



AIML CA 2

Predicting Energy Consumptions

Time Series

Goh Kun Ming

iL

DAAA/FT/1B/02



Data Cleaning and Analysis

Checking For Missing Data

Missing Values	
DATE	0
Gas Consumption (tons)	0
Electricity Consumption (MWh)	0
Water Consumption (tons)	0

Checking For Duplicated Data

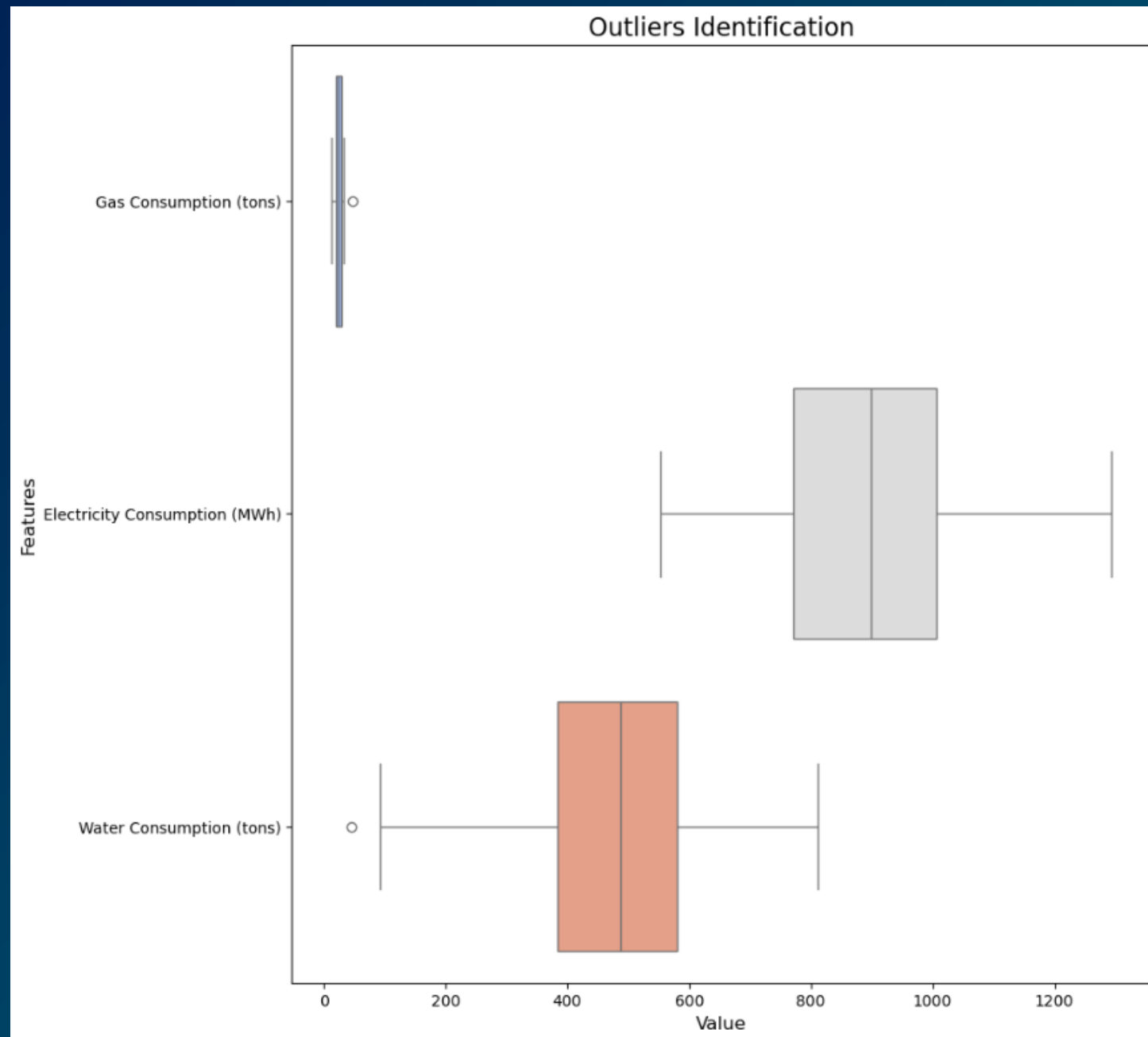
Count	
Duplicated Rows	0

Correcting Data Types and Setting Index

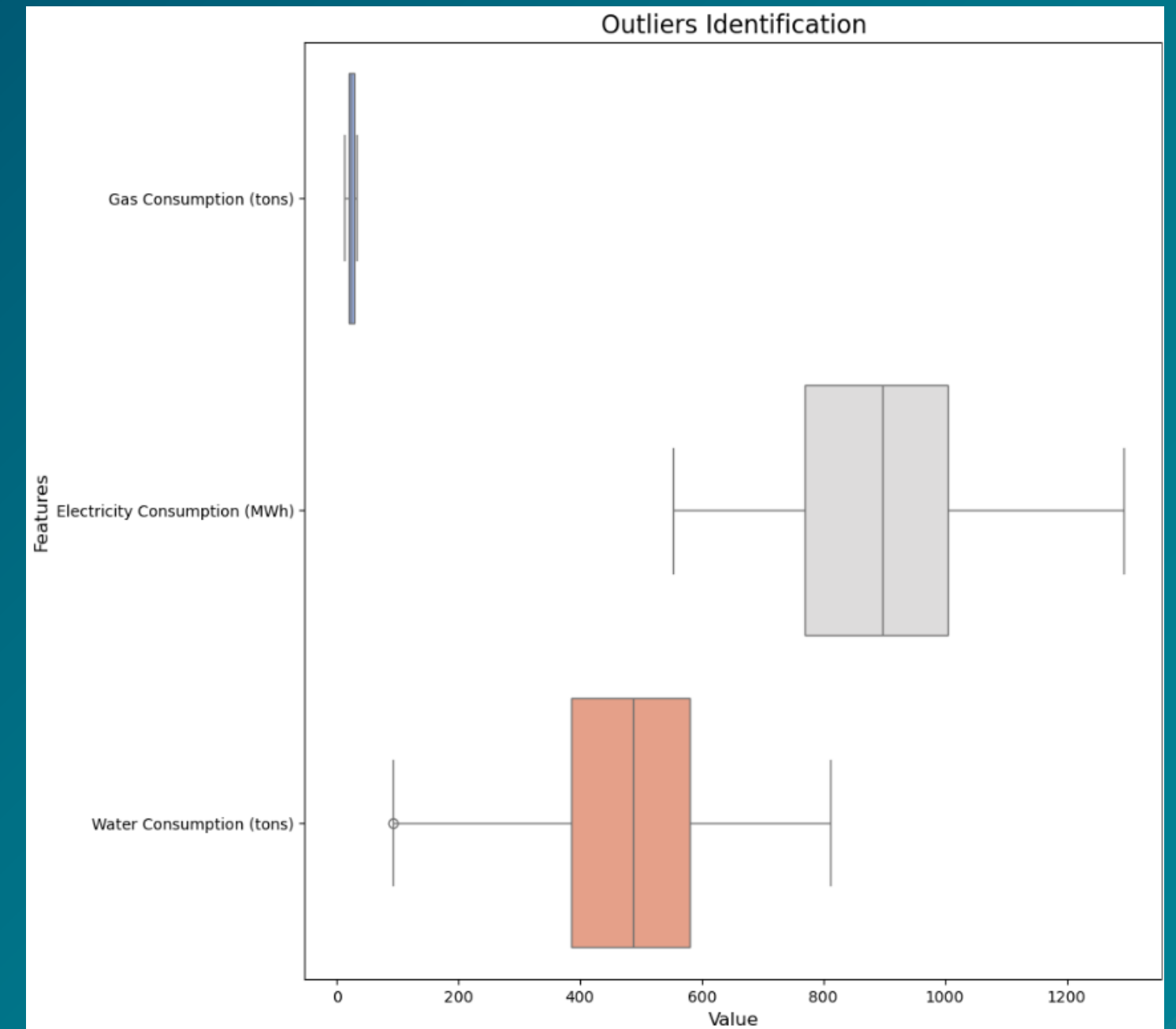
DATE	Gas Consumption (tons)	Electricity Consumption (MWh)	Water Consumption (tons)
1990-01-01	18.0	725.1	548.8
1990-02-01	15.8	706.7	640.7
1990-03-01	17.3	624.5	511.1
1990-04-01	18.9	574.7	515.3
1990-05-01	22.0	553.2	488.4

