INDIVIDUAL ASSESSMENT COVER SHEET

Faculty of Design and Creative Technologies

AUT
TE WĀNANGA ARONUI O TĀMAKI MAKAU RAU

First Name	**Thandar**	**Family Name**	**Kyi**	**Student ID No**	24251155
Course Name	**Mathematical Modelling and Simulation**	**Course Code**	**MATH803**	**Assignment Due Date**	11/06/2025
Lecturer	**Dr. Nuttanan Wichitaksorn/ Dr. Shu Su**	**Tutorial Day**	**April ~ May 2025**	**Date Submitted**	08/06/2025
Tutor	**Dr. Shu Su**	**Tutorial Time**		**No.Words/Pages**	**12**

In order to ensure fair and honest assessment results for all students, it is a requirement that the work that you hand in for assessment is your own work. If you are uncertain about any of these matters, then please discuss them with your lecturer.

Plagiarism and Dishonesty are methods of cheating for the purposes of General Academic Regulations (GAR) http://www.aut.ac.nz/calendar

Assignments will not be accepted if this section is not completed and signed.

Please read the following and **tick** ✓ to indicate your understanding:

1. I understand it is my responsibility to keep a copy of my assignment.	☒ **Yes**	☐ **No**
2. I have signed and read the **Student's Statement below**.	☒ **Yes**	☐ **No**
3. I understand that a software programme (Turnitin) that detects plagiarism and copying may be used on my assignment.	☒ **Yes**	☐ **No**

Student's Statement:

This assessment is entirely my own work and has not been submitted in any other course of study. I have submitted a copy of this assessment to Turnitin, if required. In this assessment I have acknowledged, to the best of my ability:

- The source of direct quotes from the work of others.
- The ideas of others (includes work from private or professional services, past assessments, other students, books, journals, cut/paste from internet sites and/or other materials).
- The source of diagrams or visual images.

Student's Signature: Thr **Date: 08/06/2025**

The information on this form is collected for the primary purpose of submitting your assignment for assessment. Other purposes of collection include receiving your acknowledgement of plagiarism polices and attending to administrative matters. If you choose not to complete all questions on this form, it may not be possible for the Faculty of Design and Creative Technologies to accept your assignment.

MATH803: Mathematical Modelling and Simulation
Project Part B: Forecast of Electricity Consumption in North Island, Auckland
Student ID: 24251155

1. Data Organization and Preliminary Assessment

1.(a) Import the data.

The data from NorthIslandHourlyElec2-1.xlsx has been imported as per below SAS codes:

```
/* Import Excel data */
proc import datafile="/home/u64175686/sasuser.v94/MATH803/NorthIslandHourlyElec2-1.xlsx"
   out=elec_data
   dbms=xlsx
   replace;
   sheet="Sheet1";
   getnames=yes;
run;
```

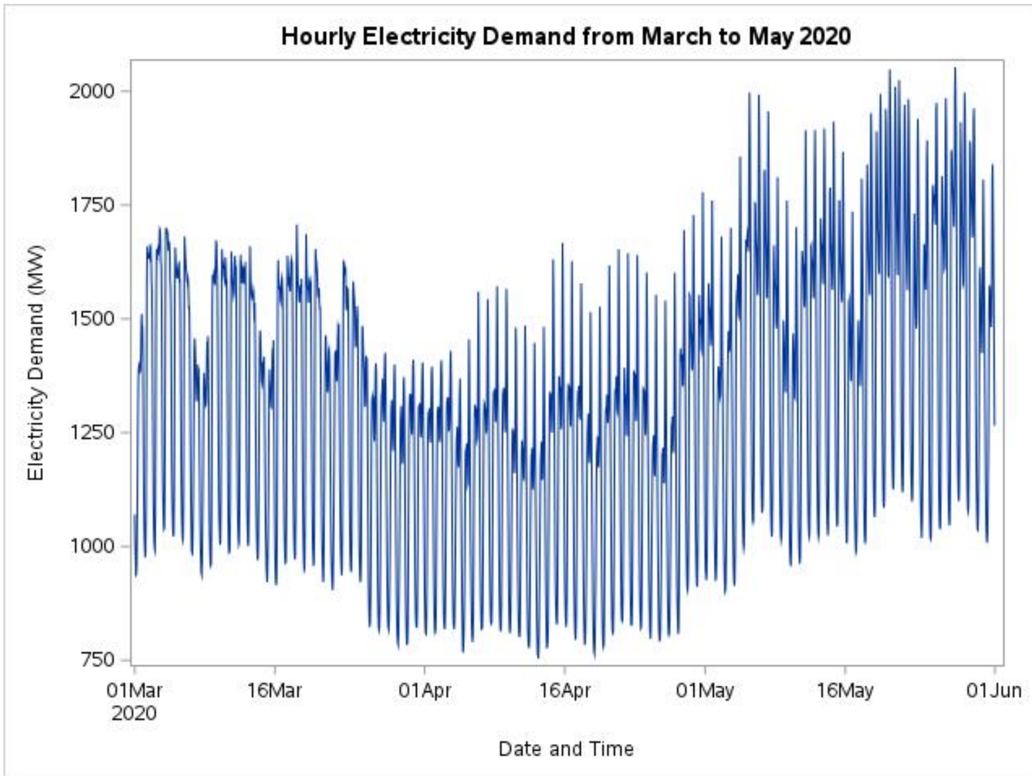
1.(b) An Appropriate Plot and Discussion of the Main Features of the Hourly Electricity Demand

