Water Quality Monitoring of Maryland's Tidal Waterways

REU Site: Interdisciplinary Program in High Performance Computing

Rosemary K. Le[1] Christopher V. Rackauckas[2], Anne S. Ross[3], Nehemias Ulloa[4]

[1]Department of Applied Mathematics, Brown University

[2]Department of Mathematics, Oberlin College

[3]Department of Computer Science, Department of Statistics, Colorado State University

[4]Department of Mathematics, California State University, Bakersfield

Abstract

The Chesapeake Bay and its surrounding tributaries are home to over 3,600 species of plants and animals. In order to assess the health of the region, the Maryland Department of Natural Resources (DNR) monitors various parameters, such as dissolved oxygen, with monitoring stations located throughout the tidal waterways. Utilizing data provided by DNR, we assessed the waterways for areas of water quality concern. We analyzed the percentage of the readings taken for each parameter that failed to meet the threshold values and used the Wilcoxon Signed-Rank Test to determine the statuses of the stations. In order to assess the applicability of the Wilcoxon Test given the positive skew in the data, a simulation was performed. This simulation demonstrated that log-transforming the data prior to performing the Wilcoxon Test helped reduce the Type I Error. Furthermore, we ranked the stations using a set of multiple comparison methods: a version of the Tukey Test on variance-transformed proportions, the Bonferroni adjustment method, a Bayesian method, and the Benjamini-Hochberg rejection method. Also, we conducted a trend analysis for the five stations in the Corsica River. To aid in the presentation of our analyses, we constructed an interactive GUI (Graphical User Interface) for use by researchers and the general public.

Key Words: Chesapeake Bay, Water Quality Monitoring, Wilcoxon Signed-Rank Test, Multiple Comparison Tests, Simulation, Benjamini-Hochberg method

1 Introduction

The Chesapeake Bay and its surrounding waterways provide a habitat for over 3,600 species of plants and animals. It is a valuable resource, both recreationally and commercially, to those who live in the basin. DNR operates 35 continuous monitoring stations with some of the stations taking readings at multiple depths for a total of 38 stations of interest. The majority of these stations take readings every 15 minutes, although some stations take readings every 10 or every 20 minutes. Various parameters such as water clarity, dissolved oxygen, and chlorophyll, are used to help determine the health of the water [3]. Past analyses of this data have been used to determine trends in different regions of the tidal waterways and aid in assessing the success of projects to clean.

1

Figure 1.1: The Chesapeake Bay and its tributaries are located on the East Coast, spanning six states (MD, VA, WV, PA, NY, DE) and Washington DC and emptying into the Atlantic Ocean.

Three parameters are of particular importance in water quality monitoring: dissolved oxygen, chlorophyll, and turbidity. Chlorophyll is the measure by which algae levels are evaluated. An increase in algae levels corresponds to a decrease in water clarity and has a negative impact on dissolved oxygen levels as the algae decomposes. Turbidity measurements are an important measure to monitor water clarity which is necessary for light to reach submerged aquatic vegetation (SAV) and promote growth. Turbidity is measured in NTUs. With both of these parameters (chlorophyll and turbidity) a higher reading corresponds to less healthy water. Contrastingly, when considering dissolved oxygen concentrations, higher readings are preferable. While a reading of 5mg/L is a widely considered failure threshold, a threshold of 3mg/L is often used to test for waters that are severely oxygen-deficient [6].

Summer is a time of particular interest for analyzing these parameters. When evaluating the concentration of dissolved oxygen, the time frame considered is June through September, the warmest months of the year. Deficient oxygen levels are most prevalent during these warm summer months because warmer water holds less dissolved oxygen. When evaluating chlorophyll and turbidity, April through September is the time frame of most concern because this period is the growing season of aquatic vegetation. Thus, high levels of chlorophyll and turbidity can have the most detrimental effect on the ecosystem during these months.

Our project focused on four parameters including those listed above and pH. We included

Table 1.1: Failure threshold and time frame of greatest interest.

Parameter	Failure Threshold	Time Frame
Dissolved Oxygen (severe)	$< 3\text{mg/L}$	June to September
Dissolved Oxygen	$< 5\text{mg/L}$	June to September
Turbidity	> 7 NTU	April to September
Chlorophyll	$> 30\mu\text{g/L}$	April to September
pH	< 5.5 or > 8.3	April to September

pH, a measure of the acidity of water, because extreme pH levels are detrimental to aquatic wildlife. The failure threshold(s) of each parameter and its time frame of greatest interest are summarized in Table 1.1.

In this project, we assessed the station performances using the Wilcoxon Signed-Rank Test. This test assumes the distribution from which the data is taken to be symmetric. However, portions of our data set appear to violate this assumption. To assess the impact of these violations, we simulated the performance of the Wilcoxon Signed-Rank Test on a skewed distribution. After assessing the stations' statuses, we ranked the stations with respect to their performance using multiple ranking techniques: a version of the Tukey Test on variance-transformed proportions, the Bonferroni adjustment method, a Bayesian method, and the Benjamini-Hochberg rejection method. From these multiple ranking results, we attempted to deduce the ranking technique that is most applicable to our data. We also performed trend analyses on data from the Corsica River in order to better understand how this area has been faring over the past six years. We constructed a a Graphical User Interface (GUI) to aid in the exploratory analysis and visualization of this data set. The GUI can be used to an exploratory tool that can be made avaialbel to all Bay scientists.

This technical report is organized as follows. Section 2 outlines the statistical methodologies used in the assessment. A brief descritpion of each method used is given. Since the assessment involves numerous pairwise comparisons, special care is needed to conserve the overall Type I Error (False Positive) rates are needed. Improved methods are also illustarted in Section 2. Section 3 considers ranking the stations in order of the severity of degradation. This is proposed as a way of agumenting the assessment approach and perhaps helpful in the planning of Bay-wide restoration efforts. We conclude the report with final remarks in Section 4.

2 Methodologies

2.1 Percent Failure and Classification of Stations

We define percent failure as the number of readings that did not meet the DNR-provided threshold divided by the total number of readings. For each station, we determined the percent failure in our timespan for each parameter: dissolved oxygen, chlorophyll, turbidity and pH. Percent failure gave us a basic value to work with. We then proceeded to use statistical tests that provide more complex insights and allow us to make more statistically sound conclusions.

One such test was the Wilcoxon's Signed-Rank Test. The Wilcoxon Signed-Rank Test is a non-parametric test that was used to compare the median of the station's data for a given parameter against the threshold value. The test assumes the the probability distribution from which the data is taken is symmetric. Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$ be the data vector. The test statistic is given by

$$S = \left| \sum_{i=1}^{m} \left[R_i \cdot \text{sign}(x_i - thresh) \right] \right|,$$

where R_i is the rank of $|x_i - thresh|$ in ascending order and the sign function is defined as

$$\text{sign}(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ \text{-1} & \text{if } x < 0. \end{cases}$$

The test rejects the null hypothesis that the median of the data is equal to the threshold if S is sufficiently large. For the purposes of our study, the Wilcoxon Signed-Rank test was used to test the hypothesis:

$$H_0 : \text{median}(\boldsymbol{x}) = thresh$$

$$H_A : \text{median}(\boldsymbol{x}) \neq thresh.$$

The station statuses were defined as "Good" or "Bad" if the Wilcoxon Test rejected the null hypothesis, depending upon if the station's median for that parameter fell on the "unhealthy" or "healthy" side of the threshold value. The station was assigned "Borderline" if the null hypothesis could not be rejected. Our tests were conducted with $\alpha = 0.01$. In order to ensure the familywise Type I Error was α, we utilized the Benjamini-Hochberg rejection method. This method ranks the p-values in ascending order and rejects the null hypotheses corresponding to the lowest p-values until the cumulative sum of the rejected p-values is equal to α. Intuitively, this method rejects the null hypothesis of the tests where the null hypothesis is most likely to be false, that is the tests that have the lowest p-values, until the cumulative p-value is the chosen α.

2.2 Simulation

The Wilcoxon Signed-Rank Test is implemented under the assumption that the distribution from which the samples are drawn is symmetric. However, examples such as Figure 2.1
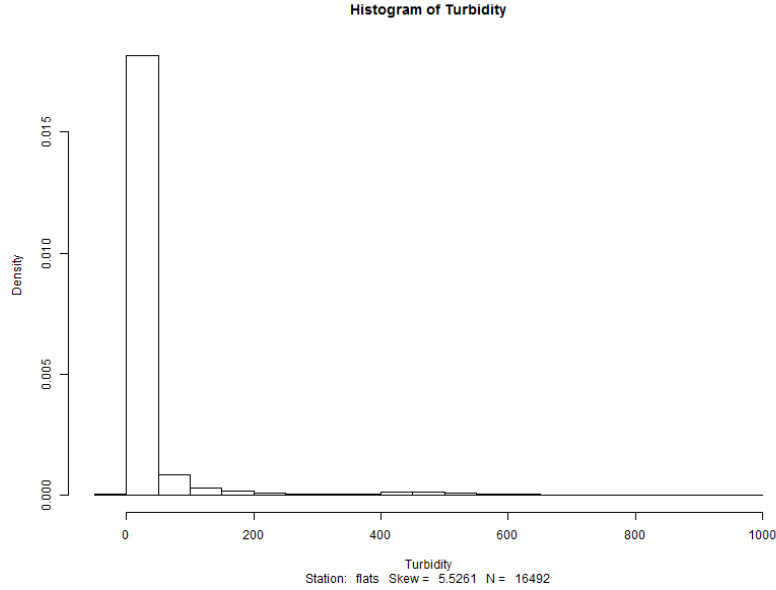
Figure 2.1: Turbidity readings from Flats station is an example of skewness in the data.

show that certain parameters such as turbidity and chlorophyll displayed postively-skewed distributions. We decided to run a simulation to assess the impact of the data's skew on the Type I Error rate for the Wilcoxon Signed-Rank Test. For the purposes of our study, we used sample data $(x_1, \ldots, x_n) \sim \Gamma(\alpha_{shape}, \beta)$, where the gamma distribution is defined as having the probability density function

$$f(\alpha_{shape}, \beta) = \frac{\beta^{\alpha_{shape}}}{\Gamma(\alpha_{shape})} x^{\alpha_{shape}-1} e^{-\beta x},$$

where Γ is the gamma function. This simulation is performed with sample size $n = 1000$ on each gamma distribution with various values for shape and rate. For accuracy, the mean of the Type I Errors from 10,000 simulations was taken as the estimate for the Type I Error.

