Time series Analysis and Modeling

Energy Demand Forecasting

DATS 6313

Brooklyn Chen

Table of content

Table of figures and tables

Table of figures and tables

Abstract

In this study, I aimed to forecast kW, a dependent variable, using three electricity-related variables: KWh, kVARh, and kVAR. These variables measure energy, power, and reactive power and are commonly used in power system analysis. I used different models to predict kW, evaluated their performance, identified the optimal model, and developed the forecasting function for kW.

Introduction

Time series analysis and modeling process

1. Understand dataset
2. Stationary Test
3. Time Series Decomposition
4. Feature Selection
5. Predictions using different methods and models
6. Diagnostic Testing
7. Final Model Selection
8. Forecasting Function

Description of the dataset

The independent variables

- **kWh** (kilowatt-hour) :

used to measure electricity consumption. One kilowatt-hour is equal to the energy consumed by using one kilowatt of power for one hour.

- **kVARh** (kilovolt-ampere reactive hour)

a unit of reactive power of a power system. Reactive power is power generated by capacitors or inductors for regulating voltage and current.

- **kVAR** (kilovolt-ampere reactive)

a unit of reactive power used to measure the reactive power in a power system for maintaining stable voltage and preventing overload in the power system.

The dependent variable

- **kW** (kilowatt)

used to measure the rate at which electricity is used. One kilowatt is equal to a power consumption or production rate of 1000 watts per second.

Description of the dataset

Table 1 Smart Meter Energy(kW) – Raw Data

serial	kWh	kW	kVARh	kVAR	Time_stamp
3098000032	10.854	0	7.814	0.002	2018/3/1 15:30
3098000032	10.75	0	7.813	0	2018/3/1 14:30
3098000032	12.325	0.086	8.302	0.076	2018/3/5 9:00
3098000032	12.372	0.094	8.345	0.086	2018/3/5 9:30
3098000032	12.415	0.086	8.386	0.082	2018/3/5 10:00
…	…	…	…	…	…
3098000032	6781.4	0.982	1721.24	0.406	2019/3/31 22:00
3098000032	6781.92	1.04	1721.32	0.46	2019/3/31 22:30
3098000032	6782.52	1.192	1721.39	0.498	2019/3/31 23:00
3098000032	6783.12	0	1721.47	0.436	2019/3/31 23:30
3098000032	6783.84	0	1721.55	0	2019/4/1 0:00

References

Zahid, A. (2020, January 24). *Smart Meter Energy(kW) demand forecasting*. Kaggle. Retrieved May 4, 2023, from https://www.kaggle.com/datasets/asimzahid/smart-meter-energykw-demand-forecasting

Description of the dataset

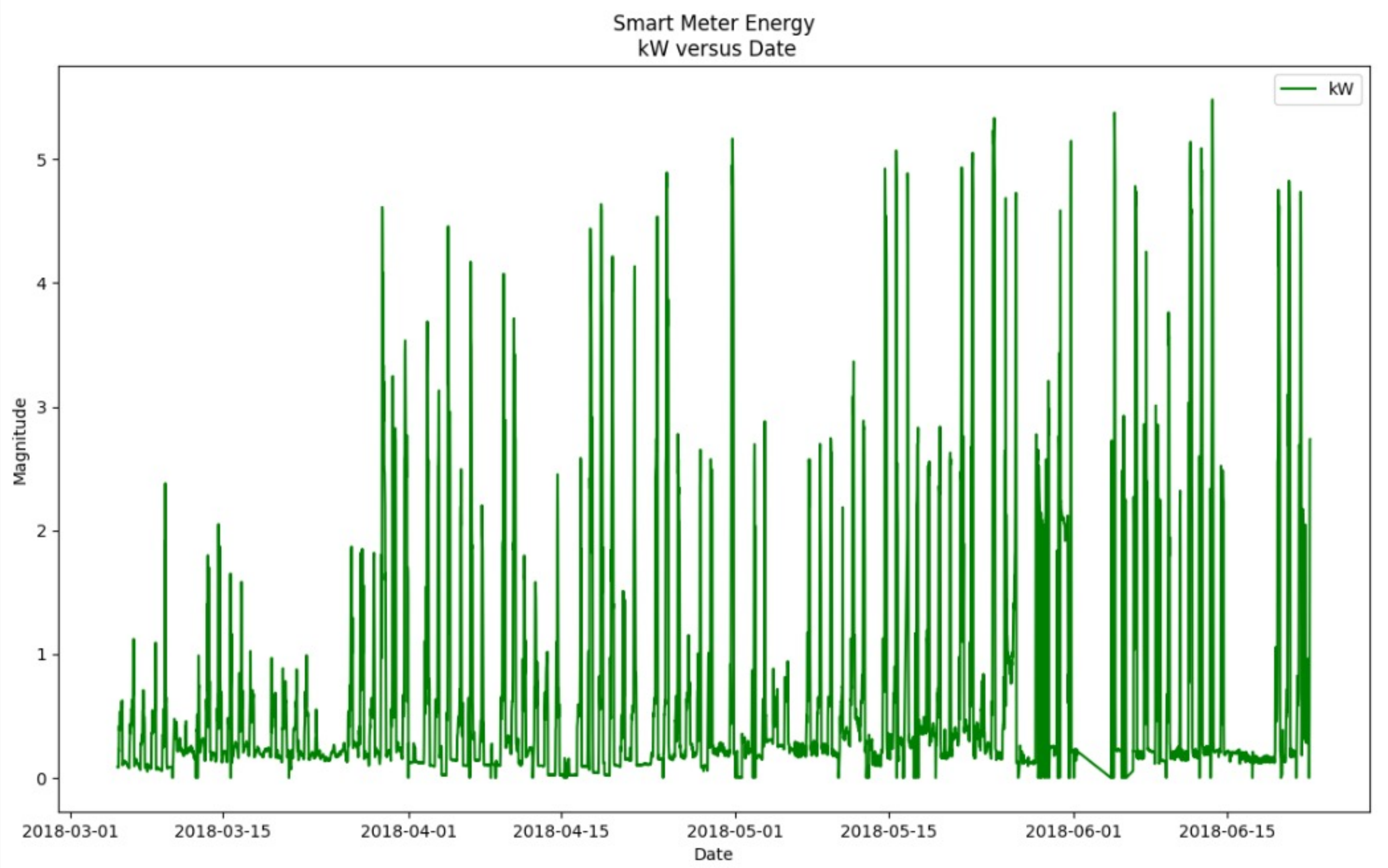

Figure 1.1 kW versus date

No obvious trend or seasonality can be seen in this plot, which needs further analysis.

Description of the dataset

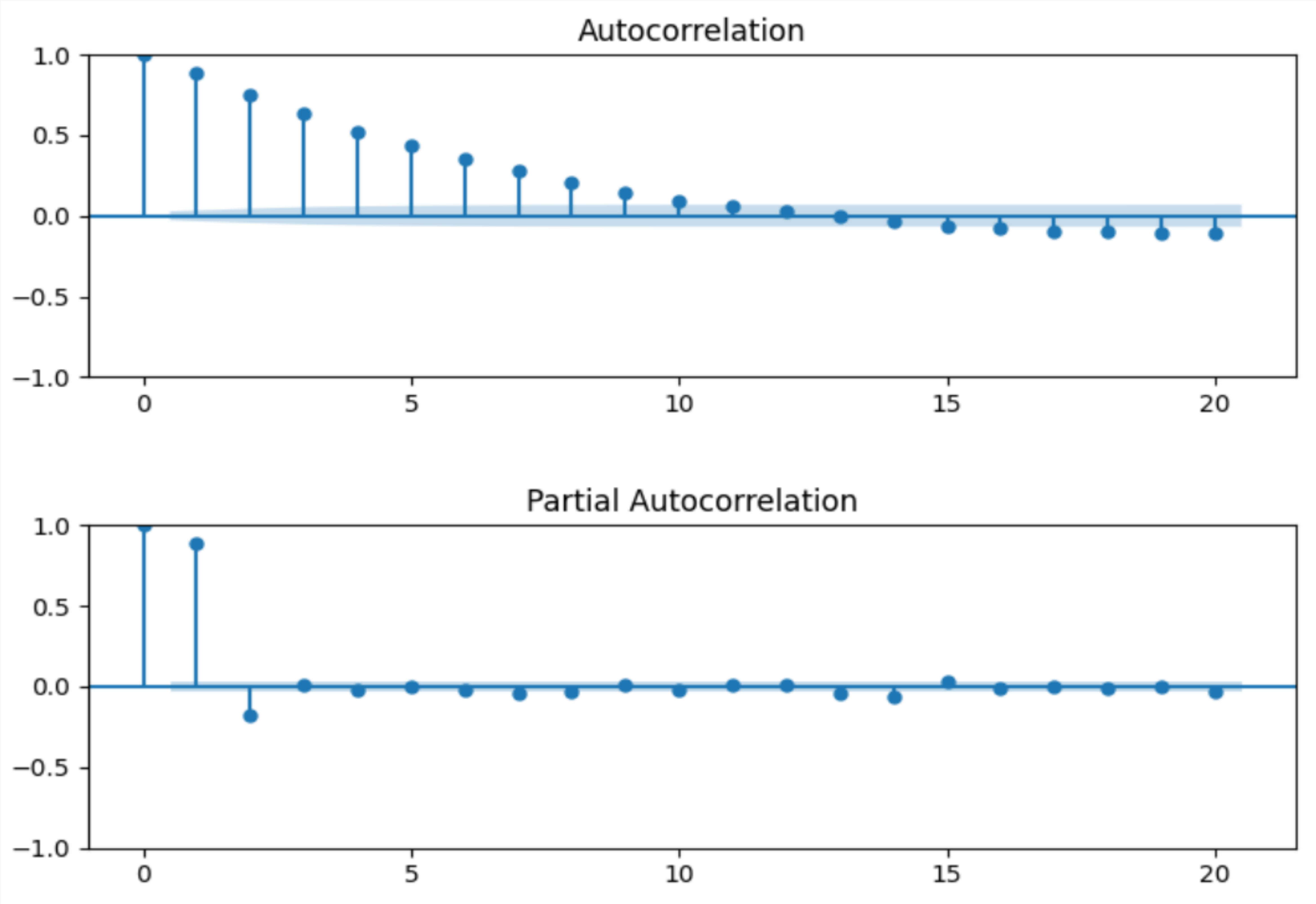

Figure 1.2 ACF/PACF of kW

The graphs suggest that ARMA (1, 0) or ARMA (2, 1) would be appropriate for the time series.

