1 Dear Reviewer,

We appreciate your careful consideration of our manuscript. We have carefully responded to all of your **point-by-point** comments and issues and have revised the manuscript accordingly. These revisions are described in detail below.

5

6 **Reviewer 1#**

7 This is an interesting study, using machine learning to estimate the ozone formation 8 sensitivity. The idea is not novel (a few previous studies with similar scope are cited in 9 this manuscript). The method, using reactivity-corrected VOC measurements (i.e., 10 initial VOC concentrations), sheds some insights into ozone production in an urban 11 environment.

Reply: Thanks for your positive comments. We have carefully responded to all of your
point-by-point comments and issues and have revised the manuscript accordingly.
These revisions are described in detail below.

15

16 However, there are several major issues:

Q1: (1) The machine learning workflow described in this manuscript does not include a robust or systematic solution to mitigate overtraining. I will elaborate on this later but the measures described in this work absolutely do not guarantee that overtraining is/can be avoided.

Reply: Thank you for your good suggestion. According to your suggestion, we 21 performed a 12-fold cross-validation after data-normalization, i.e., by randomly 22 dividing the dataset into 12 subsets and alternately taking one subset as testing data and 23 24 the rest as training data. By doing this, every data point has an equal chance being trained and tested. In lines 148-153 in the revised manuscript, we added a short 25 paragraph "To avoid over-fitting, we trained the random forest model using cross-26 validation for the normalized data, which can improve the robustness of the model. 27 Briefly, we randomly divided the normalized data into 12 subsets, then alternately took 28 29 one subset as testing data along with the rest as training data. By doing this, every data point has an equal chance being trained and tested.". 30

31 We added the RF model workflow to Text S3 in the revised Supporting 32 Information.

33 "Text S3. Workflow of RF model and the calculation of Relative Importance (RI)
34 The workflow of RF model used in this study was established through the following
35 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_{2.5} and
meteorological parameters (including temperature, relative humidity, solar radiation,
wind speed and direction) were used as input variables and O₃ 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.

53 (4) Model uncertainty estimation. The uncertainty of the model was estimated 54 according to the predicted and observed O₃ concentrations. The performance of the 55 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, 61 the first step is to identify the out-of-bag observations and the indices of the predictor 62 63 variables that are split to a growing tree t (from 1 to T). Then, one can estimate the outof-bag error (ε_t) for each tree. For a predictor variable x_i (j: from 1 to p), one can estimate 64 the model error $(\varepsilon_{t, j})$ again corroding to the out-of-bag observations after randomly 65 permuting the observations of x_j. Thus, the difference of the model error $(d_{t,j} = \varepsilon_{t,j} - \varepsilon_t)$ 66 is obtained. If the predictor variables are not split, the difference of a growing tree t is 67 0. The second step is to calculate the mean difference of the model errors (\overline{d}_j) , and the 68 standard deviation (σ_i) of the differences for all the learners and each predictor variable 69 in the training data. Finally, the out-of-bag relative importance (RI) for x_i is calculated 70 by dividing the difference of the model errors by the standard deviation $(\overline{d}_j/\sigma_j)$. 71 72 (6) EKMA curves. The Empirical Kinetic Modeling Approach (EKMA) was used to 73 assess the O₃ 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. 74 75 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 76 NOx were varied $\pm 10\%$ (or from 90% to 110%) of their mean values with a step of 1% 77 in a two-dimensional matrix along with other inputs unchanged. This matrix was used 78 79 as the testing data, while all the measured data were taken as the training data in the RF model to simulate O3 concentrations under different scenarios of VOCs and NOx 80 concentrations. To decrease the model uncertainty, we set relatively small variations of 81 82 VOCs and NOx $(\pm 10\%)$ compared to the observed values in this study. The mean 83 relative error of simulated O3 concentrations between RF model and Box model (within 15.6%, Figure S8) suggests that the RF model can well predict O₃ concentrations during 84 our observations." 85

Q2: (2) Random forest depends heavily on the training dataset. The authors do not provide an overview of the comprehensiveness of the training dataset: for instance, does the dataset cover all major chemical regimes in the EKMA plot, i.e., NOx-limited,

90 VOC-limited, NO titration? The authors claim that ozone production in Beijing, China 91 is mostly VOC-limited, which is consistent with previous studies. If the training set 92 collected in Beijing does not have sufficient coverage in the NOx-limited regime, then 93 the trained algorithm essentially attempts to extrapolate in that regime, which is 94 dangerous and prone to overtraining. I would then question the if this random forest 95 model can make meaningful forecast in that regime at all.

96 **Reply**: Thank you for your valuable suggestion. We added the description of training 97 dataset in Text S3 in the revised SI. This point has been replied in the aforementioned 98 question. The mean values (±standard deviation) of the input and output parameters for 99 the training data set are shown in Table R1. This Table was also added as Table S1 in 100 the revised SI.

period.

- 101 Table R1. An overview of training dataset from 2014 to 2016 during the observation
- 102

2014 2015 2016 Measured Measured Measured Initial VOC Initial VOC Initial VOC species / unit VOC VOC VOC std. aver std. std. aver std. std. aver std. aver aver aver age dev.* age dev. dev. age dev. dev. dev. age age age 0.95 1.05 0.95 1.05 0.00 0.00 0.27 0.29 0.27 0.29 Cyclopentane 0.00 0.00 ppbv 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 1.31 Acetylene / ppbv 0.13 0.33 0.14 0.33 0.32 0.30 0.32 0.30 1.64 1.31 1.65 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 0.35 0.36 4.59 4.23 0.44 0.61 0.47 0.47 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 0.02 0.02 2,2-0.00 0.01 0.00 0.01 0.00 0.00 0.00 0.00 0.00 0.00 Dimethylbutane / ppbv 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 2,2,4-0.04 0.04 0.07 0.07 0.02 0.01 0.01 0.04 0.05 0.02 0.02 0.02 Trimethylpentane / ppbv 0.38 0.00 0.00 0.00 0.18 iso-Pentane 0.11 0.11 0.40 0.00 0.16 0.16 0.18 ppbv 2,3-0.07 0.08 0.07 0.08 0.06 0.08 0.06 0.08 0.02 0.03 0.02 0.03 Dimethylpentane / ppbv

3-Methylhexane /	0.06	0.07	0.06	0.07	0.04	0.05	0.04	0.05	0.01	0.02	0.01	0.02
ppbv												
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
2,3-	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.06	0.08	0.06	0.08
Dimethylbutane /												
ppbv												
n-Propyl benzene	0.01	0.02	0.01	0.02	0.01	0.03	0.01	0.03	0.04	0.11	0.05	0.11
/ ppbv												
iso-Propyl	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.05	0.01	0.06
benzene / ppbv												
2,3,4-	0.12	0.29	0.12	0.31	0.06	0.10	0.06	0.11	0.01	0.02	0.02	0.02
trimethylpentane /												
ppbv												
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 /	0.03	0.03	0.03	0.04	0.02	0.04	0.02	0.04	0.01	0.01	0.01	0.01
ppbv												
3-methylhexane /	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.00	0.01	0.00	0.01
ppbv												
cyclohexane /	0.04	0.05	0.05	0.05	0.03	0.05	0.04	0.05	0.04	0.10	0.04	0.12
ppbv												
ethylbenzene /	0.33	0.31	0.34	0.32	0.21	0.23	0.23	0.25	0.10	0.15	0.10	0.16
ppbv												
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
methylcyclohexa	0.01	0.03	0.01	0.03	0.01	0.03	0.01	0.04	0.02	0.03	0.02	0.04
ne / ppbv												
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 /	0.06	0.08	0.06	0.08	0.02	0.03	0.03	0.04	0.07	0.10	0.07	0.11
ppbv												
p-diethyl benzene	0.01	0.04	0.01	0.04	0.01	0.02	0.01	0.02	0.09	0.17	0.11	0.22
/ ppbv												
o-ethyl toluene /	0.03	0.04	0.04	0.04	0.01	0.03	0.01	0.03	0.08	0.28	0.09	0.32
ppbv												
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 /	0.02	0.07	0.02	0.07	0.04	0.09	0.04	0.09	0.03	0.04	0.03	0.05
ppbv												
m-diethyl	0.01	0.03	0.01	0.03	0.00	0.01	0.00	0.01	0.00	0.02	0.00	0.02
benzene / ppbv												
m/p-Xylene /	0.61	0.64	0.68	0.65	0.45	0.51	0.54	0.59	0.22	0.38	0.25	0.41
ppbv					-					- *	-	
propene / ppby	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	0.01	0.08	0.01	0.08	0.08	0.09	0.11	0.12	0.05	0.05	0.06	0.09
benzene/ ppbv												
1,2,3-trimethyl	0.00	0.01	0.00	0.01	0.03	0.05	0.04	0.08	0.05	0.28	0.05	0.28
benzene/ ppbv												
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 /	0.34	0.70	0.85	2.67	0.66	0.85	1.77	4.56	0.04	0.05	0.11	0.29
ppbv												
1,3,5-	0.05	0.07	0.08	0.11	0.03	0.05	0.07	0.14	0.25	0.56	1.07	4.11
Trimethylbenzene												
/ ppbv												
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-	0.18	0.37	0.72	2.94	0.04	0.04	0.26	1.68	0.02	0.09	0.03	0.12
pentene / ppbv												
trans-2-Butene /	0.08	0.16	0.24	1.15	0.09	0.11	0.34	0.74	0.02	0.02	0.04	0.08
ppbv												
cis-2-Pentene /	0.15	0.20	0.37	0.93	0.17	0.17	0.91	4.24	0.01	0.02	0.02	0.08
ppbv												
1,3-Butadiene /	0.09	0.10	0.19	0.34	0.04	0.05	0.12	0.38	0.02	0.03	0.05	0.25
ppbv												
trans-2-Pentene /	0.03	0.08	0.06	0.27	0.01	0.02	0.11	0.89	0.01	0.02	0.01	0.05
ppbv												
β -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	18.7	0.34	0.43	6.40	21.5	0.13	0.17	2.12	7.46
				8				6				
NO / ppbv	7.03	17.02	7.03	17.0	3.38	5.59	3.38	5.59	5.28	10.3	5.28	10.3
				2						5		5
NO ₂ / ppbv	15.5	15.79	15.5	15.7	19.1	12.6	19.1	12.6	18.7	12.4	18.7	12.4
	0		0	9	1	8	1	8	2	0	2	0
T/°C	22.5	6.28	22.5	6.28	22.7	5.24	22.7	5.24	22.3	4.85	22.3	4.85
	6		6		0		0		7		7	
RH / %	50.9	23.88	50.9	23.8	41.4	23.2	41.4	23.2	36.2	21.5	36.2	21.5
	3		3	8	9	3	9	3	3	8	3	8
SR / W m ⁻²	162.	222.9	162.	222.	153.	205.	153.	205.	150.	199.	150.	199.
1	92	5	92	95	29	01	29	01	81	35	81	35
WS / m s ⁻¹	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.	105.0	162.	105.	175.	101.	175.	101.	184.	108.	184.	108.
	42	7	42	07	38	87	38	87	21	06	21	06
PM _{2.5} /μg m ⁻³	67.1	53.47	67.1	53.4	63.1	56.4	63.1	56.4	61.0	48.6	61.0	48.6
	6	0.40	6	7	3	6	3	6	5	4	5	4
CO /mg m ⁻³	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 ₃ / ppbv	44.3	32.38	44.3	32.3	42.7	27.9	42.7	27.9	44.0	29.6	44.0	29.6
	2		2	8	4	4	4	4	1	4	1	4

