Capstone Project Final Report

ESCI 492

3/11/2022

Fecal Coliform in Freshwater and Marine Systems

By Derek Delong & Maggie Evans

Introduction

Fecal coliform is a naturally occurring bacteria found within the gastrointestinal tract of warm-blooded animals, such as humans, birds, and other mammals (USGS, 2017). These enteric bacterias are generally not harmful to other organisms that become exposed to them. However, high concentrations of fecal coliform within the watershed show that fecal contamination has occurred, thus increasing the likelihood that harmful pathogens have entered the watershed (Gregory et al., 2000). Fecal coliform is used as an indicator species to measure the quality of a body of water because it's abundant, easy to observe, and behaves similarly to most pathogenic bacteria found in the environment (Cisneros, 2011). Associated harmful pathogens include Giardia, Cryptosporidium, Shigella sonnei, and Campylobacter (Arnone et al., 2007). In addition to being used as an indicator species, fecal coliform is a facultatively anaerobic organism that can alter the biological and chemical processes of the local environment (Liu et al., 2012). High fecal coliform concentrations have a negative relationship with dissolved oxygen levels in the water column. Therefore, increased concentrations of fecal coliform results in decreased levels of dissolved oxygen, making them a useful metric to determine the overall quality of the watershed (Aslan-Yilmaz, et al, 2004).

Land-use changes, such as agriculture and urbanization, are becoming increasingly prevalent sources of fecal coliform (Laurent et al., 2012). Growing populations lead to a higher rate of urbanization, agriculture, deforestation, sewage, and pet waste. As urbanization expands, so does the total surface area of impervious surfaces: surfaces that don't allow water to permeate into the ground. These impervious physical alterations increase the rate of surface runoff entering the local watershed (Nair et al., 2015). Growing populations produce more sewage, which can leak from pipes and overflow from waste treatment plants, contaminating the watershed. Pet waste in urban areas is also increasing, and with it, an increase in the amount of subsequent waste transported from impervious surfaces (Mostaghimi et al., 2011). Agriculture and livestock waste are other primary sources of fecal contaminants. Agricultural practices such as manure spreading and effluent ponds substantially contribute to the distribution of enteric pathogens found within the watershed (Ferguson et al., 2010). In addition, deforestation weakens the overall soil retention of an area and thereby increases the amount of sediment transported into nearby water sources (Khanal et al., 2013). Sediments transported and suspended within the water column increase the survival rate of enteric bacteria, such as fecal coliform, by increasing the nutrient load and providing a shielding effect from UV radiation and predation by other microorganisms (Irvine et al., 1995).

Studies show that fecal coliform does well in high-turbidity, low-salinity environments (Soueidan et al., 2021), and stormflow often leads to such conditions (Boyer et al., 2007). Large precipitation events result in high rates of surface runoff, sediment transportation, turbidity, and decreases the salinity of marine environments. The Pacific Northwest is affected by climate change in a multitude of ways, and the expected increases in atmospheric rivers throughout the region and their associated extreme precipitation events (Collow et al., 2020) could result in an influx of fecal coliform contamination from surface runoff. In addition, these climatic shifts could result in more favorable environmental conditions for enteric bacterias and pathogens to persist (Anderson et al., 1979).

Various parameters that determine the health of an aquatic ecosystem include monitoring the water and sediment quality by taking flow measurements that observe the physical and chemical conditions as well as using biological indicators, such as shellfish (NWT Water Stewardship). The Washington Department of Health (DOH) uses shellfish as a biological indicator to determine if local Adverse Pollution Conditions (APC) have occurred at monitored sites. Shellfish filter large quantities of water through feeding, thus biomagnifying any toxins and/or pathogens present within the aquatic environment (FDA, 2019). Approved shellfish growing areas with fecal coliform contamination greater than a geometric mean of 14 FC/100mL or an estimated 90th percentile of 43 FC/100 mL are susceptible to closure or restricted access, either indefinitely or for a set period. If Adverse Pollution Conditions occur, the surrounding areas are downgraded to “Prohibited.” If the sources of fecal coliform contamination can be determined and managed, then the “Prohibited” area can be reclassified as “Conditionally Approved” (FDA, 2019).

Different agencies across the State have collected watershed and estuary water quality datasets using different parameters, and as a result, data collections and spreadsheet formatting differ. The discrepancies between data formatting and standardization prevent cross-agency analysis. Standardizing scientific research is critical for proper data analysis because it increases the validity and reliability of research by ensuring consistency and replicability of the study. Standardization of research creates uniformity between various studies and ensures that a certain baseline quality of research is maintained (Weinberg, 2007). Additionally, knowledge gaps can persist through a lack of interagency communication and information sharing, inhibiting meta-analyses (USGS, 2009). The goal of our capstone project was to communicate with four different agencies that have collected regional fecal coliform data over the past 30 years; Padilla Bay, Samish Bay, Stream Team, and Skagit. Datasets were aggregated into two streamlined master sheets, where we could then perform various statistical analyses. However, due to a lack of overlapping variables between agencies and date inconsistencies, interagency analysis was limited to surface water temperature (SWT), and fecal coliform concentrations (FC). We determined that a site-by-site approach would yield better results until further data transformation could be determined.

To obtain meaningful statistical results, we needed to approach our analysis in a way that was immune to date gaps and/or date-clustering. We determined that subsetting the data into seasonality would achieve this, as it would assign an archetype to the data; Spring, Summer, Fall, Winter. In addition, we hypothesized that increased precipitation events would cause increased fecal coliform concentrations within the watershed, and as such, we could expect to see higher fecal coliform concentrations in the Samish River watershed during seasons with high precipitation. We were provided with GPS locations from the Samish Bay and Stream Team sites that allowed us to focus our analysis on the Samish Bay estuary and the surrounding area. Single-site analysis was done at three locations using the Kruskal Wallis Rank-Sum test to determine if seasonality plays a significant role in fecal coliform concentrations. We examined site LS4, where the Samish River meets the Samish Bay, and two additional sites categorized as "Prohibited" according to DOH Adverse Pollution Condition standards; Stations 88 and 89. Spatial temporal analysis was used to detect if land-use changes had observational correlation to any measured spikes in fecal coliform concentrations along the Samish River. Finally we looked at simulations of sediment discharge rates from the Samish River into the Samish Bay that incorporated bay bottom topography and tidal mixing. We did this to determine if there were larger environmental interactions that could potentially provide a better explanation for the prohibited status of Stations 88 and 89 and their spatial relationship to the Samish River.

Methods

Communication with Stakeholders

In order to identify what our stakeholders' needs were, we organized a zoom meeting with Sylvia Yang from Padilla Bay, Kevin Jackman and Karen DuBose from Skagit County Public Works, and Amberose Kelley, to assess their needs and discuss what we could deliver through the aggregation and analysis of their datasets. We discussed focusing on WA DOH marine sites, particularly the prohibited stations 83, 88, and 89, and the freshwater sites monitored by Stream Team and/or Skagit along the Samish River that feed into Samish Bay. In addition, we discussed how best to inventory the datasets to determine if analysis on overlapping time periods was possible. Skagit, Padilla Bay, Samish Bay and Stream Team shared their datasets with Sylvia Yang, who then uploaded and shared the files with us through the Padilla Bay Google Drive.

Data Harmonization

Fecal coliform collection data consisted of 8 to 22 categorical/numerical variables collected from 1994 to 2021. After receiving data sets from all agencies, two master sheets were created in Google Sheets/Google Drive that would host all of the information. No transformations were made to the data (i.e., not filtered for possible human error). These two master sheets were differentiated only in being either: *Single-tab* or *Multi-tab*. Data aggregation was focused primarily on the multi-tab approach. Upon completion of the multi-tab master sheet, the data was copied into the single-tab master sheet using the *control+shift+v* function to ensure no formatting was carried over.

