Environmental Science & Technology Letters

pubs.acs.org/journal/estlcu

Letter

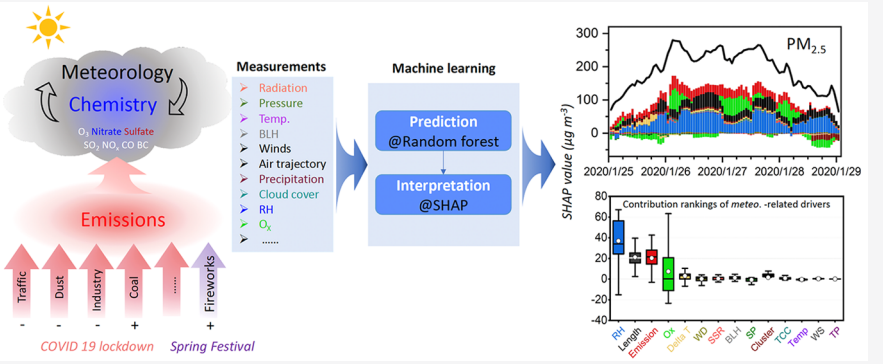
Revealing Drivers of Haze Pollution by Explainable Machine Learning

Linlu Hou, Qili Dai,* Congbo Song, Bowen Liu, Fangzhou Guo, Tianjiao Dai, Linxuan Li, Baoshuang Liu, Xiaohui Bi, Yufen Zhang, and Yinchang Feng

Cite This: https://doi.org/10.1021/acs.estlett.1c00865

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Many places on earth still suffer from a high level of atmospheric fine particulate matter ($PM_{2.5}$) pollution. Formation of a particulate pollution event or haze episode (HE) involves many factors, including meteorology, emissions, and chemistry. Understanding the direct causes of and key drivers behind the HE is thus essential. Traditionally, this is done via chemical transport models. However, substantial uncertainties are introduced into the model estimation when there are significant changes in the emissions inventory due to interventions (e.g., the COVID-19 lockdown). Here we applied a Random Forest model coupled with a Shapley additive explanation algorithm, a *post hoc* explanation technique, to investigate the roles of major meteorological factors, primary emissions, and chemistry in five severe HEs that occurred before or during the COVID-19 lockdown in China. We discovered that, in addition to the high level of primary emissions, $PM_{2.5}$ in these haze episodes was largely driven by meteorological effects (with average contributions of 30−65 $\mu g\ m^{-3}$ for the five HEs), followed by chemistry (~15−30 $\mu g\ m^{-3}$). Photochemistry was likely the major pathway of formation of nitrate, while air humidity was the predominant factor in forming sulfate. Our results highlight that the machine learning driven by data has the potential to be a complementary tool in predicting and interpreting air pollution.

INTRODUCTION

The world still carries a heavy burden of severe air pollution events, which is often linked to the high concentration of airborne fine particulate matter ($PM_{2.5}$).[1−3] Generally, air pollution is a function of enormous emissions and meteorology.[4,5] A severe pollution event can be formed through a rapid increase in emission rates, atmospheric conditions that favor the formation of secondary PM from gaseous pollutants, meteorological stagnation (weak surface winds, shallow mixing layer, etc.) that limits the dispersion of pollutants, and strong transport of pollutants.[1,6−9] Examining the roles of emissions and meteorology as well as chemistry in processing haze formation is essential before regulatory strategies to reduce air pollution can be formulated.

Traditionally, chemical transport models (CTMs) are the most popular tools for performing such analysis via the simulation of scenarios and process analysis,[10−13] and the model performance heavily relies on the accuracy and up-to-date availability of an emission inventory. Recently, a machine learning (ML)-based meteorological normalization technique (also called "deweathering") has been intensively used in atmospheric research.[14−17] It is a promising alternative to account for the effects of meteorology on air pollutants. The model performance of ML is generally better than that of traditional statistical analysis (e.g., linear models) and CTMs in predicting, for example, the $PM_{2.5}$ level;[17,18] however, these results are less robust in physical interpretability due to the "black box" nature of most ML models. With the development

Received: October 25, 2021
Revised: December 21, 2021
Accepted: December 22, 2021

of feature attribution techniques, tools that allow researchers to explore the importance of features are gradually becoming available.[19−22] Lundberg and Lee[23] proposed a SHapley Additive exPlanation (SHAP) approach based on game theory to quantify the global and local impacts of input features on model predictions, which has been utilized in many fields[24−28] but rarely applied to the field of atmospheric science.[29,30] Such a tool has the potential to enhance the physical interpretation of the contributors and/or drivers of air pollution.

Unexpected severe haze episodes occurred in northern China during the COVID-19 lockdowns in early 2020,[15,31] which attracted a great deal of public attention.[32] Existing studies based on field measurements and conventional CTMs suggested that the pandemic hazes were formed with extremely unfavorable meteorological conditions compared to a year before,[33,34] together with enhanced secondary nitrate chemistry[32,35] as well as increased festival emissions and residential burning.[36] Given that emission rates changed dramatically due to the lockdowns, using CTMs with an outdated inventory to reproduce the haze episodes can thus lead to large degrees of uncertainty. Quantifying the major drivers behind the HEs remains a challenge. Dai et al.[37] reported that the five HEs observed before and during the lockdown in Tianjin, China, were formed with distinctive emissions under different meteorological conditions; thus, these HEs provide a unique opportunity to examine the feasibility of coupling the SHAP approach with the ML technique in identifying key contributors.

Here we first applied a ML-based deweathering technique to investigate changes in emissions by decoupling meteorology from the observed concentrations of air pollutants. The SHAP approach was then used to quantify the roles of meteorological variables and chemistry in processing the five HEs.

MATERIALS AND METHODS

Source of the Data. A six-year (2015−2020) hourly data set, including air quality data and meteorological data, was collected and compiled into training models. Hourly concentrations of sulfur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3), and $PM_{2.5}$ were downloaded from the China National Environmental Monitoring Center (http://106.37.208.233:20035/). Hourly surface meteorological variables, including air temperature (Temp), relative humidity (RH), wind direction (WD), and wind speed (WS), recorded at Tianjin Binhai International Airport were retrieved from the "worldMet" R package (https://github.com/davidcarslaw/worldmet). Hourly ERA5 data, including the boundary layer height (BLH), total cloud cover (TCC), surface net solar radiation (SSR), surface pressure (SP), and total precipitation (TP), were collected from ECMWF. The 24 h temperature difference (DeltaT) for each measured temperature was calculated. The major secondary $PM_{2.5}$ species, sulfate and nitrate, in the studied period in 2019 and 2020 were measured via ion chromatography (URG 9000D, Thermo) at the Air Quality Research Supersite at Nankai University (NKUAQS, 38°59′N and 117°20′E).[36] The 72 h backward air mass trajectories (100 m AGL) were calculated for each measurement using the hybrid single-particle Lagrangian integrated trajectory model (HYSPLIT).[38] Trajectories were subjected to cluster analysis using the Euclidian distance to produce 12 clusters for subsequent analysis.

