SARIMA parameter selection Report - 西条

SARIMA$(p,d,q)(P,D,Q)_s$:

$$(1-\sum_{i=1}^{p}\phi_i L^i)(1-\sum_{I=1}^{P}\Phi_I L^{sI})(1-L)^d(1-L^s)^D y_t =$$

$$c+(1+\sum_{i=1}^{q}\theta_i L^i)(1+\sum_{I=1}^{Q}\Theta_I L^{sI})\varepsilon_t$$

$$\hat{y}_t = c+\sum_{i=1}^{p}\phi_i y_{t-i}+\sum_{I=1}^{P}\Phi_I y_{t-sI}+\sum_{j=1}^{q}\theta_j\varepsilon_{t-j}+\sum_{J=1}^{Q}\Theta_J\varepsilon_{t-sJ}$$

y_t: observed value at time t

$\hat{y}_t$: predicted value at time t

ϕ_i, Φ_I: non-seasonal and seasonal AR coefficients

θ_j, Θ_J: non-seasonal and seasonal MA coefficients

d, D: differencing orders (non-seasonal and seasonal)

s: seasonal period (here 24 for hourly data)

L: lag operator ($Ly_t = y_{t-1}$)

ε_t: white noise (random error term)

Objective = minimize residual variance σ_ε^2 to fit observed series.

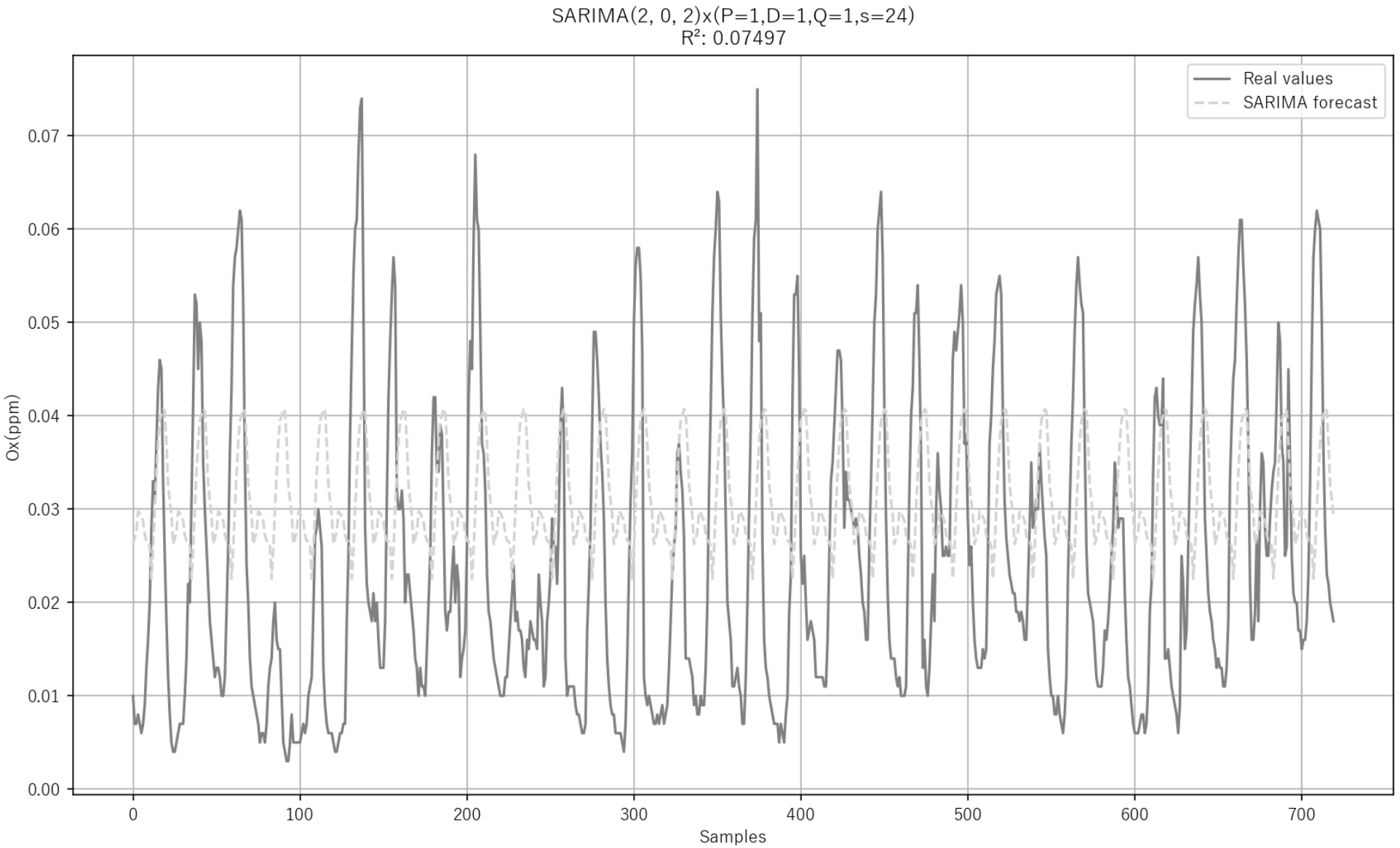
Prefecture code	38
Station code	38206050
Station name	西条
Target item	Ox(ppm)
Number of training samples	15768
Number of testing samples	6758
Model	SARIMA
SARIMA order	(2, 0, 2)
Seasonal order	(P=1,D=1,Q=1,s=24)
Parameter Grid (tested) p	[1, 2]
Parameter Grid (tested) d	[0, 1]
Parameter Grid (tested) q	[1, 2]
Predictions mean	0.03103258895564225
Predictions std	0.005293818667149114
Real mean	0.03427160402485943
Real std	0.018249610182089367
Ljung-Box residuals autocorrelation, Prob(Q)	0.0
Residuals skew	0.6132690783004816
Residuals kurtosis	3.13677390820583

