Supplement of

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

* Corresponding author: liuyc@buct.edu.cn; lihong@craes.org.cn

Text S1. Data Information

VOCs were measured by an online commercial instrument (GC-866, Chromatotec, France), which consisted of two independent analyzers for detecting C_2-C_6 and C_6-C_{12} hydrocarbon components. Both analyzers were equipped with a preconcentration system, a chromatographic column, and a flame ionization detector. In our study, total of 51 VOCs (including 21 alkanes, 13 alkenes, 1 alkyne and 16 aromatics) were analyzed within a limit of quantification of 0.1-100 ppbv. O_3 and NO_X were measured using a UV photometric O_3 analyzer (model 49i, Thermo-Fisher Scientific, USA) with a detection limit of 1.0 ppbv, and a chemiluminescence NO_X analyzer (model 17i, Thermo-Fisher Scientific, USA) with a detection limit of 1.0 ppbv, respectively. $PM_{2.5}$ was detected by synchronized hybrid ambient real-time particulate monitor (Model 5030, Thermo-Fisher Scientific, USA). CO was detected by CO analyzer (model 48i, Thermo-Fisher Scientific, USA). Meteorological data (T, RH, WS&WD and solar irradiation) were measured by an automatic weather station (MAWS301, Vaisala, Finland). This instrument output data hourly and was checked and calibrated weekly. It should be noted that meteorological data and solar radiation were missing in May and June 2014, May 25 to July 1, 2015 and June 2016. Therefore, the meteorological data use Beijing Capital Airport data, and solar radiation data are from the Copernicus Services (www.copernicus.eu/en).

Text S2. Calculation of initial VOCs

Changes in VOCs concentration as a function of time due to photochemical reaction are described in equation S1.

$$\int c_{\mathrm{OH}}\mathrm{d}t = \frac{1}{k_{\mathrm{A,OH}} - k_{\mathrm{B,\,OH}}}\left[\ln\left(\frac{\mathrm{VOC_A}}{\mathrm{VOC_B}}\right)_{\mathrm{initial}} - \ln\left(\frac{\mathrm{VOC_A}}{\mathrm{VOC_B}}\right)\right] \tag{S1}$$

The initial concentration of species *i* can be calculated using Equation S2.

$$\mathrm{VOC_{i,\,initial}} = \frac{\mathrm{VOC_i}}{\exp(-\mathrm{k_{i,OH}})\exp(\int \mathrm{c_{OH}dt})} \tag{S2}$$

Substituting equation 1 into equation 2, then we can get equation S3.

$$\mathrm{VOC}_{i,\,\mathrm{initial}} = \frac{\mathrm{VOC}_i}{\exp(-\mathrm{k}_{i,\,\mathrm{OH}})\exp\left(\frac{1}{k_{\mathrm{A,OH}} - k_{\mathrm{B,\,OH}}}\left[\ln\left(\frac{\mathrm{VOC_A}}{\mathrm{VOC_B}}\right)_{\mathrm{initial}} - \ln\left(\frac{\mathrm{VOC_A}}{\mathrm{VOC_B}}\right)\right]\right)} \tag{S3}$$

In this study, the ethylbenzene/xylene pair was used to calculate ambient OH exposure [1]. The hourly concentrations of ethylbenzene and xylene were used to calculate the concentration of initial VOCs.

Text S3. Relative Importance (RI)

The central idea of RI, also known as “Breiman-Cutler importance”[2], is to measure the reduction in the accuracy of RF model. The method provides a ranking for the RI of the different features and is associated with a predictive performance. In particular, the RI for feature *i* is derived as follows. Firstly, we record the prediction error for decision tree *n* as $errOOB_1$. Secondly, for feature *i*, add data noise (less than 1%, usually 0.1% was used)[2, 3] and then place the OOB data under decision tree *n* again, recording the new error as $errOOB_2$. Finally, repeat steps 1 and 2, averaging the errors for all decision trees N to obtain the RI for feature *i*.

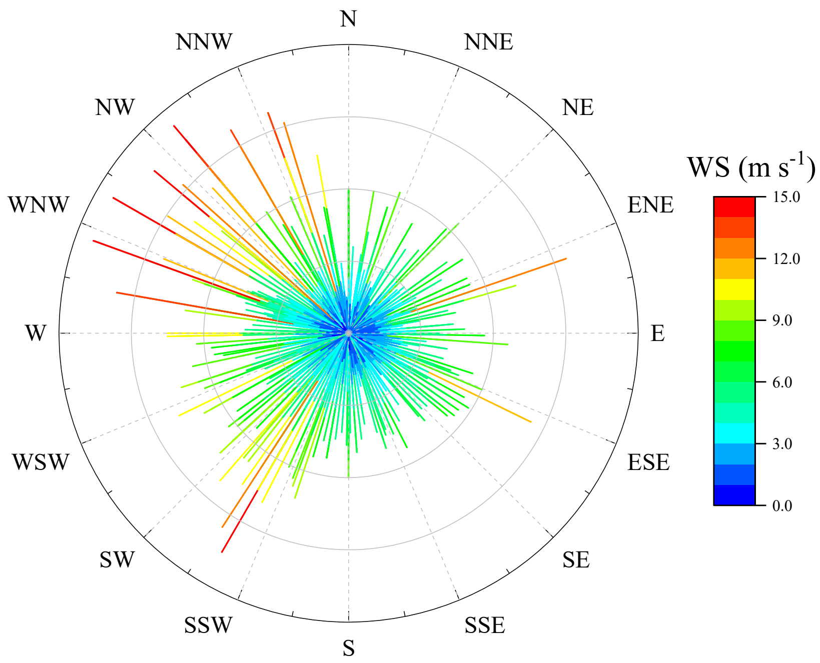

Figure S1. Map of wind rose during the observation.

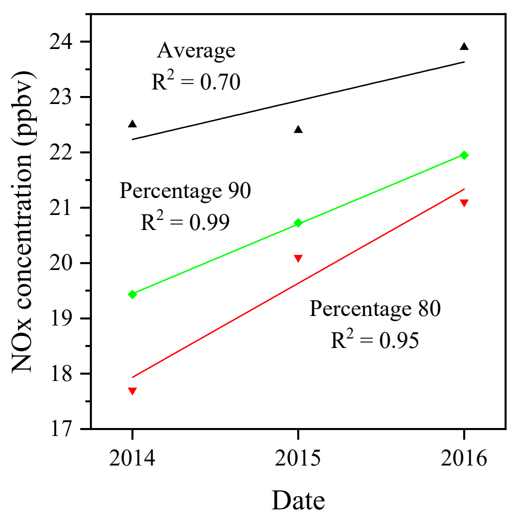

Figure S2. Change in NOx concentration from 2014-2016.

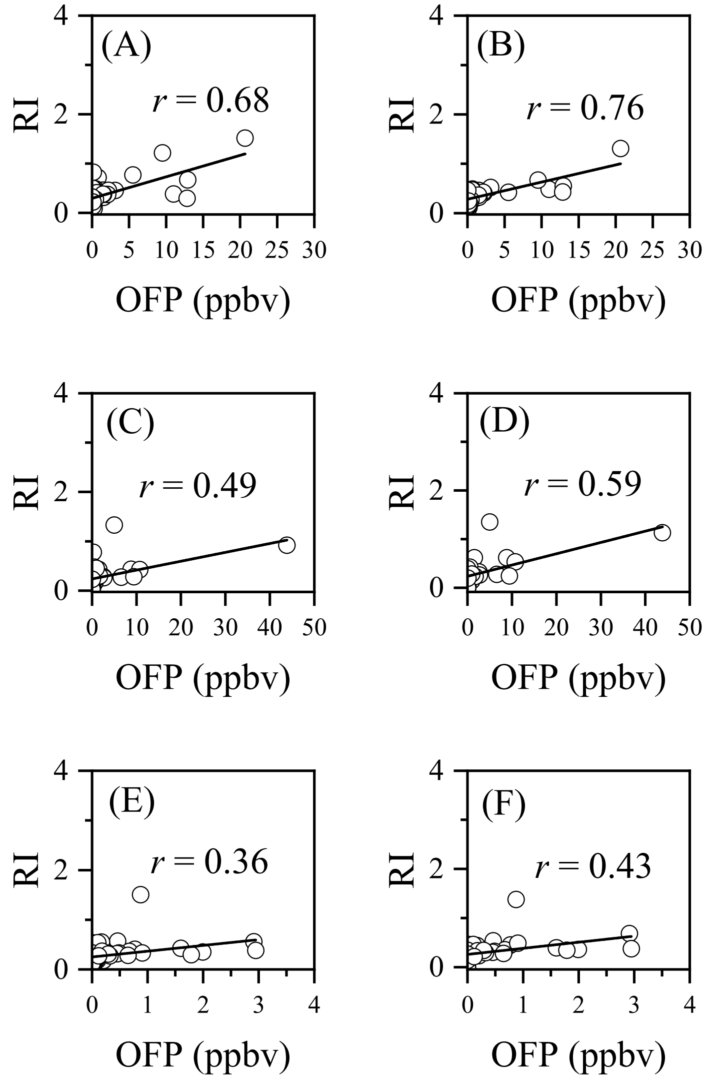

Figure S3. The relationship between RI and OFP of different VOC species in 2015 and 2016. (A, B: measured and initial concentrations of VOCs in 2014, respectively; C, D: measured and initial concentrations of VOCs in 2015, respectively; E, F: measured and initial concentrations of VOCs in 2016, respectively)

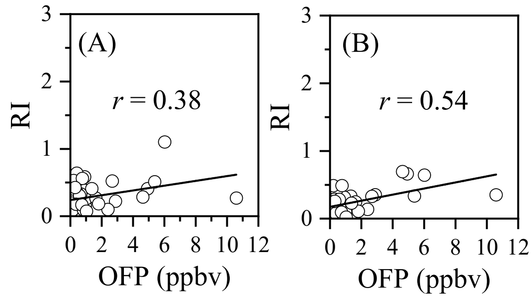

Figure S4. The relationship between RI and OFP in 2019 summer in Daxing region*.(A, B: measured and initial concentrations of VOCs in 2019, respectively)

* The information of sampling site and VOCs characteristics can be seen in our previous study [4].

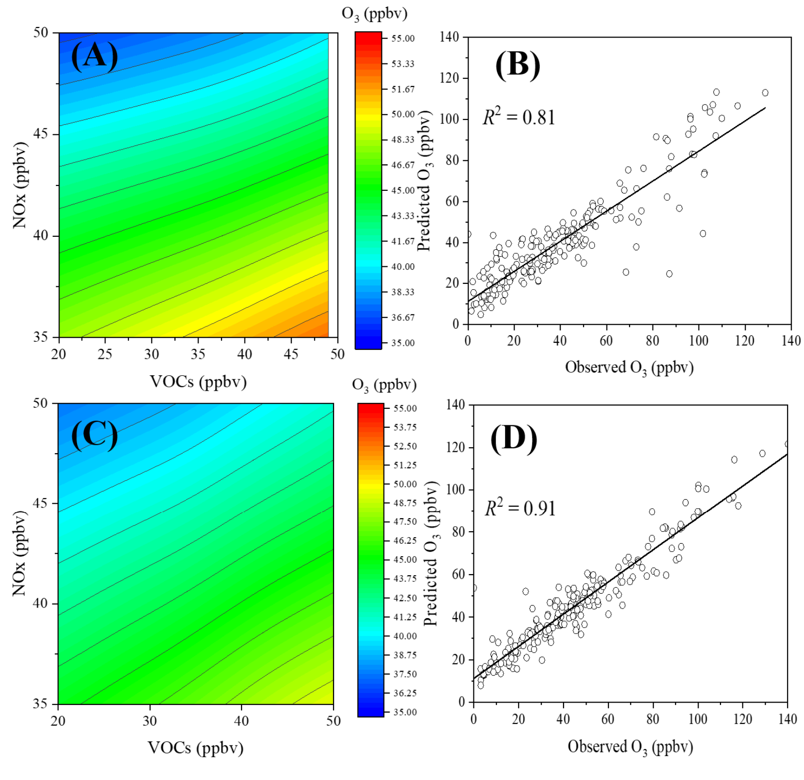

Figure S5. Comparison of O_3 formation sensitivity curves calculated using TVOCs (A) and initial concentrations of VOC species (C) in 2015. (B and D show the prediction performance using total VOCs and initial VOCs species, respectively)

Table S1. RF model parameters and input parameters

RF model parameters		RF model inputs	
Type	Value	Type	Parameters
training data	80%	Figure 2A	Total VOC concentration, $PM_{2.5}$, NO, NO_2, CO, SR, RH, WD, WS, T
testing data	20%		
leaf number	5	Figure 2B	51 measured VOC species, $PM_{2.5}$, NO, NO_2, CO, SR, RH, WD, WS, T
trees number	500		
fboot	1	Figure 2C	51 initial VOC species, $PM_{2.5}$, NO, NO_2, CO, SR, RH, WD, WS, T
method	regression		
sampling	random		

Table S2. Changes in RIs using measured and initial concentrations of VOC species

specie	2014		2015		2016	
	$RI_{measured}$	$RI_{initial}$	$RI_{measured}$	$RI_{initial}$	$RI_{measured}$	$RI_{initial}$
NOx	7.49	13.01	11.15	11.17	9.10	8.30
T	2.01	2.35	5.29	5.18	8.85	9.37
SR	5.09	7.33	5.65	5.21	2.68	2.66
isoprene	2.65	2.08	6.04	5.87	6.65	5.95
propene	5.99	4.95	4.18	5.08	1.55	1.58
CO	2.10	2.74	3.18	3.48	3.01	2.79
PM2.5	2.34	2.70	5.54	4.30	4.17	3.96
toluene	4.80	2.54	1.25	1.24	1.32	1.49
benzene	1.62	1.84	2.17	1.90	2.47	2.73
RH	2.60	2.50	2.78	2.74	1.93	2.06
ethylene	1.51	1.83	1.97	2.77	1.88	1.73
trans-2-butene	1.83	1.29	1.96	2.77	2.43	1.82

Unit of RI: %

REFERENCE

[1] Shao M, Bin W, Sihua L, et al. Effects of Beijing Olympics Control Measures on Reducing Reactive Hydrocarbon Species. Environ. Sci. Technol., 45, 514-519, 2011.

[2] Breiman L. Random Forests. Mach. Learn., 45, 5-32, 2001.

[3] Wang T, Xue L, Brimblecombe P, et al. Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects. Sci. Total Environ., 575, 1582-1596, 2017.

[4] Zhan J, Feng Z, Liu P, et al. Ozone and SOA formation potential based on photochemical loss of VOCs during the Beijing summer. Environ. Pollut. 285, 117444, 2021.