Both Master Sheets

Data transformation of the sample dates was necessary inorder for a seasonality analysis to be achieved. This required subsetting the sample dates across all agencies into a season: Winter, Spring, Summer and Fall. The “Date” column was formatted in Google Sheets using the following methods of transformation for all agencies:

- Format → Number → Date

This was necessary so that Google Sheets would interpret the “Date” column as an actual date. A second column titled “Month” was then created and formatted in a similar manner.

- Format → Number → Plain Text

The “Plain Text” formatting is necessary so that a simple numerical variable representing the month the sample was taken could be derived from the Date column. This was achieved with the following Google Sheets function:

- =MONTH(A2)
    - *Example*: Date Column → (08/08/1994) → returns → 8

A third column titled “Season” was then created and a nested function was used that could transform a numerical variable into a categorical variable.

- =IF(B2=12, "Winter", IF(B2=1, "Winter", IF(B2=2, "Winter", IF(B2=3, "Spring", IF(B2=4, "Spring", IF(B2=5, "Spring", IF(B2=6, "Summer", IF(B2=7, "Summer", IF(B2=8, "Summer", IF(B2=9, "Fall", IF(B2=10, "Fall",IF(B2=11, "Fall"))))))))))))
    - *Example*: If “Month” is “12” then return → “Winter”
- All commas were removed from column data using search and replace function.
    - RStudio will interpret numerical data as a character if commas are present.
- *Ctrl+F* was used to ensure all titles and names were exactly identical across tabs.
- A combination of pastel coloring and selective line bolding was implemented to make the data easier to interpret and differentiate.

Multi-tab

A unique tab was generated for each agency. Columns per tab were only generated for applicable agency-by-agency variables.

- *Example*: “Tide” column only exists in the Samish and Padilla Bay tabs.
- Conditional formatting was used to quickly identify blanks.
    - If cell is “Empty” → Fill “Red”
- “NA” was inserted to fill blanks
    - *Ctrl+F* → ^\s*$ → Replace with → NA
    - Conditional formatting → If cell is *exactly* “NA” → Text “Red”

Single-tab

All agency data was contained in a single-tab with the appropriate columns generated for each variable across agencies. Where the variables were not shared across agencies, the row/column was left blank (i.e., “NA” was not inserted to fill blanks).

- An additional column was added titled, “Agency” for the purpose of subsetting.

Multi-tab Summaries

Once all of the data sets had been sufficiently formatted so that simple subsetting could be applied in Rstudio, a second approach of aggregation was used to make general summary observations on a site-by-site and agency-by-agency level. The Samish Bay and Padilla Bay data sets used a site summary box that contained the means and geometric means of each variable as well as other useful metrics such as site classification, site code, etc,.

We adapted this summary method to each site, within each agency, reflecting agency specific variables, and then aggregated all of the summary boxes into a separate tab. This aggregated summary tab was titled, “Gray Box Site Summaries”. The *Gray Box Site Summaries* tab allows for quick, human-level site interpretation of each site without the need to scroll through the longform data, but the box style is inadequate for data analysis. To solve this, the

Gray Box Site Summaries was reaggregated into another tab titled, “Gray Box Vertical Summaries”. The *Gray Box Vertical Summaries* tab was arranged so that the summary data was vertically aligned for subsetting in Rstudio.

To develop a visual of what each agency’s fecal coliform concentrations looked like relative to one another we replicated the *Gray Box Vertical Summaries* tab into a new tab titled, “Gray Boxes Refined Dataset” but removed data that did not meet our standards for the purpose of generating a simple pie chart. Removed data was quantified as follows:

- Removed: Sites with less than 5 years of data
- Removed: Sites with less than 20 samples

These remaining sites were then summarized at an agency level using the gray box method and a final tab titled, “Ultra Simplified Gray Boxes” was generated. This tab summarized the following data for each agency:

- Geometric Mean of Fecal coliform Concentrations per 100 mL
- Mean Surface Water Temperature in °C
- Mean Salinity (PPT)
- Mean Dissolved Oxygen (%SAT)
- Mean Turbidity (NTU)
- Mean pH
- Mean Years Total
- Total Sites

Spatial Temporal Observational Analysis of the Samish River

Using the latitude and longitude coordinates we received from the Department of Health: Skagit and Samish Bay, we made an approximate conversion to GPS coordinates and mapped the results into Google Earth. For the Skagit sites we designated a color that corresponded to the current status of each monitoring station as determined by the DOH.

- Red: Prohibited Classification
- Yellow: Conditionally Approved Classification
- Green: Approved Classification
- Gray: Undetermined Status

A red zone, signifying the “Prohibited Zone” was added as a layer by approximating the boundaries found in the image “WA DOH Samish Bay.jpg" located in the Google Drive for Padilla Bay Projects. The Stream Team sites were then added to the Google Earth map as general points and the Samish River was traced in blue bold lines to show the connection between Samish Bay and the Samish River watershed.

Using Rstudio, we did a simple regression analysis of fecal coliform concentrations over time for Skagit stations in the prohibited zone; Stations 83,88,89 and the Samish River Stream Team sites; LS1, LS2, LS3, LS4, US1, US2, US3, US4. We then overlaid each regression line onto a single graph to identify any positive trends in fecal coliform concentrations over time which would allow us to focus our analysis. We identified Stations 88, 89 and Stream Team sites US2, LS2 as having minor positive trends in addition to being connected via the Samish River. A strictly observational analysis was then taken to determine if land-use changes could be found in relation to spikes in fecal coliform concentrations, plotted by year, in sites US2 and LS2. Landsat time lapses were generated for both sites from the time period of 2009 through 2019. Changes in land-use were determined by large red-to-green vegetation transitions and then compared to our plot of fecal coliform concentrations by year.

- Landsat Timelapse generation procedure
    - https://streamlit.geemap.org/

Interagency Analysis of SWT vs Fecal Coliform Concentrations

As surface water temperature(°C) and fecal coliform concentrations(/100 mL) were the only consistent inter-agency variables, the following simple linear regression analysis was performed on Rstudio to test for any correlation.

- lm.FC.SWT = lm(FC ~ SWT, data = Mastersheet)

Analysis of Fecal Coliform Concentrations vs. SWT, Time, Turbidity or Salinity

Simple linear regression analysis was performed on Rstudio for:

- LS1-4: FC vs SWT, FC vs Turbidity, FC vs Time as days from sample to 02/24/2022
- US1-4: FC vs SWT, FC vs Turbidity, FC vs Time as days from sample to 02/24/2022
- Stations 83, 88, 89: FC vs SWT, FC vs Salinity, FC vs Time as days from sample to 02/24/2022

Statistical Analysis of the Samish Bay Estuary

Kruskal wallis tests examining fecal coliform concentrations by season were performed on Skagit prohibited Stations 88, 89 and Stream Team site LS4. Rstudio was then used to identify fecal coliform concentration outliers with the following script:

- KW.S83.outliers <- boxplot(S83$FC, plot=FALSE)$out
    - Repeated for each site

The identified outliers were then manually removed from a duplicated master sheet and reuploaded to Rstudio for a second Kruskal Wallis test to validate the results. Boxplots were then generated for each of the sites.

Results

Data Harmonization

Data collected from each agency was successfully aggregated into two separate mastersheets, Single-tab & Multi-tab. Each mastersheet was developed according to the need expressed by the stakeholders per the stakeholders Zoom meeting. Single-tab explicitly aggregates the data into a vertical layout for exportation as a CSV file and ease of subsetting/analysis in Rstudio. The Multi-tab mastersheet explicitly aggregates the data into an agency-by-agency format with summary tabs containing site-by-site information. The Multi-tab is human-centric, designed for uploading additional collected data and keeping the data separated for easy visual interpretation.

