



## Supplement of

# Ozone formation sensitivity study using machine learning coupled with the reactivity of volatile organic compound species

Junlei Zhan et al.

Correspondence to: Yongchun Liu (liuyc@buct.edu.cn) and Hong Li (lihong@craes.org.cn)

The copyright of individual parts of the supplement might differ from the article licence.

### 1 Text S1. Field observations

VOCs were measured in May and June from 2014 to 2016 by an online commercial instrument (GC-866, Chromatotec, France), which consisted of two independent analyzers for C<sub>2</sub>-C<sub>6</sub> and C<sub>6</sub>-C<sub>12</sub> hydrocarbons. Both analyzers were equipped with a preconcentration system, a chromatographic column, and a flame ionization detector. The analyzers are located in an air-conditioned room and the sample tubes are wrapped with a heating jackets and insulation to ensure that the temperature remains stable between 22 and  $27^{\circ}$ C.

9 The samples were injected into the low carbon (C<sub>2</sub>-C<sub>6</sub>) analyzer and the high 10 carbon (C<sub>6</sub>-C<sub>12</sub>) analyzer, respectively. Isoprene was detected in the components of C<sub>2</sub>-C<sub>6</sub>, while  $\alpha$ -pinene and  $\beta$ -pinene were detected along with other VOCs of C<sub>6</sub>-C<sub>12</sub>. In the 11 12 low carbon analyzer, the samples were adsorbed by a capture tube at -8 °C. The capture tube was then rapidly heated to 220 °C. The samples were introduced into a 13 chromatographic column (id=0.53mm, length=25m) with hydrogen as the carrier gas 14 15 and detected by a flame ionization detector (FID) detector. In the high carbon analyzer, the samples were adsorbed by a capture tube at room temperature; then the capture tube 16 was heated to 380 °C, introduced into a chromatographic column (id=0.28mm, 17 length=30m) with hydrogen as the carrier gas and finally detected by the same FID 18 detector. The material in the column was Al<sub>2</sub>O<sub>3</sub>/Na<sub>2</sub>SO<sub>4</sub>. 19

The signals of VOCs were converted into chromatograms for qualitative and quantitative analysis. Before quantitative analysis, the retention time of each component was carefully checked using the chromatographic analysis software. The

instruments were calibrated using both internal and external calibrations. Internally 23 calibration was carried out twice every 24 hours using n-butane, n-hexane and benzene 24 25 at different flow rates. External calibration was performed monthly using standard gas mixtures of volatile organic compounds (PAMS and TO-14, Linde gas, USA). The 26 27 concentrations of each species were calculated according to the corresponding working curves with six concentration levels. In this study, total of 51 VOCs (including 21 28 alkanes, 13 alkenes, 1 alkyne and 16 aromatics) were analyzed within a limit of 29 quantification of 0.002-0.05 ppbv as shown in Table S5. The relative standard 30 31 derivations (RSDs) were within 10% for each compound among seven replicates.

O3 and NOx were measured using a UV photometric O3 analyzer (model 49i, 32 Thermo-Fisher Scientific, USA) with a detection limit of 1.0 ppbv, and a 33 34 chemiluminescence NO<sub>X</sub> analyzer (model 17i, Thermo-Fisher Scientific, USA) with a detection limit of 1.0 ppbv, respectively. PM2.5 was detected by synchronized hybrid 35 ambient real-time particulate monitor (Model 5030, Thermo-Fisher Scientific, USA). 36 CO was detected by CO analyzer (model 48i, Thermo-Fisher Scientific, USA). 37 Meteorological data (T, RH, WS&WD and solar irradiation) were measured by an 38 automatic weather station (MAWS301, Vaisala, Finland). This instrument output data 39 hourly and was checked and calibrated weekly. It should be noted that meteorological 40 data and solar radiation were missing in May and June 2014, May 25 to July 1, 2015 41 and June 2016. Therefore, the meteorological data use Beijing Capital Airport data, and 42 solar radiation data are from the Copernicus Services (www.copernicus.eu/en). 43

44 Text S2. Calculation of initial VOCs concentrations

Photochemical initial concentration (PIC) proposed by Shao et al. (2011), which
is calculated based on the photochemical-age approach and has been applied to evaluate
the effect of photochemical processing on measured VOC levels. Equation S1
essentially describes the integrated OH exposure (Shao et al., 2011).

49 
$$\int c_{\text{OH}} dt = \frac{1}{k_{\text{A,OH}} - k_{\text{B,OH}}} \left[ \ln\left(\frac{\text{VOC}_{\text{A}}}{\text{VOC}_{\text{B}}}\right)_{\text{initial}} - \ln\left(\frac{\text{VOC}_{\text{A}}}{\text{VOC}_{\text{B}}}\right) \right]$$
(S1)

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

51 
$$\operatorname{VOC}_{i, \text{ initial}} = \frac{\operatorname{VOC}_{i}}{\exp(-k_{i, OH})\exp(\int c_{OH} dt)}$$
 (S2)

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

53 
$$\operatorname{VOC}_{i, \text{ initial}} = \frac{\operatorname{VOC}_{i}}{\exp(-k_{i, \text{ OH}}) \exp(\frac{1}{k_{A, \text{OH}} - k_{B, \text{ OH}}} \left[\ln(\frac{\operatorname{VOC}_{A}}{\operatorname{VOC}_{B}})_{\text{ initial}} - \ln(\frac{\operatorname{VOC}_{A}}{\operatorname{VOC}_{B}})\right])}$$
(S3)

54 Where C<sub>OH</sub> represents the ambient OH concentration;  $k_{A,OH}$  and  $k_{B,OH}$  represent the 55 reaction rate of compound A and B with OH radical, respectively; t represents the 56 reaction time of species *i* in the ambient.