Random Forest (RF)-Based Meteorological Normalization. A model trained with only regular meteorological variables cannot reproduce the observed pollutant level well .[29] Surface meteorological variables are typically used together with a cluster of air mass trajectories and time variables, such as Unix time (number of seconds since January 1, 1970) as a linear trend term, Gregorian date (day of the year) as a seasonal term, day of the week, and hour of the day.[14,16,17,39] Time variables are usually considered as proxies for time-related drivers (e.g., emission rates). Here we fed the RF model with the aforementioned meteorological variables, the cluster and length of air mass trajectories, an additional time variable (lunar date[15]), and the total gaseous oxidant (i.e., $O_X = NO_2 + O_3$, a proxy representing atmospheric photooxidation conditions[16,40]). Details of the RF modeling are provided in Text S1 of the Supporting Information. The added variables greatly improved the performance of the $PM_{2.5}$ predictive model (e.g., r^2 increased from 0.624 ± 0.003 to 0.906 ± 0.001 from 100 repeated predictions with different seeds) (Table S1).

To decouple the impacts of meteorology and chemistry on the observed $PM_{2.5}$ as much as possible, the meteorologically normalized $PM_{2.5}$ concentration at a particular time was calculated by averaging 1000 predictions from the RF model with meteorological variables randomly resampled from the studied period (2015−2020).[14−16] All meteorological variables and the air mass cluster and length were replaced while the time variables were retained during the "deweathering" process. The deweathered $PM_{2.5}$ ($PM_{2.5,dew}$) at a particular time was thus interpreted as the ambient level of emissions (termed "$PM_{2.5,emission}$") under averaged meteorological conditions (eq 1).

$$PM_{2.5,emission} = PM_{2.5,dew} = \frac{1}{1000}\sum_{i=1}^{1000} C_{i,pred} \quad (1)$$

where $C_{i,pred}$ is the RF model-predicted concentration of $PM_{2.5}$ for a given meteorological condition at time i.

SHapley Additive ExPlanation (SHAP) Approach. The average importance of the input variables and their overall effect on RF model output can be determined via a partial dependence plot and a global importance plot, but the relationship between each variable and every prediction is not clear.[14] To address this, the SHAP approach,[25,26] which distributes the total gains among the players based on coalitional game theory,[41] was applied. Briefly, the difference in model prediction with a variable (e.g., j) against the prediction without j is attributed to the marginal contribution of variable j. Considering the interactive effects between variables, differences are computed for every possible variable subset combination of each sample.[23] For each predicted sample (x_i) that has K variables generating a predicted value [$f(x_i)$], the explanatory model f is a linear function of feature attribution (eq 2).

$$f(x_i) = \emptyset_0(f, x) + \sum_{j=1}^{K} \emptyset_j(f, x_i) \quad (2)$$

where $\emptyset_j(f, x_i)$ is the SHAP value representing the impact of variable j on the prediction of model f for input x_i. The base value, $\emptyset_0(f, x) = E[f(x)]$, is the expected value of the model output over the data set.

The SHAP value [$\emptyset_j(f, x)$] is the weighted average of $\emptyset_j$ values across all possible variable subset combinations.[23,26]

https://doi.org/10.1021/acs.estlett.1c00865
Environ. Sci. Technol. Lett. XXXX, XXX, XXX−XXX