Spatial Temporal Observations of the Samish River

A simple regression analysis of fecal coliform concentrations over time for Skagit stations in the prohibited zone; Stations 83,88,89 and the Samish River Stream Team sites; LS1, LS2, LS3, LS4, US1, US2, US3, US4 did not yield conclusive results due to misalignments in sampling dates that skewed the data. The regression lines were used to identify positive trends in fecal coliform concentrations over time in Stream Team sites US2 and LS2 for our spatial temporal analysis (Figures 2.1, 2.2). A LANDSAT vegetation reflectance time lapse of US2 exhibited a deforestation event on a hillside above the site having occurred between 2011 and 2012 (Figure 1.1). This aligned with spikes in fecal coliform concentrations found in a general plot of the US2 data. No major land use changes were found in LANDSAT data at LS2 during the same period, although the LS2 fecal coliform plot did have similar spikes in concentrations from 2011 to 2012 (Figure 1.2). No other sites down river exhibited similar spikes during the same period of time.

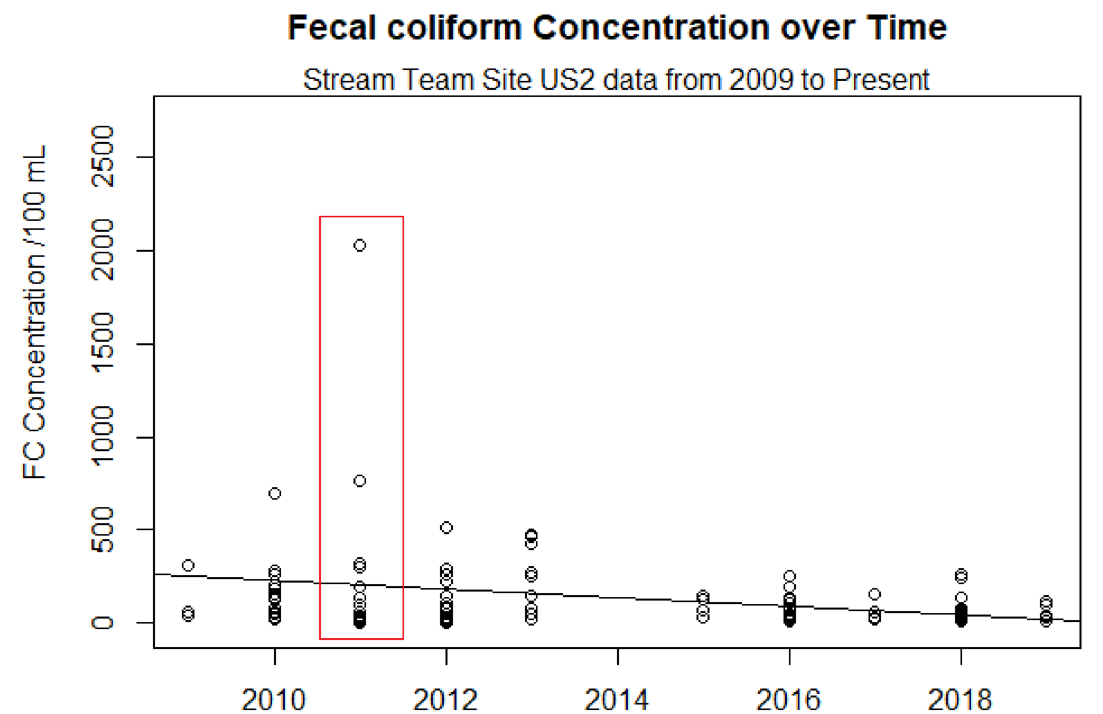

Figure 1.1 A general plot of Stream Team site US2 that exhibited large spikes in fecal coliform concentrations between 2011 and 2012.

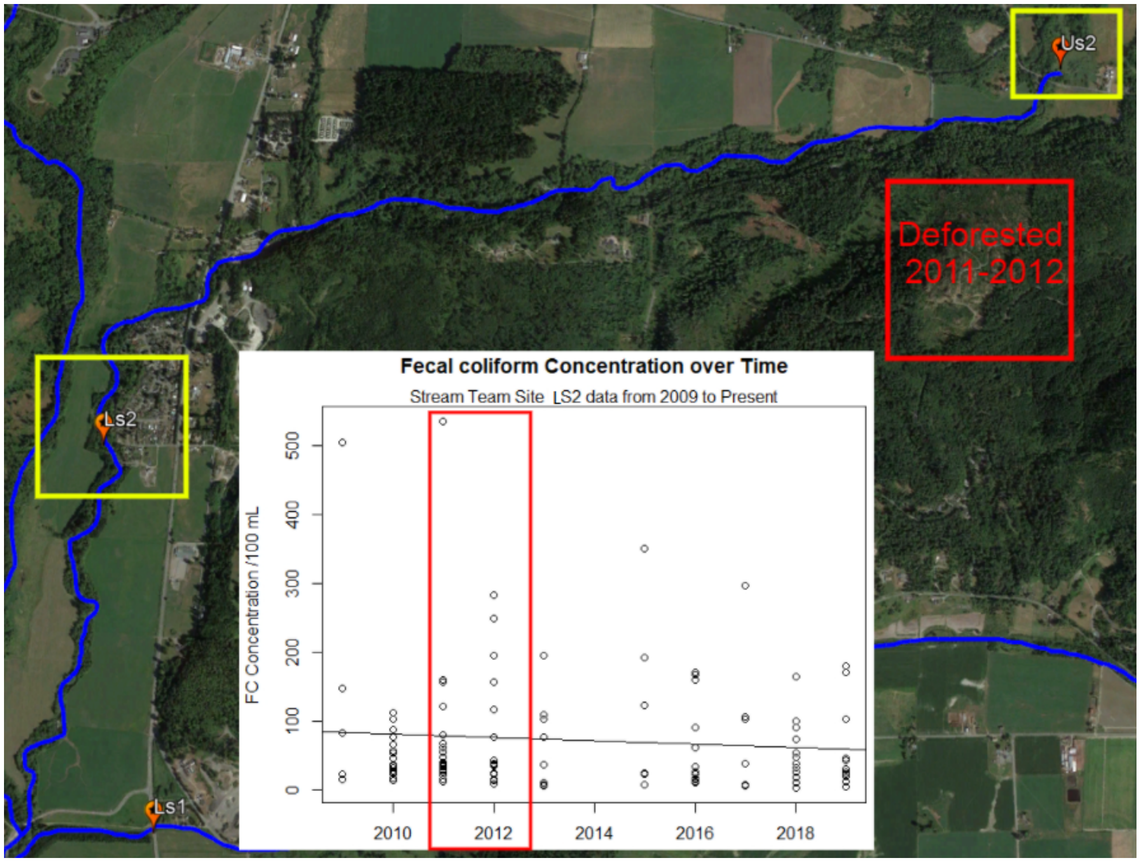

Figure 1.2 A view of the spatial relationship between US2 and LS2. The large red box outlines a deforestation event that occurred some time between 2011 and 2012. Any such event would be expected to increase local fecal coliform concentrations in the river and the plotted data generally reflects that. LS2 data was found to have similar patterns to the US2 data. Fecal coliform introduced into the river appears to have traveled down river and persisted long enough to be detected by sampling. The deforestation event may have continued to influence fecal coliform concentrations post 2011.

Interagency Analysis of SWT vs Fecal Coliform Concentrations

An interagency simple linear regression analysis of surface water temperature and fecal coliform concentrations was found to be significant, with a p-value of $4.9*10^{-15}$ and a R^2 value of 0.001966. This indicates that the independent variable, SWT, does not explain the variation in our dependent variable, FC.

Analysis of Fecal Coliform Concentrations vs. SWT, Turbidity or Salinity, & Time

Site LS1 had a statistically significant p-value and an R^2 value of 0.13, meaning that 13% of the dependent variable, FC, can be explained by the independent variable, SWT. Turbidity at this site had a statistically significant p-value and an R^2 value of 0.09, meaning that 9% of the dependent variable, FC, can be explained by the independent variable, turbidity. Site LS2 had a statistically significant p-value and an R^2 value of 0.23, meaning that 23% of the dependent variable, FC, can be explained by the independent variable, SWT. Turbidity at this site showed no statistical significance. Site LS3 had a statistically significant p-value and an R^2 value of 0.07, meaning that 7% of the dependent variable, FC, can be explained by the independent variable, SWT. Turbidity at this site showed no statistical significance. Site LS4 did not show any statistical significance between fecal coliform and all measured variables.