The simulation was performed using the **R** statistical programming environment. We generated random samples from the gamma distribution in **R** using the `rgamma()` function and calculated its median using the `qgamma()` function. On each sample, the Wilcoxon Signed-Rank Test was used to test the sample median against the true median of the gamma distribution. We tested the null hypothesis that the sample median does not differ from the true median with the alternative hypothesis that the medians were different. To summarize,

$$H_0 : \mathrm{median}(\boldsymbol{x}) = \mathrm{median}(\Gamma(\alpha_{shape}, \beta))$$

and

$$H_A : \mathrm{median}(\boldsymbol{x}) \neq \mathrm{median}(\Gamma(\alpha_{shape}, \beta)).$$

The Type I Error rate was the percentage of tests which rejected the null hypothesis. We performed the simulation using a claimed significance level $\alpha = 0.01$. However, with a skewed

5

distribution, such as that of the gamma function, we suspected that the Wilcoxon test will have a true Type I Error rate greater than the claimed 0.01. If the calculated Type I Error rate is less than our claimed $\alpha = 0.01$, then this would imply that the Wilcoxon Signed-Rank test is conservative for postively skewed data and therefore inefficient when applied to skewed distributions. If the Type I Error rate is greater than the claimed $\alpha = 0.01$, this would imply that the outcome of the Wilcoxon test is adversely affected by the lack of symmetry. In this case, we may find that the Wilcoxon Signed-Rank test may not be the most appropriate test to use to assess the status of the stations. Results of simulation study are given in Section 3.

2.3   Rankings of Salinity Regimes

In addition to classifying the stations' performance, we used the Tukey Honestly Significant Difference (HSD) test to see whether the salinity content of the station affects the station's performance. The Tukey HSD test is implemented by dividing all stations into salinity regimes and calculating the mean value of the percent fail (in terms of a particular parameter) within each salinity regime. The mean percent fail of each regime is then compared to all the others to test for significant differences between regimes. This comparison is based on the studentized range distribution. Significant differences between regimes indicate that salinity content significantly impacts stations' performance. The Tukey HSD test is conducted under the assumption that all observations are independent and that there is equal variation across variations (homoscedasticity). Classifications of each station's salinity regime was provided by DNR.

To implement the Tukey HSD we paired each station's percent failure with their corresponding salinity regime. Once these two were paired, we fit them to the ANOVA model using the `aov` function in **R** with the percent failures as the response variables and salinity regimes as the explanatory variables. After fitting the ANOVA model, we used the `TukeyHSD` function in **R**.

2.4   Rankings of Stations

In addition to analyzing the statuses of the stations, we ranked the stations' performance. The result of the ranking techniques presents a set of stations that are tied in performance, thus giving groupings of stations that are linked by their similarities. To rank the stations, one must compare each station to all other stations and thus perform $\binom{n}{2}$ tests, where n is the number of stations. In order to bound the Type I Error rate, we utilized multiple comparison tests. We implemented four different ranking methods based on: the Tukey Test, the Bonferonni Test, and the Bayesian Ranking method using the stations' percent failures, and the Benjamini-Hochberg method using the stations' percent failures and means. Table 2.1 provides a brief summary of the ranking techniques, with more details in the following sub-sections.

Table 2.1: A summary of ranking techniques used.

Method	Description
Tukey	Performs variance transform on the data; obtains the q statistic in the studentized range distribution
Bonferroni	Tests each of the $\binom{n}{2}$ hypotheses at a statistical significance level of $\frac{\alpha}{\binom{n}{2}}$, where n is the number of stations
Benjamini-Hochberg	For a given α, finds the largest k such that $P_{(k)} \leq \frac{k}{m}\alpha$, where m is the number of statistical tests
Bayesian Ranking	Transforms an assumed prior distribution to a posterior distribution; parameter estimates are then obtained from the posterior to rank the stations

2.4.1 Tukey Test

We implemented one of our ranking methods by performing the Tukey-Type Multiple Comparison Testing among the stations. This testing involves a comparison of each possible combination of the stations. In order to make the comparison, the percent fail first must be transformed by this function,

$$p' = \frac{1}{2}\left[\arcsin\sqrt{\frac{X}{n+1}} + \arcsin\sqrt{\frac{X+1}{n+1}}\right],$$

where X is the number of readings above the threshold and n is the number of observations in the sample (station). Once the percent failures are transformed, we rank them based on their numerical value from smallest to largest. Then, differences are computed between all pairs of stations: the largest and smallest, the largest and second smallest, etc. A standard error (SE) is then computed for each of the pairs using the following formula,

$$SE = \sqrt{\frac{410.35}{n_A + 0.5} + \frac{410.35}{n_B + 0.5}},$$

where n_A is the sample size of one of the stations and n_B is the sample size of the other station in the pair being compared. After the difference and SE are computed for the pair, they are used to compute the test statistic:

$$q = \frac{p_A - p_B}{SE}.$$

This 'q' statistic is then compared to a $q_{0.01}$ critical value, which comes from the q-distribution attributed to the well-known statistician John Tukey, with $\alpha = 0.01$, degrees of freedom, v, equal to ∞ and $k =$ the number of groups (in this case $\binom{38}{2}$) given in Table B.5 of [7]. If the 'q' statistic $> q_{0.01}$ critical value, then we reject H_0, which assumed the stations were not

7

equal, and conclude the stations are the same; otherwise, we fail to reject H_0 and look to the original percent failure values to see which station is faring worse. From this, a pair-wise comparison method is used to determine the ranking.

2.4.2 Benjamini-Hochberg Method

A more recently developed Benjamini-Hochberg method of multiple comparisons is based the idea of controling the overall Flase Discovery Rate (FDR) because it controls the expected proportion of falsely rehjected hypotheses [2]. In this task, we have $\binom{38}{2}$ stations or $\binom{4}{2}$ salinity regimes. Thus, detting $\alpha = 0.01$, the target is to reject no more than 10% of all tested null hypotheses, under the assumption that all compared pairs are indeed equal.

The implementation of Benjamini-Hochberg method proceeds as follows [5]:

1. Sort the P-values $P_{(1)} \ldots P_{(m)}$ where m is the number of tests

2. Finding the largest k such that $P_{(k)} \leq \frac{k}{m} \alpha$

3. Reject $P_{(1)} \ldots P_{(k)}$.

For pairs of stations that are found significant, the winner for the test is the station with the lower percent fail, or the mean corresponding to the 'healthier' value. This process is repeated, keeping track of each station's number of wins. The stations are then listed from most to least wins, creating a rank from best to worst condition.

2.4.3 Bayesian Method

Our analysis also included ranking the stations using a Bayesian ranking method. It consists of estimating the ranks of certain unknown distribution parameters by ranking corresponding sample estimates. We model the distribution of the sample estimates for each station, x_i, by

$$x_i | \boldsymbol{\theta} \overset{\text{ind}}{\sim} N(\theta_i, \sigma_i^2), \qquad i = 1, \ldots, k,$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ is the vector of unknown quantities (such as the true percent failures). Let $\boldsymbol{\sigma} = (\sigma_1^2, \ldots, \sigma_k^2)$ denote the vector of variances of the estimates $\boldsymbol{x} = (x_1, \ldots, x_k)$. Since percent of failures are usually computed from a large sample, this assumption of normality should be defnsible. Bayesian ranking procedure is based on the compoutation of posterior probabilities for all possible rankings. It determines the rank of θ_i by ordering the sample estimates and associating each x_i with a corresponding rank, $\widehat{r}_i$.

To implement this method, we let $\boldsymbol{S}$ be an $n \times 38$ matrix where columns correspond to 38 stations. Let $r = 1$. For each θ_i in $\boldsymbol{S}$, where $i \in \{1, \ldots, 38\}$, compute the probability that θ_i is rank r. This implementation is based on simulating the posterior distribution and n corresponds to the simulation sample size which is set to a large number such as 1000. To compute the posterior probability, we find the maximum element in each row and set it equal to 1 while all other elements in the row are set to 0, and then compute the column sums. Once the column sums are obtained, we divide each sum by the total number of rows

8

in S. Let θ_j represent the station with the highest probability and associate the rank, r, with θ_j. Then eliminate that column from matrix S so that S is now an $n \times (38-r)$ matrix. Continue this process, incrementing r, until all the stations are ranked.

For more on this method and other Bayesian ranking systems, please see [1].

2.5   Trend Analysis

Given that the distribution of the data is unknown, non-parametric techniques are used in the trend analysis. The two non-parametric tests for trend used are the Seasonal Mann-Kendall Test and the Theil-Sen Estimator (also known as Sen's Estimate). The Seasonal Mann-Kendall Test is used to establish the existence of trend while Sen's Estimate is an estimation of the trend value.

To understand the basis for the Seasonal Mann-Kendall Test, one must first be familiar with the Mann-Kendall Test. The Mann-Kendall Test is a non-parametric trend test. A non-parametric test requires no assumptions about the distribution. The Mann-Kendall test for trend utilizes the following test statistic:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sign}(x_j - x_i).$$

This S is used to compute what is known as Kendall's tau statistic which is then tested for significance. The Seasonal Mann-Kendall test for trend accounts for seasonal trends by grouping the data by seasons and taking

$$S = \sum_{j=1}^{p} S_j,$$

where S_j is the Kendall statistic for season j and p is the number of seasons. Our seasons are grouped as months, a decision based off the Department of Natural Resources methodology [4]. Significance in the Seasonal Mann-Kendall Test denotes that there is a correlation between the values and time while controlling for seasonal variations of the value. The Theil-Sen Estimator of slope $\hat{\beta}$ is defined as

$$\hat{\beta} = \text{median}_{i \neq j} \left(\frac{x_j - x_i}{t_j - t_i} \right).$$

Note that if the mean is used instead of the median then $\hat{\beta}$ is the estimator for the least squares technique.

3   Results

3.1   Stations' Statuses

A table of our results from the Wilcoxon Signed-Rank Test can be found in Table 3.1. Each row represents a station. They are listed in alphabetical order. The columns display the stations' statues as assigned by the Wilcoxon Test for dissolved oxygen concentration at both the 3mg/l and 5mg/l threshold, turbidity, chlorophyll concentration, and pH.

According to the Wilcoxon Signed-Rank Test we ran on the continuous monitoring stations' data from the summer of 2011, all stations are classified as exhibiting turbidity levels significantly above the threshold level, and all but the Bishopville station, Little Monie Creek station and the stations located at the bottom of Goose and Mansonville received a 'good' status for dissolved oxygen using the 5mg/l threshold. When the more critical dissolved oxygen concentration threshold of 3mg/l is used, the Bishopville station's and Little Monie Creek station's status raised from 'bad' to 'good.' This suggests that the Bishopville station and Little Monie Creek station, while not as healthy as the rest of the bay, may not be in as critical condition as the two bottom stations whose statuses remain 'bad' regardless of which dissolved oxygen threshold is used. DNR mentions that low dissolved oxygen readings during summer months are most prominent in measurements taken at bottom stations due to decomposing algae that has sunk and the lack of mixing between surface and bottom waters [3]. These explanations support our conclusions that two of the bottom stations are faring the worst.

