Temperature Forecasting

DATS 6313: Time Series and Analysis

Fall 2023

December 11, 2023

By

Shubham Jadhav
G30570862

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Table of Contents

Objective

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Objective

- To create a model to forecast temperature based on the features present in the dataset i.e., sensor data.
- To perform Time Series Analysis on the selected dataset.
- To evaluate model performance based on various parameters

Data Description

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

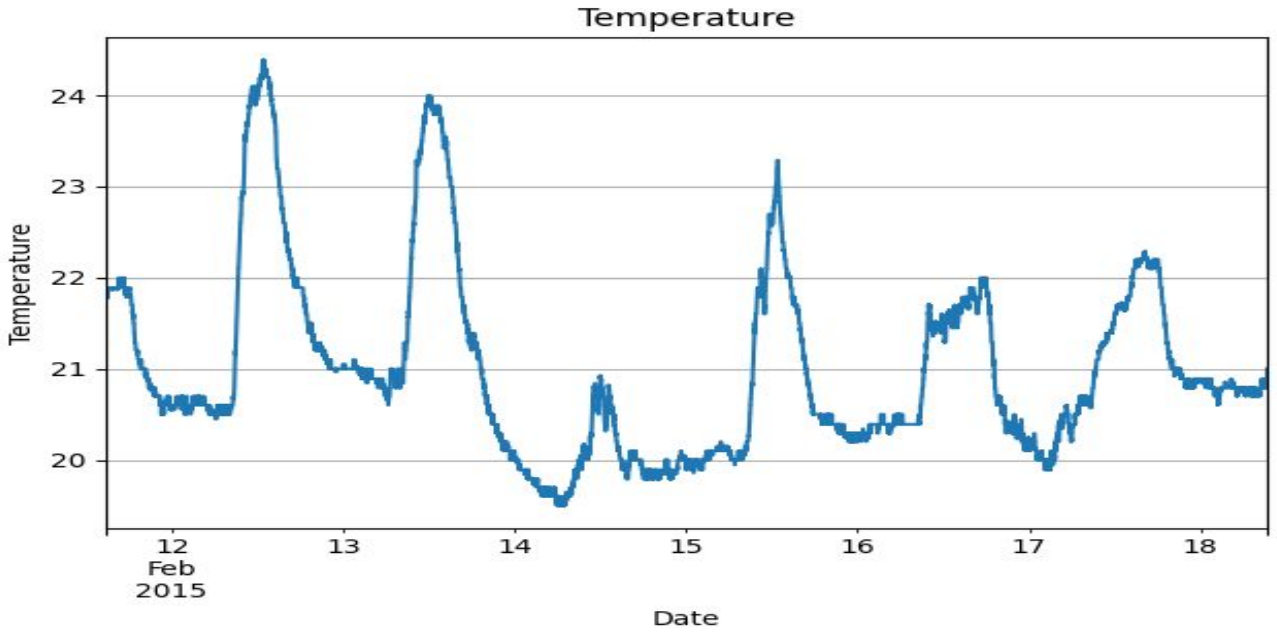
Data Description

- The Occupation dataset consists of contains data collected from an office room over a period of 7 days.
- The dataset was collected by using several sensors.
- 9752 observations and 7 numerical features.
- Features include time, temperature, humidity, light intensity, CO2 levels, humid- ity ratio, and occupancy.

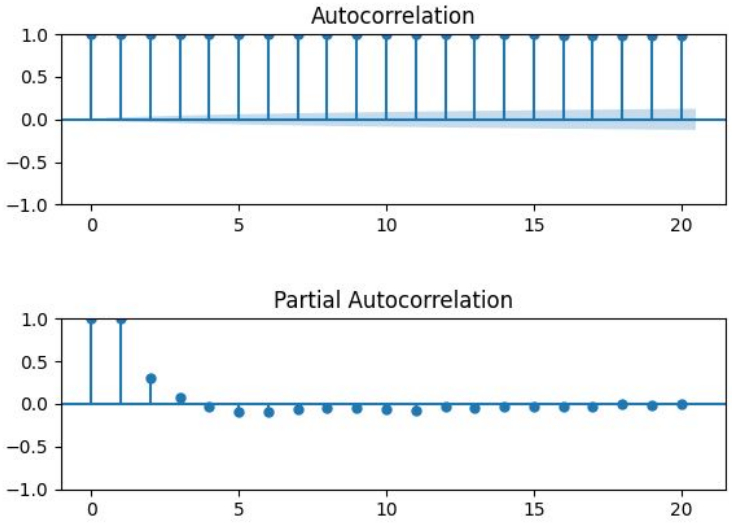
Sample observations from dataset

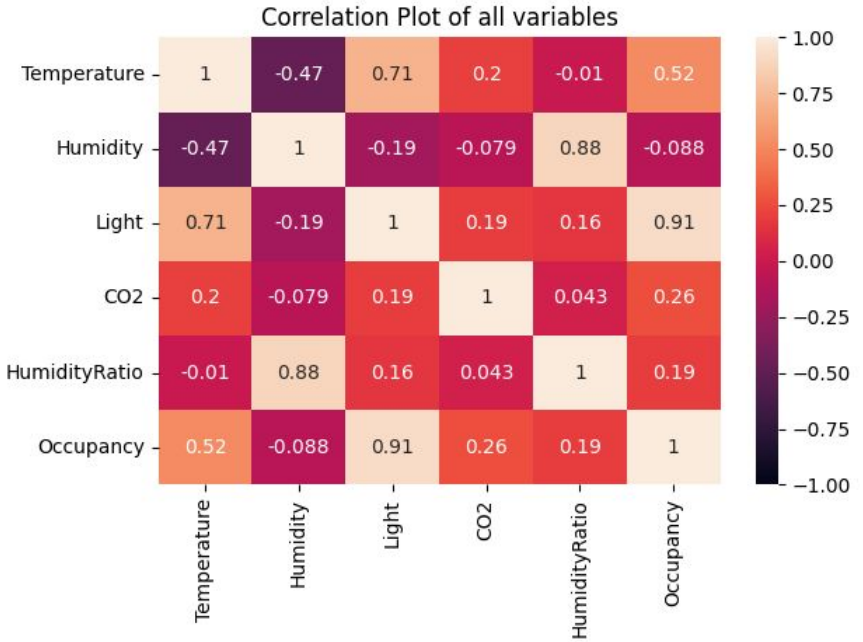
date	Temperature	Humidity	Light	CO2	HumidityRatio	Occupancy
2015-02-11 14:48:00	21.76	31.1333	437.333	1029.67	0.00502101	1
2015-02-11 14:49:00	21.79	31	437.333	1000	0.00500858	1
2015-02-11 14:50:00	21.7675	31.1225	434	1003.75	0.00502157	1
2015-02-11 14:51:00	21.7675	31.1225	439	1009.5	0.00502157	1
2015-02-11 14:52:00	21.79	31.1333	437.333	1005.67	0.0050303	1

