Real-time modeling and reporting of beach water quality on the Great Lakes

Laura De Cicco¹, David Sibley¹
Steve Corsi², Rebecca Carvin²
Nathaniel Booth³

¹ USGS Center for Integrated Data Analytics (CIDA)
² USGS Wisconsin Water Science Center
³ USGS Office of Water Information

Speaker: Daniel Sullivan²
Introduction

- The overall mission is to provide science-based information and methods to:
  - More accurately make beach closure and advisory decisions
  - Understand the sources and physical processes affecting beach contaminants
  - Understand how science-based information can be used to mitigate and restore beaches and protect the public.
- Empirical models have been developed to predict *E. coli* concentrations (a surrogate for water quality) in nearshore waters based on environmental data (such as rainfall, water current, turbidity, and temperature)
An amazing amount of data is available for creating environmental models. However:

- many sources
- many formats
- many period of records and resolutions

In order to create and implement real-time models, a tool was needed to efficiently discover, acquire, and process these diverse data sets.

**Environmental Data Discovery and Transformation (EnDDaT)** was designed to fit this need.

http://cida.usgs.gov/enddat/
Sorting data from multiple web sources can be daunting

Thredds / OPeNDAP / NetCDF

XML (WaterML2, WQX)

- EnDDaT deals with the issues that tend to always come up:
  - Parsing
  - Sorting
  - Missing data
  - Time zones
  - Daylight savings
EnDDaT Introduction

- Discover available data within a specified area around a point location
- Gather data from multiple sources
- Sort data
- Process data
  - resolve vectors based on orientation
  - moving window: mean, minimum, maximum, summation
- Export data (tab or csv delimited, interactive graph)
• http://cida.usgs.gov/enddat/
• Home page lets you jump right into Data Discovery
• Or, go to the User Guide
Data Discovery

- Choose a central project location
- Set a bounding box area
- Choose data to discover

Points show up on the map at the location of available data.
Information on the available data is displayed in a table.
Choose data to discover.
Currently Available Data

**USGS NWIS: daily/continuous (‘real-time’)**

<table>
<thead>
<tr>
<th>Process</th>
<th>Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily mean</td>
<td>Discharge/gage</td>
</tr>
<tr>
<td>Daily minimum</td>
<td>Specific conductance</td>
</tr>
<tr>
<td>Daily maximum</td>
<td>pH</td>
</tr>
<tr>
<td>Daily summation</td>
<td>Turbidity</td>
</tr>
<tr>
<td>Continuous (~15 minute interval) – last 120 days</td>
<td>etc.</td>
</tr>
</tbody>
</table>

**USGS NWIS: water quality**

<table>
<thead>
<tr>
<th>Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suspended solids</td>
</tr>
<tr>
<td>Ammonia and ammonium</td>
</tr>
<tr>
<td>Organic nitrogen</td>
</tr>
<tr>
<td>Phosphorus</td>
</tr>
<tr>
<td>etc.</td>
</tr>
</tbody>
</table>

**NOAA: Great Lakes Coastal Forecasting System**

<table>
<thead>
<tr>
<th>Hourly Data</th>
<th>3-Hour Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height above model sea level</td>
<td>Sea water temperature</td>
</tr>
<tr>
<td>E/N water velocity at surface</td>
<td>E/N water velocity</td>
</tr>
<tr>
<td>Air temperature, dew point, cloud cover</td>
<td></td>
</tr>
<tr>
<td>Wave direction, period, height</td>
<td></td>
</tr>
<tr>
<td>E/N air velocity</td>
<td></td>
</tr>
</tbody>
</table>

**National RFC QPE Mosaic**

<table>
<thead>
<tr>
<th>Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation</td>
</tr>
</tbody>
</table>

* National River Forecasting Center Quantitative Precipitation Estimation: - radar-indicated, rain verified, and corrected precipitation estimates
Refine data request

- Choose requested time range
- Pick output format
- Pick output time zone
- Optionally upload temporal filter file
  - data output could match survey data
- Review data choices
Processing Options

Refine data request:
- Choose raw data
- Choose temporal processing
- Choose vector processing
- Choose to download the data now

Or:
- Generate the URL to call the data in other programs

This final URL generation is the key output for running real-time models. Users do not need to use the user-interface each time to gather data, just know and understand the URL.
Interactive Option

Interactive graphs:
- While not useful for running real-time models, EnDDaT also has an option to view the requested data with an interactive graph.
Modeling Capabilities:
Using diverse data in nowcast models

- The user interface on the web is one part of the overall EnDDaT functionality
- The tools on the previous slides are all very useful for discovering and obtaining the data to create nowcast models
- A ‘modeling calculator’ connected to the Great Lakes Beach Health Database was created to efficiently run models
- Output is an email sent to all interested parties
Typical email:

```
Using data from 07/02/2012 10:10:00-05:00 (CDT)

Log10(E.coli): 2.11
Estimated E.coli (MPN/100 mls): 129

Probability of Exceedance: 29.8%

Model Inputs:
GLCFS data: Mean over 24 hours, cloud cover : 0.10833776249
NWIS data: Minimum over 72 hours, river discharge at Manitowoc (04085427) : 70
GLCFS data: Mean over 1 hour, significant wave height : 0.15544589
NWIS data: Minimum over 504 hours, river discharge at Manitowoc (04085427) : 48
GLCFS data: Mean over 336 hours water velocity at surface perpendicular to beach : 0.0314
Manually entered data: Turbidity at beach : 1.99
GLCFS data: St. deviation over 672 hours air temperature : 2.4817095203339127885
Manually entered data: Specific conductance at south storm water outfall : 780

Questions, comments, or concerns? Send them to enddat@usgs.gov
We appreciate your help in making a better tool!
The EnDDAT Team
```

- Students go to sampling site, collect data, input it in a database, and click a button to run the model
- This email is sent to the student and beach manager
- If probability of exceedance is greater than 50%, an advisory sign is displayed at the beach
Virtual Beach 3.0 Integration

- Virtual Beach (VB) is software developed by the EPA and USGS used to develop and run environmental models.
- EnDDaT data import was integrated in the latest version of Virtual Beach (3.0), which will be publicly released later this summer.
VB 3.0 screen shot with EnDDaT link
Operational Models

- The USGS currently runs 9 beach models to predict the probability of threshold exceedence for *E. coli* using Great Lakes Beach Health Database.

- Wisconsin DNR runs 10 models using Virtual Beach 3.0.
Sample Results: Observed vs. Predicted

Evaluation of 2013 Maslowski Beaches Beach Models
Number of Observations = 16

Trending line = good

Cloud of points = not-so-good

Modeling using EnDDaT services

Traditional approach using ‘yesterday’s laboratory result’
Conclusions and Future Work

• Gathering, parsing, sorting, and processing data from multiple sources can be a tricky task...especially for running models on a daily or real-time basis
• EnDDaT simplifies this process for web-accessible data
• *E.coli* models on Great Lake beaches are implemented using EnDDaT to gather data
• Future work:
  – We would like to continue to add data sources as resources become available
  – Improve the efficiencies of the data transfer with existing sources
Acknowledgements

ORPP: Ocean Research Priority Plan
GLRI: Great Lakes Restoration Initiative
USGS Beach Health Program
  -OH, IN, MI, WI,
  NY, PA, IL
U.S. EPA: ORD

Wisconsin DNR

Many local cooperators

Questions:
Laura DeCicco: ldecicco@usgs.gov