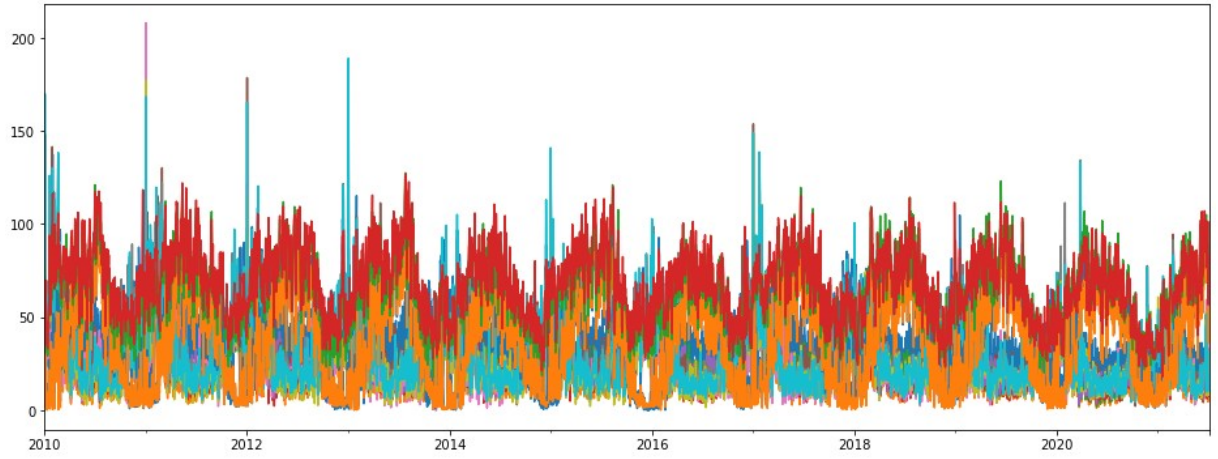
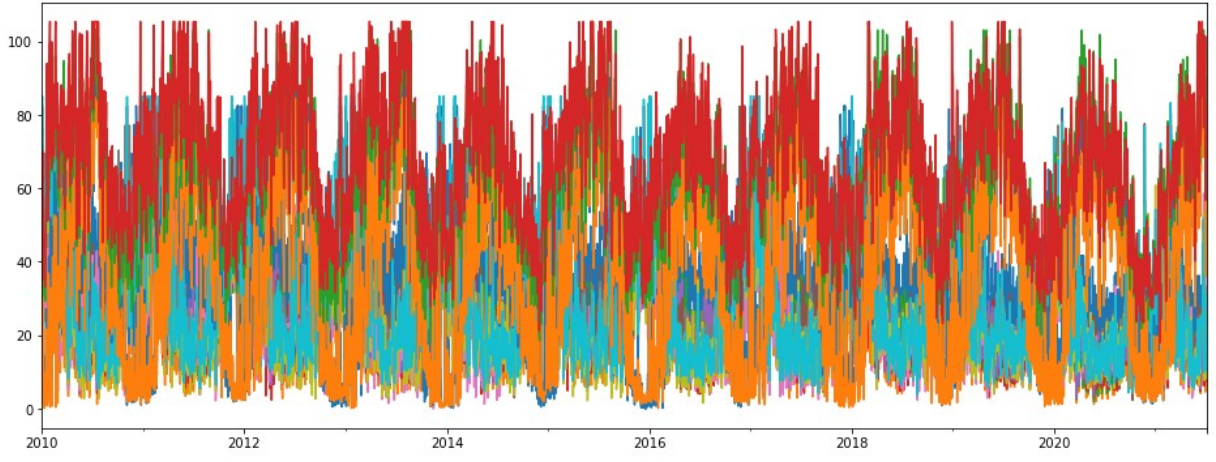
Graz air pollution

