PROCEEDINGS A

rspa.royalsocietypublishing.org

Research

Cite this article: Zhang S, Guo B, Dong A, He J, Xu Z, Chen SX. 2017 Cautionary tales on air-quality improvement in Beijing. *Proc. R. Soc. A* **473**: 20170457.
http://dx.doi.org/10.1098/rspa.2017.0457

Received: 29 June 2017
Accepted: 11 August 2017

Subject Areas:
statistics, atmospheric science

Keywords:
air-quality assessment, meteorological confounding, observational study, spatial–temporal adjustment

Authors for correspondence:
Bin Guo
e-mail: guobin@swufe.edu.cn
Song Xi Chen
e-mail: csx@gsm.pku.edu.cn

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3865483.

THE ROYAL SOCIETY PUBLISHING

Cautionary tales on air-quality improvement in Beijing

Shuyi Zhang[1], Bin Guo[4,5], Anlan Dong[2], Jing He[5], Ziping Xu[3], Song Xi Chen[1,2]

[1]Guanghua School of Management, [2]Center for Statistical Science, [3]Yuanpei College, Peking University, Beijing 100871, People's Republic of China
[4]Center of Statistical Research, [5]School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, People's Republic of China

SXC, 0000-0002-2338-0873

The official air-quality statistic reported that Beijing had a 9.9% decline in the annual concentration of $PM_{2.5}$ in 2016. While this statistic offered some relief for the inhabitants of the capital, we present several analyses based on Beijing's $PM_{2.5}$ data of the past 4 years at 36 monitoring sites along with meteorological data of the past 7 years. The analyses reveal the air pollution situation in 2016 was not as rosy as the 9.9% decline would convey, and improvement if any was rather uncertain. The paper also provides an assessment on the city's $PM_{2.5}$ situation in the past 4 years.

1. Introduction

A substantial part of China is experiencing chronic air pollution with severe fine particulate matter (PM) concentration and $PM_{2.5}$ in particular [1,2]. $PM_{2.5}$ refers to fine PM with aerodynamic diameter of less than 2.5 µm. The north China Plain (NCP) that surrounds Beijing endures the most severe air pollution in the country with excessive $PM_{2.5}$ concentration. In a move to clear up the smog, China's State Council has set a 25% $PM_{2.5}$ reduction target for the NCP by 2017 relative to the 2012 level, and a specific target of no more than 60 $\mu g\,m^{-3}$ for Beijing's annual average.

Beijing's official air-quality statistics are composed using hourly $PM_{2.5}$ data from 11 Guokong (state controlled) monitoring sites by calculating the simple averages of hourly readings. Guokong sites are required to transmit their data in real time to the Ministry of Environment and Protection (MEP) to avoid potential

© 2017 The Author(s) Published by the Royal Society. All rights reserved.

data interference. The consistency of $PM_{2.5}$ data between US diplomatic missions in five Chinese cities and the neighbouring Guokong sites was confirmed in [3].

Just a few days into the new year of 2017, the Beijing Municipal Environmental Monitoring Center (BMEMC) released the official annual $PM_{2.5}$ average of $72.6\,\mu g\,m^{-3}$ for 2016, which represents a 9.9% reduction from the previous year. While this statistic was quite comforting, it was in a sharp contrast to many locals' experience, especially as in the last three months of the year the capital had experienced repeated episodes of severe smog.

In this study, we conduct several statistical analyses on the $PM_{2.5}$ data of the past 4 years from Beijing's 36 monitoring sites in conjunction with 7 years' meteorological records at 15 stations; see the electronic supplementary material, table S1, for specific information of these sites. We want to provide meaning and insight to the official statistics and a broader understanding of the air pollution situation in Beijing. To this end, we consider (i) two types of years: the calendar year and the seasonal year, which runs from March to February the following year; (ii) two types of monitoring sites: the 11 Guokong sites and more sites in central Beijing to provide wider spatial coverage, and (iii) two types of averages: the simple average and an adjusted average constructed under a standardized baseline meteorological condition. The latter is based on a consideration that the observed air pollution data are confounded by the meteorological variation [3–6], and the adjustment largely removes the meteorological confounding and makes the adjusted averages comparable both spatially and temporally. Having these three perspectives in the analyses leads to a fuller view on Beijing's $PM_{2.5}$ pollution in the past 4 years and 2016 in particular.

2. Data

The city of Beijing established an air pollution monitoring network in January 2013 as part of the national monitoring network. There are 36 air-quality monitoring sites in Beijing, 35 of which are BMEMC sites and one at the US Embassy in Beijing. The locations of the air-quality monitoring sites along with the meteorological sites are marked in figure 1, and more specific site information is available in the electronic supplementary material, table S1.

As the observed $PM_{2.5}$ level is ordinarily impacted by the meteorological condition, we consider meteorological data at 15 weather observing stations of the China Meteorological Administration (CMA; marked in figure 1). The meteorological data consist of 6 hourly observed variables: air temperature, wind direction (WD) and speed, pressure, relative humidity (or dew point temperature, DEW) and precipitation, from March 2010 to February 2017. The reason for using three more years' meteorological data is for a better construction of a spatial and temporal baseline weather condition over the study region. This is to obtain a spatially and temporally adjusted average of $PM_{2.5}$ concentration which is comparable over space and time; see [3,7] for temporally adjusted averages of $PM_{2.5}$. The rationale for adding the spatial adjustment is that we have multiple meteorological stations which provide data of spatial variation.

As there were many missing values in January and February of 2013 in most sites when Beijing's air-quality monitoring network was first put in operation, we consider the seasonal year, namely hourly $PM_{2.5}$ data ranging from March 2013 to February 2017 that makes up four seasonal years. As mentioned earlier, an advantage of using the seasonal year is that it keeps the winter season intact without breaking it into two separate years. The time unit of the study is season, which consists of segments of three months starting from March, June, September and December, which represent the four seasons of spring, summer, autumn and winter, respectively. This takes advantage of Beijing's distinct seasons as each season provides a homogeneous weather pattern. A season offers more data than a month which is important for our adjustment method based on the non-parametric regression approach [8,9].

3. Meteorological variation and need for adjustment

The rationale of the meteorological adjustment is that there is much spatial and temporal variation that confounds the air-quality observations. As a result, we have to compare $PM_{2.5}$ concentration

