A framework for delineating the regional boundaries of PM$_{2.5}$ pollution: A case study of China

Jianzheng Liu $^{a, b}$, Weifeng Li $^{a, b, *}$, Jiansheng Wu $^{c, d}$

$^a$ Department of Urban Planning and Design, Faculty of Architecture, The University of Hong Kong, Hong Kong, China
$^b$ Shenzhen Institute of Research and Innovation, The University of Hong Kong, Shenzhen, China
$^c$ Key Laboratory of Human Environmental Science and Technology, Peking University Shenzhen Graduate School, Shenzhen 518055, China
$^d$ Key Laboratory for Earth Surface Processes, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China

**Article info**

**Article history:**
Received 26 September 2017
Received in revised form 15 December 2017
Accepted 17 December 2017

**Keywords:**
Fine particulate matter
Area of interaction
Regional boundaries
Air pollution in China
Time series clustering

**Abstract**

Fine particulate matter (PM$_{2.5}$) pollution has been a major issue in many countries. Considerable studies have demonstrated that PM$_{2.5}$ pollution is a regional issue, but little research has been done to investigate the regional extent of PM$_{2.5}$ pollution or to define areas in which PM$_{2.5}$ pollutants interact. To allow for a better understanding of the regional nature and spatial patterns of PM$_{2.5}$ pollution, this study proposes a novel framework for delineating regional boundaries of PM$_{2.5}$ pollution. The framework consists of four steps, including cross-correlation analysis, time-series clustering, generation of Voronoi polygons, and polygon smoothing using polynomial approximation with exponential kernel method. Using the framework, the regional PM$_{2.5}$ boundaries for China are produced and the boundaries define areas where the monthly PM$_{2.5}$ time series of any two cities show, on average, more than 50% similarity with each other. These areas demonstrate straightforwardly that PM$_{2.5}$ pollution is not limited to a single city or a single province. We also found that the PM$_{2.5}$ areas in China tend to be larger in cold months, but more fragmented in warm months, suggesting that, in cold months, the interactions between PM$_{2.5}$ concentrations in adjacent cities are stronger than in warmer months. The proposed framework provides a tool to delineate PM$_{2.5}$ boundaries and identify areas where PM$_{2.5}$ pollutants interact. It can help define air pollution management zones and assess impacts related to PM$_{2.5}$ pollution. It can also be used in analyses of other air pollutants.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Fine particulate matter (PM$_{2.5}$) is one of the air pollutants most detrimental to human health (Samet et al., 2000; Yuan et al., 2012). In China, it has been estimated that approximately 1.6 million deaths per year, about 17% of total annual deaths, are attributable to PM$_{2.5}$ (Rohde and Muller, 2015). Public calls for government action to address the problem have been growing, but a limited understanding of the regional nature of PM$_{2.5}$ pollution could probably undermine the efforts of government agencies to formulate effective policies and measures. In February 2016, the Beijing planning authority proposed construction of five 500-m wide ventilation corridors, in an attempt to blow air pollution away from the city (Gan, 2016). While this policy might work in the short term, clearly the rationale behind this policy reflects a “Not-In-My-Backyard” mentality, i.e., “blow the pollution away from my city and let it harm other places”. However, evidence is mounting that local PM$_{2.5}$ concentrations are susceptible to regional influences such as transport of external pollution emissions (Chen et al., 2016; Khuzestani et al., 2017; Li et al., 2015; Wu et al., 2015; Yuan et al., 2012).

Once it is recognized that PM$_{2.5}$ pollution is a regional issue that requires regional cooperation, the next question for researchers is, how regional are the PM$_{2.5}$ pollutants? That is, how large is the area within which PM$_{2.5}$ pollutants interact? Is there a method to delineate the boundaries of PM$_{2.5}$ pollution? The answers to these questions are imperative because they could help city managers to develop effective measures for pollution mitigation. The answers to these questions are also useful because they would advance our understanding of the regional nature of PM$_{2.5}$ pollution, and could help define management zones for air pollution management.
control. However, research on PM2.5 pollution is still for the most part focused on the spatiotemporal characterization of PM2.5 pollution (Chen et al., 2015; He et al., 2017; Hu et al., 2014; Huang et al., 2015; Luo et al., 2017; Yang and Christakos, 2015), factors influencing PM2.5 pollution (Chen et al., 2017; He et al., 2017; Liu et al., 2016b; Pearce et al., 2011), PM2.5 forecasting (Li et al., 2017b; Zhan et al., 2017), source apportionment (Lv et al., 2015; Zhang et al., 2013), and the health effects of PM2.5 pollution (Lin et al., 2016), little research has been done to address these important questions.

