

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

3     Junlei Zhan<sup>1</sup>, Yongchun Liu<sup>1\*</sup>, Wei Ma<sup>1</sup>, Xin Zhang<sup>2</sup>, Xuezhong Wang<sup>2</sup>, Fang Bi<sup>2</sup>,  
4                   Yujie Zhang<sup>2</sup>, Zhenhai Wu<sup>2</sup>, Hong Li<sup>2\*</sup>

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

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

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

## 10 **Abstract**

11 The formation of ground-level ozone ( $O_3$ ) is dependent on both atmospheric chemical  
12 processes and meteorological factors. In this study, a random forest (RF) model coupled  
13 with the reactivity of volatile organic compound (VOC) species was used to investigate  
14 the  $O_3$  formation sensitivity in Beijing, China, from 2014 to 2016, and evaluate the  
15 relative importance (RI) of chemical and meteorological factors to  $O_3$  formation. The  
16 results showed that the  $O_3$  prediction performance using concentrations of  
17 measured/initial VOC species ( $R^2 = 0.82/0.81$ ) was better than that using total VOCs  
18 (TVOCs) concentrations ( $R^2 = 0.77$ ). Meanwhile, the RIs of initial VOC species  
19 correlated well with their  $O_3$  formation potentials (OFPs), which indicate that the model  
20 results can be partially explained by the maximum incremental reactivity (MIR) method.  
21  $O_3$  formation presented a negative response to nitrogen oxides ( $NO_x$ ) and relative  
22 humidity (RH), and a positive response to temperature (T), solar radiation (SR) and  
23 VOCs. The  $O_3$  isopleth calculated by the RF model were generally comparable with  
24 those calculated by the box model.  $O_3$  formation shifted from a VOC-limited regime to  
25 a transition regime from 2014 to 2016. This study demonstrates that the RF model  
26 coupled with the initial concentrations of VOC species could provide an accurate,  
27 flexible, and computationally efficient approach for  $O_3$  sensitivity analysis.

28

29

30

## 31 **1. Introduction**

32 Ground-level ozone ( $O_3$ ) pollution, which can cause adverse human health effects  
33 such as cardiovascular and respiratory diseases, has received increasing attention in  
34 recent decades (Cohen et al., 2017). Oxidation of volatile organic compounds (VOCs)  
35 will produce peroxy radicals ( $RO_2$ ) and hydroperoxy radicals ( $HO_2$ ). The  $RO_2/HO_2$   
36 can accelerate the conversion from NO to  $NO_2$ , subsequently, formation of  $O_3$  by  
37 photolysis of  $NO_2$  in the presence of  $O_2$  (Wang et al., 2017a). The production and loss  
38 of  $RO_2$  and  $HO_2$  are highly dependent on the concentration ratio of VOCs and  $NO_x$  in  
39 the atmosphere. Hence, atmospheric  $O_3$  concentrations or production rates show a  
40 nonlinear relationship with VOCs and  $NO_x$ . Moreover, the  $O_3$ -VOC- $NO_x$  sensitivity is  
41 readily influenced by VOC species (Tan et al., 2018), meteorological parameters (Liu  
42 et al., 2020a; Liu et al., 2020), and even atmospheric particulate matter (Li et al., 2019),  
43 thus, exhibits high temporal and spatial variability. Therefore, it is urgent to develop an  
44 accurate and highly efficient method for timely assessing the sensitivity regime of  $O_3$   
45 production and evaluating the effectiveness of a potential measure on  $O_3$  pollution  
46 control.

47 The sensitivity of  $O_3$  formation can usually be analysed using observed indicators,  
48 such as ozone production efficiency (OPE,  $\Delta O_3/\Delta NO_z$ ) (Wang et al., 2010; Lin et al.,  
49 2011),  $HCHO/NO_y$  (Martin et al., 2004), and  $H_2O_2/NO_z$  (or  $H_2O_2/HNO_3$ ) (Sillman 1995;  
50 Hammer et al., 2002; Wang et al., 2017a), observation-based model (OBM) (Vélez-  
51 Pereira et al., 2021) and chemical transport models including community multiscale air

52 quality (CMAQ) (Djalalova et al., 2015) and Weather Research and Forecasting with  
53 Chemistry (WRF-Chem) model (Wang et al., 2020a).

54 The observed indicators can be utilized to quickly diagnose the sensitivity regime  
55 of O<sub>3</sub> production. However, the accuracy is sensitive to the precision of tracer  
56 measurements. OBMs combine *in-situ* field observations, **remote sensing**  
57 **measurements and** chemical box models, which are built on widely-used chemistry  
58 mechanisms (e.g., MCM, Carbon Bond, RACM or SAPRC), and applied to the  
59 observed atmospheric conditions to simulate the *in-situ* O<sub>3</sub> production rate (Mo et al.,  
60 2018). The sensitivity of O<sub>3</sub> production to various O<sub>3</sub> precursors, including **NO<sub>x</sub>** and  
61 VOCs can be diagnosed based on the empirical kinetic modeling approach (EKMA) or  
62 quantitatively assessed with the relative incremental reactivity (RIR). Chemical  
63 transport models, which are driven by meteorological dynamics and incorporated with  
64 the emissions of pollutants and the complex atmospheric chemical mechanism, provide  
65 a powerful tool for simulating various atmospheric processes, including spatial  
66 distribution, regional transport *vs.* local formation, source apportionment and  
67 production rates of pollutants and so on (Sayeed et al., 2021). At present, OBMs are  
68 widely used to investigate O<sub>3</sub> formation sensitivity in China. Previous studies indicated  
69 that O<sub>3</sub> formation in urban areas of China is located in a VOC-limited or a transition  
70 regime and varies with time and location (Ou et al., 2016; Wang et al., 2017a; Zhan et  
71 al., 2021). Although both OBMs and chemical transport models can assess the  
72 sensitivity of O<sub>3</sub> production and predict the O<sub>3</sub> pollution level in a scenario of control

73 measures, the calculation accuracy is affected by the uncertainty of input parameters  
74 (Tang et al., 2011; Yang et al., 2021b). Thus, they are mostly applied to sampling cases  
75 with a short time span (days or weeks) (Xue et al., 2014; Ou et al., 2016).

76 Compared to traditional methods, machine learning (ML) is able to capture the  
77 main factors affecting atmospheric O<sub>3</sub> formation in a timely manner with great  
78 flexibility (without the constraints of time and space) and high computational efficiency  
79 (Wang et al., 2020c; Grange et al., 2021; Yang et al., 2021a). Although attentions should  
80 be paid to the robustness of machine learning because it depends on the input dataset  
81 (observations or outputs of chemical transport models), previous studies have  
82 demonstrated that cross-validation and data-normalization can well reduce the  
83 dependence of the model on input data and improve the robustness of the model (Wang  
84 et al., 2016; Wang et al., 2017b; Liu et al., 2021; Ma et al., 2021a). Thus, it is a  
85 promising alternative to account for the effects of meteorology on air pollutants and has  
86 been intensively used in atmospheric studies (Liu et al., 2020a; Hou et al., 2022).

