Quantifying the influence of natural and socioeconomic factors and their interactive impact on PM$_{2.5}$ pollution in China* 

Dongyang Yang $^a$, Xiaomin Wang $^b$, Jianhua Xu $^a$,*, Chengdong Xu $^c$,**, Debin Lu $^a$, $^d$, Chao Ye $^e$, Zujing Wang $^f$, Ling Bai $^f$

$^a$ School of Geographic Sciences & Institute of Eco-Chongming, East China Normal University, Shanghai 200241, China
$^b$ School of Geography, Beijing Normal University, Beijing 100875, China
$^c$ State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Science and Natural Resource Research, Chinese Academy of Sciences, Beijing 100101, China
$^d$ Department of Tourism and Geography, Tongren University, Tongren, Guizhou Province 554300, China
$^e$ School of Environmental Science and Engineering, Suzhou University of Science and Technology, Suzhou, Jiangsu Province 215009, China
$^f$ School of Economics and Management, Nanchang University, Nanchang 330031, China

** Article info

Article history:
Received 7 February 2018
Received in revised form 14 March 2018
Accepted 14 May 2018

Keywords:
PM$_{2.5}$
Natural and socioeconomic factors
Interaction effect
GeogDetector

** Abstract

PM$_{2.5}$ pollution is an environmental issue caused by multiple natural and socioeconomic factors, presenting with significant spatial disparities across mainland China. However, the determinant power of natural and socioeconomic factors and their interactive impact on PM$_{2.5}$ pollution is still unclear. In the study, the GeogDetector method was used to quantify nonlinear associations between PM$_{2.5}$ and potential factors. This study found that natural factors, including ecological environments and climate, were more influential than socioeconomic factors, and climate was the predominant factor ($q = 0.56$) in influencing PM$_{2.5}$ pollution. Among all interactions of the six influencing factors, the interaction of industry and climate had the largest influence ($q = 0.66$). Two recognized major contaminated areas were the Tarim Basin in the northwest region and the eastern plain region; the former was mainly influenced by the warm temperate arid climate and desert, and the latter was mainly influenced by the warm temperate semi-humid climate and multiple socioeconomic factors. The findings provided an interpretation of the influencing mechanisms of PM$_{2.5}$ pollution, which can contribute to more specific policies aimed at successful PM$_{2.5}$ pollution control and abatement.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

PM$_{2.5}$ pollution's negative effects on public health have burdened China severely in recent years (Butt et al., 2017; Lu et al., 2015). As a geographical phenomenon, PM$_{2.5}$ pollution is inevitably influenced by the comprehensive geographical environment, including natural and socioeconomic conditions (Guo et al., 2014; Timmermans et al., 2017; Wang et al., 2014). China is characterized as a vast territorial area, and the natural and socioeconomic conditions and their interactions differ widely across regions, simultaneously, PM$_{2.5}$ pollution presents geographically discrepant distribution in China (Ma et al., 2016).

Multidisciplinary researchers have conducted numerous studies on the sources and driving factors of PM$_{2.5}$. Researchers have recognized industry, fossil fuel combustion, motor vehicles, and building yards as general anthropogenic sources of PM$_{2.5}$ in many cities in the world (Chowdhury et al., 2007; de Miranda et al., 2012; Huang et al., 2014; Marcazzan et al., 2003; Timmermans et al., 2017). Social and economic activities can affect the generation of and changes in these sources. Thus, some researchers have examined the relationships between PM$_{2.5}$ pollution and socioeconomic factors and indicated that socioeconomic factors such as economic growth, urbanization, industrialization, and others drove increases in PM$_{2.5}$ concentrations (Guan et al., 2014; Hao and Liu, 2016; Li et al., 2016; Meng et al., 2015).

