Exploring spatiotemporal variation characteristics of exceedance air pollution risk using social media big data

Zheng Cao¹ ², Zhifeng Wu¹ ², Shaoying Li¹ ², Wenjun Ma¹, Yujiao Deng⁴, Hui Sun¹ ² and Wenchuan Guan¹ ²

¹ School of Geographical Sciences, Guangzhou University, Guangzhou 510006, People’s Republic of China
² Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, People’s Republic of China
³ Guangdong Provincial Institute of Public Health, Guangdong Provincial Center for Disease Control and Prevention, Guangzhou 511430, People’s Republic of China
⁴ Ecological Meteorological Center of Guangdong Province, Guangzhou 510080, People’s Republic of China

E-mail: gzuwzf@163.com and zfwu@soil.gd.cn

Keywords: PM₂.₅, risk assessment, population mobility, TUD data, daily cumulative variation

Abstract

Air pollution in the form of PM₂.₅ decreases life expectancy considering its contribution to morbidity and mortality. Therefore, scientific and accurate PM₂.₅ exposure risk assessment is essential. However, the considering daily/hourly mean PM₂.₅ concentration and overlooking population mobility in exposure risk assessments result in underestimation of its adverse effects. Thus, using social media data and exceedance PM₂.₅ concentration, two novel indicators named hourly exceedance PM₂.₅ exposure (HEPE) and daily cumulative variation of exceedance PM₂.₅ exposure (DCEPE) are developed in our study. Spatiotemporal variation analysis of HEPE showed that the first exceedance PM₂.₅ exposure risk was observed at 10:00; this lasted till the end of the day. According to the standard deviation ellipse (SDE) method analysis results, at first, the major spatial tendency direction was northeast-southwest, with the average center (AE) located in the Yuangang Township. Then, the HEPE in the western study area increased dramatically. The major spatial tendency direction shifted from northeast-southwest to east-west. Consequently, the AE of HEPE shifted to central study area. The spatiotemporal variation characteristics led us to investigate the mechanisms. A bivariate LISA was applied to detect the spatial association between DCEPE and city functional zones (CFZs). Results showed that highly spatial associations were found between the DCEPE and CFZs in the southwest of the study area. Residential neighborhoods and transportation services showed a closer relationship with the spatial distribution of DCEPE. Based on these results, we found increasing public health threats posed by PM₂.₅. Thus, HEPE is an essential factor to assess air pollution exposure risk. Moreover, more attention should be paid to the spatial association between DCEPE and city functions, which is important for the development of air pollution mitigation strategies.

Abbreviation interpretations:

EP exceedance concentration of PM₂.₅
HEPE hourly exceedance PM₂.₅ exposure
DCEPE daily cumulative variation of exceedance PM₂.₅ exposure
TUD Tencent user density
SDE standard deviation ellipse
CFZs city functional zones

1. Introduction

Inhalation of PM₂.₅ has been widely documented in order to determine the adverse effects on public health (Lu et al 2015, Shi et al 2015b, Xie et al 2016, Ruan et al 2019, Zhang et al 2019). In 2010, Institute for Health Metrics and Evaluation and World Health Organization conducted the Global Burden of Disease study. Results revealed that over 3.2 million premature deaths were associated with PM₂.₅ pollution, because of its increasing of morbidity and mortality from cardiovascular and respiratory diseases (Chen et al 2017a). Future climate projections suggest that the potential PM₂.₅ risk will be exacerbated, considering its enhanced intensity and frequency. Moreover, the urbanization and natural disaster, such as wildfire, are also treated as potential driving factors of
PM$_{2.5}$ increasing (Ikeda and Tanimoto 2015, Vadrevu et al 2017). Therefore, PM$_{2.5}$ exposure risk assessment is of vital importance for development of preventive measures to alleviate its adverse effects on public health. Current investigation selected daily/hourly mean PM$_{2.5}$ concentration values as major indicator to evaluate the exposure risk (Holloman et al 2004, Talbott et al 2014, Weber et al 2016, Lu et al 2019, Desouza et al 2019). However, application of mean PM$_{2.5}$ concentration value ignores the hourly variation of PM$_{2.5}$ exposure risk also leads to underestimate of its adverse health influences, when PM$_{2.5}$ concentration is higher than the standards developed by WHO (Dellino et al 2002, Lin et al 2018). For example, daily mean concentration of PM$_{2.5}$ is 25 µg m$^{-3}$, while hourly PM$_{2.5}$ concentration within this day might larger than 25 µg m$^{-3}$. Applying daily mean concentration results in the underestimation of PM$_{2.5}$ exposure risk. Therefore, hourly exceedance PM$_{2.5}$ concentration should be a more effective predictor for assessment of PM$_{2.5}$ public health risk, instead of the hourly mean value.

