Objective: We are looking at the New Orleans weather high temperature from 1958 to 2015, we want to fit a good SARIMA (Autoregressive integrated moving average) model in order to predict the future observation.

Model selection procedure: after changing the high temperature variable to be a time series object using the ts function. We plot the data (Figure 1.). The best model selection should have the smallest AICc compared to other fitted models. Also, it should have the smallest ME, RMSE …, obtained from the model accuracy metrics. We will look at the Ljung- Box-lack as well to calculate the p-value for the fitted models, the bigger the p-value the better the model is. Moreover, Analyzing the residuals. There should not be any trend in the residuals. Based on ACF and PACF plot of the residuals should look like the white noise.

High temperatures time series plot:

.

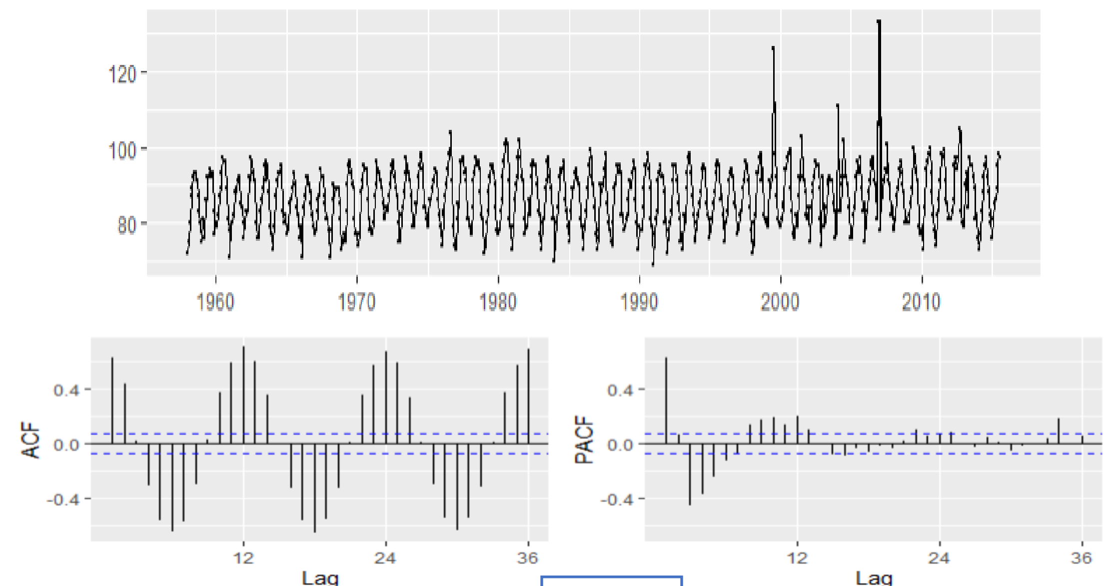


Figure1

The data are clearly stationary, with some seasonality, seen from the ACF and the PACF plot. Also, we can see that the variance is constant meaning that no need for a box cox transformation. Non-seasonal difference is not needed since the time series is stationary.

First, we will take a seasonal difference. The seasonally differenced data are shown in Figure 2

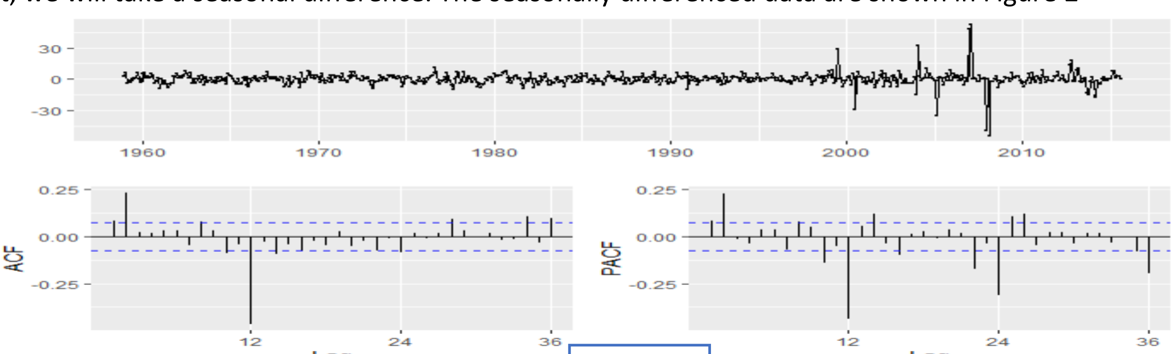


Figure2

Possible ARIMA models:

-Model1: ARIMA (2,0,2) (3,1,1) [12]

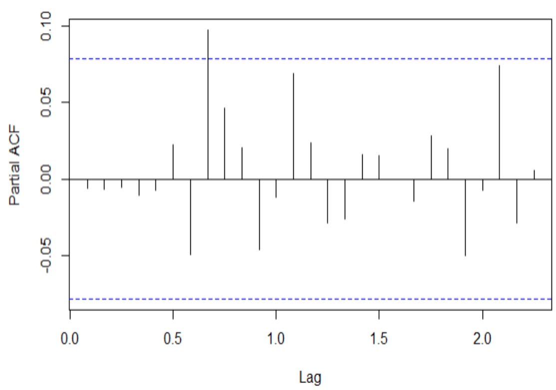
-Model2: ARIMA (3,0,2) (3,1,1) [12]

-Model3: ARIMA (2,0,2) (3,1,0) [12]

-Model4: ARIMA (2,0,1) (3,1,1) [12]

-Model5: ARIMA (2,0,1) (3,1,0) [12]

Our aim now is to find an appropriate ARIMA model based on the ACF and PACF shown in Figure 2. The significant spikes at lag 1 and lag2 in the ACF suggests a non-seasonal MA(2) component, and the significant spike at lag 12 in the ACF suggests a seasonal MA(1) component. Also, The significant spikes at lag 1 and lag2 in the PACF suggests a non-seasonal AR(2) component, and the significant spike at lag 12,24,36 in the PACF suggests a seasonal AR(3)component. Consequently, we begin with an ARIMA(2,0,2)(3,1,1)12 model. The residuals for the fitted model are shown in Figure 3.



