Developing and Implementing Predictive Models for Estimating Recreational Water Quality at Great Lakes Beaches

Scientific Investigations Report 2013–5166

Developing and Implementing Predictive Models for Estimating Recreational Water Quality at Great Lakes Beaches

By Donna S. Francy, Amie M.G. Brady, Rebecca B. Carvin, Steven R. Corsi, Lori M. Fuller, John H. Harrison, Brett A. Hayhurst, Jeremiah Lant, Meredith B. Nevers, Paul J. Terrio, and Tammy M. Zimmerman

Scientific Investigations Report 2013–5166

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2013

Acknowledgments

Contents

Appendixes

Appendixes 4 and 5 follow immediately in this document. Appendixes 1, 2, and 3 are separate documents, available for downloading from the indicated Web address.

Figures

Tables

Conversion Factors

Multiply	By	To obtain
Length		
inch (in)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
Area		
square meter (m^2)	10.76	square foot (ft^2)
Volume		
milliliter (mL)	0.03382	ounce, fluid (fl. oz)
Precipitation and flow rate		
inch per hour (in/h)	25.4	millimeter per hour (mm/h)
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

°F=(1.8×°C)+32

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

°C=(°F-32)/1.8

Bacteria concentrations are given in colony-forming units per 100 milliliters (CFU/100 mL) or most-probable number per 100 milliliters (MPN/100 mL).

Developing and Implementing Predictive Models for Estimating Recreational Water Quality at Great Lakes Beaches

By Donna S. Francy, Amie M.G. Brady, Rebecca B. Carvin, Steven R. Corsi, Lori M. Fuller, John H. Harrison, Brett A. Hayhurst, Jeremiah Lant, Meredith B. Nevers, Paul J. Terrio, and Tammy M. Zimmerman

Abstract

Predictive models have been used at beaches to improve the timeliness and accuracy of recreational water-quality assessments over the most common current approach to water-quality monitoring, which relies on culturing fecal-indicator bacteria such as *Escherichia coli (E. coli.)*. Beach-specific predictive models use environmental and water-quality variables that are easily and quickly measured as surrogates to estimate concentrations of fecal-indicator bacteria or to provide the probability that a State recreational water-quality standard will be exceeded. When predictive models are used for beach closure or advisory decisions, they are referred to as "nowcasts." During the recreational seasons of 2010–12, the U.S. Geological Survey (USGS), in cooperation with 23 local and State agencies, worked to improve existing nowcasts at 4 beaches, validate predictive models at another 38 beaches, and collect data for predictive-model development at 7 beaches throughout the Great Lakes. This report summarizes efforts to collect data and develop predictive models by multiple agencies and to compile existing information on the beaches and beach-monitoring programs into one comprehensive report.

Local agencies measured *E. coli* concentrations and variables expected to affect *E. coli* concentrations such as wave height, turbidity, water temperature, and numbers of birds at the time of sampling. In addition to these field measurements, equipment was installed by the USGS or local agencies at or near several beaches to collect water-quality and metrological measurements in near real time, including nearshore buoys, weather stations, and tributary staff gages and monitors. The USGS worked with local agencies to retrieve data from existing sources either manually or by use of tools designed specifically to compile and process data for predictive-model development.

Predictive models were developed by use of linear regression and (or) partial least squares techniques for 42 beaches that had at least 2 years of data (2010–11 and sometimes earlier) and for 1 beach that had 1 year of data. For most models, software designed for model development by the U.S. Environmental Protection Agency (Virtual Beach) was used. The selected model for each beach was based on a combination of explanatory variables including, most commonly, turbidity, day of the year, change in lake level over 24 hours, wave height, wind direction and speed, and antecedent rainfall for various time periods. Forty-two predictive models were validated against data collected during an independent year (2012) and compared to the current method for assessing recreational water quality—using the previous day's *E. coli* concentration (persistence model). Goals for good predictive-model performance were responses that were at least 5 percent greater than the persistence model and overall correct responses greater than or equal to 80 percent, sensitivities (percentage of exceedances of the bathing-water standard that were correctly predicted by the model) greater than or equal to 50 percent, and specificities (percentage of nonexceedances correctly predicted by the model) greater than or equal to 85 percent. Out of 42 predictive models, 24 models yielded overall correct responses that were at least 5 percent greater than the use of the persistence model. Predictive-model responses met the performance goals more often than the persistence-model responses in terms of overall correctness (28 versus 17 models, respectively), sensitivity (17 versus 4 models), and specificity (34 versus 25 models). Gaining knowledge of each beach and the factors that affect *E. coli* concentrations is important for developing good predictive models. Collection of additional years of data with a wide range of environmental conditions may also help to improve future model performance. The USGS will continue to work with local agencies in 2013 and beyond to develop and validate predictive models at beaches and improve existing nowcasts, restructuring monitoring activities to accommodate future uncertainties in funding and resources.

Introduction

Currently, swim advisories or closings issued by beach managers in the United States are based on standards for concentrations of *Escherichia coli* (*E. coli*) or enterococci. Standard culture methods for these fecal-indicator bacteria (FIB) take at least 18–24 hours before results are available. The beach is posted with an advisory or closing or is determined to be acceptable for swimming on the basis of the previous day's bacteria concentration (persistence model); however, sanitary conditions may change overnight and even throughout the day (Boehm and others, 2002). Because of the time lag inherent in this approach, water-resource managers are seeking solutions that provide near-real-time estimates of recreational water quality.

One solution to improve the timeliness and accuracy of recreational water-quality assessments is to use predictive models. Predictive models employ environmental and water-quality variables that are easily and quickly measured to either estimate the probability that the State standard will be exceeded or directly estimate concentrations of FIB, such as *E. coli*. In new recreational water-quality criteria recently promulgated by U.S. Environmental Protection Agency (EPA), the use of predictive models to facilitate rapid notification of potential water-quality problems and supplement beach notification programs is encouraged (U.S. Environmental Protection Agency, 2012a).

Model Definitions and Measures of Performance

- **Persistence models** use the previous day's *E. coli* concentration to estimate the current day's *E. coli* concentration.

- **Predictive models** are statistical models that use environmental and water-quality variables to estimate the probability that the State standard will be exceeded or to directly estimate concentrations of *E. coli*.

- Measures of model performance include the following:

  - The overall percentage of **correct** responses (exceedances and nonexceedances) that are predicted by the predictive or persistence model.

  - The **sensitivity** of the model; that is, the percentage of exceedances of the bathing-water standard that are correctly predicted by the predictive or persistence model.

  - The **specificity** of the model; that is, the percentage of nonexceedances of the bathing-water standard that are correctly predicted by the predictive or persistence model.

Existing Nowcast Systems

When predictive models are used for beach closure or advisory decisions, they are referred to as "nowcasts." In 2010–11, operational nowcasts (those used daily by beach managers to inform the public) were used at only a few locations within the Great Lakes, including three beaches in Ohio (*http://www.ohionowcast.info*), three beaches in Illinois (Lake County Health Department, 2013), and two beaches in Wisconsin (*http://www.wibeaches.us/*). Of these operational nowcast systems, the U.S. Geological Survey (USGS) was involved in development and testing of predictive models in Ohio and Wisconsin. The USGS has also been involved in past nowcasts in Indiana. Nowcasts with USGS involvement have used water-quality and environmental variables in predictive models developed using statistical techniques to estimate current water-quality conditions based on *E. coli* concentrations.

The Ohio Nowcast, a cooperative project between the USGS and local agencies, has been providing near-real-time beach advisories to the public on the basis of predictive models since 2006 (*http://www.ohionowcast.info*). The probability of exceeding the bathing-water standard is posted on the Ohio Nowcast Web site by 9:30 a.m. during the recreational season. Local agencies that collect daily data and run the nowcast include the Cuyahoga County Board of Health, Northeast Ohio Regional Sewer District, and the University of Toledo, with postings for Huntington, Edgewater, Maumee Bay State Park, and Villa Angela beaches. Predictive models have included such variables as turbidity, wave height, antecedent rainfall, change in lake level in the past 24 hours, and day of the year. At Huntington—in Bay Village, Ohio, a suburb of Cleveland—nowcasts were provided to the public for 581 days during the recreational seasons of 2006–11. A hindsight examination of the Huntington nowcast results in which the same-day cultured *E. coli* concentration in beach-water samples was used as the criterion for correct response indicated that the nowcast yielded a correct response 84.2 percent of the time. By comparison, use of the previous day's *E. coli* concentration (the persistence model) resulted in a correct response 76.1 percent of the time. The predictive model exhibited greater sensitivity and specificity (54.9 and 89.6 percent) than the persistence model (22.4 and 85.9 percent). Nowcasts for other Ohio beaches yielded similar results, showing that they can better estimate current water-quality conditions than can the persistence model used by most beach managers, especially in regard to sensitivity (*http://www.ohionowcast.info*).

In Wisconsin, various predictive models have been used operationally since 2008. A predictive model was implemented in 2008 for Upper Lake Park Beach (Port Washington, Wis.) using rainfall, turbidity and discharge thresholds to predict *E. coli* concentrations. In 2009, the Wisconsin Department of Natural Resources (WDNR) began assisting the Ozaukee County Health Department by developing a predictive model

for Upper Lake Park Beach. For the 2011 season, WDNR extended their modeling assistance, working with the City of Racine Health Department to include North Beach in Racine. Although both of these beaches had few exceedances of the bathing-water standard during 2011, the predictive model still yielded slightly more accurate information than the persistence model.

In Indiana, the Swim Advisory Forecast Estimate (Project SAFE), providing real-time data on *E. coli* concentrations based on predictive models, was first used during the recreational season of 2005 (U.S. Geological Survey, 2010). The USGS created models for five beaches (Ogden Dunes, West, Wells Street, Lake Street, and Marquette Beaches) in Lake and Porter Counties in Indiana using water-quality data collected in 2004 (Nevers and Whitman, 2005). These beaches are directly affected by fecal contamination from nearby Burns Ditch, particularly during prevailing north winds. Because of the similar impact of Burns Ditch, wind direction, and rainfall on *E. coli* concentrations at these beaches, data from all five beaches were combined for model development. Depending on the prevailing wind direction, different weather and water variables were used. Predicted *E. coli* concentrations by stepwise linear regression techniques were more accurate during north winds (R^2 approximately 0.7 compared to approximately 0.3 during south winds), and the model incorporated parameters from the lake (chlorophyll and turbidity), from Burns Ditch (gage height), and rainfall in the previous 48 hours. SAFE was used for nowcasting in 2005, and beach managers were notified of probabilities for beach closures through email by 10 a.m. each day. In subsequent years, the program has been managed by Gary Sanitary District and the town of Ogden Dunes, with varying numbers of beaches included in the prediction, depending on funding and interagency cooperation.

The USGS developed an amended predictive model for Ogden Dunes in 2011, using water quality and detailed hydrometeorological data collected in 2008 (unpublished data). The Ogden Dunes beaches are made up of three adjacent beaches. A complex empirical model using a complete set of hydrometeorological variables, including wave height, wind directions, wind velocity, current velocity, and rainfall, turned out to have a similar predictive capacity as models using turbidity only. The simple turbidity model yielded an adjusted R^2 of 0.55 when the three beaches at Ogden Dunes were considered as independent locations and an adjusted R^2 of 0.67 when all three beaches were averaged. The turbidity model (averaged for three beaches) was used for nowcasting by the Town of Ogden Dunes in 2011 for the first time and yielded a correct response 90 percent of the time. The persistence model resulted in 83.7 percent of correct responses. The predictive model had a higher sensitivity and specificity (61.2 and 97.4 percent, respectively) than the persistence model (59 and 90.1 percent).

Expanding Nowcasts

Given the success of existing operational nowcasts, expanding the use of predictive models to other areas is seen as a welcome addition to many beach monitoring programs. Coordinated leadership, technical expertise, funding, and automated systems are needed to accomplish this expansion. Work is also needed to identify new variables to improve existing operational models and install and maintain onsite equipment for automated measurements of model variables. Automated measurements of wave height or turbidity by use of nearshore buoys offer real-time measurements of model variables without the need to send personnel to the beach.

The 2010–12 predictive-modeling efforts by the USGS described in this report were designed to provide technical expertise on predictive-model development and data collection whenever requested by local and State agencies. These USGS efforts were funded by the EPA through the Great Lakes Restoration Initiative and by the USGS Coastal and Marine Geology Program through the Oceans Research Priority Plan. The cooperation of State and local agencies was made possible through funding from a variety of sources including EPA Beaches Environmental Assessment and Coastal Health (BEACH) Act Funding (U.S. Congress, 2000) and the Great Lakes Restoration Initiative and from services provided by the agencies themselves.

Purpose and Scope

This report describes results of efforts by the USGS and State and local agencies to improve and expand operational nowcasting at beaches throughout the Great Lakes region. During the recreational seasons (May–September) of 2010–12, the USGS, in cooperation with 23 local and State agencies, worked to improve existing operational nowcast systems at 4 beaches, validate predictive models at another 38 beaches, and collect data for predictive-model development at 7 beaches in Illinois, Indiana, Michigan, New York, Ohio, Pennsylvania, and Wisconsin. Data were collected during 2010–11 to identify and compare the explanatory variables that best described *E. coli* concentrations at each beach. For some beaches, data collected before 2010 were also used. Real-time measures of turbidity, wave height, specific conductance, water temperature, and other variables were obtained from equipment installed in swimming areas at several beaches. Water-level and water-quality data in tributaries expected to affect beach water quality also were collected by use of similar equipment. Additionally, data on photosynthetically active radiation measured by a sensor and shallow groundwater levels measured in a piezometer were evaluated to determine whether these variables could improve existing operational nowcasts at Ohio beaches.

Software routines or spreadsheets were designed by the USGS to automate the compilation of data from other agencies including lake level, rainfall, wind direction and speed, predicted current direction, and stream discharge. A free software tool designed by EPA for model development, called Virtual Beach, was used to develop or update models by means of multiple linear regression techniques (U.S. Environmental Protection Agency, 2012b). An alternative statistical technique, partial least squares regression (Brooks and others, 2013), was tested and incorporated into Virtual Beach. The selected model for each beach was based on a combination of explanatory variables including (most commonly) turbidity, day of the year, change in lake level over 24 hours, wave height, wind direction and speed, and antecedent rainfall for various time periods. Updated or new models for 42 beaches were validated during an independent year (2012). Model performance was compared to use of the persistence model and was examined to determine whether model results met established criteria and could be used for future public notifications.

This report summarizes efforts to collect data and develop predictive models by multiple agencies. Some models have been published previously, and this report compiles existing information into one comprehensive report on past and current USGS modeling efforts throughout the Great Lakes. Background information on study areas was obtained from previous studies whenever available and through discussions with beach managers and public health officials, resulting in different amounts of information among the study beaches. The report is somewhat modular in approach. In several sections, information is presented by State to help the reader focus on specific areas without having to read the report from cover to cover.

Study Areas

Studies were done at 49 Great Lakes beaches in Illinois, Indiana, Michigan, New York, Ohio, Pennsylvania, and Wisconsin (fig. 1) in cooperation with 23 local agencies (table 1). The USGS Site Identification Numbers (table 2) correspond to latitudes and longitudes of each site (with two additional digits) and can be used to pinpoint site locations by using any digital map system. Altnernatively, detailed site information can be obtained for most beaches through EPA's Beach Web site (*http://watersgeo.epa.gov/beacon2/*). All of the study sites are popular swimming beaches, and their locations span a variety of land-use settings. Beaches were placed in the following categories based on nowcast status during the recreational season of 2012 (table 1):

- *Operational.*—An existing nowcast was used to issue advisories or closings to the public before 2012 and continued to be used during 2012.

- *Recently implemented.*—Models were developed and used to issue advisories or closings to the public for the first time in 2012.

- *Validation.*—At least 2 years of data were collected, and models were developed and validated during 2012 but not used to issue advisories or closings to the public.

- *Development.*—At least 1 year of data was collected, and agencies continued to collect data during 2012 for model development.
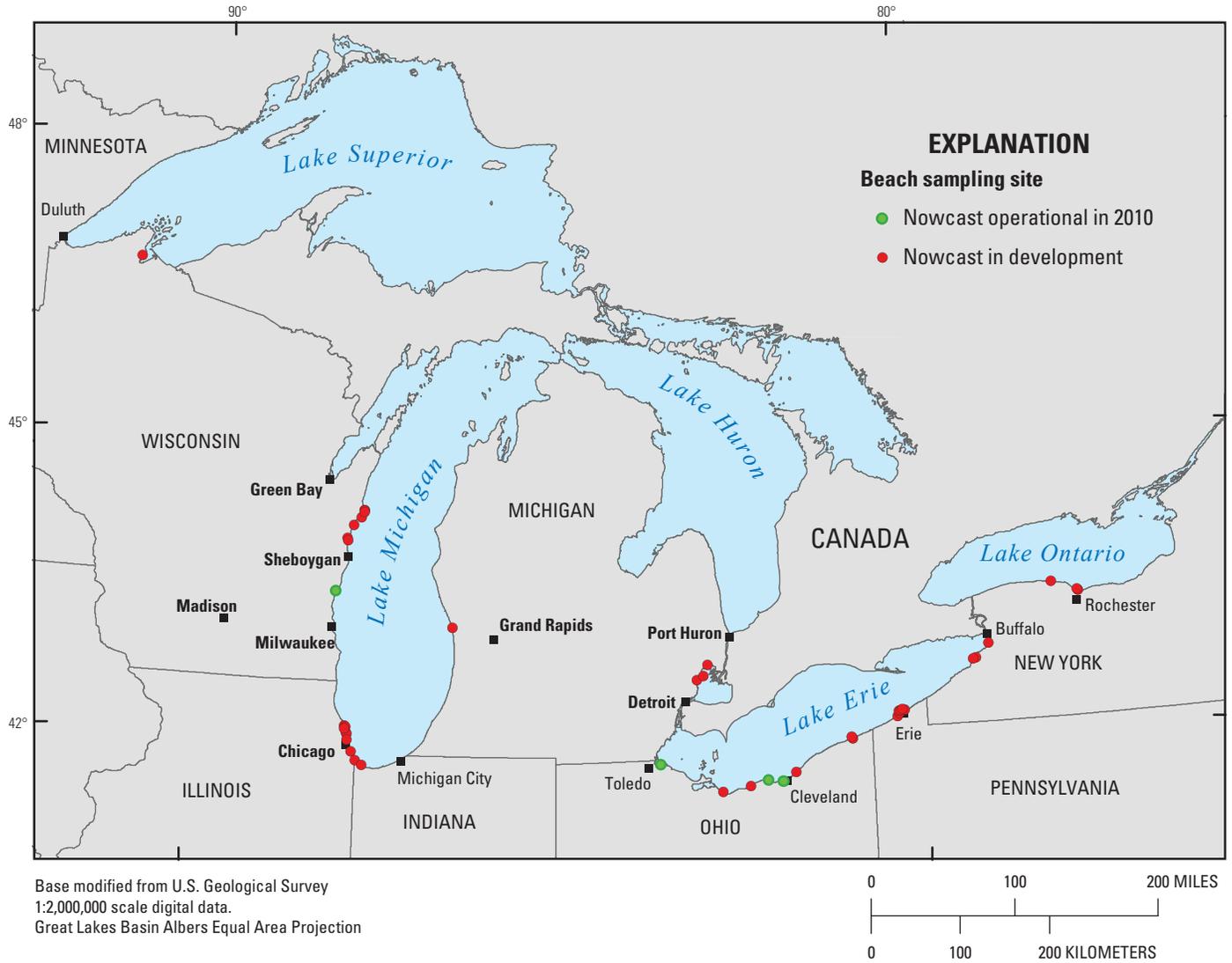
Figure 1. Locations of 49 beach sampling sites in the Great Lakes region of Illinois, Indiana, Michigan, New York, Ohio, Pennsylvania, and Wisconsin (site details are listed in table 1).

Table 1. List of 49 study beaches, cooperating agencies, and nowcast status in the Great Lakes region of Illinois, Indiana, Michigan, New York, Ohio, Pennsylvania, and Wisconsin.

Beach name and location	Beach short name (for this report)	Cooperating agencies	Nowcast status in 2012[a]
Illinois			
Evanston South, Evanston	South Beach	City of Evanston Health Dept.	Validation
Greenwood, Evanston	Greenwood Beach	City of Evanston Health Dept.	Validation
Lee, Evanston	Lee Beach	City of Evanston Health Dept.	Validation
Montrose Beach, Chicago	Montrose	Chicago Park District	Recently implemented
Jackson Park/63rd Street Beach, Chicago	63rd Street	Chicago Park District	Recently implemented
Foster Beach, Chicago	Foster	Chicago Park District	Recently implemented
Calumet Beach, Chicago	Calumet	Chicago Park District	Recently implemented
Oak Street Beach, Chicago	Oak	Chicago Park District	Recently implemented
Indiana			
Jeorse Park, East Chicago	Jeorse Park	Indiana Dept. of Environmental Management; East Chicago Health Dept.	Validation
Michigan			
Metropark Beach, Mount Clemens	Metropark	Macomb County Health Dept.	Validation
Memorial Beach, St. Clair Shores	Memorial	Macomb County Health Dept.	Validation
New Baltimore Beach, New Baltimore	New Baltimore	Macomb County Health Dept.	Development
Grand Haven State Park Beach, Grand Haven	Grand Haven	County of Ottawa Health Dept.	Development
New York			
Hamlin Beach State Park Area 3, Hamlin	Hamlin SP Area 3	New York State Office of Parks, Recreation, and Historic Preservation; Monroe County Public Health Laboratory	Development
Hamlin Beach State Park Area 4, Hamlin	Hamlin SP Area 4	New York State Office of Parks, Recreation, and Historic Preservation; Monroe County Public Health Laboratory	Development
Ontario Beach Park, Rochester	Ontario	Monroe County Dept. of Health	Validation
Durand-Eastman Park, Rochester	Durand	Monroe County Dept. of Health	Validation
Woodlawn Beach State Park, Hamburg	Woodlawn SP	New York State Office of Parks, Recreation, and Historic Preservation; Erie County Public Health Laboratory	Validation
Evans Town Park, Evans Center	Evans TP	Erie County Health Dept.	Validation
Lake Erie Beach, Lake Erie Beach	Lake Erie	Erie County Health Dept.	Validation
Ohio			
Edgewater, Lakefront Reservation, Cleveland	Edgewater	Northeast Ohio Regional Sewer District	Operational
Huntington Reservation, Bay Village	Huntington	Cuyahoga County Board of Health	Operational
Maumee Bay State Park, Oregon	Maumee Bay SP	University of Toledo	Operational
Villa Angela, Cleveland	Villa Angela	Northeast Ohio Regional Sewer District	Recently implemented
Vermillion West, Vermillion	Vermillion West	Erie County Health Dept.	Validation
Nickel Plate, Huron	Nickel Plate	Erie County Health Dept.	Validation

Table 1. List of 49 study beaches, cooperating agencies, and nowcast status in the Great Lakes region of Illinois, Indiana, Michigan, New York, Ohio, Pennsylvania, and Wisconsin.—Continued.

Beach name and location	Beach short name (for this report)	Cooperating agencies	Nowcast status in 2012[a]
Ohio—Continued			
Lake Front, Huron West Beach, Huron	Lakefront	Erie County Health Dept.	Validation
Fairport Harbor Lakefront Park Beach, Fairport Harbor	Fairport Harbor	Lake County General Health District	Validation
Headlands Beach State Park, Mentor	Mentor Headlands	Lake County General Health District	Validation
Lakeview, Lorain	Lakeview	Cuyahoga County Board of Health	Validation
Lakeshore, Ashtabula	Lakeshore	Lake County General Health District	Development
Walnut Beach, Ashtabula	Walnut	Lake County General Health District	Development
Pennsylvania			
Presque Isle State Park Beach 1 West Extension, Erie	PISP Beach 1 West Extension	Erie County Department of Health; Presque Isle State Park; Regional Science Consortium	Validation
Presque Isle State Park Beach 2, Erie	PISP Beach 2	Same as above	Validation
Presque Isle State Park Beach 6, Erie	PISP Beach 6	Same as above	Validation
Presque Isle State Park Mill Road Beaches, Erie	PISP Mill Rd Beach	Same as above	Validation
Presque Isle State Park Beach 9, Erie	PISP Beach 9	Same as above	Validation
Presque Isle State Park Beach 11, Erie	PISP Beach 11	Same as above	Validation
Wisconsin			
Upper Lake Park Beach, Port Washington	Upper Lake Park	Ozaukee County	Operational
Hika Park Bay Beach, Cleveland	Hika Park	University of Wisconsin–Manitowoc	Validation
Fisher Park Beach, Cleveland	Fischer	University of Wisconsin–Manitowoc	Validation
Red Arrow Park Beach, Manitowoc	Red Arrow	University of Wisconsin–Manitowoc	Validation
Neshota Beach, Two Rivers	Neshotah	University of Wisconsin–Manitowoc	Validation
Point Beach State Forest Lighthouse Picnic Area South	Point South	University of Wisconsin–Manitowoc	Validation
Point Beach State Forest Lakeshore Picnic Area Central	Point Central	University of Wisconsin–Manitowoc	Validation
Point Beach State Forest concession Stand North	Point North	University of Wisconsin–Manitowoc	Validation
Kreher Park Beach, Ashland	Kreher	Ashland County	Validation
Maslowski Beach, Ashland	Maslowski	Ashland County	Validation
Thompson's West End Park, Washburn	Thompson	Bayfield County	Validation

[a] Operational—an existing nowcast was used to issue advisories or closings to the public before 2012 and continued to be used during 2012.

Recently implemented—models were developed and were used to issue advisories or closings to the public for the first time in 2012.

Validation—at least 2 years of data were collected and models were developed and validated during 2012, but not used to issue advisories or closings to the public.

Development—at least 1 year of data was collected and agencies continued to collect data during 2012 for model development.

Table 2.   Beach sites and identification numbers, sampling locations, and equipment installed by the U.S. Geological Survey (USGS) or local agencies for predictive-model data collection, 2010–12, at 49 beaches in the Great Lakes region.

Beach short name	USGS station name	USGS site identification number (incorporates latitude and longitude)	Sampling locations and equipment
Illinois			
South Beach	South Beach at Evanston, IL	420140087400301	Single sampling location.
Greenwood Beach	Dempster-Greenwood Beach at Evanston, IL	420235087401601	Single sampling location.
Lee Beach	Lee Beach at Evanston, IL	420211087401001	Single sampling location.
Montrose	Lake Michigan at Montrose Beach at Chicago, IL	415800087381301	Geometric mean *E. coli* for two sampling locations; buoy for turbidity, waves, and water temperature (data transmitted in near real-time).
Calumet	Lake Michigan at Calumet Park at Chicago, IL	414250087314001	Same as above.
63rd Street	Lake Michigan at 63rd St Beach at Chicago, IL	414657087342501	Geometric mean *E. coli* for two sampling locations; buoy for turbidity, waves, and water-temperature (data transmitted in near real time); weather station for winds, rainfall, air temperature, solar radiation, relative humidity, barometric pressure.
Foster	Lake Michigan at Foster Ave at Chicago, IL	415844087385601	Same as above.
Oak	Lake Michigan at Oak St Beach at Chicago, IL	415412087371901	Same as above.
Indiana			
Jeorse Park	Lake Michigan at Jeorse Park at East Chicago, IN	413902087260001	Single sampling location.
Michigan			
Metropark	Metropark Beach on Lk St Clair at Mount Clemens, MI	423415082474601	Geometric mean *E. coli* for multiple sampling locations as required by State of Michigan.
Memorial	Memorial Beach on Lk St Clair at St. Clair Shores, MI	423138082521601	Same as above.
New Baltimore	New Baltimore Beach on Lk St Clair at New Baltimore, MI	424039082440201	Same as above.
Grand Haven	Grand Haven State Park Beach on Lk Michigan at Grand Haven, MI	430307086144301	Same as above.
New York			
Hamlin SP Area 3	Hamlin Beach State Park at Area 3, NY	432155077572101	Single sampling location.
Hamlin SP Area 4	Hamlin Beach State Park at Area 4, NY	432203077574701	Single sampling location.
Ontario	Ontario Beach Park, Rochester, NY	431535077363101	Single sampling location.
Durand	Durand-Eastman Park, Rochester, NY	431428077341901	Single sampling location.
Woodlawn SP	Woodlawn Beach State Park, Hamburg, NY	424722078510901	Single sampling location; Rush Creek staff gage.
Evans TP	Evans Town Park, NY	423837079035301	Single sampling location; Fern Brook staff gage.
Lake Erie	Lake Erie Beach, NY	423729079054501	Single sampling location; Muddy Creek staff gage.

