AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2017JD027913

Key Points:
- A new method is developed to determine the spatial representation area of $PM_{2.5}$ at single station using spatial variability and correlation
- The $PM_{2.5}$ representation area for single station ranges from 0.25 to 16.25 km^2, but less than 3 km^2 for more than half of the stations
- $PM_{2.5}$ observation at 10 optimal stations would have a good representation of those obtained from 169 stations for the study period

Correspondence to:
C. Zhao,
czhao@bnu.edu.cn

Citation:
Shi, X., Zhao, C., Jiang, J. H., Wang, C., Yang, X., & Yung, Y. L. (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

Received 18 OCT 2017
Accepted 27 FEB 2018
Accepted article online 5 MAR 2018
Published online 25 MAR 2018

Spatial Representativeness of $PM_{2.5}$ Concentrations Obtained Using Observations From Network Stations

Xiaoqin Shi[1], Chuanfeng Zhao[1,2], Jonathan H. Jiang[3], Chunying Wang[4], Xin Yang[1], and Yuk L. Yung[2]

[1]State Key Laboratory of Earth Surface Processes and Resource Ecology, and College of Global Change and Earth System Science, Beijing Normal University, Beijing, China, [2]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA, [3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, [4]Hebei Sailhero Environmental Protection Hi-tech., Ltd, Shijiazhuang, China

Abstract Haze has been a focused air pollution phenomenon in China, and its characterization is highly desired. Aerosol properties obtained from a single station are frequently used to represent the haze condition over a large domain, such as tens of kilometers, which could result in high uncertainties due to their spatial variation. Using a high-resolution network observation over an urban city in North China from November 2015 to February 2016, this study examines the spatial representativeness of ground station observations of particulate matter with diameters less than 2.5 μm ($PM_{2.5}$). We developed a new method to determine the representative area of $PM_{2.5}$ measurements from limited stations. The key idea is to determine the $PM_{2.5}$ spatial representative area using its spatial variability and temporal correlation. We also determine stations with large representative area using two grid networks with different resolutions. Based on the high spatial resolution measurements, the representative area of $PM_{2.5}$ at one station can be determined from the grids with high correlations and small differences of $PM_{2.5}$. The representative area for a single station in the study period ranges from 0.25 to 16.25 km^2 but is less than 3 km^2 for more than half of the stations. The representative area varies with locations, and observation at 10 optimal stations would have a good representativeness of those obtained from 169 stations for the 4 month time scale studied. Both evaluations with an empirical orthogonal function analysis and with independent data set corroborate the validity of the results found in this study.

1. Introduction

Haze pollution is atmospheric pollution phenomenon with large fine aerosol amount. On 8 December 2015, Beijing released the first haze pollution of red warning in history and more red warnings have been released from then. Heavy haze pollution has affected many aspects of our society, such as weather and climate, public health, economics, and social activities (Peters et al., 1997a, 1997b; Pope III et al., 2002; Schwartz & Neas, 2000; Q. Wang et al., 2013; Yin & Chen, 2007; Zhao & Garrett, 2015). For example, aerosol influences cloud properties, alters the radiation budget of the earth-atmosphere system, affects atmospheric circulation patterns, and causes changes in surface temperature and precipitation (Kaufman et al., 2002). At the same time, heavy haze pollution has also aroused widespread concerns from the country and even the world. Significant efforts have been made to monitor, predict, and control the haze pollution.

Satellite remote sensing has been widely used for the estimation of the mass concentration of ground particulate matter with diameters less than 2.5 μm ($PM_{2.5}$) (Kaufman & Fraser, 1983; Van Donkelaar et al., 2006, 2010, 2012; Zheng et al., 2017). Since the aerosol optical depth (AOD) measured from satellite reflects the integrated amount of aerosol particles in the vertical column with a relatively large horizontal area (1 km^2 or even larger), and the relationship between AOD and $PM_{2.5}$ often varies a lot with locations (Ma et al., 2014; Ramachandran, 2005; J. Wang & Christopher, 2003; Zhang et al., 2009), the derived relationship between AOD from satellite and surface $PM_{2.5}$ may have large uncertainties. Recently, Zheng et al. (2017) have investigated the various influential factors that can affect the relationship between AOD and $PM_{2.5}$, which include the different domain representativeness of AOD.

Nowadays, the high-density observations of China's ground hourly data allow us to understand the regional distribution of air pollutants by ground observation alone (Rohde & Muller, 2015). Several ground-based observation networks are currently operated for aerosol measurements, including the Aerosol Robotic

Network (Holben et al., 2001), the Interagency Monitoring of Protected Visual Environment network (Malm et al., 1994), and about 2,400 environmental observation stations in China. While the measurements from these stations are well calibrated and widely used for evaluation of satellite aerosol observations (Bréon et al., 2011; Chu et al., 2003; Engel-Cox et al., 2004; Levy et al., 2010; Li et al., 2015), these stations still have limited representativeness in space and have a relatively large distance from each other. In practice, the aerosol observation is often used to represent the pollution over a large area and used to compare with the satellite observations within a domain with a radius like 1 km or 5 km. However, when the regional change of the local emission sources is large, the limited number of city monitoring stations are unable to accurately provide the air quality information for the whole city.

With the continuous expansion of large cities, the internal structure of the city is more and more complex. Urban winds near surface and inversions of atmospheric temperature within the boundary layer make it difficult to disperse (Y. Wang et al., 2014). The aerosol pollution over a location is the sum of both local emission sources and long-range transport (Garrett et al., 2010; Sun et al., 2014; Zhao et al., 2009). When the pollution is highly related to local emission sources, the spatial variation of air pollution is large and observations from limited stations are not reliable to represent the air quality over a large domain. Scaperdas and Colvile (1999) studied the representative problems of the environmental monitoring stations at the crossroads in central London, where they argued that the different layouts of high-rise buildings and streets in cities had a significant impact on the diffusion of microscale emissions (such as traffic emissions). Thus, it is very important to understand the temporal and spatial distribution characteristics of $PM_{2.5}$ and its representativeness in large cities.

During the last 2 or 3 years, several cities in North China have established high spatial resolution $PM_{2.5}$ observation network, which includes hundreds of monitoring stations in a single city. Using these observations, this study first examines the spatial representativeness of $PM_{2.5}$ observed at a single station or a limited number of stations by exploring their spatial heterogeneities. We then develop a method to determine the representative area of measurements from a network of limited stations and validate the accuracy of this method using the higher density monitoring stations.

The paper is organized as follows. Section 2 describes the measurements and the method used in this study. Section 3 shows the analysis and results. Section 4 gives a summary and discussion.

2. Data and Method

2.1. Data

We use a network observation of air pollution within a city in North China, for the region with latitudes from 37.95°N to 38.15°N and longitude from 114.40°E to 114.65°E. The instrument used for the measurement is called XHAQSN-808, manufactured by the Sailhero cooperation. It includes several modes to measure different variables, including $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, and O_3. Here we briefly describe the mode used for measuring $PM_{2.5}$. It takes use of the β ray absorption method, with a measurable range set as 0–2000 $\mu g/m^3$. Observation is done at every hour. The hourly minimum detection amount is 5 $\mu g/m^3$, and the resolution/accuracy is 0.01 $\mu g/m^3$.

There are totally 169 stations in the study area, and then distribution is shown in Figure 1. The area with latitudes less than 38.06°N is the most densely populated inner city region with most observation stations there. The domain size shown in Figure 1 is 21.89 km (zonal) × 22.24 km (poleward). The data period examined in this study is from 2 November 2015 to 29 February 2016.

We have made further data quality control to the network observations. First, for the observation data with value 0, they are set as missing. Second, if there are repeated values for no less than 3 times, it is likely that the instrument has stopped working and the reporting system is simply repeating the last valid measurement received (Rohde & Muller, 2015), and the data are also set as missing. Third, if there are only one or two continually missing data, the $PM_{2.5}$ value at the missing data time will be filled using a linear interpolation. Fourth, we remove the unreasonable extreme values observed from ground stations using a variability check, which is described as follows. For every station at a given time, we compare the $PM_{2.5}$ observation (x) with the average $PM_{2.5}$ observation (y) from all stations within a 1 km circular domain around the station. If $x > 2y$ or $x < y/2$, the observations are assumed as extreme data and discarded. This method is similar as that used by Rohde and Muller (2015).

