Tools for Assessing and Managing Environmental Waters

Marirosa Molina, USEPA, ORD, National Exposure Research Laboratory

The findings and conclusions in this presentation have not been formally disseminated by the EPA and should not be construed to represent any Agency determination or policy.
Introduction

- Water resources across the nation are regularly contaminated by unacceptably high levels of fecal indicator bacteria (FIB) that exceed EPA water quality criteria levels.
National Summary of Water Quality Assessments

Total miles assessed 1,110,961
- Pathogens ≠ indicators
- No source information
- Increased economic burden
- Increased health risk

Cost of Chesapeake Bay TMDL for Virginia

<table>
<thead>
<tr>
<th>Source</th>
<th>Projected Total Cost ($ in billions)</th>
<th>Who Pays</th>
<th>Potential State Costs ($ in billions)</th>
<th>Potential Sources of Funding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wastewater (Including CSOs)</td>
<td>$1.4</td>
<td>State Govt./Local Govt./Rate-payers</td>
<td>$0.3 (plus $78 million for CSOs?)</td>
<td>WQIF, State GF, Bonds /Local GF, Bonds/Tax Assessments, Sewer Rates</td>
</tr>
<tr>
<td>Agriculture</td>
<td>$1.2+</td>
<td>State Govt./Farmers</td>
<td>$0.8+</td>
<td>WQIF, State GF/Agribusinesses</td>
</tr>
<tr>
<td>Stormwater</td>
<td>$9.4 to $11.5 (including VDOT)</td>
<td>Local Govt./Property Owners/VDOT</td>
<td>$2.1 (VDOT Share)</td>
<td>Local GF, Bonds/Utility Fees, Assessments/Transportation Trust Fund</td>
</tr>
<tr>
<td>Onsite/Septic Systems</td>
<td>$1.6</td>
<td>Property Owners</td>
<td>Unknown What Role State May Play</td>
<td>“Betterment loans”, Potential for Tax Credits or Grants</td>
</tr>
<tr>
<td>Bay TMDL Total</td>
<td>$13.6 to $15.7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From Senate Finance Committee Report Nov 18, 2011

---

**NEED:** Develop Microbial Risk Assessment Tools

**Understanding the risk of exposure to waterborne pathogens**

**Information may help states and stakeholders prioritize use of resources**
2012/2017 RWQC Revision
More tools for assessing and managing recreational waters

Rapid qPCR Assays
• Enterococcus qPCR technique (EPA Method 1611),
• use of qPCR on a site-specific basis for monitoring.

Microbial Source Tracking (MST):
• allows development of alternative site-specific criteria and identifies opportunities for source remediation.

Predictive models
• offer states, territories, and tribes the potential for same-day notification and public health protection
• considerably lower capital investment and unit costs.
2012/2017 RWQC Revision
More tools for assessing and managing recreational waters

**Rapid qPCR Assays**
- Enterococcus qPCR technique (EPA Method 1611),
- use of qPCR on a site-specific basis for monitoring.

**Microbial Source Tracking (MST):**
- allows development of alternative site-specific criteria and identifies opportunities for source remediation.

**Predictive models**
- offer states, territories, and tribes the potential for same-day notification and public health protection
- considerably lower capital investment and unit costs.
Culture vs. Real Time PCR Based Fecal Indicator Bacteria Measurements to Determine Beach Water Quality

- Filter Water Sample
- Extract DNA from Filter
- Amplify & Quantify Indicator DNA by PCR to Determine Water Quality
- Count Indicator Colonies on Filter to Determine Water Quality
- Grow Indicators on Filter Membranes

24 Hours
2-6 Hours

Pathogen (Virus, Parasite, or Bacteria) Indicator (Fecal Bacteria)
qPCR Enterococci - better predictor of illness

- *Enterococcus* spp. measured by qPCR is more predictive of swimming-associated health effects.
- Results are more timely than traditional methods used for culturing bacterial indicators.

Predicted probabilities of GI illness as a function of *Enterococcus* QPCRCE, predicted from the logistic regression model, adjusted for age and beach.

Wade et al., 2006
Rapid qPCR Assays
- Enterococcus qPCR technique (EPA Method 1611),
- use of qPCR on a site-specific basis for monitoring.

Microbial Source Tracking (MST):
- allows development of alternative site-specific criteria and identifies opportunities for source remediation.