Description of the dataset

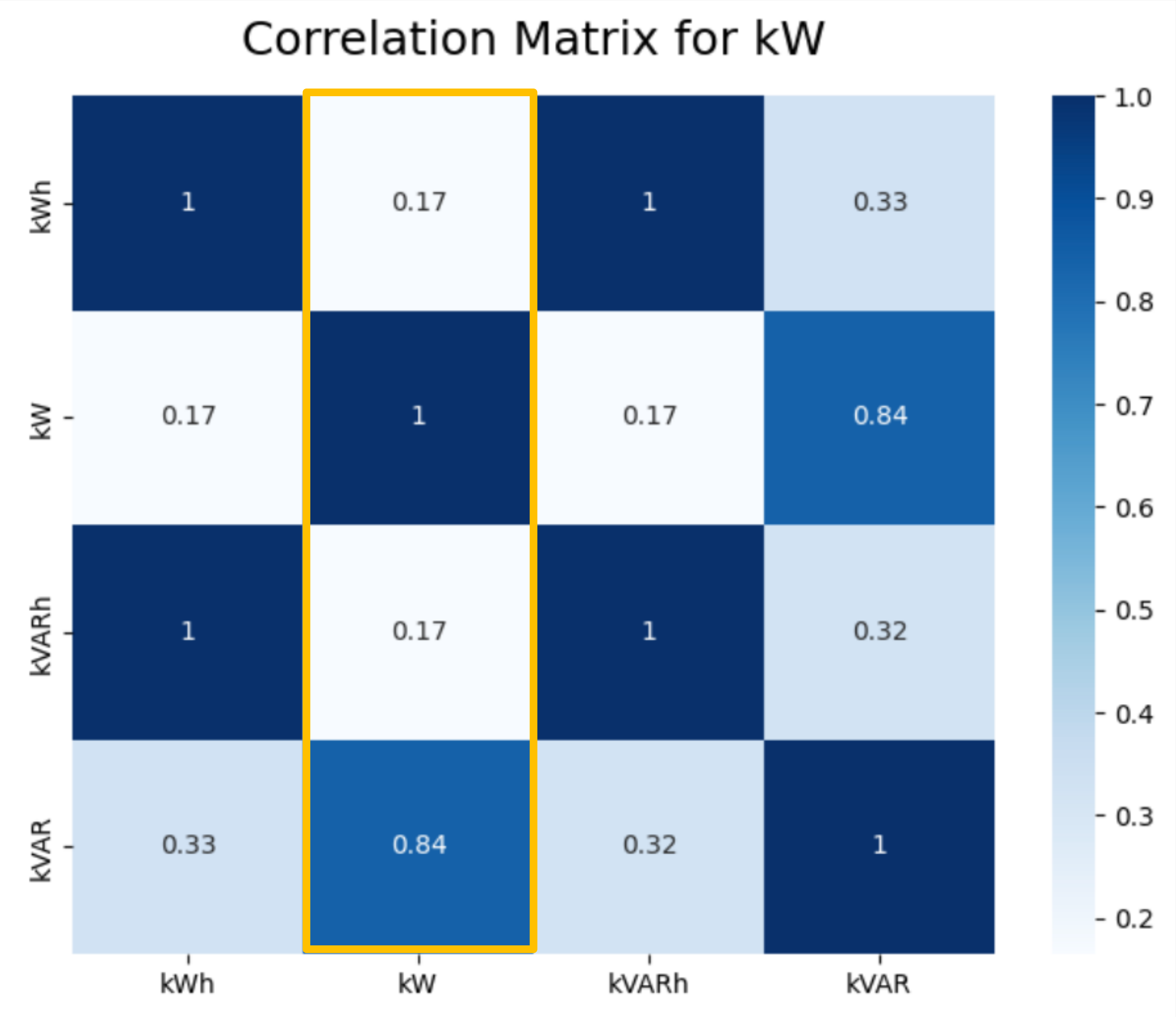

Figure 1.3 Correlation Matrix with the Pearson's correlation coefficient

Stationary
raw data

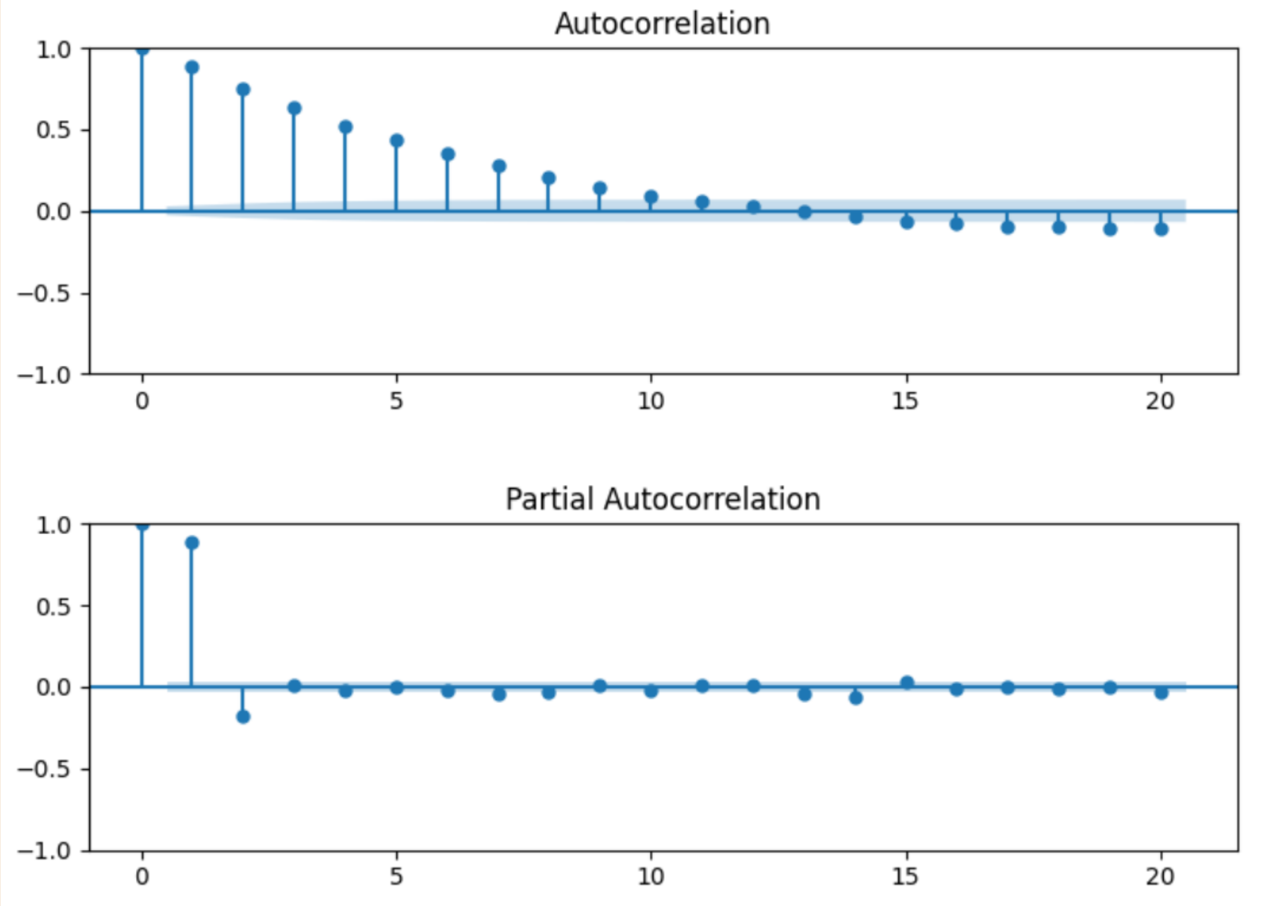

Figure 2.1 ACF/PACF of Raw Data

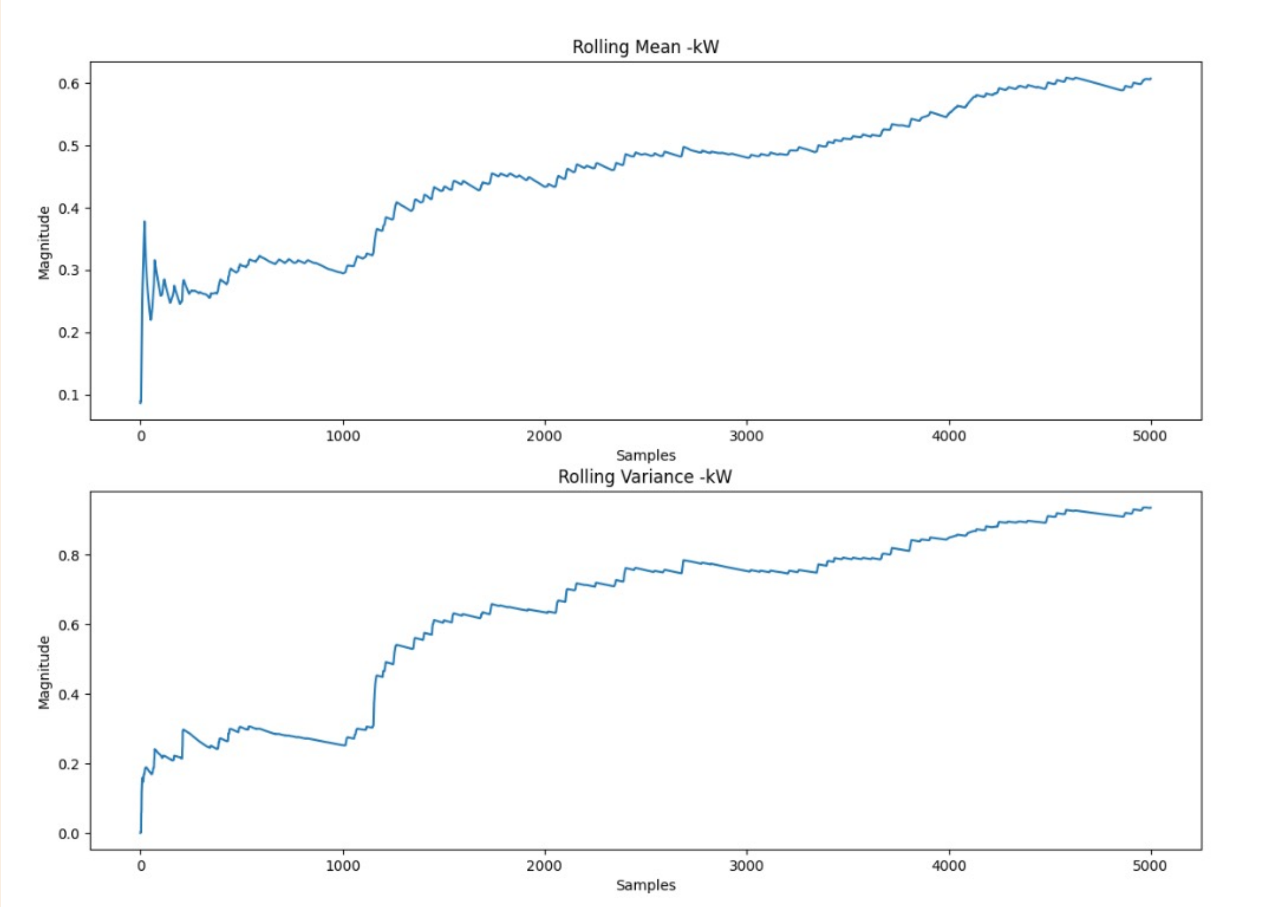

Figure 2.2 Rolling Mean and Variance of Raw Data

```
ADF Statistic: -16.17
p-value: 0.00
Critical Values:
    1%: -3.43
    5%: -2.86
    10%: -2.57
Results of KPSS Test:
Test Statistic                2.108622
p-value                       0.010000
Lags Used                    37.000000
Critical Value (10%)          0.347000
Critical Value (5%)           0.463000
Critical Value (2.5%)         0.574000
Critical Value (1%)           0.739000
dtype: float64
```