The statuses for turbidity,are largely 'bad' in that many stations fell above the benchmark level. On the otherhand, the station statuses for chlorophyll faired quite well. This indicates that the algee blooms may have been a less of a concern during the summer of 2011 than dissolved oxygen or turbidity. Unlike turbidity, the statuses in pH are mostly 'good.' The exceptions are the Mataponi and Little Monie Creek stations which both received classifications of 'bad.' This shows that most of the stations fell within the benchmark range for pH. Some possible reasons for the 'bad' status could be because of an algae bloom or low salinity. Overall, in terms of pH the stations performed well. As our future research investigates into the appropriateness of the current methodology, namely the use of the Wilcoxon Signed-Rank to determine stations' statuses, we are hesitant to draw strong conclusions regarding the heath of the bay based on current statuses. The results of the methodology exploration is discussed further in Section 3.2. Our status assignment results are also presented visually in Figures 3.1–3.5. The stations are marked as circles on the map of the Chesapeake Bay and surrounding area. Each parameter is presented on a separate map. Green circles represent stations that were assigned a status of 'good' and red circles indicate stations that were assigned a status of 'bad.'

3.2   Simulation Results

In this section we provide the results of the simualtion we ran to investigate the sensitivity of the Wilcoxon Signed-Rank to the non-symmetry in the data. This simualtion study is

Table 3.1: Wilcoxon Signed-Ranked test assignment of station statuses as "Good," "Borderline," or "Bad" of Dissolved oxygen (3mg), Dissolved oxygen (5mg), Chlorophyll, Turbidity, and pH.

| Station Name	DO5	DO3	Turbidity	Chlorophyll	pH
Annapolis	Good	Good	Bad	Good	Good
Betterton	Good	Good	Bad	Good	Good
Big Annemessex	Good	Good	Good	Good	Good
Bishopville	Bad	Good	Bad	Bad	Good
Budds Landing	Good	Good	Bad	Bad	Good
Chesapeake Y. Club	Good	Good	Bad	Good	Good
Corisca River	Good	Good	Bad	Border	Good
Downs Park	Good	Good	Bad	Good	Good
Flats	Good	Good	Bad	Good	Good
Fort Armistead	Good	Good	Bad	Good	Good
Fort Howard	Good	Good	Bad	Good	Good
Fort Smallwood	Good	Good	Bad	Good	Good
Goose (bottom)	Bad	Bad	Good	Good	Good
Goose (surface)	Good	Good	Good	Good	Good
Gratitude Marina	Good	Good	Bad	Good	Good
Greys Creek	Good	Good	Bad	Good	Good
Harness Creek (down)	Good	Good	Bad	Good	Good
Havre de Grace	Good	Good	Bad	Good	Good
Indian Head	Good	Good	Bad	Good	Good
Iron Plot Landing	Good	Good	Bad	Good	Good
Little Monie	Bad	Good	Good	Good	Good
Love Point	Good	Good	Bad	Good	Good
Manokin	Good	Good	Good	Good	Good
Masonville (bottom)	Bad	Bad	Bad	Good	Good
Masonville Cove	Good	Good	Bad	Good	Good
Mataponi	Good	Good	Bad	Good	Good
Mattawoman	Good	Good	Bad	Good	Good
McHenry	Good	Good	Good	Good	Good
Newport Creek	Good	Good	Bad	Good	Good
Otter Point Creek	Good	Good	Bad	Good	Good
Possum (bottom)	Good	Good	Bad	Good	Good
Possum Point	Good	Good	Bad	Good	Good
Public Landing	Good	Good	Bad	Good	Good
Railroad Bridge	Good	Good	Bad	Good	Good
Sandy Point South	Good	Good	Bad	Good	Good
Sill	Good	Good	Bad	Good	Good
Sill (bottom)	Good	Good	Bad	Good	Good
St. George's Creek	Good	Good	Good	Good	Good
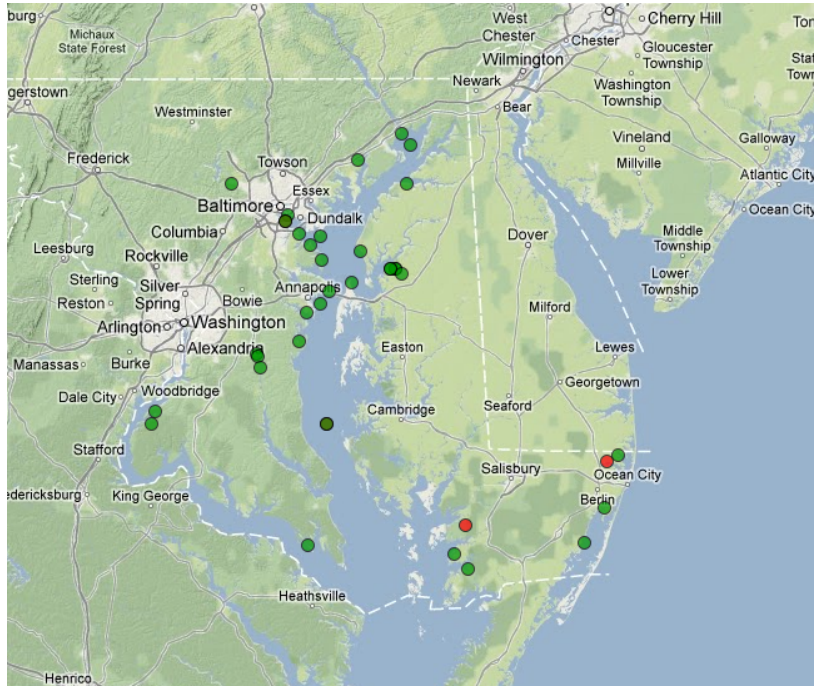
Figure 3.1: A map of the station statuses for dissolved oxygen (5mg/L).
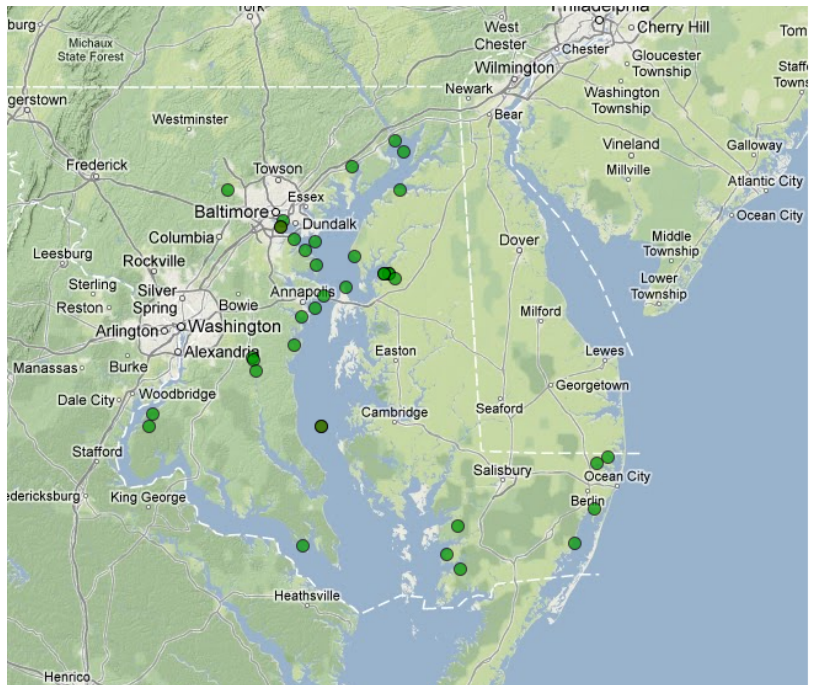

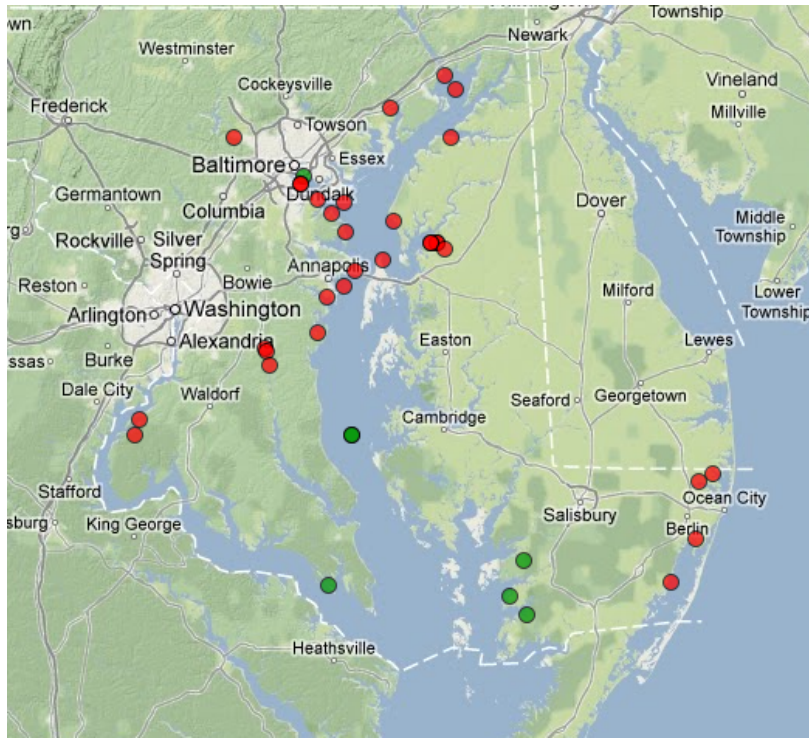Figure 3.2: A map of the station statuses for dissolved oxygen (3mg/L).