<u>SAS codes and outputs are as follows:</u>

```
/* Step 1: Clean the data */
data elec_data_clean;
   set elec_data;

   /* Convert string to datetime, strip quotes */
   DateTime = input(compress(Date, '"'),
anydtdtm.);
   format DateTime datetime20.;

   /* Extract DateOnly and Hour */
   DateOnly = datepart(DateTime);
   format DateOnly date9.;
   Hour = hour(DateTime);
run;

/* Step 2: Ensure non-missing MW values */
proc means data=elec_data_clean n nmiss;
   var MW DateTime;
run;

/* Step 3: Sort before plotting */
proc sort data=elec_data_clean;
   by DateTime;
run;

/* Step 4: Plot */
proc sgplot data=elec_data_clean;
   series x=DateTime y=MW /
lineattrs=(thickness=1);
   xaxis label="Date and Time";
   yaxis label="Electricity Demand (MW)";
   title "Hourly Electricity Demand from March to
May 2020";
run;
```

Main Features Observed in the Electricity Demand Time Series:

1. **Strong Daily Seasonality**: The plot exhibits consistent up-and-down cycles within each 24-hour period, typical of human activity patterns (e.g., lower demand at night, peaks in the morning and evening).
2. **Mild Weekly Patterns**: There may be weaker recurring trends every 7 days, hinting at possible weekly seasonality influenced by workweek vs. weekend usage.
3. **Non-Stationary Behavior**: The overall level of demand appears to fluctuate slightly across months, suggesting a possible trend component or dependence on external factors (e.g., temperature).



2. Time Series Regression and Exponential Smoothing

2.(a) Time Series Regression

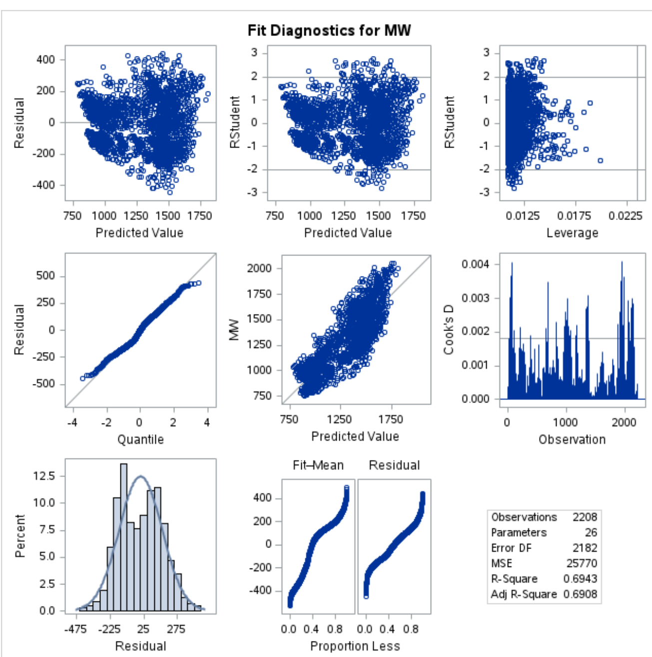
A multiple linear regression model was fitted to hourly electricity demand (MW) using temperature, wind speed, and time of day (Hour) as predictors. The Hour variable was treated as a categorical variable (via class Hour), allowing the model to capture cyclical daily effects using 23 hour-specific dummy variables, as represented in the model equation:

$$MW_t = \beta_0 + \beta_1 \,.Temperature_t + \beta_2 \,.Windspeed_t + \sum_{h=1}^{23} \beta_h \,.Hour_t + \varepsilon_t$$

SAS codes and outputs are as follows:

```
/* Step 1: Fit multiple linear regression model */
proc glm data=elec_data_clean plots=diagnostics;
   class Hour;
   model MW = Temperature WindSpeed Hour / solution;
   output out=reg_out p=Predicted r=Residual;
   title "Time Series Regression Model for Hourly Electricity Demand";
run;

/* Step 2: Generate 7-day (168-hour) forecast input manually */
data forecast_input;
   format DateTime datetime20. DateOnly date9.;
   do i = 1 to 168;
      DateTime = '01JUN2020:00:00:00'dt + (i - 1)*3600; /* 1-hour step */
      Hour = hour(DateTime);
      DateOnly = datepart(DateTime);
```

```
        Temperature = 15; /* Replace with assumed or average values */
        WindSpeed = 5;    /* Replace with assumed or average values */
        output;
    end;
run;

/* Step 3: Store model from PROC GLM */
proc glm data=elec_data_clean;
    class Hour;
    model MW = Temperature WindSpeed Hour;
    store out=reg_model;
run;

/* Step 4: Apply stored model to forecast_input */
proc plm restore=reg_model;
    score data=forecast_input out=reg_forecast predicted=Forecast_MW;
run;

/* Step 5: Plot the forecast */
proc sgplot data=reg_forecast;
    series x=DateTime y=Forecast_MW / lineattrs=(thickness=2);
    xaxis label="DateTime";
    yaxis label="Forecasted Electricity Demand (MW)";
    title "7-Day Forecast Using Time Series Regression Model";
run;
```



The GLM Procedure

Dependent Variable: MW MW

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	25	127738544.7	5109541.8	198.27	<.0001
Error	2182	56230949.4	25770.4		
Corrected Total	2207	183969494.1			

