Atmospheric Environment 267 (2021) 118737

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Relative importance of meteorological variables on air quality and role of boundary layer height

Yaxuan Huang [a], Bin Guo [b,*], Haoxuan Sun [c], Huijie Liu [d], Song Xi Chen [e,**]

[a] Yuanpei College, Peking University, Beijing, 100871, China
[b] Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, 611130, China
[c] Center for Data Science, Peking University, Beijing, 100871, China
[d] The Experimental High School Attached To Beijing Normal University, Beijing, 100032, China
[e] School of Mathematical Sciences, Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, 100871, China

HIGHLIGHTS

- Stable orders of meteorological effects on six air pollutants in North China.
- Boundary layer height's effects can be modeled by the surface meteorological variables.
- $PM_{2.5}$, PM_{10}, SO_2 and CO shared a common order of variable importance.

ARTICLE INFO

Keywords:
Air quality assessment
Statistical learning
Meteorological effects
Variable selection

ABSTRACT

To gain insight on the meteorological effects of air pollution, we study the relative importance of surface meteorological variables and boundary layer height (BLH) on six major air pollutants according to the orders of variables being selected in a forward variable selection algorithm. It is found that there was a strong agreement in the orders of relative importance for the major pollutants among six major cities in North China, which implies regularities in the meteorological processes of air pollution in that region. In particular, $PM_{2.5}$, PM_{10}, SO_2 and CO shared a common variable importance order and were mostly impacted by the dew point temperature and air pressure, while the NO_2 and O_3 were mostly influenced by the boundary layer height (BLH) and temperature, respectively. We evaluate the impacts of BLH on the pollution levels given the surface meteorological variables. It is found that BLH can be well modeled by the surface meteorological variables. Thus, air quality assessment without using BLH would also produce adequate results.

1. Introduction

Air pollution is an enduring challenge encountered by many countries including China. A substantial part of China has experienced severe air pollution in the last two decades. Epidemiological studies have shown a high correlation between exposure to air pollutants and human health (Donaldson et al., 1998; Schwartz, 2000; Pope III et al., 2002; Chen et al., 2013); research has also found that air pollution may cause social and economic problems (Xie et al., 2016; Zhang et al., 2017b; Feng et al., 2019).

Mitigation and control for air pollution are needed in these countries, which requires understanding the mechanisms of the air pollution with respect to its main drivers.

Air pollution is impacted by both the underlying emission and the meteorological condition. As one of the two main drivers, it is important to understand the meteorological aspects of air pollution based on the air quality and meteorological data which are quite available in this era of data.

The relationship between meteorological factors and the pollutant concentration has be considered in a range of studies. Tai et al. (2010) studied the correlation of $PM_{2.5}$ and its components with meteorological variables in the United States by applying multiple linear regression and

* Corresponding author. Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, 611130, China.
** Corresponding author. School of Mathematical Sciences and Guanghua School of Management, Peking University, Beijing, 100871, China.
E-mail addresses: huangyaxuan@pku.edu.cn (Y. Huang), guobin@swufe.edu.cn (B. Guo), hxsun@pku.edu.cn (H. Sun), liuhuijie16@pku.edu.cn (H. Liu), songxichen@pku.edu.cn (S.X. Chen).

https://doi.org/10.1016/j.atmosenv.2021.118737
Received 22 July 2021; Received in revised form 15 September 2021; Accepted 15 September 2021
Available online 2 October 2021

found meteorological variables, can explain up to 50% of the variation of $PM_{2.5}$. Liang et al. (2015, 2016) revealed that $PM_{2.5}$ concentration was highly influenced by meteorological conditions in Beijing and found that 75% of the hourly variation of $PM_{2.5}$ was driven by meteorological factors. Shen et al. (2018) studied the monthly $PM_{2.5}$ levels in Beijing's winters during 2010–2017 and found that 81% of the variation can be explained by the first principal component of 850 hPa meridian wind velocity and the relative humidity. Vu et al. (2019) applied the random forest to assess the impact of clean air action on six major pollutants in Beijing by decoupling the impact of meteorology on ambient air quality.

Stafoggia et al. (2019) estimated the daily PM_{10} and $PM_{2.5}$ concentrations in Italy during 2013–2015 using the random forest with similar meteorological variables as well as the planetary boundary layer height (BLH). Choubin et al. (2020) studied the hazard prediction of PM_{10} in Barcelona Province, Spain and found seven critical features among a total of thirteen features, which included topographic wetness index and precipitation.

In an attempt to construct timely statistical measures to reflect the underlying emission based on the air quality data, Liang et al. (2015) and Zhang et al. (2020) proposed meteorologically adjusted pollution measures (average and quantile) under the meteorological baseline distribution constructed from the readily available surface meteorological data. They showed that the meteorologically adjusted pollution measures can reflect the underlying changes in the emission, while the unadjusted raw average pollution levels are subject to meteorological confounding and cannot objectively reflect the underlying emission.

Despite the wide range of studies on the meteorological effects of the air quality motivated from different contexts, there is still a lack of in-depth understanding on the relative importance of the meteorological variables on different air pollutant species. Specifically, given the inter-dependence among the meteorological variables, is there an established order of relative meteorological importance for an air pollutant in a region? An established order of relative importance would imply a common mechanism in the meteorological process of the air pollutant, which would be useful for the modeling and prediction for the air pollutant.

This paper tries to answer the above question by analysing air quality data in six major northern China cities. The orders of the variable importance are determined by the ranks of variables being selected in a step-wise forward selection procedure (Hastie et al., 2009) in the regression of pollution concentration on the meteorological variables. The forward variable selection orders take into account the inter-dependence among the meteorological variables relative to the target pollutant. The study finds that there was a strong agreement in the orders of meteorological variables in affecting the pollution concentration among the six major cities in North China.

In particular, $PM_{2.5}$, PM_{10}, SO_2 and CO shared a common order of meteorological importance led by the dew point temperature, while NO_2 and O_3 were mostly influenced by the boundary layer height (BLH) and air temperature, respectively. These suggest there is regularity in the meteorological processes for the six species of air pollutants in North China.

The planetary boundary layer height (BLH) defines the vertical dispersion property and is known to be influential to the ground-level concentration of air pollutants. Studies have used BLH as a key factor in the formation and evolution of heavy air pollution (Liu et al., 2013; Quan et al., 2014; Tang et al., 2016; Miao and Liu, 2019). A negative correlation between BLH and pollution concentration from observed data was found in Zhang et al. (2009); Miao et al. (2017); Gui et al. (2019). Xiang et al. (2019) revealed a significant negative correlation between BLH and $PM_{2.5}$ in Beijing during the winter heavy pollution. Miao et al. (2021) found that the relationships between planetary boundary layer and $PM_{2.5}$ and O_3 in the summer of Beijing and Shanghai were different with heavy pollution being associated with lower BLH in Beijing, but higher BLH in Shanghai. That higher pollution being associated with higher BLH in Shanghai was due to confounding of transported pollution from warmer inland under westerly wind, which indicates the importance of treating the multivariate meteorological variables collectively.

However, unlike the surface meteorological variables which are readily available, there is a delay of around three months in the BLH data assimilated by the European Centre for Medium-Range Weather Forecasts (ECMWF). There are two pending questions regarding the role of BLH in the air pollution processes. One is how much information contained in BLH can be explained by the surface meteorological variables; and the other is the impacts of BLH on the meteorological adjusted average pollution levels. The study finds that, despite a significant negative correlation with the pollutants, BLH was not highly ranked for the five non-NO_2 species after considering the surface variables in the modeling.

Furthermore, more than 77% of the variation in the BLH can be explained by the surface variables in the six cities. Even for NO_2, where BLH was the most important variable, the effects of BLH on NO_2 was less than 1.24% of the meteorologically adjusted average concentration. Therefore, using only the surface meteorological variables for air quality assessment would produce adequate results.

The paper is organized as follows. Section 2 outlines the data and the study region. Research methods are outlined in Section 3, which include the nonparametric regression, the forward variable selection and the meteorological adjustment. Section 4 reports results of the analysis. Additional information, extra tables and figures are provided in the supplementary materials (SM).

2. Study region and data

The study region contains six major cities in the North China, which include two mega-cities of Beijing and Tianjin, and four cities: Shijiazhuang, Jinan, Zhengzhou and Taiyuan, which are the provincial capitals of Hebei, Shandong, Henan and Shanxi provinces, respectively. The region encompassed by Beijing, Tianjin and the four provinces has been the main battle ground for the air pollution mitigation campaign in China; see Fig. S1 in SM for the geographical location of the six cities. The time range of the study was from March 1, 2013 to February 28, 2021, which contained 8 seasonal years from spring to winter. The unit of our study is season, consisting of spring from March to May, summer from June to August, fall from September to November, and winter from December to February next year.

