Improving predictions of bacterial water quality with real-time networked sensors

Benjamen Wetherill
UMass Boston
School of Marine Sciences
Outline

• Methods for monitoring water quality
  • Study sites & models developed
  • Benefits of networked real-time sensors
  • Results & Conclusions
Current methods for monitoring water quality

1. Lab analysis of water samples
   - Requires 24 hours
   - Handicapped by high variability of bacteria
   - Little correlation with next day conditions

2. Modeling based on environmental conditions
   - Positive correlation with rainfall, wind, turbidity ...
   - Negative correlation with solar radiance, salinity ...
   - 50-80% sensitivity
Current practices employed by beaches

- Most beach managers use previous day sample results [1]
- Some combine that with simple rainfall thresholds
  - Mass DCR – sample positive or 24-hour rainfall > threshold
- Some use regression models based on hydro-meteorological data
  - CRWA – rain, temperature, wind, river flow
  - Ohio “Nowcasting” – turbidity, rainfall, lake stage, day of year
- Could not find any examples of real-time forecasts based on real-time data

[ 1. NRDC 2004 ]
Opportunity for Sensor Networks?

• What are sensor networks?

• With current technologies, it is completely feasible to automate data collection, analysis, and reporting.

• Could water quality modeling benefit from automated data?

• Real-time data
  • Always updated with latest conditions

• Continuous data
  • Moving average trends instead of discrete points

• Localized data
  • Local data is always more relevant

• Connectable data
  • Data from multiple sources can be combined

• Automatable
  • Not dependent on a person being at work
Outline

• Methods for monitoring water quality
  • Sites studied & models developed
  • Benefits of networked real-time sensors
  • Results & Conclusions
Two sites studied

- **Charles River**
  - Fresh water
  - River
  - *E.coli* bacteria
  - Samples 2x per week
  - 1½ seasons of data
  - Not open to swimming

- **Wollaston Beach**
  - Salt water
  - Bay
  - *Enterococcus* bacteria
  - Daily samples
  - 1 season of data
  - Open to swimming
Wollaston Beach

- Frequent beach closures
- Daily bacteria count data from DCR 6/25-8/30
- Environmental data sourced from...
  - CESN buoy
  - Boston tide gage
  - Umass weather station
- Modeled maximum daily bacteria from 4 sites

Image credit: MWRA Environmental Quality Dept. [http://www.mwra.state.ma.us/harbor/graphic/wollaston.jpg](http://www.mwra.state.ma.us/harbor/graphic/wollaston.jpg)
Wollaston: Explanatory Variable Statistics

- Statistical analysis of each variable highlighted...
  - Wind direction
  - PAR
  - Air temp
  - Tide phase
  - Tidal range
  - Rain

<table>
<thead>
<tr>
<th>Variable</th>
<th>Corr.</th>
<th>$R^2$</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind Direction (1.5x cst.)</td>
<td>24hr</td>
<td>0.47</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>6hr</td>
<td>0.44</td>
<td>0.20</td>
</tr>
<tr>
<td>PAR</td>
<td>24hr</td>
<td>-0.37</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>12hr</td>
<td>-0.29</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>3hr</td>
<td>-0.25</td>
<td>0.06</td>
</tr>
<tr>
<td>Sample Light/Dark</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air Temperature</td>
<td>24hr</td>
<td>-0.38</td>
<td>0.15</td>
</tr>
<tr>
<td>Water Temperature</td>
<td>24hr</td>
<td>-0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>Sample Ebb/Flood</td>
<td>48hr</td>
<td>0.29</td>
<td>0.12</td>
</tr>
<tr>
<td>Tidal Range</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample Tide Level</td>
<td></td>
<td>-0.08</td>
<td>0.01</td>
</tr>
<tr>
<td>Rain (log)</td>
<td>72hr</td>
<td>0.31</td>
<td>0.10</td>
</tr>
<tr>
<td>Days since rain</td>
<td>24hr</td>
<td>0.31</td>
<td>0.09</td>
</tr>
<tr>
<td>Wind Speed (average)</td>
<td>48hr</td>
<td>-0.15</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>24hr</td>
<td>-0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>Salinity</td>
<td>24hr</td>
<td>-0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>Turbidity</td>
<td>24hr</td>
<td>-0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Sample Time</td>
<td></td>
<td>-0.16</td>
<td>0.03</td>
</tr>
</tbody>
</table>

$$\ln\left(\frac{p}{1-p}\right) = 3.96 - 2.6T_p + 0.21 \ln\left(R_{72} + 10^{-4}\right) - 0.40W_{24} + 0.22T_{24}M_{24}$$
Linear Regression
True Positive: 50%; True Negative: 98%

Logistic Regression (30% prob.)
True Positive: 79%; True Negative: 85%
Sensitivity  Specificity

Combined
True Positive: 86%; True Negative: 85%

Previous Day Bacteria
True Positive: 21% (2013), 14% (2012)
• Methods for monitoring water quality
• Sites studied & models developed
  • Benefits of networked real-time sensors
• Results & Conclusions
Multiple updates per day

• Some conditions vary dramatically from hour to hour.
• Time of model calculation can significantly change results.

