Appliance Energy Use

Alexis Kaldany

Time Series Term Project

Overseen by Professor R. Jafari

April 22,2022

Table of Contents

- Descriptive Statistics
- Differencing and Log-Diff
- Seasonality
- Base Models
- Linear Regression and Backwards Selection
- GPAC
- ARMA (3,0)
- ARMA(3,3)
- SARIMA (ARIMA(3,0,0)x(0,3,0,12))
- LSTM
- Overview of Model Results
- Conclusion
- Citations

- Dataset is composed on 26 variables.
- Dependent variable is “Appliances”, which measures energy use of appliances in the home.
- All variables have a 10 minute time step.
- 144 time step is equal to one day.
- Rv1 and rv2 are random variables and do not have predictive power.
- Temperature and humidity are observed in various places in the home.
- Additional variables pertaining to the weather are calculated outside.

Table of Variables

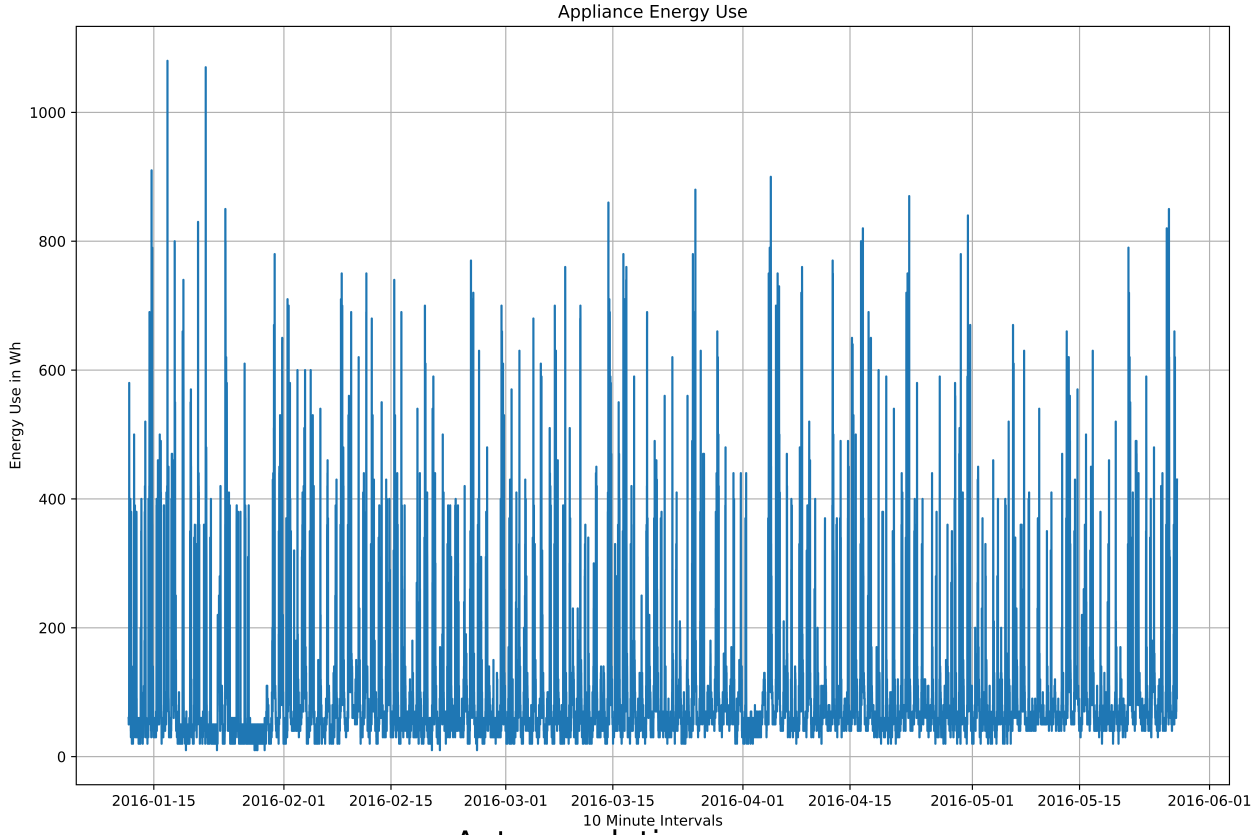
- All temperatures in Celsius
- All humidity is in % terms
- All energy use in watts per hour

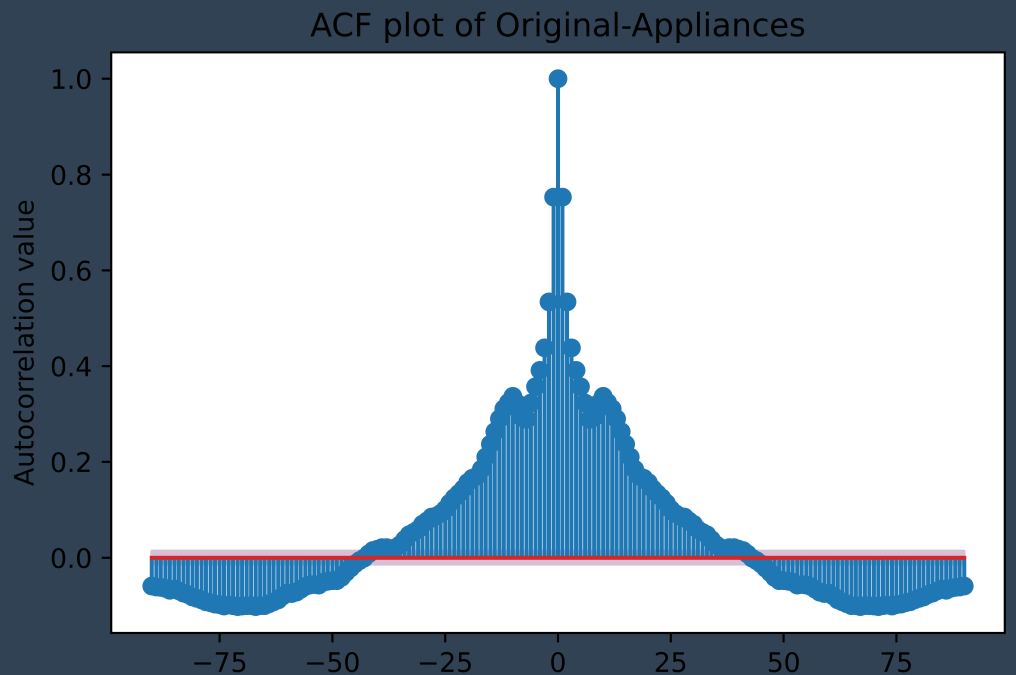
Variable	Description	Variable	Description
date	10 minute intervals	Appliances	energy use of appliances
lights	energy use of light fixtures	Press_mm_hg	
RH_1	Humidity in kitchen	T1	Temp in kitchen
RH_2	Humidity in living room	T2	Temp in living room
RH_3	Humidity in laundry area	T3	Temp in laundry area
RH_4	Humidity in office room	T4	Temp in office room
RH_5	Humidity in bathroom	T5	Temp in bathroom
RH_6	Humidity in north area	T6	Temp in north area
RH_7	Humidity in ironing room	T7	Temp in ironing room
RH_8	Humidity in teenager room	T8	Temp in teenager room
RH_9	Humidity in parents room	T9	Temp in parents room
RH_out	Humidity outside	T_out	Temp outside
Windspeed	in m/s	Visibility	kilometers
rv1	random variable 1	rv2	andom variable 2
Tdewpoint	A*C		

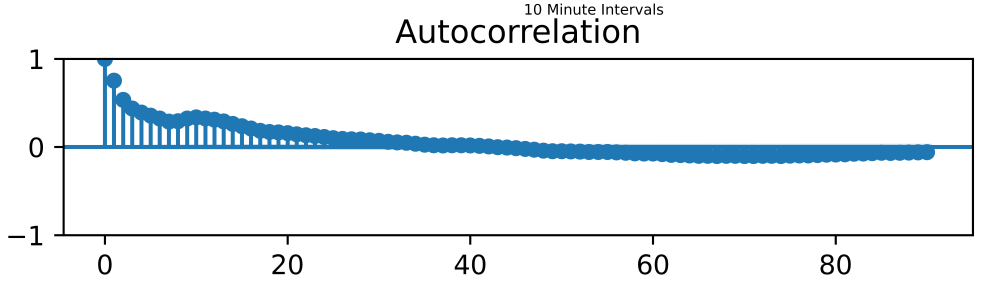
Location of variables within the home

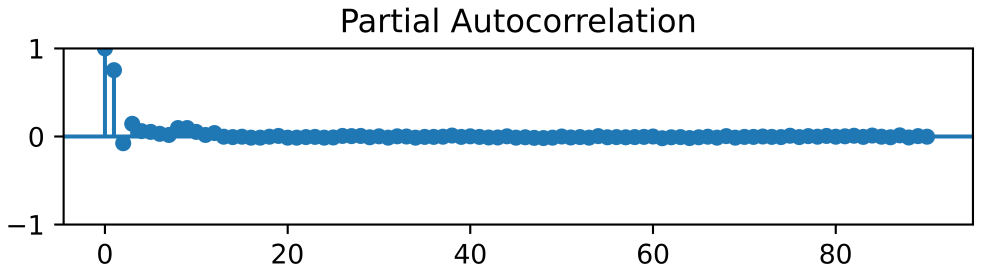
1. Original Undifferenced plot
2. ACF and PACF of undifferenced data

- Based on the plot there is clearly spikes in energy usages, and then a return to zero use.
- Energy use sits at ~0.0 then spikes all the way to 1000 at times.
- The ACF looks like there is some AR terms because of the relative smoothness over the 90 lags.
- The PACF shows a little bit of MA ordering.

