

Supplementary information for
Prediction and cause investigation of ozone based on a
double-stage attention mechanism recurrent neural
network

Summary:

Page 2-3: Table S1. Part of the training set

Page 3-4: Table S2. Feature importance of the data set

Page 4: Figure S1 Importance of each species in VOCS

Page 5: Figure S2 line graph of the performance of the
DA-RNN model on the training set and the prediction set

Page 5: Figure S3 Loss function graph during training

Page 6: Figure S4 WRF-CMAQ forecast performance
graph in July

Page 7: Figure S5 WRF-CMAQ forecast performance
graph in November

Table S1. Part of the training set

Feature name	2020-01-01	2020-01-01	2020-01-01	2020-01-01	2020-01-01
	00	01	02	03	04
NOx	69	59	47	50	50
CO	1.9	1.8	1.8	1.7	1.6
PM10	234	236	237	228	196
PM2.5	214	214	207	200	190
Ethane	7.71	7.61	6.75	6.16	5.93
Propane	18.29	23.57	19.47	19.75	18.15
N-butane	7.68	7.37	7.73	7.76	9.19
Isobutane	3.99	4.07	4.15	3.84	4.54
N-pentane	3.9	3.55	3.55	3.26	2.48
Isopentane	3.87	3.32	3.35	3.06	3.16
N-hexane	25.05	24.55	10.81	3.5	3.46
N-heptane	0.58	0.63	0.67	0.54	0.76
N-octane	2.6	1.94	2.29	1.78	2.55
2,2,4-Trimethyl pentane	0	0	0.05	0	0
2,3,4-Trimethyl pentane	0.51	0.41	0.51	0.41	0.46
Ethylene	6.5	7.32	4.77	4	3.76
Acetylene	2.48	2.19	1.83	1.93	2
Propylene	5.56	7.72	4.04	1.77	1.69
cis-2-Butene	0.3	0.4	0.4	0.38	0.28
trans-2-Butene	0.25	0	0.05	0.3	0.3
1,3-Butadiene	0.53	0.39	0.97	0	0.34
Isoprene	4.29	3.74	3.89	3.77	6.02
benzene	9.17	8.72	8.19	8.06	7.6
Toluene	13.16	9.78	9.82	9.03	14.03
Ethyl benzene	4.74	3.65	3.6	3.08	3.65
ortho-xylene	2.8	1.94	1.9	1.56	1.9
Styrene	0.79	0.65	0.56	0.46	0.51
1,2,3- Trimethylbenzene	0.16	0.32	0.32	0.27	0.27
1,2,4- Trimethylbenzene	0	0	0	0	0
1,3,5- Trimethylbenzene	0.48	0.27	0.21	0.27	0.21
Cumene	0.54	0.43	0.43	0.38	0.38
n-propylbenzene	1.34	0.86	0.75	0.54	0.64
1,1-dichloroethane	30.75	29.51	17.72	11.04	13.47
1,2-Dichloroethane	1.33	1.59	0.57	0.35	0.57

Trichloroethylene	0	0	0	0	0
Perchloroethylene	0	0	0	0.22	0
Chlorobenzene	0	0	0.05	0	0.05
1,2,4-trichlorobenzene	0	0	0	0	0
Wind speed	0.6	0.6	0.7	0.6	0.8
Temperature	5	4.6	4.4	4.1	4
Humidity	94	96	97	97	94
Radiation intensity	0	0	0	0	0
Ozone	26	33	42	38	42

Table S2. Feature importance of the data set

Feature name	Feature importance
NOx	0.083
CO	0.006
PM10	0.047
PM2.5	0.058
Ethane	0.008
Propane	0.006
N-butane	0.004
Isobutane	0.011
N-pentane	0.013
Isopentane	0.008
N-hexane	0.008
N-heptane	0.004
N-octane	0.007
2,2,4-Trimethyl pentane	0.009
2,3,4-Trimethyl pentane	0.002
Ethylene	0.008
Acetylene	0.018
Propylene	0.007
cis-2-Butene	0.011
trans-2-Butene	0.009
1,3-Butadiene	0.008
Isoprene	0.028
benzene	0.014
Toluene	0.006
Ethyl benzene	0.006
ortho-xylene	0.013
Styrene	0.041
1,2,3-Trimethylbenzene	0.011
1,2,4-Trimethylbenzene	0.015
1,3,5-Trimethylbenzene	0.006

