Estimating Long-Term Indoor $PM_{2.5}$ of Outdoor and Indoor Origin using Low-Cost Sensors

Tongling Xia [a], *Yue* Qi [b], *Xilei* Dai [c], *Ruoyu* You [d], *Dayi* Lai [e], *Junjie* Liu [c], *Chun* Chen [a,f]

[a] Department of Mechanical and Automation Engineering, The Chinese University, tlxia@mae.cuhk.edu.hk

[b] Shanghai Research Institute of Building Sciences (Group) Co., Ltd, Shanghai 201108, China, qiyue@sribs.com

[c] Tianjin Key Lab of Indoor Air Environmental Quality Control, School of Environmental Science and Engineering, Tianjin University, Tianjin, 300072, China, xileidai@tju.edu.cn, jjliu@tju.edu.cn

[d] Department of Building Services Engineering, The Hong Kong Polytechnic University, Kowloon, 999077, Hong Kong SAR, China, ruoyu.you@polyu.edu.hk

[e] School of Design, Shanghai Jiao Tong University, Shanghai 200240, China, dayi_lai@sjtu.edu.cn

[f] Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen 518057, China, chunchen@mae.cuhk.edu.hk

Abstract. To evaluate the separate impacts on human health and establish indoor control strategies, it is crucial to estimate the contribution of outdoor infiltration and indoor emission to indoor $PM_{2.5}$ in the built environment. This study applied an algorithm to automatically estimate the long-term time-resolved indoor $PM_{2.5}$ of outdoor and indoor origin in real apartments with natural ventilation. The inputs for the algorithm were only the time-resolved indoor/outdoor $PM_{2.5}$ concentrations and occupants' window actions, which were easily obtained from the low-cost sensors. This study first applied the algorithm in an apartment in Tianjin, China. The indoor/outdoor contribution to the gross indoor exposure and time-resolved infiltration factor were automatically estimated using the algorithm. Due to the year-round monitoring, the probabilistic distribution of the time-resolved $PM_{2.5}$ infiltration factor and indoor $PM_{2.5}$ emission can be given over a year. The influence of outdoor $PM_{2.5}$ data source on the estimated results was compared using the data from the low-cost light-scattering sensor and official monitoring station. Besides, the sensitive parameters to the algorithm were analyzed and their effects on the indoor emission contribution and estimated infiltration factor were investigated. Through the analysis, this study identified the practical applications that robust long-term outdoor $PM_{2.5}$ monitoring for a specific building can use the data from nearest official monitoring station. This study demonstrated an algorithm for estimating long-term time-resolved indoor $PM_{2.5}$ of outdoor and indoor origin in real naturally ventilated apartments with only the time-resolved indoor/outdoor $PM_{2.5}$ concentrations and window behaviors.

Keywords. Indoor $PM_{2.5}$ exposure, indoor emission, real building monitoring, I/O ratio, year-round distribution, natural ventilation
DOI: https://doi.org/10.34641/clima.2022.367

1. Introduction

Exposure to particulate air pollution poses one of the greatest risks to human health around the world [1]. In recent decades, $PM_{2.5}$ (particulate matter with a diameter less than 2.5 μm) has been proven to have a strong association with various diseases, on the basis of a large amount of epidemiological data [2-4]. Given that people spend a significant fraction of their time in indoor environments [5], it is essential to reduce indoor exposure to $PM_{2.5}$. Many studies have used the outdoor $PM_{2.5}$ concentration as an indicator to estimate the indoor exposure to $PM_{2.5}$ [6-7]. However, even when indoor $PM_{2.5}$ originates outdoors, the concentration of outdoor $PM_{2.5}$ is not a suitable indicator, because building-specific parameters such as air tightness and window-opening behavior would also influence the exposure [8-9]. Moreover, the existence of indoor $PM_{2.5}$ emissions, such as those from cooking and smoking, would further differentiate ambient $PM_{2.5}$ and indoor $PM_{2.5}$. Therefore, it is crucial to estimate the contribution of outdoor infiltration and indoor emissions to indoor $PM_{2.5}$ for evaluating the separate risk effects on human health.

With the rapid development of low-cost light-scattering $PM_{2.5}$ sensors, it is now straightforward to monitor time-resolved outdoor and indoor $PM_{2.5}$ concentrations. Our previous study proposed an algorithm that automatically differentiates the indoor $PM_{2.5}$ of outdoor and indoor origin using time-resolved indoor-to-outdoor $PM_{2.5}$ concentration ratio and window status. The method was validated in a small-scale chamber in a laboratory with a low relative error of 0.32% [10]. However, its performance in real buildings is still unclear. Therefore, to facilitate practical applications, it is worthwhile to assess the performance of the method in differentiating indoor $PM_{2.5}$ of outdoor and indoor origin in real buildings.