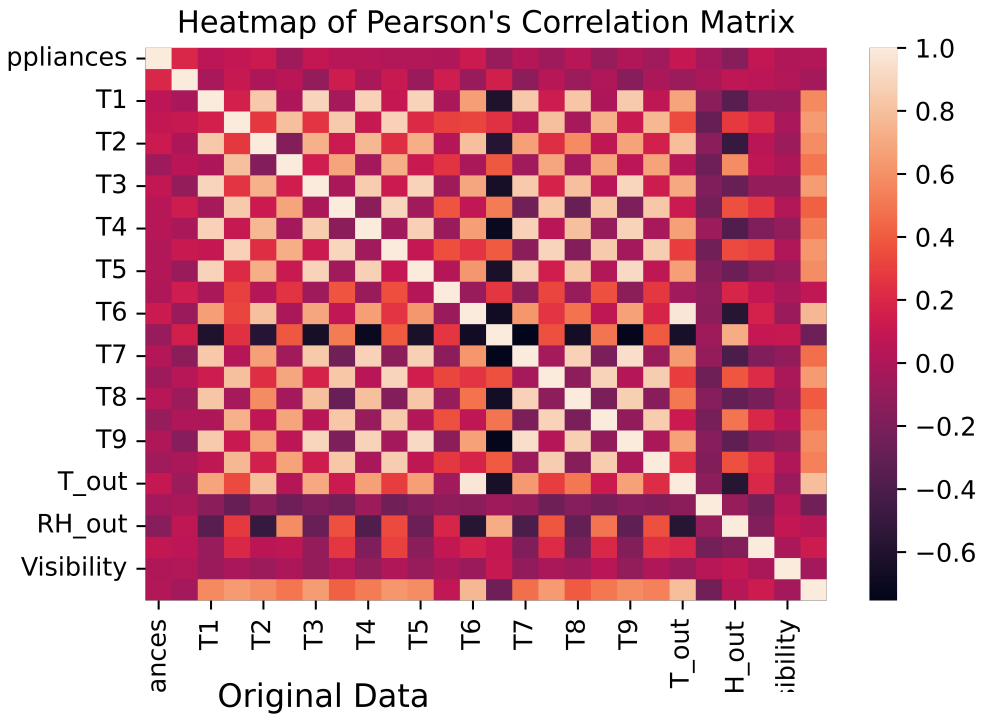


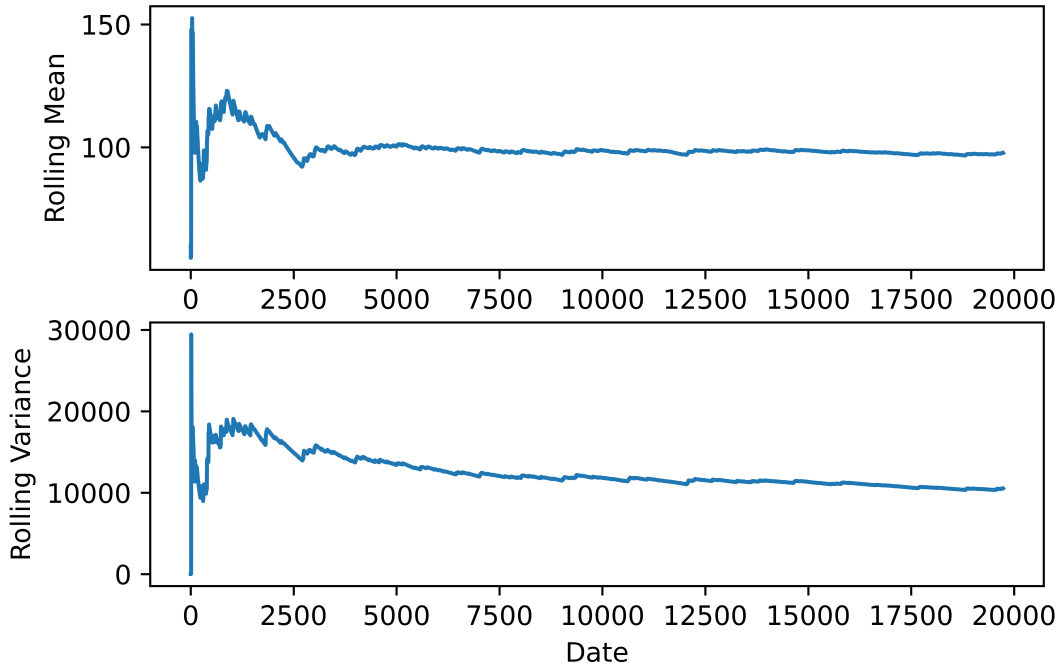






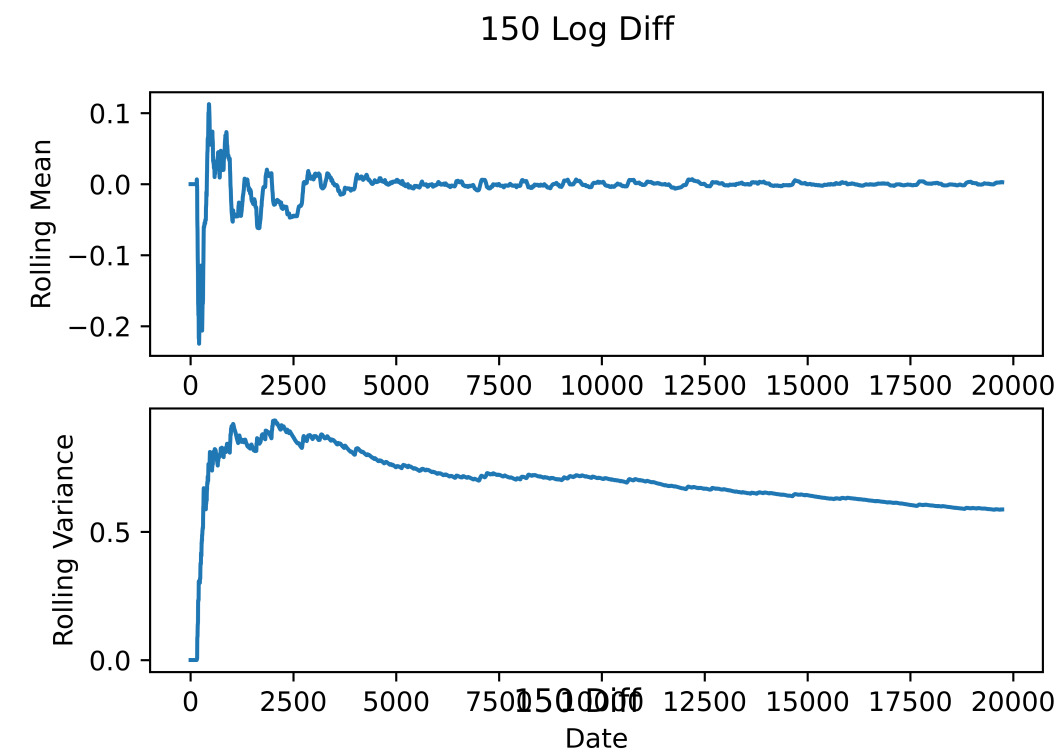
- Some of the variable names didn't make it into the heat plots labels because there are so many.
- Clearly we see the classic check marked scheme in the correlation matrix, indicating lower levels of collinearity.
- Rolling mean/variance of the original dependent variable shows stationarity in the values but not necessarily stationarity in the variance, but pretty close.

