PLOS ONE

RESEARCH ARTICLE

Spatial–seasonal characteristics and critical impact factors of $PM_{2.5}$ concentration in the Beijing–Tianjin–Hebei urban agglomeration

Tianhang Huang[1], Yunjiang Yu[2]*, Yigang Wei[3,4]*, Huiwen Wang[3,4], Wenyang Huang[3], Xuchang Chen[5]

1 School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing, China, **2** International Business School, Shanghai Lixin University of Accounting and Finance, Shanghai, China, **3** School of Economics and Management, Beihang University, Beijing, China, **4** Beijing Key Laboratory of Emergency Support Simulation Technologies for City Operation, Beijing, China, **5** School of Economics and Management, University of Chinese Academy of Sciences, Beijing, China

* 20159016@lixin.edu.cn (YY); weiyg@buaa.edu.cn (YW)

Citation: Huang T, Yu Y, Wei Y, Wang H, Huang W, Chen X (2018) Spatial–seasonal characteristics and critical impact factors of $PM_{2.5}$ concentration in the Beijing–Tianjin–Hebei urban agglomeration. PLoS ONE 13(9): e0201364. https://doi.org/10.1371/journal.pone.0201364

Editor: Krishna Prasad Vadrevu, University of Maryland at College Park, UNITED STATES

Received: November 3, 2017

Accepted: July 13, 2018

Published: September 20, 2018

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by National Social Science Foundation of P.R. China (13CGL103), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (No.18YJC840041) and Natural Science Foundation of China (Grant No.71420107025; No.51478025) and Energy Economy Major

Abstract

As China's political and economic centre, the Beijing–Tianjin–Hebei (BTH) urban agglomeration experiences serious environmental challenges on particulate matter (PM) concentration, which results in fundamental or irreparable damages in various socioeconomic aspects. This study investigates the seasonal and spatial distribution characteristics of $PM_{2.5}$ concentration in the BTH urban agglomeration and their critical impact factors. Spatial interpolation are used to analyse the real-time monitoring of $PM_{2.5}$ data in BTH from December 2013 to May 2017, and partial least squares regression is applied to investigate the latest data of potential polluting variables in 2015. Several important findings are obtained: (1) Notable differences exist amongst $PM_{2.5}$ concentrations in different seasons; January (133.10 mg/m^3) and December (120.19 mg/m^3) are the most polluted months, whereas July (38.76 mg/m^3) and August (41.31 mg/m^3) are the least polluted months. $PM_{2.5}$ concentration shows a periodic U-shaped variation pattern with high pollution levels in autumn and winter and low levels in spring and summer. (2) In terms of spatial distribution characteristics, the most highly polluted areas are located south and east of the BTH urban agglomeration, and $PM_{2.5}$ concentration is significantly low in the north. (3) Empirical results demonstrate that the deterioration of $PM_{2.5}$ concentration in 2015 is closely related to a set of critical impact factors, including population density, urbanisation rate, road freight volume, secondary industry gross domestic product, overall energy consumption and industrial pollutants, such as steel production and volume of sulphur dioxide emission, which are ranked in terms of their contributing powers. The findings provide a basis for the causes and conditions of $PM_{2.5}$ pollution in the BTH regions. Viable policy recommendations are provided for effective air pollution treatment.

Development Project of "Double First-Class" Initiative of China (ZG211S1816).

Competing interests: "The authors have declared that no competing interests exist."

Introduction

The rapid urbanisation, industrialisation and modernisation that China has undergone have improved household income and living standards but have also led to extensive and severe issues of air pollution. In particular, air pollution has become a major environmental challenge for most Chinese cities [1–5]. Various particulate matters (PMs) are recognised as key air pollutants that have serious adverse impacts on human health. Scientific reports indicate that excessive exposure to high PM concentrations reduces the expected human lifespan by 1–5.5 years [6]. In 2013, the WHO identified PMs as the leading cause of human cancer.

Atmospheric PM can be divided into several categories on the basis of their aerodynamic diameters, namely, total suspended PM, PM with a particle size below 10 μm (PM_{10}) and fine PM ($PM_{2.5}$) [7–8]. Fine $PM_{2.5}$ with a diameter less than 2.5 mm is widely recognised as one of the most detrimental airborne PMs because small particle pollutants can enter the lungs and alveolar macrophage through breathing [9]. Long-term exposure to $PM_{2.5}$ leads to high incidences of cardiovascular and respiratory diseases and chronic bronchitis cases. According to the Global Burden of Disease 2010 Comparative Risk Assessment, high $PM_{2.5}$ concentrations have caused 3.1 million deaths worldwide [10]. Despite the apparent reductions on 'traditional pollutants' (e.g. NO_2 and SO_2) in recent years, $PM_{2.5}$ has become a major air pollutant that threatens human morbidity and mortality in developing countries [11–15]. Severe and frequent smog and haze weather (high $PM_{2.5}$ concentration) have also become major barriers for China in attracting foreign investments and talents [16]. Although $PM_{2.5}$ has been incorporated in the official monitoring and appraisal system in Europe and North America for more than 30 years [17–19], China only included $PM_{2.5}$ as a routine monitoring indicator in 2013. Numerous studies have focused on $PM_{2.5}$ pollution in developed and emerging economies [20–26]. Some studies showed that 20%–30% of $PM_{2.5}$ in Chinese cities originated from coal combustion [20, 22, 27, 28]. For example, Aldabe et al. (2011) [29] investigated the chemical compositions and source apportionments of $PM_{2.5}$ in North Spain, Cesari et al. (2018) [26] studied seasonal variability in Southern Italy and Lonati et al. (2005) [30] examined $PM_{2.5}$ composition in Milan, Italy. Ram et al. (2008) [31] confirmed that the contribution of secondary organic carbon to $PM_{2.5}$ is relatively high in developing countries, such as India. Donkelaar et al. (2010) [32] developed the first global $PM_{2.5}$ distribution map based on satellite data. The data indicated that most polluted regions are mainly located in North Africa and East Asia. In conclusion, relevant studies show that chemical compositions and source apportionments of $PM_{2.5}$ significantly vary in different regions and countries.

At the end of the last century, China officially incorporated PM_{10} and $PM_{2.5}$ in its official environmental monitoring and appraisal system [33]. Several studies showed that Beijing's annual average PM_{10} mass concentration maintained a consecutive upward trend, increasing from 162 $\mu g/m^3$ to 166 $\mu g/m^3$ from 2000 to 2002 (Beijing Environmental Protection, 2004, 2005) [34, 35]. Beijing's annual average $PM_{2.5}$ concentration, which was measured at five sites, was relatively high in 2000, varying from 87.6 $\mu g/m^3$ to 111.9 $\mu g/m^3$ [36]. From January 2004 to December 2012, high levels of PM_{10} and $PM_{2.5}$ were obtained in Beijing, with annual mean values of 138.5 ± 92.9 $\mu g/m^3$ and 72.3 ± 54.4 $\mu g/m^3$, respectively [37]. The Beijing–Tianjin–Hebei (BTH) region is the largest urban agglomeration in North China and is the core economic zone in China. According to China Environmental Status Bulletin 2015 [38], 70 cities at the prefecture level and above in BTH and surrounding regions experienced 1710 episodes of severe pollution in 2014, which accounted for 41% of the national total [39].

Mainstream literature on China's air quality has focused on the chemical compositions of air pollutants, namely, elemental constituents, organic compounds and major inorganic ions [27, 28]. Several studies based on remote sensing and modelling techniques revealed that $PM_{2.5}$ concentrations in China's urban areas are significantly higher than in rural areas [40,

41]. Zhang et al. (2013) [22] indicated that motor vehicle ownership is the key contributing factor to Beijing's air pollution, accounting for 63% of the carbonaceous components on $PM_{2.5}$, whereas coal combustion is the main air pollutant source in Tangshan, accounting for 30% of $PM_{2.5}$ compositions. Ma and Zhang (2014) [16] investigated the $PM_{2.5}$ distribution characteristics in China from 2001 and 2010 based on satellite data developed by Battelle Memorial Institute. Their study confirmed that the spatial aggregation effects of $PM_{2.5}$ are apparent in China. Specifically, highly polluted areas are mainly concentrated in the BTJ region, Yangtze River Delta (YRD) and the linking zones.

Several studies have focused on the $PM_{2.5}$ characteristics in the BTH region. Using data from 1497 station-based monitoring sites, Shen and Yao (2017) [42] investigated the effects of demographic and economic factors on $PM_{2.5}$ concentration in four urban agglomerations, namely, BTH, YRD, Pearl River Delta (PRD) and Chengdu–Chongqing (CC). The estimated results indicated that a high correlation exists amongst population density, economic affluence and $PM_{2.5}$ concentration. Geography is another important determining factor for $PM_{2.5}$ concentration because high altitudes are usually associated with high PM2.5 concentrations. Zhou et al. (2017) [43] investigated the impact of economic and ecological factors on $PM_{2.5}$ concentrations by using a two-stage distribution lag model. Their estimated results indicated that the emission of atmospheric pollutants causes hysteresis effects on $PM_{2.5}$ concentrations. Specifically, coal consumption, industrial exhaust, value-added from heavy pollution industry and the ownership of 'yellow label car', which are heavy-polluting vehicles, are the key sources of $PM_{2.5}$ pollutant emissions. On the basis of panel data in the last 10 years, Li and Yin (2017) [44] utilised a panel threshold model to investigate the nonlinear changing patterns between socioeconomic development and $PM_{2.5}$ concentration. The study confirmed that the development of the manufacturing and construction sectors and the growth of automobile volumes aggravate $PM_{2.5}$ pollution when the value-added of the tertiary industry is below the threshold value of 6,080 billion. Moreover, the development of the second and third industries was noted to be an effective roadmap to alleviate $PM_{2.5}$ pollution when the value-added of the tertiary industry exceeds 6,080 billion. Su and Zhong (2015) [45] analysed the natural and man-made contributing factors of $PM_{2.5}$ in nine key economic circles in China, such as BTH, YRD, PRD and CC, by using a factor analysis method. They concluded that the effects of human activities are more significant than those of natural factors, and industrial activities are the important contributing factors of $PM_{2.5}$.

These findings also stimulated another key question regarding the seasonal and spatial characteristics and critical impact factor of $PM_{2.5}$ concentrations. However, considering the insufficient long-term and large-scale $PM_{2.5}$ concentration data [46], especially the lack of real-time monitoring data [47], few studies have quantitatively estimated these factors in BTH regions [5, 48]. Although the BTH urban agglomeration is recognised as one of the most severely polluted regions in China [24, 49], analyses on the seasonal and spatial characteristics of PM2.5 spanning a long period are limited [50–52].

The present study aims to estimate the $PM_{2.5}$ concentration characteristics in the BTH urban agglomeration. Specifically, the key objectives of this study are as follows: 1) to investigate the relationships between seasons and $PM_{2.5}$ concentrations (i.e. seasonal variation characteristics), 2) to evaluate the spatial distribution of $PM_{2.5}$ concentrations in different cities and 3) to investigate the critical impact factors of $PM_{2.5}$ concentrations. Real-time monitoring data spanning December 2013 to May 2017 are collected from 80 atmospheric physics observation points by the China Meteorological Administration. Several studies have investigated Beijing's $PM_{2.5}$ concentration [8, 27, 53]. However, the data sources of these studies were mainly based on satellite images [47], and few studies collected $PM_{2.5}$ real-time monitoring data to significantly improve the estimation accuracy. The current study derives a reliable and accurate estimation on the spatial and seasonal characteristics of $PM_{2.5}$ concentration.

