Evaluating the Efficacy of TMDL Implementation Actions on Fecal Bacteria Concentrations in Mill Neck Creek, NY

By Thomas J. Vogel

Honors Essay in Geography

Spring 2016

Department of Global Studies and Geography

Hofstra University

Advisor:

Dr. Craig M. Dalton, Department of Global Studies and Geography

Committee:

Dr. Kevin Bisceglia, Department of Chemistry

Dr. Linda Longmire, Department of Global Studies and Geography

<u>**Table of Contents**</u>

Abstract

In 2003, New York State's Department of Environmental Conservation (NYSDEC) established a total maximum daily load (TMDL) for pathogens in Oyster Bay-Mill Neck Creek, on the North Shore of Long Island. The TMDL used fecal indicator bacteria as surrogates for enteric pathogens believed to be entering the waterbody through improper or illegal sewage disposal, and subsequently diminishing its recreational and commercial value. Multiple pollution reduction strategies have since been implemented to comply with the TMDL, including a $13.2 million project to re-sewer houses near Mill Neck Creek's headwaters. To date, however, there has been little consistent study of the impact made by these expensive efforts on the pathogen levels in Oyster Bay-Mill Neck Creek – something this study aims to redress. To do so, we conducted a combined geospatial and temporal analysis of fecal bacteria concentrations in the waterbody over the past fifteen years. Although major reduction strategies targeted the headwaters of Mill Neck Creek, bacterial concentrations in the waterbody continue to exhibit pronounced spatial dependence that is indicative of elevated and persistent pathogen loading from that region. On the other hand, parametric and nonparametric analysis indicate that creek-wide bacterial concentrations are decreasing with time, and are significantly lower ($p < 0.05$) after 2012, when major sewer infrastructure upgrades were completed. Currently, efforts are underway to identify the cause of these reductions by accounting for natural variations attributable to flow, season, and tidal mixing.

Key Words: Mill Neck Creek, TMDL, fecal coliform, "The Birches", tidal creek, septic tanks

Introduction

Oyster Bay's beaches, marinas, and wetlands provide an essential economic, recreational and conservational service to the North Shore of Long Island. Shellfishing was once a prevalent industry in the Bay and provided substantial income to the region along with the influx of money associated with beach season and water sports activities in the area. In 2003, the US EPA issued a Total Maximum Daily Loads (TMDL) report for Pathogen Contamination in Oyster Bay. The TMDL was issued with a notice of closure for Mill Neck Creek, a section of Oyster Bay, for both shellfish harvest and recreational activities due to the excessive concentration of fecal coliform bacteria in the waters. Other sections of Oyster Bay were closed on a seasonal basis to prevent contamination during peak loading periods.[1] The intent of this study is to assess the impact of the TMDL on Mill Neck Creek's water quality, with particular focus placed on the re-sewering of Continental Villas, a 30 home community known colloquially as "The Birches". The infrastructure renovation and connection to Glen Cove municipal sewer system cost approximately $13.2 million and was completed in December of 2011.[2] Prior to the renovation the treatment of sewage for the community was handled locally and in an ineffective manner. The local system did not meet the demands of the community and ultimately release untreated sewage waste directly into Mill Neck Creek through the adjacent outfall. After the renovations to the infrastructure were completed in 2012 "the Birches" was connected to the Glen Cove municipal sewer system which had a functional wastewater treatment facility capable of filtering the additional volume, this is hypothesized to have created a decrease in fecal coliform contamination within Mill Neck Creek.

With numerous agencies, at all political levels, having monitored water quality in Mill Neck Creek it was first essential to audit the available data and understand the extent to which each agency performed testing. Federal and state agencies such as the Environmental Protection Agency (EPA) and Department of Environmental Conservation (DEC) provide the most stringent monitoring requirements and standards however the monitoring performed by these agencies was often on a rotating or semi-regular basis. In addition, many of the government level monitoring sites have been inactive for several years. Because the Friends of the Bay only have consistent records dating back to 2004 the scope of this study was restricted to a period of 11 years from 2004 through 2014. While the Friends of the Bay are consistent, the TMDL sites National Shellfish Sanitation Program which is federally regulated and performs testing through the New York State DEC. The National Shellfish Sanitation Program has actively monitored this area since 1987 however, the agency is only required to make 5 measurements annually at each site. Outside of these agencies there are numerous others that must be considered as potential data sources. **Figure 1** represents all of the monitoring agencies, both active and inactive, within Mill Neck Creek and Oyster Bay with the exception of the National Shellfish Sanitation Program. With little apparent communication between agencies in this area it is questionable whether the monitoring is being performed effectively and whether duplication of efforts is limiting the effectiveness of monitoring efforts. The audit of monitoring agencies also included a reconstruction of representative figures from the Friends of the Bay *Annual Water Quality Report*, a publication describing the state of the watershed and surface water in Oyster Bay. The figures presented in the report are not statistically driven and it is necessary to understand the methodology driving analysis and any assumptions made during processing, in particular regarding detection limits.

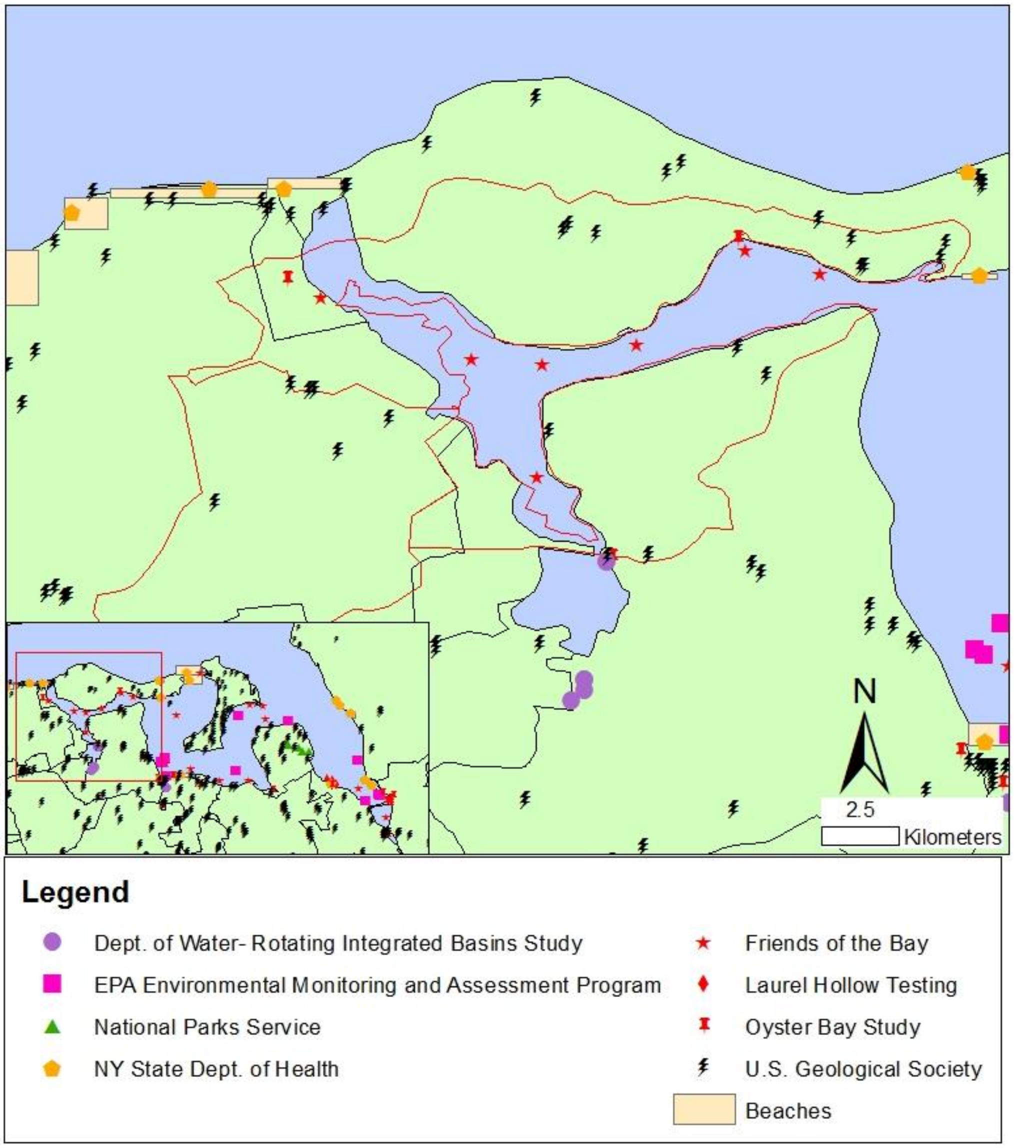


Figure 1: Map of monitoring sites within the Mill Neck Creek and Oyster Bay study area (excluding the NSSP)[3,4,5,6]

Mill Neck Creek, and all of the Oyster Bay water system, is subject to a complex web of political constraints beginning with federal regulations and expanding downward through the New York State and both Nassau and Suffolk County levels before reaching the local municipalities. Within the small study area there are several autonomous villages contained within the town of Oyster Bay, which can implement policy at the town level. Additionally the villages of Bayville, Locust Valley, Oyster Bay, Lattingtown, and Mill Neck also have the potential to craft policy within the area. Mill Neck Creek itself is subject to all of these influences with the exception of Suffolk County. The political complexity can be seen in **Figure 2**, a map of the study area designating the local political boundaries with direct influence over Mill Neck Creek. Along with the political boundaries, **Figure 3** designates the subwatershed boundaries which directly influence Mill Neck Creek. Water bodies such as Beaver Brook and Oyster Bay Harbor are also delineated in the figure to clarify potential freshwater sources and points of entry for fecal coliform contamination. **Figures 4 and 5** are respectively a land use map of Mill Neck Creek and a topographic and bathymetric map overlaying an aerial image of the study area. These maps were designed to assist in analyzing potential sources of fecal coliform contamination on a qualitative level. **Table 1** approximately quantifies the land use in Mill Neck Creek; this can be used with **Map 1** to better understand the distribution over the study area. Land use assessment will be useful in estimating the pathogen contributions outside of "the Birches" sewer outfall. Impervious surface, marine vessels, and wildlife all provide fecal coliform contributions which are not assessed directly but contribute to the loading indirectly. The map of political boundaries in the area will assist in speculation concerning potential inter-municipality policies and communications which may assist in improving the water quality in Mill Neck Creek.

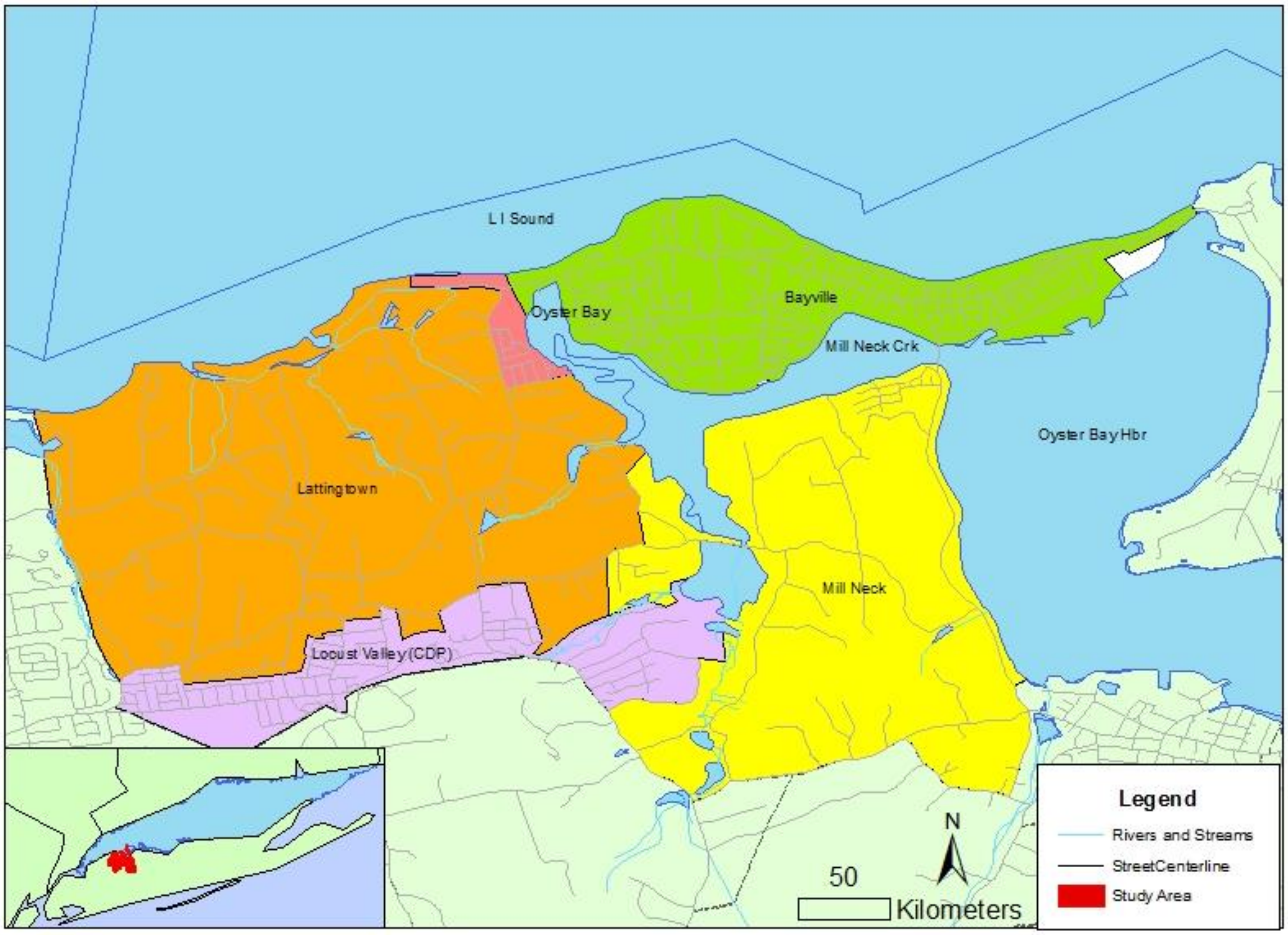


Figure 2: Map of the political boundaries within the Mill Neck Creek study area. The inset map indicates the boundaries of the Oyster Bay watershed in the broader Long Island Sound environment. Nassau County encompasses all of Oyster bay Harbor and Mill Neck Creek, the boundary between Nassau and Suffolk Counties occurs at the base of Cold Spring Harbor.[3,4,6]

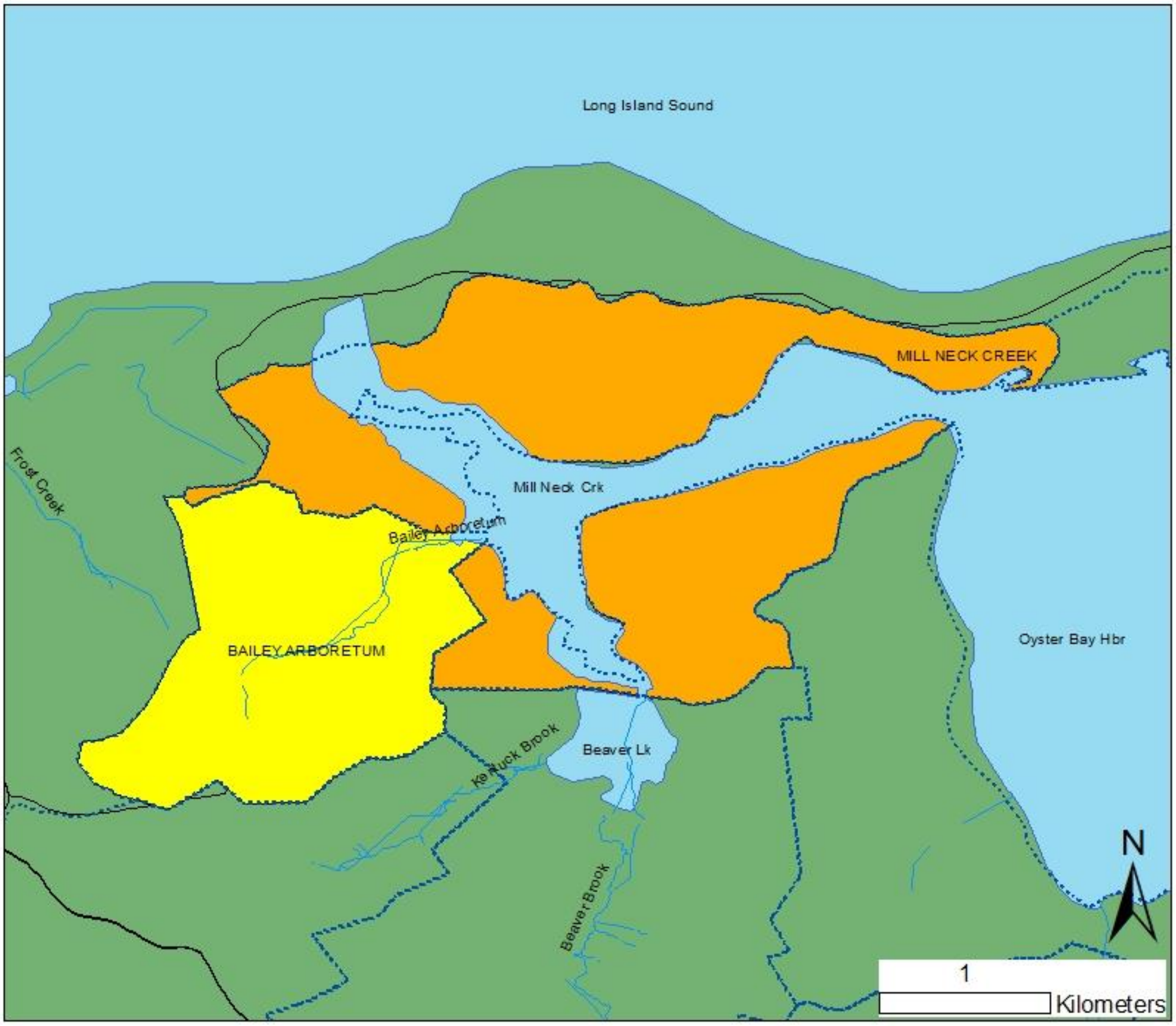


Figure 3: A map delineating the subwatersheds, Bailey Arboretum and Mill Neck Creek, which directly influence Mill Neck Creek. Freshwater sources such as Beaver Brook and significant water bodies, such as the Long Island Sound, are also identified.[3,4,6]

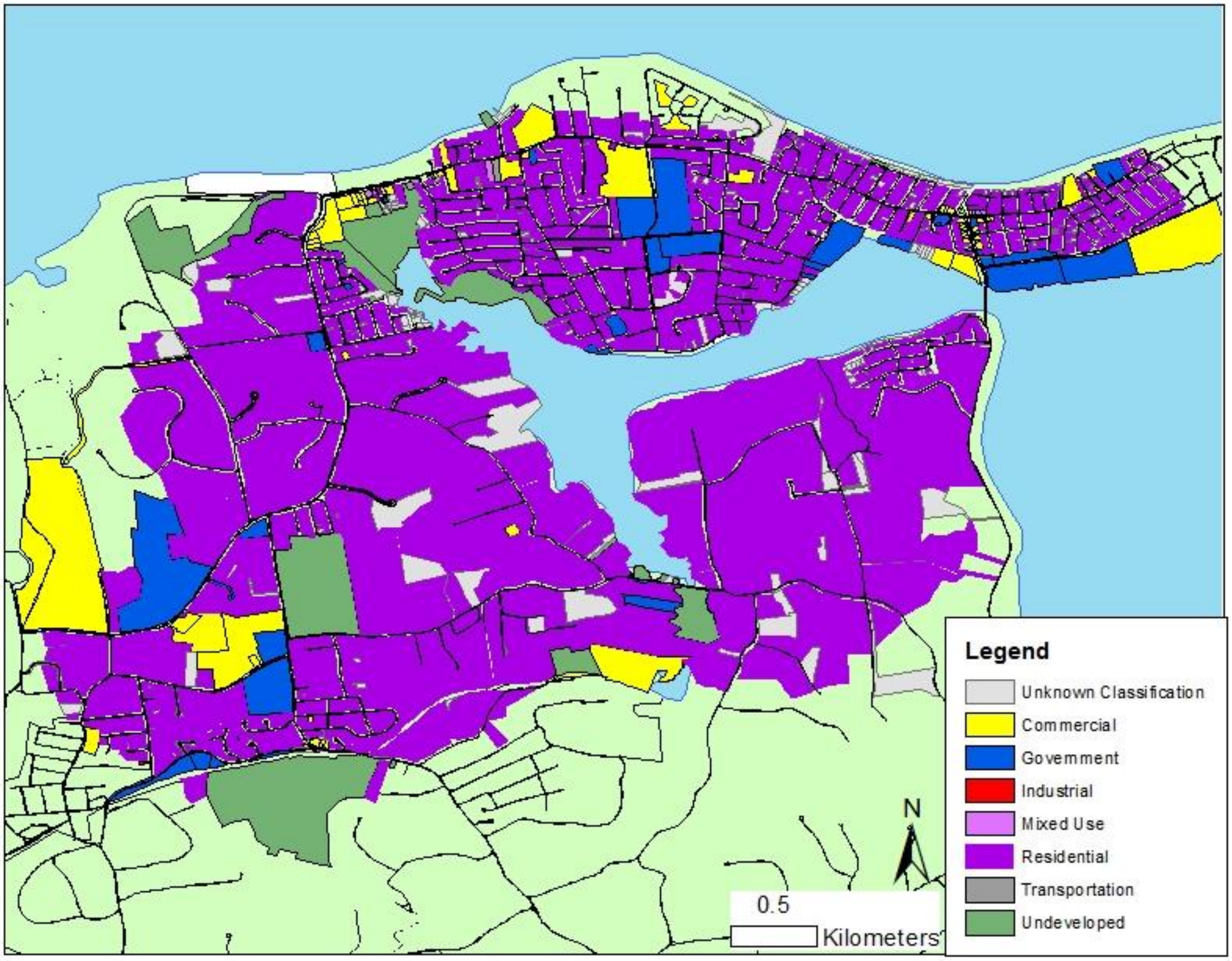


Figure 4: Land use map of Mill Neck Creek. This map highlights the highly residential nature of the subwatersheds.[3,4,6]

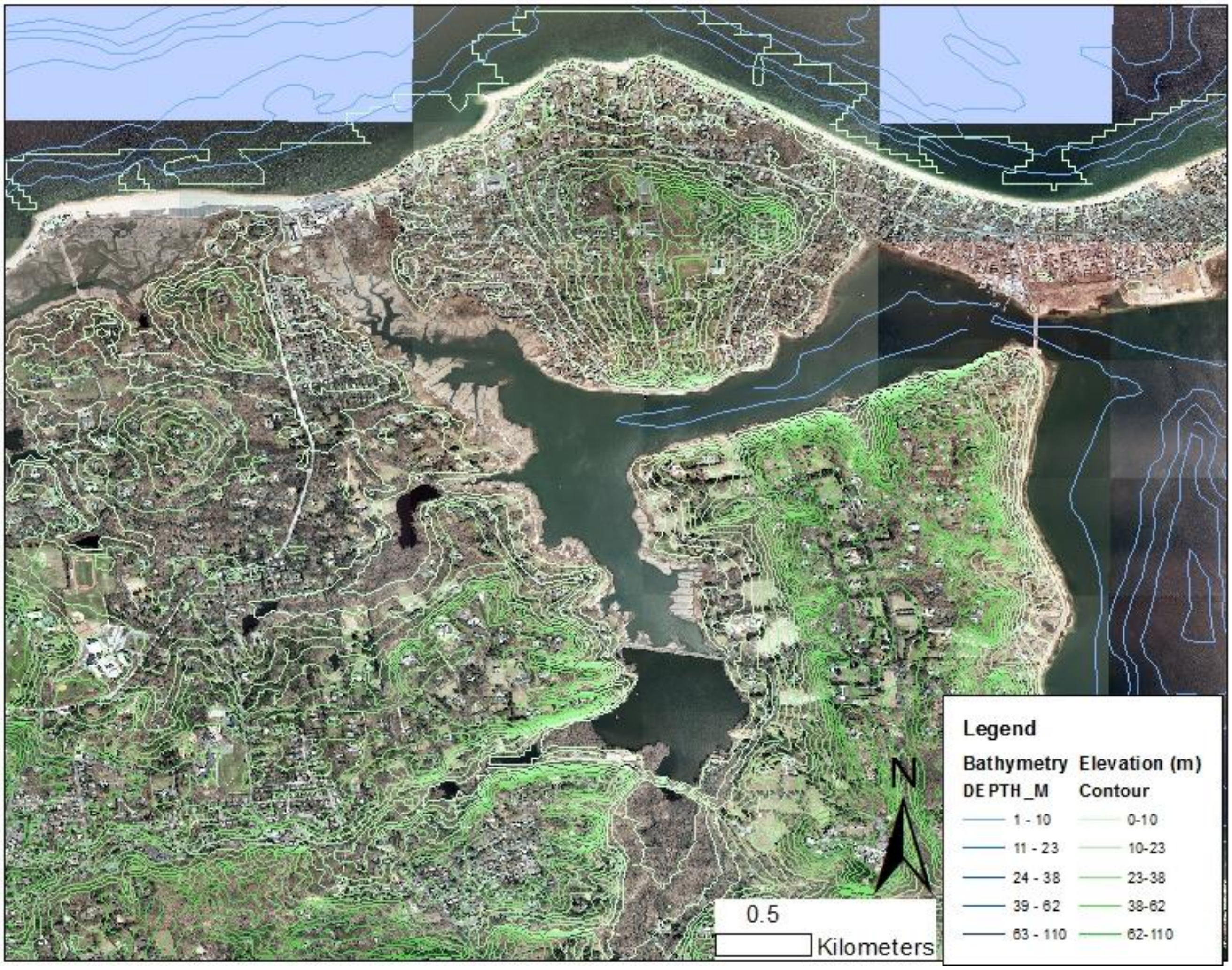


Figure 5: A topographic and bathymetric map of the Mill Neck Creek study area. Aerial imagery is also included to assist in land use assessment.[3,4,5,6]