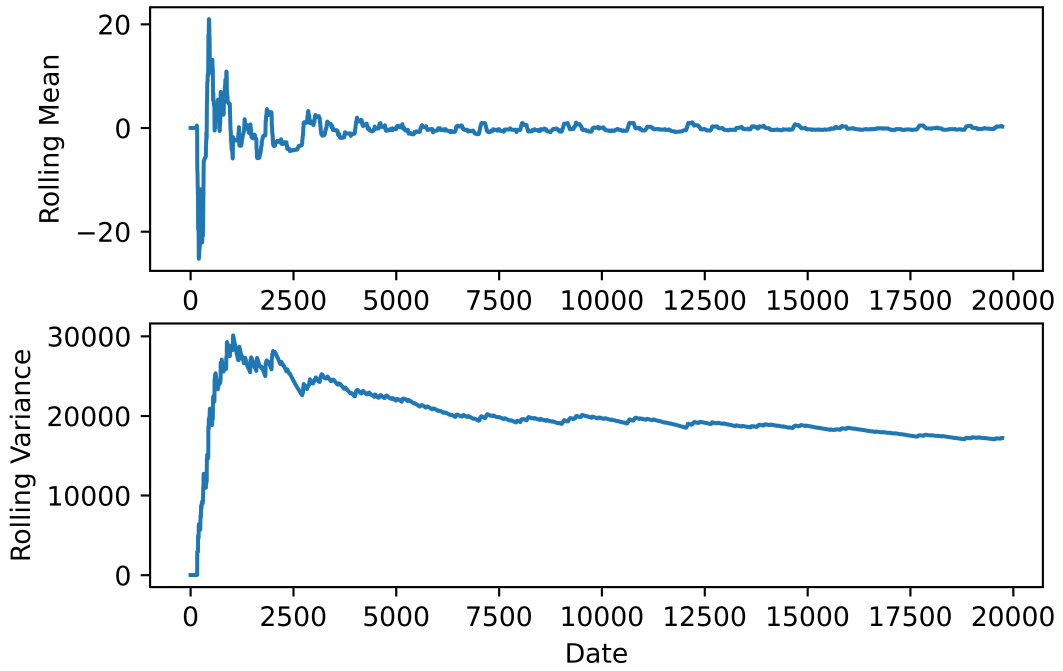




Differencing and Log-Diff

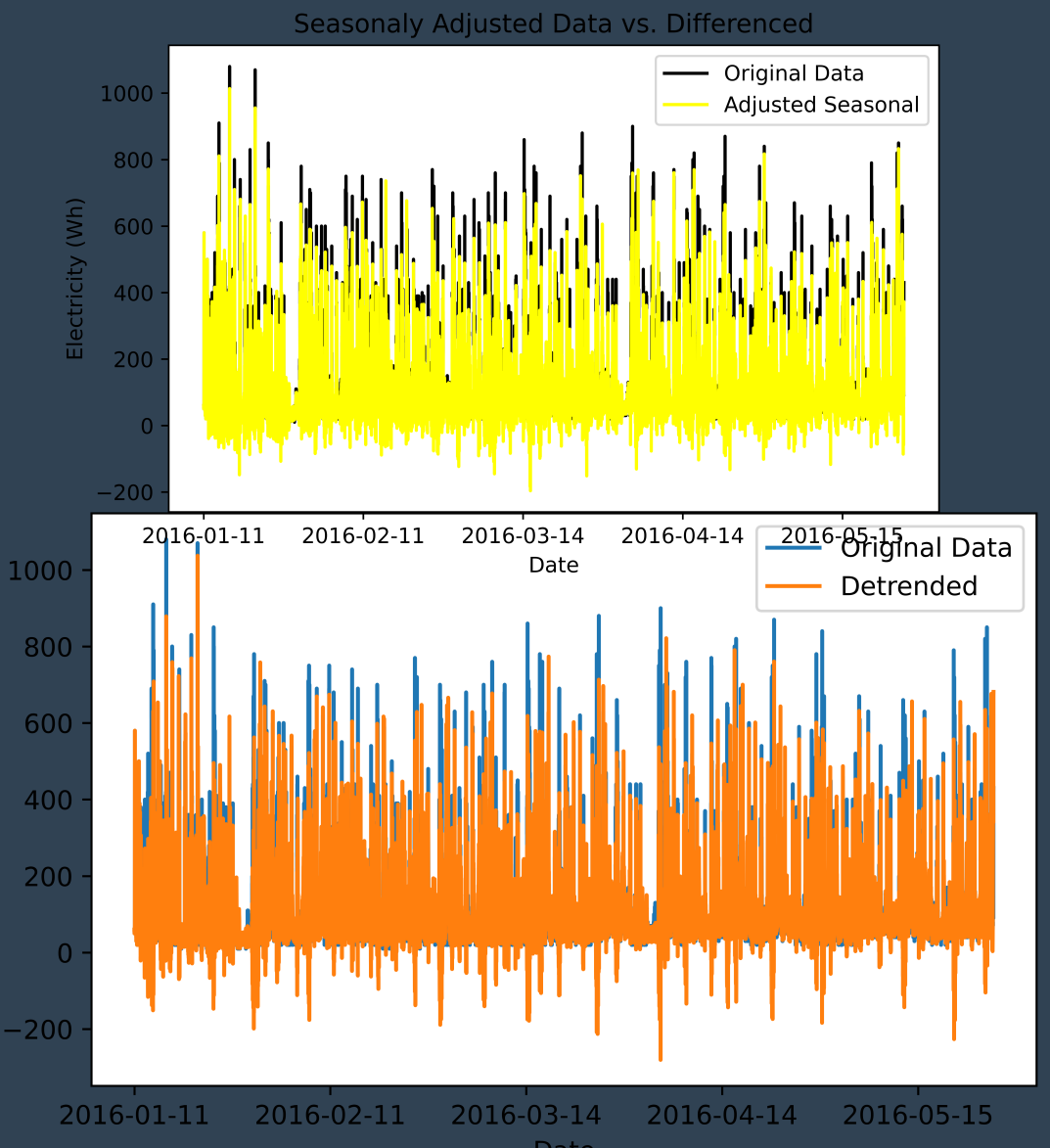
- I chose 150 as a differencing interval as it is near a full day (which is actually 144 steps as the data moves at 10 minute intervals)
- Differencing at that interval does make the rolling mean even more stationary but doesn't help with making the variance stationary
- Taking the log-diff at that interval makes the variance of the mean even worse
- I decided to stick with undifferenced data
- Both the ADF and KPSS tests indicated stationarity.
- ADF test statistic for original was -21, KPSS was 0.03, both indicate stationarity.

