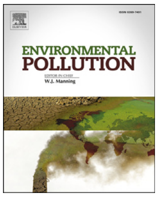
Environmental Pollution 240 (2018) 839–847

Contents lists available at ScienceDirect

Environmental Pollution

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



Using big data from air quality monitors to evaluate indoor $PM_{2.5}$ exposure in buildings: Case study in Beijing[☆]

JinXing Zuo [a, b], Wei Ji [c, 1], YuJie Ben [d], Muhammad Azher Hassan [a, b], WenHong Fan [a, b], Liam Bates [e], ZhaoMin Dong [a, b, *]

[a] School of Space and Environment, Beihang University, China
[b] Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, Beijing, 100191, China
[c] State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographical Sciences and Natural Resources Research, China
[d] School of Environmental Science and Engineering, South University of Science and Technology, Shenzhen, 518055, China
[e] Kaiterra Ltd., Beijing, China

ARTICLE INFO

Article history:
Received 22 February 2018
Received in revised form
10 May 2018
Accepted 10 May 2018

Keywords:
Indoor $PM_{2.5}$
Infiltration factor
Indoor/outdoor ratio
Beijing

ABSTRACT

Due to time- and expense- consuming of conventional indoor $PM_{2.5}$ (particulate matter with aerodynamic diameter of less than 2.5 μm) sampling, the sample size in previous studies was generally small, which leaded to high heterogeneity in indoor $PM_{2.5}$ exposure assessment. Based on 4403 indoor air monitors in Beijing, this study evaluated indoor $PM_{2.5}$ exposure from 15th March 2016 to 14th March 2017. Indoor $PM_{2.5}$ concentration in Beijing was estimated to be $38.6 \pm 18.4\ \mu g/m^3$. Specifically, the concentration in non-heating season was $34.9 \pm 15.8\ \mu g/m^3$, which was 24% lower than that in heating season ($46.1 \pm 21.2\ \mu g/m^3$). A significant correlation between indoor and ambient $PM_{2.5}$ ($p < 0.05$) was evident with an infiltration factor of 0.21, and the ambient $PM_{2.5}$ contributed approximately 52% and 42% to indoor $PM_{2.5}$ for non-heating and heating seasons, respectively. Meanwhile, the mean indoor/outdoor (I/O) ratio was estimated to be 0.73 ± 0.54. Finally, the adjusted $PM_{2.5}$ exposure level integrating the indoor and outdoor impact was calculated to be $46.8 \pm 27.4\ \mu g/m^3$, which was approximately 42% lower than estimation only relied on ambient $PM_{2.5}$ concentration. This study is the first attempt to employ big data from commercial air monitors to evaluate indoor $PM_{2.5}$ exposure and risk in Beijing, which may be instrumental to indoor $PM_{2.5}$ pollution control.

1. Introduction

Human exposure to Particulate Matter (PM), is a major public health concern. In 2015, more than 4 million premature deaths worldwide were attributed to ambient air pollution, which made it the fifth leading global public health risk factor for humans (Cohen et al., 2017). Mounting publications have indicated that smaller particles may result in stronger health hazards, since they can get deep into lungs, bloodstream and other organs. Thus, the research on exposure and risk raised by fine particles with a diameter of 2.5 μm ($PM_{2.5}$) or less is growing. In China, approximately 1.36 million deaths have been caused by $PM_{2.5}$ in 2010 (Lelieveld et al., 2015). To date, most risk estimation for $PM_{2.5}$ exposure has been conducted by using ambient $PM_{2.5}$ levels (Burnett et al., 2014). However, since most people spent more than 80% of their time indoors (Drew et al., 2010), calculation only riled on ambient $PM_{2.5}$ concentrations may yield mis-estimation bias on the exposure and risk raised by $PM_{2.5}$.

