Regional Air-Quality Assessment That Adjusts for Meteorological Confounding

Shuyi Zhang[1], Song Xi Chen[1,2*], Bin Guo[4,5], Hengfang Wang[6], Wei Lin[2,3]

[1]Guanghua School of Management, [2] Center for Statistical Science, [3] School of Mathematical Sciences, Peking University, Beijing 100871, China.

[4]Center of Statistical Research, [5]School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, China.

[6] Department of Statistics, Iowa State University, Ames, IA 50011, USA

Abstract

Although air pollution is caused by emission of pollutants to the atmosphere, the observed pollution levels are confounded by meteorological conditions, which largely determine the dispersion of the pollutants. Hence, effective air-quality management requires statistical measures that are immune to meteorological confounding and reflect changes in pollutant concentrations accurately and objectively. Motivated by the task of assessing changes in the underlying emission in a region near Beijing, we propose a spatial and temporal adjustment approach to remove meteorological confounding. The adjusted average pollutant concentration over space and time can capture changes in the underlying emission by controlling the meteorological variation. Estimation of the adjusted average is proposed together with theoretical and numerical analysis. We apply the approach to conducting air-quality assessments in the Beijing region, which reveals some intriguing patterns and trends that are useful for air-quality management.

Key words: Air-quality assessment; Meteorological confounding; Nonparametric regression; Spatio-temporal adjustment; Treatment effect.

1 Introduction

China has experienced severe air pollution as it rapidly industrializes in the last two decades. The cause of the air pollution is due to a steady increase in the emission of pollutants as the country becomes a global manufacturing hub. While this enormous increase has propelled a spectacular economic growth, it has also led to widespread air pollution in a substantial part of the country. The region around Beijing is the most affected. The primary air pollutants in Chinese cities are particulate matters $PM_{2.5}$ and PM_{10} [Zhang et al. (2012); Guo et al. (2014)], which represent airborne particles with aerodynamic diameters less than 2.5μm and 10μm, respectively. In recent years, the ground-level ozone (O_3) has been on the rise in China [Chen et al. (2018)].

The key in improving air quality is to reduce emissions, which requires a timely and accurate account of emission. Emission inventory is a commonly used tool for emission measurement which collects industrial data and downscales them to a finer resolution [Kuykendal (2017)]. This inventory is usually at yearly or smaller temporal frequencies and is subject to measurement and reporting errors. In China, although there are quite a number of emission inventories, they are typically three or four years behind and are not generally available.

This paper proposes using hourly air-quality data for emission quantification. An immediate challenge is the fact that the observed pollutant concentrations are confounded by meteorological conditions, for instance by the wind direction and speed and the relative humidity as demonstrated in Liang et al. (2015) and Finazzi et al. (2013). The meteorological confounding to the air pollution is similar to that in observational studies [Rosenbaum (2002); Qin (2017)] where the bias due to pre-treatment covariates needs to be adjusted in the evaluation of treatment effects. However, our

setting differs from observational studies where covariates follow the same baseline distribution [Huang et al. (2008) and Qin (2017)]. In our study, the baseline covariate distribution should be constructed properly by considering meteorological variations. Another major difference from the existing treatment effect literature is that there is a lack of random treatment assignment for applying the propensity-based approach, since the treatment variable associated with air-quality management is year and thus is fixed.

We propose a meteorological adjustment approach to removing weather confounding from the observed concentration. The adjustment is carried out both temporally and spatially to provide temporally and spatially comparable means and quantiles regarding the pollutant concentration at a time horizon. The adjusted means at different years can be compared to gain information about whether there is a reduction in the emission. **Temporally meteorological adjustment can also be conducted via the trend analysis as proposed in Thompson** et al. **(2001). The trend analysis is included as a special case in the proposed adjustment framework, which corresponds to linear regression. The other one was the three-year moving average method advocated by US Environmental Protection Agency (EPA). An advantage of the proposed adjustment is that it allows a general form of regression models. Moreover, it takes into account spatial variations in temporal adjustment.**

The paper is structured as follows. Section 2 describes the study region, the data that motivate our study, and models accounting for meteorological confounding. Section 3 outlines the spatial and temporal adjustment approach and its ability in gauging the underlying emission. Nonparametric estimators of the adjusted regional air-quality measures and their theoretical properties are given in Section 4. The variance estimation and hypothesis testing procedures are provided in Section 5. Section 6 demonstrates the proposed approach by empirically assessing air quality around Beijing. We defer technical conditions, proofs of theoretical results, simulation studies, and additional empirical results to the supplementary information (SI).

2 Study region, data and models

China established an air-quality monitoring network in January 2013 in 74 cities with 496 Guokong (nationally controlled) monitoring sites, which was extended to 1438 monitoring sites in January 2015 in 338 cities. These data are generally of high quality as shown in Liang et al. (2016) which cross-compared $PM_{2.5}$ data from the US diplomatic posts in five Chinese cities with the neighboring Guokong sites. The US Embassy in Beijing started to report hourly $PM_{2.5}$ concentrations from April 2008. As a part of the national network, Beijing Municipal Environmental Monitoring Center (BMEMC) administrates a monitoring network that consists of 35 air-quality monitoring sites, which collects hourly concentrations of $PM_{2.5}$ and five other pollutants: PM_{10}, **sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO)**, and ozone (O_3). The US Embassy measures only $PM_{2.5}$. Instead of the calendar year, we consider using the seasonal year which runs from March to February the following year that covers a set of four seasons from spring to winter.

We focus on the **North China Plain** (NCP) portion of Beijing as shown in Figure 1, which occupies a land area of 5180km^2 from 116.0°E to 116.8°E in longitude and from 39.5°N to 40.2°N in latitude. The study region has 28 monitoring sites including the US Embassy site, which encloses the urban core of Beijing confined by the Sixth-Ring Road plus the southern area between the Sixth-Ring Road and the border with Hebei Province. The "Southern Area" has 3 sites, while the area with the other 25 sites is termed as "Central Area". Hebei Province is known for having the worst air quality in China due to its enormous iron and steel consumptions together with other high emission industries. Including "Southern Area" serves to understand pollution transportation.

To adjust for meteorological confounding, we use data at 11 weather stations of the Central Meteorological Agency (CMA) in the study region. The locations of weather stations are marked in Figure 1. The meteorological variables contain hourly measurements of the air temperature, the air pressure, the relative humidity, the dew point temperature, the wind direction, the cumulative wind speed, and the cumulative precipitation. **The wind direction is an un-ordered discrete**

variable having 5 categories: northwest (NW), northeast (NE), southeast (SE), southwest (SW), and calm and variable (CV). According to the Magnus formula [Alduchov and Eskridge (1996)], the dew point temperature can be mathematically expressed by the relative humidity and the air temperature via a known nonlinear function. Hence, any two of the three variables can determine the third one. To reduce the dimensionality of covariates, we would like to drop one variable. Since the relative humidity is bounded within values between $[0, 1]$, including it in nonparametric regression will create the so-called boundary bias [see Page 202 in Fan and Yao (2003)]. To avoid the boundary issue, we drop the relative humidity in the analysis.

Suppose there are L air-quality monitoring sites in the study region $\mathcal{R}$, and S meteorological sites whose locations are collected in $\mathcal{W}$. At an air-quality monitoring site $\boldsymbol{s}$, let $Y_{ijt}(\boldsymbol{s})$ be the concentration of a pollutant at hour t of season j in year i, where $j = 1, \cdots, 4$ for spring, summer, fall, and winter, respectively, and $\boldsymbol{X}_{ijt}(\boldsymbol{s})$ be a **6-dimensional** vector of meteorological variables, which consist of the air pressure, the air temperature, the dew point temperature, the wind direction, the cumulative wind speed under a wind direction, and the cumulative precipitation, from a weather station which is the closest to the air-quality monitoring site $\boldsymbol{s}$.

Let $\boldsymbol{U}_{ijt}(\boldsymbol{s})$ be the level of emission which is regarded as latent as economic statistics are compiled at much coarser frequencies, which prevents a timely emission inventory. Nevertheless, an underlying model that describes the relationship between $Y_{ijt}(\boldsymbol{s})$ and $\{\boldsymbol{X}_{ijt}(\boldsymbol{s})^{\mathrm{T}}, \boldsymbol{U}_{ijt}(\boldsymbol{s})\}^{\mathrm{T}}$ is

$$Y_{ijt}(\boldsymbol{s}) = \tilde{m}_j\{\boldsymbol{X}_{ijt}(\boldsymbol{s}), \boldsymbol{U}_{ijt}(\boldsymbol{s})\} + \tilde{\epsilon}_{ijt}(\boldsymbol{s}), \tag{2.1}$$

for $t = 1, \ldots, n_{ij}$, where $\tilde{m}_j\{\boldsymbol{X}_{ijt}(\boldsymbol{s}), \boldsymbol{U}_{ijt}(\boldsymbol{s})\} = \mathrm{E}\{Y_{ijt}(\boldsymbol{s})|\boldsymbol{X}_{ijt}(\boldsymbol{s}), \boldsymbol{U}_{ijt}(\boldsymbol{s})\}$, and $\tilde{\epsilon}_{ijt}(\boldsymbol{s})$ are residuals, and n_{ij} is the number of hourly observations in season j and year i.

As $\boldsymbol{U}_{ijt}(\boldsymbol{s})$ is latent, we take the expectation on both sides of (2.1) conditioning on the observed weather covariates $\boldsymbol{X}_{ijt}(\boldsymbol{s})$, which gives rise to

$$Y_{ijt}(\boldsymbol{s}) = m_{ij}\{\boldsymbol{X}_{ijt}(\boldsymbol{s}), \boldsymbol{s}\} + \sigma_{ij}\{\boldsymbol{X}_{ijt}(\boldsymbol{s}), \boldsymbol{s}\}e_{ijt}(\boldsymbol{s}), \tag{2.2}$$

where $m_{ij}(\boldsymbol{x},\boldsymbol{s}) = \mathrm{E}\{Y_{ijt}(\boldsymbol{s})|\boldsymbol{X}_{ijt}(\boldsymbol{s}) = \boldsymbol{x}\}$, $\sigma^2_{ij}(\boldsymbol{x},\boldsymbol{s}) = \mathrm{Var}\{Y_{ijt}(\boldsymbol{s})|\boldsymbol{X}_{ijt}(\boldsymbol{s}) = x\}$ and $e_{ijt}(\boldsymbol{s})$ are standardized residuals. **Moreover, by taking the conditional expectation of the right hand side in (2.1) given $\boldsymbol{X}_{ijt}(\boldsymbol{s})$, we can obtain another formula for $m_{ij}(\boldsymbol{x},\boldsymbol{s})$ as $m_{ij}(\boldsymbol{x},\boldsymbol{s}) = \mathrm{E}[\tilde{m}_j\{\boldsymbol{x},\boldsymbol{U}_{ijt}(\boldsymbol{s})\}|\boldsymbol{X}_{ijt}(\boldsymbol{s}) = \boldsymbol{x}]$, which is essentially determined by the conditional distribution of $\boldsymbol{U}_{ijt}(\boldsymbol{s})$ given $\boldsymbol{X}_{ijt}(\boldsymbol{s})$.** Model (2.2) is the model that we will use for inference. It should be noted that the yearly index i and site index $\boldsymbol{s}$ appear in the regression function $m_{ij}(\boldsymbol{x},\boldsymbol{s})$ as $\boldsymbol{U}_{ijt}(\boldsymbol{s})$'s distribution may change yearly and spatially.

Let $\mathbb{X}_{ijt} = \{\boldsymbol{X}_{ijt}(\boldsymbol{s}_1)^{\mathrm{T}},\ldots,\boldsymbol{X}_{ijt}(\boldsymbol{s}_L)^{\mathrm{T}}\}^{\mathrm{T}}$ and $\boldsymbol{e}_{ijt} = \{e_{ijt}(\boldsymbol{s}_1),\ldots,e_{ijt}(\boldsymbol{s}_L)\}^{\mathrm{T}}$, where L is the number of air-quality monitoring sites. **Then $\mathbb{X}_{ijt}$ and $\boldsymbol{e}_{ijt}$ are $6L$- and L-dimensional, respectively, and collect the meteorological variables and standardized residuals of all the sites in the region**. We assume the multivariate time series $\{\mathbb{X}_{ijt}\}_{t=1}^{n_{ij}}$ and $\{\boldsymbol{e}_{ijt}\}_{t=1}^{n_{ij}}$ are temporally stationary and weakly dependent satisfying the α-mixing condition (see the SI for details), while leaving the spatial dependence unspecified to allow generality.

