AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2016JD024877

Key Points:

- Air quality assessment using data from the U.S. diplomatic posts and the nearby MEP sites shows highly consistent results
- Significant and consistent winter heating effects in Bejing and Shenyang are detected using both the U.S. and the MEP data
- $PM_{2.5}$ concentrations in the five cities have showed moderate decline since 2014

Supporting Information:

- Supporting Information S1

Correspondence to:
S. X. Chen and H. Huang,
csx@gsm.pku.edu.cn;
huanghui@math.pku.edu.cn

Citation:
Liang, X., S. Li, S. Zhang, H. Huang, and S. X. Chen (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.

Received 3 FEB 2016
Accepted 17 AUG 2016
Accepted article online 23 AUG 2016
Published online 13 SEP 2016

$PM_{2.5}$ data reliability, consistency, and air quality assessment in five Chinese cities

Xuan Liang[1], Shuo Li[2], Shuyi Zhang[1], Hui Huang[3,4], and Song Xi Chen[1,3,5]

[1]Guanghua School of Management, Peking University, Beijing, China, [2]Department of Statistics, Tianjin University of Finance and Economics, Tianjin, China, [3]Center for Statistical Science, Peking University, Beijing, China, [4]Department of Probability and Statistics, Peking University, Beijing, China, [5]Department of Statistics, Iowa State University of Science and Technology, Ames, Iowa, USA

Abstract We investigate particulate matter ($PM_{2.5}$) data reliability in five major Chinese cities: Beijing, Shanghai, Guangzhou, Chengdu, and Shenyang by cross-validating data from the U.S. diplomatic posts and the nearby Ministry of Environmental Protection sites based on 3 years' data from January 2013. The investigation focuses on the consistency in air quality assessment derived from the two data sources. It consists of studying (i) the occurrence length and percentage of different $PM_{2.5}$ concentration ranges; (ii) the air quality assessment for each city; and (iii) the winter-heating effects in Beijing and Shenyang. Our analysis indicates that the two data sources produced highly consistent air quality assessment in the five cities. This is encouraging as it would inject a much needed confidence on the air pollution measurements from China. We also provide air quality assessments on the severity and trends of the fine particulate matter pollution in the five cities. The assessments are produced by statistically constructing the standard monthly meteorological conditions for each city, which are designed to minimize the effects of confounding factors due to yearly variations of some important meteorological variables. Our studies show that Beijing and Chengdu had the worst air quality, while Guangzhou and Shanghai faired the best among the five cities. Most of the five cities had their $PM_{2.5}$ concentration decreased significantly in the last 2 years. By linking the air quality with the amount of energy consumed, our study suggests that the geographical configuration is a significant factor in a city's air quality management and economic development.

1. Introduction

Accompanying China's rapid industrialization in the last two decades, there has been a sharp increase in energy consumption. Official statistics [*National Bureau of Statistics*, 2015] indicate that China's total energy consumption increased from 1555 million tons (coal equivalent) in 2001 to 4260 million tons coal equivalent in 2014, representing a 1.7-fold increase since China joined the World Trade Organization in December 2001. This staggering increase in energy consumption accompanied with lax environmental regulations has led to severe and chronic air pollution in the near earth atmosphere over a substantial part of China [*Tian et al.*, 2007].

One of the main constituents of air pollution is fine particulate matter ($PM_{2.5}$) [*Guo et al.*, 2014]. $PM_{2.5}$ consists of airborne particles with aerodynamic diameters less than 2.5 μm. They are known to adversely affect human health [*Pope et al.*, 2009] and potentially have lasting negative effects on agriculture [*Burney and Ramanathan*, 2014], climate, ecosystem, and many aspects of society [*Wang et al.*, 2014; *Prajapati*, 2012; *Zhao et al.*, 2013].

The U.S. Embassy in Beijing started releasing hourly $PM_{2.5}$ readings in April 2008, followed by the consulates in Guangzhou, Shanghai, Chengdu, and Shenyang in November and December 2011, June 2012, and April 2013, respectively. China's Ministry of Environmental Protection (MEP) began to report hourly $PM_{2.5}$ data in 74 cities (which include the five cities with the U.S. diplomatic posts) from January 2013, which was extended to 338 cities in January 2015.

How reliable are the $PM_{2.5}$ data in China? Data reliability is a prerequisite in the campaign to combat the chronic air pollution in China. There have been reports of discrepancies between MEP's and the U.S. Embassy's readings on $PM_{2.5}$ prior to 2013 [*Spegele*, 2012]. Studies have revealed manipulations of PM_{10} data in some Chinese cities from 2001 to 2010 [*Chen et al.*, 2012; *Ghanem and Zhang*, 2014; *Andrews*, 2008]. PM_{10} was the measure used to count "blue sky days" in Chinese cities before $PM_{2.5}$ was adopted in 2013 and was one of the performance measures for city mayors. *Stoerk* [2016] compared data of three sources in Beijing from year

2008 to year 2013: daily air quality measurements from Beijing Municipal Environmental Protection Bureau and the U.S. Embassy and aerosol optical depth from AERONET measurement station in Beijing. They found misreporting of the official PM_{10} data prior to 2013. The need for checking $PM_{2.5}$ data reliability is profound. If the data are confirmed reliable, it would provide much needed confidence on the air pollution measurements in China.

We investigate the reliability of hourly $PM_{2.5}$ data by cross-validating data from the U.S. diplomatic posts (U.S. posts) with those of the nearby MEP sites in the five major Chinese cities based on 3 year's data since January 2013. Because the U.S. posts and the MEP sites do not have the same levels of local emission and circumstances with the monitoring sites and devices, the hourly $PM_{2.5}$ readings between the U.S. post and the MEP are not directly comparable. Instead, we focus on the consistency in the air quality assessment derived from these two sources of $PM_{2.5}$ data. Our investigation on the $PM_{2.5}$ data reliability consists of three analyses: (i) comparing the frequencies of occurrence of various $PM_{2.5}$ concentration ranges and their duration, (ii) checking on the air quality assessment obtained from the U.S. posts and the MEP sites, and (iii) investigating the effect of winter heating in Beijing and Shenyang, the two cities which have centrally controlled heating.

As $PM_{2.5}$ concentration is strongly influenced by meteorological conditions [*Liang et al.*, 2015], we employ a statistical approach (section 2.3) that produces weather-adjusted monthly averages and percentiles of $PM_{2.5}$ concentration at each site. The weather adjustment [*Liang et al.*, 2015] makes the monthly averages and percentiles more comparable over different years, which makes it an important tool for our investigation. Our analysis extends the analyses of *San Martini et al.* [2015] which only considered data from the U.S. posts without the MEP sites and *Stoerk* [2016] which mainly focused on the PM_{10} data in Beijing up to year 2013.

Our study on the data reliability and consistency between the U.S. posts and the MEP sites also provides much needed air quality assessments on the severity and the trend of the air pollution in the major Chinese cities in the 3 years from 2013 to 2015. The five cities are located in the economically most vibrant regions of China, which account collectively for more than 50% of China's Gross Domestic Product (GDP). Hence, assessing the air quality in the five metropolises and understanding their patterns and trends provide highly representative summary on the $PM_{2.5}$ pollution in China.

Our manuscript is organized as follows. Sections 2.1 and 2.2 provide the data information and analyses on descriptive statistics on the percentage and duration of different $PM_{2.5}$ concentration ranges from the two data sources: the U.S. posts and the MEP in the five cities. Section 2.3 outlines the statistical approach for adjusting meteorological conditions so that the statistical estimates for the average and percentile $PM_{2.5}$ concentration are comparable over different years. By employing this weather adjustment approach, we investigate the data reliability and consistency between the U.S. posts and the nearby MEP sites by comparing their monthly averages, quantile concentrations and their yearly changes in sections 3.1 and 3.2. Section 3.3 evaluates the winter heating effects for Beijing and Shenyang. A general air quality assessment based on the weather-adjusted $PM_{2.5}$ concentrations is reported in section 3.4. Section 4 summarizes our findings with added discussions.

2. Data and Descriptive Statistics

2.1. Data Description

We consider $PM_{2.5}$ data from the U.S. posts (http://www.stateair.net/web/historical/1/1.html) and 11 MEP sites (https://wat.epmap.org/) which are geographically close to the U.S. posts in the five cities from January 2013, when MEP started releasing $PM_{2.5}$ data, to December 2015. The U.S. consulate in Shenyang began to collect $PM_{2.5}$ data in April 2013. The locations of the 11 MEP sites and the U.S. posts are marked in Figure 1. Among the 11 MEP sites in the study, the furthest distance to a U.S. post is 8 kilometers (km) for City Station in Guangzhou. The other 10 MEP sites are all within 6 km of a U.S. post. The MEP sites also report hourly readings on other five pollutants (PM_{10}, CO, NO_2, SO_2, and O_3), while the U.S. posts only have $PM_{2.5}$.

Meteorological conditions are known to have significant influences on $PM_{2.5}$ [*Liang et al.*, 2015]. We use weather data recorded at the airports for Beijing (http://weather.nocrew.org/), Shanghai, Chengdu, and Shenyang. For Shanghai, Chengdu, and Shenyang, the data are from https://weatherspark.com/. For Guangzhou, as Baiyun Airport and the U.S. consulate are on different sides of a mountain range (see Figure 1), we use data from a Central Meteorological Agency (CMA) site on the same side of the mountain range with the consulate. The meteorological data contain hourly measures of temperature, pressure, relative humidity, dew point, wind direction and speed, and precipitation.