This study aimed to differentiate the indoor $PM_{2.5}$ of outdoor and indoor origin in real apartments with natural ventilation to demonstrate the robustness of the differentiation method proposed in our previous study [10]. Three inputs, the time-resolved concentrations of outdoor $PM_{2.5}$ and indoor $PM_{2.5}$ and window/door status, were monitored for a one-year period in 2017. The concentrations of indoor and outdoor $PM_{2.5}$ were monitored by a low-cost light-scattering $PM_{2.5}$ sensor, while official data from the national monitoring station near the target building were also obtained as alternative for the concentrations of outdoor $PM_{2.5}$. The occupants' window and door action was also monitored using low-cost sensors. Based on the differentiation method, the time-resolved indoor $PM_{2.5}$ of outdoor and indoor origin and their contributions to the total indoor exposure were estimated. The time-resolved infiltration factors were also obtained.

2. Research Methods

2.1 Original data

This study first focused on a naturally ventilated apartment located in Tianjin, China. The apartment was on the 16th floor of an 18-floor residential building. The indoor and outdoor $PM_{2.5}$ concentrations were recorded in the living room and the neighborhood, respectively, from January to December of 2017 using two low-cost light-scattering sensors with a time resolution of 1 min. In addition, the window/door-opening/closing actions were monitored with window/door sensors. The details of the monitoring setup can be found in a previous study [11]. However, it was found that a significant amount of data were missing from the outdoor low-cost $PM_{2.5}$ sensor due to bad weather and unstable power supply. Therefore, as an alternative, the outdoor $PM_{2.5}$ concentrations with a time resolution of 2 h were also obtained from the nearest official monitoring station, Binshui West Road station, operated by the China National Environmental Monitoring Center.

2.2 Data pre-processing

The low-cost light-scattering sensor for indoor $PM_{2.5}$ monitoring was first calibrated by a standard gravimetric instrument under a controlled environment. Since previous studies found that an increase in relative humidity can result in an increase in $PM_{2.5}$ concentration as measured by a light-scattering sensor, the indoor $PM_{2.5}$ concentrations were further calibrated based on the relative humidity [11].

Note that the original indoor $PM_{2.5}$ data were recorded once every minute. However, our previous study found that a time step size smaller than 10 minutes would result in significant errors in the differentiation algorithm [10]. Therefore, after the calibration, the concentrations of indoor $PM_{2.5}$ were averaged every 10 minutes, so that the time step size was in line with that of the differentiation algorithm. The sensitivity analysis about the time step size will be discussed in Section 3.4. In general, the low-cost sensor used indoors was stable. Over 94% of the indoor $PM_{2.5}$ data were successfully recorded throughout the year. Due to the relatively harsh environment, around 60% of the outdoor $PM_{2.5}$ data from the low-cost sensor were missing. Only were the data in February to April relatively complete for the analysis. As an alternative, this study also obtained the outdoor $PM_{2.5}$ data recorded once every two hours from the official monitoring station. Linear interpolation was used to convert the official monitoring outdoor $PM_{2.5}$ data to that with a time step size of 10 minutes. This study first used both outdoor $PM_{2.5}$ data from the low-cost sensor and the official monitoring station to estimate the indoor $PM_{2.5}$ of outdoor and indoor origin in February to April, and discussed the differences. We then used the outdoor $PM_{2.5}$ data from the official monitoring station to calculate the year-round indoor $PM_{2.5}$ of outdoor and indoor origin to demonstrate of the proposed algorithm.

Since the time step size was set at 10 minutes, this study counted effective window behavior with a time interval longer than 10 minutes from the previous and subsequent actions.

Besides, any time periods with mechanical ventilation or air cleaners turned on were removed since the algorithm is applicable for natural ventilation. Furthermore, this study applied the differentiation algorithm to each day individually. The results would be unsatisfactory for periods shorter than four hours [10]. Thus, this study removed the data that were recorded in any period shorter than four hours. Furthermore, if the results from the differentiation algorithm indicated that indoor $PM_{2.5}$ emissions occurred continuously throughout a whole day, there was no way to estimate the indoor $PM_{2.5}$ of outdoor and indoor origin, then such days were also removed. With these considerations, there were 40 days with valid input data for the analysis using the outdoor $PM_{2.5}$ data measured by the low-cost sensor in February to April. For the year-round estimation using the official monitoring outdoor $PM_{2.5}$ data, there were 275 days with valid inputs in this study.

2.3. Differentiation algorithm

After the data pre-processing, the three inputs, concentrations of outdoor $PM_{2.5}$ and indoor $PM_{2.5}$ and window action, were used with the differentiation algorithm to estimate indoor $PM_{2.5}$ of outdoor and indoor origin. This sub-section briefly describes the differentiation algorithm developed in our previous study [10].

