

---

# **Particle Matters Composition**

# **Content**

**1 Extraction**

**2 Composition**

**3 EDA**

**4 Quality Control**

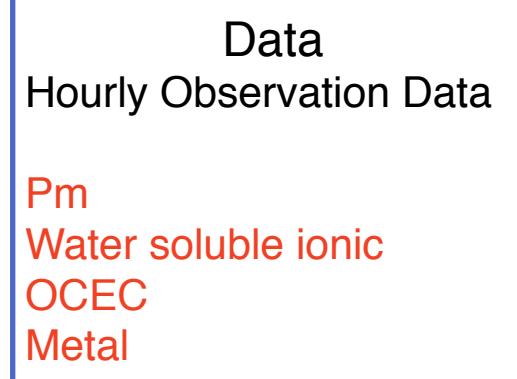
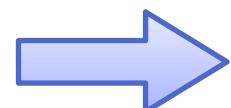
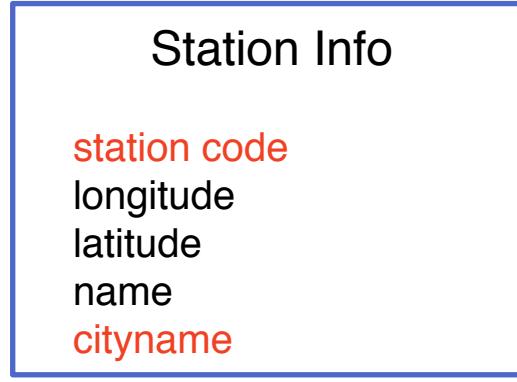
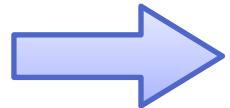
**5 WorkFlow**

# Extraction

- DB — CSV



data\_yyyymm  
parameter  
state  
qccode  
site  
instrument



- Obs\_com\_stations.txt

```
stationcode,longitude,latitude,name,cityname
110000006,116.04110,39.59716,琉璃河,北京
371500001,115.97389,36.49306,周公河站,聊城
150100001,111.75199,40.84149,市站,呼和浩特
1320100001,118.75394,32.05505,江苏省南京市草场门站,南京
410200001,114.33111,34.76111,金明西街站,开封
140100002,112.48248,37.71918,景明南路站,太原
340100001,117.11720,31.47000,环保局,合肥
```

- obs\_com\_yyyymmddhh.txt

# Composition

- Main Composition

- Water Soluble Ionic
  - $\text{Ca}^{2+}$   $\text{Mg}^{2+}$   $\text{K}^+$   $\text{NH}_4^+$   $\text{Na}^+$   $\text{SO}_4^{2-}$   $\text{NO}_3^-$   $\text{Cl}^-$
- OC
- EC
- Metal
  - Hg Br As Si Se Te V Ca Ti Ba Sc Pd Co Mo
  - K Fe Pb TI Cu Cr Cs Ag Zn Sb Sn Mn Cd Ni Ga

## • 2.2 Water Soluble Ionic

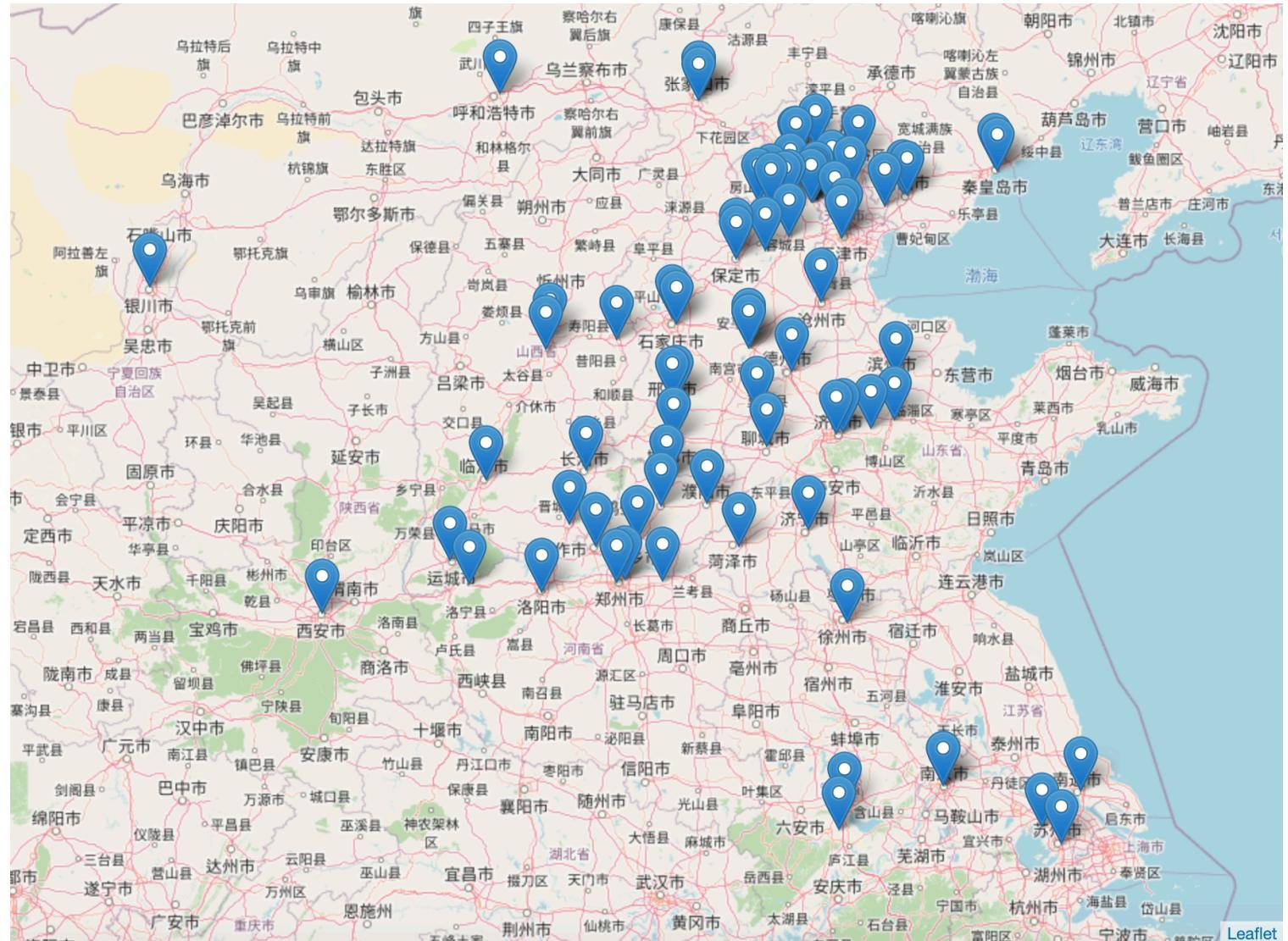
- Toal / PM25 30~40%
- CE(Cation Equivalent)
- (Anion Equivalent)
- SNA = SO<sub>4</sub><sup>2-</sup> + NO<sub>3</sub><sup>-</sup> + NH<sub>4</sub><sup>+</sup>
- (SO<sub>4</sub><sup>2-</sup>\*2/96 + NO<sub>3</sub><sup>-</sup>/62) / (NH<sub>4</sub><sup>+</sup>/18)

$$CE = \frac{Na^+}{23} + \frac{NH_4^+}{18} + \frac{K^+}{39} + \frac{Mg^{2+}}{12} + \frac{Ca^{2+}}{20} \quad (1)$$

$$AE = \frac{SO_4^{2-}}{48} + \frac{NO_3^-}{62} + \frac{Cl^-}{35.5} + \frac{F^-}{19} + \frac{NO_2^-}{46} \quad (2)$$

EDA

# • Monitor



## • Sample

time	stationcode	longitude	latitude	PM10	PM25	Cl-	NO3-	SO42-	Ca2+	Na+	K+	NH4+	Mg2+	OC	EC	Hg	Br	As	
2019-01-01:00	110000012	116.426	40.048	97.92	66.830	-999.000	-999.000	-999.000	-999.000	-999.000	-999.000	-999.000	-999.000	1.525	-999.000	-999.0	-999.000	-999.000	
2019-01-01:01	110000012	116.426	40.048	95.92	59.340	4.777	7.389	0.101	0.190	0.495	0.770	7.297	0.034	8.569	2.035	-999.0	24.363	2.447	
2019-01-01:02	110000012	116.426	40.048	78.62	51.570	3.188	7.599	0.075	0.177	0.422	0.583	6.282	0.032	6.164	1.269	-999.0	24.815	5.972	
2019-01-01	110000012	116.426	40.048	95.27	44.566	3.130	7.335	2.560	0.167	0.410	0.505	6.277	0.033	6.397	1.177	-999.0	21.289	6.292	
C	EC	Hg	Br	As	Si	Se	Te	V	Ca	Ti	Ba	Sc	Pd	Co	Mo	K	Fe	Pb	TI
5	-999.000	-999.0	-999.000	-999.000	-999.000	-999.000	-999.0	-999.0	-999.000	-999.000	-999.000	-999.000	-999.0	-999.000	-999.000	-999.000	-999.000	-999.000	
9	2.035	-999.0	24.363	2.447	582.994	9.484	-999.0	-999.0	532.644	24.715	66.903	1.146	-999.0	10.852	84.933	826.137	756.589	81.246	4.575
↓	1.269	-999.0	24.815	5.972	769.016	11.844	-999.0	-999.0	771.753	30.422	93.461	2.479	-999.0	3.585	26.758	918.143	838.080	86.893	-999.000
Ba	Sc	Pd	Co	Mo	K	Fe	Pb	TI	Cu	Cr	Cs	Ag	Zn	Sb	Sn	Mn	Cd	Ni	Ga
.000	-999.000	-999.0	-999.000	-999.000	-999.000	-999.000	-999.000	-999.000	-999.000	-999.000	-999.000	-999.0	-999.000	-999.0	-999.000	-999.000	NaN	NaN	NaN
.903	1.146	-999.0	10.852	84.933	826.137	756.589	81.246	4.575	135.414	15.796	11.278	-999.0	243.554	-999.0	-999.0	85.677	-999.0	15.792	10.030
.461	2.479	-999.0	3.585	26.758	918.143	838.080	86.893	-999.000	47.587	12.914	7.485	-999.0	244.243	-999.0	-999.0	60.340	-999.0	-999.000	9.985

- Split Merge

- obs ionic

```
obs_ionic.sample(3)
```

	time	PM10	PM25	Cl-	N03-	S042-	Ca2+	Na+	K+	NH4+	Mg2+
5044	2019-07-30 04:00:00	19.62	18.637	0.152	7.116	3.348	0.040	0.302	0.146	4.921	0.03
5300	2019-08-09 20:00:00	7.28	13.110	0.180	4.460	3.130	0.022	0.302	0.046	3.010	0.01
3666	2019-06-02 18:00:00	58.22	55.550	0.110	1.860	3.680	0.080	0.080	NaN	5.580	NaN

- obs ocec

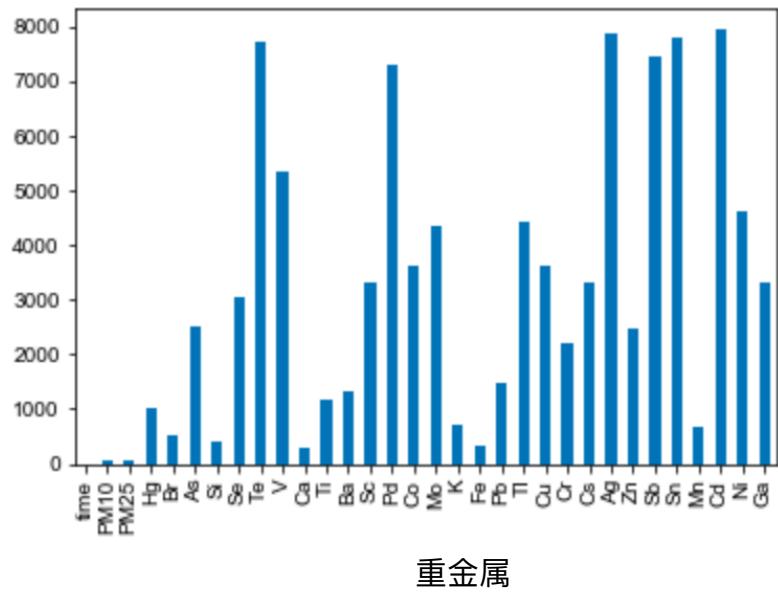
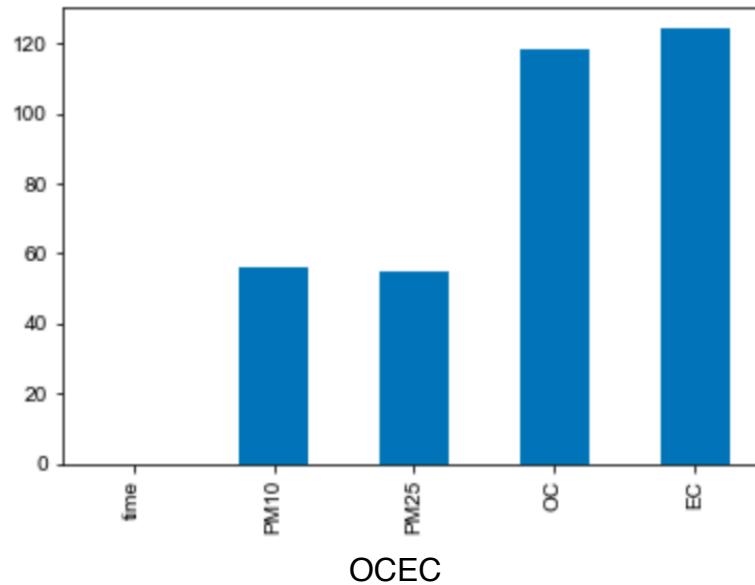
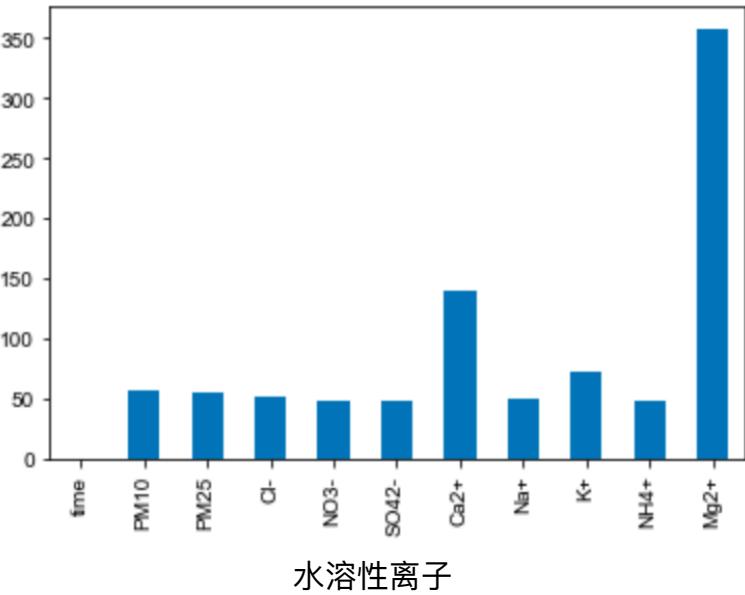
```
obs_ocec.sample(3)
```

		time	PM10	PM25	OC	EC
4778	2019-07-19	02:00:00	37.74	58.34	2.460	1.219
4667	2019-07-14	11:00:00	6.02	12.37	2.154	0.525
6810	2019-10-11	18:00:00	20.23	24.56	2.542	0.504