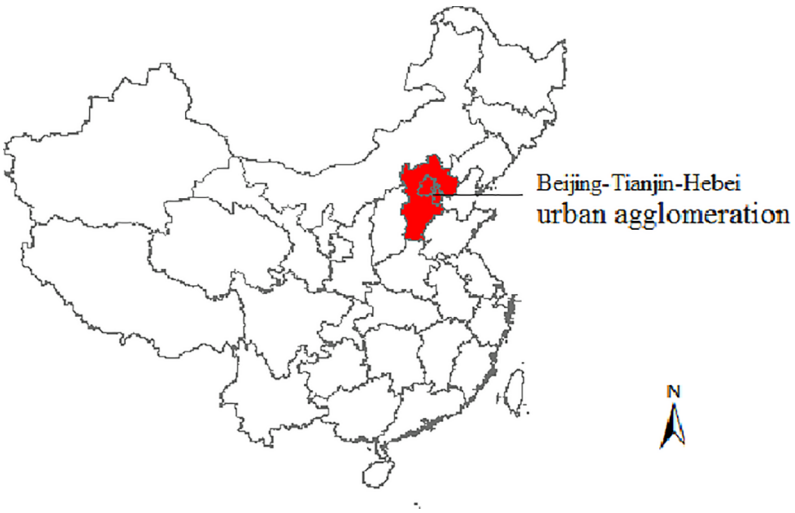

Fig 1. Geographic information of the BTH urban agglomeration.

https://doi.org/10.1371/journal.pone.0201364.g001

This study provides novel contributions to existing literature as follows: 1) A unique $PM_{2.5}$ data source with a long-time span from December 2013 to April 2017 is used. Real-time monitoring data are consistent with the reliable estimation of satellite data. 2) An array of novel and promising techniques, namely, remote sensing techniques, spatial interpolation and partial least squares (PLS) regression are employed, all of which provide an insightful analysis on $PM_{2.5}$ characteristics in the BTH region. Specifically, spatial interpolation is a powerful approach to replace the traditional inversion method in reflecting $PM_{2.5}$ concentrations at near-ground level [54]. PLS regression is an effective approach to cope with the multicollinearity issue when a large number of independent variables are included in the estimation. 3) Various variables are considered in identifying the critical impact factors of $PM_{2.5}$. Considering that peripheral or distant sources commonly affect the air pollution of a city, this study provides a complete analysis on $PM_{2.5}$ characteristics of a city and a region.

Data and research methods

Study area

The study area used is the BTH urban agglomeration, which is also called the Greater Beijing region (Fig 1). This area is located in Northeast China, at longitude 113˚04' to 119˚53' east and 36˚01' to 42˚37 ' north. It measures 218,000 km^2 and had more than 100 million residents as of 2016 (National Bureau of Statistics of China, 2016). The BTH urban agglomeration is the largest urban agglomeration and the most developed economic centre in northern China. Beijing is the political capital, cultural and information centre of China and is one of the largest megacities worldwide, with more than 21 million people and 5.7 million vehicles in 2016 [55]. Given the importance of the Greater Beijing region, severe air pollution has been the leading environmental challenge, with frequent occurrences of fog and haze. Related statistical consensus indicate that the total annual mean values of Beijing's PM_{10} and $PM_{2.5}$ concentrations from 2012 to 2015were 138.5 ± 92.9 μg/m^3 and 2.3 ± 54.4 μg/m^3, respectively [37].

Data sources

In terms of urban distribution and prefectural boundary, prefectural boundary layers at a scale of 1:4,000,000 are obtained from the National Geomatics Centre of China.

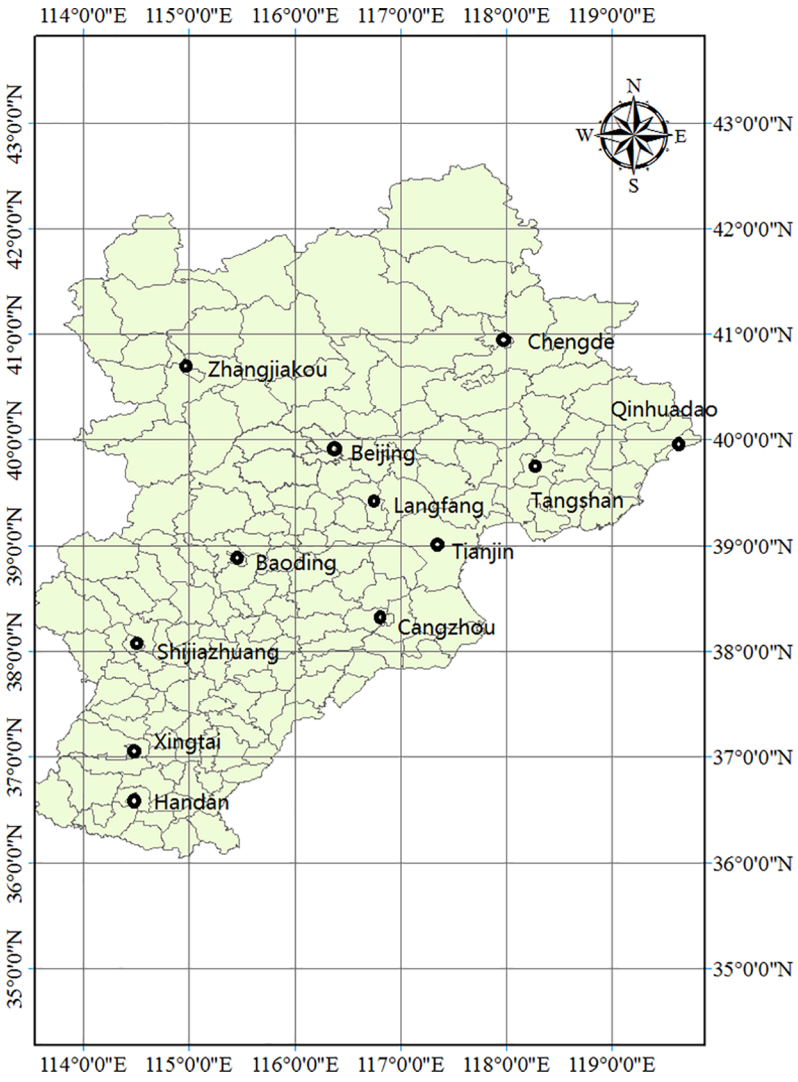

Fig 2. Geographic allocation of the 12 cities in the BTH urban agglomeration.

https://doi.org/10.1371/journal.pone.0201364.g002

$PM_{2.5}$ concentration data in the BTH urban agglomeration from December 2013 to May 2017 cover 12 cities, namely, Beijing, Tianjin, Tangshan, Zhangjiakou, Baoding, Handan, Chengde, Qinhuangdao, Xingtai, Cangzhou, Langfang and Shijiazhuang (For details, see Fig 2 and S2 Table).

This monitoring dataset is obtained from the atmospheric physics sites of the 12 cities by the Ministry of Environmental Protection of China. The remote sensing data from the Atmospheric Composition Analysis Group (2016) [56] are initially used in BTH region. Compared with the remote sensing data on $PM_{2.5}$ (Atmospheric Composition Analysis Group, 2016) [56], spatial interpolation has higher accuracy than remote sensing data in reflecting $PM_{2.5}$ concentrations at near-ground level in this study.

These data are measured by 80 monitoring stations distributed throughout the BTH region (Fig 3). Each monitoring station automatically measures monthly $PM_{2.5}$ concentrations. Cangzhou, which is a small city, has three air quality monitoring stations. Other cities have more than five air quality monitoring stations that are distributed from the suburbs to downtown. The annual average of $PM_{2.5}$ concentration for each monitoring station is calculated

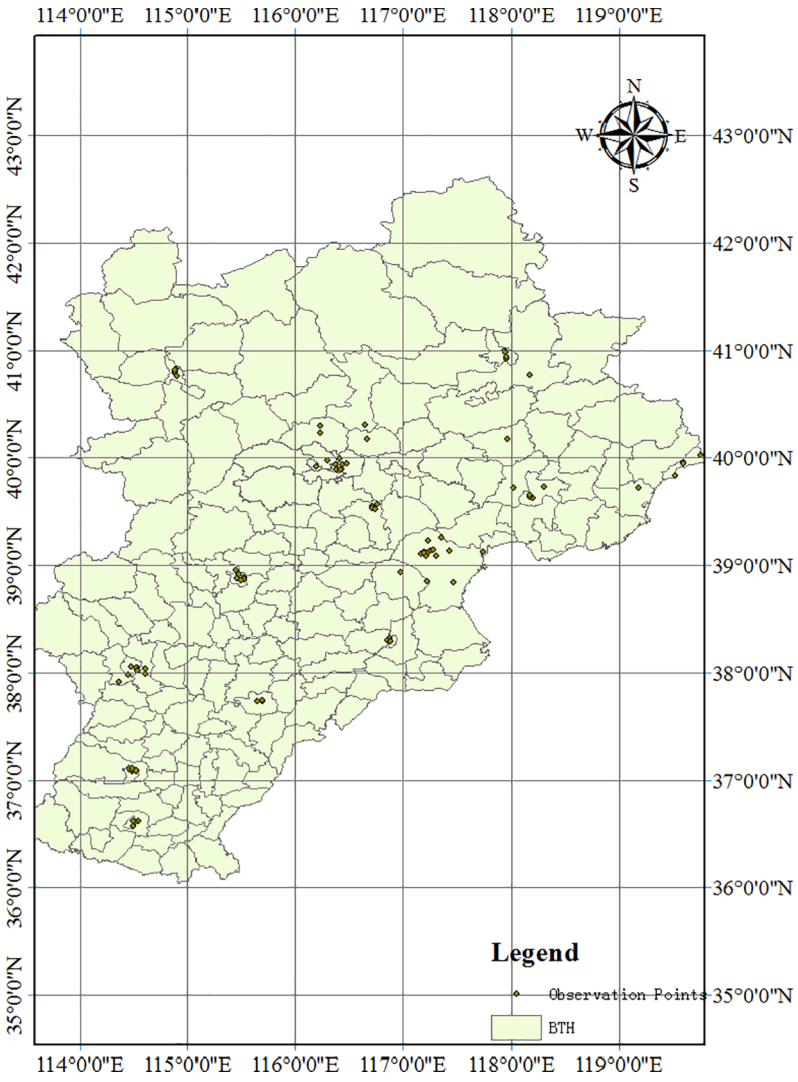

Fig 3. Geographic allocation of 80 atmospheric physics observation points.

https://doi.org/10.1371/journal.pone.0201364.g003

based on the hourly real-time data. Beijing, Tianjin and Hebei have 12, 15 and 53 atmospheric physics observation points, respectively. Table 1 describes the geographic information on several atmospheric physics observation points. The geographic information of all observation points is shown in S1 Table. In terms of data standardisation, the collected $PM_{2.5}$ values at 2:00, 8:00, 14:00 and 20:00 from different observation points are averaged to derive the daily

Table 1. Geographic information on some atmospheric physics observation points.

City	Observation Point	Longitude	Latitude
Beijing	Wanshougong West	116.3747278	39.88565298
Tianjin	Municipal inspection centre	117.1655924	39.10485468
Baoding	Huadian	115.520781	38.8918471

Source: China Meteorological Administration

https://doi.org/10.1371/journal.pone.0201364.t001

Table 2. Potential critical impact factors for $PM_{2.5}$ concentration.

Abbreviation	Variable	Unit	Reference
PD	Population density	Persons/KM^2	[1, 2, 4, 5]
UR	Urbanisation rate	%	[1, 2, 4, 5]
RFV	Road freight volume	10000 tons	[26, 29, 61–63]
SIGDP	Secondary industry GDP	100 millions	[30]
OEC	Overall energy consumption	10000 tons of standard coal	[20, 22, 27, 28, 31, 41]
SP	Steel production	10000 tons	[20, 22, 27, 28, 30, 31, 33]
VOSDE	Volume of sulphur dioxide emission	Ton	[20, 22, 27, 28, 30, 31, 33]
VOISE	Volume of industrial soot (dust) emission	Ton	[20, 22, 27, 28, 30, 31, 33]
CP	Cement production	10000 tons	[20, 22, 27, 28, 30, 31, 33]
MVO	Motor vehicle ownership	10000 units	[29, 33, 61]
RPTV	Road passenger traffic volume	10000 persons	[29, 61]
RNGC	Residential natural gas consumption	10000 cubic metres	[60]

https://doi.org/10.1371/journal.pone.0201364.t002

and monthly $PM_{2.5}$ concentrations of a city. The $PM_{2.5}$ concentration of 12 sites is the average of 80 sites.