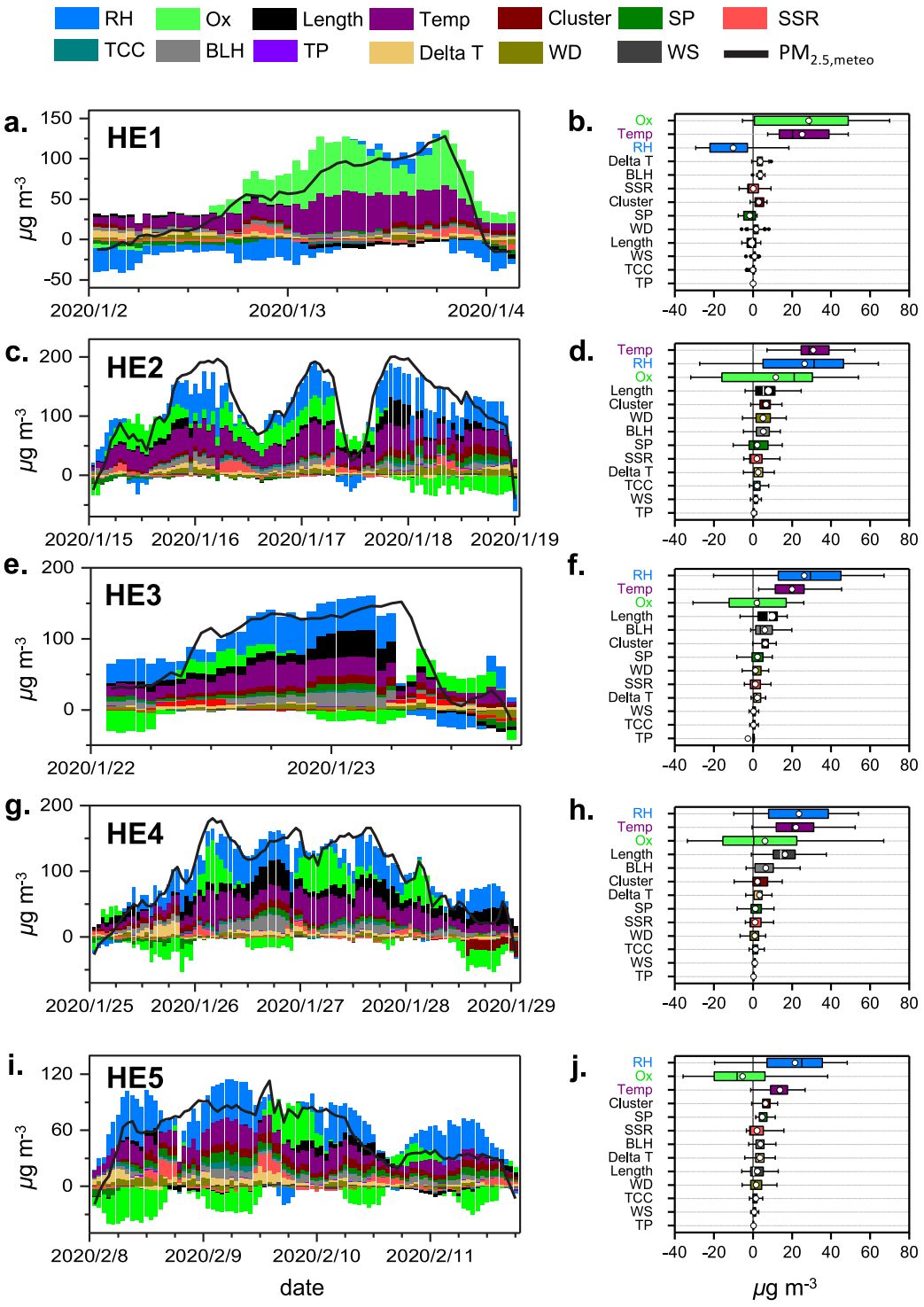


Figure 1. SHAP values from the $PM_{2.5,meteo}$ Random Forest model. Time series of each variable's SHAP values during the five HEs are shown in panels a, c, e, g, and i, and data are also shown in box plots with the order of the corresponding absolute average |SHAP| values (b, d, f, h, and j). Left and right box boundaries represent the 25th and 75th percentiles, respectively. Line and circular inside boxes represent median and mean values, respectively. Left and right error lines represent the 1.5 * IQR (interquartile range) below the third quartile and above the first quartile, respectively.

$$\Phi_j(f, x) = \sum_{S \subseteq K \setminus \{j\}} \frac{|S|!(K - |S| - 1)!}{K!} [f_x(S \cup \{j\}) - f_x(S)] \quad (3)$$

where $S \subseteq \{0, 1\}^K$ and K is the set of all K input variables. $|S|$ is the number of non-zero entries in S.

If $\Phi_j(f, x) > 0$ [$\Phi_j(f, x) < 0$], it refers to the positive (negative) effect of the variable j that increases (decreases) the prediction above (below) the base value, to infer the specific process driving the sign of the change in the pollutant concentration for each sample.[23,29]

After estimating the $PM_{2.5,emission}$ from a built RF model for $PM_{2.5}$, we rebuilt the RF model for the meteorologically driven

https://doi.org/10.1021/acs.estlett.1c00865
Environ. Sci. Technol. Lett. XXXX, XXX, XXX−XXX

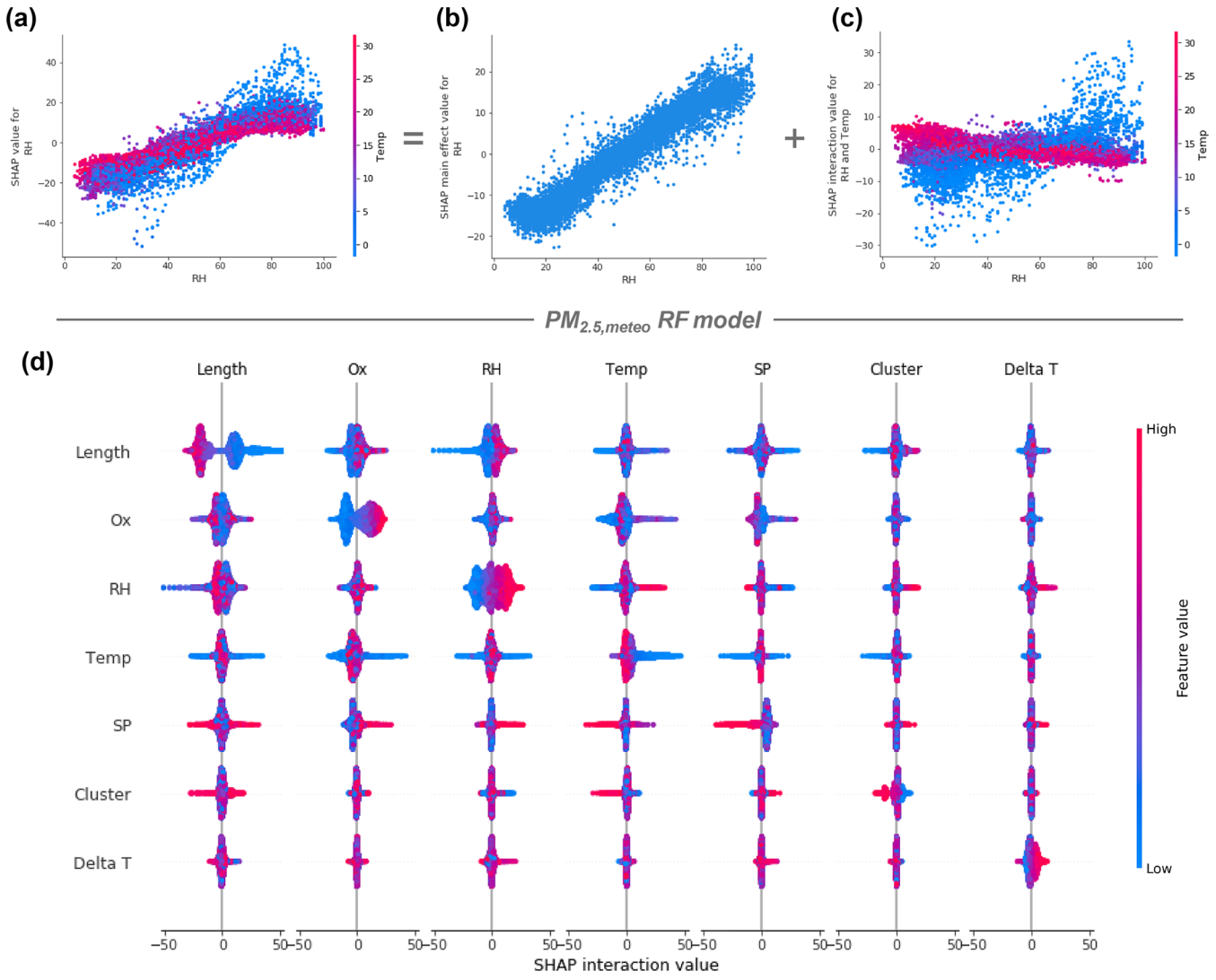


Figure 2. Main and interaction effects of major variables on $PM_{2.5,meteo}$ from SHAP ($\mu g\ m^{-3}$). The results are combined from local explanations based on a $PM_{2.5,meteo}$ Random Forest model. (a) SHAP dependence plot of RH vs its SHAP value. The dependence plots of other major variables are presented in Figure S7. The vertical dispersion of the data in the dependence plot (a) indicates a clear interaction effect between RH and temperature. (b) The main effect of RH shows a nearly linear increase in the effect of RH on $PM_{2.5,meteo}$. (c) The interaction SHAP value between RH and temperature shows how the effect of RH on $PM_{2.5,meteo}$ varies with temperature. An increase in temperature increased $PM_{2.5,meteo}$ concentrations for periods with relatively low RH ($\lesssim$50%) and decreased $PM_{2.5,\ meteo}$ concentrations for $\gtrsim$50% RH. (d) Summary plots of the SHAP interaction matrix values for $PM_{2.5,\ meteo}$. The main effects are on the diagonal, and the interaction effects off the diagonal.

