https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

Atmospheric Measurement Techniques
Discussions

Ozone formation sensitivity study using machine learning coupled with the reactivity of VOC species

Junlei Zhan[1], Yongchun Liu[1*], Wei Ma[1], Xin Zhang[2], Xuezhong Wang[2], Fang Bi[2], Yujie Zhang[2], Zhenhai Wu[2], Hong Li[2*]

1. Aerosol and Haze Laboratory, Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China

2. State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China

Correspondence: liuyc@buct.edu.cn; lihong@craes.org.cn

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

Atmospheric Measurement Techniques Discussions

Abstract

The formation of ground-level ozone (O_3) is dependent on both atmospheric chemical processes and meteorological factors. Traditional models have difficulty assessing O_3 formation sensitivity in a timely manner due to the limitations of flexibility and computational efficiency. In this study, a random forest (RF) model coupled with the reactivity of volatile organic compound (VOC) species was used to investigate the O_3 formation sensitivity in Beijing from 2014 to 2016, and evaluate the relative importance (RI) of chemical and meteorological factors to O_3 formation. The results showed that the O_3 prediction performance using initial concentrations of VOC species ($R^2 = 0.87$) was better than that using total VOCs (TVOCs) concentrations ($R^2 = 0.77$). Meanwhile, the RIs of VOC species correlated well with their O_3 formation potentials (OFPs). O_3 formation presented a negative response to NOx, $PM_{2.5}$ and relative humidity, and a positive response to temperature, solar radiation and VOCs. The O_3 isopleth curves calculated by the RF model were generally comparable with those calculated by the box model. O_3 formation shifted from a VOC-limited regime to a transition regime from 2014 to 2016. This study demonstrates that the RF model coupled with the initial concentrations of VOC species could provide an accurate, flexible, and computationally efficient approach for O_3 sensitivity analysis.

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

1. Introduction

Ground-level ozone (O_3) pollution, which can cause adverse human health effects such as cardiovascular and respiratory diseases, has received increasing attention in recent decades (Cohen et al., 2017). As important precursors of O_3, volatile organic compounds (VOCs) in the atmosphere are oxidized to produce peroxyl radicals (RO_2) and hydroperoxyl radicals (HO_2), which will accelerate the NO-O_3-NO_2 cycle, thus leading to the accumulation of O_3 (Wang et al., 2017a). The production and loss of RO_2 and HO_2 are highly dependent on the concentration ratio of VOCs and NOx in the atmosphere. Hence, atmospheric O_3 concentrations or production rates show a nonlinear relationship with VOCs and NOx. Moreover, the O_3-VOC-NOx sensitivity is readily influenced by VOC species (Tan et al., 2018), meteorological parameters (Liu et al., 2020a; Liu & Wang 2020), and even atmospheric particulate matter (Li et al., 2019), thus, exhibits high temporal and spatial variability. Therefore, it is urgent to develop an accurate and highly efficient method for timely assessing the sensitivity regime of O_3 production and evaluating the effectiveness of a potential measure on O_3 pollution control.

The sensitivity of O_3 formation can usually be analysed using observed indicators, such as ozone production efficiency (OPE, $\Delta O_3/\Delta NOz$) (Wang et al., 2010; Lin et al., 2011), HCHO/NO_y (Martin et al., 2004), and H_2O_2/NOz (or H_2O_2/HNO_3) (Sillman 1995; Hammer et al., 2002; Wang et al., 2017a), observation-based model (OBM) (Vélez-Pereira et al., 2021) and chemical transport models including community multiscale air

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

quality (CMAQ) (Djalalova et al., 2015) and Weather Research and Forecasting with Chemistry (WRF-Chem) model (Wang et al., 2020a). The observed indicators can be utilized to quickly diagnose the sensitivity regime of O_3 production. However, the accuracy is sensitive to the precision of tracer measurements. In addition, this method lacks the predictability of O_3 concentrations for policy-making. OBMs combine *in-situ* field observations and chemical box models, which are built on widely-used chemistry mechanisms (e.g., MCM, Carbon Bond, RACM or SAPRC), and applied to the observed atmospheric conditions to simulate the *in-situ* O_3 production rate (Mo et al., 2018). The sensitivity of O_3 production to various O_3 precursors, including NOx and VOCs can be diagnosed based on the empirical kinetic modeling approach (EKMA) or quantitatively assessed with the relative incremental reactivity (RIR). Chemical transport models, which are driven by meteorological dynamics and incorporated with the emissions of pollutants and the complex atmospheric chemical mechanism, provide a powerful tool for simulating various atmospheric processes, including spatial distribution, regional transport *vs.* local formation, source apportionment and production rates of pollutants and so on (Sayeed et al., 2021). At present, OBMs are widely used to investigate O_3 formation sensitivity in China. Previous studies indicated that O_3 formation in urban areas of China is located in a VOC-limited or a transition regime and varies with time and location (Ou et al., 2016; Wang et al., 2017a; Zhan et al., 2021).

Although both OBMs and chemical transport models can assess the sensitivity of

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

Atmospheric Measurement Techniques Discussions

O_3 production and predict the O_3 pollution level in a scenario of control measures, the calculation accuracy is affected by the uncertainty of input parameters (Tang et al., 2011; Yang et al., 2021b). In addition, both of them are time-consuming and expensive when computational resources are considered. Thus, they are mostly applied to sampling cases with a short time span (days or weeks) (Xue et al., 2014; Ou et al., 2016), and identifying O_3 formation sensitivity in a timely manner is difficult. Compared to traditional methods, machine learning (ML) is able to capture the main factors affecting atmospheric O_3 formation in a timely manner with great flexibility (without the constraints of time and space) and high computational efficiency (Wang et al., 2020c; Grange et al., 2021; Yang et al., 2021a). Recently, ML based on convolutional neural network (CNN), random forest (RF) and artificial neural network (ANN) models has been applied in simulating atmospheric O_3 and shown good performance in O_3 prediction (Ma et al., 2020; Xing et al., 2020). For example, Ma et al. (Ma et al., 2021a) simulated O_3 concentrations in the Beijing-Tianjin-Hebei (BTH) region from 2010-2017 using an RF model that considered meteorological variables and output variables from chemical transport models, and the correlation coefficient (R^2) between the observed and modelled O_3 concentrations was greater than 0.8. Liu et al. (Liu et al., 2021) also reported a high accuracy (80.4%) for classifying pollution levels of O_3 and $PM_{2.5}$ at 1464 monitoring sites in China using an RF model. According to these previous studies, the RF model has shown good performance in terms of prediction accuracy and computational efficiency (Wang et al., 2016; Wang et al., 2017b).