Data Cleaning and Analysis

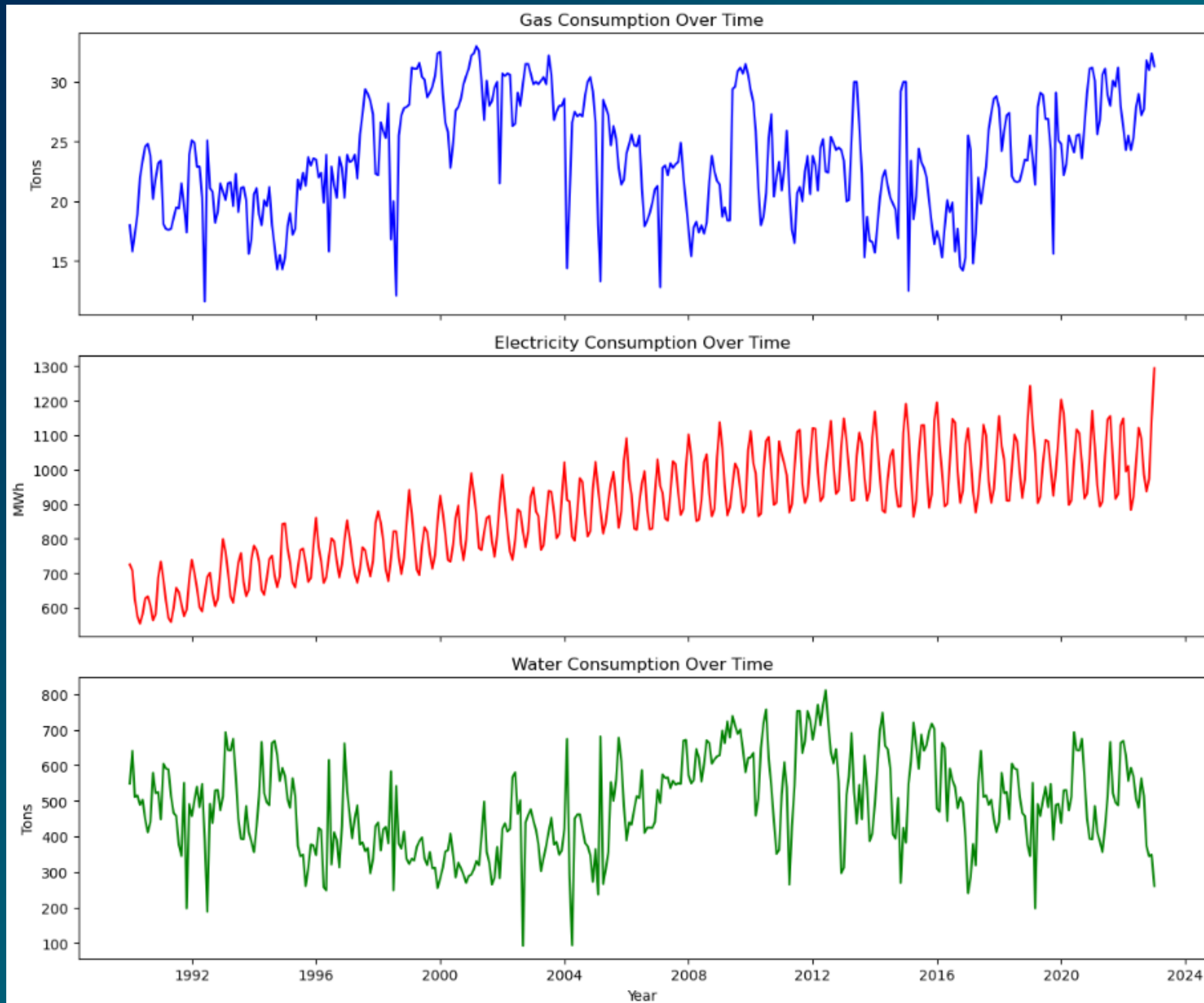
Checking For Outliers



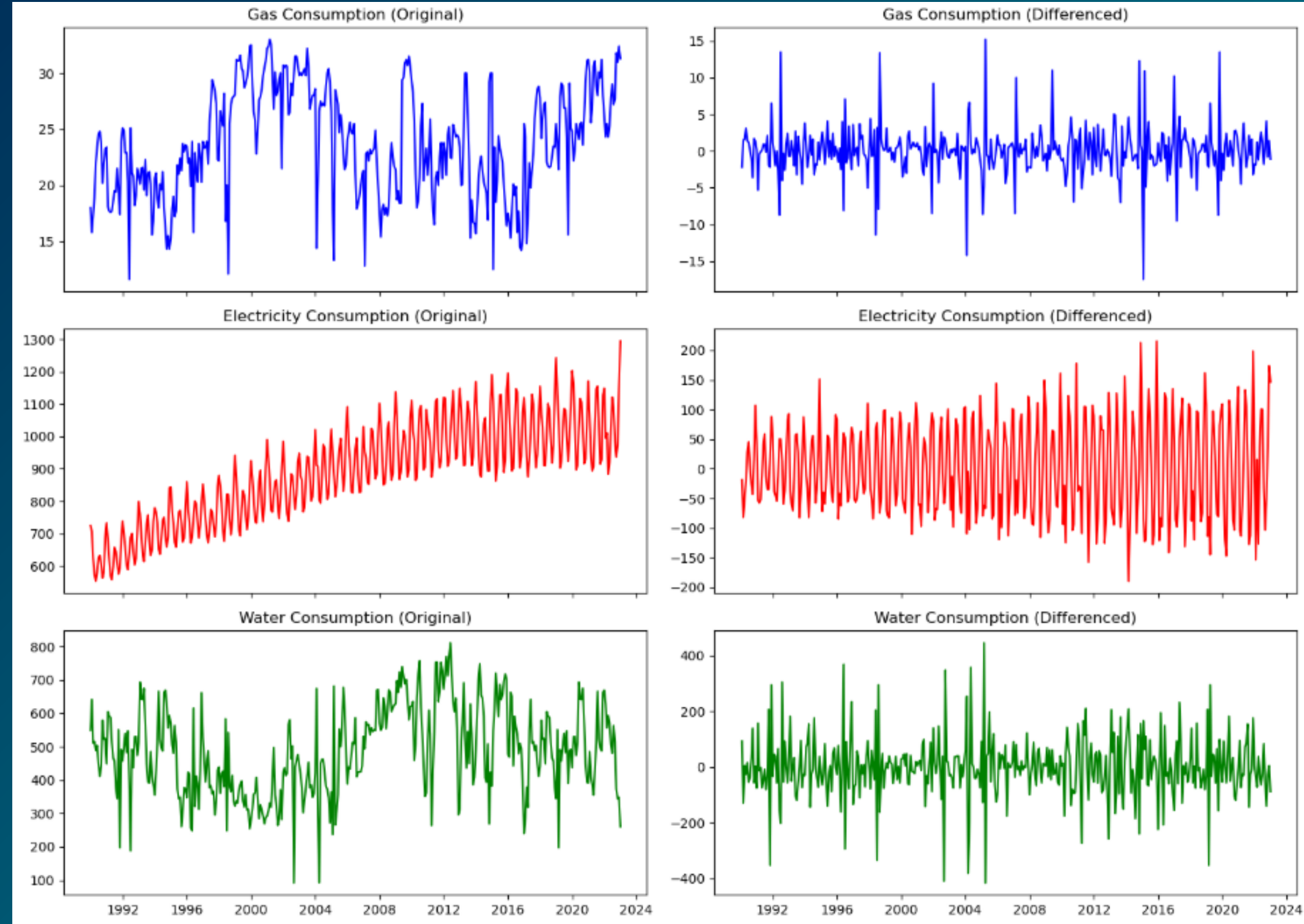
Addressing Outliers



Univariate Analysis

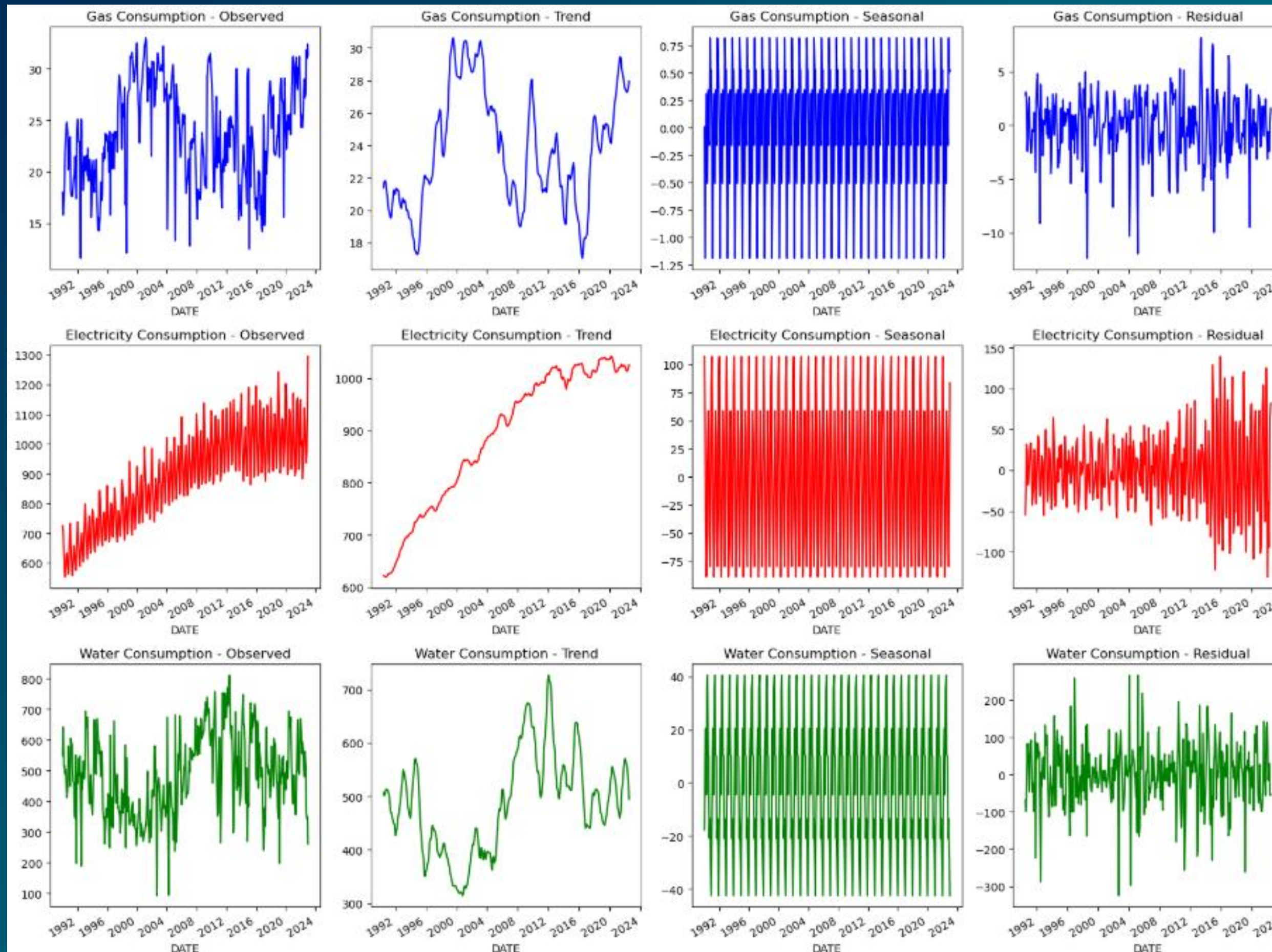


Time Series Analysis (ADF)

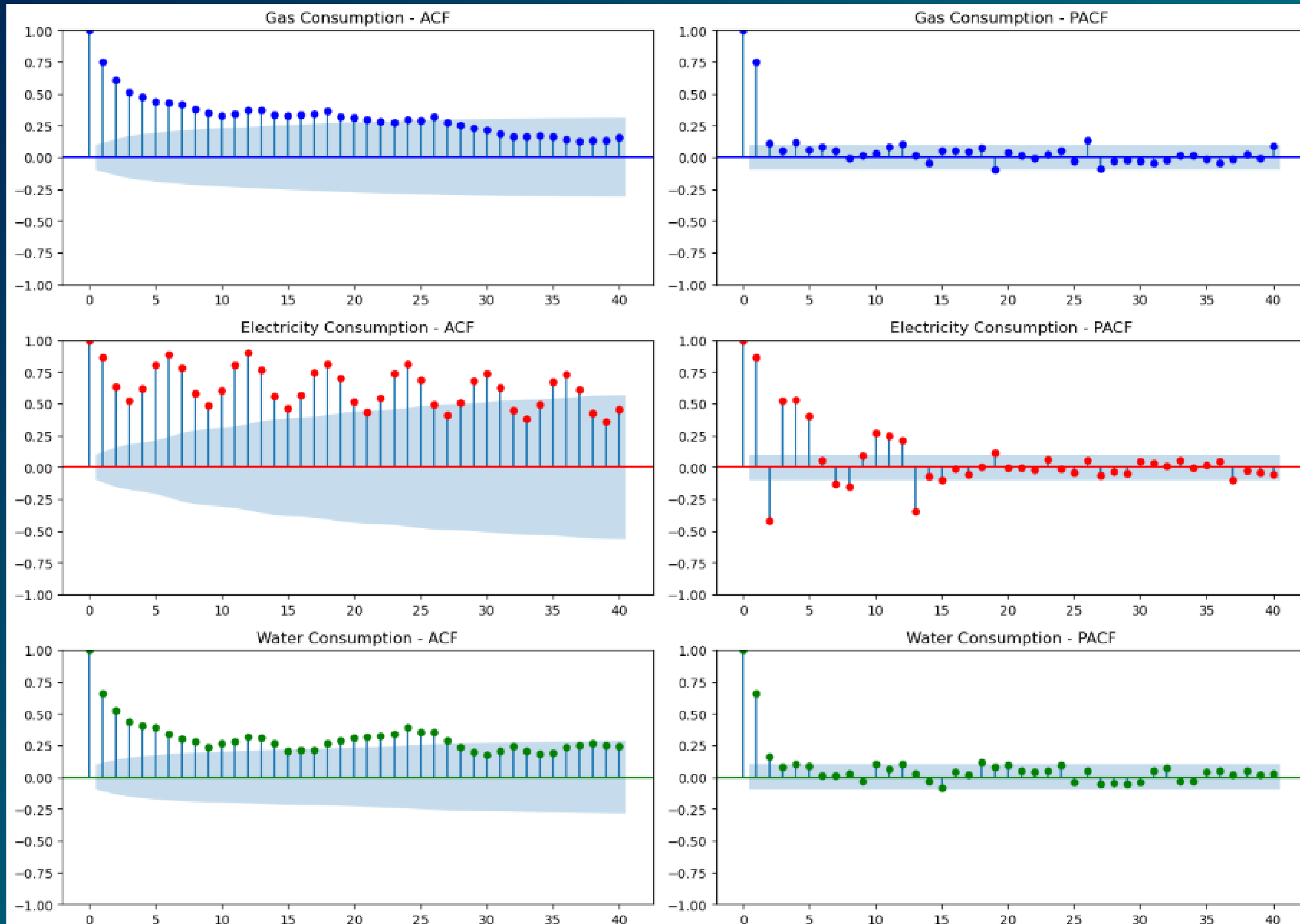


Variable	p-value (Before Differencing)	Stationary (Before Differencing)	p-value (After Differencing)	Stationary (After Differencing)
Gas Consumption	4.32e-03	True	1.94e-15	True
Electricity Consumption	1.67e-01	False	2.71e-08	True
Water Consumption	2.22e-04	True	1.50e-10	True

Time Series Analysis (Seasonal Decomp.)



