Assignment-1

Group 22

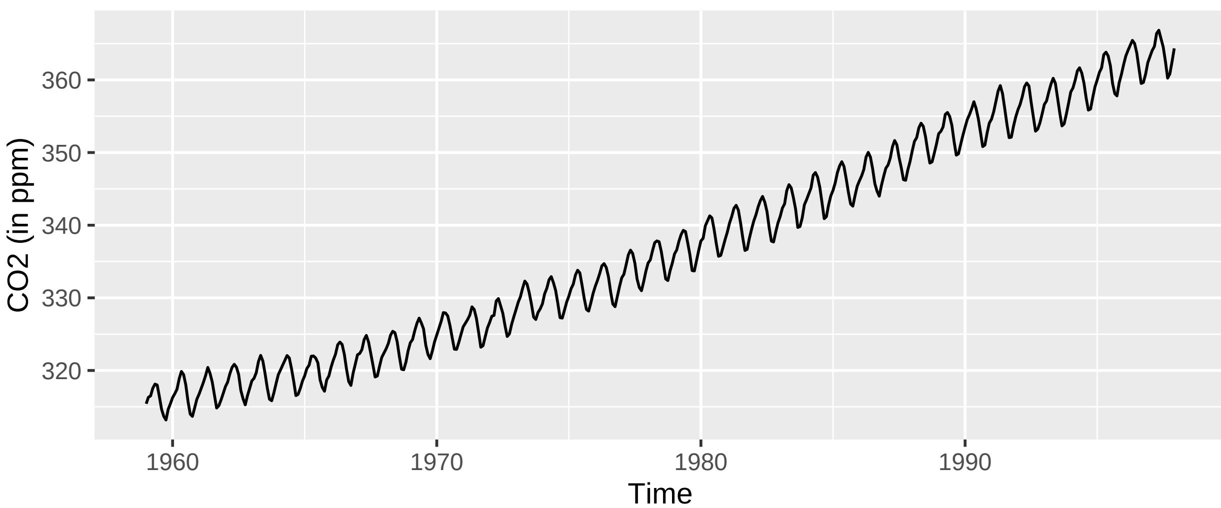
Imanbayeva Sofya, Mazzi Lapo, Piras Mattia, Srivastava Dev

2023-03-01

Exercise 1

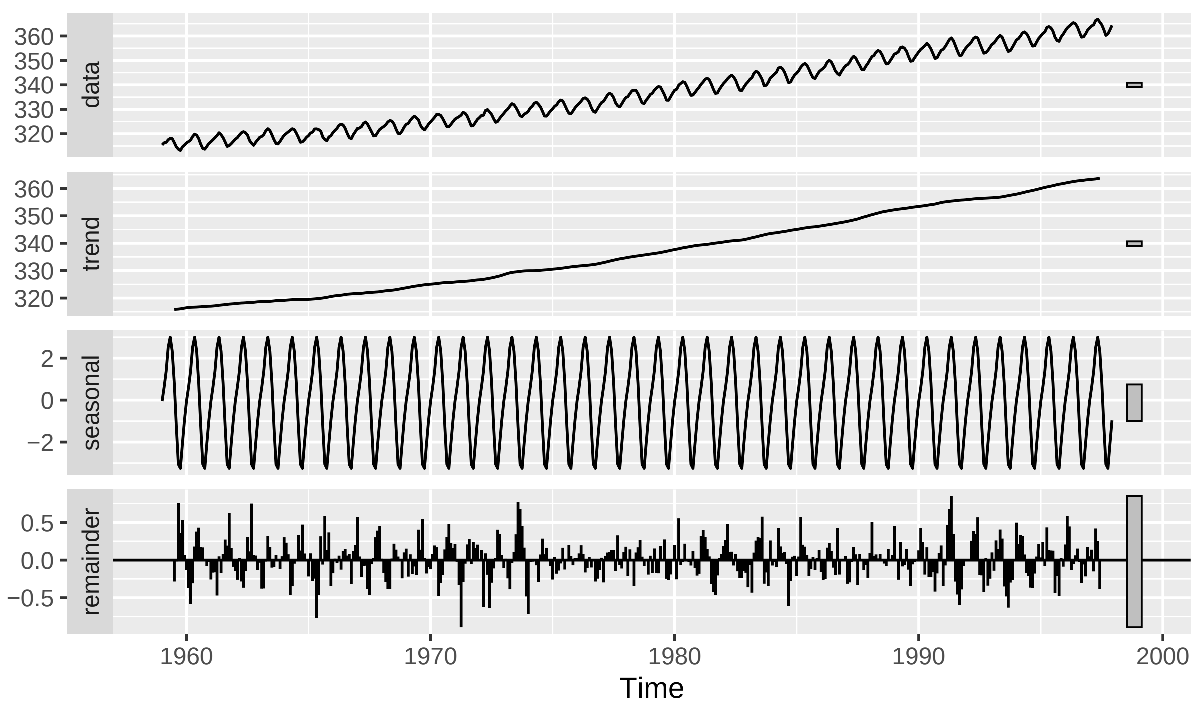
Figure 1 shows the concentration of CO2 (in parts per million) in the atmosphere between 1959 and 1997.