In previous work (Shao et al., 2011; Zhan et al., 2021), the selection of ethylbenzene and xylene as tracers was justified for calculating ambient OH exposure under the following conditions: 1) the concentrations of xylene and ethylbenzene were well correlated (Figure S9), which indicated that they were simultaneously emitted; 2) they had different degradation rates in the atmosphere; and 3) the calculated PICs were in good agreement with those calculated using other tracers (Shao et al., 2011; Zhan et al., 2021).

In this study, the ethylbenzene/xylene pair was used to calculate ambient OH exposure. As shown in Figure S9, the concentrations of xylene and ethylbenzene are well correlated, which indicates that they are simultaneously emitted. In addition, we compared the PICs according to xylene/ethylbenzene with that using toluene/benzene (Figure S10). The calculated PICs ratio (PIC  $_{Xylene/Ethylbenzene}$  / PIC  $_{Toluene/Benzene}$ ) varied from 0.5 to 1.5 with a mean value of 0.96. This means the calculated initial VOCs was in good agreement when using different tracers. The mean ratio (0.52, from 0.45 to 0.66) of ethylbenzene/xylene before sunrise was taken as the initial ratio of ethylbenzene/xylene. Sensitivity tests showed that the uncertainty of PICs caused by the OH exposure (from -10% to +10%) ranged from 0.55 to 1.57 (Table S5).

Variations of air mass may also affect the VOC ratio. Figure S11 A-D shows the 74 75 mean concentration distribution of ethylbenzene and xylene in the early morning and the whole day based on potential source contribution function (PSCF) analysis. Xylene 76 showed similar patterns to ethylbenzene in different air mass trajectories and different 77 78 periods. These results indicate that the emissions of xylene and ethylbenzene were constant throughout the day and variations of air mass should have little influence on 79 the initial ratio of VOCs. The hourly concentrations of ethylbenzene and xylene were 80 used to calculate the concentration of initial VOCs. The initial VOC was calculated by 81 adding the measured VOC concentration and the calculated photochemical loss. Figure 82 S12 shows the diurnal variations of the observed and initial VOCs concentrations from 83 2014 to 2016. Photochemical loss of VOC occurred mainly during the daytime. 84

It should be noted that the lifetimes  $(1/k_{2}c_{OH})$  of highly reactive VOCs, such as isoprene, greatly depend on the OH exposure. The photochemical ages of isoprene were 0.01–6.21 h (1.26 ± 1.12 h). This value is comparable with previously reported photochemical ages (Shao et al., 2011; Gao et al., 2018). However, the initial

89	concentrations of highly reactive VOCs may be overestimated due to their short
90	lifetimes and should be taken as the upper limits. On the other hand, isoprene is a
91	biogenic VOC, while xylene and ethylbenzene are anthropogenic VOCs. If they do not
92	share the same air mass histories, an additional uncertainty is inevitable for the PICs of
93	isoprene. As shown in Figure S11, isoprene showed similar patterns to that of xylene
94	and ethylbenzene, which means VOC emissions are evenly distributed in Beijing during
95	our observations. This can be ascribed to the fact that our observation site is a typical
96	urban station. Although isoprene and xylene/ethylbenzene different sources, both them
97	are non-point sources on a city scale. Therefore, the photochemical clock calculated
98	using xylene and ethylbenzene is able to correct the photochemical loss of biogenic
99	VOCs to some extent. It should be noted that uncertainty is inevitable when we
100	estimating the photochemical age (Parrish et al., 2007). However, the aim of this work
101	is to test whether the ML-model can reflect the influence of photochemical loss of
102	VOCs species on O <sub>3</sub> modelling. The PICs should provide additional information for
103	understanding O <sub>3</sub> formation in the atmosphere.

Text S3. Workflow of RF model and the calculation of Relative Importance (RI)
The workflow of RF model used in this study was established through the following
steps.

(1) Data description. The length of the input data from 2014 to 2016 were 1190, 1062
and 872 rows, respectively, in which different types of VOCs, NOx, CO, PM<sub>2.5</sub> and
meteorological parameters (including temperature, relative humidity, solar radiation,
wind speed and direction) were used as input variables and O<sub>3</sub> was the output variable.
The mean values (±standard deviation) of input/output parameters are shown in Table
S1.

(2) Data process. After the extreme values were removed, all data were normalized (between 0 and 1) in order to decrease the sample distribution range, accelerate calculation efficiency and improve the robustness of the RF model. Then, the dataset was randomly divided into 12 subsets. Thus, a 12-fold cross-validation was performed by alternately taking one subset as testing data and the rest as training data to ensure that each data point has an equal chance being trained and tested.

(3) Hyper-parameters optimization. All network configuration parameters (i.e., leaf
number, number of trees, algorithm, and so on) were modified by a trial and error
method to obtain the optimized network structure. The optimized RF model parameters
are shown in Table S2. Figures S13 and S14 show the examples to optimize the number
of minimal samples split and trees, respectively.

(4) Model uncertainty estimation. The uncertainty of the model was estimatedaccording to the predicted and observed O<sub>3</sub> concentrations. The performance of the

127 model was evaluated using R square ( $R^2$ ) and Root Mean Squared Error (RMSE).

(5) Relative importance (RI) calculation: The influence of an input variable on model
performance was evaluated by changes in the accuracy of the model by variable
permuting. Briefly, a change of prediction error was resulted from permuting a variable
across the observations. The magnitude of the response was estimated using out-of-bag
error of a predictor according to following steps.

