Modeling Temporal Variations in *Escherichia coli* at Villa Angela Beach in Cleveland, Ohio

Alex Johnson, Sho Tanaka, Jemima Ukwela, Alexa Roberts
Faculty Advisor: Dr. Barbara Margolius
NEORSD: Eric Soehnlen

NORTHEAST OHIO REGIONAL SEWER DISTRICT

The Northeast Ohio Regional Sewer District (NEORSD) is a public utility district in Cuyahoga and Summit County. NEORSD manages three wastewater treatment plants serving 62 communities, over one million residents, and 90 billion gallons of wastewater a year. The mission of the NEORSD is to protect public health and the environment while providing effective services. The district works to improve combined sewer overflows in the Greater Cleveland area's old sewers through its engineering and construction programs. The district opened an environmental & and maintenance services center(EMSC) in 1990. The departments in this building are focused on industrial discharges, storm water management, and analytical services. Also located at EMSC is the water quality and industrial surveillance (WQIS) department where investigators monitor stream water quality, inspect and sample industrial users discharges, and respond to chemical spills and other hazardous emergencies. Overall, WQIS is a pretreatment program to comply with the districts National Pollution Discharge Elimination System.

BEACH PROGRAM

The Sewer District monitors water quality conditions at several area beaches to protect the public from harmful bacteria during the recreation season. The program has been supported since 1992 through the U.S EPA's Beaches Environmental Assessment and Coastal Health (BEACH) program as well as several other agencies.

- Villa Angela Beach
- Edgewater Beach
- Euclid Beach

Water samples are taken daily and tested for *Escherichia coli* (E. coli) by analytical services at EMSC. Results take 24 hours to receive. EPA approved modeling software Virtual Beach 3.0 is used to make an empirical model to predict E. coli levels to provide the public with timely beach water quality. The purpose of this project is to investigate Villa Angela Beach's model and see how E. coli grows throughout time

Photos from NEORSD

VIRTUAL BEACH

Data starting in 2013 was compiled into a spreadsheet with all the beach parameters and imported into virtual beach including the lab results of E. coli. Virtual Beach transforms every parameter to the transformation with the highest correlation coefficient to achieve the best model. **Only 80%** of the data was used.

Dependent Variable: $[\log_{10}]$E. Coli
Independent Variables: Transformed Flow, Lake Turbidity, Wave Height

Virtual Beach runs every combination of the model and evaluates the models based on a criteria set by the user.

The Bathing Water Recreational Criteria for E. coli is **235 CFU/100mL**. In order to maximize model sensitivity and accuracy, a new threshold, the decision criteria, is set. The remaining **20%** of data was used as test data to see how well each model performed. Model performance is based on accuracy and sensitivity of the predictions. A percent exceedance (of 235 CFU/100mL) is also generated because the decision criteria is not representative of beach water quality.

METHODS

Parameter	R^2	P-Value
Flow	0.4004	< 0.001
Weather	0.3026	< 0.001
Air Temperature	0.1923	< 0.001
Lake Turbidity	0.544	< 0.001
Wave Height	0.5367	< 0.001
Specific Conductivity	0.1221	< 0.001

Figure 1. Possible model independent variables correlation to the combined model. The variables analyzed were chosen because of their high correlation with the combined model.

Transform dependent and independent variables

Variables

1. Mean Flow from Euclid Creek (MF)
2. Lake Turbidity (LT)
3. Wave Height (WH)

Break the model into multiple models to see how it performs in different parts of the recreation season

	Flow	Turbidity	Wave Height
May	0.5993	0.5670	0.5947
June	0.4830	0.6530	0.5730
July	0.3165	0.5976	0.6603
August	0.2679	0.5394	0.5052
September	0.6410	0.3087	0.7029
Combined	0.4004	0.5440	0.5367

Figure 2. Model broken down into months and the R^2 values of the three independent variables. Combined refers to all the months together as one model. The flow was dramatically lower in July and August.

RESULTS

Made a July and August model and a May-June September model to see if the isolating July and August yielded better overall model performance. These models were tested using their respective months test data.

Equations and Standardized Coefficients of the models

Combined: Ecoli_cfu_100ml = 0.6715 + 0.2417*(LN(FLOW_Mean_Euclid_Creek)) + 0.3084*(QUADROOT(Lake_Turb)) + 0.6577*(SQUAREROOT(WaveHT_Ft))

MF = 0.3112
WH = 0.3768
LT = 0.1700

July-August:
Ecoli_cfu_100ml = 1.325 + 0.05657*(SQUAREROOT(FLOW_Mean_Euclid_Creek)) + 0.5421*(SQUAREROOT(WaveHT_Ft)) + 0.1876*(LN(Lake_Turb))

MF = 0.2018
WH = 0.3205
LT = 0.2626

May-June, September:
Ecoli_cfu_100ml = 1.749 - 7.396*(INVERSE(FLOW_Mean_Euclid_Creek,2.65)) + 0.1794*(QUADROOT(Lake_Turb)) + 1.111*(QUADROOT(WaveHT_Ft))

MF = 0.4428
WH = 0.4194
LT = 0.0918

Standardized coefficients confirmed that flow from Euclid Creek is not as significant for E. coli growth in July-August as it is for May-June and September. Wave height gains more importance in July-August Model relative to flow.

