

SCIENTIFIC REPORTS



OPEN

Quantitative association analysis between PM_{2.5} concentration and factors on industry, energy, agriculture, and transportation

Nan Zhang¹, Hong Huang², Xiaoli Duan³, Jinlong Zhao² & Boni Su⁴

Rapid urbanization is causing serious PM_{2.5} (particulate matter $\leq 2.5 \mu\text{m}$) pollution in China. However, the impacts of human activities (including industrial production, energy production, agriculture, and transportation) on PM_{2.5} concentrations have not been thoroughly studied. In this study, we obtained a regression formula for PM_{2.5} concentration based on more than 1 million PM_{2.5} recorded values and data from meteorology, industrial production, energy production, agriculture, and transportation for 31 provinces of mainland China between January 2013 and May 2017. We used stepwise regression to process 49 factors that influence PM_{2.5} concentration, and obtained the 10 primary influencing factors. Data of PM_{2.5} concentration and 10 factors from June to December, 2017 was used to verify the robustness of the model. Excluding meteorological factors, production of natural gas, industrial boilers, and ore production have the highest association with PM_{2.5} concentration, while nuclear power generation is the most positive factor in decreasing PM_{2.5} concentration. Tianjin, Beijing, and Hebei provinces are the most vulnerable to high PM_{2.5} concentrations caused by industrial production, energy production, agriculture, and transportation (IEAT).

PM_{2.5} (particulate matter $\leq 2.5 \mu\text{m}$) has recently emerged as a serious pollutant in many countries. Some Asian countries such as India¹, Japan², and Malaysia³ have severe PM_{2.5} pollution problems. The situation has been very challenging in China^{4,5}. China is experiencing extremely rapid urbanization which leads to high risk of PM_{2.5} pollution. This problem has attracted wide attention since 2013⁶. In Beijing, the capital of China, the highest daily average PM_{2.5} concentration measured was more than $500 \mu\text{g}/\text{m}^3$, which is 20-fold higher than the WHO guideline value⁷. Similar high concentrations have occurred several times in the past four years. According to a report from the Asian-development Bank, less than 1% of the 500 largest cities in China meet the WHO air quality guideline values ($10 \mu\text{g}/\text{m}^3$ for the annual mean and $25 \mu\text{g}/\text{m}^3$ for the daily mean)⁸. Long-term exposure to PM_{2.5} has been consistently linked to heart and lung disease⁹, and reduces personal resistance¹⁰. For every PM_{2.5} concentration increase of $10 \mu\text{g}/\text{m}^3$, the risk of emergency hospital admissions for cerebrovascular diseases increased by 1.29%¹¹. An aggressive global program of PM_{2.5} mitigation in line with WHO guidelines could avoid 750,000 (23%) of the current 3.2 million associated deaths per year¹². Ways to efficiently control PM_{2.5} under rapid urbanization and aging populations need to urgently be found.

The Chinese government has placed great emphasis on PM_{2.5} control. The Chinese State Council released the 'Atmospheric Pollution Prevention and Control Action Plan' on September 10, 2013 which aimed to reduce PM_{2.5} by up to 25% by 2017 relative to the 2012 level¹³. At an APEC meeting in October 2014, Chinese President Xi Jinping stated that China will make a concerted effort to reduce air pollution¹⁴. In the 13th Five-Year Plan (2016–2020), environmental pollution control is one of the first tasks specified. The plan also advocates the use of 'big' data to support pollution mitigation¹⁵.

Satellite and ground monitoring data are usually used to predict PM_{2.5} in the short term^{16,17}. Models such as the WRF-Chem model¹⁸, the GEOS-Chem model¹⁹, and the CMAQ model²⁰ are also used to analyze PM_{2.5}

¹Department of Mechanical Engineering, The University of Hong Kong, Hong Kong, China. ²Institute of Public Safety Research, Department of Engineering Physics, Tsinghua University, Beijing, China. ³School of Energy and Environmental Engineering, University of Science and Technology Beijing, Beijing, China. ⁴Electric Power Planning & Engineering Institute, Beijing, China. Correspondence and requests for materials should be addressed to B.S. (email: bnsu@eppei.com)

concentrations. Many studies have been devoted to finding the influencing factors for $PM_{2.5}$ concentration. Meteorological factors such as temperature²¹, wind speed²², and rainfall²³, and human factors such as industrial processes²⁴, energy production and consumption²⁵, and transportation²⁶ are the most important. Previous research has shown that many human activities severely impact air pollution such as crop straw burning²⁷, coal burning²⁸, and vehicle exhaust emissions²⁹. Many research has been carried out on $PM_{2.5}$ concentration using Chemistry Transport Models (CTMs). CTMs can simulate the distribution of $PM_{2.5}$ concentration in certain regions. Currently, however, this method still has some limitations. First, more work should be done on the mechanism of $PM_{2.5}$ formation and development³⁰. Second, data on pollution sources in a whole nation are often not accurate enough and will lead to significant simulation error³¹.

In this paper, $PM_{2.5}$ concentration is studied through a different approach. We tried to figure out the most important influencing factors of industrial production, energy production, agriculture, and transportation (IEAT) on $PM_{2.5}$ concentration based on available statistics data. Based on millions of collected $PM_{2.5}$ data, we determined the spatial and temporal characteristics of $PM_{2.5}$ distribution in mainland China, and analyzed trends in $PM_{2.5}$ for different provinces. We also collected multisource statistical data that included meteorological and IEAT-factors for each month from January 2013 to May 2017. We found associations between $PM_{2.5}$ and IEAT-factors, and developed a regression formula $PM_{2.5}$ concentration based on 10 primary factors. In addition, we calculated the meteorological and IEAT contributions to $PM_{2.5}$ levels in the different provinces. The results are helpful for governments providing macroscopic mitigation plans for controlling $PM_{2.5}$ in the future.