For a random forest model that has T learners and p predictors in the training data, 133 the first step is to identify the out-of-bag observations and the indices of the predictor 134 135 variables that are split to a growing tree t (from 1 to T). Then, one can estimate the outof-bag error ( $\varepsilon_t$ ) for each tree. For a predictor variable  $x_i$  (i: from 1 to p), one can estimate 136 the model error  $(\varepsilon_{t, j})$  again corroding to the out-of-bag observations after randomly 137 permuting the observations of x<sub>j</sub>. Thus, the difference of the model error  $(d_{t,j} = \varepsilon_{t,j} - \varepsilon_t)$ 138 is obtained. If the predictor variables are not split, the difference of a growing tree t is 139 0. The second step is to calculate the mean difference of the model errors  $(\overline{d}_i)$ , and the 140 141 standard deviation ( $\sigma_i$ ) of the differences for all the learners and each predictor variable 142 in the training data. Finally, the out-of-bag relative importance (RI) for x<sub>i</sub> is calculated by dividing the difference of the model errors by the standard deviation  $(\overline{d}_i/\sigma_i)$ . 143

(6) EKMA curves. The Empirical Kinetic Modeling Approach (EKMA) was used to
assess the O<sub>3</sub> formation mechanism regime. Both the RF model and a box model with
Master Chemical Mechanism (MCM, 3.3.1) were used to calculate the EKMA curves.
For the RF model simulations, the observed point data was chosen as the mean values
of the input parameters during our observations, then the concentrations of VOCs and

NOx were varied  $\pm 10\%$  (or from 90% to 110%) of their mean values with a step of 1% 149 in a two-dimensional matrix along with other inputs unchanged. This matrix was used 150 as the testing data, while all the measured data were taken as the training data in the RF 151 model to simulate O<sub>3</sub> concentrations under different scenarios of VOCs and NOx 152 concentrations. To decrease the model uncertainty, we set relatively small variations of 153 VOCs and NOx (±10%) compared to the observed values in this study. The mean 154 relative error of simulated O<sub>3</sub> concentrations between RF model and Box model (within 155 15.6%, Figure S8) suggests that the RF model can well predict O<sub>3</sub> concentrations during 156 our observations. 157





Figure S1. Map of wind rose during the observation.



Figure S2. Variations of NO<sub>x</sub> concentrations from 2014-2016.



Figure S3. Comparison of the predicted and measured O<sub>3</sub> 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)



Figure S4. Comparison of the predicted and measured O<sub>3</sub> concentrations in Beijing in
the summer of 2016. (A and D: TVOC concentrations; B and E: measured
concentrations of VOC species; C and F: initial concentrations of VOC species)



Figure S5. The relationship between RI and OFP of different VOC species in 2015 and
2016. (A, D: measured and initial concentrations of VOCs in 2014, respectively; B, E:
measured and initial concentrations of VOCs in 2015, respectively; C, F: measured and
initial concentrations of VOCs in 2016, respectively)



178 Figure S6. The relationship between RI and OFP in 2019 summer in Daxing region\*

179 (A, B: measured and initial concentrations of VOCs in 2019, respectively).

180 \* The information of sampling site and VOCs characteristics can be seen in our previous

181 study (Zhan et al., 2021).



**Figure S7.** Time series of biogenic and anthropogenic VOCs during the observation period. (Biogenic VOCs: including isoprene,  $\alpha$ -pinene and  $\beta$ -





201 Figure S8. The relative error of simulated O<sub>3</sub> concentrations between the RF model and



202 the box model in 2015.

Figure S9. The relationship between xylene and ethylbenzene.



Figure S10. Comparison of the initial VOCs calculated using the ratio of xylene/ethylbenzene with that using the ratio of toluene/benzene in 2015. (Error bars are standard deviations.)

















Figure S11. The potential source contribution function (PSCF) maps for ethylbenzene (A and B), xylene (C and D), ratio of xylene to ethylbenzene (E and F), and isoprene (G and H) arriving in the observation site. The figures A, C, E and G are the results for the morning (05:00 and 06:00), and the figures of B, D, F and H are the results of the whole day (00:00-23:00).

217

218





Figure S12. The daily variation of VOCs concentration. (A and D for 2014; B and E

221 for 2015; C and F for 2016)





Figure S13. The relationship between min sample split and mean squared error (MSE).



Figure S14. The relationship between trees number and mean squared error (MSE).



Figure S15. Sensitivity curves of O<sub>3</sub> formation and distribution of training data in 2015.

	2014					20	15		2016			
species / unit	Measu	red VOC	Initial	I VOC	Measur	ed VOC	Initial	I VOC	Measur	ed VOC	Initial	VOC
	average	std. dev.*	average	std. dev.								
Cyclopentane / ppbv	0.95	1.05	0.95	1.05	0.00	0.00	0.00	0.00	0.27	0.29	0.27	0.29
Ethane / ppbv	2.38	0.98	2.39	0.98	1.84	0.88	1.85	0.89	1.07	0.51	1.07	0.51
Acetylene / ppbv	1.64	1.31	1.65	1.31	0.13	0.33	0.14	0.33	0.32	0.30	0.32	0.30
Propane / ppbv	2.44	1.60	2.46	1.61	2.42	1.75	2.45	1.76	1.35	0.93	1.36	0.93
Benzene / ppbv	0.60	0.44	0.61	0.44	0.47	0.35	0.47	0.36	4.59	4.23	4.64	4.29
iso-Butane / ppbv	0.95	0.66	0.96	0.67	0.35	0.53	0.35	0.54	0.24	0.18	0.24	0.19
2,2-Dimethylbutane / ppbv	0.00	0.01	0.00	0.01	0.00	0.02	0.00	0.02	0.00	0.00	0.00	0.00
n-Butane / ppbv	1.57	1.11	1.60	1.11	0.67	0.87	0.69	0.89	0.85	0.73	0.87	0.74
2,2,4-Trimethylpentane / ppbv	0.01	0.04	0.01	0.04	0.04	0.07	0.05	0.07	0.02	0.02	0.02	0.02
iso-Pentane / ppbv	0.11	0.38	0.11	0.40	0.00	0.00	0.00	0.00	0.16	0.18	0.16	0.18
2,3-Dimethylpentane / ppbv	0.07	0.08	0.07	0.08	0.06	0.08	0.06	0.08	0.02	0.03	0.02	0.03
3-Methylhexane / ppbv	0.06	0.07	0.06	0.07	0.04	0.05	0.04	0.05	0.01	0.02	0.01	0.02
Toluene / ppbv	1.28	1.02	1.32	1.04	0.88	1.55	0.93	1.57	0.30	0.34	0.32	0.37

