Estimation of daily PM$_{2.5}$ concentration and its relationship with meteorological conditions in Beijing

Qian Yin$^1$, Jinfeng Wang$^{1,2,*}$, Maogui Hu$^1$, Hoting Wong$^1$

1. LREIS, Institute of Geographic Sciences and Nature Resources Research, Chinese Academy of Sciences, Beijing 100101, China. E-mail: yinq@lreis.ac.cn
2. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

ARTICLE INFO

Article history:
Received 9 October 2015
Revised 5 March 2016
Accepted 7 March 2016
Available online 24 May 2016

Keywords:
PM$_{10}$ concentration
PM$_{2.5}$ concentration estimation
Wind speed
Wind direction

ABSTRACT

When investigating the impact of air pollution on health, particulate matter less than 2.5 μm in aerodynamic diameter (PM$_{2.5}$) is considered more harmful than particulates of other sizes. Therefore, studies of PM$_{2.5}$ have attracted more attention. Beijing, the capital of China, is notorious for its serious air pollution problem, an issue which has been of great concern to the residents, government, and related institutes for decades. However, in China, significantly less time has been devoted to observing PM$_{2.5}$ than for PM$_{10}$. Especially before 2013, the density of the PM$_{2.5}$ ground observation network was relatively low, and the distribution of observation stations was uneven. One solution is to estimate PM$_{2.5}$ concentrations from the existing data on PM$_{10}$. In the present study, by analyzing the relationship between the concentrations of PM$_{2.5}$ and PM$_{10}$, and the meteorological conditions for each season in Beijing from 2008 to 2014, a U-shaped relationship was found between the daily maximum wind speed and the daily PM concentration, including both PM$_{2.5}$ and PM$_{10}$. That is, the relationship between wind speed and PM concentration is not a simple positive or negative correlation in these wind directions; their relationship has a complex effect, with higher PM at low and high wind than for moderate winds. Additionally, in contrast to previous studies, we found that the PM$_{2.5}$/PM$_{10}$ ratio is proportional to the mean relative humidity (MRH). According to this relationship, for each season we established a multiple nonlinear regression (MNLR) model to estimate the PM$_{2.5}$ concentrations of the missing periods.

© 2016 The Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences. Published by Elsevier B.V.

Introduction

With people becoming increasingly keen on healthy living, the levels of atmospheric particulate matter (PM) have received considerable attention. Although the detailed effects of each type of PM on health have not yet been established, epidemiological studies (Ostro et al., 2010; Zhang et al., 2002; Cairncross et al., 2007) have identified a strong exposure–response relationship between PM and both short-term (coughing, sneezing, runny nose, shortness of breath and lung irritation) and long-term health effects (lung cancer, cardiovascular and cardiopulmonary diseases).

PM$_{10}$ is particulate matter of less than 10 μm in aerodynamic diameter. It can penetrate into the air passages of the lungs. Different from PM$_{10}$, PM$_{2.5}$ is less than 2.5 μm in aerodynamic diameter. It can reach the deeper alveolar spaces and is more harmful than particulate matter of other sizes (Hossain and Easa, 2012; Fan et al., 2013; Chun et al., 2014; Park and Kim, 2014).
Moreover, it has been reported that PM$_{2.5}$ can easily penetrate into other organs, and its removal takes longer. Thus, within the research on the health-related effects of air pollution, PM$_{2.5}$ has attracted widespread attention.

Air pollution has long been a serious problem in China, especially in eastern areas such as the Yangtze River Delta, southern areas such as the Pearl River Delta (Wang and Christopher, 2003) as well as in the Jing-Jin-Ji (Beijing, Tianjin and Hebei provinces) region (Xu et al., 2005; Ding et al., 2008; Zhang et al., 2013). Beijing, the capital of China, has a large population and high energy consumption, both of which have led to serious air pollution. The weather in Beijing is typical of a temperate monsoon climate: summers are hot and rainy, while winters are cold and dry. Particulates in Beijing are very complicated in terms of sources, pollution characteristics, and their effects on the environment and health. Among these particulates, PM$_{2.5}$ is one of the main air pollutants, and its concentration is extremely high. For example, January 2013 witnessed the most frequent and serious episodes of heavy haze to date, which caused consternation among the public (Che et al., 2014; Wang et al., 2014; Zhang et al., 2014). In 2014, Beijing residents suffered seven periods of heavy pollution that lasted more than 3 days, including a week-long period in February.