Table 2 describes an array of $PM_{2.5}$-related variables summarised from an extensive literature review. These variables are tested to identify the critical impact factors for $PM_{2.5}$ concentration. Existing studies on identifying the key contributing factors of $PM_{2.5}$ concentration in China have mainly focused on demographic and economic aspects and other chemical air pollutants. Yang and Chen (2017) [57] used independent variables, namely, coal consumption, cement production, automobile volume, population and gross domestic product (GDP). Lu et al. (2017) [58] incorporated the following variables in their estimation, namely, population density, annual volume of bus passengers, road freight, proportion of secondary industry to overall GDP, volume of SO_2 emissions and volume of industrial soot emission. Ma and Xiao (2017) [59] considered urbanisation, energy consumption structure[60] and construction areas in their investigation. On the basis of an extensive literature review, 12 potential contributing factors for $PM_{2.5}$ concentrations are identified (Table 2).

The possible critical impact factors of $PM_{2.5}$ concentration are selected (Tables 2 and 3), discussed and included in the estimation model. Since independent variable data in 2016 and

Table 3. Statistic description (2015).

Variable	Std. Deviation	Mean	Minimum	Maximum
PD	264.25	578.89	96.73	870.29
UR	25.77	44.23	7.10	86.51
RFV	12913.84	20361.00	4152.00	38704.00
SIGDP	2134.57	2182.93	445.09	7723.60
OEC	123147.06	39054.27	472.52	430000.00
SP	3587.67	2800.52	31.85	11179.00
VOSDE	57232.31	81347.50	22070.00	214723.00
VOISE	47176.04	63778.92	12987.00	191713.00
CP	784.35	889.35	141.06	2781.00
MVO	165.84	209.15	60.56	561.90
RPTV	13441.63	9055.83	1163.00	49931.00
RNGC	53139.42	22124.42	666.00	189188.00

Data Resource: National Bureau of Statistics of the People's Republic of China (2016 a, b)

https://doi.org/10.1371/journal.pone.0201364.t003

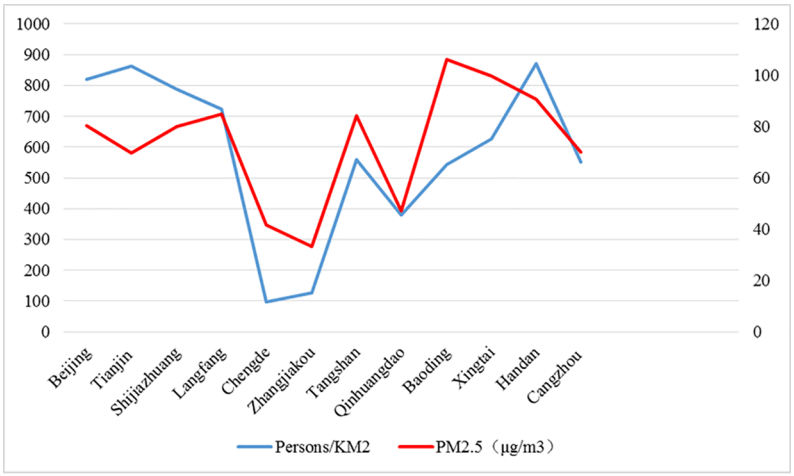

Fig 4. $PM_{2.5}$ concentration and population density per km^2 in 2015. (Unit: $\mu g/m^3$ refers to the left axis, and Persons/km^2 refers to the right axis).

https://doi.org/10.1371/journal.pone.0201364.g004

2017 has yet been published by the statistic consensus, this study used the latest data of 2015 in the PLS model to analyse the critical factors for PM2.5 concentration (For details, see S3 Table).

Population density. A comparison of $PM_{2.5}$ concentrations with population reveals interesting findings that should be considered. Shijiazhuang, which is the most polluted city, features a relatively high population density of 788.14 persons/km^2. Handan, the second most polluted city, has the highest population density of 870.29 persons/km^2. Zhangjiakou, which has the best air quality, presents a low population density of 127.46 persons/km^2. Chengde, with a similar pollution level as Shijiazhuang, also manifests a low population density of 96.73 persons/km^2 (Fig 4). A common pattern exists in which population density forms a certain positive relationship with $PM_{2.5}$ concentration. However, this pattern is affected by various critical impact factors that lead to certain variations. Therefore, a detailed investigation on critical $PM_{2.5}$ factors should be conducted for a thorough analysis of such particles.

Industrial and energy aspects. The estimation results reveal that $PM_{2.5}$ concentration in the BTH urban agglomeration exhibits a distinctive spatial distribution characteristic. Related literature shows that coal combustion accounts for 20%–30% of $PM_{2.5}$ pollution in Chinese cities [20, 22, 27, 28]. In winter, $PM_{2.5}$ concentration is usually high because coal is used as the main energy for winter heating. In summer, the situation significantly differs in the BTH urban agglomeration. For example, motor vehicles account for 63% of the carbonaceous components of $PM_{2.5}$ in Beijing, while coal combustion accounts for 30.3% of $PM_{2.5}$ compositions because it is used as the major energy source for industrial production in the city [22]. Therefore, the present study uses industrial dust and industrial SO_2 emissions as parameters to investigate the air pollution contributions of heating and industrial development. Tangshan shows the highest volumes of industrial SO_2 emission, which amounted to 214,723 t in 2016, followed by Shijiazhuang and Handan with 113,652 and 110,193 t, respectively. Xingtai and Handan represent the top two contributors of industrial dust emissions, accounting for 191,713 and 100,738 t, respectively. Fig 5 shows the volume of industrial soot (dust) and sulphur dioxide emissions of 12 cities in BTH.

Transportation. Several studies argue that transportation exerts a significant adverse influence on air pollution [29]. On the basis of available data from statistical consensus, 'passenger and freight volume of highway traffic' are used as a parameter for measuring $PM_{2.5}$

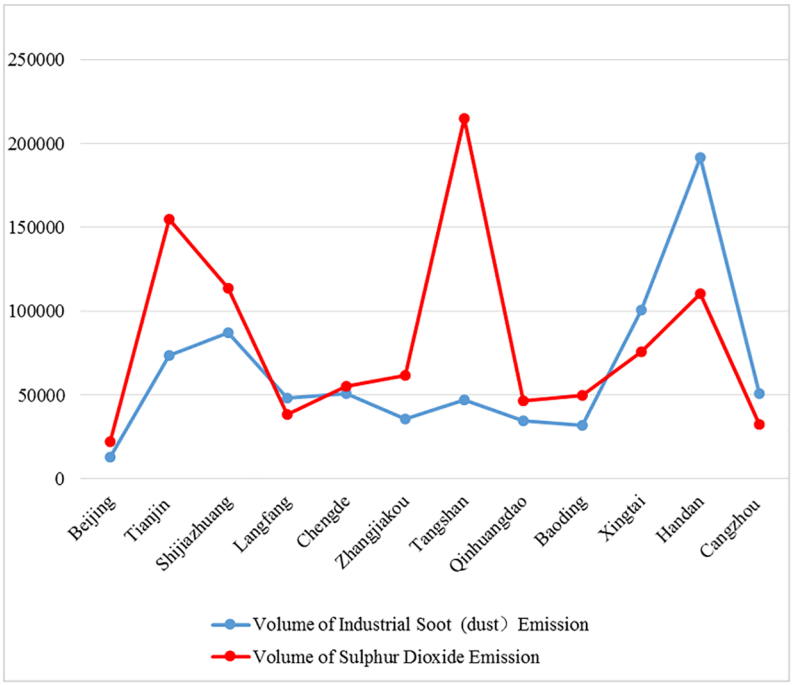

Fig 5. Volume of industrial soot (dust) and sulphur dioxide emissions in 2015 (Unit: Ton).

https://doi.org/10.1371/journal.pone.0201364.g005

pollution from transportation. Data suggest that polluted cities are generally associated with high road freight volume. For example, Shijiazhuang, Tangshan and Handan, which are the most polluted cities, are associated with relatively high freight volumes with 3,695,410,000, 363,580,000 and 387,040,000 t, respectively.

Research methods

The research framework and main research steps are illustrated in Fig 6.

Spatial interpolation. $PM_{2.5}$ concentration is a scalar description of atmospheric state significantly affected by local human activities. Although remote sensing has been improved by techniques such as regional correlations in recent years, several studies indicate that spatial interpolation is a powerful approach to replace the inversion method, leading to higher accuracy than remote sensing data in reflecting $PM_{2.5}$ concentrations at near-ground level [40, 54, 64–67]. To address this limitation, spatial interpolation is employed and the results of the

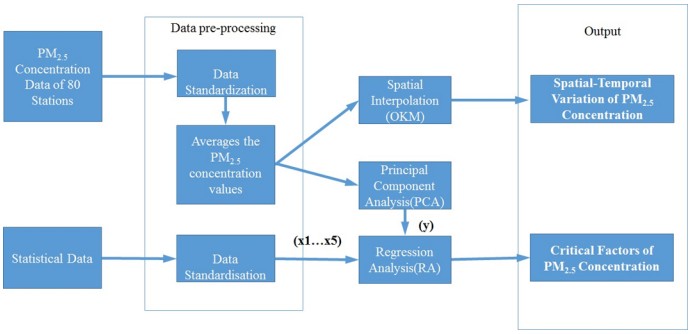

Fig 6. Research framework.

https://doi.org/10.1371/journal.pone.0201364.g006

inversion method are considered as references. Interpolation methods used in regional-scale factors include inverse distance interpolation (IDW) and Kriging interpolation method (OKM). OKM is a more widely recognised method for dealing with interpolation points than IDW [22].

This study uses OKM to simulate seasonal variations of $PM_{2.5}$ in the 12 cities of the BTH urban agglomeration. The supporting concept of OKM is that the interpolation results at the target point are the weighted sum of known attribute values of the samples [68]. In the study area, x represents the spatial location of point x. $z(x_i)$ (i = 1, 2, ⋯, n) represents the property value of sampling point x_i (i = 1, 2, ⋯, n), and annual mean $PM_{2.5}$ concentration is the property value of point x_i. Then, the interpolation result at target point x_0 is $z(x_0)$:

$$z(x_0) = \sum_{i=1}^{n} \lambda_i z(x_i).$$

Where λ_i (i = 1, 2, ⋯, n) depends on undetermined coefficients. Assuming that the entire study area satisfies the second-order stationary assumption, that is, 'the mathematical expectation of z(x) exists and is equal to the constant, that is, E[z(x)] = m', the covariance function of variables z(x) exists and only depends on lag value (h), that is, $Cov[z(x),z(x+h)] = E[z(x)z(x+h)] - m^2 = C(h)$.

On the basis of unbiased expectation $E[z^*(x_0)] = E[z(x_0)]$, $E[z^*(x_i)]$ refers to the spatial variation of $PM_{2.5}$ concentration in BTH by OKM in point x_i, $E[z^*(x_0)]$ denotes the spatial variation of $PM_{2.5}$ concentration in BTH by OKM in point x_0, and $z(x_0)$ is the $PM_{2.5}$ concentration in point x_0. We can conclude that $\sum_{i=1}^{n} \lambda_i = 1$. For regionalised variables that satisfy the second-order stationary conditions, the estimated variance can be calculated using the following formula:

$$\sigma_E^2 = E[z^*(x_0) - z(x_0)]^2 - \{E[z^*(x_0) - z(x_0)]\}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C_{i,j} - 2\sum_{i=1}^{n} \lambda_i C_{i,0} + C_{0,0}.$$

To obtain the minimum variance estimation under unbiased conditions, that is, $\mathrm{Min}\{Var[z^*(x_0) - z(x_0)] - 2\mu\sum_{i=1}^{n}(\lambda_i - 1)\}$.

