




Article

Real-Time Estimation of Population Exposure to PM_{2.5} Using Mobile- and Station-Based Big Data

Bin Chen ^{1,2}, Yimeng Song ³, Tingting Jiang ¹, Ziyue Chen ⁴ , Bo Huang ^{3,*} and Bing Xu ^{1,4,5,*}

¹ Ministry of Education Key Laboratory for Earth System Modelling, Department of Earth System Science, Tsinghua University, Beijing 100084, China; bin.chen792@gmail.com (B.C.); ecnu_jtt@163.com (T.J.)

² Department of Land, Air and Water Resources, University of California, Davis, CA 95616, USA

³ Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, Hong Kong, China; yimengsong@link.cuhk.edu.hk

⁴ State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China; zychen@bnu.edu.cn

⁵ Department of Geography, University of Utah, 260 S. Central Campus Dr., Salt Lake City, UT 84112, USA

* Correspondence: bohuang@cuhk.edu.hk (B.H.); bingxu@tsinghua.edu.cn (B.X.);

Tel.: +852-3943-6536 (B.H.); +86-10-6279-3906 (B.X.)

Received: 5 March 2018; Accepted: 16 March 2018; Published: 23 March 2018



Abstract: Extremely high fine particulate matter (PM_{2.5}) concentration has been a topic of special concern in recent years because of its important and sensitive relation with health risks. However, many previous PM_{2.5} exposure assessments have practical limitations, due to the assumption that population distribution or air pollution levels are spatially stationary and temporally constant and people move within regions of generally the same air quality throughout a day or other time periods. To deal with this challenge, we propose a novel method to achieve the real-time estimation of population exposure to PM_{2.5} in China by integrating mobile-phone locating-request (MPL) big data and station-based PM_{2.5} observations. Nationwide experiments show that the proposed method can yield the estimation of population exposure to PM_{2.5} concentrations and cumulative inhaled PM_{2.5} masses with a 3-h updating frequency. Compared with the census-based method, it introduced the dynamics of population distribution into the exposure estimation, thereby providing an improved way to better assess the population exposure to PM_{2.5} at different temporal scales. Additionally, the proposed method and dataset can be easily extended to estimate other ambient pollutant exposures such as PM₁₀, O₃, SO₂, and NO₂, and may hold potential utilities in supporting the environmental exposure assessment and related policy-driven environmental actions.

Keywords: air pollution exposure; human mobility; mobile phone data; dynamic assessment

1. Introduction

Air pollutants, especially fine particulate matters such as PM_{2.5} (particles with an aerodynamic diameter less than 2.5 μm), have been the focus of increasing public concern because of its strong relation with health risks [1,2]. Numerous epidemiologic studies have established robust associations between long-term exposure to PM_{2.5} and premature mortality associated with various health conditions—such as heart disease, cardiovascular and respiratory diseases, and lung cancer—that substantially reduce life expectancy [2–7]. With the unprecedented economic development and urbanization over the past three decades, the severe and widespread PM_{2.5} pollution has been one of the biggest health threats in China [8,9]. The Ministry of Environmental Protection reported that only eight of the 74 monitored cities meet China’s ambient air quality standards (annual mean: 35 μg/m³; and 24-h mean: 75 μg/m³) in 2014 [10], and the number of cities was only three in 2013 [11].

The country environmental analysis report from the Asian Development Bank shows that only <1% of 500 largest cities in China could meet the air quality guidance [12] (annual mean: $10 \mu\text{g}/\text{m}^3$; and 24-h mean: $25 \mu\text{g}/\text{m}^3$) suggested by the World Health Organization (WHO) [13].

Numerous studies have attempted to estimate ground $\text{PM}_{2.5}$ concentration levels over the past decade. As ground monitoring stations provide temporally continuous records of air pollutant concentrations, the most straightforward method applied in previous researches is using the station-based $\text{PM}_{2.5}$ observations directly to interpolate point- or surface-based $\text{PM}_{2.5}$ concentration levels [14,15], thereby offering the near real-time estimations of $\text{PM}_{2.5}$ pollution levels from local to regional scales. However, these stations are always limited in number and unevenly distributed, resulting in potential biases from interpolating local point-based measurements to surface-based estimations at a large spatial scale. Fortunately, the satellite-derived atmospheric aerosol optical depth (AOD) [16–18] has greatly advanced our understanding of spatially- and temporally- explicit changes of $\text{PM}_{2.5}$ concentrations at both regional and global scales. Over the past decade, a number of pioneering works have been devoted to quantifying the relationship between satellite-based AOD retrievals and ground-measured $\text{PM}_{2.5}$ concentrations. Here we categorize them into three major groups, (i) the chemical transport models. This type of models is based on characteristics of the vertical distribution and dispersal of aerosols, and it can further integrate aerosols' components and the effects of other pollutants to predict ground-level $\text{PM}_{2.5}$ concentrations. For example, Liu et al. [19] coupled the global atmospheric chemistry model (GEOS-CHEM) with AOD retrieved by the Multiangle Imaging Spectroradiometer (MISR) to map annual mean ground-level $\text{PM}_{2.5}$ concentrations over the contiguous United States. By simulating factors that affect the relation between AOD and $\text{PM}_{2.5}$, van Donkelaar et al. [17] estimated a global field of surface $\text{PM}_{2.5}$ concentrations with the AOD retrieved from both the Moderate-resolution Imaging Spectroradiometer (MODIS) and MISR observations. (ii) The semi-empirical models. This type of models is generally based on the modeling of the AOD- $\text{PM}_{2.5}$ relationships by incorporating environmental factors. For example, several semi-empirical models have been developed ranging from simple linear relationships to complex nonlinear relationships involving meteorological and geographic variables [20,21]. (iii) The statistical regression models. This type of models is based on statistical regressions by regarding ground-based $\text{PM}_{2.5}$ measurements as the dependent variables, and the satellite-based AOD retrievals and other factors including topography, land cover/use types, humidity, temperature, wind speed, wind direction, vertical visibility, and the height of boundary layer, etc., as the independent variables [22–27]. Despite the integration of satellite- and station-based observations has proven to be useful in improving the retrieval accuracy of $\text{PM}_{2.5}$ concentrations, the available datasets are still with a coarse temporal resolution from daily, to monthly or yearly scales, rather than depicting the spatiotemporal variation of $\text{PM}_{2.5}$ concentrations within a day.

