



# Assessment of German population exposure levels to PM<sub>10</sub> based on multiple spatial-temporal data

Xiansheng Liu<sup>1,2</sup> · Haiying Huang<sup>3,4</sup> · Yiming Jiang<sup>3,4</sup> · Tao Wang<sup>5</sup> · Yanling Xu<sup>6</sup> · Gülçin Abbaszade<sup>1</sup> · Jürgen Schnelle-Kreis<sup>1</sup> · Ralf Zimmermann<sup>1,2</sup>

Received: 21 May 2019 / Accepted: 14 November 2019 / Published online: 24 December 2019  
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

Particulate matter is the key to increasing urban air pollution, and research into pollution exposure assessment is an important part of environmental health. In order to classify PM<sub>10</sub> air pollution and to investigate the population exposure to the distribution of PM<sub>10</sub>, daily and monthly PM<sub>10</sub> concentrations of 379 air pollution monitoring stations were obtained for a period from 01/01/2017 to 31/12/2017. Firstly, PM<sub>10</sub> concentrations were classified using the head/tail break clustering algorithm to identify locations with elevated PM<sub>10</sub> levels. Subsequently, population exposure levels were calculated using population-weighted PM<sub>10</sub> concentrations. Finally, the power-law distribution was used to test the distribution of PM<sub>10</sub> polluted areas. Our results indicate that the head/tail break algorithm, with an appropriate segmentation threshold, can effectively identify areas with high PM<sub>10</sub> concentrations. The distribution of the population according to exposure level shows that the majority of people is living in polluted areas. The distribution of heavily PM<sub>10</sub> polluted areas in Germany follows the power-law distribution well, but their boundaries differ from the boundaries of administrative cities; some even cross several administrative cities. These classification results can guide policymakers in dividing the country into several areas for pollution control.

**Keywords** PM<sub>10</sub> air pollution · Spatio-temporal distribution · Head/tail · Power-law distribution · Population exposure · Germany

## Introduction

Particulate matter (PM) air pollution can be significantly detrimental to human health (Boldo et al. 2006). The World Health Organization reported that about 0.8 million people are dying each year due to air pollution (WHO 2002). Samet (2000) have found that the relative mortality rate, caused by cardiovascular

and respiratory diseases, increases by 0.68 percent per each 10 µg/cm<sup>3</sup> increase in the PM<sub>10</sub> level. As PM air pollution events occur frequently, this topic has aroused high social attention and academic interest. Furthermore, PM is currently considered to be the best indicator for health effects by ambient air pollution (Burnett et al. 2014; WHO 2014). Therefore, knowledge of the present state and the spatio-temporal distribution of PM and

Responsible editor: Gerhard Lammel

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11356-019-07071-0>) contains supplementary material, which is available to authorized users.

✉ Jürgen Schnelle-Kreis  
juergen.schnelle@helmholtz-muenchen.de

<sup>1</sup> Joint Mass Spectrometry Center, Cooperation Group Comprehensive Molecular Analytics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

<sup>2</sup> Joint Mass Spectrometry Center, Chair of Analytical Chemistry, University of Rostock, 18059 Rostock, Germany

<sup>3</sup> Institute of Virology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

<sup>4</sup> Institute of Virology, Technical University of Munich, Trogerstr. 30, 81675 München, Germany

<sup>5</sup> Shanghai Key Laboratory of Atmospheric Particle Pollution and Prevention, Department of Environmental Science & Engineering, Fudan University, Shanghai 200433, China

<sup>6</sup> College of Plant Health and Medicine, Qingdao Agricultural University, Qingdao 266109, China

evaluation of the adverse health effects, caused by PM<sub>10</sub> pollution, are important for understanding this air pollution problem and protecting human health and establishing pollution control policies.

So far, it has been shown that frequent exceedances of the limit for PM<sub>10</sub> concentrations have been observed widely across Western Europe from the end of the last century to the beginning of this century, particularly in Switzerland, Belgium, Germany, Italy, Norway, and the Czech Republic (Harrison et al. 2008). According to reports, the assumption that exposure is uniform within each region may have resulted in errors in the estimation of exposure, especially for pollutants with important local sources (Hoek et al. 2002). One way to solve this problem is to use spatial interpolation to address the data collected from pollutant detection stations in the study area and to obtain a continuous plane of pollutant concentrations. The Geographic Information System (GIS) has achieved broad application as an interdisciplinary tool, especially in Environmental Science. Johansson et al. (2007) analyzed how differences in emissions, background concentrations, and meteorology affect the temporal and spatial distribution of PM<sub>10</sub> based on measurements and dispersion modelling in Stockholm, Sweden, and found that up to 90% of the locally emitted PM<sub>10</sub> may be due to road abrasion. Sampson et al. (2013) successfully predicted PM<sub>2.5</sub> concentrations at a fine spatial scale across the USA using regionalized universal kriging. Long et al. (2018) estimated sub district-level daily PM<sub>2.5</sub> concentrations by means of the block cokriging approach.

The population exposure study mainly includes two methods, the sampling method and the mathematical model method. The sampling method can be accurate to the individual exposure in different micro-environments, but the method has limited sampling points and large local environmental differences, resulting in a narrower range of applications. The mathematical model method is an exposure assessment method that combines pollution concentration distribution data and population distribution data and finally obtains a model with strong applicability. Jerrett et al. (2005) reviewed six classes of intraurban exposure models, including (i) proximity-based assessments, (ii) statistical interpolation, (iii) land use regression models, (iv) line dispersion models, (v) integrated emission-meteorological models, and (vi) hybrid models. Anenberg (2010) estimated the global burden of mortality due to O<sub>3</sub> and PM<sub>2.5</sub> from anthropogenic emissions using global atmospheric chemical transport model simulations of preindustrial and present-day (2000) concentrations to derive exposure estimates. Eeftens (2012) used land use regression (LUR) models for modeling small-scale spatial variation in air pollution concentrations and estimating individual exposure for participants of cohort studies. The more intraurban concentration studies and health studies are described in detail elsewhere (Jerrett et al. 2005).

However, studying human exposure to ambient air pollution, demographic data is also calculated mainly for the total

population of the examined city, ignoring the non-uniformity of the spatial distribution of pollutants and populations and their inherent time-varying characteristics. Kousa et al. (2002) presented a mathematical model for the determination of exposure to ambient air pollution in an urban area to evaluate the spatial and temporal variation of the average exposure of an urban population to ambient air pollution in different microenvironments with reasonable accuracy. The model developed has been designed to be utilized by municipal authorities in urban planning. Therefore, in this paper, the total population of the study area was not used as the exposed population. Rather, interpolated spatial data of the population after an investigation was used to obtain a continuous plane of population density. In addition, the spatial distribution of residents is inconsistent with the spatial distribution of air pollution. If air pollution concentrations are simply used to characterize pollution exposures in large spaces, true levels of exposure will not be reflected. Therefore, a population-weighted air pollution exposure model was proposed. The pollutant concentration was superimposed on the population data to obtain a population-weighted concentration. Then the population-weighted exposure level (PWEL) was used to estimate the potential exposure to PM<sub>10</sub> in Germany in 2017. In this way, different numbers of people in the grid were exposed to different concentrations of atmospheric pollutants.