Site US1 had a statistically significant p-value and an R^2 value of 0.11, meaning that 11% of the dependent variable, FC, can be explained by the independent variable, SWT. Turbidity at this site had a statistically significant p-value and an R^2 value of 0.06, meaning that 6% of the dependent variable, FC can be explained by the independent variable, turbidity. Site US2 had a statistically significant p-value and an R^2 value of 0.08, meaning that 8% of the dependent variable, FC can be explained by the independent variable, SWT. Turbidity at this site had a statistically significant p-value and an R^2 value of 0.07, meaning that 7% of the dependent variable, FC, can be explained by the independent variable, turbidity. Site US3 did not show any statistical significance between FC and SWT. Turbidity at this site had a statistically significant p-value and an R^2 value of 0.03, meaning that 3% of the dependent variable, FC, can be explained by the independent variable, turbidity. Site US4 had a statistically significant p-value and an R^2 value of 0.078, meaning that ~ 8% of the dependent variable, FC can be explained by the independent variable, SWT. Turbidity at this site showed no statistical significance.

Station 83 had a statistically significant p-value and an R^2 value of 0.26, meaning that 26% of the dependent variable, FC can be explained by the independent variable, SWT. Salinity at this site had a statistically significant p-value and an R^2 value of 0.07, meaning that 7% of the dependent variable, FC, can be explained by the independent variable, salinity. Station 88 had a statistically significant p-value and an R^2 value of 0.07, meaning that 7% of the dependent variable, FC, can be explained by the independent variable, SWT. Salinity at this site had a statistically significant p-value and an R^2 value of 0.18, meaning that 18% of the dependent variable, FC, can be explained by the independent variable, salinity. Station 89 had a statistically significant p-value and an R^2 value of 0.09, meaning that 9% of the dependent variable, FC, can be explained by the independent variable, SWT. Salinity at this site had a statistically significant p-value and an R^2 value of 0.22, meaning that 22% of the dependent variable, FC, can be explained by the independent variable, salinity.

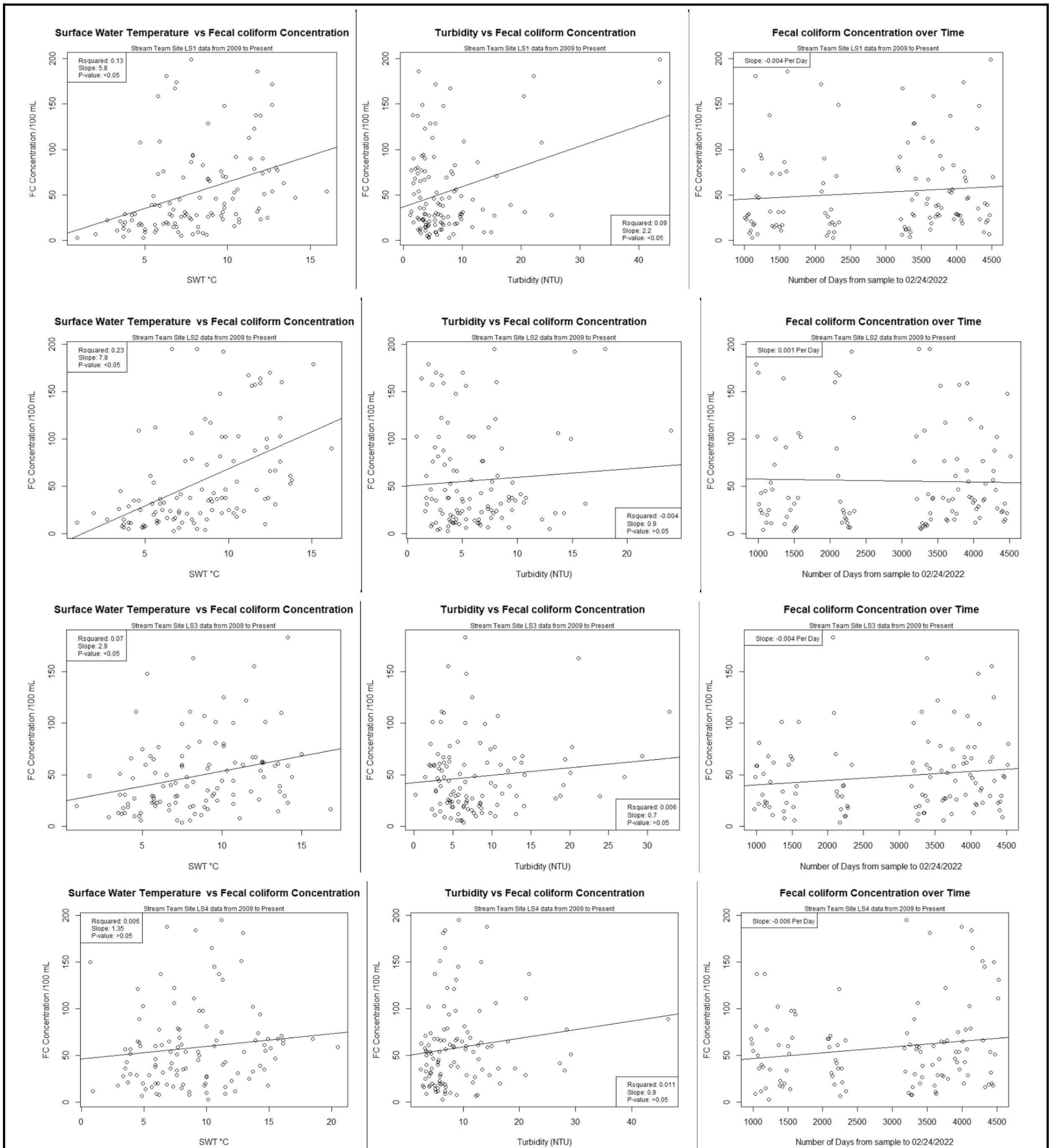

Figure 2.1 Linear Regression Analysis at Sites LS1-4. Analysis shows no significant relationship between FC and time. SWT shows significance at most site locations, but with low R^2 values, indicating that SWT is not a good explanation for the variation within FC. Turbidity shows significance at some site locations, but with low R^2 values, indicating that turbidity is not a good explanation for the variation within FC.