Another critical issue relating to the estimation of population exposure to $\text{PM}_{2.5}$ pollutants is that most of existing exposure assessments always regard population as static, without considering the temporal dynamics of population distribution [14,28]. Currently, demographic data based on administrative units are the most widely used data source for estimating air pollution exposure risks. It provides accurate population census information over a certain time period based on the smallest administrative unit (i.e., census block) and often includes kinds of socio-economic attributes such as age, gender, education, and income. However, such kind of data has limitations for estimating the real-time exposure risks to air pollutants since it just regards population as a homogeneous entity for each census block without diving into the spatial heterogeneity of population distribution. More importantly, it does not consider spatiotemporal dynamics of the human mobility due to the very low updating frequency. In contrast, recent studies have demonstrated the necessity of considering spatiotemporal variability of air pollution and human mobility in exposure assessments [14,29–31]. That is because, first, air pollution concentrations are not only spatially varied but also changing across temporal scales from minutes to hours, and second, population exposure to air pollutants is actually determined by both the specific location and how much time spent on that location, rather than the assumption that

people move within regions of generally the same air quality throughout a day or other time periods. Thus, how to obtain real-time estimations of population exposure to $PM_{2.5}$ concentrations is urgently needed for instant or short-time assessments (e.g., hourly or short-term $PM_{2.5}$ concentrations are more relevant to vulnerable population groups than the daily or monthly concentrations on average [14]) and cumulative exposure effects (the aggregation of short-term assessments is more robust than the monthly or annual average).

Addressing these ubiquitous challenges, more information on human space-time location is required. Some of previous studies have tried to use surveying data, such as travel questionnaire surveys, personal GPS or smart sensor based devices [14,31,32] to delineate how an individual move in the city during his/her daily life. For example, Lu and Fang [32] used the GPS-equipped portable air sensor to measure air pollutant intakes in individual's immediate surroundings and space-time movement trajectories in Huston, Texas. However, their high expenses and limited samples within local areas barricade the data availability. The alternative approaches are to use mathematical models to simulate population mobility patterns, such as gravity model [33] and radiation model [34]. This kind of methods allow us to draw more quantitative conclusions from a larger population size, but their results are only valid for situations with similar initial parameters in the simulation process [29]. Recently, Park and Kwan [14] simulated 80 possible daily movement trajectories based on daily trip distribution data from the Congestion Management Program Report to reflect the actual commuting tendency of Los Angeles (USA) county residents, and estimated exposure risks by considering the interactions between air pollution and individuals' location. However, such kind of studies are still constrained to limited spatial and temporal scales. With the rapid growth of mobile internet, especially the location-based services of applications (apps) in the smartphones, it makes us possible to access direct spatiotemporal records of human activities [35,36]. Additionally, the high correlation between the mobile-phone locating-request records and the spatiotemporal characteristics of human activities has been revealed by many studies [37–39]. A growing number of studies have started to use mobile phone data in the field of environmental exposure assessments [29,30,40]. For example, Dewulf et al. [29] collected mobile phone data of approximately five million mobile users in Belgium to calculate the daily exposure to NO_2 . Gariazzo et al. [30] conducted a dynamic city-wide air pollution (NO_2 , O_3 , and $PM_{2.5}$) exposure assessment by using time resolved population distributions derived from mobile phone traffic data, and modelled air pollutants concentrations. Yu et al. [40] combined cell phone location data from 9886 SIMcard IDs in Shenzhen, China to assess the misclassification errors in air pollution exposure estimation. Although all these pioneering studies highlight the promising advantages of incorporating population dynamics in estimating air pollution exposure, the available datasets are still limited to sample sizes and spatiotemporal scales due to the cost and time for collecting fine-resolution data, data privacy and confidentiality issues, and computational complexities [41].

To investigate the nationwide $PM_{2.5}$ concentration risks for population in China, spatially explicit and temporally continuous studies are needed to detect hotspots, estimate vulnerability, and assess population exposure at finer temporal scales. In this paper, we propose a novel approach to achieve the real-time estimation of population exposure to $PM_{2.5}$ by integrating mobile-phone locating-request (MPL) big data and station-based $PM_{2.5}$ observations. Compared with previous studies regarding ambient pollution exposure assessments, it has the following highlights. First, the proposed method introduces the dynamics of population distribution into the nationwide exposure estimation, thereby providing an improved way to better assess the actual exposure risk to $PM_{2.5}$ at different temporal scales. Second, to the best of our knowledge, it is the first time to provide the real-time estimation of nationwide population exposure to $PM_{2.5}$ at pixel-based level (~1.2 km) in China. Third, the proposed method and dataset can be easily extended to estimate other ambient pollutant exposures such as PM_{10} , O_3 , SO_2 , and NO_2 , and may hold potential utilities in supporting the environmental exposure assessments and related policy-driven environmental actions.

2. Materials and Methods

2.1. Ground-Station $PM_{2.5}$ Measurements

Hourly ground-station $PM_{2.5}$ measurements from 1 March to 31 March 2016 were collected from the official website of the China Environmental Monitoring Center (<http://113.108.142.147:20035/emcpublish/>). According to the Chinese National Ambient Air Quality Standard (CNAAQS), the station-based $PM_{2.5}$ data in China were obtained using the tapered element oscillating microbalance method (TEOM) or the beta-attenuation method, combined with the periodic calibration. In this study, we used a total of 1465 monitoring stations (Figure 1) that have been established in all provinces for monitoring ambient air quality.

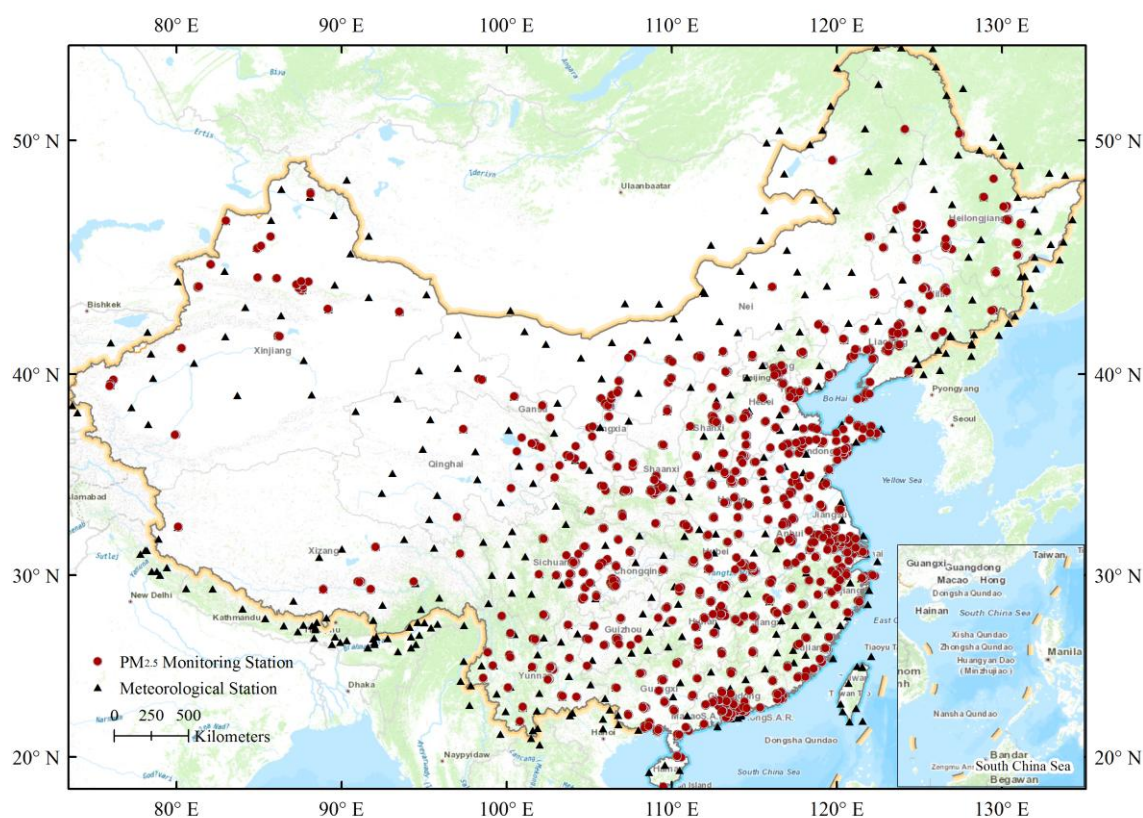


Figure 1. Spatial distribution of nationwide monitoring stations for $PM_{2.5}$ concentrations (red dots) and meteorological stations (black triangles) in China.