Based on these knowledge, the objective of this study are to classify PM<sub>10</sub> air pollution for each month of 2017 based on the head/tail breaks classification method; to explore the population exposure to the distribution of PM<sub>10</sub> using geospatial statistical tools systematically; to portray the present situation of PM<sub>10</sub> air pollution with high PM<sub>10</sub> levels; and to test the distribution of PM<sub>10</sub> heavily polluted areas using the power-law distribution. These estimates may be useful in assessing health impacts through linkage studies and in communicating with the public and policymakers for potential intervention.

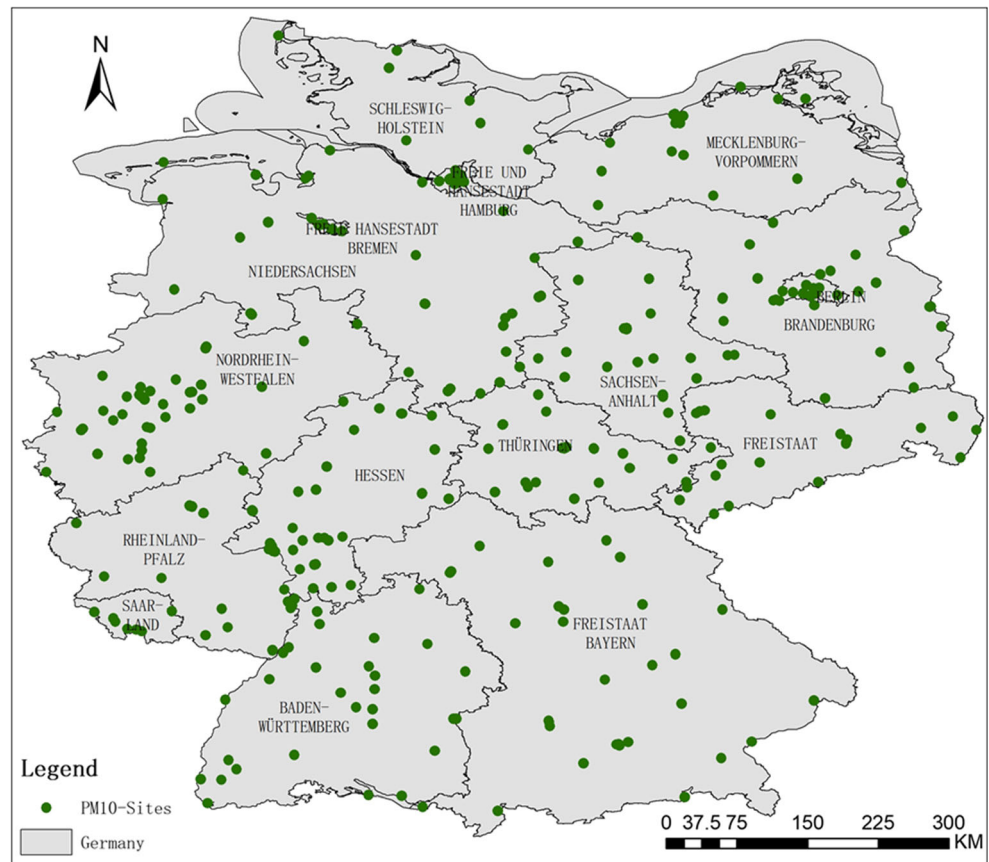
## Methods

### Monitoring data

Daily PM<sub>10</sub> concentration data was obtained from the German Federal Environment Agency (UBA) networks (<https://www.umweltbundesamt.de/en/>) and used to calculate monthly and annual means for each monitoring location. Only data that met the minimum inclusion criteria (within the detection limit) were included. The observations included the daily PM<sub>10</sub> concentrations at 379 monitoring sites in Germany in 2017 (Fig. 1), which were used to characterize the spatial variability of particle concentrations.

Data analysis was performed with the Data Processing System 9.5 (DPS) (Tang and Zhang 2013). For the spatial autocorrelation and agglomeration analysis, the GeoDA1.4.0

**Fig. 1** Locations of surface air pollution PM<sub>10</sub> monitoring stations in Germany during 2017



and Arc GIS 10.2 were used. Results are displayed with ArcView 4.0.

**Spatial analysis**

Spatial distributions of the average atmospheric pollutant concentrations for 2017 were simulated using ordinary kriging. Liao et al. (2006) showed that investigation of GIS approaches for estimating daily mean geocoded location-specific air pollutant concentrations supports the use of a spherical model to perform lognormal ordinary kriging on a national scale.

**Spatial autocorrelation**

Observations made at different locations may not be independent. For example, measurements made at nearby locations may be closer in value than measurements made at locations farther apart. This phenomenon is called spatial autocorrelation, which can be calculated using Moran’s I (Moran 1948; Geary 1954).

Spatial autocorrelation measures the correlation of a variable with itself through space. Spatial autocorrelation can be positive or negative. Positive spatial autocorrelation appears when similar values occur near one another, while there is negative spatial autocorrelation when dissimilar values occur near one another.

The spatial autocorrelation method for identifying the correlation of atmospheric pollutant profiles is described in detail elsewhere (Fang et al. 2016; Xu et al. 2016). Fang et al. (2016) covers global spatial autocorrelation (Global Moran’s I (GMI)) and local spatial autocorrelation (Local Moran’s I (LMI)) (Anselin 1995), which are defined as follows:

$$GMI = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n W_{i,j} Z_i Z_j}{\sum_{i=1}^n Z_i^2} \tag{1}$$

where  $Z_i$  is the deviation of an attribute for feature  $i$  from its mean ( $\chi_i - \bar{X}$ ),  $W_{i,j}$  is the spatial weight between features  $i$  and  $j$ ,  $n$  is equal to the total number of features, and  $S_0$  is the aggregate of all the spatial weights:

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n W_{i,j} \tag{2}$$

The local Moran’s I statistic is given as

$$LMI_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^n w_{i,j} (x_j - \bar{x}) \tag{3}$$

where  $x_i$  is an attribute for feature  $j$ ,  $\bar{x}$  is the mean of the corresponding attribute,  $W_{ij}$  is the spatial weight between feature  $i$  and  $j$ , and

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n W_{i,j}}{n-1} \bar{X}^2 \quad (4)$$

In general, Moran's  $I$  is similar but not equivalent to a correlation coefficient. Its value varies between  $-1$  and  $1$  and represents negative and positive spatial autocorrelation, respectively. Positive and significant Moran's  $I$  values indicate that nearby areas have similar spatial patterns, whereas negative values indicate the contrary. If the Moran's  $I$  is zero, the values are arranged randomly. GMI is used to judge the spatial agglomeration degree of Germany's  $PM_{10}$ , and LMI is used to examine the spatial distribution of the "hot spots" and "cold spots."