Table 2. Beach sites and identification numbers, sampling locations, and equipment installed by the U.S. Geological Survey or local agencies for predictive-model data collection, 2010–12, at 49 beaches in the Great Lakes region.—Continued.

Beach short name	USGS station name	USGS site identification number (incorporates latitude and longitude)	Sampling locations and equipment
Ohio			
Edgewater	Lake Erie at Edgewater, Main Beach East and West, 3 ft	412922081442303 412920081442603	Average for two sampling locations.
Edgewater	Lake Erie at Edgewater Beach	412923081442600	Buoy for turbidity, waves, and water-temperature (data transmitted in near real time).
Huntington	Lake Erie at Huntington Reservation 2 (central location) and West	412928081560220 412929081561100	Average *E. coli* for two sampling locations; buoy for turbidity, waves, and temperature (central location, data transmitted in near real time).
Huntington	Huntington Beach Tower	412926081560100	Sensor for photosynthetically active radiation.
Maumee Bay SP	Maumee Bay at Maumee Bay State Park Cove 2, Cove 3, and Cove 4	414110083222700 414111083223200 414110083223900	Average *E. coli* for three sampling locations.
Villa Angela	Lake Erie at Villa Angela beach east, OH	413510081340100	Single sampling location.
Vermilion West	Lake Erie at Vermilion West Beach nr Vermilion, OH	412531082215800	Single sampling location.
Nickel Plate	Lake Erie at Nickel Plate Beach at Huron, OH	412348082323900	Single sampling location.
Lake Front	Lake Erie at Huron River West Beach near Huron OH	412355082331800	Single sampling location.
Fairport Harbor	Lake Erie at Fairport Harbor Beach	414530081163100	Single sampling location.
Mentor Headlands	Lake Erie at Headlands SP Beach East and West	414527081172800 414524081173100	Average *E. coli* for two sampling locations[a]; buoy for turbidity, waves, and temperature (east location, data transmitted in near real time).
Mentor Headlands	Weather station near Headlands State Park, OH	414514081174400	Weather station for precipitation, wind speed and direction, barometric pressure, air temperature, net solar radiation, and photosynthetically active radiation (data transmitted in near real time).
Lakeview	Lake Erie at Lakeview Beach East	412753082114200	Single sampling location.
Lakeshore	Lake Erie at Lakeshore Central Beach	415430080462903	Single sampling location.
Walnut	Lake Erie at Walnut Beach at Ashtabula, OH	415407080483200	Single sampling location.
Pennsylvania			
PISP Beach 1 West Extension	Presque Isle 1 West Ext Center	420652080091901	Single sampling location.
PISP Beach 2	Presque Isle Beach 2 East	420755080084501	Buoy for waves, water currents, turbidity, pH, dissolved oxygen, conductivity, water temperature; weather station for winds, rainfall, air temperature, solar radiation, relative humidity, barometric pressure (data transmitted in near real time).
PISP Beach 6	Presque Isle Beach 6 Center	420839080081801	Single sampling location; buoy for turbidity, pH, dissolved oxygen, conductivity, and water temperature (data transmitted in near real time).

Table 2.   Beach sites and identification numbers, sampling locations, and equipment installed by the U.S. Geological Survey or local agencies for predictive-model data collection, 2010–12, at 49 beaches in the Great Lakes region.—Continued.

Beach short name	USGS station name	USGS site identification number (incorporates latitude and longitude)	Sampling locations and equipment
		Pennsylvania—Continued	
PISP Mill Rd Beach	Presque Isle Millroad Beach Center	420944080071501	Single sampling location.
PISP Beach 9	Presque Isle Pine Tree Beach Center	421012080062101	Single sampling location.
PISP Beach 11	Presque Isle Beach 11 Center	420945080044201	Single sampling location.
		Wisconsin	
Upper Lake Park	Lake Park Beach at Port Washington, WI	432345087514501	Single sampling location.
Hika Park	Hika Park Beach at Cleveland, WI	435457087432401	Single sampling location; rain gage, Centerville Creek staff gage.
Fischer	Fischer Cr Conservation Area Beach nr Cleveland, WI	435613087430801	Single sampling location; Fischer Creek staff gage.
Red Arrow	Red Arrow Park Beach at Manitowoc, WI	440434087392001	Single sampling location; rain gage; storm sewer staff gage.
Neshotah	Neshotah Beach at Two Rivers, WI	440903087331001	Single sampling location; rain gage; storm sewer staff gage.
Point South	Point Beach State Forest Rawley Point Beach, WI	441243087302401	Single sampling location; Molash Creek staff gage.
Point Central	Point Beach State Forest Lakeshore Picnic Beach, WI	441303087302401	Single sampling location.
Point North	Point Beach State Forest Concession Area Beach, WI	441323087302401	Single sampling location; rain gage.
Kreher	Kreher Park Beach at Ashland, WI	463549090525401	Single sampling location; Bay City Creek staff gage.
Maslowski	Maslowski Park Beach at Ashland, WI	463502090551801	Single sampling location; Fish Creek and Whittlesey Creek staff gages.
Thompson's West End Park	Thompson's West End Park Beach at Washburn, WI	463957090541601	Single sampling location; rain gage; Thompson Creek staff gage.

[a] If a sample from one sampling location exceeded the single-sample standard of 235 colony-forming units per 100 milliliters (235 CFU/100 mL) and the average was less than 235 CFU/100 mL, the highest value among the two locations was used to err on the safe side.

Illinois—Evanston Beaches

The City of Evanston, Ill., adjacent to the northern border of Chicago, operates six public beaches on Lake Michigan. All of the beaches have been formed or built up with groins at the southern ends. These groins generally extend 350 to 500 feet (ft) out from the shoreline. One beach is designated for dog activity and is not used for public bathing. The most southern beach, aptly named South Beach, has historically been closed more often than the other beaches because of elevated *E. coli* concentrations. The Evanston Health Department has been collecting samples for *E. coli* concentrations and environmental data at its beaches since 2007. Data collection for predictive models, in cooperation with the USGS, began in 2010 at three beaches: South Beach, Greenwood Beach, and Lee Beach.

Illinois—Chicago Beaches

Chicago's Lake Michigan beaches have been the subject of many previous investigations to identify causes for periodic beach closures. All of the beaches are in an urban environment with no known point-source contamination. The situation of these beaches is varied, with many partially enclosed by breakwaters, a factor that may facilitate the accumulation of FIB. The 63rd Street Beach has elevated concentrations of FIB more frequently than the other 22 Chicago beaches and is largely enclosed by groins. Attempts to develop predictive models for 63rd Street Beach have met with varying success (Olyphant and Whitman 2004), but more recent research has identified the hydrodynamic processes that may be influencing this beach (Ge and others, 2010, 2012) and could lead to the refinement of predictive capabilities. Other beaches that have elevated bacteria concentrations and may be potential candidates for predictive models are Foster, Montrose, Calumet, and Oak Street. Predictive models have been developed for these beaches by using historical monitoring data (Nevers and Whitman, 2011), and a regional model that attempts to explain bacteria fluctuations along the entire Chicago coast was developed that highlighted relationships of FIB with wave height, day of year, and barometric pressure (Whitman and Nevers, 2008). One year of data to develop predictive models was collected for Chicago beaches (2010), and a second year of data was collected with slight modifications to the methods supported by a cooperative grant that facilitated the installation of automated buoys and weather systems (2011). Predictive models were used to issue advisories at five Chicago beaches for the first time in 2012, with considerable testing and refinement expected to be needed.

Indiana

Jeorse Park Beach on Lake Michigan in East Chicago, Ind., is one of the most frequently closed beaches in the Great Lakes. Since beach monitoring was initiated in 2004, the frequency of beach closings has continued to increase each year. Bird populations have been identified as a potential source of contamination (M.N. Byappanahalli, U.S. Geological Survey, written commun., 2011), but source-tracking attempts have been inconclusive. Predictive-model testing was initiated in 2010 as part of this study, but because of poor model performance and other reasons, the effort was discontinued in 2011. Although the beach was never a strong candidate for predictive modeling given the high rate of exceedances (78.1 percent in 2010), predictive modeling was explored to determine whether any particularly strong variable was related to *E. coli* concentrations. Predictive models do not work as well as the persistence model at beaches where the standard is exceeded on the majority of days (Francy, 2009). Initiating the modeling helped with focusing activities toward eliminating potential contamination sources.

Michigan

In Macomb County, Mich., three beaches are occasionally impaired because of high *E. coli* concentrations: Metropark, Memorial, and New Baltimore, all on Lake St. Clair. The USGS conducted a previous study at two of these beaches, where investigators found that rainfall, water temperature, and turbidity were related to *E. coli* concentrations (Holtschlag and others, 2008). In Ottawa County, Grand Haven State Park Beach on Lake Michigan is a popular beach for citizens and tourists, with attendance in the hundreds on most summer days. High *E. coli* concentrations are rare at Grand Haven, but because of its popularity, there remains interest in protecting the public by the most cost effective and timely methods. Data collection and model validation were done by Macomb County Health Department at two beaches during this study: Metropark and Memorial. At New Baltimore and Grand Haven State Park, data were collected for model development by Macomb County and the County of Ottawa Health Department, respectively.

New York

Hamlin State Park (SP) is on Lake Ontario in Monroe County, north of the town of Hamlin and west of the city of Rochester. Two beaches were included for predictive-model development during this study, both operated by New York State Office of Parks, Recreation, and Historic Preservation: Hamlin SP Area 3 and Hamlin SP Area 4. The two beaches are separated by jetties and Spring Brook Marsh. Hamlin SP Area 3 has 2 miles (mi) of lakefront with swim areas bounded by

stone jetties. Yanty Creek Marsh, to the east of Area 3, drains a predominantly agricultural watershed. Hamlin SP Area 4 has approximately 1.6 mi of shoreline west of Area 3. Potential bacteria sources and modeling parameters for Hamlin SP were documented through a detailed sanitary survey (Kirsten Husson, New York State Office of Parks, written commun., 2012).

Monroe County Department of Health operates two public beaches on Lake Ontario, northwest of the city of Rochester: Ontario and Durand beaches. Both beaches were included for model validation in 2012. Ontario is a popular urban beach located west of the mouth of the Genesee River. Durand is a smaller urban beach at Durand Eastman Park and 1.8 mi east from the mouth of the Genesee River. The Genesee River is likely the greatest contributor of fecal contamination and sediments to these beaches, with inputs from septic systems, wastewater-treatment-plant effluents, stormwater runoff, and agriculture. Monroe County has been using decision-tree models for beach closure decisions since 1976. These models are based on thresholds for Genesee River discharge (greater than 5,000 cubic feet per second (ft^3/s)) and rainfall totals (greater than 0.7 inches (in.) for 24 hours or greater than 0.4 inch per hour (in/h) of peak rainfall intensity) and include qualitative estimates of algal mats and water clarity. The current decision-tree model produced accurate predictions 60 to 80 percent of the time over the operational periods (1976 to 2011) for *E. coli* exceedances (Charles Knauf, Monroe County Department of Health, oral commun., 2011).

Three public beaches on Lake Erie in New York were included for predictive-model validation during this study: Woodlawn SP, Evans Town Park (TP), and Lake Erie Beach. Woodlawn SP contains approximately 1 mi of Lake Erie shoreline and is approximately 7.2 mi south of the city of Buffalo in the town of Hamburg. The natural areas of the park consist of vegetated sand dunes, wetlands, and wooded areas. The day-use park is bordered by heavy industry and residential apartments and a wastewater treatment plant south of the park. Rush Creek and Blasdell Creek, directly north of the park, are considered the major sources of fecal contamination to the beach from sewage overflows, stormwater, and urban runoff (Eric Wiegert, New York State Health Department, written commun., 2012). Potential pollution sources and possible modeling parameters for Woodlawn SP were documented previously through a detailed sanitary survey (Terbush and Husson, 2009). During 2000–2008, closures based on rainfall amounts (greater than 0.5 in. for 24 hours) resulted in 16 percent false negatives (E. Costanzo, New York State Office of Parks, oral commun., 2011). Woodlawn SP beach closures are currently based on the persistence model. The Evans TP and Lake Erie beaches are south of Buffalo and less than a mile apart between the moderately populated towns of Evans Center and Lake Erie Beach. Evans TP beach is a recreational area with a small swimming section along the shoreline surrounded by residential housing and wooded areas. Fern Brook contains stormwater runoff and seepage from rural septic systems and is likely the main source of fecal contamination to Evans TP. Fern Brook empties directly on to the sands along the south

side of the swimming area and is often channelized during beach operation and after a storm event. Lake Erie Beach is a small beach directly north of Evangola State Park encompassing two small tributaries that discharge directly adjacent to the swimming area. Muddy Creek to the south is the larger of the two and the likely source of fecal contamination at this beach. Muddy Creek flow contains inputs from septic systems, wastewater-treatment-plant effluent, combined-sewer overflows, storm runoff, and agricultural runoff. Beach closures for Evans TP and Lake Erie beaches are based on the persistence model and (or) a rainfall threshold (greater than 0.5 in. for 24 hours). The beach operators may also close the beaches if a chemical substance sheen or physical deposition that may affect water quality is present, as described in the New York State Department of Health Sanitary Code (Scott Zimmerman, Erie County Department of Health, oral commun., 2011).

Ohio

Studies were done at 12 Ohio Lake Erie beaches: 3 with existing operational nowcasts, 1 with a recently implemented nowcast, 6 with model validation, and 2 with model development during 2012.

Edgewater, Huntington, and Maumee Bay State Park (Maumee Bay SP) have been part of the Ohio Nowcast since 2008, 2006, and 2011, respectively. Edgewater is an urban recreational area in Lakefront Reservation where water circulation is restricted by groins to the west and east of the main swimming area. Earlier work at Edgewater indicated that local sources of fecal contamination were dominant, the Edgewater outfall (to the west of the beach) was not a major source of fecal contamination, and elevated *E. coli* concentrations were found in shallow groundwater. The infiltration of bird feces through beach sands and interaction of shallow groundwater and waves were identified as a likely mechanism of *E. coli* storage and accumulation in beach sands and transport to lake water (Francy and others, 2006a). At Huntington in Bay Village, a western suburb of Cleveland, two outfalls discharge stormwater runoff from a parking lot into the lake, and a creek to the east drains a heavily populated area (Francy and others, 2006b). Wave height, turbidity, antecedent rainfall, and day of the year have typically been used in the Ohio Nowcast for Edgewater and Huntington. Maumee Bay SP is in the southwest corner of Lake Erie near the cities of Toledo and Oregon. This Lake Erie beach is in a popular state park and consists of five separate coves used for swimming. In an earlier study, remote sources were identified as minor contributors of *E. coli* to the beach, and a ditch to the east of the beach (Berger Ditch) was identified as a principal source of *E. coli* (Francy and others, 2005). At Maumee Bay SP, wind direction has been a dominant explanatory variable used in the Ohio Nowcast. Wind directions from the north and east (sweeping contamination from Berger Ditch into the beach area) have been associated with higher *E. coli* concentrations at Maumee Bay SP.

Villa Angela is part of Lakefront Reservation and is on the east side of Cleveland. At Villa Angela, four breakwaters and a groin to the west stabilize the beach area. Overflow from a sanitary-sewage pump station near Euclid Creek, directly to the east of the beach, is a probable source of fecal contamination. Previous work at Villa Angela indicated that predictive models did little to improve upon the persistence model (Francy and others, 2003). As a result, Villa Angela was not included in the Ohio Nowcast before 2012. Recent work, however, showed that a rapid analytical method for *E. coli* concentrations, quantitative polymerase chain reaction (qPCR), could be used to derive a measure that improved correct responses over using the persistence model (Mark Citriglia, Northeast Ohio Regional Sewer District, written commun., 2012). With qPCR results as a predictive variable, Villa Angela was added to the Ohio Nowcast in 2012.

Two years of data were collected for model validation in 2012 at three beaches in Erie County (Vermillion West, Nickel Plate, and Lake Front) by the Erie County Health Department, two beaches in Lake County (Fairport Harbor and Mentor Headlands) by the Lake County General Health District, and one beach in Lorain County (Lakeview) by the Cuyahoga County Board of Health. The Erie County beaches are in two small cities along the Lake Erie shoreline: Vermillion and Huron. Vermillion West lies between the Vermillion River to the west, which is affected by combined-sewer overflows, and a groin to the east. Nickel Plate is a ½-mi-long popular recreational area east of the Huron River. Lake Front is a small swimming area in a Huron city park, west of the mouth of the Huron River. The Erie County beaches are potentially affected by stormwater runoff and discharges of wastewater. Fairport Harbor and Mentor Headlands are east of Cleveland in a less populated area. Fairport Harbor is a small beach protected by breakwaters with a shallow swimming area and little wave action. Mentor Headlands is a 1-mi-long, heavily used beach that is open to the lake and susceptible to high waves. Possible sources of fecal contamination to Lake County beaches include effluents from a wastewater treatment plant, sanitary-sewer overflows, and sanitary-sewer exfiltration to the stormwater drainage system (Susan Bell, Lake County General Health District, oral commun., 2012). The Grand River drains into Lake Erie directly west of Fairport Harbor and east of Mentor Headlands. In an earlier study (Francy and others, 2003), a model was developed for Mentor Headlands and was based on wave height, rainfall, and turbidity; however, because of too few exceedances that year, a model was not developed for Fairport Harbor. At Lakeview, an urban beach in Lorain, Ohio, three breakwaters restrict water circulation, and large populations of waterfowl frequent the swimming area. In an earlier study, data indicated that wind direction, day of the year, and turbidity were significantly related to *E. coli* concentrations at Lakeview (Francy and others, 2006b).

One year of data were collected by the Lake County General Health District for development of models at Lakeshore Park and Walnut Beach. Walnut Beach is in the city of Ashtabula in northeast Ohio, and Lakeshore Park is just east of Ashtabula. The mouth of the Ashtabula River lies to the east of Walnut and to the west of Lakeshore. Sources of fecal contamination to Lakeshore include a waterfowl pond that drains to the west of the beach and stormwater from a parking lot outfall that drains directly to the beach (Francy and others, 2006a). Rainfall, turbidity, and water temperature were found to be significantly related to *E. coli* concentrations at Lakeshore (Francy and others, 2006b). Walnut Beach has a small swimming area open to the lake and has not been previously investigated by the USGS.

Pennsylvania

Pennsylvania has 40 mi of Lake Erie shoreline, 8 mi of which are permitted public bathing beaches (Dorfman and Rosselot, 2011). Monitored public bathing beaches at Presque Isle State Park in Erie, Pennsylvania (Pa.), make up less than 1 mi. Previous studies by the Erie County Department of Health (ECDH) and other agencies (Erie County Department of Health, 2007) identified sources of fecal contamination at the Presque Isle beaches that included stormwater runoff and combined-sewer overflows. These are likely the same type of sources that continue to influence concentrations of FIB at Presque Isle beaches, especially during periods of heavy rainfall.

A model was developed as part of a previous study using data collected at one Presque Isle State Park (PISP) beach during three recreational seasons (2004–6). Investigators found that $\log_{10}$ turbidity, bird count, and wave height were related to *E. coli* concentrations at PISP Beach 2 (Zimmerman, 2008). In 2010–11, data were collected at PISP Beach 2 and five additional PISP beaches (Beach 1 West Extension, Beach 6, Mill Road Beach, Beach 9, and Beach 11) for model development and validation in 2012.

Wisconsin

Studies were done at eight Lake Michigan beaches and three Lake Superior beaches in Wisconsin. Lake Michigan beaches are Upper Lake Park in Port Washington and seven beaches in Manitowoc County. The Manitowoc County beaches are, from south to north, Hika, Fischer, Red Arrow, Neshotah, and three beach areas along the shore of Point Beach State Park (Point South, Point Central, and Point North). The Lake Superior beaches are all within Chequamegon Bay, with Kreher and Maslowski in the City of Ashland and Thompson immediately south of the city of Washburn.

At Upper Lake Park, a predictive model had already been developed by WDNR and the Ozaukee County Health Department. Previous predictive models included variables for rainfall, water temperature, cloud conditions, wave height, currents, and date. Influences on Upper Lake Park include two nearby small tributaries to the south with a mix of urban and agricultural land use, and one small tributary to the north with primarily agricultural land use. Predictive models were

developed for nowcast validation in 2012 for the remaining 10 Wisconsin beaches.

The Manitowoc County beaches represent a range of different land use influences from the surrounding drainage areas. Hika and Fischer are south of the city of Manitowoc near the small community of Cleveland, Wis. Both have small streams draining to the lake adjacent to each beach that are dominated by agricultural land use. Centerville Creek drains at the northern boundary of Hika, and Fischer Creek drains within the boundaries of Fischer. Red Arrow is within the city of Manitowoc. This beach has numerous potential influences on water quality, including two storm sewers that drain urban areas from Manitowoc, the mouth of the Manitowoc River 1 mi north of the beach, and a sewage treatment plant that discharges 0.75 mi to the northeast of the beach. The Manitowoc River is dominated by agricultural land use, but there is some urban influence from Manitowoc. Neshota is in the small community of Two Rivers. Small storm sewers drain to the north and to the south directly adjacent to the beach boundaries, and the mouth of the Twin River is 0.5 mi south of Neshota. The Twin River drains an agricultural watershed. Point Beach State Park is approximately 11 mi north of Manitowoc. These three beaches are situated approximately along a 1-mi section of shore that is approximately 2.5 mi north from the mouth of Molash Creek, whose watershed encompasses a mix of agricultural land use and wetland area. The mouth of Twin River is 6.3 mi south and the mouth of the Kewaunee River is 16 mi north of Point Beach State Park. The Kewaunee River is also dominated by agricultural land use.

The three Chequamegon Bay/Lake Superior beaches are influenced by nearby streams that drain into Chequamegon Bay, as well as by urban runoff from Ashland and Washburn. Chequamegon Bay is approximately 12 mi long and ranges from 2 to 6 mi in width, with a maximum depth of 35 ft. Numerous small streams draining to the bay have potential to affect the water quality at these three beaches. The closest streams to the study beaches are the following: City Creek is adjacent to Kreher and is influenced by urban land use and forested areas, Fish Creek and Whittlesey Creek are near Maslowski, and Thompson Creek is adjacent to Thompson. All three of these streams are influenced by a mix of agricultural land use and forested areas.

Methods

Data were collected during the recreational seasons (May–September) of 2010–12 for most beaches, with earlier years' data included for some beaches. Daily data were collected by local agencies (table 1) and included analysis of water samples for *E. coli* concentrations that were part of existing monitoring programs designed to conform to State water-quality standards, described below. At the time of *E. coli* sample collection, local agencies made field measurements and recorded observations for model development

and validation. Nearshore buoys and other equipment were installed or monitored by the USGS and (or) local agencies specifically for this project to collect site-specific water-quality and meterological data. Data on explanatory variables from other sources were compiled by local agencies or by the USGS.

Recreational Water-Quality Standards and Criteria for Closings and Advisories

Water-quality standards for *E. coli* concentrations and associated actions taken (advisories and closings) differ by State and (or) locality but were all based on the 1986 criteria at the time of this study (2010–12) (U.S. Environmental Protection Agency, 1986). The EPA published new recreational water-quality criteria in December 2012 (U.S. Environmental Protection Agency, 2012a). New recreational water-quality criteria include 30-day geometric means and statistical threshold values (STV) for two estimated illness rates. The EPA also recommended that States could use a Beach Action Value (BAV) as a "precautionary tool for making beach notification decisions." The BAV is not a "component of EPA's recommended criteria" but could be used by States to trigger a notification if any single sample goes above the BAV. Two BAVs were identified for *E. coli*—235 colony-forming units per 100 milliliters (CFU/100 mL) and a stricter, 190 CFU/100 mL—both corresponding to the same illness rates as in the new geometric mean and STV criteria. At the time of this study (2010–12) and as described below, most States used the single-sample bathing water standard of 235 CFU/100 mL as a lower threshold for posting beach advisories or closings. Therefore, models based on culturable *E. coli* data collected during the present study would fit into new BAV State water-quality standards, and model thresholds could be adjusted if the stricter BAV was adopted into a State standard.

Illinois

In Chicago, a swim advisory is posted at a beach when the geometric mean *E. coli* concentration of two water samples from the previous day is equal to or greater than 235 CFU/100 mL. A swim ban is posted at a beach when the geometric mean is equal to or greater than 1,000 CFU/100 mL. This system was revised during the course of this study, such that swimming advisories were issued when *E. coli* concentrations exceeded 235 CFU/100 mL, but swim bans were eliminated. Precautionary swim bans are also issued for 2 days at all beaches following any storm event that initiates the opening of the locks that separate Lake Michigan from the Chicago River system for flood control. Updated beach information is available at the Chicago Park District Web site or by calling 312–74–BEACH or by sending a text message with the name of the beach to 312–715–SWIM. In Evanston, the beach is closed to swimming when the *E. coli* concentration from the previous day is equal to or greater than 235 CFU/100 mL.

Indiana

In Indiana, beaches are generally monitored and managed by use of the single-sample bathing water standard for *E. coli* of 235 CFU/100 mL. If this concentration is exceeded, determinations for issuing a swimming advisory or closing for the beach are at the discretion of the beach manager. Results and advisories using the previous day's *E. coli* concentration are posted through the BeachGuard system (Indiana Department of Environmental Management, 2012).

Michigan

In Michigan, county health departments are not required to conduct beach monitoring; but if they do, they are required to collect a minimum of three samples each time a beach area is monitored (Michigan Department of Environmental Quality, 2012a). The daily geometric mean for *E. coli* concentrations calculated from these three samples must be below 300 CFU/100 mL for the water to be considered safe for swimming. The county health departments frequently sample more than once a month. A minimum of five sampling events (consisting of at least three samples per event) must be collected within a 30-day period for the results to be considered a reliable indicator of water quality. After 30 days, a geometric mean is calculated for all the individual samples collected within that timeframe. The 30-day geometric mean for *E. coli* concentrations must be below 130 CFU/100 mL for the water to be considered safe for swimming. A beach is closed if either the single-day or 30-day concentration exceeds the established limit. If a beach closure is due to bacterial contamination, county health departments will continue to monitor the water quality at the beach and will permit the beach to reopen when bacteria levels fall to within acceptable levels. It is possible that a beach could be closed for swimming but that other recreational activities at the beach may still be available. The State maintains a Web site where each public beach is identified, and monitoring data, as well as beach contact information, are provided throughout the swimming season (Michigan Department of Environmental Quality, 2012b).

New York

In New York, swim advisories are issued by beach managers if the previous day's *E. coli* concentration is greater than or equal to the single-sample bathing water standard of 235 CFU/100 mL. The condition of the beach is also considered in the decision process. An advisory may also be issued if there are other indications of chemical contamination, visible deposits or oil sheens, or hazardous algal or other aquatic growth (New York State Department of Health, 2010). If the *E. coli* concentration is equal to or greater than 1,000 CFU/100 mL, beaches are posted with a swimming ban, and swimming is not permitted. In addition, beaches are closed when a threshold of 0.5 in. of rain in a 24-hour period is exceeded at locations with nearby tributary or outflow influences.

Ohio

In Ohio, a swimming advisory is issued at all beaches (except the Ohio Nowcast beaches) if the previous day's *E. coli* concentration exceeds the single-sampling bathing water standard of 235 CFU/100 mL (Ohio Department of Health, 2012a). Results and advisories using the previous day's *E. coli* concentration or the nowcast are posted through a BeachGuard system (Ohio Department of Health, 2012b) by local agencies that monitor their beaches.

Pennsylvania

In Pennsylvania, swimming advisories are issued by beach managers if the previous day's *E. coli* concentration is greater than or equal to the single-sample bathing water standard of 235 CFU/100 mL but less than 1,000 CFU/100 mL (Erie County Department of Health, 2012). If the *E. coli* concentration is equal to or greater than 1,000 CFU/100 mL, beaches are posted with a swimming restriction, and swimming is not permitted. In addition, precautionary advisories are issued at the beach manager's discretion prior to receiving any bacteria results when conditions are observed that historically produced elevated *E. coli* concentrations.

