STAT-S 650 TIME SERIES ANALYSIS

ELECTRIC TRANSFORMER TEMPERATURE FORECASTING

Final Project Report

Spring 2023

Authored by:

Nihaar Nandedkar

1. INTRODUCTION

Electricity is a fundamental requirement for modern society, and its distribution is a critical process that impacts every aspect of life. However, ensuring a reliable and stable supply of electricity is a challenging task for power plant managers. The issue of efficient power distribution is exacerbated by the constant variability of power demand across different regions and seasons. Moreover, it is further complicated by the unpredictable and fluctuating nature of factors like weather, temperature, and other external load factors, which make it difficult to accurately predict the amount of power that will be required in a given period. Even with the advanced analytical methods available today no existing method can perform a long-term prediction for the forecast based on super long-term real-world data with high precision.

The traditional methods of forecasting power demand based on empirical data have proven to be insufficient and often lead to overproduction, leading to wastage of electricity and other resources. The overproduction also results in a financial burden for the electricity provider, leading to higher costs for consumers. The maintenance of power transmission equipment, especially transformers, is also impacted by the production of excess electricity.

One way to tackle this problem is to focus on forecasting the oil temperature of electrical transformers, which is a reliable indicator of their condition and efficiency. Maintaining optimal oil temperature can help prevent the degradation of the transformer and extend its life. However, forecasting oil temperature is also a challenging task that requires accurate and sophisticated models that consider a wide range of factors.

ARMAX models are a promising approach for forecasting oil temperature in electrical transformers. These models can account for various factors external power load metrics, holidays, weekends, temperature, and weather conditions. By analyzing historical data on transformer oil temperature, power demand, and external load factors during times of useful load (actual demand for power supply) and useless load (no demand for power supply), ARMAX models can identify patterns and trends that can help predict future oil temperature of the transformer with a high degree of accuracy. By building multiple ARMAX models with varying levels of complexity, power plant managers can obtain accurate forecasts of transformer oil temperature and maintain it within an efficient range. This will help reduce the wastage of electricity, prevent the degradation of transmission equipment, and ultimately lead to cost savings for both electricity providers and consumers.

2. DATASET DESCRIPTION

The dataset used in the implementation of this project is a real-world dataset collected over 2 years from a power generation facility in China. The data was collected for the purpose of addressing the problem of oil temperature forecasting as a result of a study that was conducted by a group of researchers and the Beijing Guowang Fuda Science & Technology Development Company. It is publicly available on github.com.

The dataset contains the following variables –

Field	date	HUFL	HULL	MUFL	MULL	LUFL	LULL	OT
Description	The recorded date	High UseFul Load	High UseLess Load	Middle UseFul Load	Middle UseLess Load	Low UseFul Load	Low UseLess Load	Oil Temperature (target)

Here high, middle(mean) and low correspond to the reading that was observed on the electrical transformer in **kilovolt-amperes**. Useful load refers to the load that the transformer takes when there's an actual demand for power supply while Useless load refers to the load that the transformer takes when there's no demand for powerful supply. Oil temperature is measured in **Celsius**.

For ease of modelling in this project, the dataset is aggregated from hourly to daily by taking the mean of all the variables over 24 hours.

3. HYPOTHESIS TO TEST

Through the implementation of this project, the basic hypothesis that I will be testing out are as follows -

1. The target variable oil temperature follows a seasonal trend over the years as well as a daily and weekly trend.

2. External power load factors which function as the predictor variables follow a daily and weekly trend.

3. The value of the dependent variable oil temperature at any given day is influenced by the oil temperature value from previous days.

4. The value of the dependent variable oil temperature at any given day is influenced by the values of the external power load metrics of the transformer from previous days.

4. METHODOLOGY & RESULTS

4.1 Importing Data and Necessary Libraries:

The step involves importing the relevant data from the CSV into R along with the required packages like tidyr, tseries, ggplot2, forecast, and lmtest to support the functionality for performing the analysis and implementing this project.

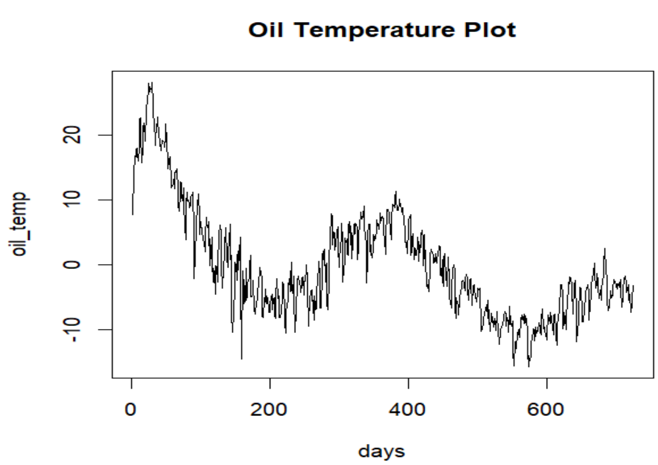
4.2 Data Wrangling:

Once the data is imported, it may require some preprocessing for it to be useful for the analysis. This involves removing missing values, aggregating the data from hourly to daily by taking the mean of the observations across all variables over 24 hours, and standardizing the dataset by subtracting the mean of all the variables from its series. Standardizing a time series helps to ensure that the data is prepared for analysis in a way that removes any irrelevant factors and makes it easier to identify the patterns and trends that are relevant to the analysis.

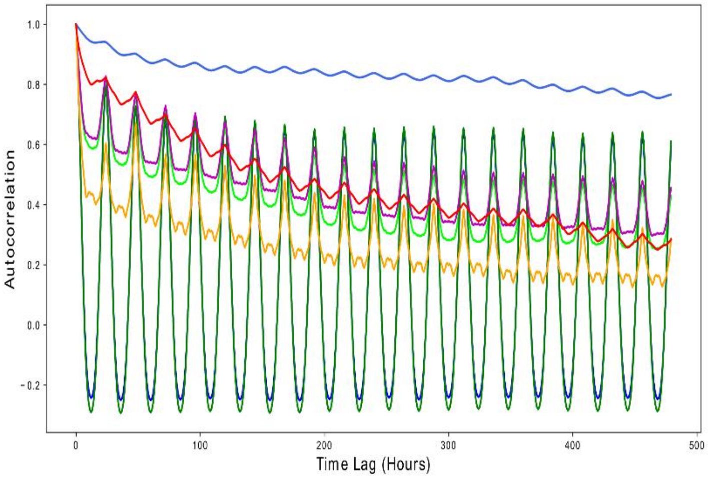
4.3 Exploratory Data Analysis:

This step involves analyzing the time series plots and autocorrelation plots of all the variables involved. Through these plots the hypothesis regarding the presence of any trends in the data can be tested out. The plots can also give an idea about the stationarity of the series.