Since the Moran's  $I$  statistic follows a random distribution or a nearly normal distribution, the significance test can be converted into a  $Z$  value representing the normal distribution statistic. At the 5% significance level, if the  $Z$  value is greater than  $1.96$  or less than  $-1.96$ , there is a spatially significant positive or negative correlation between the observations, while if the  $Z$  value is between  $-1.96$  and  $1.96$ , the spatial correlation is not significant. The correlation is extremely significant when  $Z$  is greater than  $2.58$ . The standard statistic  $Z$  is calculated by following formula:

$$Z = \frac{I - E(I)}{\sqrt{\text{VAR}(I)}} \quad (5)$$

where  $E(I)$  and  $\text{VAR}(I)$  are expectation and variance of Moran's  $I$ , respectively, and  $E(I) = (-1)/(n-1)$ . When  $Z$  is positive and significant, it indicates a positive spatial correlation and the observed values are aggregated. On the contrary, when  $Z$  is negative and significant, indicating a negative correlation, the observations tend to be discretely distributed.  $Z = 0$  means that the observations are randomly distributed.

## Two-dimension graphics cluster

The graph theory clustering method, also known as the largest (small) support tree clustering algorithm, was first proposed by Zahn (1971). The traditional graph-theoretical clustering methods process the local area's joint characteristics of sample data as the main information. These methods then lead to most joint data in clusters not being settled. Initializing each grid cell as a class makes a large body of processing data. Based on spatial relation, a two-dimension graphics cluster could solve this problem. The method not only considers the relation between grid cells and their neighborhood regions but also makes classes depending on the uniform gray level of grid cells. The  $PM_{10}$  spatial distribution analysis based on the

two-dimensional graph theory clustering method combines the spatial analysis of ArcGIS with the tree algorithm of graph theory. The specific methods are as follows: first, the spatial coordinates of the monitoring points are extracted and the spatial autocorrelation analysis is performed. Then, in the DPS software, the minimum spanning tree is divided into  $n$  subtrees by using the two-dimensional graph theory minimum tree method, that is, the research area is divided into  $n$  regions. A network diagram of the relationship between the partition units is generated by means of the DPS (Data Processing System). Finally, the directly connected places indicate that the adjacency of each unit space and the comprehensive conditional similarity of each aspect are relatively high. When the  $PM_{10}$  pollution is controlled, a common control method can be adopted.

## Evaluation of the effect of $PM_{10}$ on human health based on GIS

### The rule of the head/tail division and power-law distribution

The head/tail breaks classification can be helpful for finding the ht-index of the data, which in turn is helpful for finding the inherent structure of the data. The ht-index based on this kind of classification approach is a special indicator to describe the complexity of the data, and it is an indicator of the underlying hierarchy of the data (Liu 2014). Liu (2014) has used nighttime lighting data and ht-index classification methods to extract the naturally defined urban boundaries constitute that we call natural cities. The method has been described in detail in Jiang and Yin (2014) and Liu (2014). Briefly, the rules for head/tail splitting determine the class or hierarchy of data using a heavy-tailed distribution, which can then be referred to as a scaling, hierarchy, or zoom level. The head of the heavy-tailed distribution contains a small number of large values, while the tail of the heavy-tailed distribution contains the majority of small values. For example, if a given data value is heavy-tailed, the average value of the data value can be divided into all the values to include the head portion of the data value equal to or greater than the average value. Otherwise, the data value should be included in the tail, and then the steps should be repeated until the data values in the head are no longer heavy-tailed. The main focus of the head and tail classification is low-frequency events. These events are more important than high-frequency events because low-frequency events have a larger impact than high-frequency events.

In summary, the head/tail classification ability captures the inherent level of data better than the natural rest classification. The head/tail classification is suitable for discovering the underlying hierarchy and the level of detail inherent in the data. The generalization of the data is now based more efficiently on the captured hierarchy.

Also, a heavy-tailed distribution is a common distribution pattern in dynamic and unbalanced nature (Mohajeri 2013). In this case, the data is heavily right skewed, with a minority of large values in the head and a majority of small values in the tail, and is commonly characterized by a power-law, a lognormal, or an exponential function (Jiang 2013). Therefore, to determine if a variable obeys a power-law distribution, it can be judged by plotting the data as a scatter plot on a double logarithmic scale. If the random variable fitting result is a linear function, then the cumulative probability obeys the power-law distribution. For more detail on power-law distribution, see the published literature by Anderson (2005).

### Estimation of the population exposure to PM<sub>10</sub>

Population is an important indicator for measuring the PWEL (Hao et al. 2012). The population data, obtained from the Federal Statistical Office of Germany, was initially analyzed (Bundesamt 2016). However, the spatial distribution of residents in space is inconsistent with the spatial distribution of atmospheric pollution (Ivy et al. 2008). So, if air pollution concentrations are simply used to characterize pollution exposure in a large space, the true level of exposure to residents will not be captured. Therefore, the population-weighted exposure level model is used (Sun et al. 2013). Given grid (1.0 km•1.0 km), *i*, the population weighted exposure equation is as follows:

$$PWEL = \sum(P_i \times C_i) / \sum P_i \tag{6}$$

where *P<sub>i</sub>* is the population in the grid cell, *i*, and *C<sub>i</sub>* is its average PM<sub>10</sub> concentration in the grid cell.

In our study, we used 50 μg/m<sup>3</sup> as pollution concentration, which is the PM10 limit value corresponding to the mean daily PM<sub>10</sub> concentration. With this value, we calculated the number of exposure days exceeding the standard during the research period (which can indicate the exposure intensity in different parts of the area).

### Spatial distribution of PM<sub>10</sub> pollution exposure

Population PM<sub>10</sub> exposure intensity equals the duration of population exposure per unit area exposed to severe PM<sub>10</sub> pollution. The formula for PM<sub>10</sub> exposure intensity is as follows:

$$N_{ij}^T = \sum_{t=t_1}^{t_2} 1(C_{ijt} \geq STD) \tag{7}$$

$$EI_{ij} = N_{ij}^T \frac{P_{ij}}{A_{ij}} \tag{8}$$

where *A<sub>ij</sub>* is the area of the (i,j)th grid cell (1.0 km•1.0 km), *P<sub>ij</sub>* is the population of the (i,j)th grid cell, and *N<sub>ij</sub><sup>T</sup>* is the accumulated pollution days of the (i,j)th grid cell. *EI<sub>ij</sub>* is the

pollution exposure intensity of the (i,j)th grid cell at time *T*, in day-person/km<sup>2</sup>. STD = 50 μg/m<sup>3</sup> for the EU standard. As shown in the formula above, the intensity of pollution exposure is directly related to the accumulated days of pollution and population density.

