RESEARCH NEEDS FOR MARINE BEACHES FINAL REPORT

Contributing Authors: Dave McLaughlin, Sherry Poucher, Eva Touhey, Jillian Chopy and Sabrina Pereira

Federal Award Identification Number: CE00A00004

NBEP: Section 320 Funds

NEI Job Code: 318-001

Project Code: S-2018-005

2018/2019 Grant Program

June 13, 2019

This project work was funded by an agreement (CE00A0004) awarded by the Environmental Protection Agency to the New England Interstate Water Pollution Control Commission on behalf of the Narragansett Bay Estuary Program. Although the information in this document has been funded wholly or in part by the United States Environmental Protection Agency under agreement CE00A0004 to NEIWPCC, it has not undergone the Agency's publications review process and therefore, may not necessarily reflect the views of the Agency and no official endorsement should be inferred. The viewpoints expressed here do not necessarily represent those of the NBEP, NEIWPCC, or U.S. EPA nor does mention of trade names, commercial products, or causes constitute endorsement or recommendation for use.

THIS PAGE IS INTENTIONALLY LEFT BLANK.

EXECUTIVE SUMMARY

This research project was developed in response to needs identified in the 2017 State of Narragansett Bay and its Watershed Technical Report to improve our understanding of current marine beach water quality status and trends within the Narragansett Bay watershed. This technical report described conditions of beaches in terms of annual closures, and site-specific advisories dictated by pathogen risks. It recognized that there was a wealth of Fecal Indicator Bacteria data which could be expected to provide more resolution with respect to status and trends than closure data alone could provide.

The following report describes the work completed regarding two Rhode Island beaches through case study research to test new tools to enhance our ability to characterize, manage and find remedies to pathogen load problems. Specifically, the project applied state-of-the-science methodologies to evaluate data with respect to the United States Environmental Protection Agency's criteria for use of recreational waters and used advanced statistical modeling to better describe the environmental factors affecting the highly variable monitoring data. The most current approach to water quality monitoring relies on culturing the fecal-indicator bacteria, *Enterococcus*, but a delay of greater than 24 hours to perform the analysis and publish results often translates to poor protection of public health or unnecessary beach closures. Beach-specific predictive statistical models use a variety of environmental parameters to predict water quality variables in relation to state-specific recreational water quality standards. The models can be used not only to better understand root causes of contamination, but to initiate more timely and appropriate management actions to protect public health.

This project created beach-specific predictive models for two marine beaches in Rhode Island that have historically experienced high levels of *Enterococcus*. Utilizing bacteria data collected by the Rhode Island Department of Health and non-profit organization Clean Ocean Access from 2015 through 2018, along with over 50 environmental parameters, predictive models were developed using Multiple Linear Regression, Generalized Boosted and Partial Least Squares modeling techniques. All models were developed using the United States Environmental Protection Agency's Virtual Beach software.

Virtual Beach is a powerful modeling software package that uses tools to generate multiple transformations of explanatory variables, along with second order interaction terms to improve the model fit for prediction of Fecal Indicator Bacteria (Enterococcus) concentrations. Virtual Beach guidance follows a prescribed set of steps to incrementally increase the predictive capabilities of the model using state-of-the-science statistical approaches. Preferred models were selected by evaluating their benefits and limitations. The selection process limited false positives, while maintaining a good model fit with the small number of easily accessed environmental data. The variables found to be statistically significant within each model were also evaluated against various optimization criteria for plausibility and practicality. Chosen models were cross validated and inputted into Virtual Beach's prediction tool to determine how successful the 2015 – 2017 models were at predicting bacteria results from 2018. Ultimately, the models created for both case studies demonstrated the power of statistical modeling to determine the strongest relationships between environmental variables and the pathogen indicator.

The environmental variables found to be statistically significant predictors of *Enterococcus* were different for each beach. At Oakland beach, rain and tide variables along with the groundwater depth of the Warwick Neck were found to be statistically significant predictors, while the Easton's beach models more heavily emphasize weather variables like current, prior rain events and temperature. Operation of Newport's UV Disinfection Treatment Plant was consistently an important predictor of *Enterococcus* bacteria at Easton's beach as well. Although the generated models using these significant predictors demonstrated strong relationships between the predictors and the pathogen indicator, they were not successful at predicting 2018 data, potentially due to the lack of high *Enterococcus* exceedances in 2018. Individual case study reports were developed to support this summary.

After the completion of the case study modelling, the results were presented to officials working in both the City of Warwick and Newport. From these meetings, both parties were interested and excited about this project and water quality work. The mayor of Warwick approved using the Virtual Beach models on a trial basis, where RIDOH staff agreed to exercise the model for potential closure triggers in 2019 and follow-up with a review of the results to re-evaluate the utility of the model at the end of the season. In Newport, the information was received but another year of data validation was suggested before using it in a real-time setting.

TABLE OF CONTENTS

TABLES

FIGURES

BACKGROUND

The *2017 State of Narragansett Bay and Its Watershed Technical Report: Chapter 23* clearly identifies historic and continuing health risks at marine beaches throughout Rhode Island's Narragansett Bay. However, the assessment in the technical report was constrained by a limited scope for identification of temporal trends or comparison of health risks and the root causes of pathogenic waters across the spatial extent of the bay. The Research Needs for Marine Beaches project sought to test methods that would improve the assessments described in Chapter 23 of the technical report.

The overarching goal of the Research Needs for Marine Beaches project was to improve our understanding of the water quality of beaches situated within Narragansett by employing new tools and enhance our ability to manage and find remedies. Specifically, the project applied state-of-the-science methodologies to conduct case study research at two beaches located within the Narragansett Bay watershed that present health risks to recreators. Ultimately, this project sought to inform and improve management actions so that they may become more effective at protecting public health and minimizing unnecessary beach closures. Looking forward to 2022, the project will provide a stronger technical basis for status and trend evaluations for the next State of Narragansett Bay and Its Watershed Technical Report for beaches, with improved insight into the relationship of precipitation and other environmental variables in a predictive schema to protect human health and minimize unnecessary beach closures. While this research focused on predictive modeling at two specific locations (Oakland Beach, Warwick RI and Easton's Beach, Newport RI), the efforts aimed to create a project deliverable that is scalable and transferable to other beaches in the Narragansett Bay watershed; scalable in the sense of allowing more (or less) variables to be integrated into the predictive schema, and transferable in the sense of allowing beach managers (or their staff) to readily follow the Rhode Island-specific model constructs as a starting point that allows easier development, implementation, and usage of similar models in Rhode Island and Massachusetts. The current modeling efforts employed Virtual Beach (VB), a software package provided by the United States Environmental Protection Agency (USEPA) specifically to apply advanced statistical analysis to predict water quality from sets of potentially explanatory environmental variables.

Oakland Beach, in Warwick, RI was selected as one of two case study sites for predicting Fecal Indicator Bacteria (FIB, i.e. *Enterococcus*) using existing environmental variables, along with Easton's Beach in Newport, RI. Oakland Beach is a desirable case study site, given its history of serious water quality problems and beach closures, its iconic status as a historic recreation area in the state, and for its representation of urban beaches in upper Narragansett Bay. Oakland Beach is situated at the edge of Greenwich Bay, and as a part of the upper bay this coastline contrasts with the high energy environment of the other study site, Easton's Beach in Newport, RI. Both beaches have historic and continuing water quality problems with uncertain root causes, as sources and pathways are not sufficiently understood. Oakland Beach is located at the end of a peninsular sub-watershed to Greenwich Bay and Narragansett Bay. Here it is flanked by Brushneck Cove to the west and Warwick Cove to the east. Warwick Cove is a highly trafficked waterway with five marinas while recreational uses are more moderate in Brushneck Cove.

In addition to the potential contaminant sources from the Oakland Beach peninsula and the coves, currents and tidal exchange within Greenwich Bay to the south and west and Narragansett Bay to the east, present other potential contaminant inputs. Monitoring to characterize pathogen risks in the waters at Oakland Beach involves analysis for *Enterococcus*, an indicator of warm-blooded animal waste. Monitoring data are available for the years 2006-2018 through the Rhode Island Department of Health (RIDOH). The intermittent, but chronic impairment of Oakland Beach waters due to FIB contamination has been established through analysis of the existing data set relative to Rhode Island state water quality standards. Yet the real-time management to protect public health is limited by the 24-36-hour delay between sample collection and reportable laboratory results. The predictive modeling undertaken for this study addresses the priority to protect human health from pathogen exposures by providing a timelier estimate of *Enterococcus* concentrations, and, to better understand sources and transport processes of the pathogens.

Easton's Beach, in Newport, RI was selected as the other case study site for predicting FIB using existing environmental variables. Easton's Beach is a desirable case study site due to high closure rates from bacteria, high recreational use during the summer months (and year-round) and is one of the most-studied recreational waters in the Narragansett Bay watershed. Easton's Beach is a high energy environment with greater exchange rates than Oakland Beach. Like Oakland Beach, Easton's Beach has historic and continuing water quality issues with uncertain sources and pathways of pathogens. Newport, RI has a large resident and tourist population that use the beaches during the summer months, while over 3.5 million people visit the City of Newport on an annual basis of which tens of thousands of visitors recreate at Easton's Beach. Additionally, ocean enthusiasts continue to enjoy the recreational benefits of the shoreline year-round. This case study fits nicely with the on-going project "Stormwater Pathogens: Find It and Fix It" that works to determine if bacteria in sediment of the Easton's Bay watershed that drains to Easton's Beach poses a risk to water quality at the beaches. The "Find It and Fix It" project also studies possible remediation plans and improvements in conveyance systems. The Aquidneck Island watershed generates approximately 62 million gallons of stormwater per one inch of rainfall. This watershed area drains to Easton's Bay via the moat surrounding Easton's Pond. The moat is approximately eight feet wide and two feet deep. At a broader level, modeling of natural systems based on environmental inputs to form predictions sets the stage for decision makers to make more informed judgments with respect to remedial alternatives.

Protecting human health from exposure to bacteria at recreational beaches is of critical importance, as Rhode Island beaches are a main tourist attraction and undoubtably linked to the health of the state's economy. This project supports other initiatives that are working towards permanent year-round clean water for the State of Rhode Island that spans best management practice for stormwater and wastewater systems. These include projects that are mandated by consent decrees, supported by bonds, and grant funded efforts. The City of Newport and Town of Middletown have consent decrees with the EPA to address impaired waters and this project provides further insight into how the measures are resulting in the ability to predict healthy waters in the ocean and harbor.

While the project team had access to a large volume of bacteria data from 2000 to 2018, only the data collected during 2015 to 2018 were utilized in the modeling exercises. As expressed in the VB guidance document (Cyterski, 2016) and in USEPA's broader guidance for water quality modeling at beaches, the assumption of unchanging underlying sources and fate and transport processes is a critical element of the predictive model. Therefore, the project team performed the modeling exercises using the most recent data, collected from 2015 through 2018 by RIDOH and Clean Ocean Access (COA). This minimized the influence of prior remedial management actions, which could not be considered within the modeling construct.

EXPLORATORY ANALYSIS OF RAW DATA

The *2017 State of Narragansett Bay and Its Watershed Technical Report: Chapter 23* notes that the assessment was constrained by limited scope for identification of temporal trends or comparison of conditions across the spatial extent of the bay. One reason for this was the use of closure days rather than raw FIB data. The raw data provides the range of conditions at each beach, both within and across years, with individual data points ranging from the detection limit of 10 cfu/100 mL to greater than 10,000 cfu/100 mL. This range provides a greater scope for characterizing variability in condition in the short term and for potential long-term patterns that NBEP was unable to discern from beach closures alone.

The first assessment analyzed raw data from 2006 through 2018 to characterize conditions on an annual basis. While data collection began in 2000, E. coli was the monitored indicator bacteria until 2003. From 2003 to 2006, RIDOH monitored water for *Enterococcus* using the membrane filtration method. It was not until 2006 that the RIDOH state laboratory began to exclusively use the Enterolert method. Given the documented biases associated with these methods, it was most appropriate to limit the detailed statistical analysis to the years 2006 forward (ASTM, 2014).

One of the goals of Research Needs for Marine Beaches was to determine how often the case study beaches exceeded water quality standards for primary contact recreational uses. These standards, applied to the *Enterococcus* count data, included a geometric mean of 30 cfu/100 mL and a 10th percentile of 110 cfu/100 mL (EPA, 2012). Additionally, tests for upward or downward trends were conducted using the Mann-Kendall test (Meals et al., 2011). All of the above analyses were conducted using R software, including the packages “tidyverse” and “psych” (R Development Core Team, 2006).

Tables 1 through 3 summarize findings with respect to water quality standards. The tables report number of exceedances of each primary contact use standard. The water quality standards are to be applied on a monthly basis, so the number of months with exceedances per year are reported in the tables. Any year with a value greater than 0 fails to meet the criteria. For impaired waterbody status, no more than one in three years can have an exceedance. The findings support the assessment that both water bodies are impaired with respect to the designated uses (primarily contact recreation). In addition, the tables show that the waters at each beach have exceeded the Beach Action Value (BAV) of 60 cfu/100 mL many times during each year. The BAV serves as a more conservative measure than the two established Clean Water Act-based standards. In the tables, Single Sample Results greater than 60 cfu are simply the number of days with bacteria results greater than 60.

The actual monthly geometric means used to derive the number of monthly exceedances are provided in Appendix 4 for both beaches.

Table 1. Oakland Beach Center Station Enterococcus Exceedances 2006 - 2018

Year	Geomean > 30 cfu Exceedances	$10^{\%}$ > 110 cfu Exceedances	Single Sample Results > 60 cfu
2006	1	2	16
2007	1	2	6
2008	0	0	4
2009	0	1	6
2010	1	2	9
2011	0	2	5
2012	0	2	3
2013	2	2	11
2014	0	1	3
2015	3	3	9
2016	1	0	5

2017	2	3	10
2018	0	2	5

Table 2. Easton's Beach East Station Enterococcus Exceedances 2006 – 2018 (no East station data for 2010)

Year	Geomean > 30 cfu Exceedances	10% > 110 cfu Exceedances	Single Sample Results > 60 cfu
2006	3	3	20
2007	1	1	12
2008	3	3	14
2009	2	3	11
2011	2	2	15
2012	1	0	5
2013	1	1	6
2014	2	2	10
2015	2	3	12
2016	0	0	8
2017	0	1	5
2018	2	1	9

Trend analyses conducted using the Mann-Kendall test are summarized in Tables 3a and 3b below for Oakland Beach and Easton's Beach respectively. The Mann-Kendal statistic represents the proportion of up-movements against time versus the proportion of down-movements, looking at all possible pairwise time-differences. A Mann-Kendall statistic is formed by assigning a 1, -1, or 0 to each pair of data points depending on the sign (positive, negative, or equal) of the difference between the values. These values are then summed to form the Mann-Kendall statistic, which is compared to a critical value in a look-up table to determine if the null hypothesis is valid. Positive statistics, as for Oakland Beach indicate that the data would trend positive, and negative values suggest a downward trend. The p-values are the basis for determining the certainty in the relationship. Using $p \leq 0.05$ would typically support the finding of a trend with enough certainty to accept the "alternative hypothesis". Only one station, Oakland Beach East, met this criterion. The most highly used center stations at Oakland Beach and the eastern station at Easton's Beach exhibited no significant trend. Plots of the geometric mean data for each beach show the year to year variability that produced the Mann-Kendall results. These plots, as linear regressions, are shown on the graphs in Figures 1a and 1b for Oakland and Easton's beaches, respectively.

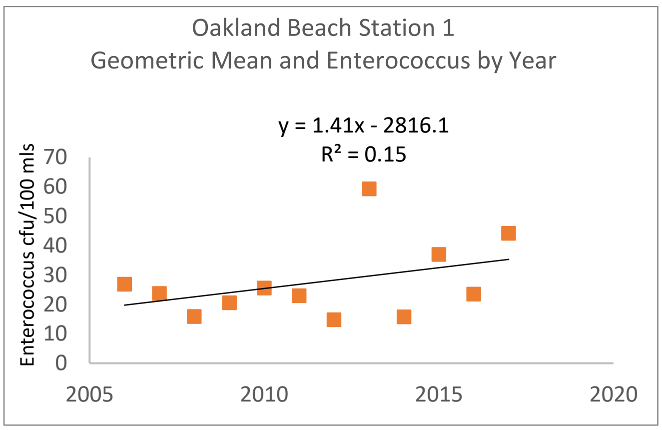
Tables 3a and 3b. Mann-Kendall Trend Analysis for three Oakland Beach and Easton's Beach Stations

3a.

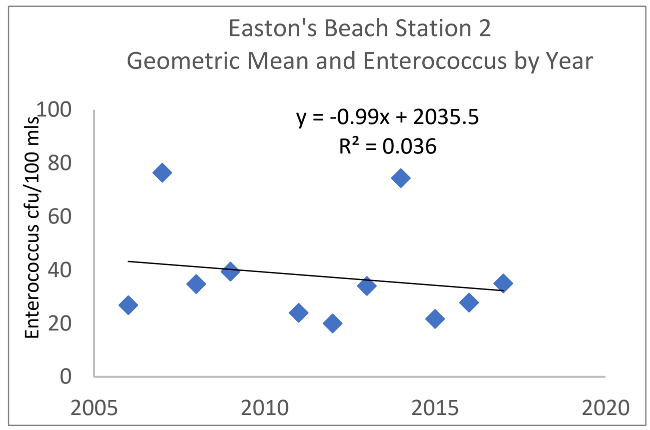
Oakland Beach Station	Mann – Kendal Statistic	P - value
Center	.480008	.631222
East	2.125749	.033524
West	1.028588	.303673

3b.

Easton's Beach Station	Mann – Kendal Statistic	P - value
Center	-.54908	.582951
East	-1.44002	.149861
West	-1.76926	.076851



1a.



1b.

Figures 1a and 1b. Geometric Means and Enterococcus Levels for Oakland and Easton's Beaches (2006 – 2018)

The exploratory analysis above provides context for the modeling work that was the central activity for the current project. Interannual variability poses questions with respect to the factors that influence that variability, and that is the intended function of the statistical modeling.

DATA ACQUISITION

The purpose and reasoning for the selection of the independent variables used in the Oakland Beach and Easton's Beach modeling case studies were derived from the EPA's guide *Six Key Steps for Developing and Using Predictive Tools at Your Beach* (2016). As stated in the data acquisition document for this project (Appendix 1), the research team used EPA's Six Step guide and prior knowledge of the beach's natural environment to determine which environmental parameters should be included as potential predictors of water quality. For both case studies, precipitation, wind, tide, air quality, water and air temperature and coastal survey data were gathered from existing sources and compiled into spreadsheets formatted to serve as input for the site-specific predictive models for each beach.