The air quality data consisted of hourly concentrations of air pollutants ($PM_{2.5}$, PM_{10}, SO_2, NO_2, CO and O_3) from 53 monitoring sites, which are directly managed by the Ministry of Ecology and Environment (MEE). For ozone, we only focused on the period from noon to 7pm (8-h O_3), when the concentrations tended to be the highest. The 53 air quality sites consisted of 11 sites in Beijing and Tianjin, 7 in Shijiazhuang, 8 from Jinan, Zhengzhou and Taiyuan, respectively.

The meteorological data contained hourly surface (2 m) meteorological observations of six variables: air temperature (TEMP) and pressure (PRES), dew point temperature (DEWP), wind direction (WD), cumulative wind speed (IWS), and cumulative precipitation (IPREC). The wind directions were grouped into five broad directions: Northwesterly (NW), Northeasterly (NE), Southwesterly (SW), Southeasterly (SE) and CV, where CV was for the calm (wind speed less than 0.5 m/s) and variable wind. The cumulative wind speed was the summation of the wind speed since a wind direction was established and was reset to zero when there was a change in the direction.

Similarly, the cumulative precipitation referred to the sum of precipitation since the hour when it rained and was reset to zero when there was an hour without precipitation. Data of the above variables were from 12 weather observing stations from China Meteorological Administration (CMA), where 5 of them were located in Beijing, 3 in Tianjin and 1 each in other four cities.

We also considered the reanalyzed BLH data from the ERA5 data set of ECMWF, which contained hourly BLH at a grid size of $0.25^\circ \times 0.25^\circ$.

We matched each CMA site to the nearest grid of the ERA5 data for BLH as well as to the nearest air quality site. The time range of surface meteorological and BLH data was from March 1, 2011 to February 28, 2021. Here, two more years of meteorological data were employed to construct the meteorological baseline in the meteorological adjustment outlined in Section 3.3. The locations of the 53 air quality sites and the meteorological stations are displayed in Fig. S2 in SM.

In additional to the reanalyzed BLH data from ECMWF, there are other approaches to estimate BLH, which include the radiosonde approach by NOAA (Durre and Yin, 2008) and China's CMA (Guo et al., 2016). The BLH from the radiosonde observations have usually two (sometimes four in summer) observations per day. Seidel et al. (2012) analysed BLH measurements from four approaches over the United States and Europe, and found it was suitable to estimate BLH based on the bulk Richardson number (Vogelezang and Holtslag, 1996), which is the commonly used approach adopted by both the radiosonde observation and the reanalysis approach that produced the ECMWF's ERA5 reanalysis BLH dataset. Based on the radiosonde data of CMA, the properties of planetary boundary layer and its relationship with other meteorological variables in China have been studied. Guo et al. (2019) observed a shift in the temporal trend of BLH in China and found BLH was negatively associated with relative humidity, while positively associated with the near-surface temperature. Zhang et al. (2018) studied the thermodynamic stability of planetary boundary layer in China in summer. They found that convective boundary layer (CBL) dominated in summer throughout China (70%), with sometimes neutral boundary layer (NBL, 26%) and stable boundary layer (SBL, 4%) and BLH of CBL and NBL was positively (negatively) associated with near-surface temperature (humidity), whereas no apparent relationship was found for SBL.

Studies (Seidel et al., 2012; Guo et al., 2016) had shown that the BLH estimates from the radiosonde and the ERA Interim, the predecessor of the ERA5, had consistent pattern. Seidel et al. (2012) evaluated the BLH data at the United States and Europe and found that the radiosonde and reanalysis datasets show similar patterns of spatial and seasonal variability though with biases that vary spatially, seasonally and diurnally. Guo et al. (2016) compared the BLH measurements from the radiosonde measurement from China's CMA and the ERA Interim and found good agreement between the two sources. Another aspect is that the BLH data from the ERA5 is at the hourly frequency which matches the frequency of the air quality data. Therefore, we choose the BLH reanalysis data from the ERA5 of ECMWF in our analysis.

3. Methods

We introduce the regression model used in the study, the forward variable selection that determines the order of the meteorological variables' importance and the meteorological adjustment approach.

3.1. Nonparametric regression models

To avoid potential model mis-specification, we consider nonparametric regression modeling of each air pollutant with respect to the meteorological variables. The nonparametric aspect of the model makes it more adaptive to the complex meteorological processes than the linear or nonlinear parametric regression models.

Let $Y_{ijt}(s)$ denote the concentration of a pollutant, say $PM_{2.5}$, at hour t in season j and year i of a monitoring site s, $\mathbf{X}_{ijt}(s)$ be a d-dimensional vector collecting d meteorological variables, and $U_{ijt}(s)$ be the underlying emission leading to the pollution. The following nonparametric model quantifies the relationship between the concentration $Y_{ijt}(s)$, the emission $U_{ijt}(s)$ and the meteorological $\mathbf{X}_{ijt}(s)$:

$$Y_{ijt}(s) = \widetilde{m}_j(U_{ijt}(s), \mathbf{X}_{ijt}(s)) + e_{ijt}(s), \quad t = 1, 2, \dots, n_{ij}, \tag{3.1}$$

where $\widetilde{m}_j(u(s), \mathbf{x}(s)) = E(Y_{ijt}(s) | U_{ijt}(s) = u(s), \mathbf{X}_{ijt}(s) = \mathbf{x}(s))$ is the regression function, $e_{ijt}(s)$ is the residual satisfying $E(e_{ijt}(s)|U_{ijt}(s), \mathbf{X}_{ijt}(s)) = 0$ and has the finite second moment, and n_{ij} is the total number of observations in season j of year i at the site s.

As the emission information is largely not available at the hourly frequency, we have to condition on the observable meteorological variable $\mathbf{X}_{ijt}(s)$ to attain the following observable model

$$Y_{ijt}(s) = m_{ij}(\mathbf{X}_{ijt}(s)) + \varepsilon_{ijt}(s), \quad t = 1, 2, \dots, n_{ij}, \tag{3.2}$$

where $m_{ij}(\mathbf{x}(s)) = E(Y_{ijt}(s)|\mathbf{X}_{ijt}(s) = \mathbf{x}(s))$ is the regression function by taking the conditional expectation on the observable meteorological data only, $\varepsilon_{ijt}(s)$ is the residual satisfying $E(\varepsilon_{ijt}(s)|\mathbf{X}_{ijt}(s)) = 0$, $\mathrm{var}(\varepsilon_{ijt}(s)|\mathbf{X}_{ijt}(s)) = \sigma_{ij}^2(\mathbf{X}_{ijt}(s))$. Despite $U_{ijt}(s)$ is latent, its information has been reflected in the regression function $m_{ij}(\cdot)$ via the yearly index i which was not appeared in Model (3.1). See Liang et al. (2015) and Zhang et al. (2020) for more details on the linkage between Models (3.1) and (3.2).

In Model (3.2), specific parametric form of the regression function $m_{ij}(\cdot)$ is not assumed to allow model flexibility and non-linear pattern in the pollution generation processes. The yearly seasonal regression function $m_{ij}(\cdot)$ can be estimated by the nonparametric kernel estimator (Härdle, 1990). Specifically, let $k(\cdot)$ be a univariate kernel function which is a symmetric probability density function, for instance the Gaussian kernel $k(u) = (2\pi)^{-1/2}\exp(-u^2/2)$, and define the multivariate product kernel

$$K_{\mathbf{h}}(\mathbf{x}) = (h_1 h_2 \cdots h_d)^{-1} k(x_1 / h_1) k(x_2 / h_2) \cdots k(x_d / h_d),$$

where $\mathbf{x} = (x_1, x_2, \dots, x_d)'$ and $\mathbf{h} = (h_1, h_2, \dots, h_d)'$ is a vector of smoothing bandwidths. The kernel estimator of $m_{ij}(\mathbf{x})$ is

$$\widehat{m}_{ij}(\mathbf{x}) = \frac{\sum_{t=1}^{n_{ij}} K_{\mathbf{h}}(\mathbf{x} - \mathbf{X}_{ijt}(s)) Y_{ijt}(s)}{\sum_{t=1}^{n_{ij}} K_{\mathbf{h}}(\mathbf{x} - \mathbf{X}_{ijt}(s))}. \tag{3.3}$$

The smoothing bandwidths $\mathbf{h}$ can be selected by the cross validation method (Härdle, 1990; Liang et al., 2015).

3.2. Forward variable selection

An important question in the air quality assessment is the relative importance of the meteorological variables on the average pollution level through the regression function $m_{ij}(\cdot)$. We consider the forward variable selection procedure for the nonparametric regression model, which selects the most influential variable once at a time that achieves the best fitting performance (Hastie et al., 2009).

