Environment International 173 (2023) 107861

Contents lists available at ScienceDirect

Environment International

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

Full length article

Contributions of various driving factors to air pollution events: Interpretability analysis from Machine learning perspective

Tianshuai Li [a], Qingzhu Zhang [a], Yanbo Peng [a,b,*], Xu Guan [b], Lei Li [a], Jiangshan Mu [a], Xinfeng Wang [a], Xianwei Yin [c], Qiao Wang [a]

[a] Big Data Research Center for Ecology and Environment, Environment Research Institute, Shandong University, Qingdao 266003, PR China
[b] Shandong Academy for Environmental Planning, Jinan 250101, PR China
[c] Zibo Ecological Environment Monitoring Center of Shandong Province, Zibo 255040, PR China

ARTICLE INFO

Handling Editor: Xavier Querol

Keywords:
$PM_{2.5}$
Air pollution
Machine learning
Permutation importance
PDP
SHAP

ABSTRACT

The air quality in China has been improved substantially, however fine particulate matter ($PM_{2.5}$) still remain at a high level in many areas. $PM_{2.5}$ pollution is a complex process that is attributed to gaseous precursors, chemical, and meteorological factors. Quantifying the contribution of each variable to air pollution can facilitate the formulation of effective policies to precisely eliminate air pollution. In this study, we first used decision plot to map out the decision process of the Random Forest (RF) model for a single hourly data set and constructed a framework for analyzing the causes of air pollution using multiple interpretable methods. Permutation importance was used to qualitatively analyze the effect of each variable on $PM_{2.5}$ concentrations. The sensitivity of secondary inorganic aerosols (SIA): SO_4^{2-}, NO_3^- and NH_4^+ to $PM_{2.5}$ was verified by Partial dependence plot (PDP). Shapley Additive Explanation (Shapley) was used to quantify the contribution of drivers behind the ten air pollution events. The RF model can accurately predict $PM_{2.5}$ concentrations, with determination coefficient (R^2) of 0.94, root mean square error (RMSE) and mean absolute error (MAE) of 9.4 μg/m^3 and 5.7 μg/m^3, respectively. This study revealed that the order of sensitivity of SIA to $PM_{2.5}$ was $NH_4^+ > NO_3^- > SO_4^{2-}$. Fossil fuel and biomass combustion may be contributing factors to air pollution events in Zibo in 2021 autumn–winter. NH_4^+ contributed 19.9–65.4 μg/m^3 among ten air pollution events (APs). K, NO_3^-, EC and OC were the other main drivers, contributing 8.7 ± 2.7 μg/m^3, 6.8 ± 7.5 μg/m^3, 3.6 ± 5.8 μg/m^3 and 2.5 ± 2.0 μg/m^3, respectively. Lower temperature and higher humidity were vital factors that promoted the formation of NO_3^-. Our study may provide a methodological framework for precise air pollution management.

1. Introduction

In the past few years, many cities have suffered severe air pollution (Andersson et al., 2015; Peng et al., 2016) in China, which seriously affected human health and life (Bourdrel et al., 2017; Lee et al., 2015; Xing et al., 2016). Therefore, the Chinese government implemented the Air Pollution Prevention and Control Action Plan in 2013, and the Three-year Action Plan for Winning the Blue Sky Defense War in 2018 to solve the air pollution problems, which led to the significant improvement of air quality (Ma et al., 2019; Zhang et al., 2019). However, air pollution problems related to high concentrations of fine particulate matter ($PM_{2.5}$) still occur in many areas. Therefore, it is essential to analyze the drivers affecting $PM_{2.5}$ formation during air pollution periods.

In previous studies, chemical transport models (CTMs) have been widely used to study $PM_{2.5}$ pollution by scenario simulation and pollution process analysis (Chen et al., 2019; Weagle et al., 2018; Zheng et al., 2015). However, CTMs are often subject to large biases due to uncertainties in emission inventories, and physical and chemical parameters (Hu et al., 2017; Lam et al., 2021). Recently, machine learning models have been widely used to study air pollution because of their excellent performance (Geng et al., 2021; Wang et al., 2020; Wei et al., 2019). For example, machine learning models generally outperform CTMs and traditional statistical analysis (e.g., linear models) in predicting $PM_{2.5}$ (Vu et al., 2019; Yang et al., 2021). However, the black-box nature and complex system mechanisms of many machine

* Corresponding author at: Big Data Research Center for Ecology and Environment, Environment Research Institute, Shandong University, Qingdao 266003, PR China.
E-mail address: pengyanbo@mail.sdu.edu.cn (Y. Peng).

https://doi.org/10.1016/j.envint.2023.107861
Received 31 December 2022; Received in revised form 9 February 2023; Accepted 1 March 2023
Available online 4 March 2023

learning models make the results hard to interpret. With the development of technology, more and more interpretable tools are becoming available. The Partial dependence plot (PDP) method was proposed and extended (Friedman, 2001; Goldstein et al., 2015) to test machine learning models, and was also applied to the atmospheric science field (Wei et al., 2022; Yang et al., 2022; Zhang et al., 2022). To overcome the barriers of black box models, a feature attribution technique – Shapley Additive Explanation (Shapley) (Lundberg and Lee, 2017) has been proposed to quantify the overall and local impact of variables on the results, which has been applied in the medical field (Alfi et al., 2022; Lewin-Epstein et al., 2021), but rarely in the atmospheric field (Hou et al., 2022). These interpretable tools can assist machine learning models to explain the causes of air pollution well.

Zibo, a heavily polluted industrial city, is often listed as one of the top ten most polluted cities in China. Many ceramic, heavy and light industries have led to serious air pollution in Zibo. As the capital city of China, Beijing has received widespread attention in recent years due to frequent and serious haze events (Lu et al., 2019; Wang et al., 2020; Wang et al., 2021; Xu et al., 2019). Zibo is located in the southeast of Beijing, upwind of Beijing. Considering the regional transport, Beijing will be affected by pollutants from Zibo. In addition, Zibo will also be affected by the surrounding urban environment. Therefore, Zibo is a donor and a receptor of regional pollutants transport. Therefore, quantifying the contributions of different variables to air pollution in Zibo can provide pollution control ideas for Zibo itself, Shandong Province and even Beijing.

In this study, we analyzed four-month online data measured by the Zibo Atmospheric Environment Super Monitoring Station in 2021 autumn–winter and quantified the contributions of precursor emissions, meteorological conditions, and chemical and elemental composition to air pollution events. Firstly, we conducted a qualitative analysis of various drivers associated with $PM_{2.5}$ using Permutation importance. On this basis, the sensitivity of the three main secondary inorganic aerosols (SIA), SO_4^{2-}, NO_3^- and NH_4^+, to $PM_{2.5}$ was analyzed using the PDP method. The $PM_{2.5}$ concentrations higher than 75 $\mu g/m^3$ were divided into ten air pollution events (APs), and the SHAP method was used to quantify the contributions of various variables at each AP stage. Finally, we explored the formation mechanism of NO_3^-. The use of multiple interpretable methods may provide a set of ideas for analyzing the causes of air pollution.

2. Materials and methods

2.1. Study area and data source

The study area was in Zibo, Shandong Province, which is one of the most polluted industrial cities in China. All data (hourly resolution) from September 1 to December 22 in 2021 were measured by the Atmospheric Environment Super Monitoring Station (Fig. 1, 36.86°N, 118.14°E). The dataset consists of air pollutants ($PM_{2.5}$, carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3)), meteorological data (wind speed (WS), wind direction (WD), temperature (T), pressure (P), relative humidity (RH)), ion data (Cl^-, NH_4^+, NO_3^-, etc.), carbon data (organic carbon (OC), element carbon (EC)) and element data (Na, Fe, Ni, K, etc.).

$PM_{2.5}$, NO_2, SO_2, CO and O_3 were measured by 5014i Beta Continuous Ambient Particulate Monitor, Model 42i NO-NO_2-NO_x Analyzer, Model 43i SO_2 Analyzer, Model 48i CO Analyzer and Model 49i O_3 Analyzer, respectively. Ion species were measured by in-situ gas and aerosol compositions monitor (IGAC). An atmospheric heavy metal analyzer (ZSTK-HMCA-3200) and an atmospheric carbon composition analyzer (sunset OC_EC model4) were used to measure metallic elements (Na, Fe, Ni, K, etc.) and carbon data (OC, EC), respectively. Meteorological data (T, RH, P, WS, WD) was measured by a miniature portable weather station (WXT520).

2.2. Machine learning method

In this study, five types of experimentally measured data were entered into the Random Forest (RF) model as input. The results of the model were analyzed for interpretability using Permutation importance, PDP and SHAP.

2.2.1. Random Forest

RF model is a representative bagging integration algorithm consisting of decision trees as base evaluators, which was proposed (Breiman, 2001) in 2001 to solve classification and regression problems. The central idea of the RF model is to construct multiple independent