Full data

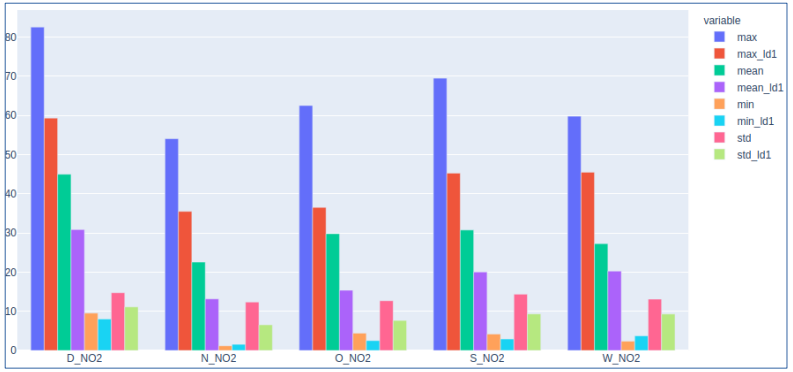
Cutoffs – 0.98

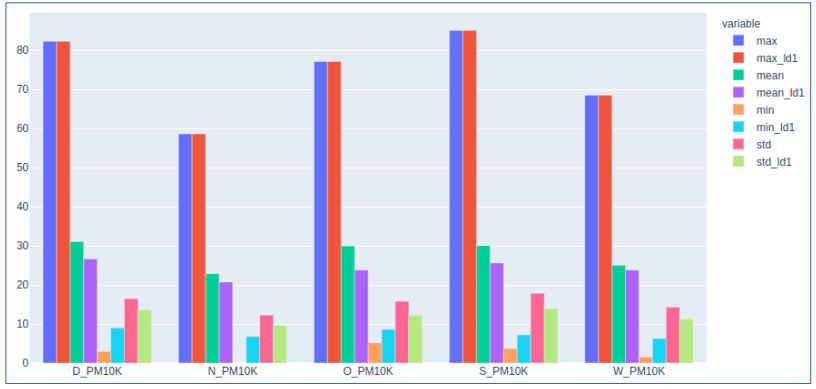

200
150
100
50
0
2010
2012
2014
2016
2018
2020

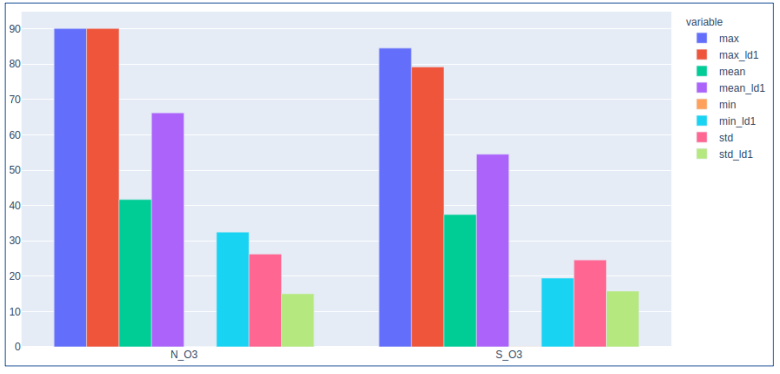

100
80
60
40
20
0
2010
2012
2014
2016
2018
2020

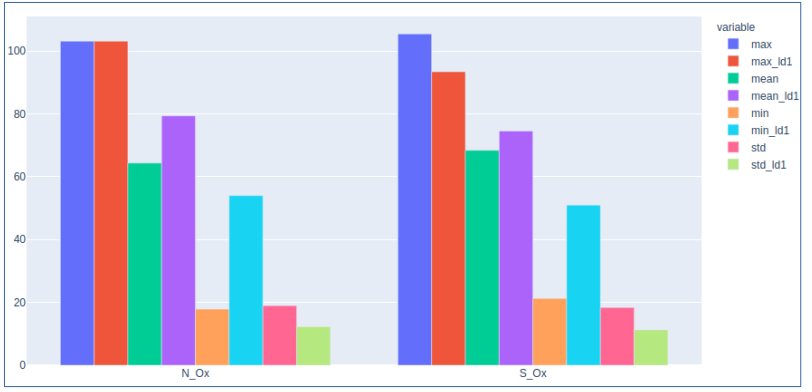
Lockdown1 vs. train

Lockdown1 vs. train

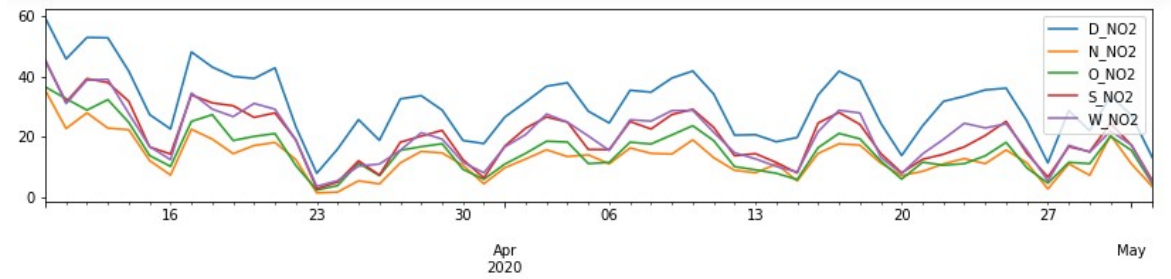


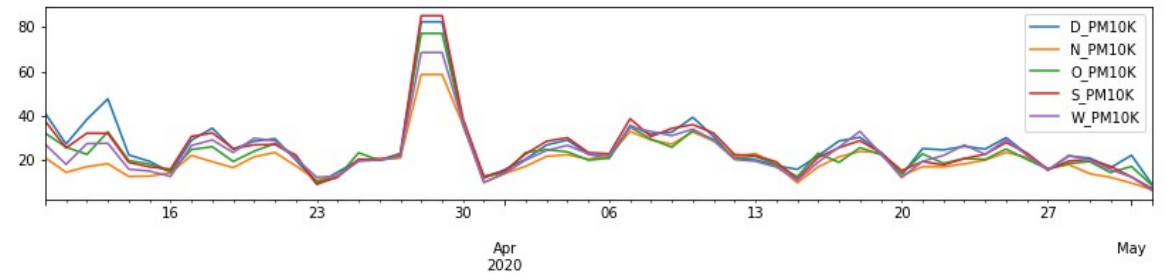


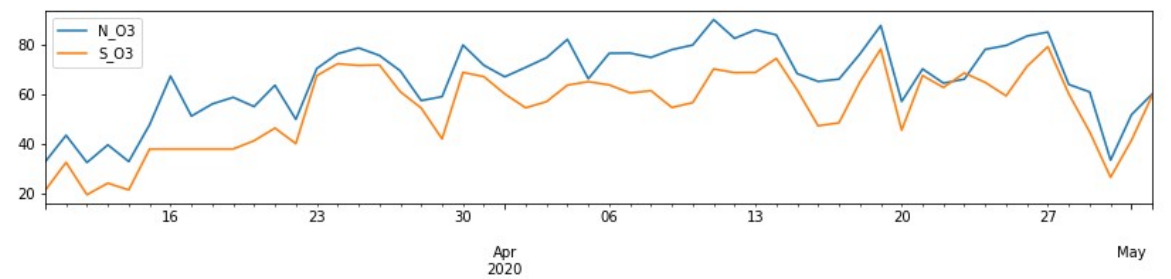


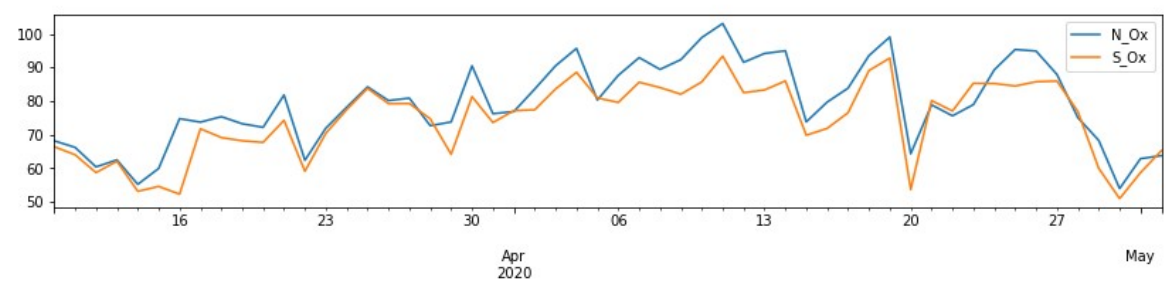


- Test data for NO2 is more simple than PM10
  - max_test ~ max_train


D_NO2
N_NO2
O_NO2
S_NO2
W_NO2
16
23
30
06
13
20
27
Apr
2020
May


D_PM10K
N_PM10K
O_PM10K
S_PM10K
W_PM10K
16
23
30
06
13
20
27
Apr
2020
May


N_O3
S_O3
16
23
30
06
13
20
27
Apr
2020
May


N_Ox
S_Ox
16
23
30
06
13
20
27
Apr
2020
May

test

Methods

1. Base additive: f0_add
2. Base multiplicative: f0_mul
3. + features: f1
4. + features + features2: f2
5. + holidays + features: f3
6. + holidays + featues + features2 + features3: f4
7. 6. + crossed_features: f5

```
features=['Temp', 'RH', 'Pressure',
          'Winddirection', 'Windspeed', 'Precip']

features2=['weekday_Friday', 'weekday_Monday',
       'weekday_Saturday', 'weekday_Sunday', 'weekday_Thursday',
       'weekday_Tuesday', 'weekday_Wednesday']

features3 = ['season_fall', 'season_spring',
             'season_summer', 'season_winter'
]
```

7. feature cross

```
weekday_convert = {
    'weekday_Sunday': 0,
    'weekday_Monday': 1,
    'weekday_Tuesday': 2,
    'weekday_Wednesday': 3,
    'weekday_Thursday': 4,
    'weekday_Friday': 5,
    'weekday_Saturday': 6
}
for c in features2:
    df_org[c] = [weekday_convert[c] if i else i for i in df_org[c]]
df_org['dayoftheweek'] = df_org[features2].sum(axis=1)
df_org[['dayoftheweek']]
```

```
season_convert = {
    'season_spring': 1,
    'season_summer': 2,
    'season_fall': 3,
    'season_winter': 4
}

for c in season_convert.keys():
    df_org[c] = [season_convert[c] if i else i for i in df_org[c]]
df_org['season'] = df_org[season_convert.keys()].sum(axis=1)
df_org[['season']]
```

```
df_org['cross1'] = df_org['season']*df_org['dayoftheweek']
df_org['cross2'] = df_org['season']**2 + df_org['dayoftheweek']**2
df_org['cross3'] = df_org['season']**2
df_org['cross4'] = df_org['dayoftheweek']**2
crosses = ['cross1', 'cross2', 'cross3', 'cross4']
```

Results

	D_NO2	N_NO2	O_NO2	S_NO2	W_NO2	D_PM10K	N_PM10K	O_PM10K	S_PM10K	W_PM10K	N_O3	S_O3	N_Ox	S_Ox
RMSE_f0_add	10.14	5.97	7.34	7.41	7.59	7.44	6.23	6.63	7.35	6.96	12.72	13.09	11.46	10.27
MAE_f0_add	7.58	4.37	5.58	5.31	5.85	6.31	5.10	5.04	5.88	5.82	10.67	10.73	9.52	8.76
RMSE_f0_mul	10.52	6.21	8.63	7.77	7.58	7.33	6.12	6.88	7.41	6.89	12.57	13.00	11.45	11.09
MAE_f0_mul	8.09	4.53	7.09	5.69	5.79	6.04	4.95	5.28	5.90	5.77	10.53	10.66	9.53	9.50
RMSE_f1	9.52	6.43	8.76	6.43	6.94	7.70	5.96	7.12	6.41	6.53	9.33	9.70	8.09	8.43
MAE_f1	7.20	4.91	7.89	4.97	5.23	6.34	5.06	5.69	5.17	5.50	7.24	7.81	6.68	6.75
RMSE_f2	9.59	6.43	8.79	6.44	6.93	7.71	5.95	7.09	6.35	6.53	9.31	9.68	8.12	8.49
MAE_f2	7.29	4.91	7.92	4.98	5.22	6.35	5.04	5.65	5.12	5.51	7.22	7.80	6.70	6.80
RMSE_f3	9.05	6.39	8.77	6.19	7.10	7.47	5.90	6.03	5.65	6.13	9.27	9.12	8.05	7.71
MAE_f3	7.06	4.84	7.82	4.83	5.46	6.14	4.97	4.79	4.41	4.95	7.31	7.45	6.66	5.89
RMSE_f4	9.10	6.44	8.79	6.18	7.15	7.49	5.90	5.86	5.66	6.11	9.18	9.03	8.05	7.66
MAE_f4	7.11	4.86	7.84	4.84	5.50	6.15	4.97	4.65	4.41	4.93	7.25	7.34	6.62	5.83
RMSE_f5	8.93	6.42	8.83	6.18	7.09	7.47	5.84	5.76	5.55	6.11	9.17	8.97	8.01	7.60
MAE_f5	7.01	4.85	7.91	4.85	5.46	6.19	4.88	4.59	4.24	4.91	7.25	7.31	6.62	5.76

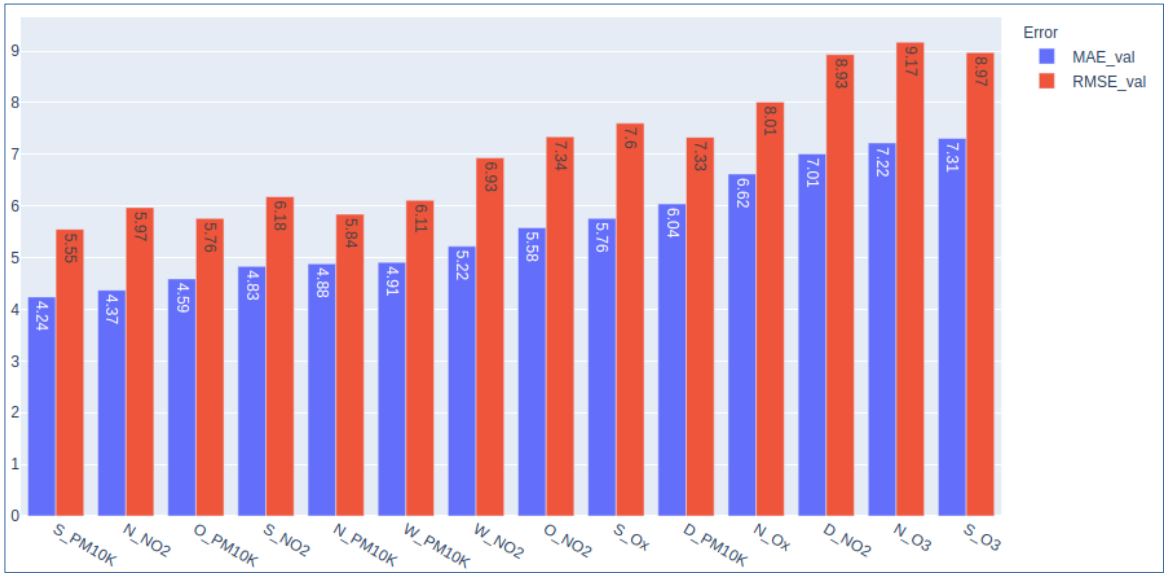
The best

Without [26.03 – 30.03]