2.2. Ground-Station Meteorological Measurements

Ground-station meteorological variables, including air temperature (AT), surface wind speed (WS), and horizontal visibility (VIS) were used from Global Telecommunication System (GTS) established by World Meteorological Organization (<https://rda.ucar.edu/datasets/ds461.0/>). In this study, the 3-h measurements (from 2:00 a.m. to 23:00 p.m. local time) from 411 stations in China and 128 stations within the 0.01-degree buffer zones around the boundary of China (Figure 1) were collected from 1 March to 31 March 2016.

2.3. Mobile Phone Locating-Request Big Data

By retrieving real-time locating requests from mobile phone users' activities in apps, the mobile phone locating-request (MPL) data was used in this study to monitor human movement. The MPL data are from Tencent big data platform in China, which is one of the largest Internet service providers both nationwide and worldwide. All of the MPL data are produced by active smartphone users using

apps, which have been enabled to report real-time locations from the mobile devices. Due to the widespread usage of Tencent apps (e.g., WeChat, QQ, Tencent Map, etc.) and their location-based services, the daily locating records have reached 36 billion from more than 450 million users globally in 2016 [42]. Thus, the MPL big data can be represented as an indicator to characterize human activities and population distribution in a fine spatiotemporal scale. The Tencent MPL dataset used in this study was collected from 1 March to 31 March 2016 via the application program interface (API) from the Tencent big data platform (<http://heat.qq.com>). The original Tencent MPL dataset was recorded by aggregating the real-time locations of active apps users every five minutes within a mesh grid at a spatial resolution of 30 arc-second (~1.2 km). All the information regarding users' identities and privacies were removed in this publicly available dataset.

2.4. Population Census Data

The latest city-level population census of China in 2014 obtained from the national scientific data sharing platform for population and health (<http://www.ncmi.cn/>) was used in this study. This dataset was established and maintained by infectious disease network reporting system, and it was derived based on population census released by the State Statistics Bureau. It collected all population census including permanent resident and registered resident at the county level by gender and age group since 2004.

2.5. Estimation of Spatiotemporal Continuous $PM_{2.5}$ Concentrations

Due to the difference in geographic locations between $PM_{2.5}$ monitoring stations and meteorological stations, all datasets were processed to be consistent in spatial and temporal domains. The meteorological variables were first interpolated by ordinary Kriging method [43] to obtain data that covering the entire study area with a spatial resolution of 30 arc-second (~1.2 km). To mitigate the interpolation biases, we averaged all meteorological observations with a 30 arc-second search radius around each $PM_{2.5}$ monitoring station, and then assigned the result to the corresponding $PM_{2.5}$ monitoring station. In addition, the widely used Geographically Weighted Regression (GWR) model [44] with adaptive Gaussian bandwidth was adopted to build the statistical relationship between meteorological variables and $PM_{2.5}$ concentrations. Specifically, we grouped all variables within a month into 8 time points (i.e., from 2:00 a.m., 5:00 a.m., . . . , 23:00 p.m.), and then developed 8 GWR models for each time point in this study as follows:

$$PM_{2.5,i,t} = \beta_{0,i,t} + \beta_{1,i,t}VIS_{i,t} + \beta_{2,i,t}AT_{i,t} + \beta_{3,i,t}WS_{i,t} \quad (1)$$

where $PM_{2.5,i,t}$ denotes the $PM_{2.5}$ concentration at the location i at time t , $VIS_{i,t}$, $AT_{i,t}$, and $WS_{i,t}$ denote the visibility (m), air temperature ($^{\circ}C$), and surface wind speed (m/s), respectively, at location i at time t . $\beta_{0,i,t}$, $\beta_{1,i,t}$, $\beta_{2,i,t}$, and $\beta_{3,i,t}$ are corresponding regression coefficients at location i at time t .

A 10-fold validation analysis [45] was adopted to evaluate the modeling performance by comparing the estimated and measured $PM_{2.5}$ concentrations (details can be found in Supplementary Materials). With the iterative cross validations, the optimal coefficients in each time point were retrieved to interpolate the entire study areas with a spatial resolution of 30 arc-second (~1.2 km), and then were used to estimate gridded $PM_{2.5}$ concentrations.

2.6. Estimation of Real-Time Population Distribution by Integrating MPL and Census Data

The mobile phone locating-request (MPL) data can be served as an indicator to delineate the spatiotemporal pattern of population distribution, however, the MPL data do not represent the actual population sizes. In this study, we first aggregated the 5-min MPL data into 3-h MPL data, making its temporal resolution consistent with that of the estimated $PM_{2.5}$ concentrations, and calculated the pixel-based population density using the MPL data, and then applied the MPL-based population density map to downscale the census data. Consequently, we can obtain the 3-h pixel-based population

approximations. Given the difference of physical environment and socio-economic development in various areas of China, downscaling the MPL data with population census at the national scale will undoubtedly result in the underestimation of population in under- and less-developed areas and overestimation of population in those developed areas. To solve this problem, we decided to estimate real-time population distribution by integrating MPL and census data at the city level. The 3-h MPL map was used to redistribute the census data for each city by Equations (2) and (3), under the assumption that the inter-city mobility will not dramatically influence the total population of a city within a short time window. Finally, we could obtain the 3-h pixel-based population approximation for each city, and then conducted the image mosaic to produce the 3-h national-scale population distribution map in China.

$$W_{ij} = \frac{p_{ij}}{\sum_{i=1}^n p_{ij}} \quad (2)$$

$$pop_{ij} = TR \times W_{ij} \quad (3)$$

where $p_{i,j}$ is the amount of locating-request times within the i -th pixel at the hour j , n is the total number of pixels within a city, $W_{i,j}$ is the weight for redistributing population and TR is the total population in the city from the census data. $Pop_{i,j}$ denotes the population approximation in the i -th pixel at the hour j .

2.7. Real-Time Estimation of Population Exposure to $PM_{2.5}$

Since the levels of $PM_{2.5}$ concentration and population distribution are spatially and temporally varied, here we adopted the population-weighted metric (Equation (4)) to estimate the real-time exposure risks to $PM_{2.5}$ concentrations, which was likely to be more representative of population exposure to $PM_{2.5}$ across different temporal scales [46]:

$$PWP = \frac{\sum_{i=1}^N (pop_i \cdot pm_i)}{\sum_{i=1}^N pop_i} \quad (4)$$

where pop_i and pm_i denote the population and $PM_{2.5}$ concentration level in the i -th pixel, N is the total number of pixels within the corresponding administrative unit. PWP is the population-weighted $PM_{2.5}$ concentration level for the targeted administrative unit.