Temperature over time



ACF/PACF and Correlation plot

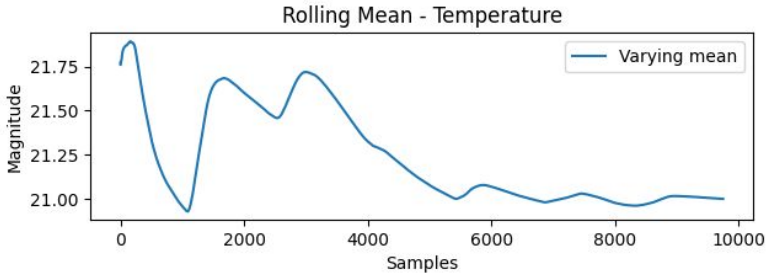


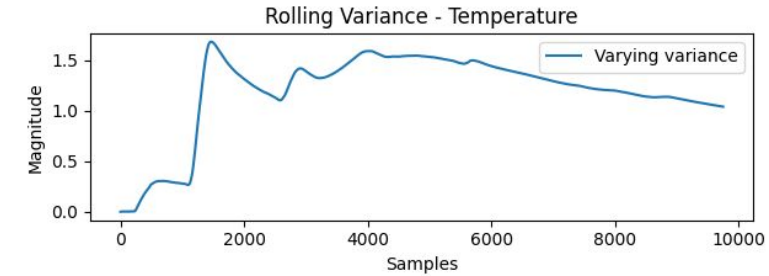


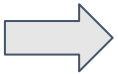
Stationarity

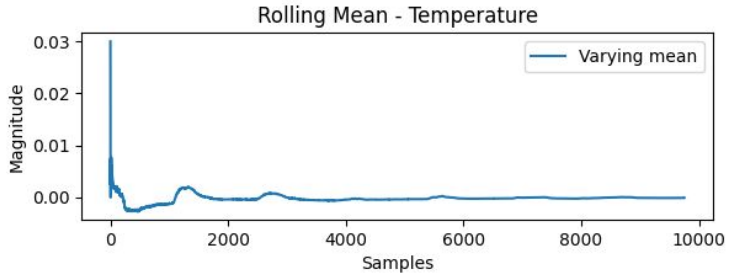
THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

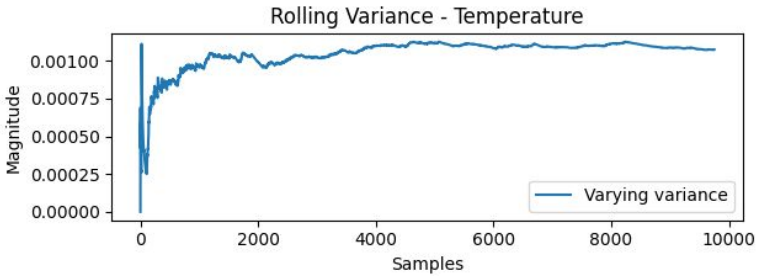
Rolling Mean and Variance











ADF and KPSS Test Results

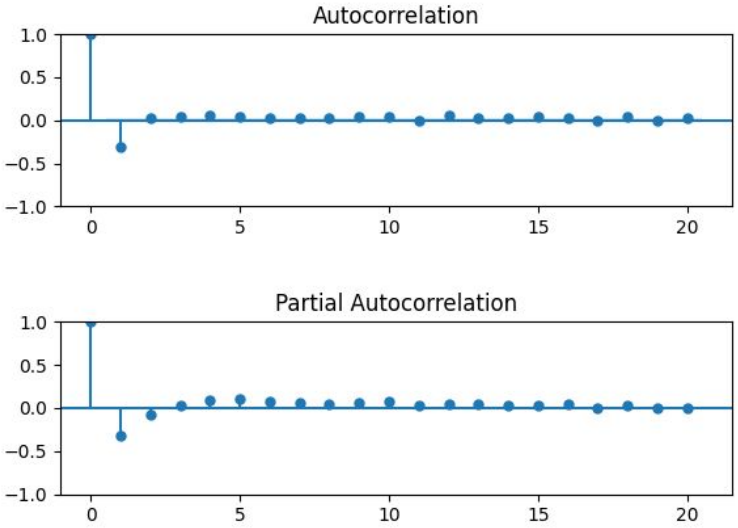
```
ADF Statistic: -9.684629
p-value: 0.000000
Critical Values:
    1%: -3.431023
    5%: -2.861838
    10%: -2.566928
```

```
Results of KPSS Test:
Test Statistic            0.082141
p-value                   0.100000
Lags Used                32.000000
Critical Value (10%)      0.347000
Critical Value (5%)       0.463000
Critical Value (2.5%)     0.574000
Critical Value (1%)       0.739000
```

According to ADF test p-value is less than 0.05, hence the time series is Stationary.

According to KPSS Test p-value is greater than 0.05, hence the time series is Stationary.

ACF/PACF plot of Stationary Data

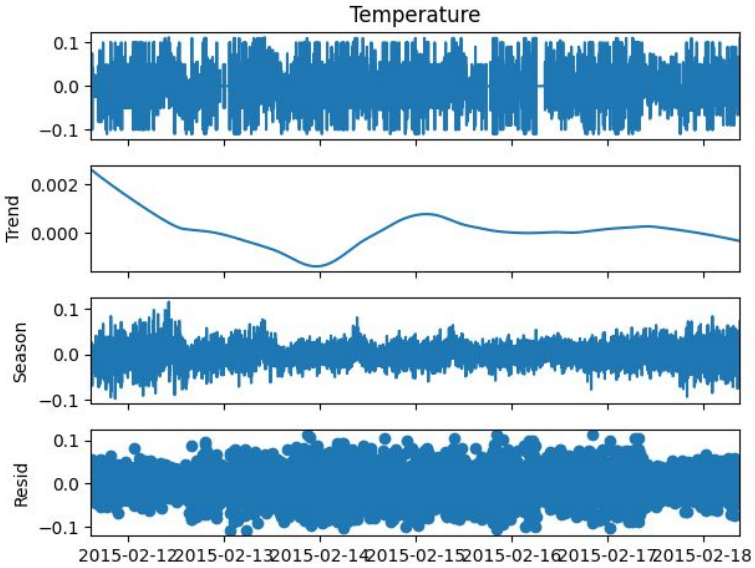


We can observe from the ACF/PACF plot that differenced data is now stationary.