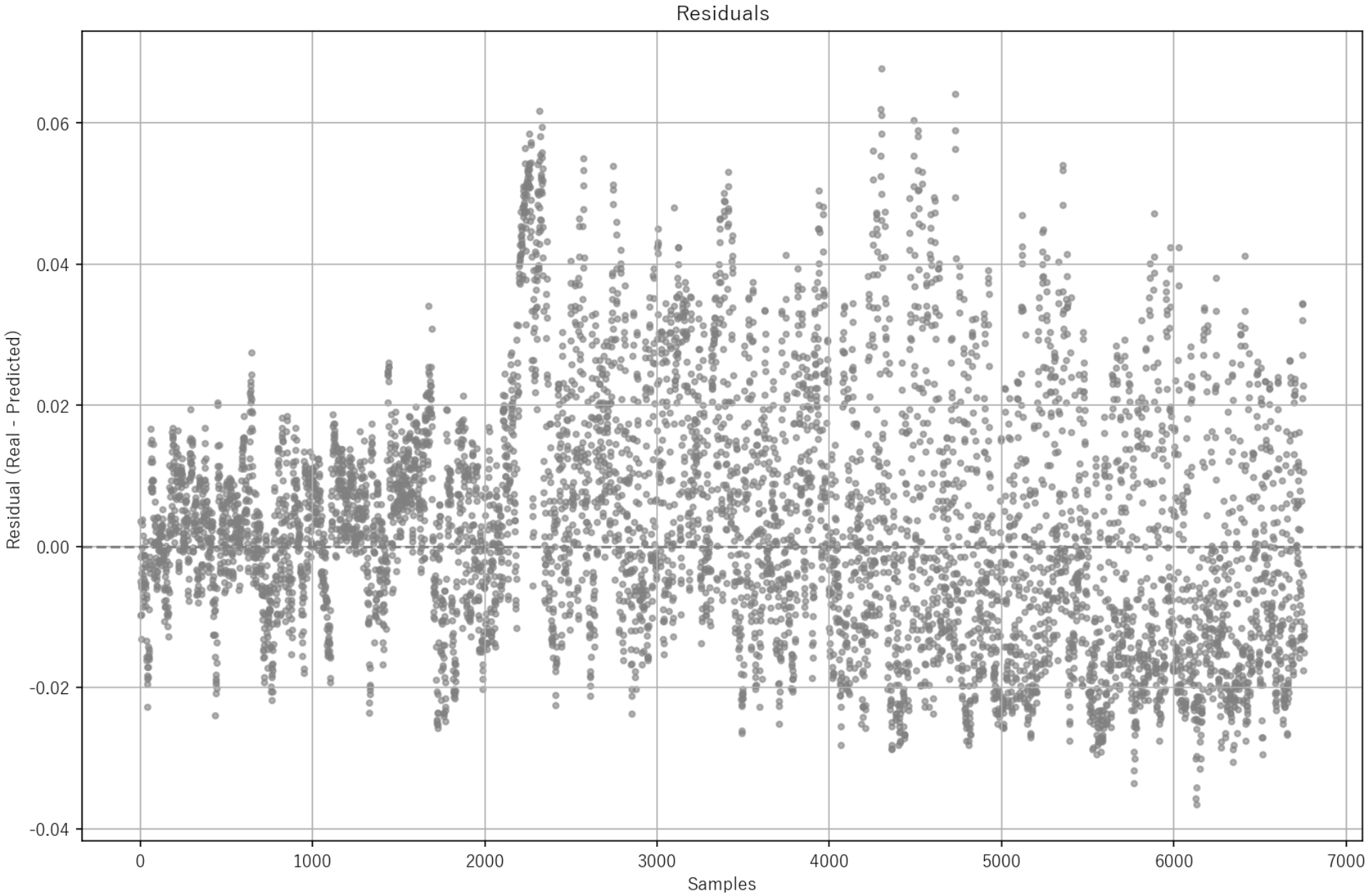
Features used for prediction

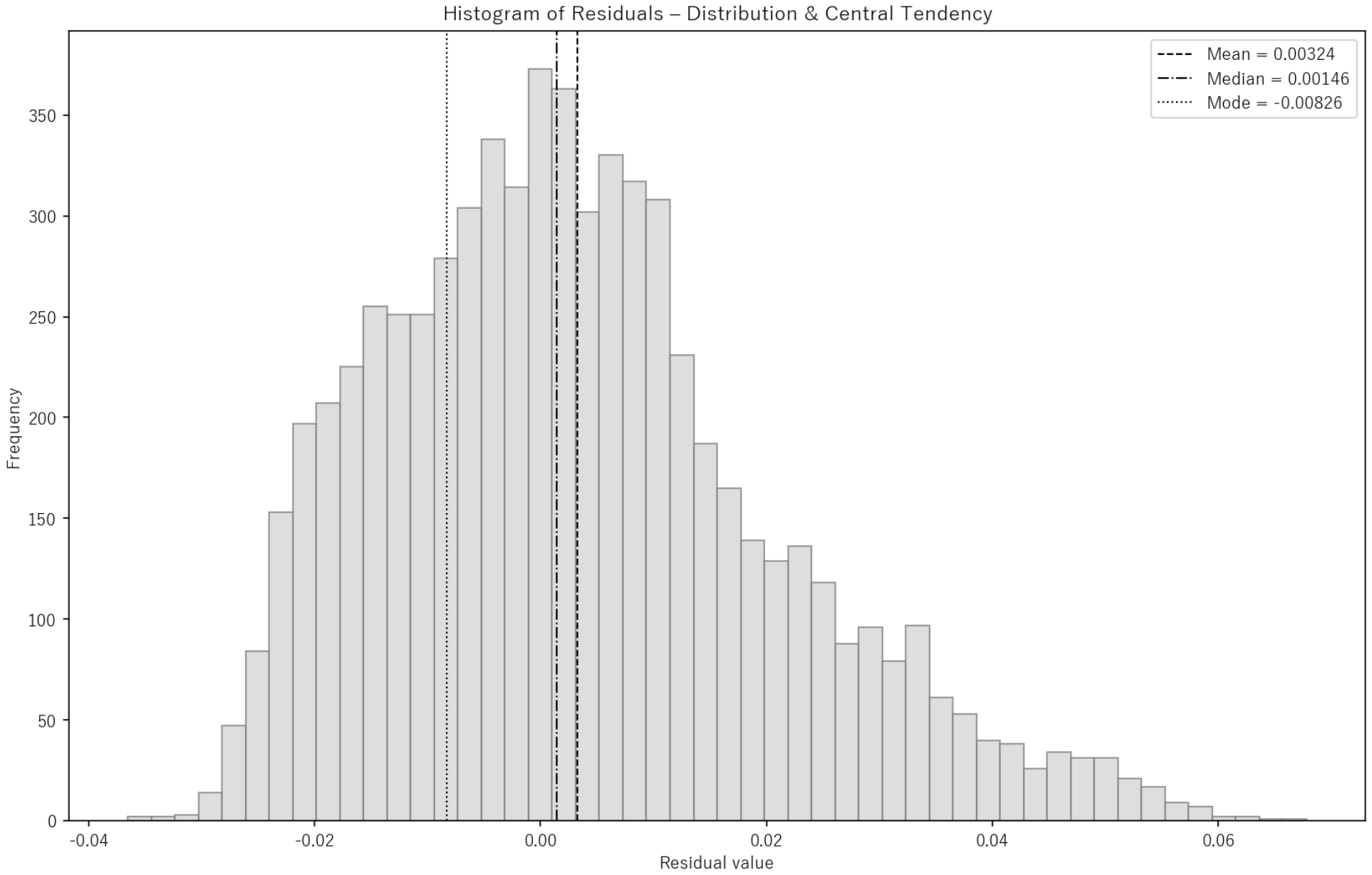
Ox(ppm)

Model accuracy

Target	R^2	MAE	RMSE
Ox(ppm)	0.0750	0.0136	0.0176


SARIMA(2, 0, 2)x(P=1,D=1,Q=1,s=24)
R²: 0.07497
Real values
SARIMA forecast
Ox(ppm)
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0
100
200
300
400
500
600
700
Samples


Residuals
Residual (Real - Predicted)
0.06
0.04
0.02
0.00
-0.02
-0.04
0
1000
2000
3000
4000
5000
6000
7000
Samples


Histogram of Residuals – Distribution & Central Tendency
Mean = 0.00324
Median = 0.00146
Mode = -0.00826
Frequency
0
50
100
150
200
250
300
350
-0.04
-0.02
0.00
0.02
0.04
0.06
Residual value