Data and Methods

Data source. We summarized the collected data into 49 influencing factors, as listed in Table S1. These factors can be grouped into 5 categories: meteorology, industrial production, energy production, agriculture, and transportation. Since 2013, $PM_{2.5}$ data were collected hourly by 391 monitoring stations managed by the Ministry of Environment Protection, China (MEP); $PM_{2.5}$ data from January 2014 through May 2017 were collected daily by 190 monitoring cities from Air Quality Inspection Platform of China³²; meteorological data between 2013 and 2016 were collected by 195 weather stations of the China Meteorological Administration (CMA)³³; meteorological data between January and May 2017 are from the National Oceanic and Atmospheric Administration (NOAA)³⁴; all data for industrial and energy production are from the National Bureau of Statistics of China (NBSC)³⁵; all transportation data are from the Ministry of Transport, China (MOT)³⁶; data for straw burning are from the Ministry of Environmental Protection, China (MEP)³⁷; data for geographic division of mainland China are from the Ministry of Civil Affairs, China (MCA)³⁸.

Data processing. All data have different temporal collection frequency. $PM_{2.5}$ and meteorological data were collected hourly and daily, while other data for human factors were collected monthly. To unify these data sets, the monthly value for each factor is used in this study, therefore a monthly averaged value is calculated for $PM_{2.5}$ and meteorological data. $PM_{2.5}$ and meteorological data, which were obtained by stations or cities, are averaged by province to obtain a provincial value.

To obtain the regression formula, all data are from the 53 months between January 2013 and May 2017. There are 34 provinces in China. Since there is no available data for Hong Kong, Macau, and Taiwan, we only considered the 31 provinces of mainland China. Therefore, the total data volume for each index should be 1643 (31×53) if no data was missing. In reality, there are data missing as shown in Table S1. We used linear interpolation to obtain this missing data. Most of the meteorological factors, and some of the human factors vary periodically by year. We therefore give priority to linear interpolation by year rather than by month. For example, if the data (X, M, Y) for province X in month M , Year Y is missing, we determine the data by linear interpolation based on data ($X, M, Y - 1$) and data ($X, M, Y + 1$). If data ($X, M, Y - 1$) or data ($X, M, Y + 1$) is missing, we use data ($X, M - 1, Y$) and data ($X, M + 1, Y$) to interpolate the missing data. We do not fill data outside the periods for which they were measured, for example, if data was only recorded from May 2015 to May 2017, then data before May 2015 cannot be determined via interpolation. Finally, 3.1% of the data was determined using linear interpolation.

To remove the differences due to geometrical area variation among the 31 provinces and the number of days in a month, all accumulative data (industrial production, energy production, agriculture, and transportation) were divided by province area and number of days in the month.

In this study, we used stepwise regression^{39,40} to process data to obtain the primary influencing factors that contribute the most to $PM_{2.5}$ concentration. In statistics, stepwise regression is a method of fitting regression models in which the choice of predictive variables is made by an automatic procedure⁴¹. The variables ending up in the final equation signify the best combination of independent variables for predicting the dependent variable⁴². Stepwise regression is frequently used in the statistical analysis of air pollution and has the advantage of being able to avoid collinearity⁴³. SPSS software (IBM SPSS, version 20) was used for the statistical analysis⁴⁴.

All the 49 influencing factors (as listed in Table S1) are used as input variables. Using stepwise regression, 10 factors having the highest impact on $PM_{2.5}$ are found. In addition, the regression formula for $PM_{2.5}$ concentration is obtained.

Results

General analysis for $PM_{2.5}$ in China. Because of periodicity of meteorological impacts and human behavior, $PM_{2.5}$ variation is also periodic. Figure 1 shows the pollution level and mean concentration of $PM_{2.5}$ in China for 12 months based on data from the past four years. The month with the best air quality is August, with a mean concentration of $PM_{2.5}$ of $36.0 \mu\text{g}/\text{m}^3$. 81.1% of the days had good air quality, and 17.5% of the days had moderate air quality. January has the worst air quality with an average $94.3 \mu\text{g}/\text{m}^3$ of $PM_{2.5}$. In January, only 27.6% of the days have good air quality, and 37.1% days have light, moderate, heavy, or severe pollution. Generally speaking,

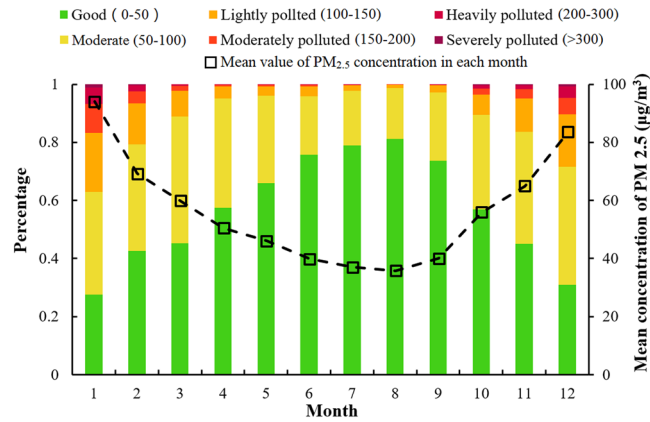


Figure 1. Pollution level analysis for 12 months (average value by 31 provinces) (data from January 1st, 2013 to December 31th, 2016; unit: µg/m³).