Figure 2.3 ADF and KPSS of Raw Data

The raw dataset is not stationary

Stationary

first order differencing

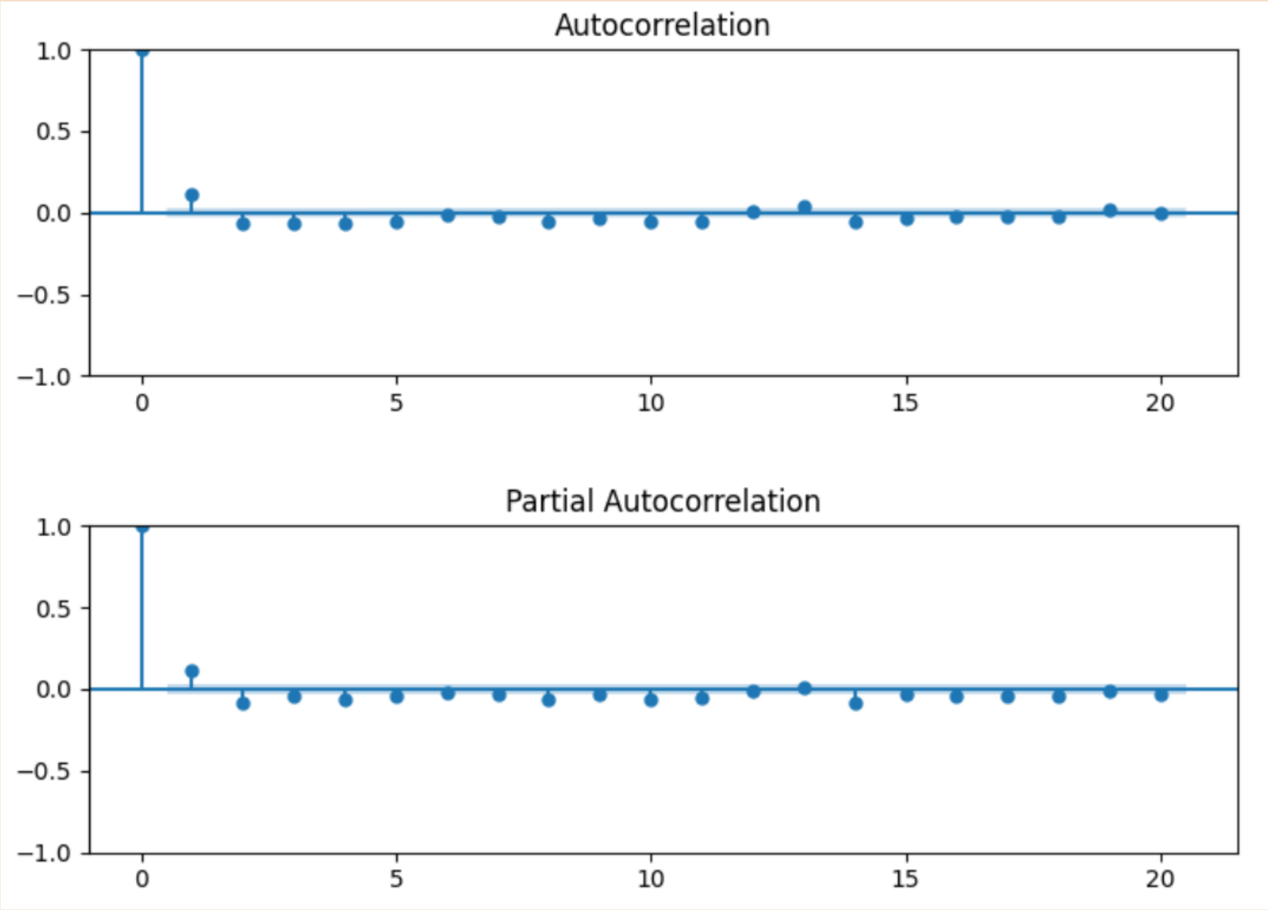

Figure 2.4 ACF/PACF of First Order Differencing

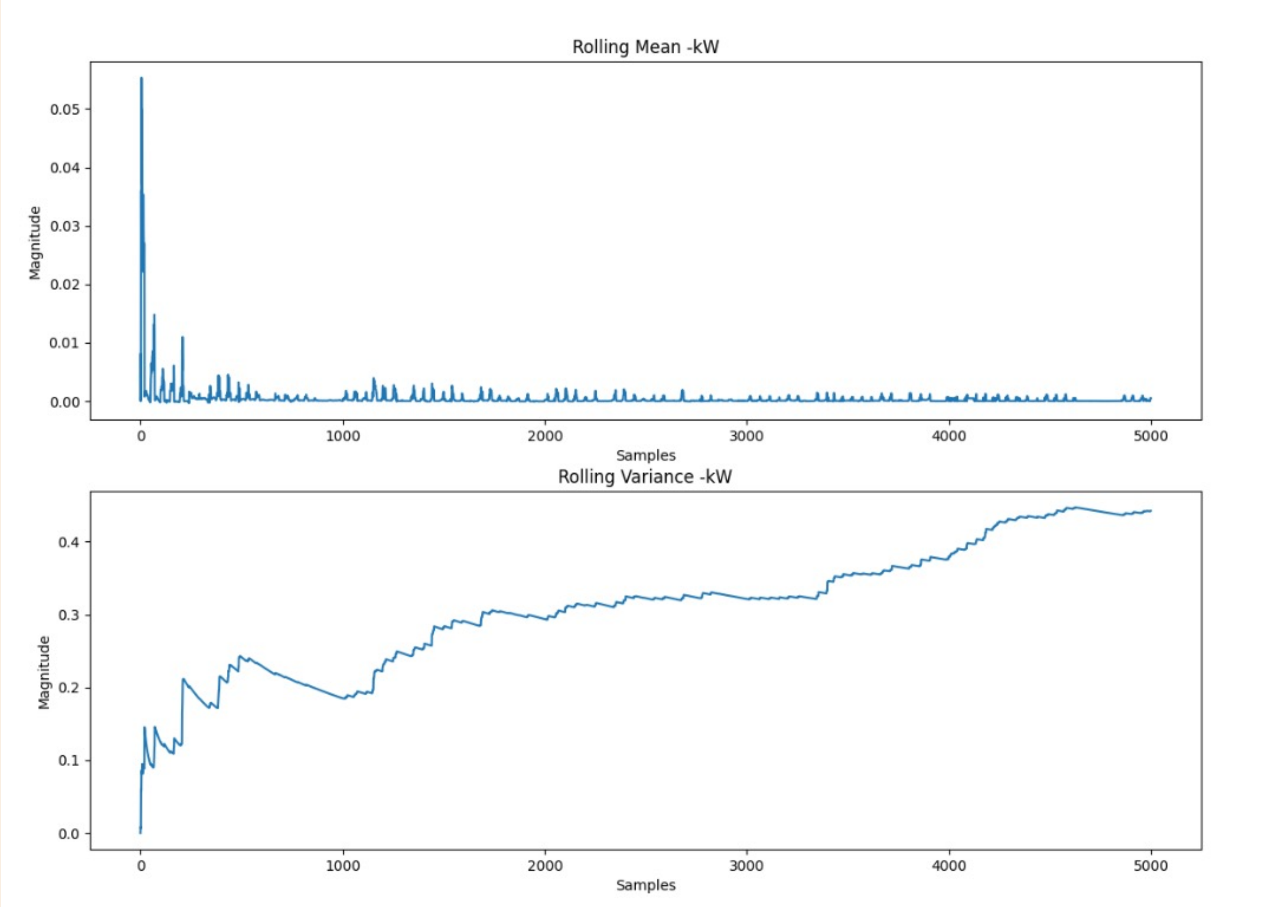

Figure 2.5 Rolling Mean and Variance of First Order Differencing

```
ADF Statistic: -20.60
p-value: 0.00
Critical Values:
    1%: -3.43
    5%: -2.86
    10%: -2.57
Results of KPSS Test:
Test Statistic              0.007547
p-value                     0.100000
Lags Used                  35.000000
Critical Value (10%)        0.347000
Critical Value (5%)         0.463000
Critical Value (2.5%)       0.574000
Critical Value (1%)         0.739000
dtype: float64
```

Figure 2.6 ADF and KPSS of First Order Differencing

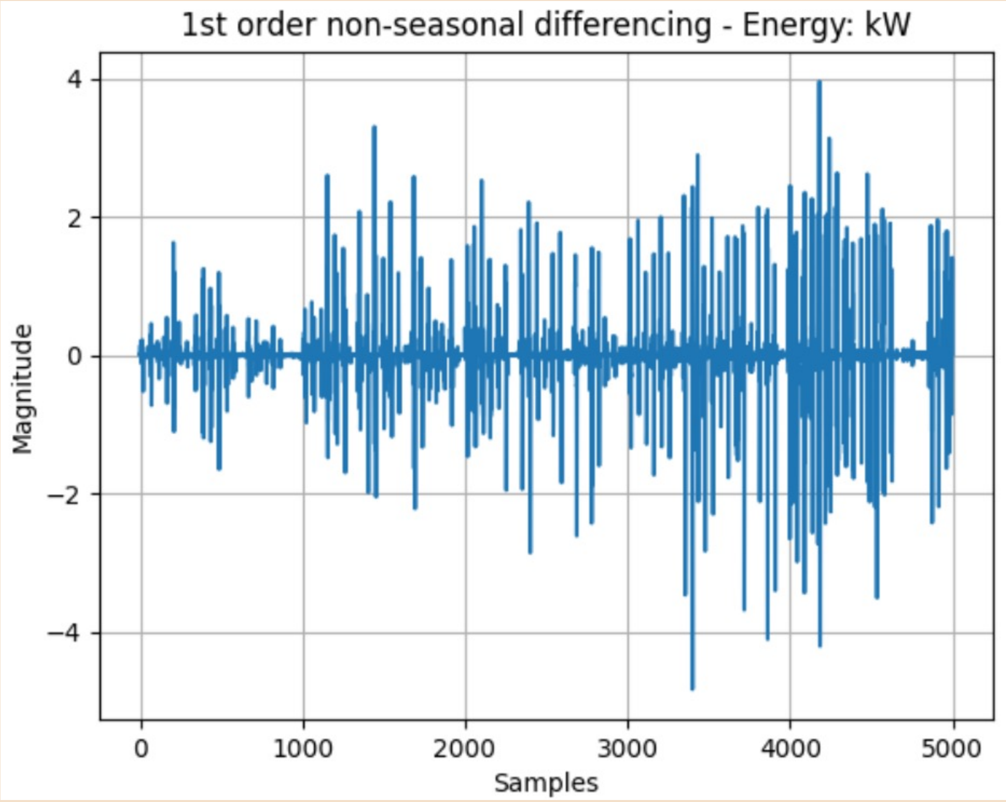

Figure 2.7 1st Oder Differencing of kW

The dataset is stationary after the transformation.

Time series Decomposition

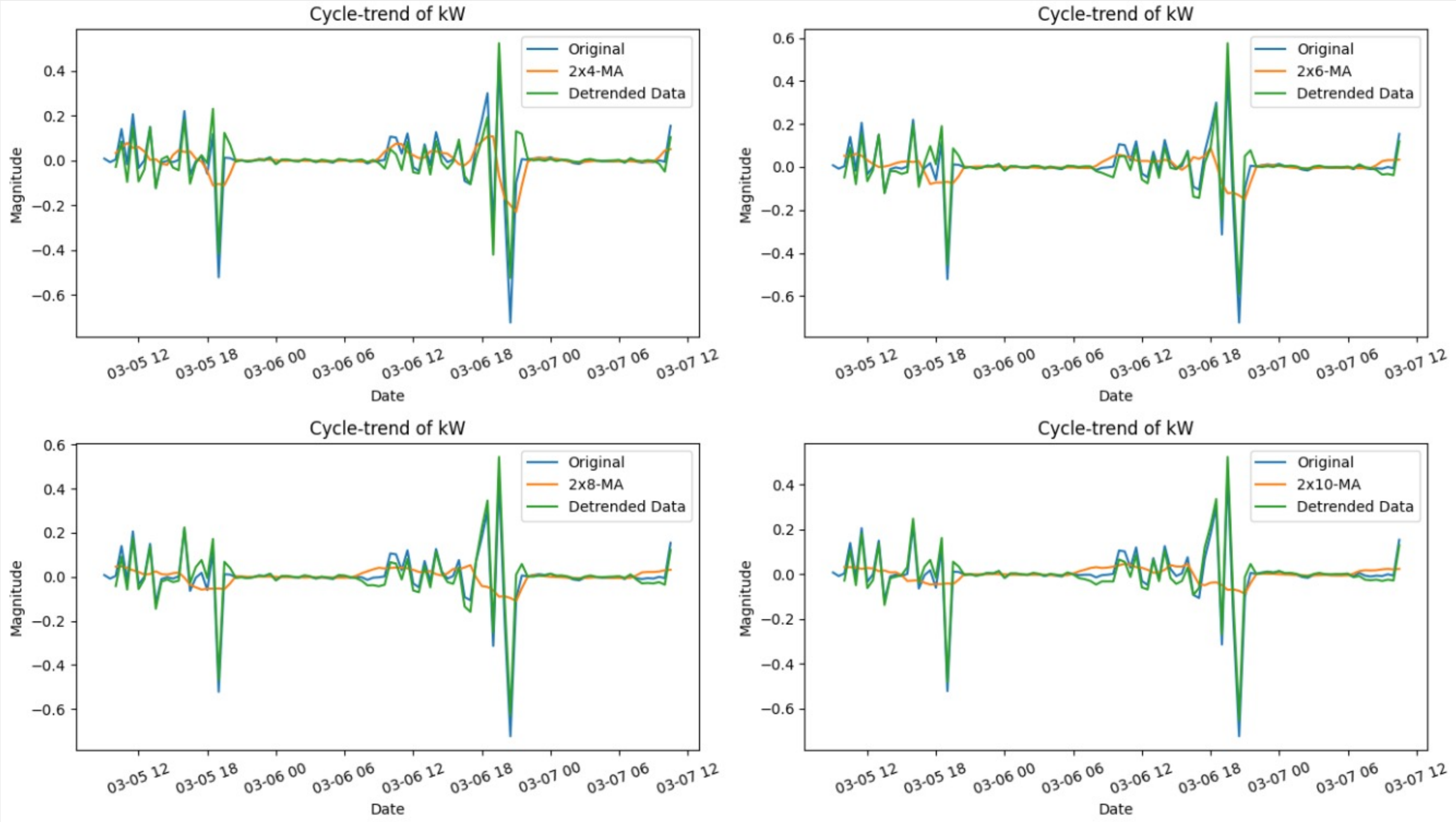

Figure 3.1 Cycle-Trend of kW

There seems no trend-cycle in the data after differencing

Time series Decomposition

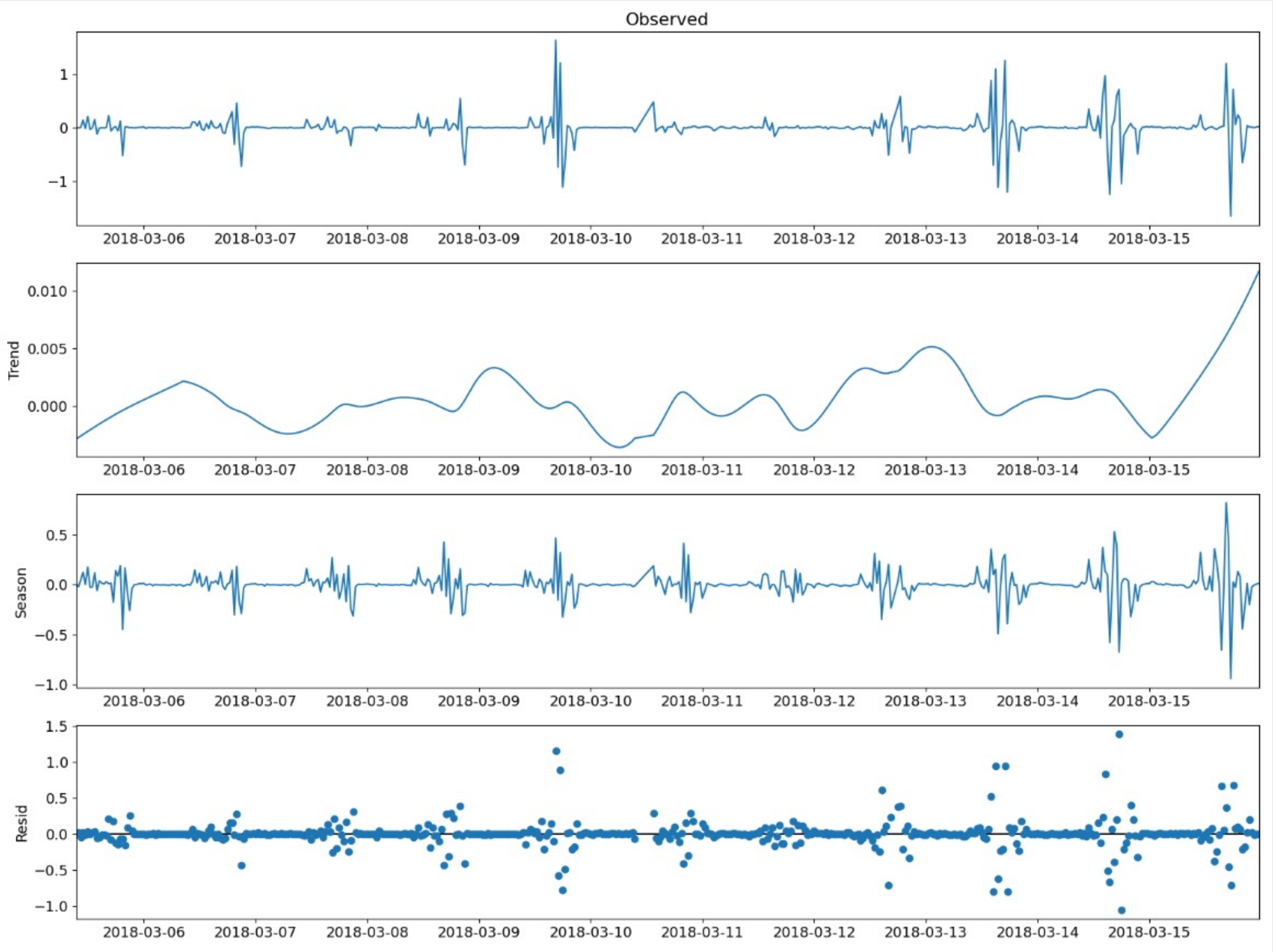

Figure 3.2 STL applied to kW(1)

