

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

Laura De Cicco<sup>1</sup>, David Sibley<sup>1</sup>  
Steve Corsi<sup>2</sup>, Rebecca Carvin<sup>2</sup>  
Nathaniel Booth<sup>3</sup>

<sup>1</sup> USGS Center for Integrated Data Analytics (CIDA)

<sup>2</sup> USGS Wisconsin Water Science Center

<sup>3</sup> USGS Office of Water Information

Speaker: Daniel Sullivan<sup>2</sup>

# 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)

# Introduction

- 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

```
Dataset {
  Float32 lat[ny = 11][nx = 11];
  Float32 dp[time = 601][ny = 11][nx = 11];
} glos/glcfs/michigan/force;
-----
lat[11][11]
[0], 41.606766, 41.607094, 41.6074, 41.607708, 41.608012, 41.608316, 41.60862, 41.608924, 41.609228, 41.609532, 41.609836
[1], 41.624714, 41.625034, 41.62535, 41.625656, 41.625958, 41.62626, 41.626562, 41.626864, 41.627166, 41.627468, 41.62777
[2], 41.64267, 41.64299, 41.643303, 41.643612, 41.643921, 41.64423, 41.644539, 41.644848, 41.645156, 41.645465, 41.645774
[3], 41.660618, 41.66094, 41.661255, 41.66156, 41.661865, 41.66217, 41.662475, 41.66278, 41.663084, 41.663388, 41.663692
[4], 41.67857, 41.678883, 41.6792, 41.679504, 41.679808, 41.680112, 41.680416, 41.68072, 41.681024, 41.681328, 41.681632
[5], 41.69652, 41.696842, 41.69715, 41.697456, 41.69776, 41.698064, 41.698368, 41.698672, 41.698976, 41.69928, 41.699584
[6], 41.71447, 41.71479, 41.715107, 41.715412, 41.715717, 41.716022, 41.716327, 41.716632, 41.716937, 41.717242, 41.717547
[7], 41.732414, 41.732742, 41.733055, 41.73336, 41.733665, 41.73397, 41.734275, 41.73458, 41.734885, 41.73519, 41.735495
[8], 41.75037, 41.750687, 41.751007, 41.751312, 41.751617, 41.751922, 41.752227, 41.752532, 41.752837, 41.753142, 41.753447
[9], 41.76832, 41.76864, 41.76895, 41.769264, 41.76957, 41.769875, 41.77018, 41.770485, 41.77079, 41.771095, 41.7714
[10], 41.786266, 41.786587, 41.7869, 41.78721, 41.787514, 41.787818, 41.788122, 41.788426, 41.78873, 41.789034, 41.789338
dp[601][11][11]
[0][0], -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0
[0][1], -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0
[0][2], -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0, -99999.0
```

csv, tab-delimited

```
OrganizationIdentifier,OrganizationFormalName,ActivityStartTime/Time,ActivityStartTime/Time:
ureValue,ActivityDepthHeightMeasure/Measure
thHeightMeasure/MeasureUnitCode,ActivityBot
vityConductingOrganizationText,MonitoringLoc
hod/MethodIdentifier,SampleCollectionMethod,
nditionText,CharacteristicName,ResultSample
StatisticalBaseCode,ResultValueTypeName,Resu
ue,ResultCommentText,USGSPCode,ResultDepthHe
subjectTaxonomicName,SampleTissueAnatomyName,
thod/MethodName,MethodDescriptionText,Labora
```