Stationarity in a time series data refers to a property of the data where the statistical properties such as the mean, variance and covariance are constant over time. A stationary time series has no trend, and no seasonal patterns making it easier to model and forecast. On the other hand, non-stationary time series have trends, seasonality, and/or changing statistical properties over time, making it more difficult to model and forecast accurately.

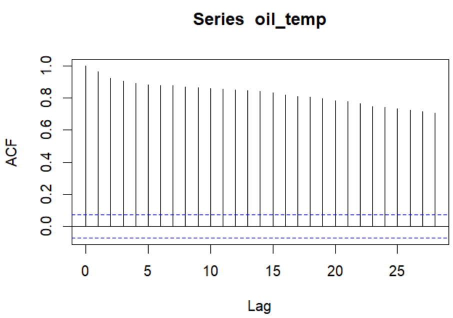


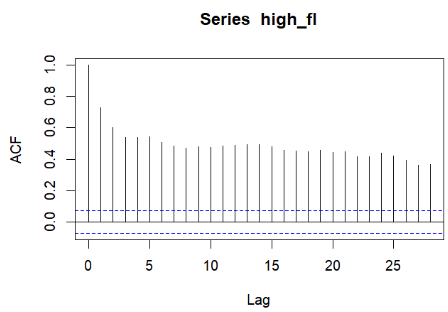
From the time series plot of the target variable oil temperature, peaks are observed roughly every 365 days. This indicates that oil temperature has a seasonal/yearly trend.

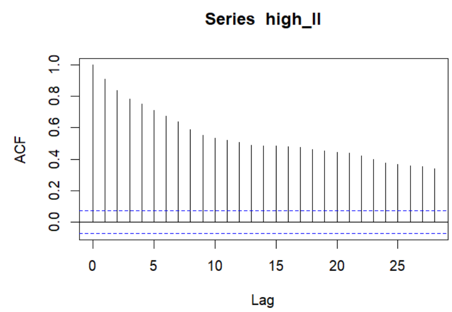


From the autocorrelation plot of all the predictor variables and the target variable (light blue line on top) before aggregation we can observe that all the predictor variables (external power load metrics) show a short-term daily pattern and a long-term weekly pattern.

The autocorrelation plots for each variable post aggregation also show the presence of trends.

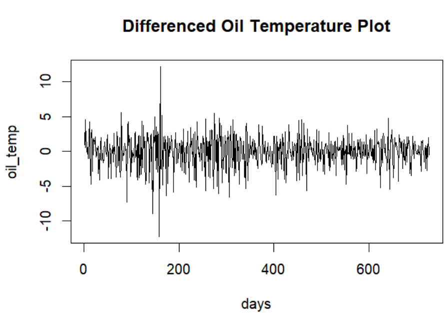


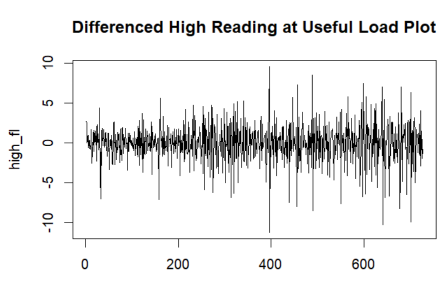


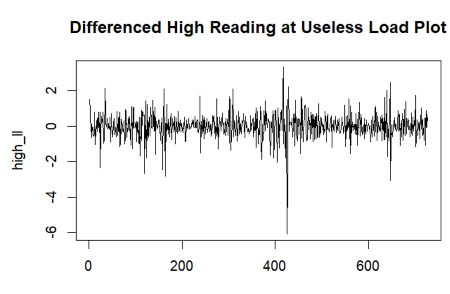


All the autocorrelation plots *(Refer Image 1-4)* show a slow decay of autocorrelation over multiple lags indicating the presence of a trend. This is because a trend will lead to a slow decay in autocorrelation as the values of the series are correlated with past values.

Presence of trends indicates that the time series are not stationary and thus require some form of transformation. Common techniques to transform a non-stationary time series include differencing, logarithmic transformation, and seasonal adjustment. Once a time series is stationary, it can be more easily modeled and forecasted using time series forecasting methods. Using differencing on the time series results in the following plots –





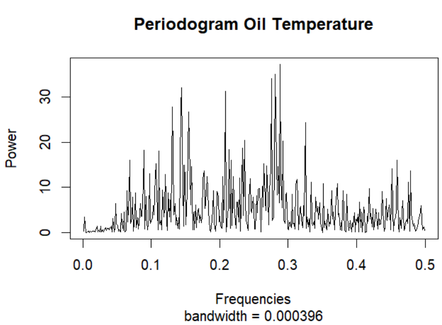


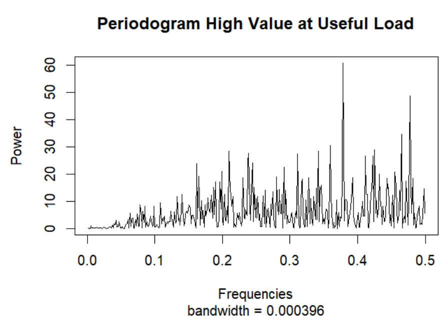
From all the time series plots *(Refer Image 5-8)* we can say that all the time series satisfy the conditions of constant mean, constant variance, and constant covariance hence, the series have become stationary post differencing.

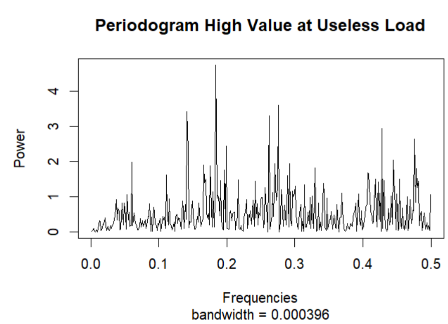
An ADF (Augmented Dickey-Fuller) test can also be used as a check for stationarity. The test produces a test statistic and a p-value. If the p-value is less than the significance level, then the null hypothesis is rejected and the time series is considered stationary.

VARIABLE	**P-Value**	**CONCLUSION**
Diff. Oil Temperature	< 0.01	Stationary
Diff. High Val at Useful Load	< 0.01	Stationary
Diff. High Val at Useless Load	< 0.01	Stationary
Diff. Mean Val at Useful Load	< 0.01	Stationary
Diff. Mean Val at Useless Load	< 0.01	Stationary
Diff. Low Val at Useful Load	< 0.01	Stationary
Diff. Low Val at Useless Load	< 0.01	Stationary

A periodogram of the spectral analysis of a time series is a tool used to analyze the frequency components of a time series. While it is not a direct test of stationarity, we can use the periodogram to identify whether a time series has any significant frequency components that could indicate non-stationarity. The spectral analysis of all the time series is as follows –





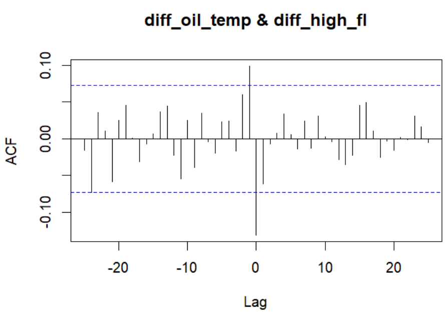


The peaks correspond to the patterns that occur most often in the time series. Since none of the series *(Refer Image 9-12)* show any significant spikes indicating a higher power at a particular point that could potentially indicate a trend, we can say that all the time series are stationary.