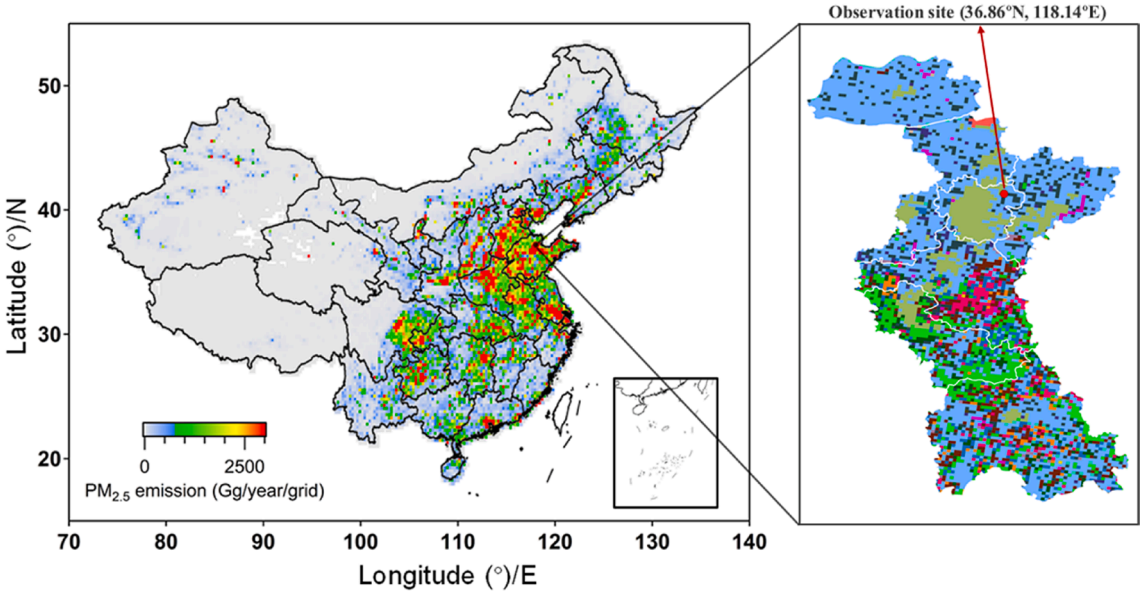

Fig. 1. The black dot on the left panel showed the location of Zibo, a city with high industrial pollution, and observation site was shown on the right panel. The left panel was colored by $PM_{2.5}$ emission inventory (Li et al., 2017b; Zheng et al., 2018) in 2017 while land-use types were used to color the right panel, i.e. light blue for dry land, and dark green for urban land, etc. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

evaluators and then average or majority vote on their predictions to determine the outcome of the integrated evaluator results. Compared with other machine learning models, RF model is fast to train, highly accurate and easy to implement, and have been widely used in other studies (Chen et al., 2018; Rahman et al., 2021; Wang et al., 2022a; Wang et al., 2022b). In order to train a model in line with the content of the study, we ordered the model to divide the data into a training set and a testing set in a 7:3 ratio, and a more specific explanation can be found in Text S1. Text S1-S2, Figure S1-S2 and Table S1 detailed the model training and hyper-parameters selecting processes.

2.2.2. Model evaluation

Three evaluation metrics were employed to evaluate the regression performance of the RF model: determination coefficient (R^2), mean absolute error (MAE) and root mean square error (RMSE).

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \quad (1)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \widehat{y}_i| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2} \quad (3)$$

where N represents the total number of data points. i is the i_{th} data-point. y_i is the observed value (μg/m^3) of the $PM_{2.5}$. $\widehat{y}_i$ is the predicted value (μg/m^3). $\overline{y}$ is the average of the observed value (μg/m^3).

2.2.3. Model interpretation

(1) Permutation importance method.

Permutation importance was performed by the Eli5 toolkit to calculate the average importance of each feature to $PM_{2.5}$ after the column data was disrupted. More information about Permutation importance and Eli5 toolkit were listed in Text S3. The theory calculation formula is as follows:$i_j = s - \frac{\sum_{k=1}^{k} s_{k,j}}{k}$(4).

The feature j was rearranged and then repeated k times to construct the polluted dataset $Dc_{k,j}$.i_j is the weight of feature j. s is the performance score of the RF model on test dataset D and k is the number of iterations. j represents each feature of the test dataset and $s_{k,j}$ is the performance score of the RF model on test dataset $Dc_{k,j}$.

(2) Partial dependence plot.

PDP shows the dependence of one or two features on the predicted results of the machine learning models (Friedman, 2001), which is defined as:

$$\widehat{f}_{x_S}(x_S) = Ex_C[\widehat{f}(x_S, x_C)] = \int \widehat{f}(x_S, x_C)dP(x_C) \quad (5)$$

X_S represents a set containing only one or two features (in our study, S refers to two features). The features in set S are those that this study focuses on to reveal their impact on the prediction results. X_C is the set of other features. X_S and X_C constitute the total features and are used as input features for the model. $\widehat{f}$ is the Random Forest model.

(3) Shapley Additive Explanation method.

SHAP (Lundberg et al., 2020; Lundberg et al., 2018) is an approach based on the coalitional game theory which was used to allocate total revenue to game players (Shapley, 1953). Briefly, each feature variable in the dataset can be treated as a player. The predictions obtained by training the model using this dataset can be considered as the benefit of the collaboration of many participants to complete the project. Shapley value distributes the benefits of cooperation fairly by considering the contributions made by each player, that is, the relationship between each characteristic variable and each predicted value can be calculated by shapley value. It has the following functional expressions (eq (6)).

$$f(x_i) = \phi_0(f, x) + \sum_{j=1}^{N}\phi_j(f, x_i) \quad (6)$$

where $f(x_i)$ is the predicted value generated for each sample (x_i) with N features. $\phi_0(f, x)$ is the base value representing the expected value of RF model output over the dataset. $\phi_j(f, x_i)$ is the Shapley value of the impact of the feature j in the sample (x_i) on the predicted outcome of the sample.

Further, $\phi_j(f, x_i)$ represents the Shapley value of each feature in each sample, which is a weighted average over all possible combinations of variable subsets (Lundberg and Lee, 2017).

$$\phi_j(f, x) = \sum_{s \subseteq \{x_1, x_2, \cdots x_n\} \setminus \{x_j\}} \frac{|S|(N - |S| - 1)!}{N!}\left(f_x(S \cup \{x_j\}) - f_x(S)\right) \quad (7)$$

where $\phi_j(f, x)$ is the Shapely value of j feature. S is a subset of the features used in the model. $x_1, x_2 \cdots x_n$ are the features. $|S|$ is the number of non-zero entries in S. $f_x(S)$ is the predicted value of subset S.

3. Results and discussion

3.1. Characteristics of the detected data

Based on the Air Quality Index (AQI) and previous studies (Li et al., 2017a; Zheng et al., 2015b), an hourly concentration threshold of 75 μg/m^3 was used to distinguish polluted hours from clean hours: clean hours (C, $PM_{2.5}$ < 75 μg/m^3), polluted hours (P, 75 ≤ $PM_{2.5}$ ≤ 250 μg/m^3), and highly polluted hours (H, $PM_{2.5}$ > 250 μg/m^3). According to this criterion, ten APs (air pollution) including AP1 (10:00, September 16), AP2 (9, October) AP3 (14, October), AP4 (21–31, October), AP5 (3–5, November), AP6 (13–20, November), AP7 (24–29, November), AP8 (4–7, December), AP9 (14–16, December) and AP10 (20–22, December) were considered as air pollution periods (shaded parts of Fig. 2) in our study.

The hourly measurements of $PM_{2.5}$, gaseous pollutant concentrations, and meteorological conditions during the observation period (September 1 to December 22, 2021) were presented in Fig. 2. The hourly average concentration of $PM_{2.5}$ in the ten air pollution events was 111.9 ± 30.6 μg/m^3, which was much higher than the ambient air quality standard in China (35 μg/m^3, GB 3095–2012).

The hourly average concentrations of three gas pollutants, CO, SO_2 and NO_2, were 0.9 ± 0.4 μg/m^3, 13.1 ± 8.0 μg/m^3 and 39.1 ± 20.8 μg/m^3, respectively. The three gaseous pollutants had a similar trend, which was easy to understand considering that they were the primary emission from coal combustion, factory exhaust emissions, etc. However, the O_3 (hourly average concentration 55.3 ± 43.3 μg/m^3) showed an opposite trend to that of NO_2, which was due to the existence of a photochemical cycle between NO_x and O_3 according to the previous studies (Crutzen, 1979). The hourly average temperature during the data collection period was 13.4 ± 8.3℃ and the relative humidity was 72.1 ± 23.0%.

NO_3^-, NH_4^+, SO_4^{2-}, OC and EC were the main components of $PM_{2.5}$ (Table 1), with hourly average concentrations of 13.1 ± 13.8 μg/m^3, 9.4 ± 6.2 μg/m^3, 7.3 ± 5.6 μg/m^3, 4.5 ± 2.9 μg/m^3 and 1.6 ± 1.1 μg/m^3, accounting for 27.6%, 19.6% 15.3%, 9.5% and 3.3%, respectively. Compared with the data collected in September, the hourly average concentrations of ionic species and carbon species both showed a gradient increase in the latter three months, especially in December. This is due to the fact that Zibo is located in the north of China, where the temperature of winter is low and temperature inversions are frequent, and gaseous pollutants emitted from coal-fired heating and vehicle exhaust are not easily diffused, resulting in lower air quality in winter. In addition, along the C-P-H order, the specie concentrations of components such as carbonaceous species and SIA increased significantly. During the polluted hours, the hourly average concentrations of

