Time Series Analysis of a Pressure Cell installed on the Big Springs Bridge

by

Ramin Ziaei

Final Project

Under the Supervision of Dr. Cheng

Department of Mathematics

University of Nebraska - Omaha

Omaha, Nebraska

August 10^{th}, 2021

Contents

List of Figures

1 Introduction

There are more than 9,000 integral abutment bridges and 4,000 semi-integral abutment bridges in the U.S., which increased dramatically in the past two decades. Nebraska is no exception – there are hundreds of integral and semi-integral abutment bridges in the state of Nebraska, and thus guidelines and specifications for these structures listed on the Bridge Office Policies and Procedures. The obvious advantage of using integral abutment bridges is their reduced construction and maintenance costs by eliminating bearings and expansion joints that make the bridge "joint-less". There has not been much research on these types of bridges especially in the state of Nebraska. Most of the previous researches were on straight joint-less bridges. In this work, in collaboration with the Nebraska Department of Transportation (NDOT), a horizontally curved bridge was selected in the city of Big Springs, Nebraska to monitored. Among all the instrumentations installed on this bridge, one of the pressure cells was chosen to be investigated in this work. On each abutment, there are 3 pressure cells installed, one of which from the North side was selected as the case study in this project. Figure 1 shows the original time series of the pressures in the selected pressure cell. As can be seen, there are several spikes throughout the plot that are outliers and need to be taken care of.

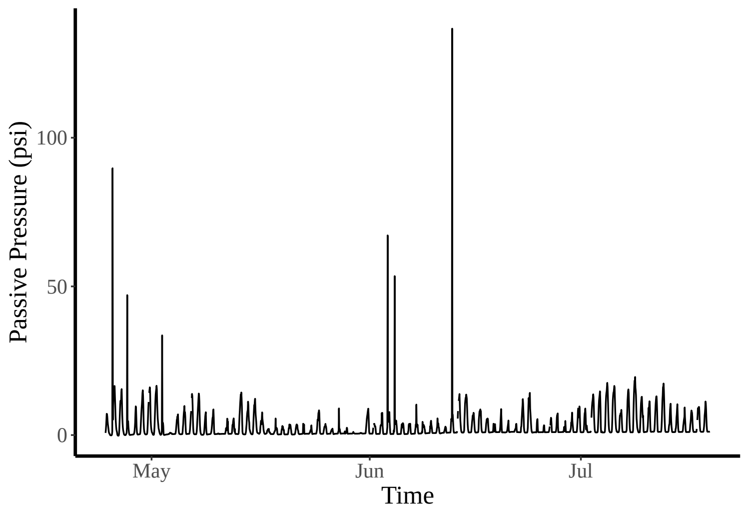

Figure 1: Original time series of the selected pressure cell

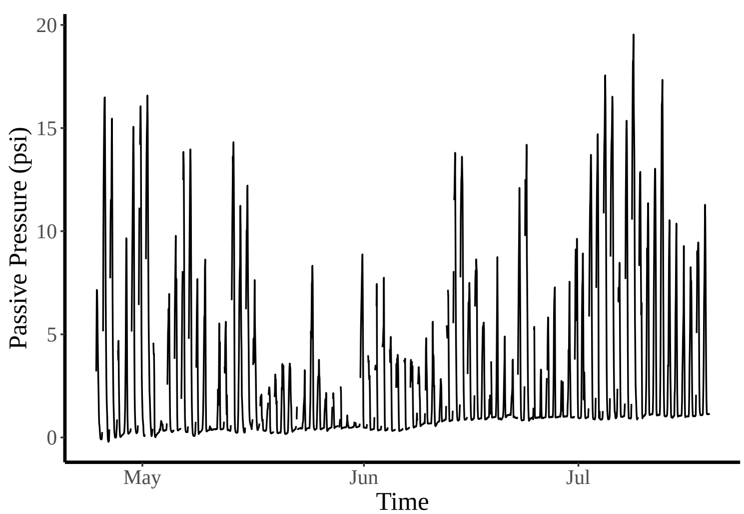

Figure 2: Time series of the selected pressure cell after removing noises

It was decided to define a threshold, and if the change from one data point to the next data point was over that threshold, that data point should be removed. The threshold was chosen as 100% increase in successive pressures. So, if the change from one pressure point to the next one is more than 100%, that point will be replaced by NAs which is shown in figure 2. Finally, a linear interpolation missing values implementation was used to fill the NAs. Figure 3 displays the final times series of this pressure cell after all missing values are implemented. It needs to be mentioned that the data is recorded every one hour. The last step to finalize the pressure cell data is to remove all negative values. As we know, this type of pressure cells are not capable of capturing negative values, and if there is any negative recorded pressure, they are just errors, and the actual pressure must be replaced by zero in those cases. This pressure cell represents the passive pressure generated from the bridge expansion. As the bridge expands, abutments move away form their original position, and the passive pressure is generated. We also know that bridge expansion and contraction is due to daily thermal fluctuations. So, we expect to see a daily period for the pressures as the temperature fluctuations repeat itself almost exactly every day. From 6-7 in the morning, the pressure starts to raise where the maximum occures at around 5-6 PM. After that, it comes down until 10-11 PM when there is no pressure anymore. It also can be seen from figure 3 that there is a drift towards the positive pressures in the graph showing some type of build-up pressure as time passes.

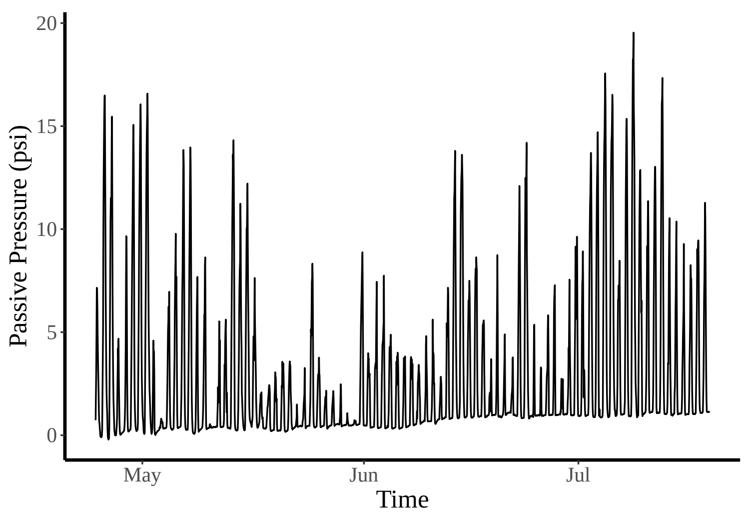

Figure 3: Final time series of the selected pressure cell after missing values imputations

2 Model Specification

After finalizing the time series data, it is time to come up with models that can appropriately fit the the pressure data. Based on figure 3 from the previous chapter, this times series is a non-stationary and seasonal series and needs to be dealt with accordingly. There are several ways to make non-stationary data sets stationary, and then using ARMA models to model them three of which are Logarithm transformation, Box-Cox transformation and differencing. Figures 4 to 16 show the transfored data and the corresponding ACF and PACF only based on differencing. Based on these graphs, taking the first difference will remove the upward trend of the data, but the seasonality is still visible in the time series plot. So, the first difference of the non-stationary part was also taken. Compared to only the first difference of the stationary part of the time series, this one seems to have the seasonality removed, but still there is some seasonality that can be seen in the plot. ACF and PACF of these differences are shown in figures 9 and 10. Autocorrelation looks much better for most parts but it is still not perfect. That is why the second and third difference of the stationary part were also calculated. According to figure 15 which shows the ACF of the the third difference of the stationary part with the first difference of the seasonal part, All lags are insignificant except for lags 1, 23, 24 and 25 which makes perfect sense. These are hourly data, and the trend repeats itself every 24 hours. We also know that it is normal to see high correlation for the neighboring points of the lag 24. So, considering all these plots, it seems that the third difference of the stationary part is the best option so far.

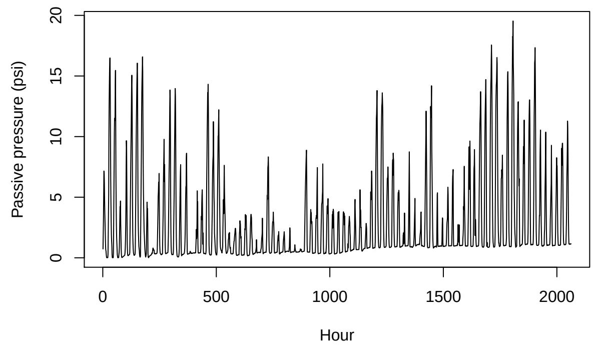

Figure 4: Finalized time series of the pressure cell

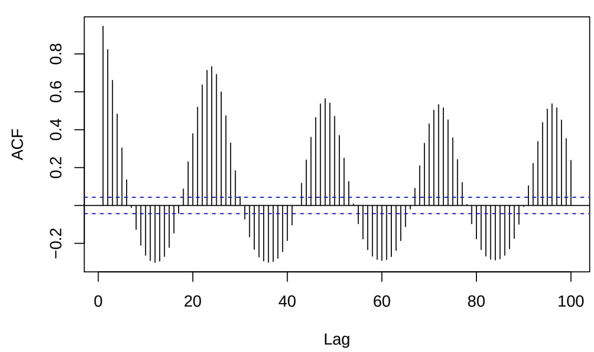

Figure 5: Auto-correlation of the time series

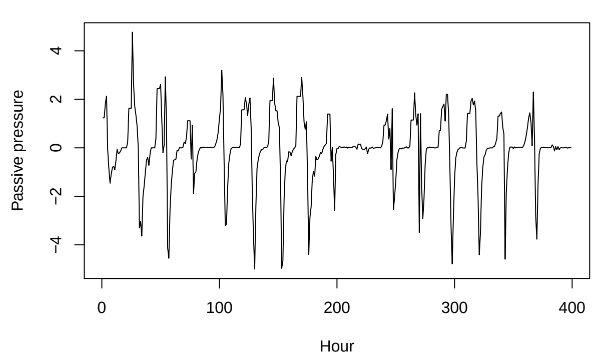

Figure 6: Time series of the first difference of the non-stationary part of the data

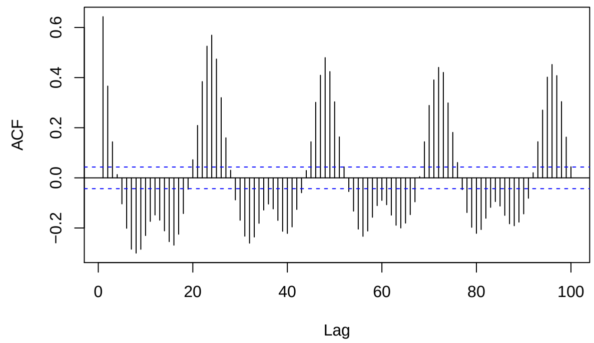

Figure 7: ACF of the first difference of the non-stationary part of the data

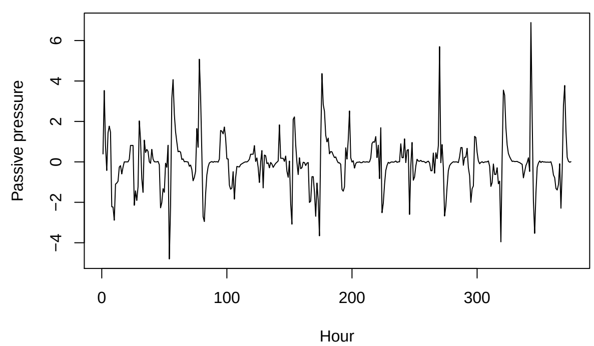

Figure 8: Time series of the first difference of both stationary and non-stationary parts of the data

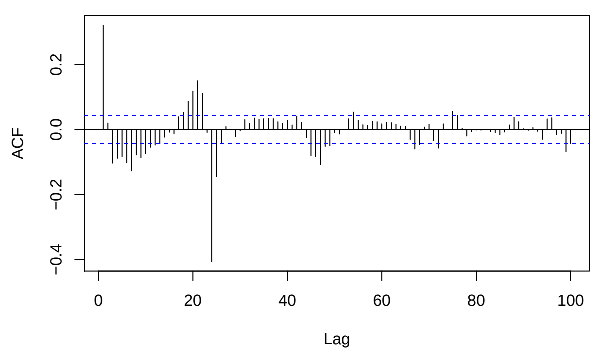

Figure 9: ACF of the first difference of both stationary and non-stationary parts of the data

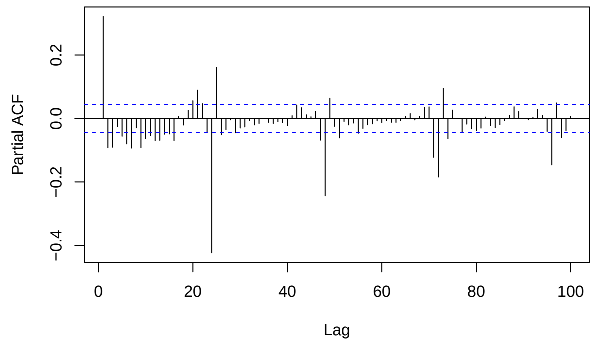

Figure 10: Partial ACF of the first difference of both stationary and non-stationary parts of the data