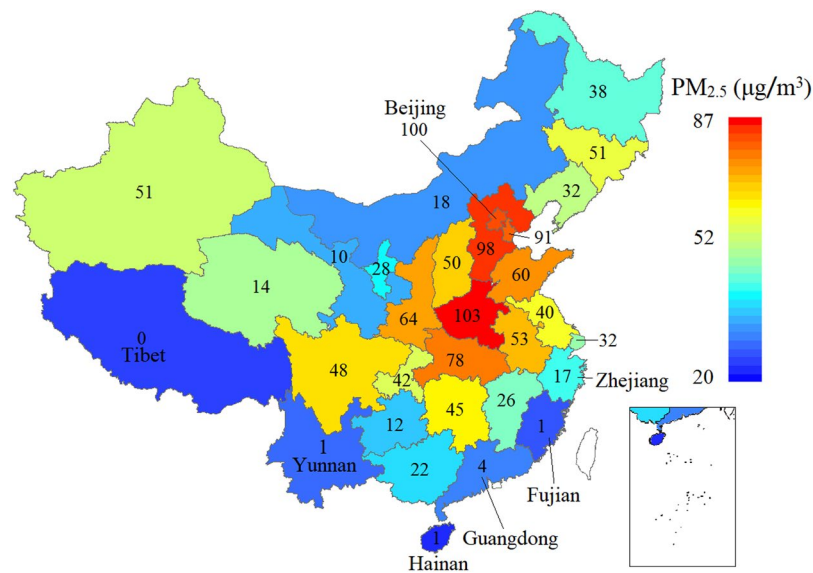


Figure 2. PM_{2.5} concentration distribution and number of yearly polluted days in 31 provinces of China (data from January 1st 2013 to December 31th 2016). (The picture was generated by ArcGIS 10.0, <https://www.esri.com/en-us/home>).

concentrations of PM_{2.5} are much higher in winter than in summer. Therefore, better control of PM_{2.5} in winter is critical.

Figure 1 shows the temporal distribution of PM_{2.5}, while Fig. 2 shows its spatial distribution. The number in Fig. 2 shows the average number of polluted days (PM_{2.5} concentration > 100 µg/m³) per year. From Fig. 2, Henan, Hebei, Beijing, Tianjin, and Hubei ranked in the top five for average concentration of PM_{2.5}. Beijing experiences about 100 polluted days on average every year. The average PM_{2.5} concentration in Beijing was 79.1 µg/m³ for the past 4 years. The five provinces with best air quality are Hainan, Tibet, Fujian, Yunnan, and Guangdong. Hainan has on average 20.0 µg/m³ of PM_{2.5} while there was not one polluted day in Tibet from 2013 to 2016. Generally, the highly polluted areas are North, Central and some parts of East China. The Coastal province and Northwest China have better air quality.

Combining the results of spatial and temporal distribution of PM_{2.5}, Fig. 3 shows the concentration of PM_{2.5} in five typical provinces from April 2013 to March 2017. Beijing (black dashed line) experiences serious air pollution during these years. The PM_{2.5} concentration in Beijing decreased between 2013 and 2015, but increased slightly again to 76.7 µg/m³ in 2016. Shanghai, with the highest income per capita in China, experienced a gradual decrease in PM_{2.5} concentration from 60.8 to 41.4 µg/m³ between April 2013 and March 2017. Henan, the province with the highest population, experienced the most serious air pollution in China. The PM_{2.5} concentration in Henan gradually decreased from 107.0 to 79.5 µg/m³ in the same period. Because of the impact of heavy industry, the PM_{2.5} concentration in Tianjin increased by 11.0% year-on-year. Ningxia, a representative city with low economic development, saw only a slight increase in PM_{2.5} concentration for the period covered by the data.

Rank	Factor	Coefficient	Standard error	Unit	Sig.	Accumu-lative R ²
1	Temperature	-1.57×10^0	7×10^{-2}	$\mu\text{g}\cdot\text{m}^{-3}\cdot^\circ\text{C}^{-1}$	0.000	0.358
2	Air pressure	8.3×10^{-2}	6×10^{-3}	$\mu\text{g}\cdot\text{m}^{-3}\cdot\text{hPa}^{-1}$	0.000	0.518
3	Wind speed	-1.2×10^1	1×10^0	$\mu\text{g}\cdot\text{m}^{-3}\cdot\text{m}^{-1}\cdot\text{s}$	0.000	0.577
4	Natural gas production	4.3×10^2	5×10^1	$10^{-4} \mu\text{g}\cdot\text{m}^{-3}\cdot\text{m}^{-3}\cdot\text{km}^2\cdot\text{d}$	0.000	0.634
5	Industrial boiler production	6.0×10^{-1}	9×10^{-2}	$\mu\text{g}\cdot\text{m}^{-3}\cdot\text{t}(\text{vapor})^{-1}\cdot\text{km}^2\cdot\text{d}$	0.000	0.673
6	Ore production	3.0×10^0	4×10^{-1}	$\mu\text{g}\cdot\text{m}^{-3}\cdot\text{t}^{-1}\cdot\text{km}^2\cdot\text{d}$	0.000	0.694
7	Tractor production	1.1×10^3	2×10^2	$10^4 \mu\text{g}\cdot\text{m}^{-3}\cdot\text{km}^2\cdot\text{d}$	0.000	0.707
8	Nuclear power generation	-7×10^1	2×10^1	$10^{-4} \mu\text{g}\cdot\text{m}^{-3}\cdot\text{kW}\cdot\text{h}\cdot\text{km}^2\cdot\text{d}$	0.000	0.713
9	Locomotive production	2.4×10^0	8×10^{-1}	$10^4 \mu\text{g}\cdot\text{m}^{-3}\cdot\text{km}^2\cdot\text{d}$	0.003	0.717
10	24-hour rainfall	-6×10^{-1}	2×10^{-1}	$\mu\text{g}\cdot\text{m}^{-3}\cdot\text{mm}^{-1}$	0.004	0.721

Table 1. Top 10 influencing factors highly associated with PM_{2.5} concentration. Constant = $1.37 \times 10^1 \mu\text{g}\cdot\text{m}^{-3}$.

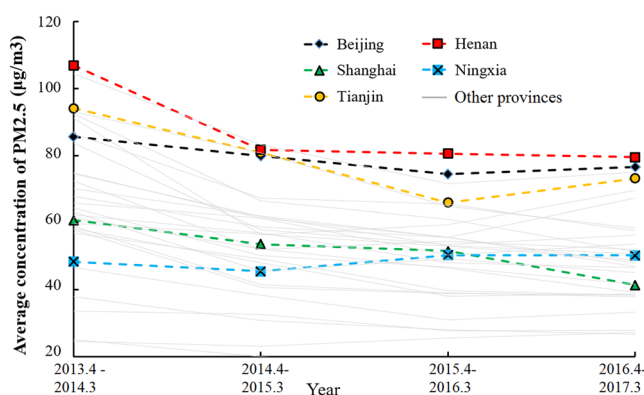


Figure 3. Concentration trends of PM_{2.5} in Beijing, Shanghai, Tianjin, Henan, and Ningxia provinces (data from April 1st 2013 to March 31th 2017).