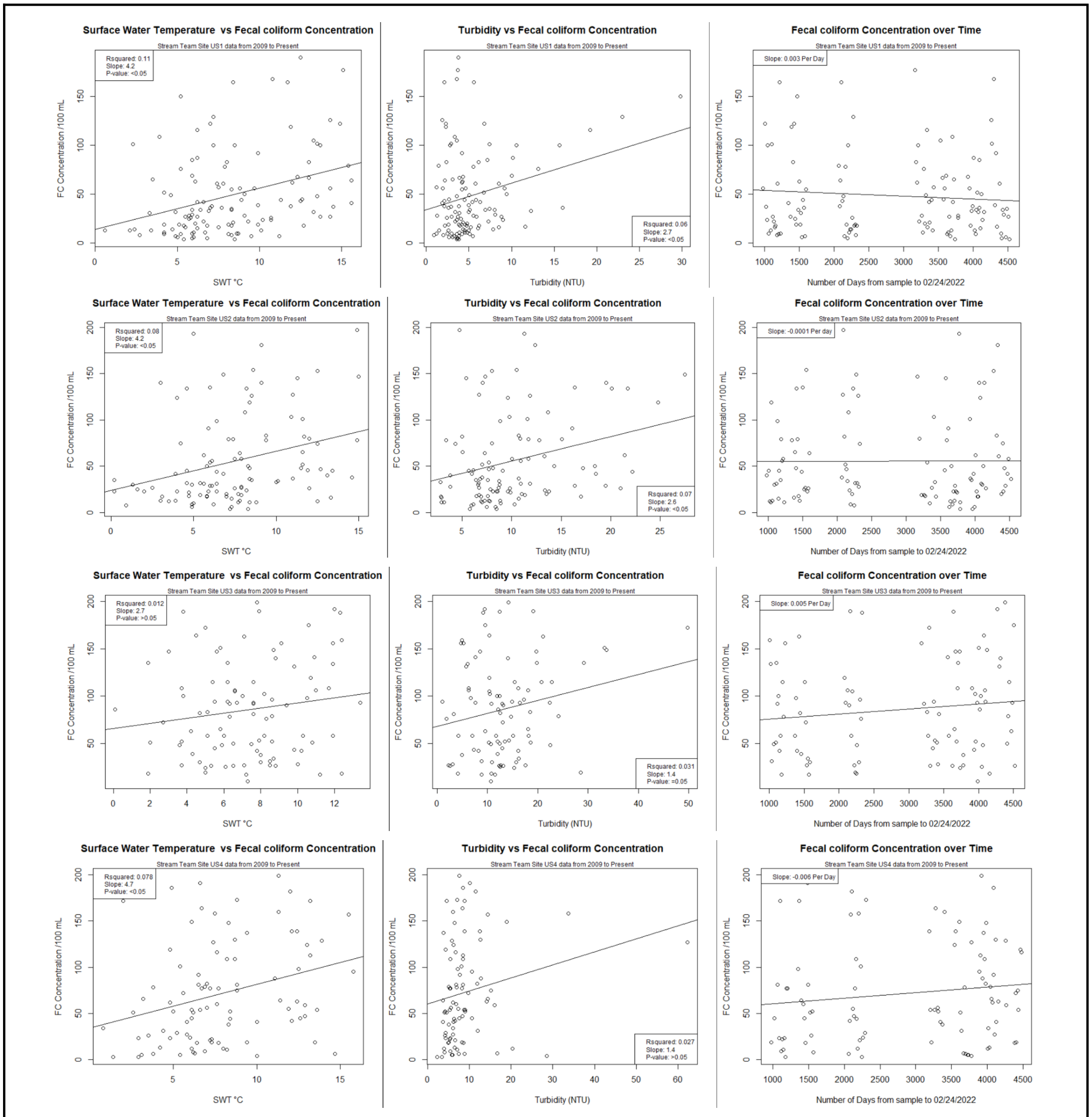

Figure 2.2 Linear Regression Analysis at Sites US1-4. Analysis shows no significant relationship between FC and time. SWT shows significance at most site locations, but with low R^2 values, indicating that SWT is not a good explanation for the variation within FC. Turbidity shows significance at some site locations, but with low R^2 values, indicating that turbidity is not a good explanation for the variation within FC.

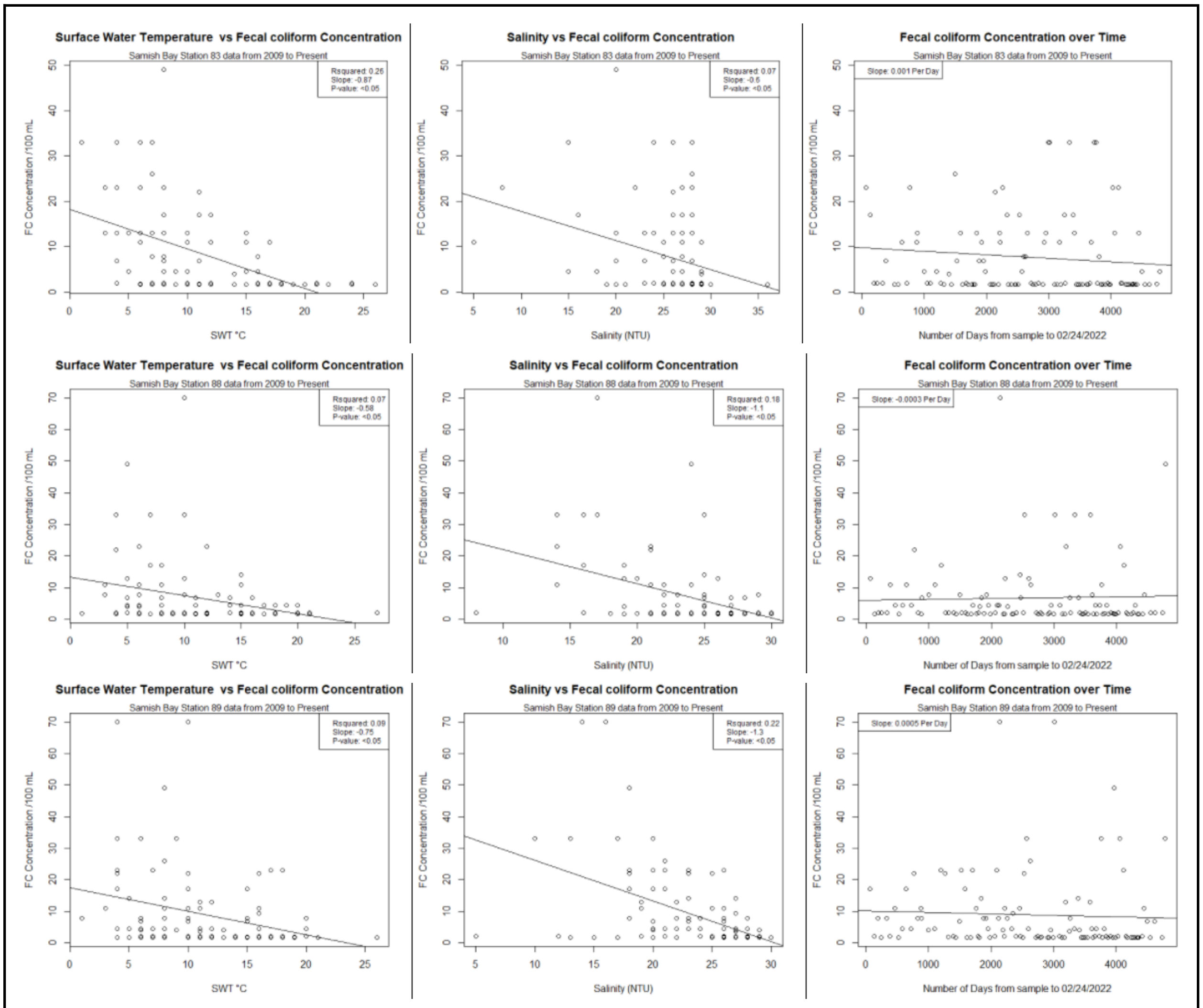

Figure 2.3 Linear Regression Analysis at Stations 83, 88, and 89. Analysis shows no significant relationship between FC and time. SWT shows significance at all site locations, but with low R^2 values, indicating that SWT is not a good explanation for the variation within FC. Salinity shows significance at all site locations, but with low R^2 values, indicating that salinity is not a good explanation for the variation within FC.

Statistical Analysis of the Samish Bay Estuary

A Kruskal-Wallis rank sum test of fecal coliform concentration and seasonality at site LS4 and Stations 88 and 89 found significant p-values and the chi-squared values were greater than the critical value, indicating that seasonality does influence fecal coliform concentration at these locations. Site LS4 had 3 degrees of freedom, a significant p-value of 0.03346, and a chi-squared value of 8.7065. Station 88 had 3 degrees of freedom, a significant p-value of 0.000336, and a chi-squared value of 18.566. Station 89 had 3 degrees of freedom, a significant p-value of $3.51*10^{-6}$, and a chi-squared value of 28.07. Additional Kruskal-Wallis rank sum tests with outliers in fecal coliform concentrations removed found site LS4 had 3 degrees of freedom, a significant p-value of 0.0007754, and a chi-squared value of 16.804. Station 88 had 3 degrees of freedom, a significant p-value of 0.004065, and a chi-squared value of 13.282. Station 89 had 3 degrees of freedom, a significant p-value of 0.00002395, and a chi-squared value of 24.087.