Time Series Decomposition

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Time Series Decomposition

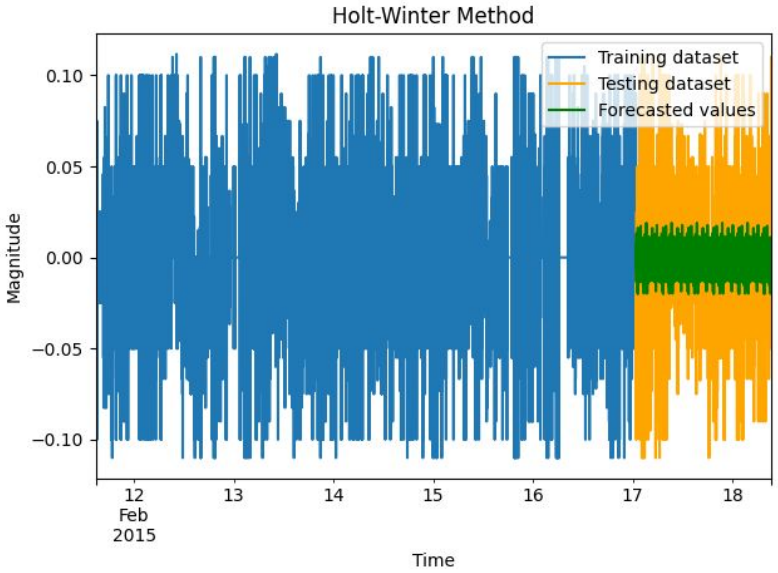


- The strength for this data is 0.0012 or 0.12% which is very low.
- The strength of this is 0.4107 or 41.07% which is significant.

Holt-Winters Method

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Holt-Winters Method

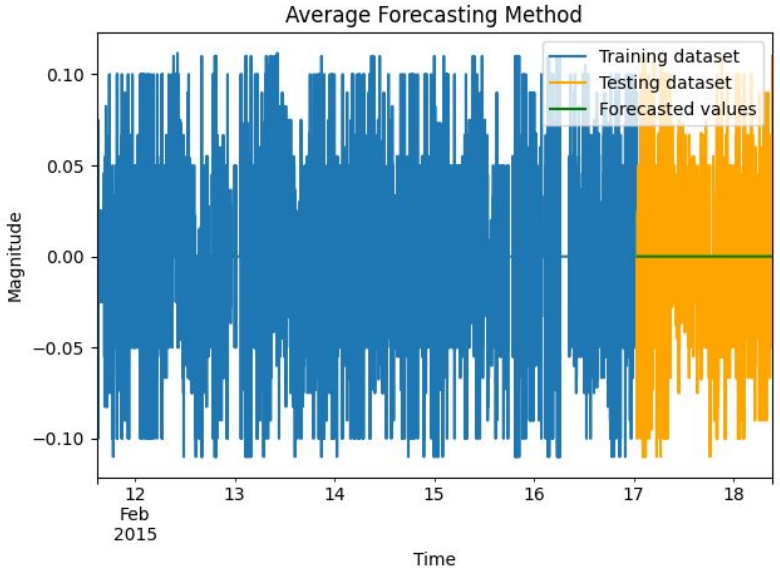


- Q-Value: 957.818 (Critical: 36.19)
- Mean of residual error: 0
- MSE of residual error: 0.001
- Mean of forecast error: 0.002
- MSE of forecast error: 0.001
- Variance of residual errors versus forecast errors: 1.13

Base Models

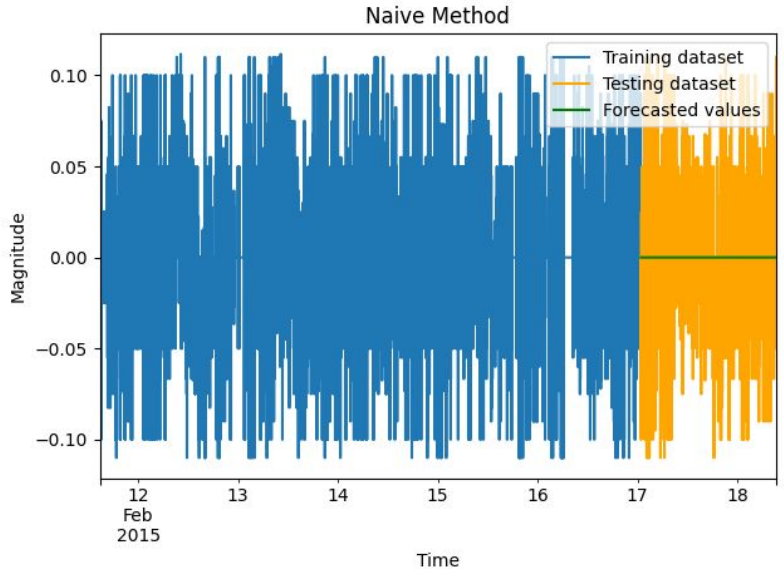
THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Average Method



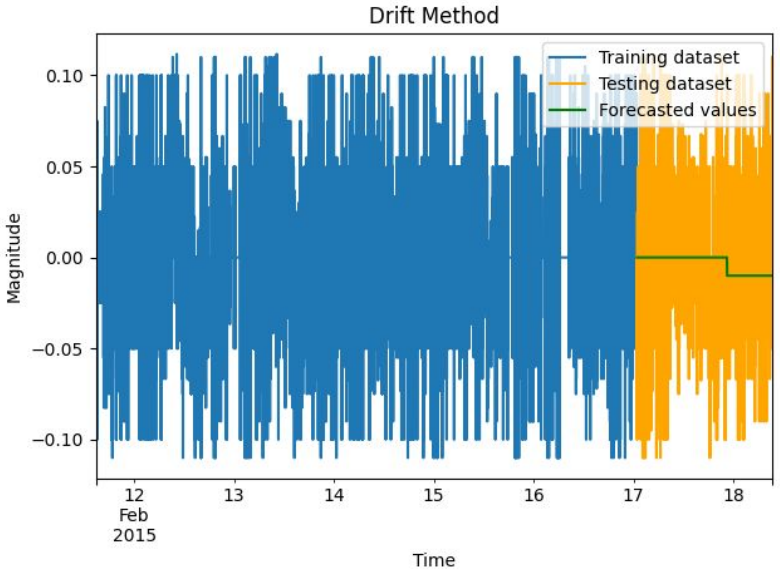
- Q-Value: 968.28 (Critical: 36.19)
- Mean of residual error: 0
- MSE of residual error: 0.001
- Mean of forecast error: 0.004
- MSE of forecast error: 0.001
- Variance of residual errors versus forecast errors: 1.14

Naive Method



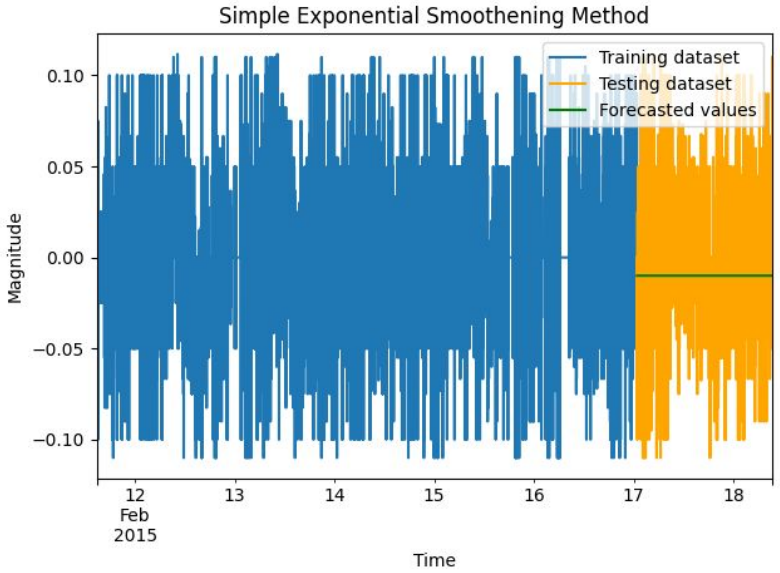
- Q-Value: 957.818 (Critical: 36.19)
- Mean of residual error: 0
- MSE of residual error: 0.003
- Mean of forecast error: 0.004
- MSE of forecast error: 0.001
- Variance of residual errors versus forecast errors: 2.98