- obs metal

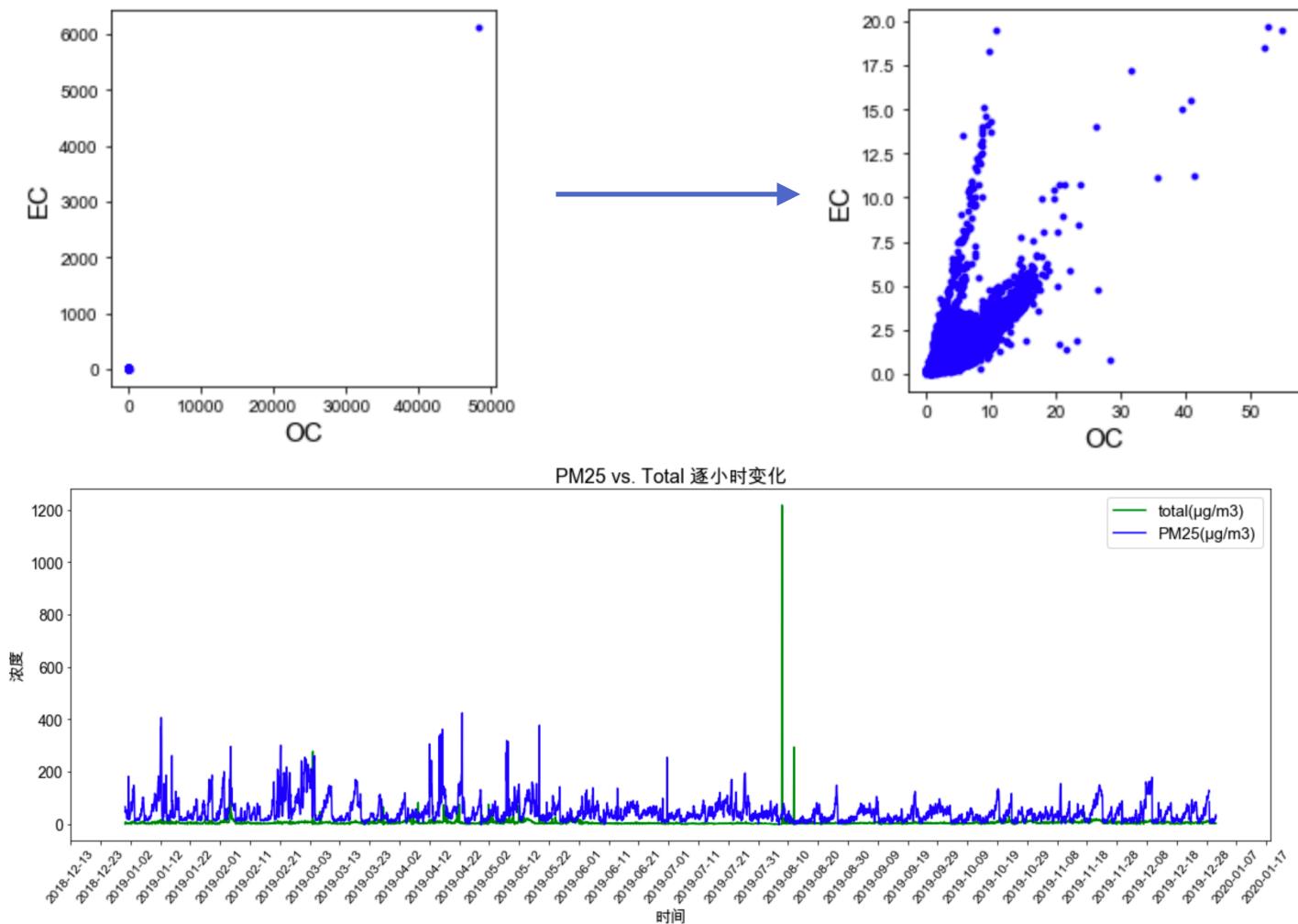
```
obs_metal.sample(3)
```

- Missing



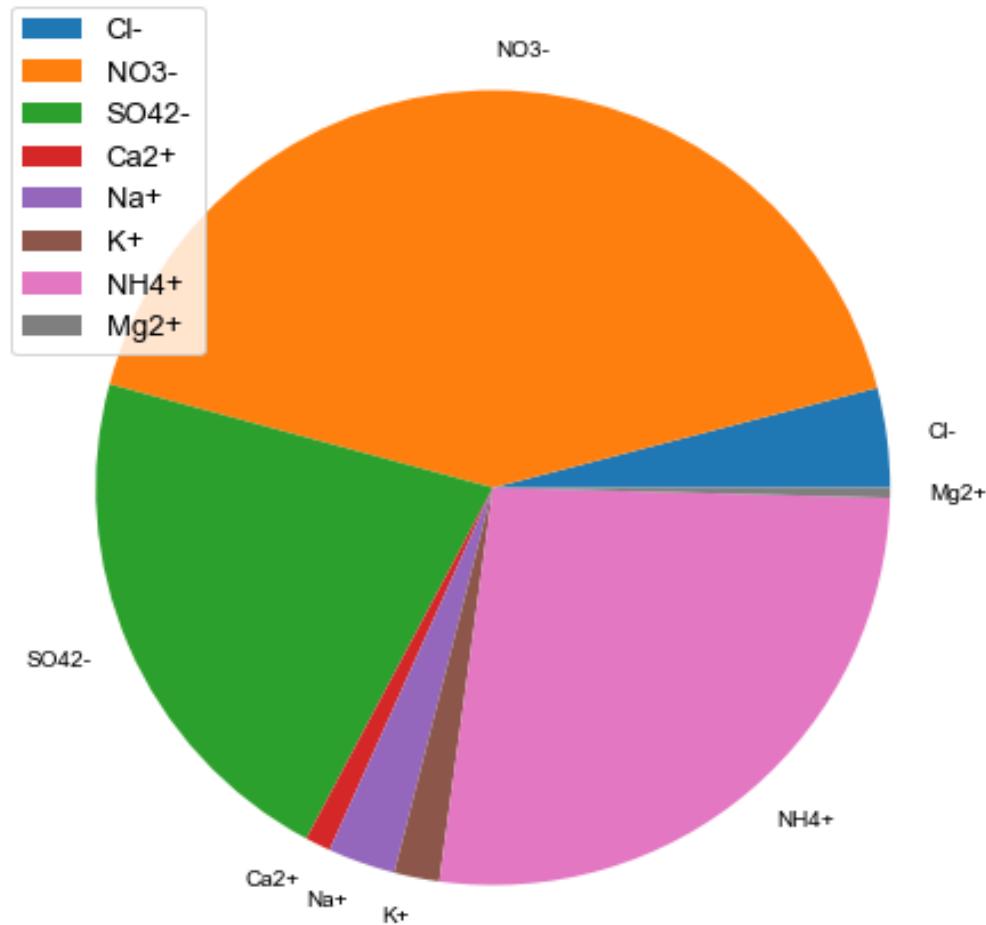
- Error

obs_metal.sample(3)				
	time	PM10	PM25	Hg
3445	2019-05-24 13:00:00	81.18	71.87	0.00200
7902	2019-11-26 06:00:00	54.66	29.96	0.00013
1257	2019-02-22 09:00:00	106.49	118.56	0.00231

