

US EPA ARCHIVE DOCUMENT

Friday, April 15  
10:15 a.m.–11:45 a.m.

**Session 10:  
Predictive Modeling and Forecasting**



# Implementing Predictive Models: Practical Advice and New Tools

**Adam Mednick, PhD**

University of Wisconsin Sea Grant Institute

## Abstract

Over the past 5 years, the practice of developing and implementing predictive models at coastal beaches has increased several-fold, particularly in the Great Lakes. During the first 3 years of the Great Lakes Restoration Initiative (GLRI), the adoption of the U.S. Environmental Protection Agency's (EPA's) Virtual Beach decision-support software, among other tools for implementing operational nowcasts, expanded from a handful of sites to over 50 beaches. Whether this expansion will continue in the absence of centralized model-building services previously supported by GLRI remains to be seen. At issue is whether a typical local government (e.g., public health or parks department) can develop, operate, and/or maintain nowcast models without additional funding or specialized staff. Based on past experience and research, the presenter will argue that the answer is a conditional "yes" and will provide practical suggestions on how EPA and its state, local, and academic partners can overcome both real and perceived barriers, such as the lack of adequate data, technical know-how, clear decision criteria, managerial confidence, and time. The presentation will highlight issues relevant to marine beaches, where adoption to date has been minimal, and will conclude with an updated look at the suite of resources and tools being developed to make the process easier and more sustainable over time.

## Biosketch

Dr. Adam Mednick is a postdoctoral fellow at the University of Wisconsin (UW) Sea Grant Institute. He received his bachelor of science degree in natural resources from the University of Minnesota, his master of forest science degree from Yale University, and his doctorate in urban and regional planning from UW-Madison. Dr. Mednick has worked in conservation policy and planning, spatial analysis, research, outreach, and education on a range of issues at the state, local, and national levels. Prior to joining UW Sea Grant in 2014, he worked for the National Parks and Conservation Association in Washington, DC; the New Jersey Conservation Foundation in Far Hills, New Jersey; and the Wisconsin Department of Natural Resources in Madison. Dr. Mednick is an elected member of the Great Lakes Beach Association board of directors, a founding cochair of the Wisconsin Coastal Beaches Workgroup, and the manager of the Virtual Beach Users' Group. His current professional interests include how best to develop and deploy environmental data and modeling systems to the benefit of real-world decision making; and, more generally, how to make academic and government research more useful through collaboration and cooperative extension.



### Implementing Predictive Models: Practical Advice and New Tools

Adam Mednick — University of Wisconsin Sea Grant Institute  
EPA Recreational Waters Conference (New Orleans, April 16, 2016)

### Why Predictive Models (Nowcasts)?

1. To Reduce Unnecessary & Missed Advisories

	Sampled Advisories:		Sampled Open:		Sampled All:	
	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%)
<b>Total</b>	<b>4,498</b>	<b>2,958 (66%)</b>	<b>39,850</b>	<b>4,564 (11%)</b>	<b>44,348</b>	<b>7,522 (17%)</b>

Data from EPA BEACON (2008-'12)

### Why Predictive Models (Nowcasts)?

2. To Increase the Frequency of Monitoring

### Communities using Nowcasts

### Case Study Communities

- Community A (The "Innovator" – adopted 2009)  
*"The intent [is to] rely on Virtual Beach more fully, so we don't spend as much time and money on testing."*
- Community B (The "Early Adopter" – adopted 2011)  
*"We [experimented with] it year-to-year [to see] how well it performed... how many times it [was] right or wrong..."*
- Community C (The "Early Majority" – adopted 2013)  
*"I went to my Administration and said... 'There's some guy from Madison who's really pushing it and I've heard [another community] is using it'..."*

From Mednick (2014)

### Theoretical "Diffusion" of Nowcast Models

Based on Rogers (1962)



UNIVERSITY OF WISCONSIN SEA GRANT INSTITUTE

### Critical Question:

Can a typical local health department *develop, operate, and maintain* its own predictive model...

- Without special funding?
- Without specialized staff?

- Mednick & Watermolen (2014)

**Expanded Beach "Nowcast"**  
Modeling across Wisconsin

Workshop on Coastal Science  
Wisconsin Department of Natural Resources  
1750 University Avenue  
Madison, WI 53707-7025  
www.dnr.wisconsin.gov

UNIVERSITY OF WISCONSIN SEA GRANT INSTITUTE

### 'Typical' Health Departments

From Rockwell et al. (2014):

- In nearly **40%** of reporting depts, **interns responsible for over half** the beach-related work.
- Among depts responsible for > 5 beaches, over **75%** devote < **10%** of **overall time** to beach-related work.
- Among primary staff responsible for beaches, over **85%** spend < **1/4** of **their time** on beach-related work.