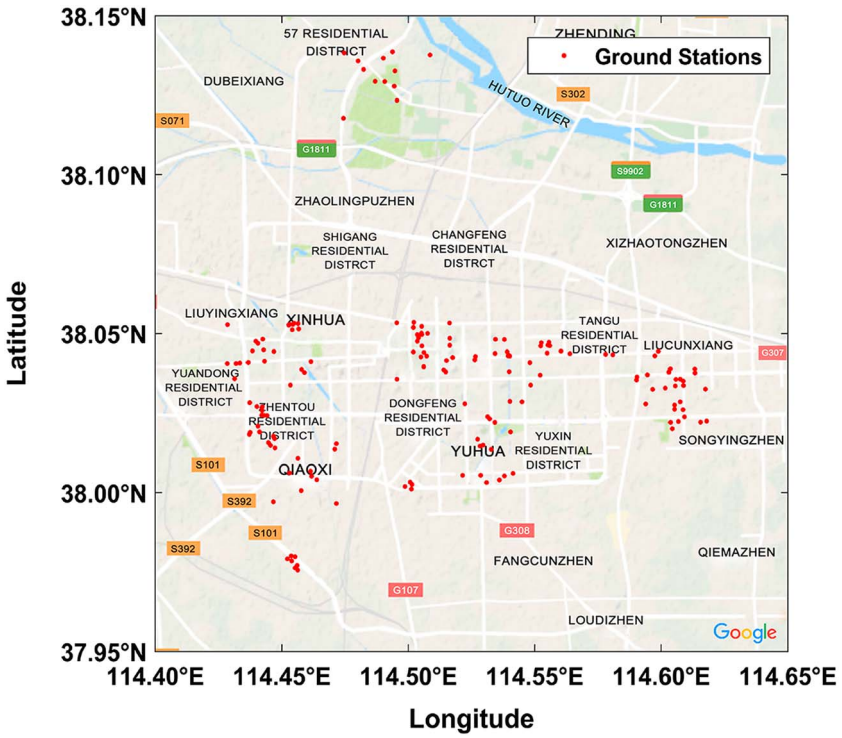

Figure 1. The 169 ground observation stations which lie within a domain with latitudes from 37.95°N to 38.15°N and longitude from 114.40°E to 114.65°E. The map is from google map.

2.2. Method

First, for the considered data set, we introduce empirical variogram to estimate the magnitude of the spatial heterogeneity over different region size. For observations z_i ($i = 1, \ldots, k$) at locations $x_1, x_2, \ldots, x_k$, the empirical variogram $\gamma(h)$ is defined as (Cressie, 1992)

$$\gamma(h) = \frac{1}{2|N(h)|}\Sigma_{(i,j)\in N(h)}|z_i - z_j|^2 \quad (1)$$

where $N(h)$ denotes the set of pairs of observations i, j, such that $|x_i - x_j| = h$, and $|N(h)|$ is the number of pairs in the set. h is often denoted as lag distance. The empirical variogram $\gamma(h)$ varies with the distance h among stations. Generally, but not always, $\gamma(h)$ increases with h. Based on the variation of $\gamma(h)$ with h, we can estimate a distance R with relatively weak change of $\gamma(h)$ to h for $h \leq R$ and large increase of $\gamma(h)$ to h for $h > R$. The distance R varies with the time scale considered. Thus, R is determined as a minimum of values with different time scales. The spatial representativeness of $PM_{2.5}$ obtained from a single station is determined based on the $PM_{2.5}$ spatial variability over the study period.

Second, we divide the whole study area into $R \times R$ grids. For each grid which includes stations no less than 3, we analyze the coefficient of variation (CV) among the stations for their time-averaged $PM_{2.5}$ concentration, which is defined as,

$$CV = {}^{Std}/_{Mean} \quad (2)$$

where the Std and Mean are the standard deviation and mean of the time-averaged $PM_{2.5}$ concentration among the stations within a grid, respectively. Equation (2) implies that CV can represent the relative variability of $PM_{2.5}$ within a grid. We need set up a CV threshold value to classify low and high spatial variability of $PM_{2.5}$ for a grid. The CV threshold value is defined here as the ratio between the standard deviation and the mean of all measurements for the study region, which is ~0.15. When CV is less than the threshold value of 0.15 or the spatial variability is less than 15% of the mean within a grid, the grid is identified as a Grid with nearly Uniform $PM_{2.5}$ (GUP) concentration. Generally, the spatial representativeness of $PM_{2.5}$ observed at GUP might be larger than that at other grids and will be investigated and quantified in next few steps.

We carry out the analysis of second step using all $PM_{2.5}$ observations, using the $PM_{2.5}$ observations at different pollution levels, such as 0–35 $\mu g/m^3$, 35–75 $\mu g/m^3$, 75–115 $\mu g/m^3$, 115–150 $\mu g/m^3$, 150–250 $\mu g/m^3$, 250–500 $\mu g/m^3$, and >500 $\mu g/m^3$. The air pollution level is classified based on the technical regulation on ambient quality index (HJ633–2012) from the Ministry of Environmental Protection of the People's Republic of China (2012).

Third, for each $R \times R$ grid, we divide it into $r \times r$ ($r < R$) subgrids with a higher spatial resolution. Each subgrid generally contains a very small number of stations, such as less than five. We then make regional average within each subgrid to obtain a time series of averaged $PM_{2.5}$. For each subgrid, we can define an index of correlation (IC), which is

$$\mathrm{IC} = \frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{d_i} \tag{3}$$

where r_i and d_i are the correlation coefficient and distance between a given subgrid and others within a distance R of the given subgrid, respectively; n is the number of subgrids within a distance R to the given subgrid ($n \geq 3$). For these subgrids, the higher the IC value, the better the spatial representativeness. Those subgrids within GUP or with an IC value higher than a threshold are denoted as the sub-Grids with spatial Representativeness to be Determined (GRD). The spatial representativeness of GRD is most likely larger than that of other subgrids.

Fourth, for a GRD (A), we examine every subgrid (B) within the study area to figure out if the $PM_{2.5}$ measured at A can represent that measured at B. We believe that high correlation of time series and small time-averaged difference are the representative criteria. So we introduce two indicators, the correlation (R) between the time series of $PM_{2.5}$ observations at A and B, and the differences (D) in time-averaged $PM_{2.5}$ mass concentration at A and B. When R is high and D is small, $PM_{2.5}$ observations at A will have a good representativeness to that measured at B. In this study, we adopted the threshold values of 0.9 and 30 $\mu g/m^3$ for R and D, respectively. $R > 0.9$ is adopted to make sure the high correlation, and 30 $\mu g/m^3$ is actually about 15% of the averaged $PM_{2.5}$ mass concentration during study period which can ensure the weak spatial variation. So, when $R > 0.9$ and $D < 30$ $\mu g/m^3$, the $PM_{2.5}$ mass concentration of subgrid B could be represented by that at subgrid A. We should note that the threshold values for R and D could be subjective and affect the regional representativeness we obtained. When more rigorous (loose) threshold values are adopted, there will be less (more) GRDs identified with a good representativeness.

The subgrid A is also called sub-Grid of Representativeness (GR), and all subgrids that A can represent form a subgrid collection. Considering that the subgrid collection of every GR is often discontinuous in space, we need to add some subgrids into the collection to make them more continuous in space, which is somewhat subjective and could introduce extra uncertainties to our study. The uncertainties caused by the filling could vary by cases and are not discussed quantitatively here. The total subgrid collection obtained is the representative area of the examined GR or a station in GR.

After knowing the spatial representativeness of GRs, we can further estimate the minimum number of GRs that can represent the study area as large as possible by analyzing the overlapping of the domains that all GRs represent. Here is a rule: if the representative domain of GR_A contains that of GR_B, GR_B will be removed and only GR_A will remain. By subtracting some GRs whose representative areas can be represented by other GRs, we can find the minimum number of GRs that can represent the study area as large as possible. These GRs are denoted as the Important Grids of Representativeness (IGR) in this study.