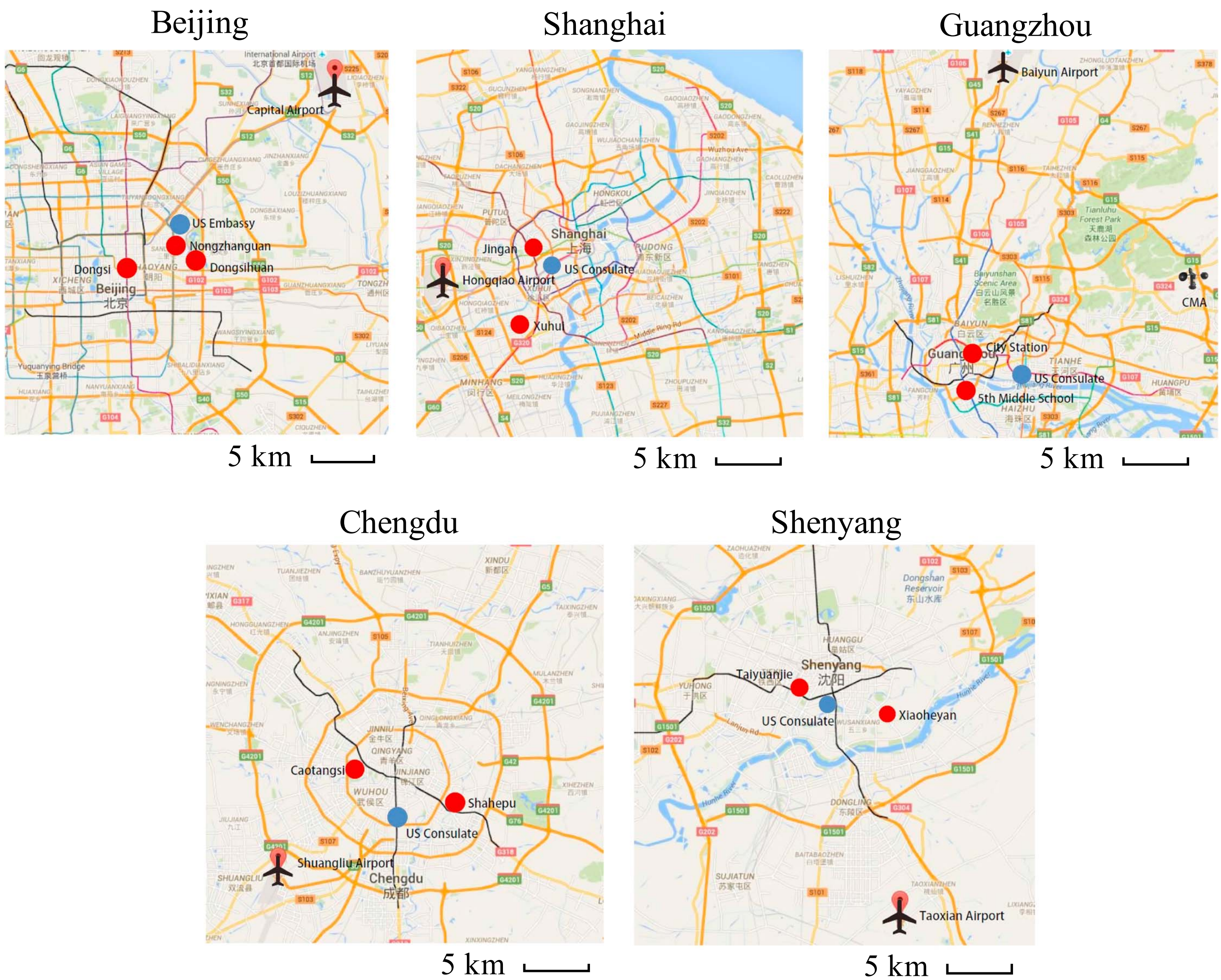

Figure 1. Geographical locations of the U.S. diplomatic posts (blue), China's MEP sites (red), and the meteorological sites in the five cities (black). The locations of the five cities are given in Figure 6.

2.2. Descriptive Statistics

Our first check on data consistency is conducted by comparing frequencies of occurrence of various $PM_{2.5}$ concentration ranges and their duration in hours between the U.S. posts and the nearby MEP sites. China uses 75 µg/m^3 as the highest $PM_{2.5}$ for acceptable air quality, we use 35 µg/m^3 since it is the first interim annual average set by the World Health Organization (WHO) [*World Health Organization*, 2006]. In our study, we define three $PM_{2.5}$ concentration ranges: "$PM_{2.5} \leq 35$ µg/m^3," "$PM_{2.5} > 35$ µg/m^3," and "$PM_{2.5} > 150$ µg/m^3" to indicate the states of clear air, polluted air, and severely polluted air, respectively.

To classify the hourly $PM_{2.5}$ time series into segments of different ranges of concentration, we compute the moving kernel weighted averages [*Härdle*, 1990] with 3 h bandwidth to reduce the variability in the raw $PM_{2.5}$ data and to impute missing values. The original $PM_{2.5}$ data are used in all subsequent analyses. To respect the annual seasonal effects, we use data in multiples of 12 consecutive months when computing the statistics.

Table 1 provides the average duration and time percentages of the three $PM_{2.5}$ ranges of concentration along with the time ranges (in the caption) of data in the multiples of 12 months used in the comparative study. The best agreement between the U.S. post and the MEP sites in the statistics tended to happen at the closest MEP

Table 1. Average Duration (Standard Deviation) in Hours and Percentage of Time Under the Three $PM_{2.5}$ Ranges of Concentration for the U.S. Posts and the MEP Sites in the Five Cities[a]

			Average Duration			Percentage of Time		
City	Site	Distance to U.S. Post	$PM_{2.5} \leq 35$	$PM_{2.5} > 35$	$PM_{2.5} > 150$	$PM_{2.5} \leq 35$	$PM_{2.5} > 35$	$PM_{2.5} > 150$
Beijing	U.S. Post	-	20.0(1.3)	55.6(4.2)	24.5(2.6)	28%	72%	18%
	Nongzhanguan	1.2 km	21.9(1.4)	56.2(3.7)	24.0(2.6)	30%	70%	17%
	Dongsihuan	2 km	18.4(1.2)	53.0(3.7)	23.8(2.6)	28%	72%	19%
	Dongsi	6 km	23.0(1.4)	56.6(3.9)	24.3(2.6)	31%	69%	17%
Shanghai	U.S. Post	-	31.2(3.1)	50.8(4.9)	15.5(2.2)	38%	62%	4%
	Jingan	2.5 km	28.4(2.4)	50.9(4.5)	16.6(2.0)	36%	64%	5%
	Xuhui	5 km	31.0(2.8)	54.2(5.1)	17.0(2.0)	37%	63%	6%
Guangzhou	U.S. Post	-	26.6(3.3)	46.4(6.8)	14.6(2.3)	38%	62%	1%
	5th Middle School	6 km	24.1(2.7)	46.5(7.2)	7.8(1.3)*	35%	65%	0%
	City Station	8 km	31.9(4.2)	42.9(5.6)	10.0(2.3)	43%	57%	0%
Chengdu	U.S. Post	-	12.1(1.2)	92.4(11.8)	22.9(4.3)	12%	88%	9%
	Caotangsi	5 km	17.0(1.6)***	66.5(8.2)	16.6(2.3)	21%	79%	5%
	Shahepu	5 km	18.1(1.6)***	60.7(6.8)***	15.7(1.6)	24%	76%	4%
Shenyang	U.S. Post	-	14.0(0.9)	52.4(5.1)	19.7(1.8)	22%	78%	11%
	Taiyuanjie	2 km	13.6(0.9)	53.5(4.9)	17.0(1.4)	21%	79%	13%
	Xiaoheyan	3 km	16.8(1.1)*	50.0(4.7)	16.6(1.6)	26%	74%	10%

[a]The time ranges are from April 2013 to March 2015 for Beijing, Shanghai and Guangzhou, from July 2013 to June 2015 for Chengdu, from June 2013 to May 2015 for Shenyang. The unit of $PM_{2.5}$ ranges is μg/m^3. The number of * represents the level of significance for testing the equality of duration statistics between the U.S. Posts and the MEP sites.

*$0.025 \leq p$ value < 0.05.

**$0.01 \leq p$ value < 0.025.

***p value < 0.01.

site to the U.S. post. Larger differences from the U.S. post were registered for further away MEP sites. The MEP sites in Guangzhou (within 8 km) were the furthest away from the U.S. post among the five cities, followed by Chengdu with its two MEP sites being 5 km away from the consulate.

We conduct the two-sample t tests for the multiple hypotheses which maintain that the MEP sites have the same average durations for the three $PM_{2.5}$ concentration ranges with the U.S. post in each city. The p values of the tests indicated that among a total of 33 t tests, only five rejected the hypothesis with the p values less than 5%, which are marked in Table 1 by the number of stars. The two MEP sites in Chengdu all had highly (p value $< 1\%$) significantly longer $PM_{2.5} \leq 35$ μg/m^3 duration and one site had highly (p value $< 1\%$) significantly shorter $PM_{2.5} > 35$ μg/m^3 than that of the U.S. consulate. There was one significantly (p value $< 5\%$) shorter $PM_{2.5} > 150$ μg/m^3 period for one MEP site in Guangzhou and one significant longer $PM_{2.5} \leq 35$ μg/m^3 range at a site in Shenyang. We note that conducting the t test for the multiple hypotheses between the MEP sites and the U.S. post in a city can generate false significance (overestimation of the significance), which implies that the number of MEP sites which were significantly different from the U.S. posts may be less than five, the number marked in Table 1. The Bonferroni correction [*Dunn*, 1961] is an approach to account for the multiplicity. However, it can result in underestimation of the significance. Hence, we would only view the multiple t test as a screening tool for detecting the difference between the MEP and the U.S. posts.

In several MEP sites of Beijing and Shanghai, the percentage time under the range $PM_{2.5} \leq 35$ μg/m^3 was shorter and $PM_{2.5} > 150$ μg/m^3 was longer than the corresponding U.S. post. The average time length under $PM_{2.5} > 150$ μg/m^3 from the U.S. posts in Chengdu, Shenyang, and Guangzhou were all longer than those of the corresponding MEP sites. However, Chengdu was the only city that had the statistics from the U.S. post not laid in between the ones from the MEP sites for all the three $PM_{2.5}$ ranges. Apart from Chengdu, we do not see consistent lower MEP readings than the U.S. posts. The significant differences in the duration and the proportions of the $PM_{2.5}$ ranges between the U.S. site and the two MEP sites in Chengdu will be further analyzed in the next section. Given the fact that only 5 out of a total of 33 MEP duration entries in Table 1 were significantly different from the U.S. posts at the 5% level, we would say the statistics were generally agreeable between the two data sources.