87 Recently, ML based on convolutional neural network (CNN), random forest (RF)  
88 and artificial neural network (ANN) models have been applied in simulating  
89 atmospheric O<sub>3</sub> and shown good performance in O<sub>3</sub> prediction (Ma et al., 2020; Xing  
90 et al., 2020). For example, Ma et al. (2021a) simulated O<sub>3</sub> concentrations in the Beijing-  
91 Tianjin-Hebei (BTH) region from 2010-2017 using an RF model that considered  
92 meteorological variables and output variables from chemical transport models, and the  
93 correlation coefficient ( $R^2$ ) between the observed and modelled O<sub>3</sub> concentrations was

94 greater than 0.8. Liu et al. (2021) also reported a high accuracy (80.4%) for classifying  
95 pollution levels of O<sub>3</sub> and fine particulate matter with aerodynamic diameter less than  
96 2.5 μm (PM<sub>2.5</sub>) at 1464 monitoring sites in China using an RF model. Thus, the RF  
97 model has shown good performance in terms of prediction accuracy and computational  
98 efficiency (Wang et al., 2016; Wang et al., 2017b).

99 It should be noted that physical interpretability of the results is an important  
100 question when ML models are applied in atmospheric studies due to the “black box”  
101 nature of most ML models (Hou et al., 2022). At the present, many ML studies have  
102 used total VOCs (TVOCs) to simulate O<sub>3</sub> formation and rarely considered the effect of  
103 VOC species on O<sub>3</sub> formation sensitivity (Feng et al., 2019; Liu et al., 2021; Ma et al.,  
104 2021a). Thus, they were unable to identify the chemical reactivity of a single species to  
105 O<sub>3</sub> formation, which may lead to underestimations or even misunderstandings of the  
106 role of VOCs in O<sub>3</sub> formation because the same concentration of TVOCs with different  
107 compositions may lead to different OPEs. In addition, VOCs react with OH radicals  
108 during atmospheric transport, which is the most important sink of VOCs (Carlo et al.,  
109 2004; Liu et al., 2020b). Makar et al. (1999) reported that the isoprene emissions were  
110 underestimated by up to 40% if the OH oxidation is not considered. Other studies  
111 indicated that the initial concentrations of VOCs, which account for the photochemical  
112 loss of VOCs during transport, were more representative of pollution levels in the  
113 sampling area than the observed VOCs (Yuan et al., 2013; Zhan et al., 2021). However,  
114 whether the ML model can identify the connection between the reactivity of VOC

115 species and O<sub>3</sub> formation sensitivity has not been clarified.

116 Although ML is widely used to understand air pollution, explanations of ML  
117 results (e.g., RI) are somewhat vague because ML is a “black-box” model from the  
118 point view of chemical mechanism (Hou et al., 2022; Taoufik et al., 2022). In this study,  
119 we used the RF model to evaluate the prediction performance of atmospheric O<sub>3</sub> using  
120 the TVOCs, measured VOC species and photochemical initial concentration (PIC) of  
121 VOC species, which is calculated based on the photochemical-age approach (Shao et  
122 al., 2011). We compared the relative importance (RI) of the precursors (VOC species,  
123 NO<sub>x</sub>, PM<sub>2.5</sub>, CO) and the meteorological parameters (temperature, solar radiation,  
124 relative humidity, wind speed and direction) on O<sub>3</sub> formation in the summer of Beijing  
125 from 2014 to 2016. We also discussed the possibility of connecting the RIs of VOCs  
126 with their OFPs and the changes in O<sub>3</sub>-VOC-NO<sub>x</sub> sensitivity based on the RF model  
127 from 2014 to 2016. Our study indicates that the RF model combined with initial  
128 concentrations of VOC species can simulate O<sub>3</sub> concentrations well and provides a  
129 flexible and efficient tool for O<sub>3</sub> modelling in a near real-time way.

## 130 **2. Methods**

### 131 **2.1 Sampling site and data**

132 The sampling site (40.04°N, 116.42°E) is located at the campus of Chinese  
133 Research Academy of Environmental Sciences and was described in our previous work  
134 (Zhang et al., 2021). Briefly, the station is located two kilometers from the north 4<sup>th</sup> ring  
135 road and surrounded by a mixed residential and commercial area. The concentrations  
136 of VOCs, NO<sub>x</sub>, CO, O<sub>3</sub> and PM<sub>2.5</sub> were measured at 8 m above ground level at this

137 location. Meteorological parameters, including temperature (T), relative humidity (RH),  
138 wind speed and direction (WS&WD), solar radiation (SR), were monitored at 15 m  
139 above ground level. VOCs were measured by an online commercial instrument (GC-  
140 866, Chromatotec, France), which consisted of two independent analysers for detecting  
141 C<sub>2</sub>-C<sub>6</sub> and C<sub>6</sub>-C<sub>12</sub> hydrocarbon components. More details about the observations can be  
142 found in the Supplemental Materials (S1). The calculation of initial VOCs and  
143 sensitivity tests can be found in the Supplemental Materials (S2).

## 144 **2.2 Random forest model**

145 The random forest (RF) is a type of ensemble decision tree that can be used for  
146 classification and regression (Breiman 2001). During the training process, the model  
147 creates a large number of different decision trees with different sample sets at each node,  
148 and then averages the results of all decision trees as its final results (Breiman 2001). To  
149 avoid over-fitting, we trained the random forest model using cross-validation for the  
150 normalized data, which can improve the robustness of the model. Briefly, we randomly  
151 divided the normalized data into 12 subsets, then alternately took one subset as testing  
152 data along with the rest as training data. By doing this, every data point has an equal  
153 chance being trained and tested. The length of the input data from 2014 to 2016 were  
154 1190, 1062 and 872 rows, respectively, in which different types of VOCs, NO<sub>x</sub>, CO,  
155 PM<sub>2.5</sub> and meteorological parameters (including temperature, relative humidity, solar  
156 radiation, wind speed and direction) were used as input variables and O<sub>3</sub> as output  
157 variables. The mean values ( $\pm$ standard deviation) of input/output parameters are shown



158 in Table S1. Approximately one-third of the samples are excluded from the sample,  
159 when the decision tree is built and used to calculate the out-of-bag data error. Hence,  
160 RF can evaluate the RI of variables via the changes in out-of-bag (OOB) data error  
161 (Svetnik et al., 2003),

$$162 \quad RI_i = \sum (\text{errOOB2}_i - \text{errOOB1}_i) / N \quad (1)$$

163 where N represents the number of decision trees, and **errOOB1** and **errOOB2** represent  
164 the out-of-bag data error of feature *i* before and after randomly permuting the  
165 observation, respectively. The  $RI_i$  used to evaluate the importance and sensitivity of  
166 feature *i* to O<sub>3</sub> formation in this study. More details about workflow of RF model and  
167 the hyperparameter tuning can be found in the Text S3. The optimized parameters are  
168 shown in Table S2. To verify the stability of the model, we performed a significance  
169 test on the model results. The results showed that there was no significant difference  
170 among the different tests ( $P > 0.05$ ,  $R^2 > 0.98$ ).