3. Analysis and Results

3.1. Spatial Heterogeneity

Figure 2 shows the time series boxplots of the medians of daily average $PM_{2.5}$ mass concentration among all stations in the study area from 2 November 2015 to 29 February 2016. We can see clearly the distribution range of daily averaged $PM_{2.5}$ mass concentration in the whole study area. Obviously, the serious pollution period with large temporal (orange line) and spatial (gray lines) variations of $PM_{2.5}$ mass concentration is from 26 November 2015 to 5 January 2016, during which the maximum $PM_{2.5}$ mass concentration value for all of the medians among all stations is more than 600 $\mu g/m^3$. By contrast, there are two relatively clean air periods

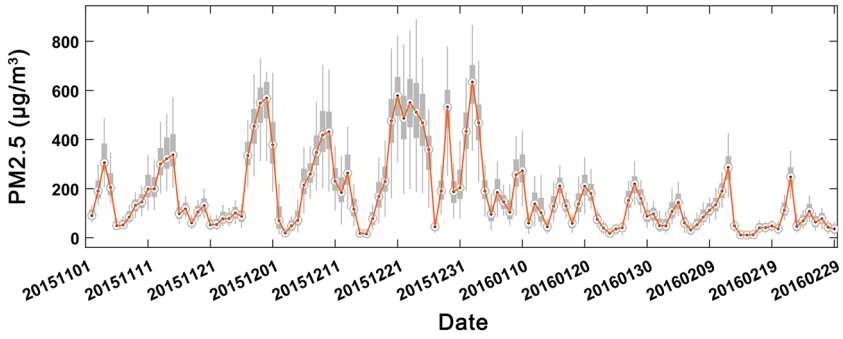

Figure 2. The time series boxplot of the medians of daily averaged $PM_{2.5}$ mass concentration in the study area from 2 November 2015 to 29 February 2016. The gray boxes represent the upper and lower quartile data range, and gray lines represent the data range of the most extreme values that are considered to be not outliers that exceed 1.5 times interquartile range from the upper and lower quartile.

with more than 7 days each, which are from 16 to 25 November 2015 and from 12 to 20 February 2016. For these two clean air periods, the median values of daily averaged $PM_{2.5}$ mass concentration among all stations are generally less than 100 $\mu g/m^3$.

Figure 2 shows approximately periodic variations of $PM_{2.5}$ mass concentrations. By defining a $PM_{2.5}$ pollution event as that $PM_{2.5}$ increase from low to high followed by a decrease from high to low values, we can find more than 10 $PM_{2.5}$ pollution events in the study period. Except for the two relatively clean air periods, a $PM_{2.5}$ pollution event develops very fast in no more than 3 days from a clean start. The dissipation of the pollution event is even faster, which is strongly related to the meteorological conditions such as strong wind speed. Among the $PM_{2.5}$ pollution events, there are serious pollution periods from the end of November 2015 to early January 2016. For these heavy pollution cases, there are strong variations of $PM_{2.5}$ mass concentration in both time and space, and the regional representativeness of $PM_{2.5}$ observations at a single station might be smaller during these periods compared to others.

The spatial distribution of $PM_{2.5}$ varies with the time scale considered. We first examine the spatial distribution of $PM_{2.5}$ on a monthly time scale. Figure 3 shows the spatial distributions of monthly averaged $PM_{2.5}$ along with their variogram calculated by equation (1) for the study area in November–February. The monthly region-averaged $PM_{2.5}$ mass concentrations are 195 $\mu g/m^3$, 268 $\mu g/m^3$, 168 $\mu g/m^3$, and 85 $\mu g/m^3$ in November–February, respectively. Spatially, the $PM_{2.5}$ mass concentration is in general slightly larger in north region (>38.1°N) and southeast edge. The variograms shows the similar trend with the lag distance (h) for all months: the variogram is relatively small and varies slowly with the lag distance when $h < 4$ km and is large and varies quickly with the lag distance (h) when $h > 4$ km. Figure 4 further shows the spatial distributions of daily averaged $PM_{2.5}$ and hourly $PM_{2.5}$ along with their variograms. For the daily time scale analyses (Figures 4a and 4b), we chose a relatively clean day on 21 November and a polluted day on 21 December with region-average $PM_{2.5}$ of 57 $\mu g/m^3$ and 566 $\mu g/m^3$, respectively. The variograms show the different change with lag distance (h) as that found for monthly analyses in Figure 3. On the clean day, the variogram changes drastically with lag distance compared with the polluted day. For the hourly observation analyses (Figures 4c and 4d), we choose the noontime (12:00 Beijing time) of the clean and polluted days shown in Figures 4a and 4b. The change of variogram with lag distance (h) on the clean (polluted) day noontime is similar to (larger than) that from daily time scale analyses. Roughly, the variogram has weak change with lag distance $h < 2$ km in all cases except for clean days. Considering this, we next analyze the spatial variation of $PM_{2.5}$ observations within 2 km × 2 km grids.

3.2. Determination of Spatial Representativeness

3.2.1. The Spatial Distribution of $PM_{2.5}$

The CV analysis is for spatial variation within a given domain/grid so that we choose the domain/grid without high spatial variation as the GUP. To do this kind of CV analysis, we need make the domain/grid not too large that we can recognize relatively clear spatial variability of CV and not too small that there are enough stations (3–15 stations) within one grid for the calculation of CV. Based on the results shown in section 3.1, we classify the study area into 2 km × 2 km grids and calculate the CV value for each grid that includes no less than three

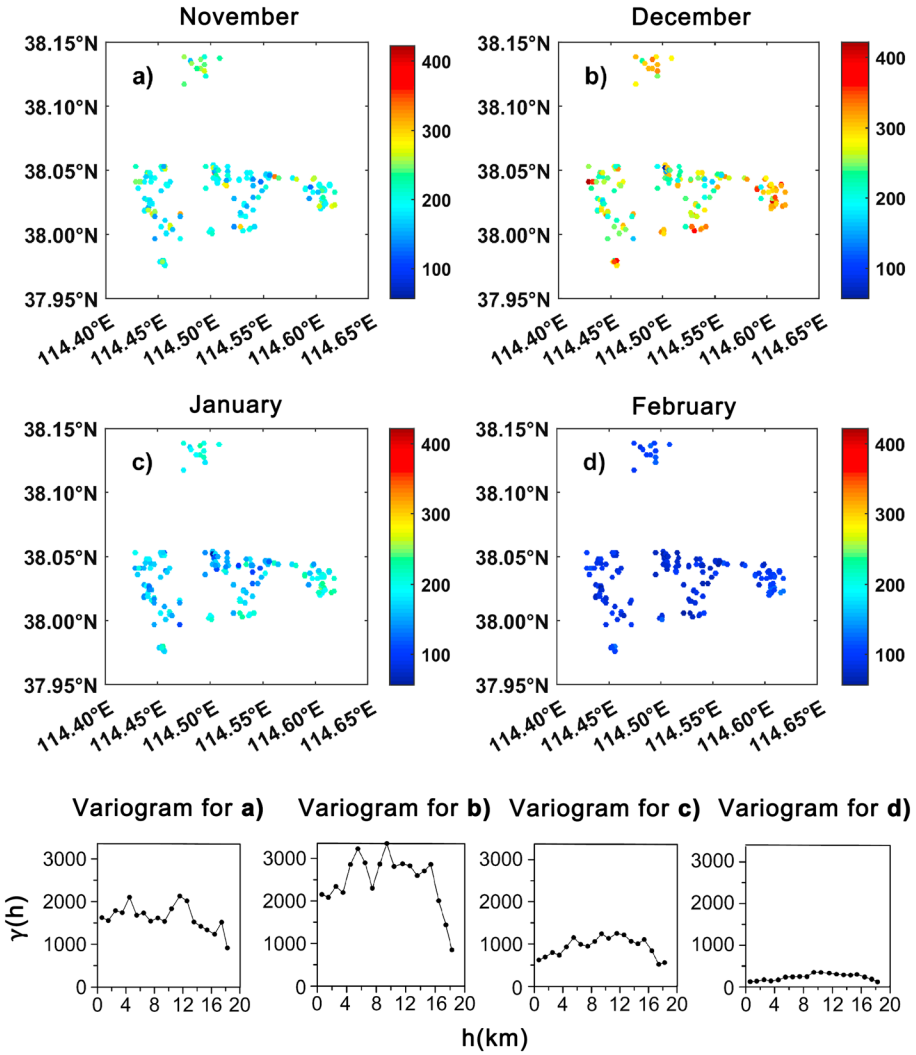

Figure 3. The spatial distribution of monthly averaged $PM_{2.5}$ along with their variograms for the study area in November (a), December (b), January (c), and February (d).

stations and at most 15 stations. According to the average value of $PM_{2.5}$ mass concentration of the whole study area, we classify the observation time into seven types with different pollution levels, which are <35 μg/m^3, 35–75 μg/m^3, 75–115 μg/m^3, 115–150 μg/m^3, 150–250 μg/m^3, 250–500 μg/m^3, and >500 μg/m^3. Figure 5 shows the spatial distributions of the grid CV values under seven different pollution conditions (Figures 5a–5g) along with that for all observation time (Figure 5h). Here we have adopted the Universal Transverse Mercator system for the study region and set the origin of coordinates at the location with longitude 114.4°E and latitude 37.95°N. The relative time frequencies for the classified seven pollution levels during the study period are 18%, 17%, 12%, 8%, 18%, 21%, and 6%, respectively.