2.3. Method for Statistical Adjustment

In addition to comparing the $PM_{2.5}$ range occurrence frequency and duration, we examine the consistency of monthly air quality measures between the U.S. post and MEP sites in a city. The air quality at a site is assessed based on 3 monthly metrics: the average, the median (50th percentile), and the 90th percentile of $PM_{2.5}$ in a month, which are constructed after adjusting for meteorological conditions.

The observed $PM_{2.5}$ is known to be impacted by emission of pollutants, meteorological conditions, and their interaction [*Sun et al.*, 2014; *Liang et al.*, 2015]. Supporting information documents the impacts of wind and precipitation in the five cities; see *Liang et al.* [2015] for evidence of other meteorological variables. Indeed, an increased emission accompanied by favorable weather condition for diffusion of pollutants can result in lower $PM_{2.5}$ readings than a lowered emission regime with unfavorable conditions. This highlights the need to account for the meteorological confounding factors in the observed $PM_{2.5}$, similar to commonly practised in statistical observational studies [*Rosenbaum*, 2002]. Statistical models with corrections for confounding factors have been investigated in the context of air pollution for PM_{10} [*Smith et al.*, 2000]. Some studies incorporated differential equations in the modeling of spatio-temporal processes [*Wikle and Hooten*, 2010], which can be applied to reveal the dynamic nature of the emission behaviours once its mechanism is understood.

We extend the approach proposed in *Liang et al.* [2015] trying to remove the weather influence and produce weather-adjusted $PM_{2.5}$ averages and percentiles, which are used to assess the air quality in the five cities. The purpose of doing meteorological adjustment is to minimize the meteorological effects and to acquire standardized measures to the underlying emission at a site or a city. The adjustment makes the measures comparable over years at each city/site since they are obtained under the standardized weather conditions of the city/site. If the purpose of the study is on the impacts of the air pollution on human health, there is no need to carry out the meteorological adjustment as the amount of human exposure of $PM_{2.5}$ is what matters. However, for assessing the impacts of emission and the effectiveness of air quality management in a city, the city-wide adjusted measures are more objective.

Let Y^s_{ijt} be the $PM_{2.5}$ reading at hour t of a site s in month j and year i in a city, and $X_{ijt} = (P_{ijt}, T_{ijt}, D_{ijt}, C_{ijt}, R_{ijt})$ be the continuous meteorological variables consisting of air pressure (P), dew point (D), temperature (T), cumulative wind power (C), and cumulative precipitation (R). We also grouped the 16 wind directions into the discrete categorical variable W_{ijt}, which contains five levels. The cumulative wind power (C) is the cumulative wind speed over time under a wind direction, and the cumulative precipitation (R) is similarly defined. See SI for the specifics of these variables. It is noted that airmass trajectories can be used to analyze the atmospheric transport of pollutants [*Kulshrestha and Kumar*, 2014]. Although local wind may not fully capture the overall transportation, it can still significantly impact the local $PM_{2.5}$ concentration. We document this issue in the SI. Vertical mixing is another factor that affects the $PM_{2.5}$ concentration. However, unlike the six meteorological variables we have chosen in the study, re-assimilated boundary layer height (BLH) data are only available four times a day over quite sparsely distributed grid points. As BLH is associated with atmospheric temperature, pressure, and relative humidity [*Von Engeln and Teixeira*, 2013], the vertically mixing is partially reflected in the study. As we only employ the six meteorological variables due to data limitation, the meteorological adjustment can only reduce the meteorological confounding effects caused by these variables.

We consider the following nonparametric regression model for a site s in a city:

$$Y^s_{ijt} = m^s_{ij}(X_{ijt}, W_{ijt}) + e^s_{ijt}, \quad t = 1, \dots, n_{ij}, \tag{1}$$

where $m^s_{ij}(x, w) = E(Y^s_{ijt}|X_{ijt} = x, W_{ijt} = w)$ is the regression function and $\{e^s_{ijt}\}$ are the errors which are assumed to be stationary and weakly dependent. Here we use different regression functions to capture the fixed effect in different sites. For the purpose of air quality assessment, which is retrospective, a detailed parametric model for the regression function can be avoided and the above nonparametric model is sufficient, given that we have quite numbers of data over several monitoring sites. The model also offers robustness against potential model mis-specification with parametric models. We note that model (1) does not involve the emission as the latter is only observed over much coarser spatial grids at much lower frequency. However, the changes in the underlying emission may be recovered by the changes of the adjusted average concentration provided all the meteorological influences could be accounted for, as shown in *Liang et al.* [2015].

The quantities we rely on for air quality assessment are the monthly average μ_{ij}^s, the median $\xi_{ij}^s(0.5)$, and the 90th percentile $\xi_{ij}^s(0.9)$ of $PM_{2.5}$ concentration for a site s at month j of year i. In particular, the 90th percentile quantifies the top 10% severe pollution. Let us illustrate how to obtain the weather-adjusted monthly average μ_{ij}^s, and those for the median and the 90th percentile are given in the SI.

The key steps of the meteorological adjustment to the monthly averages are the following: (1) estimate the nonparametric regressions $m_{ij}^s(x, w)$ of $PM_{2.5}$ on the meteorological variables X and W at site s for each month j of year i and denote the estimate as $\hat{m}_{ij}^s(x, w)$; (2) substitute the meteorological data for (X, W) of month j of all years to the estimated $\hat{m}_{ij}^s(x, w)$ and sum them over and divide by the total number of observations in month j, and this gives the adjusted average for month j of year i at site s; and (3) employ the block bootstrap method to evaluate the variation of the adjusted estimate to preserve the spatial and temporal dependence.

Let us provide the rationale of the meteorological adjustment. Suppose that there are n years of meteorological data. Let $f_{aj}(x, w)$ be the probability density function of (X, W) for month j of year a, and

$$f_j(x, w) = n^{-1} \sum_{a=1}^{n} f_{aj}(x, w) \tag{2}$$

be the average (standard) meteorological condition for month j, representing the equilibrium state of the meteorological condition for the month. In this paper, we have meteorological data for each city from year 2010 to year 2015. So the standard weather condition is constructed based $n = 6$ years' data. To take into account the weather confounding effects, we adapt $\mu_{ij}^s = \int m_{ij}^s(X_{ijt}, W_{ijt}) f_j(x, w)\mathrm{d}x\mathrm{d}w$ where $f_j(x, w)$ is used rather than $f_{ij}(x, w)$. Substitute (2), we get

$$\mu_{ij}^s = n^{-1} \sum_{a=1}^{n} \int m_{ij}^s(x, w) f_{aj}(x, w)\mathrm{d}x\mathrm{d}w. \tag{3}$$

The above expression differs from

$$\int m_{ij}^s(x, w) f_{ij}(x, w)\mathrm{d}x\mathrm{d}w, \tag{4}$$

which uses only meteorological information in month j of year i. The approach in (4) leads to using the average of raw $PM_{2.5}$, which is confounded by the weather and is commonly used in air quality assessment in some countries including China. So the merit of (3) is that it includes the $n-1$ terms $\int m_{ij}^s(x, w) f_{aj}(x, w)\mathrm{d}x\mathrm{d}w$ for $a \neq i$, which are the averages by applying the $PM_{2.5}$ function of month j, year i on the weather conditions of the other years. Since they are not directly observable in reality and are thus called counterfactual averages, borrowing a terminology from the casual inference.

It is often the case that the precipitation $R_{ijt} = 0$ with high probability. The probability exceeds 0.8 even in the three southern cities Shanghai, Guangzhou, and Chengdu, which have more rainfall than Beijing and Shenyang. By defining $X_{ijt,1} = (P_{ijt}, T_{ijt}, D_{ijt}, C_{ijt})$ and $X_{ijt,2} = R_{ijt}$, the regression function $m_{ij}^s(x, w)$ can written as

$$m_{ij}^s(x, w) = m_{ij0}^s(x_1, 0, w) I(x_2 = 0) + m_{ij+}^s(x_1, x_2, w) I(x_2 > 0), \tag{5}$$

where $I(A)$ is the indicator function that assumes value 1 when the event A is satisfied and 0 otherwise.

To estimate $m_{ij}^s(x, w)$, we first estimate $m_{ij0}^s(x_1, 0, w)$ under each wind direction using data with zero precipitation. With the Nadaraya-Watson kernel estimator [*Härdle*, 1990; *Fan and Gijbels*, 1996], we estimate $m_{ij0}^s(x_1, 0, w)$ by

$$\hat{m}_{ij0}^s(x_1, 0, w) = \frac{\sum_{t=1}^{n_{ij}} K\left(\frac{X_{ijt,1}-x_1}{h_1}\right) I(X_{ijt,2} = 0, W_{ijt} = w) Y_{ijt}^s}{\sum_{t=1}^{n_{ij}} K\left(\frac{X_{ijt,1}-x_1}{h_1}\right) I(X_{ijt,2} = 0, W_{ijt} = w)}, \tag{6}$$

where $K(\cdot)$ is a product of the univariate Gaussian kernel function and h_1 is the smoothing bandwidth.

To estimate $m_{ij+}^s(x_1, x_2, w)$ under non-zero precipitation, we pool all the wind directions together to increase the sample size. We consider a version of the kernel estimator that constructs an extra smoothing over the wind directions using a discrete kernel [*Aitchison and Aitken*, 1976; *Chen and Tang*, 2011]

$$\hat{m}_{ij+}^s(x_1, x_2, w) = \frac{\sum_{t=1}^{n_{ij}} K_1\left(\frac{X_{ijt,1}-x_1}{h_1}\right) K_2\left(\frac{X_{ijt,2}-x_2}{h_2}\right) l(W_{ijt}, w, \lambda) I(X_{ijt,2} > 0) Y_{ijt}^s}{\sum_{t=1}^{n_{ij}} K_1\left(\frac{X_{ijt,1}-x_1}{h_1}\right) K_2\left(\frac{X_{ijt,2}-x_2}{h_2}\right) l(W_{ijt}, w, \lambda) I(X_{ijt,2} > 0)}, \tag{7}$$

where $K_1(\cdot)$ is a product of the univariate Gaussian kernel function, $K_2(\cdot)$ is the univariate Gaussian kernel function, and $l(W_{ijt}, w, \lambda)$ is a kernel for the discrete variable: $l(W_{ijt}, w, \lambda) = 1$ if $W_{ijt} = w$, $l(W_{ijt}, w, \lambda) = \lambda$ otherwise and λ is the smoothing parameter. The smoothing bandwidths are h_1, h_2, and λ. We use the cross-validation method to choose h_1, h_2, and λ. Then, substitute (6) and (7) into (5) to get $\hat{m}^s_{ij}(x, w)$, the estimator of $m^s_{ij}(x, w)$.