The weight coefficients should satisfy the following equations:

$$\begin{cases} \sum_{i=1}^{n} \lambda_i Cov(x_i, x_j) + \mu = Cov(x_0, x_i) \\ \sum_{i=1}^{n} \lambda_i = 1 \end{cases}.$$

Then, we can calculate the value of λ_i (i = 1, 2, ⋯, n) and obtain the attribute value $z^*(x_0)$ at sample point x_0.

1. The degree of correlation between t_1 and u_1 should be the maximum.

The two conditions can be summarised as follows:

$$\mathrm{Var}(t_1) \rightarrow max$$

$$\mathrm{Var}(u_1) \rightarrow max$$

$$\mathrm{Var}(t_1, u_1) \rightarrow max$$

After the first principal components t_1 and u_1 are extracted from X and Y, PLS performs linear regressions of X and Y on t_1. In the PLS estimation, components t_1 and u_1 have typical component characteristics. A significant linear relationship between t_1 and u_1 indicates that X has a notable correlation with Y, and PLS is appropriate for estimating the contribution of X to

Y. The algorithm is terminated when the regression equations reach satisfactory levels. Otherwise, the residuals of X and Y after regression on t_1 are used to extract the next principal component. The algorithm iterates until the results reach satisfactory levels.

Cross-validation (Q_h^2) is used as the measurement criterion to determine whether the regression results reach the satisfactory level. For the number of extracted principal components h, rounding observation i (i = 1,2, ⋯, n) for each time, (i = 1,2, ⋯, n), the PLS model is built with the remaining (n−1) observations. Then, observation i is substituted in the fitted PLS regression equation to obtain the predicted value of y_j (j = 1,2, ⋯, q) at observation i, and the predicted value is recorded as $\widehat{y_{(i)j}}(h)$. The above calculation is repeated for each i (i = 1,2, ⋯, n). The sum of the squared errors (SSE) for dependent y_j is obtained when h principal components are extracted and $PRESS_j(h) = \sum_{i=1}^{n}(y_{ij} - \widehat{y_{(i)j}}(h))^2$ is recorded. Then, the sum of SSE for Y = [y_1, y_2, ⋯, y_q] is obtained and $\mathrm{PRESS}(h) = \sum_{i=1}^{q} PRESS_j(h)$ is recorded. All observations are likewise used to fit the regression equation with h principal components. At this time, the prediction value for observation i is noted as $\widehat{y_{(i)j}}(h)$. The sum of SSE for y_j is defined as $SS_j(h) = \sum_{i=1}^{n}(y_{ij} - \widehat{y_{(i)j}}(h))^2$, and the sum of SSE for Y is defined as $\mathrm{SS}(h) = \sum_{j=1}^{q} SS_j(h)$. Cross-validation is defined as $Q_h^2 = 1 - PRESS(h)/SS(h-1)$ Thus, a cross-validation test is performed before the end of each modelling step. The model estimation reaches a satisfactory level of precision and the extraction of components is stopped if $Q_h^2 < 1 - 0.95^2 = 0.0975$ is satisfied at step h. If $Q_h^2 \geq 0.0975$ is satisfied at step h, then the marginal contribution of the extracted principal component t_h is significant, and step (h+1) should be calculated.

After m principal components t_1, t_2, ⋯, t_m are finally extracted from X, PLS first performs a regression of y_k on t_1, t_2, ⋯, t_m and converts it in the regression equation of y_k on x_1, x_2, ⋯, x_p.

The specific procedures of the PLS algorithm are summarised as follows:

Step 1. To simplify the calculation and eliminate the effects of different units of variables, this study first standardises the original data matrices (X and Y), which are denoted by E_0 and F_0.

Step 2. Let t_1 be the first principal component extracted from E_0. The regression of E_0 and F_0 on t_1 is performed as follows:

$$E_0 = t_1 p_1' + E_1, F_0 = t_1 r_1' + F_1.$$

Where p_1 and r_1 refer to regression coefficient vectors, and E_1 and F_1 represent the corresponding residual matrices. The accuracy of the regression equation is calculated. The algorithm is terminated when the regression equations reach satisfactory levels. Otherwise, let $E_0 = E_1$ and $F_0 = F_1$, and iterate the component extraction and regression analysis. Cross validation (Q_h^2) is used to evaluate the model until the expected accuracy is obtained.

Step3. The number of regression components should be selected. The number of regression components included in the PLS model is important because it directly affects the fitting accuracy of the model. It should be carefully selected based on cross validation (Q_h^2). If Q_h^2 is higher or equal to 0.0975, then the marginal contribution of component t_h is significant and contributes to the precision of estimation results.

Step4. The regression equation of E_0 and F_0 on t_1, t_2, ⋯, t_m is derived if the model extracts m principal components. The following regression equation is developed through inverse transformation.

In the calculation of PLS, the principle component t_h should both represent the variation information of X (x_j (j = 1, 2, . . ., p)) and explain the information of Y (y_k (k = 1, 2, ⋯, q)) as much as possible. To measure the explanatory power of t_h for interpreting X and Y, we define various explanatory powers of t_h as follows:

1. The explanatory power of t_h to interpret X: $\mathrm{Rd}(\mathrm{X}; t_h) = \frac{1}{p}\sum_{j=1}^{p}\mathrm{Rd}(x_j; t_h)$;
2. The cumulative explanatory power of $t_1, t_2, \cdots, t_m$ to interpret X: $\mathrm{Rd}(\mathrm{X}; t_1, t_m) = \sum_{h=1}^{m}\mathrm{Rd}(\mathrm{X}; t_h)$;
3. The explanatory power of t_h to interpret Y: $\mathrm{Rd}(\mathrm{Y}; t_h) = \frac{1}{q}\sum_{k=1}^{q}\mathrm{Rd}(y_k; t_h)$;
4. The cumulative explanatory power of $t_1, t_2, \cdots, t_m$ to interpret Y: $\mathrm{Rd}(\mathrm{Y}; t_1, t_m) = \sum_{h=1}^{m}\mathrm{Rd}(\mathrm{Y}; t_h)$.

A significant advantage of PLS regression is the reliable choice of variables. When independent variable x_j is used to explain the set of dependent variables Y, the variable importance in projection VIP_j can be used to measure the importance of x_j in interpreting Y [69].

The expression of VIP_j is $VIP_j = \sqrt{\frac{p}{\mathrm{Rd}(\mathrm{Y};t_1,t_m)}\sum_{h=1}^{m}\mathrm{Rd}(\mathrm{Y}; t_h)w^2_{hj}}$, where p represents the number of independent variables, and w_{hj} is the linear combination coefficient of the principal component t_h. For principle component t_h, $t_h = w_{h1}x_1 + w_{h2}x_2 + \cdots + w_{hp}x_p$. For h = 1,2, ⋯,m, $\sum_{j=1}^{p} w^2_{hj} = 1$. The explanatory power of x_j to Y is transferred by t_h. Formula $VIP_j^2 = \frac{p\sum_{h=1}^{m}\mathrm{Rd}(\mathrm{Y};t_h)w^2_{hj}}{\sum_{h=1}^{m}\mathrm{Rd}(\mathrm{Y};t_h)}$ indicates that when the values of $\mathrm{Rd}(\mathrm{Y};t_h)$ and w^2_{hj} are large, VIP^2_j will also gain a large value.

Formula $\sum_{j=1}^{p} VIP_j^2 = \sum_{j=1}^{p}\frac{p\sum_{h=1}^{m}\mathrm{Rd}(\mathrm{Y};t_h)w^2_{hj}}{\sum_{h=1}^{m}\mathrm{Rd}(\mathrm{Y};t_h)} = \frac{p\sum_{h=1}^{m}\mathrm{Rd}(\mathrm{Y};t_h)}{\sum_{h=1}^{m}\mathrm{Rd}(\mathrm{Y};t_h)}\sum_{j=1}^{p} w^2_{hj} = p$ indicates that if the VIP_j of all independent variables x_j(j = 1,2, . . .,p) equals 1, then they all play the same role in interpreting Y. Otherwise, x_j exerts a significant effect on interpreting Y when VIPj> 1.

Results and discussion

Seasonal variation characteristics of $PM_{2.5}$ concentration

The average $PM_{2.5}$ concentration in the BTH urban agglomeration shows a notable periodical U-shaped variation from December 2013 to May 2017. The annual average $PM_{2.5}$ concentration in all cities is 77.79 mg/m^3. The monthly variation of $PM_{2.5}$ concentrations in the BTH region (Fig 7) shows that the $PM_{2.5}$ concentration is below 35 mg/m^3 for only 8.3% of the time (Interim target-1 of WHO, 2005). Specifically, $PM_{2.5}$ concentration is high in autumn and winter and low in spring and summer. This finding is consistent with other findings on China's $PM_{2.5}$ [51, 66, 70]. $PM_{2.5}$ concentration from January to May shows a downward trend and from June to September maintains a stable level at 35–65 μg/m^3, which is slightly lower than those from January to May. $PM_{2.5}$ concentration from October to December exhibits an upward trend, with an increase from 58.95 μg/m^3 to 144.8 μg/m^3. The highest average $PM_{2.5}$ concentration occurs in February 2014 at 142.43 μg/m^3, whereas the lowest average $PM_{2.5}$ concentration occurs in August 2016 at 35.46 μg/m^3. The highest $PM_{2.5}$ concentrations are recorded in December in Shijiazhuang (276.30 mg/m^3), whereas the lowest $PM_{2.5}$ concentrations are recorded in September in Qinhuangdao (14.90 mg/m^3). Therefore, a high $PM_{2.5}$ concentration condition generally occurs from October to January, especially in December. The change in $PM_{2.5}$ concentration every year is regular from January to December. Estimation results in the BTH urban agglomeration are consistent with observations.

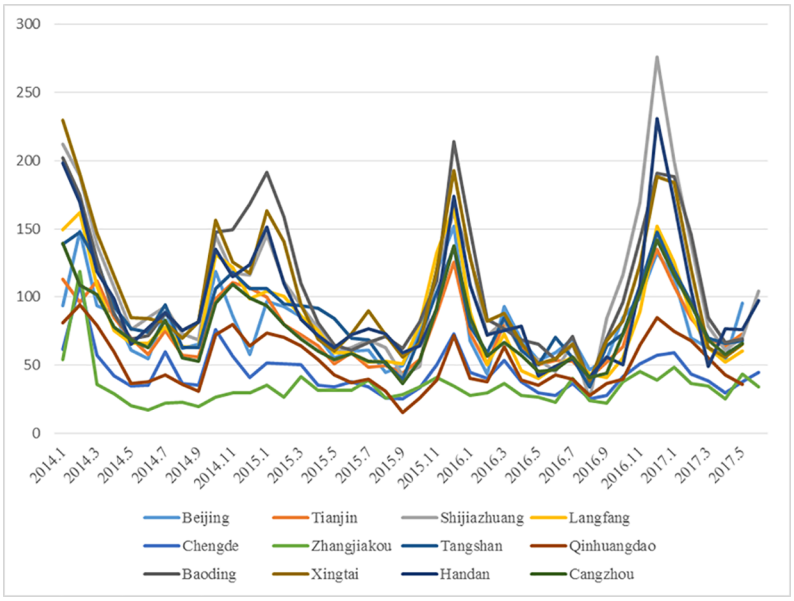

Fig 7. Seasonal variations of $PM_{2.5}$ concentration of 12 cities in the BTH urban agglomeration (Unit: μg/m^3).

https://doi.org/10.1371/journal.pone.0201364.g007

Interpolation results add evidence that $PM_{2.5}$ concentration in the BTH urban agglomeration exhibits a pronounced characteristic of seasonal variation. From December 2013 to May 2017, $PM_{2.5}$ concentration is significantly high in autumn and winter and low in spring and summer. Coal burning in Northern China for winter warming can be expected to be the main contributor to the highest concentration of $PM_{2.5}$ in the country [20, 22, 27, 28]. Average $PM_{2.5}$ concentration measures 135.6 μg/m^3 in winter and 64.1 μg/m^3 in summer, and the total mean values of $PM_{2.5}$ in spring and autumn are 81.5 and 89.6 μg/m^3, respectively.

