

# SARIMAX parameter selection Report - 西条

## SARIMAX Model Representation

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{l=1}^P \Phi_l y_{t-sl} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{j=1}^Q \Theta_j \varepsilon_{t-sj} + \beta^T X_t + \varepsilon_t$$

SARIMAX model = SARIMA model extended with exogenous variables ( $X_t$ ).

- $y_t$ : observed value at time  $t$
- $X_t$ : exogenous variables (external predictors)
- $\beta$ : coefficients for  $X_t$  (impact of external factors)
- $\phi_i, \Phi_i$ : autoregressive coefficients (non-seasonal, seasonal)
- $\theta_j, \Theta_j$ : moving-average coefficients (non-seasonal, seasonal)
- $d, D$ : differencing orders (non-seasonal, seasonal)
- $s$ : seasonal period (e.g., 24 for hourly data)
- $\varepsilon_t$ : white noise (random error)

Goal: model both temporal dynamics and the effect of external regressors.

Prefecture code	38
Station code	38206050
Station name	西条
Target item	Ox(ppm)
Number of training samples	8760
Number of testing samples	720
Model	SARIMAX
SARIMAX order (non-seasonal)	(2, 0, 0)
Seasonal order (P,D,Q,s)	(0,0,0,24)
Predictions mean	0.0253976193625562
Predictions std	0.01606772370313761
Real mean	0.0247277777777778
Real std	0.016101356242239538
Ljung-Box residuals autocorrelation, Prob(Q)	0.012300560862369516
Residuals skew	-0.9769511888742474
Residuals kurtosis	9.57901220658165