103 * Standard Deviation (std. Dev.)

In Text S3 in the revised SI, we added a short paragraph "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, $PM_{2.5}$ and meteorological parameters (including temperature, relative humidity, solar radiation, wind speed and direction) were used as input variables and O₃ was the output variable. The mean values (±standard deviation) of input/output parameters are shown in Table S1"

111 As shown in Figure R1 or Figure S15, the training dataset were located in VOClimited, NOx-limited, and transition regimes, while most of the training data were 112 located in the VOC-limited regime. To avoid overtraining, we performed a 12-fold 113 cross-validation, i.e., by randomly dividing the observed data into 12 subsets and 114 alternately taking one subset as testing data and the rest as training data, to ensure that 115 each data point has an equal chance of being trained and tested. Figure R2 (Figure 2 in 116 the revised manuscript) shows the comparisons between the measured and predicted O₃ 117 concentrations using different VOC inputs. The curves of the predicted O₃ 118 119 concentrations were spliced using the testing datasets in all runs. Thus, both the training data and the testing data actually covered all the sensitivity regimes of O₃ formation. 120 We think that the model is robust in the revised version according to your good 121 122 suggestion.



124 **Figure R1**. Sensitivity curves of O₃ formation and distribution of training data in 2015.



125

Figure R2. Comparison of the predicted and measured O₃ concentrations in Beijing in the summer of 2015. (A and D: TVOC concentrations; B and E: measured concentrations of VOC species; C and F: initial concentrations of VOC species). The testing data curves were spliced using the testing datasets in all runs during the 12-fold

130 cross-validation.

131

When plotting the O₃ formation sensitivity curves, we made a virtual matrix of 132 inputs by varying the concentrations of NOx and VOCs from 0.9 to 1.1 times (with a 133 step of 0.01) of their mean values (observed point data) while keeping all other inputs 134 unchanged (i.e., the mean values during our observations). Then, the new matrix was 135 used as testing data, while all the measured data were taken as training data. Thus, the 136 137 testing data should represent the mean sensitivity regime of O₃ in Beijing, while the training data actually covered all the sensitivity regimes of O₃ formation. As shown in 138 Figure R3 or Figure 4B, the sensitivity of O₃ formation was located in VOC-limited 139 regime in 2015. The mean relative errors of simulated O₃ between RF model and Box 140 model was 15.6% (Figure R4 or Figure S8), which means that the RF model can well 141 predict the O₃ concentrations and the sensitivity regime of O₃ formation. Although we 142 set relatively small variations of VOCs and NOx ($\pm 10\%$) compared to the observed 143 values to decrease the model uncertainty when depicting the EKMA curves, the training 144 145 data represent the real conditions in Beijing during our observations. Therefore, we think our results should be reliable and meaningful. 146



Figure R3. 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)



151

152 Figure R4. The relative error of simulated O₃ concentrations between the RF model

and the box model in 2015.

In lines 232-240 the revised manuscript, we added a paragraph "It should be 154 155 pointed out that if the training dataset does not have sufficient coverage in the NOxlimited regime, then the trained algorithm essentially attempts to extrapolate in that 156 regime, which is prone to overtraining. To avoid such overtraining, a 12-fold cross-157 158 validation by randomly dividing the observation data in each day into 12 subsets and alternately taking one subset as testing data and the rest as training data ensures that 159 160 each data point has an equal chance of being trained and tested. The curves of the predicted O₃ concentrations in Figure 2 were spliced using the testing datasets in all 161 runs. Thus, our results actually covered all the sensitivity regimes of O₃ formation. This 162 means that the model is robust". 163

In lines 171-178 in the revised manuscript, we added a paragraph "When plotting 164 the O₃ formation sensitivity curves, we made a virtual matrix of inputs by varying the 165 concentrations of NOx and VOCs from 0.9 to 1.1 times (with a step of 0.01) of their 166 mean values while keeping all other inputs unchanged (i.e., the mean values). Then, the 167 168 new matrix was used as testing data, while all the measured data were taken as training data. Thus, the testing data should represent the mean sensitivity regime of O₃ in Beijing, 169 while the training data actually covered all the sensitivity regimes of O₃ formation to 170 171 guarantee a sufficient coverage in the NOx-limited regime for the RF model 172 simulations".

173

Q3: (3) The calculation of the initial VOC concentrations is problematic: the method 174 depends heavily on initial/source ratio which is not discussed at all in this work; the 175 176 method assumes biogenic VOCs share the same air mass histories as the anthropogenic VOCs which is not supported by any evidence. For these reasons, I do not recommend 177 the current form of the manuscript for publication in Atmospheric Measurement and 178 Techniques. Given the substantial amount of work needed to demonstrate the 179 robustness of the machine learning workflow, to outline key details in a transparent 180 181 manner, and to revise the initial VOC calculation, resubmission is recommended. Please see my specific and minor/technical comments below. 182

Reply: Thank you so much for your good suggestions. In previous studies (Shao et al., 2011; Zhan et al., 2021), it has been justified for selecting the pair of ethylbenzene/xylene as the tracers when calculating ambient OH exposure in terms of the following rules: 1) the concentrations of xylene and ethylbenzene are well correlated, which indicates that they are simultaneously emitted; 2) they have different degradation rates in the atmosphere; 3) the calculated initial VOCs are in good agreement with those calculated using other tracers, such as toluene/benzene.

190 As shown in Figure R5 or Figure S9, the concentrations of xylene and ethylbenzene correlated well during our observations in this work. In addition, we 191 compared the ratio of the initial concentrations calculated according to the ratio of 192 xylene/ethylbenzene with that using the ratio of toluene/benzene (Figure R6 or S9). 193 Except for several compounds, the ratio of the PICs for most of these VOCs varied 194 within 1.0±0.1. This means the calculated photochemical initial concentrations (PICs) 195 are in good agreement when using different tracers. Sensitivity tests showed that the 196 uncertainty caused by the OH exposure (from -10% to +10%) ranged from 0.55 to 1.57 197 198 (Table R2 or Table S4). Figure R7 or Figure S12 shows the calculated diurnal curves of the PICs from 2014 to 2016. Photochemical losses of VOCs occurred prominently 199 during the daytime. 200



Figure R5. The relationship between xylene and ethylbenzene.





Figure R6. 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.)