PM$_{2.5}$ is a type of air pollutant and the transformation, diffusion,
and even the generation of areas severely polluted with PM$_{2.5}$ are inevitably affected by meteorological conditions. Generally, PM$_{2.5}$ pollution deteriorates and forms a thick haze in continuous and stable weather conditions (Green et al., 2015; Tao et al., 2014; Zhang and Cao, 2015). Long-term meteorological conditions are also related to the spatial and seasonal variability of PM$_{2.5}$, especially in China (Guo et al., 2017; Pant et al., 2015; Yang et al., 2017a; Zhang and Cao, 2015). The ecology or land use environment is the spatially common carrier of pollution sources and potential factors of PM$_{2.5}$; different ecological systems can also affect PM$_{2.5}$. Human settlements are undoubtedly concentrated primary and secondary anthropogenic emission sources (Cao et al., 2016; Guo et al., 2014; Zhang et al., 2015). Environmental chemists have also documented that fertilizer use on farmlands is the largest emission source of ammonia (NH$_3$), which mainly contributes to urban PM$_{2.5}$ by combining with urban emission pollutants, such as volatile organic compounds, NO$_x$, and SO$_2$ (Gu et al., 2014; Wang et al., 2016a). Furthermore, air pollutants in city suburbs aggregate into the city and enhance PM$_{2.5}$ pollution under the urban heat islands effect (Aslam et al., 2017; Bloomer et al., 2009; Zhang et al., 2009). Other ecological systems also have different effects on PM$_{2.5}$; for example, desert areas are natural sources of sand and dust in PM$_{2.5}$ (Lu et al., 2014), and forests can absorb and purify PM$_{2.5}$ pollution (Nowak et al., 2014).

PM$_{2.5}$ is comprehensively affected by natural and socioeconomic factors, and the use of related factors have been highlighted with effectiveness in improving PM$_{2.5}$ mapping (Liu et al., 2018). However, the discrepant influences of the factors and their interactions have rarely been examined. Investigators are usually concerned with unilateral factors such as natural or socioeconomic factors, while a quantification of the influences of integrated natural and socioeconomic multi-factors has been overlooked. Traditional methods are disadvantaged in examining the interactions of factors influencing air pollution. The interaction of two factors can be multiple coupling forms in reality, while it is usually the product of two factors in traditional regression methods. Additionally, statistical models of interactions are usually created using local regions and are limited in reflecting the spatial global variability of influences (Pearce et al., 2011). Coefficients with spatial differences can be derived using GIS-based regression methods or machine learning algorithms, but they have poor large-scale explanatory capability because of the existence of spatial stratified heterogeneity (Lin et al., 2014; Wang et al., 2016b).

China is experiencing tremendous economic growth and has increasingly become an important engine for world economic growth with its rapid urbanization and industrialization; it is also one of the main regions facing serious PM$_{2.5}$ pollution. Anthropogenic emissions drive the increases in PM$_{2.5}$, but not all concentrated areas of anthropogenic emissions are highly polluted because of the differences in natural conditions (Jin et al., 2017; Lu et al., 2017). Quantifying the influences of potential factors and their interactions are important for understanding the spatial pattern of PM$_{2.5}$ pollution and the driving mechanisms, which can contribute to successful policy making for controlling and reducing PM$_{2.5}$ pollution. This study aimed to quantitatively investigate the influence of potential factors, including industry, construction, traffic, coal combustion, ecological environments, and climate and their interactive effects on PM$_{2.5}$ pollution in China.

2. Materials and methods

2.1. PM$_{2.5}$ data

PM$_{2.5}$ data is a subset of a published global PM$_{2.5}$ concentration dataset for the study region in 2015 from the Atmospheric Composition Analysis Group at Dalhousie University (http://fizz.phys.dal.ca/~atmos/martin/?page_id=140) (Fig. 1). The spatial resolution of the raw data is 0.1° × 0.1°. This dataset has good accuracy with a high cross-validated R$^2$ of 0.81 (Van Donkelaar et al., 2016) and has been used in many research studies (Donkelaar et al., 2015; Lu et al., 2017; Mcquinn et al., 2017; Xie et al., 2016). The cross-validated R$^2$, between the estimated annual average PM$_{2.5}$ and observation values in 313 cities of China in 2015, is 0.72.

2.2. Potential driving factors, proxies, and data

Given that socioeconomic factors, including industry, construction, traffic, coal combustion, and natural factors, including ecological environment and climate, were related to aspects of the natural and anthropogenic sources, influences, and secondary influences of PM$_{2.5}$, we integrated these factors in this study. The corresponding proxies of these potential factors are shown in Fig. 2.