Table S1. A	An overview o	f training	dataset	from	2014	to	2016	during	the o	observati	on p	period	Ι.
-------------	---------------	------------	---------	------	------	----	------	--------	-------	-----------	------	--------	----

2,3-Dimethylbutane / ppbv	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.06	0.08	0.06	0.08
n-Propyl benzene / ppbv	0.01	0.02	0.01	0.02	0.01	0.03	0.01	0.03	0.04	0.11	0.05	0.11
iso-Propyl benzene / ppbv	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.05	0.01	0.06
2,3,4-trimethylpentane / ppbv	0.12	0.29	0.12	0.31	0.06	0.10	0.06	0.11	0.01	0.02	0.02	0.02
n-hexane / ppbv	0.37	0.30	0.39	0.31	0.05	0.18	0.06	0.20	0.18	0.27	0.19	0.30
n-heptane / ppbv	0.08	0.09	0.09	0.10	0.06	0.06	0.06	0.07	0.02	0.02	0.02	0.02
2-methylhexane / ppbv	0.03	0.03	0.03	0.04	0.02	0.04	0.02	0.04	0.01	0.01	0.01	0.01
3-methylhexane / ppbv	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.00	0.01	0.00	0.01
cyclohexane / ppbv	0.04	0.05	0.05	0.05	0.03	0.05	0.04	0.05	0.04	0.10	0.04	0.12
ethylbenzene / ppbv	0.33	0.31	0.34	0.32	0.21	0.23	0.23	0.25	0.10	0.15	0.10	0.16
n-octane / ppbv	0.04	0.11	0.04	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ethene / ppbv	2.15	1.36	2.31	1.43	1.72	1.16	1.90	1.25	0.39	0.30	0.41	0.31
methylcyclohexane / ppbv	0.01	0.03	0.01	0.03	0.01	0.03	0.01	0.04	0.02	0.03	0.02	0.04
n-nonane / ppbv	0.03	0.04	0.03	0.04	0.02	0.02	0.02	0.03	0.02	0.04	0.02	0.04
n-decane / ppbv	0.02	0.04	0.02	0.05	0.02	0.03	0.02	0.03	0.00	0.01	0.00	0.01
p-ethyltoluene / ppbv	0.06	0.08	0.06	0.08	0.02	0.03	0.03	0.04	0.07	0.10	0.07	0.11
p-diethyl benzene / ppbv	0.01	0.04	0.01	0.04	0.01	0.02	0.01	0.02	0.09	0.17	0.11	0.22

o-ethyl toluene / ppbv	0.03	0.04	0.04	0.04	0.01	0.03	0.01	0.03	0.08	0.28	0.09	0.32
o-xylene / ppbv	0.09	0.18	0.10	0.18	0.16	0.18	0.19	0.20	0.14	0.26	0.15	0.27
m-ethyl toluene / ppbv	0.02	0.07	0.02	0.07	0.04	0.09	0.04	0.09	0.03	0.04	0.03	0.05
m-diethyl benzene / ppbv	0.01	0.03	0.01	0.03	0.00	0.01	0.00	0.01	0.00	0.02	0.00	0.02
m/p-Xylene / ppbv	0.61	0.64	0.68	0.65	0.45	0.51	0.54	0.59	0.22	0.38	0.25	0.41
propene / ppbv	2.07	1.18	2.83	2.26	4.40	2.61	6.60	6.12	0.28	0.41	0.34	0.45
1-Butene / ppbv	0.10	0.14	0.13	0.17	0.04	0.10	0.08	0.25	0.03	0.03	0.04	0.06
1-Pentene / ppbv	0.03	0.09	0.04	0.09	0.03	0.07	0.05	0.12	0.02	0.06	0.02	0.07
1,2,4-trimethyl benzene/ ppbv	0.01	0.08	0.01	0.08	0.08	0.09	0.11	0.12	0.05	0.05	0.06	0.09
1,2,3-trimethyl benzene/ ppbv	0.00	0.01	0.00	0.01	0.03	0.05	0.04	0.08	0.05	0.28	0.05	0.28
a-pinene / ppbv	0.01	0.03	0.02	0.03	0.01	0.03	0.01	0.03	0.18	0.46	0.84	3.48
cis-2-Butene / ppbv	0.34	0.70	0.85	2.67	0.66	0.85	1.77	4.56	0.04	0.05	0.11	0.29
1,3,5-Trimethylbenzene/ ppbv	0.05	0.07	0.08	0.11	0.03	0.05	0.07	0.14	0.25	0.56	1.07	4.11
styrene / ppbv	0.18	0.27	0.30	0.61	0.00	0.03	0.01	0.08	0.27	0.79	0.57	2.08
2-methyl-1-pentene / ppbv	0.18	0.37	0.72	2.94	0.04	0.04	0.26	1.68	0.02	0.09	0.03	0.12
trans-2-Butene / ppbv	0.08	0.16	0.24	1.15	0.09	0.11	0.34	0.74	0.02	0.02	0.04	0.08
cis-2-Pentene / ppbv	0.15	0.20	0.37	0.93	0.17	0.17	0.91	4.24	0.01	0.02	0.02	0.08