Drift Method



- Q-Value: 957.818 (Critical: 36.19)
- Mean of residual error: 0
- MSE of residual error: 0.003
- Mean of forecast error: 0.004
- MSE of forecast error: 0.001
- Variance of residual errors versus forecast errors: 1.72

SES Method



- Q-Value: 957.818 (Critical: 36.19)
- Mean of residual error: 0
- MSE of residual error: 0.002
- Mean of forecast error: 0.002
- MSE of forecast error: 0.010
- Variance of residual errors versus forecast errors: 1.13

Base Model Comparison

Model	Q-Value	Critical-Value	White-Residual	m_res	mse_res	var_res	m_pred	mse_pred	var_pred
Average	968.24	36.1909	No	-0.0002	0.0011	0.0011	0.0004	0.001	0.001
Naive	3282.9	36.1909	No	0	0.0029	0.0029	0.0004	0.001	0.001
Drift	3270.29	36.1909	No	0.0002	0.0029	0.0029	0.0038	0.001	0.001
SES	1854.69	36.1909	No	-0	0.0017	0.0017	0.0104	0.0011	0.001

Multiple Linear Regression

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Feature Selection

We performed feature selection in three ways.

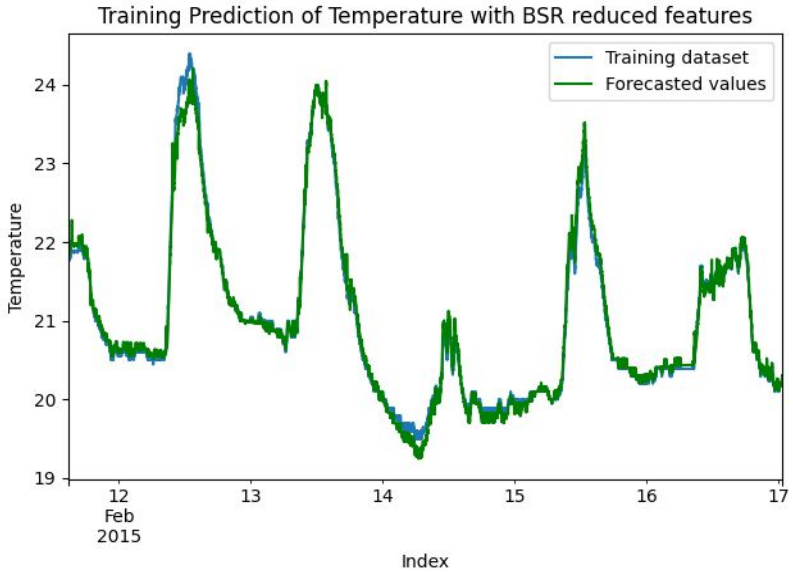
- Backward Stepwise Regression: Using this method, the initial OLS model did not have any feature whose p-value is less than 0.05. Therefore, no features were dropped.
- Variance Inflation Factor: Using VIF method the highest VIF score was observed for Humidity feature. But removing Humidity, degrades the OLS performance. Therefore, no features were dropped.
- Principal Component Analysis: Using PCA method, we received 4 components and executed OLS model on those 4 components.

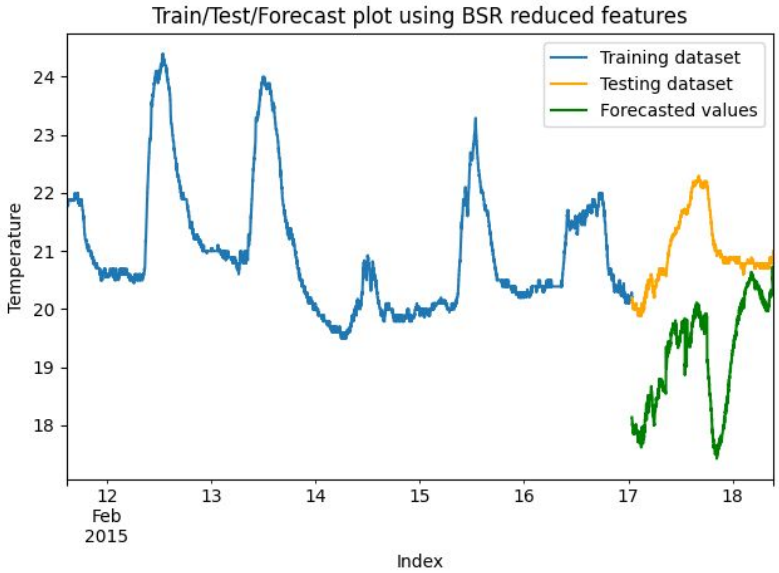
Model results from Feature Selection

	Model	R2	Adjusted-R2	AIC	BIC	F-test	Mean Cross-val
0	All	0.9911	0.9911	-13196.3	-13154.5	0	0.7615
1	BSR	0.9911	0.9911	-13196.3	-13154.5	0	0.7615
2	VIF	0.9911	0.9911	-13196.3	-13154.5	0	0.7615
3	PCA	0.9429	0.9429	1296.01	1330.82	0	-5.2516

	Model	Q-Value	Critical-Value	White-Residual	mse_res	var_res	mse_pred	var_pred
0	All	134490	36.1909	No	0.0108	0.0108	4.0355	0.7771
1	BSR	134490	36.1909	No	0.0108	0.0108	4.0355	0.7771
2	VIF	134490	36.1909	No	0.0108	0.0108	4.0355	0.7771
3	PCA	74914.9	36.1909	No	0.069	0.069	1.2287	1.2287

Plots for results using features from BSR





ARMA

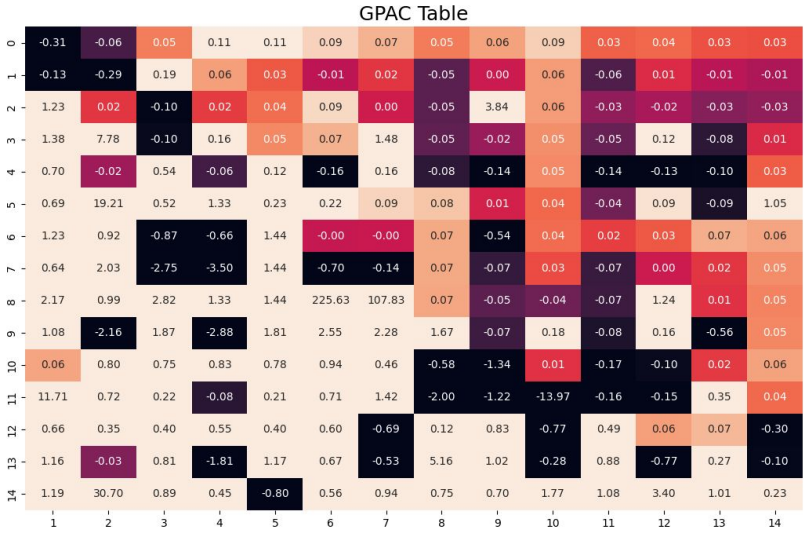
THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Order Determination using GPAC

