Atmospheric Environment 285 (2022) 119242

Contents lists available at ScienceDirect

Atmospheric Environment

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

Episode based air quality assessment

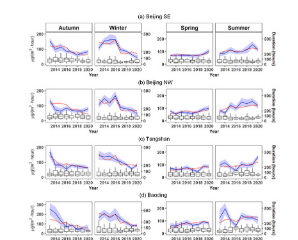
Shanshan Luo [a], Yuru Zhu [b], Song Xi Chen [b,c,*]

[a] *Center for Big Data Research, Peking University, Beijing, 100871, China*
[b] *Center for Statistical Science, Peking University, Beijing, 100871, China*
[c] *School of Mathematical Sciences and Guanghua School of Management, Peking University, Beijing, 100871, China*

HIGHLIGHTS

- Episode based air quality assessment to evaluate the average pollution loading.
- Pollution episode selection with a data driven algorithm.
- Downward trend in the episodic SO_2 and $PM_{2.5}$, while that for NO_2 was uncertain.
- Increased episodic O_3 in all seasons with that in autumn and winter at higher rates.

GRAPHICAL ABSTRACT



Adjusted (blue) and original (red) average episodic pollution loading of $PM_{2.5}$ in autumn and winter (left two columns) and O_3 in spring and summer (right two columns) for the four site clusters in three North China cities from 2013 to 2020, where the 95% confidence bands of the adjusted averages are indicated by the colored shade. The box plots are on the distributions of the episodic duration in hours.

ARTICLE INFO

Keywords:
Air-quality assessment
Pollution episode
Longitudinal data
Random forest
Meteorological adjustment

ABSTRACT

Air pollution naturally comes in episodes due to the repetition of the meteorological processes that remove and accumulate the pollutants. We present an episodic air quality assessment approach that consists of two steps. The first step selects the pollution episodes via a data driven algorithm that controls the wind conditions after a thorough removal process. The second step analyzes the episodic average and total pollution loading with respect to the within episode and prior episode meteorological conditions via the linear and random forest regression. A meteorological adjustment is conducted to remove the yearly meteorological variation in the estimated episodic air quality measures. Empirical analyses using the episode based assessment on four site clusters in three North China cities revealed that it offered sharper analysis and different perspectives on the underlying air quality than those based on the conventional full sample method due to the proper control in the episode selection.

1. Introduction

Air pollution is both an environmental and a public health problem, especially in developing countries. It can cause serious adverse effects on human health by either short-term or long-term exposure to air pollutants, and air pollution is responsible for increased mortality and hospital admissions, reduced life expectancy, and the prevalence of asthma and allergies (Brunekreef and Holgate, 2002; Kampa and Castanas, 2008; Kim et al., 2013; Ghorani-Azam et al., 2016). Air pollution is also an important factor in affecting climate changes and the ecohydrological processes (Duan et al., 2017). Research has found that air pollution can have huge economic costs on individuals and society (Zhang et al., 2017b; Taghizadeh-Hesary and Taghizadeh-Hesary, 2020).

Measuring pollution emissions and their impacts on the air quality has been an urgent but challenging task in air quality management. Emission inventory has been an important tool for emission

* Corresponding author. Center for Statistical Science, Peking University, Beijing, 100871, China.
E-mail addresses: lshanshan@stu.pku.edu.cn (S. Luo), yuruzhu@pku.edu.cn (Y. Zhu), csx@gsm.pku.edu.cn (S.X. Chen).

https://doi.org/10.1016/j.atmosenv.2022.119242
Received 9 March 2022; Received in revised form 14 June 2022; Accepted 15 June 2022
Available online 20 June 2022

measurement, which identifies and quantifies different sources of emission in a region based on economic and energy statistics as well as the estimates for the underlying economic and human activities; see Bouwman et al. (1997); Pacyna et al. (2006); Shi et al. (2014) for inventories of specific pollution species and Streets et al. (2003); Ohara et al. (2007); Zhang et al. (2009); Kuenen et al. (2014) for specific regions. Major shortcomings of the emission inventories are the measurement errors and the time delay in composing them, which can reduce their accuracy in the air quality studies.

Comprehensive numerical models with meteorological, emissions and chemical components have been established to conduct forecasting and regional air quality assessment. The Community Multiscale Air Quality (CMAQ) and the Nested Air Quality Prediction Modeling System (NAQPMS) are two such models (Byun and Schere, 2006; Liu et al., 2010; Wang et al., 2001), which can conduct scenario analysis to simulate the air quality under various meteorological and emission scenarios. A limitation of the approach is that the aforementioned measurement errors associated with the emission inventory would make the generated scenarios subject to substantial errors and thus affect the quality of the air quality assessment. Statistical models have been employed to model the meteorological effects on the air quality (Briggs et al., 2000). Liang et al. (2015) proposed a framework that provides air quality measures under a meteorological baseline condition, constructed based on historic data, to remove the meteorological confounding on the air quality measures; see Zhang et al. (2017a) for an extension to remove both the temporal and spatial meteorological variation.

It is known that the observed air quality is much affected by the meteorological conditions especially in the high temporal scales, as shown in Tai et al. (2010); Whiteman et al. (2014); Cheng et al. (2015), while Huang et al. (2021) provided ranks of variable importance for different pollutant species for North China cities. Zheng et al. (2015); Su et al. (2017) found a decrease of planetary boundary layer height hindered the vertical mixing of pollutants, resulting in a faster accumulation and higher concentrations. In addition to the general meteorological effects, wind condition is known to facilitate regional transportation of air pollutants, causing pollution to increase or decline under different wind directions. In the North China Plain (NCP), especially in the Beijing-Tianjin-Hebei region, regional transportation has played an important role in creating heavy pollution episodes (Lang et al., 2013; Zheng et al., 2015; Chang et al., 2019). To measure locally generated air pollution without the interference of regional transportation, Zhu et al. (2021) employed a data-driven algorithm to select calm periods after sustained northerly cleaning but before the arrival of the transported pollutants, and proposed the air quality measures that reflect the local emission in three cities in NCP. A panel data regression model was used to model the accumulation of the pollutant with respect to the meteorological variables during the calm periods, which led to estimates of the growth rates of different pollutants during the periods.

Naturally, air pollution comes in episodes due to regular repetitions of the weather cycle. A typical episode begins from a low pollution regime with calm wind after a cleaning, followed by a growth period caused by the local accumulation and regional transportation, and then another calm period usually accompanied with excessive pollution levels, and finally ends with a weather system that removes the pollution. As most of the severe pollution events happen in such episodes (Wang et al., 2018), there is a great need to conduct air quality assessment from the episodic aspects in order to gain more understanding on the episodic nature of the air pollution in a location. Such episodic assessment will supplement the commonly practised assessment using the entire data without an episodic design.

However, unlike the assessment with the entire data, the pollution episode and the accompanied meteorological data have to be selected via an algorithm based on meteorological and concentration related conditions. We will consider a more relaxed condition for $PM_{2.5}$ than that of Wang et al. (2018) which aimed at the severe pollution episode only with the 24-h average $PM_{2.5}$ concentration being above 150 μg/m^3 for at least two days. We use instead of $PM_{2.5}$ at 35 or 50 μg/m^3 as the threshold in the context of three North China cities, which would make the episodes include both severe and less severe pollution episodes and is more consistent with the World Health Organization's guideline for long term pollution exposure with respect to human health. There is no doubt that the episode selection criteria depends on the geophysical configuration of a city or region, and thus has to be city or region specific.

The episodic air quality assessment is made by first establishing the regression function of the total episodic pollution loading on regressors which quantifies the meteorology before and during the episodes. Both the linear regression and the non-parametric regression with the random forest method are considered, which show that the linear regression offered better in-sample fitting and the out-sample forecasting performance than the random forest, and thus was selected for later assessment. In order to remove potential bias caused by the meteorological variations, we formulate the meteorological adjusted total and average episodic pollution loading, and present ways to estimate these measures and their variations.

The paper is organized as follows. Details of data and methods are presented in Section 2, which consists of three subsections. Section 2.1 outlines the data and variables used in the analysis. The selection criteria for the pollution episode are given in Section 2.2. Section 2.3 provides modeling of the total episodic pollution loading via the linear and the non-parametric random forest regression, and the meteorological adjustment method. Section 3 reports the results of the episodic analysis on the average pollution loading of pollution episodes in the three cities from 2013 to 2020, which also reports the descriptive statistics on aspects of the episodes and the variable selection results. Section 4 concludes the paper. Additional information, extra tables and figures are provided in the supporting information (SI).

2. Material and methods

2.1. Data and variables

We considered hourly concentrations of six air pollutants $PM_{2.5}$, NO_2, CO, SO_2, O_3 and PM_{10} in three cities in North China Plain: Beijing, Tangshan and Baoding from March 2013 to February 2021, which contained 8 seasonal years. A seasonal year encompasses spring (March to May), summer (June to August), autumn (September to November) and winter (December to February of the following year). The air quality data were from twelve monitoring sites, which reported the hourly data directly to China National Environmental Monitoring Center in real time. The twelve monitoring sites made up four site clusters with each cluster having three sites, which were Beijing SE, Beijing NW, Tangshan, Baoding, respectively; see Table S1 in the SI for more specific information of the site clusters. To reduce the measurement errors, we applied a five-point moving average filter over the hourly time series data with weights 0.1, 0.2, 0.4, 0.2 and 0.1 for $t-2$, $t-1$, t, $t+1$ and $t+2$, respectively.

The considered meteorological variables were the air temperature (TEMP), the dew point temperature (DEWP), the relative humidity (HUMI), the air pressure (PRES), the wind speed (WS), direction (Wd) and the boundary layer height (BLH). The cumulative precipitation (R) referred to the sum of precipitation since the hour when it rained or snowed and was reset to zero when there was an hour without precipitation. Among the meteorological variables, the BLH data were obtained from the Global Reanalysis dataset ERA5 provided by the European Center for Medium-Range Weather Forecasts (ECMWF) at a grid resolution of 0.5 × 0.5 (latitude by longitude), and the other meteorological variables were surface measurements from the China Meteorological Administration (CMA) monitoring sites. We matched the BLH grid data to the closest CMA monitoring site. And then each air quality monitoring site cluster was matched to the nearest meteorological site from CMA for the meteorological conditions corresponding to the hourly pollution data. We took the logarithm of humidity (LogHUMI) and boundary layer height (LogBLH) to reduce the skewness of the variables. These variables are generally available and have been used in existing air quality studies such as Tai et al. (2010); Liang et al. (2015); Zheng et al. (2015); Wang et al. (2014).

According to Liang et al. (2015), in North China Plain, the 16 wind directions can be grouped into five broad categories: northwest (NW) which includes W, WNW, NW, NNW and N; northeast (NE) for NNE, NE and ENE; southeast (SE) covering E, ESE, SE, SSE and S; southwest (SW) having SSW, SW and WSW; and the calm and variable (CV). We merge NE and NW to form the northerly wind, and combine SE, SW and CV for the southerly wind. Furthermore, we define the cumulative northerly wind speed (CNWS) as the sum of hourly northerly wind speed, and the cumulative southerly wind speed (CSWS) is similarly defined. In both cases, whenever the wind direction changes, the cumulation is reset to 0. Having the cumulative wind speeds for the two broad wind directions is to model the cleaning and the accumulation of pollution as the northerly (southerly) wind tends to remove (accumulate) pollutants in North China Plain due to the geographical and economic configurations; see Liang et al. (2015) for more details.

There are variables associated with the pollution episodes, which will be defined after we define the episodes in the next section.

2.2. Pollution episodes selection

Our study is aimed at assessing the air pollution in the pollution episodes starting from a low level of $PM_{2.5}$ after a sustained cleaning, followed by a calm growth period due to the local emission accumulation, and then the regional transportation and air stagnation which usually accompanied with high and even severe pollution level, and finally the removal process by the northerly wind. We define the pollution episodes by extending the calm periods studied in Zhu et al. (2021) to cover the whole episodes as mentioned above. The purpose of Zhu et al. (2021) was for measuring the local emission's effects on the air quality, which considered only the beginning portion of the episodes before the regional transportation of pollutants.