Populations are exposed to different levels of air pollution, which is partially attributed to different population mobility patterns. Population mobility patterns vary geographically due to the spatial distribution of residential and working zones, the selection of traffic modes, and social backgrounds. Therefore, population mobility patterns are not only associated with the PM$_{2.5}$ exposure levels but also the adaptive measures (Kang et al 2013, Shi et al 2015a, Kwan and Schwanen 2016, Ma et al 2020). For example, houses with ventilation and filtration facilities protected the residents from the adverse effect of high PM$_{2.5}$ exposure. Moreover, older people (>65 years old) are more vulnerable to high exposure of PM$_{2.5}$ (Vlachokostas et al 2014, Kioumourtzoglou et al 2015, Ben-David and Waring 2018, Barkjohn et al 2020). Thus, population mobility should be a major indicator to be considered in air pollution exposure risk assessment, instead of simply relying on air pollution monitoring data. However, improvement of air pollution assessment was limited due to the lack of spatial fine-grained population mobility data. Traditionally, survey data shows good spatial coverage; however, spatial and temporal resolution is low. Temporal scale of survey data is five year due to its large social and economic costs. This data is hard to obtain due to its confidentiality. Hence, hourly or daily air pollution assessment cannot be achieved. Big data, including social media data, such as Twitter, Wechat, or Weibo data, are defined as volunteered geographic information (VGI) data (Mennis and Guo 2009, Elwood et al 2012, Fan et al 2014). These data allow the platforms to record geographical information, when GPS-enabled smartphone users uses relative applications (Liu et al 2015; Song et al 2019; Yao et al 2017a). With the use of this type of data, investigations on hourly or daily air pollution exposure risk assessment can be conducted.

Thus, air pollution assessment solely using mean PM$_{2.5}$ concentration data, without considering exceedance PM$_{2.5}$ concentration and population mobility leads to underestimation of the PM$_{2.5}$ health burden. Consequently, a more scientific indicator considering exceedance PM$_{2.5}$ concentration and population mobility is defined called hourly exceedance PM$_{2.5}$ exposure (HEPE). Furthermore, the net variation characteristics and daily cumulative variation results are also investigated in the highly urbanized area, Tianhe District. The findings from this study will help us gain new insights into air pollution assessment and help governors develop proper mitigation strategies.

2. Study area

Tianhe District is located between 23°6’ N to 23°15’ N and 113°16’ E and 113°26’ E (figure 1). As an economic and political center, more than 20% of the total gross domestic product and 10% of the population in Guangzhou is attributed to Tianhe District. Great economic development and population explosion results in significant energy demands, which lead to exacerbation of air pollution. As a consequence, annual mean PM$_{2.5}$ is more than three times greater than the annual mean concentration standard set by the World Health Organization (WHO) (Organization 2006). Here, the population faces a public health threat due to heavy air pollution, particularly PM$_{2.5}$. Therefore, detailed air pollution risk assessment methods should be developed to provide high spatiotemporal PM$_{2.5}$ exposure variation results.

3. Data and methodology

3.1. Data

3.1.1. PM$_{2.5}$ monitoring data

The hourly concentration of PM$_{2.5}$ was collected from 11 monitoring stations on May 17, 2019. PM$_{2.5}$ concentrations have been regularly monitored in these 11 monitoring stations, since 2014. Data quality control is carried out before analysis, with the invalid and missing values being delaminated. Data is provided by Guangdong Provincial Meteorological Bureau. PM$_{2.5}$ concentration value is recorded every 5 min. Data used in our study is the mean value within one hour.

In this study, exceedance concentration of PM$_{2.5}$ (EP) is applied instead of mean concentration PM$_{2.5}$ value. The EP is defined as the hourly PM$_{2.5}$ concentration exceeding 25 µg m$^{-3}$. For example, if the hourly PM$_{2.5}$ concentration in one monitoring station is 40 µg m$^{-3}$, then the EP value is 15 µg m$^{-3}$. If hourly PM$_{2.5}$ concentration value is less than 25 µg m$^{-3}$, then the EP value is 0. The standard value
of 25 μg m$^{-3}$ is defined as the safe PM$_{2.5}$ exposure level as per WHO (Organization 2006). In this study, EP is calculated based on the interpolated hourly PM$_{2.5}$ data.

3.1.2. Population mobility data

In this study, hourly Tencent user density (TUD) dataset on May 17, 2019 obtained from the Tencent platform (www.qq.com) was used for population mobility data. The TUD dataset records the locations of smart phone users when they are checking in or checking out Tencent products, such as Wechat (social media chatting application), QQ (immediate message application), Tencent Map (mobile navigation software). In our study, TUD data recorded relative long-ins numbers per hour. As one of the largest social media platform providers, more than 93% of residents in Guangzhou are TUD creators. Four advantages were observed: (1) real-time performance: this data was obtained from the mobile platform with a high accuracy of geographic coordinates every hour. (2) High spatial resolution: the spatial resolution of the data was 25 m, which is larger than that of the common social media data, for example, the cellular signaling data (spatial resolution is 100–500 m) or check-in data of Wechat or Tencent (spatial resolution is 1000 m). (3) Large coverage area for the popular used of Tencent applications. (4) Easy acquirement: Tencent provides specific ports for users to obtain this data. Therefore, the TUD dataset is widely acknowledged for population mobility related analysis, such as population dynamic mapping (Yao et al. 2017a, Chen et al. 2017b), rail transit ridership analysis (Li et al. 2020), urban park visitors analysis (Chen et al. 2018). Further, owing to its broadest coverage and high spatiotemporal resolution, it is used in this study.