XML (WaterML2, WQX)

```
<wml2:Collection xsi:schemaLocation="http://www.opengis.net/wml2"
  <gml:identifier codeSpace="http://nwis.waterdata.usgs.gov/Nwis/collection"
  <gml:name codeSpace="http://nwis.waterdata.usgs.gov/Nwis/collection"
- <wml2:metadata>
  - <wml2:DocumentMetadata gml:id="doc.USGS.MP.USGS"
    - <gml:metaDataProperty xlink:href="contact">
      <gml:GenericMetaData>http://cida.usgs.gov/</gml:GenericMetaData>
    </gml:metaDataProperty>
    <wml2:generationDate>2012-10-09T10:37:58</wml2:generationDate>
    <wml2:version xlink:href="http://www.opengis.net/wml2"
  </wml2:DocumentMetadata>
</wml2:metadata>
- <wml2:observationMember>
  - <om:OM_Observation gml:id="obs.USGS.01446500">
    - <om:metadata>
      - <wml2:ObservationMetadata>
```

- 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)

# EnDDaT Home Page

**USGS**  
science for a changing world

**USGS Home**  
Contact USGS  
Search USGS

## Environmental Data Discovery and Transformation - Version 1.3.15

Access and Integrate Environmental Observations for Coastal Decision Support

### EnDDaT Resources

- Data Discovery
- Introduction
- User Guide
- NWIS Plotting Tool
- Version Updates

### Data Sources

- NWIS Web
- NWIS WaterML2
- GLCFS Data
- Water Quality Portal
- NDBC
- NCDC
- GDP
- Precipitation

### Related Resources

- GLRI
- GLOS

**CIDA**  
Center for Integrated Data Analytics

### Overview

Welcome to the Environmental Data Discovery and Transformation (EnDDaT) service. EnDDaT is a tool used to discover data from our natural environment. This tool accesses data from a variety of data sources, compiles and processes the data, and performs common transformations. The end result is that environmental data from multiple sources is sorted into a single table. See the user guide for step-by-step instructions on obtaining data, specifying transforms, and processing data.

### Motivation

As environmental models have become more intricate and comprehensive, the amount of data necessary to build and run the models has increased significantly. As a result, efficient data discovery, aggregation and processing can be a barrier to environmental modeling efforts. For example, in order to develop near-shore water quality forecasting models, which are often times used to predict bacteria concentrations at recreational beaches, two to five years of historical data is commonly needed for model driven and model predicted parameters. Furthermore, real-time or near real-time data is necessary to run models for accurate and time-relevant forecasting. In order to run the model from the previous example, real-time data with as little lag time as possible (< 6 hours) is necessary in order to predict the bacteria concentrations for that day at a particular beach. To meet these needs associated with environmental modeling, the Environmental Data Discovery and Transformation (EnDDaT) tool was developed with the capabilities of retrieving publicly available data resources through standard Web services, aggregating the disparate data sources, and processing the data through a single Web-accessible user interface. In addition, the tool provides a variety of output formats and data visualization tools. Therefore, these capabilities aid in model development and implementation by allowing scientists to efficiently obtain, aggregate and manipulate the data necessary for these purposes.

### Data Sources

EnDDaT is not the owner or provider of any data. Instead, EnDDaT gathers data from a variety of data providers. The data providers are listed on the left side bar. EnDDaT has been designed especially to gather data that uses recognized web standards such as SOS, WQX, and Thredds. However, if data is deemed useful for environmental modeling, custom data gathering tools and data parsers can be included.

### Feedback

Questions, comments, and requests are welcome. Please email [enddat@usgs.gov](mailto:enddat@usgs.gov)

- <http://cida.usgs.gov/enddat/>
- Home page lets you jump right into Data Discovery
- Or, go to the User Guide

# Data Discovery

The screenshot shows the EnDDaT web application interface. At the top, there are three tabs: "Choose Data", "Create Project Location", and "EnDDaT Information". The "Choose Data" tab is active, showing a list of data sources with checkboxes. The "Create Project Location" tab is also visible. Below the tabs, there is a search bar with a "5" in a box, indicating a 5-mile bounding box. The main area is a map of the Great Lakes region, showing various data points and locations. A blue arrow points from the "5" in the search bar to the map. Another blue arrow points from the "Active single point" tab to the "Available Data" table.