171 When plotting the O<sub>3</sub> formation sensitivity curves, we made a virtual matrix of  
172 inputs by varying the concentrations of NO<sub>x</sub> and VOCs from 0.9 to 1.1 times (with a  
173 step of 0.01) of their mean values while keeping all other inputs unchanged (i.e., the  
174 mean values). Then, the new matrix was used as testing data, while all the measured  
175 data were taken as training data. Thus, the testing data should represent the mean  
176 sensitivity regime of O<sub>3</sub> in Beijing, while the training data actually covered all the  
177 sensitivity regimes of O<sub>3</sub> formation to guarantee a sufficient coverage in the NO<sub>x</sub>-  
178 limited regime for the RF model simulations. The EKMA curves were plotted using the

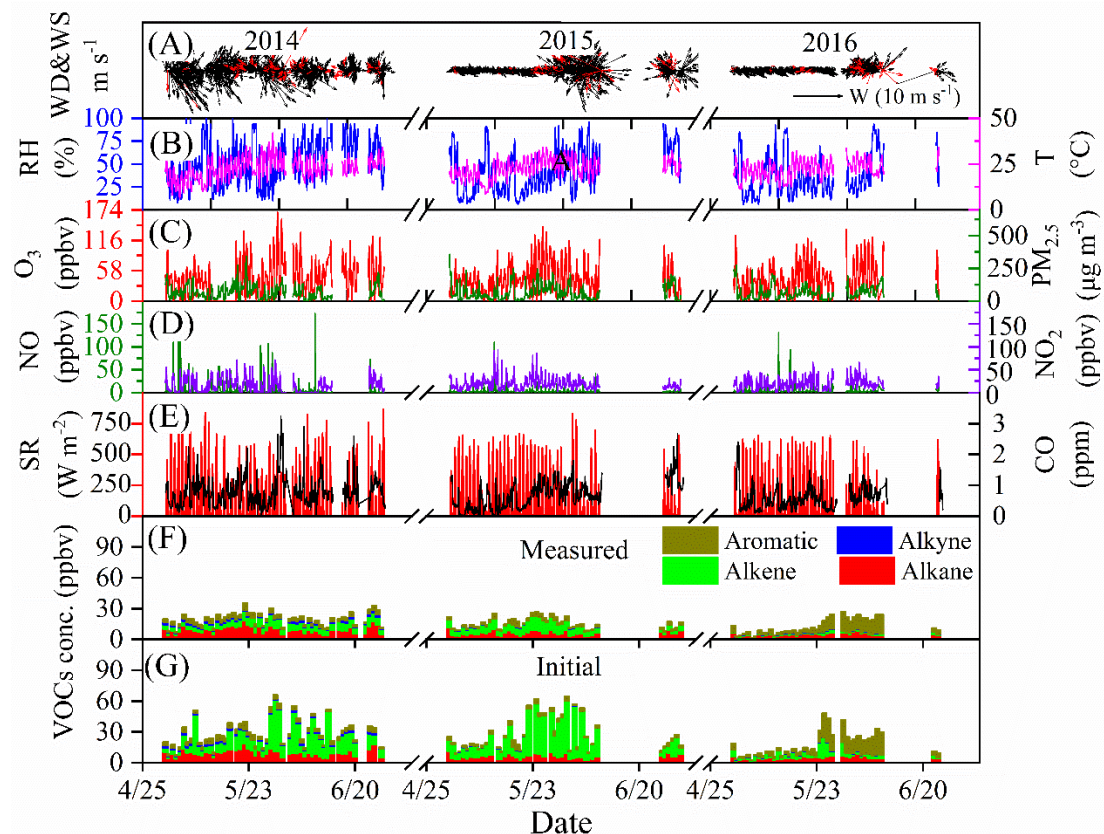
179 daily maximum 8-h (MDA8) O<sub>3</sub>. More details can be found in the SI.

## 180 **3. Results and discussion**

### 181 **3.1 Overview of air pollutants and meteorological conditions**

182 Figure 1 shows the time series of air pollutants and meteorological parameters  
183 during the observations from 2014 to 2016. In 2014, 2015 and 2016, the wind direction  
184 was dominated by northwest winds (Figure S1), with mean wind speeds of  $3.1 \pm 2.7$  m  
185  $s^{-1}$ ,  $2.3 \pm 2.2$  m  $s^{-1}$ , and  $1.3 \pm 1.2$  m  $s^{-1}$ , respectively, and the mean daytime temperature  
186 were  $22.3 \pm 5.8$ ,  $23.9 \pm 5.0$  and  $24.0 \pm 4.4$  °C, respectively. The average value of SR  
187 decreased from 162.9 to 150.8 W  $m^{-2}$  during the observation period. As shown in Figure  
188 1F-G, in 2014, 2015 and 2016, the mean VOC concentrations were  $20.3 \pm 10.9$ ,  $15.8 \pm$   
189  $8.3$  and  $12.1 \pm 7.7$  ppbv, respectively, while the mean initial VOC concentrations were  
190  $28.1 \pm 25.7$ ,  $27.2 \pm 32.6$  and  $16.4 \pm 16.1$  ppbv, respectively. The calculation of initial  
191 VOCs and sensitivity tests can be found in the Supplemental Materials (S2). Both the  
192 measured VOCs and initial VOCs showed a decline along with a decrease in PM<sub>2.5</sub>  
193 concentration from  $67.2 \pm 53.5$  to  $61.1 \pm 48.6$   $\mu g m^{-3}$  due to the Air Pollution Prevention  
194 and Control Action Plan in China (Zhao et al., 2021). However, O<sub>3</sub> concentrations  
195 showed a slight downward trend from  $44.3 \pm 32.4$  to  $42.7 \pm 27.9$  ppbv from 2014 to  
196 2015 and then reach to  $44.0 \pm 29.6$  ppbv in 2016. A slight upward trend was observed  
197 for NO<sub>x</sub> concentrations (Figure S2). As shown in Figure 1F-G, the concentrations of  
198 four types (alkanes, alkenes, alkynes, and aromatics) of VOCs showed significant  
199 differences from 2014 to 2016 due to the variations in emission sources (Zhang et al.,

200 2021). In addition to VOC species, the variations in other parameters, such as  
 201 meteorological conditions and PM<sub>2.5</sub>, should have a complex influence on O<sub>3</sub>-VOC-  
 202 NO<sub>x</sub> sensitivity (Li et al., 2019; Ma et al., 2021b).



203  
 204 **Figure 1.** Time series of air pollutants and meteorological parameters during  
 205 observations in Beijing. (In A, the red arrows represent the O<sub>3</sub> concentration exceed  
 206 74.6 ppbv according to the national ambient air quality standard.)