With a threshold value of 0.15 for CV in each grid, we determine the grid with nearly uniform $PM_{2.5}$ mass concentration, that is, GUP. The GUPs, which are shown as empty triangles in Figure 5, vary with the pollution conditions. Under the condition that region-averaged $PM_{2.5}$ mass concentration is less than <35 μg/m^3, the CV value can be as large as 0.47 and there are only eight GUPs. The numbers of GUPs under other conditions are 14, 17, 21, 19, 13, and 14 with increasing $PM_{2.5}$ levels as defined above. We can see that the number of GUPs increases with $PM_{2.5}$ mass concentration when it is less than 115–150 μg/m3 and decreases when $PM_{2.5}$ mass concentration is larger. Under the condition that region-averaged $PM_{2.5}$ mass concentration is 115–150 μg/m^3, there are the most GUPs (the number is 21) which make up a proportion of 87.5% of the total effective grids in the study area, whereas the GUPs (16 grids) are about 66.7% of the effective

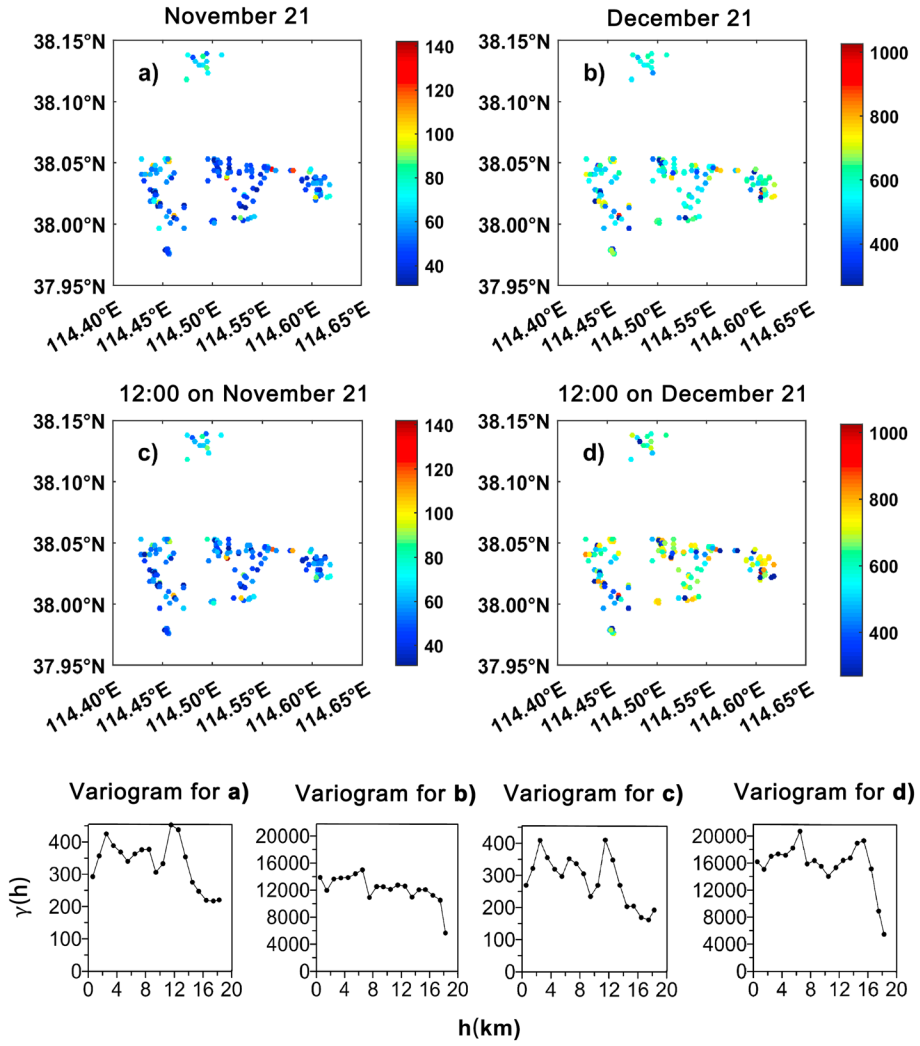

Figure 4. The spatial distribution of daily averaged $PM_{2.5}$ along with their variograms for the study area in a clean day of 21 November (a) and a polluted day of 21 December (b), and the spatial distribution of instantaneous $PM_{2.5}$ along with their variograms for the study area at 12:00 (Beijing time) on 21 November (c) and 21 December (d).

grid points for the condition with all observation time. More GUPs represents that there are more grids with uniform $PM_{2.5}$ mass concentration, implying less local emission impacts for the study area. Thus, the results shown above potentially suggests the shift from local source influence to regional background/long-range transport influence to back to local to nearby high-emitting source influence when $PM_{2.5}$ concentration increase from low to high.

3.2.2. Temporal Correlation Analysis

The major purpose of IC analysis is to figure out the stations/subgrids that have good temporal correlation of $PM_{2.5}$ with surrounding regions. Thus, the subgrids for IC analysis should be as small as possible. We could simply use the stations. Considering that most surface station observations have spatial representative area larger than 0.5 km × 0.5 km and most of these subgrids contain only 1 or 2 (maximum is 5) stations, we adopted the subgrid with a resolution of 0.5 km. For every 2 km × 2 km grid, we divide it into subgrids with higher resolution, 0.5 km × 0.5 km. For every subgrid, it contains 1 to 5 ground stations, and the averaged $PM_{2.5}$ measurements from these stations are used as the $PM_{2.5}$ mass concentrations at one subgrid. For all subgrids, we calculate their IC values using equation (3). As indicated earlier, the higher the IC value, the better the spatial representativeness. Figure 6a shows the determined IC values for all subgrids in the study area. We identify subgrids in 16 GUPs and the subgrids outside GUPs with lager IC values (>0.4817, and 0.4817 is the median of all IC values) as the GRD. Figure 6b