R-Square	Coeff Var	Root MSE	MW Mean
0.694346	12.24062	160.5315	1311.465

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Temperature	1	3837342.5	3837342.5	148.91	<.0001
WindSpeed	1	1488666.7	1488666.7	57.77	<.0001
Hour	23	122412535.5	5322284.2	206.53	<.0001

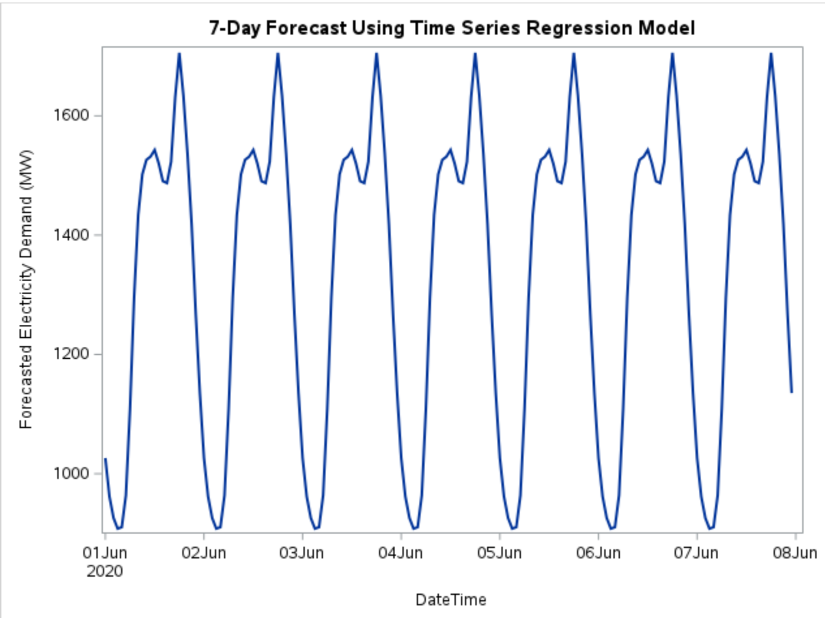
Source	DF	Type III SS	Mean Square	F Value	Pr > F
Temperature	1	6545674.5	6545674.5	254.00	<.0001
WindSpeed	1	5997.9	5997.9	0.23	0.6295
Hour	23	122412535.5	5322284.2	206.53	<.0001

Model Fit and Residual Analysis:

- **R-squared = 0.6943, Adjusted R-squared = 0.6908**, indicating that approximately 69% of the variation in electricity demand is explained by the model.
- **Temperature and Hour** were significant predictors **(p < 0.0001)**, while **WindSpeed** was not statistically significant **(p = 0.6295)** based on the Type III SS.
- The F-statistic for the full model was 198.27 (**p < 0.0001**), confirming overall model significance.

Residual Diagnostics

- Residual vs. predicted plots indicate some non-constant variance, with slightly fan-shaped patterns.
- The Q-Q plot suggests that residuals are approximately normally distributed, though with minor deviations at the tails.
- Leverage and Cook's D plots show a few influential points but nothing extreme.
- Histogram and cumulative distribution plots indicate reasonable residual spread.



7-Day Forecast: The 7-day forecast was generated using assumed average temperature (15°C) and wind speed (5 knots). The forecast displays clear and realistic daily cycles, consistent with the historical hourly demand structure captured by the dummy-coded hour terms.

In conclusion, the regression model accurately captures intraday seasonality via categorical Hour, demonstrates strong explanatory power (Adj. $R^2 \approx 0.69$), and provides interpretable, stable short-term forecasts.

2.(b) Exponential Smoothing

An additive Holt-Winters exponential smoothing model was applied to the hourly electricity demand data from March to May 2020. The model was trained using PROC ESM with a 24-hour seasonal cycle and used to forecast the next 168 hours (7 days) of demand.

SAS codes and outputs are as follows:

```
/* Step 1: Sort and Prepare Data */
proc sort data=elec_data_clean;
   by DateTime;
run;

proc timeseries data=elec_data_clean out=ts_hourly;
   id DateTime interval=hour;
   var MW;
run;

/* Step 2: Fit Additive Seasonal Exponential Smoothing Model */
proc esm data=ts_hourly
```

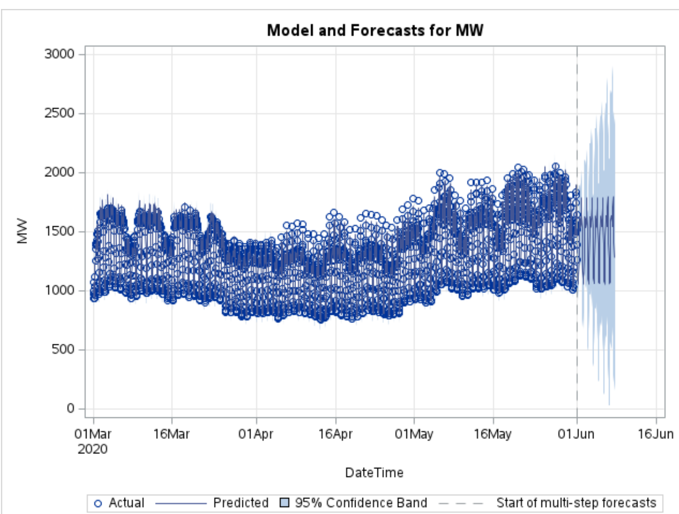
```
        out=esm_hw_add
        outfor=esm_forecast
        outest=esm_betas
        lead=168
        print=all
        plot=(forecasts modelforecasts);
    id DateTime interval=hour;
    forecast MW / model=addwinters;
run;

/* Step 3: Plot Forecasted Values */
title1 "Additive Holt-Winters Forecast – Electricity Demand";
proc sgplot data=esm_forecast;
    series x=DateTime y=Predict / lineattrs=(color=blue);
    series x=DateTime y=Actual / lineattrs=(color=red);
    refline '01JUN2020:00:00:00'dt / axis=x;
    yaxis label="MW";
    xaxis label="DateTime";
run;
```