Cumene	0.006
n-propylbenzene	0.014
1,1-dichloroethane	0.0065
1,2-Dichloroethane	0.0073
Trichloroethylene	0.0052
Perchloroethylene	0.0278
Chlorobenzene	0.0094
1,2,4-trichlorobenzene	0.0105
Wind speed	0.009
Temperature	0.078
Humidity	0.062
Radiation intensity	0.292

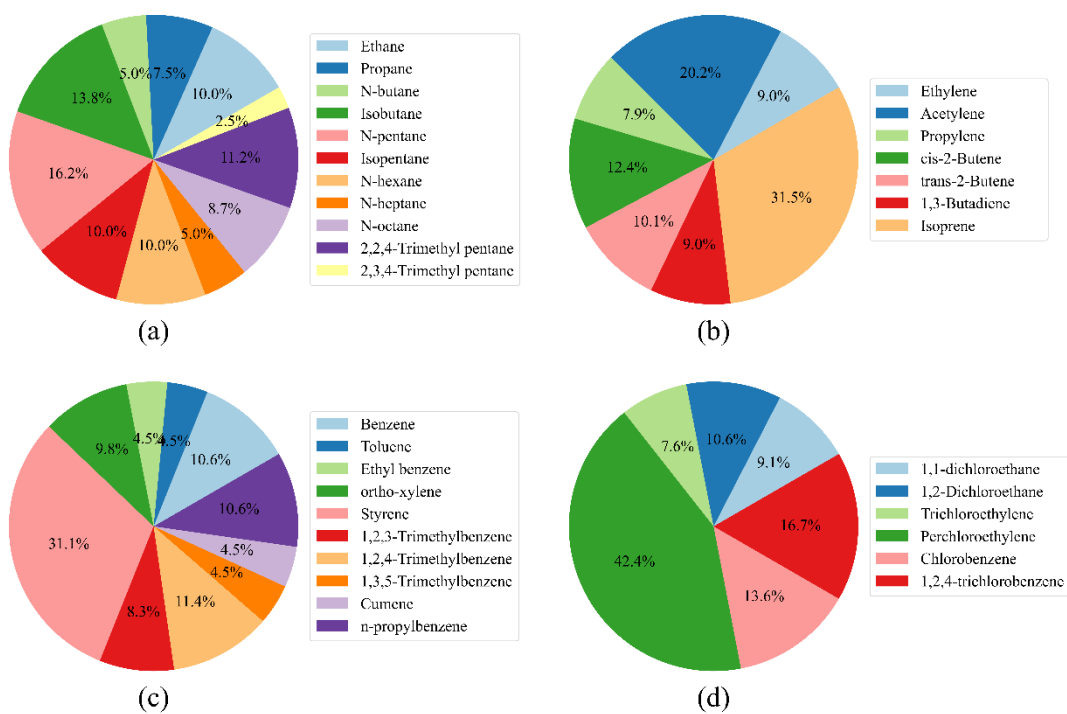


Figure S1 Importance of each species in VOCs

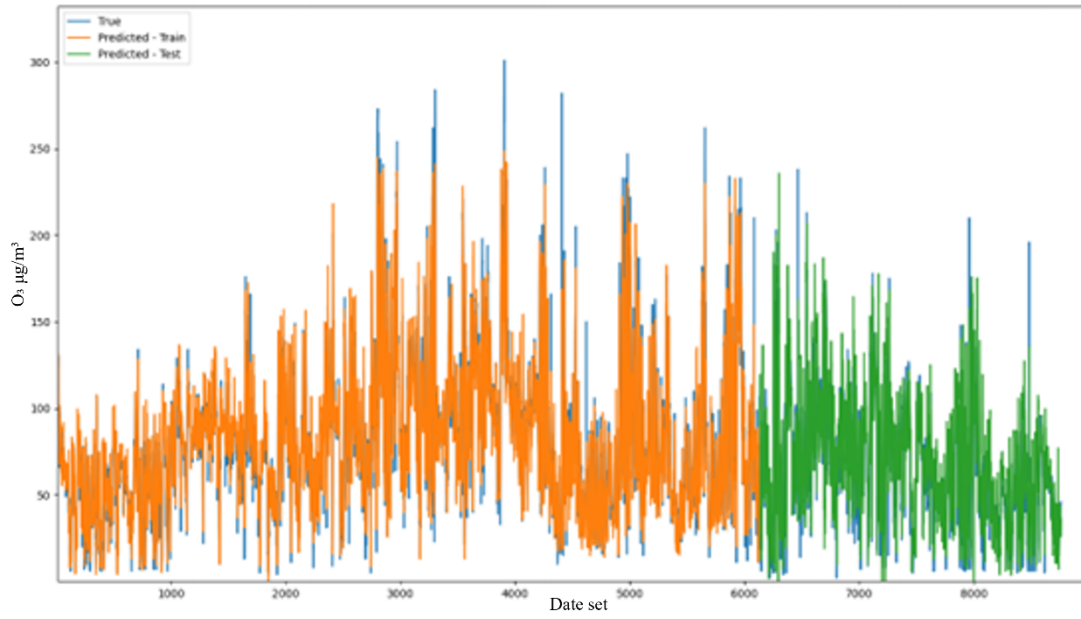


Figure S2 line graph of the performance of the DA-RNN model on the training set and the prediction set