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

However, many ML studies have used total VOCs (TVOCs) to simulate O_3 formation and rarely considered the effect of VOC species on O_3 formation sensitivity (Feng et al., 2019; Liu et al., 2021; Ma et al., 2021a). Thus, they were unable to identify the chemical reactivity of a single species to O_3 formation, which may lead to underestimations or even misunderstandings of the role of VOCs in O_3 formation because the same concentration of TVOCs with different compositions may lead to different OPEs. In addition, VOCs react with OH radicals during atmospheric transport, which is the most important sink of VOCs (Carlo et al., 2004; Liu et al., 2020b). Makar et al. (Makar et al., 1999) reported that highly reactive species, such as isoprene, were underestimated by 40% when the OH reactions were ignored. Other studies indicated that the initial concentrations of VOCs, which account for the photochemical loss of VOCs during transport, were more representative of pollution levels in the sampling area than the observed VOCs (Yuan et al., 2013; Zhan et al., 2021). However, whether the ML model can identify the connection between the reactivity of VOC species and O_3 formation sensitivity has not been clarified.

In this study, we used the RF model to evaluate the prediction performance of atmospheric O_3 using the TVOCs, measured VOC species and photochemical initial concentration (PIC) of VOC species. We compared the relative importance (RI) of the precursors (VOC species, NOx, $PM_{2.5}$, CO) and the meteorological parameters (temperature, solar radiation, relative humidity, wind speed and direction) on O_3 formation in the summer of Beijing from 2014 to 2016. We also discussed the

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

possibility of connecting the RIs of VOCs with their OFPs and the changes in O_3-VOC-NOx sensitivity based on the RF model from 2014 to 2016. Our study indicates that the RF model combined with initial concentrations of VOC species can simulate O_3 concentrations well and provides a flexible and efficient tool for O_3 modelling in a near real-time way.

2. Methods

2.1 Sampling site and data

The sampling site (40.04°N, 116.42°E) is located at the campus of Chinese Research Academy of Environmental Sciences and was described in our previous work (Zhang et al., 2021). Briefly, the station is located two kilometers from the north 4th ring road and surrounded by a mixed residential and commercial area. The concentrations of VOCs, NOx, CO, O_3 and $PM_{2.5}$ were measured at 8 m above ground level at this location. Meteorological parameters, including temperature (T), relative humidity (RH), wind speed and direction (WS&WD), solar radiation (SR), were monitored at 15 m above ground level. VOCs were measured by an online commercial instrument (GC-866, Chromatotec, France), which consisted of two independent analysers for detecting C2-C6 and C6-C12 hydrocarbon components. More details about the observations can be found in the Supplemental Materials (S1). The PICs of VOCs were calculated according to the method reported in our previous work (Zhan et al., 2021) and the Supplemental Materials (S2).

2.2 Random forest model

The random forest (RF) is a type of decision tree that can be used for classification

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

and regression (Breiman 2001). During the training process, the model creates a large number of different decision trees with different sample sets at each node, and then averages the scores of each decision tree as its final score to obtain more accurate results that avoid large bias and overfitting (Breiman 2001). Approximately one-third of the samples are excluded from the sample when the decision tree is built and used to calculate the out-of-bag data error. Hence, RF can evaluate the RI of variables via out-of-bag (OOB) data error (Svetnik et al., 2003),

$$RI_i = \sum (errOOB2_i - errOOB1_i)/N \quad (1)$$

where N represents the number of decision trees, and errOOB1 and OOB2 represent the out-of-bag data error of feature *i* before and after adding tiny data noise (Kohavi & John 1997; Breiman 2001), respectively. The RI_i reflects the response of the RF model to feature *i* after adding tiny data noise. It was used to evaluate the importance and sensitivity of feature *i* to O_3 formation in this study. More details about RI can be found in the Supplemental Materials (S3). To verify the stability of the model, we interrupted the continuity of the time series, fed the randomly arranged inputs to the model, and performed a significance test on the RI. The results showed that there was no significant difference among the different tests (P>0.05, R^2>0.97).

3. Results and discussion

3.1 Overview of air pollutants and meteorological conditions

Figure 1 shows the time series of air pollutants and meteorological parameters during the observations from 2014 to 2016. In 2014, 2015 and 2016, the wind direction

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

Atmospheric Measurement Techniques Discussions

was dominated by northwest winds (Figure S1), with mean wind speeds of 3.1 ± 2.7 m s^{-1}, 2.3 ± 2.2 m s^{-1}, and 1.3 ± 1.2 m s^{-1}, respectively, and the mean daytime temperature were 22.3 ± 5.8, 23.9 ± 5.0 and 24.0 ± 4.4 °C, respectively. The average value of SR decreased from 162.9 to 150.8 W m^{-2} during the observation period. As shown in Figure 1F-G, in 2014, 2015 and 2016, the mean VOC concentrations were 20.3 ± 10.9, 15.8 ± 8.3 and 12.1 ± 7.7 ppbv, respectively, while the mean initial VOC concentrations were 28.1 ± 25.7, 27.2 ± 32.6 and 16.4 ± 16.1 ppbv, respectively. Both the measured VOCs and initial VOCs showed a decline along with a decrease in $PM_{2.5}$ concentration from 67.2 ± 53.5 to 61.1 ± 48.6 μg m^{-3} due to the Air Pollution Prevention and Control Action Plan in China (Zhao et al., 2021). However, O_3 concentrations showed a slight upward trend from 38.7 ± 33.4 to 42.7 ± 27.9 ppbv from 2014 to 2015 and then to 44.0 ± 29.6 ppbv in 2016. A similar trend was observed for NOx concentrations (Figure S2). As shown in Figure 1F-G, the concentrations of four types (alkanes, alkenes, alkynes, and aromatics) of VOCs showed significant differences from 2014 to 2016 due to the variations in emission sources (Zhang et al., 2021). In addition to VOC species, the variations in other parameters, such as meteorological conditions and $PM_{2.5}$, should have a complex influence on O_3-VOC-NOx sensitivity (Li et al., 2019; Ma et al., 2021b).

