atmosphere

Supplementary Materials: Observation-Based Ozone Formation Rules by Gradient Boosting Decision Trees Model in Typical Chemical Industrial Parks

Nana Cheng [1,2], Deji Jing [3], Zhenyu Gu [1,2], Xingnong Cai [3], Zhanhong Shi [3], Sujing Li [3], Liang Chen [4], Wei Li [3,*] and Qiaoli Wang [5,*]

Table S1. Statistics of VOC species monitored in the park.

Groups	Number	species	Groups	Number	species
Alkanes	v1	Ethane	Aromatics	v36	Benzene
	v2	Propane		v37	Toluene
	v3	n-Butane		v38	Ethyl benzene
	v4	Isobutane		v39	o-Xylene
	v5	Isopentane		v40	Isopropyl benzene
	v6	2,2-Dimethyl butane		v41	n-Propyl benzene
	v7	2,3-Dimethyl butane		v42	1,2,3-Trimethyl benzene
	v8	Cyclopentane		v43	1,2,4-Trimethyl benzene
	v9	2-Methyl pentane		v44	1,3,5-Trimethyl benzene
	v10	n-Hexane		v45	m-Ethyl toluene
	v11	Cyclohexane		v46	p-Ethyl toluene
	v12	Methyl cyclopentane		v47	o-Ethyl toluene
	v13	3-Methyl hexane		v48	m-Diethyl benzene
	v14	2,2,4-Trimethyl pentane		v49	Styrene
	v15	2,3,4-Trimethyl pentane		v50	Methyl chloride
	v16	n-Heptane		v51	Vinyl chloride
	v17	Methyl cyclohexane		v52	Trifluorotrichloroethane
	v18	2-Methyl heptane		v53	Dichloromethane
	v19	3-Methyl heptane		v54	1,1-Dichloroethane
	v20	n-Octane	Halocarbons	v55	cis-1,2-Dichloroethene
	v21	n-Nonane		v56	Chloroform
	v22	n-Decane		v57	1,1,1-Trichloroethane
	v23	n-Undecane		v58	1,2-Dichloroethane
	v24	n-Dodecane		v59	Carbon tetrachloride
	v25	Ethene		v60	Trichloroethylene

Alkenes and Ethyne	v26	Ethyne	v61	1,2-Dichloropropane
	v27	Propene	v62	cis-1,3-Dichloropropene
	v28	1-Butene	v63	trans-1,3-Dichloropropene
	v29	trans-2-Butene	v64	1,1,2-Trichloroethane
	v30	cis-2-Butene	v65	Perchloroethylene
	v31	1,3-Butadiene	v66	1,2-Dibromoethane
	v32	1-Pentene	v67	Monochlorobenzene
	v33	trans-2-Pentene	v68	m-Dichlorobenzene
	v34	cis-2-Pentene	v69	p-Dichlorobenzene
	v35	Isoprene	v70	o-Dichlorobenzene