With the $PM_{2.5}$ concentrations and population distribution estimated in previous sections, we could integrate them based on Equation (4) to provide the estimation of population exposure to $PM_{2.5}$ with a 3-h updating frequency, thereby being able to track the real-time dynamics of exposure risks by considering the spatiotemporal variation of $PM_{2.5}$ concentration and population distribution.

2.8. Estimation of Cumulative Inhaled $PM_{2.5}$

$PM_{2.5}$ concentration causes acute and chronic adverse effects on human health mainly by means of inhalation exposure. To our understanding, deriving the estimations of cumulative inhaled $PM_{2.5}$ masses will be one of the most important prerequisites to model the accurate relationship between $PM_{2.5}$ exposure and human health [47–49]. Thus, we proposed to incorporate human respiratory volume and the spatiotemporal variation of $PM_{2.5}$ concentration and population density to present a better estimation of cumulative inhaled $PM_{2.5}$:

$$InPM_{2.5} = \sum_{i=1}^N \sum_{t=1}^T p_i(t) \cdot h_i \cdot d_i(t) \cdot m(t) + p_i(t) \cdot h_i \cdot (1 - d_i(t)) \cdot m(t) \cdot \alpha \quad (5)$$

where p_i and h_i denote the population and the inhaled volume of air for the i -th age group, N is the total number of the age group. t denotes the time (hours in this study), $m(t)$ denotes the $PM_{2.5}$ concentration level at time t , T is the target temporal period, d_i is the percentage of outdoor population, α is the outdoor-indoor ratio of $PM_{2.5}$ concentration.

However, recent advances regarding the outdoor-indoor ratio of $PM_{2.5}$ concentrations are all limited to local scales for the purpose of experimental tests [50], as it is difficult to acquire such valid observations relating this ratio on a large scale. More importantly, the outdoor-indoor ratio is influenced by several factors such as geographic location, building structures, and living habits. In addition, the inhaled volume of air is also different, not only in terms of age differences but of physical activities, gender, and size, all of these factors would affect the inhaled value [51,52]. Thus, we have to simplify the ideal model in Equation (5) for being suitable to nationwide estimates of cumulative inhaled $PM_{2.5}$ masses by neglecting the difference between outdoor and indoor $PM_{2.5}$ concentration exposure and the inhaled volume of air among different age groups, gender, and other related factors. In this way, we can directly obtain the estimation of cumulative inhaled $PM_{2.5}$ masses using the following equation:

$$InPM'_{2.5} = \sum_{t=1}^T p_i(t) \cdot h \cdot m(t) \quad (6)$$

where $InPM'_{2.5}$ denotes the cumulative inhaled $PM_{2.5}$ mass from the simplified model, and h denotes the empirical inhaled volume of air. A measurement conducted by Adams [51] based on 200 individuals showed that the hourly average volume of air breathed by adults when they are sitting or resting were ranging from 0.42 to 0.63 m^3 (i.e., 10.08 to 15.12 m^3 /day), and the volumes for walking were from 1.20 to 1.44 m^3 /h, and for running were from 3.10 to 3.48 m^3 /h. Thus, the average inhaled volume of air for an individual is assumed to be 15 m^3 /day in this study [52].

2.9. Comparison of Exposure Assessments from the MPL-Based and Census-Based Methods

In order to investigate whether the improvement of incorporating dynamic population distributions does make a difference in the exposure assessment, we intuitively compared the MPL-based and census-based calculations of cumulative inhaled $PM_{2.5}$ masses and population-weighted $PM_{2.5}$ exposure concentrations in China's 359 cities across different temporal scales (i.e., 3-h, 1-day, 1-week, and 1-month). For each city, the population from the census data was directly used in the census-based method, while the redistributed population dynamics was used in the MPL-based method.

3. Results

3.1. Different Facets of Population Exposure to $PM_{2.5}$

The spatiotemporal integration of $PM_{2.5}$ concentration and population density was used to produce thematic information that document different facets of population exposure to $PM_{2.5}$. Figure 2 shows an extracted example from the time-series analysis of population exposure to $PM_{2.5}$ in China.

Figure 2a shows the real-time nationwide estimation of population distribution (11:00 a.m.) on 1 March 2016, which is derived by integrating MPL and census data at a city-level scale in Section 2.6. The intensity represents the specific population number in each gridded pixel with stretched colors from blue to red denoting varied population size. Figure 2b shows the real-time nationwide estimation of $PM_{2.5}$ concentrations (11:00 a.m.), which is derived from incorporating ground-station $PM_{2.5}$ measurements and meteorological variables based on GWR models in Section 2.5. Figure 2c shows the nationwide estimation of 24-h cumulative inhaled $PM_{2.5}$ masses. Figure 2d shows the estimation of 24-h cumulative inhaled $PM_{2.5}$ masses based on the census data. Figure 2e–h show insets from Figure 2a–d for part of the Northern China as a zoomed visualization in different facets of population exposure to $PM_{2.5}$ concentrations.

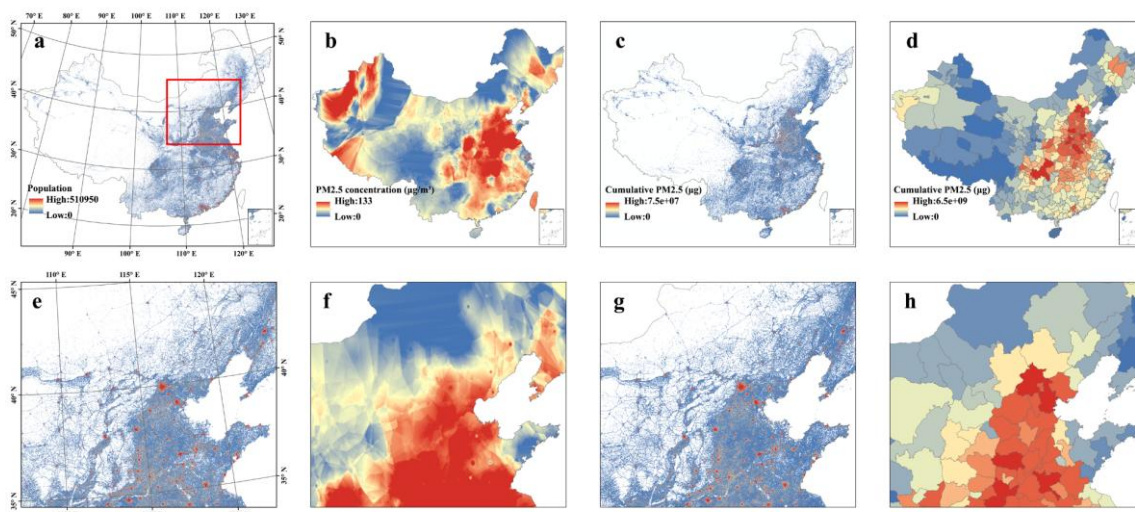


Figure 2. Different facets of population exposure to $PM_{2.5}$. (a) Map of population distribution in China on 1 March 2016 (11:00 a.m.). (b) Map of $PM_{2.5}$ concentration levels in China on 1 March 2016 (11:00 a.m.). (c) Map of cumulative inhaled $PM_{2.5}$ masses in China based on the MPL data on 1 March 2016. (d) Map of cumulative inhaled $PM_{2.5}$ in China based on the census data on 1 March 2016. (e–h) show the insets from (a–d) for part of the Northern China.