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

Atmospheric Measurement Techniques Discussions

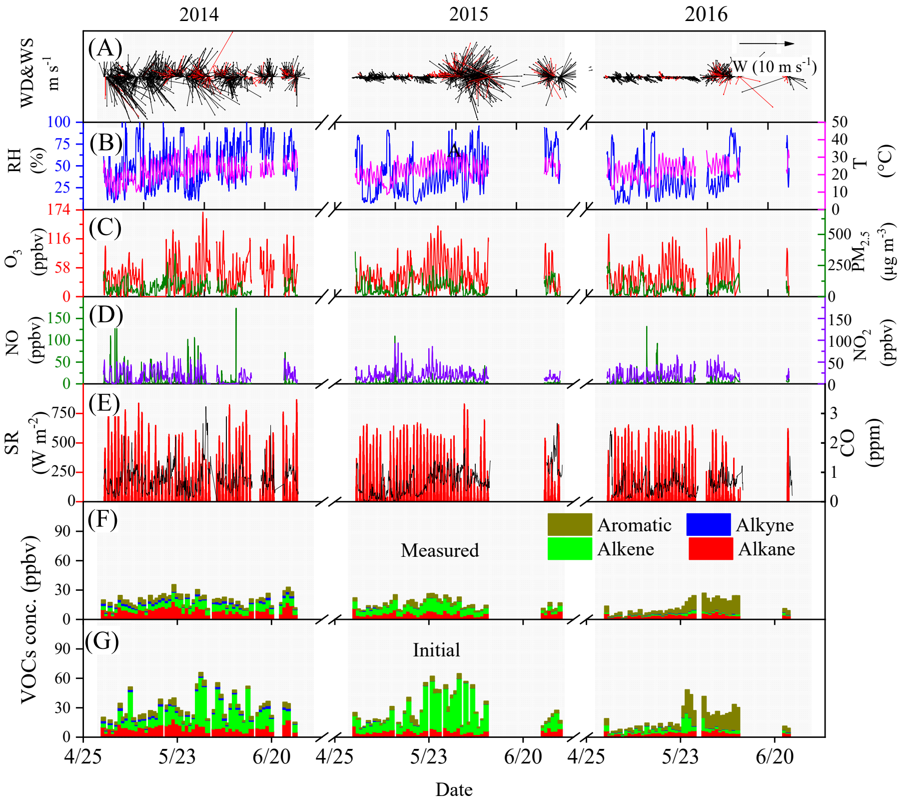

Figure 1. Time series of air pollutants and meteorological parameters during observations in Beijing.

3.2 Prediction performance of the model.

To build a robust model, we evaluated the prediction performance of the RF model for the ambient O_3 simulation. Figure 2 shows the O_3 prediction performance when chemical species (including VOCs, NOx, $PM_{2.5}$, CO) and meteorological factors (i.e., WS, WD, SR, T and RH) were used as inputs in the RF model. The details of the modelling and input parameters are shown in Table S1. Figure 2A-C shows the time series of the measured and modelled O_3 concentrations, which were simulated using

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

the TVOCs, measured VOC species and initial VOC species as input variables along with the same set of other parameters. The correlation coefficients (R^2) of the training data were 0.88, 0.94 and 0.94 for the TVOCs, measured VOC species and initial VOC species, respectively. The corresponding root mean squared errors (RMSEs) for the predicted O_3 concentrations were 9.9, 9.3 and 9.1. Figure 2D-F shows the prediction performance of the testing dataset under these three circumstances. When the TVOCs were split into VOC species, the R^2 increased from 0.77 to 0.86 as the number of data features increased. Therefore, the VOC composition has a significant influence on O_3 prediction using the RF model. Thus, our model has good prediction performance (R^2 = 0.87) when combined with the initial VOC species. In previous studies using TVOCs, the influence of VOC composition was neglected (Liu et al., 2021; Ma et al., 2021a). Therefore, our results indicate that the RF model can accurately predict O_3 concentrations when the concentrations of VOC species are considered and identify the connection between the reactivity of VOC species and O_3 formation in the atmosphere.

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

Atmospheric Measurement Techniques Discussions

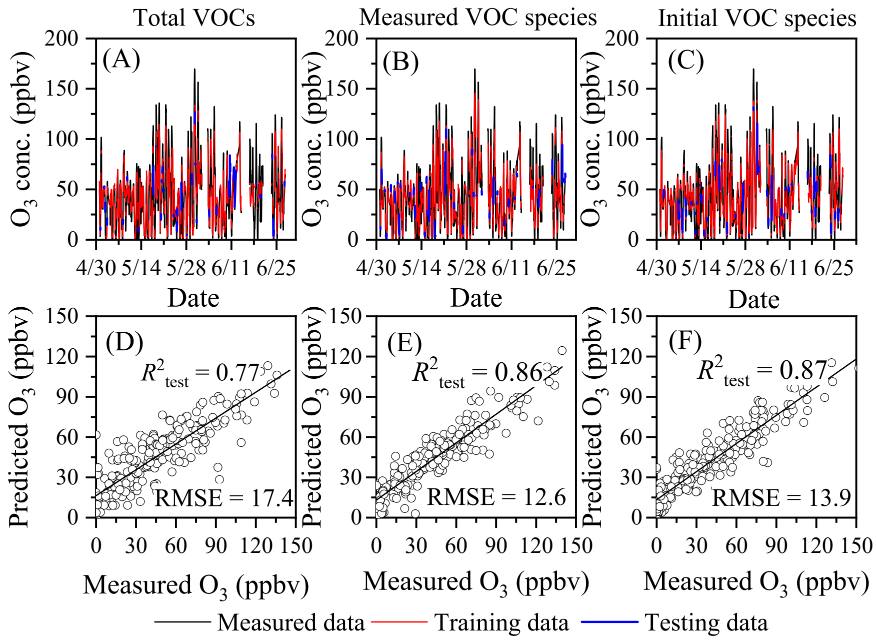

Figure 2. Comparison of the predicted and measured O_3 concentrations in Beijing in the summer of 2014. (A and D: TVOC concentrations; B and E: measured concentrations of VOC species; C and F: initial concentrations of VOC species)