Figure 1: Carbon Dioxide Atmospheric Concentration 1959−1997



As the seasonal component seems to be constant, we can do an additive decomposition whose plot is:

Figure 2: CO2 Time Series Decompostion (ppm)

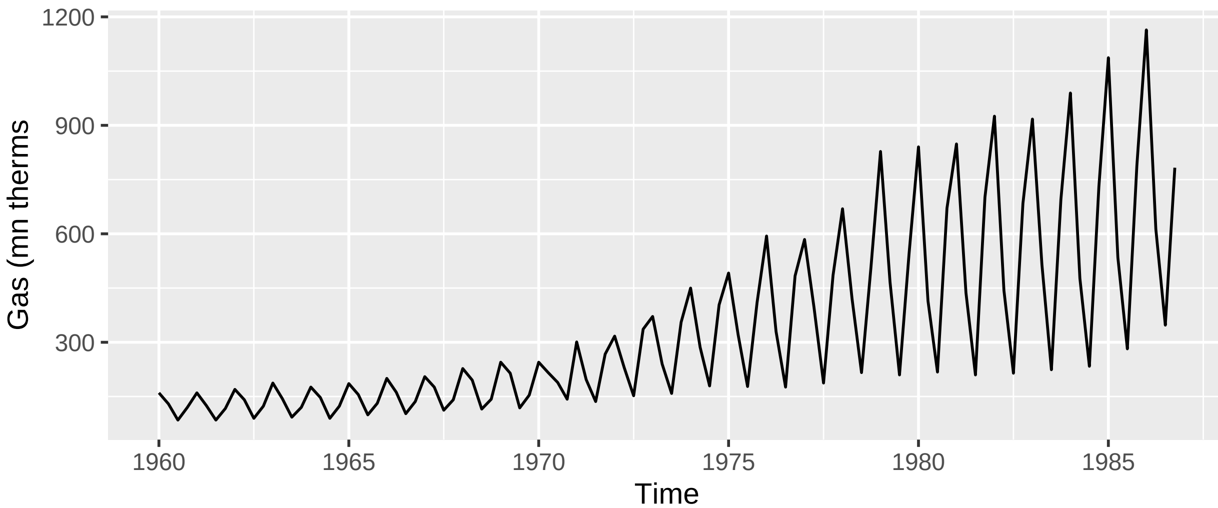


By looking at the decomposition of the time series (Figure 2) we can clearly see that there is a linear increasing trend, while seasonality is regular and its magnitude doesn't change over time.