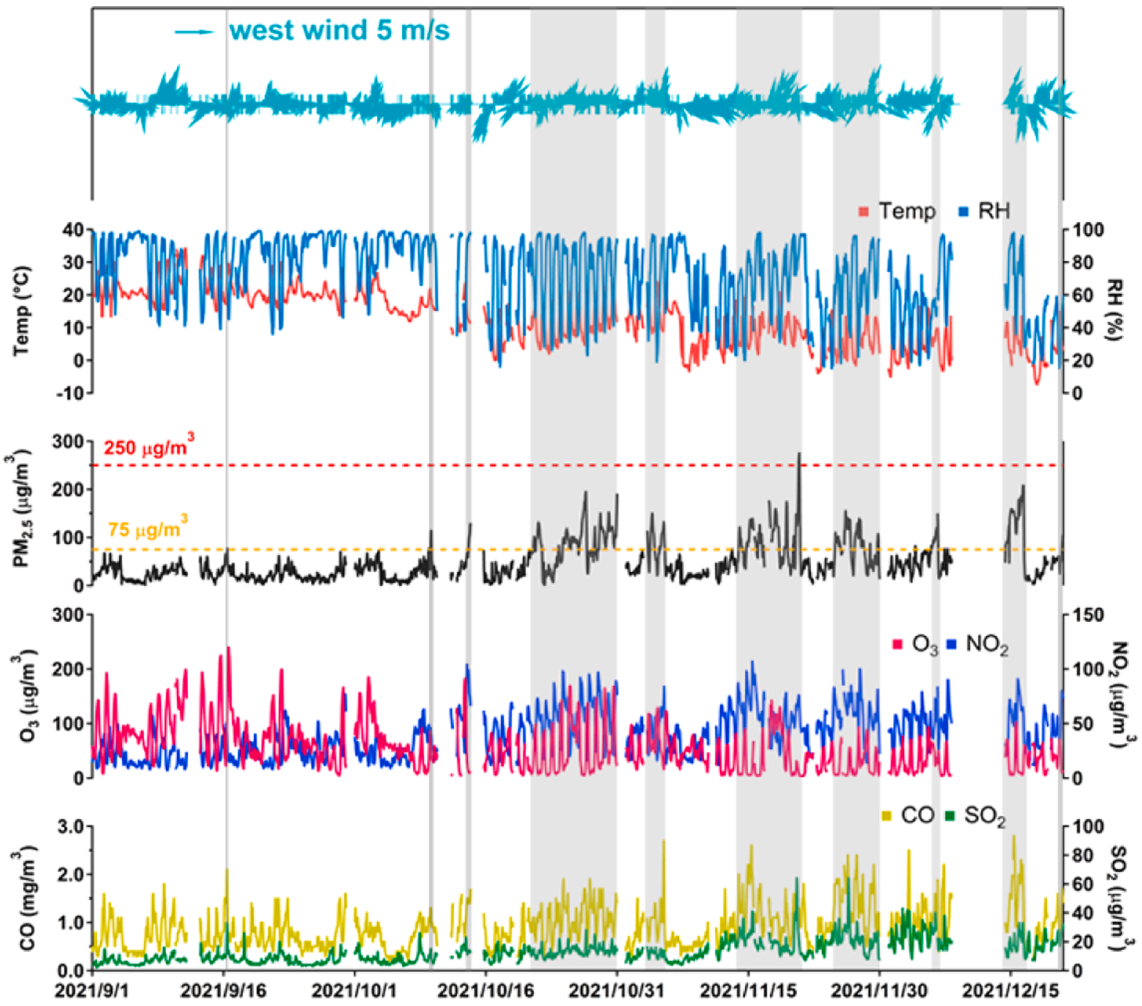

Fig. 2. Time series of hourly data of gaseous pollutants ($PM_{2.5}$, O_3, NO_2, CO, SO_2) and meteorological parameters (WS, WD, RH, T) were measured at Atmospheric Environment Super Monitoring Station from September 1 to December 22 in 2021.The orange and red dashed lines indicated concentration of 75 and 250 μg/m^3, respectively. The gray shaded area represented the ten air pollution events that occurred during the observation period. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
The hourly average concentrations of ionic species and carbon species during the observation period (September 1 to December 22, 2021).

Species	Concentrations (μg/m^3)			
	Sep.	Oct.	Nov.	Dec.
OC	3.3 ± 1.9	4.3 ± 2.4	5.3 ± 3.2	5.8 ± 3.5
EC	1.0 ± 0.7	1.5 ± 1.1	1.8 ± 1.2	2.1 ± 1.3
SO_4^{2-}	5.9 ± 3.9	7.0 ± 4.8	9.1 ± 7.0	7.1 ± 5.9
NO_3^-	5.4 ± 5.2	16.9 ± 15.6	15.6 ± 13.7	16.8 ± 15.6
NH_4^+	7.1 ± 3.2	9.4 ± 5.5	10.7 ± 7.0	11.4 ± 8.4
Cl^-	1.2 ± 1.7	1.9 ± 2.0	2.7 ± 3.2	2.0 ± 1.7
K^+	0.2 ± 0.2	0.4 ± 0.2	0.5 ± 0.9	0.4 ± 0.3
Na^+	0.2 ± 0.2	0.3 ± 0.2	0.3 ± 0.5	0.3 ± 0.4
Mg^{2+}	0.1 ± 0.1	0.1 ± 0.1	0.2 ± 0.8	0.1 ± 0.1
NO_2^-	0.3 ± 0.2	0.3 ± 0.2	0.6 ± 1.9	0.5 ± 0.4
Ca^{2+}	0.4 ± 0.2	0.6 ± 0.4	0.8 ± 0.8	0.7 ± 0.5
F^-	0.1	0.1	0.4	0.7

NO_3^-, NH_4^+, SO_4^{2-}, OC and EC increased to 33.5 μg/m^3, 18.0 μg/m^3, 12.6 μg/m^3, 8.2 μg/m^3, and 3.0 μg/m^3, respectively, which were 4.2, 2.5, 2.1, 2.3, and 2.5 times higher than the those measured during the clean period. In the heavily polluted period, the hourly average concentrations of SO_4^{2-}, NO_3^-, and NH_4^+ were 10.1, 7.8, and 5.3 times higher than those in the clean period, respectively, indicating that these substances may be the main cause of $PM_{2.5}$ pollution.

The concentrations of elements in the Zibo region during the four months were presented in Table 2. As can be seen, the total monthly average concentrations of the 11 elements were 88.9, 161.5, 220.9 and 229.2 ng/m^3, accounting for 4.1%, 3.5%, 4.0% and 4.1% of the total $PM_{2.5}$ mass concentration, respectively. The total average concentration of crustal elements (Al, K, Ca and Fe) was 413.2 ng/m^3, accounting for 89.8% of total element mass while trace elements (Cr, Co, Cu, Mn, Ni, V

Table 2
The hourly average concentrations of elemental composition of $PM_{2.5}$ (September 1 to December 22, 2021).

Species	Concentrations (ng/m^3)			
	Sep.	Oct.	Nov.	Dec.
Al	59.4 ± 118.5	116.1 ± 171.9	200.7 ± 272.2	188.3 ± 205.3
K	230.4 ± 193.2	442.8 ± 319.7	593.1 ± 362.7	655.1 ± 368.2
Ca	254.6 ± 235.3	562.9 ± 557.1	790.3 ± 684.5	822.4 ± 507.9
Cr	6.6 ± 5.9	6.6 ± 6.1	6.3 ± 5.4	7.6 ± 9.1
Mn	18.2 ± 13.5	29.1 ± 20.5	37.7 ± 25.3	40.1 ± 24.5
Fe	307.5 ± 187.4	480.3 ± 291.8	598.6 ± 347.5	629.0 ± 314.5
Cu	1.9 ± 3.8	3.1 ± 4.3	3.7 ± 5.3	3.8 ± 5.8
Zn	93.3 ± 117.7	128.9 ± 115.6	193.0 ± 198.0	168.2 ± 140.8
Co	1.0 ± 0.7	1.0 ± 0.8	1.1 ± 0.8	1.1 ± 1.0
Ni	1.3 ± 1.5	1.4 ± 1.5	1.6 ± 1.5	1.6 ± 3.4
V	3.8 ± 2.9	3.9 ± 3.4	4.0 ± 3.3	4.0 ± 3.3

and Zn) accounted for only 10.2%. In winter, anthropogenic emissions sources such as biomass burning, coal combustion, and other human activities increased (Tian et al., 2015). Additionally, the mixed layer height was low and relatively stable in winter, which inhibited both the vertical and horizontal diffusion of the air pollutants, resulting in pollution accumulation. All these reasons led to the highest concentrations of trace elements occurring in December.

3.2. Evaluation of model output results

A total of 1582 and 678 hourly data were used for the RF model training and testing, respectively (Text S1). The scatter plot (Figure S3) showed the measured values of the $PM_{2.5}$ concentrations and the predicted values based on the RF model for the training set (Figure S3a) and testing set (Figure S3b). The predicted and measured values were in good agreement on the training set, with R^2 of 0.99, RMSE and MAE of 3.4 μg/m^3 and 2.1 μg/m^3, respectively. A fairly satisfactory accuracy on the testing set was achieved, with R^2 of 0.94, RMSE and MAE of 9.4 μg/m^3 and 5.7 μg/m^3, respectively.

Permutation importance and feature importance built into the RF model were used to assess the effect of variables on $PM_{2.5}$ concentrations during the observation period. As shown in Figure S4, in terms of permutation importance, NH_4^+ had the greatest influence on the model prediction results with a weight of 0.28, followed by NO_3^- and K, together with the following EC, OC, Al, etc. The results showed that the secondary inorganic aerosols and heavy metals may be the main factors causing high concentrations of $PM_{2.5}$, which was generally consistent with the previous studies (Gao et al., 2018; Ming et al., 2017).

3.3. Sensitivity of SIA to $PM_{2.5}$ during the observation period

In our study, SIA accounted for up to 62.5% of the total $PM_{2.5}$ mass concentrations during the whole observation. Therefore, it is of great significance to explore the highly nonlinear relationship between SIA and $PM_{2.5}$. In this work, RF-PDP method was used to evaluate the synergistic effects of SIA on $PM_{2.5}$ by sensitivity analysis of two driving factors. The results of sensitivity analysis (NO_3^--NH_4^+, SO_4^{2-}-NO_3^-, NH_4^+-SO_4^{2-}) by RF-PDP (Fig. 3) reflected the average trend of the synergistic effects of two driving factors on $PM_{2.5}$.

According to the RF-PDP (Fig. 3a) results, the increase of NH_4^+ concentration had a stronger effect on $PM_{2.5}$ when NO_3^- was below about 33.5 μg/m^3, while the phenomenon was more significant when NO_3^- was above about 33.5 μg/m^3. The synergistic effect of NO_3^- and NH_4^+ can contribute to $PM_{2.5}$ (15–60 μg/m^3) when their concentrations were higher than about 33.5 μg/m^3 and 17.9 μg/m^3, respectively. The results showed that NH_4^+ had a greater impact on the formation of $PM_{2.5}$. When the NO_3^- concentration was lower than about 8.5 μg/m^3, the $PM_{2.5}$ was determined by SO_4^{2-}, and with the increase of the SO_4^{2-} concentration, the $PM_{2.5}$ concentration changed from 42.5 to 47.5 μg/m^3 (Fig. 3b). However, when the NO_3^- concentration was higher than 8.5 μg/m^3, the concentration of $PM_{2.5}$ was completely determined by NO_3^-. Overall, the sensitivity of $PM_{2.5}$ to NO_3^- was higher than SO_4^{2-}. Several studies have confirmed that NO_3^- concentrations have generally surpassed SO_4^{2-} in the North China Plain (NCP) (Wang et al., 2019; Xu et al., 2019). According to Fig. 3c, the concentration of $PM_{2.5}$ was basically determined by NH_4^+ while SO_4^{2-} had little effect on the concentration of $PM_{2.5}$.