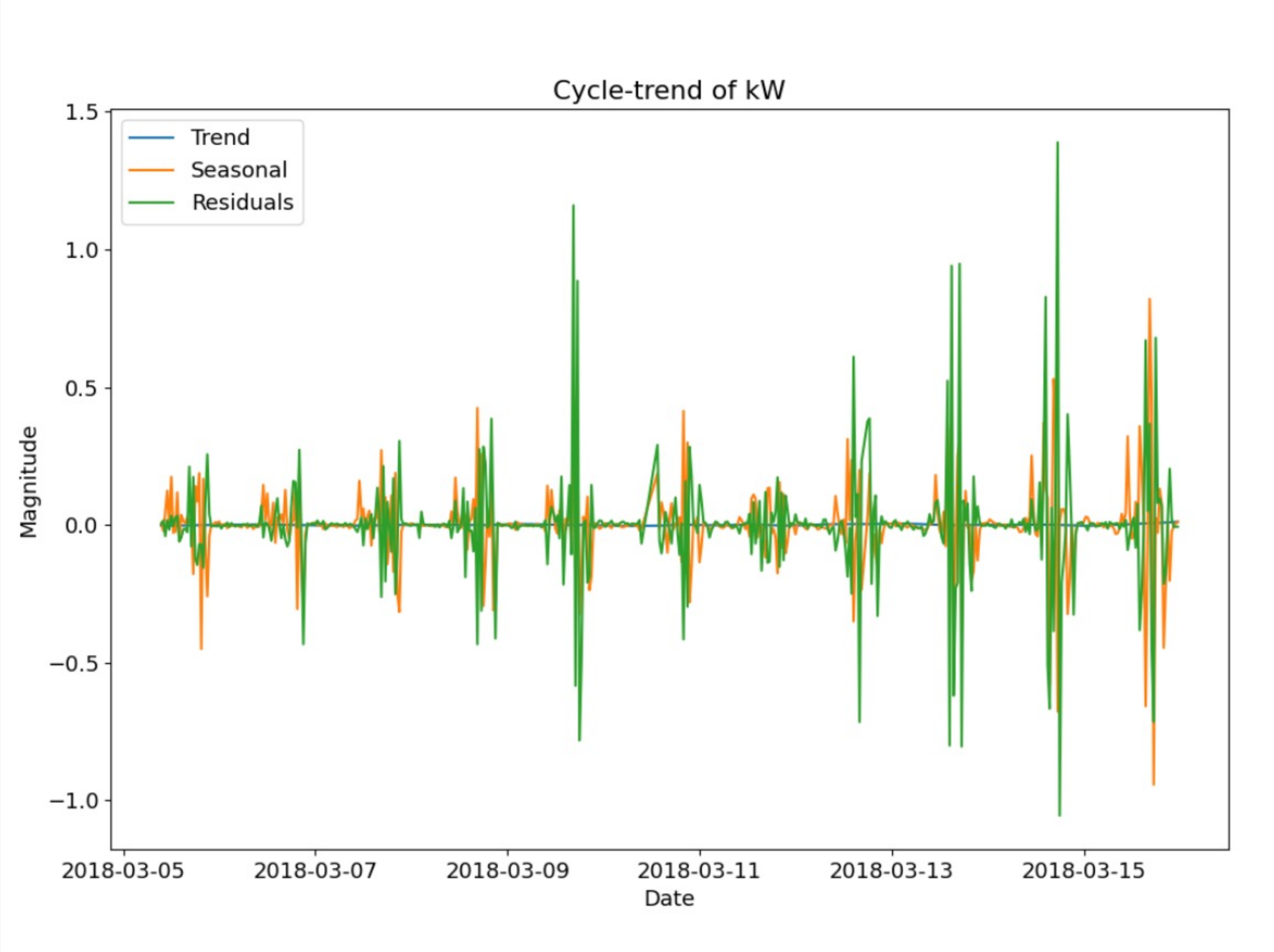

Figure 3.3 STL applied to kW(2)

The plots above show the STL decomposition where the seasonal component repeated patterns in the data that occur at fixed intervals and changed slightly over time. On the other hand, there is little to no trend in the data.

Time series Decomposition

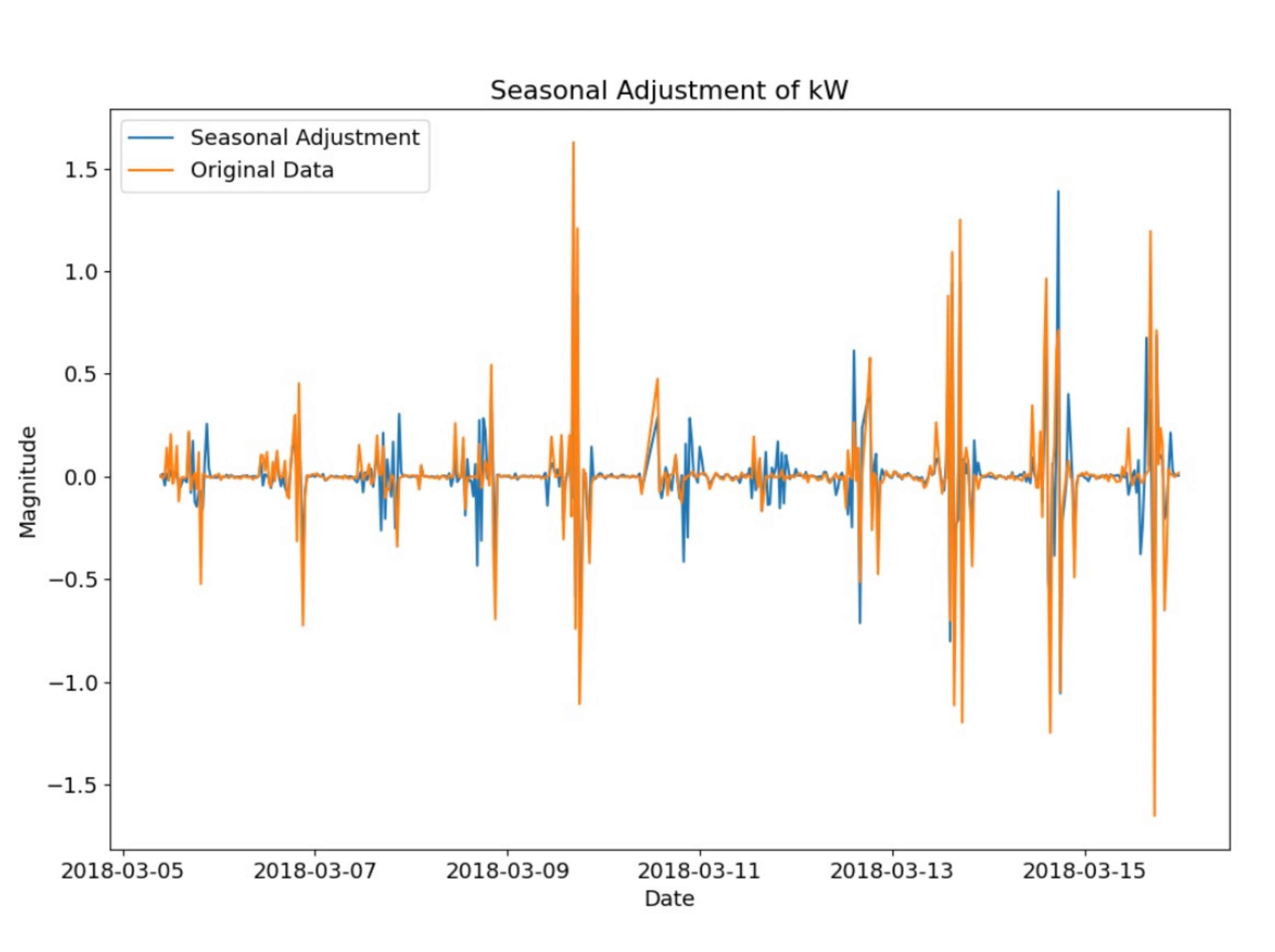

Figure 3.4 Seasonal Adjusted of kW

After first-order differencing, the strength of its trend and seasonality became weak

The strength of trend for this data set is 0.0
The strength of seasonality for this data set is 0.349

Feature Selection

	Backward Stepwise Regression	VIF	PCA	Simple Linear Regression
Selected Features	kVAR, kWh	kVAR, kVARh	kWh, kVAR	kVAR
AIC	3943.71	4055.25	-7105.85	7313.22
BIC	3968.88	4067.83	-7086.30	7326.26
Adjusted R-squared	0.78	0.84	0.72	0.71
SVD Analysis (Singular Values)	[5.73950677e+04 1.51333983e+01]	[2.32236685e+04 1.53012682e+01]	[5.73950677e+04 1.51333983e+01]	NA
Condition Number (k)	3792.6093	1517.761	3792.6093	1
Collinearity	v	v	v	x

Table 2 Feature Selection

Holt-Winters method

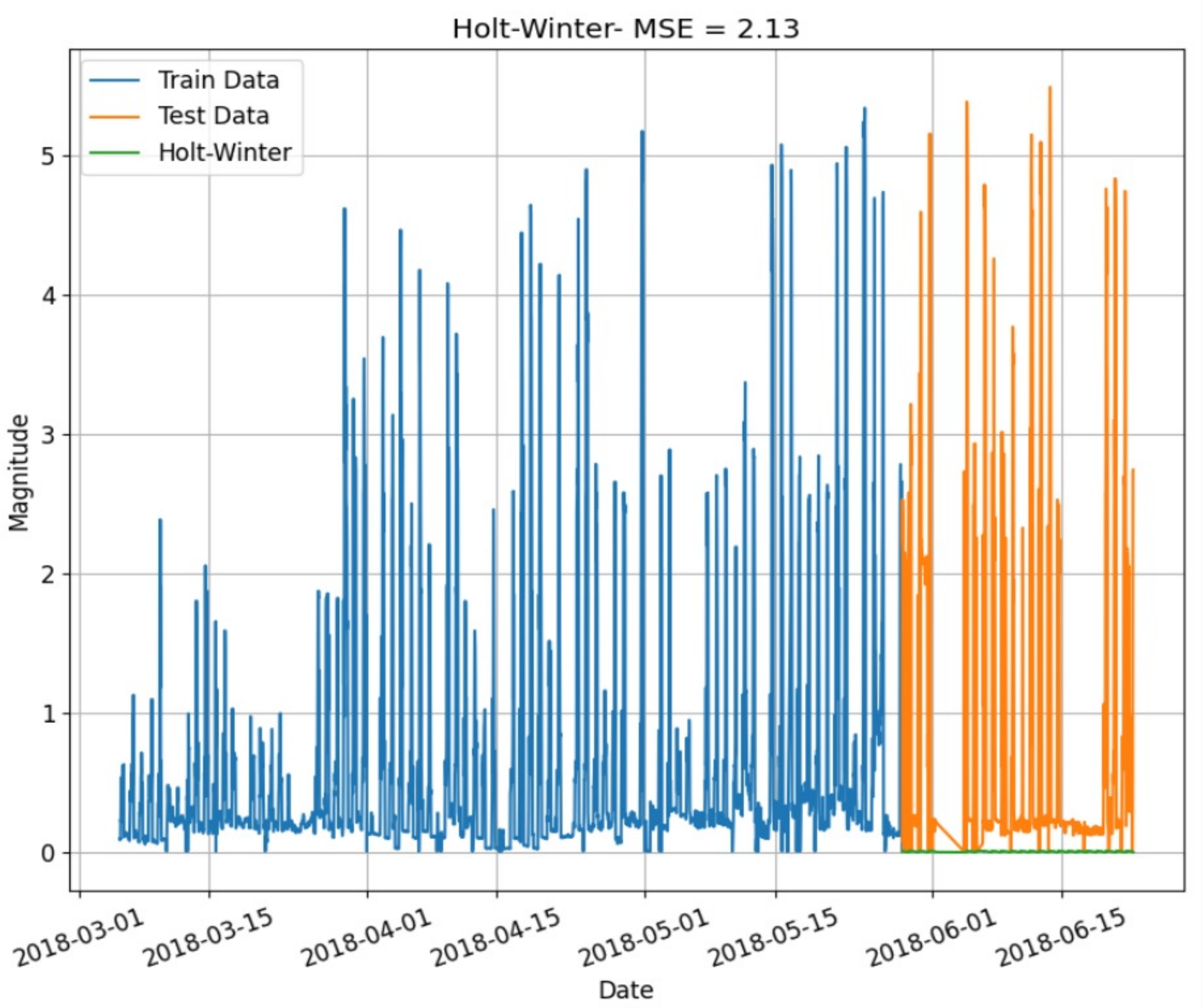

Figure 4.1 Holt-Winter Method

Mean square error for Holt-Winter is 2.13
Residual Mean: 0.8359
Residual Variance: 1.4276