To better discuss the combined effects of emission and weather, we can assume an additive structure to $\tilde{m}_j\{\boldsymbol{X}(\boldsymbol{s}),\boldsymbol{U}(\boldsymbol{s})\}$ so that

$$\tilde{m}_j\{\boldsymbol{X}(\boldsymbol{s}),\boldsymbol{U}(\boldsymbol{s})\} = \tilde{m}_{j,1}\{\boldsymbol{X}(\boldsymbol{s})\} + \tilde{m}_{j,2}\{\boldsymbol{U}(\boldsymbol{s})\} + \tilde{m}_{j,3}\{\boldsymbol{X}(\boldsymbol{s}),\boldsymbol{U}(\boldsymbol{s})\}, \tag{2.3}$$

where the main effects and the interaction are homogeneous with respect to the year and location, but are seasonally specific. The corresponding version for **the observed data** is

$$m_{ij}(\boldsymbol{x},\boldsymbol{s}) = \tilde{m}_{j,1}(\boldsymbol{x}) + \mathrm{E}[\tilde{m}_{j,2}\{\boldsymbol{U}_{ijt}(\boldsymbol{s})\}|\boldsymbol{X}_{ijt}(\boldsymbol{s}) = \boldsymbol{x}] + \mathrm{E}[\tilde{m}_{j,3}\{\boldsymbol{x},\boldsymbol{U}_{ijt}(\boldsymbol{s})\}|\boldsymbol{X}_{ijt}(\boldsymbol{s}) = \boldsymbol{x}]. \tag{2.4}$$

Both (2.3) and (2.4) serve to untangle the emission effect in the next section.

For our purpose, there is no need to build an elaborate parametric version of (2.2). It is also not necessary to include the temporal lagged or the spatial neighbors' responses. This is because the aim of the study is in assessing the pollutant concentration rather than predicting it. For the purpose of assessment, the nonparametric model (2.2) would be sufficient.

3 Spatially and temporally adjusted measures

We propose a spatial and temporal adjustment approach to account for meteorological confounding in statistical measures like the means or the quantiles under Model (2.2). This is motivated by the practice in China and other countries where the averaged pollutant concentration over a time horizon is used as measures of air quality. Thompson et al. (2001) considered a trend analysis under linear regression models in the context of ground ozone pollution. We will show that the adjustment offered by the trend analysis can be viewed as a special case of our proposed approach.

Ordinarily, the mean of $Y_{ijt}(\boldsymbol{s})$ is $\mathrm{E}\{Y_{ijt}(\boldsymbol{s})\} = \mathrm{E}[\mathrm{E}\{Y_{ijt}(\boldsymbol{s})|\boldsymbol{X}_{ijt}(\boldsymbol{s})\}] = \mathrm{E}[m_{ij}\{\boldsymbol{X}_{ijt}(\boldsymbol{s}), \boldsymbol{s}\}]$. The key is which probability density is used in the last expectation. A generic form for the mean is

$$\mathrm{E}\{Y_{ijt}(\boldsymbol{s})\} = \int m_{ij}(\boldsymbol{x}, \boldsymbol{s}) f_j(\boldsymbol{x}, \boldsymbol{s}) d\boldsymbol{x}, \tag{3.1}$$

where $f_j(\boldsymbol{x}, \boldsymbol{s})$ denotes a generic density for $\boldsymbol{X}_{ijt}(\boldsymbol{s})$ in season j at site $\boldsymbol{s}$. Different forms of $f_j(\boldsymbol{x}, \boldsymbol{s})$ lead to different measures as shown below.

Let $f_{ij}(\boldsymbol{x}, \boldsymbol{s})$ be the density of $\boldsymbol{X}_{ijt}(\boldsymbol{s})$ for season j of year i at site $\boldsymbol{s}$. If one chooses $f_j(\boldsymbol{x}, \boldsymbol{s}) = f_{ij}(\boldsymbol{x}, \boldsymbol{s})$, the mean in (3.1) is denoted as $\mu^0_{ij}(\boldsymbol{s})$. The commonly used air-quality measure is the simple average $\bar{Y}_{ij}(\boldsymbol{s}) = n_{ij}^{-1} \sum_{t=1}^{n_{ij}} Y_{ijt}(\boldsymbol{s})$. By the law of large numbers for weakly dependent processes, $\bar{Y}_{ij}(\boldsymbol{s}) \xrightarrow{P} \mu^0_{ij}(\boldsymbol{s})$ as $n_{ij} \to \infty$. While $\mu^0_{ij}(\boldsymbol{s})$ and $\bar{Y}_{ij}(\boldsymbol{s})$ measure the pollution exposure for the health purpose, they are confounded by weather conditions of different years and locations.

Another version of $f_j(\boldsymbol{x}, \boldsymbol{s})$ is offered by averaging $\{f_{aj}(\boldsymbol{x}, \boldsymbol{s})\}_{a=1}^{A_j}$ for season j over A_j years. Specifically, let $f_{\cdot j}(\boldsymbol{x}, \boldsymbol{s}) = A_j^{-1} \sum_{a=1}^{A_j} f_{aj}(\boldsymbol{x}, \boldsymbol{s})$, which defines the temporal baseline weather condition. Choosing $f_j(\boldsymbol{x}, \boldsymbol{s}) = f_{\cdot j}(\boldsymbol{x}, \boldsymbol{s})$ in (3.1), we arrive at

$$\tilde{\mu}_{ij}(\boldsymbol{s}) = \int m_{ij}(\boldsymbol{x}, \boldsymbol{s}) f_{\cdot j}(\boldsymbol{x}, \boldsymbol{s}) d\boldsymbol{x} = A_j^{-1} \sum_{a=1}^{A_j} \int m_{ij}(\boldsymbol{x}, \boldsymbol{s}) f_{aj}(\boldsymbol{x}, \boldsymbol{s}) d\boldsymbol{x}. \tag{3.2}$$

The terms $\int m_{ij}(\boldsymbol{x}, \boldsymbol{s}) f_{aj}(\boldsymbol{x}, \boldsymbol{s}) d\boldsymbol{x}$ for $a \neq i$ are counter-factuals [Rosenbaum and Rubin (1983)], which provide the potential averages under other years' weather but with year i's pollution-weather mechanism $m_{ij}(\boldsymbol{x}, \boldsymbol{s})$. We call $\tilde{\mu}_{ij}(\boldsymbol{s})$ the temporally adjusted average.

The $\{\tilde{\mu}_{aj}(\boldsymbol{s})\}_{a=1}^{A_j}$ are comparable over different years as they are formulated under the temporal baseline $f_{\cdot j}(\boldsymbol{x}, \boldsymbol{s})$. Specifically, the difference

$$\mu_{ij}^0(\boldsymbol{s}) - \tilde{\mu}_{ij}(\boldsymbol{s}) = \int m_{ij}(\boldsymbol{x}, \boldsymbol{s})\{f_{ij}(\boldsymbol{x}, \boldsymbol{s}) - f_{\cdot j}(\boldsymbol{x}, \boldsymbol{s})\}d\boldsymbol{x}$$

measures the amount of the weather confounding in year i, and

$$\tilde{\mu}_{ij}(\boldsymbol{s}) - \tilde{\mu}_{kj}(\boldsymbol{s}) = \int \{m_{ij}(\boldsymbol{x}, \boldsymbol{s}) - m_{kj}(\boldsymbol{x}, \boldsymbol{s})\}f_{\cdot j}(\boldsymbol{x}, \boldsymbol{s})d\boldsymbol{x}$$

measures the temporal treatment effect due to different emission levels between years i and k.

Thompson et al. **(2001) considered a trend analysis in the linear regression model to gain information on the trend of the ground level ozone pollution. The trend analysis can be included in our proposed framework under the linear regression setting. To appreciate this, suppose the regression function $m_{ij}(\boldsymbol{x}, \boldsymbol{s})$ is linear such that**

$$Y_{ijt}(\boldsymbol{s}) = \alpha_{ij}(\boldsymbol{s}) + \boldsymbol{\beta}_{ij}^T(\boldsymbol{s})\tilde{\boldsymbol{X}}_{ijt}(\boldsymbol{s}) + \epsilon_{ijt}(\boldsymbol{s}), \tag{3.3}$$

where $\tilde{\boldsymbol{X}}_{ijt}(\boldsymbol{s})$ are temporally centered covariates over A_j years. Then, it is readily shown that $\tilde{\mu}_{ij}(\boldsymbol{s}) = \alpha_{ij}(\boldsymbol{s})$ due to temporal centering of covariates, which was not explicitly stated in Thompson et al. **(2001). Moreover, $\alpha_{ij}(\boldsymbol{s})$ can be used for spatio-temporal comparison, and $\alpha_{ij}(\boldsymbol{s}) = \mu_{ij}(\boldsymbol{s})$, the spatially and temporally adjusted average which we will define below. It should be noted that our proposed adjustments by $\tilde{\mu}_{ij}(\boldsymbol{s})$ and $\mu_{ij}(\boldsymbol{s})$ allow more general regression models with linear regression employed by the trend analysis as a special case.**

The temporally adjusted means $\tilde{\mu}_{ij}(\boldsymbol{s})$ are not comparable spatially for $\{\tilde{\mu}_{aj}(\boldsymbol{s}_1)\}_{a=1}^{A_j}$ and $\{\tilde{\mu}_{aj}(\boldsymbol{s}_2)\}_{a=1}^{A_j}$ at two different sites $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ since the two sites can bear different weather distributions. In the following, we takes into account spatial variations into temporal adjustment. As indicated earlier, the study region $\mathcal{R}$ has S meteorological sites whose locations are collected in $\mathcal{W}$. **We can define the spatial and temporal baseline as a weighted version of $f_{\cdot j}(\boldsymbol{x}, \boldsymbol{s})$. Specifically, let $p(\boldsymbol{s})$ be a probability density function over the study region. We can**

construct a weighted spatio-temporal weather baseline as

$$f_{.j}^{p}(\boldsymbol{x}) = \int_{\boldsymbol{s}\in\mathcal{R}} f_{.j}(\boldsymbol{x},\boldsymbol{s})p(\boldsymbol{s})d\boldsymbol{s}.$$

Under the fixed design sampling over space, if we select $p(\boldsymbol{s})$ as the uniformly distributed density over the meteorological sites, we obtain

$$f_{.j}(\boldsymbol{x}) = S^{-1}\sum_{\boldsymbol{s}'\in\mathcal{W}} f_{.j}(\boldsymbol{x},\boldsymbol{s}'). \tag{3.4}$$

We call $f_{.j}(\boldsymbol{x})$ the spatio-temporal weather baseline over $\mathcal{R}$ for season j. In the following, we will use this unweighted baseline $f_{.j}(\boldsymbol{x})$ to simplify the analysis. Using $f_{.j}(\boldsymbol{x})$ in (3.4), we arrive at the spatially and temporally adjusted average:

$$\mu_{ij}(\boldsymbol{s}) = \int m_{ij}(\boldsymbol{x},\boldsymbol{s})f_{.j}(\boldsymbol{x})d\boldsymbol{x} = S^{-1}A_j^{-1}\sum_{\boldsymbol{s}'\in\mathcal{W}}\sum_{a=1}^{A_j}\int m_{ij}(\boldsymbol{x},\boldsymbol{s})f_{aj}(\boldsymbol{x},\boldsymbol{s}')d\boldsymbol{x}, \tag{3.5}$$

where those terms with $a \neq i$ or $\boldsymbol{s}' \neq \boldsymbol{s}$ are spatial and temporal counter-factuals.