Socioeconomic data, including industrial output (IO), construction output (CO), and coal consumption (CC) in 2015 were obtained from the Provincial Statistical Yearbook on China National Knowledge Infrastructure (CNKI: http://data.cnki.net/Yearbook). The unit area mean values of the three types of socioeconomic data, which are the original values divided by the administrative area, were calculated and used. Data on roads were gathered from OpenStreetMap (https://www.openstreetmap.org). The roads included urban roads and multistage highways such as expressways; national, provincial, and county highways; and township roads. Road density (RD) was calculated using “line density” in spatial analysis tools in ArcGIS software. To match the following modeling needs, industrial output, construction output, and road density were discretized and classified using the geometrical interval method in ArcGIS according to their quantitative values, and coal consumption was classified using the natural breaks method (Fig. 3a–d).

Data on ecosystem types (ET) in 2015 were provided by the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (http://www.resdc.cn). The spatial resolution of the raw data was 1 km × 1 km. To obtain the ecosystem types in each 10 km × 10 km grid, we extracted the most common ecosystem type in each 10 km × 10 km grid as the ecosystem type for each grid except human settlement. The area of human settlement was small compared with other ecosystem types, but human activity space and manmade pollutants, including PM$_{2.5}$, were not just confined to the spatial range of the human settlement. Thus, the ecosystem type where the total area of the human settlement was larger than 25 km$^2$ was considered a human settlement ecosystem (Fig. 3e).

China has diverse climatic conditions. Zheng et al. (2010) proposed a new scheme for generating China’s climate regionalization (CR) based on daily meteorological observation data during 1971–2000. We regenerated the climate regionalization by vectorizing Zheng’s scheme and merging some small partitions (Fig. 3f).

2.3. GeogDetector model

The GeogDetector model is a novel spatial variation analysis method, which can be used to explore the driving force of a responding variable under the assumption that the two variables are associated if their spatial stratified heterogeneity tend to be consistent (Wang et al., 2010; Wang and Hu, 2012). Compared to traditional linear models, GeogDetector is capable of handling categorical dependent variables, finding the dominant driving force, and investigating the interaction between two variables with no assumption of linearity and immunity to the collinearity of dependent variables.

It comprises four modules: factor detector, interaction detector,
risk detector, and ecological detector (Wang and Hu, 2012). The factor detector uses a $q$ value to quantify the influences of factors ($X_s$) on $Y$; $q$ is given by following formula:

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N \sigma^2}$$

$$SSW = \sum_{h=1}^{L} N_h \sigma_h^2, \quad SST = N \sigma^2$$

where $h = 1, \ldots, L$ is the number of subregions (or subclasses) of factors $X$ and $N_h$ and $N$ donates the number of samples in subregion $h$ and the total number of samples over the whole study region, respectively; $\sigma_h^2$ and $\sigma^2$ denotes the variance of samples in subregion $h$ and the global variance of $Y$ over the entire study region. SSW and SST are the within sum of squares and the total sum of squares, respectively. The value of $q$ ranges from 0 to 1. The larger the $q$ value, the stronger the influence of variable $X$ on $Y$.

The interaction detector can examine the interaction of different factors ($X_s$) and reveals whether the interaction of factors ($X_1$ and $X_2$) weaken or enhance the influence on $Y$ or whether they are independent in influencing $Y$. The $q$ value of factors $X_1$ and $X_2$ calculated from a factor detector can be marked as $q(X_1)$ and $q(X_2)$. A new factor layer and subregions can be generated by overlaying the factor layer $X_1$ and $X_2$ spatially; it can be marked as $X_1 \cap X_2$. 

Fig. 1. Spatial pattern of PM$_{2.5}$ concentrations in 2015 for mainland China.

Fig. 2. Factors and their proxies.
Fig. 3. Spatial distribution of six underlying factors: (a) industrial output, (b) construction output, (c) road density, (d) coal consumption, (e) ecosystem type, and (f) climate regionalization.
\( r \) denotes the intersection between the factor layer \( X_1 \) and \( X_2 \). Then, the \( q \) value of \( X_1 \times X_2 \), represented as \( q \) (\( X_1 \times X_2 \)), can be calculated. The interactive relationship can be interpreted as five categories by comparing the interactive \( q \) value of the two factors and the \( q \) value of each of the two factors (Wang and Hu, 2012; Wu et al., 2017). The five categories are presented in Table 1.