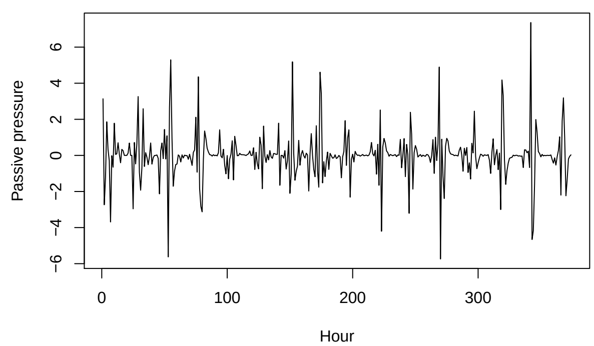

Figure 11: Time series of the second difference of the stationary part and first difference of the non-stationary part of the data

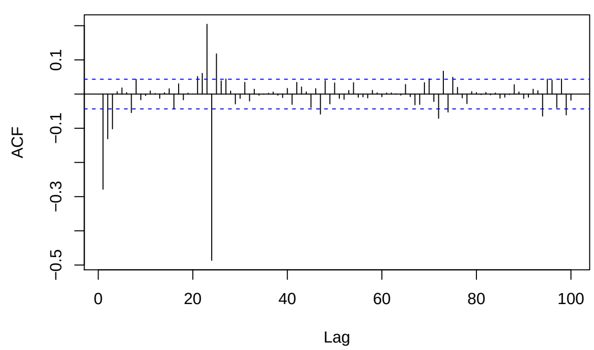

Figure 12: ACF of the second difference of the stationary part and first difference of the non-stationary part of the data

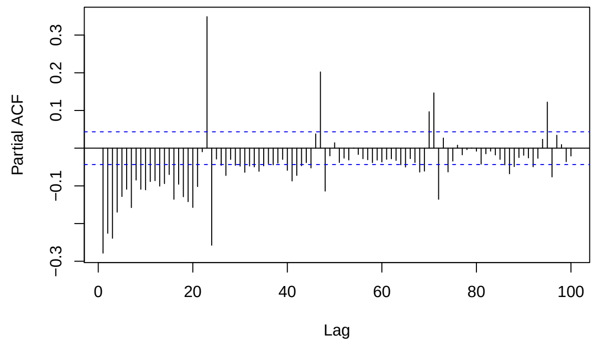

Figure 13: PACF of the second difference of the stationary part and first difference of the non-stationary part of the data

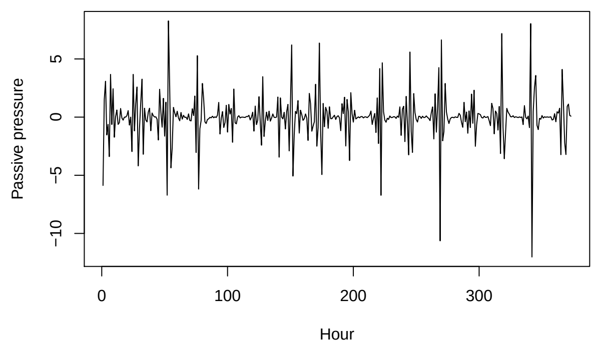

Figure 14: Time series of the third difference of the stationary part and first difference of the non-stationary part of the data

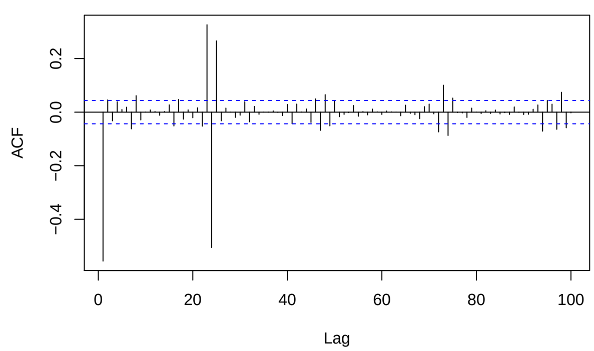

Figure 15: ACF of the third difference of the stationary part and first difference of the non-stationary part of the data

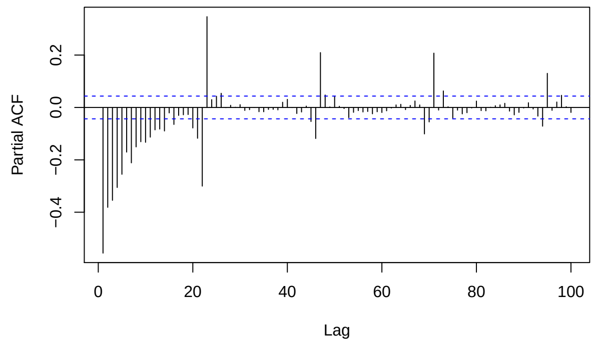

Figure 16: PACF of the third difference of the stationary part and first difference of the non-stationary part of the data

Next, the Box-Cox transformation is considered. Since there is not a specific type of data transformation for any given time series data, different options need to be considered in order to be able choose the best transformation. Using the Box-Cox transformation function, the λ was selected as 0.2062893. According to figure 17, the Box-Cox transformation does not get rid of the upward trend of the data. For that, we take the first difference of the data which is shown in figure 19 where it can be seen that the the upward trend is gone, but the seasonality is still noticeable. Figure 21 shows the plot where both upward trend and seasonality are gone by taking the first difference of both of them. Just like the previous case, since the first difference is not ideal, second and third differences were also taken to see if there is any difference. Third difference is again the best here, but when it is compared to the previous case which did not have any transformation, There are still some autocorrelations in lags 7, 8, 31, 32, ... in the Box-Cox transformed data (figure 28). So, this transformation is not as good as the the one without transformation so far.

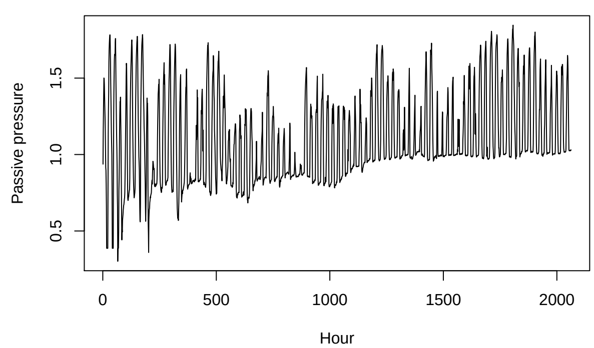

Figure 17: Time series of the Box-Cox transformed data

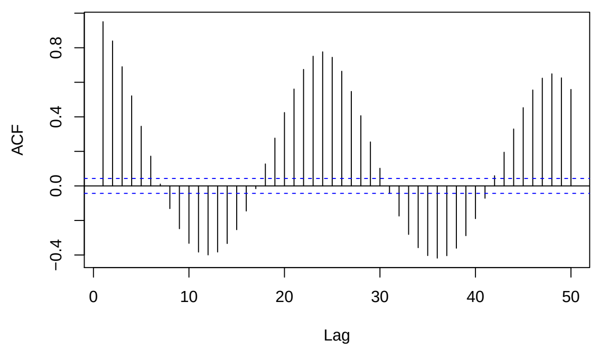

Figure 18: ACF of the Box-Cox transformed data

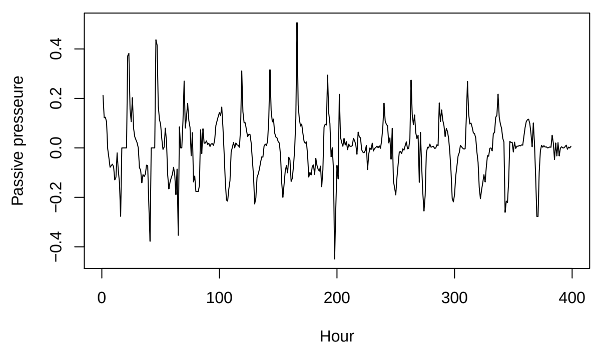

Figure 19: Time series of the first difference of the stationary part of the Box-Cox transformed data

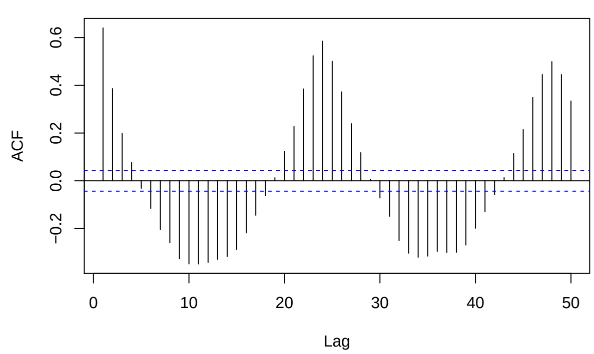

Figure 20: ACF of the first difference of the stationary part of the Box-Cox transformed data

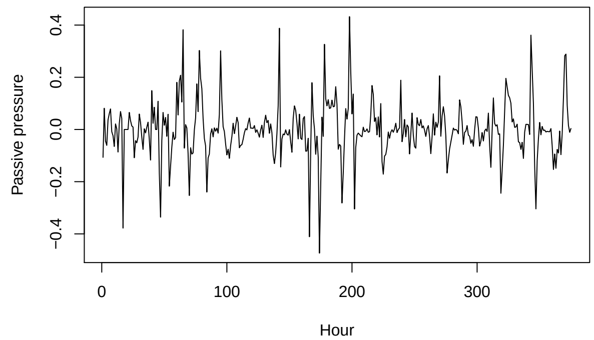

Figure 21: Time series of the first difference of the stationary part and the first difference of the non-stationary part of the Box-Cox transformed data

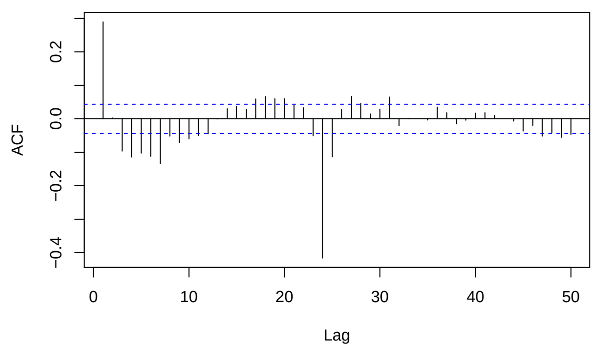

Figure 22: ACF of the first difference of the stationary part and the first difference of the non-stationary part of the Box-Cox transformed data

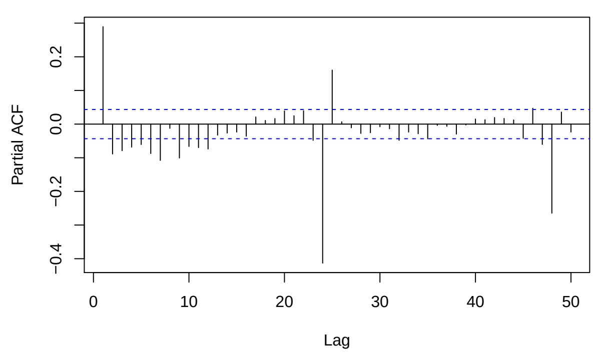

Figure 23: PACF of the first difference of the stationary part and the first difference of the non-stationary part of the Box-Cox transformed data

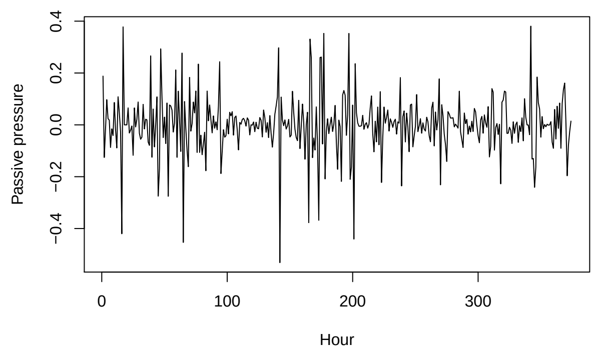

Figure 24: Time series of the second difference of the stationary part and the first difference of the non-stationary part of the Box-Cox transformed data

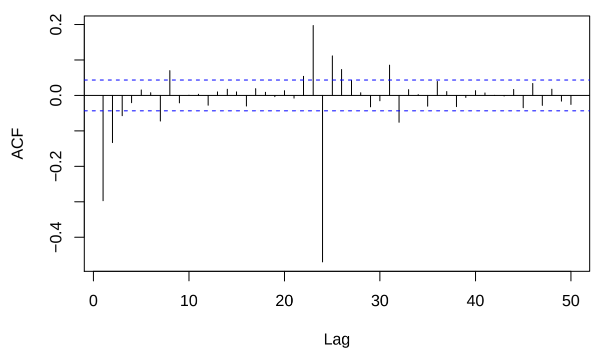

Figure 25: ACF of the second difference of the stationary part and the first difference of the non-stationary part of the Box-Cox transformed data