Exercise 2

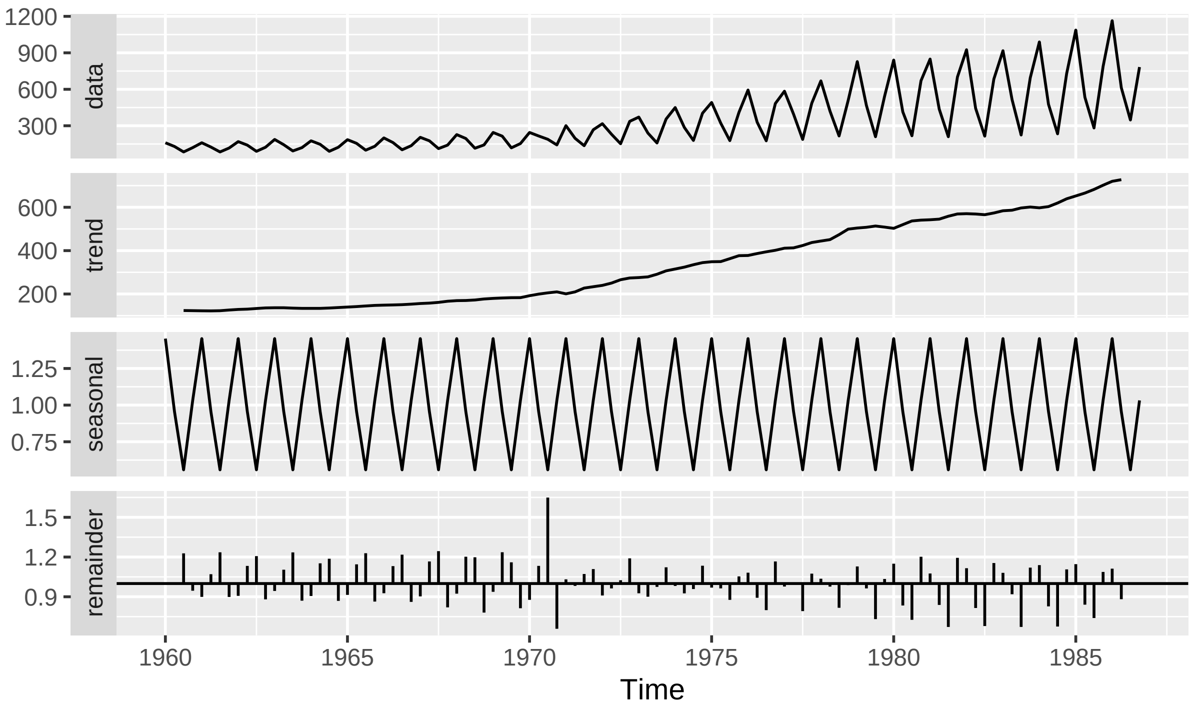
Below is the time series plot for the quarterly UK gas consumption from the first quarter of 1960 to the forth quarter of 1986, in millions of therms.

Figure 3: Quarterly UK Gas Consumption (in mn therms)



As we can see from the plotted data (Figure 3), there is a change in the seasonal component from the begin of the 70's. At the same time there is also a change in the trend, which becomes steeper around the same year. In this case, a multiplicative decomposition is the more appropriate approach, since we do not have a constant seasonal component.

Figure 4: Multiplicative Decomposition of Gas Consumption



Classical time decomposition models assume constant seasonal effect and tend to over-smooth sharp increases and falls in the data. It seems that in the 1970s there was a change in UK gas consumption patterns. Due to that change in the seasonal component, the multiplicative decomposition over-smoothed the sharp increase in data which is reflected in the residuals. Additive decomposition showed even worse results. One of the possible solutions to improve the results might be to use alternative methods that allow for slight change in the seasonal component, for example X11 decomposition.

If we take the log of the time series we can see that the heteroskdasticity of the data diminishes. By using log(UKgas) we can use an additive time series decomposition and still get a good representation of the trend and seasonal components.