Association between PM_{2.5} concentration and IEAT-factors. The ten most important factors influencing PM_{2.5} concentration, determined by stepwise regression, are listed in Table 1. Meteorological factors including temperature, air pressure, and wind speed are the most strongly associated with PM_{2.5} concentration. The top six IEAT factors having a strong association with PM_{2.5} concentration include the production of: natural gas; industrial boilers; ore; tractors; nuclear power, and locomotives.

Shaanxi, Tianjin, and Beijing have the highest production of natural gas per unit area. Henan, Shanghai, and Jiangsu are the three biggest industrial boiler producers per unit area. Hebei, Beijing, and Liaoning have the highest production of ore per unit area. Shandong, Henan, and Chongqing are the top three provinces in tractor production per unit area. Zhejiang, Tianjin, and Beijing rank highest in the production of locomotives per unit area. Nuclear power generation is the only positive human factor among the top ten factors. Using clean energy can efficiently reduce PM_{2.5} air pollution. Just 7 out of 31 provinces have nuclear power generation. The average PM_{2.5} concentrations per year in Zhejiang, Guangdong, and Fujian provinces (which have high nuclear power generation) are 49.5, 37.2, and 30.4 $\mu\text{g}/\text{m}^3$, respectively. Constant in this formula is 13.7 which means the background value of PM_{2.5} concentration is 13.7 $\mu\text{g}/\text{m}^3$ in China. In addition, R² of the linear fitting by these 10 factors is 0.721.

According to values in Table 1, the regression formula for PM_{2.5} concentration can be written as (Eq. 1):

$$C_{PM_{2.5}} = -1.57T + 8.3 \times 10^{-2}P_A - 12S_W - 0.6R_{24} + 4.3 \times 10^2P_{NG} + 0.60P_{IB} + 3.0P_O + 1.1 \times 10^3P_T - 7 \times 10P_{NP} + 2.4P_{RL} + 13.7 \quad (1)$$

where $C_{PM_{2.5}}$ is the PM_{2.5} concentration ($\mu\text{g}/\text{m}^3$), T average temperature ($^\circ\text{C}$), P_A average air pressure (hPa), S_W average wind speed ($\text{m}\cdot\text{s}^{-1}$), R_{24} average rainfall for 24 hours (mm); P_{NG} , P_{IB} , P_O , P_T , P_{NP} , and P_{RL} are production rates (production per square kilometer per day) of natural gas ($10^4 \text{m}^3\cdot\text{km}^{-2}\cdot\text{d}^{-1}$), industrial boilers ($\text{t}(\text{vapor})\cdot\text{km}^{-2}\cdot\text{d}^{-1}$), ore ($\text{t}\cdot\text{km}^{-2}\cdot\text{d}^{-1}$), tractors ($\text{km}^{-2}\cdot\text{d}^{-1}$), nuclear power ($10^4 \text{kW}\cdot\text{h}\cdot\text{km}^{-2}\cdot\text{d}^{-1}$), and locomotives ($10^{-4} \text{km}^{-2}\cdot\text{d}^{-1}$) per square kilometer per day, respectively.

Meteorological and IEAT contributions to PM_{2.5} concentration. In this study, we divided the influencing factors into meteorological and IEAT elements. Meteorological and IEAT contributions to PM_{2.5} concentration are calculated using Eq. 1 with meteorological and human factors, respectively.

Figure 4(a,b) show the meteorological and IEAT contributions to PM_{2.5}. In Fig. 4(a) we see that the three northeastern provinces of China and North China (except Inner Mongolia and Tianjin) have a high meteorological contribution to PM_{2.5}, while the south of China has a lower meteorological contribution to PM_{2.5}.

