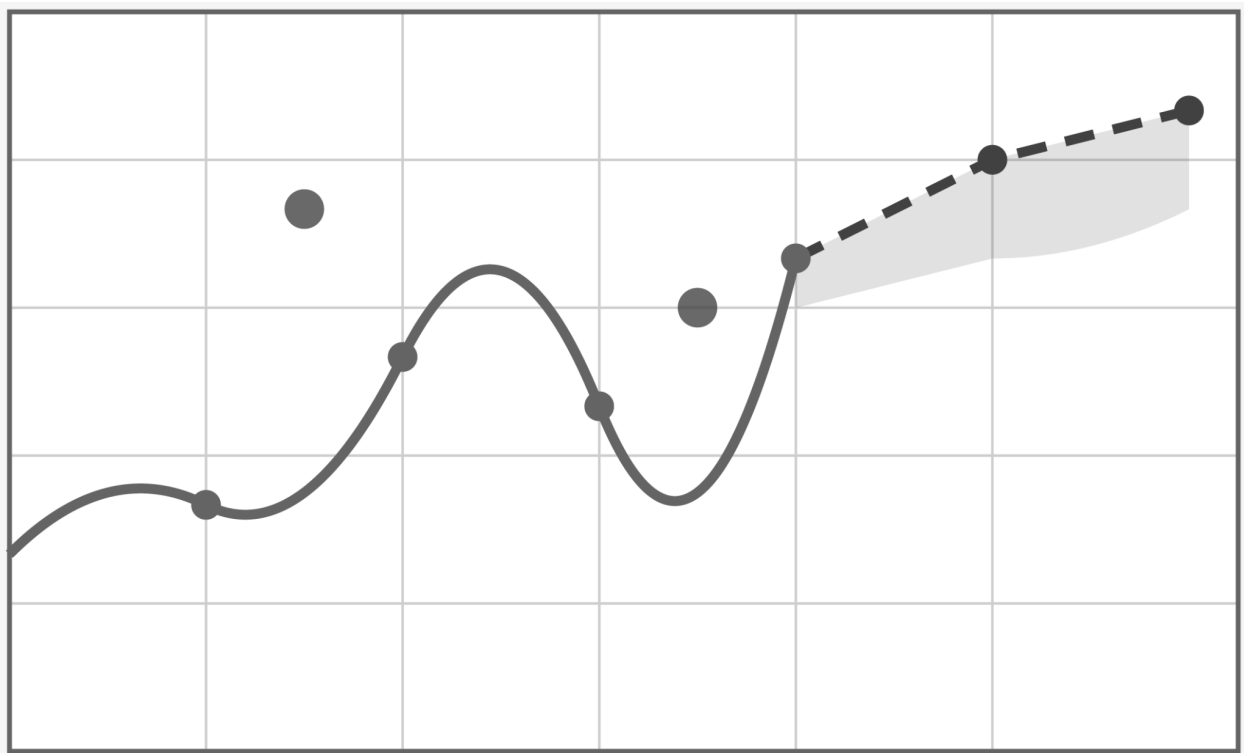# Exercises 4
(Prediction, regression AR)