To find the most important (the first one to be selected) variable for a pollutant, we start with the univariate model

$$Y_{ijt}(s) = m_{ij}^{(1)}(X_{ijt}^{(1)}(s)) + \varepsilon_{ijt}^{(1)}(s), \quad t = 1, 2, \dots, n_{ij}, \tag{3.4}$$

where $X_{ijt}^{(1)}(s)$ is a candidate covariate, say DEWP or BLH, $m_{ij}^{(1)}(X_{ijt}^{(1)}(s)) = E(Y_{ijt}(s)|X_{ijt}^{(1)}(s))$ is the conditional mean given the covariate and $\varepsilon_{ijt}^{(1)}(s)$ is the residual. It is noted here and throughout this subsection that the function $m_{ij}^{(1)}(\cdot)$ changes with respect to different covariate $X_{ijt}^{(1)}(s)$ and target cities.

Let $\widehat{m}_{ij}^{(1)}(\cdot)$ be the kernel estimate of $m_{ij}^{(1)}(\cdot)$ according to the generic estimator (3.3). The fitting Mean Square Error (MSE) of each variable $X_{ijt}^{(1)}(s)$ at site s for year i and season j is

$$\mathrm{MSE}_{ij}^{(1)}(X_{ijt}^{(1)}(s), s) = \frac{1}{n_{ij}} \sum_{t=1}^{n_{ij}} \left\{ Y_{ijt}(s) - \widehat{m}_{ij}^{(1)}(X_{ijt}^{(1)}(s)) \right\}^2. \tag{3.5}$$

To evaluate the explanation power of the meteorological variables, we use the coefficient of determination R^2 to assess fitting performance of the covariate. For nonparametric regression, Doksum and Samarov (1995) proposed to use

$$R_{ij}^2(1,s)=\frac{\left[\sum_{t=1}^{n_{ij}}\{Y_{ijt}(s)-\overline{Y}_{ij}(s)\}\{\widehat{m}_{ij}^{(1)}(X_{ijt}^{(1)}(s))-\overline{Y}_{ij}(s)\}\right]^2}{\sum_{t=1}^{n_{ij}}\{Y_{ijt}(s)-\overline{Y}_{ij}(s)\}^2\sum_{t=1}^{n_{ij}}\{\widehat{m}_{ij}^{(1)}(X_{ijt}^{(1)}(s))-\overline{Y}_{ij}(s)\}^2}, \quad (3.6)$$

where $\overline{Y}_{ij}(s)=\sum_{t=1}^{n_{ij}}Y_{ijt}(s)/n_{ij}$ is the average of $Y_{ijt}(s)$ in season j of year i.

We use the MSE to evaluate the importance of explanatory variables. By taking the average of $\text{MSE}_{ij}^{(1)}(X_{ijt}^{(1)}(s),s)$ for all the monitoring sites in the city, we obtain the MSE of the city for using the univariate regressor $X_{ijt}^{(1)}$ at year i and season j. The most important explanatory variable for the city, denoted as $X_{ijt}^{(1)*}$, at year i and season j is the one that produced the smallest MSE of the city.

To select the second most important regressor, we consider the bivariate regression model

$$Y_{ijt}(s)=m_{ij}^{(2)}(\mathbf{X}_{ijt}^{(2)}(s))+\varepsilon_{ijt}^{(2)}(s), \quad t=1,2,\ldots,n_{ij}, \quad (3.7)$$

where $\mathbf{X}_{ijt}^{(2)}(s)=(X_{ijt}^{(1)*}(s),X_{ijt}^{(2)}(s))'$ for a $X_{ijt}^{(2)}(s)$ other than the already selected $X_{ijt}^{(1)*}(s)$, $m_{ij}^{(2)}(\mathbf{X}_{ijt}^{(2)}(s))=E(Y_{ijt}(s)|\mathbf{X}_{ijt}^{(2)}(s))$ is the conditional mean given $\mathbf{X}_{ijt}^{(2)}(s)$ and $\varepsilon_{ijt}^{(2)}(s)$ is the corresponding residual. Similar to selecting the first variable, the MSE with $X_{ijt}^{(2)}(s)$ as covariates is

$$\text{MSE}_{ij}^{(2)}(\mathbf{X}_{ijt}^{(2)}(s),s)=\frac{1}{n_{ij}}\sum_{t=1}^{n_{ij}}\{Y_{ijt}(s)-\widehat{m}_{ij}^{(2)}(\mathbf{X}_{ijt}^{(2)}(s))\}^2, \quad (3.8)$$

where $\widehat{m}_{ij}^{(2)}(\cdot)$ is also attained via (3.3). We select the second important variable by minimizing the average MSE in (3.8) over the sites in the city. The coefficient of determination $R^2(2, s)$ by using $\mathbf{X}_{ijt}^{(2)}(s)$ is

$$R_{ij}^2(2,s)=\frac{\left[\sum_{t=1}^{n_{ij}}\{Y_{ijt}(s)-\overline{Y}_{ij}(s)\}\{\widehat{m}_{ij}^{(2)}(\mathbf{X}_{ijt}^{(2)}(s))-\overline{Y}_{ij}(s)\}\right]^2}{\sum_{t=1}^{n_{ij}}\{Y_{ijt}(s)-\overline{Y}_{ij}(s)\}^2\sum_{t=1}^{n_{ij}}\{\widehat{m}_{ij}^{(2)}(\mathbf{X}_{ijt}^{(2)}(s))-\overline{Y}_{ij}(s)\}^2}. \quad (3.9)$$

It is shown in Theorem S1 of SM that the variance of $\varepsilon_{ijt}^{(2)}(s)$ in Model (3.7) is less than or equal to the variance of $\varepsilon_{ijt}^{(1)}(s)$ in Model (3.4) with the $X_{ijt}^{(1)*}(s)$ as the regressor. This informs the benefits of adding more regressors to the regression model as it can reduce the variations of the residual term and hence makes the regression model having better fitting performance, although it is another matter for prediction as having more variables may not lead to better forecast. An implication of the result is that asymptotically $\text{MSE}_{ij}^{(2)}(\mathbf{X}_{ijt}^{(2)}(s),s)\le\text{MSE}_{ij}^{(1)}(X_{ijt}^{(1)*}(s),s)$ and the R^2 monotonically increases with the nested regressors.

We proceed the nested forward procedure to select the third, the fourth important covariates and beyond until all the candidate covariates are considered. The order of the variables being selected contains useful information on the relative importance of the meteorological factors on the pollutant concentration and has taken into consideration of the inter-dependence of the meteorological variables.

To find out how much information in the BLH that can be explained by the surface meteorological variables, we carry on the same nested forward selection procedure by substituting the pollution concentration with BLH as the response variable and the surface variables as the regressors. The order of the variables being selected are used as the order of relative importance for BLH.

Furthermore, to avoid over-fitting of the nonparametric regression, we used a 10-fold cross validation (CV) procedure to gain information on the out-of-sample performance with the selected variables. The 10-fold CV randomly divides the data into 10 segments and alternatively uses any 9 segments at a time to fit the regression model and the remaining one to gain validation performance of the model. As the validation is made on the data set which is not used in the model fitting, it is more objective and is called the out-of-sample validation. Specifically, we randomly divided the data of a season at a station into 10 segments. Then, for each data segment (as the validation data set), we fit the nonparametric regression model on the remaining 9 segments of data (the training set) and evaluate its performance on the validation data set by calculating the out-of-sample $R_{CV,k}^2$

$$R_{CV,k}^2=\frac{\left\{\sum_{t=1}^{n_k}(Y_{k,t}-\overline{Y}_k)(\widehat{m}_{-k}(\mathbf{X}_{k,t})-\overline{Y}_k)\right\}^2}{\sum_{t=1}^{n_k}(Y_{k,t}-\overline{Y}_k)^2\sum_{t=1}^{n_k}\{\widehat{m}_{-k}(\mathbf{X}_{k,t})-\overline{Y}_k\}^2}, \quad k=1,2,\ldots,10, \quad (3.10)$$

where $\widehat{m}_{-k}(\cdot)$ is the fitted regression function without the kth segment, $\mathbf{X}_{k,t}$ and $Y_{k,t}$ denote the tth observation in the kth segment. The cross-validated coefficient of determination R_{CV}^2 is

$$R_{CV}^2=\frac{1}{10}\sum_{k=1}^{10}R_{CV,k}^2, \quad (3.11)$$

3.3. Meteorological adjustment

To gain information on the underlying emission, the meteorological effects on the pollution concentration has to be removed. There were methods to adjust for the meteorological variation, including the trend analysis approach proposed in Thompson et al. (2001) based on the linear regression and the three-year moving average method proposed by the US Environmental Protection Agency (EPA). Chen et al. (2018) showed that the latter method can not remove the meteorological confounding. Liang et al. (2015) proposed a general framework to adjust for the meteorological confounding, see Zhang et al. (2020) for comprehensive theoretical analysis.