Related studies have recorded that coal consumption for heating in autumn and winter is the main reason for the difference in values [66]. Coal combustion remains the main way of heating for residents in winter. Thus, government in the BTH regions are proactively promoting coal substation schemes, such as urban gasification projects. Coal is 'dirty' energy, and emissions of particle pollutants are enormous in winter. Meteorological factors also contribute to high $PM_{2.5}$ in winter [39]. Specifically, rainfall in winter is scarce relative to the other seasons, the flushing effect of rainfall to air is little, inhalable particles are easily suspended in air and low temperature in winter is not conducive to the diffusion of $PM_{2.5}$ particles.

This study employed OKM to analyse the seasonal variation of $PM_{2.5}$ concentration of 12 cities in the BTH urban agglomeration from December 2013 to December 2014. The Ministry of Environmental Protection of the People's Republic of China (2018) [38] categorises $PM_{2.5}$ concentration into five classes: (1) 0–50 ug/m^3, (2) 50–100 ug/m^3, (3) 101–200 ug/m^3, (4) 200–300 ug/m^3 and (5) 300–500 ug/m^3. Fig 8 shows that $PM_{2.5}$ concentration for every month is below the fourth classification of 200–300 ug/m^3. The most heavy air pollution areas are recorded in the southern and eastern parts of the BTH region, especially Shijiazhuang City.

Spatial variation characteristics of $pm_{2.5}$ concentration

$PM_{2.5}$ concentration in the BTH urban agglomeration shows a significant spatial variation (Fig 9), which is also recorded in some studies [51, 52, 70]. From 2014 to 2016, $PM_{2.5}$ concentration is significantly high in the south and east of BTH urban agglomeration, particularly in Shijiazhuang. The mean value of $PM_{2.5}$ concentration in Shijiazhuang amounts to 104 μg/m^3. $PM_{2.5}$

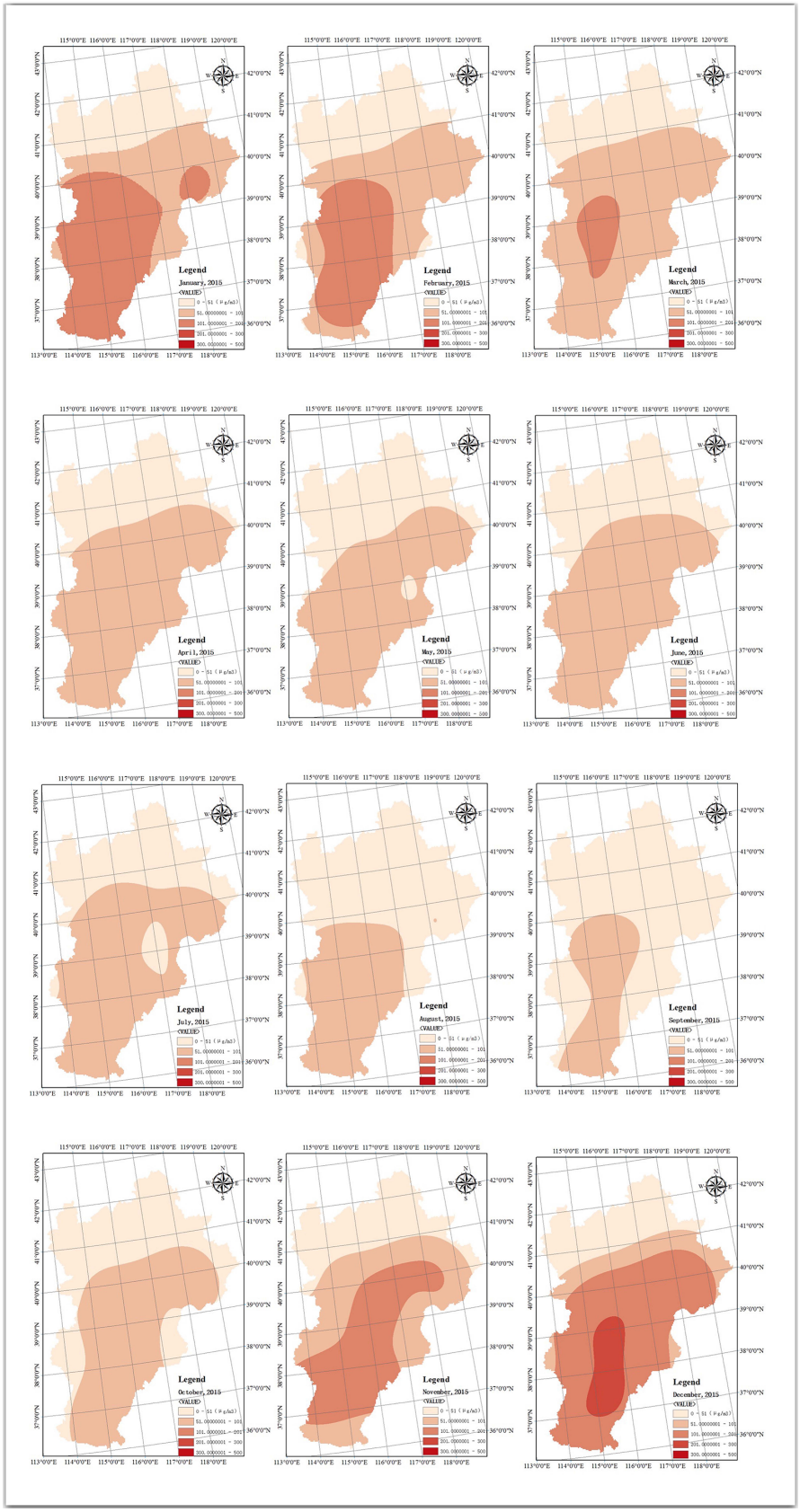

Fig 8. Spatial–temporal variation of $PM_{2.5}$ concentration from January 2015 to December 2015 (Unit: μg/m^3).

https://doi.org/10.1371/journal.pone.0201364.g008

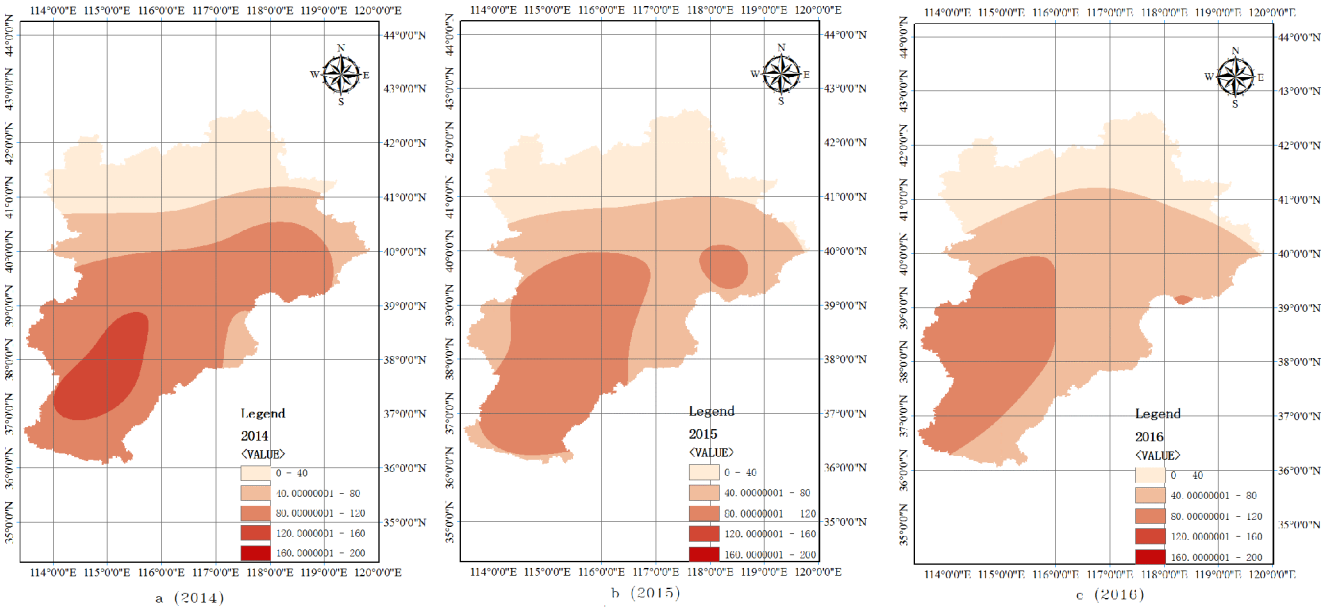

Fig 9. Spatial variation of $PM_{2.5}$ concentration in 2014, 2015 and 2016 (a, b and c) (Unit: μg/m^3).

https://doi.org/10.1371/journal.pone.0201364.g009

concentration is significantly low in the north, particularly in Zhangjiakou. The mean value of $PM_{2.5}$ concentration in Zhangjiakou measures 34 μg/m^3. From 2014 to 2016, the annual $PM_{2.5}$ concentration in the BTH urban agglomeration slightly decreases. By means of 3km satellite aerosol optical depth (AOD) database and geographically and temporally weighted regression (GTWR) modelling, He et al., (2018) [66] also add evidence to the fact that air pollution in the BTH urban agglomeration generally showed a decreasing trend from January 2013 to December 2015.

As shown in Fig 10, the remote sensing data is used to analyse the relationship between $PM_{2.5}$ concentration and population density, landform, geography and elevation. Remote sensing data (Fig 10) shows that $PM_{2.5}$ concentration is low in mountainous and basin regions, such as Zhangjiakou and Chengde, which are situated more than 1000 feet above sea level. High sea level is an advantageous landform for blocking the invasion of $PM_{2.5}$ from peripheral regions [71]. Regions at high sea levels are also subject to northern and northwest winds, which benefit the dissemination of $PM_{2.5}$ pollutants.

Estimation results of PLS regression

Two key plots are considered when examining the performance of PLS regression. The first is t_1/t_2 oval plot.

As the first two extracted linear combinations of $x_1, x_2, \ldots, x_p$, t_1 and t_2 represent the key information of X variables and exhibit remarkable explanatory power for Y variables [71,72]. The underpinning logic is straightforward: when all t_1/t_2 points are covered in the oval, raw data are homogenous and appropriate for model calculations [72]. Fig 11 shows that all the 12 observations of sample points are covered in the oval, indicating that the PLS model is suitable for use in this study.

Another plot that should be considered is the t_1/u_1 scatter plot. When t_1/u_1 of the sample data shows a nearly linear relationship, PLS is appropriate for studying the issue [73]. As shown in Fig 12, the scattering of sample data generally shows a linear relationship. Thus, the PLS model is suitable for use in this study.