Time Series Analysis (ACF / PACF)



Time Series Analysis (Ljung-Box Test)

Variable	p-value (First Differencing)	Autocorrelated (First Differencing)	p-value (Second Differencing)	Autocorrelated (Second Differencing)
Gas Consumption	2.23e-04	True	2.23e-23	True
Electricity Consumption	6.00e-244	True	5.18e-145	True
Water Consumption	4.95e-07	True	1.41e-27	True

Data Pre-Processing

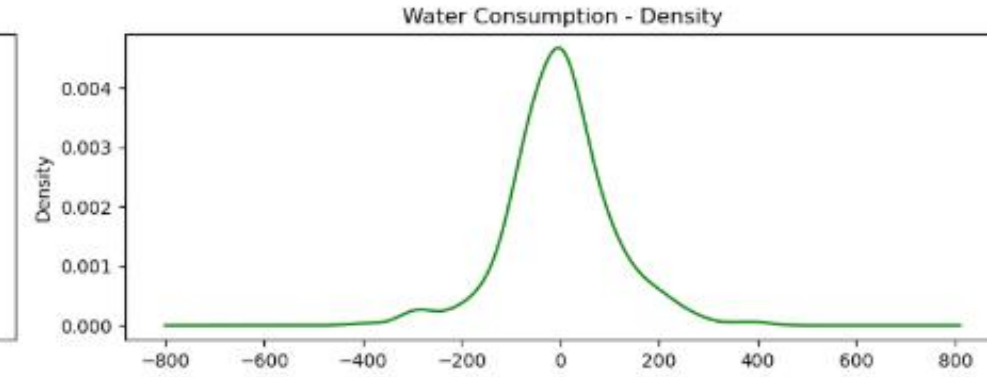
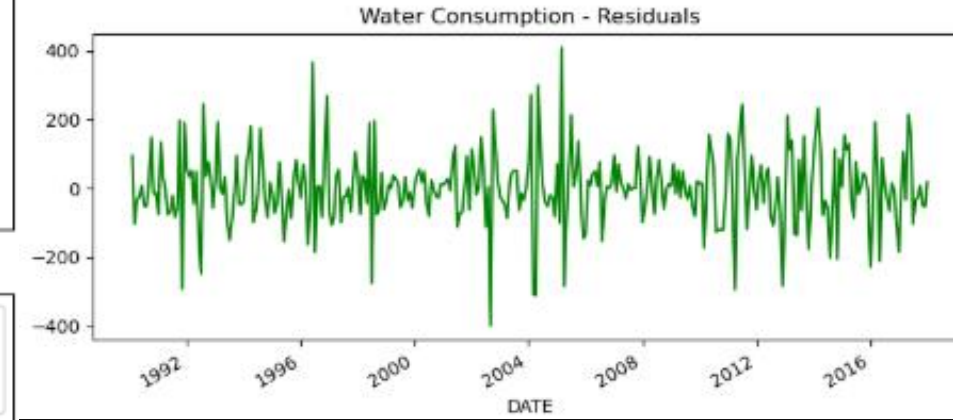
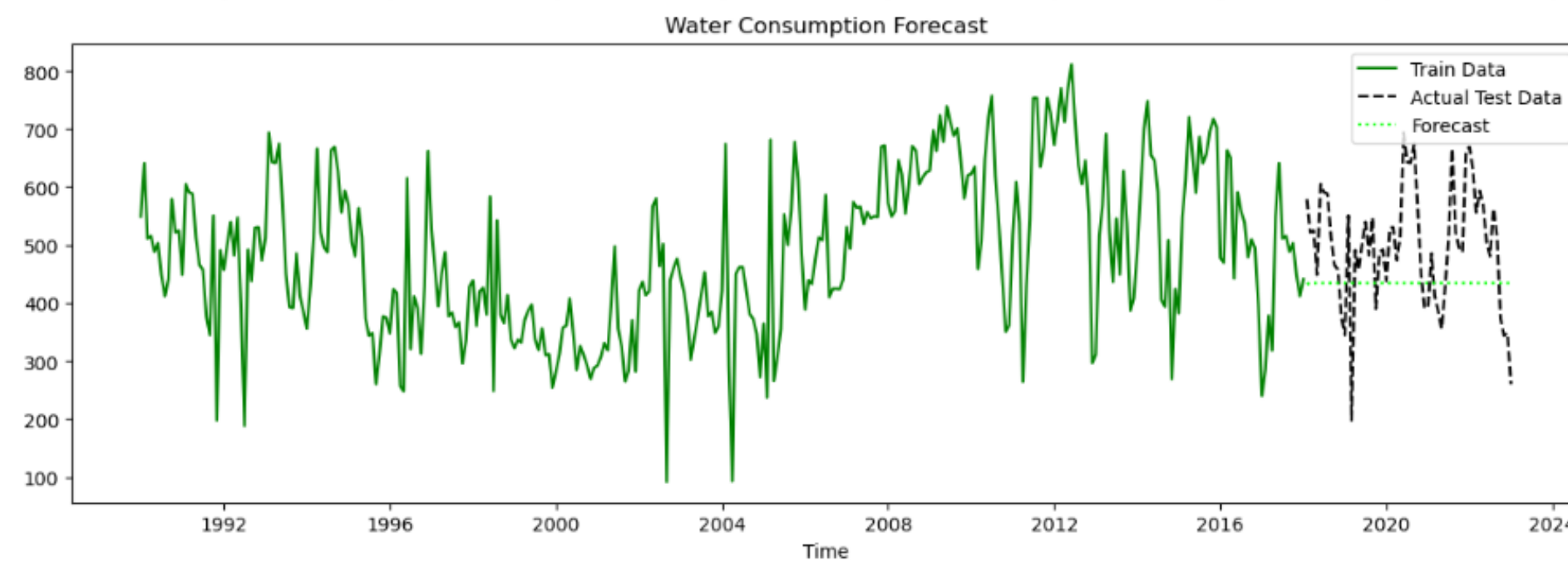
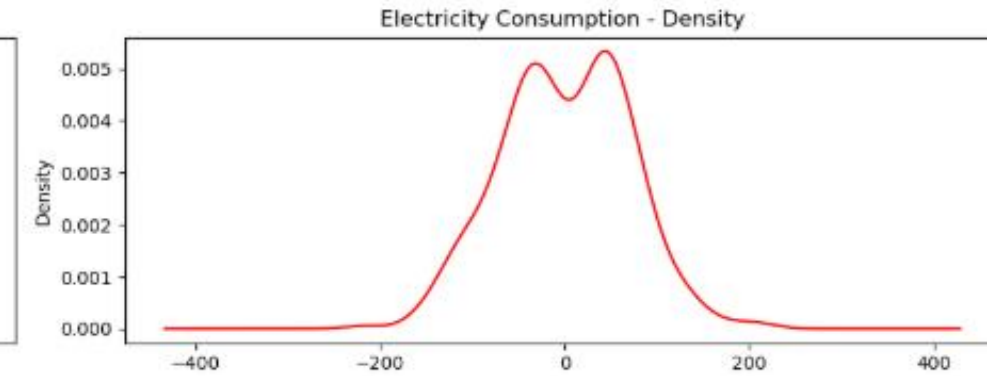
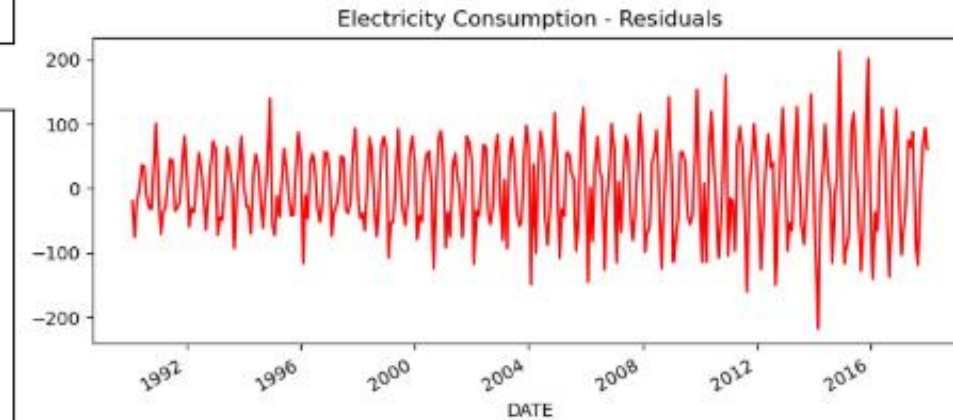
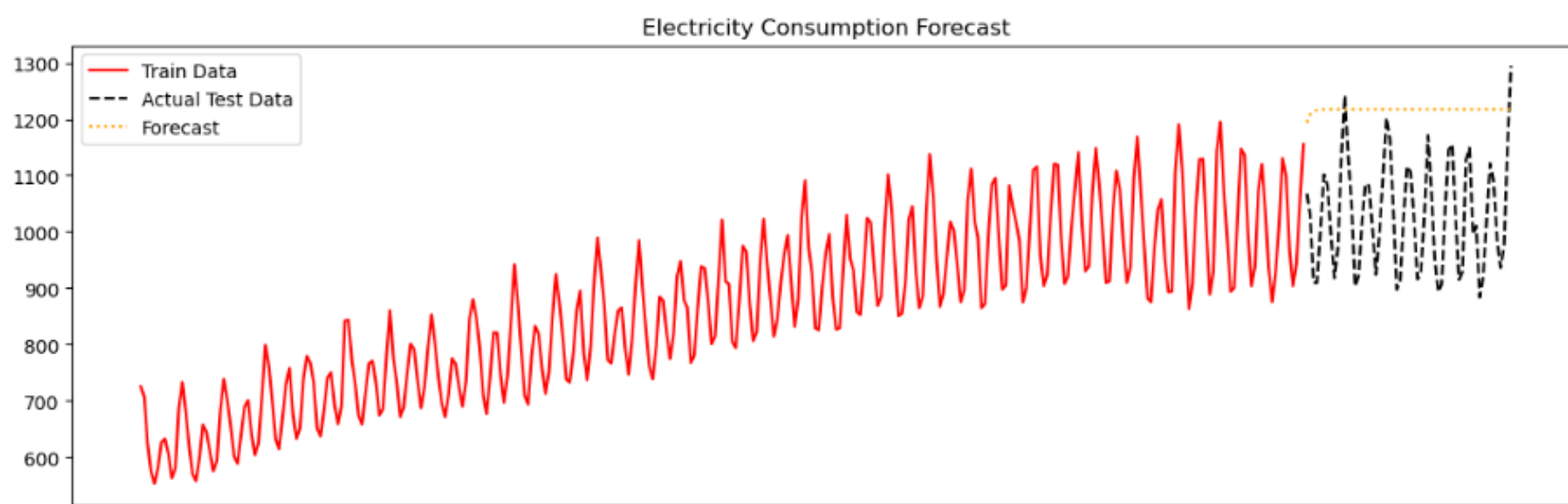
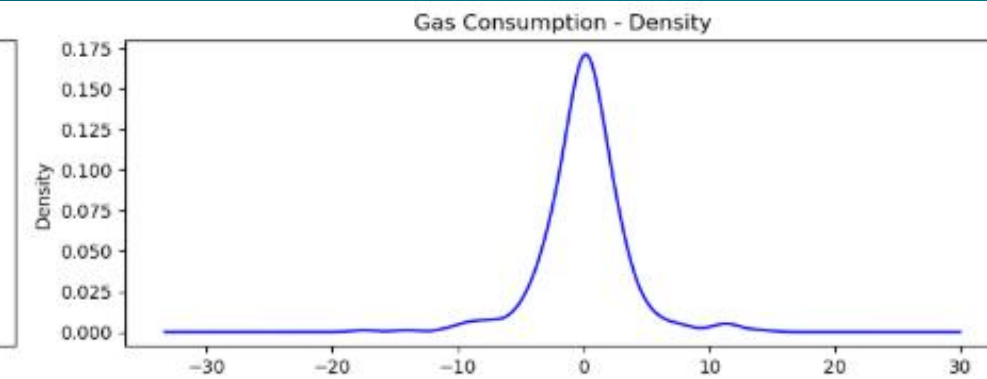
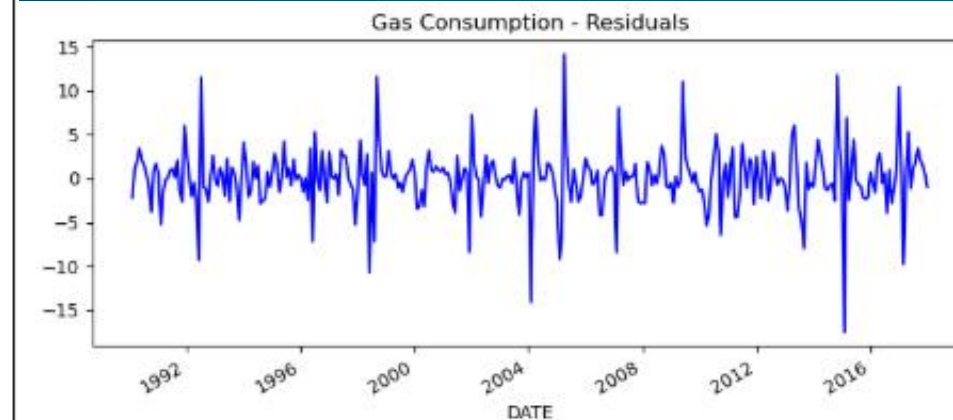
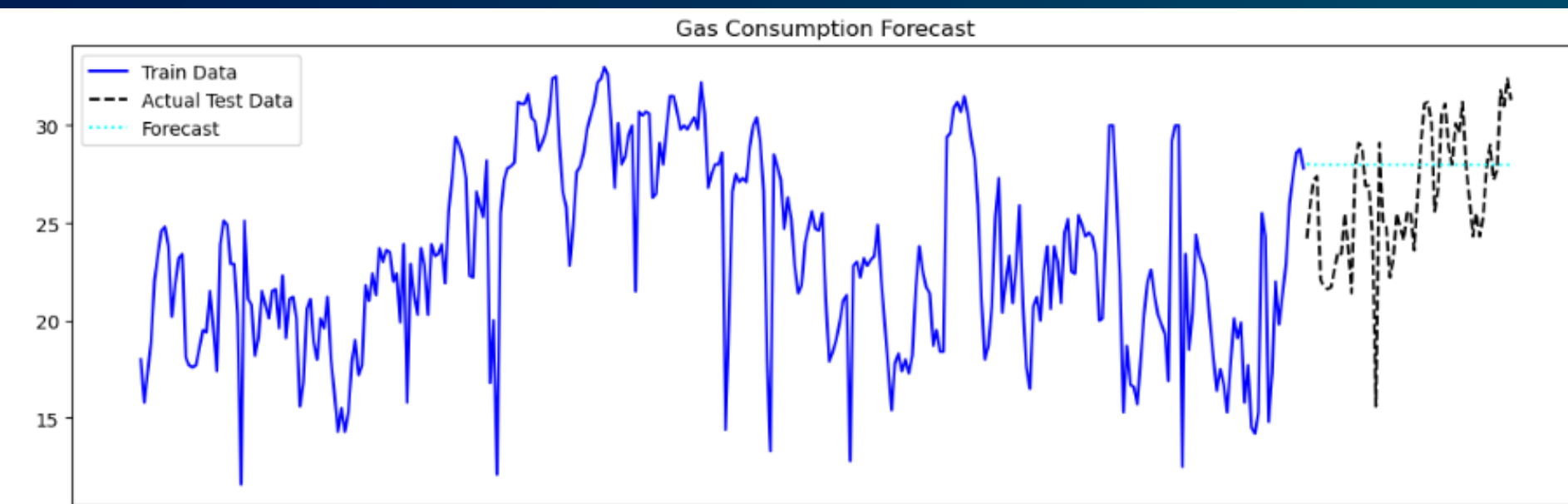
Data Logging

```
train_log = np.log(train)
test_log = np.log(test)
```