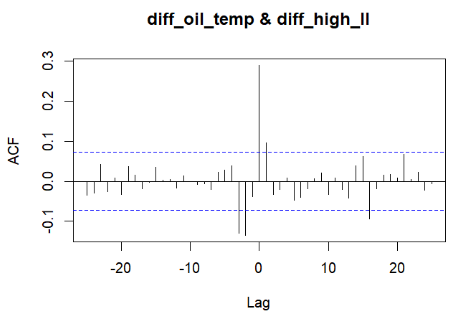
Using the observations from the time series plots and spectral analysis of the differenced time series as well as the p-values obtained from the ADF test, we can conclude that the differenced time series are stationary and can be used for modelling.

4.3 Interpreting CCF, ACF and PACF of the Differenced Series :

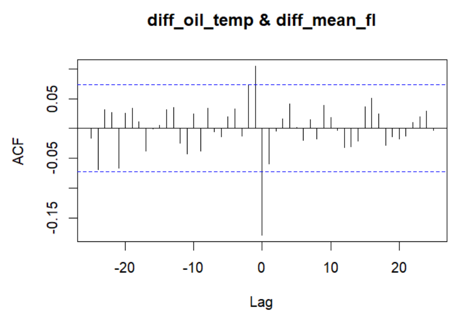
The cross-correlation function (CCF) of two time series measures the linear relationship between the two series at different lags. The CCF measures the correlation between the values of the first time series at time t and the values of the second time series at time t+k, where k is the lag between the two time series. The direction of the cross-correlation indicates the lead-lag relationship between the two time series.



high_fl -vely drives oil_temp at lag 0
oil_temp +vely drives high_fl at lag 1



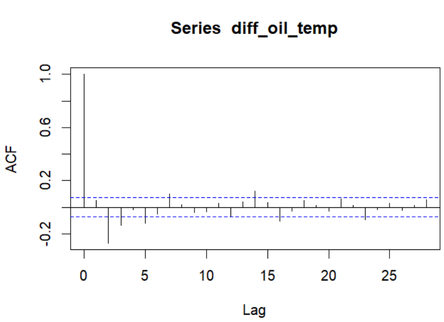
high_ll +vely drives oil_temp at lead 0,1
oil_temp -vely drives high_ll at lag 2,3



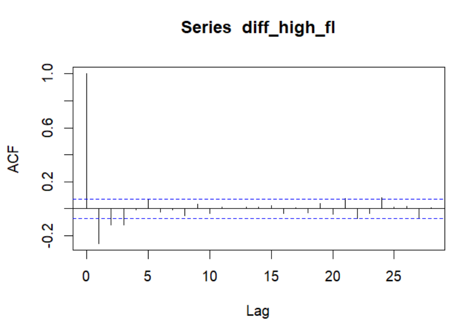
mean_fl -vely drives oil_temp at lag 0
oil_temp +vely drives mean_fl at lag 1

Overall, the CCF plots *(Refer Image 13-15)* are a useful tool to visualize the linear relationship between two time series at different lags. However, it should be used in conjunction with other statistical tools to make a conclusive decision about the nature of the relationship between the two time series.

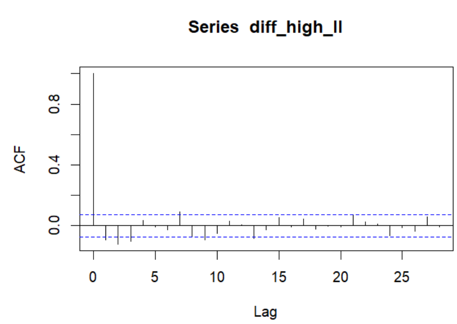
ACF (Autocorrelation Function) plots are a tool used to visualize the correlation between observations in a time series at different lags. An ACF plot displays the correlation coefficients for a range of lags. The autocorrelation at lag k is significant if it falls outside the dashed lines on the ACF plot. These lines represent the confidence intervals for the autocorrelation, and they are calculated based on the assumption that the time series is white noise. The ACF plots for the differenced series are as follows –



Significant spikes up to lag 3



Significant spikes up to lag 3

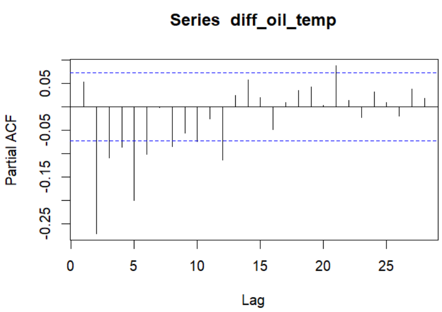


Significant spikes up to lag 2

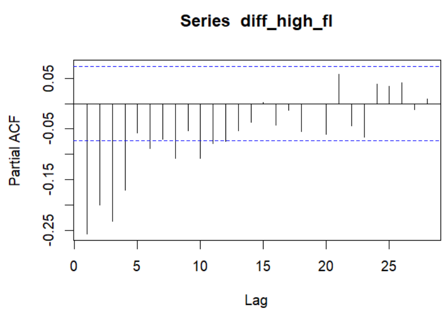
From the ACF plots *(Refer Image 16-19)* we can say that an MA model of order 1, 2 or 3 could be suitable to forecast the target variable oil temperature.

Overall, the ACF plot is a useful tool to visualize the correlation between observations in a time series and to identify non-random patterns in the data. However, it should be used in conjunction with other statistical tools to make a conclusive decision about the appropriate ARMA model order.

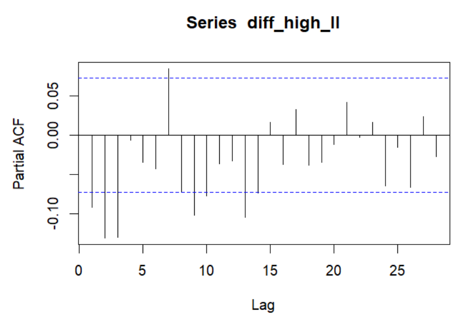
The Partial Autocorrelation Function (PACF) plot can provide additional information to help interpret the ACF plot. The PACF plot shows the correlation between two observations in a time series while controlling for the influence of all the other observations between them. Therefore, it can help identify the order of an ARMA model more precisely than the ACF plot alone.