### 207 3.2 Prediction performance of the model.

208 To build a robust model, we evaluated the prediction performance of the RF model  
 209 for the ambient O<sub>3</sub> simulation. Figure 2 shows the O<sub>3</sub> prediction performance in 2015  
 210 when chemical species (including VOCs, NO<sub>x</sub>, PM<sub>2.5</sub>, CO) and meteorological factors  
 211 (i.e., WS, WD, SR, T and RH) were used as inputs in the RF model. The prediction

212 performance of RF model for 2014 and 2016 is shown in Figures S3 and S4 respectively.

213 The details of the modelling and input parameters are shown in Table S2. Figure 2A-C

214 shows the time series of the measured and modelled O<sub>3</sub> concentrations, which were

215 simulated using the TVOCs, measured VOC species and initial VOC species as part

216 input variables along with the same set of other parameters. The correlation coefficients

217 ( $R^2$ ) of the training data were 0.77, 0.82 and 0.81 for the TVOCs, measured VOC

218 species and initial VOC species, respectively. The corresponding root mean squared

219 errors (RMSEs) for the predicted O<sub>3</sub> concentrations were 17.4, 12.6 and 13.9. Figure

220 2D-F shows the prediction performance of the testing dataset under these three

221 circumstances. When the TVOCs were split into measured or initial VOC species, the

222  $R^2$  increased obviously as the number of data features increased. Therefore, the VOC

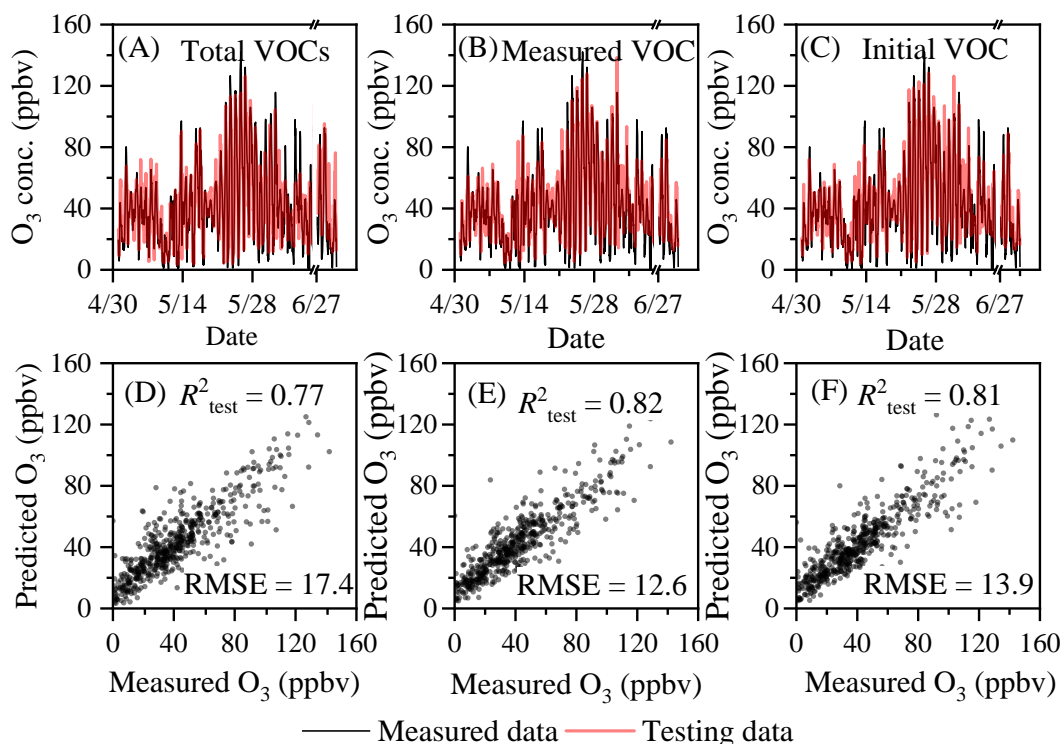
223 composition has a significant influence on O<sub>3</sub> prediction using the RF model. In

224 previous studies using TVOCs, the influence of VOC composition was neglected (Liu

225 et al., 2021; Ma et al., 2021a). Our results indicate that the RF model can accurately

226 predict O<sub>3</sub> concentrations when the concentrations of measured/initial VOC species are

227 considered.



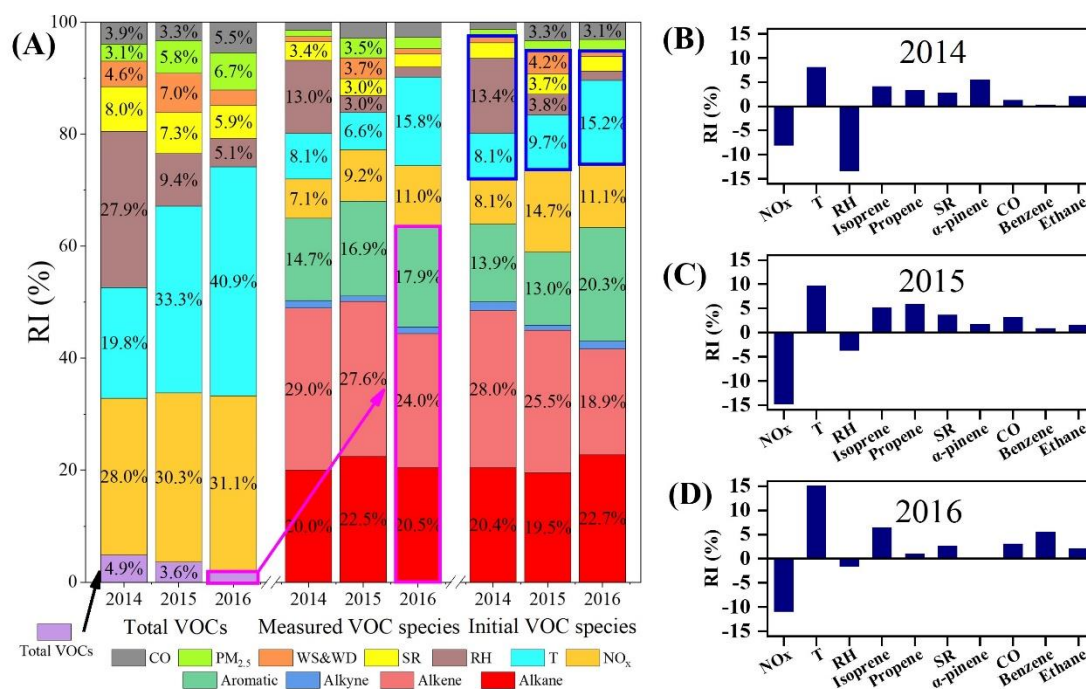
228

229 **Figure 2.** Comparison of the predicted and measured O<sub>3</sub> concentrations in Beijing in  
 230 the summer of 2015. (A and D: TVOC concentrations; B and E: measured  
 231 concentrations of VOC species; C and F: initial concentrations of VOC species)