Figure 3.3: A map of the station statuses for turbidity (7 NTU).
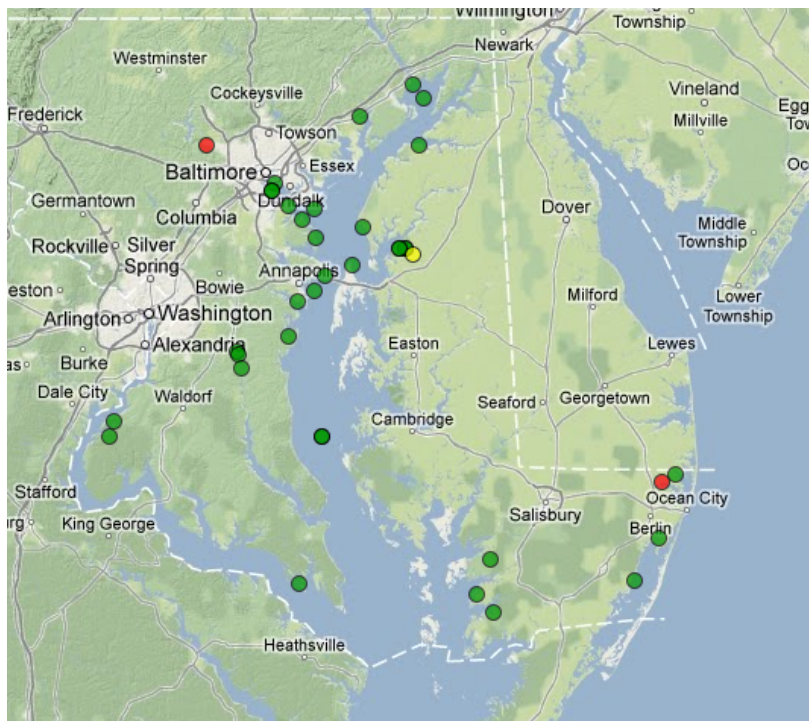
Figure 3.4: A map of the station statuses for chlorophyll (30μg/L).
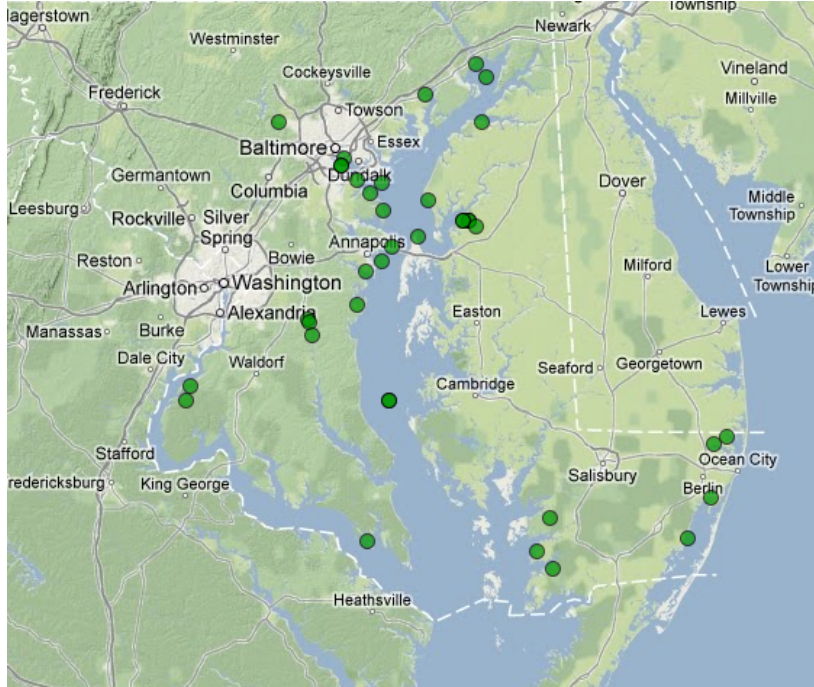
Figure 3.5: A map of the station statuses for pH (8.5/5.5).

described in 2.2. Tables 3.2 and 3.3 show the results of the simulation. In both tables, the columns represent the values used for the β_{rate} and the rows represent a set of the values used for the α_{shape} in the simulation. Tables 3.4 shows the average skew of each parameter (listed in the columns) at each station (listed in the rows alphabetically). Table 3.5 shows the skew of the $\Gamma(\alpha_{shape}, \beta_{rate})$ distribution with varying β_{rate} (listed in the columns) and α_{shape} (listed in the rows) values. From Table 3.5 we can see that as either the α_{shape} or β_{rate} parameter increases, the skew decreases, with the α_{shape} having a larger impact. By looking at Table 3.2 and Table 3.3 we see a similar trend with the Type I Error, that is it decreases as α_{shape} and β_{rate} increase. We conclude that as the skew decreases, the Type I Error rate obtained with the Wilcoxon Signed-Rank Test also decreases. Although we see similar behavior using data with and without the log-transformation, the inflation of the Type I Error is less severe with the transformation.

This gives strong evidence that the use of the Wilcoxon Signed-Rank test on the log-transformed data over the raw data is more appropriate for the highly-skewed distributions we see in the data. Comparing the skew values for the gamma distributions found in Table 3.5 to the skew values of our station data found in Table 3.4 we see that dissolved oxygen's and pH's positive skew values are in the range of the gamma distributions'. For turbidity and chlorophyll, we see even larger skew values than those covered with our gamma distribution. Even with the lower skew values, we see in Table 3.2 that the Wilcoxon Test sill yields a Type I Error rate much larger than our claimed significance level of 0.01 with the trend indicating the Type I Error will continue to inflate with larger skew values like those of turbidity and chlorophyll. Notice the simulation results indicate that the log-transformation does not does

Table 3.2: Type I Error of the Wilcoxon Test applied to samples drawn from the Gamma distribution with parameters α and β using a significance level of 0.01.

	β_{rate}	
	1	10
$\alpha_{shape} = 2$	0.8737	0.8692
$\alpha_{shape} = 4$	0.5054	0.5042
$\alpha_{shape} = 10$	0.1716	0.1701
$\alpha_{shape} = 50$	0.0304	0.0297
$\alpha_{shape} = 100$	0.0204	0.0205

Table 3.3: Type I Error of the Wilcoxon Test applied to the log-transform of samples drawn from the Gamma distribution with parameters α and β using a significance of 0.01.

	β_{rate}	
	1	10
$\alpha_{shape} = 2$	0.2183	0.2207
$\alpha_{shape} = 4$	0.1003	0.0977
$\alpha_{shape} = 10$	0.0407	0.0335
$\alpha_{shape} = 50$	0.0131	0.0145
$\alpha_{shape} = 100$	0.0116	0.0128

not significantly help reduce the Type I Error for the Wilcoxon Signed-Rank Test for data with this larger amount of skew.

3.3 Salinity Regimes

The Tukey Test results in Table 3.6 display the p-value for each of the comparisons between the regimes. Noticing that all the p-values are greater then 0.01, we fail to reject H_0 and conclude that none of the regimes are significantly different from each other. Because of this fact, the ranking methods are not applicable or significant to the salinity regimes.

3.4 Ranking

Tables 3.7 – 3.11 show the results of our ranking methodologies. The first column shows the percent failure of the given station. The rows are ordered by ascending percent failures. Each column shows the results of the respective ranking methodology.

The Tukey Test using the transformed percent failures was conservative, meaning it assigned ties even when station performances were dissimilar. The Bonferoni ranking method produced more conservative groupings than the Tukey Test. The Bayesian Ranking method is the only method that does not readily lend itself to groupings and as is shown in the

Table 3.4: Average skew values for each station's data for dissolved oxygen, chlorophyll, turbidity and pH. Stations are listed alphabetically.

Station Name	DO	Turbidity	Chlorophyll	pH
Annapolis	0.8600	3.5520	5.1955	0.1123
Betterton	0.7221	3.9284	4.0422	0.6824
Big Annemessex	-0.0837	4.7184	2.4859	0.2181
Bishopville	0.6247	1.5330	3.5572	0.5359
Budds Landing	0.0033	22.896	0.8347	-0.8338
Chesapeake Y. Club	0.7433	2.0293	2.6650	0.5736
Corisca River	0.6647	3.6492	1.8978	-0.0347
Downs Park	-0.4362	3.7650	1.6356	0.4968
Flats	0.4660	5.5261	1.7057	-0.1534
Fort Armistead	0.4411	3.2652	2.6704	0.2692
Fort Howard	0.0265	3.5911	1.6136	0.3341
Fort Smallwood	-0.0350	7.0761	2.1032	0.3495
Goose (bottom)	0.9488	23.037	2.0776	0.5552
Goose (surface)	0.9363	3.6947	2.5232	-0.1927
Gratitude Marina	0.3992	3.6118	1.7231	0.7324
Greys Creek	1.0849	5.4151	1.7777	-0.1125
Harness Creek (down)	0.3694	3.3329	1.6014	-0.0118
Havre de Grace	0.5920	5.3488	1.6738	0.9268
Indian Head	0.0680	4.1753	1.1078	0.0945
Iron Plot Landing	0.6860	3.7273	8.3336	0.01653
Little Monie	0.2843	2.2775	13.122	0.5784
Love Point	0.0870	3.6370	0.9614	0.1996
Manokin	0.0360	5.1246	4.1258	-0.3016
Masonville (bottom)	0.8714	10.021	4.9516	1.4185
Masonville Cove	0.6011	5.5490	1.4686	0.3826
Mataponi	0.3116	9.2984	5.4602	0.5960
Mattawoman	0.3659	9.5283	0.3484	-0.4789
McHenry	0.4703	5.2170	3.2518	0.5557
Newport Creek	0.1511	4.5250	0.5743	-0.5666
Otter Point Creek	0.0653	4.4148	0.7785	0.1441
Possum (bottom)	0.4976	3.7311	1.9859	0.5159
Possum Point	0.8535	9.6415	1.5337	0.1973
Public Landing	0.0054	4.9307	0.5873	0.0193
Railroad Bridge	0.8822	7.9900	12.955	0.6746
Sandy Point South	0.2204	4.7515	1.1912	0.7569
Sill	0.8892	1.7235	2.3680	0.5063
Sill (bottom)	0.4632	1.1784	2.2671	0.8180
St. George's Creek	-0.2798	3.4172	2.8722	0.0064

Table 3.5: Skew values for a Gamma distributions with parameters α_{shape} and β_{rate}.

| | β_{rate} | |
	1	10
$\alpha_{shape} = 2$	1.4431	1.2808
$\alpha_{shape} = 4$	0.8033	0.8793
$\alpha_{shape} = 10$	0.6059	0.6356
$\alpha_{shape} = 50$	0.2725	0.3453
$\alpha_{shape} = 100$	0.0633	0.2306

Table 3.6: Comparing stations by salinity regimeTF = Tidal Fresh (0-0.5 ppt), OH = Oligo-haline (0.5-5 ppt), MH = Mesohaline (5-18 ppt) and PH = Polyhaline (18-30 ppt).

p-value	Oxygen 3	Oxygen 5	Turbidity	Chlorophyll	pH
OH-MH	0.8528	0.6444	0.4824	0.9561	0.1745
TF-MH	0.7859	0.7499	0.9999	0.3261	0.7562
TF-OH	0.9992	0.9742	0.6079	0.4327	0.0818

ranking tables it is not reliable. The problem stems from the simplicity of the method in where the posterior probability distribution comes from a normal distribution using the percent failure as the mean and the standard error from the sample, but since our data is skewed, the assumption is violated. Therefore, the Baysian Ranking method was not suitable to accurately rank the data.