212

213 Table R2. koh, Method Detection Limit (MDL) and sensitivity test on estimation of

214 [OH]×t of different VOC species

Speci				Ratio to the initial VOC***							
Speer				2014		2015		2016			
e	species name	kon*	MDL**	-10%	+10%	-10%	+10%	-10%	+10%		
numb				[OH]×	[OH]×	[OH]×	[OH]×	[OH]×	[OH]×		
er				t	t	t	t	t	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		
(2,2-	2.27	0.005	1.00	1.00	1.00	1.00	1.00	1.00		
0	Dimethylbutane	2.27		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		
0	2,2,4-	2 2 2	0.009	1.00	1.00	1.00	1.00	1.00	1.00		
8	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
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	0.008	0.99	1.01	0.98	1.02	0.98	1.03

		.2							
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
20	1,2,4-trimethyl	22.5		1.00	1.01	0.05	1.05	0.91	1 10
39	benzene	32.5	0.008	1.00	1.01	0.95	1.05		1.10
40	1,2,3-trimethyl	22.7	0.000	0.06	1.04	0.06	1.04	0.07	1.02
40	benzene	32.7	0.009	0.96	1.04	0.96	1.04	0.97	1.05
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
42	1,3,5-	567	0.007	0.02	1.09	0.00	1 1 2	0.73	1 27
43	Trimethylbenzene	30.7	0.007	0.95	1.00	0.90	1.15	0.75	1.57
44	styrene	58	0.010	0.91	1.11	0.90	1.13	0.98	1.02
15	2-methyl-1-	62	0.002	0.91	1 25	0.70	1.40	0.81	1 20
45	pentene	03	0.002	0.01	1.23	0.70	1.49	0.81	1.20
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

215 * Unit: 10⁻¹² cm³ mole⁻¹ s⁻¹. k_{OH} values were under conditions of 300K. (Carter 2010)

*** Unit: ppb. The relative standard derivations (RSDs) were within 10% for the target compounds in all six replicates.
*** 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 VOCs and were characterized by low, medium and high k_{OH} 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.

221

Potential source contribution function (PSCF) analysis has been further carried out to evaluated the possible influence of air mass on the emission ratio of ethylbenzene

and xylene. As shown in Figure R8A-D or Figure S11, xylene showed similar pattern 224 225 to ethylbenzene in the early morning or in the whole day. These results indicate that variations of air mass should have little influence on their initial ratio. In addition, 226 isoprene showed similar patterns to that of xylene and ethylbenzene (Figure R8G-H), 227 which means VOC emissions are evenly distributed in Beijing. This can be ascribed to 228 the fact that our observation site is a typical urban station. Although isoprene and 229 xylene/ethylbenzene are from biogenic sources and anthropogenic sources, both them 230 231 are non-point sources on a city scale.

















233 Figure R8. 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

235 (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

237 whole day (00:00-23:00).

238

In the revised SI, we added the details on calculation of the initial VOCs and more
discussions in the Text S2. The changes are shown below.

241 "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).

246
$$\int c_{\text{OH}} dt = \frac{1}{k_{\text{A,OH}} - k_{\text{B,OH}}} \left[\ln\left(\frac{\text{VOC}_{\text{A}}}{\text{VOC}_{\text{B}}}\right)_{\text{initial}} - \ln\left(\frac{\text{VOC}_{\text{A}}}{\text{VOC}_{\text{B}}}\right) \right]$$
(S1)

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

248
$$VOC_{i, initial} = \frac{VOC_{i}}{exp(-k_{i,OH})exp(\int c_{OH}dt)}$$
(S2)

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

250
$$\operatorname{VOC}_{i, \text{ initial}} = \frac{\operatorname{VOC}_{i}}{\exp(-k_{i, \text{ OH}}) \exp(\frac{1}{k_{A, \text{OH}} - k_{B, \text{ OH}}} \left[\ln(\frac{\operatorname{VOC}_{A}}{\operatorname{VOC}_{B}})_{\text{ initial}} - \ln(\frac{\operatorname{VOC}_{A}}{\operatorname{VOC}_{B}})\right])}$$
(S3)

Where C_{OH} represents the ambient OH concentration; $k_{A,OH}$ and $k_{B,OH}$ represent the reaction rate of compound A and B with OH radical, respectively; t represents the 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).

261 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 262 well correlated, which indicates that they are simultaneously emitted. In addition, we 263 264 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 265 from 0.5 to 1.5 with a mean value of 0.96. This means the calculated initial VOCs was 266 in good agreement when using different tracers. The mean ratio (0.52, from 0.45 to 0.66) 267 of ethylbenzene/xylene before sunrise was taken as the initial ratio of 268 269 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 S4). 270

Variations of air mass may also affect the VOC ratio. Figure S11 A-D shows the 271 mean concentration distribution of ethylbenzene and xylene in the early morning and 272 the whole day based on potential source contribution function (PSCF) analysis. Xylene 273 showed similar patterns to ethylbenzene in different air mass trajectories and different 274 periods. These results indicate that the emissions of xylene and ethylbenzene were 275 constant throughout the day and variations of air mass should have little influence on 276 277 the initial ratio of VOCs. The hourly concentrations of ethylbenzene and xylene were used to calculate the concentration of initial VOCs. The initial VOC was calculated by 278 adding the measured VOC concentration and the calculated photochemical loss. Figure 279 S12 shows the diurnal variations of the observed and initial VOCs concentrations from 280 2014 to 2016. Photochemical loss of VOC occurred mainly during the daytime. 281

It should be noted that the lifetimes $(1/k_2c_{OH})$ of highly reactive VOCs, such as 282 283 isoprene, greatly depend on the OH exposure. The photochemical ages of isoprene were 0.01–6.21 h (1.26 \pm 1.12 h). This value is comparable with previously reported 284 285 photochemical ages (Shao et al., 2011; Gao et al., 2018). However, the initial concentrations of highly reactive VOCs may be overestimated due to their short 286 lifetimes and should be taken as the upper limits. On the other hand, isoprene is a 287 biogenic VOC, while xylene and ethylbenzene are anthropogenic VOCs. If they do not 288 289 share the same air mass histories, an additional uncertainty is inevitable for the PICs of 290 isoprene. As shown in Figure S11, isoprene showed similar patterns to that of xylene and ethylbenzene, which means VOC emissions are evenly distributed in Beijing during 291

our observations. This can be ascribed to the fact that our observation site is a typical 292 urban station. Although isoprene and xylene/ethylbenzene different sources, both them 293 294 are non-point sources on a city scale. Therefore, the photochemical clock calculated using xylene and ethylbenzene is able to correct the photochemical loss of biogenic 295 VOCs to some extent. It should be noted that uncertainty is inevitable when we 296 estimating the photochemical age (Parrish et al., 2007). However, the aim of this work 297 is to test whether the ML-model can reflect the influence of photochemical loss of 298 299 VOCs species on O₃ modelling. The PICs should provide additional information for understanding O₃ formation in the atmosphere.". 300

As for the robustness and workflow of RF model, we have replied this point in Q1
and Q2 and revised the manuscript and the SI accordingly.

303

304 Specific concerns:

305 Q4: Line 70-71: This is a valid concern. However, the machine learning based
306 approaches are also subject to this.

Reply: Thank you for your comment. We have added a paragraph to point out the shortcoming of machine learn "Although attentions should be paid to the robustness of machine learning because it depends on input dataset (observations or outputs from chemical transport models), previous studies have demonstrated that cross-validation and data-normalization can well reduce the dependence of the model on input data and improve the robustness of the model (Wang et al., 2016; Wang et al., 2017b; Liu et al.,

313 2021; Ma et al., 2021)" in Lines 79-83 in the revised manuscript.

314

Q5: Line 72-73: Respectfully, I disagree. Box models using condensed mechanisms are usually quite cheap. Near-explicit mechanisms such as MCM are more expensive, but the EKMA-type configurations are still considerably cheaper than 3D chemical transport models. Well-developed box models with MCM or other condensed mechanisms configured for ozone sensitivity (EKMA) can run on personal computers or small servers/clusters, providing timely predictions with no major demand for computational resources. OBMs generally are not considered as being time-consuming 322 or computationally expensive.

Reply: Thank you. We have deleted the sentences "In addition, both of them are timeconsuming and expensive when computational resources are considered" and "Traditional models have difficulty assessing O₃ formation sensitivity in a timely manner due to the limitations of flexibility and computational efficiency" in the revised manuscript.

328

329 **Q6:** Line 79-90: This section lists a few previous studies with vaguely portrayed methodologies and outcomes, but failed to mention any disadvantages of machine 330 learning, such as the demand for large volume of comprehensive and good quality data, 331 and of course the risk of overtraining. This section also fails to address a concern 332 brought up earlier by the authors themselves: uncertainties and biases in the input 333 dataset (observations, or outputs from chemical transport models). Please revise this 334 section and discuss the applications of machine learning in air quality studies in the 335 context of its disadvantages. Please also address how the impacts of input data (e.g., 336 337 uncertainties and biases) might be reduced.