Discussion

Data Harmonization

Interagency analysis was limited due to the lack of overlapping variables and sampling dates. It would be beneficial for all agencies to communicate with one another to avoid knowledge gaps and determine what parameters should be measured for future analysis. Skagit's datasets were extensive and thorough, thus we believe this would be a good model for all agencies to follow. Additionally, taking frequent and consistent measurements, if possible 1 time a week at all sites to avoid future data gaps, would help aid in meaningful future analysis. Additional parameters, such as taking flow rate measurement at time of sample and noting if there were precipitation events within the past 1-3 days, may also be beneficial for future analysis.

Spatial Temporal Observations of the Samish River

The LANDSAT data of US2 provided an example of land-use changes, deforestation in this case, as having possibly influenced fecal coliform concentrations at the sampling site as well as immediately downstream at site LS2. The LANDSAT data was limited to year-to-year transitions so our fecal coliform data was plotted congruently as year-to-year to try and better align with the LANDSAT data we had available. Unfortunately, this prevented empirical analysis, as we could not pinpoint the exact dates when the deforestation event occurred. Therefore, we could not cross-reference the deforestation event with the fecal coliform sampling dates. Other satellite data that has more defined time frames would likely yield more conclusive results.

The deforested area was on a hillside directly above the Samish River and as such, sediment transportation of fecal coliform from the loosened soil could have explained the spikes in fecal coliform concentrations we found in the plotted data. The LS2 fecal coliform concentrations by year showed similar spikes to US2 during the 2011-2012 period. This would be consistent with what we would expect to find in downstream environments for sediment discharge. Sites further downstream did not exhibit similar patterns in the fecal coliform data. This suggests that there is a possibility the residence time of fecal coliform is shorter in rivers with low flow rates, such as the Samish River (USGS, 2018).

Interagency Analysis of SWT vs Fecal Coliform Concentrations

Surface water temperature (SWT) and fecal coliform concentrations (FC) were the only variables consistent across all agencies. Although interagency analysis between SWT and FC yielded statistical significance, the R^2 value was very low, at 0.001966, meaning that $> 1\%$ of the time SWT can explain the variation found within FC. During the time of analysis, the data had not been transformed to reflect seasonality, thus future analysis using seasonality subsetting features in RStudio may yield more conclusive results. Human error while recording SWT may have occurred, as we suspect that although temperature was recorded in °C, temperatures may have actually been measured in °F in some instances. Stakeholder Sylvia Yang agrees with this possibility, although verification is not possible. We suggest that in future analysis, outliers are removed using the RStudio script provided in the methods section.

Analysis of Fecal Coliform Concentrations vs. SWT, Turbidity or Salinity, & Time

Site-by-site regression analysis of FC versus SWT, turbidity, and time was performed at sites LS1-4 (Figure 2.1) and US1-4 (Figure 2.2). Additionally, site-by-site regression analysis was performed at Stations 83, 88, and 89 to determine if there was a relationship between FC versus SWT, salinity, and time (Figure 2.3). No significance was detected between FC and time at all observed locations. SWT did have significance at most site locations, but with low R^2 values, indicating that SWT is not a good explanation for the variation within FC. Results showed that FC and turbidity had significance at some sites as well, but with very low R^2 values, indicating that turbidity is not a good explanation for the variation within FC. At all Stations observed, 83, 88, and 89, salinity showed significance, but with low R^2 values, indicating that salinity is not a good explanation for the variation within FC.

Statistical Analysis of the Samish Bay Estuary

Our Kruskal-Wallis rank sum tests determined that seasonality did have a significant influence on FC at site LS4 and Stations 88 and 89. Box plots were generated in RStudio, however, due to there being a large amount of outliers within the datasets, the visual representation was skewed. This made it difficult to see the relationship between seasonality and FC (Figure 3.1). To address this, we had RStudio identify fecal coliform concentration outliers in the dataset and then we manually removed the outliers from the data. We then ran another

Kruskal-Wallis rank sum test with the outliers removed and found that significance remained at all sites. The new box plots generated provided a statistically valid and clearer visual representation of the relationship between FC and seasonality (Figure 3.2). Box plots generated in Figure 3.2 are arranged to represent the spatial distances from the Samish River. Site LS4 is where the Samish River feeds into the Samish Bay estuary. Station 89 is the next closest site, and slightly further out into the bay is Station 88.

The box plot for LS4 displays larger values for FC on the y-axis compared to the box plots for Stations 89 & 88 (Figure 3.2). The fecal coliform concentrations at LS4 ranged from 0 - 250 FC/100mL, whereas Stations 89 and 88, located further out into Samish Bay, had much lower concentrations ranging from 0 - 50 FC/100mL. The higher median values at site LS4 align with expectations as it is a freshwater system that runs through extensive agricultural lands. The relatively lower median values of FC found at Stations 89 and 88 is to be expected, as the stations are located further away from terrestrial inputs. Additionally, suspended fecal coliform in the water column will be exposed to increasing levels of salinity as it travels into Samish Bay.

Site LS4 has higher median values during all four seasons relative to Stations 89 and 88, with Fall having the greatest median value at ~ 65FC/100mL and the largest distribution extending to ~ 200FC/100mL. The higher median Fall value is consistent with expected influxes of fecal coliform occurring with greater terrestrial run-off due to the rainy Fall season. Fall discharge rates in the Samish River can be as high as 5,020 cu ft/s, with a 3-5 year cycle of flooding possibly explaining the extreme outliers present in the data (USGS, 2005). Summer at site LS4 had the second highest median value at ~ 60FC/100mL, but was more range bound than all other seasons. Summer flow rates in the Samish River can be as low as 15 cu ft/s (USGS, 2005). We suspect that the relatively high median, but range bound FC during this period, may be due to the lower discharge rates of the river. Low energy discharge rates may prevent adequate estuarine mixing as the sediment, and thus fecal coliform, can persist in the surrounding near-shore area. The fecal coliform concentrations appear to drop off in relation to the distance from LS4, hence why Station 89 median values are slightly higher than Station 88. Exposure to higher salinity conditions will occur as fecal coliform travels further away from the freshwater input of the Samish River. Additionally, FC will become diluted as it travels out further into the Bay. Stations 89 and 88 both see low median values of FC, with Winter having the highest median value at ~ 3 - 5 FC/100mL, respectively. We are unsure as to why the spike in median values occurs during this time, as it does not correspond to what we see at site LS4. Future projects can do a more in depth analysis to address this phenomenon.

The Salish Sea Modeling Center Youtube channel created simulations that show the influence that Samish Bay bottom topography and tide patterns have on sediment discharge of the Samish River (Salish Sea Modeling Center, 2022). In these simulations we can see that sediment discharges, possibly containing fecal coliform, tend to cluster into clouds and linger directly over the prohibited Stations 88 and 89. These simulations can bridge the gap that our data alone cannot explain. The ebb and flow of sediment discharges, bottom bay topography, tides, and environmental conditions should all be considered when examining complex systems such as the Samish Bay estuary.