For air quality evaluation based on the meteorological variables $\mathbf{X}_{ijt}(s)$, we consider Model (3.2) to quantify the meteorological relationship with the pollutant's concentration $Y_{ijt}(s)$. To remove the meteorological confounding, Liang et al. (2015) proposed a version of the average pollution concentration

$$\mu_{ij}(s)=E(Y_{ijt}(s))=E\{m_{ij}(\mathbf{X}_{ijt}(s))\}=\int m_{ij}(\mathbf{x},s)f_j(\mathbf{x},s)d\mathbf{x}, \quad (3.12)$$

where $m_{ij}(\mathbf{X}_{ijt}(s))$ is the regression function that can be estimated by (3.3), and $f_j(\mathbf{x}, s)$ is a baseline probability density function for the meteorological variable $\mathbf{X}(s)$ in season j which can be constructed as

$$f_j(\mathbf{x},s)=A_j^{-1}\sum_{a=1}^{A_j}f_{aj}(\mathbf{x},s),$$

where $f_{aj}(\mathbf{x}, s)$ is the density function of meteorological variable $\mathbf{X}_{ajt}(s)$ in year a of season j, and A_j is the number of years that the meteorological data are available. In our study, A_j was 10 as we used 10 years' meteorological data from March 2011 to February 2021 to build the baseline density $f_j(\mathbf{x}, s)$.

An estimator for $\mu_{ij}(s)$ which is free of the meteorological variation and can reflect the emission is

$$\widehat{\mu}_{ij}(s)=\left(\sum_{a=1}^{A_j}n_{aj}\right)^{-1}\sum_{a=1}^{A_j}\sum_{t=1}^{n_{aj}}\widehat{m}_{ij}(\mathbf{X}_{ajt}(s)), \quad (3.13)$$

where n_{aj} is the number of meteorological sample size in season j of year a. The nonparametric regression estimator $\widehat{m}_{ij}(\cdot)$ is (3.3). Unlike the raw averages, the temporally adjusted averages $\widehat{\mu}_{ij}(s)$ are comparable for the specific season over different years as shown in Zhang et al. (2017a) and Chen et al. (2018).

The air quality measures $\widehat{\mu}_{ij}(\mathcal{A})$ over a region $\mathcal{A}$ occupied by a city is an average of the adjusted $\widehat{\mu}_{ij}(s)$ over the monitoring sites in the region:

$$\widehat{\mu}_{ij}(\mathcal{A})=|\mathcal{A}|^{-1}\sum_{s\in\mathcal{A}}\widehat{\mu}_{ij}(s), \quad (3.14)$$

where $|\mathcal{A}|$ denotes the number of monitoring sites in the region $\mathcal{A}$.

4. Results

We present the results of the analyzes in this section.

4.1. Variable importance for pollutants

We used the proposed methods in Section 3.2 to obtain the orders of meteorological variables for the six pollutants in each city in each season. Specifically, for the pollutant at each season and year, we attained the order of variable importance for each station in a city and then average the ranks over all stations in the city and over all years for the season.

Figs. 1–3 display the seasonal average orders of the meteorological variables for $PM_{2.5}$, NO_2 and 8-h O_3 in each city respectively, while those for the other three pollutants are shown in Figs. S3–S5. The numerical average ranks for each pollutant in the six cities were reported in Tables S1–S6.

For $PM_{2.5}$, as shown in Fig. 1, there was a consistent ordering in the variable ranks among the seasons and the cities, namely

$$\text{DEWP, PRES, TEMP, BLH, IWS, WD, IPREC.} \quad (4.1)$$

We also noticed that, the above order of variable importance for $PM_{2.5}$ was also obeyed for PM_{10}, SO_2 and CO as shown in Figs. S3–S5 in the SM. However, the orders for NO_2 and 8-h O_3, as shown in Figs. 2 and 3 were different. For NO_2, the variable order was

$$\text{BLH, DEWP, PRES, TEMP, IWS, WD, IPREC,} \quad (4.2)$$

which saw BLH jumped to the leading position while the relative order of the other variables remained unchanged from (4.1). For ozone, the order was

$$\text{TEMP, DEWP, PRES, BLH, IWS, WD, IPREC,} \quad (4.3)$$

which had the temperature (TEMP) became the leading variable. The strong consistency in the three orderings (4.1)–(4.3) among the six cities indicated consistency of the meteorological effects on the three categories of pollutants in the study region.

The dew point temperature (DEWP) was the highest ranked variable for the PM-CO-SO_2 group, and the second ranked for NO_2 and O_3. The reason for DEWP's leading roles was its ability in reflecting both temperature and humidity. It is noted that we did not consider the relative humidity as a variable since it is determined by the dew point and the air temperature. Air pressure (PRES) was the second ranked for the PM-CO-SO_2 group, and the third ranked for NO_2 and O_3. Air temperature (TEMP) was the third ranked for the PM-CO-SO_2 group, the fourth for NO_2 but risen to the top place for ozone. TEMP's leading role for ozone was due to its being a proxy for the solar radiation, a major element in the photo-chemistry process of ozone generation. The wind direction (WD) and the cumulative precipitation (IPREC) ranked the lowest for all six pollutants. Cumulative wind speed (IWS) was the third last ranked for all six species.

It should be noted that the ranking did not suggest that the lower ranking variables, say IWS, WD and IPREC, were not influential to the pollutant's concentration, but rather their relative importance after considering the other variables. In particular, the impacts of the lowly ranked variables had been represented by the higher ranking variables. Indeed, a change in the wind direction from the southerly to northerly would bring dryer, cooler and cleaner air mass from the north, which led to a lower DEWP.

It was interesting to see BLH became the leading variable for NO_2. BLH was the clear top variable for all six cities in the three non-winter seasons. The leading role of BLH to NO_2 can be understood by noting the pathway $NO_2 + O_2 \rightarrow O_3 + NO$ that converts NO_2 to ozone under solar radiation, and the fact that the reverse pathway is weaken by interference of other chemical elements in polluted air which consumed

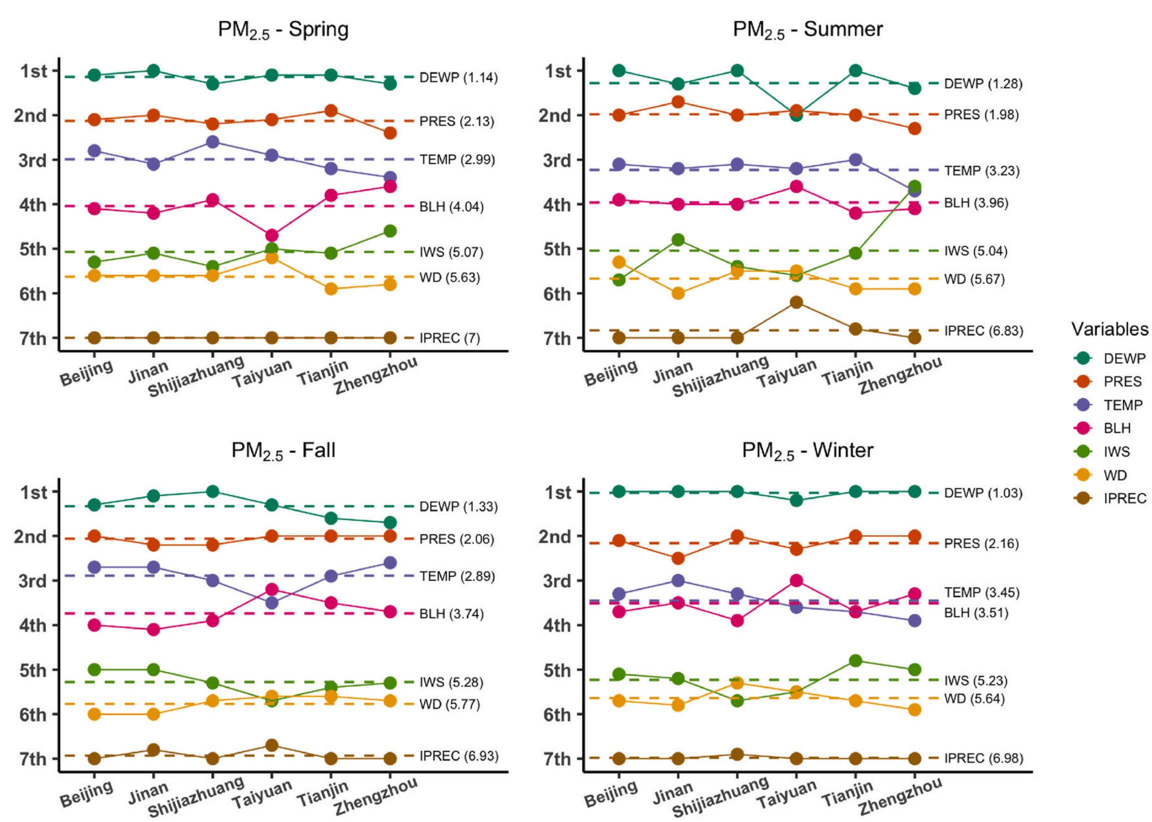

Fig. 1. Seasonal average ranks of the meteorological variables for $PM_{2.5}$ regression from 2013 to 2020. The dots with different colors represent ranks with the horizontal dashed lines marking the average ranks, reported inside the parentheses on the right edges, of the variables among the six cities based on the seasonal regression models. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