1,3-Butadiene / ppbv	0.09	0.10	0.19	0.34	0.04	0.05	0.12	0.38	0.02	0.03	0.05	0.25
trans-2-Pentene / ppbv	0.03	0.08	0.06	0.27	0.01	0.02	0.11	0.89	0.01	0.02	0.01	0.05
$\beta$ -pinene / ppbv	0.00	0.01	0.01	0.03	0.01	0.01	0.02	0.15	0.00	0.01	0.00	0.02
isoprene / ppbv	0.89	0.64	5.70	18.78	0.34	0.43	6.40	21.56	0.13	0.17	2.12	7.46
NO / ppbv	7.03	17.02	7.03	17.02	3.38	5.59	3.38	5.59	5.28	10.35	5.28	10.35
NO <sub>2</sub> / ppbv	15.50	15.79	15.50	15.79	19.11	12.68	19.11	12.68	18.72	12.40	18.72	12.40
T / °C	22.56	6.28	22.56	6.28	22.70	5.24	22.70	5.24	22.37	4.85	22.37	4.85
RH / %	50.93	23.88	50.93	23.88	41.49	23.23	41.49	23.23	36.23	21.58	36.23	21.58
SR / W m <sup>-2</sup>	162.92	222.95	162.92	222.95	153.29	205.01	153.29	205.01	150.81	199.35	150.81	199.35
WS / m s <sup>-1</sup>	3.11	2.70	3.11	2.70	2.29	2.15	2.29	2.15	1.25	1.24	1.25	1.24
WD / °	162.42	105.07	162.42	105.07	175.38	101.87	175.38	101.87	184.21	108.06	184.21	108.06
$PM_{2.5} / \mu g \ m^{-3}$	67.16	53.47	67.16	53.47	63.13	56.46	63.13	56.46	61.05	48.64	61.05	48.64
CO /mg m <sup>-3</sup>	0.78	0.49	0.78	0.49	0.68	0.44	0.68	0.44	0.57	0.36	0.57	0.36
O <sub>3</sub> / ppbv	44.32	32.38	44.32	32.38	42.74	27.94	42.74	27.94	44.01	29.64	44.01	29.64

232 \* Standard Deviation (std. Dev.)

Table S2. RF model parameters and input parameters											
_	RF model para	meters		RF model inputs							
_	Туре	Value	Туре	input parameter	output parameter						
	Cross-validation	12	Figure 2A	Total VOC concentration, PM <sub>2.5</sub> , NO, NO <sub>2</sub> , CO, SR, RH, WD, WS, T	O3						
	tree number	500	Figure 2B	51 measured VOC species, PM2.5, NO, NO2, CO, SR, RH, WD, WS, T	O3						
	min sample split	5									
	min sample leaf	1	Figure 2C	51 initial VOC species, PM <sub>2.5</sub> , NO, NO <sub>2</sub> , CO, SR, RH, WD, WS, T	O3						

Note: In this study, we optimized the number of tree and min sample split as shown in Figure S14 and S13, respectively. The min sample leaf 236

was set 1 (default), and other parameters were set to default (auto/none). 237

Min samples split: the minimum number of samples to be split; Min samples leaf: the minimum number of samples in a leaf; Trees number: the 238

number of trees during the training. 239

	201	4	201	5	2016			
specie	RI measured	RI initial	RI measured	RI initial	RI measured	RI initial		
NOx	7.06	8.08	9.24	14.79	10.96	11.01		
Т	8.10	8.12	6.60	9.76	15.79	15.14		
RH	13.04	13.36	3.02	3.78	1.79	1.64		
isoprene	2.74	4.11	5.16	5.28	9.37	6.49		
propene	2.46	3.41	4.78	5.92	0.82	1.02		
SR	3.39	2.82	3.02	3.68	2.36	2.63		
α-pinene	6.39	5.54	2.64	1.83	0.66	0.30		
СО	1.44	1.30	2.84	3.28	2.72	3.06		
Benzene	0.25	0.39	1.39	0.91	3.73	5.61		
Ethane	1.37	2.20	2.24	1.60	2.24	2.11		

Table S3. Changes in RIs using measured and initial concentrations of VOC species

Unit of RI: %

 Table S4. Independence test between RH and SR

		RI value	
name	RH and SR as input	RH as input	SR as input
RH	0.68	0.68	/
SR	0.76	/	0.76

			-	Ratio to the initial VOC***							
Seccio ayanhar		1*	MDI **	20	14	20	15	20	16		
specie number	species name	КОН	MDL	-10%	+10%	-10%	+10%	-10%	+10%		
				[OH]×t	[OH]×t	[OH]×t	[OH]×t	[OH]×t	[OH]×t		
1	Ethane	0.254	0.050	1.00	1.00	1.00	1.00	1.00	1.00		
2	Acetylene	0.756	0.022	1.00	1.00	1.00	1.00	1.00	1.00		
3	Propane	1.11	0.013	1.00	1.00	1.00	1.00	1.00	1.00		
4	Benzene	1.22	0.011	1.00	1.00	1.00	1.00	1.00	1.00		
5	iso-Butane	2.14	0.010	1.00	1.00	1.00	1.00	1.00	1.00		
6	2,2-Dimethylbutane	2.27	0.005	1.00	1.00	1.00	1.00	1.00	1.00		
7	n-Butane	2.38	0.011	1.00	1.00	1.00	1.00	1.00	1.00		
8	2,2,4-Trimethylpentane	3.38	0.008	1.00	1.00	1.00	1.00	1.00	1.00		
9	iso-Pentane	3.6	0.008	1.00	1.00	1.00	1.00	1.00	1.00		
10	Cyclopentane	5.02	0.005	1.00	1.00	1.00	1.00	1.00	1.00		
11	n-hexane	5.25	0.011	0.99	1.01	0.99	1.01	0.99	1.01		
12	Toluene	5.58	0.009	1.00	1.00	0.99	1.01	1.00	1.00		