We use the procedure in Zhu et al. (2021) to determine the beginning time t_s of an episode. The ending time t_ω of the northerly cleaning process satisfies

$$\mathrm{CNWS}_{t_\omega - 1} \geq 10.8\mathrm{m/s} \;\text{ and }\; \mathrm{CNWS}_{t_\omega} = 0. \tag{2.1}$$

The beginning time t_s of a pollution episode is located around t_ω, which corresponds to the lowest $PM_{2.5}$ in an 8-h neighborhood of t_ω within a calm, cleaned and dry period. Let $\mathcal{C}$ be the set of times when the system is calm, clean and dry satisfying

$$\mathrm{WS}_t \leq 5.4\mathrm{m/s}, \max\{\mathrm{PM2.5}_{t-1}, \mathrm{PM2.5}_t\} \leq 35\mu g/\mathrm{m}^3, \mathrm{R}_{t-1} = \mathrm{R}_t = 0. \tag{2.2}$$

It requires that $PM_{2.5}$ is not higher than 35 μg/m^3 for two consecutive hours, where 35 μg/m^3 is the 24-h primary standard according to the National Ambient Air Quality Standards. However, for Tangshan and Baoding, we replace 35 with 50 μg/m^3 in (2.2) due to more severe baseline pollution in the two cities as a result of heavier industrial installations in the two cities. Using the higher threshold was to ensure the sufficient number of episodes for the two cities.

Let $\mathcal{E}_{t_\omega}$ be the set of the ending times of the previously selected episodes that end before t_ω. The beginning time t_s of the episode is obtained by searching within an 8-h neighborhood of t_ω within $\mathcal{C}$ after the ending time of the previous episode, namely

$$t_s = \arg\min_{t \in \mathcal{B}_{t_\omega}} \mathrm{PM2.5}_t, \tag{2.3}$$

where $\mathcal{B}_{t_\omega} = [t_\omega - 8, t_\omega + 8] \cap (max\{t : t \in \mathcal{E}_{t_\omega}\}, L] \cap \mathcal{C}$, and L is the total length of observation time in a season for a site cluster; see Zhu et al. (2021) for more details.

A regular life cycle of a pollution episode starts from t_s followed by sustained pollution accumulation that makes the $PM_{2.5}$ peak at a time t_p, and then due to the arrival of the northerly cleaning process, $PM_{2.5}$ rapidly declined to a low level of $PM_{2.5}$, marking the end of the episode at a t_e. In the following, we explicitly define t_e.

To define t_e, we first locate the time, say t_{50}, when $PM_{2.5}$ has reached 50 μg/m^3 after t_s. If $PM_{2.5}$ can not reach 50 μg/m^3 before the next cleaning process arrives, the last obtained t_s is abandoned and the search for a new set of t_ω and t_s is re-started over the remaining time series. Otherwise, we search for the start of the cleaning period for the ending time t_e after t_{50}. It is noted that requiring $PM_{2.5}$ reaches 50 μg/m^3 during the episode does not contradict to the 50 μg/m^3 threshold used in defining the beginning time t_s for Baoding and Tangshan. For the purpose of locating t_e, we define two time sets $\mathcal{D}_{L1}$ and $\mathcal{D}_{L2}$ to quantify two conditions in the effort to define the ending time t_e of an episode. Moreover, Beijing, Tangshan and Baoding share the following definition of t_e which is classified into two cases associated with the strong and weaker cleaning, respectively. The $\mathcal{D}_{L1}$ describes a low $PM_{2.5}$ regime such that

$$\mathcal{D}_{L1} = \{t \,|\max\{\mathrm{PM2.5}_t, \mathrm{PM2.5}_{t+1}, \mathrm{PM2.5}_{t+2}\} \leq 35\mu g / \mathrm{m}^3\}, \tag{2.4}$$

while $\mathcal{D}_{L2}$ prescribes a slightly higher $PM_{2.5}$ state using 40 μg/m^3 as the threshold rather than 35 μg/m^3 with continued northerly wind,

$$\mathcal{D}_{L2} = \{t \,|\mathrm{PM2.5}_t \leq 40\mu g / \mathrm{m}^3 \;\text{ and }\; \max\{\mathrm{CNWS}_t, \mathrm{CNWS}_{t+1}, \mathrm{CNWS}_{t+2}\} > 0\}. \tag{2.5}$$

The ending time t_e is defined in two forms via $\mathcal{D}_{L1}$ and $\mathcal{D}_{L2}$, reflecting two ending patterns of the pollution episodes. The first one is

$$t_e = min_{t \in \mathcal{D}_{L1}}\{t : \mathrm{PM2.5}_t \leq min\{\mathrm{PM2.5}_{t-2}, \mathrm{PM2.5}_{t-1}\}, t > t_s\} \tag{2.6}$$

representing an ending with sustained lower $PM_{2.5}$ level usually accompanied by strong cleaning or removal processes. In the second form,

$$t_e = min_{t \in \mathcal{D}_{L2}}\{t : \mathrm{PM2.5}_{t-3} - \mathrm{PM2.5}_t \geq 50\mu g / \mathrm{m}^3, t > t_s\}. \tag{2.7}$$

Although this latter type of ending is also triggered by the northerly cleaning that makes $PM_{2.5}$ dropped by more than 50 μg/m^3 in the proceeding 3 h, it is different from the first type as the $PM_{2.5}$ may not necessarily drop below 35 μg/m^3 and the concentration may start to rise again and starts the next episode. Basically, the first type of ending is the result of strong and thorough cleaning while the second type is associated with weaker cleaning. Tables S2–S5 of the SI showed that 67%–98% of the episodes in the four site clusters were of the first type.

It is clear that our definition of episodes is quite different from the existing episode formulations which focused on severe pollution. For instance, Wang et al. (2018) aimed at severe pollution episodes by requesting the 24-h average $PM_{2.5}$ concentration being above 150 μg/m^3 for at least two days, and the t_s is min\{t: $PM2.5_{t:t+48} \geq 150$ μg/m^3\} and t_e is min\{t: t> t_s, $PM2.5_t < 150$ μg/m^3\}. However, this type of definition misses the initial calm and growth phase of the episodes, and does not provide a full account on the evolution of the episodes.

Fig. 1 displays two pollution episodes with the two types of ending defined in (2.6) and (2.7). In the first case, after experiencing a northerly cleaning process as shown by the cumulative northerly speeds, $PM_{2.5}$ dropped below 35 μg/m^3 and kept so for a period of time, indicating a strong and thorough cleaning process. In the second case, the cleaning process was weaker and could not reduce the concentration below 35 μg/m^3, but below 40 μg/m^3. At the t_e, the cleaning process had ended, and the concentration would not decrease further after reaching the minimum value before it would start the next episode.

After finding the ending time t_e, we went back to scan over the episode from t_s to t_e for the peak time when the maximum $PM_{2.5}$ was attained within the episode. Fig. 2 displays the time series of $PM_{2.5}$ versus the cumulative northerly and southerly wind speed (CNWS and CSWS) in Beijing's Dongsi site (in Beijing SE) and Baoding's Huadianerqu site over November 2018. Recall that we have used 50 μg/m^3

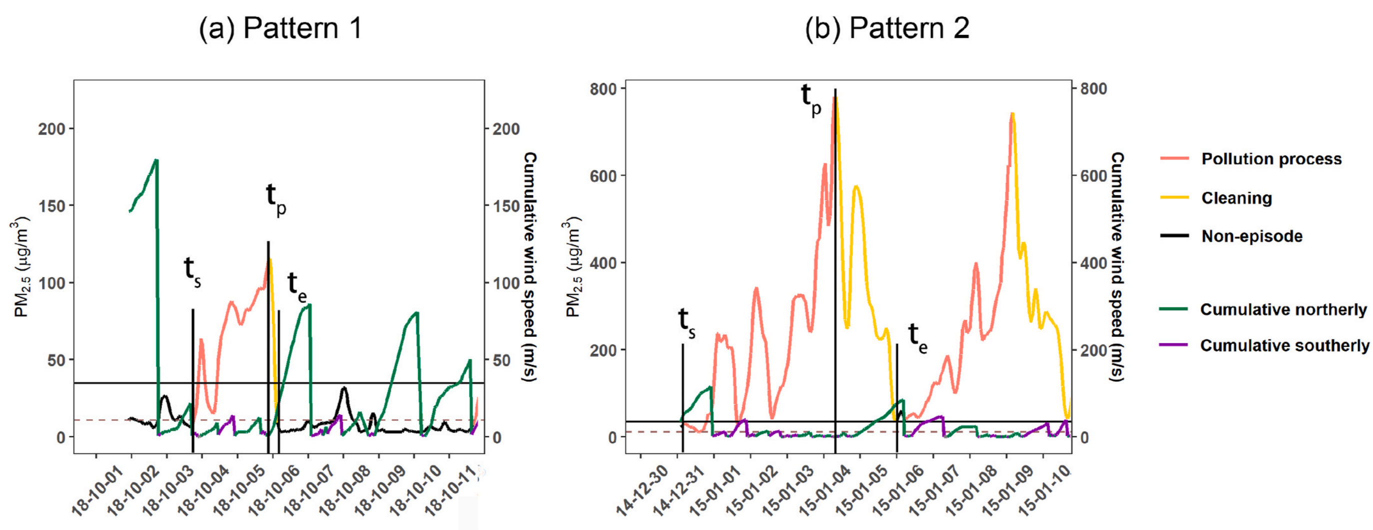


Fig. 1. Two pollution episodes with the two ending patterns caused by strong (left panel) and weak (right panel) northerly cleaning based on observations in Dongsi site in Beijing SE in the autumn of 2018 (left panel) and Huadianerqu site in Baoding in the winter of 2014 (right panel). The $PM_{2.5}(\mu g/m^3)$ series was divided to three states: pollution (red), cleaning (yellow) and non-episode (black) with the cumulative northerly (southerly) winds speed (m/s) marked by green (purple). The black solid and brown dashed lines mark 35 $\mu g/m^3$ and 10.8 m/s, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

instead of 35 $\mu g/m^3$ for Baoding and Tangshan when defining the start of episodes. These two figures confirmed the need as otherwise in Tangshan and Baoding there would be quite some episodes that could not be selected because the concentration at the start time was higher than 35 $\mu g/m^3$ but lower than 50 $\mu g/m^3$, while it was obvious that there was an episode being formulated afterward.

After having selected the episodes for a site from the original data series, we define variables to measure the wind condition in the 24 h before and 48 h after the starts of the episodes. Specifically, MCNWS (MCSWS) is the maximum of the cumulative northerly (southerly) wind speed in the 24 (48) hours before (after) the episodes, which quantifies the extent of cleaning (transportation) before (after) the episodes. Similar SNWS (SSWS) represents the sum of hourly northerly (southerly) wind speed in the 24 (48) hours before (after) the episodes, and PNWS (PSWS) is the percentage of northerly (southerly) wind in the 24 (48) hours before (after) the episodes. The total pollution loading of an episode is

$$PL = \sum_{t=t_s}^{t_e} \mathrm{PM2.5}_t \tag{2.8}$$

and the duration or length of the episode is $\mathrm{Dur} = t_e - t_s$.

As each site cluster includes multiple sites, after determining the pollution episodes for each site, we take the union of the episode time periods of the sites to form episodes of the site cluster. The summary episodic statistics and analysis in the rest of the paper are based on episodes of the site clusters.

2.3. Statistical methods

For the four site clusters, the average gap times (standard errors) between two consecutive episodes over the eight years were 46(3.4), 70 (4.6), 51(3.4) and 45 (3.6) hours in spring, summer, autumn and winter, respectively. The sufficient gap times between the adjacent episodes suggested that it is reasonable to assume that different episodes may be regarded as statistically independent, which leads to a linear model for the total pollution loading for the episodes. Specifically, for a site cluster and a season of year i, let PL_{ij} be the total pollution loading of a pollutant during the episode for $i = 1, \ldots, A$ and $j = 1, \ldots, n_i$, where $A = 8$ represents the total number of the seasonal years from 2013 to 2020, and n_i is the total number of episodes in year i of the season in the site cluster.

Let $M_{ij} = (\mathrm{DEWP}_{ij}, \mathrm{PRES}_{ij}, \mathrm{TEMP}_{ij}, \mathrm{LogBLH}_{ij}, \mathrm{LogHUMI}_{ij})$ be a 5-dimensional row vector of averages of the five meteorological variables during the j-th episode in year i, $N_{ij} = (\mathrm{MCNWS}_{ij}, \mathrm{SNWS}_{ij}, \mathrm{PNWS}_{ij})$ be the row vector of pre-episode summary on the northerly wind speed (maximum cumulative, the total and the proportion) in the 24 h before the start of the episode, and $S_{ij} = (\mathrm{MCSWS}_{ij}, \mathrm{SSWS}_{ij}, \mathrm{PSWS}_{ij})$ collects the specifics of the southerly wind in the 48 h since the start of the episode. Dur_{ij} denotes the duration $t_e - t_s$ of the episode.