The Oakland Beach case study also utilized multiple variables that were specific to this beach and surrounding environment. A gage from the Hunt River, which is a tributary to Greenwich Bay, provided flow and river height data. In addition, a permanent well located on Warwick Neck, a peninsula directly to the east of Oakland Beach, provided groundwater depth data. The Easton's Beach case study also incorporated bacteria data from Easton's Point, Easton's Stream and the South East Esplanade, sampled by COA, into its modeling exercises. These variables were organized into a spreadsheet along with data from Newport's UV Disinfection Plant, as requested and provided by the City of Newport's Director of Utilities, Julia Forgue.

Quality-assurance and quality-control (QA/QC) procedures, as detailed in the project Quality Assurance Project Plan (QAPP) (COA, 2018) were implemented to facilitate the compilation of datasets for model development. Examination of final data sheets (inputs and outputs), including the aggregated covariate data and the *Enterococcus* measurements, was completed by the Quality Assurance Manager to assure no errors (e.g. transcription errors from raw data sheets and conversion of meteorological data) were made.

MODELING METHODOLOGY

Model development and validation was completed using VB, a statistical software package provided by the USEPA. Procedures for developing VB models are described in Francy and Darner (2006) and Francy et al. (2013) and these steps generally were followed in this study. Within VB, statistical methods for model development included Multiple Linear Regression (MLR), Generalized Boosted Model (GBM) and Partial Least Squares (PLS) techniques described in the EPA's VB guidance (USEPA, 2012) and outlined in Table 4. The QAPP for Predictive Modeling of Microbial Water Quality at Jones Falls, Baltimore Harbor, using VB (Cyterski et al., 2018), was also an important reference for all phases of the project. All modeling for Oakland Beach used *Enterococcus* results from the center station as the response variable, while Easton's Beach utilized results from the east (i.e. second) station.

Table 4. Descriptions of Models Used in VB Data Analysis

Model Name	Multiple Linear Regression	Generalized Boosted Model	Partial Least Squares
Acronym	MLR	GBM	PLS
Model Parameters	- Uses 2 or more explanatory variables to understand relationship with a response variable by finding a linear fit - Provided the least false positives and best model fit in this study	- Uses decision tree algorithms to find the most accurate model fit - Models produced the most false positives and did not provide the best model fit in this study	- Like MLR, but can incorporate many response variables in a model - Models produced many false positives and did not provide the best model fit in this study -

For MLR, the relations between individual explanatory variables and *Enterococcus* concentrations were examined using Pearson's correlation analysis and by scatterplots generated in VB. Explanatory variables were transformed (e.g. log, square root, polynomial) to improve linearity, as needed (e.g. transformed values were selected when they improved the linear model fit by at least 20%). Explanatory variables that showed the best linear relationships with *Enterococcus* concentrations (Pearson's Correlation > 0.2), were selected to develop an additional set of explanatory variables from multiplicative interactions of sets of two variables (e.g. rain variable x tide variable).

Since MLR allows for selection of seven different statistical optimization criteria (e.g. Adjusted R-Squared, Corrected Akaike, Bayesian Inference) we used each and compared model outputs. We evaluated the benefits and limitations of each, selecting for those that limited false positives while maintaining a good model fit. We also evaluated the variables selected by each optimizing criterion with consideration for plausibility (e.g. does the relationship have a positive slope when expected?) and whether it would be practical for predictions (e.g. certain variables related to sample collection time would not be available or useful as predictors). Parsimony (lowest number of input variables) was another criterion for selection of a preferred model. After completing initial model runs with each optimizing criterion, the researchers conducted additional model runs with alternate Regulatory Criteria (applied against the measured observations of *Enterococcus*) and Decision Criteria (applied against the modeled, "fitted" measures of *Enterococcus*) to result in a balance that minimizes both incorrect exceedances (false positives) and model values that fail to predict true high values (false negatives). Appendices 2B and 3B provide summaries of results from each of the model runs that were conducted to better understand how the model performed under differing constraints.

After choosing a preferred MLR model based on the above selection factors, the model was cross validated. VB cross-validates by randomly extracting a percentage (~25% were extracted per VB guidance) of the input data to simulate how well the model would perform in the predictive mode. Here, the researcher found VB calculations with the lowest Mean Square Error of Predictions (MSEP) to evaluate the model that may have the best predictive capacity. Again, the resulting model explanatory variables were evaluated for plausibility, practicality and parsimony.

GBM and PLS model development requires no transformations or interactions of explanatory variables as preparation and they use a single process for determining "best fit". The *Enterococcus* counts were natural log transformed to produce the best fit.

After completing model development, researchers conducted the final step, using the model to predict results for a new data set of *Enterococcus* measurements and explanatory variables (2018 data). In this step, VB calculates predictions using the selected model algorithm. Output includes both the predicted values and binary results from a test against the decision criteria – whether the model correctly or incorrectly predicts values that exceed or fall below the selected decision criteria threshold.

RESULTS

The results from each case study are presented below:

OAKLAND BEACH RESULTS

Output from the preferred model selections for predicting *Enterococcus* at Oakland Beach using both GBM and MLR models are presented below. Unlike the GBM and MLR models, the PLS model outputs relied on explanatory variables other than rain, which would be less accessible and understandable with respect to pathogen transport. For this reason, GBM and MLR were selected as preferred model constructs. For all MLR tests, only those explanatory variables with alpha (p-values) of ≤ 0.05 were used. Initially, *Enterococcus* counts were transformed (natural log) prior to modeling, but output from non-transformed *Enterococcus* ultimately provided a better fit with respect to the above stated selection factors.

GBM model results are shown in Figure 2. For each count of *Enterococcus* the graph displays measured counts on the horizontal axis and the "fitted" GBM model results on the vertical axis. The values are natural log transformed so the Regulatory Criterion of 60 cfu/100 mL (for potential beach closure) presents as 4.1. A Decision Criteria threshold of 3.85 was selected to optimize the correct "exceedance" results, while avoiding false positive results. The graph shows that the model correctly predicted values greater than the Decision Criteria eight times (upper right quadrant). There were sixteen low predictions (lower right quadrant) and the remaining 57 "fitted" values were correctly predicted to be below the Decision Criteria. Initially many variables contributed to the GBM model. The "drop variables" option was used to find the subset of variables listed in Table 5, which achieved equivalent predictive power with less variables. This model identifies almost a third of the times when water quality exceeded the standard which would allow significantly increased protection of the public. It is noteworthy that three (circled in red) of the seven correctly predicted exceedances of the standard were for the highest actual exceedances. The GBM variables and their respective influences are presented in Table 5.

Table 5. GBM Explanatory Variables and Their Influence on The Model

Parameter	Influence (out of 100)
AM Low Tide Height (ft)	18.5621
Cumulative Rain 6 Days	17.6165
Previous Day Water Temp.	15.1844
Day 0 Rain	11.9532
Cumulative Rain Day 7	10.5890
Cloud Transmission	9.3074
Cumulative Rain Day 1	8.5897
Water Temp. minus Air Temp.	8.1976

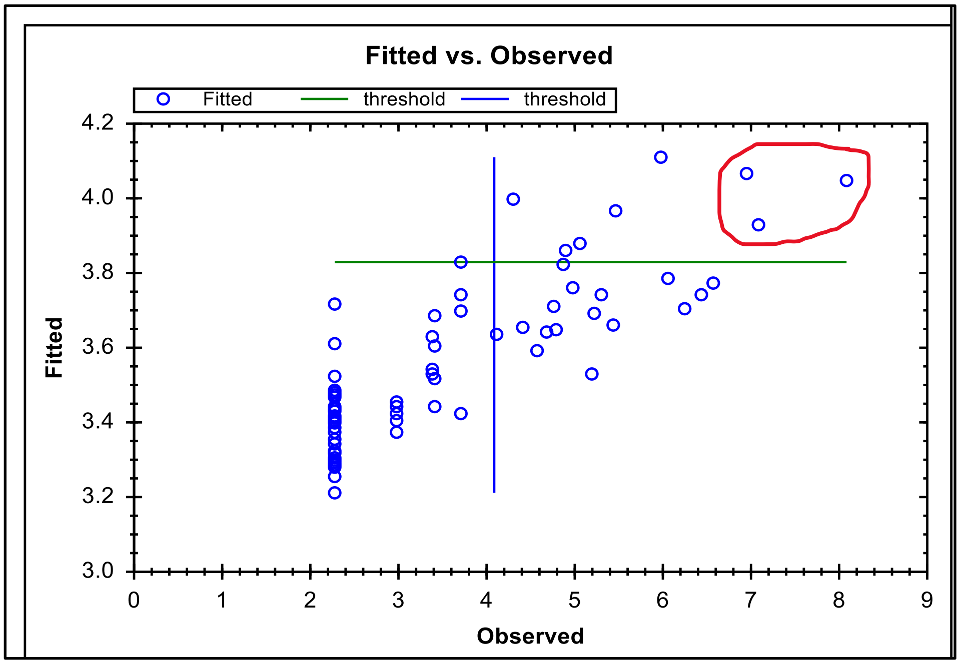


Figure 2. GBM model with eight explanatory variables; correctly predicted values greater than decision criteria (upper right quadrant) and low predictions (lower right quadrant). (Decision Criteria = 60 cfu/100 ml. Seed = 50)

The variables defined in Table 4 are as follows: the AM low tide height (AMLH) had the greatest influence, followed by cumulative inches of rainfall over the previous six days (CR_6), previous day's water temperature (P_WT), current day's rainfall (RAIN), cumulative rainfall over the previous seven days (CR_7), cumulative rainfall over the previous day and sample collection day (CR_1), and water temperature minus air temperature (WT-AT).

Figure 3 presents results from the MLR model using the Corrected Akaike Information Criteria (AICC). The model was run first with "Genetic Algorithm" selected, a VB option provided to reduce computation time. The manual model fit was used after selecting all variables that contributed to the top ten "Genetic Algorithm models," using the "Add to List" feature. By intentionally selecting a Regulatory Criteria and Decision Criteria = 300 cfu/100 mL, the figure clearly shows the range where the model performs well. Four events with observed values > 300 cfu/100 mL (above blue horizontal line) were correctly identified and four were not (below the blue horizontal line). In fact, the four which were not correctly identified appear as equivalent values below the horizontal Decision Criteria line, indicating that there were no explanatory variables in the model to elevate the predictions for those events. Notably, none of the non-predicted events were associated with more than 0.25 inches of rain over the seven-day antecedent period.

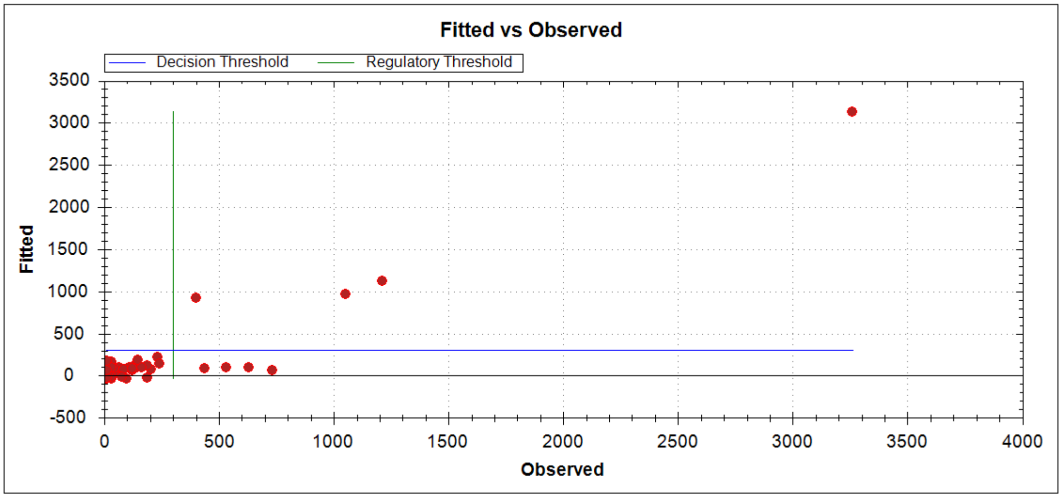


Figure 3. Fitted vs observed values from Oakland Beach MLR model, optimized using Corrected Akaike Criteria (AICC)

The statistics presented in Table 6 are defined by the following variables: product of current day RAIN and P_RD4 (four days prior rain); P_RD 1 (prior day's rain); and the product of P_RD4 (four days prior rain) and GWI (Groundwater Depth Index). The statistics indicate that the Explanatory Variable (Parameter) presented as the product of current day RAIN and P_RD4 (four days prior rain) is the most influential, with a Standardized Coefficient of 0.80, followed by a polynomial transformation of P_RD 1 (prior day's rain), and another product term P_RD4 and GWI (Groundwater Depth Index).

The algorithm for this model, where ENT_1 is the model-fitted result for *Enterococcus* is:

ENT_1 = -72.3453e00 + 21.6697e01*PROD[P_RD4, GWI] + 36.9317e02*PROD[RAIN,P_RD4] + 99.7012e-02*POLY[P_RD1,156.34711,-309.79643,184.06441]

Table 6. Explanatory Variables (Parameters) and statistics for MLR Model for Oakland Beach (2015-2017)

Parameter	Standardized Coefficient	P - Value
Product of Rain Day 0 and Previous 4 Day Rain	0.8032	0.0000
Previous Day Rain *	0.2586	4.8921e-8
Product of Previous 4 Day Rain and Groundwater Depth Index	0.1511	0.0038

*Polynomial transformed variables

For MLR, we optimized the model using the Corrected Akaike Information Criteria (AICC), a statistical approach designed to balance goodness of fit with the risk of over-fitting. The AICC also optimizes by considering the number of observations relative to the number of parameters (i.e. where the observation count is low, the number of parameters used in the model should be limited). AICC can only be used to compare a set of models produced from a specific data set, so we also used the Adjusted R-Squared values to compare models formulated with different data sets (e.g. different interactions and transformations). The model we selected had the lowest (best fit) for AICC and an Adjusted R-Squared of 0.87, which was nearly the same as if Adjusted R-Squared were used to optimize the model. While this R-Squared value is generally recognized as a very good fit (explaining 87% of the variability in the data), it is highly influenced by the high values in the (non-transformed) *Enterococcus* data. Sensitivity, Specificity and Accuracy are also useful metrics for evaluating the model. For these measures, the model still exceeds recommended goals. (Sensitivity matched the goal of 0.5, Specificity was 1.0, exceeding the ≥ 0.9 goal, and Accuracy was 0.95, also exceeding the Accuracy goal of ≥ 0.8.)

Figure 4 shows the measured data in black and the fitted model data in red in chronological order. On the vertical axis Y = *Enterococcus* cfu/100 mL). As stated by Figure 3, Figure 4 clearly shows the four times when high bacteria values (greater than benchmark blue line) are correctly predicted (when x = 3, 12, 57, 62), while also illustrating the four times there were high bacteria values but were not in fact predicted by the model (when x = 14, 36, 67, 72). These missed predications are understood due to the fact that models will never be able to predict every day of poor water quality because environmental predictors are not the only parameters for determining water quality. This illustrates the Sensitivity of 0.5 (50% correct).

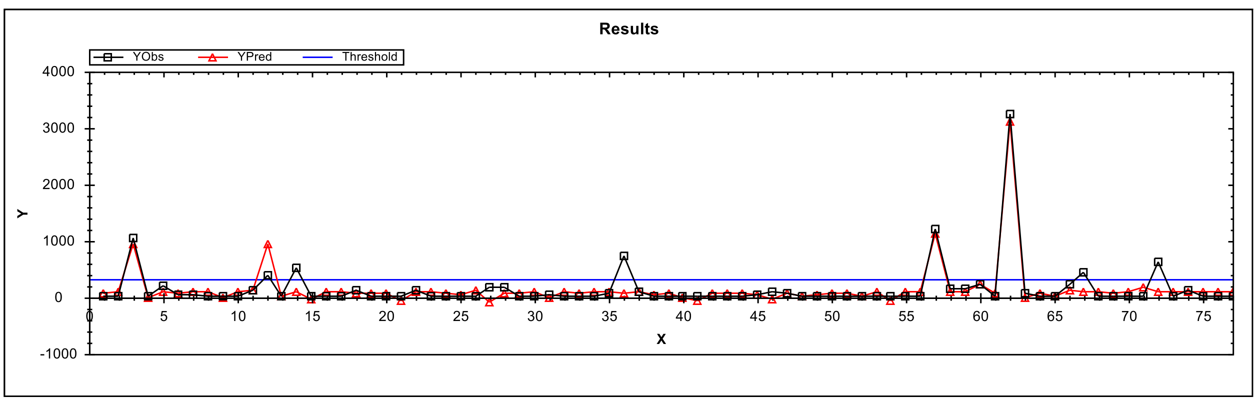


Figure 4. Chronological comparison of measured data (black) and fitted data (red) for Oakland Beach for summer samples collected in 2015 (x = 1-29), 2016 (x = 30-50), and 2017 (x = 51-77).

The next step, cross-validation, tested the model fit for 19 (25%) of the full 77 rows (dates) of data, repeating the test 5,000 times with explanatory data from different combinations of dates. For the selected AICC optimized model, the cross- validation Mean Square Error Prediction (MSEP) was calculated at 107593 while some of the other models produced in the AICC optimization produced lower MSEP (the lowest was 79879). This suggests, all other factors being equal, that a different model might produce better predictions than the model we selected. Recall, however, that we prioritized models with a low number of parameters, that produced minimal false positive results, and that used parameters that would be relatively easy to access "real-time" when we need to use the model in predictive mode. Testing both models in the prediction step provided additional insight regarding predictive success of each model.

EASTON'S BEACH RESULTS

Output from the preferred model selections for predicting *Enterococcus* at Easton's Beach using both GBM and MLR models are presented below. As seen in the Oakland Beach case study, the PLS model outputs relied on explanatory variables other than rain which would be less accessible in practical applications. For these reasons, GBM and MLR were selected as preferred model constructs. It is important to note that for all MLR tests, only those explanatory variables with alpha (p-values) of $\geq$ 0.05 were used.

Two different datasets were utilized for Easton's beach modeling exercises. The results presented first predominantly predict weekly samples collected twelve months of the year by COA from 2015 – 2017. Data from RIDOH collected in the summer months of 2015, 2016 and 2017 was utilized in these models as well. For the sake of simplicity, from this point on this data will be referred to as COA data since most of the bacteria samples being predicted were collected by COA. The results presented later only include samples collected during the summer months (late May, June, July, August and early September). Since this dataset and results include bacteria samples predominantly collected by the RIDOH, the outputs presented will be referred to as RIDOH data.

The results from an MLR model using Corrected Akaike optimization criteria can be seen in Table 7 and Figure 5. The Adjusted R Square value for this model is 0.37 (the model explains 37% of the measured *Enterococcus* variability). It can be seen that the seven statistically significant predictors of *Enterococcus* samples collected by COA from 2015 to 2017 are the product of one day previous water temperature and the time between high and low tide, the product of one day previous wind degree and cloud transmission, the product of previous air temperature and previous low tide height, the product of five days previous rain and one day previous water temperature, the product of water temperature and the wind alongshore component, three days cumulative rain and whether or not the UV disinfection light operated on the day of the sample. The negative standardized coefficient indicates that when the UV light was turned on the day of the water sampling, there were also decreased bacteria levels at Easton's Beach on the same day. Also, from Figure 5 below it can be seen that this model was able to predict seven of the highest *Enterococcus* exceedances (upper right quadrant) of the benchmark (60 cfu/100 mL) without forecasting false positive results (upper left quadrant). Additional model statistics can be found in Table 8.