This led us to conclude that the Benjamini-Hochberg method, which falls between the overly conservative Bonferoni and Tukey methods and the unreliable Bayesian Method, provides the best ranking given our data and problem. Furthermore, we can see that the Benjamini-Hochberg method using the two-proportions Z-test gives a better ranking for the percent fail values than the Benjamini-Hochberg method using the mean values. In fact, the Benjamini-Hochberg method using the mean values seems to be just as unreliable as the Bayesian method; this is because the means are impacted by the skews. In some of the tables, namely chlorophyll, turbidity, and pH, we see a similar pattern between the Benjamini-Hochberg using the means and Bayesian Method which reaffirms our previous suspicions about the skewness causing problems. Also like the Bayesian Method, the Benjamini-Hochberg using the means is not very stable, has massive jumps in rankings, and is not suitable to rank the stations.

Table 3.7: **Oxygen (5mg)** — Ranking of continuous monitoring stations (with 1 being the best) with respect to its percent failure, the Tukey Test, the Bonferroni Test, Benjamini Hochberg Method, and the Bayesian Simple Ranking Method, respectively.

Station Name	% Fail	Tukey	Bonferroni	Benjamini		Bayesian
				% Fail	Mean	
Betterton	0%	1	1	1	4	16
Havre de Grace	0%	1	1	1	5	17
Flats	0.01%	1	1	3	2	18
Goose (surface)	0.04%	1	1	3	3	15
Big Annemessex	1.34%	5	5	5	16	19
AnnapolisCIBS	2.23%	6	6	6	6	12
Manokin	2.62%	6	6	6	9	14
Sill	4.24%	8	8	8	9	10
Budds Landing	5.06%	8	8	8	1	7
Love Point	5.79%	10	10	10	8	2
Iron Plot Head	5.91%	10	10	10	23	13
Fort Howard	6.31%	10	10	10	7	11
St. George's Creek	7.62%	13	13	13	20	9
Sandy Point South	8.14%	14	13	14	15	4
Possum Point	11.98%	15	15	15	9	20
Gratitude Marina	12.60%	15	15	15	22	21
Fort Smallwood	15.09%	17	17	17	9	8
Fort Armistead	16.67%	17	17	17	14	6
Public Landing	18.58%	19	19	19	27	24
Railroad Bridge Crossing	19.50%	20	20	20	18	23
Downs Park	19.83%	20	20	20	16	1
Mattawoman	21.74%	22	22	22	25	5
Otter Point Creek	22.61%	23	23	23	9	22
Indian Head	23.73%	23	23	23	18	3
Masonville Cove	28.07%	25	25	25	20	27
Harness Creek	29.66%	25	25	25	23	26
Sill (bottom)	30.51%	25	25	25	29	25
Corisca River	37.03%	28	28	28	26	29
Mataponi	41.50%	29	29	29	33	30
Newport Creek	41.77%	29	29	29	32	32
Chesapeake Yacht Club	42.37%	29	29	29	29	28
McHenry	43.05%	39	29	29	29	31
Greys Creek	44.95%	33	33	33	27	33
Possum (bottom)	51.50%	34	34	34	34	34
Bishopville	55.78%	35	35	35	35	35
Little Monie	80.21%	36	36	36	36	37
Masonville (bottom)	80.41%	36	36	36	37	36
Goose (bottom)	89.75%	38	38	38	38	38

Table 3.8: **Oxygen (3mg)** — Ranking of continuous monitoring stations (with 1 being the best) with respect to its percent failure, the Tukey Test, the Bonferroni Test, Benjamini Hochberg Method, and the Bayesian Simple Ranking Method, respectively.

Station Name	% Fail	Tukey	Bonferroni	Benjamini		Bayesian
				% Fail	Mean	
Betterton	0%	1	1	1	4	19
Big Annemessex	0%	1	1	1	16	21
Flats	0%	1	1	1	2	22
Goose (surface)	0%	1	1	1	3	23
Havre de Grace	0%	1	1	1	5	24
Iron Plot Head	0%	1	1	1	23	25
Manokin	0%	1	1	1	9	26
Possum Point	0%	1	1	1	9	27
Railroad Bridge Crossing	0%	1	1	1	18	28
Sill	0.06%	1	1	10	9	29
AnnapolisCIBS	0.10%	1	1	10	6	20
Budds Landing	0.14%	12	1	10	1	17
Gratitude Marina	0.17%	12	1	10	22	30
St. George's Creek	0.31%	14	1	14	20	18
Public Landing	0.37%	14	1	15	27	16
Fort Armistead	0.58%	16	16	16	14	15
Indian Head	0.78%	17	17	17	18	13
Fort Howard	0.86%	18	17	17	7	14
Sandy Point South	1.05%	18	19	19	15	8
Love Point	1.09%	18	19	19	8	7
Fort Smallwood	2.10%	21	21	21	9	9
Mattawoman	3.09%	21	21	22	25	4
Sill (bottom)	3.30%	22	23	22	29	5
Possum (bottom)	5.26%	24	24	24	34	1
Otter Point Creek	5.54%	24	24	24	9	12
Corisca River	8.09%	26	26	26	26	3
Masonville Cove	8.52%	26	26	26	20	31
Harness Creek	8.61%	26	26	26	23	11
Downs Park	8.91%	26	26	26	16	2
Newport Creek	10.22%	30	30	30	32	32
Chesapeake Yacht Club	10.73%	31	31	30	29	10
McHenry	13.33%	32	32	32	29	6
Greys Creek	18.04%	33	33	33	27	34
Mataponi	19.06%	33	33	33	33	33
Bishopville	26.41%	35	35	35	35	35
Little Monie	27.86%	35	35	35	36	36
Masonville (bottom)	54.87%	37	37	37	37	37
Goose (bottom)	72.00%	38	38	38	38	38

Table 3.9: **Turbidity** — Ranking of continuous monitoring stations (with 1 being the best) with respect to its percent failure, the Tukey Test, the Bonferroni Test, Benjamini Hochberg Method, and the Bayesian Simple Ranking Method, respectively.

Station Name	% Fail	Tukey	Bonferroni	Benjamini		Bayesian
				% Fail	Mean	
Gooses (surface)	20.31%	1	1	1	3	3
Gooses (bottom)	23.65%	2	2	2	1	24
McHenry	27.01%	3	3	3	3	14
St George's Creek	32.30%	4	4	4	6	1
Manokin	35.40%	5	5	5	5	32
Little Monie	38.85%	6	6	6	1	31
Big Annemessex	42.28%	7	7	7	8	29
Flats	44.55%	7	7	8	30	37
AnnapolisCBIBS	45.45%	9	9	8	13	20
Harness Creek	52.95%	10	10	10	8	23
Fort Smallwood	53.29%	10	10	10	12	15
Havre de Grace	54.94%	10	10	10	30	5
Bishopville	56.94%	13	13	13	6	19
Love Point	57.14%	13	13	13	19	17
Indian Head	58.40%	15	15	15	10	18
Mattawoman	58.65%	15	15	15	11	26
Masonville Cove	59.77%	17	17	17	14	13
Newport Creek	68.24%	18	18	18	18	33
Grey's Creek	69.34%	18	18	18	16	35
Mataponi	73.33%	20	20	20	34	28
Downs Park	79.47%	21	21	21	22	4
Betterton	80.07%	21	21	21	30	36
Public Landing	80.67%	21	21	21	30	34
Fort Armistead	81.22%	21	21	21	16	22
Iron Pot Landing	82.06%	25	25	25	38	30
Otter Point Creek	86.52%	26	26	26	36	12
Sill	86.73%	26	26	26	14	8
Sandy Point South	87.85%	26	26	26	22	2
Possum Point	93.74%	29	29	29	21	10
Gratitude Marina	93.90%	29	29	29	27	21
Fort Howard	94.38%	29	29	29	28	16
Chesapeake Yacht Club	94.51%	29	29	29	20	9
Masonville (bottom)	98.22%	33	33	33	37	25
Corsica River	98.46%	33	33	33	24	11
Railroad Bridge Crossing	98.56%	33	33	33	34	27
Budds Landing	98.65%	33	33	33	29	38
Sill (bottom)	98.76%	33	33	33	25	7
Possum (bottom)	99.89%	38	38	38	26	6

Table 3.10: **Chlorophyll** — Ranking of continuous monitoring stations (with 1 being the best) with respect to its percent failure, the Tukey Test, the Bonferroni Test, Benjamini Hochberg Method, and the Bayesian Simple Ranking Method, respectively.

Station Name	% Fail	Tukey	Bonferroni	Benjamini		Bayesian
				% Fail	Mean	
Big Annemessex	0%	1	1	1	2	9
Mattawoman	0%	1	1	2	9	13
Public Landing	00.01%	1	1	2	11	23
Havre de Grace	00.03%	1	1	2	5	10
Betterton	00.08%	5	1	5	7	21
Manokin	00.10%	6	1	5	3	7
Flats	00.15%	7	1	7	4	3
Indian Head	00.15%	7	1	7	9	12
Love Point	00.22%	9	1	9	12	16
Otter Point Creek	00.24%	9	10	9	15	26
Iron Pot Landing	00.59%	11	11	11	1	25
Little Monie	00.72%	11	12	11	12	8
Mataponi	01.37%	13	13	13	6	20
Sandy Point South	01.44%	13	13	13	19	27
Gooses (surface)	01.62%	13	13	13	14	2
Gooses (bottom)	02.12%	13	13	13	7	1
Railroad Bridge Crossing	02.33%	17	17	17	17	33
Gratitude Marina	03.97%	18	18	18	17	31
AnnapolisCBIBS	04.15%	18	18	18	19	6
McHenry	06.41%	20	20	20	15	4
Fort Howard	08.20%	21	21	21	23	32
St. George's Creek	08.28%	21	21	21	21	5
Downs Park	09.48%	21	21	21	22	22
Masonville (bottom)	10.28%	24	24	24	26	34
Masonville Cove	10.66%	24	24	24	24	18
Fort Smallwood	13.28%	26	26	26	24	15
Sill (bottom)	17.92%	27	27	27	27	36
Fort Armistead	18.48%	27	27	27	27	24
Chesapeake Yacht Club	20.63%	29	29	29	30	30
Harness Creek	20.96%	30	30	29	29	14
Sill	22.61%	30	30	29	30	28
Possum (bottom)	26.45%	32	32	32	30	38
Possum Point	38.02%	33	33	33	33	29
Grey's Creek	40.99%	34	34	34	35	19
Corsica River	42.99%	34	34	34	37	35
Newport Creek	46.31%	36	36	36	33	17
Budds Landing	50.19%	37	37	37	36	37
Bishopville	77.97%	38	38	38	38	11

Table 3.11: **pH** — Ranking of continuous monitoring stations (with 1 being the best) with respect to its percent failure, the Tukey Test, the Bonferroni Test, Benjamini Hochberg Method, and the Bayesian Simple Ranking Method, respectively.