Step 1 is to obtain indoor-to-outdoor $PM_{2.5}$ ratio. To consider the change in concentrations of outdoor and indoor $PM_{2.5}$ simultaneously, this study utilized the indoor-to-outdoor ratio time-resolved $PM_{2.5}$ concentration (I/O ratio), IO(t), to start the differentiation method:

$$IO(t) = \frac{C_{in}(t)}{C_{out}(t)} \tag{1}$$

where $C_{in}(t)$ and $C_{out}(t)$ ($\mu g/m^3$) are the averaged indoor and outdoor $PM_{2.5}$ concentrations, respectively, in the t-th time step. The size of time step was set at 10 minutes [10].

Step 2 is to process the change-point analysis. Normally, indoor $PM_{2.5}$ emissions can affect indoor $PM_{2.5}$ concentrations. The method of change-point analysis was used to detect significant changes in the time-series I/O ratios statistically [12]. The method also provided the confidence level for each change point. The details of the change-point analysis method can be found in [10]. Note that except for indoor emission of $PM_{2.5}$, window actions and fluctuations in infiltration rate can make a difference in the time-series I/O ratios and result in the change points as well. Therefore, this step was taken mainly to identify candidates for change points due to indoor $PM_{2.5}$ emissions.

Step 3 is to handle time periods no window status change. For the periods without window actions, significant increases in the I/O ratio were ascribed to either the indoor emission of $PM_{2.5}$ or the change of infiltration rate. To differentiate these two scenarios, three criteria were employed. First, when the I/O ratio was greater than 1, the period must have had an indoor emission of $PM_{2.5}$. Second, if the outlier was more than 1.5 interquartile over the third quartile in the I/O ratio, the period was considered to contain an indoor emission. Third, a detected change might arise from either a sudden increase in infiltration rate or an indoor $PM_{2.5}$ emission. Sudden increase in infiltration rate would increase the indoor concentration smoothly, while an indoor emission would increase the concentration with relatively strong fluctuations [10]. In a large-scale simulation, Shi et al. [9] obtained the infiltration rate distribution in Beijing residences with a 15^{th} percentile of 0.09 h^{-1} and an 85^{th} percentile of 0.32 h^{-1}. Considering a relatively extreme case in which the infiltration rate suddenly increased from 0.09 to 0.32 h^{-1}, the infiltration factor would increase by 0.22 (from 0.40 to 0.62), assuming the penetration factor and deposition rate to be 0.8 and 0.09 h^{-1}, respectively. Namely, if a detected change is caused by a sudden increase in infiltration rate, it is unlikely that the I/O ratio would increase by 0.22. Therefore, when the difference between the maximum and minimum values of the I/O ratio in the period was over an empirical threshold of 0.22, the period was regarded as containing an indoor $PM_{2.5}$ emission.

Step 4 is to handle the time periods having window status change. For the periods with window behavior, the I/O ratio would follow an exponential regression deducted from the mass balance equation without indoor particle emission [13]:

$$IO(t) = c_1 + c_2 \cdot e^{-c_3(t-t_0)} \tag{2}$$

where c_1 (unitless), c_2 (unitless), and c_3 (h-1) are constants as a function of the air exchange rate, $PM_{2.5}$ deposition rate and penetration factor. If the time-series I/O ratio fitted very well with equation (2), it is likely that there was no indoor source. Therefore, the R^2 value of the regression was used to determine the existence of indoor emission of $PM_{2.5}$. If the data fitting yielded a satisfactory R^2 value above 0.8, an empirical value according to Xia and Chen [10], then the period was regarded as free of indoor-generated $PM_{2.5}$. Otherwise, there existed indoor emission of $PM_{2.5}$.

Step 5 is to estimate the indoor $PM_{2.5}$ of outdoor and indoor origin. For the periods without indoor $PM_{2.5}$ emissions, the infiltration factor, F_{in}, was equal to the I/O ratio. For the periods with indoor $PM_{2.5}$ emissions, the infiltration factor was estimated with the use of equation (2), as demonstrated by Xia and Chen [10]. The indoor $PM_{2.5}$ of outdoor origin, $C_{in,out}$, in the periods with indoor $PM_{2.5}$ emissions was then calculated by:

$$C_{in,out}(t) = F_{in}(t) \cdot C_{out}(t) \tag{3}$$

The indoor $PM_{2.5}$ of indoor origin, $C_{in,in}$, can be expressed as:

$$C_{in,in}(t) = C_{in}(t) - C_{in,out}(t) \tag{4}$$

To compare the contributions of outdoor infiltrated $PM_{2.5}$ and indoor emitted $PM_{2.5}$, we calculated the ratio of indoor exposure to $PM_{2.5}$ of outdoor and indoor origin, respectively, to the gross indoor exposure for each day, denoted as the "indoor contribution" and "outdoor contribution", respectively, as follows:

