Air Quality, Atmosphere & Health (2022) 15:1571–1581
https://doi.org/10.1007/s11869-022-01202-2

Spatial representativeness of $PM_{2.5}$ monitoring stations and its implication for health assessment

Heming Bai[1] · Rusha Yan[2] · Wenkang Gao[3] · Jing Wei[4] · Myeongsu Seong[1]

Received: 17 November 2021 / Accepted: 25 April 2022 / Published online: 9 May 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
Air pollution measurements from monitoring stations are widely used in health assessment, and it is important to take into account the spatial representativeness (SR) of stations when quantifying population exposure to air pollution measured in these stations. Using high-quality satellite-derived $PM_{2.5}$ data with 1-km spatial resolution over Yangtze River Delta (YRD) from 2016 to 2020, this study estimates SR of 213 $PM_{2.5}$ monitoring stations, and these SR estimates are further used to calculate annual population-weighted mean (PWM) $PM_{2.5}$ and deaths attributable to $PM_{2.5}$ exposure for each city in YRD. Our results show that SR areas of 213 stations totally account for 32.33% of the area of YRD, and the SR size varies greatly with stations. Additionally, we find that the city-level PWM $PM_{2.5}$ based on SR is nearly always larger than that using full coverage satellite-derived data. The difference tends to decrease as the population ratio of SR area increases. For the entire YRD, attributable deaths using PWM $PM_{2.5}$ based on SR are 182,009 (95% CI: 136,632–225,081), and are comparable to the ones derived using full-coverage satellite-derived data. Nevertheless, the relative change in attributable deaths is more than 6% in some cities due to the low population ratio of SR (less than 20%), which suggests that more monitoring stations should be deployed in these cities for human assessment.

Keywords Spatial representativeness · Satellite remote sensing · Monitoring stations · PM2.5 · Health assessment

Introduction

Spatial representativeness (SR) is defined as the extent to which the monitoring data is meaningful and useful in a spatial context, and is an important parameter for interpreting monitoring data (Righini et al. 2014). It should also be carefully considered when air quality measurements are used for data assimilation in air quality models (Elbern et al. 2007). Additionally, spatial representativeness is also useful to validate satellite-derived air pollution concentration at a certain spatial resolution. Besides, spatial representativeness can provide useful information for optimal monitoring network design (Piersanti et al. 2015).

Some different methods have been investigated to estimate the spatial representativeness of stations. The most straightforward method is based on the specific measurement campaigns with dense monitoring networks (Blanchard et al. 2014; Shi et al. 2018; Vardoulakis et al. 2005). However, this method is expensive if a large number of samplers are deployed. Additionally, results based on the measurement campaigns strongly depend on the spatial distribution of the samplers. Some methods use surrogate indicators to investigate concentration variability when measurements around sites are not available, such as emission sources (Righini et al., 2014) and land use characteristics (Janssen et al. 2012). Some methods are based on air quality models that can take the effects of emission and meteorology into account (Martin et al. 2014; Piersanti et al. 2015; Santiago et al. 2013). However, simulated results are subject to

✉ Heming Bai
hemingbai@ntu.edu.cn

[1] Research Center for Intelligent Information Technology, Nantong University, Nantong, China

[2] State Environmental Protection Key Laboratory of Formation and Prevention of Urban Air Pollution Complex, Shanghai Academy of Environmental Sciences, Shanghai, China

[3] State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

[4] Earth System Science Interdisciplinary Center, Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD, USA

emission data accuracy and different microphysical schemes (Baró et al. 2015; Cao et al. 2021; Schneider et al. 1997). Moreover, applying air quality models at a fine spatial resolution requires more computational cost, and it is not an effective way to investigate a large research area. Alternatively, using satellite-based air pollution data would be a cost-effective method since they can provide long-term data with high spatial resolution and wide coverage. Recent studies have developed the China-High-Air-Pollutants (CHAP) data set based on satellite remote sensing and machine learning (Wei et al. 2021, 2020). These high-resolution and high-quality data sets include seven major air pollutants, and have great potential for investigating the spatial representativeness.

Many previous studies have used ground-based air pollution data for human health assessment, but they did not consider the difference in population covered by spatial representativeness area of different sites (Bai et al. 2022; Chen et al. 2017; Dominici et al. 2006; Song et al. 2017); they generally directly averaged air pollution data from each site within a city to derive the exposure concentrations, which may lead to uncertainty in health assessment. Thus, it is meaningful to combine the concept of spatial representativeness and ground-based air pollution data for health assessment.

The main objective of this study is to estimate the spatial representativeness of monitoring stations in Yangtze River Delta (YRD) by using the satellite-derived CHAP data with 1-km spatial resolution. We focus on particulate matter with an aerodynamic diameter of 2.5 μm ($PM_{2.5}$) since it is the primary pollutant in most cities in China (Lv et al. 2015). Furthermore, we apply the spatial representativeness to assess the exposure concentrations and deaths attributable to $PM_{2.5}$. The remainder of this paper is organized as follows. "Data and methods" introduces data on $PM_{2.5}$, population, and mortality, as well as methods for estimating the spatial representativeness and deaths attributable to $PM_{2.5}$ pollution. The results are presented in "Results." The discussions are made in "Discussion," followed by the conclusions in "Conclusions."

Data and methods

Study area

The study region is YRD, which is located in Eastern China and contains a total of 26 cities. The YRD region only covers 2.2% of the national land area (Hu et al. 2018), but accounts for 11.4% of the national population and 20.2% of the national gross domestic product in 2020 based on the latest statistical yearbook. In addition, YRD is one of the most polluted areas in China (Bai et al. 2021); however, annual mean $PM_{2.5}$ concentrations over this region exhibited a significant downward trend in the past few years (H. Zhao et al. 2021a, b). According to the China Ecology and Enviroment Bulletion 2020 (https://www.mee.gov.cn/), the annual mean concentration of $PM_{2.5}$ in YRD in 2020 is 35 μg/m^3.

Data sets

Monitoring stations and satellite-derived $PM_{2.5}$ data

The spatial representativeness (SR) of 213 stations in YRD was assessed by using the satellite-derived 1-km-resolution daily $PM_{2.5}$ concentrations (i.e., ChinaHigh$PM_{2.5}$) collected from the CHAP data set (https://weijing-rs.github.io/product.html). The ChinaHigh$PM_{2.5}$ was generated based on a newly developed space–time extremely randomized tree model by using MODIS/Terra + Aqua MAIAC AOD products together with other auxiliary data (Wei et al. 2021, 2020). The daily $PM_{2.5}$ estimates have high accuracy with a cross-validation coefficient of determination reaching 0.90. Note that $PM_{2.5}$ estimations from the ChinaHigh$PM_{2.5}$ data set are 24-h average rather than mean values during the satellite overpassing period. One of the reasons for this is that satellite-based $PM_{2.5}$ retrieval models using the $PM_{2.5}$ 24-h average may have better performance than those using the $PM_{2.5}$ average during the satellite overpassing period (Han et al. 2018). Another reason is that $PM_{2.5}$-related health effects are usually assessed by using 24-h average concentrations. Here, we employed the ChinaHigh$PM_{2.5}$ data for the period from 2016 to 2020 to ensure that more samples can be collected.