3.3 Relative importance of major factors

Figure 3A shows the RIs of different ambient factors, including chemical and meteorological variables on O_3 formation. The difference in the RIs is also compared using the TVOCs and the VOC species as inputs. Chemical factors (including VOC species, NOx, $PM_{2.5}$ and CO) accounted for 83.1% of the contribution to O_3 production in the summer of 2016. Meanwhile, VOC species accounted for approximately 66.7% of O_3 production while the RIs using TVOC concentrations accounted for only 6.5%. Ma et al. (Ma et al., 2021b) analysed the contribution of meteorological conditions and chemical factors to O_3 formation on the North China Plain (NCP) using the CMAQ

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

model in combination with process analysis and found that chemical factors dominate O_3 formation in summer. Using probability theory, Ueno et al. (Ueno & Tsunematsu 2019) also found that VOCs/NOx dominate O_3 production compared to meteorological variables. Thus, our results are similar to those of previous studies based on chemical models (Ueno & Tsunematsu 2019; Ma et al., 2021b), which demonstrates that the RF model can reflect the contribution of VOC species to O_3 production even if the observed VOC species are used.

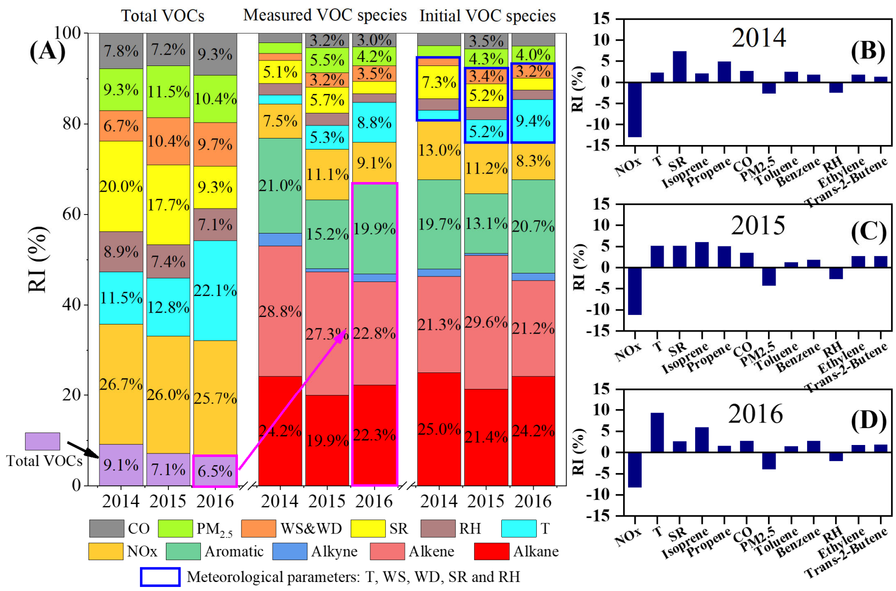

Figure 3. Percentage of RI for O_3 precursors and meteorological parameters (A) and the top 12 factors with high values of RI in 2014-2016 (B-D: using initial concentrations of VOC species).

Although ML is widely used to understand air pollution, explanations of ML results (e.g., RI) are somewhat vague because ML is a black-box model (Sayeed et al., 2021). Here, we compared the RIs of VOCs calculated using the initial VOC species and the observed VOC species with the O_3 formation potentials (OFPs). The OFPs were calculated by the maximum incremental reactivity (MIR) method (Carter 2010). As shown in Figure S3, the RIs showed good correlations with the OFP. Interestingly, the initial concentrations of VOC species improved the correlation coefficients between the RIs and OFPs. Furthermore, we calculated the RIs and OFPs of different species using the observed data during the campaign study in Daxing District in the summer of 2019 (Zhan et al., 2021), and a strong correlation was observed between the RIs of the initial VOC species and the OFPs (Figure S4). These results indicate that the RIs of the initial VOCs species in the ML model should partially reflect the chemical reactivity of VOCs to produce O_3 in the atmosphere.

Although the RIs calculated using the initial VOC species slightly changed compared to those calculated using the observed VOCs (Table S2), VOCs still dominated O_3 formation (Figure 3A). For example, the initial VOCs dominated O_3 production in 2014, 2015, and 2016, with RI values of 67.7, 64.5 and 67.7% respectively. Li et al. (Li et al., 2020a) used a multiple linear regression (MLR) model to study the contribution of anthropogenic and meteorological factors to O_3 formation in China from 2013-2019 and found that meteorological factors accounted for 36.8% and anthropogenic factors accounted for 63.2%, which is similar to our results. Figure 3B-D shows the top 12 factors having a strongly influence on O_3 production. Interestingly, NOx, $PM_{2.5}$ and RH showed negative responses to O_3 formation, while other variables, including T, SR, CO and all of the VOCs, showed positive responses.