Wisconsin

In Wisconsin, beaches are prioritized into high, medium, and low priority with monitoring conducted most frequently at high-priority beaches and least frequently at low-priority beaches (Wisconsin Department of Natural Resources, 2012a). Advisories are posted at all of these beaches when a single sample concentration for *E. coli* exceeds 235 CFU/100 mL. Advisories are also posted for high-priority beaches when a geometric mean of at least five samples collected over a 30-day period exceeds 126 CFU/100 mL. In addition, beaches are closed under any of the following four conditions: whenever a human health hazard, such as a reported illness, exists as determined by the local health department; after a major pollution event; after a significant rainfall event that is determined to affect a beach area; and whenever the *E. coli* concentration in a beach water sample exceeds 1,000 CFU/100 mL. Beach advisories are posted with signs directly on the beach as well as on the Wisconsin Beach Health Web site (Wisconsin Department of Natural Resources, 2012b). An option to be alerted by email for specific beach conditions also is available.

Data Collection and Sample Analysis

Water samples for *E. coli* concentrations and turbidity (all beaches), chlorophyll *a* (Illinois and Indiana, only), and other field observations and measurements were collected during the recreational season at 49 beaches (table 2). Other field measurements included discrete data collected at the time of water-sample collection and continuous data collected by instruments mounted on buoys, weather stations, streamgages, or other equipment near the beaches. At most beaches, USGS site identification numbers for equipment measurements were the same as site identification numbers for the sampling locations (for example, 63rd Street, Ill.). In Ohio, two USGS site identification numbers were established specifically for measurements at a buoy (Edgewater) and a weather station (Mentor Headlands) (table 2).

Field and Laboratory Measurements

Daily samples (3 to 7 consecutive days each week) were collected during the morning hours throughout the recreational seasons (May–September) from the area(s) used for swimming. Sampling techniques and locations were based on EPA guidance (U.S. Environmental Protection Agency, 2002) and State requirements. Samples were collected into sterile bottles or bags by use of grab-sampling techniques in water depths consistent for each beach, ranging from 1.5 to 3 ft. Samples were collected from one to six sampling sites at each beach (table 2), depending on the length of the beach and (or) State requirements.

Bird counts and debris assessments were made upon arrival at each beach. Wave heights were measured by placing a graduated rod at the sampling location and noting minimum and maximum water heights for 1 minute. The wave height was computed as the difference between the maximum and minimum water height. Specific conductance and water temperature were measured at the sampling location by using a digital thermometer and (or) an in situ sensor by use of standard USGS methods (Wilde, variously dated). Turbidity was measured by use of a turbidimeter and an aliquot from the *E. coli* sample bottle or a second subsample bottle, or by use of a an in situ turbidity sensor. At Ontario and Durand beaches, New York (N.Y.), Secchi disk measurements were included at sampling locations to account for transparency of the entire water column.

Water samples were kept on ice and processed within 6 hours for *E. coli* concentrations by use of the Colilert Quanti-Tray/2000 method (IDEXX Laboratories, Inc., Westbrook, Maine) or the modified mTEC method (U.S. Environmental Protection Agency, 2009). For the Colilert Quanti-Tray/2000

method, results are reported in most-probable number per 100 milliliters (MPN/100 mL). These results are considered equivalent to the CFU/100 mL results reported from the modified mTEC method.

At Villa Angela, Ohio, samples were analyzed by the local agency using the quantitative polymerase chain reaction (qPCR) method for *E. coli* as described in Noble and others (2010). The qPCR method was designed to provide rapid analytical results within 4 hours. For this method, a target genetic sequence of DNA present in *E. coli* is amplified into an amount that can be quantified. The direct output from the qPCR method is a cycle threshold (Ct), which is the intersection of the amplification curve and a designated threshold level. The Ct value was used as an explanatory variable in the model. A sample-processing control was run with each sample to detect any inhibition of the qPCR. If the results indicated inhibition, the Ct value was not used and the model was not run. Instead, the ΔΔCt method was used to calculate results (Haugland and others, 2005) in calibrator cell equivalents 100 milliliters (CCE/100 mL), which was then compared to the water-quality standard of 235 CFU/100 mL.

At Chicago beaches in Illinois and at Jeorse Park, Ind., water samples collected at the beach were analyzed in the laboratory for relative fluorescence (a surrogate for chlorophyll *a*) using a fluorometer. Results are reported as relative fluorescence units (RFU).

Nearshore Buoys

At some beaches in Illinois, Pennsylvania, and Ohio, automated measurements were made from buoys placed outside the swimming areas (table 2). The buoys were custom made by environmental equipment manufacturers for use in nearshore shallow waters. Buoys were equipped with dataloggers to store data and a battery and (or) solar panels for power. Equipment for measurements included pressure transducers to measure wave heights and wave periods; thermisters to measure water temperature; and sensors to measure pH, dissolved oxygen, conductivity, and turbidity. Most buoys were equipped with a radio, amplifier, and antenna to transmit the data for display on the Web in near real time. Information specific to each location is the following:

- In Chicago, buoys were installed and maintained by the Chicago Park District and USGS at five beaches in 2011 and seven beaches in 2012, usually at the middle point outside of the swimming area at the depth of 4–6 ft. Data were transmitted in near real time during the recreational season and displayed at *http://v4.wqdata. com/webdblink/chicagopark.php*. Parameters were measured and updated every hour.

- In Ohio, buoys were installed and maintained by the USGS at Huntington in 2008–11, Edgewater in 2005–12, and Mentor Headlands in 2011–12. Wave heights and wave periods were compiled every 30 minutes, and water-quality data were recorded every hour. Buoy data were transmitted in near real time during the recreational season and displayed on the USGS National Water Information System Web site (NWISWeb) (U.S. Geological Survey, 2012).

- At Presque Isle State Park, Erie, Pa., buoys were placed outside the Beach 2 and Beach 6 swimming areas. The buoys were installed by the Regional Science Consortium and were maintained with help from the USGS. The buoy at Beach 2 has been operational since 2008 for most parameters, and instrumentation to measure wave height and water current speed and direction were added in 2011. The buoy at Beach 6 was newly deployed in 2011. All buoy data were transmitted in near real time every 10 minutes during the recreational season and displayed at *http://v4.wqdata.com/webdblink/trec.php?tab2=overview&t=classic_web*. Parameters were measured every 10 minutes.

Weather Stations and Other Equipment Installed by the U.S. Geological Survey

A weather station at one beach in Ohio and rain gages at several beaches in Wisconsin were installed and maintained by the USGS to obtain site-specific weather data for model development. A weather station was installed near the western edge of Mentor Headlands State Park, Ohio, to measure precipitation, wind speed and direction, barometric pressure, air temperature, net solar radiation, and photosynthetically active radiation. Weather data from Mentor Headlands (table 2) were transmitted during the recreational season for near-real-time display on NWISWeb (U.S. Geological Survey, 2012). In Wisconsin, tipping-bucket rain gages yielded rainfall data on a fine time scale (5 minutes or less). These included four rain gages in Manitowoc County that measured rainfall near all seven Manitowoc County beaches and one rain gage in Washburn, Wis., to measure rainfall at Thompson West End Park Beach (table 2).

Staff gages were installed by the USGS on seven tributaries that were expected to influence water-quality at beaches in New York and Wisconsin. A gage-height measurement, to the tenth of an inch, was obtained by visible observation by the beach sampler near the time of sample collection, and the difference from the previous day's measurement was calculated. Previous-day gage-height variables represented changes in discharge over a 24-hour period up to the time of sampling.

Other types of equipment were installed and maintained by the USGS at two beaches in Ohio. A sensor to measure photosynthetically active radiation (PAR) and datalogger were mounted on top of a 50-ft water tower at Huntington to record PAR data every 3 minutes. The total PAR for the previous day (midnight to midnight) was calculated by summing all recorded measurements. At Edgewater, a temporary piezometer (shallow water well) equipped with a pressure transducer and datalogger to measure and record water levels every 30 minutes was installed 20 ft inland from the edge of water. The screened interval was from 1.3 to 4.3 ft below land surface. To aid in beach grooming, the piezometer was buried 1 ft deep; the elevation of the top of the casing was surveyed. Water-level data from the piezometer were corrected for barometric pressure and are referred to as "foreshore head." The foreshore head varies as a function of pressure changes due to the local water table, fluctuations in barometric pressure, the water level in the lake, and the pressure variance due to wave action (Francy and others, 2009). The rate of change in foreshore head, in feet per hour (ft/h), was calculated for each 30-minute time interval. Average rates were then calculated by averaging the rate of change in foreshore-head measurements over the antecedent 6-, 12-, 24-, and 48-hour periods up to the time of sampling.

Compilation of Data From Existing Sources

Data from existing sources were compiled and used to develop and validate predictive models for this study. Data were retrieved manually by the USGS and local agencies, or by use of tools described below and designed specifically to compile and process data for predictive-model development. A lake-level spreadsheet organizes hourly lake-level data retrieved from the nearest offshore buoy; PROCESSNOAA processes hourly rainfall, wind direction and speed, and barometric pressure data retrieved from the nearest airport site; and Environmental Data Discovery and Transformation (EnDDaT) retrieves and processes data from multiple sources.

Weather Data From the National Weather Service

Rainfall, wind direction and speed, and air-temperature data were obtained from the nearest National Weather Service (NWS) airport site by compiling data daily throughout the season (National Oceanic and Atmospheric Administration, 2012a) or in batch for the entire season(s) (National Oceanic and Atmospheric Administration, 2012b). Hourly rainfall values were totaled for 24 hours up to the time of sampling and lagged and weighted as described below; running 30-day totals were calculated for one site. Wind direction and speed were recorded for the approximate time of sampling (generally 8 a.m.) or compiled for the past 24 hours and summed as described below.

In order to obtain information from a more widespread area, hourly radar rainfall data were obtained from the NWS

(National Oceanic and Atmospheric Administration, 2012c) for 4-kilometer (km) grids, or "cells." The number of cells used for calculations ranged from 1 to 102, depending on the configuration of the beach and the areal extent of the watershed. For some beaches, radar rainfall values were totaled for each cell for the previous 24 hours up to the time of sampling. Data were expressed as the maximum value among all the cells (radarMax24) or the summed total for all cells (radarTotal24); data were lagged and weighted as described below in the section "PROCESSNOAA—A Tool for Compiling Weather Data." For other beaches, watershed areas contributing to the beach were delineated in a geographic information system (GIS), as were watershed boundaries for nearby tributaries. The watershed average rainfall was computed by using all radar rainfall cells that influenced these watersheds, as described below in the section "EnDDaT—A Tool for Compiling Data from Multiple Sources."

Stream Measurements From a Nearby U.S. Geological Survey Streamgage

For streams in proximity to beaches (less than 5 mi away), stream-discharge or stream-stage data were obtained from established USGS streamgages. These data are available online (U.S. Geological Survey, 2012) or can be obtained by contacting a state USGS office. The mean, minimum, or maximum stream discharge or stage for the previous 24–72 hours up to the time of sampling or the instantaneous stream discharge at the time of sampling was used in predictive models.

Measurements or Predictions From the National Oceanic and Atmospheric Administration

Lake-level data were obtained from the nearest offshore station (National Oceanic and Atmospheric Administration, 2012d). Based on the time of sampling, change over the past 24 hours or change from spring average values were used. Hourly lake-level data and change over the past 24 hours were compiled for some sites by using the lake-level spreadsheet, as discussed later in the report in the section "Lake-Level Spreadsheet—A Tool for Compiling Lake-Level Data." To compute change from the prerecreational season average (spring), hourly lake levels for April 1–May 15 for the current year were compiled, the prerecreational season average was calculated, and the instantaneous value at the time of sampling was subtracted from the prerecreational season average. Lake level was not used directly as an explanatory variable. Previous work indicated that direct lake levels may be poor predictors of *E. coli* concentrations if lake levels are very different in the calibrator years from the nowcast year (Francy and others, 2009); change in lake level over the past 24 hours or change

from prerecreational season averages were used in this study because they may be better predictors over several years.

Wind direction and speed data for the time of sampling were obtained for one site (Nickel Plate, Ohio) from the National Estuarine Research Reserve System (National Oceanic and Atmospheric Administration, 2012e).

Nowcast or forecast results from Great Lakes Coastal Forecasting System (GLCFS) were available on a 2-km grid throughout the Great Lakes (National Oceanic and Atmospheric Administration, 2012f). Data used from the GLCFS for predictive modeling, which were compiled by using the EnDDat system, included nowcast or forecast winds, wave heights, surface temperatures, and currents.

Local Sources of Weather Data

In addition to established regional or national sources of weather data, local data were obtained from nearby sources near some Illinois and Pennsylvania beaches. Hourly rainfall and wind speed and direction were collected at the Evanston, Ill., water treatment plant on the Lake Michigan shoreline just north of the study beaches. In Chicago, Ill., weather stations were installed by the Chicago Park District at 63rd Street, Oak Street, and Foster Beaches to measure precipitation, wind speed and direction, barometric pressure, air temperature, and solar radiation. Weather data from this site were transmitted in near real time during the recreational season and displayed on the Web at *http://v4.wqdata.com/webdblink/chicagopark.php*. At Presque Isle State Park, Pa., two weather stations were installed in 2008 by the Regional Science Consortium. One weather station is on a tower on top of the Tom Ridge Environmental Center (TREC) at the entrance to Presque Isle State Park; the other weather station is on a tower at Beach 2. Weather data were measured every 10 minutes and were transmitted in near real time year round and displayed on the Web at *http://v4.wqdata.com/webdblink/trec.php?tab2=overview&t=classic_web*.

Lake Level Spreadsheet—A Tool for Compiling Lake-Level Data

Instructions and a spreadsheet for compiling hourly lake-level data from the nearest offshore station (National Oceanic and Atmospheric Administration, 2011d) are included in this report as appendix 1. The spreadsheet was used by study personnel; after retrieving data from the source, instantaneous lake-level data for the time of sampling were compiled, and change in lake level from the previous day's sampling time was calculated.

PROCESSNOAA—A Tool for Compiling Weather Data

Rainfall, wind direction and speed, and barometric pressure data from National Oceanic and Atmospheric Adminsitration (NOAA) airport locations can be compiled in batch by use of a software program called PROCESSNOAA. Instructions for PROCESSNOAA (appendix 2) and the downloadable program are available at *http://pubs.usgs.gov/sir/2013/5166/.*

Rainfall totals were compiled for the 24-hour period preceding the morning sampling (R_{d-1}). The PROCESSNOAA program was used to lag rainfall data and calculate weighted rainfall variables from airport rainfall data. These calculations can also be done manually with rainfall from other sources. The "Rw48" and "Rw72" variables were 48 hours and 72 hours of cumulative rainfall, respectively, giving more weight to the most recent rainfall amount as follows:

$$Rw48 = (2* R_{d-1}) + R_{d-2}$$

$$Rw72 = (3* R_{d-1}) + (2* R_{d-2}) + R_{d-3}$$

where

R_{d-1} is the amount of rain, in inches, that fell in the 24-hour period preceding the morning sampling;

R_{d-2} is the amount of rain that fell in the 24-hour period 2 days preceding the morning sampling; and

R_{d-3} is the amount of rain that fell in the 24-hour period 3 days preceding the morning sampling.

The PROCESSNOAA software was also used to compile and process wind direction and speed data for the time of sampling (approximately 8 a.m.) and for the preceding 24-hour period. The 24-hour wind speed and direction variables were calculated by averaging hourly wind vectors for the 24-hour period preceding sampling and determining the direction and speed of the resultant vector. The instantaneous 8 a.m. and 24-hour wind speed and direction variables were used directly or were used in Virtual Beach to calculate alongshore and offshore wind components (U.S. Environmental Protection Agency, 2012b). The 24-hour wind speed and direction variable, however, could not be easily obtained on a daily basis and was determined to be impractical for an operational nowcast; this variable was omitted from the latest version of PROCESSNOAA.

For some sites, wind directions compiled from PROCESSNOAA were placed into categories by examining patterns in plots of *E. coli* concentrations in lake-water samples as a function of wind direction. Wind codes were calculated by assigning the highest value to the range of wind directions associated with the highest *E. coli* concentrations. Processes affecting *E. coli* concentrations were also considered to ensure that the wind direction categories could be reasonably explained as a function of the nature and locations of potential *E. coli* sources.

EnDDaT—A Tool for Compiling Data From Multiple Sources

A Web tool, Environmental Data Discovery and Transformation (EnDDaT), was developed by the USGS to facilitate retrieval of data from multiple Web-based sources and combine them into one dataset (*U.S. Geological Survey, 2013*). Currently, this tool can access national sources of data such as the USGS NWISWeb, the GLCFS (a 2-km grid throughout the Great Lakes), radar-indicated precipitation from the NOAA National Mosaic Quantitative Precipitation Archive (4-km grid for the United States), the National Data Buoy Center, the National Weather Service Surface Summary of Day data, and a link to nearby meteorological data from the National Climatic Data Center. In addition, this tool is currently evolving and incorporating other relevant sources of environmental data that become available.

Variables retrievable from the GLCFS included surface current magnitude and direction, depth-integrated current magnitude and direction, wave height, wave direction, deviation from average lake level, surface water temperature, and wind speed and direction. Access to these model results was facilitated by use of EnDDaT, including a mapping interface for choice of grid points, definition of sample times, specifying the beach orientation resolution of velocity vectors and direction of waves to the beach orientation, and statistical processing of the model results time series. Additional detail on the use of EnDDaT is available in the user manual (*http://cida.usgs.gov/enddat/*).

Development of Predictive Models

At least 2 years of data on *E. coli* concentrations and field measurements (except for Jeorse Park where only 1 year of data was used) and data from existing sources were compiled into spreadsheets and formatted to develop site-specific predictive models for *E. coli*. Data exploration, model development, and model validation were done by use of Virtual Beach (U.S. Environmental Protection Agency, 2012b) or, less often, by use of other statistical software. Procedures for developing models are described in Francy and Darner (2006), and these steps generally were followed whether Virtual Beach or other software was used. Briefly, because of the wide range of expected values and to improve linear relations between *E. coli* concentrations and explanatory variables, concentrations of *E. coli* were $\log_{10}$-transformed before any statistical testing and modeling. The relations between explanatory variables and *E. coli* concentrations were examined by using Pearson's correlation analysis and by associated scatterplots. Explanatory variables that showed statistically significant linear relations to *E. coli* concentrations, or nonlinear relations in scatterplots, were selected for model development. Statistical methods for model development included linear regression and partial least squares (PLS) techniques described elsewhere (U.S. Environmental Protection Agency, 2012b).

For each beach, a model was selected to be validated and (or) used in nowcasts during the 2012 recreational season. Beach managers selected one of two model outputs: (1) a predicted *E. coli* concentration or (2) the probability of exceeding the State standard. If using the probability as the output, threshold probabilities for advisories or closings were established by examining model-output sensitivities and specificities at different probability levels.

Data Management and Analysis

Daily data (from field and lab measurements or compiled from other sources) and model results (from Virtual Beach) were compiled by local agencies into spreadsheets or entered into a sanitary survey database for Wisconsin data, described below.

Example spreadsheets for compiling daily data are included in appendix 3. Daily data for sample times, field observations, model variables, *E. coli* concentrations, and exceedance probabilities of a State standard or predicted *E. coli* concentrations ("model prediction" in Virtual Beach software and in appendix 3) were entered into the spreadsheet. The exceedance probability was used in the example for Beach 1 and was obtained by entering values for explanatory variables into the Virtual Beach software and obtaining an exceedance probability as output from the model. The predicted *E. coli* concentration was used in the example for Beach 2 and was similarly obtained from Virtual Beach software. The spreadsheet was used to compile responses to facilitate comparisons of the model output (exceedance probability or predicted *E. coli* concentration), threshold (probability or *E. coli* concentration), and the measured *E. coli* concentration. Assessment of a correct response was based on the measured *E. coli* concentration determined by the culture method for the day of prediction. Model responses were classified as follows:

- *Correct nonexceedance* (*Correct -*).—The exceedance probability (or predicted *E. coli*) was below the threshold, and the measured *E. coli* concentration was below the bathing-water standard.

- *False positive* (*false +*).—The exceedance probability (or predicted *E. coli*) was equal to or above the threshold, and the measured *E. coli* concentration was below the bathing-water standard.

- *Correct exceedance* (*Correct +*).—The exceedance probability (or predicted *E. coli*) was equal to or above

the threshold, and the measured *E. coli* concentration was above the bathing-water standard.

- *False negative* (*false -*)—The exceedance probability (or predicted *E. coli*) was below the threshold, and the measured *E. coli* concentration was above the bathing-water standard.

Data on *E. coli* concentrations, wave heights, turbidity, and water temperature measured at the time of sampling were entered into the USGS database for water-quality data available to the public through NWISWeb (U.S. Geological Survey, 2012). Data were entered in batch format at the end of the season, with codes for local collecting and analyzing agencies and station identification numbers for each sampling location (table 2). Data were also available in near real time on NWISWeb during the recreational season for measurements obtained from selected water-quality sensors installed to support the development of nowcasts (table 2).

The model responses for the calibration dataset (data used to develop the model) and validation dataset (data collected during an independent year) were evaluated in terms of the correct predictions, sensitivities, and specificities. The validation dataset included a comparison to the use of the persistence model. To determine whether responses for correct predictions, sensitivities, and specificities were significantly different between the nowcast and persistence models, responses were compared by use of the paired student's t-test in SAS 9.2 (SAS Institute Inc., Cary, North Carolina).

Data from Wisconsin sites were managed through a USGS daily sanitary survey database designed specifically for nowcast data. The daily sanitary survey database was project specific and not available to the public. It included access to online forms for entering water-quality and beach-observation data. Data were entered at the "beach level" for data elements with only one sampling point per beach. Data were entered at the "monitoring point level" for data elements that were collected at more than one site per beach. Data were entered for relevant tributaries for data elements such as water level, turbidity, and specific conductance. All beach and tributary data were stored in a relational database. Access to all data residing in the database was available through online query forms by specifying the beach, choosing parameters of interest, choosing the file format, and specifying the date range. Final file format options included XML, tab delimited, comma delimited, or Microsoft Excel.

Quality Assurance and Quality Control

Quality-assurance and quality-control (QA/QC) procedures were implemented to facilitate the collection of high-quality datasets for model development. Written protocols, similar to the example presented in appendix 4, were distributed to local agencies if locally written protocols were not available. The USGS performed QA/QC checks of procedures performed by field and laboratory personnel and provided written comments and recommended corrective actions. For most sites, the USGS checked spreadsheets compiled by local agencies against daily field forms. Field quality-control (QC) samples collected for *E. coli* concentrations included field blanks and replicates. Laboratory QC samples for *E. coli* included reference samples provided by the USGS, filter blanks, and processing blanks (Myers and others, 2007). For turbidity, duplicate aliquots were measured from the same bottle, and duplicate measurements that did not agree within 15 percent (or 1.5 nephelometric turbidity units (NTU) for values less than 10 NTU) were repeated. Turbidity reference samples were sent to all participating laboratories. Results from QC samples were monitored, and retests were done and corrective measures taken when needed.

Automated measurements from buoys placed outside the swimming areas at some beaches in Illinois, Pennsylvania, and Ohio were monitored throughout the season for abnormal fluctuations. If an abnormal fluctuation occurred, the equipment was removed, checked, and replaced (if needed). At Chicago beaches, turbidity, wave height, and temperature sensor measurements were checked several times throughout the season per manufacturer's instructions. At the end of the season, data obtained from Chicago-area buoys were screened for abnormal or negative values, and these data points were removed for predictive modeling. For example, unusually high turbidity measurements randomly occurring within datasets were compared against measurements from other buoys and removed if needed. At Pennsylvania beaches, the continuous water-quality instrumentation on the buoys was recalibrated every 4 to 6 weeks.

The USGS installed and operated a weather station at Mentor Headlands, Ohio, and rain gages at several beaches in Wisconsin. Rain gages were calibrated at the beginning and end of the season in accordance with standard USGS methods (U.S. Geological Survey, 2009). The Chicago Park District installed three onsite weather stations and made repairs as needed. Weather data from the Chicago sites were screened for abnormal and negative values, which were subsequently removed from datasets. If there were long periods of missing data due to malfunctioning of the equipment, missing data were substituted with readings from the closest onsite weather station.

Identification of Variables and Development of Predictive Models

Summary statistics for *E. coli* concentrations for the sampling period for this project at 49 Great Lakes beaches and percent exceedances for the calibration and validation year(s) are listed in table 3. For this study, the number of samples collected ranged from 32 samples at Jeorse Park, Ind., to 829 samples at Edgewater, Ohio. *E. coli* concentrations ranged from less than 1 CFU or MPN/100 mL (both units will be reported hereinafter as "CFU/100 mL") at several sites to at least 13,000 CFU/100 mL at Ontario, N.Y. Median concentrations of *E. coli* ranged from 10 CFU/100 mL at Lake Erie, N.Y., to 570 CFU/100 mL at Jeorse Park, Ind. During the recreational season of the calibration year(s), the percentage of days that the *E. coli* bathing-water standard was exceeded ranged from 3.9 percent at Upper Lake Park, Wis., to 78.1 percent at Jeorse Park, Ind. During the recreational season of the validation year (2012), the percentage of days that the *E. coli* bathing-water standard was exceeded ranged from 0 percent at Lake Erie, N.Y., and Oak Street, Ill., to 51 percent at Lakeview, Ohio.

Table 3. Sampling information and summary statistics at Great Lakes beaches for period of sampling and exceedances for calibration and validation years.