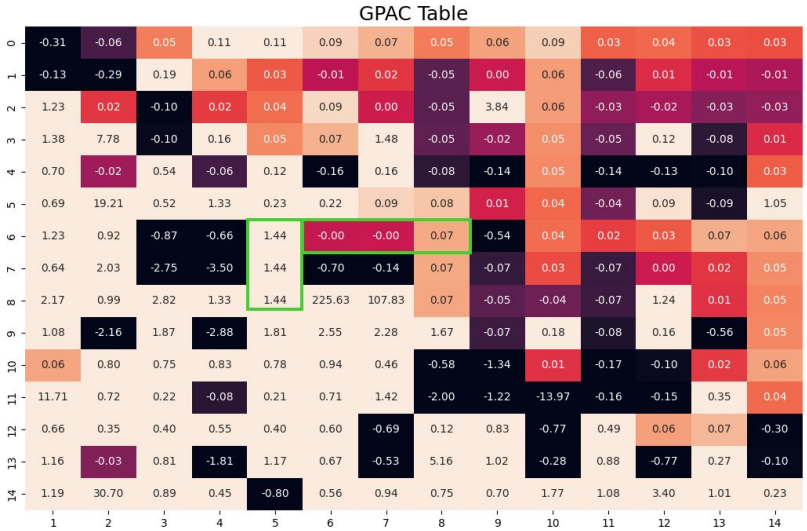
Using GPAC table we can observe four order or AR and MA.

- ARMA (5, 6)
- ARMA (8, 1)
- ARMA (10, 2)
- ARMA (11, 7)

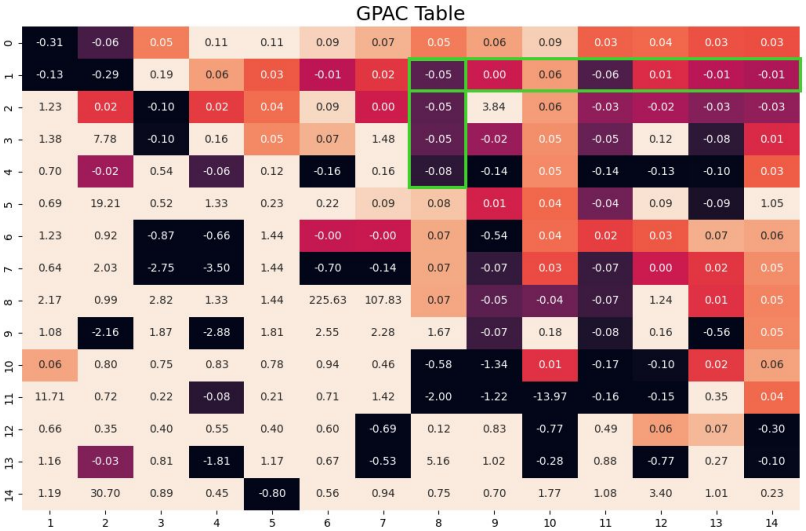
We then perform time series forecasting for the data using these three order.



ARMA order from GPAC

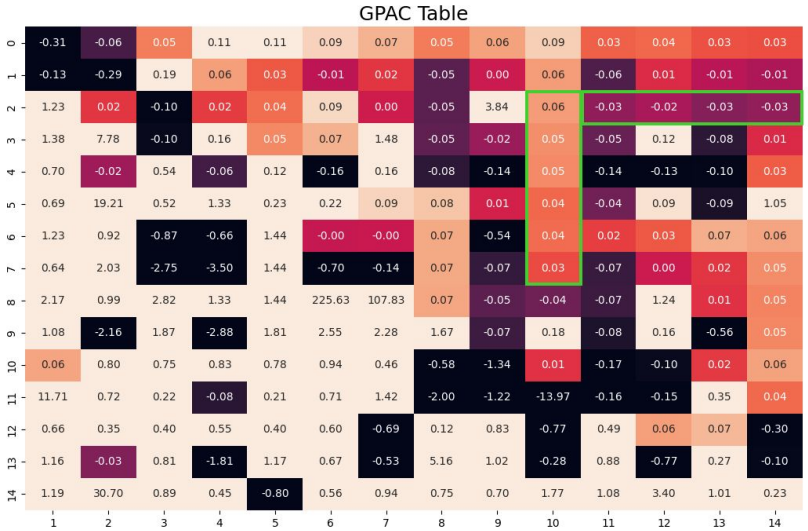


ARMA(5,6)

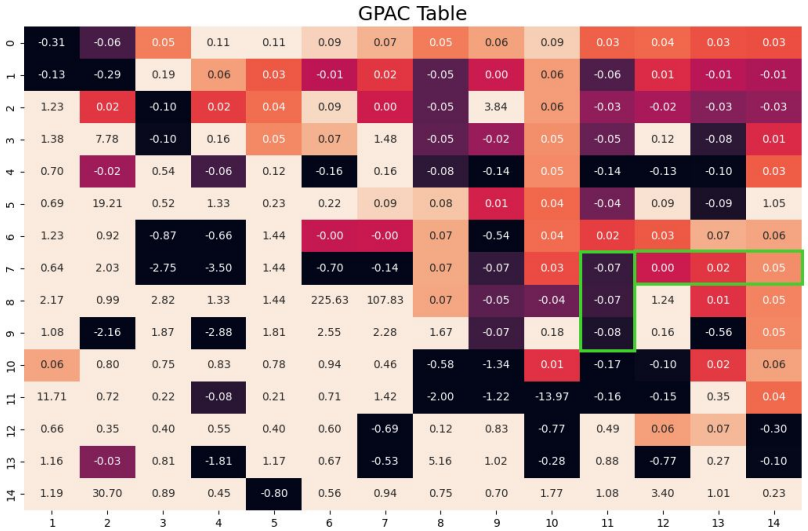


ARMA(8,1)

ARMA order from GPAC (continued)