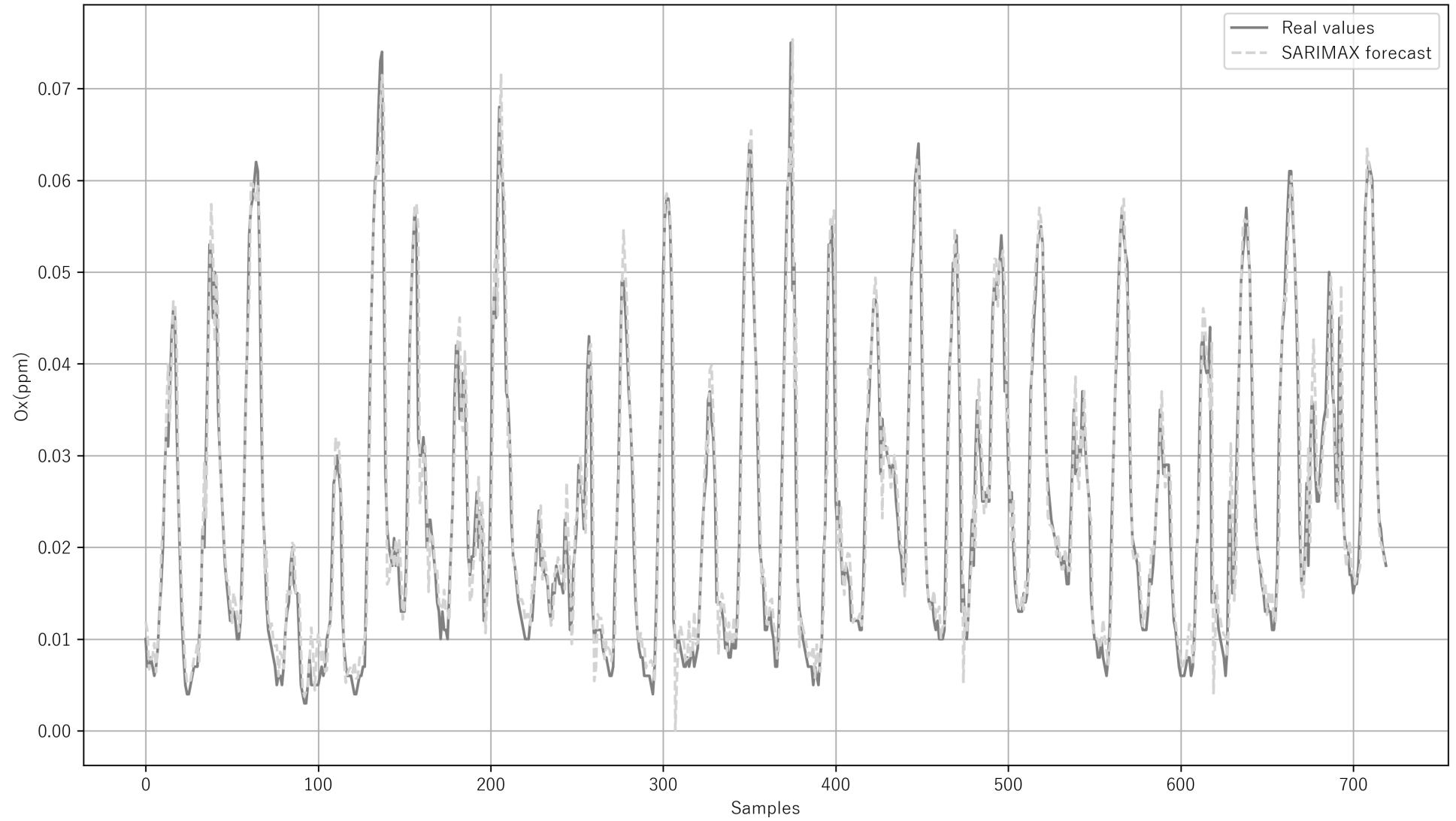
Features used for prediction

NO2(ppm)	U	V	Ox(ppm)_lag1	Ox(ppm)_lag2
Ox(ppm)_lag3	Ox(ppm)_roll_mean_3	Ox(ppm)_roll_std_6	Ox(ppm)_diff_1	hour_sin
dayofweek				

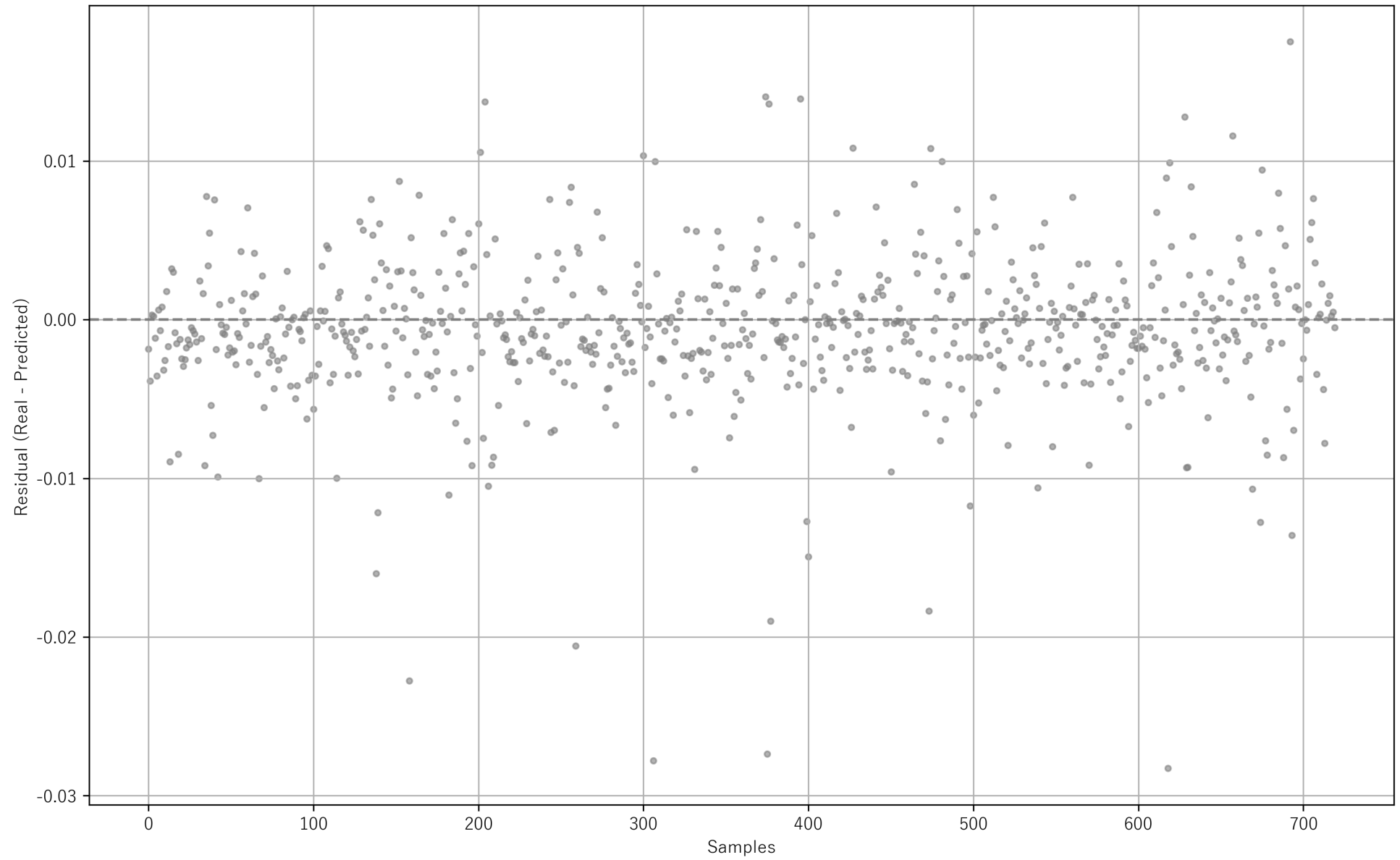
Model accuracy

Target	R <sup>2</sup>	MAE	RMSE
Ox(ppm)	0.9146	0.0032	0.0047

SARIMAX(2, 0, 0) x (P,D,Q,s)=(0,0,0,24)  
R<sup>2</sup>: 0.91455



Residuals (Out-of-sample)



Histogram of Residuals – Distribution & Central Tendency