232 It should be pointed out that if the training dataset does not have sufficient  
 233 coverage in the NO<sub>x</sub>-limited regime, then the trained algorithm essentially attempts to  
 234 extrapolate in that regime, which is prone to overtraining. To avoid such overtraining,  
 235 a 12-fold cross-validation by randomly dividing the observation data in each day into  
 236 12 subsets and alternately taking one subset as testing data and the rest as training data  
 237 ensures that each data point has an equal chance of being trained and tested. The curves  
 238 of the predicted O<sub>3</sub> concentrations in Figure 2 were spliced using the testing datasets in  
 239 all runs. Thus, our results actually covered all the sensitivity regimes of O<sub>3</sub> formation.  
 240 This means that the model is robust

### 241 3.3 Relative importance of major factors

242 Figure 3A shows the RIs of different ambient factors, including chemical and  
243 meteorological variables on O<sub>3</sub> formation. The difference in the RIs is also compared  
244 using the TVOCs and the VOC species as inputs. Chemical factors (including VOC  
245 species, NO<sub>x</sub>, PM<sub>2.5</sub> and CO) accounted for 79.1% of the contribution to O<sub>3</sub> production  
246 in the summer of 2016. Meanwhile, VOC species accounted for approximately 63.4%  
247 of O<sub>3</sub> production while the RIs using TVOC concentrations accounted for only 2.1%.  
248 Ma et al. (2021b) analysed the contribution of meteorological conditions and chemical  
249 factors to O<sub>3</sub> formation on the North China Plain (NCP) using the CMAQ model in  
250 combination with process analysis and found that chemical factors dominate O<sub>3</sub>  
251 formation in summer. Using probability theory, Ueno et al. (2019) also found that  
252 VOCs/NO<sub>x</sub> dominate O<sub>3</sub> production compared to meteorological variables. Thus, our  
253 results are similar to those of previous studies based on chemical models (Ueno et al.,  
254 2019), which demonstrates that the RF model can reflect the contribution of VOC  
255 species to O<sub>3</sub> production even if the observed VOC species are used.



256

257 **Figure 3.** Percentage of RI for O<sub>3</sub> precursors and meteorological parameters (A) and  
 258 the top 10 factors with high values of RI in 2014-2016 (B-D: using initial concentrations  
 259 of VOC species).

260 Here, we compared the RIs of VOCs calculated using the initial VOC species and  
 261 the observed VOC species with the O<sub>3</sub> formation potentials (OFPs). The OFPs were  
 262 calculated by the maximum incremental reactivity (MIR) method (Carter 2010). As  
 263 shown in Figure S5, the RIs showed good correlations with the OFP. Interestingly, the  
 264 initial concentrations of VOC species improved the correlation coefficients between the  
 265 RIs and OFPs. Furthermore, we calculated the RIs and OFPs of different species using  
 266 the observed data during the campaign study in Daxing District in the summer of 2019  
 267 (Zhan et al., 2021), and a **stronger** correlation was observed between the RIs of the  
 268 initial VOC species and the OFPs (Figure S6). These results indicate that the RIs of the  
 269 initial VOCs species in the ML model should partially reflect the chemical reactivity of

270 VOCs to produce O<sub>3</sub> in the atmosphere.

271 Although the RIs calculated using the initial VOC species slightly changed  
272 compared to those calculated using the observed VOCs (Table S3), VOCs still  
273 dominated O<sub>3</sub> formation (Figure 3A). For example, the initial VOCs dominated O<sub>3</sub>  
274 production in 2014, 2015, and 2016, with RI values of 64.0, 59.0 and 63.3%  
275 respectively. Li et al. (2020a) used a multiple linear regression (MLR) model to study  
276 the contribution of anthropogenic and meteorological factors to O<sub>3</sub> formation in China  
277 from 2013-2019 and found that meteorological factors accounted for 36.8% and  
278 anthropogenic factors accounted for 63.2%, which is similar to our results. Figure 3B-  
279 D shows the top 10 factors having a strongly influence on O<sub>3</sub> production. Interestingly,  
280 NO<sub>x</sub> and RH showed negative responses to O<sub>3</sub> formation, while other variables,  
281 including T, SR, CO and all of the VOCs, showed positive responses. Thus, a decrease  
282 in NO<sub>x</sub> or RH will lead to an increase in O<sub>3</sub> concentration while a decrease in T, SR,  
283 CO and VOCs will lead to a decrease in O<sub>3</sub> concentration. Although O<sub>3</sub> formation is  
284 highly related to the photolysis of NO<sub>2</sub>, a previous study demonstrated that it is VOC-  
285 limited in summer in Beijing (Zhan et al., 2021). This finding is consistent with the  
286 observed negative response of O<sub>3</sub> to NO<sub>x</sub> in this work. High RH usually coincides with  
287 low surface O<sub>3</sub> concentrations in field observations, which can be ascribed to the  
288 inhibition of O<sub>3</sub> formation by the transfer of NO<sub>2</sub>/ONO<sub>2</sub>-containing products into the  
289 particle phase and the promotion of dry deposition of O<sub>3</sub> on the surface (Kavassalis et  
290 al., 2017; Yu 2019). These previous works can well explain the observed negative