3.1.3. Points of interest (POIs) data

In this study, POI data were provided by Baidu Map (https://map.baidu.com/). A total record of 65,671 points was collected. Six categories were included: catering, education and cultural, financial, medical, residential community, and transportation services. These data were applied to extract city functional zones (CFZs) to investigate the spatial association between PM$_{2.5}$ exposure and CFZs.

3.2. Methodology

3.2.1. Spatial pattern detection method

Spatial distribution patterns of hourly exceedence PM$_{2.5}$ was measured by using the standard deviation ellipse (SDE) method. Average center (AE) and standard distance of other points away from the AE are two calculated results obtained from this method. The AE describes the center location of PM$_{2.5}$ concentration. A long diameter expresses the discretization of major directional distributions of PM$_{2.5}$ concentration, while the short diameter represents the distribution range of the secondary directional distribution of PM$_{2.5}$ concentration. A large difference between
long diameter and short diameter represents a more evident directional distribution of PM$_{2.5}$ concentration. In this study, the SDE method is applied to not only detect the directional distribution of PM$_{2.5}$, but also determine the hourly moving trace of local PM$_{2.5}$ concentration, as this trace cannot be revealed at a pixel level.

### 3.2.2. Exceedance exposure assessment

A novel indicator was developed to quantitatively address the HEPE. This indicator is calculated as follows:

$$\text{HEPE} = \frac{p_{mi}}{\sum_{i=1}^{n} p_{mi}} \times 100\% \times EP_i$$  \hspace{1cm} (1)

where $p_{mi}$ represents the population mobility value of grid $i$ obtained from TUD, $n$ represents the number of grids within the study area, and $EP_i$ represents the exceedance PM$_{2.5}$ exposure value. Six PM$_{2.5}$ monitoring stations were located within the study area. We used PM$_{2.5}$ data from 11 monitoring stations to conduct interpolation to obtain spatial distribution maps of $EP_i$ result. Then HEPE was calculated based on the $EP$ result.

### 3.2.3. Exceedance exposure change detection

To address the change of exceedance exposure, two indicators were developed: (1) net variation of exceedance PM$_{2.5}$ exposure (NEPE) and (2) daily cumulative variation of exceedance PM$_{2.5}$ exposure (DCEPE). These two indicators are calculated as follows:

$$\text{NEPE}_i = \text{HEPE}^{t+1}_i - \text{HEPE}^t_i$$  \hspace{1cm} (2)

where $\text{HEPE}^{t+1}_i$ represents the HPE assessment result at analysis unit $i$ at time $t+1$ and $\text{HEPE}^t_i$ represents the hourly exceedance PM$_{2.5}$ exposure assessment result at analysis unit $i$ at time $t$.

$$\text{DCEPE} = \sum_{i=0}^{23} \text{NEPE}_i$$  \hspace{1cm} (3)

where $\text{NEPE}_i$ represents the net variation of exceedance PM$_{2.5}$ exposure result at hour $i$.

### 3.2.4. CFZs extraction

In this study, a measure called POI density is applied to determine the CFZs. In this measure, the POI density is defined by the occurrence of a POI type in the neighborhood of a location relative to the occurrence of these POI types in the study area (Verburg et al 2004). This is calculated as follows:

$$F_{i,l} = \frac{n_{i,l}/n_i}{N_{i}/N}$$  \hspace{1cm} (4)

where $F_{i,l}$ characterizes the enrichment of type $l$ POIs in type $i$ neighborhood, $n_{i,l}$ represents the number of type $l$ POIs in neighborhood $i$, $n_i$ represents the total number of POIs in neighborhood $i$, $N_{i}$ represents the total number of type $l$ POIs in whole study area, and $N$ represents the total number of POIs in the study area.

#### 3.2.5. Potential determining factor investigation

The mechanism driving the spatial distribution of DCEPE was investigated using bivariate LISA spatial autocorrelation. This method is used to detect the spatial patterns of two associated factors: DCEPE and CFZs. As shown in figure 2, four types of patterns were defined: High-High (H-H), Low-Low (L-L), High-Low (H-L), and Low-High (L-H) (Anselin 1993, 1995). H-H represents the increase of DCEPE and one type of CFZ. Therefore, this type of CFZ is defined as a potential determining factor.

### 4. Results

#### 4.1. Temporal variation characteristics of exceedance air pollution

The temporal variation characteristics of PM$_{2.5}$ concentrations at hourly levels is illustrated in figure 3. The linear regression model showed that PM$_{2.5}$ concentration value increased with time; the slope was larger than 0.8. This indicated that PM$_{2.5}$ concentrations within the study periods were higher in evening than in the morning. At hourly levels, the first exceedance PM$_{2.5}$ concentration was observed at 10:00. Then, the PM$_{2.5}$ concentration value was above 25.00 $\mu$g m$^{-3}$. The lowest exceedance PM$_{2.5}$ concentration was observed at 17:00, with a value of 0.25 $\mu$g m$^{-3}$. The highest exceedance PM$_{2.5}$ concentration was observed at 22:00, with a value of 14.09 $\mu$g m$^{-3}$.