We now outline the benefits of the spatial and temporal adjustment under the additive regression framework (2.3) and (2.4). Define

$$\mu_j^M = \int \tilde{m}_{j,1}(\boldsymbol{x})f_{.j}(\boldsymbol{x})d\boldsymbol{x}, \quad \mu_{ij}^E(\boldsymbol{s}) = \int \mathrm{E}[\tilde{m}_{j,2}\{\boldsymbol{U}_{ijt}(\boldsymbol{s})\}|\boldsymbol{X}_{ijt}(\boldsymbol{s}) = \boldsymbol{x}]f_{.j}(\boldsymbol{x})d\boldsymbol{x} \text{ and}$$

$$\mu_{ij}^{ME}(\boldsymbol{s}) = \int \mathrm{E}[\tilde{m}_{j,3}\{\boldsymbol{x},\boldsymbol{U}_{ijt}(\boldsymbol{s})\}|\boldsymbol{X}_{ijt}(\boldsymbol{s}) = \boldsymbol{x}]f_{.j}(\boldsymbol{x})d\boldsymbol{x}.$$

From (3.5), $\mu_{ij}(\boldsymbol{s}) = \mu_j^M + \mu_{ij}^E(\boldsymbol{s}) + \mu_{ij}^{ME}(\boldsymbol{s})$. It is noted that the meteorological effect μ_j^M, due to the spatial and temporal adjustment, is the same for all the years and locations at a given season j. However, the emission and the interaction effects $\mu_{ij}^E(\boldsymbol{s})$ and $\mu_{ij}^{ME}(\boldsymbol{s})$ can vary yearly and spatially as the distribution of $\boldsymbol{U}_{ijt}(\boldsymbol{s})$ can differ in years and sites at season j.

Hence, the yearly difference in two consecutive years at season j is

$$\mu_{ij}(\boldsymbol{s}) - \mu_{i-1,j}(\boldsymbol{s}) = \mu_{ij}^E(\boldsymbol{s}) - \mu_{i-1,j}^E(\boldsymbol{s}) + \mu_{ij}^{ME}(\boldsymbol{s}) - \mu_{i-1,j}^{ME}(\boldsymbol{s}).$$

Consider the yearly change related to the emission

$$\mu_{ij}^E(\boldsymbol{s}) - \mu_{i-1,j}^E(\boldsymbol{s}) = \int\int \tilde{m}_{j,2}(\boldsymbol{u})\{g_{ij}(\boldsymbol{u},\boldsymbol{s}|\boldsymbol{x}) - g_{i-1,j}(\boldsymbol{u},\boldsymbol{s}|\boldsymbol{x})\}f_{.j}(\boldsymbol{x})d\boldsymbol{u}d\boldsymbol{x}, \tag{3.6}$$

where $g_{ij}(\boldsymbol{u},\boldsymbol{s}|\boldsymbol{x})$ is the conditional density of $\boldsymbol{U}_{ijt}(\boldsymbol{s})$ given $\boldsymbol{X}_{ijt}(\boldsymbol{s}) = \boldsymbol{x}$. Similarly, the yearly

change related to the interaction is

$$\mu_{ij}^{ME}(\boldsymbol{s}) - \mu_{i-1,j}^{ME}(\boldsymbol{s}) = \int\int \tilde{m}_{j,3}(\boldsymbol{x},\boldsymbol{u})\{g_{ij}(\boldsymbol{u},\boldsymbol{s}|\boldsymbol{x}) - g_{i-1,j}(\boldsymbol{u},\boldsymbol{s}|\boldsymbol{x})\}f_{\cdot j}(\boldsymbol{x})d\boldsymbol{u}d\boldsymbol{x}. \tag{3.7}$$

Both yearly changes in (3.6) and (3.7) are dependent on $g_{ij}(\boldsymbol{u},\boldsymbol{s}|\boldsymbol{x}) - g_{i-1,j}(\boldsymbol{u},\boldsymbol{s}|\boldsymbol{x})$, the yearly change in the conditional densities. This is only possible with the employment of the spatio-temporal baseline weather condition $f_{\cdot j}(\boldsymbol{x})$. Without the baseline, we can not attribute the yearly differences to that in the emission as it may be due to yearly change in the meteorological condition.

The same analysis can be made when we compare $\mu_{ij}(\boldsymbol{s}_1)$ and $\mu_{ij}(\boldsymbol{s}_2)$ at two locations, and we can attribute the difference as the difference in the emission profiles at two locations since the meteorological variables have been standardized spatially.

With the spatially and temporally adjusted measure $\mu_{ij}(\boldsymbol{s})$, we can construct the average pollutant concentration in an area $\mathcal{A}$. The average pollution over $\mathcal{A}$ is

$$\mu_{ij}(\mathcal{A}) = |\mathcal{A}|^{-1}\sum_{\boldsymbol{s}\in\mathcal{A}}\mu_{ij}(\boldsymbol{s}), \tag{3.8}$$

where $|\mathcal{A}|$ denotes the number of air-quality monitoring sites in $\mathcal{A}$. **This version of the regional air-quality measure $\mu_{ij}(\mathcal{A})$ is a simple average of $\mu_{ij}(\boldsymbol{s})$ over air-quality monitoring sites in $\mathcal{A}$, corresponding to the design commonly practiced in China's air-quality management. It may be viewed as conditioning on the locations of the monitoring sites, which mirrors the fix design survey sampling approach. In practice, the distribution of the monitoring sites may not be evenly distributed with certain area (for instance north of Beijing) having a higher density of the sites relative to another region (south of Beijing). The region with less number of monitoring sites will encounter higher variation, as compared to a region with more sites, while other things being equal.**

To counter the uneven site distributions, we can introduce a weight function $w_{\mathcal{A}}(\boldsymbol{s})$ to attain a weighted version of the regional mean:

$$\mu_{ij}^{w}(\mathcal{A}) = |\mathcal{A}|^{-1}\sum_{\boldsymbol{s}\in\mathcal{A}}\mu_{ij}(\boldsymbol{s})w_{\mathcal{A}}(\boldsymbol{s})d\boldsymbol{s},$$

where $w_{\mathcal{A}}(\boldsymbol{s})$ may re-allocate weights to attain spatial balance. For ease of expedition, we will consider (3.8) for $\mu_{ij}(\mathcal{A})$ in both theoretical and empirical studies.

4 Estimation and theoretical properties

A key step in constructing estimators for $\mu_{ij}(\boldsymbol{s})$ in (3.5) and $\mu_{ij}(\mathcal{A})$ in (3.8) is to estimate $m_{ij}(\boldsymbol{x}, \boldsymbol{s})$. In this paper, we adopt the nonparametric kernel method [Härdle (1990); Fan and Yao (2003)] for estimating the regression function $m_{ij}(\boldsymbol{x}, \boldsymbol{s})$.

We partition $\boldsymbol{X}_{ijt}(\boldsymbol{s}) = \{\boldsymbol{Z}_{ijt}(\boldsymbol{s})^{\mathrm{T}}, W_{ijt}(\boldsymbol{s})\}^{\mathrm{T}}$ so that $W_{ijt}(\boldsymbol{s})$ is the categorical wind direction and $\boldsymbol{Z}_{ijt}(\boldsymbol{s})$ contains the remaining continuous covariates of d-dimension. Let $K(\cdot)$ be a d-dimensional symmetric kernel function (see the SI for details). Define

$$K_{\boldsymbol{H}}(\boldsymbol{z}) = (h_1 h_2 \cdots h_d)^{-1} K(z_1/h_1, \ldots, z_d/h_d),$$

where $\boldsymbol{z} = (z_1, \ldots, z_d)^{\mathrm{T}}$, and $\boldsymbol{H} = (h_1, \ldots, h_d)^{\mathrm{T}}$ is a vector of smoothing bandwidths. The kernel estimator [Fan and Yao (2003)] of $m_{ij}(\boldsymbol{x}, \boldsymbol{s})$ using data of season j of year i at site $\boldsymbol{s}$ under wind direction w is

$$\hat{m}_{ij}(\boldsymbol{z}, w; \boldsymbol{s}) = \frac{\sum_{t=1}^{n_{ij}} K_{\boldsymbol{H}}\{\boldsymbol{z} - \boldsymbol{Z}_{ijt}(\boldsymbol{s})\} Y_{ijt}(\boldsymbol{s}) I\{W_{ijt}(\boldsymbol{s}) = w\}}{\sum_{t=1}^{n_{ij}} K_{\boldsymbol{H}}\{\boldsymbol{z} - \boldsymbol{Z}_{ijt}(\boldsymbol{s})\} I\{W_{ijt}(\boldsymbol{s}) = w\}}, \tag{4.1}$$

where n_{ij} is the sample size, and $I(\cdot)$ is the indicator function, and $W_{ijt}(\boldsymbol{s}) = 1, \ldots, 5$ correspond to wind directions CV, NE, NW, SE, and SW.

The smoothing bandwidths are chosen based on the cross-validation method [Härdle (1990); Fan and Yao (2003)] for each given wind direction. In some seasons when the sample size under a wind direction was small, they were merged with data of another direction that had the similar effect on the pollution. For instance, the three pollution-enhancing wind directions SW, CV and SE can be combined, so are the two pollution-reducing directions NW and NE. It is noted that the targets of inference are about $\mu_{ij}(\boldsymbol{s})$ and $\mu_{ij}(\mathcal{A})$ which are integrated versions of $m_{ij}(\boldsymbol{x}, \boldsymbol{s})$. As shown in Theorems 1 and 2, estimators of $\mu_{ij}(\boldsymbol{s})$ and $\mu_{ij}(\mathcal{A})$ enjoy the root-n convergence rate.

This means they are less sensitive to the smoothing bandwidths than the estimators $\hat{m}_{ij}(\boldsymbol{x}, \boldsymbol{s})$.

For any $\boldsymbol{x} = (\boldsymbol{z}^{\mathrm{T}}, w)^{\mathrm{T}}$, let $\hat{F}_{\cdot j}(\boldsymbol{x})$ be the empirical distribution function corresponding to the distribution $F_{\cdot j}(\boldsymbol{x})$ with $f_{\cdot j}(\boldsymbol{x})$ as the density. $\hat{F}_{\cdot j}(\boldsymbol{x})$ could be obtained based on A_j years' data at all sites for season j. According to the law of large numbers, there is no need to explicitly construct $\hat{F}_{\cdot j}(\boldsymbol{x})$ and the proposed estimator of $\mu_{ij}(\boldsymbol{s})$ is

$$\begin{aligned}\hat{\mu}_{ij}(\boldsymbol{s}) &= \int \hat{m}_{ij}(\boldsymbol{x}, \boldsymbol{s}) d\hat{F}_{\cdot j}(\boldsymbol{x}) \\ &= S^{-1} \left(\sum_{a=1}^{A_j} n_{aj} \right)^{-1} \sum_{w=1}^{5} \sum_{\boldsymbol{s}' \in \mathcal{W}} \sum_{a=1}^{A_j} \sum_{t=1}^{n_{aj}} \hat{m}_{ij}\{\boldsymbol{Z}_{ajt}(\boldsymbol{s}'), w, \boldsymbol{s}\} I\{W_{ajt}(\boldsymbol{s}') = w\}, \end{aligned} \tag{4.2}$$

where $\hat{m}_{ij}(\boldsymbol{x}, \boldsymbol{s})$ is given in (4.1). The regional average $\mu_{ij}(\mathcal{A})$ is estimated by

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

We can extend the above framework to meteorologically adjusted distribution of the pollutant, which can produce adjusted quantiles to provide information about extreme levels of concentrations. Similar to the adjusted average in (3.5), we define the adjusted distribution function for season j of year i at site $\boldsymbol{s}$ as

$$G_{ij}(y, \boldsymbol{s}) = \sum_{w=1}^{5} \int F_{ij}(y, \boldsymbol{s} | \boldsymbol{z}, w) f_{\cdot j}(\boldsymbol{z}, w) d\boldsymbol{z},$$

where $F_{ij}(y, \boldsymbol{s}|\boldsymbol{z}, w) = P\{Y_{ijt}(\boldsymbol{s}) \leq y | \boldsymbol{Z}_{ijt}(\boldsymbol{s}) = \boldsymbol{z}, W_{ijt}(\boldsymbol{s}) = w\}$ is the conditional distribution. Similar to (4.2), the estimator of $G_{ij}(y, \boldsymbol{s})$ is

$$\hat{G}_{ij}(y, \boldsymbol{s}) = S^{-1} \left(\sum_{a=1}^{A_j} n_{aj} \right)^{-1} \sum_{w=1}^{5} \sum_{\boldsymbol{s}' \in \mathcal{W}} \sum_{a=1}^{A_j} \sum_{t=1}^{n_{aj}} \hat{F}_{ij}\{y, \boldsymbol{s} | \boldsymbol{Z}_{ajt}(\boldsymbol{s}'), w\} I\{W_{ajt}(\boldsymbol{s}') = w\}, \tag{4.4}$$

where

$$\hat{F}_{ij}(y, \boldsymbol{s}|\boldsymbol{z}, w) = \frac{\sum_{t=1}^{n_{ij}} K'_{\boldsymbol{H}'}\{\boldsymbol{z} - \boldsymbol{Z}_{ijt}(\boldsymbol{s})\} R_{h_0}\{Y_{ijt}(\boldsymbol{s}) - y\} I\{W_{ijt}(\boldsymbol{s}) = w\}}{\sum_{t=1}^{n_{ij}} K'_{\boldsymbol{H}'}\{\boldsymbol{z} - \boldsymbol{Z}_{ijt}(\boldsymbol{s})\} I\{W_{ijt}(\boldsymbol{s}) = w\}}$$

is the kernel estimator of $F_{ij}(y, \boldsymbol{s}|\boldsymbol{z}, w)$. Here $R_{h_0}(y) = \int_0^{y/h_0} k(u) du$ is the integration of the univariate kernel $k(\cdot)$ and h_0 is the smoothing bandwidth. For any $q \in (0, 1)$, the adjusted q-th percentile is estimated by $\hat{G}_{ij}^{-1}(q, \boldsymbol{s})$, the inverse of the estimated adjusted distribution function.