[CFU, colony-forming units; MPN, most-probable number; mL, milliliters; na, not applicable because model was not validated during 2012]

Beach short name	Sampling years and frequency/week	Number of samples	Daily *E. coli* concentrations during calibration and validation years in CFU or MPN/100 mL			Percentage of days standard was exceeded during calibration years (2010, 2010–11, or earlier)	Percentage of days standard was exceeded during validation year (2012)	Difference between calibration and validation years for percentage of days standard was exceeded
			Median	Minimum	Maximum			
Illinois								
South Beach	2010-12, 5 days	186	72	1	1,000	20.0	16.2	3.8
Greenwood Beach	2010-12, 5 days	179	27	1	1,100	11.0	6.2	4.8
Lee Beach	2010-12, 5 days	186	32	1	960	7.0	4.4	2.6
Montrose	2010-12; 3–5 days	192	76	3	>2,400	22.3	26.3	-4.0
63rd Street	2010-12; 3–5 days	184	47	1	>2,400	15.6	24.0	-8.4
Foster	2010-12; 3–5 days	182	26	1	1,800	6.5	10.8	-4.3
Calumet	2010-12; 3–5 days	190	57	2	1,800	16.4	16.3	0.1
Oak Street	2011-12; 5 days	138	14	1	1,800	4.5	0.0	4.5
Indiana								
Jeorse Park	2010; 3 days	32	570	15	>2,400	78.1	na	na
Michigan								
Metro	2010-12, 2–4 days	187	27	1	>2,400	16.0	1.5	14.5
Memorial	2010-12, 2–4 days	179	63	1	>2,400	20.7	4.4	16.3
New Baltimore	2010-12, 2–4 days	181	52	1	>2,400	7.0	na	na
Grand Haven	2011, 4 days	65	77	6	1,400	12.3	na	na
New York								
Hamlin SP Area 3	2011-12; 3-5 days	107	84	1	>3000	27.1	na	na
Hamlin SP Area 4	2011-12; 3-5 days	104	99	4	>3000	29.8	na	na
Ontario	2010-12; 7 days	253	125	1	13,000	36.4	31.8	4.6
Durand	2010-12; 7 days	229	43	1	4,200	20.6	15.9	4.7
Woodlawn SP	2010-12; 5 days	266	80	4	>7,600	30.8	12.1	18.7
Evans TP	2010-12; 5 days	201	20	4	1,000	13.4	3.5	9.9
Lake Erie	2010-12; 5 days	201	10	4	4,800	16.9	0.0	16.9

[CFU, colony-forming units; MPN, most-probable number; mL, milliliters; na, not applicable because model was not validated during 2012]

Beach short name	Sampling years and frequency/week	Number of samples	Daily *E. coli* concentrations during calibration and validation years in CFU or MPN/100 mL			Percentage of days standard was exceeded during calibration years (2010, 2010–11, or earlier)	Percentage of days standard was exceeded during validation year (2012)	Difference between calibration and validation years for percentage of days standard was exceeded
			Median	Minimum	Maximum			
Ohio								
Edgewater	2005-12; 5–7 days	829	67	3	4,800	23.8	13.0	10.8
Huntington	2005-12; 4–7 days	692	45	1	5,600	15.5	19.0	-3.5
Maumee Bay SP	2008-12; 4 days	157	57	4	3,600	16.2	22.4	-6.2
Villa Angela	2011-12, 7 days	190	190	2	12,000	48.5	44.8	3.7
Vermillion West	2010-12; 3–4 days	143	45	1	>2,400	29.7	19.2	10.5
Nickel Plate	2010-12; 3–4 days	143	45	1	>2,400	15.4	17.3	-1.9
Lake Front	2010-12; 3–4 days	143	96	3	>2,400	28.6	19.2	9.4
Fairport Harbor	2010-12; 7 days	310	49	1	>2,400	14.6	17.1	-2.5
Mentor Headlands	2010-12; 7 days	310	33	1	1,700	16.6	14.3	2.3
Lakeview	2005, 11-12; 4, 7 days	242	240	11	6,000	50.7	51.0	-0.3
Lakeshore	2011-12, 4 days	111	140	1	>2,400	35.2	na	na
Walnut	2011-12, 4 days	111	23	<1	870	9.3	na	na
Pennsylvania								
Beach 1 West Extension	2010-12; 4–7 days	223	38	<1	1,800	18.3	18.5	-0.2
Beach 2	2004-6, 10-12; 4, 4–7 days	372	20	<1	1,200	7.5	12.0	-4.5
Beach 6	2010-12; 4–7 days	223	40	<1	1,300	8.4	14.1	-5.7
Mill Road	2010-12; 4–7 days	221	28	<1	720	7.8	15.6	-7.8
Beach 9	2010-12; 4–7 days	219	28	<1	700	10.6	7.9	2.7
Beach 11	2010-12; 4–7 days	221	22	<1	5,500	5.6	10.0	-4.4
Wisconsin								
Upper Lake Park	2009-2012; 4 days	140	12	<1	1,300	3.9	25.0	-21.1
Hika Park	2010-2012; 4 days	128	79	<1	>2,400	28.2	18.6	9.6
Fischer	2010-2012; 4 days	132	52	<1	>2,400	19.0	20.8	-1.8
Red Arrow	2010-2012; 4 days	140	112	<1	>2,400	34.2	25.0	9.2
Neshotah	2010-2012; 4 days	127	29	<1	1,300	6.8	1.9	4.9

[CFU, colony-forming units; MPN, most-probable number; mL, milliliters; na, not applicable because model was not validated during 2012]

Beach short name	Sampling years and frequency/week	Number of samples	Daily *E. coli* concentrations during calibration and validation years in CFU or MPN/100 mL			Percentage of days standard was exceeded during calibration years (2010, 2010–11, or earlier)	Percentage of days standard was exceeded during validation year (2012)	Difference between calibration and validation years for percentage of days standard was exceeded
			Median	Minimum	Maximum			
Wisconsin—Continued								
Point Beach South	2010-2012; 4 days	136	46	<1	>2,400	16.7	15.5	1.2
Point Beach Central	2010-2012; 4 days	136	41	<1	>2,400	15.4	13.8	1.6
Point Beach North	2010-2012; 4 days	127	33	<1	>2,400	10.1	12.1	-2.0
Thompson's West End Park	2010-2012; 4 days	128	22	<1	>2,400	16.9	5.9	11.0
Kreher Park	2010-2012; 4 days	116	29	<1	>2,400	18.8	10.6	8.2
Maslowski	2010-2012; 4 days	137	76	2	>2,400	11.5	32.0	-20.5

Identification of Variables Used to Improve Operational Nowcasts in Ohio and Wisconsin

Data were examined to determine the potential for new variables to improve performance of three nowcasts in Ohio and Wisconsin that were operational at the start of the study in 2010. In Ohio, correlations between *E. coli* concentrations and photosynthetically active radiation (PAR) and foreshore-head variables were evaluated. In Wisconsin, correlations between *E. coli* concentrations and cloud cover, air temperature, water temperature, and directional wave height variables were evaluated.

Ohio

The data representing the linear regression relations between $\log_{10}$ *E. coli* concentrations and sum of the PAR for the previous day are shown for Huntington (fig. 2*A*) and Edgewater (fig. 2*B*) for 2008–12. The PAR sensor was operating at Huntington; however, data for Edgewater were included because Edgewater is approximately 10 mi east of Huntington along the shore of Lake Erie. The *E. coli* concentrations were negatively related to PAR at Huntington and Edgewater with similar slopes and y-intercepts; however considerable scatter is present in these relations.

Pearson's correlations between $\log_{10}$ *E. coli* concentrations and the sum of the PAR for the previous day are listed in table 4 for each year and for all years combined. At Huntington, the correlation was stronger during 2010 (r = −0.57) than for the other years, and the only year where a significant correlation was not found was in 2012 (r = 0.02). At Edgewater, correlations were strongest during 2011 (r = −0.37), and the only year where a significant correlation was not found was in 2008 (r = −0.15). By using datasets with observations that had PAR data, the correlations between *E. coli* concentrations and PAR were compared qualitatively to the relations between *E. coli* and the variables used in predictive models. At Huntington and Edgewater, these relations were not as strong for PAR as they were for sum of the log turbidity and wave height or for radar rain 48 hours weighted (Rw48). At Huntington,

the relation between *E. coli* and PAR was similar to that for the wind variable or R_{d-1} for most years and was stronger than that for airport rainfall R_{d-2} for all years except 2012. Airport rainfall R_{d-2} was a variable used in the Huntington backup model. At Edgewater, the relation between *E. coli* and PAR was stronger than day of the year (a weakly related variable, if at all) and was stronger than lake-level change for most years.

A piezometer with a pressure transducer to measure water levels in shallow groundwater ("foreshore head") was installed at Edgewater during 2008–09 and 2011–12. An example of one time period (May 23–June 30, 2012) is shown in figure 3. Elevated *E. coli* concentrations (greater than the bathing-water standard of 235 CFU/100 mL) were usually associated with periods of increasing foreshore head. Because the foreshore head can be influenced by rainfall, lake level elevations, and wave heights, these ancillary measurements are included on the plot. The patterns of lake levels and wave heights were similar to those of foreshore head, but significant rainfall events were not usually associated with spikes in foreshore head.

Pearson's r correlations between $\log_{10}$ *E. coli* concentrations and the average rate of change in foreshore head for different time intervals are listed in table 5 for each year and for all years combined. Over all years, the 12-hour average was most highly correlated with *E. coli*, with the highest single-year correlation for the average rate of 12 hours in 2009 (r = 0.53). Significant correlations were found for each year and all years combined using the average rates of 6, 12, 24, and 48 hours, except during 2008 using the average 6 hour rate (r = 0.14). Using datasets with foreshore-head data, the correlations between *E. coli* concentrations and foreshore head were compared qualitatively to the relations between *E. coli* and the variables used in predictive models. The relations between *E. coli* concentrations and the 12-, 24-, and 48-hour average rates were comparable to or slightly weaker than the relations between *E. coli* and the radar rain 48 hours weighted (Rw48) or sum of the log turbidity and wave height variables. Stronger correlations were found between *E. coli* concentrations and these three foreshore lake variables than with day of the year or lake-level change.

A.

B.
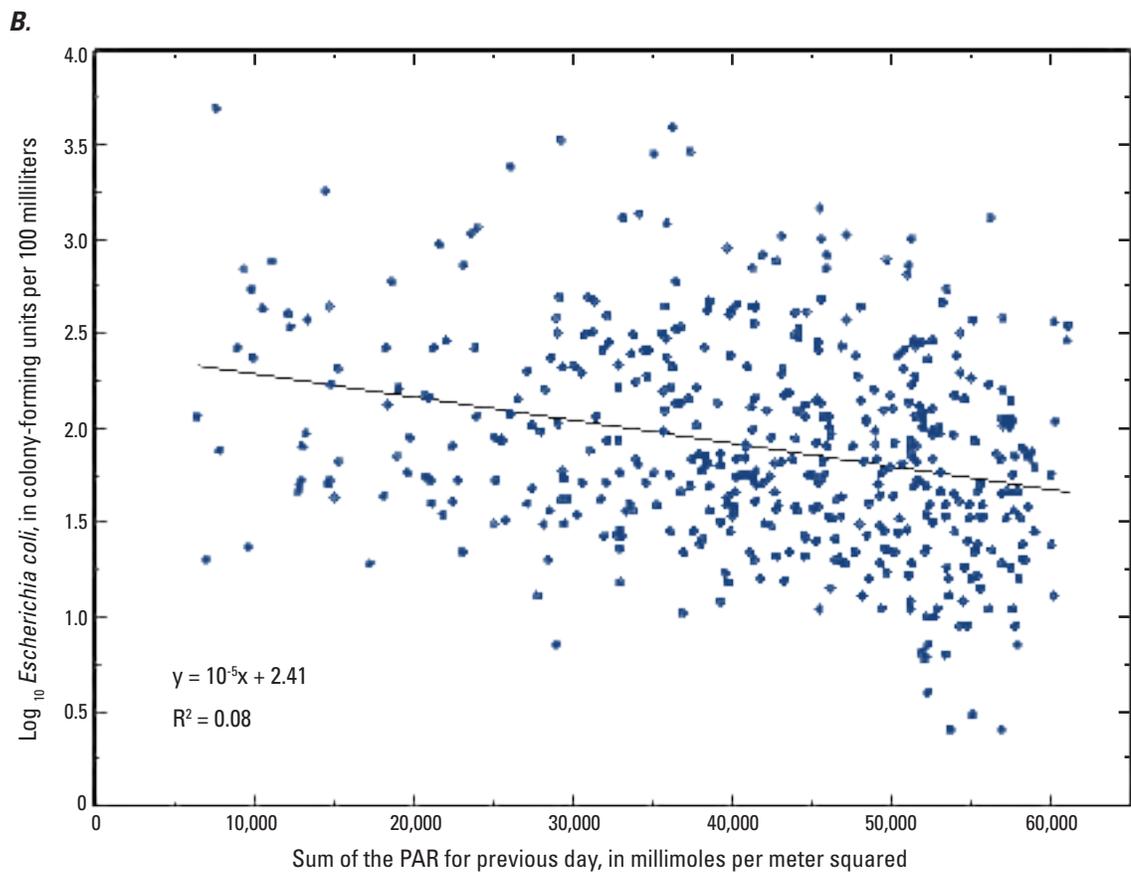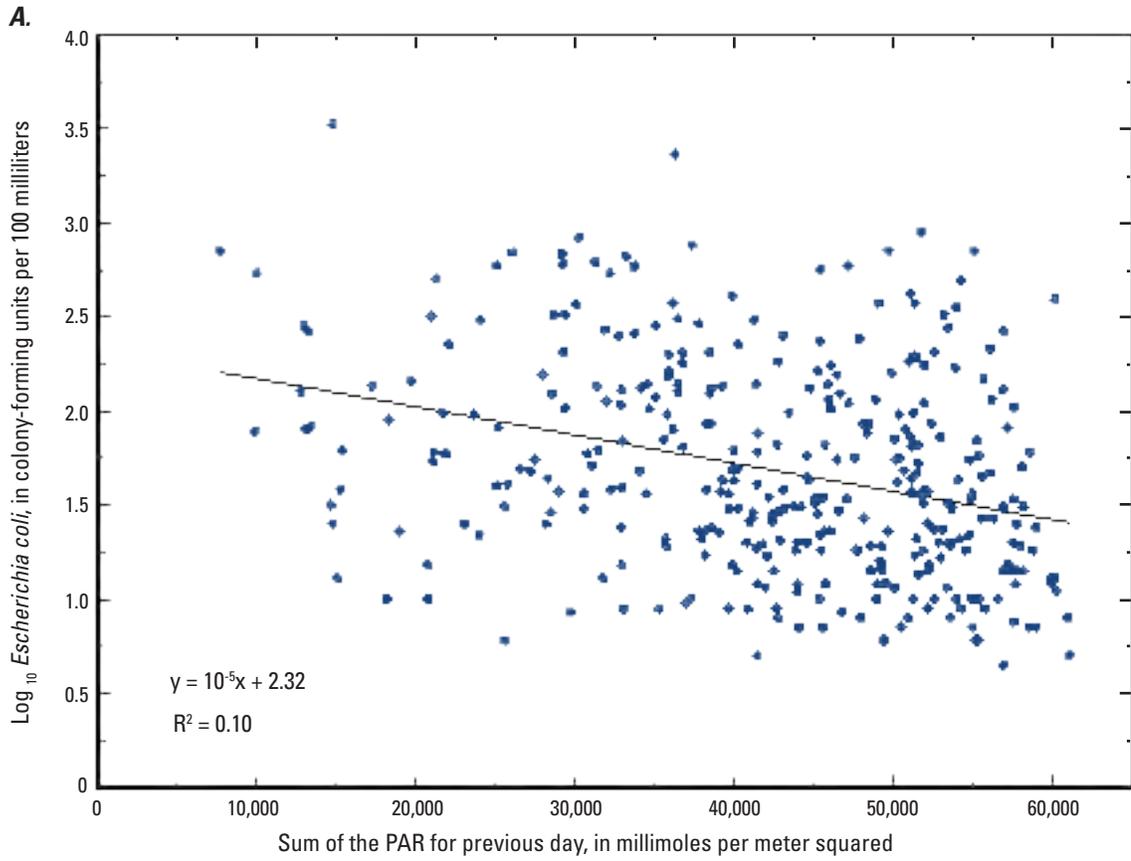
Figure 2. Relations between $\log_{10}$ *Escherichia coli* concentrations and photosynthetically active radiation (PAR) for 2008–12 at two Ohio beaches. *A*, Huntington. *B*, Edgewater. (r, Pearson's correlation coefficient)

Table 4.   Pearson's correlations between $\log_{10}$ *Escherichia coli* concentrations and photosynthetically active radiation (PAR) or explanatory variables used in predictive models at Huntington, Bay Village, and Edgewater, Cleveland, Ohio, 2008–12.

Variable	2008	2009	2010	2011	2012	All years
Huntington						
Number of samples	81	103	57	54	58	353
PAR, sum of the previous day	-0.35 (0.0012)	-0.39 (<0.0001)	-0.57 (<0.0001)	-0.33 (0.0136)	0.02 (0.8554)	-0.32 (<0.0001)
Sum, log turbidity and wave height	0.59 (<0.0001)	0.68 (<0.0001)	0.63 (<0.0001)	0.60 (<0.0001)	0.65 (<0.0001)	0.62 (<0.001)
Radar rain, 48 hours weighted (Rw48)[a]	0.45 (<0.0001)	0.59 (<0.0001)	0.32 (0.0184)	0.62 (<0.0001)	0.28 (0.0325)	0.46 (<0.0001)
Wind code x wind speed, instantaneous	0.42 (0.0001)	0.39 (<0.0001)	0.25 (0.0615)	0.46 (0.0004)	0.20 (0.1229)	0.34 (<0.0001)
Airport rainfall, Rd-1[b]	0.31 (0.0048)	0.51 (<0.0001)	0.57 (<0.0001)	0.53 (<0.0001)	0.31 (0.0188)	0.40 (<0.0001)
Airport rainfall, Rd-2[cd]	0.07 (0.5131)	0.27 (0.0067)	0.18 (01888)	0.20 (0.1484)	0.19 (0.1555)	0.16 (0.0022)
Edgewater						
Number of samples	88	117	86	59	114	464
PAR, sum of the previous day	-0.15 (0.1560)	-0.30 (0.0009)	-0.30 (0.0048)	-0.37 (0.0035)	-0.23 (0.0123)	-0.28 (<0.0001)
Sum, log turbidity and wave height	0.42 (<0.0001)	0.58 (<0.0001)	0.53 (<0.0001)	0.54 (<0.0001)	0.32 (0.0006)	0.48 (<0.0001)
Radar rain, 48 hours weighted (Rw48)[a]	0.35 (0.0008)	0.39 (<0.0001)	0.25 (0.0221)	0.65 (<0.0001)	0.38 (<0.0001)	0.39 (<0.0001)
Day of the year	0.00 (0.9374)	-0.07 (0.4778)	0.10 (0.3637)	0.14 (0.2761)	0.31 (0.0009)	0.16 (0.0004)
Lake level, change in 24 hours	0.30 (0.0040)	0.17 (0.0658)	0.31 (0.0042)	0.31 (0.0154)	0.20 (0.0311)	0.22 (<0.0001)

[a] Radar rainfall sum from multiple cells for the 24-hour periods 1 and 2 days before sampling. The most recent rainfall receives the most weight.

[b] The total rainfall, in inches, for the 24-hour period before sampling .

[c] The total rainfall, in inches, for the 24-hour period 2 days before sampling.

[d] Variable was used in the backup model when radar rainfall data were not available.
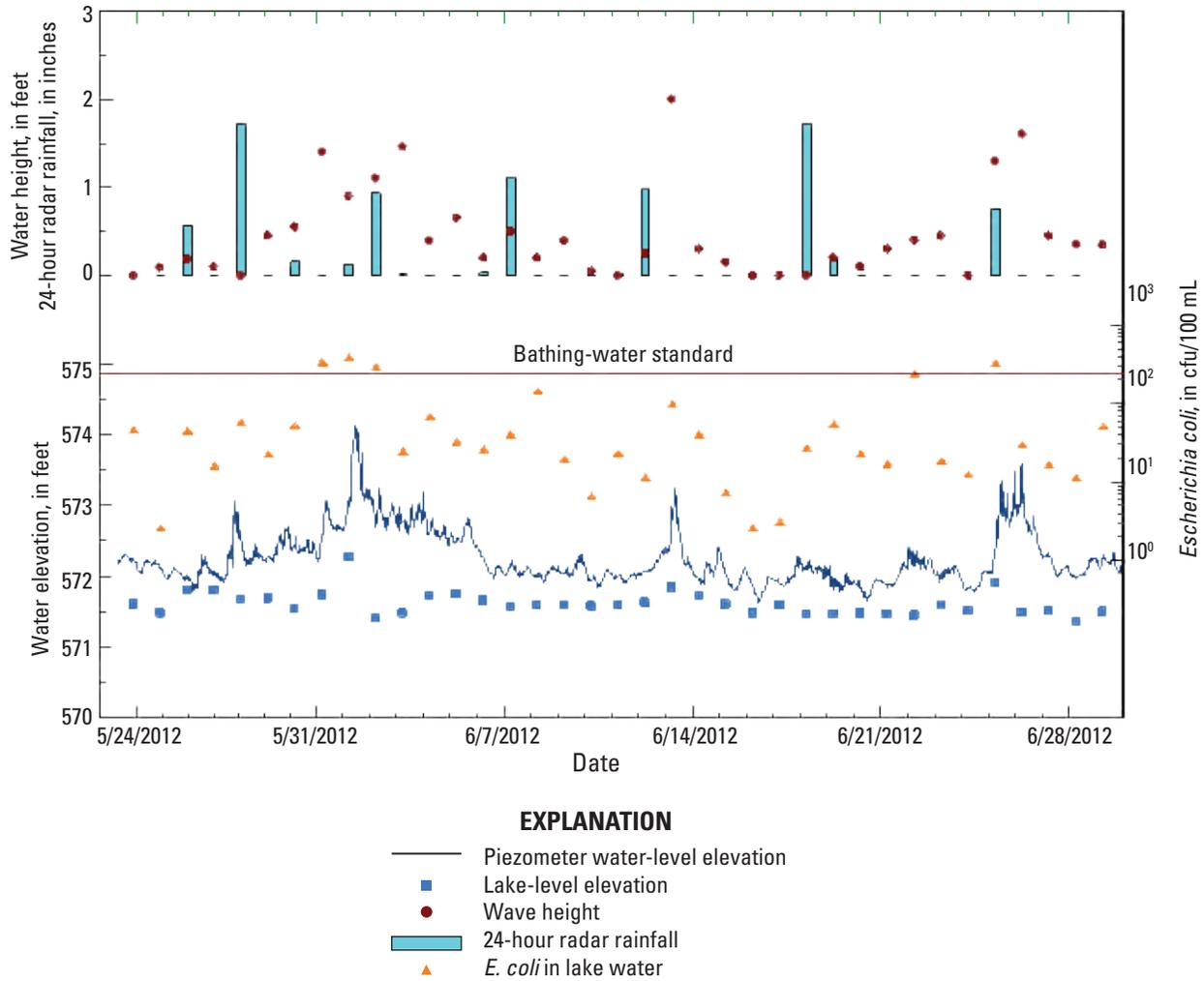
Figure 3.   Piezometer water level elevations ("foreshore head"), lake-level elevations, wave heights, radar rainfall, and concentrations of *Escherichia coli* at Edgewater, Cleveland, Ohio, May 23–June 30, 2012. (CFU/100 mL, colony-forming units per 100 milliliters)

Table 5.   Pearson's correlations between $\log_{10}$ *Escherichia coli* concentrations and foreshore head or to explanatory variables used in the predictive model at Edgewater, Cleveland, Ohio, 2008–12.

[The p values are in parenthesis; the highest correlations among foreshore head variables for each year are in bold]

Variable	2008	2009	2011	2012	All years
Number of samples	71	96	70	100	337
Foreshore head, average rate of change, 6 hours	0.14 (0.2392)	0.34 (0.0007)	0.33 (0.0053)	0.30 (0.0025)	0.27 (<0.0001)
Foreshore head, average rate of change, 12 hours	0.25 (0.0326)	**0.53 (<0.0001)**	**0.46 (<0.0001)**	0.22 (0.0298)	**0.35 (<0.0001)**
Foreshore head, average rate of change, 24 hours	**0.36 (0.0020)**	0.44 (<0.0001)	0.37 (0.0016)	0.26 (0.0077)	0.34 (<0.0001)
Foreshore head, average rate of change, 48 hours	0.25 (0.0391)	0.40 (<0.0001)	0.37 (0.0018)	**0.35 (0.0003)**	0.33 (<0.0001)
Sum, log turbidity and wave height	0.43 (0.0002)	0.51 (<0.0001)	0.55 (<0.0001)	0.36 (0.0002)	0.47 (<0.0001)
Radar rain, 48 hours weighted (Rw48)	0.31 (0.0086)	0.42 (<0.0001)	0.66 (<0.0001)	0.38 (0.0001)	0.41 (<0.0001)
Day of the year	0.11 (0.3737)	-0.06 (0.5641)	0.27 (0.0221)	0.27 (0.0064)	0.13 (0.0192)
Lake-level, change in 24 hours	0.21 (0.0813)	0.12 (0.2473)	0.28 (0.0180)	0.20 (0.0452)	0.20 (0.0002)

Wisconsin

Several new variables from the GLCFS were explored for potential utility as predictor variables in the *E. coli* models for Upper Lake Park, Port Washington, Wis. The current model used in 2012 at Upper Lake Park included 10 different variables with various transformations of those variables (table 6). New variables from GLCFS explored for use in the Upper Lake Park predictive model included time-based means of cloud cover and directional wave height (parallel and perpendicular components of a vector using wave height as the magnitude along with wave direction) and the standard deviation of air temperature and water temperature over several time spans (table 6). Time-averaged cloud cover was used as a surrogate for UV-light exposure and subsequent adverse impact of UV light on *E. coli* survival. The standard deviation of air and water temperature was used to explore potential impacts on *E. coli* concentrations resulting from fluctuations in temperature. For Upper Lake Park data, correlations of $\log_{10}$ *E. coli* concentrations to these new variables varied depending on the year and timespan. It did not appear that any of the new variables would result in improvement over those used in the current model; however, several of them proved to be valuable as predictors in other Wisconsin beach models (table 7).

Development of Predictive Models

Predictive models were developed by using data collected through 2011 for 43 beaches. Models were not developed for six beaches that were in "development" nowcast status (table 1). The same types of data were not available at all beaches, accounting for some differences in the selection of variables for the predictive models.

Among all variables, turbidity (measured with a turbidimeter) was used most often (79.1 percent of models) and in models from all six States (table 7). Wave heights (wave rod) were used in models from all States, and day of the year was used in models from all States except Wisconsin. Because field measurements and observations are easy to obtain and were were included in monitoring programs at all beaches, they were used in 95.3 percent of models, the highest percentage among general categories of variables. Lab measurements of quantitative polymerase chain reaction or chlorophyll *a* were made and used at only one beach each, or 2.3 percent of the models. Weather data from the NWS nearest airport site were used in 44.2 percent of models; these data were relied on in

models for several beaches in Michigan, New York, Ohio, and Pennsylvania. Radar rainfall data were used in 18.6 percent of the models and only at beaches in Ohio and Wisconsin. Radar data were available for only Ohio and Pennsylvania beaches by request to the USGS through a special computer program. Although radar data were also available through EnDDat, these data were accessed only for Wisconsin beaches. Data from nearby USGS gages were used in 32.6 percent of models. Among the measurements and predictions from NOAA, 24-hour lake-level change was used most often (34.9 percent). The variables based on NOAA predictions were used often in models for Wisconsin beaches, and local sources of data were used in Illinois/Indiana and Pennsylvania, owing to the availability of these data.

The equations for the model(s) selected for each beach are presented in appendix 5. Table 8 lists the variables, adjusted R^2 values, predicted *E. coli* concentration or threshold probability, number of observations, and response characteristics for the calibration datasets. The adjusted R^2 (coefficient of multiple determination) is a measure of the fraction of the variability in *E. coli* concentrations that can be explained by each model. The adjusted R^2 values ranged from 0.22 at Lake Front, Ohio, to 0.71 at Thompson, Wis., with a median adjusted R^2 of 0.43 (median data not listed in table 8). Beach managers in Illinois and Indiana used the predicted *E. coli* concentration as output from the models, along with a threshold of 235 CFU /100 mL. At all other beaches, the probability of exceeding the water-quality standard was computed, and a threshold probability was established from the calibration dataset. At Wisconsin beaches, a threshold probability of 50 percent was used for all beaches. Excluding Illinois, Indiana, and Wisconsin, threshold probabilities were established to maximize the correct percentage while, at the same time, obtaining a compromise between false positives and false negatives. The threshold probabilities established ranged from 15 percent at Metropark, Mich., to 45 percent at Villa Angela, Ohio, with a median of 28 percent. For all beaches, the percentages of correct responses ranged from 71.9 percent at Jeorse Park, Ind., to 97.3 percent at Neshotah, Wis., with a median of 86.7 percent. Sensitivities were 0 percent or very low for most of the beaches in Illinois and Indiana; at all of the other beaches except for Kreher, sensitivities were greater than or equal to 50 percent. Low specificities were found at Jeorse Park, Ind. (40 percent), and at Lakeview, Ohio (74.6 percent); otherwise specificities were greater than or equal to 82 percent.

Table 6.   Pearson's correlations between $\log_{10}$ *Escherichia coli* concentrations and new variables or explanatory variables used for the current predictive model at Upper Lake Park, Port Washington, Wisconsin.