By substituting all meteorological data to $\hat{m}^s_{ij}(x, w)$, we obtain an empirical version of (3), which is the weather-adjusted estimator for the mean concentration of $PM_{2.5}$ at site s

$$\hat{\mu}^s_{ij} = \left(\sum_{a=1}^{n} n_{aj} \right)^{-1} \sum_{a=1}^{n} \sum_{t=1}^{n_{aj}} \hat{m}^s_{ij}(X_{ajt}, W_{ajt}). \tag{8}$$

Weather-adjusted percentiles $\hat{\xi}^s_{ij}(0.5)$ and $\hat{\xi}^s_{ij}(0.9)$ can be obtained by inverting the weather-adjusted $PM_{2.5}$ distribution function estimates as outlined in the SI.

To gauge on the variation in $\hat{\mu}^s_{ij}$ and the percentiles we employ the block bootstrap method [*Davison and Hinkley*, 1997] to generate large number of pseudo-samples for both the meteorological and $PM_{2.5}$ data. The data blocking with respect to time is to preserve the temporal dependence in the data. At the same time, $PM_{2.5}$ data at different sites in a city are re-sampled jointly to maintain the spatial dependence. See section S3.3 in the SI for details.

3. Consistency in Air Quality Assessment

We perform weather-adjusted monthly estimation for three metrics (average, median, and 90th percentile) to quantify the characteristics of the $PM_{2.5}$ distribution for all sites in the five cities. The adjustment evens out most of the meteorological variation over different years and hence makes the estimates for the three characteristics more comparable over the years. It is noted that the 90th percentile describes the worst 10% level of the pollution. As more than 90% of $PM_{2.5}$ observations were missing from the MEP sites in Guangzhou in July, August, October, and November of 2013, we did not assess the MEP sites for the 4 months. As the U.S. consulate in Shenyang started collecting $PM_{2.5}$ on 22 April 2013, we analyzed the U.S. post data from May 2013.

3.1. Comparing Mean Concentration

It is observed from Figure 2 that the averages at the U.S. posts were not necessarily higher than those at the nearby MEP sites. Often the MEP sites had slightly higher averages, for instance, in February 2015 in Beijing and summer of 2014 in Shanghai. However, there were noticeable differences between the U.S. post and the two MEP sites in Chengdu since the summer of 2014, which confirms the discrepancy in the duration statistics reported in Table 1.

Figure 2 displays the adjusted means for the U.S. posts and the MEP sites in the five cities. The most striking feature of the figure is the high synchronization in the monthly average $PM_{2.5}$ among all the sites in each city. Additional figures given in the SI also show similar synchronization in the median and the 90th percentile as well. Dongsihuan in Beijing had more than 85% missing data since October 2015, which caused large difference with other sites in Beijing. Except these months in Dongsihuan, the averages among the sites in a city were largely comparable to each other in general and between the U.S. post and the MEP sites in particular.

Table 2 reports the Pearson's correlation coefficients in the adjusted means, medians, and 90th percentiles between the U.S. posts and the averages of the MEP sites, respectively, based on 36 monthly values of the three metrics. The correlation coefficients were overwhelmingly high. The lowest two correlations were 0.71 and 0.72 happened in the medians and 90th percentiles for Guangzhou. All others were above 0.92, with those of Shanghai, Chengdu and Shenyang all above 0.95. The high correlation between the U.S. post and the MEP sites in Chengdu indicates that despite the significant difference in the frequencies of occurrence of $PM_{2.5}$ concentration ranges and their duration in Table 1 and monthly adjusted means in Figure 2, both data streams reacted to the overall $PM_{2.5}$ concentrations resulting from emissions, sinks, and dilution, as well as highly correlated meteorological conditions of the city in a highly compatible manner. We also report the slopes of linearly regressing the MEP adjusted values versus their U.S. counterparts in Table S6 of the SI, which were consistent with the magnitude of correlation coefficients reported in Table 2.

3.2. Compatibility in Yearly Changes

Some differences between the U.S. posts and the MEP sites in Figure 2 may be attributed to the so-called fixed effect. The fixed effect summarizes the statistical effect due to unique local factors to a site, for instance,

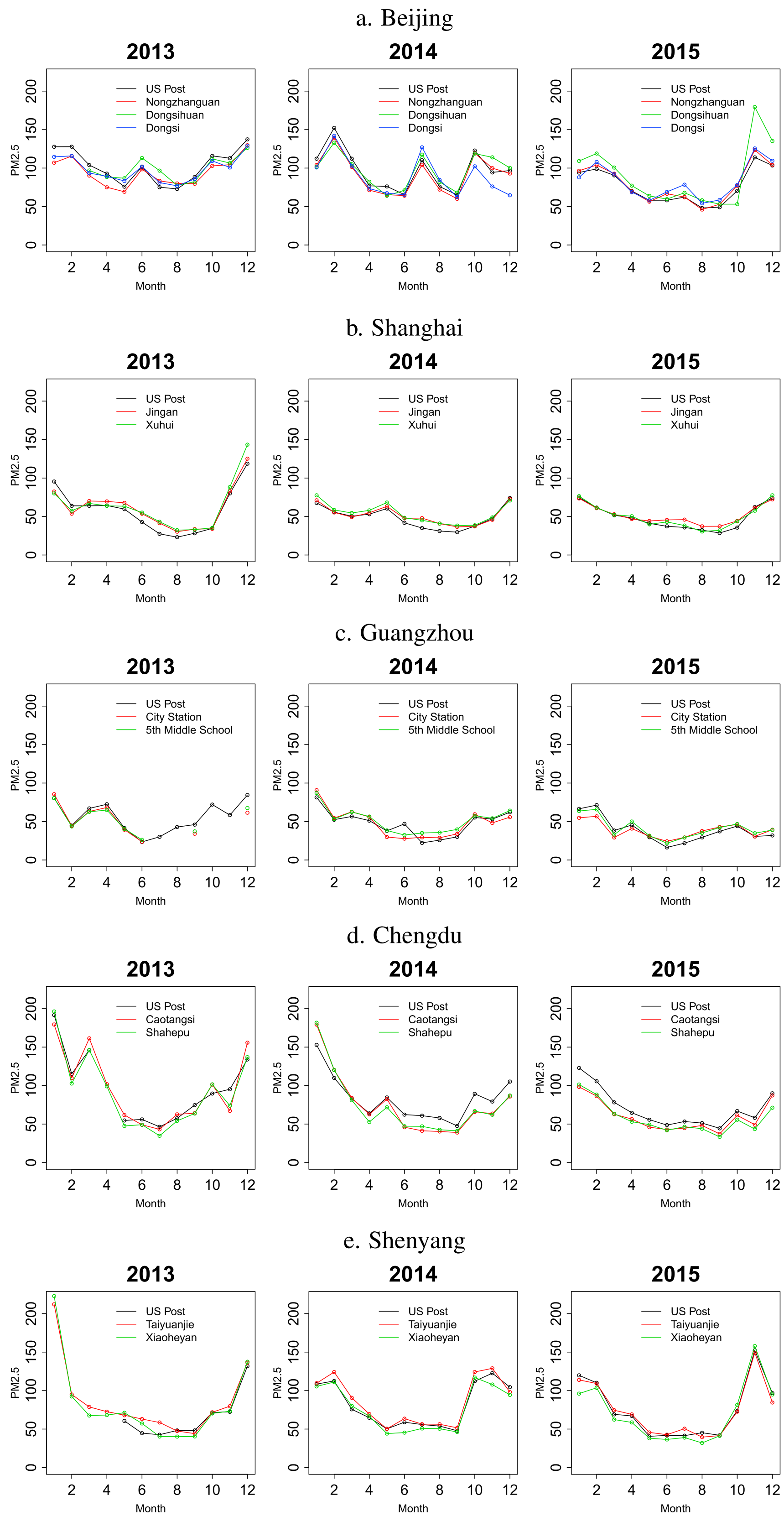

Figure 2. Meteorologically adjusted averages in the U.S. posts and the MEP sites in the five cities from January 2013 to December 2015.

Table 2. Pearson's Correlation Coefficients in the Three Air Quality Metrics Between the U.S. Posts and the Averages of the Nearby MEP Sites[a]

Metrics	Beijing	Shanghai	Guangzhou	Chengdu	Shenhang
Mean	0.96 (0.06)	0.96 (0.05)	0.92 (0.07)	0.96 (0.05)	0.97 (0.04)
Median	0.93 (0.05)	0.95 (0.05)	0.71 (0.14)	0.96 (0.05)	0.97 (0.05)
90th Percentile	0.97 (0.04)	0.97 (0.04)	0.72 (0.14)	0.95 (0.05)	0.97 (0.04)

[a]The figures in the parentheses are the standard errors.

different local emission exposures, distance to the traffic or emission sources, and the configurations of the $PM_{2.5}$ measuring apparatus. We model the fixed effect nonparametrically via the regression function $m_{ij}^{s}(x)$ which is specific for site *s*. Hence, there is no need to add a fixed effect term as commonly adapted in parametric modeling. As the targeted local emission is slowly varying over time, taking the yearly differences (yearly changes) in the adjusted means, medians, and percentiles would largely remove the fixed effects.