We present asymptotic properties of estimators $\hat{\mu}_{ij}(\boldsymbol{s})$ and $\hat{\mu}_{ij}(\mathcal{A})$ in the rest of this section. To simplify the presentation, we consider the case where the covariates in $\boldsymbol{X}_{ijt}(\boldsymbol{s})$ are continuous, which essentially considers adjusted averages at each wind direction. The overall results for (4.2) and (4.3) can be obtained by combining the results of all wind directions. Under such arrangement,

$$\hat{m}_{ij}(\boldsymbol{x},\boldsymbol{s}) = \frac{\sum_{t=1}^{n_{ij}} K_{\boldsymbol{H}}\{\boldsymbol{x}-\boldsymbol{X}_{ijt}(\boldsymbol{s})\}Y_{ijt}(\boldsymbol{s})}{\sum_{t=1}^{n_{ij}} K_{\boldsymbol{H}}\{\boldsymbol{x}-\boldsymbol{X}_{ijt}(\boldsymbol{s})\}} \quad \text{and}$$

$$\hat{\mu}_{ij}(\boldsymbol{s}) = \int \hat{m}_{ij}(\boldsymbol{x},\boldsymbol{s})d\hat{F}_{\cdot j}(\boldsymbol{x}) = S^{-1}\left(\sum_{a=1}^{A_j} n_{aj}\right)^{-1} \sum_{a=1}^{A_j}\sum_{\boldsymbol{s}'\in\mathcal{W}}\sum_{t=1}^{n_{aj}} \hat{m}_{ij}\{\boldsymbol{X}_{ajt}(\boldsymbol{s}'),\boldsymbol{s}\}.$$

We first introduce some notations. Recall that $\mathbb{X}_{ijt} = \{\boldsymbol{X}_{ijt}(\boldsymbol{s}_1)^{\mathrm{T}},\ldots,\boldsymbol{X}_{ijt}(\boldsymbol{s}_L)^{\mathrm{T}}\}^{\mathrm{T}}$ and $\boldsymbol{e}_{ijt} = \{e_{ijt}(\boldsymbol{s}_1),\ldots,e_{ijt}(\boldsymbol{s}_L)\}^{\mathrm{T}}$. Similarly, define $\mathbb{U}_{ijt} = \{\boldsymbol{U}_{ijt}(\boldsymbol{s}_1)^{\mathrm{T}},\ldots,\boldsymbol{U}_{ijt}(\boldsymbol{s}_L)^{\mathrm{T}}\}^{\mathrm{T}}$. Under the temporal stationarity assumptions of $\mathbb{X}_{ijt}$ and $\boldsymbol{e}_{ijt}$ (see the SI for details), define $q_{ij,t-t'}(\boldsymbol{x},\boldsymbol{x}';\boldsymbol{s},\boldsymbol{s}')$ as the joint density of $\boldsymbol{X}_{ijt}(\boldsymbol{s})$ and $\boldsymbol{X}_{ijt'}(\boldsymbol{s}')$, $\rho(|t-t'|;\boldsymbol{s},\boldsymbol{s}') = \mathrm{E}\{e_{ijt}(\boldsymbol{s})e_{ijt'}(\boldsymbol{s}')|\mathcal{F}_{ij}\}$ and

$$C^a_{i_1i_2,j,t_1-t_2}(\boldsymbol{s}_1',\boldsymbol{s}_2';\boldsymbol{s}_1,\boldsymbol{s}_2) = \mathrm{Cov}[m_{i_1j}\{\boldsymbol{X}_{ajt_1}(\boldsymbol{s}_1'),\boldsymbol{s}_1\}, m_{i_2j}\{\boldsymbol{X}_{ajt_2}(\boldsymbol{s}_2'),\boldsymbol{s}_2\}].$$

The assumptions needed for the theorems in this section along with their proofs are given in the SI. The major ones are for a given pair of i and j: (i) the emission $\{\mathbb{U}_{ijt}\}_{t=1}^{n_{ij}}$ are identically distributed; (ii) the weather variables $\{\mathbb{X}_{ijt}\}_{t=1}^{n_{ij}}$ and the standardized residuals $\{\boldsymbol{e}_{ijt}\}_{t=1}^{n_{ij}}$ are both temporally strictly stationary and α-mixing, but not necessarily spatially stationarity to allow more flexible spatial dependence. Under regularity conditions, we can define

$$\gamma_{ij}(\boldsymbol{s}_1,\boldsymbol{s}_2) = \sum_{k=-\infty}^{\infty}\rho(|k|;\boldsymbol{s}_1,\boldsymbol{s}_2)\iint \sigma_{ij}(\boldsymbol{x}_1,\boldsymbol{s}_1)\sigma_{ij}(\boldsymbol{x}_2,\boldsymbol{s}_2)\frac{q_{ij,k}(\boldsymbol{x}_1,\boldsymbol{x}_2;\boldsymbol{s}_1,\boldsymbol{s}_2)}{f_{ij}(\boldsymbol{x}_1,\boldsymbol{s}_1)f_{ij}(\boldsymbol{x}_2,\boldsymbol{s}_2)}dF_{\cdot j}(\boldsymbol{x}_1)dF_{\cdot j}(\boldsymbol{x}_2),$$

$$\lambda_{i_1i_2,j}(\boldsymbol{s}_1,\boldsymbol{s}_2) = S^{-2}A_j^{-2}\sum_{a=1}^{A_j}\sum_{\boldsymbol{s}_1',\boldsymbol{s}_2'\in\mathcal{W}}\sum_{k=-\infty}^{\infty} C^a_{i_1i_2,j,k}(\boldsymbol{s}_1',\boldsymbol{s}_2';\boldsymbol{s}_1,\boldsymbol{s}_2),$$

and their corresponding regional versions

$$\gamma_{ij}(\mathcal{A},\mathcal{B}) = |\mathcal{A}|^{-1}|\mathcal{B}|^{-1}\sum_{\boldsymbol{s}_1\in\mathcal{A},\boldsymbol{s}_2\in\mathcal{B}}\gamma_{ij}(\boldsymbol{s}_1,\boldsymbol{s}_2) \quad \text{and} \quad \lambda_{i_1i_2,j}(\mathcal{A},\mathcal{B}) = |\mathcal{A}|^{-1}|\mathcal{B}|^{-1}\sum_{\boldsymbol{s}_1\in\mathcal{A},\boldsymbol{s}_2\in\mathcal{B}}\lambda_{i_1i_2,j}(\boldsymbol{s}_1,\boldsymbol{s}_2).$$

Theorem 1. *Under Assumptions 1 – 9 given in the SI, as $n_{ij}\to\infty$,*

$$\sqrt{n_{ij}}\,\{\hat{\mu}_{ij}(\boldsymbol{s})-\mu_{ij}(\boldsymbol{s})\} \xrightarrow{d} N(0,\tilde{\sigma}^2_{ij}(\boldsymbol{s},\boldsymbol{s})) \quad \textit{and} \quad \sqrt{n_{ij}}\,\{\hat{\mu}_{ij}(\mathcal{A})-\mu_{ij}(\mathcal{A})\} \xrightarrow{d} N(0,\tilde{\sigma}^2_{ij}(\mathcal{A},\mathcal{A})),$$

where $\tilde{\sigma}^2_{ij}(\boldsymbol{s},\boldsymbol{s}) = \gamma_{ij}(\boldsymbol{s},\boldsymbol{s}) + \lambda_{ii,j}(\boldsymbol{s},\boldsymbol{s})$ *and* $\tilde{\sigma}^2_{ij}(\mathcal{A},\mathcal{A}) = \gamma_{ij}(\mathcal{A},\mathcal{A}) + \lambda_{ii,j}(\mathcal{A},\mathcal{A})$.

We note that the bias commonly associated with the kernel estimation vanishes in Theorem 1 due to the under-smoothing entailed under Assumption 8 as elaborated in the SI.

To compare the adjusted averages between two years i_1 and i_2 or between two different regions $\mathcal{A}$ and $\mathcal{B}$, we need to derive the asymptotic distributions of $\hat{\mu}_{i_2 j}(\mathcal{A}) - \hat{\mu}_{i_1 j}(\mathcal{A})$ and $\hat{\mu}_{ij}(\mathcal{A}) - \hat{\mu}_{ij}(\mathcal{B})$, respectively. The following theorem provides the needed results. Define

$$\phi_{i_1 i_2,j}(\boldsymbol{s}_1,\boldsymbol{s}_2) = \lambda_{i_1 i_1,j}(\boldsymbol{s}_1,\boldsymbol{s}_2) + \lambda_{i_2 i_2,j}(\boldsymbol{s}_1,\boldsymbol{s}_2) - \lambda_{i_1 i_2,j}(\boldsymbol{s}_1,\boldsymbol{s}_2) - \lambda_{i_2 i_1,j}(\boldsymbol{s}_1,\boldsymbol{s}_2) \text{ and}$$

$$\phi_{i_1 i_2,j}(\mathcal{A},\mathcal{B}) = |\mathcal{A}|^{-1}|\mathcal{B}|^{-1} \sum_{\boldsymbol{s}_1\in\mathcal{A},\boldsymbol{s}_2\in\mathcal{B}} \phi_{i_1 i_2,j}(\boldsymbol{s}_1,\boldsymbol{s}_2).$$

Theorem 2. *Under Assumptions 1 – 9 in the SI, (i) for* $i_1 \neq i_2$, *as* $n_{i_1 j}, n_{i_2 j} \to \infty$,

$$\sqrt{n_{i_1 j}}[\{\hat{\mu}_{i_2 j}(\mathcal{A}) - \hat{\mu}_{i_1 j}(\mathcal{A})\} - \{\mu_{i_2 j}(\mathcal{A}) - \mu_{i_1 j}(\mathcal{A})\}] \xrightarrow{d} N(0, \tilde{\sigma}^2_{i_2 i_1,j}(\mathcal{A})),$$

where $\tilde{\sigma}^2_{i_2 i_1,j}(\mathcal{A}) = \sum_{p=1}^2 \gamma_{i_p j}(\mathcal{A},\mathcal{A}) + \phi_{i_1 i_2,j}(\mathcal{A},\mathcal{A})$*; and (ii) for* $\mathcal{A}\cap\mathcal{B} = \emptyset$, *as* $n_{ij} \to \infty$,

$$\sqrt{n_{ij}}[\{\hat{\mu}_{ij}(\mathcal{A}) - \hat{\mu}_{ij}(\mathcal{B})\} - \{\mu_{ij}(\mathcal{A}) - \mu_{ij}(\mathcal{B})\}] \xrightarrow{d} N(0, \tilde{\sigma}^2_{ij}(\mathcal{A}-\mathcal{B})),$$

where $\tilde{\sigma}^2_{ij}(\mathcal{A}-\mathcal{B}) = \tilde{\sigma}^2_{ij}(\mathcal{A},\mathcal{A}) - 2\tilde{\sigma}^2_{ij}(\mathcal{A},\mathcal{B}) + \tilde{\sigma}^2_{ij}(\mathcal{B},\mathcal{B})$ *with* $\tilde{\sigma}^2_{ij}(\mathcal{A},\mathcal{B}) = \gamma_{ij}(\mathcal{A},\mathcal{B}) + \lambda_{ii,j}(\mathcal{A},\mathcal{B})$.

The asymptotic normality in Theorem 2 allows us to assess the statistical significances in spatial and temporal differences of the adjusted averages in the empirical study in Section 6.