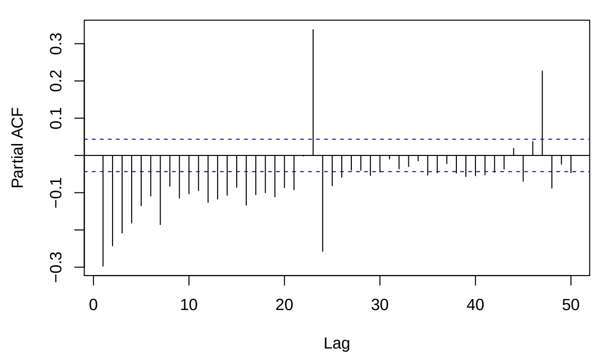

Figure 26: PACF of the second difference of the stationary part and the first difference of the non-stationary part of the Box-Cox transformed data

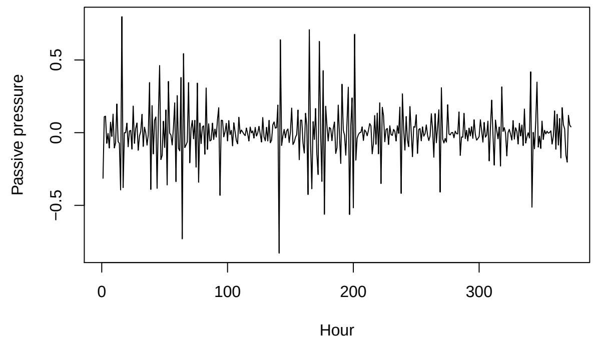

Figure 27: Time series of the third difference of the stationary part and the first difference of the non-stationary part of the Box-Cox transformed data

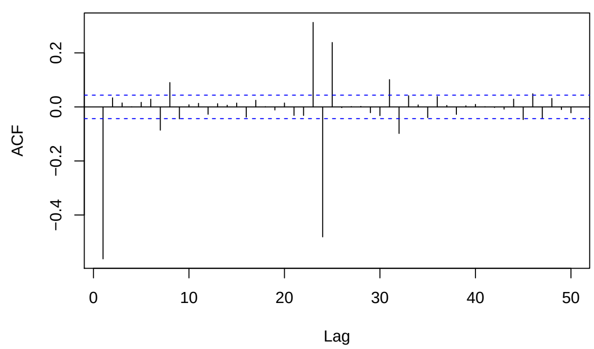

Figure 28: ACF of the third difference of the stationary part and the first difference of the non-stationary part of the Box-Cox transformed data

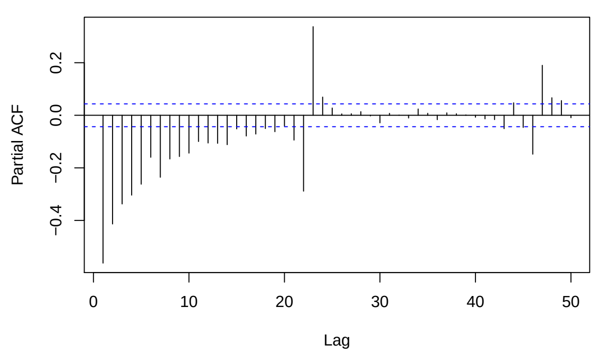

Figure 29: PACF of the third difference of the stationary part and the first difference of the non-stationary part of the Box-Cox transformed data

Finally, we do a Log transformation. Figure 30 shows that using the log deformation can take care of the upward trend of the data, and it will oscillate around the origin. The process is the same as for the previous two. Figure 41 shows that the third difference of the stationary part along with the first difference of the seasonal part will remove most of the lag correlations in the data. However, this one has more remaining lags even compared to the Box-Cox transformation. So, all in all, after comparing all three cases, the first one which was just the third difference of the stationary part of the time series without any transformation is the best option.

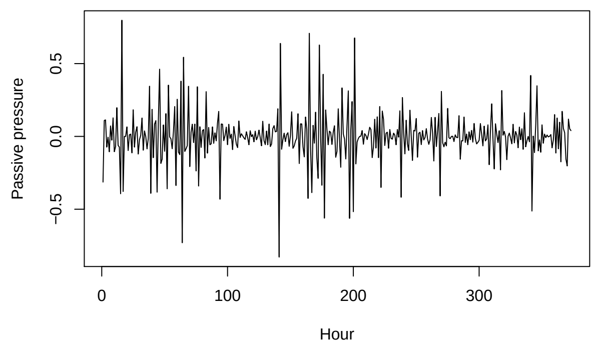

Figure 30: Time series of the Log transformed data

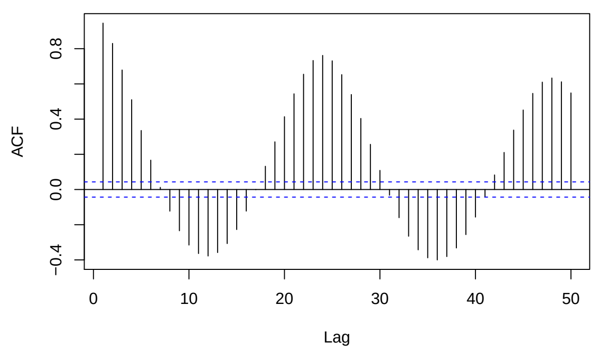

Figure 31: ACF of the Log transformed data

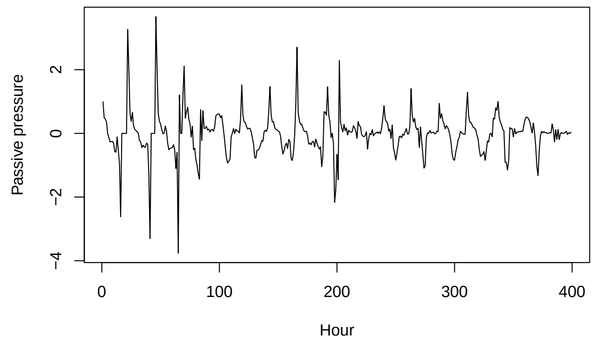

Figure 32: Time series of the first difference of the stationary part of the Log transformed data

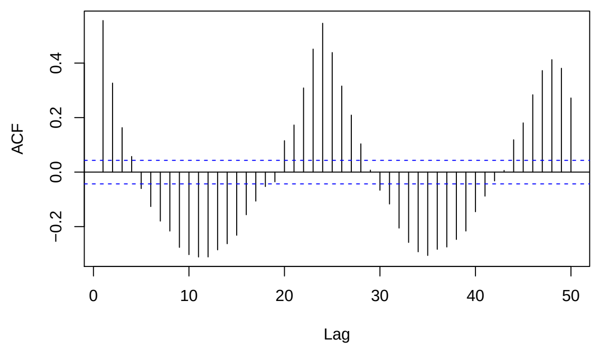

Figure 33: ACF of the first difference of the stationary part of the Log transformed data

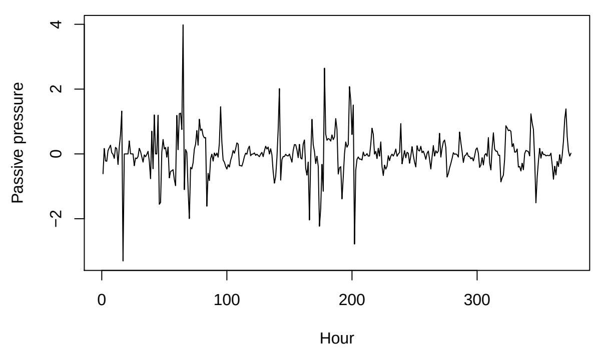

Figure 34: Time series of the first difference of the stationary part and the first difference of the non-stationary part of the Log transformed data

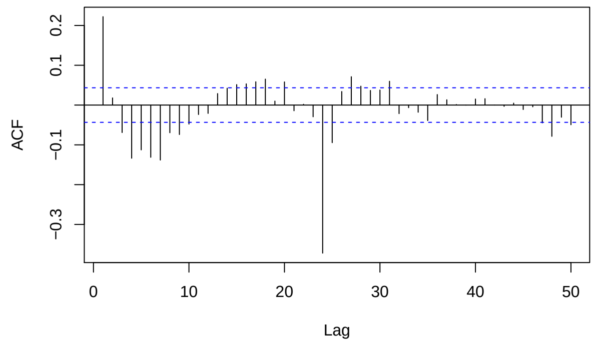

Figure 35: ACF of the first difference of the stationary part and the first difference of the non-stationary part of the Log transformed data

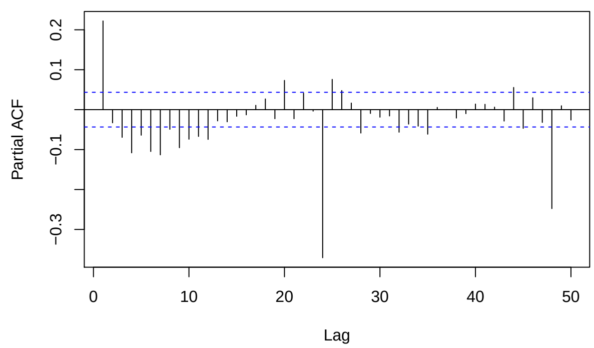

Figure 36: PACF of the first difference of the stationary part and the first difference of the non-stationary part of the Log transformed data

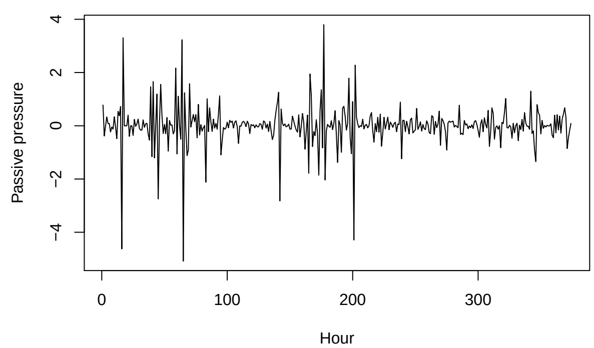

Figure 37: Time series of the second difference of the stationary part and the first difference of the non-stationary part of the Log transformed data

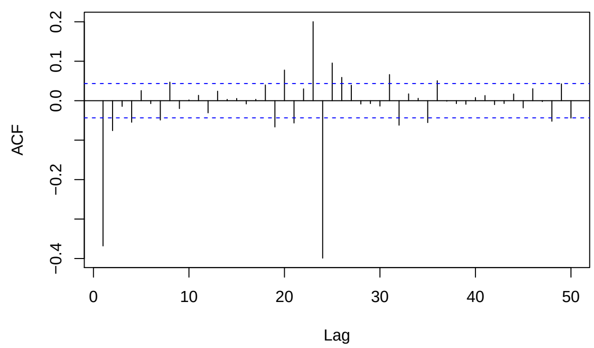

Figure 38: ACF of the second difference of the stationary part and the first difference of the non-stationary part of the Log transformed data

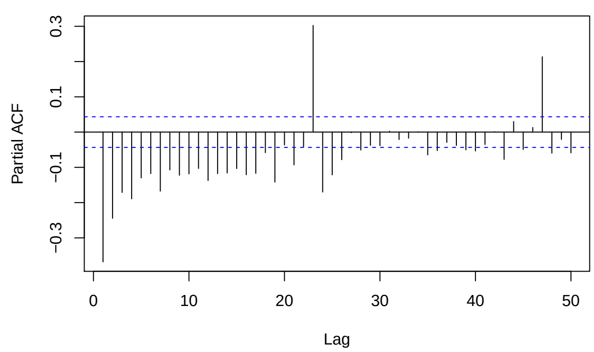

Figure 39: PACF of the second difference of the stationary part and the first difference of the non-stationary part of the Log transformed data

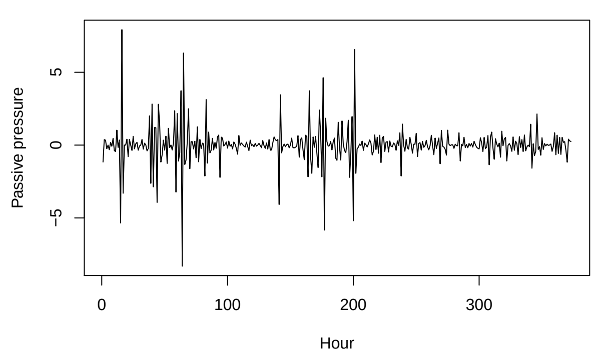

Figure 40: Time series of the third difference of the stationary part and the first difference of the non-stationary part of the Log transformed data

