Department of the Government of the Hong Kong Special Administrative Region.

HKEPD (2013). *Air quality monitoring stations info.* Hong Kong: Hong Kong Environmental Protection Department (HKEPD).

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment, 42*(33), 7561–7578.

Isakov, V., Touma, J. S., & Khlystov, A. (2007). A method of assessing air toxics concentrations in urban areas using mobile platform measurements. *Journal of the Air & Waste Management Association, 57*(11), 1286–1295.

Janhäll, S. (2015). Review on urban vegetation and particle air pollution – Deposition and dispersion. *Atmospheric Environment, 105*(Suppl. C), 130–137.

Kim Oanh, N., Upadhyay, N., Zhuang, Y.-H., Hao, Z.-P., Murthy, D., Lestari, P., ... Dung, N. (2006). Particulate air pollution in six Asian cities: Spatial and temporal distributions, and associated sources. *Atmospheric Environment, 40*(18), 3367–3380.

Landsberg, H. E. (1981). *The urban climate.* London: Academic press.

Lau, A., Lo, A., Gray, J., Yuan, Z., & Loh, C. (2007). *Relative significance of local vs. regional sources: Hong Kong's air pollution.* Civic Exchange.

Lau, S. S. Y., Giridharan, R., & Ganesan, S. (2005). Multiple and intensive land use: Case studies in Hong Kong. *Habitat International, 29*(3), 527–546.

Leung, Y., Mei, C.-L., & Zhang, W.-X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A: Economy and Space, 32*(1), 9–32.

Li, H., & Reynolds, J. F. (1993). A new contagion index to quantify spatial patterns of landscapes. *Landscape Ecology, 8*(3), 155–162.

Li, Z., Che, W., Frey, H. C., & Lau, A. K. H. (2018). Factors affecting variability in PM2.5 exposure concentrations in a metro system. *Environmental Research, 160*(Suppl. C), 20–26.

Li, Z., Che, W., Frey, H. C., Lau, A. K. H., & Lin, C. (2017). Characterization of PM2.5 exposure concentration in transport microenvironments using portable monitors. *Environmental Pollution, 228*(Suppl. C), 433–442.

Lightowlers, C., Nelson, T., Setton, E., & Keller, C. P. (2008). Determining the spatial scale for analysing mobile measurements of air pollution. *Atmospheric Environment, 42*(23), 5933–5937.

Lin, H., Ma, W., Qiu, H., Wang, X., Trevathan, E., Yao, Z., Dong, G.-H., Vaughn, M. G., Qian, Z., & Tian, L. (2017). Using daily excessive concentration hours to explore the short-term mortality effects of ambient PM2.5 in Hong Kong. *Environmental Pollution, 229*(Suppl. C), 896–901.

McGarigal, K. (2006). *Landscape pattern metrics. Encyclopedia of environmetrics*. John Wiley & Sons Ltd.

McGarigal, K., Cushman, S. A., & Ene, E. (2012). *FRAGSTATS: Spatial pattern analysis program for categorical maps.* Amherst, Massachusetts, USA: The University of Massachusetts.

Mills, G., Ching, J., See, L., Bechtel, B., & Foley, M. (2015). *An introduction to the WUDAPT project. ICUC9 – 9th International conference on urban climate jointly with 12th symposium on the urban environment.* France: The International Association for Urban Climate (IAUC) and the American Meteorological Society (AMS) Toulouse.

Neel, M. C., McGarigal, K., & Cushman, S. A. (2004). Behavior of class-level landscape metrics across gradients of class aggregation and area. *Landscape Ecology, 19*(4), 435–455.

Ng, E. (2009). Policies and technical guidelines for urban planning of high-density cities–air ventilation assessment (AVA) of Hong Kong. *Building and Environment, 44*(7), 1478–1488.

Nowak, D. J., Crane, D. E., & Stevens, J. C. (2006). Air pollution removal by urban trees and shrubs in the United States. *Urban Forestry & Urban Greening, 4*(3–4), 115–123.

Orfanidis, S. J. (1995). *Introduction to signal processing.* Prentice-Hall Inc.

Pielke, R. A., & Avissar, R. (1990). Influence of landscape structure on local and regional climate. *Landscape Ecology, 4*(2), 133–155.

Pielke, R. A., Marland, G., Betts, R. A., Chase, T. N., Eastman, J. L., Niles, J. O., & Running, S. W. (2002). The influence of land-use change and landscape dynamics on the climate system: Relevance to climate-change policy beyond the radiative effect of greenhouse gases. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 360*(1797), 1705–1719.

Ramachandran, G., Adgate, J. L., Pratt, G. C., & Sexton, K. (2003). Characterizing indoor and outdoor 15 minute average PM 2.5 concentrations in urban neighborhoods. *Aerosol Science & Technology, 37*(1), 33–45.

Ren, C., Wang, R., Cai, M., Xu, Y., Zheng, Y., & Ng, E. (2016). *The accuracy of LCZ maps generated by the world urban database and access portal tools (WUDAPT) method: A case study of Hong. 4th Int. Conf. Countermeasure Urban Heat Islands, Singapore*.

Ross, Z., English, P. B., Scalf, R., Gunier, R., Smorodinsky, S., Wall, S., & Jerrett, M.