We compute the yearly changes for each month in the three metrics, which results in 36 pair-wise yearly changes for each metric in each city. The yearly changes provide another check on the data consistency between the MEP and U.S. post. Figure 3 displays the yearly differences together with the 95% confidence intervals in the averages for the U.S. post and the MEP sites (average of the MEP sites in a city is used), respectively, and those for the medians and the 90th percentiles are given in the SI. A noticeable aspect of Figure 3 is that the discrepancies between the U.S. post and MEP sites in Chengdu shown in Table 1 and Figure 2 are largely disappeared. This confirmed the existence of the fixed effect mentioned above.

Like Figure 2, Figure 3 shows a consistent pattern in the yearly changes between the U.S. post and the MEP sites. To better summarize the results of Figure 3, we say the yearly change from the U.S. post to be "agreeable" with the corresponding yearly change from the MEP sites if they reach the same statistical decision (significant/insignificant increase/decrease) at 5% level. The yearly changes from the two data sources are said to be "close" if one significant increase (decrease) is matched with one insignificant increase (decrease), or insignificant increase is matched with insignificant decrease. The yearly changes are classified as "sharply different" if they are neither "agreeable" nor "close", for instance, one significant decrease is matched with either a insignificant or significant increase. Among the 36 pairs of yearly changes in the mean $PM_{2.5}$ concentration for Beijing, there were 24 being "agreeable" and 10 being "close". Only two yearly changes were "sharply different": one in May between 2014 and 2013 when the MEP projected a significant decrease, while the U.S. post projected an insignificant increase; and one in March between 2015 and 2013, when the U.S. post asserted a significant decrease but the MEP indicated an insignificant increase. The results for all the five cities showed similar pattern and are given in Table 3.

Different from the adjusted metrics, the yearly changes can be a proxy of the changes of local emissions. See the discussion in section 2.3. High correlations of yearly changes between MEP sites and U.S. posts indicate similar patterns of change in the emission. It is fair to say that the yearly changes based on the MEP sites were not necessarily rosier than those based on the U.S. posts. In fact the opposite was often the case. We note in particular a very high agreement in the air quality assessment between the two data sources in Beijing and Shanghai, the two leading metropolises in China. Table S5 in the SI reports the Pearson's correlation coefficients in the yearly changes in the three monthly air quality metrics between the U.S. post and the MEP sites. Although the correlations were lower after taking the yearly changes, the level of the dependence was still quite high. For instance, they were still well above 0.85 for Beijing, Shanghai, and Shenyang. Chengdu's correlations for the yearly changes stayed above 0.68 and those in Guangzhou above 0.69. The results for Chengdu based on the yearly changes are also supportive to consistency between the U.S. post and the MEP sites. It suggests that the difference between the U.S. post and the MEP sites in Chengdu was likely caused by the local fixed effects. The slopes of linearly regressing the MEP yearly changes versus their U.S. counterparts in Table S6 of the SI were also consistent with the correlation coefficients reported in Table S5.

3.3. Comparing Winter Heating Effect

Centralized winter heating is provided in the northern cities in China, which include Beijing and Shenyang. The winter heating season runs from 1 November to 31 March in Shenyang and 15 November to 15 March in Beijing. The heating periods can be extended if low temperature is encountered before (after) the starting

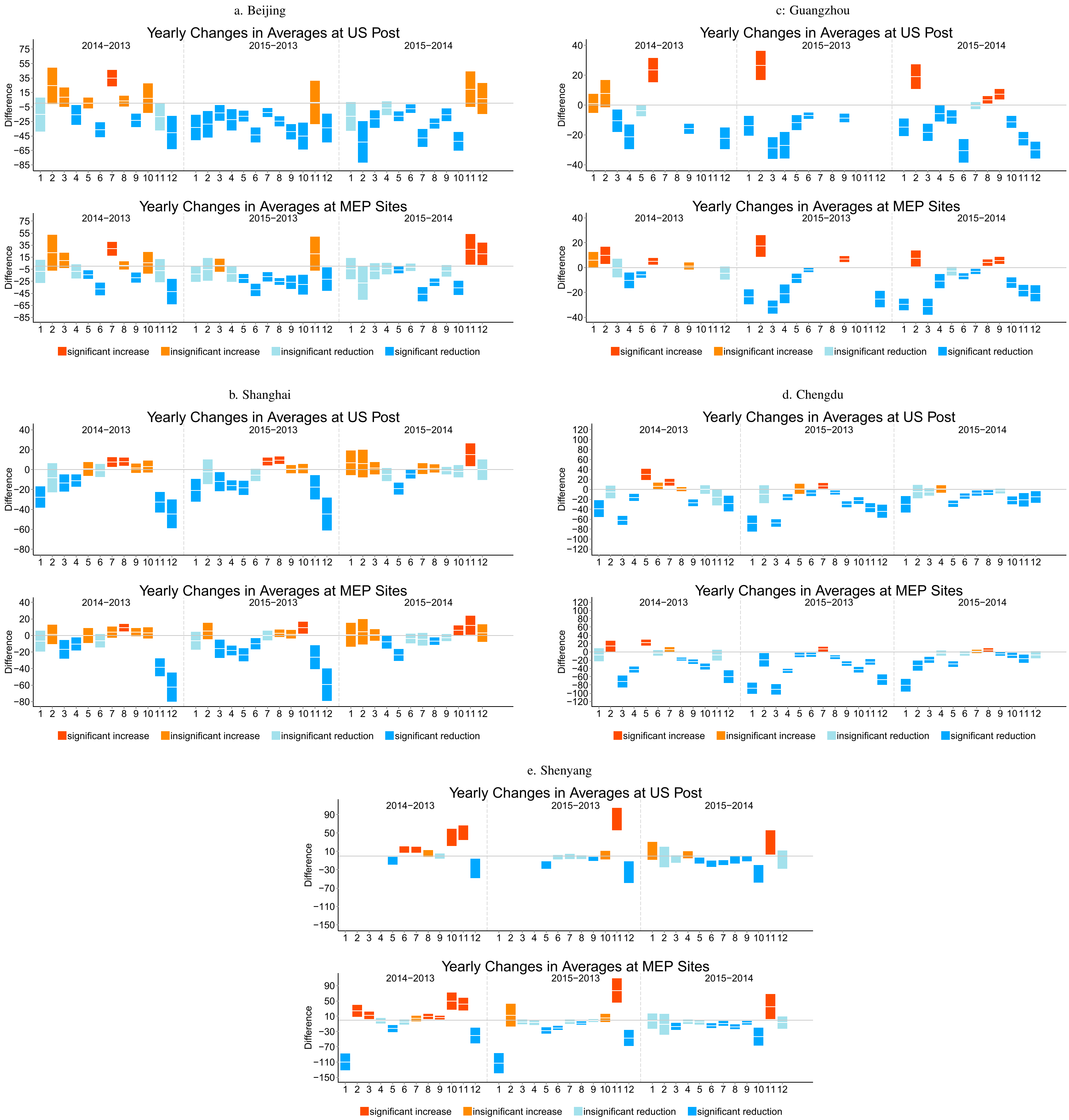

Figure 3. Yearly changes in the meteorologically adjusted means in the U.S. posts and the average of the MEP sites in five cities. The boxes are 1.96 times the standard deviations above and below the differences (white line). Significant increases (decreases) correspond to boxes which are entirely above (below) the horizontal line of zero, and insignificant increases (decreases) correspond to boxes which intercept the horizontal line. No estimates for the MEP sites in Guangzhou in July, August, October, and November of 2013 and the Shenyang's U.S. post from January to April 2013 due to severe missing values.

Table 3. Number of Months With Yearly Changes in the Monthly Average $PM_{2.5}$ Which Lead to the "Agreeable," "Close," and "Sharply Different" Statistical Decisions Between the U.S. Posts and MEP Sites[a]

	Beijing	Shanghai	Guangzhou	Chengdu	Shenhang
Total	36	36	28	36	28
"Agreeable"	24	24	20	20	16
"Close"	10	12	6	12	10
"Sharply different"	2	0	2	4	2

[a] See text for definitions of the categories.

date (ending date). The fuel used to generate the heating is predominantly coal in North China, although there has been a gradual transition to the natural gas in Beijing in recent years.

The winter heating event can be regarded as a quasi-experiment [*Dominici et al.*, 2014] that allows measurement of the heating effect in the two cities. We use winter heating, a regional emission phenomenon, to check if comparable heating effects can be obtained between the U.S. posts and the nearby MEP sites in Beijing and Shenyang.

We define non-heating and heating periods in November (March), respectively, as the two weeks before (after) and after (before) the start (end) of a heating season, respectively. We use the weather conditions of the four weeks that cover the non-heating and heating periods, respectively, over all the years of observations as the baseline. By regressing $PM_{2.5}$ on the meteorological variables separately over the two periods, we obtain the weather-adjusted averages of $PM_{2.5}$ in the heating and non-heating periods, respectively.

Figure 4 displays the original and the weather-adjusted means for the heating and non-heating periods at the U.S. posts and the MEP sites in Beijing and Shenyang. The heating effect was statistically significant at 5% level at both the start (November) and the end (March) of the heating season in Beijing for all the sites (U.S. and MEP) except at Dongsihuan in November 2013 (the p value was 0.186). For Shenyang, we did not find significant heating effect in November 2013 and 2014 for both the U.S. post and the MEP sites but found the heating was significant (at 5% level) in March for all sites in all 3 years. The likely cause for the insignificance November heating effect in Shenyang is the confounding effects with biomass burning over the non-heating period (October 18–31). Biomass burning of farm residuals is at its peak in October over a vast area of China [*Zheng et al.*, 2005] including the Northeast where Shenyang is situated. For Beijing, the non-heating control period is November 1–14, which narrowly misses the biomass burning.

Figure 4 shows highly consistent heating effects between the U.S. post and the MEP sites in Beijing and Shenyang. The amount of heating effect (the difference in the averages of the $PM_{2.5}$ concentration between the heating and non-heating periods) was highly comparable among all the sites in the two cities. This provides another support on the consistency of the $PM_{2.5}$ data among the U.S. posts and MEP sites in the two cities.