Table 1: Land Use estimations in the Mill Neck Creek study area[1]		
Land Use	Area (acres)	Relative Distribution (%)
Residential	3185.51	65.30
Commercial	209.09	4.29
Industrial	0.00	0.00
Mixed Urban or Build-up Land	151.93	3.11
Forest Land	981.91	20.13
Agricultural Land	248.06	5.08
Water and Wetland	101.86	1.09
Barren Land	0.00	0.00
Total	4878.36	100.00

Several variables are discussed in this study, the most important being fecal coliform bacteria. Fecal coliform is a relatively innocuous bacteria however, it serves as an indicator for more severe pathogens. Sources of fecal coliform are animal and human fecal matter as well as human sewage and septage containing the same. **Figure 6** represents depicts, in a simplified way, the sources of fecal coliform contamination. Locally this can be sourced directly from

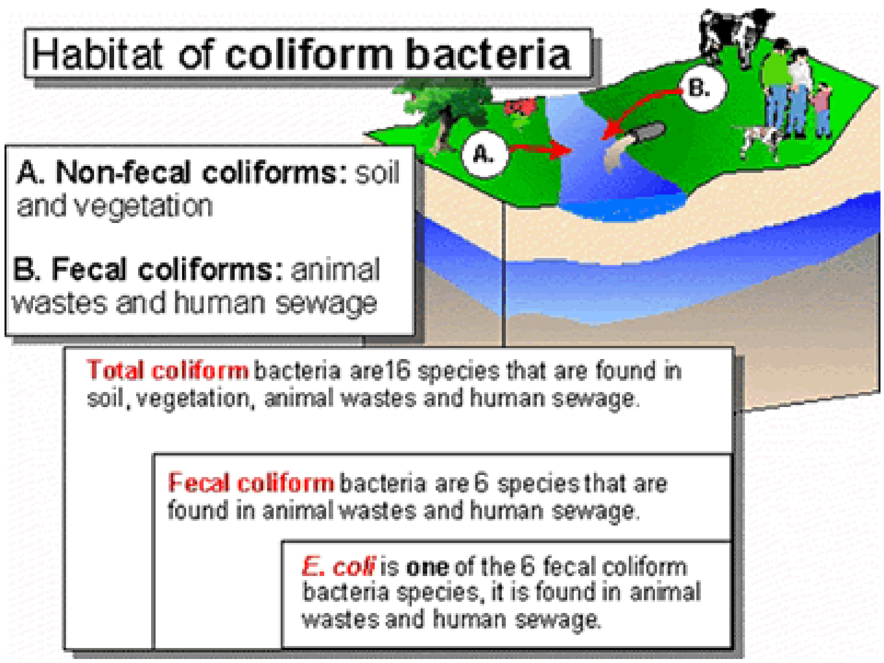


Figure 6: A representation of the sources of fecal coliform and simplified depiction of their transport methods to water sources.[7]

outfalls, septic tanks, animal feces deposited on grass or impervious surface, and direct deposition of feces into the water. Fecal coliform is most relevant to the study because it serves as the metric for swimming and shellfish harvest closures by the EPA. Impervious surface, a surface which cannot be effectively penetrated by water, is not assessed directly in this study but is a measure of human development and is used to make some assumptions concerning potential sources of fecal coliform contamination in Mill Neck Creek. Blacktop, cement, roofing, and roadways are all considered impervious surface. Salinity is discussed as an influence on the concentration of fecal coliform bacteria within a water body. Higher salinity makes the pathogen concentration decrease by making the environment less hospitable for the bacteria. Suspended solids are discussed in terms of future work and speculation but are not directly analyzed in this study, the exception being reconstruction of data. Secchi depth is a common metric for turbidity which in turn is a measurement of the suspended particles in water. This is relevant as fecal coliform has a high affinity for sediment and a portion of the concentration will reside in the sediment and suspended particles. **Figure 7** depicts common sources of water pollution in a mixed land use environment. From the image the most relavent aspects to this study are seepage which is contributed by septic tanks mainly, wastewater which is contributed by sewer systems and outfalls, and runoff which occurs due to rainfall.

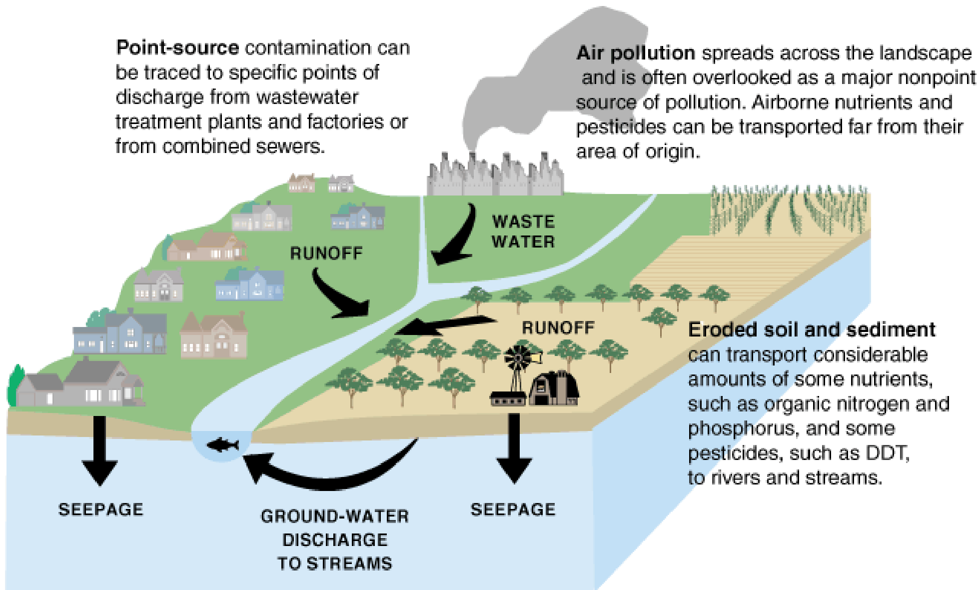


Figure 7: A depiction of common sources of water contamination in a mix land use environment. In this study runoff, seepage, and wastewater are the most relevant sources of contamination.[7]

Median fecal coliform concentration was selected as the representative value for statistical evaluation. Rigorous nonparametric statistical examination will be performed to determine if the data fits a lognormal distribution, as would be expected for environmental variables. Additionally, the statistical significance of fecal coliform change associated with "the Birches" re-sewering will be assessed using similar methods. If concentration change is determined to be significant surrounding the infrastructure update it is possible to attribute it to "the Birches". Goodness of fit analysis, linear rank sum, and test for tail shift are all rigorous methods of assessing the statistical significance of concentration changes. Using these tests it is possible to assess the impact of "the Birches" reform on Mill Neck Creek. This analysis is performed at the request of Oyster Bay/ Cold Spring Harbor Watershed Protection Committee, a joint committee designed for inter-town communications of water quality issues in Oyster Bay. The Oyster Bay/ Cold Spring Harbor Watershed Protection Committee is also connected with the Friends of the Bay.

Literature Review

Suburban development within tidal creek watersheds and estuaries results in a unique set of both point and non-point source contaminants to the water body. In 2003, the final TMDL was issued for fecal coliform contamination in Mill Neck Creek and Oyster Bay Harbor. The regions were cited as out of compliance with pathogen safety standards, Mill Neck Creek exceeded both the Shellfishing and Swimming standards issued by the EPA. The TMDL stipulates that the minimum safety standards were not met in Mill Neck Creek consistently and in specific regions of Oyster Bay Harbor on a temporary basis. The Waters of Oyster Bay Harbor are largely rehabilitated and, with the exception of Beekman Beach, meet the Swimming standard and throughout most of the season meet the Shellfishing standard as well. Outside of the DEC closures related to pathogen contamination, the Town of Oyster Bay maintains the closure of all shellfishing zones nearest the sewage treatment outfalls. This measure is meant to increase the acceptable response time to an incident rather than as a response to pathogen loading directly.[1] The assumptions made in the TMDL report are based solely on the available National Shellfish Sanitation Program data and do not encompass groups such as the Friends of the Bay. The TMDL stated that the Friends of the Bay data was excluded because it was both unavailable to the assessors and could not verify analysis by a DEC approved laboratory. The Friends of the Bay water quality reports were published annually, however these publications began after the TMDL activation. **Table 1** indicates the assessment of the USDEC concerning land use in the study area, Mill Neck Creek and Oyster Bay Harbor, and serves as a basis for the studies The sampling requirements by the National Shellfish Sanitation Program to define an analysis in any area is 5 APC measurements taken annually for 3 consecutive years. This provides the minimum number of measurements required to make a sound assessment of the data.[1] The use of only 5 measurements annually does not provide a detailed analysis of the regions pathogen characteristics, it only allows the extrapolation of a trend to determine the safety of the region. Total coliform level is also listed as the metric for contamination limits, however since the onset of the TMDL all monitoring agencies in the region have changed methods to consider fecal coliform due to greater accuracy and lower costs.

The TMDL establishes a precedent of pathogen contamination in both Oyster Bay Harbor and Mill Neck Creek as a result of both point and non-point sources of fecal coliform bacteria. In the study area there have been no published studies assessing the impact of stormwater runoff or sewage outfalls as sources of fecal pathogen contamination. In the absence of local studies, there have been several studies assessing stormwater, septic systems, sewer outfalls as pathogen source in Southwest Brunswick County, North Carolina. Brunswick County encompasses several tidal creeks which, when considered as a whole, provide a great deal of insight into the methods of pathogen contamination in tidal creek systems. The land use in Brunswick County is similar to that of Mill Neck Creek, largely residential and with a strong presence of septic rather than sanitary sewer infrastructure. For this reason, studies in this region will provide insight into the contamination methods in Mill Neck Creek. [8,9]

Cahoon et al. provided two studies concerning water quality in Southwest Brunswick, the first in 2006 followed by an additional analysis of the area published in 2016. In the 2006 study Cahoon et al. assessed the impact of septic tank failure against stormwater runoff as the chief cause of fecal pathogen contamination in the tidal creeks.

"Wastewater treatment systems in the estuarian watershed consisted almost entirely of septic systems for individual residences or businesses, with the exception of one small NPDES-permitted discharger (an oyster packing plant) and one non-discharge treatment facility permitted at 0.5mg/d serving a golf course residential community."[8]

Based upon this description, the treatment infrastructure is remarkably similar to that in place in Mill Neck Creek, with numerous septic systems and only a single SPDES located the Continental Villas, colloquially "the Birches", housing community in Mill Neck. The estuary, like Mill Neck Creek, was previously designated as a shellfish harvesting ground but had been closed prior to and throughout the duration of the study.[8] Also similar to Mill Neck Creek, "the large number of septic tanks in the SBWSA area precluded any possible inspection program to assess the overall frequency of poorly performing systems."[8] The statement indicates a strong uncertainty in the functionality of septic systems resulting from the inability to properly inspect all of them individually. This supports the conclusion of Cahoon et al. that stormwater was neither the most significant nor most important cause of fecal coliform contamination in the study area.[8]

The study performed regression analysis comparing the fecal coliform concentration during periods of rain within 48 hours (2 days) of rainfall. This analysis indicated very little correlation between the amount of rainfall prior to analysis and the concentration of fecal coliform at monitoring locations.[8] The small difference in concentration relative to rainfall indicates that runoff is not the main source of contamination. The lack of sanitary sewer systems in place, coupled with the suburban nature of the regions has led to the overuse of land for septic leaching , creating a lack of proper space for natural filtration of the human waste products.

"When septic tanks are closely packed, lateral movement of septage into drainage feature, such as man-made ditches or natural channels, such as the many small streams in the mainland portion of Sunset Beach, may be facilitated by rainfall but may continue even during dry weather. Stormwater drains, such as those at Sunset Beach, may drain septage directly if they are placed deep enough to lie below groundwater tables in areas where septic tank densities are high"[8]

This statement is likely applicable to Mill Neck Creek as well in that there are regions of higher suburban density which rely entirely on septic drainage. The improper development of stormwater infrastructure most likely augmented the already prevalent problem with septic failures. With overflow and lateral transfer of septage, the shortcomings of an improperly established stormwater management system would provide additional pathways for the septage to reach the receiving waters, increasing the coliform load beyond what would be provided by natural runoff. Impervious surface is a common metric for the effect of stormwater runoff on receiving waters. Cohoon et al. employed this technique with strong correlations where septic failures minimally influenced the environment but decreased reliability when septic failure was a stronger influence on pathogen loading. In the case of site ML 28, 7.8% impervious surface was present in the region and a concentration of 10 CFU/100ml, a very close approximation to the observed value of 11 CFU/100ml. In this case there was very little impact outside of stormwater and runoff into the waters at the monitoring location. In contrast to this location ML 10 and ML 11, with 8.3% impervious surface, had a predicted concentration of 16 CFU/100ml, in compliance with the state and federal standards, whereas the observed value was 46

CFU/100ml.[8] The disparity between observed and predicted concentrations is likely a result of failing septic systems and improperly established stormwater infrastructure. The leakage could be transported from the septic systems to the receiving waters in the estuary rather than being properly filtered via the leaching fields and passive filtration methods which occur naturally in soil and vegetation.

A later study by Cohoon et al. analyzed multiple potential sources of fecal coliform contamination in the Southwest Brunswick County including but no longer limited to septic systems and stormwater infrastructure. In this study a linear regression analysis was applied to the coliform concentrations in order to discern trends. In contrast to the previous study, a heavier emphasis was placed on land use and alternative methods of fecal contamination. Sewer and septic failures in the study area still provided a substantial portion of the contamination, and the stormwater infrastructure in place provided additional outlets for the waste products to enter the waterway as previously described but with increased regard for the environment as a whole. In the secondary study Cohoon et al. determined that the impact of both rainfall within 24 hours and within 72 hours was of little significance to the fecal coliform concentration.[9] In this case a regression analysis was performed on the data resulting in poor correlation coefficient, indicating that despite small contributions, other routes of contamination provide a better explanation for the trend. Cohoon et al. also explore the possibility of increases in groundwater outflow containing coliform contamination through a silicate metric, observing that the relationship between rainfall and silicate concentration does not follow any overt trends. An explanation offered in the publication is that the stormwater is partially absorbed by the soil, increasing outflow from groundwater near the surface with a moderate change in fecal coliform and silicate concentration noted.[9] In each of these cases the fecal coliform concentrations were independent of the rainfall for any duration studied, indicating that the majority of pathogen loading is provided by sources such as direct deposition and leaching rather than runoff. This trend is location specific but as a whole for the water body holds true. Cohoon et al. note that "Consequently, direct impacts of human waste generation appear difficult to avoid in coastal ecosystems, although the worst may be limited by effective system construction, performance, and enforcement."[9] indicating that though the impact of human development can be minimized, it is impossible to entirely remove pathogen contamination as a result human development.

Human development within the watersheds of tidal creeks and estuaries has significant impacts on the surface water quality in the region. In the case of suburban and more specifically urban development increases the amount of impervious surface dramatically through the construction of roadways, parking facilities, and construction of buildings while concurrently destroying vegetated green spaces which act as filters for pathogens and other contaminants in runoff. Mallin et al. assess the specific impact of impervious surface on fecal coliform They also considered the impact of salinity on the observed fecal coliform levels in the study. Mallin et al. determined that "acceptable microbiological water quality for these costal ecosystems occurs when percentage of impervious surface of a watershed is less than 10%, impaired microbiological water quality occurs above 10% impervious surface, and highly degraded water quality occurs above 20% impervious surface."[9] This analysis is fairly consistent with the conclusions of Cohoon et al. when considering only the impact of land use on microbial contaminants however, Cohoon et al. stipulate that while the impact is important to consider, it is far less substantial than the role of septic systems and stormwater infrastructure have greater impact[9]. Mallin et al. have determined the latter impacts are minimal for their watershed and therefore believe it is of optimal conditions to determine land use impacts directly.[10] Mallin et al. speculate that in conjunction with the higher percentages of impervious surfaces, domestic animals provide substantial contribution to the fecal coliform loading in his study area, encompassing estuary regions of New Hanover and Pender Counties, North Carolina.

> "Results of a 1990 census indicated that there were at least 60000 pets in New Hanover County, which roughly translates to about 1360kg of manure produced per day. Weiskel et al. (1996) found 10^3 fecal coliforms/g of dog feces; thus, dog manure represents a sizable potential fecal bacterialload to the receiving waters. A large portion of this manure is deposited on the landscape. Visual observation indicates that much of pet fecal matter is deposited adjacent to impervious surfaces such as roads, sidewalks, driveways, etc well as on public and private lawns near creeks and drainage ditches"[10]

This observation concerning animal waste deposition on or adjacent to impervious surfaces explains well the increased impact impervious surfaces in the region have on fecal coliform loads in surface water. The findings in this region are consistent with the estimated animal contributions in Mill Neck Creek, based on the TMDL report.[1] With a significant portion of the area serving as a suburban community there is an increased number of domestic animals depositing waste in the public parks and other regions which have direct interactions with the surface water. Mill Neck Watershed is approximately 20.13% forest land[1], this indicates that outside of domestic animal fecal depositions there is also an expectation of substantial wild mammalian deposition onto the landscape. These depositions could account for some of the background fecal coliform concentration not associated with septic and sewer failures.

Mallin et al. also briefly discusses the impact of salinity and nutrient presence on observed fecal coliform bacteria, indicating that there is an inverse relationship between the salinity of receiving waters and the measured fecal coliform concentration at the associated monitoring locations. With regard to nutrients, Mallin et al. found that the concentration on nitrate and other nutrients followed a consistently positive relationship with fecal coliform concentration. The most likely explanation of this relationship is that the origins of both the nutrients and pathogens are the same and therefore the two are carried in the same transport channels.[10] A common source of both nutrient and pathogen loading is mammalian waste which contains high concentrations of fecal coliform bacteria and nutrients such as nitrate and ammonia.[10] With regard to salinity, it was observed that the fecal coliform bacteria was more sensitive to salinity concentration changes than alternative pathogen indicators such as *Escherichia coli* indicating that the results of fecal coliform monitoring at sampling locations with moderate salinity are likely to project lower concentrations than if alternative pathogen indicators were used. Mallin et al. find that :

> "The population distribution of *Escherichia coli* was generally similar to that of fecal coliform bacteria. At high salinities, concentrations were very similar between the two indicators. However at oligohaline salinities *E. coli* abundances were noticeably higher than fecal coliform concentrations."[10]

The above findings indicate that while at high and low salinity values *E. coli* and fecal coliform bacterial indicators behave in very similar ways, at moderate salinities, such as regions of an estuary where greater tidal and freshwater mixing occur, the fecal coliform indicator provides a lower measurement due to greater sensitivity to changes in salinity. These observations are important to consider in analysis as they could provide insight into the variations in coliform concentration within the tidal creek that are not entirely consistent with estimations based on anthropogenic activities.

The impact of salinity gradients discussed by Mallin et al. could indicate that the decreasing fecal coliform concentrations observed as sites approach Oyster Bay Harbor from within Mill Neck Creek partially reflect a change in the aqueous environment rather than a decrease in contamination. Fecal coliform appears to have an increased sensitivity to salinity, more so than other pathogen indicators, resulting in decreased observed concentrations.[10] In the case of Mill Neck Creek it is likely that the influence of a salinity gradient is minimal as the salt content in Mill Neck Creek is very similar to that of Oyster bay Harbor but cannot be neglected as a potential influence of fecal coliform concentration.

Schoonover and Lockaby further discuss the relationship between land use and fecal contamination, though the setting of their study is the Piedmont region of Georgia rather than North Carolina. Piedmont, Georgia is not of a similar topography or geography to Mill Neck Creek but is useful to consider because of the urbanizing activity occurring near water sources. The inland location does not have an estuary nature and is therefore more useful in evaluating the impact of land use directly without the confounding effects of a salinity gradient[10]. The study concluded that the most important land use factors for predicting fecal coliform and nutrient trends are the amount of impervious surface within the study area and the amount of managed and unmanaged forest land within the watershed.[11] Highly urbanized or growing regions have the greatest percent impervious surface. Schoonover and Lockaby found that:

> "27% - 100% of the samples collected from 22 tributaries to the Chattahoochee River near Atlanta, Georgia exceeded the US EPA review criterion for fecal coliform… highest fecal coliform concentrations occur in sub basins with the greatest proportions of commercial and mixed urban land use/cover in Piedmont"[11]

indicating that percent impervious surface is a strong indicator of pathogen contamination susceptibility. It was also determined that in the Piedmont region, 95% of the fecal coliform variability could be explained by the percentage impervious surface. The analysis was also consistently accurate when applied to areas within Piedmont with very low impervious surface percentages.

> "Streams with high proportions of unmanaged forests and low IS (i.e., <5%) may reflect an estimate of natural wildlife fecal coliform inputs to the streams. However, our results indicate that streams draining watersheds dominated by managed forests had fecal coliform concentrations slightly lower than those of unmanaged, forest-dominated watersheds during stormflow but not base flow."[11]

This observed relationship indicates that in the case of low percent impervious surface, there is less fecal coliform deposition into the streams in both base and storm flow periods. The slight decrease in fecal coliform deposition in stormflow periods is likely due to habitat restrictions resulting from the management techniques for the forest. With a more controlled environment, it is likely that there is less influence of natural wildlife than in an unmanaged forest area. An explanation for the overall trend is related to the filtering potential of grass and forestland in opposition to that of impervious materials such as concrete and blacktop. Findings of the study indicated that pasturelands, like forestland displayed long-term trends of lower fecal coliform concentration in nearby water sources with small, short-term spikes related to stormwater deposition. Schoonover and Lockaby believe that the consistency of these lower concentrations is related to the dense grasses, which encompass pasture lands and flora living on the forest floors.[11] They report:

> "Coliform bacteria have also been reported to preferentially bind with sediment, thus the dense grasses may have facilitated bacterial settling in the terrestrial environment (Schillinger and Gannon, 1982). Moreover, the Center for Watershed Protection (2000) reported that 15% - 30% of fecal coliform is adsorbed to sediment or larger suspended particles, whereas, approximately 50% of the FC remains in suspension and effectively functions like fine clay particles."[11]

The above statement indicates two important considerations in analysis. The first is that sediment rich vegetated areas are better able to filter fecal coliform prior to it entering surface water, this reduces concentrations and can mitigate some effects that impervious surface causes if proper consideration is taken when planning the development of costal suburban and urban developments. The second important fact is that the analytical methods, though effective for the suspended fecal coliform concentration, do not account for a significant part of the potential pathogen contamination because they do not analyze the settled sediment concentrations. For this reason it is important to consider the measured calculations as accurate but incomplete. They provide an underestimate of the fecal coliform contamination in surface water.[11]

As a whole Mill Neck Creek is a residential area with varying medium to low housing density throughout the regions. Several parks and conservation areas are also present in the area, including Mill Neck Preserve, which occupies a large portion of the creek's northern branch. A small number of commercial properties are present in close proximity to the study area; these encompass restaurants, the Bayville Adventure Park, a motel, marinas, and rental facility for water sports equipment. Septic systems are prevalent throughout the area, acting as a barrier to higher population density. The subwatershed of Mill Neck Creek are estimated to have approximately 10% impervious surface with the exception of Bayville and the Hamlet of Oyster Bay which have an estimated 30% impervious surface.[1] Based on this analysis it is likely that impervious surface will have a minor influence within the study area as there is minimal impervious cover in the region. Bayville and the Hamlet of Oyster Bay will have a greater impact with decreased natural filtration however, this is still a small area of influence relative to the areas . Despite increased green space in the study area, a portion of this provides pathogen contamination through indirect contact with the small areas of impervious surface. According to the TMDL, 46% of households in the study area have pet dogs.[1] Not accounting for other domestic and wild animals, dog feces contribute 10^3 fecal coliforms/g of waste most commonly near road or walkways.[10] The direct deposition of fecal waste in such close proximity to impervious surfaces increases the potential impact through the removal of natural filtration channels. The conservation areas, in particular Mill Neck Preserve, are likely to act as a natural filter for runoff which passes through them as reported in previous studies concerning land cover impact on pathogen contamination.[8,9,10] Grass acts as a particularly important filtering agent as fecal coliform has a strong affinity for sediment particles. In the case of runoff this means that barren land or sand without vegetation facilitates greater transport of pathogens than grass fields or forests where the particles can be captured in the leaves.[11] Oyster Bay's TMDL indicates that no barren land exists in Mill Neck Creek. This represents a risk of minimal deposition of sediment adsorbed fecal coliform contamination. With regard to "the Birches", it is likely that the sandy terrain increased pathogen contribution due to sewer failure. Satellite imagery of the land between the community and the waters of Mill Neck Creek appears show sand with minimal vegetation. This environment, likely washed over by the rising tide, would introduce additional fecal coliform that is adsorbed to the sediment because of tidal mixing.[11]

Another non-point source discussed in all articles is direct deposition of waste into surface water. The TMDL for Oyster Bay addresses this directly with regard to private vessels discharging waste directly into Mill Neck Creek and Oyster Bay. Officially, this practice is illegal and prohibited in the entirety of Oyster Bay however, during holidays the waters near marinas are closed to bathers in anticipation of such behavior.[1] There is also a seasonal contribution from domestic water fowl which is most substantial between late spring and fall. The contribution of water fowl is consistent and should be considered minor when compared to the contributions of previously examined sources. Illicit discharge seasonally provides a sizable load of pathogens to the region during the peak seasons but has little impact throughout the year, as the weather is not conducive to recreational boating. In terms of the analysis both of these sources can be considered background as they should be relatively constant through the years and therefore impact the periods before and after 2012 approximately equally.

