XGBoost forecasting Report - 西条

$$\hat{y}_i = \sum_{m=1}^{M} \eta\, f_m(X_i)$$
$$\mathcal{L} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{m=1}^{M} \Omega(f_m)$$

$\hat{y}_i$: predicted value for sample i

M: total number of trees (number of estimators, eg.:400)

η: learning rate (eg.:0.04), controlling the contribution of each tree

$f_m(x_i)$: output of the m-th regression tree with depth, eg.:¥leq 6

$\mathcal{L}$: total loss function minimized during training

$\Omega(f_m)$: regularization term penalizing tree complexity (depth, leaf weights)

Objective = minimize the total cost: prediction error + regularization penalty.

The XGBoost model builds an ensemble of decision trees,
where each tree tries to correct the residual errors made by the previous ones.
At each iteration, the model adds a new function f(x) that minimizes
a global objective function composed of two terms:
1. The **prediction loss** (in this case, the squared error between actual and predicted values).
2. A **regularization term** Ω(f) that penalizes complex trees to prevent overfitting.

The squared error loss is used as the primary cost to minimize,
while the regularization controls model complexity.
The optimization proceeds additively, building the model tree by tree, guided by the learning rate.

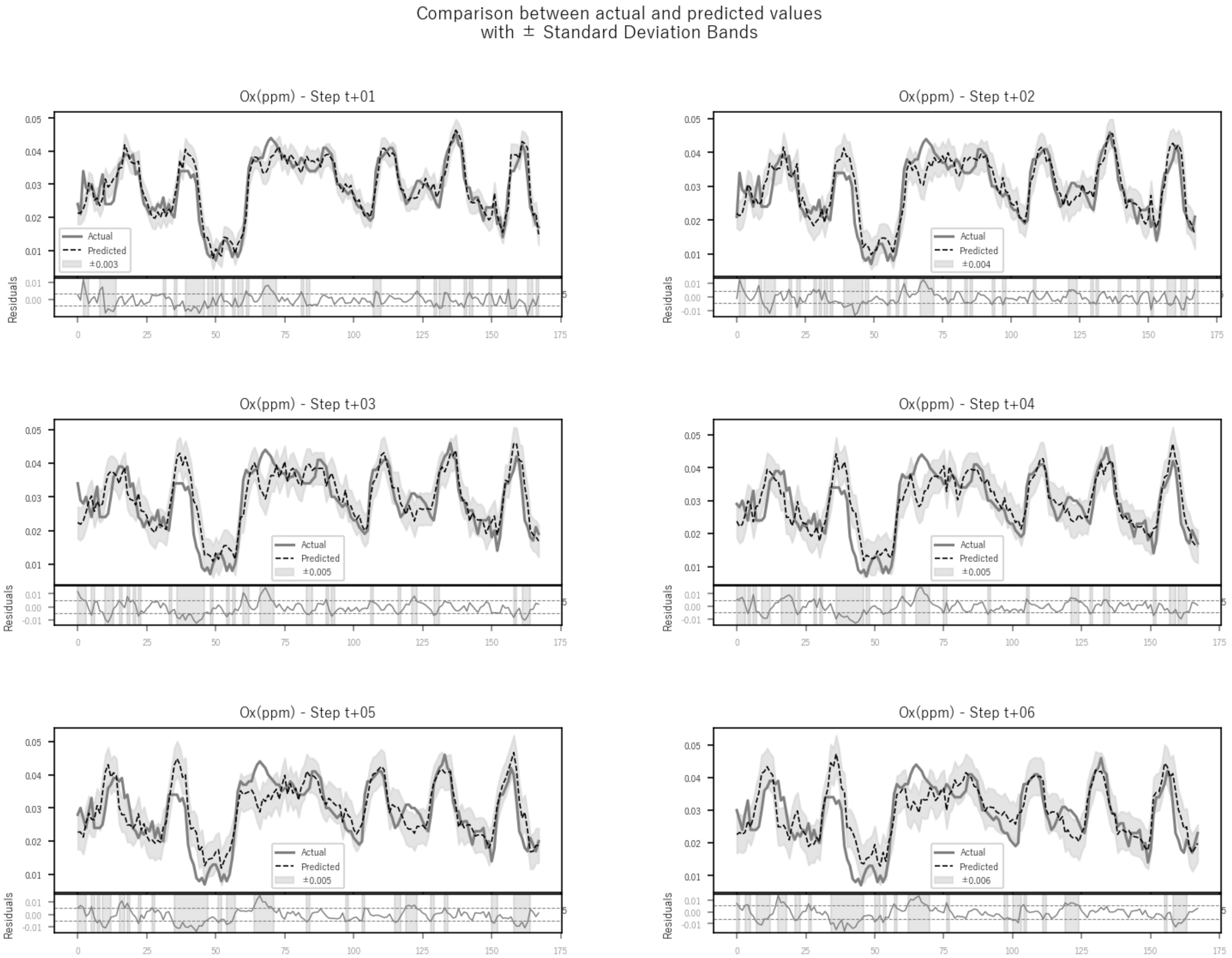
Prefecture code	38
Station code	38206050
Station name	西条
Target item	Ox(ppm)
Number of data points in the train set	15735
Number of data points in the test set	6744
Forecast horizon (hours)	24
Model	XGBoost
Objective	reg:squarederror
Booster	None
Number of estimators	1000
Learning rate	0.045
Elapsed time	3 min 9 sec
Number of features used	140