In an attempt to address these questions, this paper proposes a framework to delineate the regional boundaries of PM2.5 pollution. The framework uses techniques including cross correlation, time series clustering, Voronoi polygons generation, and polynomial approximation with exponential kernel (PAEK). We apply this framework to delineate the PM2.5 boundaries in China using ground-based PM2.5 measurements collected in 157 cities. It is anticipated that this framework will identify areas within which PM2.5 pollutants interact, allowing a better understanding of the heterogeneity and spatial patterns of PM2.5 pollution, and helping to define management zones for air pollution control and assess impacts related to PM2.5 pollution. We also hope that the identified PM2.5 boundaries for China in this study could be used to support further investigations into the delineation of air pollution management zones in China.

The following section introduces the framework for delineation of the boundaries of PM2.5 pollution in detail. Section 3 presents the application of this framework in China, including the data and the resulting boundaries of PM2.5 pollution. Section 4 presents the interpretations of the results, and describes potential applications of the framework, and limitations on PM2.5 boundaries identified in China. The final section summarizes the findings of the study.

2. The framework for delineating PM2.5 boundaries

The proposed framework is shown diagrammatically in Fig. 1. The first step is key to understanding this framework, as the framework is built on the significant interactions between PM2.5 pollution in adjacent cities. In this step, the strengths of these interactions are calculated using a cross correlation method. Using the strengths of these interactions as the measure of similarity, time-series clustering is performed using unweighted pair group method using arithmetic averages (UPGMA). In the third step, Voronoi polygons are produced based on the clustering results of the cities. In the fourth step, polygons are smoothed for better cartographic presentation. Each step is described in more detail below.

2.1. Significant interactions between PM2.5 concentrations in adjacent cities

In the first step, we will show how to calculate the strengths of these interactions using cross correlation method, and demonstrate the significant interactions between PM2.5 pollution in adjacent cities. Details on the data used and the study area in this section can be found in section 3.1.

We found strong and significant interactions between the PM2.5 time series of adjacent cities. For example, consider the PM2.5 time series from Beijing and Tianjin in December 2014. As shown in Fig. 2a, the PM2.5 time series of Beijing and Tianjin had very similar trends and there was a strong correlation and an obvious time lag between the two time series.

To measure the strengths of these interactions between PM2.5 time series, this study employed the cross-correlation method, a technique used in the field of signal processing to measure the similarity of two time series as a function of the lag of one relative to the other (Rhudy et al., 2009). The cross-correlation method is a two-step process. First, the correlation coefficients between two time series are calculated at progressively varying time lags. Second, the maximum correlation coefficient is identified and the time lag corresponding to that maximum correlation coefficient is noted. This maximum correlation coefficient occurs at the time shift for which the two time series are best aligned. The process can be expressed mathematically using the following equations:

\[ R(\tau) = \text{Corr}(S_1(t), S_2(t - \tau)) \]

\[ R_{\text{max}} = \max(R(\tau)) \]

\[ T_{\text{delay}} = \text{argmax}_\tau(R(\tau)) \]

where \( R(\tau) \) is the Pearson correlation coefficient between two time series at a specific time lag value \( \tau \), and \( S_1 \) and \( S_2 \) are the two time series to be analyzed. \( R_{\text{max}} \) is the maximum correlation coefficient found in the analysis, and \( T_{\text{delay}} \) is the time lag that generates \( R_{\text{max}} \).

