Random Forest Parameter Selection Report - 西条

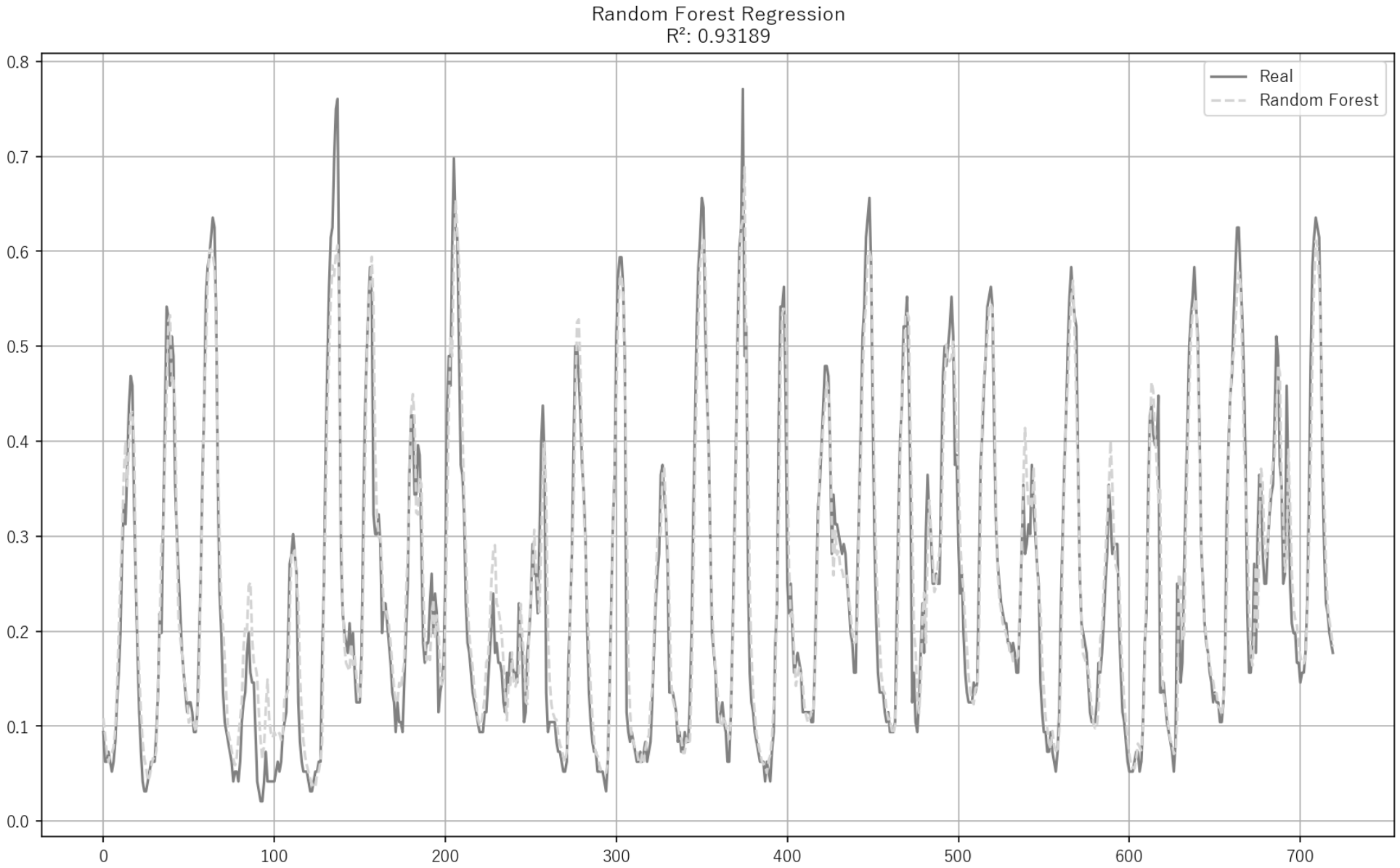
Prefecture code	38
Station code	38206050
Station name	西条
Target item	Ox(ppm)
Model	RandomForestRegressor
Number of features used	140
Train samples	18002
Validation samples	4501
Parameter Grid n_estimators	[900, 1000, 1100]
Parameter Grid max_depth	[10, 20, 30]
Parameter Grid min_samples_split	[2, 3, 4]
Parameter Grid min_samples_leaf	[1, 2]
Parameter Grid max_features	['sqrt', 'log2']
Best Parameter max_depth	20
Best Parameter max_features	sqrt
Best Parameter min_samples_leaf	1
Best Parameter min_samples_split	3
Best Parameter n_estimators	1000
R^2	0.9318887790580441
MAE	0.040026505942976874
RMSE	0.055391164118419316
Residuals Prob(Q)	3.375323248535871e-228
Residuals skew	0.382
Residuals kurtosis	2.365