## Results and discussion

### Pollution area development and changes in Germany

Firstly, in order to reasonably divide the area of high pollution and the area of severely polluted areas reflecting rich urban information, this section used the average of the first and last divisions of PM<sub>10</sub> pollution concentration in 2017 to extract the high pollution area of the PM<sub>10</sub> air pollution concentration. The extraction results were compared and analyzed. As shown in Fig. 2, the results of the concentration distribution area were extracted using the first layer to the third layer of the average concentration of PM<sub>10</sub>. According to the comparison and analysis, when the extraction value was 17.20 μg/m<sup>3</sup>, the extraction range was too wide, and many cities were connected. When the extraction value was 18.47 μg/m<sup>3</sup>, the extraction effect captured the details of individual cities. However, when the extraction value was 19.38 μg/m<sup>3</sup>, the extraction effect was poor, and more details were ignored. Therefore, the average value of 18.47 μg/m<sup>3</sup> was selected as the PM<sub>10</sub> pollution area division value. In order to better compare the extracted values, this paper also analyzed the values based on the WHO standard (20 μg/m<sup>3</sup> for PM<sub>10</sub>).

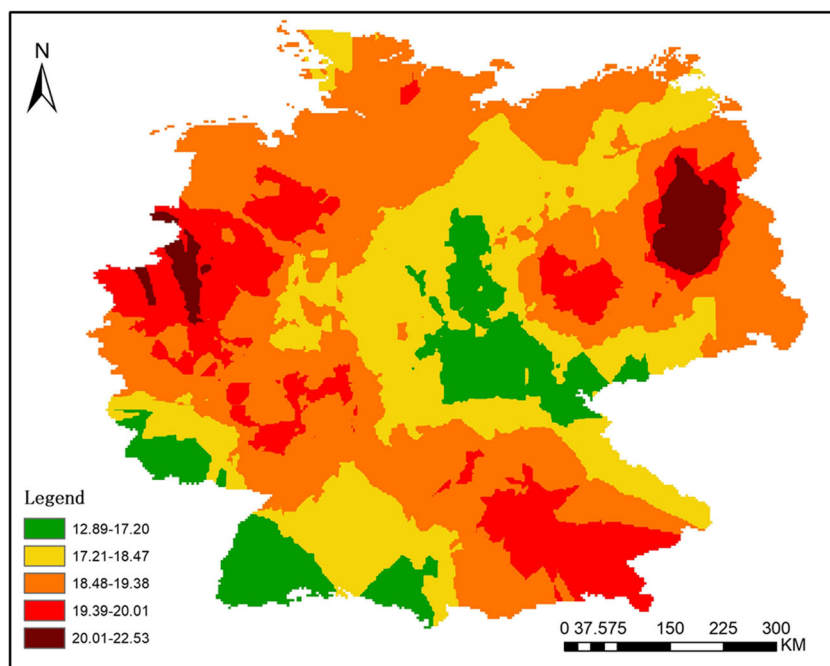
As shown in Fig. 2, higher PM<sub>10</sub> concentrations occurred in the states of Bavaria, North Rhine-Westphalia, Saxony-Anhalt, and Berlin. In these areas, PM<sub>10</sub> concentrations exceeded the value of 18.47 μg/m<sup>3</sup>.

Secondly, 18.47 μg/m<sup>3</sup> was used as the PM<sub>10</sub> pollution area division value to extract the PM<sub>10</sub> pollution area range of all Germany for each month in 2017 (Fig. 3). Among them, the pollution range from January to March was relatively wide, and the pollution distribution range was relatively small in the other months. In June, the PM<sub>10</sub> concentration was below the value throughout all of Germany. In 2017, PM<sub>10</sub> pollution was mainly concentrated in North Rhine-Westphalia, Brandenburg, and Berlin; in these areas, the pollution was the most widely distributed and lasted permanently.

### Spatial autocorrelation of PM<sub>10</sub> concentration

To better reflect the spatio-temporal changes of the PM<sub>10</sub> concentrations in Germany, the Global Moran’s *I* was employed to identify the spatial autocorrelation of the PM<sub>10</sub> concentrations. Further, the analysis of spatial autocorrelation could help to determine whether sparse ground monitoring sites can meet the requirements.

**Fig. 2** Distribution of PM<sub>10</sub> concentrations based on different extraction values



Moran's I scatter plots of the PM<sub>10</sub> concentrations in Germany using GeoDA are presented in Table 1. Results showed that, for PM<sub>10</sub>, Moran's I exceeded 0 for monthly values in 2017 and ranged from 0.228 to 0.443, suggesting there was a significant positive spatial autocorrelation of PM<sub>10</sub> concentrations in Germany. The average Z(I) of 12 months was 9.65, which exceeded 2.58, suggesting that there was a significant positive spatial autocorrelation of the PM<sub>10</sub> concentrations in the cities of Germany (high-high or low-low agglomeration).

In the analysis of the spatial autocorrelation model based on ArcGIS10.2, the confidence level bin (Gi\_Bin) was used to identify statistically significant hot spots and cold spots. A confidence interval of +3 to −3 elements reflected a confidence level of 99% statistical significance; a confidence interval of +2 to −2 elements reflected a confidence level of 95% statistical significance; a confidence interval of −1 to +1 elements reflected a confidence level of 90% statistical significance; and a confidence interval of 0 clustering was not distinctly significant. As shown in Fig. 4, analysis of annual and monthly agglomerations indicated that hot spots occurred mainly in North Rhine-Westphalia, Brandenburg, and Berlin. Therefore, these areas should be considered when formulating air pollution control measures. Cold spots occurred mostly in the south of Bremen Niedersachsen and Thüringen and in the south-west of Baden-Wuerttemberg. No significant spots were found in Schleswig-Holstein, Mecklenburg-Vorpommern, Sachsen, and Bavaria.

### Two-dimension graphics cluster in elevated pollution districts

In five states of Germany, Bavaria, North Rhine-Westphalia, Saxony-Anhalt, Brandenburg, and Berlin, the PM<sub>10</sub> levels were relatively much higher than other states in 2017. Henceforth, this paper analyzed these five states based on the theory of two-dimension graphics cluster.

In Fig. 5, the significance of these connections indicates the spatial adjacency and intrinsic similarity between the various points (monitoring sites). Spatial adjacency represents geographic location information, while intrinsic similarity expresses the degree of similarity of pollution concentrations. With this, a relatively large-scale comprehensive consolidation project can be formed upon the determination of PM<sub>10</sub> pollution levels and management plans. Meanwhile, using the GIS spatial analysis method, the mathematical clustering analysis method, and the graph theory tree algorithm combined with the regional division analysis technology, four functional pollution areas in Germany were obtained. First, the cities of Burg, Bitterfeld, Bernburg, and Magdeburg constituted the Saxony-Anhalt pollution region; second, the cities of Berlin, Potsdam, Hasenholz, Eberswalde, Bernau, and Blankenfelde constituted the Berlin-Brandenburg pollution region; third, the cities of Essen, Hürth, Düsseldorf, Dortmund, and Solingen Konradas constituted the North Rhine-Westphalia pollution region; and lastly, the cities of Augsburg, Regensburg, Munich and Sulzbach-Rosenberg constituted the Bavaria pollution region.