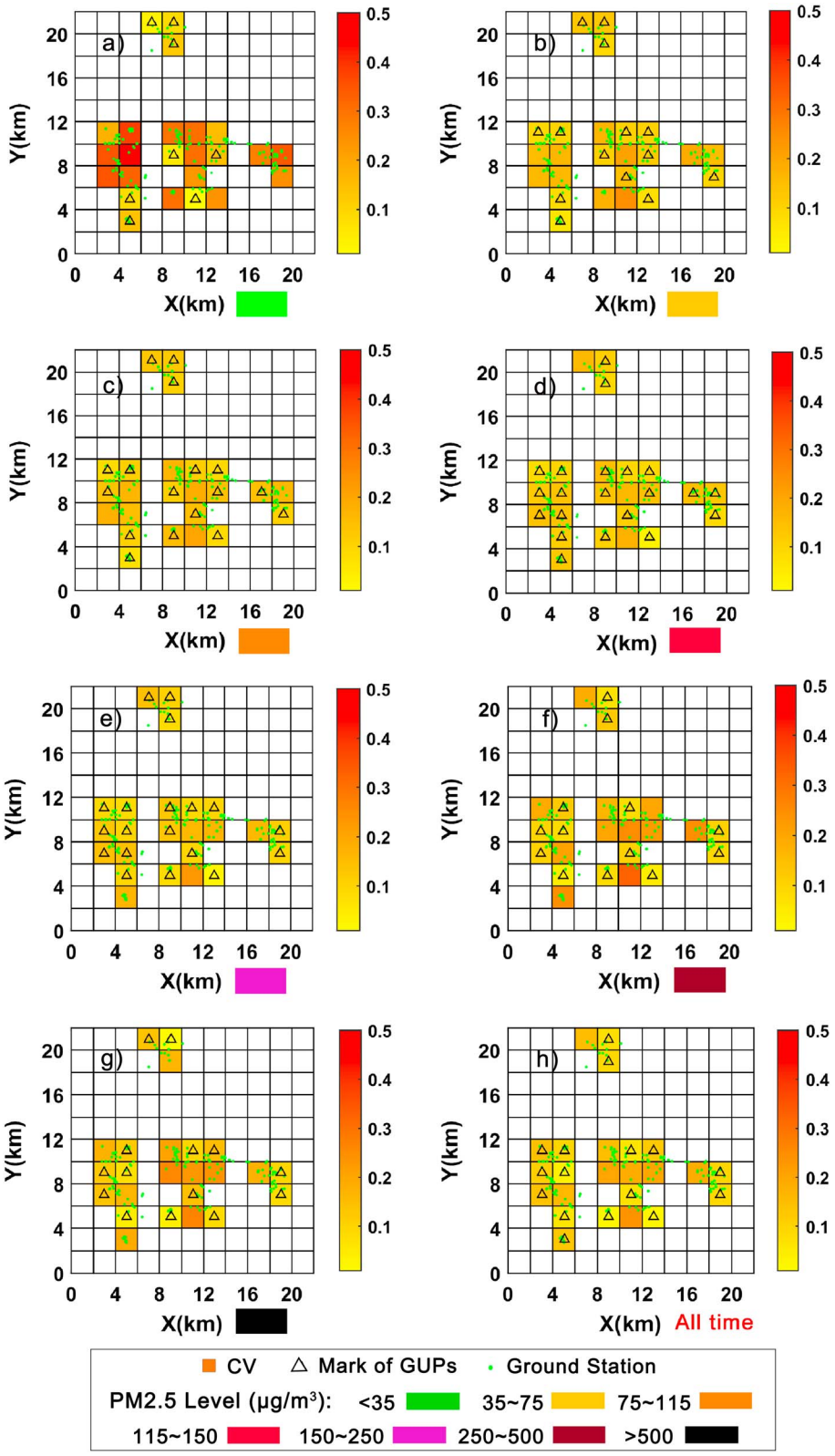

Figure 5. The spatial distribution of coefficient of variation (CV) for each grid that includes no less than 3 stations and at most 15 stations. Figures 5a–5g are for seven different pollution conditions with $PM_{2.5}$ mass concentration as <35 μg/m^3, 35–75 μg/m^3, 75–115 μg/m^3, 115–150 μg/m^3, 150–250 μg/m^3, 250–500 μg/m^3, and >500 μg/m^3, respectively; Figure 5h is for observations of all time. In a–h, the GUP is marked by the black empty triangle.

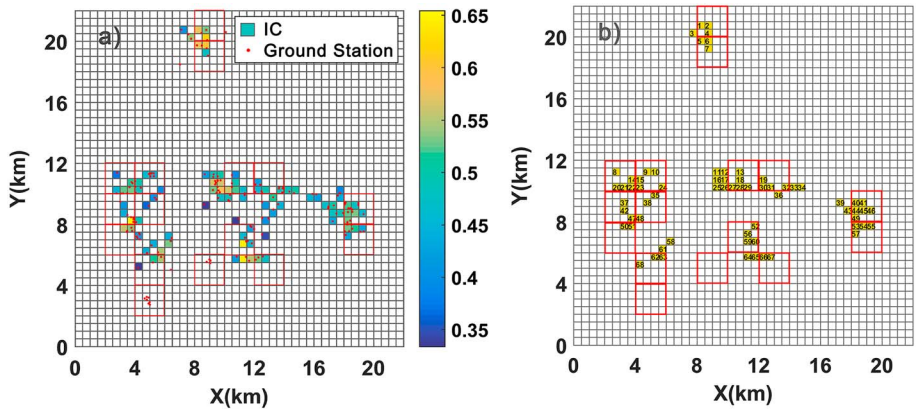

Figure 6. The determined index of correlation (IC) values for all subgrids in the study area (a) and the subgrids in 16 GUPs and the subgrids outside GUPs with IC $>$ 0.4817 (b). The subgrids that are identified as GRDs are those yellow grids in (b) with a total number of 68, which are numbered from 1 to 68.

shows the subgrids that are identified as GRDs. In total, 68 subgrids have been identified as GRDs in which the $PM_{2.5}$ observations could have a large spatial representativeness, and they have been numbered as 1 to 68 from upper left to bottom right in Figure 6b.

3.2.3. Determination of the Representative Domain

Using the method indicated in section 2, we determine the GRDs that have good spatial representativeness. Note that we have adopted the threshold values of 0.9 and 30 $\mu g/m^3$ for R and D, respectively. It is found that there are 50 GRDs, that is, those grids with serial numbers in Figure 7, which can represent their surrounding subgrids with a total number of from 1 to 20. These 50 GRDs are denoted as GRs in this study. Figure 8 further illustrated the spatial representativeness results using eight examples after filling the gaps among discontinuous subgrids in space. After filling, the 50 GRs have better spatial representativeness, with total representative area containing up to 49 subgrids. Thus, the representative domain area of these 50 GRs ranges from 0.25 km^2 to 12.25 km^2. Considering that other types of subgrids or stations could have a much weaker spatial representativeness, we could roughly conclude that $PM_{2.5}$ observations over most stations in the study area generally have a spatial representativeness less than 12.25 km^2 in winter. Another interesting result is that the representative domain is not a circle, but with directional preference. This is most likely related to the local meteorology and surface condition, which is worthy of further examination in the future. One thing we should note is that the representative domain is also limited by the observation stations we have over the study area.

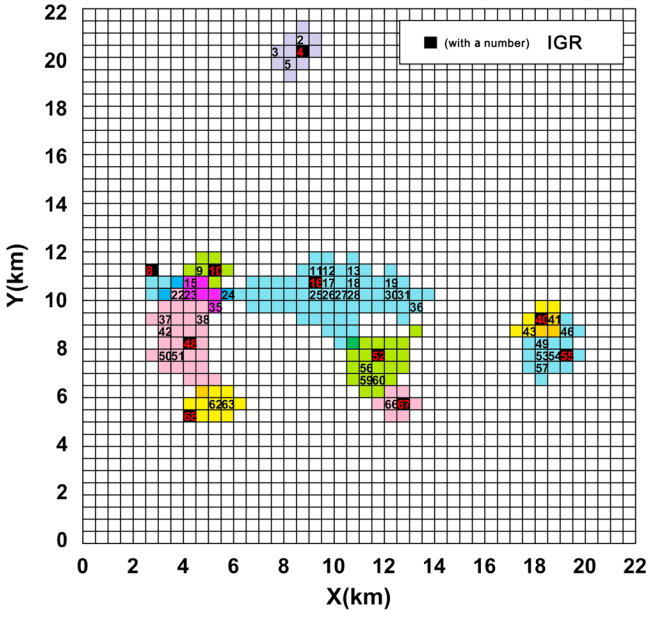

Figure 7. The 50 GRs determined in this study and the representative domains of the 10 Important Grids of Representativeness (IGRs), which are shown by different colors. The overlapped area between two domains is represented by a deeper color than the one that represents one domain (e.g., deep blue, deep green, and deep yellow). Every grid with a number represents a grid of representativeness.

3.2.4. Determination of IGR

Considering the overlapping of the representative domains among the 50 GRs, we have carried out a merging analysis by removing those GRs with small domains, which are mostly covered by larger domains represented by other GRs. We found that $PM_{2.5}$ mass concentration at 10 IGRs can represent that at all stations during the study period. The GRD serial number, number of representative subgrids, and areas of representative domain of these 10 IGRs have been listed in Table 1. The representative domain of these 10 IGRs have also been illustrated in Figure 7, from which we can see that these 10 IGRs have the ability to represent almost all the subgrids observations in this study in winter 2015. Table 1 shows that the minimum and maximum representative domain area for these 10 IGRs are 1.25 km^2 and 16.25 km^2, respectively.