Table S5. k<sub>OH</sub>, Method Detection Limit (MDL) and sensitivity test on estimation of [OH]×t of different VOC species

13	2,3-Dimethylbutane	5.79	0.004	1.00	1.00	1.00	1.00	0.99	1.01
14	n-Propyl benzene	5.8	0.008	1.00	1.00	1.00	1.00	0.99	1.01
15	iso-Propyl benzene	6.3	0.007	1.01	1.01	0.99	1.01	0.97	1.03
16	2,3,4-trimethylpentane	6.6	0.008	0.99	1.01	0.99	1.01	1.00	1.00
17	n-heptane	6.81	0.009	0.99	1.01	0.99	1.01	0.99	1.01
18	ethylbenzene	7	0.009	0.99	1.01	0.99	1.01	0.99	1.01
19	cyclohexane	7.02	0.011	1.00	1.00	0.99	1.01	0.99	1.01
20	2,3-Dimethylpentane	7.15	0.009	1.00	1.00	1.00	1.00	1.00	1.00
21	3-Methylhexane	7.17	0.009	1.00	1.00	0.99	1.01	1.00	1.00
22	ethene	8.15	0.021	0.99	1.01	0.99	1.01	0.99	1.01
23	n-octane	8.16	0.008	0.99	1.01	1.00	1.00	1.00	1.00
24	2-Methylheptane	8.31	0.008	1.00	1.00	0.99	1.01	0.99	1.01
25	3-Methylheptane	8.59	0.008	1.00	1.00	1.00	1.01	0.99	1.01
26	methylcyclohexane	9.64	0.005	0.99	1.01	0.99	1.01	0.99	1.01
27	n-nonane	9.75	0.006	0.99	1.01	0.99	1.01	0.98	1.02
28	n-decane	11	0.007	0.99	1.01	0.99	1.01	0.99	1.01
29	p-ethyl toluene	11.8	0.007	0.99	1.01	0.98	1.02	0.98	1.02

30	p-diethyl benzene	-	0.008	1.00	1.00	0.99	1.01	0.97	1.03
31	o-ethyl toluene	11.9	0.007	0.99	1.01	0.99	1.01	1.00	1.00
32	o-xylene	13.6	0.007	0.99	1.01	0.98	1.02	1.00	1.00
33	m-ethyl toluene	18.6	0.010	0.99	1.01	0.99	1.01	0.97	1.03
34	m-diethyl benzene	-	0.009	0.99	1.01	0.99	1.01	0.98	1.02
35	m/p-Xylene	23.1/14.2	0.008	0.99	1.01	0.98	1.02	0.98	1.03
36	propene	26	0.015	0.96	1.04	0.95	1.05	0.96	1.05
37	1-Butene	31.1	0.010	0.97	1.04	0.90	1.12	0.92	1.10
38	1-Pentene	31.4	0.009	0.98	1.02	0.93	1.09	0.93	1.08
39	1,2,4-trimethyl benzene	32.5	0.008	1.00	1.01	0.95	1.05	0.91	1.10
40	1,2,3-trimethyl benzene	32.7	0.009	0.96	1.04	0.96	1.04	0.97	1.03
41	a-pinene	51.8	0.010	0.97	1.04	0.96	1.05	0.75	1.35
42	cis-2-Butene	55.8	0.019	0.87	1.16	0.86	1.17	0.77	1.32
43	1,3,5-Trimethylbenzene	56.7	0.007	0.93	1.08	0.90	1.13	0.73	1.37
44	styrene	58	0.010	0.91	1.11	0.90	1.13	0.98	1.02
45	2-methyl-1-pentene	63	0.002	0.81	1.25	0.70	1.49	0.81	1.28
46	trans-2-Butene	63.2	0.014	0.84	1.22	0.82	1.25	0.76	1.35

47	cis-2-Pentene	65	0.006	0.86	1.19	0.74	1.42	0.83	1.24
48	1,3-Butadiene	65.9	0.024	0.88	1.16	0.82	1.26	0.87	1.18
49	trans-2-Pentene	67	0.009	0.88	1.16	0.63	1.63	0.75	1.38
50	β-pinene	73.5	0.010	0.90	1.12	0.81	1.26	0.92	1.10
51	isoprene	99.6	0.009	0.73	1.40	0.67	1.50	0.55	1.57

<sup>247</sup> \* Unit:  $10^{-12}$  cm<sup>3</sup> mole<sup>-1</sup> s<sup>-1</sup>. k<sub>OH</sub> values were under conditions of 300K. (Carter 2010)

<sup>248</sup> \*\* Unit: ppb. The relative standard derivations (RSDs) were within 10% for the target compounds in all six replicates.

249 \*\*\* All species were selected for sensitivity tests of initial VOCs to [OH]×t. The reaction rates of these species with OH covered the range of 51

250 VOCs and were characterized by low, medium and high k<sub>OH</sub> levels. The sensitivity test results showed that the uncertainty in the estimation of

initial VOCs caused by the [OH]×t estimation uncertainty ranged from 0.55 to 1.57.