	RMSE	MAE	RMSE_val	MAE_val
S_PM10K	f5	f5	5.549522	4.242137
O_PM10K	f5	f5	5.758059	4.593578
N_PM10K	f5	f5	5.835789	4.875501
N_NO2	f0_add	f0_add	5.973273	4.369344
W_PM10K	f5	f5	6.108925	4.909600
S_NO2	f4	f3	6.180596	4.827288
W_NO2	f2	f2	6.929287	5.217883
D_PM10K	f0_mul	f0_mul	7.329732	6.039981
O_NO2	f0_add	f0_add	7.337199	5.575182
S_Ox	f5	f5	7.595386	5.755069
N_Ox	f5	f5	8.005494	6.615279
D_NO2	f5	f5	8.929430	7.012986
S_O3	f5	f5	8.967940	7.305033
N_O3	f5	f2	9.169308	7.223101

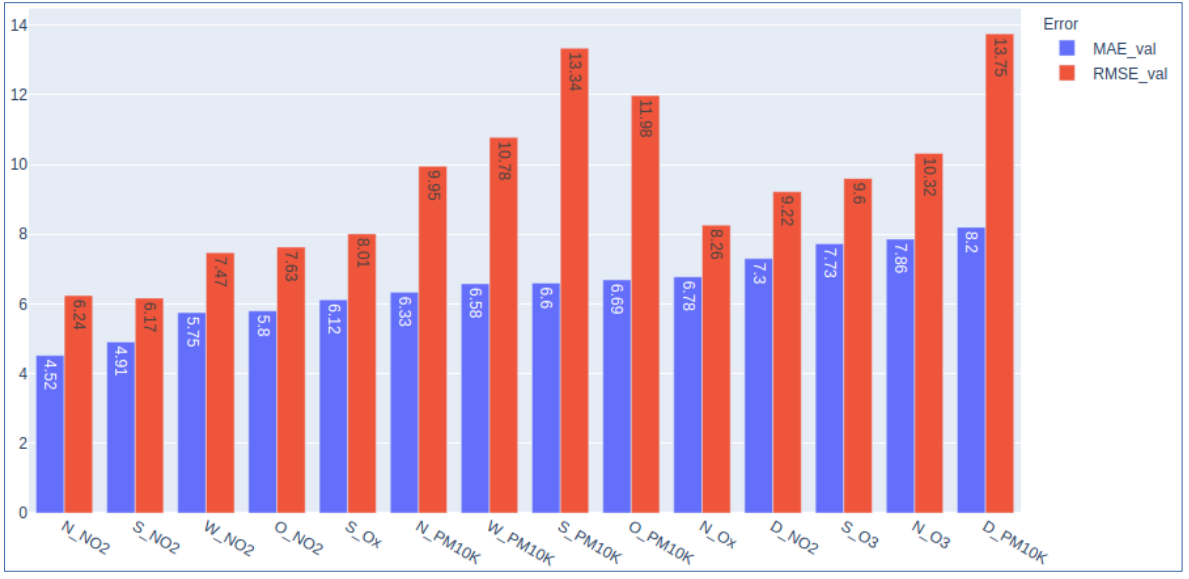
With [26.03 – 30.03]

	RMSE	MAE	RMSE_val	MAE_val
S_NO2	f5	f5	6.168732	4.906557
N_NO2	f0_add	f0_add	6.243970	4.522569
W_NO2	f2	f2	7.466454	5.749956
O_NO2	f0_add	f0_add	7.626506	5.801559
S_Ox	f5	f5	8.006745	6.115079
N_Ox	f5	f5	8.260435	6.781672
D_NO2	f5	f5	9.223593	7.303620
S_O3	f5	f5	9.604191	7.732554
N_PM10K	f0_add	f0_mul	9.949572	6.331526
N_O3	f4	f2	10.323822	7.855165
W_PM10K	f5	f5	10.775693	6.576765
O_PM10K	f5	f5	11.980793	6.691246
S_PM10K	f5	f5	13.339639	6.603996
D_PM10K	f0_mul	f0_mul	13.747833	8.201998



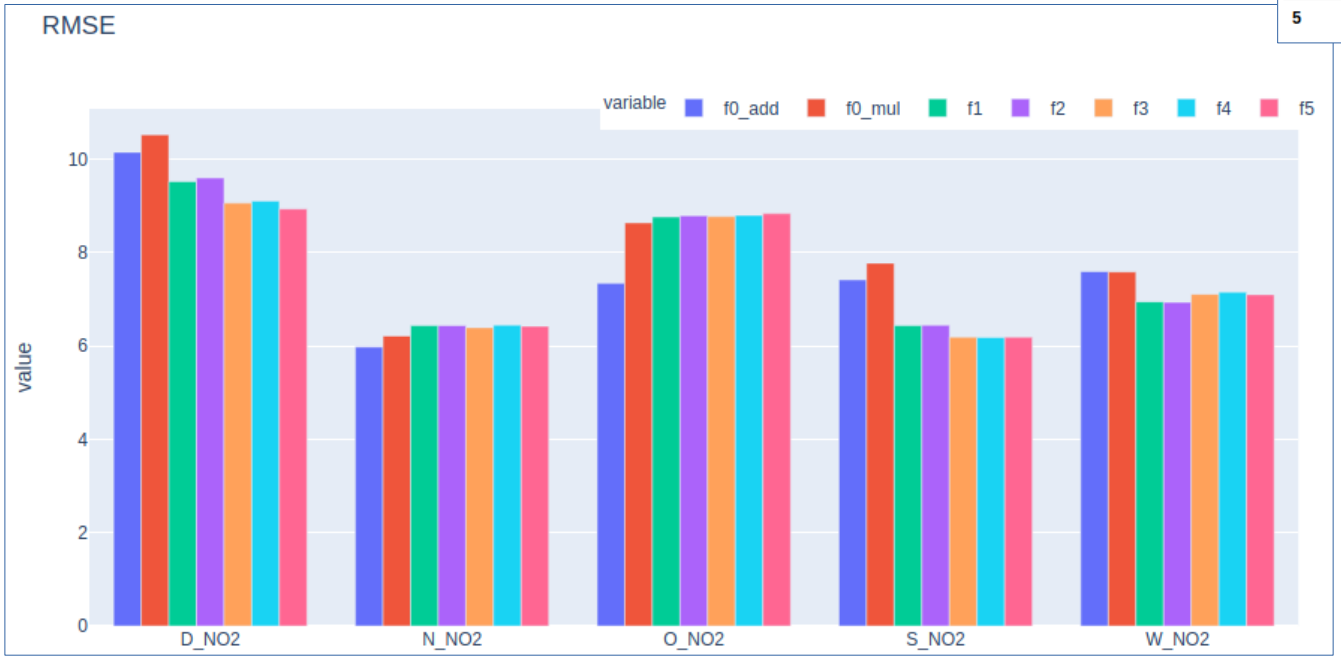
Should use MAE to measure,
b/c some abnormal points?

With [26.03 – 30.03]



1. Base additive: f0_add
2. Base multiplicative: f0_mul
3. + features: f1
4. + features + features2: f2
5. + holidays + features: f3
6. + holidays + featues + features2 + features3: f4
7. 6. + crossed_features: f5

	D	N	O	W	S
0	DonBosco_RH	Nord_Precip	Ost_Pressure	West_RH	Sud_RH
1	DonBosco_Temp	Nord_Pressure	Ost_RH	West_Temp	Sud_Temp
2	NaN	Nord_RH	Ost_Temp	West_Winddirection	Sud_Winddirection
3	NaN	Nord_Temp	Ost_Winddirection	West_Windspeed	Sud_Windspeed
4	NaN	Nord_Winddirection	Ost_Windspeed	NaN	NaN
5	NaN	Nord_Windspeed	NaN	NaN	NaN