(2006). Nitrogen dioxide prediction in Southern California using land use regression modeling: Potential for environmental health analyses. *Journal of Exposure Science and Environmental Epidemiology, 16*(2), 106–114.
Roy, H.-Y., & Mark, C. (1996). Quantifying landscape structure: A review of landscape indices and their application to forested landscapes. *Progress in Physical Geography, 20*(4), 418–445.
Schwela, D., Haq, G., Huizenga, C., Han, W. J., & Fabian, H. (2012). *Urban air pollution in Asian cities: Status, challenges and management.* Taylor & Francis.
Shi, Y., Lau, K. K.-L., & Ng, E. (2016). Developing street-level PM2.5 and PM10 land use regression models in high-density Hong Kong with urban morphological factors. *Environmental Science & Technology, 50*(15), 8178–8187.
Shi, Y., Lau, K. K.-L., & Ng, E. (2017). Incorporating wind availability into land use regression modelling of air quality in mountainous high-density urban environment. *Environmental Research, 157*, 17–29.
Shi, Y., Lau, K. K.-L., Ren, C., & Ng, E. (2018). Evaluating the local climate zone classification in high-density heterogeneous urban environment using mobile measurement. *Urban Climate, 25*, 167–186.
Simpson, J. E. (1994). *Sea breeze and local winds.* New York, USA: Cambridge University Press.
Southworth, J., Nagendra, H., & Tucker, C. (2002). Fragmentation of a landscape: Incorporating landscape metrics into satellite analyses of land-cover change. *Landscape Research, 27*(3), 253–269.
Stewart, I. D., & Oke, T. R. (2012). Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society, 93*(12), 1879–1900.
Swietlicki, E. (2004). *Influence of particle-bound water on aerosol mass measurements. KIKU seminar – Sampling, real-time monitoring and conditioning of air samples for determination of particle mass, Copenhagen.*
Taylor, B. (1986). Geography in Hong Kong. *The Professional Geographer, 38*(4), 419–423.
Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological), 58*(1), 267–288.
Tischendorf, L., & Fahrig, L. (2000). On the usage and measurement of landscape connectivity. *Oikos, 90*(1), 7–19.
Turner, M. G., & Gardner, R. H. (2015). *Landscape ecology in theory and practice: Pattern and process.* New York: Springer.
UN (2014). *World urbanization prospects, the 2014 revision.* New York: UN, Department of Economic and Social Affairs United Nations, Population Division.
UN (2016). *Vast majority of world – 6.76 billion people – living with excessive air pollution – UN report.* UN News Centre.
UNEP (2012). *Global Environment Outlook 5 (GEO-5).* United Nations Environment Programme.
van Donkelaar, A., Martin, R. V., Spurr, R. J. D., & Burnett, R. T. (2015). High-resolution satellite-derived PM2.5 from optimal estimation and geographically weighted regression over North America. *Environmental Science & Technology, 49*(17), 10482–10491.
Vienneau, D., de Hoogh, K., Bechle, M. J., Beelen, R., van Donkelaar, A., Martin, R. V., ... Marshall, J. D. (2013). Western European land use regression incorporating satellite- and ground-based measurements of NO_2 and PM10. *Environmental Science & Technology, 47*(23), 13555–13564.
Vos, P. E. J., Maiheu, B., Vankerkom, J., & Janssen, S. (2013). Improving local air quality in cities: To tree or not to tree? *Environmental Pollution, 183*, 113–122.
Wang, R., Ren, C., Xu, Y., Lau, K. K.-L., & Shi, Y. (2017). Mapping the local climate zones of urban areas by GIS-based and WUDAPT methods: A case study of Hong Kong. *Urban Climate* in press.
Weaver, C. P., & Avissar, R. (2001). Atmospheric disturbances caused by human modification of the landscape. *Bulletin of the American Meteorological Society, 82*(2), 269–281.
Westerdahl, D., Fruin, S., Sax, T., Fine, P. M., & Sioutas, C. (2005). Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. *Atmospheric Environment, 39*(20), 3597–3610.
WHO (2016). *World health statistics 2016: Monitoring health for the SDGs sustainable development goals.* Geneva, Switzerland: World Health Organization.
WHO, & UNAIDS (2006). *Air quality guidelines: Global update 2005.* World Health Organization.
Wong, T. W., Tam, W. S., Yu, T. S., & Wong, A. H. S. (2002). Associations between daily mortalities from respiratory and cardiovascular diseases and air pollution in Hong Kong, China. *Occupational and Environmental Medicine, 59*(1), 30–35.
Wu, J., Xie, W., Li, W., & Li, J. (2015). Effects of urban landscape pattern on PM2.5 pollution—A Beijing case study. *PLoS One, 10*(11), e0142449.
Xian, G. (2007). Analysis of impacts of urban land use and land cover on air quality in the Las Vegas region using remote sensing information and ground observations. *International Journal of Remote Sensing, 28*(24), 5427–5445.
Xu, W., Riley, E. A., Austin, E., Sasakura, M., Schaal, L., Gould, T. R., ... Vedal, S. (2017). Use of mobile and passive badge air monitoring data for NOX and ozone air pollution spatial exposure prediction models. *Journal of Exposure Science and Environmental Epidemiology, 27*(2), 184–192.
Yuan, Z., Lau, A. K. H., Zhang, H., Yu, J. Z., Louie, P. K. K., & Fung, J. C. H. (2006). Identification and spatiotemporal variations of dominant PM 10 sources over Hong Kong. *Atmospheric Environment, 40*(10), 1803–1815.
Zhang, X. (1996). Issues in aerosol measurement for optics assessments. *Journal of Geophysical Research, 101*(D14), 19189–19197.