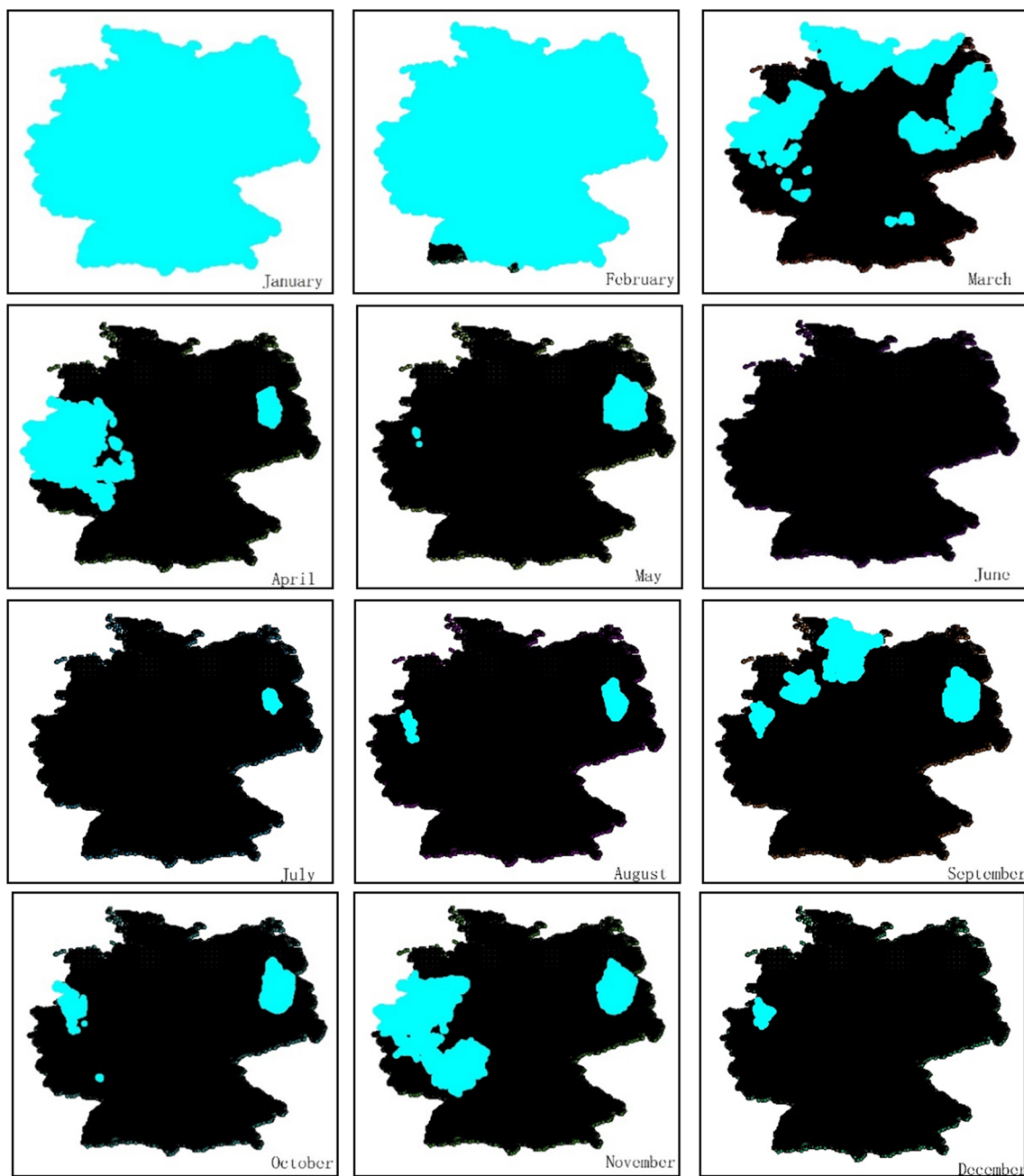


Fig. 3 The PM<sub>10</sub> contaminated area range for each month of 2017 in Germany, based on the value of 18.47 μg/m<sup>3</sup>

**Spatialization of demographic data**

According to demographic data and district-county-level boundary data, statistics on population density at the

county level in Germany show that the population density of most districts and counties in the country are relatively small; only a few counties have high population density. In terms of spatial distribution, the Moran’s I index can be

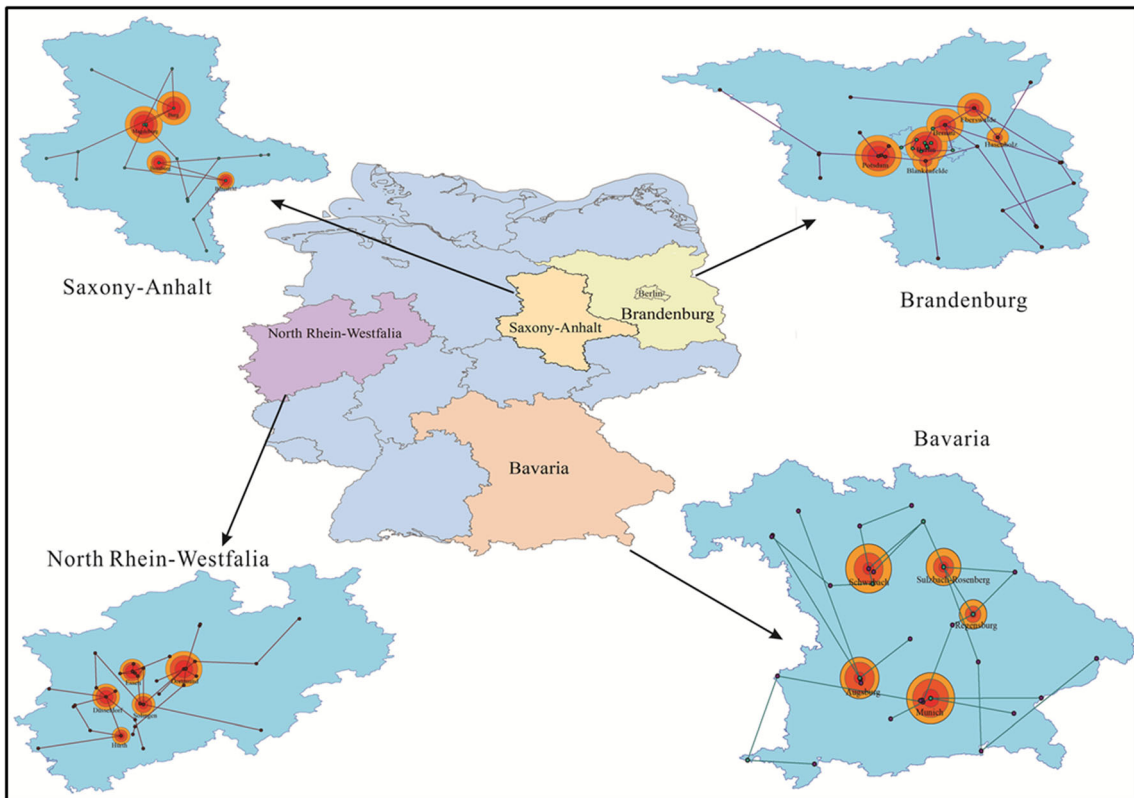
**Table 1** Spatial autocorrelation index of PM<sub>10</sub> concentrations in Germany in 2017