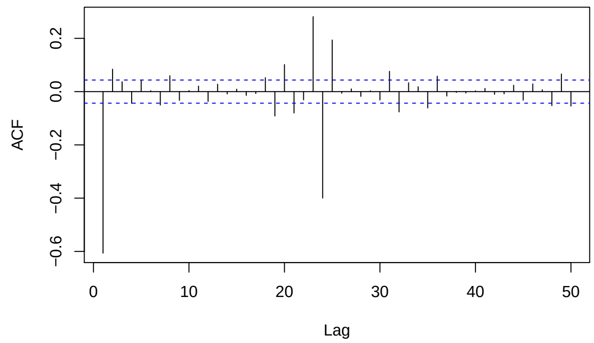

Figure 41: ACF of the third difference of the stationary part and the first difference of the non-stationary part of the Log transformed data

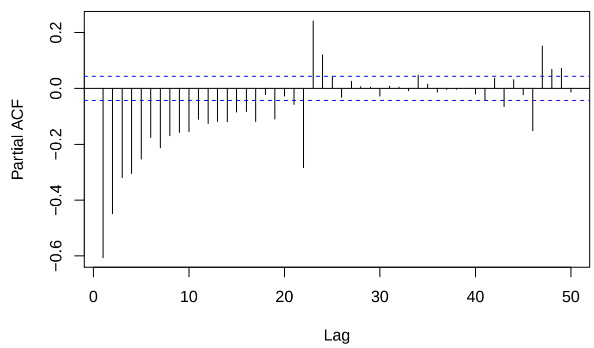

Figure 42: PACF of the third difference of the stationary part and the first difference of the non-stationary part of the Log transformed data

3 Model Fitting

Having specified a proper transformation to make the seasonal data stationary, now it is time to fit a tentative model to the time series data. According to figures 15 and 16, and since there is only one spike at lag 24 and its neighboring lags in the ACF plot, and the PACF plot shows decaying correlations at lags 24, 48, 72 and so on, it can be concluded that this model has an MA component in it. Therefore, the data can be modeled as an $ARIMA(0,3,1)(0,1,1)_{24}$. However, this is not the only model to consider here. Based on the their performances, different models are considered. The seasonality part of the model is taken care of by the first difference. But, for the stationary part, third difference seems to be more suitable for this model. The level of MA seems to be 1 based on figure 15, but we can play with MA and AR to see if there is any room for improvement in the model or not. Eventually, 9 different models were considered in this study for further investigation.

```
## Series: as.vector(dat)
## ARIMA(0,3,1)(0,1,1)[24]
##
## Coefficients:
##           ma1     sma1
```

```
##       -1.0000  -0.9514
## s.e.   0.0018   0.0112
##
## sigma^2 estimated as 0.8549:  log likelihood=-2764.85
## AIC=5535.71   AICc=5535.72   BIC=5552.56
```

```
## Series: as.vector(dat)
## ARIMA(0,3,0)(0,1,1)[24]
##
## Coefficients:
##          sma1
##       -0.9606
## s.e.   0.0104
##
## sigma^2 estimated as 2.121:  log likelihood=-3686.82
## AIC=7377.65   AICc=7377.65   BIC=7388.88
```

```
## Series: as.vector(dat)
## ARIMA(0,3,2)(0,1,1)[24]
##
## Coefficients:
##           ma1     ma2     sma1
##       -1.5322  0.5322  -0.9195
## s.e.   0.0358  0.0358   0.0122
##
## sigma^2 estimated as 0.7634:  log likelihood=-2643.83
## AIC=5295.65   AICc=5295.67   BIC=5318.12
```

```
## Series: as.vector(dat)
## ARIMA(1,3,1)(0,1,1)[24]
##
## Coefficients:
##           ar1      ma1     sma1
##       -0.2448  -1.0000  -0.9458
## s.e.   0.0212   0.0018   0.0114
##
## sigma^2 estimated as 0.805:  log likelihood=-2702.1
## AIC=5412.21   AICc=5412.23   BIC=5434.68
```

```
## Series: as.vector(dat)
## ARIMA(0,2,1)(0,1,1)[24]
##
## Coefficients:
##           ma1     sma1
##       -0.5371  -0.9206
## s.e.   0.0363   0.0122
##
## sigma^2 estimated as 0.7625:  log likelihood=-2636.66
## AIC=5279.31   AICc=5279.32   BIC=5296.17
```

```
## Series: as.vector(dat)
## ARIMA(0,2,2)(0,1,1)[24]
##
## Coefficients:
##           ma1      ma2     sma1
##       -0.6262  -0.3738  -0.8969
```

```
## s.e.   0.0180   0.0179   0.0121
##
## sigma^2 estimated as 0.6247:  log likelihood=-2435.78
## AIC=4879.56   AICc=4879.58   BIC=4902.04
```

```
## Series: as.vector(dat)
## ARIMA(1,3,2)(0,1,1)[24]
##
## Coefficients:
##          ar1      ma1     ma2     sma1
##       0.4250  -1.9937  0.9943  -0.9057
## s.e.  0.0193   0.0026  0.0025   0.0130
##
## sigma^2 estimated as 0.6207:  log likelihood=-2432.3
## AIC=4874.6   AICc=4874.63   BIC=4902.7
```

```
## Series: as.vector(dat)
## ARIMA(2,3,1)(0,1,1)[24]
##
## Coefficients:
##           ar1      ar2      ma1     sma1
##       -0.2939  -0.1948  -1.0000  -0.9342
## s.e.   0.0204   0.0196   0.0017   0.0111
##
## sigma^2 estimated as 0.7769:  log likelihood=-2663.22
## AIC=5336.44   AICc=5336.47   BIC=5364.53
```

```
## Series: as.vector(dat)
## ARIMA(1,2,2)(0,1,1)[24]
##
## Coefficients:
##          ar1      ma1      ma2     sma1
##       0.3239  -0.8881  -0.1119  -0.9103
## s.e.  0.0447   0.0451   0.0451   0.0121
##
## sigma^2 estimated as 0.6111:  log likelihood=-2414.54
## AIC=4839.09   AICc=4839.12   BIC=4867.18
```

4 Model Diagnostics

After forming the very first tentative models, it is time to see how good they perform in terms of their residuals and auto correlations. First, we form the residual plots for the selected models. Figures 43 to 78 show residuals, their ACF and PACF of the residuals. Based on these plots, none of the models are perfect for the time series data since they still have auto correlations in some of the lags. The final model can be chosen either based on the AIC-BIC criteria or based on the ACF plots for each model. In terms of AIC-BIC, the best 3 models are 6, 9 and 5 in order. However, when comparing the ACF of the residuals, none of them are perfect. There are still autocorrelations left in some of the lags of all models residual plots. Models 2, 3 and 8 are the ones that have the minimum number of left autocorrelations. Although it is not a perfect model, we can still try them to see if their performance is satisfying.