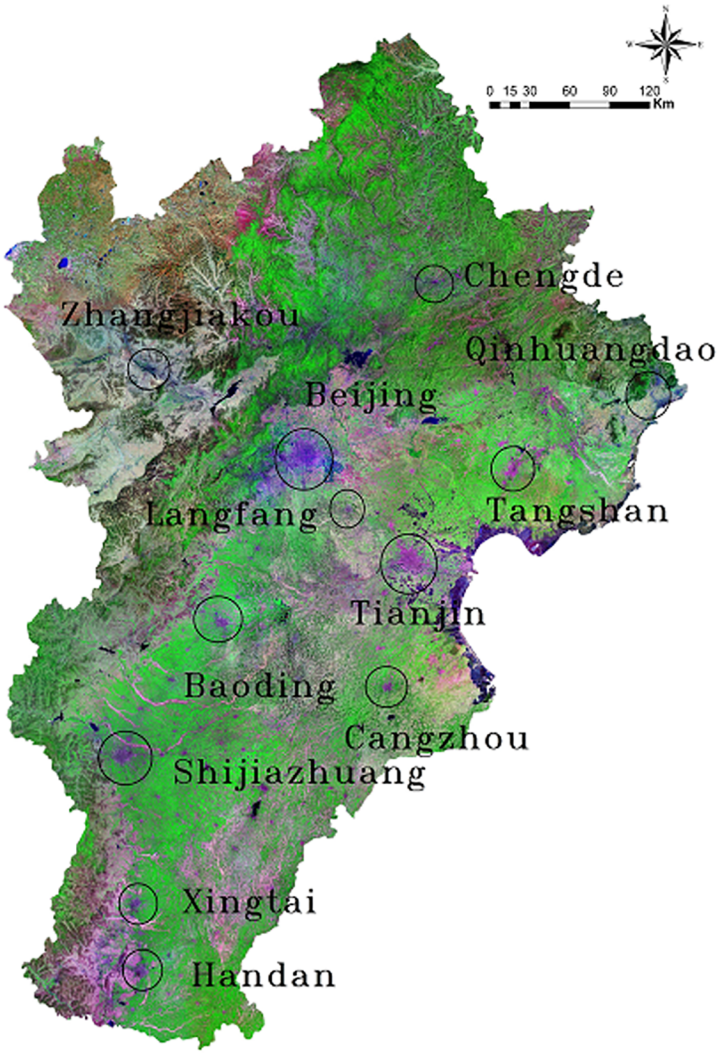

Fig 10. Remote sensing of BTH.

https://doi.org/10.1371/journal.pone.0201364.g010

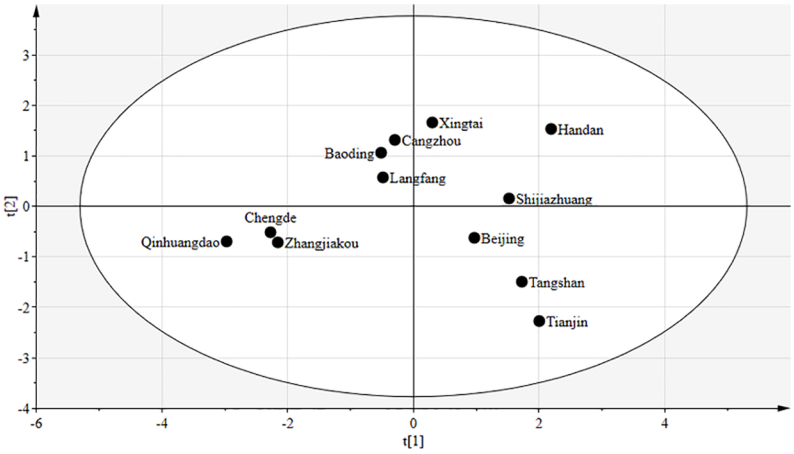

Fig 11. t_1/t_2 oval plot.

https://doi.org/10.1371/journal.pone.0201364.g011

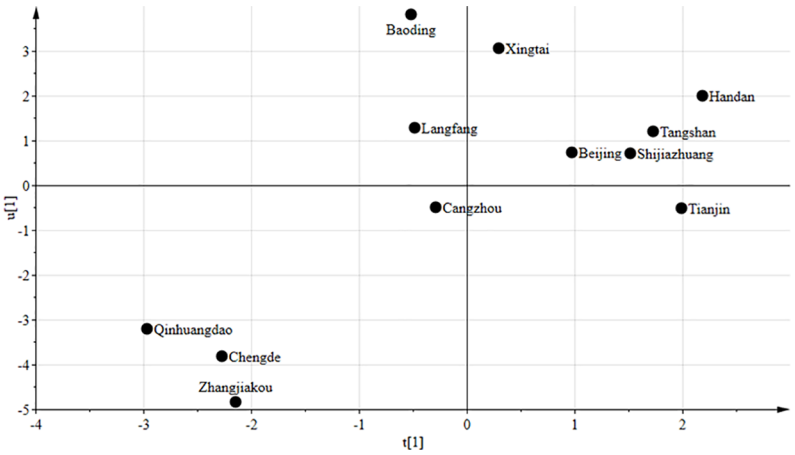

Fig 12. t_1/u_1 scatter plot.

https://doi.org/10.1371/journal.pone.0201364.g012

Table 4 illustrates the overview results of PLS regression. The PLS regression is estimated by using SIMCA-P (a data analytic software). Two principle components are extracted according to the value of cross validation. When two components are extracted, '$R^2X(cum) = 0.510$' indicates that the two components exhibit explanatory powers for 51.0% of the variance of independent variables. '$R^2Y(cum) = 0.681$' indicates that the two extracted components explain 68.1% of the information on dependent variables, indicating the acceptable explanation power of the PLS method.

VIP is a critical parameter for measuring the fitting performance of the PLS model [33, 69]. It quantifies the statistical significance of independent variables (X) in explaining dependent variables. When the VIP value of a variable is higher than 0.8, then the variable is 'important' and exhibits significant explanatory powers on the independent variables [69]. Table 5 presents the significance of 12 independent variables in evaluating the $PM_{2.5}$ index amongst 12 cities of the BTH region.

Table 5 shows that the VIP values of most variables reach more than 0.8, that is, most variables are significant in explaining the $PM_{2.5}$ concentration. The VIP values of population density, urbanisation rate and road freight volume are higher than 1.0. These results indicate that the variables are the three top contributors to $PM_{2.5}$ concentration in the 12 sample cities of the BTH urban agglomeration. The VIP values of secondary industry GDP, overall energy consumption, steel production and volume of sulphur dioxide emission are higher than 0.8. Therefore, these factors are also major drivers of $PM_{2.5}$ pollutant emission. The rest of the variables with VIP values in the spectrum of 0.5–0.8 cannot be assessed or are unimportant drivers of pollutant emission [72].

The VIP value of population density is 1.774, which is the highest explanatory power for $PM_{2.5}$ concentration. The VIP value of urbanisation rate in explaining $PM_{2.5}$ concentration is 1.476, which ranks second, as shown in Fig 13. Previous studies have demonstrated that $PM_{2.5}$ concentration is particularly high in large cities and urbanised regions [1, 2, 4, 5]. BTH is one

Table 4. Overview of PLS regression results.

Number of components	$R^2X(cum)$	$R^2Y(cum)$
1	0.339	0.424
2	0.510	0.681

https://doi.org/10.1371/journal.pone.0201364.t004

Table 5. VIP values of factors.

Abbreviation	Corresponding variable	VIP value
PD	Population density	1.774
UR	Urbanisation rate	1.476
RFV	Road freight volume	1.157
SIGDP	Secondary industry GDP	0.953
OEC	Overall energy consumption	0.890
SP	Steel production	0.889
VOSDE	Volume of Sulphur Dioxide Emission	0.864
VOISE	Volume of Industrial Soot(dust) Emission	0.784
CP	Cement production	0.762
MVO	Motor vehicle ownership	0.749
RPTV	Road passenger traffic volume	0.504
RNGC	Residential natural gas consumption	0.295

https://doi.org/10.1371/journal.pone.0201364.t005

of the most urbanised, populous and developed urban agglomerations in China, and human activities, such as transport, productions of secondary industry and energy consumption, are intensive in this region. Active human activities demand considerable resource and energy and create significant traffic daily. Numerous studies have reported that vehicular exhaust is the main source of $PM_{2.5}$ [61]. Coal and oil, as cost-effective energy, have long been used as main fuels for the secondary industry in developing countries [30]. In particular, in industrial zones of BTH, coal consumption is the main fuel for energy- and emission-intensive sectors, such as those of steel, cement and glass, which are key materials for China's remarkable infrastructure build-up in the past decades. Coal combustion is the main air pollution source in this region [59].

The VIP value of road freight volume totals 1.157 in explaining $PM_{2.5}$ concentration. Previous studies have demonstrated that road vehicular exhaust is highly related to $PM_{2.5}$ concentration [26, 29, 62, 63]. According to literature, road freight volume, particularly those of heavy-duty trucks, is the main air pollution source in the transportation sector in some cities [64]. However, in the current study, the VIP value of road passenger traffic volume reaches only 0.504, which is much less than that of road freight volume. Therefore, this factor is not a primary air pollutant source.

The VIP value of the secondary industry GDP totals 0.953 in explaining $PM_{2.5}$ concentration. In the industrial zones of BTH, such as Tangshan, coal is the main energy in different

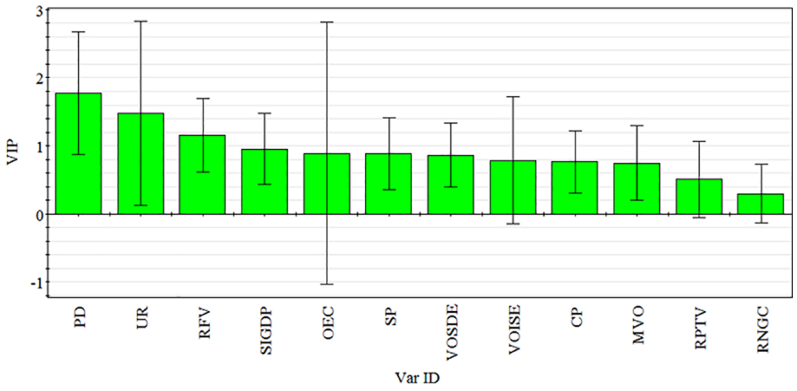

Fig 13. Variable importance plot.

https://doi.org/10.1371/journal.pone.0201364.g013

sectors because of their copious coal deposits. However, carbonaceous aerosols are the main source of $PM_{2.5}$. Coal combustion and sulphur dioxide are the dominant sources of carbonaceous aerosols [61]. Secondary industry, heating for residences in winter and electricity from coal-fired stations rely heavily on coal combustion [20, 22, 27, 28]. Thus, the VIP value of overall energy consumption in explaining $PM_{2.5}$ concentration is high at 0.890. The VIP value of steel production in BTH in explaining $PM_{2.5}$ concentration is 0.889, because metal elements correspond to other sources of $PM_{2.5}$ in China. The process of producing steel requires significant coal combustion and releases abundant metal elements and sulphur dioxide into air [61]. Studies have recorded that $PM_{2.5}$ concentrations exhibit a notable and positive relationship with associated air pollutants, such as SO_2, NO_2 and O_3, and suggested that those atmospheric pollutants can evolve from primary pollution to secondary pollution and form a vicious cycle [39, 74].

As China's political and economic centre, the particulate matter (PM) pollution in the Beijing-Tianjin-Hebei (BTH) urban agglomeration attracts extensive attention of the scholars. Many recent studies investigate $PM_{2.5}$ concentration characteristics in the BTH region by using various methods and data sources. Similarities and new evidences of empirical findings of this study are compared with those of related studies.

First, some findings of this research are generally consistent with conclusions of existing studies. For example, this study reconfirmed that the annual $PM_{2.5}$ concentration experienced a slight downturn in recent years, which is also recorded by He et al (2018) [67]. For temporal characteristics of $PM_{2.5}$ concentration, the study captured a periodic U-shaped variation pattern in BTH urban agglomeration with high pollution levels in autumn and winter and low levels in spring and summer. Yan et al. (2018) [51] also recorded a pronounced characteristic of seasonal variation. They found that the concentrations increased from late autumn to early winter and that the $PM_{2.5}$ concentration decreased rapidly from late winter to early spring. For spatial aspects, empirical results of this study find that the south and east of the BTH urban agglomeration where are densely populated are suffered with highest $PM_{2.5}$ concentration and that $PM_{2.5}$ concentration is significantly low in the north. Similar spatial characteristics are also recorded in existing studies [52, 67].

Second, this study employed data with a long period from December 2013 to May 2017, which ensures robust and complete understandings on $PM_{2.5}$ concentration characteristics in the region. Most studies on the BTH urban agglomeration used either one year cross-sectional data or daily time series data [50, 51, 70] or outdated data before 2015 [67]. For example, based on daily monitoring data from 1 January 2014 to 31 December 2014, Liu et al. (2018) [50] investigated dynamic interactions and relationships between $PM_{2.5}$ concentrations in different cities. Estimation results based on a short period of data undermined the understandings on the temporal characteristics of $PM_{2.5}$. In addition, since 2015, Chinese government has committed great efforts and resources in $PM_{2.5}$ treatment and special focuses are dedicated in the BTH region. Findings with data before 2015 cannot reflect the recent characteristic of $PM_{2.5}$ concentrations and effectiveness of $PM_{2.5}$ treatment.