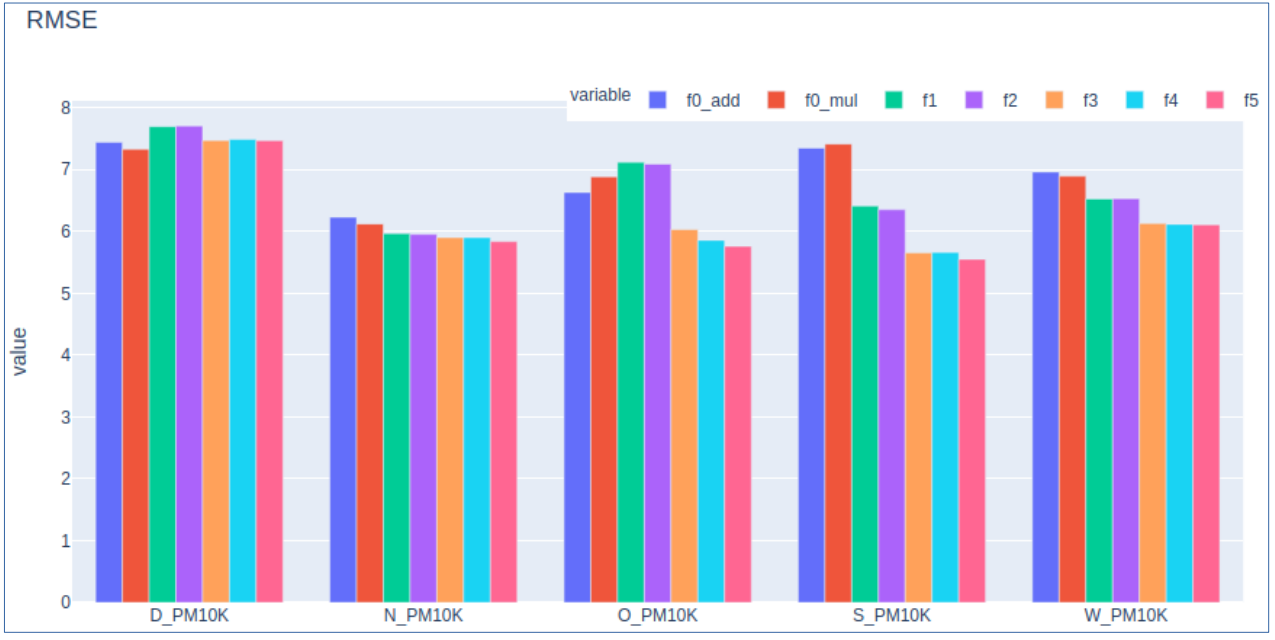


- Holidays contribute
- N_NO2 + O_NO2
  - Additive aggresors do not affect
  - Features of O_ are only available from 19.05.2017 → less data than other areas
- D_NO2
  - Improve not much due to
    - lack of extra features (only 2 available)
    - Test data is more complex than other areas
- S_NO2
  - f4 and f5 does not help to improve
- W_NO2
  - After f3, results not better

summary for NO2

➔ F2 not good → features2 do not affect

1. Base additive: f0_add
2. Base multiplicative: f0_mul
3. + features: f1
4. + features + features2: f2
5. + holidays + features: f3
6. + holidays + featues + features2 + features3: f4
7. 6. + crossed_features: f5



	D	N	O	W	S
0	DonBosco_RH	Nord_Precip	Ost_Pressure	West_RH	Sud_RH
1	DonBosco_Temp	Nord_Pressure	Ost_RH	West_Temp	Sud_Temp
2	NaN	Nord_RH	Ost_Temp	West_Winddirection	Sud_Winddirection
3	NaN	Nord_Temp	Ost_Winddirection	West_Windspeed	Sud_Windspeed
4	NaN	Nord_Winddirection	Ost_Windspeed	NaN	NaN
5	NaN	Nord_Windspeed	NaN	NaN	NaN

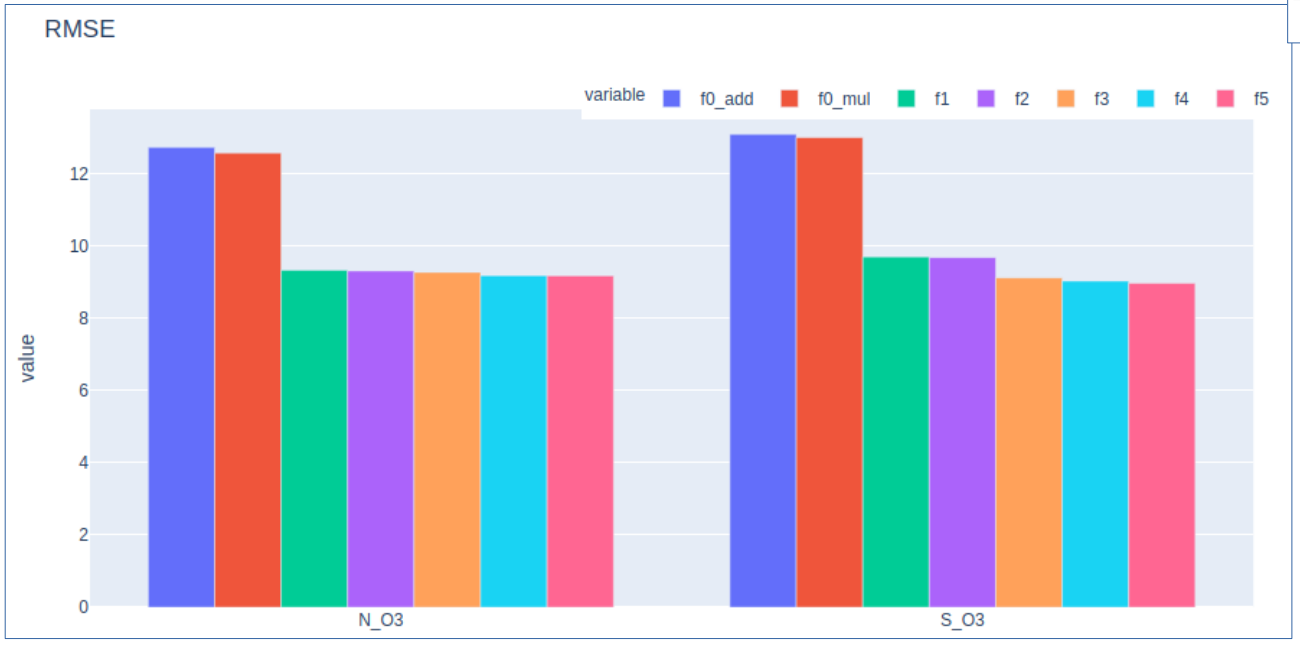
- Holidays contribute much

summary for PM10

- f2 not good → features2 do not affect
- f4 only works on O_PM10 → features3 do not affect
- cross more features

1. Base additive: f0_add
2. Base multiplicative: f0_mul
3. + features: f1
4. + features + features2: f2
5. + holidays + features: f3
6. + holidays + featues + features2 + features3: f4
7. 6. + crossed_features: f5

	D	N	O	W	S
0	DonBosco_RH	Nord_Precip	Ost_Pressure	West_RH	Sud_RH
1	DonBosco_Temp	Nord_Pressure	Ost_RH	West_Temp	Sud_Temp
2	NaN	Nord_RH	Ost_Temp	West_Winddirection	Sud_Winddirection
3	NaN	Nord_Temp	Ost_Winddirection	West_Windspeed	Sud_Windspeed
4	NaN	Nord_Winddirection	Ost_Windspeed	NaN	NaN
5	NaN	Nord_Windspeed	NaN	NaN	NaN



- The more modified algorithm and the more additional aggressors added, the better

summary for NO2

➔ F2 not good → features2 do not affect

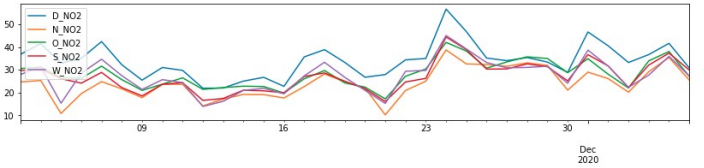
1. Base additive: f0_add
2. Base multiplicative: f0_mul
3. + features: f1
4. + features + features2: f2
5. + holidays + features: f3
6. + holidays + featues + features2 + features3: f4
7. 6. + crossed_features: f5

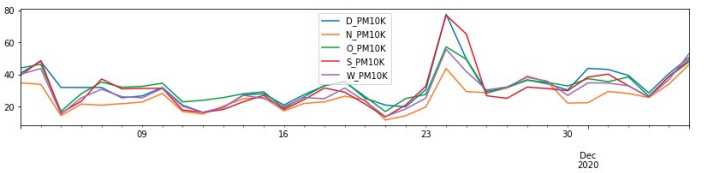
- Seems “features” to be enough for N_Ox, can not improve after f2
- Holidays work for S_Ox

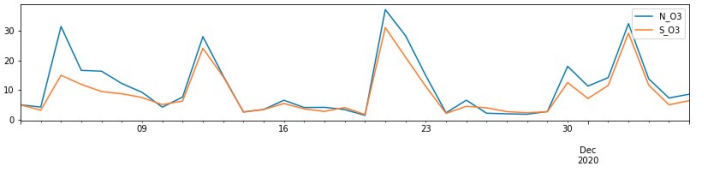
Lockdown2

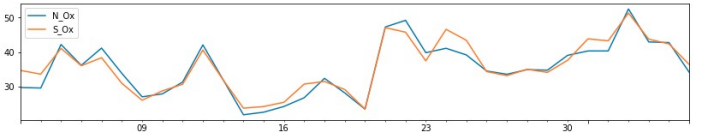
	#days
LD1	54
LD2	34
LD3	44

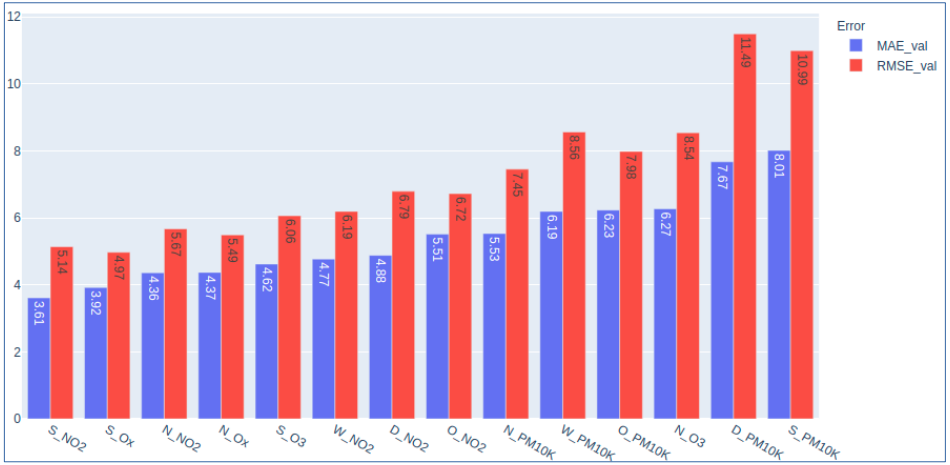
	RMSE	MAE	RMSE_val	MAE_val
S_Ox	f4	f4	4.967191	3.923796
S_NO2	f6	f6	5.143038	3.613367
N_Ox	f2	f6	5.490847	4.367127
N_NO2	f1	f6	5.674186	4.359572
S_O3	f6	f4	6.057256	4.623346
W_NO2	f6	f6	6.190127	4.774957
O_NO2	f0_add	f0_add	6.719727	5.511973
D_NO2	f6	f6	6.791037	4.883716
N_PM10K	f6	f6	7.450410	5.529715
O_PM10K	f6	f6	7.983604	6.229738
N_O3	f6	f6	8.544976	6.274708
W_PM10K	f6	f6	8.555882	6.187934
S_PM10K	f6	f6	10.992120	8.007369
D_PM10K	f0_add	f6	11.494510	7.667573