Station Name	% Fail	Tukey	Bonferroni	Benjamini		Bayesian
				% Fail	Mean	
Big Annemessex	0%	1	1	1	18	29
Iron Pot Landing	0%	1	1	1	4	30
Little Monie	0%	1	1	1	2	31
Manokin	0%	1	1	1	14	32
Mataponi	0%	1	1	1	1	28
Newport Creek	0%	1	1	1	7	33
Public Landing	0%	1	1	1	12	34
Railroad Bridge Crossing	00.05%	1	1	8	3	27
Gooses (bottom)	00.41%	9	9	9	9	24
Grey's Creek	00.43%	9	10	9	8	35
Masonville (bottom)	02.63%	11	11	11	6	25
Bishopville	04.62%	12	12	12	10	19
Havre de Grace	05.84%	13	13	13	11	5
Indian Head	07.99%	14	14	14	5	18
Betterton	08.15%	14	14	14	12	36
Gratitude Marina	08.39%	14	14	14	20	21
Love Point	09.44%	17	17	17	26	17
Sandy Point South	09.64%	18	17	18	21	2
Mattawoman	13.53%	19	19	19	14	26
McHenry	13.88%	19	19	19	17	14
AnnapolisCBIBS	14.45%	19	19	19	28	20
Possum (bottom)	17.49%	22	22	22	19	6
Sill (bottom)	18.39%	22	22	22	24	7
Downs Park	19.40%	22	22	24	26	4
Fort Armistead	20.26%	25	25	24	29	22
Masonville Cove	20.61%	26	26	26	21	13
Corsica River	20.92%	26	26	26	24	11
Sill	24.92%	28	28	28	30	8
Fort Smallwood	25.13%	28	28	28	32	15
Chesapeake Yacht Club	26.21%	28	28	28	21	9
Harness Creek	26.36%	28	28	28	30	23
Gooses (surface)	26.82%	28	28	28	35	3
St George's Creek	28.26%	33	33	33	35	1
Fort Howard	28.37%	33	33	33	32	16
Possum Point	30.89%	35	35	35	32	10
Otter Point Creek	31.07%	35	36	35	16	12
Flats	48.94%	37	37	37	38	37
Budds Landing	59.38%	38	38	38	37	38

3.5 Trend Analysis

Table 3.12 shows the results of the trend analysis tests on the continuous monitoring stations. The stations are listed in alphabetic order. Each column denotes a different parameter and each row a different station. As expected, most of the results show a significant difference as many of the parameters are affected by the change of temperature and rainfall which would be expected to vary with change of seasons. However, there are a few stations which stand out as not showing such a trend. Bishopville on oxygen is one example. This is particularly a problem given that Bishopville is one of the only failing stations and this result indicates the the area Bishopville is monitoring is not expected to improve after the summer. Other stations showing no trends are Downs Park, Flats, Fort Armistead, and Love Point, and Sill. However, since these stations are not failing in the areas where no trend is seen there is no concern for these outcomes.

Using data from 2005 to 2011, we conduct trend analysis on the long-term monitoring stations in the Corsica River. The long-term monitoring stations includes: XHH3851, XHH4528, XHH4742, XHH4916, XHH4931. In Table 3.13we see a table of trend analysis results conducted at station XHH3951 for chlorophyll. The first column indicates the name of the station, the second column indicates the Seasonal Mann-Kendall P-value, and the third column indicates the Sen's estimate value. Each row corresponds to a particular station. If the Seasonal Mann-Kendall p-value is less than 0.01 for a particular station, then there is strong evidence a trend exists at a rate indicated in the Sen's Estimate. If the entire first column has p-values less than or close to 0.01, then it is likely the entire Corsica River exhibits a trend in that parameter.

Overall, there is evidence of a decreasing chlorophyll trend in the Corsica River except for XHH4528. Here, the Sen's test indicates a positive trend, most likely the result of algae blooms. There also appears to be an overall decreasing trend in salinity at all five stations. All five stations are very similar in that they increase slightly between 2005 and 2008, then decrease at a much greater rate between 2008 and 2011. Conductivity performs the same way — at all stations, conductivity readings increase from 2005 to 2008, then decrease much more drastically between 2008 and 2011. Water's ability to conduct electricity depends on the amount of dissolved salt in the water, thus the similarity in trends between salinity and conductivity is not surprising. Total nitrogen is another nutrient that shows an overall decrease, most likely because nitrate nitrogen as N, nitrite nitrite as N, and particulate nitrogen show a decreasing trend. Tables 3.13 through 3.43 can be consulted for p-values and estimated slopes for all stations and parameters.

Table 3.12: p-values of the Seasonal Mann-Kendall test. The Seasonal Mann-Kendall Test tests for the existence of a trend. A low p-value indicates a high chance of a trend being present in the data.

Station Name	DO	Chlorophyll	Turbidity	pH
Annapolis	<0.0001	<0.0001	<0.0001	<0.0001
Betterton	<0.0001	<0.0001	<0.0001	0.0342
Big Annemessex	<0.0001	<0.0001	<0.0001	<0.0001
Bishopville	0.0196	<0.0001	<0.0001	<0.0001
Budds Landing	<0.0001	<0.0001	.6181	<0.0001
Chesapeake Y. Club	<0.0001	<0.0001	<0.0001	<0.0001
Corisca River	<0.0001	<0.0001	<0.0001	<0.0001
Downs Park	0.4461	<0.0001	<0.0001	<0.0001
Flats	0.1387	<0.0001	<0.0001	<0.0001
Fort Armistead	0.0056	.8869	.4478	<0.0001
Fort Howard	<0.0001	<0.0001	<0.0001	<0.0001
Fort Smallwood	<0.0001	<0.0001	<0.0001	<0.0001
Goose (bottom)	<0.0001	<0.0001	<0.0001	<0.0001
Goose (surface)	<0.0001	<0.0001	<0.0001	<0.0001
Gratitude Marina	0.0785	<0.0001	<0.0001	<0.0001
Greys Creek	<0.0001	<0.0001	<0.0001	<0.0001
Harness Creek	<0.0001	<0.0001	<0.0001	<0.0001
Havre de Grace	<0.0001	<0.0001	<0.0001	<0.0001
Indian Head	<0.0001	<0.0001	<0.0001	<0.0001
Iron Plot Landing	<0.0001	<0.0001	<0.0001	<0.0001
Little Monie	<0.0001	<0.0001	<0.0001	<0.0001
Love Point	0.0206	<0.0001	<0.0001	<0.0001
Manokin	0.00042	<0.0001	<0.0001	<0.0001
Masonville (bottom)	<0.0001	<0.0001	<0.0001	0.0002
Masonville Cove	<0.0001	<0.0001	<0.0001	<0.0001
Mataponi	<0.0001	0.0001	<0.0001	<0.0001
Mattawoman	<0.0001	<0.0001	<0.0001	<0.0001
McHenry	<0.0001	<0.0001	<0.0001	<0.0001
Newport Creek	<0.0001	<0.0001	<0.0001	<0.0001
Otter Point Creek	<0.0001	<0.0001	<0.0001	<0.0001
Possum (bottom)	<0.0001	<0.0001	<0.0001	<0.0001
Possum Point	<0.0001	<0.0001	<0.0001	<0.0001
Public Landing	<0.0001	<0.0001	.0013	0.0003
Railroad Bridge	<0.0001	<0.0001	<0.0001	<0.0001
Sandy Point South	0.0002	<0.0001	<0.0001	<0.0001
Sill	<0.0001	<0.0001	<0.0001	<0.0001
Sill (bottom)	<0.0001	<0.0001	.0453	<0.0001
St. George's Creek	0.0053	<0.0001	<0.0001	<0.0001

Table 3.13: Long-term monitoring station results for chlorophyll A.

Station	Seasonal Mann-Kendall P-Value	Sen's Estimate (μg/L)/Year
XHH3851	<.0001	-0.0386
XHH4528	<.0001	3.7316
XHH4742	<.0001	-1.4967
XHH4916	<.0001	-0.4689
XHH4931	<.0001	-1.1474

Table 3.14: Long-term monitoring station results for dissolved inorganic nitrogen.

Station	Seasonal Mann-Kendall P-Value	Sen's Estimate (mg/L)/Year
XHH3851	<.0001	-0.0018
XHH4528	<.0001	-0.0043
XHH4742	0.05195	0.0009
XHH4916	<.0001	0.0003
XHH4931	<.0001	0.0003

Table 3.15: Long-term monitoring station results for dissolved organic carbon.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	0.2333	-0.0518
XHH4528	<0.0001	-0.0087
XHH4742	0.0134	-0.1150
XHH4916	0.9411	-0.0363
XHH4931	0.3417	-0.1359

Table 3.16: Long-term monitoring station results for dissolved oxygen.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	<0.0001	-0.0607
XHH4528	<0.0001	0.1037
XHH4742	0.4379	-0.2571
XHH4916	0.1486	<0.0001
XHH4931	0.0067	<0.0001

Table 3.17: Long-term monitoring station results for dissolved organic nitrogen.

Station	Seasonal Mann-Kendall P-Value	Sen's Estimate (mg/L)/Year
XHH3851	0.5227	-0.0035
XHH4528	<0.0001	-0.0044
XHH4742	0.1082	-0.0099
XHH4916	0.0864	-0.0038
XHH4931	0.0443	-0.0016

Table 3.18: Long-term monitoring station results for dissolved organic phosphorus.

Station	Seasonal Mann-Kendall P-Value	Sen's Estimate (mg/L)/Year
XHH3851	<0.0001	-0.0014
XHH4528	<0.0001	-0.0008
XHH4742	0.3877	-0.0016
XHH4916	<.0001	-0.0014
XHH4931	0.0772	-0.0014

Table 3.19: Long-term monitoring station results for light attenuation.

Station	Seasonal Mann-Kendall P-Value	Sen's Estimate (1/m)/Year
XHH3851	0.8149	0.0849
XHH4528	<.0001	0.1121
XHH4742	–	–
XHH4916	0.0560	0.0522
XHH4931	<0.0001	0.0943

Table 3.20: Long-term monitoring station results for ammonium nitrogen as N (filtered).

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	0.0423	<0.0001
XHH4528	<0.0001	-0.0015
XHH4742	0.3023	<0.0001
XHH4916	0.0042	-0.0007
XHH4931	0.4625	0.0006

Table 3.21: Long-term monitoring station results for nitrite nitrogen as N (filtered).