The spatial distribution of 213 stations in YRD is shown in Fig. 1. These sites are the state-controlled and non-background sites. This site category is selected because they are officially used to assess air quality at the city level or larger scale and are widely used in health assessment. To confirm the SR results, a case analysis in Zhoushan City was carried out based on ground-based $PM_{2.5}$ measurements. Hourly $PM_{2.5}$ concentrations data of 3 stations (the site codes are 1258A, 1259A, and 1260A, respectively) in Zhoushan City from 2016 to 2020 were obtained from the China National Environmental Monitoring Center (CNEMC, http://www.cnemc.cn). Daily $PM_{2.5}$ concentrations were computed from the hourly time series when more than 18 measurements in a day were available (Barrero et al., 2015). Daily mean ground-based $PM_{2.5}$ data within 24-h measurements are used to verify SR results from satellite-based $PM_{2.5}$ concentrations.

Population and mortality data

In this study, population data for 2020 were obtained from the Gridded Population of the World, Version 4

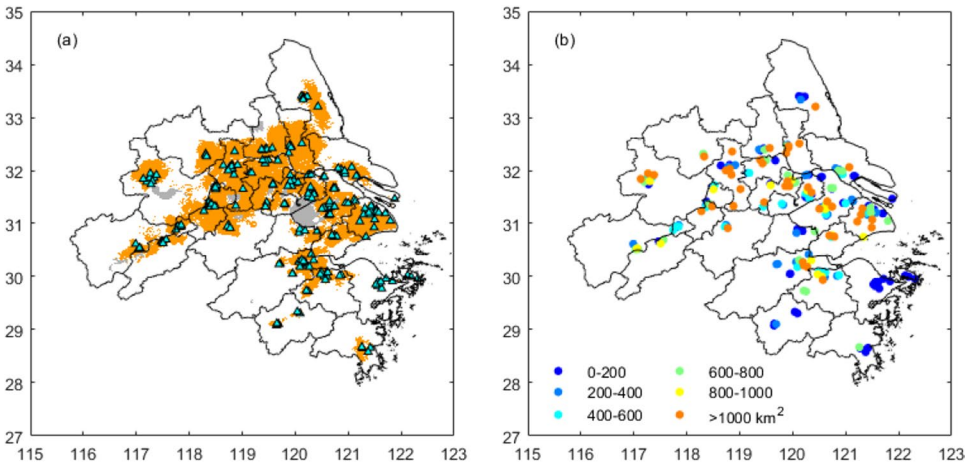

Fig. 1 **a** The spatial representativeness (SR) areas of 213 $PM_{2.5}$ monitoring stations in Yangtze River Delta. SR sizes for these stations are categorized into 8 bins, and shown in panel **b**

(Doxsey-Whitfield et al. 2015). Population estimates with 1 km spatial resolution, available at https://sedac.ciesin.columbia.edu/data/collection/gpw-v4, were used. The population estimates for 2020 at city level from the Seventh China Census were also used to adjust the gridded population data. Specifically, for a given city, we scaled the gridded values by the ratio of the city's total population from census data to gridded data. Data on the age structure at the national level for 2019, as well as the age-specific and disease-specific mortality were obtained from Global Burden of Disease Study 2019 (GBD 2019) dataset (https://vizhub.healthdata.org/gbd-compare/).

Methods

Spatial representativeness

Spatial representativeness is defined as the spatial area over which the air quality data for a given monitoring site can be considered representative (Zoroufchi Benis and Fatehifar, 2015). We combine the spatial coverage of SR and population distribution for health impact assessment. SR analysis should also consider functional representation (e.g., industrial, port and traffic sites). However, such sites are limited and corresponding information is not currently public except background station. Hence, this study only compares SR estimations between non-background and background stations (see "Discussion"). To assess the SR of the monitoring locations for $PM_{2.5}$, we adopted a method of Concentration Similarity Frequency (CSF) defined as the following formula (Piersanti et al. 2015):

$$F_{site}(x,y)=\frac{\sum_{k=1}^{N_t} Flag}{N_t}, whereFlag=\begin{cases}1, \frac{|PM2.5(X_{site},Y_{site},t_k)-PM2.5(x,y,t_k)|}{PM2.5(X_{site},Y_{site},t_k)}<0.2\\ 0, \frac{|PM2.5(X_{site},Y_{site},t_k)-PM2.5(x,y,t_k)|}{PM2.5(X_{site},Y_{site},t_k)}>0.2\end{cases} \quad (1)$$

where F_{site} (x, y) is a frequency function to determine whether the grid point (x, y) is included in the SR area of the monitoring site $(X_{\text{site}}, Y_{\text{site}})$. PM2.5$(x, y, t)$ represents the surface concentration field from satellite-derived $PM_{2.5}$ data. N_t is the number of pairs of $PM_{2.5}$ data. Flag is the concentration similarity at time t_k by comparing $\Delta PM_{2.5}/PM_{2.5}$ with a threshold of 20% (Piersanti et al. 2015). This work mainly focuses on state-controlled and non-background stations. According to the technical regulation for selection of ambient air quality monitoring stations in China (MEPC, 2013), a state-controlled and non-background station generally represents an area within the 0.5- to 4-km radius of this station. The maximum radius can be expanded to dozens of kilometers over regions with slight spatial variation in air pollutant concentrations (MEPC, 2013). Therefore, we assumed that the maximum SR area is a box of 100 km × 100 km centered on a given site. After calculating F_i (x, y) for each grid point in the box, the SR area of the site was assessed as the area where the condition F_{site} $(x, y) > 0.9$ is met on a multi-year basis (Piersanti et al. 2015).

Note that satellite-based $PM_{2.5}$ used in this study suffers from data gaps due to AOD missing, which may influence the SR estimation. However, using 5-year satellite-based $PM_{2.5}$ data produces enough samples to alleviate this issue. Specifically, the mean sample size of grid-site pairs within the 100 km × 100 km box is higher than 300 for 85% of stations in YRD. Thus, we would not expect this data gap issue to influence the estimated SR significantly. Furthermore, we plan to assess SR by using gap-free $PM_{2.5}$ data in the future.

Estimating deaths attributable to $PM_{2.5}$ pollution

Following a similar method of the GBD 2019 project (GBD 2019 Risk Factors Collaborators), we estimated deaths attributable to exposure to annual mean $PM_{2.5}$ using the following equation:

$$M_i = \sum_a \sum_d \left(\mathrm{POP}_i \times \mathrm{Age}P_{i,a} \times MB_{i,a,d} \times \frac{RR_{a,d} - 1}{RR_{a,d}} \right) \quad (2)$$

where POP_i stands for the total population for city i; $\mathrm{Age}P_{i,a}$ is the proportion of population with age a in the city i; $MB_{i,a,d}$ is the baseline mortality of disease d for people with age a in the city i; $RR_{a,d}$ is the relative risk for disease d in a population with age a. Five mortality endpoints associated with $PM_{2.5}$ pollution were estimated in this study, including stroke, ischemic heart disease (IHD), chronic obstructive pulmonary disease (COPD), lung cancer (LC), and lower respiratory infection (LRI). We used the updated RR from a recent study (McDuffie et al. 2021), which was based on a Meta Regression-Bayesian, Regularized, Trimmed (MR-BRT) spline from the GBD 2019. MR-BRT RR uses splines with Bayesian priors in order to avoid using relative risk estimates for active smoking (McDuffie et al. 2021). RR is a function of annual population-weighted mean (PWM) $PM_{2.5}$ in each city. Note that we used national-level Age$P$ and MB from GBD 2019 given that these data at city level are unavailable.