To illustrate the cross-correlation analysis, consider again the two PM2.5 time series from Beijing and Tianjin in December 2014. First, the correlation coefficients were calculated at different time lags as shown in Fig. 2b; then the maximum correlation coefficient was identified as 0.706, and the time lag that creates the maximum correlation coefficient can be determined as 7 h. As can be seen from Fig. 2a, the best alignment between the two time series can be obtained by shifting the Tianjin PM2.5 time series to the left by approximately 7 h, which is consistent with the results of cross-correlation analysis. In this study, the maximum correlation coefficients identified in the cross-correlation analysis are used to measure the strength of the interactions between PM2.5 time series.

To demonstrate the significant interactions between PM2.5 concentrations in adjacent cities, the intercity correlation coefficients among the cities in the Beijing–Tianjin–Hebei region were calculated using the cross-correlation method. As shown in Fig. 3, there were strong associations between the PM2.5 time series not only between Beijing and Tianjin, but also among many other cities. This demonstrates that there are significant interactions between PM2.5 pollution in adjacent cities, which is consistent with previous studies concluding that strong bidirectional coupling of
PM$_{2.5}$ pollution exists among neighboring cities (Chen et al., 2016; Yang and Christakos, 2015). Fig. 3 also shows that the strengths of these interactions vary between different cities. Clearly these features of the interactions between PM$_{2.5}$ time series across a region make it possible to delineate the boundaries of PM$_{2.5}$ pollution in that region. In fact, using the intercity correlation coefficient as the measure of similarity, the time series clustering used in the second step can group adjacent cities which have strong interactions into the same cluster, and separate cities which have weak interactions into different clusters.

2.2. Time-series clustering

Clustering is a process of partitioning a set of data objects (in this case, the PM$_{2.5}$ time-series from each city) into subsets or clusters (Austin et al., 2013; Malley et al., 2014). Objects within a cluster are similar to each other, but dissimilar to objects in other clusters. There are two critical components that must be established prior to application of a clustering technique: (1) the clustering algorithm defining how to cluster; (2) the distance measure defining the degree of similarity (Wang et al., 2006).

A clustering algorithm describes the procedures by which similar objects are clustered. There is a wide range of clustering algorithms available for selection, including agglomerative hierarchical clustering and K-means algorithms. This study uses UPGMA as the clustering method because, compared with K-means, it does not require a predetermined number of clusters and generates repeatable and consistent results (Aghabozorgi et al., 2015); moreover, it is able to produce more robust cluster results than many other hierarchical clustering methods (Rodrigues and Diniz-Filho, 1998).

The distance measure employed in cluster analysis is used to establish the degree of similarity between the objects. This study uses the time lag-adjusted intercity correlation coefficient as the measure of similarity. The distance measure used in time-series clustering can be mathematically expressed as follows:

$$D(S_1, S_2) = 1 - R_{\text{max}},$$

where $D(S_1, S_2)$ is the distance between the two time series $S_1$ and $S_2$, and $R_{\text{max}}$ is the maximum correlation coefficient computed from Equation (2).

When two objects are more similar, the intercity correlation coefficient of the two objects will be closer to one, and therefore the distance measure will be closer to zero. When two data objects are
less similar, the distance measure will be further from zero.

Tests for statistical significance are needed to determine whether the maximum correlation coefficient $R_{\text{max}}$ is significantly larger than the correlation coefficient without alignment $R(0)$. A value of $R_{\text{max}}$ that was significantly larger than $R(0)$ would mean that the difference between the two coefficients is not due to random chance and the maximum correlation coefficients can be used in further analysis. To determine whether the maximum correlation coefficient $R_{\text{max}}$ identified by cross-correlation analysis was significantly larger than the correlation coefficient without alignment $R(0)$, the correlations were transformed using Fisher’s r-to-z transformation (Fisher, 1921). This transformation and its method of calculation are described in detail by Kenny (1987).

When the maximum correlation coefficient $R_{\text{max}}$ was significantly larger than the correlation coefficient without alignment $R(0)$, the distance measure was calculated using $R_{\text{max}}$. When $R_{\text{max}}$ was not significantly larger than $R(0)$, the distance measure was computed using $R(0)$ as follows:

$$D(S_1, S_2) = \begin{cases} 1 - R_{\text{max}}, & R_{\text{max}} \text{ is significantly larger than } R(0) \\ 1 - R(0), & R_{\text{max}} \text{ is not significantly larger than } R(0) \end{cases}$$

(5)

where the definitions of $D(S_1, S_2)$ and $R_{\text{max}}$ are the same as in equation (4), and $R(0)$ is the correlation coefficient without cross-correlation alignment.