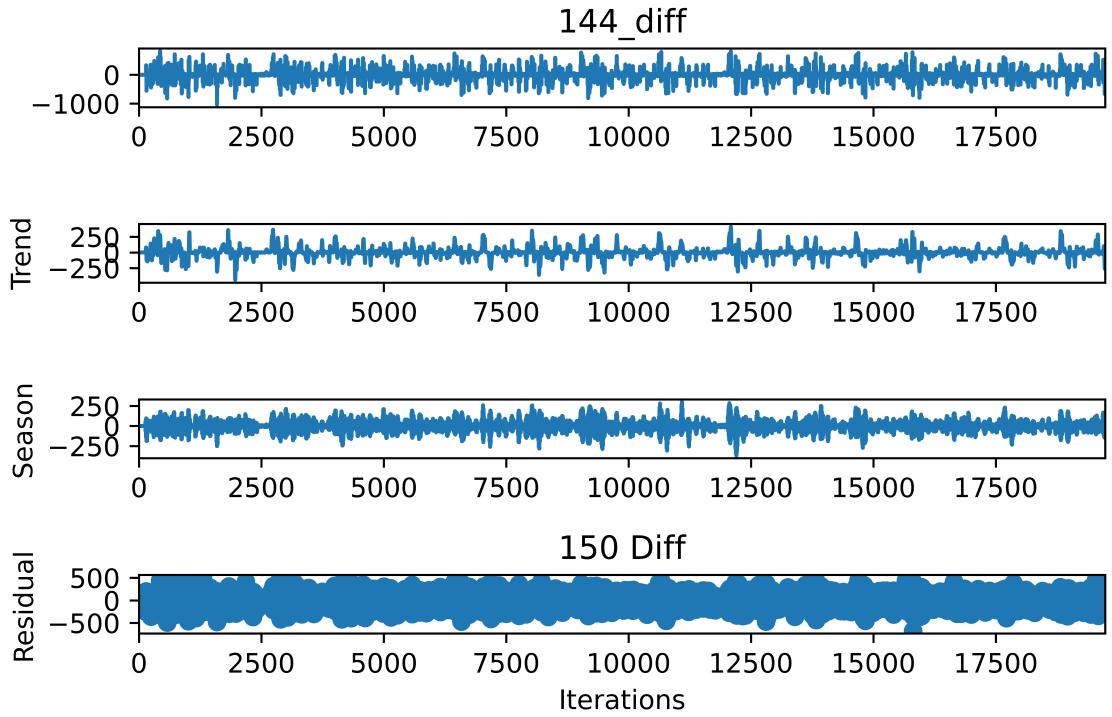


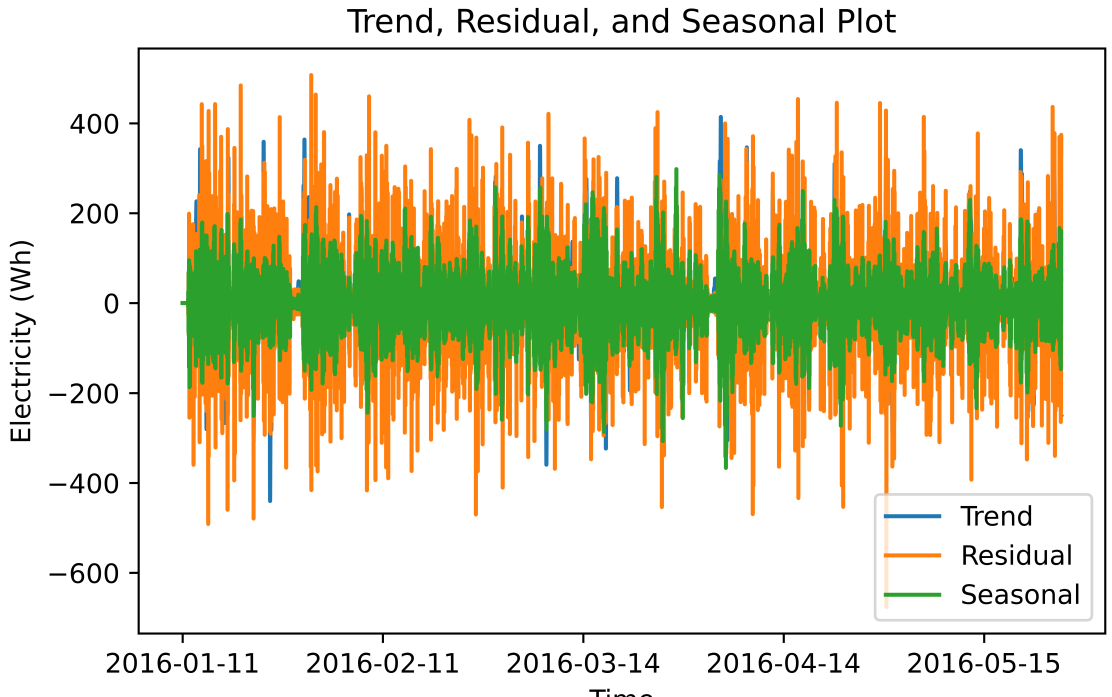


Decomposition Analysis

- Strength of Trend = 0.6
- Strength of Seasonality = 0.3

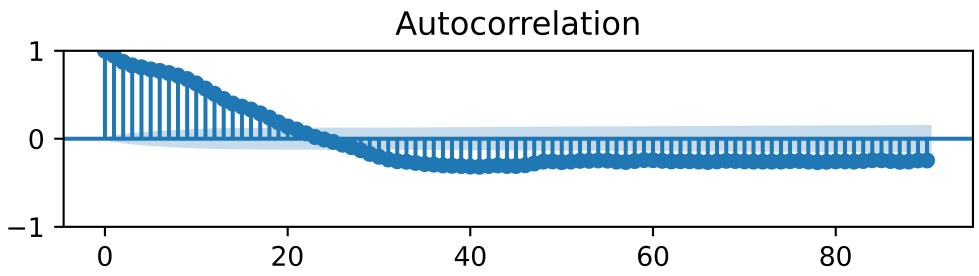


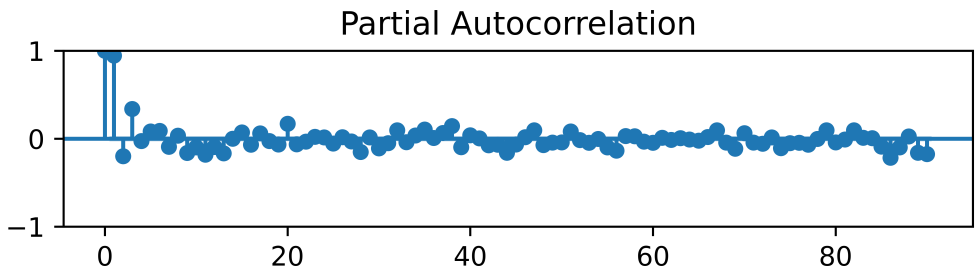


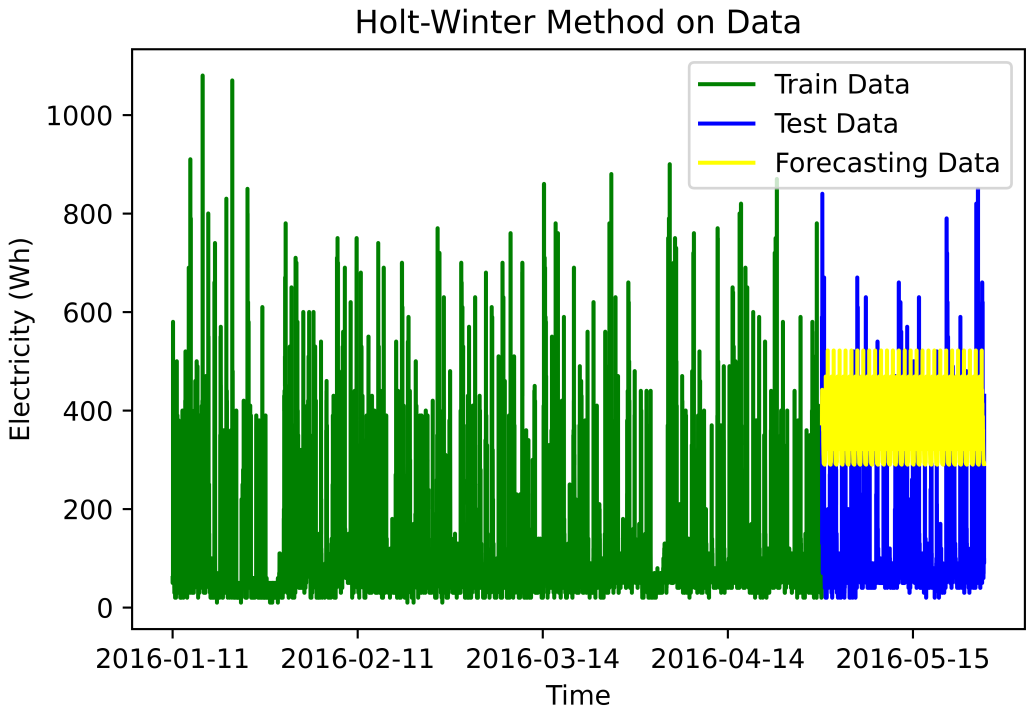


Holt-Winters Model

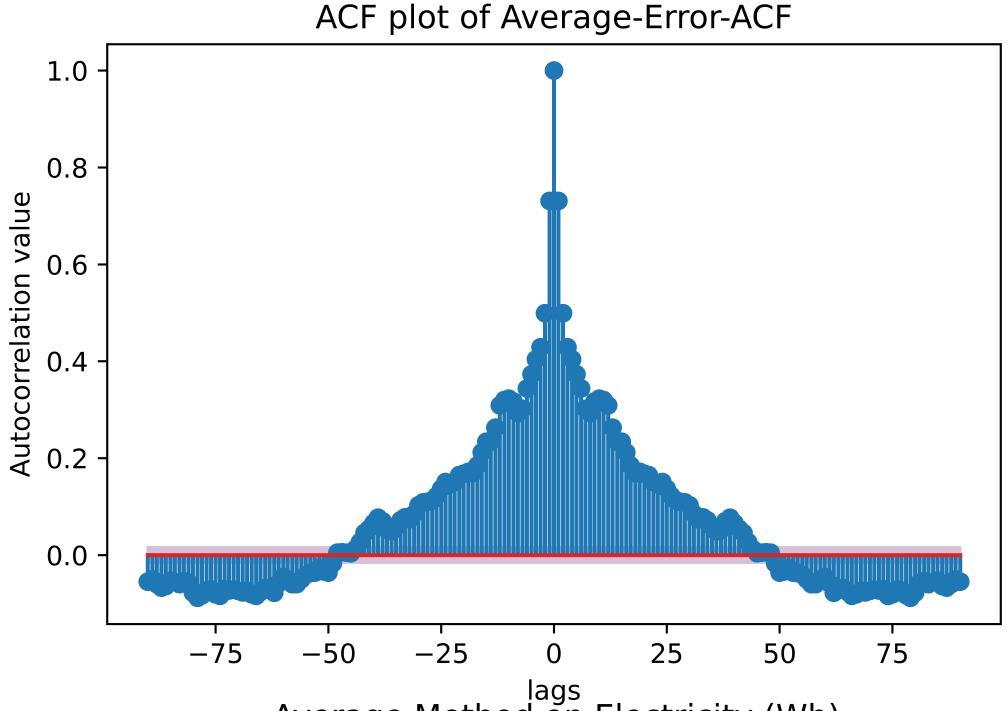
- MSE = 80324
- Q = 17055
- Error Var = 10564

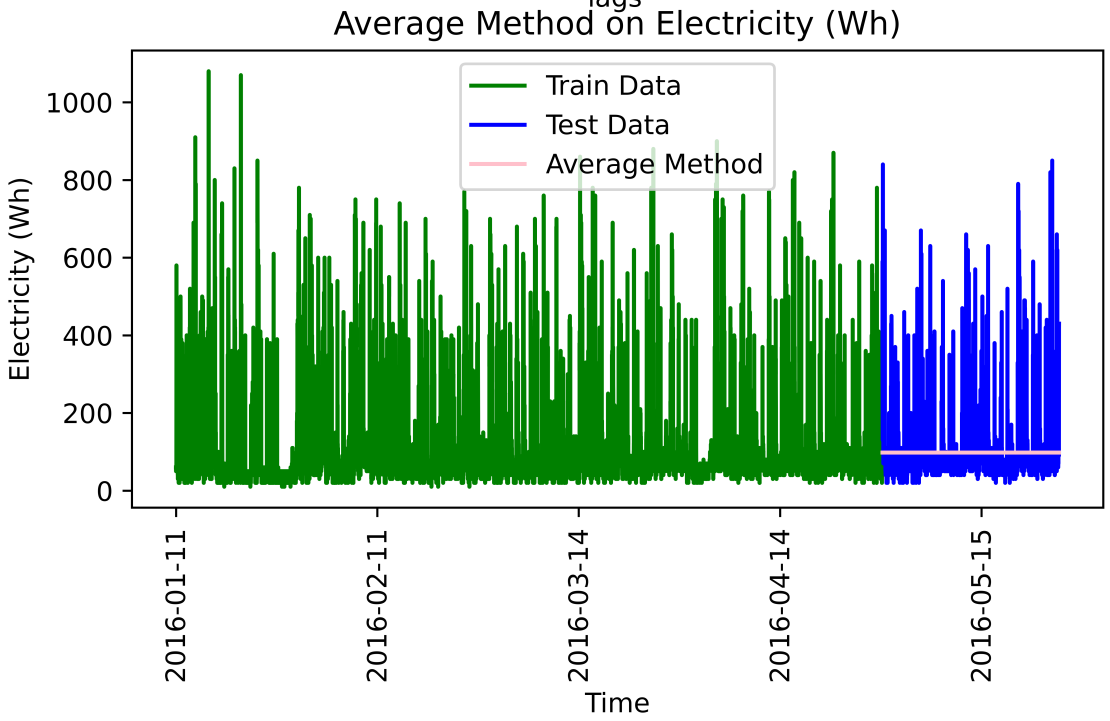




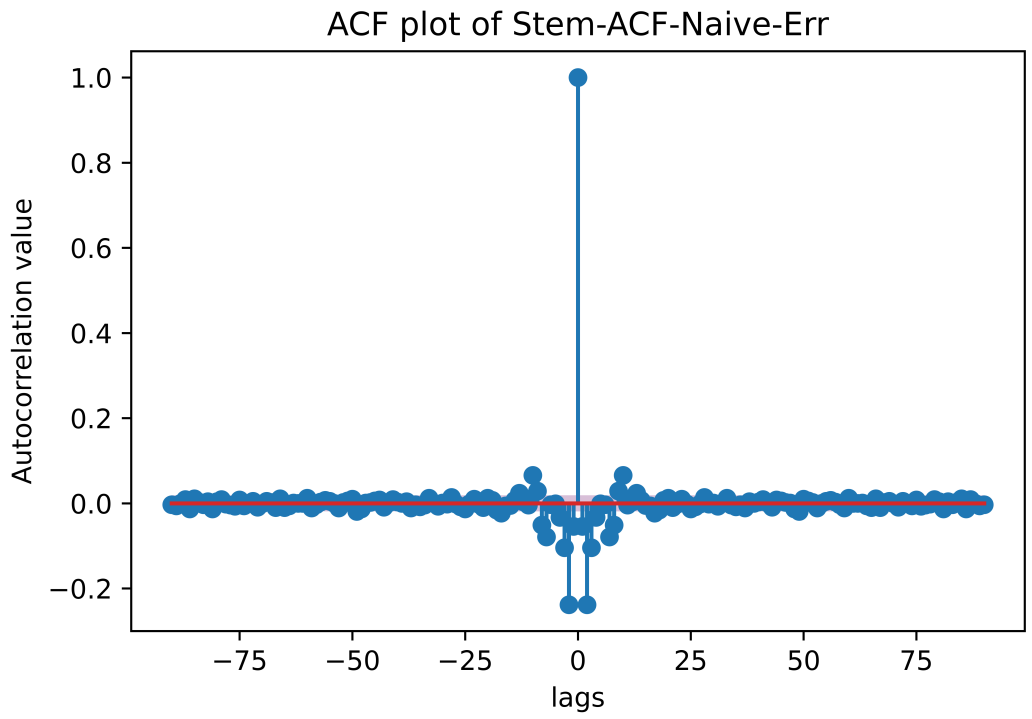


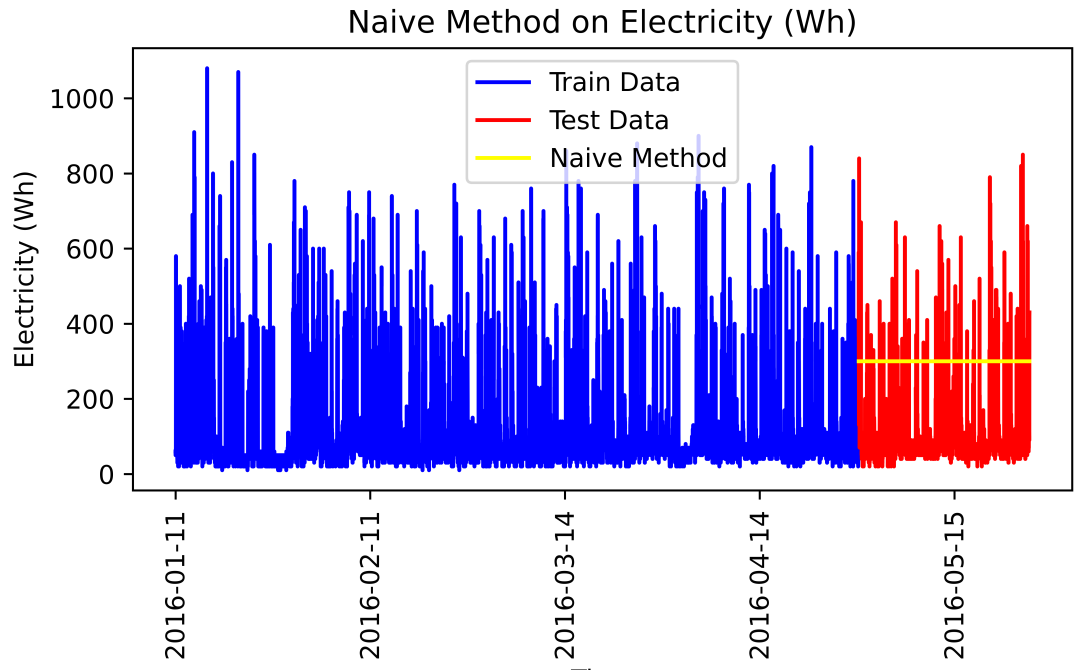
Average Method
MSE = 11074
Q value = 21037
Error Var = 5381