Train-Test Split

	Dataset	Size (Months)
0	Train Set	336
1	Test Set	60

Baseline Model (ARIMA)

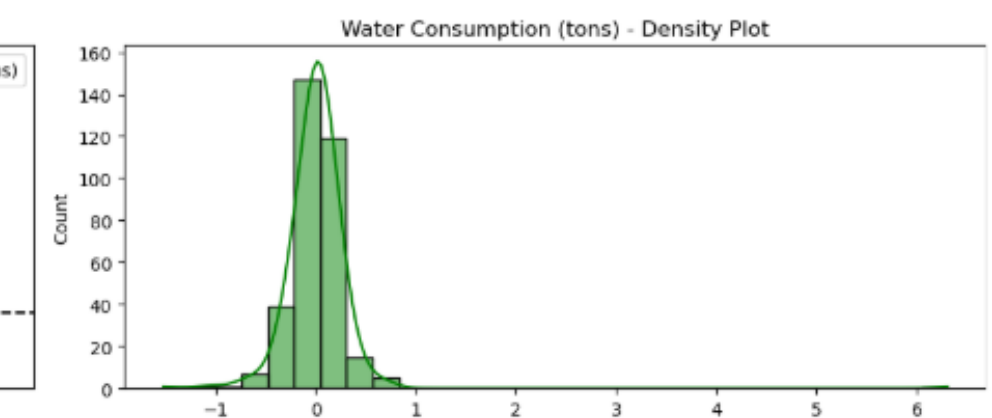
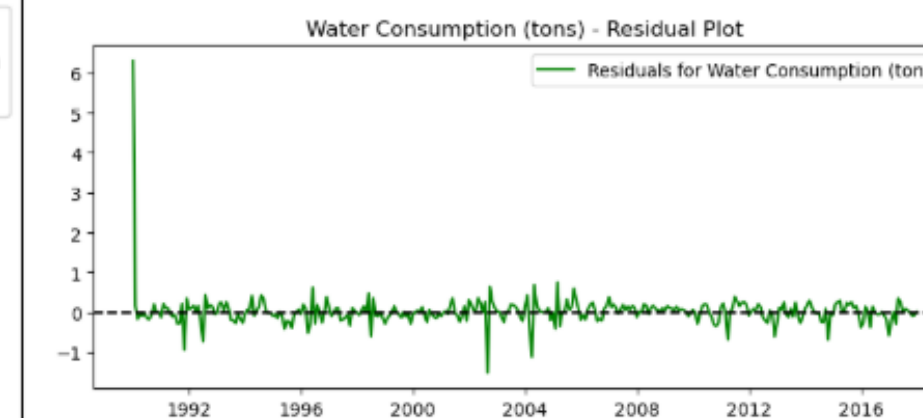
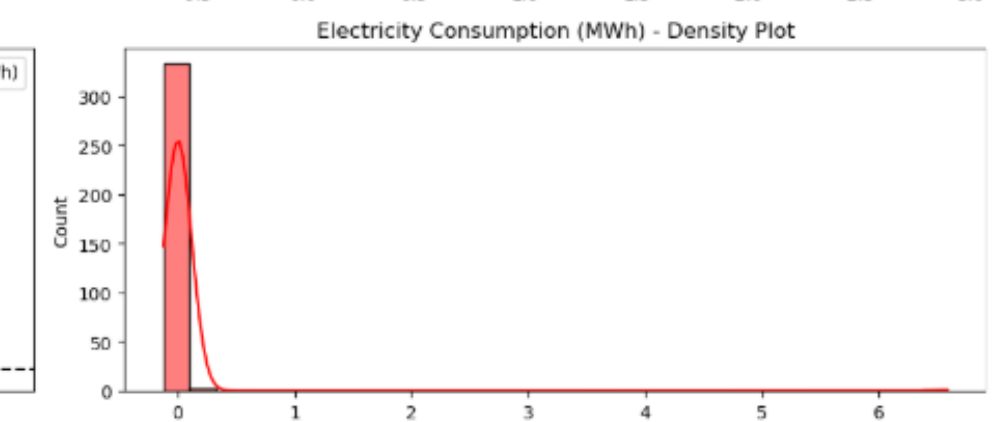
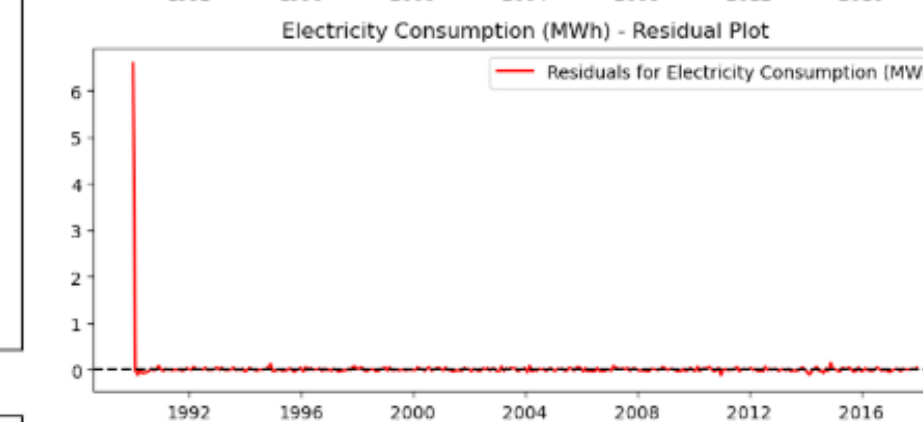
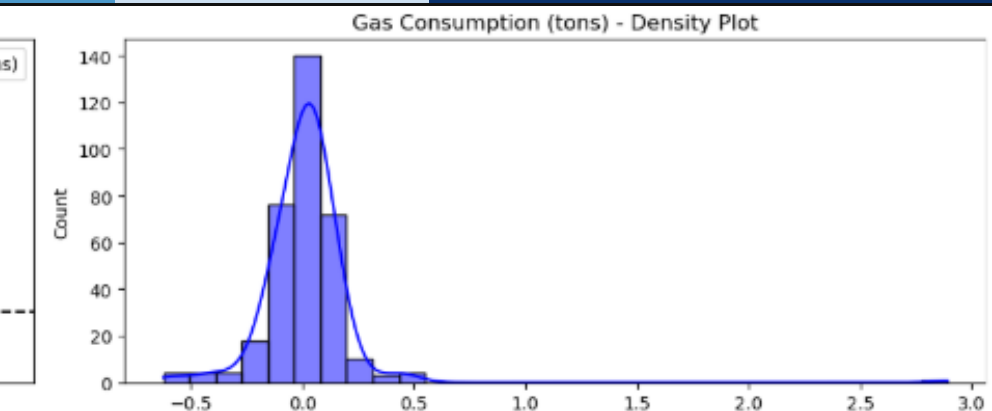
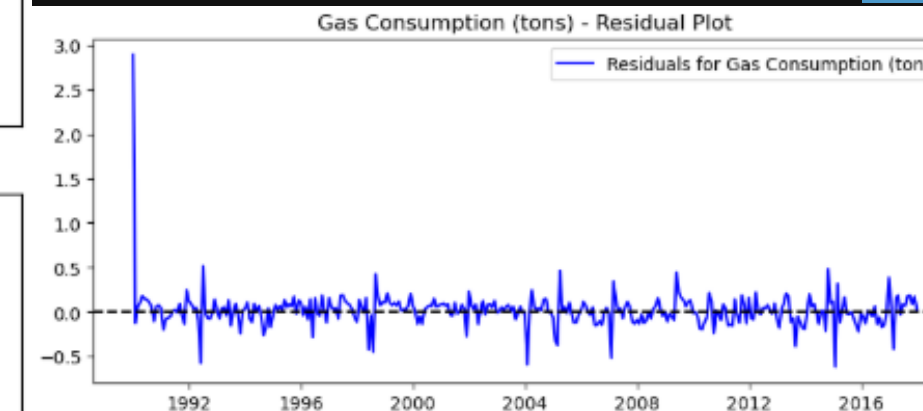
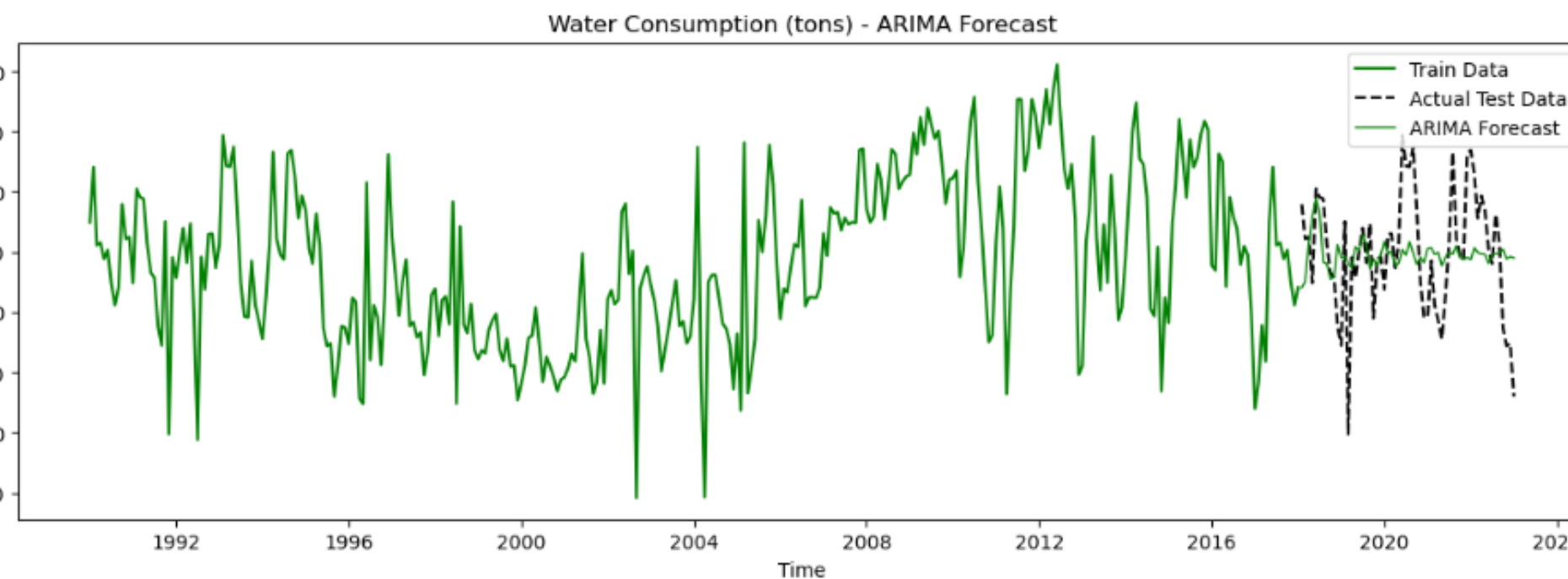