The previous studies relied on ANOVA and regression analysis to determine trends in fecal coliform concentration and assess relationships with other variables such as salinity, nutrients, and impervious surface. These regression techniques are valid in a normalized environment but do not account for the natural variability, which is inherent to the tidal environment of estuaries and tidal creeks, such as Mill Neck Creek. Beck and Hagy have adapted the Weighted Regression with regard to Time, Discharge, and Seasonality written by Hirsch to account for these tidal variations. Hirsch developed his techniques to analyze environmental variables in streams with reported flow data, the method can interact with STORET and the USGS repository as well as accepting personal data sources entered manually.[12] Where Hirsch falls short is in accounting for tidal variations associated with the estuary ecosystem. Beck and Hagy designed an adaptation of the Hirsch model to analyze chlorophyll in Tampa Bay.[13] The most significant change made by Beck and Hagy to the Hirsch model is the ability to use a salinity metric in place of reported flow data. To use a salinity metric the conversion formula

$$Sal_{ff} = 1 - \frac{Sal_{obs}}{Sal_{ref}}$$

must be applied which accounts for flow rather than simple salinity gradient. For consistency of analysis it is also important that measurements be taken at the same time or at the same tidal stage to ensure that all flow metrics are self-consistent.[13] By using the salinity metric in place of flow data it is possible to observe trends associated with the estuary flow and specifically the influence of stream flow rather than simply accounting for the discharge of the stream. . It also better accounts for mixing which occurs in a tidal environment that is not present in an inland stream environment. The use of quantile regression techniques rather and matrix analysis also better accounts for natural variability than the linear regression techniques applied in other studies because there is a greater ability to cope with left censorship and non-normal distributions. Survival regression is the applied matric method, which is used to minimize the impact of censorship on the dataset.[13] An important assumption to note with Beck and Hagy's analysis is that salinity is uniformly distributed in the study area.[13] This assumption implies that salinity variation is only due to the influence of the tidal and freshwater flows and not a gradient throughout the area. If salinity in the study area is not approximately uniform it is better to create segments of the estuary and analyze each individually instead of treating the entire region as a single, uniform tidal entity.[13]

> "The improvement in performance emphasizes that traditional regression assumes model parameters are constant throughout the time series, whereas WRTDS allows for dynamic interactions between response and predictor variables. Improved model fit through flexible parameterization increases the ability of the model to describe historical patterns, but reduces applications to predict future chl-*a*. If drivers of chl-*a* are changing over time, predicting future chl-*a* while assuming the drivers are not changing could be of limited value.[13]

The statement by Beck and Hagy identifies the strengths of the adapted regression analysis to determining patters in a long-term existing dataset, such as the impact of "the Birches" new sewer infrastructure or existing land use changes. In contrast to analyzing existing trends, applying this analytical method to make predictions for the future would not produce strong results as the main assumption for the analysis indicates that influencing factors are changing throughout time and therefore should not be treated as constants for the purpose of trend extrapolation.[13]

Considering the waste disposal infrastructure in place in Mill Neck Creek with regard to New York City, it is apparent that there are many modifications which would make the system more sustainable and decrease pathogen contributions into the environment. In 2006, New York City convened a committee consisting of representatives from several water and environmental agencies within the city with the intent of planning sewer and water reforms which will be sustainable with regard to the imminent and approaching climate changes. Existing infrastructure was constructed with the intent of longevity and adaptability to the eventuality of climate change however, it was also determined that reevaluation of the systems and development of a long-term plan was essential.[14] Proposed modifications to the system include improvements to the tidal gates which minimize direct release of sewage and stormwater into the waters surrounding New York City during weather events which raise tidal levels above standard.[14] Application of a similar system would be beneficial in Mill Neck Creek as it is likely that as sea-level rises and tides rise to greater heights, water will infiltrate the existing septic structures. Cohoon et al. discuss the implications of closely packed septic tanks on leachability into ground and surface water[8] indicating that the close proximity of septic tanks leads to increased risk of fecal coliform contamination in nearby waters. Replacement of the septic systems which are likely beyond the ability of any agency to adequately inspect for proper function[1] the installation of a sewer system, as New York City has developed, would provide many environmental benefits. Sewers would also increase the longevity of waste disposal in the home and decrease the maintenance related to backups, failures, and leakages. While a large portion of the region has adequate land for proper leaching fields due to the larger plots in wealthier areas, towns such as Bayville will have a larger septic density without necessarily providing proper leaching space.[8] The adoption of New York City's system of analysis and course of action would benefit both the citizens and environment in Mill neck Creek substantially in both the near and distant future.

The retention of septic systems in the Mill Neck Creek area is most likely motivated by two forces. The first is the potential for increased population density following a sewer connection. The current septic systems are not capable of accommodating the volume of sewer waste produced by apartments and larger multifamily dwellings, restricting the potential growth in the area to low density suburban orientation. With the installation of a public sewer system the increased volume could be processed and therefore more dense housing could be created. The second, and more substantial, motivation in retaining septic systems is the cost of installation and increased annual costs towards maintenance.[2] Burden et al. discuss in detail the initial and annual costs of common septic and sewer systems as well as minimal physical requirements for both. The study occurs in Sarasota County, Florida which is costal and of comparable physical conditions to Mill Neck Creek. Detailed tables of costs for both systems, as well as municipal costs for Nassau County, are provided in **Appendix 1** for reference. According to Glen Cove town code, the costs associated with sewer installation are the responsibility of the residents requesting connection including installation, applications, and repairs of any related damage to town property.[15] **Appendix** 1 also contains the table of Nassau County applications and associated fees required for public works, including applications for sewer installation and inspection.[16] Despite residents bearing the burden of construction costs, Burden et al. reported that at property sizes smaller than 0.5 acres the cost of all three sewer alternatives was less than or equal to that of the septic options examined. The exception to this being low density developments where property sizes are greater than 0.5 acres, in those cases it was cost prohibitive to connect to public sewer systems.[15] The decreased cost associated with sewer conversion is not indicative of decreased service and maintenance, the opposite is likely true. Septic service and replacement is a responsibility placed on the homeowner including any repair costs associated with backup or failure of the system. In contrast, a public sewer system is maintained entirely by the town, partially financed by the municipal connection fees levied on the homeowner.

The most likely reason for coordinated increase of septic costs with population density is the demand for increasingly complex systems to accommodate the smaller space which the system can occupy. In low density regions a traditional and minimalist cesspool system is easily accommodated within the property, the article recommends a 500 square foot drainage field for a conventional system, without requiring additional filtration or pumping measures. As population density decreases the demands on the system are increased as organic filtration of the septage is no longer possible due to space constraints. In this case the complexity and cost of onsite septic systems increased with the addition of additional filtration media and pumping requirements. Many existing septic systems within the study area did not meet the requirements, particularly in devoting appropriate space to the drainage field.[15] Another issue associated with septic tanks is the reliance on homeowners to inspect and maintain the systems without routine inspections. The burden of care for septic systems rests entirely with the homeowner and it is therefore the possibility for a malfunction to go unnoticed or unaddressed increases. With a public sewer system the municipality is responsible for maintenance and as a whole the public is more vocal with their concerns than if the burden rested with them personally. The annual maintenance cost in medium to high density regions is also substantially lower for sewer systems than septic systems[15], largely because the cost is furnished by many home owners rather than a single individual. In Mill Neck Creek, specifically with regard to Bayville and the Hamlet of Oyster Bay, the population density is sufficiently large to warrant sewer infrastructure. Resistance in these areas is likely due to the immediate material cost, the cost to complete the renovation of "the Birches" was $13 million[2], is daunting but ultimately will provide increased reliability, less susceptibility to flooding, and an overall decrease in cost.[15]

Methodology

Description of All Active and Inactive Monitoring Entities Within Mill Neck Creek

Federal Organizations

The Environmental Protection Agency (EPA) monitors in a semi-regular manner the water quality of the long island sound and Oyster Bay region. Due to steady declines in funding for the organization the EPA does relatively little sampling in recent years however they were responsible for a significant quantity of data obtained between the years 1990 and 2006 through their Environmental Monitoring and Assessment Project (EMAPs). After 2006 water quality studies were transferred to the EPA's National Aquatic Resource Survey (NARS).

1. EMAPS monitors for both biological and chemical factors in waterbodies however the data is not consistently obtained over several years. The data obtained for the long island sound consists only of samples between January of 1993 and December of 1994 which does not provide a long term analysis of the water quality in the region. The chemical data includes the monitoring of persistent organic pollutants such as PCB's and pesticides as well as metal concentrations. The biological analysis measures the presence of many strains of bacteria, microorganisms, invertebrates, and fish species present in the water body.

   EPA data including EMAPS and NARS information are obtained through the EPA database STORET which is a repository for unprocessed data from many governmental and private organizations performing water quality analysis. The data is unprocessed and therefore the analytical methods used by EPA organizations are not made apparent through their data sets.

2. The National Parks Service monitors for physical data such as temperature and tidal conditions, chemical properties such as pH and salinity, and levels of fecal coliform bacteria in the water. There is a very limited number of National Parks Service testing sites in the study area however the additional data will assisting in understanding long term patterns in the study area since legacy data as well as more recent data is available through the EPA STORET database. Legacy data is defined as published and unpublished water quality measurements taken up to the year 1998 when the data was brought to the active STORET database. The legacy data is no longer updated but contains a large repository of data not otherwise available to the public.

3. The National Shellfish Sanitation program (NSSP) is a responsible for monitoring and certifying the safety of seasonal and permanent shellfish harvesting sites throughout the United States. The NSSP works in conjunction with the Department of Environmental Conservation to measure fecal coliform, and in the past total coliform, pathogen levels which are an indicator of human and mammalian fecal contamination in the water body. The NSSP has been monitoring in Oyster Bay and Mill Neck Creek for approximately 29 years continuously through the DEC. The NSSP uses the swim standard of 200MP/100ml and shellfish harvesting standard of 14MPN/100ml. Exceeding the shellfish harvesting standard as a 30-day running geometric mean will enact a closure of the location as a shellfish farm until further analysis is completed,. Exceeding the swim standard will

enact a closure of the water body for both shellfish harvesting and recreational activity as controlled by the EPA BEACHES program. With the exception of the years 1998-2001 there is a continuous table of monitoring data for fecal coliform, 1998-2001 were years in which monitoring procedures were only capable of recording total coliform concentrations. During the period of analysis the NSSP measurements consistently monitor for fecal coliform bacteria only and are therefore compatible with the Friends of the Bay dataset.

State Level Agencies

4. The New York State Department of Environmental Conservation is also responsible for monitoring through the Department of Water Rotating Integrated Basins Study (DOW-RIBS) which assesses water quality of all New York State water basins on a five year rotation. This program takes measurements on a regular schedule for each basin including the Long Island Sound, including some data for Oyster Bay and Mill Neck Creek, however the five year rotation results in infrequent and disjointed results. The lack of consistent measurements makes the data less meaningful in the study since no patterns can be drawn over such a long and infrequent period. The data in the study area specifically is also very limited since the data encompasses all water bodies in Northern Long Island and mainly concerns the main body of the Long Island Sound. The DEC monitors for physical parameters such as pH and Salinity, as well as benthic organisms, and bacteria through the Department of Health.

5. The New York State Department of Fish and Wildlife monitors water bodies including Mill Neck Creek and Oyster Bay for their suitability as shell fishing locations. Fecal Coliform bacteria levels are used to determine shell fishing suitability and therefore obtaining this data would increase the number of data points for bacterial monitoring substantially. This data is monitored on a five year rotation similar to the RIBS measurements and is not readily available. In future it would be beneficial to attempt to obtain the measurements from the last 10-15 years in order to increase the spatial resolution of data in the study area.

6. The New York State Department of Health (DOH) monitors bacterial levels for the EPA Beaches Program. Nassau County is responsible for monitoring the beaches and therefore provides the samples to the Department of Health, however all data is then reported to the EPA for publication in the annual beach closure reports record the closure of beaches with duration and approximate reasons for the closure.

Local Organizations

7. Nassau County is responsible for the testing and monitoring of water quality of beaches in the area and tests regularly for fecal coliform bacteria in these locations. The data is then reported to the EPA Beaches Program to be recorded in database with other New York State beach closures. While information about beach closure data, duration, and approximate reason for closure is reported there are no direct reports of Fecal Coliform bacteria levels that can be used to quantify the closure or provide data resolution in contour production or statistical analysis other than knowledge that the measured level exceeded the NYS standard level to ensure swimmer safety.

Non-Governmental Organizations

8. The Long Island Sound Study is a larger Non-Governmental Organization which monitors the Long Island Sound. Their aim is to maintain and improve the quality of water within the Sound and to attempt to understand what causes the changes in chemical, bacterial, and toxin levels within the water body. Their second major purpose is to provide funding through research grants to other education and research organizations studying the Sound. The Long Island Sound Study is not only active on the Long Island side of the Sound, it also encompasses the greater New York area and the Connecticut coast which also have direct contact and reason for concern over the water quality of the Long Island Sound.

9. Friends of the Bay is a volunteer based organization focused on monitoring water quality in Oyster Bay. Samples are taken on a weekly basis, weather permitting, for the duration of the sampling season. The exception to this practice is the nitrogen compound series which is sampled and analyzed on a monthly basis through an outside laboratory. The sampling season for Friends of the Bay is defined as April through October each year and includes data from 19 testing locations throughout the Cold Spring Harbor, Oyster Bay, and Mil Neck Creek region.

   Friends of the Bay monitors weather conditions, water surface and tidal conditions, and wind parameters on a qualitative level with values defined from a numerical ranking system. Conditions include cloud cover, weather, wind speed and direction, water surface conditions and color, wave height, tidal stage, and rainfall in the past 24 hours.

   Quantitative measurements are taken for salinity, dissolved oxygen, pH, and water temperature for each testing location. The measurements for these parameters are taken at three depths: 0.5m from the surface, 1.0m from the surface, and 0.5m from the bottom. Secchi Depth is also measured using an average of two measurements for each testing location. One reading of the Secchi depth is defined as the average of the depth at which the disk is no longer visible during descent and the point at which it becomes distinctly visible again upon returning to the surface. Fecal Coliform and Enterococci presence is also monitored on a weekly basis and the quantity observed is defined as the log of the most probable number per 100ml [log(mpn/100ml)] in both cases. This method of reporting bacterial results is consistent with state and federal agencies and therefore permits the comparison of bacterial data from Friends of the Bay to that of the NYS Department of Health, EPA, and The Department of Environmental Conservation.

   Nitrogen data includes the monitoring of ammonia, nitrate and nitrite collectively, and Total Kjeldahl Nitrogen (TKN) as well as the calculation of Organic nitrogen by subtraction of the ammonia concentration from the TKN value and total nitrogen which is defined as the sum of the TKN value and the concentration of nitrate and nitrite collectively. The analysis of nitrogen compound concentrations is performed by an outside laboratory on a monthly rather than weekly basis.

   Friends of the Bay has several methods of analysis for the measured data. The major form of analysis is seasonal arithmetic or geometric means where a season is defined as the period from April to October where testing is performed. For physical parameters such as water temperature, pH, etc an arithmetic average is taken of all testing sites for the duration of the testing season. The nitrogen data is also analyzed by an arithmetic mean for the season but is considered on the scale of defined subdivisions of the bay (Cold Spring Harbor, Oyster Bay, and Mill Neck Creek) which have a defined set of testing locations associated with them. Bacterial data is considered on the level of individual testing locations by producing a running 30-day geometric mean of the data. The definition of this is the $\sqrt[13]{(x_1 * x_2 \ldots x_n)}$ for a period of 30 days prior to each data point. The geometric mean of the data is also compared at the level of each subdivision of the larger body Oyster Bay, in this case the geometric mean of all bacterial data is taken for each testing location and then from these geometric means a geometric mean is taken for all stations defined as a particular subdivision (i.e. Cold Spring Harbor= FB1, FB2, FB3, and FB4 collectively)

A table of the testing locations for friends of the bay is included with the site name and decimal degree latitude and longitude location for all 19 Friends of the Bay (FB) sites. This table includes the location of 19 permanent friends of the bay testing locations, three locations in Laurel Hollow (LH) included in the study for specific years at the request of Nassau County, and the Oyster Bay Study (OBS) which monitors streams and outfalls in the immediate area of oyster bay at eight permanent locations and one rotating locations. The Oyster Bay Study sites are monitored by Friends of the Bay but at a less frequent interval.

After considering the consistency, duration, and quantity of measurements made by each organization and the fact that many are now inactive for several years it was essential to use data from the Friends of the Bay and the National Shellfish Sanitation Program. Both of these agencies have a monitoring record greater than 10 years and monitor on a regular basis for multiple sites in the Mill Neck Creek study area. The Friends of the Bay monitors weekly ten months annually and the National Shellfish Sanitation Program monitors on a random sampling basis throughout the year. Salinity data will be obtained using STORET for a Connecticut Department of Energy and Environmental Protection automated monitoring site.

Preliminary Evaluation of Data and Reconstruction of Figures by the Friends of the Bay

The Friends of the Bay provide the most consistent source of water quality monitoring in the study area therefore a substantial effort was made to understand the results published in the organizations state of the water shed report. Recreating the figures provided in the report provided greater insight into the monitoring methods, data processing, viability of results, and systematic treatment performed on all data sets used in the Friends of the Bay studies. In this study, this data was used to assess the monitoring activity of the Friends of the Bay for auditing purposes as well as to guide the hypothesis formation concerning “the Birches”. The most conclusive method used to assess the results was a recreation of representative figures and charts taken form the *Annual Water Quality Reports* for each variable and method of processing. There are three major methods used by Friends of the Bay to describe and analyze their data applied to approximately three different groupings of data. The first mathematical treatment is a straight arithmetic average of the values for each variable defined by the formula:

$$(1) arithmetic\ mean = \frac{\sum_1^n x}{n}$$

. This is used for all physical and chemical variables measured. The second method is a straight geometric mean of all values for a particular variable. A geometric mean is defined as

$$(2) geometric\ mean = \sqrt[n]{x_1 x_2 \ldots x_n}$$

The final mathematical treatment is a 30-day running geometric mean taken for bacterial data. The geometric mean was taken in the same form as Equation (2) with the exception that the mean was taken using for each individual data using measurements occurring up to 30 days prior to the measured value. This is the standard method of measurement used by state and federal agencies to determine impairment of a water body in terms of shell fishing viability and swimming safety. The data evaluated was presented in three ways throughout the report; as data for each individual site used to compare the 19 testing locations in the greater Oyster Bay area, as sectional means (both arithmetic and geometric depending on the variable) to assess Mill neck Creek, Oyster Bay Harbor, and Cold Spring Harbor in relation to each other, and as a total seasonal average for all sites. In the case of the Friends of the Bay a seasonal average is defined as the mean of all data for the given variable over the course of the entire testing cycle, loosely defined as April to October.

The data obtained by the Friends of the Bay is both the most abundant and most consistent source of information concerning the Mill Neck Creek and Oyster Bay regions of the Long Island Sound. For this it was imperative to understand the methods used by the Friends of the Bay to analyze and report their data in the *Annual Water Quality Report*. Using the data from 2006, 2007, and 2009 a set f representative figures was selected to base the understanding of systematic treatment of data and reporting methods on. This set included one representative table or chart from each type of data analysis provided. The comparison was performed by first digitizing the desired plots in order to obtain an approximate value for points or bars when the label was not provided. The data tables for the desired parameters were then exported to a separate excel file to make manipulation of the tables simpler. Through this method it was possible to understand the treatment of non-detect values and values which were exceeding or below the detection limit. It also provided an understanding of some terminology used to describe the data and its treatment.

Geospatial Analysis using ArcGIS

The data availability is not the only concern with this small region of the Long Island Sound. Despite several organizations taking measurements with varying degrees of activity over the last 15 years there is little inter-agency communication and the result is a fragmented dataset and incomplete understanding of the results. The GIS component of this study has two purposes; the first is to represent the Mill Neck Creek and Oyster Bay study area in all of its aspects including the natural physical features, the features of its development, and political boundaries. The second aspect of the geospatial analysis is representation of the study area through the presentation of data sources and whenever possible representations of the data itself. Layers of point data for the Friends of the Bay are joined to annual data tables for each year enabling interpolation in future work if it is deemed useful to analysis.

Maps delineating both the watersheds and subwatersheds were also produced by layering the watershed and subwatershed boundaries above a map of land and water features. These maps also name and designate the streams and waterbodies in the study area and are clipped specifically to the Mill Neck Creek study area. A map of elevation and bathymetric contours was also produced using a land elevation raster produced by the USGS and the bathymetric contours were added to the map but were produced by the University of Connecticut and downloaded from the DEEP database. The elevation and bathymetry map was set with a background aerial image of the study area to better demonstrate the terrain and development in the area.

A land use map was produced using a census tax block map and the title information provided by municipalities in the study area. This map is classified at three levels to ensure that all use aspects are represented at the clearest level. The attached figure has buildings classified as Class I which is the broadest level incorporating the categories Residential, Industrial, Commercial, Government, Transportation, Undeveloped, Mixed Use, and Unknown while Class II and Class III go into greater detail about each subset for example the level of government for which a building functions (federal, state, Town of Oyster Bay, etc).

Data Compilation

For the purpose of this project, fecal coliform is used as a metric for human waste contamination in the water supply. While there are an abundance of agencies which provide short-term monitoring at various intervals, only two agencies provide a reasonable amount of resolution within an appropriate temporal range to assess "The Birches" hypothesis, the Friends of the Bay and the National Shellfish Sanitation Program.

The national Shellfish Sanitation program dates back to 1987 and monitors for fecal coliform, salinity, and temperature. These sites are monitored on a consistent, random basis but at a low frequency of collection. The spatial resolution of this data is good within the region of Mill Neck Creek where it joins with Oyster Bay Harbor, however there are a limited number of points within the two arms of the creek. Despite having a fair number of sites within Mill Neck Creek, there are less than ten samples taken annually for each site which indicates that since being designated as impaired, there is little effort being made to assess improvements made within Mill Neck Creek. A significant benefit to the National Shellfish sanitation program dataset is that it is federally controls, meaning that all measurements are held to a strict standard and are required to follow a highly specific methodology including guidelines for reporting non-detects and unmeasured values.

The Friends of the Bay began monitoring later, consistent records for Mill Neck Creek date back to 2004 however, the Friends of the Bay monitor all of their sites weekly (weather permitting) approximately nine months annually. While there is a gap for the winter months each year, the data collected has a greater volume for each specific site, which increases the efficacy of the analysis. The Friends of the Bay also have a broader range of variables being monitored including but not limited to depth, salinity, fecal coliform level, and temperature. Limitations to the measurements taken by the Friends of the Bay are largely related to its status as a volunteer agency. Having a volunteer basis provides a basis for a larger collection of staff however, there is a much larger disparity between experience levels. This disparity is most commonly observed in the reporting of fecal coliform levels which are reported below the limit of detection in a significant portion of the data or above the limit of quantification.