$$\frac{E_{in,in}}{E_{in}} = \frac{\int_{t_{start,d}}^{t_{end,d}} C_{in,in}(t)dt}{\int_{t_{start,d}}^{t_{end,d}} C_{in}(t)dt} \tag{5}$$

$$\frac{E_{in,out}}{E_{in}} = \frac{\int_{t_{start,d}}^{t_{end,d}} C_{in,out}(t)dt}{\int_{t_{start,d}}^{t_{end,d}} C_{in}(t)dt} \tag{6}$$

where $E_{in,out}$ and $E_{in,in}$ ((μg·10min)/m³) are the daily indoor exposure to $PM_{2.5}$ of outdoor and indoor origin, respectively, and Ein ((μg·10min)/m³) is the daily gross indoor exposure to $PM_{2.5}$. The start time, $t_{start,d}$, and end time, $t_{end,d}$, of the daily gross indoor exposure are the start and end of an effective day. Here it was assumed that the occupant stayed indoors all the time. For different occupancy schedules, the corresponding exposures can be calculated accordingly based on the estimated concentrations of indoor $PM_{2.5}$ of outdoor and indoor origin.

Note that the indoor $PM_{2.5}$ sensor was placed in the living room. The indoor $PM_{2.5}$ emissions that occurred in other rooms, e.g., the kitchen and bedroom, may have contributed to the $PM_{2.5}$ concentration in the living room. In the differentiation algorithm, although the non-living-room $PM_{2.5}$ emissions were also detected, the estimated emission strength was equivalent to the portion that actually influenced the living room. In other words, these $PM_{2.5}$ emissions were also regarded as indoor sources that were located in the living room.

3. Results and discussion

3.1. Examples of estimated indoor $PM_{2.5}$ of outdoor and indoor origin

Fig. 1 illustrates the estimated time-resolved concentrations of indoor $PM_{2.5}$ of outdoor and indoor origin and the infiltration factor on Feb 21, Feb 14, and Mar 19. The inputs, outdoor and indoor $PM_{2.5}$ concentrations, are also shown in the figure. The area under the estimated indoor $PM_{2.5}$ of outdoor origin line (in orange) represents the daily indoor exposure to $PM_{2.5}$ of outdoor origin ($E_{in,out}$). The area under the indoor $PM_{2.5}$ line (in green) represents the daily gross indoor exposure to $PM_{2.5}$ (E_{in}). The total area of the four purple shaded zones represents the daily indoor exposure to $PM_{2.5}$ of indoor origin ($E_{in,in}$). equations (5) and (6) were used to calculate the daily indoor and outdoor contribution to the total indoor exposure, respectively. On Feb 21, four indoor emission events were detected with the differentiation algorithm, as shown in Fig. 1(a). The latter three detected $PM_{2.5}$ emissions were likely attributed to cooking considering the normal time periods for preparing breakfast, lunch, and dinner. In this apartment, the occupants often prepare late-night snacks. Therefore, the first emission might be from a late-night cooking activity. Based on the algorithm, the daily indoor and outdoor contribution was 32.5% and 67.5%, respectively. However, it should be noted that the first detected emission was weak, which might be a misclassification. If this weak emission was not considered as a real emission, the daily indoor and outdoor contribution would be altered by only 1.7%. Namely, the detected small emission did not alter the results of indoor/outdoor contribution in a major way. By characterizing the indoor $PM_{2.5}$ emission, the differentiation algorithm can then calculate the time-resolved infiltration factor ($F_{in}(t)$). As shown in Fig. 1(a), the real-time infiltration factor fluctuated in a wide range from 0.16 to 0.51. The daily averaged infiltration factor was 0.34±0.09. Similar results can be found on Feb 14, as shown in Fig. 1(b). The indoor emission events were also likely from cooking activities for late-night snack, breakfast, lunch, and dinner. After breakfast and lunch, there might be cleaning or other activities leading to emissions. The estimated daily indoor and outdoor contribution was 19.1% and 80.9%, respectively, and the averaged infiltration factor was 0.23±0.08. On Mar 19, the algorithm only detected one indoor emission event, which was likely from preparing the lunch. Since Mar 19 was a weekend, the occupants might get up late and skip the breakfast, and have their dinner in a restaurant. The estimated daily indoor and outdoor contribution was 3.0% and 97.0%, respectively, and the averaged infiltration factor was 0.45±0.11. The plausible explanation for the detected indoor emissions in these examples can partially support the feasibility of the algorithm.