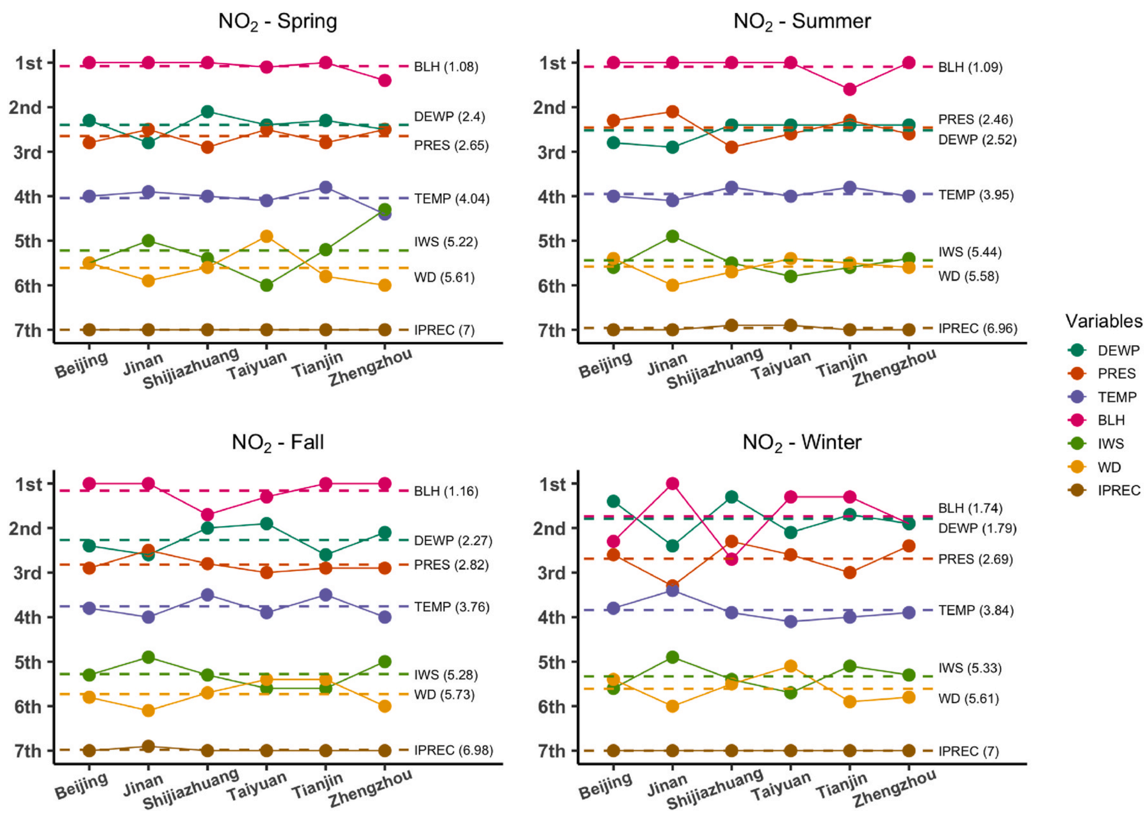

Fig. 2. Seasonal average ranks of the meteorological variables for NO_2 regression from 2013 to 2020. The dots with different colors represent ranks with the horizontal dashed lines marking the average ranks, reported inside the parentheses on the right edges, of the variables among the six cities based on the seasonal regression models. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

much NO leading to a weakening of the reverse process. As BLH is highly correlated with the solar radiation, BLH became a proxy for the radiation and hence the leading meteorological variable for NO_2. Indeed, as shown in Figs. S6 and S7, NO_2 was the pollutant which BLH had both the highest correlation and the partial correlation after conditioning on the other meteorological variables. The partial correlation has filtered out the effects of the other variables on the pollutant (Anderson, 2003). While the partial correlation of BLH with the other five pollutants were much reduced after considering the other variables, its partial correlation with the NO_2 remained at a relative high level as shown in Figs. S12 and S13, which provides a statistical understanding for the BLH's leading role in NO_2.

The temperature's leading role to Ozone can be appreciated by its close relationship with the solar radiation, and the latter is directly involved in the ozone generation. As shown in Figs. S10 and S11, the temperature had the highest partial correlation with the O_3 which explained its leading role among the meteorological variables. It is interesting to see a large change in the correlation between BLH and ozone after we conditioned on the other variables in Figs. S6 and S7, which provided an explanation for BLH's dropping to the fifth or sixth place for ozone in the non-winter seasons in Fig. 3.

Fig. 4 reports the cumulative R^2 curves in Beijing as the meteorological variables were added according to the forward selection procedure. Similar plots for the other five cities are available in Figs. S14–S16. These figures showed a rapid increase in the R^2 as the first three or four ranked variables were added in the models, while the growth in the cumulative R^2 was much slower after the fourth variables. These suggest that the top three or four variables in (4.1)-(4.3) contain high percentages of the meteorological information.

4.2. *Role of boundary layer height*

The analysis in the last subsection showed that, except for NO_2, the BLH ranked around the fourth or fifth place in the overall variable ranking for the pollutants. In particular, it ranked behind DEWP, PRES, TEMP for the five non-NO_2 pollutants in both the ranking and the contributions to the R^2 in Fig. 4. These did not necessarily imply that BLH was not influential to these species, but rather its role was relatively less important when the other meteorological variables were considered.

For better viewing on the correlation between BLH and the pollutants, Figs. S6 and S7 display the heat-map of the correlations between BLH and the pollutants and the corresponding partial correlations by conditioning on the surface meteorological variables. The partial correlations filter out the role of the surface meteorological variables, and tended to reduce the magnitude of the usual correlations. Both forms of the correlations were largely negative for the five pollutants other than the 8-h O_3 which were positive. The negative correlations between BLH and NO_2 were the highest among the six species. This was consistent with the finding in (4.2) that the BLH was the leading variable for NO_2. Now we try to understand the importance of BLH from the aspect that how much data information contained in BLH can be explained by the other meteorological variables.

To address this question, we conduct analysis based on the nonparametric regression model (3.2) but with the response variable Y being BLH and the regressors $\widetilde{\mathbf{X}}$ being the other meteorological variables:

$$\mathrm{BLH}_{ijt} = m_{ij}(\widetilde{\mathbf{X}}_{ijt}) + \varepsilon_{ijt}, \quad t = 1, 2, \ldots, n_{ij}. \tag{4.4}$$

Fig. S23 in SM reports the variable ranks from the forward selection algorithm, which shows that the temperature (TEMP), dew point (DEWP), and pressure (PRES) were the top three factors for BLH in three