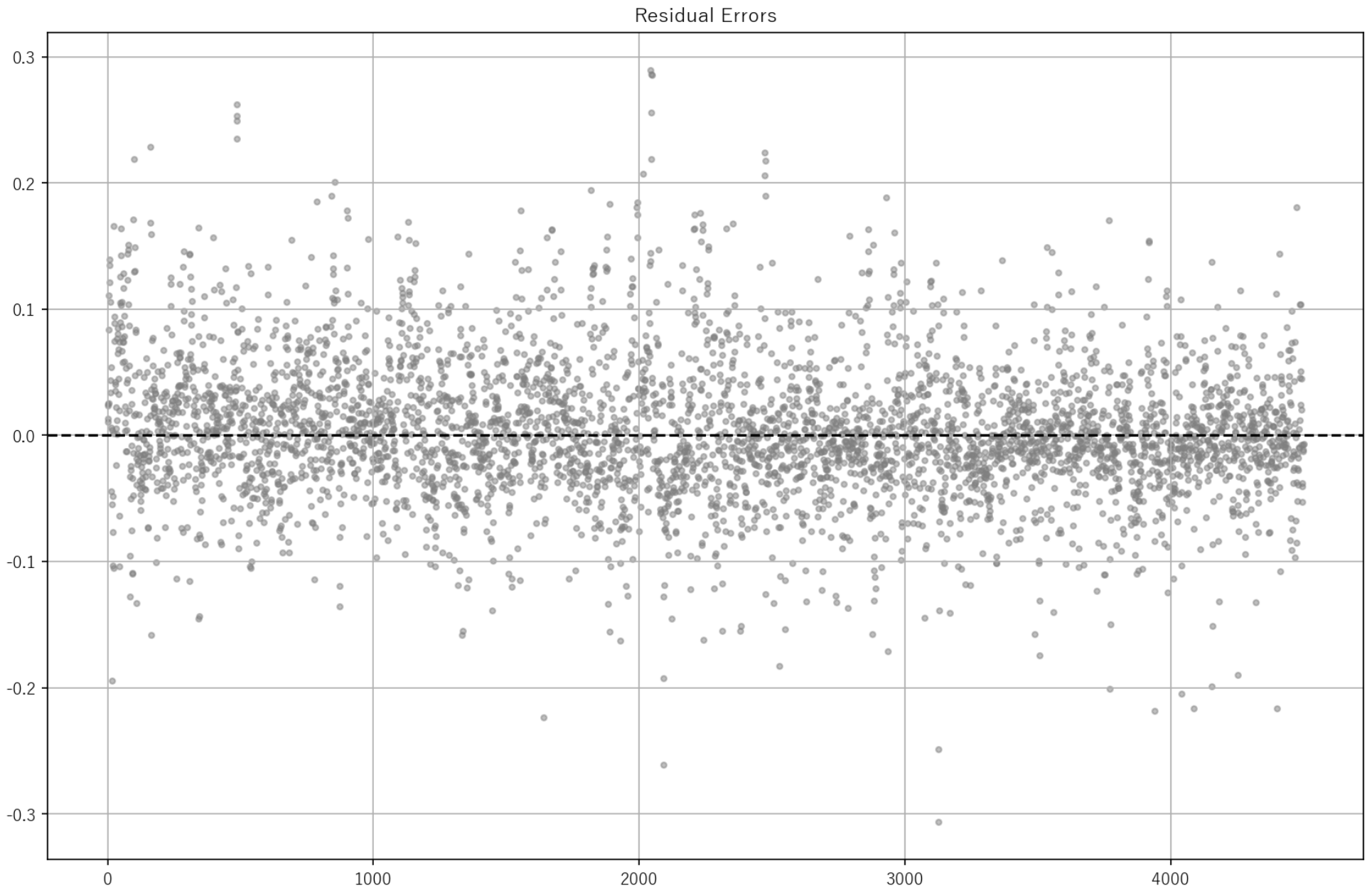
Features used for prediction

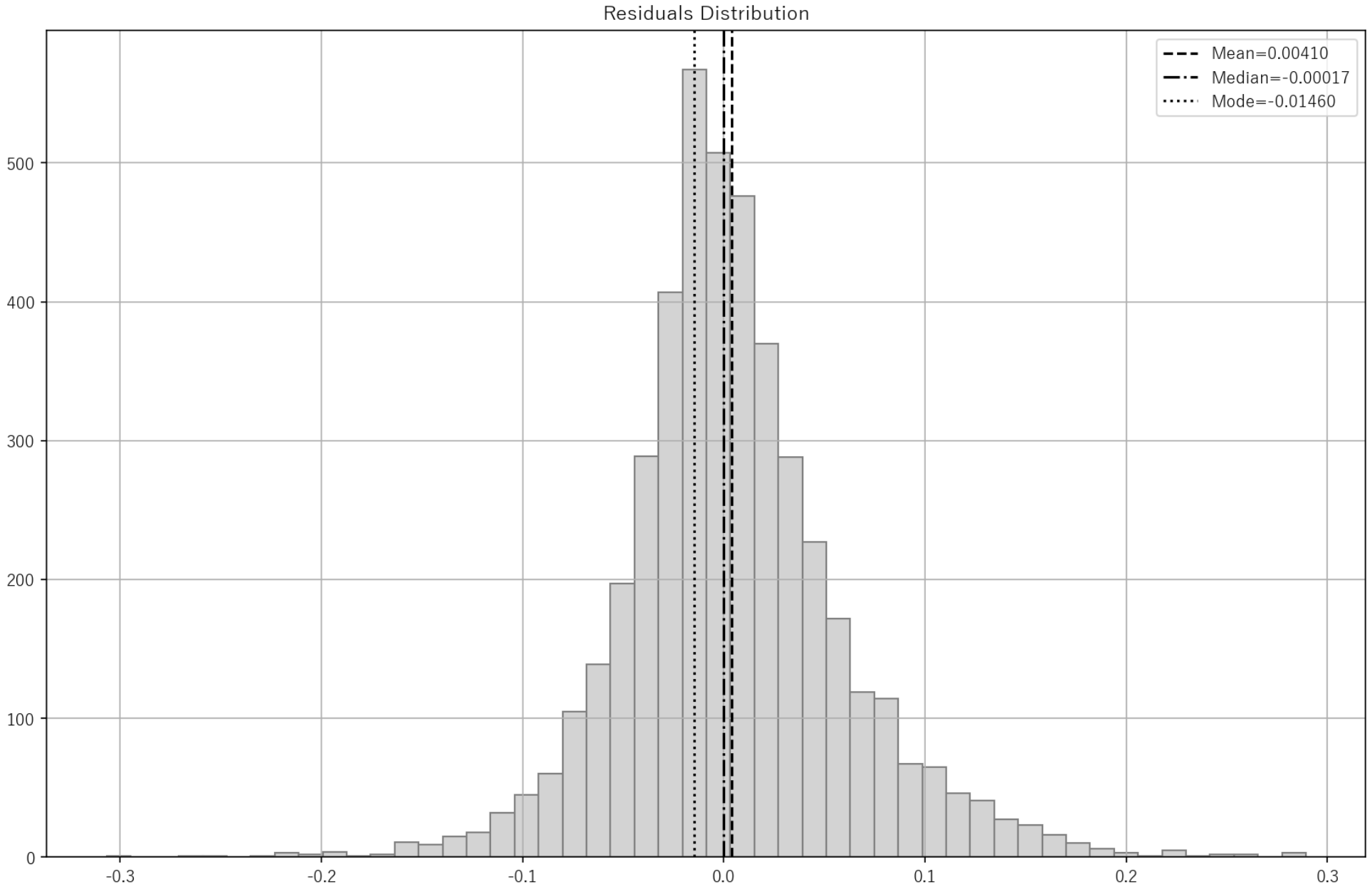
NO(ppm)	NO2(ppm)	U	V	Ox(ppm)_lag1
Ox(ppm)_lag2	Ox(ppm)_lag3	Ox(ppm)_lag4	Ox(ppm)_lag5	Ox(ppm)_lag6
Ox(ppm)_lag7	Ox(ppm)_lag8	Ox(ppm)_lag9	Ox(ppm)_lag10	Ox(ppm)_lag11
Ox(ppm)_lag12	Ox(ppm)_lag13	Ox(ppm)_lag14	Ox(ppm)_lag15	Ox(ppm)_lag16
Ox(ppm)_lag17	Ox(ppm)_lag18	Ox(ppm)_lag19	Ox(ppm)_lag20	Ox(ppm)_lag21
Ox(ppm)_lag22	Ox(ppm)_lag23	NO(ppm)_lag1	NO(ppm)_lag2	NO(ppm)_lag3
NO(ppm)_lag4	NO(ppm)_lag5	NO(ppm)_lag6	NO(ppm)_lag7	NO(ppm)_lag8