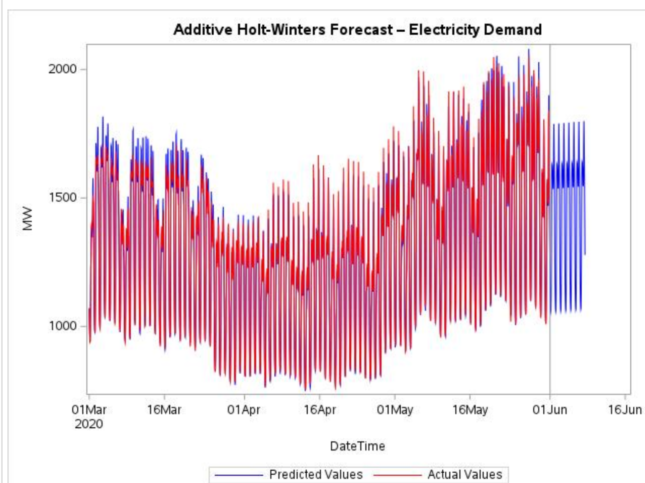
Winters Method (Additive) Parameter Estimates

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Level Weight	0.99900	0.01531	65.27	<.0001
Trend Weight	0.0010000	0.0039286	0.25	0.7991
Seasonal Weight	0.99900	15.20472	0.07	0.9476

Forecast Summation

Variable	Forecast	Standard Error	Confidence Limits	
MW	243302	54548	136391	350214





Model Fit:

- **Level weight**: 0.999 (highly responsive to recent level changes)
- **Trend weight**: 0.001 (near-zero, suggesting no strong long-term trend)
- **Seasonal weight**: 0.999 (strong emphasis on repeating seasonal components)

These values imply the model gives high priority to **recent data and seasonal patterns**, while smoothing out long-term trends.

Forecast Performance:

- The left figure ("Model and Forecasts for MW") shows predicted values closely following the actual observations during the training period, and forecasted values exhibit realistic daily oscillations with expanding confidence bands over time.
- The right figure ("Additive Holt-Winters Forecast – Electricity Demand") visualizes both actual (red) and forecasted (blue) values, clearly showing repeating daily cycles in the forecast period.
- The total 7-day forecasted demand (summed across all hours) is approximately **243,302 MW**, with a **standard error of 54,548**. The 95% confidence interval ranges from **136,391 to 350,214 MW**, reflecting uncertainty over a longer horizon.

In conclusion, the Holt-Winters additive model provides a robust short-term forecast by effectively capturing the strong daily demand cycles in the data. It adapts to recent level shifts and does not rely on external variables. While the regression model in 2(a) offers interpretability and incorporates weather effects, the exponential smoothing model in 2(b) is better suited for **purely data-driven short-term forecasting**, especially when autocorrelation and seasonality are dominant.

3. ARIMA and ARIMAX

3.(a) ARIMA Model for Hourly Electricity Demand

SAS codes and outputs are as follows:

```
/* Step 1: Sort the data by time */
proc sort data=elec_data_clean;
   by DateTime;
run;

/* Step 2: Identification – Check Stationarity with ADF Test */
proc arima data=elec_data_clean;
   identify var=MW stationarity=(adf=1);
run; quit;
```

Interpretation: The hourly electricity demand series (MW) is **stationary**, based on strong rejection of the null hypothesis in the ADF test (p value shows < 0.0001). Therefore, no non-seasonal differencing is required, i.e., d = 0.

```
/* Step 3: Model Identification via ACF/PACF */
proc arima data=elec_data_clean; /* Identification – ACF/PACF */
   identify var=MW nlag=48;
run;
quit;

proc arima data=elec_data_clean; /* Estimation - Initial Attempt with  ARMA(1,1)  */
   identify var=MW;
   estimate p=1 q=1 method=ML;
run;
```

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	4934.41	6	<.0001	0.944	0.806	0.629	0.446	0.277	0.136
12	4969.88	12	<.0001	0.032	-0.029	-0.054	-0.060	-0.061	-0.063
18	5031.88	18	<.0001	-0.064	-0.066	-0.063	-0.041	0.016	0.116
24	9999.99	24	<.0001	0.253	0.418	0.595	0.764	0.895	0.945
30	9999.99	30	<.0001	0.892	0.758	0.585	0.405	0.239	0.099
36	9999.99	36	<.0001	-0.003	-0.063	-0.088	-0.094	-0.096	-0.098
42	9999.99	42	<.0001	-0.099	-0.100	-0.097	-0.076	-0.021	0.077
48	9999.99	48	<.0001	0.211	0.372	0.545	0.709	0.835	0.884