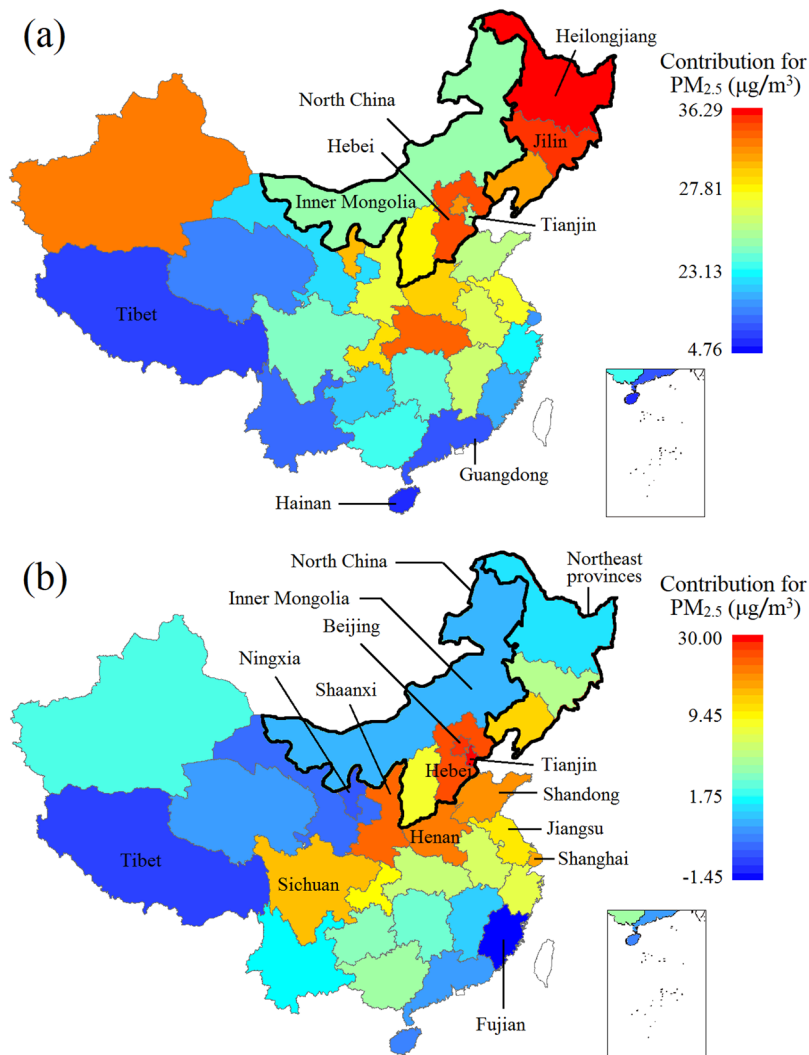


Figure 4. Contribution map for PM_{2.5}. (a) Meteorological contribution. (b) IEAT contribution. (The picture was generated by ArcGIS 10.0, <https://www.esri.com/en-us/home>).

Heilongjiang (38.0 µg/m³), Jilin (36.3 µg/m³), and Hebei (33.5 µg/m³) provinces have the highest meteorological contribution, while Hainan (4.8 µg/m³), Tibet (14.0 µg/m³), and Guangdong (14.0) provinces have the lowest meteorological contribution. In Fig. 4(b) we see that the three northeastern provinces of China are not bad in regard to IEAT contribution to PM_{2.5}. North China (except Inner Mongolia), Sichuan, Shaanxi, Henan, and Shandong provinces are high IEAT contributors to PM_{2.5}. Some southern coastal provinces such as Shanghai and Jiangsu also have high IEAT contributions. Tianjin (30.5 µg/m³), Beijing (27.5 µg/m³), and Hebei province (26.7 µg/m³) have the highest IEAT contribution to PM_{2.5} while Fujian (−1.5 µg/m³), Tibet (0.0 µg/m³), and Ningxia (0.2 µg/m³) rank lowest. More detailed information can be found in Table S2.

Verification. Another group of data from June 2017 to December 2017 was used to verify the robustness of regression formula obtained above (Eq. 1). In Fig. 5, a total of 217 points show the monthly averaged PM_{2.5} data for 7 months of 31 provinces. From comparison between the measured data and calculated values by regression formula based on ten influencing factors from June 2017 to December 2017 (Fig. 5), R² is 0.62, which is slightly lower than regression formula's accumulative R² (0.72). The average errors between the measured data and the calculated value is 11.1 µg/m³.

Discussion

In this study we investigated 49 influencing factors including meteorology, industrial production, energy production, agriculture, and transportation on PM_{2.5} concentrations. We determined a quantitative association between PM_{2.5} concentration and IEAT factors, and obtained a formula for PM_{2.5} concentration considering 10 primary factors based on stepwise regression. Stepwise regression is used because it is suitable for processing collinear data. We have tried to use common linear regression to process data, and found that the results are not as good as results obtained by stepwise regression (see *Supplementary Information* for details).

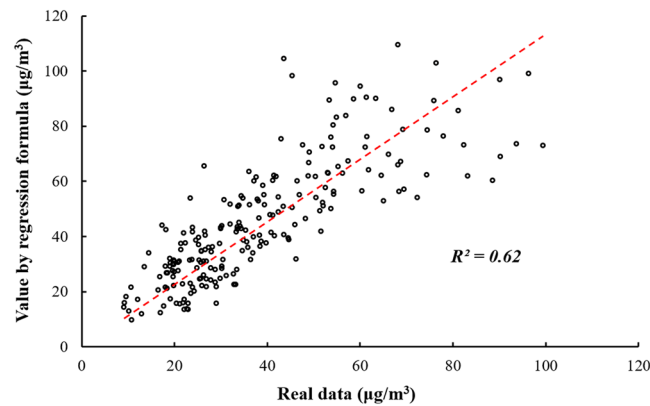


Figure 5. Comparison between the measured data of PM_{2.5} concentration (from June 2017 to December, 2017) and the values obtained by regression formula (Eq. 1).

In this study, we analyzed PM_{2.5} concentrations in China from January 2013 to May 2017. The average PM_{2.5} concentration over 12 months shows an upward parabola. Severe pollution (>300 µg/m³) appeared in winter in many areas of China. Using clean energy such as nuclear power to replace coal burning power plants is a very efficient way to reduce the number of severely polluted days in China. North and Central China have serious PM_{2.5} pollution problems. Research has shown that a 10 µg/m³ increase over a previous day's PM_{2.5} level results in a 1.78% increase in respiratory related mortality and a 1.03% increase in stroke related mortality⁴⁵. In China, residents in high-PM_{2.5} concentration areas look forward to gale force winds to reduce pollution. However, reducing PM_{2.5} generated by human activity is the key solution. One approach would be a reasonable distribution of PM_{2.5} sources to help balance PM_{2.5} concentration between highly-populated areas and rural areas. In addition, efforts should be made to reduce the number of severely polluted days rather than just reducing the average PM_{2.5} concentration.

Meteorological contributions to PM_{2.5} are high in the three northeastern provinces of China and North China (excluding Inner Mongolia and Tianjin). These areas are cold in winter, resulting in more coal consumption. Less rainfall in these inland areas is another meteorological reason for high PM_{2.5} levels. North China (excluding Inner Mongolia), Central China, and some provinces of East China have high IEAT contributions to PM_{2.5}. Beijing is a typical polluted city which had an average PM_{2.5} concentration of 76.7 µg/m³ from April 2016 to March 2017. In the 13th Five-Year Plan, The Ministry of Environment Protection, China (MEP) set a target for the concentration of PM_{2.5} in Beijing to be reduced to 56 µg/m³ by 2020. This is a considerable challenge. Shanghai provides a good example where in the past four years, the PM_{2.5} concentration has gradually decreased even though urban construction continues and the economy keeps improving.