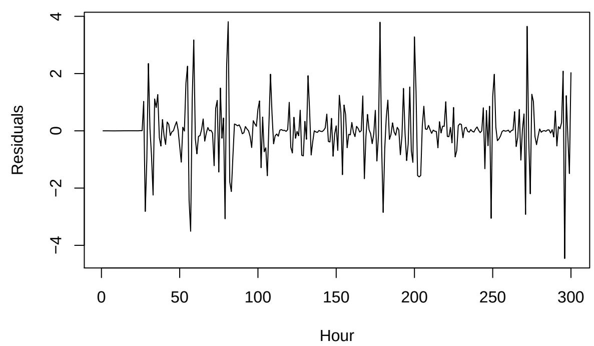

Figure 43: Residuals of model 1

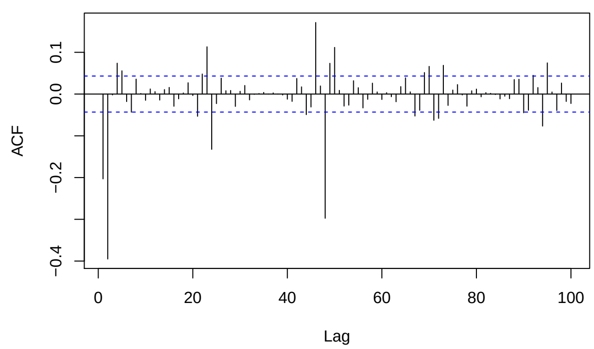

Figure 44: ACF of the residuals of model 1

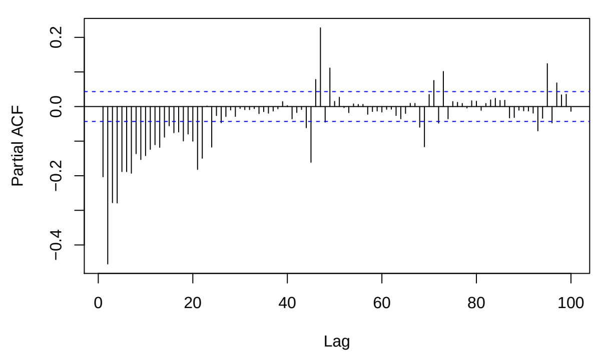

Figure 45: PACF of the residuals of model 1

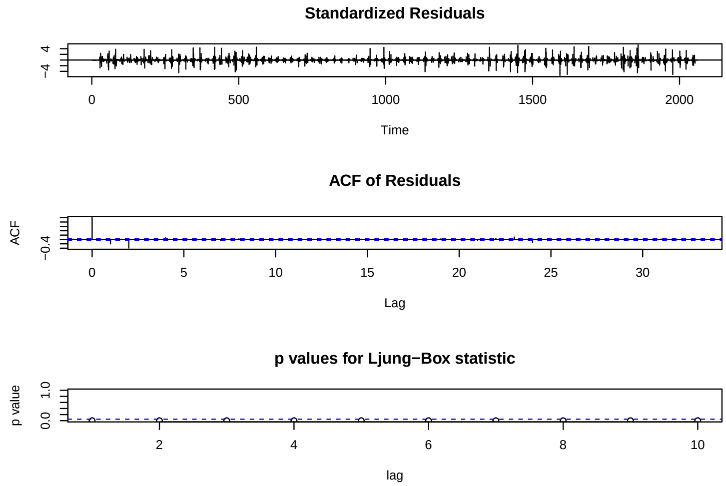

Figure 46: Diagnostic display for model 1

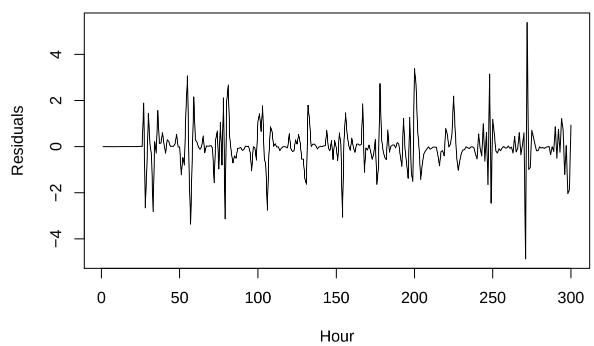

Figure 47: Residuals of model 2

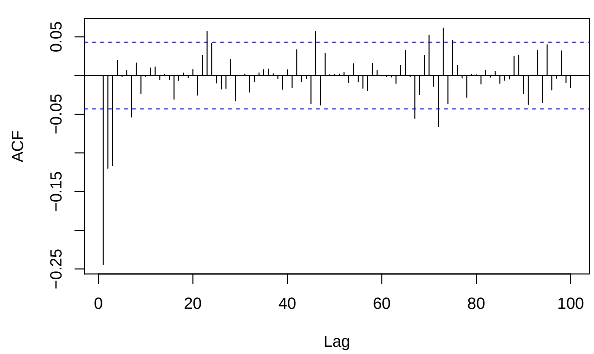

Figure 48: ACF of the residuals of model 2

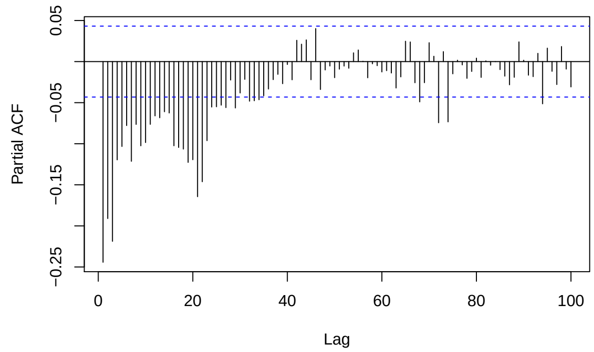

Figure 49: PACF of the residuals of model 2

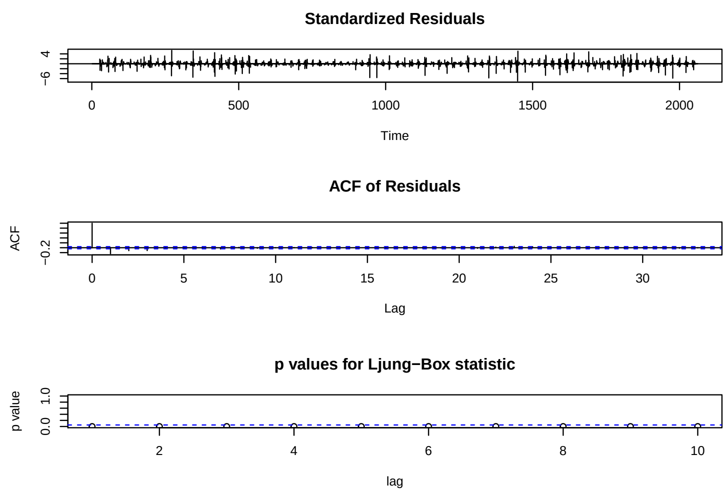

Figure 50: Diagnostic display for model 2

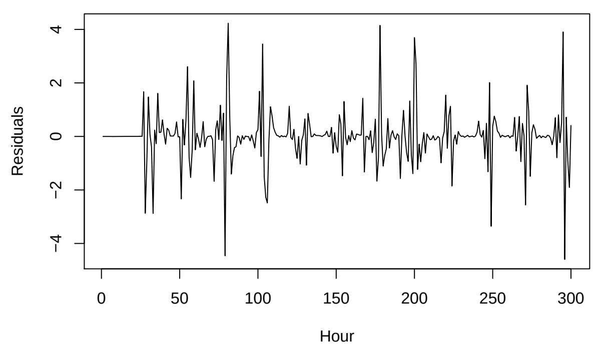

Figure 51: Residuals of model 3

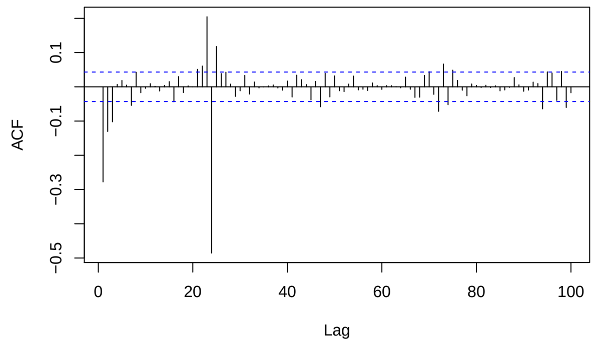

Figure 52: ACF of the residuals of model 3

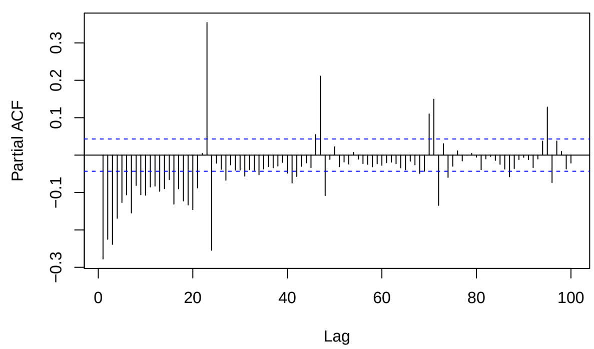

Figure 53: PACF of the residuals of model 3

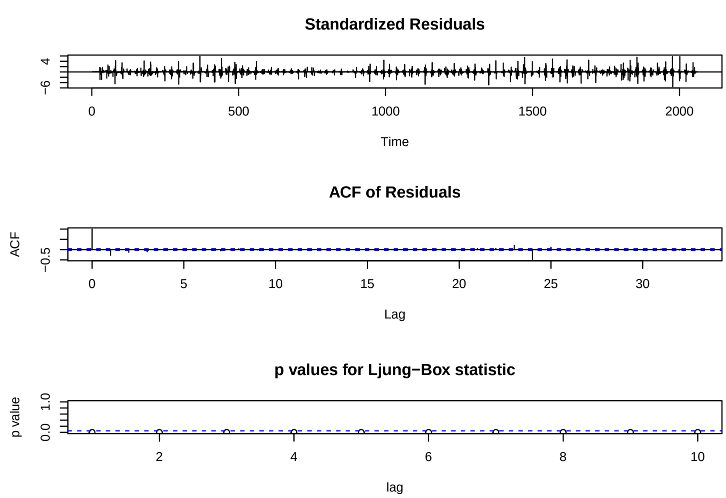

Figure 54: Diagnostic display for model 3

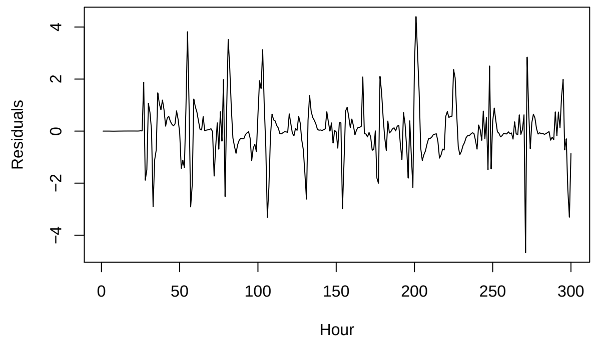

Figure 55: Residuals of model 4

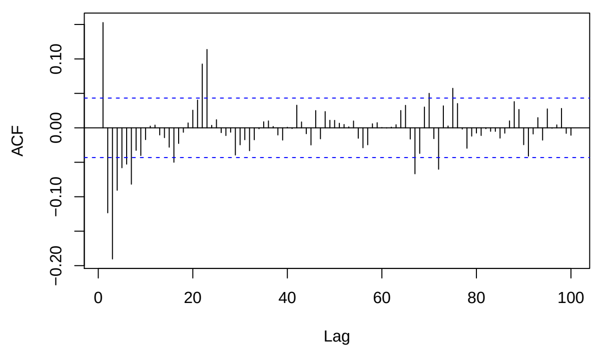

Figure 56: ACF of the residuals of model 4

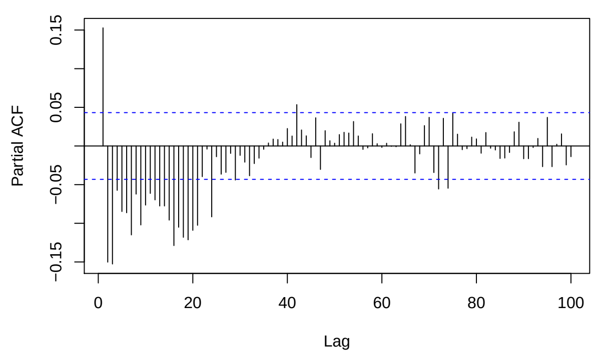

Figure 57: PACF of the residuals of model 4

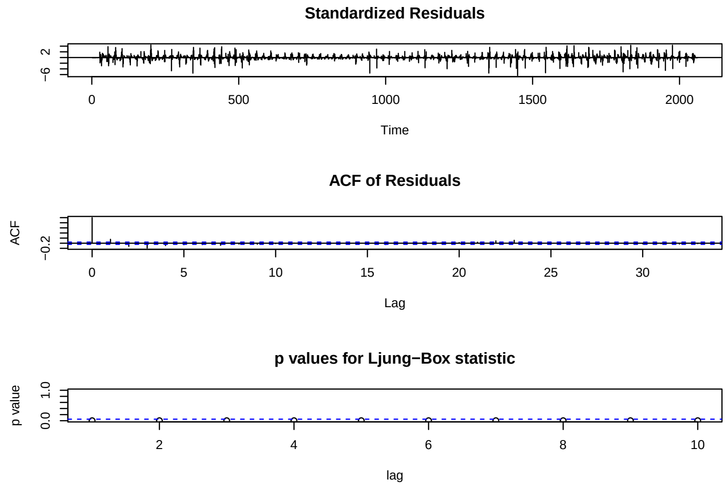

Figure 58: Diagnostic display for model 4

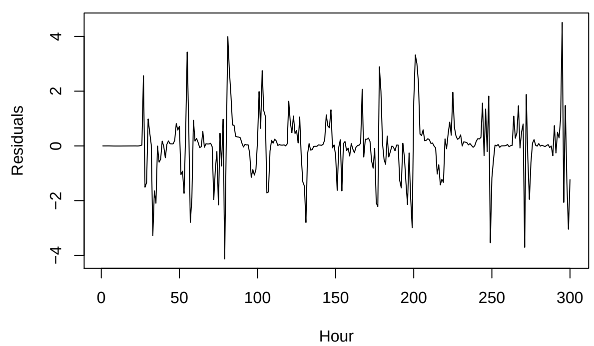

Figure 59: Residuals of model 5

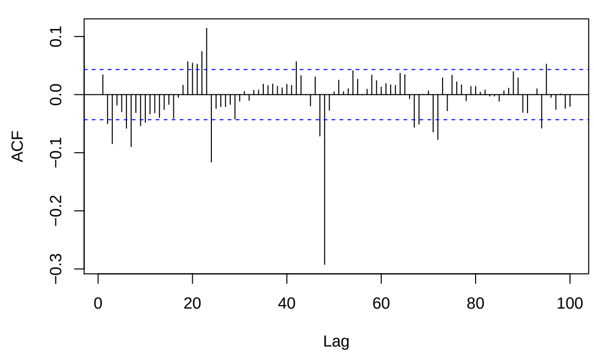

Figure 60: ACF of the residuals of model 5

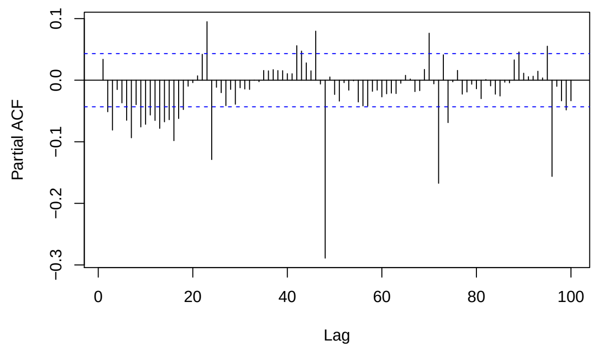

Figure 61: PACF of the residuals of model 5

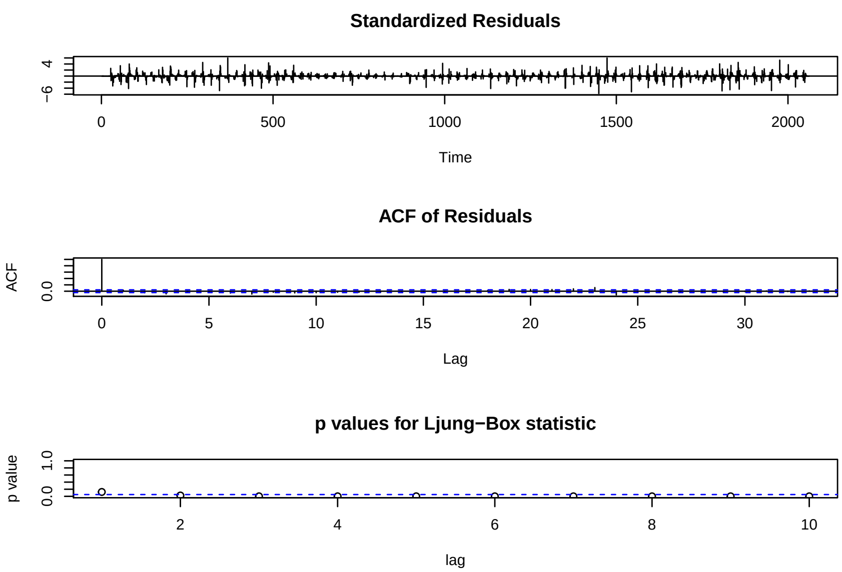

Figure 62: Diagnostic display for model 5