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	819.40	4	<.0001	0.427	0.330	-0.027	-0.020	-0.221	-0.171
12	1221.13	10	<.0001	-0.338	-0.194	-0.161	0.040	-0.035	0.027
18	1687.74	16	<.0001	-0.036	0.039	-0.158	-0.197	-0.332	-0.181
24	4222.61	22	<.0001	-0.202	-0.030	0.007	0.298	0.453	0.894
30	5060.32	28	<.0001	0.452	0.304	0.002	-0.031	-0.211	-0.179
36	5439.91	34	<.0001	-0.324	-0.193	-0.155	0.034	-0.035	0.016
42	5894.38	40	<.0001	-0.031	0.034	-0.148	-0.198	-0.321	-0.189
48	8180.77	46	<.0001	-0.195	-0.033	0.022	0.289	0.443	0.833

Interpretation: The autocorrelation plots show strong dependencies at 1st lag 6 (0.944), tapering but persisting beyond lag 24. This indicates a **daily seasonal pattern**, with evidence of autocorrelation at lags 24 and 48, suggesting the presence of a 24-hour cycle in electricity demand.

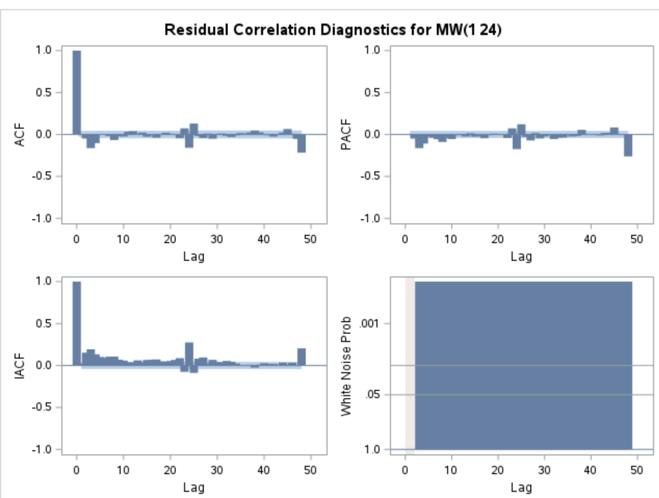
```
/* Step 4: Refine the Model with ARIMA(1,0,1)(1,0,1)[24] */
proc arima data=elec_data_clean;
   identify var=MW(1,24) nlag=48; /* Apply both non-seasonal and seasonal differencing */
   estimate p=1 q=1 P=1 Q=1 method=ML;
run;
quit;
```

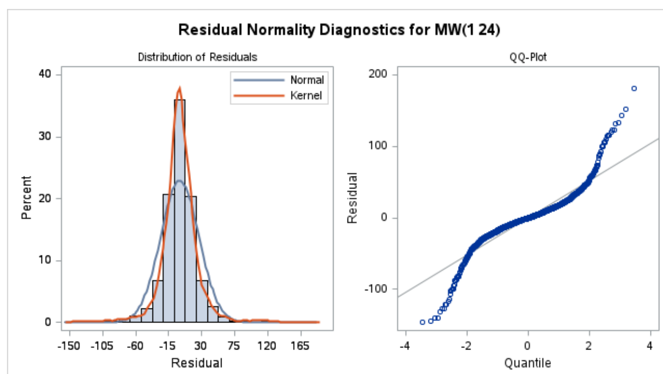
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.0035354	0.91825	-0.00	0.9969	0
MA1,1	-0.52807	0.03167	-16.67	<.0001	1
AR1,1	0.06947	0.03715	1.87	0.0615	1

Constant Estimate	-0.00329
Variance Estimate	682.9667
Std Error Estimate	26.13363
AIC	20445.72
SBC	20462.78
Number of Residuals	2183

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	625.21	6	<.0001	0.434	-0.092	-0.230	-0.180	-0.055	-0.015
12	663.89	12	<.0001	-0.049	-0.084	-0.059	-0.018	0.039	0.053
18	672.78	18	<.0001	0.026	0.010	-0.020	-0.035	-0.040	-0.004
24	686.19	24	<.0001	0.026	0.019	-0.011	-0.014	-0.016	-0.067
30	703.63	30	<.0001	0.051	0.019	-0.043	-0.032	-0.044	-0.011
36	710.79	36	<.0001	0.005	-0.025	-0.041	-0.016	0.014	0.022
42	734.78	42	<.0001	0.048	0.070	0.048	0.006	-0.022	-0.027
48	953.55	48	<.0001	0.006	0.054	0.077	0.009	-0.146	-0.260

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	86.32	4	<.0001	0.004	-0.049	-0.163	-0.103	-0.002	0.001
12	104.38	10	<.0001	-0.018	-0.066	-0.019	-0.024	0.033	0.039
18	110.31	16	<.0001	-0.002	0.022	-0.026	-0.007	-0.038	0.005
24	183.47	22	<.0001	0.021	0.009	0.003	-0.045	0.073	-0.159
30	231.77	28	<.0001	0.131	-0.017	-0.042	0.011	-0.050	0.006
36	235.51	34	<.0001	0.012	-0.016	-0.030	-0.010	0.017	0.002
42	244.64	40	<.0001	0.024	0.047	0.025	-0.003	-0.008	-0.025
48	366.28	46	<.0001	0.006	0.022	0.065	0.009	-0.052	-0.217