Significant spike at lag 0 only



Significant spikes up to lag 4



Significant spikes up to lag 3

From the PACF plots *(Refer Image 20-23)* we can say that an AR model of order 1, 2, 3 or 4 could be suitable to forecast the target variable oil temperature.

4.4 Determining Best ARMAX Model :

It is important to keep in mind that there is no one-size-fits-all approach to selecting the order of an ARMAX model, and the appropriate order will depend on the specific data and context of the problem. Therefore, it is recommended to use multiple models and diagnostic tests in the selection of the final model.

Based on the observations from the ACF and PACF plots of the differenced series, lagged time series (from lags 1-4) of the target variable as well as the external power load factor variables are created. These lagged series will serve as exogenous variables for the ARMAX models. To account for the seasonal trend in the oil temperature we also include cyclic sin and cos components with a period of 365 as predictor variables.

The different ARMAX models tested are as follows –

- Simple White Noise Model
- Average Model 1 – Containing 1 lagged terms of the external power load factors and the oil temperature, and cyclic sin and cos components as exogenous variables.

- Complex Model 2 – Containing up to 2 lagged terms of the external power load factors and the oil temperature, and cyclic sin and cos components as exogenous variables.
- Complex Model 3 – Containing up to 3 lagged terms of the external power load factors and the oil temperature, and cyclic sin and cos components as exogenous variables.
- Complex Model 4 – Containing up to 4 lagged terms of the external power load factors and the oil temperature, and cyclic sin and cos components as exogenous variables.

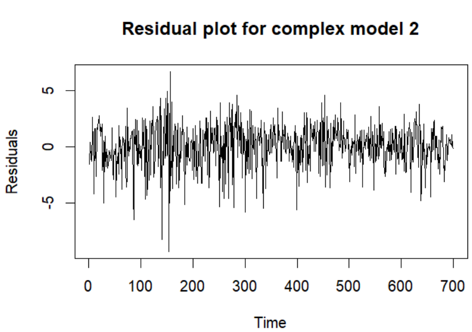
To determine the best order for each ARMAX model listed above, we loop through all possible values for p,q from 1-5 and choose the values for which the model gives the lowest AIC and BIC scores *(Refer Code Output 1-5)*. The best value for the order of each model along with their model selection criteria scores (AIC and BIC) are as follows –

MODEL	**ORDER**	**AIC**	**BIC**
White Noise	ARMA(5,4)	2987	3047
Average 1	ARMA(3,3)	2974	3051
Complex 2	ARMA(4,4)	2969	3087
Complex 3	ARMA(3,5)	2976	3126
Complex 4	ARMA(3,4)	2972	3150

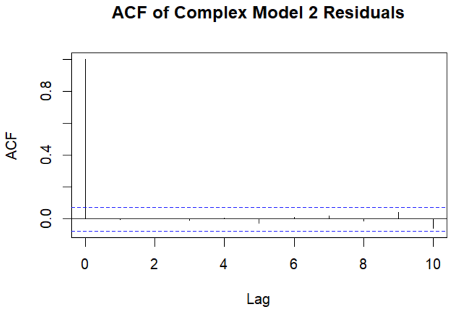
From the above table, based on the comparatively lower AIC and BIC scores, model 'Complex 2' with order ARMA(4,4) is determined to be the best model for forecasting oil temperature.

4.5 Evaluating Goodness of Fit for Complex Model 2:

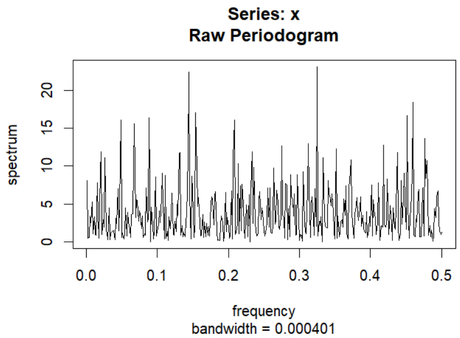
After fitting an ARMAX model, it's important to assess the model's goodness of fit to determine whether it adequately captures the patterns and variability in the data. One way to do this is to examine the residuals.



A random scatter of points around zero indicates that the model is a good fit for the data. The mean of the residuals should be close to zero, and the variance of the residuals should be constant over time.



The ACF of the residuals can help detect any remaining patterns in the data that are not captured by the model. Since there an no significant spikes in the ACF plot, it indicates that the model captured all the temporal dependence in the data.

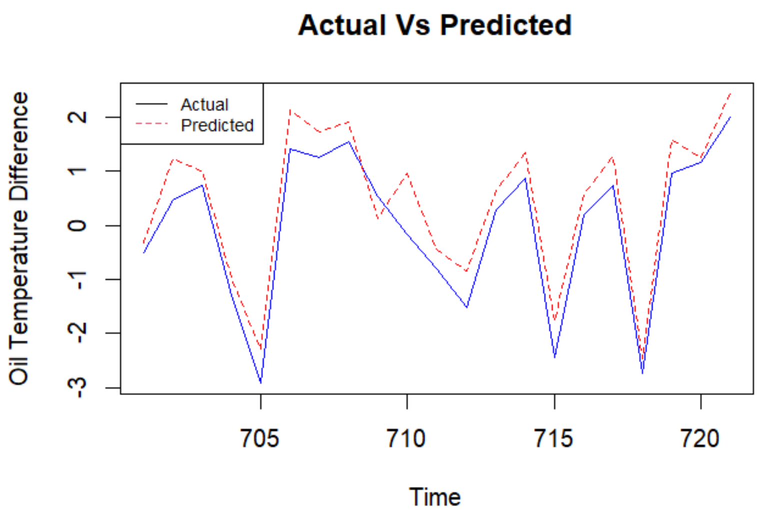


Since there are no significant peaks in the periodogram, it suggests that there are no patterns or trends in the residual time series. Hence, the residuals are a random walk.

There are several statistical tests that can be performed on the residuals to test the goodness of fit of the ARMAX model. One commonly used test is the Ljung-Box test, which tests the null hypothesis that the residuals are independently and identically distributed against the alternative that they are not. For the residuals of Complex Model 2 we obtain a p-value of 0.992 on performing the Box Test *(Refer Code Output 6)*. This suggests that the residuals are independent.

From the results of the time series plot, ACF plot, spectral analysis, and Ljung-Box test, we conclude that the model is a good fit and can be used for forecasting.

4.6 Forecasting and Model Metrics



The plot shows the models forecast of change in oil temperature for the next 20 days.

From the plot of the actual vs predicted values, we see the predicted values follow a similar trend to the actual values, and the magnitude of the predictions is also very close to that of the actual values indicating that the model does a good job at forecasting.