The spatial difference of an influencing factor should have different influences on \( Y \) in different regions. The risk detector can determine whether there is a significant difference in influence on \( Y \) in two subregions via a \( t \)-test. Its formula is:

\[
I_{fl} = \frac{\overline{Y}_{h,1} - \overline{Y}_{h,2}}{\sqrt{(\text{Var}(\overline{Y}_{h,1})/n_h_1) + (\text{Var}(\overline{Y}_{h,2})/n_h_2)}}^{1/2}
\]

where \( \overline{Y}_h \) represents the average of \( Y \) in the subregion \( h \); \( n_h \) is the size of samples in subregion \( h \), and Var is variance.

The ecological detector is used to compare whether \( X_1 \) has a significantly greater influence or contribution than \( X_2 \). It is measured using the statistics \( F \):

\[
F = \frac{N_{X_1}(N_{X_2} - 1)\text{SSW}_{X_1}}{N_{X_2}(N_{X_1} - 1)\text{SSW}_{X_2}}
\]

where \( N_{X_1} \) and \( N_{X_2} \) represent the number of samples of the two factors \( X_1 \) and \( X_2 \), respectively; \( \text{SSW}_{X_1} \) and \( \text{SSW}_{X_2} \) are the within sum of squares in the subregion generated by factor layers \( X_1 \) and \( X_2 \), respectively. \( L_1 \) and \( L_2 \) represent the number of subregions of \( X_1 \) and \( X_2 \), respectively. The null hypothesis is defined as \( H_0 \):

\[
\text{SSW}_{X_1} = \text{SSW}_{X_2}
\]

The rejected \( H_0 \) at the significance level \( \alpha \) indicates that it is statistically significant.

In this research, the geographical detector model is used to examine the influence of six potential influencing factors, including industry, construction industry, traffic, coal combustion, ecological environment, and climate and their interaction effects on \( \text{PM}_{2.5} \) concentrations.

3. Results

3.1. The influence of potential driving factors on \( \text{PM}_{2.5} \) concentrations

The influence (\( q \) values) of the six driving factors on \( \text{PM}_{2.5} \) concentrations were calculated using the factor detector and are presented in Table 2. Among the six potential driving factors, \( \text{CR} \) had the greatest influence on the pattern of \( \text{PM}_{2.5} \) concentrations (\( q = 0.56 \)), followed by another natural factor, \( \text{ET} \) (\( q = 0.30 \)). Compared with natural factors, socioeconomic factors presented smaller influences on \( \text{PM}_{2.5} \) concentrations. Among the socioeconomic factors, \( \text{IO} \) presented as having a dominant influence (\( q = 0.14 \)) on \( \text{PM}_{2.5} \) concentrations, while \( \text{RD} \) presented as having the least influence (\( q = 0.10 \)).

### Table 2

<table>
<thead>
<tr>
<th>Factors</th>
<th>( q ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{IO} \ (10^6 \text{Yuan/Km}^2) )</td>
<td>0.14***</td>
</tr>
<tr>
<td>( \text{CO} \ (10^3 \text{Yuan/Km}^2) )</td>
<td>0.12***</td>
</tr>
<tr>
<td>( \text{RD} \ (\text{Km/K m}^2) )</td>
<td>0.10***</td>
</tr>
<tr>
<td>( \text{CC} \ (10^5 \text{Kg/Km}^2) )</td>
<td>0.13***</td>
</tr>
<tr>
<td>( \text{ET} )</td>
<td>0.30***</td>
</tr>
<tr>
<td>( \text{CR} )</td>
<td>0.56***</td>
</tr>
</tbody>
</table>

*Note: ***denotes that \( q \) value is significant at the 0.001 level (\( p < 0.001 \)).