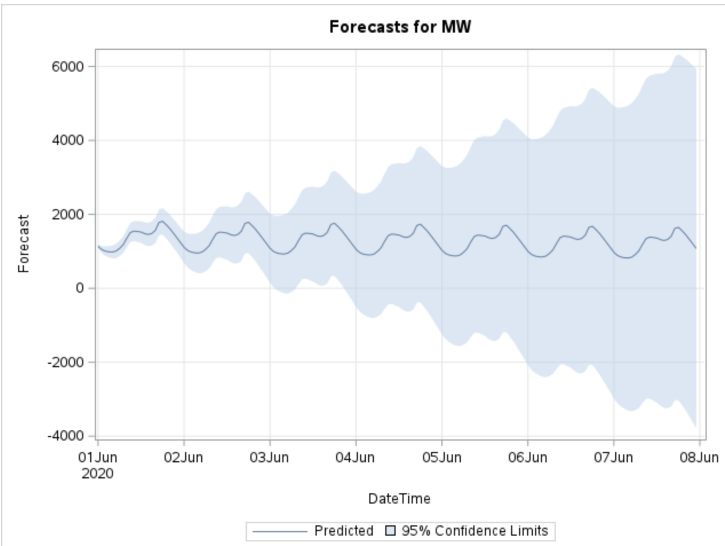


Interpretation: To address the strong seasonal autocorrelation, we apply seasonal differencing at lag 24, resulting in an ARIMA(1,0,1)(1,0,1)[24] model. This model effectively captures both:

- **Short-term fluctuations** via AR(1) and MA(1) components,
- **Daily seasonality** through seasonal AR(1) and MA(1) components.
- AIC = **20445.72** (useful for model comparison).
- MA(1) is **strongly significant** ($p < 0.0001$).
- AR(1) is **marginally insignificant** ($p = 0.0615$).

Residual diagnostics reveal that **autocorrelations have been substantially reduced**, with residuals approximating white noise, satisfying model adequacy conditions.

```
/* Step 5: Forecast 7 Days Ahead Using the Fitted Seasonal ARIMA Model */
proc arima data=elec_data_clean;
   identify var=MW(1,24);
   estimate p=1 q=1 P=1 Q=1 method=ML;
   forecast lead=168 interval=hour id=DateTime out=forecast_MW;
run; quit;
```

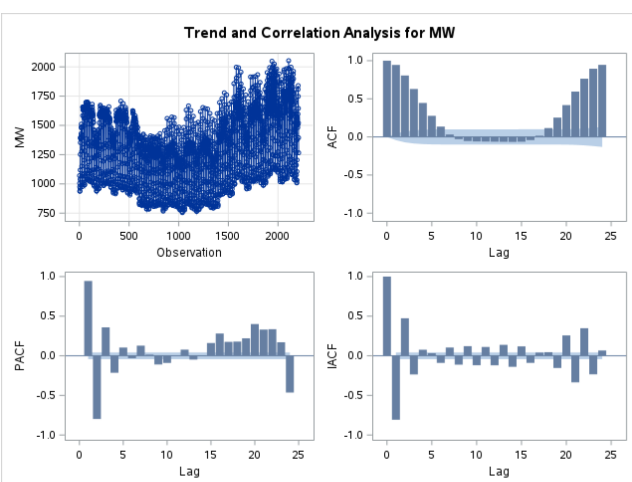


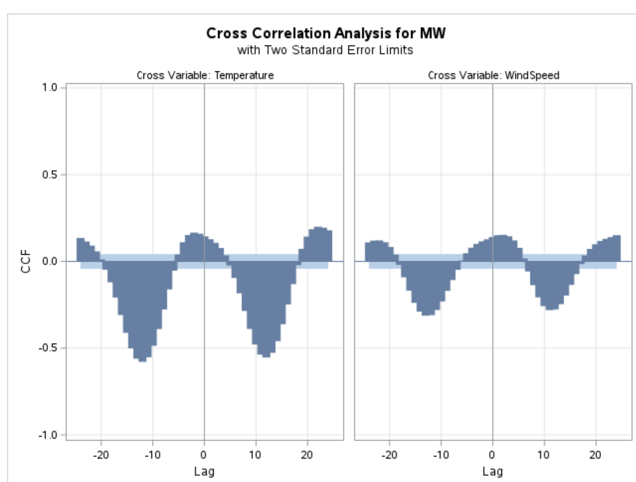
The 7-day forecast (see *Forecasts for MW* plot) shows clear **daily seasonality** with periodic peaks and troughs across each 24-hour cycle. The **95% confidence intervals widen** as the forecast horizon increases, appropriately reflecting forecast uncertainty over time.