According to stepwise regression, we found that of the 49 influencing factors (meteorology, industrial production, energy production, agriculture, and transportation), the production of natural gas, industrial boilers, ore, tractors, nuclear power, and locomotives are the top six IEAT factors contributing to PM_{2.5} concentrations. Since the production of natural gas, energy, ore, locomotives and tractors do not need to be concentrated in one area, future planning should include spreading these industries over wider areas to avoid creating areas with high population densities and heavy pollution. Finally, the regression formula was verified by the data from another seven months from June 2017 to December 2017.

Some researchers are focusing on the relationship between human factors and PM_{2.5} pollution. Previous research has suggested that clean fuels such as natural gas should replace the coal used for small domestic boilers to reduce air pollution⁴⁶ because combustion for natural gas is cleaner. However, some recent studies noticed that several air pollutants, including VOCs, NO_x, PM_{2.5} and SO₂, will be emitted during production stage of the natural gas^{47–49}. Our results also show the production of natural gas is strongly related to PM_{2.5} concentrations. Using clean energy (such as solar power, wind power, and nuclear power) to replace fossil energy, rather than using natural gas to replace coal, would be a better solution for PM_{2.5} control. Industrial boilers, which are usually used for burning coal⁵⁰, are bad for air quality⁵¹. We also found that provinces with high production of industrial boilers, usually have high PM_{2.5} concentrations. Industrial boilers are usually heavy and not convenient for long-range transportation, so most of the produced boilers are likely to be installed and used locally, leading to this phenomenon. Through analysis of collected 24-h PM_{2.5} samples in Brazil, researchers found that environmental contamination is led by ore mining and related activities such as the transport of products to and from the mines⁵². In our study, we found that production of ironstone and phosphate ore were the third biggest generators of PM_{2.5} ($P < 0.001$), followed by the production of natural gas and industrial boilers. Turkish researchers found that the PM_{2.5} concentration around tractors can reach thousands µg/m³⁵³. We found that the production of tractors is also strongly related to PM_{2.5} concentration ($P < 0.001$). New tractors are usually transported to nearby regions, and tractor operation will also generate PM_{2.5}. In New York, the levels of PM_{2.5} rapidly reached a peak when a diesel-powered locomotive passed⁵⁴. In China, locomotives are widely used. We found that producing locomotives will increase PM_{2.5} concentration. High production of locomotives generally reflects a high level of heavy industry, which lead to serious air pollution. As for meteorological factors, many studies have shown that the PM_{2.5} concentration is negatively correlated with wind speed and rainfall^{55–58} which agrees with our findings.

Because many human factors such as coal burning are strongly related to meteorological factors, the impacts of meteorological factors on $PM_{2.5}$ concentration are difficult to quantitatively measure.

There are some limitations to this study. Because of data limitation, we considered only some influencing factors in meteorology, industrial production, energy production, agriculture, and transportation. Other meteorological and IEAT factors may also influence $PM_{2.5}$ levels. For example, energy consumption is very important on emissions and air pollution. However, accurate monthly fuel consumption data at provincial level is unavailable, mainly because there are numerous distributed consumers. In the future, we should take these factors into consideration, if advanced statistical method is developed and accurate energy consumption data are available. In addition, some unconsidered human factors are strongly related to meteorological factors. For example, the consumption of coal and fireworks is high when the temperature is low (Spring Festival which consumes many fireworks is around January and February)⁵⁹. Therefore, the meteorological contribution in this study is mainly comprised of both unconsidered human and meteorological factors. In addition, all data is from January 2013 to May 2017. The precision could be improved if we could collect more data over longer periods. We used the province as the spatial unit in this study. City-scale data should be analyzed in the future once more precise data is obtained. In our data processing, some data was obtained using yearly and monthly linear interpolation, and this may slightly influence the precision of the regression formula for $PM_{2.5}$ concentration. Moreover, since all factors and results are related to industrial technologies, personal living habits, and a few other characteristics, the association should be adjusted when being used for analysis of $PM_{2.5}$ concentrations in other countries.

Conclusions

$PM_{2.5}$ in China in spatial and temporal dimensions was analyzed for data from January 2013 to May 2017. We quantitatively obtained the impacts of meteorology, industrial production, energy production, agriculture, and transportation on $PM_{2.5}$ concentration over an extended period. We found that production of natural gas, industrial boilers, ore production, tractors, and locomotives were the five human factors with the strongest association with $PM_{2.5}$ concentration. The model and the results provide efficient references for governments to make better plans on controlling $PM_{2.5}$ concentrations.

In the future, more types of data, longer time periods, and more detailed regionalization should be considered to improve the precision of the association analysis for $PM_{2.5}$ concentration.