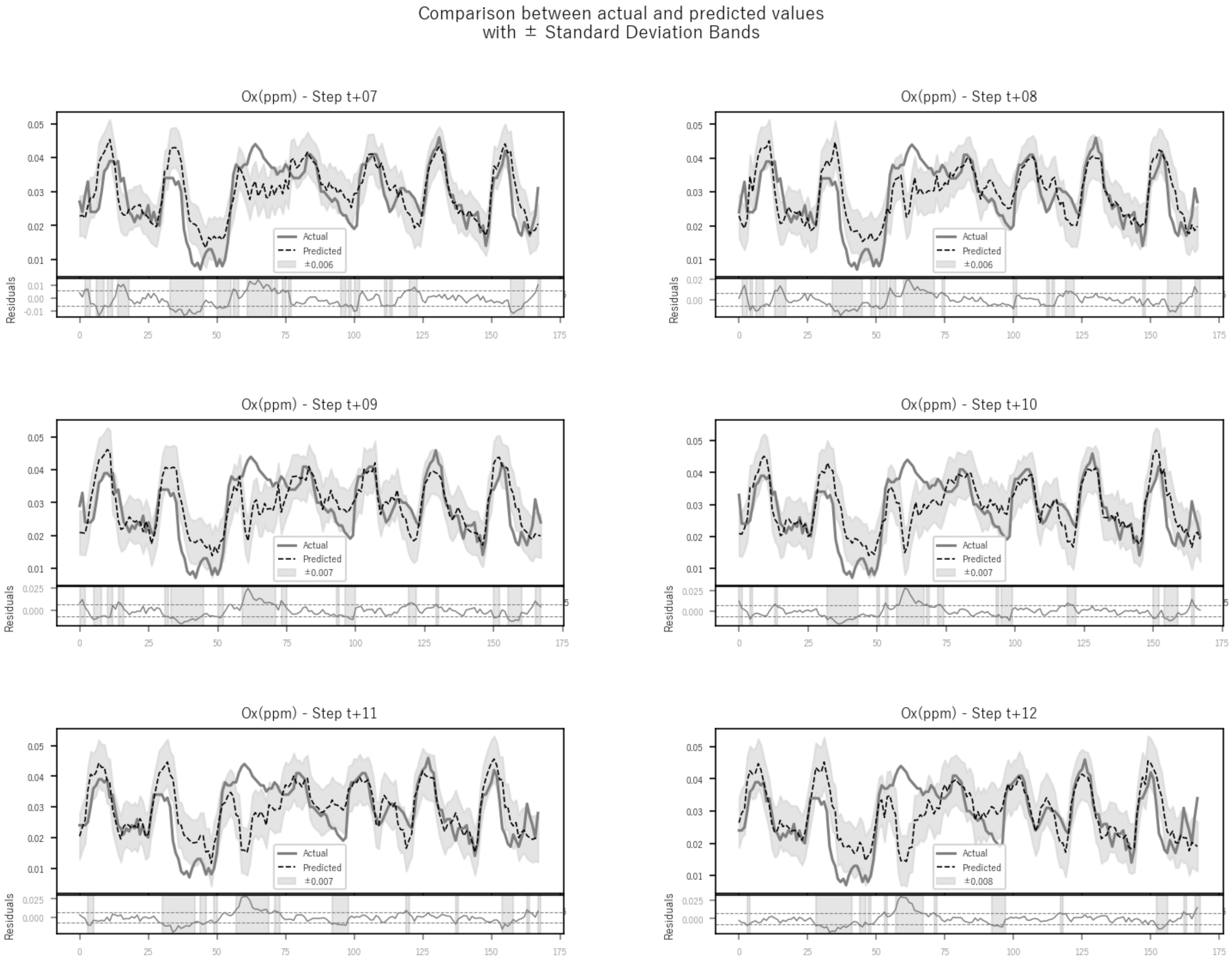
Features used for prediction

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

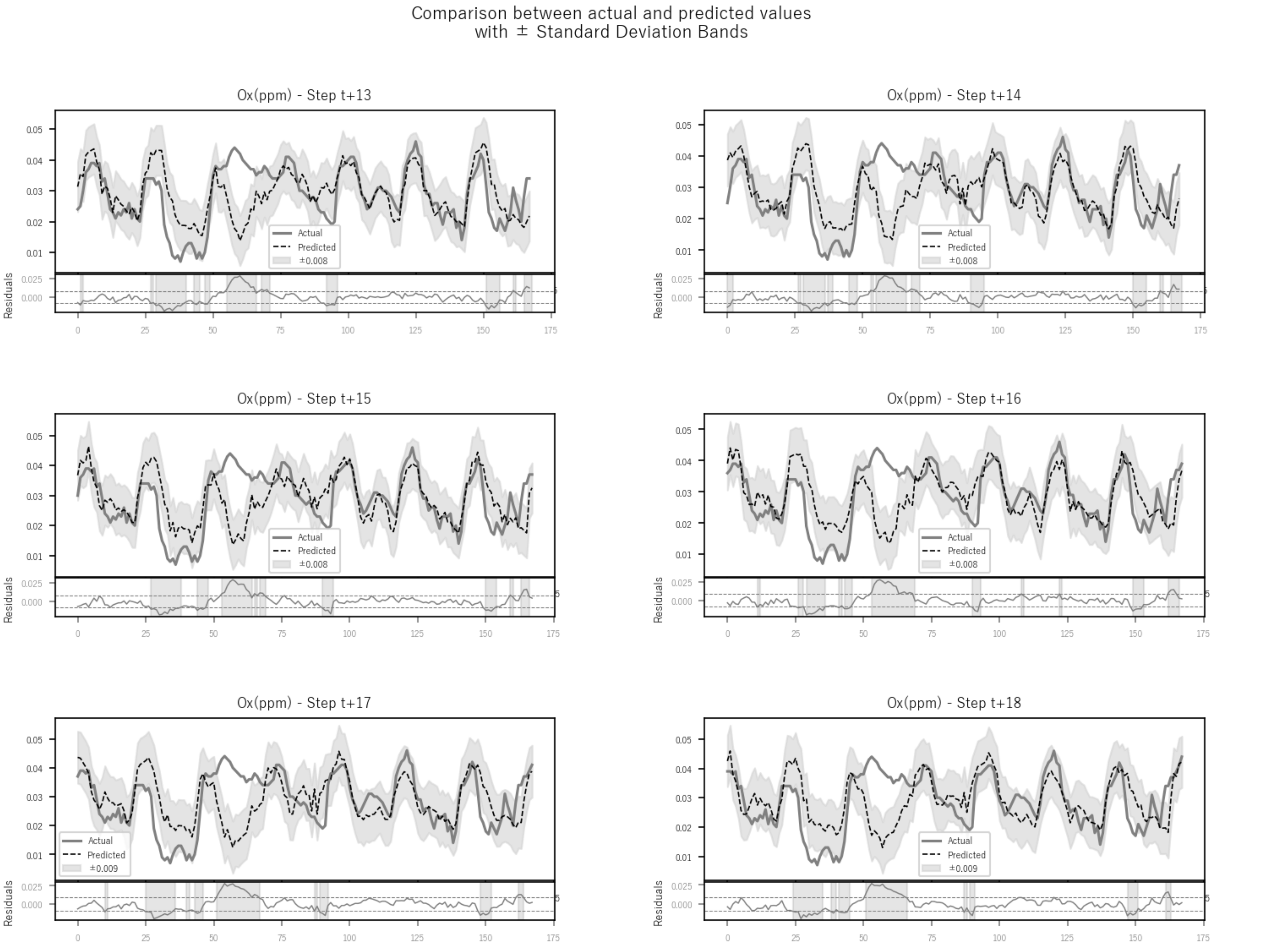
Model accuracy

Target	R^2	MAE	RMSE
Ox(ppm)_t+01	0.8945	0.0042	0.0059
Ox(ppm)_t+02	0.8404	0.0053	0.0073
Ox(ppm)_t+03	0.7878	0.0061	0.0084
Ox(ppm)_t+04	0.7428	0.0068	0.0093
Ox(ppm)_t+05	0.7054	0.0073	0.0099
Ox(ppm)_t+06	0.6745	0.0077	0.0104
Ox(ppm)_t+07	0.6531	0.0080	0.0107
Ox(ppm)_t+08	0.6269	0.0083	0.0111
Ox(ppm)_t+09	0.6046	0.0085	0.0115
Ox(ppm)_t+10	0.5954	0.0086	0.0116
Ox(ppm)_t+11	0.5816	0.0088	0.0118
Ox(ppm)_t+12	0.5669	0.0090	0.0120
Ox(ppm)_t+13	0.5576	0.0091	0.0121
Ox(ppm)_t+14	0.5422	0.0092	0.0124
Ox(ppm)_t+15	0.5283	0.0093	0.0125
Ox(ppm)_t+16	0.5344	0.0093	0.0125
Ox(ppm)_t+17	0.5302	0.0094	0.0125
Ox(ppm)_t+18	0.5137	0.0095	0.0127
Ox(ppm)_t+19	0.5159	0.0095	0.0127
Ox(ppm)_t+20	0.5007	0.0095	0.0129
Ox(ppm)_t+21	0.4915	0.0096	0.0130
Ox(ppm)_t+22	0.5033	0.0096	0.0129
Ox(ppm)_t+23	0.4944	0.0097	0.0130
Ox(ppm)_t+24	0.4926	0.0098	0.0130