Selecting a cutting threshold distance in the dendrogram of the hierarchical cluster analysis is a critical step, because it controls how many clusters will be formed. There are many methods that can accomplish this task, including the Elbow method and Silhouette method (Aghabozorgi et al., 2015), but no method is universally applicable because the number of clusters depends on expert knowledge of the subject and study area, and on the context of the data in question (Estivill-Castro, 2002). A value of 0.5 was selected as the cutting threshold distance for this study, meaning that objects will be grouped as a cluster only at those dendrogram nodes where the average correlation distance between data objects (in this case, PM$_{2.5}$ time series) is less than 0.5. The implication of this threshold is that the average cross-correlation coefficient between pairs of PM$_{2.5}$ time series in the cluster is greater than 0.5. In other words, the PM$_{2.5}$ time series of each city within a cluster has, on average, more than 50% similarity with the PM$_{2.5}$ time series of each of the other cities in the cluster.

The length of temporal duration in clustering analysis also matters. It is inappropriate to conduct clustering analysis over the entire year because the correlations between PM$_{2.5}$ time series among different cities were assumed to vary monthly. It is also unrealistic to carry out the analysis on a daily basis because there would be insufficient numbers of PM$_{2.5}$ measurements for the clustering analysis to separate cities into different clusters. Analysis on a monthly basis is a reasonable compromise between adequate sample size, a reasonably fine temporal resolution, and practical uses of the results.

Python 2.7.5 was used to perform the clustering analyses for this
study.

2.3. Delineating the PM$_{2.5}$ boundaries

Tobler’s First Law of Geography states that “Everything is related to everything else, but near things are more related than distant things” (Tobler, 1970). Ideally, all cities falling within the same cluster should be spatially adjacent. Accordingly, after cluster analysis was complete, those cities that were not spatially adjacent to other cities in the same cluster, or that were standing alone as a singleton cluster, were considered to be outliers and were removed from further analysis.

Voronoi polygons were then created using the city point features, encompassing the areas that were closest to each city relative to all other cities. The city polygons were then spatially merged based on the clustering results. Finally, polynomial approximation with exponential kernel (PAEK) was applied to smooth the merged Voronoi polygons for better cartographic presentation.

3. Delineation of PM$_{2.5}$ boundaries in China

3.1. PM$_{2.5}$ data in China

The ground-based PM$_{2.5}$ concentration data used in this study were collected by the national air quality monitoring network from January 1 to December 31, 2014; this network is run and maintained by the Chinese Ministry of Environmental Protection (MEP). The MEP has been publishing hourly concentration measurements of six air pollutants, including PM$_{2.5}$, through an online reporting portal (http://113.108.142.147/20035/emppublish) since early 2013. However, this official online reporting portal does not provide access to historical data. Fortunately, third parties, including AQISTUDY.cn and EPMAP.org, have been crawling these data since late 2013; this study obtained air quality data from 1 January 2014 to 31 December 2014 from the two third parties. Since there were missing hourly measurements in both data sources, the two data-sets were combined to fill these data gaps and obtain a more complete 24-h PM$_{2.5}$ measurement dataset for each day of 2014.

A comprehensive quality check of the raw data was conducted to reduce the impact of problematic data points, including duplicated data records, missing measurements with placeholders, implausible zeros, and data points with unreasonably high PM$_{2.5}$ concentrations (>1000 $\mu$g/m$^3$). The data show that 1074 stations were collecting PM$_{2.5}$ data in 190 Chinese cities in 2014. Ideally, each station would collect 8760 hourly measurements per year. However, 15 stations collected less than two thirds, and 194 stations collected less than one third, of this ideal total. To ensure a sufficient data size for monthly analysis, data from these stations were removed from further analysis. The remaining data contain a total of 865 stations in 161 cities.

Following the quality check, the air quality monitoring data were aggregated for each city by averaging the hourly data from all the stations within each city. This was done to reduce computational complexity of the clustering analysis and resulted in 161 PM$_{2.5}$ time series, corresponding to 161 cities. Since there are not a sufficient number of cities with ground-based PM$_{2.5}$ measurements in western China, the study area is limited to locations east of 98°E, which includes 157 cities and over 98% of the total population of China. Fig. 4 shows the locations of cities in this area.