Monthly	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
Moran’s I-PM <sub>10</sub>	0.347	0.443	0.228	0.345	0.342	0.286	0.298	0.244	0.369	0.248	0.282	0.296
P	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Z	10.456	13.478	7.181	10.637	10.652	8.747	9.386	7.630	11.545	7.864	8.843	9.325



**Fig. 4** Spatial agglomeration of annual and monthly PM<sub>10</sub> concentrations in the monitoring sites in Germany, in 2017





**Fig. 5** Two-dimension graphics cluster in districts with higher pollution, based on the data of PM<sub>10</sub> in Germany in 2017

used to analyze the spatial correlation of population distribution across the county.

Moran’s I scatter plots of the population density in Germany using GeoDA are presented following Fig. 6. Results showed that, for population density, Moran’s I was 0.555, suggesting that the area with high population density has a certain aggregation effect.

**Variation of the total PWEL and MEAN concentration**

Combining the PM<sub>10</sub> concentrations and population spatialization results, the daily total population exposure to PM<sub>10</sub> (PWEL) and the average of the PM<sub>10</sub> concentrations were calculated (shown in Fig. 7a). The trend of PWEL was similar to the PM<sub>10</sub> average in all Germany areas. Meanwhile, the monthly PWEL and PM<sub>10</sub> average were also calculated to show their monthly trend characteristics more intuitively (Fig. 7b). The concentrations were higher in January and February, reaching average values of 29.14 and 28.93 μg/m<sup>3</sup>, respectively, indicating humans could be adversely affected by exposure to air pollutants in winter. Certainly, it is worth mentioning that the population exposures to PM<sub>10</sub> in other months were lower than 20 μg/m<sup>3</sup>, suggesting a high ambient air quality in Germany.

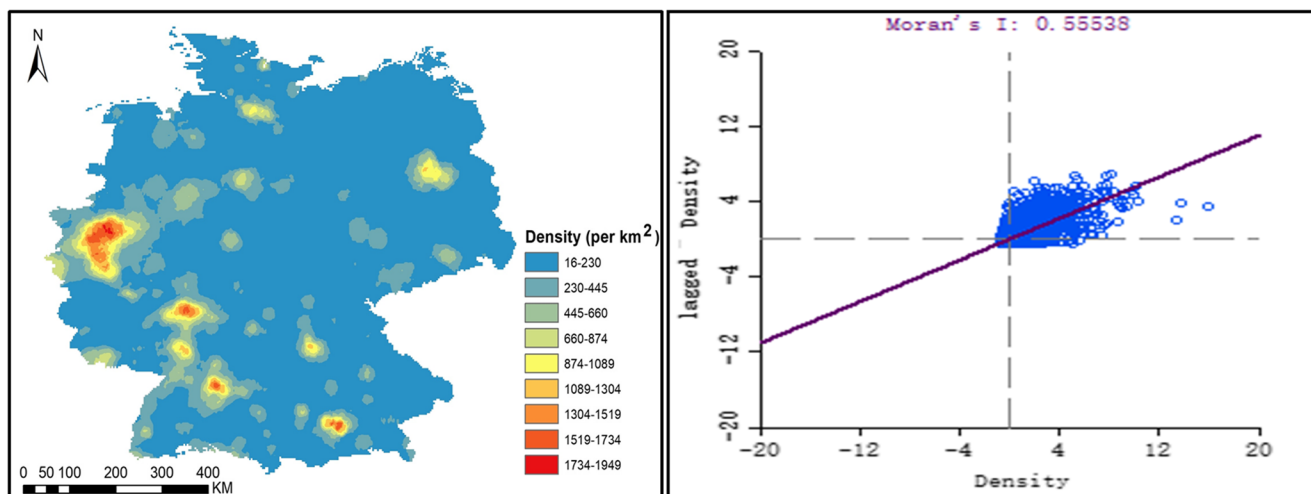
Figure 1S shows the cumulative exposure days exceeding the standard value provided by the EU and the

cumulative population exposure to PM<sub>10</sub> concentrations. Figure 1Sa demonstrates days with extremely high exposure in south and east of Germany, mainly belonging to Bavaria, Brandenburg, and Berlin, reaching 17 exposure max days. Figure 1Sb shows that the distribution of cumulative population exposure to PM<sub>10</sub> was similar to the distribution of cumulative exposure time.

**Spatial distribution of PM<sub>10</sub> pollution exposure in Germany**

In Section 3.5, the time evolution of the German PM<sub>10</sub> pollution exposure level in 2017 was discussed. This section introduces the concept of population pollution exposure intensity to further explore the spatial interaction and distribution of the population PM<sub>10</sub> pollution distributions.

The results show that long-term exposure to ambient particulate matter air pollution is associated with natural-cause mortality, even in concentration ranges far below the present European annual mean limit value (Beelen et al. 2014). In addition, the original air pollution concentrations cannot clearly reflect the spatial extent of population exposure to air pollution. The clustered regions extracted by the ht-index method are not affected by human subjective cognition based on legal and administrative factors. Their border ranges differ from the administrative border. In this case, they can show the high concentration ranges of



**Fig. 6** Population density in Germany and the corresponding Moran's I scatter plots (data obtained from the Federal Statistical Office of Germany in 2016 and population density is expressed as inhabitants per km squared)

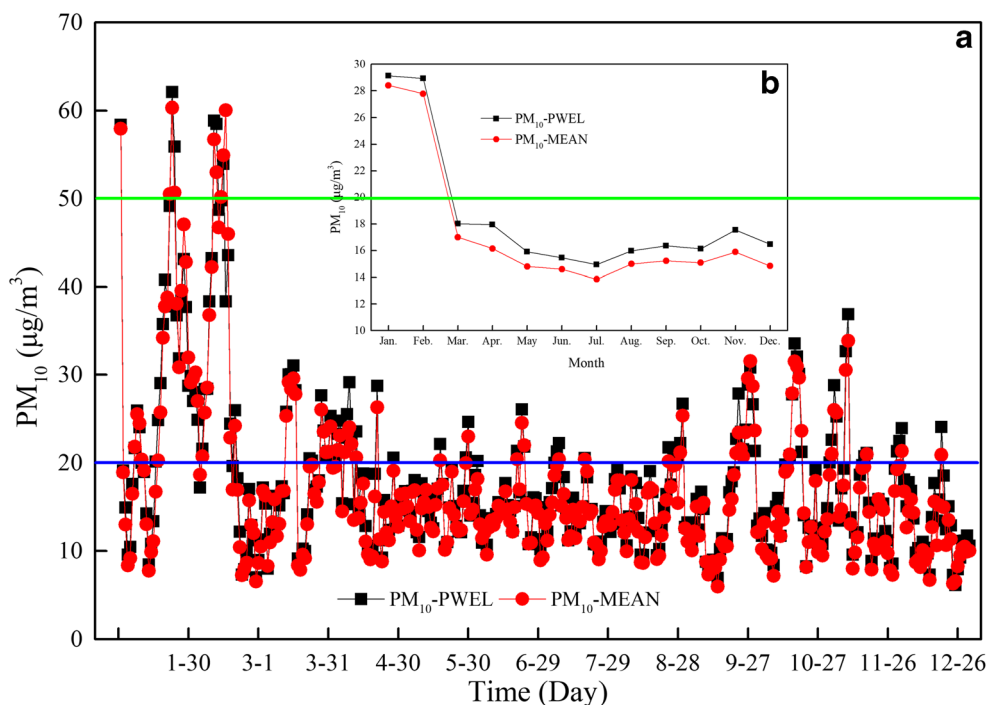
human activity and pollution concentration distribution. For this reason, the distribution of exposure intensity was analyzed based on the head/tail breaks classification method.