This observation is further supported by the low values of different model metrics like Mean Absolute Error **0.48** and Root Mean Square Error **0.53.** The accuracy of the ARMAX model on the test dataset is **~87.5%** *(Refer Code Output 7)*.

From the coefficients and standard errors of the model, the statistically significant components are found to be **AR2**, **AR3**, **MA1**, **MA2**, **MA3**, **MA4**, **intercept**, **oil_temp_lag1**, **high_fl_lag2**, **high_ll_lag2**, **mean_fl_lag2**, and cyclic component **sin** *(Refer Code Output 8)*. The statistically insignificant components are AR1, AR4, oil_temp_lag2, high_fl_lag1, high_ll_lag1, mean_fl_lag1, mean_ll_lag1, mean_ll_lag2, low_fl_lag1, low_fl_lag2, low_ll_lag1, low_ll_lag2 and cyclic component cos.

Of the statistically significant components of the model, oil_temp_lag1 and high_fl_lag2 have positive coefficients indicating that an increase in their values will increase the value of the target as well. On the other hand, high_ll_lag2, mean_fl_lag2 and cyclic sin component have negative coefficients indicating an increase in their values will decrease the value of the target *(Refer Code Output 8)*.

5. CONCLUSION

The following conclusions can be drawn from the results of the exploratory data analysis, model fitting and evaluation –

1. The observations from the time series plot of the target variable as well as the ACF plots of the other variables pre and post aggregation support the initial hypothesis that there is a daily/weekly/seasonal trend in the data.

2. The model metrics suggest that complex model 2, with an order of ARMA(4,4), that contains up to 2 lagged terms of each series as exogenous variables, provides a reasonable fit to the data and is a suitable model for forecasting the oil temperature that should be maintained in the electric transformer.

3. Based on the ARMAX model that was developed, it is evident that there is a significant correlation between the variables of lagged oil temperature, high value at useful load, high value at useless load, and mean value at useful load with the target variable.

4. Therefore, the initial hypothesis that the oil temperature of the electric transformer is influenced by the lagged oil temperature as well as the lagged values of other external power load factors associated with the transformer is supported by the ARMAX model.

6. REFERENCES

Links:

- https://github.com/zhouhaoyi/ETDataset/blob/main/ETT-small/ETTh1.csv
- https://arxiv.org/abs/2012.07436
- https://towardsdatascience.com/interpreting-acf-and-pacf-plots-for-time-series-forecasting-af0d6db4061c
- https://jbusemey.pages.iu.edu/time/time_series.htm
- https://cprosenjit.medium.com/4-more-mv-time-series-forecasting-we-should-know-auto-arima-sarimax-varmax-prophet-ca4a704c1848

Images:

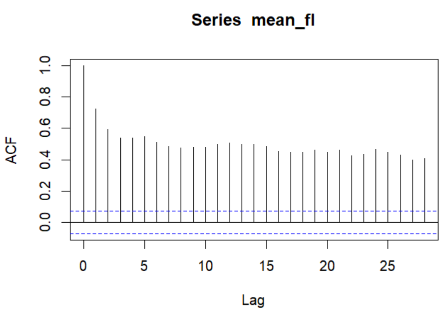


Image 1

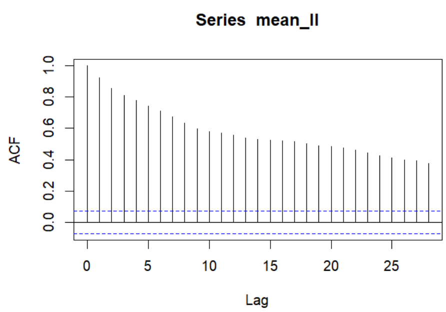


Image 2

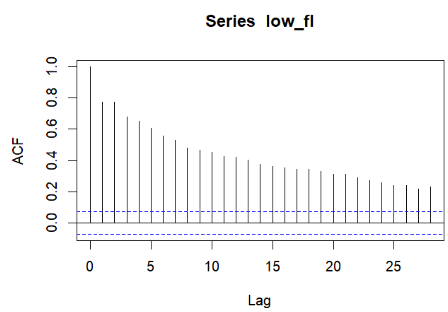


Image 3

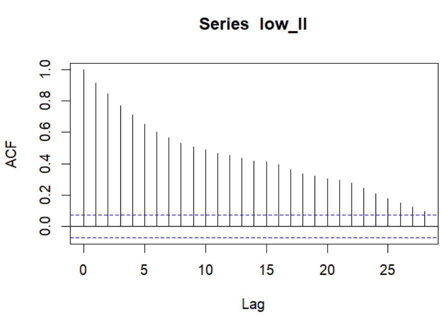


Image 4

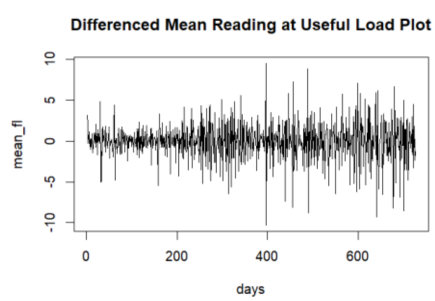


Image 5

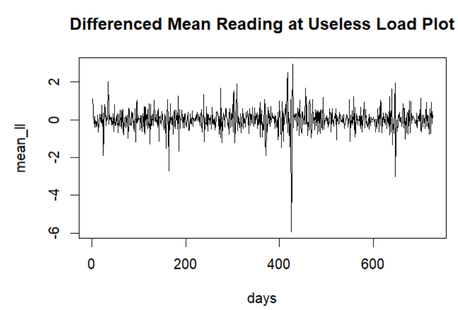


Image 6

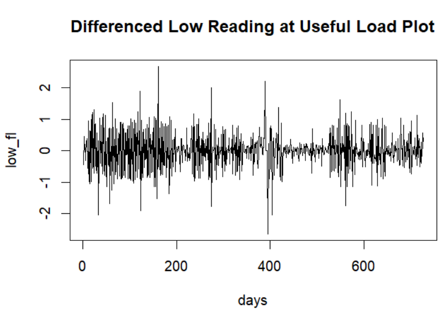


Image 7

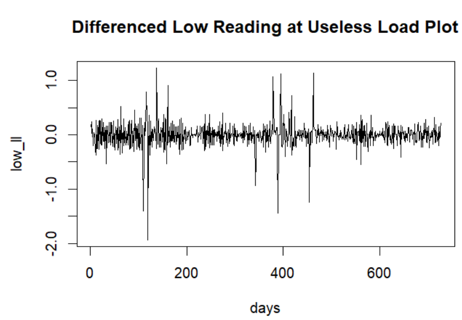


Image 8

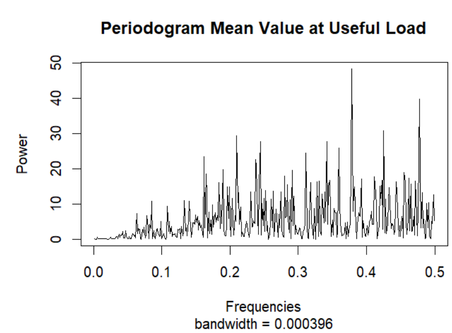


Image 9

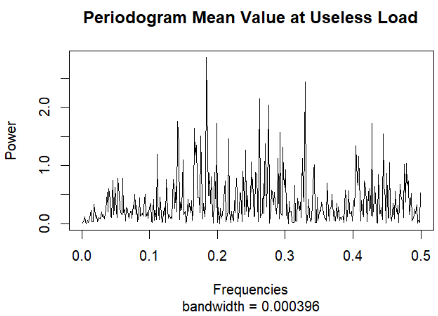


Image 10

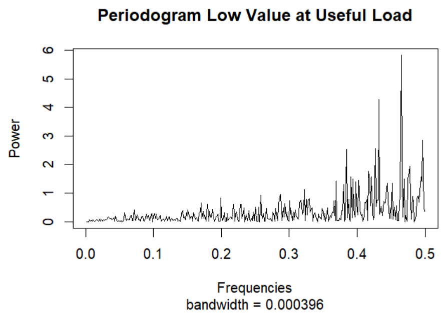


Image 11

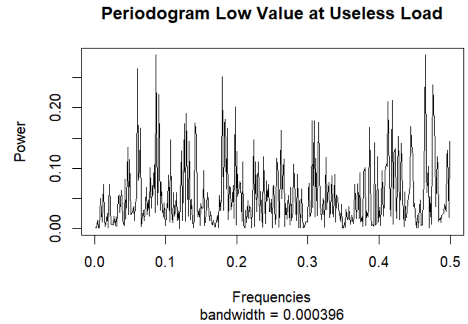


Image 12

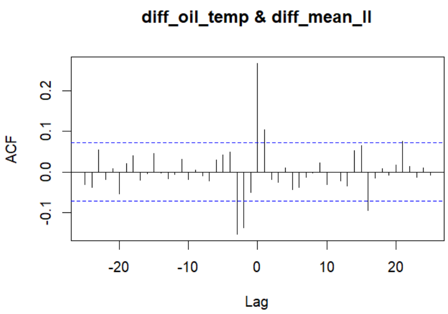


Image 13

mean_ll +vely drives oil_temp at lead 0,1
oil_temp -vely drives mean_ll at lag 1,2