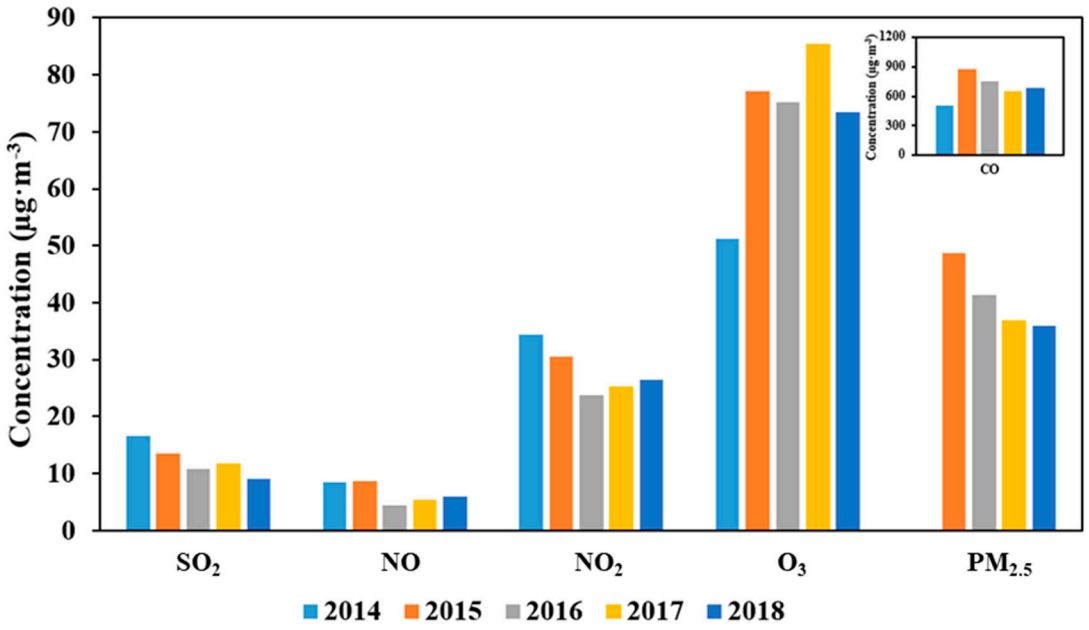

Figure S1. Average annual variations for 6 conventional atmospheric pollutants.

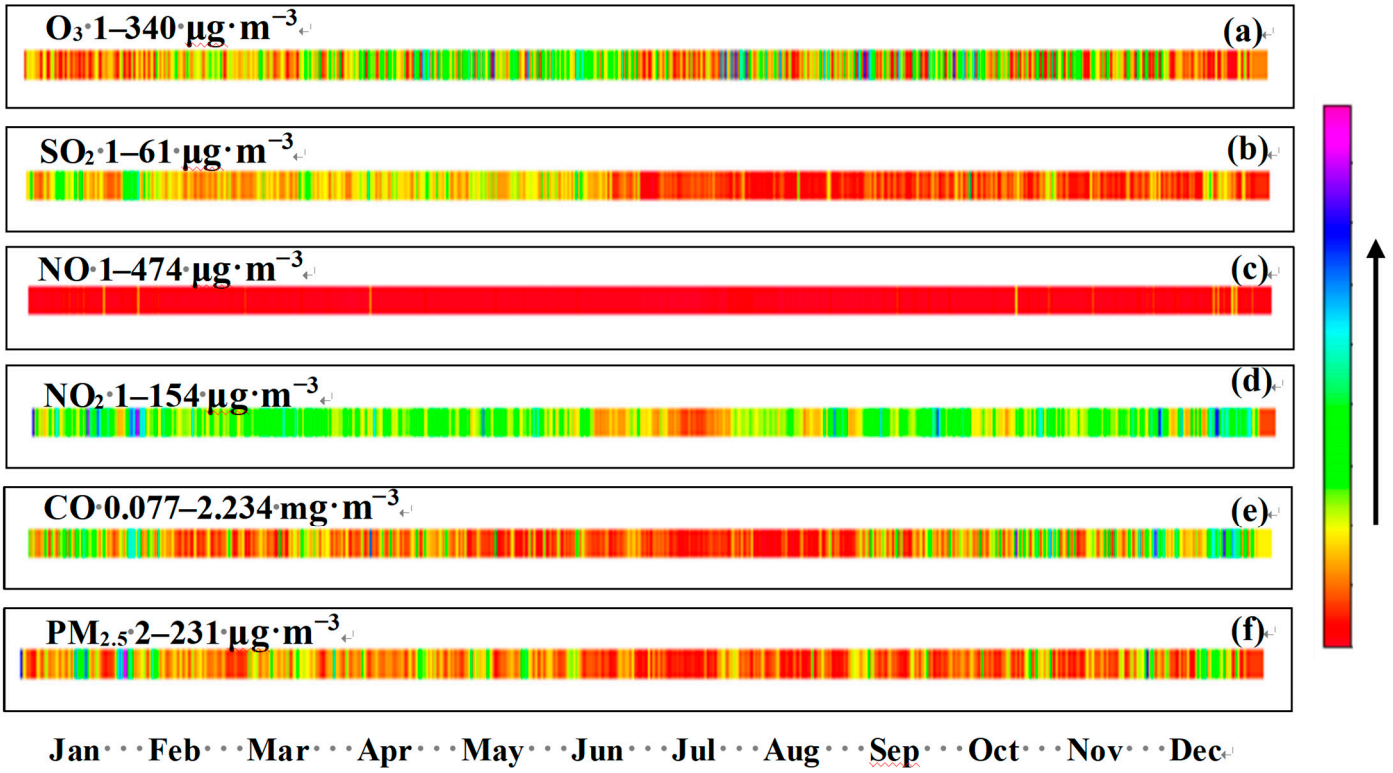

Figure S2. Temporal variations in O_3 (a), SO_2 (b), NO (c), NO_2 (d), CO (e), and $PM_{2.5}$ (f) levels in 2018.