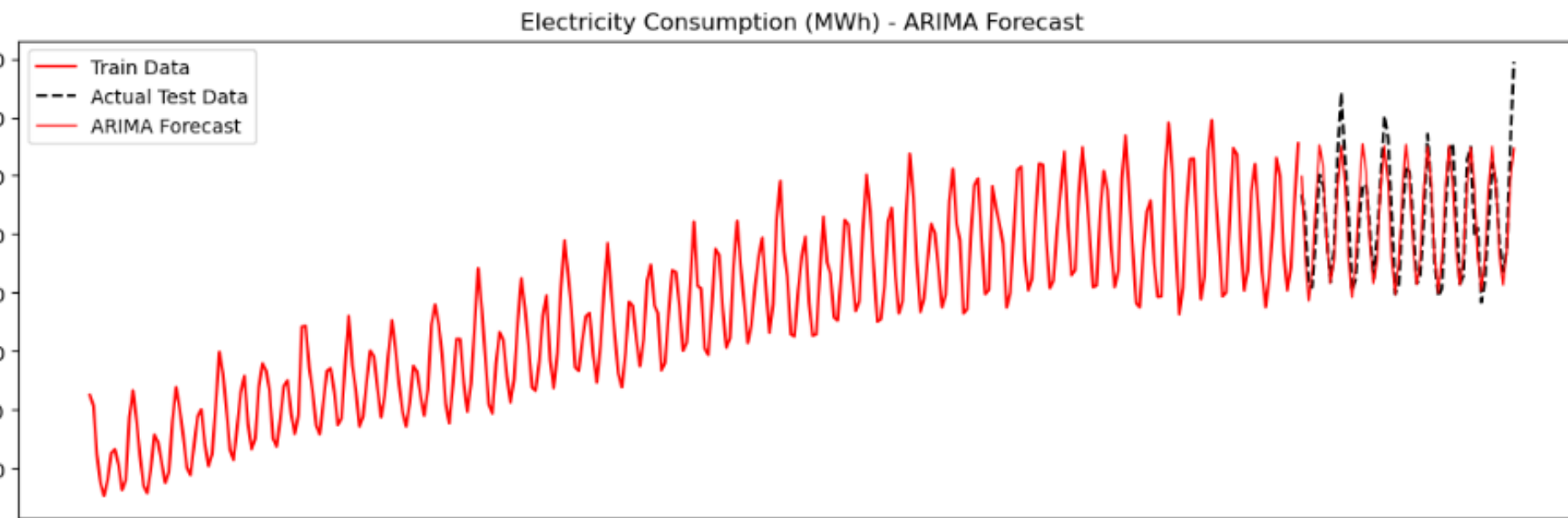
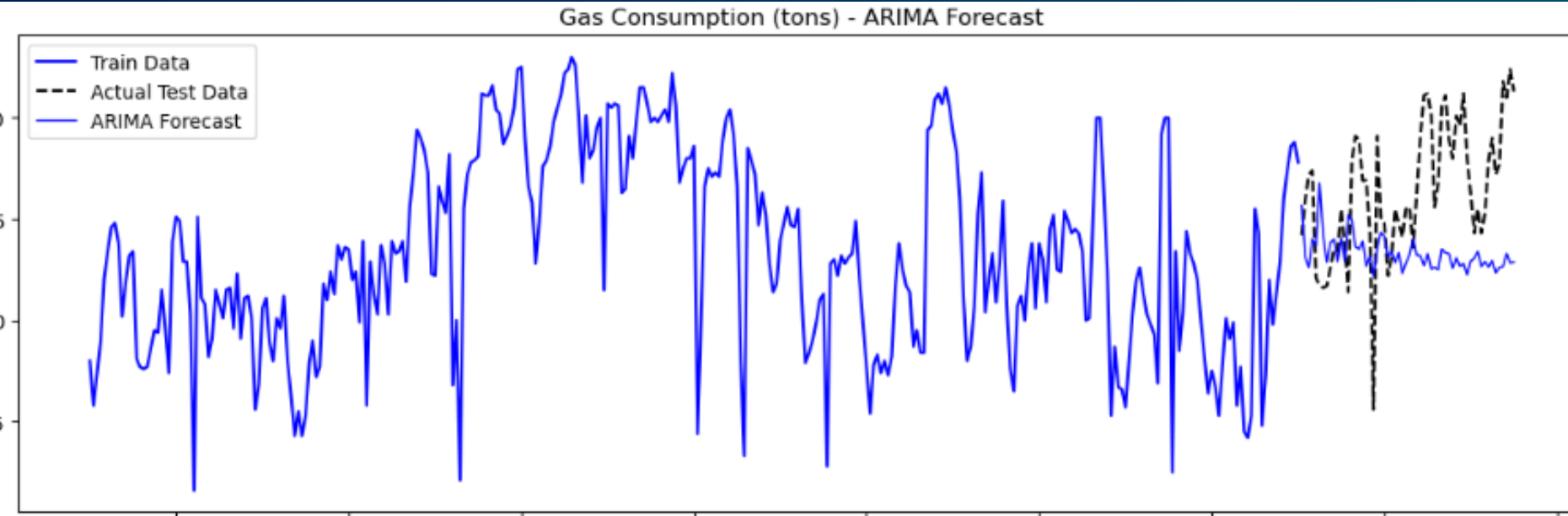


	MAE	RMSE	R2 Score	Explained Variance Score	MAPE
Gas Consumption (tons)	2.942169	3.638249	-0.220467	-0.000273	0.122218
Electricity Consumption (MWh)	189.930799	210.507290	-3.663881	-0.002225	0.194153
Water Consumption (tons)	100.197518	120.676837	-0.395805	-0.000493	0.204181