U.S. Environmental Monitoring System (EMS)

Beach Health Information Needs Assessment  
Final Report  
Beach Health Information Needs Assessment  
Final Report

U.S. Environmental Monitoring System (EMS)  
Beach Health Information Needs Assessment  
Final Report  
Beach Health Information Needs Assessment  
Final Report

U.S. Environmental Monitoring System (EMS)  
Beach Health Information Needs Assessment  
Final Report  
Beach Health Information Needs Assessment  
Final Report

UNIVERSITY OF WISCONSIN SEA GRANT INSTITUTE

### Other Perceived Barriers:

**Lack of Data**

- 86% of beach managers said **location-specific, Web-accessible data** would be 'very useful' or 'extremely useful' (#1 out of 5)

**Lack of Tools**

- 86%... said **improved predictive modeling tools** would be 'helpful' or 'very helpful' (#1 out of 9)

NOAA Technical Memorandum OLEB-143

Beach Health Information Needs Assessment  
Final Report  
Beach Health Information Needs Assessment  
Final Report

U.S. Environmental Monitoring System (EMS)  
Beach Health Information Needs Assessment  
Final Report  
Beach Health Information Needs Assessment  
Final Report

UNIVERSITY OF WISCONSIN SEA GRANT INSTITUTE

### "Virtual Beach"

Model Information

Variable	Statistics - SelectedModel	Model Statistics - SelectedModel
Value	0.5138	
Adjusted R Squared	0.5077	
Multiple Information Criteria	-31.6493	
Correlated AIC	-32.9452	
Bayesian Info Criterion	-38.1068	
RMSE	52.2026	
RMSE	0.5422	
Transformed DC	2.3711	

UNIVERSITY OF WISCONSIN SEA GRANT INSTITUTE

[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 separate 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

CONTACT  
Adam Mednick  
Post-Doctoral Fellow: Beach Health  
Moira Harvathson  
Communications Manager

**On This Page:**  
[About Virtual Beach](#)  
[Virtual Beach Users](#)  
[Online Data](#)  
[Training](#)

**About Virtual Beach**

When compromised by poor water quality, beaches can provide coastal communities with substantial economic, social and health benefits. Along the U.S. coast, lakes exist, 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 unreliable, inconsistent and/or incomplete. Virtual Beach (VB) addresses these shortcomings.

UNIVERSITY OF WISCONSIN SEA GRANT INSTITUTE

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

[Back to Virtual Beach \(main page\)](#)

### Online Data

To develop a Virtual Beach (VB) model, one needs historical data on water quality from two or more beach seasons, together with potentially related environmental conditions, such as stream discharge, antecedent rainfall, wave height, waterlevel, etc... Depending on the beach, many if not all of these data will be available online for free.

### Water Quality & Beach Conditions

Historical water quality can be downloaded from EPA's BEACHCOOL website – or from the state beach websites listed below. Depending on the state, field-collected sanitary monitoring data, such as turbidity, algae levels and pH, may also be available. Note: If you have sanitary survey data for your beach, it is to your advantage to upload them to your state's beach website. This will make building and maintaining VB models significantly easier.

- Illinois Beach Guard:** Viewable tables of E. coli data. Single-day summaries of sanitary survey results.
- Indiana Beach Guard:** Viewable tables of E. coli data. Single-day summaries of sanitary survey results.
- Michigan Beach Guard:** Viewable tables of E. coli data. Single-day summaries of sanitary survey results.

**1.) Lab Data + Sanitary Conditions**  
**2.) Hydro-Meteorological Data**



xy	Latitude	Longitude	Start Date	End
585:713	43.890	-87.890	2002-01-01	2011
284:714	43.720	-87.952	2002-01-01	2011
584:714	43.767	-87.896	2002-01-01	2011
583:714	43.804	-87.870	2002-01-01	2011
584:714	43.841	-87.854	2002-01-01	2011
284:714	43.879	-87.838	2002-01-01	2011
279:713	43.811	-87.892	2002-01-01	2011

From Zhang (2013)

File Type / Priority	File Name/Format/Size	File Description
Install / Required	Virtual VB 3.0.6 (2.05.57 KB) / Ziped File	Ziped file for vcrun3.0.6 of the VB application.
	Virtual Beach v 3.0.6 User Guide	



## Virtual Beach 3.0.6

1. 'Gradient Boosting Machine' (GBM) Method
  - **More efficient** model-building (regression trees)
2. Direct connection to 'EnDDaT'
  - **BIG, Easy data:** River Discharge, Waves, Currents, etc. (Spatiotemporally-matched/processed)

## Traditional (MLR) Model-building

## GBM Model-building

## Download data from EnDDaT

1	RESPONSE	PPHVAL	ATEMP	LLULV	SETEMP
1	0.625	10.88	6.686	0.9021	7.856

## Other Perceived Barriers:

- Limited Technical Know-How**
  - Over 60% said training on **predictive models** would be 'helpful' or 'very helpful' (Rockwell et al. 2014)
- Lack of Comprehensive Guidelines/ "Best Practices"**
  - Under Development (UW Sea Grant)
- Lack of Confidence on the part of Administrators and Decision-Makers**

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



UNIVERSITY OF WISCONSIN SEA GRANT INSTITUTE

## 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 VR 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 Users](#) is a secure site for users who wish to share VR files with their collaborators. Typically too large to email, VR 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 (e.g., lab results) from two or more beach seasons, together with water quality related environmental conditions such as stream