Mounting studies have addressed indoor $PM_{2.5}$ exposure in recent years (Shao et al., 2017; Wang et al., 2016). In 2015, Huang et al. reported indoor $PM_{2.5}$ level in Beijing's non-heating season was 4–193 $\mu g/m^3$ with a median of 34 $\mu g/m^3$ (n = 41), and indicated the indoor $PM_{2.5}$ level was significantly related to outdoor $PM_{2.5}$ level ($R^2 > 0.9$) (Huang et al., 2015). This indoor-outdoor correlation was confirmed by another study, which also pointed out $PM_{2.5}$ level lagged behind its outdoor counterpart with a 75–115 min delay (Han et al., 2015). Apart from outdoor $PM_{2.5}$, the impact of indoor human activities on indoor $PM_{2.5}$ was also determined. Emission rates from cooking (0.03 mg/min-2.78 mg/

[☆] This paper has been recommended for acceptance by Eddy Y. Zeng.

[*] Corresponding author. No.37, XueYuan Road, HaiDian District, Beijing, 100191, China.

E-mail address: dongzm@buaa.edu.cn (Z. Dong).

[1] These authors contributed to the paper equally.

https://doi.org/10.1016/j.envpol.2018.05.030
0269-7491/

min) and smoking (8–20 mg/cigarette, n = 8) were reported (Klepeis et al., 2003), while air purifier can also significantly decrease indoor $PM_{2.5}$ (He et al., 2004). However, conventional indoor $PM_{2.5}$ sampling is time- and expense-consuming since it involves household survey, which restricted sample size in most studies (Chen and Zhao, 2011). Considering the indoor $PM_{2.5}$ can be influenced by quite a number of factors (such as outdoor $PM_{2.5}$ levels, building types, ventilation, human activities and others), the limitation of small sample size was usually resulted in inconsistent conclusions amongst studies. For example, the indoor/outdoor (I/O) $PM_{2.5}$ ratio varied considerably with a wide range of 0.12–3.36 (Chen and Zhao, 2011). Also, the infiltration factor (the fraction of ambient $PM_{2.5}$ that can penetrate indoors and remain suspended) was only 0.35 (n = 114) based on random component superposition model (Meng et al., 2005), while the residential infiltration factor was reported 77% higher (0.62 ± 0.21) in the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA AIR) study (Allen et al., 2012). By regressing the indoor $PM_{2.5}$ and outdoor $PM_{2.5}$, the contribution of outdoor $PM_{2.5}$ to indoor $PM_{2.5}$ was estimated from 54% to 96% (n = 90) (Ji and Zhao, 2015). Therefore, the heterogeneity in indoor $PM_{2.5}$ exposure in large population is still high.

To address indoor $PM_{2.5}$ exposure in a large-scale context, scientists attempted to develop some mechanism models. For example, many factors (such as the influence of filtration intervention, energy efficiency, ventilation, stove use, presence and operability of exhaust fans) were included in the CONTAM toolbox, a multizone indoor air quality and ventilation analysis computer program (Walton and Dols, 2006). This toolbox was designed to determine the contaminant concentrations and personal exposure indoors. Based on CONTAM, Fabin et al. compared simulated $PM_{2.5}$ levels and their correlates with previous literature in Boston, which illustrated that CONTAM simulation could be readily explained by available parameters with an R^2 of 0.89–0.98 (Fabian et al., 2012). A similar study performed in London concluded the indoor $PM_{2.5}$ in detached or semi-detached properties (compared to flats and apartments) is more easily influenced by outdoor pollution infiltration, due to their greater externally exposed surface area-to-volume (Taylor et al., 2014). Nonetheless, these models require quite a number of parameters for various buildings, which are always unavailable (Milner et al., 2011). In addition, most model validations are conducted at low ambient $PM_{2.5}$ levels ($<10\ \mu g/m^3$). The model performance at high ambient $PM_{2.5}$ remains uncertain, which restricts model application in some highly polluted regions.

In recent years, given that concerns on $PM_{2.5}$ pollution are growing, people have increasingly purchased air monitors to measure the level of indoor pollution, and these monitoring data were commonly stored on a cloud server. The big data provides an avenue to reduce heterogeneity in indoor $PM_{2.5}$ exposure. Using Beijing as case study, this study aims to address indoor $PM_{2.5}$ exposure profile. Specifically, the objectives are to: 1) reveal spatial and temporal differences of indoor $PM_{2.5}$ concentrations; 2) estimate the I/O ratio, infiltration factor and the outdoor contribution to indoor $PM_{2.5}$; and 3) calculate $PM_{2.5}$ exposure and population risk. To the best of our knowledge, this study is the first attempt to integrate the data collected from thousands of air sensors into indoor $PM_{2.5}$ research, which may be informative to indoor $PM_{2.5}$ pollution control.