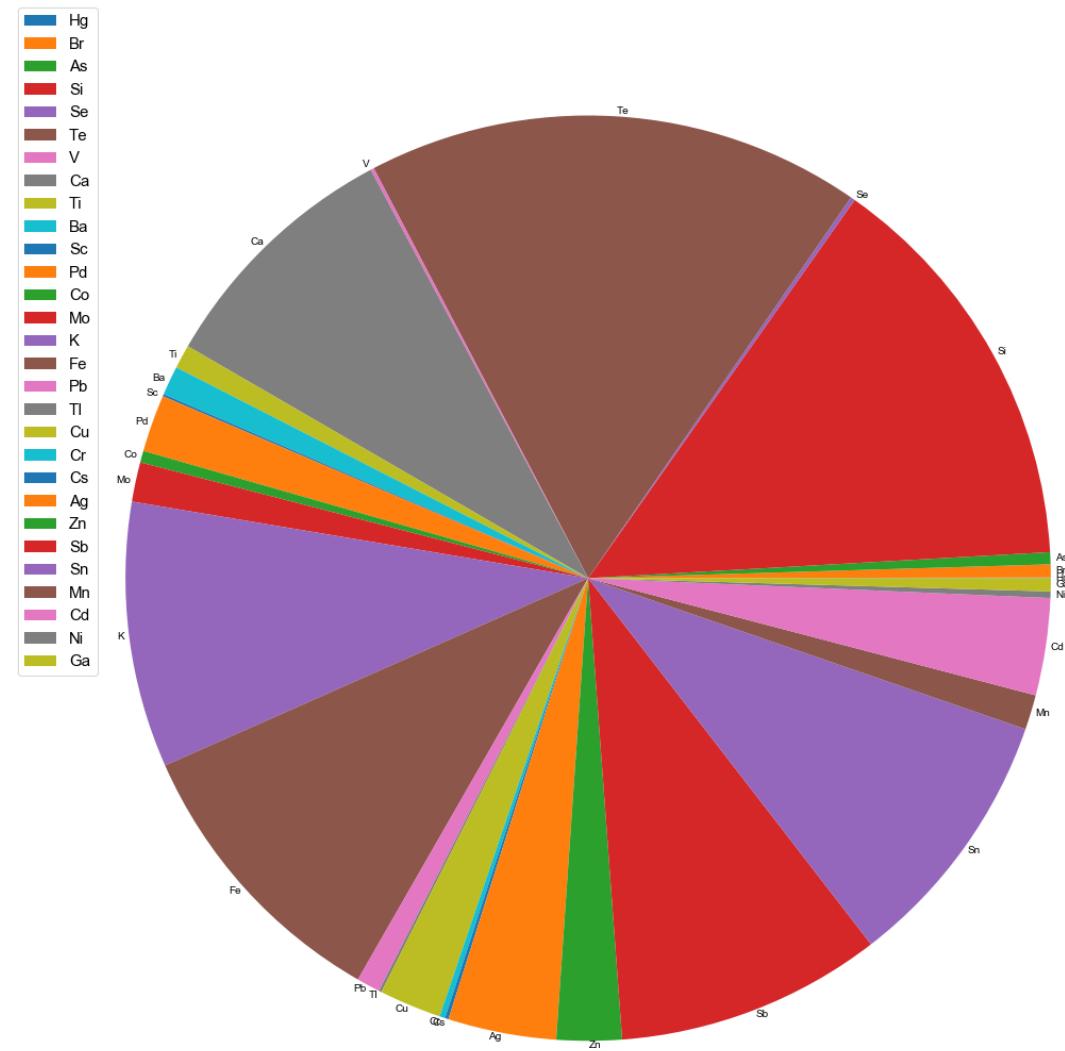


## • Distribution

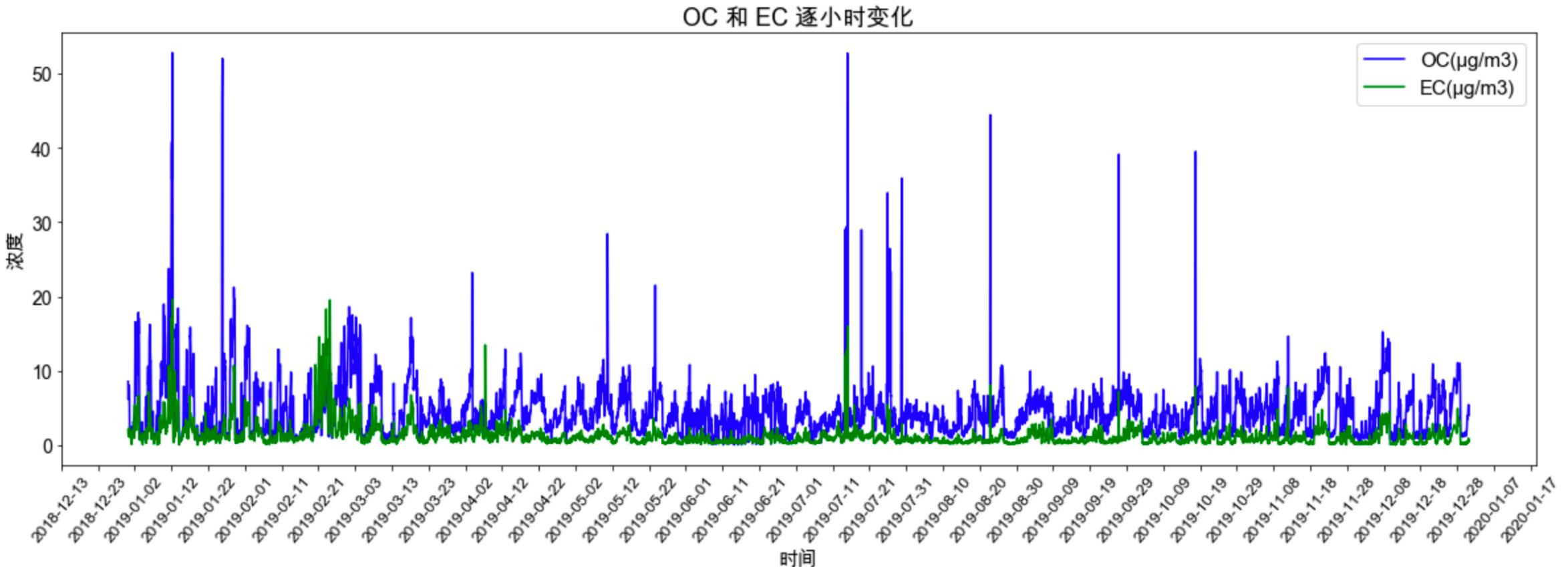
水溶性离子 均值 分布



在线重金属 均值 分布



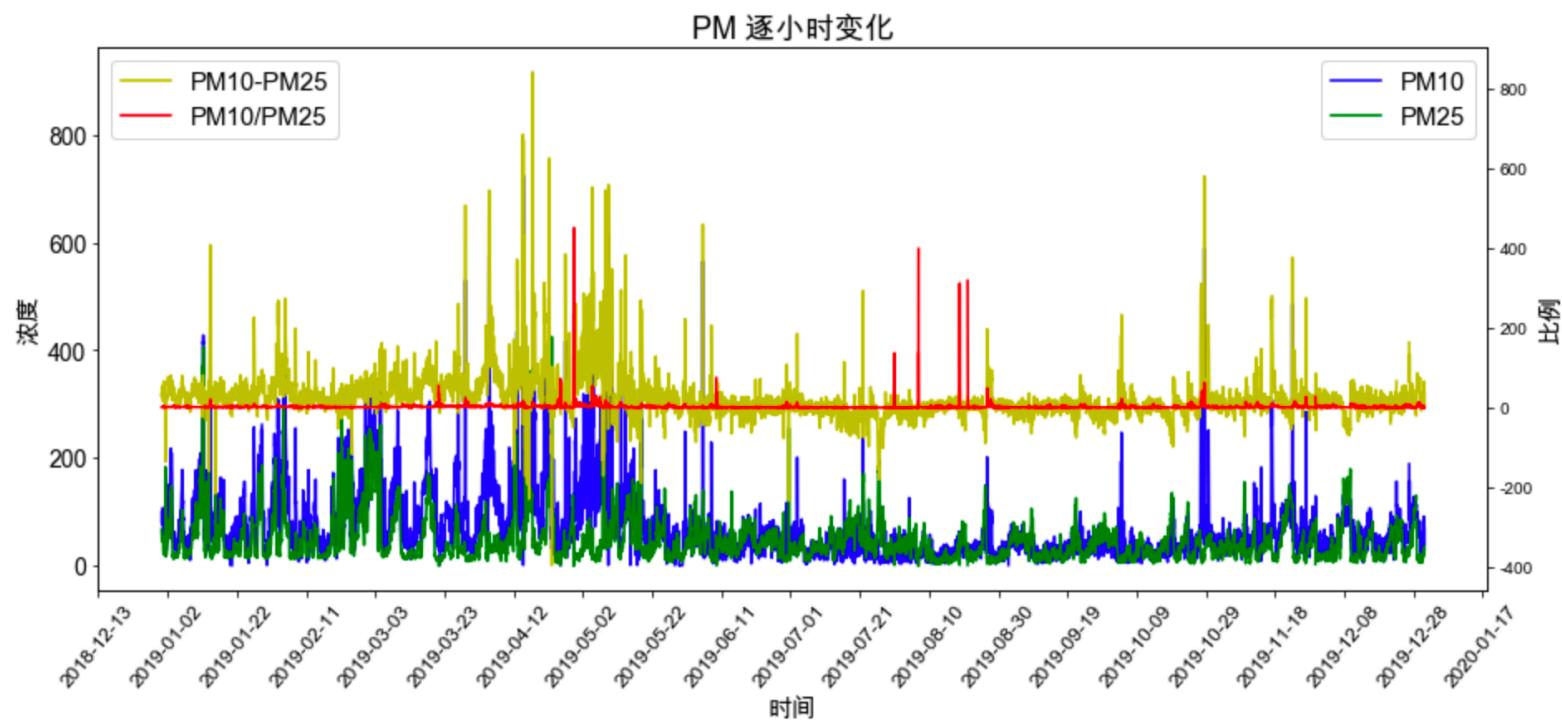
- OCEC



# 4 Quality Control

- PM

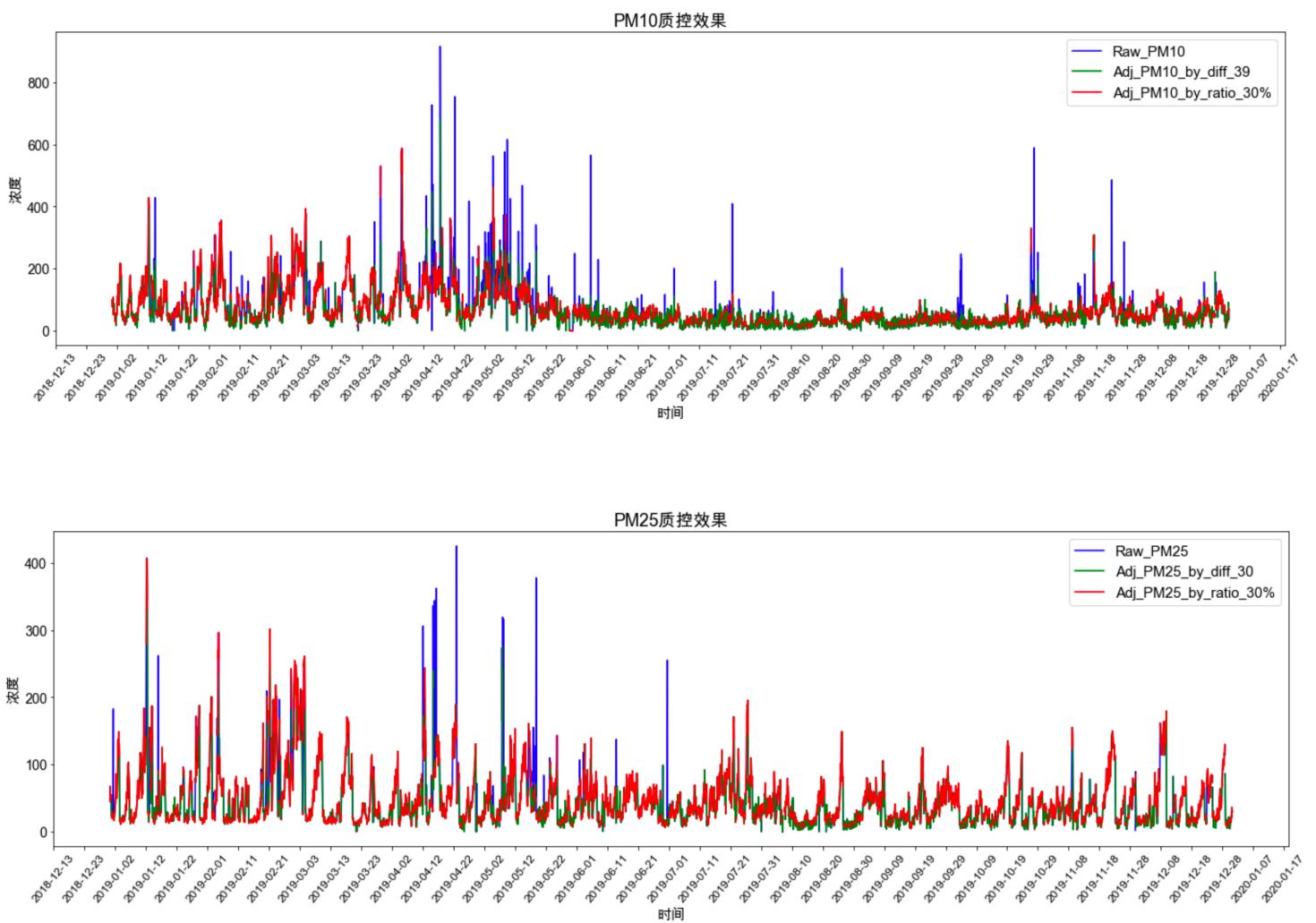
- Abs diff
- 95% index



	count	mean	std	min	2%	5%	10%	20%	30%	40%	50%	80%	90%	95%	97%	99%	max
PM10	8672.0	16.448745	30.429946	0.0	0.32	0.8155	1.7000	3.63	5.483	7.65	10.01	21.13	31.663	47.0980	63.78310	130.6029	727.506
PM25	8673.0	7.049660	14.172167	0.0	0.14	0.3400	0.6792	1.36	2.170	3.06	4.04	9.11	13.914	20.9248	27.86008	52.0056	378.550

- ratio diff method

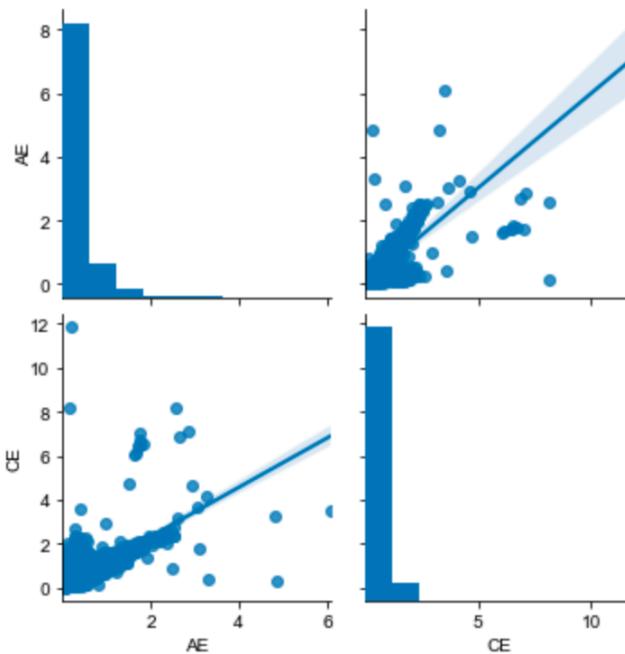
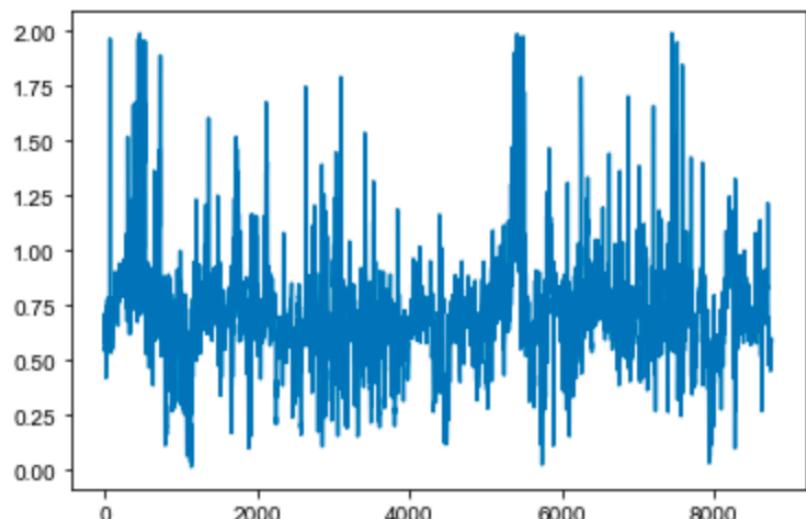
	PM10	PM25	PM10/PM25	PM10-PM25	diff_PM10	diff_PM25	PM10_ratio	PM25_ratio
count	8704.00000	8705.00000	8693.00000	8693.00000	8672.00000	8673.00000	8672.00000	8673.00000
mean	68.93929	45.263637	2.379388	23.660967	-0.006229	0.00260	2.370677	0.398026
std	65.32186	40.918336	9.450552	55.465864	34.591530	15.82889	90.059293	8.358597
min	0.00700	0.039000	0.000412	-394.880000	-727.506000	-378.55000	0.000000	0.000000
2%	7.32060	6.783200	0.257582	-53.415600	-55.297200	-24.94344	0.006951	0.004261
5%	12.56600	9.456400	0.437419	-33.292000	-32.848000	-14.11400	0.017322	0.010794
10%	18.19200	12.340000	0.591460	-19.146000	-21.349000	-9.06800	0.033678	0.021059
50%	47.61500	32.940000	1.451844	14.540000	-0.135000	0.21200	0.202093	0.122297
80%	100.52200	65.660000	2.644588	45.188000	12.800000	5.45600	0.443169	0.278226
90%	149.09000	91.640000	3.940394	72.298000	21.010000	9.12000	0.647608	0.421960
95%	193.97300	128.688000	5.933142	104.096000	30.794500	13.58420	0.921192	0.595552
97%	229.33240	147.350800	7.931307	139.047200	39.529600	17.73360	1.232938	0.754387
99%	310.06930	196.977600	12.836278	240.203200	83.855400	33.83400	2.487479	1.263521
max	916.60000	425.020000	450.576923	839.790000	724.316000	305.51000	6995.250000	596.179487



