MATH58012 REPORT

Mohammed Fahaidur Rahman

November 2022

1 Introduction

X is a time series which holds the data on the water level of a pond, in rural Hampshire, in feet over the course of 50 years with 600 data points. The data was recorded at the beginning of each month. In this report I conclude an AR (2) model process is best to model the time series with a seasonal component $s(t)$ and white noise ε_t , therefore $X_t = s\,(t) + 0.668X_{t-1} + 0.260X_{t-2} + \varepsilon_t$. The plot of X is in Figure 1.

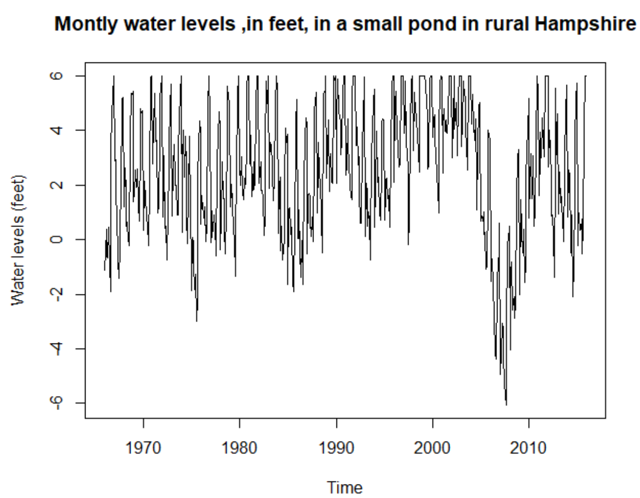

Figure 1: Monthly water levels, in feet, in a small pond in rural Hampshire

2 Main Report

There seems to be aseasonal effect as there are rhythmic fluctuations that seem to be repeated. This is also shown in Figure 2 which is the ACF of X as there is high autocorrelation of lag 1 and lag 2. There seems to be no trend as the values of the water levels seem to fluctuate around the value of 2 and this has been shown by fitting a linear in Figure 3. As you can see the linear trend is just a horizontal line with the y intercept at 2.5175.

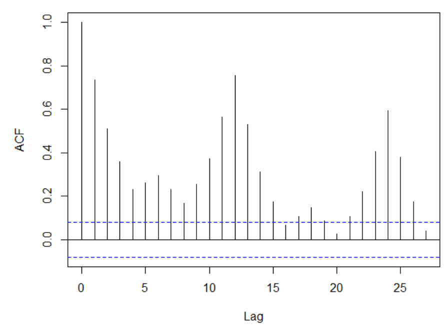

Figure 2: acf of X

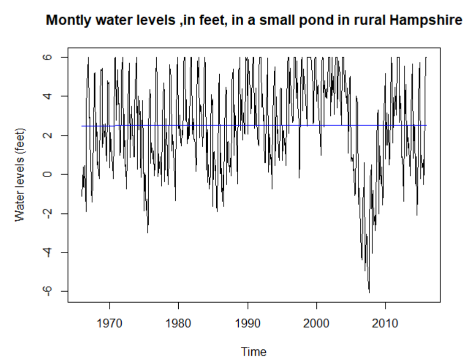

Figure 3: Monthly water levels, in feet, in a small pond in rural Hampshire with a linear trend fitted on

Therefore no linear trend will be removed to find a suitable model for X. The seasonal effect was removed by creating dummy variables for each month and residual variable was created which was called Y. The plot can be seen in Figure 4.

To determine if an AR process or MA process is suitable to describe Y the ACF and partial ACF (PACF) were plotted and this can be seen in Figure 5. Looking at the ACF of Y we can see damped oscillations suggesting an AR(p) model where p>1. Looking at the PACF of Y we can see that it cuts off at lag 2 suggesting that the model is AR(2). instead of an exponential decay or