The results indicated that the formation of $PM_{2.5}$ was governed by the synergistic effect of SIA during the observation periods. Gaseous precursors in the atmosphere transform from gas to particles through atmospheric chemical reactions, thus generating secondary particles such as sulfate and nitrate, which constitute chemical components of $PM_{2.5}$ (Zhang et al., 2012). In addition, according to the RF-PDP results, the sensitivity order of $PM_{2.5}$ to SIA was $NH_4^+ > NO_3^- > SO_4^{2-}$ and the conclusion was consistent with the result of permutation importance in Section 3.2 and previous study (Twigg et al., 2015).

3.4. Contributions of various variables in ten pollution events

The contributions of each variable in AP1-3 and AP4-10 were shown in Figure S5 and Fig. 4, respectively. Among the ten APs, the average contribution of NH_4^+ ranked first, ranging from 19.9 to 65.4 μg/m^3. In contrast, the average contribution of NO_3^- (0.4–12.8 μg/m^3) ranked in the top 5 except the 10th air pollution event. The results suggested that NH_4^+ and NO_3^- were the prominent factors causing $PM_{2.5}$ pollution. In addition, K, EC and OC were the other main drivers with average contributions of 8.7 ± 2.7 μg/m^3, 3.6 ± 5.8 μg/m^3 and 2.5 ± 2.0 μg/m^3, respectively.

NH_4^+ is formed by the reaction of ammonia (NH_3) in the atmosphere with acidic compounds such as sulfuric acid and nitric acid (Griffith et al., 2015). It was the most important factor leading to air pollution events with an average contribution of 37.2 ± 23.5 μg/m^3. ^{15}N-Stable Isotope was used to confirm that severe haze was mainly caused by fossil fuel combustion-related emissions in the Beijing area (Pan et al., 2016a). Zibo is a famous industrial city in Shandong Province, with a GDP of 367.4 billion in 2020. Among them, the total output value of the primary industry (agriculture, forestry, animal husbandry and fishery) was 15.7 billion, accounting for only 4.3% of the GDP. However, the total output value of the secondary industry (manufacturing, electricity, heat, gas, etc.) was 177.7 billion, accounting for 48.5% of the GDP, which was the highest percentage among all cities in Shandong Province. The growth rate of industrial investment was 13.8%, and the export of petrochemical products was 892.8 million. Energy consumption was 32,277,100

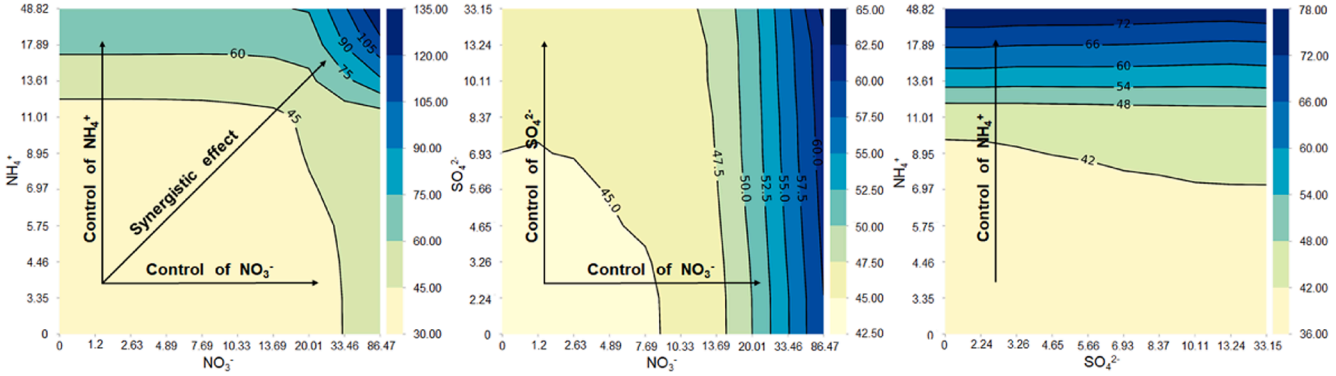

Fig. 3. Synergistic effects of SIA NH_4^+(, NO_3^-, SO_4^{2-}) on $PM_{2.5}$ concentration during the whole online observation period (from September 1 to December 22, 2021). The horizontal and vertical coordinates indicated the concentrations of NO_3^-, SO_4^{2-}, and NH_4^+ (μg/m^3) and color bar indicated the concentration of $PM_{2.5}$. (a) NO_3^--NH_4^+; (b)NO_3^--SO_4^{2-}; (c) SO_4^{2-}-NH_4^+.

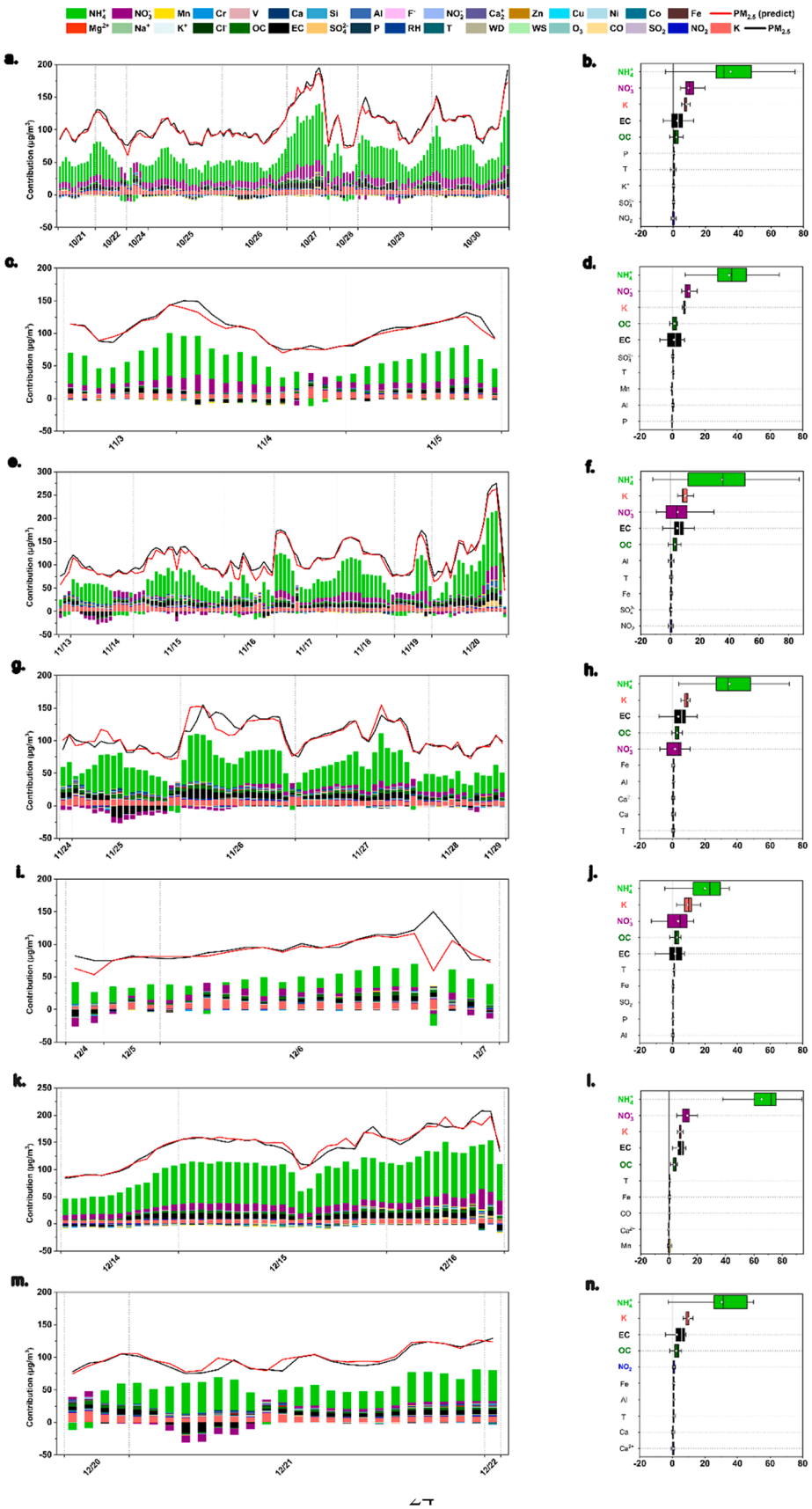

Fig. 4. Shapley values were calculated using Shapley Additive Explanation algorithm based on the Random Forest model. Time series of all variables' Shapley values during the seven APs (Figure S5 for AP1-3) were shown in picture a, c, e, g, i, k, and m, and the mean absolute value of each variable's Shapley values (| Shapley | values) were shown in the right box plots b, d, f, h, j, l, and n (the top 10 variables). Both ends are error bars, and the inner box are the lower quartile, mean values (circular), median values (Weagle et al.), and upper quartile from left to right, respectively.

tons of standard coal, of which coal fuel and fuel oil accounted for 78.5% (Shandong Statistical Yearbook, 2021). Therefore, the ten air pollution events were mainly caused by fossil fuel combustion emissions rather than agriculture. K is often considered as an indicator of biomass burning (Chantara et al., 2019; Rahman et al., 2020; Tao et al., 2014; Zhang et al., 2020; Zhang et al., 2015). Coal combustion, biomass combustion and vehicle exhaust are often considered to be the main sources of OC and EC (Cao et al., 2005; Cheng et al., 2013; Gao et al., 2018).

The highest value of $PM_{2.5}$ concentration (275 μg/m^3) was observed at 20:00 on November 20, when the WS, RH, NO_2 and SO_2 were 1.6 m/s, 90%, 70 μg/m^3 and 17 μg/m^3, respectively. High humidity, low wind speed, and high concentrations of precursors are often the main causes of high concentrations of $PM_{2.5}$ (Sun et al., 2013; Tai et al., 2010; Xu et al., 2011). Details of some variables were presented in Fig. 5.