3.2. The interaction effects of factors on \( \text{PM}_{2.5} \) concentrations (interaction detector)

A total of 15 pairs of interactions between the 6 factors were detected using the interaction detector. The interactive \( q \) value of each pair of factors was found to be more than both the two factors’ \( q \) values but was less than the sum of the two factors’ \( q \) values. Thus, the interactive relationship between each pair of factors was bivariate enhanced each other in influencing \( \text{PM}_{2.5} \) concentrations. Fig. 4 shows the specific comparisons between the interactive \( q \) value and the two factors’ \( q \) values. Among the interactions of socioeconomic factors, \( q \) (\( \text{IO} \times \text{CC} \)) was the maximum (0.21), indicating that the interaction between \( \text{IO} \) and \( \text{CC} \) was strongest. Also, the interaction was strongest between \( \text{IO} \) and \( \text{ET} \) among the interactions between socioeconomic factors and \( \text{ET} \), and the interaction between \( \text{IO} \) and \( \text{CR} \) was strongest among all the interactions between socioeconomic factors and natural factors. Additionally, the interaction between natural factors, \( \text{ET} \) and \( \text{CR} \), was not significantly enhanced compared with both the original \( q \) values of \( \text{ET} \) and \( \text{CR} \).

3.3. The leading impact areas (subregions) of factors in influencing \( \text{PM}_{2.5} \) concentrations

The average \( \text{PM}_{2.5} \) concentrations in all subregions of the six factors were calculated and the significance of influence differences was captured using the risk detector. The socioeconomic factors were divided according to their numerical values from small to large. \( \text{PM}_{2.5} \) concentrations in all subregions of the socioeconomic factors presented increasing trends along with increases in these factors generally (Fig. 5), indicating that larger the values of these factors...
factors were associated with higher PM$_{2.5}$ concentrations. The average PM$_{2.5}$ concentration was highest (58.15 $\mu$g/m$^3$) in the subregion of human settlement in ET, followed by desert (49.04 $\mu$g/m$^3$) and farmland (46.89 $\mu$g/m$^3$). Among subregions of CR, the average PM$_{2.5}$ concentration was the highest (69.67 $\mu$g/m$^3$) in the subregion with a warm temperate arid climate, followed by warm temperate semi-humid climate (54.02 $\mu$g/m$^3$), warm temperate humid climate (53.52 $\mu$g/m$^3$), and north subtropics (50.74 $\mu$g/m$^3$).

The area where the PM$_{2.5}$ concentrations were high can be identified as a high-risk area for PM$_{2.5}$ pollution. The first several subregions of each factor where the average PM$_{2.5}$ concentrations were relatively high (more than 40 $\mu$g/m$^3$ generally) were mapped to inform the areas with high risk of PM$_{2.5}$ pollution (Fig. 6). The eastern plain area, especially the North China Plain, was affected by multiple factors and was a high risk area for PM$_{2.5}$ pollution, while the northwest region, especially the Tarim Basin region which is mainly affected by deserts and a warm temperate arid climate, was also a high-risk area for PM$_{2.5}$ pollution.

The significance of varying influence among different subregions of IO showed that there were no significant differences between subregions 5 and 6, and there were significant differences between other pairs of subregions (Table A.2). Similarly, there were no significant differences between subregions 5 and 6, and there were significant differences between other pairs of subregions of CO (Table A.3). For factor RD, there were significant differences between all pairs of subregions except the pair of subregions 2 and 4 (Table A.4). The results of significant differences were all true between all pairs of subregions of both CC and ET (Table A.5 and A.6). The significance of varying influences of CR was true except the pair of warm temperate semi-humid climate region and warm temperate humid climate region (Table A.7).

### 3.4. Statistical significance of differences among influencing factors

The significance of varying influence among the six factors was examined using the ecological detector. The results (Table 3) showed that more than two-thirds were not statistically significant, while there were statistically significant differences between IO and other socioeconomic factors (CC, RD, and CO) and between CO and RD. By combining the findings from the factor detector, it could be concluded that IO has a significantly stronger effect on PM$_{2.5}$ than CO, RD, and CC; and CC has a significantly stronger effect on PM$_{2.5}$ than RD.
4. Discussion

Integrated natural and socioeconomic factors and their interactions drove the discrepant geographical pattern of PM$_{2.5}$ concentrations in China. This paper quantified the influences of the socioeconomic and natural factors, as well as their interactive impact on PM$_{2.5}$ concentrations using the GeogDetector model. The results showed that natural factors, including ecosystem and climate, were more influential than socioeconomic factors in driving PM$_{2.5}$ pollution. Among the influencing factors, climate was the dominant factor, and industry was more influential than the other socioeconomic factors. Among all the interactions of the six influencing factors, the interaction between industry and climate presented the largest influence, which was larger than the interaction of ecosystem and climate. The Tarim Basin was a severely contaminated region located mainly in the warm temperate arid climate and desert ecosystem. Another severely contaminated region located mainly in the warm temperate semi-humid climate and desert ecosystem. The Tarim Basin was a severely contaminated region located mainly in the warm temperate arid climate, warm temperate humid climate, warm temperate semi-humid climate, warm temperate humid climate, and north subtropics were also found to have high PM$_{2.5}$ concentrations (Fig. 5).