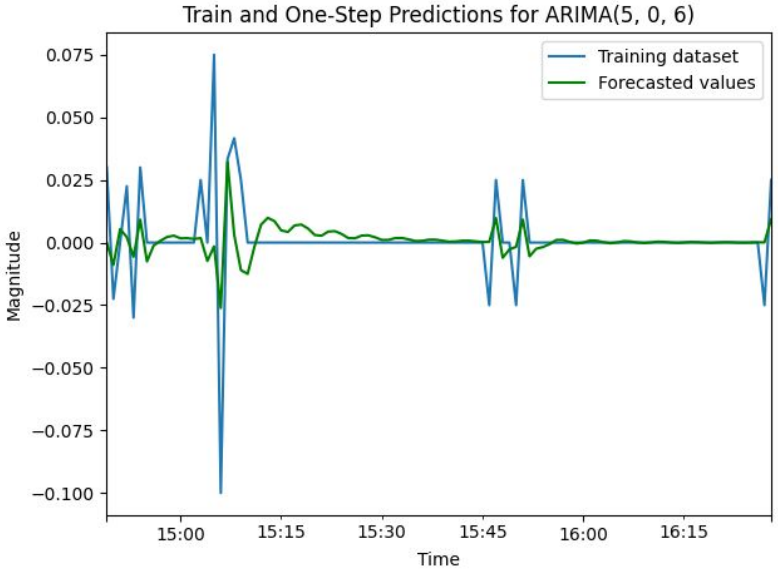


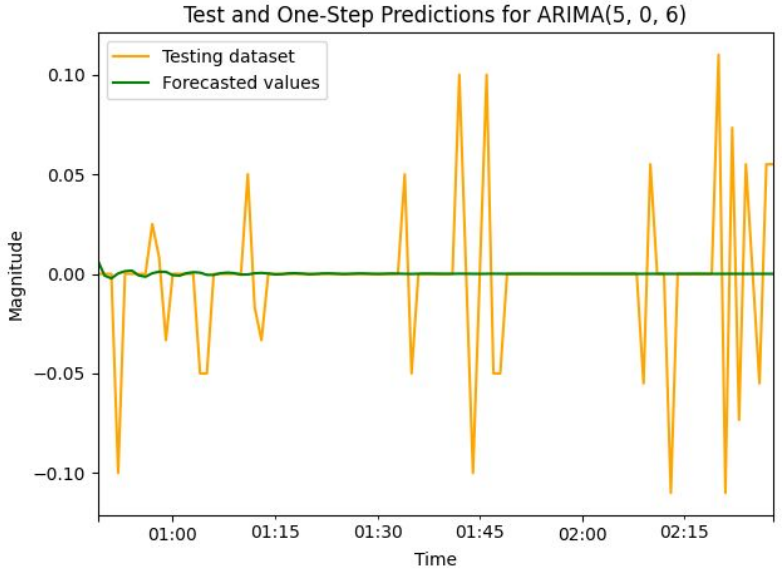
ARMA(10,2)



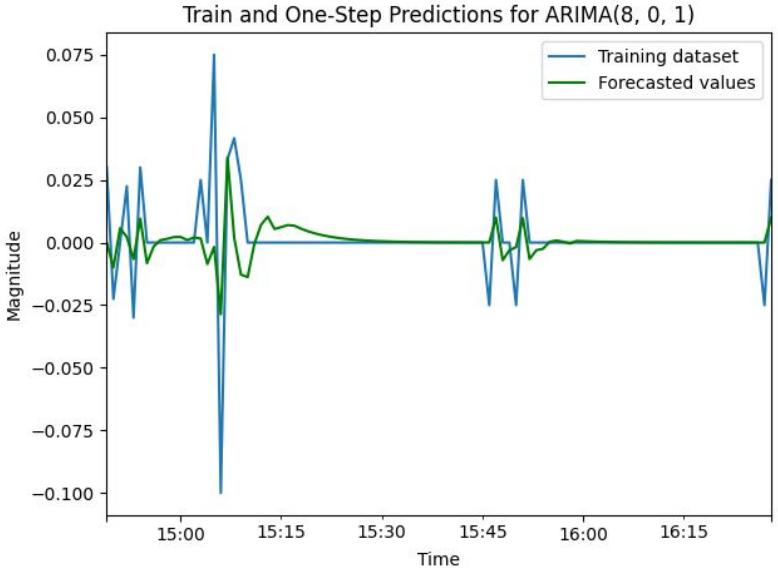
ARMA(11,7)

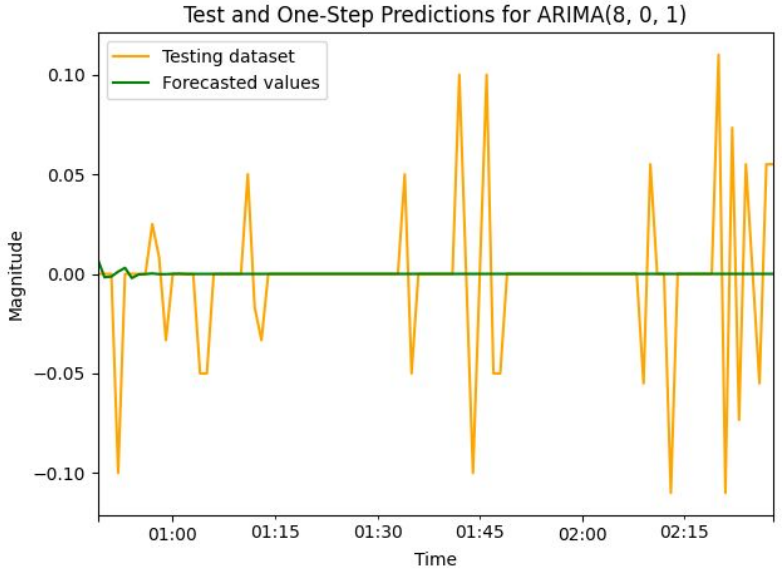
ARMA (5,6)


Train and One-Step Predictions for ARIMA(5, 0, 6)
Training dataset
Forecasted values
Magnitude
Time
0.075
0.050
0.025
0.000
−0.025
−0.050
−0.075
−0.100
15:00
15:15
15:30
15:45
16:00
16:15

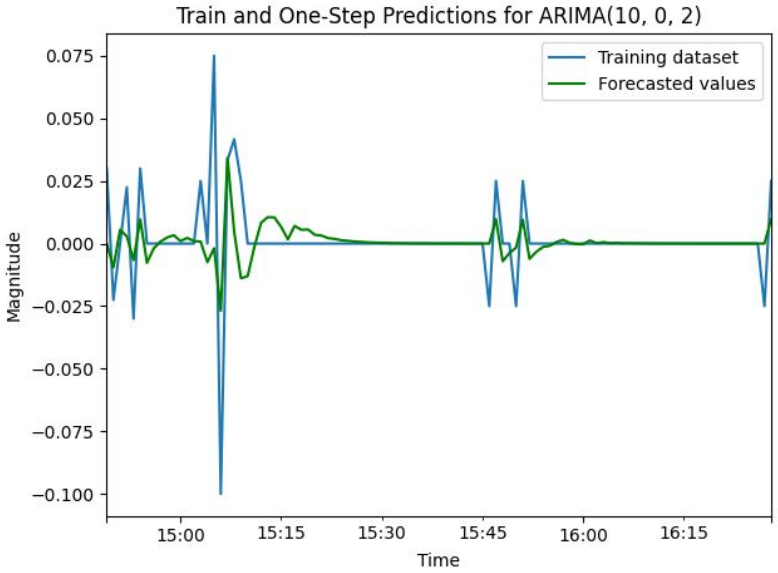

Test and One-Step Predictions for ARIMA(5, 0, 6)
Testing dataset
Forecasted values
Magnitude
Time
0.10
0.05
0.00
−0.05
−0.10
01:00
01:15
01:30
01:45
02:00
02:15

