Time Series Lab 2

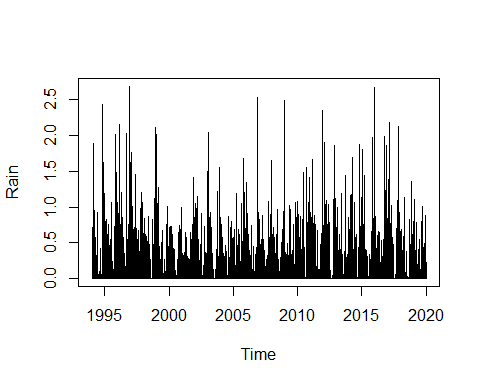
Subha Varadarajan

11/21/2020

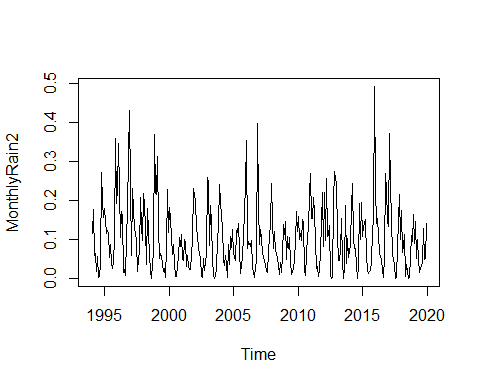
**Create a plot of the time series**

The below two plots refers to the time series plot.

Plot 1:



Plot 2:



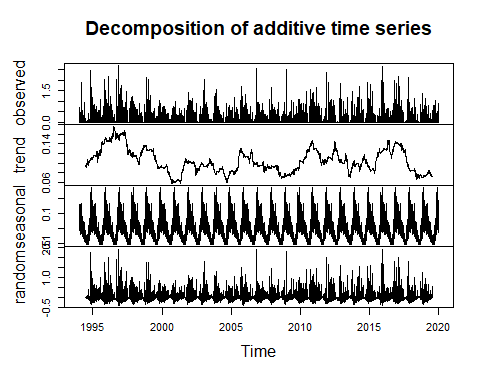
**Adjust for the seasonality issues by collapsing the data into monthly averages**

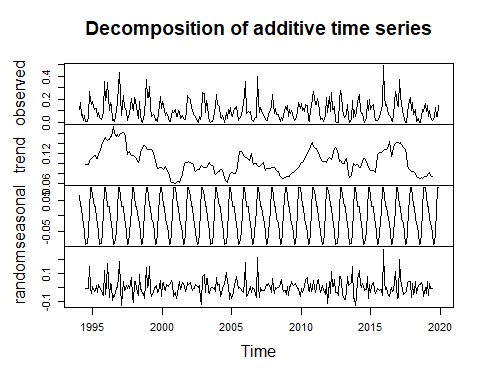
Text

Description automatically generated

**Decompose Model:**

Identify the trend/season/random components We will be using decompose function to decompose the dataset. Plot1: Represents the decomposition of full data Plot2: Represents the decomposition of the monthly data Both the plots do not have any trend but the seasonal variations are constant. We do have some irregularity.There are some random spikes

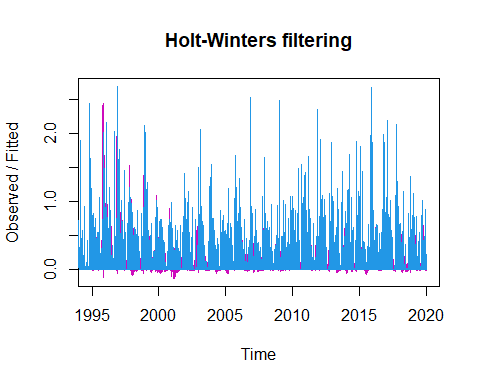




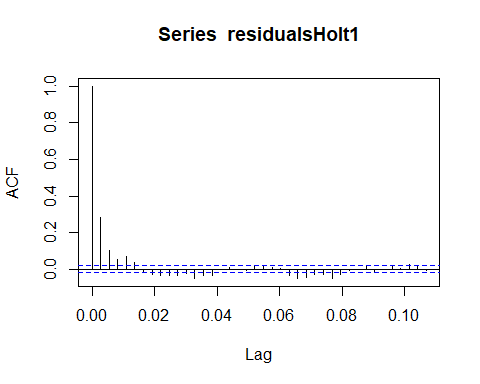
**Modeling using exponential smoothing – Full data:**

It compares the forecasted value with the actual value and takes difference with the factor

The actual data and predicted data are plotted in the same plot. The Predicted model (pink) does not have much higher spikes. It has only few. It has negative values too as it predicts based on the coefficient of the formula and hence the negative values.