$PM_{2.5}$ ($PM_{2.5,meteo}$, eq 4) with the aforementioned input variables (excluding time variables) and then utilized the SHAP value estimated with the "shap" python package (https://github.com/slundberg/shap) to quantify each variable's contribution to each $PM_{2.5,meteo}$.

$$PM_{2.5,meteo} = PM_{2.5,obs} - PM_{2.5,emission} \quad (4)$$

Similarly, the impacts of factors affecting the formation of nitrate and sulfate were also investigated. The models were built with inputs from only meteorological variables, air mass trajectories, and O_X. Pairwise interactive effects between features were also calculated (eqs S1−S3 and Text S2).

RESULTS AND DISCUSSION

The haze episodes were selected according to a 24 h moving average $PM_{2.5}$ concentration of >150 $\mu g\ m^{-3}$.[37] The duration for a given episode was calculated using another criterion, an hourly concentration threshold of 75 $\mu g\ m^{-3}$, to separate clean hours from polluted hours.[42] Statistically, there were five severe regional HEs observed in northern China in early 2020. The statistical description and contributing sources of the five HEs are available in ref 37. The first three HEs (HE1−HE3) occurred before the Chinese Lunar New Year's Eve (January 24) under "business as usual" conditions, while the last two HEs (HE4 and HE5) overlapped with the Chinese Spring Festival (CSF, January 25−30) and Lantern Festival (February 8), respectively. The COVID-19 lockdown in Tianjin started at the beginning of the CSF, leading to significant changes in emissions thereafter (Figure S1).[15,36]

Distinctive Drivers of the Five Haze Events. On average, primary emissions far exceeded the recently updated World Health Organization guidance level (i.e., 15 $\mu g\ m^{-3}$ for daily $PM_{2.5}$), contributing $PM_{2.5}$ concentrations of 91 ± 2, 89 ± 3, 86 ± 2, 97 ± 4, and 85 ± 3 $\mu g\ m^{-3}$ during HE1−HE5, respectively. This suggests that these HEs, particularly HE1 and HE4, were largely driven by primary emissions.[37] Variables

https://doi.org/10.1021/acs.estlett.1c00865
Environ. Sci. Technol. Lett. XXXX, XXX, XXX−XXX

associated with air mass trajectories (cluster and length) in HE1 contributed the least among the five HEs [~8 ± 3 μg m^{-3} $PM_{2.5}$ (Figure 1)], suggesting that the level of regional transport was the lowest during HE1. However, it was relatively high in the two HEs around New Year's Eve (~20 ± 14 μg m^{-3} in HE3 and ~23 ± 15 μg m^{-3} in HE4), which is in accordance with the local aerosol lidar measurement[37] and CTM modeling.[43]

O_X ranked first among the variables in HE1 (30 ± 24 μg m^{-3} $PM_{2.5}$). In contrast, RH was the leading meteorological factor that enhanced the formation of hazes excluding HE1. The SHAP value of O_X peaked from noon to afternoon (Figure 1 and Figure S2) due to the increase in radiation intensity and temperature.[44] The periodic increase in the SHAP value of SSR was greater than the increase in that of daytime O_X (Figure 1). Such diel patterns of and the time lag between SHAP values for SSR and O_X are indicative of the daytime photochemistry. The SHAP value of RH presented consistent diel patterns in all HEs, with a negative impact in the daytime and the strongest positive impact at night (Figures S2). Such a wide range of RH's SHAP values indicates its complicated role in contributing to a HE (e.g., aqueous-phase chemistry, hygroscopic growth, and/or wet removal).[34,45−47] The contribution of RH was positively correlated with that of BLH and negatively correlated with that of SSR (Figure S3), suggesting that RH acted as aqueous-phase chemistry that coincided mostly with dark conditions. Water vapor was less available when the RH was below ~50−60% like during HE1. As opposed to the predominant role of photochemistry in forming HE1, the enhanced $PM_{2.5}$ driven by RH in the remaining HEs was more likely related to aqueous-phase chemistry.

Surface temperature is another important factor, accounting for 14−31 μg m^{-3} $PM_{2.5}$ across the five HEs. These strong positive contributions occurred only when the temperature was in the range of ~0 ± 8 °C, which coincided with ≳60% RH (Figure S6).[48] A low temperature together with a high RH would enhance the pollution. The major fractions of $PM_{2.5}$ during HE2 were nitrate and sulfate.[37] HE2 was largely driven by factors associated with temperature, RH, and O_X, implying the importance of both aqueous processing and photochemistry in increasing $PM_{2.5}$. The air masses were relatively short in length (Figure S4b), bringing emissions from surroundings to the receptor site,[37] thus leading to the second and fourth increases in $PM_{2.5}$ concentration during HE2 (Figure 1c). The final increase in $PM_{2.5}$ before the end of HE3 was also due to strong regional transport (Figure 1e), as confirmed by Dai et al.,[37] who showed that the increment of $PM_{2.5}$ was carried by air flows from the upwind polluted areas. It was found that the meteorological condition during the CSF was worse than that of the previous year,[33,34] which was further verified by Xue et al.,[49] who identified that unfavorable weather was an important cause of haze pollution in northern China during early 2020 on the basis of WRF-CMAQ modeling. In addition to the enhanced formation of secondary aerosol during HE4,[43,45] festival-related emissions also facilitated the formation of hazes.[15,36] It was estimated that festival-related emissions contributed ~29 and ~16 μg m^{-3} $PM_{2.5}$ on the first lunar day (January 25) in Beijing and Tianjin, respectively,[15] resulting in the maximum primary emissions being recorded on January 25 (Figure S1). Similar to the cases for HE3 and HE4, $PM_{2.5}$ in HE5 was also largely governed by RH (23 ± 15 μg m^{-3}). The contribution of regional transport was relatively weak (~11 ± 6 μg m^{-3}). Overall, wintertime HEs were generally driven by a combination of shallow BLH, weak winds, low temperatures, and high RH (Figure 2 and Figure S7). Details of the relationships between variables are presented in Figure 2, Figure S3, and Text S3.