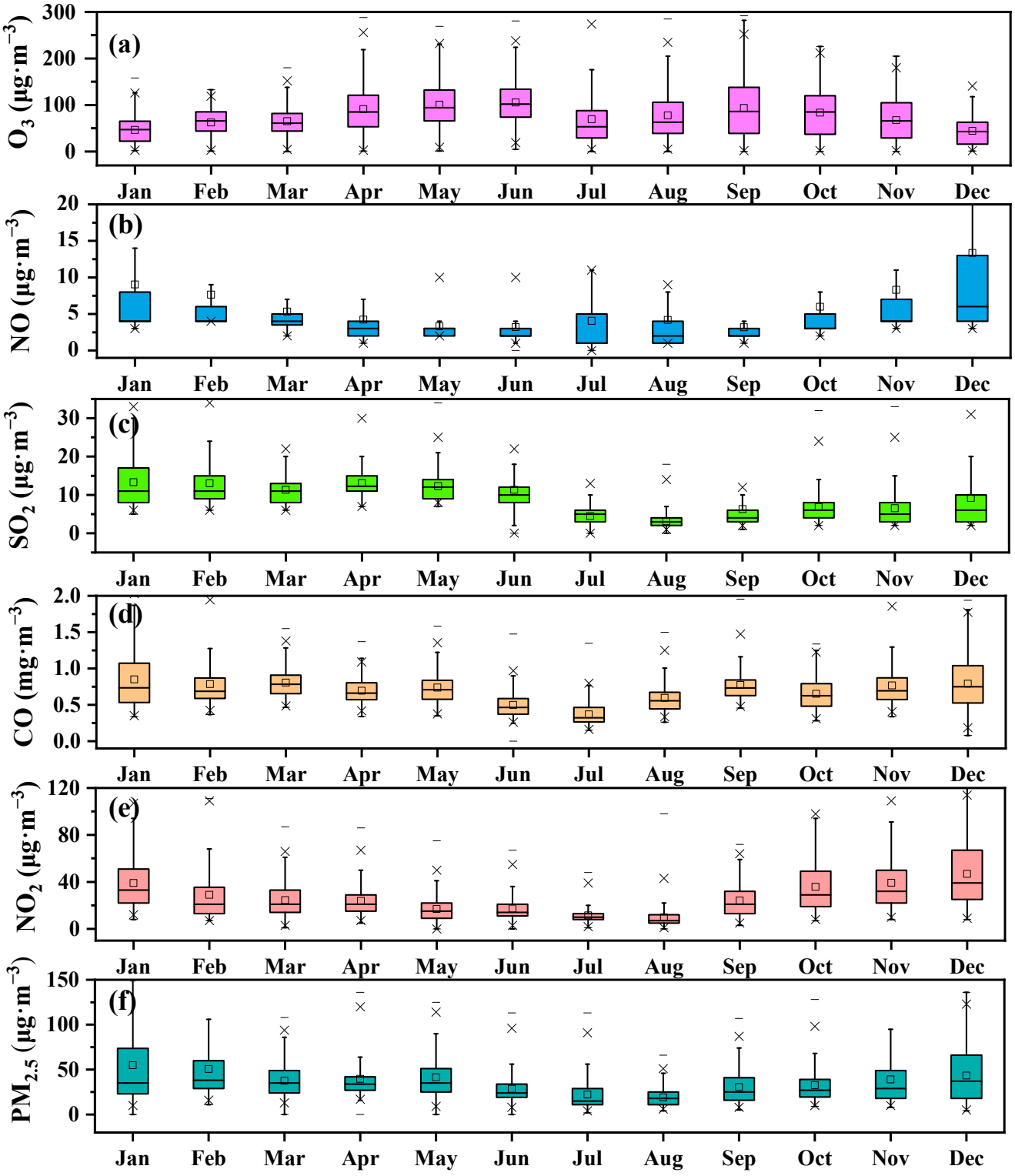

Figure S3. Average monthly variations in O_3 (a), SO_2 (b), NO (c), NO_2 (d), CO (e), and $PM_{2.5}$ (f) levels in 2018.