[The p values are in parenthesis]

Category	Variable	2009	2010	2011	2012	All
	Variables used in current model					
Upper Lake Park, 2012 model variables	High algae (binary variable)	0.17 (0.27)	NA (NA)	NA (NA)	0.15 (0.4)	0.12 (0.1)
	Onshore current (quadratic)	0.15 (0.35)	0.12 (0.41)	0.23 (0.1)	0.09 (0.59)	0.08 (0.27)
	Gull count (quadratic)	-0.14 (0.39)	-0.04 (0.78)	-0.71 (<0.01)	-0.22 (0.2)	-0.35 (<0.01)
	Turbidity (1/3 power)	0.51 (<0.01)	0.56 (<0.01)	0.49 (<0.01)	0.64 (<0.01)	0.47 (<0.01)
	Tubidity x 48-hour radar rainfall	0.39 (0.02)	0.49 (<0.01)	0.43 (<0.01)	0.52 (<0.01)	0.25 (<0.01)
	24-hour radar rainfall (1/4 power)	0.45 (<0.01)	0.19 (0.19)	0.23 (0.11)	0.33 (0.05)	0.26 (<0.01)
	Water temperature (manual)	0.16 (0.31)	0.39 (0.01)	0.21 (0.14)	0.23 (0.18)	0.33 (<0.01)
	Wave height	0.5 (<0.01)	0.25 (0.08)	0.4 (<0.01)	0.29 (0.09)	0.38 (<0.01)
	Perpendicular wave height	0.32 (0.04)	0.23 (0.11)	0.19 (0.2)	0.3 (0.08)	0.28 (<0.01)
	Water temperature (square; Great Lakes Coastal Forecasting System.)	0.34 (0.03)	0.37 (0.01)	0.23 (0.12)	0.37 (0.03)	0.46 (<0.01)
	Potential new variables					
Cloud cover, mean	1-hour	0.44 (<0.01)	-0.04 (0.79)	0.21 (0.15)	0.24 (0.16)	0.15 (0.05)
	2-hour	0.46 (<0.01)	0.01 (0.95)	0.13 (0.36)	0.3 (0.08)	0.14 (0.06)
	6-hour	0.5 (<0.01)	0.06 (0.68)	0.08 (0.6)	0.42 (0.01)	0.16 (0.04)
	12-hour	0.41 (0.01)	0.08 (0.59)	0.01 (0.96)	0.42 (0.01)	0.13 (0.09)
	24-hour	0.33 (0.03)	0 (0.98)	0.1 (0.5)	0.37 (0.03)	0.1 (0.18)
Air temperature, standard deviation	6-hour	-0.23 (0.14)	0.28 (0.05)	0.1 (0.48)	0 (0.99)	0.01 (0.93)
	12-hour	-0.36 (0.02)	0.08 (0.58)	0.11 (0.44)	0.08 (0.65)	0.01 (0.85)
	24-hour	-0.34 (0.03)	-0.15 (0.29)	-0.17 (0.24)	-0.02 (0.91)	-0.13 (0.09)
	48-hour	-0.14 (0.39)	-0.31 (0.03)	-0.16 (0.28)	0.02 (0.92)	-0.12 (0.12)
	72-hour	-0.14 (0.36)	-0.31 (0.03)	0 (1)	0.33 (0.05)	-0.02 (0.77)
	120-hour	-0.2 (0.2)	-0.25 (0.08)	-0.03 (0.81)	0.41 (0.01)	0 (1)
Water temperature, standard deviation	6-hour	-0.16 (0.32)	0.14 (0.34)	-0.02 (0.91)	0.11 (0.52)	-0.01 (0.86)
	12-hour	-0.08 (0.6)	0.03 (0.84)	-0.21 (0.14)	-0.27 (0.11)	-0.15 (0.05)
	24-hour	-0.03 (0.83)	0.15 (0.3)	-0.26 (0.07)	-0.12 (0.47)	-0.06 (0.41)
	48-hour	-0.07 (0.64)	-0.09 (0.52)	-0.15 (0.32)	-0.1 (0.55)	-0.09 (0.22)
	72-hour	-0.01 (0.95)	-0.02 (0.91)	-0.06 (0.67)	-0.09 (0.62)	-0.06 (0.45)
	120-hour	0.12 (0.46)	0.15 (0.29)	-0.08 (0.59)	0.11 (0.52)	0.01 (0.88)
Directional wave height, mean	Parallel: 1-hour	0.04 (0.8)	0.33 (0.02)	-0.36 (0.01)	-0.23 (0.17)	-0.14 (0.07)
	2-hour	0.04 (0.78)	0.22 (0.13)	-0.35 (0.02)	-0.27 (0.12)	-0.16 (0.04)
	6-hour	0.05 (0.74)	0.19 (0.18)	-0.4 (<0.01)	-0.36 (0.03)	-0.24 (<0.01)
	12-hour	-0.12 (0.46)	-0.01 (0.93)	-0.4 (<0.01)	-0.36 (0.03)	-0.31 (<0.01)
	24-hour	-0.2 (0.2)	-0.09 (0.55)	-0.28 (0.05)	-0.4 (0.01)	-0.33 (<0.01)
	Perpendicular: 1-hour	-0.31 (0.05)	-0.25 (0.08)	-0.24 (0.1)	-0.3 (0.08)	-0.29 (<0.01)
	2-hour	-0.31 (0.04)	-0.17 (0.23)	-0.22 (0.13)	-0.27 (0.11)	-0.27 (<0.01)
	6-hour	-0.33 (0.03)	-0.24 (0.09)	-0.15 (0.3)	-0.33 (0.05)	-0.3 (<0.01)
	12-hour	-0.32 (0.04)	-0.26 (0.06)	-0.22 (0.13)	-0.23 (0.17)	-0.32 (<0.01)
	24-hour	-0.23 (0.14)	-0.2 (0.16)	-0.19 (0.2)	-0.25 (0.14)	-0.3 (<0.01)

Table 7. Summary of variables, by State, used in beach-specific predictive models.

[Rd-1, Rd-2, Rd-3: the total rainfall for the 24-hour period before sampling, 2 days before sampling, or 3 days before sampling; Rw48, Rw72: the amount of rainfall for the 48-hour or 72-hour period before sampling with the most recent rainfall receiving the most weight; h, hrs: hour(s); min: minimum; max: maximum; USGS, U.S. Geological Survey; NOAA, National Oceanic and Atmospheric Administration]

	Ill./ Ind.	Mich.	N.Y.	Ohio	Pa.	Wis.	Totals	Used in all models (percent)
Number of beaches	**9**	**2**	**5**	**10**	**6**	**11**	**43**	
Field measurements and observations								
Day of the year	7	1	3	3	2	0	16	37.2
Subseason category	0	0	0	2	0	0	2	4.7
Water temperature	3	0	0	0	0	1	4	9.3
Wave height (wave rod)	2	1	2	7	1	1	14	32.6
Turbidity, turbidimeter, lake or nearby tributary	1	2	5	9	6	11	34	79.1
Specific conductance, lake or nearby tributary	0	0	0	0	0	2	2	4.7
Debris category	0	0	1	0	0	0	1	2.3
Algae category	2	0	0	0	0	1	3	7.0
Bird count, time of sampling	2	0	1	0	0	1	4	9.3
People count, time of sampling	2	0	0	0	0	0	2	4.7
Weather categories	3	0	0	0	0	0	3	7.0
Total number of models that included field measurements and observations	**8**	**2**	**5**	**9**	**6**	**11**	**41**	**95.3**
Laboratory measurements								
Quantiative polymerase chain reaction	0	0	0	1	0	0	1	2.3
chlorophyll *a* (fluorometric sensor)	1	0	0	0	0	0	1	2.3
Total number of models that included lab measurements	**1**	**0**	**0**	**1**	**0**	**0**	**2**	**4.7**
Weather data from the National Weather Service nearest airport site								
Rainfall: Rd-1, Rd-2, and (or) Rd-3	0	1	4	5	2	0	12	27.9
Rainfall, weighted: Rw48 or Rw72	0	0	1	3	1	0	5	11.6
Rainfall: 30-day sum	0	1	0	0	0	0	1	2.3
Air temperature: 24 h mean	0	0	1	0	0	0	1	2.3
Wind direction and (or) speed, instantaneous	1	2	4	5	0	0	12	27.9
Wind direction and (or) speed, 24 h	0	2	2	0	1	0	5	11.6
Total number of models that included weather data from National Weather Service nearest airport site	**1**	**2**	**5**	**8**	**3**	**0**	**19**	**44.2**
Radar rainfall (sum from several 4-kilometer grids)								
Radar rainfall: 12 h, Rd-1, and (or) Rd-2	0	0	0	0	0	5	5	11.6
Radar rainfall, weighted: Rw48	0	0	0	2	0	0	2	4.7
Radar rainfall: 5 day sum	0	0	0	0	0	1	1	2.3
Total number of models that included radar rainfall	**0**	**0**	**0**	**2**	**0**	**6**	**8**	**18.6**
Measurements from a nearby USGS gage or buoy								
Rainfall: Rd-1 and (or) Rd-2	0	0	0	0	0	3	3	7.0
Rainfall: 5-day sum	0	0	0	0	0	1	1	2.3
Wave height or period, nearshore buoy	0	0	0	0	0	0	0	0.0
Turbidity, nearshore buoy	1	0	0	0	0	0	1	2.3
Water temperature, nearshore buoy	0	0	0	0	0	0	0	0.0
Stream discharge, time of sampling	0	0	0	4	0	0	4	9.3

Table 7. Summary of variables, by State, used in beach-specific predictive models.—Continued

[Rd-1, Rd-2, Rd-3: the total rainfall for the 24-hour period before sampling, 2 days before sampling, or 3 days before sampling; Rw48, Rw72: the amount of rainfall for the 48-hour or 72-hour period before sampling with the most recent rainfall receiving the most weight; h, hrs: hour(s); min: minimum; max: maximum; USGS, U.S. Geological Survey; NOAA, National Oceanic and Atmospheric Administration]

	Ill./ Ind.	Mich.	N.Y.	Ohio	Pa.	Wis.	Totals	Used in all models (percent)
Number of beaches	9	2	5	10	6	11	43	
Measurements from a nearby USGS gage or buoy—Continued								
Stream discharge, mean 24, 48, or 72 h	0	0	0	1	0	0	1	2.3
Stream discharge, min or max, 24–72 h	0	0	0	0	0	1	1	2.3
Stream discharge, change, 6 h	0	0	0	0	0	3	3	7.0
Stream turbidity, change in 24 h	0	0	1	0	0	0	1	2.3
Total number of models that included measurements from a nearby USGS gage or buoy	**1**	**0**	**1**	**5**	**0**	**7**	**14**	**32.6**
Measurements or predictions from NOAA								
Lake level, 24 h change	2	1	3	4	5	0	15	34.9
Lake level, change from spring average	0	0	0	1	0	0	1	2.3
Wind direction and speed, time of sampling	0	0	0	1	0	0	1	2.3
Current direction, 24 h prediction	0	0	1	0	0	0	1	2.3
Air temperature, mean, 1–120 h prediction	0	0	0	0	0	3	3	7.0
Air temperature, st. deviation, 1–144 h prediction	0	0	0	0	0	3	3	7.0
Cloud cover, mean, 1–48 h prediction	0	0	0	0	0	3	3	7.0
Depth averaged current vector, mean 1–120 h prediction	0	0	0	0	0	1	1	2.3
Lake height above sea level, mean 1 h prediction	0	0	0	0	0	2	2	4.7
Lake height above sea level, mean, max, min or change, 24 h prediction	0	0	0	0	0	5	5	11.6
Surface current vector, mean, 1–120 hrs prediction	0	0	0	0	0	6	6	14.0
Water temperature, mean, 1-24 h prediction	0	0	1	0	0	4	5	11.6
Water temperature, st. deviation, 1–120 h prediction	0	0	0	0	0	2	2	4.7
Wave direction, mean, 1–24 h prediction	0	0	0	0	0	3	3	7.0
Wave height, mean, 1–24 h prediction	0	0	2	0	0	5	7	16.3
Wind vector, mean, 1–24 h prediction	0	0	0	0	0	3	3	7.0
Total number of models that included measurements or predictions from NOAA	**2**	**1**	**3**	**5**	**5**	**10**	**26**	**60.5**
Local sources, including on-site weather stations								
Rainfall: 4 or 6 h	2	0	0	0	0	0	2	4.7
Rainfall: Rd-1, Rd-2, and (or) Rd-3	3	0	0	0	0	0	3	7.0
Solar radiation, 4–6 h	5	0	0	0	0	0	5	11.6
Wind speed and (or) direction	2	0	0	0	1	0	3	7.0
Offshore or alongshore wind component	4	0	0	0	0	0	4	9.3
Air temperature	2	0	0	0	0	0	2	4.7
Humidity	1	0	0	0	0	0	1	2.3
Barometric pressure	1	0	0	0	5	0	6	14.0
Total number of models that included local sources	**8**	**0**	**0**	**0**	**6**	**0**	**14**	**32.6**

Table 8. Variables and statistics for beach-specific predictive models in calibration datasets.

[R^2, fraction of the variation of *E. coli* concentrations that is explained by the model; threshold probability is based on meeting and exceeding the single-sample bathing standard for *E. coli*; sensitivity is the percentage of exceedances of the bathing-water standard that are correctly predicted by the model; specificity is the percentage of nonexceedances of the bathing-water standard that are correctly predicted by the model; model performance is based on the calibration dataset]

Beach short name	General variables	Adj. R^2	Predicted *E. coli* concentration or threshold probability[a]	Number (calibration dataset)					Percentage		
				Obser-vations	Correct +	Correct −	False +	False −	Correct	Sensitivity	Specificity
Illinois											
South Beach	Day of the year, water temperature, people count, bird count, weather categories, local rainfall	0.43	235	109	4	84	4	17	80.7	19.0	95.4
Greenwood Beach	Day of the year, wave-height rod, bird count, people count, algae, weather category, lake-level change, local rainfall, local air temperature	0.36	235	110	0	99	0	11	90.0	0	100
Lee Beach	Day of year, water temperature, turbidity, lake-level change, algae, weather category, local wind	0.40	235	112	3	102	0	7	93.7	30	100
Montrose	Day of the year, local solar radiation, local offshore wind component, local humidity	0.36	235	79	8	61	2	8	87.3	50.0	96.8
63rd Street	Day of the year, local rainfall, local solar radiation, local offshore wind component	0.38	235	76	3	64	3	6	88.2	33.3	95.5
Foster	Day of the year, local solar radiation, local offshore wind component, local air temperature	0.33	235	76	0	73	0	3	96.1	0	100
Calumet	Day of the year, local rainfall, local solar radiation, local offshore wind component	0.38	235	78	2	64	4	8	84.6	20.0	94.1
Oak	Turbidity from nearshore buoy, local rainfall, local solar radiation, local wind speed, local barometric pressure	0.28	235	67	0	64	0	3	95.5	0	100
Indiana											
Jeorse Park	Field water temperature, wave-height rod, chlorophyll *a* (fluorometric sensor), airport wind speed	0.28	235	32	19	4	6	3	71.9	86.4	40.0

Table 8. Variables and statistics for beach-specific predictive models in calibration datasets.—Continued

[R², fraction of the variation of *E. coli* concentrations that is explained by the model; threshold probability is based on meeting and exceeding the single-sample bathing standard for *E. coli*; sensitivity is the percentage of exceedances of the bathing-water standard that are correctly predicted by the model; specificity is the percentage of nonexceedances of the bathing-water standard that are correctly predicted by the model; model performance is based on the calibration dataset]

Beach short name	General variables	Adj. R²	Predicted *E. coli* concentration or threshold probability[a]	Number (calibration dataset)					Percentage		
				Obser-vations	Correct +	Correct −	False +	False −	Correct	Sensitivity	Specificity
Michigan											
Metropark	Wave-height rod, turbidity, airport rainfall, airport wind, lake-level change	0.33	15	70	10	53	4	3	90.0	76.9	93.0
Memorial	Day of the year, turbidity, airport rainfall, airport wind	0.55	39	72	7	57	4	4	88.9	63.6	93.4
New York											
Ontario	Day of the year, wave-height rod, turbidity, airport wind, lake-level change, current direction prediction, wave-height prediction	0.67	27	153	45	84	15	9	84.3	83.3	84.8
Durand	Day of the year, turbidity, debris category, bird count, airport rainfall, airport wind, lake-level change, wave-height prediction	0.59	25	131	24	89	15	3	86.3	88.9	85.6
Woodlawn SP	Day of the year, wave-height rod, turbidity, airport rainfall, airport wind, stream turbidity change	0.39	40	156	34	89	19	14	78.8	70.8	82.4
Evans TP	Turbidity, airport rainfall, airport wind, lake-level change, water temperature prediction	0.58	20	95	12	70	11	2	86.3	85.7	86.4
Lake Erie	Turbidity, airport rainfall, airport wind, airport air temperature	0.42	27	94	10	70	6	8	85.1	55.6	92.1
Ohio											
Edgewater (radar available)	Day of the year, subseason, wave-height rod, turbidity, radar rainfall, lake-level change	0.41	30	681	99	448	67	67	80.3	59.6	87.0
Edgewater (no radar available)	Day of the year, subseason, wave-height rod, turbidity, airport rainfall, lake-level change	0.39	30	714	98	462	82	72	78.4	57.6	84.9
Huntington (radar available)	Subseason, wave-height rod, turbidity, airport rainfall, airport wind, radar rainfall	0.53	27	606	61	464	49	32	86.6	65.6	90.4
Huntington (no radar available)	Subseason, wave-height rod, turbidity, airport rainfall, airport wind	0.52	31	606	59	461	47	39	85.8	60.2	90.7

Table 8. Variables and statistics for beach-specific predictive models in calibration datasets.—Continued

[R^2, fraction of the variation of *E. coli* concentrations that is explained by the model; threshold probability is based on meeting and exceeding the single-sample bathing standard for *E. coli*; sensitivity is the percentage of exceedances of the bathing-water standard that are correctly predicted by the model; specificity is the percentage of nonexceedances of the bathing-water standard that are correctly predicted by the model; model performance is based on the calibration dataset]

Beach short name	General variables	Adj. R^2	Predicted *E. coli* concentration or threshold probability[a]	Number (calibration dataset)					Percentage		
				Obser-vations	Correct +	Correct −	False +	False −	Correct	Sensitivity	Specificity
Ohio—Continued											
Maumee Bay SP (positive flow in ditch)	Airport rainfall, airport wind	0.43	20	70	6	58	5	1	91.4	85.7	92.1
Maumee Bay SP (negative flow in ditch)	Turbidity, airport wind, stream discharge, lake-level change	0.45	25	29	7	18	2	2	86.2	77.8	90.0
Villa Angela	Quantiative polymerase chain reaction (qPCR) for *E. coli*	0.47	45	257	92	119	26	20	82.1	82.1	82.1
Vermillion West	Wave-height rod, turbidity, airport rainfall, stream discharge	0.63	39	87	19	56	5	7	86.2	73.1	91.8
Nickel Plate	Day of the year, wave-height rod, turbid-ity, wind speed and direction, stream discharge	0.26	29	87	7	63	10	7	80.5	50.0	86.3
Lake Front	Wave-height rod, turbidity, lake-level change	0.22	37	86	15	56	5	10	82.6	60.0	91.8
Lakeview	Day of the year, turbidity, airport rainfall, airport wind, lake-level change	0.43	41	144	64	53	18	9	81.3	87.7	74.6
Fairport Harbor	Wave-height rod, turbidity, airport rainfall, airport wind, stream discharge	0.38	28	204	16	158	16	14	85.3	53.3	90.8
Mentor Headlands	Wave-height rod, turbidity, airport rainfall, airport wind, stream discharge	0.47	22	204	22	154	16	12	86.3	64.7	90.6
Pennsylvania											
Beach 1 West Extension	Turbidity, lake-level change, local wind	0.52	29	129	13	96	10	10	84.5	56.5	90.6
Beach 2	Turbidity, lake-level change, local baromet-ric pressure	0.51	28	130	6	112	10	2	90.8	75.0	91.8
Beach 6	Turbidity, lake-level change, local baromet-ric pressure	0.36	21	130	7	106	13	4	86.9	63.6	89.1
Mill Road	Wave height, turbidity, airport rainfall, lake-level change, local barometric pressure	0.39	26	127	5	106	11	5	87.4	50.0	90.6

Table 8.   Variables and statistics for beach-specific predictive models in calibration datasets.—Continued

[R^2, fraction of the variation of *E. coli* concentrations that is explained by the model; threshold probability is based on meeting and exceeding the single-sample bathing standard for *E. coli*; sensitivity is the percentage of exceedances of the bathing-water standard that are correctly predicted by the model; specificity is the percentage of nonexceedances of the bathing-water standard that are correctly predicted by the model; model performance is based on the calibration dataset]

Beach short name	General variables	Adj. R^2	Predicted *E. coli* concentration or threshold probability[a]	Number (calibration dataset)					Percentage		
				Obser-vations	Correct +	Correct −	False +	False −	Correct	Sensitivity	Specificity
				Pennsylvania—Continued							
Beach 9	Day of the year, turbidity, airport rainfall, airport wind, lake-level change, local barometric pressure	0.39	29	136	14	104	11	7	86.8	66.7	90.4
Beach 11	Day of the year, turbidity, airport rainfall, local barometric pressure	0.36	18	131	6	114	9	2	91.6	75.0	92.7
				Wisconsin							
Upper Lake Park	Turbidity, algae category, bird count, wave height, radar rainfall, surface-current vec-tor, water temperature, and wind-vector predictions	0.49	30	104	2	99	1	2	97.1	50.0	99.0
Red Arrow	Turbidity, cloud cover, water temperature, and wave-height predictions	0.31	50	76	16	43	7	10	77.6	61.5	86.0
Neshotah	Turbidity, air temperature, cloud cover, lake height above sea level, surface-current vector, water temperature, and wave-height predictions	0.57	50	73	4	67	1	1	97.3	80.0	98.5
Point Beach South	Turbidity, radar rainfall, stream discharge, lake height above sea level, surface-current vector, water temperature, wave direction, and wave-height predictions	0.57	50	225	19	190	3	13	92.9	59.4	98.5
Point Beach Central	Same variables as Point Beach South	0.57	50	225	19	190	3	13	92.9	59.4	98.5
Point Beach North	Same variables as Point Beach South	0.57	50	225	19	190	3	13	92.9	59.4	98.5
Fischer	Turbidity, gage rainfall, depth-averaged-current vector and surface-current vector predictions	0.51	50	79	8	61	3	7	87.3	53.3	95.3
Hika	Turbidity, specific conductance, gage rainfall (PLS model)	0.29	50	85	14	52	9	10	77.6	58.3	85.2

Table 8. Variables and statistics for beach-specific predictive models in calibration datasets.—Continued

[R^2, fraction of the variation of *E. coli* concentrations that is explained by the model; threshold probability is based on meeting and exceeding the single-sample bathing standard for *E. coli*; sensitivity is the percentage of exceedances of the bathing-water standard that are correctly predicted by the model; specificity is the percentage of nonexceedances of the bathing-water standard that are correctly predicted by the model; model performance is based on the calibration dataset]

Beach short name	General variables	Adj. R^2	Predicted *E. coli* concentration or threshold probability[a]	Number (calibration dataset)					Percentage		
				Obser-vations	Correct +	Correct −	False +	False −	Correct	Sensitivity	Specificity
Wisconsin—Continued											
Thompson	Turbidity, water temperature, gage rainfall, radar rainfall, air temperature, cloud cover, and wind-vector predictions	0.71	50	77	10	64	0	3	96.1	76.9	100
Kreher	Turbidity, stream discharge, gage rainfall, air temperature, and lake height above sea level predictions	0.49	50	82	6	67	0	9	89.0	40.0	100
Maslowski	Turbidity, tributary specific conductance, radar rainfall, air temperature, lake height above sea level, and wind-vector predictions	0.43	50	74	4	62	4	4	89.2	50.0	93.9

[a] The predicted E. coli concentration is 235 colony-forming units per 100 milliliters; the other values are threshold probabilities, in percent.

Validation of Predictive Models and Implementation of Nowcast Systems

Models for 42 beaches were validated during an independent year (table 9). At Jeorse Park, the predictive model was not validated and the modeling effort was discontinued in 2011, owing to poor model performance and various other reasons. Models that performed at least 5 percent better than the persistence model in yielding correct responses, sensitivities, or specificities were considered promising models; these responses are highlighted in bold on table 9. Out of 42 predictive models, 24 models (57 percent) yielded overall correct responses that were at least 5 percent greater than the use of the persistence model. For the purposes of this report, overall correct responses greater than or equal to 80 percent, sensitivities greater than or equal to 50 percent, and specificities greater than or equal to 85 percent were goals for good model performance.

The predictive-model responses were compared to use of the persistence model during validation in 2012 in terms of overall percent correct, sensitivities, and specificities (fig. 4). Scatterplots of the data as compared to the 1:1 lines show that predictive-model responses were often higher than persistence-model responses. Using the percentage goals set forth in the previous paragraph, the plots can be split into quadrants where (1) the predictive model met the goal but the persistence model did not, (2) both models met the goal, (3) neither model met the goal, and (4) the persistence model met the goal but the predictive model did not. For overall percent correct, the predictive model met the goal while the persistence model did not at 13 beaches (fig 4*A*, quadrant 1), whereas the opposite was true at only 2 beaches (fig 4*A*, quadrant 4). The persistence model met the goal for sensitivity only 4 times (fig. 4*B*, quadrants 2 and 4), whereas the predictive model met the goal 17 times (fig. 4*B*, quadrants 1 and 2). The specificity goal was met by both models 22 times (fig. 4*C*, quadrant 2), but the predictive model met the goal more often when the persistence model did not (fig. 4*C*, quadrants 1 and 3, respectively).

The differences between predictive- and persistence-model paired responses (that is, the predictive-model percentage minus the persistence-model percentage) for each beach model are shown in figure 5. In all cases, the median difference was above zero, indicating typically better performance of the predictive models. By use of a paired student's t-test, mean percentages of overall correct responses, sensitivities, and specificities were significantly higher for the predictive model than for the persistence model ($p = 0.0018$, 0.0006, and 0.0241, respectively).

For all models, data from 2012 will be added to previous years' datasets to refine the models for use in 2013. During the recreational season of 2013, plans are underway to continue operational nowcasts at 10 beaches, initiate new nowcasts at 3 beaches, and refine and retest models at 29 beaches (table 9). Validation results are discussed below for beaches by State.

Illinois—Evanston Beaches

Predictive models for South, Greenwood, and Lee Beaches in Evanston, Ill., yielded slightly higher percentages of correct predictions than the persistence models, ranging from 77.3 to 96.9 percent correct (table 9). Although overall correct predictions by the nowcast and the persistence models for South Beach were similar, the predictive model yielded a lower sensitivity and a higher specificity than the persistence model.

For both Greenwood Beach and Lee Beach, the predictive models yielded higher percentages correct (93.8 and 96.9 percent, respectively) but resulted in sensitivities of 0 percent and specificities of 100 percent. The persistence models for these beaches also yielded sensitivities of 0 percent and slightly lower specificities than the predictive models. These results indicate that elevated *E. coli* concentrations are likely ephemeral at these beaches, accounting for low sensitivities of the persistence models and the possibility that some factor or factors are not being accounted for in the predictive models.

The 2012 validation period used for the predictive models represented a period with less rainfall and lower stream discharge and direct runoff into Lake Michigan than normal. Therefore, 2012 validation datasets might not adequately represent the more normal conditions during the calibration years (2010–11). Refinement of the predictive models for Evanston, Ill., beaches will continue, and refined models will be validated during 2013.

Illinois—Chicago Beaches

Predictive models for four Chicago beaches (Montrose, 63rd Street, Foster, and Calumet) generally performed better than persistence models in terms of percentage of correct responses and specificity (table 9). In both of these categories, however, the predictive model for Oak Street Beach had a slightly lower response than the persistence model. All models for Chicago beaches had lower sensitivities than the persistence models, and none of the Chicago models reached the targeted sensitivity level of 50 percent. Only the model for Foster had ≥80 percent correct responses (86.1 percent), and all models met the goal for specificity (≥85 percent).

The high number of false negative responses from all Chicago predictive models in 2012 was of concern, but several factors should be considered. Three out of five Chicago models included rainfall as a variable (appendix 5), and the 2012 summer was very dry in comparison to the 2010–11 calibration years for predictive models. Lack of rain may have contributed to the lower number of exceedances, indicating the importance of rainfall in the success of these models. Beaches in Chicago are generally influenced by nonpoint sources of FIB (Whitman and Nevers, 2008) because no river flows into Lake Michigan in the Chicago area. Thus, presumptive sources of high *E. coli* concentrations are beach sands (Whitman and Nevers, 2003), shorebirds, and nearshore processes

such as current circulation, wave action, resuspension, and transport (Ge and others, 2010, 2012).

Interestingly, 63rd Street Beach had a higher percentage of *E. coli* exceedances in the validation year (2012) than in the calibration years (2010–11) (24 and 15.6 percent, respectively, table 3). Historically, high concentrations of *E. coli* have occurred at the 63rd Street Beach, but the annual average concentrations fluctuate, and several management approaches in recent years have likely helped to decrease the number of swimming advisories. Gull fecal droppings contribute to higher *E. coli* concentrations in the sand, which in turn becomes a nonpoint source of *E. coli* to nearshore beach water (Whitman and Nevers, 2003). Any bacterial contamination that enters the beach water tends to remain in the beach embayment; current and wave actions limit water exchange with the open lake, leading to frequent closures at this beach (Ge and others, 2010, 2012).

Of the five Chicago Beaches, Montrose had the highest number of *E. coli* exceedances in calibration and validation years (22.3 and 26.3 percent, respectively; table 3). The model in 2012 did not perform well, missing *E. coli* exceedances on 17 out of 21 occasions. The current model contains three meteorological variables measured at a neighboring onsite weather station and day of the year. More studies are needed to determine causes for *E. coli* exceedances at this beach.

Michigan

Predictive models were created for Metropark and Memorial Beaches, both on Lake St. Clair in Macomb County, Mich., by using data collected in 2010–11. There were too few exceedances during 2012 to adequately evaluate the Metropark model. The Memorial predictive model performed better than the persistence model in terms of sensitivity (table 9). There were a high number of false positives for Memorial, however, which lowered the percentage of correct responses.