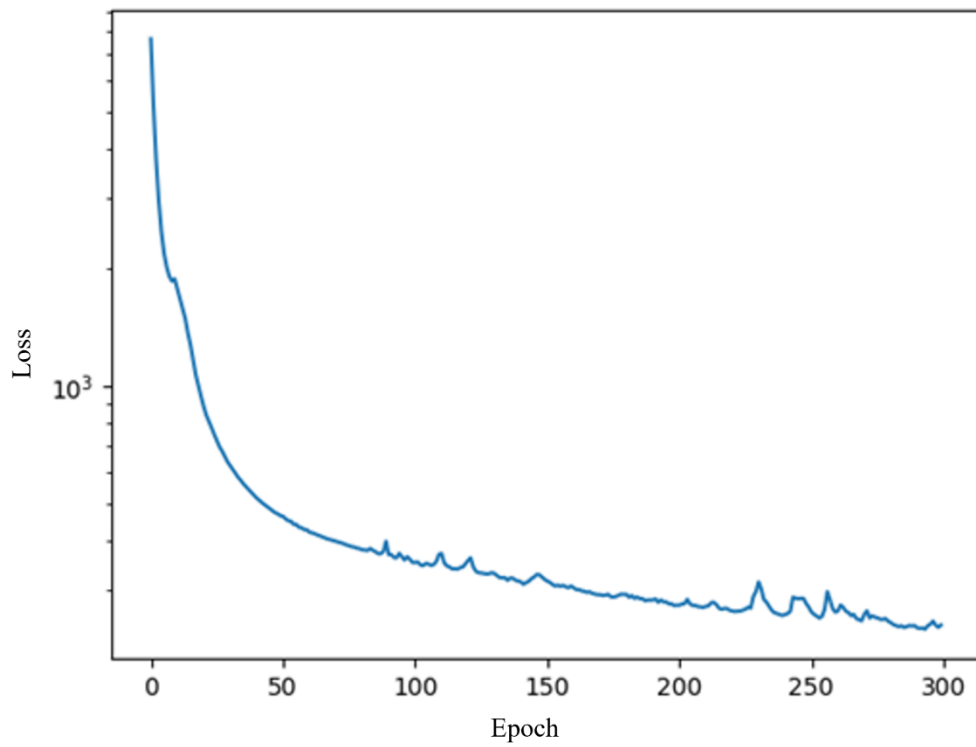


Figure S3 Loss function graph during training

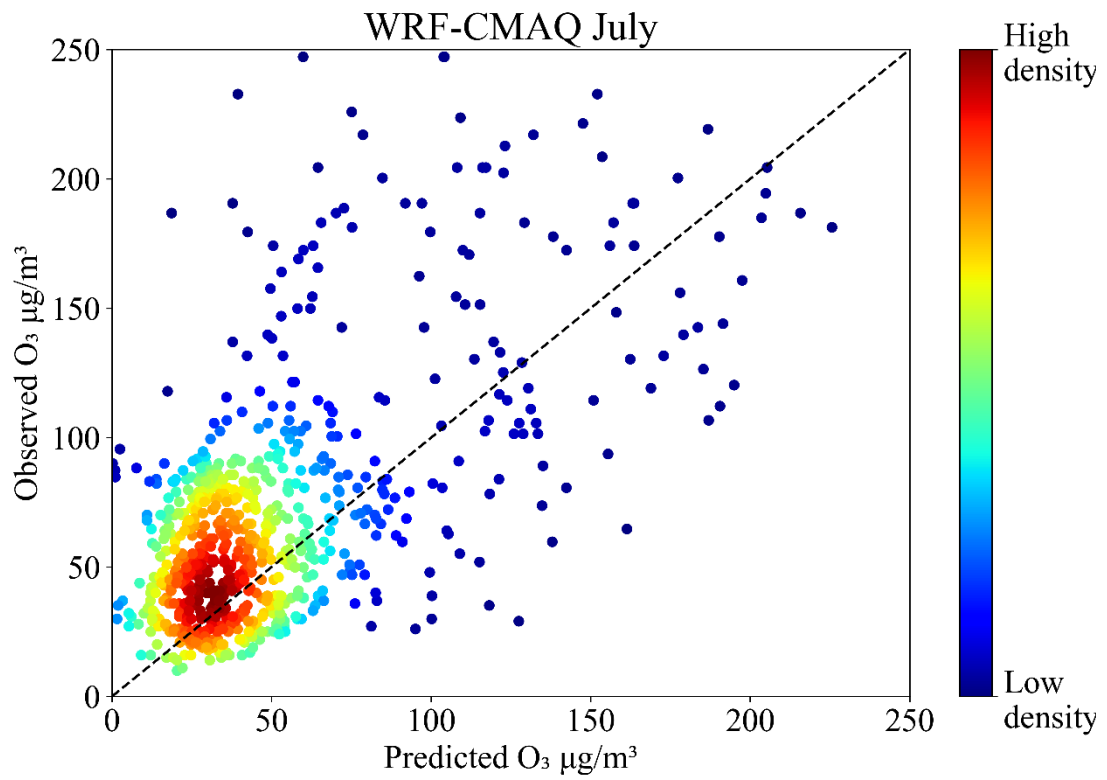