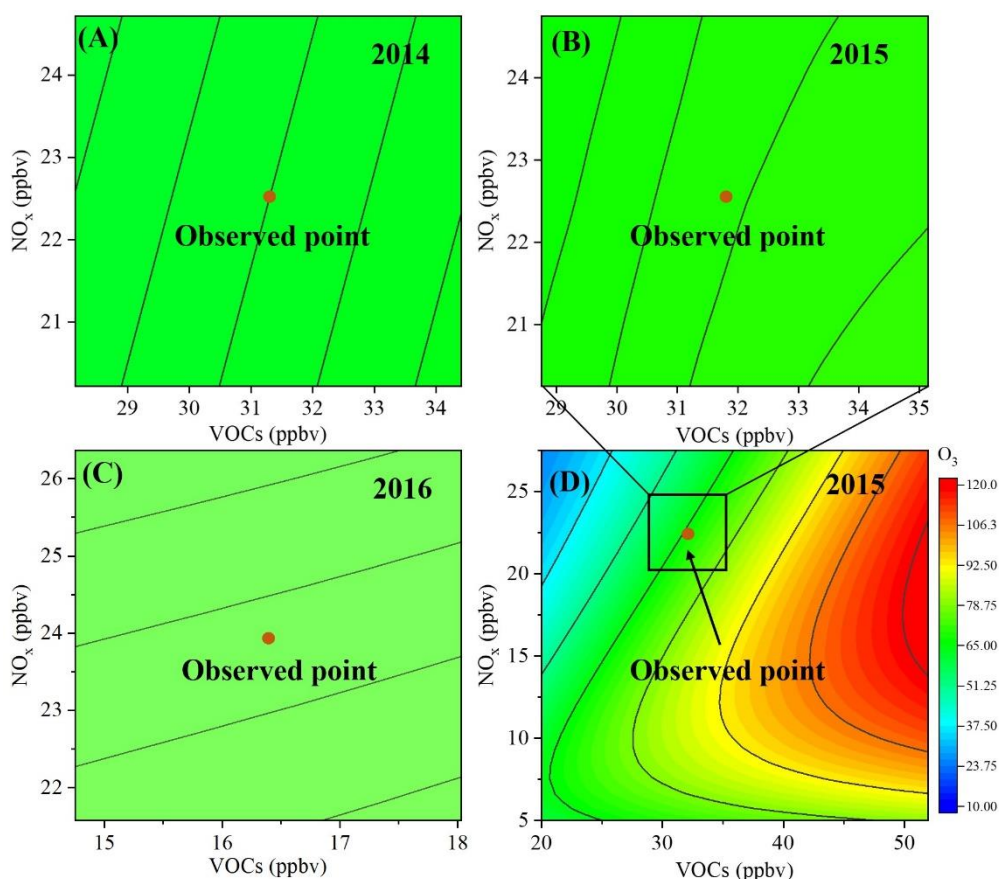


291 response of O<sub>3</sub> to RH in **Figure 3B-D**. Previous studies have observed a positive  
292 correlation between the O<sub>3</sub> concentration and T or SR (Steiner et al., 2010; Paraschiv  
293 et al., 2020; Li et al., 2021). Temperature can directly affect the chemical reaction rate  
294 of O<sub>3</sub> formation (Fu et al., 2015), and SR can promote the photolysis of NO<sub>2</sub> (Hu et al.,  
295 2017; Wang et al., 2020b), thus accelerating O<sub>3</sub> formation. As mentioned above, O<sub>3</sub>  
296 formation is VOC-limited in Beijing; thus, a positive response of O<sub>3</sub> concentration to  
297 VOCs is observed in Figure 3B. Interestingly, the RIs of isoprene showed an increasing  
298 trend from 2014 to 2016 because of the obvious reduction in anthropogenic VOCs  
299 (Figure S7) (Zhang et al., 2021). In the context of global warming, studies should focus  
300 on the factors that affect O<sub>3</sub> formation, including biogenic emissions, T and SR. Thus,  
301 additional efforts will be required to reduce anthropogenic pollutants in the future.

### 302 **3.4 Ozone formation sensitivity**

303 To further analyse the sensitivity of O<sub>3</sub> to VOCs and **NO<sub>x</sub>** from 2014 to 2016, we  
304 plotted sensitivity curves for O<sub>3</sub> generation using the RF model, and the results are  
305 shown in Figure 4A-C. Moreover, EKMA curves in 2015 were also obtained using the  
306 OBM (Figure 4D). As shown in Figure 4A-C, O<sub>3</sub> formation was sensitive to VOCs in  
307 the summer of Beijing during our observations, which is consistent with previous  
308 studies that used box models (Li et al., 2020b) and chemical transport models (Shao et  
309 al., 2021). This result is also consistent with the RIs of VOCs or **NO<sub>x</sub>** to O<sub>3</sub> formation  
310 (**Figure 3B-D**). Interestingly, the O<sub>3</sub> formation sensitivity to VOCs decreases or  
311 gradually shifts from the observed point to the transition regime from 2014 to 2016

312 (Figure 4A-C), which is similar to that reported by Zhang et al. (2021). These  
313 phenomena can be ascribed to the increased **relative** importance of meteorological  
314 factors, such as T, SR, and RH, for O<sub>3</sub> formation and the variation in anthropogenic  
315 VOC emissions (Steiner et al., 2010; Ma et al., 2021b).



316

317 **Figure 4.** Ozone formation sensitivity curves from 2014-2016. (A, B, C: calculated by  
318 the RF model for 2014, 2015, and 2016, respectively. D: calculated by the OBM for  
319 2015.)

320 We compared **the relative error of simulated MDA8 O<sub>3</sub> calculated** using the RF  
321 **and OBM model** in 2015, as shown in Figure S8. **The mean relative error of simulated**  
322 **MDA8 O<sub>3</sub> between RF model and Box model was 15.6%.** Hence, a combination of the  
323 RF model and initial VOCs species can accurately depict the sensitivity regime of O<sub>3</sub>

324 formation, while the calculated RIs correlate well with the OFPs.

#### 325 **4. Conclusions**

326 In summary, this work investigated O<sub>3</sub> formation sensitivity in the summer from  
327 2014-2016 in Beijing using the RF model coupled with the reactivity of VOC species.  
328 The results show that the prediction performance of O<sub>3</sub> by the RF model was  
329 significantly improved when measured/initial VOC species were considered compared  
330 to TVOCs. Furthermore, after the photochemical loss of VOC species during transport  
331 was corrected, the RIs of the VOC species were well correlated with the OFPs of VOC  
332 species calculated using the MIR method, thus indicating that the RIs in the ML model  
333 reflect the chemical reactivity of VOCs. Meanwhile, both NO<sub>x</sub> and highly reactive  
334 species (such as isoprene, propene, benzene) played an important role in O<sub>3</sub> formation.  
335 An increased contribution of temperature to O<sub>3</sub> production was observed, which  
336 implied the importance of temperature to O<sub>3</sub> pollution in the context of global warming  
337 conditions. Both the RF model and the box model results showed that O<sub>3</sub> formation was  
338 sensitive to VOCs in Beijing, although the sensitivity regime shifted from VOC-limited  
339 regime to a transition regime from 2014 to 2016. Due to the high computational  
340 efficiency of ML, the O<sub>3</sub> formation sensitivity plotted by the RF model coupled with  
341 the reactivity of VOC species can provide an accurate, flexible and efficient approach  
342 for analysing O<sub>3</sub> sensitivity in a near real-time way.