Here, for a given city, SR-based PWM $PM_{2.5}$ was calculated by weighting the annual $PM_{2.5}$ concentration at each station in the city and population covered by the SR area of the corresponding station. To examine their reliability, SR-based PWM $PM_{2.5}$ values were further compared to those by using full coverage satellite-derived $PM_{2.5}$ and gridded population data (referred to as full coverage PWM $PM_{2.5}$).

Results

Spatial representativeness analysis

The coverage of SR areas for 213 stations in YRD is shown as orange areas in Fig. 1a. The orange areas are mainly located in the central area of YRD. The SR areas only account for 32.33% of the total area of YRD, and mainly correspond to urban areas. Figure 1 b shows SR area sizes for each station in YRD. About 70% of the stations correspond to SR size less than 700 km^2. Generally, there is no obvious spatial pattern for SR sizes across 213 stations in YRD (Fig. 1b), except for coastal areas where SR sizes are typically small and less than 200 km^2.

We find a large variability in the size and shape of SR (Fig. 1). This variability is most likely related to the local meteorology, surface condition, and emissions, and has also been reported by previous studies (Martin et al. 2014; Piersanti et al. 2015; Shi et al., 2018). For instance, Piersanti et al. (2015) found a coast-inland difference in SR size, with lower values for sites in coastal areas. This finding is consistent with our results (Fig. 1b). The low SR estimations in coastal areas are partly because these areas are influenced by land-sea breeze recirculation and thus may exhibit a large pollutant horizontal gradient (Ding et al. 2004; Russo et al., 2016). Additionally, due to a large intra-urban spatial variability of $PM_{2.5}$ concentrations in some cities in YRD (Liu et al., 2016), these regions may exhibit a large difference in SR estimations among different sites at a city level. Furthermore, previous studies have shown that SR areas have no fixed shape but with directional preference (Piersanti et al. 2015; Shi et al. 2018). They also found that SR areas were spatially discrete for some stations. These findings of the SR shape are in accordance with our results. Note that, by using the same method as ours, Piersanti et al. (2015) showed good performances in describing the size and the shape of the SR area. Hence, we would expect that the method used to estimate SR in this study is reasonable. Nevertheless, the difference in SR between different methods still needs to be examined in the future.

SR sizes are very small and less than 10 km^2 in some cases. For instance, SR sizes for 3 stations in Zhoushan city located in the east of Zhejiang province range from 3 to 6 km^2. To verify such cases based on satellite-derived $PM_{2.5}$ data, we used ground-based $PM_{2.5}$ measurements to analyze the concentration similarity among these 3 adjacent stations in Zhoushan city. Specifically, the daily time series of $PM_{2.5}$ at station 1260A is treated as the reference series (Fig. 2a) given that this station is located in the center of the other two stations. The relative change in daily $PM_{2.5}$ between other stations and the reference station was then calculated (Fig. 2b), thereby estimating CSF (see " Spatial representativeness" for details). As shown in Fig. 2b, the relative changes exceed the threshold ($\pm$ 20%) in many cases. The CSF values for paired stations of 1258A–1260A and 1259A–1260A are 0.80 and 0.89, respectively, and both are less than the threshold of 0.9. Therefore, stations 1258A and 1259A are not covered by the SR area of the 1260A station, despite the fact that their distances are very small (7 km for 1258A–1260A and 12 km for 1259A–1260A). These results could support our conclusion about the low SR size for station 1260A using satellite-derived data.

Given that SR areas for two or more stations may overlap, we propose an indicator R_i to estimate how redundant stations are for a given city defined as the following formula:

$$R_i = \frac{\sum_{j=2}^{n} (j-1) \times S_{ij}}{(n-1) \times S_i} \quad (3)$$

where S_i is the overall area of SR in a city i, $\underline{n}$ is the number of stations in which the SR area is fully or partially contained in the city i. S_{ij} stands for the overlap SR area contributed by j stations ($j \geq 2$). R_i ranges from 0 to 1, 0 indicating

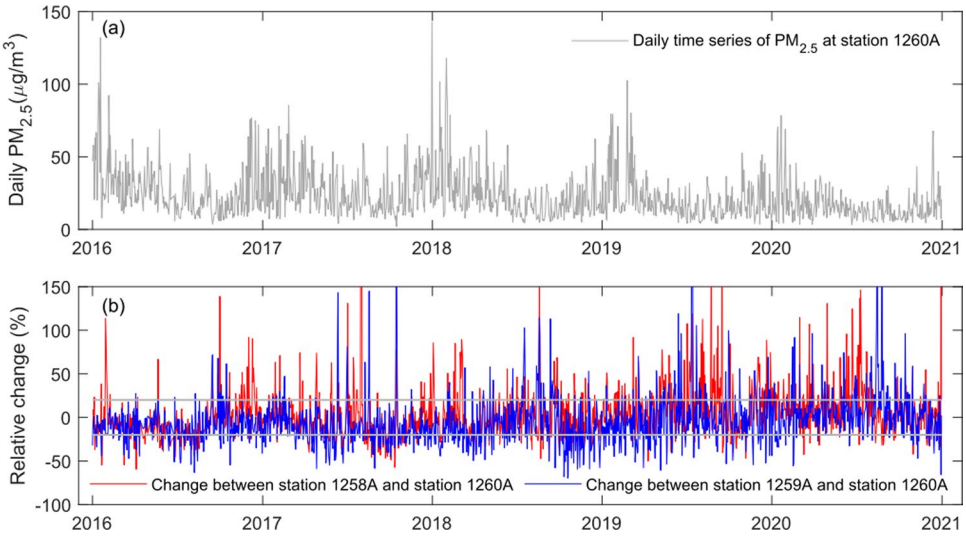

Fig. 2 **a** Daily time series of $PM_{2.5}$ concentration from 2016 to 2020 at station 1260A. Red (blue) line in **b** presents the difference in daily $PM_{2.5}$ concentration between station 1258A (1259A) and station 1260A

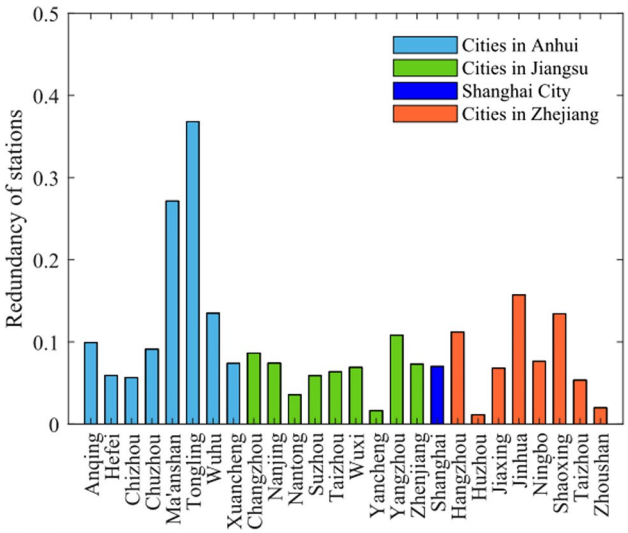

Fig. 3 The redundancy of stations for 26 cities in Yangtze River Delta. Different color bars correspond to cities in different provinces or municipality