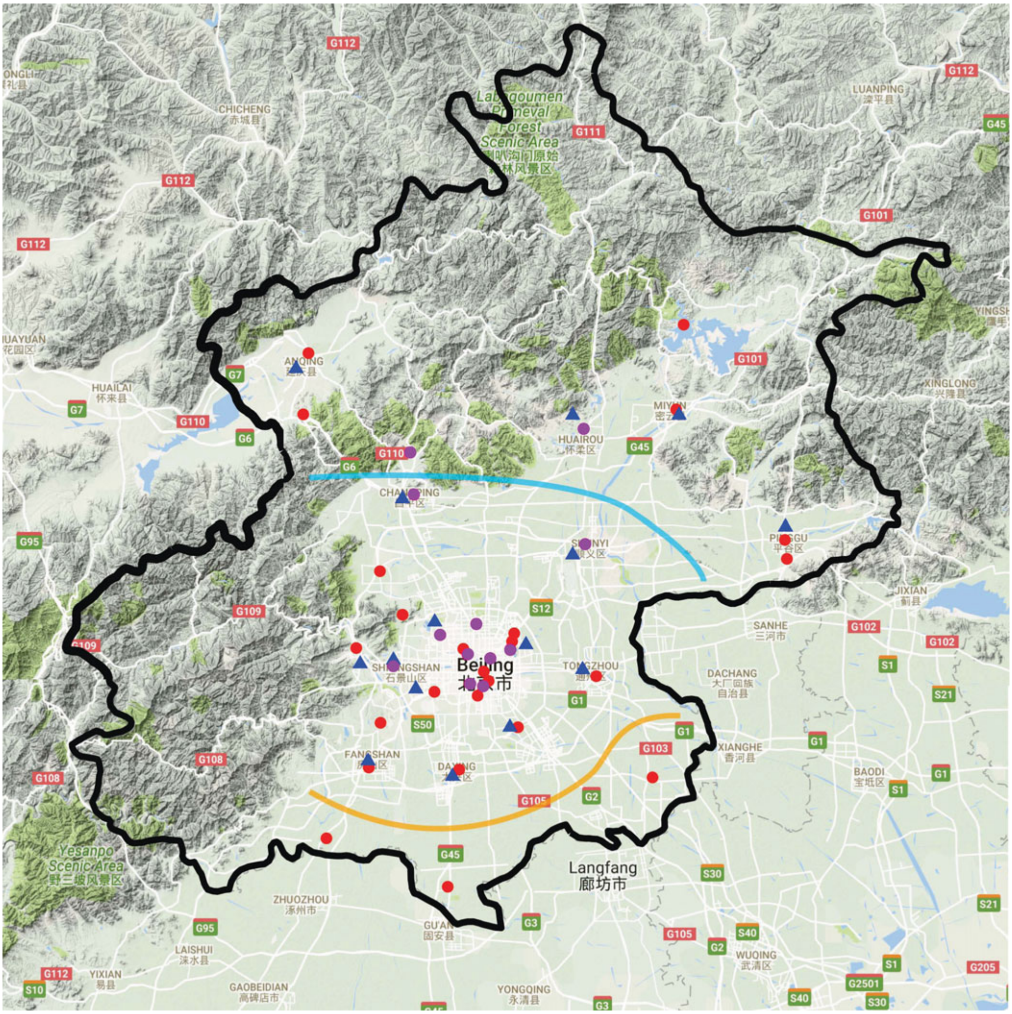

Figure 1. Locations of the 36 air-quality monitoring sites (red or purple dots) and 15 meteorological sites (blue triangles) in Beijing. The purple dots mark the 12 Guokong sites and the red dots show the other 24 sites. The region above the blue line is the north region, and that below the yellow line is the south region, the area in between is the central region.

under a comparable baseline weather equilibrium. To demonstrate the meteorological variation, we consider the probabilistic distributions of WD, cumulative wind speed (CWS) and DEW for January and February 2016 and 2017. The three variables are known to be important in the modelling of the meteorological confounding of $PM_{2.5}$ [3,7]. Specifically, the dew point reflects both relative humidity and air temperature [10].

The marginal and joint distributions (via the box plots and the contour plots) of the three variables are displayed in figure 2. To gain contrast, we also present the spatial–temporal baseline distributions of these three variables using seven (seasonal) years' data from March 2010 to February 2017.

Figure 2 shows that January and February of 2016 had quite favourable dispersion conditions. Relative to the two months in 2017 and the 7 years' baseline, January and February 2016 had a much higher percentage of northerly wind, larger CWS and lower DEW. The latter reflects a lower humidity and temperature regime that is usually associated with the clean, colder and dryer northerly wind flow, which removes the $PM_{2.5}$ in the city. This explains the low $PM_{2.5}$ averages in the first two months of 2016 (electronic supplementary material, figure S2). It is noted that the mountainous north has little emission loading, whereas the NCP region that neighbours the