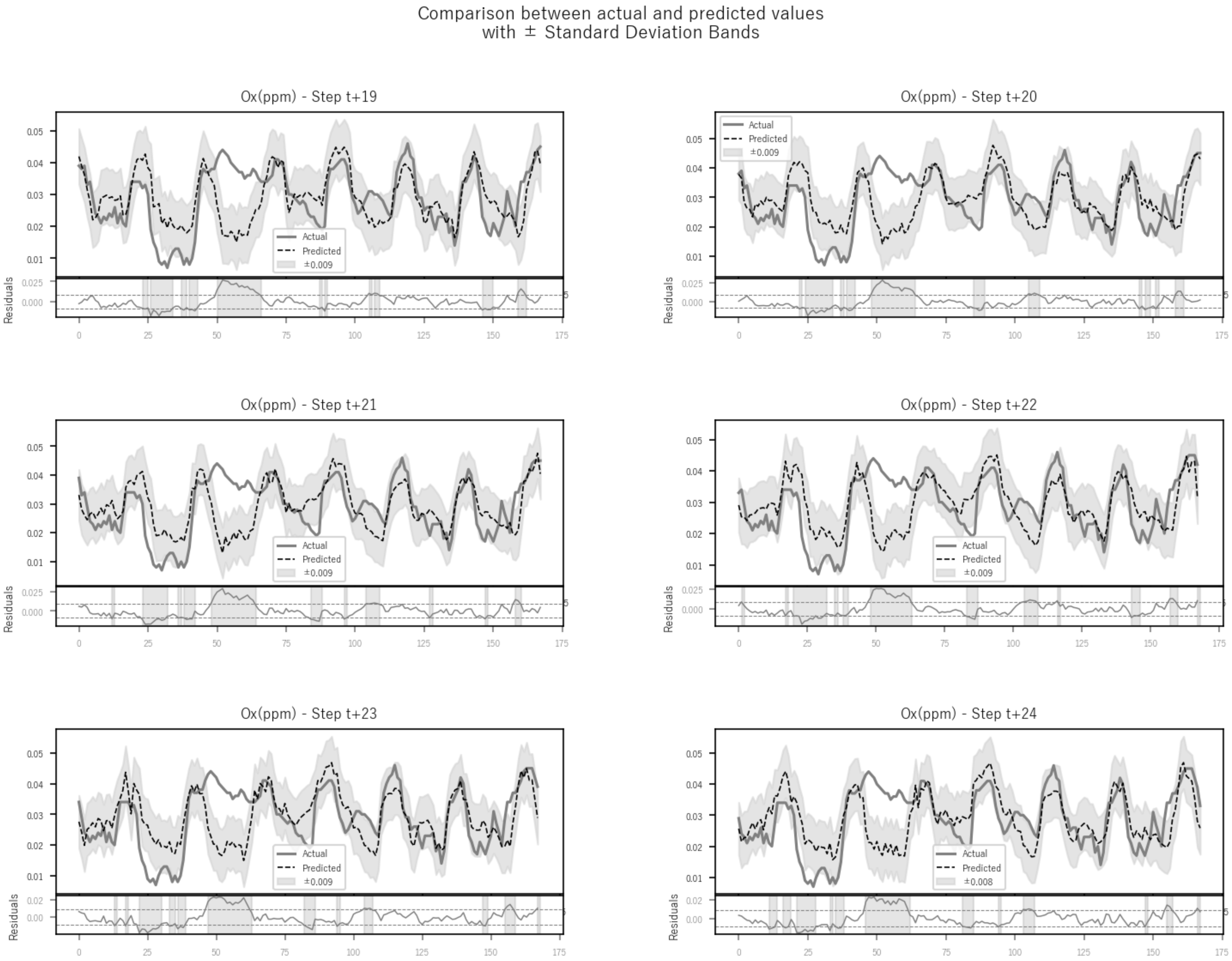

Comparison between actual and predicted values
with ± Standard Deviation Bands
Ox(ppm) - Step t+01
Actual
Predicted
±0.003
Ox(ppm) - Step t+02
Actual
Predicted
±0.004
Ox(ppm) - Step t+03
Actual
Predicted
±0.005
Ox(ppm) - Step t+04
Actual
Predicted
±0.005
Ox(ppm) - Step t+05
Actual
Predicted
±0.005
Ox(ppm) - Step t+06
Actual
Predicted
±0.006
Residuals