Metropark Beach had only one closure in 2012 (1.5 percent of the sample events exceeding the daily standard of 300 MPN/100 mL or the 30-day geometric mean standard of 130 MPN/100 mL) compared to 19 closures (16 percent) in 2010 and 2011. Memorial Beach was closed only 3 times in 2012 (4.4 percent) compared to 23 times (20.7 percent) in 2010 and 2011. The 2012 swimming season was marked by very low precipitation, which likely contributed to fewer beach closures than in previous years.

Future plans for Metropark and Memorial Beaches are to refine and retest the models during the 2013 season, with plans to pursue operational nowcast systems at both beaches. The county health department is also looking at developing a predictive model for a third Lake St. Clair beach in New Baltimore.

New York

During the 2012 season, predictive models at Ontario and Durand beaches performed better than the persistence models, with percentages of correct responses of 77.0 and 83.5, respectively (table 9). Performance in the calibration years (table 8, 2010–11) was comparable to performance during the 2012 validation. Both models will be used for nowcasting in 2013.

Woodlawn SP has been a difficult model to define because of the potential input of *E. coli* loads from beach sands. Additionally, combined-sewer and stormwater overflows have contributed to fecal contamination in Rush Creek, a tributary that discharges north of the bathing area. Woodlawn beach has a gentle slope eastward into small dunes separating the beach from a low-lying wetland area that may have direct interactions with Rush Creek. Most likely, the bathing area is influenced by both groundwater interactions and direct bacterial loading from Rush Creek. The model validated during 2012 resulted in an increase in sensitivity as compared to the persistence model; however, the percentage correct and specificity for the predictive model were 4 to 5 percent lower than for the persistence model. Work will continue in 2013 to refine the Woodlawn model by considering variables for 24- and 48-hour changes to both lake turbidity and Rush Creek turbidity, as well as multiday changes in lake levels, wave directions, and wave heights.

The Evans TP and Lake Erie models performed well considering that there were only three exceedances at Evans TP and no exceedances at Lake Erie. A period of drought affected western New York, drying up the tributaries to both beaches. The model sensitivity at Evans TP was 100 percent and the overall correctness was 95.3 percent, but this level of performance may change during a season of nondrought conditions. Plans are to include the Evans TP model in a new nowcast and refine and retest the Lake Erie model during 2013.

Ohio

Predictive models for three of the four beaches with operational nowcasts in Ohio performed better than the persistence models during 2012 in all categories. Overall correct predictive-model responses were at least 5 percentage points higher than those for the persistence model at Huntington, Maumee Bay SP, and Villa Angela, but not at Edgewater, owing to a lower specificity at Edgewater (table 9). Edgewater has been a challenge to model because of the probable influence of *E. coli* in beach sands on lake water quality (Francy and others, 2006a). The beach has a gentle slope, and beach sands are often saturated. Change in lake level and subseason variables have improved nowcast predictions at Edgewater, but more work needs to be done to quantify the interaction of shallow groundwater and waves and their influence on *E. coli* concentrations in the lake. As more data become available, adding foreshore head as a variable may improve model

performance at Edgewater. The Ohio Nowcast for Huntington yielded correct predictions more than 83 percent of the time since 2006. The inclusion of a subseason variable and radar rainfall improved model performance from earlier years (Francy and others, 2009). Water quality at Maumee Bay SP is influenced by Berger Ditch to the east of the beach (Francy and others, 2005). After an unsuccessful test of the nowcast in 2011, it was determined that the inclusion of two models—one for negative flow and one for positive flow in Berger Ditch— resulted in better nowcast predictions during 2012. Because of the high percentage of exceedances of the bathing-water standard at Villa Angela (greater than 45 percent), models developed in past studies based on environmental variables did not perform any better than the persistence model (Francy and others, 2003). As a result, a model based on the Ct value from qPCR as the sole predictive variable was validated during 2012. Although the predictive model yielded 81.6 percent correct responses at Villa Angela, 40 percent of the responses were based on the ΔΔCt and were not based on the Ct model because of inhibition of the qPCR reaction (data not shown). Using the ΔΔCt exclusively would have yielded more correct responses (88.5 percent) than using the Ct model. Plans are to continue the nowcast for these four beaches in 2013.

The predictive models for beaches in Erie County (Vermillion West, Lake Front, and Nickel Plate) yielded mixed responses. Sensitivities were low at Vermillion West and Lake Front, and specificity was low at Nickel Plate. The R^2 values for the Lake Front and Nickel Plate models were also low (0.26 and 0.22; table 8). Before nowcast systems are implemented, more work will be needed to understand the processes that affect *E. coli* concentrations at Erie County beaches in order to identify variables to improve model performance. The predictive models for beaches in Lake County (Fairport Harbor and Mentor Headlands) yielded over 80 percent correct responses but had low sensitivities. Inclusion of site-specific weather data from the Mentor Headlands weather station may help to improve future model performance.

The predictive model for Lakeview in Lorain County did not perform well in 2012, most notably in terms of specificity. Lakeview is similar to Edgewater in that it is a gently sloping beach with frequently saturated sands, that it has breakwaters that restrict water circulation, and that it is frequented by large populations of waterfowl in the swimming area. At Lakeview during 2012, the bird count at the time of sampling was as high as 3,000, with counts of over 1,000 birds on 15 days. The median bird count for the season was 500. By comparison, bird counts at Edgewater were over 1,000 on only one day (2,100 birds), and the median was much lower (165). It may be that a model with good performance will be difficult to develop for Lakeview unless a way to account for the influence of birds can be identified.

Pennsylvania

Predictive models based on 2010–11 datasets and tested during validation in 2012 performed better than the persistence models for all six Pennsylvania beaches at Presque Isle State Park (table 9). Correct responses for the predictive models were well over 80 percent in all cases, whereas correct responses for the persistence models were above 80 percent at only three of the six beaches. Specificities for the predictive models were greater than 85 percent at all six beaches; sensitivities were higher than those for persistence models, but none were greater than 50 percent (ranged from 23.1 to 42.9 percent). Specificities for the persistence models were greater than 85 percent for all but one beach (Beach 1 West Extension) but the sensitivities were less than or equal to 22.2 percent and were 0 percent at two beaches (Beach 2 and Beach 9). Turbidity, barometric pressure and change in lake level were important variables in predicting when the *E. coli* standard would be exceeded (explanatory variables in at least five of the six predictive models). Refinement of the predictive models for Presque Isle State Park beaches will continue, and refined models will be validated during 2013.

Wisconsin

Environmental conditions may have influenced how well some of the models performed throughout the 2012 season in Wisconsin. During the beginning of the 2012 beach season in Wisconsin, there was a substantial rainfall event that generated large stream discharges in tributaries, followed by a long drought period. Neither of these conditions were represented in the calibration datasets from 2010–11 for either Manitowoc County or the Chequamegon Bay area of Ashland and Bayfield Counties. It is hoped that adding 2012 data to the calibration datasets and redeveloping models will improve the accuracy of future model predictions given that more data are available for model calibration and that the extra year of data will provide a broader suite of environmental conditions.

The Chequamegon Bay, Lake Superior models had variable results in prediction accuracy, even though they are less than 6 mi apart (table 9). The model used in 2012 at Thompson had a very high sensitivity and specificity, performing better than the persistence model. The predictive model for Kreher also performed better than the persistence model for sensitivity but was nearly the same in overall accuracy and specificity. The model for Maslowski yielded only 66 percent overall correct responses and a lower sensitivity than the persistence model. For use in operational decisions, Thompson and Kreher models appear to have promise, the Kreher model may benefit from recalibration with the 2012 data, and the Maslowski predictive model will need recalibration and improvements to be useful.

Given the substantial differences in model results at these three beaches and given their proximity, exploration of site information is warranted to help explain potential factors involved. A few likely sources appear to be contributors to these differences. The two largest tributaries near these beaches are closest to Maslowski. The distances from Maslowski to the mouths of Fish Creek and Whittlesey Creek are 0.94 mi and 1.5 mi, respectively. Longshore-current direction varies, so the influence of these creeks on water quality at Maslowski is variable. Adding to that variability, North Fish Creek is affected by seiche from the bay, so flow may be from North Fish Creek into the bay or from the bay back into North Fish Creek, depending on the exact time. Kreher is a mile to the east of Maslowski, resulting in extra traveltime and additional dilution (relative to Maslowski) before waters from these two tributaries reach Kreher. Thompson is 5 mi to the North of Whittlesey Creek. There is more direct influence from a much smaller stream, Thompson Creek, and a small artesian well (very clean water) that discharges adjacent to the beach.

Birds are also much more prevalent at Maslowski (average number of birds = 39) than at Kreher and Thompson (average number of birds = 6 and 8, respectively). Currents at the beach, flow direction in North Fish Creek, and the number of birds influencing the beach are all variables that could help improve predictions at Maslowski. All three of these variables can vary quickly, and discrete measurements four times per week are not always representative of the influence from tributaries, birds, and currents on beach water quality.

Prediction accuracy of models among the Lake Michigan beaches in Wisconsin were also quite variable. Upper Lake Park, Red Arrow, and the three Point Beach predictive-model results were better than the persistence model for all three primary performance measures (total correct, sensitivity, and specificity), and other predictive models were better than the persistence model in one or two of these performance measures. On the low end, predictive models for Hika and Fischer yielded 71.9 and 79.1 percent correct predictions, respectively. Both predictive models had the same (Fischer) or better (Hika) overall correct predictions as the persistence model. Both beaches had substantial *Cladophora* accumulation at times during the season that increased turbidity (one of the explanatory variables) substantially, even though *E. coli* concentrations did not consistently exceed the decision criteria. In general, Wisconsin predictive-model-development protocol dictated that ordinary least squares regression models from Virtual Beach be considered first to be consistent with other models throughout the Great Lakes. For Hika, ordinary least squares analyses did not result in an acceptable model, but performance was improved by using partial least squares analyses for development of this model. Red Arrow had the most exceedances of the Wisconsin beaches in 2012 (31 percent of sampled days), most due to influence from multiple potential contamination sources. At Red Arrow, the predictive model was a substantial improvement over the persistence model. The predictive model for Neshotah predicted the only exceedance of 2012 correctly. The Point Beach models were developed by using data from all three beaches. The only difference in explanatory variables between the three beaches is that local measurements of turbidity were used in each model. Results for the Point Beach predictive models were better than those from persistence models, but they still have potential for some improvement.

For Wisconsin predictive models in the 2013 season, plans are to explore the use of a two-tiered modeling approach with tier I as the "standard" (using field-measured data) model and tier II as the "automated" model for nonsampled days. The tier II model would use only data that do not require a visit to the beach.

Table 9.   Predictive-model responses as compared to use of previous day's *E. coli* concentration (persistence model) during validation in 2012.

[The persistence model is the previous day's *E. coli* concentration, the current method for assessing water quality; sensitivity is the percentage of exceedances of the bathing-water standard that are correctly predicted by the model; specificity is the percentage of nonexceedances of the bathing-water standard that are correctly predicted by the model; responses that were at least approximately 5 percent higher for the predictive model than for the persistence model are in bold]

Beach short name	Model used	Number					Percentage			Nowcast Status for 2013
		Observations	Correct +	Correct −	False +	False −	Correct	Sensitivity	Specificity	
					Illinois					
South Beach	Predictive	66	1	50	5	10	77.3	9.1	90.9	Refine and retest
	Persistence	68	3	49	8	8	76.5	27.3	86.0	
Greenwood Beach	Predictive	64	0	60	0	4	**93.8**	0	**100.0**	Refine and retest
	Persistence	65	0	53	4	4	87.7	0	93.4	
Lee Beach	Predictive	65	0	63	0	2	**96.9**	0	100.0	Refine and retest
	Persistence	68	0	62	3	3	91.2	0	95.4	
Montrose	Predictive	78	4	52	5	17	**71.8**	19.0	**91.2**	Continue nowcast
	Persistence	65	6	32	15	12	58.5	33.3	68.1	
63rd Street	Predictive	72	1	53	3	15	75.0	6.3	**94.6**	Continue nowcast
	Persistence	59	8	35	10	6	72.9	57.1	77.8	
Foster	Predictive	72	0	62	2	8	**86.1**	0	**96.9**	Continue nowcast
	Persistence	57	0	42	8	7	73.7	0	84	
Calumet	Predictive	77	2	58	6	11	77.9	15.4	**90.6**	Continue nowcast
	Persistence	65	2	47	10	6	75.4	25.0	82.5	
Oak Street	Predictive	70	0	67	3	0	95.7	0	95.7	Continue nowcast
	Persistence	54	0	54	0	0	100.0	0	100	
					Michigan					
Metropark	Predictive	61	0	60	0	1	98.4	0	100.0	Refine and retest
	Persistence	68	0	67	1	0	98.5	0	98.5	
Memorial	Predictive	64	1	47	14	2	75.0	**33.3**	77.0	Refine and retest
	Persistence	68	0	65	3	0	95.6	0	95.6	
					New York					
Ontario	Predictive	87	18	49	10	10	**77.0**	**64.3**	**83.1**	New nowcast
	Persistence	88	11	43	17	17	61.4	39.3	71.7	
Durand	Predictive	79	6	60	10	3	**83.5**	**66.7**	85.7	New nowcast
	Persistence	113	3	81	14	15	74.3	16.7	85.3	
Woodlawn SP	Predictive	72	2	51	11	8	73.6	**20**	82.3	Refine and retest
	Persistence	107	1	82	12	12	77.6	7.7	87.2	
Evans TP	Predictive	85	3	78	4	0	95.3	**100**	95.1	New nowcast
	Persistence	85	1	80	2	2	95.3	33.3	97.5	
Lake Erie	Predictive	86	0	84	2	0	97.7	0	97.7	Refine and retest
	Persistence	86	0	86	0	0	100.0	0	100.0	
					Ohio					
Edgewater	Predictive	115	9	85	15	6	81.7	**60.0**	85.0	Continue nowcast
	Persistence	111	5	89	8	9	84.7	35.7	91.8	
Huntington	Predictive	58	7	44	3	4	**87.9**	**63.6**	**93.6**	Continue nowcast
	Persistence	39	2	26	5	6	71.8	25.0	83.9	
Maumee Bay SP	Predictive	58	7	38	7	6	**77.6**	**53.8**	**84.4**	Continue nowcast
	Persistence	43	1	23	9	10	55.8	9.1	71.9	

Table 9.   Predictive-model responses as compared to use of previous day's *E. coli* concentration (persistence model) during validation in 2012.—Continued

[The persistence model is the previous day's *E. coli* concentration, the current method for assessing water quality; sensitivity is the percentage of exceedances of the bathing-water standard that are correctly predicted by the model; specificity is the percentage of nonexceedances of the bathing-water standard that are correctly predicted by the model; responses that were at least approximately 5 percent higher for the predictive model than for the persistence model are in bold]

| Beach short name | Model used | Number | | | | | Percentage | | | Nowcast |
		Observations	Correct +	Correct −	False +	False −	Correct	Sensitivity	Specificity	Status for 2013
Ohio—Continued										
Villa Angela	Predictive	87	24	47	1	15	**81.6**	**61.5**	**97.9**	Continue nowcast
	ΔΔ CT[a]	87	31	46	2	8	**88.5**	**79.5**	**95.8**	
	Persistence	69	14	24	13	18	55.1	43.8	64.9	
Vermillion West	Predictive	52	1	42	0	9	82.7	10.0	**100.0**	Refine and retest
	Persistence	30	4	21	3	2	83.3	**66.7**	87.5	
Nickel Plate	Predictive	52	6	23	20	3	55.8	66.7	53.5	Refine and retest
	Persistence	30	3	21	3	3	80.0	50.0	87.5	
Lake Front	Predictive	52	4	35	7	6	75.0	40.0	83.3	Refine and retest
	Persistence	30	2	21	4	3	76.7	40.0	84.0	
Fairport Harbor	Predictive	105	5	80	7	13	**81.0**	**27.8**	**92.0**	**Refine and retest**
	Persistance	104	3	71	15	15	71.2	16.7	82.6	
Mentor Headlands	Predictive	105	3	86	4	12	**84.8**	20.0	95.6	Refine and retest
	Persistance	104	4	78	11	11	78.8	26.7	87.6	
Lakeview	Predictive	95	39	19	28	9	**61.1**	**81.3**	40.4	Refine and retest
	Persistence	93	28	24	20	21	55.9	57.1	54.5	
Pennsylvania										
Beach 1 West Extension	Predictive	92	7	73	2	10	**87.0**	**41.2**	**97.3**	Refine and retest
	Persistence	89	3	58	14	14	68.5	17.6	80.6	
Beach 2	Predictive	92	4	79	2	7	**90.2**	**36.4**	**97.5**	Refine and retest
	Persistence	88	0	67	10	11	76.1	0.0	87.0	
Beach 6	Predictive	92	3	74	5	10	**83.7**	**23.1**	**93.7**	Refine and retest
	Persistance	88	2	66	10	10	77.3	16.7	86.8	
Mill Road	Predictive	90	3	77	3	7	**88.9**	**30.0**	**96.3**	Refine and retest
	Persistance	85	2	69	7	7	83.5	22.2	90.8	
Beach 9	Predictive	89	3	79	3	4	**92.1**	**42.9**	**96.3**	Refine and retest
	Persistence	84	0	70	7	7	83.3	0.0	90.9	
Beach 11	Predictive	90	3	76	5	6	**87.8**	**33.3**	**93.8**	Refine and retest
	Persistence	85	1	68	8	8	81.2	11.1	89.5	
Wisconsin										
Upper Lake Park	Predictive	36	6	26	1	3	**88.9**	**66.7**	**96.3**	Continue nowcast
	Persistence	36	2	16	7	7	50.0	22.2	69.6	
Red Arrow	Predictive	64	13	41	7	3	**84.4**	**81.3**	85.4	Refine and retest
	Persistence	64	7	40	8	9	73.4	43.8	83.3	
Neshotah	Predictive	54	1	50	3	0	94.4	**100.0**	94.3	Refine and retest
	Persistence	54	0	52	1	1	96.3	0.0	98.1	
Point Beach South	Predictive	58	4	45	4	5	**84.5**	**44.4**	**91.8**	Refine and retest
	Persistence	58	3	43	6	6	79.3	33.3	87.8	
Point Beach Central	Predictive	58	4	46	4	4	**86.2**	**50.0**	92.0	Refine and retest
	Persistence	58	2	44	6	6	79.3	25.0	88.0	

Table 9.   Predictive-model responses as compared to use of previous day's *E. coli* concentration (persistence model) during validation in 2012.—Continued

[The persistence model is the previous day's *E. coli* concentration, the current method for assessing water quality; sensitivity is the percentage of exceedances of the bathing-water standard that are correctly predicted by the model; specificity is the percentage of nonexceedances of the bathing-water standard that are correctly predicted by the model; responses that were at least approximately 5 percent higher for the predictive model than for the persistence model are in bold]

| Beach short name | Model used | Number | | | | | Percentage | | | Nowcast |
		Observations	Correct +	Correct −	False +	False −	Correct	Sensitivity	Specificity	Status for 2013
					Wisconsin—Continued					
Point Beach North	Predictive	58	4	46	5	3	86.2	**57.1**	90.2	Refine and retest
	Persistence	58	2	46	5	5	82.8	28.6	90.2	
Fischer	Predictive	53	10	28	13	2	71.7	**83.3**	68.3	Refine and retest
	Persistence	53	5	33	8	7	71.7	41.7	80.5	
Hika	Predictive	43	2	32	3	6	**79.1**	25.0	**91.4**	Refine and retest
	Persistence	43	2	29	6	6	72.1	25.0	82.9	
Thompson	Predictive	51	3	47	1	0	**98.0**	**100.0**	97.9	Refine and retest
	Persistence	51	1	46	2	2	92.2	33.3	95.8	
Kreher Park	Predictive	47	3	37	5	2	85.1	**60.0**	88.1	Refine and retest
	Persistence	47	1	38	4	4	83.0	20.0	90.5	
Maslowski	Predictive	50	3	30	4	13	66.0	18.8	**88.2**	Refine and retest
	Persistence	50	7	26	8	9	66.0	43.8	76.5	

[a] A delta delta Ct calculation, described in Haugland and others (2005) was used instead of the predictive model.

Figure 4.   Model responses as compared to using previous day's *Escherichia coli* concentration (persistence model) during validation in 2012: *A*, Overall percent correct. *B*, Sensitivities. *C*, Specificities. (Each plot was split into quadrants where (1) the predictive model met the goal but the persistence model did not, (2) both models met the goal, (3) neither model met the goal, and (4) the persistence model met the goal but the predictive model did not. Sensitivity is the percentage of exceedances of the bathing-water standard that are correctly predicted by the model. Specificity is the percentage of nonexceedances of the bathing-water standard that are correctly predicted by the model.)
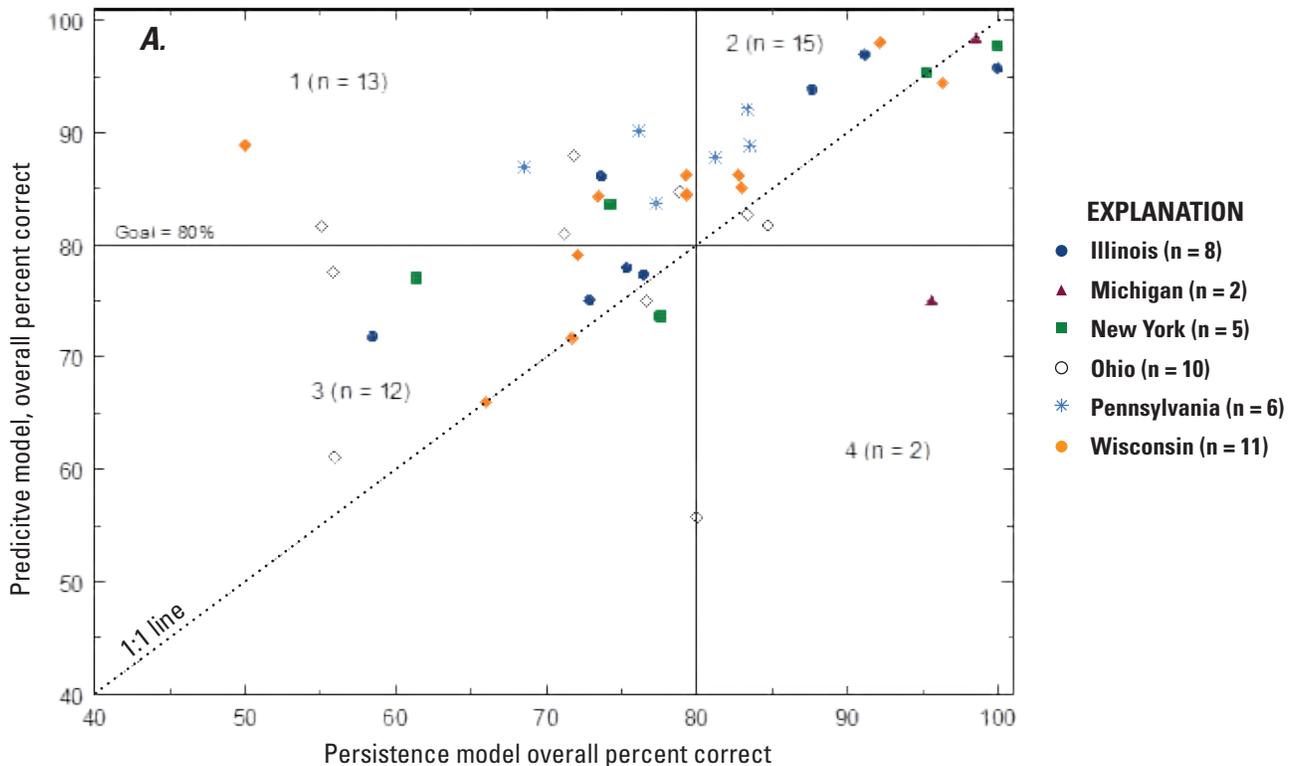
Figure 4.—Continued   Model responses as compared to using previous day's *Escherichia coli* concentration (persistence model) during validation in 2012: *A*, Overall percent correct. *B*, Sensitivities. *C*, Specificities. (Each plot was split into quadrants where (1) the predictive model met the goal but the persistence model did not, (2) both models met the goal, (3) neither model met the goal, and (4) the persistence model met the goal but the predictive model did not. Sensitivity is the percentage of exceedances of the bathing-water standard that are correctly predicted by the model. Specificity is the percentage of nonexceedances of the bathing-water standard that are correctly predicted by the model.)
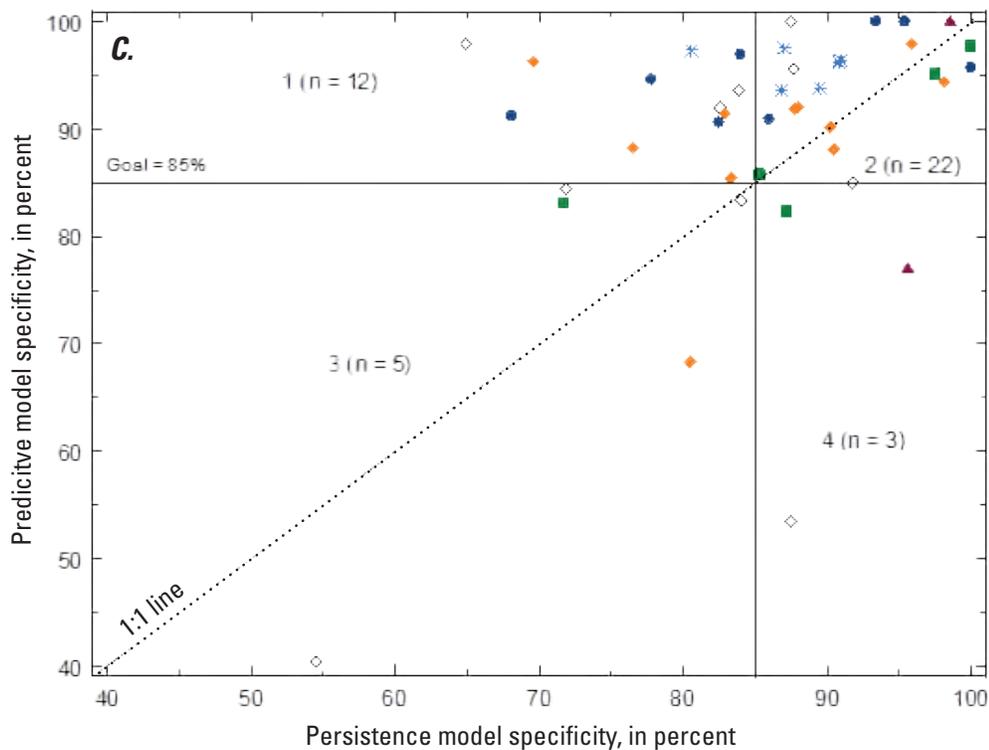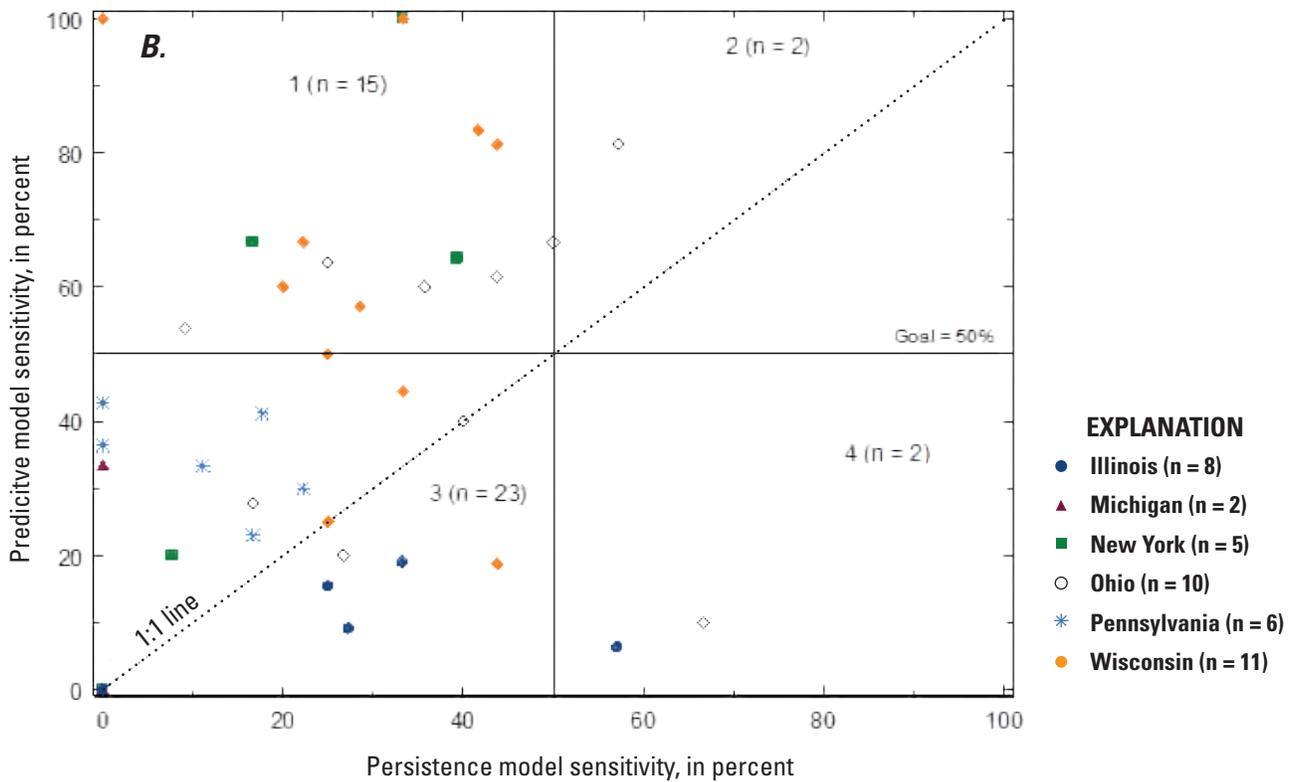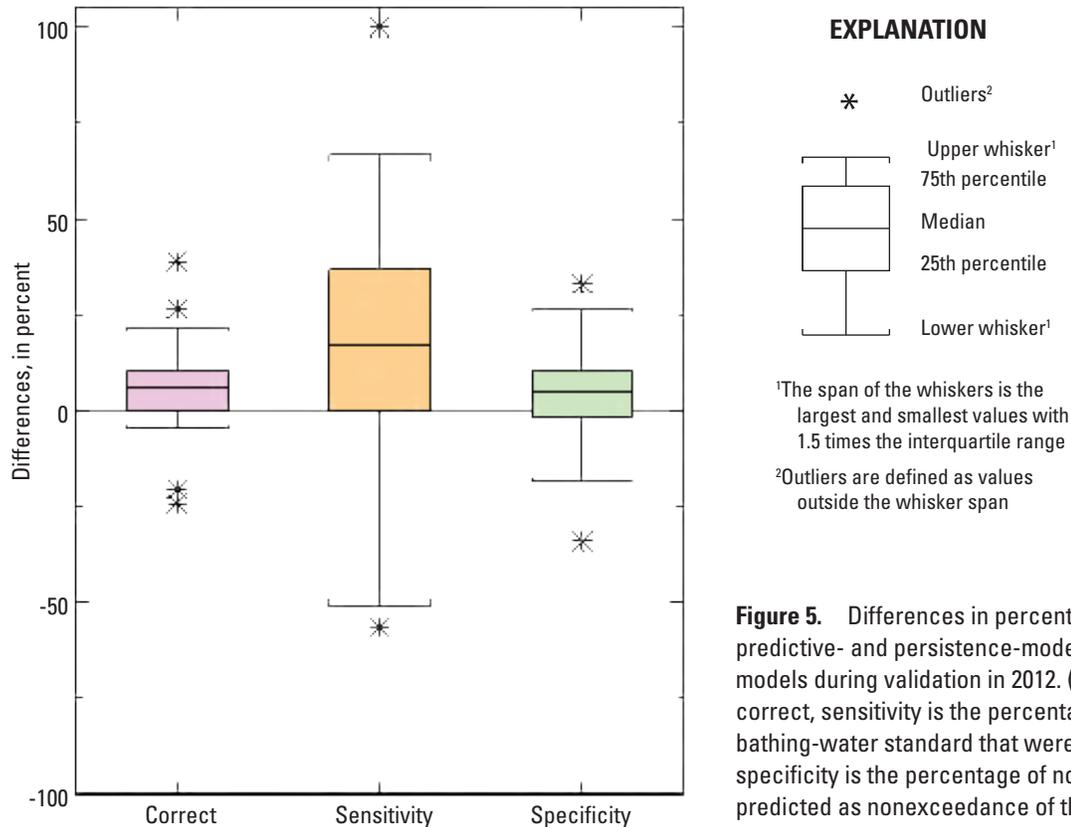
EXPLANATION