2.3.1. Linear regression

We first consider the linear regression of the total pollution loading PL of an episode on the meteorological variables before and during the episode as well as its duration. The model in year i for a cluster of a season is

$$PL_{ij} = M_{ij}\alpha_i + N_{ij}\gamma_i^1 + S_{ij}\gamma_i^2 + \mathrm{Dur}_{ij}\eta_i + b_i + \epsilon_{ij} = U_{ij}\beta_i + \mathrm{Dur}_{ij}\eta_i + \epsilon_{ij} \tag{2.9}$$

where $U_{ij} = (M_{ij}, N_{ij}, S_{ij}, 1)$ is the m-variate vector of explanatory variables, m is the number of regressors other than Dur_{ij}, and $\alpha_i, \gamma_i^1, \gamma_i^2$ are the coefficient vectors corresponding to the variables M, N, S, respectively, and $\beta_i = (\alpha_i^\top, \gamma_i^{1\top}, \gamma_i^{2\top}, b_i)^\top$ is the m-dimensional coefficient vector to U_{ij}. Here b_i is the intercept, and ϵ_{ij} is possibly heterogeneous random error with zero conditional mean and finite conditional variance given the explanatory variables.

It is noted that the above model is different from Zhu et al. (2021), which used the panel-data regression model. The current proposal is based on the episode-wise variables as shown above to model the pollution level over the entire pollution episodes, while Zhu et al. (2021) is on hourly data as the purpose was to measure the hourly growth rate of pollutants in the initial calm period of the episodes for local emission measurement.

To avoid model over-fitting, we select the important variables by the forward step-wise method based on the Bayesian information criterion (BIC) (Schwarz, 1978), which chooses one variable at each step that leads to the largest reduction in the BIC until no more variables can be added to reduce the BIC; a similar forward selection method based on the mean square errors was used in Huang et al. (2021). Specifically, for each year and site cluster, we conduct the forward variable selection and record the rank of the selected variables. The ranks of the unselected variables are given the average of the remaining ranks of the unselected

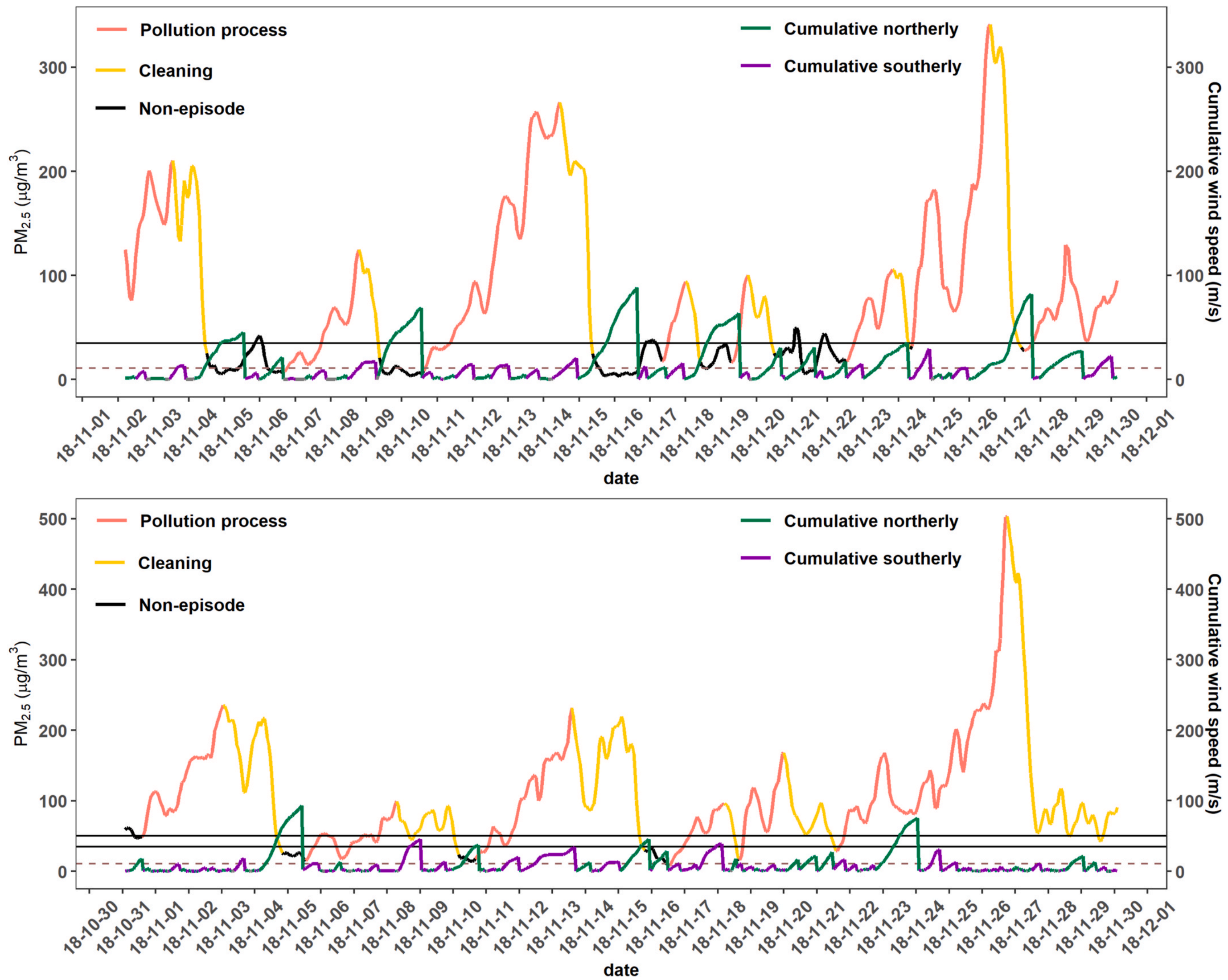


Fig. 2. Time series of $PM_{2.5}$($\mu g/m^3$), cumulative northerly (green) or southerly (purple) wind speed (m/s) in November 2018 in Dongsi site (top) in Beijing SE and Huadianerqu site (bottom) in Baoding with the during episode (cleaning) $PM_{2.5}$ marked in red (yellow), and otherwise in black. The black solid line and brown dashed line mark 35 $\mu g/m^3$ and 10.8 m/s, respectively. In Baoding, the 50 $\mu g/m^3$ concentration is also marked with a solid black line. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

variables. We then attain an initial variable importance order by averaging the yearly ranks over the eight years.

The final variable importance ordering is attained by calculating the cumulative R^2 and BIC criteria when we successively add one variable according to the initial variable order attained above. The coefficient of determination

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{2.10}$$

measures the goodness of the model and the explaining power of the selected variables, where SS_{tot} and SS_{res} refer to the total and the residual sums of squares of the regression, respectively. The variable selection ends at the variable that attains the minimum BIC values within the variables whose cumulative R^2s are within 0.03 of the R^2 attained by using all variables.

2.3.2. Random forest

Random forest (RF) (Ho, 1995; Breiman, 2001) offers a regression estimation and prediction in a model-free non-parametric fashion alternative to the linear regression. It is based on averaging over ensembles of regression trees. We consider applying the random forest on the same episode data and compare its performance with the linear regression to offer a wider perspective. In this study, the number of random forest trees is set to 500. A similar variable selection procedure for the linear regression was employed for the RF regression with the detail reported in the SI.

2.3.3. Meteorological adjustment

Since meteorological conditions vary from year to year, in order to remove the effect of the annual meteorological variation from the estimated total pollution loading, we construct a baseline distribution to adjust for the meteorological variation. Liang et al. (2015) proposed a framework to adjust for the meteorological confounding via a meteorological baseline to make the pollutant concentration comparable temporally. A similar meteorological adjustment for estimating the growth rate was used in Zhu et al. (2021). Let $f_i\,(u|l)$ be the conditional density of $U_{i\cdot}$ given $\mathrm{Dur}_{i\cdot} = l$, where $U_{i\cdot} = (M_{i\cdot}, N_{i\cdot}, S_{i\cdot}, 1)$ is the regression variables with intercept, and $p_{il} = \frac{n_{il}}{n_i}$ for $l = 1, \ldots, T_i$ be a set of weights, where T_i is the maximum duration of episodes in year i, and n_{il} is the number of episodes whose duration is l in year i. A = 8 represents the total number of the seasonal years from 2013 to 2020. The baseline probability density function from 2013 to 2020 for the meteorological variables is

$$f.(u,l) = \frac{1}{A}\sum_{a=1}^{A} f_a(u,l) = \frac{1}{A}\sum_{a=1}^{A} p_{al} f_a(u|l) \tag{2.11}$$

From (2.9) of the linear regression,

$$E(PL_{ij}|U_{ij} = u_{ij}, \mathrm{Dur}_{ij} = l_{ij}) = u_{ij}\beta_i + l_{ij}\eta_i. \tag{2.12}$$

The adjusted average pollution loading PL in year i is the mean of PL_{ij} for (U_{ij}, Dur_{ij}) distributed with $f.(u, l)$ as the probability density, that is

$$\begin{aligned} PL_i^* &= \int_{u,l} E(PL_{ij}|u,l) f.(u,l)\,dudl \\ &= \frac{1}{A}\sum_{a=1}^{A}\int_l p_{al}\int_u (u\beta_i + l\eta_i) f_a(u|l)\,dudl \\ &= \frac{1}{A}\sum_{a=1}^{A}\{E(U_{aj})\beta_i + E(\mathrm{Dur}_{aj})\eta_i\}. \end{aligned} \tag{2.13}$$

As in Liang et al. (2015), the meteorologically adjusted mean PL_i^* can be estimated by

$$\widehat{PL}_i = \frac{1}{A}\sum_{a=1}^{A} n_a^{-1}\sum_{j=1}^{n_a}(U_{aj}\widehat{\beta}_i + \mathrm{Dur}_{aj}\widehat{\eta}_i). \tag{2.14}$$

We use the method of Zhu et al. (2021) to obtain the robust variance estimator of $\widehat{PL}_i$:

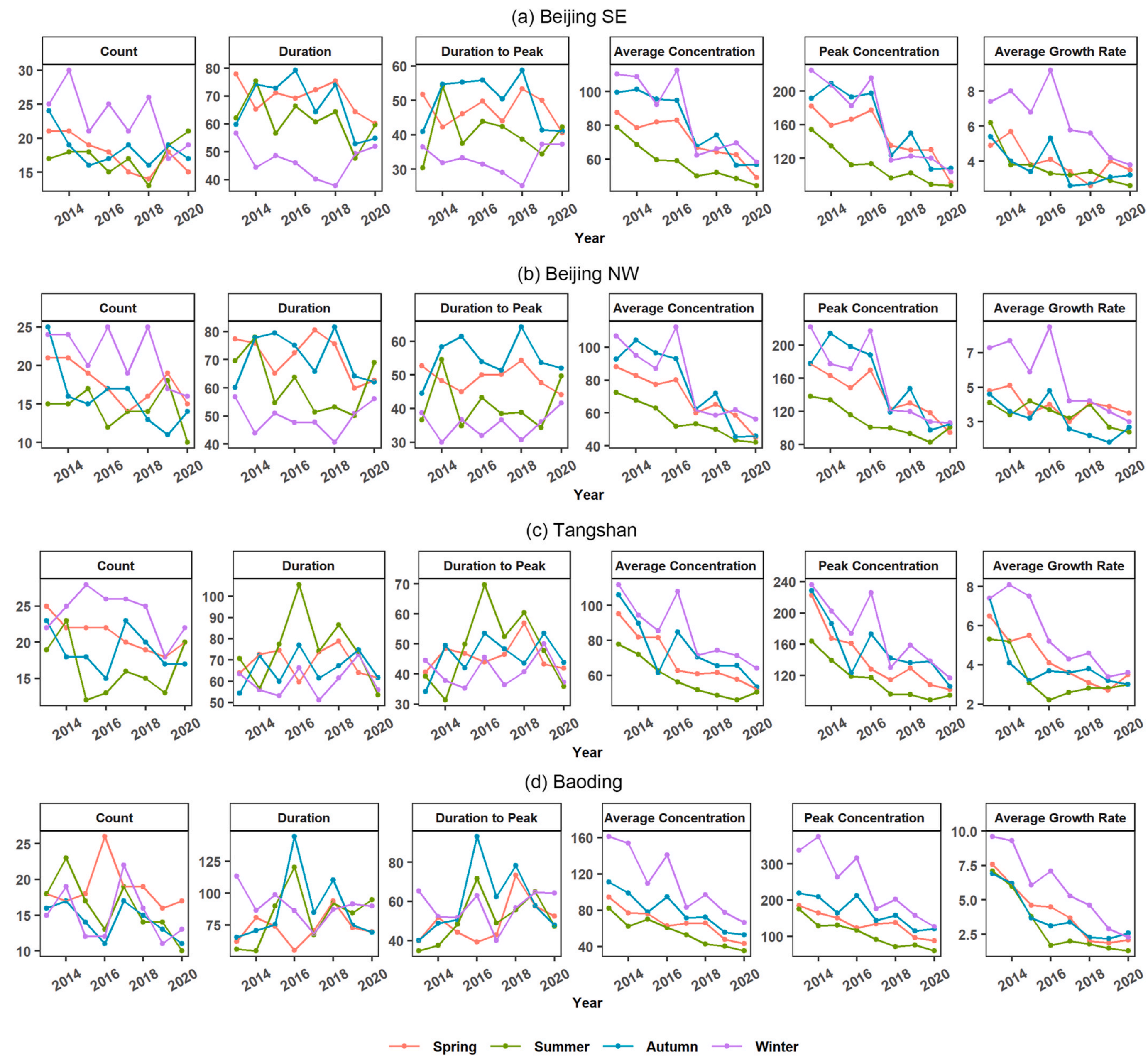


Fig. 3. Summary statistics of the selected episodes in (a) Beijing SE, (b) Beijing NW, (c) Tangshan and (d) Baoding for each season from March 2013 to February 2020 including the average number of episodes (Count), the average concentration of the episode, the maximum (Peak Concentration), the average length of the episodes in hours (Duration), the average length from t_s to t_p in hours (Duration to Peak), the average growth rate of pollution process from t_s to t_p.

