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

Laura De Cicco[1], David Sibley[1]
Steve Corsi[2], Rebecca Carvin[2]
Nathaniel Booth[3]

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

Speaker: Daniel Sullivan[2]

USGS science for a changing world

CIDA Center for Integrated Data Analytics

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
- **En**vironmental **D**ata **D**iscovery **a**nd **T**ransformation (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
[1], 41.624714, 41.625034, 41.62535, 41.625656, 4
[2], 41.64267, 41.64299, 41.643303, 41.643612, 41
[3], 41.660618, 41.66094, 41.661255, 41.66156, 41
[4], 41.67857, 41.678883, 41.6792, 41.679504, 41.
[5], 41.69652, 41.696842, 41.69715, 41.697456, 41
[6], 41.71447, 41.71479, 41.715107, 41.715412, 41
[7], 41.732414, 41.732742, 41.733055, 41.73336, 4
[8], 41.75037, 41.750687, 41.751007, 41.751312, 4
[9], 41.76832, 41.76864, 41.76895, 41.769264, 41.
[10], 41.786266, 41.786587, 41.7869, 41.78721, 41

dp[601][11][11]
[0][0], -99999.0, -99999.0, -99999.0, -99999.0, -
[0][1], -99999.0, -99999.0, -99999.0, -99999.0, -
[0][2], -99999.0, -99999.0, -99999.0, -99999.0, -
```

csv, tab-delimited

```
OrganizationIdentifier,OrganizationFormalNar
tivityStartTime/Time,ActivityStartTime/Time
ureValue,ActivityDepthHeightMeasure/Measure
thHeightMeasure/MeasureUnitCode,ActivityBot
vityConductingOrganizationText,MonitoringLo
hod/MethodIdentifier,SampleCollectionMethod,
nditionText,CharacteristicName,ResultSample
StatisticalBaseCode,ResultValueTypeName,Res
ue,ResultCommentText,USGSPCode,ResultDepthH
ubjectTaxonomicName,SampleTissueAnatomyName,
thod/MethodName,MethodDescriptionText,Labor
```

XML (WaterML2, WQX)

```
- <wml2:Collection xsi:schemaLocation="http://www.opengis.ne
    <gml:identifier codeSpace="http://nwis.waterdata.usgs.gov/N
    <gml:name codeSpace="http://nwis.waterdata.usgs.gov/NJ/nv
  - <wml2:metadata>
    - <wml2:DocumentMetadata gml:id="doc.USGS.MP.USGS
      - <gml:metaDataProperty xlink:href="contact">
          <gml:GenericMetaData>http://cida.usgs.gov</gml:G
        </gml:metaDataProperty>
        <wml2:generationDate>2012-10-09T10:37:58</wml2:g
        <wml2:version xlink:href="http://www.opengis.net/wate
      </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

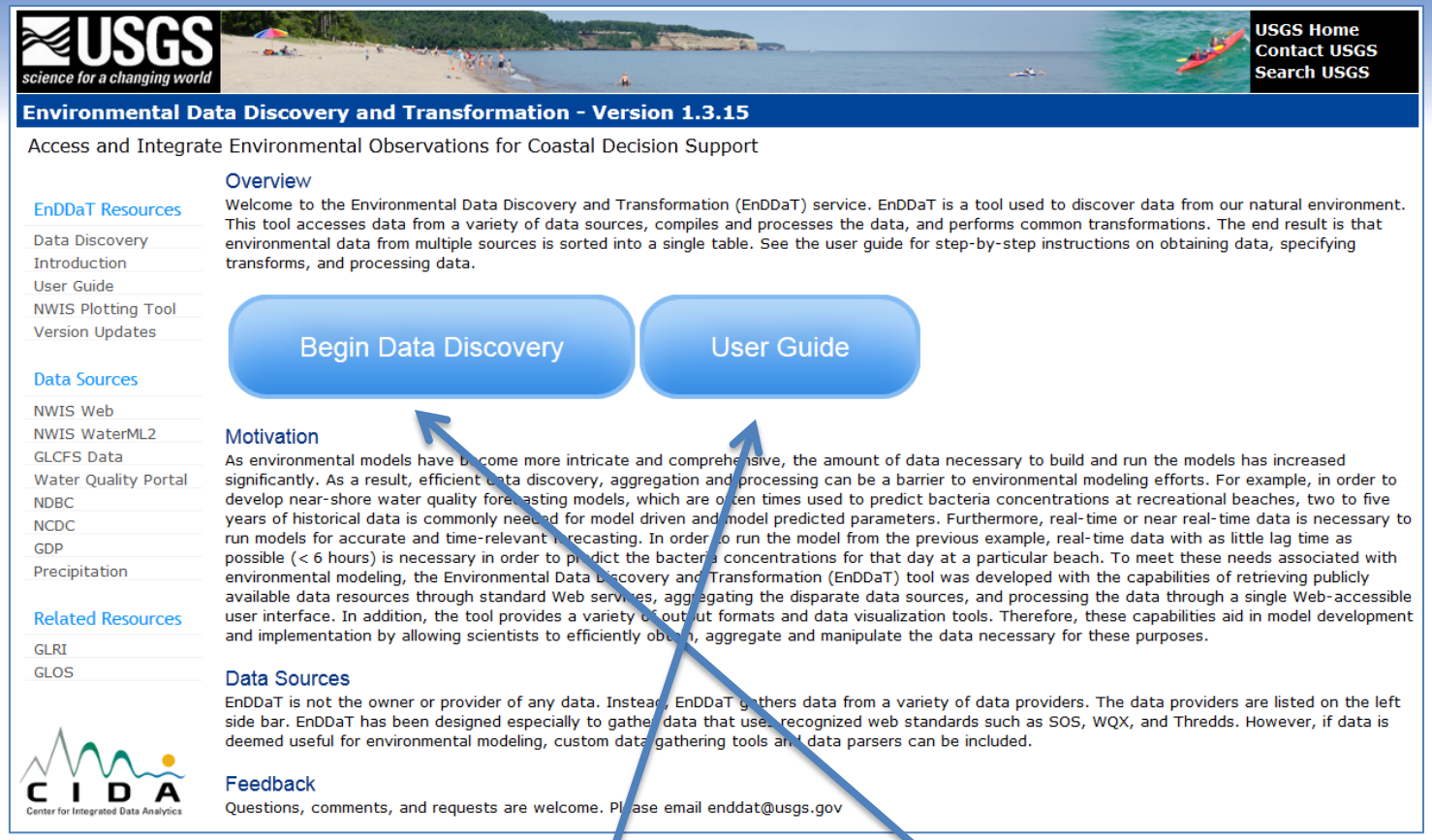
USGS science for a changing world

CIDA Center for Integrated Data Analytics

EnDDaT Introduction

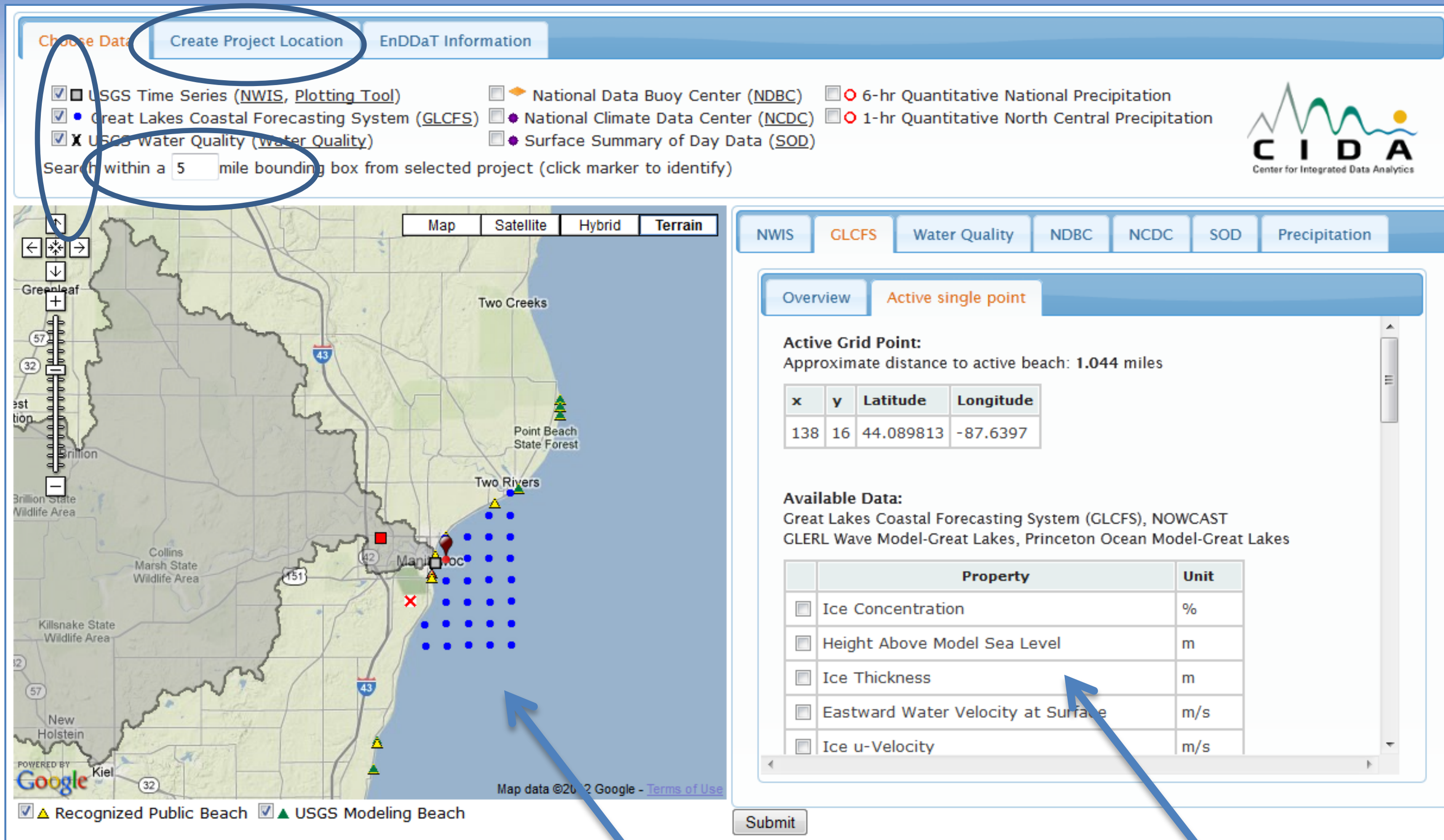
- Discover available data within a specified area around a point location
- Gather data from multiple sources
- Sort data
- Process data
  - resolve vectors based on orientation
  - moving window: mean, minimum, maximum, summation
- Export data (tab or csv delimited, interactive graph)

EnDDaT Home Page



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

Data Discovery



- Choose a project location
- Points shown 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

USGS science for a changing world

C I D A Center for Integrated Data Analytics

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*

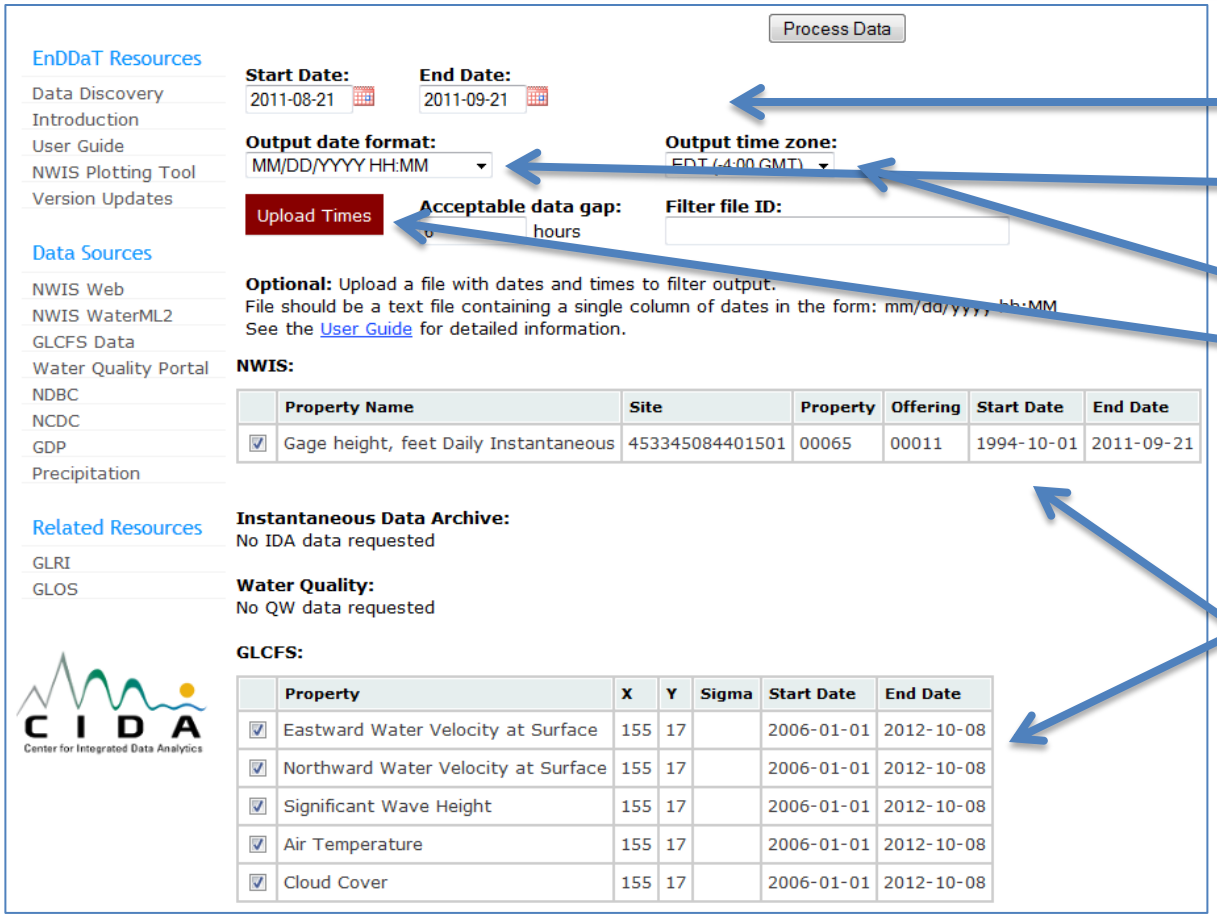
Property
Precipitation

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

USGS science for a changing world

CIDA Center for Integrated Data Analytics

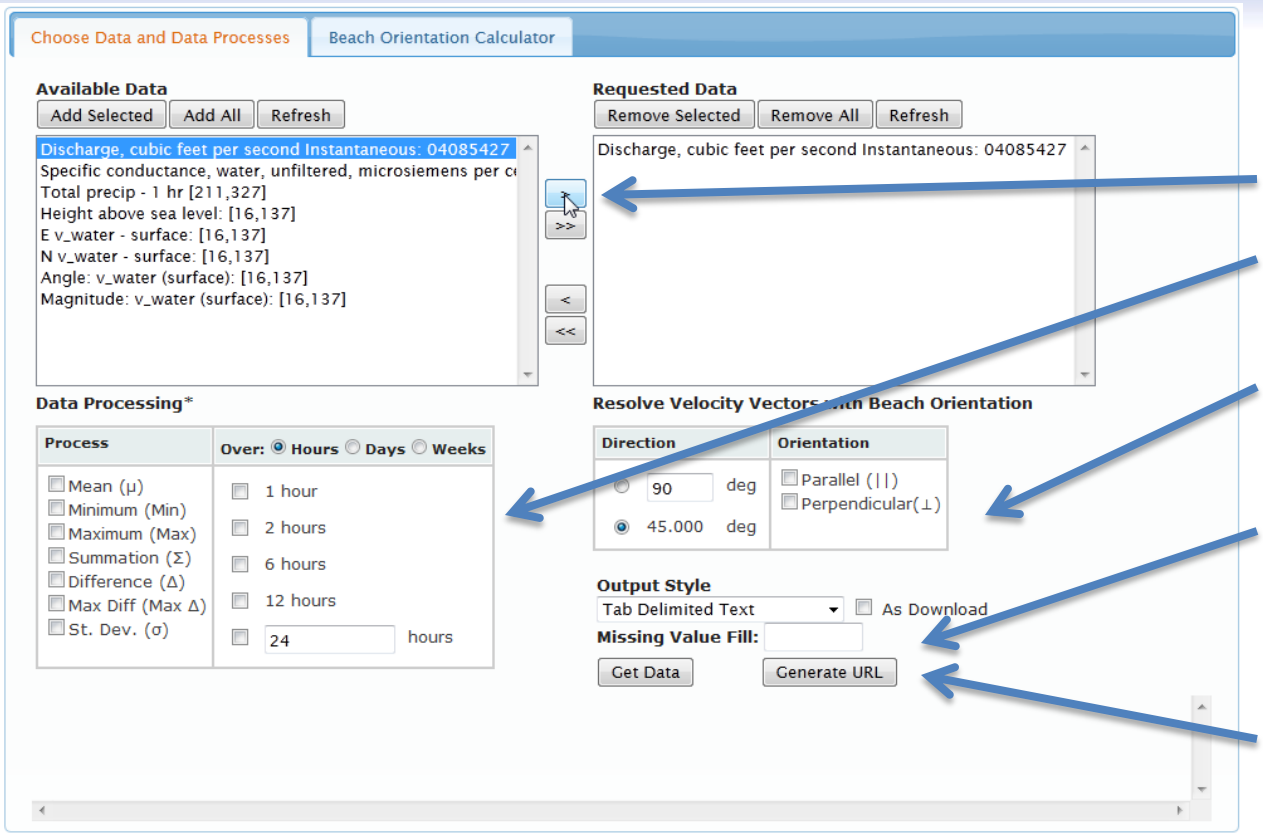
Refine data request



- 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

USGS science for a changing world

Processing Options



Refine data request:

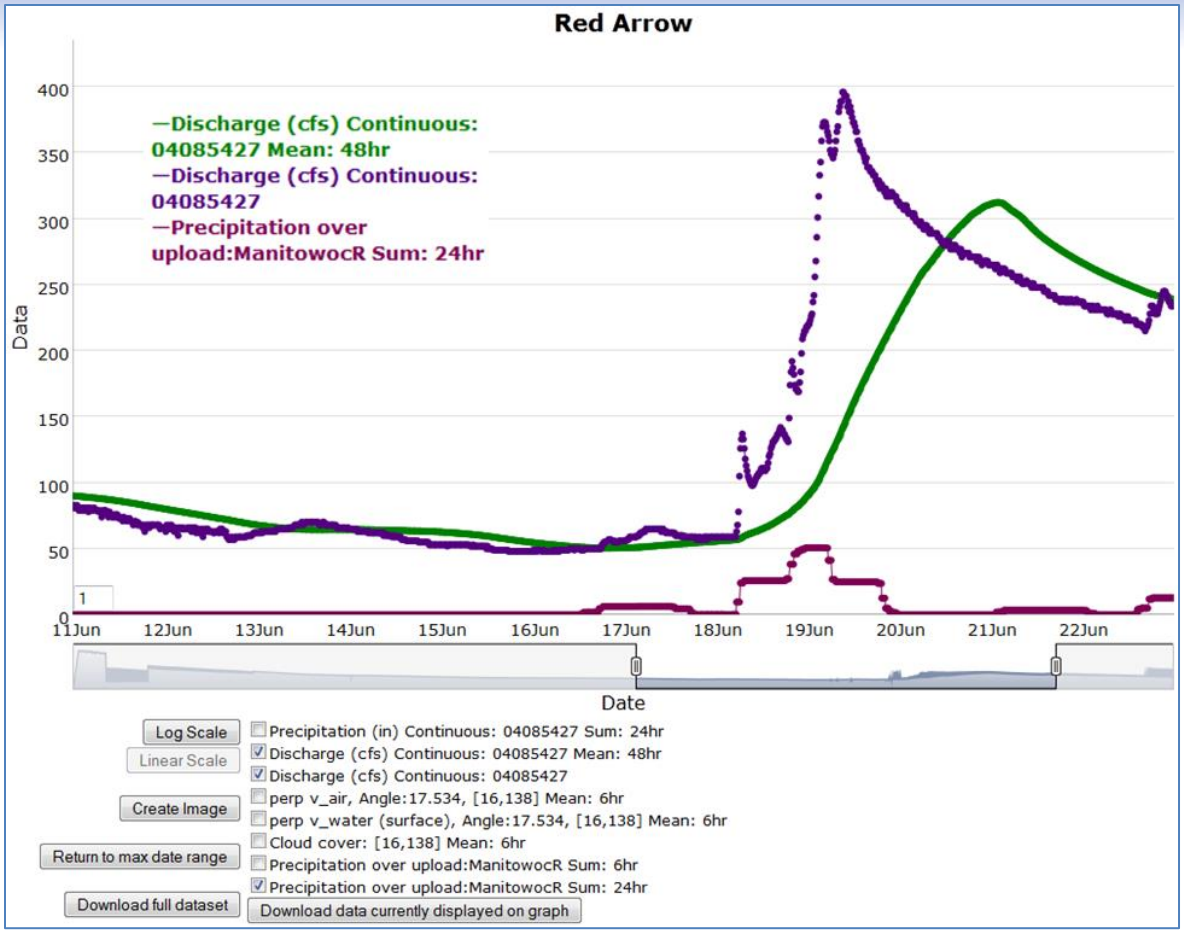
- Choose raw data
- Choose temporal processing
- Choose vector processing
- Choose to download the data now

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

USGS science for a changing world

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

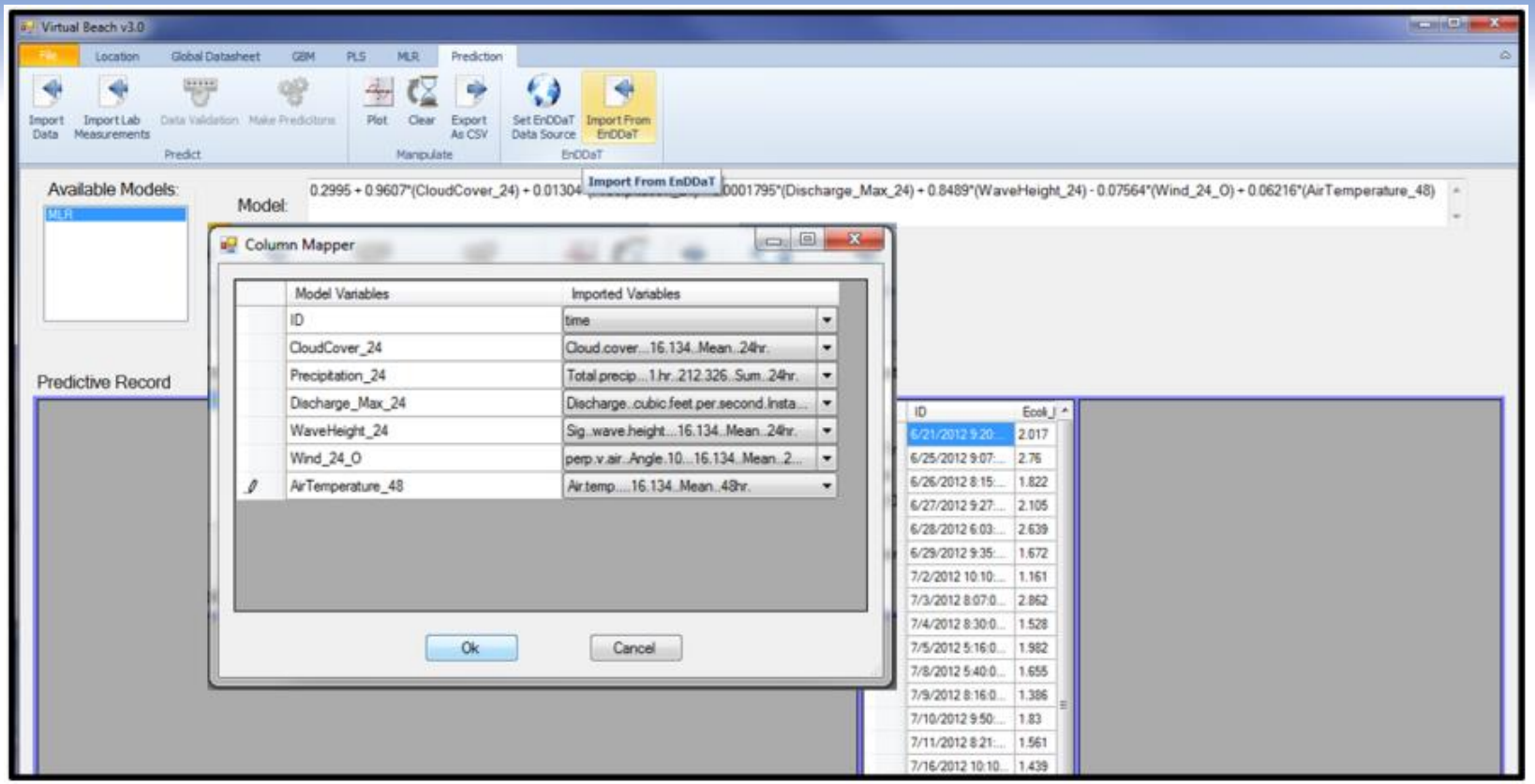
Beach Model Email

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
```

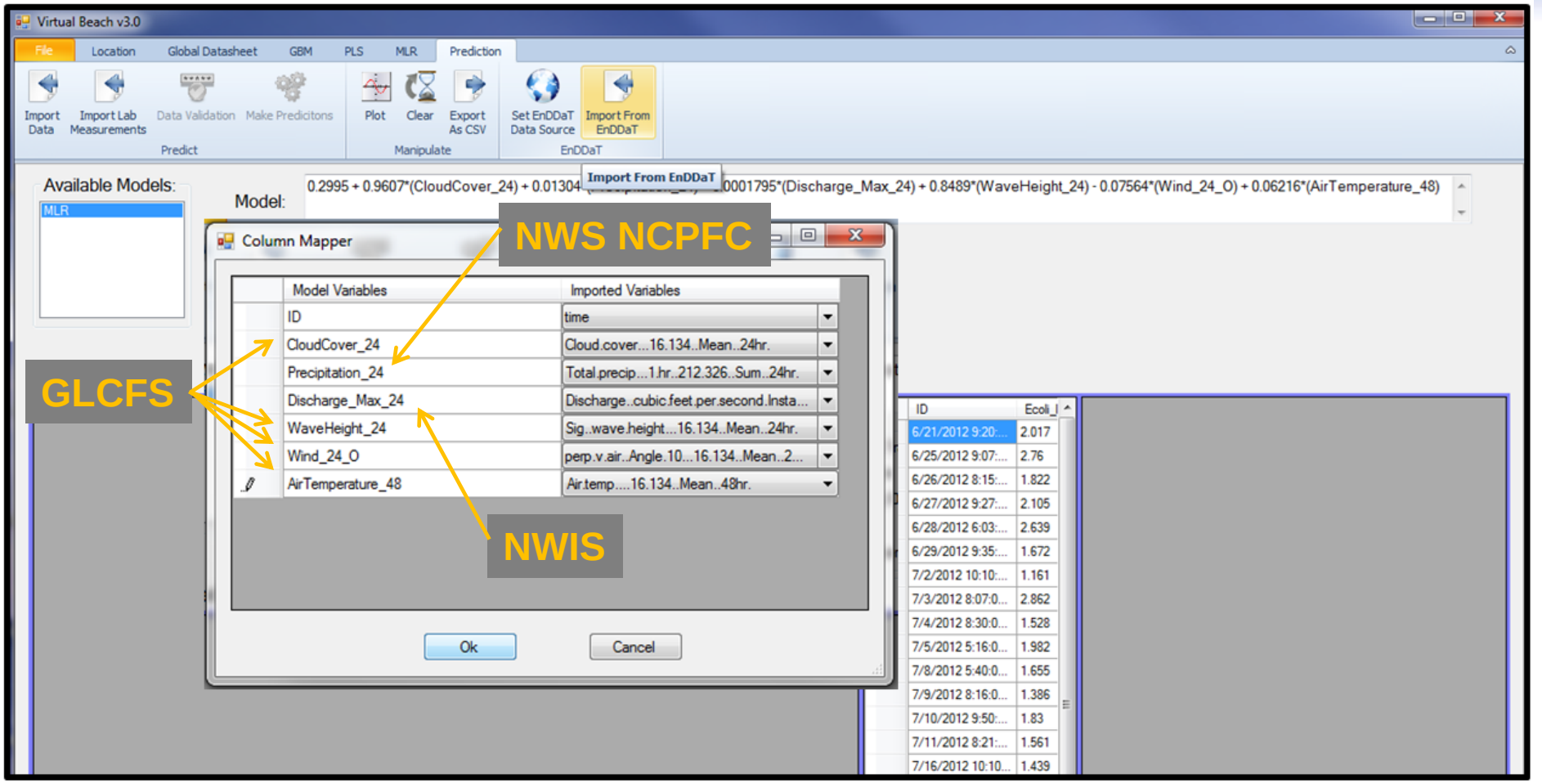
- Students go to sampling site, collect data, input it in a database, and click a button to run the model
- This email is sent to the student and beach manager
- If probability of exceedance is greater than 50%, an advisory sign is displayed at the beach

Virtual Beach 3.0 Integration



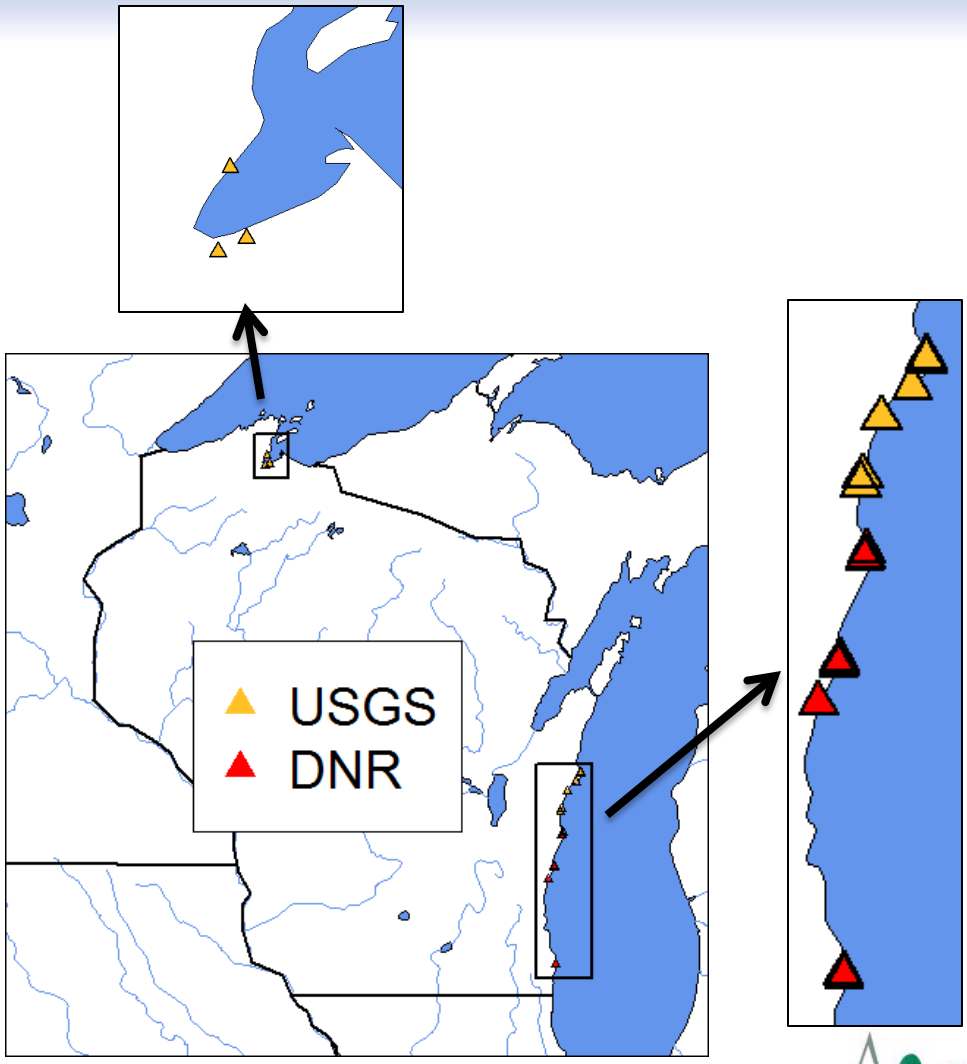
- Virtual Beach (VB) is software developed by the EPA and USGS used to develop and run environmental models.
- EnDDaT data import was integrated in the latest version of Virtual Beach (3.0), which will be publicly released later this summer

VB 3.0 screen shot with EnDDaT link



Operational Models

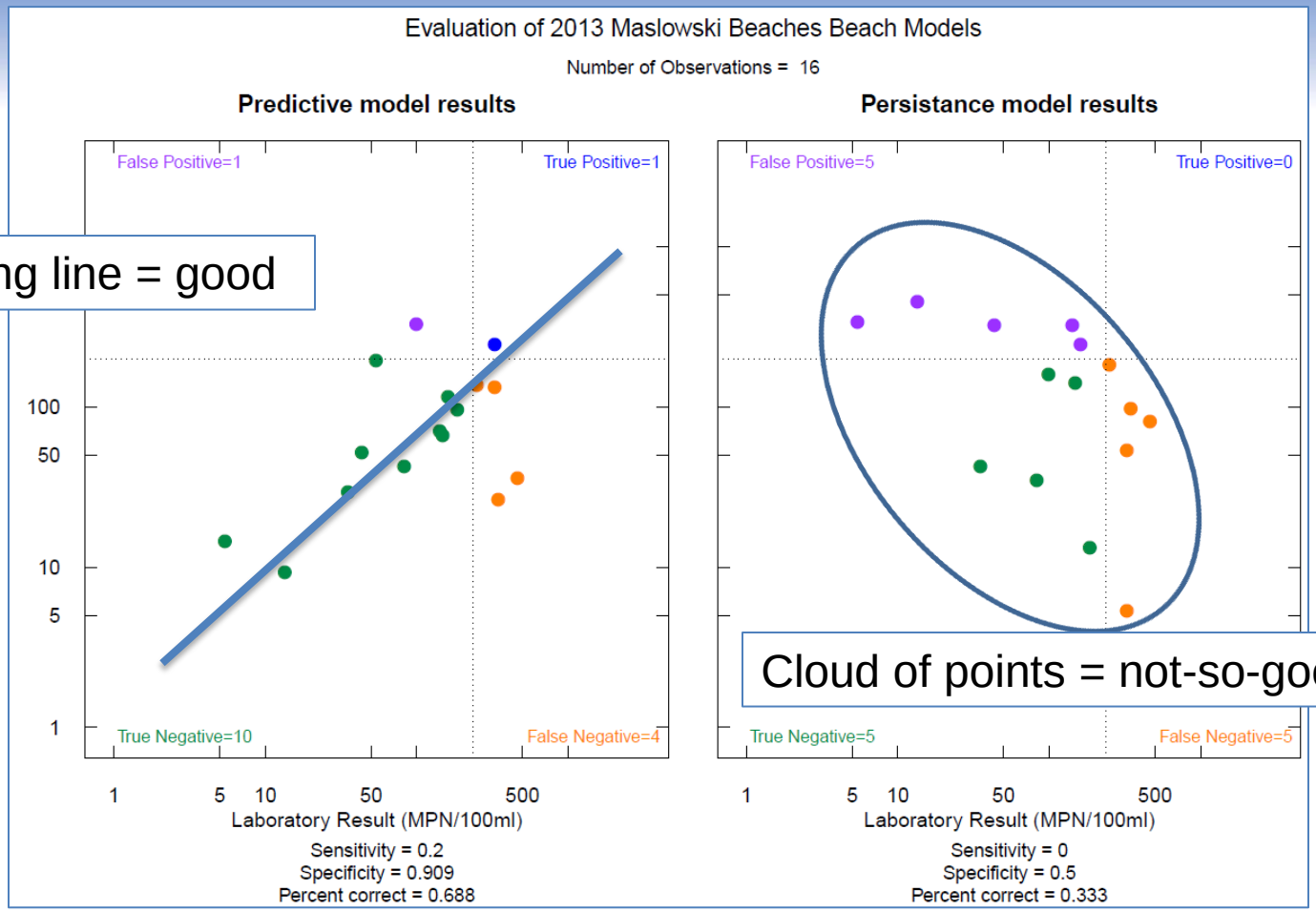
- The USGS currently runs 9 beach models to predict the probability of threshold exceedence for *E.coli* using Great Lakes Beach Health Database
- Wisconsin DNR runs 10 models using Virtual Beach 3.0



USGS science for a changing world

CIDA Center for Integrated Data Analytics

Sample Results: Observed vs. Predicted



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

USGS science for a changing world

CIDA Center for Integrated Data Analytics