Unraveling environmental justice in ambient PM2.5 exposure in Beijing: A big data approach

Unraveling Environmental Justice in Ambient $PM_{2.5}$ Exposure in Beijing: A Big Data Approach

Yanyan Xu,[1,2,3] Shan Jiang,[4,5] Ruiqi Li,[6,7] Jiang Zhang,[6]
Jinhua Zhao,[8] Sofiane Abbar,[9] Marta C. González[1,2,3*]

[1]Department of Civil & Environmental Engineering, MIT, Cambridge, MA, 02139, USA
[2]Department of City and Regional Planning, University of California, Berkeley, CA, 94720, USA
[3]Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA
[4]Department of Urban and Environmental Policy and Planning, Tufts University, Medford, MA 02155, USA
[5]Department of Civil & Environmental Engineering, Tufts University, Medford, MA 02155, USA
[6]School of Systems Science, Beijing Normal University, Beijing 100875, China
[7]College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China
[8]Department of Urban Studies & Planning, MIT, Cambridge, MA 02139, USA
[9]Qatar Computing Research Institute, HBKU, Doha 5825, Qatar

*To whom correspondence should be addressed; E-mail: martag@berkeley.edu.

Air pollution imposes significant environmental and health risks worldwide and is expected to deteriorate in the coming decade as cities expand. Measuring population exposure to air pollution is crucial to quantifying risks to public health. In this work, we introduce a big data analytics framework to model residents' stay and commuters' travel exposure to outdoor $PM_{2.5}$ and evaluate their environmental justice, with Beijing as an example. Using mo-

bile phone and census data, we first infer travel demand of the population to derive residents' stay activities in each analysis zone, and then focus on commuters and estimate their travel routes with a traffic assignment model. Based on air quality observations from monitoring stations and a spatial interpolation model, we estimate the outdoor $PM_{2.5}$ concentrations at a 500-meter grid level and map them to road networks. We then estimate the travel exposure for each road segment by multiplying the $PM_{2.5}$ concentration and travel time spent on the road. By combining the estimated $PM_{2.5}$ exposure and housing price harnessed from online housing transaction platforms, we discover that in the winter, Beijing commuters with low wealth level are exposed to 13% more $PM_{2.5}$ per hour than those with high wealth level when staying at home, but exposed to less $PM_{2.5}$ by 5% when commuting the same distance (due to lighter traffic congestion in suburban areas). We also find that the residents from the southern suburbs of Beijing have both lower level of wealth and higher stay- and travel- exposure to $PM_{2.5}$, especially in the winter. These findings inform more equitable environmental mitigation policies for future sustainable development in Beijing. Finally, or the first time in the literature, we compare the results of exposure estimated from passive data with subjective measures of perceived air quality (PAQ) from a survey. The PAQ data was collected via a mobile-app. The comparison confirms consistencies in results and the advantages of the big data for air pollution exposure assessments.

Keywords— Environmental justice, $PM_{2.5}$ exposure, travel exposure, urban mobility, mobile phone data, Beijing

1 Introduction

With rapid urbanization and industrialization, air pollution has become a global threat to human health, especially for large and dense cities in developing countries (*Kampa and Castanas, 2008, Pope III et al., 2009, Lelieveld et al., 2015, World Health Organization, 2016, Kelly and Zhu, 2016*). According to the World Health Organization (WHO), around 3 million people died in 2012 as a result of ambient air pollution exposure, which makes it the largest environmental risk to the health of human beings worldwide. The particulate matter of a diameter of less than 2.5 μm ($PM_{2.5}$) is a major public concern in recent years, especially for cities suffering from severe hazes. Taking Beijing as an example, in 2015, the outdoor $PM_{2.5}$ concentration of 179 days were higher than 75 $\mu g/m^3$ and the average annual level reached 80.6 $\mu g/m^3$ (*Beijing Municipal Environmental Monitoring Center, 2016*). Such concentrations represent severe threats to public health, particularly to those vulnerable to heart or respiratory diseases, such as the young and elderly population (*Di et al., 2017*).

Due to these threats to health, quantifying human exposure to air pollutants has received considerable attention, and various data and methods have been introduced to estimate their exposure in space and time (*Jerrett et al., 2005, Steinle et al., 2013, Quiros et al., 2013, Levy et al., 2015, Smith et al., 2016, Nyhan et al., 2016, Dewulf et al., 2016, Shekarrizfard et al., 2017*). Many previous studies infer pollutants concentration from stationary air quality monitoring networks using spatial interpolation techniques and estimate air pollution exposure with spatial distributions of population at aggregated level (*Steinle et al., 2013*). This approach ignores human mobility and the time spent at various places in a day (*Setton et al., 2011, Zhang et al., 2013*). The development of GPS-enabled mobile monitors and location-aware instruments addresses this issue by measuring individual exposure during their activities (*Gerharz et al., 2009, Hatzopoulou et al., 2013, Quiros et al., 2013, Deville Cavellin et al., 2015*). Travel survey data were also employed to examine the dynamic exposure of participating interviewees (*Smith et al., 2016*). However, such datasets can only be used to estimate exposure for limited groups of individuals, and it is challenging to expand the samples to the population.

More recently, researchers have employed new data sources and methods to study the dynamic exposure to air pollutants at the city scale (*Saraswat et al., 2016, Nyhan et al., 2016, Dewulf et al., 2016*). For example, based on travel survey data, Shekarrizfard et al. developed an integrated transportation and emission model to assess the individual stay and travel exposure to NO_2 in Montreal, Canada (*Shekarrizfard et al., 2016*). Saraswat et al. simulated the travel demand of population in New Delhi with a Gravity model and estimated the human exposure for different activities (such as staying at home, work and commuting in the road network) (*Saraswat et al., 2016*). However, they neglected travel behavior and speed when estimating travelers' exposures.

With the development of information and communication technologies (ICT), large scale geo-located mobile phone data have been increasingly useful to model human mobility in cities (*Gonzalez et al., 2008, Deville et al., 2014, Jiang et al., 2016, Dong et al., 2016*). For instance, Nyhan et al. (2016) adopted mobile phone data to model the active population weighted exposure to $PM_{2.5}$ at aggregated spatial levels in the New York City (*Nyhan et al., 2016*), and quantified individual exposures to $PM_{2.5}$ by considering both home and work locations (*Nyhan et al., 2018*). Dewulf et al. (2016) used mobile phone data to track a user's visited locations and duration of stays to estimate dynamic exposure at individual level in Belgium (*Dewulf et al., 2016*). Although these works took advantages of the large scale mobile phone data, they only considered the samples without expanding them to the population level in the city (*Gauderman et al., 2004, Devlin et al., 2003*). By only focusing on the exposures of stationary activities, these studies also ignored exposures of human travel in cities. With the increased traffic congestion and long commuting distances in large cities, commuters have spent more time on the road than ever before (*TomTom, 2017*). As a result, people are exposed to non-negligible air pollution while traveling. Therefore, it is important to estimate the air pollution exposure on the road networks. More importantly, by estimating exposure to $PM_{2.5}$, it will offer opportunities for researchers to examine the environmental justice for the economically disadvantaged population (*Marmot, 2005*). The evaluation of environmental justice in cities will be useful to inform policymakers to develop equitable strategies for sustainable urban futures (*Brugge et al., 2015*).

Towards this end, this study presents a framework that incorporates urban mobility derived from massive and passive mobile phone data to evaluate the environmental justice of $PM_{2.5}$ exposure of commuters with different wealth levels in Beijing at the population level. First, we infer the seasonal average $PM_{2.5}$ concentration per hour (at a 500-meter grid level) based on air monitor observations and a spatial interpolation algorithm. We then map the $PM_{2.5}$ concentration to the road networks. Second, by using mobile phone data and census data together with a traffic assignment model, we derive human mobility (including individual stay location, duration, travel routes and travel time) at both individual and aggregated levels for the metropolitan area without expensive travel surveys (*Deville et al., 2014, Xu and González, 2017, Jiang et al., 2016, Jiang et al., 2017, Li et al., 2017*). Moreover, we estimate the residents' outdoor stay exposure to $PM_{2.5}$ by weighting population density, and model the travel exposure of commuters by accumulating the $PM_{2.5}$ exposure on each traversed road segment, taking travel time in traffic into account. Third, we investigate the environmental justice of residents by connecting human exposure to $PM_{2.5}$ with housing prices (a proxy for wealth) in Beijing, which is important for policymakers to develop equitable environmental mitigation policies for the city. Finally, by comparing our results with a mobile-phone based survey on individual perception of air quality, we assess the feasibility of using large-scale mobile phone data to measure human exposure to the air pollution in the city.