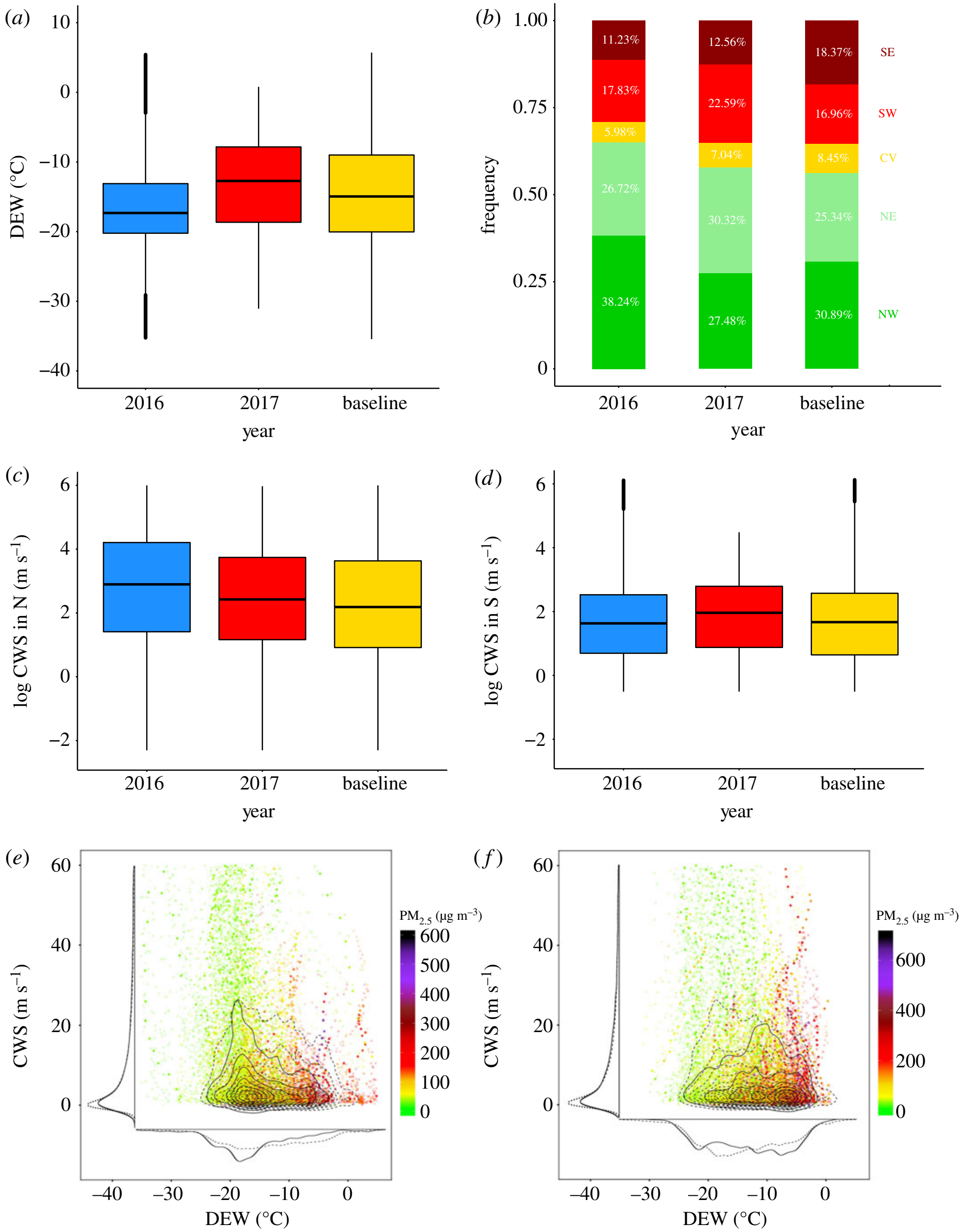

Figure 2. (*a–d*) Distributions of three key meteorological variables in January and February of 2016, 2017 and the 7 years' baseline from 2011 to 2017. (*a*) DEW. (*b*) Percentages of WD in 2016 and 2017 and also under the 7 years' baseline. WD are combined into five categories (from bottom up): northwest (NW), northeast (NE), calm and variable (CV), southwest (SW) and southeast (SE). (*c*,*d*) Log CWS in north (*c*) and south (*d*) WD. (*e*,*f*) Contour plots of the joint distribution of CWS and DEW in 2016 (*e*) and 2017 (*f*). The colour of the dots reflects the $PM_{2.5}$ concentration at the time of the observation.

south of Beijing is heavily installed with iron, steel and cement industries in Hebei province. By contrast, the dispersion condition in January and February 2017 was less favourable, as shown by less frequent northerly wind (more southerly wind) and higher dew point (hence more humidity and warmer temperature). The latter is known to encourage secondary generation of $PM_{2.5}$ [4,6].

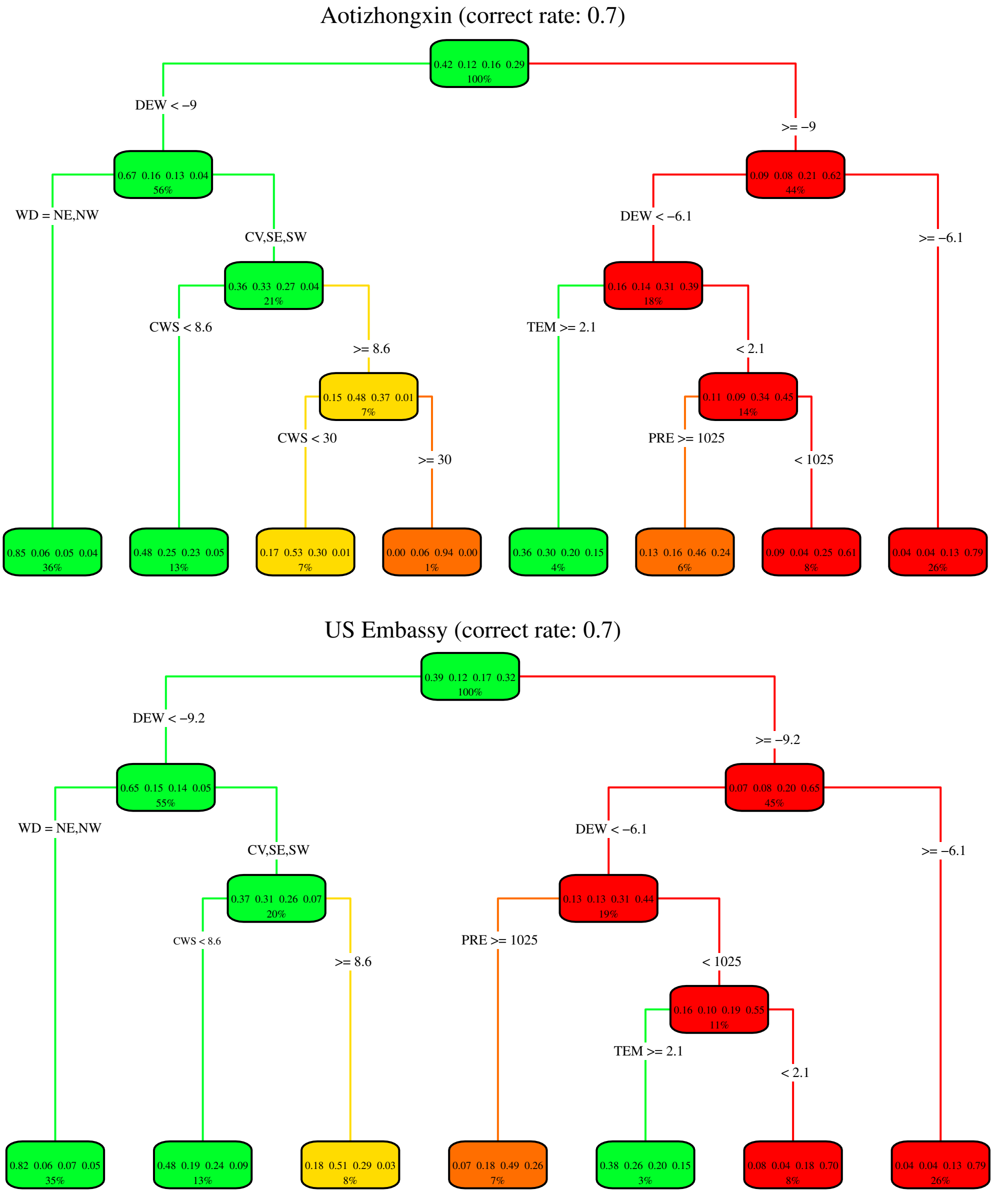

Figure 3. $PM_{2.5}$ classification trees with respect to five meteorological variables (without cumulative precipitation) in two air-quality monitoring sites, Aotizhongxin (*a*) and US Embassy (*b*), in the winter of 2016. The response $PM_{2.5}$ is classified into four categories: $PM_{2.5} \leq 35\,\mu g\,m^{-3}$ (green), $35\,\mu g\,m^{-3} < PM_{2.5} \leq 75\,\mu g\,m^{-3}$ (yellow), $75\,\mu g\,m^{-3} < PM_{2.5} \leq 150\,\mu g\,m^{-3}$ (orange) and $PM_{2.5} > 150\,\mu g\,m^{-3}$ (red). The four numbers inside each coloured node indicate the proportions of the four $PM_{2.5}$ categories at each layer of the branch, and the percentage represents the marginal proportion of the sample at the node.

The less favourable condition in the first two months of 2017 was also the case relative to the 7 years' baseline.

