Time Series Coursework

Short summary

This report research on the time series data of water levels, which were recorded in feet on the first day monthly between 1966 and 2015, in tiny pond in countryside of Hampshire. There were both a linear trend and seasonal effect in the data. After removing these effects, it will explain why the data is appropriate to AR model with p value equaling to two among one to three values. As a MATH5802M student, this report will describe further analysis on periodograms of the result of raw pond's data defined as X_t, the residual after removing trend and seasonal effect assigned as Y_t, and the last residual which may be a white noise set as Z_t. All processes are performed in R program.

Removing trend and seasonal effect

In order to model time series from pond data, the first step of the process is to import the data into R and plot the pond time series data showing in the Figure 1

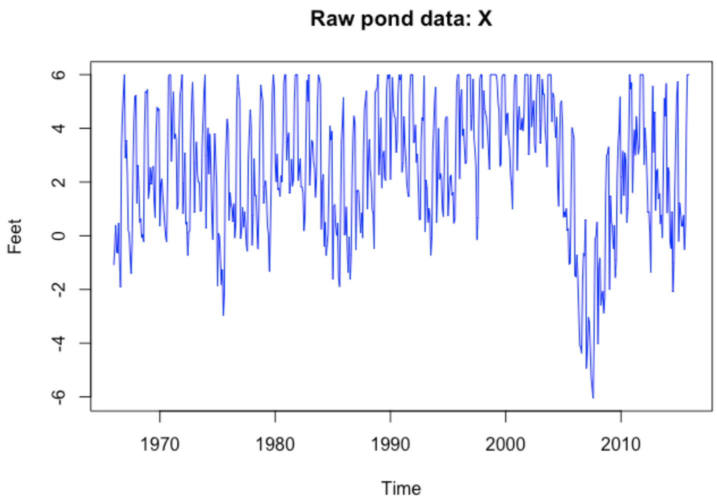

Figure 1: Raw pond's data (X)

After examining the data in Figure 1, the common information of the data is maximum value (6 feet), minimum value (-6.05 feet), mean (2.52 feet), variance (6.14 $feet^2$) and the number of X_t accounting for 600. Moreover, the graph may contain the linear trend and seasonal effect therefore the second approach is to eliminate the trend via fitting the linear model using `lm()` function by inputting "$X_t \sim x + x^2$" to create the trend. As a result, it provides little rising trend which is almost flat line (equation 2) representing in Figure 2, and then utilize this line to remove the bias values from the raw data showing as Figure 3 by these equations:

$$Residual1 = X_t - u(t) \tag{1}$$

Where u(t) is trend and defined as

$$u(t) = 2.497533 + 0.000066 * t \tag{2}$$

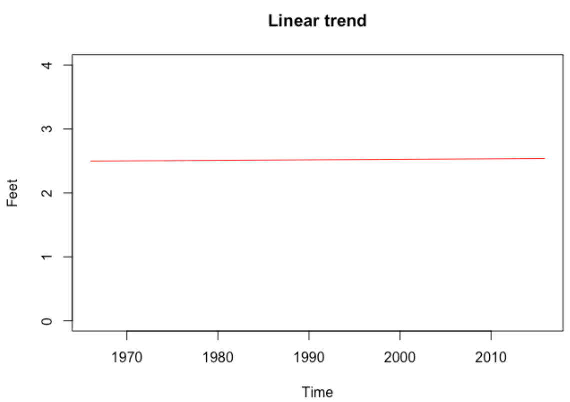

Figure 2: Linear trend

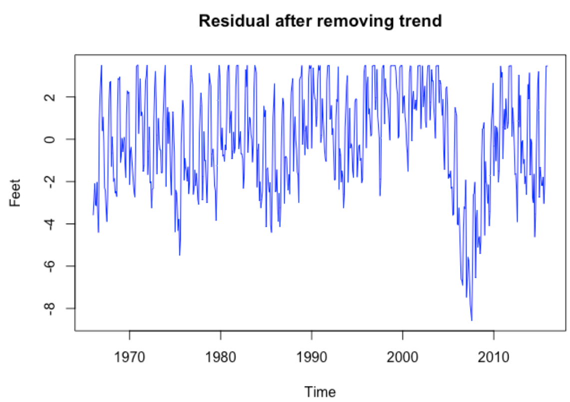

Figure 3: Residual after removing trend (Residual1)

According to Figure 3, the graph seems to contain the seasonal effect because there is a wild fluctuation in the same pattern yearly. This might be because it contains seasonal effect. To determine this effect, the ACF plot is needed and created as Figure 4.