252

253

254

**Table S6**. The RI values of 2015 in all 12 folds

						Fold	number					
species number	1	2	3	4	5	6	7	8	9	10	11	12
Ethane	1.39	1.15	1.69	2.28	1.70	1.70	1.97	1.80	1.20	1.35	1.71	1.26
ethene	1.60	1.97	1.01	1.01	1.52	2.15	0.65	2.01	1.44	2.18	2.17	1.64
Propane	0.24	0.85	0.38	1.20	1.23	1.18	0.49	0.01	1.19	0.54	0.28	1.61
propene	5.70	5.95	6.30	5.88	6.06	5.14	6.59	6.18	5.94	5.99	6.10	5.28
iso-Butane	1.45	0.96	1.58	1.55	1.67	1.22	1.77	1.97	1.26	1.80	1.64	1.56
n-Butane	0.05	1.08	0.35	0.32	0.35	0.61	0.27	0.24	0.60	0.05	0.09	0.05
Acetylene	1.50	1.11	1.01	1.03	0.66	0.00	0.96	0.96	0.23	1.05	1.02	1.81
trans-2-Butene	1.30	2.26	2.37	2.09	1.51	2.22	2.23	2.30	1.93	2.42	2.60	2.34
1-Butene	1.71	1.29	0.89	1.13	1.20	1.34	1.31	1.17	1.70	1.20	0.93	1.42
Cyclopentane	0.00	0.00	0.25	0.00	0.25	0.25	0.24	0.24	0.27	0.00	0.25	0.24
cis-2-Butene	0.86	1.05	1.33	1.09	1.43	1.65	1.65	1.23	1.60	0.81	1.44	1.15
iso-Pentane	0.00	0.00	0.30	0.00	0.25	0.25	0.25	0.24	0.23	0.00	0.23	0.27
1,3-Butadiene	0.58	1.31	0.86	1.10	0.85	1.17	0.69	0.99	1.36	0.91	1.08	0.93
trans-2-Pentene	0.99	0.75	1.26	0.78	1.16	1.18	1.03	1.01	0.69	1.17	1.12	1.23

1-Pentene	0.28	0.70	0.81	0.77	0.59	1.06	0.61	0.92	0.86	0.56	1.07	0.49
cis-2-Pentene	1.29	1.37	1.16	1.21	1.26	0.84	1.23	1.76	1.55	1.09	1.25	0.98
2,3-Dimethylbutane	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.00	0.26	0.00	0.00
n-hexane	1.27	1.42	1.69	1.46	1.42	1.47	1.30	1.62	0.82	2.00	1.08	0.91
isoprene	5.30	4.16	4.96	5.25	4.98	5.39	5.98	5.25	5.44	5.06	5.34	6.24
2-methyl-1-pentene	2.27	2.51	2.08	1.93	2.36	2.09	2.10	1.74	2.35	2.28	2.26	2.17
2,2-Dimethylbutane	0.78	0.63	0.65	0.68	0.87	0.63	0.41	0.57	0.81	0.71	0.63	0.60
Benzene	0.35	0.79	0.96	1.37	0.84	1.12	0.86	0.83	0.80	1.10	1.10	0.79
cyclohexane	1.71	1.40	1.18	1.20	1.46	1.06	1.55	1.09	0.86	1.21	1.22	1.61
2,3-Dimethylpentane	2.26	1.61	2.39	1.65	2.23	1.93	1.68	1.56	1.89	2.16	2.01	2.21
3-Methylhexane	1.05	0.36	0.63	0.98	0.48	0.92	0.84	0.69	0.54	0.45	0.90	0.66
2,2,4-Trimethylpentane	1.59	1.64	1.62	1.72	1.28	1.33	1.46	1.59	1.85	2.07	1.84	1.71
n-heptane	1.19	1.13	1.15	1.16	1.36	0.97	0.95	0.60	1.06	1.27	1.17	0.83
methylcyclohexane	0.34	0.56	0.97	0.69	0.50	0.55	0.75	0.42	0.57	0.70	0.81	0.86
2,3,4-trimethylpentane	0.76	0.76	0.87	1.11	0.89	0.90	0.81	0.67	0.95	0.81	0.88	0.97
Toluene	0.79	0.47	0.63	0.50	0.75	0.39	0.49	0.64	0.91	0.71	0.68	0.52
2-methylhexane	1.56	1.69	1.39	1.32	1.61	1.73	1.37	1.50	1.02	1.93	1.08	1.25