343

#### 344 **Code and data availability**

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

#### 349 **Supplement**

350 Supplementary information is available for this paper.

#### 351 **Author contributions**

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

#### 357 **Competing interest**

358 The authors declare that they have no conflict of interest.

#### 359 **Acknowledgments**

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

365

366 **References**

- 367 Breiman, L. Random Forests. *Machine Learning*, 45, 5-32, 10.1023/A:1010933404324, 2001.
- 368 Carlo, P.D., Brune, W.H., Martinez, M., Harder, H., Leshner, R., Ren, X., Thornberry, T., Carroll,  
369 M.A., Young, V., Shepson, P.B., Riemer, D., Apel, E., Campbell, C. Missing OH Reactivity  
370 in a Forest: Evidence for Unknown Reactive Biogenic VOCs. *Science*, 304, 722-725,  
371 doi:10.1126/science.1094392, 2004.
- 372 Carter, W. Updated maximum incremental reactivity scale and hydrocarbon bin reactivities for  
373 regulatory applications. California Air Resources Board Contract, 1, 07-339, 2010.
- 374 Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., Balakrishnan, K.,  
375 Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling,  
376 A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., Pope, C.A., Shin, H., Straif, K.,  
377 Shaddick, G., Thomas, M., van Dingenen, R., van Donkelaar, A., Vos, T., Murray, C.J.L.,  
378 Forouzanfar, M.H. Estimates and 25-year trends of the global burden of disease attributable  
379 to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015.  
380 *The Lancet*, 389, 1907-1918, [https://doi.org/10.1016/S0140-6736\(17\)30505-6](https://doi.org/10.1016/S0140-6736(17)30505-6), 2017.
- 381 Djalalova, I., Delle Monache, L., Wilczak, J. PM<sub>2.5</sub> analog forecast and Kalman filter post-  
382 processing for the Community Multiscale Air Quality (CMAQ) model. *Atmos. Environ.*,  
383 108, 76-87, <https://doi.org/10.1016/j.atmosenv.2015.02.021>, 2015.
- 384 Feng, R., Zheng, H.-j., Gao, H., Zhang, A.-r., Huang, C., Zhang, J.-x., Luo, K., Fan, J.-r. Recurrent  
385 Neural Network and random forest for analysis and accurate forecast of atmospheric  
386 pollutants: A case study in Hangzhou, China. *J. Clean. Prod.*, 231, 1005-1015,  
387 <https://doi.org/10.1016/j.jclepro.2019.05.319>, 2019.
- 388 Fu, T.-M., Zheng, Y., Paulot, F., Mao, J., Yantosca, R.M. Positive but variable sensitivity of August  
389 surface ozone to large-scale warming in the southeast United States. *Nat. Clim. Change*, 5,  
390 454-458, 10.1038/nclimate2567, 2015.
- 391 Grange, S.K., Lee, J.D., Drysdale, W.S., Lewis, A.C., Hueglin, C., Emmenegger, L., Carslaw, D.C.  
392 COVID-19 lockdowns highlight a risk of increasing ozone pollution in European urban  
393 areas. *Atmos. Chem. Phys.*, 21, 4169-4185, 10.5194/acp-21-4169-2021, 2021.
- 394 Hammer, M.-U., Vogel, B., Vogel, H. Findings on H<sub>2</sub>O<sub>2</sub>/HNO<sub>3</sub> as an indicator of ozone sensitivity  
395 in Baden-Württemberg, Berlin-Brandenburg, and the Po valley based on numerical  
396 simulations. *Journal of Geophysical Research: Atmospheres*, 107, LOP 3-1-LOP 3-18,  
397 <https://doi.org/10.1029/2000JD000211>, 2002.
- 398 Hou, L., Dai, Q., Song, C., Liu, B., Guo, F., Dai, T., Li, L., Liu, B., Bi, X., Zhang, Y., Feng, Y.  
399 Revealing Drivers of Haze Pollution by Explainable Machine Learning. *Environ. Sci.*  
400 *Technol. Lett.*, 10.1021/acs.estlett.1c00865, 2022.
- 401 Hu, B., Zhao, X., Liu, H., Liu, Z., Song, T., Wang, Y., Tang, L., Xia, X., Tang, G., Ji, D., Wen, T.,  
402 Wang, L., Sun, Y., Xin, J. Quantification of the impact of aerosol on broadband solar  
403 radiation in North China. *Sci. Rep.*, 7, 44851, 10.1038/srep44851, 2017.
- 404 Kavassalis, S.C., Murphy, J.G. Understanding ozone-meteorology correlations: A role for dry  
405 deposition. *Geophys. Res. Lett.*, 44, 2922-2931, <https://doi.org/10.1002/2016GL071791>,  
406 2017.
- 407 Li, J., Cai, J., Zhang, M., Liu, H., Han, X., Cai, X., Xu, Y. Model analysis of meteorology and

408 emission impacts on springtime surface ozone in Shandong. *Sci. Total Environ.*, 771,  
409 144784, <https://doi.org/10.1016/j.scitotenv.2020.144784>, 2021.

410 Li, K., Jacob, D.J., Liao, H., Zhu, J., Shah, V., Shen, L., Bates, K.H., Zhang, Q., Zhai, S. A two-  
411 pollutant strategy for improving ozone and particulate air quality in China. *Nat. Geosci.*,  
412 12, 906-910, 10.1038/s41561-019-0464-x, 2019.

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

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

419 Lin, W., Xu, X., Ge, B., Liu, X. Gaseous pollutants in Beijing urban area during the heating period  
420 2007–2008: variability, sources, meteorological, and chemical impacts. *Atmos. Chem.*  
421 *Phys.*, 11, 8157-8170, 10.5194/acp-11-8157-2011, 2011.

422 Liu, H., Liu, J., Liu, Y., Ouyang, B., Xiang, S., Yi, K., Tao, S. Analysis of wintertime O<sub>3</sub> variability  
423 using a random forest model and high-frequency observations in Zhangjiakou—an area  
424 with background pollution level of the North China Plain. *Environ. Pollut.*, 262, 114191,  
425 <https://doi.org/10.1016/j.envpol.2020.114191>, 2020a.

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

430 Liu, Y., Wang, T. Worsening urban ozone pollution in China from 2013 to 2017 – Part 1: The  
431 complex and varying roles of meteorology. *Atmos. Chem. Phys.*, 20, 6305-6321,  
432 10.5194/acp-20-6305-2020, 2020.

433 Liu, Z., Qi, Z., Ni, X., Dong, M., Ma, M., Xue, W., Zhang, Q., Wang, J. How to apply O<sub>3</sub> and PM<sub>2.5</sub>  
434 collaborative control to practical management in China: A study based on meta-analysis  
435 and machine learning. *Sci. Total Environ.*, 772, 145392,  
436 <https://doi.org/10.1016/j.scitotenv.2021.145392>, 2021.

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

440 Ma, R., Ban, J., Wang, Q., Zhang, Y., Yang, Y., He, M.Z., Li, S., Shi, W., Li, T. Random forest model  
441 based fine scale spatiotemporal O<sub>3</sub> trends in the Beijing-Tianjin-Hebei region in China,  
442 2010 to 2017. *Environ. Pollut.*, 276, 116635, <https://doi.org/10.1016/j.envpol.2021.116635>,  
443 2021a.

444 Ma, S., Shao, M., Zhang, Y., Dai, Q., Xie, M. Sensitivity of PM<sub>2.5</sub> and O<sub>3</sub> pollution episodes to  
445 meteorological factors over the North China Plain. *Sci. Total Environ.*, 792, 148474,  
446 <https://doi.org/10.1016/j.scitotenv.2021.148474>, 2021b.

447 Makar, P.A., Fuentes, J.D., Wang, D., Staebler, R.M., Wiebe, H.A. Chemical processing of biogenic  
448 hydrocarbons within and above a temperate deciduous forest. *J. Geophys. Res. Atmos.*, 104,  
449 3581-3603, <https://doi.org/10.1029/1998JD100065>, 1999.

450 Martin, R.V., Fiore, A.M., Van Donkelaar, A. Space-based diagnosis of surface ozone sensitivity to  
451 anthropogenic emissions. *Geophys. Res. Lett.*, 31, <https://doi.org/10.1029/2004GL019416>,  
452 2004.

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

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

459 Paraschiv, S., Barbuta-Misu, N., Paraschiv, S.L. Influence of NO<sub>2</sub>, NO and meteorological  
460 conditions on the tropospheric O<sub>3</sub> concentration at an industrial station. *Energy Rep.*, 6,  
461 231-236, <https://doi.org/10.1016/j.egy.2020.11.263>, 2020.