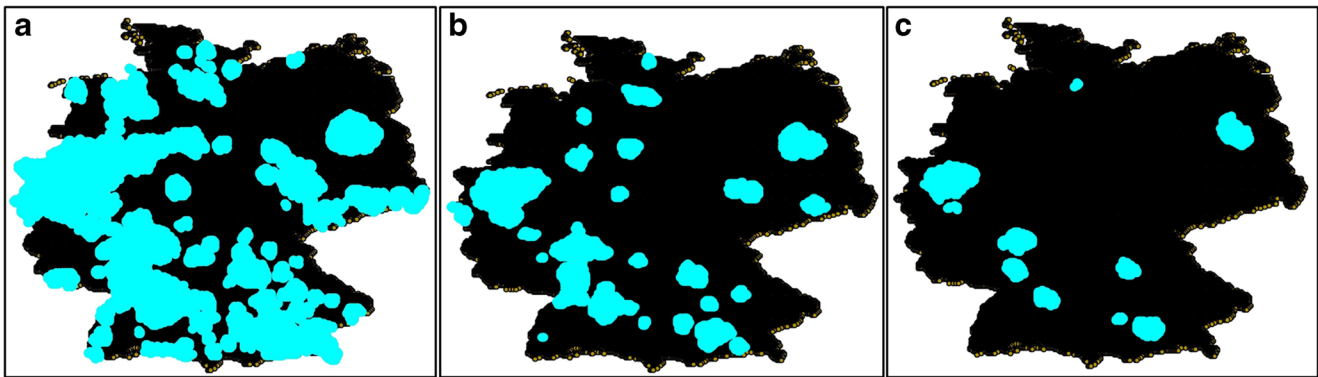
Firstly, using 50% as threshold, the population distribution of PM<sub>10</sub> pollution was statistically analyzed with the head/tail breaks classification method. According to the calculations, the ht-index of the PM<sub>10</sub> pollution exposure of the entire German population is 12. Subsequently, the appropriate aggregation area was determined as the extraction value. According to the mean value of each level, the distribution

of the population's PM<sub>10</sub> pollution exposure was classified, and the above-average range was extracted. In line with the daily standard of WHO and EU, Fig. 8a, b, and c show the effect of clustering regions at means 1912.71 day·person/km<sup>2</sup>, 4307.37 day·person/km<sup>2</sup>, and 7957.98 day·person/km<sup>2</sup>, respectively.

Comparing the extraction results of the three maps, the ranges of the aggregated areas shown in Fig. 8a are quite vague. The urban details shown in Fig. 8c are too few, and the boundaries of the clustered regions are significantly less than the boundaries of the administrative borders.

**Fig. 7** The daily (a) and monthly (b) variations in the total PWEL and MEAN concentration during 2017





**Fig. 8** Population exposure intensity distribution (a, value = 1912.71 day-person/km<sup>2</sup>; b, value = 4307.37 day-person/km<sup>2</sup>; c, value = 7957.98 day-person/km<sup>2</sup>)

The city boundaries in Fig. 8b show a quality extraction effect. Therefore, this section selects 4307.37 (day-person/km<sup>2</sup>) for the daily WHO and EU standard as the final aggregation area extraction values. As seen in Fig. 8b, the details are similar. The two largest areas of PM<sub>10</sub> pollution are located in the surrounding areas of Berlin and Düsseldorf.

Liu (2014) examined the extraction of natural cities based on nightlight data in 2014 and suggested that the scope of natural cities is closely related to the scope of the population’s living and activities. Based on the head/tail breaks classification method, the extraction of nocturnal natural cities was realized. Liu (2014) found that in most cities there is some overlap between the natural city boundaries and administrative boundaries. Combining the concepts of natural cities with Fig. 8, the polluted areas in Germany can be used to describe the “PM<sub>10</sub> pollution exposure and aggregation area of the population.”

**Power-law distribution test**

Various phenomena governed by power-law distributions are ubiquitous in nature and in society; thus, their study carries broad and far-reaching significance. In this section, power-law distribution tests are performed on the area distribution of PM<sub>10</sub> pollution accumulation. By using Pareto distribution analysis, PM<sub>10</sub>-gathered area data can be fitted by a power-law distribution in log-log coordinates. The power-law distribution is simulated as  $f(x) = 14.25x^{-0.82}$  ( $R^2 = 0.7$ ). From the power-law detection results, it can be concluded that the area distribution of PM<sub>10</sub> clustering area has a significant heavy-tailed distribution and obeys the power-law distribution.

**Conclusions**

In summary, this paper investigated the spatio-temporal distribution of PM<sub>10</sub> air pollution with GIS techniques, analyzed the multi-source spatio-temporal data collected from UBA networks, and obtained additional meaningful

discoveries according to PM<sub>10</sub> spatialization results. The national average PM<sub>10</sub> amounted 17.20 µg/m<sup>3</sup> in 2017 and was much lower than the annual mean PM<sub>10</sub> concentration of 40 µg/m<sup>3</sup> as declared by the European Union, and only slightly lower than the standard concentration of the WHO, which amounts for 20 µg/m<sup>3</sup>, indicating humans could be adversely affected by exposure to air pollutants in the ambient air. In terms of spatial distribution, PM<sub>10</sub> pollution was mainly concentrated in North Rhine-Westphalia, Brandenburg, and Berlin. In terms of time distribution, PM<sub>10</sub> pollution exhibited a distinct seasonal characteristic, i.e., PM<sub>10</sub> pollution in winter was higher than in summer and lowest in June.