3.2. Temporal Dynamics of Population Exposure to $PM_{2.5}$

In the form of Figure 2a–c, we can also provide the temporal variation of population, $PM_{2.5}$ concentrations, and cumulative inhaled $PM_{2.5}$ masses with a 3-h temporal resolution from 1 March to 31 March 2016. In this way, the pixel-based dynamics of population exposure to $PM_{2.5}$ concentrations at the nationwide scale with a nearly real-time updating frequency (i.e., 3-h in this study) were retrieved. In order to better present the experimental results with an entire month in March 2016, we further aggregated the pixel-based estimations into 359 cities in this study. Results demonstrate that both the population-weighted $PM_{2.5}$ concentrations (Figure 3a) and cumulative inhaled $PM_{2.5}$ masses (Figure 3b) exist distinguished diurnal and daily variations, which also verify the necessity of considering the spatiotemporal variability of both air pollution and population distribution in air pollution exposure assessments.

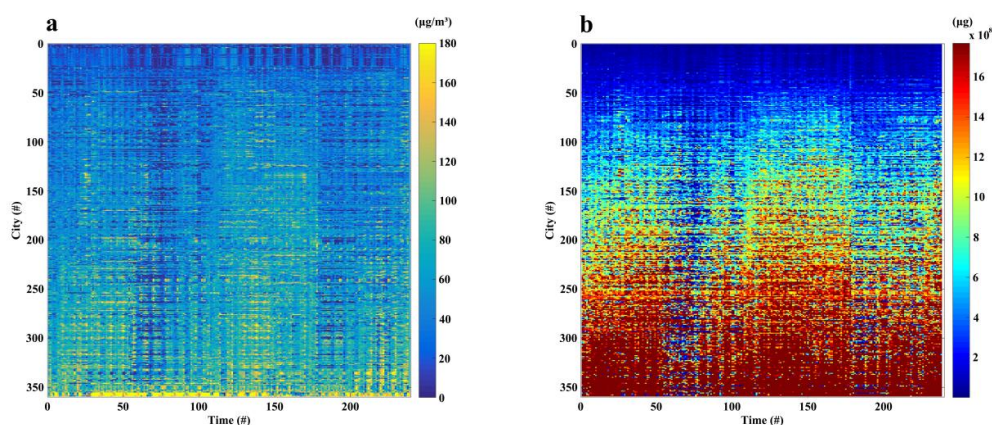


Figure 3. The estimated population-weighted $PM_{2.5}$ concentrations (a) and cumulative inhaled $PM_{2.5}$ masses (b) for 359 cities in China with every 3 h from 1 March to 31 March 2016. Note that the x axis represents the time from the first 3-h (2:00 a.m. 1 March 2016) to the last 3-h (23:00 p.m. 31 March 2016), and y axis represents the order of 359 cities.

3.3. Comparison of Exposure Assessment Methods

From the visual inspection from Figure 2c,d, it can be found out that the MPL-based method yields the gridded cumulative inhaled $PM_{2.5}$ masses, whereas the census-based assessments are only based on administrative units (cities in this study), which informs us that the MPL-based method improves the spatial resolution of basic cells from administrative units to gridded pixels in exposure assessments. In addition, by comparing the cumulative inhaled $PM_{2.5}$ masses and population-weighted $PM_{2.5}$ exposure concentrations in China's 359 cities across different temporal scales, results in Figure 4 show that without introducing the dynamics of population distribution into the exposure assessment, the maximum biases (over- or under- estimation) of cumulative inhaled $PM_{2.5}$ mass reach to over 100% across different temporal scales. Meanwhile, the maximum biases of population-weighted $PM_{2.5}$ concentrations will be approximately $30 \mu\text{g}/\text{m}^3$. By aggregating the experimental tests in China's 359 cities from 1 March to 31 March 2016, the biased percentage between the MPL-based and the census-based estimations will be the level of 14.9% (3-h), 5.8% (1-day), 4.7% (1-week), and 3.9% (1-month) on average.

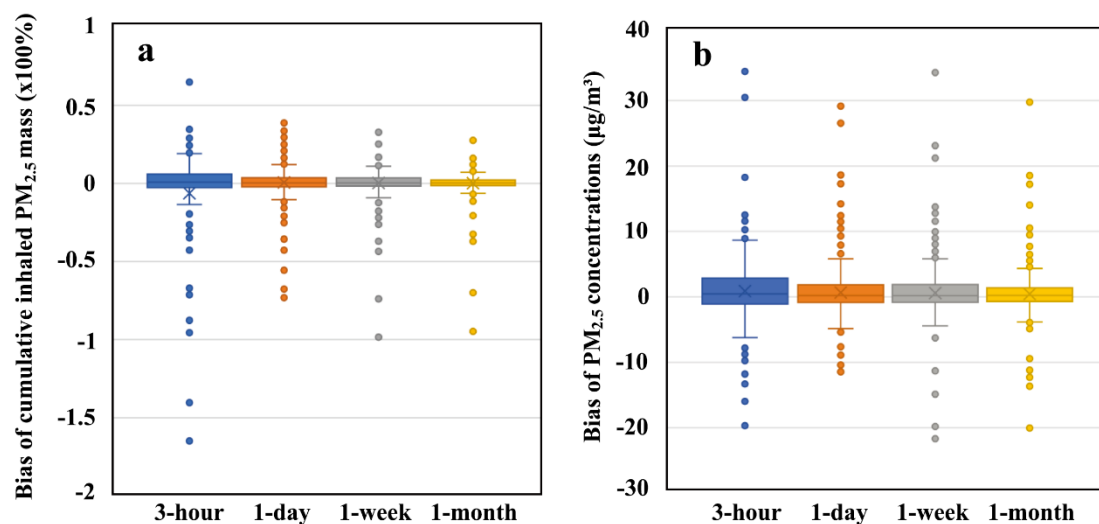


Figure 4. The biases of cumulative inhaled $PM_{2.5}$ mass (a) and the per capita $PM_{2.5}$ exposure concentration (b) between the MPL-based estimations and the census-based estimations in China's cities across different temporal scales.