**Choose Data** **Create Project Location** EnDDaT Information

USGS Time Series (NWIS, Plotting Tool)  National Data Buoy Center (NDBC)  6-hr Quantitative National Precipitation  
 Great Lakes Coastal Forecasting System (GLCFS)  National Climate Data Center (NCDC)  1-hr Quantitative North Central Precipitation  
 USGS Water Quality (Water Quality)  Surface Summary of Day Data (SOD)

Search within a 5 mile bounding box from selected project (click marker to identify)

**Map** Satellite Hybrid Terrain

**NWIS** **GLCFS** Water Quality NDBC NCDC SOD Precipitation

**Overview** **Active single point**

**Active Grid Point:**  
Approximate distance to active beach: 1.044 miles

x	y	Latitude	Longitude
138	16	44.089813	-87.6397

**Available Data:**  
Great Lakes Coastal Forecasting System (GLCFS), NOWCAST  
GLERL Wave Model-Great Lakes, Princeton Ocean Model-Great Lakes

Property	Unit
<input type="checkbox"/> Ice Concentration	%
<input type="checkbox"/> Height Above Model Sea Level	m
<input type="checkbox"/> Ice Thickness	m
<input type="checkbox"/> Eastward Water Velocity at Surface	m/s
<input type="checkbox"/> Ice u-Velocity	m/s

Submit

- Points show up on the map at the location of available data
- Set a bounding box area
- Information on the available data is displayed in a table
- Choose data to discover

# Currently Available Data

## USGS NWIS: daily/continuous ('real-time')

Process	Property
Daily mean	Discharge/gage
Daily minimum	Specific conductance
Daily maximum	pH
Daily summation	Turbidity
Continuous (~15 minute interval) – last 120 days	etc.

## USGS NWIS: water quality

Property
Suspended solids
Ammonia and ammonium
Organic nitrogen
Phosphorus
etc.

## NOAA: Great Lakes Coastal Forecasting System

Hourly Data	3-Hour Data
Height above model sea level	Sea water temperature
E/N water velocity at surface	E/N water velocity
Air temperature, dew point, cloud cover	
Wave direction, period, height	
E/N air velocity	

## National RFC QPE Mosaic\*

Property
Precipitation

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



# Refine data request

**Start Date:** 2011-08-21   
**End Date:** 2011-09-21

**Output date format:** MM/DD/YYYY HH:MM  
**Output time zone:** EDT (-4:00 GMT)

**Acceptable data gap:** 0 hours  
**Filter file ID:**

**Optional:** Upload a file with dates and times to filter output.  
File should be a text file containing a single column of dates in the form: mm/dd/yyyy, hh:mm  
See the [User Guide](#) for detailed information.

**NWIS:**

	Property Name	Site	Property	Offering	Start Date	End Date
<input checked="" type="checkbox"/>	Gage height, feet Daily Instantaneous	453345084401501	00065	00011	1994-10-01	2011-09-21

**Instantaneous Data Archive:**  
No IDA data requested

**Water Quality:**  
No QW data requested

**GLCFS:**

	Property	X	Y	Sigma	Start Date	End Date
<input checked="" type="checkbox"/>	Eastward Water Velocity at Surface	155	17		2006-01-01	2012-10-08
<input checked="" type="checkbox"/>	Northward Water Velocity at Surface	155	17		2006-01-01	2012-10-08
<input checked="" type="checkbox"/>	Significant Wave Height	155	17		2006-01-01	2012-10-08
<input checked="" type="checkbox"/>	Air Temperature	155	17		2006-01-01	2012-10-08
<input checked="" type="checkbox"/>	Cloud Cover	155	17		2006-01-01	2012-10-08