The algorithm for this model, where ent2eastCOA is the model-fitted result for *Enterococcus* is:

ent2eastCOA = -7.555 - 0.0009013*(Ent2COAPriorConc) +
0.0002395*(PROD(PrevAirTemp,INVERSE(PrevLTH,-0.005))) + 0.4123*(cum3) +
0.2776*(PROD(POLY(PrevWaterTemp,-0.10576175,0.11242022,-
0.00091965705),POLY(timebetweentides,-19.924902,188.44461,-378.84368))) - 1.219*(UVDisinfection) +
0.239*(PROD(POLY(PrevWindDeg,2.2032174,0.011636093,-2.7617565e-
05),POLY(cloudtransmission,1.6766236,0.06281225,-0.00049995742))) -
0.3501*(PROD(rain5,POLY(PrevWaterTemp,-0.10576175,0.11242022,-0.00091965705))) +
0.000911*(PROD(watertemp,WindA_comp(wind_deg,windspeed,74))) +
1.753*(POLY(rain4,3.1192897,0.88834218,-0.58304864))

Table 7. MLR Results on COA Data Optimized Using Corrected Akaike Criteria

Parameter	Standardized Coefficient	P - Value
Product of Previous Water Temp and Time Between Tides*	0.2922	5.872e-5
Product of Previous Wind Degree and Cloud Transmission*	0.2711	0.0002

Product of Previous Air Temp and Previous Low Tide Height**	0.2408	0.0007
Product of Rain 5 and Previous Water Temp	-0.2075	0.0032
Product of Water Temp and Wind Alongshore	0.2066	0.0057
Cumulative Rain 3 Days	0.1966	0.0064
UV Disinfection Plant	-0.1896	0.0072

*Polynomial transformed variables **Previous low tide height inverse transformed

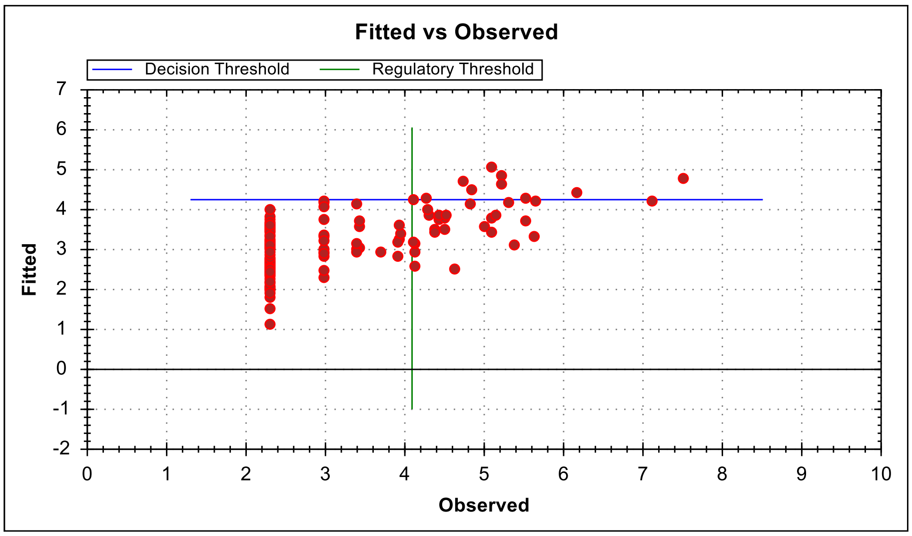


Figure 5. Graph of Fitted Samples vs. Observed for COA data; correctly predicted values greater than decision criteria (upper right quadrant) (Decision Criteria = 60 cfu/100mL).

Table 8. MLR Model Statistics Optimized Using the Corrected Akaike Information Criteria and COA data for Easton's Beach (2015-2017)

Criteria	Value
R Squared:	0.3772
Adjusted R Squared:	0.3683
Akaike Info Criterion:	142.8978
Corrected AIC:	143.8978
Bayesian Info Criterion:	23.6262
PRESS:	141.6164
RMSE:	0.9634
Sensitivity:	0.25
Specificity:	1.00
Accuracy:	0.8125

This MLR model that incorporated COA data was selected for parsimony, although it still includes seven statistically significant variables, this model utilizes less variables than other VB outputs for Easton's Beach models. The model was mainly chosen because it does not falsely predict positive *Enterococcus* values. The cross validation run on this model found that this model had the lowest MSEP of other potential models. The cross validation tested the model fit for 36 (25%) of the 144 rows (dates) of data, repeating the test 1,000 times with explanatory data from different combinations of dates. This model was utilized in the VB prediction step as well to predict 2018 *Enterococcus* data (collected by COA), and the outputs can be seen in Appendix 3C. What is noteworthy in that output is the final column, which indicates the types of errors predicted for this model. 11 false negatives, one false positive and no true positives were predicted for this model, indicating that this model, although accurate at prediction for the years 2015 through 2017, is not reliable for prediction of 2018 data. This could be since 2018 was different from the years 2015 through 2017 as it did not have many high exceedances or great rain amounts throughout the year. All independent variables excluding the NOAA UV data (cloud UV, clear sky UV, cloud transmission etc.) were included in the 2018 prediction step since there was missing data for the UV variables from October to December of 2018.

Results from another MLR model that utilized Corrected Akaike optimization criteria can be seen below in Table 9 and Figure 6. This model was run using RIDOH data throughout the summers of 2015, 2016 and 2017.

Table 9. MLR Results on RIDOH Data Optimized Using Corrected Akaike Criteria.

Parameter	Standardized Coefficient	P - Value
Product of Previous Clear Sky UV and Wind Onshore*	0.3526	3.9369e-5
Product of Rain 4 and Stream/Days*	0.3148	0.0002
Product of Rain 0 and Prior Dry	0.2609	0.0017
Product of Cumulative Rain 2 Days and Tide Difference*	0.2273	0.0089
Product of Rain 6 and Previous Air Temp*	-0.2043	0.0149
Previous UV Disinfection Plant	-0.1979	0.0221

*polynomial-transformed variable

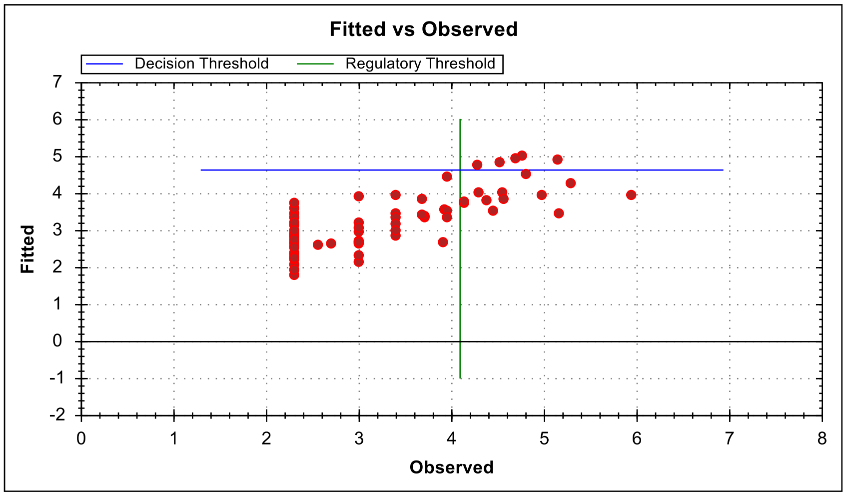


Figure 6. Graph of Fitted Samples vs. Observed for RIDOH data; correctly predicted values greater than decision criteria (upper right quadrant) (Decision Criteria = 60 cfu/100mL).

This second MLR model that incorporated RIDOH data (in Table 9 and Figure 6) has an Adjusted R Square value of 0.56, and it correctly predicted five true positive (upper right quadrant) and no false positive results (upper left quadrant). The variables with the most statistical influence in this model are the product of previous day clear sky UV and Wind Onshore, the product of rain from four days prior to sampling and Stream/Days, the product of the sampling day's rain amount and the number of days that had no rain prior to sampling, the product of two days cumulative rain and the difference between high and low tide heights, the product of six days previous rain and previous air temperature and whether or not the UV light was on the day prior to sampling. It is of note that the top five variables were all interacted products of two variables and that Stream/Days is a linearly interpolated variable that divides the last collected Easton's Stream sample by the number of days that passed between Easton's Stream sampling. Additional statistics for this model can be found in Table 10 below.

The algorithm for this model, where ent2east is the model-fitted result for *Enterococcus* is:

ent2east = -4.087 + 0.284*(PROD(POLY(PrevCSUV,38.126609,-7.8045923,0.433165),POLY(WindO_comp(wind_deg,windspeed,74),3.2173559,-0.11380623,0.020980051))) + 0.2518*(PROD(rain0,priordry)) + 0.2273*(PROD(POLY(rain4,3.1009785,2.9298767,-2.209452),POLY(Stream/days,2.9958221,0.0018240562,-6.6213746e-07))) - 0.4594*(PrevUVD) + 0.2084*(PROD(POLY(cum2,3.3868733,-1.7809224,1.4609812),POLY(tidediff,9.5371728,-3.389398,0.4377199))) - 0.2684*(PROD(rain6,POLY(PrevAirTemp,-18.22652,0.63186573,-0.0046186144)))

Table 10. MLR Model Statistics Optimized Using the Corrected Akaike Information Criteria and RIDOH Data for Easton's Beach (2015-2017)

Criteria	Value
R Squared:	0.5754
Adjusted R Squared:	0.5639
Akaike Info Criterion:	22.3028
Corrected AIC:	23.3028
Bayesian Info Criterion:	-42.6343
PRESS:	36.5696
RMSE:	0.6714
Sensitivity:	0.2941
Specificity:	1.00
Accuracy:	0.8442

This MLR model of RIDOH data (Tables 9 and 10 and Figure 6) was selected for parsimony, although it still includes six statistically significant variables, this model utilizes less variables than other VB outputs for Easton's Beach models. The model was mainly chosen because it does not falsely predict positive *Enterococcus* values. The cross validation run on this model found that this model had the lowest MSEP of other potential models generated by VB. The cross validation tested the model fit for 19 (25%) of the 77 rows (dates) of data, repeating the test 1,000 times with explanatory data from different combinations of dates.

GBM was also used for both sets of *Enterococcus* data collected by COA and RIDOH. As with MLR, the GBM decision-tree generating statistical approach, resulted in different predictors for COA and RIDOH bacteria datasets. Table 11 and Figure 7 show the relationship between measured *Enterococcus* counts and the "fitted" GBM model results produced using the COA collected data. The vertical axis is natural log transformed, so the Decision Criterion of 60 cfu/100 mL (for potential beach closure) presents as 4.1. The graph shows that the model correctly predicted values greater than the Decision Criteria 27 times (upper right quadrant), but also predicts nine values greater than the Decision Criteria incorrectly (upper left quadrant). Initially many variables contributed to the GBM model. The "drop variables" option was used to find the subset of variables listed in Table 11, which achieved equivalent predictive power with less variables. The resulting variables, in order from greatest influence to least influence are presented in Table 11: Time between tides, Solar zenith angle, Wind Onshore, High tide height, Previous time between tides, Prior Stream/Days, Previous Temperature Change, Wind Alongshore, Previous Wind Degree, Cloud Transmission, Clear Sky UV, Previous Wind Speed, and Temperature Change.

Table 11. COA GBM Results (seed = 50)

Parameter	Influence (out of 100)
Time Between Tides	16.1213
Solar Zenith Angle	11.5996
Wind Onshore	9.2841
High Tide Height	8.2524
Previous Time Between Tides	7.0027
Prior Stream / Days	6.8075
Previous Temperature Change	6.7970
Wind Alongshore	6.5434
Previous Wind Degree	6.3266
Cloud Transmission	6.1133
Clear Sky UV	5.7771
Previous Wind Speed	5.5651
Temperature Change	3.8100

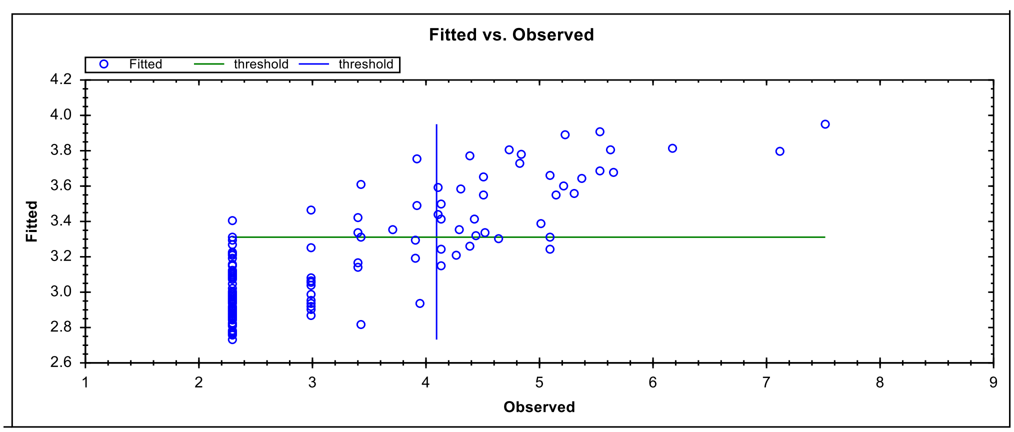


Figure 7. COA Fitted Samples vs. Observed - GBM Model for COA data; correctly predicted values greater than decision criteria (upper right quadrant) with incorrectly predicted values greater than the decision criteria (upper left quadrant) (Decision Criteria = 60 cfu/100mL).

Table 12 and Figure 8 present the results of GBM models that tested covariates against RIDOH collected data. Here, the vertical axis is again, natural log transformed, and the graph shows that the model correctly predicted values greater than the Decision Criteria 14 times (upper right quadrant), but incorrectly predicts high values three times (upper left quadrant). Like what was done for the COA model, the "drop variables" option was used to find the subset of variables listed in Table 12 that ultimately achieved equivalent predictive power. The resulting variables, in order from greatest influence to least influence are presented in Table 12: Previous high tide height, Wind Onshore, Stream/Days, Previous Clear

Sky UV, Previous Prior Dry, Cumulative Rain (7 Days), Rain 0, Stream Prior Concentration, Cumulative Rain (3 days), Wind Alongshore, Cloud UV, and Previous Could Transmission.

Table 12. RIDOH GBM Results (seed = 50)

Parameter	Influence (out of 100)
Previous High Tide Height	13.3403
Wind Onshore	10.8528
Stream/Days	9.9204
Previous Clear Sky UV	9.2910
Previous Prior Dry	9.1666
Cumulative Rain (7 Days)	8.0036
Rain 0	7.6289
Stream Prior Concentration	7.2777
Cumulative Rain (3 Days)	6.8062
Wind Alongshore	6.6918
Cloud UV	6.1836
Previous Cloud Transmission	4.8371

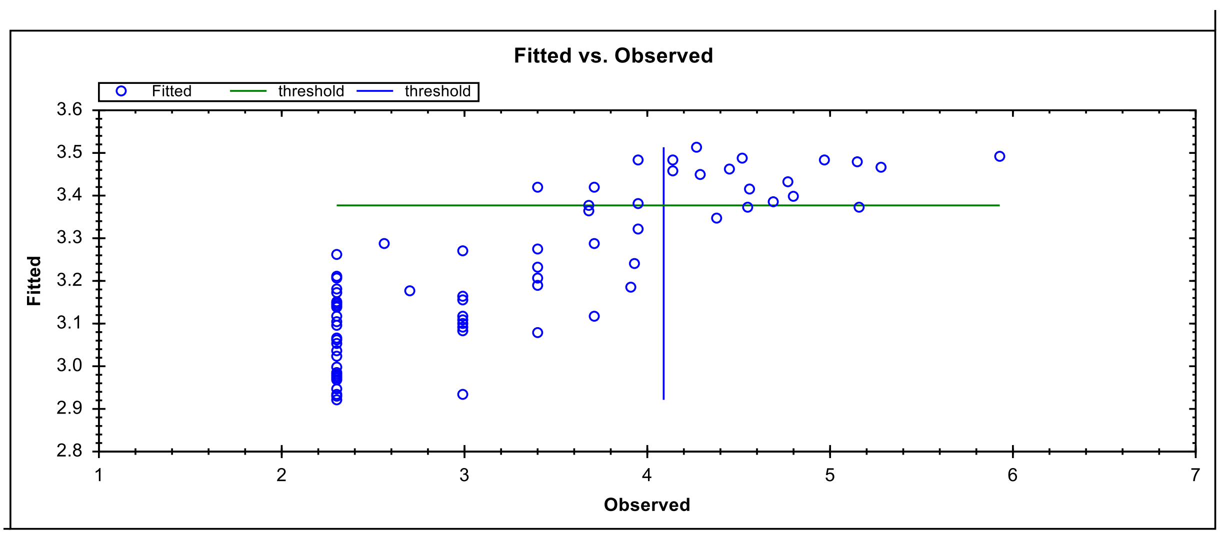


Figure 8. RIDOH Fitted Samples vs. Observed – GBM Model with RIDOH data; correctly predicted values greater than decision criteria (upper right quadrant) with incorrectly predicted values greater than the decision criteria (upper left quadrant) (Decision Criteria = 60 cfu/100mL).

DISCUSSION

OAKLAND BEACH CASE STUDY

Decisions regarding which model formulation may have the greatest benefit to management depend on the intended purpose. One purpose is to identify clues to what the greatest influences on water quality impairment may be, and another is to use the model for prediction to protect public health on a day-to-day basis. For the former, both the GBM and MLR models indicated that antecedent rain (going back four to six days) may be as important as rain that occurs on the day of sampling. This has implications for better understanding of fate and transport pathways. Two possible scenarios consistent with delay in pathogens reaching Narragansett Bay water several days after a rain is that perched groundwater may be a contributing factor – rain collected in swales may travel through subterranean pathways over the course of several days. Another possibility is that waters from the boundaries of the site (e.g., Greenwich Bay or Upper Narragansett Bay) may be contributing.

With respect to improved protection of the public, the selected MLR model for the Oakland beach study demonstrated a good fit 50% of the time sample results were greater than 300 cfu/100 mL. Applying this model (closing the beach when the model predicts values over this threshold) would theoretically provide protection to the public four times out of eight when conditions are worst at Oakland Beach. This is four times more than current practices, due to the two-day effective delay from the time samples are collected to when analytical results are available to make beach closure decisions. The City of Warwick has expressed support for wanting to use this model while continuing to test its predictive capabilities with sample analytical results. The decision to opt for the MLR model over the GBM model is due to the parsimony of the three-variable model over an eight-variable model. During 2019, researchers will exercise both models to continue to track the performance of each.