UNIVERSITY OF WISCONSIN SEA GRANT INSTITUTE

## Best Practices (DRAFT)

- Model Operation
  - ✓ Daily, preferably between 8:30–10:30 am EDT
  - ✓ Operate in conjunction with regular data reporting
  - ✓ Report 'Model' as the reason for beach actions
- Minimum Field Data (required in Wisconsin)
  - ✓ 'Clarity' (categories)
  - ✓ Cloud Cover (categories)
  - ✓ Turbidity (NTUs/ Secchi cm)
  - ✓ Algae in water (categories)
  - ✓ Water Temperature
  - ✓ Algae beach (categories)
  - ✓ Wave Height
  - ✓ Gulls (estimated number)

UNIVERSITY OF WISCONSIN SEA GRANT INSTITUTE

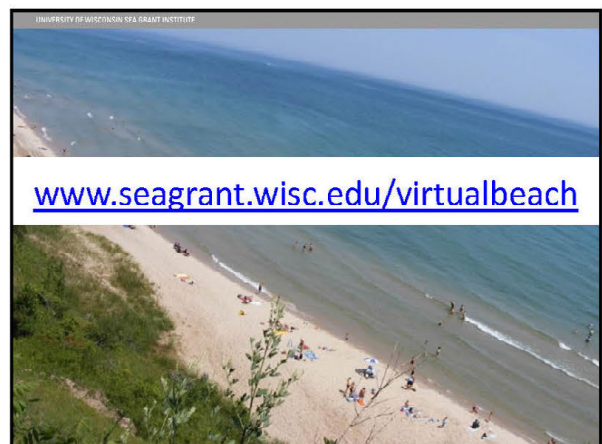
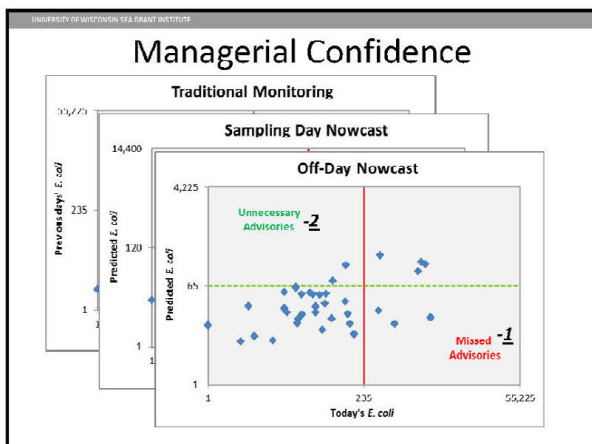
## Best Practices (DRAFT)

- Sampling Frequency
  - ✓ 2 or more Samples per Week
  - ✓ Less Frequent (?) for Year-Round beaches
- Validation Frequency
  - ✓ Every 1-2 Months
  - ✓ Don't over-validate (~~1x/week~~)
- Validation Metrics (Francy et al. 2013)
  - ✓ 50% 'sensitivity' (correct advisories) and 90% 'specificity' (correct beach-open)

UNIVERSITY OF WISCONSIN SEA GRANT INSTITUTE

## Best Practices (DRAFT)

- Model Building (*re-calibration*) Frequency:
  - ✓ Annual (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 Nowcast Model cannot be run...
  - Default to "Persistence" (i.e., most recent lab results)







# California Beach Water Quality Nowcasting

**Leslie Griffin**

Heal the Bay

## Abstract

Traditional beach management that uses concentrations of cultivatable fecal indicator bacteria (FIB) may lead to delayed notification of unsafe swimming conditions. Predictive, “Nowcast” models of beach water quality may help reduce beach management errors and enhance protection of public health. This study compared the performances of five different types of statistical, data-driven predictive models—multiple linear regression model, binary logistic regression model, partial least-squares regression model, artificial neural network, and classification tree—in predicting health advisories due to FIB contamination at 25 beaches along the California coastline. In total, over 700 models were developed and evaluated. Multiple linear regression with threshold tuning performed well, along with binary logistic regression with threshold tuning and classification trees. On average, models outperformed the current method based on day-old FIB concentrations by capturing 25% more poor water quality days while maintaining equivalent false negative results. Beaches with well-performing models usually have a rainfall/flow-related dominating factor affecting beach water quality, while beaches having a deteriorating water quality trend or low FIB exceedance rates are less likely to have a well-performing model. Based on the results of this study, we carried out a pilot study at three Californian beaches with beach managers in the summer of 2015 to use daily nowcasting for public notification of beach water quality. Due to the success of the pilot program, the State of California has funded the development of a Nowcasting system to provide daily

information to local beach managers in an effort to help inform public notification decisions for up to 25 separate beach locations over the next 3 years.