The risk detector also confirmed that two highly polluted regions, the Tarim Basin (and its surrounding areas) and the eastern plain, were affected by multiple leading impact subregions of either natural or socioeconomic factors, or both. In the Tarim Basin, the high PM$_{2.5}$ concentrations were mainly attributable to the Takla Makan Desert and warm temperate arid climate (Fig. 6). The lack of plant-covered geographical surfaces formed under the influence of dry weather strengthened the effect of the wind and further promoted dust storm outbreaks (Ma et al., 2013; Wang et al., 2009). Thus, the interaction of the desert ecosystem and climate resulted in long-term highly polluted areas of PM$_{2.5}$.

In the eastern plain, especially the North China Plain, is the main agricultural area of China and has high concentrations of heavy industries, high road density, and accelerating urbanization. This general regional environment contributed a substantial amount to primary emissions and secondary inorganic PM$_{2.5}$ pollutants (Guo et al., 2014). Adverse weather conditions in the planetary boundary layer, such as a continuous strong temperature inversion, downdrafts, and weak surface wind speeds are frequent and continual in the winter, which can further cause pollutants to accumulate in a shallow layer (Zhang et al., 2014; Zhao et al., 2013). Moreover, coal-fueled heating and less vegetation coverage (deciduous broad-leaf forests) were related to substantial increases in pollution sources in winter (Wu et al., 2015; Xiao et al., 2015; Yu et al., 2013; Zheng et al., 2005). Hence, the interactions between multiple leading impact subregions of natural and socioeconomic factors led to the formation of severe PM$_{2.5}$ pollution, especially during winter in the North China Plain.

Previous studies had similar findings and determined that the air pollution over eastern China was associated with both anthropogenic emissions and meteorological conditions (Yang et al., 2017b; Zhang and Cao, 2015). Although there were also massive cities, industries, and transportation in other regions, such as in the Greater Changsha Metropolitan Region, the Chengdu-Chongqing Urban Agglomeration, and the Pearl River Delta, the emissions of industrial waste gas, such as SO$_2$ and soot (He et al., 2014), and coal combustion were less than the North China Plain, and natural conditions (climate and ecology) were not conducive to forming severe haze, and their PM$_{2.5}$ concentrations were much lower than in the North China Plain. No other region had as many anthropogenic emission sources and such favorable natural conditions for developing severe PM$_{2.5}$ pollution as the North China Plain (Fig. 6). The ecological detector’s results indicated that industry has a significantly stronger effect on PM$_{2.5}$ than coal consumption, traffic, and the building industry, and coal consumption has a significantly stronger effect on PM$_{2.5}$ than traffic on the national scale. Socioeconomic factors differ widely across regions, and empirical studies have indicated varying contributions of these factors to PM$_{2.5}$ pollution in different cities or regions (Wang et al., 2013; Wei et al., 1999; Yu et al., 2013). Some previous studies on the national scale have also explored the effected factors of PM$_{2.5}$. Huang et al. (2014) indicated that China’s emitted PM$_{2.5}$ was coarser mainly due to strong industry emissions; however, nonindustrial sources had a larger contribution in developed countries. Timmermans et al. (2017) indicated that industry and traffic were the largest contributors during summer months, and residential combustion was

Additionally, the leading impact areas (high-risk areas of PM$_{2.5}$ pollution) were revealed using the risk detector. As the hypothesis that the PM$_{2.5}$ concentrations were generally high in subregions with high values of the socioeconomic variables, and the subclass of ecosystem, mainly human settlements, desert, and farmland and the subclass of climate, including the warm temperate arid climate, warm temperate semi-humid climate, warm temperate humid climate, and north subtropics were also found to have high PM$_{2.5}$ concentrations (Fig. 5).