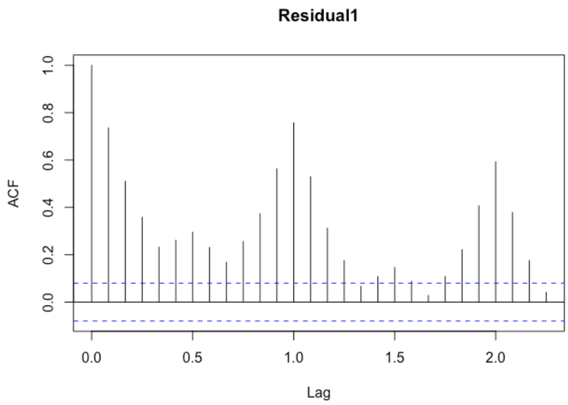

Figure 4: ACF plot of Residual1

Looking at Figure 4, it's obviously seen that the graph comprises seasonal in every year. Therefore, `lm()` function is also adopted in this method by separating twelve month of the year to be variables and fitting via "`Residual1 ~ 0 + jan + feb + mar + apr + may + jun + jul + aug + sep + oct + nov + dec`". Noted that, for jan ($\delta_{t,0}$) to dec ($\delta_{t,11}$) variables, each of those will equal to 1 if X_t is their month such as jan ($\delta_{t,0}$) = 1 if X_t is in January showing as equation 4. After fitting the seasonality, it provides seasonal effect values. In order to examine the seasonal effect easily, the period (x-axis) in Figure 5 is featured just between the year 1995 and 2000 but the other period has also the same pattern and values of this period. In conclusion of this process, the residual after removing trend and seasonal effect (Y_t) illustrating in Figure 6 is followed by this equation:

$$Y(t) = Residual1 - s(t) \tag{3}$$

Where s(t) is seasonal effect and defined as

$$\begin{aligned} s(t) = &- 1.395 * \delta_{t,0} - 0.148 * \delta_{t,1} + 0.35 * \delta_{t,2} - 0.791 * \delta_{t,3} \\ &- 0.905 * \delta_{t,4} - 1.134 * \delta_{t,5} - 2.153 * \delta_{t,6} - 2.108 * \delta_{t,7} \\ &+ 0.877 * \delta_{t,8} + 2.349 * \delta_{t,9} + 2.444 * \delta_{t,10} + 2.614 * \delta_{t,11} \end{aligned} \tag{4}$$