However, in China the time spent on PM$_{2.5}$ ground observation has been quite short, only about a few years, and the density of the PM$_{2.5}$ ground observation network is relatively low. Moreover, the distribution of observation stations is uneven. Particularly before 2013, information about the concentration of PM$_{2.5}$ in Beijing was not released. The limited number of observation stations meant that the PM$_{2.5}$ characteristics throughout Beijing could not be well established. One solution to this problem is to estimate the PM$_{2.5}$ concentrations using the existing PM$_{10}$ concentration data.

In this respect, the existing research has focused on estimating the concentration of PM$_{2.5}$ based on the aerosol optical depth (AOD), NO$_2$, SO$_2$, CO or O$_3$ concentration (Che et al., 2015; Huo et al., 2011; Zhang et al., 2015; Wang et al., 2013). Furthermore, since PM$_{2.5}$ is a part of PM$_{10}$ (Wang et al., 2006), many researchers of previous studies believed that the PM$_{2.5}$ concentration has a linear relationship with PM$_{10}$ concentration, therefore they conducted a linear regression analysis to estimate PM$_{2.5}$ concentration according to the PM$_{10}$ concentration. In the present study, we first analyzed the relationships between PM$_{2.5}$ and PM$_{10}$, as well as the meteorological conditions, including mean relative humidity (MRH), sunshine duration, wind speed, and wind direction. Accordingly, we found that the concentrations of PM$_{2.5}$ and PM$_{10}$ are nonlinear with respect to MRH. Secondly, since different seasons have different weather conditions, their influences on PM concentrations are also different. Therefore, according to the above nonlinear relationship, for each season we conducted a multiple nonlinear regression (MNLR) model. Since our PM$_{10}$ concentration data covered an entire seven-year period from 2008/01/01 to 2014/12/31, but the PM$_{2.5}$ data only covered about 4 years and 5 months (from 2010/01/01 to 2014/12/31, with the exception of 2012/05/16 to 2012/12/31), we used the model to estimate the PM$_{2.5}$ concentration data of the missing period. In addition, we analyzed the relationship between the wind speed from different directions and PM concentrations in detail.

1. Materials and methods

1.1. Study area

The study was conducted in Beijing. As shown in Fig. 1, the red area represents the urban region of Beijing and the black dot indicates the location of the observation point.

1.2. Materials

Two datasets were used in the research. The first dataset consisted of hourly observations of PM$_{2.5}$ concentrations at the United States (US) Embassy station from 2010/01/01 to 2014/12/31 (with the exception of 2012/05/16 to 2012/12/31), which were obtained from the website of the Embassy of the United States in China (http://chinese.usembassy-china.org.cn/070109air.html), and the daily mean concentrations of PM$_{10}$ at the same station from 2008/01/01 to 2014/12/31, which were obtained from the website of the Beijing Municipal Environmental Monitoring Center (http://www.bjmemc.com.cn/).

The second dataset consisted of daily meteorological data from 2008/01/01 to 2014/12/31 in Beijing, which was obtained from the website of the China Meteorological Data Sharing Service System (http://cdc.nmic.cn/home.do). The meteorological parameters are listed in Table 1.