Mechanisms of the Formation of Nitrate and Sulfate during the Haze Events. The formation of nitrate and sulfate was enhanced during the lockdown periods (HE4 and HE5),[33−35,45,49,50] which would be related to chemistry-favorable meteorological conditions. Similar to previous studies,[44,51,52] photochemistry played an important role in forming daytime nitrate, as O_X explained 10 ± 6, 6 ± 5, 4 ± 4, 6 ± 4, and 3 ± 3 μg m^{-3} nitrate on average during HE1−HE5, respectively (Figure S5). A number of pieces of evidence prove that the O_X-driven daytime nitrate resulted from photochemistry rather than its precursor (NO_x, part of the O_X composition), particularly during the lockdown period. First, there was no significant correlation between the time series of O_X and NO_2. Second, O_X was positively correlated with nitrate (r^2 = 0.22), whereas there was no discernible correlation between NO_2 and nitrate. In addition, O_X explained the majority of nitrate in the daytime but not the nighttime. The RH-driven nitrate levels during HE1−HE5 were 2 ± 1, 6 ± 4, 6 ± 3, 5 ± 3, and 4 ± 3 μg m^{-3}, respectively. The low temperature and high-RH conditions were conducive to nitrate formation because they enhanced the N_2O_5 heterogeneous hydrolysis reaction and strengthened the partitioning of gas into particles.[45−47,52,53]

The major pathway of formation of sulfate was different from that of nitrate. The aqueous chemistry was likely the dominant pathway of formation for sulfate because RH accounted for 3 ± 2, 5 ± 4, 6 ± 2, 5 ± 3, and 4 ± 3 μg m^{-3} sulfate in HE1−HE5, respectively (Figure S6). Interestingly, a decrease in temperature (<0 °C) for >60% RH promoted sulfate formation (Figure S12c). It makes sense because a low temperature enhanced SO_2 dissolution,[46] resulting in a relatively high initial aerosol pH that in turn accelerated the aqueous oxidation rate.[54] Photochemistry was less important for sulfate formation (3 ± 2, 3 ± 3, 2 ± 1, 2 ± 2, and 2 ± 1 μg m^{-3} for HE1−HE5) as opposed to its dominant contribution to nitrate. Although it is unlikely to have strong photochemical activity in the winter, photochemistry still played an important role in the formation of secondary particles,[44] especially for nitrate.[55−57] The maximum SHAP values of O_X for nitrate and sulfate reached 25 and 18 μg m^{-3}, respectively (Figures S5 and S6). Here we assumed that sulfate was totally from secondary formation; however, recent studies found that residential coal combustion was a major source of sulfate in northern China.[58,59] Regional transport has a stronger impact on nitrate and sulfate than the total $PM_{2.5}$ mass (Figure 1 and Figures S5 and S6), as secondary particles are more regional than local in nature. Previous studies also demonstrated that these large-scale severe air pollution events (herein HE3 and HE4) were associated with strong regional transport of nitrate and sulfate.[43,60]

Our results are consistent with previous publications,[31−35,37,44,49] further verifying the existing conclusions and implying the feasibility of ML in interpreting haze formation. ML-based deweathering coupled with SHAP analysis seems to be an improvement and complementary to current CTMs. The results should be interpreted recognizing the limitations as detailed in Text S4.

https://doi.org/10.1021/acs.estlett.1c00865
Environ. Sci. Technol. Lett. XXXX, XXX, XXX−XXX

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.estlett.1c00865.

Random Forest modeling and deweathering (Text S1); calculation and discussion of the pairwise interactive effects between major features (Texts S2 and S3); limitations of the tree-based ML feature attribution (Text S4); improvement of the RF model performance by adding new variables (Table S1); time series of the observed and meteorologically normalized SO_2, NO_2, CO, and $PM_{2.5}$ (Figure S1); diurnal variations of $PM_{2.5,emission}$ and SHAP values of RH and O_X for $PM_{2.5,meteo}$ during the five HEs (Figure S2); correlation analysis of the variable's SHAP values (Figure S3); time series of the observed air temperature, relatively humidity, air trajectory length, and O_X during the study period (Figure S4); SHAP values from the RF model of nitrate and sulfate (Figures S5 and S6); SHAP dependence plots of $PM_{2.5,\ meteo}$, nitrate, and sulfate from Random Forest models (Figures S7−S9); main and interaction effects of major variables on nitrate and sulfate (Figures S10 and S12); summary plots of the SHAP interaction matrix values for nitrate and sulfate (Figures S11 and S13); and comparison of the feature importance of $PM_{2.5,meteo}$ estimated via the RF model and SHAP algorithm (Figure S14) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Qili Dai − *State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China; CMA-NKU Cooperative Laboratory for Atmospheric Environment-Health Research, Tianjin 300350, China; Center for Urban Transport Emission Research, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China;* orcid.org/0000-0001-9534-2887; Email: daiql@nankai.edu.cn

Authors

Linlu Hou − *State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China; CMA-NKU Cooperative Laboratory for Atmospheric Environment-Health Research, Tianjin 300350, China*

Congbo Song − *School of Geography, Earth and Environment Sciences, University of Birmingham, Birmingham B15 2TT, U.K.;* orcid.org/0000-0001-7948-4834

Bowen Liu − *Department of Economics, University of Birmingham, Birmingham B15 2TT, U.K.*

Fangzhou Guo − *Department of Civil and Environmental Engineering, Rice University, Houston, Texas 77005-1827, United States;* orcid.org/0000-0003-3854-038X

Tianjiao Dai − *State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China; CMA-NKU Cooperative Laboratory for Atmospheric Environment-Health Research, Tianjin 300350, China*

Linxuan Li − *State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China; CMA-NKU Cooperative Laboratory for Atmospheric Environment-Health Research, Tianjin 300350, China*

Baoshuang Liu − *State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China; CMA-NKU Cooperative Laboratory for Atmospheric Environment-Health Research, Tianjin 300350, China; Center for Urban Transport Emission Research, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China*

Xiaohui Bi − *State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China; CMA-NKU Cooperative Laboratory for Atmospheric Environment-Health Research, Tianjin 300350, China; Center for Urban Transport Emission Research, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China*

Yufen Zhang − *State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China; CMA-NKU Cooperative Laboratory for Atmospheric Environment-Health Research, Tianjin 300350, China; Center for Urban Transport Emission Research, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China*

Yinchang Feng − *State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China; CMA-NKU Cooperative Laboratory for Atmospheric Environment-Health Research, Tianjin 300350, China; Center for Urban Transport Emission Research, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China;* orcid.org/0000-0002-6014-5258

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.estlett.1c00865

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was financially supported by the National Natural Science Foundation of China (42177085), the Tianjin Science and Technology Plan Project (PTZWHZ00120), the Tianjin Research Institute for the Development Strategy of China's Engineering Science and Technology (2020C0-0002), and the National Key R&D Program of China (2016YFC0208505).