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	0.0184	<0.0001
XHH4528	<0.0001	-0.0002
XHH4742	0.1574	<0.0001
XHH4916	0.0002	<0.0001
XHH4931	<0.0001	<0.0001

Table 3.22: Long-term monitoring station results for nitrate nitrogen as N (filtered).

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	<0.0001	-0.0002
XHH4528	<0.0001	-0.0017
XHH4742	0.0181	0.0001
XHH4916	<0.0001	0.0002
XHH4931	<0.0001	0.0003

Table 3.23: Long-term monitoring station results for nitrite+nitrite as N (filtered).

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	<0.0001	-0.0001
XHH4528	<0.0001	-0.0024
XHH4742	0.0225	-0.0001
XHH4916	<0.0001	<0.0001
XHH4931	<0.0001	0.0002

Table 3.24: Long-term monitoring station results for particulate carbon.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	<0.0001	-0.0881
XHH4528	<0.0001	0.0858
XHH4742	<0.0001	-0.2880
XHH4916	<0.0001	-0.0721
XHH4931	<0.0001	-0.1042

Table 3.25: Long-term monitoring station results for Pheophytin.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (μg/L)/Year
XHH3851	0.0003	0.0217
XHH4528	<0.0001	0.5670
XHH4742	0.0746	0.4832
XHH4916	0.2675	0.4570
XHH4931	0.5967	0.6677

Table 3.26: Long-term monitoring station results for pH.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (SU)/Year
XHH3851	0.0828	-0.0079
XHH4528	<0.0001	-0.0280
XHH4742	0.2220	-0.1000
XHH4916	0.2735	-0.0307
XHH4931	0.0147	-0.0383

Table 3.27: Long-term monitoring station results for particulate inorganic phosphorus.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	0.9851	0.0028
XHH4528	<0.0001	0.0003
XHH4742	–	–
XHH4916	0.5161	0.0010
XHH4931	0.0020	0.0008

Table 3.28: Long-term monitoring station results for particulate nitrogen.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	<0.0001	-0.0165
XHH4528	<0.0001	0.0030
XHH4742	<0.0001	-0.0483
XHH4916	<0.0001	-0.0153
XHH4931	<0.0001	-0.0290

Table 3.29: Long-term monitoring station results for orthophsophate phosphorus as P (filtered).

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	0.4071	<0.0003
XHH4528	<0.0001	-0.0002
XHH4742	0.9602	-0.0002
XHH4916	<0.0001	<0.0002
XHH4931	0.0005	<0.0004

Table 3.30: Long-term monitoring station results for particulate phosphorus.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	0.0292	0.0024
XHH4528	<0.0001	0.0014
XHH4742	<0.0001	-0.0027
XHH4916	<0.0001	0.0015
XHH4931	<0.0001	0.0009

Table 3.31: Long-term monitoring station results for salinity.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (PPT)/Year
XHH3851	<0.0001	-0.2477
XHH4528	<0.0001	-0.3727
XHH4742	0.0002	-0.4163
XHH4916	<0.0001	-0.4166
XHH4931	<0.0001	-0.4431

Table 3.32: Long-term monitoring station results for Secchi.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (m)/Year
XHH3851	0.8049	<0.0001
XHH4528	<0.0001	<0.0001
XHH4742	<0.0001	<0.0001
XHH4916	<0.0001	<0.0001
XHH4931	<0.0001	<0.0001

Table 3.33: Long-term monitoring station results for silica as SI (filtered).

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	0.7106	-0.1396
XHH4528	<0.0001	-0.6650
XHH4742	0.0556	-0.5200
XHH4916	0.3206	-0.7771
XHH4931	0.0421	-1.2841

Table 3.34: Long-term monitoring station results for water density.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (/Year)
XHH3851	0.8232	-0.1832
XHH4528	<0.0001	-0.3622
XHH4742	0.0468	-0.3547
XHH4916	0.0013	-0.4367
XHH4931	0.0010	-0.3784

Table 3.35: Long-term monitoring station results for conductivity.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (UMHOS/CM)/Year
XHH3851	<0.0001	-407.3057
XHH4528	<0.0001	-600.00
XHH4742	0.0002	-675.00
XHH4916	<0.0001	-663.1361
XHH4931	<0.0001	-720.6667

Table 3.36: Long-term monitoring station results for dissolved nitrogen.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	<0.0001	-0.0100
XHH4528	<0.0001	-0.02
XHH4742	0.0146	-0.015
XHH4916	<0.0001	<0.0001
XHH4931	<0.0001	<0.0001

Table 3.37: Long-term monitoring station results for total dissolved phosphorus.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	0.0039	-0.0009
XHH4528	<0.0001	-0.0008
XHH4742	0.9584	-0.0023
XHH4916	<0.0001	-0.0008
XHH4931	0.00017	-0.0020

Table 3.38: Long-term monitoring station results for total nitrogen.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	<0.0001	-0.0049
XHH4528	<0.0001	-0.0190
XHH4742	<0.0001	-0.0770
XHH4916	<0.0001	-0.0163
XHH4931	<0.0001	-0.0307

Table 3.39: Long-term monitoring station results for total organic nitrogen.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	<0.0001	-0.0146
XHH4528	<0.0001	-0.0121
XHH4742	<0.0001	-0.0761
XHH4916	<0.0001	-0.0147
XHH4931	<0.0001	-0.0325

Table 3.40: Long-term monitoring station results for total phosphorus.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	0.9581	-0.0007
XHH4528	<0.0001	-0.0003
XHH4742	0.0030	-0.0072
XHH4916	0.1454	0.0004
XHH4931	0.6717	-0.0008

Table 3.41: Long-term monitoring station results for total suspended solids.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	0.0731	-0.0318
XHH4528	<0.0001	0.3900
XHH4742	0.0013	<0.0001
XHH4916	0.1111	1.1837
XHH4931	0.0253	1.6727

Table 3.42: Long-term monitoring station results for volatile suspended solids.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (mg/L)/Year
XHH3851	<0.0001	<0.0001
XHH4528	<0.0001	<0.0001
XHH4742	–	–
XHH4916	<0.0001	-0.2015
XHH4931	<0.0001	-0.3067

3.6 Visualizations

Throughout the project we generated a variety of graphs that proved helpful in visualizing the data. To make such graphs easily replicable with this data as well as data from other years, we created a Graphical User Interface (GUI). This GUI reads in databases in the form of the Microsoft Access databases DNR currently uses for the monitoring stations data. One available plot to generate is the scatter plot of a single parameter at a single station over time. It also plots a horizontal red line indicating the threshold level we used for that parameter as well as printing out the percent of readings that fell on the unhealthy side of the threshold. An example of such a graph can be seen in Figure 3.6.

An example of the scatter plot matrix option can be seen in Image 3.7.

The scatterplot matrix provides a set of scatterplots, one for each parameter pairing for all parameters in a user-chosen station. These graphs allow users to visually see potential interactions between different parameters. If a plot from the scatterplot matrix is of partic-

Table 3.43: Long-term monitoring station results for water temperature.

Station	Seasonal Mann-Kendall P-value	Sen's Estimate (C°)/Year
XHH3851	<0.0001	-0.1325
XHH4528	<0.0001	0.3500
XHH4742	0.5987	<0.0001
XHH4916	0.0109	0.1266
XHH4931	0.0125	<0.0001
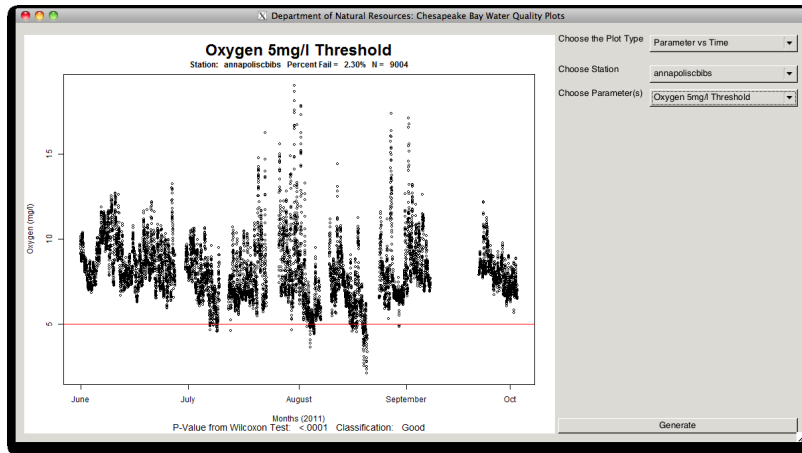
Figure 3.6: GUI Example Screenshot: Plot of oxygen readings from AnnapolisCIBS in 2011 from June through October.
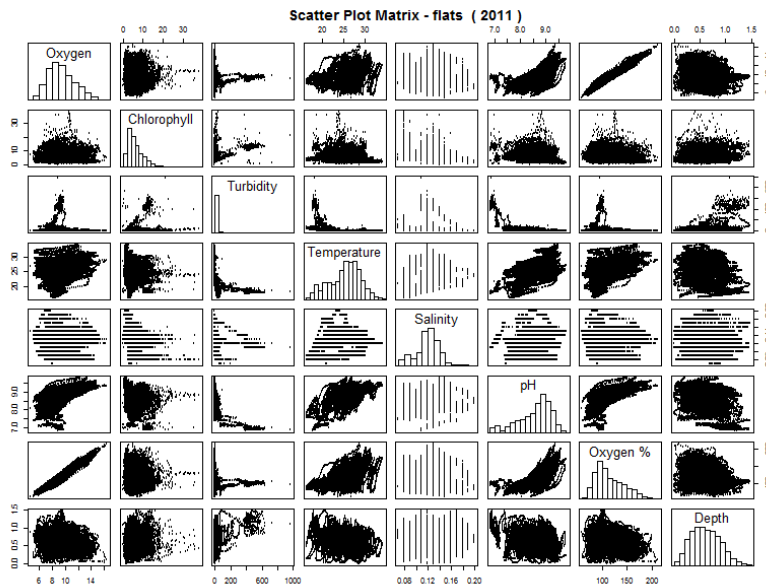
Figure 3.7: GUI Example Screenshot: Readings from Flat's in 2011 scatterplot matrix with 8 parameters.

ular interest, it can be enlarged by using the Scatter Plot option such as in Figure 3.8. It provides a plot between two user-chosen parameters at an individual station.

Other plot functions available in our GUI include two time series options. Two time series settings are available: stacked time series by parameter and stacked time series by month. An example of the time series by parameter can be seen in Figure 3.9. This allows users to observe all the parameters as they change over the full specified time frame. This may aid in noticing what time in the year each parameter is in most critical condition.