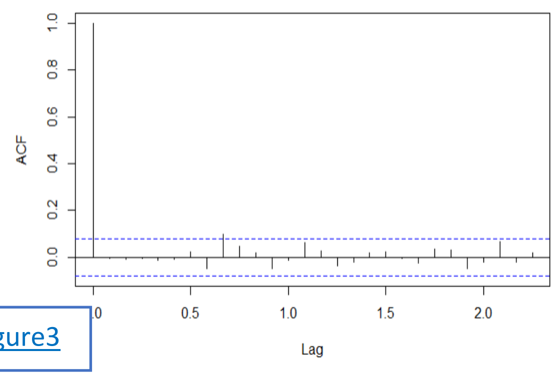


Figure3

Both the ACF and PACF show significant spikes at lag 8, indicating that some additional non-seasonal terms need to be included in the model. The AICc of the ARIMA (2,0,2) (3,1,1)[12] model is -2023.83, while that for the ARIMA(3,0,2)(3,1,1)[12] model is -2025.27. We tried other models with different AR terms and MA terms as well, but none that gave a smaller AICc value(model3, model4,model5…). Consequently, we choose the ARIMA (3,0,2)(3,1,1)[12] model. Its residuals are plotted in Figure 4. All the spikes are now within the significance limits, so the residuals appear to be white noise. The Ljung-Box test also shows that the residuals have no remaining autocorrelations also it is very high for model 2meaning the model does not show a lack of fit, so it fits the data well. We can also look at the Log Likelihood, the higher the better, the highest value was also Model 2. Finally, we will look at the RMSE which is better when it is closer to 0 which is also Model 2.

```
        Box-Ljung test

data:  Model$residuals
X-squared = 10.288, df = 12, p-value = 0.5907
```

Box-Ljung p-value for Model2: ARIMA (3,0,2) (3,1,1) [12]

ACF and PACF for the selected model Model2: ARIMA (3,0,2) (3,1,1) [12]

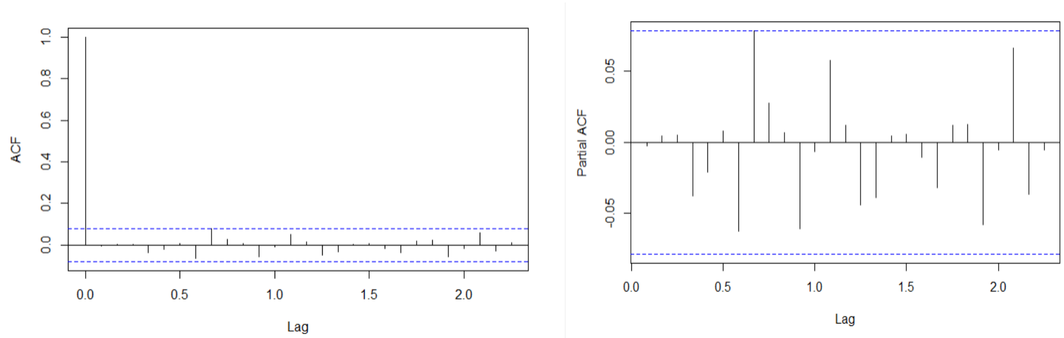


Figure4

Forcasting:

Thus, we now have a seasonal ARIMA model that passes the required checks and is ready for forecasting. Forecasts for the model from 2011 to 2015 are shown in Figure 5.

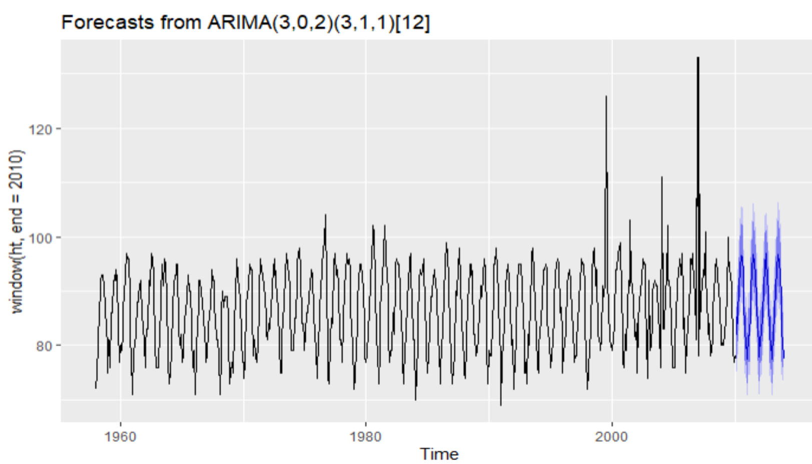


Figure5