5 Variance estimation and hypothesis testing

Since the asymptotic variance in Theorem 2 is quite involved, we propose a bootstrap procedure to obtain their estimation. To begin with, it may be shown that as $n_{ij} \to \infty$,

$$\hat{\mu}_{ij}(\boldsymbol{s}) - \mu_{ij}(\boldsymbol{s}) = T_{ij,1}(\boldsymbol{s}) + T_{ij,2}(\boldsymbol{s}) + o_P(n_{ij}^{-1/2}),$$

where the two leading terms that determine the asymptotic variance are

$$T_{ij,1}(\boldsymbol{s}) = \int \{\hat{m}_{ij}(\boldsymbol{x},\boldsymbol{s}) - m_{ij}(\boldsymbol{x},\boldsymbol{s})\} dF_{\cdot j}(\boldsymbol{x}) \quad \text{and}$$

$$T_{ij,2}(\boldsymbol{s}) = \int m_{ij}(\boldsymbol{x},\boldsymbol{s}) d\{\hat{F}_{\cdot j}(\boldsymbol{x}) - F_{\cdot j}(\boldsymbol{x})\}$$
$$= S^{-1} \sum_{\boldsymbol{s}' \in \mathcal{W}} \sum_{a=1}^{A_j} n_{aj}^{-1} \sum_{t=1}^{n_{aj}} \left[n_{aj} \left(\sum_{a=1}^{A_j} n_{aj} \right)^{-1} m_{ij}\{\boldsymbol{X}_{ajt}(\boldsymbol{s}'), \boldsymbol{s}\} - A_j^{-1} \int m_{ij}(\boldsymbol{x},\boldsymbol{s}) dF_{aj}(\boldsymbol{x},\boldsymbol{s}') \right],$$

where $F_{\cdot j}(\boldsymbol{x})$ and $F_{aj}(\boldsymbol{x},\boldsymbol{s}')$ are the distributions matched to $f_{\cdot j}(\boldsymbol{x})$ and $f_{aj}(\boldsymbol{x},\boldsymbol{s}')$, respectively.

The forms of $T_{ij,1}(\boldsymbol{s})$ and $T_{ij,2}(\boldsymbol{s})$ suggest a bootstrap strategy that combines the temporal block bootstrap [Carlstein (1986)] on meteorological data with the wild bootstrap [Liu (1988); Härdle and Mammen (1993)] that resamples the residuals of the regression model (2.2). In order to keep the spatial dependence in the residuals, we resample estimators of residual vectors $\{\boldsymbol{e}_{ijt}\}_{t=1}^{n_{ij}}$. An underlying reason for separating the temporal and spatial bootstrap is that the temporal dependence has negligible contributions to the variance of $T_{ij,1}(\boldsymbol{s})$ due to the whitening effect of the kernel smoothing which retains the leading order term as shown in Kreiss et al. (2008).

The temporal dependence in $T_{ij,2}(\boldsymbol{s})$ is handled by the temporal block bootstrap method. To this end, we combine meteorological data from all sites to form the time series $[\mathbb{X}_{ijt} = \{\boldsymbol{X}_{ijt}(\boldsymbol{s}_1)^{\mathrm{T}}, \ldots, \boldsymbol{X}_{ijt}(\boldsymbol{s}_L)^{\mathrm{T}}\}^{\mathrm{T}},\ t = 1, \ldots, n_{ij}]$ in season j and year i. Define $\boldsymbol{B}_1 = (\mathbb{X}_{ij1}^{\mathrm{T}}, \ldots, \mathbb{X}_{ijl}^{\mathrm{T}})^{\mathrm{T}}, \ldots,$ $\boldsymbol{B}_{n_{ij}-l+1} = (\mathbb{X}_{ij,n_{ij}-l+1}^{\mathrm{T}}, \ldots, \mathbb{X}_{ij,n_{ij}}^{\mathrm{T}})^{\mathrm{T}}, \boldsymbol{B}_{n_{ij}-l+2} = (\mathbb{X}_{ij,n_{ij}-l+2}^{\mathrm{T}}, \ldots, \mathbb{X}_{ij,n_{ij}}^{\mathrm{T}}, \mathbb{X}_{ij,1}^{\mathrm{T}})^{\mathrm{T}}, \ldots, \boldsymbol{B}_{n_{ij}} = (\mathbb{X}_{ij,n_{ij}}^{\mathrm{T}},$ $\mathbb{X}_{ij,1}^{\mathrm{T}}, \ldots, \mathbb{X}_{ij,l-1}^{\mathrm{T}})^{\mathrm{T}}$ as a series of circular moving blocks [Davison and Hinkley (1997)] with length l, which makes every observation have the same chance to be selected in resampled data. We choose $l = 12$ (hours) based on experience with the data. For the b-th replication, we randomly sample n_{ij}/l blocks from $\{\boldsymbol{B}_t\}_{t=1}^{n_{ij}}$ with replacement and combine them to obtain a resampled weather series $[\mathbb{X}_{ijt}^{*b} = \{\boldsymbol{X}_{ijt}^{*b}(\boldsymbol{s}_1)^{\mathrm{T}}, \ldots, \boldsymbol{X}_{ijt}^{*b}(\boldsymbol{s}_L)^{\mathrm{T}}\}^{\mathrm{T}}, t = 1, \ldots, n_{ij}]$ for season j and year i.

To generate bootstrap samples of the response variables $Y_{ijt}^*(\boldsymbol{s})$, we still need to resample the standardized residuals, whose core idea is the wild bootstrap or the regression bootstrap advocated by Liu (1988) and Kreiss et al. (2008). Given the estimated regression function $\hat{m}_{ij}\{\boldsymbol{X}_{ijt}(\boldsymbol{s}), \boldsymbol{s}\}$ in (4.1), the conditional variance $\sigma_{ij}^2\{\boldsymbol{X}_{ijt}(\boldsymbol{s}), \boldsymbol{s}\}$ can be estimated by applying the kernel smoothing

approach on $\hat{\epsilon}^2_{ijt}(\boldsymbol{s}) = [Y_{ijt}(\boldsymbol{s}) - \hat{m}_{ij}\{\boldsymbol{X}_{ijt}(\boldsymbol{s}), \boldsymbol{s}\}]^2$, so that for $\boldsymbol{x} = (\boldsymbol{z}^{\mathrm{T}}, w)^{\mathrm{T}}$,

$$\hat{\sigma}^2_{ij}(\boldsymbol{x}, \boldsymbol{s}) = \frac{\sum_{t=1}^{n_{ij}} \tilde{K}_{\tilde{\boldsymbol{H}}}\{\boldsymbol{z} - \boldsymbol{Z}_{ijt}(\boldsymbol{s})\}\hat{\epsilon}^2_{ijt}(\boldsymbol{s}) I\{W_{ijt}(\boldsymbol{s}) = w\}}{\sum_{t=1}^{n_{ij}} \tilde{K}_{\tilde{\boldsymbol{H}}}\{\boldsymbol{z} - \boldsymbol{Z}_{ijt}(\boldsymbol{s})\} I\{W_{ijt}(\boldsymbol{s}) = w\}}. \tag{5.1}$$

Here the bandwidths are selected afresh by applying the cross-validation method. This leads to the estimators of standardized residuals

$$\hat{e}_{ijt}(\boldsymbol{s}) = \hat{\epsilon}_{ijt}(\boldsymbol{s}) / \hat{\sigma}_{ij}\{\boldsymbol{X}_{ijt}(\boldsymbol{s}), \boldsymbol{s}\} \tag{5.2}$$

and $\hat{\boldsymbol{e}}_{ijt} = \{\hat{e}_{ijt}(\boldsymbol{s}_1), \ldots, \hat{e}_{ijt}(\boldsymbol{s}_L)\}^{\mathrm{T}}$. Let

$$\hat{\boldsymbol{\Sigma}}_{ij} = n_{ij}^{-1} \sum_{t=1}^{n_{ij}} \hat{\boldsymbol{e}}_{ijt} \hat{\boldsymbol{e}}_{ijt}^{\mathrm{T}} - \left(n_{ij}^{-1} \sum_{t=1}^{n_{ij}} \hat{\boldsymbol{e}}_{ijt} \right) \left(n_{ij}^{-1} \sum_{t=1}^{n_{ij}} \hat{\boldsymbol{e}}_{ijt} \right)^{\mathrm{T}}.$$

We generate resamples of the standardized residual by $\hat{\boldsymbol{e}}^{*b}_{ijt} \overset{iid}{\sim} N_L(\boldsymbol{0}, \hat{\boldsymbol{\Sigma}}_{ij})$, which together with the resampled weather process lead to the resampled responses

$$Y^{*b}_{ijt}(\boldsymbol{s}) = \hat{m}_{ij}\{\boldsymbol{X}^{*b}_{ijt}(\boldsymbol{s}), \boldsymbol{s}\} + \hat{\sigma}_{ij}\{\boldsymbol{X}^{*b}_{ijt}(\boldsymbol{s}), \boldsymbol{s}\} \hat{\boldsymbol{e}}^{*b}_{ijt}(\boldsymbol{s}), \tag{5.3}$$

for $t = 1, \cdots, n_{ij}$. We re-compute the adjusted average for each bootstrap replication by

$$\hat{\mu}^{*b}_{ij}(\boldsymbol{s}) = S^{-1} \left(\sum_{a=1}^{A_j} n_{aj} \right)^{-1} \sum_{w=1}^{5} \sum_{\boldsymbol{s}' \in \mathcal{W}} \sum_{a=1}^{A_j} \sum_{t=1}^{n_{aj}} \hat{m}^b_{ij}\{\boldsymbol{Z}^{*b}_{ajt}(\boldsymbol{s}'), w, \boldsymbol{s}\} I\{W^{*b}_{ajt}(\boldsymbol{s}') = w\} \quad \text{and}$$

$$\hat{\mu}^{*b}_{ij}(\mathcal{A}) = |\mathcal{A}|^{-1} \sum_{\boldsymbol{s} \in \mathcal{A}} \hat{\mu}^{*b}_{ij}(\boldsymbol{s}).$$

The bootstrap standard deviations of $\hat{\mu}_{ij}(\boldsymbol{s})$, $\hat{\mu}_{ij}(\mathcal{A})$, $\hat{\mu}_{i_2 j}(\mathcal{A}) - \hat{\mu}_{i_1 j}(\mathcal{A})$ and $\hat{\mu}_{ij}(\mathcal{A}) - \hat{\mu}_{ij}(\mathcal{B})$ can be obtained via Monte-Carlo simulation, which are denoted as $\hat{\hat{\sigma}}_{ij}(\boldsymbol{s}, \boldsymbol{s}), \hat{\hat{\sigma}}_{ij}(\mathcal{A}, \mathcal{A}), \hat{\hat{\sigma}}_{i_2 i_1, j}(\mathcal{A})$ and $\hat{\hat{\sigma}}_{ij}(\mathcal{A} - \mathcal{B})$, respectively. These standard errors together with the spatial and temporal differences in the adjusted averages are used for assessing changes in regional air quality.

For testing the yearly difference hypotheses $H_0 : \mu_{i_2 j}(\mathcal{A}) = \mu_{i_1 j}(\mathcal{A})$ versus $H_1 : \mu_{i_2 j}(\mathcal{A}) > (<) \mu_{i_1 j}(\mathcal{A})$, we use the test statistic $\{\hat{\mu}_{i_2 j}(\mathcal{A}) - \hat{\mu}_{i_1 j}(\mathcal{A})\} / \hat{\hat{\sigma}}_{i_2 i_1, j}(\mathcal{A})$. For detecting any regional difference, we consider testing $H_0 : \mu_{ij}(\mathcal{A}) = \mu_{ij}(\mathcal{B})$ versus $H_1 : \mu_{ij}(\mathcal{A}) > (<) \mu_{ij}(\mathcal{B})$ with the test statistic $\{\hat{\mu}_{ij}(\mathcal{A}) - \hat{\mu}_{ij}(\mathcal{B})\} / \hat{\hat{\sigma}}_{ij}(\mathcal{A} - \mathcal{B})$. Both statistics asymptotically follow the standard normal distribution by Theorems 1 and 2, which allows obtaining the p-values for statistical significance.