#### 4.2. Spatial pattern of exceedance air pollution

Spatial variation characteristics in the AE and error ellipse of HEPE in Tianhe District is illustrated in figure 4. At 10:00, the AE was located in Yuangang Township, northwest of Tianhe District. At 23:00, the AE was located in Changxing Township, central west of Tianhe District. During this period, the AE moved southeast. This indicated that the west of the study area experienced more severe exceedance PM$_{2.5}$ exposure than the east. The long and short diameters of the error ellipse characterized the spatial tendency. From 10:00 to 23:00, the long diameter decreased from 16.39 km to 10.81 km, whereas the short diameter increased from 1.12 km to 7.51 km. The decrease in the long diameter and increase in the short diameter indicated the uneven spatial variation of HEPE. The HEPE of the area outside the error ellipse increased faster than that inside the error ellipse. Moreover, the variation of the long and short diameters also implied that the major spatial tendency direction shifted from northeast-southwest to
east-west. This showed that the HEPE variation during the study period was more affected by the eastern area of Tianhe District.

4.3. Exceedance air pollution exposure risk assessment
Temporal variation in net variation of exceedance PM$_{2.5}$ exposure (NEPE), spatial distribution of daily mean of NEPE, DCEPE, and CFZs extraction results are illustrated in figure 5. As shown in figure 5 A, NEPE variation exhibited two peaks and two troughs. The first peak was observed at 13:00 with an NEPE value of 0.18, while the other one was observed at 21:00, with an NEPE value of 0.83. The two troughs were observed at 15:00 and at 19:00, respectively. The NEPE values at the two troughs were 0.00006 and 0.70, respectively.

As showed in figure 5 B and C, spatial distribution of NEPE and DCEPE varied geographically. The daily mean of NEPE ranged from 0 to 0.13, whereas the DCEPE value ranged from 0 to 1.65. This indicated that the DCEPE was positive over the entire study area. We overlaid the township boundary and the DCEPE; high DCEPE areas were located in the west.
and central study area, including Xinghua Township, Shadong Township, Shahe Township, Wushan Township, Linhe Township, Tianhan Township, Shipai Township, Tangxia Township, Tianyuan Township, and Yuancun Township. To validate the DCEPE results, we collected the respiratory disease mortality case data for these townships in 2015 from Guangdong Provincial Center for Disease Control and Prevention. Results showed that more than 36% of respiratory disease mortality cases were observed in the above townships.

To quantitatively validate the DCEPE results, the partial regression model was used, controlling population factor. Result showed that correlation coefficient between NEPE and mortality was 0.71. This indicated that the potential association between NEPE and mortality exits.

The spatial distribution of DCEPE led us to investigate the mechanism driving this phenomenon. Hence, we linked the CFZ and DCEPE results. As shown in figure 6, four types of CFZs dominated the spatial distribution of DCEPE, including catering services, financial services, residential neighborhoods, and transportation services. Two major types of spatial associations were detected: H-H and L-L. The areal percentage of H-H for catering services, financial services, residential neighborhoods, and transportation services accounted for 14.36%, 15.43%,
Figure 5. Temporal variation of NEPE and spatial distribution of NEPE, DCEPE, and CFZ.

Figure 6. Spatial associations between different types of CFZ and DCEPE. A. Spatial association between catering services and DCEPE, B. spatial association between financial services and DCEPE, C. spatial association between living neighborhoods and DCEPE, and D. spatial association between transportation services and DCEPE.
Increasing evidence has been demonstrated the adverse influence of air pollution on public health, for accelerating the progression of respiratory and cerebrovascular diseases (Friedman 2001, Leitte et al 2009, Liu et al 2017). Therefore, PM$_{2.5}$ exposure risk assessment has attracted tremendous attentions. Various new data and new methods have been applied in this research field (Haikerwal et al 2015, Huang et al 2017, Shupler et al 2018, Vodonos et al 2018).

However, the considering daily/hourly mean PM$_{2.5}$ concentration and overlooking population mobility lead to underestimation of adverse effects of PM$_{2.5}$. Therefore, by considering population mobility, we proposed two novel indicators: HEPE and DCEPE. Then, we analyzed the spatiotemporal variation characteristics of these indicators. Finally, the potential determining factors were detected using bivariate LISA. This helped us gain insights into the complex association between social activities and public health issues. This research is valuable in terms of two aspects: theoretical implications and management implications.

5. Discussion and conclusion

Increasing evidence has been demonstrated the adverse influence of air pollution on public health, for accelerating the progression of respiratory and cerebrovascular diseases (Friedman 2001, Leitte et al 2009, Liu et al 2017). Therefore, PM$_{2.5}$ exposure risk assessment has attracted tremendous attentions. Various new data and new methods have been applied in this research field (Haikerwal et al 2015, Huang et al 2017, Shupler et al 2018, Vodonos et al 2018).

However, the considering daily/hourly mean PM$_{2.5}$ concentration and overlooking population mobility lead to underestimation of adverse effects of PM$_{2.5}$. Therefore, by considering population mobility, we proposed two novel indicators: HEPE and DCEPE. Then, we analyzed the spatiotemporal variation characteristics of these indicators. Finally, the potential determining factors were detected using bivariate LISA. This helped us gain insights into the complex association between social activities and public health issues. This research is valuable in terms of two aspects: theoretical implications and management implications.