NO(ppm)_lag9	NO(ppm)_lag10	NO(ppm)_lag11	NO(ppm)_lag12	NO(ppm)_lag13
NO(ppm)_lag14	NO(ppm)_lag15	NO(ppm)_lag16	NO(ppm)_lag17	NO(ppm)_lag18
NO(ppm)_lag19	NO(ppm)_lag20	NO(ppm)_lag21	NO(ppm)_lag22	NO(ppm)_lag23
NO2(ppm)_lag1	NO2(ppm)_lag2	NO2(ppm)_lag3	NO2(ppm)_lag4	NO2(ppm)_lag5
NO2(ppm)_lag6	NO2(ppm)_lag7	NO2(ppm)_lag8	NO2(ppm)_lag9	NO2(ppm)_lag10
NO2(ppm)_lag11	NO2(ppm)_lag12	NO2(ppm)_lag13	NO2(ppm)_lag14	NO2(ppm)_lag15
NO2(ppm)_lag16	NO2(ppm)_lag17	NO2(ppm)_lag18	NO2(ppm)_lag19	NO2(ppm)_lag20
NO2(ppm)_lag21	NO2(ppm)_lag22	NO2(ppm)_lag23	U_lag1	U_lag2
U_lag3	U_lag4	U_lag5	U_lag6	U_lag7
U_lag8	U_lag9	U_lag10	U_lag11	U_lag12
U_lag13	U_lag14	U_lag15	U_lag16	U_lag17
U_lag18	U_lag19	U_lag20	U_lag21	U_lag22
U_lag23	V_lag1	V_lag2	V_lag3	V_lag4
V_lag5	V_lag6	V_lag7	V_lag8	V_lag9
V_lag10	V_lag11	V_lag12	V_lag13	V_lag14
V_lag15	V_lag16	V_lag17	V_lag18	V_lag19
V_lag20	V_lag21	V_lag22	V_lag23	Ox(ppm)_roll_mean_3
Ox(ppm)_roll_std_6	NO(ppm)_roll_mean_3	NO(ppm)_roll_std_6	NO2(ppm)_roll_mean_3	NO2(ppm)_roll_std_6
U_roll_mean_3	U_roll_std_6	V_roll_mean_3	V_roll_std_6	Ox(ppm)_diff_1
Ox(ppm)_diff_2	Ox(ppm)_diff_3	NO(ppm)_diff_3	NO2(ppm)_diff_3	U_diff_3
V_diff_3	hour_sin	hour_cos	dayofweek	is_weekend