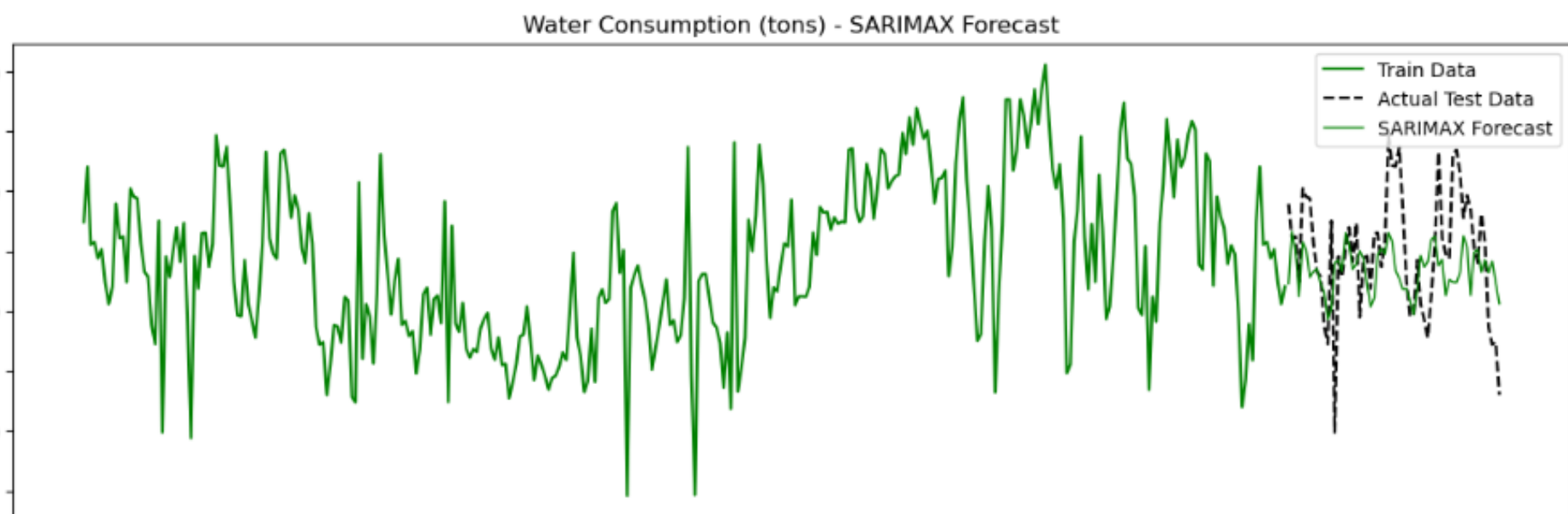
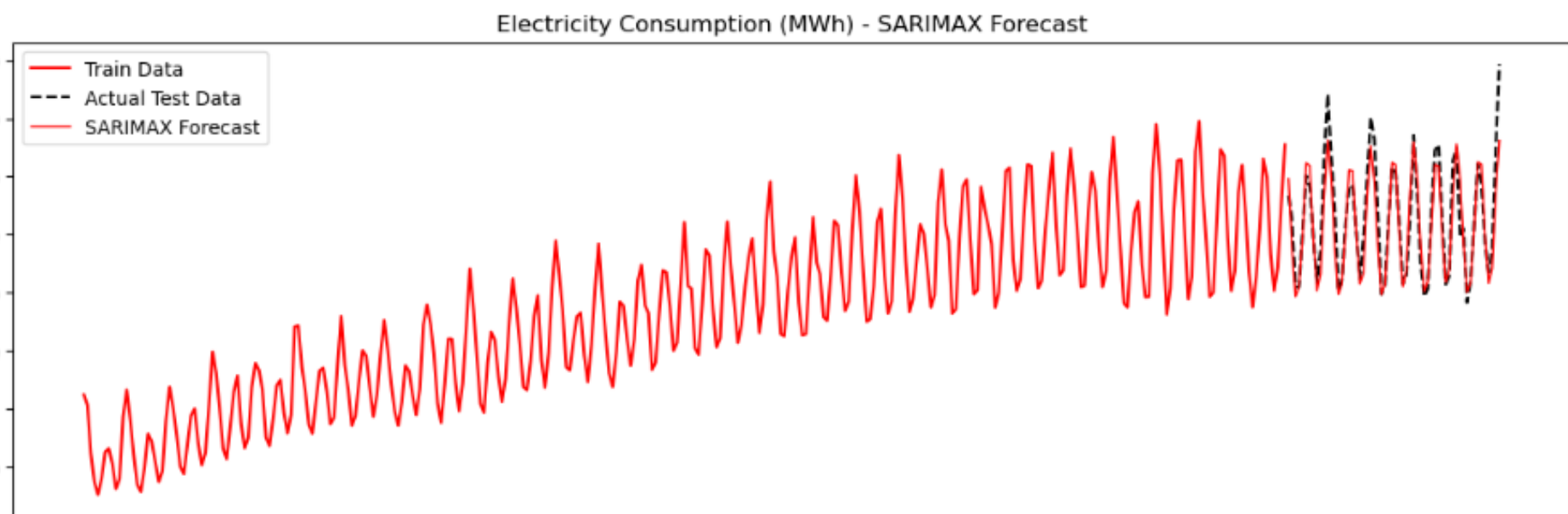
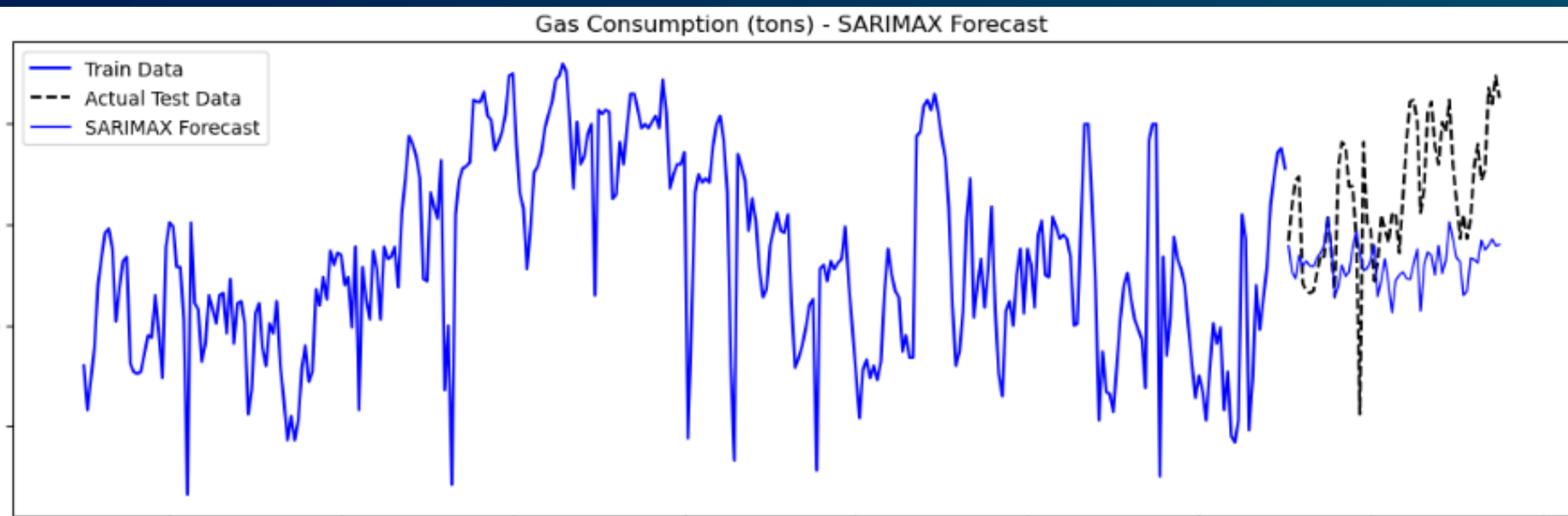
ARIMA Model

Grid Search's Best Parameter

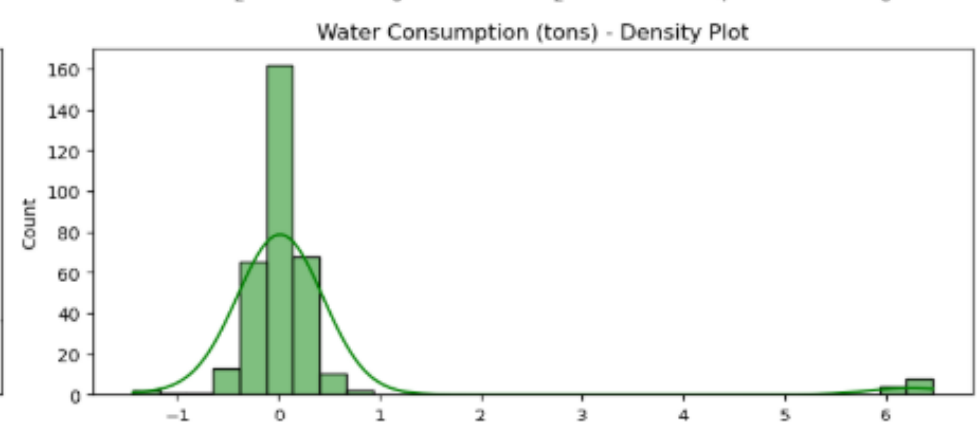
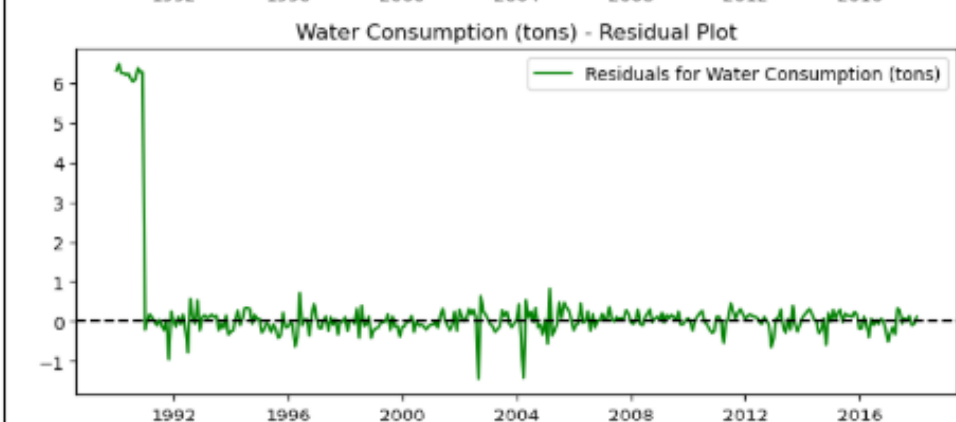
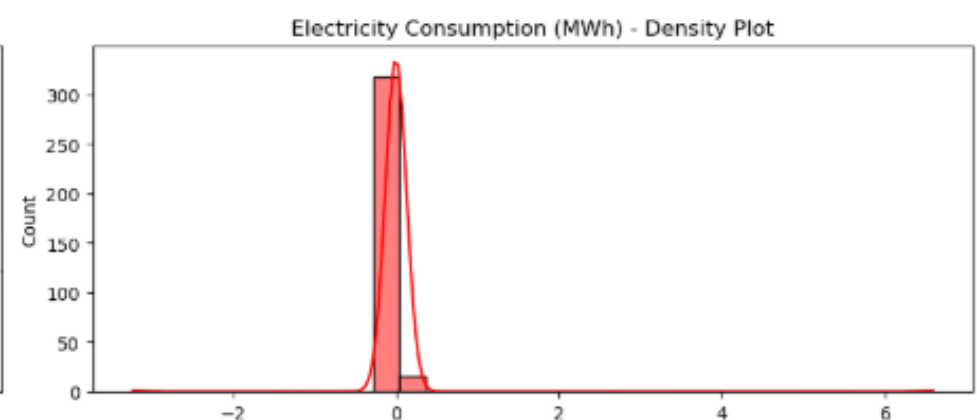
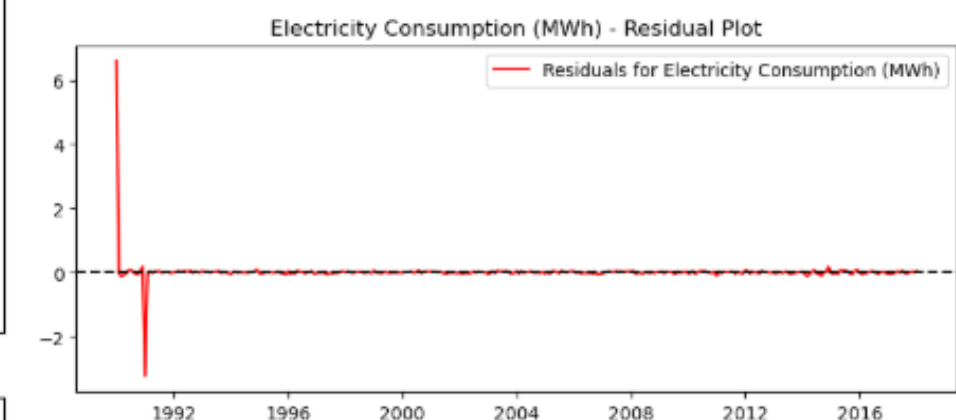
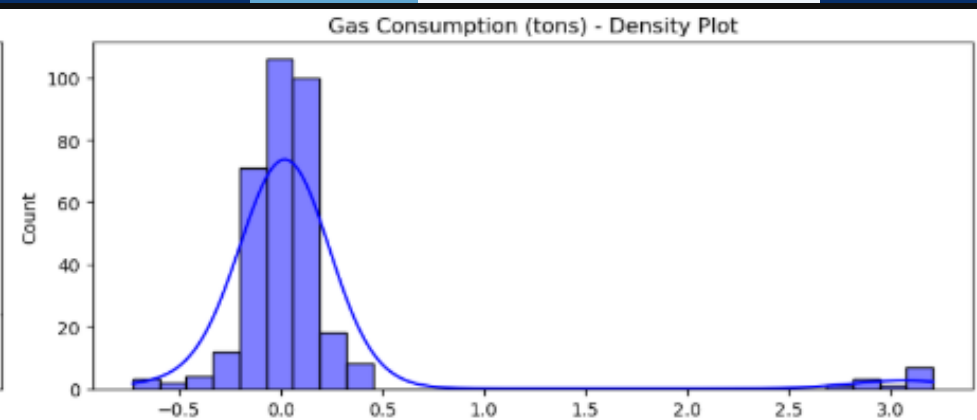
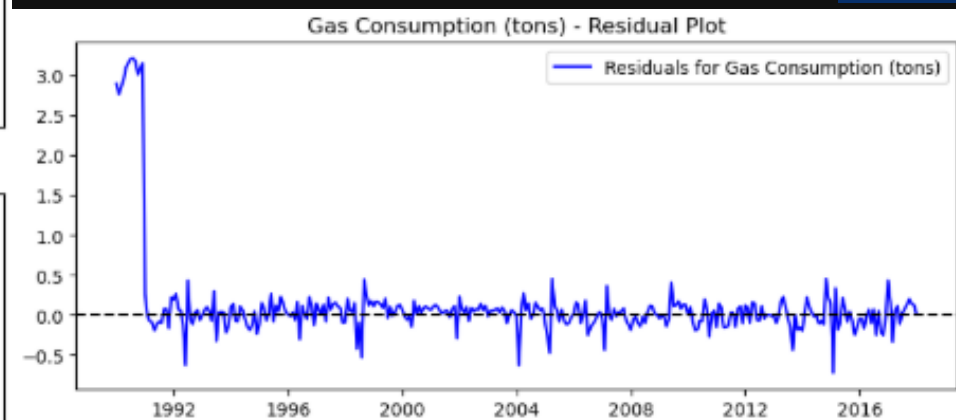
	Variable	Best ARIMA Order	R ² Score	Explained Variance	MSE	MAE	MAPE
0	Gas Consumption (tons)	(18, 1, 9)	-1.110402	-0.199329	22.888862	3.998905	0.145554
1	Electricity Consumption (MWh)	(8, 1, 8)	0.776596	0.779166	2122.643862	36.697694	0.034733
2	Water Consumption (tons)	(9, 1, 17)	0.014433	0.015053	10282.749063	77.724024	0.179988