Figure 5: Log Quarterly UK Gas Consumption

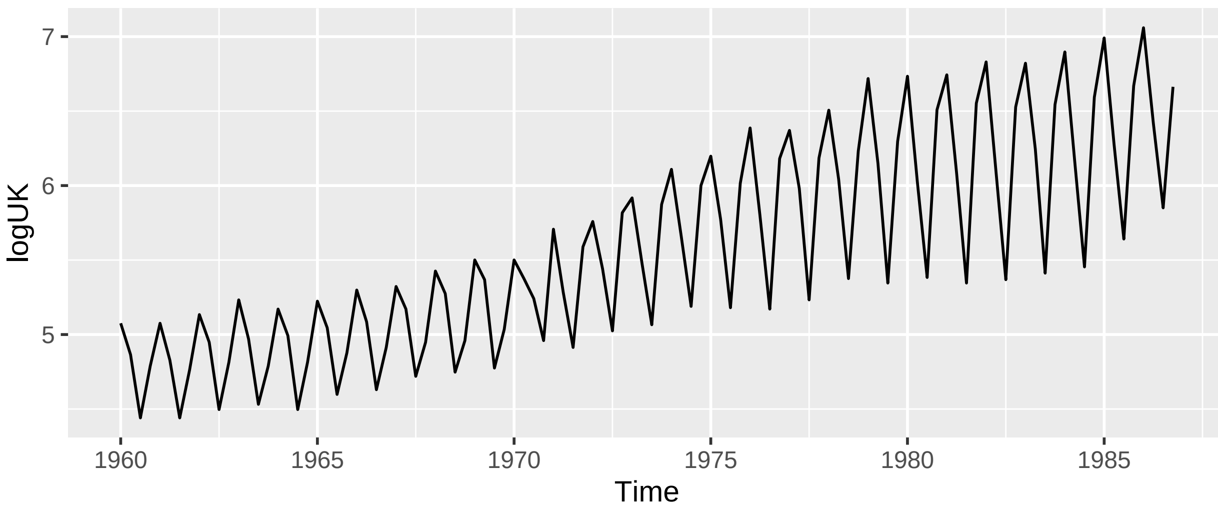
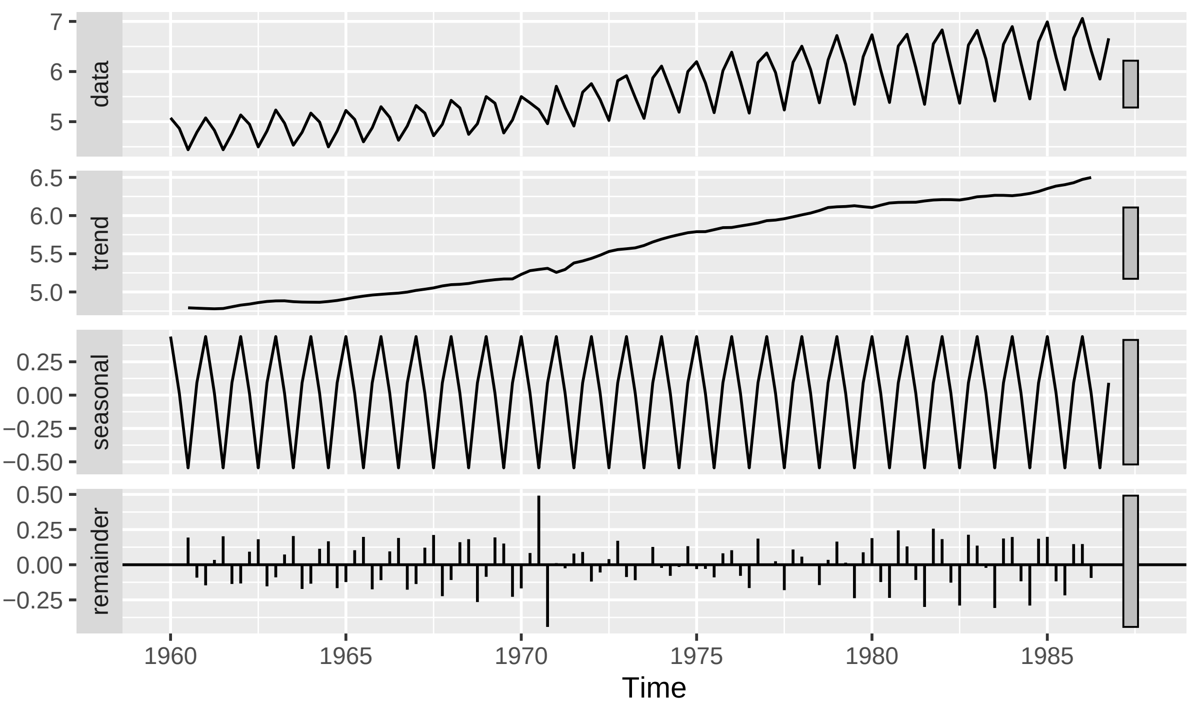
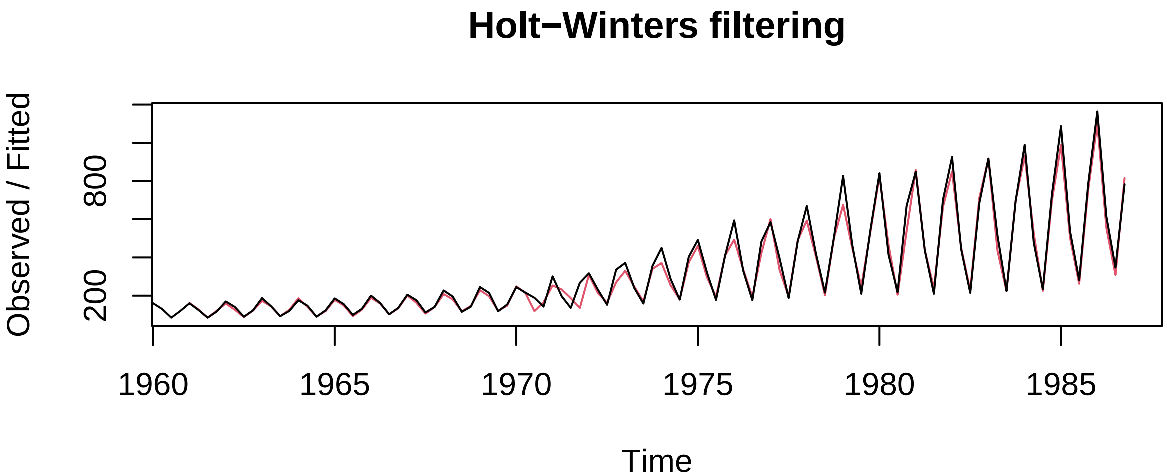