In the following sections, we will discuss in detail our methods to estimate $PM_{2.5}$ concentrations, urban mobility, population stay and travel exposure to $PM_{2.5}$ concentrations, and their relation to environmental justice in Beijing. We obtained data of the $PM_{2.5}$ concentrations in Beijing from 2015 to 2016. Ideally, we would like to collect data on human mobility, air quality, and wealth level of residents for the same period; however, such efforts have been challenging. In particular, world-wide urban mobility studies in the past relied mostly on expensive travel surveys collected by cities and government agencies in very low frequency to estimate travel demand for a given year, by adjusting samples to a current population. For example, the national household travel surveys in the U.S. are conducted every ten years. One of the advantages of this study is that by using more recent mobile phone data we can estimate more up-to-date travel demand models compared to those using traditional expensive travel surveys. However, due to data availability limitation, we could only obtain mobile phone data from 2013. With the assumption that the routine travel demand of Beijing residents in 2013 were similar to those in 2015 and 2016, the mobile phone data in 2013 can be a good proxy to estimate the stay and travel exposure of residents in 2015. We collected housing price data harnessed from an on-line platform in 2016 as a proxy for wealth distribution, assuming that correlation between wealth level and housing price is relatively stable in metropolitan areas such as Beijing.

2 Materials and methods

2.1 Estimating the spatial concentration of $PM_{2.5}$

As one of the most crowded cities in the world, Beijing accommodates 21.5 million residents in an area of $16,410$ km^2. Since 75% of Beijing residents live in the urban area within the Sixth Ring Road, a 20% of the total land area of the Beijing metropolitan area (Figure 1A), this study is focused in this area. Beijing Municipal Environmental Monitoring Center (BJMEMC) (*Beijing Municipal Environmental Monitoring Center, 2016*) collects concentrations of major air pollutants on an hourly basis from 35 air quality monitoring stations, among which 24 are located within the Sixth Ring Road (Figure 1A). The $PM_{2.5}$ concentration in Beijing displays strong seasonality, in both climate characteristics and economic activities such as coal heating in the winter (*Guo et al., 2014, Liu et al., 2016*). To derive the representative seasonal air quality data, we average the $PM_{2.5}$ concentration for each monitoring station during a given hour in the summer (from June 1st to August 31st, 2015) and winter (from December 1st, 2015 to February 28th, 2016), respectively. Figure 1B shows the average $PM_{2.5}$ concentration per hour for the 24 stations, and their respective average value in the summer and winter. The average $PM_{2.5}$ concentrations in the summer were stable and below $75\ \mu g/m^3$; while in the winter, the values were relatively higher at night than those during the day. The $PM_{2.5}$ concentrations in the winter at most monitoring stations were higher than $100\ \mu g/m^3$ and ranked as either unhealthy or very unhealthy based on standards defined by the U.S. Environmental Protection Agency (EPA) (*U.S. Environmental Protection Agency, 2017*).

Estimating the $PM_{2.5}$ spatial concentrations is the first step to quantify human exposure to air pollutants in the city. The Ordinary Kriging (OK) spatial interpolation method (*Wong et al., 2004, Zou et al., 2015*) and the land use regression (LUR) method are widely used for mapping the spatial $PM_{2.5}$ concentration with sparse monitoring data. LUR utilizes the geographic characteristics to refine the estimation of $PM_{2.5}$ concentration. Zou et al. evaluated the performance of LUR and OK interpolation in Houston, and revealed that LUR and OK generate similar end-result accuracy for their comparable error rates (6.13% and 7.01%, respectively) in the estimation of $PM_{2.5}$ concentration (*Zou et al., 2015*). By applying the OK method implemented with the python function of GEOMS2, an open-source geostatistics and geosciences modeling software (*CERENA Research Center, 2017*), we estimate the $PM_{2.5}$ concentration at a 500 meter grid level in sub-regions of the city based on the seasonal average values of the 24 monitoring stations per hour.

We predict the values at the target locations based on the distance and spatial distribution

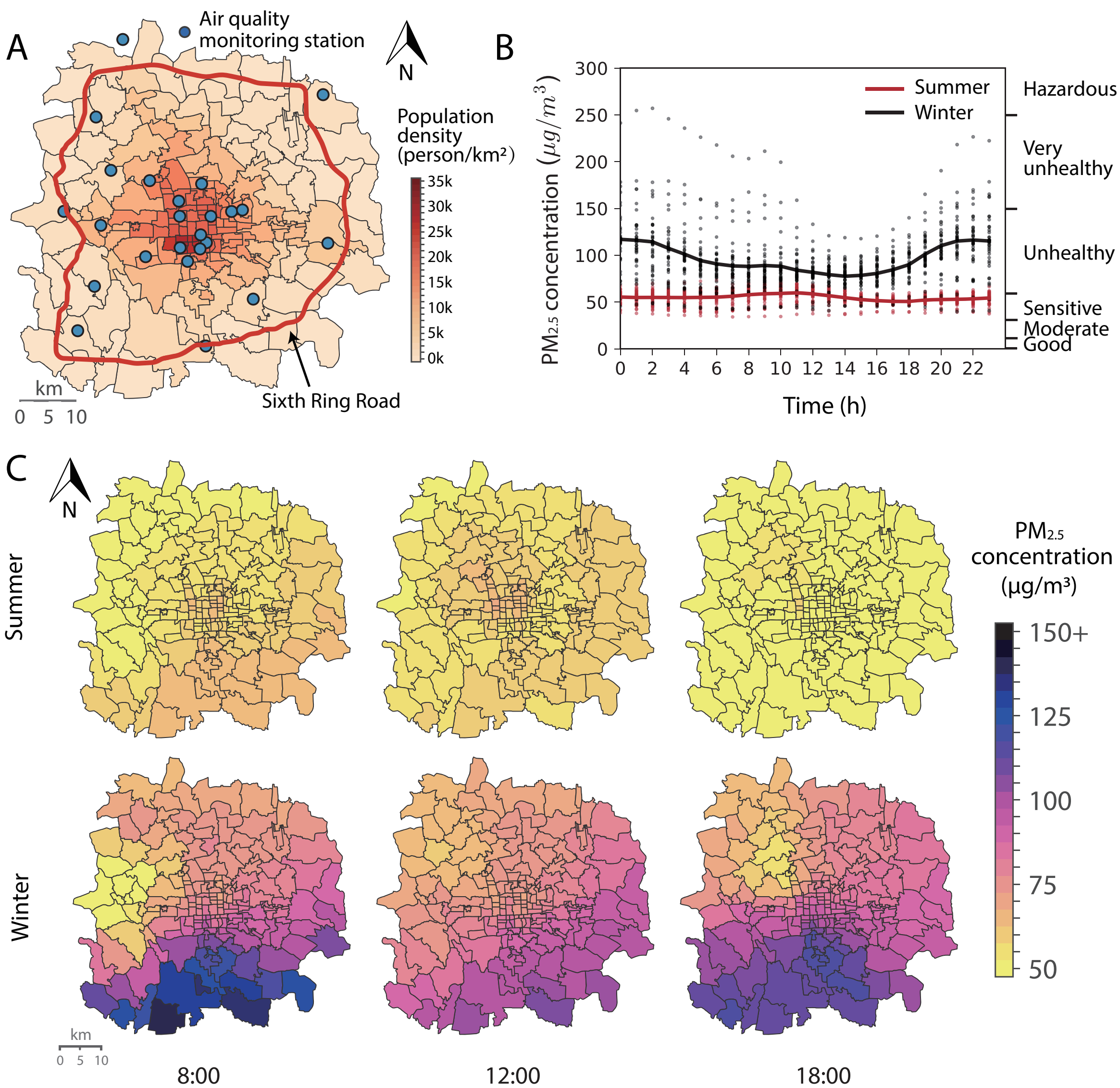

Figure 1: **Observed and estimated $PM_{2.5}$ concentration in Beijing.** **A** Population density in 2015 and the distribution of air quality monitoring stations within the Sixth Ring Road of Beijing. **B** Average hourly $PM_{2.5}$ concentration during the summer and winter of 2015. The scatter dots denote the concentration of the monitoring stations and the solid line shows the average concentration of all monitoring stations. **C** Estimated average $PM_{2.5}$ concentration in each zone at selected hours, 8:00, 12:00 and 18:00, in summer and winter. [Figure created with the Python library Matplotlib Basemap Toolkit]

of the target location using the OK method. Each of the (500 meter by 500 meter) grid is then assigned a $PM_{2.5}$ concentration per hour for the summer and winter. Figure 1C exhibits the estimated $PM_{2.5}$ concentration in each zone *Jiedao* (a community-level zone in China) during the morning peak hour, the midday off-peak hour, and the evening peak hour in both summer and winter. As shown in the figure, $PM_{2.5}$ in the winter were more concentrated in the southeast of Beijing, where industrial activities played an important role.