## • Calc index

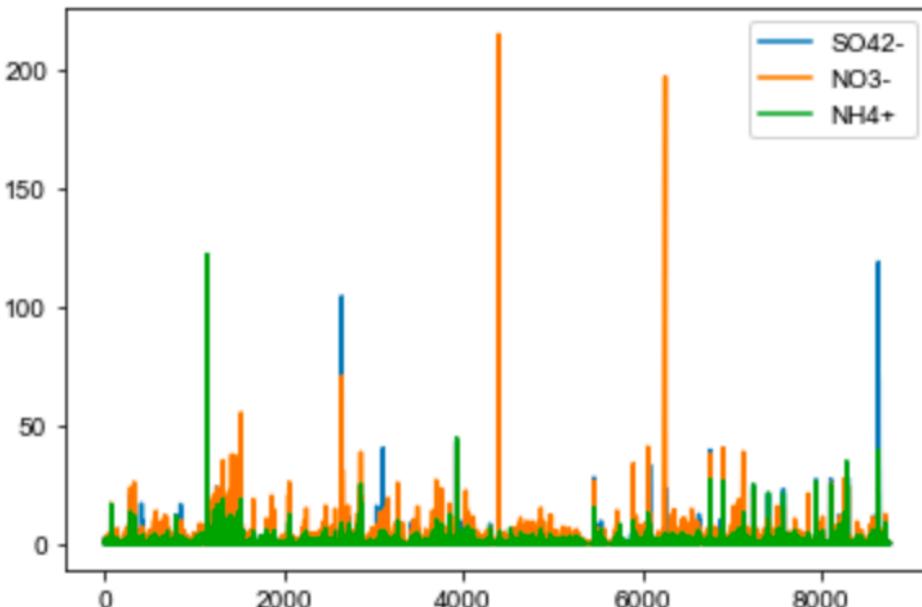
	time	PM10	PM25	Cl-	N03-	SO42-	Ca2+	Na+	K+	NH4+	Mg2+	ionic	SNA	others	SO42-N03-_mol	NH4+_mol	AE	CE	AE/CE	SNA_ratio
6909	2019-10-15 21:00:00	25.24	22.670	0.152	7.376	1.694	0.060	0.336	0.130	4.956	0.050	14.754	14.026	0.728	0.154259	0.275333	0.158541	0.300442	0.527693	0.560264
3555	2019-05-29 03:00:00	65.49	20.570	0.320	1.860	0.850	0.326	0.252	0.519	0.710	0.030	4.867	3.420	1.447	0.047708	0.039444	0.056722	0.082509	0.687472	1.209507
7782	2019-11-21 06:00:00	64.01	69.260	3.887	15.451	3.174	0.413	0.660	0.759	11.822	0.070	36.236	30.447	5.789	0.315335	0.656778	0.424828	0.731418	0.580827	0.480124
1842	2019-03-18 18:00:00	203.42	122.320	1.765	46.114	7.504	0.370	0.658	0.811	19.924	0.037	77.183	73.542	3.641	0.900108	1.106889	0.949826	1.177876	0.806389	0.813187
7561	2019-11-12 01:00:00	43.66	27.770	0.708	6.652	1.607	0.164	0.413	0.277	3.564	0.070	13.455	11.823	1.632	0.140769	0.198000	0.160713	0.237092	0.677850	0.710957
5090	2019-08-01 02:00:00	29.00	30.320	0.100	4.690	3.769	0.098	0.291	0.090	2.460	0.060	11.558	10.919	0.639	0.154166	0.136667	0.156983	0.161527	0.971871	1.128044
4601	2019-07-11 17:00:00	18.92	27.380	0.110	3.850	4.234	0.031	0.465	0.096	3.460	0.020	12.266	11.544	0.722	0.150305	0.192222	0.153404	0.218118	0.703307	0.781934
1861	2019-03-19 13:00:00	294.29	162.150	3.075	57.767	11.727	0.400	0.709	1.043	25.014	0.049	99.784	94.508	5.276	1.176038	1.389667	1.262658	1.471320	0.858181	0.846274
6339	2019-09-22 03:00:00	27.92	86.330	0.540	41.300	12.680	0.040	0.483	0.670	17.540	0.050	73.303	71.520	1.783	0.930296	0.974444	0.945507	1.018791	0.928068	0.954693
5735	2019-08-27 23:00:00	38.05	12.570	0.070	3.400	0.581	0.050	0.224	0.040	1.590	0.040	5.995	5.571	0.424	0.066943	0.088333	0.068915	0.104931	0.656759	0.757844
5632	2019-08-23 16:00:00	16.44	13.230	0.120	0.600	0.491	0.050	0.235	0.231	0.880	0.050	2.657	1.971	0.686	0.019907	0.048889	0.023287	0.071696	0.324800	0.407180
3148	2019-05-12 04:00:00	92.12	106.690	0.650	34.527	10.066	0.117	0.524	0.294	17.234	0.080	63.492	61.827	1.665	0.766595	0.957444	0.784905	1.000282	0.784684	0.800668
3298	2019-05-18 10:00:00	28.30	47.470	0.861	9.770	6.177	0.146	0.437	0.106	8.900	0.040	26.437	24.847	1.590	0.286268	0.494444	0.310522	0.526796	0.589454	0.578969
1173	2019-02-18 21:00:00	55.71	35.261	1.573	6.398	1.739	0.065	0.267	0.271	4.323	0.070	14.706	12.460	2.246	0.139423	0.240167	0.183733	0.267807	0.686062	0.580525
8344	2019-12-14 16:00:00	29.93	16.550	0.254	2.220	1.108	0.031	0.270	0.081	1.575	0.040	5.579	4.903	0.676	0.058890	0.087500	0.066045	0.106199	0.621894	0.673026

- AE CE balance



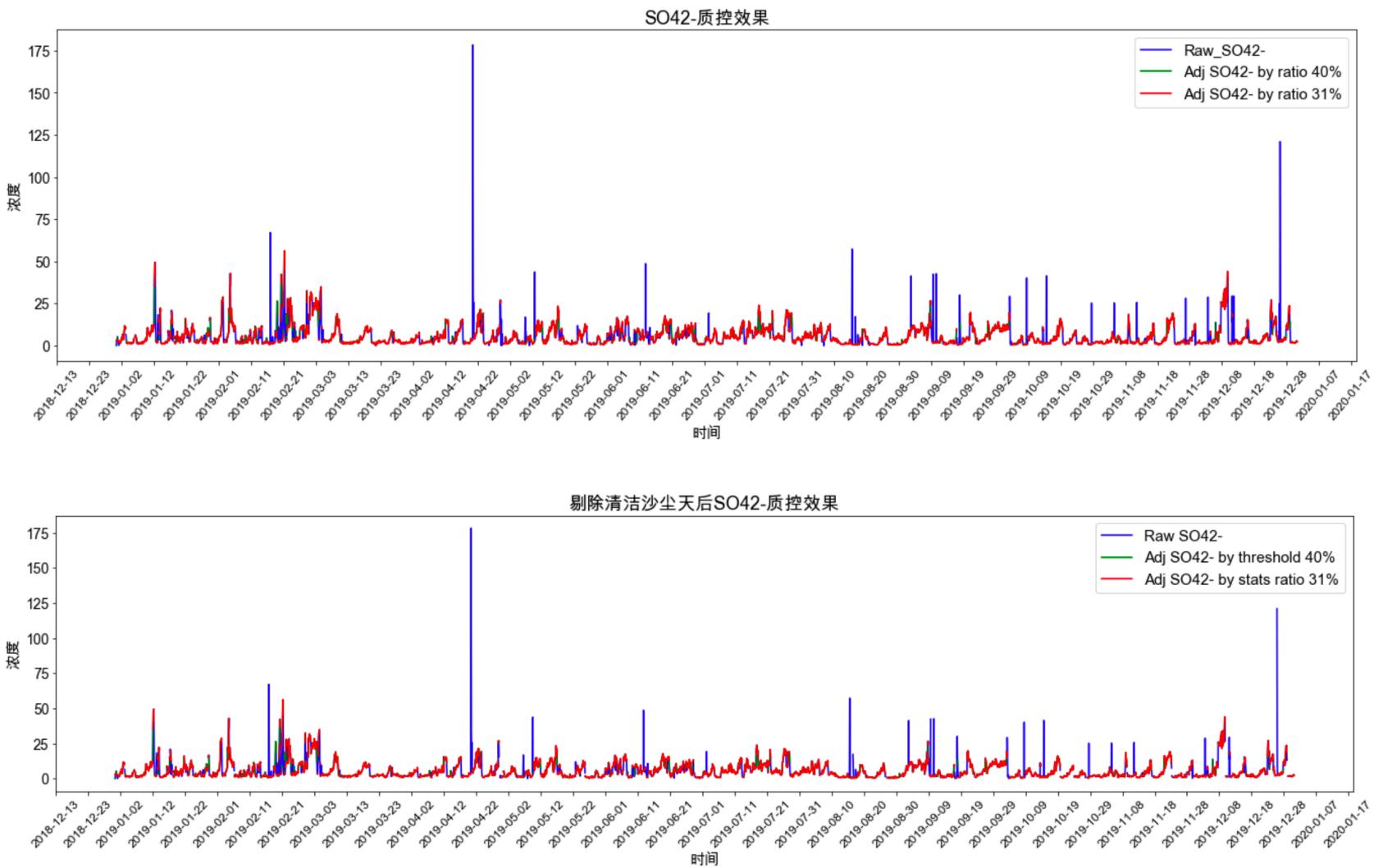
	AE	CE	AE/CE
count	8259.000000	8259.000000	8259.000000
mean	0.330146	0.464282	0.718196
std	0.360282	0.494923	0.315860
min	0.010207	0.022433	0.014987
10%	0.057251	0.087098	0.449303
20%	0.082613	0.128099	0.569265
30%	0.112350	0.181583	0.628339
40%	0.154750	0.251620	0.673229
50%	0.207685	0.332369	0.708250
80%	0.509084	0.701241	0.828492
90%	0.735450	0.984162	0.930055
95%	1.050163	1.330122	1.105334
max	6.073102	11.873694	14.129988

- SNA balance

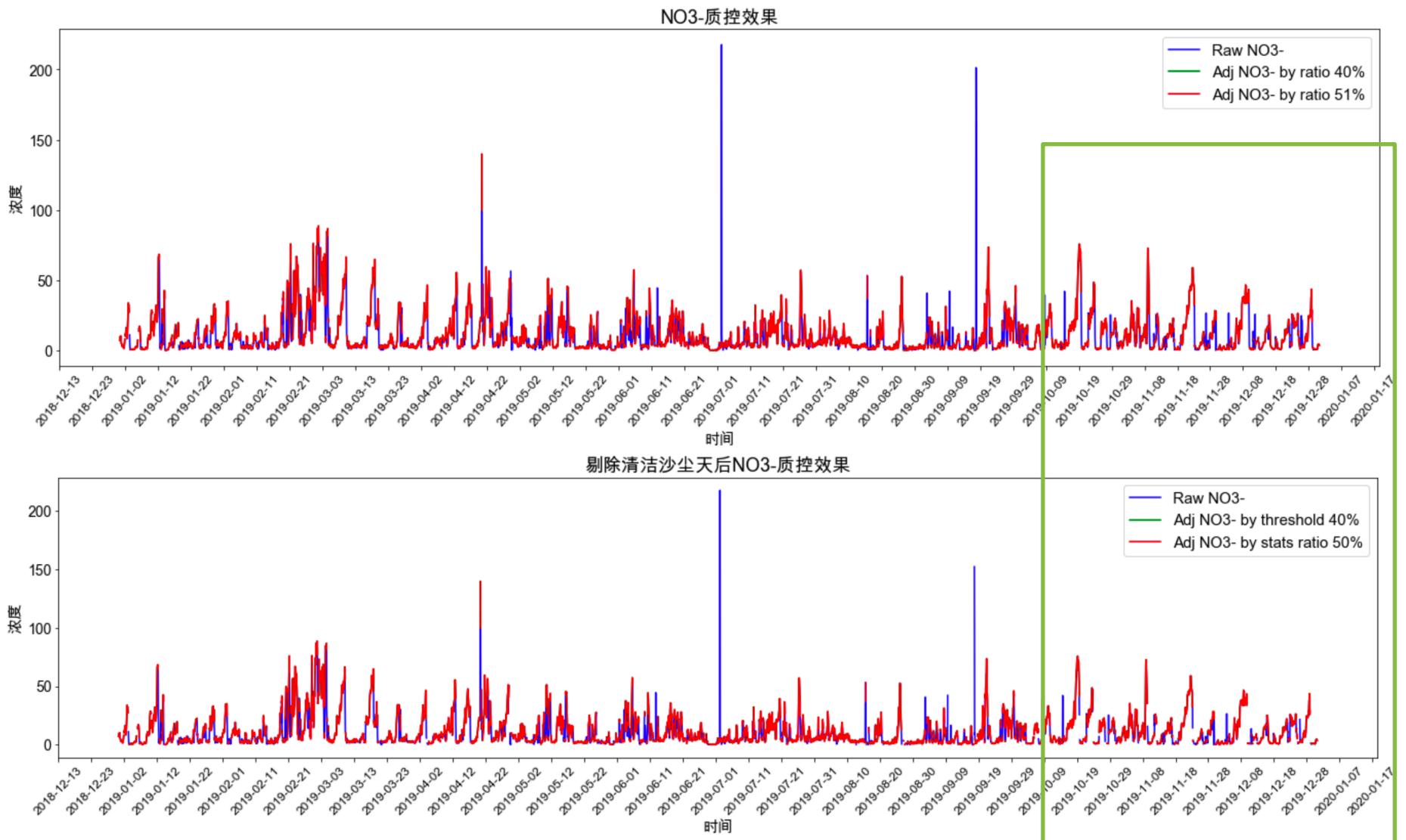


	SO42-	NO3-	NH4+	diff_SO42-	diff_NO3-	diff_NH4+	SO42_-_ratio	NO3_-_ratio	NH4+_ratio
count	8712.000000	8712.000000	8712.000000	8682.000000	8681.000000	8681.000000	8682.000000	8681.000000	8681.000000
mean	5.570272	10.868177	6.960854	0.000723	0.006042	0.003023	0.290591	0.261027	0.678715
std	6.499598	13.433143	6.830384	3.478198	5.522560	2.821363	3.781662	0.936071	20.114600
min	0.020000	0.056000	0.017000	-85.640000	-206.413000	-122.162000	0.000000	0.000000	0.000000
25%	1.697000	2.487500	1.980000	-0.241750	-0.640000	-0.340000	0.031377	0.048443	0.033708
28%	1.848160	2.690000	2.260000	-0.190000	-0.500000	-0.270000	0.035899	0.055557	0.038098
30%	1.950000	2.850000	2.470000	-0.159000	-0.420000	-0.226000	0.038760	0.060811	0.041522
40%	2.555000	3.910000	3.526400	-0.050000	-0.126000	-0.080000	0.055392	0.087629	0.057923
50%	3.430000	5.551500	4.859500	0.010000	0.030000	-0.003000	0.075978	0.122126	0.080214
80%	8.746800	17.836200	11.128400	0.402000	1.108000	0.550000	0.193080	0.323810	0.199776
90%	12.155500	27.009000	15.325900	0.873900	2.243000	1.140000	0.312487	0.518420	0.313696
95%	16.034700	37.629350	20.763000	1.602950	3.858000	2.009000	0.475848	0.749788	0.471016
97%	20.642840	45.913600	24.218040	2.396810	5.512800	2.753600	0.679177	0.954895	0.626307
99%	28.619590	63.102920	31.748010	5.679220	10.256800	5.485600	1.306976	1.826050	1.089343
max	178.330000	217.690000	123.410000	118.892000	211.705000	75.790000	188.882353	39.088843	1501.882353