Base–models

Average

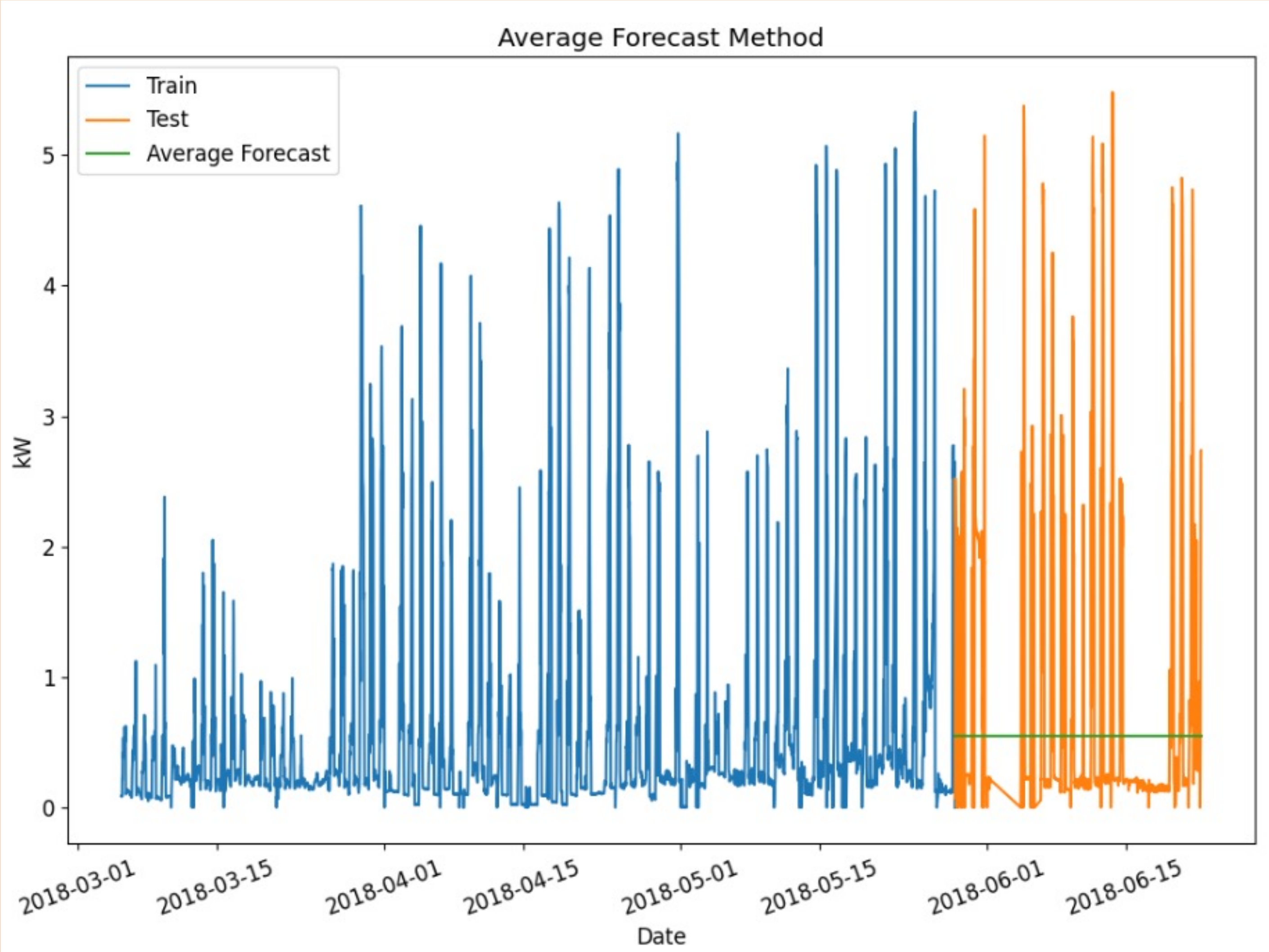

Figure 4.2 Average Forecast Method

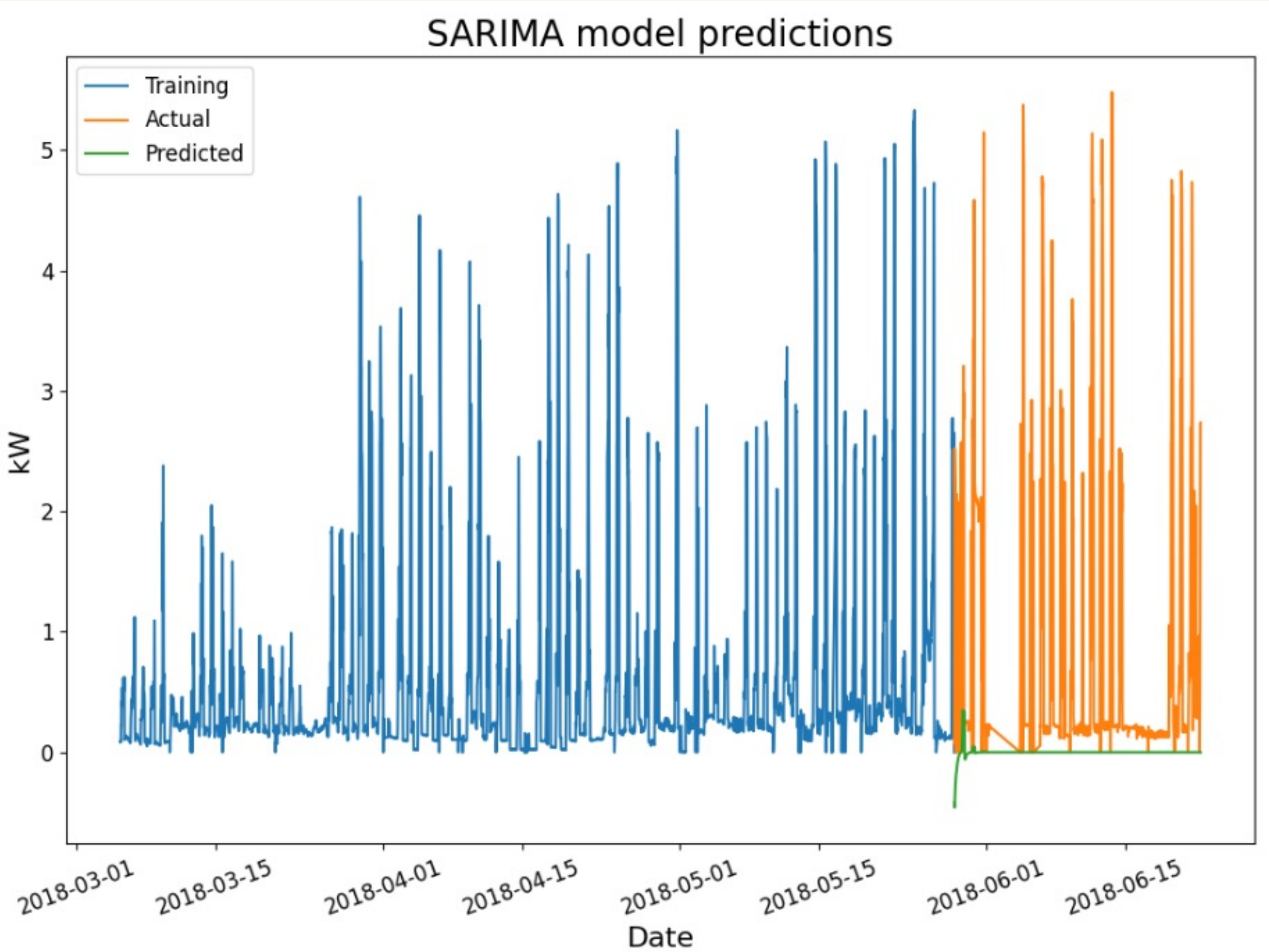

Figure 4.3 ARIMA(2, 0, 2) Model

RMSE: 1.2284
Residual Mean: 0.2835
Residual Variance: 1.4286

RMSE: 0.42
Residual Mean: 0.0000
Residual Variance: 0.1785

Base-models

Naïve

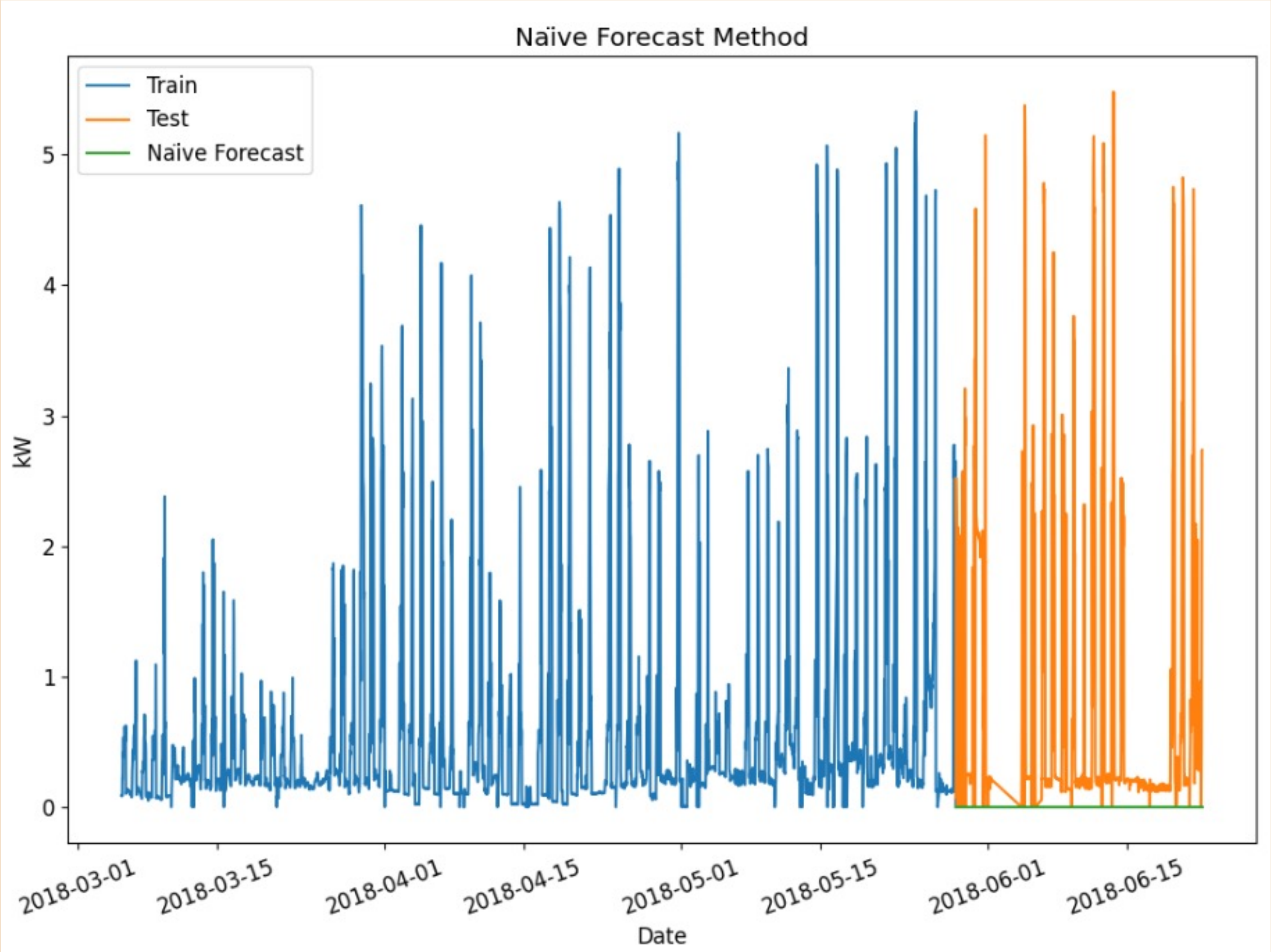

Figure 4.4 Naïve Forecast Method

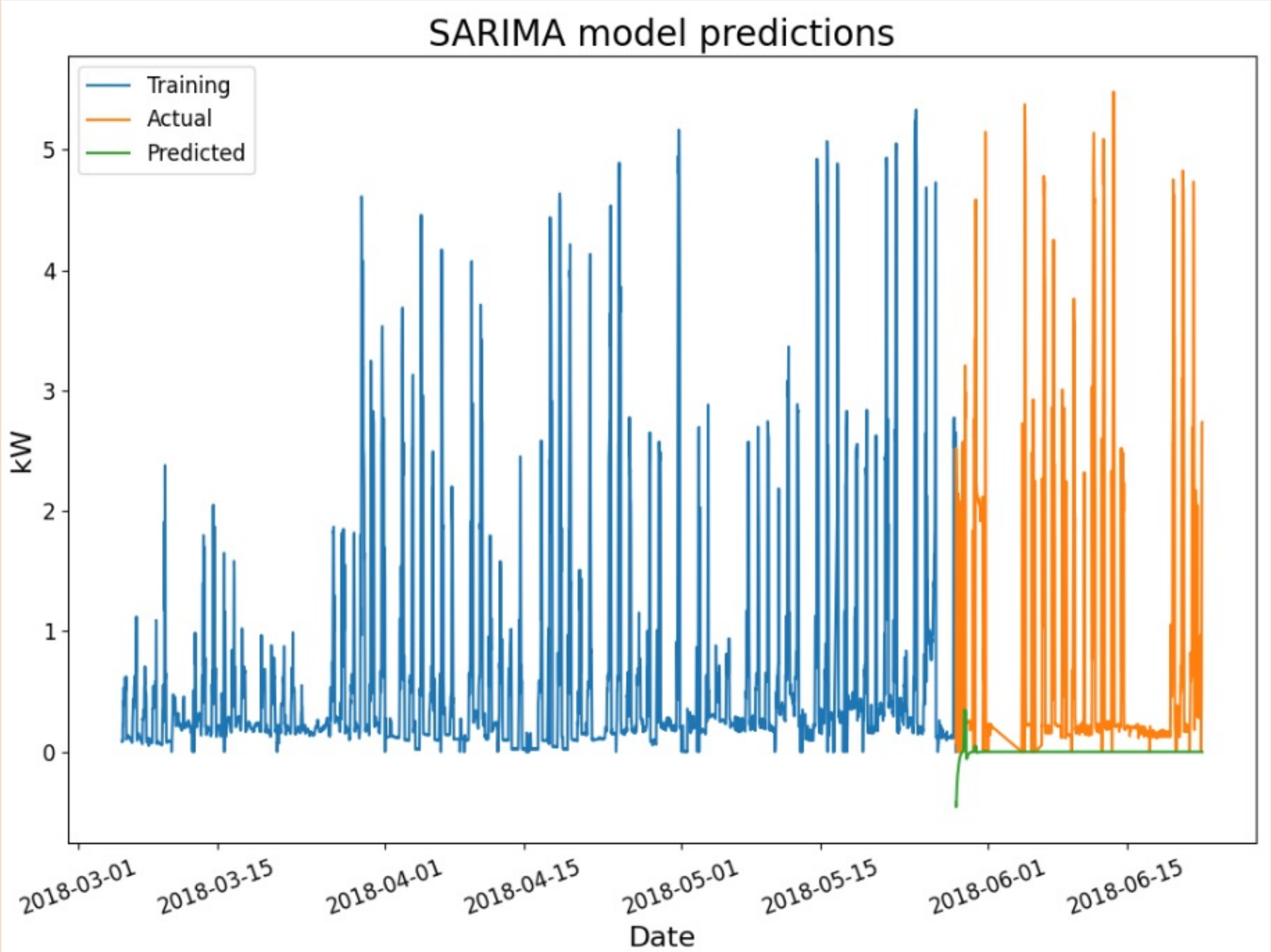

Figure 4.3 ARIMA(2, 0, 2) Model

RMSE: 1.4575
Residual Mean: 0.8342
Residual Variance: 1.4286

RMSE: 0.42
Residual Mean: 0.0000
Residual Variance: 0.1785

Base–models
Drift

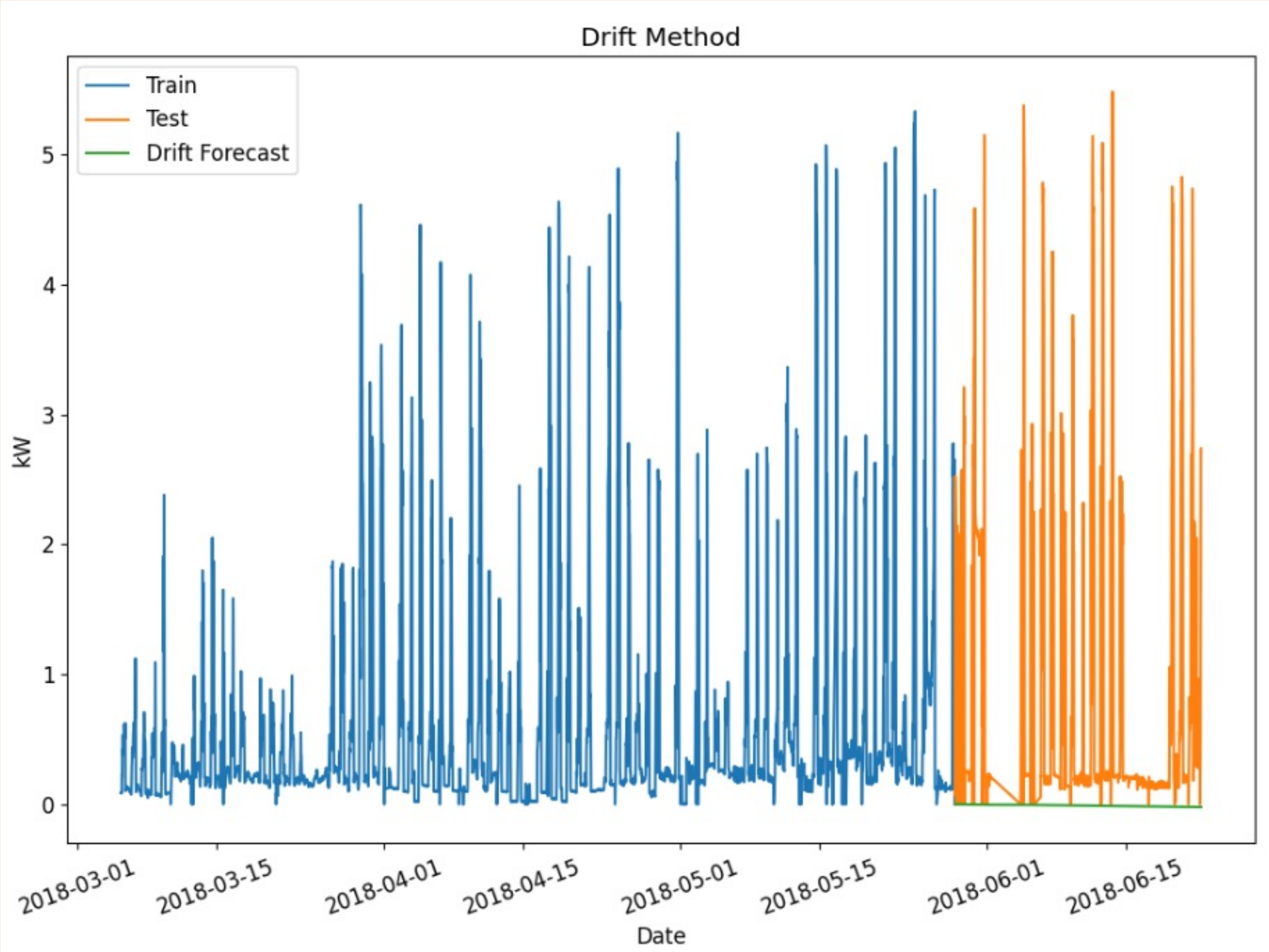

Figure 4.5 Drift Method

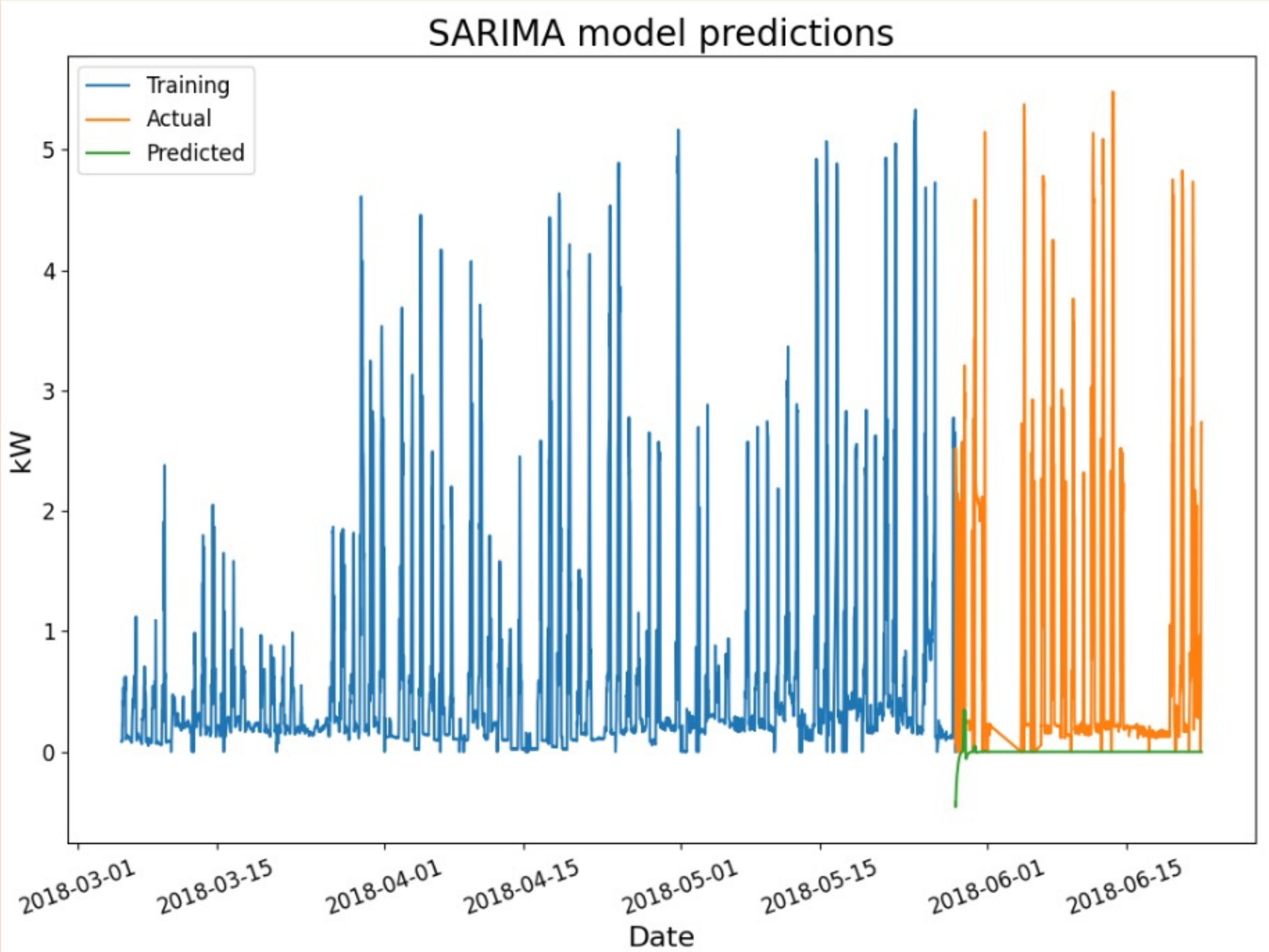

Figure 4.3 ARIMA(2, 0, 2) Model

RMSE: 1.4631
Residual Mean: 0.8449
Residual Variance: 1.4266

RMSE: 0.42
Residual Mean: 0.0000
Residual Variance: 0.1785

Base–models

SES

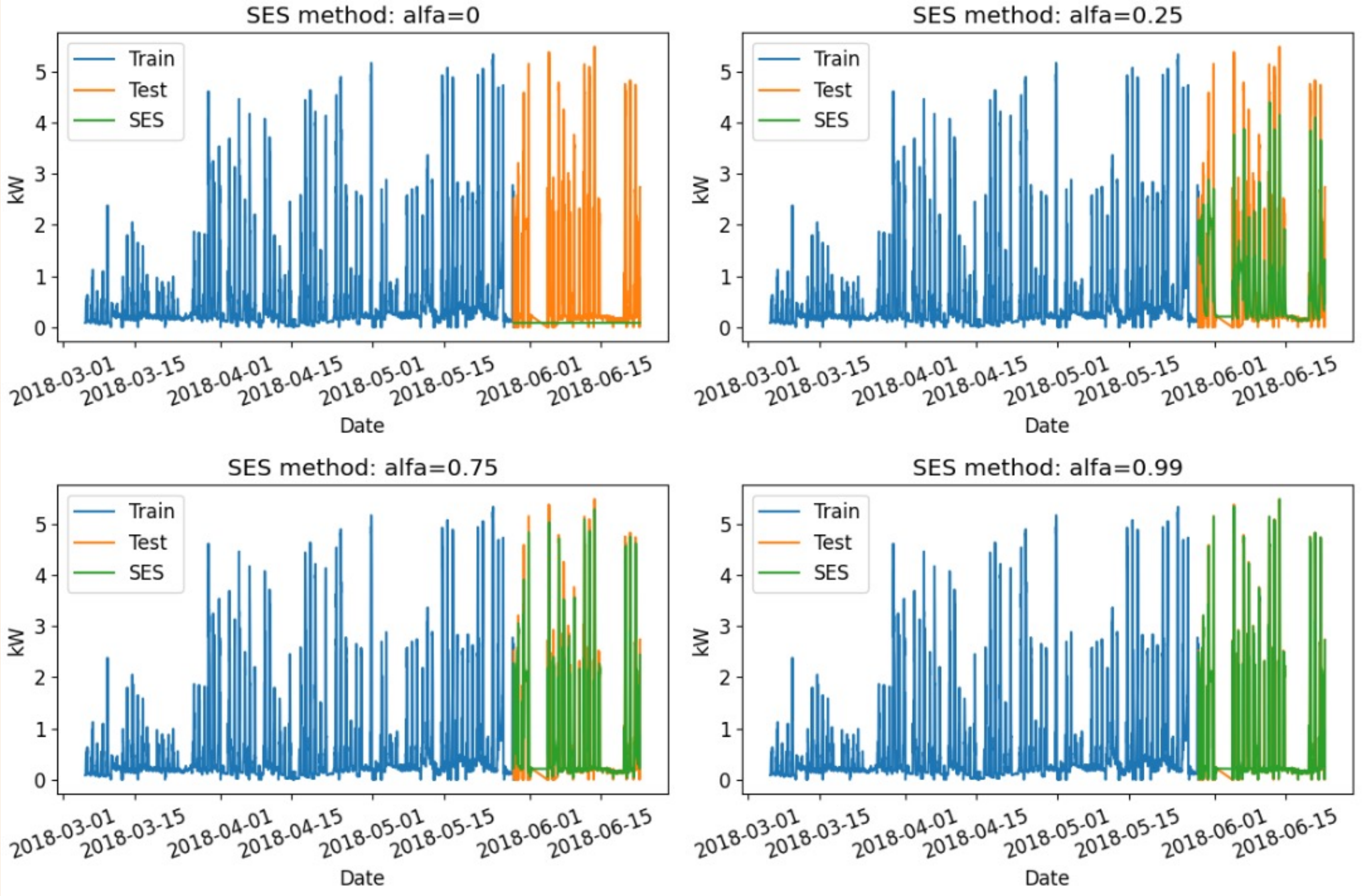

Figure 4.6 SES Method

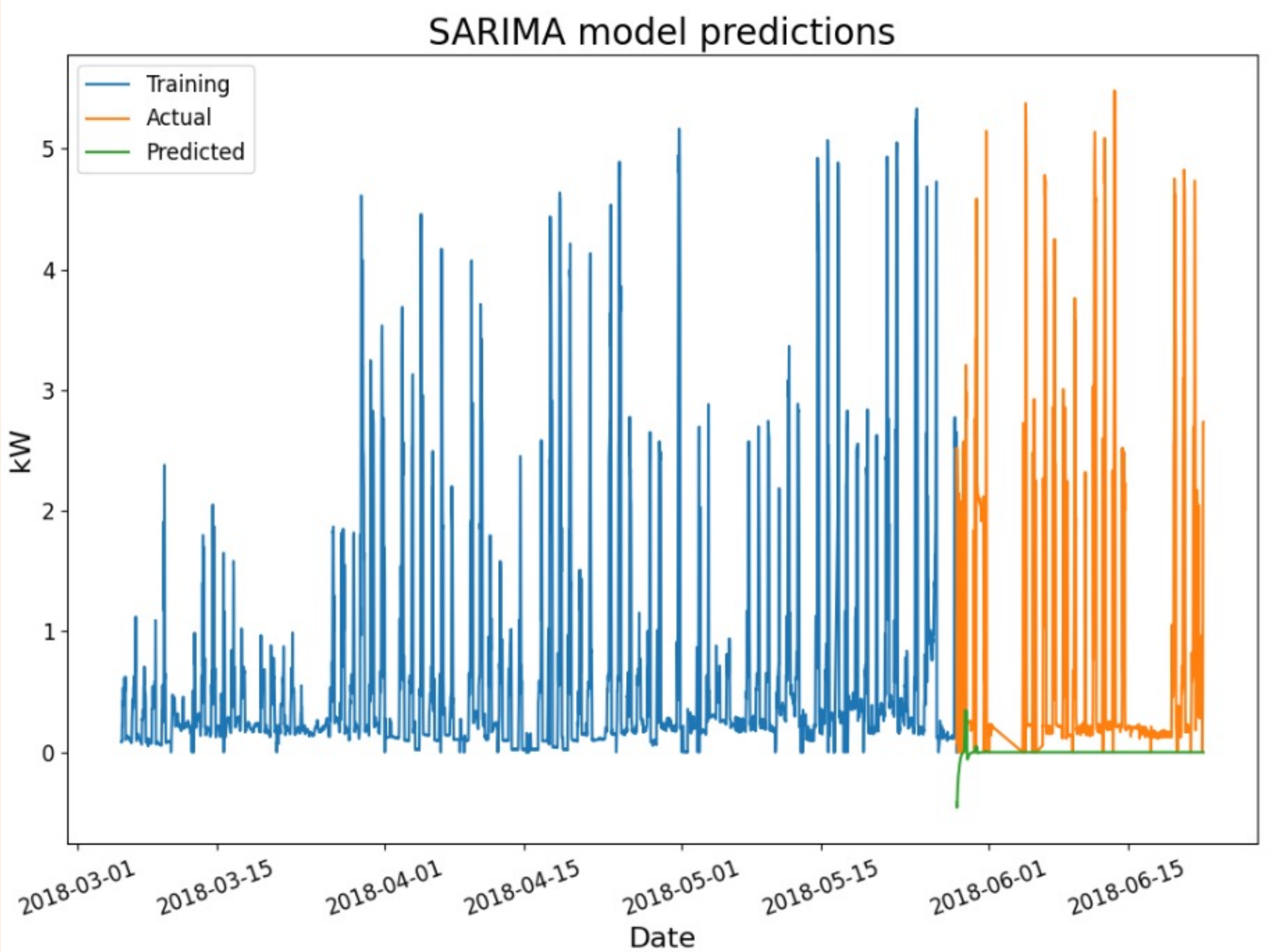

Figure 4.3 ARIMA(2, 0, 2) Model

RMSE: 1.4101
Residual Mean: 0.7482
Residual Variance: 1.4286

RMSE: 1.6559
Residual Mean: 0.0027
Residual Variance: 2.7421

RMSE: 1.5157
Residual Mean: -0.0008
Residual Variance: 2.2974

RMSE: 1.6882
Residual Mean: 0.0027
Residual Variance: 2.8500

RMSE: 0.42
Residual Mean: 0.0000
Residual Variance: 0.1785

Linear Regression

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     kW   R-squared:                       0.780
Model:                            OLS   Adj. R-squared:                  0.780
Method:                 Least Squares   F-statistic:                     7104.
Date:                Tue, 09 May 2023   Prob (F-statistic):               0.00
Time:                        00:33:24   Log-Likelihood:                -1985.1