244 Thus, a decrease in NOx, $PM_{2.5}$ or RH will lead to an increase in O_3 concentration while
245 a decrease in T, SR, CO and VOCs will lead to a decrease in O_3 concentration. Although
246 O_3 formation is highly related to the photolysis of NO_2, a previous study demonstrated
247 that it is VOC-limited in summer in Beijing (Zhan et al., 2021). This finding is
248 consistent with the observed negative response of O_3 to NOx in this work. High
249 concentrations of $PM_{2.5}$ can reduce solar radiation and increase the sinks of reactive
250 radicals (HOx and ROx) (Li et al., 2019). In addition, high RH usually coincides with
251 low surface O_3 concentrations in field observations, which can be ascribed to the
252 inhibition of O_3 formation by the transfer of NO_2/ONO_2-containing products into the
253 particle phase and the promotion of dry deposition of O_3 on the surface (Kavassalis &
254 Murphy 2017; Yu 2019). These previous works can well explain the observed negative
255 response of O_3 to $PM_{2.5}$ and RH in Figure 3B. Previous studies have observed a positive
256 correlation between the O_3 concentration and T or SR (Steiner et al., 2010; Paraschiv
257 et al., 2020; Li et al., 2021). Temperature can directly affect the chemical reaction rate
258 of O_3 formation (Fu et al., 2015), and SR can promote the photolysis of NO_2 (Hu et al.,
259 2017; Wang et al., 2020b), thus accelerating O_3 formation. As mentioned above, O_3
260 formation is VOC-limited in Beijing; thus, a positive response of O_3 concentration to
261 VOCs is observed in Figure 3B. Interestingly, the RIs of isoprene showed an increasing
262 trend from 2014 to 2016 because of the obvious reduction in anthropogenic VOCs
263 (Figure 1) (Zhang et al., 2021). In the context of global warming, studies should focus
264 on the factors that affect O_3 formation, including biogenic emissions, T and SR. Thus,

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

additional efforts will be required to reduce anthropogenic pollutants in the future.

3.4 Ozone formation sensitivity

To further analyse the sensitivity of O_3 to VOCs and NOx from 2014 to 2016, we plotted sensitivity curves for O_3 generation using the RF model, and the results are shown in Figure 4A-C. Moreover, EKMA curves in 2015 were also obtained using the OBM (Figure 4D). As shown in Figure 4A-C, O_3 formation was sensitive to VOCs in the summer of Beijing during our observations, which is consistent with previous studies that used box models (Li et al., 2020b) and chemical transport models (Shao et al., 2021). This result is also consistent with the RIs of VOCs or NOx to O_3 formation (Figure 3B). Interestingly, the O_3 formation sensitivity to VOCs decreases or gradually shifts from the observed point to the transition regime from 2014 to 2016 (Figure 4A-C), which is similar to that reported by Zhang et al. (Zhang et al., 2021). These phenomena can be ascribed to the increased importance of meteorological factors, such as T, SR, and RH, for O_3 formation and the variation in anthropogenic VOC emissions (Steiner et al., 2010; Ma et al., 2021b).

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

Atmospheric Measurement Techniques Discussions

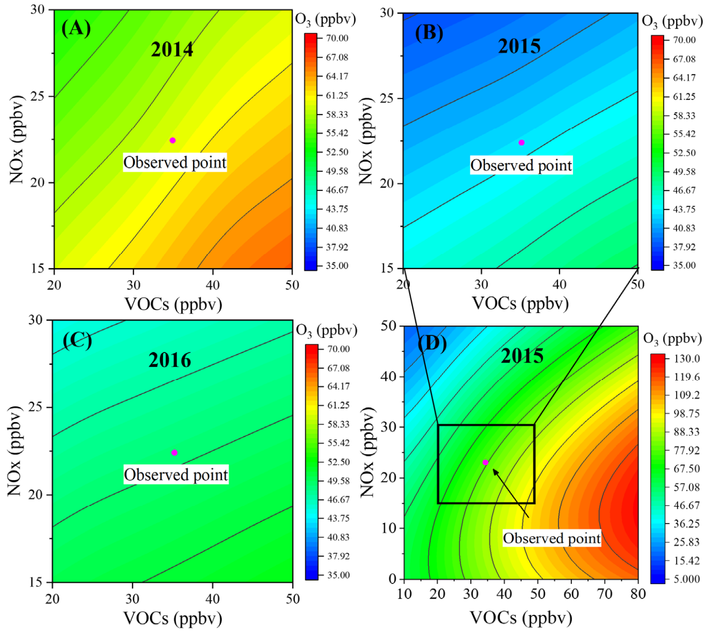

Figure 4. Ozone formation sensitivity curves from 2014-2016. (A, B, C: calculated by the RF model for 2014, 2015, and 2016, respectively. D: calculated by the OBM for 2015)

We compared O_3 sensitivity using the RF model based on the TVOCs and the initial VOC species in 2015. As shown in Figure S5, the O_3 concentrations predicted using the initial concentrations of VOC species were more accurate after correcting the reactivity during transport than those predicted using the TVOCs. Hence, a combination of the RF model and initial VOCs species (Figure 4B) can accurately depict the sensitivity regime of O_3 formation in comparison to the box model (Figure 4D), although a difference is observable between the predicted O_3 concentrations using these two models. In the box model, the O_3 isopleth plot was drawn with the maximum O_3 concentrations, while in the RF model, this plot was drawn with the real O_3

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

Atmospheric Measurement Techniques Discussions

concentrations.

4. Conclusions

In summary, this work investigated O_3 formation sensitivity in the summer from 2014-2016 in Beijing using the RF model coupled with the reactivity of VOC species. The results show that the prediction performance of O_3 by the RF model was significantly improved when VOC species were considered compared to TVOCs. Furthermore, after the photochemical loss of VOC species during transport was corrected, the RIs of the VOC species were well correlated with the OFPs of VOC species calculated using the MIR method, thus indicating that the RIs in the ML model reflect the chemical reactivity of VOCs. Meanwhile, both NOx and highly reactive species (such as isoprene, propene, benzene, and toluene) played an important role in O_3 formation. An increased contribution of temperature to O_3 production was observed, which implied the importance of temperature to O_3 pollution in the context of global warming conditions. Both the RF model and the box model results showed that O_3 formation was sensitive to VOCs in Beijing, although the sensitivity regime shifted from VOC-limited regime to a transition regime from 2014 to 2016. Due to the high computational efficiency of ML, the O_3 formation sensitivity plotted by the RF model coupled with the reactivity of VOC species can provide an accurate, flexible and efficient approach for analysing O_3 sensitivity in a near real-time way.

Code and data availability

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

Atmospheric Measurement Techniques Discussions

The code and datasets of VOCs and meteorology are available and will be provided by the corresponding authors Yongchun Liu (liuyc@buct.edu.cn) and Hong Li (lihong@craes.org.cn) upon request. The solar radiation data are publicly available via www.copernicus.eu/en.