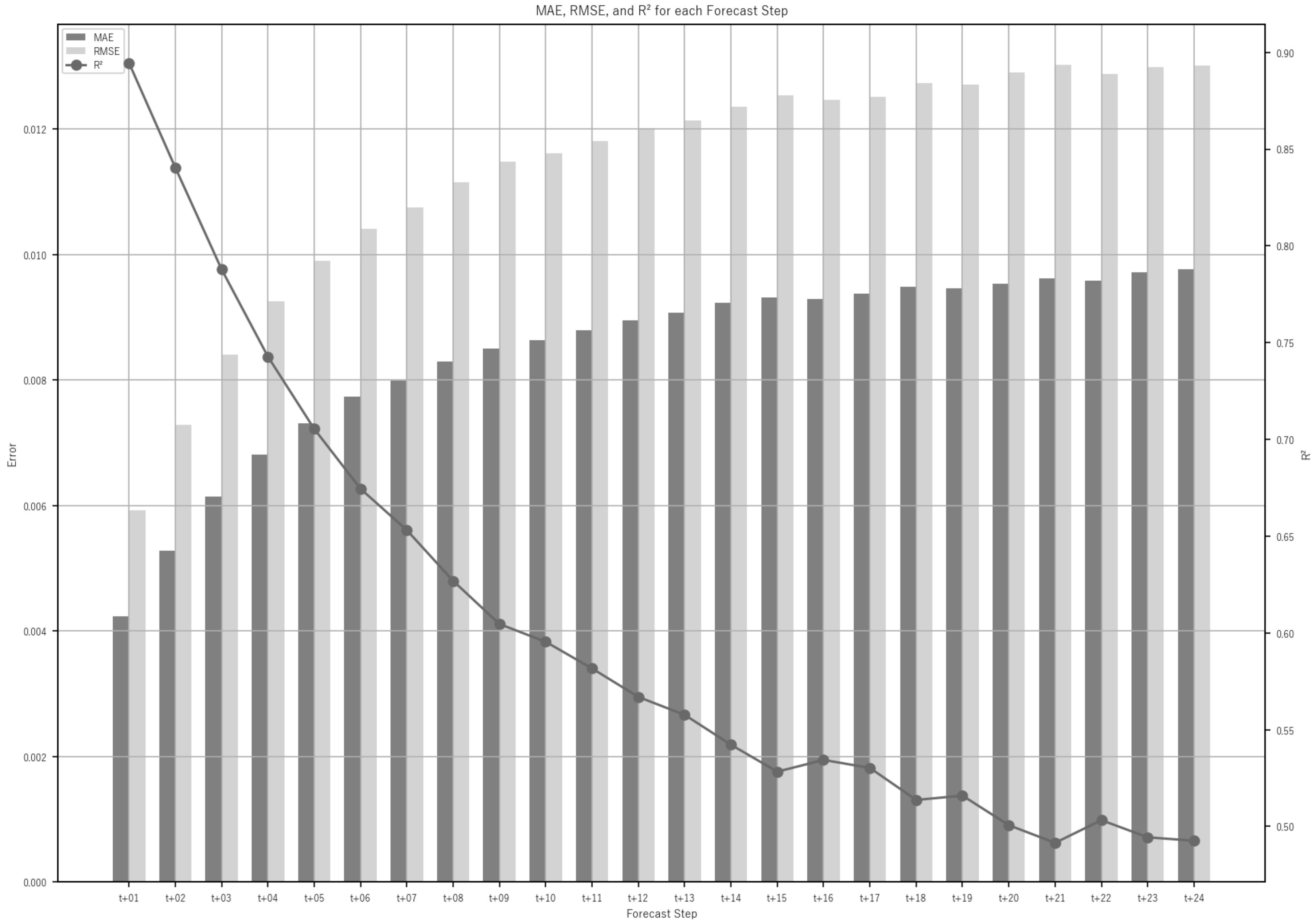


Comparison between actual and predicted values
with ± Standard Deviation Bands
Ox(ppm) - Step t+07
Ox(ppm) - Step t+08
Ox(ppm) - Step t+09
Ox(ppm) - Step t+10
Ox(ppm) - Step t+11
Ox(ppm) - Step t+12
Actual
Predicted
±0.006
±0.006
±0.007
±0.007
±0.007
±0.008
Residuals