$$\widehat{\mathrm{Var}}\left(\widehat{\theta}_i\right)=\left(\sum_{j=1}^{n_i}X_{ij}X_{ij}^{\top}\right)^{-1}\left[\sum_{j=1}^{n_i}(X_{ij}\widehat{\epsilon}_{ij})(X_{ij}\widehat{\epsilon}_{ij})^{\top}\right]\left(\sum_{j=1}^{n_i}X_{ij}X_{ij}^{\top}\right)^{-1} \tag{2.15}$$

where $X_{ij}=(U_{ij},\mathrm{Dur}_{ij})^{\top}$ is the vector of selected covariates for episode j in year i, and $\theta_i=(\beta_i^{\top},\eta_i)^{\top}$ be the regression coefficient vector and $\widehat{\epsilon}_{ij}=PL_{ij}-X_{ij}^{\top}\widehat{\theta}_i$ is the residual. The consistency and the asymptotic normality of $\widehat{PL}_i$, and the validity of the variance estimation under some assumptions can be found in the SI of Zhu et al. (2021).

A similar meteorologically adjusted RF estimator for the total pollution loading PL_i is

$$\widehat{PL}_i^{RF}=\frac{1}{A}\sum_{a=1}^{A}n_a^{-1}\sum_{j=1}^{n_a}\widehat{RF}_i(U_{aj},\mathrm{Dur}_{aj}) \tag{2.16}$$

by substituting the past years variable information $\{(U_{aj},\mathrm{Dur}_{aj})\}$ to the estimated regression function by the RF. Let $\widehat{\mathrm{Var}}\left(\widehat{RF}_i(U_{aj},\mathrm{Dur}_{aj})\right)$ be the variance of the RF regression obtained based on 500 replications. Then,

$$\widehat{\mathrm{Var}}\left(\widehat{PL}_i^{RF}\right)=\frac{1}{A^2}\sum_{a=1}^{A}n_a^{-2}\sum_{j=1}^{n_a}\widehat{\mathrm{Var}}\left(\widehat{RF}_i(U_{aj},\mathrm{Dur}_{aj})\right). \tag{2.17}$$

Considering that the duration of the episode may change (in fact a downward trend in the four site cluster as shown in Fig. 3), a more objective measure for the episode assessment is the meteorologically adjusted average episodic pollution loading $I_i^*=PL_i^*/E(\mathrm{Dur}_{i\cdot})$, which is the ratio of the meteorologically adjusted mean episodic total pollution loading and the mean episodic duration. It can be estimated by

$$\widehat{I}_i^*=\widehat{PL}_i\Big/\overline{\mathrm{Dur}}_i=\frac{1}{A}\sum_{a=1}^{A}n_a^{-1}\sum_{j=1}^{n_a}(U_{aj}\widehat{\beta}_i+\mathrm{Dur}_{aj}\widehat{\eta}_i)\Big/\overline{\mathrm{Dur}}_i \tag{2.18}$$

where $\overline{\mathrm{Dur}}_i$ is the average duration of all episodes in a certain season of the i-th year. The standard deviation of $\widehat{I}_i^*$ can be formulated accordingly based on that for $\widehat{PL}_i$ and the delta method, with the detailed formula given in the SI.

3. Results and discussion

3.1. Features and trend of episodes

We report the main features and trend of the selected pollution episodes to gain insights on the air quality from the episodic point of view rather than the conventional approach based on the entire data of a season. The episodic data mirror those collected via a statistical design of experiments in many fields of natural science as it has exercised controls in the sample selection to remove the less important aspects of the pollution process. Fig. 3 displays the seasonal summary statistics on the pollution episodes of the four site clusters in Beijing, Tangshan and Baoding from 2013 to 2020, while Tables S2–S5 in the SI provide their numeric values and the standard errors.

The figure and the tables show that the number of episodes was the highest in winter for Beijing SE, Beijing NW and Tangshan as the episodes of the northerly were the most frequently in winter in Beijing and Tangshan. In contrast, Baoding's autumn and winter had a lower number of episodes than the other three site clusters. In fact, there was less variation in the number of episodes in Baoding among the four seasons although the number of episodes in spring tended to be the most frequent. The fewer numbers of episodes in Baoding reflect its being less influenced by the northerly wind in terms of frequency and velocity as shown in Table 1, largely because it is geographically located to the south of the other two cities and more to the center of the North China Plain. Fig. 3 shows a clear negative correlation between the number of episodes and their duration, and the duration to the peak $PM_{2.5}$ in the episodes. These are expected as fewer (more) episodes means prolonged (shortened) pollution period and fewer numbers of the northerly wind processes to remove the pollutants.

Despite the winter season had the most pollution episodes with less duration due to more frequent northerly removal in Beijing and Tangshan, the episodic average and maximum $PM_{2.5}$ concentration tended to be the highest in winter, which was also the case for Baoding. This reflected the increased emission coupled with lower boundary layer height in winter. The autumn season ranked just behind winter in terms of the average and maximum episodic $PM_{2.5}$, followed by spring and summer, which were consistent with the seasonal pattern using the entire data as shown in Chen et al. (2018). Both the episodic average and maximum $PM_{2.5}$ showed an overall downward trend for each season over the eight years from 2013 to 2020, reflecting an overall improvement in the air quality in North China Plain. By the year 2020, the average hourly episodic growth rates were all below 4 μg/(m^3 ·hour) for each site cluster in all seasons. The growth rates in 2020 ranged between 1/3 to 1/2 of the corresponding levels in 2013.

Seasonal occupancy rates of the pollution episodes to the total hours are reported in Tables S2–S5 of the SI. The annual average occupancy rate from 2013 to 2020 at four site clusters and the rates of decline in 2020 relative to those in 2013 can be found in Table 2. In Beijing SE and NW site clusters, the occupancy rates have been stable to be less than 50% since 2017; while in Tangshan and Baoding, the occupancy rates in most years were higher than those in Beijing. In the four site clusters, compared to their respective occupancy rates in 2013, the occupancy rates in 2020 decreased by 27.1% for Beijing SE, 38.1% for Beijing NW, 18.6% for Tangshan and 15.3% for Baoding respectively. Although these were substantial decreases, they were far lower than the proportion of the decline in the episodic growth rate, average concentration and peak concentration. The reduced occupancy rates were consistent with the increased average gap time between two episodes also reported in Tables S2–S5. In Beijing SE and Beijing NW, by 2020, except for summer, the average gap times had been more than doubled as compared to those in 2013. In Tangshan, the average gap time for the four seasons in 2013 and 2020 were 26.2 (2.2) and 48.4 (3.3), respectively, and those in Baoding were 46.4 (5.8) and 75.0 (9.5), respectively, representing 84.7% and 61.6% increase, respectively.

3.2. Variable selection results and fitting performance of the models

In linear regression for $PM_{2.5}$, Table 3 reports the incremental change of R^2 and BIC after adding a variable each time in each season in the Beijing SE and NW, while those for the other two site clusters are presented in Table S6 in the SI. It can be seen that when the selection of variables was stopped, the R^2 had largely peaked. Table S7 in the SI reports the final selected variables and the order in all seasons in the four

Table 1
Seasonal average northerly wind speed (m/s) and proportion from 2013 to 2020 at four site clusters. The numbers inside the parentheses are the standard errors.

Site Cluster	Spring		Summer		Autumn		Winter	
	Average	Percentage	Average	Percentage	Average	Percentage	Average	Percentage
Beijing SE	2.16 (0.010)	45.3% (0.22%)	1.52 (0.006)	43.9% (0.22%)	1.69 (0.008)	59.4% (0.22%)	2.13 (0.008)	69.3% (0.20%)
Beijing NW	1.74 (0.008)	52.6% (0.22%)	1.10 (0.005)	52.6% (0.22%)	1.28 (0.006)	65.6% (0.21%)	1.80 (0.007)	74.6% (0.19%)
Tangshan	3.11 (0.014)	42.9% (0.22%)	1.97 (0.009)	39.8% (0.22%)	2.16 (0.010)	55.9% (0.22%)	2.41 (0.010)	61.7% (0.22%)
Baoding	2.88 (0.012)	42.2% (0.22%)	2.27 (0.009)	45.9% (0.22%)	2.05 (0.010)	48.7% (0.23%)	2.13 (0.010)	51.1% (0.22%)

Table 2
Annual average occupancy rate is from 2013 to 2020 at the four site clusters and the reduction percentages in the occupancy rate (R_{Occu}), hourly growth rate (R_{Growth}), average concentration (R_{AVE}) and the peak concentration (Red_{Peak}) of $PM_{2.5}$ in 2020 relative to those in 2013, respectively.

Site Cluster	2013	2014	2015	2016	2017	2018	2019	2020	R_{Occu}	R_{Growth}	R_{AVE}	R_{Peak}
Beijing SE	63.5%	63.0%	51.2%	53.8%	48.4%	45.7%	44.9%	46.3%	27.1%	45.4%	44.6%	48.2%
Beijing NW	63.8%	58.7%	49.8%	51.4%	44.0%	45.9%	40.9%	39.5%	38.1%	44.2%	47.5%	43.2%
Tangshan	64.2%	63.9%	58.1%	63.3%	61.6%	64.8%	53.6%	52.3%	18.6%	50.6%	43.7%	50.8%
Baoding	55.3%	62.8%	58.2%	62.6%	63.5%	70.2%	49.0%	46.8%	15.3%	73.3%	55.9%	57.0%

Table 3
The cumulative R^2 and BIC for successively adding a variable each time according to the initial variable ordering in each season using eight-year data in Beijing SE and Beijing NW site clusters, where the variables marked with an asterisk were the last variables in the final variable ordering in the linear regression model for episode-wise $PM_{2.5}$ loading.

(a) Beijing SE

Spring	variables	Dur	LogBLH	PNWS	TEMP	LogHUMI	MCSWS	SNWS(*)	MCNWS	SSWS	PRES	DEWP	PSWS
	R^2	0.84	0.87	0.87	0.89	0.90	0.90	0.91	0.92	0.92	0.93	0.93	0.94
	BIC	973	966	967	967	965	966	965	965	968	968	965	968
Summer	variables	Dur	MCNWS	DEWP	LogBLH	PNWS	MCSWS(*)	SNWS	TEMP	PSWS	PRES	SSWS	LogHUMI
	R^2	0.88	0.89	0.91	0.91	0.92	0.93	0.93	0.94	0.94	0.94	0.94	0.94
	BIC	878	876	871	872	872	870	873	874	875	874	876	880
Autumn	variables	Dur	MCSWS	LogBLH	TEMP	PRES	MCNWS	DEWP	SNWS	PNWS(*)	PSWS	LogHUMI	SSWS
	R^2	0.77	0.79	0.81	0.82	0.83	0.84	0.86	0.86	0.87	0.88	0.89	0.89
	BIC	1021	1020	1018	1017	1017	1017	1016	1017	1016	1017	1017	1018
Winter	variables	Dur	SNWS	LogBLH	MCNWS	SSWS	PNWS	MCSWS	PSWS(*)	DEWP	PRES	LogHUMI	TEMP
	R^2	0.80	0.84	0.84	0.85	0.86	0.87	0.88	0.89	0.89	0.89	0.89	0.91
	BIC	1267	1259	1260	1261	1262	1261	1262	1260	1262	1264	1266	1261

(b) Beijing NW

Spring	variables	Dur	TEMP	LogBLH	SNWS	LogHUMI	MCNWS	PRES	MCSWS(*)	PSWS	PNWS	DEWP	SSWS
	R^2	0.83	0.86	0.87	0.88	0.89	0.90	0.91	0.92	0.92	0.92	0.93	0.93
	BIC	987	983	983	983	981	979	980	978	979	981	980	981
Summer	variables	Dur	PNWS	TEMP	DEWP	PRES	LogHUMI	SNWS	SSWS(*)	MCNWS	LogBLH	MCSWS	PSWS
	R^2	0.85	0.87	0.89	0.91	0.92	0.93	0.93	0.94	0.94	0.95	0.95	0.95
	BIC	743	738	738	735	733	734	735	729	732	734	735	736
Autumn	variables	Dur	PNWS	TEMP	LogBLH	SNWS	MCNWS	PSWS	DEWP	LogHUMI(*)	MCSWS	SSWS	PRES
	R^2	0.69	0.71	0.75	0.78	0.82	0.83	0.84	0.85	0.87	0.88	0.88	0.89
	BIC	912	911	909	907	902	903	903	904	902	902	904	902
Winter	variables	Dur	DEWP	PNWS	SNWS	PSWS	LogBLH	LogHUMI	MCNWS	MCSWS(*)	SSWS	TEMP	PRES
	R^2	0.80	0.82	0.83	0.84	0.85	0.86	0.86	0.87	0.88	0.88	0.89	0.90
	BIC	1176	1171	1171	1172	1173	1175	1177	1180	1179	1181	1182	1182

site clusters. The duration "Dur" was always selected first. Regarding the wind direction and speed variables, the three variables related to the northerly wind were more important and the two variables related to the southerly wind tended to be selected last. The boundary layer height and temperature were the two most important meteorological variables.