SARIMAX Model



	Variable	SARIMAX Order	MAE	RMSE	R2 Score	Explained Variance Score	MAPE
0	Gas Consumption (tons)	((5, 0, 3), (9, 1, 6, 12))	3.904889	4.674279	-1.014514	0.018138	0.142284
1	Electricity Consumption (MWh)	((3, 1, 4), (5, 1, 6, 12))	28.479077	39.064938	0.839385	0.855936	0.026516
2	Water Consumption (tons)	((2, 0, 1), (9, 1, 6, 12))	79.174317	103.453577	-0.025812	0.055082	0.171447



Model Evaluation and Selection

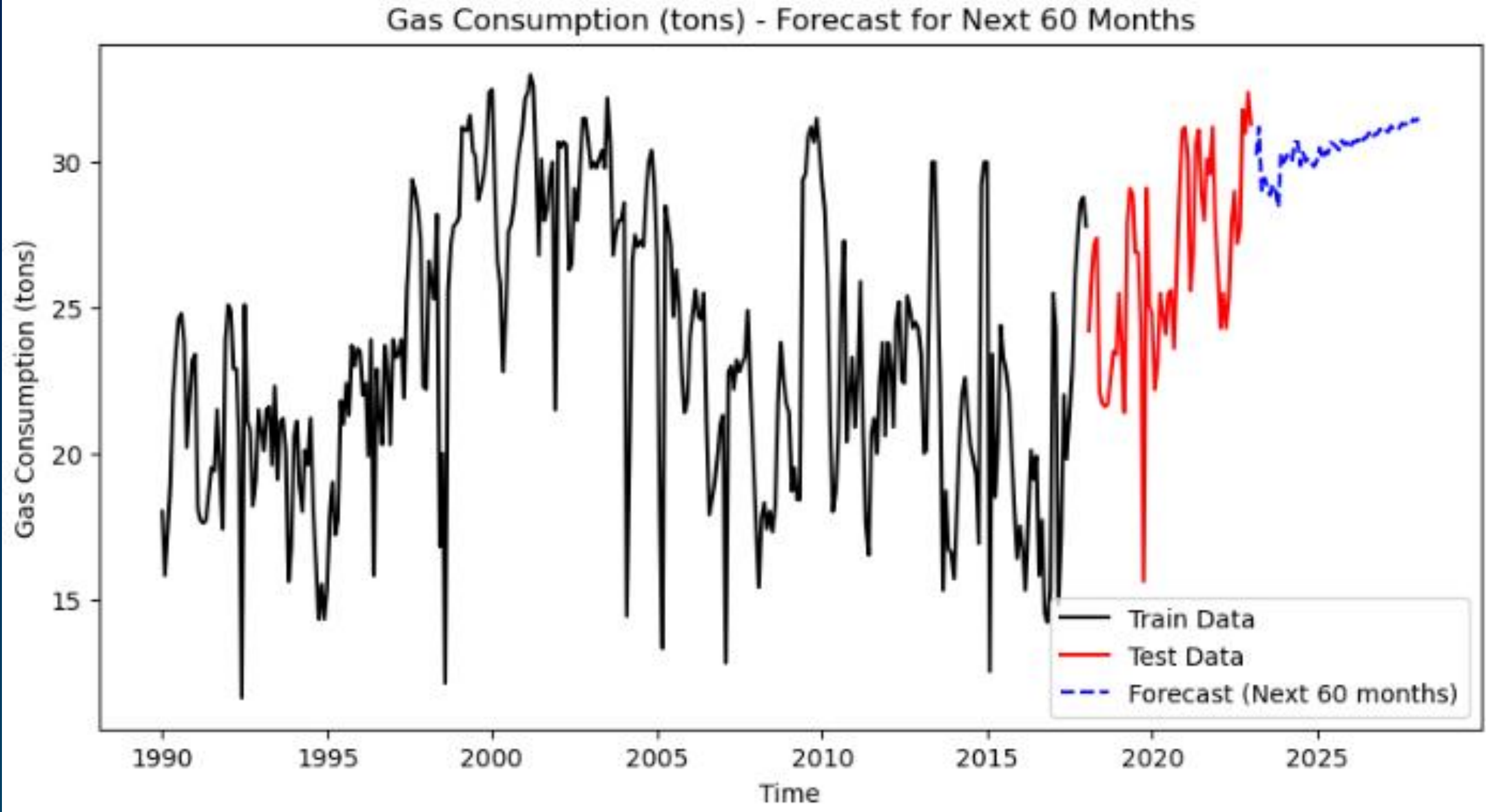
Evaluation

Model	Variable	Order	MAE	RMSE	R2 Score	Explained Variance Score	MAPE
Baseline Model	Gas Consumption (tons)	(1, 1, 0)	2.954328	3.652395	-0.229976	-0.000398	0.122795
Baseline Model	Electricity Consumption (MWh)	(1, 1, 0)	197.827313	218.059507	-4.004529	-0.002740	0.201997
Baseline Model	Water Consumption (tons)	(1, 1, 0)	100.947112	121.437014	-0.413446	-0.000753	0.205248
ARIMA	Gas Consumption (tons)	(18, 2, 9)	2.688515	3.307355	-0.008562	0.105145	0.101671
ARIMA	Electricity Consumption (MWh)	(8, 1, 8)	36.697694	46.072159	0.776596	0.779166	0.034733
ARIMA	Water Consumption (tons)	(9, 1, 17)	77.724024	101.403891	0.014433	0.015053	0.179988
SARIMAX	Gas Consumption (tons)	((5, 0, 3), (9, 1, 6, 12))	3.904889	4.674279	-1.014514	0.018138	0.142284
SARIMAX	Electricity Consumption (MWh)	((3, 1, 4), (5, 1, 6, 12))	28.479077	39.064938	0.839385	0.855936	0.026516
SARIMAX	Water Consumption (tons)	((2, 0, 1), (9, 1, 6, 12))	79.174317	103.453577	-0.025812	0.055082	0.171447

Selection

Model	Variable	Order	MAE	RMSE	R2 Score	Explained Variance Score	MAPE
SARIMAX	Electricity Consumption (MWh)	((3, 1, 4), (5, 1, 6, 12))	28.479077	39.064938	0.839385	0.855936	0.026516
ARIMA	Electricity Consumption (MWh)	(8, 1, 8)	36.697694	46.072159	0.776596	0.779166	0.034733
Baseline Model	Electricity Consumption (MWh)	(1, 1, 0)	197.827313	218.059507	-4.004529	-0.002740	0.201997
ARIMA	Gas Consumption (tons)	(18, 2, 9)	2.688515	3.307355	-0.008562	0.105145	0.101671
Baseline Model	Gas Consumption (tons)	(1, 1, 0)	2.954328	3.652395	-0.229976	-0.000398	0.122795
SARIMAX	Gas Consumption (tons)	((5, 0, 3), (9, 1, 6, 12))	3.904889	4.674279	-1.014514	0.018138	0.142284
ARIMA	Water Consumption (tons)	(9, 1, 17)	77.724024	101.403891	0.014433	0.015053	0.179988
SARIMAX	Water Consumption (tons)	((2, 0, 1), (9, 1, 6, 12))	79.174317	103.453577	-0.025812	0.055082	0.171447
Baseline Model	Water Consumption (tons)	(1, 1, 0)	100.947112	121.437014	-0.413446	-0.000753	0.205248

Model Deployment (Gas)

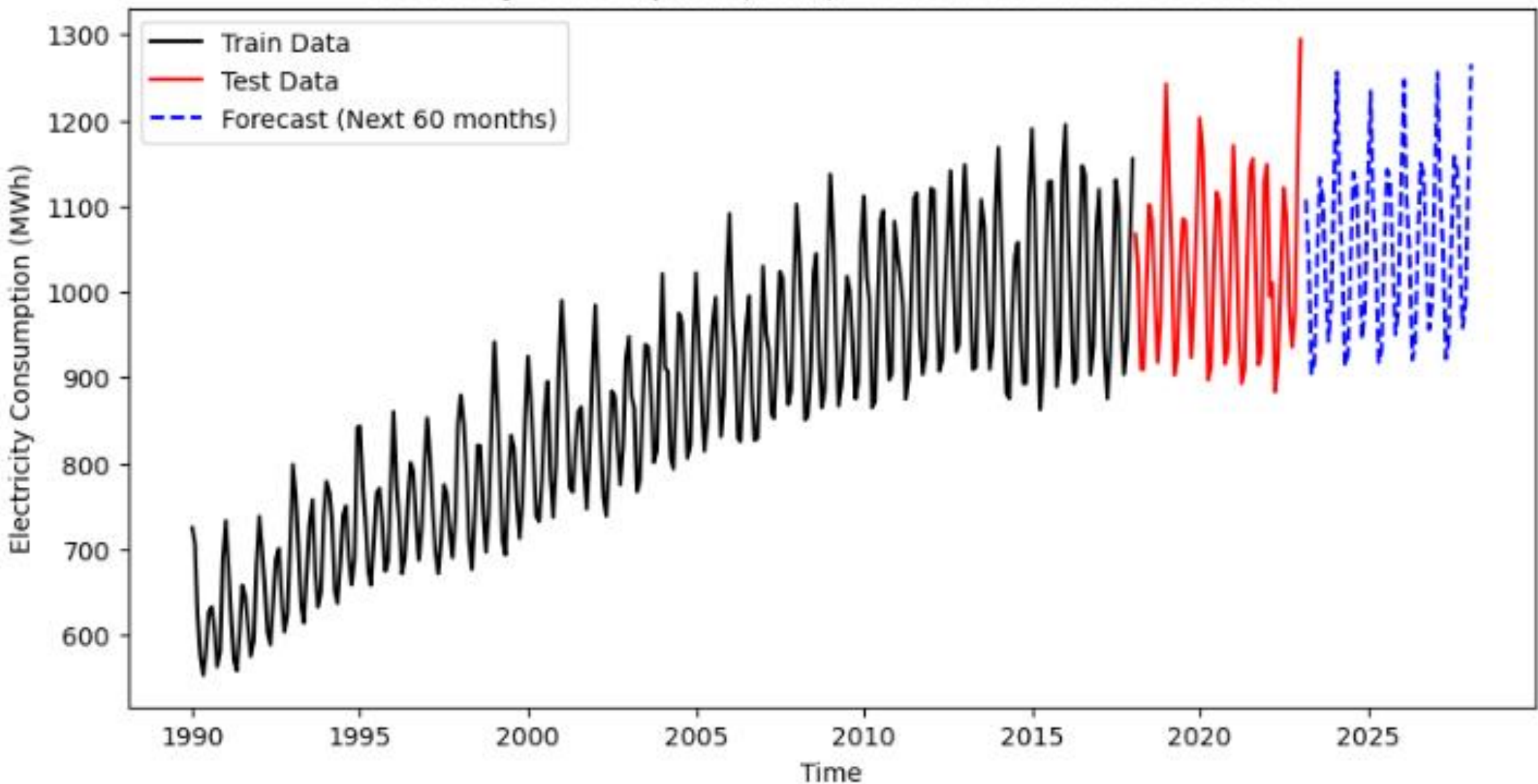


SARIMAX Results						
=====						
Dep. Variable:	Gas Consumption (tons)		No. Observations:		396	
Model:	ARIMA(18, 2, 9)		Log Likelihood		201.573	
Date:	Fri, 14 Feb 2025		AIC		-347.146	
Time:	22:52:16		BIC		-235.809	
Sample:	0		HQIC		-303.029	
	- 396					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