One final exercise was completed with the Oakland Beach model input. Using the set of explanatory data compiled for Oakland Beach, Warren Town Beach *Enterococcus* data for the same 2015 to 2017 sample days were substituted into the input file in place of Oakland Beach data. Figure 9 on the following page shows that the resulting GBM model is promising, given that it correctly fit acceptable water quality (upper right quadrant) and was also able to fit about half of the impaired water quality data. It is notable that the variables of highest influence were measures of light transmission (CLSUV and tide). While most of the weather, tide and atmospheric explanatory variable values that populate the Oakland Beach model are expected to be valid for Warren Town Beach, it has a very different location (within the Palmer River estuary), and orientation with respect to water movement. This preliminary exercise does indicate promising opportunities to extend the work conducted for the current study toward applications at other beaches.

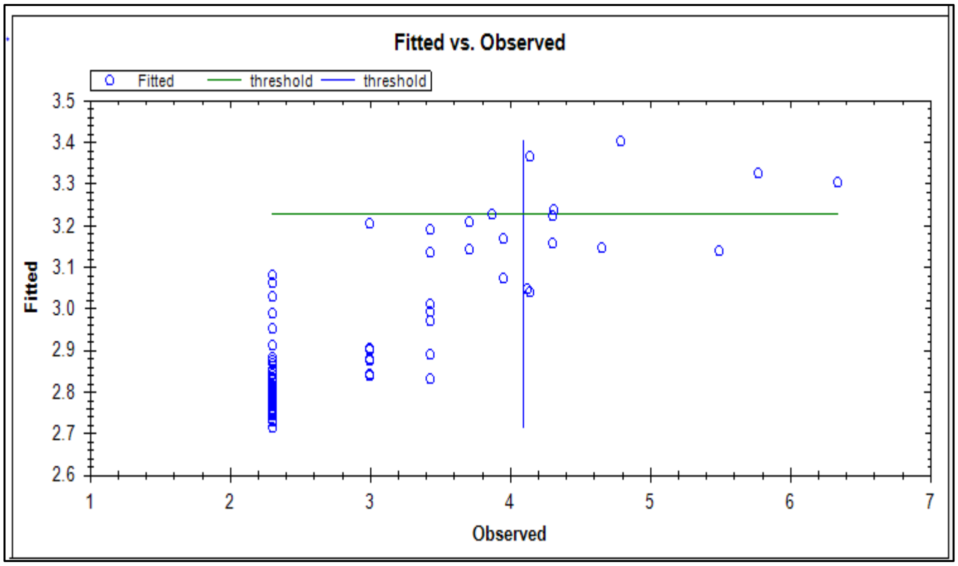


Figure 9. Warren Town Beach GBM Model Fitted vs Observed Graph Derived from Explanatory Variables Used for Oakland Beach (Substituting Warren Town Beach Enterococcus Data); correctly predicted values greater than decision criteria (upper right quadrant) (Decision Criteria = 60 cfu/100mL).

Table 13. Warren Town Beach GBM Model Result Parameters Derived from Explanatory Variables Used for Oakland Beach (Substituting Warren Town Beach Enterococcus Data).

Parameter	Influence (out of 100)
Cloudy Sky UV	23.4376
AM Low Tide Height (ft)	11.0186
AM High Tide Height (ft)	6.6129
Change in Tide Height (ft)	4.5407
Time Between High and Low tide	4.4151
Cloud Transmission	4.3483
Previous Hunt River Flow	3.6058
Previous Clear Sky UV	3.1934
Wind Speed	2.8635
Total Column Ozone	2.7834
Clear Sky UV	2.4081
AM low tide	2.3649

EASTON'S BEACH CASE STUDY

The results from the Easton's Beach case study indicate that MLR and GBM models present different sets of explanatory variables to obtain the best fits for prediction. The MLR models from both COA and RIDOH data emphasize weather variables. Wind was more prominent for the COA model while natural UV light transmission factored more heavily in the RIDOH model. In both, current and prior rain events, temperature, and operation of the UV Disinfection Treatment Plant, were important predictors of *Enterococcus* bacteria at Easton's beach.

In GBM models for both data sets, tide variables were the most influential factors, in addition to wind, rain and natural UV light. While the GBM models, as configured, resulted in several false positive results (wherein a decision to close the beach would be in error), this problem can be addressed by adjusting the Decision Criteria upward. Taking this step would also reduce the number of correct predictions for closure, making the outcome similar to the MLR model (with a smaller number of closure predictions)

Consideration of the value of each Easton's Beach model formulation will depend on the intended purpose. One goal was to use the model for prediction to protect public health on a day-to-day basis. Another was to use it to identify clues regarding the factors influencing water quality impairment. It could be an option to use one of the models produced with Easton's Beach data for "worst case" predictions only. However, given that no true positive results were obtained in the prediction of 2018 data, it may be best to test model performance with another year of new data, and to re-evaluate at that time.

LESSONS LEARNED

The statistical models do present the opportunity for better protection of public health than current state-level practices. However, more information is needed from management in order to determine the level of risk beach managers are willing to accept for both false positives and false negatives since false positives risk closing beaches when they should not be, while false negatives risk not protecting public health.

There were two main findings regarding clues to obtaining a better understanding of bacteria sources, fate and transport:

1. Existing paradigms that emphasize the importance of same day or immediately prior rain were tempered by our findings. In some model runs, VB found that more temporally distant rain (up to six days antecedent) contributed importantly to the predictive outcomes. This was especially true for Oakland beach.
2. UV light intensity was inversely correlated with *Enterococcus* counts; more use or improved use of UV treatment could improve water quality. This was particularly true for Easton's Beach.

An important lesson learned from both case studies was the large time commitment for development of the input file of explanatory variables. The QA/QC steps revealed that meteorological data almost always had gaps where data for a subset of the requisite dates were missing. Retaining as many dates (rows) of data as possible required identification and review of multiple sources to find the one(s) with the fewest data gaps, while also having confidence in the quality of the data. Once the input dataset was completed, the next challenge was to become 'fluent' in the capabilities of VB. There are numerous decisions required regarding selection of tools to estimate the predictive capacity of any data set/model formulation. For instance, in the set of optimizing criteria that includes Adjusted R squared, Corrected Akaike and Bayesian Inference, PRESS is another option. The latter uses a random sampler to sequentially remove one row of data at a time from the model to determine if the model could predict the outcome for that day. The PRESS method produces a set of explanatory parameters that theoretically could represent a better "predictive" capacity than the other criteria, which focuses on best fits of the entire "training" data set. Through another VB tool, the Cross-Validation step takes a larger subset (~25%) of the per-date data set to test how well the remaining data predicts the subset of data (again in a random sampling approach). The final prediction step is the only way to test a real-world prediction on "untrained" data. In this study, neither of the "training" data sets (for Oakland or Easton's Beaches) were able to predict the unusually low number of elevated *Enterococcus* results in 2018. This emphasizes the important point made in all the guidance for beach water quality modeling – that the potential to develop a good model is always greater when the water body has many poor water quality samples.

CONCLUSIONS AND NEXT STEPS

The case studies have served as a successful learning exercise. COA and RIDOH now have in-house capabilities to use analysis of raw data and VB modeling tools to evaluate temporal trends in beach water quality data, and to better evaluate health risks and their root causes at Rhode Island beaches. While the current models were not predictive of poor water quality results obtained during 2018, this result was expected, given that there were few high values in 2018.

The municipal partners who are responsible for managing each of the case study beaches have been supportive and helpful in providing feedback regarding their understanding of the results from the study. The research team's experience with these partners and with public outreach for this study show an increased understanding and willingness to grapple with the difficult challenges of health protection and management to reduce pathogen loads. As of currently, the City of Warwick is interested and excited to advance the use of Virtual Beach modeling as part of their monitoring process, whereas the City of Newport would like more exploration of the modeling tool using more recent data.

Principally, this project has clearly documented the need to face the complicated questions posed by impaired waters. The study findings are a warning that oversimplified planning can result in ineffective protection of the pubic on a day-to-day basis, and ineffective management strategies to remove or reduce pathogen loads. The project provides a framework for the next steps that will be needed to deliver the benefits of improved water quality and associated values of the beaches of Narragansett Bay.

Specifically, the research team recommends:

- Using the dataset of explanatory data to develop models for other Narragansett Bay and Aquidneck Island beaches (Oakland Beach for the Upper Bay and Easton's for Lower Bay and Aquidneck Island)
- Testing the models using 2019 data
- Improving practical implementation by substituting variables that can be easily obtained on a real time basis (e.g. natural UV light variables obtained from NOAA might be replaced by on-site measurements at Easton's Beach)
- Continuing interpretation of model results to find additional clues to better understand sources and their fate and transport processes; Use these findings to develop a conceptual model for mechanistic models (e.g. SWAT) (e.g., for both Oakland and Easton's Beaches)
- Making immediate changes to UV treatment protocols (time of year, levels of rain) for Easton's Beach
- Querying results from ongoing studies (e.g., Find it and Fix it) to harvest additional data to improve model performance

- Model validation by collecting a water sample on the same day of the predicted closure by the model

One immediate consequence of the study is that the City of Warwick plans to use VB to inform beach closure advisories during the summer of 2019. Given the remaining uncertainties, the model results will only trigger beach advisories when *Enterococcus* levels exceed 300 cfu/100 mL, well above the Beach Action Value of 60 cfu/100 mL. This is a conservative approach to begin to integrate the new capability under conditions that are most likely to impact public health, and not in the range where predictions are less certain. For Easton's Beach, 2019 data will provide critical additional information and tests of the existing models. At this high-use, high-visibility location, other ongoing studies of pathogen sources will provide additional information to improve the predictive capabilities of the VB model. The resources are coming together to meet the critical need for improved management.

APPENDICES

Appendix 1 – Data Acquisition Document

METHODOLOGY

At the onset of this project, it was recognized that all sources of environmental data to be used in statistical analyses needed to be identified and documented. The proposed variables for the Oakland Beach and Easton's Beach case studies were identified prior to the Virtual Beach modeling, and the purpose, source, method of extraction and means of quality control were documented. The purpose and reasoning for the selection of the independent variables used in the Oakland Beach and Easton's Beach case studies were derived from the EPA's guide *Six Key Steps for Developing and Using Predictive Tools at Your Beach* (2016). When describing how to choose independent variables for analysis, the EPA notes, "…your Beach Team should become familiar with the likely sources of bacteria that affect the beach, how they are transported to the beach area, and conditions that tend to increase or decrease Fecal Indicator Bacteria (FIB) density in the swimming area" (16). Based on these guidelines, the research team for the Oakland Beach and Easton's Beach case studies used prior knowledge of Oakland Beach and Easton's Beach and their surrounding natural environments of the beaches, page 16 of the EPA guide *Six Key Steps…*, to determine which environmental parameters should be used for data analyses.

The following statements from the approved Quality Assurance Project Plan (QAPP) serve as a standard guide for quality assurance and control: "For any data that requires processing (from a simple task such as consistency of significant digits to a transformation such as converting from Fahrenheit to Celsius) the quality assurance effort will prescribe that these data files will include checking at least 5% of the data in each of the independent and response variable columns by staff other than the resource that performed the initial data processing. If more than one error is found in the 5% sample size, the entire column will be re-checked for accuracy" (McLaughlin & Poucher, 2018). Therefore, for every variable provided in this document, at least 5% of data points must be quality checked according to the specified means of quality control for that variable.

The sections below detail the kinds of bacteria and environmental data used for modeling, along with sources, extraction methods and quality control measures for each. In total across both case studies, 34 independent variables have been extracted from various sources, 53 independent variables have been derived or calculated from those that were extracted, and there are 4 Fecal Indicator Bacteria (FIB) response variables. Some variables are specific to only one case study and these are identified within each section.

INDEPENDENT ENVIRONMENTAL PARAMETERS

The following environmental parameters were identified as potential independent variables for Virtual Beach analyses.

Primary Source Variables		
Date	Air Temperature (F)	Hunt River Flow (ft^3/sec) [O]
Precipitation	Wind Direction	Hunt River Tide Gage Height (ft) [O]
AM High Tide Time	Wind Degree	Clear Sky UV Index
AM Low Tide Time	Wind Speed (mph)	Cloudy Sky UV Index
PM High Tide Time	Weather Condition	Cloud Transmission
PM Low Tide Time	Surf Condition	Solar Zenith Angle
High Tide Height(ft)	Tide (High or Low)	Aerosol Transmission
Low Tide Height (ft)	Vegetation in Water (Percent Covered)	Total Column Zone
Enterococcus Station 1 Sample Time	Visible Sewage or Odor	South Esplanade Enterococcus [E]
Enterococcus Station 2 Sample Time	Water Condition	Easton's Stream Enterococcus [E]
Enterococcus Station 3 Sample Time	Seagull Count	Easton's Point Enterococcus [E]
Water Temperature (F)	Warwick Neck Groundwater Depth (ft) [O]	Newport UV Disinfection Light [E]

[O] Specific to Oakland Beach case study

[E] Specific to Easton's Beach case study These parameters were all directly transcribed or downloaded from the internet or manually collected by the Rhode Island Department of Health (RIDOH) and Clean Ocean Access (COA) for the Oakland Beach and Easton's Beach case studies. The main sources for these parameters are Weather Underground (WU), the National Oceanic and Atmospheric Administration (NOAA) and Coastal Survey Sheets used by RIDOH and COA to record environmental data and the time of the collected water samples at the beaches. At both beaches, there are 3 stations for which FIB bacteria is collected: Center station 1, East station 2 and West station 3. RIDOH and COA also are the sources for the bacteria data at Easton's Point, Easton's Stream and the South Esplanade in Middletown, RI that were used as independent, environmental parameters in the Easton's beach case study. The United States Geological Survey (USGS) site was also used to obtain groundwater and Hunt River data for the Oakland beach study. Below, the sources and extraction procedures for independent environmental parameters are detailed and are separated by data source. To *quality check* any of these variables, 5% of data points should be checked against the original source named below.

WEATHER UNDERGROUND

Weather Underground (WU) was used to obtain **Air Temperature**, **Windspeed**, **Wind Direction** and **Precipitation** data for every day examined in the Oakland beach case study and for every day of November 2018 in the Easton's beach case study. Weather Underground was only used to supplement the Easton's study in November of 2018 because the principal source used to extract these variables for Easton's beach, the NOAA climate data website detailed below, was missing data for November 2018. To extract these data points, custom dates are entered on the WU page. The following links were used to access data closest to Oakland and Easton's beaches:

- Oakland Beach (KPVD station): https://www.wunderground.com/history/airportfrompws/KPVD/2016/6/18/CustomHistory.html?dayend=31&monthend=6&yearend=2016&req_city=&req_state=&req_statename=&reqdb.zip=&reqdb.magic=&reqdb.wmo= .

- Easton's Beach (KUUU station): https://www.wunderground.com/history/airportfrompws/KUUU/2015/1/1/MonthlyHistory.html?req_city=&req_state=&req_statename=&reqdb.zip=&reqdb.magic=&reqdb.wmo=

Alternatively, a researcher can visit the WU website and follow the subsequent procedure: (1) Search for the desired location on the homepage. (2) On the station's page, "history" should be clicked. (3) Custom dates can then be entered for the days that this data is sought. (4) Scroll down to view the resulting data table of average air temperature, dew point, humidity, sea level pressure, visibility, windspeed and precipitation. (5) Copy and paste the whole table into Notepad. (6) Save it as text file and open it in Excel. (7) Choose "Delimited by space." (8) Click "General" on the last screen. (9) Click finish. (10) Save as an excel file. (11) Copy and paste the desired variables into the Virtual Beach input file.

It should be noted that for precipitation data, data was gathered for 7 days before the first FIB "result" date available to the researchers in both case studies. For instance, the first FIB result date for Easton's beach was 1/8/2015, therefore, precipitation data was gathered from 1/1/2015 to 12/31/2018.

In order to obtain **wind direction** data for the Oakland beach case study: (1) The above link was used to access weather underground. (2) "Custom" mode had to be used to enter each date manually to receive the wind direction of that day. (3) Each wind direction value was manually transcribed into a new "wind direction" column of the Virtual Beach input file. The Easton's case study also utilized this procedure for missing **wind direction** data for the dates 7/30/2015, 7/9/2016, 6/7/2017, 7/25/2017 and from 11/1/2018 - 11/30/2018.

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

For both case studies multiple National Oceanic and Administration (NOAA) sites were consulted for various variables. The NOAA *tides and currents* site was used in both case studies for **Water Temperature** data and also used in the Easton's beach study for **AM Low Tide (Time)**, **PM Low Tide (Time)**, **AM High Tide (Time)**, **PM High Tide (Time)**, **High Tide Height** and **Low Tide Height**. In the Oakland study, the NOAA *Mobile Geographic* site was used to obtain **High and Low Tide Height** data. In the Easton's study, **Air Temperature**, **Precipitation**, **Wind Direction** and **Wind Speed** were all gathered from NOAA's *Climate* data site. Finally, for both studies the NOAA *Ultraviolet* data site was consulted for the **Clear Sky UV Index**, **Cloudy Sky UV Index**, **Cloud Transmission**, **Solar Zenith Angle**, **Aerosol Transmission** and **Total Column Zone** data points. Specific extraction procedures are detailed in each website's section below.

I. TIDES AND CURRENTS

The following links were used to access data closest to Oakland beach and Easton's beach respectively:

- Oakland Beach (station: 8452944 Conimicut Light, RI): https://tidesandcurrents.noaa.gov/physocean.html?bdate=20150526&edate=20150904&units=standard&timezone=GMT&id=8452944&interval=h&action=data

- Easton's Beach (station: 8452660 Newport, RI): https://tidesandcurrents.noaa.gov/noaatidepredictions.html?id=8452660&units=standard&bdate=20150828&edate=20150904&timezone=LST/LDT&clock=12hour&datum=MLLW&interval=hilo&action=data

To gather **water temperature** data, from each of these links: (1) The "phys. oceanography" tab was selected. (2) Then desired dates were entered. (3) "1 hr." interval was clicked. (4) Within the "time zone" drop down menu "LST/LDT" was selected. (5) The "data only" button must be clicked to refresh the page. (6) Water temperature at every hour of the day at local times can be seen by scrolling down the page.

For Oakland Beach, the water temperature at 8:00 AM was chosen and manually recorded in the researcher's spreadsheet since this is around the average time when the RIDOH sampled at Oakland Beach. For Easton's beach, the water temperature at 7:00 AM was chosen and manually recorded in the researcher's spreadsheet since this is around the average time when COA sampled at Easton's beach.

For this procedure to be replicated at other beaches, the above links can be accessed but the researcher will have to choose his or her station under "Options for" after clicking the "phys. oceanography" tab. It should be noted that for the Oakland beach study, on the days when NOAA had missing data, water temperature data from RIDOH's daily sampler sheets were used as a supplement. Similarly, for the Easton's beach study, COA and RIDOH coastal survey sheet water temperature data were used to supplement the missing NOAA data.