Our calculation shows that on average, the $PM_{2.5}$ concentration in the heating period was about 87.9% (standard deviation: 68.0%) higher than that in the non-heating period in Beijing, based on the analyses for November (2013–2015) and March (2013–2015). The Asia-Pacific Economic Cooperation (APEC) meeting was held on November 8–10 in Beijing in the non-heating period. The government implemented rather aggressive measures to suppress emission before and during the APEC meeting, which would largely inflate the November heating effect in 2014. Therefore, if we exclude November 2014 of the four sites, the average heating effect (standard deviation) was lower to 75.2% (68.1%). The large standard deviations were due to more than 200% heating effect in November 2015. The cause for this large increase is unknown and needs further investigation. If we avoid the extreme high heating effect in November 2014–2015, we get 47.5% (18.9%) average heating effect, which was close to the average heating effect using only the U.S. Embassy data from 2010 to 2014 given in *Liang et al.* [2015]. Similar calculation shows that the heating effect in Shenyang for March (2013–2015) was 37% (15.6%), and we did not compute the November heating effect for Shenyang due to the confounding effects with the biomass burning.

Suggested by a referee, we also check if there are some "heating effect" for the three southern cities although centralized winter heating is not provided. We use 15 November and 15 March as the hypothetical beginning

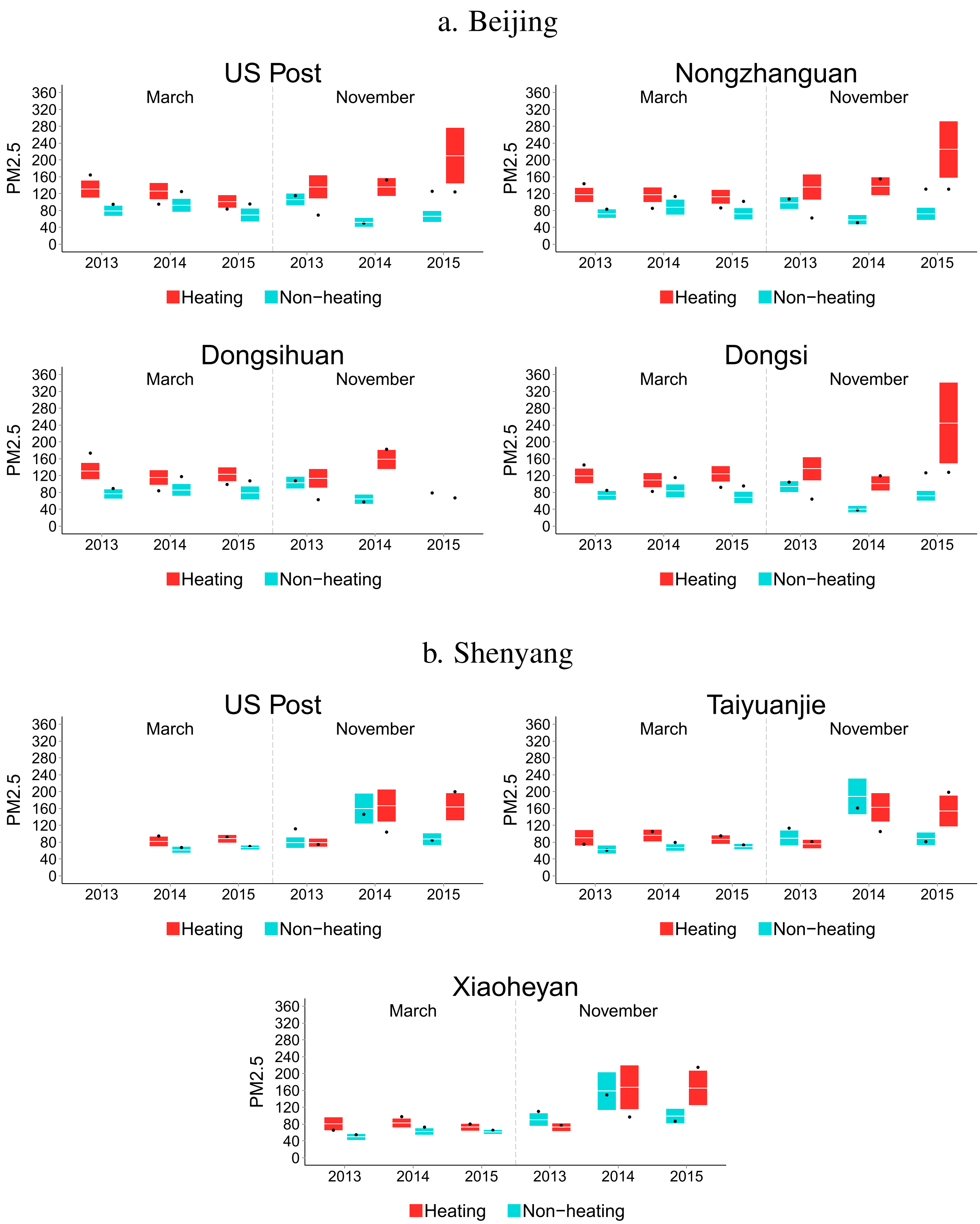

Figure 4. Heating effect: meteorologically adjusted averages in the non-heating and heating periods in November and March in Beijing and Shenyang. The black dots are the averages without the adjustment. The data of the U.S. consulate in Shenyang were missing before 22 April 2013.

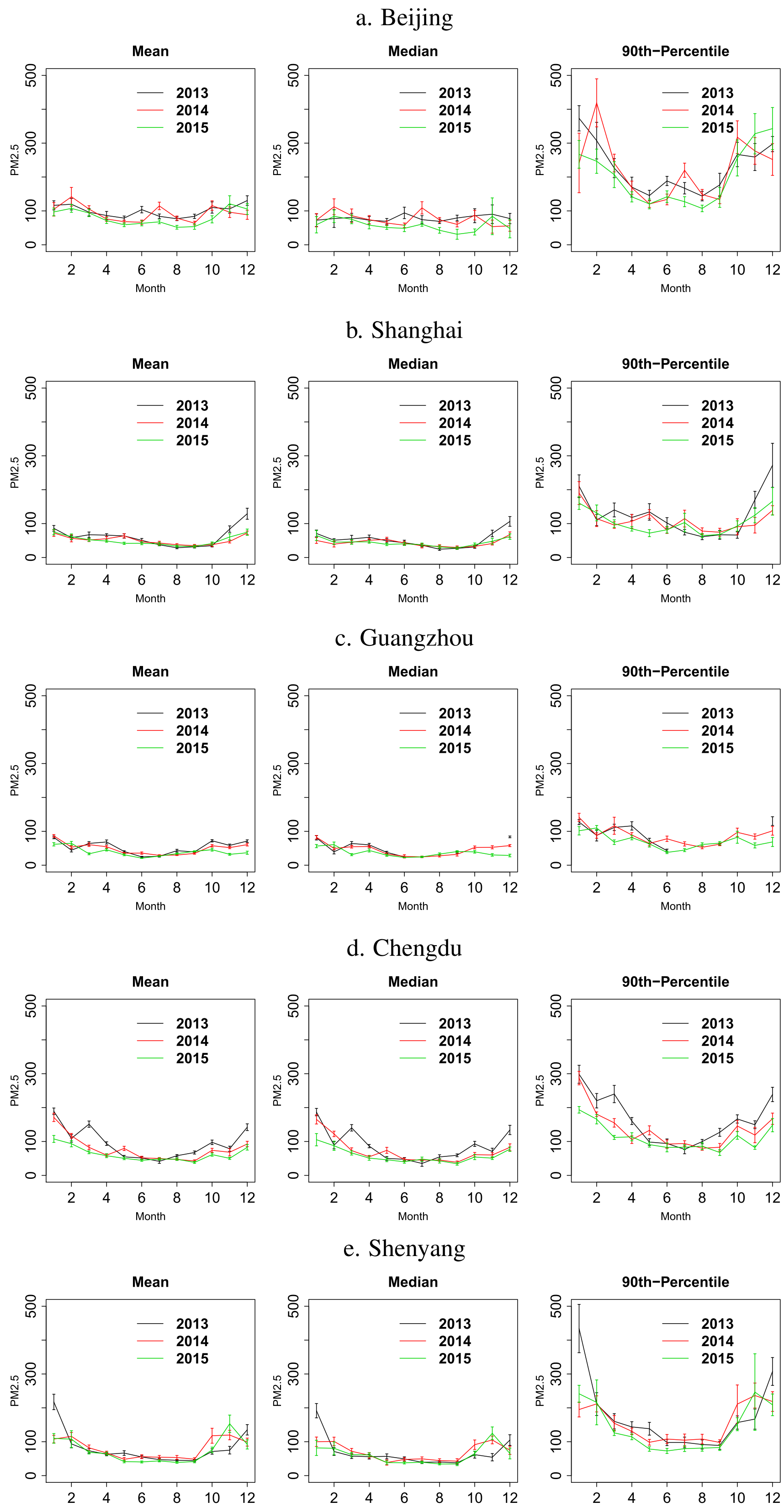

Figure 5. Comparison of the monthly average of the meteorologically adjusted averages, medians, and 90th percentiles in the five cities in different years. The bars are 1.96 times the standard deviations above and below the estimates.

Table 4. Annual Averages of the Meteorologically Adjusted Averages, Medians, and 90th Percentiles in the Five Cities From 2013 to 2015[a]

City	Year	Mean	Median	90th Percentile
Beijing	2013	99.4(1.71)	78.6(2.45)	226.8(4.55)
	2014	93.4(1.98)*	75.5(2.26)	222.7(6.16)
	2015	80.6(1.7)***	56.7(3.14)***	202.1(5.47)**
Shanghai	2013	61.4(1.09)	52.1(1.15)	127.2(3.80)
	2014	51.5(0.90)***	43.3(0.90)***	109.0(2.87)**
	2015	50.3(0.88)	43.7(2.45)	104.1(2.87)
Guangzhou	2013	54.3(0.73)	53.5(2.59)	93.4(3.42)
	2014	49.1(0.51)***	45.6(1.95)**	86.4(2.68)
	2015	39.4(0.50)***	36.7(1.75)***	70.1(2.42)***
Chengdu	2013	94.6(1.06)	87.2(1.25)	164.2(2.44)
	2014	78.0(1.00)***	72.5(1.09)***	136.9(2.05)***
	2015	62.6(0.84)***	58.7(1.24)***	112.0(1.62)***
Shenyang	2013	82.3(1.50)	68.5(1.52)	174.8(4.44)
	2014	80.8(1.54)	69.0(1.44)	156.3(3.78)*
	2015	73.2(1.81)***	60.5(1.74)***	141.7(5.96)*

[a]The figures in the parentheses are the standard errors obtained by taking the square root of the average variance over the 12 months period. The number of * represents the level of significance for testing if there is a decrease from the same period in the previous year.