DATA.ML.450 Time Series Analysis using Machine Learning (Autumn 2025)
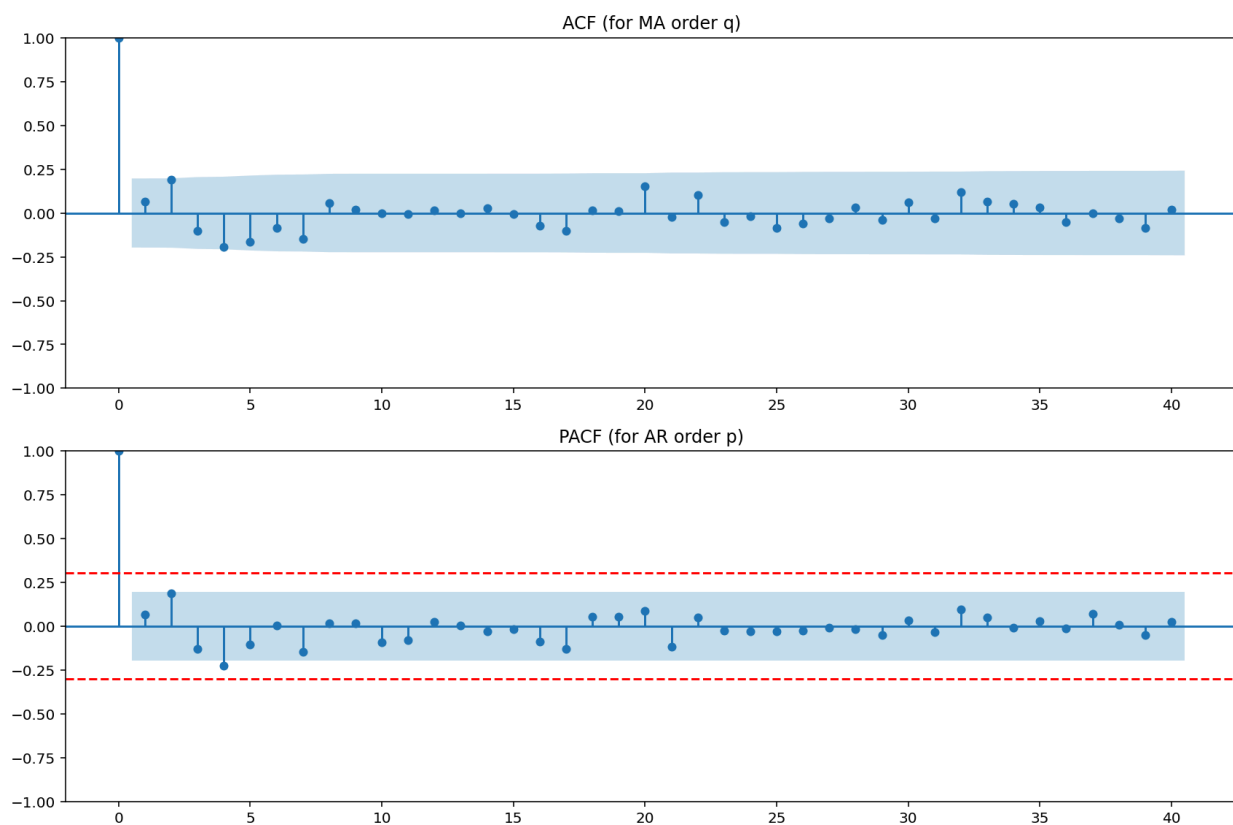