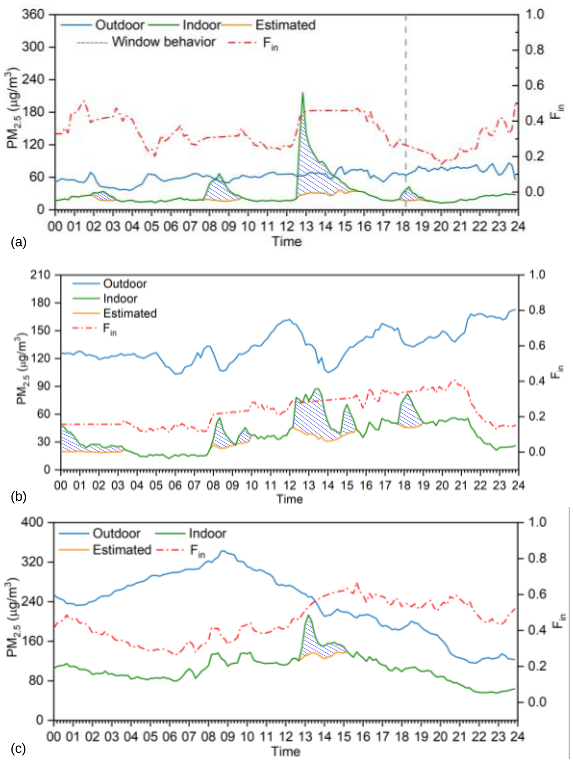

Fig. 1 - Estimated concentrations of indoor $PM_{2.5}$ of outdoor and indoor origin and infiltration factor on (a) Feb 21, (b) Feb 14, and (c) Mar 19 using the differentiation algorithm. (The purple shading represents indoor exposure to $PM_{2.5}$ of indoor origin.)

3.2. Comparison of results based on outdoor low-cost sensor and official monitoring station

As discussed, due to bad weather and unstable power supply, the outdoor $PM_{2.5}$ data from the low-cost light-scattering sensor were only available in February to April. The alternative was the outdoor $PM_{2.5}$ data recorded from the nearest official monitoring station. The mean ± standard deviation of the outdoor $PM_{2.5}$ data measured by the light-scattering sensor in February to April (68.9 ± 71.4 μg/m^3) was close to that measured by the official monitoring station (69.4 ± 60.7 μg/m^3). This study first compared the results based on the 1-min outdoor low-cost sensor and 2-h official monitoring station in February to April. Both datasets were averaged or interpolated to a 10-min resolution. Fig. 2 compares the probabilistic distribution of the daily indoor/outdoor contribution estimated based on the two outdoor $PM_{2.5}$ datasets. The general distributions were similar, but discrepancies can also be observed. It was estimated that, on average, the indoor $PM_{2.5}$ emissions and outdoor $PM_{2.5}$ infiltration contributed 23.2% and 76.8% of the daily total indoor exposure, respectively, if the outdoor $PM_{2.5}$ data from the low-cost sensor were used. When using the data from the official monitoring station, the average indoor and outdoor contribution was estimated to be 17.8% and 82.2%, respectively. Interestingly, the indoor contribution over 70% only occurred when the low-cost light-scattering sensor was used and the outdoor $PM_{2.5}$ concentrations were low. The outdoor $PM_{2.5}$ concentrations at the low level tended to be under-reported by the light-scattering sensor, which would result in a higher indoor contribution. The under-reported outdoor $PM_{2.5}$ concentration by the low-cost light-scattering sensor could be another possible reason for the long continuous indoor emission. In conclusion, the proposed algorithm can effectively differentiate indoor $PM_{2.5}$ of outdoor and indoor origin and estimate their contributions to the total indoor exposure. However, the average results of daily indoor/outdoor contributions estimated based on the two outdoor $PM_{2.5}$ datasets had an around 5% difference.

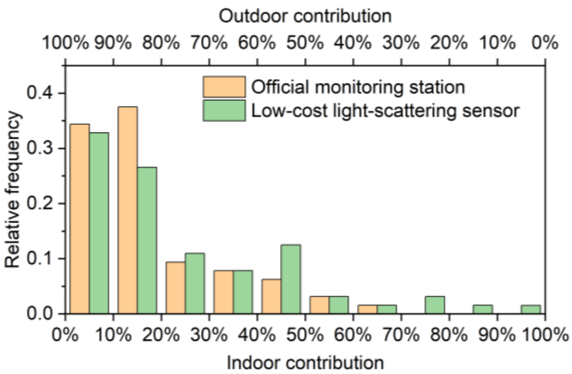

Fig. 2 - Comparison of the probabilistic distribution of the daily indoor/outdoor contribution estimated based on the outdoor $PM_{2.5}$ data from the low-cost sensor and official monitoring station in February to April 2017.

Fig. 3 compares the probabilistic distribution of the time-resolved infiltration factor estimated based on the outdoor $PM_{2.5}$ data from the low-cost sensor and official monitoring station in February to April. Again, the general distributions were similar, but discrepancies can be observed. The average infiltration factor estimated using the outdoor $PM_{2.5}$ data from the low-cost sensor was 0.46, which was equal to that estimated based on the official monitoring data, 0.46. Therefore, the proposed algorithm can effectively calculate the time-resolved infiltration factor using only the inputs of time-series indoor/outdoor $PM_{2.5}$ concentrations and window behavior. Furthermore, the average infiltration factors obtained from the two outdoor $PM_{2.5}$ datasets were similar, but discrepancies can be observed in terms of the probabilistic distribution.