It is noted that some regressors are correlated, for instance DEWP, TEMP and LogHUMI. According to the Magnus-Tetens formula (Lawrence, 2005; Tetens, 1930), DEWP, LogHUMI and TEMP admit a non-linear relationship

$$\mathrm{DEWP} = -B_1 + B_1\left\{1 - \mathrm{Log}\left(\frac{\mathrm{HUMI}}{100}\right)\Big/A_1 - \frac{\mathrm{TEMP}}{B_1 + \mathrm{TEMP}}\right\}^{-1} \tag{3.1}$$

where $A_1 = 17.27$ and $B_1 = 237.3$ °C. With the data used in this study, we conduct two regression analyses of the DEWP. One is the linear regression of the DEWP on the other two variable, another is the non-linear regression according to (3.1). Fig. S1 of the SI displays the estimated residual plots of the two regression, which show that the residual plots of the linear regression of the DEWP displayed a clear non-linear curvature indicating a lack of fit of the linear regression. In contrast, the non-liner regression according to the Magnus-Tetens formula offered almost perfect residual fits. As the proposed episode based air quality analysis is based on the linear model (2.9), we include all three variables in the analysis.

The variable selection procedure is designed to remove the potential redundancy among the regressors. Indeed, only 5 of the 16 site cluster-season combinations selected DEWP and LogHUMI at the same time. For the other five pollutants, the final selected variable results are shown in Tables S8–S12 which imply the important meteorological factors are different among different pollutants.

Fig. S2 displays the yearly standardized residual plots for each season obtained by fitting the linear regression with the selected variables in the four site clusters. It can be seen that the 1st and the 3rd quartiles largely fluctuated within [− 0.3, 0.3]. This together with the corresponding plots of the model fitted versus the observed pollution loading in Figs. S3–S6 of the SI suggest reasonable fits of the linear regression model.

Fig. S7 in the SI reports the estimated linear regression coefficients. It shows that the coefficients to the episodic duration (Dur) were all significantly positive, those to the dew point temperature (DEMP) and the humidity (LogHUMI) were largely positive, and those to the air pressure (PRES) and boundary layer height (LogBLH) were mainly negative. These were consistent with the underlying geophysics and the existing research.

As mentioned in the definition of the pollution episodes, the greater the intensity of the northerly wind before the pollution episode, the more thorough the cleaning and lower the pollution intensity; while the greater intensity of the southerly wind during the episode can increase the intensity of pollution. Consequently, the coefficients of MCNWS, SNWS and PNWS related to northerly wind were mainly negative. Moreover, these three variables were important for the two site clusters in Beijing. In the Beijing SE, except for MCNWS in spring and SNWS in summer, all three variables were selected into the model in all four seasons; in the Beijing NW, except for MCNWS in summer, PNWS in spring, all three variables were selected in all four seasons. The coefficients of MCSWS, SSWS and PSWS were mainly positive, which was also consistent with the known effect of southerly winds.

Table S13 reports the selected variables in all seasons of the four clusters by the RF. Compared with the linear regression, fewer variables were selected by the RF, which was especially the case in Tangshan and Baoding. The duration was still the most important one and was always selected first, and the humidity, dew point temperature and boundary layer height were ranked immediately after the duration as the most important meteorological variables.

In the linear regression, the northerly wind related variables were ranked high while those related to the southerly wind were ranked low. However, this was not the case for the variables selected by the RF.

Fig. F8 of the SI provides the yearly standardized residual plots for the random forest regression in the four site clusters. Comparing the residual plots with those of the linear regression in Figure S2, there were more outliers associated with the RF, indicating its inferior fitting performance.

To compare the performance of the linear and the random forest regression, both the in-sample fitting and out-of-sample prediction RMSE were considered as the performance measures, see SI for comparison based on R^2 whose conclusion was the same. Fig. 4 displays both the standardized in-sample and out-sample RMSEs for each season from 2013 to 2020, while an overall eight-year average can be found in Table S14. Among the 32 season-year combinations over the eight years in each site cluster, the linear regression had much smaller in-sample and out-of-sample RMSEs than those of the RF in the majority of the combinations. This was the case for Beijing SE in 22 out of 32 season-year combinations, and 19, 29, and 27 season-year combinations for Beijing NW, Tangshan and Baoding, respectively. The average standardized in-sample RMSEs (standard errors) of the linear regression were 0.30(0.018), 0.30(0.014), 0.29(0.013), 0.27(0.016) for the four site clusters, while those for the random forest were 0.35(0.015), 0.39

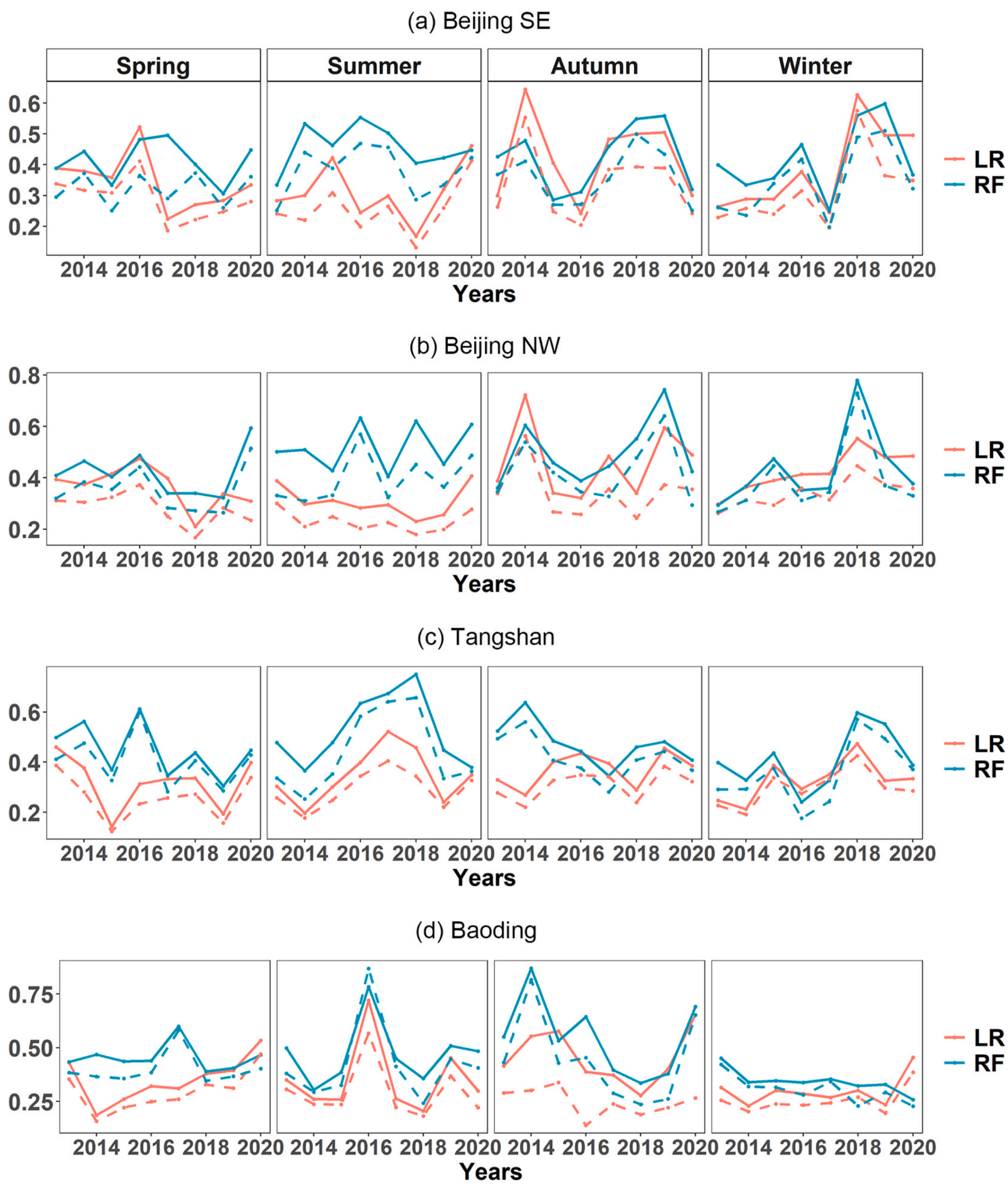


Fig. 4. Avearge in-sample (dashed lines) and out-of-sample (solid lines) root mean square errors (RMSE) of the linear (LR, red) and random forest (RF, blue) regression for four seasons from 2013 to 2020 in the four site clusters. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

(0.020), 0.40(0.022), 0.39(0.027), respectively. The average out-of-sample RMSEs in the four site clusters for the linear regression were 0.37(0.021), 0.39(0.019), 0.34(0.016), 0.36(0.023), and those for the random forests were 0.43(0.016), 0.47(0.021), 0.46(0.021), 0.45 (0.025), respectively. These suggested the linear regression was better suited for the study, while offering an easy geophysical interpretation.

3.3. Trends and changes in average pollution loading

We report the results of the episode based air quality assessment using the proposed adjustment approach to remove the potential bias caused by the meteorological anomaly. According to the performance of the linear and random forest regression reported in Section 3.2, we considered only the linear regression in the episode based assessment.

Figs. 5–7 display the seasonal meteorologically adjusted average pollution loading $\widehat{I}_i^*$ for $PM_{2.5}$, NO_2 and SO_2, and O_3 (only spring and summer), respectively in the four site clusters from 2013 to 2020; while the results of CO and PM_{10} are reported in Figs. S9 and S10. The distributions of the episodic duration are displayed in Fig. 5. The episodes selected based on the $PM_{2.5}$ were used as the episodes for the other species. It is observed that there was an overall downward trend in the average episodic pollution loading of $PM_{2.5}$ and SO_2 in all site clusters and all seasons, and the downward trend was well established since 2016 for $PM_{2.5}$ and earlier for SO_2. The very substantial reduction in the SO_2 in all seasons and especially in the winter season was very striking, which reflected the sustained effort in improving the way that coal is consumed throughout North China Plain since 2013. The declining trend for NO_2 was less pronounced as compared with those of $PM_{2.5}$ and SO_2. The episodic average ozone was largely on an increasing trend in the spring and summer in all four site clusters, which was consistent with the previous assessment using the entire sample (Chen et al., 2018; Li et al., 2021).

It is noted that there were substantial differences between the raw and the meteorologically adjusted episodic average concentration for all the pollutants, and many of the raw averages were outside the 95% confidence intervals of the adjusted averages. This suggests the differences between the raw and the adjusted averages were statistically significant, and the need for the meteorological adjustment.