■ REFERENCES

(1) Morawska, L.; Zhu, T.; Liu, N.; Amouei Torkmahalleh, M.; de Fatima Andrade, M.; Barratt, B.; Broomandi, P.; Buonanno, G.; Carlos Belalcazar Ceron, L.; Chen, J.; Cheng, Y.; Evans, G.; Gavidia, M.; Guo, H.; Hanigan, I.; Hu, M.; Jeong, C. H.; Kelly, F.; Gallardo, L.; Kumar, P.; Lyu, X.; Mullins, B. J.; Nordstrøm, C.; Pereira, G.; Querol,

X.; Yezid Rojas Roa, N.; Russell, A.; Thompson, H.; Wang, H.; Wang, L.; Wang, T.; Wierzbicka, A.; Xue, T.; Ye, C. The state of science on severe air pollution episodes: Quantitative and qualitative analysis. *Environ. Int.* **2021**, *156*, 106732.

(2) Burnett, R.; Chen, H.; Szyszkowicz, M.; Fann, N.; Hubbell, B.; et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (38), 9592−9597.

(3) McDuffie, E. E.; Martin, R. V.; Spadaro, J. V.; Burnett, R.; Smith, S. J.; O'Rourke, P.; Hammer, M. S.; van Donkelaar, A.; Bindle, L.; Shah, V.; Jaeglé, L.; Luo, G.; Yu, F.; Adeniran, J. A.; Lin, J.; Brauer, M. Source sector and fuel contributions to ambient $PM_{2.5}$ and attributable mortality across multiple spatial scales. *Nat. Commun.* **2021**, *12*, 3594.

(4) Zhang, X.; Xu, X.; Ding, Y.; Liu, Y.; Zhang, H.; Wang, Y.; Zhong, J. The impact of meteorological changes from 2013 to 2017 on $PM_{2.5}$ mass reduction in key regions in China. *Sci. China. Earth. Sci.* **2019**, *62* (12), 1885−1902.

(5) Zhai, S.; Jacob, D. J.; Wang, X.; Shen, L.; Li, K.; Zhang, Y.; Gui, K.; Zhao, T.; Liao, H. Fine Particulate Matter ($PM_{2.5}$) Trends In China, 2013−2018: Separating contributions from anthropogenic emissions and meteorology. *Atmos. Chem. Phys.* **2019**, *19* (16), 11031−11041.

(6) Li, Z.; Guo, J.; Ding, A.; Liao, H.; Liu, J.; Sun, Y.; Wang, T.; Xue, H.; Zhang, H.; Zhu, B. Aerosol and boundary-layer interactions and impact on air quality. *Natl. Sci. Rev.* **2017**, *4*, 810−833.

(7) 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. Exploring the severe winter haze in Beijing: the impact of synoptic weather, regional transport and heterogeneous reactions. *Atmos. Chem. Phys.* **2015**, *15* (6), 2969−2983.

(8) Lin, P.; Hu, M.; Deng, Z.; Slanina, J.; Han, S.; Kondo, Y.; Takegawa, N.; Miyazaki, Y.; Zhao, Y.; Sugimoto, N. Seasonal and diurnal variations of organic carbon in $PM_{2.5}$ in Beijing and the estimation of secondary organic carbon. *J. Geophys. Res.* **2009**, *114* (D2), D00G11.

(9) Wang, G.; Zhang, R.; Gomez, M. E.; Yang, L.; Levy Zamora, M.; Hu, M.; Lin, Y.; Peng, J.; Guo, S.; Meng, J.; Li, J.; Cheng, C.; Hu, T.; Ren, Y.; Wang, Y.; Gao, J.; Cao, J.; An, Z.; Zhou, W.; Li, G.; Wang, J.; Tian, P.; Marrero-Ortiz, W.; Secrest, J.; Du, Z.; Zheng, J.; Shang, D.; Zeng, L.; Shao, M.; Wang, W.; Huang, Y.; Wang, Y.; Zhu, Y.; Li, Y.; Hu, J.; Pan, B.; Cai, L.; Cheng, Y.; Ji, Y.; Zhang, F.; Rosenfeld, D.; Liss, P. S.; Duce, R. A.; Kolb, C. E.; Molina, M. J. Persistent sulfate formation from London Fog to Chinese haze. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (48), 13630−13635.

(10) 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. Drivers of improved $PM_{2.5}$ air quality in China from 2013 to 2017. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (49), 24463−24469.

(11) Chen, Z.; Chen, D.; Kwan, M.-P.; Chen, B.; Gao, B.; Zhuang, Y.; Li, R.; Xu, B. The control of anthropogenic emissions contributed to 80% of the decrease in $PM_{2.5}$ concentrations in Beijing from 2013 to 2017. *Atmos. Chem. Phys.* **2019**, *19* (21), 13519.

(12) Xiao, Q.; Zheng, Y.; Geng, G.; Chen, C.; Huang, X.; Che, H.; Zhang, X.; He, K.; Zhang, Q. Separating emission and meteorological contributions to long-term $PM_{2.5}$ trends over eastern China during 2000−2018. *Atmos. Chem. Phys.* **2021**, *21* (12), 9475−9496.

(13) Zhao, B.; Wu, W.; Wang, S.; Xing, J.; Chang, X.; Liou, K.-N.; Jiang, J. H.; Gu, Y.; Jang, C.; Fu, J. S.; Zhu, Y.; Wang, J.; Lin, Y.; Hao, J. A modeling study of the nonlinear response of fine particles to air pollutant emissions in the Beijing−Tianjin−Hebei region. *Atmos. Chem. Phys.* **2017**, *17* (19), 12031−12050.

(14) Grange, S. K.; Carslaw, D. C.; Lewis, A. C.; Boleti, E.; Hueglin, C. Random forest meteorological normalisation models for Swiss PM_{10} trend analysis. *Atmos. Chem. Phys.* **2018**, *18*, 6223.

(15) Dai, Q.; Hou, L.; Liu, B.; Zhang, Y.; Song, C.; Shi, Z.; Hopke, P. K.; Feng, Y. Spring Festival and COVID-19 Lockdown: Disentangling PM Sources in Major Chinese Cities. *Geophys. Res. Lett.* **2021**, *48* (11), e2021GL093403.

(16) Shi, Z.; Song, C.; Liu, B.; Lu, G.; Xu, J.; Van Vu, T.; Elliott, R. J. R.; Li, W.; Bloss, J. W.; Harrison, M. R. Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns. *Sci. Adv.* **2021**, *7* (3), eabd6696.

(17) Vu, T. V.; Shi, Z.; Cheng, J.; Zhang, Q.; He, K.; Wang, S.; Harrison, R. M. Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique. *Atmos. Chem. Phys.* **2019**, *19* (17), 11303.

(18) Yang, J.; Wen, Y.; Wang, Y.; Zhang, S.; Pinto, J. P.; Pennington, E. A.; Wang, Z.; Wu, Y.; Sander, S. P.; Jiang, J. H.; Hao, J.; Yung, Y. L.; Seinfeld, J. H. From COVID-19 to Future Electrification: Assessing Traffic Impacts on Air Quality by a Machine Learning Model. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, e2102705118.