Figure 6: Additive Decomposition of Log Gas Consumption



We can try to create a forecast of our time series and compare it with the original dataset. Using Holt-Winters exponential smoothening we get

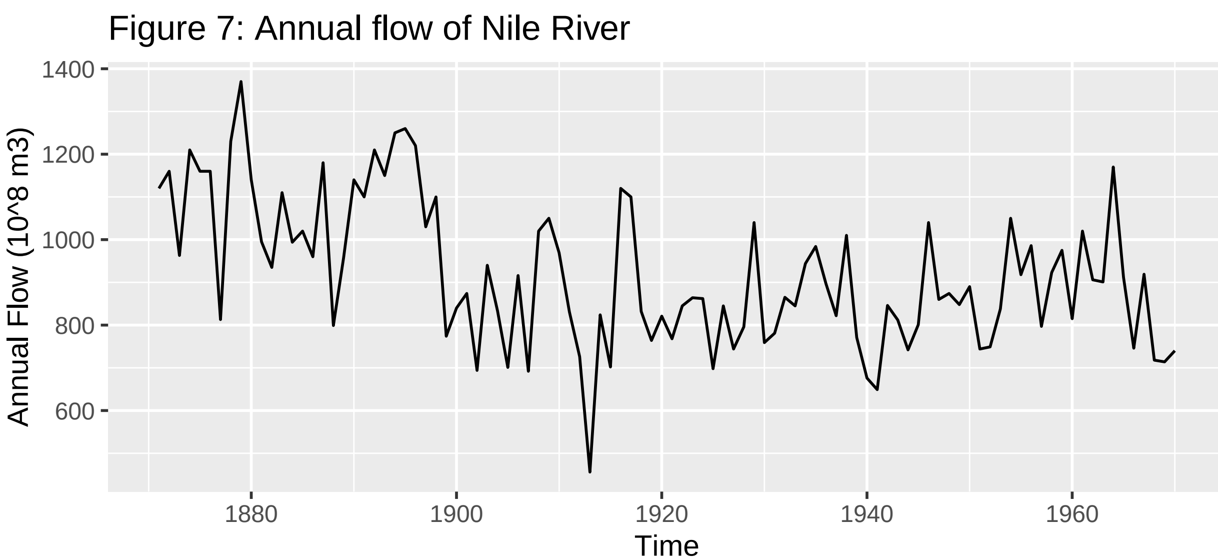


From the plot comparison, it seems that H&W is a good predictor of the raw data. However, we can get a more accurate value of the predictive performance by using the MAPE.