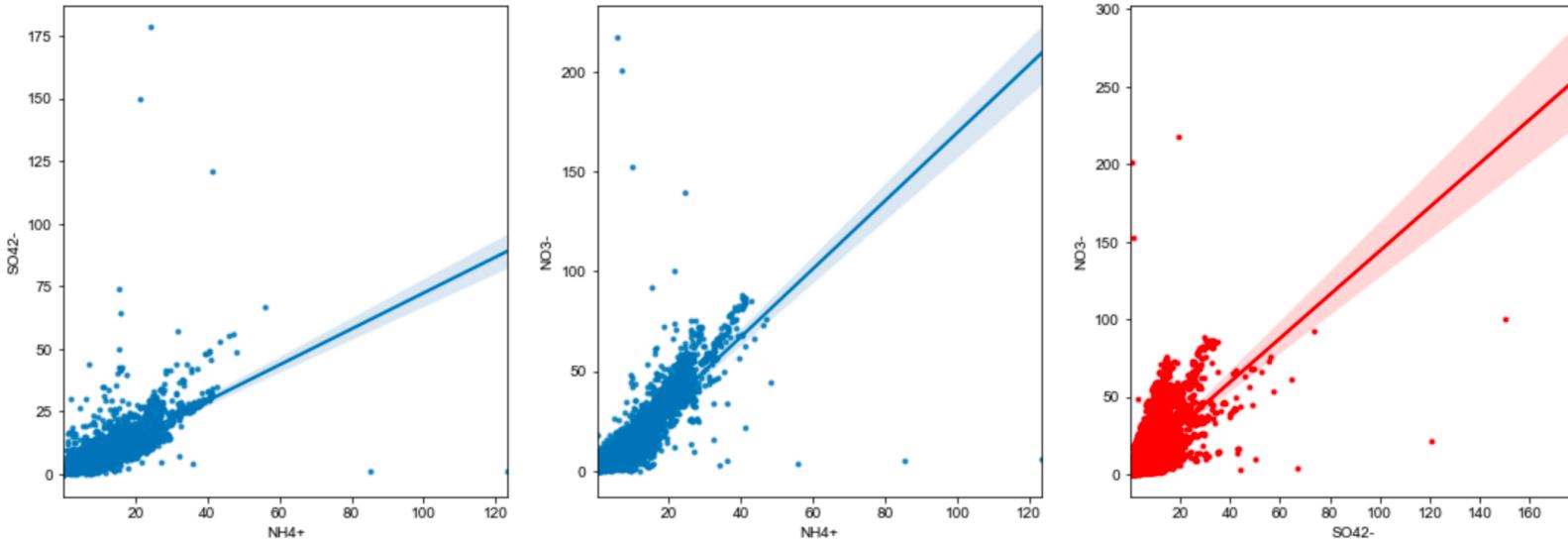
## • Result



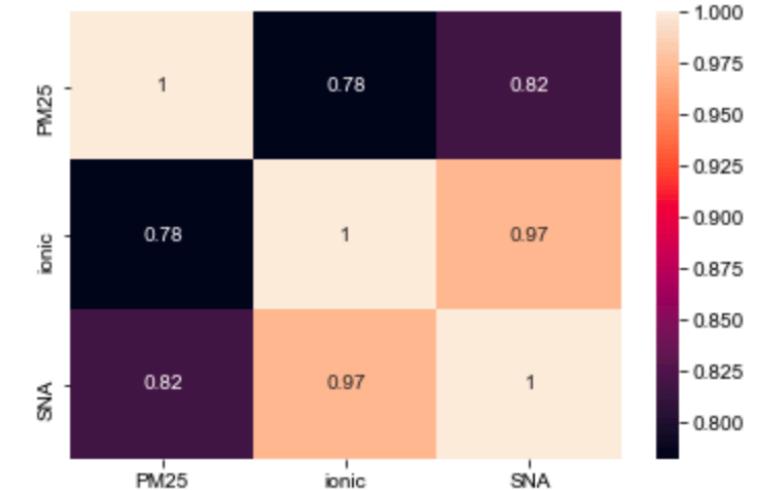
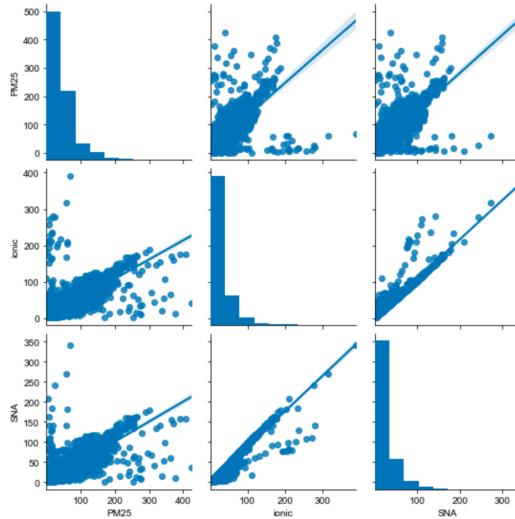
# Result



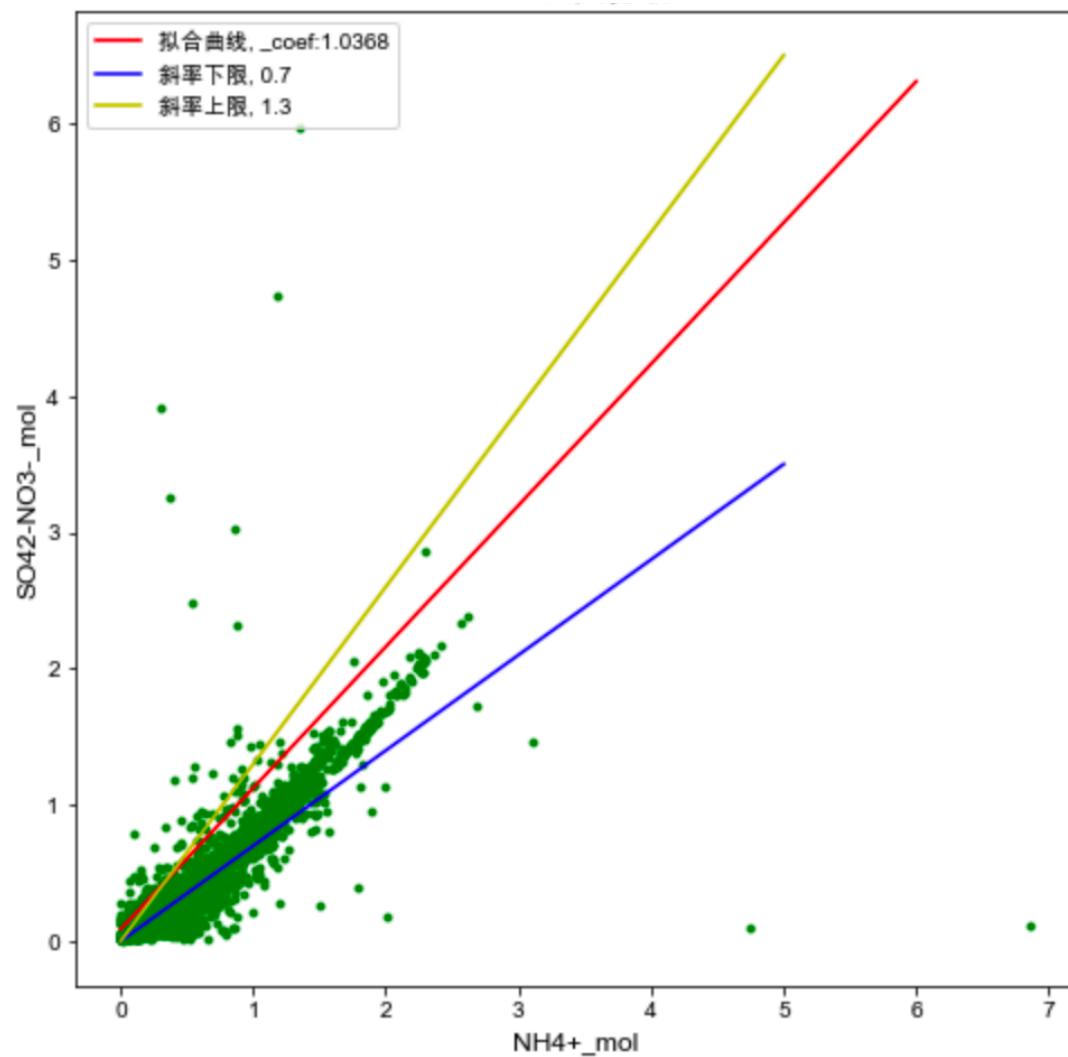
- SNA linear
  - SNA composition



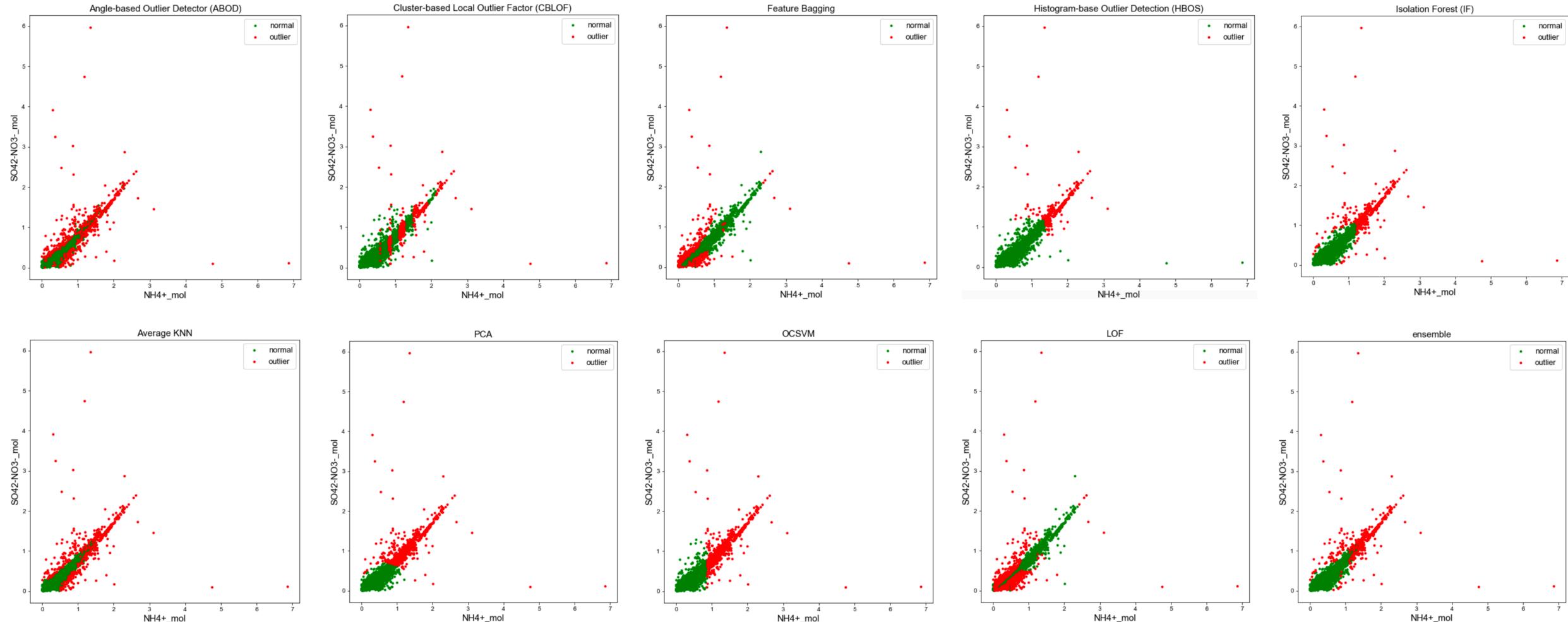
- SNA ionic PM25



- SNA Regression



# • Unsupervised Learning



- Correlation



# • Stats ML AI

		time	PM10	PM25	Cl-	N03-	S042-	Ca2+	Na+	K+	NH4+	Mg2+
24	2019-01-02 00:00:00	37.69	22.969	NaN		NaN	NaN	NaN	NaN	NaN	3.095	NaN
144	2019-01-07 00:00:00	161.62	102.870	NaN		NaN	NaN	NaN	NaN	NaN	10.001	NaN
384	2019-01-17 00:00:00	133.66	71.910	NaN		NaN	NaN	NaN	NaN	NaN	5.444	NaN
6237	2019-09-17 21:00:00	40.35	51.840	0.3		8.101	NaN	0.092	0.556	0.303	6.287	0.042

		time	PM10	PM25	Cl-	N03-	S042-	Ca2+	Na+	K+	NH4+	Mg2+
24	2019-01-02 00:00:00	37.69	22.969	NaN		4.355	2.832	NaN	NaN	NaN	3.095	NaN
144	2019-01-07 00:00:00	161.62	102.870	NaN		15.924	7.708	NaN	NaN	NaN	10.001	NaN
384	2019-01-17 00:00:00	133.66	71.910	NaN		8.290	4.491	NaN	NaN	NaN	5.444	NaN
6237	2019-09-17 21:00:00	40.35	51.840	0.3		9.702	5.086	0.092	0.556	0.303	6.287	0.042

		time	PM10	PM25	Cl-	N03-	S042-	Ca2+	Na+	K+	NH4+	Mg2+
0	2019-01-01 00:00:00	97.92	66.830	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN
24	2019-01-02 00:00:00	37.69	22.969	NaN		NaN	NaN	NaN	NaN	NaN	3.095	NaN
39	2019-01-02 15:00:00	61.55	35.035	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN
72	2019-01-04 00:00:00	180.96	130.330	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN
120	2019-01-06 00:00:00	49.92	30.530	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN
134	2019-01-06 14:00:00	87.26	52.470	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN
135	2019-01-06 15:00:00	109.17	61.730	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN
136	2019-01-06 16:00:00	109.38	66.930	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN
137	2019-01-06 17:00:00	123.86	76.230	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN
138	2019-01-06 18:00:00	121.77	81.430	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN
139	2019-01-06 19:00:00	136.71	74.100	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN
140	2019-01-06 20:00:00	140.12	81.230	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN
144	2019-01-07 00:00:00	161.62	102.870	NaN		NaN	NaN	NaN	NaN	NaN	10.001	NaN
168	2019-01-08 00:00:00	31.23	17.679	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN

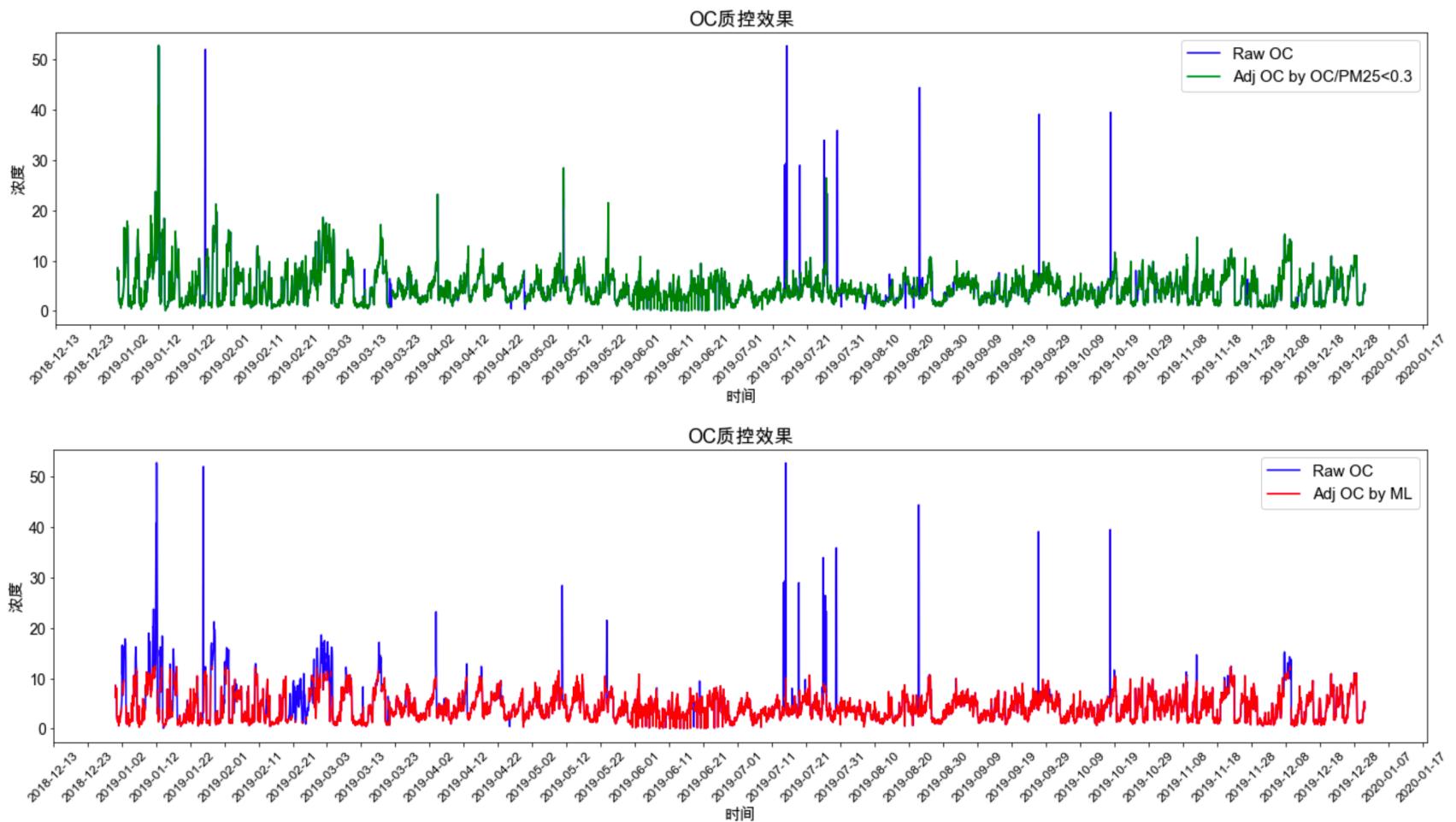
- OCEC

- OC/PM25
- EC/PM25
- (OC+EC)/PM25
- OC/EC

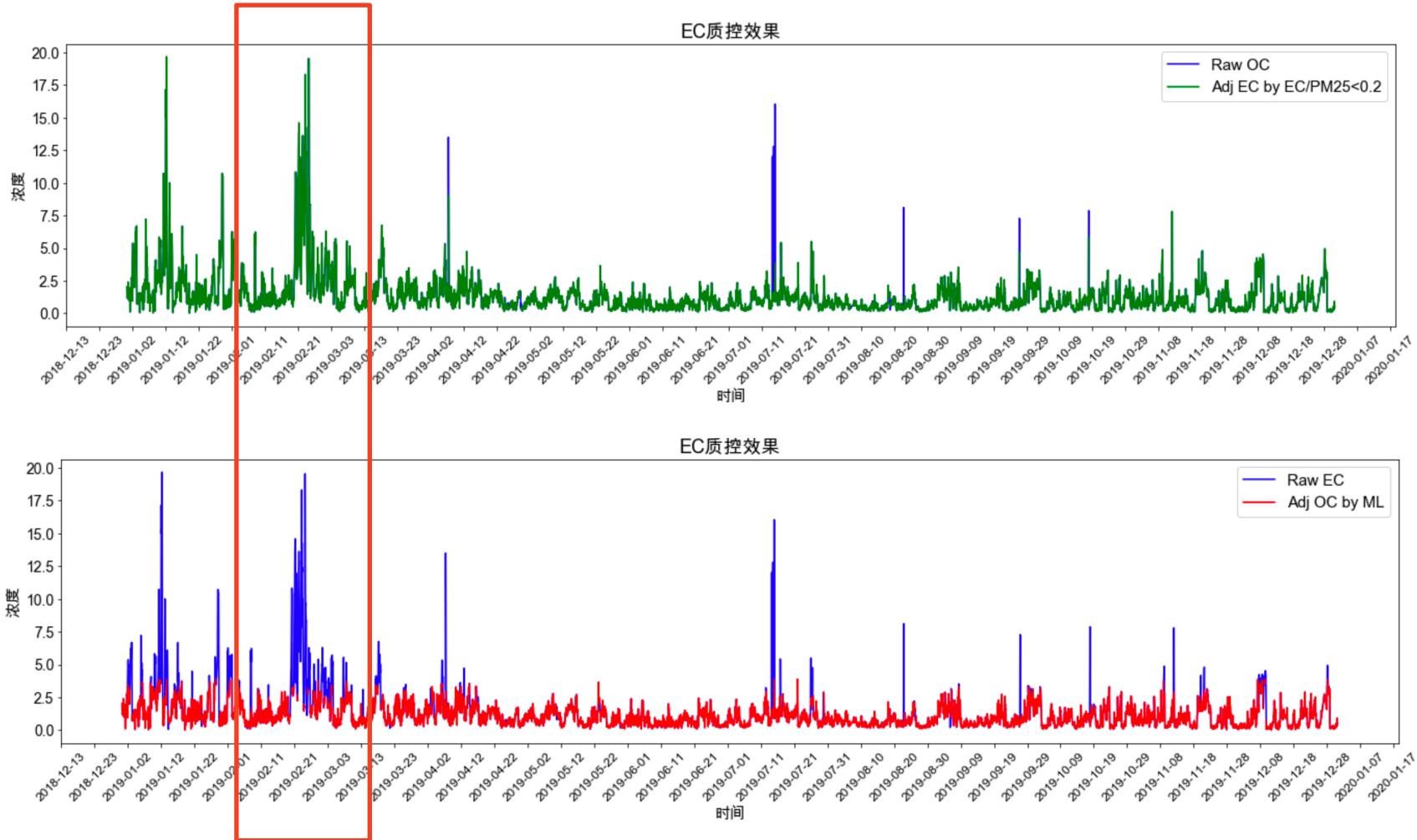
	time	PM10	PM25	OC	EC	oc/pm25	ec/pm25	ocec/pm25	oc/ec
1844	2019-03-18 20:00:00	217.17	113.050	11.600	2.741	0.102609	0.024246	0.126855	4.232032
1373	2019-02-27 05:00:00	127.37	93.840	7.703	1.559	0.082087	0.016613	0.098700	4.940988
8016	2019-12-01 00:00:00	43.98	16.340	1.210	0.170	0.074051	0.010404	0.084455	7.117647
4516	2019-07-08 04:00:00	23.66	38.940	2.687	0.395	0.069004	0.010144	0.079147	6.802532
717	2019-01-30 21:00:00	43.72	17.746	1.379	0.606	0.077708	0.034149	0.111856	2.275578

	PM10	PM25	OC	EC	oc/pm25	ec/pm25	ocec/pm25	oc/ec
count	8583.000000	8583.000000	8583.000000	8583.000000	8583.000000	8583.000000	8583.000000	8583.000000
mean	68.871765	45.334546	10.227934	1.994792	0.488306	0.082334	0.570640	4.724572
std	65.520722	40.981893	519.936839	66.040908	32.702307	4.157273	36.856863	2.801153
min	0.007000	0.039000	0.040000	0.010000	0.000816	0.000666	0.003467	0.166667
5%	12.551000	9.491000	1.096400	0.230000	0.046671	0.009911	0.065042	1.203337
10%	18.150000	12.370000	1.355200	0.287000	0.058786	0.012650	0.079423	1.586516
50%	47.430000	32.950000	3.860000	0.915000	0.110442	0.026792	0.141794	4.427640
80%	100.428000	65.736000	6.451600	1.752600	0.160160	0.044839	0.198405	6.306906
85%	119.517000	75.967000	7.140700	2.046700	0.174645	0.050517	0.214325	6.817113
90%	149.152000	91.672000	8.211600	2.482800	0.197793	0.058787	0.239045	7.617373
95%	194.317000	128.901000	10.199600	3.466900	0.247447	0.073531	0.294184	8.943691
97%	229.643200	147.670800	12.205620	4.460640	0.287893	0.084069	0.349394	9.997490
99%	311.280400	197.476800	16.190140	6.893560	0.429472	0.109014	0.503344	12.711072
max	916.600000	425.020000	48171.970000	6118.159000	3029.683648	384.789874	3414.473522	90.600000