Note that: $\delta_{t,i}$ = 1 if t is in month i, and = 0 for else

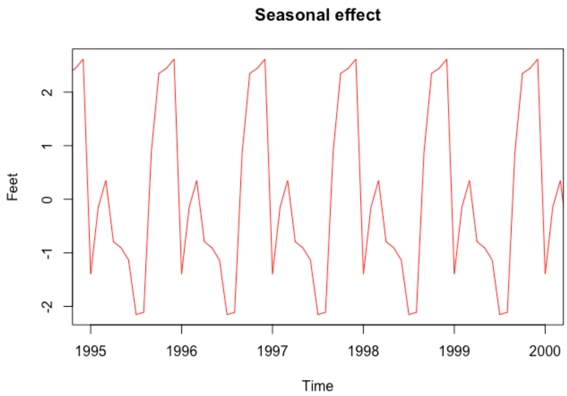

Figure 5: Seasonal effect

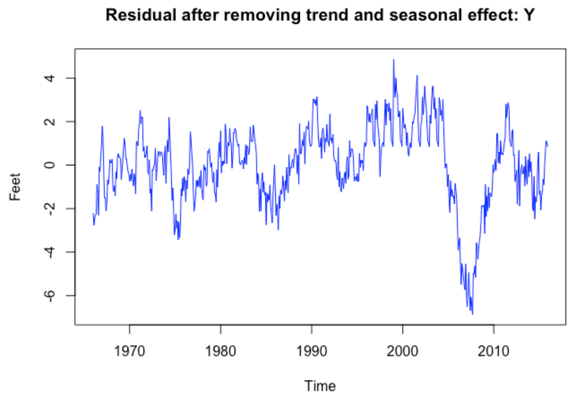

Figure 6: Residual after removing trend and seasonal effect (Y)

Choosing MA or AR model

To determine an appropriate model between MA and AR, the first step is to plot auto-correlation of time series Y whether an ACF plot (Figure 7) or a PACF plot (Figure 8).

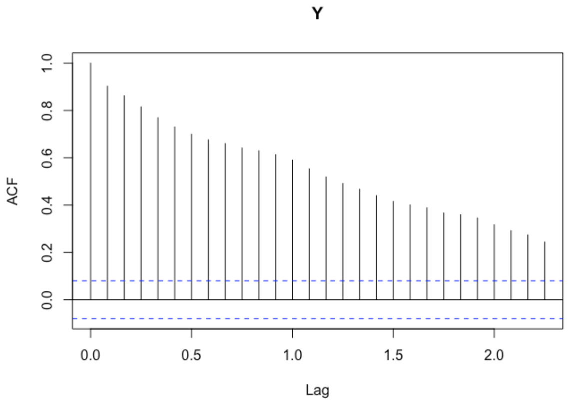

Figure 7: ACF plot of time series Y

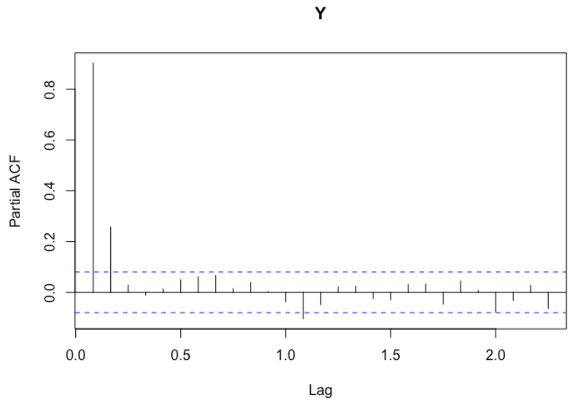

Figure 8: PACF plot of time series Y

In Figure 7, it seems that there are some correlations between time steps (lag) even in the long preceding time. In another word, there is no cut-off at any lag. For Figure 8, it's clearly seen that the PACF is cut off at lag 2 so it's highly likely to be AR(2) model.

<u>Fitting AR model</u>

For AR(p) model which p equals to 1, 2 and 3, this report uses *`ar()`* function and inputting the essential parameters to compute the Coefficients and sigma^2 of residual by the Yule-Walker equations, for example the coding line is *`ar(Y, method = 'yule-walker', aic = FALSE, order = 1)`* for AR(1). Consequently, the fitting parameters of AR model namely

- For AR(1): Coefficient 1 = 0.9025, and Sigma^2 of residual = 0.6292
- For AR(2): Coefficient 1 = 0.6708, Coefficient 2 = 0.2568, and Sigma^2 of residual = 0.5887
- For AR(3): Coefficient 1 = 0.6634, Coefficient 2 = 0.2374, Coefficient 3 = 0.0288, and Sigma^2 of residual = 0.5892

<u>Choosing the best AR model</u>

After plotting the ACF of residual of AR(1) (Figure 9), AR(2) (Figure 10) and AR(3) (Figure 11) models, for Figure 8, there is still correlation at lag 1 so the AR(1) model is not good fit enough. For Figure 9 and Figure 10, the residuals of AR(2) and AR(3) models seem to be white noise because there is no correlation at any lag except lag 0, and both variance of residuals are constant. Moreover, not only the values of both residuals look quite similar, but the parameters of AR(2) and AR(3), such as Coefficients and Sigma^2 of residual, are also so close each other which makes the result be no different. Therefore, to reduce the time consuming and the work load of computation, the AR(2) is the best model for the time series Y.