Fig. 5 also suggested that the average episodic duration had been stable over the eight years for all seasons and all site clusters. This is

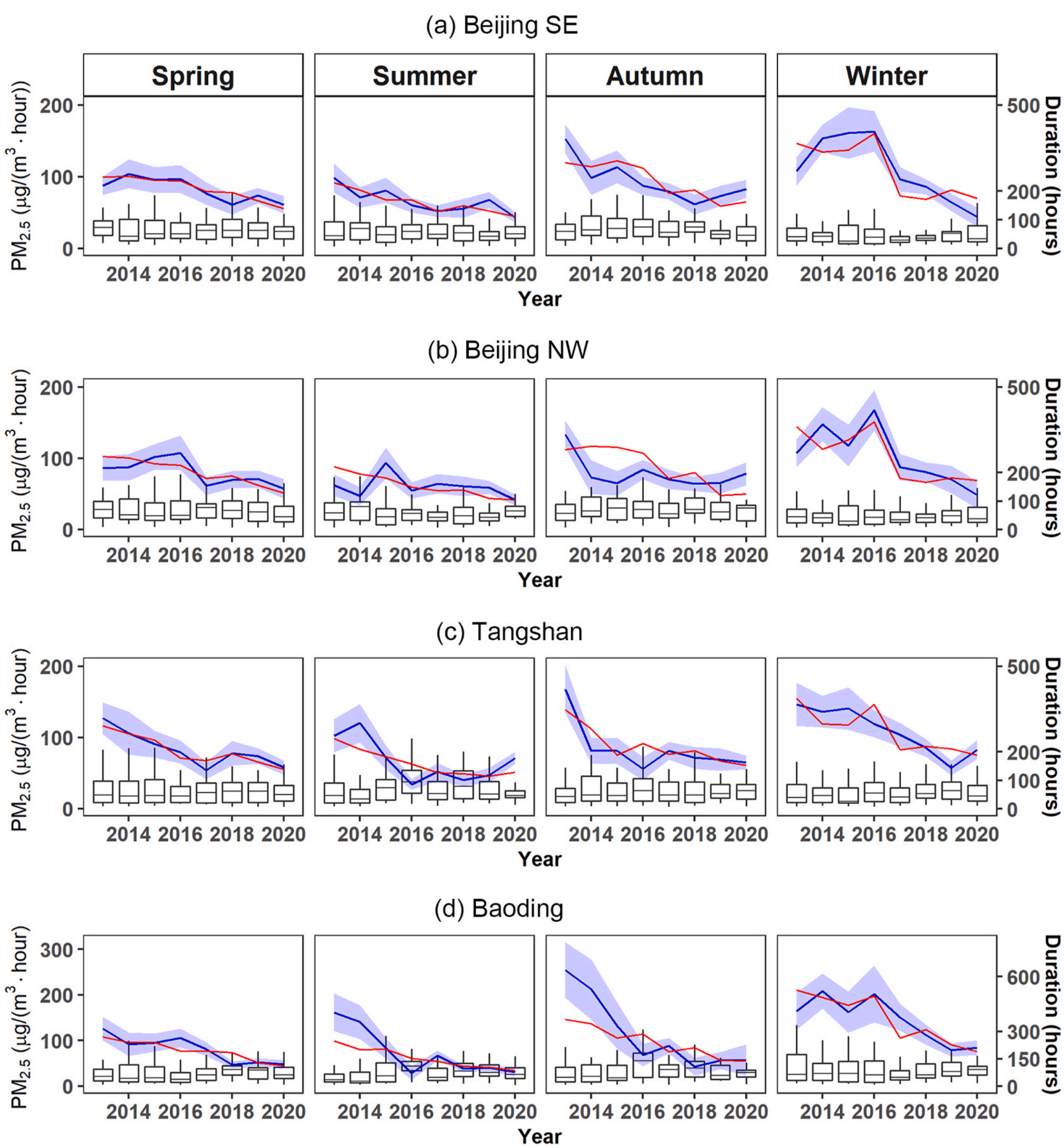


Fig. 5. Adjusted (blue) and original (red) average episodic pollution loading of $PM_{2.5}$ for the four site clusters from 2013 to 2020, where the 95% confidence bands of the adjusted averages are indicated by the colored shade. The box plots depict the distributions of the episodic duration. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

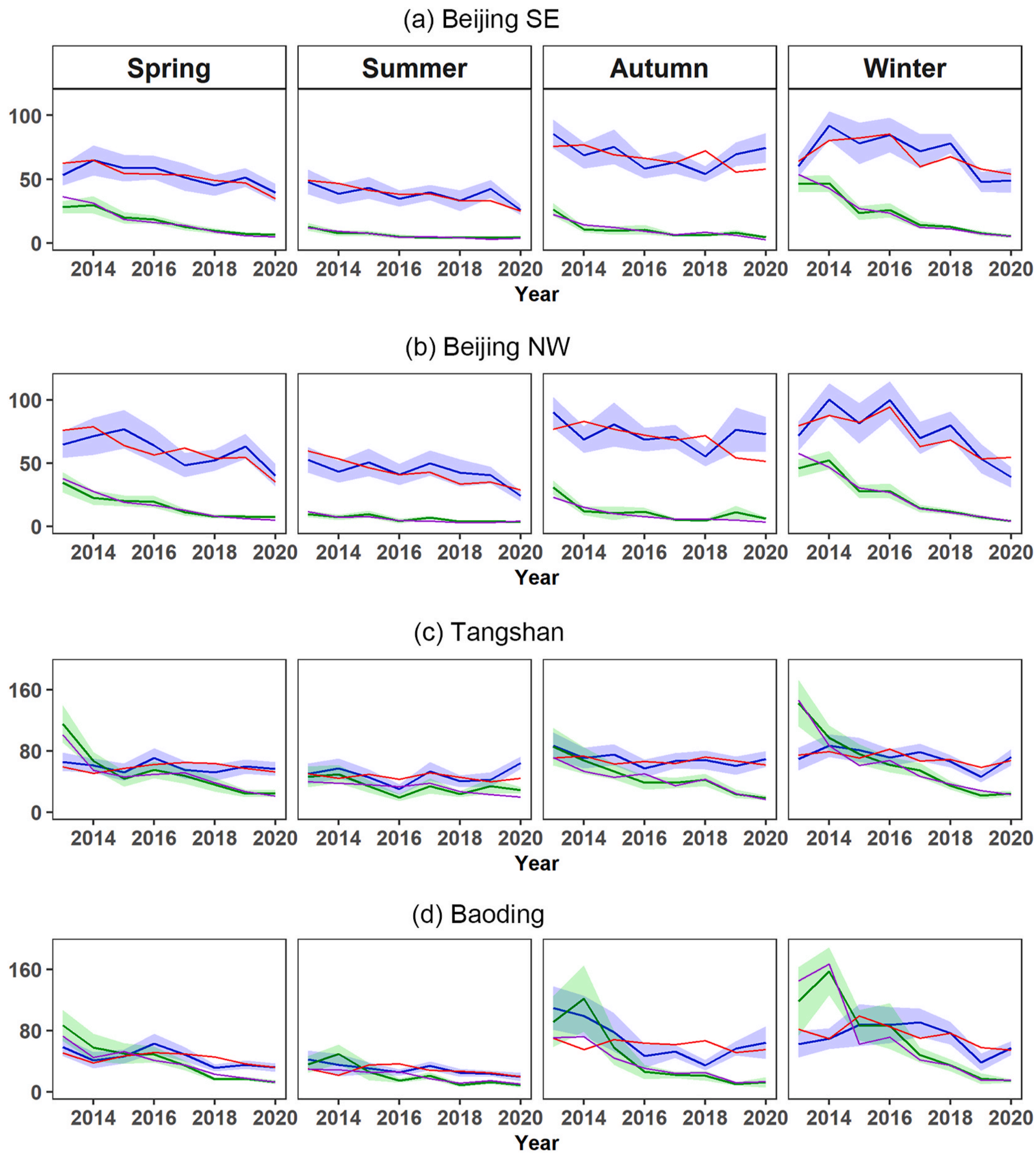


Fig. 6. Adjusted NO_2 (blue) and original NO_2 (red), and the adjusted SO_2 (green) and original SO_2 (purple) average episodic pollution loading (in $\mu g/(m^3)$ for the four site clusters from 2013 to 2020, where the 95% confidence bands of the adjusted averages are indicated by the colored shade. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

understandable as the episodic duration is largely driven by the distribution of the strong northerly cleaning processes, which is a naturally occurring phenomenon and would not easily show any trend. Figs. S12–S17 display the yearly meteorologically adjusted average total pollution loading $\widehat{PL}_i$ for the six pollutants. The results conveyed in these figures were consistent with the average episodic concentration reported in Figs. 5–7. This is understandable as the average episodic duration had been stable over the eight years from 2013 to 2018 as shown in Fig. 5.

Fig. 8 displays seasonal relative reduction in the average episodic pollution loading for $PM_{2.5}$, NO_2, SO_2 and O_3 from 2014 to 2020 relative to their respective average levels in 2013 in the four site clusters, while the SI reports the results for CO and PM_{10}. The reduction in SO_2 was the most significant and persistent with the reduction well established in almost all site clusters in every season since 2015. By 2020, the relative reductions were close to 100%. The relative reduction to the 2013 level came later with a smaller magnitude for $PM_{2.5}$, and the significant reduction in NO_2 tended to be even smaller and later than those of $PM_{2.5}$. Significant reduction over the 2013 NO_2 level in winter only happened since 2019 for the two Beijing site clusters while Tangshan and Baoding had not reached that yet in 2020. The situation for the spring NO_2 was largely the same as the winter, which saw Tangshan and Beijing only made a significant reduction in 2020, while Baoding performed the best among the four site clusters for spring, summer and autumn. Baoding's reduction amount in both $PM_{2.5}$ and NO_2 tended to be higher than the other three site clusters, which were partially due to Baoding's more severe pollution and having more room to decline.

The situation for the ozone was sharply different from those for the other three pollutants. All four site clusters had increased average episodic loading with those in Beijing SE being more persistent while the other three site clusters had more yearly variations. While this was consistent with the general assessment results using the entire data (Chen et al., 2018; Sun et al., 2021), it reveals that the ozone increase in the spring and summer had not been stopped yet, while the analysis using the whole data had shown some sign of leveling off in the ozone in spring and summer when the ozone level is high and the situation is

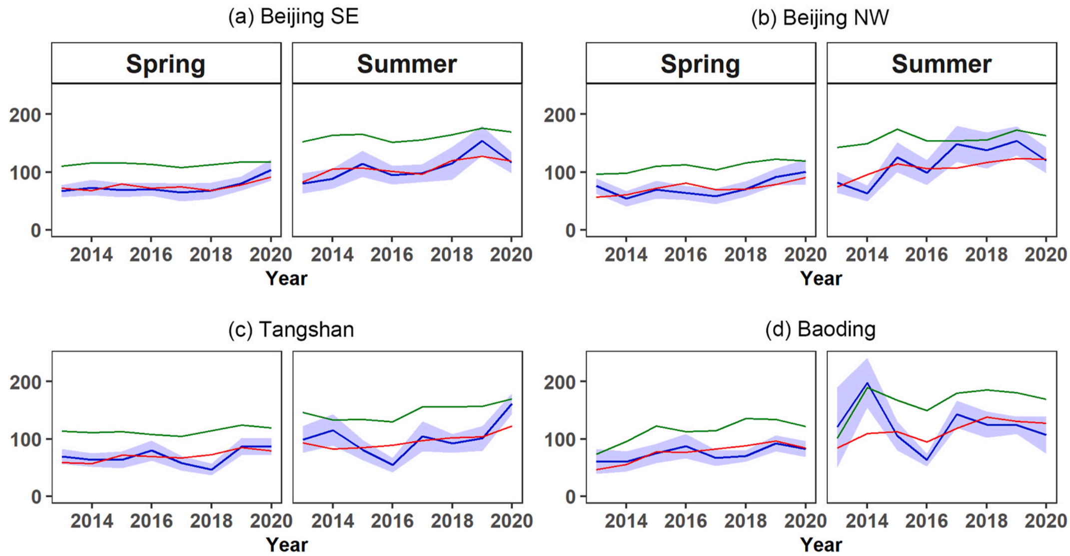


Fig. 7. Adjusted (blue), original (red) average episodic pollution loading and adjusted average concentration using all hourly data (green) of O_3 for the four site clusters from 2013 to 2020 in spring and summer, where the 95% confidence bands of the adjusted averages are indicated by the colored shade. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

more severe. It is alarming to see the increase in the episodic ozone levels in autumn and winter in the four site clusters although there were some uncertainties with the increase in autumn in Beijing NW and Tangshan. The large relative increase in winter was partly due to the low ozone level in 2013. For instance, the winter average episodic ozone was only 9.1 μg/m^3 in 2013 and was increased to 34.9 μg/m^3 in 2020, leading to the 283.8% increase. Although the absolute ozone pollution was still relatively less severe in autumn and winter, the persistent increasing trends in autumn and winter, together with the wide spread increase in the spring and summer, suggest that the underlying ozone generation pathways had not been diminished in the region, and more efforts are needed.