4. Discussion

Compared with previous methods for air pollution exposure assessment, the proposed method in this study considered well the spatiotemporal variability of both population distribution and $PM_{2.5}$ concentration levels, thereby contributing to a better exposure assessment. The relative reasonability of our method may be due to the following strengths. First, the spatiotemporal variability of $PM_{2.5}$ concentrations and population distribution are incorporated in air pollution exposure assessments. Given that the level of $PM_{2.5}$ concentrations is continuously changing over space and time and human beings are also mobile across spatiotemporal scales [14], both of these dynamic characteristics and their interactions at finer spatiotemporal scales should be well considered to estimate population exposure risks. However, many previous studies always used the census data with the assumption that people are non-mobile or moving within regions of generally the same air quality throughout a day or other time periods, thus leading to considerable biases in actual air pollution exposure assessments. In reality, people in different areas experience different levels of $PM_{2.5}$ concentrations across different temporal scales. In order to characterize the interaction between population dynamics and $PM_{2.5}$ concentrations, here we used the mobile-phone locating-request (MPL) big data to quantify the dynamics of population

distribution. By integrating the MPL and census data, we then derived real-time pixel-based population dynamics at the nationwide scale. Combing this nationwide population dynamic information and surface-based $PM_{2.5}$ concentrations simultaneously will be of great importance to assess the actual population exposure to $PM_{2.5}$ at different temporal scales. Second, the characterized dynamics of $PM_{2.5}$ concentrations and population dynamics in the proposed method keep a consistent spatiotemporal scale. The MPL data used in this study were initially retrieved at a 5-min updating temporal resolution from the Tencent big data platform. We further aggregated the 5-min updating MPL data into 3-h synthetic data, making it temporally comparable to the updating frequency of the nationwide surface-based $PM_{2.5}$ concentrations. Meanwhile, the spatial resolution of $PM_{2.5}$ concentrations is also set to be with a 30 arc-second (~1.2 km) spatial resolution, which is the same with that of MPL data. These efforts contribute much to achieving near real-time (3-h) estimates of national population exposure to $PM_{2.5}$ at the pixel-based level in China. Third, the presented model incorporated human respiratory volume and the spatiotemporal variation of $PM_{2.5}$ concentration and population density to estimate cumulative inhaled $PM_{2.5}$ masses. It will contribute to advancing the development of modelling the relationship between $PM_{2.5}$ exposures, health risks, and life expectancies quantitatively.

Besides $PM_{2.5}$, the ground monitoring stations are always coupled with sensors measuring other air pollutants such as PM_{10} , SO_2 , NO_2 , and O_3 . With the similar framework by integrating mobile phone big data and air pollutant concentrations, the proposed method can also be customized to estimate population exposure risks to these ambient pollutants in China. Compared with the census-based method, the MPL-based method can yield near real-time estimations of population exposure to ambient pollutants. That is, we can achieve the estimation of air pollution exposure risks at any specific location and time on a large scale by combining the spatiotemporal variability of population distribution and air pollutant concentrations. By aggregating the short-term exposure assessments into longer temporal scales, we can also derive more robust and reliable estimations related to the chronic effects from air pollutants. Additionally, the proposed framework can be also applied to estimate the real-time number of people exposed to poor air quality as a result of updating the population distribution and air pollutant concentrations.

Meanwhile, some potential concerns regarding the implementation of the proposed method should be pointed out. First, in order to redistribute the census data to derive real-time population dynamics using the MPL data, we assume that the total population of each administrative unit (359 cities in this study) is constant since the inter-city mobility (the trade-off of inflow and outflow population) will not dramatically influence the total population of a city within a short time window. Thus, human movements and migrations across administrative units are neglected in this study. Second, volunteer-produced geospatial big data, such as MPL records in this study tend to leave out some population groups of the society because the children, the elderly, and the poor are less-frequent active users. Nevertheless, such data can still well quantify actual population distribution patterns [35,37,38] because of the massive volumes of data records. Here we take the MPL records in China on 1 March 2016 for example, the total number of locating-request records reaches 1.71 billion. By aggregating all MPL records from 1 March to 31 March 2016, the total number of locating-request records will be approximately 60 billion, thereby providing a robust measurement of population dynamics. Third, although the nationwide $PM_{2.5}$ concentrations used in this study are estimated by incorporating the meteorological variables and ground-based $PM_{2.5}$ measurements with the GWR models, the spatial interpolations are still the limits to affect the estimation accuracy in areas without sufficient inputs of station-based variables. As a result, even there is much greater spatial variations in the population data, there will be relatively less spatial variations in $PM_{2.5}$ concentrations, which may lead to no significant impacts on the exposure assessments. However, with the comparison of exposure assessments between the MPL-based and the census-based methods, we can still figure out considerable differences. Thus, if we can further improve the estimation of $PM_{2.5}$ concentrations, such as developing spatial-temporal integrated method by combing satellite-based and station-based observations guided with the diurnal change pattern of $PM_{2.5}$ concentrations, land

cover/use types, landscape topography, and related meteorological variables, the combination of the mobile phone big data and the improved air pollutant concentrations will contribute to a more reliable exposure assessment. Finally, the simplified model without considering outdoor-indoor ratio of PM_{2.5} concentrations and the difference of inhaled volumes of air among different population groups may be biased to the assessment of actual cumulative inhaled PM_{2.5} masses. As the Tencent-based MPL dataset was recorded by aggregating the real-time locations of active apps users within a mesh grid at a spatial resolution of 30 arc-second (~1.2 km) without differentiating individual's moving trajectories and population groups, it was impractical to apply empirical parameters into the exposure assessment at a nationwide scale since the outdoor-indoor ratio of PM_{2.5} concentrations is influenced by several factors such as geographical locations, building materials, living habits, and so on. Similarly, the gridded MPL data without tracking individuals' trajectories also prevented us from considering the commuting patterns or choices of different transports. However, the MPL dataset represents the unique data source having the best spatial resolution with real-time updating population distribution we can access right now. Meanwhile, the estimates in the experimental test also represent the trade-off between over- and under-estimated cumulative inhaled PM_{2.5} masses. On the one hand, these estimates are the highest estimates of cumulative inhaled PM_{2.5} masses since we do not consider the situations that people are with indoor environments or commuting transportations. On the other hand, the cumulative inhaled PM_{2.5} masses could be even higher because we use the constant value representing a low level of the inhaled air volume for an adult without considering factors such as physical activity, gender, and size [51]. Thus, these over and under estimates help balance each out in terms of cumulative inhaled PM_{2.5} masses to provide the general assessment at large scales.

5. Conclusions

This study sought to combine mobile phone big data and station-based PM_{2.5} measurements to achieve real-time estimations of population exposure to PM_{2.5} concentrations in China. The results showed that the proposed method can well quantify dynamics of the real-time population distribution and yield the estimation of population exposure to PM_{2.5} concentrations and cumulative inhaled PM_{2.5} masses with a 3-h updating frequency. This study provides a novel framework for environmental exposure assessments by considering the spatiotemporal variability of both population distribution and PM_{2.5} concentrations, which can also be customized to estimate other ambient pollutant exposure risks. These findings and methods may hold potential utilities in supporting the environmental exposure assessment and related policy-driven environmental actions.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1660-4601/15/4/573/s1>, Table S1. Accuracy of the fitting and 10-fold cross-validation for eight periods. Figure S1: Scatterplots of the observed and predicted PM_{2.5} for eight-time periods.