- 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

Choose Data and Data Processes | Beach Orientation Calculator

**Available Data**  
Add Selected | Add All | Refresh

Discharge, cubic feet per second Instantaneous: 04085427  
Specific conductance, water, unfiltered, microsiemens per cm  
Total precip - 1 hr [211,327]  
Height above sea level: [16,137]  
E v\_water - surface: [16,137]  
N v\_water - surface: [16,137]  
Angle: v\_water (surface): [16,137]  
Magnitude: v\_water (surface): [16,137]

**Requested Data**  
Remove Selected | Remove All | Refresh

Discharge, cubic feet per second Instantaneous: 04085427

**Data Processing\***

Process	Over: <input checked="" type="radio"/> Hours <input type="radio"/> Days <input type="radio"/> Weeks
<input type="checkbox"/> Mean ( $\mu$ )	<input type="checkbox"/> 1 hour
<input type="checkbox"/> Minimum (Min)	<input type="checkbox"/> 2 hours
<input type="checkbox"/> Maximum (Max)	<input type="checkbox"/> 6 hours
<input type="checkbox"/> Summation ( $\Sigma$ )	<input type="checkbox"/> 12 hours
<input type="checkbox"/> Difference ( $\Delta$ )	<input type="checkbox"/> 24 hours
<input type="checkbox"/> Max Diff (Max $\Delta$ )	
<input type="checkbox"/> St. Dev. ( $\sigma$ )	

**Resolve Velocity Vectors with Beach Orientation**

Direction	Orientation
<input type="radio"/> 90 deg	<input type="checkbox"/> Parallel (  )
<input checked="" type="radio"/> 45.000 deg	<input type="checkbox"/> Perpendicular( $\perp$ )

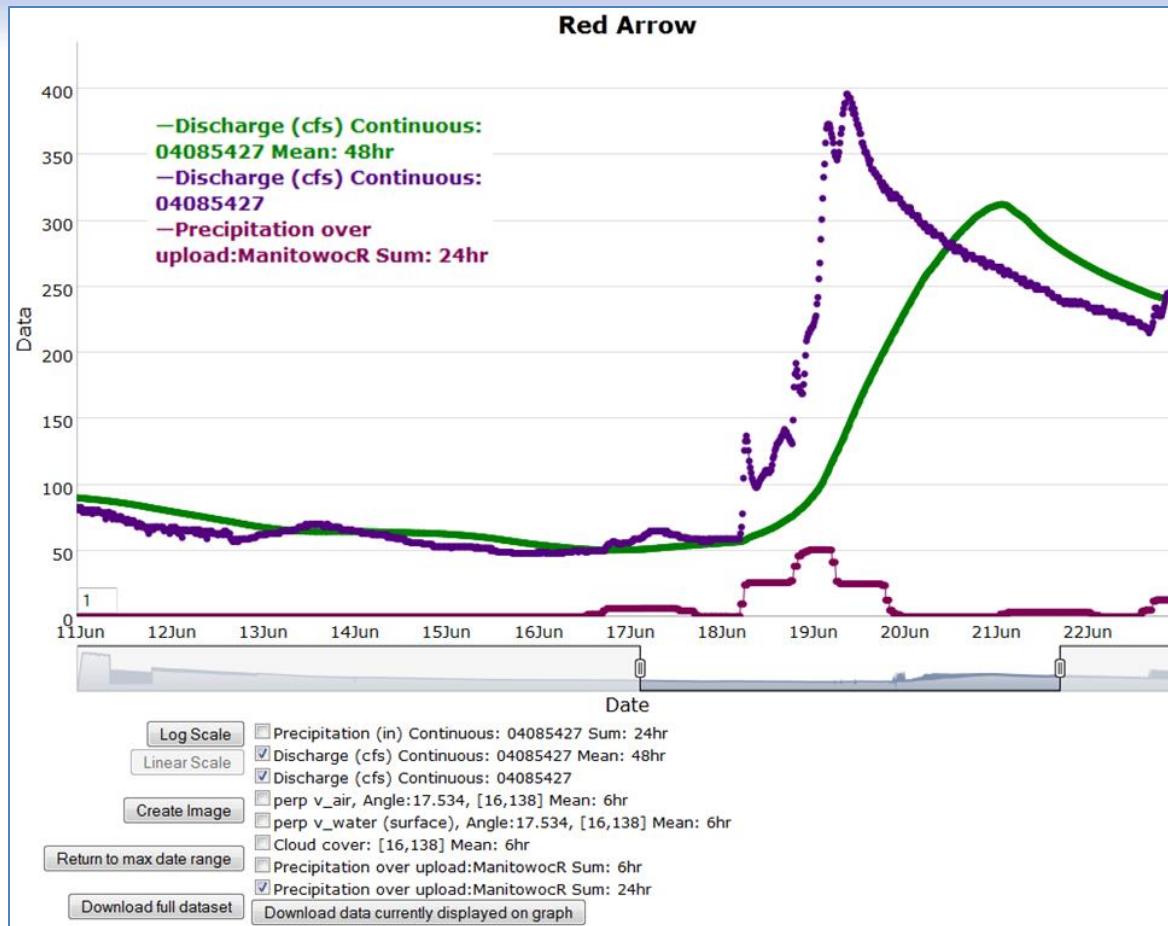
**Output Style**  
Tab Delimited Text |  As Download