Chart

Description automatically generated

Chart

Description automatically generated

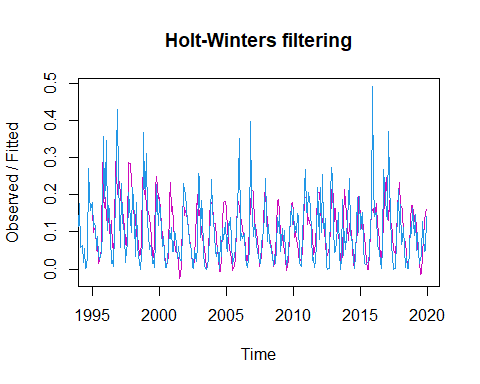
**Residuals:** Residuals have been calculated and plotted. Residuals should be evenly spaced around the mean.We do not want too high or too low.Since it has more high spikes, the model underpredicts (more residuals at the top part). This may be due to smoothing of the spikes. In the ACF and PACF plot we can clearly confirm seasonality.

From the above observation of plots it is clear that the model using exponential smoothing for full data underpredicts. Thus, it is not an accurate model.

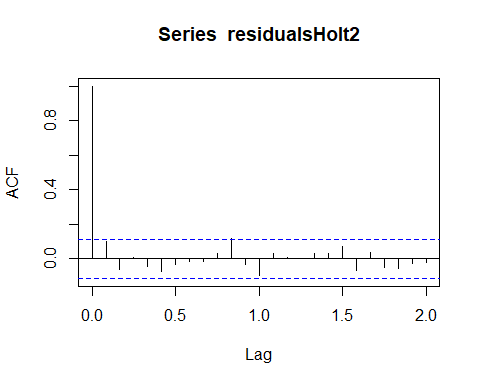
**Modeling using exponential smoothing - Monthly data:**

As the exponential smoothing model for whole data was not satisfactory, we now perform the exponential smoothing for the monthly data.

On contrast with the full data model, the difference between actual and predictive is much lower. The predictive spikes up in most of the cases but still, it underpredicts as it hits spikes in 5 places since we used the smoothing model. Yet, the prediction is better than the full data model (it underpredicts not as much as in the full data)

Chart, histogram

Description automatically generated

Chart, box and whisker chart

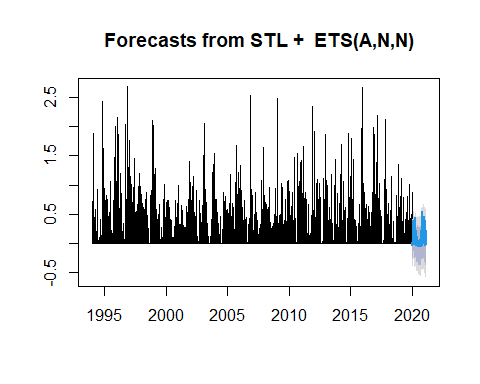
Description automatically generated

**Residuals:** In Residual plot, the residuals are more evenly spaced around the zero line when compared to the full data model. Though they are underpredicting but they arent as bad as the full data model.It may be because of the less data points (monthly average is easier to predict than daily). In ACF there is only one line above the blue line while in PACF, there are no lines above the blue line ( All the lags are within the blue line) which means this model is really good. Seasonality can also be seen.

Hence, compared to the above full data model monthly data model is more accurate.

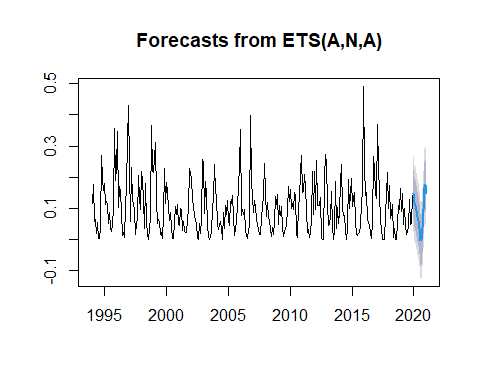
**Forecasting using exponential smoothing - Full Data**

We are using a full year data for forecasting.



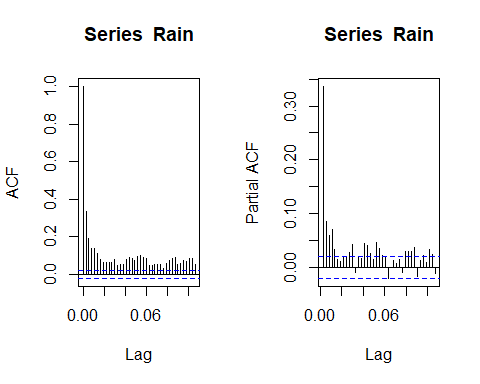
In prediction, we get a seasonal pattern. The forecast shows a good pattern of rain. Though it does not have huge spikes to let us know how much it is going to rain, but from this prediction we can figure out when it is going to rain and when it does not.