References

- Kulshrestha, A., Satsangi, P. G., Masih, J. & Taneja, A. Metal concentration of $PM_{2.5}$ and PM_{10} particles and seasonal variations in urban and rural environment of Agra, India. *Science of the Total Environment* **407**, 6196–6204 (2009).
- Goto, D. *et al.* Estimation of excess mortality due to long-term exposure to $PM_{2.5}$ in Japan using a high-resolution model for present and future scenarios. *Atmospheric Environment* **140**, 320–332 (2016).
- Fujii, Y., Mahmud, M., Tohno, S., Okuda, T. & Mizohata, A. A case study of $PM_{2.5}$ characterization in Bangi, Selangor, Malaysia during the Southwest monsoon season. *Air Quality, Atmosphere & Health* **16**, 2685–2691 (2016).
- Han, L., Zhou, W. & Li, W. Increasing impact of urban fine particles ($PM_{2.5}$) on areas surrounding Chinese cities. *Scientific Reports* **5**, 12467 (2015).
- Ma, Z. *et al.* Satellite-based spatiotemporal trends in $PM_{2.5}$ concentrations: China, 2004–2013. *Environmental Health Perspectives* **124**, 184–192 (2016).
- Xu, P., Chen, Y. & Ye, X. Haze, air pollution, and health in China. *Lancet* **382**, 2067 (2013).
- Cao, C. *et al.* Inhalable microorganisms in Beijing's $PM_{2.5}$ and PM_{10} pollutants during a severe smog event. *Environmental Science & Technology* **48**, 1499–1507 (2014).
- Zhang, Y. L. & Cao, F. Fine particulate matter ($PM_{2.5}$) in China at a city level. *Scientific Reports* **5**, 14884 (2015).
- Kioumourtzoglou, M. A. *et al.* Long-term $PM_{2.5}$ exposure and neurological hospital admissions in the northeastern United States. *Environmental Health Perspectives* **124**, 23–29 (2016).
- Zhang, N., Huang, H., Duarte, M. & Zhang, J. Dynamic population flow based risk analysis of infectious disease propagation in a metropolis. *Environment International* **94**, 369–379 (2016).
- Leiva, G. M. A., Santibañez, D. A., Ibarra, E. S., Matus, C. P. & Seguel, R. A five-year study of particulate matter ($PM_{2.5}$) and cerebrovascular diseases. *Environmental Pollution* **181**, 1–6 (2013).
- Apte, J. S., Marshall, J. D., Cohen, A. J. & Brauer, M. Addressing global mortality from ambient $PM_{2.5}$. *Environmental Science & Technology* **49**, 8057–8066 (2015).
- Huang, R. J. *et al.* High secondary aerosol contribution to particulate pollution during haze events in China. *Nature* **514**, 218–222 (2014).
- Xinhua. President Xi: Carry out pollution control with great efforts. Available at: http://www.qstheory.cn/yaowen/2014-11/13/c_1113239067.htm, (Accessed 2017 July 1).
- CEN (China Environmental News). Strengthen the environmental impact assessment during 13th Five-Year. Available at: http://www.qstheory.cn/zoology/2017-02/21/c_1120501829.htm (Accessed 2017 July 1).
- Ma, Z., Hu, X., Huang, L., Bi, J. & Liu, Y. Estimating ground-level $PM_{2.5}$ in China using satellite remote sensing. *Environmental Science & Technology* **48**, 7346–7444 (2014).
- Seltenrich, N. A clearer picture of China's air: using satellite data and ground monitoring to estimate $PM_{2.5}$ over Time. *Environmental Health Perspectives* **124**, A38 (2016).
- Zhang, H. *et al.* Evaluation of a seven-year air quality simulation using the Weather Research and Forecasting (WRF)/Community Multiscale Air Quality (CMAQ) models in the eastern United States. *Science of the Total Environment* **473**, 275–285 (2014).
- Lee, H. M. *et al.* $PM_{2.5}$ source attribution for Seoul in May from 2009 to 2013 using GEOS-Chem and its adjoint model. *Environmental Pollution* **221**, 377–384 (2017).
- Chemel, C. *et al.* Application of chemical transport model CMAQ to policy decisions regarding $PM_{2.5}$ in the UK. *Atmospheric Environment* **82**, 410–417 (2014).
- Zhang, R. *et al.* Nanofiber air filters with high-temperature stability for efficient $PM_{2.5}$ removal from the pollution sources. *Nano Letters* **16**, 3642–3649 (2016).
- Yu, L. *et al.* Characterization and source apportionment of $PM_{2.5}$ in an urban environment in Beijing. *Aerosol and air quality research* **13**, 574–583 (2013).
- Ozbek, N., Baltaci, H. & Baysal, A. Investigation of fluorine content in $PM_{2.5}$ airborne particles of Istanbul, Turkey. *Environmental Science and Pollution Research* **23**, 13169–13177 (2016).