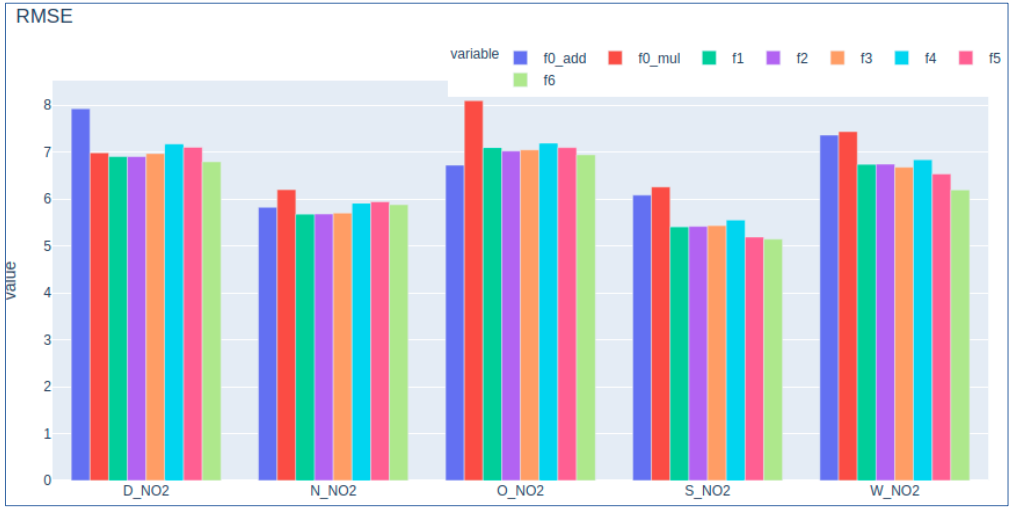


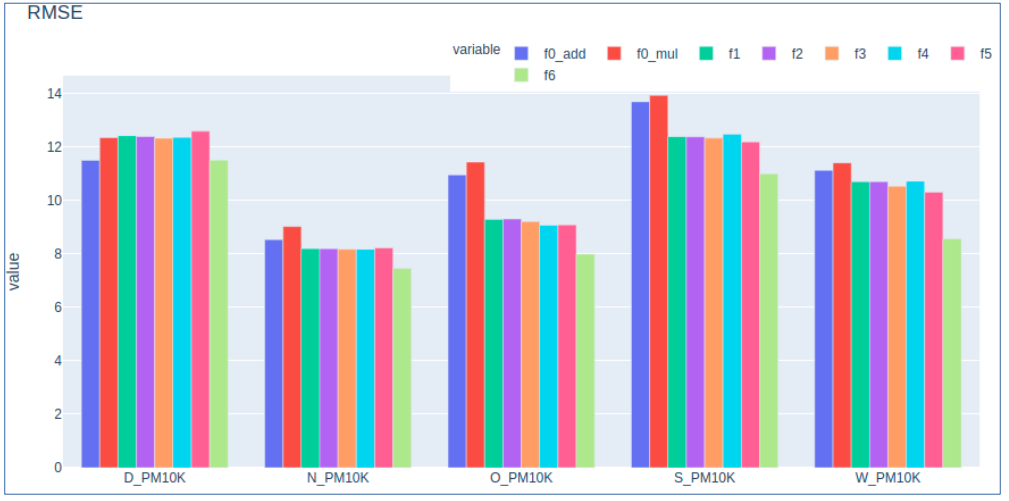









RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
f6
value
0
1
2
3
4
5
6
7
8
D_NO2
N_NO2
O_NO2
S_NO2
W_NO2


RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
f6
value
0
2
4
6
8
10
12
14
D_PM10K
N_PM10K
O_PM10K
S_PM10K
W_PM10K

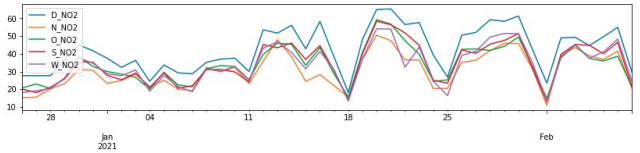

RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
f6
value
0
2
4
6
8
10
N_O3
S_O3

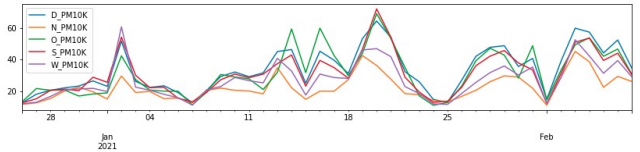

RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
f6
value
0
1
2
3
4
5
6
7
8
N_Ox
S_Ox

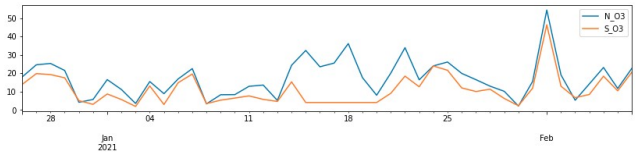
Lockdown3

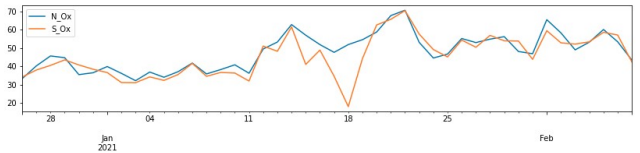
	#days
LD1	54
LD2	34
LD3	44

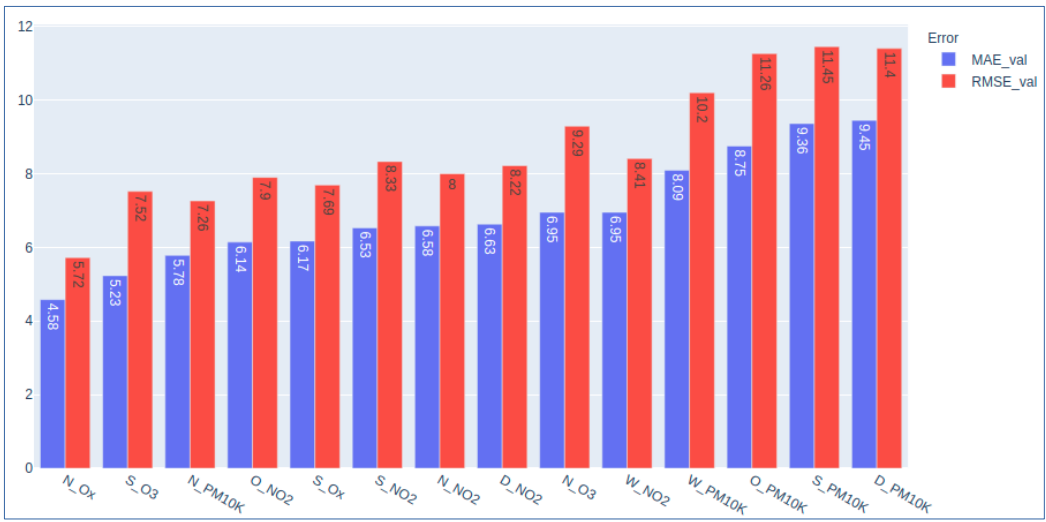
	RMSE	MAE	RMSE_val	MAE_val
N_Ox	f3	f6	5.719025	4.576951
N_PM10K	f6	f6	7.261158	5.784290
S_O3	f5	f3	7.522663	5.234595
S_Ox	f5	f5	7.687741	6.167953
O_NO2	f0_add	f6	7.895804	6.138974
N_NO2	f0_add	f0_add	7.995718	6.576368
D_NO2	f0_add	f0_add	8.216432	6.632510
S_NO2	f6	f6	8.332976	6.530794
W_NO2	f6	f6	8.414429	6.948812
N_O3	f6	f6	9.294481	6.946582
W_PM10K	f6	f0_mul	10.198871	8.093743
O_PM10K	f6	f6	11.262424	8.754989
D_PM10K	f6	f6	11.399489	9.452694
S_PM10K	f6	f3	11.451918	9.360340