Initial combination of the datasets was performed using Excel. Both the Friends of the Bay and the National Shellfish Sanitation Program have unique systems for naming data columns, this made it essential to make a single naming convention which would be applicable to both datasets. Above the necessity of unifying the naming conventions, it was also necessary to add additional columns which contain remarks concerning data. In the Friends of the Bay dataset remarks for the fecal coliform levels were combined with the reported measurement, ie. > 16000 MPN/100ml or < 3 MPN/100ml. Data remarks provide significant insight into the degree of censorship exhibited by a dataset because it represents the total number of values which are reported at or below the limit of detection, revealing the validity of a measurement. Reporting limits are an important feature in environmental data because the concentrations of sample can be very low and therefore may not be detectable outside of a trace analysis method. Tidal stage is also critical to the performance of WRTDSTidal because using salinity as a flow metric is only valid to the extent that all measurements were taken at a consistent tidal stage. If tidal stages are not recorded it is not possible to make the assumption that all measurements were taken at a consistent point and therefore the assumption that tidal mixing is approximately uniform throughout all time at a single site for a particular stage is false. After accounting for the naming changes it was also essential to ensure that all columns were appropriately aligned to make the merge occur without disagreement between the data locations. A naming convention for all of the sites was also established to distinguish the monitoring agencies and locations without overlaps in site names.

After completion of modifications to each individual dataset in Excel, Rstudio was used to merge the two agency datasets into a single database that can be drawn from. First each dataset was read into the Rstudio software as a comma separated file, allowing all aspects of the data to be accessible within the program. Using the tidyr gather function the two datasets were merged using the column headings, ensuring that all data points were transferred in their original format. The library lubridate was then loaded to manipulate all dates into a uniform and readable format. The function "ymd()" converts all dates which are already in the year-month-day format into a POSIX format which can be manipulated and read fully by Rstudio. The merged dataset was then exported to a new csv file to ensure that the original data remains pure.

Spatial and Temporal Trend Analysis

Outside of qualitative mapping a set of spatially arranged boxplots were produced in an attempt to confirm that the fecal coliform data approximately follows our hydrologic based understanding of the Mill Neck Creek water system. The sites were ordered primarily from Beaver Brook, the primary freshwater source, located in the lower branch of Mill Neck Creek and also from "the Birches" outfall located in the upper branch of Mill Neck Creek. The sites are numbered One through 21 based on the hydrologic flow map produced during TMDL creation, see **Figures 8 & 9** and **Table 2**, and will be use to compare the change in median fecal coliform concentration. Lines were also added at log(14)MPN/100ml and log(200)MPN/100ml which are respectively the Shellfish and Swim safety standards[17]. The addition of these lines gives a reference to determine annual variations and exceedances of safety standards.

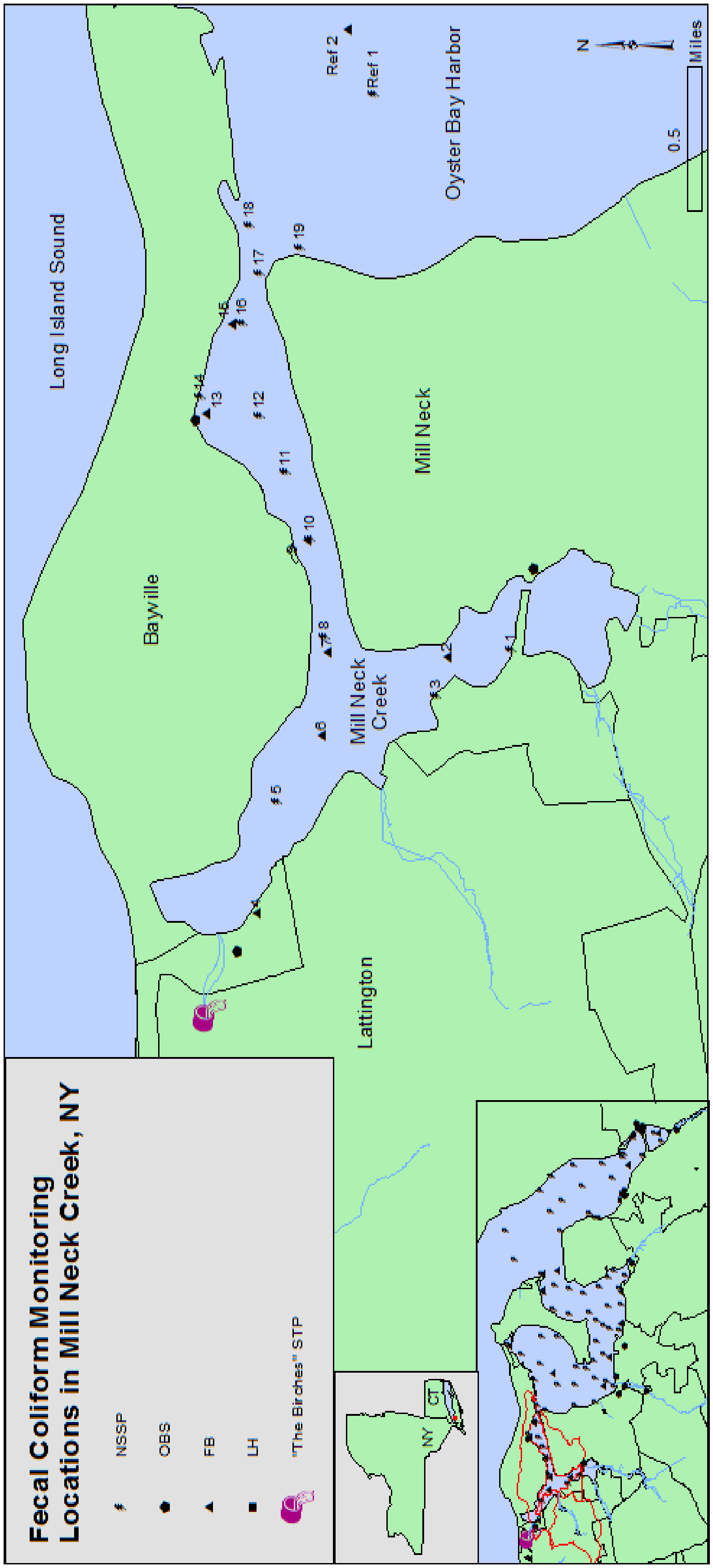


Figure 8: Map of ordered sites used for hypothesis testing and statistical analysis.[3,4,5,6,18]

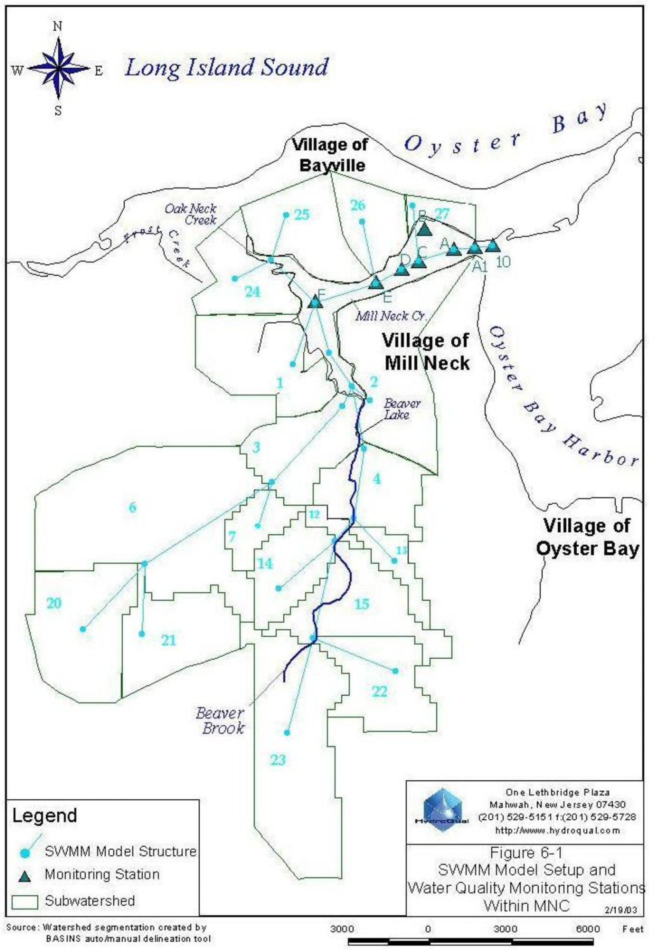


Table 2: Hydrologic Order of Sites	
Order Number	Site ID
1	47-FOB15
2	FB15
3	47-FOB16
4	FB17
5	47-FOB17
6	FB16
7	FB14
8	47-F
9	47-E
10	FB13
11	47-D
12	47-C
13	FB18
14	47-B
15	FB19
16	47-A
17	47-A1
18	47-10
19	47-9
20	47-8
21	FB11
47-8 and FB11 are reference sites outside of the study area	

Figure 9: Estimated hydrologic flow in Mill Neck Creek as assessed by the NYS DEC[1]

Goodness of Fit Analysis[19]

The new csv file created in the previous stage was reread into the project to be clipped exclusively to the Mill Neck Creek study area. From this dataset a new data frame was created by subsetting all sites in the Mill Neck Creek study area, see **Figure 8** for a map of all sites in the study area, and saved to a new csv denoting that it is only data for the study area. The Mill Neck Creek dataset was reentered into the project for further manipulation. Two new columns were created; Agency, which denotes whether the Friends of the Bay or the National Shellfish Sanitation program is responsible for the site, and Birches, which denotes whether the data was taken before or after 1-1-2012 which is the date at which "The Birches" new sewer infrastructure was considered active. All measurements of fecal coliform containing a value of 0 were also removed at this stage, this is because a measurement of 0 indicates that no measurement was made and therefore no value can be placed in the cell. A log transform was also performed on the fecal coliform data, this process allowed curve fitting to a log distribution and ensured a more manageable range of measurement values ranging from 0 up to approximately 4 log(MPN/100ml) rather than a range extending into the thousands of MPN/100ml.

Primary analysis of both the Friends of the Bay and National Shellfish Sanitation Program occurred separately for both datasets. This analysis was meant to determine if each series conformed to a log-normal distribution as expected for pathogen data. The same systematic analysis was also applied to the merged dataset to ensure that both data series functioned cohesively once merged. A Kolmogorov-Smirnov Goodness of Fit test was performed on both individual data sets and the merged file. The Kolmogorov-Smirnov test was selected because it is considered a non-parametric test for normality and is regarded as making no assumptions about the data's distribution. This is ideal for the circumstances in Mill Neck Creek because there is an extremely limited body of work for the area and a largely censored dataset. The test functions by setting both a hypothesis, normal or lognormal distribution of data, and a null hypothesis, the data does not follow a distinct distribution. The test encompasses several individual statistical analysis which produce a graphical output. First a Quantile-Quantile plot is produced to compare the median fecal coliform concentrations was performed. Secondly a Tukey Mean Difference plot was produced which plots the mean quantile difference against a horizontal zero line to show the difference between two datasets. The third component is an Empirical CDF plot which plots the value of each point versus the percentage of points below that value. The points are connected in a stepwise manner and optionally a best fit curve can be included to compare the points to an estimated normal distribution. The final plot included in this plot can be either a strip plot or boxplot which simply plots all data points for each set side by side. When using a boxplot it also allows for a comparison of the quantile distribution, outliers, and a comparison of the median values. The Kolmogorov-Smirnov test was also performed for the merged dataset as well as a test of "the Birches" hypothesis directly.

Testing "the Birches" Hypothesis[19]

Following the characterization and Two Sample Wilcox Rank Sum test was performed to examine "the Birches" hypothesis specifically. The Wilcox test is rigorously nonparametric and calculates whether there is a statistically significant difference between two data sets. The Wilcox hypothesis testing is, again, a rigorous nonparametric examination of the data and makes minimal assumptions about distribution and normality. The Wilcox Rank Sum test examines if there is a statistically significant difference between two given data sets. In this case "the Birches" Hypothesis is written such that the median value of the data from January 1, 2004 until December 31, 2011 is lower than the median value of the data values between January 1, 2012 and December 31, 2014. January 1, 2012 was selected as the breaking point because it is the first year in which the new sewer infrastructure in "the Birches" community was completely integrated and functional. The null hypothesis in this test is that the median value before updating the infrastructure is equal to the value after infrastructure was upgraded and properly connected. The alternative hypothesis is such that the median value after the proper installation of sewer infrastructure the median value of fecal coliform concentration was statistically lower than before the modifications were made.

A Quantile Test for Tail Shift was also performed to assess the degree to which the data is censored. Environmental variables have a tendency to be left censored, meaning that a significant portion of the data falls at or below the detection limit. The Tail Shift analysis provides detail as to whether the tail shift will impact the methods of analysis and if adjustments must be made to combat the influence of tail shift. This test is important to this particular analysis as the data is approximately 5-10% left censored which is consistent with observations of environmental variables. Because the censorship is less than 30% it is likely that a standard analytical method can be applied without significantly hampering the results. For the Tail Shift examination the null hypothesis is such that the difference between the series is zero and therefore there is no or negligible censorship of the data. The alternative hypothesis indicates that the tail is shifted to the right, indicating that the values of Data Y are closer to the reporting limit or more heavily censored than Data X.

Table 3: "The Birches" Hypothesis Testing Parameters		
	Quantile Test	**Two-Sample Wilcox Rank Sum Test**
Null Hypothesis	e=0	Fy(t) = Fx(t)
Alternative Hypothesis	Tail of Fx Shifted to Right of Tail of Fy. 0 < e <= 1, where Fx(t) = (1-e)*Fy(t) + e*Fz(t), Fz(t) <= Fy(t) for all t, and Fy != Fz	Fy(t) > Fx(t) for at least one t
Data x	value[Birches == "Before"]	value[Birches == "Before"]
Data y	value[Birches == "After"]	value[Birches == "After"]
Sample Size x	1941	1942
Sample Size y	604	604
Test Statistic	k (# x obs of r largest) = 9; r = 9	z = 3.31247
Test Statistic Parameters	m = 1941; n = 604 quantile.ub = 0.996465	NA

Table ##: Detailed description of all parameters for hypothesis testing. Variable X refers to data collected between January 1, 2004 and December 31, 2011. Variable Y refers to data collected between January 1, 2012 and December 31, 2014.

A Note on the Treatment of Data

The statistical analysis of data for this study was performed using the ENVStats package in Rstudio. The package was constructed to meet the analytical specifications of the United States EPA and is considered comparable to their own methods of analysis. This is the reason that the ENVStats package was selected. The Package has a litany of statistical and characterization functions available to users which enable clean and precise figure and chart creation.[19] Copies of the annotated coding for all statistical analyses are provided in **Appendix 3**. A full performance of WRTDSTidal[16] was not completed due to time constraints. A copy of the trial run is included in **Appendix 3** however, no results have been included in this study. The Tidal model requires a more complete dataset for Salinity including the designation of a reference site in the main body of Long Island sound. It also requires a conversion of the salinity data into a flow metric arithmetically.

Results

Preliminary Results

Analysis of the Friends of the Bay data in the preliminary study indicates a consistent and reliable reporting method. The information provided by the Friends of the Bay is uniform in analysis and processing by the Fuss & O'Neil analytical company and the existing results are consistent with reported methodology and detection limits. Some inconsistency was discovered in reporting values at both the lower and upper detection limit in that values are reported between 1 and 3 MPN/100ml. This issue is fairly minor to the extent that the reporting limit for the measurement methods is 3MPN/100ml, as is reported for the EPA guidelines for the methodology used. A similar reporting error occurred at the maximum reporting limit where values were reported to varying degrees exceeding the 2500 MPN/100ml reporting limit. For the purpose of this analysis values exceeding the reporting limit were recorded in the remarks column and then replaced with the reporting limit of 2500MPN. This compromise in processing is consistent with the treatment applied by the Fuss and O'Neil analysis and provides consistent results.

Outside of discovering the inconsistency in treatment of non-detects the preliminary analysis revealed the superficial nature of previous analysis in Mill Neck Creek. The Fuss & O'Neil data processing examined all of the variables in accordance with their stated QAPP however it did not examine trends for fecal coliform in detail. The data examines trends in fecal coliform concentration for the entire Oyster Bay water system over the course of a single monitoring season, however it does not compare the results throughout the monitoring lifetime. Two major methods were applied to the fecal coliform variable, a line chart plotting the 30 Day Running Geometric Mean of fecal coliform concentration taken for the geometric mean of all monitoring sites in Oyster Bay and a bar chart of the geometric mean of fecal coliform concentrations for Mill Neck Creek, Oyster Bay Harbor, and Cold Spring harbor. The 30 Day Running Geometric Mean plot depicts trends for the entire water body for a year but does not take into account the large disparity in coliform levels between sampling locations in Mill neck Creek and those in the remainder of the bay. The chart is also not overlayed for multiple years making it more difficult to examine temporal trends. The regional bar charts are somewhat more helpful in examining concentrations but still do little to examine trends. Overall the existing analysis does little to further understanding of any changes to the water quality in Mill Neck Creek over the lifetime of the study.

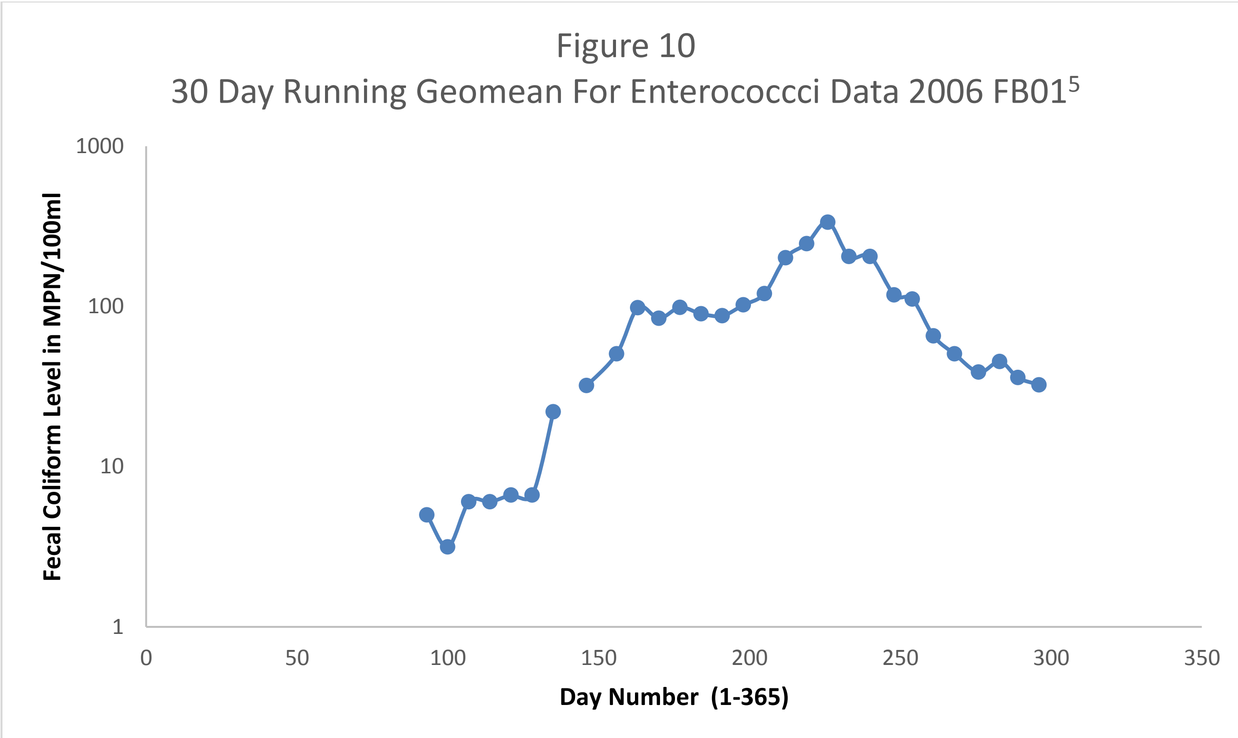


Figure 10: 30 Day Running Geometric Mean replication from the Friends of the Bay 2006 data for site FB1[6]

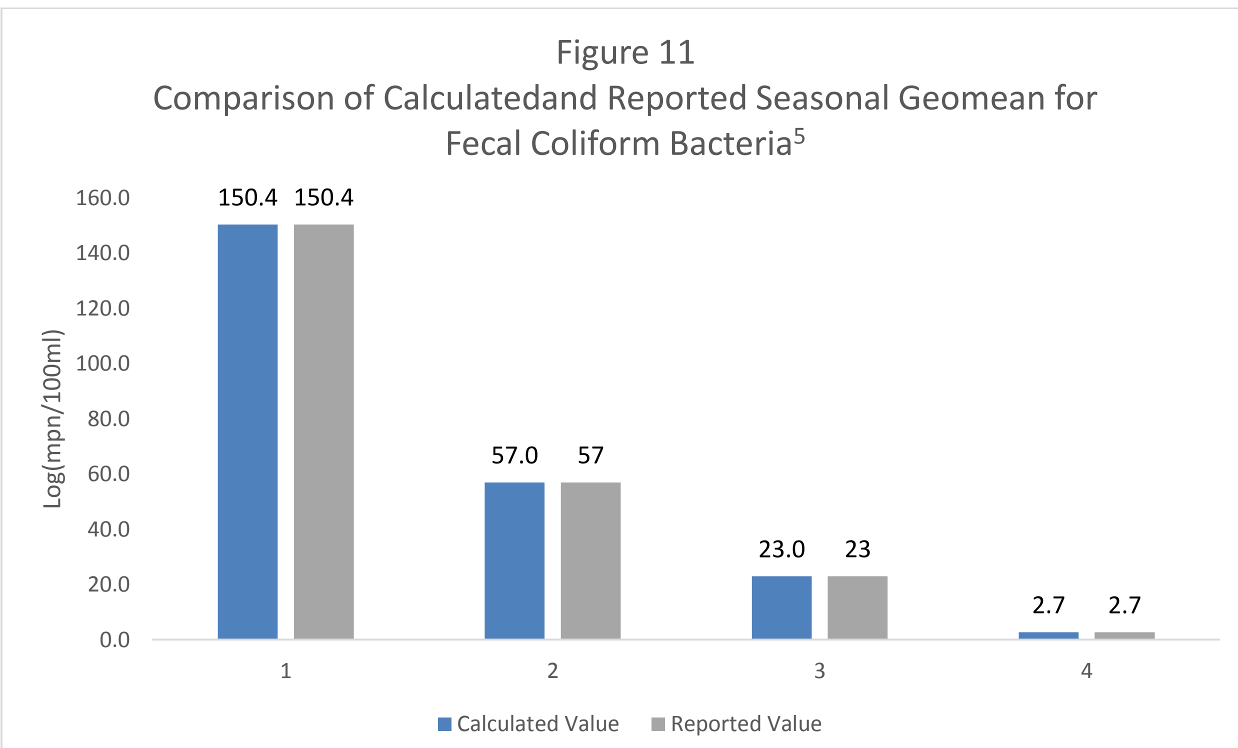


Figure 11: Replication of the Friends of the Bay Seasonal Geometric Mean bar chart for sites FB1-4[6]

Spatial and Temporal Trends

Boxplots were created according to the site order described in the methodology. The trends in the boxplots are consistent with both the spatial and flow dynamics predicted for the Mill Neck Creek system as well as the temporal trends predicted for “the Birches”. For purpose of clarity boxplots of even years for the duration of the study have been included in **Figure 12** to depict annual trends from 2004 through 2014.

First examining the spatial trends of the data with regard to predicted hydrologic flow, there is a strong apparent trend that the estimated ordering was correct. Based upon **Figure 12**, flow originates in the lower branch of Mill Neck Creek where Beaver Brook feeds into the estuary, a secondary smaller input begins in the upper branch of the creek at approximately the location of “the Birches” outfall. Water then flows outward towards Oyster Bay Harbor and ultimately Long Island Sound. This is supported by the decreasing median fecal coliform concentration observed at monitoring locations closer to Oyster Bay Harbor. The fluctuations in fecal coliform concentration at sites outside of the entry points indicate two other influences in the study area. The first being the role of tidal mixing within Mill Neck Creek, tidal influence is more relevant with regard to the sites closer to Oyster Bay Harbor where stream and outfall inputs are less influential. Sites 1-6 are subject to greater influence by stream and outfall influences, however they are not independent of the tidal changes. In contrast Sties 15-21 are more dependent on tidal influences due to increased distance from the freshwater sources. Future work in this area includes the analysis of data in Mill Neck Creek using WRTDSTidal, an Rscript which takes into account tidal mixing. The second influence that is not quantitatively accounted for in this study is runoff water from the nearby suburban and conservation land which will contribute a certain degree of coliform contamination based upon the wildlife living on the land itself. Natural fertilizing agents could also contribute to this but have a greater impact on nutrient concentrations, which are outside the scope of this study.

Examining the data with regard to “the Birches” hypothesis provides some superficial evidence of the impact re-sewering “the Birches” has had on Mill Neck Creek. Comparing the median fecal coliform concentrations over the time series, a decrease in the median values is distinguishable. This evidence does not have a strong, quantitative standing, however it does provide additional information in support of “the Birches” hypothesis. A decreasing trend for median fecal coliform concentration after the 2012 completion of sewer upgrade and connection to a proper filtration system does indicate that the efforts were successful, at a minimum on the immediate median concentration.