## Biosketch

Ms. Leslie Griffin is the beach water quality scientist at the Los Angeles-based environmental organization, Heal the Bay. Native to the East Coast, she relocated across country to receive her bachelor and master of science degrees in environmental science with an emphasis in water quality from Loyola Marymount University. She worked on passive sampling of PAHs for 2 years while obtaining her master’s degree. While pursuing her education, Ms. Griffin interned at Heal the Bay as an aquarist and a watershed educator. In 2015, she began working full time with the organization as the data analyst for the Beach Report Card program. Currently, Ms. Griffin manages the Beach Report Card program—working to ensure accurate and timely dissemination of weekly beach water quality info for over 600 locations along the West Coast, as well as implementing a daily predictive modeling—or “nowcasting”—program for five beaches in Southern California.



# CALIFORNIA BEACH WATER QUALITY PREDICTIVE MODELING PROJECT

HEAL THE BAY, STANFORD UNIVERSITY, AND UCLA  
 U. GRIFFIN, J. ALAMILLO, M. TAGGART, A. THOE, A. BOEHM, M. GOLD

## Project Goal

To study the feasibility of using predictive models as a public notification tool at California beaches

## Project Outline

- Phase I: Proof of concept
  - Can models be developed for CA marine beaches?
- Phase II: Pilot at 3 beaches
  - Can models be readily integrated in existing M&PN programs?
- Phase III: Program Implementation
  - Developing and Applying models for 20-25 beaches in CA

## Why do we need predictive models in California?

- Our current monitoring and public notification (M&PN) programs leave the public at risk:
  - 24-48 hours from sample to posting
  - Rapid detection methods still take hours
  - Of the ~500 beaches monitored in CA:
    - ~30 sampled 5x per week
    - ~30 2x per week
    - >400 only once per week

## Phase I: Proof of Concept

- Completed 2012-2014 at 25 beaches in CA
- 6 years of historical data
- Input factors: rain, tide, wind, solar radiation, etc.
- 5 model types
- 3 TIB, 2 seasons
  - Summer (S) April to October
  - Winter (W) November to March
- Over 700 models developed and tested
- Calibration (2006-2010) and validation (2011-2012)

## Phase I: Conclusions

- Models can improve sensitivity while maintaining a reasonable specificity
  - Sensitivity: the ability of a model to accurately predict beach postings
  - Specificity: The accurate prediction of open beach days
- Two peer-reviewed scientific papers were published based on Phase I results



US EPA ARCHIVE DOCUMENT

## Phase II: Pilot at Three Beaches

- Completed 2015
- Objectives:
  - Optimizing models from Phase I
  - Feasibility of using models w/in M&PN programs at CA beaches



## Pilot Design

- Prediction tool: optimized MLR model in an Excel spreadsheet
- Prediction of post/no-post daily by 10 am
- Study period: Memorial Day to Labor Day
- Three Beaches
  - LADPH/City of LA Santa Monica Pier
  - OCHCA Doheny
  - SBCEHS Arroyo Burro




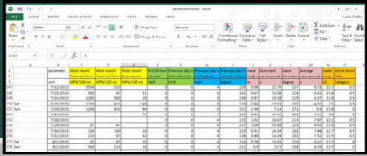


## Current Monitoring and Public Notification at Pilot Beaches

Site	County	Monitoring Agency	Public Notification Agency	Monitoring Frequency	Posting Protocol	Log Time
Arroyo Burro beach	Santa Barbara	SBCEHS	SBCEHS	1/per week	Post on Single Sample	2
Doheny State Beach	Orange	OCHCA SOCWA	OCHCA	2/per week	Post on Single Sample or Geo Mean	2
Santa Monica Pier Beach	Los Angeles	LADPH-FMD	LADPH	5/per week	Post on Single Sample	1/Total and 2/Entero

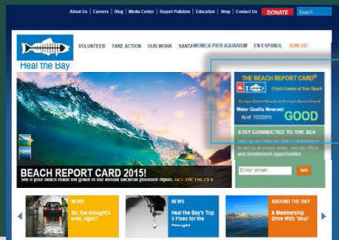
## Pilot Daily Modeling Steps

- Obtaining FIB data
- Collect online environmental data
- Run each FIB model
- Cross-check between with agencies, H&B, and Stanford
- Posting results online





## Posting Predictive Model Results Online

Heal the Bay Main Webpage



Orange County Website

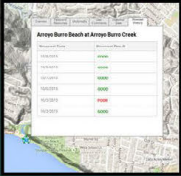
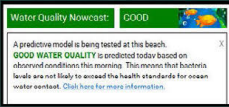
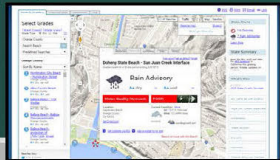


## Posting Predictive Model Results Online

Beach Report Card website:

- Beachgoers could find wet and dry grades, Advisories, Nowcast results, and historical information by beach


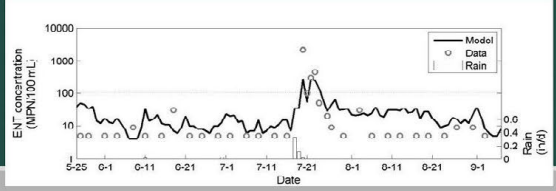
Both the BRC and OC websites also had FAQ sections for questions/concerns