ar.L1	-1.2956	4.399	-0.295	0.768	-9.917	7.326
ar.L2	-1.2258	4.886	-0.251	0.802	-10.802	8.350
ar.L3	-1.2024	4.885	-0.246	0.806	-10.776	8.371
ar.L4	-1.2116	4.903	-0.247	0.805	-10.820	8.397
ar.L5	-1.1339	4.850	-0.234	0.815	-10.639	8.371
ar.L6	-0.9735	4.446	-0.219	0.827	-9.687	7.740
ar.L7	-1.2645	3.754	-0.337	0.736	-8.622	6.093
ar.L8	-1.0170	4.990	-0.204	0.839	-10.797	8.763
ar.L9	-0.8272	3.471	-0.238	0.812	-7.630	5.976
ar.L10	-0.8135	3.048	-0.267	0.790	-6.787	5.160
ar.L11	-0.7579	3.072	-0.247	0.805	-6.779	5.264
ar.L12	-0.6185	2.835	-0.218	0.827	-6.175	4.938
ar.L13	-0.4687	2.305	-0.203	0.839	-4.986	4.048
ar.L14	-0.3811	1.786	-0.213	0.831	-3.881	3.119
ar.L15	-0.3383	1.460	-0.232	0.817	-3.199	2.523
ar.L16	-0.2721	1.295	-0.210	0.834	-2.810	2.266
ar.L17	-0.1907	1.023	-0.186	0.852	-2.196	1.815
ar.L18	-0.0236	0.708	-0.033	0.973	-1.411	1.364
ma.L1	-0.1212	4.423	-0.027	0.978	-8.790	8.548
ma.L2	-0.3341	1.423	-0.235	0.814	-3.124	2.455
ma.L3	-0.1831	0.852	-0.215	0.830	-1.852	1.486
ma.L4	0.0013	0.749	0.002	0.999	-1.467	1.470
ma.L5	-0.1419	0.261	-0.544	0.586	-0.653	0.369
ma.L6	-0.1777	0.773	-0.230	0.818	-1.692	1.336
ma.L7	0.4798	0.738	0.650	0.516	-0.966	1.926
ma.L8	-0.3614	2.162	-0.167	0.867	-4.599	3.876
ma.L9	-0.1572	2.147	-0.073	0.942	-4.365	4.051
sigma2	0.0202	0.002	9.330	0.000	0.016	0.024
=====						
Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	290.68			
Prob(Q):	0.97	Prob(JB):	0.00			
Heteroskedasticity (H):	1.19	Skew:	-0.55			
Prob(H) (two-sided):	0.33	Kurtosis:	7.06			
=====						

Model Deployment (Energy)

Electricity Consumption (MWh) - Forecast for Next 60 Months



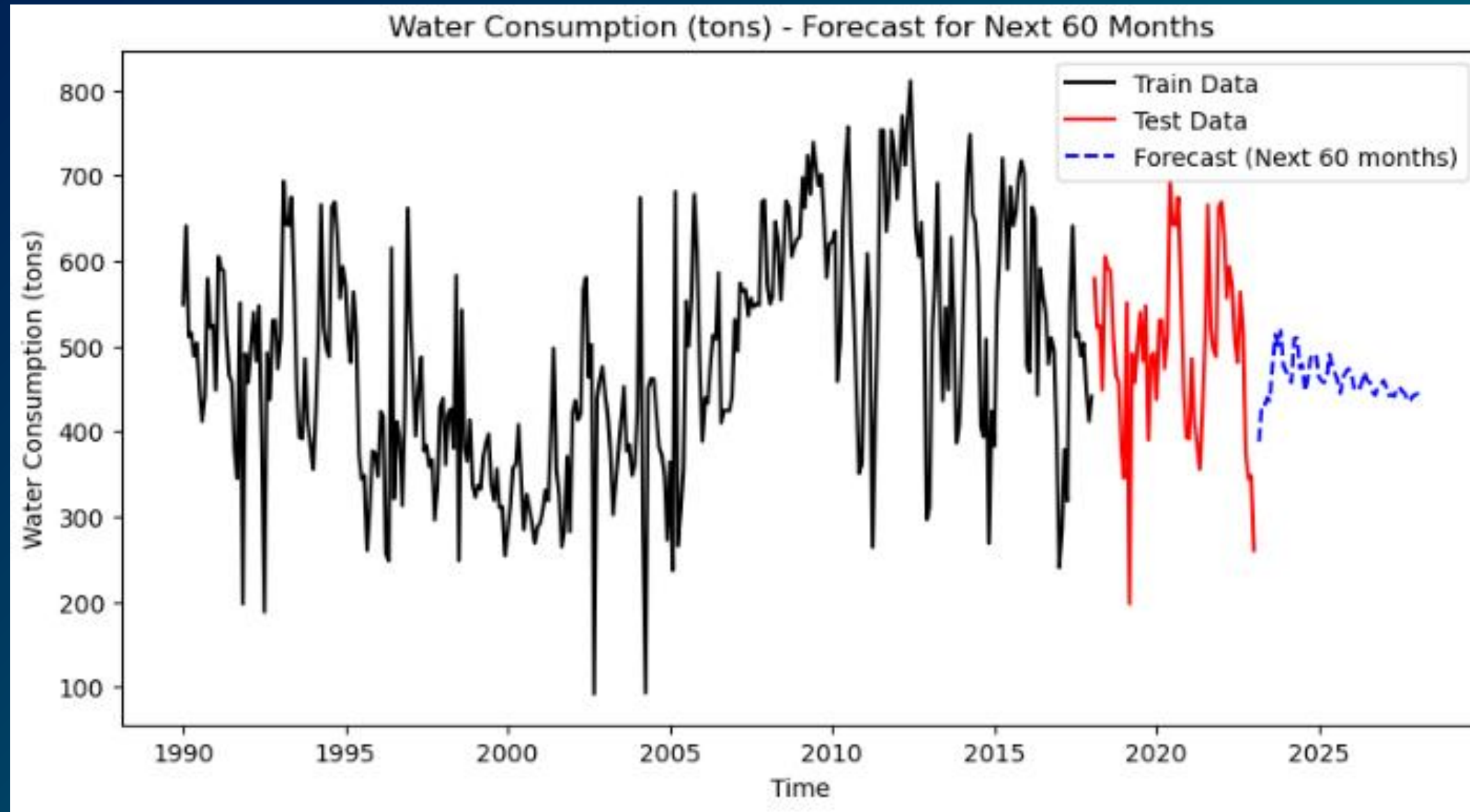
SARIMAX Results

```
Dep. Variable: Electricity Consumption (MWh) No. Observations: 396
Model: SARIMAX(3, 1, 4)x(5, 1, [1, 2, 3, 4, 5, 6], 12) Log Likelihood: 801.401
Date: Fri, 14 Feb 2025 AIC: -1564.802
Time: 22:54:10 BIC: -1489.789
Sample: 0 HQIC: -1535.046
Covariance Type: opg
```

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.2320	0.270	-0.860	0.390	-0.761	0.297
ar.L2	-0.2501	0.257	-0.974	0.330	-0.753	0.253
ar.L3	-0.2291	0.123	-1.859	0.063	-0.471	0.012
ma.L1	-0.1178	0.270	-0.436	0.663	-0.647	0.412
ma.L2	-0.2450	0.243	-1.010	0.313	-0.721	0.231
ma.L3	-0.0892	0.244	-0.365	0.715	-0.568	0.389
ma.L4	-0.3456	0.126	-2.750	0.006	-0.592	-0.099
ar.S.L12	-0.3880	0.960	-0.404	0.686	-2.269	1.493
ar.S.L24	-0.4432	0.826	-0.537	0.592	-2.062	1.176
ar.S.L36	-0.3074	0.752	-0.409	0.683	-1.781	1.166
ar.S.L48	-0.2531	0.556	-0.456	0.649	-1.342	0.836
ar.S.L60	-0.3781	0.334	-1.133	0.257	-1.032	0.276
ma.S.L12	-0.0550	0.963	-0.057	0.954	-1.942	1.832
ma.S.L24	0.0435	0.758	0.057	0.954	-1.443	1.530
ma.S.L36	0.0699	0.795	0.088	0.930	-1.488	1.628
ma.S.L48	0.1102	0.680	0.162	0.871	-1.224	1.444
ma.S.L60	0.2754	0.319	0.862	0.389	-0.351	0.901
ma.S.L72	-0.1878	0.216	-0.871	0.384	-0.611	0.235
sigma2	0.0009	5.88e-05	14.740	0.000	0.001	0.001

```
Ljung-Box (L1) (Q): 0.08 Jarque-Bera (JB): 191.88
Prob(Q): 0.78 Prob(JB): 0.00
Heteroskedasticity (H): 2.83 Skew: 0.24
Prob(H) (two-sided): 0.00 Kurtosis: 6.44
```

Model Deployment (Water)



SARIMAX Results						
=====						
Dep. Variable:	Water Consumption (tons)		No. Observations:		396	
Model:	ARIMA(9, 2, 17)		Log Likelihood		3.182	
Date:	Fri, 14 Feb 2025		AIC		47.637	
Time:	22:54:15		BIC		154.998	
Sample:	0		HQIC		90.179	
	- 396					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

ar.L1	-1.6806	0.354	-4.745	0.000	-2.375	-0.986
ar.L2	-2.6320	0.624	-4.218	0.000	-3.855	-1.409
ar.L3	-3.1297	0.848	-3.693	0.000	-4.791	-1.468
ar.L4	-3.5788	0.971	-3.684	0.000	-5.483	-1.675
ar.L5	-3.2682	1.122	-2.912	0.004	-5.468	-1.068
ar.L6	-2.7778	0.954	-2.912	0.004	-4.647	-0.908
ar.L7	-1.9556	0.796	-2.458	0.014	-3.515	-0.396
ar.L8	-1.1374	0.547	-2.079	0.038	-2.210	-0.065
ar.L9	-0.4898	0.295	-1.662	0.097	-1.067	0.088
ma.L1	0.0818	0.364	0.225	0.822	-0.631	0.795
ma.L2	0.4417	0.299	1.479	0.139	-0.143	1.027
ma.L3	-0.2097	0.253	-0.829	0.407	-0.705	0.286
ma.L4	-0.0118	0.253	-0.047	0.963	-0.508	0.484
ma.L5	-0.7101	0.264	-2.691	0.007	-1.227	-0.193
ma.L6	-0.4597	0.398	-1.155	0.248	-1.240	0.320
ma.L7	-0.6257	0.258	-2.423	0.015	-1.132	-0.120
ma.L8	-0.3533	0.291	-1.212	0.226	-0.925	0.218
ma.L9	-0.1716	0.254	-0.676	0.499	-0.669	0.326
ma.L10	0.0103	0.247	0.042	0.967	-0.473	0.494
ma.L11	0.4020	0.204	1.972	0.049	0.002	0.802
ma.L12	0.2371	0.087	2.716	0.007	0.066	0.408
ma.L13	0.1707	0.094	1.818	0.069	-0.013	0.355
ma.L14	0.2255	0.085	2.647	0.008	0.059	0.392
ma.L15	0.0705	0.103	0.684	0.494	-0.132	0.273
ma.L16	0.1304	0.070	1.868	0.062	-0.006	0.267
ma.L17	-0.1038	0.073	-1.429	0.153	-0.246	0.039
sigma2	0.0560	0.003	18.325	0.000	0.050	0.062
=====						
Ljung-Box (L1) (Q):	0.03	Jarque-Bera (JB):	702.89			
Prob(Q):	0.86	Prob(JB):	0.00			
Heteroskedasticity (H):	0.99	Skew:	-1.19			
Prob(H) (two-sided):	0.95	Kurtosis:	9.10			
=====						

Overall Deployment

	Mean Forecast	Min Forecast	Max Forecast
Gas Consumption (tons)	30.50	28.44	31.51
Electricity Consumption (MWh)	1055.06	904.81	1265.85
Water Consumption (tons)	460.42	387.92	518.04



AIML CA 2

Thank You For Your Attention