Table 3

<table>
<thead>
<tr>
<th>Difference</th>
<th>CR</th>
<th>ET</th>
<th>CC</th>
<th>RD</th>
<th>CO</th>
<th>IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR</td>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ET</td>
<td></td>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>IO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Y means the difference of the influencing factors is significant with the confidence of 95%, while N means no significant difference.

<table>
<thead>
<tr>
<th>Region</th>
<th>PM$_{2.5}$ concentration (µg/m$^3$)</th>
<th>Industrialization</th>
<th>Traffic</th>
<th>Building</th>
<th>Coal consumption</th>
<th>Ecosystem</th>
<th>Climate</th>
<th>Other socioeconomic factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>North China Plain</td>
<td>96.3 (±15.2)</td>
<td>65%</td>
<td>40%</td>
<td>10%</td>
<td>56%</td>
<td>49%</td>
<td>32%</td>
<td>21%</td>
</tr>
<tr>
<td>Eastern Plain</td>
<td>121.4 (±20.7)</td>
<td>75%</td>
<td>50%</td>
<td>20%</td>
<td>65%</td>
<td>59%</td>
<td>38%</td>
<td>12%</td>
</tr>
<tr>
<td>Western Plain</td>
<td>58.7 (±9.8)</td>
<td>50%</td>
<td>25%</td>
<td>10%</td>
<td>45%</td>
<td>38%</td>
<td>22%</td>
<td>20%</td>
</tr>
</tbody>
</table>

There are fewer emission sources in a forest system and other socioeconomic factors were related to the direct sources of PM$_{2.5}$ pollution, the results documented that socioeconomic factors had less influence than natural factors. Meteorological conditions usually have a direct and real-time impact on PM$_{2.5}$ pollution (Yang et al., 2011; Zhao et al., 2013). There are fewer emission sources in a forest system and other socioeconomic factors were related to the direct sources of PM$_{2.5}$ pollution, the results documented that socioeconomic factors had less influence than natural factors. Meteorological conditions usually have a direct and real-time impact on PM$_{2.5}$ pollution (Yang et al., 2011; Zhao et al., 2013).
the largest during winter months in China.

Spatial stratified heterogeneity is a spatial representation of natural and socioeconomic phenomena, including PM2.5 pollution. Theoretically, the GeogDetector in this study is based on a spatial variance analysis of the spatial consistency between PM2.5 concentrations and suspect geographical strata (subregions or subclasses). The key assumption is that if the spatial distribution of two variables tends to be consistent, then there is a statistical association between the two variables. Thus, the statistical association can imply potential mechanisms within distinct spatial strata and suggest potential factors involved in the observed process (Wang et al., 2010; Wang and Hu, 2012). This assumption reflects a general understanding of natural phenomena.

In this paper, discrepant influences and interactions among possible factors involved in PM2.5 pollution were examined. Although anthropogenic emissions related to socioeconomic factors are the primary cause for increases in PM2.5 concentrations, this study verified that natural factors had a stronger influence in driving regional severe and persistent PM2.5 pollution, especially in the Tarim Basin and the North China Plain. The existence of spatial stratified heterogeneity within PM2.5 pollution determined that the association between PM2.5 concentrations and the possible factors was usually weak on a large scale; relationships derived from the regression model were explicable on a limited local scale, but lack explanatory power on a large scale (Lin et al., 2014; Lu et al., 2017).

Some limitations of the study should be clarified. The first limitation is that the results are statistical and do not prove causality, the specific factors for PM2.5 pollution are quite complex and diverse in different regions. However, the results can screen out highly suspicious factors for confirming the causality with further analysis. The second limitation is from the raw data of PM2.5. The raw data of PM2.5 are estimated based on aerosol optical depth (AOD). The accuracy is considerably high, but it also shows uncertainty in different regions (Van Donkelaar et al., 2016). The error in the raw PM2.5 data would affect the results inevitably. In addition, different industrial structure may contribute to different amount of polluting emissions and produce different effects on PM2.5 pollution. The potential impacts were not fully considered, which would also introduce some uncertainties. Thus, data for emission inventory with high spatial resolution can be used in further studies to improve the accuracy and efficiency of the quantification of anthropogenic influences on PM2.5 pollution.