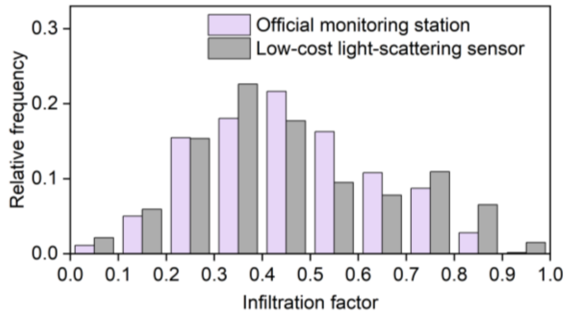

Fig. 3 - Comparison of the probabilistic distribution of the time-resolved $PM_{2.5}$ infiltration factor estimated based on the outdoor $PM_{2.5}$ data from the low-cost sensor and official monitoring station in February to April 2017.

Theoretically, using the same light-scattering sensors with careful calibration for both indoor and outdoor $PM_{2.5}$ monitoring would yield more accurate results than using the official monitoring outdoor data since the light-scattering sensor can effectively capture the peak outdoor $PM_{2.5}$ concentration with a 10-min time step size but the official monitoring data with a 2-h sampling interval cannot. In addition, the nearest official monitoring station may still be far away from the target building, which would result in inaccurate input of outdoor $PM_{2.5}$. Therefore, from the theoretical perspective, we would recommend using the same low-cost light-scattering sensors with careful calibration for both indoor and outdoor $PM_{2.5}$ monitoring.

However, using the low-cost light-scattering sensor to monitor long-term outdoor $PM_{2.5}$ in a neighborhood is practically challenging for general customer use. Several problems were identified in the monitoring of this study. First, the current low-cost light-scattering sensors available on the market suffers from severe data loss due to bad weather, unstable power supply, or even accidental damage. The general users, such as the participants in this study, would not spend time on regular maintenance for the outdoor sensor. Furthermore, most of them do not have the technical skills to fix a light-scattering sensor. Second, the outdoor sensor should be placed somewhere in the neighborhood which is a public area. Thus, the outdoor monitoring requires the permission from the neighborhood committee, which would become impractical if a lot of residents request a public area for outdoor monitoring. Therefore, from the practical perspective, we would recommend using the outdoor $PM_{2.5}$ data from the nearest official monitoring station as the input for the algorithm if the year-around results are to be obtained.

In the future, efforts should be made in the following aspects to facilitate the practical application of the proposed algorithm by using low-cost light-scattering sensor for outdoor $PM_{2.5}$ monitoring. First, the sensors should be further developed for robust and stable long-term measurements in relatively hash environments. Currently, there are some well-designed $PM_{2.5}$ monitors specifically for outdoor monitoring available on the market. However, the cost would be too high for general customer use. Low-cost solutions would significantly facilitate the practical applications. Second, the neighborhood-based outdoor $PM_{2.5}$ monitoring should be conducted by the neighborhood property manager and the data should be shared in real-time with all the residents in the neighborhood. This would require both technical development in terms of data sharing and policy development in terms of neighborhood-based air quality monitoring.

3.3. Year-round results

Based on the discussion above, this study then used the outdoor $PM_{2.5}$ data from the official monitoring station to calculate the year-round indoor $PM_{2.5}$ of outdoor and indoor origin as a full demonstration of the proposed algorithm. Furthermore, the seasonal characteristics were analyzed in addition to the year-round results. Division of the time into four seasons was based on the five-day moving average temperature according to the definition of climatic season in Chinese national standard QX/T 152-2012 [14]. According to this standard, in 2017, spring in Tianjin was from March 16 to May 10, summer from May 11 to October 2, autumn from October 3 to November 11, and winter from January 1 to March 15 and November 12 to December 31.

Fig. 4 displays the year-round distribution of the daily indoor/outdoor contribution for the 275 days, with a wide range from 0 to 94.2%. On average, the indoor $PM_{2.5}$ emissions and outdoor $PM_{2.5}$ infiltration contributed 26.3% and 73.7% of the daily total indoor exposure, respectively. In other words, for most of the time, the outdoor $PM_{2.5}$ infiltration contributed to the indoor $PM_{2.5}$ more than the indoor emission did. The results demonstrated that the proposed algorithm can automatically differentiate indoor $PM_{2.5}$ of indoor and outdoor origin and estimate their contributions to the total indoor exposure, even for a whole year. The automated estimation of indoor and outdoor contributions would support the large-scale exposure and health risk assessment as well as the development of effective strategies for controlling indoor particulate air pollution.