5.1. Theoretical implications

Compared with mean PM$_{2.5}$ concentrations, HEPE and DCEPE have a few advantages. The temporal and cumulative variations of PM$_{2.5}$ at hourly levels were fully considered. The health risk above the standard guided by WHO was also quantitatively addressed. Overall, the application of HEPE is more appropriate to evaluate the influences of environmental exposure on public health. This was also proved via conclusions from previous research, showing that a 500 $\mu$g m$^{-3}$ increase in PM$_{2.5}$ daily exceedance concentration hours at 3-day lags was associated with an increase of 4.55% in cardiovascular mortality (Lin et al 2017, 2018). Moreover, some studies attempted to investigate the health risk caused by exceedance PM$_{2.5}$ using the daily highest PM$_{2.5}$ concentration. However, overestimations of its adverse effects and lack of daily cumulative variation were not resolved. Instead of daily mean or daily peak PM$_{2.5}$ concentrations, the use of the novel indicators can ensure reconsideration of the WHO air pollution exposure standard. Further, mitigation strategies developed based on the application of daily mean PM$_{2.5}$ concentrations may impact the effectiveness of public health protection.

Moreover, here, population mobility data is also introduced in air pollution risk assessment research. Unlike previous related studies, PM$_{2.5}$ exposure risk assessment considers population movement patterns using TUD data. Previous studies illustrated that similar PM$_{2.5}$ exposure may also lead to differing health outcomes. This may account for the different social backgrounds of different population groups. Further, population mobility not only revealed the individual footprint but also related to the real-time PM$_{2.5}$ exposure and adaption measures. For example, people with higher education levels or better economic backgrounds, they tend to take more adaptive measures to reduce the adverse influence of PM$_{2.5}$, such as equipped with air cleaners. This kind of people usually work indoors. Indoor social media footprints were the most commonly observed. While low backgrounds or physical laborers usually work outdoors. Outdoor social media footprints were the most commonly observed. Therefore, population mobility should be considered in environmental exposure risk assessment research. In addition, social media data provides the means to investigate this issue, especially in China, which has the largest number of internet users. Social media data has been widely used in transportation planning, CFZs extraction, extreme temperature exposure risk assessment. This is its first use in exceedance PM$_{2.5}$ exposure risk analysis. This research provides a method to couple environment exposure risk analysis with population mobility analysis. Unlike previous studies, population survey data is not used in this study. Compared with population survey data, TUD dataset reflects individual PM$_{2.5}$ exposure risk level. In previous studies, the application of aggregated-level population data (population survey data) is under one assumption that population within one fixed geographical unit experience the same PM$_{2.5}$ exposure concentration. However, air pollution usually varies over space and time within the same area. This results in errors of individual PM$_{2.5}$ exposure risk assessment. Moreover, ignoring individual mobility generates the neighborhood effect averaging problems. Previous studies focused on long term PM$_{2.5}$ exposure risk assessment, while multi-temporal assessment investigations are rarely seen. However, whether individual PM$_{2.5}$ exposure at short-term or mid-term shows the similar characteristics is far from clear. TUD dataset is an assistant to couple with data used in previous studies for reflecting individual mobility at hourly level, daily level or annual level. However, the limitations of TUD should also be addressed. This type of geo-referred social media data are collected from smart phone users. However, some elders and children do not use smart phones. Therefore, use of these data can be regarded as a bias sampling of the total
population dynamics distribution, as per some studies. However, TUD have been successfully used in pioneering studies on fine-scale population mapping (Yao et al 2017a) and rail transit ridership analysis (Li et al 2020). Thus, the application of TUD can reflect population mobility patterns; moreover, its application in environmental exposure risk assessment links public health and social issue analyses.

The application of TUD data is under one assumption that all TUD data recorded outdoor users’ online footprints. Indoor and outdoor users’ online footprint were all recorded in TUD data. However, people within constructions were exposed to lower exceedance PM$_{2.5}$ exposure. HEPE and DCEPE were overestimated when all TUD data were considered. Therefore, this resulted in uncertainty. To reduce this uncertainty, urban construction information should be introduced into this kind of research. With this data, individual online footprint located within constructions are not considered. As a result, HEPE and DCEPE assessment are more accurate and precise.

Future challenges should be addressed here. In our study, HEPE and DCEPE were assessed with the coupling of air pollution monitoring data and social media data. Due to the spatial uneven of air pollution monitoring stations, we only conducted our research focusing on central downtown areas of Guangzhou. The amounts and spatial distribution of air pollution station is one major restriction to apply our methodology in rural areas. By applying of remote sensing technology, this barriers might be broken. However, the temporal resolution of remote sensing is usually larger than one day. Hourly variations of HEPE and DCEPE can also not be achieved. One good news is that China is developing air pollution monitoring networks. More automatic monitoring stations are constructing. With this development, we can transfer this challenge to one opportunity to introduce our methodology into more relative studies.