*$0.025 \leq p$ value < 0.05.

**$0.01 \leq p$ value < 0.025.

***p value < 0.01.

and ending dates of "heating" for the three southern cities, as 15 November to 15 March (same with Beijing) represents the shortest heating season for cities north of Huai River which is the southern boundary of the centrally supplied domestic heating region in China. For each heating period, we adopt the same study design and the testing method as that for Beijing. Table S7 in the SI provides details on the analysis and Table S8 summarizes the testing results for the three cities along with Beijing and Shenyang. These tables show that it is inconclusive to confirm the heating effect for the three southern cities as the amount of evidence was not as consistent as those for Beijing and Shenyang.

3.4. Five Cities Assessment

Given the data consistency between the U.S. posts and the MEP sites in the five cities, we first average the fitted regression and distribution functions among all sites (U.S. and MEP) in a city and then carry out the meteorological adjustment to get city-wide averages, medians, and 90th percentiles. See section S3 in the SI for technical details.

Figure 5 shows the city-wide monthly averages, medians, and 90th percentiles for the five cities, along with their 95% confidence intervals. Both the adjusted averages and the 90th percentiles displayed in the figure show that $PM_{2.5}$ tends to be the highest in the winter months (November to February) for each city. The winter increase was most notable in the 90th percentiles. This was expected for Beijing and Shenyang due to the winter heating. However, there was quite substantial winter effect for the other three cities as well. The increase in November and December 2013 in Shanghai demonstrated that the $PM_{2.5}$ can reach a very high level there as well. The monthly medians exhibited much less variation than the averages and the 90th percentiles. This is not surprising as the median is known to be a robust statistical measure.

Table 4 composes the yearly averages of the three metrics for the five cities since 2013. It shows that in terms of the average and extreme pollution (90th percentile), Beijing was the worst, followed by Shenyang and Chengdu. Comparing the latter two cities, Shenyang had higher 90th percentiles, while Chengdu had worse averages.

Guangzhou had the best air quality among the five cities, and Shanghai ranked the second best. Guangzhou was the only city that had the 90th percentile below 100 μg/m^3, while the worst three cities (Beijing, Shenyang, and Chengdu) had the 90th percentiles well above 130 μg/m^3 except Chengdu in 2015, and Beijing's were

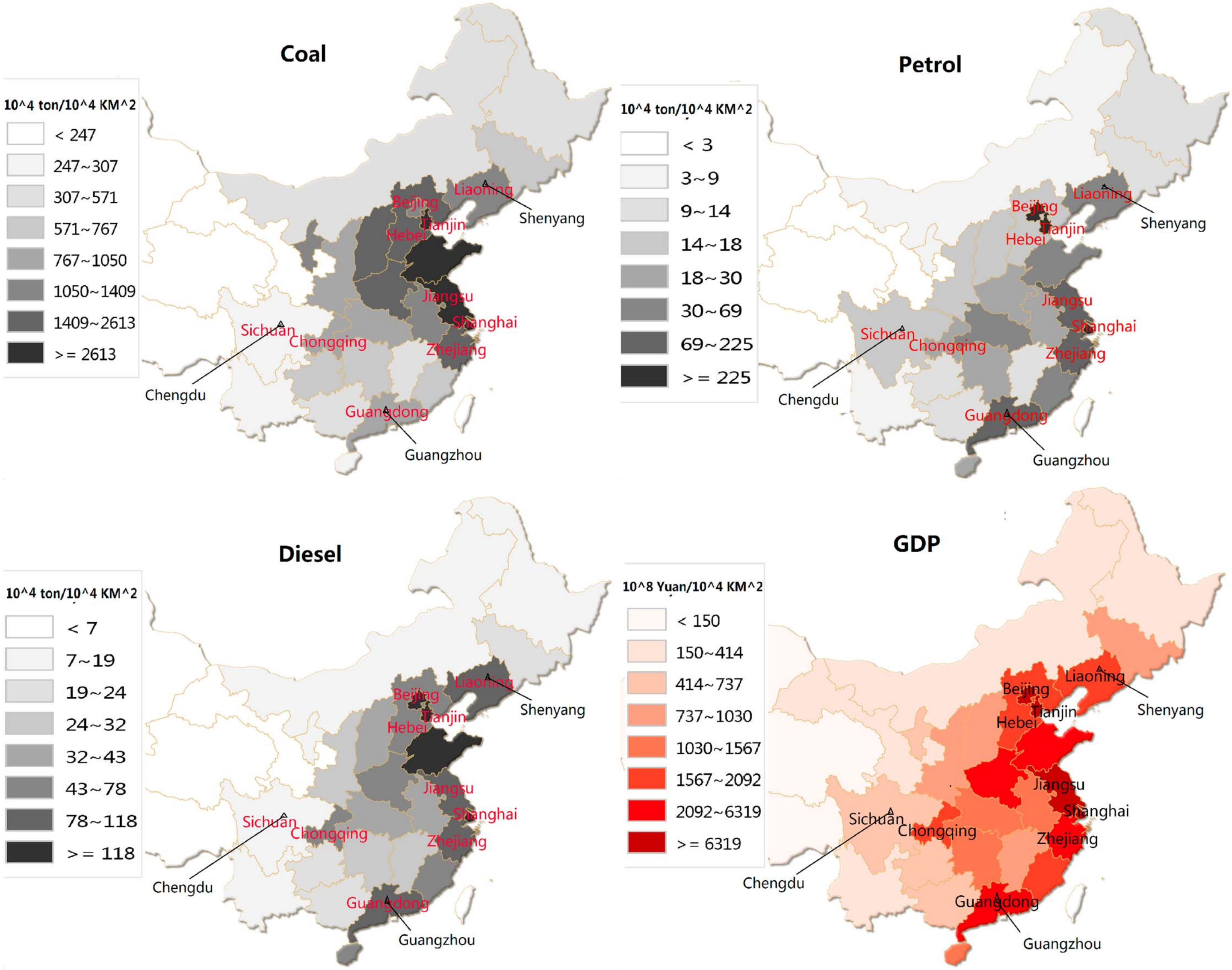

Figure 6. Intensities of energy (coal, petrol, and diesel) consumption and Gross Domestic Product (GDP) of the five regions where the five cities are located, embedded within the intensity profiles of other provinces in China. The intensities are standardized consumption and GDP amount by 10,000 km^2 of land area. The energy data were from year 2012, and the GDP data were the 2014 data.

more than 200 μg/m^3. The annual averages in Guangzhou and Shanghai were mostly in the 40–60 μg/m^3 range and those for Beijing, Chengdu, and Shenyang were in the 70–100 μg/m^3 range. These were all well above WHO's first interim target of 35 μg/m^3 for annual mean concentration [*World Health Organization*, 2006]. It is noted that WHO's ultimate annual air quality guideline is 10 μg/m^3.

Table 4 suggests that there have been some reductions in the three metrics in the five cities since 2014, as indicated by the level of statistical significance marked in the table by comparing the 12 pairs of monthly averages in the two neighboring years. It was clear that $PM_{2.5}$ pollution in 2013 was the worst for most metrics in the five cities except annual median in Shenyang. The situation of the five cities appeared to have improved in 2014, with further significant improvement in 2015 except Shanghai. Chengdu had statistically significant yearly improvement in all the three metrics since 2014, followed by Guangzhou, Beijng, and Shenyang. Shanghai had no improvement in 2015, but the decrease from 2013 to 2014 was notable.

To gain insight on the relative $PM_{2.5}$ concentration in the five cities, we analyze the energy consumption for the five regions where the five cities are situated. According to Figure S10 in the SI, Yangtze River Delta (YRD) region that encompasses Shanghai was the top consumer of energy, while Liaoning (Shenyang) and

Sichuan-Chongqing (Chengdu) regions were the least users. Figure S10, however, may be misleading as it does not take into account of the land area of the regions. We then compute two energy intensity measures: one standardized by the land area and the other by the total GDP of the regions.

Figure 6 displays the energy intensity for coal, petrol, and diesel of the five regions standardized by land area (in 10,000 km^2) and the detailed numbers are in Table S10 of supporting information. The figure shows YRD (Shanghai) had the highest intensity in all three forms of energy, and Sichuan-Chongqing (Chengdu) had the least intensity. NCP (Beijing) had the second highest intensity in coal.

If standardized by GDP (detailed numbers in Table S11 of SI), YRD (Shanghai)'s energy intensity was the lowest in diesel, and the second lowest in coal and petrol, and Liaoning (Shenyang) had the highest intensity in coal, diesel, and petrol, and Sichuan-Chongqing region (Chengdu) moved up to the third place in coal and the second places in diesel and petrol. These figures showed that YRD (Shanghai) had the most energy efficient economy, followed by Pearl River Delta (PRD) (Guangzhou). Liaoning, NCP, and Sichuan-Chongqing were lagged behind.

Above discussion on the energy consumption since 2002 reveals that the YRD (Shanghai) was the largest energy user in the five regions, and yet Shanghai had the second best air quality among the five cities. In contrast, Sichuan-Chongqing consumed the least energy and yet Chengdu's air quality was one of the worst. After standardizing the meteorological condition and considering the energy profile of each city, the differential levels of the $PM_{2.5}$ concentration in the five cities highlight the role of geographic configuration of a city that defines the diffusion condition of pollution. In particular, Chengdu is located in the western edge of the Sichuan Basin which are known for lack of wind and high humidity. Located at the northwest corner of the NCP, Beijing is hemmed in by two mountain ranges, one on its west and one on its north and is heavily dependent on the northerly wind to clean up the pollution. In contrast, the other three cities lay on much more open terrain with better condition of diffusion. In particular, Shanghai and Guangzhou are quite close to the sea.