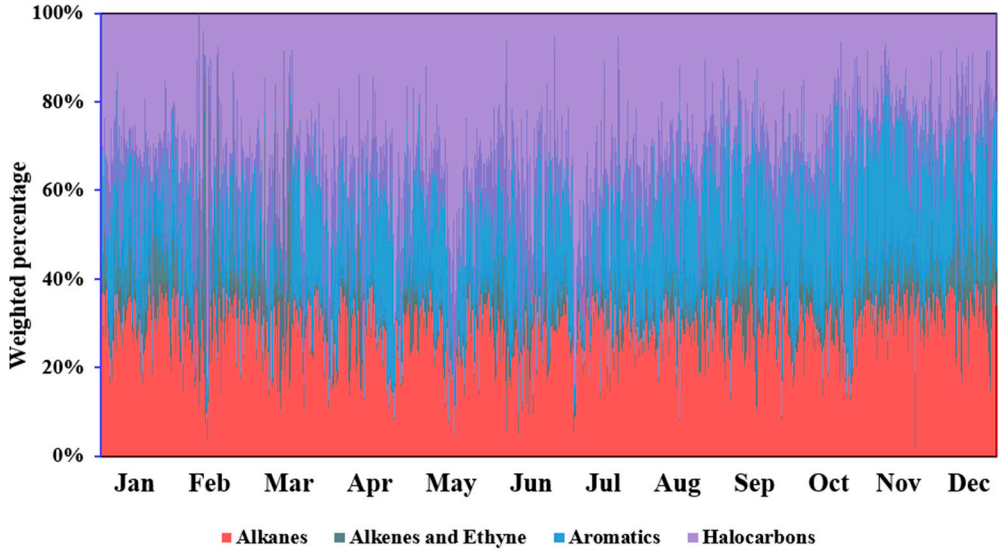

Figure S4. Temporal variations in VOC composition in 2018.