*   Outliers[2]

Upper whisker[1]
75th percentile

Median

25th percentile

Lower whisker[1]

[1]The span of the whiskers is the largest and smallest values with 1.5 times the interquartile range

[2]Outliers are defined as values outside the whisker span

Figure 5. Differences in percentages between paired predictive- and persistence-model responses for 42 beach models during validation in 2012. (Correct is the overall percent correct, sensitivity is the percentage of exceedances of the bathing-water standard that were correctly predicted, and specificity is the percentage of nonexceedances correctly predicted as nonexceedance of the standard.)

Future Work

The results of this study indicate that predictive models can predict exceedances of *E. coli* standards with accuracies as good as or better than the persistence model at most, but not all beaches. Drier conditions during 2012 as compared to 2010–11 (most calibration periods) may explain the poor model performance in 2012 at Illinois, Michigan, New York Lake Erie, and Wisconsin beaches, suggesting possible differences in factors affecting bacteria concentrations in dry and wet years. Additional years of data with a wide range of environmental conditions may help to improve future model performance.

A few beaches were described as gently sloping, with reservoirs of *E. coli* (presumably from bird populations) in the sand and shallow groundwater acting as a potential source of contamination to the lake; in particular, 63rd Street (Chicago, Ill.), Woodlawn SP (N.Y.), and Edgewater and Lakeview (Ohio). Inclusion of a variable that quantifies the interaction of shallow groundwater with lake water, such as foreshore head, may help to improve model performance at this type of beach.

Most importantly, many of the predictive models that met goals for percentages of correct responses were those where data had been collected for a number of years and the beach managers and modelers had gained an understanding of the factors affecting *E. coli* concentrations. For example,

at Ontario and Durand, N.Y., Monroe County has been using decision-tree models for beach closure decisions since 1976. The experience gained in applying logic models to beach management decisions was the backbone for developing linear regression models in the current study that performed well. The knowledge gained in developing and testing models over several years at Ohio Nowcast beaches resulted in more accurate predictions during 2012 than at those Ohio beaches where only 2 years of data were collected. At Ohio Nowcast beaches, model performance was improved by the addition of radar rainfall data and subseason variables at Huntington, development of two models at Maumee Bay SP (one for negative flow and one for positive flow in Berger Ditch), and development of a model based on qPCR at Villa Angela. Knowing the specific characteristics of each beach is the key to developing models that perform well in assessing water quality.

Predictive models are intended to augment existing beach monitoring programs, not to replace them. Models must be continuously tested and refined, especially if changes are made to improve water quality at a particular beach. The enhancement of Virtual Beach software has made it possible for beach managers with limited statistical backgrounds to more easily take over developing and running predictive models from those with more advanced statistical backgrounds. With budget uncertainties for beach monitoring activities in the future, especially in light of potential reduced funding under

the Beaches Environmental Assessment and Coastal Health (BEACH) Act (U.S. Congress, 2000), the future of predictive models is also uncertain. At beaches with many years of model data (at least 5 years) and (or) with strong understandings of the factors that affect *E. coli* concentrations, it may be possible to reduce sampling to once or twice a week (using field measurement data) and use another predictive model (using automated data) for beach advisory decisions on nonsampling days. However, at beaches with only a few years of data and (or) where there is a weak understanding of the factors affecting *E. coli* concentrations, it would be more difficult to maintain modeling activities with infrequent monitoring. Initiating a two-tiered modeling approach, as described previously for Wisconsin beaches, may be one solution to reduced funding. In the two-tiered system, tier I would be a standard model (using field-measured data) and tier II would be an automated model for nonsampled days.

In future years, the USGS plans to continue working with local partners to develop and validate predictive models at beaches and improve existing nowcasts, restructuring monitoring activities as budget issues arise.

Summary

Predictive models have been used to overcome the time-lag issue that limits the ability of 18–24 hour bacterial culture results to accurately assess current recreational water-quality conditions at beaches. Predictive models use environmental and water-quality variables to compute the probability that the State standard will be exceeded or to estimate concentrations of fecal-indicator bacteria (FIB), such as *Escherichia coli* (*E. coli*). They are recommended for use by the U.S. Environmental Protection Agency (EPA) (2012a) to provide same-day public health decisions. When predictive models are used for beach closure or advisory decisions, they are referred to as "nowcasts." In 2010–11, operational nowcasts (those used daily by beach managers to inform the public) were used at a few locations within the Great Lakes. Prior to 2010 and the current study, the U.S. Geological Survey (USGS) was involved in developing and testing predictive models in Ohio, Wisconsin, and Indiana. These models were able to provide more accurate water-quality information than estimates based on the previous day's *E. coli* concentration (persistence model).

Building on the success of nowcast systems, the USGS worked with 23 local and State agencies during the recreational seasons of 2010–12 to improve existing operational nowcast systems and to expand the use of operational predictive models throughout the Great Lakes. This report summarizes existing information on past and current USGS modeling efforts, compiles background information on the study areas, and presents efforts to collect data and develop predictive models by multiple agencies. Studies were done at 49 Great Lakes beaches in Illinois, Indiana, Michigan, New York, Ohio,

Pennsylvania, and Wisconsin. At the start of this overall study, the beaches consisted of four with operational nowcasts, six with recently implemented nowcasts, 33 in model validation, and six in model development. By using data collected during the recreational seasons of 2010–11 for most beaches (with earlier years' data included for some beaches), predictive models were refined or developed for 43 beaches, and 42 of these models were validated during an independent year (2012). Data collection continued during 2012 at the remaining six sites to develop predictive models.

Data on *E. coli* concentrations were collected to conform to State water-quality standards. Swim advisories were issued if the concentration in a single sample or the geometric mean of multiple samples from the same beach on the most recent monitoring day (usually the previous day) exceeded 235 colony-forming units per 100 milliliters (CFU/100 mL) in Illinois, Indiana, New York, Ohio, and Pennsylvania or 300 CFU/100 mL in Michigan. Swim bans were posted in Illinois, Indiana, New York, Pennsylvania, and Wisconsin if the concentration was greater than 1,000 CFU/100 mL. The 235 CFU/100 mL and 300 CFU/100 mL standards were used as thresholds in model development.

Data on explanatory variables for model development and validation included field measurements and observations, data from nearshore buoys and other equipment installed specifically for this project, and data compiled from other sources. Field measurements included wave height, turbidity, water temperature, bird and people counts, and categorical measures of debris and algae. In nearby laboratories, samples from one beach in Ohio (Villa Angela) were analyzed by use of the quantitative polymerase chain reaction (qPCR) method for *E. coli*, and samples from several beaches in Chicago and one beach in Indiana were analyzed for relative fluorescence as a surrogate for chlorophyll *a*. Nearshore buoys were installed by the USGS or local agencies at beaches in Illinois, Pennsylvania, and Ohio to measure wave heights and wave periods, water temperature, pH, dissolved oxygen, conductivity, and (or) turbidity in near real time. A weather station was installed at one beach in Ohio (Mentor Headlands), three beaches in Chicago, and one beach in Pennsylvania (Presque Isle Beach 2), and rain gages were installed at several beaches in Wisconsin to obtain site-specific weather data. Equipment was also installed on nearby tributaries in New York and Wisconsin to more directly investigate their influence on beach water quality; these measurements in tributaries included water level (staff gages), turbidity, specific conductance, and water temperature. A device to measure photosynthetically active radiation and a piezometer to measure foreshore head were installed at selected Ohio beaches to determine whether these variables could improve the performance of operational predictive models.

Data from other sources were compiled manually by the USGS or by use of tools designed specifically to facilitate the compilation and processing of data for predictive-model development. A lake-level spreadsheet developed during this study can be used to compile hourly lake-level data retrieved

from the nearest offshore buoy; PROCESSNOAA compiles and processes hourly rainfall, wind direction and speed, and barometric pressure data collected at the nearest National Oceanic and Atmospheric Administration (NOAA) airport site; and Environmental Data Discovery and Transformation (EnDDaT) retrieves and processes data from multiple sources.

Models were developed by use of Virtual Beach, a software program designed by USEPA specifically for beach model development (U.S. Environmental Protection Agency, 2012b) or by use of other statistical software. Most models were developed by use of multiple linear regression techniques. As one alternative statistical technique, the USGS developed software to construct predictive models based on partial least squares regression (Brooks and others, 2013). The model outputs were the predicted *E. coli* concentration or the probability of exceeding the State standard. If using the probability as the output, threshold probabilities for advisories were established by examining model output sensitivities and specificities at different probability levels. Assuming that the actual *E. coli* concentration measured by the culture method determined the correct response, calibration and validation datasets were examined in terms of overall correct responses, sensitivities (percentage of exceedances of the bathing-water standard that were correctly predicted), and specificities (percentage of nonexceedances of the bathing-water standard that were correctly predicted).

Data on photosynthetically active radiation (PAR) at Huntington, Ohio, and foreshore head at Edgewater, Ohio, were collected for several years to determine whether they could be used to improve performance of models used in the Ohio Nowcast. The use of the sum of the PAR for the previous day was examined for both Huntington and Edgewater because of the proximity of these two beaches. There was a significant negative correlation between PAR and *E. coli* concentrations for each of the five years tested except for one year at each beach. At both beaches, the relations between *E. coli* and PAR were not as strong as those between *E. coli* and two other model variables, but were equal to or stronger than three weaker model variables. Whether or not PAR can improve model performance is still uncertain. Water level in a shallow ground water well (piezometer) was used as measurement of foreshore head at Edgewater. The rate of change in foreshore head (in feet per hour) over the antecedent 12-hour period up to the time of sampling showed the strongest relations to *E. coli* concentrations among the foreshore-head variables (6, 12, 24, and 48 hours). The relations between foreshore-head variables and *E. coli* concentrations were about as strong or stronger than most other model variables at Edgewater.

More than 50 different variables or groups of variables were used in 43 beach-specific models. The number of times each variable was used in models was summed by State, and the frequency of use for each variable and group of variables was calculated as a percentage of all models. The same types of data were not available at all beaches, accounting for some differences in the selection of variables for the predictive models. Among individual variables, turbidity measured with a turbidimeter was used most often (79.1 percent), followed by day of the year (37.2), lake-level change over 24 hours (34.9 percent), and wave height measured with a wave rod (32.6 percent). As a variable group, weather data from the nearest NOAA airport sites were used in 44.2 percent of models from all States except Wisconsin, and radar rainfall data were used less often (18.6 percent) and only in models from Ohio and Wisconsin. The variables based on NOAA predictions were used in 2.3–16.3 percent of models, all at beaches in Wisconsin and New York. Local sources of data were used in 32.6 percent of models, all at beaches in Illinois/Indiana and Pennsylvania.

Goals for good model performance were set for overall correct responses greater than or equal to 80 percent, sensitivities greater than or equal to 50 percent, and specificities greater than or equal to 85 percent. By using these goals, the model responses during validation (2012) were compared to results from the persistence model. For 42 beach models, predictive-model responses met the performance goals more often than the persistence-model responses in terms of overall correct (31 versus 18 models, respectively), sensitivity (18 versus 4 models), and specificity (34 versus 26 models). Results from the paired student's t-test showed that overall correct responses, sensitivities, and specificities were significantly higher for the predictive model than for the persistence model.

Although predictive models performed better than most persistence models, there is still room for improvement. Drier conditions during the 2012 validation year as compared to 2010–11 calibration years may have led to poor model performance at Illinois, Michigan, New York Lake Erie, and Wisconsin beaches. Measuring foreshore head may improve model performance at Edgewater, Ohio, and other gently sloping beaches with sand as reservoirs of *E. coli*. Most importantly, gaining knowledge of each beach and the factors that affect *E. coli* concentrations is important for developing good predictive models. Collection of additional years of data with a wide range of environmental conditions may also help to improve future model performance. In times of budget restrictions, it may be possible to reduce sampling to once or twice a week at beaches with good-performing models and use a predictive model for beach advisory decisions on nonsampling days. The USGS will continue to work with local agencies in 2013 and beyond to develop and validate predictive models at beaches and improve existing nowcasts, restructuring monitoring activities as budget issues arise.

References Cited

Boehm, A.B., Grant, S.B., Kim, J.H., Mowbray, S.L., McGee, C.D., Clark, C.D., Foley, D.M., and Wellman, D.E., 2002, Decadal and shorter period variability of surf zone water quality at Huntington Beach, California: Environmental Science & Technology, v. 36, no. 18, p. 3885–3892.

Brooks, W.R., Fienen, M.N., and Corsi, S.R., 2013, Partial least squares for efficient models of fecal indicator bacteria on Great Lakes beaches: Journal of Environmental Management, v. 114, p. 470–475.

Dorfman, Mark, and Rosselot, K.S., 2011, Testing the waters—A guide to water quality at vacation beaches: Natural Resources Defense Council, accessed December 2012 at *http://www.nrdc.org/water/oceans/ttw/ttw2011.pdf*.

Erie County Department of Health, 2007, Bacterial quality of waters near Presque Isle—A compilation of relevant publications, 1987–2007, accessed December 2012 at *http://ecdh.org/water-quality.php/Bacterial-Water-Quality/40/2144/446/1453*.

Erie County Department of Health, 2012, Environmental Health Services, Beach Sampling, Protocol for public notification, accessed December 2012 at *http://ecdh.org/water-quality.php/Beach-Sampling/40/2144/446/1420*.

Francy, D.S., 2009, Use of predictive models and rapid methods to nowcast bacteria levels at coastal beaches: Aquatic Ecosystem Health & Management, v. 12, no. 2, p. 177–182.

Francy, D.S., Gifford, A.M., and Darner, R.A., 2003, *Escherichia coli* at Ohio bathing beaches—Distribution, sources, wastewater indicators, and predictive modeling: U.S. Geological Survey Water-Resources Investigations Report 02–4285, 120 p. (also available at *http://oh.water.usgs.gov/reports/Abstracts/wrir02-4285.html*).

Francy, D.S., Struffolino, P., Brady, A.M.G., and Dwyer, D.F., 2005, A spatial, multivariable approach for identifying proximate sources of *Escherichia coli* to Maumee Bay, Lake Erie, Ohio: U.S. Geological Survey Open File Report 2005–1386, 20 p. (also available at *http://pubs.usgs.gov/of/2005/1386/*).

Francy, D.S., Bertke, E.E., Finnegan, D.P., Kephart, C.M., Sheets, R.A., Rhoades, J., and Stumpe, L., 2006a, Use of spatial sampling and microbial source-tracking tools for understanding fecal contamination at two Lake Erie beaches: U.S. Geological Survey Scientific Investigations Report 2006–5298, 29 p., available at *http://pubs.usgs.gov/sir/2006/5298/*.

Francy, D.S., Darner, R.A., and Bertke, E.E., 2006b, Models for predicting recreational water quality at Lake Erie beaches: U.S. Geological Survey Scientific Investigations Report 2006–5192, 13 p., available at *http://pubs.usgs.gov/sir/2006/5192/*.

Francy, D.S., and Darner, R.A., 2006, Procedures for developing models to predict exceedance of recreational water-quality standards at coastal beaches: U.S. Geological Survey Techniques and Methods book 6, chap. B5, 34 p., available at *http://pubs.usgs.gov/tm/2006/tm6b5/*.

Francy, D.S., Bertke, E.E., and Darner, R.A., 2009, Testing and refining the Ohio Nowcast at two Lake Erie beaches—2008: U.S. Geological Survey Open-File Report 2009–1066, 20 p., available at *http://pubs.usgs.gov/of/2009/1066/*.

Ge, Z., Nevers, M.B., Schwab, D.J., and Whitman, R.L., 2010, Coastal loading and transport of *Escherichia coli* at an embayed beach in Lake Michigan: Environmental Science & Technology, v. 44, p. 6731–6737.

Ge, Z., Whitman, R.L., Nevers, M.B., Phanikumar, M.S., and Byappanahalli, M.N., 2012, Evaluating the role of an embayed beach as a reservoir and a net source of fecal contamination: Limnology and Oceanography v. 57, no. 1, p. 362–381.

Haugland, R.A., Siefring, S.C., Wymer, L.J., Brenner, K.P., and Dufour, A.P., 2005, Comparison of Enterococcus measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis: Water Research, v. 39, p. 559–568.

Holtschlag, D.J., Shively, Dawn, Whitman, R.L., Haack, S.K., and Fogarty, L.R., 2008, Environmental factors and flow paths related to *Escherichia coli* concentrations at two beaches on Lake St. Clair, Michigan, 2002–2005: U.S. Geological Survey Scientific Investigations Report 2008–5028, 38 p.

Indiana Department of Environmental Management, 2012, BeachGuard, accessed December 2012 at *https://extranet.idem.in.gov/BeachGuard/Default.aspx*.

Lake County (Illinois) Health Department, 2013, SwimCast-Data, accessed June 2013 at *http://health.lakecountyil.gov/Population/LMU/Pages/SwimCast-Data.aspx*.

Michigan Department of Environmental Quality, 2012a, About beach monitoring in Michigan, accessed March 2012 at *http://www.michigan.gov/deq/0,1607,7-135-3313_3686_3730-11005--,00.html*.

Michigan Department of Environmental Quality, 2012b, About beach monitoring in Michigan, accessed March 2012 at *http://www.deq.state.mi.us/beach*.

Myers, D.N., Stoeckel, D.M., Bushon, R.N., Francy, D.S., and Brady, A.M.G., 2007, Fecal indicator bacteria: U.S. Geological Survey Techniques of Water Resources Investigations, book 9, chap A7, section 7.1 (version 2.0), available at *http://pubs.water.usgs.gov/twri9A/*.

National Oceanic and Atmospheric Administration, 2012a, National Weather Service, accessed December 2012 at *http://www.nws.noaa.gov/*.

National Oceanic and Atmospheric Administration, 2012b, National Climatic Data Center—Unedited local climatological data, accessed December 2012 at *http://cdo.ncdc.noaa.gov/ulcd/ULCD*.

National Oceanic and Atmospheric Administration, 2012c, National Weather Service—Advanced Hydrologic Prediction Service, accessed December 2012 at *http://water.weather.gov/precip/download.php*.

National Oceanic and Atmospheric Administration, 2012d, Center for Operational Oceanographic Products and Services—Tides and currents, accessed December 2012 at *http://tidesandcurrents.noaa.gov/*.

National Oceanic and Atmospheric Administration, 2012e, National Estuarine Research Reserve System—Centralized Data Management Office, accessed December 2012 at *http://cdmo.baruch.sc.edu/get/realTime.cfm*.

National Oceanic and Atmospheric Administration, 2012f, Great Lakes Environmental Research Laboratory—Great Lakes Coastal Forecasting System, accessed December 2012 at *http://www.glerl.noaa.gov/res/glcfs*.

Nevers, M.B., and Whitman, R.L., 2005, Nowcast modeling of *Escherichia coli* concentrations at multiple urban beaches of southern Lake Michigan: Water Research, v. 39, p. 5250–5260.

Nevers, M.B., and Whitman, R.L., 2011, Efficacy of monitoring and empirical predictive modeling at improving public health protection at Chicago beaches: Water Research, v. 45, p. 1659–1668.

New York State Department of Health, 2010, Part 6—Subpart 6-2 Bathing Beaches, accessed December 2012 at *http://www.health.ny.gov/regulations/nycrr/title_10/part_6/subpart_6-2.htm#operate*.

Noble, R.T., Blackwood, A.D., Griffith, J.F., McGee, C.D., and Weisberg, S.B., 2010, Comparison of rapid quantitative PCR-based and conventional culture-based methods for enumeration of *Enterococcus* spp. and *Escherichia coli* in recreational waters: Applied and Environmental Microbiology, v. 76, no. 22, p. 7437–7443.

Ohio Department of Health, 2012a, Beach monitoring, accessed December 2012 at *http://www.odh.ohio.gov/odh-programs/eh/bbeach/beachmon.aspx*.

Ohio Department of Health, 2012b, BeachGuard, accessed December 2012 at *http://publicapps.odh.ohio.gov/BeachGuardPublic/Default.aspx*.

Olyphant, G.A., and Whitman, R.L., 2004, Elements of a predictive model for determining beach closures on a real time basis—The case of 63rd Street Beach Chicago: Environmental Monitoring and Assessment v. 98, p. 175–190.

Terbush, K., and Husson, K., 2009, Woodlawn Beach State Park Beach sanitary survey report: Albany, N.Y., New York State Office of Parks, Recreation and Historic Preservation Environmental Management Bureau, 26 p.

U.S. Congress, 2000, Beaches Environmental Assessment and Coastal Health Act of 2000: U.S. Congress, 106th, Public Law 106–284.

U.S. Environmental Protection Agency, 1986, Ambient water quality criteria for bacteria—1986: Washington, D.C., EPA 440/5–84–002, 18 p.

U.S. Environmental Protection Agency, 2002, National beach guidance and required performance criteria for grants: Washington, D.C., EPA–823–B–02–004, 178 p.

U.S. Environmental Protection Agency, 2009, Method 1603— *Escherichia coli* in water by membrane filtration using modified membrane-thermotolerant *Escherichia coli* agar (modified mTEC): Washington, D.C., EPA 821–R–09–007, 36 p., accessed December 2012 at *http://www.epa.gov/waterscience/methods/method/biological/*.

U.S. Environmental Protection Agency, 2012a, Recreational water quality criteria: Washington, D.C., EPA–820–F–12–058, 63 p., accessed December 2012 at *http://water.epa.gov/scitech/swguidance/standards/criteria/health/recreation/index.cfm*.

U.S. Environmental Protection Agency, 2012b, Exposure assessment models—Virtual Beach: Center for Exposure Assessment Modeling, accessed May 2013 at *http://www2.epa.gov/exposure-assessment-models/virtual-beach-vb*.

U.S. Geological Survey, 2009, Collection, quality assurance, and presentation of precipitation data: Office of Surface Water Technical Memorandum 2006.01 (revised December 2009), accessed March 2013 at *http://water.usgs.gov/osw/pubs/memo.summaries.html*.

U.S. Geological Survey, 2010, Swimming Advisory Forecast Estimate, accessed June 2013 at *http://www.glsc.usgs.gov/projectSAFE.php*.

U.S. Geological Survey, 2012, National Water Information System—Web interface, USGS Water Data for the Nation, accessed December 2012 at *http://waterdata.usgs.gov/nwis/.*

U.S. Geological Survey, 2013, Environmental Data Discovery and Transformation, accessed June 2013 at *http://cida.usgs.gov/enddat/.*

Whitman, R.L., and Nevers, M.B., 2003, Foreshore sand as a source of *Escherichia coli* in nearshore water of a Lake Michigan beach: Applied and Environmental Microbiology, v. 69, p. 5555–5562.

Whitman, R. L. and Nevers, M. B., 2008, Summer *E. coli* patterns and responses along 23 Chicago beaches: Environmental Science & Technology, v. 42, p. 9217–9224.

Wilde, F.D., ed., variously dated, Field measurements: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A6, sec. 6.1, 6.3, and 6.7, available at *http://pubs.water.usgs.gov/twri9A6/.*

Wisconsin Department of Natural Resources, 2012a, Wisconsin Beaches, accessed December 2012 at *http://dnr.wi.gov/topic/beaches/.*

Wisconsin Department of Natural Resources, 2012b, Wisconsin Beach Health, accessed December 2012 at *http://www.wibeaches.us/apex/f?p=BEACH:HOME:447869626819978.*

Zimmerman, T.M., 2008, Modeling to predict *Escherichia coli* at Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania: U.S. Geological Survey Scientific Investigations Report 2008–5039, 13 p.

Appendixes

Appendixes 4 and 5 follow immediately in this document. Appendixes 1, 2, and 3 are separate documents, available for downloading from the indicated Web address.

Appendix 1.   Obtaining National Oceanic and Atmospheric Administration Lake-Level Data

Appendix 1 documents in spreadsheet form are available for downloading from the indicated Web address: *http://pubs.usgs.gov/sir/2013/5166.*

Appendix 2.   Graphical User Interface: PROCESSNOAA

Appendix 2 documents, including instructions, software, and source code are available for downloading from the indicated Web address: *http://pubs.usgs.gov/sir/2013/5166*.