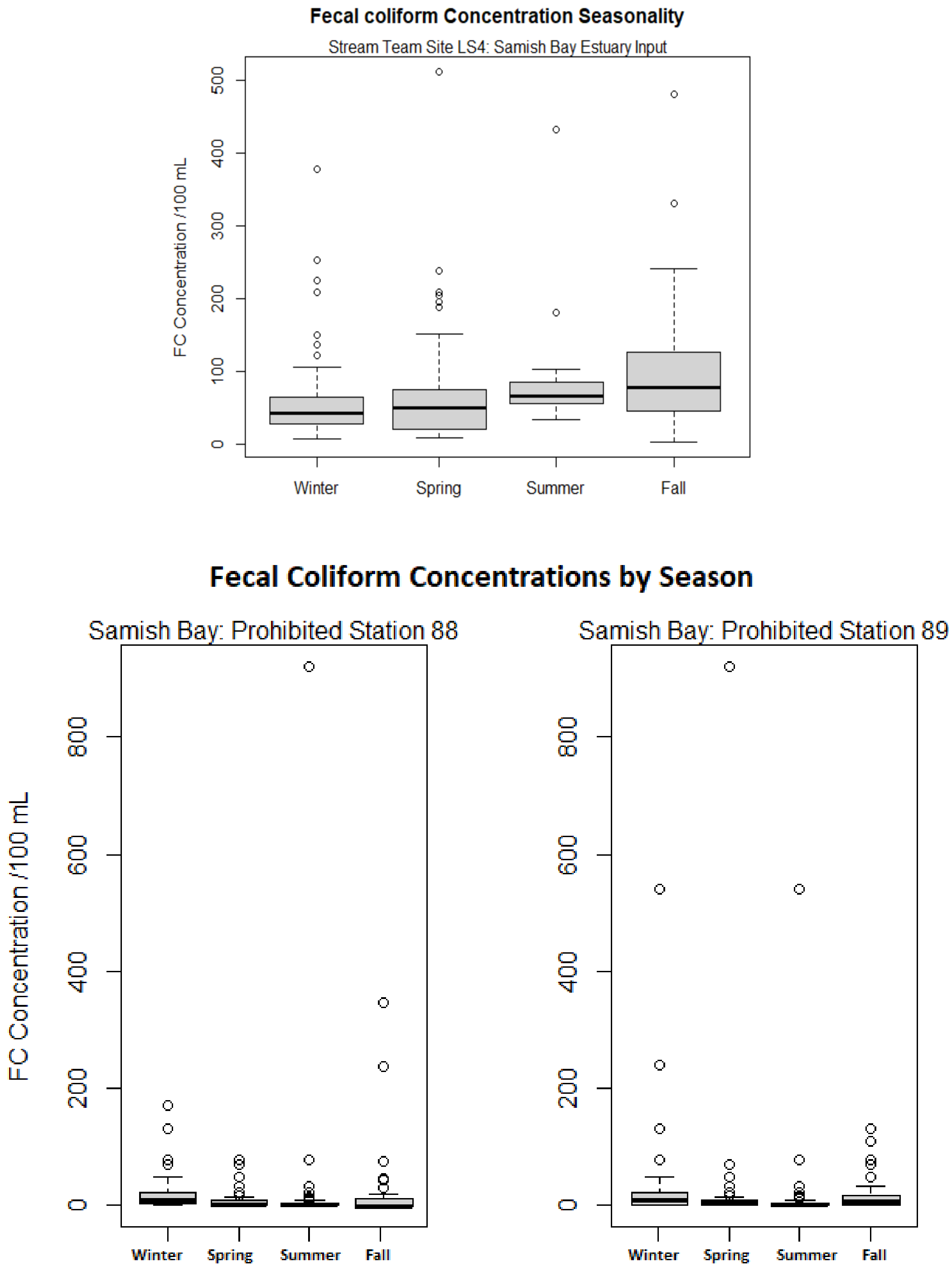

Figure 3.1 Fecal coliform concentrations by seasonality at site LS4 and Stations 88 and 89 with no outliers removed. Stations 88 and 89 are very skewed, making it difficult to draw a meaningful conclusion on the relationship between FC and seasonality from the data plotted above.

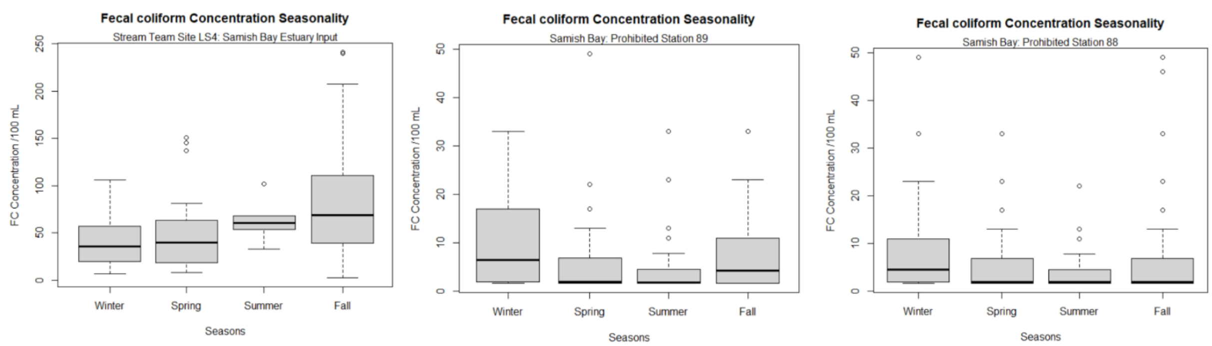

Figure 3.2 Fecal coliform concentrations by seasonality at site LS4 and Station 88 and 89 with outliers removed. The median fecal coliform concentration is higher in LS4 in the Summer, but not at Stations 89 and 88. Low discharge rates of the Samish river not pushing the freshwater out into the bay for mixing with the saline ocean water may be a plausible explanation.

Conclusions

Through communication with various stakeholders, we assessed their needs and developed two harmonized datasets. These harmonized datasets took two separate approaches to data aggregation; human-centric (Multi-tab) and statistical exportation (Single-tab). Multi-tab allowed for one to look at the aggregated data on an agency-by-agency and site-by-site level for summary conclusions to be made. While the Single-tab is a longform data aggregation that is difficult to visually interpret, but is set up to be easily analyzed in Rstudio with little need for additional subsetting. Inconsistencies in sampling dates proved to be the greatest challenge when trying to perform meaningful statistical analysis, and interagency meta-analysis was limited due to the lack of overlapping variables across agencies. Efforts to mitigate future hindrance when performing interagency statistical analysis should be taken by all agencies. Agencies should begin to gather standardized data frequently and consistently to yield better results in the future. Students should harmonize new incoming data sets onto the master sheets as well as continue our transformation techniques (i.e., Month, Seasonality). Furthermore, students should perform site-by-site analysis on locations that have yet to be analyzed such as the Sloughs of Skagit and

Samish Bay. Using LANDSAT technologies, students and/or stakeholders should also take note of any regional land-use changes that may be influencing fecal coliform concentrations.

Fecal coliform concentrations within the watershed are influenced by a multitude of dynamic environmental, spatial, and temporal factors. These can range from land-use changes such as agriculture, increased impervious surfaces, or deforestation, leading to influxes of fecal coliform within the watershed. Additional spatial and temporal factors such as the topography of a bay and/or tidal patterns can have strong influences on the mixing rates of high salinity with low salinity water. This can influence the residence time of fecal coliform in the water due to the destructive effect high salinity water has on fecal coliform (Florini et al., 2020). Precipitation events also have been shown to be related to influxes of fecal coliform in watersheds. Therefore, efforts should be made to note the proximity of sampling dates to recent precipitation events. Finally, a large influx of fecal coliform could occur at any point in a system, such as the Samish River, and if sampling is inconsistent or infrequent, this may skew the data on spatially related sites. Any approach of data analysis should take factors such as these into consideration before coming to any conclusions.

Consistent interagency data sharing improves the chances of identifying what variables could be directly responsible for the influx of fecal coliform within the local watershed. Standardized data collection methods, as well as interagency communication, are equally important in determining the source and fate of fecal contamination, as they minimize knowledge gaps and allow for better meta-analysis. Interagency cooperation, beyond our project focus of fecal coliform, is critical to enhance knowledge, increase monitoring efforts, and develop effective management strategies to address any potential impacts on the ecosystem. Fecal coliform contamination within the watershed will likely persist and compound over time, as urban environments expand, land-use patterns change, and regional precipitation rates increase due to unabated climate change. Increased interagency harmonization; the ways in which agencies communicate, collect, and share their knowledge, should be pursued as collaborative efforts can be viewed as force multipliers. Without interagency harmonization, unoptimized solitary efforts might be additive at best, redundant, or ultimately unsuccessful.