Both case studies also supplemented missing water temperature data with data collected from a buoy the University of Rhode Island's Graduate School of Oceanography. The researchers contacted and received this data, for 2015-2018, from Heather Stoffel.

To gather **AM and PM high and low tide times** for both case studies and **high and low tide heights** for Easton's beach: (1) The above links were clicked. (2) Then the "tide/water levels" tab was clicked. (3) "NOAA tide predictions" was selected. (4) Click the drop-down menu under "options for" to select the station for which data is sought. (5) Input the preferred dates. (6) Scroll to the bottom of page to choose the greater tide height of the high tides given for a day, the lowest low tide height and the times for AM high tide, AM low tide, PM high tide and PM low tide. (7) All these values can be directly transcribed into the researcher's spreadsheet in different columns for each type of data.

It should be noted that if only one high tide was given for a day, then that tide's height was used for high tide height. The same applies to low tide height as well.

II. MOBILE GEOGRAPHICS

For Oakland beach the following link was used to gather **high and low tide height** information:

- https://tides.mobilegeographics.com/

(1) Click the above link. (2) The station Nyatt Point, Narragansett Bay was typed into the search bar, alternatively enter the station from which data is sought. (3) Click search. (4) Scroll down to the bottom of the new page. (5) Enter the year and month for which data is desired. (6) Copy and paste the whole table into Notepad. (7) Save it as text file and open it in Excel. (8) Choose "Delimited by space." (9) Click "General" on the last screen. (10) Click finish. (11) Save as an excel file. (12) Copy and paste the high tide height and low tide height columns into Virtual Beach input file (one column for each type of data).

It should be noted that the research team ran into issues using this data source for Oakland beach tide height data and then decided to use the NOAA Tides and Currents site, the process detailed in the previous section, instead for the Easton's beach study.

III. NOAA CLIMATE DATA

The following link was used to gather **air temperature**, **precipitation**, **wind degree** and **wind speed** data for Easton's beach. This can also be used for other beach modeling studies by entering a different station in the eleventh step of the provided procedure.

- https://www.ncdc.noaa.gov/

To compile this information for Easton's beach, the following procedure was used.

(1) Click source link. (2) Click data access. (3) Click land-based station. (4) Click climate data online. (5) Click browse datasets. (6) Click daily summaries. (7) Click search tool. (8) Enter the date range for which you want data (it is advised to start 7 days before the first date for which FIB results are available if obtaining precipitation information). (9) Click apply. (10) In the "search for" drop down menu click stations. (11) Type in station in the

enter a search term box (Newport airport station is KUUU). (12) Click search. (13) Click add to cart next to the desired station. (14) Click on the cart in the top right. (15) Click view all items. (16) Click custom GHCN-Daily csv file. (17) Click continue. (18) Confirm station name is checked off at the top of the webpage. (19) Scroll down and make sure that precipitation, air temperature, wind and weather type are all checked off (if some of these variables are not desired then do not check them off). (20) Click continue. (21) Type in the email address that the data should be sent to. (22) Click submit order. (23) Check email and download csv file of desired data.

Once downloaded, open the downloaded file and look under the "PRCP" column for precipitation, "AWND" for average wind speed and "WDF5" for the direction of the fastest 5-second wind gusts on a day. The "TAVG" column may be empty. If it is, use the average function in Excel to take the average of the "TMAX" and "TMIN" values. These columns were selected and then copied and pasted into the research assistant's Virtual Beach input spreadsheet. It should be noted that "WDF5" gives wind direction in degrees, therefore, this data was copied into a wind degree column for the Easton's beach study's Virtual Beach input file. For Oakland beach, wind direction data (from Weather Underground) was converted into degrees and this Excel calculation will be detailed in the Derived Environmental Parameters section of this document.

IV. NOAA ULTRAVIOLET DATA

For both case studies the following link was used to extract the ultraviolet data variables **Clear Sky UV Index**, **Cloudy Sky UV Index**, **Cloud Transmission**, **Solar Zenith Angle**, **Aerosol Transmission** and **Total Column Zone**.

- http://ftp.cpc.ncep.noaa.gov/long/uv/cities/

(1) Click the source link. (2) Click on the year for which data is needed. (3) Then scroll to and click on the airport closest to the beach for which you need data. The PVD station was used for both the Oakland and Easton's data sets since this is the only available UV source for Rhode Island. (4) Data for every day of the year will be displayed in 6 columns from left to right: Clear Sky UV, Cloud UV, Cloud Transmission, Solar Zenith Angle and Total Column Ozone respectively. (5) Copy and paste all the data into Notepad. (6) Save it as text file and open it in Excel. (7) Choose "Delimited by space." (8) Click "General" on the last screen. (9) Click finish. (10) Save as an excel file. (11) Copy and paste the data columns into the Virtual Beach spreadsheet. This process must be repeated for each year that UV data is desired.

V. NOAA NATIONAL DATA BUOY CENTER

For the Easton's case study the following link was used to extract supplementary **Water Temperature** data.

- https://www.ndbc.noaa.gov/station_history.php?station=nwpr1

(1) Click the source link and then click on the year for which data is sought. (2) Click the text file and download it. (3) Copy and paste all columns of data into a word pad or word processing document. (4) Save this document as a CSV file. (5) Open the CSV file in Excel. (6) Choose "Delimited by space." (7) Click "General" on the last screen and then click finish. (8) Save as an Excel file. (9) Copy and paste the water temperature (or other desired variables into the Virtual Beach spreadsheet). This process must be repeated for each year that data from this buoy is desired.

COASTAL WATER AREA SURVEY SHEETS

Both RIDOH and COA samplers collect the following environmental parameters at Oakland and Easton's beaches and record them on "Coastal Water Area Survey" sheets: **Weather Conditions** (clear, cloudy or rainy), **Water Conditions** (clear or cloudy/murky), **Tide** (low, mid or high), how much of the water's surface is covered by **Vegetation** (<25%, 25%, 50%, 75% or 100%), the number of **Seagulls**, **Water Temperature** and whether or not there is **Visible Sewage or Odor**. **Surf Conditions** (slow, moderate or high) were also collected by RIDOH. Other data is collected with these sheets but only the variables listed were utilized for statistical analyses since the reliability of the other variables are questionable. For instance, wind strength could be difficult to determine for a citizen scientist volunteer for Clean Ocean Access, it is not being used in the marine beaches case studies.

The information collected on these sheets were recorded in yearly spreadsheets for RIDOH and COA by interns and employees. These spreadsheets were made available to the researchers for the Oakland and Easton's beach case studies. For both case studies, data was copied and pasted from the yearly spreadsheets into the VB input file.

UNITED STATES GEOLOGICAL SURVEY (USGS)

The following link was used to obtain **Warwick Neck Groundwater Depth** data from a USGS well.

- https://waterdata.usgs.gov/nwis/inventory?agency_code=USGS&site_no=414106071223901

(1) Click the above link. (2) Click field groundwater level measurements. (3) Click reselect period. (4) Enter the desired dates. (5) Click tab-separated file as a format. (6) Click submit. (7) Once downloaded, the values under the "10s" column were manually transcribed into the Virtual Beach input file.

Since there was only one value provided for each month, all other values had to be linearly interpolated in Excel. Values were interpolated by subtracting two data points from the originally downloaded file and then dividing by the number of days between them. This value is added to the first data point and then added again for each corresponding day up until the second data point. Then the calculation is performed again between the second and third real data points (from the downloaded file) and the process is repeated until the column is complete.

To find data for a different well or stream site: (1) Click the link above. (2) Under the data category click site information. (3) Then find the desired site by choosing a geographic area (a state or territory). (4) Select different identifiers or attributes (these are given on the website) of the well or stream site. (5) Once the site is found, click access data and all available data will be found through different links on that page.

The following link was used to obtain **Hunt River Flow** and **Hunt River Tide Gage Height** data.

- https://waterdata.usgs.gov/nwis/dv?cb_00060=on&format=html&site_no=01117000&referred_module=sw&period=&begin_date=2015-05-25&end_date=2015-09-04

For the **river flow** data: (1) Click the above link. (2) Enter the desired dates. (3) Select tab-separated under "Output Format." (4) Check off daily discharge to give the river flow data. (5) The data table should be copied and pasted into a notepad file. (6) Save this as a .csv. (7) Open the .csv in Excel. (8) Choose "Delimited by spaces." (9) Copy and paste the river flow data into the Virtual Beach input file.

For the **tide gage height** data: (1) Click the above link. (2) Select "time series: current observations," from the drop-down menu next to "Available data for this site." (3) After the page loads, the gage height at a particular time on each day can be recorded in the Virtual Beach input file. For Oakland beach, the gage height at 8:00 am was recorded for each day since 8:00 am is the time when bacteria samples are collected at Oakland beach by RIDOH.

RHODE ISLAND DEPARTMENT OF HEALTH

RIDOH and COA collect bacteria samples at Oakland and Easton's beaches and record these sample times on the coastal survey sheets mentioned previously that RIDOH then gathers and records in chain of custody forms (Note that RIDOH collects samples at both beaches from 3 different stations while COA only collects samples from the East station of Easton's beach).

All sample information for both beaches is then documented in an Excel spreadsheet by RIDOH. **Station 1, 2 and 3 sample times** for both case studies (collected by RIDOH) were taken from this Excel spreadsheet to perform tide to low tide calculations (described in the next section of this document) for summer bacteria samples. Also taken from this spreadsheet was the **Easton's Stream** bacteria data to be looked at as independent environmental variables. COA also provided Easton's Stream, **Easton's Point** and **South Esplanade** bacteria data from a spreadsheet that is updated annually by COA employees and Sherry Poucher from RIDOH. For the Easton's beach study, 2 different columns were created for the Easton's Stream bacteria variable to differentiate between the RIDOH and COA samples.

Alternatively, sample times and bacteria data can be found on the EPA's website: https://watersgeo.epa.gov/BEACON2/ by: (1) Clicking the link. (2) Searching for the desired beach, stream or other type of station. (3) Selecting "water quality" data. (4) Inputting the desired dates for which bacteria sample data is sought.

NEWPORT UV DISINFECTION TREATMENT PLANT

In May 2011, the Newport Department of Utilities began operation of a UV Disinfection treatment plant at the Newport-Middletown line. In Newport, stormwater from the moat around Easton's Pond is pumped through a screen that removes debris and trash before the water flows through a chamber with 336 light bulbs that cast the ultraviolet rays. The light is only activated when a quarter inch of rain falls into the moat and Pond. These rays kill more than 99 percent of all bacteria, viruses and other waterborne pathogens before the water is discharged into the stream that flows into Easton Bay, and later Easton's beach. The ultraviolet plant is in operation from Memorial Day to Labor Day and can treat up to 62 million gallons of water per day, or 40,000 gallons a minute.

Data regarding whether the UV Disinfection treatment light was on during the summers of 2015-2018 was sought from Newport's Director of Utilities Julia Forgue.

DERIVED ENVIRONMENTAL PARAMETERS

The following environmental parameters were used as independent variables for Virtual Beach analyses but have been calculated or derived from the variables listed in the first section.

Derived Variables
Number of Previous Dry Days
Previous Day Values For All Variables
Cumulative Rain Values
Time Between High and Low Tide
Time Between Collection and Low tide (Station 1)
Time Between Collection and Low tide (Station 2)
Time Between Collection and Low tide (Station 2 - COA) [E]
Time Between Collection and Low tide (Station 3)
Wind Degree
Change in Tide Height
Difference in Temperature
Groundwater Indicator [O]

[O] Specific to Oakland Beach case study

[E] Specific to Easton's Beach case study

The procedures to derive each of the variables, or groups of variables, are listed within the following three categories.

I. Previous Day Variables

II. Cumulative Variables

III. Other Calculated Variables

To *quality check* any of the “Previous Day” or “Cumulative” variables, 5% of data points should be checked against the original, independent variables that they were derived from. All “Calculated” variables should be quality checked by manually performing the calculations for 5% of all data points.

PREVIOUS DAY VARIABLES

This section describes how to create “previous day” columns for all independent variables 1 – 34 listed in the independent environmental parameters section and for all calculated variables detailed below. For all of these variables, a “one day previous” column was created to see if a variable’s value on the day prior to a given day affects bacteria concentration at marine beaches.

To populate a “one day previous” column for any of the 34 independent variables and the calculated variables, the following process was used. Previous day air temperature is used as an example below.

(1) Create new column. (2) Click the cell that corresponds to the second date of the whole dataset within the new “previous air temperature” column. (3) Click function box. (4) Type "=". (5) Click the cell of the original air temperature column that corresponds to the first date of the dataset. (6) Hit enter. (7) Click the working cell of the “previous air temperature” column, now with a value in it, and drag the right bottom corner all the way down the rest of the column to fill in the whole column with 1-day previous air temperature values.

For precipitation data, 7 “previous day” variables were created to test how precipitation from 1 day to 7 days prior to a given date might affect bacteria levels at the beaches. To create columns for “2 days previous precipitation,” “3 days previous precipitation” etc. the procedure listed above can be referenced for guidance. In the function box, instead of clicking the cell from one day before a given date, click on cells that correspond to two days, three days etc. prior to the given date. Note that for the previous precipitation variables (1 day previous, 2 day previous etc.) the first cells that should be populated are those that correspond to the first day for which FIB bacteria was collected, not necessarily the first dates in the dataset. Therefore, precipitation data should be gathered from 7 days prior to the first date for which FIB bacteria was collected.

The number of previous dry days is a different variable that was derived by counting the number of previous days, before a given date, with rain less than 0.05 inches. The precipitation column was visually consulted to count the previous dry days and this number was manually recorded for each day within the research assistant’s spreadsheet.

CUMULATIVE VARIABLES

Cumulative rain amounts were also calculated for both case studies. The following cumulative rain variables were created by using the add function in Excel to find the sum of a given day’s precipitation and a certain number of days before it. Day 0 represents the precipitation on a given day.

1. **Cumulative Rain Day 0 and 1 day previous**
2. **Cumulative Rain Days 0 through 2 days previous**
3. **Cumulative Rain Days 0 through 3 days previous**
4. **Cumulative Rain Days 0 through 4 days previous**
5. **Cumulative Rain Days 0 through 5 days previous**
6. **Cumulative Rain Days 0 through 6 days previous**
7. **Cumulative Rain Days 0 through 7 days previous**

The procedure to calculate any of these variables is based off the following example. To calculate Cumulative Rain Days 0 through 3 days previous: (1) Use the add function in Excel to enter the sum of the precipitation for a given day, 1-day previous precipitation for that same day, 2-day previous precipitation and 3-days previous precipitation in the function box for the first data row that has a value for 3-days previous precipitation. Alternatively, the sum of "cumulative rain days 0 through 2 days previous" and 3 days previous precipitation could be entered into the function box since both summation values are equivalent. (2) Then drag the right bottom corner to fill the remaining rows once the calculation has been made. Similar calculations can be performed for all the cumulative rain variables.

OTHER CALCULATIONS

1. Time to Low Tide (for each station)

This variable can only be calculated once AM and PM low tides and the bacteria stations' sample times have been recorded. In order to perform calculations with time, make sure AM and PM times are clearly indicated in the spreadsheet with the necessary "AM" or "PM" after the time. All PM times must be written as either 1:00 PM or in military time, 13:00.

The following procedure was used to calculate "time between collection and low tide (station 1)" for both case studies. It was repeated for stations 2 and 3 as well. For the Easton's beach case study, the "time between collection and low tide (station 2 - COA)" variable was also created using the same procedure but using the sample times from COA bacteria collection not RIDOH since COA only collects at station 2 on Easton's beach.

(1) Create a new column titled "time between collection and AM Low Tide Station 1." (2) Click the second cell of this column. (3) In the function box type: =IF(ISBLANK(station 1 sample time),"",ABS($AM Low Tide - station 1 sample time)). Here, the second cell for each variable in the if statement should be clicked. (4) Hit enter. (5) Drag the bottom right corner of this cell (now with the time difference between the sample time and morning low tide time) down the remainder of the column until the column is filled with values. (6) Now create a new column titled "time between collection and PM Low Tide Station 1" and repeat the previous procedure but instead of using the AM Low Tide column for the calculation IF statement, use the PM Low Tide column. (7) Create a new column titled "Minimum Time Between Sample and Low Tide Station 1". (8) Click the second cell of this new column and in the function box type: =IF(ISBLANK(station 1 sample time),"",MIN(Time Between Collection and AM Low Tide Station 1:Time Between Collection and PM Low Tide Station 1)) This will give you the shortest time between the sample time and a low tide for that day. (9) Drag the bottom right corner of the cell (now with a value in it) down the remainder of the column until it is filled in. (10) Highlight this whole column, right click on it and click "copy." 11. Now create a new column titled "time to low 1". (12) Highlight this whole, new column and paste the values from the "Minimum Time Between Sample and Low Tide Station 1" column. (13) Highlight the "time to low 1" column, right click on it and select "Format Cells". It should already be on the "time" tab but now click the "number" tab on the left side. (14) In the decimal places box type "4." (15) Click "Ok." Now the whole column's values, originally times, should be replaced with number values.

2. Wind Degree

For the Oakland beach case study, wind direction data (from WU) was converted to wind degrees using the following procedure.

(1) Create a new column called "wind degree." (2) Click the row of this column that corresponds to the first day of wind direction data. (3) Type the following IF statement into the function box but replace "WIND DIRECTION" with the first cell in the wind direction column in the input spreadsheet:

=IF(WIND DIRECTION="S",180,IF(WIND DIRECTION="SSW",202.5,IF(WIND DIRECTION="SW",225,IF(WIND DIRECTION="WSW",247.5,IF(WIND DIRECTION="W",270,IF(WIND DIRECTION="WNW",292.5,IF(WIND DIRECTION="NW",315,IF(WIND DIRECTION="NNW",337.5,IF(WIND DIRECTION="N",0,IF(WIND DIRECTION="NNE",22.5,IF(WIND DIRECTION="NE",45,IF(WIND DIRECTION="ENE",67.5,IF(WIND

DIRECTION="E",90,IF(WIND DIRECTION="ESE",112.5,IF(WIND DIRECTION="SE",135,IF(WIND DIRECTION="SSE",157.5)))))))))))))))

(4) Hit enter. (5) Drag the right bottom corner of the now completed cell down the rest of the column to fill in the rest of the rows with wind degree data.

3. Change in Tide Height

This is the difference between high and low tide heights. The subtraction: high tide height – low tide height can be performed using the difference function in Excel once the high tide height and low tide height columns are populated.

4. Difference in Temperature

This is the difference between air temperature and water temperature on a particular day. The subtraction: water temperature – air temperature can be performed using the difference function in Excel once the air and water temperature columns are populated.

5. Time Between Tides

This is the difference between am high and low tide times. Using the difference function in Excel the time between tides can be calculated by subtracting the earlier tide time from the later tide time.