We can use tools of statistical learning to explore the relationship between $PM_{2.5}$ and meteorological variables and plot the classification trees [11] to visualize the result. Figure 3 and electronic supplementary material, figure S1, display the top four layers of the classification trees for $PM_{2.5}$ with respect to five meteorological variables at four monitoring sites in the winter of 2016. We did not consider the cumulative precipitation (CP) as Beijing has little of it in the winter. The response $PM_{2.5}$ is partitioned into four categories: $PM_{2.5} \leq 35\,\mu g\,m^{-3}$, $35\,\mu g\,m^{-3} < PM_{2.5} \leq$

Table 1. Empirical weights of the five meteorological covariates (without cumulative precipitation) in the top four layers of the classification trees in the entire region (36 sites) and three subregions for the winter of 2016. Meteorological covariates used are DEW, CWS, temperature (TEM), WD and pressure (PRE). For each covariate and each layer of a tree, the weight are obtained by the ratios of the numbers of monitoring sites where the covariate appears in the layer to the number of all monitoring sites available in the region.

		empirical weight							empirical weight				
region	layer	DEW	WD	CWS	PRE	TEM	region	layer	DEW	WD	CWS	PRE	TEM
entire	1	1.00	0.00	0.00	0.00	0.00	north	1	1.00	0.00	0.00	0.00	0.00
	2	0.44	0.50	0.04	0.02	0.00		2	0.56	0.33	0.00	0.11	0.00
	3	0.16	0.00	0.34	0.29	0.21		3	0.21	0.00	0.07	0.43	0.29
	4	0.28	0.02	0.11	0.31	0.28		4	0.33	0.11	0.11	0.22	0.23
central	1	1.00	0.00	0.00	0.00	0.00	south	1	1.00	0.00	0.00	0.00	0.00
	2	0.40	0.54	0.06	0.00	0.00		2	0.50	0.50	0.00	0.00	0.00
	3	0.11	0.00	0.47	0.25	0.17		3	0.33	0.00	0.17	0.17	0.33
	4	0.32	0.00	0.13	0.30	0.25		4	0.00	0.00	0.00	0.50	0.50

$75\,\mu g\,m^{-3}$, $75\,\mu g\,m^{-3} < PM_{2.5} \leq 150\,\mu g\,m^{-3}$ and $PM_{2.5} > 150\,\mu g\,m^{-3}$. Figure 3 and electronic supplementary material, figure S1, show that in the winter of 2016, nearly 70% of data can be correctly classified by trees with four layers. Table 1 summarizes the relative importance of the weather covariates at each layer of the trees by the relative frequency of the classification variable (variables) chosen by the trees standardized by the number of monitoring sites involved for the entire region with 36 monitoring sites, and the three sub-regions. The table shows that the dew point (DEW) is the dominating variable as it classifies the first layer of the tree in all cases. WD is the second most important classifier, followed by the CWS, pressure (PRE) and temperature (TEM). That the dew point is the most important factor that influences the $PM_{2.5}$ rather than the commonly perceived wind can be understood by the fact that the WD is the main factor that impacts the dew point, as the northly wind brings cooler and drier air from the cleaner north which reduces both the DEW and the $PM_{2.5}$. By contrast, the southerly wind brings warmer and more humid air from the more polluted south that elevates the $PM_{2.5}$ concentration. This is consistent with the previous findings reported in [7]. Despite the classification power of the trees, for the purpose of more accurate air-quality assessment, we will employ the non-parametric regression approach as it treats $PM_{2.5}$ as a continuous random variable.

These suggest a need to adjust the observed $PM_{2.5}$ under a baseline meteorological condition that reflects the equilibrium weather in the study region. Our adjustment method is one such method whose details are outlined later.

4. Spatially and temporally meteorological adjustment

The official air-quality statistics in China are based on the simple averages of hourly $PM_{2.5}$ readings. However, the observed $PM_{2.5}$ levels are known to be influenced by emission of pollutants, meteorological conditions and their interaction [4,6]. The secondary generation of fine PMs is an act of the interaction, which is, for instance, much encouraged by high humidity with high temperature and calm wind. It is found using data from the US Embassy in Beijing that more than 70% of the variation in $PM_{2.5}$ can be explained by the variation in meteorological variables [7]. Indeed, increased emissions accompanied with a favourable dispersion condition can result in lower $PM_{2.5}$ readings than a lowered emission regime but with unfavourable weather conditions. This indicates a similar situation to statistical observational studies [12], where bias due to pre-treatment variables needs to be adjusted in the evaluation of treatment effects.

At each $PM_{2.5}$ observational site s, let $Y_{ijt}(s)$ be the $PM_{2.5}$ concentration at hour t of year i in season j, where $j=1$ for spring (March–May), $j=2$ for summer (June–August), $j=3$ for autumn (September–November), and $j=4$ for winter (December to February of the following year); and $X_{ijt}(s)$ denotes the vector of the six meteorological variables: air temperature (TEM), DEW, air pressure (PRE), WD, CWS at a direction and CP. We classify the wind into five directions: NW, NE, SE, SW and CV. CWS adds up hourly wind speed at one direction until the direction changes. The same applies to CP.

We use the following non-parametric regression model [8,9]:

$$Y_{ijt}(s) = m_{ij}(X_{ijt}(s), s) + \epsilon_{ijt}(s),$$

where $\epsilon_{ijt}(s)$ is the residual, and $m_{ij}(x, s)$ is the regression function of $Y_{ijt}(s)$ at site s with weather variables $X_{ijt}(s)$ being conditioned at x. Mathematically, $m_{ij}(x, s) = E\{Y_{ijt}(s) \mid X_{ijt}(s) = x\}$, where $E(\cdot \mid \cdot)$ denotes the conditional expectation. The above non-parametric model does not assume an explicit form of the regression function that offers model robustness and flexibility in the relationship between the pollution level and the meteorological variables. This is advantageous as it can avoid the risk of model misspecification. It is also noted that the purpose of this study is air-quality assessment, which does not require a detailed parametric or semiparametric model since the purpose of the kernel smoothing is obtaining the adjusted averages or quantiles of the response variable.

The approach we use to adjust for the meteorological confounding consists of four steps.

Step 1. Construct a spatially and temporally meteorological equilibrium probability density for a particular season j based on the 7 years' (March 2010 to February 2017) historic weather data of all available weather stations. Let $f_{ij}(x, s)$ be the probability density function of weather variables $X_{ijt}(s)$ at the location s of season j in year i, which can be estimated based on the meteorological data at the location s of the particular season and year via the kernel density estimation approach [13]. The spatial–temporal baseline probability density function for season j is

$$f_{\cdot j}(x) = 7^{-1}|\mathcal{M}|^{-1} \sum_{i=1}^{7} \sum_{s \in \mathcal{M}} f_{ij}(x, s), \tag{4.1}$$

where $\mathcal{M}$ denotes the collection of all available meteorological stations and $|\mathcal{M}|$ denotes their size (which is 15 in the study), and the number 7 corresponds to 7 years of weather data. The spatial and temporal baseline is attained by averaging the weather density with respect to the years and sites at a fixed season j. Let $\hat{f}_{\cdot j}(x)$ be the empirical estimate, which can be obtained by the kernel density estimator $\hat{f}_{ij}(x, s)$ [13]. However, in the practical calculation of the adjusted averages, explicit estimation of $f_{\cdot j}(x)$ can be avoided as shown below.

Step 2. For a season j of year i, regress hourly $PM_{2.5}$ values $Y_{ijt}(s)$ with respect to the weather condition $X_{ijt}(s)$ at site s to obtain an estimate $\hat{m}_{ij}(x, s)$ for the regression function $m_{ij}(x, s)$ via the non-parametric regression method [8,9]. In this step, only data of the specific season j of year i at site s is used.