3.2. Results

The results of the time-series clustering in the framework are clusters of cities with similar PM$_{2.5}$ trends within the study area during the 12 months of 2014. Fig. 5 shows the mapped results of time-series clustering for February, May, August, and November of 2014. As can be seen, the clustering results fit the First Law of Geography very well, as nearly all cities within each cluster are spatially adjacent to one another.

Fig. 6 shows the mapped PM$_{2.5}$ boundaries for February, May, August, and November of 2014, which were delineated based on the clustering results in Fig. 5. The average PM$_{2.5}$ concentration of each PM$_{2.5}$ area is calculated and displayed in Fig. 6. The breakpoints of the PM$_{2.5}$ concentrations shown in the legend are set based on the revised air quality standards for particle pollution administered by the U.S. Environmental Protection Agency (Environmental Protection Agency, 2012).

4. Discussion

4.1. Interpretation of the PM$_{2.5}$ boundaries in China

As explained in section 2.2, a PM$_{2.5}$ area defined by the boundaries is an area where the PM$_{2.5}$ time series of any two cities have more than 50% similarity with each other on average. For example, in February 2014 (Fig. 6a), North China and Northeast China, including Beijing, Tianjin, Hebei, Shanxi, Inner Mongolia, Shaanxi, Jilin, Heilongjiang and Liaoning are all within the same PM$_{2.5}$ area, indicating that, in February 2014, all cities within this large PM$_{2.5}$ area have PM$_{2.5}$ time series that show more than 50% similarity with one another on average, and the PM$_{2.5}$ pollution in different cities in this area interact with each other. These large PM$_{2.5}$ areas demonstrate straightforwardly that the problem of PM$_{2.5}$ pollution is not limited to a single city or a single province, but is a regional issue requiring pollution mitigation policies and measures at regional level. This result is consistent with the conclusions of many previous studies (China Council for International Cooperation on Environment and Development, 2014; He et al., 2017; Khuzestani et al., 2017; Li et al., 2017a).

As shown in Fig. 6 and Fig. S1, the boundaries of PM$_{2.5}$ areas vary considerably month by month. We speculate that the variable boundaries of the PM$_{2.5}$ areas reflect the volatile nature of synoptic meteorological conditions, because the meteorological factors play a dominant role in affecting PM$_{2.5}$ concentrations (Chen et al., 2017; Cheng and Li, 2010; He et al., 2016, 2017; Huang et al., 2015; Jia et al., 2008; Liu et al., 2016a; Luo et al., 2017; Pearce et al., 2011), and the meteorological factors change frequently. Fig. 6 and Fig. S1 also show that the PM$_{2.5}$ areas tend to be larger in cold months such as January, February, November, and December, while in warm months they are more fragmented. This suggests that in cold months the interactions between PM$_{2.5}$ concentrations in adjacent cities are probably stronger than the interactions in warm months. The reason behind might be that there are stronger influences from meteorological conditions and pollution emissions in cold months, while in warm months the influences are weaker, resulting in fragmented PM$_{2.5}$ areas (Liu et al., 2016a). Specifically, PM$_{2.5}$ emissions are much greater in cold months than other months (Wang et al., 2014; Zhang and Cao, 2015) and the effects of meteorological conditions (e.g., air pressure and wind speeds) on PM$_{2.5}$ concentrations in cold months tend to be stronger than during the rest of the year (Liu et al., 2016a). These conditions probably contribute to formation of larger PM$_{2.5}$ areas in cold months than in warm months.

4.2. Potential applications of the framework

The framework proposed in this paper provides an approach that can be used in several potential applications, including identification of areas of PM$_{2.5}$ pollution interactions, understanding the spatial patterns of PM$_{2.5}$ pollution, impact assessment related to
PM$_{2.5}$ pollution, and zoning for air pollution control. This framework can also be used to analyze various properties of other air pollutants, including sulfur dioxide, nitrogen dioxide, ozone, and carbon monoxide.