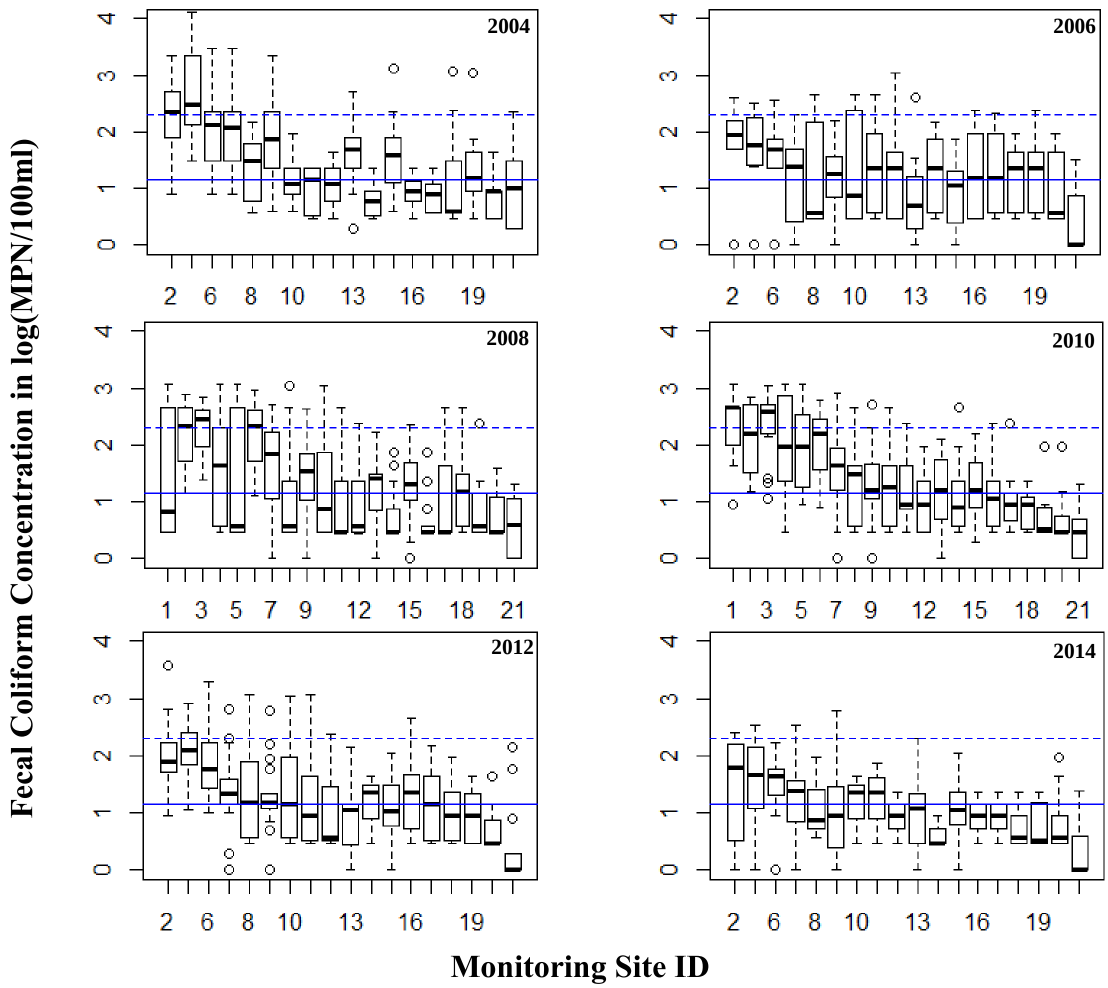


Figure 12: Annual boxplots of fecal coliform concentration for selected years with sites ordered by estimated hydrologic flow[1,6,20]

Goodness of Fit Analysis

After analyzing both the Friends of the Bay and National Shellfish Sanitation Program datasets independently it was possible to conclude that both follow approximately a lognormal distribution with 5-10% left censorship. By providing evidence that both datasets follow a lognormal distribution it also supports the claim that the datasets can be merged and treated as a single entity. Based on the Goodness of Fit analysis both data series are left censored, indicating that there is a significant portion of each dataset that is at or below the detection limit. Moving forward from this all data points that have a value of 0 were removed because they were indicative of a non-measured sample rather than a non-detect. Values reported between 1 and 3 MPN/100ml were reported as 2.9 MPN/100ml with “<” placed as the remark in contrast to the Fuss and O’Neil convention of using the detection limit for non-detects. For values exceeding the 2500MPN/100ml maximum, a value of 2501 MPN/100ml. These choices substitutions reflect the conventions of the National Shellfish Sanitation Program. **Figures 13 & 14** reflect the

Goodness of Fit test examining the distribution for The Friends of the Bay and National Shellfish Sanitation Program respectively. Based on these individual tests it is clear that the Friends of the Bay data follows a lognormal distribution very closely. In contrast the apparent distribution of the National Shellfish Sanitation Program is heavily left censored indicating that there is a substantial portion of the data that is at or below the detection limit. Despite the poor distribution agreement when examined separately it was essential to test the data as a complete set to determine how it functions as a whole.

Looking specifically at the National Shellfish Sanitation program data, the fit to a lognormal distribution is approximately 85% (W=0.861) indicating that a significant portion of the data falls outside of the distribution. Examining the histogram and QQ Plot however, it becomes clear that the majority of this deviation from normal distribution can be accounted for as a result of left censorship of the dataset. The data for the National Shellfish Sanitation Program has a much higher degree of censorship than the Friends of the Bay. This censorship shifts the curve outside of a standard lognormal distribution by extending the left end. The stepwise nature of this dataset is most likely related to the limited number of samplings when compared to the Friends of the Bay. Proceeding from this point it was likely that by merging both datasets the large portion of non-detect values would be diluted due to the near doubling of measurements being evaluated. **Figure 15** is a boxplot indicating the distribution of data for Friends of the Bay and the National Shellfish Sanitation Program. Using this to augment the Goodness of Fit results it is safe to conclude that the non-detect values present in the National Shellfish Sanitation Program data could improve in quality by merging with the more cohesive Friends of the Bay data.

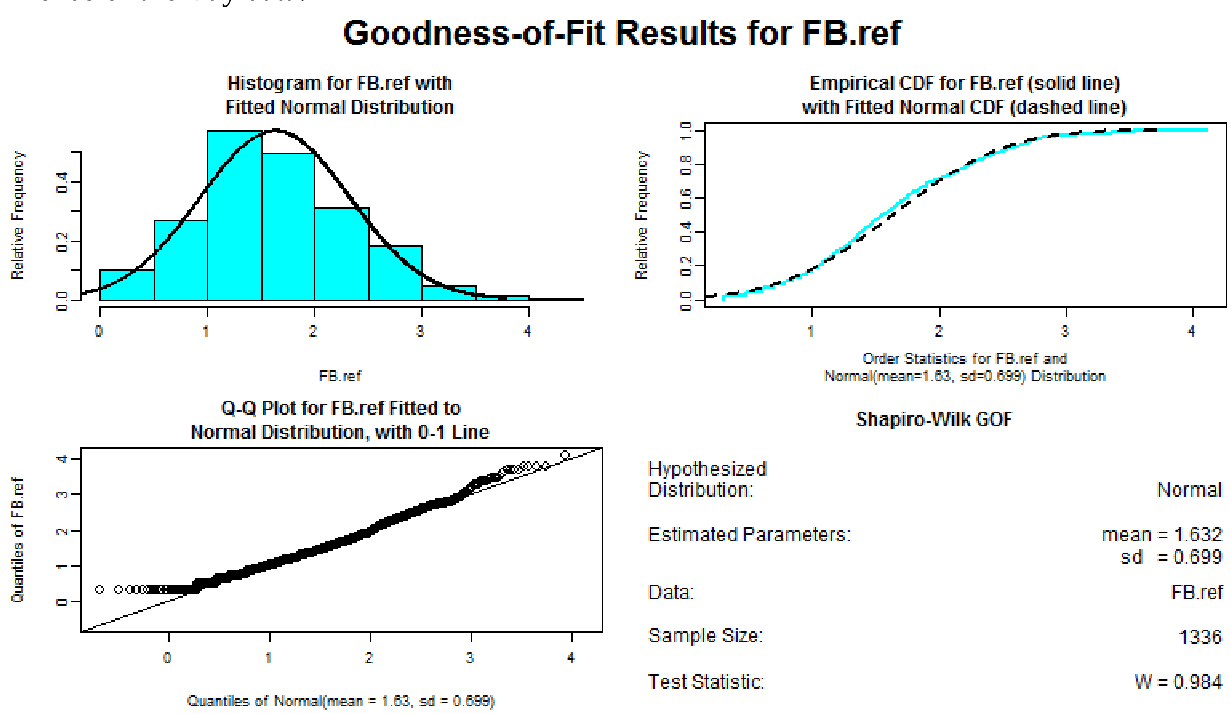


Figure 13:Goodness of Fit results for the Friends of the Bay Data[6,]

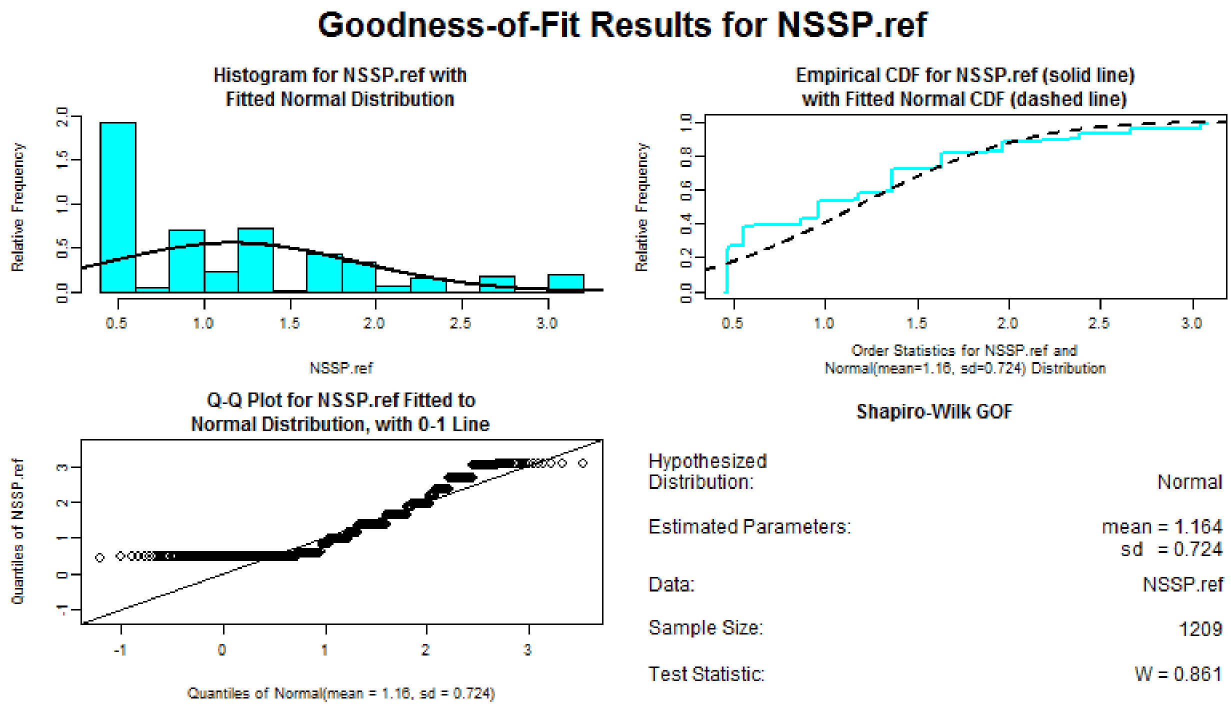


Figure 14:Goodness of fit results for the National Shellfish Sanitation Program data.[20]

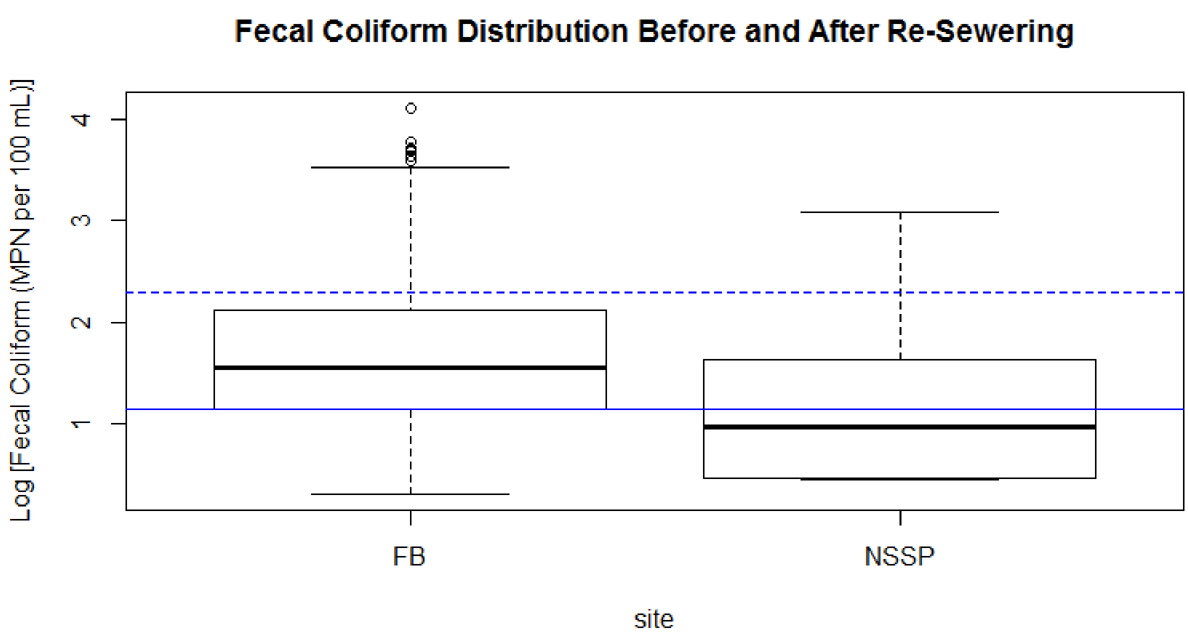


Figure 15: Boxplot of all data for 2004-2014 displayed as before or after "the Birches" was resewered[6,20]

Examining the Goodness of fit results for the combined dataset it becomes clear that when the data is considered in its completeness it can be considered a cohesive entity. Examining the Goodness of Fit test for the combined dataset there is a 95% (W=0.95) confidence interval at which the data is aligned with the lognormal distribution that was predicted. There is still some left censoring but the influence is much smaller with the increased number of measurements being considered, approximately twice As many points are examined in the merged data. The included QQ Plot indicates that beyond the values reported at the detection limit, there is a very strong correlation with a normal distribution. The left censoring is a common phenomenon for environmental studies due to the relatively low concentrations being examined under normal, safe conditions. Examining the Histogram and QQ Plot it becomes apparent that previous assumptions concerning non-detects were correct. By increasing the number of measurements the percentage of values below the detection limit is substantially decreased resulting in a distribution substantially closer to lognormal. Examining the Empirical CDF provides a visual verification of the close relationsxhip between observed values and the calculated curve.

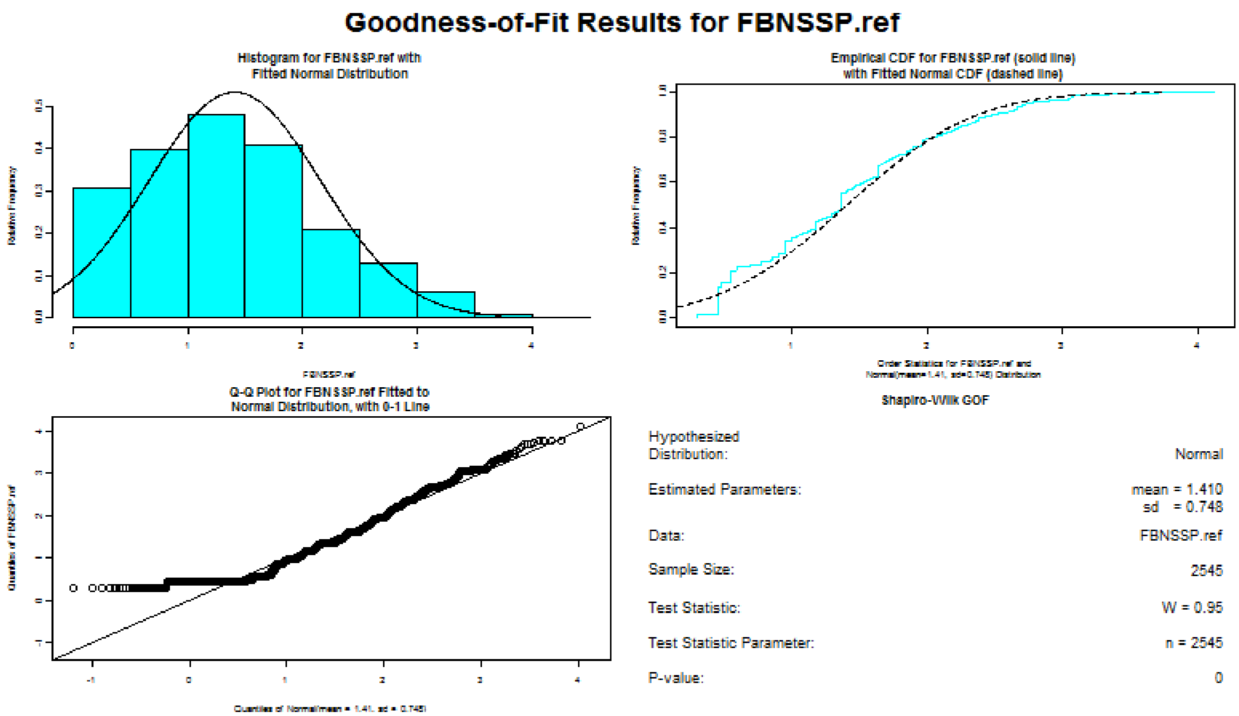


Figure 16:Goodness of fit result for the combined dataset.[6,20]

"The Birches" Hypothesis

With the lognormal distribution of the merged data confirmed it was possible to examine the impact re-sewering "the Birches" had on fecal coliform concentration. Below is a boxplot depicting the distribution of measurements Before and After the infrastructure. There are a greater number of measurements taken prior to January 1, 2012 which is largely indicative of the time span of the study and the duration since re-sewering was completed. Based on the boxplots a decrease in median fecal coliform concentration can be discerned since the re-sewering occurred. The simplest and most probable explanation is that "the Birches" served as a substantial point source of fecal coliform and correcting the sewer infrastructure resolved this contribution. The included QQ Plot shows that as a whole the data follows an approximate one-to-one trend line indicating that the change is slight but is apparent. The Tukey mean Difference plot indicates a deviation from a zero difference line. Points falling on the line indicate that the mean value for both Before and After the re-sewering are equal resulting in a difference of zero between the two means. With a more prominent deviation of the points below the zero line, the data test suggests a decrease in fecal coliform concentration as the difference places the points in the domain of Before the re-sewering. The Empirical CDF comparison supports this, showing that the median fecal coliform concentration After re-sewering is lower than that for Before the infrastructure upgrade.

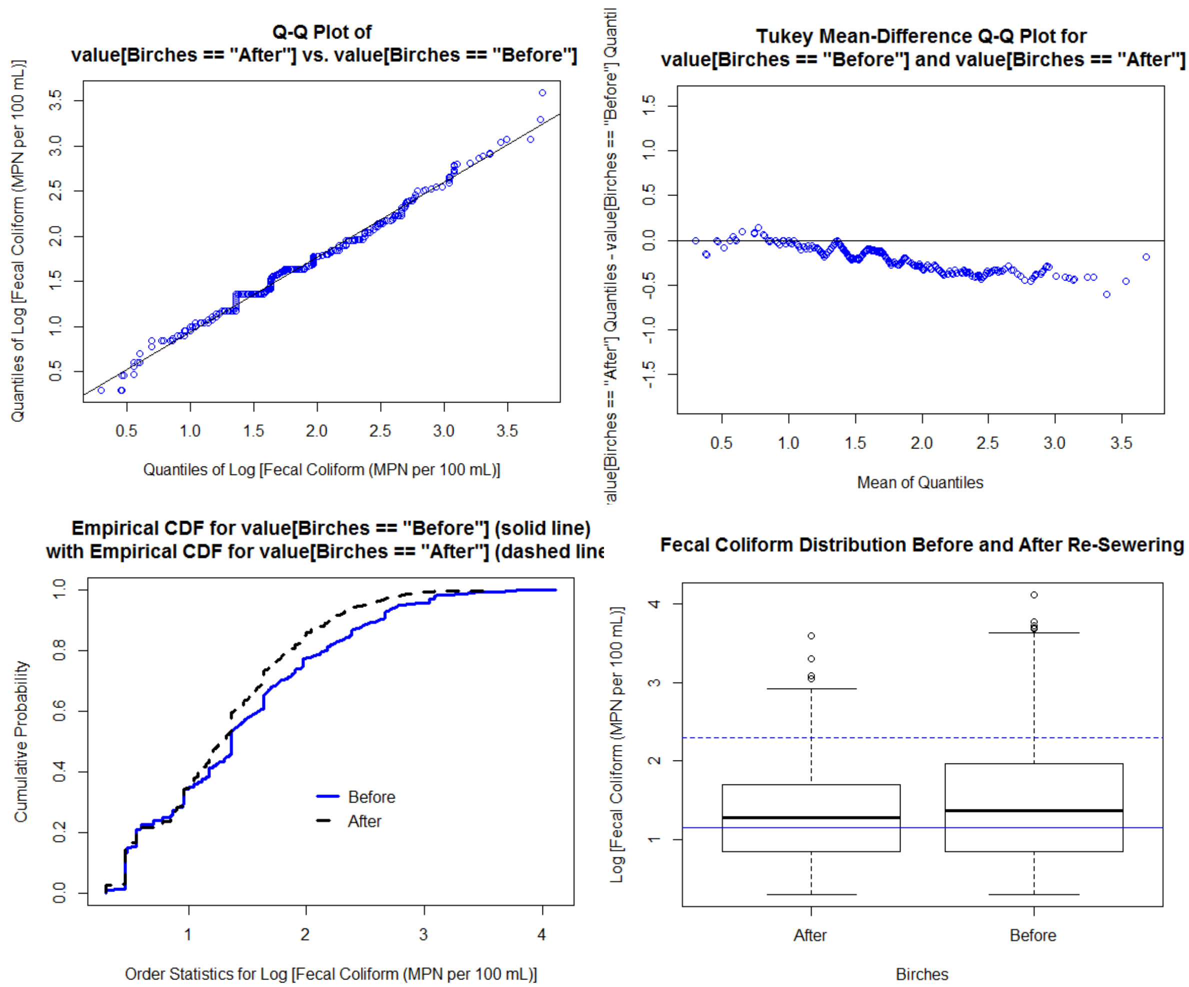


Figure 17:Hypothesis testing for "the Birches" hypothesis including QQ Plot, Tukey Mean Difference Plot, Comparison of Empirical CDF, and Boxplot for distribution of data Before and After re-sewering of "the Birches"[6,20]

Both the Two-Sample Wilcox Rank Sum test and the Quantile test for Tail Shift support "the Birches" hypothesis as well. In both of these statistical tests "the Birches" hypothesis is presented as the alternative hypothesis and the null hypothesis predicts no difference between the two selections of data. The Quantile test also assesses the degree to which censoring impacted the data analysis.

Examining the Quantile Test first, the tail shift observed in the calculations is in agreement with the alternative hypothesis. This means that the concentration of fecal coliform before "the Birches" was properly sewered is statistically greater than the concentration after based on the indication that a right shift results in higher median concentration. The test also indicates a large degree of left censorship causing the tail shift to be larger due to an increased number of low measurements in the study area. The certainty in this result is approximately 90%, as indicated by the p-value (p=0.0869),

The Two Sample Rank Sum test indicates that there is a statistical difference between the two data sets and that the difference follows the alternative hypothesis. In the case of the Wilcox test this indicates that the data After re-sewering reached the median value at a lower concentration than Before infrastructure changes were completed. Agreement of the data with the alternative hypothesis confirmed "the Birtches" hypothesis with greater than 95% confidence based on the produced p-value (p=0.00462)

Table 4
"The Birches" Hypothesis Testing

	Quantile Test	**Two-Sample Wilcox Rank Sum Test**
Null Hypothesis	e=0	Fy(t) = Fx(t)
Alternative Hypothesis	Tail of Fx Shifted to Right of Tail of Fy. 0 < e <= 1, where Fx(t) = (1-e)*Fy(t) + e*Fz(t), Fz(t) <= Fy(t) for all t, and Fy != Fz	Fy(t) > Fx(t) for at least one t
Data x	value[Birches == "Before"]	value[Birches == "Before"]
Data y	value[Birches == "After"]	value[Birches == "After"]
Sample Size x	1941	1942
Sample Size y	604	604
Test Statistica	k (# x obs of r largest) = 9; r = 9	z = 3.31247
Test Statistic Parameters	m = 1941; n = 604 quantile.ub = 0.996465	NA
P-value	0.08692069	0.004623796

Conclusions

There were several goals to this study which collectively assess the efficacy of the pathogen TMDL placed on Mill Neck Creek. Auditing the monitoring agencies provided insight into the scope and timeframe of monitoring and assisted in determining whether there was a duplication of efforts. Reconstruction of figures published in the Friends of the Bay *Annual Water Quality Report* assisted in making preliminary assessments and guided the hypothesis formation. Finally and most importantly, the statistical testing of "the Birches" hypothesis was used as the determining factor as to whether the TMDL was effective in improving the water quality of Mill Neck Creek. The quantitative results are limited to assessing the impact of "the Birches" outfall before and after the sewer system was updated and connected to the Glen Cove municipal sewer system. Assessment of land use impact, animal and septic contribution, and the contribution of suspended sediment borne fecal coliform remain speculation based upon the literature.