Step 3. Compute the spatially and temporally adjusted average $PM_{2.5}$ for season j of year i at the site s by

$$\hat{\mu}_{ij}(s) = \int \hat{m}_{ij}(x, s)\hat{f}_{\cdot j}(x)\, \mathrm{d}x. \tag{4.2}$$

As $\hat{f}_{\cdot j}(x)$ is a probability density, the integral in (4.2) can be replaced by summations so that

$$\hat{\mu}_{ij}(s) = |\mathcal{M}|^{-1} \left(\sum_{a=1}^{7} n_{aj} \right)^{-1} \sum_{a=1}^{7} \sum_{s' \in \mathcal{M}} \sum_{t=1}^{n_{aj}} \hat{m}_{ij}(X_{ajt}(s'), s),$$

where n_{aj} is the size of meteorological data in season j of year a at each site.

Step 4. From the site specific averages $\hat{\mu}_{ij}(s)$, we can derive the weather adjusted average over a region $\mathcal{R}$:

$$\hat{\mu}_{ij}(\mathcal{R}) = |\mathcal{R}|^{-1} \sum_{s\in\mathcal{R}} \hat{\mu}_{ij}(s), \tag{4.3}$$

where $|\mathcal{R}|$ denotes the number of monitoring sites in $\mathcal{R}$. Specific examples of $\mathcal{R}$ in our studies are central, south and north regions as well as the region encompassed by the 11 Guokong sites.

It is noted that there are around 2160 hourly observations per season in the non-parametric estimation of the regression function. With six covariates, this largely reduces the curse of dimension for the non-parametric approach in Step 2 when compared with employing monthly data. As mentioned a few paragraphs earlier, the purpose of the current study is air-quality assessment, which allows us to bypass a detailed parametric or semiparametric model since the adjusted average is an integration of the product of the regression function and the density function, whose estimation enjoys a faster rate of convergence ($\sqrt{n_{ij}}$, see [14]) as compared to that of estimating the regression function $m_{ij}(x, s)$. As a result, the issue of curse-of-dimensionality associated with kernel smoothing is less impactful in the current context of the air-quality assessment.

To obtain the standard error for $\hat{\mu}_{ij}(\mathcal{R})$, we employ the spatial and temporal block bootstrap method [15]. The method has been applied in the air-quality study of five Chinese cities [3] to supply the standard errors to the adjusted monthly $PM_{2.5}$ estimates. The standard errors together with $\hat{\mu}_{ij}(\mathcal{R})$ are used to test for various statistical significance in the study. In this study, much of the interest is whether the adjusted averages in two consecutive years are significantly increased or decreased statistically.

Specifically, in season j of year i_1 and $i_2 = i_1 + 1$ over region $\mathcal{R}$, we first obtain the adjusted averages $\hat{\mu}_{i_1j}(\mathcal{R})$ and $\hat{\mu}_{i_2j}(\mathcal{R})$ by the above spatial–temporal adjustment approach, and then implement a spatial–temporal bootstrap method to acquire the standard error of $\hat{\mu}_{i_2j}(\mathcal{R}) - \hat{\mu}_{i_1j}(\mathcal{R})$, denoted by $\hat{\sigma}_{i_2-i_1,j}(\mathcal{R})$. If $\hat{\mu}_{i_2j}(\mathcal{R}) - \hat{\mu}_{i_1j}(\mathcal{R}) \geq (<)0$; we shall consider testing the hypotheses $H_0 : \mu_{i_2j} - \mu_{i_1j} = 0$ versus $H_1 : \mu_{i_2j} - \mu_{i_1j} > (<)0$, where H_0 is the null hypotheses and H_1 the alternative hypotheses. Under the null hypothesis H_0, the test statistic

$$\frac{\hat{\mu}_{i_2j}(\mathcal{R}) - \hat{\mu}_{i_1j}(\mathcal{R})}{\hat{\sigma}_{i_2-i_1,j}(\mathcal{R})},$$

asymptotically follows the standard normal distribution, which allows us to obtain the p-values and the corresponding statistical significances to see whether the air quality in season j of year i_2 is better than that in season j of year i_1 or not.

5. Prognosis of $PM_{2.5}$ concentration from 2013 to 2016

By conducting the spatially and temporally meteorological adjustment (details in §4), we first obtain adjusted averages at each monitoring site for each season of a year. By spatially smoothing [9], the site-specific averages with the Gaussian kernel and a bandwidth $h = 0.15$ (in latitude and longitude degrees), we provide in figure 4 a seasonal $PM_{2.5}$ concentration map over Beijing for each season from 2013 to 2016. Here, we use the seasonal year that runs from spring (March–May), summer (June–August), autumn (September–November) to winter (December–February of the following year). We choose a season as the basic time unit of study to take advantage of Beijing's distinct seasons and the fact that each season has a quite homogeneous meteorological condition. The latter provides stability for the meteorological adjustment we employ in the study.

The average $PM_{2.5}$ concentration has quite large seasonal and spatial variations as shown in figure 4. On the seasonal variation, winter has the highest concentration followed by autumn and spring, and summer is the season with the lowest $PM_{2.5}$. However, the city-wide averages in the best summer season were still above $60\,\mu\mathrm{g\,m^{-3}}$ in 2015 and 2016, Beijing's annual target