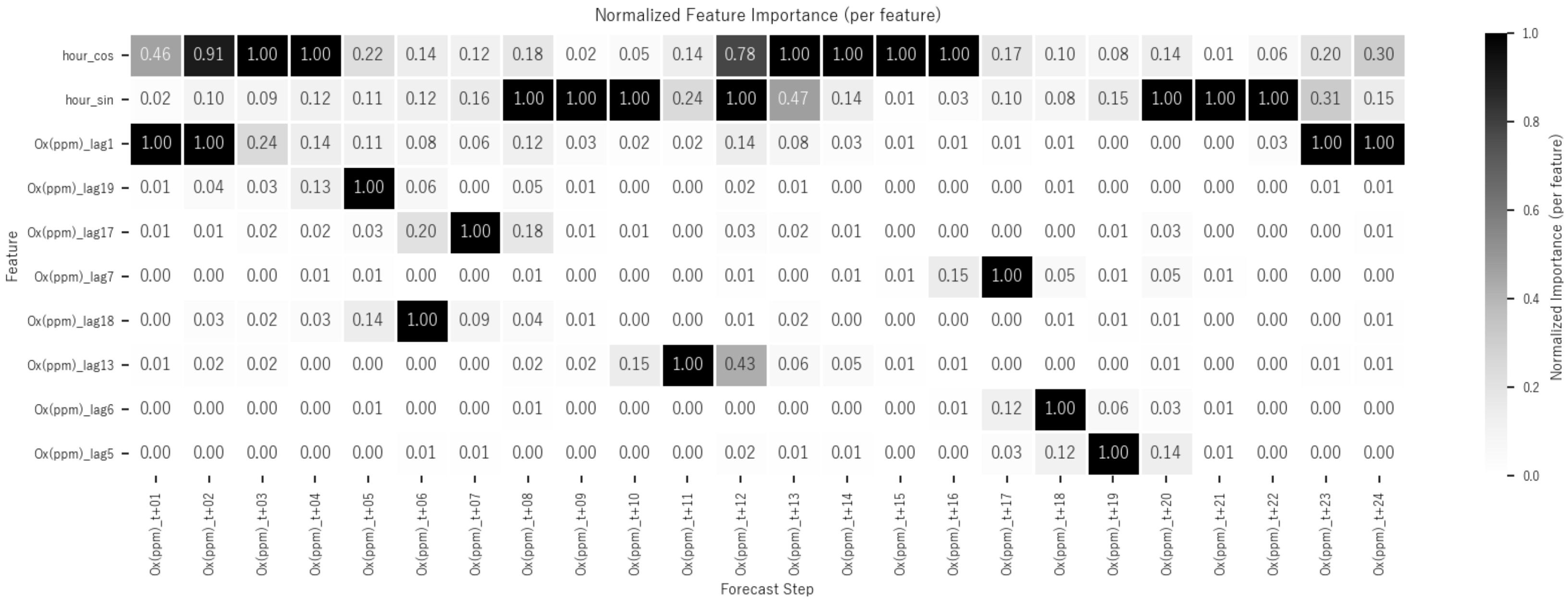


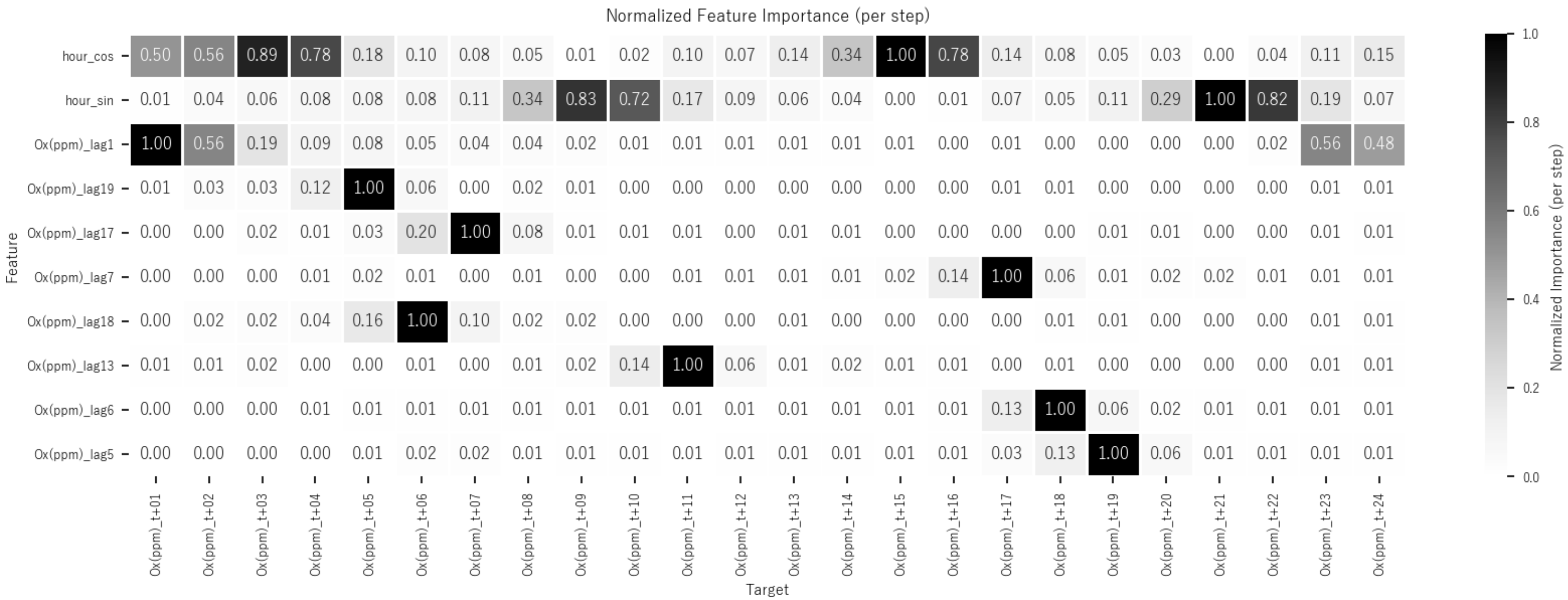
Comparison between actual and predicted values
with ± Standard Deviation Bands
Ox(ppm) - Step t+13
Ox(ppm) - Step t+14
Ox(ppm) - Step t+15
Ox(ppm) - Step t+16
Ox(ppm) - Step t+17
Ox(ppm) - Step t+18
Actual
Predicted
±0.008
±0.009
Residuals


Comparison between actual and predicted values
with ± Standard Deviation Bands
Ox(ppm) - Step t+19
Ox(ppm) - Step t+20
Ox(ppm) - Step t+21
Ox(ppm) - Step t+22
Ox(ppm) - Step t+23
Ox(ppm) - Step t+24
Actual
Predicted
±0.009
±0.008
Residuals
0.05
0.04
0.03
0.02
0.01
0.025
0.000
0
25
50
75
100
125
150
175



MAE, RMSE, and R² for each Forecast Step
MAE
RMSE
R²
Error
Forecast Step
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
t+01
t+02
t+03
t+04
t+05
t+06
t+07
t+08
t+09
t+10
t+11
t+12
t+13
t+14
t+15
t+16
t+17
t+18
t+19
t+20
t+21
t+22
t+23
t+24


Normalized Feature Importance (per feature)
hour_cos 0.46 0.91 1.00 1.00 0.22 0.14 0.12 0.18 0.02 0.05 0.14 0.78 1.00 1.00 1.00 1.00 0.17 0.10 0.08 0.14 0.01 0.06 0.20 0.30
hour_sin 0.02 0.10 0.09 0.12 0.11 0.12 0.16 1.00 1.00 1.00 0.24 1.00 0.47 0.14 0.01 0.03 0.10 0.08 0.15 1.00 1.00 1.00 0.31 0.15
Ox(ppm)_lag1 1.00 1.00 0.24 0.14 0.11 0.08 0.06 0.12 0.03 0.02 0.02 0.14 0.08 0.03 0.01 0.01 0.01 0.01 0.00 0.00 0.00 0.03 1.00 1.00