24. Huang, Y. *et al.* Quantification of global primary emissions of PM_{2.5}, PM₁₀, and TSP from combustion and industrial process sources. *Environmental Science & Technology* **48**, 13834–13843 (2014).
25. Xu, Y. *et al.* Field measurements on the emission and removal of PM_{2.5} from Coal-Fired power stations: 4. PM removal performance of wet electrostatic precipitators. *Energy Fuels* **230**, 7465–7473 (2016).
26. Ji, S. *et al.* Environmental justice aspects of exposure to PM_{2.5} emissions from electric vehicle use in China. *Environmental Science & Technology* **49**, 13912–13920 (2015).
27. Xu, H. *et al.* Inter-annual variability of wintertime PM_{2.5} chemical composition in Xi'an, China: Evidences of changing source emissions. *Science of the Total Environment* **545**, 546–555 (2016).
28. Yang, L. *et al.* Source identification and health impact of PM_{2.5} in a heavily polluted urban atmosphere in China. *Atmospheric Environment* **75**, 265–269 (2013).
29. Zhang, L., Liu, Y. & Hao, L. Contributions of open crop straw burning emissions to PM_{2.5} concentrations in China. *Environmental Research Letters* **11**, 014014 (2016).
30. Hu, M., Guo, S., Peng, J. & Wu, Z. Insight into characteristics and sources of PM_{2.5} in the Beijing-Tianjin-Hebei region, China. *National Science Review* **2**, 257–258 (2015).
31. Murray, N., Chang, H. H., Holmes, H. & Liu, Y. Combining Satellite Imagery and Numerical Model Simulation to Estimate Ambient Air Pollution: An Ensemble Averaging Approach. arXiv preprint arXiv:1802.03077 (2018).
32. AQIPC (Air Quality Inspection Platform of China). Available at: <https://www.aqistudy.cn/historydata/> (Accessed 2017 July 1).
33. CMA (China Meteorological Administration). Available at: <http://data.cma.cn/> (Accessed 2017 July 1).
34. NOAA (National Oceanic and Atmospheric Administration). Climate Data Online: Dataset Discovery. Available at: <https://www.ncdc.noaa.gov/cdo-web/datasets> (Accessed 2017 July 1).
35. NBSC (National Bureau of Statistics of China). Available at: [data.stats.gov.cn.](http://data.stats.gov.cn/) (Accessed 2017 July 1).
36. MOT (Ministry of Transport of the People's Republic of China). Available at: <http://zizhan.mot.gov.cn/zfxgk/> (Accessed 2017 July 1).
37. MEP (Ministry of Environment Protection of the People's Republic of China). Monitoring of straw burning. Available at: <http://hj.mep.gov.cn/jgjs/> (Accessed 2017 July 1).
38. MCA (Ministry of Civil Affairs of the People's Republic of China). National administrative division information inquiry platform. Available at: <http://xzqh.mca.gov.cn/map> (Accessed 2017 July 1).
39. Lv, J., Liu, Y., Zhang, Z. & Dai, J. Factorial kriging and stepwise regression approach to identify environmental factors influencing spatial multi-scale variability of heavy metals in soils. *Journal of Hazardous Materials* **261**, 387–397 (2013).
40. Lee, H. J., Liu, Y., Schuartz, J. & Koutrakis, P. A novel calibration approach of MODIS AOD data to predict PM_{2.5} concentrations. *Atmospheric Chemistry and Physics* **11**, 7991–8002 (2011).
41. Hocking, R. R. A Biometrics invited paper. The analysis and selection of variables in linear regression. *Biometrics* **32**, 1–49 (1976).
42. Yu, T., Yu, G., Li, P. Y. & Wang, L. Citation impact prediction for scientific papers using stepwise regression analysis. *Scientometrics* **101**, 1233–1252 (2014).
43. Chen, Y., Shi, R., Shu, S. & Gao, W. Ensemble and enhanced PM₁₀ concentration forecast model based on stepwise regression and wavelet analysis. *Atmospheric Environment* **74**, 346–359 (2013).
44. Voutsas, D., Samara, C., Manoli, E., Lazarou, D. & Tzoumaka, P. Ionic composition of PM_{2.5} at urban sites of northern Greece: secondary inorganic aerosol formation. *Environmental Science and Pollution Research* **21**, 4995–5006 (2014).
45. Franklin, M., Zeka, A. & Schwartz, J. Association between PM_{2.5} and all-cause and specific-cause mortality in 27 US communities. *Journal of Exposure Science and Environmental Epidemiology* **17**, 279–287 (2007).
46. Yao, X. *et al.* The water-soluble ionic composition of PM_{2.5} in Shanghai and Beijing, China. *Atmospheric Environment* **36**, 4223–4234 (2002).
47. Moore, C. W., Zielinska, B., Petron, G. & Jackson, R. B. Air impacts of increased natural gas acquisition, processing, and use: a critical review. *Environmental Science & Technology* **48**, 8349–8359 (2014).
48. Gilman, J. B., Lerner, B. M., Kuster, W. C. & Gouw, J. A. Source signature of volatile organic compounds from oil and natural gas operations in northeastern Colorado. *Environmental Science & Technology* **47**, 1297–1305 (2013).
49. Litovitz, A., Curtright, A., Abramzon, S., Burger, N. & Samaras, C. Estimation of regional air-quality damages from Marcellus Shale natural gas extraction in Pennsylvania. *Environmental Research Letters* **8**, 014017 (2013).
50. Dai, B. Q., Wu, X., De Girolamo, A. & Zhang, L. Inhibition of lignite ash slugging and fouling upon the use of a silica-based additive in an industrial pulverised coal-fired boiler. Part 1. *Changes on the properties of ash deposits along the furnace. Fuel* **139**, 720–732 (2015).
51. Li, Q. *et al.* Semi-coke briquettes: towards reducing emissions of primary PM_{2.5}, particulate carbon, and carbon monoxide from household coal combustion in China. *Scientific Reports* **6**, 19306 (2016).
52. de Fatima Andrade, M. *et al.* Vehicle emissions and PM_{2.5} mass concentrations in six Brazilian cities. *Air Quality, Atmosphere & Health* **5**, 79–88 (2012).
53. Arslan, S., Aybek, A. & Ekerbicer, H. C. Measurement of personal PM₁₀, PM_{2.5} and PM₁ exposures in tractor and combine operations and evaluation of health disturbances of operators. *Journal of Agricultural Sciences* **16**, 104–115 (2010).
54. Vilcassim, M. J. R., Thurston, G. D., Peltier, R. E. & Gordon, T. Black carbon and particulate matter (PM_{2.5}) concentrations in New York City's subway stations. *Environmental Science & Technology* **48**, 14738–14745 (2014).
55. Maraziotis, E., Sarotis, L., Marazioti, C. & Marazioti, P. Statistical analysis of inhalable (PM₁₀) and fine particles (PM_{2.5}) concentrations in urban region of Patras, Greece. *Global Nest Journal* **10**, 123–131 (2008).
56. Ni, X. Y., Huang, H. & Du, W. P. Relevance analysis and short-term prediction of PM_{2.5} concentrations in Beijing based on multi-source data. *Atmospheric Environment* **150**, 146–161 (2017).
57. Pillai, P. S., Babu, S. S. & Moorthy, K. K. A study of PM, PM₁₀ and PM_{2.5} concentration at a tropical coastal station. *Atmospheric Research* **61**, 149–167 (2002).
58. Stortini, A. M. *et al.* An evaluation of the PM_{2.5} trace elemental composition in the Venice Lagoon area and an analysis of the possible sources. *Atmospheric Environment* **43**, 6296–6304 (2009).
59. Kong, S. *et al.* Variation of polycyclic aromatic hydrocarbons in atmospheric PM_{2.5} during winter haze period around 2014 Chinese Spring Festival at Nanjing: Insights of source changes, air mass direction and firework particle injection. *Science of the Total Environment* **520**, 59–72 (2015).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 71473146, 71741023).

Author Contributions

B.S., N.Z., and H.H. designed the contents. N.Z. wrote the main manuscript and prepared all figures and tables. B.S., X.D., and N.Z. collected the data. N.Z. and B.S. analyzed the data. N.Z., B.S., and H.H. developed the model. B.S., H.H., and J.Z. supervised the theoretical analysis and writing. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-27771-w>.

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018