Figure S4 WRF-CMAQ forecast performance graph in July

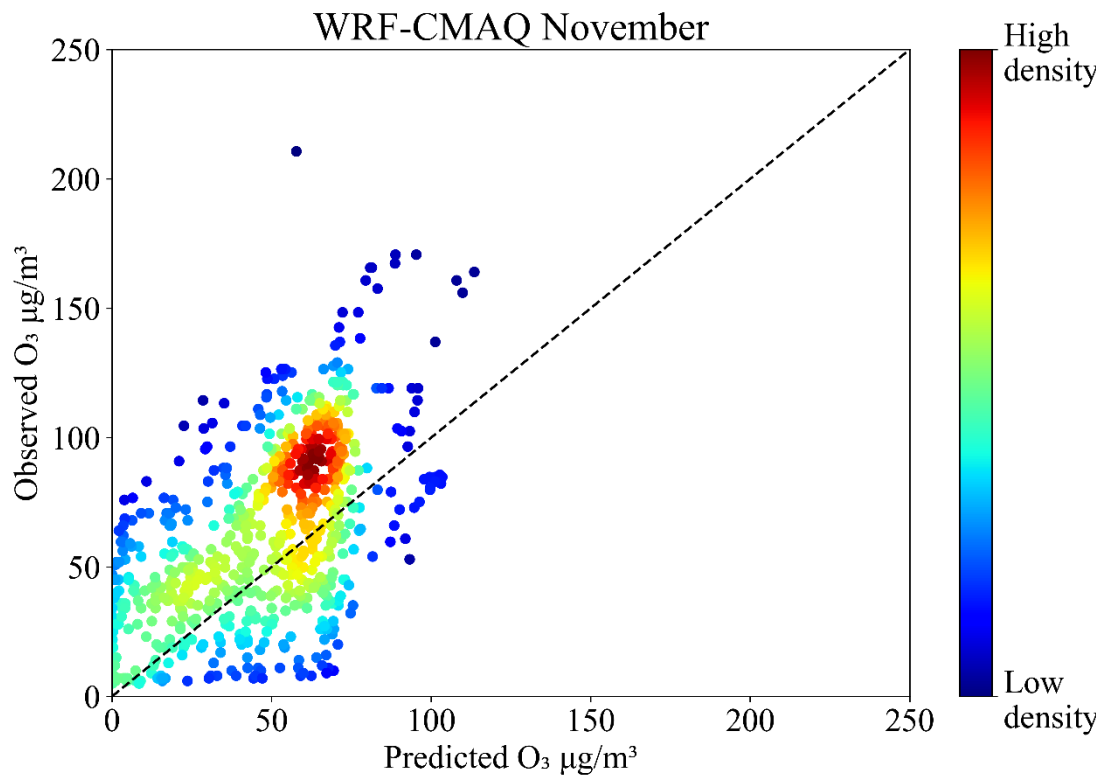


Figure S5 WRF-CMAQ forecast performance graph in November