Reply: Thank you. We have revised this section and added the discussion in lines 79-338 86 in the reversed manuscript "Although attentions should be paid to the robustness of 339 machine learning because it depends on input dataset (observations or outputs from 340 341 chemical transport models), previous studies have demonstrated that cross-validation and data-normalization can well reduce the dependence of the model on input data and 342 improve the robustness of the model (Wang et al., 2016; Wang et al., 2017b; Liu et al., 343 2021; Ma et al., 2021). Thus, it is a promising alternative to account for the effects of 344 345 meteorology on air pollutants and has been intensively used in atmospheric study (Liu et al., 2020; Hou et al., 2022).". 346

347

Q7: Line 135-137: "... and then averages the scores of each decision tree as its final score..." This is a very vague description of the algorithm. Indeed, the ensemble prediction made by the entire forest is usually more accurate and robust than the predictions made by individual trees, relatively speaking. However, this (averaging

across all the decision trees) ABSOLUTELY DOES NOT guarantee that large biases 352 and overfitting can be avoided. The splitting might help with mitigating the risk of 353 overfitting but it is still FAR FROM BEING SUFFICIENT to guarantee the algorithm 354 is not overfitted. Generally, much more comprehensive and rigorous measures than 355 what is shown in this work are needed, for instance, multifold cross validation is a good 356 idea. To further test the robustness of the machine learning workflow in real-world 357 physics-driven problems, sometimes it is recommended to perform the cross validation 358 359 with each fold being the data from a specific time period or geographic region.

Reply: Thank you so much for your good suggestion. We agree with you that the 360 splitting might help with mitigating the risk of overfitting but it is still not sufficient to 361 guarantee the algorithm is not overfitted. As replied in above questions, we performed 362 multifold cross-validation according to your suggestion. A 12-fold cross-validation was 363 carried out to mitigate the risk of overfitting. Briefly, we randomly divided the 364 normalized data into 12 subsets, then alternately took one subset as testing data along 365 with the rest as training data. By doing this, every data point has an equal chance being 366 367 trained and tested.

In lines 145-153 in the revised manuscript, we have revised the sentence "... and 368 then averages the scores of each decision tree as its final score..." to "The random forest 369 (RF) is a type of ensemble decision tree that can be used for classification and 370 regression (Breiman 2001). During the training process, the model creates a large 371 number of different decision trees with different sample sets at each node, and then 372 373 averages the results of all decision trees as its final results (Breiman 2001). To avoid 374 over-fitting, we trained the random forest model using cross-validation for the 375 normalized data, which can improve the robustness of the model. Briefly, we randomly 376 divided the normalized data into 12 subsets, then alternately took one subset as testing data along with the rest as training data. By doing this, every data point has an equal 377 chance being trained and tested.". 378

379

Q8: Line 132-: this section does not provide any details on whether the authors have
performed any sort of hyperparameter tuning, which is important. How can the readers

be convinced that the performance is optimized? Information outlined in Table S1 is not at all sufficient to described how the algorithm is configured. And frankly, certain information in that table is practically useless (e.g., the method is "regression", and the sampling is "random"). Please also clarify if the authors implemented random forest by themselves or used that from a certain package (e.g., R, python). If latter, please specify which package is used.

Reply: Thank you for you good comments. In the revised SI, we added the process of hyperparameter tuning, such as the influence of the min sample split (Figure R9 or Figure S13) and tree number (Figure R10 or Figure S14) on the mean squared error (MSE). We also revised Table S2 (Table R3). We implemented random forest simulation by ourselves. The code was written by ourselves based on a MATLAB platform.



394

Figure R9. The relationship between min sample split and mean squared error (MSE)



397 Figure R10. The relationship between trees number and mean squared error (MSE).

398

396

399

Table R3. RF model parameters and input parameters

RF model para	neters		RF model inputs					
Туре	Value	Туре	input parameter	output parameter				
Cross-validation	12	Figure 2A	Total VOC concentration, PM _{2.5} , NO, NO ₂ , CO, SR, RH, WD, WS, T	O ₃				
tree number	500	Figure 2B	51 measured VOC species, PM _{2.5} , NO, NO ₂ , CO, SR, RH, WD, WS, T	O3				
min sample split	5							
min sample leaf	1	Figure 2C	51 initial VOC species, PM _{2.5} , NO, NO ₂ , CO, SR, RH, WD, WS, T	O3				

400 Note: In this study, we optimized the number of tree and min samples split as shown

401 in Figure S14 and S13, respectively. The min sample leaf was set 1 (default), and

402 other parameter were set to default (auto/none).

403 Min samples split: the minimum number of samples to be split; Min samples leaf: the
404 minimum number of samples in a leaf; Trees number: the number of trees during the
405 training.

406

In lines 166-168 in the revised manuscript, we added a sentence "More details about workflow of RF model and the hyperparameter tuning can be found in the Text S3. The optimized parameters are shown in Table S2".

In the revised SI, we added a short paragraph to describe the process of hyperparameter tuning "(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 S12 and S13 show the examples to optimize the number of min sample split and tree, respectively.".

416

417 **Q9:** Line 143, 145: please clarify how this "tiny data noise" is added.

418 **Reply**: Thank you. The "tiny data noise" was through randomly permutes the 419 observations of feature *i*. We have revised it to "errOOB1 and errOOB2 represent the 420 out-of-bag data error of feature *i* before and after randomly permuting the observation, 421 respectively." in line 163-165 in the revised manuscript.

422

423 Q10: Line 147-150: I am not sure what the purpose is. Given the atmospheric lifetime 424 of ozone, the measurements may well carry some "memory effects". Do the authors 425 suggest that the ozone records can be reconstructed with randomly arranged inputs and 426 hence any transport footprint (and factors alike) is already captured by the suite of 427 instantaneous measurements?

Reply: Thank you. we agree with you that the measurements (especially O₃) may carry some "memory effects". However, it is inevitable that some "memory effects" cannot be considered. In order to decrease the dependence of the model on the data, we fed the randomly arranged inputs to the model by interrupting the continuity of the time series to check the robustness of the model. It is actually a cross-validation but with a smaller fold (not systematic) than that used in the revised manuscript, in which we carried out a 12-fold cross-validation to split the training and testing data according to your suggestion. Therefore, the results are more robust in the revised manuscript. We deleted the sentence "we interrupted the continuity of the time series, fed the randomly arranged inputs to the model" in the revised manuscript, and added the detailed workflow in the Text S3.

439

440 Q11: Line 160: It is unclear how the initial VOC concentrations are derived. I do notice that Text S2 in the supplement information is about initial VOCs. Please refer to SI 441 contents whenever necessary. I also have several major concerns regarding the initial 442 VOC calculations in Text S2: (1) these formulations require initial/source ratios. It 443 remains absolutely unclear how the authors derive the initial or source ratio of 444 ethylbenzene/xylene. (2) The initial/source ratios may vary among different sources 445 (e.g., gasoline, diesel, combustion, ...). How would the authors account for the impacts 446 from different emission sources? Or the mix with air masses with different 447 448 photochemical ages? (3) Ethylbenzene and xylene are primarily of anthropogenic origin. The underlying assumption here is that all VOCs experience similar transport and aging 449 processes. This may not be a bad assumption for other anthropogenic VOCs, but I doubt 450 this could be applied for biogenic VOCs, which, according to the authors (Line 261 and 451 Figure 3), is quite important. To sum up, I am not convinced that the initial VOC 452 calculation presented in this manuscript can accurately describe the VOC oxidation 453 454 during the transport receptor site.

Reply: Thanks for your comments. In the revised manuscript, we referred to SI contents
accordingly, such as, "The calculation of initial VOCs and sensitivity tests can be found
in the Supplemental Materials (S2)" in lines 142-143 and 190-191.

We took the mean value (0.52) of ethylbenzene/xylene from 0.45 to 0.66 before sunrise during our observations as the initial ratio of ethylbenzene/xylene. The protocol is the same as that reported in literatures (Shao et al., 2011; Zhan et al., 2021).

We agree with you that the initial/source ratios may vary among different sources(e.g., gasoline, diesel, combustion). Here, we used the ambient concentrations of

ethylbenzene and xylene to calculate the initial source ratio. Thus, it actually reflects a mean or mixing value of different sources instead of a specific one. There is a basic assumption that emission patterns of different sources are relatively stable in a day when calculating the photochemical initial concentrations of VOCs using this method (Yuan et al., 2013). As replied in the third question (Q3), we checked this assumption based on PSCF analysis. The results suggest that the emission patterns of xylene and ethylbenzene are highly similar in the early morning to that in the whole day.

470 Although xylene and ethylbenzene are anthropogenic VOCs, while isoprene is a biogenic VOC, PSCF analysis indicated that the spatial pattern of isoprene was similar 471 to that of xylene and ethylbenzene during our observations (Figure R8). This means 472 VOC emissions were evenly distributed in Beijing. This can be ascribed to the fact that 473 our observation site is a typical urban station and VOCs are emitted from non-point 474 sources on a city scale. Therefore, the photochemical clock calculated using xylene and 475 ethylbenzene is able to correct the photochemical loss of biogenic VOCs to some extent. 476 In Text S2 in the revised SI, we added more details to calculate the initial 477 478 concentrations of VOCs.

479 "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).