33

Oxygen vs Turbidity
Station Name:havredegrace (2011)
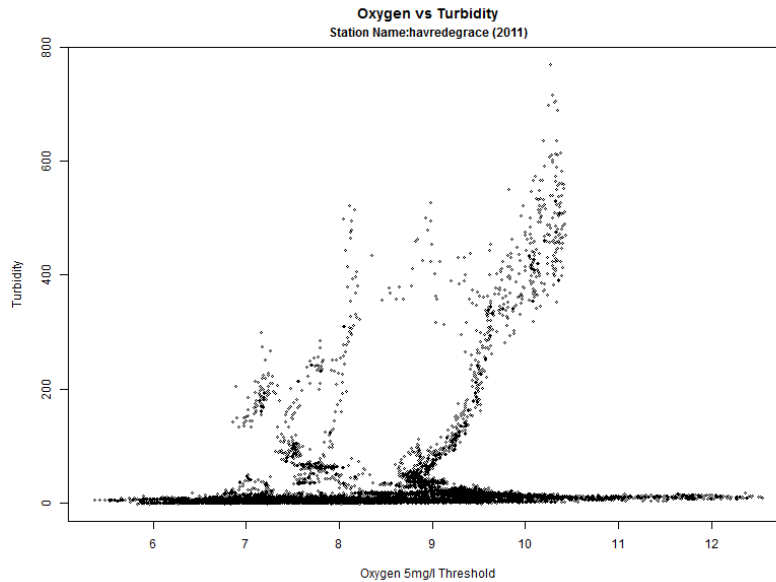
Figure 3.8: GUI Example Screenshot: Plot of two parameters, oxygen and turbidity, against each other from Havre de Grace in 2011.

The stacked time series by month focuses on a single, user chosen parameter. An example can be seen in Figure 3.10. It breaks the time series down by months. This allows for a closer look at how each parameter fluctuates over time.

4   Limitations of Project

It is important to remember that the results in this tech report only apply to the specific tests we conducted on the specific data set we used (shallow-water stations) and by no means applies to the Chesapeake Bay as a whole. There are many statistical tests one could perform to assess the water quality of the Chesapeake Bay, we chose to focus on a select number of them. If you would like to find out more about the different methods of assessing the Chesapeake Bay or a more general view of how Maryland's tidal waterways are doing, please visit the Eyes on the Bay website [3].

Acknowledgments
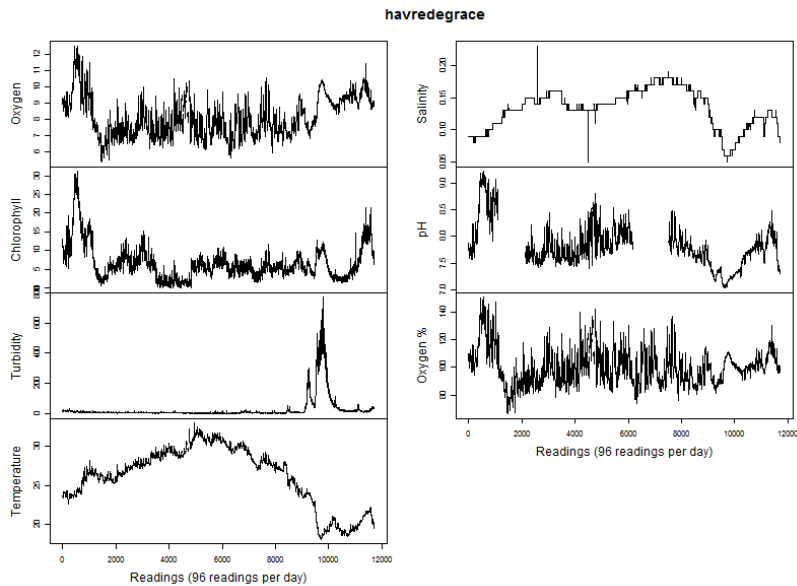
Figure 3.9: GUI Example Screenshot: Time series plots of data taken from Havre de Grace presented by parameter.
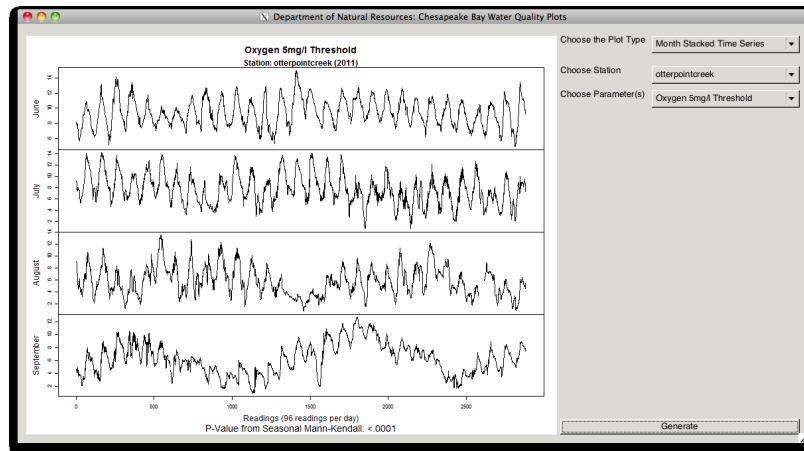
Figure 3.10: GUI Example Screenshot: Time series plots of data taken from Otter Point Creek presented by month.

References

[1] Samantha Allen, Dorthy Kirlew, Neil Obetz, Derek Wade, April Albertine, Nagaraj K. Neerchal, and Martin Klein. Assesment of simple and alternative bayesian ranking methods utilizing parallel computing. Technical Report HPCF–2011–11, UMBC High Performance Computing Facility, University of Maryland, Baltimore County, 2011.

[2] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, pp. 289–300, 1995.

[3] Eyes on the bay. `www.eyesonthebay.net`.

[4] Elizabeth Ebersole, Mike Lane, Marcia Olson, Elgin Perry, and Bill Romano. Assumptions and procedures for calculating water quality status and trends in tidal waters of the chesapeake bay and its tributaries: A cumulative history. Eyes on the Bay website, June 2002.

[5] The false discovery rate (fdr) procedure. NAEP Technical Documentation Website, November 2009.

[6] Brian R. Smith. Email correspondence, July 2012.

[7] Jerrold H. Zar. *Biostatistical Analysis*. Pearson Prentice Hall, fifth edition, 2010.

A   GUI Documentation

A.1   Overview and Installation

The GUI was developed using the R statistical package. However, there is no installation of R required in order to run this GUI, one must simply unpack the compressed folder and then there will be an executable file to run the specified GUI. The GUI has three modes:

- 2011 Mode: Runs the GUI using the premade plots for the summer of 2011

- MDB Mode: Runs the GUI using the Microsoft Access database provided by DNR's Eye's on the Bay site

- Batch Mode: Runs MDB Mode but instead of displaying a plot, it generates all of the plots from the given database and saves them to the GUIplots folder.

By driver limitations, MDB Mode and Batch Mode are only available for Windows machines. In order to run the specified GUI, simply click the executable file which corresponds to the mode you wish to use. These are respectively named "run2011.exe", "run.exe" (MDB Mode), and "runBatch.exe". On Mac and Linux distributions you will only see the respective "MacRun.command" and "LinuxRun.desktop" files giving the appropriate system the double-click execution functionality to run the 2011 Mode GUI.

A.2   Using The GUI

Once the GUI is opened you will see a standard GUI. If MDB or Batch mode is used, you will be first asked to select a Microsoft Access database. After this point, both GUIs will enter the plotting mode. For those using MDB or Batch mode, at this time enter the year to which the database corresponds, and then click refresh lists. This will set up the dates so that dates can be dynamically chosen (dates cannot be dynamically chosen in 2011 Mode). At this point, choose the desired station, the desired plot, and (if MDB or Batch Mode) the start and end dates. After these parameters have been set, click generate. The plot should generate and be placed in the window. Also, if in MDB or Batch mode, the location to which the plot is saved will be shown in the terminal window. These plots will all be saved in the "GUIplots" folder in the main directory of the program. For those using the 2011 Mode, note that all of the plots are saved within the "plots" folder of the main directory of the program.

A.3   Types of Plots

The GUI creates various types of plots depending on one's need. All of them can be accessed by using the "Plot Type" dropdown menu to select the plot type one would wish to have. Listed below is a brief description of each type:

- *Parameter vs Time*: Gives a simple plot of all of the data values on the y-axis vs time on the x-axis.

- *Parameter vs Time (means)*: Gives a plot of the means of the data values for the number of days set by the box titled "delimiter" vs time on the x-axis

- *Parameter vs Time (Overlayed)*: Gives the parameter vs time plot with a line overlayed showing the local regression line

- *Scatter Plot*: Gives a scatter plot which plots one parameter vs another parameter

- *Scatter Plot Matrix*: Gives a matrix of scatter plots whose diagonal shows the histogram of the given parameter.

- *Month Stacked Time Series*: Gives a time series plot of each month with the Seasonal Kendall p-value on the bottom to check for seasonality.

- *Parameter Stacked Time Series*: Gives a time series plot of all of the requested parameters for the station. This defaults to using all of the parameters the station measures.

A.4   Error Handling Behavior

The GUI is designed to be robust to errors. For example, the only dates that are given in the drop-down menu in MDB and 2011 mode are dates that the program is able to produce data. In Batch Mode, the dates that the user can select from represent all of the possible

choices given the set of all stations active in that year (in the dataset). Thus the dates act as a range to which the station will adjust to. For example, the start date chosen is the minimum start date, and thus for each station the GUI will find the closest available start date above (or equal to) that date. This is to allow users to select a large range of dates and then have the plots automatically adjust the range to the range where the station was active. Thus one may pick a range of dates such as 2/2–10/20 and if the station is only active from 5/3–9/30 the station will make the plots for 5/3–9/30. All such range adjustments are stated in the command line and thus also replicated in the command line's output file at "Output/BatchOutput.txt".

A.5 Plot Not Building In Batch Mode? Common Sources Of Error

If your plot is not building in Batch Mode, some common causes of error are:

1. The station may not be set to active in the database. Go into the database and check the stations table. Make sure active_####, where #### denotes the year, is set to active

2. Make sure that the name for the station is the same as the table name. A notable example of this error was StGeorges(Creek) in the 2009 database.

3. Make sure that the station has some datapoints in that timeslot. If the timeslot is too narrow, then there may just be no datapoints for that station in that time due to many various reasons. Check the dataset to make sure there are actually datapoints in there.

4. Make sure you chose the right database.

A.6 Default Behaviors

The default assumptions for the code are:

1. A significance alpha of .01 is used.

2. The parameter columns reference the same columns as the 2011 data.

3. It is not leap year.

4. The test values are the values given by the DNR in 2011.

To change any of these assumptions, the user just has to change the respective values in "src/assets/globals.R".