3.5. Mechanisms of the formation of nitrate

In recent years, the proportion of NO_3^- mass concentration has increased significantly (Pan et al., 2016b; Sun et al., 2015; Teng et al., 2017; Zhang et al., 2013) and it has become the main component of $PM_{2.5}$. Previous studies have confirmed that SIA accounted for about 30%-60% of the total mass concentration of $PM_{2.5}$ in the North China Plain (NCP) with a large population and a high degree of industrialization (Huang et al., 2014; Tao et al., 2017; Zheng et al., 2015b). Similar to the results of previous studies, SIA accounted for 62.5% of the total $PM_{2.5}$ mass concentration in our study, with NO_3^- ranking first (27.6%). The observation period was divided into daytime and nighttime based on sunrise and sunset schedules (https://richurimo.bmcx.com) to explore the effects of gaseous precursors and meteorological conditions (Wang et al., 2009) on NO_3^-.

As shown in Fig. 6, the higher NO_2 and O_3 concentrations and higher relative humidity contributed positively to NO_3^- during the daytime and nighttime. NO_2 is oxidized by OH generated by photochemistry to form gaseous nitric acid, which then reacts with alkaline substances (primarily NH_3) on the surface of atmospheric particles to form particulate NO_3^-(Calvert and Stockwell, 1983; Guo et al., 2016) during the daytime. A recent study (Fu et al., 2020) indicated that the high production of photochemical oxidants O_3 and OH accelerated the formation of nitrate in the NCP. At nighttime, the conversion of NO_2 was significantly accelerated when the O_3 concentration was above 60 μg/m^3 (Figure S6a, Fig. 6). NO_2 is oxidized to NO_3·, which reacts with NO_2 to form N_2O_5. N_2O_5 undergoes a non-homogeneous hydrolysis reaction at the particle surface to form NO_3^- (Pathak et al., 2011; Pathak et al., 2009). A small quantity of NO_3· can also react directly in a non-homogeneous manner to form NO_3^- (Hallquist et al., 1999).

Temperature and humidity are also important drivers of NO_3^- concentration variation. At nighttime, the higher humidity (Figure S6b, Fig. 6) contributed more to NO_3^- formation, which was due to moisture promoting the hydrolysis of N_2O_5 to NO_3^- (Liu et al., 2020; Wang et al., 2017; Wu et al., 2018). The concentration of NO_3^- decreased with the increasing temperature (Figure S7). At noontime (Figure S7, 11:00 ~ 12:00), the concentration of NO_3^- increased because high OH concentration and strong photochemical activity promoted the conversion of NO_2 to NO_3^-. A recent study reported that OH played an important role in pollution in Beijing, with midday OH concentrations reaching to 2.4 × 10^6 cm^{-3}, almost an order of magnitude higher than that predicted by global or regional models in the NCP (Fu et al., 2020; Lelieveld et al., 2016; Tan et al., 2018; Zheng et al., 2015b). However, the concentration decreased after 12:00, probably due to the decomposition of gaseous nitric acid caused by higher temperature. Fig. 7 presented the decision process of the model from the base value to the predicted value and the specific contribution of each variable to NO_3^-. It can be seen from the figure that O_3 contributed the most to NO_3^- because O_3 acted as a photochemical oxidant promoted the gas-phase reaction of NO_2 and OH to form NO_3^-.

4. Conclusions

During the pollution period, the average $PM_{2.5}$ concentration was 111.9 μg/m^3 in Zibo, far exceeding the air quality standard in China (35 μg/m^3, GB 3095–2012).

The RF model can accurately predict the concentration of $PM_{2.5}$, with R^2 of 0.94, RMSE and MAE of 9.4 μg/m^3 and 5.7 μg/m^3, respectively. A preliminary qualitative analysis of the various drivers influencing $PM_{2.5}$ concentrations by Permutation importance showed that NH_4^+ had the highest weight, followed by NO_3^- and K, and then EC, OC, Al, etc.

The results of PDP showed that the order of the sensitivity of $PM_{2.5}$ was $NH_4^+ > NO_3^- > SO_4^{2-}$. In addition, SO_4^{2-} had a strong influence on $PM_{2.5}$ concentrations when NO_3^- concentration was less than 8.5 μg/m^3. When the concentration of NO_3^- was between 8.5 μg/m^3 and 33.5 μg/m^3, NH_4^+ was the primary driver to influence $PM_{2.5}$ concentrations. However, when NH_4^+ was greater than about 17.9 μg/m^3 and NO_3^- was greater than 33.5 μg/m^3, both of the two species together promoted

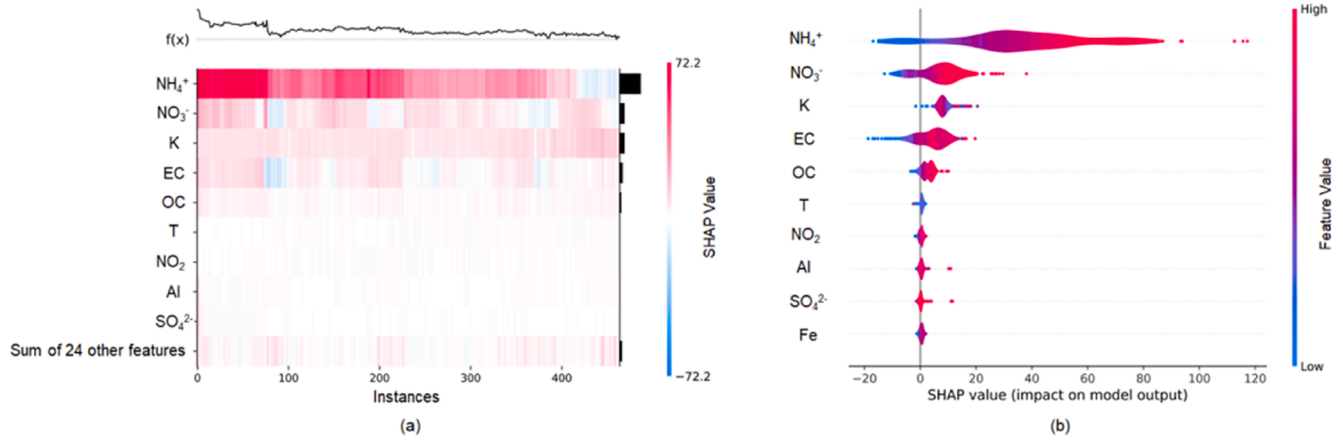

Fig. 5. Local explanations of some major variables based on the Random Forest models on $PM_{2.5}$ during air pollution events (A portion of the variables). (a) The heatmap plot of the major variables during the whole pollution period. The f(x) value represents the sum of the SHAP values of all the features of each sample (the sum of the sample dimension SHAP values), indicating the degree of deviation from the expected value. The x-axis is the sample sequence. The left side of the y-axis is the feature name and the right side represents the sum of the SHAP values of the feature (the sum of the feature dimension SHAP values). The red and blue stripes represent the SHAP value size of each feature for each data sample. (b) Feature density scatter plot (bee swarm plot) of the SHAP values for $PM_{2.5}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

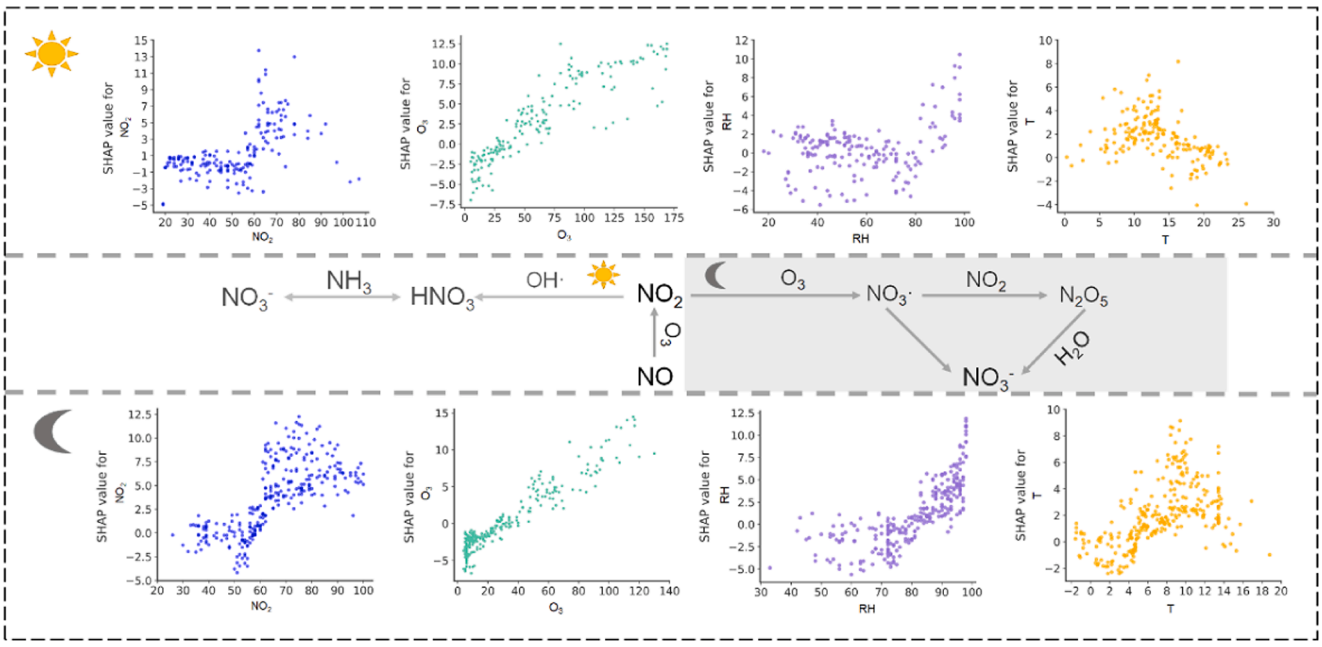

Fig. 6. SHAP dependence plot of some main factors for NO_3^- during day and night. The larger the absolute value of Shapley values, the greater the positive or negative contributions (impact) of the features to the NO_3^-.