3(b) ARIMA Model with Explanatory Variables (ARIMAX) – Temperature and Wind Speed

<u>SAS codes and outputs are as follows:</u>

```
/* Step 1: Check cross-correlation */
proc arima data=elec_data_clean;
   identify var=MW crosscorr=(Temperature WindSpeed);
run; quit;
```



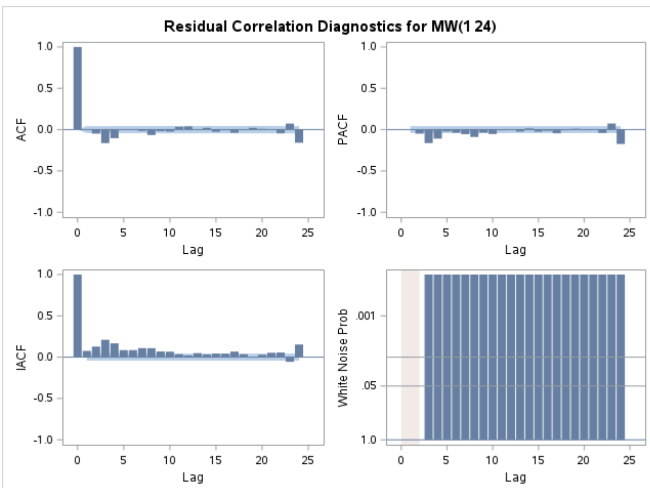


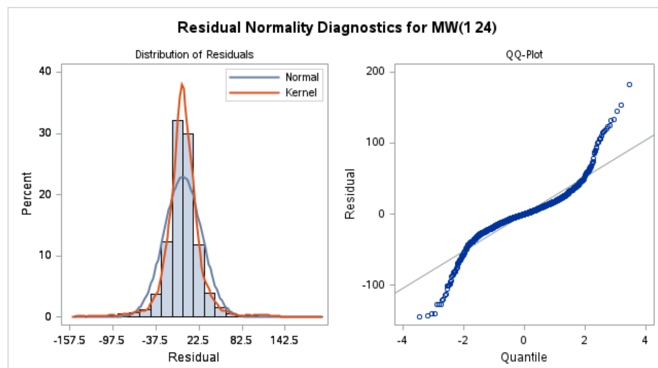
Interpretation: The cross-correlation plots indicate strong negative correlations between **MW** and both **Temperature** and **WindSpeed**, especially at lag 0, justifying their inclusion in the model.

```
/* Step 2: Differencing for stationarity */
proc arima data=elec_data_clean;
   identify var=MW(1,24) crosscorr=(Temperature WindSpeed);
run; quit;
```

Seasonal and non-seasonal differencing [i.e., **ARIMA(1,0,1)(1,0,1)[24]**] was applied to stabilize the variance and remove persistent autocorrelation.

```
/* Step 3: Estimate ARIMAX model with inputs */
proc arima data=elec_data_clean;
   identify var=MW(1,24) crosscorr=(Temperature WindSpeed);
   estimate p=1 q=1 input=(Temperature WindSpeed) method=ML;
   forecast lead=168 interval=hour id=DateTime out=forecast_arimax;
run; quit;
```





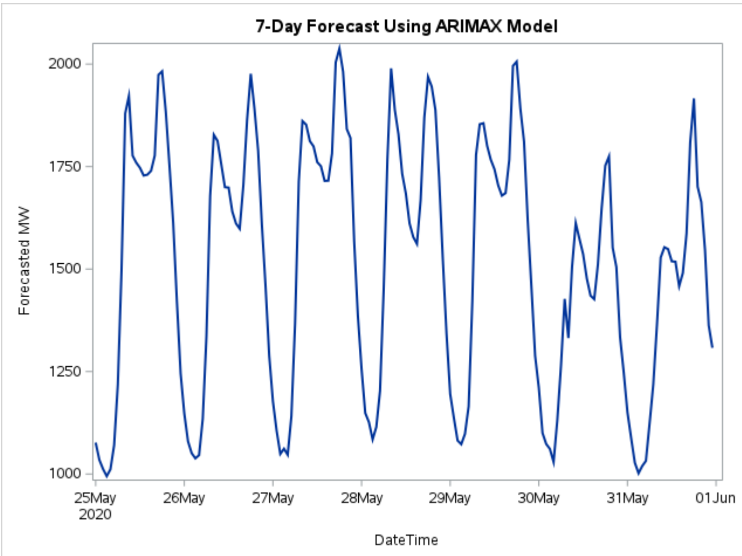
Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	-5.51373	4.16060	-1.33	0.1851	0	MW	0
MA1,1	-0.52863	0.03166	-16.70	<.0001	1	MW	0
AR1,1	0.06903	0.03716	1.86	0.0632	1	MW	0
NUM1	0.27662	0.27041	1.02	0.3063	0	Temperature	0
NUM2	0.13891	0.17005	0.82	0.4140	0	WindSpeed	0

Constant Estimate	-5.1331
Variance Estimate	682.7475
Std Error Estimate	26.12944
AIC	20447.02
SBC	20475.46
Number of Residuals	2183

Model Summary:

- **MA(1)** is highly significant ($p < 0.0001$).
- **AR(1)** is marginally insignificant ($p = 0.0632$).
- Although Temperature ($p = 0.3063$) and WindSpeed ($p = 0.4140$) are not statistically significant individually, the ARIMAX model performs similarly to the seasonal ARIMA model. However, because ARIMAX has a **slightly higher AIC (20447.02 vs. 20445.72)** and includes additional parameters, the simpler **ARIMA(1,0,1)(1,0,1)[24]** model is preferred for forecasting due to better efficiency.
- Residuals show low autocorrelation and approximate normality (see residual diagnostic plots), confirming model adequacy.

```
/* Step 4: Plot 7-Day Forecast from ARIMAX Model */
proc sgplot data=forecast_arimax;
   where DateTime >= '25MAY2020:00:00:00'dt and DateTime < '01JUN2020:00:00:00'dt;
   series x=DateTime y=forecast / lineattrs=(thickness=2);
   xaxis label="DateTime";
   yaxis label="Forecasted MW";
   title "7-Day Forecast Using ARIMAX Model";
run;
```



The 7-day ARIMAX forecast shows clear daily cycles in electricity demand, capturing recurring peaks and troughs influenced by temperature and wind speed.