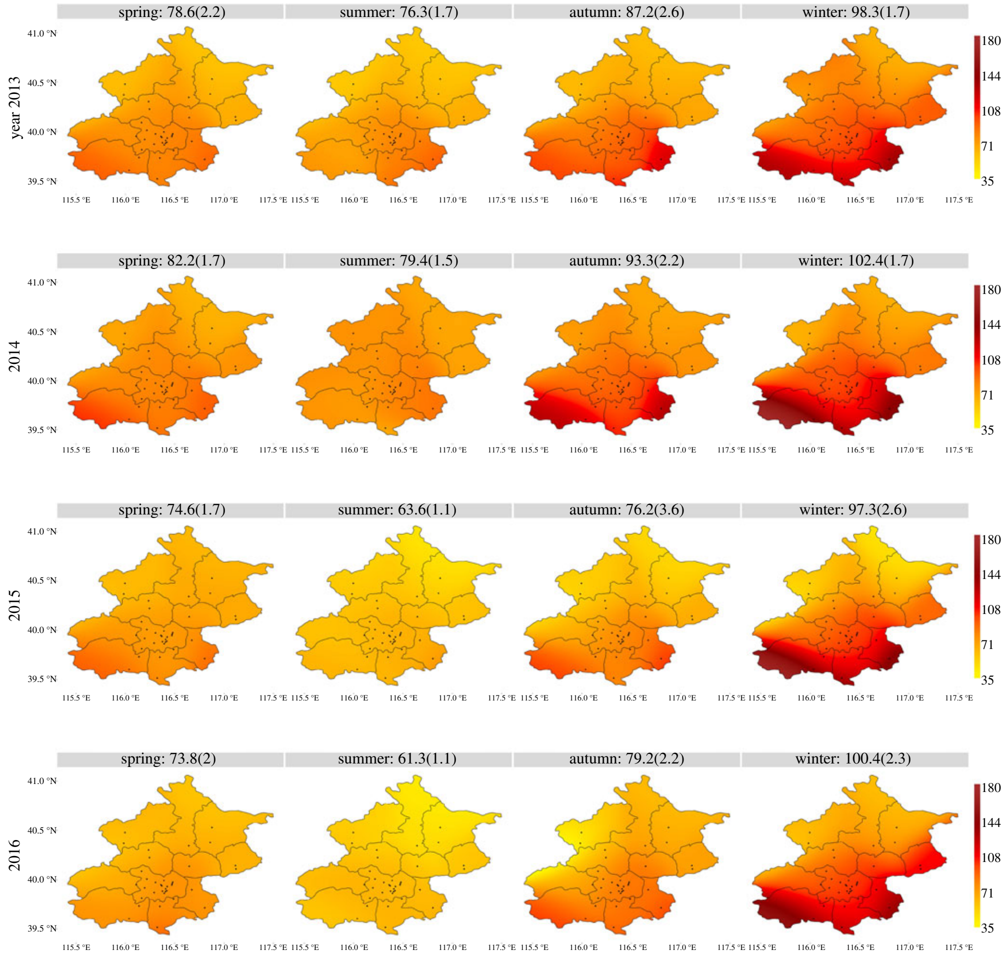

Figure 4. Spatially and temporally adjusted average $PM_{2.5}$ concentration in Beijing from 2013 to 2016. The numbers above the panels are the seasonal averages obtained by averaging over the weather adjusted averages of the 36 monitoring sites in Beijing with the numbers inside the parentheses were their standard errors. The kernel smoothing bandwidth used for generating the maps is 0.15 (in degree), and the seasonal year is used.

$PM_{2.5}$ level set by the State Council for 2017 [16]. The winter average $PM_{2.5}$ was in the range of 97–103 $\mu g\,m^{-3}$, which was more than 50% higher than that of the summer.

The spatial differences are well illustrated in figure 5 by the kernel smoothed average $PM_{2.5}$ curves [9] in the south–north direction (latitude) in 2016; those for 2013, 2014 and 2015 are given in electronic supplementary material, figure S3. The smoothed average curves show clearly a south-high and north-low characteristic in the spatial distribution of $PM_{2.5}$ in Beijing. The spatial variation was the largest in winter and autumn. In winter, the $PM_{2.5}$ level in the southern edge of the city can be 50% higher than that in the city centre around Changan Avenue. This pattern was little changed in the past 4 years, which was partially due to the pollutant transportation from the industrial Hebei province in the south.

To conduct finer analyses, we group 36 air-quality monitoring sites in Beijing into three regions: central, north and south, as displayed in figure 1 and detailed in electronic supplementary material, table S1. The central region largely encompasses the sixth ring road that encloses the central Beijing with 80% of population living within it. It has 25 monitoring sites, including 10 of the 11 Guokong sites, and is the most similar to the area encompassed by the 11 Guokong sites used by BMEMC.

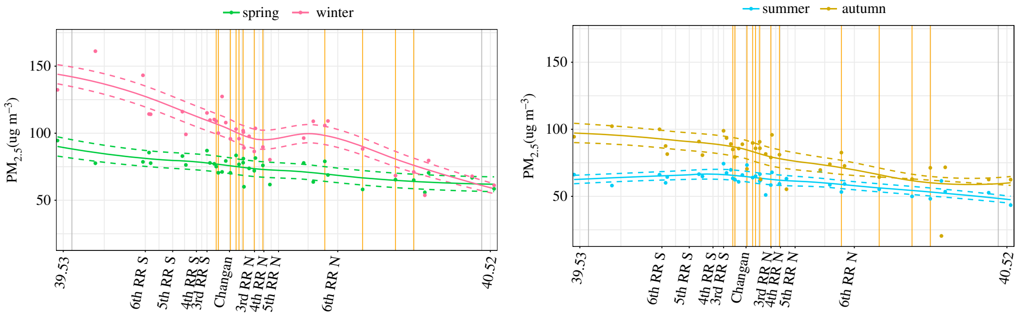

Figure 5. Adjusted average $PM_{2.5}$ curves (solid line) in the south–north direction (latitude) and the 95% confidence intervals (dashed lines) for four seasons in 2016. Vertical yellow lines mark the latitudes of Guokong sites. Dots indicate the site-specific adjusted averages. Horizontal label: 4th RR S (N) means 4th Ring Road south (north) and Changan is for Changan Avenue.

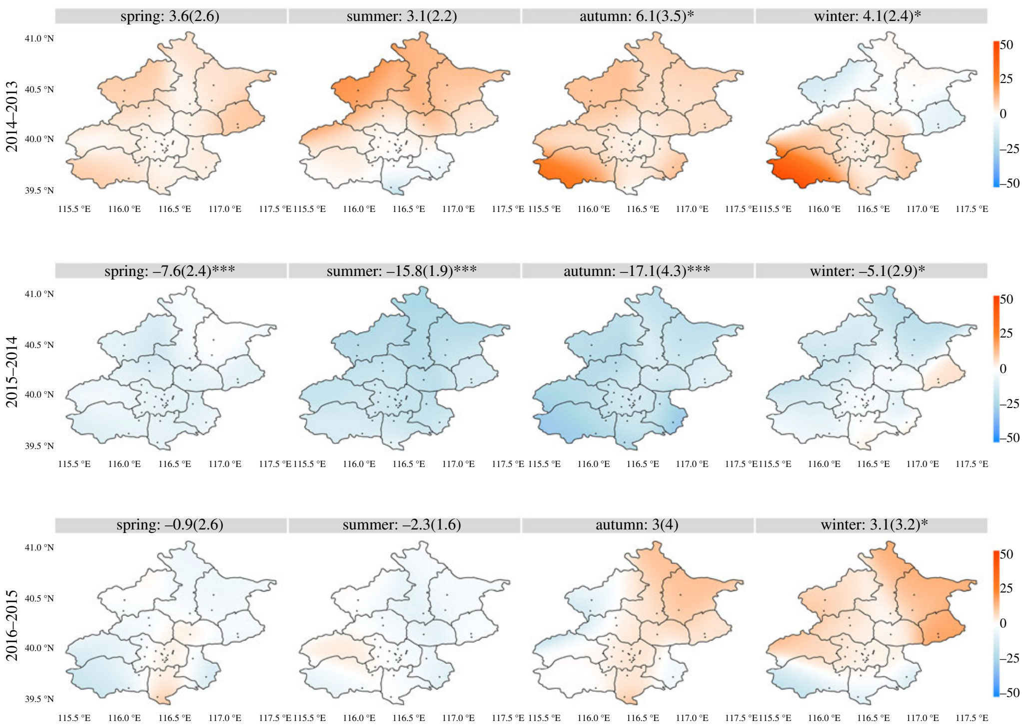

Figure 6. Annual seasonal differences in the adjusted average $PM_{2.5}$ concentration in Beijing. The numbers above the maps are the annual seasonal changes over entire Beijing and the numbers inside the parentheses are their standard errors. The asterisks represent the level of significance for testing the increase or the decrease of the annual changes between two consecutive years (no asterisk means the increase/decrease was not significant; *$0.025 \le p < 0.05$; **$0.01 \le p < 0.025$; ***$p < 0.01$). The figure shows insignificant changes in all four seasons of 2015 and 2016.