There appears to be little communication between agencies past and present which acts to exacerbate the problems of monitoring and ultimately results in duplication of efforts. The audit revealed that the Friends of the Bay and the National Shellfish Sanitation Program in the Mill Neck Creek area only carry out long term monitoring despite the presence of numerous inactive testing sites from other agencies. Some agencies with inactive sites were decommissioned in the past and had no usable data within the study area for the period of 2004 through 2014, EMAPS is an example of this type of agency. The Friends of the Bay are the most consistent monitoring agency, though their compliance with EAP/DEC standard monitoring procedures would permit greater functionality of their monitoring. A response in the TMDL comments section indicated that the Friends of the Bay data was not readily available at the time of writing and that even published data could not be used by the EPA/DEC in determining compliance without being processed by an associated laboratory facility. For this reason interagency cooperation would greatly increase the remediation potential for Mill Neck Creek.

Examination of the preliminary data and figures provided by Fuss & O'Neil on behalf of the Friends of the Bay lead to several conclusions concerning the existing systems. The first is a need for more detailed analysis of trends for each variable, in particular fecal coliform and enterococci bacteria as these are directly related to the TMDL and closure of Mill Neck Creek. The figures and tables provided in the Fuss & O'Neil analysis provide superficial trends occurring in Mill Neck Creek or the broader Oyster Bay as a whole but do not examine temporal trends effectively. For example, the 30-day running geometric mean of fecal coliform bacteria presents trends for all of Oyster Bay over the course of one year. These figures provide very broad, single year assessments about the state of Oyster Bay and its three regions however, the detail in regional trends is less than desirable when examining the long-term trends. In addition, the inconsistent reporting limits used can produce undue skew of results if not properly accounted for. In this study, the DEC recommended practices for measurements outside of the detection limit were followed for consistency.

Evaluation of the statistical results indicates that “the Birches” hypothesis is valid with regard to collected data. Goodness of fit analysis shows that the data fits a lognormal distribution as expected for environmental data where the majority of measurements approach the lower detection limit. “The Birches” hypothesis predicts that after completing the re-sewering of the housing community there was a statistically significant decrease in fecal coliform concentration, indicating that the outfall provided a substantial source of pathogen loading. The null hypothesis predicted that no statistical change occurred after completion of the infrastructure update and therefore the outfall was not a major contributor to pathogen loading in Mill Neck Creek. The statistically significant decrease in median fecal coliform concentration indicates that the re-sewering of “the Birches” is coincident with the small improvement of water quality in Mill Neck Creek. Outside of the infrastructure changes introduced to the 30 home community, there were no major changes to the suburban environment which impacted impervious surface or directly contributed to the pathogen load. The vast majority of properties bordering Mill Neck Creek and its contributing water sources are residential with a small number of commercial properties mainly in the Bayville and Mill Neck creating an environment which has strong potential to filter pathogens before they enter the waterway. Barring any other major developments in the region, the most probable cause for the decreased fecal coliform load was “the Birches” renovation. A single point source which directly contributed sewage to the ecosystem would provide substantial independent pathogen loading, in particular when deposited in a fairly closed system with minimal opportunities for outflow. In the case of Mill Neck Creek, the only pathway for water to leave is through the channel into Oyster Bay and ultimately the Long Island Sound. The creek itself is small and the channel is no exception, outflow would be restricted largely to tidal interactions as the sources of fresh water are streams and creeks with water flow that is insufficient to force the load into the larger water body. Since there are no large, multi-occupant residential, hotel, or motel properties in the area it is unlikely that another point source of the same caliber was modified in the same timeframe. It is also unlikely that there were any dramatic changes in non-point sources within the Mill Neck Creek study area which would decrease the median fecal coliform concentration.

Trends observed in the boxplots support the conclusion that “the Birches” was a major point source in Mill Neck Creek. The sites are ordered beginning in the southern branch of Mill Neck Creek (sites 1-3) followed by the northern branch (sites 4-6) and then ordered outward towards Oyster Bay Harbor. Sites 4-6 consistently had higher median fecal coliform concentrations than other areas of Mill Neck Creek, indicating that a point source was likely located in the area. Site 4 represented “the Birches” outfall and examining the median value in that area indicates that the median value for both the creek itself and the outfall area decreased supports the conclusion that the community outfall was a significant source of pathogen loading. High median values in the southern fork does indicate that other pathogen sources remain unaddressed however, the plots also indicate that Mill Neck Creek is making great strides towards compliance with the EPA swim and shellfish standards. Examining the 2014 boxplot, all median and third quartile fecal coliform concentrations fall below the swim standard for pathogen levels. This dramatic change occurred only two years after the infrastructure renovation was completed. It is likely that infrastructure change in “the Birches” can explain a portion of this concentration change, small decreases in fecal coliform level have occurred in previous years. It is important to recognize that the smaller decreases can likely be attributed directly to the TMDL legislation but cannot be addressed by the current analysis. The TMDL was implemented in 2003, one year prior to the earliest consistent monitoring records by the Friends of the Bay and therefore data prior to the implementation was not examined in this study. Based on the small increments of improvement in the boxplots it is likely that the TMDL caused a positive change in the waters of Mill Neck Creek.

With all of the results considered, it is clear that the water quality in Mill Neck Creek has improved noticeably. The median fecal coliform for the region has decreased since the completion of “the Birches” connection to Glen Cove indicating that the housing community was a significant contribution to the pathogen loading. The TMDL’s stated goal was to bring pathogen concentrations below the shellfish harvesting standard, 14MPN/100ml. At this time the water quality does not meet this criteria but has decreased steadily since the re-sewering of “the Birches”. The results had a high degree of confidence in this concentration change indicating that there was strong correlation with “the Birches” construction. Residual concentrations and fecal coliform load farther from the outfall cannot be accounted for directly in this study but it is likely that it is due to non-point source contamination from the surrounding environment.

Moving Forward

Based on the analysis, it was apparent that the state of Mill Neck Creek is improving steadily. The shellfish standard remains out of reach for the present but is an attainable goal for the creek. TMDL regulations appear to have impacted the watershed in a positive manner, as has the implementation of proper sanitary sewers in “the Birches” community. If implementation of a single, functional sewer system in “the Birches” was capable of enacting notable changes in pathogen loading, it is logical to assume that a dramatic improvement would be the result of widespread sewer implementation in the Mill Neck Creek region. Burden et al. discuss the cost effective nature of sewer conversions, particularly for areas with suburban development patterns similar to Bayville, were property size is less than 0.5 acres per lot.[15] Despite lowering costs, the upfront materials cost is likely daunting to residents however, the increased immediate costs provide an increase in service and maintenance for new sewer infrastructure. Homeowners will no longer be responsible for maintaining, updating, or servicing septic systems and are also no longer responsible for damages associated with systematic failure, in septic systems all of these responsibilities are delegated to the homeowner. A municipal push towards sewer upgrades would benefit residents and the water body in the long term as well. Sewer systems are less likely to fail than private septic systems due to their role as a public utility. Resistance to sewer systems is likely based in the immediate cost, as a public utility there are increased fees and taxes associated with installation and maintenance. As a public utility these fees are used directly for maintenance and come with accountability, in contrast a private septic system places the burden of update, maintenance, routine service on the owner. Since there is no system which can hold owners accountable for neglect of these systemic requirements, it is more likely that the systems will fail due to neglect or unobserved issues. Ultimately the maintenance and costs of septic systems are greater but this is misunderstood due to the immediate cost of installing sewer infrastructure. A directed program would not be sufficient to calm the public resistance to sewer infrastructure, with the legislation it would be necessary to educate the citizens of benefits associated with sewer infrastructure. These education programs would be most effective if curated by representatives from all five autonomous regions surrounding Mill Neck Creek.

Interagency cooperation with regard to monitoring and policy would likely increase the effect of existing changes dramatically. At present there are nine agencies with past and present monitoring activities in Mill Neck Creek, the merging of data from these sources into a single repository would provide substantial benefits to municipalities as public users. By placing all data in a single locally managed database it will be more user friendly and easier to search. Data from federal and state agencies such as the National Shellfish Sanitation Program will be available in the same location as Friends of the Bay data saving both time and effort on the part of the user. This change will likely encourage further studies of water quality in Mill Neck Creek and decrease the risk of duplicated efforts. With this also comes the recommendation that the Friends of the Bay contact the National Shellfish Sanitation Program and work towards a unified monitoring technique. If the NSSP were able to use the Friends of the Bay data to assess the

TMDL compliance it would increase the temporal resolution of information dramatically. The Friends of the Bay monitor more often than the NSSP, therefore a unified testing and analytical method would allow the data to be used by all agencies rather than being excluded for procedural differences or lab certifications. Another benefit would be the ability of municipalities to monitor water quality with ease. A unified database would provide all available data in a searchable format making access faster and easier. It would ultimately be beneficial to include a section for GIS data. The GIS data would save time and effort by compiling it in a single place, more so if the monitoring data was already linked to point or polygon features and ready for use. Two major challenges to this assessment were related to the data sourcing. The first was finding all data for fecal coliform available and determining what time frame it was available from. The second was merging the data due to variations in reporting styles and detection limits. With several sites providing data for different agencies, it remains extremely difficult to obtain all data for the region as there is no major communication between monitoring agencies. This was addressed in the TMDL comments section where inquiries were made regarding the Friends of the Bay data, the DEC did not include the Friends of the Bay data because they were unable to locate it or verify its laboratory credentials.[1] Cooperation between the monitoring agencies would minimize this issue because there would be a comparable reporting system resulting in a more unified method of reporting results.

The current set of policies in place is also assisting but likely could be improved upon. While there are at least five autonomous municipalities with influence on Mill Neck Creek, it would be beneficial for the region if a committee was formed by members of all of these municipalities. By forming a committee, more uniform policy will be in place throughout the area and therefore it will be easier to enforce a higher standard for Mill Neck Creek. The consistent enforcement of all policies will also be important to the rehabilitation of Mill Neck Creek; though it is considered a no-dumping zone for onboard boat septic tanks[1] the enforcement could be better. Improved enforcement of the boat dumping policy will likely present another significant improvement to the water quality. Existing policy on removing pet fecal waste on public ground is also an important existing measure in minimizing fecal coliform contamination. Also an increase in frequency and consistency of septic tank inspections would likely improve the water quality, this would be moving forward and hopefully for the present while a sewer construction plan is negotiated and implemented. Improving the inspection policy for the area's septic systems would lead to better maintenance and a decrease uncorrected failing or ineffective tanks. Since, in areas like Bayville, a conventional tank has insufficient space to filter the waste in the event of a breach it is likely that unnoticed septic failures are contributing to the fecal coliform load[8].

At this time, the data only validates that the fecal coliform concentration has decreased since the re-sewering of "the Birches" and that the improvement has brought the median concentration below the swimming standard. Speculations related to policies and alternative sources are made using the literature and attempts to apply the results to other regions of Mill

Neck Creek. Future work to verify this analysis can include the implementation of WRTDSTidal properly as it is a more rigorous test for changes in the estuary environment. Also a comparison of fecal coliform concentration to Secchi depth and salinity would provide insight into the effect adsorbed fecal coliform has on Mill Neck Creek and if the salinity is masking any of the contamination. Research into the failure rate of septic systems in the area would also provide insight into the contamination contributed from those sources.

Acknowledgements

Thank you to Dr. Craig Dalton, the chair of my committee. Without your guidance and quick answers to my questions the geographic portion of this study would not hold together. Repeated editing of my drafts and sections provided irreplaceable input that has guided the final product and allowed to exceed previous expectations.

Thank you to Dr. Kevin Bisceglia, my primary advisor in the chemistry department and the guide for all statistical work in this project. Without your guidance I would not have been able to complete the rigorous analysis which determined out conclusions.

Thank you to Dr. Linda Longmire for her continued support and guidance throughout this entire project. In particular for her guidance provided while abroad both presently and before the onset of this research.

Thank you to Dr. Sandra Garren for your guidance in the preliminary work and production of maps for this project. Without your guidance the analysis would not have occurred. Thank you also for connecting us to the Friends of the Bay and facilitating communications between us.

Thank you to Catherine Fischer for her assistance in loading Esri software and Rstudio onto computers in Berliner Hall, these programs were essential to the analysis.

Thank you to the Department of Chemistry for funding this research in the Summer 2015 sessions. This funding allowed me to devote additional effort to this project without having to work additional time at another job, this greatly increased my ability to focus on the work.

Thank you to the Department of Global Studies and Geography for the continued support in this effort. In particular for the use of the Esri software and plotter.

Thank you to Hofstra University and all of its staff for assisting in various ways throughout this project

Thank you to the Friends of the Bay Executive Director Paul DeOrsay for providing access to the organization’s historical water quality data.

Work on this study was completed independently by the author at the request of the Oyster Bay/ Cold Spring Harbor Watershed Protection Committee. The analysis provided reflects only the author’s conclusions and does not necessarily reflect the opinions of the Oyster Bay/ Cold Spring Harbor Watershed Protection Committee, their staff, or any of their affiliates.

References

1. T New York State. Pathogen Total Maximum Daily Loads For Shellfish Waters in Oyster Bay Harbor and Mill Neck Creek Nassau County, New York; NYSDEC, 2003.
2. Enriquez, S. Locust Valley Sewage Woes Coming to an End Newsday [online] April 16, 2009 http://www.newsday.com/long-island/nassau/locust-valley-sewer-woes-coming-to-an-end-1.1219105?pts=428060 (Accessed December 8, 2015)
3. Environmental Protection Agency. EPA Storet Central Warehouse http://ofmpub.epa.gov/storpubl/dw_pages.querycriteria (June 10, 2015)
4. New York State GIS Clearinghouse https://gis.ny.gov/gisdata/inventories/member.cfm?organizationID=529 (June 15, 2015)
5. A Connecticut Department of Energy and Environmental Protections (DEEP) http://www.ct.gov/deep/cwp/view.asp?a=2698&q=322898&deepNav_GID=1707%20 (August 1, 2015)
6. Learning About Water A presentation about water and the world we live in http://hunstem.uhd.edu/learningaboutwater/learningaboutwater.swf (accessed May 9, 2016).
7. Friends of the Bay. Friends of the Bay Water Quality Data; http://friendsofthebay.org/ 2004 – 2015 Water Quality Data (Accessed June 1, 2015)
8. Cahoon, L. B.; Hales, J. C.; Carey, E. S.; Loucaides, S.; Rowland, K. R.; Nearhoof, J. E. Shellfishing Closures in Southwest Brunswick County, North Carolina: Septic Tanks vs. Storm-Water Runoff as Fecal Coliform Sources. Journal of Coastal Research **2006**, 222, 319–327.
9. Cahoon, L. B.; Hales, J. C.; Carey, E. S.; Loucaides, S.; Rowland, K. R.; Toothman, B. R. Multiple modes of water quality impairment by fecal contamination in a rapidly developing coastal area: southwest Brunswick County, North Carolina. Environmental Monitoring and Assessment Environ Monit Assess **2016**, 188 (2).
10. Mallin, M. A.; Williams, K. E.; Esham, E. C.; Lowe, R. P. Effect of Human Development on Bacteriological Water Quality in Coastal Watersheds. Ecological Applications **2000**, 10 (4), 1047.
11. Schoonover, J. E.; Lockaby, B. G. Land cover impacts on stream nutrients and fecal coliform in the lower Piedmont of West Georgia. Journal of Hydrology **2006**, 331 (3-4), 371–382.
12. Hirsch, R.M., and De Cicco, L.A., 2015, User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval—R packages for hydrologic data (version 2.0, February 2015): U.S. Geological Survey Techniques and Methods book 4, chap. A10, 93 p., http://dx.doi.org/10.3133/tm4A10.
13. Beck MW, Hagy JD. 2015. Adaptation of a weighted regression approach to evaluate water quality trends in an Estuary. Environmental Modelling and Assessment. 20(6):637-855. http://dx.doi.org/10.1007/s10666-015-9452-8
14. Rosenzweig, C.; Major, D. C.; Demong, K.; Stanton, C.; Horton, R.; Stults, M. Managing climate change risks in New York City’s water system: assessment and adaptation planning. Mitig Adapt Strat Glob Change Mitigation and Adaptation Strategies for Global Change **2007**, 12 (8), 1391–1409.
15. City of Glen Cove, NY Building Sewers and Connections http://ecode360.com/12086081 (accessed Apr 27, 2016).

16. Permits & Fees | Nassau County, NY - Official Website https://www.nassaucountyny.gov/1874/permits-fees (accessed Apr 27, 2016).
17. Burden, D. G.; Anderson, D. L.; Zoeller, P. Septic Vs. Sewer: A Cost Comparison For Communities In Sarasota County, Florida. Proceedings of the Water Environment Federation proc water environ fed **2003**, 2003 (7), 319–343.
18. National Shellfish Sanitation Program, Fecal Coliform Monitoring Data. Obtained by FOIL Request. (Accessed January 12, 2015)
19. Millard, S.P., Package for Environmental Statistics, Including EPA Guidance— An R Package for Environmental Statistics, Springer-Verlag: New York, 2015; pp http://cran.r-project.org/web/packages/EnvStats/EnvStats.pdf
20. Bacterial Water Quality Standards. New York State Codes, Rules, and, Regulations (NYCRR) Part 47.
21. The Long Island Sound Study http://longislandsoundstudy.net/ (June 1, 2015)

Appendix 1

Estimated costs associated with both sewer and septic tank installation and maintenance. According to the City of Glen Cove, the sewer district to which Mill Neck Creek homes will be connected, requires homeowner payment towards the cost of construction of and connection to the new sewer lines. Tables include permitting costs dictated by Nassau County[16] as well as Construction, Operation and Management, and Total Annualized costs for sewer and septic systems in Sarasota County, Florida in estimated 1999 dollars.[15] The line chart provided visualizes the changes in costs of septic systems and sewer systems as population density increases.[15]

Nassau County Permits and Associated Fees[16]

Permit Type	Fee Amount
Aerial Photos (Scale of 1 inch to 200 inches) (per sheet)	$20.00
Building Permit Review - 239F (Initial Review)	$1500.00
Building Permit Review - 239F (Resubmission)	$740.00
Dumpster Permit - (Must provide copy of insurance policy)	$418.00
Dye Test (Lowest possible)	$160.00 to $800.00
Dye Test (Highest possible)	$800.00
GPS Monument Books (Benchmarks)	$550.00
Hauling Permit (Must provide copy of insurance policy)	$150.00
Industrial Waste Discharge Permit	$213.00
Plans and Specifications	$300.00 Each project
Reproduction of Maps (per square foot)	$18.00
Reproduction of Sewer Maps (per sheet)	$126.00
Road Opening Permit Application - **Residential** (Including curb cuts) (Fee includes deposit) (Mail completed Road Opening Form) Road Opening Permit Application Form-Residential (PDF file) 2016 Rules & Regulation Updates (PDF)	$275.00
Road Opening Inspection - Residential	$460.00 + deposit
Road Opening Permit Application - **Commercial** (Including curb cuts) (Fee includes deposit) (Mail completed Road Opening Form) Road Opening Permit Application Form-Commercial (PDF file) 2016 Rules & Regulation Updates(PDF)	$550.00
Road Opening Inspection - Commercial	$920.00 + deposit
Scavenger Waste (Bay Park only) (per 1,000 gallons)	$53.00
Sewer Permit (Mail completed Sewer Permit Form) Sewer Permit Application Form (PDF file)	$140.00
Sewer Permit Inspection	$460.00
Sidewalk Permit - Residential ONLY (Mail completed Sidewalk Permit Form) Sidewalk Construction Application Form (PDF file)	N/A
Special Permits (Commercial) (Lowest possible)	$160.00 to $800.00 +2% estimated construction cost
Special Permits (Commercial) (Highest possible; 2% estimated construction cost included)	$800.00
Standard Specifications (Civil engineering and site development construction)	$520.00 Each
Standard Specifications (Traffic and sewer)	$160.00
Subdivision Plan Review (Initial review; additional $300 per lot)	$7,000.00 + $300.00 per lot
Subdivision Plan Review (Resubmission)	$700.00
Verification of Permit Connection or Spur Location (each)	$95.00
** + .5% of construction costs for projects over $250,000.000	

Estimated Construction Cost of Sewer and OWTS Alternatives (per connection)[4, 9]

Alternative	Low Density[1] >0.5 acre lots	Medium Density[2] 0.25 - 0.5 acre lots	High Density[3] <0.25 acre lots
Low Pressure GP[9,10]	$8,700	$6,800	$6,700
Vacuum[9,10]	$10,600	$6,000	$5,100
Gravity[9,10]	$14,900	$7,500	$6,400
OWTS[5]			
0' WT	$7,070[6]	$8,580[7]	$10,280[8]
1' WT	$5,930[6]	$7,860[7]	$9,780[8]
2' WT	$5,050[6]	$7,340[7]	$9,410[8]
>3' WT	$4,890[6]	$6,690[7]	$8,960[8]

Notes:

1. Cost based on screening analysis of an area with 226 connections and average lot size of 0.93 acres
2. Cost based on screening analysis of an area with 810 connections and average lot size of 0.33 acres
3. Cost based on screening analysis of an area with 1,368 connections and average lot size of 0.21 acres
4. Construction costs include 20% contingency and On-Lot costs
5. Water table (WT) depth shown is the estimated SHWT depth below ground surface
6. Cost is for OWTS Alternative I: Septic tank with mounded drainfield
7. Cost is for OWTS Alternative II: Septic tank with SDI disposal
8. Cost is for OWTS Alternative IV: Advanced Secondary Treatment System
9. Costs do not include engineering, legal, and administrative costs

Estimated Annual O&M Cost (per EDU)[1, 2]

Alternative	Low Density[3] >0.5 Acre Lots	Medium Density[4] 0.25-0.5 acre lots	High Density[5] <0.25 acre lots
Low Pressure GP	$190	$190	$190
Vacuum	$140	$70	$60
Gravity	$90	$50	$50
OWTS[6]			
0' WT	$140[7]	$420[8]	$1,080[9]
1' WT	$140[7]	$420[8]	$1,080[9]
2' WT	$140[7]	$420[8]	$1,080[9]
>3' WT	$140[7]	$420[8]	$1,080[9]

Notes:

1. Costs include O&M and replacement
2. Replacement costs are annualized at an interest rate of 7% over 20 years
3. Costs based on screening analysis of an area with 226 connections and average lot size of 0.93 acres
4. Costs based on screening analysis of an area with 810 connections and average lot size of 0.33 acres
5. Costs based on screening analysis of an area with 1,368 connections and average lot size of 0.21 acres
6. Water table (WT) depth shown is estimated SHWT depth below ground surface
7. Cost is for OWTS Alternative I: Septic tank with mounded drainfield
8. Cost is for OWTS Alternative II: Septic tank with SDI disposal
9. Cost is for OWTS Alternative IV: Advanced Secondary Treatment

Table 14. Estimated Uniform Annual Cost (per connection)[9, 10, 13]

Alternative	Treatment and Transmission Cost ($/Connection)	Low Density[2] >0.5 acre lots	Medium Density[3] 0.25-0.5 acre lots	High Density[4] <0.25 acre lots
Low Pressure GP[11]	$105	$1,270[1]	$1,080[1]	$1,070[1]
Vacuum[11]	$105	$1,400[1]	$900[1]	$800[1]
Gravity[11]	$105	$1,760[1]	$1,020[1]	$910[1]
OWTS[5]				
0' WT	N/A	$800[6]	$1,230[7]	$2,050[8]
1' WT	N/A	$700[6]	$1,160[7]	$2,000[8]
2' WT	N/A	$610[6]	$1,110[7]	$1,970[8]
>3' WT	N/A	$600[6]	$1,050[7]	$1,930[8]

Notes:

1. Estimated treatment and transmission cost applied to all collection system alternatives
2. Costs based on screening analysis of an area with 226 connections and average lot size of 0.93 acres
3. Costs based on screening analysis of an area with 810 connections and average lot size of 0.33 acres
4. Costs based on screening analysis of an area with 1,368 connections and average lot size of 0.21 acres
5. Water table (WT) depth shown is below ground surface
6. Cost is for OWTS Alternative I: Septic tank with mounded drainfield
7. Cost is for OWTS Alternative II: Septic tank with SDI disposal
8. Cost is for OWTS Alternative IV: Advanced Secondary Treatment
9. Uniform Annual Cost includes materials, construction, operation and maintenance, and replacement costs
10. Costs do not include Engineering, Legal and Administrative Costs
11. Includes Capacity Fee of $1,642 per EDU
12. Includes Treatment Cost of $105 per EDU per year (Sarasota County Utilities)
13. Annualized construction cost, impact fees, and replacement costs are based on an interest rate of 7% over 20 years

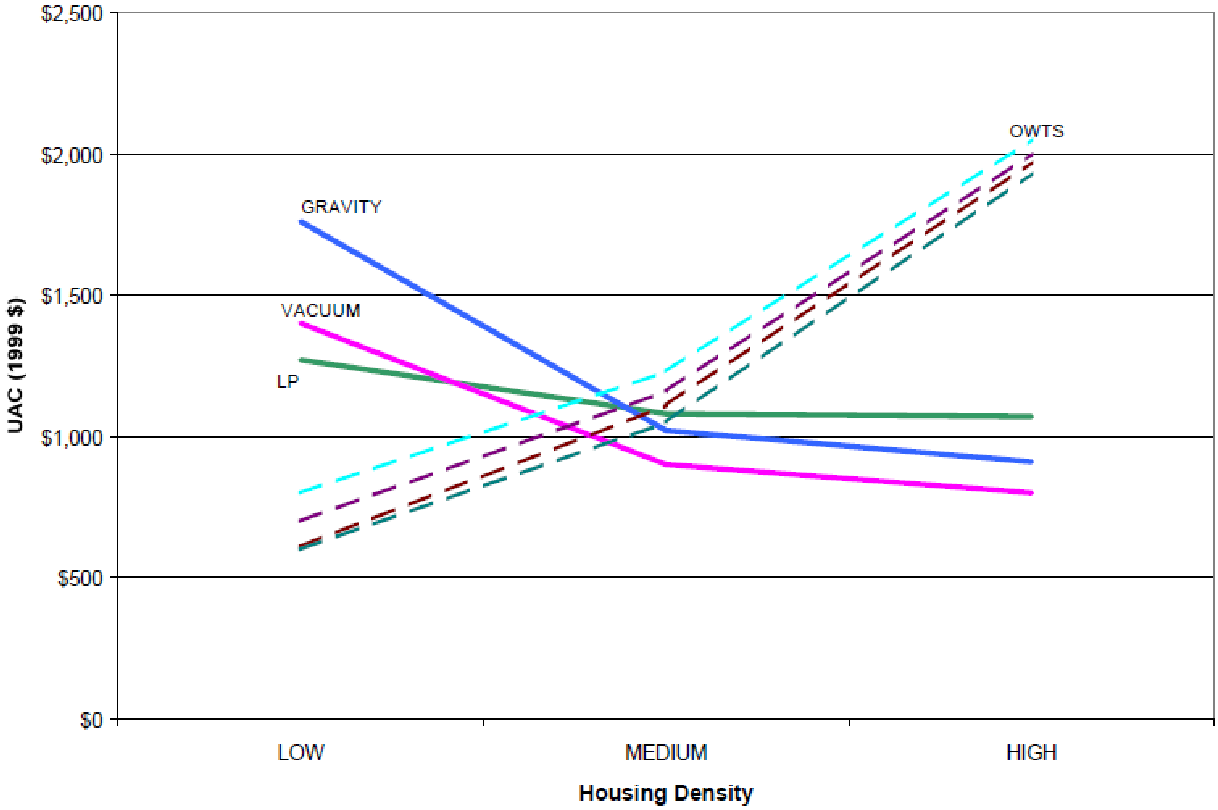


Figure 2. Cost vs. Density Comparison of OWTS and Collection Alternatives for Communities in the Phillippi Creek Project Area

Appendix 2

Data tables and compilations of the Friends of the Bay data for 2004 through 2014. Tables contain all variables and any necessary calculations to replicate figures provided in the *Annual Water Quality Report*. All tables in this section provide access to the associated spread sheets when double clicked for convenience and functionality.