Supplement

Supplementary information is available for this paper.

Author contributions

Junlei Zhan designed the idea and wrote this manuscript; Yongchun Liu and Hong Li provided useful advice and revised the manuscript; Wei Ma performed box model simulations; and Xin Zhang, Xuezhong Wang, Fang Bi, Yujie Zhang and Zhenhai Wu conducted the campaign and compiled the data. All authors contributed to the discussion of the results and writing of the manuscript.

Competing interest

The authors declare that they have no conflict of interest.

Acknowledgments

This research was financially supported by the Ministry of Science and Technology of the People's Republic of China (2019YFC0214701), the National Natural Science Foundation of China (41877306 and 92044301) and the programs from Beijing Municipal Science & Technology Commission (No. Z181100005418015). We thank Yizhen Chen for providing the meteorological parameter data for campaign studies.

References

Breiman, L. Random Forests. Machine Learning, 45, 5-32, 10.1023/A:1010933404324, 2001.

Carlo, P.D., Brune, W.H., Martinez, M., Harder, H., Lesher, R., Ren, X., Thornberry, T., Carroll, M.A., Young, V., Shepson, P.B., Riemer, D., Apel, E., Campbell, C. Missing OH Reactivity in a Forest: Evidence for Unknown Reactive Biogenic VOCs. Science, 304, 722-725, doi:10.1126/science.1094392, 2004.

Carter, W. Updated maximum incremental reactivity scale and hydrocarbon bin reactivities for regulatory applications. California Air Resources Board Contract, 1, 07-339, 2010.

Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., Pope, C.A., Shin, H., Straif, K., Shaddick, G., Thomas, M., van Dingenen, R., van Donkelaar, A., Vos, T., Murray, C.J.L., Forouzanfar, M.H. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. The Lancet, 389, 1907-1918, https://doi.org/10.1016/S0140-6736(17)30505-6, 2017.

Djalalova, I., Delle Monache, L., Wilczak, J. $PM_{2.5}$ analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model. Atmospheric Environment, 108, 76-87, https://doi.org/10.1016/j.atmosenv.2015.02.021, 2015.

Feng, R., Zheng, H. J., Gao, H., Zhang, A. R., Huang, C., Zhang, J. X., Luo, K., Fan, J. R. Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in Hangzhou, China. J. Clean. Prod., 231, 1005-1015, https://doi.org/10.1016/j.jclepro.2019.05.319, 2019.

Fu, T.-M., Zheng, Y., Paulot, F., Mao, J., Yantosca, R.M. Positive but variable sensitivity of August surface ozone to large-scale warming in the southeast United States. Nat. Clim. Change, 5, 454-458, 10.1038/nclimate2567, 2015.

Grange, S.K., Lee, J.D., Drysdale, W.S., Lewis, A.C., Hueglin, C., Emmenegger, L., Carslaw, D.C. COVID-19 lockdowns highlight a risk of increasing ozone pollution in European urban areas. Atmos. Chem. Phys., 21, 4169-4185, 10.5194/acp-21-4169-2021, 2021.

Hammer, M.-U., Vogel, B., Vogel, H. Findings on H_2O2/HNO_3 as an indicator of ozone sensitivity in Baden-Württemberg, Berlin-Brandenburg, and the Po valley based on numerical simulations. J. Geophys. Res. Atmos., 107, LOP 3-1-LOP 3-18, https://doi.org/10.1029/2000JD000211, 2002.

Hu, B., Zhao, X., Liu, H., Liu, Z., Song, T., Wang, Y., Tang, L., Xia, X., Tang, G., Ji, D., Wen, T., Wang, L., Sun, Y., Xin, J. Quantification of the impact of aerosol on broadband solar radiation in North China. Sci. Rep., 7, 44851, 10.1038/srep44851, 2017.

Kavassalis, S.C., Murphy, J.G. Understanding ozone-meteorology correlations: A role for dry deposition. Geophys. Res. Lett., 44, 2922-2931, https://doi.org/10.1002/2016GL071791, 2017.

Kohavi, R., John, G.H. Wrappers for feature subset selection. Artif. Intell., 97, 273-324, https://doi.org/10.1016/S0004-3702(97)00043-X, 1997.

Li, J., Cai, J., Zhang, M., Liu, H., Han, X., Cai, X., Xu, Y. Model analysis of meteorology and emission impacts on springtime surface ozone in Shandong. Sci. Total Environ., 771,

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

144784, https://doi.org/10.1016/j.scitotenv.2020.144784, 2021.

Li, K., Jacob, D.J., Liao, H., Zhu, J., Shah, V., Shen, L., Bates, K.H., Zhang, Q., Zhai, S. A two-pollutant strategy for improving ozone and particulate air quality in China. Nat. Geosci., 12, 906-910, 10.1038/s41561-019-0464-x, 2019.

Li, K., Jacob, D.J., Shen, L., Lu, X., De Smedt, I., Liao, H. Increases in surface ozone pollution in China from 2013 to 2019: anthropogenic and meteorological influences. Atmos. Chem. Phys., 20, 11423-11433, 10.5194/acp-20-11423-2020, 2020a.

Li, Q., Su, G., Li, C., Liu, P., Zhao, X., Zhang, C., Sun, X., Mu, Y., Wu, M., Wang, Q., Sun, B. An investigation into the role of VOCs in SOA and ozone production in Beijing, China. Sci. Total Environ., 720, 137536, https://doi.org/10.1016/j.scitotenv.2020.137536, 2020b.

Lin, W., Xu, X., Ge, B., Liu, X. Gaseous pollutants in Beijing urban area during the heating period 2007–2008: variability, sources, meteorological, and chemical impacts. Atmos. Chem. Phys., 11, 8157-8170, 10.5194/acp-11-8157-2011, 2011.

Liu, H., Liu, J., Liu, Y., Ouyang, B., Xiang, S., Yi, K., Tao, S. Analysis of wintertime O_3 variability using a random forest model and high-frequency observations in Zhangjiakou—an area with background pollution level of the North China Plain. Environ. Pollut., 262, 114191, https://doi.org/10.1016/j.envpol.2020.114191, 2020a.