5.2. Management implications

Urbanization leads to large artificial land coverage and rapid economic development. Consequently, great energy demands and an overall increase in individual wealth. As a result, industrial consumption energy and vehicles increase. Air pollution emissions such as PM$_{2.5}$, NO$_x$, SO$_2$ become public health as well as social concerns. Moreover, artificial land coverage accelerates the urban heat island effect, which prevents the dispersion of air pollution. Therefore, air pollution control measures are essential for both environmental protection and public health (Ding et al 2019). However, to the spatiotemporal details of implementing low cost air pollution control measures were still unclear. In this study, we extracted the spatial variation pattern of HEPE, and investigated potential determining factors. Our study showed that the AE of HEPE moved from the northwest to the center of the study area. The HEPE extending direction shifted from northeast-southwest to east-west, accounting for population mobility and CFZs. As Tianhe District is a mixed residential and working area, most people working in this area reside outside the area. During weekdays, workers travel to Tianhe District via individual vehicles or public transportation, especially after 10:00, which is the beginning of the work day. Heavy vehicle use releases PM$_{2.5}$, increasing the public PM$_{2.5}$ exposure risk. Related research showed that PM$_{2.5}$ released from vehicles accounted for more than half of the total amount of PM$_{2.5}$ (Chan et al 1999, Gillies et al 2001, Vijayan and Kumar 2010, Anenberg et al 2017). During weekends, people travel to Tianhe District for commercial, catering, or leisure services. Among these services, catering services may release 2.5–9.6 times higher PM$_{2.5}$ than other services (Li et al 2015). Thus, suitable PM$_{2.5}$ control measures should be implemented in public service areas from the beginning of the work day, especially at transportation stations and eateries.

In this study, exceedance concentration of PM$_{2.5}$ (EP) was used instead of PM$_{2.5}$ concentration. EP is calculated based on the daily concentration guideline by the WHO. Daily concentration guided by WHO is 25µg m$^{-3}$, while this guideline in China is 75µg m$^{-3}$. It is believed safety even with the daily PM$_{2.5}$ concentration larger than 60µg m$^{-3}$ in China. However, as one heavy polluted area, public health in China is now still under threats. Most of the previous studies tended to assess PM$_{2.5}$ exposure risk using daily mean concentration. One US study observed that each 10µg m$^{-3}$ increase in 2-day lagged PM$_{2.5}$ was associated with 2.14% increase in mortality. This implies that daily PM$_{2.5}$ concentration is not suitable for PM$_{2.5}$ exposure risk assessment. EP is coupled with PM$_{2.5}$ concentration guideline. With the development of this indicator, PM$_{2.5}$ concentration guideline should be adjusted. Hourly variation of PM$_{2.5}$ concentration can also be quantitatively measured. As new guideline developed, EP, HEPE and DCEPE mapping could provide guidelines for commuting, travelling, and the application of protective measures. For example, high risk area of HEPE and DCEPE reminds people wear masks or the utility of air cleaners. For urban planning, lower risk area of HEPE and DCEPE can be used as recreational entertainment places. Above all, the application of EP, HEPE and DCEPE provides two major management implications: (1) this promotes the adjustment of PM$_{2.5}$ concentration guideline to further improve the exposure risk assessment. (2) Mapping of EP, HEPE and DCEPE provides suggestion for public commuting, protective measures application and urban planning.

5.3. Limitation and prospects

These methods are subject to certain limitations; constrained by the availability of population mobility
data, HEPE and DCEPE on only one specific day were investigated. Long-term exposure of PM$_{2.5}$, such as one year or longer, is supposed to have more influences on public health. With sufficient population mobility data, HEPE of one year or longer can be achieved. Therefore, the HEPE and DCEPE in different seasons or years should be discussed. Moreover, population mobility data used in this study does not provide individual characteristic information. Hence, personal vulnerability cannot be studied, which is essential for comprehensive study of environmental exposure risk assessment. Therefore, other types of social media data, such as cellular signaling data, which has individual characteristics information, should be used in such studies.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (No. 41901219, No. 41671430, No. 41801326) and Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (No. GML2019ZD0301). The authors would like to thank the editor and the anonymous reviewers for their helpful comments and suggestions.

Declaration of Competing Interest

All the authors declare no financial and personal relationships with other people or organizations that can inappropriately influence the work. There are no professional or other personal interests of any nature or type in any product, service, and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, 'Exploring spatiotemporal variation characteristics of exceedance air pollution risk using social media big data.'