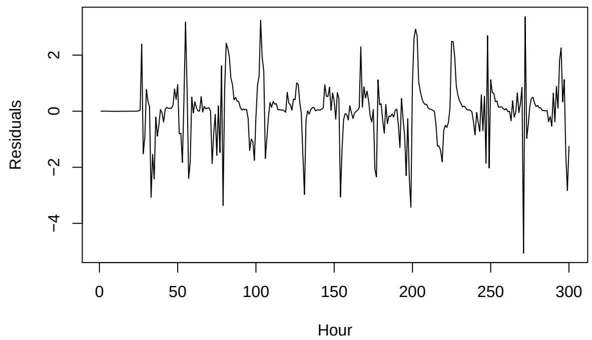

Figure 63: Residuals of model 6

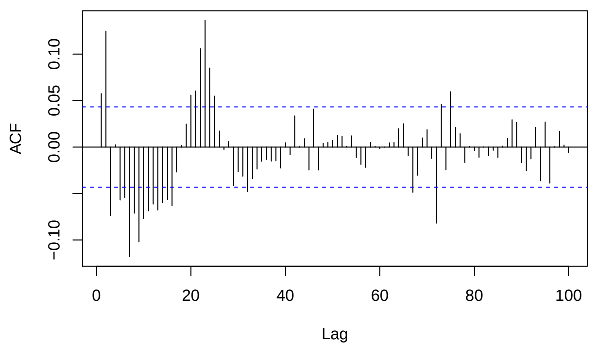

Figure 64: ACF of the residuals of model 6

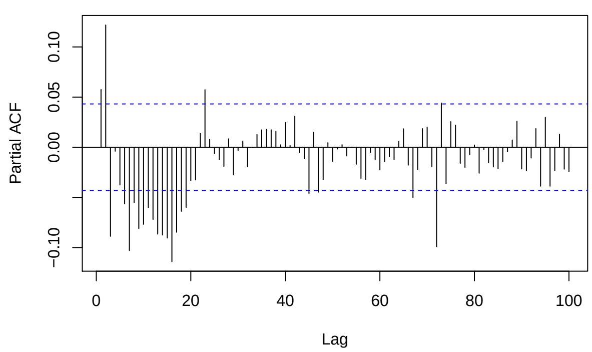

Figure 65: PACF of the residuals of model 6

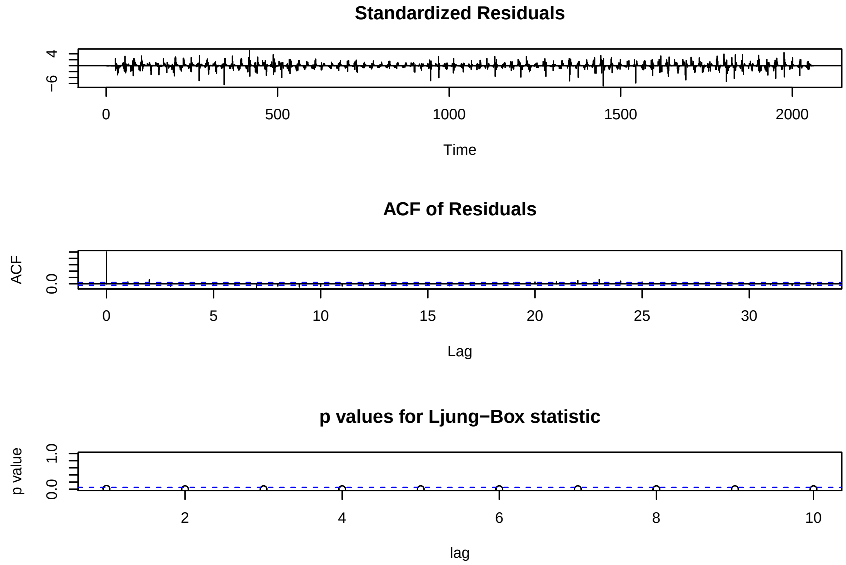

Figure 66: Diagnostic display for model 6

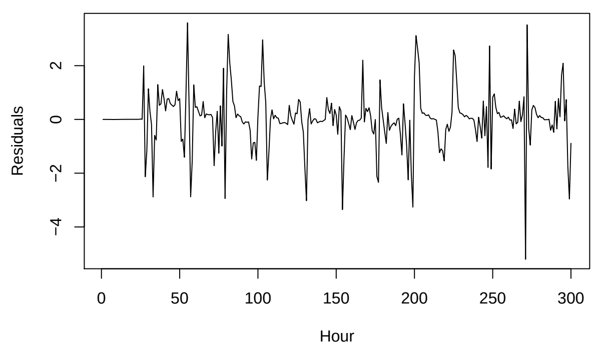

Figure 67: Residuals of model 7

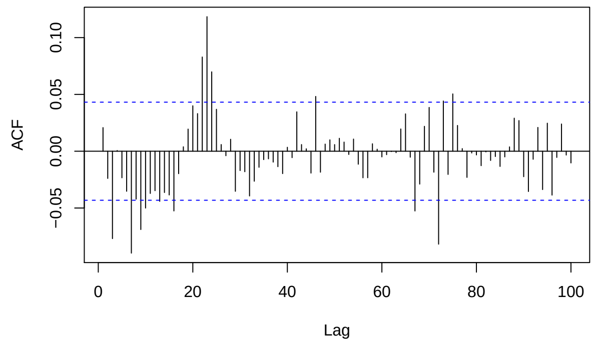

Figure 68: ACF of the residuals of model 7

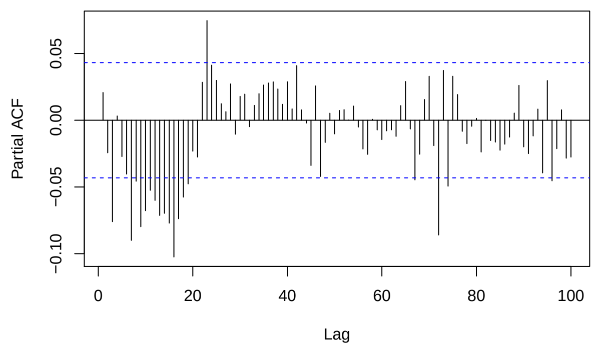

Figure 69: PACF of the residuals of model 7

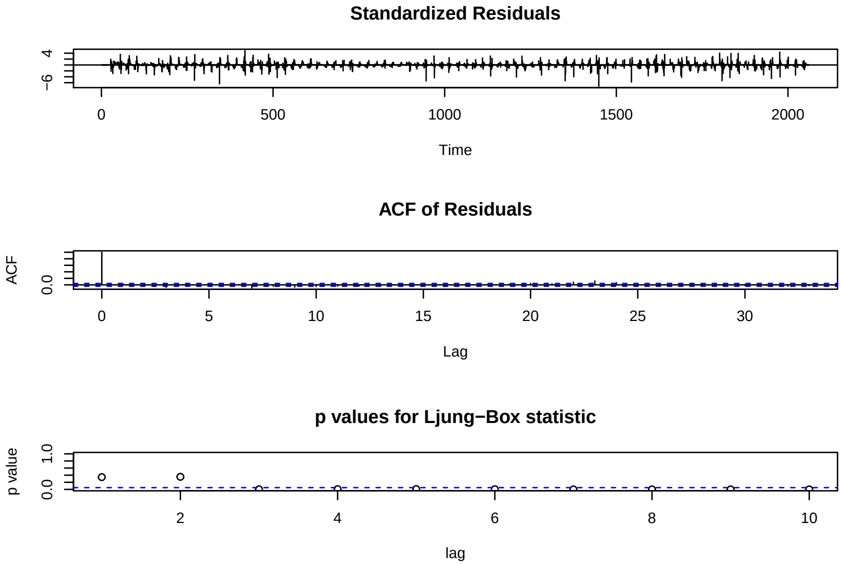

Figure 70: Diagnostic display for model 7

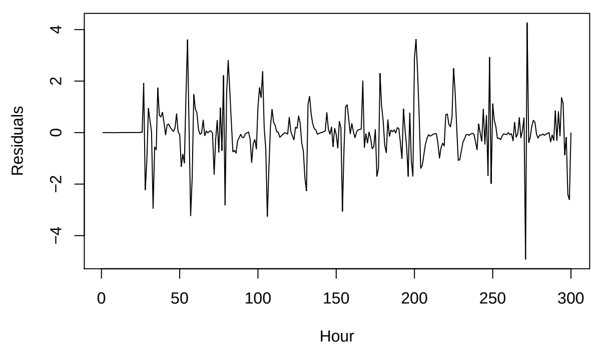

Figure 71: Residuals of model 8

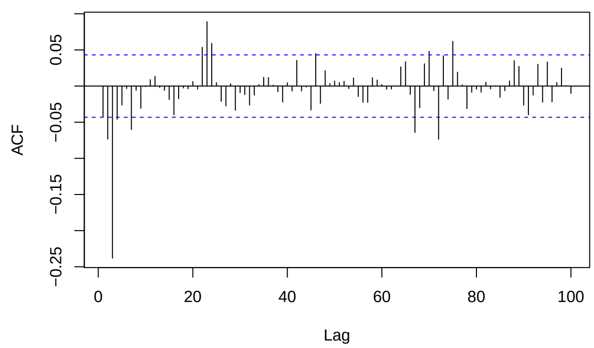

Figure 72: ACF of the residuals of model 8

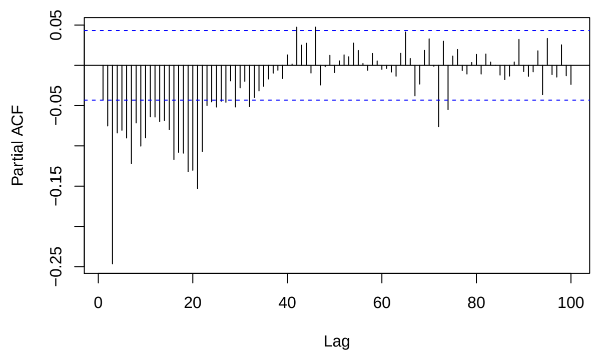

Figure 73: PACF of the residuals of model 8

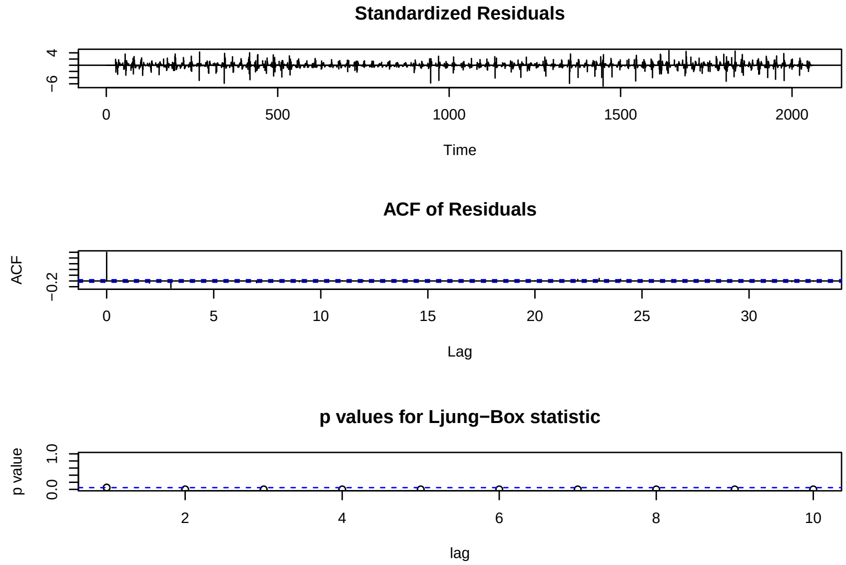

Figure 74: Diagnostic display for model 8

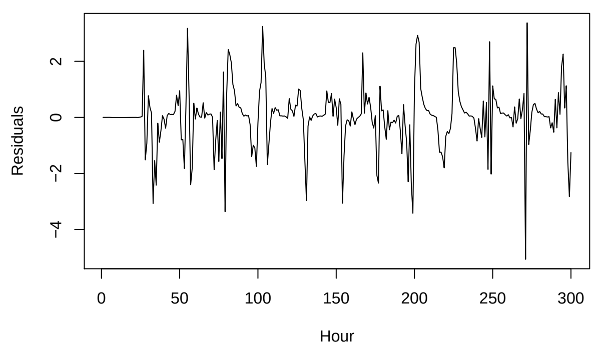

Figure 75: Residuals of model 9

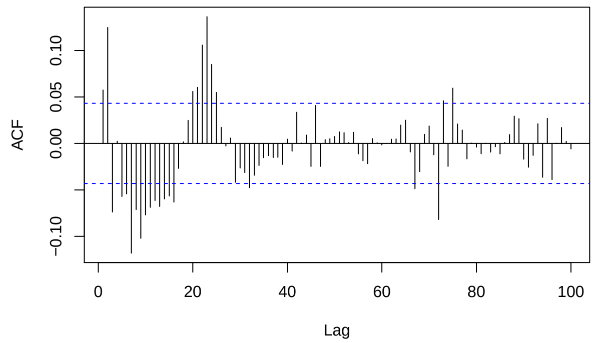

Figure 76: ACF of the residuals of model 9

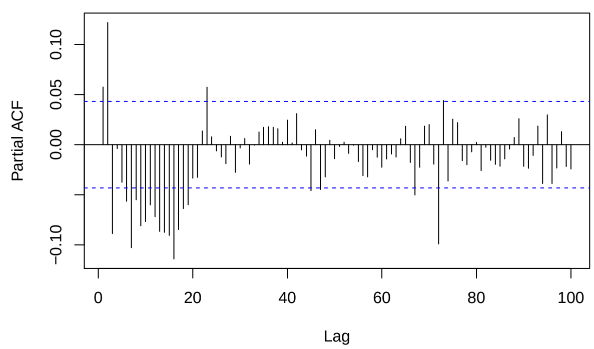

Figure 77: PACF of the residuals of model 9

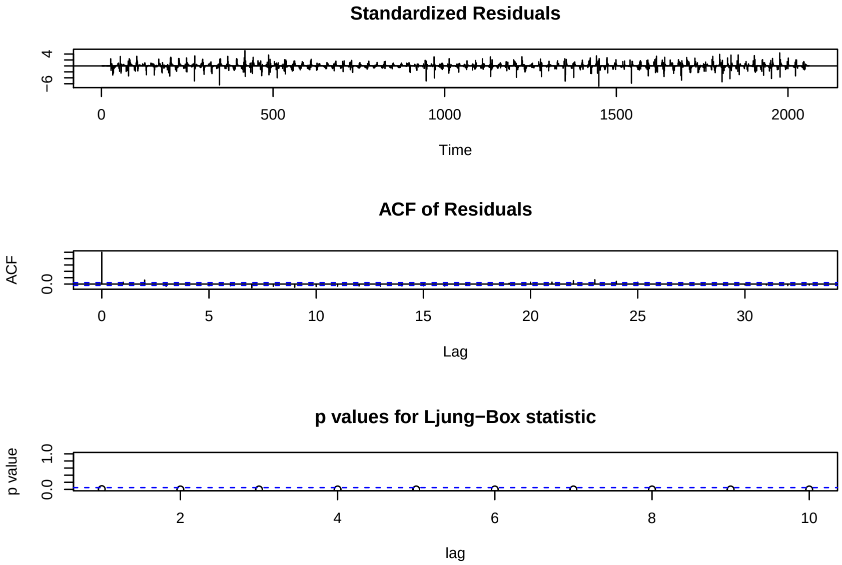

Figure 78: Diagnostic display for model 9

To check the normality of the residuals, the QQ plot and the histogram of them is provided here. According to figures 79 to 96, the histogram of all the models look normal. However, when the QQ graph is plotted, it shows that all models residuals are far from normal especially at their tails.