Anas Uddin

05.10.2025

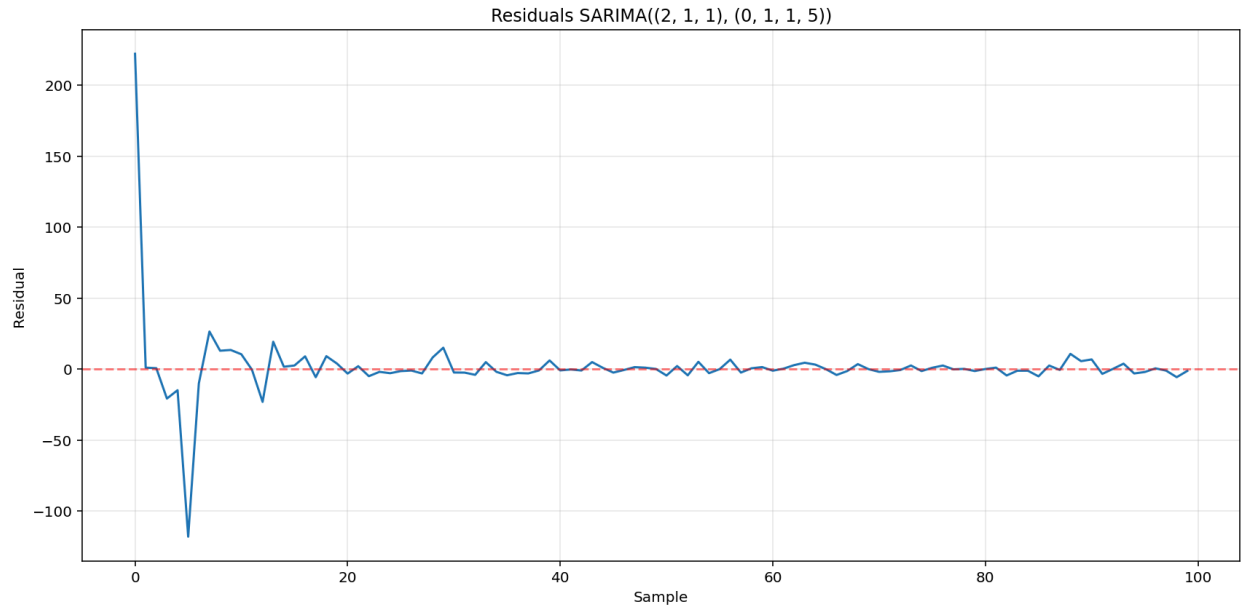MSc in Data Science

Exercise 1

I analyzed AAPL (Apple Inc.) stock data from the last 6 months. Downloaded as a CSV file from MarketWatch's 'Historical Quotes' (filename: Download Data - STOCK_US_XNAS_AAPL.csv). After splitting the data 80/20 into training and testing, I applied ARIMA and SARIMA models, using ACF/PACF and ADF tests to guide differencing