2.2 Inferring urban mobility from mobile phone data

The travel demand for the 16.3 million residents living in the urban area of Beijing is estimated using mobile phone data, provided by a mobile operator in China. The mobile phone dataset contains $100,000$ active users with their call detail records (CDRs) and data detail records (DDRs) for December 2013. The communication activities of CDRs comprise the incoming and outgoing phone call and the sending and receiving of a text message, while DDRs comprise the using of Internet data. Each record of the CDRs and DDRs data has a hashed ID, start time-stamp of the activity, type of activity, duration of the activity, longitude and latitude of the cell tower that communicated with the phone. The hashed ID is unique for each mobile phone devise, so that we are able to analyze the anonymized user when she is interacting with the phone which communicates with the nearest cell tower. In rare cases, the second nearest tower will be used if the nearest one is fully loaded. The cell phone in use will be switched to the closest cell tower when the user is moving. The average distance between cell towers is 332 meters (with a median of 254 meters), representing the spatial resolution of the study. Alexander *et al.* and Colak *et al.* outlined a general framework to obtain the travel demand, a.k.a. Origin-Destination (OD) matrices, from massive mobile phone data (*Alexander et al., 2015, Çolak et al., 2015*). We apply the same method here to extract daily trips from mobile phone samples in Beijing. These trips are then combined with the resident population of each zone within the Sixth Ring Road to estimate representative urban travel demand at the zonal level. Before modeling the travel behavior of each anonymous users, we first eliminate non active users. Then, we extract stay locations of the active users from their raw records. We improve upon the stay-point algorithm presented by Zheng and Xie (*Zheng and Xie, 2011*) and Jiang et al. (*Jiang et al., 2013*) as follows: (i) We apply a temporal agglomeration algorithm, such that temporally consecutive records within a certain radius (e.g., 500 meters) are bundled together with an updated stay duration from the start time of the first record to the end time of last one. (ii) We label the records as pass-by points and stays, based on the stay duration threshold (e.g.,10 minutes). In the analysis hereafter, we only focus on the stays. We then combine all the spatially adjacent stay points for a user (within a threshold) as his or her stay regions. For this spatial agglomeration, we use a spatial search balancing tree, R-tree, to accelerate the computation (see Figure S1 in Supplementary Material (SM) (*Guttman, 1984*)).

After the stay locations are detected for each user, the stays are labeled as *home*, *work*, or *other*. The most frequently visited location during weekday nights and weekends are labeled as *home*, and the most frequently visited one during weekday working hours (at least 500 meters away from home) is labeled as *work*, if one exists, and the rest are labeled as *other*. Each trip can be then labeled as one of the three categories, (i) home-based-work (HBW), which refers to the trips between home and work, a.k.a. commuting; (ii) home-based-other (HBO), which refers to the trips between home and other places; (iii) non-home-based (NHB), which refers to the trips between work and other places or two other places. Among the three categories of trips, commuting flow among zones are the most stable as the urban population and economic structure are relatively stable in the metropolitan area, such as Beijing. Eventually, as the mobile phone users only cover a part of the entire population, we expand the travel portrait of the mobile phone samples with expansion factor by zone. The expansion factor is defined as the ratio between the actual resident population and the number of mobile phone users whose *home* are located in the given zone. We aggregate the population data in 2015 at the 100 by 100 meter grid level obtained from WorldPop (*Tatem, 2017*) to the $Jiedao$ level. After aggregating the

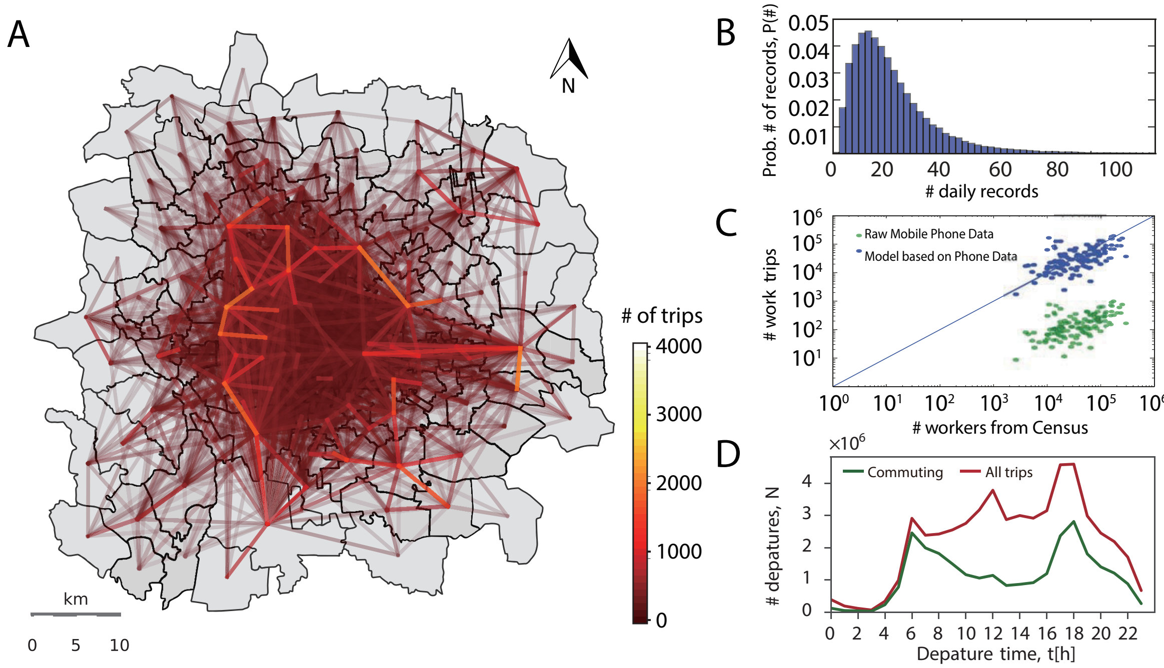

Figure 2: **Estimation and validation of urban mobility in Beijing. A** The top 5,000 origin-destination commuting flows among zones during the weekday morning peak hour. **B** Distribution of the number of daily records of each active user in the mobile phone dataset. Most users interact with the mobile phone for 10-30 times per day, including phone calls, text messages and Internet data. **C** Number of workers in each zone before and after expansion from users to the total population. The green circles show the comparison between recognized work trips from mobile phone data and the number of workers from census data in 2014. The blue circles show the number of work trips after expanding the mobile phone data to the whole population. The Pearson correlation coefficient $C_{pearson} = 0.745$ after expansion. **D** Number of trip departures for commuters and all population by time of the day.

trips at the zone level, we estimate the OD matrices by hour for an average weekday and 24 OD matrices were derived for the population and commuters, respectively.

Figure 2A shows the OD pairs with the large commuting flows between zones for the morning peak hour, obtained from the above discussed mobility model. Figure 2B shows the average number of phone usage records per day per user during a month. Majority of users are active, with an average of 15 daily records. As we observe the travel demand of individuals during one month, the records cover most places that they visited in their daily life, especially the work places for commuters. For the validation of estimated commuting travel demand, we compare with the census employment statistics at the *Jiedao* level in the Beijing 2^{nd} Economic Census in 2014. We find the Pearson correlation coefficient $C_{pearson} = 0.745$, which indicates a good agreement between our employment estimation and the survey data (see Figure 2C) (*Beijing Municipal Bureau Statistics, 2016*). The number of commuting and all trips per hour are shown in Figure 2D. The morning and evening peaks can be observed from the commuting trips, while there are three peaks for all trips during the morning, mid-day and evening on an average weekday. While a travel survey from Beijing is not available for this study, this method has been validated in many other cities with travel surveys, and traditional travel demand models developed by planning agencies (*Çolak et al., 2015, Toole et al., 2015, Alexander et al., 2015*). More details of the urban mobility model results can be found in SM Note 1.