## Doheny State Results

Number of samples: 34  
 Number of predictions: 106  
 Observed exceedances: 3  
 Model captured: 3 vs 2 (current)  
 False alarms: 1 vs 3 (current)


## Phase II Outcomes

- Models successfully run 106 consecutive days w/ results ready by 10 am
- Daily notification to beach-goers in the morning everyday **including weekends**
- 3 agencies voluntarily participated & donated staff/resources to the pilot
- Models can be successfully integrated into existing M&PN programs



## Benefits of Predictive Models

- Improved accuracy in public notification over current method
- Improved understanding of FIB pollution at the beach and how to mitigate sources
- Easy and flexible model implementation can be run by the health agencies or a third party (Heal the Bay)



## Phase III: Develop CA Nowcast System

- SWRCB grant to build permanent CA Beach Nowcast system:
  - Heal the Bay, Stanford, and UCLA
  - 3 year roll-out
  - 20 summer AB-411 beaches
  - 5 winter surf beaches



## Phase III: Develop CA Nowcast System

- Technical Advisory Committee and Implementation Advisory Committee
- Public outreach
- On-the-beach public notification program
- Webpage and Mobile Apps



## Acknowledgements

Santa Barbara County Environmental Health Services (SBCEHS)  
 Willie Brummett, Lawrence Fay, and David Brummond

Orange County Health Care Agency (OCHCA)  
 Larry Brenner, Lauren Hatch, Juan Amador, Joe Gutzman, and Dan Yokoyama

City of Los Angeles, Environmental Monitoring Division (EMD)  
 Ioannice Lee (Victor Ruiz, Amber Kuhn, Tommy Nguyen), Mas Dojiri, and Stan Asato

Los Angeles County Department of Public Health  
 Jacqueline Taylor, Maurice Pantolija, and Nick Brakband (Zepur Ohannessian)

USGS Ohio Water Science Center  
 Donna Francy



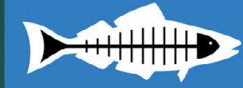
## Want more info?

Email: [LGriffin@healthebay.org](mailto:LGriffin@healthebay.org)

FAQ: <http://brc.healthebay.org/default.aspx?tabid=4>



## Thank You



Heal the Bay



STANFORD  
UNIVERSITY



# Predictive Modeling and Forecasting of Water Quality at Recreational Beaches along Gulf of Mexico Coast

**Zhiqiang Deng**

Louisiana State University

## Abstract

A series of predictive models has been developed by Louisiana State University for recreational beaches that have experienced frequent advisories over the past 10 years. The beaches used in the project, which was funded by the National Aeronautics and Space Administration (NASA), were Siesta Key Beach and Venice Beach in Florida, Orange Street Pier/Park Beach in Alabama, Harrison County Beach in Mississippi, Holly Beach in Louisiana, and Galveston Bay Beach and Corpus Christi Bay Beach in Texas. The models were constructed using an artificial neural networks toolbox in the MATLAB program and can predict either daily enterococci levels in beach waters or risks of water quality standard violations at a beach site as long as daily data are available for the environmental parameters (e.g., rainfall, salinity, temperature, wind, tide [or gage height], and solar radiation). Some models require less data and some of the data can be replaced with NASA satellite data. The models were able to explain 70–86% of the variations in observed enterococci levels or recreational water quality advisories issued by state beach monitoring programs. User manuals for state beach monitoring personnel explain how to use the models for real-time monitoring of recreational water quality. This presentation will provide an overview of the models and their performance in predicting water quality at the beaches. It is expected that the adoption and sustained use of the models will significantly improve the effectiveness of recreational water programs and provide better protection of public health in the Gulf of Mexico states and the nation.

## Biosketch

Dr. Zhiqiang Deng is a professor of water resources engineering at Louisiana State University. He specializes in predicting and preventing the contamination of water bodies with high public health and economic impacts (primarily recreational beach waters, oyster harvesting waters, and rivers) through sensor network-based monitoring, watershed-based modeling, and sustainability-based mitigation. Dr. Deng has published over 50 refereed journal papers in those areas.



# Using Probabilities of Enterococci Exceedance and Logistic Regression to Evaluate Long-Term Weekly Beach Monitoring Data

**Jay Fleisher, PhD**

Nova Southeastern University

## Abstract

Recreational water quality surveillance involves comparing bacterial levels to set threshold values to determine beach closure. Bacterial levels can be predicted through models which are traditionally based on multiple linear regression. The objective of this study was to evaluate exceedance probabilities—as opposed to bacterial levels—as an alternate method to express beach risk. Data were incorporated into a logistic regression to identify environmental parameters most closely correlated with exceedance probabilities. The analysis was based on 7,422 historical sample data points from the years 2000–2010 for 15 beach sample sites in south Florida. Probability analyses showed which beaches in the data set were most susceptible to exceedances. No yearly trends were observed nor were any relationships to monthly rainfall or hurricanes apparent. Results from logistic regression analyses found that among the environmental parameters evaluated, tide was most closely associated with exceedances, with exceedances 2.475 times more likely to occur at high tide than at low tide. The logistic regression methodology proved useful for predicting future exceedances at a beach location in terms of probability and modeling water quality environmental parameters with dependence on a binary response. Beach managers can use this methodology for allocating resources when sampling more than one beach.