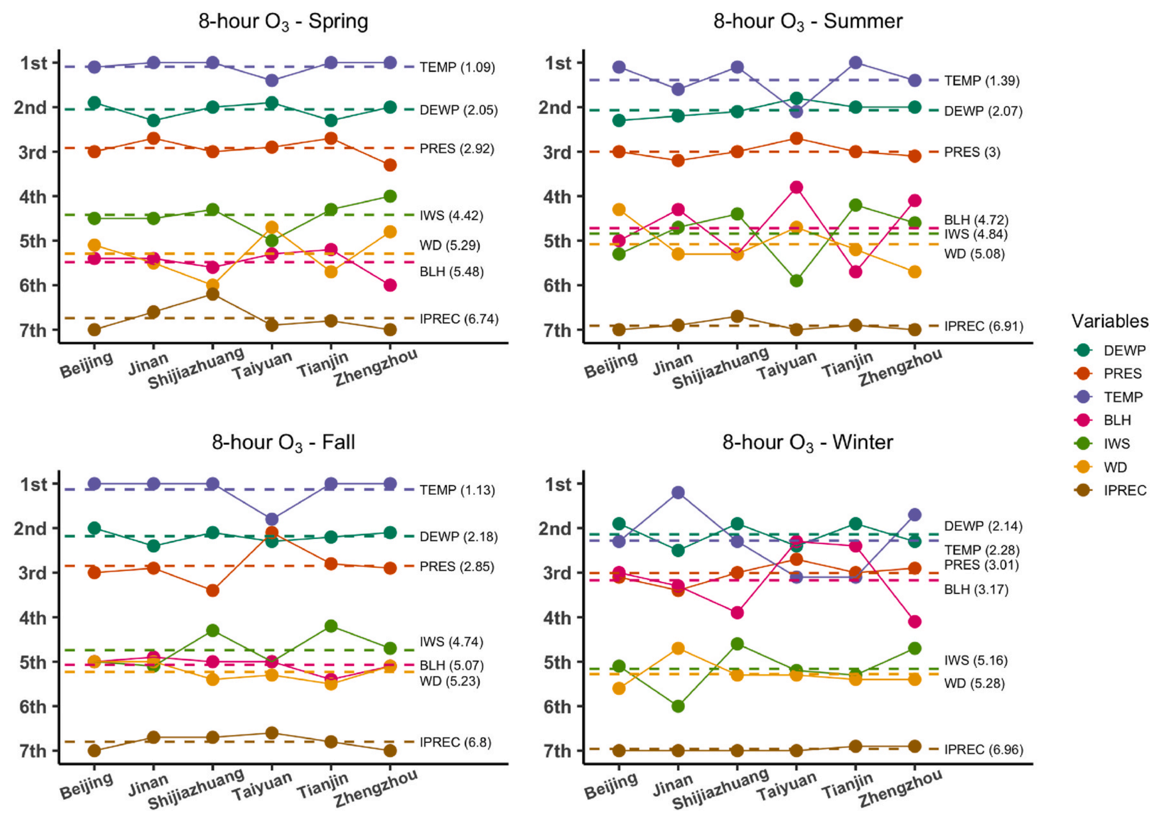

Fig. 3. Seasonal average ranks of the meteorological variables for 8-h O_3 regression from 2013 to 2020. The dots with different colors represent ranks with the horizontal dashed lines marking the average ranks, reported inside the parentheses on the right edges, of the variables among the six cities based on the seasonal regression models. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

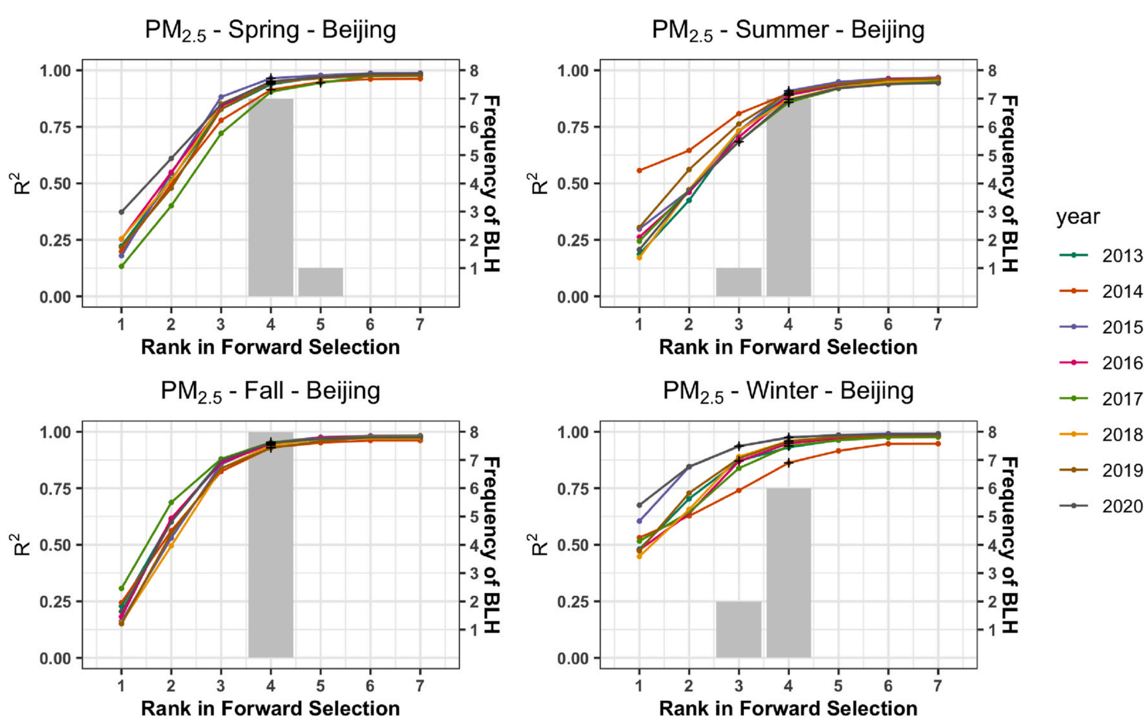

Fig. 4. Seasonal step-wise cumulative in-sample R^2 (segmented lines) with variables added to the nonparametric regression of $PM_{2.5}$ in Beijing according to their ranks selected in the forward procedure, and the histograms (grey color bars) for the distributions of the ranks occupied by the BLH (scaled on the right). The ranks of BLH are also marked by black crosses. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

seasons other than fall, where in fall IWS replaced pressure as the third one. We consider two different subsets of $\widetilde{\mathbf{X}}$. One was the five variables without IPREC as it was the last ranked in the rankings reported in (4.1)-(4.3). And the other subset contains only the three top variables, DEWP, TEMP and PRES, which were the most important three ones among most cities and seasons and they ranked in front of BLH in most cases.

Table 1 reports both the seasonal in-sample R^2 and the out-of-sample R^2_{CV} when regressing BLH with respect to the three and the five variable sets, respectively, for the six cities. The results showed that the average in-sample R^2 with the three top variables can attain at least 77% of the total variation in BLH among the four seasons and the six cities, which was risen to at least 92% with the five variables' set of the meteorological factors as the regressors. As expected, the out-of-sample R^2_{CV}s were lower than the corresponding R^2. However, the minimum R^2_{CV} was still 57% with the three top variables and risen to 68% for the five variables' set.

The results conveyed in Table 1 suggest that much of the information in BLH can be modeled by the other meteorological variables, and by the top three variables DEWP, TEMP and PRES in particular. The results provide explanation on why BLH was not ranked high for most of the pollutant species as conveyed in Figs. 1 and 3, as well as on the marginal contribution to the R^2 by the BLH (Figs. 4 and 5).

To gain further insights on the role of the BLH, we studied its influence on the air quality measures by calculating the meteorologically adjusted average concentration (3.13) with BLH versus those without BLH as a covariate. Specifically, for a city that encompasses region $\mathcal{A}$, let $\widehat{\mu}_{ij}(\mathcal{A})$ and $\widehat{\mu}_{ij}^{-BLH}(\mathcal{A})$ be the adjusted average concentrations with and without BLH, respectively, for year i and season j via (3.14). The BLH effect is measured by relative change caused by without BLH

$$\widehat{r}_{ij}(\mathcal{A}) = |\widehat{\mu}_{ij}(\mathcal{A}) - \widehat{\mu}_{ij}^{-BLH}(\mathcal{A})| \Big/ \widehat{\mu}_{ij}(\mathcal{A}). \quad (4.5)$$

Table 2 summarizes the average and the maximum BLH effects (in percentage) over the 32 seasons from 2013 to 2020 for the six pollutants in the six cities. Among the 36 pollutant-city combinations for the estimated average BLH effects, 16 of them were less than 1% and 15 between 1% and 1.5%, and only 5 of them above 1.5%, indicating rather mild BLH effects on the pollutants among the six cities. However, the BLH effects appeared not evenly distributed among the species and cities. Beijing and Tianjin had smaller average and maximum effects while those for Taiyuan and Jinan were larger. Taiyuan had the largest average and maximum effects among the six cities, which may be due to its higher elevation (average 800 m) as the other five cities are all situated in the North China Plain.

To gain information on the significance of the BLH effects reported in Table 2, we conducted statistical testing for the hypothesis $H_0: \mu_{ij}(\mathcal{A}) = \mu_{ij}^{-BLH}(\mathcal{A})$ versus $H_1: \mu_{ij}(\mathcal{A}) \neq \mu_{ij}^{-BLH}(\mathcal{A})$ at 5% significance level in each season for each city and pollutant. The p-value of each testing was obtained based on 300 bootstrap resampling of the test statistics (Liang et al., 2015; Zhang et al., 2017a). Table 3 summarizes the overall testing results by providing the frequencies of rejecting the above H_0 among the 32 seasons from 2013 to 2020 for each city and each pollutant. It shows that only Taiyuan, Jinan and Zhengzhou had significant BLH effects. Taiyuan had the most number of significant differences for three species (SO_2, NO_2 and O_3), but still the numbers of significant seasons ranged from 2 to 4 seasons. The results of Table 3 were consistent to those in Table 2, showing the BLH's effect was largely small as far as the average air quality measures were concerned.

5. Conclusion

Existing studies on the meteorological effects on air pollution tended to consider one variable at a time as showed in the cited works of the paper. This study is designed to evaluate the collective meteorological effects that takes into account the mutual dependence among the meteorological variables. And it finds that there were much agreement in the meteorological effects on the air pollution in North China, as reflected by the much agreed orders of variable importance on the six major air pollutants. These suggest stable meteorological processes with respect to the air pollution in North China, which may be used for meteorological modeling and prediction of air quality.