Liu, Y., Cheng, Z., Liu, S., Tan, Y., Yuan, T., Yu, X., Shen, Z. Quantitative structure activity relationship (QSAR) modelling of the degradability rate constant of volatile organic compounds (VOCs) by OH radicals in atmosphere. Sci. Total Environ., 729, 138871, https://doi.org/10.1016/j.scitotenv.2020.138871, 2020b.

Liu, Y., Wang, T. Worsening urban ozone pollution in China from 2013 to 2017 – Part 1: The complex and varying roles of meteorology. Atmos. Chem. Phys., 20, 6305-6321, 10.5194/acp-20-6305-2020, 2020.

Liu, Z., Qi, Z., Ni, X., Dong, M., Ma, M., Xue, W., Zhang, Q., Wang, J. How to apply O_3 and $PM_{2.5}$ collaborative control to practical management in China: A study based on meta-analysis and machine learning. Sci. Total Environ., 772, 145392, https://doi.org/10.1016/j.scitotenv.2021.145392, 2021.

Ma, R., Ban, J., Wang, Q., Li, T. Statistical spatial-temporal modeling of ambient ozone exposure for environmental epidemiology studies: A review. Sci. Total Environ., 701, 134463, https://doi.org/10.1016/j.scitotenv.2019.134463, 2020.

Ma, R., Ban, J., Wang, Q., Zhang, Y., Yang, Y., He, M.Z., Li, S., Shi, W., Li, T. Random forest model based fine scale spatiotemporal O_3 trends in the Beijing-Tianjin-Hebei region in China, 2010 to 2017. Environ. Pollut., 276, 116635, https://doi.org/10.1016/j.envpol.2021.116635, 2021a.

Ma, S., Shao, M., Zhang, Y., Dai, Q., Xie, M. Sensitivity of $PM_{2.5}$ and O_3 pollution episodes to meteorological factors over the North China Plain. Sci. Total Environ., 792, 148474, https://doi.org/10.1016/j.scitotenv.2021.148474, 2021b.

Makar, P.A., Fuentes, J.D., Wang, D., Staebler, R.M., Wiebe, H.A. Chemical processing of biogenic hydrocarbons within and above a temperate deciduous forest. J. Geophys. Res. Atmos., 104, 3581-3603, https://doi.org/10.1029/1998JD100065, 1999.

Martin, R.V., Fiore, A.M., Van Donkelaar, A. Space-based diagnosis of surface ozone sensitivity to anthropogenic emissions. Geophys. Res. Lett., 31, https://doi.org/10.1029/2004GL019416, 2004.

Mo, Z., Shao, M., Liu, Y., Xiang, Y., Wang, M., Lu, S., Ou, J., Zheng, J., Li, M., Zhang, Q., Wang, X., Zhong, L. Species-specified VOC emissions derived from a gridded study in the Pearl River Delta, China. Sci. Rep., 8, 2963, 10.1038/s41598-018-21296-y, 2018.

Ou, J., Yuan, Z., Zheng, J., Huang, Z., Shao, M., Li, Z., Huang, X., Guo, H., Louie, P.K.K. Ambient Ozone Control in a Photochemically Active Region: Short-Term Despiking or Long-Term Attainment? Environ. Sci. Technol., 50, 5720-5728, 10.1021/acs.est.6b00345, 2016.

Paraschiv, S., Barbuta-Misu, N., Paraschiv, S.L. Influence of NO_2, NO and meteorological conditions on the tropospheric O3 concentration at an industrial station. Energy Rep., 6, 231-236, https://doi.org/10.1016/j.egyr.2020.11.263, 2020.

Sayeed, A., Choi, Y., Eslami, E., Jung, J., Lops, Y., Salman, A.K., Lee, J.-B., Park, H.-J., Choi, M.-H. A novel CMAQ-CNN hybrid model to forecast hourly surface-ozone concentrations 14 days in advance. Sci. Rep., 11, 10891, 10.1038/s41598-021-90446-6, 2021.

Shao, M., Wang, W., Yuan, B., Parrish, D.D., Li, X., Lu, K., Wu, L., Wang, X., Mo, Z., Yang, S., Peng, Y., Kuang, Y., Chen, W., Hu, M., Zeng, L., Su, H., Cheng, Y., Zheng, J., Zhang, Y. Quantifying the role of $PM_{2.5}$ dropping in variations of ground-level ozone: Inter-comparison between Beijing and Los Angeles. Sci. Total Environ., 788, 147712, https://doi.org/10.1016/j.scitotenv.2021.147712, 2021.

Sillman, S. The use of NOy , H_2O_2, and HNO_3 as indicators for ozone-NOx-hydrocarbon sensitivity in urban locations. J. Geophys. Res. Atmos., 100, 14175-14188, https://doi.org/10.1029/94JD02953, 1995.

Steiner, A.L., Davis, A.J., Sillman, S., Owen, R.C., Michalak, A.M., Fiore, A.M. Observed suppression of ozone formation at extremely high temperatures due to chemical and biophysical feedbacks. P. Natl. Acad. Sci. USA, 107, 19685-19690, 10.1073/pnas.1008336107, 2010.

Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. J. Chem. Inf. Comput. Sci., 43, 1947-1958, 10.1021/ci034160g, 2003.

Tan, Z., Lu, K., Jiang, M., Su, R., Dong, H., Zeng, L., Xie, S., Tan, Q., Zhang, Y. Exploring ozone pollution in Chengdu, southwestern China: A case study from radical chemistry to O_3-VOC-NOx sensitivity. Sci. Total Environ., 636, 775-786, https://doi.org/10.1016/j.scitotenv.2018.04.286, 2018.

Tang, X., Zhu, J., Wang, Z.F., Gbaguidi, A. Improvement of ozone forecast over Beijing based on ensemble Kalman filter with simultaneous adjustment of initial conditions and emissions. Atmos. Chem. Phys., 11, 12901-12916, 10.5194/acp-11-12901-2011, 2011.