1.3. Methods

In this research, some common statistical methods were employed, for example Spearman correlation analyses and an MNLR model which was utilized to fit the PM$_{2.5}$ concentration...
according to the PM\(_{10}\) concentration and meteorological variables shown in Table 1. The assumption of MNLR models is that the relationship between the dependent variable \(y\) and the regressor \(x_i\) is nonlinear. It was defined by the following (Eq. (1)):

\[
y = \alpha + \sum_{j=0}^{15} \beta_j x_j + \sum_{j>i} \beta_{ji} x_j x_i + \sum_{j=0}^{15} \gamma_j x_j^2
\]

where \(y\) is the daily mean concentration of PM\(_{2.5}\) and \(x_i\) is the intercept. \(x_i\) is the daily mean concentration of PM\(_{10}\). \(x_{j}, \ldots, x_{15}\) are the meteorological variables in Table 1, while \(\beta_0, \ldots, \beta_j\) are the regression coefficients. For each meteorological variable, the regression was done stepwise to add and delete terms based on Akaike Information Criterion (AIC) statistics to obtain the best model fit.

In addition, the mean absolute percentage error (MAPE) was used to measure the accuracy of the model. This is a statistical forecasting method to measure prediction accuracy. It is defined by the formula (Eq. (2)):

\[
\text{MAPE} = 100\% \times \frac{\sum_{i=1}^{n} |(A_i - F_i) / A_i|}{n}
\]

where \(A_i\) is the actual value and \(F_i\) is the estimated value. The absolute value in this calculation is summed for every forecasted point in time and divided by the number of fitted points \(n\). Multiplying by 100 gives the percentage error.

\(R\) provides comprehensive support for MNLR. All the statistical analyses were conducted using \(R\) statistical software (version 3.1.2) and the Openair package (version 1.4; Carslaw, 2015).

2. Results and discussion

2.1. Descriptive statistics

We collected and analyzed the concentration data of PM\(_{2.5}\) and PM\(_{10}\) at the observation point from January 1, 2008 to December 31, 2014 (with the exception of May 16, 2012 to December 31, 2012). The results showed that the daily mean concentration of PM\(_{10}\) ranged from 8 to 600 \(\mu\text{g/m}^3\), while the concentration of PM\(_{2.5}\) ranged from 3 to 480 \(\mu\text{g/m}^3\).

Fig. 2 shows the plots of the daily PM\(_{2.5}\) and PM\(_{10}\) concentrations at the US Embassy station from January 2010 to December 2010. Fig. 3 shows the daily MRH of Beijing from 2010 to 2012.

---

**Table 1 - Meteorological parameters considered in the statistical analysis.**

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Meteorological parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_1)</td>
<td>Daily mean air temperature (°C)</td>
</tr>
<tr>
<td>(x_2)</td>
<td>Daily maximum air temperature (°C)</td>
</tr>
<tr>
<td>(x_3)</td>
<td>Daily minimum air temperature (°C)</td>
</tr>
<tr>
<td>(x_4)</td>
<td>Mean relative humidity (%)</td>
</tr>
<tr>
<td>(x_5)</td>
<td>Minimum relative humidity (%)</td>
</tr>
<tr>
<td>(x_6)</td>
<td>Extreme max. wind speed (m/sec)</td>
</tr>
<tr>
<td>(x_7)</td>
<td>Direction of extreme max. wind speed (16 directions)</td>
</tr>
<tr>
<td>(x_8)</td>
<td>Maximum wind speed (m/sec)</td>
</tr>
<tr>
<td>(x_9)</td>
<td>Direction of maximum wind speed (16 directions)</td>
</tr>
<tr>
<td>(x_{10})</td>
<td>Mean wind speed (m/sec)</td>
</tr>
<tr>
<td>(x_{11})</td>
<td>Mean air pressure (hPa)</td>
</tr>
<tr>
<td>(x_{12})</td>
<td>Maximum air pressure (hPa)</td>
</tr>
<tr>
<td>(x_{13})</td>
<td>Minimum air pressure (hPa)</td>
</tr>
<tr>
<td>(x_{14})</td>
<td>Daily total precipitation (from 20:00 to 19:59) (mm)</td>
</tr>
<tr>
<td>(x_{15})</td>
<td>Sunshine duration (hr)</td>
</tr>
</tbody>
</table>

---

Fig. 2 – Plots of daily mean PM\(_{2.5}\) and PM\(_{10}\) concentrations at the US Embassy station in 2010. PM\(_{2.5}\): particulate matter less than 2.5 \(\mu\text{m}\); PM\(_{10}\): particulate matter less than 10 \(\mu\text{m}\).