and order selection. I compared multiple configurations and chose the best model based on test RMSE (residual amplitudes). The results showed that incorporating weekly seasonality with SARIMA provided the most accurate predictions, with low residual errors and good alignment with the actual stock prices.
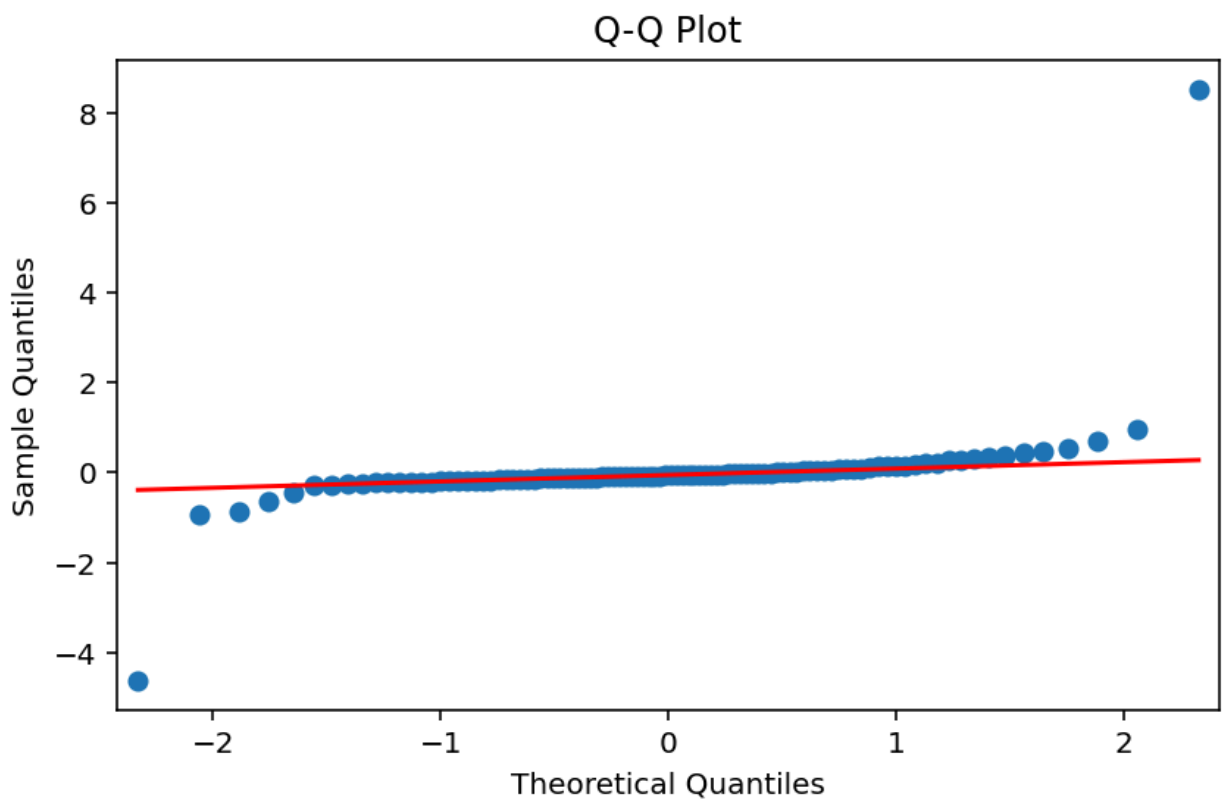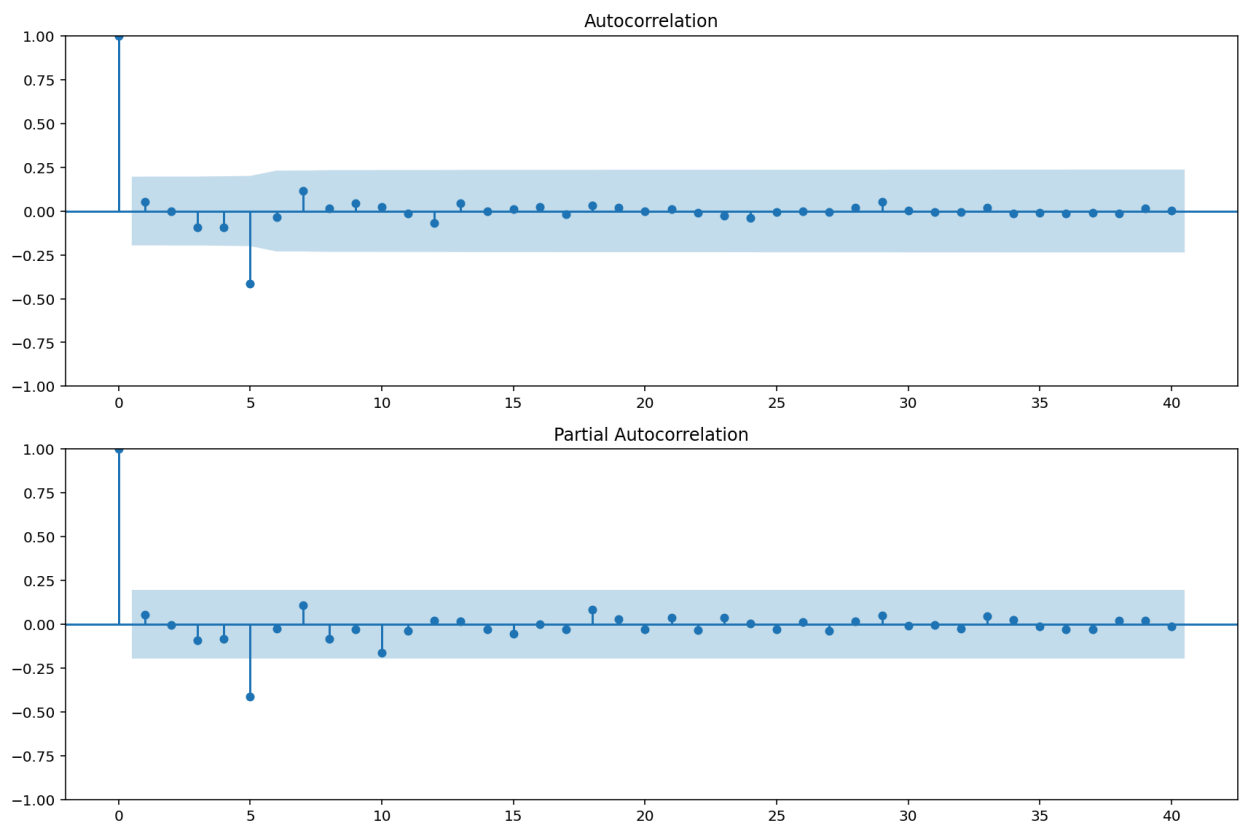
Output from the analysis:

Best model: SARIMA((2, 1, 1), (0, 1, 1, 5)), RMSE=14.6934
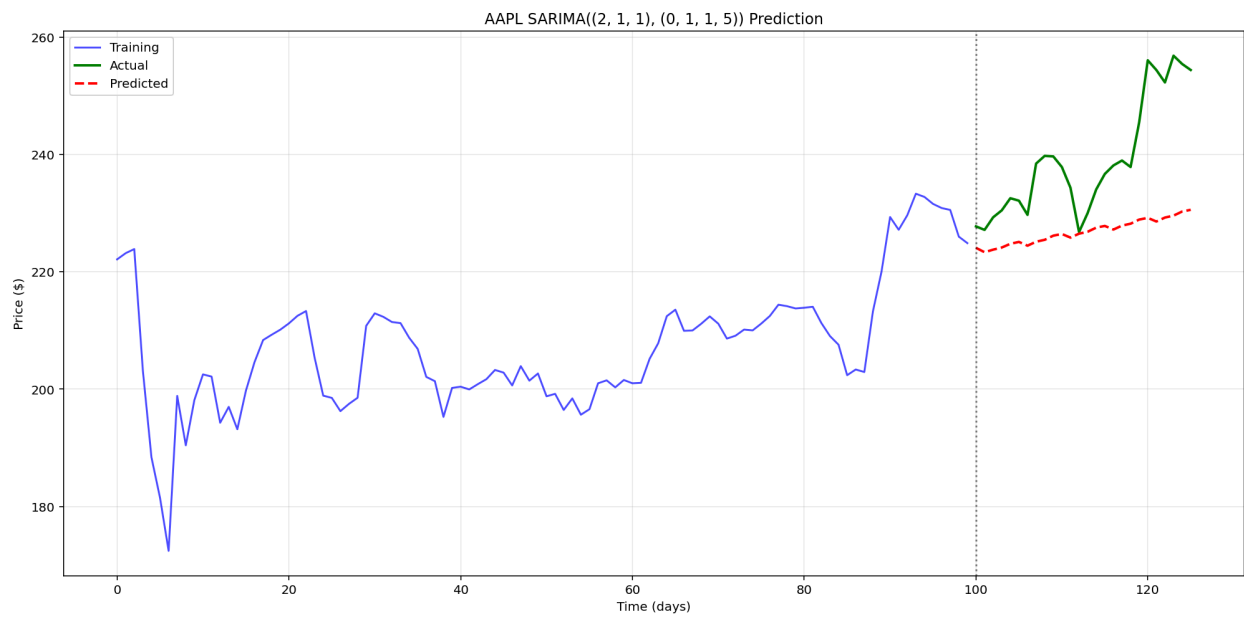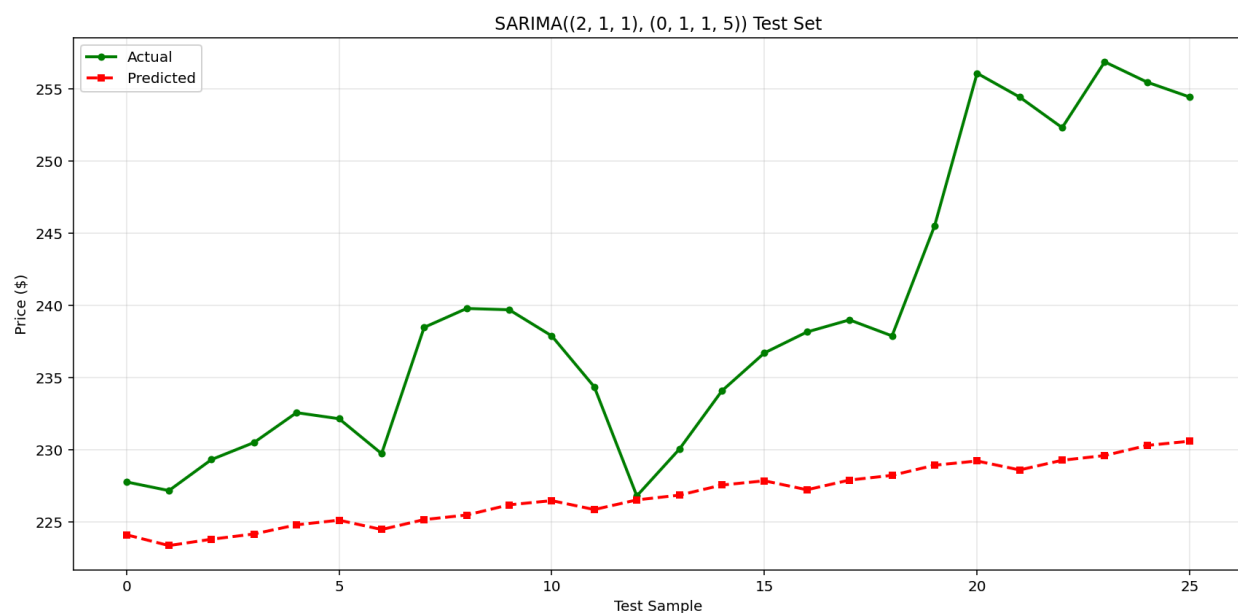
Durbin-Watson: 1.1638

Residuals SARIMA((2, 1, 1), (0, 1, 1, 5))

Normality test p-value: 0.0000

Q-Q Plot

2

Autocorrelation

Partial Autocorrelation

MAE: 12.2938, RMSE: 14.6934, MAPE: 5.02%

AAPL SARIMA((2, 1, 1), (0, 1, 1, 5)) Prediction

SARIMA((2, 1, 1), (0, 1, 1, 5)) Test Set

Top 5 configurations by RMSE:

type	config	AIC	BIC	RMSE
SARIMA ((2, 1, 1),	(0, 1, 1, 5))	595.780894	608.497368	14.693425
ARIMA	(4, 1, 1)	600.651000	616.221719	15.461147
ARIMA	(4, 1, 2)	602.521925	620.687764	15.872641
ARIMA	(3, 1, 2)	601.431015	617.001735	15.913331
SARIMA ((1, 1, 1),	(1, 0, 1, 5))	612.716309	625.691908	16.444482

Exercise 2

In this analysis, I used the given Nepal electricity consumption dataset (Nepal_electricity_consumption_in_MWh.pkl) to build an ARIMA-based time series model. I sorted the data by date and split it into 80% training and 20% testing sets. I tried and explored a few ARIMA and seasonal ARIMA configurations to capture the clear seasonality in the data. The final model was selected based on the lowest test RMSE, which ensures minimal residual amplitudes and the best fit for forecasting. The results show that the model follows the consumption trend and seasonal patterns.