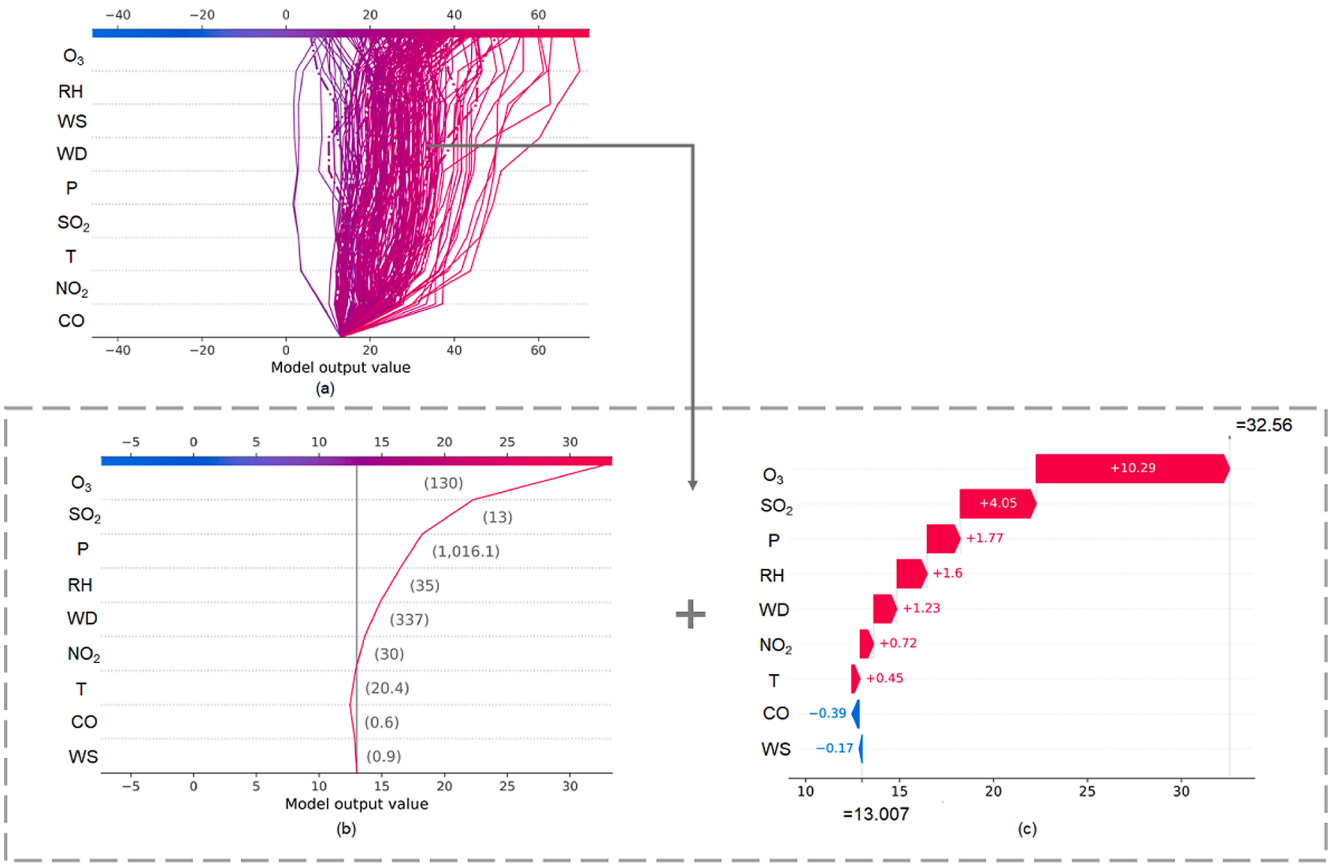

Fig. 7. Local interpretation based on the Random Forest model. (a) SHAP decision plot for NO_3^- during the daytime. The highlighted parts (dotted line) are 11:00–12:00 am during the pollution period. The upper color bar in the figure represents the predicted value of the model, the middle gray vertical line is the expected value of the model, and the colored lines indicate that the output value of each feature is higher or lower than the expected value of the model. Starting at the bottom of the graph, the colored lines represent the accumulation of Shapley values from the base value to the predicted value of the final output of the model at the top of the graph. (b) Enlarged view of the one of the highlighted parts (11:00–12:00, October 26). The values next to the colored line represent the measured value of the variables. (c) SHAP waterfall plot shows the contribution of each variable to the result and the model predicted value is 32.56.

$PM_{2.5}$ generation (~15–60 μg/m^3). NH_4^+ contributed 19.9–65.4 μg/m^3 during the ten air pollution events and was the most influential driver of $PM_{2.5}$ concentration increase. K, NO_3^-, EC and OC were the other main drivers, contributing $PM_{2.5}$ concentrations of 8.7 ± 2.7 μg/m^3, 6.8 ± 7.5 μg/m^3, 3.6 ± 5.8 μg/m^3, and 2.5 ± 2.0 μg/m^3, respectively. Fossil fuel combustion and biomass combustion were mainly the factors that contributed to the air pollution events in the region.

NO_3^- was the most abundant chemical component in $PM_{2.5}$, with an average concentration of 13.1 μg/m^3, accounting for 27.6% of the total mass concentration of $PM_{2.5}$. O_3 played an important role in the production of nitrate. Lower temperature and higher humidity were the other two drivers that promoted the formation of nitrate.

This work qualitatively analyzed the key drivers influencing air pollution, and then characterized the contribution of each driver. Here, we first used decision plot to map the decision process of the RF model, which showed how the model made decisions at specific points in time as well as to see the specific contributions of variables to air pollution. This study may provide a frame for data-driven analysis based on interpretable algorithms for air pollution prevention and control.

CRediT authorship contribution statement

Tianshuai Li: Conceptualization, Data curation, Formal analysis, Writing – original draft. **Qingzhu Zhang:** Formal analysis, Writing – original draft. **Yanbo Peng:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing. **Xu Guan:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing. **Lei Li:** Formal analysis, Writing – original draft. **Jiangshan Mu:** Formal analysis, Writing – original draft. **Xinfeng Wang:** Investigation, Resources. **Xianwei Yin:** Investigation, Resources. **Qiao Wang:** Investigation, Resources.

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.

Data availability

Data will be made available on request.

Acknowledgements

The work was financially supported by National Natural Science Foundation of China (project No. 22236004).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2023.107861.

References

Alfi, I., Rahman, M., Shorfuzzaman, M., Nazir, A., 2022. A non-invasive interpretable diagnosis of melanoma skin cancer using deep learning and ensemble stacking of machine learning models. Diagnostics 12.

Andersson, A., Deng, J., Du, K., Zheng, M., Yan, C., Skold, M., Gustafsson, O., 2015. Regionally-varying combustion sources of the January 2013 severe haze events over eastern China. Environmental Science & Technology 49, 2038–2043.

Bourdrel, T., Bind, M., Bejot, Y., Morel, O., Argacha, J., 2017. Cardiovascular effects of air pollution. Archives of Cardiovascular Diseases 110, 634–642.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Calvert, J., Stockwell, W., 1983. Acid generation in the troposphere by gas-phere chemistry. Environmental Science & Technology 17, A428–A443.

Cao, J., Wu, F., Chow, J., Lee, S., Li, Y., Chen, S., An, Z., Fung, K., Watson, J., Zhu, C., Liu, S., 2005. Characterization and source apportionment of atmospheric organic and elemental carbon during fall and winter of 2003 in Xi'an, China. Atmospheric Chemistry and Physics 5, 3127–3137.

Chantara, S., Thepnuan, D., Wiriya, W., Prawan, S., Tsai, Y., 2019. Emissions of pollutant gases, fine particulate matters and their significant tracers from biomass burning in an open-system combustion chamber. Chemosphere 224, 407–416.

Chen, Z., Chen, D., Kwan, M., Chen, B., Gao, B., Zhuang, Y., Li, R., Xu, B., 2019. The control of anthropogenic emissions contributed to 80% of the decrease in $PM_{2.5}$ concentrations in Beijing from 2013 to 2017. Atmospheric Chemistry and Physics 19, 13519–13533.

Chen, G., Li, S., Knibbs, L., Hamm, N., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M., Guo, Y., 2018. A machine learning method to estimate $PM_{2.5}$ concentrations across China with remote sensing, meteorological and land use information. Science of The Total Environment 636, 52–60.

Cheng, Y., Engling, G., He, K., Duan, F., Ma, Y., Du, Z., Liu, J., Zheng, M., Weber, R., 2013. Biomass burning contribution to Beijing aerosol. Atmospheric Chemistry and Physics 13, 7765–7781.

Crutzen, P., 1979. Role of NO and NO_2 in the chemistry of the troposphere and stratosphere. Annual Review of Earth and Planetary Sciences 7, 443–472.

Friedman, J., 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics 29, 1189–1232.

Fu, X., Wang, T., Gao, J., Wang, P., Liu, Y., Wang, S., Zhao, B., Xue, L., 2020. Persistent heavy winter nitrate pollution driven by increased photochemical oxidants in northern China. Environmental Science & Technology 54, 3881–3889.

Gao, J., Wang, K., Wang, Y., Liu, S., Zhu, C., Hao, J., Liu, H., Hua, S., Tian, H., 2018. Temporal-spatial characteristics and source apportionment of $PM_{2.5}$ as well as its associated chemical species in the Beijing-Tianjin-Hebei region of China. Environmental Pollution 233, 714–724.

Geng, G., Xiao, Q., Liu, S., Liu, X., Cheng, J., Zheng, Y., Xue, T., Tong, D., Zheng, B., Peng, Y., Huang, X., He, K., Zhang, Q., 2021. Tracking air pollution in China: near real-time $PM_{2.5}$ retrievals from multisource data fusion. Environmental Science & Technology 55, 12106–12115.

Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. Journal of Computational and Graphical Statistics 24, 44–65.

Griffith, S., Huang, X., Louie, P., Yu, J., 2015. Characterizing the thermodynamic and chemical composition factors controlling $PM_{2.5}$ nitrate: Insights gained from two years of online measurements in Hong Kong. Atmospheric Environment 122, 864–875.

Guo, H., Sullivan, A., Campuzano-Jost, P., Schroder, J., Lopez-Hilfiker, F., Dibb, J., Jimenez, J., Thornton, J., Brown, S., Nenes, A., Weber, R., 2016. Fine particle pH and the partitioning of nitric acid during winter in the northeastern United States. Journal of Geophysical Research-Atmospheres 121, 10355–10376.