---

Fig. 3 – Plot of daily mean relative humidity (MRH) of Beijing from 2010 to 2012.
2010 to 2012. Fig. 4 shows the annual mean of the PM$_{2.5}$ concentration, PM$_{10}$ concentration, PM$_{2.5}$/PM$_{10}$ ratio, and the MRH from 2010 to 2014 (with the exception the period of 2012/05/16 to 2012/12/31).

2.2. Relationship between PM$_{2.5}$ and PM$_{10}$ and meteorological factors

To better understand the factors influencing PM$_{2.5}$, we investigated the relationships between the daily mean concentration of PM$_{2.5}$ and the meteorological factors for each season. Spearman correlations between them were calculated (Table 2). The seasons were defined as: spring (March–May), summer (June–August), autumn (September–November) and winter (December–February).

From Table 2, it can be seen that PM$_{2.5}$, as a part of PM$_{10}$, has a strong correlation with PM$_{10}$ in respect to their concentrations. In addition to PM$_{10}$, most of the meteorological factors also had correlations with PM$_{2.5}$ concentrations. Among these meteorological factors, relative humidity (RH, including MRH and minimum relative humidity) and sunshine duration had stronger correlations with PM$_{2.5}$ concentrations than the other meteorological factors. The correlation coefficients between MRH and PM$_{2.5}$ concentrations were 0.59, 0.32, 0.55, and 0.69 for spring, summer, autumn and winter respectively. Similarly, the correlation coefficients between sunshine duration and PM$_{2.5}$ concentration were $-0.51$, $-0.39$, $-0.51$, and $-0.52$ for the four seasons respectively. The correlations all had obvious statistical significance. Since MRH and sunshine duration were highly correlated in this data set (their correlation coefficients were 0.64, 0.63, 0.58, and 0.63 for the four seasons respectively), and their p-values were all less than 0.001, we only needed to consider one of them.

Fig. 5 shows the scatter plots of the daily mean concentrations of PM$_{2.5}$, the daily mean concentration of PM$_{10}$ and the daily MRH in four seasons.

The above significant positive and negative correlations both had certain practical implications. For MRH, the significant positive correlation showed that higher MRH was often associated with windless, cloudy, and insufficiently sunny days, which aggravated the accumulation and chemical reaction of pollutants (Kang et al., 2013; Csavina et al., 2014). In contrast, the significant negative correlation with sunshine duration showed that with the increase of sunshine hours, the concentration of PM$_{2.5}$ decreased. A day with more sunshine hours often indicated better weather conditions with little cloud cover or strong winds, which are favorable for the diffusion and elimination of air pollutants (Yang et al., 2009; Sanchez-Romero et al., 2014).

Beyond that, a poor correlation was found between PM$_{2.5}$ concentration and wind speed. This means that the relationship between wind speed and PM concentration may not be a simple positive or negative correlation in different seasons. The impacts of wind speed and wind direction on PM concentration will be discussed in more detail in the next section.

Similarly, a poor correlation was found between PM$_{2.5}$ concentration and air pressure in spring, autumn, and winter.