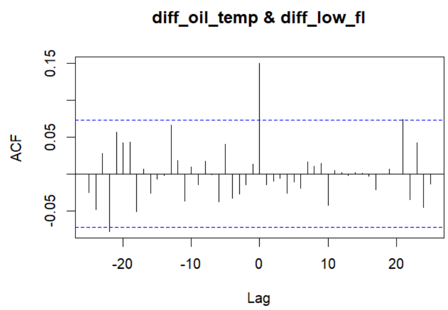


Image 14

low_fl +vely drives oil_temp at lag 0

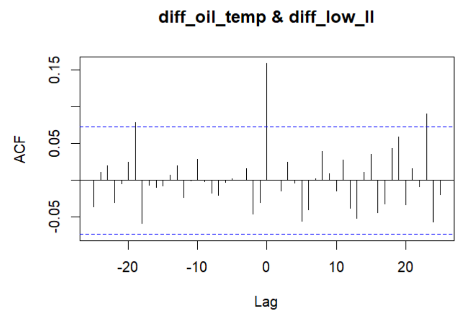


Image 15

low_ll +vely drives oil_temp at lag 0

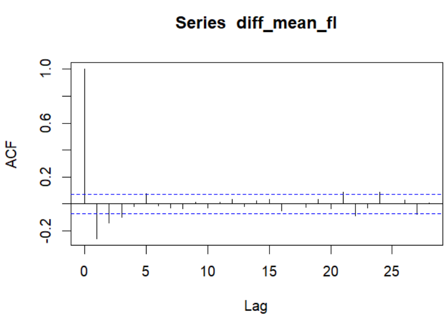


Image 16

Significant spikes up to lag 3

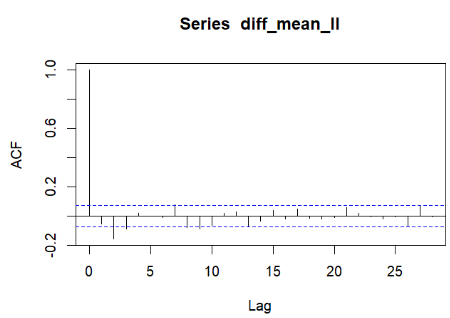


Image 17

Significant spikes up to lag 2

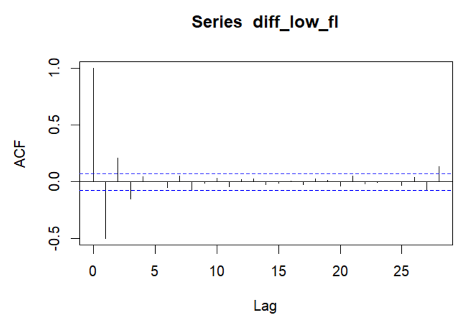


Image 18

Significant spikes up to lag 3

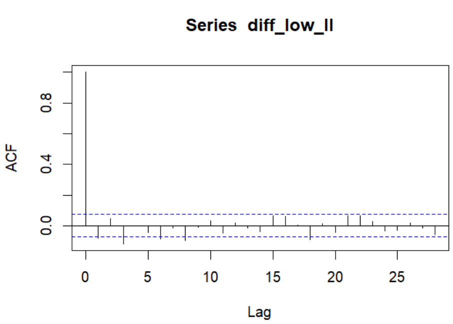


Image 19

Significant spike at lag 0 only

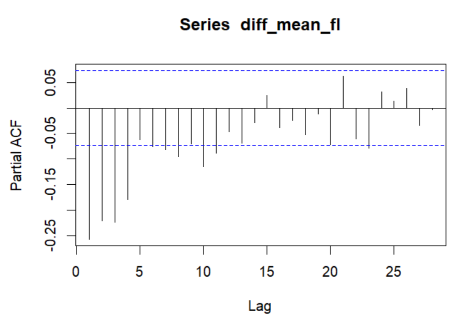


Image 20

Significant spikes up to lag 4

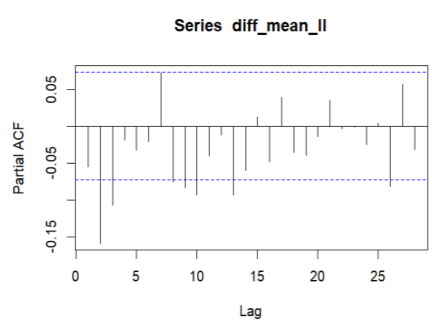


Image 21

Significant spike at lag 0 only

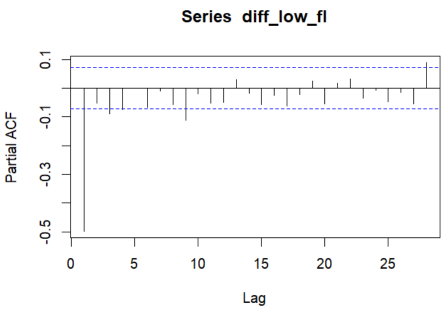


Image 22
Significant spikes up to lag 1

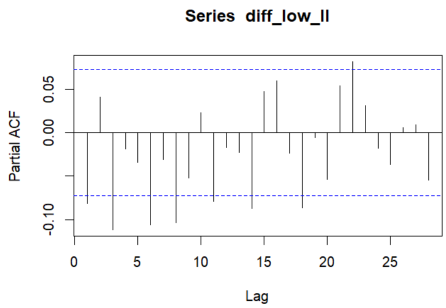


Image 23
Significant spikes up to lag 1

Code Outputs:

```
AIC: 2991.74 BIC: 3146.476 p: 1 q: 1
AIC: 2992.791 BIC: 3152.079 p: 1 q: 2
AIC: 2994.748 BIC: 3158.587 p: 1 q: 3
AIC: 2989.227 BIC: 3157.617 p: 1 q: 4
AIC: 2991.844 BIC: 3164.785 p: 1 q: 5
AIC: 2989.835 BIC: 3149.122 p: 2 q: 1
AIC: 2991.826 BIC: 3155.665 p: 2 q: 2
AIC: 2987.714 BIC: 3156.104 p: 2 q: 3
AIC: 2979.872 BIC: 3152.813 p: 2 q: 4
AIC: 2972.045 BIC: 3149.537 p: 2 q: 5
AIC: 2993.279 BIC: 3157.118 p: 3 q: 1
AIC: 2989.476 BIC: 3157.866 p: 3 q: 2
AIC: 2990.863 BIC: 3163.804 p: 3 q: 3
AIC: 2972.565 BIC: 3150.057 p: 3 q: 4
AIC: 2975.059 BIC: 3157.102 p: 3 q: 5
AIC: 2990.159 BIC: 3158.549 p: 4 q: 1
AIC: 2991.383 BIC: 3164.324 p: 4 q: 2
AIC: 2989.492 BIC: 3166.984 p: 4 q: 3
AIC: 2993.142 BIC: 3175.185 p: 4 q: 4
AIC: 2996.039 BIC: 3182.633 p: 4 q: 5
AIC: 2992.068 BIC: 3165.009 p: 5 q: 1
AIC: 2994.186 BIC: 3171.678 p: 5 q: 2
AIC: 2991.346 BIC: 3173.389 p: 5 q: 3
AIC: 2991.673 BIC: 3178.267 p: 5 q: 4
AIC: 2990.381 BIC: 3181.527 p: 5 q: 5
```

Output 1
Scores for Complex Model 4

```
AIC: 2995.001 BIC: 3117.88 p: 1 q: 1
AIC: 2986.696 BIC: 3114.126 p: 1 q: 2
AIC: 2994.086 BIC: 3126.067 p: 1 q: 3
AIC: 2996.086 BIC: 3132.619 p: 1 q: 4
AIC: 2979.123 BIC: 3120.206 p: 1 q: 5
AIC: 2987.2 BIC: 3114.631 p: 2 q: 1
AIC: 2988.527 BIC: 3120.508 p: 2 q: 2
AIC: 2996.069 BIC: 3132.602 p: 2 q: 3
AIC: 2995.65 BIC: 3136.733 p: 2 q: 4
AIC: 2987.516 BIC: 3133.151 p: 2 q: 5
AIC: 2988.216 BIC: 3120.198 p: 3 q: 1
AIC: 2989.105 BIC: 3125.638 p: 3 q: 2
AIC: 2989.843 BIC: 3130.926 p: 3 q: 3
AIC: 2988.763 BIC: 3134.398 p: 3 q: 4
AIC: 2976.792 BIC: 3126.977 p: 3 q: 5
AIC: 2988.959 BIC: 3125.491 p: 4 q: 1
AIC: 2987.665 BIC: 3128.748 p: 4 q: 2
AIC: 2988.679 BIC: 3134.314 p: 4 q: 3
AIC: 2991.01 BIC: 3141.196 p: 4 q: 4
AIC: 2986.795 BIC: 3141.532 p: 4 q: 5
AIC: 2983.788 BIC: 3124.871 p: 5 q: 1
AIC: 2984.279 BIC: 3129.914 p: 5 q: 2
AIC: 2987.917 BIC: 3138.102 p: 5 q: 3
AIC: 2982.571 BIC: 3137.308 p: 5 q: 4
AIC: 2986.559 BIC: 3145.847 p: 5 q: 5
```

Output 2
Scores for Complex Model 3

```
AIC: 2987.687 BIC: 3078.708 p: 1 q: 1
AIC: 2986.929 BIC: 3082.502 p: 1 q: 2
AIC: 2988.517 BIC: 3088.641 p: 1 q: 3
AIC: 2978.301 BIC: 3082.976 p: 1 q: 4
AIC: 2981.62 BIC: 3090.846 p: 1 q: 5
AIC: 2981.98 BIC: 3077.553 p: 2 q: 1
AIC: 2974.645 BIC: 3074.769 p: 2 q: 2
AIC: 2977.442 BIC: 3082.116 p: 2 q: 3
AIC: 2974.562 BIC: 3083.787 p: 2 q: 4
AIC: 2978.669 BIC: 3092.446 p: 2 q: 5
AIC: 2994.642 BIC: 3094.766 p: 3 q: 1
AIC: 2995.618 BIC: 3100.293 p: 3 q: 2
AIC: 2975.137 BIC: 3084.363 p: 3 q: 3
AIC: 2977.332 BIC: 3091.109 p: 3 q: 4
AIC: 2989.588 BIC: 3107.916 p: 3 q: 5
AIC: 2992.354 BIC: 3097.029 p: 4 q: 1
AIC: 2993.739 BIC: 3102.965 p: 4 q: 2
AIC: 2980.768 BIC: 3094.545 p: 4 q: 3
AIC: 2969.025 BIC: 3087.353 p: 4 q: 4
AIC: 2969.311 BIC: 3092.19 p: 4 q: 5
AIC: 2977.328 BIC: 3086.554 p: 5 q: 1
AIC: 2973.236 BIC: 3087.013 p: 5 q: 2
AIC: 2975.343 BIC: 3093.671 p: 5 q: 3
AIC: 2969.24 BIC: 3092.12 p: 5 q: 4
AIC: 2976.93 BIC: 3104.361 p: 5 q: 5
```