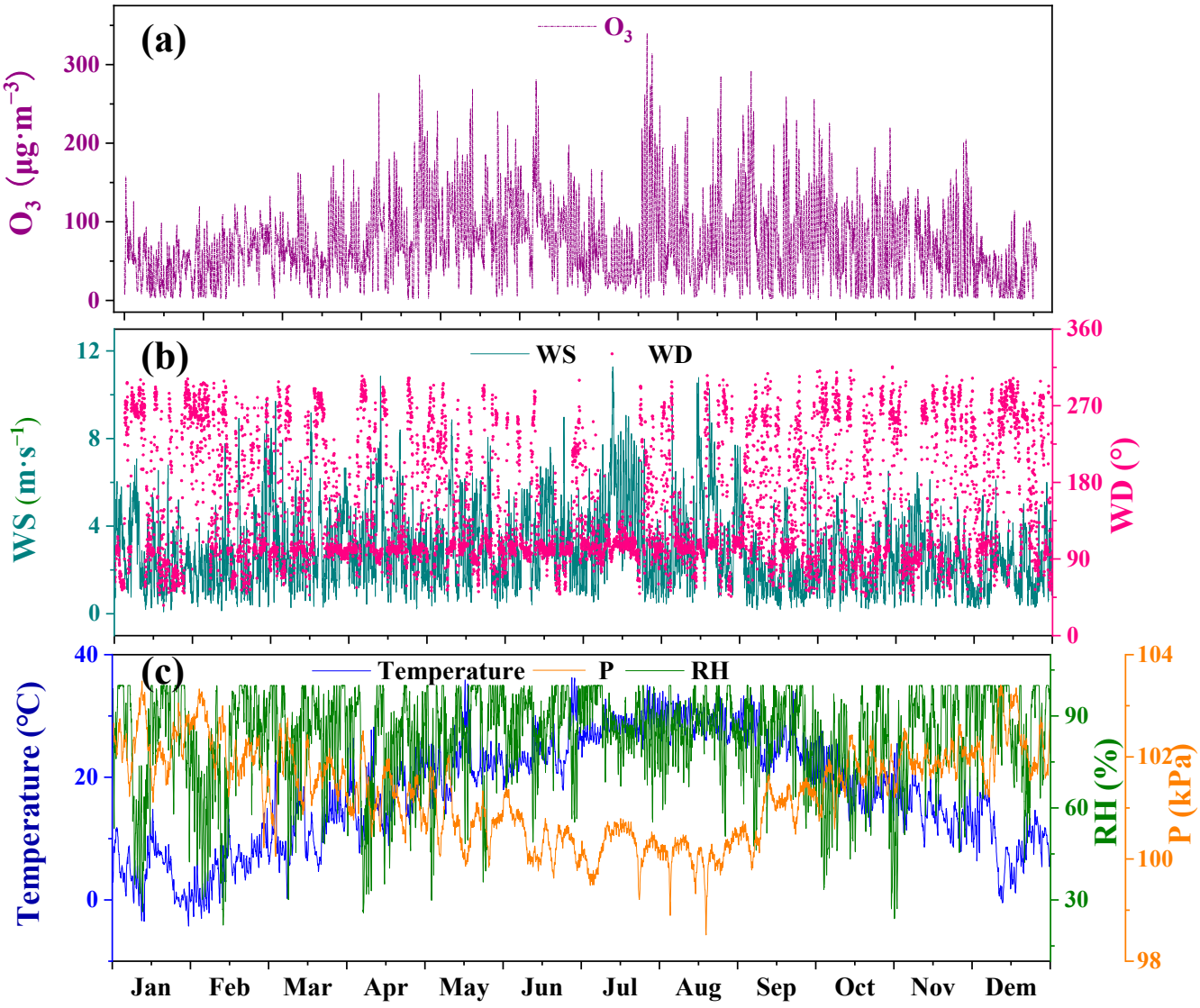

Figure S5. Time series of O_3 (a) and meteorological parameters, including WS, WD (b) and Temperature, RH, and P (c) in 2018.

The Pearson correlation coefficients (PCCs) for the 70 VOC species are shown in **Fig. S5**. The PCCs among 2-methylpentane, n-hexane, and methylcyclopentane were all approximately 0.99. At the same time, the PCCs among the C8-C12 alkanes were approximately 0.8, and even reached approximately 0.9 within isomers. In addition, the PCCs between C8-C12 alkanes and C8-C9 aromatics were strong, with most values reaching 0.8. Additionally, the PCCs of the C8-C9 aromatics were larger than 0.84 and mostly close to 1. However, the correlations of halocarbons were relatively weak. The PCCs for 1,2-dichloromethane, carbon tetrachloride, dimethylbutane and 1,3-dichloropropene were approximately 0.8, and the PCCs were 0.82, 0.98 and 0.80 for 2,2-dimethylbutane and isoprene, cis-2-butene and cyclopentane, benzene and 1,2-dichloroethane, respectively. Then, the VOC species ultimately used for fittings with the GBDT model, which were selected based on the stability of the species, the integrity of the monitoring data and the correlation analysis, were n-butane, n-hexane, n-undecane, ethane, cis-2-pentene, 1,3-butadiene, isoprene, toluene, methyl chloride, dichloromethane, and 1,2-dichloroethane.