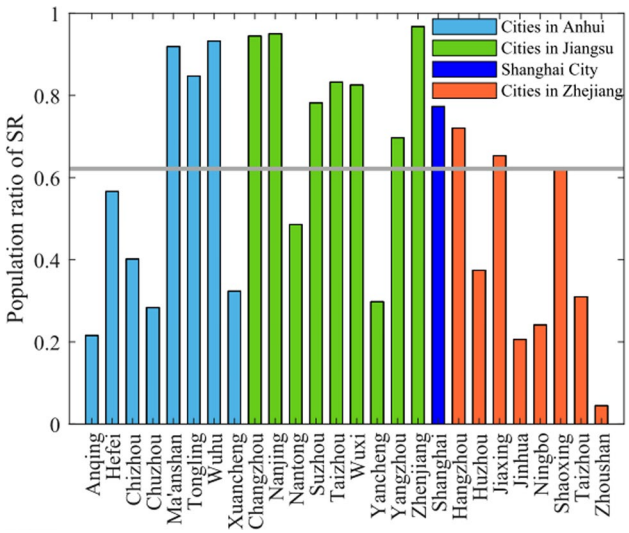

Fig. 4 Population ratio of spatial representativeness (SR) area to city area for each city in Yangtze River Delta. Different color bars correspond to cities in different provinces or municipality

no overlap area and thus no redundancy in SR, and 1 indicating that the SR areas among n stations are exactly the same.

Figure 3 shows the R_i for 26 cities in YRD. As can be seen, the redundancy of stations for 19 cities is less than 0.1, which suggests that the spatial distribution of stations is reasonable for most cities from a redundancy perspective. Note that the redundancy is relatively high for Ma'anshan (0.27) and Tongling (0.37), which is likely due to a combination of a concentrated spatial distribution of stations and the large SR sizes for these stations.

Appling SR for health assessment

Combining SR areas with the spatial distribution of population, we can estimate the population ratio of SR areas to city areas for each city in YRD, as shown in Fig. 4. The population ratio varies greatly with cities: the minimum value of 4.42% appears in Zhoushan city; the values are less than 30% for the additional five cities (i.e., Anqing, Chuzhou, Yancheng, Jinhua, and Ningbo); and only five cites have the population ratios above 90%, including Ma'anshan, Wuhu, Changzhou, Nanjing, and Zhenjiang. For the entire YRD region, the SR areas of all monitoring stations can only cover 62.16% of the population (see the gray line in Fig. 4). We further examined the relationship between the population

ratio of SR and the area ratio of SR. As shown in Fig. 5, the population ratio tends to increase with increasing area ratio. Note that there is one exceptional case in Hangzhou, where the SR area covers only 16.83% of the city area, but the population ratio of SR is up to 72.03%.

SR estimates were also used to calculate annual PWM $PM_{2.5}$ for each city in YRD. Figure 6 a shows a scatter of SR-based PWM $PM_{2.5}$ and full-coverage PWM $PM_{2.5}$ (see "Estimating deaths attributable to PM2.5 pollution" for details). As can be seen, SR-based PWM $PM_{2.5}$ values are higher than full-coverage PWM $PM_{2.5}$ values for all cities in YRD except Nanjing (Fig. 6a). Compared to full-coverage PWM $PM_{2.5}$ for the entire YRD region (43.35 μg/m^3), SR-based PWM $PM_{2.5}$ overestimates by 6.30%. The overestimate is partly because most stations are located in urban areas with a high pollution level (Gao et al. 2020). Furthermore, the overestimation may also be related to a positive relationship between $PM_{2.5}$ concentration at each station and population covered by SR of each station, as shown in Fig. 7. Due to this relationship, SR-based PWM $PM_{2.5}$ is mainly contributed by those stations that represent both high levels of population and $PM_{2.5}$ concentration.

We further estimate deaths attributable to $PM_{2.5}$ from different sources (i.e., full coverage PWM $PM_{2.5}$ and SR-based PWM $PM_{2.5}$). Figure 8 a shows attributable deaths by using full-coverage PWM $PM_{2.5}$ for each city in YRD. Such estimations vary greatly with cities. More than 10,000 attributable deaths are estimated for Hefei, Nanjing, Suzhou, Shanghai, and Hangzhou. For Chizhou, Tongling, and Zhoushan, attributable deaths are less than 2000 (Fig. 8a). Additionally, there are a total of 176,908 (95% CI: 131,664–220,023)

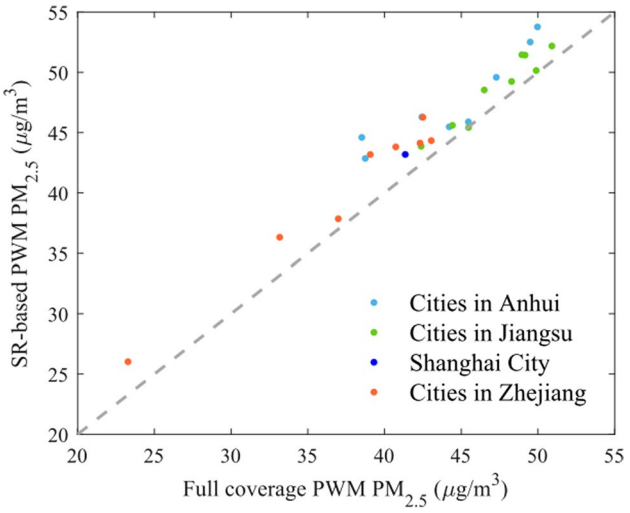

Fig. 6 Scatter plot of SR-based population-weighted mean (PWM) $PM_{2.5}$ versus full coverage PWM $PM_{2.5}$ for each city in Yangtze River Delta. Different color points correspond to cities in different provinces or municipality. See text for details

attributable deaths for the entire YRD by using full coverage PWM $PM_{2.5}$. Attributable deaths are mainly from Stroke and IHD (66.53%), followed by COPD, LC, and LRI.