To gain information on the yearly changes in $PM_{2.5}$, we difference the average $PM_{2.5}$ maps of a season in figure 4 between two consecutive years and report them in figure 6. Figure 6 suggests that the $PM_{2.5}$ level was substantially reduced in all four seasons of 2015, but there was no improvement in 2016 as conveyed by the associated p-values. Figure 7 supplements figure 6 by providing the seasonal $PM_{2.5}$ averages for the three regions in Beijing and their yearly changes. It indicates that the decline in $PM_{2.5}$ was significant at the 5% level from spring to autumn of

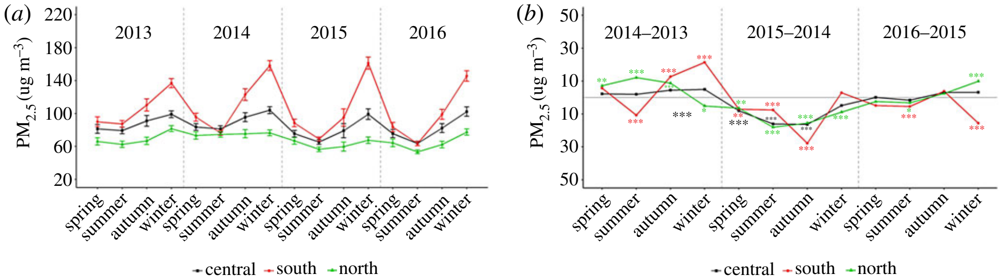

Figure 7. (*a*) Adjusted averages with bars indicating the 95% confidence intervals. The averages for a region are obtained by averaging the adjusted average $PM_{2.5}$ at all monitoring sites in the region. (*b*) Yearly differences in the adjusted averages where the asterisks indicate the level of significance in the yearly increase/decrease as specified in the caption of figure 6.

2015 in all three regions. The winter of 2015 was mixed with the decline in the central and north, but a non-significant increase in the south. The story for 2016 was a lack of a clear trend. This was particularly the case for the central region with all four seasons' changes being statistically insignificant.

6. Situation in 2016

The conclusion of no significant improvement in 2016 as conveyed in figure 6 is sharply different from the official 9.9% decline in $PM_{2.5}$ for calendar year 2016. Specifically, our weather adjusted annual average for the central region in 2016 was 80.9 $\mu g\,m^{-3}$, which was 1.4% higher than the 2015 average of 79.8 $\mu g\,m^{-3}$. If the 9.9% reduction were the case, it would put Beijing's 2016 concentration to 72.6 $\mu g\,m^{-3}$ and make the task of reaching the 60 $\mu g\,m^{-3}$ annual target imposed by the State Council much easier in 2017. For the sake of effective air-quality management, there is a great need to settle the difference.

There are three major differences in the composition of the statistics between BMEMC and our assessment. Firstly, the BMEMC statistic is based on the calendar year, while we use the seasonal year. Secondly, the BMEMC uses simple averages while we employ the adjustment that removes the meteorological confounding and makes the averages comparable spatially and temporarily. Lastly, the BMEMC statistic is based on the 11 Guokong sites, while ours is based on the 25 sites in the central region.

To reconcile the statistics from the city and our assessment, we report in table 2 the simple and adjusted averages for the 11 Guokong sites and the central region, respectively. Electronic supplementary material, figure S4 is a visual display for part of table 2.

Table 2 shows that if we adopt the simple average but based on the seasonal year instead of the calendar year, the annual average for the Guokong sites became 79.4 $\mu g\,m^{-3}$, which was 6.7% more than the corresponding average of 74.4 $\mu g\,m^{-3}$ for 2015. This means that the improvement conveyed by the 9.9% reduction in the calendar year 2016 was actually rather fragile as a simple two-month time shift could produce a sharply different result. A reason behind this dramatic turning around was extremely low $PM_{2.5}$ recorded in February and January 2016, which can be shown in electronic supplementary material, figure S2. Both months are in the winter, which usually endures high $PM_{2.5}$ in the range of 90–100 $\mu g\,m^{-3}$ caused by winter heating combined with a much lower boundary layer height. However, February 2016 saw that the simple average $PM_{2.5}$ fell to an unprecedented 43.6 $\mu g\,m^{-3}$, and the average for January 2016 was 68 $\mu g\,m^{-3}$, also well below the winter norm.

If we carry out the spatial and temporal meteorological adjustment for the 11 Guokong sites, the size of the decrease in calendar year 2016 was reduced from 9.9% to a mild 4.4%, indicating more than 55% of the 9.9% decline was due to more favourable weather conditions in calendar year 2016. At the same time, the size of the increase in the seasonal year 2016 was also reduced

Table 2. Annual averages of $PM_{2.5}$ concentration based on two types of years (calendar year versus seasonal year), two collections of monitoring sites (Guokong versus central region with 25 sites) and two kinds of statistical averages (simple averages versus (weather) adjusted averages). The asterisks (*) represent the level of significance for testing the increase/decrease of the annual change between two consecutive years (no asterisk means the increase/decrease was not significant; *$0.025 \leq p < 0.05$; **$0.01 \leq p < 0.025$; ***$p < 0.01$).

			average				annual percentage change (%)		
sites	form of average	year type	2013	2014	2015	2016	2014–2013	2015–2014	2016–2015
Guokong	simple	calendar	81.3	87.1	80.9	72.6	7.1	−7.1	−10.3[a]
		seasonal	88.2	82.3	74.4	79.4	−6.7	−9.6	6.7
	adjusted	calendar	80.6	84.9	76.8	73.4	5.3**	−9.5***	−4.4*
		seasonal	82.7	85.1	75.5	77.2	2.9	−11.3***	2.3
central	simple	calendar	86.5	93.5	84.9	76.4	8.1	−9.2	−10.0
		seasonal	94.2	89.1	78.0	83.1	−5.4	−12.5	6.5
	adjusted	calendar	86.0	90.9	80.6	76.9	5.7***	−11.3***	−4.6*
		seasonal	87.7	91.1	79.8	80.9	3.9*	−12.4***	1.4

[a] The −10.3% differs from the −9.9% figure from the BMEMC as we have combined several data sources to lower the rates of missing values.

from 6.7 to 2.3% after adjusting for the weather conditions. The 2.3% increase was not statistically significant at the 5% level, while the 4.4% decrease was significant at the 5% level but not at the 2.5% level.

If we focus on the central region that has 25 air-quality monitoring sites, we find the simple calendar year average in 2016 was $76.4\,\mu g\,m^{-3}$ and the simple seasonal year average was $83.1\,\mu g\,m^{-3}$, representing an annual 10.0% reduction and 6.5% increase, respectively. However, after meteorological adjustment, these two percentages became a 4.6% reduction and a 1.4% increase, respectively, and the latter was not statistically significant at the 5% level while the former was not significant at 2.5%.

Table 2 also demonstrates the impacts of the spatial and temporal meteorological adjustment in mitigating the effect of extreme meteorological variation and in producing more robust air-quality assessments.

This smoothing effect of the adjustment is clearly shown by the adjusted seasonal average curves for both the Guokong sites and the central region to largely correct the highs and lows of the simple averages.