Third, this research combines the satellite sensing data and monitoring sites data. Some academics carried out their studies based on satellite sensing data [52,67,75], meanwhile others used the data obtained from surface monitoring stations [51, 70]. In this study, the surface monitoring data of 12 cities in the regions are measured by 80 monitoring stations distributed throughout the BTH region, which is used for statistical modeling. In addition, remote sensing data is processed by OKM method to uncover the temporal and spatial characteristics. The remote sensing data are also used to analyze the relationship between $PM_{2.5}$ concentration and landform, geography and elevation. Remote sensing data (Fig 10) shows that $PM_{2.5}$ concentration is low in mountainous and basin regions, which are situated more than 1000 feet above sea level.

Fourth, this study focused on the driving factors of $PM_{2.5}$ concentration, besides one goal of investigating the temporal and spatial characteristics. Related studies on the BTH regions are generally salient on diving facts of $PM_{2.5}$ concentration. For example, Zheng et al. (2018) [70] focused on improvement of the real-time forecast. Liu et al.(2018) [50] aim to visualize the dynamic interactions and relationships between $PM_{2.5}$ in different cities in the BTH regions. Yan et al. (2018) [51] investigated the spatiotemporal pattern of $PM_{2.5}$ concentrations in China. In this study, a large number of socioeconomic factors are included in the investigation. Partial least squares (PLS) method to explore the critical driving factors of PM2.5, which is rarely used in $PM_{2.5}$ research. Estimation bias caused by multicollinearity of raw data can be avoided and it can lead to a robust estimation results. Empirical results demonstrate that the deterioration of $PM_{2.5}$ concentration in 2015 is closely related to a set of critical impact factors, including population, transport, industry production, and energy consumption aspects.

Conclusion

$PM_{2.5}$ is a challenging and urgent air pollutant that should be fully treated in China. The BTH region is heavily exposed to serious $PM_{2.5}$ pollution. Previous studies are limited by the lack of real-time monitoring data and poor consideration of multicollinearity issues amongst dependent variables. This study aims to quantitatively measure the spatial–seasonal concentration characteristics of $PM_{2.5}$ and identify critical impact factors. Empirical findings reveal the following. (1) Notable differences in $PM_{2.5}$ concentrations exist amongst different seasons. Specifically, a periodical U-shaped variation trend with high pollution levels is found in autumn and winter and one with low levels is observed in spring and summer. (2) An apparent spatial distribution pattern exists in which $PM_{2.5}$ concentration in the south is higher than in the northern regions in BTH. (3) The deterioration of air pollution is closely related to several critical impact factors, including population density, urbanisation rate, road freight volume, secondary industry GDP, overall energy consumption and industrial pollutants (e.g. steel production and volume of sulphur dioxide emission).

Numerous viable and concrete police recommendations are provided for effective $PM_{2.5}$ treatment. 1) Decentralisation policy is a viable alternative policy for improvement of $PM_{2.5}$ treatment. Decentralisation policy is an effective approach for downsizing urban population and relieving congestion [76, 77]. Empirical studies verify a tight relation between population density and $PM_{2.5}$ concentration. For example, rapid population influx in Beijing leads to $PM_{2.5}$ concentration deterioration due to associated energy consumption and production. Population and emission-intensive industries should be decentralised outward under the cooperated plan of BTH integration. 2) Expansions of high-energy consumption and emission industries should be constrained. Non-coal or renewable energies, such as natural gas, solar, biomass and wide and regular hydro power, should be encouraged for wide commercialisation and deployment. Statuary standards for energy use efficiency and pollutant emission should be reinforced to raise the consciousness and capability of enterprises on $PM_{2.5}$ treatments. 3) Mass transportation should be encouraged by the government and the public. For example, a complete and convenient mass transportation facility should be developed by the government to improve public ridership rate; the number of fuel-based motor vehicle ownership should be restricted; non-fuel cars, such as pure electric and plug-in new energy, should be encouraged; and diverse low-emission transportation modes, such as walking, cycling, car sharing and subway, should be used by urban residents. 4) Winter is the most heavily polluted season in the BTH region. Heating and vast coal combustion are associated, to a large extent, to contributing to high $PM_{2.5}$ emission. In addition, meteorological conditions in winter are not conducive to the purification and diffusion of $PM_{2.5}$ particles, thus adding to $PM_{2.5}$ pollution. Ma and

Zhang (2014) [16] emphasised that the wide use of imported low-calorie coal and lignite in industrial and domestic sectors are particularly adverse for treatments of $PM_{2.5}$ pollution. The Chinese government should reduce the use of low-calorie coal and promote energy substations for clean heating energy. 5) $PM_{2.5}$ pollution is more serious in the south than in the north. The government should formulate an industrial restructuring scheme to avoid excessive concentration of polluting industries in certain regions, plan scientific reallocation for highly polluting industries and upgrade the energy use efficiency of industrial sectors.

Supporting information

S1 Table. Geographic information of atmospheric physics observation points. The longitude and latitude information of 80 atmospheric physics observation points are reported.
(DOC)

S2 Table. PM2.5 concentration data of 12 cities in the beijing-tianjin-hebei urban agglomeration. $PM_{2.5}$ concentration data covers the period from December 2013 to May 2017 cover 12 cities.
(DOC)

S3 Table. The possible critical impact factors of PM2.5 concentration in 2015. Consensus data of 12 potential contributing factors for $PM_{2.5}$ concentrations are reported.
(DOC)

Author Contributions

Conceptualization: Tianhang Huang.

Data curation: Xuchang Chen.

Methodology: Huiwen Wang.

Software: Wenyang Huang.

Visualization: Wenyang Huang.

Writing – original draft: Tianhang Huang, Yigang Wei.

Writing – review & editing: Yunjiang Yu, Yigang Wei.

References

1. Zhu YG, Ioannidis JPA, Li H, Jones KC, Martin FL. Understanding and harnessing the health effects of rapid urbanization in China. Environmental Science & Technology. 2011; 45(12):5099.
2. Gong P, Liang S, Carlton EJ, Jiang Q, Wu J, Wang L, et al. Urbanisation and health in China. Lancet. 2012; 379(9818):843–852. https://doi.org/10.1016/S0140-6736(11)61878-3 PMID: 22386037
3. Wei Y, Huang C, Lam PTI, Sha Y, Feng Y. Using urban-carrying capacity as a benchmark for sustainable urban development: an empirical study of Beijing. Sustainability. 2015; 7(3):3244–3268.
4. Han L, Zhou W, Li W, Li L. Impact of urbanization level on urban air quality: a case of fine particles ($PM_{2.5}$) in Chinese cities. Environmental Pollution. 2014; 194C(1), 163.
5. Han L, Zhou W, Li W. Increasing impact of urban fine particles (PM2.5) on areas surrounding Chinese cities. Scientific Reports. 2015; 5:12467 https://doi.org/10.1038/srep12467 PMID: 26219273
6. World Health Organization (WHO). Ambient (outdoor) air pollution in cities database 2014. 2014 [cited 2017-09-18]. Available from: http://www.who.int/phe/health_topics/outdoorair/databases/cities/en/.]
7. Ulrich POschl. Atmospheric aerosols: composition, transformation, climate and health effects. Cheminform. 2005; 44(46):7520.
8. Chan CK, Yao X. (2008). Air pollution in mega cities in China. Atmospheric Environment. 2008; 42 (1):1–42.

9. Kampa M, Castanas E. Human health effects of air pollution. Environmental Pollution. 2008; 151 (2):362. https://doi.org/10.1016/j.envpol.2007.06.012 PMID: 17646040
10. Lim S, Vos T, Bruce N. The burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions 1990–2010: a systematic analysis. Lancet. 2012; 380(9859):2224–60. https://doi.org/10.1016/S0140-6736(12)61766-8 PMID: 23245609
11. Pope C III. Dockery D. Health effects of fine particle air pollution: Lines that connect. Journal of the Air & Waste Management Association. 2012; 56, 709–742.
12. Querol X, Alastuey A, Moreno T, Viana MM, Castillo S, Pey J, et al. Spatial and temporal variations in airborne particulate matter (pm and pm2.5) across Spain 1999–2005. Atmospheric Environment. 2008; 42(7):3964–3979.
13. Pateraki S, Maggos T, Michopoulos J, Flocas HA, Asimakopoulos DN, Vasilakos C. Ions species size distribution in particulate matter associated with vocs and meteorological conditions over an urban region. Chemosphere. 2008; 72(3):496. https://doi.org/10.1016/j.chemosphere.2008.02.061 PMID: 18440047
14. Anttila P. Trends of primary and secondary pollutant concentrations in Finland in 1994–2007. Atmospheric Environment. 2010; 44(1):30–41.
15. Barmpadimos I, Hueglin C, Keller J, Henne S, T ASHP. Influence of meteorology on PM_{10} trends and variability in Switzerland from 1991 to 2008. Atmospheric Chemistry & Physics & Discussions. 2011; 11 (4):1813–1835.
16. Ma L, Zhang X. The spatial effect of China's haze pollution and the impact from economic change and energy structure. China Industrial Economics. 2014; 4:19–31.
17. Rodes CE, Evans EG. Preliminary assessment of 10 μm particulate sampling at eight locations in the United States. Atmospheric Environment. 1985; 19(2):293–303.
18. Darlington TL, Kahlbaum DF, Heuss JM, Wolff GT. Analysis of PM_{10} trends in the United States from 1988 through 1995. Journal of the Air & Waste Management Association. 1997; 47(10):1070–1078.
19. Lefèvre RA, Ionescu A, Desplat J, Kounkou-Arnaud R, Perrussel O, Languille B. Quantification and mapping of the impact of the recent air pollution abatement on limestone and window glass in Paris. Environmental Earth Sciences. 2016; 75(20): 1359.
20. Cao JJ, Shen ZX, Chow JC, Watson JG, Lee SC, Tie XX, et al. Winter and summer $PM_{2.5}$ chemical compositions in fourteen Chinese cities. Journal of the Air & Waste Management Association. 2012; 62 (10):1214–26.
21. Ward T. Source apportionment of $PM_{2.5}$ in a subarctic Airshed—Fairbanks, Alaska. Aerosol & Air Quality Research. 2012; 12(4): 536–543.
22. Zhang R, Jing J, Tao J, Hsu SC. Chemical characterization and source apportionment of PM2.5 in Beijing: seasonal perspective. Atmospheric Chemistry &. 2013;(14):7053–7074.
23. Tao J, Gao J, Zhang L, Zhang R, Che H, Zhang Z, et al. $PM_{2.5}$ pollution in a megacity of southwest China: source apportionment and implication. Atmospheric Chemistry & Physics. 2014; 14(4): 8679–8699.
24. Wang G, Cheng S, Li J, Lang J, Wen W, Yang X, et al. Source apportionment and seasonal variation of $PM_{2.5}$ carbonaceous aerosol in the beijing-tianjin-hebei region of China. Environmental Monitoring & Assessment. 2015; 187(3):143.
25. Wang SR, Yu YY, Wang QG, Lu Y, Yin LN, Zhang YY, et al. Source apportionment of $PM_{2.5}$ in Nanjing by PMF. China Environ. Sci. 2015; 35 (12): 3535–3542. (in Chinese)
26. Cesari D, De Benedetto G, Bonasoni P, Busetto M, Dinoi A, Merico E, et al. Seasonal variability of $PM_{2.5}$ and PM_{10} composition and sources in an urban background site in Southern Italy. Science of the Total Environment. 2018; 612:202–213. https://doi.org/10.1016/j.scitotenv.2017.08.230
27. He K, Yang F, Ma Y, Zhang Q, Yao X, Chan CK, et al. The characteristics of $PM_{2.5}$ in Beijing, China. Atmospheric Environment. 2001; 35(29): 4959–4970.
28. Ye B, Ji X, Yang H, Yao X, Chan CK, Cadle SH, et al. Concentration and chemical composition of $PM_{2.5}$, in Shanghai for a 1-year period. Atmospheric Environment. 2003; 37(4): 499–510.
29. Aldabe J, Elustondo D, Santamaría C, Lasheras E, Pandolfi M, Alastuey A, et al. Chemical characterisation and source apportionment of $PM_{2.5}$ and PM_{10} at rural, urban and traffic sites in Navarra (north of Spain). Atmospheric Research. 2011; 102(1):191–205.
30. Lonati G, Butelli GP, Romele L, Tardivo R. Major chemical components of $PM_{2.5}$ in Milan. Atmospheric Environment. 2005; 39(10): 1925–1934. Italy.
31. Ram K, Sarin MM, Hegde P. Atmospheric abundances of primary and secondary carbonaceous species at two high-altitude sites in India: sources and temporal variability. Atmospheric Environment. 2008; 42(28):6785–6796.

