**Solution to Exercise 2**

WATR 404/604
First Semester 2018

Prepared by David R. Maidment

**Question 1**

*To be Turned In: A plot of E Coli as a function of time for the given data. Determine the mean, median, maximum and minimum value of these data.*

**Solution**

The plot of E. Coli as a function of time is shown in Figure 1. This is for weekly sampling done during summer periods when swimming is popular.

Figure 1. Sampled E. Coli values (CFU/100mL) for Mangatainoka River at Pahiatua Town Bridge for the summers of 2016-2017, and 2017-2018.

The summary statistics of the E. Coli data are shown in Table 1. These were determined using the AVERAGE, MEDIAN, MAX and MIN functions in Excel. Notice how the mean is much higher than the median – this is because there are a few very high E. Coli values that disproportionately affect the mean. The median is a better measure of the “middle” of the data.

|  |  |
| --- | --- |
| Statistic | E. Coli Concentration (CFU/100mL) |
| Mean | 722 |
| Median | 302 |
| Maximum | 6900 |
| Minimum | 86 |

Table 1. Statistics of the E. Coli data

**Question 2**

*To be Turned In: A plot of rainfall and discharge for a shorter interval so you can see the effect of individual rain events on the discharge at Town Bridge. Plot the Turbidity and Discharge as a function of time so you can see how sediment transport responds to changes in flow. Plot the Water Temperature as a function of time – what is the range of water temperature from summer to winter at this location? Plot the DO saturation as a function of time – what do you notice about the values in the summer?*

**Solution**

A chart of daily rainfall and streamflow for 60 days in the Mangatainoka Catchment is shown in Figure 2. The response of streamflow to rainfall events is readily apparent, especially when there are several rainy days in sequence.

Figure 2. Daily Rainfall and Streamflow in the Mangatainoka Catchment, February 2016 to April 2016

The time patterns of discharge and turbidity are shown in Figure 3. There is a readily observable correlation between spikes in the flow and spikes in the turbidity, which makes sense because the extra turbulence in swiftly flowing water stirs up the sediment and increases the turbidity. There is some evidence of high turbidity at the start of the data period during the summer of 2016 when there are few high flows. This might mean that there is biofouling of the turbidity meter. This obscures the penetration of its light beam into the water to measure turbidity and may mean that the turbidity values are erroneously high in this period.

Figure 3. Discharge and Turbidity in the Mangatainoka Catchment, February 2016 to February 2018.

There is a fair amount of fluctuation in the daily water temperature data, as shown in Figure 4.

Figure 4. Daily water temperature in the Mangatainoka at Pahiatua Town Bridge, February 2016 to February 2018.

I attempted to smooth this variation out by taking a 7-day average of the water temperature data centered on day 4 (ie the smoothed temperature on that day is the average of that value and the 3 days prior and 3 days after that day). Figure 5 shows the result.

Figure 5. 7-Day Average Smoothed Water Temperature, February 2016 to February 2018

This is a little bit easier to interpret. It is apparent that the summer of 2016-2017 wasn’t as warm as either the summer before or the summer after, while the winter minimum temperatures are more consistent. It appears that during the summer, the water temperature varies from about 8°C in winter to approximately 20°C in the summer.

The dissolved oxygen saturation is shown in Figure 6. During February 2016 and again in December and January of 2018 there is supersaturation of dissolved oxygen when photosynthesis of vegetation growing in the river produces excess oxygen during the day. This doesn’t happen during the cooler summer of 2016-2017.

Figure 6. Dissolved Oxygen Saturation in Mangatainoka River at Pahiatua Town Bridge, February 2016 to February 2018.

**Question 3**

*To be turned in: Use Regression to prepare a statistical model for the prediction of E. Coli levels in the Mangatainoka River at Pahiatua Town Bridge. Prepare an estimate of the mean concentration of E. Coli for the year at this location using this model. How does this compare with the mean value calculated from the instantaneous sampled values of E. Coli? On a daily basis, multiply the discharge and estimated concentration to get an E Coli load value and find mean daily load in CFU/day for this location.*

**Solution**

I have shown in the text for the assignment that Water Temperature and DO Saturation are not useful variables for describing E. Coli concentrations but that both discharge and turbidity are useful predictors. In particular,

**ECOLI = -256.685 + 0.049438\* DISHARGE + 67.90047 \* TURBIDITY (1)**

This equation R2 = 0.63, which means that it explains about 63% of the variation of the E. Coli values. Given that E. Coli is a highly variable quantity, as the time series chart in Question 1 shows, this is actually quite a good result. If we consider only discharge as an explanatory variable, then:

**ECOLI = -34.3929 + 0.061779\* DISHARGE (2)**

This equation R2 = 0.46, which means that it explains about 63% of the variation of the E. Coli values. The drop in R2 values from 0.63 to 0.46 shows that discharge alone describes 46% percent of the variation of E. Coli and adding turbidity as a predictor variable explains about 17% more of the variation of E. Coli.

To reconstruct the predicted daily average E. Coli values, I used the following procedure:

1. Where both daily mean discharge and turbidity are measured, employ Eq. (1)
2. Where daily mean discharge only is measured, employ Eq. (2)
3. Where the predicted values for E. Coli were negative, set them to 0.

This is an approximate approach to estimating E. Coli values and other statistical models might be formed from these data. The resulting values are shown for the first 9 days of data in Table 2. I have computed the loadings using the formula I presented in the exercise text. For the first row of the data:

Load = 1430 (L/s) \* 1184.9 (CFU/100mL) \* 10 (100mL/1Liter)\* 86,400 (s/day) = 1.46 \* 1012 CFU/day.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Q(l/s)** | **Turbidity(FNU)** |  | **Coliform(CFU/100mL)** | **Load (CFU/Day)** |
| 1430 | 20.19 |  | 1184.922 | 1.46E+12 |
| 1468 |  |  | 53.95107 | 6.84E+10 |
| 1422 |  |  | 56.29867 | 6.92E+10 |
| 1362 |  |  | 53.45684 | 6.29E+10 |
| 1312 | 4.72 |  | 128.6679 | 1.46E+11 |
| 1298 |  |  | 46.66115 | 5.23E+10 |
| 1538 | 34.35 |  | 2151.732 | 2.86E+12 |
| 2553 | 68.39 |  | 4513.243 | 9.96E+12 |
| 5023 | 279.34 |  | 18958.96 | 8.23E+13 |
| ….. | ….. |  | ….. | ….. |

Table 2. Estimated E. Coli concentrations (Col. 3) and loading (CFU/Day) (Col. 4)

Taking the average values of the estimated E. Coli values in Table 2 (for the complete period of the data), the results come out to be 1812 CFU/100mL for the concentration and 9.48 x 1013 CFU/Day for the loading.

The mean concentration of E. Coli computed this way (1812 CFU/100mL) is quite a lot higher than the mean of the observed E. Coli values in the summer time determined in Question 1 (722 CFU/100mL). This means that more investigation is needed to get good estimates of E. Coli from a predictive approach using regression equations. In particular, E. Coli is a quantity that varies logarithmically rather than linearly so regressions using logarithms of the data may help.

In any case, a loading estimate of 9.48 x 1013 CFU/Day is approximately 10 x 106 x 106 or **10 Billion Billion Coliform Units per Day** going past the Pahiatua Town Bridge in the Mangatainoka River. This gives you a sense of how large the bacterial contamination problem is in this river and the degree of effort that will be required to reduce it a level where the river can again support swimming.