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

References

Anenberg S C et al 2017 Impacts and mitigation of excess diesel-related NOX emissions in 11 major vehicle markets Nature 545 467–71
Anselin L 1993 Linking GIS and spatial data analysis in practice Geogr. Syst. 1 3–23
Anselin L 1995 Local Indicators of Spatial Association—LISA Geogr. Anal. 27 93–115
Barjokhin K K, Bergin M H, Norris C, Schauer J J, Zhang Y, Black M, Hu M and Zhang J 2020 Using low-cost sensors to quantify the effects of air filtration on indoor and personal exposure relevant PM$_{2.5}$ concentrations in Beijing, China Aerosol Air Qual. Res. 20 297–313
Ben-David T and Waring M S 2018 Interplay of ventilation and filtration: differential analysis of cost function combining energy use and indoor exposure to PM$_{2.5}$ and ozone Build. Environ. 128 320–35
Chan Y C, Simpson R W, Mctainsh G H, Vowles P D, Cohen D D and Bailey G M 1999 Source apportionment of PM$_{2.5}$ and PM$_{10}$ aerosols in Brisbane (Australia) by receptor modelling Atmos. Environ. 33 3251–68
Chen L, Shi M, Gao S, Li S, Mao J, Zhang H, Sun Y, Bai Z and Wang Z 2017a Assessment of population exposure to PM$_{2.5}$ for mortality in China and its public health benefit based on BenMAP Environ. Pollut. 221 311–7
Chen Y, Liu X, Gao W, Wang R Y, Li Y and Tu W 2018 Emerging social media data on measuring urban park use Urban Forestry Urban Greening 31 130–41
Chen Y, Liu X, Li X, Liu X, Yao Y, Hu G, Xu X and Pei F 2017b Delineating urban functional areas with building-level social media data: a dynamic time warping (DTW) distance based k-medoids method Landsc. Urban Plan. 160 45–60
Deltino R J, Zeiger R S, Seltzer J M, Street D H and Mclaren C E 2002 Association of asthma symptoms with peak particulate air pollution and effect modification by anti-inflammatory medication use Environ. Health Perspect. 110
Desouza P et al Assessing the short-term effect of PM$_{2.5}$ on cardiovascular hospitalizations in the Medicaid population: a case-crossover study 2019 Environ. Epidemiol. 3 96
Ding D, Xing J, Wang S, Liu K and Hao J 2019 Estimated contributions of emissions controls, meteorological factors, population growth, and changes in baseline mortality to reductions in ambient PM$_{2.5}$ and PM$_{2.5}$-related mortality in China, 2013–2017 Environ. Health Perspect. 127
Elwood S, Goodchild M F and Sui D Z 2012 Researching volunteered geographic information: spatial data, geographic research, and new social practice Ann. Assoc. Am. Geogr. 102 571–90
Fan H, Zipf A, Fu Q and Neis P 2014 Quality assessment for building footprints data on OpenStreetMap Int. J. Geogr. Inf. Sci. 28 700–19
Friedman M S 2001 Impact of changes in transportation and commuting behaviors during the 1996 summer olympic games in Atlanta on air quality and childhood asthma Jama 285 897
Gillies J A, Gentler A W, Sagebiel J C and Dipped W A 2001 On-road particulate matter (PM$_{2.5}$ and PM$_{10}$) emissions in the Sepulveda Tunnel, Los Angeles, California Environ. Sci. Technol. 35 1054–63
Haikerwal A, Akram M, Del Monaco A, Smith K, Sim M R, Meyer M, Tonkin A M, Abramson M J and Dennekamp M 2015 Impact of fine particulate matter (PM$_{2.5}$) exposure during wildfires on cardiovascular health outcomes J. Am. Heart Assoc. 4 1–10
Holloman C H, Bortnick S M, Morara M, Strauss W J and Calder C A 2004 A bayesian hierarchical approach for relating PM$_{2.5}$ exposure to cardiovascular mortality in North Carolina Environ. Health Perspect. 112 1288–2
Huang L, Rao C, van der Kuip T J, Bi J and Liu Y 2017 A comparison of individual exposure, perception, and acceptable levels of PM$_{2.5}$ with air pollution policy objectives in China Environ. Res. 157 78–96
Ikeda K and Tanimoto H 2015 Exceedances of air quality standard level of PM$_{2.5}$ in Japan caused by Siberian wildfires Environ. Res. Lett. 10 105001
Kang C et al 2013 Exploring human movements in Singapore: a comparative analysis based on mobile phone and taxicab usages UrbComp ’13: Proc. 2nd ACM SIGKDD Int. Workshop on Urban Computing (August 2013) pp 1–8
Kioumourtzoglou M-A, Schwartz J D, Weisskopf M G, Melly S J, Wang Y, Dominici F and Zanobetti A 2015 Long-term PM$_{2.5}$ exposure and neurological hospital admissions in the northeastern United States Environ. Health Perspect. 124 23–29
Kwan M P and Schwane T 2016 Geographies of mobility Ann. Assoc. Am. Geogr. 106 243–56
Leitte A M, Petrescu C, Franck U, Richter M, Sicu O, Ionovici R, Herbarth O and Schlink U 2009 Respiratory health, effects