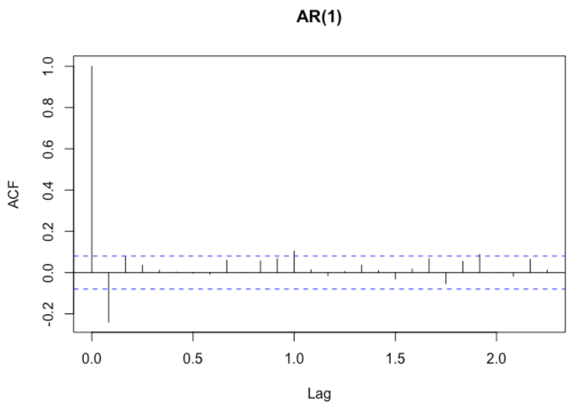

Figure 9: ACF plot of residual of AR(1) model

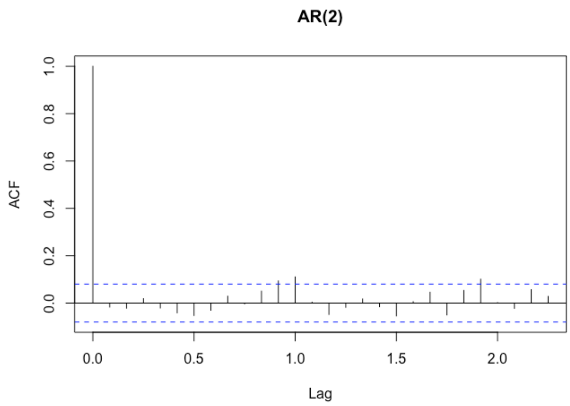

Figure 10: ACF plot of residual of AR(2) model

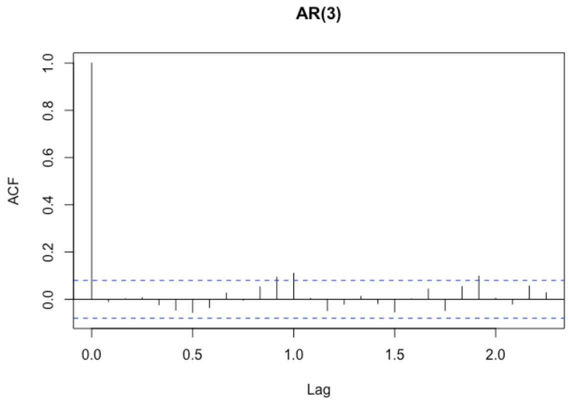

Figure 11: ACF plot of residual of AR(3) model

<u>Plotting periodograms of time series X, Y and Z</u>

In order to find the meaning of X, Y and Z in frequency domain, the function of `fft()` is used to plot the periodograms of time series X (Figure 12), Y (Figure 13) and Z (Figure 14) in this approach. However, they are showed just only first half frequency period in each graph because of symmetricity. For Figure 12, it seems that the data contain seasonality relation, which are the sprite values, and the trend since at $f_j = 0$, the spectral density is high value. For Figure 13, there are plummet values occurring because the trend and seasonal effect have already been removed. For the last graph in the Figure 14, the overall values of the data seem to be very small and flat which means it would be a white noise.