The study also reveals that, due to the inter-dependence of the meteorological variables, the top three variables can provide quite satisfactory modeling for the meteorological effects of a pollutant, as reflected by the rapid increase in the R^2 among the first three variables. The inter-dependence of the meteorological variables means that the less ranked variable's information and their effects on the pollutants can be well represented by the top ranked meteorological variables, as demonstrated in the case of BLH. The inter-dependence also means that adequate surface air quality measures can be well approximated by utilizing the surface meteorological variables as much of the data information in the mixing layer may be well represented by the surface variables.

Our study has considered only one non-surface variable. Other variables at the vertical pressure layers within the mixing layer can be considered and their importance can be assessed using the method demonstrated in the study. Given the high correlation between the variables on the surface and the vertical layer, it would not be surprising to see the meteorological information in the vertical layer can be well substituted by the surface variables for most of the pollutants.

Table 1
Seasonal in-sample R^2 and the out-sample R^2_{CV} for BLH with the top three meteorological variables (DEWP, PRES, TEMP) or five variables (DEWP, PRES, TEMP, IWS, WD) for the six cities (averaged from 2013 to 2020). The numbers inside the parentheses are the standard errors of the average R^2 above.

City	Number of	Spring		Summer		Fall		Winter	
	Variables	R^2	R^2_{CV}	R^2	R^2_{CV}	R^2	R^2_{CV}	R^2	R^2_{CV}
Beijing	3	0.85	0.64	0.77	0.57	0.88	0.71	0.78	0.60
	5	0.96	0.76	0.91	0.67	0.95	0.77	0.93	0.69
Jinan	3	0.82	0.56	0.78	0.57	0.86	0.64	0.80	0.56
	5	0.95	0.68	0.94	0.69	0.96	0.76	0.95	0.71
Shijiazhuang	3	0.83	0.59	0.76	0.54	0.85	0.63	0.78	0.54
	5	0.96	0.73	0.92	0.66	0.95	0.73	0.93	0.68
Taiyuan	3	0.78	0.58	0.78	0.63	0.82	0.66	0.76	0.58
	5	0.94	0.71	0.90	0.70	0.93	0.72	0.91	0.66
Tianjin	3	0.83	0.58	0.76	0.55	0.85	0.67	0.79	0.57
	5	0.96	0.73	0.92	0.67	0.96	0.76	0.94	0.71
Zhengzhou	3	0.82	0.58	0.77	0.57	0.85	0.63	0.81	0.58
	5	0.96	0.74	0.94	0.71	0.96	0.74	0.95	0.70
Average	3	0.82	0.59	0.77	0.57	0.85	0.66	0.79	0.57
		(0.009)	(0.011)	(0.003)	(0.012)	(0.008)	(0.012)	(0.007)	(0.008)
	5	0.96	0.72	0.92	0.68	0.95	0.75	0.93	0.69
		(0.003)	(0.010)	(0.006)	(0.009)	(0.005)	(0.008)	(0.006)	(0.008)

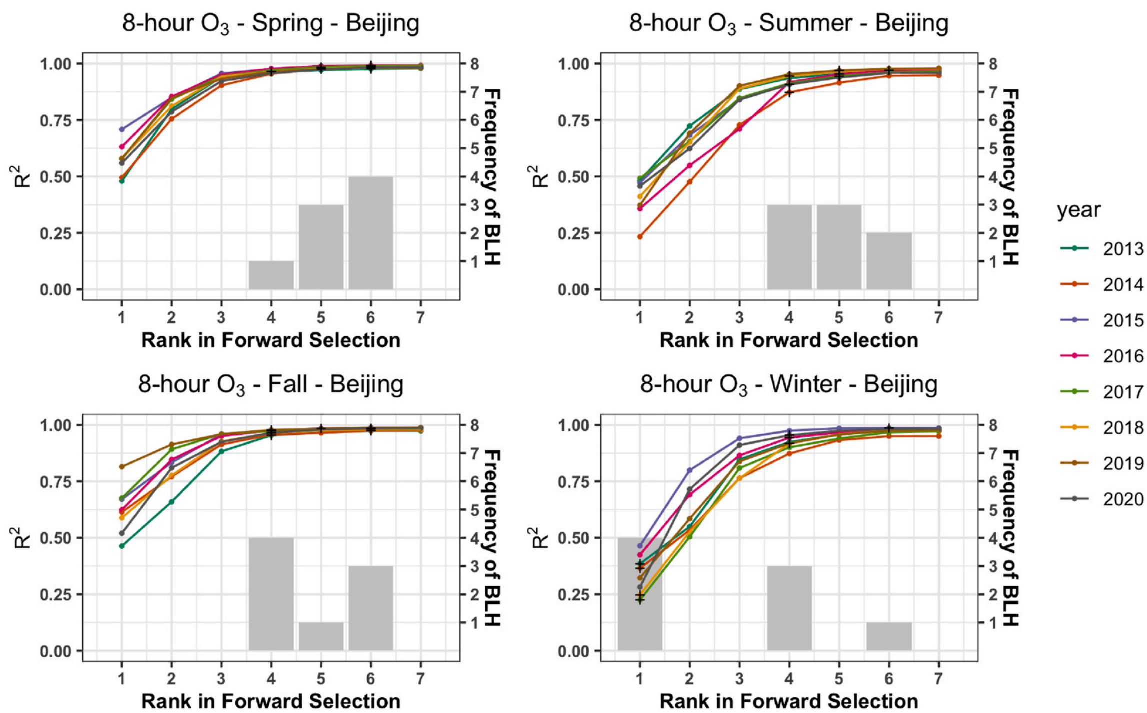

Fig. 5. Seasonal step-wise cumulative in-sample R^2 (segmented lines) with variables added to the nonparametric regression of 8-h O_3 in Beijing according to their ranks selected in the forward procedure, and the histograms (grey color bars) for the distributions of the ranks occupied by the BLH (scaled on the right). The ranks of BLH are also marked by black crosses. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
The average relative BLH effects (standard deviations) and the maximum BLH effects (the second row) $\hat{r}_{ij}(\mathcal{A})$ in (4.5) in percentage term for the six pollutants and the six cities over the 32 seasons from 2013 to 2020.

Pollutants	Beijing	Jinan	Shijiazhuang	Taiyuan	Tianjin	Zhengzhou	Average
$PM_{2.5}$	0.83(0.12)	1.29(0.15)	0.82(0.12)	1.57(0.24)	1.02(0.13)	0.73(0.09)	1.04(0.12)
	2.54	4.02	2.74	5.52	2.89	1.98	3.28
PM_{10}	0.65(0.12)	1.07(0.14)	0.67(0.10)	1.34(0.22)	0.93(0.11)	0.70(0.08)	0.89(0.10)
	2.82	4.06	2.27	4.31	2.15	1.80	2.90
SO_2	1.20(0.13)	1.37(0.19)	1.19(0.21)	2.32(0.36)	1.15(0.13)	0.96(0.14)	1.37(0.18)
	2.74	4.38	6.06	7.90	3.25	3.22	4.59
NO_2	0.94(0.13)	1.58(0.19)	1.29(0.15)	1.39(0.18)	1.13(0.13)	1.13(0.28)	1.24(0.08)
	2.96	4.29	4.06	3.77	2.63	8.55	4.37
CO	0.78(0.14)	1.24(0.18)	1.10(0.18)	0.99(0.17)	0.80(0.10)	0.72(0.15)	0.94(0.08)
	2.86	4.04	3.71	4.16	2.39	4.37	3.59
8-h O_3	0.56(0.08)	2.72(0.38)	1.29(0.21)	3.79(0.47)	0.79(0.11)	0.73(0.12)	1.65(0.49)
	1.76	7.19	5.13	10.62	2.27	2.34	4.88
Average	0.83(0.09)	1.55(0.24)	1.06(0.11)	1.90(0.42)	0.97(0.06)	0.83(0.07)	1.19(0.10)
	2.61	4.66	3.99	6.05	2.60	3.71	3.94

Table 3
Frequencies (percentages) of significant differences (at 5% level) in the seasonal adjusted average pollution concentrations that included BLH versus those without BLH for the six pollutants and the six cities over the 32 seasons from 2013 to 2020.

Pollutants	Beijing	Jinan	Shijiazhuang	Taiyuan	Tianjin	Zhengzhou
$PM_{2.5}$	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
PM_{10}	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
SO_2	0 (0)	0 (0)	0 (0)	2 (6.25%)	0 (0)	0 (0)
NO_2	0 (0)	0 (0)	0 (0)	3 (9.38%)	0 (0)	2 (6.25%)
CO	0 (0)	2 (6.25%)	0 (0)	0 (0)	0 (0)	1 (3.13%)
8-h O_3	0 (0)	2 (6.25%)	0 (0)	4 (12.5%)	0 (0)	0 (0)

CRediT authorship contribution statement

Yaxuan Huang: Software, Formal analysis, Data curation, Writing – original draft. **Bin Guo:** Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft. **Haoxuan Sun:** Software, Formal analysis, Data curation, Writing – original draft. **Huijie Liu:** Software, Formal analysis, Data curation. **Song Xi Chen:** Conceptualization, Methodology, Writing – original draft, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The research was partially supported by National Natural Science Foundation of China Grants 92046021, 12071013, 71973005 and 11971390, and LMEQF at Peking University. Bin Guo acknowledges support from Fundamental Research Funds for the Central Universities JBK1806002.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2021.118737.