Compilation of All Friends of the Bay Data For All years and All Sites[15]							
Date	H20 Temp TOP (0.5m)	H2O Temp 1.0 m	H20 Temp 0.5 m from BTM	Salinity TOP	Salinity 1.0 m	Salinity BTM	PH Top
4/3/2006	8.80	8.60	8.00	23.95	24.35	25.07	
4/10/2006	8.25	8.22	7.59	23.67	24.27	24.98	
4/17/2006	11.33	11.51	10.02	22.23	22.51	24.83	
4/24/2006	run cancelled due to weather conditions	0	0	0	0	0	

FOB unlinked2.0 (Recovered).xlsx

Compiled Raw Data from all Laurel Hollow Sites for All Years[15]						
	Date	H20 Temp TOP (0.5m)	H2O Temp 1.0 m	H20 Temp 0.5 m from BTM	Salinity TOP	Salinity 1.0 m
LH01	7/19/2010	25.60	25.45	24.50	29.83	29.90
LH01	7/26/2010	26.25	26.06	23.77	26.77	27.05
LH01	8/2/2010	23.46	23.65	21.46	27.53	27.95
LH01	8/9/2010	24.16	24.01	23.35	30.85	31.00
LH01	8/16/2010	22.88	22.75	22.65	31.23	31.44
LH01	8/23/2010	Run cancelled due to small craft advisories	XXX	XXX	XXX	XXX
LH01	8/30/2010	23.18	23.11	22.62	30.09	30.73
LH01	9/7/2010	22.82	22.84	23.28	30.79	30.94

OBS02	Beaver Lake		W	0.12	42117.00	838.00
				0.06		
OBS03	Beekman Creek		W	0.69	41500.00	0.50
OBS03	Beekman Creek		W	0.03	41613.00	0.40
				0.36		
OBS03	Beekman Creek		W	1.76	41836.00	0.38
OBS03	Beekman Creek		D	0.00	41941.00	0.42
				0.88		
OBS03	Beekman Creek		D	0.00	42026.00	1104.00
OBS03	Beekman Creek		W	0.12	42117.00	927.00

\

COMPILED OBS.xlsx

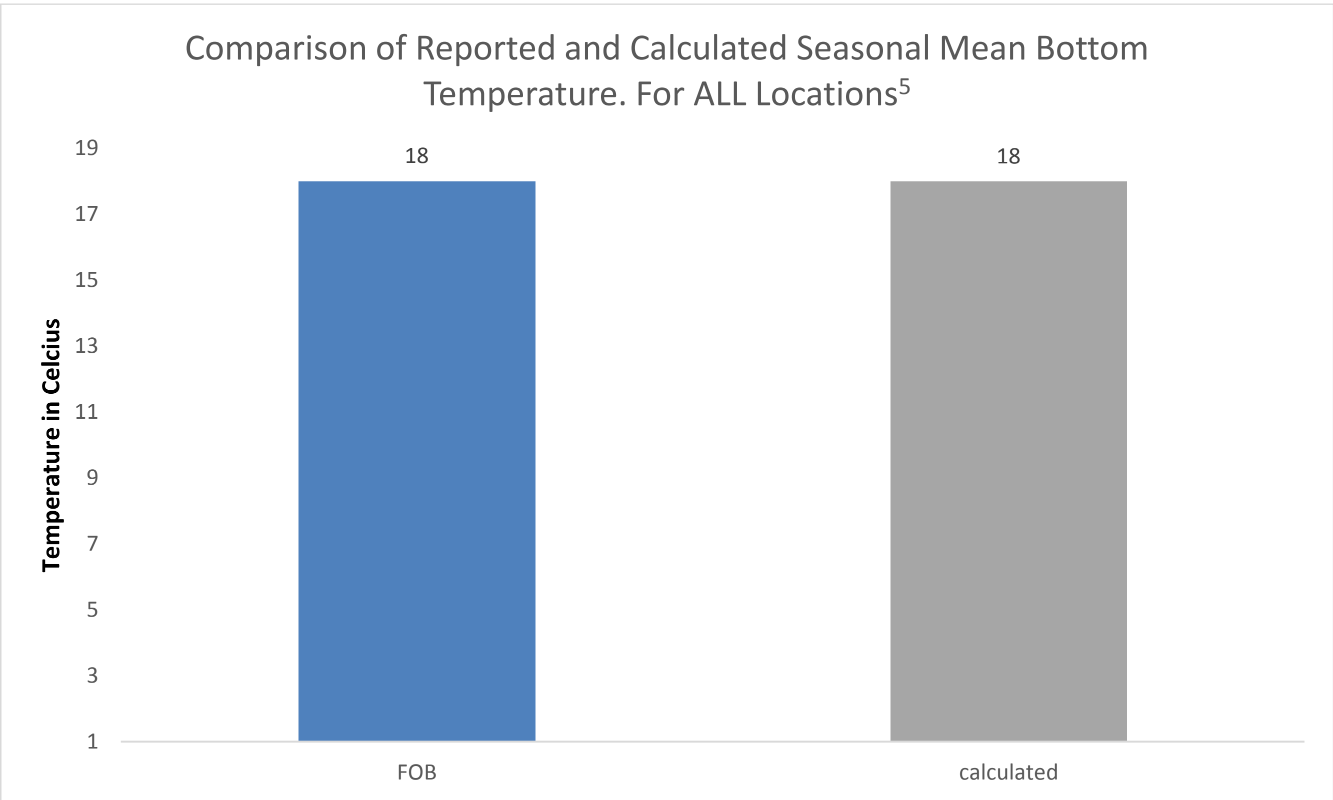
Mastersheet Containing the Mean and Geomean data for all sites organized by year[15]									
Date	ID1	W_TEMP_TOP	W_TEMP_MID	W_TEMP_BTM	Salinity_TOP	Salinity_MID	Salinity_BTM	PH_Top	PH_MID
2001	FB19	20.21	0.00	20.53	25.55	0.00	25.80	0.00	0.00
2001	FB18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001	FB17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001	FB16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001	FB15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001	FB14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001	FB13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001	FB12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

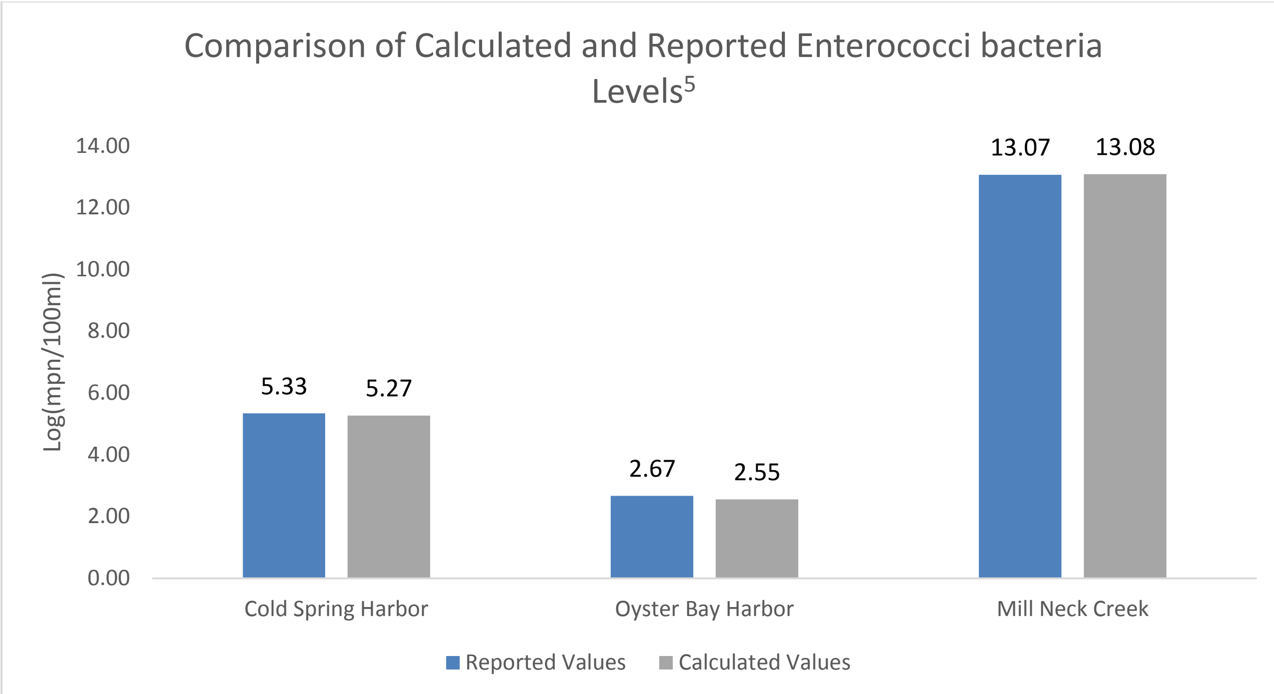
master
neglidgeable.xlsx

Locations of Testing Sites for All Agencies [15,19]						
ID	ID	Site Name	Latitude	Longitude		
FB	FB01	south cold	40.863	-73.464		
FB	FB02	CSH north	40.869	-73.463		
FB	FB03	csh south	40.873	-73.474		
FB	FB04	CSH north	40.896	-73.506		
FB	FB05	plum poin	40.901	-73.506		
FB	FB06	seawanha	40.901	-73.511		
FB	FB07	oyster bay	40.875	-73.512		
FB	FB08	white cree	40.875	-73.521		

Building Classifications[15]		
FID	OBJECTID	OWNER
5218	236165	TOWN OF OYSTER BAY
13522	229237	SWAN LAKE AT WOODBURY HOMEOWNERS ASSN
13766	405519	SCHNEIDER JEFFREY G
242	6491	PERONI ESTATE OF WOODBURY
2434	50694	NORTH SHORE WILDLIFE SANCTUARY
5538	255975	NORTH SHORE WILDLIFE SANCTUARY
13258	162619	NORTH SHORE WILDLIFE SANCTUARY
13594	171073	NORTH SHORE WILDLIFE SANCTUARY
443	297646	TOWN OF OYSTER BAY
2014	210816	INC VILLAGE OF UPPER BROOKVILLE
7986	120028	INC VILLAGE OF UPPER BROOKVILLE
593	386522	TOWN OF OYSTER BAY
775	394470	TOWN OF OYSTER BAY
1454	351012	TOWN OF OYSTER BAY
5150	321973	TOWN OF OYSTER BAY
6152	372109	TOWN OF OYSTER BAY
6310	419176	TOWN OF OYSTER BAY

Data tables used for recreation of FOB Charts[15]			
Date	Day Number	Fecal Coliform Bacteria	30-Day Geometric Mean
4/3/2006	93	5	5
4/10/2006	100	2	3.16
4/17/2006	107	22	6.04
4/24/2006	114		6.04
5/1/2006	121		6.63
5/8/2006	128		6.63
5/15/2006	135		22
5/22/2006	142		
5/26/2006	146	32	32
6/5/2006	156	80	50.6
6/12/2006	163	370	98.21
6/19/2006	170	53	84.17
6/26/2006	177	60	98.5
7/3/2006	184	62	89.79
7/10/2006	191	70	87.42
7/17/2006	198	800	102




Comparison of Calculated and Reported Enterococci bacteria Levels5
Log(mpn/100ml)
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
5.33
5.27
2.67
2.55
13.07
13.08
Cold Spring Harbor
Oyster Bay Harbor
Mill Neck Creek
Reported Values
Calculated Values

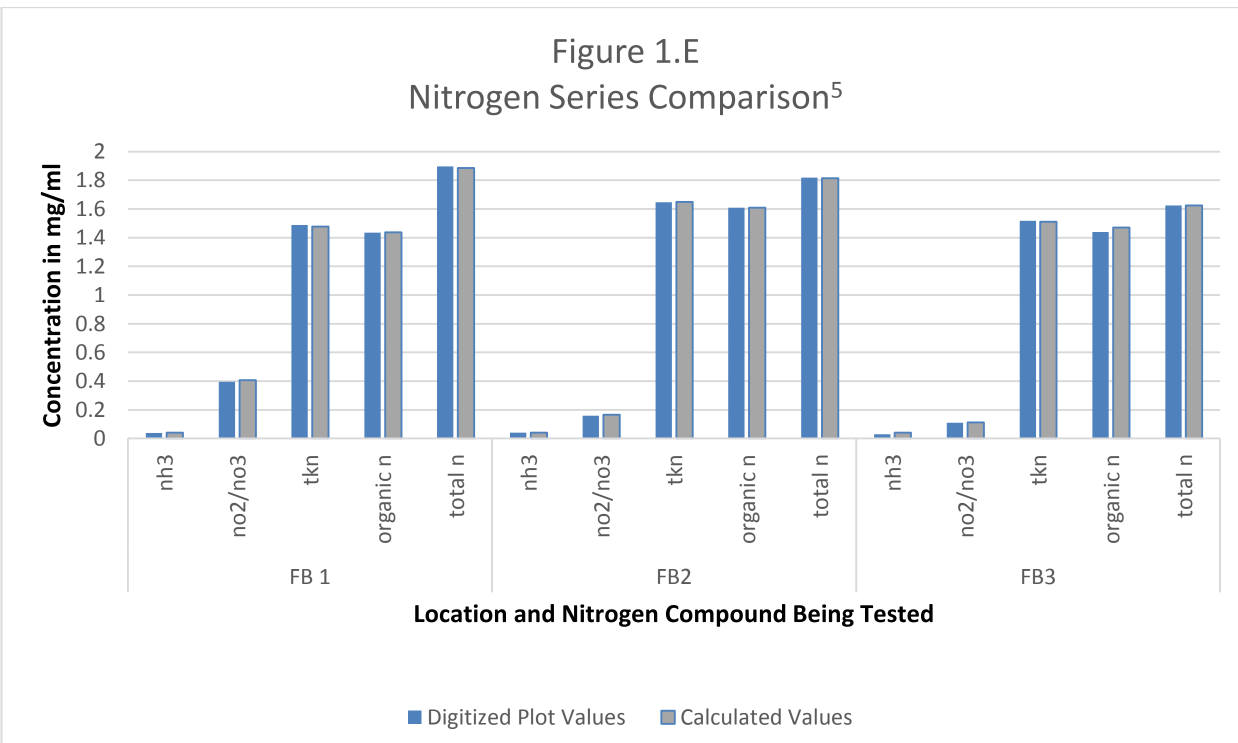
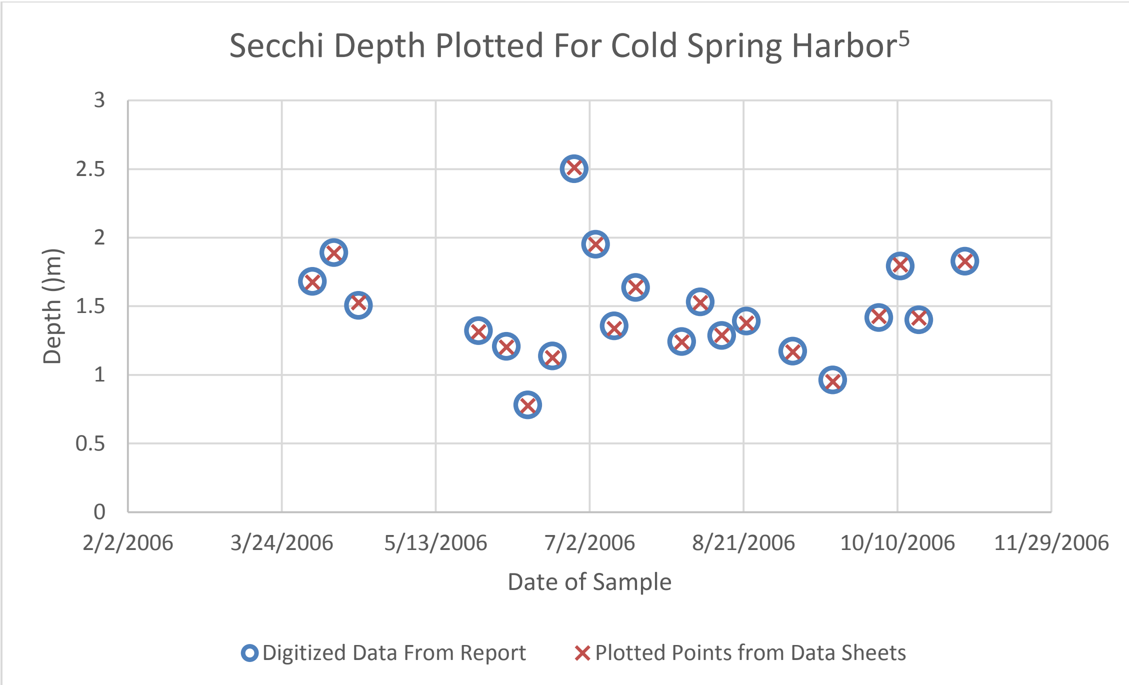


Figure 1.E
Nitrogen Series Comparison5
Concentration in mg/ml
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
nh3
no2/no3
tkn
organic n
total n
FB 1
FB2
FB3
Location and Nitrogen Compound Being Tested
Digitized Plot Values
Calculated Values


Secchi Depth Plotted For Cold Spring Harbor[5]
Depth ()m)
3
2.5
2
1.5
1
0.5
0
2/2/2006
3/24/2006
5/13/2006
7/2/2006
8/21/2006
10/10/2006
11/29/2006
Date of Sample
Digitized Data From Report
Plotted Points from Data Sheets

Appendix 3

Annotated coding for all calculations carried out in Rstudio are provided here to increase understanding of the processes and functions. These transcripts include the compilation of data, boxplot construction, goodness of fit testing[19], hypothesis testing[19], and preliminary Tidal[2] analysis. The compiled spreadsheet is also provided in this appendix.

---
title: "Processing OB_MNC and CSH Data"
author: "Thomas J Vogel"
date: "January 14, 2016"
output: html_document

This project is the location where DEC data from the NSSP and teh FOD data will be merged into a single data repository. the following are the gaols for the above taks:

Merging Data Sets
Merging FOB and DEC NSSP Data sets
Wednesday, January 13, 2016
11:35 AM
Goal:
 * Combine FOB and DEC NSSP data sets in one file that can be analyzed in R

Approach:

 * Uploaded DEC NSSP data to new folder in Google Drive
(https://drive.google.com/open?id=0B43SCEe4f5fxc2lkN09zcWN0V1E)

Files:
 * FOB 2004-2014 excel file. Link: https://drive.google.com/open?id=0B43SCEe4f5fxbzBqWTVkOGxIOXM
 * DEC NSSP Data excel file: https://drive.google.com/open?id=0B43SCEe4f5fxU3Y1d3QyTmMxQTg

Data Manipulation:
 1. Process the DEC NSSP Data so that one observation (one measurement at one site on one date) per row
 * Columns a-k are metadata. Column F denotes the sampling date.
 * Columns L through BH are sample locations
 2. Notes about FC data:
 * DEC data below LDL recorded as 2.9 MPN/100 mL. Do global search for 2.9. Leave value as 2.9 but add < in remark column to denote below DL
 * DEC data above 2400 MPN/100 mL is reported as 2501. Do global search for 2501, leave value alone but add > to remark column to indicate above 2400
 * Values where no measurements were made are left blank. This is not changed
 3. Include trip ID in merged sheet

Merging Data
 * Create merged file by saving FOB spreadsheet as new file
 * Place FC values in Merged FOB spreadsheet

* Place FC remark code in Merged sheet
* Place sampling date and location (site ID) in merged sheet
* Place rainfall in merged sheet
* Add remark code column to merged sheet
* Place rainfall value from NSSP into Rainfall column
* Merge 0-24, 24-48, 48-72, 72-96 h rainfall columns into one column, using ";" as delimiter, and place in rainfall column
* Convert tidal stage in NSSP data to code system used by FOB
* Conversion key: Tidal Terms
* Include the FOB value in the Tidal Column and move the original tidal text to the remark column

Tidal Terms:

FOB:
1 = high slack
2 = ebbing/falling
3 = low slack
4 = flooding/rising

DEC NSSP
Low Slack
EBB
Flood
High Slack
Low slack/flood
Low Ebb
High Ebb

Change NSSP as follows:
Low Slack = 3
Ebb = 2
Flood = 4
High Slack = 1
Low slack/flood = 3
Low Ebb = 2
High Ebb = 2
LOEBB/LOW SLACK=2

Files:
Copy of FC Data Compilation Oyster Bay-Mill Neck Creek.FOIL request.csv
Copy of FCDataCompilation - Cold Spring Harbor.FOIL Request.csv
FOB_WQ_2004-2014_Cleaned_KJB_20150826.csv

**these are all copies of the downloaded files and are not linked to the originals. All changes made to these files are local and will not result in alterations to the original files.

First I am calling calling the package dplyrand tidyr to help with the data manipulation and the files to be read by Rstudio.

I am now begining the process of uniting the Fecal Coliform data into a single column, adding a remarks column to the coliform data, collecting all rainfall data into a single ";" delimited column so that it is accessible but compact since it is not presently neccessary to utilize all rainfall data, add remarks column to rainfall, and converting tidal stages in NSSP dataset into the same format as FOB with addition of remarks column containing original text. First I am performign the conversion tasks on the DEC_CSH data to ensure everything functions properly.