Combined with the concept of a natural city, this paper proposes to use the PM<sub>10</sub> air pollution accumulation area to describe the current status of PM<sub>10</sub> air pollution. The German PM<sub>10</sub> air pollution accumulation area extraction results show that the German PM<sub>10</sub> pollution accumulation area distribution meets the power-law distribution characteristics; that is, “a small area of pollution agglomeration area is much larger than the number of large areas of pollution agglomeration area.” The largest contaminated areas of concentration are in Berlin, Dusseldorf, and Munich but are not entirely within the city limits.

The population distribution by exposure level shows that the majority of people is living in polluted areas. Also, the exposure level changes greatly from North to South, and each sub-district maintains similarity to neighboring sub-districts. In addition, PM<sub>10</sub> pollution aggregation area boundaries and administrative boundaries do not coincide, even across multiple administrative city boundaries. Therefore, PM<sub>10</sub> air pollution control is not only the task of a single city but should be a joint effort by multiple cities in order to govern effectively.

**Acknowledgments** The authors also like to thank Mr. Thomas Himpel from Federal Environment Agency, for providing the data of air pollution in Germany.

**Funding information** Xiansheng Liu's Ph.D. work is funded by the China Scholarship Council (CSC) under the State Scholarship Fund (File No.201706860028).

## References

- Anderson G, Ge Y (2005) The size distribution of Chinese cities. *Reg Sci Urban Econ* 35(6):756–776
- Anenberg S C, Horowitz L W, Tong D Q, West J J (2010) An estimate of the global burden of anthropogenic ozone and fine particulate matter on premature human mortality using atmospheric modeling. *Environ Health Perspect* 118(9):1189–1195
- Anselin L (1995) Local indicators of spatial association: LISA. *Geogr Anal* 2(27):93–115
- Beelen R, Raaschou-Nielsen O, Stafoggia M, Andersen Z J, Weinmayr G, Hoffmann B, Vineis P (2014) Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. *Lancet* 9919(383):785–795
- Boldo E, Medina S, Le Tertre A, Hurley F, Mücke HG, Ballester F, Aguilera I, Eilstein D, Apheis Group (2006) Apheis: Health impact assessment of long-term exposure to PM(2.5) in 23 European cities. *Eur J Epidemiol* 21(6):449–458
- Bundesamt S (2016) Gemeinden in Deutschland nach Fläche, Bevölkerung und Postleitzahl am 30.09.2016. Erscheinungsmonat: August 2017 (*in German*)
- Burnett R T, Pope III C A, Ezzati M, Olives C, Lim S S, Mehta S, Anderson H R (2014) An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environ Health Perspect* 122(4):397–403
- Eeftens M, Beelen R, de Hoogh K, Bellander T, Cesaroni G, Cirach M, Dimakopoulou K (2012) Development of land use regression models for PM<sub>2.5</sub>, PM<sub>2.5</sub> absorbance, PM<sub>10</sub> and PM<sub>coarse</sub> in 20 European study areas; results of the ESCAPE project. *Environ Sci Technol* 46(20):11195–11205
- Fang C, Wang Z and Xu G (2016) Spatial-temporal characteristics of PM<sub>2.5</sub> in China: A city-level perspective analysis. *J Geogr Sci* 26(11):1519–1532
- Geary R C (1954) The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician* 5 (3):115
- Hao Y, Flowers H, Monti M M, Qualters J R (2012) US census unit population exposures to ambient air pollutants. *Int J Health Geogr* 11(1):3
- Harrison RM, Stedman J, Derwent D (2008) New directions: why are PM<sub>10</sub> concentrations in Europe not falling? *Atmos Environ* 42:603–606
- Hoek G, Brunekreef B, Goldbohm S, Fischer P, van den Brandt P (2002) Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. *Lancet* 360(9341):1203–1209
- Ivy D, Mulholland J A, Russell A G (2008) Development of ambient air quality population-weighted metrics for use in time-series health studies. *J Air Waste Manage Assoc* 58(5):711–720
- Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahsuvaroglu T, Morrison J, Giovis C, (2005) A review and evaluation of intraurban air pollution exposure models. *J Air Waste Manage Assoc* 15(2):185–204
- Jiang B (2013) Head/tail breaks: a new classification scheme for data with a heavy-tailed distribution. *Prof Geogr* 65(3):482–494
- Jiang B and Yin J (2014) Ht-Index for quantifying the fractal or scaling structure of geographic features. *Ann Assoc Am Geogr* 3(104):530–540
- Johansson C, Norman M, Gighagen L (2007) Spatial & temporal variations of PM<sub>10</sub> and particle number concentrations in urban air. *Environ Monit Assess* 1–3(127):477–487
- Kousa A, Kukkonen J, Karppinen A, Aarnio P, Koskentalo T (2002) A model for evaluating the population exposure to ambient air pollution in an urban area. *Atmos Environ* 36(13):2109–2119
- Liao D, Peuquet DJ, Duan Y, Whitsel EA, Dou J, Smith RL, Lin HM, Chen JC, Heiss G (2006) GIS approaches for the estimation of residential-level ambient PM concentrations. *Environ Health Perspect* 114(9):1374–1380
- Liu Q (2014) A Case Study on the Extraction of the Natural Cities from Nightlight Image of the United States of America. Master thesis, University of Gävle Malevergne Y, Pisaren
- Long Y, Wang J, Wu K, Zhang, J (2018) Population exposure to ambient PM<sub>2.5</sub> at the subdistrict level in China. *International journal of environmental research and public health* 15(12):2683
- Mohajeri N, French J R, Batty M (2013) Evolution and entropy in the organization of urban street patterns. *Ann GIS* 19(1):1–16
- Moran (1948) The interpretation of statistical maps. *J R Stat Soc B* 2(10):243–251
- World Health Organization (2002) The World Health Report 2002: Reducing Risks, Promoting Healthy Life, WHO, Geneva, Switzerland
- Samet J M, Dominici F, Currier F C, Coursac I, Zeger S L (2000) Fine particulate air pollution and mortality in 20 US cities, 1987–1994. *N Engl J Med* 343(24):1742–1749
- Sampson P D, Richards M, Szpiro A A, Bergen S, Sheppard L, Larson T V, Kaufman J D (2013) A regionalized national universal kriging model using Partial Least Squares regression for estimating annual PM<sub>2.5</sub> concentrations in epidemiology. *Atmos Environ* 75:383–392
- Sun Z, An X, Tao Y, Hou Q (2013) Assessment of population exposure to PM<sub>10</sub> for respiratory disease in Lanzhou (China) and its health-related economic costs based on GIS. *BMC Public Health* 13(1):891
- Tang QY, Zhang CX (2013) Data Processing System (DPS) software with experimental design, statistical analysis and data mining developed for use in entomological research. *Insect Sci* 2(20):254–260
- World Health Organization (2014) Burden of disease from ambient air pollution for 2012. Copenhagen: WHO Regional Office for Europe
- Xu G, Jiao L, Zhao S, Cheng J (2016) Spatial and temporal variability of PM<sub>2.5</sub> concentration in China. *Wuhan Univ J Nat Sci* 21(4):358–368
- Zahn CT (1971) Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Trans Comput* 1(20):68–86

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