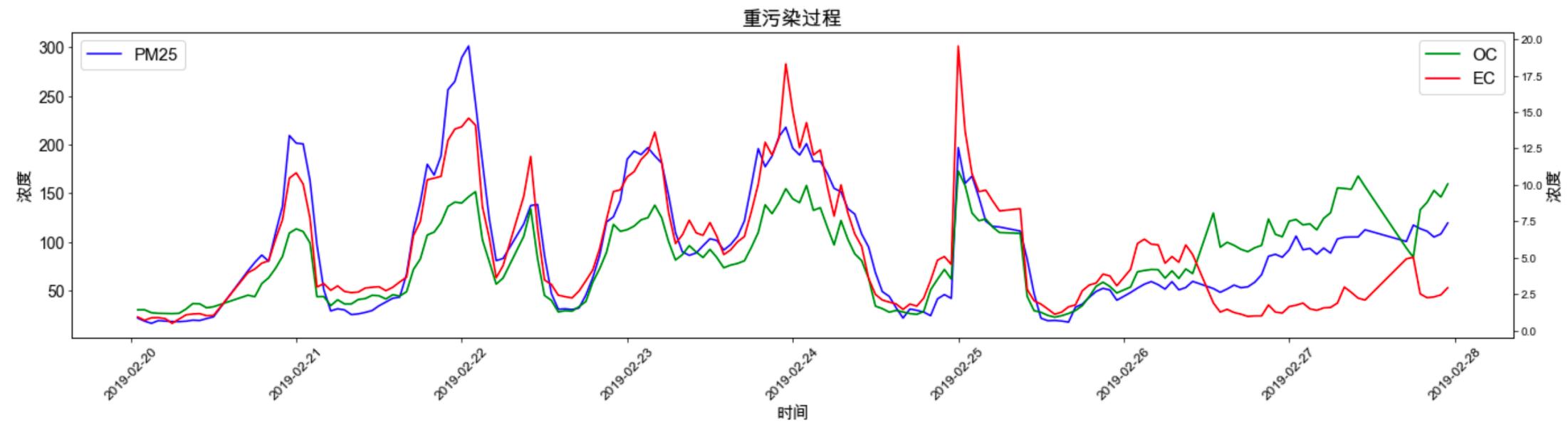
- Result



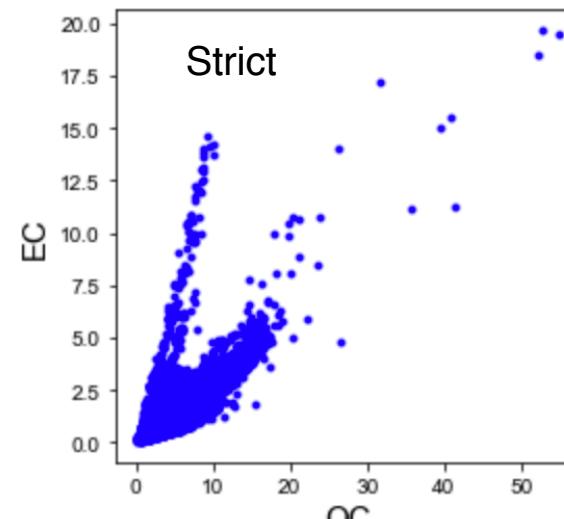
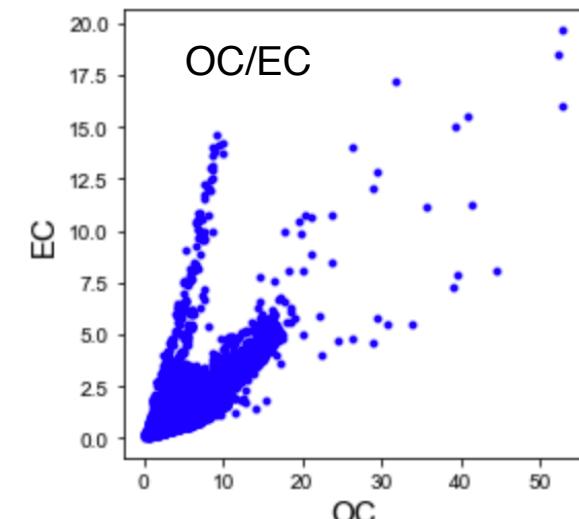
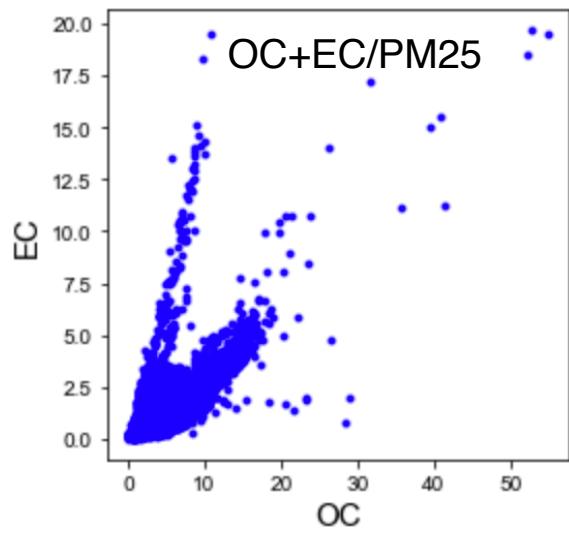
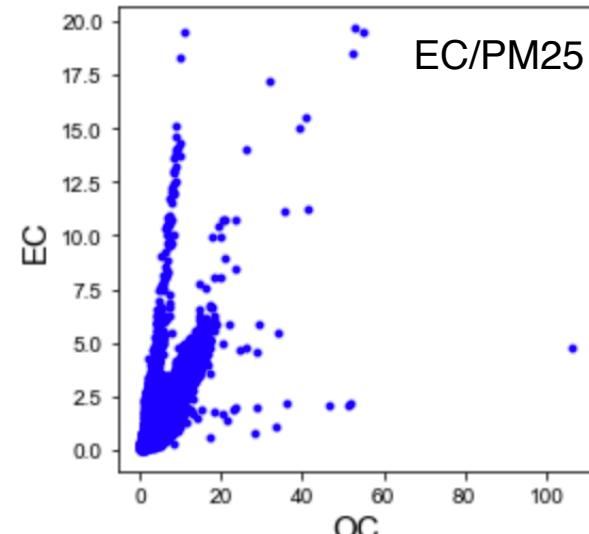
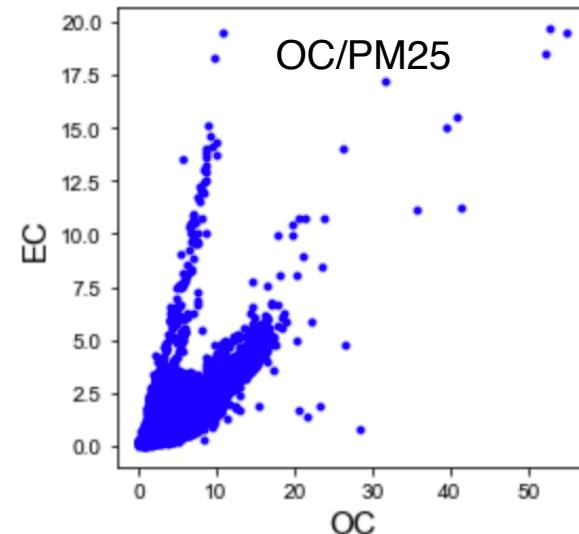
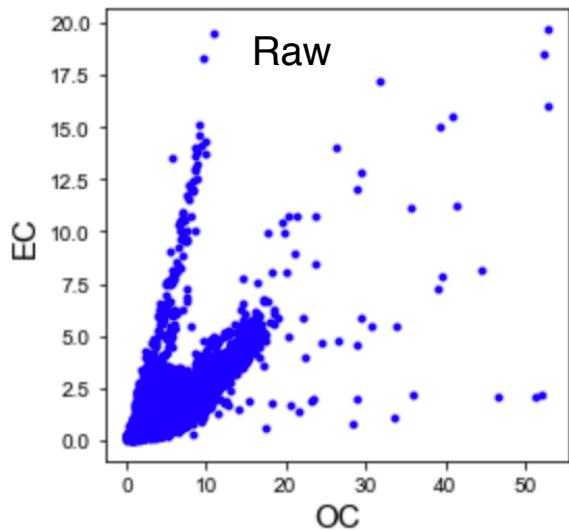
- Result