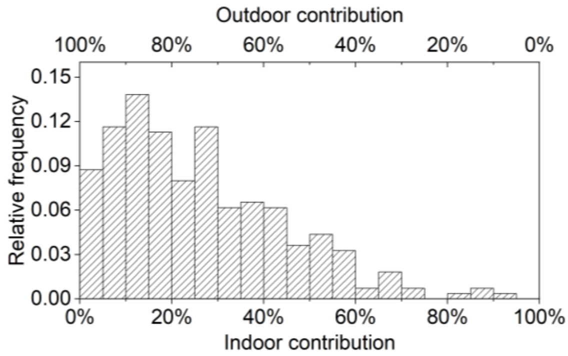

Fig. 4 - Year-round probabilistic distribution of the daily indoor/outdoor contribution to total indoor exposure in 2017.

Fig. 5 shows the year-round probabilistic distribution of the time-resolved $PM_{2.5}$ infiltration factor with a mean value ± standard deviation of 0.56 ± 0.22. The median value was 0.55. The results were comparable to the annual-averaged infiltration factor for residences in Beijing, 0.48 ± 0.07 [9]. The year-round infiltration factor showed a great span ranging from 0.001 to 0.993 in the same apartment. The great variation in the infiltration factor was attributed to the window behavior, outdoor wind speed, etc. Note that measuring time-resolved infiltration factor in a real building with indoor $PM_{2.5}$ sources is very challenging using the existing methods in the literature [13]. However, with the approach proposed in this study, the real-time $PM_{2.5}$ infiltration factors can be obtained by using only the concentrations of outdoor $PM_{2.5}$ and indoor $PM_{2.5}$ and window actions.

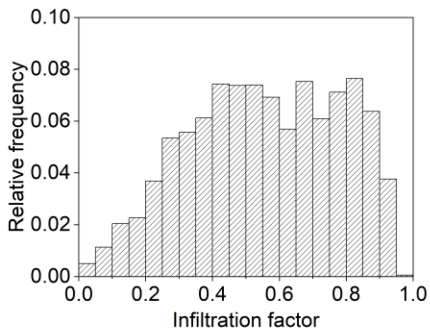

Fig. 5 - Year-round probabilistic distribution of the time-resolved $PM_{2.5}$ infiltration factor in 2017.

3.4. Sensitivity analysis

The sensitivity analysis in this study was to test how the empirically determined setting parameters, i.e., time step size, infiltration factor range (in Step 3), and R^2 value (in Step 4), would alter the results of indoor/outdoor contribution and infiltration factor based on the outdoor low-cost sensor data shown in Section 3.2.

Time step size: In general, a smaller time step size would result in greater uncertainties due to data fluctuation, while a larger time step size would result in greater error in quantifying indoor emission. The validation using the ground truth data in the laboratory tests in our previous study [10] indicated that the time step size of 10 min yielded the best estimation of the indoor/outdoor contribution. This study further tested how the time step size of 2, 5, 10, 15, and 20 min affected the results in the Tianjin apartment. The average daily indoor contribution increased with the time step size, while the infiltration factor decreased. The change point analysis in Step 2 may generate more change points with a smaller time step size due to data fluctuation. If a real indoor emission was divided into several time periods by the additional change points, the algorithm might misclassify these short time periods as without emission. Consequently, the average daily indoor contribution would be underestimated, while the infiltration factor would be overestimated. On the other hand, a larger time step size might result in the overestimation of indoor contribution when calculating the integral in the numerator of equation (5). As a result, the average daily indoor contribution would be overestimated, while the infiltration factor would be underestimated. When the time step size ranged from 5 to 15 min, the average daily indoor contribution in the range of 21.3% to 24.1% with an absolute difference of 2.8%, and the average infiltration factor was in the range of 0.47 to 0.50 with an absolute difference of 3%. Therefore, if an uncertainty of 5% in indoor/outdoor contribution is acceptable, the time step size can be set between 5 to 15 min. Furthermore, the time step size between 5 to 15 min is also suitable considering the light-scattering $PM_{2.5}$ sensor performance and typical indoor emission duration [10].

Infiltration factor range threshold: The infiltration factor range threshold of 0.22 in Step 3 was determined according to the reasonable inputs from the literature. The validation using the ground truth data in the laboratory tests [10] also indicated that the threshold of 0.22 yielded the best estimation of the indoor/outdoor contribution. This study tested how the threshold of 0.15, 0.19, 0.22, 0.28, and 0.35 affected the results in the Tianjin apartment in February to April. These values corresponded to the $75^{th}/25^{th}$, $80^{th}/20^{th}$, $85^{th}/15^{th}$, $90^{th}/10^{th}$, and $95^{th}/5^{th}$ percentiles of the upper/lower limits of the infiltration rate, respectively [9]. A larger infiltration factor range threshold resulted in a lower the indoor contribution and a higher infiltration factor. When the infiltration factor range threshold ranged from 0.15 to 0.35, the average daily indoor contribution was in the range of 21.7% to 24.8% with an absolute difference of 3.1%, and the average infiltration factor was in the range of 0.47 to 0.48 with an absolute difference of only 0.01. Therefore, the results were insensitive to the infiltration factor range threshold between 0.15 and 0.35.