No. Observations:                4000   AIC:                             3976.
Df Residuals:                    3997   BIC:                             3995.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.1046      0.012     -9.042      0.000      -0.127      -0.082
kVAR           3.9198      0.033    117.245      0.000       3.854       3.985
kVARh         -0.0012   5.05e-05    -24.434      0.000      -0.001      -0.001
==============================================================================
Omnibus:                     1596.327   Durbin-Watson:                   0.319
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            13804.747
Skew:                           1.667   Prob(JB):                         0.00
Kurtosis:                      11.469   Cond. No.                     1.25e+03
==============================================================================
                             Test for Constraints
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
c0            -0.1046      0.012     -9.042      0.000      -0.127      -0.082
c1             3.9198      0.033    117.245      0.000       3.854       3.985
c2            -0.0012   5.05e-05    -24.434      0.000      -0.001      -0.001
==============================================================================
<F test: F=7104.386860063783, p=0.0, df_denom=4e+03, df_num=2>
```

Figure 5.1 OLS Regression Results(1)

- Independent Variables: kVAR, kVARh
- RMSE: 0.8312
- The residual is NOT white
- lb_stat lb_pvalue
  2609.156268 0.0
- Residuals variance: 0.5331
- Residuals mean: 0.3973

Linear Regression

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     kW   R-squared:                       0.748