As previous studies have shown that exposure to $PM_{2.5}$ can also be positively associated with increased psychological distress and affect human health (*Sass et al., 2017*), to compare the objective estimate with individual subjective perception of air quality, we collected a smartphone based survey from individuals in Beijing in this study. To our knowledge, this is the first survey in a Chinese city that is dedicated to capturing the perceived air quality by local residents. In the perceived air quality (PAQ) experiment, 860 individuals downloaded our smartphone-based survey application to track their daily trajectory, and 256 of them finished the survey by rating the PAQ for home, workplace, and worst spot during their commuting during the two week study period in the winter of 2015 (although some of them didn't complete the whole period). The perception runs from 0 to 5, with 0 being the best air quality and 5 being the worst. When comparing the daily $PM_{2.5}$ with the daily average PAQ for these respondents during the study period, we find that they are highly correlated with a Pearson correlation coefficient of 0.831. More detailed description of PAQ data can be found in SM Note 2.

2.3 Modeling stay and travel exposure to $PM_{2.5}$

With our inferred mobility at the urban population scale, we account the hourly dynamic stay of the population and the travel time and route choice of commuters assuming all residents starts their daily trips from home. We then estimate the stay exposure to $PM_{2.5}$ of population and the travel exposure to $PM_{2.5}$ of commuters per hour in the summer and winter, respectively. We define population density weighted exposure (PDWE) to represent the total outdoor exposure of population per square kilometer per hour. For a given zone z during an hour h, its PDWE is defined as $E_{h,z} = C_{h,z} \cdot P_{h,z}/S_z$, where $C_{h,z}$ denotes the $PM_{2.5}$ concentration during that hour, $P_{h,z}$ denotes the dynamic population staying in the zone during the same hour and S_z denotes the area of the zone. The unit of PDWE is $person \cdot \mu g \cdot m^{-3} \cdot km^{-2} \cdot h$. Unlike the population weighted exposure (PWE) introduced by Nyhan et al. (*Nyhan et al., 2016*), which is density-independent and tied to the total population of the zone, PDWE highlights the areas with higher population density and heavier air pollution. As shown in Figure 1A, the size of zones are widely different, which causes the total population of low density zones in suburban area is higher than that of high density zones in the central area. However, air pollution impose more serious threats to zones with high density.

The travel exposure to $PM_{2.5}$ of a commuter depends on her selected route, travel time, and $PM_{2.5}$ concentration on the road segments. In this work, we focus on commuters who travel between zones either with cars or buses as they are the main travel modes for long-distance above-ground commuting in Beijing. Based on the travel demand of commuters inferred from mobile phone data and census data, we use the travel mode share reported in the Travel Survey of Beijing Residents to estimate trips made by passenger cars and buses per hour (*Beijing Transportation Development & Research Center, 2007, Huang et al., 2008*). Regarding the traffic conditions in the road networks, we derived vehicle trips from the vehicle usage rate per origin zone. We then estimated the driving time per route by assigning the vehicle OD matrices to the road network using a traffic assignment model based on user equilibrium (UE). Further details are discussed in Ref. (*Çolak et al., 2015, Çolak et al., 2016*). The road network is extracted from OpenStreetMap (*OpenStreetMap, 2016*). The UE model gives each OD pair its shortest travel time and paths. We validated our estimates of travel times with Gaode Map (*AutoNavi, 2016*), a widely used travel navigation platform in China. The Pearson correlation coefficient between the travel times estimated by our model and Gaode Map is 0.84 and the relative accuracy is 79.17% by regarding Gaode Map as the ground truth. The comparison of the distribution of travel time indicates good estimates, shown in Figure S2B and .S2C in SM. For simplicity, we assume that buses use similar routes as cars, but with longer travel times. According to a travel survey report, the travel time for a trip made by bus equals to 1.57 times for a car trip on average (*Beijing Transportation Development & Research Center, 2007*).

With the traffic assignment model, each road segment in the road network is associated with a travel time with traffic. Further, we can estimate the $PM_{2.5}$ concentration on each road segment by mapping the grid map of $PM_{2.5}$ concentration to the roads. When a road segment covers more than one grid, the average concentration of all covered grids is regarded as the $PM_{2.5}$ concentration of the road segment (see Figure S3 in the SM). The estimated $PM_{2.5}$ concentration of the road network during the morning peak hour, mid-day, and the evening peak hour in summer and winter are shown in Figure S4 in the SM. Finally, we calculate the route travel exposure to $PM_{2.5}$ of a commuter by aggregating the road segment exposure, $E_h^T = \sum_{r=1}^{R} C_h^r T_h^r$, where C_h^r and T_h^r denote respectively the $PM_{2.5}$ concentration and travel time on the rth road segment in the route during the hth hour and R is the number of road segments forming the route. The unit of travel exposure is $\mu g \cdot m^{-3} \cdot h$.

3 Results

3.1 Population density weighted exposure

As most travel activities happen during the daytime (*e.g.*, work, school), we calculate the stay exposure to $PM_{2.5}$ in two periods, work hours and non-work hours. The former covers hours between 9:00 am to 5:00 pm; the latter covers the rest of the day. Figure 3A illustrates the average hourly PDWE during non-work and work hours in the summer and winter. By observing the spatial distribution of PDWE, we identify exposure by zones during non-work and work hours. The disparity of PDWE between work and non-work hour is mainly caused by the travel activities of residents, and the disparity between summer and winter is mainly caused by

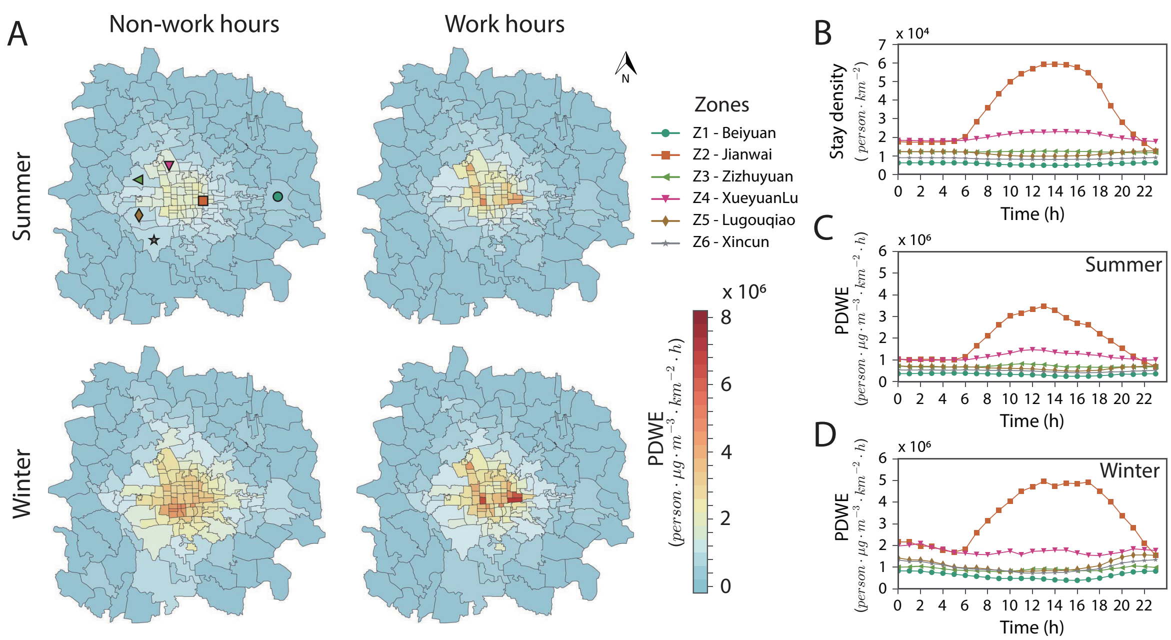

Figure 3: **Population density weighted exposure (PDWE) to $PM_{2.5}$ per zone. A** Average PDWE in each zone during work (from 9:00 am to 5:00 pm) and non-work hours during the summer and winter of 2015. The zones in downtown are exposed to heavier $PM_{2.5}$ concentration during work hours than non-work hours. **B** Hourly population density of six selected zones. The higher population density during daytime in Zone Z2 and Z4 captures the incoming flow during work hours. **C** Hourly PDWE of selected zones in the summer. **D** Hourly PDWE of selected zones in the winter.

the seasonal variations in $PM_{2.5}$ concentration in Beijing. Specifically, some central areas in Beijing experience more severe PDWE during work hours than non-work hours as a significant portion of the population is gathering into the central area of the city during daytime for work and/or business. The disparity of stay exposure in space and time is more evident in winter than in summer.