4. Out-of-Sample Forecasts

4(a) The Last Seven Days as The Test Set and the rest as The Training Set

We do the test set by holding the last seven days and use the rest as the training set. SAS codes are as follows:

```
/* Create training and test datasets */
data train_set test_set;
   set elec_data_clean;
   if DateTime < '25MAY2020:00:00:00'dt then output train_set;
   else if DateTime < '01JUN2020:00:00:00'dt then output test_set;
run;
```

4(b) Evaluation of the Forecast Accuracy and Choosing the Best Forecasting Model/ Method.

ARIMA Model - SAS codes and outputs are as follows:

```
/* ARIMA */
data forecast_arima_7d;
   set forecast_MW;
   if '25MAY2020:00:00:00'dt <= DateTime < '01JUN2020:00:00:00'dt;
run;

data compare_arima;
   merge test_set(in=a) forecast_arima_7d(in=b);
```

```
   by DateTime;
   if a and b;
run;
%forecast_accuracy(modelname=ARIMA, compare_ds=compare_arima);
```

ARIMAX Model - SAS codes and outputs are as follows:

```
/* ARIMAX */
data forecast_arimax_7d;
   set forecast_arimax;
   if '25MAY2020:00:00:00'dt <= DateTime < '01JUN2020:00:00:00'dt;
run;

data compare_arimax;
   merge test_set(in=a) forecast_arimax_7d(in=b);
   by DateTime;
   if a and b;
run;
%forecast_accuracy(modelname=ARIMAX, compare_ds=compare_arimax);
```

Time Series Regression Model - SAS codes and outputs are as follows:

```
/* Time Series Regression */
data forecast_input_holdout;
   format DateTime datetime20. DateOnly date9.;
   do i = 1 to 168;
      DateTime = '25MAY2020:00:00:00'dt + (i - 1)*3600;
      Hour = hour(DateTime); DateOnly = datepart(DateTime);
      Temperature = 15; WindSpeed = 5; output;
   end;
run;

proc plm restore=reg_model;
   score data=forecast_input_holdout out=reg_forecast_7d predicted=forecast;
run;

data compare_reg;
   merge test_set(in=a) reg_forecast_7d(in=b);
   by DateTime;
   if a and b;
run;
%forecast_accuracy(modelname=Regression, compare_ds=compare_reg);
```

Exponential Smoothing Method - SAS codes and outputs are as follows:

```
/* Exponential Smoothing */
data esm_forecast_7d;
   set esm_forecast;
   if '25MAY2020:00:00:00'dt <= DateTime < '01JUN2020:00:00:00'dt;
   rename Predict=forecast;
run;

data compare_esm;
   merge test_set(in=a) esm_forecast_7d(in=b);
```

```
   by DateTime;
   if a and b;
run;
%forecast_accuracy(modelname=ExpSmoothing, compare_ds=compare_esm);
```

Summary Table for Methods Comparison - SAS codes and outputs are as follows

```
data metrics_arima;
   length Method $20;
   Method = "ARIMA";  /* Create first to force it as the first column */
   set compare_arima end=last;
   retain mae mape mse;
   mae + abs(err);
   mape + abs_pct_err;
   mse + err_sq;
   if last then do;
      mae = mae / _N_;
      mape = mape / _N_;
      mse = mse / _N_;
      rmse = sqrt(mse);
      output;
   end;
   keep Method mae mape mse rmse;
run;
```

```
data all_metrics;
   set metrics_ARIMA metrics_ARIMAX
metrics_Regression metrics_ExpSmoothing;
run;

proc print data=all_metrics noobs label;
   title "Comparison of Forecasting
Models'/Methods' Accuracy ";
   label
      Method = "Forecasting Method"
      MAE = "Mean Absolute Error (MAE)"
      MAPE = "Mean Absolute Percentage Error
(MAPE)"
      MSE = "Mean Squared Error (MSE)"
      RMSE = "Root Mean Squared Error
(RMSE)";
run;
```

The forecast accuracy of four models—ARIMA, ARIMAX, Regression, and Exponential Smoothing—was assessed using the 7-day holdout dataset of test period (25 May – 31 May 2020). Accuracy metrics include Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Results are summarized below:

Comparison of Forecasting Models'/Methods' Accuracy

Forecasting Method	Mean Absolute Error (MAE)	Mean Absolute Percentage Error (MAPE)	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)
ARIMA	21.300	1.4118	1143.22	33.812
ARIMAX	21.229	1.4063	1136.58	33.713
Regression	189.737	12.2627	47534.91	218.025
ExpSmoothing	33.779	2.1783	2214.67	47.060

Table: Summary Table for Forecasting Methods Comparison

ARIMAX outperforms the others across all metrics, achieving the lowest MAE, MAPE, MSE, and RMSE, indicating the best forecast accuracy. ARIMA performs similarly well, while Exponential Smoothing shows moderate accuracy. In contrast, the Regression model performs poorly due to its inability to fully capture the seasonal dynamics in the electricity demand data.

To conclude, **ARIMAX model** is the most accurate and robust forecasting method among those tested. It effectively captures both seasonal patterns and the influence of exogenous variables (temperature and wind speed), making it the preferred model for forecasting hourly electricity demand.