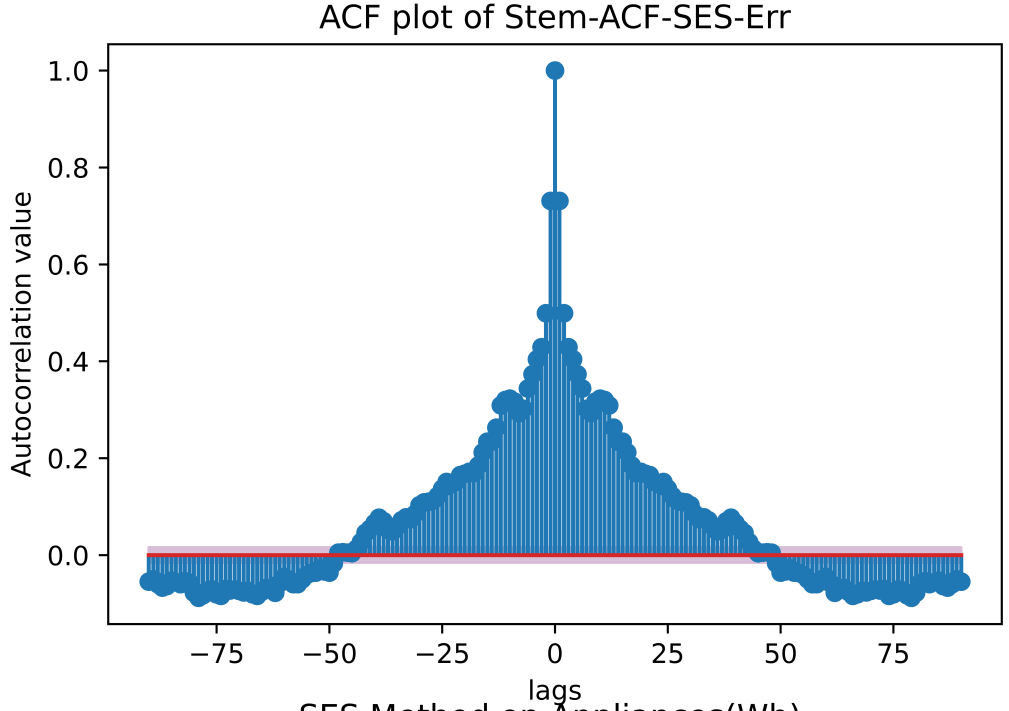
Naïve Method
MSE = 5381
Q – Value = 1129

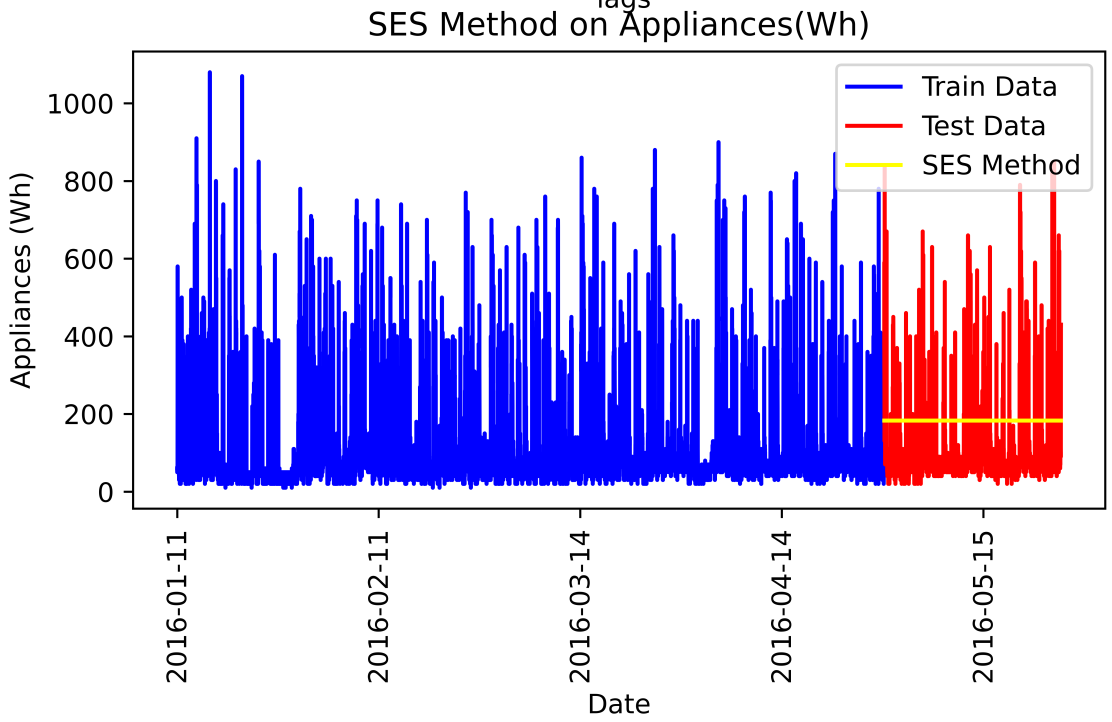




SES Method

MSE = 5717
Q Value = 2705





	coef	std err	t	P>|t|	[0.025	0.975]
lights	2.1405	0.105	20.446	0.000	1.935	2.346
T1	-6.3577	2.239	-2.840	0.005	-10.746	-1.969
RH_1	12.8278	0.757	16.946	0.000	11.344	14.312
T2	-13.2161	2.097	-6.304	0.000	-17.326	-9.107
RH_2	-12.4343	0.912	-13.636	0.000	-14.222	-10.647
T3	27.6756	1.242	22.278	0.000	25.241	30.111
RH_3	8.6016	0.839	10.250	0.000	6.957	10.246
T4	-3.7388	1.090	-3.429	0.001	-5.876	-1.601
RH_4	-2.4344	0.735	-3.313	0.001	-3.874	-0.994
T5	-1.6770	1.509	-1.111	0.266	-4.635	1.281
RH_5	0.1043	0.102	1.026	0.305	-0.095	0.303
T6	7.5826	0.829	9.143	0.000	5.957	9.208
RH_6	0.3347	0.080	4.178	0.000	0.178	0.492
T7	1.2152	1.527	0.796	0.426	-1.777	4.208
RH_7	-0.5249	0.493	-1.066	0.287	-1.490	0.441
T8	8.9150	1.086	8.208	0.000	6.786	11.044
RH_8	-5.9552	0.429	-13.875	0.000	-6.797	-5.114
T9	-14.6278	2.028	-7.212	0.000	-18.604	-10.652
RH_9	-0.9481	0.468	-2.024	0.043	-1.866	-0.030
T_out	-9.3318	2.563	-3.641	0.000	-14.356	-4.308
Press_mm_hg	0.1630	0.059	2.763	0.006	0.047	0.279
RH_out	-0.6707	0.460	-1.459	0.145	-1.572	0.230
Windspeed	1.1960	0.374	3.195	0.001	0.462	1.930
Visibility	0.1909	0.063	3.032	0.002	0.068	0.314
Tdewpoint	4.5302	2.399	1.889	0.059	-0.171	9.232

Omnibus:	11048.846	Durbin-Watson:	0.622
Prob(Omnibus):	0.000	Jarque-Bera (JB):	160765.586
Skew:	3.283	Prob(JB):	0.00

OLS Regression Results

Dep. Variable:	Appliances	R-squared (uncentered):	0.557
Model:	OLS	Adj. R-squared (uncentered):	0.557
Method:	Least Squares	F-statistic:	793.7
Date:	Fri, 22 Apr 2022	Prob (F-statistic):	0.00
Time:	14:45:06	Log-Likelihood:	-94410.
No. Observations:	15788	AIC:	1.889e+05
Df Residuals:	15763	BIC:	1.891e+05
Df Model:	25		
Covariance Type:	nonrobust		

Linear Regression with all variables

	coef	std err	t	P>|t|	[0.025	0.975]
lights	2.1491	0.104	20.614	0.000	1.945	2.353
T1	-6.7459	2.221	-3.038	0.002	-11.099	-2.393
RH_1	12.9590	0.751	17.263	0.000	11.488	14.430
T2	-13.3953	2.070	-6.470	0.000	-17.453	-9.337
RH_2	-12.5307	0.895	-14.001	0.000	-14.285	-10.776
T3	27.4552	1.215	22.602	0.000	25.074	29.836
RH_3	8.4737	0.827	10.249	0.000	6.853	10.094
T4	-3.8113	1.070	-3.563	0.000	-5.908	-1.714
RH_4	-2.6735	0.702	-3.810	0.000	-4.049	-1.298
T6	7.3525	0.801	9.177	0.000	5.782	8.923
RH_6	0.3185	0.078	4.084	0.000	0.166	0.471
T8	9.4176	0.980	9.608	0.000	7.496	11.339
RH_8	-6.3048	0.362	-17.402	0.000	-7.015	-5.595
T9	-15.1285	1.697	-8.916	0.000	-18.454	-11.803
RH_9	-0.8216	0.443	-1.855	0.064	-1.690	0.047
T_out	-5.7160	0.927	-6.165	0.000	-7.533	-3.899
Press_mm_hg	0.1025	0.031	3.260	0.001	0.041	0.164
Windspeed	1.2841	0.372	3.456	0.001	0.556	2.012
Visibility	0.1816	0.063	2.904	0.004	0.059	0.304
Tdewpoint	0.9710	0.680	1.428	0.153	-0.362	2.304

Omnibus:	11050.350	Durbin-Watson:	0.622
Prob(Omnibus):	0.000	Jarque-Bera (JB):	160868.059
Skew:	3.283	Prob(JB):	0.00
Kurtosis:	17.192	Cond. No.	3.06e+03