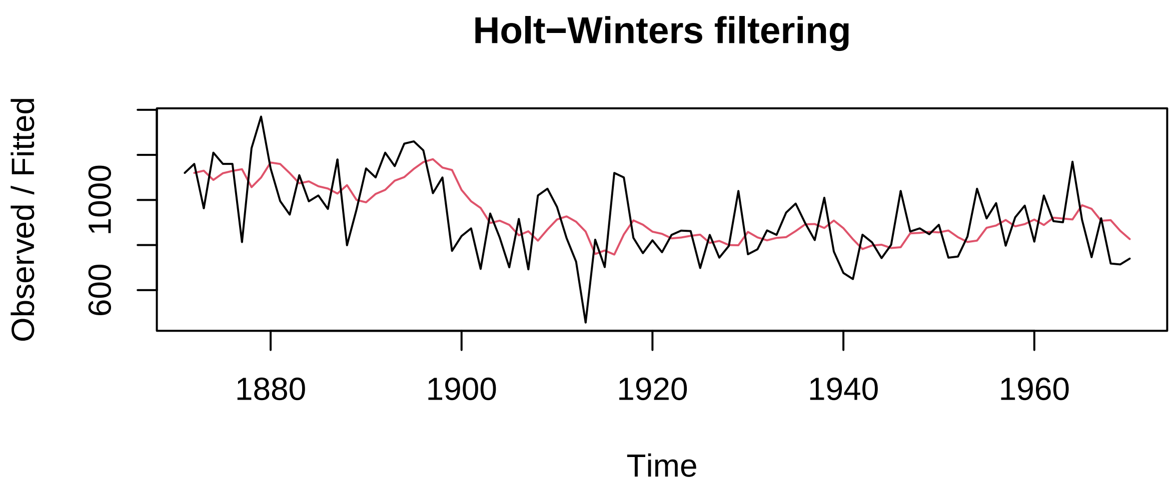
Using MAPE, we find that the predictive performance of the H&W estimates is poor: on average the predictions are 42.45% away from the target.

Exercise 3

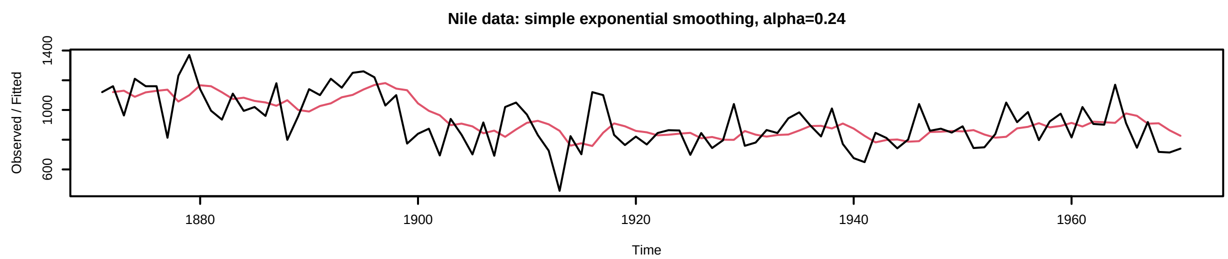
Here is the time series plot for the annual flow of the river Nile at Ashwan, for the period 1871-1970

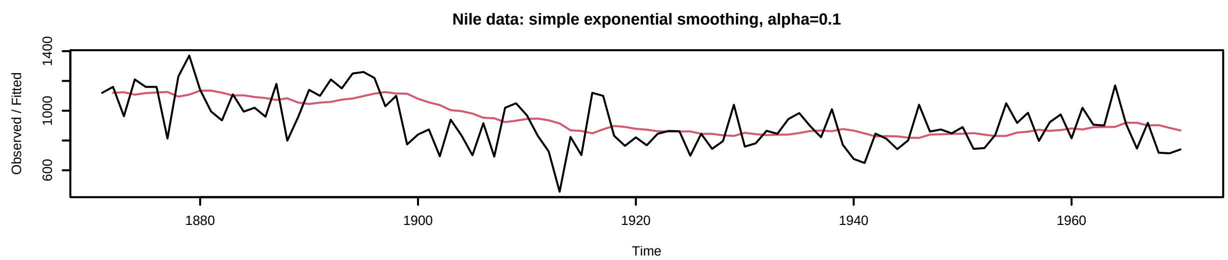


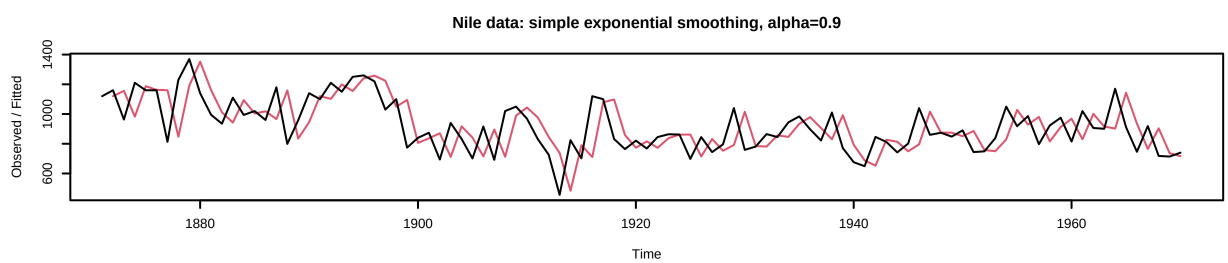
Using Holt-Winters exponential smoothing we get the following plot. The value of α chosen (by default) is 0.24



Now doing a comparison with $\alpha = 0$ and $\alpha = 1$ we get the following plots.