![Wind Direction - Wollaston Beach (high-frequency)](chart)
Multiple updates per day

• Hourly model updates raised sensitivities above 90% at both sites (> 5% increase).
  • Charles: daily – 87% TP; hourly – 93% TP
  • Wollaston: daily – 79% TP; hourly – 93% TP

• Frequent updates may also provide unknown information about non-sampled times of day.
Access to absolute latest data

- Ability to use latest data was key to Charles River model improvement.
  - Calendar day updates miss early morning events
  - Real-time updates capture early morning events

Old method
Calendar-day updates

Rain

Real-time updates
Access to absolute latest data

- Can value of data timeliness be quantified?
- Testing adjusted time buckets resulted in > 5% loss of accuracy with only 4-hour old data

Real-time updates (aligned with sampling)

Median sampling time (7:33 AM)

<table>
<thead>
<tr>
<th>Logistic Model</th>
<th>True positive</th>
<th>True negative</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floating</td>
<td>79%</td>
<td>85%</td>
<td>84%</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>64%</td>
<td>70%</td>
<td>69%</td>
</tr>
<tr>
<td>7:00 AM</td>
<td>79%</td>
<td>74%</td>
<td>75%</td>
</tr>
<tr>
<td>5:00 AM</td>
<td>86%</td>
<td>70%</td>
<td>74%</td>
</tr>
<tr>
<td>3:00 AM</td>
<td>64%</td>
<td>70%</td>
<td>69%</td>
</tr>
</tbody>
</table>

(Based on Wollaston 2013 data)
Local data from local sensors

• Study evaluated benefit of local sensors
• Wollaston rain data collected from...
  • airport (6 mi) – larger big storms, smaller small storms
  • local station (3 mi) – consistently stronger correlations (~5% greater model accuracy)

<table>
<thead>
<tr>
<th></th>
<th>Local Rain (UMass library)</th>
<th>Logan Airport Rain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation with bacteria (24hr)</td>
<td>0.31</td>
<td>0.20</td>
</tr>
<tr>
<td>Correlation with bacteria (72hr)</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>Regression predictive strength (24hr)</td>
<td>0.09</td>
<td>0.04</td>
</tr>
<tr>
<td>Regression predictive strength (72hr)</td>
<td>0.011</td>
<td>0.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear model</td>
<td>50%</td>
<td>87%</td>
<td>43%</td>
<td>82%</td>
</tr>
<tr>
<td>Logistic model</td>
<td>79%</td>
<td>84%</td>
<td>71%</td>
<td>79%</td>
</tr>
<tr>
<td>Logistic model built with Logan rain</td>
<td>79%</td>
<td>80%</td>
<td>79%</td>
<td>80%</td>
</tr>
</tbody>
</table>

Some data can be sourced from existing sensors...
• River Flow – USGS gage 9 miles up river, but no tributaries in between
• Tide Height – NOAA gauge 6 miles away, but matches local tide within 10 minutes
Outline

• Methods for monitoring water quality
• Sites studied & models developed
• Benefits of networked real-time sensors

• Results & Conclusions
Results & Conclusions

• Modeling with environmental variables can provide significantly better predictions than previous day water samples

• Hourly updates provide higher sensitivity and useful data about sub-24-hour changes
  • >5% improvement in sensitivities over daily updates

• Real-time latest data results in more accurate models
  • >5% loss of accuracy when data is 4 hours old

• Some variables must be collected locally
  • 5% loss of accuracy when rain data sourced from airport instead of locally
Acknowledgements

• Committee
  • Robert F. Chen, Crystal Schaaf, Bernie Gardner, Michael Shiaris
• Lab mates – Hayley Schiebel, Jill Arriola, Keith Cialino
• Francesco Peri - CESN
• Julie Wood – CRWA
• Kelly Coughlin – MWRA
Questions?

Image credit: Indiana Beaches http://www.in.gov/dem/In.gov/beaches

Water Quality Today
GOOD
BASED ON RECENT MODELING
FOR E.COLI BACTERIA
For More Information Visit: www.dem.in.gov/beaches

Image credit: Indiana Beaches http://www.in.gov/dem/In.gov/beaches