484
$$\int c_{\text{OH}} dt = \frac{1}{k_{\text{A,OH}} - k_{\text{B,OH}}} \left[\ln\left(\frac{\text{VOC}_{\text{A}}}{\text{VOC}_{\text{B}}}\right)_{\text{initial}} - \ln\left(\frac{\text{VOC}_{\text{A}}}{\text{VOC}_{\text{B}}}\right) \right]$$
(S1)

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

486
$$VOC_{i, initial} = \frac{VOC_{i}}{exp(-k_{i,OH})exp(\int c_{OH}dt)}$$
(S2)

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

488
$$\operatorname{VOC}_{i, \text{ initial}} = \frac{\operatorname{VOC}_{i}}{\exp(-k_{i, \text{ OH}}) \exp(\frac{1}{k_{A, \text{OH}} - k_{B, \text{ OH}}} \left[\ln(\frac{\operatorname{VOC}_{A}}{\operatorname{VOC}_{B}})_{\text{ initial}} - \ln(\frac{\operatorname{VOC}_{A}}{\operatorname{VOC}_{B}})\right])}$$
(S3)

Where C_{OH} represents the ambient OH concentration; $k_{A,OH}$ and $k_{B,OH}$ represent the reaction rate of compound A and B with OH radical, respectively; t represents the 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).

499 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 500 well correlated, which indicates that they are simultaneously emitted. In addition, we 501 compared the PICs according to xylene/ethylbenzene with that using toluene/benzene 502 (Figure S10). The calculated PICs ratio (PIC Xylene/Ethylbenzene / PIC Toluene/Benzene) varied 503 from 0.5 to 1.5 with a mean value of 0.96. This means the calculated initial VOCs was 504 in good agreement when using different tracers. The mean ratio (0.52, from 0.45 to 0.66) 505 of ethylbenzene/xylene before sunrise was taken as the initial ratio of 506 507 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 S4). 508

509 Variations of air mass may also affect the VOC ratio. Figure S11 A-D shows the mean concentration distribution of ethylbenzene and xylene in the early morning and 510 511 the whole day based on potential source contribution function (PSCF) analysis. Xylene showed similar patterns to ethylbenzene in different air mass trajectories and different 512 513 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 514 515 the initial ratio of VOCs. The hourly concentrations of ethylbenzene and xylene were used to calculate the concentration of initial VOCs. The initial VOC was calculated by 516 adding the measured VOC concentration and the calculated photochemical loss. Figure 517 S12 shows the diurnal variations of the observed and initial VOCs concentrations from 518 2014 to 2016. Photochemical loss of VOC occurred mainly during the daytime. 519

520 It should be noted that the lifetimes $(1/k_2c_{OH})$ of highly reactive VOCs, such as 521 isoprene, greatly depend on the OH exposure. The photochemical ages of isoprene were

0.01–6.21 h (1.26 \pm 1.12 h). This value is comparable with previously reported 522 photochemical ages (Shao et al., 2011; Gao et al., 2018). However, the initial 523 concentrations of highly reactive VOCs may be overestimated due to their short 524 lifetimes and should be taken as the upper limits. On the other hand, isoprene is a 525 biogenic VOC, while xylene and ethylbenzene are anthropogenic VOCs. If they do not 526 share the same air mass histories, an additional uncertainty is inevitable for the PICs of 527 isoprene. As shown in Figure S11, isoprene showed similar patterns to that of xylene 528 529 and ethylbenzene, which means VOC emissions are evenly distributed in Beijing during our observations. This can be ascribed to the fact that our observation site is a typical 530 urban station. Although isoprene and xylene/ethylbenzene different sources, both them 531 are non-point sources on a city scale. Therefore, the photochemical clock calculated 532 using xylene and ethylbenzene is able to correct the photochemical loss of biogenic 533 VOCs to some extent. It should be noted that uncertainty is inevitable when we 534 estimating the photochemical age (Parrish et al., 2007). However, the aim of this work 535 is to test whether the ML-model can reflect the influence of photochemical loss of 536 537 VOCs species on O₃ modelling. The PICs should provide additional information for understanding O₃ formation in the atmosphere.". 538

539

540 Q12: Line 176-177: Simply comparing RF outputs to measurements ABSOLUTELY 541 DOES NOT guarantee that the RF model is robust. An over-trained model will also 542 show good performance when evaluated with measurements. Figure 2 says nothing 543 about the robustness of the model.

Reply: Thank you so much. We agree with you that simply comparing RF outputs to measurements absolutely does not guarantee that the RF model is robust. We have carried out cross-validation in the revised manuscript according to your good suggestion. The details on the robustness of the model have been replied in Q1 and Q2. All the figures (Figure 2, Figure S3 and Figure S4) have been updated in the revised manuscript and the revised SI.

550

551 Q13: Line 221-222: The authors stated that "ML is a black-box model" then cited

Sayeed et al. (2021). Do the authors then imply that the convolutional neural network 552 model developed by and described in Sayeed et al is a "black-box"? If this is the case, 553 please elaborate how the Sayeed et al model appears to be a black-box to the authors, 554 since this is a somewhat strong accusation. I also do not fully agree with this statement. 555 Random forest is actually fairly transparent compared to some other types of machine 556 learning algorithms, as one can certainly examine the trees and see how a certain feature 557 is used for splitting the nodes and how the overall importance is propagated, if they 558 559 wish.

Reply: Thank you. We agree with you that when compared with other types of machine 560 learning algorithms, random forest is actually fairly transparent because one can 561 examine every node. However, when compared with chemical models, such as box 562 model and chemical transport model. It is still a "black box" from the point view of 563 564 chemical mechanism. In lines 116-118 in the revised manuscript, we revised this sentence to "Although ML is widely used to understand air pollution, explanations of 565 ML results (e.g., RI) are somewhat vague because ML is a black-box model from the 566 567 point view of chemical mechanism".

568

Q14: Line 267-270, Line 287-293: Please clarify how the EKMA plots were generated 569 using the RF model. It is confusing that Figures 4A-C and Figure 4D use different color 570 scales. It also appears to me that Figure 4B and Figure 4D show considerable 571 discrepancy: the "observed point" in Figure 4B indicates ~44 ppb ozone but in Figure 572 4D the ozone level at the "observed point" is ~60 ppb. Please also define this "observed 573 point". It could be that the OBM predicts some sort of "maximum ozone production 574 575 potential" driven by chemistry and impacts like transport would not be captured. But this would be inconsistent with one of the conclusions of this work (e.g., Line 206) that 576 ozone production seems to be dominated by local chemistry. Either way, the 577 discrepancy needs to be elaborated. 578

Reply: Thank you for your good comment. In the revised SI, we have added the calculation process of EKMA in Text S3. This has also been replied in **Q2**. In lines 171-179 in the revised manuscript, we added a paragraph "When plotting the O₃ formation

sensitivity curves, we made a virtual matrix of inputs by varying the concentrations of 582 NOx and VOCs from 0.9 to 1.1 times (with a step of 0.01) of their mean values while 583 584 keeping all other inputs unchanged (i.e., the mean values). Then, the new matrix was used as testing data, while all the measured data were taken as training data. Thus, the 585 testing data should represent the mean sensitivity regime of O₃ in Beijing, while the 586 training data actually covered all the sensitivity regimes of O₃ formation. The EKMA 587 curves were plotted using the daily maximum 8-h (MDA8) O₃. More details can be 588 589 found in the SI".

In the initial submission version of the manuscript, the EKMA curves based on the 590 RF model were plotted using the mean O₃ concentrations in a day with different color 591 scale, while the EKMA curves from the Box model were plotted using the maximum 592 O₃ concentrations in a day. This leads to the difference in O₃ concentrations between 593 594 Figure 4B and 4D. In the revised manuscript, we plotted the EKMA curves from both models using the daily maximum 8-h (MDA8) O₃ with the same scale. As shown in 595 Figure R3 or Figure 4. The predicted O3 concentrations using two different models are 596 597 comparable. In addition, we calculated the relative errors (within 15.6%) of O₃ concentrations from the RF model compared with that of the Box model (Figure R4 or 598 Figure S8). The observed point has been added in the revised Figure 4 or Figure R3. 599

600 In Text S3 in the revised SI, we have added the details for plotting EKMA curves: "(6) EKMA curves. The Empirical Kinetic Modeling Approach (EKMA) was used to assess 601 the O₃ formation mechanism regime. Both the RF model and a box model with Master 602 603 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 604 605 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% in a 606 two-dimensional matrix along with other inputs unchanged. This matrix was used as 607 the testing data, while all the measured data were taken as the training data in the RF 608 model to simulate O3 concentrations under different scenarios of VOCs and NOx 609 610 concentrations. To decrease the model uncertainty, we set relatively small variations of VOCs and NOx $(\pm 10\%)$ compared to the observed values in this study. The mean 611

612 relative error of simulated O₃ concentrations between RF model and Box model (within

613 15.6%, Figure S8) suggests that the RF model can well predict O₃ concentrations during
614 our observations.".

615

616 **Q15:** Line 299-301: I actually fail to see this. Figures 2E and 2F show that the 617 improvement is small.

Reply: Thank you. In the revised manuscript, we have revised the sentence "Thus, our 618 model has good prediction performance ($R^2 = 0.87$) when combined with the initial 619 VOC species...and identify the connection between the reactivity of VOC species and 620 O₃ formation in the atmosphere" to "When the TVOCs were split into measured or 621 initial VOC species, the R^2 increased obviously as the number of data features increased. 622 Therefore, the VOC composition has a significant influence on O₃ prediction using the 623 RF model." in Lines 221-223 in the revised manuscript. Meanwhile, we added "To 624 verify the stability of the model, we performed a significance test on the model results. 625 The results showed that there was no significant difference among the different tests 626 $(P>0.05, R^2>0.98)$," in Lines 168-170 in the revised manuscript. 627

628

629 Minor/technical comments:

630 **Q16:** Line 16: Beijing, China.