Acknowledgments: The authors thank Tencent Inc. for making the mobile phone location data publicly available. This work was supported by the Ministry of Science and Technology of China under the National Key Research and Development Program (2016YFA0600104) and was also supported by a project funded by the China Postdoctoral Science Foundation (2017M620739). The authors also thank three anonymous reviewers and editors for providing valuable suggestions and comments, which have greatly improved this manuscript.

Author Contributions: Bin Chen and Bing Xu conceived and designed the experiments; Bin Chen and Yimeng Song performed the experiments and wrote the paper; Tingting Jiang, Bo Huang, Ziyue Chen and Bing Xu contributed to the data analysis and manuscript revision.

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

References

1. Kampa, M.; Castanas, E. Human health effects of air pollution. *Environ. Pollut.* **2008**, *151*, 362–367. [[CrossRef](#)] [[PubMed](#)]
2. Pope, C.A., III; Dockery, D.W. Health effects of fine particulate air pollution: Lines that connect. *J. Air Waste Manag. Assoc.* **2006**, *56*, 709–742. [[CrossRef](#)] [[PubMed](#)]

3. Apte, J.S.; Marshall, J.D.; Cohen, A.J.; Brauer, M. Addressing global mortality from ambient *Environ. Sci. Technol.* **2015**, *49*, 8057–8066. [[CrossRef](#)] [[PubMed](#)]
4. Brook, R.D.; Rajagopalan, S.; Pope, C.A.; Brook, J.R.; Bhatnagar, A.; Diez-Roux, A.V.; Holguin, F.; Hong, Y.; Luepker, R.V.; Mittleman, M.A. Particulate matter air pollution and cardiovascular disease. *Circulation* **2010**, *121*, 2331–2378. [[CrossRef](#)] [[PubMed](#)]
5. Liu, C.; Yang, C.; Zhao, Y.; Ma, Z.; Bi, J.; Liu, Y.; Meng, X.; Wang, Y.; Cai, J.; Kan, H. Associations between long-term exposure to ambient particulate air pollution and type 2 diabetes prevalence, blood glucose and glycosylated hemoglobin levels in China. *Environ. Int.* **2016**, *92*, 416–421. [[CrossRef](#)] [[PubMed](#)]
6. Pope, C.A., III; Burnett, R.T.; Thun, M.J.; Calle, E.E.; Krewski, D.; Ito, K.; Thurston, G.D. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* **2002**, *287*, 1132–1141. [[CrossRef](#)] [[PubMed](#)]
7. Pope, C.A., III; Ezzati, M.; Dockery, D.W. Fine-particulate air pollution and life expectancy in the United States. *N. Engl. J. Med.* **2009**, *2009*, 376–386. [[CrossRef](#)] [[PubMed](#)]
8. Xu, B.; Yang, J.; Zhang, Y.; Gong, P. Healthy cities in China: A lancet commission. *Lancet* **2016**, *388*, 1863–1864. [[CrossRef](#)]
9. Xu, P.; Chen, Y.; Ye, X. Haze, air pollution, and health in China. *Lancet* **2013**, *382*, 2067. [[CrossRef](#)]
10. Ministry of Environmental Protection of the People’s Republic of China. *China’s Environmental Bulletin in 2014*; Ministry of Environmental Protection: Beijing, China, 2014.
11. Babu, S.S.; Manoj, M.; Moorthy, K.K.; Gogoi, M.M.; Nair, V.S.; Kompalli, S.K.; Satheesh, S.; Niranjan, K.; Ramagopal, K.; Bhuyan, P. Trends in aerosol optical depth over Indian region: Potential causes and impact indicators. *J. Geophys. Res. Atmos.* **2013**, *118*. [[CrossRef](#)]
12. World Health Organization; UNAIDS. *Air Quality Guidelines: Global Update 2005*; World Health Organization: Geneva, Switzerland, 2006.
13. Zhang, Q.; Crooks, R. *Toward an Environmentally Sustainable Future: Country Environmental Analysis of the People’s Republic of China*; Asian Development Bank: Mandaluyong, Philippines, 2012.
14. Park, Y.M.; Kwan, M.-P. Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. *Health Place* **2017**, *43*, 85–94. [[CrossRef](#)] [[PubMed](#)]
15. Zhang, A.; Qi, Q.; Jiang, L.; Zhou, F.; Wang, J. Population exposure to PM_{2.5} in the urban area of Beijing. *PLoS ONE* **2013**, *8*, e63486. [[CrossRef](#)] [[PubMed](#)]
16. Van Donkelaar, A.; Martin, R.V.; Brauer, M.; Kahn, R.; Levy, R.; Verduzco, C.; Villeneuve, P.J. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environ. Health Perspect.* **2010**, *118*, 847–855. [[CrossRef](#)] [[PubMed](#)]
17. Van Donkelaar, A.; Martin, R.V.; Park, R.J. Estimating ground-level PM_{2.5} using aerosol optical depth determined from satellite remote sensing. *J. Geophys. Res. Atmos.* **2006**, *111*. [[CrossRef](#)]
18. Wang, J.; Christopher, S.A. Intercomparison between satellite-derived aerosol optical thickness and PM_{2.5} mass: Implications for air quality studies. *Geophys. Res. Lett.* **2003**, *30*. [[CrossRef](#)]
19. Liu, Y.; Park, R.J.; Jacob, D.J.; Li, Q.; Kilaru, V.; Sarnat, J.A. Mapping annual mean ground-level PM_{2.5} concentrations using multiangle imaging spectroradiometer aerosol optical thickness over the contiguous united states. *J. Geophys. Res. Atmos.* **2004**, *109*. [[CrossRef](#)]
20. Lin, C.; Li, Y.; Yuan, Z.; Lau, A.K.; Li, C.; Fung, J.C. Using satellite remote sensing data to estimate the high-resolution distribution of ground-level PM_{2.5}. *Remote Sens. Environ.* **2015**, *156*, 117–128. [[CrossRef](#)]
21. Tian, J.; Chen, D. A semi-empirical model for predicting hourly ground-level fine particulate matter (PM_{2.5}) concentration in southern ontario from satellite remote sensing and ground-based meteorological measurements. *Remote Sens. Environ.* **2010**, *114*, 221–229. [[CrossRef](#)]
22. Chu, Y.; Liu, Y.; Li, X.; Liu, Z.; Lu, H.; Lu, Y.; Mao, Z.; Chen, X.; Li, N.; Ren, M. A review on predicting ground PM_{2.5} concentration using satellite aerosol optical depth. *Atmosphere* **2016**, *7*, 129. [[CrossRef](#)]
23. Hu, Z. Spatial analysis of modis aerosol optical depth, PM_{2.5}, and chronic coronary heart disease. *Int. J. Health Geogr.* **2009**, *8*, 27. [[CrossRef](#)] [[PubMed](#)]
24. Kloog, I.; Koutrakis, P.; Coull, B.A.; Lee, H.J.; Schwartz, J. Assessing temporally and spatially resolved PM_{2.5} exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos. Environ.* **2011**, *45*, 6267–6275. [[CrossRef](#)]