6. Groundwater Indicator

For the Oakland beach case study, the binary variable "groundwater indicator" was created to identify the days for which the groundwater depth of the Warwick Neck area was greater than 10 feet. After populating the groundwater depth column with data, a column was created for groundwater indicator and the following if statement was used to populate the groundwater indicator column.

If(GROUNDWATER DEPTH COLUMN > 10, 1, 0)

Here, if the groundwater depth was greater than 10 on a given day then the indicator was 1. If less than or equal to 10, the indicator was 0.

RESPONSE VARIABLES

The response variables below are the bacteria samples that were collected at 3 stations on Oakland beach and 3 stations on Easton's beach by RIDOH. The variables for each stations' bacteria results are named "Station 1 Results," "Station 2 Results," and "Station 3 Results" at both beaches. The "Clean Ocean Access Station 2 Bacteria" variable consists of bacteria samples collected by COA at station 2 on Easton's beach. Both the RIDOH and COA samples from station 2 were looked at for the Easton's beach case study.

1. Station 1 (Lab 5) Results
2. Station 2 (Lab 5) Results
3. Station 3 (Lab 5) Results
4. Clean Ocean Access Station 2 Bacteria**

* indicates variables specific to Oakland Beach case study

** indicates variables specific to Easton's Beach case study

RIDOH and COA collect bacteria samples at Oakland and Easton's beaches and send in the water samples, along with the coastal survey sheets that include sample time information, to RIDOH who then records sample information in chain of custody forms. All sample information for both beaches and from both sources, RIDOH and COA, is then documented in an Excel spreadsheet by RIDOH. From this excel spreadsheet, bacteria data for each station is then manually recorded into the Virtual Beach input spreadsheet in 3 different columns (one for each station).

The Clean Ocean Access Station 2 Bacteria was derived from a spreadsheet that COA employees and Sherry Poucher from RIDOH update every year with sample information collected only by COA. The Easton's beach bacteria data is copied and pasted from this spreadsheet into the Virtual Beach input file.

Alternatively, bacteria data can be found on the EPA's website: https://watersgeo.epa.gov/BEACON2/ by: (1) Clicking the link. (2) Searching for the desired beach, stream or other type of station. (3) Selecting "water quality" data. (4) Inputting the desired dates for which bacteria sample data is sought.

The bacteria data can be quality checked by reviewing the original excel spreadsheets they came from or the EPA link provided above.

Appendix 2 – Oakland Beach

2A – Variables and Acronyms used for modeling

Appendix A. Variables	
Primary Source Variables	**Derived Variables**
0.103896104	Number of Previous Dry Days
Rain_0 *	Rain_Day-1
AM - High Time	Rain_Day-2
AM-High Height (ft) *	Rain_Day-3
AM- Low Time	Rain_Day-4
AM-Low Height(ft) *	Rain_Day-5
PM-High Time	Rain_Day-6
PM-High Height (ft)	Rain_Day-7
PM-Low Time	Cumrain_0-1
PM-Low Height	Cumrain_0-2
Water Temp (F) *	Cumrain_0-3
Air Temp (F) *	Cumrain_0-4
Wind Direction *	Cumrain_0-5
Wind Speed (mph) *	Cumrain_0-6
Weather-S **	Cumrain_0-7
Surf-S **	Time btw high and low tide
Tide-S **	Prevday time btw high and low tide
Vegetation in Water-S **	Time btw collection and low tide
Visible Sewage or Odor-S **	Change in Tide Height (high minus low)
Condition of Water-S **	Prevday change in tide height
Seagull count-S*	PrevdayAirtemp
Clear Sky UV Index *	winddegree
Prev day clear sky UV *	prevday winddegree
Cloudy Sky UV Index *	Prevday wind speed
Cloud Transmission *	Wind A
Solar Zenith Angle *	Wind O
Aerosol Transmission *	Prevday Wind A
Total Column Zone *	Prevday Wind O
Collection Time	Water temp - air temp
	Groundwater depth-Interpolated
	Groundwater indicator (> than 10 feet deep)
	Hunt River flow
	Prevday hunt flow
	Hunt River Tide Gage Height
	Predvay hunt river tide guage
	Prev day cloudy UV

2B - Summary tables of model output in sequential order of model runs

2B1. Without transformations or interactions (2015-2017)

Model	GBM	Adj R Squared	AICC	Bay
Decision Criteria		100	100	100
Reg. Standard		60	60	60
True Pos	4			
True Neg	55			
False Pos	13	2	1	3
False Neg	5	19	17	19
R Squared		0.4291	0.4154	0.3568
Adj R2		0.4137	0.3996	0.3394
AkAICe Info		112.718	110.5382	113.9015
Corr. AIK		114.718	111.5382	113.9015
Bayesian		52.4685	45.6011	44.2767
PRESS		127.1797	118.8745	124.1905
RMSE		1.1939	1.1907	1.2315
ROC Decision Criterion		3.8883	3.7832	4.7141
False Non-Exceed		8	6	19
False Exceed		6	8	1
Variable 1	TBCL	AMLH	TBCL	CR_1
Variable 2	RAIN	P_RD1	P_RD4	P_RD4
Variable 3	CR_6	GWI	CR_1	TBCL
Variable 4	AMLH	P_RD4	AMLH	AMLH
Variable 5	CR_7	RAIN	GWI	N/A

2B2. With interactions (2015-2017)

Model	Adj R Squared	AICC	Bay
Decision Criteria	100	100	100
Reg. Standard	60	60	60
False Pos	1	1	2
False Neg	16	16	18
R Squared	0.5377	0.4896	0.4601
Adj R2	0.5252	0.4758	0.4455
Akaike Info	96.4694	98.088	100.415
Corr. AIK	98.4694	99.088	100.415
Bayesian	36.2198	30.807	30.7902
PRESS	111.3609	102.2489	108.0336
RMSE	1.0743	1.1047	1.1283
ROC Decision Criterion	4.5895	4.4796	4.743
False Non-Exceed	16	15	18
False Exceed	1	1	1
Variable 1	PROD(P_WT,WTMAT)	PROD(RAIN,P_RD4)	PROD(P_RD4,GWI)
Variable 2	PROD(P_RD4,WD)	PROD(TBCL,GWI)	PROD(TBCL, GWI)
Variable 3	Prod (P_RD1,WT)	PROD(WS,WT-AT)	PROD(WTMAT,GWI)
Variable 4	PROD (TBCL,GWI)	PROD(P_RD4, GWI)	P_RD2
Variable 5	AMLH	PROD(WD,GWI)	N/A

2B3. With transformations and interactions (2015-2017)

Model	dj R Square	AICC	Bay	PRESS	RMSE	Sensitivity	Specificity	Accuracy
Decision Criteria	100	100	100	100	100	100	100	100
Reg. Standard	60	60	60	60	60	60	60	60
False Pos	2	1	2	1	0	1	0	0
False Neg	18	20	19	20	19	14	20	11
R Squared	0.5565	0.4989	0.5524	0.5116	0.5235	0.3499	0.1084	0.4289
Adj R2	0.5445	0.4823	0.5403	0.4984	0.5106	0.3324	0.0843	0.4135
Akaike Info	93.2702	99.1269	91.9803	98.6919	96.7946	114.7163	135.0439	108.7385
Corr. AIK	95.2702	100.1269	93.9803	100.6919	98.7946	114.7163	135.0437	109.7385
Bayesian	33.0206	34.1897	29.387	36.0985	34.2012	45.0916	60.7315	43.8013
PRESS	96.27	123.0359	102.8919	97.7704	124.4764	132.2676	162.5568	124.4971
RMSE	1.0522	1.1056	1.0494	1.0962	1.0827	1.2381	1.4302	1.1768
ROC Decision Criterion	4.0884	3.7818	3.9259	3.7	3.9408	4.3082	4.1572	4.5999
False Non-	11	8	8	7	9	14	20	11
False Exceed	4	7	6	8	7	1	0	0
Variable 1	POLY[PROD[P_WT,WD]	PROD[TBCL,CR_7]	AMLH	POLY[WT]	POLY[PRO	POLY[PROD[RAIN,GWI]	PROD[P_WT,GWI]	POLY[WT]
Variable 2	PROD[P_RD4,GWI]	PROD [P_RD4,CR_7]	PROD[RAIN,P_RD4]	PROD[CR_4,TBCL]	POLY[PROD[P_RD1,P_WT[	PROD[TBCL,GWI]	POLY[PROD[CR_6,WD]	POLY[PROD[WT,WS]
Variable 3	INVERSE[PROD[CR_6,CLTR]	POLY[PROD[P_WT,WD]	POLY[PROD[WT,P_WT]	PROD[P_RD4,GWI]	POLY[PROD[TBCL,WD[	POLY[PROD[CR_6,AMLT]	N/A	POLY[PROD[P-WT,WD]
Variable 4	AMLH	AMLH	PROD[P_RD1,WT-AT]	AMLH	P_CITH	PROD[P_RD1,CR_1]	N/A	PROD[RAIN,CLSUV]
Variable 5	PROD[TBCL,GWI]	POLY[PROD[P_RD1,WT]	POLY[PROD[TBCL,WD]	POLY[PROD[RAIN,WD]	PROD[P_RD4,CR_4]	N/A	N/A	P_RD2

2B4. Interactions and transformations (2015 & 2017 only)

Model	dj R Square	AICC	Bay	PRESS	RMSE	Sensitivity	Specificity	Accuracy
Decision Criteria	100	100	100	100	100	100	100	100
Reg. Standard	60	60	60	60	60	60	60	60
False Pos	1	0	0	0	0	1	0	0
False Neg	10	13	11	11	13	8	16	8
R Squared	0.6075	0.5906	0.5573	0.5415	0.5923	0.3585	0.2186	0.4231
Adj R2	0.5927	0.5752	0.5406	0.5242	0.5769	0.3343	0.1891	0.4013
AkAICe Info	67.8616	68.2136	70.5986	72.5636	67.9855	91.3703	98.4176	87.4271
Corr. AIC	69.8616	69.2136	71.5986	73.5636	68.9855	92.3703	98.4176	88.4271
Bayesian	22.0137	20.3404	20.7	22.665	20.1152	41.4717	44.4683	39.5539
PRESS	65.8347	74.6971	71.4433	71.0416	65.8805	104.9963	122.0843	107.0819
RMSE	1.0488	1.0603	1.0918	1.1111	1.0582	1.3143	1.4229	1.2588
ROC Decision Criterion	4.806	4.558	4.6483	4.545	4.6116	4.9285	4.4551	4.5521
False Non-Exceed	12	13	11	11	13	11	15	8
False Exceed	0	0	0	0	0	0	0	0
Variable 1	POLY[WS]	POLY[WS]	POLY[WS]	POLY[PROD[CR_1,RAIN]	POLY[WS]	POLY[WS]	PROD[CR_1,TBCL]	P_WS
Variable 2	PROD[P_RD4,CR_1]	PROD[P_RD4,RAIN]	PROD[P_RD4,RAIN]	PROD[P_HRG,AMLH]	PROD[CR_1,TBCL]	PROD[P_HRG,CR_7]	POLY[P_WD]	PROD[TBCL,CR_7]
Variable 3	POLY[P_AT]	PROD[P_HRG,AMLH]	PROD[TBCL,CR_6]	PROD[P_RD4,WT]	POLY[P_AT]	PROD[TBCL,P_HRG]	N/A	PROD[P_HRG,AMLH]
Variable 4	PROD[CR_1,TBCL]	POLY[P_AT]	PROD[P_HRG,AMLH]	PROD[TBCL,CR_7]	PROD[P_HRG,AMLH]	PROD[GWI,AMLH]	N/A	POLY[P_RD1]
Variable 5	[RROD[P_HRG,AMLH]	PROD[CR_1,TBCL]	N/A	N/A	PROD[P_RD4,CR_1]	N/A	N/A	GWI

2B5. Interactions and transformations (2015 & 2017 only) with prior Enterococcus

Model	dj R Square	AICC	Bay	PRESS	RMSE	Sensitivity	Specificity	Accuracy
Decision Criteria	100	100	100	100	100	100	100	100
Reg. Standard	60	60	60	60	60	60	60	60
False Pos	0	1	0	0	1	0	0	0
False Neg	14	15	17	16	14	12	23	12
R Squared	0.6108	0.5649	0.5019	0.542	0.5807	0.379	0.0825	0.4344
Adj R2	0.6002	0.5532	0.4885	0.5296	0.5694	0.3623	0.0577	0.4191
Akaike Info	83.2226	89.7967	96.2058	93.7491	86.9504	117.187	137.2497	111.9997
Corr. AIC	85.2226	91.7967	97.2058	95.7491	88.9504	119.187	137.2497	113.9997
Bayesian	22.9731	27.2033	28.8249	31.1558	24.357	54.5936	62.9373	51.7501
PRESS	85.0186	100.7962	99.1146	90.3427	87.2144	129.0668	171.1963	122.6345
RMSE	0.9585	1.0346	1.0913	1.0615	1.0157	1.2336	1.4508	1.1883
ROC Decision Criterion	4.6243	5.0665	4.5854	4.5965	4.9228	4.8858	4.4275	4.3865
False Non-Exceed	14	18	16	16	19	18	20	12
False Exceed	0	0	0	0	0	0	0	0
Variable 1	POLY[PROD[P_ENT1_D,RAIN]	PROD[P_RD4,GWI]	PROD[TBCL,CR_6]	Log10[PROD[P_ENT1,AMLH]	PROD[GWI,TBCL]	POLY[P_WT]	P-CLTR	PROD[P_RD1,CR_6]
Variable 2	POLY[PROD[GWI,WD]	WindA_Comp[WD,WS]	PROD[AMLH,P_HRG]	POLY[PROD[P_ENT1_D,GWI]	PROD[GWI,AMLH]	PROD[CR_6,P_HRG]	PROD[CR_7,P_HRG]	WUADROOT[PROD[P_ENT1,TBCL]]
Variable 3	PROD[GWI,TBCL]	PROD[P_WT,CR_1]	PROD[P_RD4,RAIN]	PROD[P_RD4,CR_1]	WindA_comp[WD,WS]	PROD[TBCL,P_HRG]	N/A	WindA_comp[WD,WS]
Variable 4	PROD[P_RD4,P_HRG]	PROD[GWI,TBCL]	POLY[PROD[WD,P_WT]	PROD[P_RD1,TBCL]	PROD[CR_1,CR_6]	POLY[WD]	N/A	Log10{PROD[P_ENT1,AMLH[
Variable 5	POLY[WS]	P_CLSUV	POLY[WS]	POLY[PROD[GWI,RAIN]	PROD[P_RD4,GWI]	INVERSE[P_ENT1,5]	N/A	INVERSE[WindO_comp[WD,WS]

2B6. Plus prior Enterococcus no transformations or interactions

Model	GBM	Adjusted R	AICC	Bay	PRESS	RMSE	Sensitivity	Specificity	Accuracy
Decision Criteria	100	100	100	100	100	100	100	100	100
Reg. Standard		60	60	60	60	60	60	60	60
True Pos	42								
True Neg	15								
False Pos	12	2	2	2	1	2	1	0	0
False Neg	8	16	16	18	19	16	12	24	14
R Squared		0.4721	0.4544	0.4317	0.445	0.4707	0.3807	0.0441	0.3278
Adj R2		0.4579	0.4396	0.4164	0.43	0.4564	0.364	0.0183	0.3097
Akaike Info		104.6787	105.2322	106.3577	106.5433	104.8832	116.9783	140.4041	119.2908
Corr. AIC		106.6787	106.2322	107.3577	107.5433	106.8832	118.9783	140.4041	120.2908
Bayesian		42.0854	40.295	39.0767	41.6062	42.2898	54.385	66.0917	52.0098
PRESS		110.1056	112.8713	115.0223	112.401	111.236	132.6847	171.6013	137.4357
RMSE		1.1396	1.1504	1.1656	1.1602	1.1411	1.2344	1.4809	1.2678
ROC Decision Criterion		5.3733	5.1471	5.0995	4.9424	5.056	4.9167	3.0576	4.5886
False Non-Exceed		20	20	20	19	18	16	49	14
False Exceed		0	0	0	0	0	0	0	0
Variable 1	TBCL(13.5597)	P_RD4	TBCL	AMLH	TBCL	TBCL	TBCL	P_RD5	WT-AT
Variable 2	WindA_comp(8.6007)	CR_1	CR_1	TBCL	P_RD7	AMLH	WindA_comp[WD,WS]	CITH	WindA_comp[WD,WS]
Variable 3	Rain(6.0393)	WindA_comp (WD,WS)	WindA_comp (WD,WS)	WindA_Comp(WD,WS)	CR_1	WindA_comp (WD,WS)	CR_6	N/A	CR_4
Variable 4	AMLH(5.0952)	AMLH	AT	P_RD4	WindA_comp (WD,WS)	CR_1	AT	N/A	TBCL
Variable 5	CR_6(4.7275)	AT	AMLH	CR_1	AMLH	P_RD4	P_RD3	N/A	P_RD6

2B7. Plus Prior Enterococcus Without Time Between Collection and Low Tide (TBCL) and Changing Decision Criteria Based on ROC Optimization

Model	Adj R Squa	AICC	Bay	PRESS	RMSE	Sensitivity	Specificity	Accuracy
Decision Criteria	100	100	100	100	100	100	100	100
Reg. Standard	60	60	60	60	60	60	60	60
False Pos	0	1	1	1	0	1	0	0
False Neg	17	15	17	19	17	13	24	13
R Squared	0.5538	0.5215	0.54	0.5191	0.5586	0.3687	0.0958	0.4604
Adj R2	0.5418	0.5086	0.5276	0.5061	0.5467	0.3516	0.0714	0.4458
Akaike Info	93.7341	95.1147	92.0793	97.5118	90.9	116.463	134.1212	104.3728
Corr. AIC	95.7341	96.1147	93.0793	99.5118	92.9	117.463	134.1212	105.3728
Bayesian	33.4846	30.1775	27.1421	34.9184	28.3066	51.5259	57.465	39.4357
PRESS	90.8137	92.6529	95.3071	92.6111	92.7219	130.0045	160.9409	114.2778
RMSE	1.0554	1.0772	1.0562	1.0878	1.0421	1.2374	1.4306	1.144
ROC Decision Criterion	4.7017	4.7422	4.5547	4.9056	4.5384	5.2047	2.5986	4.4065
False Non-Exceed	18	18	16	20	17	20	50	13
False Exceed	0	0	0	0	0	0	0	0
Variable 1	POLY[PROD[GWI,P_RD1]	PROD[AMLH,P_HRG]	PROD[P_RD4,GWI]	PROD[P_RD4,AMLH]	PROD[P_RD4,RAIN]	PROD[P_ENT1,P_RD4]	POLY[PROD[P_ENT1_D,GWI]	PROD[AMLH,WD]
Variable 2	POLY[PROD[GWI,RAIN]	POLY[PROD[P_ENT1_D,P_RD1]	POLY[PROD[GWI,P_RD1]	POLY[PROD[P_WT,RAIN]	PROD[P_RD4,AMLH]	POLY[PROD[GWI,RAIN]	N/A	POLY[WT]
Variable 3	PROD[AMLH,P_HRG]	POLY[PROD[GWI,RAIN]	POLY[PROD[GWI,RAIN]	POLY[PROD[WD,AT]	POLY[PROD[GWI,P_RD1]	POLY[WS]	N/A	POLY[PROD[GWI,RAIN]
Variable 4	POLY[PROD[GWI,WD]	PROD[P_RD4,GWI]	PROD[AMLH,P_HRG]	PROD[P_RD4,CR_1]	POLY[PROD[GWI,RAIN]	PROD[P_ENT1,WD]	N/A	PROD[P_ENT1,P_RD4]
Variable 5	PROD[P_RD4,P_HRG]	PROD[P_RD4,GWI]	WindA_comp[WD,WS]	PROD[AMLH,P_HRG]	POLY[AT]	PROD[CR_6,P_HRG]	N/A	INVERSE[WindO_comp[WD,WS]