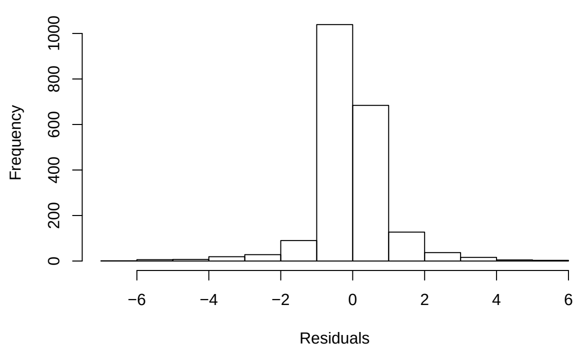

Figure 79: Histogram of the residuals of model 1

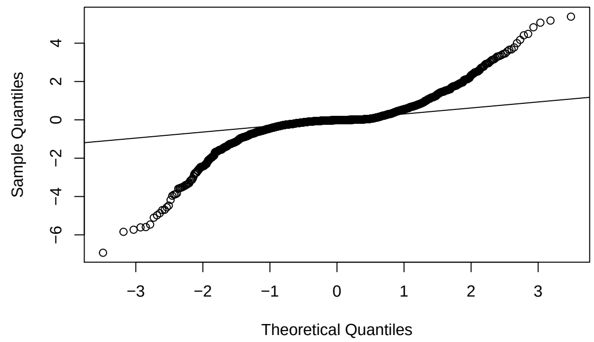

Figure 80: QQ plot of model 1

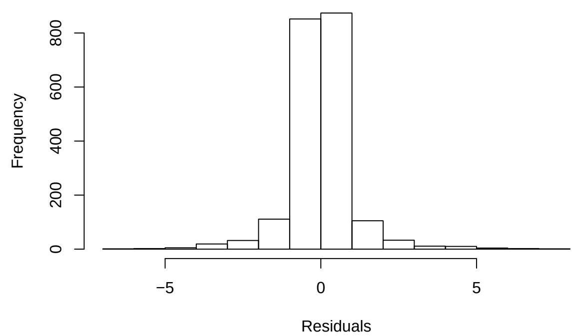

Figure 81: Histogram of the residuals of model 2

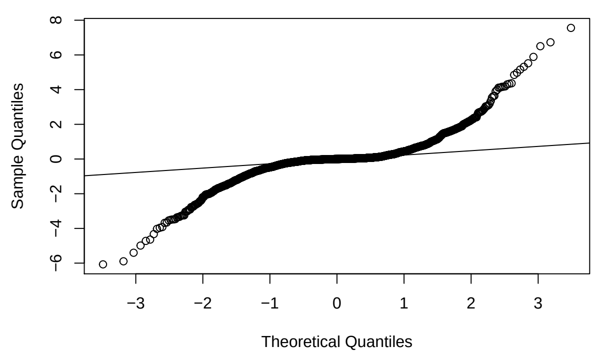

Figure 82: QQ plot of model 2

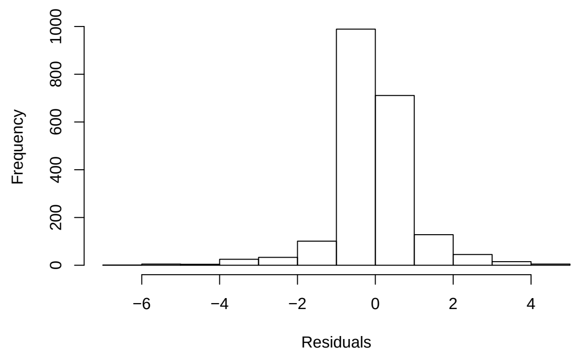

Figure 83: Histogram of the residuals of model 3

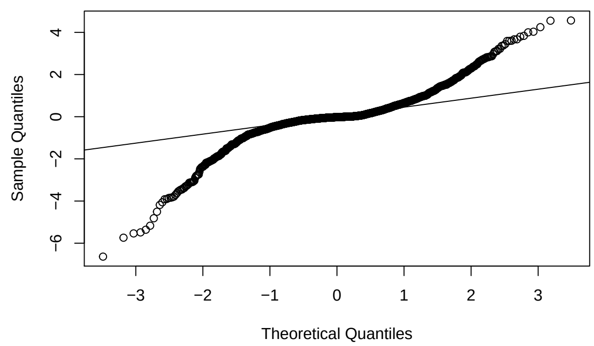

Figure 84: QQ plot of model 3

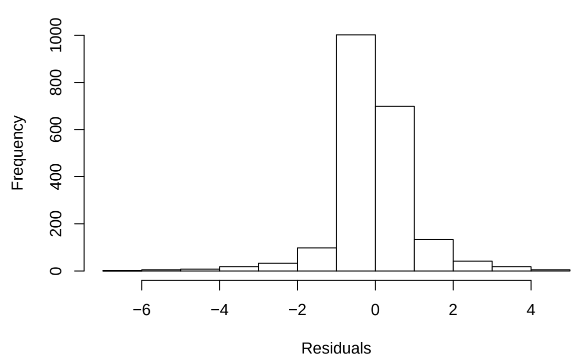

Figure 85: Histogram of the residuals of model 4

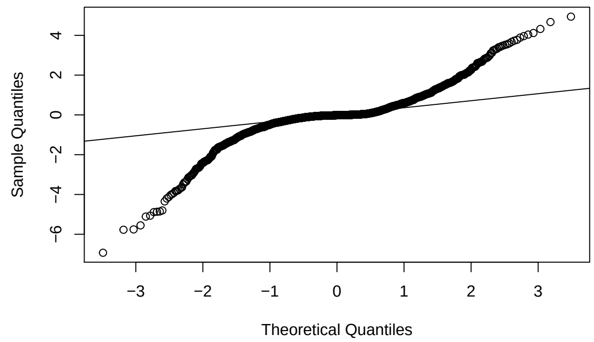

Figure 86: QQ plot of model 4

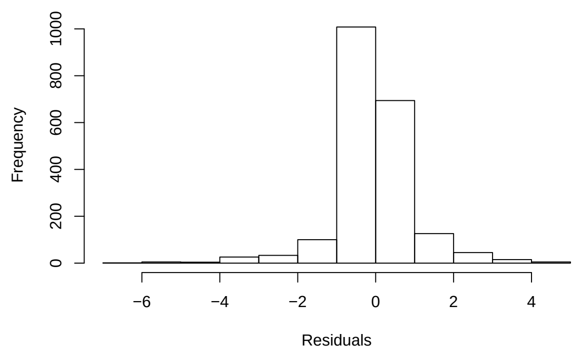

Figure 87: Histogram of the residuals of model 5

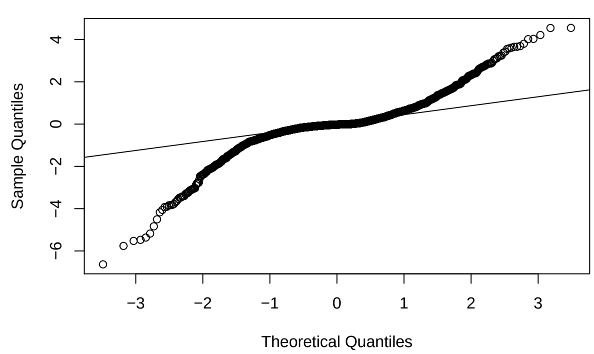

Figure 88: QQ plot of model 5

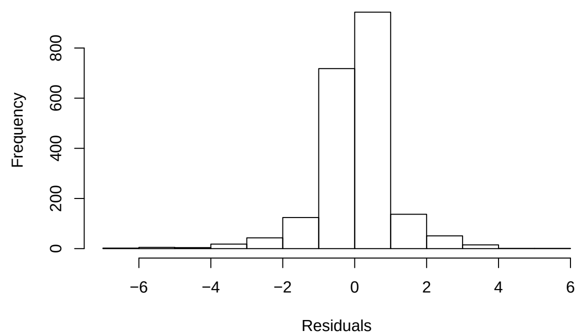

Figure 89: Histogram of the residuals of model 6

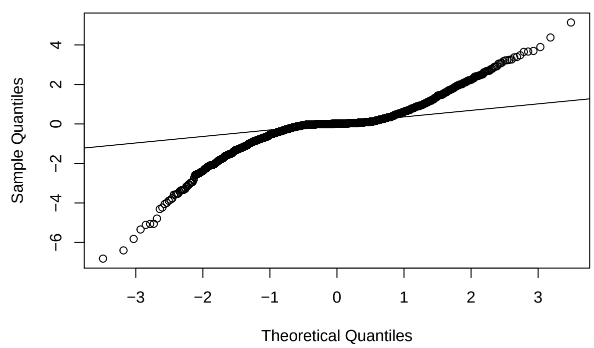

Figure 90: QQ plot of model 6

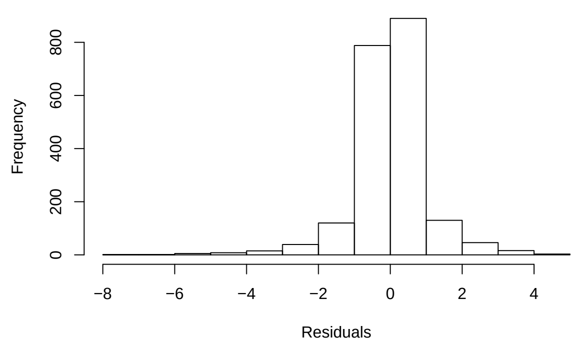

Figure 91: Histogram of the residuals of model 7

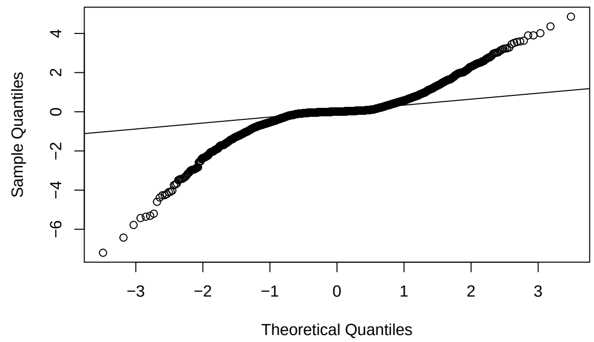

Figure 92: QQ plot of model 7

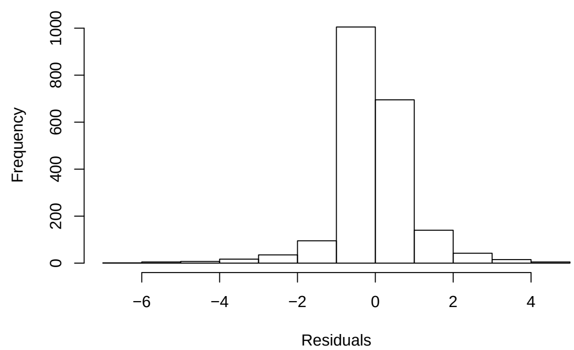

Figure 93: Histogram of the residuals of model 8

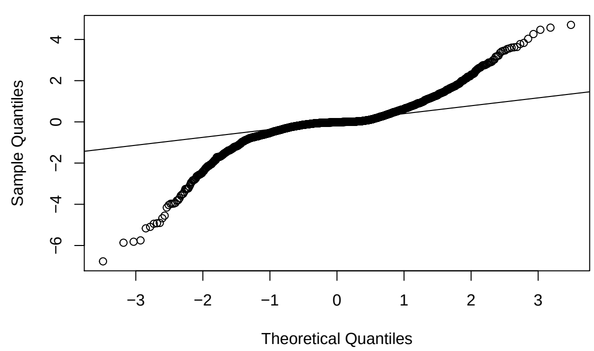

Figure 94: QQ plot of model 8

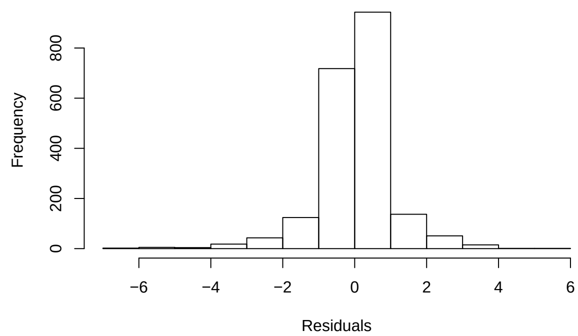

Figure 95: Histogram of the residuals of model 9

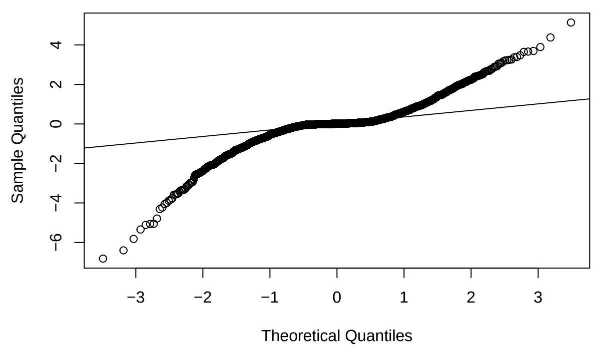

Figure 96: QQ plot of model 9

5 Forecasting and Model Validation

Since none of the models seem to work for this data in terms of their residual plots and ACF, forecasting is done for all models to investigate further which model suits this time series data best. Figures 97 to 105 show the time series of interest, the fitted data and the 72-hour forecast along with the actual data of that 72 hours. The black line is the actual data acquired from the bridge, the dashed red line is the fitted model to the train data, and finally, the dashed green line is the the forecasted data. Almost in all graphs, the first 48-hour is overestimated while the third 24-hour is underestimated. For some of the plots like figures 97 