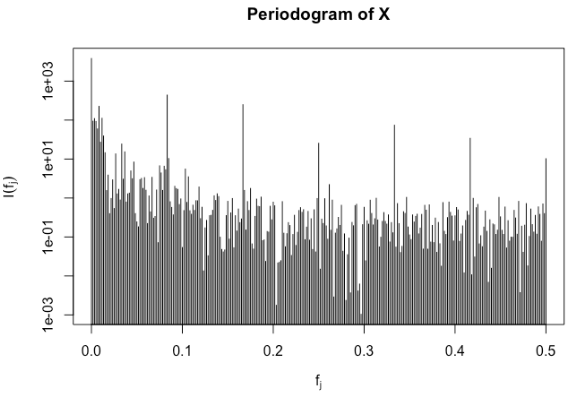

Figure 12: Periodogram plot of time series X

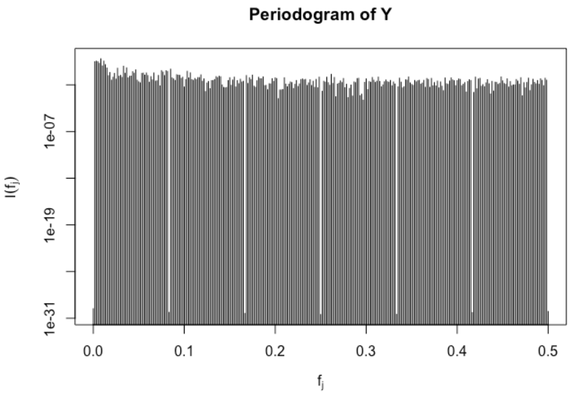

Figure 13: Periodogram plot of time series Y

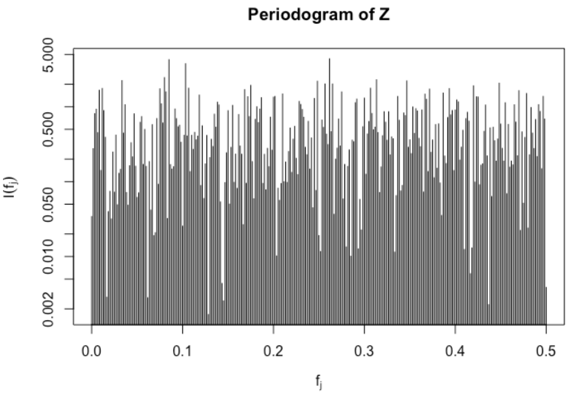

Figure 14: Periodogram plot of time series Z

Final summary

After intense computation of pond's time series (X), there are some processes occurring in order to model the time series X such as eliminating the trend (u(t)) and seasonal effect (s(t)) from the raw data (X_t), followed by extracting the residual (Z_t) via using AR(2) model. These approaches provide a full equation representing as equation 5.

$$X(t) = u(t) + s(t) + AR(2) + Z_t \tag{5}$$

Where:

- X(t) is raw pond time series data which t $\in \{0, 1, \ldots, 599\}$
- u(t) is trend which

$$u(t) = 2.497533 + 0.000066 * t \tag{6}$$

- s(t) is seasonal effect which

$$\begin{aligned} s(t) = &- 1.395 * \delta_{t,0} - 0.148 * \delta_{t,1} + 0.35 * \delta_{t,2} - 0.791 * \delta_{t,3} \\ &- 0.905 * \delta_{t,4} - 1.134 * \delta_{t,5} - 2.153 * \delta_{t,6} - 2.108 * \delta_{t,7} \\ &+ 0.877 * \delta_{t,8} + 2.349 * \delta_{t,9} + 2.444 * \delta_{t,10} + 2.614 * \delta_{t,11} \end{aligned} \tag{7}$$

  Note that: $\delta_{t,i}$ = 1 if t is in month i, and = 0 for else

- AR(2) is autoregressive model at process 2 which

$$AR(2) = 0.6708 * Y_{t-1} + 0.2568 * Y_{t-2} \tag{8}$$

- Z_t is residual of AR(2) being white noise ~ ^{iid} N(0, 0.5887)

Moreover, there are some important results from each process illustrating in the graph as Figure 15