10
of ambient air pollution and its modification by air humidity in Drobeta-Turnu Severin, Romania Sci. Total Environ. 407 4004–11
Li S, Lyu D, Huang G, Zhang X, Gao F, Chen Y and Liu X 2020 Spatially varying impacts of built environment factors on rail transit ridership at station level: a case study in Guangzhou, China J. Transp. Geogr. 82 102631
Li Y C, Shu M, Ho S S H, Wang C, Cao -J, Wang G-H, Wang -Y-X, Wang K and Zhao X-Q 2015 Characteristics of PM2.5 emitted from different cooking activities in China Atmos. Res. 166 83–91
Lin H et al 2018 Daily exceedance concentration hours: a novel indicator to measure acute cardiovascular effects of PM 2.5 in six Chinese subtropical cities Environ. Int. 111 117–23
Lin H, Ma W, Qiu H, Wang X, Trevathan E, Yao Z, Dong G-H, Vaughn M G, Qian Z and Tian L 2017 Using daily excessive concentration hours to explore the short-term mortality effects of ambient PM2.5 in Hong Kong Environ. Pollut. 229 896–901
Liu H, Tian Y, Xu Y, Huang Z, Huang C, Hu Y and Zhang J 2017 Association between ambient air pollution and hospitalization for ischemic and hemorrhagic stroke in China: a multicity case-crossover study Environ. Pollut. 230 234–41
Liu Y, Liu X, Gao S, Gong L, Kang C, Zhi Y, Chi G and Shi L 2015 Social sensing: a new approach to understanding our socioeconomic environments Ann. Assoc. Am. Geogr. 105 512–30
Lu F, Xu D, Cheng Y, Dong S, Guo C, Jiang X and Zheng X 2015 Systematic review and meta-analysis of the adverse health effects of ambient PM2.5 and PM10 pollution in the Chinese population Environ. Res. 136 196–204
Lu Y, Lin S, Fatmi Z, Malashock D, Hussain M M, Siddique A, Carpenter D O, Lin Z and Khojja H A 2019 Assessing the association between fine particulate matter (PM2.5) constituents and cardiovascular diseases in a mega-city of Pakistan Environ. Pollut. 252 1412–22
Ma J, Tao Y, Kwan M-P and Chai Y 2020 Assessing mobility-based real-time air pollution exposure in space and time using smart sensors and GPS trajectories in Beijing Ann. Assoc. Am. Geogr. 110 434–48
Mennis J and Guo D 2009 Spatial data mining and geographic knowledge discovery—An introduction Comput. Environ. Urban Syst. 33 403–8
Organization W H 2006 WHO air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide Global Update 2005 Summary of Risk Assessment Ruan Z et al 2019 Ambient fine particulate matter and ozone higher than certain thresholds associated with myopia in the elderly aged 50 years and above Environ. Res. 177 108581–7
Shi L et al 2015a Human mobility patterns in different communities: a mobile phone data-based social network approach Ann. GIS Geogr. Inf. Sci. 21 15–26
Shi L, Zanobetti A, Kloo I, Coul B A, Koutrakis P, Melly S J and Schwartz J D 2015b Low-concentration PM2.5 and mortality: estimating acute and chronic effects in a population-based study Environ. Health Perspect. 124 46–52
Shupler M, Godwin W, Frostad J, Gustafson P, Arku R E and Brauer M 2018 Global estimation of exposure to fine particulate matter (PM2.5) from household air pollution Environ. Int. 120 354–63
Song Y, Huang B, He Q, Chen B, Wei J and Mahmood R 2019 Dynamic assessment of PM2.5 exposure and health risk using remote sensing and geo-spatial big data Environ. Pollut. 253 288–96
Tamboli E O, Rager J R, Benson S, Ann Brink L, Bilonick R A and Wu C 2014 A case-crossover analysis of the impact of PM2.5 on cardiovascular disease hospitalizations for selected CDC tracking states Environ. Res. 134 455–65
Vadrevu K, Ohara T and Justice C 2017 Land cover, land use changes and air pollution in Asia: a synthesis Environ. Res. Lett. 12 120201–17
Verburg P H, de Nijs T C M, Ritsema van Eck J, Visser H and de Jong K 2004 A method to analyse neighbourhood characteristics of land use patterns Comput. Environ. Urban Syst. 28 667–90
Vijayan A and Kumar A 2010 Experimental and statistical analyses to characterize in-vehicle fine particulate matter behavior inside public transit buses operating on B20-grade biodiesel fuel Atmos. Environ. 44 4209–18
Vlachokostas C, Banias G, Athanasiadis A, Achillas C, Akylas V and Moussiopoulos N 2014 Cense: a tool to assess combined exposure to environmental health stressors in urban areas Environ. Int. 63 1–10
Vodonos A, Awad Y A and Schwartz J 2018 The concentration-response between long-term PM2.5 exposure and mortality; A meta-regression approach Environ. Res. 166 677–89
Weber S A, Insaf T Z, Hall E S, Talbot T O and Huff A K 2016 Assessing the impact of fine particulate matter (PM2.5) on respiratory-cardiovascular chronic diseases in the New York City Metropolitan area using Hierarchical Bayesian Model estimates Environ. Res. 151 399–409
Xie R, Sabel C E, Lu X, Zhu W, Kan H, Nielsen C P and Wang H 2016 Long-term trend and spatial pattern of PM2.5 induced premature mortality in China Environ. Int. 97 180–6
Yao Y et al 2017a Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data Int. J. Geogr. Inf. Sci. 31 1220–44
Zhang S, Lv J, Meng R, Yang Y, Achariya B K, Sun X, Lin H, Hu Q and Ruan Z 2019 The association between ambient air pollution control and stroke mortality during the 2010 Asian Games in Guangzhou, China Atmos. Environ. 217 116965–71