Ox(ppm)_lag19 0.01 0.04 0.03 0.13 1.00 0.06 0.00 0.05 0.01 0.00 0.00 0.02 0.01 0.00 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00 0.01 0.01
Ox(ppm)_lag17 0.01 0.01 0.02 0.02 0.03 0.20 1.00 0.18 0.01 0.01 0.00 0.03 0.02 0.01 0.00 0.00 0.00 0.00 0.01 0.03 0.00 0.00 0.00 0.01
Ox(ppm)_lag7 0.00 0.00 0.00 0.01 0.01 0.00 0.00 0.01 0.00 0.00 0.00 0.01 0.00 0.01 0.01 0.15 1.00 0.05 0.01 0.05 0.01 0.00 0.00 0.00
Ox(ppm)_lag18 0.00 0.03 0.02 0.03 0.14 1.00 0.09 0.04 0.01 0.00 0.00 0.01 0.02 0.00 0.00 0.00 0.00 0.01 0.01 0.01 0.00 0.00 0.00 0.01
Ox(ppm)_lag13 0.01 0.02 0.02 0.00 0.00 0.00 0.00 0.02 0.02 0.15 1.00 0.43 0.06 0.05 0.01 0.01 0.00 0.00 0.00 0.01 0.00 0.00 0.01 0.01
Ox(ppm)_lag6 0.00 0.00 0.00 0.00 0.01 0.00 0.00 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.12 1.00 0.06 0.03 0.01 0.00 0.00 0.00
Ox(ppm)_lag5 0.00 0.00 0.00 0.00 0.00 0.01 0.01 0.00 0.00 0.00 0.00 0.02 0.01 0.01 0.00 0.00 0.03 0.12 1.00 0.14 0.01 0.00 0.00 0.00
Feature
Ox(ppm)_t+01 Ox(ppm)_t+02 Ox(ppm)_t+03 Ox(ppm)_t+04 Ox(ppm)_t+05 Ox(ppm)_t+06 Ox(ppm)_t+07 Ox(ppm)_t+08 Ox(ppm)_t+09 Ox(ppm)_t+10 Ox(ppm)_t+11 Ox(ppm)_t+12 Ox(ppm)_t+13 Ox(ppm)_t+14 Ox(ppm)_t+15 Ox(ppm)_t+16 Ox(ppm)_t+17 Ox(ppm)_t+18 Ox(ppm)_t+19 Ox(ppm)_t+20 Ox(ppm)_t+21 Ox(ppm)_t+22 Ox(ppm)_t+23 Ox(ppm)_t+24
Forecast Step
Normalized Importance (per feature)
0.0 0.2 0.4 0.6 0.8 1.0