631 **Reply**: Thank you. It has been corrected in Line 14 in the revised manuscript.

632

Q17: Line 21: Define abbreviations upon first appearance. NOx and PM_{2.5} are not
defined. Also, x should be subscripted throughout.

635 Reply: Thank you. We have defined the abbreviations when it first appears, such as

- 636 "nitrogen oxides (NO_x) and relative humidity (RH), and a positive response to
- 637 temperature (T), solar radiation (SR)" and "fine particulate matter with aerodynamic
- diameter less than 2.5 μm (PM_{2.5})" in Lines 22 and 95-96. The "x" in NO_x has been

639 subscripted throughout the paper.

- 640
- 641 **Q18:** Line 22: Delete curves

642 **Reply**: Thank you. We have deleted the "curves"

643

644 **Q19:** Line 33: "NO-O₃-NO₂ cycle" please elaborate. The audience will appreciate 645 clearly described concepts.

646 **Reply**: Thank you. The "NO-O₃-NO₂ cycle" can be expressed by eq. R (1) and R (2).

647

$$HO_2/RO_2 + NO \rightarrow OH/RO + NO_2$$
 R (1)

648

$$NO_2 + O_2 \xrightarrow{hv} NO + O_3$$
 R (2)

649 We have revised "NO-O₃-NO₂ cycle" to "conversion from NO to NO₂, subsequently,

formation of O₃ by photolysis of NO₂ in the presence of O₂" in Line 36 in the revised
manuscript.

652

653 **Q20:** Line 31-34: Very busy and ill-formed sentence. Please rephrase.

Reply: Thank you. We have revised this sentence to "Oxidation of volatile organic compounds (VOCs) will produce peroxyl radicals (RO₂) and hydroperoxyl radicals (HO₂). The RO₂/HO₂ can accelerate conversion from NO to NO₂, subsequently, formation of O₃ by photolysis of NO₂ in the presence of O₂ (Wang et al., 2017a) " in lines 34-37 in the revised manuscript.

659

660 **Q21:** Line 52-53: "In addition, this method lacks the predictability of O₃ concentrations

for policy-making." This is a bizarre statement. By definition, observations do not havepredictability.

663 **Reply**: Thank you. We have deleted this statement in the revised manuscript.

664 **Q22:** Line 53: OBM can use remote sensing measurements too.

Reply: Thank you. We have revised to "OBMs combine *in-situ* field observations,
remote sensing measurements and chemical box models, which…" in Line 56 in the
revised manuscript.

668

669 **Q23:** Line 80: has -> have.

670 Reply: Thank you. We have revised "has" to "have" in Line 88 in the revised

671 manuscript.

672

673 **Q24:** Line 99: please fix citation style as per the journal requirements.

674 **Reply**: Thank you. We have fixed citation style as per the journal requirements in the675 revised manuscript.

676

Q25: Line 98-100: "Makar et al (1999) reported that highly reactive species, such as isoprene, were underestimated by 40% when the OH reactions were ignored" This is confusing as currently written, which is also a misinterpretation of the paper. Makar et al. clearly stated that it is the isoprene emissions that are underestimated if the OH oxidation is not considered.

Reply: Thank you. We have revised it to "the isoprene emissions were underestimated
by up to 40% if the OH oxidation is not considered." in Lines 109-110 in the revised
manuscript.

685

686 Q26: Line 101: Please clearly define what exactly "initial concentration" is. It is my 687 opinion that this concept is vague: if one attempts to measure this "initial concentration", 688 how close shall the sensor be placed? When it comes to biogenic compounds like 689 isoprene, shall the sensor be placed at the leaf level, within the canopy, or what?

Reply: Thank you. The "initial concentration" is a concept based on reaction time
instead of location or space. It is defined as the concentrations of VOCs before
degradation has occurred in a day and actually extrapolated according to a "chemical
clock" (OH exposure).

In lines 118-121 in the revised manuscript, we clarified it "In this study, we used the RF model to evaluate the prediction performance of atmospheric O₃ using the TVOCs, measured VOC species and photochemical initial concentration (PIC) of VOC species, which is calculated based on the photochemical-age approach (Shao et al., 2011)" and in the revised SI, we added a sentence "Photochemical initial concentration (PIC) proposed by Shao et al. (2011), which is calculated based on the photochemicalage approach and has been applied to evaluate the effect of photochemical processing

- 701
- on measured VOC levels".
- 702

Q27: Line 133: Random forest is not a type of decision tree. It is a collection of a
number of decision trees.

Reply: Thank you. We have revised it to "The random forest (RF) is a type of ensemble
decision tree that can be used for classification and regression" in Line 145 in the
revised manuscript.

708

Q28: Line 132-: this section does not mention anything about how these input variables are pre-processed: are extreme values removed? Do the authors apply any standardization or normalization? These variables are on very different numerical scales and sometimes standardization could improve the stability and/or performance.

Reply: Thank you. We carried out data-normalization and removed the extreme values 713 before modelling. In the revised SI, we clarified it and added the data processing in Text 714 S3 "(2) Data process. After the extreme values were removed, all data were normalized 715 716 (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 717 718 was randomly divided into 12 subsets. Thus, a 12-fold cross-validation was performed 719 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.". 720

721

722 **Q29:** Line 136: Please define "score".

Reply: Thank you. To describe the random forest more accurately, we have revised
"score" to "result" in line 146-148 in the revised manuscript.

"725 "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 results of all

- 727 decision trees as its final results".
- 728

729 **Q30:** Figure 1: what are those red arrows in the top panel?

730 **Reply**: Thank you. The red arrows represent the O₃ concentration exceeding 74.6 ppbv

according to the national ambient air quality standard. We have added this explanation
in Figure 1 or Figure R11 in Lines 205-206 in the revised manuscript.



Figure R11. Time series of air pollutants and meteorological parameters during
observations in Beijing. (In A, the red arrows represent the O₃ concentration exceed
736 74.6 ppbv according to the national ambient air quality standard.)

737

733

738 Q31: Figure 1: it would be interesting to separate biogenic VOCs from anthropogenic739 VOCs.

Reply: Thank you. In this figure, we want to show the general characteristics of
different kinds of VOCs. It is usually classified according to the structures of VOCs.

- Figure R12 shows the time series of biogenic and anthropogenic VOCs. We also added
- this figure in the revised SI (Figure S16).





Figure R12. Time series of biogenic and anthropogenic VOCs during the observation
period. (Biogenic VOCs: including isoprene, α-pinene and β-pinene; Anthropogenic
VOCs: including all detected VOCs except isoprene, α-pinene and β-pinene)

748

749 Q32: Figure 2: all top panels are not readable. Please consider extending the time series750 plots.

Reply: Thank you. We have revised Figure 2 (or Figure R2) in the revised manuscript.

Q33: Text S1: "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". Well this is a very wide range for limit of quantification. Please include a table and list the limits of quantification/detection for all VOCs used in this work. I'd argue that if the LOQ/LOD of an ambient VOC is on the order of 100 ppb, the data is practically useless. The total VOC levels are less than ~30 ppb (Figure 1F).
Reply: Thank you. In the submission version, we made a mistake about the number. It

- was 0.002-0.05 ppbv instead of 0.1-100 ppbv. We have added the method detection
- 161 limit (MDL) for all measured species in Table R2 or Table S4. In Text S1, we have

revised "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" to "total of 51 VOCs
- (including 21 alkanes, 13 alkenes, 1 alkyne and 16 aromatics) were analyzed within a
- limit of quantification of 0.002-0.05 ppbv as shown in Table S4".
- 766
- Q34: Text S2: please define all variables. CoH, t, k, are not defined. Also, Equation S1
 essentially describes the integrated OH exposure, rather than the "changes in VOC
 concentration as a function of time due to photochemical reaction".