and 99, both minimum and maximum pressures of each of the forecasted days have an increasing trend. On the other hand, some figures like figure 102 and 105 show almost a similar minimum and maximum pressures for the three consecutive days. The actual behavior of the time series is something in between meaning that the minimum pressure of all three days is almost the same, while the maximum pressure shows an increasing trend. All in all, it seems that model 6, ARIMA$(0,2,2)(0,1,1)_{24}$, has the best prediction among all 9 models.

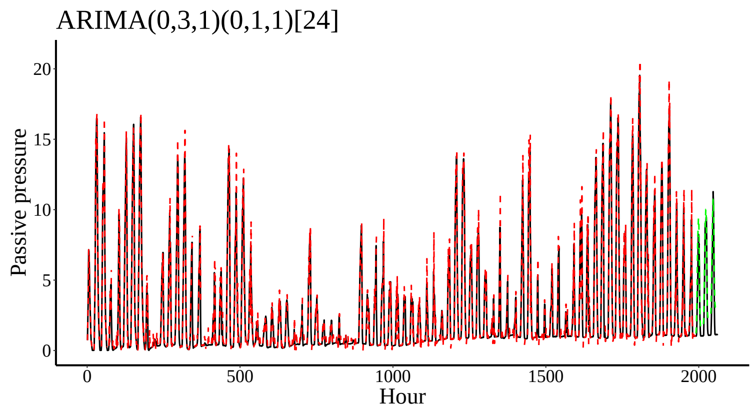

Figure 97: 72-hour forecast based on model 1

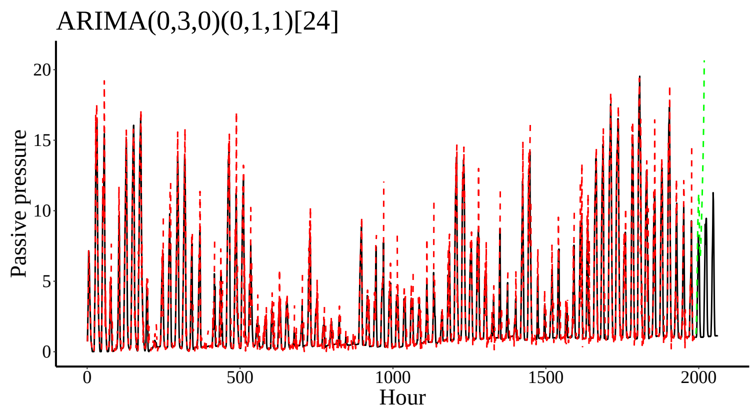

Figure 98: 72-hour forecast based on model 2

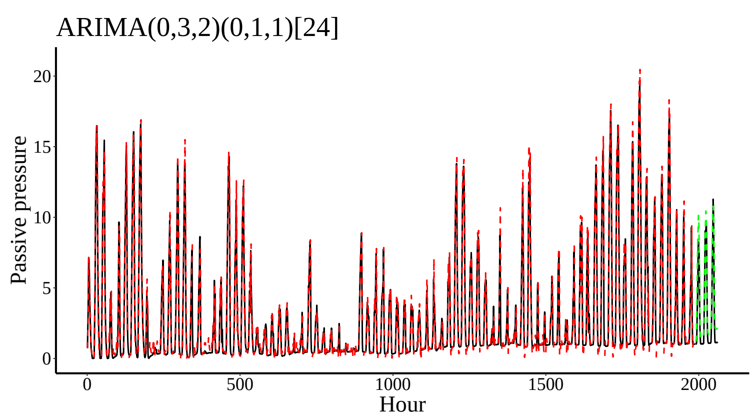

Figure 99: 72-hour forecast based on model 3

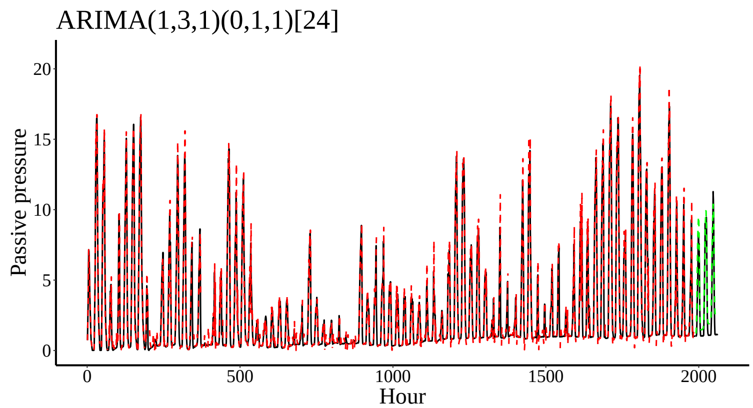

Figure 100: 72-hour forecast based on model 4

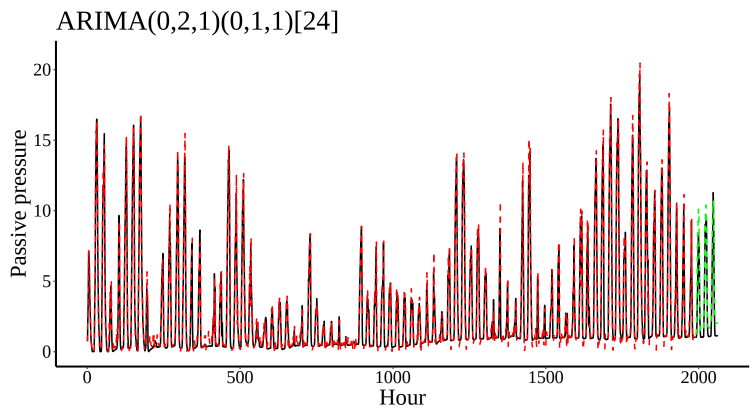

Figure 101: 72-hour forecast based on model 5

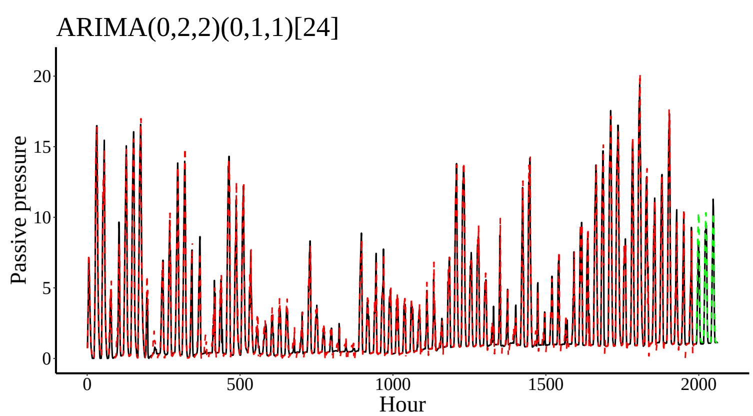

Figure 102: 72-hour forecast based on model 6

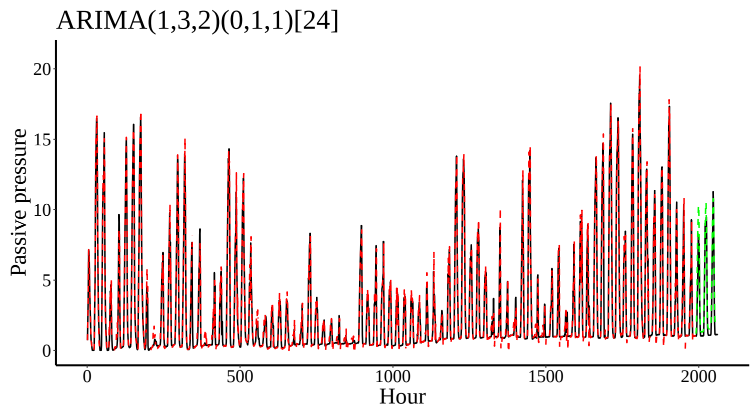

Figure 103: 72-hour forecast based on model 7

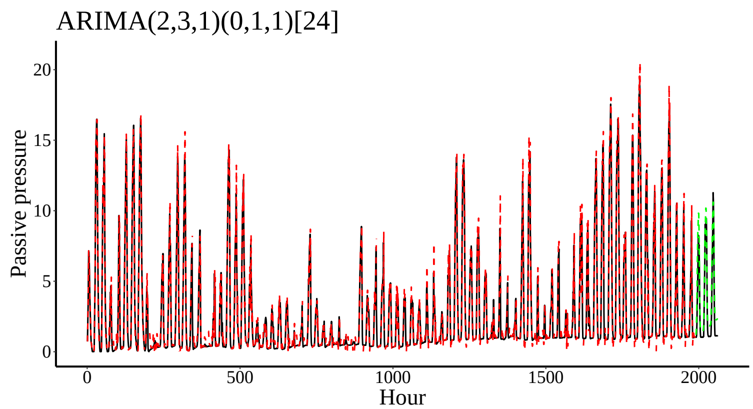

Figure 104: 72-hour forecast based on model 8

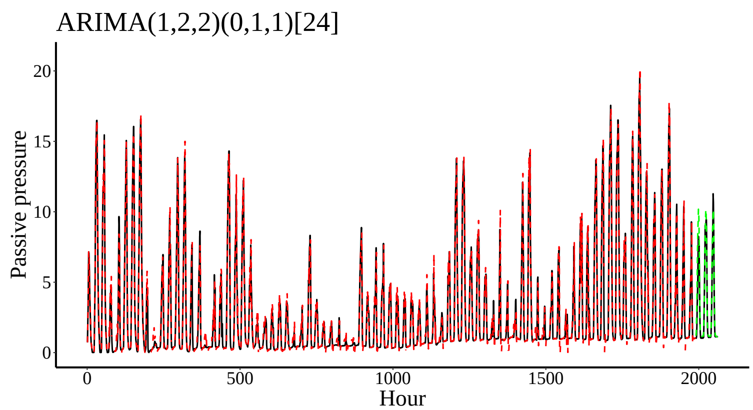

Figure 105: 72-hour forecast based on model 9

6 Conclusion

A pressure cell time series data was analyzed in this work. Different ARIMA models with various parameters were considered to find the best model that represents the times series data. none of the residual plots or ACF/PACF plots are perfect for a time series data set, there are still lags remaining in those plots that are not explained with any data transformation or differencing. Although none of them is perfect, we can still choose the best option among others. Best option can be either based on AIC-BIC criteria or the ACF plots, and since none was perfect, all were considered. Forecasting plots show that model 6 which is an $\text{ARIMA}(0,2,2)(0,1,1)_{24}$ can be considered as the best model representing the pressure cell time series. It seems that model 6 with the smallest AIC/BIC has also the best prediction.