R^2 value: The algorithm used an R^2 value of 0.8 in the data regression analysis in Step 4 for identifying whether there was $PM_{2.5}$ emission in a period with a window-opening action. Again, the validation using the ground truth data in the laboratory tests [10] also indicated that the R^2 value of 0.8 yielded the best estimation of the indoor/outdoor contribution. This study tested how the R^2 value of 0.65, 0.7, 0.8, 0.9, and 0.95 affected the results in the Tianjin apartment in February to April. when the R^2 value range threshold ranged from 0.65 to 0.95, both the average daily indoor contribution and the average infiltration factor almost remain unchanged. Therefore, the results were insensitive to the R^2 value in the range of 0.65 to 0.95.

4. Conclusions

This study used an indoor/outdoor $PM_{2.5}$ differentiation algorithm in real residential apartments to automatically estimate the long-term time-resolved indoor $PM_{2.5}$ of outdoor and indoor origin. The inputs for the differentiation algorithm were only the concentration values of outdoor and indoor $PM_{2.5}$ and occupants' window actions, which were easily obtained from the low-cost sensors. The indoor/outdoor contribution to the gross indoor exposure and the time-resolved infiltration factor were calculated using the algorithm. Within the scope of this study, the following conclusions can be made:

1. The proposed algorithm can automatically estimate the long-term time-resolved indoor $PM_{2.5}$ of outdoor and indoor origin in naturally ventilated buildings using only the inputs of time-resolved indoor/outdoor $PM_{2.5}$ concentrations and window behavior.

2. The indoor/outdoor contribution to the gross indoor exposure and time-resolved infiltration factor can also be automatically estimated using the algorithm.

3. This study identified several directions for further development, such as robust long-term outdoor $PM_{2.5}$ monitoring using low-cost light-scattering sensors, which would facilitate the practical applications of the algorithm.

5. Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 51708474) and the Early Career Scheme of Research Grants Council of Hong Kong SAR, China (Grant No. 24208518).

6. References

[1] GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990 – 2017: a systematic analysis for the Global Burden of Dis. Lancet. 2018;392:1923–94.

[2] Guo Y, Zeng H, Zheng R, Li S, Barnett AG, Zhang S, et al. The association between lung cancer incidence and ambient air pollution in China: A spatiotemporal analysis. Environ Res. 2016;144:60–5.

[3] Pope CA, Ezzati M, Dockery DW. Fine particulate air pollution and life expectancies in the United States: The role of influential observations 2013;63:129–32.

[4] Weichenthal S, Hatzopoulou M, Goldberg MS. Exposure to traffic-related air pollution during physical activity and acute changes in blood pressure, autonomic and micro-vascular function in women: A cross-over study. Part Fibre Toxicol. 2014;11.

[5] Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, et al. The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. J Expo Anal Environ Epidemiol. 2001;11:231–52.

[6] Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, et al. An Association between Air Pollution and Mortality in Six U.S. Cities. N Engl J Med. 1993;329:1753–9.

[7] Vanos JK, Hebbern C, Cakmak S. Risk assessment for cardiovascular and respiratory mortality due to air pollution and synoptic meteorology in 10 Canadian cities. Environ Pollut. 2014;185:322–32.

[8] Chen C, Zhao B, Weschler CJ. Indoor Exposure to "Outdoor PM10": Assessing Its Influence on the Relationship Between PM10 and Short-term Mortality in U.S. Cities. Epidemiology. 2012;23:870–8.

[9] Shi S, Chen C, Zhao B. Modifications of exposure to ambient particulate matter: Tackling bias in using ambient concentration as surrogate with particle infiltration factor and ambient exposure factor. Environ Pollut. 2017;220:337–47.

[10] Xia T, Chen C. Differentiating between indoor exposure to PM2.5 of indoor and outdoor origin using time-resolved monitoring data. Build Environ. 2019;147:528–39.

[11] Liu J, Dai X, Li X, Jia S, Pei J, Sun Y, et al. Indoor air quality and occupants' ventilation habits in China: Seasonal measurement and long-term monitoring. Build Environ. 2018;142:119–29.

[12] Taylor WA. Change-Point Analysis: A Powerful New Tool For Detecting Changes. Analysis. 2000;1–19.

[13] Chen C, Zhao B. Review of relationship between indoor and outdoor particles: I/O ratio, infiltration factor and penetration factor. Atmos Environ. 2011;45:275–88.

[14] China Meteorological Administration, QX/T 152-2012, Definition of Climatic Season 2012.