References

Anderson, T.W., 2003. An Introduction to Multivariate Statistical Analysis, third ed. John Wiley and Sons, New York.
Chen, L., Guo, B., Huang, J., He, J., Wang, H., Zhang, S., Chen, S.X., 2018. Assessing air-quality in Beijing-Tianjin-Hebei region: the method and mixed tales of $PM_{2.5}$ and O_3. Atmos. Environ. 193, 290–301.
Chen, R., Zhao, Z., Kan, H., 2013. Heavy smog and hospital visits in Beijing, China. Am. J. Respir. Crit. Care Med. 188, 1170–1171.
Choubin, B., Abdolshahnejad, M., Moradi, E., Querol, X., Mosavi, A., Shamshirband, S., Ghamisi, P., 2020. Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. Sci. Total Environ. 701, 134474.
Doksum, K., Samarov, A., 1995. Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. Ann. Stat. 23, 1443–1473.
Donaldson, K., Li, X., MacNee, W., 1998. Ultrafine (nanometre) particle mediated lung injury. J. Aerosol Sci. 29, 553–560.
Durre, I., Yin, X., 2008. Enhanced radiosonde data for studies of vertical structure. Bull. Am. Meteorol. Soc. 89, 1257–1262.
Feng, Z., De Marco, A., Anav, A., Gualtieri, M., Sicard, P., Tian, H., Fornasier, F., Tao, F., Guo, A., Paoletti, E., 2019. Economic losses due to ozone impacts on human health, forest productivity and crop yield across China. Environ. Int. 131, 104966.
Gui, K., Che, H., Wang, Y., Wang, H., Zhang, L., Zhao, H., Zheng, Y., Sun, T., Zhang, X., 2019. Satellite-derived $PM_{2.5}$ concentration trends over Eastern China from 1998 to 2016: relationships to emissions and meteorological parameters. Environ. Pollut. 247, 1125–1133.
Guo, J., Li, Y., Cohen, J.B., Li, J., Chen, D., Xu, H., Liu, L., Yin, J., Hu, K., Zhai, P., 2019. Shift in the temporal trend of boundary layer height in China using long-term (1979–2016) radiosonde data. Geophys. Res. Lett. 46, 6080–6089.
Guo, J., Miao, Y., Zhang, Y., Liu, H., Li, Z., Zhang, W., He, J., Lou, M., Yan, Y., Bian, L., Zhai, P., 2016. The climatology of planetary boundary layer height in China derived from radiosonde and reanalysis data. Atmos. Chem. Phys. 16, 13309–13319.
Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed. Springer.
Liang, X., Li, S., Zhang, S., Huang, H., Chen, S.X., 2016. $PM_{2.5}$ data reliability, consistency, and air quality assessment in five Chinese cities. J. Geophys. Res.: Atmosphere 121, 10220–10236.
Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., Huang, H., Chen, S.X., 2015. Assessing Beijing's $PM_{2.5}$ pollution: severity, weather impact, APEC and winter heating. Proc. Royal Soc. A 471, 20150257.
Liu, X.G., Li, J., Qu, Y., Han, T., Hou, L., Gu, J., Chen, C., Yang, Y., Liu, X., Yang, T., Zhang, Y., Tian, H., Hu, M., 2013. Formation and evolution mechanism of regional haze: a case study in the megacity Beijing, China. Atmos. Chem. Phys. 13, 4501–4514.
Miao, Y., Che, H., Zhang, X., Liu, S., 2021. Relationship between summertime concurring $PM_{2.5}$ and O_3 pollution and boundary layer height differs between Beijing and Shanghai, China. Environ. Pollut. 268, 115775.
Miao, Y., Guo, J., Liu, S., Liu, H., Li, Z., Zhang, W., Zhai, P., 2017. Classification of summertime synoptic patterns in Beijing and their associations with boundary layer structure affecting aerosol pollution. Atmos. Chem. Phys. 17, 3097–3110.
Miao, Y., Liu, S., 2019. Linkages between aerosol pollution and planetary boundary layer structure in China. Sci. Total Environ. 650, 288–296.
Pope III, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurston, G.D., 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. J. Am. Med. Assoc. 287, 1132–1141.
Quan, J., Tie, X., Zhang, Q., Liu, Q., Li, X., Gao, Y., Zhao, D., 2014. Characteristics of heavy aerosol pollution during the 2012–2013 winter in Beijing, China. Atmos. Environ. 88, 83–89.
Schwartz, J., 2000. The distributed lag between air pollution and daily deaths. Epidemiology 11, 320–326.
Seidel, D.J., Zhang, Y., Beljaars, A., Golaz, J.C., Jacobson, A.R., Medeiros, B., 2012. Climatology of the planetary boundary layer over the continental United States and Europe. J. Geophys. Res.: Atmosphere 117, 17106.
Shen, L., Jacob, D.J., Mickley, L.J., Wang, Y., Zhang, Q., 2018. Insignificant effect of climate change on winter haze pollution in Beijing. Atmos. Chem. Phys. 18, 17489–17496.
Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., de'Donato, F., Gariazzo, C., Lyapustin, A., Michelozzi, P., Renzi, M., Scortichini, M., Shtein, A., Viegi, G., Kloog, I., Schwartz, J., 2019. Estimation of daily PM_{10} and $PM_{2.5}$ concentrations in Italy, 2013-2015, using a spatiotemporal land-use random-forest model. Environ. Int. 124, 170–179.
Tai, A.P., Mickley, L.J., Jacob, D.J., 2010. Correlations between fine particulate matter ($PM_{2.5}$) and meteorological variables in the United States: implications for the sensitivity of $PM_{2.5}$ to climate change. Atmos. Environ. 44, 3976–3984.
Tang, G., Zhang, J., Zhu, X., Song, T., Münkel, C., Hu, B., Schäfer, K., Liu, Z., Zhang, J., Wang, L., Xin, J., Suppan, P., Wang, Y., 2016. Mixing layer height and its implications for air pollution over Beijing, China. Atmos. Chem. Phys. 16, 2459–2475.
Thompson, M.L., Reynolds, J., Cox, L.H., Guttorp, P., Sampson, P.D., 2001. A review of statistical methods for the meteorological adjustment of tropospheric ozone. Atmos. Environ. 35, 617–630.
Vogelezang, D.H.P., Holtslag, A.A.M., 1996. Evaluation and model impacts of alternative boundary-layer height formulations. Boundary-Layer Meteorol. 81, 245–269.
Vu, T.V., Shi, Z., Cheng, J., Zhang, Q., He, K., Wang, S., Harrison, R.M., 2019. Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique. Atmos. Chem. Phys. 19, 11303–11314.
Xiang, Y., Zhang, T., Liu, J., Lv, L., Dong, Y., Chen, Z., 2019. Atmosphere boundary layer height and its effect on air pollutants in Beijing during winter heavy pollution. Atmos. Res. 215, 305–316.
Xie, Y., Dai, H., Dong, H., Hanaoka, T., Masui, T., 2016. Economic impacts from $PM_{2.5}$ pollution-related health effects in China: a provincial-level analysis. Environ. Sci. Technol. 50, 4836–4843.
Zhang, Q., Ma, X., Tie, X., Huang, M., Zhao, C., 2009. Vertical distributions of aerosols under different weather conditions: analysis of in-situ aircraft measurements in Beijing, China. Atmos. Environ. 43, 5526–5535.
Zhang, S., Chen, S.X., Guo, B., Wang, H., Lin, W., 2020. Regional air-quality assessment that adjusts for meteorological confounding. Sci. Sinica Math. 50, 527–558. The English translation is available from: https://songxichen.com/Uploads/Files/Publication/SCM-2019-0368R2_Proof_hi.pdf.
Zhang, S., Guo, B., Dong, A., He, J., Xu, Z., Chen, S.X., 2017a. Cautionary tales on air-quality improvement in Beijing. Proc. Royal Soc. A 473, 20170457.
Zhang, W., Guo, J., Miao, Y., Liu, H., Song, Y., Fang, Z., He, J., Lou, M., Yan, Y., Li, Y., Zhai, P., 2018. On the summertime planetary boundary layer with different thermodynamic stability in China: a radiosonde perspective. J. Clim. 31, 1451–1465.
Zhang, X., Zhang, X., Chen, X., 2017b. Happiness in the air: how does a dirty sky affect mental health and subjective well-being? J. Environ. Econ. Manag. 85, 81–94.