32. Donkelaar AV, Martin RV, Brauer M, Kahn R, Levy R, Verduzco C, Villeneuve PJ. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. Environmental Health Perspectives. 2010; 118 (6):847–855. https://doi.org/10.1289/ehp.0901623 PMID: 20519161

33. Wei YG, Gu J,Wang HW, Tang Y, Yao T. Uncovering the Culprits of Air Pollution: Evidence from China's Economic Sectors and Regional Heterogeneities, Journal of Cleaner Production. 2018; 171:1481–1493.

34. Beijing Environmental Protection Bureau. Report on the State of the Environment in Beijing in 2002. 2004 [cited 2017-08-16]. Available from: www.bjepb.gov.cn/bjhb/tabid/68/InfoID/2618/Default.aspx.

35. Beijing Environmental Protection Bureau Report on the State of the Environment in Beijing in 2004. 2005 [cited 017-08-16]. Available from: www.bjepb.gov.cn/bjhb/tabid/68/InfoID/2620/Default.aspx.

36. Zheng M, Salmon LG, Schauer JJ, Zeng L, Kiang CS, Zhang Y, et al. Seasonal trends in $PM_{2.5}$ source contributions in Beijing, China. Atmospheric Environment. 2005; 39(22): 3967–3976.

37. Liu Z, Hu B, Wang L, Wu F, Gao W, Wang Y. Seasonal and diurnal variation in particulate matter (PM_{10} and $PM_{2.5}$) at an urban site of Beijing: analyses from a 9-year study. Environmental Science & Pollution Research. 2015; 22(1):627–642.

38. Ministry of Environmental Protection of the People's Republic of China. China Environmental Status Bulletin 2016. 2016 [cited 2018-01-10]. Available from: http://www.mep.gov.cn/hjzl/zghjzkgb/lnzghjzkgb/201706/P020170605833655914077.pdf.

39. Yang M, Wang Y. Spatial temporal characteristics of $PM_{2.5}$ and its influencing factors in the Yangtze River economic belt. China Population, Resources and Environment. 2017; 27(1): 91–100. Chinese

40. Aaron VD, Martin RV, Michael B, Ralph K, Robert L, Carolyn V, et al. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. Environmental Health Perspectives. 2010; 118(6):847. https://doi.org/10.1289/ehp.0901623 PMID: 20519161

41. Tao J, Zhang L, Engling G, Zhang R, Yang Y, Cao J, et al. Chemical composition of $PM_{2.5}$, in an urban environment in Chengdu, China: importance of springtime dust storms and biomass burning. Atmospheric Research. 2013; 122(3), 270–283.

42. Shen Y, Yao L. $PM_{2.5}$, Population exposure and economic effects in urban agglomerations of China using ground-based monitoring data. Environmental research and public health. 2017; 14(7): 716.

43. Zhou S, Ouyang W, Ge J. Study on the main influencing factors and their intrinsic relations of $PM_{2.5}$ in Beijing-Tianjin-Hebei. China Population, Resources and Environment. 2017; 27(4):102–109. Chinese.

44. Li Y, Yin C. Threshold effect of socio-economic factors influencing $PM_{2.5}$ in Beijing-Tianjin-Hebei. China Environmental Science. 2017; 37(4):1223–1230. Chinese.

45. Sun J, Zhong Y. Economic analysis on the factors influencing $PM_{2.5}$ in China's metropolises: An empirical study based on city-level panel data. Ecological Economy. 2015; 31(3): 62–65. Chinese.

46. Boys BL, Martin R V, Van DA, Macdonell RJ, Hsu NC, Cooper MJ, et al. Fifteen-year global time series of satellite-derived fine particulate matter. Environmental Science & Technology. 2014; 48(19):11109.

47. Song Y, Wang X, Maher BA, Li F, Xu C, Liu X, et al. The spatial-temporal characteristics and health impacts of ambient fine particulate matter in China. Journal of Cleaner Production. 2016; 112:1312–1318.

48. Gang L, Fu J, Dong J, Hu W, Dong D, Huang Y, et al. Spatio-temporal variation of PM2.5 concentrations and their relationship with geographic and socioeconomic factors in China. International Journal of Environmental Research & Public Health.2014; 11(1): 173–86. https://doi.org/10.3390/ijerph14020156

49. Tian S, Pan Y, Liu Z, Wen T, Wang Y. Size-resolved aerosol chemical analysis of extreme haze pollution events during early 2013 in urban Beijing, China. Journal of Hazardous Materials. 2014; 279: 452–460. https://doi.org/10.1016/j.jhazmat.2014.07.023 PMID: 25106045

50. Liu J, Li W, Wu J, Liu Y. Visualizing the intercity correlation of $PM_{2.5}$ time series in the Beijing-Tianjin-Hebei region using ground-based air quality monitoring data. PloS one.2018; 13(2):p.e0192614. https://doi.org/10.1371/journal.pone.0192614 PMID: 29438417

51. Yan D, Lei Y, Shi Y, Zhu Q, Li L, Zhang Z. Evolution of the spatiotemporal pattern of $PM_{2.5}$ concentrations in China–A case study from the Beijing-Tianjin-Hebei region. Atmospheric Environment.2018; 183: 225–233.

52. Zhai L, Li S, Zou B, Sang H, Fang X and Xu S. 2018. An improved geographically weighted regression model for $PM_{2.5}$ concentration estimation in large areas. Atmospheric Environment.2018; 181:pp.145–154.

53. An Z, Qi Q, Jiang L, Fang Z, Wang J. Population exposure to $PM_{2.5}$ in the urban area of Beijing. Plos One. 2013; 8(5): e63486. https://doi.org/10.1371/journal.pone.0063486 PMID: 23658832

54. Lee SJ, Serre ML, Van DA, Martin RV, Burnett RT, Jerrett M. Comparison of geostatistical interpolation and remote sensing techniques for estimating long-term exposure to ambient $PM_{2.5}$ concentrations across the continental united states. Environmental Health Perspectives. 2012; 120(12):1727. https://doi.org/10.1289/ehp.1205006 PMID: 23033456

55. Beijing Municipal Government. 2016 [cited 2017-02-16]. Available from: http://www.bjstats.gov.cn/tjsj/tjgb/ndgb/201702/t20170227369467.html.

56. Atmospheric Composition Analysis Group Satellite-Derived PM2.5, 2015, at 35% RH [ug/m3]. 2016 [cited 2017-08-16]. Available from: http://fizz.phys.dal.ca/~atmos/martin/?page_id=140.

57. Yang S, Chen B. Driving Forces of Particulate Matter Emissions in China. Energy Procedia. 2017;4601–4606.

58. Lu D, Xu J, Yang D, Zhao J. Spatio-temporal variation and influence factors of PM2.5 concentrations in China from 1998 to 2014. Atmospheric Pollution Research.2017. https://doi.org/10.1016/j.apr.2016.10.009

59. Ma Z, Xiao H. The research on a spatial differentiation of influence factors of regional $PM_{2.5}$ in China—The empirical analysis based on geographically weighted regression model. Journal of Shanxi University of Finance and Economics.2017; 5: 14–26. Chinese.

60. Yin P. Research of combined cooling heating and power systems: Haze and DeNOx. Heating Ventilating & Air Conditioning. 2014; (6): 49–55. Chinese.

61. Wang H, Gu J, Chen M. Pollution industry concentration and air pollution emission comprehensive assessment. System Engineering. 2016; (11):59–63.

62. Park SS, Bae MS, Kim YJ. Chemical composition and source apportionment of $PM_{2.5}$ particles in the Sihwa area, Korea.Journal of the Air & Waste Management Association. 2001; 51(3), 393.

63. Franco V, Kousoulidou M, Muntean M, Ntziachristos L, Hausberger S, Dilara P. Road vehicle emission factors development: A review. Atmospheric Environment.2013; 70(70):84–97.

64. Fameli KM, Assimakopoulos VD. Development of a road transport emission inventory for greece and the greater athens area: effects of important parameters. Science of the Total Environment. 2015; 505:770–786.Okabe A, Boots B, Sugihara K, Chiu SN, Kendall DG.Spatial Interpolation. Springer US.2008. https://doi.org/10.1016/j.scitotenv.2014.10.015 PMID: 25461080

65. Hase H, & Pedreros F. The most remote point method for the site selection of the future ggos network. Journal of Geodesy. 2014; 88(10):989–1006.

66. Wang Z, Fang C, Guang XU, Pan Y. Spatial-temporal characteristics of the $PM_{2.5}$ in China in 2014. Acta Geographica Sinica. 2015; 70(11):1720–1734.

67. He Q, Huang B. Satellite-based high-resolution PM2. 5 estimation over the Beijing-Tianjin-Hebei region of China using an improved geographically and temporally weighted regression model. Environmental Pollution. 2018; 236:1027–1037. https://doi.org/10.1016/j.envpol.2018.01.053 PMID: 29455919

68. Yang GL, Zhang GM, Shi-Xin LI. Application of universal Kriging interpolation in geomagnetic map. Journal of Chinese Inertial Technology. 2008.

69. Palermo G, Piraino P, Zucht HD. Performance of PLS regression coefficients in selecting variables for each response of a multivariate PLS for omics-type data. Advances & Applications in Bioinformatics & chemistry Aabc. 2009; 2(2):57–70.

70. Zheng H, Liu J, Tang X, Wang Z, Wu H, Yan P, Wang W. Improvement of the Real-time $PM_{2.5}$ Forecast over the Beijing-Tianjin-Hebei Region using an Optimal Interpolation Data Assimilation Method. Aerosol and Air Quality Research. 2018; 18:1305–1316.

71. Wang ZB, Zou B, Qiu YH, Chen JW. Geographical characterization of relationship between aerosol optical depth and $PM_{2.5}$ in China. Remote Sensing Information. 2016; 31(6):26–35.

72. Wang HW. Linear and nonlinear methods of partial least squares regression. National Defense Industry Press, Beijing. 2006. Chinese.

73. Jia JS, Deng HB, Duan J, Zhao JZ. Analysis of the major drivers of the ecological footprint using the STIRPAT model and the PLS method-a case study in Henan Province, China. Ecological Economics. 2009; 68: 2818–2824.

74. Lv B, Cai J, Xu B, Bai Y. Understanding the rising phase of the $PM_{2.5}$ concentration evolution in large China cities. Scientific Reports.2017; 7: 46456. https://doi.org/10.1038/srep46456 PMID: 28440282

75. Huang TH, Dang AR, Liu RL, Zhu ZQ, Zhang DM, Li ST. Relationship between Chinese Weihai urbanization and climate change. Applied Mechanics & Materials. 2013; 368–370, 386–389.

76. Wei YG, Huang C, Lam PTI, Yuan ZY. Sustainable Urban Development: A Review on Urban Carrying Capacity Assessment. Habitat International. 2015; 46:64–71.

77. Wei Y, Huang C, Li J, Xie L. An evaluation model for urban carrying capacity: A case study of China's mega-cities. Habitat International.2016; 53:87–96.