2. Materials and methods

2.1. Population study

Beijing is well recognized as the capital of China and located in the country's north-east region. The area of Beijing covers 16411.0 km^2. In this study, 6 major districts, namely DongCheng, XiCheng, HaiDian, ChaoYang, ShiJingShan and FengTai, were selected as target areas. The 6 major districts contribute approximately 59% of the population in Beijing, which approached 22 million in 2015.

2.2. Indoor $PM_{2.5}$ data

It should be noted indoor defined in this study specifically means indoor buildings. Actually, exposure to indoor PM includes also the exposure to polluted air inside various transportation modes, like cars (Grana et al., 2017), buses (Yan et al., 2015) and trains, where humans also spend significant time. However, to determine the exposure in these indoor transportation modes is not the scope of this study. The indoor $PM_{2.5}$ levels in present study were measured by Laser Egg, which is produced by Kaiterra company. As a real-time measurement equipment, the Laser Egg measures the $PM_{2.5}$ concentration by using the Laser-based light scattering technique. Briefly, the measurement principle is laser based Mie scattering, and a laser with 650 nm wavelength is used. Particles are pulled into the sensor by a fan, where they pass the laser beam. The scattered light is detected by a photodiode placed in a 90° angle to the beam. The peaks of the diffracted light are used to count the particles. By analysing the intensity of the scattered light, the particles' size and mass are estimated.

The products purchased by the residents lived in 6 urbans of Beijing were selected as the candidate air monitors. Furthermore, only air monitors with sufficient running time (average running time > 8 h per day) were utilized. It should be noted the location of each air monitor was identified by IP address, which was used to transfer the local $PM_{2.5}$ concentrations to the cloud server. The $PM_{2.5}$ levels measured from each sold equipment were recorded by the Kaiterra App, and uploaded to the cloud server.

To protect individual privacy, location information was encrypted by Geohash encode (Balkić et al., 2012). Briefly, the geographic location was encoded into a short string of letters and digits. This short string is a hierarchical spatial data structure which subdivides space into buckets of grid shape. The Geohash code was encrypted at 5 km^2 resolution in present study, and a thus 64 Geohash codes were generated as tabulated in Supplementary Material (SM) Table S1. As shown in Fig. 1, a total of 4403 air monitors were eligible to measure indoor $PM_{2.5}$ levels in the target areas. In particular, 2366 samplers were located in Chaoyang district, followed by 518 in HaiDian, 827 in DongCheng, 366 in XiCheng, 240 in ShiJingShan and 86 samplers in FengTai district.

2.3. Data quality

As a novel method to obtain indoor $PM_{2.5}$ data, the data quality is the first priority. During the period from 25th October, 2016 to 20th November 2016, four Laser Eggs (ID: 8ca9, 824b, 3c72 and 167e) were co-located with a Tapered Element Oscillating Microbalance (TEOM) at the Shanghai Qingpu Environmental monitoring station. During the co-location, data from the Laser Eggs and the TEOM were collected. Fig. S1 shows the course of the raw Laser Egg $PM_{2.5}$ mass concentration and the TEOM $PM_{2.5}$ mass concentration. Although some deviations can be observed, all four measurements follow the same trend.

Fig. S2 shows the correlation between the raw $PM_{2.5}$ mass concentration of the four co-located Laser Eggs and the TEOM monitor. The results demonstrated the Laser Egg could be readily measured the $PM_{2.5}$. However, the slope was estimated with a range of 1.02–1.25, which indicated that the Laser Eggs may overestimate $PM_{2.5}$ levels. The reason is at high relative humidities (RH), particles experience hygroscopic growth. If not controlled, optical sensors will overestimate the particles mass, as the particle