## Biosketch

Dr. Jay Fleisher received his bachelor of science degree in environmental health science and master of science degree in environmental science from the City University of New York, his master of science degree in epidemiology from Columbia University's School of Public Health, and his doctorate in environmental epidemiology/biostatistics from the Institute of Environmental Medicine, New York University. Dr. Fleisher holds faculty positions at Florida's Nova Southeastern University and University of Miami. Dr. Fleisher's research interests are in the fields of chronic and infectious illnesses. He has focused his research efforts on the health effects of exposure to waters contaminated with domestic sewage, indicator organism variability, indicator organism-pathogen relationships, risk assessment, statistical water quality sampling protocols, assessing compliance, setting of microbial water quality standards, population health burden assessment, risk perception, and risk vs. current standards. Dr. Fleisher has advised numerous international committees, organizations, and government agencies on various aspects of these recreational water quality issues. In addition, he has authored over 70 peer-reviewed publications and six book chapters.



Multiple Least Squares Linear Regression Vs. Multiple Logistic Regression  
A Comparison

**METHODS:**  
888 samples were utilized in this analysis. 10 major environmental variables and several FIO's were collected on each sample date. Both types of models were run on these data.

**ENVIRONMENTAL VARIABLES**

- > pH
- > Salinity Water
- > Water Temperature
- > Tidal Stage
- > Turbidity
- > Amount of Rainfall in the preceding 6 hours prior to sampling
- > Amount of Rainfall in the preceding 24 hours of sampling
- > Wind Direction
- > Wind Speed
- > Solar Radiation

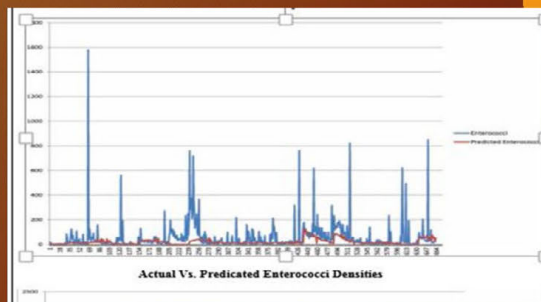
\*All Environmental Variables entered in Both Models and Backward Selection Procedure used in all models

**Results Least Squares Model**

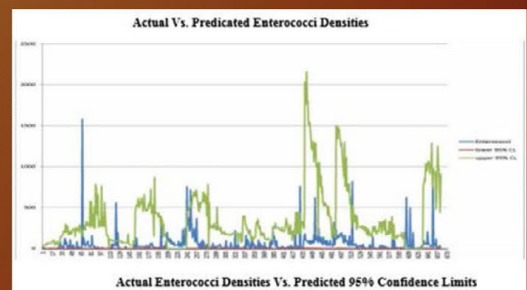
Variable	Estimate	Error	SS	F Value	Pr > F
*Intercept	-0.28651	0.346244	0.24437	0.62	0.4296
*Temperature	0.07383	0.01320	12.91371	32.99	<.0001
*Tide	1.05945	0.12393	26.02290	66.49	<.0001
*Rain 24 Hr Prior	-0.02254	0.00428	10.85001	27.72	<.0001
*Wind direction	0.00133	0.00035045	5.61027	14.33	0.0002
*Wind Speed	-0.11523	0.01280	31.70014	80.99	<.0001
*Solar radiation	-0.00108	0.001634	24.74859	63.23	<.0001

Model R<sup>2</sup> square = 0.26

Multiple Linear Least Squares Regression



Multiple Linear Least Squares Regression







### Multiple Logistic Regression

### ENVIRONMENTAL VARIABLES

- > pH
- > Salinity Water
- > Water Temperature
- > Tidal Stage
- > Turbidity
- > Amount of Rainfall in the preceding 6 hours prior to sampling
- > Amount of Rainfall in the preceding 24 hours of sampling
- > Wind Direction
- > Wind Speed
- > Solar Radiation

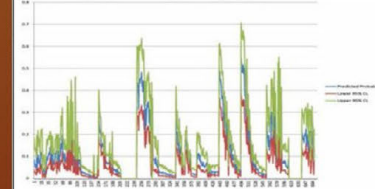
\*All Environmental Variables entered in Both Models and Backward Selection Procedure used in all models

### Results Logistic Regression Above or Below Single Sample Criteria

Parameter	DF	Standard		Wald	
		Estimate	Error	Chi-Square	Pr > ChiSq
Intercept	1	-19.1244	4.9439	14.9637	0.0001
Salinity	1	0.3526	0.1170	9.0784	0.0026
Temperature	1	0.1886	0.0663	8.1041	0.0044
Tide	1	2.8824	0.7645	14.2163	0.0002
Solar radiation	1	-0.00530	0.000860	38.0301	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Salinity	1.423	1.131	1.790
Temperature	1.208	1.061	1.373
Tide	17.858	3.991	79.903
Solar radiation	0.992	0.992	0.994



Predicted Probabilities with 95% Confidence Limits  
Sensitivity of 72.9% and a maximum Specificity of 65.9% at a cut point = 0.1

Our Best Multiple Least Squares Regression was computed with a R Square value of 0.26, while the Multiple Logistic Regression Model yielded a maximum Sensitivity of 72.9% and a maximum Specificity of 65.9% at a cut point = 0.1. A backward selection routine was used in both the Logistic and Least Squares Model.