Predictive models
- offer states, territories, and tribes the potential for same-day notification and public health protection
- considerably lower capital investment and unit costs.
• Animal feeding operations (CAFOS, AFOs, feedlots produce > 335 million tons of manure/yr.
• Humans produce ~ 7 million tons of feces/yr (or $3.9 \times 10^{12}$ kg/yr)

Berendes et al., 2018
MST Definition

Microbial source tracking (MST) refers to methodology designed to determine sources of fecal pollution in environmental waters by relating fecal microorganisms with an associated host.
MST tools can be applied for:

- Development of Total Maximum Daily Loads (TMDL)
- Evaluation of the effectiveness of best management practices.

From Stewart et al., 2013
Table 5 – Performance of quantitative MST assays normalized by CFU ENT. The number and percent of non-target samples measuring copies per CFU in the range measured in target samples is provided in column labeled ‘non-target in target range’; assays for which this is 0 are deemed specific. The median copies/CFU measured in target challenge samples is provided. The second number for the human assays (the number following the semi-colon) shows the median if DNQs and NDS are not included in the median calculation. Assays with a median greater than 50 copies/CFU are deemed sensitive. Both sensitive and specific assays are indicated. An “*” indicates that the N would change to Y if the median that does not include DNQs and NDS is used. The number of laboratories (N) that ran the assay is provided. Assays run by a single lab will require further testing.

<table>
<thead>
<tr>
<th>Assay</th>
<th>Host</th>
<th>N</th>
<th># (%) non-target in target range</th>
<th>Median in target (copies/CFU)</th>
<th>Specific</th>
<th>Sensitive</th>
<th>Spec &amp; sens</th>
</tr>
</thead>
<tbody>
<tr>
<td>BacH</td>
<td>Human</td>
<td>1</td>
<td>0 (0)</td>
<td>87; 375</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>BacHum</td>
<td>Human</td>
<td>7</td>
<td>20 (11)</td>
<td>331; 374</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>BsterI1</td>
<td>Human</td>
<td>4</td>
<td>10 (10)</td>
<td>123; 130</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Btheta</td>
<td>Human</td>
<td>1</td>
<td>0 (0)</td>
<td>11; 13</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>gyrB</td>
<td>Human</td>
<td>1</td>
<td>0 (0)</td>
<td>0.003; 279</td>
<td>Y</td>
<td>N*</td>
<td>N*</td>
</tr>
<tr>
<td>IF183SYBR</td>
<td>Human</td>
<td>4</td>
<td>7 (7)</td>
<td>52; 71</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>HF183Taqman</td>
<td>Human</td>
<td>5</td>
<td>0 (0)</td>
<td>138; 140</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>HumM2</td>
<td>Human</td>
<td>6</td>
<td>0 (0)</td>
<td>7; 48</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>mifH</td>
<td>Human</td>
<td>5</td>
<td>24 (19)</td>
<td>33; 167</td>
<td>N</td>
<td>N*</td>
<td>N</td>
</tr>
<tr>
<td>BacCow</td>
<td>Cow</td>
<td>5</td>
<td>12 (7)</td>
<td>12,490</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>CowM2</td>
<td>Cow</td>
<td>5</td>
<td>0 (0)</td>
<td>15</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>CowM3</td>
<td>Cow</td>
<td>1</td>
<td>0 (0)</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>BacR</td>
<td>Ruminant</td>
<td>2</td>
<td>0 (0)</td>
<td>955</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Rum2Bac</td>
<td>Ruminant</td>
<td>1</td>
<td>0 (0)</td>
<td>832</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Gull2SYBR</td>
<td>Gull</td>
<td>4</td>
<td>14 (10)</td>
<td>0.4</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Gull2Taqman</td>
<td>Gull</td>
<td>6</td>
<td>21 (10)</td>
<td>7</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>LeeSeaGull</td>
<td>Gull</td>
<td>1</td>
<td>0 (0)</td>
<td>55</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>
2012/2017 RWQC Revision
More tools for assessing and managing recreational waters

Rapid qPCR Assays
- Enterococcus qPCR technique (EPA Method 1611),
- use of qPCR on a site-specific basis for monitoring.

Microbial Source Tracking (MST):
- allows development of alternative site-specific criteria and identifies opportunities for source remediation.

Predictive models
- offer states, territories, and tribes the potential for same-day notification and public health protection
- considerably lower capital investment and unit costs.
Virtual Beach

A decision-support tool that facilitates the process of building and implementing a statistical model for predicting microbial water quality.