It is clear that each metropolis or region has an acceptable upper limit on the amount of emission per unit area in order to maintain acceptable air quality. The limit is related with the geography of the city and region at large which influences the dispersion of the local and transported air pollutants. The NCP where Beijing is located is an example of ignoring geographical limitations by installing excessive industrial capacity in a region that is not well dispersed. Respecting the geographical reality should be an important consideration for any regional planning and development.

4. Conclusion

Our analyses compared $PM_{2.5}$ data from the U.S. posts and the nearby MEP sites by looking at $PM_{2.5}$ range occurrence frequency and duration, the monthly air quality statistics, and the domestic heating effect (for Beijing and Shenyang). It is showed that the data from the U.S. posts and MEP stations were highly consistent in the five cities. The consistency is established by evaluating both the absolute concentration and the relative changes in the three years since January 2013. We take this as an important assurance in the fight against the chronic air pollution in China, as it shows the MEP had been fair in dealing with the data as far as the five cities are concerned. While this is encouraging, it does not mean that the data reliability will be maintained in the future given that the $PM_{2.5}$ became a performance measure for local government officials in January 2015. Continued scrutiny of the integrity of China's air pollution data should be an important aspect in the air quality management.

Our assessment on the air quality in the five cities has detected a decline in the $PM_{2.5}$ concentration in the five cities, and the most significant decline happened in 2015. While this is encouraging, we note that the levels of $PM_{2.5}$ concentration in the five cities are still well above the WHO recommended levels [*World Health Organization*, 2006]. The WHO's first interim target of annual concentration is 35 μg/m^3, and the second and third interim targets are 25 μg/m^3 and 15 μg/m^3, respectively. The ultimate annual air quality guidelines of WHO and the U.S. Environmental Protection Agency [*U.S. Environmental Protection Agency*, 2013] are 10 μg/m^3 and 12.5 μg/m^3, respectively.

The averages in 2015 were at least 60 μg/m^3 in Beijing, Chengdu, and Shenyang. Beijing's average in 2015 of 80 μg/m^3 was more than double the WHO's first interim standard and was 8 times the WHO's annual air quality guideline of 10 μg/m^3. The 90th percentiles in these three cities in 2015, despite the recent decline, were larger than 112 μg/m^3 with Beijing having a staggering 202 μg/m^3. Even in the two cities with the lowest $PM_{2.5}$

concentrations, Guangzhou and Shanghai, the annual averages were, respectively, 39 μg/m^3 and 50 μg/m^3. And Shanghai's 90th percentile in 2015 was 104 μg/m^3. Therefore, there is a long way to go to achieve even the first interim target set by the WHO.

Despite our analysis is based on only 3 year's data, the modest declines in $PM_{2.5}$ in 2015 were consistent with the decrease of the energy consumption. Hence, the moderate improvement in the air quality may be due to the economic slow-down that China is experiencing since 2014. According to official statistics, production of steel, cast iron, cement, and the amount of thermal power generation in 2015 decreased by 2.3%, 3.5%, 4.9% and 2.8%, respectively, and the nation-wide coal consumption decreased by 3.7% in 2015. As the National Bureau of Statistics of China has not released the energy data of each province, unofficial figures from a coal industrial website (http://www.sxcoal.com) showed that the coal consumption in NCP, Liaoning, YRD, Sichuan-Chongqing, and Guangdong declined by 5.4%, 0.6%, 3.2%, 13.7% and 16.8%, respectively, in 2015. These data would explain part of the $PM_{2.5}$ reduction in 2015. It is concerning that when the economy recovers, so would the air pollution. It appears timely for China to take a full evaluation on the effectiveness of the air pollution mitigation approach adopted since 2013 as the level of pollution is still very severe even for the cities with lowest $PM_{2.5}$ concentration. Our study suggests an approach that respects regional geographical characteristics is needed.

Acknowledgments
The research was supported by China's National Key Research Special Program grant SQ2016ZY01002112, National Key Basic Research Program grant 2015CB856000, Center for Statistical Science at Peking University, and National Natural Science Foundation of China grants 11131002, 71532001, and 71371016. Liang was partially supported by a Caizhai Scholarship (CZ201507) at Peking University. The $PM_{2.5}$ data of the U.S. posts are from http://www.stateair.net/web/historical/1/1.html, and the $PM_{2.5}$ data of the MEP sites are from https://wat.epmap.org/ (in Chinese). The weather data in Beijing are from http://weather.nocrew.org, and the weather data at the airports of Shanghai, Chengdu, and Shenyang are from https://weatherspark.com/. Data used in the paper are available upon request from Chen (csx@gsm.pku.edu). We have no competing interests.

References

Aitchison, J., and C. G. Aitken (1976), Multivariate binary discrimination by the kernel method, *Biometrika*, *63*, 413–420.

Andrews, S. Q. (2008), Inconsistencies in air quality metrics: 'Blue sky' days and PM_{10} concentrations in Beijing, *Environ. Res. Lett.*, *3*, 034009.

Burney, J., and V. Ramanathan (2014), Recent climate and air pollution impacts on Indian agriculture, *Proc. Natl. Acad. Sci. U. S. A.*, *111*, 16,319–16,324.

Chen, S. X., and C. Y. Tang (2011), Nonparametric regression with discrete covariates and missing values, *Stat. Interface*, *4*, 463–474.

Chen, Y., G. Z. Jin, N. Kumar, and G. Shi (2012), Gaming in air pollution data? Lessons from China, *B. E. J. Econ. Anal. Policy*, *13*(3), 1–43.

Davison, A. C., and D. V. Hinkley (1997), *Bootstrap Methods and Their Application*, vol. 1, Cambridge Univ. Press, New York.

Dominici, F., M. Greenstone, and C. R. Sunstein (2014), Particulate matter matters, *Science*, *344*, 257–259.

Dunn, O. J. (1961), Multiple comparisons among means, *J. Acoust. Soc. Am.*, *56*, 52–64.

Fan, J., and I. Gijbels (1996), *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

Ghanem, D., and J. Zhang (2014), 'Effortless perfection': Do Chinese cities manipulate air pollution data?, *J. Environ. Econ. Manage.*, *68*, 203–225.

Guo, S., et al. (2014), Elucidating severe urban haze formation in China, *Proc. Natl. Acad. Sci. U. S. A.*, *111*, 17,373–17,378.

Härdle, W. (1990), *Applied Nonparametric Regression*, vol. 27, Cambridge Univ. Press, New York.

Kulshrestha, U., and B. Kumar (2014), Airmass trajectories and long range transport of pollutants: Review of wet deposition scenario in South Asia, *Adv. Meteorol.*, *2014*, 596041.

Li, Q., and J. Racine (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton Univ. Press, Princeton, N. J.

Liang, X., et al. (2015), Assessing Beijing's $PM_{2.5}$ pollution: Severity, weather impact, APEC and winter heating, *Proc. R. Soc. A.*, *471*, 20150257.

National Bureau of Statistics, People's Republic of China (2015), National Bureau of Statistics, National data.

Pope, C. A., III, M. Ezzati, and D. W. Dockery (2009), Fine-particulate air pollution and life expectancy in the United States, *N. Engl. J. Med.*, *360*, 376–386.

Prajapati, S. K. (2012), Ecological effect of airborne particulate matter on plants, *Environ. Skep. Crit.*, *1*, 12–22.

Rosenbaum, P. R. (2002), *Observational Studies*, Springer, New York.

San Martini, F. M., C. A. Hasenkopf, and D. C. Roberts (2015), Statistical analysis of $PM_{2.5}$ observations from diplomatic facilities in China, *Atmos. Environ.*, *110*, 174–185.

Smith, R. L., J. M. Davis, J. Sacks, P. Speckman, and P. Styer (2000), Regression models for air pollution and daily mortality analysis of data from Birmingham, Alabama, *Environmetrics*, *11*, 719–743.

Spegele, B. (2012), Comparing Pollution Data: Beijing vs. US Embassy on $PM_{2.5}$, The Wall Street Journal (January 23th).

Stoerk, T. (2016), Statistical corruption in Beijing's air quality data has likely ended in 2012, *Atmos. Environ.*, *127*, 365–371.

Sun, Y., et al. (2014), Investigation of the sources and evolution processes of severe haze pollution in Beijing in January 2013, *J. Geophys. Res. Atmos.*, *119*, 4380–4398, doi:10.1002/2014JD021641.

Tian, H. Z., J. M. Hao, M. Y. Hu, and Y. F. Nie (2007), Recent trends of energy consumption and air pollution in China, *J. Energy Eng.*, *133*, 4–12.

U.S. Environmental Protection Agency (2013), National ambient air quality standards for particulate matter, final rule, *Fed. Regist.*, *78*, 3086–3287.

Von Engeln, A., and J. Teixeira (2013), A planetary boundary layer height climatology derived from ECMWF reanalysis data, *J. Clim.*, *26*, 6575–6590.

Wang, X., C. Wang, L. Gu, Y. Wang, and Q. Wang (2014), Concentration variations of atmospheric particulate matters in street greenbelts under typical weather conditions in spring, *Chin. J. Ecol.*, *33*, 2889–2896.

Wikle, C., and M. Hooten (2010), A general science-based framework for dynamical spatio-temporal models, *Test*, *19*, 417–451.

World Health Organization (2006), *Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide: Global Update 2005*, WHO Press, Geneva.

Zhang, X. Y., Y. Q. Wang, T. Niu, X. C. Zhang, S. L. Gong, Y. M. Zhang, and J. Y. Sun (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.

Zhao, H., H. Che, X. Zhang, Y. Ma, Y. Wang, H. Wang, and Y. Wang (2013), Characteristics of visibility and particulate matter (PM) in an urban area of Northeast China, *Atmos. Pollut. Res.*, *4*, 427–434.

Zheng, M., L. G. Salmon, J. J. Schauer, L. Zeng, C. S. Kiang, Y. Zhang, and G. R. Cass (2005), Seasonal trends in $PM_{2.5}$ source contributions in Beijing, China, *Atmos. Environ.*, *39*, 3967–3976.