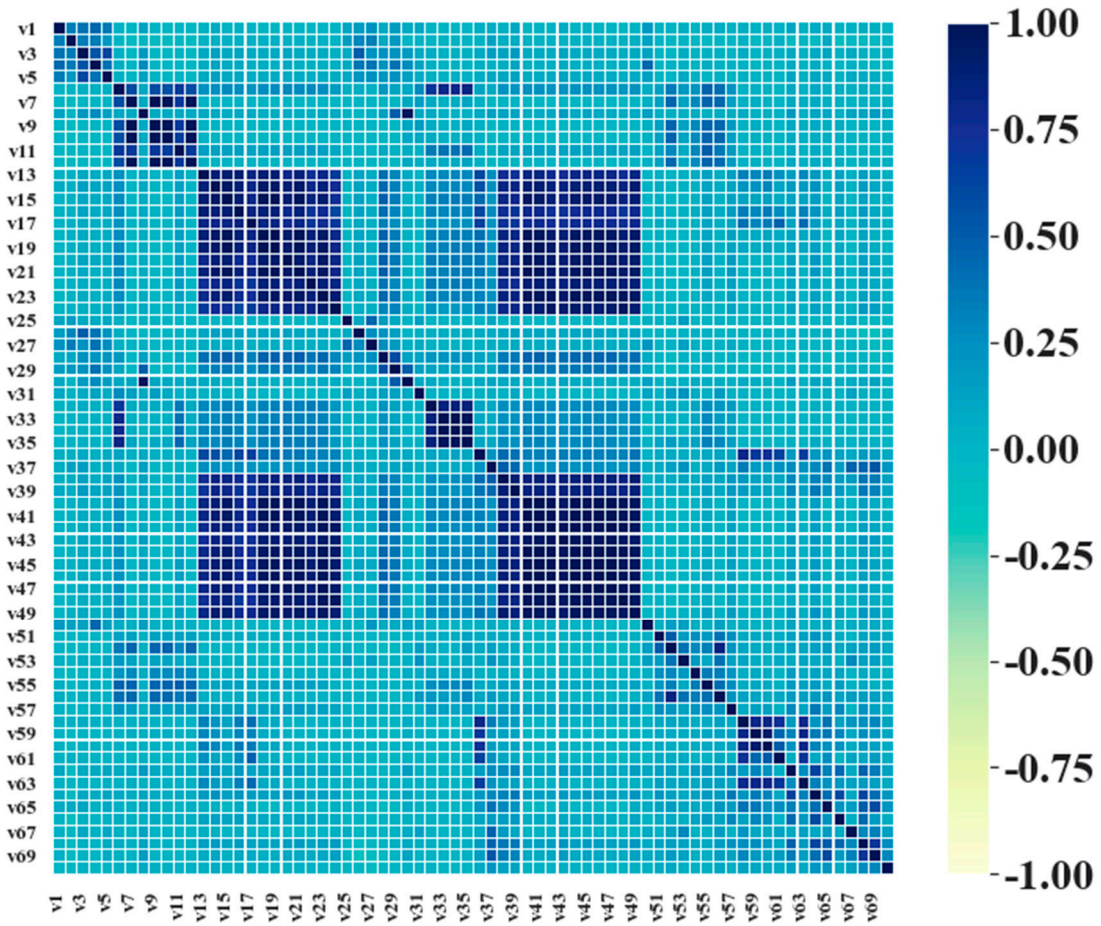

Figure. S6. Heat map for the correlation of 70 VOC species. Note: v1-v70 in this map corresponds to v1-v70 in Table S1.

Table S2. Parameters with the best performance indices under different conditions for LTP and HTP. Note: "/" indicates failed fittings for small datasets after deletion.

Missing values processing		Deletion			Interpolation		
Feature selection process		Without feature selection	Feature selection		Without feature selection	Feature selection	
			Functional groups	Correlation analysis		Functional groups	Correlation analysis
$M_{4\text{-}10}$	Learning rate	/	0.1	0.05	0.05	0.05	0.05
	Max depth		6	5	5	6	5
	n-estimators		2000	2000	2000	2000	2000

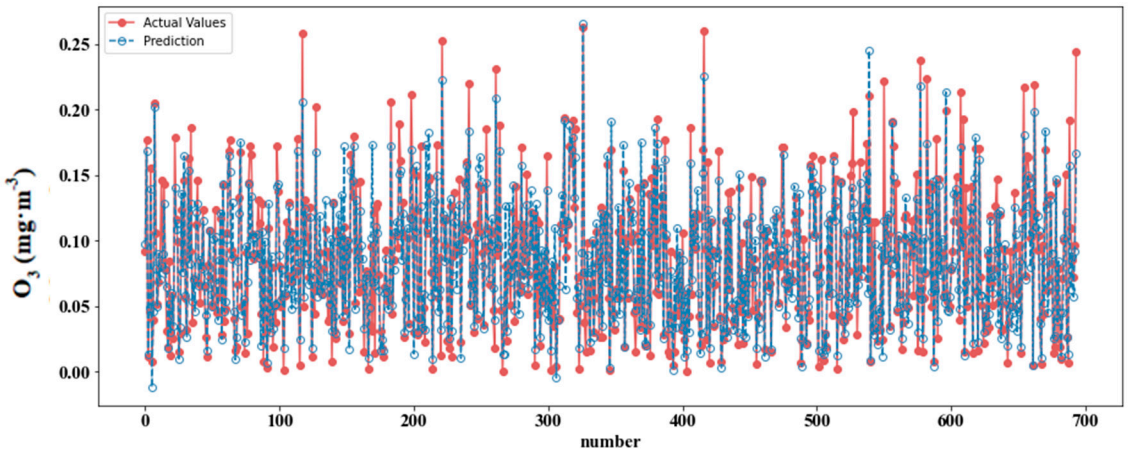

Figure S7. Concentrations of ozone predicted by the GBDT model and the corresponding actual values.