4.2.1. Identification of areas of interaction of PM$_{2.5}$ pollution

The boundaries of PM$_{2.5}$ areas generated using this framework essentially delineate areas of interaction of PM$_{2.5}$ pollutants in adjacent cities. Within one area of interaction, the PM$_{2.5}$ time series of any two cities have more than 50% similarity with each other on average, indicating a strong and dynamic relationships and interactions between PM$_{2.5}$ pollution in adjacent cities. These areas of interaction demonstrate the regional nature of PM$_{2.5}$ pollution, and show the regional extent and magnitude of the PM$_{2.5}$ pollution problem.

4.2.2. Understanding the spatial pattern of PM$_{2.5}$ pollution

The framework allows for better understanding of the heterogeneity and spatial patterns of PM$_{2.5}$ pollution in China. The PM$_{2.5}$ boundaries provide information on where and when similar PM$_{2.5}$ time series and strong interactions between PM$_{2.5}$ pollutants occur in adjacent cities. These variable PM$_{2.5}$ boundaries reflect the differences of the influences of meteorological conditions and other factors in impacting the distribution of PM$_{2.5}$ pollutants in different geographic locations.

4.2.3. Defining management zones for air pollution control

The PM$_{2.5}$ boundaries produced using this framework can provide useful references to define management zones for regional air pollution control. The current zone designation policy for air pollution control is based on political administrative boundaries, which do not consider the complex trans-boundary transport of air pollutants, the spatial and temporal distribution of these pollutants, and the influence of meteorological conditions (China Council for International Cooperation on Environment and Development, 2014). A new zone designation policy based on scientific evidence should be implemented to provide more effective regional air pollution control. Although the maps of the PM$_{2.5}$ boundaries may not be directly used for this purpose, they provide a good and useful reference for the formulation of management zones for regional air pollution control.

4.2.4. Impact assessment

The PM$_{2.5}$ boundaries produced using this framework might
Fig. 5. Clusters of cities generated through time-series clustering analysis for (a) February, (b) May, (c) August and (d) November in 2014. Different colors of the dots denote different city clusters. These maps were produced using ArcGIS 10.2.2 (www.esri.com). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
Fig. 6. Boundaries of PM$_{2.5}$ areas and mean PM$_{2.5}$ concentration for each area for (a) February, (b) May, (c) August, and (d) November in 2014. The breakpoints of the PM$_{2.5}$ concentration in the legend are set based on the revised air quality standards for particle pollution administered by the U.S. Environmental Protection Agency (Environmental Protection Agency, 2012). Maps of PM$_{2.5}$ boundaries for the remaining months of 2014 are provided in the Supporting Information (Fig. S1). These maps were produced using ArcGIS 10.2.2 (www.esri.com).
also support impact assessments related to PM$_{2.5}$ pollution. As described in section 2.2, a PM$_{2.5}$ area is an area in which the PM$_{2.5}$ time series of adjacent cities show, on average, more than 50% similarity with each other, suggesting that this area is probably subject to the same or similar influences from meteorological and emission factors. Different PM$_{2.5}$ areas imply that these areas are under different influences from meteorological and emission factors, and possibly different impacts on the environment, economy, and health. Impact assessments could be possibly performed on a PM$_{2.5}$ area basis to allow rapid estimates of impact on environment, economy and health.

4.3 Notes on using the framework

The focus of this study is the framework, while the delineation of PM$_{2.5}$ boundaries in China is an application of the framework. This section describes two points that users need to consider when using the framework; limitations of the specific application to China will be discussed in section 4.4.

The first point is related to the level of analysis. In delineation of the PM$_{2.5}$ boundaries in China, the framework uses the city as the level of analysis. The PM$_{2.5}$ time series data are aggregated at the city level by averaging the hourly data from all stations within each city, and Voronoi polygons are generated based on the point locations of the cities. However, the framework can be applied to finer levels of analysis, as long as there are time series data at those levels. For example, the framework can use data measured at the level of a grid (cell) in the remote sensing images.