The step below is reading the csv file into Rstudio and converting it into a table dataframe that can be manipulated.

```{r}
library(dplyr)
library(tidyr)
FB04_14 <- read.csv("FOB_2004_2014.csv")
DEC_CSH <- read.csv("DEC_CSH_FC.csv")
DEC_OB_MNC <- read.csv("DEC_OB_MNC_FC.csv")
```

After several attempts it became apparent that the cather function would only operate properly if only the desired data was in the csv. I have created a new csv for the cold spring harbor data containing only the date, SiteID's, and Coliform Data and attempted the function below. This yielded the correct table of compiled FC data in a 3 column table (Date, SiteID, and FC_Data)

repeated attempts to include all data have resulted in the following solutions. First, Rstudio reads teh rainfall information headers as data because of the dash (-) in the name, using excel and a simple find and replace I changed these dashes to underscores (_) which made the names recognizable. The Second part of the solution is that all columns that we want to be considered contants must be named in the function with a dash (-) preceding their heading ex gather(StationID, FC_Data, -SampleDate, GrowingAreaID,...)

after completing these steps a usable and complete table was produced with N/A is used by Rstudio in cells which have no measurement.

```{r}
library(dplyr)
library(tidyr)
FB04_14 <- read.csv("FOB_2004_2014.csv")
DEC_CSH <- read.csv("DEC_CSH_FC5.csv")
DEC_OB_MNC <- read.csv("DEC_OB_MNC_FC.csv")
```

I am now attemptign to gather all of the FC data for all sites into a single column in order to fit the format of FOB_2004-2014

removed column StationID, worked well

```{r}
CSH_FC <- tbl_df(read.csv("DEC_CSH_FC5.csv", stringsAsFactors=FALSE, check.names=FALSE))
CSH_FC
CSH_FC1 <- CSH_FC %>%
```

  gather(StationID, FC_Data, -SampleDate, -GrowingAreaID, -GrowingAreaNumber, -tripID, -SamplingStrategy, -SamplingCondition, -Tide, -Rainfall0_24, -Rainfall24_48, -Rainfall48_72, -Rainfall72_96)
View(CSH_FC1)
```

Below is the begining attempt to unite all rainfall data into a single coumn. To this point there was little success using the tidyr::unite() function. the most significant error encountered that I have nto been able to resolve is that the output cannot occur because the data "does not conform to an integer column and cannot be written" I am unsure why this error occurs because I made every effort to successfully transform the data in the Rainfall columns form dbl to int format, this effort succeeded upon checking however the output becomes numeric and I was unable as of yet to change this and produce a united column of rainfall data.

below is the code to convert the data from DEC_CSH_FC into integer values and store it in a name CSH. First I stored teh information from DEC_CSH_FC into the name CSH. Next I am converting the rainfall data into integer values one column at a time using the name CSH and storign the updated file into itself

```{r}
library(dplyr)
CSHU <- unite_(CSH_FC1, "Rainfall", c("Rainfall0_24", "Rainfall24_48", "Rainfall48_72",
"Rainfall72_96"), sep=";")
View(CSHU)
```
```{r}
write.csv(CSHU, file="CSHU")
```

I am now repreating the data processing for MNC and OB using the same procedure described above
first calling all necessary files and packages for the processing

```{r}
library(dplyr)
library(tidyr)
FB04_14 <- read.csv("FOB_2004_2014.csv")
DEC_CSH <- read.csv("DEC_CSH_FC5.csv")
DEC_OB_MNC <- read.csv("DEC_OB_MNC_FC.csv")
```

gathering all of the coliform data into a single column

```{r}
OB_MNC_FC <- tbl_df(read.csv("DEC_OB_MNC_FC.csv", stringsAsFactors=FALSE,
check.names=FALSE))
CSH_FC
OB_MNC_FC1 <- OB_MNC_FC %>%
  gather(StationID, FC_Data, -SampleDate, -GrowingAreaID, -GrowingAreaNumber, -tripID, -
SamplingStrategy, -SamplingCondition, -Tide, -Rainfall0_24, -Rainfall24_48, -Rainfall48_72, -Rainfall72_96)
View(OB_MNC_FC1)
```

uniting all of the rainfall data into a ";" separated column

```{r}
library(dplyr)

OB_MNCU <- unite_(OB_MNC_FC1, "Rainfall", c("Rainfall0_24", "Rainfall24_48", "Rainfall48_72", "Rainfall72_96"), sep=";")
View(OB_MNCU)
```
writing a new csv containing the processed data.
```{r}
write.csv(OB_MNCU, file="OB_MNCU")
```
---
title: "Remake Annual Plots"
author: "Thomas J Vogel"
date: "February 21, 2016"
output: html_document

First I read in the data from FOB and NSSP.
```{r}
OBCSH <- read.csv("FOB_NSSP_1987_2014_ORDERED.csv", strip.white = TRUE)
```
Next I subset the data for all sites in the Mill Neck Creek study area as well as three reference sites in the broader Oyster Bay and Cold Spring Harbor system. FB-11 and 47-8 are the designated reference sites. FB-11 and 47-8 are located approximately at teh cenbter of Oyster Bay.
```{r}
MNC_Sites <- subset(OBCSH, Site %in% c("FB-11", "FB-15","47-FOB15","FB-17","47-FOB17","FB-16","47-FOB16","FB-14","FB-13","FB-18","FB-19","47-A","47-A1","47-B","47-C","47-D","47-E","47-F", "47-8", "47-9", "47-10"))
MNC_Sites$Order <- factor(MNC_Sites$Order)
MNC_Sites$FC <- log10(MNC_Sites$FC)
```
```{r}
write.csv(MNC_Sites, file="MNC_Sites", sep=",")
```
Next I read the MNC sites csv back into the document and designated it as the object MNC. Rstudio is not writing csv's correctly so i added teh data to an excel sheet in resaved it as a new csv.
```{r}
MNC <- read.csv("MNC.csv", strip.white = TRUE)
```
I am opening the lubridate libnrary and formattign the date into year only form to make annual boxplot production easier.
```{r}
library(lubridate)
MNC$Year <- as.character(year(MNC$Date))
MNC$Month <- month(MNC$Date, label=TRUE, abbr=TRUE)
```
below I am subsetting each year individually
```{r}
y1987 <- subset(MNC, Year %in% c("1987"))
y1988 <- subset(MNC, Year %in% c("1988"))
y1989 <- subset(MNC, Year %in% c("1989"))
y1990 <- subset(MNC, Year %in% c("1990"))
y1991 <- subset(MNC, Year %in% c("1991"))
y1992 <- subset(MNC, Year %in% c("1992"))
y1993 <- subset(MNC, Year %in% c("1993"))

```
y1994 <- subset(MNC, Year %in% c("1994"))
y1995 <- subset(MNC, Year %in% c("1995"))
y1996 <- subset(MNC, Year %in% c("1996"))
y1997 <- subset(MNC, Year %in% c("1997"))
y1998 <- subset(MNC, Year %in% c("1998"))
y2001 <- subset(MNC, Year %in% c("2001"))
y2002 <- subset(MNC, Year %in% c("2002"))
y2003 <- subset(MNC, Year %in% c("2003"))
y2004 <- subset(MNC, Year %in% c("2004"))
y2005 <- subset(MNC, Year %in% c("2005"))
y2006 <- subset(MNC, Year %in% c("2006"))
y2007 <- subset(MNC, Year %in% c("2007"))
y2008 <- subset(MNC, Year %in% c("2008"))
y2009 <- subset(MNC, Year %in% c("2009"))
y2010 <- subset(MNC, Year %in% c("2010"))
y2011 <- subset(MNC, Year %in% c("2011"))
y2012 <- subset(MNC, Year %in% c("2012"))
y2013 <- subset(MNC, Year %in% c("2013"))
y2014 <- subset(MNC, Year %in% c("2014"))
```

I am now creating the annual boxplots of fecal coliform concentration organized by site

```{R}
boxplot(FC ~ Order, data=y1987, main = "1987 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y1988, main = "1988 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y1989, main = "1989 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y1990, main = "1990 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y1991, main = "1991 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y1992, main = "1992 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y1993, main = "1993 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y1994, main = "1994 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y1995, main = "1995 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y1996, main = "1996 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y1997, main = "1997 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y1998, main = "1998 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y2001, main = "2001 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y2002, main = "2002 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y2003, main = "2003 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
```

boxplot(FC ~ Order, data=y2004, main = "2004 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y2005, main = "2005 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y2006, main = "2006 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y2007, main = "2007 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y2008, main = "2008 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y2009, main = "2009 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y2010, main = "2010 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y2011, main = "2011 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y2012, main = "2012 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y2013, main = "2013 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
boxplot(FC ~ Order, data=y2014, main = "2014 Fecal Coliform Concentration by Site", xlab = "Site", ylab = "Fecal Coliform Concentration in log(MPN/100ml)")
```
below I am formatting selected boxplots for use in reports and posters
```{r}
par(mfrow=c(1,2))
boxplot(FC ~ Order, data=y2004, ylim=c(0.0, 4.0))
abline(h=c(1.146,2.30), lty=c(1,2), col="blue")
boxplot(FC ~ Order, data=y2006, ylim=c(0.0, 4.0))
abline(h=c(1.146,2.30), lty=c(1,2), col="blue")
```
```{r}
par(mfrow=c(1,2))
boxplot(FC ~ Order, data=y2008, ylim=c(0.0, 4.0))
abline(h=c(1.146,2.30), lty=c(1,2), col="blue")
boxplot(FC ~ Order, data=y2010, ylim=c(0.0, 4.0))
abline(h=c(1.146,2.30), lty=c(1,2), col="blue")
```
```{r}
par(mfrow=c(1,2))
boxplot(FC ~ Order, data=y2012, ylim=c(0.0, 4.0))
abline(h=c(1.146,2.30), lty=c(1,2), col="blue")
boxplot(FC ~ Order, data=y2014, ylim=c(0.0, 4.0))
abline(h=c(1.146,2.30), lty=c(1,2), col="blue")
```
```{r}
par(mfrow=c(1,2))
boxplot(FC ~ Order, data=y2004, main = "2004",  ylim=c(0.0, 4.0)))
boxplot(FC ~ Order, data=y2007, main = "2007",  ylim=c(0.0, 4.0)))
boxplot(FC ~ Order, data=y2010, main = "2010",  ylim=c(0.0, 4.0)))
boxplot(FC ~ Order, data=y2014, main = "2014",  ylim=c(0.0, 4.0)))
```
```

---
title: "test for normality"
author: "Thomas J Vogel"
date: "February 22, 2016"
output: html_document

First i will call the main file. I am also adding a new column containing only the year in which the sample was taken.

```{r}
FBNSSP <- read.csv("FBNSSP_87_14.csv", strip.white = TRUE)
FBNSSP$FC <- log10(FBNSSP$FC)
OBCSH <- subset(FBNSSP, Site %in% c("FB-15","47-FOB15","FB-17","47-FOB17","FB-16","47-FOB16","FB-14","FB-13","FB-18","FB-19","47-A","47-A1","47-B","47-C","47-D","47-E","47-F", "47-9", "47-10"))
```

Next I am loading the water quality package and use it to fitn the data for analysis in ENVStats

```{r}
library(wq)
OBCSH_FC_wqd <- wqData(OBCSH, c(1,3,5), 33, site.order = FALSE, time.format = "%Y-%m-%d", type = "wide")
```

Below I have made addittional subsets to the data. The first two lines of code designate testing as either during the period of FOB monitoring, a much smaller period than the NSSP, and subsets only the period of approximately 10 years that both organizations were monitoring. Following this I have created a new subset that contains data only if the FC Value is greater than zero. This measure was taken to limit the number of nondetect values included in calculations and therefore reduces noise and tail distortion.

```{r}
OBCSH_FC_wqd$FB_Monitoring <- ifelse(OBCSH_FC_wqd$time < as.Date("2004-01-01"), "No","Yes")
OBCSH_FC_wqd$FB_Monitoring <- factor(OBCSH_FC_wqd$FB_Monitoring)
MNC_FB <- subset(OBCSH_FC_wqd, FB_Monitoring %in% "Yes")
MNC_FB_wqd <- subset(MNC_FB, value > 0)
write.csv(MNC_FB_wqd, file = "MNC_WQD")
```

I am now creating a new column to designate if the sample was taken before or after the sewering of "the Birches". This is for use if the datasets prove to be sonsistent with each other and therefore able to act as a single sampling directory.

```{r}
MNC_FB_wqd$Birches <- ifelse(MNC_FB_wqd$time < as.Date("2012-01-01"), "Before","After")
MNC_FB_wqd$Birches <- factor(MNC_FB_wqd$Birches)
```

Next I am separating adding a designation of either FOB or NSSP data to test if the two can function as a single cohesive dataset. This was completed using a simple search and find function in excel and yielded the column Agency

I am now writing th enew dataframe to a separate table and reentering it into the program

```{r}
```

write.table(MNC_FB_wqd, file="MNC_FB_wqd.df")
```
```{r}
OBCSH.birches <- read.table("MNC_FB_wqd.df")
str(OBCSH.birches)
```
next i am activating ENVStats and entering the data into the analysis. I am also doing a quick analysis of the dataset
```{r}
library(EnvStats)
attach(OBCSH.birches)
summaryFull(value, data = OBCSH.birches)
```
Next I will create the Quantile plots. First for the FOB.
```{r}
ecdfPlot(value, xlab = "Log [Fecal Coliform (MPN per 100 mL)]")
abline(v=c(1.146,2.30), lty=c(1,2), col="blue")
```
and testing the distribution of data and a comparison of the fits
```{r}
cdfCompare(value, dist="norm", xlab="Order Statistics for Log [Fecal Coliform (MPN per 100 mL)] FB")
legend(2.0,0.4,legend=c("Empirical CDF", "Fitted Normal CDF"), lty = 1:2, col=c(4,1), lwd=3, bty="n")
```
Next teh QQ Plots
```{r}
qqPlot(value, dist="norm", add.line = TRUE, points.col = "blue", ylab = "Quantiles of Log [Fecal Coliform (MPN per 100 mL)]")
qqPlot(value, dist="norm", plot.type="Tukey", estimate.params = TRUE, add.line = TRUE, points.col = "blue")
```
and goodness of fit analysis
```{r}
FBNSSP.ref <- value
sw.norm <- gofTest(FBNSSP.ref, dist = "norm")
sw.norm
plot(sw.norm, digits=3)
sw.norm <- gofTest(FBNSSP.ref)
sw.norm
plot(sw.norm, digits=3)
```
```

---
title: "”the Birches” hypothesis testing"
author: "Thomas J Vogel"
date: "February 22, 2016"
output: html_document

First i will call the main file. I am also adding a new column containing only the year in which the sample was taken.

```{r}
FBNSSP <- read.csv("FBNSSP_87_14.csv", strip.white = TRUE)
FBNSSP$FC <- log10(FBNSSP$FC)
OBCSH <- subset(FBNSSP, Site %in% c("FB-15","47-FOB15","FB-17","47-FOB17","FB-16","47-FOB16","FB-14","FB-13","FB-18","FB-19","47-A","47-A1","47-B","47-C","47-D","47-E","47-F", "47-9", "47-10"))
```

Next I am loading the water quality package and use it to fitn the data for analysis in ENVStats

```{r}
library(wq)
OBCSH_FC_wqd <- wqData(OBCSH, c(1,3,5), 33, site.order = FALSE, time.format = "%Y-%m-%d", type = "wide")
```

Below I have made addittional subsets to the data. The first two lines of code designate testing as either during the period of FOB monitoring, a much smaller period than the NSSP, and subsets only the period of approximately 10 years that both organizations were monitoring. Following this I have created a new subset that contains data only if the FC Value is greater than zero. This measure was taken to limit the number of nondetect values included in calculations and therefore reduces noise and tail distortion.

```{r}
OBCSH_FC_wqd$FB_Monitoring <- ifelse(OBCSH_FC_wqd$time < as.Date("2004-01-01"), "No","Yes")
OBCSH_FC_wqd$FB_Monitoring <- factor(OBCSH_FC_wqd$FB_Monitoring)
MNC_FB <- subset(OBCSH_FC_wqd, FB_Monitoring %in% "Yes")
MNC_FB_wqd <- subset(MNC_FB, value > 0)
write.csv(MNC_FB_wqd, file = "MNC_WQD")
```

I am now creating a new column to designate if the sample was taken before or after the sewering of "the Birches". This is for use if the datasets prove to be sonsistent with each other and therefore able to act as a single sampling directory.

```{r}
MNC_FB_wqd$Birches <- ifelse(MNC_FB_wqd$time < as.Date("2012-01-01"), "Before","After")
MNC_FB_wqd$Birches <- factor(MNC_FB_wqd$Birches)
```

Next I am separating adding a designation of either FOB or NSSP data to test if the two can function as a single cohesive dataset. This was completed using a simple search and find function in excel and yielded the column Agency

I am now writing th enew dataframe to a separate table and reentering it into the program

```{r}
write.table(MNC_FB_wqd, file="MNC_FB_wqd.df")
```

```{r}
OBCSH.birches <- read.table("MNC_FB_wqd.df")
str(OBCSH.birches)
```

next i am activating ENVStats and entering the data into the analysis. I am also doing a quick analysis of the dataset
```{r}
library(EnvStats)
attach(OBCSH.birches)
summaryFull(value ~ site, data = OBCSH.birches)
```
Next I will create the Quantile plots. First for the FOB.
```{r}
ecdfPlot(value[Birches == "Before"], xlab = "Log [Fecal Coliform (MPN per 100 mL)]")
abline(v=c(1.146,2.30), lty=c(1,2), col="blue")
```
next for the NSSP
```{r}
ecdfPlot(value[Birches == "After"], xlab = "Log [Fecal Coliform (MPN per 100 mL)]")
abline(v=c(1.146,2.30), lty=c(1,2), col="blue")
```
and testing the distribution of data and a comparison of the fits
```{r}
cdfCompare(value[Birches == "Before"], dist="norm", xlab="Order Statistics for Log [Fecal Coliform (MPN per 100 mL)] FB")
legend(2.0,0.4,legend=c("Empirical CDF", "Fitted Normal CDF"), lty = 1:2, col=c(4,1), lwd=3, bty="n")
cdfCompare(value[Birches == "After"], dist="norm", xlab="Order Statistics for Log [Fecal Coliform (MPN per 100 mL)] NSSP")
legend(2.0,0.4,legend=c("Empirical CDF", "Fitted Normal CDF"), lty = 1:2, col=c(4,1), lwd=3, bty="n")
cdfCompare(value[Birches == "Before"], value[Birches == "After"], xlab="Order Statistics for Log [Fecal Coliform (MPN per 100 mL)]")
legend(2.0,0.4,legend=c("Before", "After"), lty = 1:2, col=c(4,1), lwd=3, bty="n")
```
Next the QQ Plots
```{r}
qqPlot(value[Birches == "Before"], dist="norm", add.line = TRUE, points.col = "blue", ylab = "Quantiles of Log [Fecal Coliform (MPN per 100 mL)]")
qqPlot(value[Birches == "After"], dist="norm", add.line = TRUE, points.col = "blue", ylab = "Quantiles of Log [Fecal Coliform (MPN per 100 mL)]")
qqPlot(value[Birches == "Before"], dist="norm", plot.type="Tukey", estimate.params = TRUE, add.line = TRUE, points.col = "blue")
qqPlot(value[Birches == "After"], dist="norm", plot.type="Tukey", estimate.params = TRUE, add.line = TRUE, points.col = "blue")
qqPlot(value[Birches == "Before"], value[Birches == "After"], add.line = TRUE, points.col = "blue", xlab = "Quantiles of Log [Fecal Coliform (MPN per 100 mL)]", ylab = "Quantiles of Log [Fecal Coliform (MPN per 100 mL)]")
qqPlot(value[Birches == "Before"], value[Birches == "After"], plot.type="Tukey", estimate.params = TRUE, add.line = TRUE, points.col = "blue")
```
and goodness of fit analysis
```{r}
Before.ref <- value[Birches == "Before"]
sw.norm <- gofTest(Before.ref, dist = "norm")
sw.norm
plot(sw.norm, digits=3)
After.ref <- value[Birches == "After"]
sw.norm <- gofTest(After.ref, dist = "norm")
```

sw.norm
plot(sw.norm, digits=3)
sw.norm <- gofTest(Before.ref, After.ref)
sw.norm
plot(sw.norm, digits=3)
twoSampleLinearRankTest(value[Birches == "Before"], value[Birches == "After"], alternative = "greater")
```
and a test for tail shift
```{r}
quantileTest(value[Birches == "Before"], value[Birches == "After"], alternative = "greater", target.r = 9)
```
```{r}
boxplot(value ~ Birches, data = OBCSH.birches, main = "Fecal Coliform Distribution Before and After Re-Sewering", xlab = "Birches", ylab = "Log [Fecal Coliform (MPN per 100 mL)]")
abline(h=c(1.146,2.30), lty=c(1,2), col="blue")
```
`
```

---
title: "Preliminary WTDSTidal"
author: "Thomas J Vogel"
date: "March 6, 2016"
output: html_document

In this file I am attempting to run the WRTDSTidal analysis using the Friends of the Bay Data in Mill Neck Creek. I am restricting analysis to this dataset due to limited availability of salinity data for teh NSSP testing sites and a poor fit to a normal logrithmic distribution. For this reason it is only feasable to apply the analysis to the FOB dataset which is both complete and consistent.

First I am loading the master file and subsetting only the Friends of the Bay sites, excluding reference sites. I have also performed a log of the Fecal Coliform data to make the measurements more manageable and better able to be charted. The final step included in this code chunk is to remove all samples that contain no measurements.
```{r}
FB_NSSP <- read.csv("FBNSSP_87_14.csv", strip.white = TRUE)
FB_NSSP$FC <- log10(FB_NSSP$FC)
FB_MNC <- subset(FB_NSSP, Site %in% c("FB-15","FB-17","FB-16","FB-14","FB-13","FB-18","FB-19"))
FB_MNC <- subset(FB_MNC, FB_MNC$FC > 0)
```
Next I am creating a new subset containign only the data required for WRTDSTidal, this is to ensure that all calculations are carried out correctly.
```{r}
FB_ <- data.frame(FB_MNC)
FB_$date <- FB_MNC$Date
FB_$res <- FB_MNC$FC
FB_$flo <- FB_MNC$Salinity_0.5
FB_$lim <- FB_MNC$FC_LOD
```
Finally I am subseting the 4 variables required for WRTDSTidal: date, res (log transformed FC measurements -> Log(FC)), flo (salinity serves as a metric of flow -> salinity_0.5), and lim (detection limit FC measurements -> Log(3MPN)). I also have called the lubridate library and used the ymd() function to convert the date to POSIXlt or POSIXct format which is required by Tidal to analyze data.
```{r}
FB_Tidal <- subset(FB_, select = c(date, res, flo, lim))
library(lubridate)
FB_Tidal$date <- ymd(FB_Tidal$date)
```
I am now calling the WRTDSTidal package and begining the data analysis by presenting the model fitting as described in the manual.I am also converting the dataset FB_Tidal into a tidal object which makes it compatible with the WRTDSTidal analysis and plotting the raw data.
```{r}
library(WRTDStidal)
tidfit <- modfit(FB_Tidal, tau = c(0.1, 0.5, 0.9))
tidfitmean <- modfit(FB_Tidal, resp_type = 'mean')
```

~~~
tidobj <- tidal(FB_Tidal)
obsplot(tidobj)
```
```{r}
obsplot(tidfit)
obsplot(tidfitmean)
```
```{r}
head(tidobj)
names(attributes(tidobj))
head(tidfit)
names(attributes(tidfit))
```
mod <- modfit(FB_Tidal) %>%
  wrtds %>%
  respred %>%
  resnorm
library(dplyr)
```{r}
data(tidfit)
fitplot(tidfit)
fitplot(tidfit, annuals = FALSE)
```
```{r}
sliceplot(tidfit)
```
```{r}
fitmoplot(tidfit, predicted = F)
```
seasplot(tidfit)
```{r}
prdnrmplot(tidfit)
prdnrmplot(tidfit, annuals = FALSE)
```
```{r}
dynaplot(tidfit)
```
```{r}
gridplot(tidfit)
wtsplot(tidfit, ref = '2004-01-01')
wtsplot(tidfit, ref = '2012-01-01')
```
```{r}
nobsplot(tidfit)
```
~~~

Table			
FOB_NSSP_1987-2014[15,16]			
Site	Agency	wq_sampling_depth_remark	wq_sampling_depth
FB-1	FB		0.3
FB-10	FB		0.3
FB-11	FB		0.3
FB-12	FB		0.3
FB-13	FB		0.3
FB-14	FB		0.3
FB-15	FB		0.3
FB-16	FB		0.3
FB-17	FB		0.3
FB-18	FB		0.3
FB-19	FB		0.3
FB-2	FB		0.3
FB-3	FB		0.3
FB-4	FB		0.3
FB-5	FB		0.3
FB-6	FB		0.3
FB-7	FB		0.3
FB-8	FB		0.3
FB-9	FB		0.3
FB-1	FB		0.3
FB-10	FB		0.3
FB-11	FB		0.3
FB-12	FB		0.3
FB-13	FB		0.3
FB-14	FB		0.3
FB-15	FB		0.3
FB-16	FB		0.3
FB-17	FB		0.3
FB-18	FB		0.3
FB-19	FB		0.3
FB-2	FB		0.3