Model:                            OLS   Adj. R-squared:                  0.748
Method:                 Least Squares   F-statistic:                 1.185e+04
Date:                Tue, 09 May 2023   Prob (F-statistic):               0.00
Time:                        00:36:25   Log-Likelihood:                -2263.5
No. Observations:                4000   AIC:                             4531.
Df Residuals:                    3998   BIC:                             4544.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.2687      0.010    -26.589      0.000      -0.288      -0.249
kVAR           3.6109      0.033    108.838      0.000       3.546       3.676
==============================================================================
Omnibus:                     1249.495   Durbin-Watson:                   0.256
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             8080.724
Skew:                           1.329   Prob(JB):                         0.00
Kurtosis:                       9.436   Cond. No.                         5.19
==============================================================================
                             Test for Constraints
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
c0            -0.2687      0.010    -26.589      0.000      -0.288      -0.249
c1             3.6109      0.033    108.838      0.000       3.546       3.676
==============================================================================
<F test: F=11845.80580827514, p=0.0, df_denom=4e+03, df_num=1>
```

Figure 5.2 OLS Regression Results(2)

- Independent Variables: kVAR
- RMSE: 0.7376
- The residual is NOT white
- lb_stat lb_pvalue
  2417.416218 0.0
- Residuals variance: 0.5425
- Residuals mean: -0.0393