According to the urban mobility patterns, we select six representative zones in the city, marked with different symbols in Figure 3A, to uncover the PDWE. Figure 3B displays the population density of stays per hour in the selected zones (keeping the colors of the labeled zones). The population density at noon is three times that of midnight in zone Z2, which is located in the CBD of Beijing, the Chaoyang district. Figures 3C and 3D depicts the hourly PDWE of six selected zones during summer and winter, respectively. In the summer, the worst PDWE in zone Z2 reaches $3.5 \times 10^6 person \cdot \mu g \cdot m^{-3} \cdot km^{-2} \cdot h$. While in winter it reaches $5.0 \times 10^6 person \cdot \mu g \cdot m^{-3} \cdot km^{-2} \cdot h$.

3.2 Spatial variation of travel exposure to $PM_{2.5}$

The travel exposure captures the air pollution for each commuter between home and work by car or bus. By combining the estimated traffic flow, route, and travel times between each OD pairs with the $PM_{2.5}$ concentration of the road network, we can estimate the travel exposure to $PM_{2.5}$ between any two zones in any hour of the day. We select the trips between 8:00 am and 9:00 am, which reflects commuters' trips and the spatial differences in their travel exposure. The map in Figure 4A depicts the number of commuters that travel across zones during the morning peak hour. In Figure 4B we show the average travel distance of commuters living in each zone. As expected, suburban areas display longer commuting distance as most jobs are centralized in the city center. Figure 4C and 4D display the average travel exposure of commuters in each zone during the morning peak hour in the summer and winter, respectively. The commuting

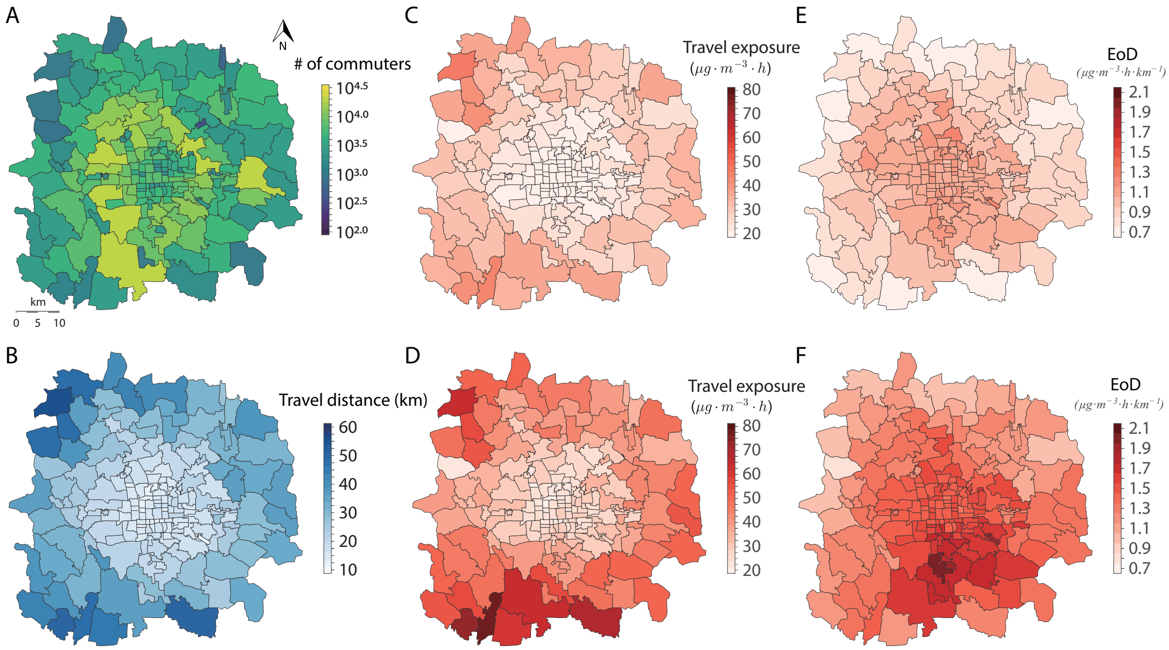

Figure 4: **Travel exposure to $PM_{2.5}$ of commuters from each zone during the morning peak hour in summer and winter.** **A** Number of commuter departures from each zone during the morning peak hour. **B** Average travel distance of commuters in each zone during the morning peak hour. **C, D** Average travel exposure to $PM_{2.5}$ of commuters during the morning peak hour in summer and winter, respectively. **E, F** Exposure-over-distance of commuters during the morning peak hour in summer and winter, respectively.

exposure in the summer and winter show some discrepancies, especially in the southern area where commuters experience higher travel exposure to $PM_{2.5}$ in the winter.

To better evaluate the spatial variation of travel exposure to $PM_{2.5}$, we define the travel exposure per kilometer in each zone as the ratio between the total travel exposure and the total travel distance for commuters in a given zone, namely exposure-over-distance ratio (EoD, $\mu g \cdot m^{-3} \cdot h \cdot km^{-1}$). EoD indicates the concentration of $PM_{2.5}$ exposures to the traveler per kilometer from the origin to the destination. Trips with larger EoD are exposed to more $PM_{2.5}$ than others even when they have the same commuting distance. Figure 4E and 4F illustrate the EoD per zone in summer and winter, respectively. In the summer, EoD displays higher values near the central area and lower values in the suburbs. This is caused by the heavy traffic congestion in the central area, as shown in Figure S2A in the SM. Regarding the distribution of EoD in winter, regions in the south show highest values due to the combined effect of both higher $PM_{2.5}$ concentration (shown in Figure 1C) caused by coal-burning plants in the south of Beijing and the heavy traffic congestion near the central area. More results are presented in Figure S5 in the SM and an on-line travel exposure visualization platform [1].

3.3 Environmental justice in $PM_{2.5}$ stay- and travel-Exposure

Environmental justice refers to "the fair treatment of all people with variant races, cultures and incomes, in development of regulations and policies." (*Marmot, 2005, Pearce et al., 2006, Brugge et al., 2015*) Here we investigate the environmental justice for commuters with different wealth levels, regarding their $PM_{2.5}$ exposure. We derive aggregated zonal housing price index from disaggregated housing price data, obtained from an online housing property listing platform in June 2016 (*Homelink, 2016*). We use housing price as a proxy for wealth level and examine its relationship with commuters' hourly stay-exposure during non-work hours (which are mostly stay-at- home activities). We then compare their travel-exposures (EoD) across space.

Figure 5A shows the average housing price in each zone, revealing higher housing value in the city center than in the suburbs. In each of the sub-figures from B1 to C2, the community zones in Beijing are separated into six groups by combing three levels of housing prices (i.e., low, middle and high) with two levels of $PM_{2.5}$ exposure (i.e., low and high). Figure 5B1 and 5B2 display the relationship between the housing prices and the hourly stay exposure of commuters during the non-work hours in the winter, with the assumption that commuters go to work in the morning and return home in the evening. We estimate that hourly stay exposure to $PM_{2.5}$ for commuters with low, middle and high wealth levels are 111.82 $\mu g \cdot m^{-3} \cdot h$,