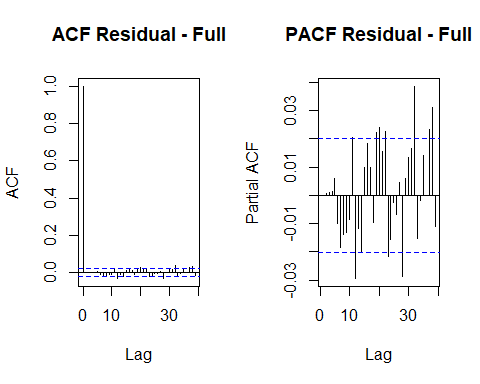
**Forecasting using exponential smoothing - Monthly Data**



In monthly model, we can see less chances of rain where the prediction dips. Though we cannot predict on which particular day but we can predict when there will be no rain.The dip is going to be when there will be no rain The error bars are not super wide. Which makes it a much better model.

**Modeling using an auto.arima model - Full Data**



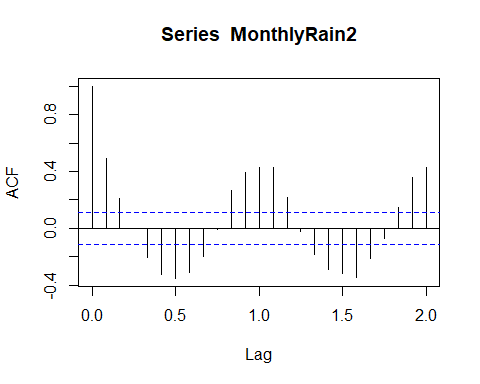


In Daily data, we cannot see daily seasonality.

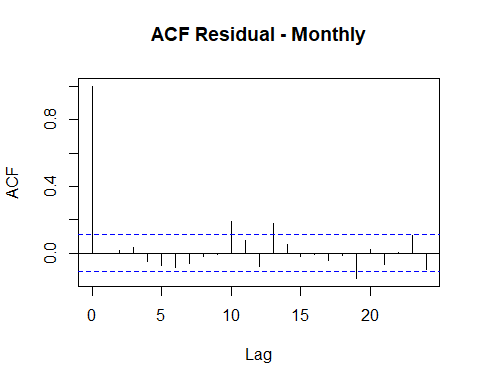
From the ACF plot we can clearly find majority of the lags are within the blue line.While in PACF plot, we can see few spikes above the blue line.

In daily data, we are unable to see daily seasonality in the ACF and PACF plots. The lags are not seasonal.

**Modeling using an auto.arima model - Monthly Data**

Chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

In Monthly data, auto arima model, the lags are lesser and we can find seasonality when compared to the full data (daily data). In the output, we can see a seasonal component (2,0,0) and also a trend (we have mean in the output).

The AIC value for monthly data is much higher than the daily data. Therefore, we can get precise prediction when we use daily data (full data).

**Forecasting:**

prediction1=predict(RainArima,n.ahead=7000)

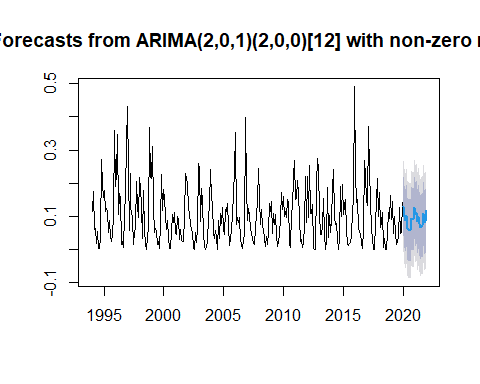
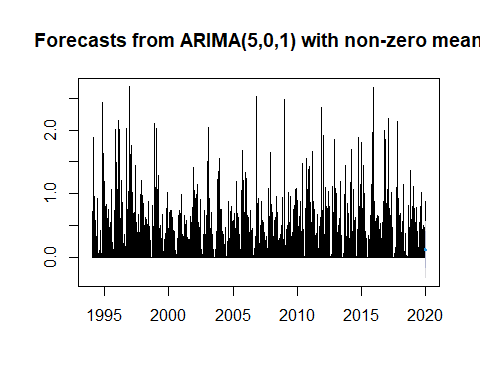
prediction1

plot(forecast(RainArima,h=24))

prediction2=predict(RainArima2,n.ahead=7000)

prediction2

plot(forecast(RainArima2,h=24))



When comparing the Holtwinters and Auto Arima model, Holtwinters monthly data seems to be a better one.