Figure 9 shows the probability distribution denoted by the subgrid number of the representative domain area for 50 GRs and 10 IGRs. For the

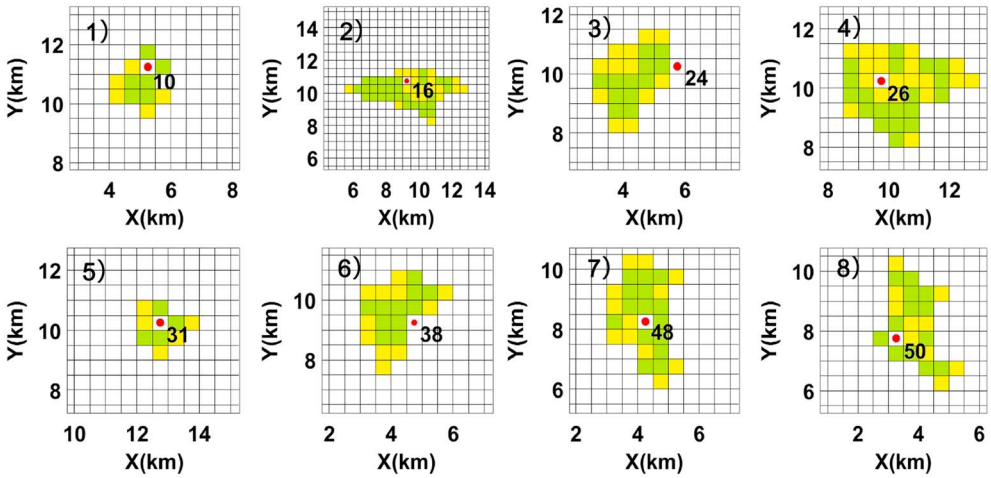

Figure 8. Eight examples for the spatial representativeness with filling subgrids. Red spots represent the grid of representativeness, yellow grids are the identified grid collection that measurement at Grid of Representativeness can represent, and the green are filled grids.

spatial representativeness of 50 GRs without merging, the spatial representative area is mostly 0.25–12.25 km^2 and only two GRs have a presentation area larger than 8 km^2 with one GR about 9–10 km^2 and one GRs about 12–13 km^2. For the IGRs with merging, the spatial representative area becomes larger while still is 1–3 km^2 for half IGRs. The representative area for the other half IGRs is between 3 and 16 km^2 with almost the same occurrence frequency. The maximum representative area for IGRs is 16.25 km^2 as a result of merging analysis, which is slightly larger than the maximum representative area of 12.25 km^2 for GRs.

Figure 10 shows the identified IGRs along with the original stations within the study area. Clearly, the number of IGRs and their representative domain are related to the availability and distribution of observation stations in the study region. This implies that the representative domain could be underestimated when there are limited number of observation stations. As shown in Figure 10, the sparse distribution in space of observation stations could cause no representative domain found for IGRs in some directions.

3.3. Evaluation

3.3.1. EOF Analysis

To evaluate the representativeness of the identified 10 IGRs above, we divide all effective subgrids into two groups, Group A with 10 IGRs and Group B including all effective grids. For the two group data sets, we carry out empirical orthogonal function (EOF) analysis. If the 10 IGRs have a good representativeness of the whole region, the spatial and temporal components in EOF decomposing of Group A and Group B will well correspond with each other.

Table 1
The GRD Serial Number, Number of Subgrids Represented, and Area of Representative Domain (km^2) of 10 IGRs

GRD serial number	Number of subgrids represented	Area of representative domain (km^2)
4	11	2.75
8	5	1.25
10	13	3.25
48	39	9.75
68	11	2.75
16	65	16.25
52	21	5.25
67	7	1.75
40	8	2.00
55	19	4.75

Note. GRD = Grids with spatial Representativeness to be Determined. IGR = Important Grids of Representativeness.

Hourly $PM_{2.5}$ value varies more significantly in time and space than daily or monthly $PM_{2.5}$ value. Daily average $PM_{2.5}$ is often used for the evaluation (Spangl et al., 2007), though hourly $PM_{2.5}$ was used in the process of determination of IGRs. To do an effective spatial-temporal EOF decomposition, we need remove the days with invalid data. There are total 93 days with valid data among 120 days, with a time fraction of 77.5% of all 120 days. The EOF decomposition shows that the first principal components of both Group A and Group B can explain the $PM_{2.5}$ variations of 98.42% and 98.11%, respectively. Figure 11a shows that the first principal components of Group A and Group B are highly correlated (r = 0.998). Similarly, the first EOF patterns of Group A and Group B can explain most of the $PM_{2.5}$ spatial variations, and Figure 11b shows high correlation (r = 0.932) between them. The results of spatial-temporal EOF analysis indicate that the $PM_{2.5}$ observations from 10 IGRs can

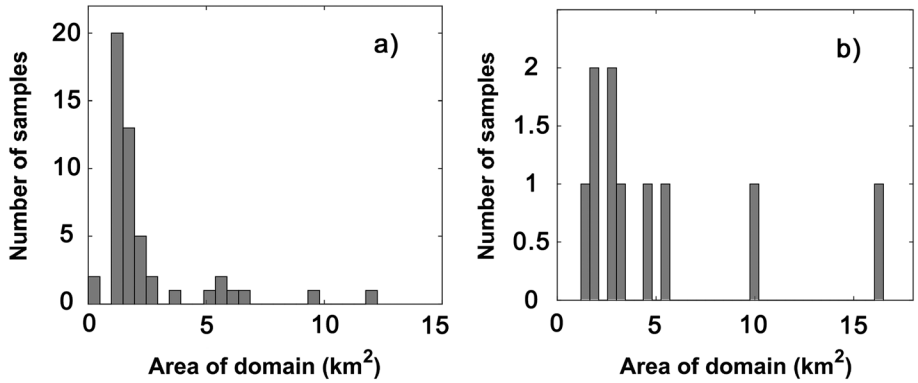

Figure 9. The probability distribution denoted by subgrid number of the representative domain area for 50 Grids of Representativeness (a) and 10 Important Grids of Representativeness (b).

well represent both spatial and temporal characteristics of $PM_{2.5}$ obtained from all sites over the study region.

3.3.2. Evaluation With Independent Data

The analysis in section 3.3.1 shows that the $PM_{2.5}$ observations at 10 IGRs can represent the spatial and temporal variations of all observations from 169 sites over the study region during the examined 4 months. We wonder whether they have the same good representativeness over other period. We here use $PM_{2.5}$ observations obtained during another period to check the representativeness of selected 10 IGRs.

The $PM_{2.5}$ observations used here are for the period from 1 to 20 December 2016. For this 20 day period, we have valid observations only at 67 stations. By regridding to 0.5 km × 0.5 km, we got 56 grids with valid data. For the selected 10 IGRs, only six of them have valid data. For other four IGRs without valid data, one IGR was removed since there are no observation stations with valid data around it, and three IGRs are replaced by nearby subgrids which belong to GRs and have valid observation data. Similarly, we set the 9 selected IGRs as Group A1 and all 56 grids as Group B1. These grids information has been shown in Figure 12a. Figure 12b shows the comparison of daily average $PM_{2.5}$ obtained between Group A1 (blue box) and Group B1 (green box) using boxplots. The observations from Group A1 agree very well with those from Group B1 in both median and quartile values for all 20 days considered here. It is also found that the daily average $PM_{2.5}$ for Group A1 and Group B1 shows almost the same temporal variation with high correlation (r = 0.999). These results suggest that the observations from selected IGRs also have a good representativeness of all grid observations over the whole area in December 2016, which indirectly evaluate the reliability of the representativeness of 10 IGRs determined earlier.