Compared with the assessment by using full-coverage PWM $PM_{2.5}$ (Fig. 8a), attributable deaths by using SR-based PWM $PM_{2.5}$ are relatively high for all cities in YRD except Nanjing, as shown in Fig. 8b. These changes are more than 200 deaths in 9 out of 26 cities. For Shanghai and Hangzhou, the differences in attributable deaths are more than 500. In addition, attributable deaths using SR-based PWM $PM_{2.5}$ totally increase by 2.80% (5101) in

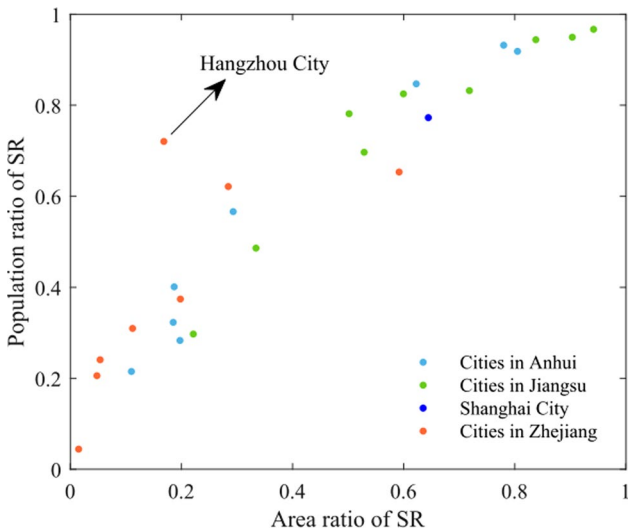

Fig. 5 Scatter plot of the population ratio of spatial representativeness (SR) versus the area ratio of SR for each city in Yangtze River Delta. Different color points correspond to cities in different provinces or municipality

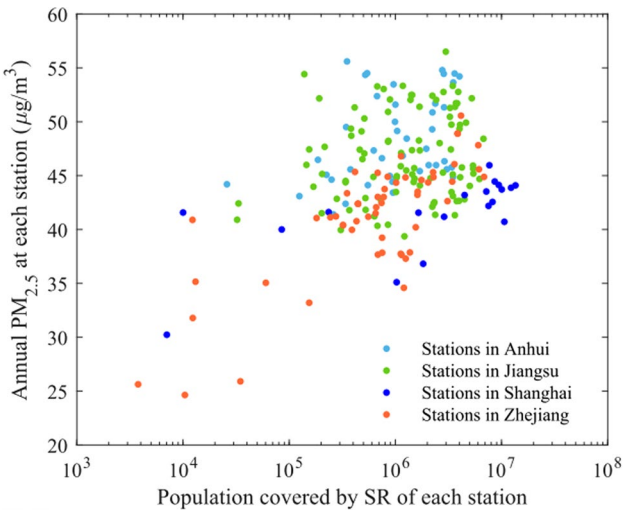

Fig. 7 Scatter plot of annual $PM_{2.5}$ at each station versus population covered by SR of each station. Different color points correspond to stations in different provinces or municipality

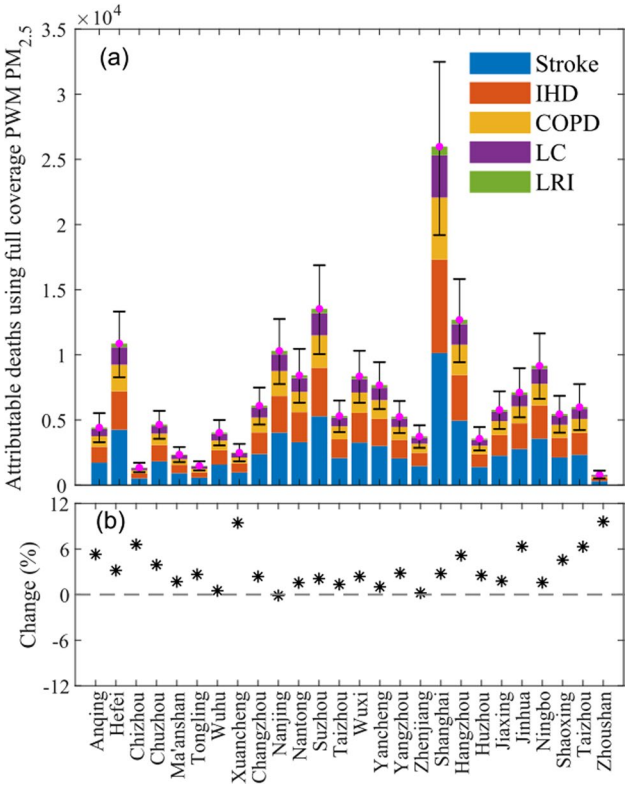

Fig. 8 **a** Deaths attributable to $PM_{2.5}$ exposure by using full coverage PWM $PM_{2.5}$ for each city in Yangtze River Delta, error bars denote 95% CI and different color bars stand for different causes of death, i.e., stroke, ischemic heart disease (IHD), chronic obstructive pulmonary disease (COPD), lung cancer (LC), and lower respiratory infection (LRI). **b** The difference in attributable deaths between using SR-based PWM $PM_{2.5}$ and full coverage PWM $PM_{2.5}$

YRD, compared to that using full-coverage PWM $PM_{2.5}$. Although the difference is low for the entire YRD region, the relative change at city level is more than 6% in five cities, including Chizhou (6.60%), Xuancheng (9.44%), Jinhua (6.34%), Taizhou (6.30%), and Zhoushan (9.60%).

Furthermore, by comparing Figs. 4 and 8b, we find that the more people SR areas of stations cover, the smaller the change in attributable deaths tends to occur. For those cities with the population ratios of SR above 90% (see Fig. 4), their relative changes are generally smaller than 1.70%. By contrast, for those cities with the low population ratios (less than 20%), their relative changes are higher than 6.30%. These results suggest that more monitoring stations should be deployed in cities with the low population ratios of SR.

Discussion

Our results show a large range of SR of stations in YRD (Fig. 1b). Thus, we should be cautious when using these site-based observations for the point-grid matching task. For example, a body of previous studies used ground-based $PM_{2.5}$ observations and satellite data to build a $PM_{2.5}$ retrieval model, and they generally directly matched a ground site with a satellite grid, without considering whether or not SR of the site is suitable for the satellite spatial resolution (Geng et al. 2015; He et al. 2018, 2016; Park et al. 2020; van Donkelaar et al., 2016). This may result in the uncertainty of satellite-derived $PM_{2.5}$ estimates in some cases when SR sizes of the stations are very low (even lower than satellite spatial resolution).

Our results show that SR performs poorly in YRD: SR areas only cover 32.33% of the total area, and 62.16% of the total population in YRD. Several reasons explain this poor SR performance. First, many stations are deployed in urban areas of YRD, but few stations are for rural areas. This situation may cause the redundancy of stations (Fig. 3) and limited rural areas covered by SR (Fig. 1a). Second, the location of the current stations is unsatisfactory. For instance, in Nantong City with the nine current stations, the SR area only covers 19.72% of the city area and 34.64% of the city population. These ratios would increase to 82.90% and 90.81% after optimizing the current stations in Nantong (not shown). The principle of the optimization approach is to iteratively find the grid where the SR area covers the largest population by using SR estimations for each grid in the city. It would be helpful to optimize the layout of the current stations in China in the future.

We find that the annual $PM_{2.5}$ concentrations at the stations generally tend to increase with increasing SR area of the stations, which is observed by comparing Fig. 7 with Fig. 5. This finding is not consistent with previous studies that suggested a large spatial heterogeneity under high pollution levels (Kikuchi et al., 2018). This inconsistency may be related to that they used aerosol optical depth as proxy for air quality, and warrants further investigation in the future. Additionally, for stations that meet the national standard of annual $PM_{2.5}$ concentrations (35 $\mu g/m^3$), their spatial representations are typically poor and less than 30 km^2 for most of these stations. Thus, it should be cautious to use $PM_{2.5}$ observations from these stations for air pollution evaluation and human health assessment.