<table>
<thead>
<tr>
<th>PM$_{2.5}$ concentration</th>
<th>PM$<em>{2.5}$/PM$</em>{10}$ ratio</th>
<th>MRH</th>
<th>Sunshine duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>x0 (PM$_{10}$)</td>
<td>0.83***</td>
<td>0.87***</td>
<td>N/A</td>
</tr>
<tr>
<td>x1 (Mean temperature)</td>
<td>0.18**</td>
<td>0.14</td>
<td>-0.09</td>
</tr>
<tr>
<td>x2 (Maximum temperature)</td>
<td>0.16</td>
<td>0.14</td>
<td>-0.09</td>
</tr>
<tr>
<td>x3 (Minimum temperature)</td>
<td>0.25**</td>
<td>0.33***</td>
<td>-0.04</td>
</tr>
<tr>
<td>x4 (Mean relative humidity)</td>
<td>0.59***</td>
<td>0.32***</td>
<td>0.55***</td>
</tr>
<tr>
<td>x5 (Minimum relative humidity)</td>
<td>0.43***</td>
<td>0.43***</td>
<td>0.51***</td>
</tr>
<tr>
<td>x6 (Direction of maximum wind speed)</td>
<td>-0.24**</td>
<td>-0.09</td>
<td>-0.40***</td>
</tr>
<tr>
<td>x7 (Maximum wind speed)</td>
<td>-0.23**</td>
<td>-0.07</td>
<td>-0.38***</td>
</tr>
<tr>
<td>x8 (Direction of maximum wind speed)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>x9 (Mean wind speed)</td>
<td>-0.26**</td>
<td>0.12</td>
<td>-0.34***</td>
</tr>
<tr>
<td>x10 (Mean air pressure)</td>
<td>-0.35**</td>
<td>-0.10</td>
<td>-0.25***</td>
</tr>
<tr>
<td>x11 (Maximum air pressure)</td>
<td>-0.33</td>
<td>-0.12</td>
<td>-0.23**</td>
</tr>
<tr>
<td>x12 (Minimum air pressure)</td>
<td>-0.35**</td>
<td>-0.06</td>
<td>-0.21**</td>
</tr>
<tr>
<td>x13 (Daily total precipitation)</td>
<td>0.22</td>
<td>0.01</td>
<td>-0.15</td>
</tr>
<tr>
<td>x14 (Sunshine duration)</td>
<td>-0.51**</td>
<td>-0.39***</td>
<td>0.11</td>
</tr>
</tbody>
</table>

PM$_{2.5}$: particulate matter less than 2.5 μm; PM$_{10}$: particulate matter less than 10 μm. N/A: Not available.

*** p-Value < 0.001.
** p-Value < 0.01.
* p-Value < 0.05.
\(-0.47 < r < -0.25, \ p\text{-value} < 0.05\); meanwhile, no obvious correlation was found between PM\(_{2.5}\) concentration and air pressure in summer (\(p\text{-value} > 0.1\)).

In contrast, for PM\(_{2.5}\), the correlations with temperature and precipitation were not obvious, which indicated that the links between PM\(_{2.5}\) pollutants and temperature or precipitation were weak in each season.

### 2.3. Relationship between PM\(_{2.5}\) and wind direction

To better understand the influence of wind speed and wind direction on PM concentration, the prevailing wind directions were divided into 16 groups: N/NNE/NE/ENE/E/ESE/SE/SSE/S/SSW/SW/WSW/W/NW/NNW. Fig. 6 shows the polar plot (Carslaw, 2015) of mean PM\(_{2.5}\) concentrations for the 16 directions. In the plot, (0, 0) is the middle of the plot and the radial distance represents the mean wind speed of each direction, the unit is \(\text{m/sec}\). From Fig. 6, we generally found that the relationship between wind speed and PM\(_{2.5}\) concentration is not a simple negative correlation in each direction. That is, at greater wind speed, the PM\(_{2.5}\) concentration is not necessarily smaller.

In order to better understand the relationships between PM concentration and maximum wind speed, in Fig. 7, the horizontal axes represent the daily maximum wind speeds (unit: \(\text{m/sec}\)) and the vertical axes represent the mean concentrations of PM\(_{2.5}\) (red lines) and PM\(_{10}\) (blue lines).

From Fig. 7, for most of the wind directions, a U-shaped relationship was found between maximum wind speed and PM concentration. This means that the relationship between wind speed and PM concentration is not a simple positive or negative correlation in these wind directions, so that their relationship is similar to a “bathtub curve”. From these U-shaped figures, we found that when the wind speed is low (<4 \(\text{m/sec}\)), the PM concentration is obviously high. In...
addition, when wind speed increases (4–8 m/sec), the PM concentration decreases. This indicates that a low wind speed could limit the spread of particulates, but increases in wind speed could restrain the accumulation of particulates and, thus, reduce the PM$_{2.5}$ and PM$_{10}$ pollutant concentrations. However, as the wind speed continues to increase, the PM concentration unexpectedly becomes higher again. The reason for this phenomenon may be that a strong wind can bring in pollution from surrounding areas.