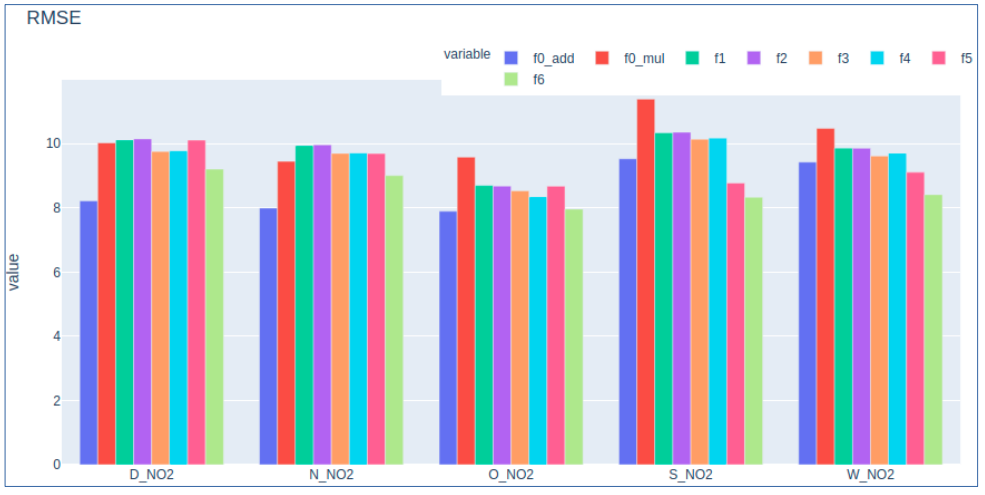


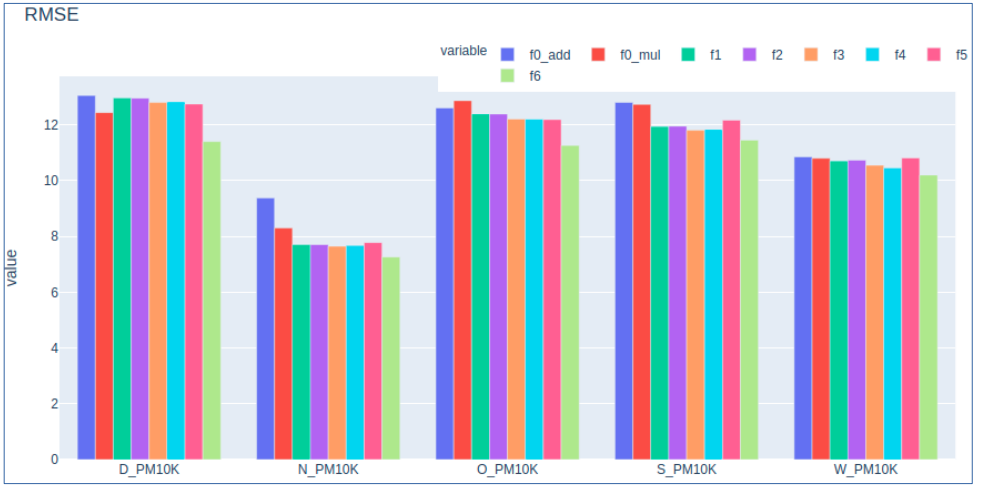


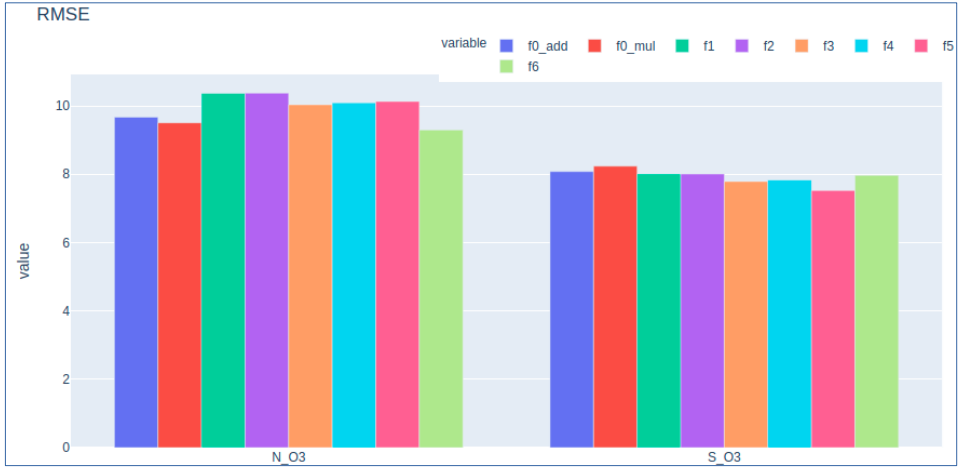


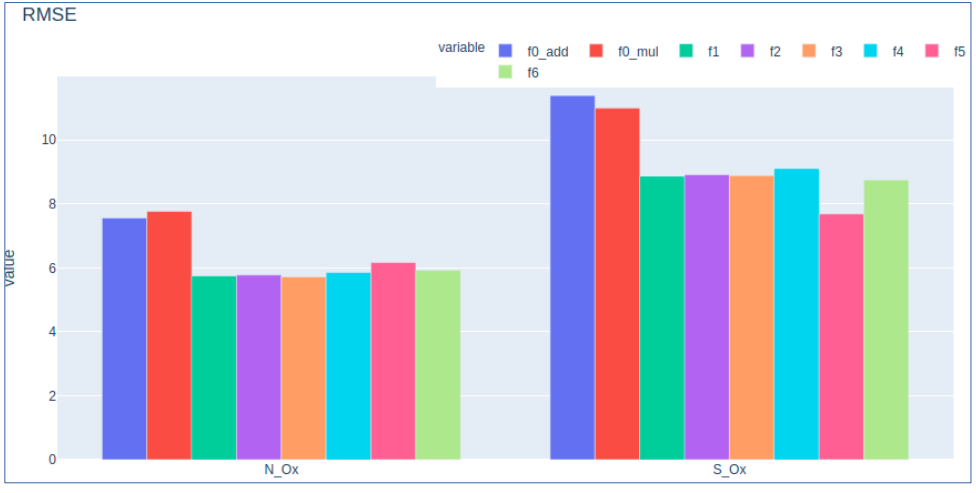





RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
f6
value
0
2
4
6
8
10
D_NO2
N_NO2
O_NO2
S_NO2
W_NO2


RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
f6
value
0
2
4
6
8
10
12
D_PM10K
N_PM10K
O_PM10K
S_PM10K
W_PM10K


RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
f6
value
0
2
4
6
8
10
N_O3
S_O3


RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
f6
value
0
2
4
6
8
10
N_Ox
S_Ox

Summary

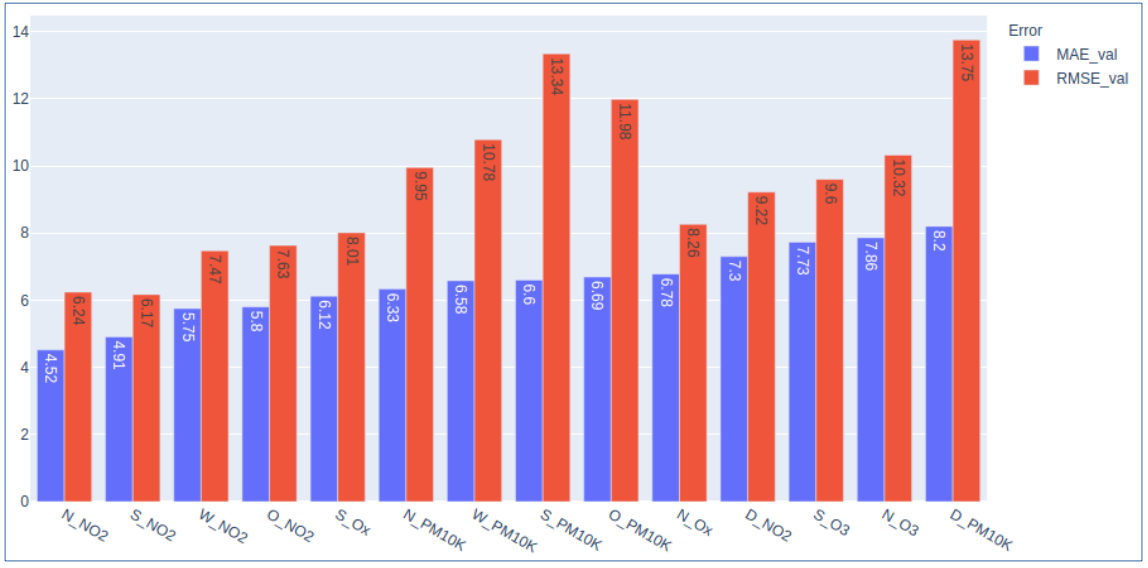
- Depend on test data → different methods apply