6 Application to Beijing's air pollution data

6.1 Model diagnostics

We conducted diagnostics to the nonparametric model (2.2) by first carrying out the nonparametric kernel estimation for the regression function given in (4.1). As a first diagnostic check, Table S1 of the SI provides the fitted R^2 for six pollutants from Spring 2013 to Winter 2016. It shows that the R^2s were mostly above 70%, indicating reasonable fit of Model (2.2).

An important notion in modeling dependence of spatial data is the semi-variogram [Cressie (1993)], which we will outline in the context of analyzing the standardized residual process. Under the temporal stationarity assumption, the semi-variogram function for the standardized residual process $\{e_{ijt}(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{R}\}$ at hour t of year i and season j is

$$\gamma_{ij}(\boldsymbol{s}, \boldsymbol{s}') = 2^{-1}\mathrm{E}\big[\{e_{ijt}(\boldsymbol{s}) - e_{ijt}(\boldsymbol{s}')\}^2\big], \quad \textbf{for any } \boldsymbol{s}, \boldsymbol{s}' \in \mathcal{R}. \tag{6.1}$$

The process $\{e_{ijt}(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{R}\}$ is said to be spatially stationary if $\gamma_{ij}(\boldsymbol{s}, \boldsymbol{s}') = \gamma_{ij}(\boldsymbol{s} - \boldsymbol{s}')$ and isotropic if $\gamma_{ij}(\boldsymbol{s}, \boldsymbol{s}') = \gamma_{ij}(\|\boldsymbol{s} - \boldsymbol{s}'\|)$ by a slight abuse of notation, where $\|\cdot\|$ denotes the Euclidean norm. The semi-variogram for other processes such as the $PM_{2.5}$ and SO_2 processes can be similarly defined.

Specifically, $\gamma_{ij}(0)$ is called the nugget effect, which denotes the variability that cannot be explained by the spatial correlation. The nugget effect is caused by measurement errors, and requires densely populated sites in order to be estimated accurately. Under the isotropic assumption, as h increases, $\gamma_{ij}(h)$ would gradually increase initially and then level off beyond a distance, commonly called the range. The value of the semi-variogram at the range is called the sill. Any two sites with a distance larger than the range would have no spatial dependence.

Under the temporal stationarity assumption, the semi-variograms between any

two sites $\boldsymbol{s}_{l_1}$ and $\boldsymbol{s}_{l_2}$ is estimated by, in the case of the standardized residual process,

$$\hat{\gamma}_{ij}(\|\boldsymbol{s}_{l_1} - \boldsymbol{s}_{l_2}\|) = \frac{1}{2n_{ij}} \sum_{t=1}^{n_{ij}} \{\hat{e}_{ijt}(\boldsymbol{s}_{l_1}) - \hat{e}_{ijt}(\boldsymbol{s}_{l_2})\}^2, \text{ for } l_1, l_2 = 1, \ldots, L, \tag{6.2}$$

which are represented by dots in Figures 2 and S1.

Motivated by Jun and Stein (2004), to gain information on the large scale spatial dependence, we display in Figure S1 of the SI the semi-variograms for the raw pollution readings, the fitted values by nonparametric regression, and the estimated residuals $\hat{\epsilon}_{ijt}(s) = Y_{ijt}(\boldsymbol{s}) - \hat{m}_{ij}\{\boldsymbol{X}_{ijt}(\boldsymbol{s}), \boldsymbol{s}\}$ for $PM_{2.5}$, SO_2, NO_2 and the 8-hour O_3 for the summers and winters of 2015 and 2016, respectively.

Figure S1 shows that the semi-variograms for the raw $PM_{2.5}$ and SO_2 displayed stronger non-stationarity and longer-range dependence, while those for the raw NO_2 and O_3 were relatively flatter even at larger distances. The latter revealed weaker spatial dependence for NO_2 and O_3 due to their shorter life expectancy as both gases are more chemically reactive, and hence cannot travel afar. The figure also shows that the semi-variograms of the raw pollutants were closely imitated by those of the fitted values, which indicated reasonable fitting performance of the kernel regression approach from the aspect of spatial dependence. The semi-variograms for the estimated residuals show much weaker dependence, which demonstrates the ability of the regression models in picking up the large scale trend and variation in observed concentrations.

Figure 2 presents **the locally estimated scatterplot smoothing** (LOESS) estimates [Cleveland and Devlin (1988)] of semi-variograms, which essentially smooth $\hat{\gamma}_{ij}(h)$ with respect to the distance h for $PM_{2.5}$, SO_2, NO_2 and 8-hour O_3 (12 noon to 7 pm). These LOESS fitted curves show there was no much spatial dependence beyond 20 kms in majority of the plots as the semi-variograms ceased to increase significantly after 20 kms, indicating Model (2.2) captures the main aspects of the spatial dependence. Figure S2 of the SI provides the autoregression functions of the standardized residuals of $PM_{2.5}$ at three monitoring sites and the corresponding long-run covariance function, showing that summer tended to have stronger temporal dependence than that of the other three **seasons** which is likely induced by Beijing's rather static weather pattern in sum-

mer. These semi-variograms in Figure 2 had comparable shapes, magnitudes and nugget effects, indicating the innovation processes $\{e_{ijt}(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{R}\}$ for the four pollutants share some common features of spatial dependence.

6.2 Concentration maps

We adopt the adjustment method to obtain $\hat{\mu}_{ij}(\boldsymbol{s})$, the estimated adjusted averages at all 28 monitoring sites for each season and year. Then we conduct the spatial kernel smoothing of the adjusted averages $\hat{\mu}_{ij}(\boldsymbol{s})$ over the study region with a bivariate productive Gaussian kernel and a smoothing bandwidth $h = 0.15$ degrees in latitude and longitude. These give rise to the seasonal concentration maps of the air pollutants for each season and year. Figure 3 display those of $PM_{2.5}$ and NO_2 from 2013 to 2016, while those SO_2 and 8-hour O_3 are provided in Figures S3 in the SI.

Figures 3 and S3 show that $PM_{2.5}$, SO_2 and NO_2 concentrations share a similar seasonal pattern of high winter and low summer with those of fall and spring situated in between. The 8-hour O_3 has a reversed seasonality such that the summer and spring were the high seasons, and winter and fall were the low seasons. This is because the photo-chemical process that governs the ground level ozone generation requires ultra-violet light (u.v.) from the sun, which is the reason for the consideration of 8-hour O_3 from 12 noon to 7 pm, the period when the O_3 concentration tended to be the highest.

On top of these seasonal patterns, $PM_{2.5}$ exhibited large spatial variations with the southern part of Beijing having much higher concentrations than the other areas, especially in the severely polluted winter season. The spatial variations of other three pollutants were much less than that of $PM_{2.5}$. Figure 3 displays elevated circular ridges of NO_2 over the city center, which were the most evident in 2014 and still quite noticeable in 2016. This was largely due to the motor vehicle emissions of NO and NO_2, especially under traffic congestion that Beijing is famous for. The peaks of the circular ridges were situated in the east part of the city between the Third and Fourth Ring

Roads, which coincides with the most congested area of the capital. The circular shape of the NO_2 distribution reflected the fact that NO_2 can not be transported afar due to its much shorter life expectancy as it is highly chemically reactive.

A close inspection of the ozone concentration map in Figure S3 shows the low concentration basin was in the area where NO_2 was high. The basin was the most apparent in the summer and spring of 2014. This trade-off between NO_2 and O_3 was a result of a chemical reaction equation $NO_2 + O_2 \overset{\text{u.v.}}{\rightleftarrows} NO + O_3$. As emissions from motor vehicles are primarily NO (and CO), the equation implies that direct emission of NO consumes O_3 for the generation of NO_2, which explains the trade-off. Of course, the inverse reaction is also valid under the condition of the ultra-violet (u.v.) radiation, which explains why the O_3 level is the highest in the afternoon and in summer.

Figure 3 reveals temporal reductions of $PM_{2.5}$ from 2014 to 2015 especially in summer and fall. However, it is hard to detect noticeable improvement from 2015 to 2016. In contrast, the figure demonstrates clear reductions in SO_2 from 2014 to 2016 in all seasons. Confirmations of the reduction being statistically significant will be made in the following subsection when we carry out inference for air-quality measures. In contrast, the improvement, if any, in NO_2 and O_3 in Figures 3 and S3 was rather unclear, and needs formal confirmation via statistical testing, again in the following subsection.

6.3 Regional air-quality assessment

In this section, we utilize the tests outlined in Section 5 to conduct assessments on the yearly and regional differences in air pollution levels. We focus on the temporal differences $\mu_{i_2 j}(\mathcal{A}) - \mu_{i_1 j}(\mathcal{A})$ for $i_2 = i_1 + 1$, and the spatial differences $\mu_{ij}(\mathcal{A}) - \mu_{ij}(\mathcal{B})$ for $\mathcal{A}$ being the Southern and $\mathcal{B}$ the Central areas, respectively.

Figure 4 displays the seasonal average concentrations in the Central and Southern areas for $PM_{2.5}$, SO_2, NO_2 and 8-hour O_3 from 2013 to 2016. It shows that the average seasonal $PM_{2.5}$

levels were persistently higher than 35$\mu g/m^3$, the Interim target-1 limit set by the World Health Organization (WHO). The NO_2 average concentrations in the Central area were also consistently over the WHO limit of 40$\mu g/m^3$ set in all seasons, while SO_2 and O_3 were relatively better with SO_2 exceeding the WHO limit only in the winter and O_3 in the summer of the last two years. The Southern area had much higher $PM_{2.5}$, but less NO_2 than the Central area. That the Southern area has higher $PM_{2.5}$ reflected the transported fine particulate matters from the heavy industrial Hebei Province. Meanwhile, the high NO_2 in the Central area was attributed to much elevated emission from motor vehicles due to congested traffic, since more than 70% of the population in the capital are resided in the Central area.

To formally check whether there were significant differences between the two areas, Table 1 gives Southern minus Central averages along with their standard errors and p-values for testing against the Southern being higher than the Central for $PM_{2.5}$, SO_2, and 8-hour O_3, and the opposite for NO_2. Numerical figures in the table reveal southern-high, central-low pattern for SO_2 and 8-hour O_3, which are not that visible in Figure 4. The table reports the p-values in four categories: those larger than 0.01, those in $(10^{-9}, 0.01]$ marked with one $*$, $(10^{-16}, 10^{-9}]$ attracting two $*$s, and those smaller than 10^{-16} with three $*$s. Our deliberate using rather smaller p-value ranges was to account for multiplicity from testing hypotheses over the 16 seasons and 4 pollutants. If counted by the number of p-values with two or three $*$s out of the total of 16, the Southern-Central difference was the most significant for NO_2 (15 out 16), followed by $PM_{2.5}$ (11 out of 16) and O_3 (9 out of 16). SO_2 exhibited the least difference with 3 out of 16 having no $*$, which were the highest among the four pollutants.

To gain information on yearly changes in air quality, we took differences in the adjusted averages between consecutive years and then employed the proposed spatio-temporal bootstrap approach to obtain standard errors and p-values for significances for four pollutants, which are displayed in Figure 5. It is observed that the temporal differences were much less significant than

the regional differences as reported in Table 1. Indeed, there were only 2 significant yearly changes out of 16 for $PM_{2.5}$ in Central, and 7 out of 16 for the Southern area, and those significant p-values were largely at one $*$ level indicating rather weak change if at all. Among the four pollutants, SO_2 was the most significant in both areas, followed by 8-hour O_3 and NO_2. $PM_{2.5}$ had the least changes in the last four years among the four pollutants.

The reduction in SO_2 in Years 2014 and 2015 was very substantial and encouraging, and was a bright spot for Beijing's air-quality management in the last four years. Comparing with SO_2, the other three pollutants' performances in the last four years were rather lackluster. It is clear that 2015 was the year that had the most significant improvement as reflected by significant reductions in $PM_{2.5}$, SO_2 and NO_2. This was largely linked to the economic slowing down in the last economic cycle. Although $PM_{2.5}$ was significantly reduced in 2015 at the 1% significance level, there was no improvement in 2016 but an insignificant increase in the Central area. This was worrying as it showed the reduction in the fine particular matters was in a stalemate in 2016. The same stalemate was also observed in NO_2 and O_3 in 2016.

The above assessments indicated the challenge faced by Beijing's air-quality management. While the SO_2 level has been reduced significantly, it has not translated to a continued $PM_{2.5}$ reduction at the time. Our analysis suggests an urgent need to reduce the NO_2 level caused by the motor vehicle emission in order to find a new driving force for $PM_{2.5}$'s decline. Cutting back NO_2 will also reduce the level of O_3 which has been on a rising curve in the broader Beijing-Tianjin-Hebei region as shown in Chen et al. (2018).