2.4. Multiple nonlinear regression

Since PM$_{2.5}$ is a part of PM$_{10}$, many researchers in previous studies believed that the PM$_{2.5}$ concentration has a linear relationship with the PM$_{10}$ concentration (Wang et al., 2006). Therefore, they conducted a linear regression analysis (Model I) to estimate PM$_{2.5}$ concentrations according to PM$_{10}$ concentrations.

However, from Fig. 4, we found that the PM$_{2.5}$/PM$_{10}$ ratio and the MRH have an obvious relationship; that is, the PM$_{2.5}$/PM$_{10}$ ratio is proportional to the MRH. Since different seasons have different MRH conditions, their influences on PM concentration also differ. According to the above relationship between them, for each season, we conducted an MNLR (Model II). Table 3 shows a comparison of the results of Models I and II:

In Table 3, $y$ represents the PM$_{2.5}$ concentration, $x_1$ represents the concentration of PM$_{10}$, and $x_4$ represents the MRH. By comparing the results of the two models, we can clearly see that in spring, the adjusted $R^2$ (0.80) of Model II was clearly higher than the adjusted $R^2$ (0.73) of Model I, and the error of Model II (13%) was remarkably smaller than the error of Model I (21.5%). In the other seasons, the results of Model II were also better than Model I.

In addition, we found that the errors were obviously larger in spring than in the other seasons. One reason for this was that in spring the impacts of climate conditions, such as dust storms, on PM$_{2.5}$ concentrations were greater than during the other seasons. If there was no dust storm, the PM$_{2.5}$/PM$_{10}$ ratio was larger and closer to the ratio of the other seasons;

Fig. 7 – In different wind directions, the mean PM$_{2.5}$ and PM$_{10}$ concentrations under different maximum wind speed levels.
however, if a dust storm occurred, then the PM$_{2.5}$/PM$_{10}$ ratio became significantly smaller. Therefore, if we only considered PM$_{10}$ concentrations and MRH as the explanatory variables when estimating the concentrations of PM$_{2.5}$, the errors would be larger.

2.5. Estimation of PM$_{2.5}$ concentrations

The results of the MNLR can be utilized to estimate the concentrations of PM$_{2.5}$. Fig. 8 shows the estimated PM$_{2.5}$ values for the missing period (May 17 to December 31) in 2012. In the figure, the dotted line represents the ground monitoring concentration of PM$_{10}$. PM$_{2.5,R}$ represents the real ground monitoring concentration of PM$_{2.5}$, while PM$_{2.5,E}$ represents the estimated PM$_{2.5}$ concentration value.

3. Conclusions

By analyzing the relationships between PM$_{2.5}$ concentration, PM$_{10}$ concentration, and the meteorological conditions for each season in Beijing from 2010 to 2014, a U-shaped relationship was found between maximum wind speed and PM concentration in most wind directions. That is, the relationship between wind speed and PM concentration is not a simple positive or negative correlation in these wind directions, with higher PM observed at low or high wind and lower PM for moderate wind speeds. In addition, in contrast to previous studies, we found that the PM$_{2.5}$/PM$_{10}$ ratio is proportionate to the MRH. According to this relationship between them, for each season, we established an MNLR model to estimate the PM$_{2.5}$ concentration of the missing period. The cross-seasonal validation suggests that our strategy for reconstructing the PM$_{2.5}$ time series using PM$_{10}$ data and meteorological factors in Beijing is satisfactory.

Acknowledgments

This work was supported by the National Science & Technology Major Program of China (No. 2012CB955503) and the National Natural Science Foundation of China (Nos. 41421001; 41301425; 41271404).

![Fig. 8 – The estimated PM$_{2.5}$ values for the missing period in 2012.](image)
REFERENCES