This study focused on the state-controlled and non-background sites in YRD. It is meaningful to compare SR results among different categories of monitoring sites. However, other sites (e.g., industrial, port and traffic stations) are limited and corresponding information is not

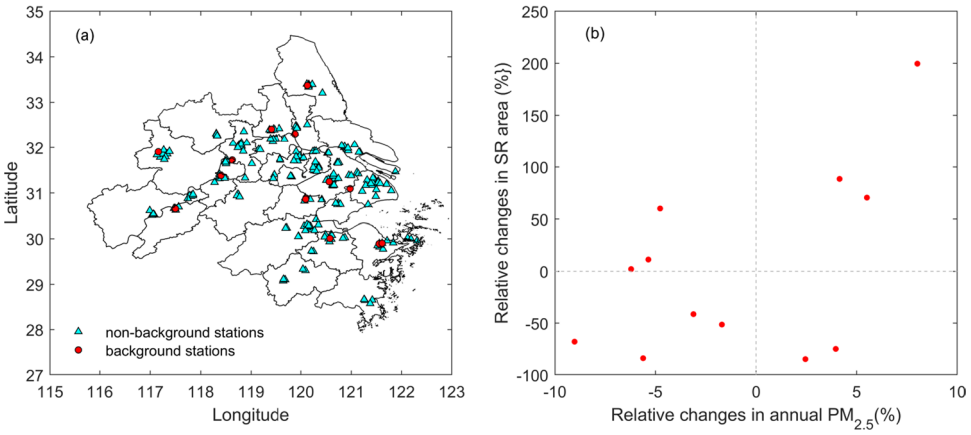

Fig. 9 **a** The spatial distribution of 213 non-background and 14 background stations in Yangtze River Delta. In **b**, the horizontal axis denotes the relative changes in annual $PM_{2.5}$ between background stations in a given city and non-background stations in the same city; the vertical axis is the same as the horizontal axis but for the spatial representative (SR) area. Note that Ningbo and Huzhou both have two background sites and arithmetic mean values are used to obtain SR area and $PM_{2.5}$ for background sites

currently public except background station. Therefore, we estimate SR of 14 background stations in YRD and further examine the difference of SR between background and non-background stations. As shown in Fig. 9a, the 14 background sites are sparsely distributed in 12 cities, and there are two background sites in Ningbo and Huzhou. Compared to the city average based on non-background sites, the annual $PM_{2.5}$ concentration and SR area for background sites exhibit an irregular variation (Fig. 9b). However, Fig. 9b generally shows a positive relationship between $PM_{2.5}$ concentration and SR area, which further supports our findings that the annual $PM_{2.5}$ concentration at a station tends to increase as the station's SR area increases.

According to "technical regulation for selection of ambient air quality monitoring stations" (on trial) (HJ 664–2013) released in China in 2013 (MEPC, 2013), one of the principles for the monitoring site layout is that the site should have a certain spatial representation. Specifically, for the state-controlled and non-background site, the technical regulation requires that SR size generally ranges from 1 to 50 km^2; for the background site, it requires that the SR size is generally larger than 31,400 km^2 (MEPC, 2013). Based on our SR results for sites in YRD, 18 out of 213 non-background sites meet the requirement, and most sites have an SR size greater than 50 km^2 (Fig. 1b). Additionally, for background sites in YRD, the maximum SR is 3973 km^2 (1208A station in Taizhou, Jiangsu province), thus failing to meet the SR size requirement in HJ 664–2013. Hence, we recommend relocating the current background sites to cover more spatial areas in the future.

Using different concentration–response functions (CRFs) may affect our health assessment. Deaths attributable to $PM_{2.5}$ exposure were also estimated by using an updated version of the Global Exposure Mortality Model (GEMM) (Burnett et al. 2018). The fractional disease contributions estimated by the GEMM are similar to those from the MR-BRT GBD2019 CRFs, whereas the absolute number of attributable deaths in each city in YRD is always larger when the GEMM is used (not shown). Moreover, based on the GEMM, the attributable deaths from SR-based PWM $PM_{2.5}$ totally increase by 7500 (3.54%) compared to that from full-coverage PWM $PM_{2.5}$. Based on the GEMM model and full-coverage PWM $PM_{2.5}$, attributable deaths in YRD are 211,927 (95% CI: 159,927–259,330). Our assessments are comparable to previous studies (Maji, 2020; Song et al. 2017). They reported about 205,000 attributable deaths in YRD in 2015. However, their assessments are supposed to be higher than our estimations due to a lower $PM_{2.5}$ concentration used in this study (5-year annual average from 2016 to 2020). This unexpected situation is partly because decreases in $PM_{2.5}$ concentrations cannot entirely offset the health impact of population aging (Yue et al. 2020).

Our analysis is at the city level, and does not further focus on urban–rural difference. Most monitoring stations are located in urban regions with a high pollution level (Gao et al. 2020), and SR has limited coverage for a rural population with low pollution exposure. This situation may result in a larger difference between SR-based PWM $PM_{2.5}$ and full-coverage PWM $PM_{2.5}$ for rural regions compared to that for urban regions. The large difference of PWM $PM_{2.5}$, combined with previous findings, shows that rural residents may face a higher air pollution–related health risk (Chen et al. 2021; Garcia et al. 2016; S. Zhao et al. 2021a, b), and may finally contribute to relatively higher uncertainty in rural health assessment by using SR-based PWM $PM_{2.5}$. We plan to explore these urban–rural differences in future work.

Conclusions

In this study, by using the multi-year daily satellite-derived $PM_{2.5}$ data with 1 km spatial resolution, we examined the spatial representativeness (SR) of 213 $PM_{2.5}$ monitoring stations in Yangtze River Delta (YRD). Based on these SR estimates, annual population-weighted mean (PWM) $PM_{2.5}$ and deaths attributable to $PM_{2.5}$ exposure were also analyzed for each city in YRD.

The SR areas of 213 stations totally account for 32.33% of the area of YRD, and the SR size varies greatly with stations. These stations with SR size higher than 1000 km^2 are mainly located in the north-central area of YRD, while SR sizes are typically low and less than 200 km^2 in coastal areas. In addition, the spatial distribution of stations is reasonable for most cities from a redundancy perspective.

The SR areas of all monitoring stations can totally cover 62.16% of the population in YRD. The population ratios of the SR area to the city area are less than 50% for about half of the cities in YRD, and most of these cities are located in Anhui and Zhejiang provinces. Partly due to the fact that most stations are located in urban areas with a high pollution level, the city-level PWA $PM_{2.5}$ estimate based on SR is nearly always larger than full-coverage PWA $PM_{2.5}$, and this difference tends to decrease with increasing population ratio of the SR area.

Attributable deaths by using SR-based PWM $PM_{2.5}$ are 182,009 (95% CI: 136,632–225,081) for the entire YRD. Although this estimate only increases by 2.80% overall compared to that by using full-coverage PWM $PM_{2.5}$, the difference is more than 6% in five cities, where the population ratio of SR is less than 20%. These results suggest that more monitoring stations should be deployed in these cities for air pollution evaluation and human health assessment, especially for rural regions.

Supplementary information.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11869-022-01202-2.