ARMA, ARIMA and SARIMA model

ARMA model order determination

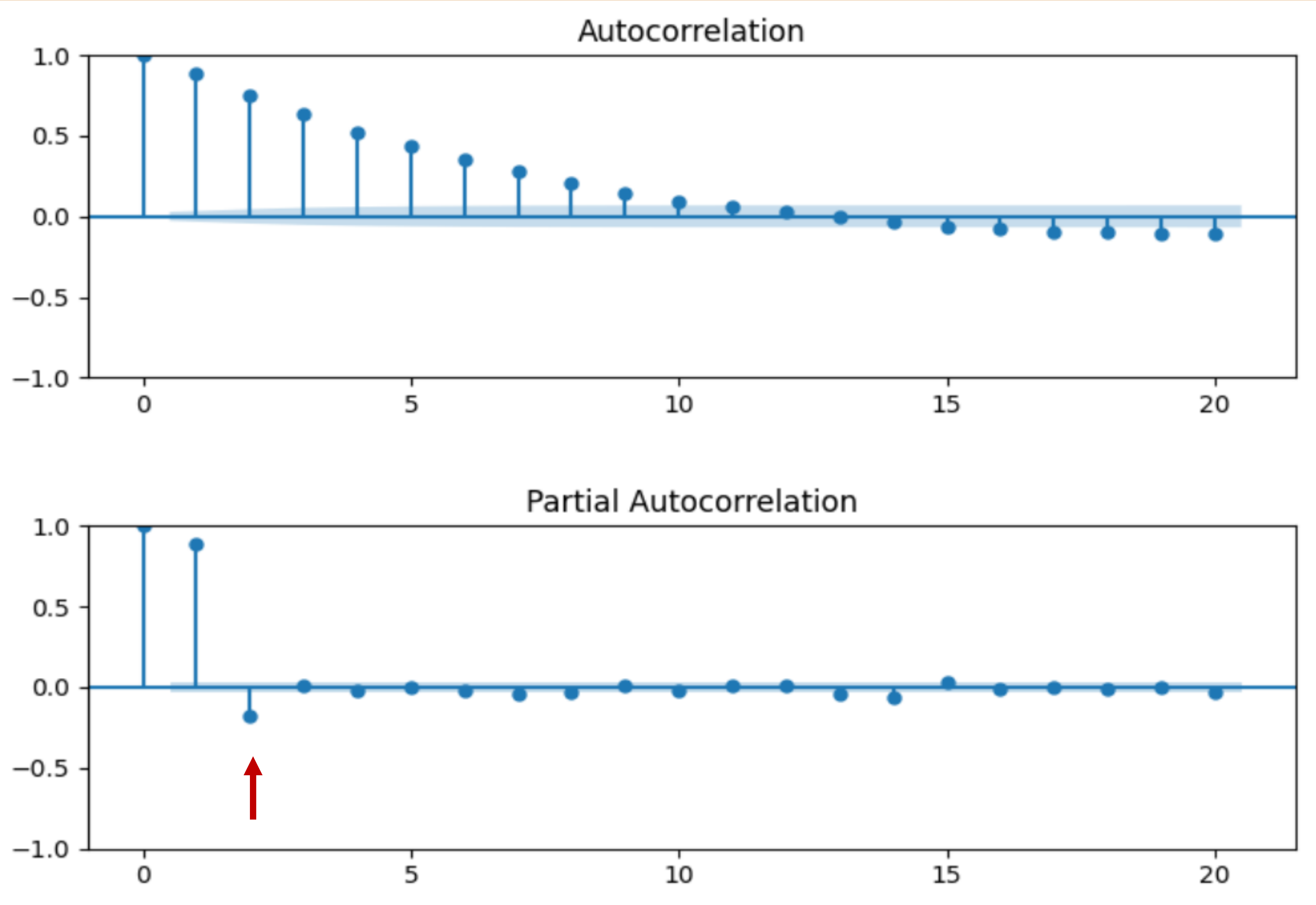

Figure 1.2 ACF/PACF of kW

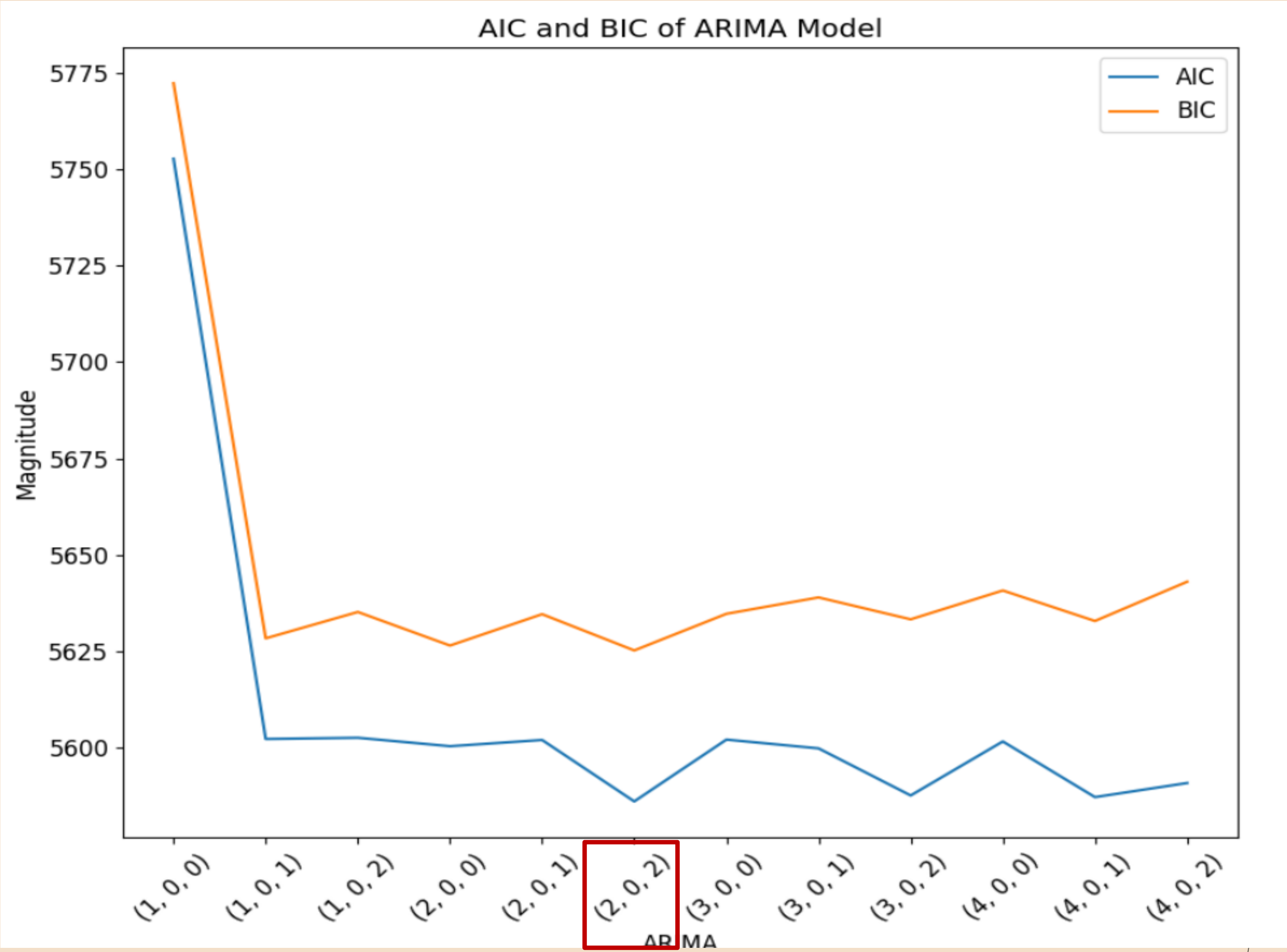

Figure 6.1 AIC and BIC of ARIMA Model

ARMA, ARIMA and SARIMA model

ARMA model order determination

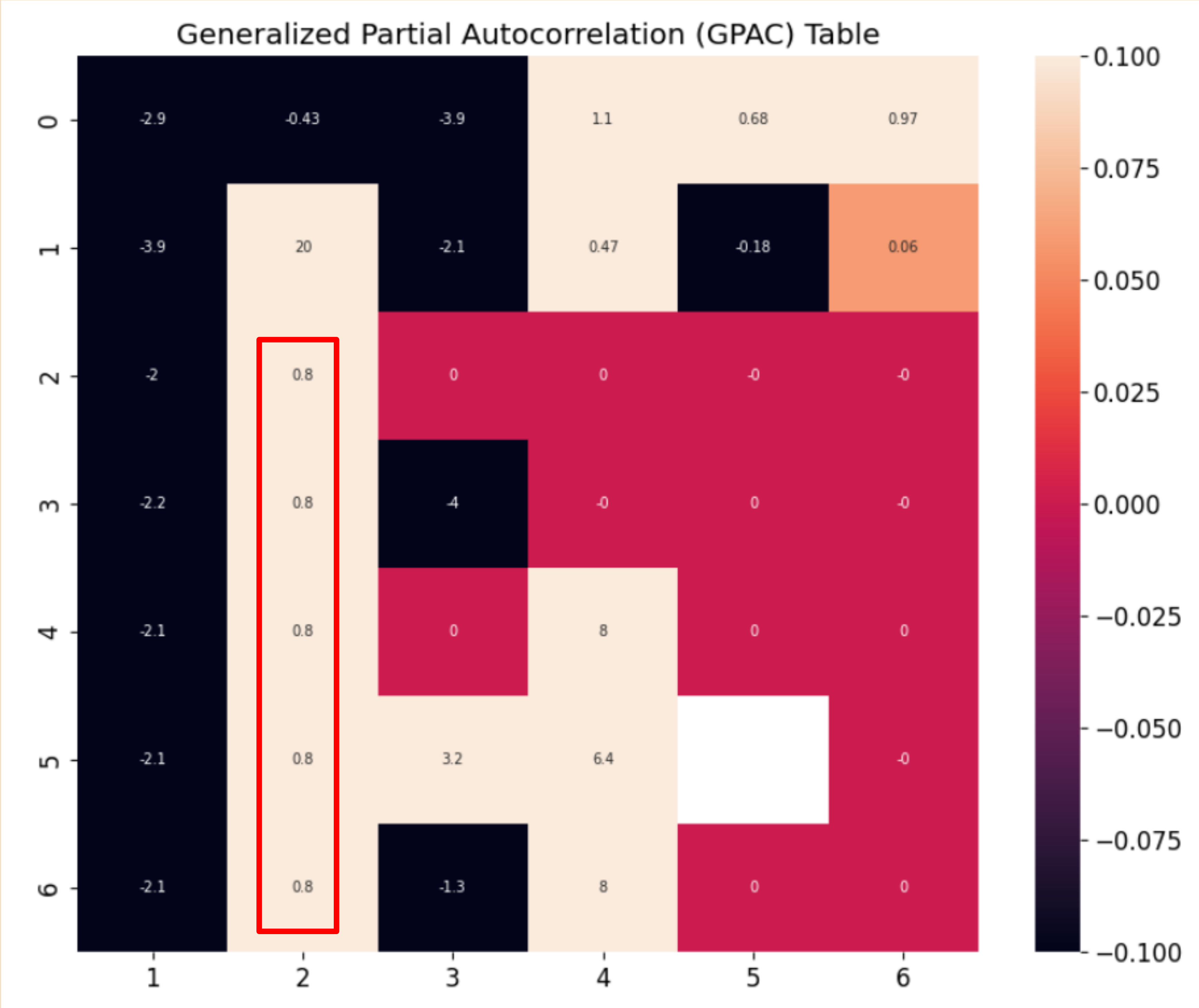

Figure 6.2 GPAC Table

Third, by GPAC table, we can estimate the ARMA order

ARMA, ARIMA and SARIMA model

ARMA model summary

Figure 6.3 ARMA model summary

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                     kW   No. Observations:                 5000
Model:                 ARIMA(2, 0, 2)   Log Likelihood               -2787.078
Date:                Tue, 09 May 2023   AIC                           5586.156
Time:                        03:17:25   BIC                           5625.260
Sample:                             0   HQIC                          5599.861
                               - 5000
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.6087      0.064      9.577      0.000       0.484       0.733
ar.L1          1.7778      0.039     45.997      0.000       1.702       1.854
ar.L2         -0.7970      0.032    -24.645      0.000      -0.860      -0.734
ma.L1         -0.7352      0.039    -18.673      0.000      -0.812      -0.658
ma.L2         -0.1495      0.012    -12.562      0.000      -0.173      -0.126
sigma2         0.1785      0.001    124.765      0.000       0.176       0.181
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):             95221.64
Prob(Q):                              0.94   Prob(JB):                         0.00
Heteroskedasticity (H):               4.60   Skew:                             0.56
Prob(H) (two-sided):                  0.00   Kurtosis:                        24.35
===================================================================================
```

ARMA(2,2)
AIC: 5586
BIC: 5625
RMSE: 0.42
Residual Mean: 0.0000
Residual Variance: 0.1785

Figure 6.4 ARIMA model summary

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                     kW   No. Observations:                 5000
Model:                 ARIMA(2, 1, 2)   Log Likelihood               -2794.049
Date:                Tue, 09 May 2023   AIC                           5598.097
Time:                        03:18:29   BIC                           5630.682
Sample:                             0   HQIC                          5609.518
                               - 5000
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0011      0.045     22.284      0.000       0.913       1.089
ar.L2         -0.1409      0.040     -3.549      0.000      -0.219      -0.063
ma.L1         -0.9573      0.044    -21.782      0.000      -1.043      -0.871
ma.L2         -0.0411      0.044     -0.935      0.350      -0.127       0.045
sigma2         0.1789      0.001    150.598      0.000       0.177       0.181
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):             94831.67
Prob(Q):                              0.97   Prob(JB):                         0.00
Heteroskedasticity (H):               4.66   Skew:                             0.47
Prob(H) (two-sided):                  0.00   Kurtosis:                        24.32
===================================================================================
```

ARIMA(2,0,2)
AIC: 5598
BIC: 5631
RMSE: 0.42
Residual Mean: 0.0088
Residual Variance: 0.1789

Figure 6.5 SARIMA model summary

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                 kW   No. Observations:                 5000
Model:             SARIMAX(2, 0, 2)x(1, 0, [], 48)   Log Likelihood               -2819.123
Date:                            Tue, 09 May 2023   AIC                           5650.247
Time:                                    03:19:03   BIC                           5689.350
Sample:                                         0   HQIC                          5663.952
                                           - 5000
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.3727      0.503      0.741      0.459      -0.613       1.358
ar.L2          0.4538      0.444      1.023      0.306      -0.416       1.323
ma.L1          0.6762      0.504      1.342      0.180      -0.311       1.664
ma.L2          0.0730      0.088      0.833      0.405      -0.099       0.245
ar.S.L48       0.1241      0.008     15.725      0.000       0.109       0.140
sigma2         0.1807      0.001    159.442      0.000       0.179       0.183
===================================================================================
Ljung-Box (L1) (Q):                   1.34   Jarque-Bera (JB):             98867.05
Prob(Q):                              0.25   Prob(JB):                         0.00
Heteroskedasticity (H):               4.63   Skew:                             0.22
Prob(H) (two-sided):                  0.00   Kurtosis:                        24.78
===================================================================================
```

SARIMA(2,0,2) D = 48
AIC: 5650
BIC: 5689
RMSE: 0.43
Residual Mean: 0.0533
Residual Variance: 0.1779

Levenberg Marquardt algorithm

	Parameter Estimates	Standard Deviation	Confidence Interval 0	Confidence Interval 1
const	0.609	0.064	0.484	0.733
ar.L1	1.778	0.039	1.702	1.854
ar.L2	-0.797	0.032	-0.860	-0.734
ma.L1	-0.735	0.039	-0.812	-0.658
ma.L2	-0.149	0.012	-0.173	-0.126
sigma2	0.178	0.001	0.176	0.181

Table 3 ARMA model by Levenberg Marquardt Algorithm

Deep Learning Model

Multivariate LSTM Model

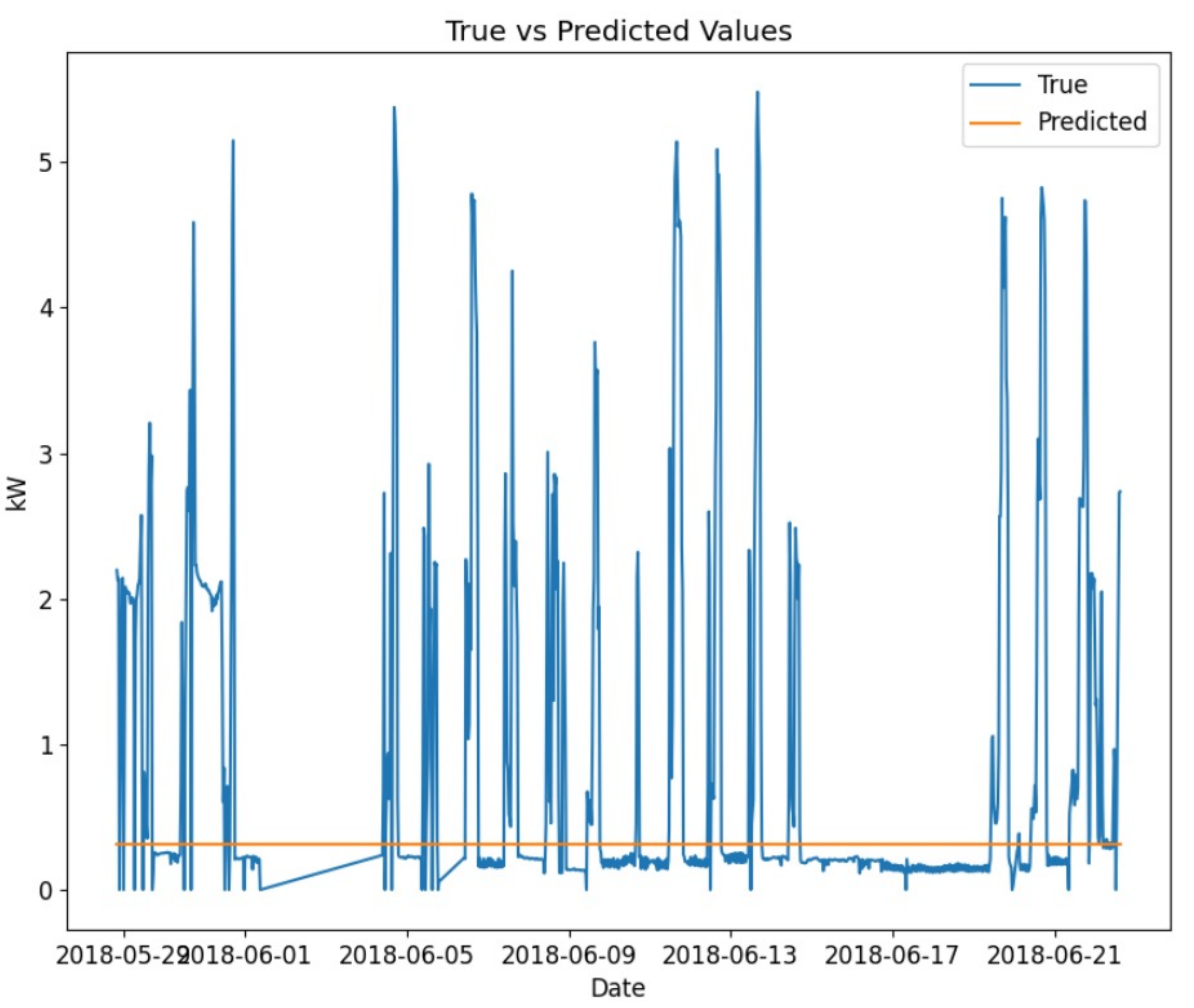

Figure 7.1 Multivariate LSTM Model

Test RMSE: 1.3098
Residual Mean: 0.5382
Residual Variance: 1.4260

Final Model selection

	Holt-Winters method	Average	naïve	drift	SES	ARIMA (2, 0, 2)	ARIMA (2, 1, 2)	SARIMA (2, 0, 2)48	LSTM
RMSE	2.13	1.2284	1.4575	1.4631	1.4101	0.42	0.42	0.43	1.299
Residual Mean	0.8359	0.2835	0.8342	0.8449	0.7482	0	0.0088	0.0533	0.5382
Residual Variance	1.4276	1.4286	1.4286	1.4266	1.4286	0.1785	0.1789	0.1779	1.4260

Table 4 Final Model Selection

Forecast Function

ARIMA(2,0,2): $\hat{Y}(t) = 0.6087 + 1.7778y(t\text{-}1) - 0.7970y(t\text{-}2) - 0.7352e(t\text{-}1) - 0.1495e(t\text{-}2)$

Summary

1. ARMA(2, 2) model fits the best in this dataset
2. The performances of base-model are similar
3. Too few independent variables can affect the analysis
4. Collinearity can significantly impact model performance

Thank You

DATS 6313 Brooklyn Chen