3-methylhexane	0.47	0.31	0.51	0.83	0.79	0.70	0.54	0.78	0.47	0.50	0.67	0.61
n-octane	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ethylbenzene	0.73	0.81	1.10	0.83	0.48	0.72	0.51	0.61	0.68	0.79	0.79	0.92
m,p-Xylene	0.52	0.31	0.83	0.22	0.32	0.24	0.45	0.36	0.01	0.53	0.57	0.22
styrene	0.31	0.41	0.23	0.27	0.25	0.26	0.33	0.33	0.37	0.17	0.29	0.28
o-xylene	0.59	0.83	0.88	0.61	0.62	0.98	0.63	0.64	0.86	0.83	0.66	0.30
n-nonane	1.19	1.30	1.38	1.68	1.34	1.57	1.30	1.14	1.48	1.27	1.27	1.73
iso-Propylbenzene	0.33	0.13	0.49	0.32	0.51	0.25	0.29	0.11	0.01	0.34	0.39	0.22
a-pinene	2.10	1.47	1.67	2.02	1.87	1.97	1.77	2.43	1.53	1.72	1.60	1.83
n-Propylbenzene	2.03	1.91	1.88	2.15	1.89	2.24	1.69	1.86	1.69	1.88	1.91	2.23
m-ethyl toluene	1.54	0.95	0.91	1.08	0.79	0.90	0.64	1.02	0.99	0.97	0.64	0.95
p-ethyltoluene	0.55	0.64	1.22	0.74	1.06	0.98	0.74	1.01	0.73	0.84	0.59	0.53
1,3,5-Trimethylbenzene	0.80	1.30	0.80	1.06	0.78	1.01	1.35	0.97	1.18	1.18	0.86	1.00
o-ethyl toluene	1.00	0.98	0.98	0.73	0.71	0.91	1.09	0.59	0.84	0.60	0.88	1.03
β-pinene	0.20	0.06	0.23	0.01	0.25	0.14	0.02	0.46	0.05	0.04	0.25	0.01
1,2,4-trimethyl benzene	1.38	1.67	1.54	1.65	2.69	1.34	1.69	1.56	1.03	1.43	1.60	0.91
n-decane	1.60	1.84	2.02	0.93	1.52	1.88	1.91	1.58	1.97	1.53	1.93	1.58

1,2,3-trimethyl benzene	1.14	1.05	1.80	1.22	0.96	1.70	1.37	1.11	0.67	0.88	1.53	1.32
m-diethyl benzene	0.26	0.00	0.25	0.26	0.24	0.36	0.00	0.34	0.00	0.26	0.00	0.00
p-diethyl benzene	1.19	0.95	0.95	0.69	0.92	0.80	0.79	0.92	0.80	0.97	0.75	0.86
NO <sub>x</sub>	16.35	16.32	14.22	15.74	15.93	11.94	14.37	14.00	15.79	14.78	13.75	14.23
Т	10.25	9.69	9.45	9.69	9.18	10.77	10.61	10.74	9.81	8.76	8.71	9.50
RH	3.93	4.27	3.36	4.31	3.85	3.68	3.63	3.62	3.96	3.61	3.11	4.10
SR	3.35	3.64	3.94	3.86	3.34	3.73	3.77	3.72	3.44	3.30	4.25	3.83
WS&WD	4.28	3.08	4.29	3.15	2.63	3.59	3.42	4.05	4.13	3.52	4.63	3.71
PM <sub>2.5</sub>	1.33	1.23	1.00	1.24	2.46	1.78	1.81	1.03	2.20	2.99	1.79	2.00
СО	3.00	4.03	2.97	2.72	2.91	3.88	3.78	3.64	3.41	3.05	3.54	2.49

257 Note: To verify the consistency between the different folds, we calculated the Pearson correlation coefficient (*r*) between the different folds as

shown in **Table S7**.

Fold number	1	2	3	4	5	6	7	8	9	10	11	12
1	1.00	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.98	0.98	0.97	0.98
2	0.98	1.00	0.98	0.98	0.97	0.95	0.97	0.97	0.98	0.98	0.96	0.96
3	0.98	0.98	1.00	0.98	0.97	0.96	0.98	0.98	0.98	0.98	0.98	0.97
4	0.98	0.98	0.98	1.00	0.98	0.96	0.98	0.98	0.98	0.97	0.97	0.98
5	0.98	0.97	0.97	0.98	1.00	0.96	0.98	0.97	0.98	0.98	0.97	0.97
6	0.95	0.95	0.96	0.96	0.96	1.00	0.98	0.98	0.96	0.95	0.97	0.96
7	0.98	0.97	0.98	0.98	0.98	0.98	1.00	0.98	0.98	0.97	0.98	0.98
8	0.98	0.97	0.98	0.98	0.97	0.98	0.98	1.00	0.98	0.97	0.98	0.97
9	0.98	0.98	0.98	0.98	0.98	0.96	0.98	0.98	1.00	0.98	0.97	0.98
10	0.98	0.98	0.98	0.97	0.98	0.95	0.97	0.97	0.98	1.00	0.97	0.97
11	0.97	0.96	0.98	0.97	0.97	0.97	0.98	0.98	0.97	0.97	1.00	0.97
12	0.98	0.96	0.97	0.98	0.97	0.96	0.98	0.97	0.98	0.97	0.97	1.00

## REFERENCE

- Carter, W.P.L. Development of the SAPRC-07 chemical mechanism. Atmos. Environ., 44, 5324-5335, https://doi.org/10.1016/j.atmosenv.2010.01.026, 2010.
- Gao, J., Zhang, J., Li, H., Li, L., Xu, L., Zhang, Y., Wang, Z., Wang, X., Zhang, W., Chen, Y., Cheng, X., Zhang, H., Peng, L., Chai, F., Wei, Y. Comparative study of volatile organic compounds in ambient air using observed mixing ratios and initial mixing ratios taking chemical loss into account A case study in a typical urban area in Beijing. Sci. Total Environ., 628-629, 791-804, https://doi.org/10.1016/j.scitotenv.2018.01.175, 2018.
- Parrish, D.D., Stohl, A., Forster, C., Atlas, E.L., Blake, D.R., Goldan, P.D., Kuster, W.C., de Gouw, J.A. Effects of mixing on evolution of hydrocarbon ratios in the troposphere. J. Geophys. Res. Atmos., 112, https://doi.org/10.1029/2006JD007583, 2007.
- Shao, M., Bin, W., Sihua, L., Bin, Y., Ming, W. Effects of Beijing Olympics Control Measures on Reducing Reactive Hydrocarbon Species. Environ. Sci. Technol., 45, 514-519, 10.1021/es102357t, 2011.
- 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.