770 Reply: Thank you. We have defined the all variables and revised the description

- "Equation S1 essentially describes the integrated OH exposure (Shao et al., 2011)" in
- Text S2 in the revised SI.
- "Where Coh represents the ambient OH concentration; $k_{A,OH}$ and $k_{B,OH}$ represent the

reaction rate of compound A and B with OH radical, respectively; t represents the

- reaction time of species *i* in the ambient."
- 776
- 777 **Q35:** Table S1: please define all parameters. What's leaf number? Is this the number of
- 1778 leaf nodes, leaf levels/depths, or what? What's fboot?
- 779 **Reply**: Thank you, we have revised the Table S2 (Table R3). The leaf number represents
- the minimum number of samples to be split. The fboot represents the minimum number
- of samples in a leaf. They are added in the footnotes of Table S2.
- 782

783 **Reference**

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

Carter, W.P.L. Development of the SAPRC-07 chemical mechanism. Atmospheric Environment, 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. Science of The Total Environment,
 628-629, 791-804, https://doi.org/10.1016/j.scitotenv.2018.01.175, 2018.
- Hou, L., Dai, Q., Song, C., Liu, B., Guo, F., Dai, T., Li, L., Liu, B., Bi, X., Zhang, Y., Feng, Y. Revealing
 Drivers of Haze Pollution by Explainable Machine Learning. Environmental Science &
 Technology Letters, 10.1021/acs.estlett.1c00865, 2022.
- Liu, H., Liu, J., Liu, Y., Ouyang, B., Xiang, S., Yi, K., Tao, S. Analysis of wintertime O3 variability using

796 a random forest model and high-frequency observations in Zhangjiakou-an area with 797 background pollution level of the North China Plain. Environmental Pollution, 262, 114191, 798 https://doi.org/10.1016/j.envpol.2020.114191, 2020. 799 Liu, Z., Qi, Z., Ni, X., Dong, M., Ma, M., Xue, W., Zhang, Q., Wang, J. How to apply O3 and PM2.5 800 collaborative control to practical management in China: A study based on meta-analysis and 801 Science of The Total machine learning. Environment, 772, 145392, 802 https://doi.org/10.1016/j.scitotenv.2021.145392, 2021. 803 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 O3 trends in the Beijing-Tianjin-Hebei region in China, 2010 to 804 805 2017. Environmental Pollution, 276, 116635, https://doi.org/10.1016/j.envpol.2021.116635, 806 2021. 807 Parrish, D.D., Stohl, A., Forster, C., Atlas, E.L., Blake, D.R., Goldan, P.D., Kuster, W.C., de Gouw, J.A. 808 Effects of mixing on evolution of hydrocarbon ratios in the troposphere. Journal of Geophysical 809 Research: Atmospheres, 112, https://doi.org/10.1029/2006JD007583, 2007. 810 Shao, M., Bin, W., Sihua, L., Bin, Y., Ming, W. Effects of Beijing Olympics Control Measures on Reducing Reactive Hydrocarbon Species. Environmental Science & Technology, 45, 514-519, 811 812 10.1021/es102357t, 2011. 813 Wang, T., Xue, L., Brimblecombe, P., Lam, Y.F., Li, L., Zhang, L. Ozone pollution in China: A review of 814 concentrations, meteorological influences, chemical precursors, and effects. Science of The 815 Total Environment, 575, 1582-1596, https://doi.org/10.1016/j.scitotenv.2016.10.081, 2017a. Wang, Y., Li, Y., Pu, W., Wen, K., Shugart, Y.Y., Xiong, M., Jin, L. Random Bits Forest: a Strong 816 817 Classifier/Regressor for Big Data. Scientific Reports, 6, 30086, 10.1038/srep30086, 2016. 818 Wang, Y., Wu, G., Deng, L., Tang, Z., Wang, K., Sun, W., Shangguan, Z. Prediction of aboveground 819 grassland biomass on the Loess Plateau, China, using a random forest algorithm. Scientific 820 Reports, 7, 6940, 10.1038/s41598-017-07197-6, 2017b. 821 Yuan, B., Hu, W.W., Shao, M., Wang, M., Chen, W.T., Lu, S.H., Zeng, L.M., Hu, M. VOC emissions, 822 evolutions and contributions to SOA formation at a receptor site in eastern China. Atmos. Chem. 823 Phys., 13, 8815-8832, 10.5194/acp-13-8815-2013, 2013. 824

825 **Reviewer 2**#

This article estimates the sensitivity of ozone formation using a random forest model 826 with not only total VOCs concentrations but also observed concentrations and initial 827 concentrations of VOC species. The result showed that the ozone prediction 828 performance using initial concentrations of VOC species was better than that using total 829 VOCs concentrations. Analytical reports using machine learning with total VOCs 830 concentrations have been published recently. This article simply indicates the 831 832 superiority of using overserved or initial concentrations of VOC species. From this aspect, the significance of this study is evident. The reviewer would recommend it for 833 834 publication.

Reply: Thank you for your positive comments. We have carefully responded to all of
your point-by-point comments and issues and have revised the manuscript accordingly.
These revisions are described in detail below.

838

Q1: However, the reliability of analytical data and explanation of the initial
concentration of VOC species are not sufficiently indicated in this article. And some
expressions seem to be somewhat inadequate for well understanding. Slight revisions
are required.

Reply: Thank you for your good suggestions. The reliability of analytical data will be 843 replied in the following question. We have added more details about the initial 844 concentration of VOC species. 1) We added the selection rules of tracers to calculate 845 846 OH exposure and confirmed these rules based on our observation data. As shown in Figure R5 (Figure S9), the concentrations of xylene and ethylbenzene are well 847 848 correlated, which indicates that they are simultaneously emitted. In addition, we compared the photochemical initial concentrations (PICs) calculated using 849 xylene/ethylbenzene with that using toluene/benzene (Figure R6 or Figure S10). 2) We 850 performed sensitivity tests about OH exposure calculation. The results showed that the 851 uncertainty caused by the OH exposure (from -10% to +10%) ranged from 0.55 to 1.57 852 (Table R2 or Table S4). 3) we compared the chemical ages in this work with those 853 reported in literatures. For example, the photochemical ages of isoprene were 0.01-6.21 854

h $(1.26 \pm 1.12 \text{ h})$. This value is comparable with previously reported photochemical ages (Shao et al., 2011; Gao et al., 2018). 4) The diurnal curves of measured and initial

857 VOC concentrations were added in the revised SI (Figure R7 or Figure S12).

In Text S2, we have added more details about initial VOC concentration calculations and data reliability.

860 "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).

865
$$\int c_{\rm OH} dt = \frac{1}{k_{\rm A,OH} - k_{\rm B,OH}} \left[\ln\left(\frac{\rm VOC_A}{\rm VOC_B}\right)_{\rm initial} - \ln\left(\frac{\rm VOC_A}{\rm VOC_B}\right) \right]$$
(S1)

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

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

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

869
$$\operatorname{VOC}_{i, \text{ initial}} = \frac{\operatorname{VOC}_{i}}{\exp(-k_{i, \text{ OH}}) \exp(\frac{1}{k_{A, \text{OH}} - k_{B, \text{ OH}}} \left[\ln(\frac{\operatorname{VOC}_{A}}{\operatorname{VOC}_{B}})_{\text{initial}} - \ln(\frac{\operatorname{VOC}_{A}}{\operatorname{VOC}_{B}})\right])}$$
(S3)

Where C_{OH} represents the ambient OH concentration; $k_{A,OH}$ and $k_{B,OH}$ represent the reaction rate of compound A and B with OH radical, respectively; t represents the 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 884 from 0.5 to 1.5 with a mean value of 0.96. This means the calculated initial VOCs was 885 in good agreement when using different tracers. The mean ratio (0.52, from 0.45 to 0.66) 886 of ethylbenzene/xylene before sunrise was taken as the initial ratio of 887 ethylbenzene/xylene. Sensitivity tests showed that the uncertainty of PICs caused by 888 the OH exposure (from -10% to +10%) ranged from 0.55 to 1.57 (Table S4). Variations 889 of air mass may also affect the VOC ratio. Figure S11 A-D shows the mean 890 891 concentration distribution of ethylbenzene and xylene in the early morning and the whole day based on potential source contribution function (PSCF) analysis. Xylene 892 showed similar patterns to ethylbenzene in different air mass trajectories and different 893 periods. These results indicate that the emissions of xylene and ethylbenzene were 894 constant throughout the day and variations of air mass should have little influence on 895 the initial ratio of VOCs. The hourly concentrations of ethylbenzene and xylene were 896 used to calculate the concentration of initial VOCs. The initial VOC was calculated by 897 adding the measured VOC concentration and the calculated photochemical loss. Figure 898 899 S12 shows the diurnal variations of the observed and initial VOCs concentrations from 2014 to 2016. Photochemical loss of VOC occurred mainly during the daytime. 900

901 It should be noted that the lifetimes $(1/k_{2COH})$ of highly reactive VOCs, such as 902 isoprene, greatly depend on the OH exposure. The photochemical ages of isoprene were 0.01–6.21 h (1.26 \pm 1.12 h). This value is comparable with previously reported 903 photochemical ages (Shao et al., 2011; Gao et al., 2018). However, the initial 904 concentrations of highly reactive VOCs may be overestimated due to their short 905 lifetimes and should be taken as the upper limits. On the other hand, isoprene is a 906 907 biogenic VOC, while xylene and ethylbenzene are anthropogenic VOCs. If they do not share the same air mass histories, an additional uncertainty is inevitable for the PICs of 908 909 isoprene. As shown in Figure S11, isoprene showed similar patterns to that of xylene and ethylbenzene, which means VOC emissions are evenly distributed in Beijing during 910 our observations. This can be ascribed to the fact that our observation site is a typical 911 912 urban station. Although isoprene and xylene/ethylbenzene different sources, both them are non-point sources on a city scale. Therefore, the photochemical clock calculated 913

using xylene and ethylbenzene is able to correct the photochemical loss of biogenic
VOCs to some extent. It should be noted that uncertainty is inevitable when we
estimating the photochemical age (Parrish et al., 2007). However, the aim of this work
is to test whether the ML-model can reflect the influence of photochemical loss of
VOCs species on O₃ modelling. The PICs should provide additional information for
understanding O₃ formation in the atmosphere.".