**Missing Value Fill:**

Get Data | Generate URL

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

# Beach Model Email

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

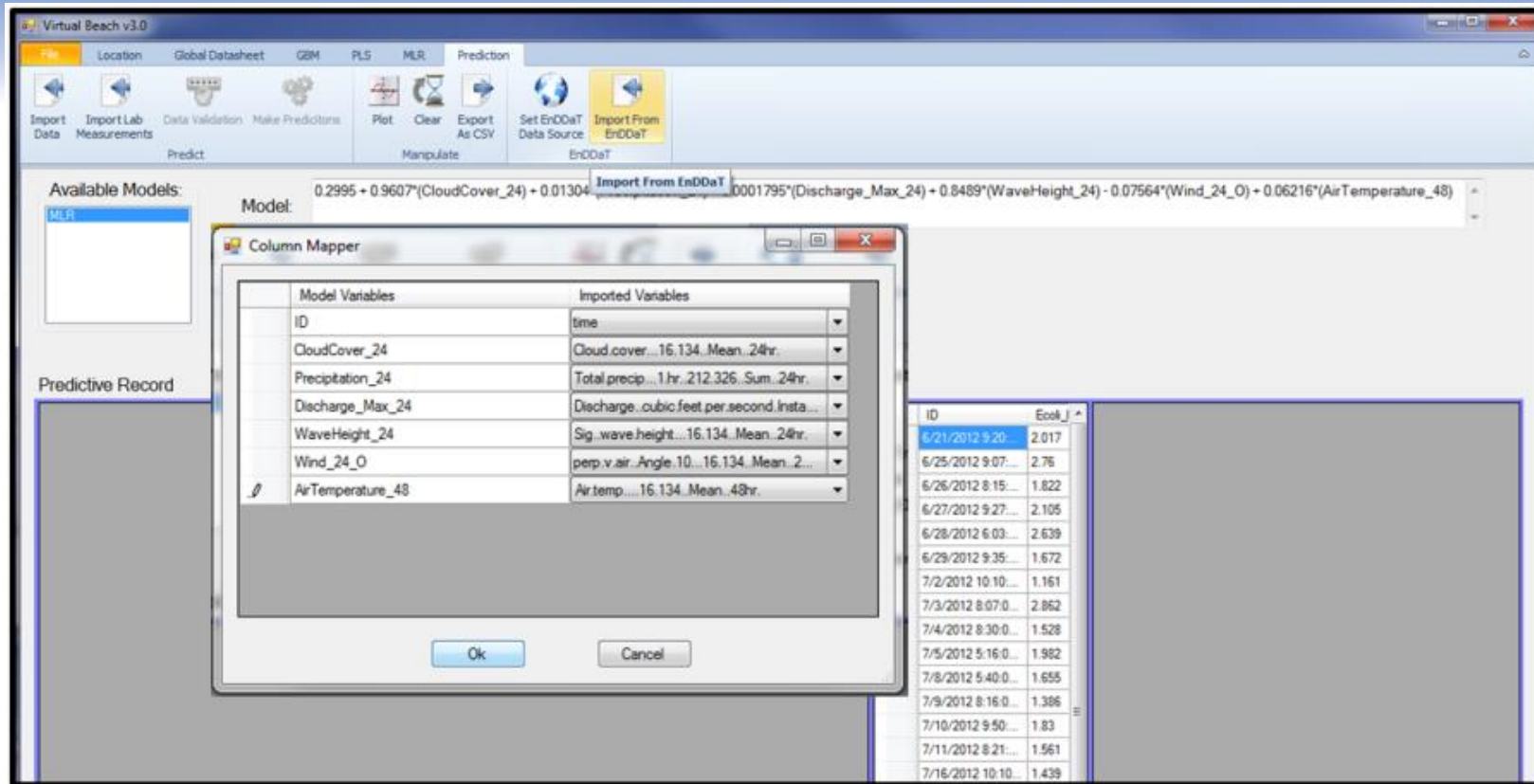
Probability of Exceedance: 29.8%

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

The screenshot displays the Virtual Beach v3.0 software interface. The main window shows a menu bar with options like File, Location, Global Datasheet, GBM, PLS, MLR, and Prediction. Below the menu is a toolbar with icons for Import Data, Import Lab Measurements, Data Validation, Make Predictions, Plot, Clear, Export As CSV, Set EnDDaT Data Source, and Import From EnDDaT. The main area shows a model equation:  $0.2995 + 0.9607 * (\text{CloudCover}_{24}) + 0.01304 * (\text{Precipitation}_{24}) + 0.0001795 * (\text{Discharge\_Max}_{24}) + 0.8489 * (\text{WaveHeight}_{24}) - 0.07564 * (\text{Wind}_{24\_O}) + 0.06216 * (\text{AirTemperature}_{48})$ . A 'Column Mapper' dialog box is open, mapping model variables to imported variables. A data table is visible on the right side of the interface.

**Model:**  $0.2995 + 0.9607 * (\text{CloudCover}_{24}) + 0.01304 * (\text{Precipitation}_{24}) + 0.0001795 * (\text{Discharge\_Max}_{24}) + 0.8489 * (\text{WaveHeight}_{24}) - 0.07564 * (\text{Wind}_{24\_O}) + 0.06216 * (\text{AirTemperature}_{48})$

**Column Mapper**

Model Variables	Imported Variables
ID	time
CloudCover_24	Cloud.cover...16.134..Mean..24hr.
Precipitation_24	Total.precip...1.hr..212.326..Sum..24hr.
Discharge_Max_24	Discharge..cubic.feet.per.second.Insta...
WaveHeight_24	Sig..wave.height...16.134..Mean..24hr.
Wind_24_O	perp.v.air..Angle.10...16.134..Mean..2...
AirTemperature_48	Airtemp....16.134..Mean..48hr.