Hallquist, M., Wangberg, I., Ljungstrom, E., Barnes, I., Becker, K., 1999. Aerosol and product yields from NO_3 radical-initiated oxidation of selected monoterpenes. Environmental Science & Technology 33, 553–559.

Hou, L., Dai, Q., Song, C., Liu, B., Guo, F., Dai, T., Li, L., Liu, B., Bi, X., Zhang, Y., Feng, Y., 2022. Revealing drivers of haze pollution by explainable machine learning. Environmental Science & Technology Letters 9, 112–119.

Hu, J., Li, X., Huang, L., Ying, Q., Zhang, Q., Zhao, B., Wang, S., Zhang, H., 2017. Ensemble prediction of air quality using the WRF/CMAQ model system for health effect studies in China. Atmospheric Chemistry and Physics 17, 13103–13118.

Huang, R., Zhang, Y., Bozzetti, C., Ho, K., Cao, J., Han, Y., Daellenbach, K., Slowik, J., Platt, S., Canonaco, F., Zotter, P., Wolf, R., Pieber, S., Bruns, E., Crippa, M., Ciarelli, G., Piazzalunga, A., Schwikowski, M., Abbaszade, G., Schnelle-Kreis, J., Zimmermann, R., An, Z., Szidat, S., Baltensperger, U., El Haddad, I., Prevot, A., 2014. High secondary aerosol contribution to particulate pollution during haze events in China. Nature 514, 218–222.

Lam, Y., Cheung, C., Zhang, X., Fu, J., Fung, J., 2021. Development of a new emission reallocation method for industrial sources in China. Atmospheric Chemistry and Physics 21, 12895–12908.

Lee, H., Honda, Y., Hashizume, M., Guo, Y., Wu, C., Kan, H., Jung, K., Lim, Y., Yi, S., Kim, H., 2015. Short-term exposure to fine and coarse particles and mortality: a multicity time-series study in East Asia. Environmental Pollution 207, 43–51.

Lelieveld, J., Gromov, S., Pozzer, A., Taraborrelli, D., 2016. Global tropospheric hydroxyl distribution, budget and reactivity. Atmospheric Chemistry and Physics 16, 12477–12493.

Lewin-Epstein, O., Baruch, S., Hadany, L., Stein, G., Obolski, U., 2021. Predicting antibiotic resistance in hospitalized patients by applying machine learning to electronic medical records. Clinical Infectious Diseases 72, E848–E855.

Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., Tong, D., Zheng, B., Cui, H., Man, H., Zhang, Q., He, K., 2017b. Anthropogenic emission inventories in China: a review. National Science Review 4, 834–866.

Li, H., Ma, Y., Duan, F., He, K., Zhu, L., Huang, T., Kimoto, T., Ma, X., Ma, T., Xu, L., Xu, B., Yang, S., Ye, S., Sun, Z., An, J., Zhang, Z., 2017a. Typical winter haze pollution in Zibo, an industrial city in China: characteristics, secondary formation, and regional contribution. Environmental Pollution 229, 339–349.

Liu, L., Bei, N., Hu, B., Wu, J., Liu, S., Li, X., Wang, R., Liu, Z., Shen, Z., Li, G., 2020. Wintertime nitrate formation pathways in the North China Plain: importance of N_2O_5 heterogeneous hydrolysis. Environmental Pollution 266, 115287.

Lu, K., Fuchs, H., Hofzumahaus, A., Tan, Z., Wang, H., Zhang, L., Schmitt, S., Rohrer, F., Bohn, B., Broch, S., Dong, H., Gkatzelis, G., Hohaus, T., Holland, F., Li, X., Liu, Y., Liu, Y., Ma, X., Novelli, A., Schlag, P., Shao, M., Wu, Y., Wu, Z., Zeng, L., Hu, M., Kiendler-Scharr, A., Wahner, A., Zhang, Y., 2019. Fast photochemistry in wintertime haze: consequences for pollution mitigation strategies. Environmental Science & Technology 53, 10676–10684.

Lundberg, S., Lee, S., 2017. A unified approach to interpreting model predictions. 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA.

Lundberg, S., Nair, B., Vavilala, M., Horibe, M., Eisses, M., Adams, T., Liston, D., Low, D., Newman, S., Kim, J., Lee, S., 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature Biomedical Engineering 2, 749–760.

Lundberg, S., Erion, G., Chen, H., DeGrave, A., Prutkin, J., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S., 2020. From local explanations to global Understanding with explainable AI for trees. Nature Machine Intelligence 2, 56–67.

Ma, Z., Liu, R., Liu, Y., Bi, J., 2019. Effects of air pollution control policies on $PM_{2.5}$ pollution improvement in China from 2005 to 2017: a satellite-based perspective. Atmospheric Chemistry and Physics 19, 6861–6877.

Ming, L., Jin, L., Li, J., Fu, P., Yang, W., Liu, D., Zhang, G., Wang, Z., Li, X., 2017. $PM_{2.5}$ in the Yangtze River Delta, China: chemical compositions, seasonal variations, and regional pollution events. Environmental Pollution 223, 200–212.

Pan, Y., Tian, S., Liu, D., Fang, Y., Zhu, X., Zhang, Q., Zheng, B., Michalski, G., Wang, Y., 2016a. Fossil fuel combustion-related emissions dominate atmospheric ammonia sources during severe haze episodes: evidence from ^{15}N-Stable Isotope in size-resolved aerosol ammonium. Environmental Science & Technology 50, 8049–8056.

Pan, Y., Wang, Y., Zhang, J., Liu, Z., Wang, L., Tian, S., Tang, G., Gao, W., Ji, D., Song, T., Wang, Y., 2016b. Redefining the importance of nitrate during haze pollution to help optimize an emission control strategy. Atmospheric Environment 141, 197–202.

Pathak, R., Wu, W., Wang, T., 2009. Summertime $PM_{2.5}$ ionic species in four major cities of China: nitrate formation in an ammonia-deficient atmosphere. Atmospheric Chemistry and Physics 9, 1711–1722.

Pathak, R., Wang, T., Wu, W., 2011. Nighttime enhancement of $PM_{2.5}$ nitrate in ammonia-poor atmospheric conditions in Beijing and Shanghai: plausible contributions of heterogeneous hydrolysis of N_2O_5 and HNO_3 partitioning. Atmospheric Environment 45, 1183–1191.

Peng, J., Chen, S., Lü, H., Liu, Y., Wu, J., 2016. Spatiotemporal patterns of remotely sensed $PM_{2.5}$ concentration in China from 1999 to 2011. Remote Sensing of Environment 174, 109–121.

Rahman, M., Begum, B., Hopke, P., Nahar, K., Thurston, G., 2020. Assessing the $PM_{2.5}$ impact of biomass combustion in megacity Dhaka, Bangladesh. Environmental Pollution 264, 114798.

Rahman, M., Azad, M., Hasanuzzaman, M., Salam, R., Islam, A., Rahman, M., Hoque, M., 2021. How air quality and COVID-19 transmission change under different lockdown scenarios? a case from Dhaka city. Bangladesh. Science of The Total Environment 762, 143161.

Shandong Bureau of Statistics, 2021. Shandong Statistical Yearbook 2021. China Statistics Press, Beijing.

Shapley, L., 1953. Stochastic games. Proceedings of the National Academy of Sciences of the United States of America 39, 1095-1100.

Sun, Z., Mu, Y., Liu, Y., Shao, L., 2013. A comparison study on airborne particles during haze days and non-haze days in Beijing. Science of The Total Environment 456, 1–8.

Sun, Y., Wang, Z., Du, W., Zhang, Q., Wang, Q., Fu, P., Pan, X., Li, J., Jayne, J., Worsnop, D., 2015. Long-term real-time measurements of aerosol particle composition in Beijing, China: seasonal variations, meteorological effects, and source analysis. Atmospheric Chemistry and Physics 15, 10149–10165.

Tai, A., Mickley, L., Jacob, D., 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. Atmospheric Environment 44, 3976–3984.

Tan, Z., Rohrer, F., Lu, K., Ma, X., Bohn, B., Broch, S., Dong, H., Fuchs, H., Gkatzelis, G., Hofzumahaus, A., Holland, F., Li, X., Liu, Y., Liu, Y., Novelli, A., Shao, M., Wang, H., Wu, Y., Zeng, L., Hu, M., Kiendler-Scharr, A., Wahner, A., Zhang, Y., 2018. Wintertime photochemistry in Beijing: observations of ROx radical concentrations in the North China Plain during the BEST-ONE campaign. Atmospheric Chemistry and Physics 18, 12391–12411.

Tao, J., Gao, J., Zhang, L., Zhang, R., Che, H., Zhang, Z., Lin, Z., Jing, J., Cao, J., Hsu, S., 2014. $PM_{2.5}$ pollution in a megacity of southwest China: source apportionment and implication. Atmospheric Chemistry and Physics 14, 8679–8699.

Tao, J., Zhang, L., Cao, J., Zhang, R., 2017. A review of current knowledge concerning $PM_{2.5}$ chemical composition, aerosol optical properties and their relationships across China. Atmospheric Chemistry and Physics 17, 9485–9518.

Teng, X., Hu, Q., Zhang, L., Qi, J., Shi, J., Xie, H., Gao, H., Yao, X., 2017. Identification of major sources of atmospheric NH_3 in an urban environment in northern China during wintertime. Environmental Science & Technology 51, 6839–6848.

Tian, H., Zhu, C., Gao, J., Cheng, K., Hao, J., Wang, K., Hua, S., Wang, Y., Zhou, J., 2015. Quantitative assessment of atmospheric emissions of toxic heavy metals from anthropogenic sources in China: historical trend, spatial distribution, uncertainties, and control policies. Atmospheric Chemistry and Physics 15, 10127–10147.

Twigg, M., Di Marco, C., Leeson, S., van Dijk, N., Jones, M., Leith, I., Morrison, E., Coyle, M., Proost, R., Peeters, A., Lemon, E., Frelink, T., Braban, C., Nemitz, E., Cape, J., 2015. Water soluble aerosols and gases at a UK background site – Part 1: Controls of $PM_{2.5}$ and PM_{10} aerosol composition. Atmospheric Chemistry and Physics 15, 8131–8145.