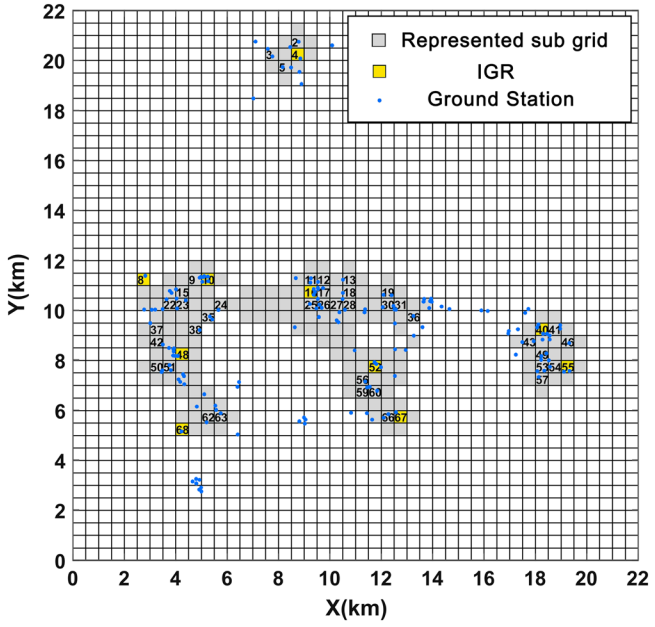

Figure 10. The identified Grids of Representativeness (GRs) (grids with numbers), Important GRs (IGRs) (yellow grids), and represented domain (gray grids) along with the original ground stations within the study area.

4. Summary and Discussion

This study examines the spatial representativeness of $PM_{2.5}$ mass concentrations observed at a single station or subgrid in a heavily polluted large city in North China. A method has been developed based on the spatial variability of $PM_{2.5}$ observations in the study area. It first identifies the stations or subgrids that are likely to have large spatial representativeness by analyzing the variogram, the CV, and the index of correlation (IC). Then the representativeness of a given station or subgrid is determined using the correlation and time-averaged difference of $PM_{2.5}$ measured between this station and others. Finally, the representative domain is determined by filling the discontinuous locations among the represented grids.

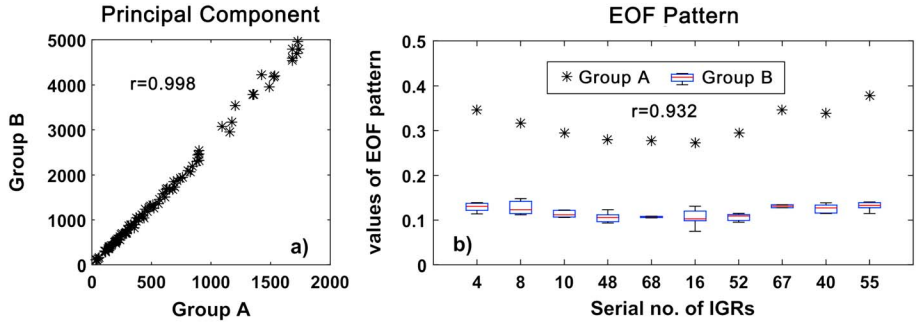

Figure 11. (a) The scatterplot of the first principal components from empirical orthogonal function (EOF) analysis between Group A and Group B; (b) the variation of first EOF patterns from Group A and Group B with 10 representative sites. The blue boxes and red lines in (b) represent the upper and lower quartile data range and the medians of the first mode results from spatial EOF pattern of Group B for grids represented by their IGRs. The black asterisk in (b) represents the first mode results from spatial EOF pattern of Group A.

Temporal variation of $PM_{2.5}$ observations shows more than 10 pollution events from November 2015 to February 2016 in the study area. The spatial variability of $PM_{2.5}$ shows a dependency on the time scale examined, as well on the pollution conditions. The spatial variability of $PM_{2.5}$ increases with the pollution severity, with the largest values in the heavy pollution period from the end of November to early January. When air is heavily polluted, there are often weak winds and weak vertical mixing making the surface pollution follows the distribution of emission sources.

The CV analysis using 2 km × 2 km grids indicates that the number of GUPs increases with pollution level when $PM_{2.5}$ mass concentration is less than 115–150 $\mu g/m^3$ and decreases when $PM_{2.5}$ mass concentration is higher. It potentially suggests the shift from local source influence to regional background/long-range transport influence to back to local to nearby high-emitting source influence when $PM_{2.5}$ concentration increase from low to high. By dividing every 2 km × 2 km grid into 0.5 km × 0.5 km subgrids, index of correlation has been analyzed for every subgrids in order to determine GRDs with a criterion. For these GRDs, we can calculate its correlation coefficient and time-averaged difference in $PM_{2.5}$ mass concentration with other surrounding subgrids. Then we identify 50 GRDs can represent some surrounding subgrids, and the 50 GRDs are called GRs. We have also filled the gaps in space among discontinuous subgrids in order to get a more reliable spatial representative domain for all GRs, and the total area of those subgrids are the spatial representative domain of the GRs. It is found that the area of spatial representative domain of the GRs for the study period is from 0.25 km^2 to 12.25 km^2.

We have also made a merging analysis by keeping as few observation stations as possible, while they can mostly represent the whole study area. We found that 10 IGRs have the ability to well represent the $PM_{2.5}$ pollution status over the same region as represented by current 169 stations at the 4 month time scale in this study. The area

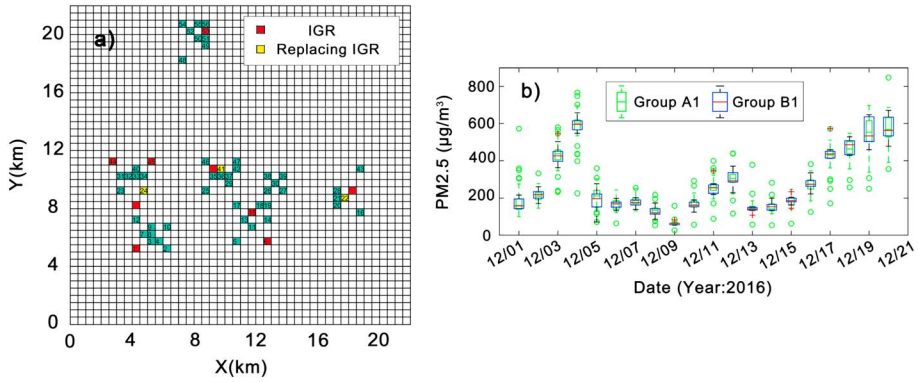

Figure 12. The grid network of 0.5 km × 0.5 km for $PM_{2.5}$ observations in December 2016 (a), and the boxplots of daily $PM_{2.5}$ observations from both Group A1 (blue box) and Group B1 (green box) (b) from 1 to 20 December 2016. In (a), all grids with numbers represent effective subgrids with $PM_{2.5}$ observation. IGR = Important Grids of Representativeness.

of representative domain is 1–3 km^2 for half of the 10 IGRs, while the maximum representative area of the 10 IGRs is 16.25 km^2. Two evaluation studies, the EOF spatial-temporal analysis and independent data analysis, both suggest the validity of the representativeness of 10 IGRs determined in this study.

This study further suggests that station redundancy exists in current observation network, and an optimal station setup framework could be made based on our proposed method here. Moreover, our analysis implies that the spatial representative domain is most likely to be underestimated due to the limited $PM_{2.5}$ observation information (sparse distribution of stations) in some directions around the GRs. In other words, the $PM_{2.5}$ mass concentration measured over the 10 IGRs could represent the $PM_{2.5}$ pollution status over a larger area than what we have indicated.

Acknowledgments
This work was supported by the Ministry of Science and Technology of China (grants 2017YFC1501403, 2013CB955802, and 2012AA120901), the National Natural Science Foundation of China (grant 41575143), the China "1000 plan" young scholar program, the State Key Laboratory of Earth Surface Processes and Resource Ecology (2017-ZY-02), and the Fundamental Research Funds for the Central Universities (2017EYT18 and 312231103). The authors also acknowledge the support by the Jet Propulsion Laboratory, California Institute of Technology, sponsored by NASA. The data used in this study are available by request to Chuanfeng Zhao through czhao@bnu.edu.cn or downloaded directly from ftp: nwpc.nmc.cn (user: pub, pswd: verygood) under directory zhao_paper_data.