**Available Models:** MLR

**GLCFS** (pointing to CloudCover\_24, Precipitation\_24, Discharge\_Max\_24, WaveHeight\_24, Wind\_24\_O)

**NWS NCPFC** (pointing to CloudCover\_24)

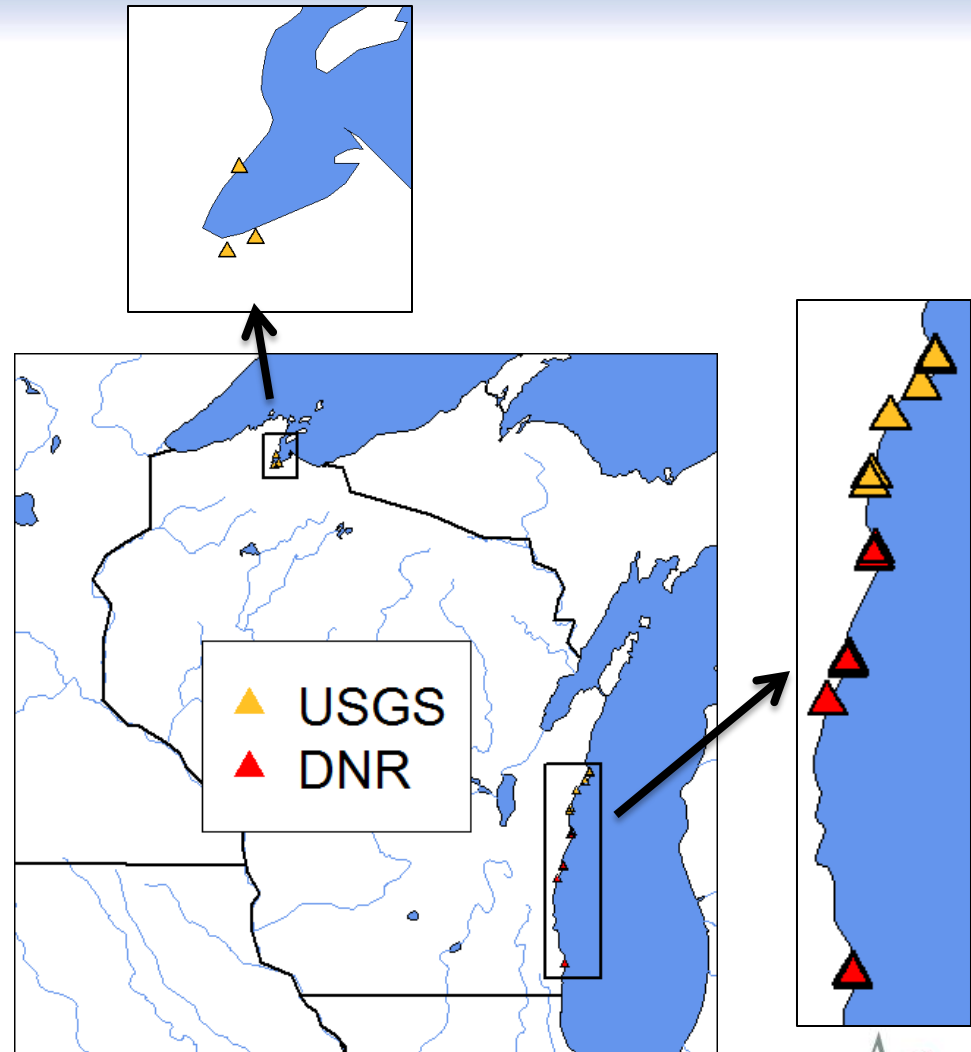
**NWIS** (pointing to AirTemperature\_48)

ID	Ecoli_J
6/21/2012 9:20:...	2.017
6/25/2012 9:07:...	2.76
6/26/2012 8:15:...	1.822
6/27/2012 9:27:...	2.105
6/28/2012 6:03:...	2.639
6/29/2012 9:35:...	1.672
7/2/2012 10:10:...	1.161
7/3/2012 8:07:0...	2.862
7/4/2012 8:30:0...	1.528
7/5/2012 5:16:0...	1.982
7/8/2012 5:40:0...	1.655
7/9/2012 8:16:0...	1.386
7/10/2012 9:50:...	1.83
7/11/2012 8:21:...	1.561
7/16/2012 10:10:...	1.439



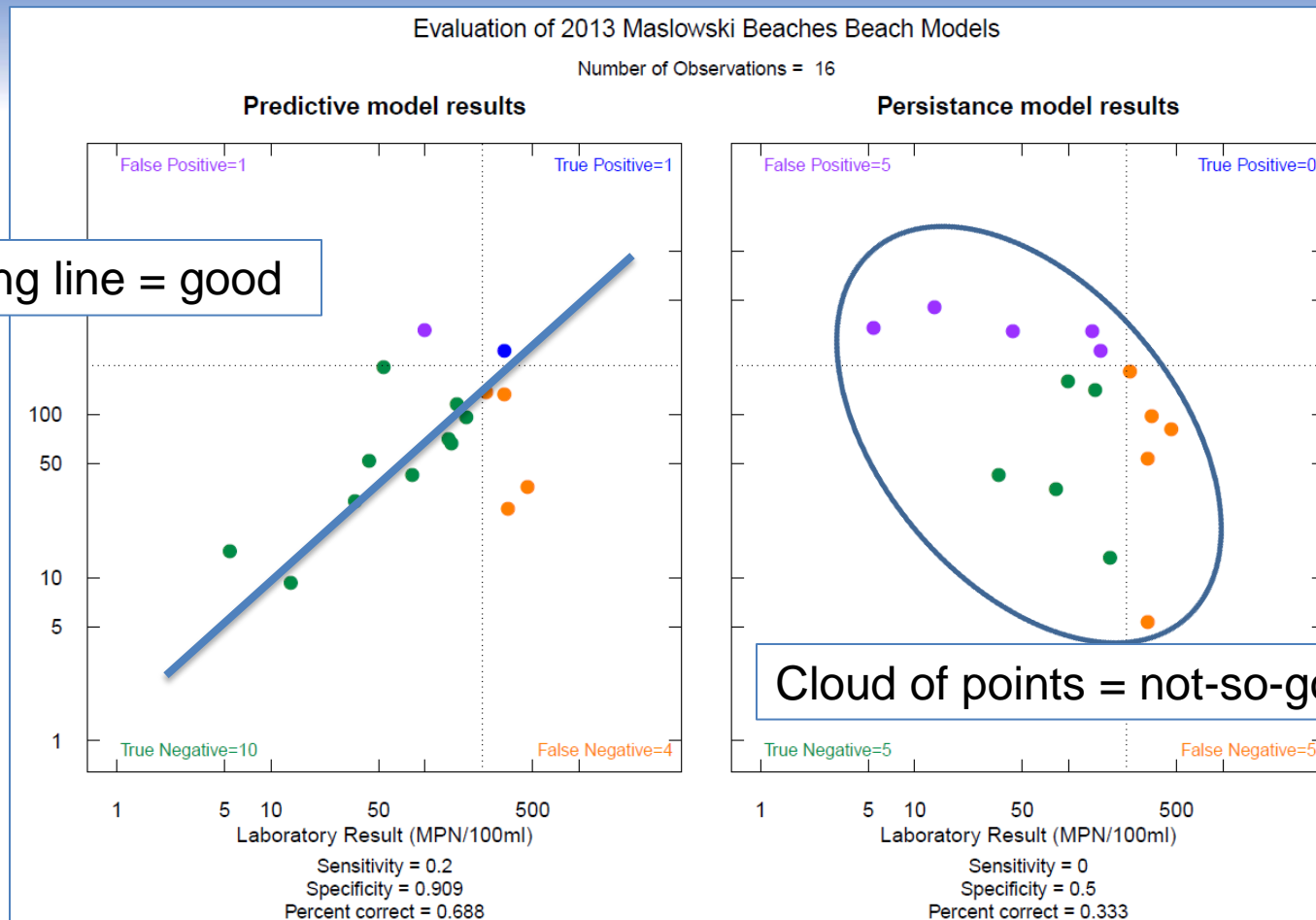
# Operational Models

- 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



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](mailto:ldecicco@usgs.gov)