(19) Auret, L.; Aldrich, C. Empirical comparison of tree ensemble variable importance measures. *Chemometr. Intell. Lab. Syst.* **2011**, *105* (2), 157−170.

(20) Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.; Samek, W. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLoS One* **2015**, *10* (7), No. e0130140.

(21) Štrumbelj, E.; Kononenko, I. Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.* **2014**, *41*, 647−665.

(22) Ribeiro, M. T.; Singh, S.; Guestrin, C. "Why Should I Trust You?" Explaining the Predictions of Any Classifier. *ACM* **2016**, 1135−1144.

(23) Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, 1705.07874.

(24) Mangalathu, S.; Hwang, S.-H.; Jeon, J.-S. Failure mode and effects analysis of RC members based on machine-learning-based Shapley additive explanations (SHAP) approach. *Eng. Struct.* **2020**, *219*, 110927.

(25) Lundberg, S. M.; Nair, B.; Vavilala, M. S.; Horibe, M.; Eisses, M. J.; Adams, T.; Liston, D. E.; Low, D. K.-W.; Newman, S.-F.; Kim, J.; Lee, S.-I. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2018**, *2*, 749−760.

(26) Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56−67.

(27) Zhong, S.; Zhang, K.; Wang, D.; Zhang, H. Shedding light on "Black Box" machine learning models for predicting the reactivity of HOradical dot radicals toward organic compounds. *Chem. Eng. J.* **2021**, *405*, 126627.

(28) Wang, C.; Feng, L.; Qi, Y. Explainable deep learning predictions for illness risk of mental disorders in Nanjing, China. *Environ. Res.* **2021**, *202*, 111740.

(29) Stirnberg, R.; Cermak, J.; Kotthaus, S.; Haeffelin, M.; Andersen, H.; Fuchs, J.; Kim, M.; Petit, J.-E.; Favez, O. Meteorology-driven variability of air pollution (PM_1) revealed with explainable machine learning. *Atmos. Chem. Phys.* **2021**, *21* (5), 3919.

(30) Gu, J.; Yang, B.; Brauer, M.; Zhang, K. M. Enhancing the Evaluation and Interpretability of Data-Driven Air Quality Models. *Atmos. Environ.* **2021**, *246*, 118125.

(31) Chang, Y.; Huang, R. J.; Ge, X.; Huang, X.; Hu, J.; Duan, Y.; Zou, Z.; Liu, X.; Lehmann, M. F. Puzzling haze events in China during the Coronavirus (COVID-19) shutdown. *Geophys. Res. Lett.* **2020**, *47* (12), e2020GL088533.

(32) Lv, Z.; Wang, X.; Deng, F.; Ying, Q.; Archibald, T. A.; Jones, L. R.; Ding, Y.; Cheng, Y.; Fu, M.; Liu, Y.; Man, H.; Xue, Z.; He, K.; Hao, J.; Liu, H. Source-Receptor Relationship Revealed by the Halted Traffic and Aggravated Haze in Beijing during the COVID-19 Lockdown. *Environ. Sci. Technol.* **2020**, *54* (24), 15660−15670.

https://doi.org/10.1021/acs.estlett.1c00865
Environ. Sci. Technol. Lett. XXXX, XXX, XXX−XXX