Vu, T., Shi, Z., Cheng, J., Zhang, Q., He, K., Wang, S., Harrison, R., 2019. Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique. Atmospheric Chemistry and Physics 19, 11303–11314.

Wang, M., Duan, Y., Zhang, Z., Huo, J., Huang, Y., Fu, Q., Wang, T., Cao, J., Lee, S., 2022a. Increased contribution to $PM_{2.5}$ from traffic-influenced road dust in Shanghai over recent years and predictable future. Environmental Pollution 313, 120119.

Wang, J., Ye, J., Zhang, Q., Zhao, J., Wu, Y., Li, J., Liu, D., Li, W., Zhang, Y., Wu, C., Xie, C., Qin, Y., Lei, Y., Huang, X., Guo, J., Liu, P., Fu, P., Li, Y., Lee, H., Choi, H., Zhang, J., Liao, H., Chen, M., Sun, Y., Ge, X., Martin, S., Jacob, D., 2021. Aqueous production of secondary organic aerosol from fossil-fuel emissions in winter Beijing haze. Proceedings of the National Academy of Sciences of the United States of America 118.

Wang, J., Li, J., Ye, J., Zhao, J., Wu, Y., Hu, J., Liu, D., Nie, D., Shen, F., Huang, X., Huang, D., Ji, D., Sun, X., Xu, W., Guo, J., Song, S., Qin, Y., Liu, P., Turner, J., Lee, H., Hwang, S., Liao, H., Martin, S., Zhang, Q., Chen, M., Sun, Y., Ge, X., Jacob, D., 2020a. Fast sulfate formation from oxidation of SO_2 by NO_2 and HONO observed in Beijing haze. Nature Communications 11, 2844.

Wang, H., Lu, K., Chen, X., Zhu, Q., Chen, Q., Guo, S., Jiang, M., Li, X., Shang, D., Tan, Z., Wu, Y., Wu, Z., Zou, Q., Zheng, Y., Zeng, L., Zhu, T., Hu, M., Zhang, Y., 2017. High N_2O_5 concentrations observed in urban Beijing: implications of a large nitrate formation pathway. Environmental Science & Technology Letters 4, 416–420.

Wang, Y., Wen, Y., Wang, Y., Zhang, S., Zhang, K., Zheng, H., Xing, J., Wu, Y., Hao, J., 2020b. Four-month changes in air quality during and after the COVID-19 lockdown in six megacities in China. Environmental Science & Technology Letters 7, 802–808.

Wang, Y., Du, Y., Fang, J., Dong, X., Wang, Q., Ban, J., Sun, Q., Ma, R., Zhang, W., He, M., Liu, C., Niu, Y., Chen, R., Kan, H., Li, T., 2022b. A random forest model for daily $PM_{2.5}$ personal exposure assessment for a Chinese cohort. Environmental Science & Technology Letters 9, 466–472.

Wang, S., Yin, S., Zhang, R., Yang, L., Zhao, Q., Zhang, L., Yan, Q., Jiang, N., Tang, X., 2019. Insight into the formation of secondary inorganic aerosol based on high-time-resolution data during haze episodes and snowfall periods in Zhengzhou, China. Science of The Total Environment 660, 47–56.

Wang, X., Zhang, Y., Chen, H., Yang, X., Chen, J., Geng, F., 2009. Particulate nitrate formation in a highly polluted urban area: a case study by single-particle mass spectrometry in Shanghai. Environmental Science & Technology 43, 3061–3066.

Weagle, C., Snider, G., Li, C., van Donkelaar, , Philip, S., Bissonnette, P., Burke, J., Jackson, J., Latimer, R., Stone, E., Abboud, I., Akoshile, C., Anh, N., Brook, J., Cohen, A., Dong, J., Gibson, M., Griffith, D., He, K., Holben, B., Kahn, R., Keller, C., Kim, J., Lagrosas, N., Lestari, P., Khian, Y., Liu, Y., Marais, E., Martins, J., Misra, A., Muliane, U., Pratiwi, R., Quel, E., Salam, A., Segev, L., Tripathi, S., Wang, C., Zhang, Q., Brauer, M., Rudich, Y., Martin, R., 2018. Global sources of fine particulate matter: interpretation of $PM_{2.5}$ chemical composition observed by spartan using a global chemical transport model. Environmental Science & Technology.

Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., Cribb, M., 2019. Estimating 1-km-resolution $PM_{2.5}$ concentrations across China using the space-time random forest approach. Remote Sensing of Environment 231, 111221.

Wei, N., Jia, Z., Men, Z., Ren, C., Zhang, Y., Peng, J., Wu, L., Wang, T., Zhang, Q., Mao, H., 2022. Machine learning predicts emissions of brake wear $PM_{2.5}$: model construction and interpretation. Environmental Science & Technology Letters 9, 352–358.

Wu, Z., Wang, Y., Tan, T., Zhu, Y., Li, M., Shang, D., Wang, H., Lu, K., Guo, S., Zeng, L., Zhang, Y., 2018. Aerosol liquid water driven by anthropogenic inorganic salts: implying its key role in haze formation over the North China Plain. Environmental Science & Technology Letters 5, 160–166.

Xing, Y., Xu, Y., Shi, M., Lian, Y., 2016. The impact of $PM_{2.5}$ on the human respiratory system. Journal of Thoracic Disease 8, E69–E74.

Xu, Q., Wang, S., Jiang, J., Bhattarai, N., Li, X., Chang, X., Qiu, X., Zheng, M., Hua, Y., Hao, J., 2019. Nitrate dominates the chemical composition of $PM_{2.5}$ during haze event in Beijing. China. Science of The Total Environment 689, 1293–1303.

Xu, W., Zhao, C., Ran, L., Deng, Z., Liu, P., Ma, N., Lin, W., Xu, X., Yan, P., He, X., Yu, J., Liang, W., Chen, L., 2011. Characteristics of pollutants and their correlation to meteorological conditions at a suburban site in the North China Plain. Atmospheric Chemistry and Physics 11, 4353–4369.

Yang, J., Wen, Y., Wang, Y., Zhang, S., Pinto, J., Pennington, E., Wang, Z., Wu, Y., Sander, S., Jiang, J., Hao, J., Yung, Y., Seinfeld, J., 2021. From COVID-19 to future electrification: assessing traffic impacts on air quality by a machine-learning model. Proceedings of the National Academy of Sciences of the United States of America 118.

Yang, N., Shi, H., Tang, H., Yang, X., 2022. Geographical and temporal encoding for improving the estimation of $PM_{2.5}$ concentrations in China using end-to-end gradient boosting. Remote Sensing of Environment 269, 112828.

Zhang, Q., Zheng, Y., Tong, D., Shao, M., Wang, S., Zhang, Y., Xu, X., Wang, J., He, H., Liu, W., Ding, Y., Lei, Y., Li, J., Wang, Z., Zhang, X., Wang, Y., Cheng, J., Liu, Y., Shi, Q., Yan, L., Geng, G., Hong, C., Li, M., Liu, F., Zheng, B., Cao, J., Ding, A., Gao, J., Fu, Q., Huo, J., Liu, B., Liu, Z., Yang, F., He, K., Hao, J., 2019. Drivers of improved $PM_{2.5}$ air quality in China from 2013 to 2017. Proceedings of the National Academy of Sciences of the United States of America 116, 24463-24469.

Zhang, Y., Huang, R., El Haddad, I., Ho, K., Cao, J., Han, Y., Zotter, P., Bozzetti, C., Daellenbach, K., Canonaco, F., Slowik, J., Salazar, G., Schwikowski, M., Schnelle-Kreis, J., Abbaszade, G., Zimmermann, R., Baltensperger, U., Prévôt, A., Szidat, S., 2015. Fossil vs. non-fossil sources of fine carbonaceous aerosols in four Chinese cities during the extreme winter haze episode of 2013. Atmospheric Chemistry and Physics 15, 1299–1312.

Zhang, R., Khalizov, A., Wang, L., Hu, M., Xu, W., 2012. Nucleation and growth of nanoparticles in the atmosphere. Chemical Reviews 112, 1957–2011.

Zhang, R., Jing, J., Tao, J., Hsu, S., Wang, G., Cao, J., Lee, C., Zhu, L., Chen, Z., Zhao, Y., Shen, Z., 2013. Chemical characterization and source apportionment of $PM_{2.5}$ in Beijing: seasonal perspective. Atmospheric Chemistry and Physics 13, 7053–7074.

Zhang, J., Liu, L., Xu, L., Lin, Q., Zhao, H., Wang, Z., Guo, S., Hu, M., Liu, D., Shi, Z., Huang, D., Li, W., 2020. Exploring wintertime regional haze in northeast China: role of coal and biomass burning. Atmospheric Chemistry and Physics 20, 5355–5372.

Zhang, Z., Xu, B., Xu, W., Wang, F., Gao, J., Li, Y., Li, M., Feng, Y., Shi, G., 2022. Machine learning combined with the PMF model reveal the synergistic effects of sources and meteorological factors on $PM_{2.5}$ pollution. Environmental Research 212, 113322.

Zheng, G., Duan, F., Su, H., Ma, Y., Cheng, Y., Zheng, B., Zhang, Q., Huang, T., Kimoto, T., Chang, D., Poschl, U., Cheng, Y., He, K., 2015b. Exploring the severe

winter haze in Beijing: the impact of synoptic weather, regional transport and heterogeneous reactions. Atmospheric Chemistry and Physics 15, 2969–2983.
Zheng, B., Zhang, Q., Zhang, Y., He, K., Wang, K., Zheng, G., Duan, F., Ma, Y., Kimoto, T., 2015a. Heterogeneous chemistry: a mechanism missing in current models to explain secondary inorganic aerosol formation during the January 2013 haze episode in North China. Atmospheric Chemistry and Physics 15, 2031–2049.
Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., Li, H., Li, X., Peng, L., Qi, J., Yan, L., Zhang, Y., Zhao, H., Zheng, Y., He, K., Zhang, Q., 2018. Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. Atmospheric Chemistry and Physics 18, 14095–14111.