We note that from table 2 and electronic supplementary material, figure S4, the simple averages of the central region were larger than those of the 11 Guokong sites with the difference being $5–6\,\mu g\,m^{-3}$ in 2013 and 2014, and $3–4\,\mu g\,m^{-3}$ in 2015 and 2016. Figure 5 reveals that all the Guokong sites are located in the middle or north part of the city where a relatively lower $PM_{2.5}$ level is registered. This means that the southern part of the city is not included in the air-quality management of the Guokong system, leading to biased assessments. We strongly suggest allocating Guokong sites to the southern part of the city to make the official statistics representative of the city.

7. Discussion

By conducting several statistical analyses based on more than 7 million pieces of $PM_{2.5}$ and meteorological records, we have shown that there was a significant reduction in the $PM_{2.5}$ level in Beijing in 2015. However, our analysis cannot confirm the improvement in 2016 as conveyed in the official 9.9% reduction in $PM_{2.5}$. By contrast, we have detected a non-significant increase in the seasonal year average $PM_{2.5}$ in 2016 after adjusting for the meteorological variation.

An underlying reason for the rather fragile and uncertain situation regarding Beijing's air quality in 2016 was the start of an economic recovery that took momentum from spring 2016 as reflected in the output of three high energy consuming products, steel, iron and cement, in the neighbouring Hebei province, according to the statistics displayed in electronic supplementary material, figure S5. The province's outputs for the three products increased at annual rates of 4.4%, 6.4% and 8.0%, respectively, in 2016. In particular, both the steel and iron output had notable increases in all the seasons since spring 2016. The cement outputs in 2016 had risen in three consecutive seasons from the spring, despite some decline in the winter of 2016. More importantly, the coal consumption in Hebei, after 3 years' steady decline, had rebounded with a 3.2% increase in the autumn of 2016. The latter correlated well with the significant decrease in $PM_{2.5}$ in 2015 and the increase in the $PM_{2.5}$ starting from autumn 2016 as shown in figure 6. The rise of coal consumption and the outputs of the three major heavy industrial products in 2016 imply that the air pollution mitigation campaign in 2017 would be an up-hill endeavour. A much tougher environmental enforcement and strategic planning are needed in order to reduce the $PM_{2.5}$ level. At the same time, our study shows that using sound and objective statistical measures are urgently needed in air-quality management in China.

Data accessibility. $PM_{2.5}$ data of the MEP sites are from https://wat.epmap.org/ at Qingyue Open Environmental Data Center and those of the US Embassy is from http://www.stateair.net/web/historical/1/1.html. The weather data in Beijing are from http://weather.nocrew.org. Data used in the paper are available upon request from S.X.C. (csx@gsm.pku.edu). The coal consumption data of Beijing and Hebei are from http://www.sxcoal.com. The steel, iron, cement output and electricity generation data are from http://www.stats.gov.cn.

Authors' contributions. S.Z. and B.G. performed most of the numerical analyses while assisted by A.D., J.H. and Z.X. S.X.C. supervised the project. S.X.C., G.B. and S.Z. wrote the manuscript with inputs from the others. All authors gave final approval for publication.

Competing interests. We have no competing interests.

Funding. The research was partially supported by China's National Key Research Special Program Grant no. 2016YFC0207701 and 2016YFC0207702, National Key Basic Research Program Grant 2015CB856000 and National Natural Science Foundation of China grant nos. 71532001 and 71371016. B.G. was supported in part by the Fundamental Research Funds for the Central Universities grant nos. JBK120509 and JBK140507, National Natural Science Foundation of China grant no. 11601356 and the Center of Statistical Research at Southwestern University of Finance and Economics. J.H. was supported in part by the National Natural Science Foundation of China grant no. 11701466.

Acknowledgements. We thank two anonymous referees for constructive comments and suggestions which have improved the content and presentation of the paper. Also, we thank the Center for Statistical Science at Peking University for various logistic support.

References

1. Zhang X, Wang Y, Niu T, Zhang X, Gong S, Zhang Y, Sun J. 2012 Atmospheric aerosol compositions in China: spatial/temporal variability, chemical signature, regional haze distribution and comparisons with global aerosols. *Atmos. Chem. Phys.* **12**, 779–799. (doi:10.5194/acp-12-779-2012)
2. Guo S *et al.* 2014 Elucidating severe urban haze formation in China. *Proc. Natl Acad. Sci. USA* **111**, 17 373–17 378. (doi:10.1073/pnas.1419604111)
3. Liang X, Li S, Zhang S, Huang H, Chen SX. 2016 $PM_{2.5}$ data reliability, consistency, and air quality assessment in five Chinese cities. *J. Geophys. Res. Atmos.* **121**, 10 220–10 236. (doi:10.1002/2016JD024877)
4. Huang K, Zhuang G, Wang Q, Fu J, Lin Y, Liu T, Han L, Deng C. 2014 Extreme haze pollution in Beijing during January 2013: chemical characteristics, formation mechanism and role of fog processing. *Atmos. Chem. Phys. Discuss.* **14**, 7517–7556. (doi:10.5194/acpd-14-7517-2014)
5. Wang L, Zhang N, Liu Z, Sun Y, Ji D, Wang Y. 2014 The influence of climate factors, meteorological conditions, and boundary-layer structure on severe haze pollution in the Beijing-Tianjin-Hebei region during January 2013. *Adv. Meteorol.* **2014**, 685971. (doi:10.1155/2014/685971)

6. Zhang R, Wang G, Guo S, Zamora ML, Ying Q, Lin Y, Wang W, Hu M, Wang Y. 2015 Formation of urban fine particulate matter. *Chem. Rev.* **115**, 3803–3855. (doi:10.1021/acs.chemrev.5b00067)
7. Liang X, Zou T, Guo B, Li S, Zhang H, Zhang S, Huang H, Chen SX. 2015 Assessing Beijing's $PM_{2.5}$ pollution: severity, weather impact, APEC and winter heating. *Proc. R. Soc. A* **471**, 20150257. (doi:10.1098/rspa.2015.0257)
8. Fan J, Yao Q. 2003 *Nonlinear time series nonparametric and parametric methods*. New York, NY: Springer.
9. Härdle W. 1990 *Applied nonparametric regression*. Cambridge, UK: Cambridge University Press.
10. Alduchov OA, Eskridge RE. 1996 Improved magnus form approximation of saturation vapor pressure. *J. Appl. Meteorol.* **35**, 601–609. (doi:10.1175/1520-0450(1996)035<0601:IMFAOS>2.0.CO;2)
11. Breiman L, Friedman J, Stone CJ, Olshen RA. 1984 *Classification and regression trees*. Boca Raton, FL: CRC Press.
12. Rosenbaum PR. 2002 *Observational studies*. New York, NY: Springer.
13. Silverman BW. 1998 *Density estimation for statistics and data analysis*. New York, NY: Chapman & Hall.
14. Chen SX, Tang CY, Qin J. 2013 Mann-whitney test with adjustments to pre-treatment variables for missing values and observational study. *J. R. Stat. Soc. B* **75**, 81–102. (doi:10.1111/j.1467-9868.2012.01036.x)
15. Davison AC, Hinkley DV. 2010 *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.
16. Chinese State Council. 2013 *Atmospheric pollution prevention and control action plan*. Beijing, China: People's Press. [In Chinese.]