The second consideration relates to handling outliers in the second step of the framework, namely the time series clustering. In the application of the framework in China described in this study, cities that were not spatially adjacent to other cities in the same cluster, or that were standing alone as a singleton cluster, were considered outliers and were removed from further analysis. This assumes that these outlier city points do not represent the regional background air quality. However, whether these outliers are representative or not is debatable, depending on one defines the scale of the region. For example, although a cluster with a single city cannot represent the background air quality in a large region encompassing several cities, it is safe to say the singleton cluster could represent the air quality within that city. Users must decide how to handle the outliers in the clustering process, depending on the specific requirements of their applications.

4.4 Limitations of the identified PM$_{2.5}$ boundaries in China

The identified PM$_{2.5}$ boundaries in China define very broad regions and this level of resolution may not be small enough to be useful for all purposes. However, compared with the air basins in California designated by the California Air Resource Board (2012) and the air zones and airsheds in Canada defined by the Canadian Council of Ministers of the Environment (2012), the PM$_{2.5}$ boundaries in China identified in this paper may not be too large. Moreover, the size of a PM$_{2.5}$ area within the boundaries is determined by the strength of the interactions between PM$_{2.5}$ pollution in adjacent cities and the threshold distance determined in the process of time series clustering. A lower threshold distance can be used to derive PM$_{2.5}$ boundaries at finer scales for other applications, as allowed by the available data.

Furthermore, the PM$_{2.5}$ boundaries do not necessarily have to match the administrative boundaries although, in practice, the administrative boundaries should be considered. A PM$_{2.5}$ area discussed in this paper can be considered as similar to a river basin, which is potentially a very large area that transcends administrative boundaries. For example, the Yangtze River basin measures approximately 2 million square kilometers, equivalent to the area of the country of Mexico (CEO Water Mandate, 2016), and this river basin does not fit the administrative boundaries.

There are two limitations regarding the PM$_{2.5}$ boundaries in China identified in this study.

The first limitation is that the PM$_{2.5}$ areas may not be accurate due to the limited number of cities with monitoring stations available for cluster analysis. However, we expect that the accuracy of the PM$_{2.5}$ boundaries will improve as more and more cities install air quality monitoring stations, and these cities can be added to the analysis in future research.

The other limitation is that the variable locations of PM$_{2.5}$ boundaries across time make it difficult to put the boundaries into practical use for defining air quality management zones. As shown in Fig. 6 and Fig. S1, the boundaries vary considerably month by month. Further, the boundaries might be subject to change with use of additional or different data. However, despite these limitations, we believe the PM$_{2.5}$ boundaries identified in this paper provide a useful reference and a sound basis for further investigations into the delineation of air zones for policy development in China. This future research may combine multiple years of PM$_{2.5}$ data and other data to derive stable and accurate boundaries for practical use.

5. Conclusions

A novel framework is proposed to delineate the PM$_{2.5}$ boundaries. This framework builds on the significant interactions between PM$_{2.5}$ pollution in adjacent cities, and consists of four steps: calculate the interaction between PM$_{2.5}$ pollution in adjacent cities using a cross-correlation method, conduct time series clustering, generate Voronoi polygons, merge those polygons based on cluster results, and finally, smooth the merged polygons.

Using the framework, this study delineated PM$_{2.5}$ boundaries in China using ground-based PM$_{2.5}$ concentration data from 2014. The results show that boundaries of PM$_{2.5}$ areas vary considerably month by month, possibly due to the variable nature of synoptic meteorological conditions. The PM$_{2.5}$ areas are larger in colder months, while in warm months the PM$_{2.5}$ areas are more fragmented, probably due to the stronger influences of meteorological conditions in cold months than in warm months. Although there are several limitations with the identified PM$_{2.5}$ boundaries in China, these boundaries have provided a sound basis for further investigations and it is expected that future research will reduce the effect of these limitations as more data become available for analysis.

The study demonstrated that the proposed framework is an approach that can be used in identification of the areas of interaction of PM$_{2.5}$ pollution, understanding the spatial patterns of PM$_{2.5}$ pollution, developing impact assessments related to PM$_{2.5}$ pollution, and defining air pollution control zoning. The framework can also be applied at finer levels of analysis than the city level, and can be used in analyses of other air pollutants such as sulfur dioxide, nitrogen dioxide, ozone, and carbon monoxide.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant No. 41471370).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.envpol.2017.12.064.