3.4. Comparison with using entire data

An important question is how the episode based air quality assessment compares with those based on the entire observations including those collected outside the pollution episodes in terms of the general trend and the yearly changes. Fig. S18 in the SI shows the meteorological adjusted seasonal means from 2013 to 2020 using the entire data in the four site clusters; see Liang et al. (2015) for the method and algorithm. Although the occupancy rates of the episodes (Table 2) varied from year to year, the two sets of results were largely consistent.

Fig. 9 provides the yearly relative reduction in the seasonal meteorologically adjusted average concentrations relative to the 2013 adjusted average levels using all data, which is a counter-part of Fig. 8 based on the episodic data. Comparing Figs. 8 and 9, it is found that there was a strong synergy between the two sets of the assessment measures in seasons and years among those weaker (statistically insignificant) reduction changes regardless of the signs of the reductions. This was most noticeable in the winters of 2014–2016 for $PM_{2.5}$, and for SO_2 for Baoding in the summer, autumn and winter of 2014. Table 4 reports the Spearman rank correlation coefficients between the episodic average pollution loading and the full data based average concentration and their respective reduction rates, while the corresponding Pearson correlation coefficients are given in Table S15. It shows that except for NO_2 in autumn and O_3 in summer and autumn, there were significant positive correlations for all other pollutants and seasons. For each season, the correlation with respect to the seasonal averages was much higher than the correlation on the relative reduction rates as both sets of averages were largely consistent on the general trend of the pollutants, and less so for the relative reduction rates. Among the four species, SO_2 had the highest correlations for almost all cases (except one), followed by $PM_{2.5}$ with quite significant correlations. The correlations for NO_2 and O_3 were much weaker, especially for the reduction rates in summer and autumn.

Our analysis showed that the episode based air quality assessment was more sensitive in reflecting the underlying changes in the concentration than that using the entire data observations. This was most reflected by the assessment on the ozone with the magnitude of the relative changes being much larger than those using the full data.

3.5. Discussion

Air quality assessment is a challenging task as it involves the complex systems of meteorology, emission and their interaction in the open atmosphere, and is governed by the atmospheric chemistry and physics processes. For such a challenging task, there is a need to develop multiple air quality measures to gauge its severity and changes. The proposed episode based air quality measure is one such effort designed to complement the assessment using the entire data sample from the viewpoint of pollution episodes. The episode based assessment provides a unique angle to the evolution of the air pollution in a location.

The episodic approach is based on a data-driven algorithm to select the start and end of the episodes, primarily based on the $PM_{2.5}$. While the three cities share a similar pattern of the cleaning process governed by the northerly wind due to their geographical location in the northern part of the North China Plain, different cleaning patterns in other cities can be learned and constructed from data with the help of statistical methods, leading to the episode selection rules suitable for the other cities. The average episodic pollution and the total pollution loading can be readily extended with the within episode meteorological adjustment almost readily carried over although the important variables may change from one city to another.

4. Conclusions

The episodic based air quality assessment was conducted in the three cities which revealed that the average pollution loading of four pollutants $PM_{2.5}$, CO, SO_2 and PM_{10} had shown a significant downward trend from year 2013 to year 2020, with the episodic averages of SO_2 declining the most significantly and persistently, followed by those of

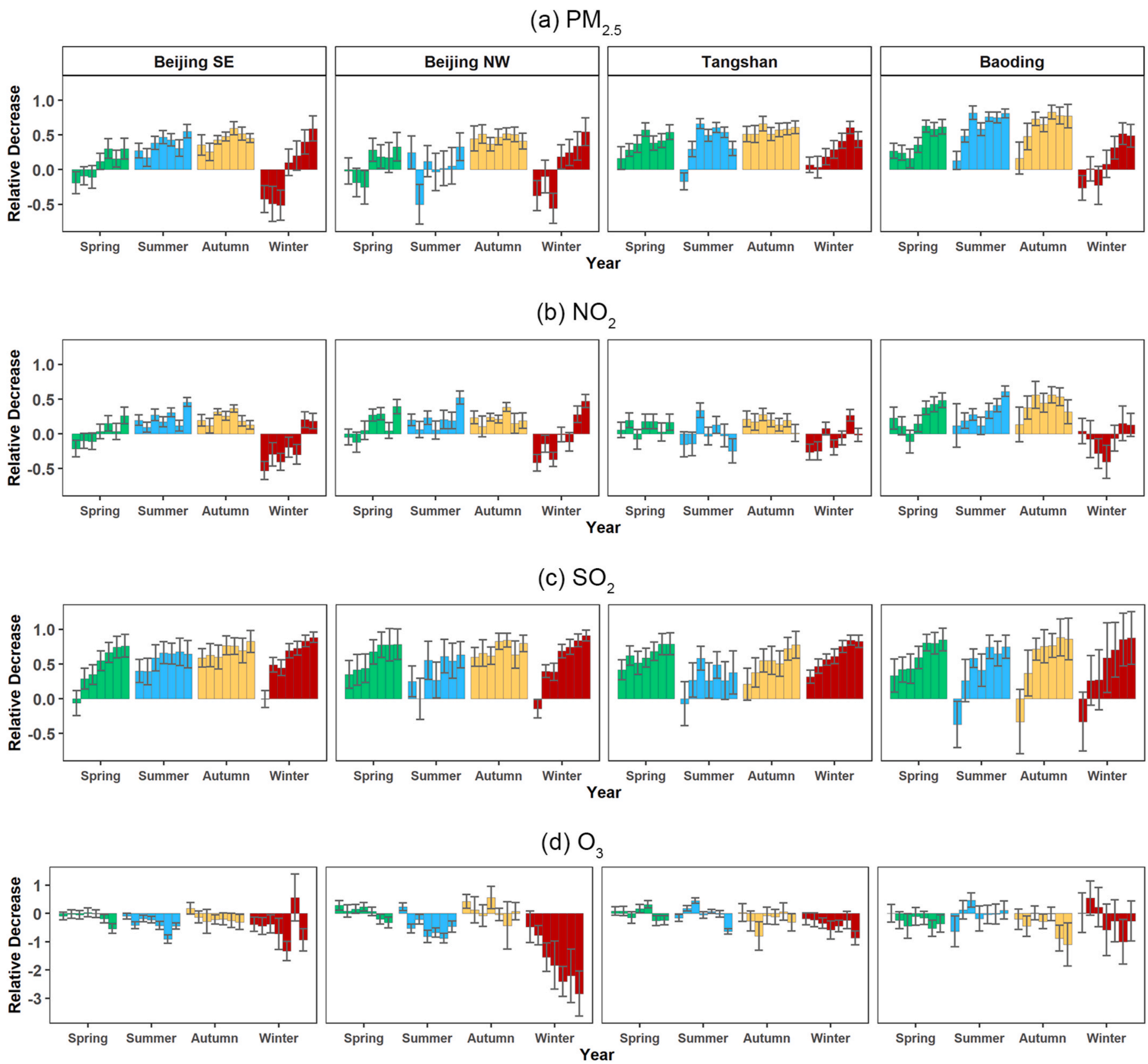


Fig. 8. The relative reduction in $PM_{2.5}$, NO_2, SO_2 and O_3 in the average episodic loading relative to those in 2013 in the four site clusters in four seasons from 2014 to 2020. The bars are the 95% confidence intervals of the relative reduction.

$PM_{2.5}$, CO and PM_{10}. Specifically, the 2020 episodic averages for SO_2 and $PM_{2.5}$ were significantly lower (at 5% significance) than the corresponding 2013 levels in all 16 season-site clusters combinations with the reduction rates ranging from 37.9% to 91.0% for SO_2 and 30.2%–81.1% for $PM_{2.5}$, while those for PM_{10} and CO were significantly lower than the corresponding 2013 levels in 15 of the total 16 season-site cluster combinations, with the significant reduction rates ranging from 17.1% to 77.9% for PM_{10} and 22.3%–89.3% for CO. The episodic averages for the NO_2 had shown a less significant decline compared to the above four species, with only 12 season-site cluster combinations in 2020 registering a significant decline relative to the 2013 levels. Among those 12 season-site clusters with significant reduction, the reduction rates in the 2020 episodic averages for NO_2 ranged from 12.7% to 61.3%, which were smaller than those of the other species.

In sharp contrast, the episodic average ozone concentration had displayed a dramatic increase in all seasons since 2014 over the 2013 levels. There were 13 season-site clusters in the three cities whose episodic average O_3 had significantly increased in 2020 as compared to 2013, with the increase rates in the 2020 episodic averages ranging from 24.7% to 63.5% in spring and summer and 26.8%–283.8% in autumn and winter. Most of the higher increase rates happened in the autumn and winter, which were due to relatively lower winter levels in 2013. The most alarming was the widespread rise in the episodic average O_3 in all seasons with no sign of subsiding despite a substantial reduction in the PM–SO_2–CO cohort in the three cities.

CRediT authorship contribution statement

Shanshan Luo: Data acquisition, Methodology, Modeling, Formal analysis, Writing – original draft, Writing – review & editing. **Yuru Zhu:** Conceptualization, Methodology, Investigation. **Song Xi Chen:** Conceptualization, Writing – original draft, Writing – review & editing,

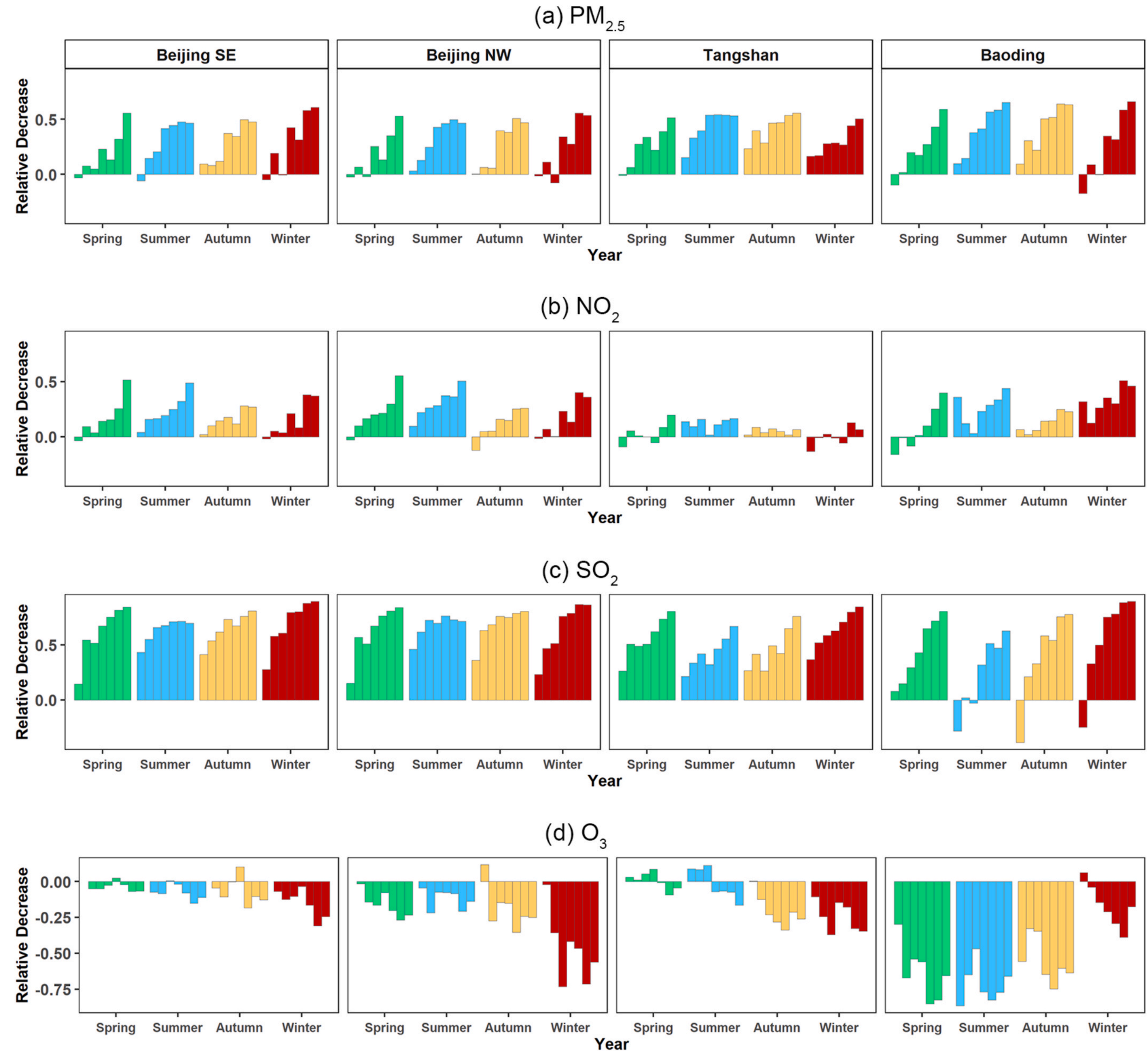


Fig. 9. The relative reduction in $PM_{2.5}$, NO_2, SO_2 and O_3 in the adjusted average concentration using all hourly data throughout the year relative to those in 2013 in the four site clusters in four seasons from 2014 to 2020.