### CONCLUSIONS

Since the Logistic regression yields a much less robust goodness of fit statistic coupled with the fact that the Beach Managers decision is a dichotomous one, more attention should be paid to research using the Multiple Logistic Model. It should also be noted that the precision of the Logistic Model seems better than both the Least Squares Model approach and the actual Culturing of Enterococci.



### Application of Logistic Regression to Historical Data

### Used Different Data than Previous

- 7,422 Samples analyzed
- Data from 2000-2010
- Data from 13 South Florida Beaches

US EPA ARCHIVE DOCUMENT



Figure 1. Beach assigned number within the FDOH Healthy Beaches Program, beach name, and GPS location of sampling point.

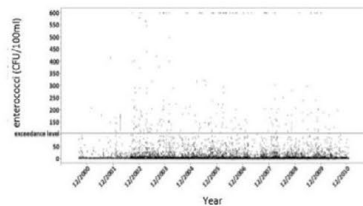


Figure 2. Historical DOH beach sampling from June 2000- December 2010 (n = 7,422). Gray line delineates any samples above the 104 CFU EPA exceedance level.

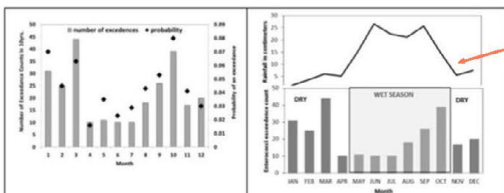


Figure 4. Monthly historical exceedance counts (bars) and their probability (diamonds) for ten years of exceedance count data.

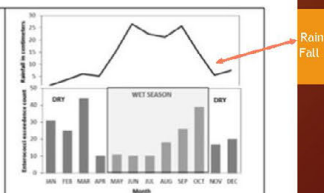


Figure 5. South Florida Water Management District (SFWMD) Dade County Average rainfall from 2000 to 2010 vs. probability and number of exceedance counts in each month. Gray area represents wet season and white area represents dry season.

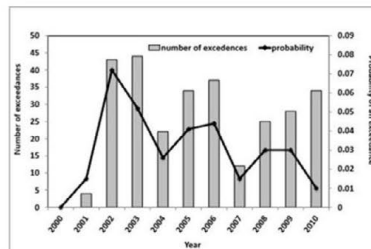


Figure 6. Historical exceedance counts (bars) and their probability (diamonds) by year. The probabilities were connected to show the variability between the years.

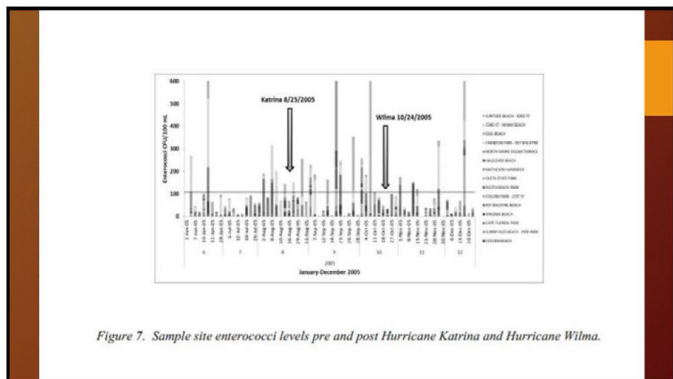


Figure 7. Sample site enterococci levels pre and post Hurricane Katrina and Hurricane Wilma.

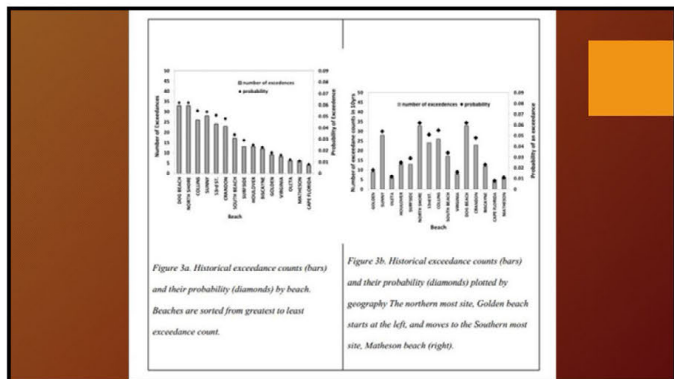


Figure 3a. Historical exceedance counts (bars) and their probability (diamonds) by beach.