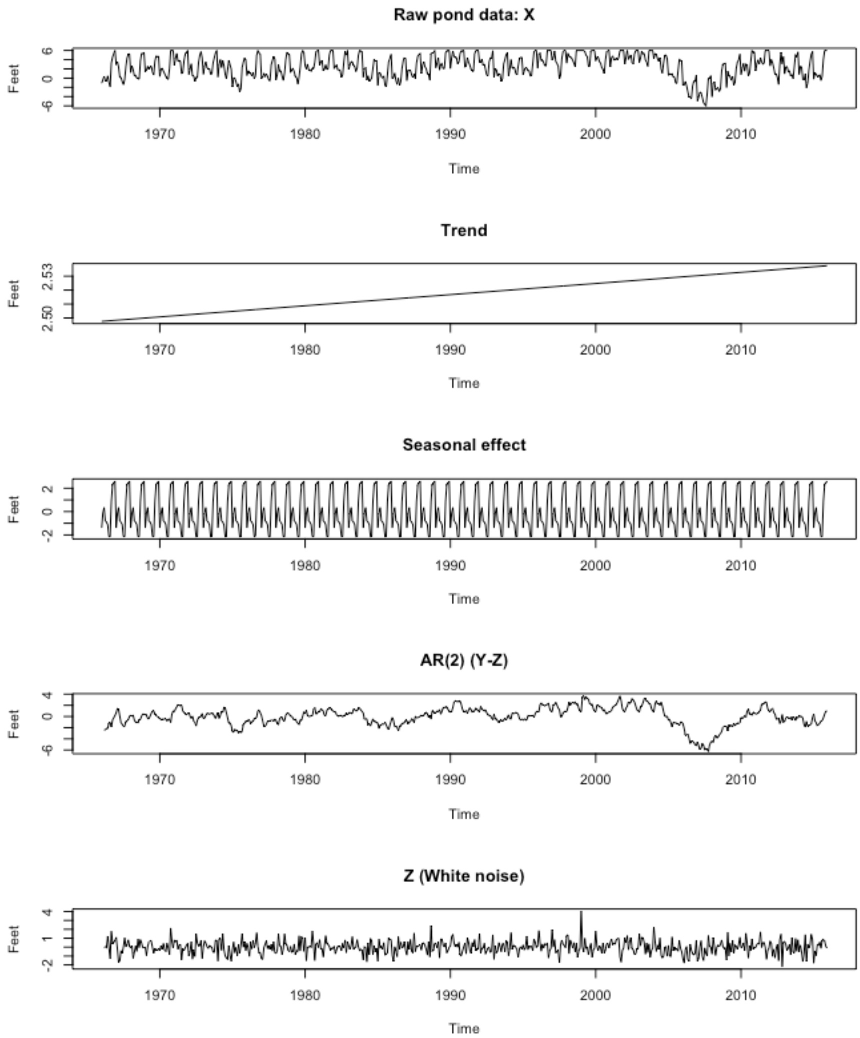

Figure 15: the results of each working step

Academic integrity statement

You must sign this (typing in your details is acceptable) and include it with each piece of work you submit.

I am aware that the University defines plagiarism as presenting someone else's work, in whole or in part, as your own. Work means any intellectual output, and typically includes text, data, images, sound or performance.

I promise that in the attached submission I have not presented anyone else's work, in whole or in part, as my own and I have not colluded with others in the preparation of this work. Where I have taken advantage of the work of others, I have given full acknowledgement. I have not resubmitted my own work or part thereof without specific written permission to do so from the University staff concerned when any of this work has been or is being submitted for marks or credits even if in a different module or for a different qualification or completed prior to entry to the University. I have read and understood the University's published rules on plagiarism and also any more detailed rules specified at School or module level. I know that if I commit plagiarism I can be expelled from the University and that it is my responsibility to be aware of the University's regulations on plagiarism and their importance.

I re-confirm my consent to the University copying and distributing any or all of my work in any form and using third parties (who may be based outside the EU/EEA) to monitor breaches of regulations, to verify whether my work contains plagiarised material, and for quality assurance purposes.

I confirm that I have declared all mitigating circumstances that may be relevant to the assessment of this piece of work and that I wish to have taken into account. I am aware of the University's policy on mitigation and the School's procedures for the submission of statements and evidence of mitigation. I am aware of the penalties imposed for the late submission of coursework.

Name Kritin Senaprame

Student ID 201684900