Ueno, H., Tsunematsu, N. Sensitivity of ozone production to increasing temperature and reduction of precursors estimated from observation data. Atmos. Environ., 214, 116818, https://doi.org/10.1016/j.atmosenv.2019.116818, 2019.

Vélez-Pereira, A.M., De Linares, C., Belmonte, J. Aerobiological modeling I: A review of predictive models. Sci. Total Environ., 795, 148783, https://doi.org/10.1016/j.scitotenv.2021.148783, 2021.

https://doi.org/10.5194/amt-2021-367
Preprint. Discussion started: 5 November 2021

Wang, P., Qiao, X., Zhang, H. Modeling $PM_{2.5}$ and O_3 with aerosol feedbacks using WRF/Chem over the Sichuan Basin, southwestern China. Chemosphere, 254, 126735, https://doi.org/10.1016/j.chemosphere.2020.126735, 2020a.

Wang, T., Nie, W., Gao, J., Xue, L.K., Gao, X.M., Wang, X.F., Qiu, J., Poon, C.N., Meinardi, S., Blake, D., Wang, S.L., Ding, A.J., Chai, F.H., Zhang, Q.Z., Wang, W.X. Air quality during the 2008 Beijing Olympics: secondary pollutants and regional impact. Atmos. Chem. Phys., 10, 7603-7615, 10.5194/acp-10-7603-2010, 2010.

Wang, T., Xue, L., Brimblecombe, P., Lam, Y.F., Li, L., Zhang, L. Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects. Sci. Total Environ., 575, 1582-1596, https://doi.org/10.1016/j.scitotenv.2016.10.081, 2017a.

Wang, Y., Gao, W., Wang, S., Song, T., Gong, Z., Ji, D., Wang, L., Liu, Z., Tang, G., Huo, Y., Tian, S., Li, J., Li, M., Yang, Y., Chu, B., Petäjä, T., Kerminen, V.-M., He, H., Hao, J., Kulmala, M., Wang, Y., Zhang, Y. Contrasting trends of $PM_{2.5}$ and surface-ozone concentrations in China from 2013 to 2017. Natl. Sci. Rev., 7, 1331-1339, 10.1093/nsr/nwaa032, 2020b.

Wang, Y., Li, Y., Pu, W., Wen, K., Shugart, Y.Y., Xiong, M., Jin, L. Random Bits Forest: a Strong Classifier/Regressor for Big Data. Sci. Rep., 6, 30086, 10.1038/srep30086, 2016.

Wang, Y., Wen, Y., Wang, Y., Zhang, S., Zhang, K.M., Zheng, H., Xing, J., Wu, Y., Hao, J. Four-Month Changes in Air Quality during and after the COVID-19 Lockdown in Six Megacities in China. Environ. Sci. Technol. Lett., 7, 802-808, 10.1021/acs.estlett.0c00605, 2020c.

Wang, Y., Wu, G., Deng, L., Tang, Z., Wang, K., Sun, W., Shangguan, Z. Prediction of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm. Sci. Rep., 7, 6940, 10.1038/s41598-017-07197-6, 2017b.

Xing, J., Zheng, S., Ding, D., Kelly, J.T., Wang, S., Li, S., Qin, T., Ma, M., Dong, Z., Jang, C., Zhu, Y., Zheng, H., Ren, L., Liu, T.-Y., Hao, J. Deep Learning for Prediction of the Air Quality Response to Emission Changes. Environ. Sci. Technol., 54, 8589-8600, 10.1021/acs.est.0c02923, 2020.

Xue, L.K., Wang, T., Gao, J., Ding, A.J., Zhou, X.H., Blake, D.R., Wang, X.F., Saunders, S.M., Fan, S.J., Zuo, H.C., Zhang, Q.Z., Wang, W.X. Ground-level ozone in four Chinese cities: precursors, regional transport and heterogeneous processes. Atmos. Chem. Phys., 14, 13175-13188, 10.5194/acp-14-13175-2014, 2014.

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. P. Natl. Acad. Sci. USA, 118, e2102705118, 10.1073/pnas.2102705118, 2021a.

Yang, L., Yuan, Z., Luo, H., Wang, Y., Xu, Y., Duan, Y., Fu, Q. Identification of long-term evolution of ozone sensitivity to precursors based on two-dimensional mutual verification. Sci. Total Environ., 760, 143401, https://doi.org/10.1016/j.scitotenv.2020.143401, 2021b.

Yu, S. Fog geoengineering to abate local ozone pollution at ground level by enhancing air moisture. Environ. Chem. Lett., 17, 565-580, 10.1007/s10311-018-0809-5, 2019.

Yuan, B., Hu, W.W., Shao, M., Wang, M., Chen, W.T., Lu, S.H., Zeng, L.M., Hu, M. VOC emissions, evolutions and contributions to SOA formation at a receptor site in eastern China. Atmos. Chem. Phys., 13, 8815-8832, 10.5194/acp-13-8815-2013, 2013.

Zhan, J., Feng, Z., Liu, P., He, X., He, Z., Chen, T., Wang, Y., He, H., Mu, Y., Liu, Y. Ozone and SOA formation potential based on photochemical loss of VOCs during the Beijing summer. Environ. Pollut., 285, 117444, https://doi.org/10.1016/j.envpol.2021.117444, 2021.

Zhang, X., Li, H., Wang, X., Zhang, Y., Bi, F., Wu, Z., Liu, Y., Zhang, H., Gao, R., Xue, L., Zhang, Q., Chen, Y., Chai, F., Wang, W. Heavy ozone pollution episodes in urban Beijing during the early summertime from 2014 to 2017: Implications for control strategy. Environ. Pollut., 285, 117162, https://doi.org/10.1016/j.envpol.2021.117162, 2021.

Zhao, H., Chen, K., Liu, Z., Zhang, Y., Shao, T., Zhang, H. Coordinated control of $PM_{2.5}$ and O_3 is urgently needed in China after implementation of the "Air pollution prevention and control action plan". Chemosphere, 270, 129441, https://doi.org/10.1016/j.chemosphere.2020.129441, 2021.