[1] http://www.mit.edu/~yanyanxu/exposure/

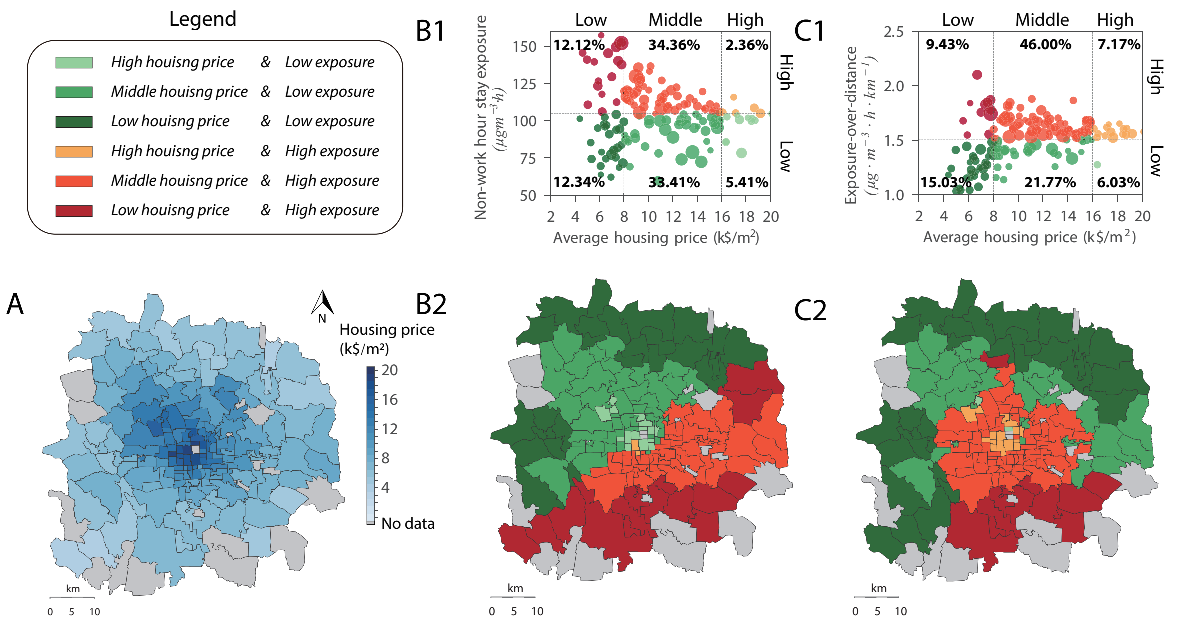

Figure 5: **Relation between PM$_{2.5}$ exposure and housing prices A** Average housing prices of each zone in thousand of US$ per square meter ($k\$ \cdot m^{-2}$) in June 2016. **B1, B2** Average housing price (of June 2016) and individual stay-exposure during non-work hours in the winter of 2015 for each zone. Darker color refers to lower housing price and the green color refers to lower exposure. We classify zones into six groups based on housing price and exposure levels. **C1, C2** Exposure-over-distance (EoD) for commuters during the morning peak hour v.s. the average hosing price in the same zone.

103.87 $\mu g \cdot m^{-3} \cdot h$ and 98.66 $\mu g \cdot m^{-3} \cdot h$ on average in winter of 2015. That is, commuters with lower wealth are exposed to more PM$_{2.5}$ than their counterparts with higher wealth by 13% per hour when they stay at home. Moreover, the different groups of zones display clear differences. For example, 12.12% of the commuters in Beijing having high stay-exposure at home are with low level of wealth. Most of these population live in the southern suburbs, depicted in dark red; another 12.34% of the commuters are with low level of wealth but the PM$_{2.5}$ concentration in their residential areas were lower as they live in the north of the city. In short, for commuters in the southeastern Beijing, those with lower wealth level experience higher level of PM$_{2.5}$ exposure than those with higher wealth levels. However, for commuters residing in the northwest of the city, those with lower wealth level are exposed to less PM$_{2.5}$.

This spatial disparity is mainly caused by the industrial and economic activity distribution in the city and can be mitigated by future spatial planning policy.

In contrast, the relationship between the housing price and the commuting travel exposure (EoD) displays large spatial disparity, as shown in Figure 5C1 and 5C2. The travel exposure to $PM_{2.5}$ per kilometer for commuters with low, middle and high wealth levels were 1.47 $\mu g \cdot m^{-3} \cdot h \cdot km^{-1}$, 1.56 $\mu g \cdot m^{-3} \cdot h \cdot km^{-1}$ and 1.55 $\mu g \cdot m^{-3} \cdot h \cdot km^{-1}$ on average in winter of 2015. This indicates that the commuters with lower wealth were exposed to 5% less $PM_{2.5}$ than commuters with higher wealth level when traveling the same distance for commuting trips. The primary reason is that the lower wealth residents are living in the suburban areas where traffic is less congested than the city center, as shown in Figure S2A in the SM. We also estimate that 9.43% of the commuters had both low wealth and high travel exposure per kilometer during their commuting trips. They were concentrated in the southern areas, colored in dark red. For commuters residing in southern Beijing, those with low wealth level were exposed to more $PM_{2.5}$ for both stay-at-home activities and travels in the winter. Moreover, 46% of the commuters have middle wealth level and experienced high travel exposure (EoD) to $PM_{2.5}$ (as shown in the in orange zones). Their high EoD is mainly caused by the heavy traffic congestion within the 5th Ring Road in the south of Beijing. The results from the summer are presented in Figure S6 in the SM.

3.4 Perceived air quality experiment as a comparison

To compare the spatial correlation of perceived air quality (PAQ) and the objective estimates of $PM_{2.5}$ exposure, we then model the population density weighted air quality perception (PDWPer) with PAQ data on the same day (February 17th, 2016) by replacing the $PM_{2.5}$ concentration in PDWE with the average perception in each zone, similar to the calculation of PDWE with mobile phone data discussed previously. PDWPer is calculated only using the perception of participants and the static population density from census data. Only 97 individual samples for PAQ were completed for this day (February 17th, 2016) among the 256 individuals who responded for the 2 week study. The PDWPer at home and work are illustrated in Figures 6A and 6B, respectively. As the sample size of the PAQ data is limited, the PDWPer does not cover all areas in Beijing. Figures 6C and 6D illustrate the PDWE of the worst hour during the non-work and work hours on the same day, respectively. Although the subjective perception of the participants does not contain the concentration of pollutants, we compare the exposures inferred from mobile phone data with PAQ data by normalizing both datasets. The r^2 between exposure from mobile phone data and PAQ equals to 0.43 at home and 0.33 at work, as shown in Figures 6E and 6F. The r^2 is relative low possibility due to the following two reasons: (i) the participants' perception to air quality might be different, as sensitive people are more vulnerable to the air pollutants; (ii) the limited number of reports per user per day loses the variation of air quality in time. In summary, given the small sample size of PAQ and the relative consistency between the two estimates from PAQ and mobile phone data, we show that when survey data are expensive to collect, combining air quality monitoring of $PM_{2.5}$ concentration and large-scale mobile phone data can be a good alternative to estimate exposure to air pollution without surveys.

4 Discussion and conclusion

Exposure to air pollution threatens public health, increasing mortality and morbidity. Within the same city, levels of exposure to air pollution differ in space and time. Among various pollution metrics, $PM_{2.5}$ concentration is the major concern for the public in Beijing as it is the main cause of haze and affects the heart and lungs when inhaled. On certain days, schools need to be closed and people are encouraged to stay at home to avoid exposure to severe haze.

Today, massive mobile phone data can help us better understand and simulate human mo-