Funding This work was supported by the National Key Research and Development Program of China (No. 2018YFC1507701), Jiangsu Provincial Double-Innovation Doctor Program (No. JSSCBS20211072), and Nantong University Scientific Research Foundation for the Introduced Talents (No. 135419629079).

Data availability Satellite-derived $PM_{2.5}$ data are available from the China-High-Air-Pollutants (CHAP) data set (https://weijing-rs.github.io/product.html). Population data were obtained from the Gridded Population of the World, Version 4 (available at https://sedac.ciesin.columbia.edu/data/collection/gpw-v4). Data on the age structure as well as the age specific and disease-specific mortality were obtained from the Global Burden of Disease Study 2019 dataset (https://vizhub.healthdata.org/gbd-compare).

Declarations

Consent for publication All authors approved the manuscript for publication.

Conflict of interest The authors declare no competing interests.

References

Bai H, Gao W, Zhang Y, Wang L (2022) Assessment of health benefit of PM2.5 reduction during COVID-19 lockdown in China and separating contributions from anthropogenic emissions and meteorology. J Environ Sci 115:422–431. https://doi.org/10.1016/j.jes.2021.01.022

Bai H, Zheng Z, Zhang Y, Huang H, Wang L (2021) Comparison of satellite-based PM2.5 estimation from aerosol optical depth and top-of-atmosphere reflectance. Aerosol and Air Quality Research 21:1–17. https://doi.org/10.4209/aaqr.2020.05.0257

Baró R, Jiménez-Guerrero P, Balzarini A, Curci G, Forkel R, Grell G, Hirtl M, Honzak L, Langer M, Pérez JL, Pirovano G, San José R, Tuccella P, Werhahn J, Žabkar R (2015) Sensitivity analysis of the microphysics scheme in WRF-Chem contributions to AQMEII phase 2. Atmos Environ 115:620–629. https://doi.org/10.1016/J.ATMOSENV.2015.01.047

Barrero MA, Orza JAG, Cabello M, Cantón L (2015) Categorisation of air quality monitoring stations by evaluation of PM10 variability. Sci Total Environ 524–525:225–236. https://doi.org/10.1016/j.scitotenv.2015.03.138

Blanchard CL, Tanenbaum S, Hidy GM (2014) Spatial and temporal variability of air pollution in Birmingham, Alabama. Atmos Environ 89:382–391. https://doi.org/10.1016/j.atmosenv.2014.01.006

Burnett RT, Chen H, Szyszkowicz M, Fann N, Hubbell B, Pope CA, Apte JS, Brauer M, Cohen A, Weichenthal S, Coggins J, Di Q, Brunekreef B, Frostad J, Lim SS, Kan H, Walker KD, Thurston GD, Hayes RB, Lim CC, Turner MC, Jerrett M, Krewski D, Gapstur SM, Diver WR, Ostro B, Goldberg D, Crouse DL, Martin RV, Peters P, Pinault L, Tjepkema M, van Donkelaar A, Villeneuve PJ, Miller AB, Yin P, Zhou M, Wang L, Janssen NAH, Marra M, Atkinson RW, Tsang H, Quoc Thach T, Cannon JB, Allen RT, Hart JE, Laden F, Cesaroni G, Forastiere F, Weinmayr G, Jaensch A, Nagel G, Concin H, Spadaro JV (2018) Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. Proc Natl Acad Sci 115:9592–9597. https://doi.org/10.1073/pnas.1803222115

Cao L, Li S, Sun L (2021) Study of different Carbon Bond 6 (CB6) mechanisms by using a concentration sensitivity analysis. Atmos Chem Phys 21:12687–12714. https://doi.org/10.5194/acp-21-12687-2021

Chen K, Zhou L, Chen X, Bi J, Kinney PL (2017) Acute effect of ozone exposure on daily mortality in seven cities of Jiangsu Province, China: no clear evidence for threshold. Environ Res 155:235–241. https://doi.org/10.1016/j.envres.2017.02.009

Chen Yun, Chen R, Chen Yue, Dong X, Zhu J, Liu C, van Donkelaar A, Martin R. V, Li H, Kan H, Jiang Q, Fu C 2021. The prospective effects of long-term exposure to ambient PM2.5 and constituents on mortality in rural East China. Chemosphere 280, 130740. https://doi.org/10.1016/j.chemosphere.2021.130740

Ding A, Wang T, Zhao M, Wang T, Li Z (2004) Simulation of sea-land breezes and a discussion of their implications on the transport of air pollution during a multi-day ozone episode in the Pearl River Delta of China. Atmos Environ 38:6737–6750. https://doi.org/10.1016/j.atmosenv.2004.09.017

Dominici F, Peng RD, Bell ML, Pham L, McDermott A, Zeger SL, Samet JM (2006) Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. JAMA 295:1127. https://doi.org/10.1001/jama.295.10.1127

Doxsey-Whitfield E, MacManus K, Adamo SB, Pistolesi L, Squires J, Borkovska O, Baptista SR (2015) Taking advantage of the improved availability of census data: a first look at the gridded population of the world, version 4. Papers in Applied Geography 1:226–234. https://doi.org/10.1080/23754931.2015.1014272

Elbern H, Strunk A, Schmidt H, Talagrand O (2007) Emission rate and chemical state estimation by 4-dimensional variational inversion. Atmos Chem Phys 7:3749–3769. https://doi.org/10.5194/acp-7-3749-2007

Gao L, Yue X, Meng X, Du L, Lei Y, Tian C, Qiu L (2020) Comparison of ozone and PM2.5 concentrations over urban, suburban, and background sites in China. Adv Atmos Sci 37:1297–1309. https://doi.org/10.1007/s00376-020-0054-2

Garcia CA, Yap P, Park H, Weller BL (2016) Association of long-term PM2.5 exposure with mortality using different air pollution exposure models: impacts in rural and urban California. Int J Environ Health Res 26:145–157. https://doi.org/10.1080/09603123.2015.1061113

GBD (2019) Risk Factors Collaborators, 2020. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet 396:1223–1249. https://doi.org/10.1016/S0140-6736(20)30752-2

Geng G, Zhang Q, Martin RV, van Donkelaar A, Huo H, Che H, Lin J, He K (2015) Estimating long-term PM2.5 concentrations in China using satellite-based aerosol optical depth and a chemical transport model. Remote Sens Environ 166:262–270. https://doi.org/10.1016/j.rse.2015.05.016

Han W, Tong L, Chen Y, Li R, Yan B, Liu X 2018. Estimation of high-resolution daily ground-level PM2.5 concentration in Beijing 2013–2017 using 1 km MAIAC AOT data. Applied Sciences 8, 2624. https://doi.org/10.3390/app8122624

He Q, Geng F, Li C, Yang S, Wang Y, Mu H, Zhou G, Liu X, Gao W, Cheng T, Wu Z (2018) Long-term characteristics of satellite-based PM2.5 over East China. Sci Total Environ 612:1417–1423. https://doi.org/10.1016/j.scitotenv.2017.09.027

He Q, Li C, Geng F, Zhou G, Gao W, Yu W, Li Z, Du M (2016) A parameterization scheme of aerosol vertical distribution for surface-level visibility retrieval from satellite remote sensing. Remote Sens Environ 181:1–13. https://doi.org/10.1016/j.rse.2016.03.016