Environmental Parameters

- Rainfall (24-hr, 48-hr)
- Turbidity, conductivity
- Wave height
- Wind speed & direction
- Air & water temperature
- UV/PAR
- Tributary discharge
- Lake currents and stage
- Bathers, birds, debris
Virtual Beach Tools: MLR, PLS, GBM

Multiple Linear Regression Models (MLR): fitting a linear equation to observed data

Generalized Boosted Regression Models (GBM): a linked chain of simple decision trees

![Fitted vs Observed Graph](Image)

- Decision Threshold
- Regulatory Threshold

![Diagram of GBM Trees](Image)

1st Tree + 2nd Tree + 3rd Tree + 4th Tree + 5th Tree + 6th Tree + ... + n-1 Tree + n-th Tree = Final Prediction
Probability of exceedance

Type I Error = False Positive; advisory issued under safe conditions

Type II Error = False Negative; no advisory issued under risky conditions
Model Application

- Application of a recent version of the software package Virtual Beach tool (VB 3.0) has been performed to predict microbial water quality in beaches located around the Great Lakes, and in South Carolina, Florida, Alabama, Mississippi, Puerto Rico.
Case Study I

Efficacy of statistical modeling tools to determine stream and river impairment due to fecal contamination.
Despite VB widespread use around the Great Lakes (>100 + beaches), application in inland waters is limited.
## Assessment of US Waters

<table>
<thead>
<tr>
<th>Category</th>
<th>Rivers (Miles)</th>
<th>Lakes, Reservoirs (Acres)</th>
<th>Bays and Estuaries (Square Miles)</th>
<th>Ocean and Near Coastal (Square Miles)</th>
<th>Wetlands (Acres)</th>
<th>Great Lakes Shoreline (Miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good Waters</td>
<td>518,293</td>
<td>5,390,570</td>
<td>11,516</td>
<td>726</td>
<td>569,328</td>
<td>106</td>
</tr>
<tr>
<td>Threatened Waters</td>
<td>4,495</td>
<td>30,309</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impaired Waters</td>
<td>588,173</td>
<td>13,208,917</td>
<td>44,625</td>
<td>6,218</td>
<td>672,924</td>
<td>4,354</td>
</tr>
<tr>
<td><strong>Total Assessed Waters</strong></td>
<td>1,110,961</td>
<td>18,629,795</td>
<td>56,141</td>
<td>6,944</td>
<td>1,242,252</td>
<td>4,460</td>
</tr>
<tr>
<td><strong>Total Waters</strong></td>
<td>3,533,205</td>
<td>41,666,049</td>
<td>87,791</td>
<td>54,120</td>
<td>107,700,000</td>
<td>5,202</td>
</tr>
<tr>
<td><strong>Percent of Waters Assessed</strong></td>
<td>31.4</td>
<td>44.7</td>
<td>63.9</td>
<td>12.8</td>
<td>1.2</td>
<td>85.7</td>
</tr>
</tbody>
</table>
Objective

Evaluate the effectiveness of Multiple Linear Regression (MLR) and Generalized Boosted Regression Models (GBM) for predicting the impairment of inland rivers and streams.
Data Collection

• Intensive field study conducted from 2012-2015

• Water samples collected during rainfall events and baseflow conditions from two sites in the watershed.

• Samples were analyzed for *E. coli*, Enterococci (culturable and qPCR), MST markers and water quality parameters.
Model Development

- Model training data were randomly selected (approximately 75% of entire dataset)

- IVs included Air Temperature, Barometric Pressure, Dewpoint, Rainfall (15min; 1, 24, 48hr), Rainfall Intensity (1 min), Stream Discharge, Total Suspended Solids (TSS), Turbidity, Water Temperature, Avg. Wind Speed
FIB Prediction at Clouds Creek

Model predictions were made for data not used in model development (approx. 25% of dataset)
## Model Predictive Performance

<table>
<thead>
<tr>
<th>Parameters</th>
<th>E.Coli</th>
<th>Enterococci</th>
<th>E.Coli</th>
<th>Enterococci</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GBM Clouds Creek</td>
<td>GBM Carlton</td>
<td>MLR Clouds Creek</td>
<td>MLR Carlton</td>
</tr>
<tr>
<td>FP</td>
<td>13</td>
<td>15</td>
<td>21</td>
<td>41</td>
</tr>
<tr>
<td>FN</td>
<td>22</td>
<td>22</td>
<td>44</td>
<td>52</td>
</tr>
<tr>
<td>TP</td>
<td>163</td>
<td>163</td>
<td>141</td>
<td>144</td>
</tr>
<tr>
<td>TN</td>
<td>172</td>
<td>170</td>
<td>164</td>
<td>133</td>
</tr>
<tr>
<td>Data Points (N)</td>
<td>370</td>
<td>370</td>
<td>370</td>
<td>370</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>93</td>
<td>92</td>
<td>89</td>
<td>78</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>88</td>
<td>88</td>
<td>76</td>
<td>72</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>91</td>
<td>90</td>
<td>82</td>
<td>75</td>
</tr>
<tr>
<td>R squared</td>
<td>0.89</td>
<td>0.91</td>
<td>0.72</td>
<td>0.67</td>
</tr>
</tbody>
</table>