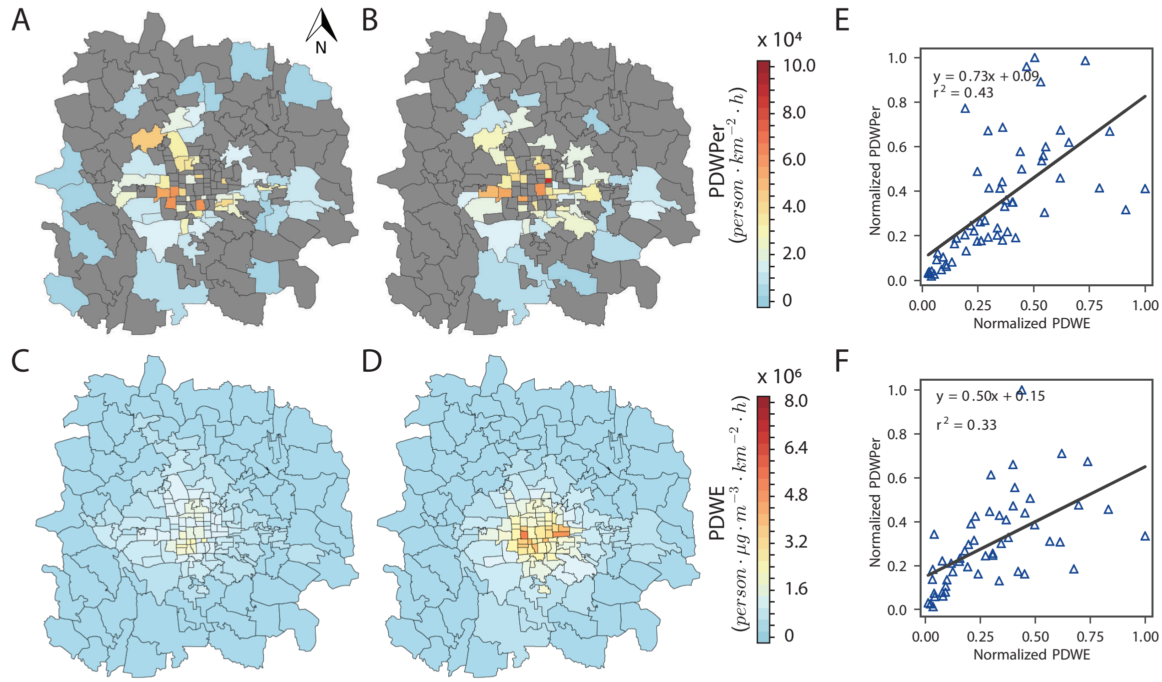

Figure 6: **Population exposure estimates from PAQ survey vs. mobile phone data. A, B** population density weighted air quality perception (PDWPer) at home and work per zone. The gray zones indicate that no participant lives or works in that zone. **C, D** population density weighted exposure (PDWE) during the worst non-work and work hour in each zone. **E, F** Normalized PDWPer and PDWE at home and work.

bility at the metropolitan scale. Still, previous works that model population exposure to $PM_{2.5}$ using mobile phone data only account for stays of the mobile phone users in each zone (*Nyhan et al., 2016*) or estimate exposure to air pollutants using the dynamic stay locations of sampled mobile phone users (*Dewulf et al., 2016, Nyhan et al., 2018*). The former study models the exposure of population in each zone at aggregated levels; the latter one focuses on the individual exposure and uses mobile phone data to track the sampled users. Both cases consider only the actual mobile phone users instead of the total population and ignore the individual travel exposure, which is nearly 10% of the total exposure if the resident spends 2 hours of travel per day.

By introducing the population census data, we expand the mobility of mobile phone users to the population at scale and incorporate their associated travel times. We estimate the $PM_{2.5}$ concentration in space and in the road networks. Without having to rely on costly travel surveys, we can estimate for the entire population their daily stay- and travel- exposure. In addition, we investigate environmental justice regarding the relationship between personal exposure to $PM_{2.5}$ and their level of wealth using housing price as a proxy. We find that commuters residing in southern Beijing are both economically disadvantaged and suffer higher static and travel exposure in the winter. This information is useful for policymakers to plan a more equitable and sustainable city. Mitigation policies may include subsidizing installation of air purifiers for low income population, regulating heavily polluting factories, and planning for urban greening projects focusing on PM2.5 control (*Brugge et al., 2015, Yang et al., 2015*). Finally, we compare exposure during stays from two diverse data sources, one passively collected via CDRs and one actively collected via a PAQ survey. We show that the $PM_{2.5}$ exposure modeled by mobile phone data is also supported by the PAQ survey. All of the results and plots in this work are implemented using Python.

The mobile phone data used for the inference of urban mobility only cover about 0.5% of the population in Beijing. Despite expanded with the actual population in census data, the small sample size might cause bias in the estimation of travel demand at urban scale. On the other hand, the low frequency of mobile phone usage may cause the loss of visited locations if the user doesn't interact with the cell phone in these places. However, these two shortcomings could be improved conveniently at low cost by expanding the sample of users and the duration of the datasets as the data of all mobile phone users have already been stored by telecommunication operators. Due to the lack of ground truth data (such as traffic counts or travel survey data), we are not able to directly validate our estimated travel demand in Beijing. The framework to generate the mobility model has been proved successful in other cities, e.g., we have validated the mobility model in Boston with the United States National Household Travel Survey (NHTS) and the Massachusetts Travel Survey (MTS) (*Alexander et al., 2015, Çolak et al., 2015*), in Bay Area with the Bay Area Transportation Survey (BATS) (*Çolak et al., 2016*). While mobile phone data is blind to the travel mode, in contrast with the Taxi GPS data and transit smart card data, it is still one of the best options to investigate the urban mobility in big cities, due to its high penetration rate. In addition, the location-based service (LBS) data collected from the mobile applications are also valuable resources to reproduce human mobility. Especially in China the high adoption rate of Wechat (the multi-purpose messaging, social media and mobile payment app) and its accurate positioning could better estimate both individual and aggregated mobility pattern of population. However, the LBS data are actively collected under the permission of users and thus lack of universality compared with the passive cellular data.

Future investigations in the following aspects would improve the estimation accuracy of population exposure to $PM_{2.5}$: (i) modeling the infiltration of vehicles and buildings would improve the estimation accuracy of personal exposure. The $PM_{2.5}$ concentration observed from the monitoring stations are adopted to model the population exposure without accounting for the ambient $PM_{2.5}$ infiltration; (ii) a chemistry-transport model would further improve the estimate of $PM_{2.5}$ concentration and other pollutants in the road networks with the consideration of land use and topography; (iii) an individual mobility model with higher resolution of data in the road segments would provide a more precise representation of personal exposure.

Acknowledgement

The authors thank Zheng Chang for his kind support, information and data provision. This work was jointly supported by the Samuel Tak Lee Faculty Seed Fund, MIT Environmental Solutions Initiative and the QCRI-CSAIL research grant. S. Jiang also thanks the support by the Knez Faculty Fellowship at Tufts University. R. Li was partially supported by NSFC under Grant Nos. 61773069 and 71731002, and also by the China Scholarship Council.

References

Alexander, L., Jiang, S., Murga, M., and González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation research part c: emerging technologies*, 58:240–250.

AutoNavi (2016). AMP direction API. `http://lbs.amap.com/api/webservice/reference/direction/`. [Online; accessed May-2016].

Beijing Municipal Bureau Statistics (2016). Beijing Regional Statistic Year Book. `http://www.bjstats.gov.cn/nj/qxnj/2014/zk/indexch.htm`. [Online; accessed January-2016].

Beijing Municipal Environmental Monitoring Center (2016). . `http://www.bjmemc.com.cn/`. [Online; accessed September-2016].

Beijing Transportation Development & Research Center (2007). The third beijing travel survey. `http://www.bjtrc.org.cn/`. [Online; accessed March-2017].

Brugge, D., Patton, A. P., Bob, A., Reisner, E., Lowe, L., Bright, O.-J. M., Durant, J. L., Newman, J., and Zamore, W. (2015). Developing community-level policy and practice to reduce traffic-related air pollution exposure. *Environmental Justice*, 8(3):95–104.

CERENA Research Center (2017). GEOMS2: geostatistics modeling software. `https://sourceforge.net/projects/geoms2/`.

Çolak, S., Alexander, L. P., Alvim, B. G., Mehndiratta, S. R., and González, M. C. (2015). Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. *Transportation Research Record: Journal of the Transportation Research Board*, (2526):126–135.

Çolak, S., Lima, A., and González, M. C. (2016). Understanding congested travel in urban areas. *Nature Communications*, 7.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., Blondel, V. D., and Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, 111(45):15888–15893.

Deville Cavellin, L., Weichenthal, S., Tack, R., Ragettli, M. S., Smargiassi, A., and Hatzopoulou, M. (2015). Investigating the use of portable air pollution sensors to capture the spatial variability of traffic-related air pollution. *Environmental Science & Technology*, 50(1):313–320.

Devlin, R., Ghio, A., Kehrl, H., Sanders, G., and Cascio, W. (2003). Elderly humans exposed to concentrated air pollution particles have decreased heart rate variability. *European Respiratory Journal*, 21(40 suppl):76s–80s.

Dewulf, B., Neutens, T., Lefebvre, W., Seynaeve, G., Vanpoucke, C., Beckx, C., and Van de Weghe, N. (2016). Dynamic assessment of exposure to air pollution using mobile phone data. *International Journal of Health Geographics*, 15(1):14.

Di, Q., Wang, Y., Zanobetti, A., Wang, Y., Koutrakis, P., Choirat, C., Dominici, F., and Schwartz, J. D. (2017). Air pollution and mortality in the medicare population. *New England Journal of Medicine*, 376(26):2513–2522.

Dong, L., Li, R., Zhang, J., and Di, Z. (2016). Population-weighted efficiency in transportation networks. *Scientific Reports*, 6:srep26377.