462 Sayeed, A., Choi, Y., Eslami, E., Jung, J., Lops, Y., Salman, A.K., Lee, J.-B., Park, H.-J., Choi, M.-  
463 H. A novel CMAQ-CNN hybrid model to forecast hourly surface-ozone concentrations 14  
464 days in advance. *Sci. Rep.*, 11, 10891, [10.1038/s41598-021-90446-6](https://doi.org/10.1038/s41598-021-90446-6), 2021.

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

470 Sillman, S. The use of NO<sub>y</sub>, H<sub>2</sub>O<sub>2</sub>, and HNO<sub>3</sub> as indicators for ozone-NO<sub>x</sub>-hydrocarbon sensitivity  
471 in urban locations. *J. Geophys. Res. Atmos.*, 100, 14175-14188,  
472 <https://doi.org/10.1029/94JD02953>, 1995.

473 Steiner, A.L., Davis, A.J., Sillman, S., Owen, R.C., Michalak, A.M., Fiore, A.M. Observed  
474 suppression of ozone formation at extremely high temperatures due to chemical and  
475 biophysical feedbacks. *P. Natl. Acad. Sci.*, 107, 19685-19690, [10.1073/pnas.1008336107](https://doi.org/10.1073/pnas.1008336107),  
476 2010.

477 Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P. Random Forest: A  
478 Classification and Regression Tool for Compound Classification and QSAR Modeling. *J.*  
479 *Chem. Inf. Comput. Sci.*, 43, 1947-1958, [10.1021/ci034160g](https://doi.org/10.1021/ci034160g), 2003.

480 Tan, Z., Lu, K., Jiang, M., Su, R., Dong, H., Zeng, L., Xie, S., Tan, Q., Zhang, Y. Exploring ozone  
481 pollution in Chengdu, southwestern China: A case study from radical chemistry to O<sub>3</sub>-  
482 VOC-NO<sub>x</sub> sensitivity. *Sci. Total Environ.*, 636, 775-786,  
483 <https://doi.org/10.1016/j.scitotenv.2018.04.286>, 2018.

484 Tang, X., Zhu, J., Wang, Z.F., Gbaguidi, A. Improvement of ozone forecast over Beijing based on  
485 ensemble Kalman filter with simultaneous adjustment of initial conditions and emissions.  
486 *Atmos. Chem. Phys.*, 11, 12901-12916, [10.5194/acp-11-12901-2011](https://doi.org/10.5194/acp-11-12901-2011), 2011.

487 Taoufik, N., Boumya, W., Achak, M., Chennouk, H., Dewil, R., Barka, N. The state of art on the  
488 prediction of efficiency and modeling of the processes of pollutants removal based on  
489 machine learning. *Sci. Total Environ.*, 807, 150554,  
490 <https://doi.org/10.1016/j.scitotenv.2021.150554>, 2022.

491 Ueno, H., Tsunematsu, N. Sensitivity of ozone production to increasing temperature and reduction

492 of precursors estimated from observation data. *Atmos. Environ.*, 214, 116818,  
493 <https://doi.org/10.1016/j.atmosenv.2019.116818>, 2019.

494 Vélez-Pereira, A.M., De Linares, C., Belmonte, J. Aerobiological modeling I: A review of predictive  
495 models. *Sci. Total Environ.*, 795, 148783, <https://doi.org/10.1016/j.scitotenv.2021.148783>,  
496 2021.

497 Wang, P., Qiao, X., Zhang, H. Modeling PM<sub>2.5</sub> and O<sub>3</sub> with aerosol feedbacks using WRF/Chem  
498 over the Sichuan Basin, southwestern China. *Chemosphere*, 254, 126735,  
499 <https://doi.org/10.1016/j.chemosphere.2020.126735>, 2020a.

500 Wang, T., Nie, W., Gao, J., Xue, L.K., Gao, X.M., Wang, X.F., Qiu, J., Poon, C.N., Meinardi, S.,  
501 Blake, D., Wang, S.L., Ding, A.J., Chai, F.H., Zhang, Q.Z., Wang, W.X. Air quality during  
502 the 2008 Beijing Olympics: secondary pollutants and regional impact. *Atmos. Chem. Phys.*,  
503 10, 7603-7615, 10.5194/acp-10-7603-2010, 2010.

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

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

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

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

516 Wang, Y., Wu, G., Deng, L., Tang, Z., Wang, K., Sun, W., Shangguan, Z. Prediction of aboveground  
517 grassland biomass on the Loess Plateau, China, using a random forest algorithm. *Sci. Rep.*,  
518 7, 6940, 10.1038/s41598-017-07197-6, 2017b.

519 Xing, J., Zheng, S., Ding, D., Kelly, J.T., Wang, S., Li, S., Qin, T., Ma, M., Dong, Z., Jang, C., Zhu,  
520 Y., Zheng, H., Ren, L., Liu, T.-Y., Hao, J. Deep Learning for Prediction of the Air Quality  
521 Response to Emission Changes. *Environ. Sci. Technol.*, 54, 8589-8600,  
522 10.1021/acs.est.0c02923, 2020.

523 Xue, L.K., Wang, T., Gao, J., Ding, A.J., Zhou, X.H., Blake, D.R., Wang, X.F., Saunders, S.M., Fan,  
524 S.J., Zuo, H.C., Zhang, Q.Z., Wang, W.X. Ground-level ozone in four Chinese cities:  
525 precursors, regional transport and heterogeneous processes. *Atmos. Chem. Phys.*, 14,  
526 13175-13188, 10.5194/acp-14-13175-2014, 2014.

527 Yang, J., Wen, Y., Wang, Y., Zhang, S., Pinto, J.P., Pennington, E.A., Wang, Z., Wu, Y., Sander, S.P.,  
528 Jiang, J.H., Hao, J., Yung, Y.L., Seinfeld, J.H. From COVID-19 to future electrification:  
529 Assessing traffic impacts on air quality by a machine-learning model. *P. Natl. Acad. Sci.*,  
530 118, e2102705118, 10.1073/pnas.2102705118, 2021a.

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



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

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

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

542 Zhang, X., Li, H., Wang, X., Zhang, Y., Bi, F., Wu, Z., Liu, Y., Zhang, H., Gao, R., Xue, L., Zhang,  
543 Q., Chen, Y., Chai, F., Wang, W. Heavy ozone pollution episodes in urban Beijing during  
544 the early summertime from 2014 to 2017: Implications for control strategy. Environ. Pollut.,  
545 285, 117162, <https://doi.org/10.1016/j.envpol.2021.117162>, 2021.

546 Zhao, H., Chen, K., Liu, Z., Zhang, Y., Shao, T., Zhang, H. Coordinated control of PM<sub>2.5</sub> and O<sub>3</sub> is  
547 urgently needed in China after implementation of the “Air pollution prevention and control  
548 action plan”. Chemosphere, 270, 129441,  
549 <https://doi.org/10.1016/j.chemosphere.2020.129441>, 2021.

550