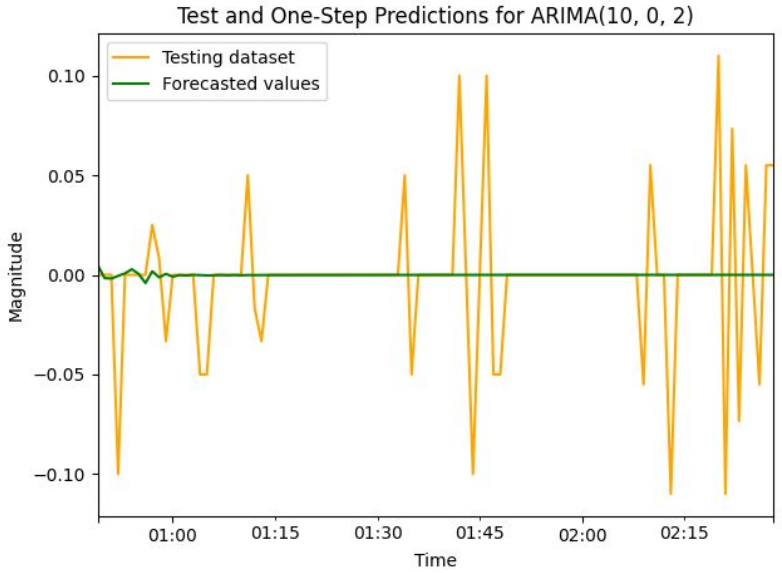
ARMA (8,1)


Train and One-Step Predictions for ARIMA(8, 0, 1)
Training dataset
Forecasted values
Magnitude
Time
0.075
0.050
0.025
0.000
−0.025
−0.050
−0.075
−0.100
15:00
15:15
15:30
15:45
16:00
16:15

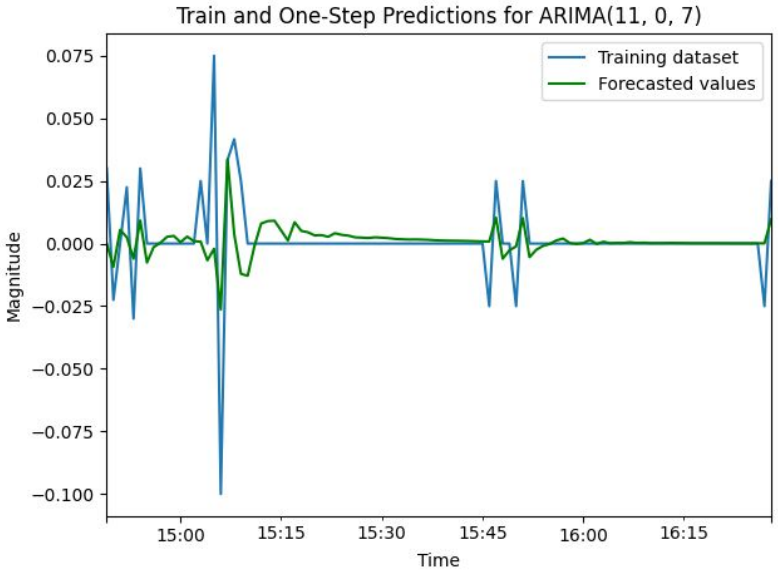

Test and One-Step Predictions for ARIMA(8, 0, 1)
Testing dataset
Forecasted values
Magnitude
Time
0.10
0.05
0.00
−0.05
−0.10
01:00
01:15
01:30
01:45
02:00
02:15

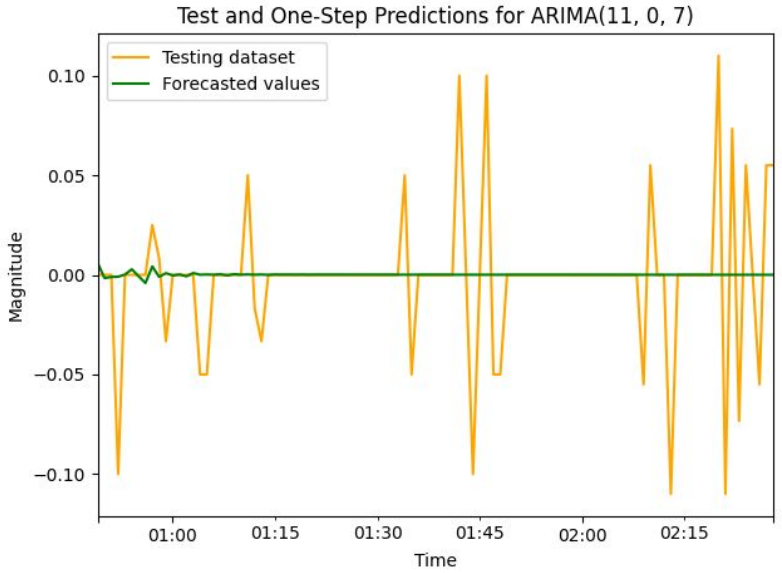
ARMA (10,2)


Train and One-Step Predictions for ARIMA(10, 0, 2)
Training dataset
Forecasted values
Magnitude
Time
0.075
0.050
0.025
0.000
−0.025
−0.050
−0.075
−0.100
15:00
15:15
15:30
15:45
16:00
16:15


Test and One-Step Predictions for ARIMA(10, 0, 2)
Testing dataset
Forecasted values
Magnitude
Time
0.10
0.05
0.00
−0.05
−0.10
01:00
01:15
01:30
01:45
02:00
02:15

ARMA (11,7)


Train and One-Step Predictions for ARIMA(11, 0, 7)
Training dataset
Forecasted values
Magnitude
Time
0.075
0.050
0.025
0.000
−0.025
−0.050
−0.075
−0.100
15:00
15:15
15:30
15:45
16:00
16:15


Test and One-Step Predictions for ARIMA(11, 0, 7)
Testing dataset
Forecasted values
Magnitude
Time
0.10
0.05
0.00
−0.05
−0.10
01:00
01:15
01:30
01:45
02:00
02:15

Model Results

Model	Q-Value	Critical-Value	White-Residual	m_res	mse_res	var_res	m_pred	mse_pred	var_pred
ARIMA(5, 0, 6)	35.3081	36.1909	Yes	-0.0001	0.0009	0.0009	0.0004	0.001	0.001
ARIMA(8, 0, 1)	40.5755	36.1909	No	-0.0001	0.0009	0.0009	0.0004	0.001	0.001
ARIMA(10, 0, 2)	17.8974	36.1909	Yes	-0.0001	0.0009	0.0009	0.0004	0.001	0.001
ARIMA(11, 0, 7)	21.3006	36.1909	Yes	-0.0001	0.0009	0.0009	0.0004	0.001	0.001