Output from the analysis:
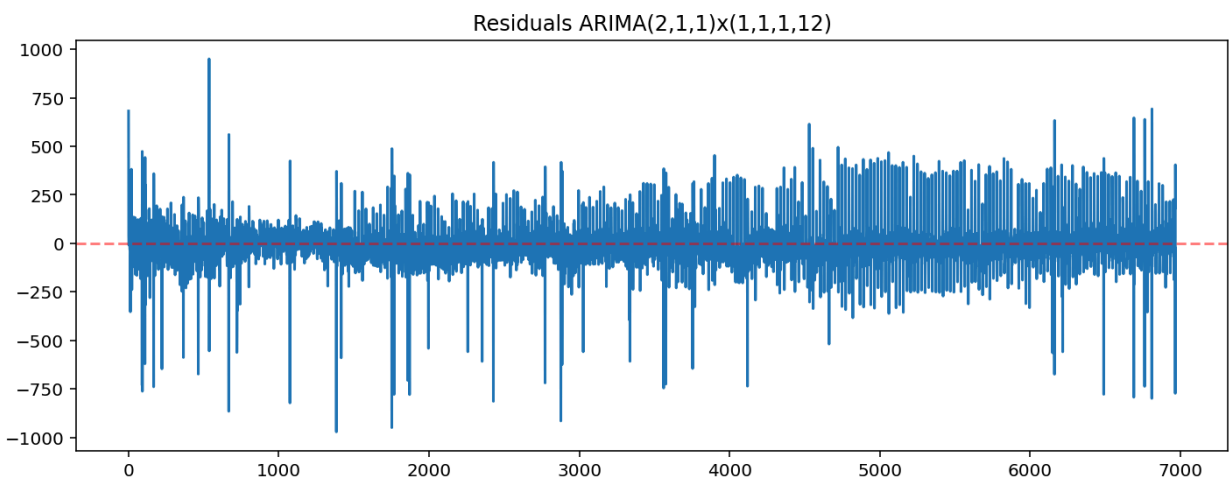

Electricity Consumption (First 500 samples)
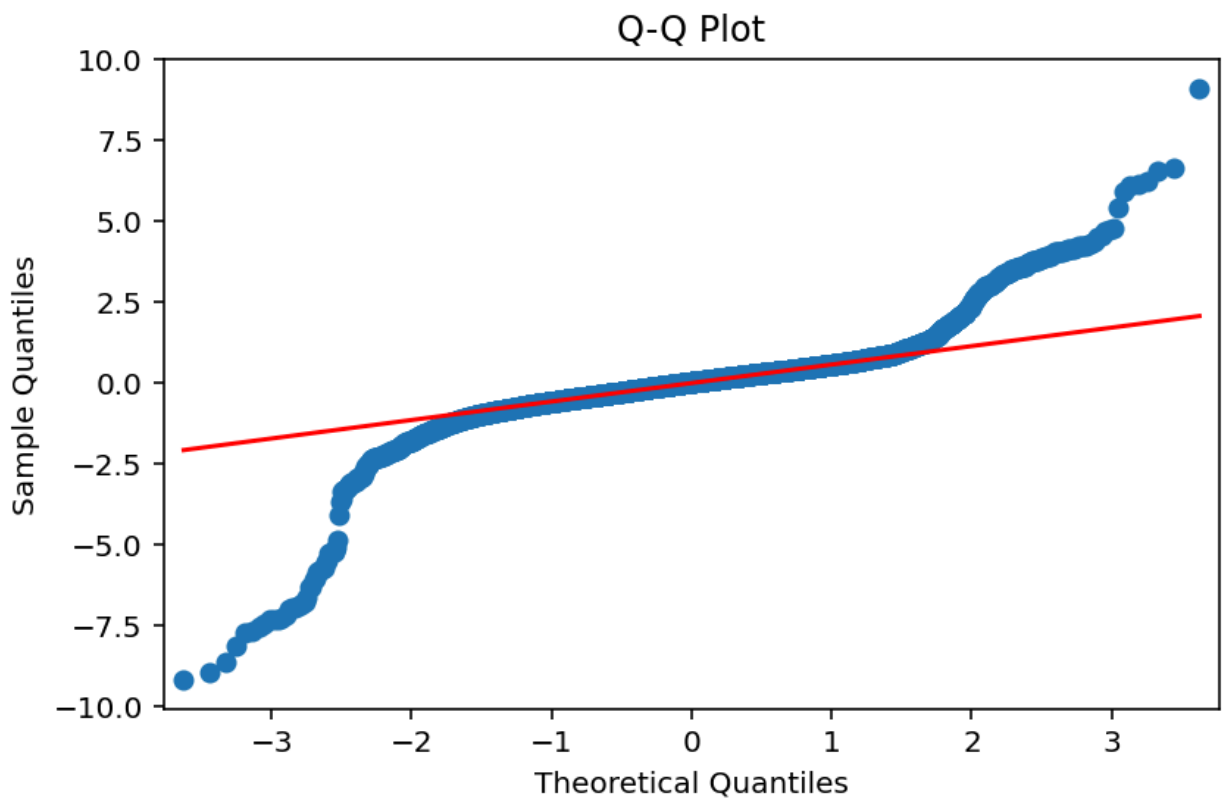
ADF p-value (original): 0.000000