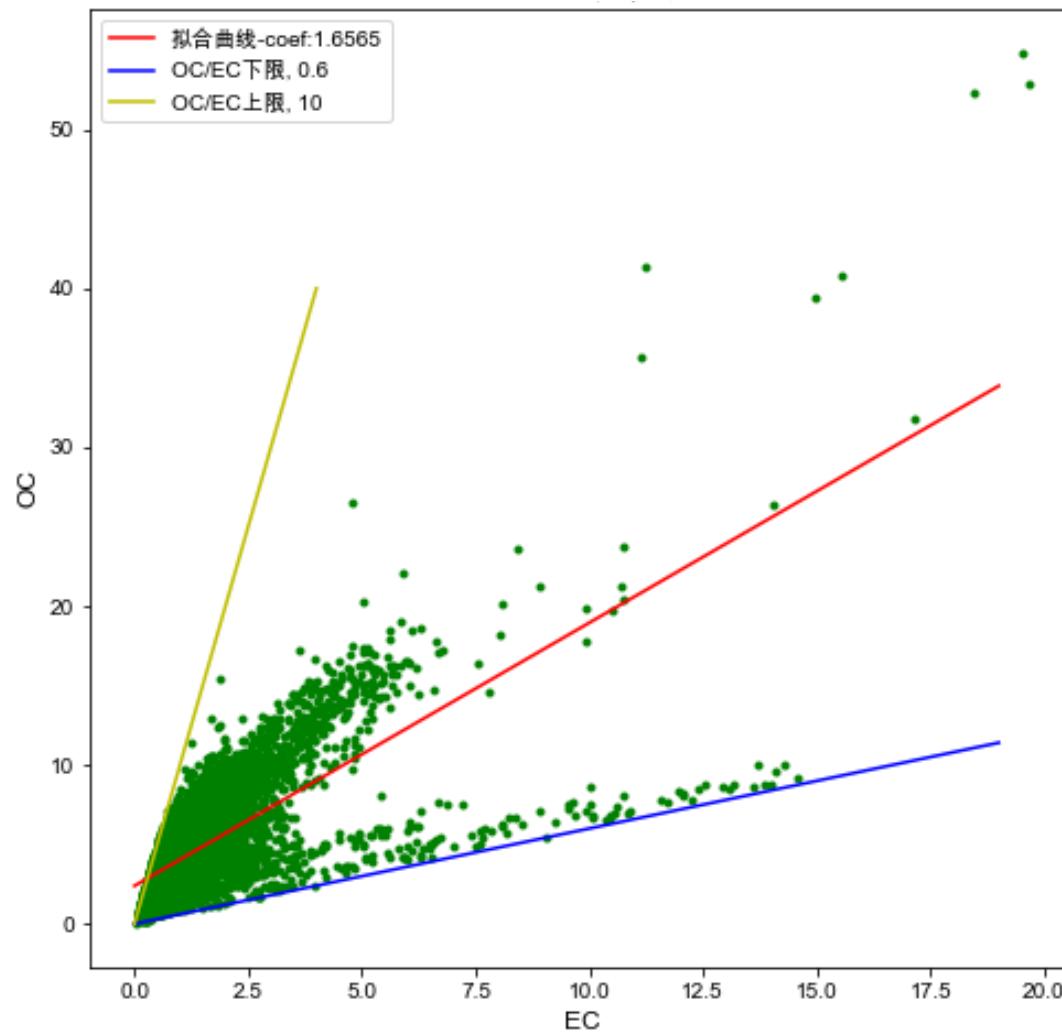
- Detail



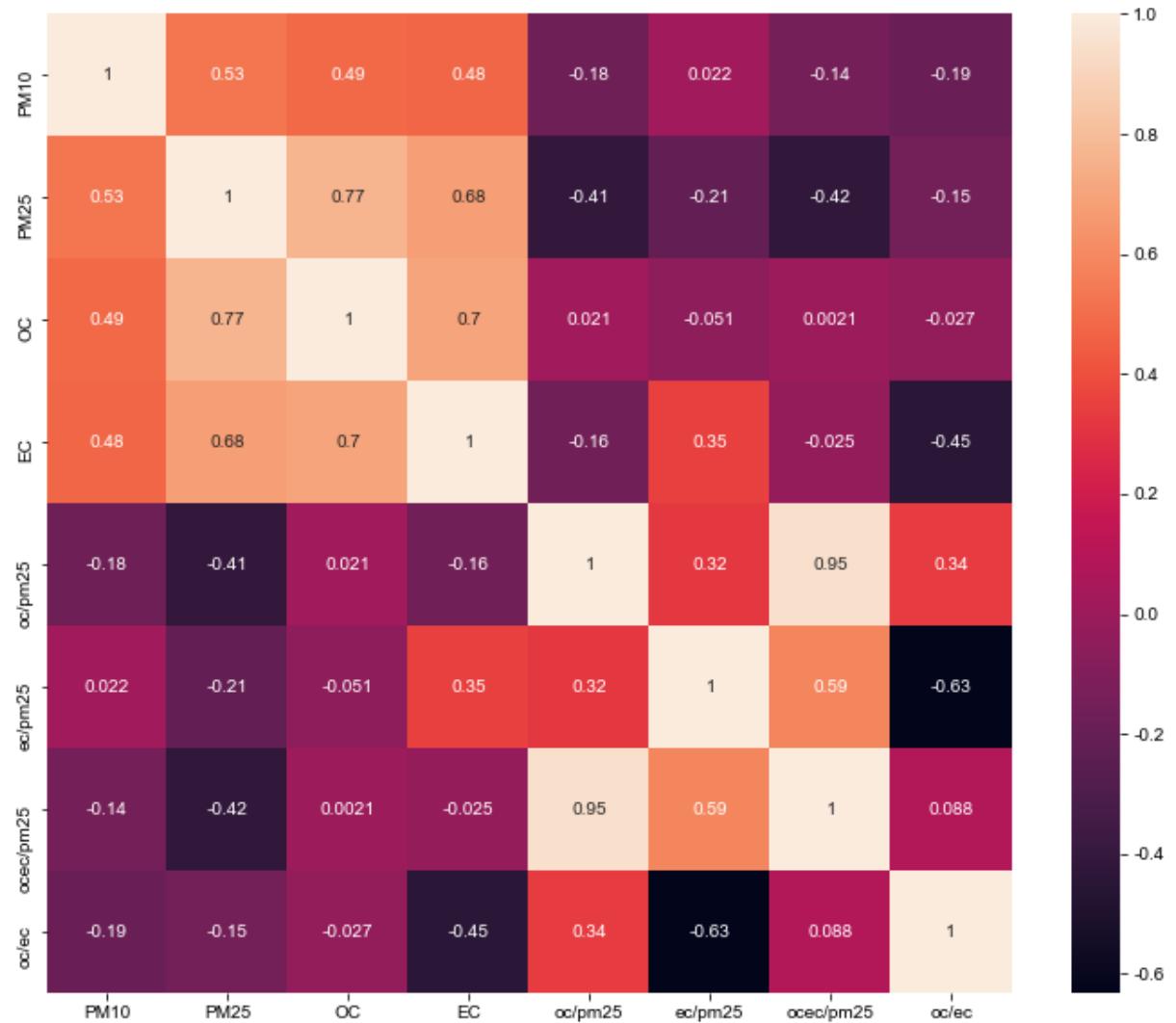
- Distribution



- Regression



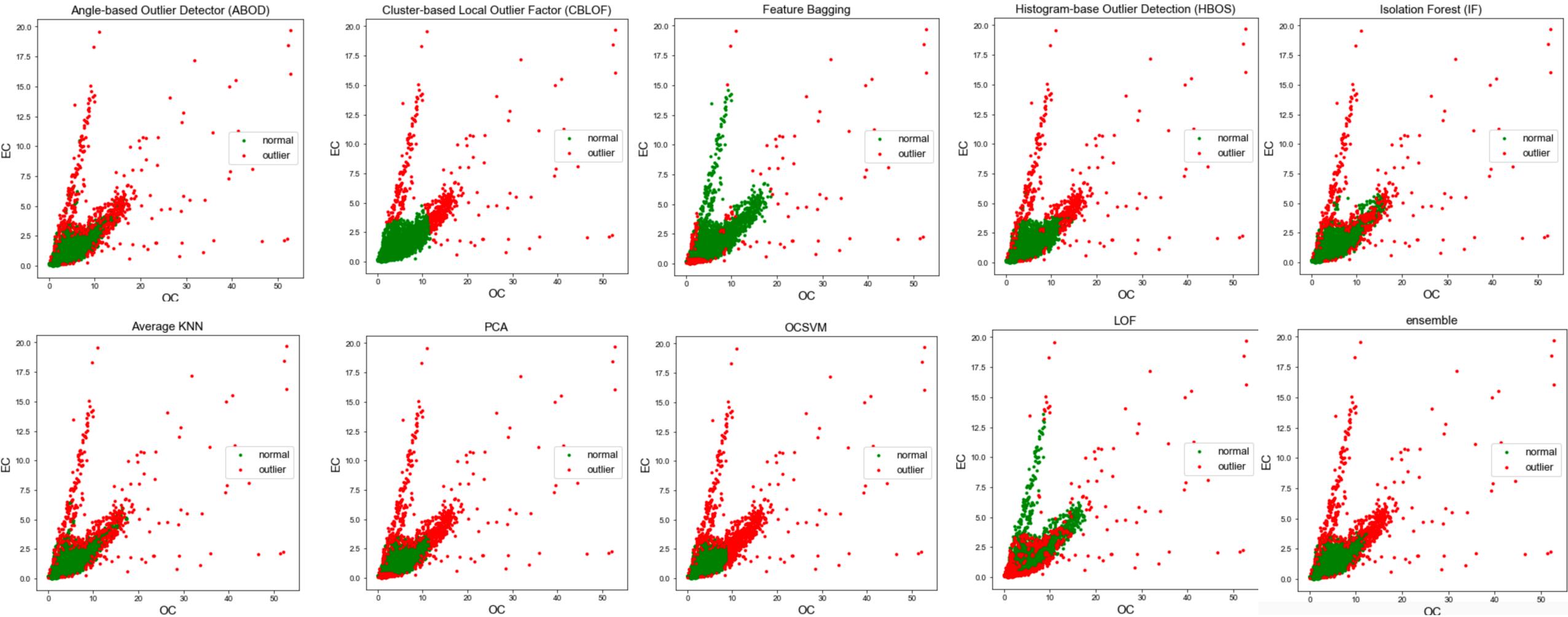
## • Correlation



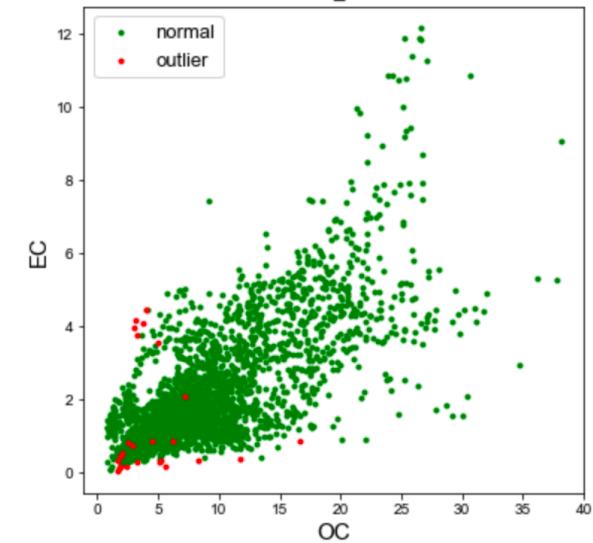
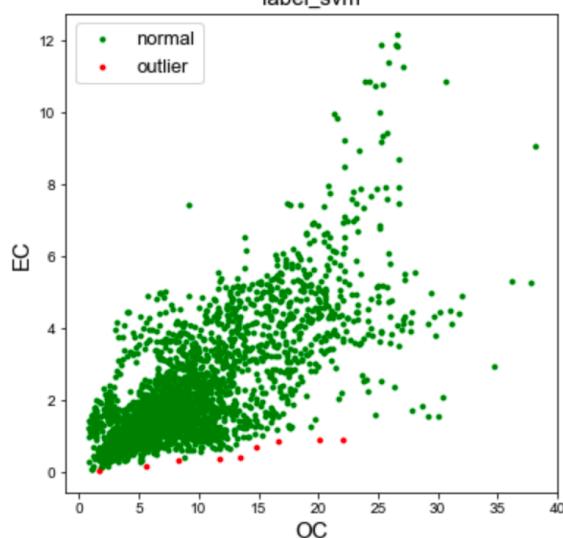
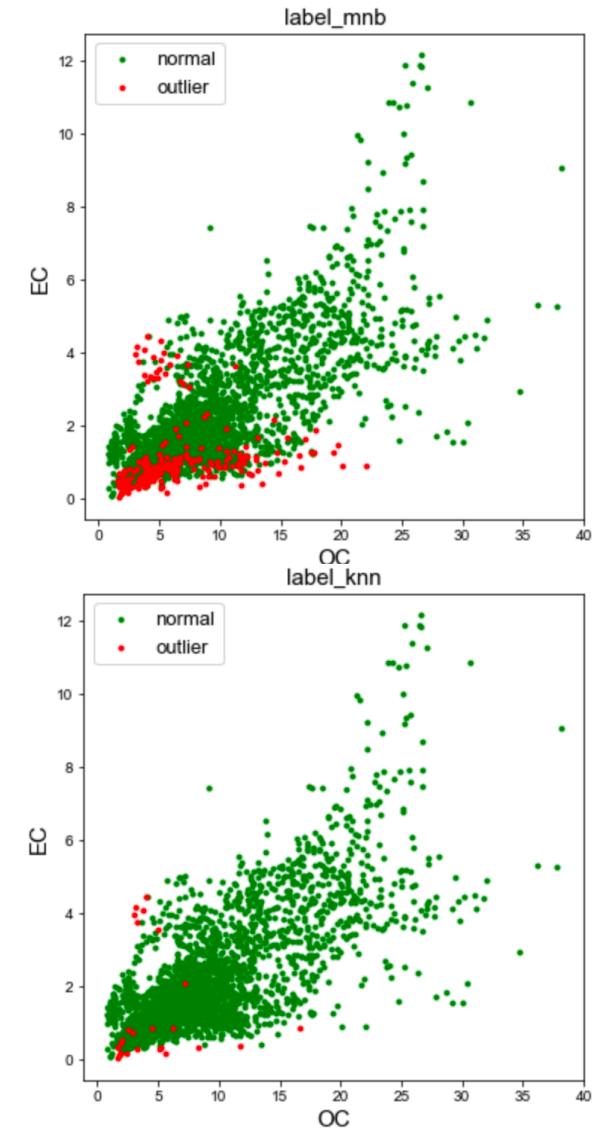
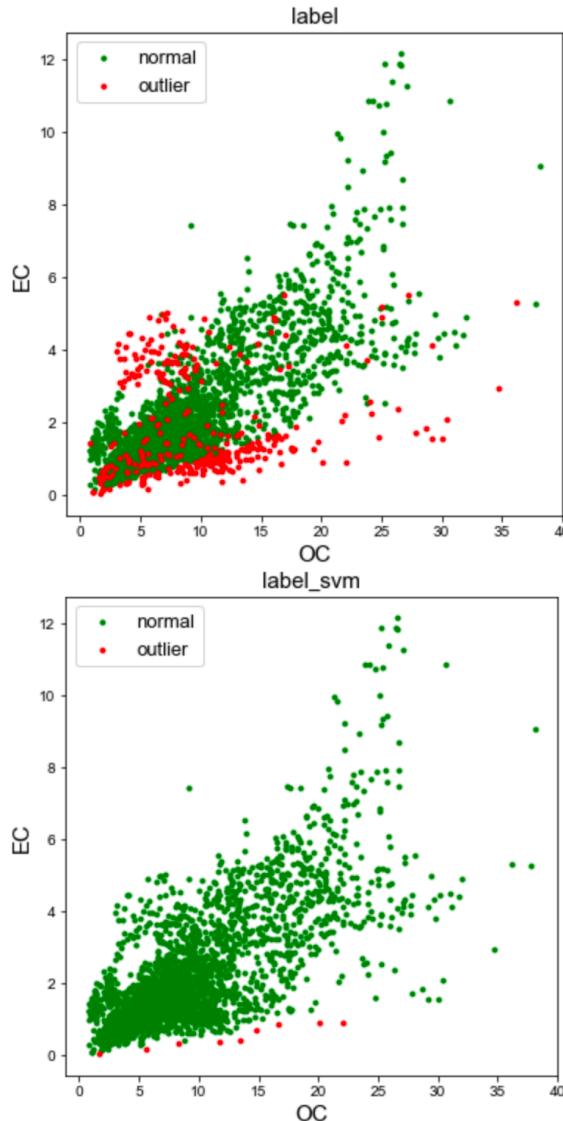
- ML-model

```
classifiers = {
    'Angle-based Outlier Detector (ABOD)': ABOD(contamination=outliers_fraction),
    'Cluster-based Local Outlier Factor (CBLOF)': CBLOF(contamination=outliers_fraction,
        check_estimator=False, random_state=random_state),
    'Feature Bagging': FeatureBagging(
        LOF(n_neighbors=35),
        contamination=outliers_fraction,
        check_estimator=False,
        random_state=random_state),
    'Histogram-base Outlier Detection (HBOS)': HBOS(contamination=outliers_fraction),
    'Isolation Forest (IF)': IForest(contamination=outliers_fraction, random_state=random_state),
    # 'K Nearest Neighbors': KNN(contamination=outliers_fraction),
    'Average KNN': KNN(method='mean',contamination=outliers_fraction),
    'PCA': PCA(svd_solver='auto'),
    'OCSVM': OCSVM(kernel='rbf', degree=4),
    'LOF': LOF()}`
```

# • Unsupervised Learning

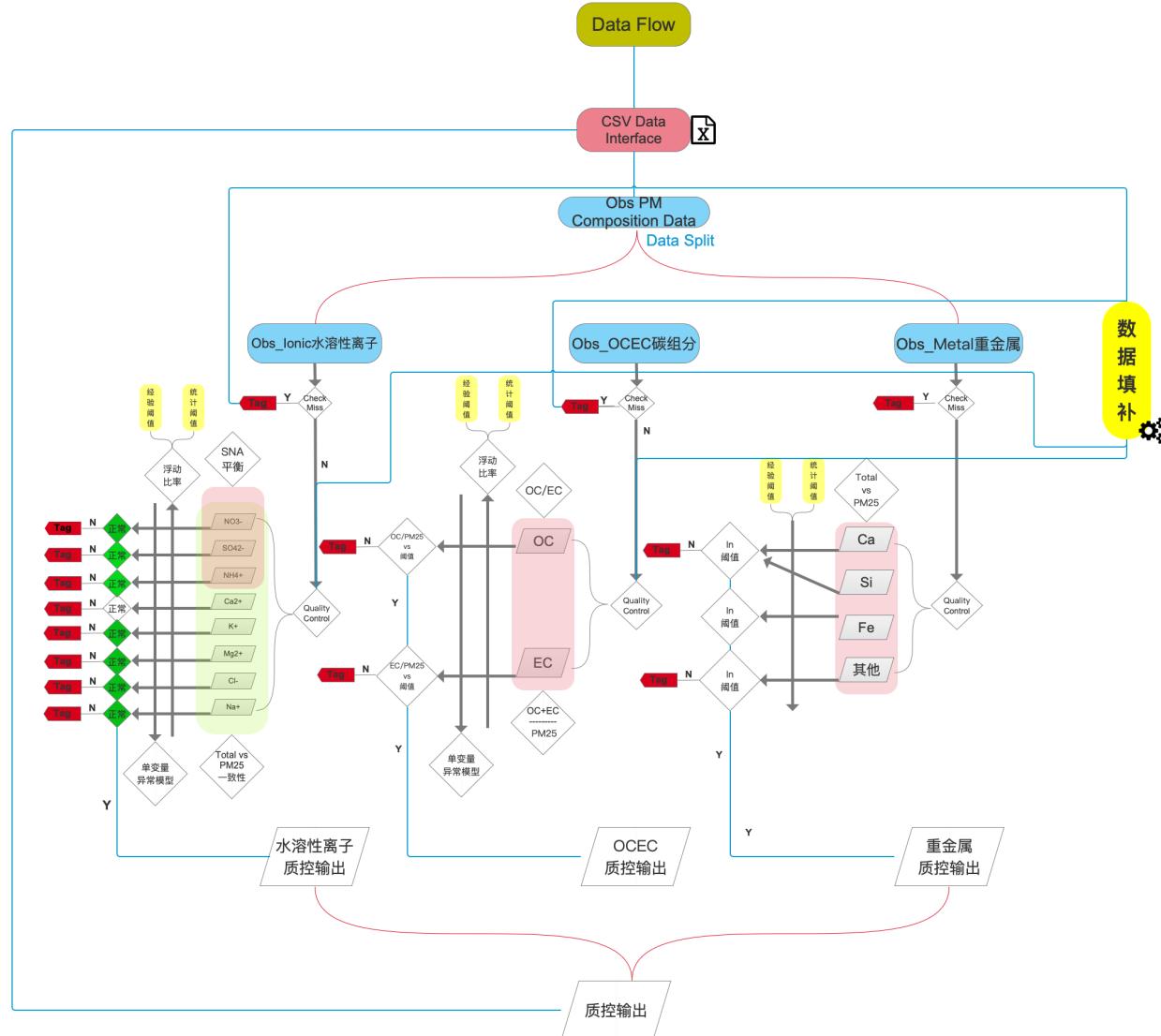


- Supervised Learning
  - Label
  - Naive Bayes
  - SVM
  - KNN



# 5 WorkFlow

## • WorkFlow



- Future Work

- AI
  - RNN
  - LSTM
  - Attention
  - GAN



Thanks