	Combined	July-August	May-June, September
Sensitivity	72.70%	60.00%	84.21%
Accuracy	79.41%	60.06%	60.00%
Specificity	75.64%	61.11%	42.31%

Figure 3. Isolated model results compared to the combined model using each models respective months as the test data. Breaking down into months did not improve the model.

	Combined	July-August	May-June, September
Sensitivity	72.70%	82.35%	82.35%
Accuracy	79.41%	76.92%	76.92%
Specificity	75.64%	63.63%	72.27%

Figure 4. Using the combined month data to build a model to predict E. coli for the isolated months. How the combined model performed in these month categories. Breaking down into months did not improve the model because these models showed improvement over isolated the models in *Figure 3*.

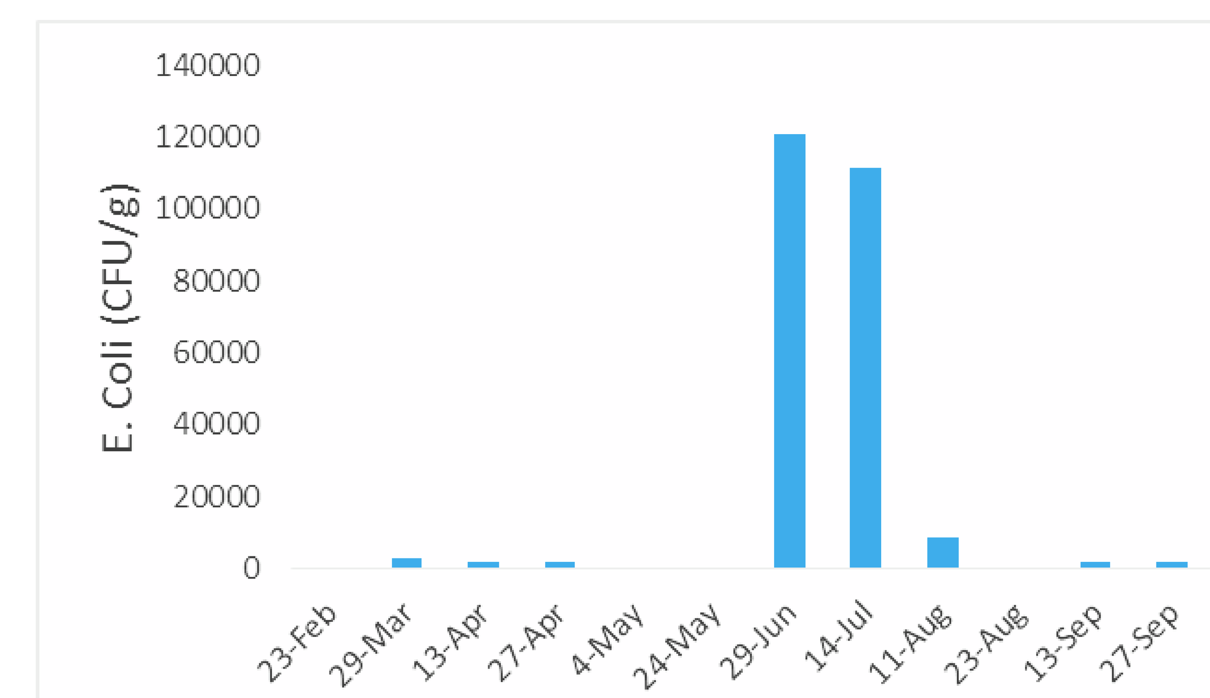

Figure 5. Villa Angela Beach sand E. coli CFU/g. At the very end of June and July E. coli is noticeably more abundant than in the rest of the year. August also has a high abundance relative to the other months. Sand sampled in June was at then end of the month close to July, so the data is not representative of June. Theoretically most of June will have low sand E. coli colonies.

CONCLUSIONS

- May-June, September: E. coli growing mechanism is the flow from Euclid Creek.
- July and August: E. coli growing mechanism is different than the flow from Euclid Creek.
- Sand data agrees with the importance the July and August model placed on wave height.
- Waves mixing sand on Villa Angela Beach are key to E. coli growth during July and August.
- E. coli concentrations are higher in the sand at Villa Angela Beach.

RECOMMENDATIONS

- Sand data for June is at the end of the month and is not representative of June.
- Sand sample more often during the recreational season to get a real representation of E. coli sand colonies.
- Make a test model for July and August during the recreational season to further understand E. coli growing mechanisms at Villa Angela Beach.

SOURCES

- Northeast Ohio Regional Sewer District. (2010). *Facts about the Sewer District* [Brochure]. Cleveland, OH