Output 3
Scores for Complex Model 2

```
AIC:  3096.137 BIC:  3155.301 p:  1 q:  1
AIC:  2981.232 BIC:  3044.947 p:  1 q:  2
AIC:  2983.954 BIC:  3052.22 p:  1 q:  3
AIC:  2979.3 BIC:  3052.118 p:  1 q:  4
AIC:  2982.578 BIC:  3059.947 p:  1 q:  5
AIC:  2980.91 BIC:  3044.625 p:  2 q:  1
AIC:  2982.621 BIC:  3050.888 p:  2 q:  2
AIC:  2984.005 BIC:  3056.822 p:  2 q:  3
AIC:  2986.778 BIC:  3064.146 p:  2 q:  4
AIC:  2981.112 BIC:  3063.032 p:  2 q:  5
AIC:  2982.438 BIC:  3050.704 p:  3 q:  1
AIC:  2984.859 BIC:  3057.676 p:  3 q:  2
AIC:  2974.173 BIC:  3051.541 p:  3 q:  3
AIC:  2974.23 BIC:  3056.15 p:  3 q:  4
AIC:  2975.774 BIC:  3062.244 p:  3 q:  5
AIC:  2984.998 BIC:  3057.815 p:  4 q:  1
AIC:  2985.219 BIC:  3062.587 p:  4 q:  2
AIC:  2974.003 BIC:  3055.923 p:  4 q:  3
AIC:  2975.761 BIC:  3062.232 p:  4 q:  4
AIC:  2977.725 BIC:  3068.747 p:  4 q:  5
AIC:  2977.388 BIC:  3054.756 p:  5 q:  1
AIC:  2978.842 BIC:  3060.762 p:  5 q:  2
AIC:  2975.668 BIC:  3062.138 p:  5 q:  3
AIC:  2977.665 BIC:  3068.687 p:  5 q:  4
AIC:  2979.657 BIC:  3075.23 p:  5 q:  5
```

Output 4
Scores for Complex Model 4

```
AIC:  3094.528 BIC:  3112.732 p:  1 q:  1
AIC:  2998.841 BIC:  3021.596 p:  1 q:  2
AIC:  2997.649 BIC:  3024.955 p:  1 q:  3
AIC:  2999.224 BIC:  3031.082 p:  1 q:  4
AIC:  2999.21 BIC:  3035.618 p:  1 q:  5
AIC:  2999.735 BIC:  3022.491 p:  2 q:  1
AIC:  2998.894 BIC:  3026.2 p:  2 q:  2
AIC:  2998.689 BIC:  3030.547 p:  2 q:  3
AIC:  3000.48 BIC:  3036.888 p:  2 q:  4
AIC:  2995.04 BIC:  3035.999 p:  2 q:  5
AIC:  2999.459 BIC:  3026.765 p:  3 q:  1
AIC:  3000.884 BIC:  3032.742 p:  3 q:  2
AIC:  2999.275 BIC:  3035.684 p:  3 q:  3
AIC:  2992.284 BIC:  3033.244 p:  3 q:  4
AIC:  2991.341 BIC:  3036.852 p:  3 q:  5
AIC:  2998.985 BIC:  3030.842 p:  4 q:  1
AIC:  3000.175 BIC:  3036.583 p:  4 q:  2
AIC:  2987.846 BIC:  3028.806 p:  4 q:  3
AIC:  2988.89 BIC:  3034.4 p:  4 q:  4
AIC:  2988.072 BIC:  3038.134 p:  4 q:  5
AIC:  2994.86 BIC:  3031.269 p:  5 q:  1
AIC:  2995.886 BIC:  3036.846 p:  5 q:  2
AIC:  2989.12 BIC:  3034.631 p:  5 q:  3
AIC:  2987.63 BIC:  3037.692 p:  5 q:  4
AIC:  2989.129 BIC:  3043.742 p:  5 q:  5
```

Output 5
Scores for Complex Model 3

```
> Box.test(complex_resids,lag=5)

        Box-Pierce test

data:  complex_resids
X-squared = 0.49686, df = 5, p-value = 0.9922
```

Output 6
Box Test on Complex Model 2 Residuals

```
> accuracy(pred_ts, test_ts)
               ME      RMSE       MAE      MPE     MAPE      ACF1 Theil's U
Test set -0.4446 0.5334875 0.4823197 4.441926 87.48906 -0.455204 0.5339796
```

Output 7
Model Metrics for Complex Model 2

```
> coeftest(complex_model_lag2)

z test of coefficients:

                Estimate Std. Error  z value  Pr(>|z|)    
ar1            0.4426728  0.2797402   1.5824  0.113549    
ar2           -0.8071982  0.1261052  -6.4010 1.544e-10 ***
ar3            0.6889954  0.1445764   4.7656 1.883e-06 ***
ar4           -0.2220024  0.1864965  -1.1904  0.233896    
ma1           -1.4142906  0.1010027 -14.0025 < 2.2e-16 ***
ma2            0.9264170  0.1418397   6.5314 6.514e-11 ***
ma3           -1.2166520  0.1082138 -11.2430 < 2.2e-16 ***
ma4            0.7304839  0.0818867   8.9207 < 2.2e-16 ***
intercept     -0.0068877  0.0029118  -2.3655  0.018007 *  
oil_temp_lag1  0.8315204  0.2658320   3.1280  0.001760 ** 
oil_temp_lag2  0.0029922  0.2335245   0.0128  0.989777    
high_fl_lag1  -0.3826939  0.2337437  -1.6372  0.101581    
high_fl_lag2   0.5891130  0.2220113   2.6535  0.007966 ** 
high_ll_lag1   0.9221942  0.6662482   1.3842  0.166309    
high_ll_lag2  -1.0953329  0.6568404  -1.6676  0.095399 .  
mean_fl_lag1   0.3114133  0.2384648   1.3059  0.191583    
mean_fl_lag2  -0.6058473  0.2246344  -2.6970  0.006996 ** 
mean_ll_lag1  -0.7639229  0.6840145  -1.1168  0.264070    
mean_ll_lag2   0.8589550  0.6815293   1.2603  0.207549    
low_fl_lag1    0.3011610  0.2032768   1.4815  0.138465    
low_fl_lag2   -0.1768557  0.2042379  -0.8659  0.386529    
low_ll_lag1   -0.4773459  0.7556655  -0.6317  0.527590    
low_ll_lag2    0.5458937  0.7611159   0.7172  0.473233    
sin_pred      -0.0205882  0.0075616  -2.7227  0.006474 ** 
cos_pred      -0.0025865  0.0033476  -0.7726  0.439733    
```

Output 8
Test of Significance for Complex Model 2 coefficients