(33) Ding, J.; Dai, Q.; Li, Y.; Han, S.; Zhang, Y.; Feng, Y. Impact of meteorological condition changes on air quality and particulate chemical composition during the COVID-19 lockdown. *J. Environ. Sci. (China)* **2021**, *109*, 45−56.
(34) Le, T.; Wang, Y.; Liu, L.; Yang, J.; Yung, L. Y.; Li, G.; Seinfeld, H. J. Unexpected air pollution with marked emission reductions during the COVID-19 outbreak in China. *Science* **2020**, *369* (6504), 702−706.
(35) Huang, X.; Ding, A.; Gao, J.; Zheng, B.; Zhou, D.; Qi, X.; Tang, R.; Wang, J.; Ren, C.; Nie, W.; Chi, X.; Xu, Z.; Chen, L.; Li, Y.; Che, F.; Pang, N.; Wang, H.; Tong, D.; Qin, W.; Cheng, W.; Liu, W.; Fu, Q.; Liu, B.; Chai, F.; Davis, S. J.; Zhang, Q.; He, K. Enhanced secondary pollution offset reduction of primary emissions during COVID-19 lockdown in China. *Natl. Sci. Rev.* **2021**, *8* (2), nwaa137.
(36) Dai, Q.; Liu, B.; Bi, X.; Wu, J.; Liang, D.; Zhang, Y.; Feng, Y.; Hopke, P. K. Dispersion normalized PMF provides insights into the significant changes in source contributions to $PM_{2.5}$ after the COVID-19 outbreak. *Environ. Sci. Technol.* **2020**, *54* (16), 9917−9927.
(37) Dai, Q.; Ding, J.; Hou, L.; Li, L.; Cai, Z.; Liu, B.; Song, C.; Bi, X.; Wu, J.; Zhang, Y.; Feng, Y.; Hopke, P. K. Haze episodes before and during the COVID-19 shutdown in Tianjin, China: Contribution of fireworks and residential burning. *Environ. Pollut.* **2021**, *286*, 117252.
(38) Stein, A. F.; Draxler, R. R.; Rolph, G. D.; Stunder, B. J. B.; Cohen, M. D.; Ngan, F. NOAA's Hysplit atmospheric transport and dispersion modeling system. *Bull. Am. Meteorol Soc.* **2015**, *96* (12), 2059−2077.
(39) Grange, S. K.; Carslaw, D. C. Using meteorological normalisation to detect interventions in air quality time series. *Sci. Total Environ.* **2019**, *653*, 578−588.
(40) Dai, Q.; Schulze, B. C.; Bi, X.; Bui, A. A. T.; Guo, F.; Wallace, H. W.; Sanchez, N. P.; Flynn, J. H.; Lefer, B. L.; Feng, Y.; Griffin, R. J. Seasonal differences in formation processes of oxidized organic aerosol near Houston, TX. *Atmos. Chem. Phys.* **2019**, *19*, 9641−9661.
(41) Shapley, L. A Value for n-Person Games. In *Contributions to Theory Games (AM-28)*, Vol. II; Princeton University Press: Princeton, NJ, 1953; Vol. *28*, pp 307−318.
(42) Zheng, G.; Duan, F.; Ma, Y.; Zhang, Q.; Huang, T.; Kimoto, T.; Cheng, Y.; Su, H.; He, K. Episode-Based Evolution Pattern Analysis of Haze Pollution: Method Development and Results from Beijing, China. *Environ. Sci. Technol.* **2016**, *50* (9), 4632−4641.
(43) Sulaymon, I. D.; Zhang, Y.; Hopke, P. K.; Hu, J.; Zhang, Y.; Li, L.; Mei, X.; Gong, K.; Shi, Z.; Zhao, B.; Zhao, F. Persistent high $PM_{2.5}$ pollution driven by unfavorable meteorological conditions during the COVID-19 lockdown period in the Beijing-Tianjin-Hebei Region, China. *Environ. Res.* **2021**, *198*, 111186.
(44) Fu, X.; Wang, T.; Gao, J.; Wang, P.; Liu, Y.; Wang, S.; Zhao, B.; Xue, L. Persistent Heavy Winter Nitrate Pollution Driven by Increased Photochemical Oxidants in Northern China. *Environ. Sci. Technol.* **2020**, *54* (7), 3881−3889.
(45) Ding, J.; Dai, Q.; Zhang, Y.; Xu, J.; Huangfu, Y.; Feng, Y. Air humidity affects secondary aerosol formation in different pathways. *Sci. Total Environ.* **2021**, *759*, 143540.
(46) 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. High N_2O_5 concentrations observed in urban Beijing: Implications of a large nitrate formation pathway. *Environ. Sci. Technol. Lett.* **2017**, *4* (10), 416−420.
(47) Wu, Z.; Wang, Y.; Tan, T.; Zhu, Y.; Li, M.; Shang, D.; Wang, H.; Lu, K.; Guo, S.; Zeng, L.; Zhang, Y. Aerosol liquid water driven by anthropogenic inorganic salts: implying its key role in haze formation over the North China Plain. *Environ. Sci. Technol. Lett.* **2018**, *5* (3), 160−166.
(48) Shao, M.; Dai, Q.; Yu, Z.; Zhang, Y.; Xie, M.; Feng, Y. Responses in $PM_{2.5}$ and its chemical components to typical unfavorable meteorological events in the suburban area of Tianjin, China. *Sci. Total Environ.* **2021**, *788*, 147814.
(49) Xue, W.; Shi, X.; Yan, G.; Wang, J.; Xu, Y.; Tang, Q.; Wang, Y.; Zheng, Y.; Lei, Y. Impact of meteorological conditions and emission changes on heavy pollution processes in North China around the 2020 Spring Festival. *Sci. China: Earth Sci.* **2021**, *64*, 329.
(50) Sun, Y.; Lei, L.; Zhou, W.; Chen, C.; He, Y.; et al. A chemical cocktail during the COVID-19 outbreak in Beijing, China: Insights from six-year aerosol particle composition measurements during the Chinese New Year holiday. *Sci. Total Environ.* **2020**, *742*, 140739.
(51) Tian, J.; Wang, Q.; Zhang, Y.; Yan, M.; Liu, H.; Zhang, N.; Ran, W.; Cao, J. Impacts of primary emissions and secondary aerosol formation on air pollution in an urban area of China during the COVID-19 lockdown. *Environ. Int.* **2021**, *150*, 106426.
(52) Zhong, H.; Huang, R.-J.; Chang, Y.; Duan, J.; Lin, C.; Chen, Y. Enhanced formation of secondary organic aerosol from photochemical oxidation during the COVID-19 lockdown in a background site in Northwest China. *Sci. Total Environ.* **2021**, *778*, 144947.
(53) Sun, Y.; Wang, Z.; Fu, P.; Jiang, Q.; Yang, T.; Li, J.; Ge, X. The impact of relative humidity on aerosol composition and evolution processes during wintertime in Beijing, China. *Atmos. Environ.* **2013**, *77*, 927−934.
(54) Pye, H. O. T.; Nenes, A.; Alexander, B.; Ault, A. P.; Barth, M. C.; Clegg, S. L.; Collett, J. L., Jr; Fahey, K. M.; Hennigan, C. J.; Herrmann, H.; Kanakidou, M.; Kelly, J. T.; Ku, I. T.; McNeill, V. F.; Riemer, N.; Schaefer, T.; Shi, G.; Tilgner, A.; Walker, J. T.; Wang, T.; Weber, R.; Xing, J.; Zaveri, R. A.; Zuend, A. The acidity of atmospheric particles and clouds. *Atmos. Chem. Phys.* **2020**, *20*, 4809−4888.
(55) Fu, X.; Wang, T.; Gao, J.; Wang, P.; Liu, Y.; Wang, S.; Zhao, B.; Xue, L. Persistent heavy winter nitrate pollution driven by increased photochemical oxidants in Northern China. *Environ. Sci. Technol.* **2020**, *54* (7), 3881−3889.
(56) Chen, X.; Wang, H.; Lu, K.; Li, C.; Zhai, T.; Tan, Z.; Ma, X.; Yang, X.; Liu, Y.; Chen, S.; Dong, H.; Li, X.; Wu, Z.; Hu, M.; Zeng, L.; Zhang, Y. Field determination of nitrate formation pathway in winter Beijing. *Environ. Sci. Technol.* **2020**, *54* (15), 9243−9253.
(57) Lu, K.; Fuchs, H.; Hofzumahaus, A.; Tan, Z.; Wang, H.; Zhang, L.; Schmitt, S. H.; Rohrer, F.; Bohn, B.; Broch, S.; Dong, H.; Gkatzelis, G. I.; 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. Fast Photochemistry in wintertime haze: consequences for pollution mitigation strategies. *Environ. Sci. Technol.* **2019**, *53* (18), 10676−10684.
(58) Dai, Q.; Bi, X.; Song, W.; Li, T.; Liu, B.; Ding, J.; Xu, J.; Song, C.; Yang, N.; Schulze, B. C.; Zhang, Y.; Feng, Y.; Hopke, P. K. Residential coal combustion as a source of primary sulfate in Xi'an, China. *Atmos. Environ.* **2019**, *196*, 66−76.
(59) Hopke, P. K.; Dai, Q. Why it makes sense that increased $PM_{2.5}$ was correlated with anthropogenic combustion-derived water. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (19), No. e2102255118.
(60) Zheng, H.; Kong, S.; Chen, N.; Yan, Y.; Liu, D.; Zhu, B.; Xu, K.; Cao, W.; Ding, Q.; Lan, B.; Zhang, Z.; Zheng, M.; Fan, Z.; Cheng, Y.; Zheng, S.; Yao, L.; Bai, Y.; Zhao, i.; Qi, S. Significant changes in the chemical compositions and sources of $PM_{2.5}$ in Wuhan since the city lockdown as COVID-19. *Sci. Total Environ.* **2020**, *739*, 140000.

https://doi.org/10.1021/acs.estlett.1c00865
Environ. Sci. Technol. Lett. XXXX, XXX, XXX−XXX