2B8. GBM Runs with Top 5 variables

Model	GBM	GBM	GBM	GBM	GBM	GBM
Reg. Standard	235	235	100	100	60	60
Decision Criteria	243.97	49.58	243.97	47.32	243.97	47.32
True Pos	4	3	6	4	6	5
True Neg	64	66	54	55	51	52
False Pos	4	2	3	2	2	1
False Neg	5	6	15	16	18	19
Sensitivity	0.44	0.33	0.25	0.2	0.25	0.21
Specificity	0.94	0.97	0.95	0.96	0.96	0.98
Accuracy	0.88	0.9	0.77	0.77	0.74	0.74
Variable 1	CR_4(30.3227)	CR_7(32.5205)	CR_4(30.3227)	CR_7(32.5205)	CR_4(30.3227)	CR_7(32.5205)
Variable 2	CR_7(20.3810)	CR_6(29.9825)	CR_7(20.3810)	CR_6(29.9825)	CR_7(20.3810)	CR_6(29.9825)
Variable 3	CR_6(17.5724)	PTCZ(18.5897)	CR_6(17.5724)	PTCZ(18.5897)	CR_6(17.5724)	PTCZ(18.5897)
Variable 4	PTCZ(15.9504)	CR_4(16.8292)	PTCZ(15.9504)	CR_4(16.8292)	PTCZ(15.9504)	CR_4(16.8292)
Variable 5	P_RD4(15.7734)	P_RD4(2.0780)	P_RD4(15.7734)	P_RD4(2.0780)	P_RD4(15.7734)	P_RD4(2.0780)

2B9. GBM Runs with Top 8 Variables

Model	GBM	GBM	GBM	GBM	GBM	GBM
Reg. Standard	235	235	100	100	60	60
Decision Criteria	259.56	83.07	259.56	60.94	259.56	55.41
True Pos	4	3	5	4	6	5
True Neg	67	66	56	55	51	52
False Pos	1	2	1	2	2	1
False Neg	5	6	15	16	18	19
Sensitivity	0.44	0.33	0.25	0.2	0.25	0.21
Specificity	0.99	0.97	0.98	0.96	0.96	0.98
Accuracy	0.92	0.9	0.79	0.77	0.74	0.74
Variable 1	CR_4(25.3511)	AMLH(19.5273)	CR_4(25.3511)	AMLH(19.5273)	CR_4(25.3511)	AMLH(19.5273)
Variable 2	CR_6(15.4961)	WindA_comp(16.0910)	CR_6(15.4961)	WindA_comp(16.0910)	CR_6(15.4961)	WindA_comp(16.0910)
Variable 3	P_RD4(14.2602)	RAIN (15.5288)	P_RD4(14.2602)	RAIN (15.5288)	P_RD4(14.2602)	RAIN (15.5288)
Variable 4	CR_7(14.2457)	PTCZ(14.0757)	CR_7(14.2457)	PTCZ(14.0757)	CR_7(14.2457)	PTCZ(14.0757)
Variable 5	RAIN(12.6187)	CR_7(14.0757)	RAIN(12.6187)	CR_7(14.0757)	RAIN(12.6187)	CR_7(14.0757)
Variable 6	PTCZ(11.6232)	CR_6(11.4467)	PTCZ(11.6232)	CR_6(11.4467)	PTCZ(11.6232)	CR_6(11.4467)
Variable 7	AMLH(3.8670)	CR_4(8.3494)	AMLH(3.8670)	CR_4(8.3494)	AMLH(3.8670)	CR_4(8.3494)
Variable 8	WindA_comp(2.5381)	P_RD4(0.9696)	WindA_comp(2.5381)	P_RD4(0.9696)	WindA_comp(2.5381)	P_RD4(0.9696)

2B10. GBM Runs Without Set Seed

Model	GBM	GBM	GBM	GBM	GBM	GBM	GBM
Reg. Standard	60	60	60	60	60	60	60
Decision Criteria	44.42	49.35	48.04	48.67	46.32	47.32	46.5
True Pos	3	4	3	7	6	5	5
True Neg	47	49	47	47	46	47	48
False Pos	6	4	6	6	7	6	5
False Neg	21	20	21	17	18	19	19
Sensitivity	0.13	0.17	0.13	0.29	0.25	0.21	0.21
Specificity	0.89	0.92	0.89	0.89	0.87	0.89	0.91
Accuracy	0.65	0.69	0.65	0.7	0.68	0.68	0.69
Variable 1	WindA_comp(8.3157)	WindA_comp(8.8422)	RAIN(9.0480)	WindA_comp(9.1284)	WindA_comp(8.5745)	RAIN(8.1742)	WindA_comp(9.3306)
Variable 2	RAIN(8.0661)	CR_6(7.3726)	WindA_comp(8.0745)	RAIN(8.1558)	CR_6(8.5551)	WindA_comp(8.0637)	CR_6(7.9467)
Variable 3	CR_6(7.7208)	RAIN(7.1894)	AMLH(7.9058)	CR_6(7.6768)	RAIN(8.1452)	AMLH(7.6797)	RAIN(7.6917)
Variable 4	AMLH(6.8916)	AMLH(6.2457)	CR_6(6.6425)	AMLH(7.1968)	AMLH(6.1272)	CR_6(6.4928)	AMLH (6.6569)
Variable 5	CR_7(4.8309)	CR_7(4.9411)	CR_7(5.2755)	PTCZ(4.7148)	PTCZ(5.1118)	CR_7(5.0918)	PTCZ(4.5649)

2B11. With transformations and interactions but untransformed ENT

Model	BAY	Adjusted R2	AICC
Decision Criteria	300	300	300
Reg. Standard	300	300	300
False Pos	0	0	0
False Neg	4	4	4
R Squared	0.8664	0.8836	0.8779
Adj R2	0.8628	0.8804	0.8746
Akaike Info	859.4589	856.8768	856.5077
Corr. AIC	859.4589	858.8768	857.5077
Bayesian	787.4903	794.2834	789.2268
PRESS	5668992	5962597.646	5461027.546
RMSE	156.9029	150.6694	152.0758
Model Fit	787.4903	0.8804	856.5077
AUC	0.59298	0.733532	0.708767
Variable 1	POLY[P_RD1] (0.2856)	P_RD3 (-0.0651)	PROD[P_RD4,GWI] (0.1367)
Variable 2	PROD[P_RD4,GWI]	PROD[RAIN,P_RD4] (0.8083)	POLY[PROD[AMLH,P_HRG] (0.0935)
Variable 3	PROD[RAIN,P_RD4] (0.8032)	POLY[WS](0.0682)	PROD[RAIN,P_RD4] (0.8046)
Variable 4	N/A	POLY[P_RD1] (0.2253)	POLY[P_RD1](0.2377)
Variable 5		POLY[P_WT]	POLY[WS](0.0754)

Appendix 3 – Easton's Beach

3A – Variables and Acronyms Used for Modeling

Abbreviation in Data Files/VB Outputs	Long Variable Name
Date	Date
rain0	Precipitation (on given date)
priordry	Prior Number of Dry Days (on given date)
PrevPriorDry	Prior Number of Dry Days (on previous date)
rain1	Previous Rain from 1 day ago
rain2	Previous Rain from 2 days ago
rain3	Previous Rain from 3 days ago
rain4	Previous Rain from 4 days ago
rain5	Previous Rain from 5 days ago
rain6	Previous Rain from 6 days ago
rain7	Previous Rain from 7 days ago
cum1	Cumulative Rain 1 Day
cum2	Cumulative Rain 2 Days
cum3	Cumulative Rain 3 Days
cum4	Cumulative Rain 4 Days
cum5	Cumulative Rain 5 Days
cum6	Cumulative Rain 6 Days
cum7	Cumulative Rain 7 Days
watertempCOA	Water Temperature (collected by COA)
watertempDOH	Water Temperature (collected by RIDOH)
watertemp	Water Temperature (gathered from other sources)
PrevWaterTemp	Water Temperature on previous day
airtemp	Air Temperature
PrevAirTemp	Air Temperature on previous day
tempchangeCOA	Temperature Change (difference in air and water temperature)
tempchangeDOH	Temperature Change (difference in air and water temperature)
tempchangeNOAA	Temperature Change (difference in air and water temperature)
PrevTempChange	Temperature change on previous day
hightideheight	High Tide Height
lowtideheight	Low Tide Height
tidediff	Difference between high and low tide height
PrevHTH	Previous day's high tide height
PrevLTH	Previous day's low tide height
PrevTideDiff	Previous day's difference between high and low tide height
ent1center	Bacteria at Easton's beach center station (station 1) collected by RIDOH
ent2eastCOA	Bacteria at Easton's beach east station (station 2) collected by COA
DaysBtwnEnt2COA	Days between ent2eastCOA samples collected
Ent2COAPriorConc	Last Easton's beach east station sample collected (by COA)
ent2east	Bacteria at Easton's beach east station (station 2) collected by RIDOH
ent3west	Bacteria at Easton's beach west station (station 3) collected by RIDOH
EastonsPoint	Easton's Point bacteria
DaysBtwnPoint	Days between Easton's Point samples collected
PointPriorConc	Last Easton's Point sample collected
EsplanadeSouth	South Esplanade bacteria
DaysBtwnEsp	Days between South Esplanade samples collected
EspPriorConc	Last South Esplanade sample collected
EastonsStreamCOA	Easton's Stream bacteria (collected by COA)
DaysBtwnStream	Days between Easton's Stream samples collected
StreamPriorConc	Last Easton's Stream sample collected

Stream/days	StreamPriorConc divided by days between Easton's Stream samples collected
EastonsStreamDOH	Easton's Stream bacteria (collected by RIDOH)
windspeed	Wind Speed
PrevWindSpeed	Previous day wind speed
wind_deg	Wind Degree
PrevWindDeg	Previous day wind degree
timetolow1	Time between collection and low tide (station 1 - DOH)
timetolow2	Time between collection and low tide (station 2 - DOH)
timetolow2COA	Time between collection and low tide (station 2 - COA)
timetolow3	Time between collection and low tide (station 3 - DOH)
Weather	Weather Conditions
weathercat	Weather Conditions Category
Surf	Surf Conditions
surfcat	Surf Conditions Category
Tide	Tide Conditions
tidecat	Tide Conditions Category
vegetation	Vegetation Category
VisibleSewageOdor	Visible Sewage or Odor
sewageodorcat	Visible Sewage or Odor Category
ConditionofWater	Water Condition
waterconditioncat	Water Condition Category
Seagulls	Seagulls (# of)
clearskyuv	Clear Sky UV
PrevCSUV	Previous day Clear Sky UV
cloudUV	Cloud UV
PrevCloudUV	Previous day Cloud UV
cloudtransmission	Cloud Transmission
PrevCTrans	Previous day cloud transmission
solarzenithangle	Solar Zenith Angle
PrevSZA	Previous day Solar Zenith Angle
aerosoltransmission	Aerosol Transmission
PrevAerTrans	Previous day Aerosol Transmission
totalcolumnzone	Total Column Zone
PrevTCZ	Previous Day Total Column Zone
UVDisinfection	UV Disinfection Light (turned on or off)
PrevUVD	Previous Day UV Disinfection
timebetweentides	Time Between Tides
PrevTBT	Previous Time Between Tides

3B1. Outputs from Modeling Exercises (Clean Ocean Access Data)

Appendix B: Summary tables of model output in sequential order of model runs

B1. Clean Ocean Access Data Model Runs (2015-2017)

Model	MLR	MLR	MLR	MLR	MLR	MLR	MLR	MLR	MLR	GBM	MLR
Decision Criteria	100	100	100	100	100	100	100	100	100	100	70
Reg. Standard	60	60	60	60	60	60	60	60	60	60	60
transformations	Yes - dep and indep	Yes - dep and indep	yes - dep and indep	yes - dep and indep	yes - dep and indep	yes - dep only	yes - dep and indep	yes - dep and indep	yes - dep and indep	no	yes - dep and indep
Interactions	no	no	yes	yes	no	yes	yes	yes	yes	no	yes
Criteria	corr akaike	corr akaike	corr akaike	adj r square	adj r square	adj r square	adj r square	adj r square	corr akaike	n/a	corr akaike
Initial	SHP	SHP	SHP	SHP	SHP	SHP	SHP	SHP	SHP	SHP	SHP
True Pos	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	4	hard to count	17	7
True Neg	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	hard to count	hard to count	76	hard to count
False Pos	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	0	0	32	0
False Neg	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	31	31	19	27
R Squared	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	0.41	0.41	n/a	0.38
Adj R2	0.3374	0.3529	0.3861	0.4168	0.36	0.2312	0.419	0.3979	0.4026	n/a	0.3683
Akaike Info	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	142.0066	136.8666	n/a	142.8978
Corr. AIK	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	144.01	138.87	n/a	143.9
Bayesian	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	31.6444	20.5647	n/a	23.6262
PRESS	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	154.9068	136.874	n/a	141.6164
RMSE	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	0.9513	0.9404	n/a	0.9634
Variable 1	poly rain 5	poly rain 5	not recorded	not recorded	not recorded	not recorded	not recorded	Inverse Prod (Prev LTH, wind A)	Poly Prod (PrevWindDeg, Prev TBT)	see case study report	see case study report
Variable 2	poly prev air trans	inverse prev LTH	not recorded	not recorded	not recorded	not recorded	not recorded	Poly Prod (Prev Wind Deg, Prev TBT)	Log10(Prod[timetolow, Prior Stream/days])	see case study report	see case study report
Variable 3	inverse prev LTH	poly TBT	not recorded	not recorded	not recorded	not recorded	not recorded	Poly (rain 2)	Ent2COA PriorConc	see case study report	see case study report
Variable 4	poly prev wind deg	poly cloud trans	not recorded	not recorded	not recorded	not recorded	not recorded	Poly Prod (air temp, cloud trans)	Poly(prod [prev LTH, cloudtrans])	see case study report	see case study report
Variable 5	poly rain 4	prev cloud trans	not recorded	not recorded	not recorded	not recorded	not recorded	Inverse (Prev LTH)	cum3	see case study report	see case study report
Notes	no stream/days or prev water temp. Used10 max variables for model	no stream/days or prev water temp. Used 15 max variables for model	no stream, interacted 8 variables from previous 2 runs	no stream, interacted 8 variables from previous 2 runs	prior stream included	prior stream included	Did not include time between collection and low tide (i.e. time to low) Included prior stream	Did not include time between collection and low tide (i.e. time to low) Included prior stream	Did not include time between collection and low tide (i.e. time to low) Included prior stream	Did not include time between collection and low tide (i.e. time to low) Included prior stream.	Did not include time between collection and low tide (i.e. time to low) Included prior stream
File named:	eastonsfilteredforCOA	eastonsfilteredforCOA	eastonsfilteredforCOA	eastonsfilteredforCOA	eastonsfilteredforCOA	eastonsfilteredforCOA	eastonsfilteredforCOA	eastonsfilteredforCOA	eastonsfilteredforCOA	eastonsfilteredforCOA	eastonsfilteredforCOA

3B2. Outputs from Modeling Exercises (RIDOH data)

B2: Department of Health Model Runns							
Model Type	MLR	MLR	MLR	MLR	MLR	MLR	GBM
Decision Criteria	100	100	100	100	100	100	100
Reg. Standard	60	60	60	60	60	60	60
transformations	yes - dep and indep	yes - dep and indep	yes - dep and indep	yes - dep and indep	yes - dep and indep	yes - dep and indep	no
Interactions	no	yes	yes	yes	yes	yes	no
Criteria	corr akaike	corr akaike	bayesian	corr akaike	corr akaike	corr akaike	n/a
True Pos	not recorded	not recorded	4	3	4	5	14
True Neg	not recorded	not recorded	not recorded	not recorded	not recorded	not recorded	too many to count
False Pos	not recorded	0	0	1	0	0	3
False Neg	not recorded	14	13	6	13		5
R Squared	not recorded	0.59	0.57	0.61	0.6	0.58	n/a
Adj R2	0.54	0.5761	0.5625	0.5936	0.5857	0.5639	n/a
Akaike Info	not recorded	-110.3116	-107.8828	-77.8204	-110.0792	not recorded	n/a
Corr. AIK	not recorded	-109.31	-106.88	-76.82	-109.08	not recorded	n/a
Bayesian	not recorded	-177.5926	-175.1638	-124.7837	-175.0164	not recorded	n/a
PRESS	not recorded	6.7525	6.9554	4.7943	6.9348	not recorded	n/a
RMSE	not recorded	0.2855	0.29	0.284	0.2842	not recorded	n/a
Variable 1	not recorded	Prod Poly (Stream/days)	Prod Poly (Stream/days)	Prod Poly rain 0	Prod Poly rain 6	not recorded	Previous HTH
Variable 2	not recorded	Prod Poly rain 6	Prod poly tide diff	prod poly (stream/days)	Prod Poly Prev CSUV	not recorded	Wind O
Variable 3	not recorded	Prod Poly rain 0	Prod Poly rain 0	Prod Poly rain 4	Prod poly rain 0	not recorded	Stream/Days
Variable 4	not recorded	prev uvd	prev uvd	none	Prev uvd	not recorded	Prev CSUV
Variable 5	not recorded	poly rain 7	poly rain 7	none	poly rain 7	not recorded	Prev Prior dry
Notes	*See VB and screenshot files dated 3/25/19			INCLUDES VEG	*This model included a variable with a bad pvalue but it had the smallest std. coeff. so I didn't put it in top 5 variables		All 12 variables In Case Study
Accuracy	not recorded	0.8182	0.8312	0.8727	0.83	not recorded	n/a
File named:	eastonsfilteredforDOHstation2	eastonsfilteredforDOHstation2	eastonsfilteredforDOHstation2	eastonsfilteredforDOHstation2_veg	eastonsfilteredforDOHstation2	eastonsfilteredforDOHstation2	eastonsfilteredforDOHstation2