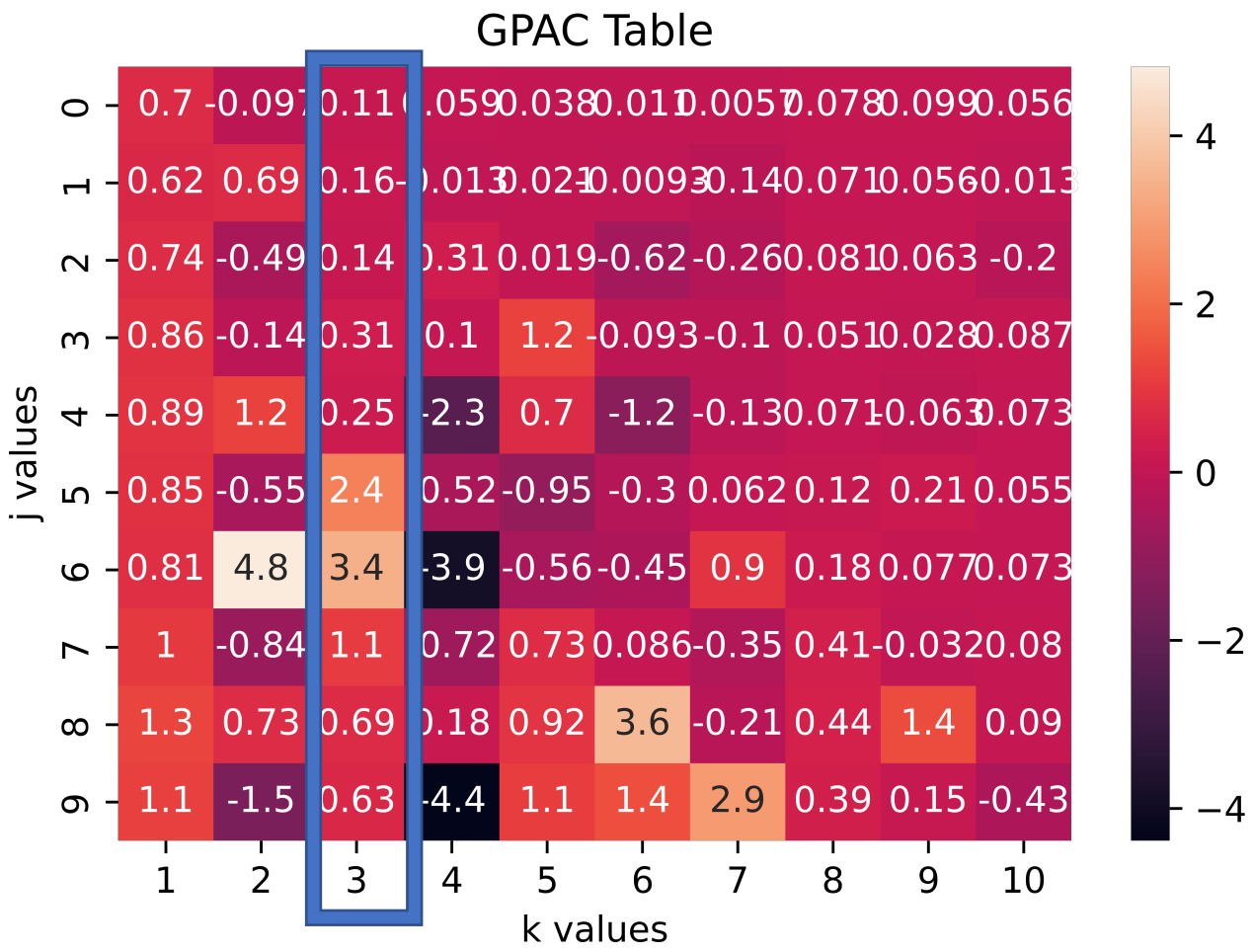
OLS Regression Results

Dep. Variable:	Appliances	R-squared (uncentered):	0.557
Model:	OLS	Adj. R-squared (uncentered):	0.557
Method:	Least Squares	F-statistic:	991.8
Date:	Fri, 22 Apr 2022	Prob (F-statistic):	0.00
Time:	14:55:19	Log-Likelihood:	-94413.
No. Observations:	15788	AIC:	1.889e+05
Df Residuals:	15768	BIC:	1.890e+05
Df Model:	20		
Covariance Type:	nonrobust		

After ['T5','RH_5','T7','RH_7','RH_out’] are removed from the model

Very little change using backwards selection. F-statistic increased somewhat.

The GPAC is a bit difficult to interpret, but a third degree AR model is my guess for the order of the ARMA model



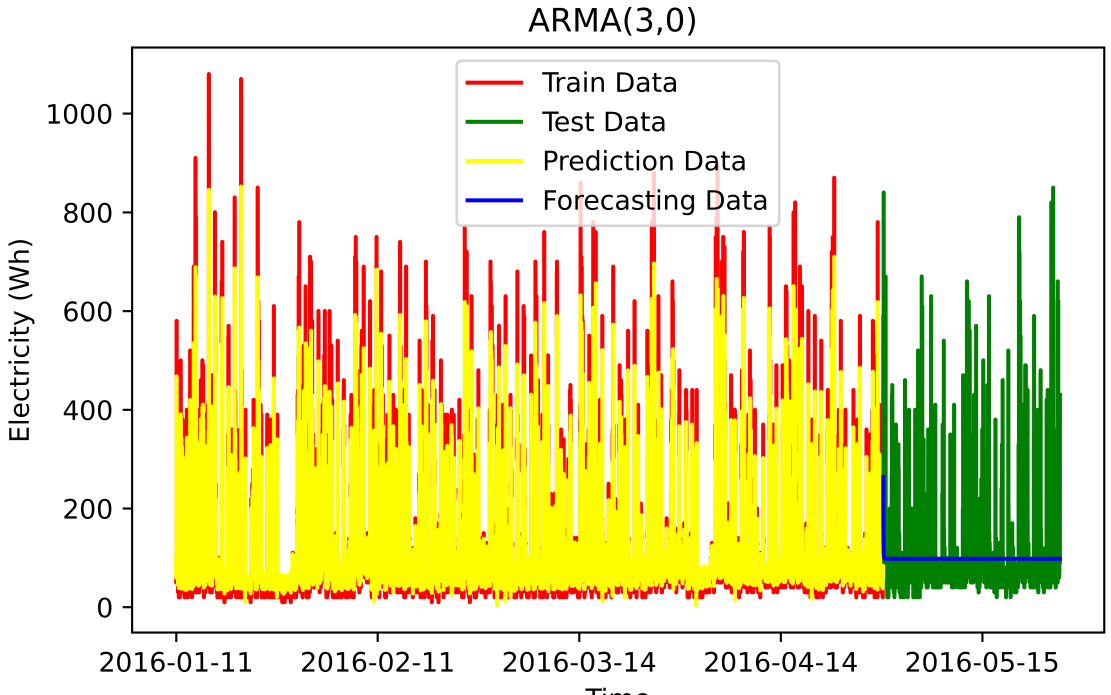
ARMA (3,0) had a MSE of 8206

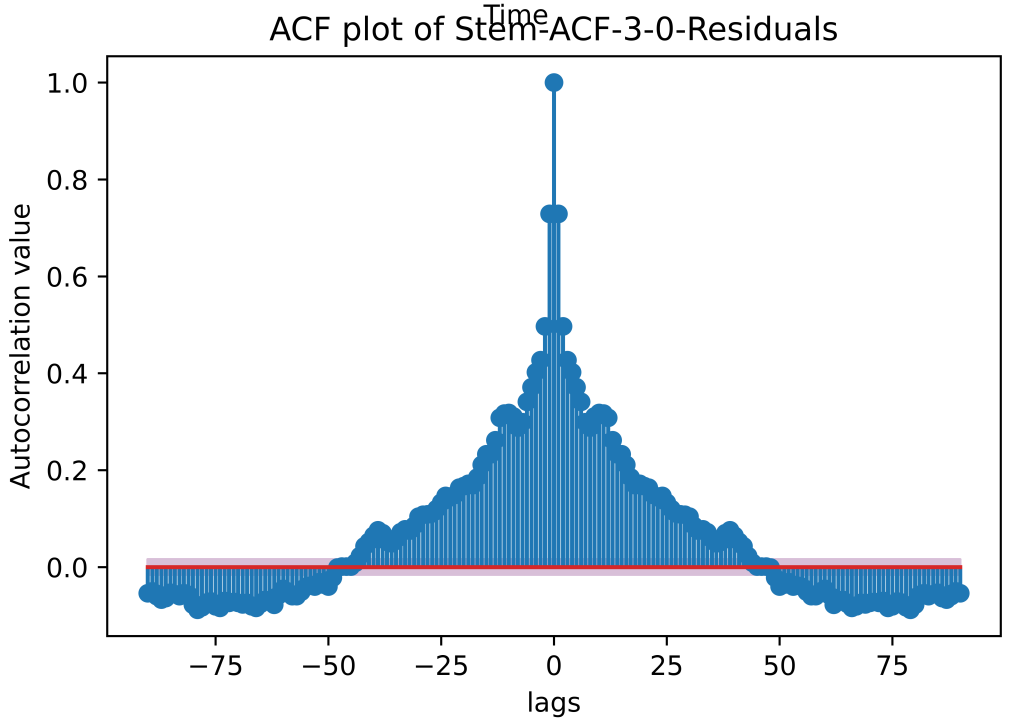
SARIMAX Results

Dep. Variable:	Appliances	No. Observations:	15788
Model:	ARIMA(3, 0, 0)	Log Likelihood	-89003.387
Date:	Fri, 22 Apr 2022	AIC	178016.775
Time:	15:16:41	BIC	178055.110
Sample:	01-11-2016	HQIC	178029.462
	- 04-30-2016		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
const	98.0244	4.526	21.660	0.000	89.154	106.894
ar.L1	0.8256	0.004	194.618	0.000	0.817	0.834
ar.L2	-0.1867	0.005	-36.561	0.000	-0.197	-0.177
ar.L3	0.1345	0.005	27.895	0.000	0.125	0.144
sigma2	4614.9424	21.953	210.222	0.000	4571.916	4657.969

Ljung-Box (L1) (Q):	1.02	Jarque-Bera (JB):	375727.84
Prob(Q):	0.31	Prob(JB):	0.00
Heteroskedasticity (H):	0.82	Skew:	3.21
Prob(H) (two-sided):	0.00	Kurtosis:	26.02