25. Kloog, I.; Nordio, F.; Coull, B.A.; Schwartz, J. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM_{2.5} exposures in the mid-atlantic states. *Environ. Sci. Technol.* **2012**, *46*, 11913–11921. [[CrossRef](#)] [[PubMed](#)]
26. Lee, H.; Liu, Y.; Coull, B.; Schwartz, J.; Koutrakis, P. A novel calibration approach of modis aod data to predict PM_{2.5} concentrations. *Atmos. Chem. Phys.* **2011**, *11*, 7991. [[CrossRef](#)]
27. Ma, Z.; Hu, X.; Huang, L.; Bi, J.; Liu, Y. Estimating ground-level PM_{2.5} in China using satellite remote sensing. *Environ. Sci. Technol.* **2014**, *48*, 7436–7444. [[CrossRef](#)] [[PubMed](#)]
28. Hu, X.; Waller, L.; Lyapustin, A.; Wang, Y.; Liu, Y. 10-year spatial and temporal trends of PM_{2.5} concentrations in the southeastern us estimated using high-resolution satellite data. *Atmos. Chem. Phys.* **2014**, *14*, 6301–6314. [[CrossRef](#)] [[PubMed](#)]
29. Dewulf, B.; Neutens, T.; Lefebvre, W.; Seynaeve, G.; Vanpoucke, C.; Beckx, C.; Van de Weghe, N. Dynamic assessment of exposure to air pollution using mobile phone data. *Int. J. Health Geogr.* **2016**, *15*, 14. [[CrossRef](#)] [[PubMed](#)]
30. Gariazzo, C.; Pelliccioni, A.; Bolignano, A. A dynamic urban air pollution population exposure assessment study using model and population density data derived by mobile phone traffic. *Atmos. Environ.* **2016**, *131*, 289–300. [[CrossRef](#)]
31. Dewulf, B.; Neutens, T.; Van Dyck, D.; De Bourdeaudhuij, I.; Panis, L.I.; Beckx, C.; Van de Weghe, N. Dynamic assessment of inhaled air pollution using gps and accelerometer data. *J. Transp. Health* **2016**, *3*, 114–123. [[CrossRef](#)]
32. Lu, Y.; Fang, B.T. Examining personal air pollution exposure, intake, and health danger zone using time geography and 3D geovisualization. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 32–46. [[CrossRef](#)]
33. Erlander, S.; Stewart, N.F. *The Gravity Model in Transportation Analysis: Theory and Extensions*; VSP: Rancho Cordova, CA, USA, 1990; Volume 3.
34. Simini, F.; González, M.C.; Maritan, A.; Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **2012**, *484*, 96–100. [[CrossRef](#)] [[PubMed](#)]
35. Lee, R.; Sumiya, K. Measuring Geographical Regularities of Crowd Behaviors for Twitter-Based Geo-Social Event Detection. In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, San Jose, CA, USA, 2 November 2010; ACM: New York, NY, USA, 2010; pp. 1–10.
36. Stefanidis, A.; Crooks, A.; Radzikowski, J. Harvesting ambient geospatial information from social media feeds. *GeoJournal* **2013**, *78*, 319–338. [[CrossRef](#)]
37. Cheng, Z.; Caverlee, J.; Lee, K.; Sui, D.Z. Exploring millions of footprints in location sharing services. *ICWSM 2011, 2011*, 81–88.
38. Frias-Martinez, V.; Soto, V.; Hohwald, H.; Frias-Martinez, E. Characterizing Urban Landscapes Using Geolocated Tweets. In Proceedings of the Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom), Amsterdam, The Netherlands, 3–5 September 2012; pp. 239–248.
39. Preoțiuc-Pietro, D.; Cohn, T. Mining User Behaviours: A Study of Check-in Patterns in Location Based Social Networks. In Proceedings of the 5th Annual ACM Web Science Conference, Paris, France, 2–4 May 2013; ACM: New York, NY, USA, 2013; pp. 306–315.
40. Yu, H.; Russell, A.; Mulholland, J.; Huang, Z. Using cell phone location to assess misclassification errors in air pollution exposure estimation. *Environ. Pollut.* **2018**, *233*, 261–266. [[CrossRef](#)] [[PubMed](#)]
41. Kwan, M.-P. How GIS can help address the uncertain geographic context problem in social science research. *Ann. GIS* **2012**, *18*, 245–255. [[CrossRef](#)]
42. Tencent. *Annual Report*; Tencent: Shenzhen, China, 2016.
43. Wackernagel, H. *Multivariate Geostatistics: An Introduction with Applications*; Springer Science & Business Media: Berlin, Germany, 2013.
44. Brunson, C.; Fotheringham, A.S.; Charlton, M.E. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geogr. Anal.* **1996**, *28*, 281–298. [[CrossRef](#)]
45. Rodriguez, J.D.; Perez, A.; Lozano, J.A. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 569–575. [[CrossRef](#)] [[PubMed](#)]
46. Chafe, Z.A.; Brauer, M.; Klimont, Z.; Van Dingenen, R.; Mehta, S.; Rao, S.; Riahi, K.; Dentener, F.; Smith, K.R. Household cooking with solid fuels contributes to ambient PM_{2.5} air pollution and the burden of disease. *Environ. Health Perspect.* **2014**, *122*, 1314–1320. [[CrossRef](#)] [[PubMed](#)]

47. Gamble, J.F. PM_{2.5} and mortality in long-term prospective cohort studies: Cause-effect or statistical associations? *Environ. Health Perspect.* **1998**, *106*, 535–549. [[CrossRef](#)] [[PubMed](#)]
48. Gavett, S.H.; Koren, H.S. The role of particulate matter in exacerbation of atopic asthma. *Int. Archiv. Allergy Immunol.* **2001**, *124*, 109–112. [[CrossRef](#)] [[PubMed](#)]
49. Wong, J.Y.; De Vivo, I.; Lin, X.; Christiani, D.C. Cumulative PM_{2.5} exposure and telomere length in workers exposed to welding fumes. *J. Toxicol. Environ. Health Part A* **2014**, *77*, 441–455. [[CrossRef](#)] [[PubMed](#)]
50. Zhang, L.; Wang, F.; Ji, Y.; Jiao, J.; Zou, D.; Liu, L.; Shan, C.; Bai, Z.; Sun, Z. Phthalate esters (paes) in indoor PM₁₀/PM_{2.5} and human exposure to paes via inhalation of indoor air in Tianjin, China. *Atmos. Environ.* **2014**, *85*, 139–146. [[CrossRef](#)]
51. Adams, W.C. *Measurement of Breathing Rate and Volume in Routinely Performed Daily Activities: Final Report*; Contract NO. A033-205; University of California: Davis, CA, USA, 1993.
52. Marty, M.A.; Blaisdell, R.J.; Broadwin, R.; Hill, M.; Shimer, D.; Jenkins, M. Distribution of daily breathing rates for use in California's air toxics hot spots program risk assessments. *Hum. Ecol. Risk Assess.* **2002**, *8*, 1723–1737. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).