3C – Outputs from Model Prediction Step in Virtual Beach

Appendix C: Results from Prediction Step for COA data

ID	Ent2COAPriorConc	PrevAirTemp	PrevLTH	cum3	PrevWaterTemp	timebetweentides	UVDisinfection	PrevWindDeg	cloudtransmission	rain5	watertemp	wind_deg	windspeed	rain4	ID	ent2eastCOA	ID	Model_Prediction	Decision_Criterion	Exceedance_Probability	Regulatory_Standard	Error_Type
1/10/2018 0:00	10	35	0.34	0.01	31.6	0.282639	0	310	100	0	33	350	4.7	0	1/10/2018 0:00	10	1/10/2018 0:00	2.304949516	4.248495	2.405466523	4.094344562	
1/18/2018 0:00	10	33.5	-0.17	0.12	34.5	0.286111	0	130	100	1.1	34.3	320	10.51	0	1/18/2018 0:00	10	1/18/2018 0:00	1.518948768	4.248495	0.374235941	4.094344562	
1/25/2018 0:00	10	34.5	-0.04	1.26	37.6	0.240278	0	280	99.9	0	35.7	300	13.42	0	1/25/2018 0:00	10	1/25/2018 0:00	3.578839785	4.248495	24.7937708	4.094344562	
2/1/2018 0:00	10	26	-0.81	0.12	34.8	0.288194	0	320	83.1	0	35	200	6.04	0.22	2/1/2018 0:00	10	2/1/2018 0:00	3.082498705	4.248495	11.87445256	4.094344562	
2/8/2018 0:00	10	37	0.36	0.96	36.2	0.263889	0	190	99.9	0	35.1	330	10.51	0.6	2/8/2018 0:00	10	2/8/2018 0:00	4.009885587	4.248495	40.5951944	4.094344562	
2/15/2018 0:00	10	38	-0.11	0.32	36.4	0.215278	0	210	78.3	0.37	36.6	200	5.82	1.3	2/15/2018 0:00	40	2/15/2018 0:00	3.26750875	4.248495	16.0669762	4.094344562	
2/22/2018 0:00	40	51	-0.22	0.19	38.8	0.294444	0	220	30.8	0.26	37.4	40	8.5	0.29	2/22/2018 0:00	10	2/22/2018 0:00	2.496855713	4.248495	3.980054954	4.094344562	
3/1/2018 0:00	10	44	-0.43	0.02	37.8	0.215278	0	220	94.7	0.25	37.4	50	6.49	0.87	3/1/2018 0:00	81	3/1/2018 0:00	3.013812316	4.248495	11.17688574	4.094344562	False Negative
3/30/2018 0:00	81	45	-0.25	0.06	39.2	0.275694	0	150	23.6	0	39.7	330	8.5	0	3/30/2018 0:00	10	3/30/2018 0:00	2.777078812	4.248495	6.481649479	4.094344562	
4/5/2018 0:00	10	46	-0.01	0.7	39.2	0.29375	0	280	100	0	39.2	320	13.87	0	4/5/2018 0:00	10	4/5/2018 0:00	1.818217206	4.248495	0.896687393	4.094344562	
4/12/2018 0:00	10	36	0.32	0.02	40.5	0.244444	0	200	82.2	0	41.4	240	8.72	0	4/12/2018 0:00	10	4/12/2018 0:00	3.947989135	4.248495	37.93255605	4.094344562	
4/26/2018 0:00	10	51.5	0.17	1.64	45.5	0.2625	0	110	100	0	43.7	210	7.16	0	4/26/2018 0:00	84	4/26/2018 0:00	4.045611563	4.248495	41.90077228	4.094344562	False Negative
5/3/2018 0:00	84	67	-0.02	0.02	48	0.285417	0	200	83.8	0	49.1	210	7.16	0.07	5/3/2018 0:00	10	5/3/2018 0:00	2.823686448	4.248495	7.803450653	4.094344562	
5/10/2018 0:00	10	54	0.62	0	52.2	0.257639	0	160	99.9	0	51.6	190	8.05	0.18	5/10/2018 0:00	10	5/10/2018 0:00	3.858469741	4.248495	34.51905212	4.094344562	
5/18/2018 0:00	10	59	-0.4	0.64	51.1	0.274306	0	220	89.8	0	51.4	30	11.63	0	5/18/2018 0:00	63	5/18/2018 0:00	3.360378802	4.248495	18.48114851	4.094344562	False Negative
5/24/2018 0:00	63	64.5	0.32	0.28	53.4	0.274306	0	340	99.9	0.17	52.3	160	8.72	0.47	5/24/2018 0:00	10	5/24/2018 0:00	3.365393248	4.248495	18.79289125	4.094344562	
5/31/2018 0:00	10	64	0.13	0.02	57.9	0.279167	0	90	99.5	0	58.1	170	6.71	0.3	5/31/2018 0:00	41	5/31/2018 0:00	3.504624015	4.248495	22.53207983	4.094344562	
6/7/2018 0:00	41	58	0.61	0.87	57.2	0.247917	0	160	89.3	0	58.8	180	7.16	0	6/7/2018 0:00	130	6/7/2018 0:00	4.238789187	4.248495	49.60474125	4.094344562	False Negative
6/14/2018 0:00	130	62.5	-0.45	0.01	59	0.265278	0	200	82.3	0	58.8	330	8.5	0	6/14/2018 0:00	20	6/14/2018 0:00	3.851437747	4.248495	34.28320722	4.094344562	
6/21/2018 0:00	20	64.5	0.32	0.09	61	0.273611	0	210	98	0	62.6	120	4.7	0	6/21/2018 0:00	101	6/21/2018 0:00	3.302647127	4.248495	16.65875586	4.094344562	False Negative
6/28/2018 0:00	101	67.5	0.3	0.21	63.9	0.275	0	150	39.6	0.04	63.5	190	9.62	0.08	6/28/2018 0:00	10	6/28/2018 0:00	3.959256215	4.248495	38.31875927	4.094344562	
7/5/2018 0:00	10	80.5	0.42	0	66.4	0.229861	0	100	100	0	69.4	200	7.38	0	7/5/2018 0:00	121	7/5/2018 0:00	3.502952087	4.248495	22.25933508	4.094344562	False Negative
7/12/2018 0:00	121	70	-0.31	0	66.4	0.263889	0	10	100	0	64.4	10	5.14	0	7/12/2018 0:00	30	7/12/2018 0:00	2.392768034	4.248495	3.06299016	4.094344562	
7/19/2018 0:00	30	73.5	0.15	0.1	66.2	0.238194	0	10	100	0	67.6	170	5.82	0	7/19/2018 0:00	31	7/19/2018 0:00	2.900604422	4.248495	8.710201177	4.094344562	
7/26/2018 0:00	31	75	0.47	0.03	69.1	0.272222	0	160	93.6	0	69.4	180	9.62	0.59	7/26/2018 0:00	20	7/26/2018 0:00	4.209692549	4.248495	48.4597428	4.094344562	
8/2/2018 0:00	20	73	0.23	0	71.1	0.292361	0	180	100	0.04	71.8	200	8.28	0	8/2/2018 0:00	80	8/2/2018 0:00	3.153922537	4.248495	13.0849848	4.094344562	False Negative
8/16/2018 0:00	80	78	-0.06	1.19	70.9	0.223611	0	280	99.3	0.06	71.8	310	4.47	0	8/16/2018 0:00	41	8/16/2018 0:00	3.407911546	4.248495	19.56669737	4.094344562	
8/23/2018 0:00	41	73.5	0.69	0.12	69.1	0.233333	0	210	99.3	0.71	69.6	320	7.61	0.02	8/23/2018 0:00	30	8/23/2018 0:00	2.913902285	4.248495	9.157643812	4.094344562	
8/30/2018 0:00	30	82	0.08	0.19	70.2	0.286111	0	190	80.1	0	70	50	6.93	0	8/30/2018 0:00	51	8/30/2018 0:00	3.354090031	4.248495	18.55965751	4.094344562	
9/6/2018 0:00	51	74.5	0.3	0.1	70.5	0.230556	0	160	100	0	70.3	320	6.04	0	9/6/2018 0:00	30	9/6/2018 0:00	3.549884156	4.248495	23.59394457	4.094344562	
9/13/2018 0:00	30	70.5	-0.17	2.53	67.1	0.295139	0	10	68.2	0	66.9	310	5.59	0	9/13/2018 0:00	122	9/13/2018 0:00	3.495773672	4.248495	23.20801238	4.094344562	False Negative
9/20/2018 0:00	122	64	0.83	0.87	68	0.231944	0	40	88.1	0	65.8	30	9.17	0	9/20/2018 0:00	262	9/20/2018 0:00	2.976002533	4.248495	9.838681682	4.094344562	False Negative
9/27/2018 0:00	262	72.5	-0.04	1.73	64.8	0.282639	0	180	63.3	0	65.8	40	9.84	0	9/27/2018 0:00	112	9/27/2018 0:00	3.263285536	4.248495	16.99769413	4.094344562	False Negative
10/4/2018 0:00	112	63.5	0.51	1.21	64.6	0.234722	0	210	96.5	0	64	180	6.04	0	10/4/2018 0:00	20	10/4/2018 0:00	4.126657262	4.248495	45.06006285	4.094344562	
10/11/2018 0:00	20	72	-0.17	0.64	64.4	0.294444	0	200	68.1	0	65.1	200	7.61	0	10/11/2018 0:00	50	10/11/2018 0:00	3.80141731	4.248495	32.4749084	4.094344562	
10/18/2018 0:00	50	52.5	0.94	0.15	58.3	0.238889	0	320	68.1	0.06	56.5	320	12.3	0	10/18/2018 0:00	41	10/18/2018 0:00	4.024496567	4.248495	40.93513916	4.094344562	
10/25/2018 0:00	41	45	-0.13	0.19	54.3	0.284722	0	330	68.1	0.01	52.9	320	11.63	0.18	10/25/2018 0:00	41	10/25/2018 0:00	3.724713419	4.248495	29.56925304	4.094344562	
11/1/2018 0:00	41	44.5	0.41	0.16	51.1	0.224306	0	190	68.1	1.31	52.5	202.5	6	0.12	11/1/2018 0:00	10	11/1/2018 0:00	2.752197178	4.248495	8.38984171	4.094344562	
11/8/2018 0:00	10	56	-0.12	0.98	53.2	0.202778	0	0	68.1	1.21	52.2	337.5	10	0	11/8/2018 0:00	120	11/8/2018 0:00	1.238151971	4.248495	0.242656343	4.094344562	False Negative
11/15/2018 0:00	120	34	0.76	2.5	49.3	0.213889	0	270	68.1	0.11	45	0	9	0	11/15/2018 0:00	20	11/15/2018 0:00	4.224836985	4.248495	49.07714108	4.094344562	
11/29/2018 0:00	20	38	0.02	0.95	45.3	0.214583	0	270	68.1	0.01	42.4	315	13	1.23	11/29/2018 0:00	30	11/29/2018 0:00	4.876417295	4.248495	73.41733252	4.094344562	False Positive
12/6/2018 0:00	30	29.5	-0.05	0	44.2	0.246528	0	10	68.1	0	45	280	6.04	1.06	12/6/2018 0:00	10	12/6/2018 0:00	3.466465435	4.248495	21.74176023	4.094344562	
12/13/2018 0:00	10	30.5	0.32	0	42.3	0.299306	0	310	68.1	0	40.1	10	6.49	0	12/13/2018 0:00	10	12/13/2018 0:00	2.681441752	4.248495	5.740539486	4.094344562	
12/20/2018 0:00	10	30.5	-0.26	0.01	38.1	0.256944	0	300	68.1	0.04	40.1	130	4.03	0.81	12/20/2018 0:00	20	12/20/2018 0:00	3.883338525	4.248495	35.96508843	4.094344562	

Appendix 4 – Geometric Means

Table 14. Oakland Beach Monthly Geometric Means

Year	Month	StationNum	geo_mean	total_obsv
2006	6	1	24.58753918	21
2006	7	1	38.55403303	16
2006	8	1	28.46825157	20
2008	6	1	28.56531873	11
2008	7	1	11.16123174	10
2008	8	1	11.69930813	7
2007	6	1	23.06594182	13
2007	7	1	32.42343126	11
2007	8	1	22.66951195	10
2009	6	1	13.25144975	9
2009	7	1	29.9727437	10
2009	8	1	21.9643172	9
2009	9	1	10	1
2010	6	1	21.63258755	11
2010	7	1	42.94839205	8
2010	8	1	19.82296541	9
2011	6	1	28.03710405	9
2011	7	1	14.91245414	7
2011	8	1	26.84952167	8
2012	6	1	12.28854369	8
2012	7	1	13.7265273	9
2012	8	1	20.65082858	8
2013	6	1	136.4410701	10
2013	7	1	43.98937044	10
2013	8	1	19.06690085	5
2014	6	1	13.9572405	9
2014	7	1	16.24524854	9
2014	8	1	19.4523989	6
2015	6	1	42.66370382	9
2015	7	1	40.9564168	8
2015	8	1	36.98019088	9
2015	9	1	10	1
2016	6	1	30.19303506	10
2016	7	1	12.53927245	5
2016	8	1	26.2888919	6
2017	6	1	58.14167284	8
2017	7	1	75.10617306	8

2017	8	1	25.4722126	9
2018	6	1	23.61155019	8
2018	7	1	16.31358097	7
2018	8	1	16.71118257	9

Table 15. Easton's Beach Monthly Geometric Means

Year	Month	StationNum	geo_mean	total_obsv
2006	6	2	52.36800432	15
2006	7	2	88.46854249	12
2006	8	2	62.01240965	10
2007	6	2	18.89876112	13
2007	7	2	38.67417694	10
2007	8	2	26.13913517	12
2008	6	2	67.806836	10
2008	7	2	56.48036976	9
2008	8	2	116.7244426	7
2009	6	2	20.59249899	12
2009	7	2	57.33722261	15
2009	8	2	35.30758269	13
2011	6	2	46.11187329	14
2011	7	2	28.25578885	10
2011	8	2	46.75363218	14
2012	6	2	14.68847067	12
2012	7	2	34.29971206	13
2012	8	2	27.01118431	12
2013	6	2	15.34127405	7
2013	7	2	30.55676064	11
2013	8	2	17.09800671	10
2014	6	2	29.68852313	13
2014	7	2	28.22111251	10
2014	8	2	46.83601461	12
2014	9	2	146	1
2015	6	2	85.01060505	10
2015	7	2	33.07354823	14
2015	8	2	22.77453649	13
2016	6	2	26.37014517	11
2016	7	2	20.05036415	17
2016	8	2	19.21441558	16
2017	6	2	26.56913346	9
2017	7	2	28.38646347	6

2017	8	2	28.40616054	8
2018	6	2	20.83979386	8
2018	7	2	32.95012332	9
2018	8	2	61.99943293	10

REFERENCES

ASTM. (2014). ASTM D6503 - 14. Retrieved from https://www.astm.org/Standards/D6503.htm

Cyterski, M. (2016). Virtual Beach 3.0.6: User's Guide. Retrieved from https://www.epa.gov/ceam/virtual-beach-306-user-guide

Cyterski, M. et al. (2018). Predictive Modeling of Microbial Water Quality at Jones Falls, Baltimore Harbor, using Virtual Beach (VB).

Francy, D. S., & Darner, R. A. (2006). Procedures for Developing Models To Predict Exceedances of Recreational Water-Quality Standards at Coastal Beaches. doi:10.3133/tm6b5

Francy, D. S., Brady, A. M., Carvin, R. B., Corsi, S. R., Fuller, L. M., Harrison, J. H., . . . Zimmerman, T. M. (2013). Developing and implementing the use of predictive models for estimating water quality at Great Lakes beaches. *Scientific Investigations Report*. doi:10.3133/sir20135166

McLaughlin, D., & Poucher, S. (2018). Quality Assurance Project Plan (QAPP) for the Research Needs for Marine Beaches Project.

Meals, D. W., Spooner, J., Dressing, S. A., & Harcum, J. B. (2011). Statistical Analysis for Monotonic Trends. *Tech Notes 6*. Retrieved from https://www.epa. gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoringtechnical-notes.

Narragansett Bay Estuary Program. (2017). The State of Narragansett Bay and Its Watershed Technical Report. Retrieved from http://nbep.org/01/wp-content/uploads/2017/09/State-of-Narragansett-Bay-and-Its-Watershed.pdf

NOAA. (n.d.). Tides and Currents. Retrieved from https://tidesandcurrents.noaa.gov/physocean.html?bdate=20150526&edate=20150904&units=standard&timezone=GMT&id=8452944&interval=h&action=data

NOAA. (n.d.). Tides and Currents. Retrieved from https://tidesandcurrents.noaa.gov/noaatidepredictions.html?id=8452660&units=standard&bdate=20150828&edate=20150904&timezone=LST/LDT&clock=12hour&datum=MLLW&interval=hilo&action=data

NOAA. (n.d.). National Centers for Environmental Information. Retrieved from https://www.ncdc.noaa.gov/

NOAA. (n.d.). Index of UV for Cities. Retrieved from http://ftp.cpc.ncep.noaa.gov/long/uv/cities/

NOAA Mobilegeographics. (n.d.). Online Tides and Currents Predictions. Retrieved from https://tides.mobilegeographics.com/

R Development Core Team. (2006). A language and environment for statistical computing. *Computing*. Retrieved December, 2018, from https://www.researchgate.net/publication/223190053_A_language_and_environment_for_statistical_computing

United States Environmental Protection Agency. (n.d.). BEACON 2.0 - Beach Advisory and Closing On-line Notification. Retrieved from https://watersgeo.epa.gov/BEACON2

United States Environmental Protection Agency. 2012. Recreational Water Quality Criteria. Office of Water, Washington, DC, 820-F-12-058.

United States Environmental Protection Agency. (2016). Six Key Steps to Developing and Using Predictive Tools at Your Beach. Retrieved from https://www.epa.gov/sites/production/files/2016-03/documents/six-key-steps-guidance-factsheet.pdf.

USGS. (n.d.). National Water Information System: Web Interface. Retrieved from https://waterdata.usgs.gov/nwis/inventory?agency_code=USGS&site_no=414106071223901

USGS. (n.d.). USGS 01117000 HUNT RIVER NEAR EAST GREENWICH, RI. Retrieved from https://waterdata.usgs.gov/nwis/dv?cb_00060=on&format=html&site_no=01117000&referred_module=sw.=&begin_date=2015-05-25&end_date=2015-09-04

Weather Underground (n.d.). Weather History for KPVD - June, 2016. Retrieved from https://www.wunderground.com/history/airportfrompws/KPVD/2016/6/18/CustomHistory.html?dayend=31&monthend=6&yearend=2016&req_city=&req_state=&req_statename=&reqdb.zip=&reqdb.magic=&reqdb.wmo=

Weather Underground (n.d.). Weather History for KUUU – Monthly History. Retrieved from https://www.wunderground.com/history/airportfrompws/KUUU/2015/1/1/MonthlyHistory.html?req_city=&req_state=&req_statename=&reqdb.zip=&reqdb.magic=&reqdb.wmo=