We also compared the proposed adjustment method with two existing methods. One was the trend analysis method given in Thompson et al. **(2001) which we have outlined in Section 3, and the other was the three-year moving average method advocated by US Environmental Protection Agency (EPA). Details and the problems with the moving average method had been documented in Chen** et al. **(2018). Figure**

6 displays the average concentrations of $PM_{2.5}$ and SO_2 using the three methods for Central and Southern areas. The plots for NO_2 and 8-hour O_3 are provided in Figure S4 in the SI. Both figures show substantial differences between the proposed approach and the other two methods especially in winter and fall seasons for $PM_{2.5}$, SO_2 and NO_2. Table 2 reports numerical average differences between the proposed approach and the other two methods, respectively. It reveals that the average differences in winter were more than $6\mu g/m^3$ and $10\mu g/m^3$ for $PM_{2.5}$ in the two areas, respectively, which showed substantial amounts of annual improvement in these four pollutants in the last a few winters in Beijing.

7 Discussion

We have proposed a spatial and temporal adjustment method for objectively assessing air quality in a region that removes meteorological confounding and produces spatially and temporally comparable air-quality estimates. The method is able to quantify underlying changes in the emission, which would be much more time consuming to measure based on the emission inventory method. **We have established the theoretical properties of the air-quality measures, and have utilized them for a comprehensive evaluation on air quality in a region around Beijing by analyzing the pattern and trend for the major air pollutants. The theoretical justification along with the simulation experiments provides the necessary guarantee for the performance of the adjustment method.**

The study reported in this paper focuses on the region of Beijing, where the air-quality monitoring sites and the meteorological stations are relatively close to each other. If they are far apart, we can use the spatial kriging method to impute meteorological variables at each air-quality site for the purpose of estimating the

regression function. However, the meteorological stations are much denser than the Guokong (nationally controlled) air-quality monitoring sites in most cities in China, so this is less of an issue. One may also wonder whether the methodology still works if the air-quality monitoring sites are not close to each other. Generally speaking, as long as the meteorological covariates from the sites share common domains to allow the definition of the spatial and temporal baseline density $f_{\cdot j}(\boldsymbol{x})$, the adjustment method can be carried out with guaranteed performance given the site configuration. Our experience suggests that the proposed approach can be used to assess air quality in quite a large region, for instance the North China Plain (NCP), since the NCP shares common meteorological characteristics.

Our assessment reveals significant reduction in SO_2 while the improvements in $PM_{2.5}$ and NO_2 were much subdued up to early 2017, the end time of the data. There has been an upward trend for the ground level ozone that deserves attention. Although the implementation of the air-quality assessment method is demonstrated using nonparametric regression in the study, a suitable parametric or semiparametric regression model can be used as well.

Acknowledgments

This research is funded by China's National Key Research Special Program Grants 2016YFC0207701, 2016YFC0207702 and 2016YFC0207703, National Key Basic Research Program Grant 2015CB856000, National Natural Science Foundation of China Grants 71532001, 71371016, 11971390 and 11671018, and Beijing Natural Science Foundation Grant Z190001. Chen acknowledges support from LMEQF at Peking University. Guo is supported by the Fundamental Research Funds for the Central Universities JBK1806002, JBK140507 and JBK1902050. Lin's work is supported by Beijing Academy of Artificial Intelligence (BAAI).

References

Alduchov, O. A. and Eskridge, R. E. (1996). Improved magnus form approximation of saturation vapor pressure. Journal of Applied Meteorology *35*, 601–609.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. The Annals of Statistics *14*, 1171–1179.

Chen, L., Guo, B., Huang, J., He, J., Wang, H., Zhang, S. and Chen, S. X. (2018). Assessing air-quality in Beijing-Tianjin-Hebei region: The method and mixed tales of $PM_{2.5}$ and O_3. Atmospheric Environment *193*, 290–301.

Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. Journal of the American Statistical Association *83*, 596–610.

Cressie, N. (1993). Statistics for Spatial Data. John Wiley & Sons, New Jersey.

Davison, A. C. and Hinkley, D. V. (1997). Bootstrap Methods and Their Application. Cambridge University Press, Cambridge.

Fan, J. and Yao, Q. (2003). Nonlinear Time Series: Nonparametric and Parametric Methods. Springer-Verlag, New York.

Finazzi, F., Scott, E. M. and Fassò, A. (2013). A model-based framework for air quality indices and population risk evaluation, with an application to the analysis of Scottish air quality data. Journal of the Royal Statistical Society: Series C *62*, 287–308.

Guo, S., Hu, M., Zamora, M. L., Peng, J., Shang, D., Zheng, J., Du, Z., Wu, Z., Shao, M., Zeng, L. et al. (2014). Elucidating severe urban haze formation in China. Proceedings of the National Academy of Sciences *111*, 17373–17378.

Härdle, W. (1990). Applied Nonparametric Regression. Cambridge University Press, Cambridge.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. The Annals of Statistics *21*, 1926–1947.

Huang, C.-Y., Qin, J. and Follmann, D. A. (2008). Empirical likelihood-based estimation of the treatment effect in a pretest–posttest study. Journal of the American Statistical Association *103*, 1270–1280.

Jun, M. and Stein, M. L. (2004). Statistical comparison of observed and CMAQ modeled daily sulfate levels. Atmospheric Environment *38*, 4427–4436.

Kreiss, J.-P., Neumann, M. H. and Yao, Q. (2008). Bootstrap tests for simple structures in nonparametric time series regression. Statistics and Its Interface *1*, 367–380.

Kuykendal, W. (2017). Emissions Inventory Guidance for Implementation of Ozone and Particulate Matter National Ambient Air Quality Standards (NAAQS) and Regional Haze Regulations. U.S. Environmental Protection Agency, Washington.

Liang, X., Li, S., Zhang, S., Huang, H. and Chen, S. X. (2016). $PM_{2.5}$ data reliability, consistency, and air quality assessment in five Chinese cities. Journal of Geophysical Research: Atmospheres *121*, 10220–10236.

Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., Huang, H. and Chen, S. X. (2015). Assessing Beijing's $PM_{2.5}$ pollution: severity, weather impact, APEC and winter heating. Proceedings of the Royal Society A *471*, 20150257.

Liu, R. Y. (1988). Bootstrap procedures under some non-iid models. The Annals of Statistics *16*, 1696–1708.

Qin, J. (2017). Biased Sampling, Over-identified Parameter Problems and Beyond. Springer, New York.

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

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika *70*, 41–55.

Thompson, M. L., Reynolds, J., Cox, L. H., Guttorp, P. and Sampson, P. D. (2001). A review of statistical methods for the meteorological adjustment of tropospheric ozone. Atmospheric Environment *35*, 617–630.

Zhang, X., Wang, Y., Niu, T., Zhang, X., Gong, S., Zhang, Y. and Sun, J. (2012). Atmospheric aerosol compositions in China: spatial/temporal variability, chemical signature, regional haze distribution and comparisons with global aerosols. Atmospheric Chemistry and Physics *12*, 779–799.

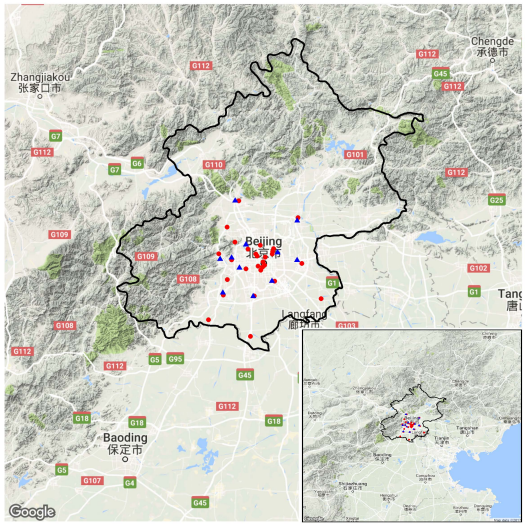

Figure 1: *Locations of air-quality monitoring sites (red circles) and meteorological stations (blue triangles) in the North China Plain portion of Beijing. Insert: the study region within the North China Plain and mountain ranges to the west and north.*

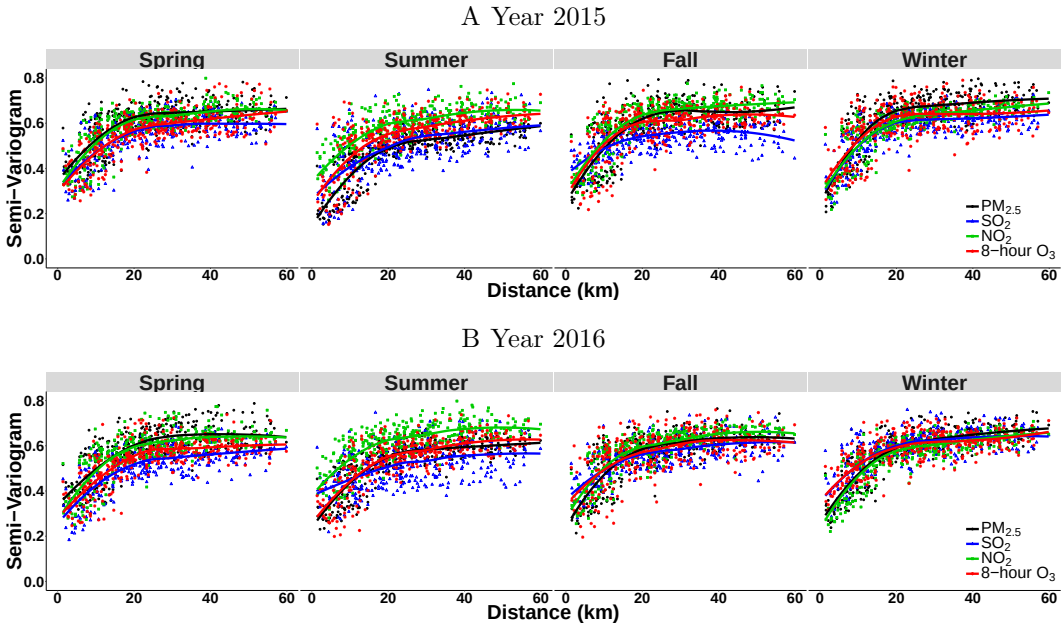

Figure 2: *Semi-variograms of the standardized residuals of the nonparametric model of $PM_{2.5}$ (black), SO_2 (blue), NO_2 (green) and 8-hour O_3 (red) in 2015 (Panel A) and 2016 (Panel B). The dots represent the empirical estimation of semi-variograms. The lines are the smoothed curves of the empirical semi-variograms by the nonparametric LOESS method.*

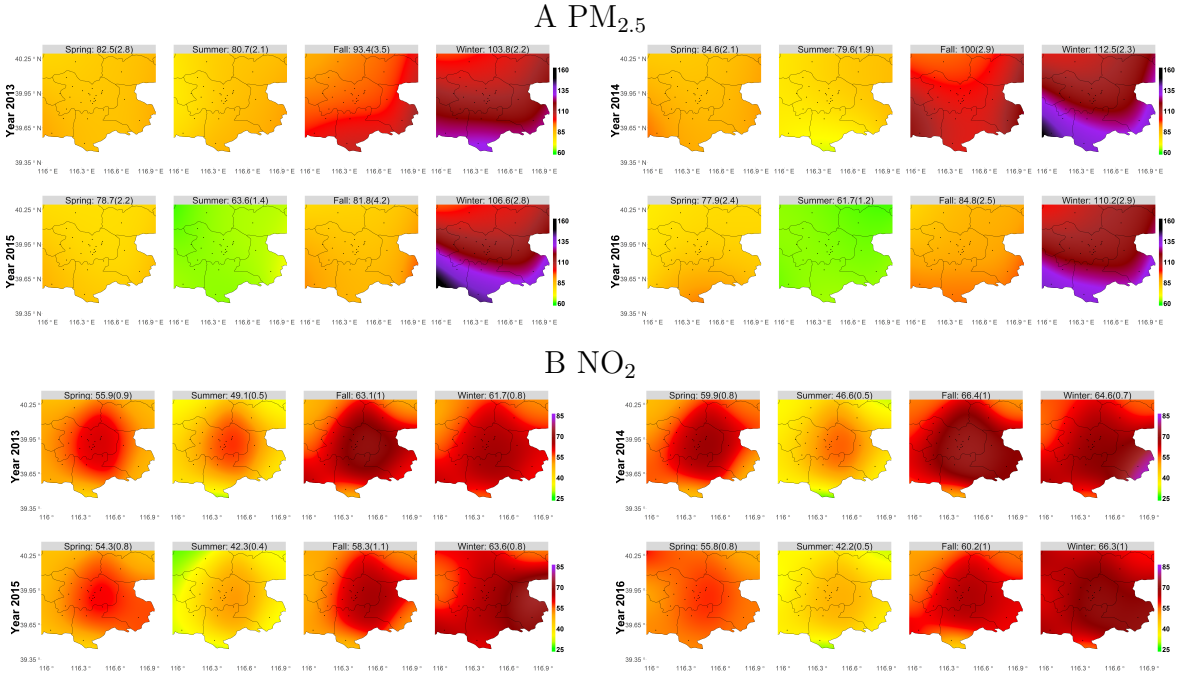

Figure 3: *Seasonal concentration maps of the spatio-temporal adjusted averaged concentration ($\mu g/m^3$) of $PM_{2.5}$ (Panel A) and NO_2 (Panel B) in the urban area of Beijing from Year 2013 to Year 2016. The number above each plot displays the regional adjusted average in Beijing while the number inside the parentheses is the standard error. The smoothing bandwidth used for generating the map is 0.15.*