Gauderman, W. J., Avol, E., Gilliland, F., Vora, H., Thomas, D., Berhane, K., McConnell, R., Kuenzli, N., Lurmann, F., Rappaport, E., and et al. (2004). The effect of air pollution on lung development from 10 to 18 years of age. *New England Journal of Medicine*, 351(11):1057–1067.

Gerharz, L. E., Krüger, A., and Klemm, O. (2009). Applying indoor and outdoor modeling techniques to estimate individual exposure to pm2. 5 from personal gps profiles and diaries: a pilot study. *Science of the Total Environment*, 407(18):5184–5193.

Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782.

Guo, S., Hu, M., Zamora, M. L., Peng, J., Shang, D., Zheng, J., Du, Z., Wu, Z., Shao, M., Zeng, L., and et al. (2014). Elucidating severe urban haze formation in china. *Proceedings of the National Academy of Sciences*, 111(49):17373–17378.

Guttman, A. (1984). *R-trees: a dynamic index structure for spatial searching*, volume 14. ACM.

Hatzopoulou, M., Weichenthal, S., Dugum, H., Pickett, G., Miranda-Moreno, L., Kulka, R., Andersen, R., and Goldberg, M. (2013). The impact of traffic volume, composition, and road geometry on personal air pollution exposures among cyclists in montreal, canada. *Journal of Exposure Science and Environmental Epidemiology*, 23(1):46.

Homelink (2016). Lianjia Beijing Housing. `https://bj.lianjia.com`. [Online; accessed October-2016].

Huang, S.-s., Song, R., and Tao, Y. (2008). Behavior of urban residents travel mode choosing and influencing factors - taking beijing as an example. *Communications Standardization*, 9(181):1–5.

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., and Giovis, C. (2005). A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Science and Environmental Epidemiology*, 15(2):185–204.

Jiang, S., Ferreira, J., and González, M. C. (2017). Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore. *IEEE Transactions on Big Data*, 3(2):208–219.

Jiang, S., Fiore, G. A., Yang, Y., Ferreira Jr, J., Frazzoli, E., and González, M. C. (2013). A review of urban computing for mobile phone traces: current methods, challenges and opportunities. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, page 2. ACM.

Jiang, S., Yang, Y., Gupta, S., Veneziano, D., Athavale, S., and González, M. C. (2016). The timegeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences*, page 201524261.

Kampa, M. and Castanas, E. (2008). Human health effects of air pollution. *Environmental Pollution*, 151(2):362–367.

Kelly, F. J. and Zhu, T. (2016). Transport solutions for cleaner air. *Science*, 352(6288):934–936.

Lelieveld, J., Evans, J., Fnais, M., Giannadaki, D., and Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 525(7569):367–371.

Levy, I., Levin, N., Schwartz, J. D., and Kark, J. D. (2015). Back-extrapolating a land use regression model for estimating past exposures to traffic-related air pollution. *Environmental Science & Technology*, 49(6):3603–3610.

Li, R., Wang, W., and Di, Z. (2017). Effects of human dynamics on epidemic spreading in côte divoire. *Physica A: Statistical Mechanics and its Applications*, 467:30–40.

Liu, J., Mauzerall, D. L., Chen, Q., Zhang, Q., Song, Y., Peng, W., Klimont, Z., Qiu, X., Zhang, S., Hu, M., and et al. (2016). Air pollutant emissions from chinese households: A major and underappreciated ambient pollution source. *Proceedings of the National Academy of Sciences*, 113(28):7756–7761.

Marmot, M. (2005). Social determinants of health inequalities. *The lancet*, 365(9464):1099–1104.

Nyhan, M., Grauwin, S., Britter, R., Misstear, B., McNabola, A., Laden, F., Barrett, S. R. H., and Ratti, C. (2016). Exposure track - the impact of mobile-device-based mobility patterns on quantifying population exposure to air pollution. *Environmental Science & Technology*, 50(17):9671–9681.

Nyhan, M., Kloog, I., Britter, R., Ratti, C., and Koutrakis, P. (2018). Quantifying population exposure to air pollution using individual mobility patterns inferred from mobile phone data. *Journal of Exposure Science and Environmental Epidemiology*.

OpenStreetMap (2016). `https://www.openstreetmap.org`. [Online; accessed April-2016].

Pearce, J., Kingham, S., and Zawar-Reza, P. (2006). Every breath you take? environmental justice and air pollution in christchurch, new zealand. *Environment and Planning A*, 38(5):919–938.

Pope III, C. A., Ezzati, M., and Dockery, D. W. (2009). Fine-particulate air pollution and life expectancy in the united states. *New England Journal of Medicine*, (360):376–386.

Quiros, D. C., Lee, E. S., Wang, R., and Zhu, Y. (2013). Ultrafine particle exposures while walking, cycling, and driving along an urban residential roadway. *Atmospheric Environment*, 73:185–194.

Saraswat, A., Kandlikar, M., Brauer, M., and Srivastava, A. (2016). Pm2. 5 population exposure in new delhi using a probabilistic simulation framework. *Environmental Science & Technology*, 50(6):3174–3183.

Sass, V., Kravitz-Wirtz, N., Karceski, S., Hajat, A., Crowder, K., and Takeuchi, D. (2017). The effects of air pollution on individual psychological distress. *Health & place*, 48:72–79.

Setton, E., Marshall, J. D., Brauer, M., Lundquist, K. R., Hystad, P., Keller, P., and Cloutier-Fisher, D. (2011). The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates. *Journal of Exposure Science and Environmental Epidemiology*, 21(1):42.

Shekarrizfard, M., Faghih-Imani, A., and Hatzopoulou, M. (2016). An examination of population exposure to traffic related air pollution: Comparing spatially and temporally resolved estimates against long-term average exposures at the home location. *Environmental Research*, 147:435–444.

Shekarrizfard, M., Faghih-Imani, A., Tetreault, L.-F., Yasmin, S., Reynaud, F., Morency, P., Plante, C., Drouin, L., Smargiassi, A., Eluru, N., and et al. (2017). Regional assessment of exposure to traffic-related air pollution: Impacts of individual mobility and transit investment scenarios. *Sustainable Cities and Society*, 29:68–76.

Smith, J. D., Mitsakou, C., Kitwiroon, N., Barratt, B. M., Walton, H. A., Taylor, J. G., Anderson, H. R., Kelly, F. J., and Beevers, S. D. (2016). London hybrid exposure model: improving human exposure estimates to no_2 and $pm_{2.5}$ in an urban setting. *Environmental Science & Technology*, 50(21):11760–11768.

Steinle, S., Reis, S., and Sabel, C. E. (2013). Quantifying human exposure to air pollution-moving from static monitoring to spatio-temporally resolved personal exposure assessment. *Science of the Total Environment*, 443:184–193.

Tatem, A. J. (2017). Worldpop, open data for spatial demography. *Scientific Data*, 4:170004.

TomTom (2017). TomTom Traffic Index, Measuring congestion worldwide. `https://www.tomtom.com/en_gb/trafficindex/`. [Online; accessed January-2017].

Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., and González, M. C. (2015). The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, 58:162–177.

U.S. Environmental Protection Agency (2017). Air quality designations for particle pollution. `https://www.epa.gov/particle-pollution-designations`. [Online; accessed June-2017].

Wong, D. W., Yuan, L., and Perlin, S. A. (2004). Comparison of spatial interpolation methods for the estimation of air quality data. *Journal of Exposure Science and Environmental Epidemiology*, 14(5):404.

World Health Organization (2016). Ambient air pollution: a global assessment of exposure and burden of disease.

Xu, Y. and González, M. C. (2017). Collective benefits in traffic during mega events via the use of information technologies. *Journal of the Royal Society, Interface*, 14(129).

Yang, J., Chang, Y., and Yan, P. (2015). Ranking the suitability of common urban tree species for controlling pm2. 5 pollution. *Atmospheric Pollution Research*, 6(2):267–277.

Zhang, A., Qi, Q., Jiang, L., Zhou, F., and Wang, J. (2013). Population exposure to pm 2.5 in the urban area of beijing. *PloS One*, 8(5):e63486.

Zheng, Y. and Xie, X. (2011). Learning travel recommendations from user-generated gps traces. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1):2.

Zou, B., Luo, Y., Wan, N., Zheng, Z., Sternberg, T., and Liao, Y. (2015). Performance comparison of lur and ok in pm2. 5 concentration mapping: a multidimensional perspective. *Scientific Reports*, 5.