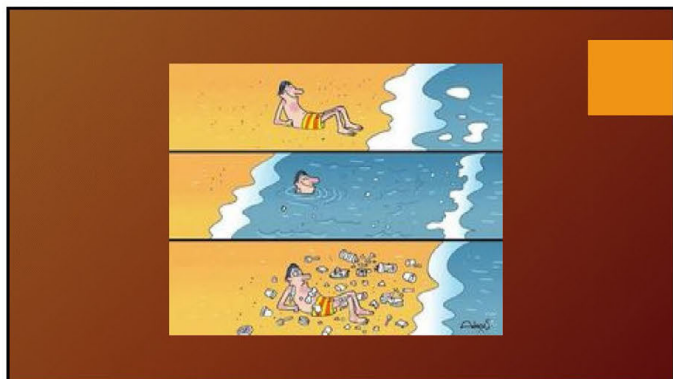
Figure 3b. Historical exceedance counts (bars) and their probability (diamonds) plotted by geography. The northern most site, Golden beach starts at the left, and moves to the Southern most site, Matheson beach (right).

US EPA ARCHIVE DOCUMENT

- Analysis identified which beaches were most susceptible to exceedances
- Logistic regression proved useful for predicting the probability of an exceedance
- Tide was most closely associated with exceedance
- Results can be used to allocated beach sampling resources

Effect	Point Estimate	95% Wald Confidence Limits	
		Lower	Upper
Tidal Conditions 1 vs 3	2.475	1.661	3.687
Tidal Conditions 2 vs 3	1.252	0.866	1.811

Table 1: Odds ratio estimates between tidal conditions as computed from Logistic Regression. Tidal conditions as reported by the FDOH are coded as 1 = High Tide, 2 = Slack Tide, and 3 = Low Tide.





# Development of a Predictive Spatial Model to Understand the Connection between Rainfall Events and Beach Water Quality

**Lance Larson, PhD**

Natural Resources Defense Council

## Abstract

Throughout coastal portions of the United States, rainfall events are physical mechanisms that deliver various urban and rural pollutants to coastal waterways, threatening human and ecosystem health. The objective of this research was to correlate historical beach water quality exceedances to rainfall events. We developed a spatial and temporal beach water quality exceedance model, which queries a database consisting of water quality sample results collected over a 10-year period (2005-2014) at over 8,000 U.S. beaches in 30 states. The model consists of a series of dynamic database queries based on a set of user-defined input parameters. In the database, each water quality sample record is associated with precipitation totals recorded on the sample collection date, as well as for each of the 3 days prior to that sample date, as measured by the nearest weather station submitting data to the National Oceanic and Atmospheric Administration's Quality Controlled Local Climatological Data (QCLCD). Our results suggest a strong connection at the national, state, county, and beach scale between increased rainfall events and beach exceedance occurrences. For example, at the national level, the failure rate increased from 9% to 21% when a rainfall event greater than 0.5 inches was observed within 10 miles within 1 day. Other states and counties observed disproportionate changes in exceedance failure rates. Our model aims to significantly increase our understanding of rainfall influences on beach water quality throughout the United States, improve water quality sampling frequencies and planning,

and examine the effectiveness of implementing watershed pollution reduction strategies.

## Biosketch

Dr. Lance Larson is a science center fellow with the Natural Resources Defense Council in Washington, DC. He earned a bachelor of science degree in environmental engineering from the California Polytechnic State University in San Luis Obispo (2008) and a master of science degree from the South Dakota School of Mines and Technology (2010). Dr. Larson received a dual doctorate in environmental engineering and biogeochemistry from Pennsylvania State University (2013). His graduate research focused on acid mine drainage, arsenic and uranium fate and transport, and biogeochemical interactions between surface and groundwater. Dr. Larson currently is working with the Land and Wildlife, Nuclear, and Water programs to protect U.S. water resources.



## Question & Answer Session

### Comment 1

*(Unknown): For Lance [Larson]. Good database and good work—I am glad you put in the lag times which are so important in rivers and runoff, and also the saturation of the soil which affects the effect of rainfall. When you have a 0.9-inch rain event, we consider the storm surge as well as the amount. It's interesting to take that into account.*

### Answer 1

*Lance Larson:* This work raises many more questions than answers. We can use it to build in other things like that.

### Question 2

*(Unknown): For Lance [Larson]. How did you make sure the rainfall is in the right area and not in another watershed?*

### Answer 2

*Lance Larson:* You could decrease that distance, so 10 miles would be your threshold. Within that, it picks it up. It's the threshold cutoff. If we can find the nearest location, we do. We can run them again at different locations. We did a sensitivity analysis as well.

### Answer 2 (follow-up)

*Adam Mednick:* You said the magic word, "tide," which is very important for incorporating into models. Also exceedance. In the best practices document we are putting out it's about probability. One use for VB [Virtual Beach] and modeling is figuring out when whether and how to test. Glad Jay [Fleisher; made that point during his presentation.

### Answer 2 (follow-up)

*Mike Cyterski:* In terms of Virtual Beach, I'd like to add some other tools, like logistical regressions and neural nets, and lasso regression (where you minimize the number of variables that you use in your regression).