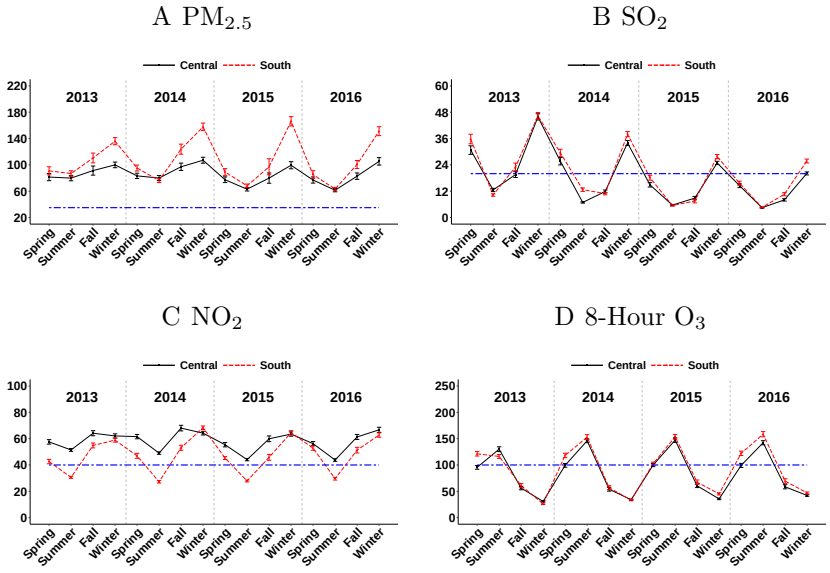

Figure 4: *Seasonally adjusted averages ($\mu g/m^3$) with the bars indicating the 95% confidence intervals. The averages for a region are obtained by averaging the adjusted averages at all sites in the region. The blue dashed line in each figure suggests the standard indicated by the WHO, which are $35\mu g/m^3$ (Interim target-1) for $PM_{2.5}$, $20\mu g/m^3$, $40\mu g/m^3$ and $100\mu g/m^3$ for SO_2, NO_2 and 8-hour O_3, respectively.*

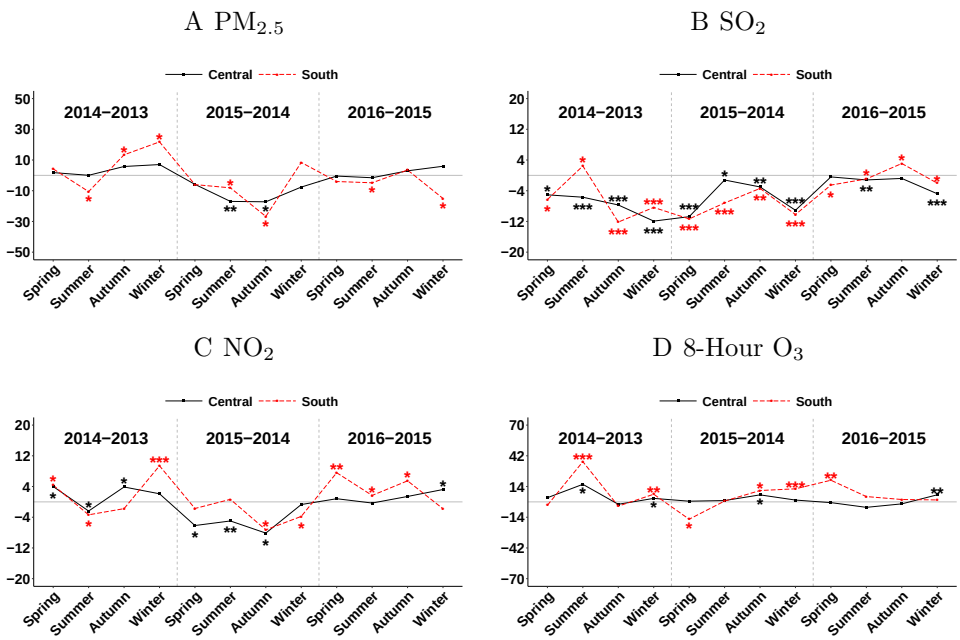

Figure 5: *Yearly differences ($\mu g/m^3$) in the adjusted averages. where the number of $*$ indicates the level of significance in the yearly increase or decrease (*: $10^{-9} \leq$ p-value$< 10^{-2}$; **: $10^{-9} \leq$ p-value$< 10^{-16}$; ***: p-value$< 10^{-16}$).*

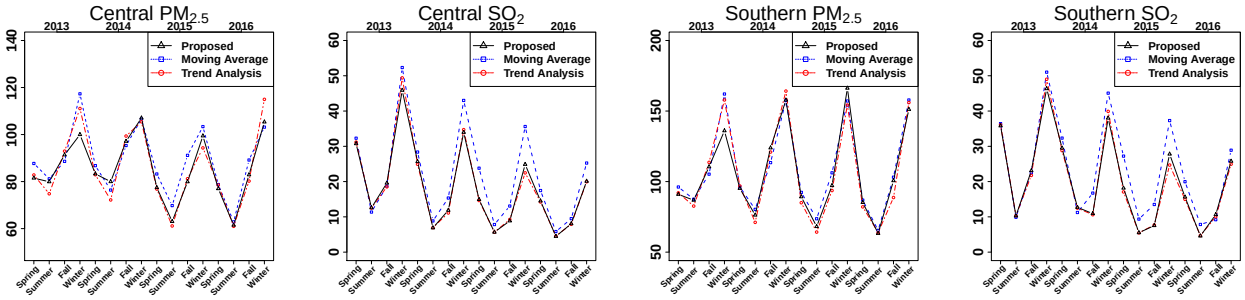

Figure 6: *Seasonally adjusted averages by using our proposed method, three-year moving average and the trend analysis for $PM_{2.5}$ and SO_2 in the Central and Southern areas.*

Table 1: *Regional differences (standard errors, $\mu g/m^3$) of seasonal and annual adjusted averages of $PM_{2.5}$, SO_2, NO_2 and 8-hour O_3 between Southern and Central areas (Southern - Central).*

Pollutant	Season	2013	2014	2015	2016
$PM_{2.5}$	Spring	9.5(1.2)**	11.9(1.1)***	11.7(1.2)***	8.2(1.6)*
	Summer	6.9(1.3)*	-3.7(1.2)*	5.1(0.8)**	1.8(1.0)
	Fall	19.1(2.1)***	26.8(2.4)***	17.2(3.3)*	17.7(1.8)***
	Winter	36.0(1.7)***	50.8(1.8)***	66.8(2.2)***	45.8(2.1)***
	Average	17.9(0.8)***	21.4(0.9)***	25.2(1.1)***	18.4(0.8)***
SO_2	Spring	5.1(0.5)***	3.8(0.5)**	3.2(0.3)***	1.1(0.3)*
	Summer	-2.3(0.3)**	5.8(0.4)***	-0.2(0.1)	0.1(0.1)
	Fall	3.5(0.6)**	-0.9(0.2)*	-1.3(0.3)*	2.6(0.2)***
	Winter	0.5(0.6)	4.0(0.5)***	2.9(0.4)***	5.7(0.3)***
	Average	1.7(0.3)**	3.2(0.2)***	1.2(0.2)**	2.4(0.1)***
NO_2	Spring	-15.0(0.5)***	-14.6(0.7)***	-10.2(0.6)***	-3.4(0.6)**
	Summer	-20.8(0.4)***	-21.8(0.4)***	-16.2(0.3)***	-14.2(0.4)***
	Fall	-9.3(0.7)***	-14.9(0.7)***	-14.1(0.7)***	-10.0(0.7)***
	Winter	-3.2(0.5)**	4.1(0.4)***	1.0(0.5)	-4.0(0.5)**
	Average	-12.1(0.3)***	-11.8(0.3)***	-9.9(0.3)***	-7.9(0.3)***
8-hour O_3	Spring	25.8(1.5)***	18.9(1.0)***	2.5(0.9)*	22.9(1.1)***
	Summer	-13.8(1.6)***	6.9(1.7)*	6.4(1.2)*	16.2(1.5)***
	Fall	4.6(1.2)*	2.9(1.0)*	6.9(1.3)*	10.9(1.1)***
	Winter	-4.5(0.5)***	-0.7(0.6)	9.5(0.7)***	4.9(0.6)***
	Average	3.0(0.6)*	7.0(0.6)***	6.3(0.6)***	13.7(0.6)***

The number of * represents the level of significance for testing the increase or decrease of the annual change between two consecutive years (*: $10^{-9} \leq$ p-value$< 10^{-2}$; **: $10^{-9} \leq$ p-value$< 10^{-16}$; ***: p-value$< 10^{-16}$).

Table 2: *The average absolute differences (standard errors, $\mu g/m^3$) between the moving average, trend analysis and our proposed method for different pollutants of Central and Southern areas in each season.*

Season	Method	Central				Southern			
		$PM_{2.5}$	**SO_2**	**NO_2**	**8-hour O_3**	**$PM_{2.5}$**	**SO_2**	**NO_2**	**8-hour O_3**
Spring	Moving Average	4.3(1.0)	4.0(0.4)	1.8(0.4)	4.1(0.8)	2.8(1.3)	4.2(0.4)	3.6(0.5)	6.3(1.2)
	Trend Analysis	1.0(0.9)	0.4(0.3)	0.8(0.4)	4.6(0.9)	2.5(1.8)	0.7(0.5)	1.6(0.6)	4.6(1.4)
Summer	Moving Average	3.3(0.8)	1.7(0.1)	1.8(0.3)	7.8(1.2)	3.1(0.9)	2.2(0.2)	1.1(0.3)	12.9(1.6)
	Trend Analysis	3.8(0.8)	0.1(0.2)	0.9(0.3)	3.0(1.3)	3.4(1.2)	0.1(0.2)	0.6(0.3)	1.4(3.2)
Fall	Moving Average	5.5(1.5)	2.5(0.2)	1.9(0.5)	1.6(0.8)	6.8(2.3)	3.4(0.3)	2.3(0.6)	3.5(1.2)
	Trend Analysis	2.0(1.3)	0.6(0.3)	0.9(0.5)	2.0(1.0)	5.4(2.4)	0.6(0.6)	2.1(0.7)	3.3(1.5)
Winter	Moving Average	6.0(1.2)	7.8(0.4)	2.4(0.4)	1.7(0.3)	10.5(2.2)	6.1(0.6)	3.5(0.6)	3.1(0.6)
	Trend Analysis	6.9(1.3)	1.7(0.3)	4.5(0.5)	2.4(0.4)	11.3(1.9)	2.1(0.4)	3.3(0.6)	2.0(0.6)
Average	Moving Average	4.8(0.6)	4.0(0.2)	2.0(0.2)	3.8(0.4)	5.8(0.9)	4.0(0.3)	2.6(0.3)	6.5(0.6)
	Trend Analysis	3.4(0.6)	0.7(0.1)	1.3(0.2)	3.0(0.4)	5.6(1.0)	0.9(0.2)	1.9(0.3)	2.8(1.4)