Literature Cited

Anderson, I C, et al. “Sublethal Stress in Escherichia Coli: A Function of Salinity.” *Applied and Environmental Microbiology*, vol. 38, no. 6, 1979, pp. 1147–1152., https://doi.org/10.1128/aem.38.6.1147-1152.1979.

Arnone, Russell D., and Joyce Perdek Walling. “Waterborne Pathogens in Urban Watersheds.” *Journal of Water and Health*, vol. 5, no. 1, 2006, pp. 149–162., https://doi.org/10.2166/wh.2006.001.

Aslan-Yılmaz, Aslı, et al. “Bacteriological Indicators of Anthropogenic Impact Prior to and during the Recovery of Water Quality in an Extremely Polluted Estuary, Golden Horn, Turkey.” *Marine Pollution Bulletin*, vol. 49, no. 11-12, 2004, pp. 951–958., https://doi.org/10.1016/j.marpolbul.2004.06.020.

Bulthuis, Douglas. “NOAA Office for Coastal Management.” Washington State Department of Ecology , NOAA, 2010, https://coast.noaa.gov/data/docs/nerrs/Reserves_PDB_SiteProfile.pdf.

Boyer, Douglas G., and Ewa Kuczynska. “Storm and Seasonal Distributions of fecal coliforms and Cryptosporidium in a Spring.” *Journal of the American Water Resources Association*, vol. 39, no. 6, 2003, pp. 1449–1456., https://doi.org/10.1111/j.1752-1688.2003.tb04430.x.

Center for Food Safety and Applied Nutrition, FDA. “National Shellfish Sanitation Program (NSSP).” U.S. Food and Drug Administration, FDA, 2019, https://www.fda.gov/food/federalstate-food-programs/national-shellfish-sanitation-program-nssp.

Cisneros, B.J. “Safe Sanitation in Low Economic Development Areas.” *Treatise on Water Science*, 2011, pp. 147–200., https://doi.org/10.1016/b978-0-444-53199-5.00082-8.

Collow, Allison B., et al. “Large-Scale Influences on Atmospheric River–Induced Extreme Precipitation Events along the Coast of Washington State.” *Journal of Hydrometeorology*, vol. 21, no. 9, 2020, pp. 2139–2156., https://doi.org/10.1175/jhm-d-19-0272.1.

Ferguson, Christobel, et al. “Fate and Transport of Surface Water Pathogens in Watersheds.” *Critical Reviews in Environmental Science and Technology*, vol. 33, no. 3, 2003, pp. 299–361., https://doi.org/10.1080/10643380390814497.

Florini, Styliani, et al. “Factors Influencing the Concentration of Fecal Coliforms in Oysters in the River Blackwater Estuary, UK.” *Water*, vol. 12, no. 4, 2020, p. 1086., https://doi.org/10.3390/w12041086.

Gregory, M. Brian, and Elizabeth A. Frick. “Fecal-Coliform Bacteria Concentrations in Streams of the Chattahoochee River National Recreation Area, Metropolitan Atlanta, Georgia, May–October 1994 and 1995.” USGS Water-Resources Investigations Report 00-4139: Fecal-Coliform Bacteria Concentrations in Streams of the Chattahoochee River National Recreation Area, Metropolitan Atlanta, Georgia, May October 1994 and 1995, U.S. GEOLOGICAL SURVEY, 2000, https://pubs.usgs.gov/wri/wri004139/.

Holland, A. F., Sanger, D. M., Gawle, C. P., Lerberg, S. B., Santiago, M. S., Riekerk, G., Zimmerman, L. E., & Scott, G. I. (2004, January 28). *Linkages between Tidal Creek ecosystems and the landscape and demographic attributes of their watersheds*. Journal of Experimental Marine Biology and Ecology. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022098103003575

Irvine, Kim, et al. “Indicator Bacteria-Sediment Relationships: Implications for Water Quality Modeling and Monitoring.” *Journal of Water Management Modeling*, 1995, https://doi.org/10.14796/jwmm.r183-14.

Khanal, Sunita, and Prem B. Parajuli. “Evaluating the Impacts of Forest Clear Cutting on Water and Sediment Yields Using SWAT in Mississippi.” *Journal of Water Resource and Protection*, vol. 05, no. 04, 2013, pp. 474–483., https://doi.org/10.4236/jwarp.2013.54047.

Lawrence, Sally. “Samish Bay Watershed fecal coliform Bacteria Total Maximum Daily Load.” Department of Ecology, State of Washington, 2009,

https://apps.ecology.wa.gov/publications/documents/0910019.pdf.

Liu, Zhenhuan, et al. “Impervious Surface Impact on Water Quality in the Process of Rapid Urbanization in Shenzhen, China.” *Environmental Earth Sciences*, vol. 68, no. 8, 2012, pp. 2365–2373., https://doi.org/10.1007/s12665-012-1918-2.

Mostaghimi, Saied, et al. “fecal coliform TMDL Development. - Opensiuc.lib.siu.edu.” OpenSIUC, Virginia Polytechnic Institute and State University, Jan. 2011, https://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1136&context=jcwre.

Nair, Dr. M. V. Harindranathan, et al. “Impacts of Impervious Surfaces on the Environment - IJESI.” Academia , International Journal of Engineering Science Invention , May 2015, http://ijesi.org/papers/Vol(4)5/E045027031.pdf.

“National Estuarine Research Reserve System.” Padilla Bay National Estuarine Research Reserve, NOAA, https://coast.noaa.gov/nerrs/reserves/Padilla-bay.html.

NOAA. “Centralized Data Management Office: Data Export System.” Centralized Data Management Office | Data Export System, NOAA, http://cdmo.baruch.sc.edu/dges/.

“Particle Release from Alice Bay Drainage for High Flow Period (Late November).” *YouTube*, Salish Sea Modeling Center, 16 Feb. 2022, https://www.youtube.com/watch?v=I0iw_1c3Boc. Accessed 12 Mar. 2022.

Soueidan, Jamileh, et al. “A Changing Estuary: Understanding Historical Patterns in Salinity and fecal coliform Levels in the May River, SC.” *Marine Pollution Bulletin*, vol. 168, 2021, p. 112384., https://doi.org/10.1016/j.marpolbul.2021.112384.

NWT Water Stewardship. “How Is Aquatic Ecosystem Health Measured?” How Is Aquatic Ecosystem Health Measured?, Government of the Northwest Territories, https://www.nwtwaterstewardship.ca/en/how-aquatic-ecosystem-health-measured.

St Laurent, Jacques, and Asit Mazumder. “The Influence of Land-Use Composition on Fecal Contamination of Riverine Source Water in Southern British Columbia.” *Water Resources Research*, vol. 48, no. 12, 2012, https://doi.org/10.1029/2012wr012455.

“USGS 12201500 Samish RIVER NEAR BURLINGTON, WA.” *USGS Current Conditions for USGS 12201500 Samish River near Burlington, WA*, USGS, 2018, https://nwis.waterdata.usgs.gov/nwis/uv?site_no=12201500.

USGS. “Fecal Indicator Bacteria and Sanitary Water Quality.” *Fecal Indicator Bacteria and Sanitary Water Quality*, USGS Michigan Water Science Center, 4 Jan. 2017, https://mi.water.usgs.gov/h2oqual/BactHOWeb.html.

USGS“ Location of Surface-Water Stations in the Samish River and Whatcom Creek Basins.” *Samish RIVER BASIN* , USGS, 2005, https://pubs.usgs.gov/wdr/2005/wdr-wa-05-1/pdf/wa00103ADR2005_Figure42.pdf.

USGS. “Why Share Your Data.” *Why Share Your Data | U.S. Geological Survey*, USGS, 2009, https://www.usgs.gov/data-management/why-share-your-data#:~:text=Data%20sharing%20allows%20researchers%20to,wider%20regional%20or%20topic%20area.

Weinberg, Darin. “Standardization.” *Wiley Online Library*, John Wiley & Sons, Ltd, 15 Feb. 2007, https://onlinelibrary.wiley.com/doi/abs/10.1002/9781405165518.wbeoss247.