Table 4
Pair-wise seasonal Spearman's rank correlation coefficients between the average (AVE)/reduction (RED) obtained by using all the hourly data and the average/reduction only on the episodes. The number of * indicates the level of significance in the association (*: p-value<0.05; **: p-value < 0.01; ***: p-value < 0.001).

Pollutants	Spring		Summer		Autumn		Winter	
	AVE	RED	AVE	RED	AVE	RED	AVE	RED
$PM_{2.5}$	0.83***	0.68***	0.69 ***	0.54**	0.57***	0.62***	0.85***	0.87***
NO_2	0.70***	0.46*	0.73***	0.47*	0.23	0.04	0.50**	0.63***
SO_2	0.98***	0.78***	0.94***	0.52**	0.94***	0.80***	0.98***	0.96***
O_3	0.67***	0.53**	0.67***	0.25	0.58***	0.26	0.46**	0.58**

Supervision, Funding acquisition.

Declaration of competing interest

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

Acknowledgment

The authors thank two reviewers and the Associate Editor for construct comments and suggestion which have improved the presentation of the paper. The research was partially supported by National Natural Science Foundation of China Grants 92046021, 12026607 and 12071013, Center for Statistical Science at Peking University; LMEQF at Peking University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2022.119242.

References

Bouwman, A., Lee, D., Asman, W., Dentener, F., Van Der Hoek, K., Olivier, J., 1997. A global high-resolution emission inventory for ammonia. Global Biogeochem. Cycles 11, 561–587.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Briggs, D.J., de Hoogh, C., Gulliver, J., Wills, J., Elliott, P., Kingham, S., Smallbone, K., 2000. A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. Sci. Total Environ. 253, 151–167. https://doi.org/10.1016/S0048-9697(00)00429-0. https://www.sciencedirect.com/science/article/pii/S0048969700004290.

Brunekreef, B., Holgate, S.T., 2002. Air pollution and health. Lancet 360, 1233–1242. https://doi.org/10.1016/S0140-6736(02)11274-8. https://www.sciencedirect.com/science/article/pii/S0140673602112748.

Byun, D., Schere, K.L., 2006. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. Appl. Mech. Rev. 59, 51–77.

Chang, X., Wang, S., Zhao, B., Xing, J., Liu, X., Wei, L., Song, Y., Wu, W., Cai, S., Zheng, H., et al., 2019. Contributions of inter-city and regional transport to PM2. 5 concentrations in the Beijing-Tianjin-Hebei region and its implications on regional joint air pollution control. Sci. Total Environ. 660, 1191–1200.

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 PM2.5 and O3. Atmos. Environ. 193, 290–301. http://www.sciencedirect.com/science/article/pii/S1352231018305685.

Cheng, Y., He, K.b., Du, Z.y., Zheng, M., Duan, F.k., Ma, Y.l., 2015. Humidity plays an important role in the PM2.5 pollution in Beijing. Environ. Pollut. 197, 68–75.

Duan, K., Sun, G., Zhang, Y., Yahya, K., Wang, K., Madden, J.M., Caldwell, P.V., Cohen, E.C., McNulty, S.G., 2017. Impact of air pollution induced climate change on water availability and ecosystem productivity in the conterminous United States. Climatic Change 140, 259–272.

Ghorani-Azam, A., Riahi-Zanjani, B., Balali-Mood, M., 2016. Effects of air pollution on human health and practical measures for prevention in Iran. J. Res. Med. Sci. Off. J. Isfahan Univ. Med. Sci. 21.

Ho, T.K., 1995. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition, vol. 1, pp. 278–282. https://doi.org/10.1109/ICDAR.1995.598994.

Huang, Y., Guo, B., Sun, H., Liu, H., Chen, S.X., 2021. Relative importance of meteorological variables on air quality and role of boundary layer height. Atmos. Environ. 267, 118737 https://doi.org/10.1016/j.atmosenv.2021.118737. https://www.sciencedirect.com/science/article/pii/S1352231021005598.

Kampa, M., Castanas, E., 2008. Human health effects of air pollution. Environ. Pollut. 151, 362–367. https://doi.org/10.1016/j.envpol.2007.06.012 proceedings of the 4th International Workshop on Biomonitoring of Atmospheric Pollution (With Emphasis on Trace Elements). https://www.sciencedirect.com/science/article/pii/S0269749107002849.

Kim, K.H., Jahan, S.A., Kabir, E., 2013. A review on human health perspective of air pollution with respect to allergies and asthma. Environ. Int. 59, 41–52. https://doi.org/10.1016/j.envint.2013.05.007. https://www.sciencedirect.com/science/article/pii/S0160412013000998.

Kuenen, J.J.P., Visschedijk, A.J.H., Jozwicka, M., Denier van der Gon, H.A.C., 2014. TNO-MACC_II emission inventory; a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling. Atmos. Chem. Phys. 14, 10963–10976. https://doi.org/10.5194/acp-14-10963-2014. https://acp.copernicus.org/articles/14/10963/2014/.

Lang, J., Cheng, S., Li, J., Chen, D., Zhou, Y., Wei, X., Han, L., Wang, H., et al., 2013. A monitoring and modeling study to investigate regional transport and characteristics of PM2.5 pollution. Aerosol Air Qual. Res. 13, 943–956.

Lawrence, M.G., 2005. The relationship between relative humidity and the dewpoint temperature in moist air: a simple conversion and applications. Bull. Am. Meteorol. Soc. 86, 225–234.

Li, S., Liu, R., Wang, S., Chen, S.X., 2021. Radiative effects of particular matters on ozone pollution in six North China cities. J. Geophys. Res. Atmos. 126 https://doi.org/10.1029/2021JD035963.

Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., Huang, H., Chen, S.X., 2015. Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating. Proc. Math. Phys. Eng. Sci. 471, 20150257 https://doi.org/10.1098/rspa.2015.0257. https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2015.0257.

Liu, X.H., Zhang, Y., Cheng, S.H., Xing, J., Zhang, Q., Streets, D.G., Jang, C., Wang, W.X., Hao, J.M., 2010. Understanding of regional air pollution over China using CMAQ, part I performance evaluation and seasonal variation. Atmos. Environ. 44, 2415–2426. https://doi.org/10.1016/j.atmosenv.2010.03.035. https://www.sciencedirect.com/science/article/pii/S135223101000261X.

Ohara, T., Akimoto, H., Kurokawa, J., Horii, N., Yamaji, K., Yan, X., Hayasaka, T., 2007. An Asian emission inventory of anthropogenic emission sources for the period 1980–2020. Atmos. Chem. Phys. 7, 4419–4444. https://doi.org/10.5194/acp-7-4419-2007. https://acp.copernicus.org/articles/7/4419/2007/.

Pacyna, E.G., Pacyna, J.M., Steenhuisen, F., Wilson, S., 2006. Global anthropogenic mercury emission inventory for 2000. Atmos. Environ. 40, 4048–4063. https://doi.org/10.1016/j.atmosenv.2006.03.041. https://www.sciencedirect.com/science/article/pii/S135223100600313X.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464. https://doi.org/10.1214/aos/1176344136, 10.1214/aos/1176344136.

Shi, Y., Xia, Y.f., Lu, B.h., Liu, N., Zhang, L., Li, S.j., Li, W., 2014. Emission inventory and trends of NOx for China, 2000–2020. J. Zhejiang Univ. - Sci. 15, 454–464.

Streets, D.G., Bond, T., Carmichael, G., Fernandes, S., Fu, Q., He, D., Klimont, Z., Nelson, S., Tsai, N., Wang, M.Q., et al., 2003. An inventory of gaseous and primary aerosol emissions in Asia in the year 2000. J. Geophys. Res. Atmos. 108.

Su, T., Li, J., Li, C., Lau, A.K.H., Yang, D., Shen, C., 2017. An intercomparison of AOD-converted PM2.5 concentrations using different approaches for estimating aerosol vertical distribution. Atmos. Environ. 166, 531–542. https://doi.org/10.1016/j.atmosenv.2017.07.054. http://www.sciencedirect.com/science/article/pii/S1352231017305034.

Sun, H., Luo, S., You, W., Wang, Y., Zhan, H., Wang, X., Huang, Y., He, J., Guo, B., Chen, S., 2021. Air Quality Assessment Report VIII: Regional Air Quality Assessment of the 3+95 Cities. Center for Statistics at Peking University. https://www.songxichen.com/Uploads/Files/Report/Air_Quality_Assessment_Report_VIII.pdf.

Taghizadeh-Hesary, F., Taghizadeh-Hesary, F., 2020. The impacts of air pollution on health and economy in Southeast Asia. Energies 13. https://doi.org/10.3390/en13071812. https://www.mdpi.com/1996-1073/13/7/1812.

Tai, A.P., Mickley, L.J., Jacob, D.J., 2010. Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: implications for the sensitivity of PM2.5 to climate change. Atmos. Environ. 44, 3976–3984. https://doi.org/10.1016/j.atmosenv.2010.06.060. https://www.sciencedirect.com/science/article/pii/S135223101000539X.

Tetens, O., 1930. Uber einige meteorologische begriffe. Z. Geophys. 6, 297–309.

Wang, L., Zhang, N., Liu, Z., Sun, Y., Ji, D., Wang, Y., 2014. The influence of climate factors, meteorological conditions, and boundary-layer structure on severe haze pollution in the Beijing-Tianjin-Hebei region during January 2013, 2014 Adv. Meteorol. 1–14. https://doi.org/10.1155/2014/685971.

Wang, X., Wei, W., Cheng, S., Li, J., Zhang, H., Lv, Z., 2018. Characteristics and classification of PM2. 5 pollution episodes in Beijing from 2013 to 2015. Sci. Total Environ. 612, 170–179.

Wang, Z., Maeda, T., Hayashi, M., Hsiao, L.F., Liu, K.Y., 2001. A nested air quality prediction modeling system for urban and regional scales: application for high-ozone episode in Taiwan. Water, Air, Soil Pollut. 130, 391–396. https://doi.org/10.1023/A:1013833217916.

Whiteman, C.D., Hoch, S.W., Horel, J.D., Charland, A., 2014. Relationship between particulate air pollution and meteorological variables in Utah's Salt Lake Valley. Atmos. Environ. 94, 742–753.

Zhang, Q., Streets, D.G., Carmichael, G.R., He, K.B., Huo, H., Kannari, A., Klimont, Z., Park, I.S., Reddy, S., Fu, J.S., Chen, D., Duan, L., Lei, Y., Wang, L.T., Yao, Z.L., 2009. Asian emissions in 2006 for the NASA INTEX-B mission. Atmos. Chem. Phys. 9, 5131–5153. https://doi.org/10.5194/acp-9-5131-2009. https://acp.copernicus.org/articles/9/5131/2009/.

Zhang, S., Guo, B., Dong, A., He, J., Xu, Z., Chen, S., 2017a. Cautionary tales on air-quality improvement in Beijing. Proc. Math. Phys. Eng. Sci. 473, 20170457 https://doi.org/10.1098/rspa.2017.0457.

Zhang, X., Ou, X., Yang, X., Qi, T., Nam, K.M., Zhang, D., Zhang, X., 2017b. Socioeconomic burden of air pollution in China: province-level analysis based on energy economic model. Energy Econ. 68, 478–489. https://doi.org/10.1016/j.eneco.2017.10.013. https://www.sciencedirect.com/science/article/pii/S0140988317303547.

Zheng, G.J., Duan, F.K., Su, H., Ma, Y.L., Cheng, Y., Zheng, B., Zhang, Q., Huang, T., Kimoto, T., Chang, D., Pöschl, U., Cheng, Y.F., He, K.B., 2015. Exploring the severe winter haze in Beijing: the impact of synoptic weather, regional transport and heterogeneous reactions. Atmos. Chem. Phys. 15, 2969–2983. https://doi.org/10.5194/acp-15-2969-2015. https://acp.copernicus.org/articles/15/2969/2015/.

Zhu, Y., Liang, Y., Chen, S.X., 2021. Assessing local emission for air pollution via data experiments. Atmos. Environ. 252, 118323. https://www.sciencedirect.com/science/article/pii/S1352231021001412.