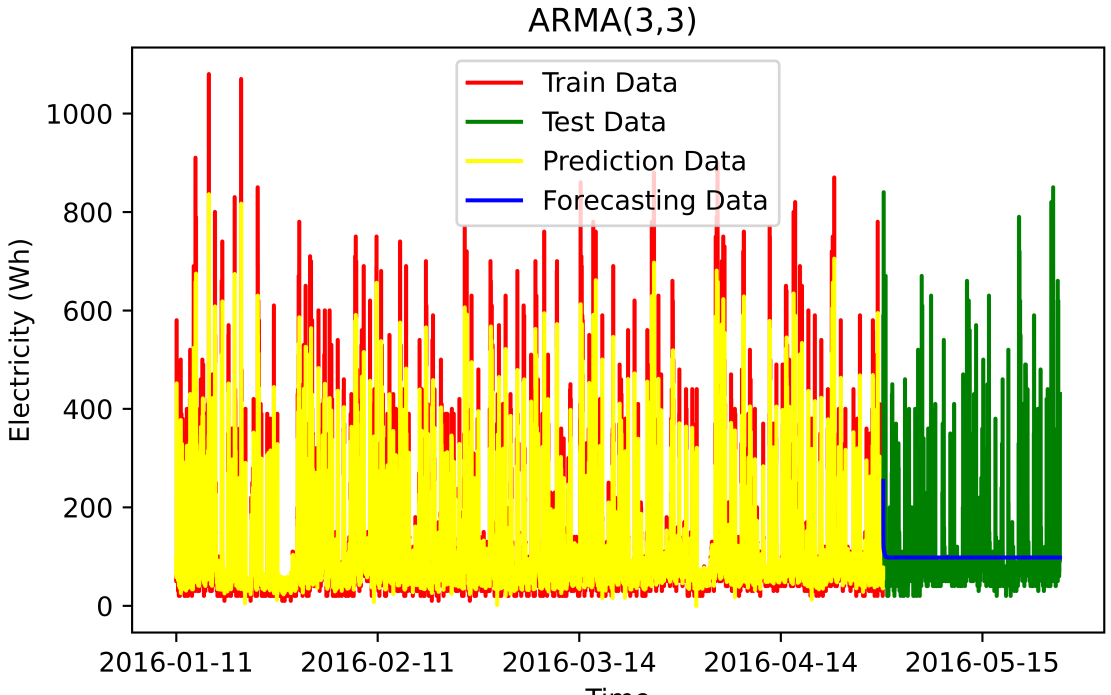
ARMA (3,3)

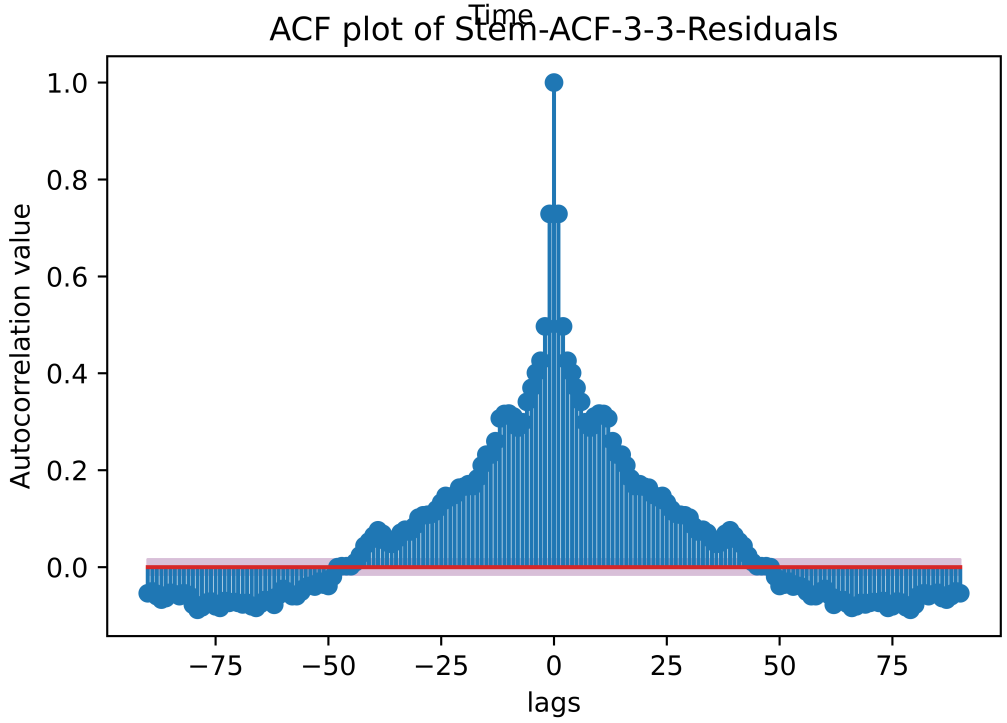
SARIMAX Results

Dep. Variable:	Appliances	No. Observations:	15788
Model:	ARIMA(3, 0, 3)	Log Likelihood	-88854.056
Date:	Fri, 22 Apr 2022	AIC	177724.112
Time:	15:17:19	BIC	177785.448
Sample:	01-11-2016	HQIC	177744.411
	- 04-30-2016		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
const	98.0277	7.277	13.471	0.000	83.766	112.290
ar.L1	0.3942	0.031	12.567	0.000	0.333	0.456
ar.L2	0.9155	0.042	21.609	0.000	0.832	0.999
ar.L3	-0.3635	0.019	-18.666	0.000	-0.402	-0.325
ma.L1	0.4060	0.031	13.031	0.000	0.345	0.467
ma.L2	-0.7763	0.023	-33.048	0.000	-0.822	-0.730
ma.L3	-0.2469	0.008	-31.047	0.000	-0.262	-0.231
sigma2	4527.9520	21.696	208.697	0.000	4485.428	4570.476

Ljung-Box (L1) (Q):	0.02	Jarque-Bera (JB):	362133.17
Prob(Q):	0.88	Prob(JB):	0.00
Heteroskedasticity (H):	0.81	Skew:	3.21
Prob(H) (two-sided):	0.00	Kurtosis:	25.57





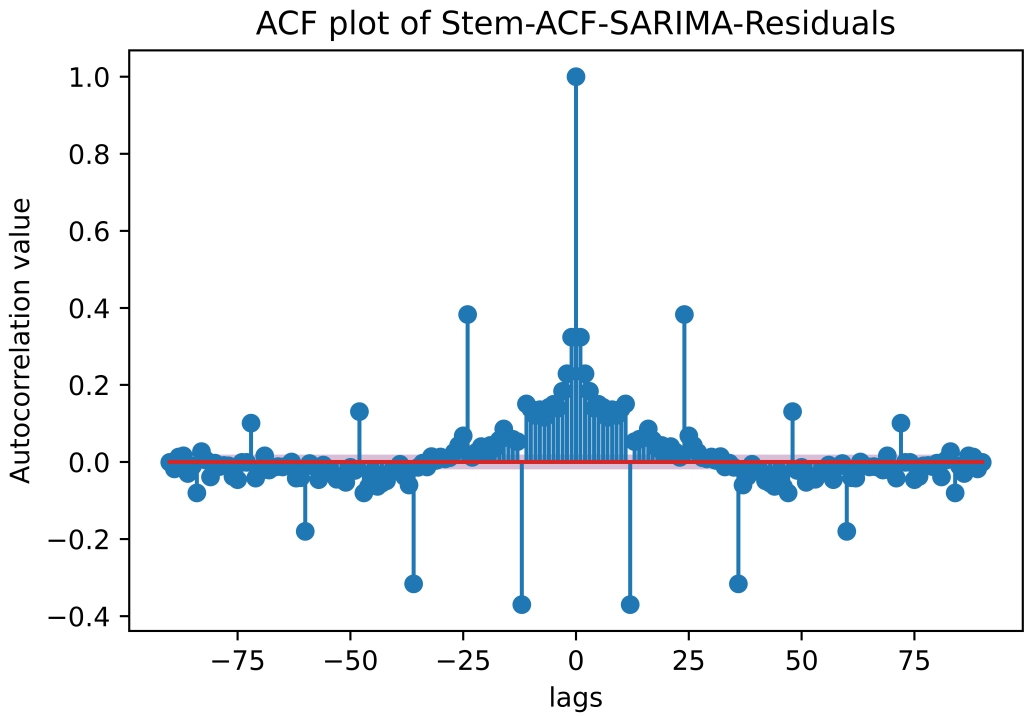
SARIMA: ARIMA(3,0,0)x(0,0,0,12)

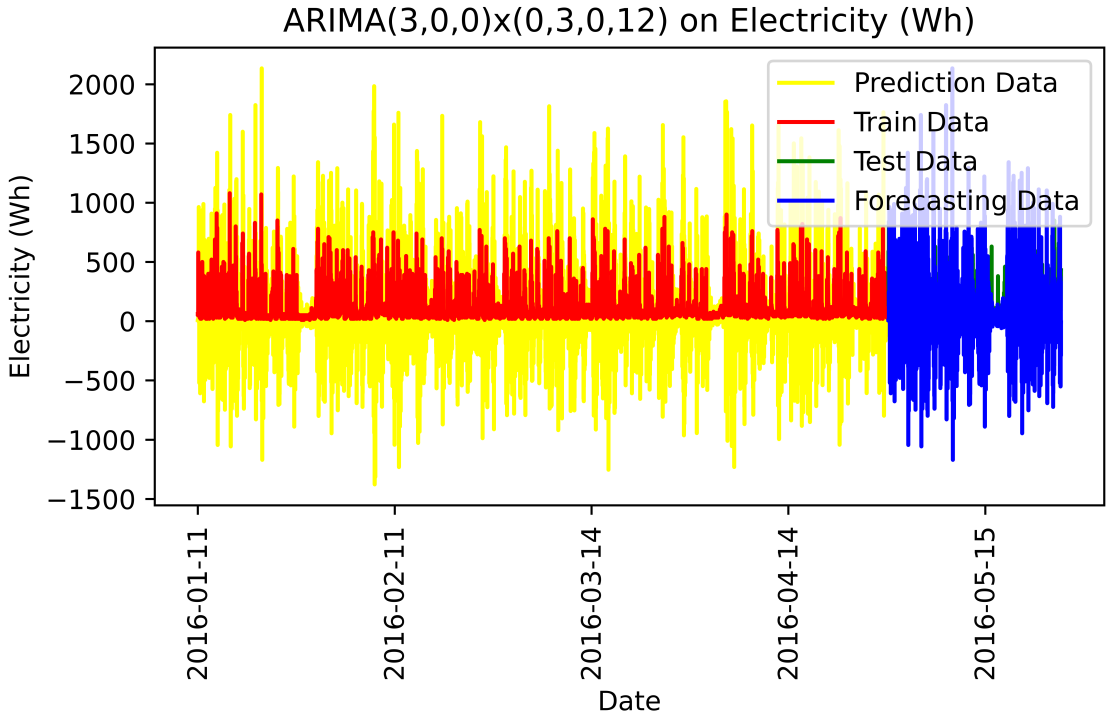
SARIMAX Results

Dep. Variable:	Appliances	No. Observations:	15788
Model:	SARIMAX(3, 0, 0)x(0, 3, 0, 12)	Log Likelihood	-108653.972
Date:	Fri, 22 Apr 2022	AIC	217315.943
Time:	16:46:08	BIC	217346.601
Sample:	01-11-2016	HQIC	217326.091
	- 04-30-2016		
Covariance Type:	opg		