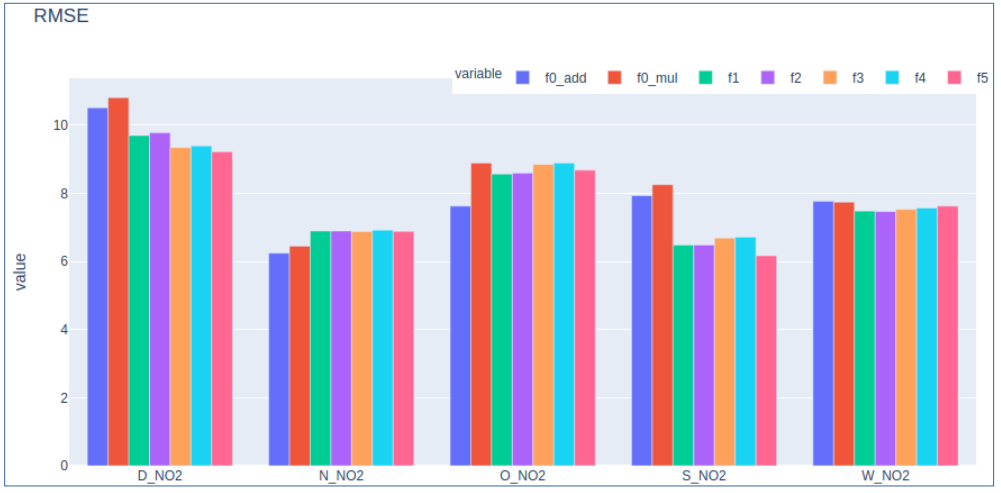
Next Steps

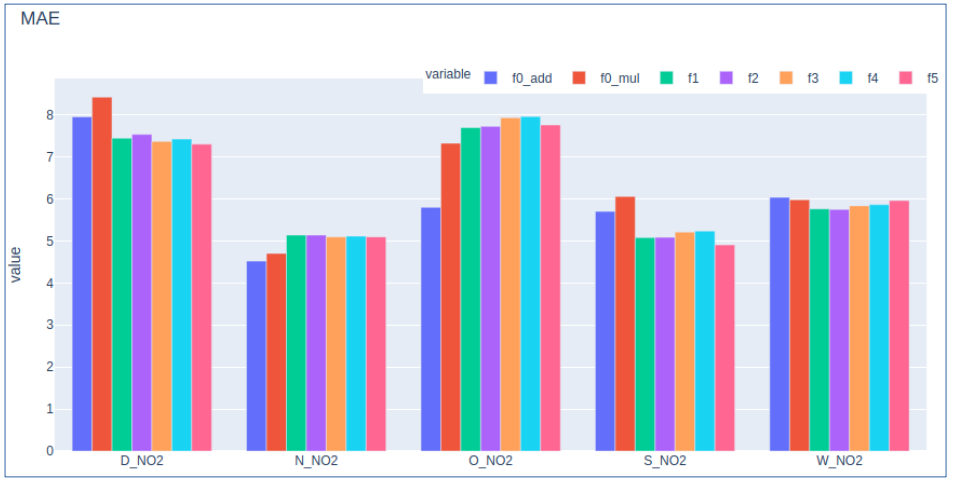
1. Cutoffs
2. Optimization for each area and factor
    1. Cross “features”
    2. PM10: remove (features + features2)
3. Lockdown 2, 3?

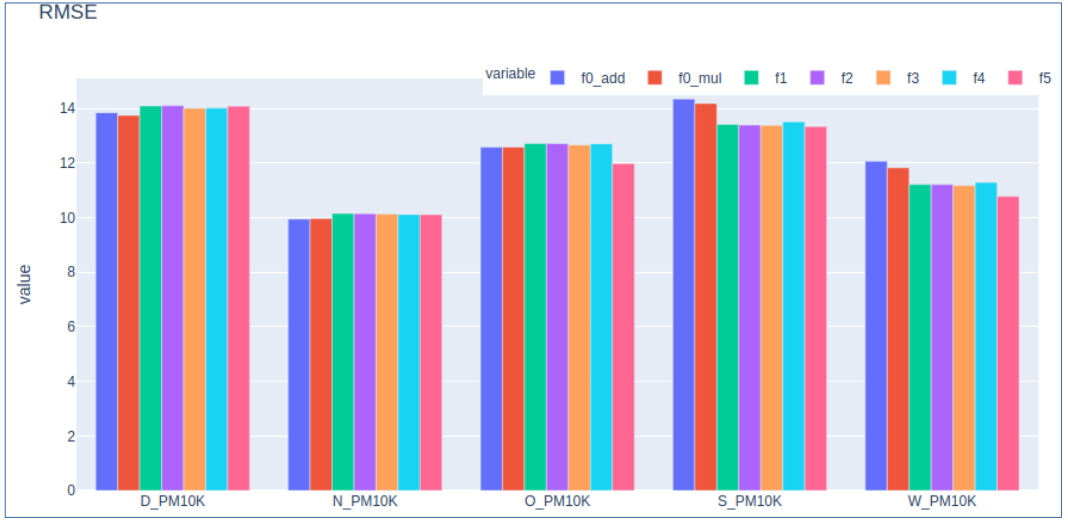
Thank you!

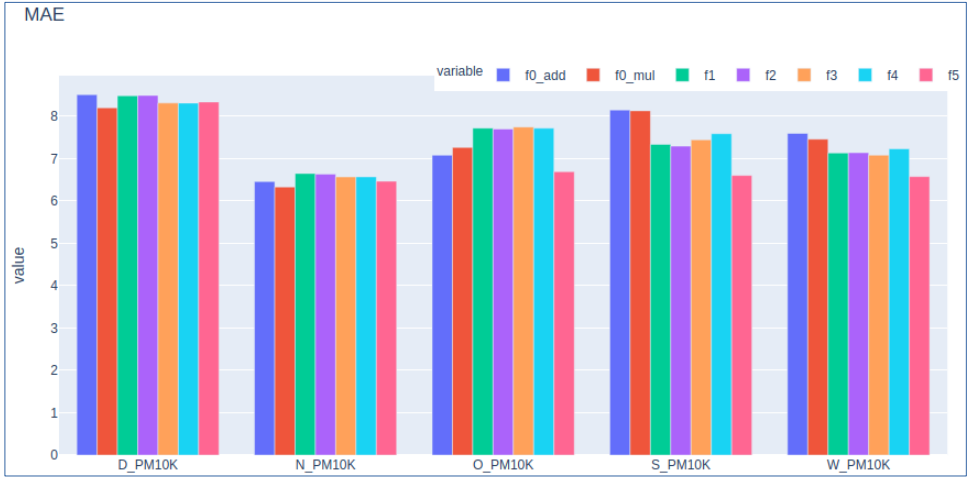
Apendix – with 26-30 March

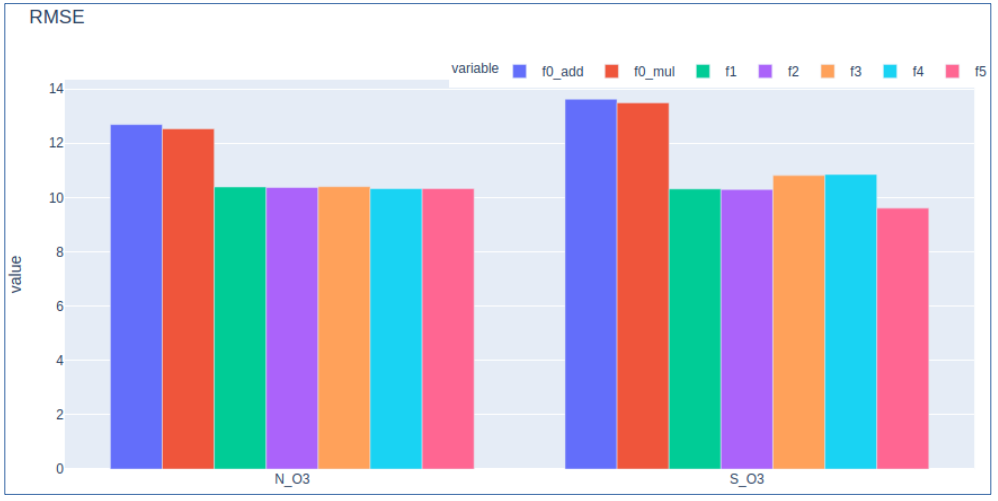



RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
value
0
2
4
6
8
10
D_NO2
N_NO2
O_NO2
S_NO2
W_NO2

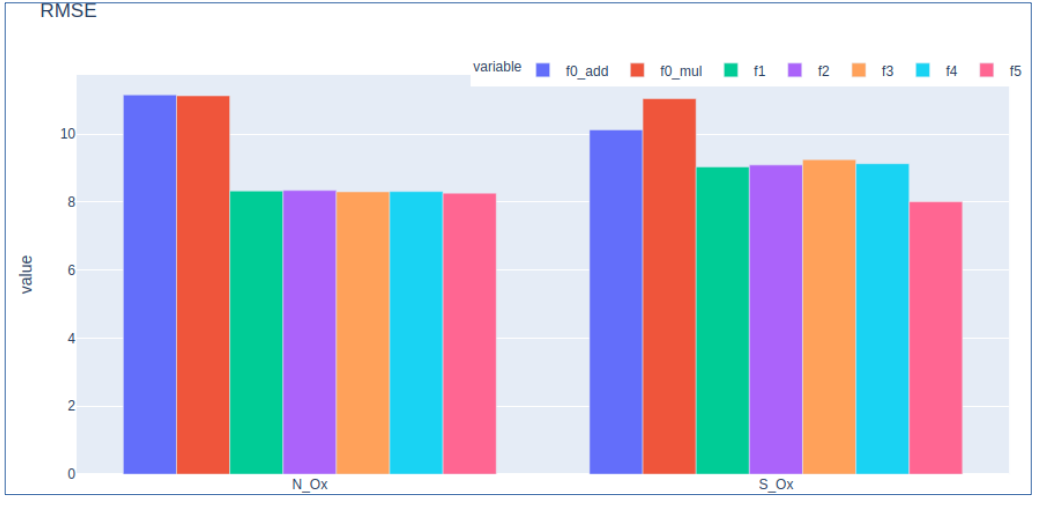

MAE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
value
0
1
2
3
4
5
6
7
8
D_NO2
N_NO2
O_NO2
S_NO2
W_NO2


RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
value
0
2
4
6
8
10
12
14
D_PM10K
N_PM10K
O_PM10K
S_PM10K
W_PM10K


MAE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
value
0
1
2
3
4
5
6
7
8
D_PM10K
N_PM10K
O_PM10K
S_PM10K
W_PM10K


RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
value
14
12
10
8
6
4
2
0
N_O3
S_O3


MAE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
value
10
8
6
4
2
0
N_O3
S_O3


RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
value
10
8
6
4
2
0
N_Ox
S_Ox


MAE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
value
8
6
4
2
0
N_Ox
S_Ox

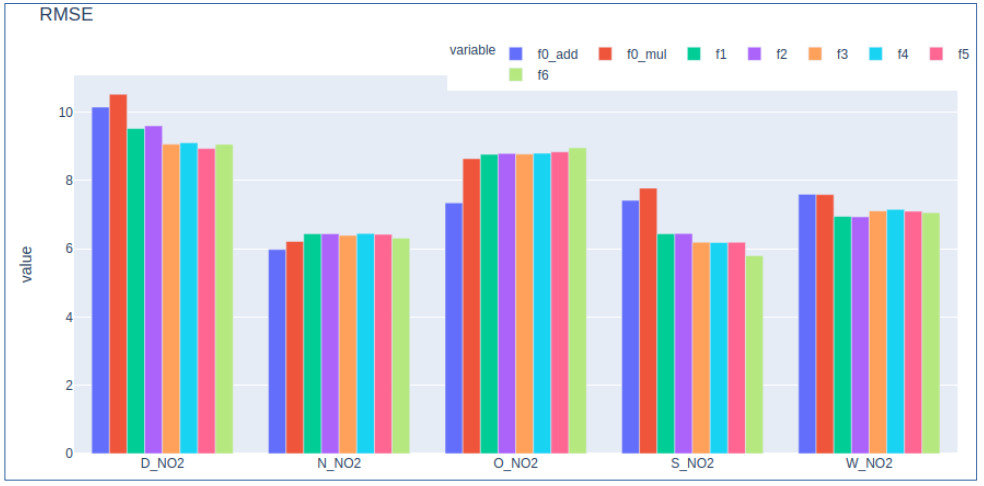
f6

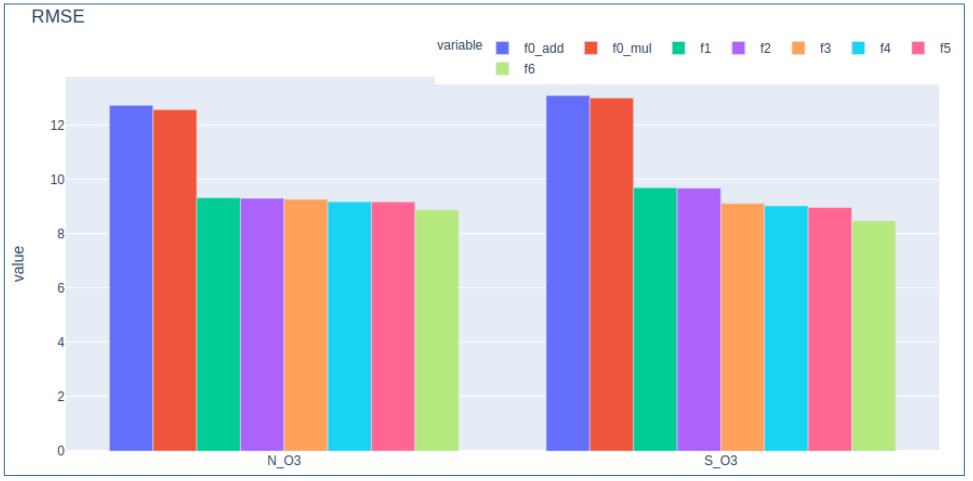
	RMSE	MAE	RMSE_val	MAE_val
S_PM10K	f5	f5	5.549522	4.242137
O_PM10K	f5	f5	5.758059	4.593578
S_NO2	f6	f6	5.793037	4.606080
N_PM10K	f5	f5	5.835789	4.875501
N_NO2	f0_add	f0_add	5.973273	4.369344
W_PM10K	f5	f5	6.108925	4.909600
W_NO2	f2	f2	6.929287	5.217883
D_PM10K	f0_mul	f0_mul	7.329732	6.039981
O_NO2	f0_add	f0_add	7.337199	5.575182
S_Ox	f6	f5	7.496641	5.755069
N_Ox	f6	f6	7.958320	6.588216
S_O3	f6	f6	8.476634	6.719293
N_O3	f6	f6	8.884625	7.189059
D_NO2	f5	f6	8.929430	6.949467

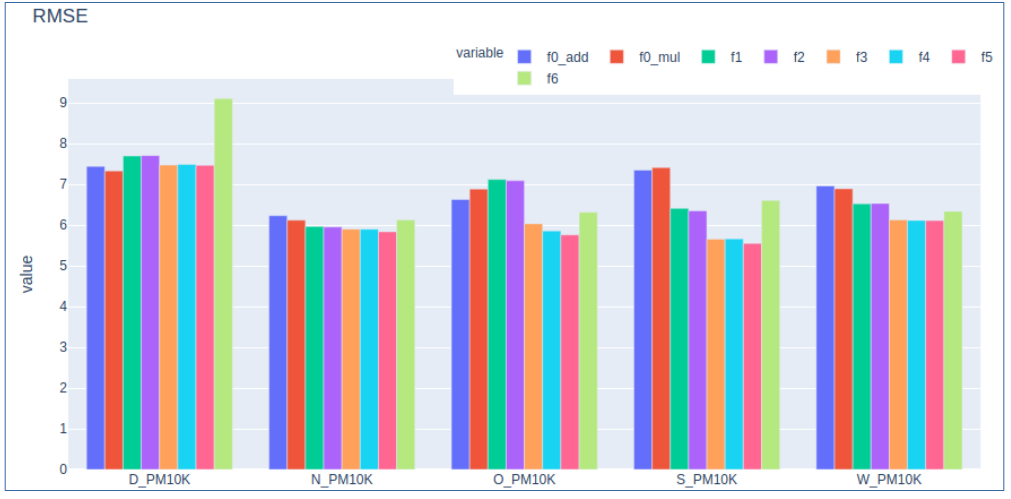
f5

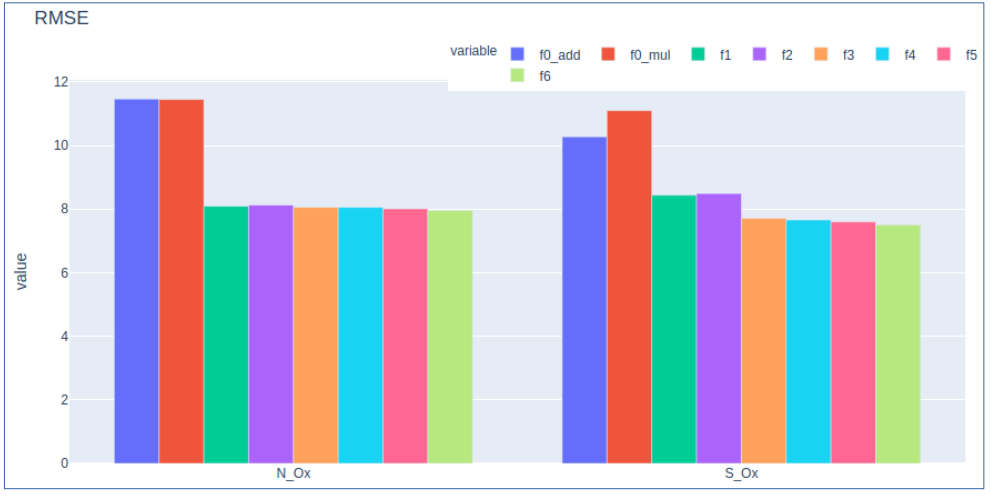
	RMSE	MAE	RMSE_val	MAE_val
S_PM10K	f5	f5	5.549522	4.242137
O_PM10K	f5	f5	5.758059	4.593578
N_PM10K	f5	f5	5.835789	4.875501
N_NO2	f0_add	f0_add	5.973273	4.369344
W_PM10K	f5	f5	6.108925	4.909600
S_NO2	f4	f3	6.180596	4.827288
W_NO2	f2	f2	6.929287	5.217883
D_PM10K	f0_mul	f0_mul	7.329732	6.039981
O_NO2	f0_add	f0_add	7.337199	5.575182
S_Ox	f5	f5	7.595386	5.755069
N_Ox	f5	f5	8.005494	6.615279
D_NO2	f5	f5	8.929430	7.012986
S_O3	f5	f5	8.967940	7.305033
N_O3	f5	f2	9.169308	7.223101

f6


RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
f6
value
D_NO2
N_NO2
O_NO2
S_NO2
W_NO2


RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
f6
value
N_O3
S_O3


RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
f6
value
D_PM10K
N_PM10K
O_PM10K
S_PM10K
W_PM10K


RMSE
variable
f0_add
f0_mul
f1
f2
f3
f4
f5
f6
value
N_Ox
S_Ox