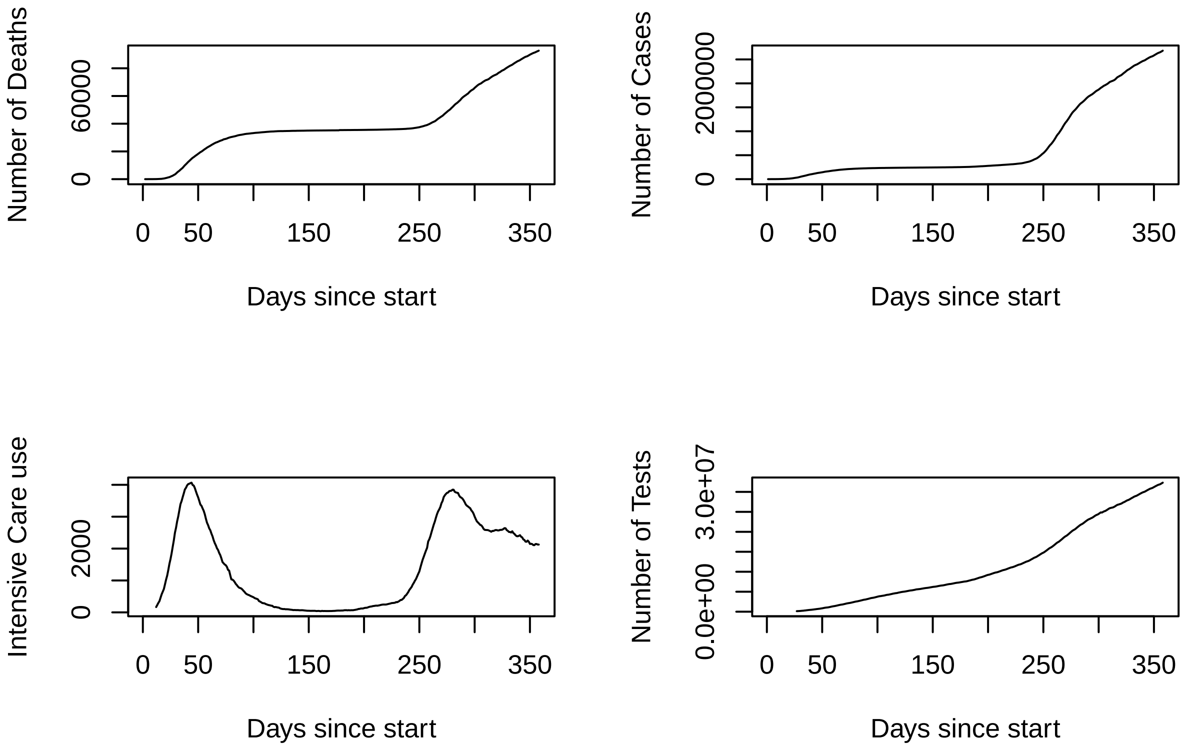
We can see from the graphs that as α approaches 1, the estimated value converges to the actual data. If instead α approaches 0, the fitted values converge to a constant.

For $\alpha = 0.24$, MAPE is 0.1307089 $\alpha = 0.1$, MAPE is 0.1348183 $\alpha = 0.9$, MAPE is 0.144959

By looking at the 3 MAPE we can see that the HW with α=0.24 is the one with the highest predicting performance, with predictions that are on average 13.07% away from the actual values.

Exercise 4

Here are some of the time series plots derived from the coronavirus dataset for Italy.


Number of Deaths
0
60000
0
50
150
250
350
Days since start
Number of Cases
0
2000000
0
50
150
250
350
Days since start
Intensive Care use
0
2000
0
50
150
250
350
Days since start
Number of Tests
0.0e+00
3.0e+07
0
50
150
250
350
Days since start