ADF p-value (1st diff): 0.000000
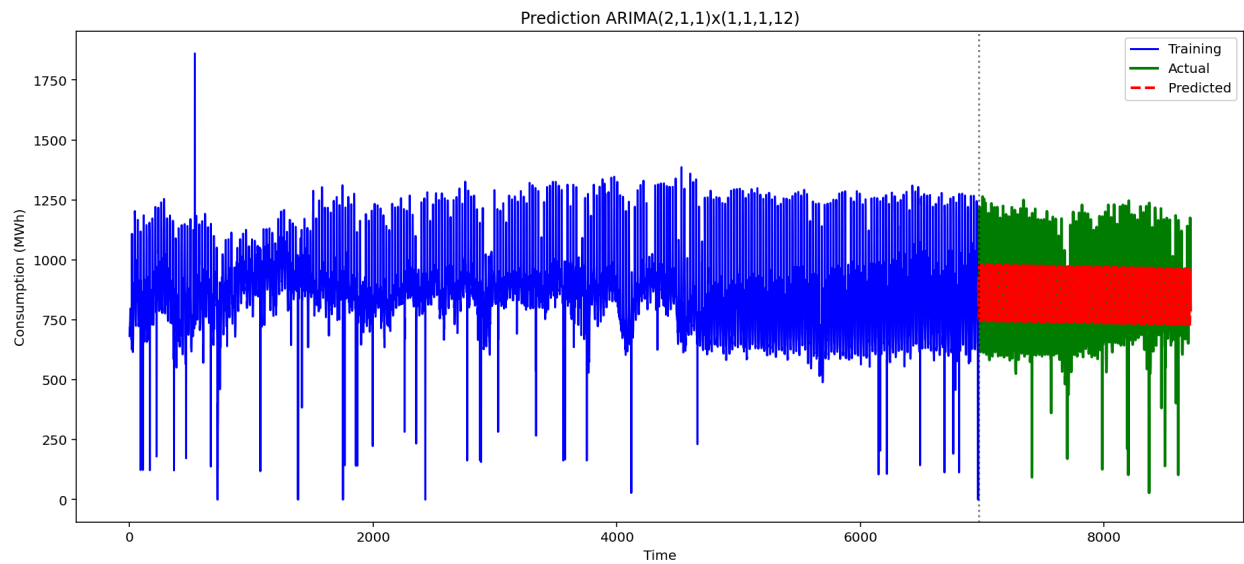
Best config (by RMSE): ARIMA(2, 1, 1, 12), RMSE=129.8747

Durbin-Watson: 1.8971

Residual normality p-value: 0.0000

Residuals ARIMA(2,1,1)x(1,1,1,12)

Q-Q Plot

MAE: 100.5141, RMSE: 129.8747, MAPE: 16.99%

Prediction ARIMA(2,1,1)x(1,1,1,12)

Top configurations (by RMSE):

config	RMSE
(2, 1, 1, 12)	129.874661
(1, 1, 1, 7)	169.007953
(2, 1, 1, 7)	169.038767
(1, 1, 1, None)	170.868064
(1, 1, 2, None)	170.871370