920

Q2: Line 50-75: It is easier to read if you divide the paragraph appropriately, e.g.,
Line 50, Line 53, Line 59, Line 64, Line 75.

923 **Reply**: Thank you. To make it is readable, we have started a new paragraph in line 54 924 before "The observed indicators can be utilized to quickly diagnose the sensitivity 925 regime…" and in line 76 before "Compared to traditional methods, machine learning 926 (ML) is able to capture the" in the revised manuscript.

927

928 Q3: Line 118-129 and S1: As the composition of VOC species varies greatly from year 929 to year as shown in Fig. 1(F), the analytical reliability is important. Information about 930 VOC measurement in this article is insufficient to understand the reliability. Further 931 information such as observation period and the reason, sampling time or cycle, 932 calibration using standard gas, and method of quality control should be described.

Reply: Thank you so much for your good suggestion. To clarify the analytical radiality,
we have added more information about VOC measurement, including sampling time,

data quality control, calibrations using standard gas. The details are shown in Text S1in the SI.

937 "Text S1. Field observations

938 VOCs were measured in May and June from 2014 to 2016 by an online 939 commercial instrument (GC-866, Chromatotec, France), which consisted of two 940 independent analyzers for C₂-C₆ and C₆-C₁₂ hydrocarbons. Both analyzers were 941 equipped with a preconcentration system, a chromatographic column, and a flame 942 ionization detector. The analyzers are located in an air-conditioned room and the sample 943 tubes are wrapped with a heating jackets and insulation to ensure that the temperature 944 remains stable between 22 and 27°C.

945 The samples were injected into the low carbon (C_2-C_6) analyzer and the high 946 carbon (C₆-C₁₂) analyzer, respectively. Isoprene was detected in the components of C₂-C₆, while α -pinene and β -pinene were detected along with other VOCs of C₆-C₁₂. In the 947 low carbon analyzer, the samples were adsorbed by a capture tube at -8 °C. The capture 948 949 tube was then rapidly heated to 220 °C. The samples were introduced into a chromatographic column (id=0.53mm, length=25m) with hydrogen as the carrier gas 950 951 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 952 was heated to 380 °C, introduced into a chromatographic column (id=0.28mm, 953 length=30m) with hydrogen as the carrier gas and finally detected by the same FID 954 detector. The material in the column was Al₂O₃/Na₂SO₄. 955

The signals of VOCs were converted into chromatograms for qualitative and 956 quantitative analysis. Before quantitative analysis, the retention time of each 957 component was carefully checked using the chromatographic analysis software. The 958 959 instruments were calibrated using both internal and external calibrations. Internally calibration was carried out twice every 24 hours using n-butane, n-hexane and benzene 960 at different flow rates. External calibration was performed monthly using standard gas 961 mixtures of volatile organic compounds (PAMS and TO-14, Linde gas, USA). The 962 963 concentrations of each species were calculated according to the corresponding working curves with six concentration levels. In this study, total of 51 VOCs (including 21 964 alkanes, 13 alkenes, 1 alkyne and 16 aromatics) were analyzed within a limit of 965 quantification of 0.002-0.05 ppbv as shown in Table S4. The relative standard 966 967 derivations (RSDs) were within 10% for each compound among seven replicates.".

968

Q4: Line 129 and S2: Explanation about PIC is insufficient in this article although PIC
is important for the results. To calculate PIC, the initial ratio of Ethylbenzene and xylene
must be constant. However, they may be emitted from several sources, e.g. painting,
mobile exhaust, etc. Please explain why you can use these compounds in this study.

973 And please describe the VOC sampling time which is also important to calculate PIC,

974 because chemical reactions in the air are different in daytime and nighttime.

975 **Reply**: Thank you very much for your suggestions. We added explanations about PIC calculations in Text S2. Meanwhile, the xylene/ethylbenzene was selected as tracer 976 for the following reasons: 1) the concentrations of xylene and ethylbenzene are well 977 correlated, which indicates that they are simultaneously emitted; 2) they have different 978 degradation rates in the atmosphere; 3) the calculated initial VOCs are in good 979 agreement with those calculated using other tracers, such as toluene/benzene. As shown 980 981 in Figure R5 or Figure S9, the concentrations of xylene and ethylbenzene correlated well during our observations in this work. In addition, we compared the ratio of the 982 initial concentrations calculated according to the ratio of xylene/ethylbenzene with that 983 using the ratio of toluene/benzene (Figure R6 or S9). Except for several compounds, 984 the ratio of the PICs for most of these VOCs varied within 1.0±0.1. This means the 985 calculated photochemical initial concentrations (PICs) are in good agreement when 986 using different tracers. Sensitivity tests showed that the uncertainty caused by the OH 987 exposure (from -10% to +10%) ranged from 0.55 to 1.57 (Table R2 or Table S4). Figure 988 989 R7 or Figure S12 shows the calculated diurnal curves of the PICs from 2014 to 2016. Photochemical losses of VOCs occurred prominently during the daytime. 990

991 Potential source contribution function (PSCF) analysis has been further carried out to evaluated the possible influence of air mass on the emission ratio of ethylbenzene 992 993 and xylene. As shown in Figure R8A-D or Figure S11, xylene showed similar pattern to ethylbenzene in the early morning or in the whole day. These results indicate that 994 variations of air mass should have little influence on their initial ratio. In addition, 995 isoprene showed similar patterns to that of xylene and ethylbenzene (Figure R8G-H), 996 997 which means VOC emissions are evenly distributed in Beijing. This can be ascribed to 998 the fact that our observation site is a typical urban station. Although isoprene and 999 xylene/ethylbenzene are from biogenic sources and anthropogenic sources, both them 1000 are non-point sources on a city scale.

Figure R7 or Figure S12 shows the diurnal variations of the observed and initial VOCs concentrations from 2014 to 2016. From Figure R7, it can be seen that photochemical loss of VOCs occurred prominently during the daytime. We have added information about the sampling time ("VOCs were measured in May and June from
2014 to 2016 by an online commercial instrument...") in Text S1.

- 1006
- 1007 **Q5:** Figure 1: What are the red lines in (A)?

1008 **Reply**: Thank you. The red lines (arrows) indicate the O₃ concentration exceed 74.6 1009 ppbv according to the national ambient air quality standard. We have added an 1010 explanation in the caption of Figure 1 "**Figure 1**. Time series of air pollutants and 1011 meteorological parameters during observations in Beijing (In A, the red arrows 1012 represent the O₃ concentration exceed 74.6 ppbv according to the national ambient air

- 1013 quality standard)".
- 1014

1015 Q6: Figure 2: It is difficult to find the difference in (A), (B), and (C). Something will1016 be needed to make it clear.

1017 **Reply**: Thank you. We revised these figures (bolded the lines and increased the color

1018 contrast) to make it clearer. Actually, Figure 2(D)-(F) can show their difference for the

1019 correlations between modeled and observed O₃ concentrations.

1020

1021 **Q7:** Line 203-205: It is unclear that this is what part about Figure 3A.

1022 **Reply**: Thank you. We have adjusted the position of the label in Figure 3 (Figure R13)

1023 and the left part is Figure 3A.



1024

Figure R13. Percentage of RI for O₃ precursors and meteorological parameters (A) and
the top 10 factors with high values of RI in 2014-2016 (B-D: using initial concentrations
of VOC species).

1028

1029 Q8: Figure 3A: Please explain why RI does not change so much even though the1030 composition of VOCs differs greatly between 2015 and 2016.

1031 **Reply**: Thank you for your good comments. VOCs contributed 64.0%, 58.9%, and 63.3% to the RI, respectively, in 2014, 2015, and 2016. From Figure R13 or Figure 3, it did 1032 1033 not change so much among the three years although VOCs concentrations changed obviously. We think this should be ascribed to the production of O₃ in Beijing was still 1034 in a VOCs-sensitive regime as shown in Figure R3. This is consistent with previous 1035 1036 studies based on transport chemical model (Li et al. 2020) which found VOCs were the dominant contributors to O₃ formation in Beijing. This means that the sensitivity of RI 1037 to VOCs concentrations might be not so high. We suppose that it should be more 1038 1039 prominent when O₃ formation shifts to a NOx-sensitive regime. However, as shown in 1040 Figure R13 (Figure 13B-D), the RI of individual VOC species changed obviously 1041 among different years.

1043 Q9: Line 287-289: It is unclear why you can describe that the RF model is better than1044 the box model from Figure S5.

Reply: Thank you. We have deleted Figure S5 and revised the statement to "We
compared the relative error of simulated MDA8 O₃ calculated using the RF and OBM
model in 2015, as shown in Figure S8. The mean relative error of simulated MDA8 O₃
between RF model and Box model was 15.6%. Hence, a combination of the RF model
and initial VOCs species can well depict the sensitivity regime of O₃ formation, while
the calculated RIs correlate well with the OFPs." in lines 319-323.

1052 **Reference**

1051

- 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.
- 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, 2020.
- 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,
 1068 10.1021/es102357t, 2011.

1069