Normalized Feature Importance (per step)
hour_cos 0.50 0.56 0.89 0.78 0.18 0.10 0.08 0.05 0.01 0.02 0.10 0.07 0.14 0.34 1.00 0.78 0.14 0.08 0.05 0.03 0.00 0.04 0.11 0.15
hour_sin 0.01 0.04 0.06 0.08 0.08 0.08 0.11 0.34 0.83 0.72 0.17 0.09 0.06 0.04 0.00 0.01 0.07 0.05 0.11 0.29 1.00 0.82 0.19 0.07
Ox(ppm)_lag1 1.00 0.56 0.19 0.09 0.08 0.05 0.04 0.04 0.02 0.01 0.01 0.01 0.01 0.01 0.01 0.00 0.01 0.00 0.00 0.00 0.00 0.02 0.56 0.48
Ox(ppm)_lag19 0.01 0.03 0.03 0.12 1.00 0.06 0.00 0.02 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.01 0.00 0.00 0.00 0.00 0.01 0.01
Ox(ppm)_lag17 0.00 0.00 0.02 0.01 0.03 0.20 1.00 0.08 0.01 0.01 0.01 0.00 0.01 0.01 0.00 0.00 0.00 0.00 0.01 0.01 0.00 0.00 0.01 0.01
Ox(ppm)_lag7 0.00 0.00 0.00 0.01 0.02 0.01 0.00 0.01 0.00 0.00 0.01 0.01 0.01 0.01 0.02 0.14 1.00 0.06 0.01 0.02 0.02 0.01 0.01 0.01
Ox(ppm)_lag18 0.00 0.02 0.02 0.04 0.16 1.00 0.10 0.02 0.02 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00 0.01 0.01 0.00 0.00 0.00 0.01 0.01
Ox(ppm)_lag13 0.01 0.01 0.02 0.00 0.00 0.01 0.00 0.01 0.02 0.14 1.00 0.06 0.01 0.02 0.01 0.01 0.00 0.01 0.00 0.00 0.00 0.00 0.01 0.01
Ox(ppm)_lag6 0.00 0.00 0.00 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.13 1.00 0.06 0.02 0.01 0.01 0.01 0.01
Ox(ppm)_lag5 0.00 0.00 0.00 0.00 0.01 0.02 0.02 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.03 0.13 1.00 0.06 0.01 0.01 0.01 0.01
Feature
Ox(ppm)_t+01 Ox(ppm)_t+02 Ox(ppm)_t+03 Ox(ppm)_t+04 Ox(ppm)_t+05 Ox(ppm)_t+06 Ox(ppm)_t+07 Ox(ppm)_t+08 Ox(ppm)_t+09 Ox(ppm)_t+10 Ox(ppm)_t+11 Ox(ppm)_t+12 Ox(ppm)_t+13 Ox(ppm)_t+14 Ox(ppm)_t+15 Ox(ppm)_t+16 Ox(ppm)_t+17 Ox(ppm)_t+18 Ox(ppm)_t+19 Ox(ppm)_t+20 Ox(ppm)_t+21 Ox(ppm)_t+22 Ox(ppm)_t+23 Ox(ppm)_t+24
Target
Normalized Importance (per step)
0.0 0.2 0.4 0.6 0.8 1.0