Hu C, Liu S, Wang Y, Zhang M, Xiao W, Wang W, Xu J (2018) Anthropogenic CO2 emissions from a megacity in the Yangtze River Delta of China. Environ Sci Pollut Res 25:23157–23169. https://doi.org/10.1007/s11356-018-2325-3

Janssen S, Dumont G, Fierens F, Deutsch F, Maiheu B, Celis D, Trimpeneers E, Mensink C (2012) Land use to characterize spatial representativeness of air quality monitoring stations and its relevance for model validation. Atmos Environ 59:492–500. https://doi.org/10.1016/J.ATMOSENV.2012.05.028

Kikuchi M, Murakami H, Suzuki K, Nagao TM, Higurashi A (2018) Improved hourly estimates of aerosol optical thickness using spatiotemporal variability derived from Himawari-8 geostationary satellite. IEEE Trans Geosci Remote Sens 56:3442–3455. https://doi.org/10.1109/TGRS.2018.2800060

Liu C, Henderson BH, Wang D, Yang X, Peng Z (2016) A land use regression application into assessing spatial variation of intra-urban fine particulate matter (PM2.5) and nitrogen dioxide (NO2) concentrations in City of Shanghai. China Science of the Total Environment 565:607–615. https://doi.org/10.1016/j.scitotenv.2016.03.189

Lv B, Liu Y, Yu P, Zhang B, Bai Y (2015) Characterizations of PM2.5 pollution pathways and sources analysis in four large cities in China. Aerosol and Air Quality Research 15:1836–1843. https://doi.org/10.4209/aaqr.2015.04.0266

Maji KJ 2020. Substantial changes in PM2.5 pollution and corresponding premature deaths across China during 2015–2019: a model prospective. Science of The Total Environment 729, 138838. https://doi.org/10.1016/j.scitotenv.2020.138838

Martin F, Fileni L, Palomino I, Vivanco MG, Garrido JL (2014) Analysis of the spatial representativeness of rural background monitoring stations in Spain. Atmos Pollut Res 5:779–788. https://doi.org/10.5094/APR.2014.087

MEPC (Ministry of Environment Protection of China), 2013. Technical regulation for selection of ambient air quality monitoring stations (on Trial). HJ 664–2013.

McDuffie EE, Martin RV, Spadaro J V, Burnett R, Smith SJ, O'Rourke P, Hammer MS, van Donkelaar A, Bindle L, Shah V, Jaeglé L, Luo G, Yu F, Adeniran JA, Lin J, Brauer M 2021. Source sector and fuel contributions to ambient PM2.5 and attributable mortality across multiple spatial scales. Nature Communications 12, 3594. https://doi.org/10.1038/s41467-021-23853-y

Park Y, Kwon B, Heo J, Hu X, Liu Y, Moon T 2020. Estimating PM2.5 concentration of the conterminous United States via interpretable convolutional neural networks. Environmental Pollution 256, 113395. https://doi.org/10.1016/j.envpol.2019.113395

Piersanti A, Vitali L, Righini G, Cremona G, Ciancarella L (2015) Spatial representativeness of air quality monitoring stations: a grid model based approach. Atmos Pollut Res 6:953–960. https://doi.org/10.1016/j.apr.2015.04.005

Righini G, Cappelletti A, Ciucci A, Cremona G, Piersanti A, Vitali L, Ciancarella L (2014) GIS based assessment of the spatial representativeness of air quality monitoring stations using pollutant emissions data. Atmos Environ 97:121–129. https://doi.org/10.1016/j.atmosenv.2014.08.015

Russo A, Gouveia C, Levy I, Dayan U, Jerez S, Mendes M, Trigo R (2016) Coastal recirculation potential affecting air pollutants in Portugal: the role of circulation weather types. Atmos Environ 135:9–19. https://doi.org/10.1016/j.atmosenv.2016.03.039

Santiago JL, Martín F, Martilli A (2013) A computational fluid dynamic modelling approach to assess the representativeness of urban monitoring stations. Sci Total Environ 454–455:61–72. https://doi.org/10.1016/J.SCITOTENV.2013.02.068

Schneider C, Kessler C, Moussiopoulos N (1997) Influence of emission input data on ozone level predictions for the Upper Rhine Valley. Atmos Environ 31:3187–3205. https://doi.org/10.1016/S1352-2310(97)00062-9

Shi X, Zhao C, Jiang JH, Wang C, Yang X, Yung YL (2018) Spatial representativeness of PM 2.5 concentrations obtained using observations from network stations. Journal of Geophysical Research: Atmospheres 123:3145–3158. https://doi.org/10.1002/2017JD027913

Song C, He J, Wu L, Jin T, Chen X, Li R, Ren P, Zhang L, Mao H (2017) Health burden attributable to ambient PM2.5 in China. Environ Pollut 223:575–586. https://doi.org/10.1016/j.envpol.2017.01.060

van Donkelaar A, Martin RV, Brauer M, Hsu NC, Kahn RA, Levy RC, Lyapustin A, Sayer AM, Winker DM (2016) Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. Environ Sci Technol 50:3762–3772. https://doi.org/10.1021/acs.est.5b05833

Vardoulakis S, Gonzalez-Flesca N, Fisher BEA, Pericleous K (2005) Spatial variability of air pollution in the vicinity of a permanent monitoring station in central Paris. Atmos Environ 39:2725–2736. https://doi.org/10.1016/J.ATMOSENV.2004.05.067

Wei J, Li Z, Cribb M, Huang W, Xue W, Sun L, Guo J, Peng Y, Li J, Lyapustin A, Liu L, Wu H, Song Y (2020) Improved 1 km resolution PM2.5 estimates across China using enhanced space–time

extremely randomized trees. Atmos Chem Phys 20:3273–3289. https://doi.org/10.5194/acp-20-3273-2020

Wei J, Li Z, Lyapustin A, Sun L, Peng Y, Xue W, Su T, Cribb M 2021. Reconstructing 1-km-resolution high-quality PM2.5 data records from 2000 to 2018 in China: spatiotemporal variations and policy implications. Remote Sensing of Environment 252, 112136. https://doi.org/10.1016/j.rse.2020.112136

Yue H, He C, Huang Q, Yin D, Bryan BA, 2020. Stronger policy required to substantially reduce deaths from PM2.5 pollution in China. Nature Communications 11, 1462. https://doi.org/10.1038/s41467-020-15319-4

Zhao H, Chen K, Liu Z, Zhang Y, Shao T, Zhang H 2021a. Coordinated control of PM2.5 and O3 is urgently needed in China after implementation of the "Air pollution prevention and control action plan." Chemosphere 270, 129441. https://doi.org/10.1016/j.chemosphere.2020.129441

Zhao S, Liu S, Hou X, Sun Y, Beazley R (2021) Air pollution and cause-specific mortality: a comparative study of urban and rural areas in China. Chemosphere 262:127884. https://doi.org/10.1016/j.chemosphere.2020.127884

Zoroufchi Benis K, Fatehifar E (2015) Optimal design of air quality monitoring network around an oil refinery plant: a holistic approach. Int J Environ Sci Technol 12:1331–1342. https://doi.org/10.1007/s13762-014-0723-8

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.