- R^2 values for all predictions were > 0.6
- R^2 values were higher for GBM predictions
- Averaged across both models (*E.coli* and enterococci) GBM prediction errors were 16% lower than MLR errors at Clouds Creek, and 7% lower at Carlton
## Selected Environmental Parameters

<table>
<thead>
<tr>
<th>Area</th>
<th>Enterococci</th>
<th>E. coli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carlton</td>
<td>Turbidity, 48 hr Rainfall, Dew Point, Humidity and wind speed</td>
<td>Turbidity, 48 hr rainfall, dewpoint</td>
</tr>
<tr>
<td>Clouds Creek</td>
<td>Turbidity, Rainfall Intensity, 48 hr rainfall, air temperature, dew point, Barometric pressure, wind speed</td>
<td>Water temperature, 1 hr rainfall, 24 hr rainfall</td>
</tr>
</tbody>
</table>
Take Home Message

• Based on the predictive performance of both GBM and MLR models for FIB at both field sites, our results show statistical models developed with VB can be effective for predicting the impairment of inland rivers and streams.

• GBM produced better *E.coli* and Enterococcci predictions than MLR regardless of site differences.
Case Study II

Using a Microbial Citizen Science Initiative As A Tool to Manage Stream Water Quality in An Urban Watershed
Objective

To provide information to local governments and the UGA River Basin Center about common sources of fecal contamination and locate hot spots within local urban streams.
Land Use Development Intensity of Athens, GA with corresponding average *E. coli* concentrations (Log GC/100mL).
Table 1. Pearson correlation and P-values between qPCR marker concentration and molecular E. coli as a function of land use development.

<table>
<thead>
<tr>
<th>Land Use Development Intensity</th>
<th>Site</th>
<th>qPCR Marker</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Rum2Bac</td>
<td>DogBact</td>
<td>HF183MGB</td>
</tr>
<tr>
<td>High</td>
<td>5</td>
<td>0.13</td>
<td>0.16</td>
<td>0.64***</td>
</tr>
<tr>
<td>High</td>
<td>6</td>
<td>-</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>High</td>
<td>7</td>
<td>0.15</td>
<td>0.50***</td>
<td>0.41**</td>
</tr>
<tr>
<td>High</td>
<td>8</td>
<td>-0.428</td>
<td>0.36</td>
<td>0.08</td>
</tr>
<tr>
<td>High</td>
<td>9</td>
<td>0.19</td>
<td>0.47**</td>
<td>0.67***</td>
</tr>
<tr>
<td>Medium</td>
<td>2</td>
<td>0.21</td>
<td>0.58***</td>
<td>0.56**</td>
</tr>
<tr>
<td>Medium</td>
<td>3</td>
<td>0.56***</td>
<td>0.18</td>
<td>0.53***</td>
</tr>
<tr>
<td>Medium</td>
<td>10</td>
<td>-</td>
<td>0.11</td>
<td>0.13</td>
</tr>
<tr>
<td>Low</td>
<td>1</td>
<td>0.07</td>
<td>0.16</td>
<td>0.34*</td>
</tr>
<tr>
<td>Low</td>
<td>4</td>
<td>-0.12</td>
<td>0.56***</td>
<td>0.49**</td>
</tr>
</tbody>
</table>

Sites categorized by Land Use Development Intensity, High >79%, Medium 50-79%, Low<50%). Positive correlation value and significant P-value (*= 0.01), (**=0.05) and (***=0.001) indicated a positive trend between both variables.
Take Home Message

- Detection of HF183MGB in highly developed areas of Athens may be an indication of failing sewer infrastructure.

- Significant correlations between source tracking markers and *E. coli* concentrations indicate that there are variable sources of contamination contributing with high *E. coli* numbers across the watershed which are independent of land use intensity.

- Little variation of *E. coli* concentration detected between methods used by volunteers and EPA scientists (average CV 3.29%), demonstrates analysis by volunteer scientists meet EPA QA/QC standards.
Next Steps

• Continue validating performance of models using MST markers
• Compare GBM vs. MLR performance across a variety of watersheds (e.g. urban systems)
• Validate model using citizen science collected data.
Acknowledgements

- Blake Snyder
- David Spidle,
- Mike Cyterski,
- Ken Bradshaw,
- Rania Georgacopolous
- Kelvin Wong
- Robert Sowah
- Kyler Herrington
- Kayleigh Hall
- Reid Brown, SSA-EPA, Athens
- Chris Fitzgerald, SSA-EPA, Athens
- Ellen Price, SSA-EPA, Athens