References

Bréon, F. M., Vermeulen, A., & Descloitres, J. (2011). An evaluation of satellite aerosol products against sunphotometer measurements. *Remote Sensing of Environment*, *115*(12), 3102–3111. https://doi.org/10.1016/j.rse.2011.06.017

Chu, D. A., Kaufman, Y. J., Zibordi, G., Chern, J. D., Mao, J., Li, C., & Holben, B. N. (2003). Global monitoring of air pollution over land from the Earth Observing System—Terra Moderate Resolution Imaging Spectroradiometer (MODIS). *Journal of Geophysical Research*, *108*(D21), 4661. https://doi.org/10.1029/2002jd003179

Cressie, N. (1992). Statistics for spatial data. *Terra Nova*, *4*(5), 613–617. https://doi.org/10.1111/j.1365-3121.1992.tb00605.x

Engel-Cox, J. A., Holloman, C. H., Coutant, B. W., & Hoff, R. M. (2004). Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmospheric Environment*, *38*(16), 2495–2509. https://doi.org/10.1016/j.atmosenv.2004.01.039

Garrett, T. J., Zhao, C., & Novelli, P. C. (2010). Assessing the relative contributions of transport efficiency and scavenging to seasonal variability in Arctic aerosol. *Tellus B*, *62*(3), 190–196. https://doi.org/10.3402/tellusb.v62i3.16525

Holben, B. N., Tanre, D., Smirnov, A., Eck, T. F., Slutsker, I., Abuhassan, N., et al. (2001). An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET. *Journal of Geophysical Research*, *106*(D11), 12,067–12,097. https://doi.org/10.1029/2001jd900014

Kaufman, Y. J., & Fraser, R. S. (1983). Light extinction by aerosols during summer air pollution. *Journal of Climate and Applied Meteorology*, *22*(10), 1694–1706. https://doi.org/10.1175/1520-0450(1983)022%3C1694:lebads%3E2.0.co;2

Kaufman, Y. J., Tanré, D., & Boucher, O. (2002). A satellite view of aerosols in the climate system. *Nature*, *419*(6903), 215–223. https://doi.org/10.1038/nature01091

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., & Eck, T. F. (2010). Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. *Atmospheric Chemistry and Physics*, *10*(6), 14,815–14,873. https://doi.org/10.5194/acpd-10-14815-2010

Li, J., Carlson, B. E., & Lacis, A. A. (2015). How well do satellite AOD observations represent the spatial and temporal variability of $PM_{2.5}$ concentration for the United States? *Atmospheric Environment*, *102*, 260–273. https://doi.org/10.1016/j.atmosenv.2014.12.010

Ma, Z., Hu, X., Huang, L., Bi, J., & Liu, Y. (2014). Estimating ground-level PM2.5 in China using satellite remote sensing. *Environmental Science & Technology*, *48*(13), 7436–7444. https://doi.org/10.1021/es5009399

Malm, W. C., Sisler, J. F., Huffman, D., Eldred, R. A., & Cahill, T. A. (1994). Spatial and seasonal trends in particle concentration and optical extinction in the United States. *Journal of Geophysical Research*, *99*(D1), 1347–1370. https://doi.org/10.1029/93jd02916

Ministry of Environmental Protection of the People's Republic of China (2012). Technical regulation on ambient air quality index (on trial). (Rep. HJ 633–2012). Beijing, CN: Ministry of Environmental Protection of the People's Republic of China. http://kjs.mep.gov.cn/hjbhbz/bzwb/jcffbz/201203/W020120410332725219541.pdf

Peters, A., Wichmann, H. E., Tuch, T., Heinrich, J., & Heyder, J. (1997a). Comparison of the number of ultra-fine particles and the mass of fine particles with respiratory symptoms in asthmatics. *The Annals of Occupational Hygiene*, *41*, 19–23. https://doi.org/10.1093/annhyg/41.inhaled_particles_viii.19

Peters, A., Wichmann, H. E., Tuch, T., Heinrich, J., & Heyder, J. (1997b). Respiratory effects are associated with the number of ultrafine particles. *American Journal of Respiratory and Critical Care Medicine*, *155*(4), 1376–1383. https://doi.org/10.1164/ajrccm.155.4.9105082

Pope, C. A. III, Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., & Thurston, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA*, *287*(9), 1132–1141. https://doi.org/10.1001/jama.287.9.1132

Ramachandran, S. (2005). $PM_{2.5}$ mass concentrations in comparison with aerosol optical depths over the Arabian Sea and Indian Ocean during winter monsoon. *Atmospheric Environment*, *39*(10), 1879–1890. https://doi.org/10.1016/j.atmosenv.2004.12.003

Rohde, R. A., & Muller, R. A. (2015). Air pollution in China: Mapping of concentrations and sources. *PLoS One*, *10*(8), e0135749. https://doi.org/10.1371/journal.pone.0135749

Scaperdas, A., & Colvile, R. N. (1999). Assessing the representativeness of monitoring data from an urban intersection site in central London, UK. *Atmospheric Environment*, *33*(4), 661–674. https://doi.org/10.1016/s1352-2310(98)00096-x

Schwartz, J., & Neas, L. M. (2000). Fine particles are more strongly associated than coarse particles with acute respiratory health effects in schoolchildren. *Epidemiology*, *11*(1), 6–10. https://doi.org/10.1097/00001648-200001000-00004

Spangl, W., Schneider, J., Moosmann, L., & Nagl, C. (2007). Representativeness and classification of air quality monitoring stations (Rep. REP-0121). Vienna, AU: Umweltbundesamt. Retrieved from http://www.umweltbundesamt.at/fileadmin/site/publikationen/REP0121.pdf

Sun, Y., Jiang, Q., Wang, Z., Fu, P., Li, J., Yang, T., & Yin, Y. (2014). Investigation of the sources and evolution processes of severe haze pollution in Beijing in January 2013. *Journal of Geophysical Research: Atmospheres*, *119*, 4380–4398. https://doi.org/10.1002/2014jd021641

Van Donkelaar, A., Martin, R. V., & Park, R. J. (2006). Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing. *Journal of Geophysical Research: Atmospheres*, *111*, D21201. https://doi.org/10.1029/2005jd006996

Van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environmental Health Perspectives*, *118*(6), 847–855. https://doi.org/10.1289/ehp.0901623

Van Donkelaar, A., Martin, R. V., Pasch, A. N., Szykman, J. J., Zhang, L., Wang, Y. X., & Chen, D. (2012). Improving the accuracy of daily satellite-derived ground-level fine aerosol concentration estimates for North America. *Environmental Science & Technology*, *46*(21), 11,971–11,978. https://doi.org/10.1021/es3025319

Wang, J., & Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: Implications for air quality studies. *Geophysical Research Letters*, *30*(21), 2095. https://doi.org/10.1029/2003gl018174

Wang, Q., Cao, J., Tao, J., Li, N., Su, X., Chen, L. A., et al. (2013). Long-term trends in visibility and at Chengdu, China. *PLoS One*, *8*(7), e68894. https://doi.org/10.1371/journal.pone.0068894

Wang, Y., Zhang, J., Wang, L., Hu, B., Tang, G., Liu, Z., et al. (2014). Researching significance, status and expectation of haze in Beijing-Tianjin-Hebei region. *Advances in Earth Science*, *29*(3), 388–396.

Yin, Y., & Chen, L. (2007). The effects of heating by transported dust layers on cloud and precipitation: A numerical study. *Atmospheric Chemistry and Physics*, *7*(13), 3497–3505. https://doi.org/10.5194/acp-7-3497-2007

Zhang, H., Hoff, R. M., & Engel-Cox, J. A. (2009). The relation between Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol optical depth and $PM_{2.5}$ over the United States: A geographical comparison by US Environmental Protection Agency regions. *Journal of the Air & Waste Management Association*, *59*(11), 1358–1369. https://doi.org/10.3155/1047-3289.59.11.1358

Zhao, C., & Garrett, T. J. (2015). Effects of Arcstic haze on surface cloud radiative forcing. *Geophysical Research Letters*, *42*, 557–564. https://doi.org/10.1002/2014gl062015

Zhao, C., Andrews, A. E., Bianco, L., Eluszkiewicz, J., Hirsch, A., MacDonald, C., et al. (2009). Atmospheric inverse estimates of methane emissions from Central California. *Journal of Geophysical Research*, *114*, D16302. https://doi.org/10.1029/2008jd011671

Zheng, C., Zhao, C., Zhu, Y., Wang, Y., Shi, X., Wu, X., et al. (2017). Analysis of influential factors for the relationship between $PM_{2.5}$ and AOD in Beijing. *Atmospheric Chemistry and Physics*, *17*(21), 13,473–13,489. https://doi.org/10.5194/acp-17-13473-2017