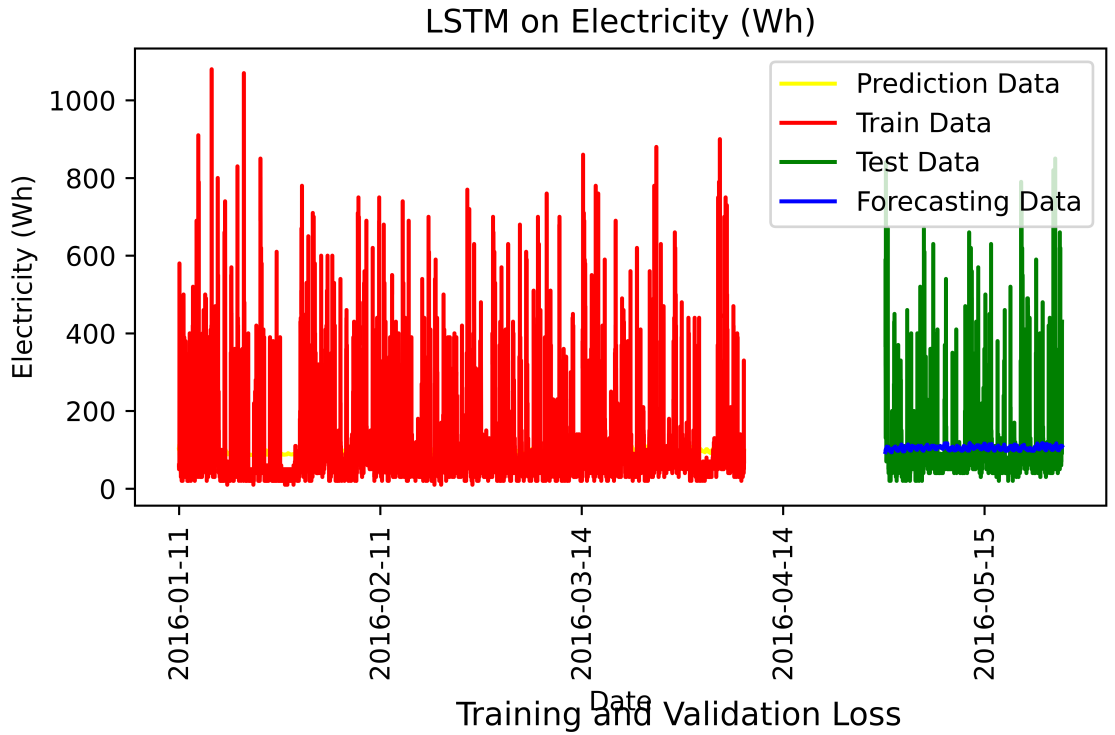
	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.7017	0.005	135.342	0.000	0.692	0.712
ar.L2	-0.2475	0.006	-40.828	0.000	-0.259	-0.236
ar.L3	0.1099	0.006	19.758	0.000	0.099	0.121
sigma2	3.276e+04	178.254	183.800	0.000	3.24e+04	3.31e+04

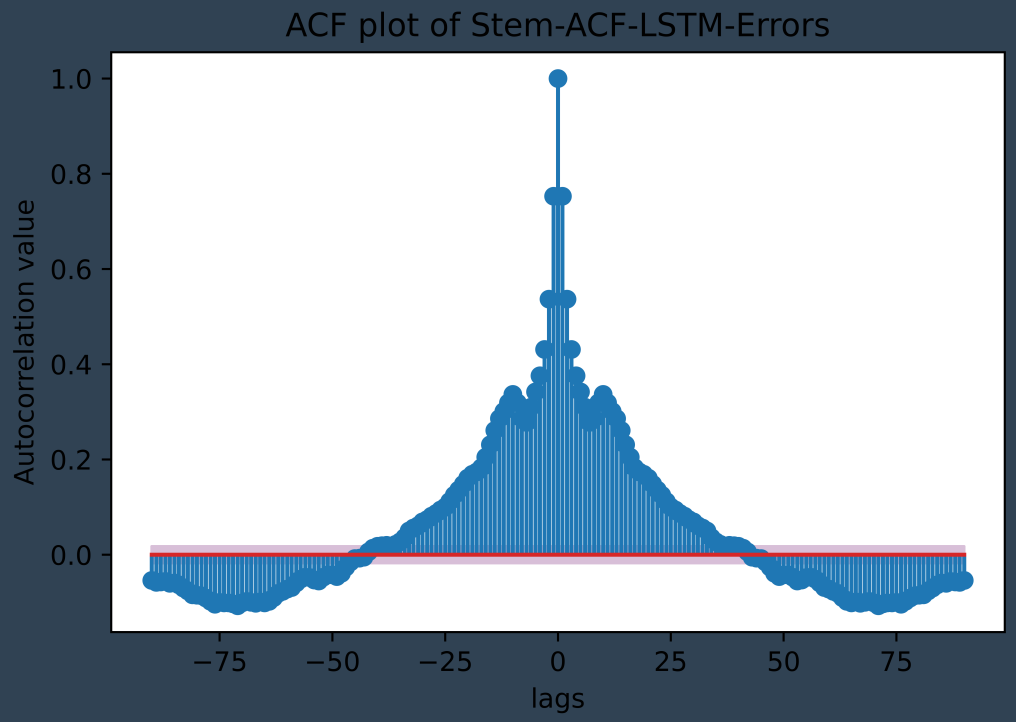
Ljung-Box (L1) (Q):	0.65	Jarque-Bera (JB):	31118.20
Prob(Q):	0.42	Prob(JB):	0.00
Heteroskedasticity (H):	0.79	Skew:	-0.49
Prob(H) (two-sided):	0.00	Kurtosis:	9.81





LSTM


LSTM on Electricity (Wh)
Prediction Data
Train Data
Test Data
Forecasting Data
Electricity (Wh)
1000
800
600
400
200
0
2016-01-11
2016-02-11
2016-03-14
2016-04-14
2016-05-15
Date


ACF plot of Stem-ACF-LSTM-Errors
Autocorrelation value
1.0
0.8
0.6
0.4
0.2
0.0
−75
−50
−25
0
25
50
75
lags

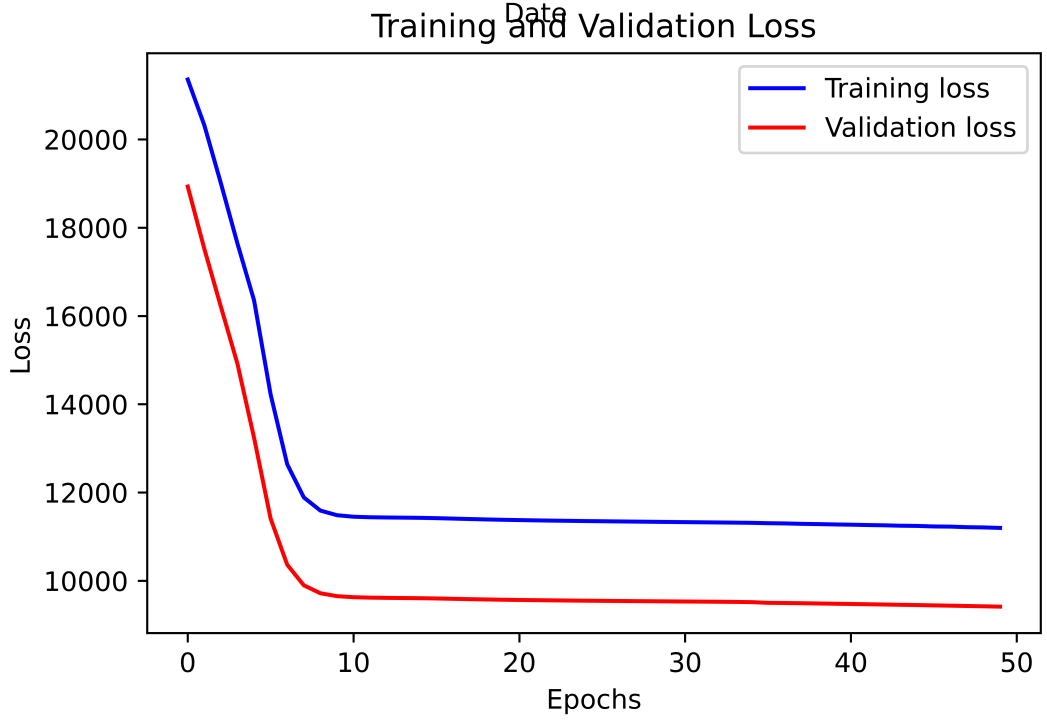

Training and Validation Loss
Training loss
Validation loss
Loss
20000
18000
16000
14000
12000
10000
0
10
20
30
40
50
Epochs

Table of Model Metrics

Winner: ARMA 3-3

models	mse	ljb
Holt-Winters	80324.1355	17055.560919
Average-Method	11074.3459	21037.197590
Naive-Method	5381.1617	1129.402962
Drift-Method	5385.1842	1128.966050
SES-Method	5717.6955	2705.998085
Linear-Regression	9156.2115	12769.092638
ARMA-3-0	4614.1882	40.116973
ARMA-3-3	4527.7083	1.145882
ARIMA(3,0,0)x(0,3,0,12)	57564.6157	NaN
LSTM	11369.3577	16878.570533

Original