Final Model Selection

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Model Comparison

Method	Q-Value	Improvement
Naive Method	3282.90	-
Drift Method	3270.29	0.38%
SES Method	1854.69	43.53%
Average Method	968.24	47.72%
Holt-Winter's Method	957.81	1.09%
ARMA (8,0)	40.58	95.76%
ARMA (5,6)	35.31	13.02%
ARMA (11,7)	21.30	39.85%
ARMA (10,2)	17.89	15.92%

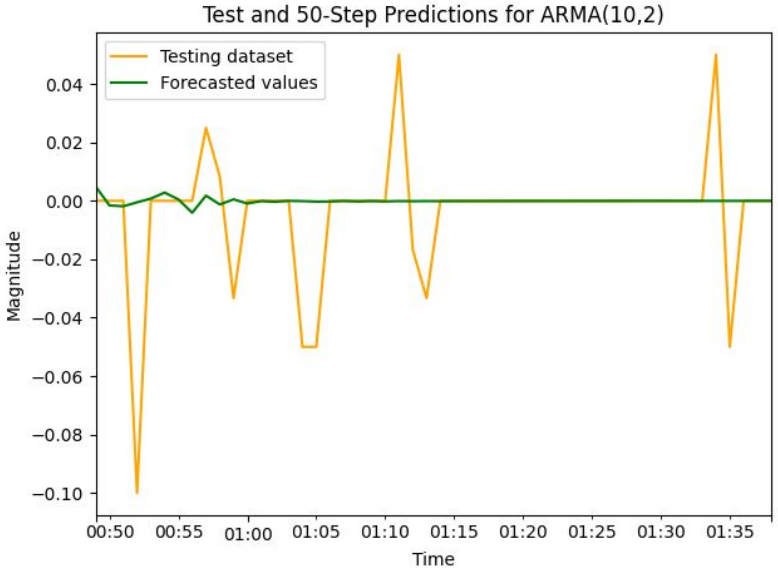
ARMA(10, 2) exhibit the lowest Q value amongst all models indicating a comparable goodness of fit, with the residuals following a white noise pattern.

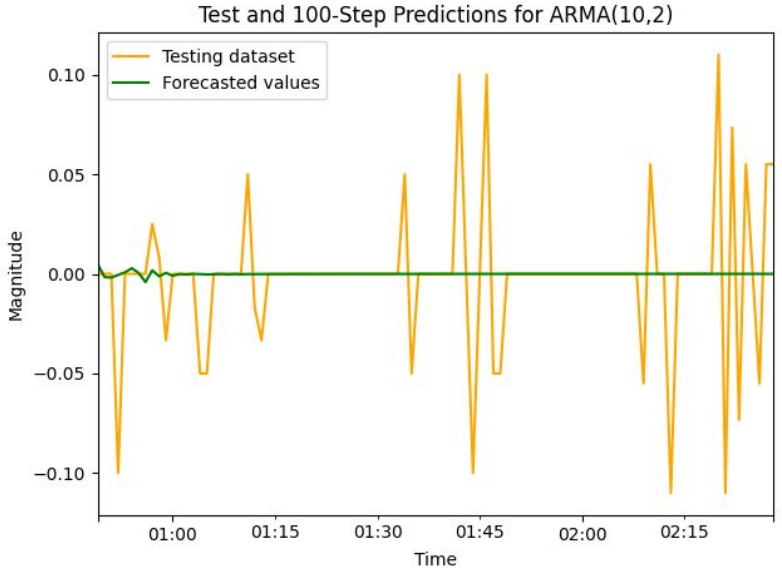
We can consider the Multiple Linear Regression model as the model explains 99% of the variance in the data, suggesting a good fit.

H-step ahead Prediction

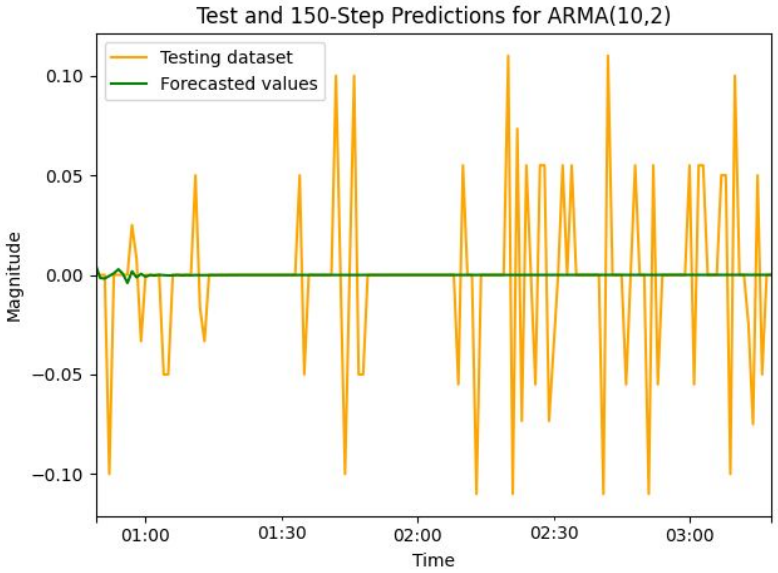
THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

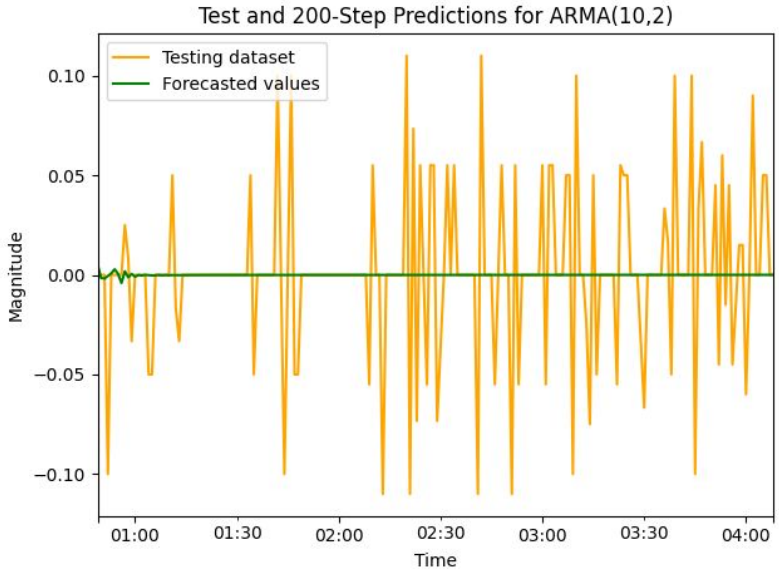
50-step and 100-step Prediction





150-step and 200-step Prediction





Thank You

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC