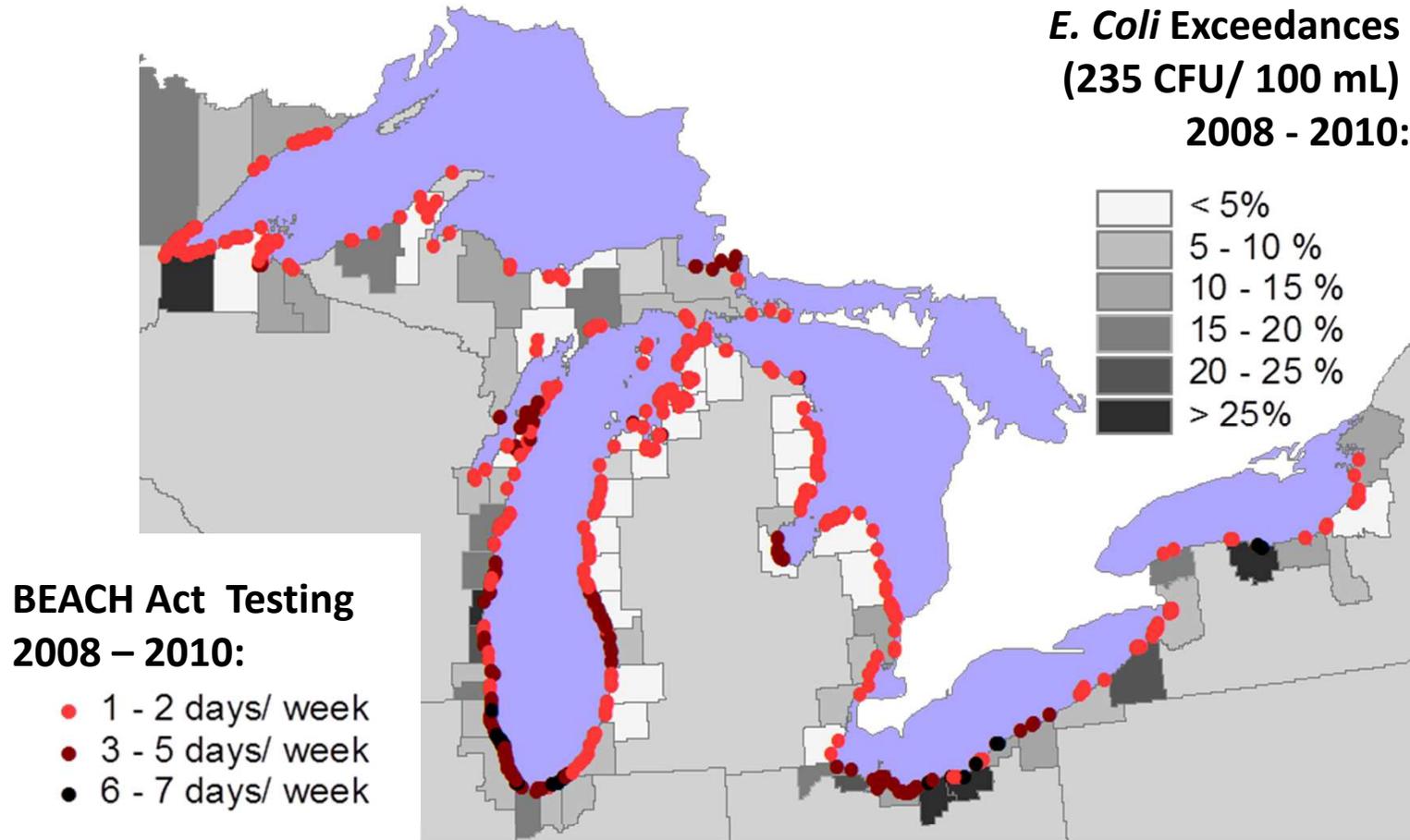


# Implementing Predictive Models: Practical Advice and New Tools

Adam Mednick — University of Wisconsin Sea Grant Institute



# Background



# Problem:

“We issue our advisories on the wrong day.”

Confirmed by BEACON data (2008-2010):

	Sampled Advisories:		Sampled Open:		All Sampled Days:	
	All	In-Error	All	In-Error	All	Total Errors
Illinois	1,107	709 (64%)	9,127	1,142 (13%)	10,234	1,851 (18%)
Indiana	716	452 (63%)	3,881	555 (14%)	4,597	1,007 (22%)
Michigan	201	137 (68%)	7,020	528 (8%)	7,221	665 (9%)
Minnesota	74	58 (78%)	1,853	90 (5%)	1,927	148 (8%)
New York	467	260 (56%)	2,322	440 (19%)	2,789	700 (25%)
Ohio	636	434 (68%)	4,842	770 (16%)	5,478	1,204 (22%)
Pennsylvania	72	59 (82%)	937	88 (9%)	1,009	147 (15%)
Wisconsin	1,225	849 (69%)	9,868	951 (10%)	11,093	1,800 (16%)



General Empirical Model



# Virtual Beach



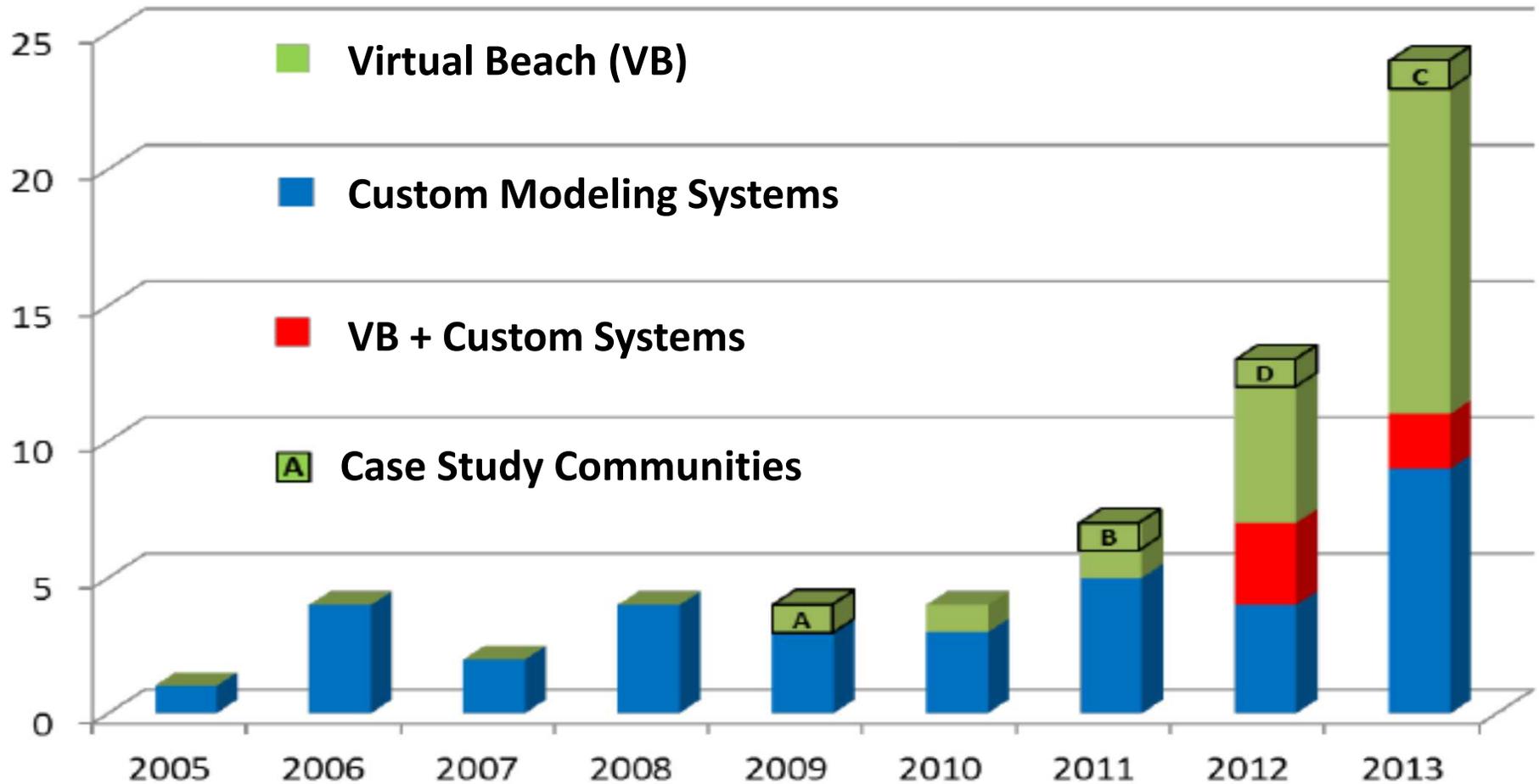
### Disclaimer:

This software has been reviewed in accordance with review policies and approved for publication. No endorsement or recommendation for use.



Palmer Luckey, 22, inventor of the Oculus Rift, is one of the visionaries making virtual reality mainstream

# Communities *using* 'Nowcast' Models



# Case Study Communities

- Community A ('Innovator')

*“The intent... is that we can rely on Virtual Beach more fully – so we don’t spend as much time and money on testing.”*

- Community B ('Early Adopter')

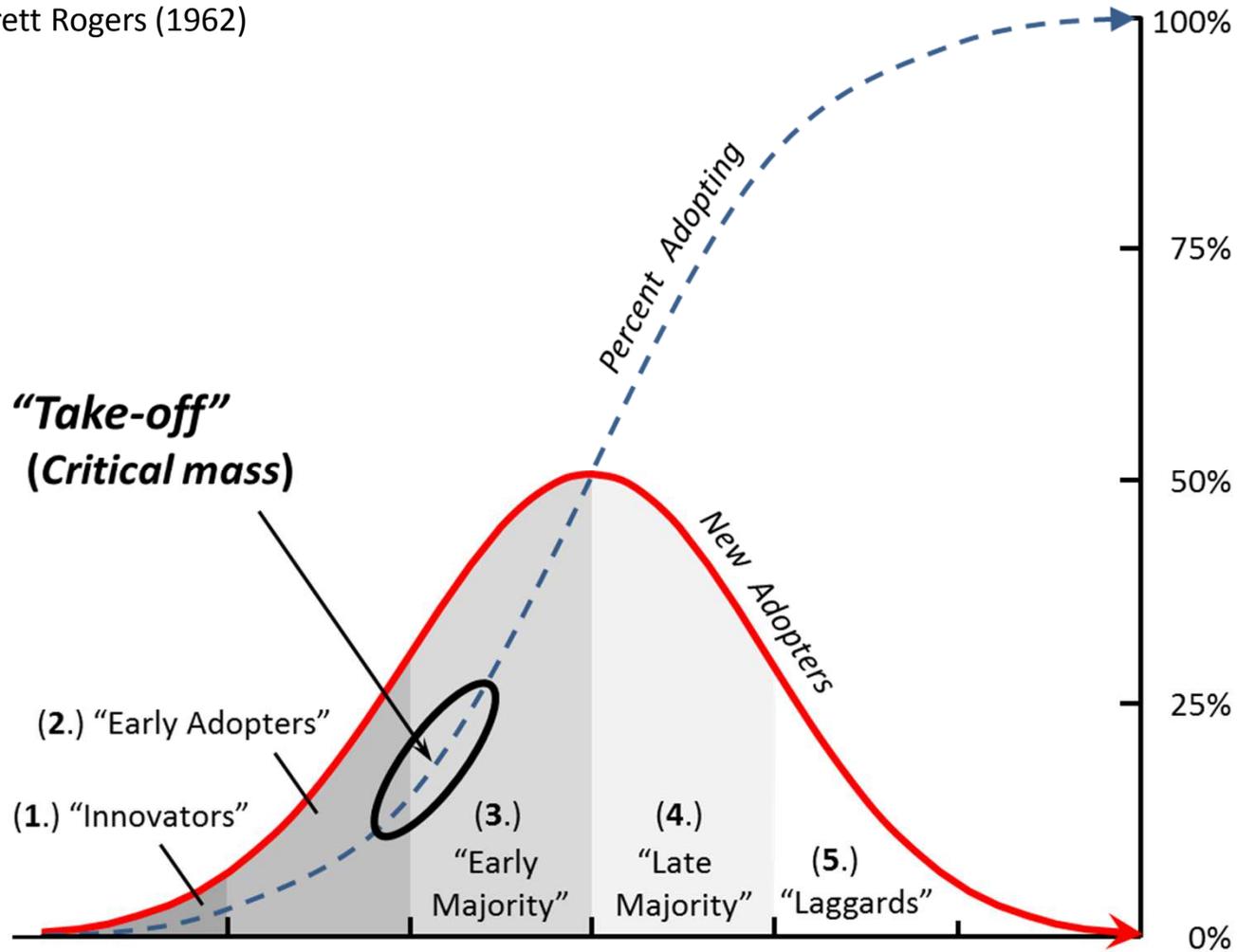
*“We [experimented with VB] year-to-year [to see] how well it performed... how many times it [was] right or wrong...”*

- Community C ('Early Majority')

*“I went to my Administration and said... ‘There’s some guy from Madison who’s really pushing [Virtual Beach], and I’ve heard [a neighboring community] is using it’...”*

# Are we there yet?

Based on Everett Rogers (1962)



# Barriers (from Rockwell et al. 2014)

- Limited Staff/ Funding
  - **86%** of ‘beach managers’ spend less than 1/4 of their time on beach-related work
  - **76%** of health dept’s covering >5 beaches devote < 10% of overall staff time to beach-related work
  - **38%**... reported that summer interns conduct > 50% of dept’s beach-related work

NOAA Technical Memorandum GLERL-163

**Beach Health Information Needs Assessment:**

9 Years Later - Results from Follow-up Survey

David Rockwell<sup>1</sup>  
Sonia Joseph Joshi<sup>2</sup>  
Adam Mednick<sup>3</sup>  
Holly Wirrick<sup>4</sup>

<sup>1</sup>Cooperative Institute for Limnology and Ecosystems Research, 4840 S. State Rd., Ann Arbor, MI 48108

<sup>2</sup>Michigan Sea Grant, and NOAA Great Lakes Environmental Research Laboratory, 4840 S. State Rd., Ann Arbor, MI 48108

<sup>3</sup>University of Wisconsin Sea Grant Institute, 1975 Willow Dr., Madison, WI 53706

<sup>4</sup>USEPA, Region V, Water Division, 77 W. Jackson Blvd., Chicago Illinois 60604

Beach Health Interagency Coordination Team in cooperation with:

NOAA, Great Lakes Environmental Research Laboratory

U.S. Environmental Protection Agency

U.S. Geological Survey

Centers for Disease Control and Prevention

Wisconsin Department of Natural Resources, Bureau of Science Services

Michigan Department of Environmental Quality, Water Resources Division

Indiana Department of Environmental Management

Minnesota Department of Health, Environmental Services

New York State Department of Health, Bureau of Community Environmental Health

Pennsylvania, Erie County, Department Of Health

Ohio Department of Health and Ohio Department of Natural Resources

October 2014



UNITED STATES  
DEPARTMENT OF COMMERCE

Penny Pritzker  
Secretary

NATIONAL OCEANIC AND  
ATMOSPHERIC ADMINISTRATION

Kathy Sullivan  
Acting, Under Secretary for Oceans & Atmosphere  
NOAA Administrator



# Barriers (from Rockwell et al. 2014)

- Limited Know-how
  - **62%** of beach managers said training on operating models would be ‘helpful’ or ‘very helpful’ (#2 out of 8)
  - **56%** of beach managers said training on developing models would be ‘helpful’ or ‘very helpful’ (#3 out of 8)



# Barriers (from Rockwell et al. 2014)

- Lack of Data/ Tools
  - 86% of beach managers said location-specific Web-accessible data would be ‘very useful’ or ‘extremely useful’ (#1 out of 5)
  - 86% of beach managers said improved predictive modeling tools would be ‘helpful’ or ‘very helpful’ (#1 out of 9)

NOAA Technical Memorandum GLERL-163

---

**Beach Health Information Needs Assessment:  
9 Years Later - Results from Follow-up Survey**

**David Rockwell<sup>1</sup>  
Sonia Joseph Joshi<sup>2</sup>  
Adam Mednick<sup>3</sup>  
Holly Wirrick<sup>4</sup>**

<sup>1</sup>Cooperative Institute for Limnology and Ecosystems Research, 4840 S. State Rd., Ann Arbor, MI 48108  
<sup>2</sup>Michigan Sea Grant, and NOAA Great Lakes Environmental Research Laboratory, 4840 S. State Rd., Ann Arbor, MI 48108  
<sup>3</sup>University of Wisconsin Sea Grant Institute, 1975 Willow Dr., Madison, WI 53706  
<sup>4</sup>USEPA, Region V, Water Division, 77 W. Jackson Blvd., Chicago Illinois 60604

Beach Health Interagency Coordination Team in cooperation with:  
NOAA, Great Lakes Environmental Research Laboratory  
U.S. Environmental Protection Agency  
U.S. Geological Survey  
Centers for Disease Control and Prevention  
Wisconsin Department of Natural Resources, Bureau of Science Services  
Michigan Department of Environmental Quality, Water Resources Division  
Indiana Department of Environmental Management  
Minnesota Department of Health, Environmental Services  
New York State Department of Health, Bureau of Community Environmental Health  
Pennsylvania, Erie County, Department Of Health  
Ohio Department of Health and Ohio Department of Natural Resources

October 2014

---

 UNITED STATES DEPARTMENT OF COMMERCE  
Penny Pritzker  
Secretary

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION  
Kathy Sullivan  
Acting, Under Secretary for Oceans & Atmosphere  
NOAA Administrator



# [www.seagrant.wisc.edu/virtualbeach](http://www.seagrant.wisc.edu/virtualbeach)

## Virtual Beach

"Virtual Beach" is [free decision-support software](#) for Windows that enables coastal beach managers, public health professionals, and applied researchers to efficiently develop and operate water-quality models to inform beach-specific decisions on:

- Whether to issue (or lift) swim advisories/closures on a given day
- What type of lab test (if any) to run on a given day
- Where and how to prioritize long-term remediation efforts

### On This Page...

[About Virtual Beach](#)

[Virtual Beach Users](#)

[Online Data](#)

[Training](#)

## About Virtual Beach



When unimpaired by poor water quality, beaches can provide coastal communities with substantial economic, social and health benefits. Along the U.S. Great Lakes coast, an estimated 8 million people visit beaches each year. Unfortunately, the types of information that are commonly used to evaluate water quality, communicate health risks and prioritize remediation efforts are often *untimely*, *inaccurate* and/or *incomplete*. Virtual Beach (VB) addresses these shortcomings.

## CONTACT

[Adam Mednick](#)

Post-Doctoral Fellow: Beach Health

[Moira Harrington](#)

Communications Manager

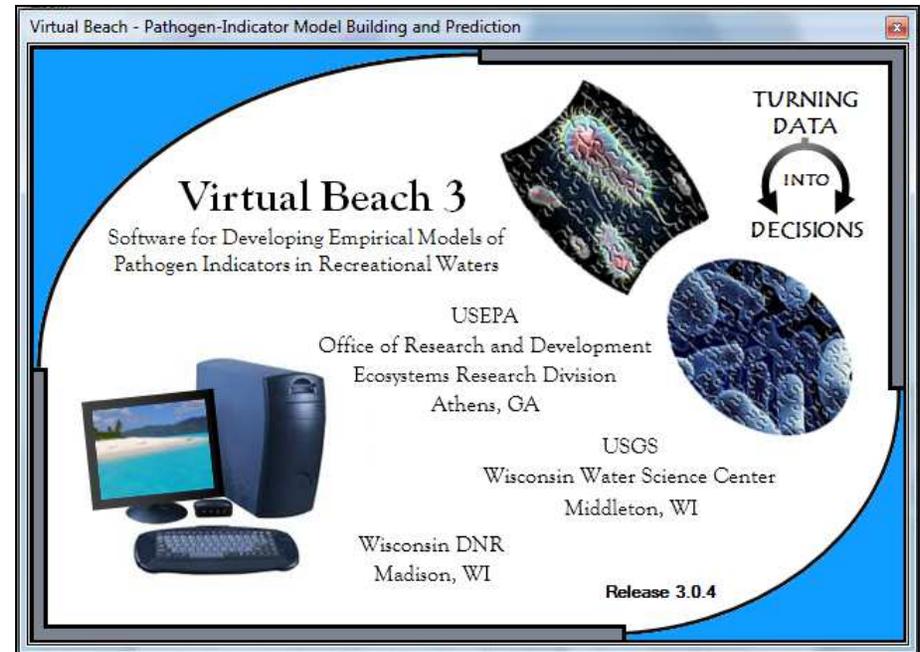


The Great Lakes are the largest freshwater system on Earth.



# Virtual Beach 3.0.5

1. 'Gradient Boosting Machine' (GBM) Method
  - Easier & faster model-building (regression trees)
2. Direct connection to 'EnDDaT'
  - Online data on rain, riv. discharge, clouds, waves, currents, etc.  
... for any beach on the Great Lakes



# GBM Model-building

Model Evaluation Threshold

235 Regulatory Standard

Threshold entry is transformed: 2013 US Regulatory Standards

Value  
 Log10 (value)  
 Loge (value)  
 Power (value) exp: 1

E. coli, Freshwater: 235  
 Enterococci, Freshwater: 104  
 Enterococci, Saltwater: 61

Decision Criterion: 80.95

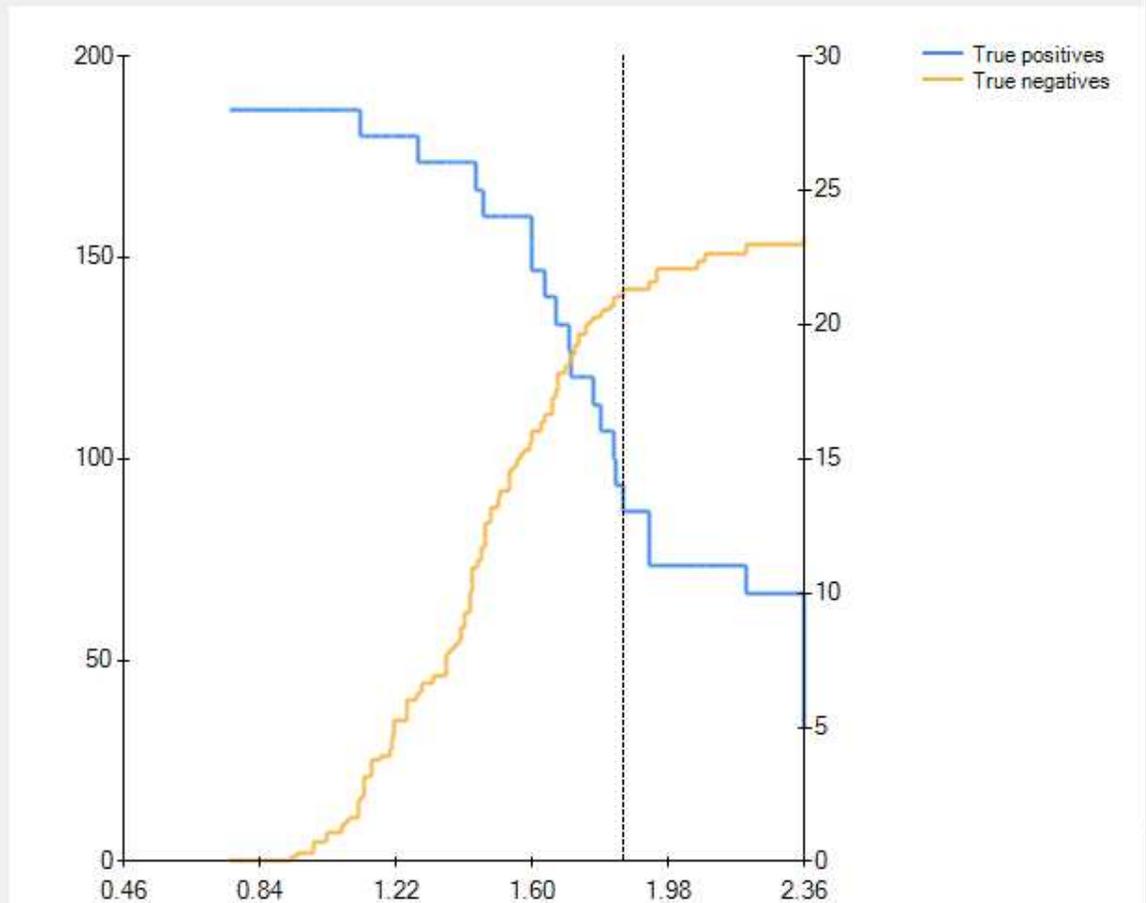
80.95 Decision Criterion

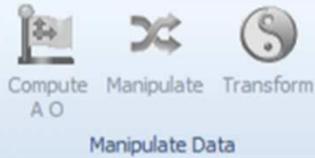
Model Summary

Variable	Coefficient	Influence
SQUAREROOT[RRAIN24]	na	8.0273
POLY[DOY,-4.4919711,0.054335752,...	na	6.8181
POLY[WTEMP,1.037126,0.12811126,...	na	6.7423
QUADROOT[RRAIN6]	na	5.4386
CPAR3	na	4.6775
WTEMPstdv12	na	3.0265
QUADROOT[CLDCV12]	na	2.9481
POLY[ATEMP,-0.85042961,0.253141,...	na	2.8220
POLY[WTEMPstdv24,0.9785591,0.9,...	na	2.7743
POLY[CLDCV24,1.1556752,2.427840,...	na	2.7296
LN[ATEMPstdv24]	na	2.6228
SQUAREROOT[RRAIN120]	na	2.6173

Model Validation

s	False Positives	False Negatives	Sensitivity	Specificity	Accuracy
	15	15	0.46	0.90	0.84





# Traditional (MLR) Model-building

### Model Settings

Variable Selection Control Options Number of Observations: 185

**Evaluation Criteria**

- Akaike Information Criterion (AIC)
- Akaike Information Criterion (AIC)
- Corrected Akaike Information Criterion (AICC)
- R Squared
- Adjusted R Squared
- PRESS**
- Bayesian Information Criterion (BIC)
- Root Mean Square Error (RMSE)
- Mc Sensitivity
- Specificity
- Accuracy

235 Regulatory Standard (Vertical)

**Threshold Transform** 2013 US Regulatory Standards

- None
- Log10
- Ln
- Power

E. coli, Freshwater: 235  
 Enterococci, Freshwater: 61  
 Enterococci, Saltwater: 104

Manual Genetic Algorithm

Set Seed Value: 1

Population Size: 100

Number of Generations: 100

Mutation Rate: 0.05

### Model Information

**Best Fits:**

- 208.2056
- 207.6627
- 207.3146
- 207.1730
- 206.5890
- 206.1195
- 205.9883

**IV Filter**

Add to List View Report

Clear List Cross Validation

**Variable Statistics - SelectedModel**

Parameter	Coefficient	Standardized Coefficient
(Intercept)	-2.7589	
POLY[WindA_comp[WDIR,WSP...	0.5851	0.1674
POLY[WTEMP,1.037126,0.12811...	0.8591	0.2113
QUADROOT[CLDCV12]	0.6463	0.1893
CLEARWATER	-0.4751	-0.3048
POLY[DOY,-4.4919711,0.054335...	1.3176	0.3218
QUADROOT[RRAIN6]	0.3390	0.2165
SQUAREROOT[CPERP12]	-0.8906	-0.1693

Progress Results Fitted vs Observed ROC Curves Residuals

### Genetic Algorithm Dynamic Fitness Update

Percent of Generations Completed	Fitness
0	-180
5	-200
10	-205
15	-206
20	-207
25	-208
30	-209
35	-210
40	-210
45	-210

# Download data from EnDDaT

File Prediction

Import Data

Available Models:

- GBM
- MLR

Model:

ECOLI = QTRSEASON + PRE\_JUNE21 + JUNE21\_JULY15 + JULY16\_AUG10 + POST\_AUG10 + POLY(DOY,-4.4919711,0.054335752,-0.00011978887) + QUADROOT(RRAIN6) + SQUAREROOT(RRAIN24) + SQUAREROOT

Model Evaluation Thresholds

- 65 Decision Criterion (Horizontal)
- 50 Exceedance Probability
- 235 Regulatory Standard (Vertical)

Threshold Transform

- None
- Log10
- Ln
- Power

Predictive Record

Save Column Order

Clear Column Order

	RRAIN48	RRAIN120	ATEMP	CLDCV	WTEMP	ID
1	0.625	10.88	6.686	0.9821	7.856	*
2						

Prediction



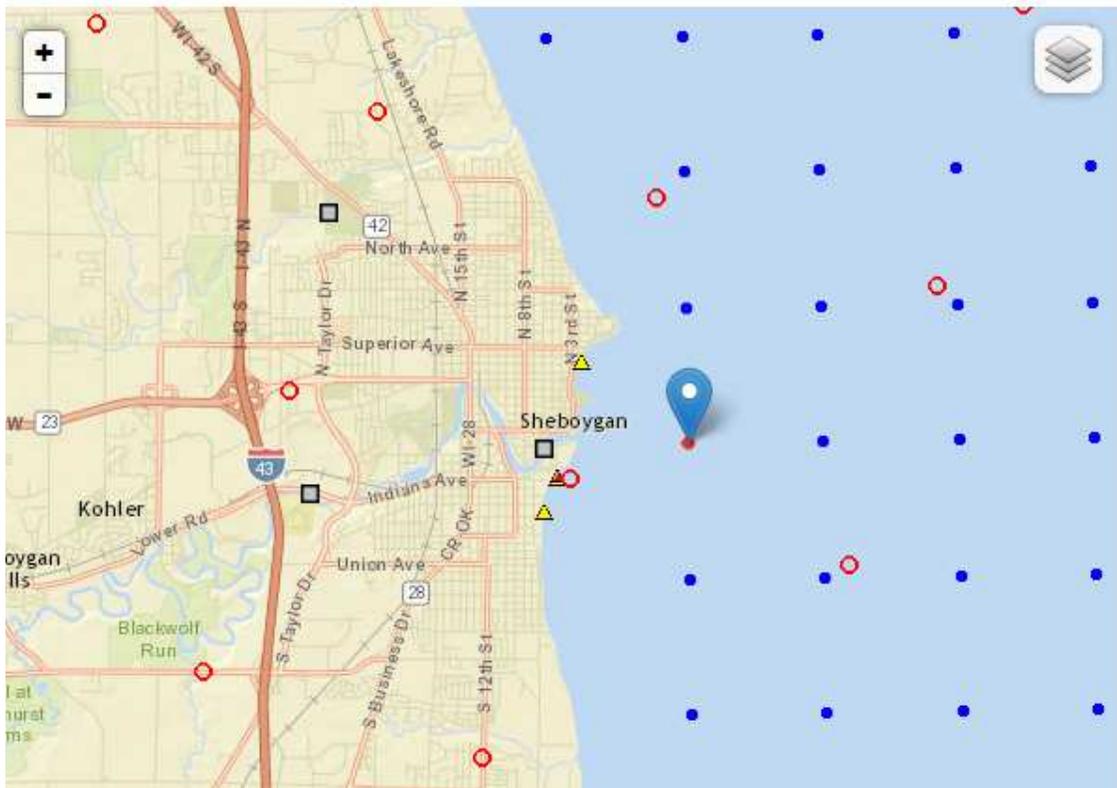
# EnDDaT

USGS Home  
Contact USGS  
Search USGS

## Environmental Data Discovery and Transformation - Version 1.3.23

Access and Integrate Environmental Observations for Coastal Decision Support

- USGS Time Series ([NWIS](#), [Plotting Tool](#))
   6-hr Quantitative National Precipitation
   National Data Buoy Center ([NDBC](#))
- Great Lakes Coastal Forecasting System ([GLCFS](#))
   1-hr Quantitative North Central Precipitation
- USGS Water Quality ([Water Quality](#))
- Search within a  mile bounding box from selected project (click marker to identify)



Active Grid Point:  
Approximate distance to active project: 1.231 miles

x	y	Latitude	Longitude
119	14	43.747955	-87.68291

Available Data:  
Great Lakes Coastal Forecasting System (GLCFS), NO GLERL Wave Model-Great Lakes, Princeton Ocean Mo

	Property	Ur
<input type="checkbox"/>	Ice Concentration	%
<input checked="" type="checkbox"/>	Height Above Model Sea Level	m
<input type="checkbox"/>	Ice Thickness	m
<input checked="" type="checkbox"/>	Eastward Water Velocity at Surface	m/c

# On-line Training





Virtual Beach 3

Location Global Datasheet

Import Data Add Validate Data Validate Compute A O Manipulate Transform Go To Model Work with Data

File	Workshop_Data.xls	DATETIME	LOG10[ECOLI]	QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG10	WATERTEMP_F	POLY[WATERTEMP_F]	W...
Column Count	148	6/11/2009 11:30	2.538	1	1	0	0	0	56.1	1.499	1.5
Row Count	195	6/12/2009 2:15	1.255	1	1	0	0	0	66.1	1.577	0.5
Date-Time Index	DATETIME	5/15/2009 11:25	1.462	1	1	0	0	0	61.5	1.598	0.5
Response Variable	LOG10[ECOLI]	6/18/2009 2:05	1.23	1	1	0	0	0	68.2	1.536	1
Disabled Row Count	0	6/22/2009 10:40	0.6021	2	0	1	0	0	70.3	1.474	1
Disabled Column Count	54	6/23/2009 11:45	1.881	2	0	1	0	0	74.5	1.29	1
Hidden Column Count	1	5/24/2009 11:55	1.176	2	0	1	0	0	71.5	1.43	0.5
Independent Variable Count	92	6/25/2009 11:35	0.4771	2	0	1	0	0	71.2	1.441	0.5
		6/29/2009 11:05	1.041	2	0	1	0	0	68.3	1.555	0.5
		6/30/2009 10:25	0.699	2	0	1	0	0	56.3	1.505	1.5
		7/1/2009 10:20	0.9451	2	0	1	0	0	56.5	1.511	1
		7/2/2009 9:25:0	1.415	2	0	1	0	0	58.5	1.559	1
		7/6/2009 10:05	1.839	2	0	1	0	0	69.3	1.506	0.5
		7/7/2009 10:25	1.322	2	0	1	0	0	64.3	1.597	1
		7/8/2009 10:40	1.255	2	0	1	0	0	64.3	1.597	1.5
		7/9/2009 10:10	0.6021	2	0	1	0	0	71.4	1.434	0.5
		7/13/2009 10:10	1.079	2	0	1	0	0	66.7	1.567	0.5
		7/14/2009 8:45	1.863	2	0	1	0	0	64.3	1.597	0.5
		7/15/2009 12:50	2.588	2	0	1	0	0	69.4	1.573	1
		7/16/2009 11:10	0.7782	3	0	0	1	0			
		7/20/2009 11:40	2.001	3	0	0	1	0			

Project File Name: C:\Users\lenovo\Desktop\VB3Training\Workshop\_Project\_Data.vb3p

Ready.



MI  
by  
Re



Av  
ab  
by  
91



(U  
Hi  
by  
Re



Re  
Th  
by  
75



Se  
Qt  
by  
16



Be  
Sp  
by  
14



A  
by  
71

## Learning Module II: Sections A - E



Subscribe 959

14 views

# Virtual Beach Users' Group



UNIVERSITY OF WISCONSIN SEA GRANT INSTITUTE

## Virtual Beach Users

### Users Group

The Virtual Beach Users Group provides a forum for sharing best practices, troubleshooting problems, and identifying priority needs. Users communicate via the Virtual Beach Users listserv (below) and periodically meet, online or in person, as needed.

### Listserv

The [Virtual Beach Users listserv](#) is intended to foster communication and mutual assistance among VB users. The list is moderated by Wisconsin Sea Grant and hosted by the Great Lakes Information Network (GLIN). [Subscribe to the listserv.](#)

### File-Sharing

The [Virtual Beach Box](#) is a secure site for users who wish to share VB files with their collaborators. Typically too large to email, VB project files include the beach-specific data and transformations used to build the model(s), as well as evaluation metrics and graphics, and the interface used to make predictions. [Request an account.](#)

---

## Online Data



To develop a VB model, one needs historical data on water quality (i.e., lab results) from two or more beach seasons, together with potentially related environmental conditions, such as stream

# Best Practices (~March 2016)

- Model Operation
  - ✓ *Daily*, preferably between 7:30–9:30 am CDT
  - ✓ Operate in conjunction with regular data reporting
  - ✓ Report 'Model' as the **reason** for beach actions
- Minimum Field Data (required in Wisconsin)
  - ✓ 'Clarity' (categories)
    - + *Turbidity (NTUs/ Secchi cm)*
  - ✓ Water Temperature
  - ✓ Wave Height
  - ✓ Cloud Cover (categories)
  - ✓ Algae in water (categories)
  - ✓ Algae beach (categories)
  - ✓ Gulls (estimated number)



# Best Practices (~March 2016)

- Sampling Frequency
  - ✓ *Minimum = 2X/ week*
- Validation Metrics
  - ✓ 50% 'sensitivity' (correct advisories)  
& 90% 'specificity' (correct beach-open)
- Validation Frequency
  - ✓ Option 1: 1X/ summer (e.g., July 15<sup>th</sup>)
  - ✓ Option 2: 2X/ summer (e.g., June 30<sup>th</sup> and July 31<sup>st</sup>)
  - ✓ Don't over-validate (~~1X/ week~~)



# Best Practices (~March 2016)

- Model Building (re-calibration) Frequency:
  - ✓ Annually (preferred); or every 2 years (minimum).
- Reasons for Overriding Model Predictions:
  - ✓ Swimmer Safety (NWS dangerous current forecasts)
  - ✓ Professional Judgment
    - Observations in the field • Weather • Rapid qPCR
    - Recent Model Performance (i.e., validation)
- When Model cannot be run...
  - Default to “persistence” (most recent lab results)





[www.seagrant.wisc.edu/virtualbeach](http://www.seagrant.wisc.edu/virtualbeach)