5. Conclusions

The study involved a quantitative analysis of natural and socioeconomic factors’ influence on PM2.5 pollution in China. The discrepant influences among natural and social-economic factors in PM2.5 pollution were revealed; the spatial differences of PM2.5 pollution was related to the spatial disparity of the factors and their interactions. The findings can contribute to better understanding the influencing mechanisms of PM2.5 pollution and the geographical factors for forming persistent and highly polluted areas, and imply that more specific control strategies need to be targeted in different regions toward successful PM2.5 pollution control and abatement.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 41601419, 41571138, 41761021) and the Collaborative Project of Henan Key Laboratory of Integrative Air Pollution Prevention and Ecological Security. The authors are grateful to the Resource and Environmental Science Data Center (http://www.resdc.cn) of the Chinese Academy of Sciences and the China Meteorological Data Sharing Service System (http://cdc.cma.gov.cn/) for providing data.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.envpol.2018.05.043.

References


Butt, E.W., Turnock, S.T., Rigby, R., Reddington, C.L., Yoshio, M., 2017. Global and regional trends in particulate air quality and attributable health burden over the past 50 years, 19, 3519.


Timmermans, R., Kranenburg, R., Manders, A., Hendriks, C., Segers, A., Dammers, E.,
process of the widespread extreme haze pollution over northern China in
January 2013: implications for regional air quality and climate. Atmos. Environ.
Timmermans, R., Kraneburg, R., Manders, A., Hendriks, C., Segers, A., Dammers, E.,
Source apportionment of PM2.5 across China using LOTOS-EUROS. Atmos.
Van Donkelaar, A., Martin, R.V., Brauer, M., Hsu, N.C., Kahn, R.A., Levy, R.C.,
matter using a combined geophysical-statistical method with information from
Environment.2012.01.015.
of particulate matter with different sizes in Lanzhou, China. Atmos. Environ.
detectors-based health risk assessment and its application in the neural tube
https://doi.org/10.1080/13658810802443547.
Xin, J.X., 2014. Mechanism for the formation of the January 2013 heavy haze
Guo, S., Meng, J., Li, J., Cheng, C., Hu, T., Ren, Y., Wang, Y., Gao, J., Cao, J., An, Z.,
Zhou, W., Li, G., Wang, J., Tian, P., Marrero-Ortiz, W., Secrest, J., Du, Z., Zheng, J.,
Kolb, C.E., Molina, M.J., 2016a. Persistent sulfate formation from London Fog to
https://doi.org/10.1073/pnas.1615640113.
Wei, F., Teng, E., Wu, G., Hu, W., Wilson, W.E., Chapman, R.S., Pau, J.C., Zhang, J.,
1999. Ambient concentrations and elemental compositions of PM10 and PM2.5 in four
es9904944.
Wu, J., Li, J., Peng, J., Li, W., Xu, G., Dong, C., 2015. Applying land use regression
model to estimate spatial variation of PM2.5 in Beijing, China. Environ. Sci.
Wu, C., Ye, X., Du, Q., Luo, P., 2017. Spatial effects of accessibility to parks on housing
1016/j.habitatint.2017.03.010.
Xiao, Q., Ma, Z., Li, S., Liu, Y., 2015. The impact of winter heating on air pollution in
trend and spatial pattern of PM2.5 induced premature mortality in China, 97,
spatio-temporal concentrations of PM2.5 using land use and meteorological
doi.org/10.1007/s00477-017-1497-6.
Dust-wind interactions can intensify aerosol pollution over eastern China. Nat.
Commun. 8, 15333. https://doi.org/10.1038/ncomms15333.
Characterization and source apportionment of PM2.5 in an urban environment
aaqr.2012.07.0192.
Zhang, Y.-L., Cao, F., 2015. Fine particulate matter (PM2.5) in China at a city level. Sci.
Rep. 5, 14884. https://doi.org/10.1038/srep14884.
2009GL039102.
Zhang, R.H., Li, Q., Zhang, R.N., 2014. Meteorological conditions for the persistent
severe fog and haze event over eastern China in January 2013. Sci. Earth China
acta.2013.10.015.
Seasonal trends in PM2.5 source contributions in Beijing, China. Atmos. Environ.
ACTA Geogr. Sin 65, 3–12.