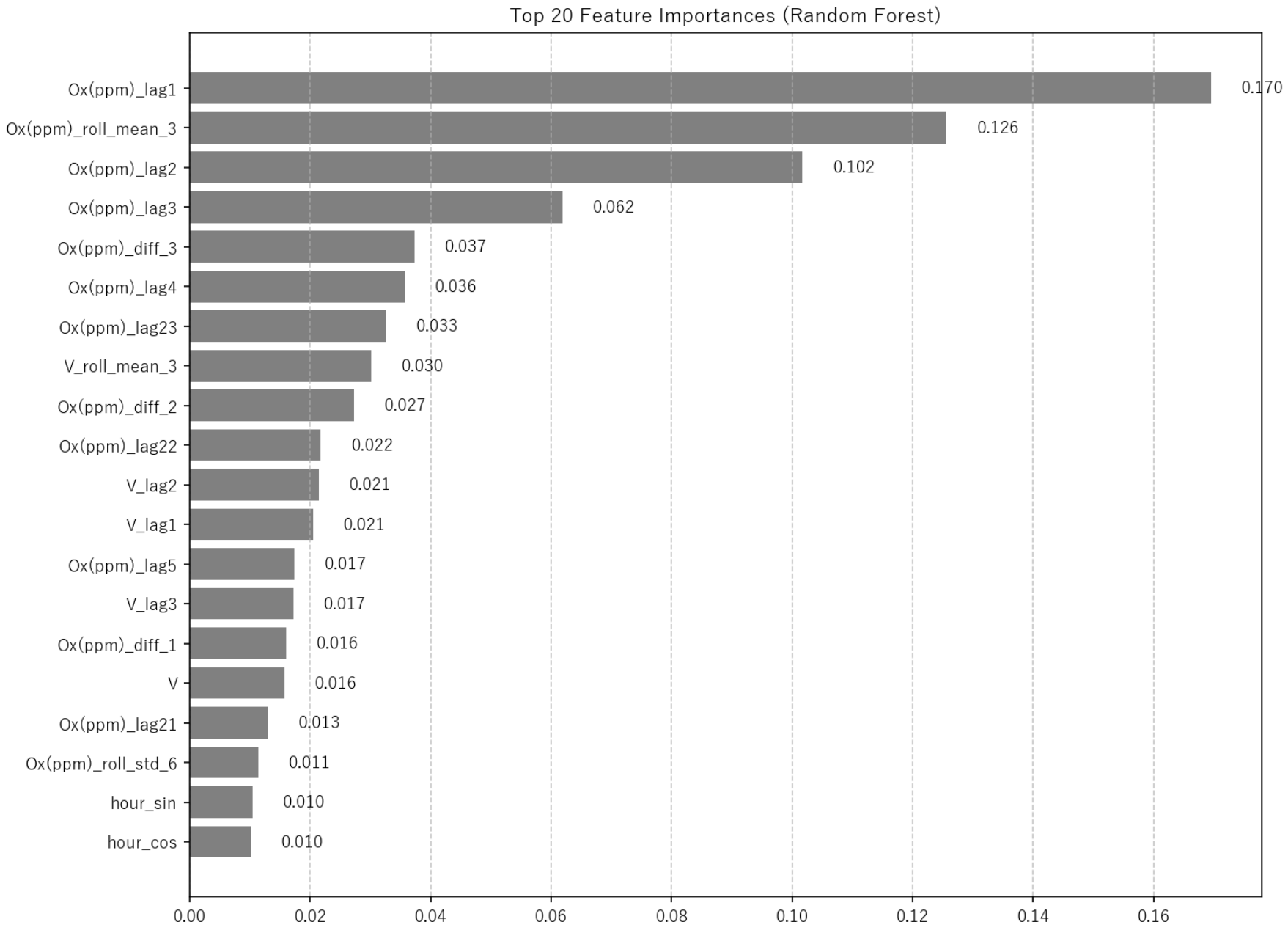
Model accuracy

Target	R^2	MAE	RMSE
Ox(ppm)	0.9319	0.0400	0.0554


Random Forest Regression
R²: 0.93189
Real
Random Forest
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
100
200
300
400
500
600
700


Residual Errors
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
0
1000
2000
3000
4000


Residuals Distribution
Mean=0.00410
Median=-0.00017
Mode=-0.01460
500
400
300
200
100
0
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3


Top 20 Feature Importances (Random Forest)
Ox(ppm)_lag1
Ox(ppm)_roll_mean_3
Ox(ppm)_lag2
Ox(ppm)_lag3
Ox(ppm)_diff_3
Ox(ppm)_lag4
Ox(ppm)_lag23
V_roll_mean_3
Ox(ppm)_diff_2
Ox(ppm)_lag22
V_lag2
V_lag1
Ox(ppm)_lag5
V_lag3
Ox(ppm)_diff_1
V
Ox(ppm)_lag21
Ox(ppm)_roll_std_6
hour_sin
hour_cos
0.170
0.126
0.102
0.062
0.037
0.036
0.033
0.030
0.027
0.022
0.021
0.021
0.017
0.017
0.016
0.016
0.013
0.011
0.010
0.010
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16