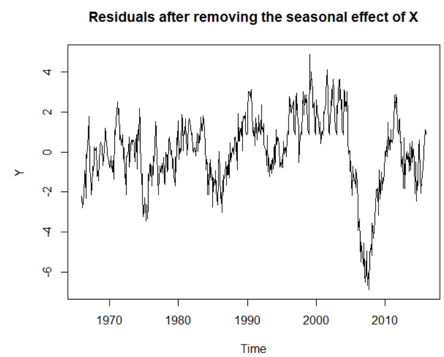

Figure 4: Plot of Residuals after removing seasonal effect from X

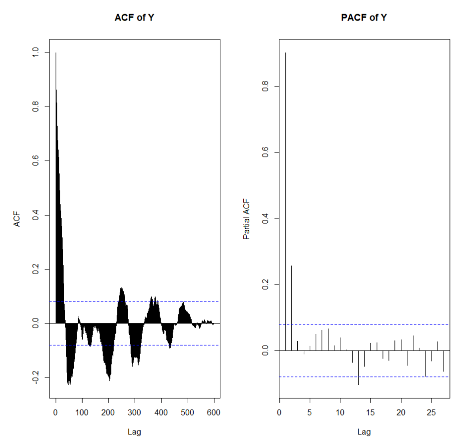

Figure 5: ACF and PACF of Y

damped oscillations so it seems that an AR process might be the best process to explain the structure present in process Y.

Next in the investigation AR(1), AR(2), and AR(3) of the time series Y were fitted. Then the residuals of these processes were plotted with their corresponding correlograms as can be seen in Figure 6.

Looking at the correlograms for AR(1), AR(2) and AR(3) it seems that the best model is AR(2) as the correlogram for AR(1) does not cut off at lag 0 and AR(2) is the first model where the correlogram cuts off after lag 0. Also looking at the AR(2) residuals the graph also looks like white noise which supports the decision to pick the AR(2) process. The residuals of the AR(2) model will be now referred to as Z.

The final part of the investigation was plotting the periodograms of X,Y and Z which are shown in Figure 7.

Looking at the periodogram of Z it looks very similar to the periodogram of white noise which further supports the AR(2) process being chosen.

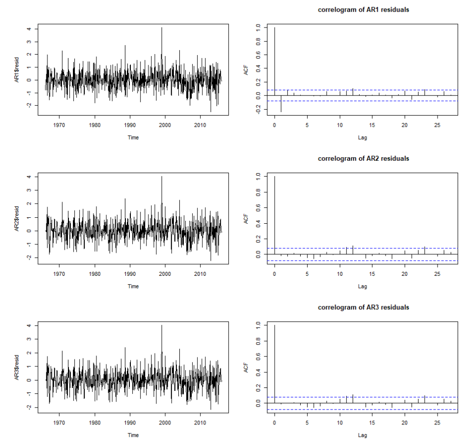

Figure 6: Residuals of AR(1),AR(2) and AR(3) and their corresponding correlograms

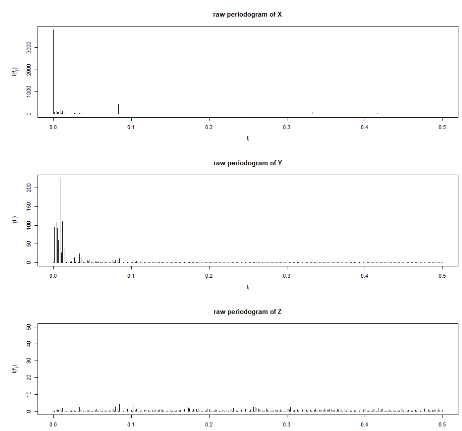

Figure 7: Periodograms of X,Y and Z

2.1 Summary

Finally, here we will summarise our findings of X. There appeared to be no trend but there was a seasonal component, $s(t)$ which will be denoted in our model of X. The seasonal effect describes the effect of each month on the data and this can be equated as:

$s(t) = 1.1217\delta_{t,0} + 2.3694\delta_{t,1} + 2.8675\delta_{t,2} + 1.7260\delta_{t,3} + 1.6127\delta_{t,4} + 1.3839\delta_{t,5} + 0.3643\delta_{t,6} + 0.4093\delta_{t,7} + 3.3945\delta_{t,8} + 4.8671\delta_{t,9} + 4.9617\delta_{t,10} + 5.1324\delta_{t,11}$

By combining all of the findings above the best model of the time series X is using an AR(2) process with a seasonal component above. When computing the ACF of AR(2) the coefficient are 0.6678 and 0.2596. The AR(3) ACF was computed also and the coefficients were 0.6601,0.22398 and 0.02967. As the 3rd coefficient for the AR(3) ACF is much smaller than the 2nd coefficient it can be taken as insignificant and the 2nd coefficient of AR(2) and AR(3) are very similar, which further suggests AR(2) process should be chosen as it is a simpler model If we look at the plots of AR(2) residuals in Figure 6 the plot looked like white noise which can be denoted as ϵ_t. Therefore the model of the time series, X is : $X_t = s(t) + 0.6678X_{t-1} + 0.2596X_{t-2} + \epsilon_t$ where $s(t)$ has been described above.