Appendix 3.   Example Spreadsheets for Compiling Beach Water-Quality and Environmental Data

Appendix 3 documents in spreadsheet form are available for downloading from the indicated Web address: *http://pubs.usgs.gov/sir/2013/5166.*

Appendix 4.   Example Protocol for Collecting Data and Running a Predictive Model for Nowcasting at Great Lakes Beaches

This document is an example of a recommended protocol for collecting and recording data for nowcasting at Great Lakes Beaches. It is intended to be used as a starting point by agencies collecting data and running predictive models for nowcasting and can be revised to meet the needs of local monitoring programs. The data collected in this protocol are entered into a spreadsheet, an example of which is Appendix 3, available for dowloading at http://pubs.usgs.gov/sir/2013/5166. This protocol is part of a larger document that describes the development and implementation of predictive models for estimating recreational water quality at Great Lakes beaches.

Field protocol

Equipment

- Personal floatation device (PFD, mandatory) and chest waders (optional)

- Sample bottles—500-milliliter (mL) or 1-liter (L) sterile plastic bottle with lid (one for each sampling location), labeled with site and date (an extra bottle is needed when a replicate or field blank is collected)

- Clipboard with field forms

- Wave rod, thermometer, and a small cooler with ice

- Camera (optional)

Sampling Locations and Frequency

Water samples are collected Monday through Thursday by 9 a.m. at Beach1 and Beach 2. Water samples are collected from approximately 1 foot (ft) below the water surface at a total water depth of 2–3 ft (not including waves). It is important that samples be collected consistently at the same depth throughout the summer. Under rough conditions (waves greater than 2 ft), the employee is to wade out to a comfortable depth, collect a sample, and record the water depth (not including waves).

Measure and Record Field Observations

- Record the number of birds and swimmers on the field form.

- Record any unusual field observations, such as excessive debris, algae, etc.

Measure Wave Height

Go to the sampling location. Place a wave rod in the water and stand there for 1 minute. Note the minimum water level and maximum height of waves during that period—record on field form. Perform the following calculation to get the wave height for the model:

Maximum height – minimum height = wave height (in feet)

If you are unable to enter the water because of unsafe conditions caused by high waves, estimate the wave height. Mark "estimate" on the field form.

Collect Water Samples

Aseptically collect water samples using a presterilized bottle. To collect a hand-dipped sample, open a sterile bottle about 12 to 18 inches (in.) below the water's surface. Grasp the bottle near the base and allow the bottle to fill with the opening pointed slightly upward. Remove the bottle with the opening pointed upward from the water and tightly cap it, *allowing some headspace for proper mixing*. Do not stir up bottom sediments while collecting the water sample.

- Measure water temperature by placing the thermometer about 12 in. below the water surface at one sampling location at each beach.

- If the sample is collected in less than 2 ft of water because of rough conditions, indicate so on the field form.

- *Immediately place all samples out of the sun in coolers with ice*. The sample needs to be kept on ice and processed for bacterial indicators and turbidity within 6 hours of sample collection.

Turbidity Protocol for Hach 2100Q[1]

Instructions for the the Hach 2100Q turbidimeter are given as an example. Other turbidimeters may be used.

Calibrate the Instrument Using the StablCal® Standards

Turbidity is measured in the laboratory. These are formazin standards that come with the instrument. Calibrate the instrument the **first time you use** it each season. After that, calibrate the instrument only **if the verification standard is off**. Use the 20, 100, and 800 nepholemetric turbidity ratio units (NTRU) standards for calibration.

1. Clean the outside of each vial using silicone oil and a lint-free cloth.

2. Read the three standards (as described in the instruction manual) and record the results in the calibration book that is kept with the instrument.

Check the Verification Standard

The verification standard is the 10 NTRU formazin standard. **Run the verification standard each week you use the instrument**.

1. Clean the outside of the 10 NTRU vial using silicone oil and a lint-free cloth.

2. Press "Verify Cal" and insert the vial.

3. Press READ. The machine will indicate if the verification passed. If it did not pass, recalibrate the instrument as described above. Press DONE.

4. Record results in the calibration log.

Read the Sample

1. Because the samples are cold, you may have to set up aliquots in the sample vials and let them reach room temperature before reading turbidity; otherwise, condensation may interfere with your reading.

2. Shake the sample bottle.

3. Immediately after shaking, pour an aliquot from the sample into a clean glass test vial. Fill to the line. Place the cap on the glass test vial.

4. Shake the sample bottle again and pour an aliquot from the sample into a <u>second</u> glass test vial. Place the cap on the second vial.

5. Hold the vials by the black lids only. Wipe the outside of the glass vials with silicone oil and a lint-free cloth.

6. Invert the vials <u>gently</u> a few times and place in the turbidimeter. Orient the vial in the well so that the diamond aligns with the raised orientation mark on the front of the well.

7. Close the turbidimeter cover.

8. Take the measurement by pressing READ.

9. Repeat for a second aliquot of the same sample. The two measurements must meet the following criteria. If they don't, take a third measurement—and a fourth if necessary—until two consecutive measurements meet the criteria.

    • For values less than 10 NTRU, they must agree within 1.5 NTRU.

    • For values greater than 10 NTRU, they must agree within 15 percent.

10. Record both turbidity values on laboratory sheets (the two that are within 15 percent) and calculate and record an average value.

11. Clean the test vials immediately after use. The glass is easily scratched, so be careful in handling and washing. Wash the vials with laboratory-grade detergent and rinse three times with deionized water. Do not allow to air dry. Store the vials filled with deionized water with the caps on.

---

[1] Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Colilert Quanti-Tray Method

Instructions for the Colilert Quanti-Tray/2000 method (IDEXX Laboratories, Inc., Westbrook, Maine) for *Escherichia coli* (*E. coli)* are given below. Other U.S. Environmental Protection Agency (EPA)-approved methods may be used.

Process the Sample

1. **Prepare lab forms and label**. Fill out appropriate lab forms for each sample and check off the dilutions to prepare, if any. Label the bottom of the tray with site name, date, and dilution (if any).

2. **Prepare sample and reagent mixture**. Measure out 100 mL of sample in a sterile graduated cylinder and pour into a sterile 125-mL bottle. Add one packet of Colilert reagent to the sample. Gently shake to distribute reagent throughout the sample.

3. **Prepare dilution (if needed)**. The system will enumerate between 1 and 2,400 most probable number (MPN)/100 mL for the undiluted sample. If you suspect the water to have greater than 2,400 MPN/100 mL, make a 1/10 dilution (10 mL of the supernatant and 90 mL of sterile deioinzed water) and mix the diluted sample with Colilert reagent. This will enumerate between 10 and 24,000 MPN/100 mL.

4. **Load and incubate the sample**. Pour the reagent/sample mixture into the Quanti-Tray. Tap the small wells to release any air and allow foam to settle. Seal the Quanti-Tray by placing in the sealer

Incubate and Count Wells

Incubate for 24 hours (24 to 28 hours max) at 35 °C. Count large and small positive wells that fluoresce under a long-wave ultraviolet light and refer to the MPN table provided by the manufacturer. For the 1:10 dilution, multiply the result in the table by 10 to get MPN/100 mL.

Models

Below are instructions for running a predictive model, using the example spreadsheet presented in appendix 3. Variable names used in appendix 3 are in parentheses below.

The variables used in the models for nowcast validation this year are as follows:

Beaches 1 and 2

1. Log turbidity (Turb_NTU) plus the wave height (Waveht_ft)

2. Airport rainfall during the previous 24 hours (AirportRainRd-1_in)

3. Lake level change in the past 24 hours (LakeLvl24chg_ft)

The thresholds for issuing an advisory are a 39 percent probability for Beach 1 and 235 MPN/100 mL for Beach 2.

Pull and Enter Data Into the Spreadsheets

1. **Enter field measurements and sample information** into the data spreadsheets (see Appendix 3 for examples of spreadsheets).

2. **Rainfall at the airport**. Go to Web page: *http://www.nws.noaa.gov/*

   a. Enter the location, click "Go," and click on "3 day history."

   b. **AirportRainRd-1_in**. Hourly data for precipitation are in the right three columns and go back 3 days. Add up the precipitation for the last 24 hours (8:53 yesterday to 7:53 today) and record the rainfall amount in the spreadsheet.

3. **Lake level**. Go to Web page: *http://tidesandcurrents.noaa.gov/station_retrieve.shtml?type=Great+Lakes+Water+Level+Data*

   a. Click on the appropriate site.

   b. Change to "feet." Change to "local time."

   c. Click on the box that says "view data."

   d. Obtain the water level value for today's date at 8:00 from the fourth column for use in the model. Record this value in the spreadsheet.

   e. The spreadsheet will calculate the change in lake level since the previous day.

Run the Models in Virtual Beach (VB)

1. Every day that you collect samples, run the models. (For running a nowcast live, this should get done by 9:30 a.m. on all sampling days.)

2. Open the VB software. Click on Model, Open, and open the Beach 1.vbmx file.

3. Enter ID (M/DD/YY) and the variables.

4. Click on the box for "IV Data Validation," click Scan, and click Return.

5. Click on the box for "Make Predictions."

6. Depending on which output you use, record the "Exceedance probability" (Beach 1) or "Model prediction" (for Beach 2) into your spreadsheet.

7. Back in the VB software, click Model, Save, change to Log10, click Save; use the same name and replace the file.

8. After obtaining the *E. coli* results from the Colilert Quanti-Tray, enter them into the "MeasuredEcoli_MPN" column in the spreadsheet. Note the tally of model results (false positive, false negatives, etc.).

9. Open the model for Beach 2 and repeat steps 2–8.

Data Entry and Management

- Enter data daily in the designated state database, as required by the State.

- Every Friday, check over the spreadsheets to be sure that all data have been entered. If you are working with the USGS, send an updated spreadsheet via email to your USGS contact.

- Enter quality-control (QC) data for *E. coli* and turbidity replicates and *E. coli* field blanks on a separate worksheet on appendix 3.

Quality Assurance and Quality Control

Duplicate Water Samples and Counting Checks

Duplicate analyses will be done for *one sample every other week*. Duplicate samples are collected from all sites, and some duplicates should be collected after significant changes in environmental or weather conditions. A duplicate sample is a second sample bottle collected immediately after and at the same location as the first bottle. Collection time is reported with a difference of 2 minutes. Both bottles are analyzed for *E. coli* and turbidity.

For counting checks, fluorescent well counts are made by a second person on approximately 10 percent of samples. If results don't agree within 20 percent, work with a laboratory supervisor to reconcile differences.

Field Blanks

One field blank is collected at the beginning of the sampling period (early June) and one later in the sampling period (July).

- Use a regular sample bottle.

- Pour about 200 mL of sterile phosphate buffer into the bottle under actual field conditions (at the beach). Store the sample with the regular sample bottles in the cooler.

- Process the field blank for *E. coli* along with the regular samples.

Reference Samples for the Culture Method

Pure culture positive and negative controls, purchased from the manufacturer, will be processed in June and again in July. In June, every analyst that is expected to process samples will analyze pure cultures. If a new analyst is brought into the project at some other time, the new analyst also will process the pure cultures.

Turbidity Measurements

The USGS will provide at least one set of reference samples for turbidity. If measured results do not agree within 15 percent of the stated turbidity value, another set of reference samples will be tested.

Quality-Assurance Checks

Someone from the USGS will perform at least one QA/QC check of field and laboratory procedures each season. USGS contacts are listed below.
Donna Francy, dsfrancy@usgs.gov
Amie Brady, amgbrady@usgs.gov

Equipment

Laboratory staff will maintain a logbook for recording equipment checks. The logbook should note the equipment model number and serial number, date and time of each check, and results and initials of the person performing each check.

Incubators. Incubators must be operated at 35+ 0.5 °C. Check the operating temperatures of the incubators by using an external thermometer weekly during the recreational season. The digital thermometer you use is checked before use against a National Institute of Standards and Technology (NIST) thermometer. *Record operating temperature in the logbook.*

Appendix 5.   Equations for the Selected Model(s) for Each Great Lakes Beach

Models were used to predict *Escherichia coli* (*E. coli*) concentrations from measurements of environmental and water-quality variables. Equations were generated using Virtual Beach software (U.S. Environmental Protection Agency, 2012). Variables are described at the end of this appendix.

Illinois

South Beach

$$\text{Ln}(E.\ coli) = -4.8 + 0.022*(\text{Dayofyear}) - 0.16*(\text{Rain\_cat}) + 0.26*(\text{Rain\_cat\_Int}) + 0.37*(\text{Cloud\_cat}) + 0.061*(\text{WaterTemp\_F}) - 0.060*(\text{People\_no}) - 0.0034*(\text{Gulls\_no}) - 0.20*(\text{Rw72})$$

Source of local rainfall data: Evanston, Ill., water treatment plant

Greenwood Beach

$$\text{Ln}(E.\ coli) = -21.5 + 2.5*\text{Ln}(\text{Dayofyear}) + 2.1*\text{Ln}(\text{AirTempF\_INST}) + 0.35*(R_{d-1})^{0.1} + 0.56*\text{Ln}(\text{WaveHt\_in}) - 0.22*(\text{Rain\_cat}) - 0.32*(\text{People\_no})^{0.25} + 0.27*\text{Ln}(\text{Gull\_no}) + 0.45*[3.4 + 1.2*(\text{LakeLevel\_48}) - 0.0020*(\text{LakeLevel\_48})^2] + 0.29*(\text{Algae\_cat})$$

Source of air temperature data: Evanston, Ill., water treatment plant

Lee Beach

$$\text{Ln}(E.\ coli) = -122.6 + 0.029*(\text{Dayofyear}) + 0.045*(\text{WaterTemp\_F}) - 0.046*(\text{Turb}) + 0.31*(\text{Cloud\_cat}) + 0.20*(\text{LakeLevel}) + 0.92*(\text{Algae\_cat}) + 0.10*(\text{WindA\_comp\_INST})$$

Source of local wind data: Evanston, Ill., water treatment plant
Beach orientation is −16.82°.

Montrose

$$\text{Log}_{10}(E.\ coli) = 4.7 - 0.0069*(\text{Dayofyear}) - 0.0044*(\text{SolarRad\_6}) - 0.008327*(\text{Humidity\_24}) + 0.083*(\text{WindO\_comp4})$$

Source of local weather and solar radiation data: City of Chicago weather station at Foster Avenue Beach
Beach orientation is −68.82°.

63rd Street

$$\text{Log}_{10}(E.\ coli) = 0.59 + 0.0067*(\text{Dayofyear}) + 0.86*(R_{6hr\_}\text{cm}) - 0.0026*(\text{SolarRad\_6}) + 0.068*(\text{WindO\_comp6})$$

Source of local weather and solar radiation data: City of Chicago weather station at 63rd Street Beach
Beach orientation is −62.24°.

Foster

$$\text{Log}_{10}(E.\ coli) = 1.8 + 0.0040*(\text{Dayofyear}) - 0.0015*(\text{SolarRad\_4}) - 0.036*(\text{AirTempC\_4}) + 0.12*(\text{WindO\_comp4})$$

Source of local weather and solar radiation data: City of Chicago weather station at Foster Avenue Beach
Beach orientation is −36.3°.

Calumet

$$\text{Log}_{10}(E.\ coli) = -1.8 + 1.6*\text{Log}_{10}(\text{Dayofyear}) + 0.60*(\text{R4hr\_cm}) - 0.0010*(\text{SolarRad\_4}) + 0.13*(\text{WindO\_comp4})$$

Source of local weather and solar radiation data: City of Chicago weather station at 63rd Street Beach
Beach orientation is −23.94°.

Oak Street

$$\text{Log}_{10}(E.\ coli) = -50.6 - 0.00061*(\text{Turb1hr}) + 0.00098*(\text{R48\_cm}) - 2.1*\text{Log}_{10}(\text{SolarRad\_24}) + 0.16*(\text{WindSp\_4}) + 0.076*(\text{BaroPress\_6})$$

Source of turbidity data: City of Chicago nearshore buoy
Source of local weather and solar radiation data: City of Chicago weather station at Foster Avenue Beach

Indiana

Jeorse Park

$$Log_{10} (E.\ coli) = 2.1 - 0.024*(Chloro-A) + 0.00076*(WaterTempC) + 0.023*(WaveHt\_cm) + 0.10*(WindSp\_INST)$$

Source of airport wind data: Gary/Chicago International Airport

Michigan

Metro

$$Log_{10} (E.\ coli) = 2.2 - 0.080*(WindSp\_24V) - 0.049*(WindSp\_INST) - 1.0*(Rw48) + 0.061*(R_{30d}) + 1.1*(LakeLevel\_24) + 0.0082*(Turb*WaveHt\_ft)$$

Source of airport weather data: Detroit City Airport and Selfridge Air Force Base

Memorial

$$Log_{10} (E.\ coli) = 3.6 - 0.0017*(WindDir\_INST) - 0.0092*(DayofYear) + 1.29*(R_{d-1}) + 0.011*(Turb*R_{d-1}) + 0.047*(WindSp\_INST) - 0.040*(WindSp\_24V) + 0.0013*(WindDir\_24V)$$

Source of airport weather data: Detroit City Airport

New York

Ontario

$$Log_{10} (E.\ coli) = -0.15 - 1.4*(CurrentA\_comp\_24) + 1.7*[-4.9 + 0.062*(DayofYear) - 0.00014*(DayofYear)^2] + 0.44*[1.6 + 1.38*(WaveHtPred\_ft) - 0.32*(WaveHtPred\_ft)^2] + 0.0014*(WindSp\_INST)^2 + 0.45*[2.0 + 2.1*(LakeLevel\_24) + 15.6*(LakeLevel\_24)^2] - 0.75*(CurrentO\_comp\_24)^{0.5} - 4.4*(WaveHtPred\_24)^{0.1} + 0.25*(Ln(Turb))$$

Source of airport weather data: Greater Rochester International Airport
Source of current and wave-height prediction data: Great Lakes Coastal Forecasting System (GLCFS)

Durand

$$Log_{10} (E.\ coli) = -7.2 + 0.69*(1.6 + 3.7*(LakeLevel\_24) + 16.8*(LakeLevel\_24)^2) + 2.7*[-2.0 + 0.034*(DayofYear) - 7.4e-05*(DayofYear)^2] + 0.91*(R_{d-1})^{0.5} + 0.55*Log_{10}(Turb) + 0.86*[1.6 - 0.025*(WindA\_comp\_Inst) + 0.0023*(WindA\_comp\_Inst)^2] + 0.12*(Debris\_Cat) - 0.60*[1.2 + 1.6*(WaveHtPred\_24) - 0.62*(WaveHtPred\_24)^2] + 0.96*[2.0 - 0.0067*(WindDir\_24V) + 2.1e-05*(WindDir\_24V)^2] + 7.3e-05*(Birds\_no)^2$$

Source of airport weather data: Greater Rochester International Airport
Source of wave-height prediction data: Great Lakes Coastal Forecasting System (GLCFS)
Beach orientation is −62.6°.

Woodlawn SP

$$Log_{10} (E.\ coli) = -4.5 + 0.54*(Turb)^{0.25} + 1.6*[6.1 - 0.043*(DayofYear) + 0.00011*(DayofYear)^2] - 0.14*(\ Turb-stream\_24)^{0.1} + 0.21*(WaveHt\_ft)^{0.5} - 0.064*(WindO\_comp24) + 0.83*[2.1 - 0.99*(R_{d-3}) + 0.99*(R_{d-3})^2]$$

Source of airport weather data: Buffalo Niagara International Airport
Beach orientation is −180°.

Evans TP

$$Log_{10} (E.\ coli) = -0.076 + 0.63*(Turb)^{0.25} + 0.73*[1.4 + 0.00031(WindA\_comp\_INST) + 0.0026*(WindA\_comp\_INST)^2] - 70.3*(1/WaterTempPred\_24) + 0.38*[1.5 + 0.39*(LakeLevel\_24) + 1.0*(LakeLevel\_24)^2] + 0.69*[1.6 - 1.6*(R_{d-3}) + 1.5*(R_{d-3})^2] + 0.26*(R_{d-1})^{0.25} + 0.33*[1.6 - 0.0042*(WindDir\_INST) + 1.83e-05*(WindDir\_INST)^2]$$

Source of airport rain and wind data: Buffalo Niagara International Airport
Source of water temperature prediction: Great Lakes Coastal Forecasting System (GLCFS)
Beach orientation is −123°.

Lake Erie

$$Log_{10} (E.\ coli) = -2.4 + 0.036*(WindSp\_INST) + 0.91*$$
$$[1.6 - 1.0*(R_{d-3}) + 0.96*(R_{d-3})^2] + 0.37*(Rw48)^{0.5} +$$
$$0.16*(Turb)^{0.5} + 1.0*[-9.6 + 0.88*(AirTemp\_mean24) -$$
$$0.17*(AirTemp\_\ mean24)^2]$$

Source of airport weather and air temperature data: Buffalo Niagara International Airport

Ohio

Edgewater

When radar rainfall data are available—

$$Log_{10}(E.\ coli) = 0.55 + 0.24*[WaveHt\_ft + Log_{10}(Turb)] +$$
$$0.0042*(DayofYear) + 0.36*(Radarxcell\text{-}sum\text{-}Rw48)^{0.5} +$$
$$0.26*(LakeLevel\_24) + 0.17*(Subseason1)$$

When radar rainfall data are not available—

$$Log_{10}(E.\ coli) = 0.52 + 0.0045*(DayofYear) + 0.32*(Rw48)^{0.5}$$
$$+ 0.19*(Subseason1) + 0.24*[WaveHt\_ft + Log_{10}(Turb)] +$$
$$0.30*(LakeLevel\_24)$$

Source of airport weather data: Cleveland Hopkins International Airport

Huntington

When radar rainfall data are available—

$$Log_{10}(E.\ coli) = 0.99 + 0.51*(R_{d-1})^{0.5} + 0.29*$$
$$[Log_{10}(Turb) + (WaveHt\_ft)^{0.5}] - 0.10*(PreJuly24) +$$
$$0.013*(WindSp\_INST*WindCode) + 0.11*$$
$$(Radarxcell\text{-}sum\text{-}Rw48)^{0.5}$$

When radar rainfall data are not available—

$$Log_{10}(E.\ coli) = 0.92 + 0.75**(R_{d-1})^{0.5} + 0.28*$$
$$[Log_{10}(Turb) + (WaveHt\_ft)^{0.5}] + 0.012*$$
$$(WindSp\_INST*WindCode) + 0.19*(R_{d-2})^{0.5} +$$
$$0.096*(PostJuly24) + 0.30*(LakeLevel\_24)$$

Source of airport weather data: Cleveland Hopkins International Airport

Maumee Bay State Park

When discharge from Berger Ditch was positive—

$$Log_{10}(E.\ coli) = 1.5 + 0.40*(R_{d-2})^{0.25} + 0.060*WindCode\_4$$

When discharge from Berger Ditch was negative (from backwater from Maumee Bay)—

$$Log_{10}(E.\ coli) = 1.1 + 10.9*Ln(LakeLevel\_24) +$$
$$0.083*[WindCode\_4*Log_{10}(Turb)] - 0.40*(Disch\_730am)^{0.25}$$

Source of airport weather data: Toledo Executive Airport (formerly Metcalf Field)

Villa Angela

$$Log_{10}(E.\ coli) = 6.4 - 0.11*Ct$$

Vermilion

$$Log_{10}(E.\ coli) = 0.50 + 0.44*[(WaveHt\_ft)^{0.5} + Log_{10}(Turb)] +$$
$$0.33*Log_{10}(Disch\_8am) + 0.25*(R_{d-1})^{0.5}$$

Source of airport weather data: Lorain County Regional Airport
Source of stream−discharge data: Vermilion River near Vermilion Ohio, USGS site 04199500

Lakefront

$$Log_{10}(E.\ coli) = 1.7 + 0.22*[WaveHt\_ft + Log_{10}(Turb) +$$
$$LakeLevel\_24]$$

Nickel Plate

$$Log_{10}(E.\ coli) = -0.037 - 0.19*(WindO\_comp\_INST)$$
$$+ 0.33*(WaveHt\_ft) + 14.5*(1/Disch\_M_{d-2}) +$$
$$0.0057*(Dayofyear) + 0.0084*(Turb)$$

Source of wind data: Old Woman Creek, Ohio, NOAA National Estuarine Research Reserve System-wide Monitoring Program site OWCOWMET
Source of stream-discharge data: Huron River at Milan Ohio, USGS site 04199000
Beach orientation is −58.7°.

Lakeview, Lorain

$$Log_{10}(E.\ coli) = 0.09 + 0.13*(WindSp\_INST)^{0.5} \\ - 0.27*(LakeLevel\_Spring)^2 + 0.32*(Rw48)^{0.5} + \\ 0.0080*(Dayofyear) + 0.14*(Turb)^{0.5}$$

Source of airport weather data: Lorain County Regional Airport

Headlands

$$Log_{10}(E.\ coli) = 0.35 - 0.026*(WindA\_comp\_INST) + \\ 0.15*(Rw72)^{0.5} + 0.36*[(WaveHt\_ft)^{0.5} + \\ Log_{10}(Turb)] + 0.19*Log_{10}(Disch\_8am)$$

Source of airport wind direction and speed data: Ashtabula County Airport
Source of airport rainfall data: Burke Lakefront Airport
Source of stream-discharge data: Grand River near Painesville Ohio, USGS site 04212100
Beach orientation is −134°.

Fairport Harbor

$$Log_{10}(E.\ coli) = 0.7242 + 0.074*(Turb)0.5 - 0.021* \\ (WindA\_comp) + 0.57*(WaveHt\_ft)^{0.5} + \\ 0.17*(Log_{10}(Disch\_8am) + 0.41*(R_{d-1})^{0.5}$$

Source of airport weather data: Ashtabula County Airport
Source of stream-discharge data: Grand River near Painesville Ohio, USGS site 04212100
Beach orientation is −76.6°.

Pennsylvania (Presque Isle State Park)

Beach 1 West Extension

$$Log_{10}(E.\ coli) = 0.60 + 1.0*Log_{10}(Turb) + \\ 0.80*(LakeLevel\_24) - 0.041*(WindA\_comp\_Inst)$$

Beach orientation is −157°.

Beach 2

$$Log_{10}(E.\ coli) = 147.6 + 1.1*Log_{10}(Turb) - 100.5*Log_{10} \\ (BaroPress\_INST) + 0.52*(LakeLevel\_24)$$

Beach 6

$$Log_{10}(E.\ coli) = 26.1 + 0.86*Log_{10}(Turb) - \\ 0.87*(BaroPress\_INST) + 0.50*(LakeLevel\_24)$$

Mill Road

$$Log_{10}(E.\ coli) = 29.6 + 0.66*Log_{10}(Turb) + \\ 0.57*(LakeLevel\_24) - 0.98*(BaroPress\_INST) + \\ 0.14*(WaveHt\_ft) - 0.44*(R72)$$

Beach 9

$$Log_{10}(E.\ coli) = 35.9 + 0.92*Log_{10}(Turb) + \\ 0.48*(LakeLevel\_24) - 1.3*(BaroPress\_INST) - 0.44*(R48) \\ + 0.44*Log_{10}(WindDir\_24V) + 0.0037*(Dayofyear)$$

Beach 11

$$Log_{10}(E.\ coli) = 27.0 + 0.076*(Turb) + 3.3* \\ (Log_{10}(Dayofyear)) - 1.2*(BaroPress\_INST) + \\ 0.1507*(Rw72)$$

Source of rainfall data: Erie International Airport
Source of local weather data: Regional Science Consortium

Wisconsin

Source of data for water temperature, air temperature, current direction, cloud cover, lake level, wave height, and surface current predictions: Great Lakes Coastal Forecasting System (GLCFS)

Upper Lake Park

$$\text{Log}_{10}\,(E.\ coli) = -3.23 + 0.069*(\text{WaveHtPred\_ft}) + 0.54*(\text{Turb})^{1/3} + 0.45*[1.14 - 0.0088*(\text{Gulls\_no}) + 9.77e{-}5*(\text{Gulls\_no})^2] + 1.04*[1.06 - 1.17* (\text{CurrentO\_comp\_1}) + 59.8*(\text{CurrentO\_comp\_1})^2] + 0.00030*(\text{Turb}*\text{RadarRain\_48}) + 0.34*(\text{WaterTempPred\_1})^{1/2} + 0.35*(\text{WaveO\_comp\_1}) + 0.00014*(\text{WaterTempPred\_F})^2 - 0.012*(\text{RadarRain\_24})^{1/4} + 0.31*(\text{Algae\_cat})$$

Beach orientation is 31.71°.

Red Arrow

$$\text{Log}_{10}\,(E.\ coli) = 1.2 + 0.025*\text{Log}_{10}\,(\text{Turb}) + 1.0*(\text{CloudCover\_24}) + 0.15* (\text{WaterTempPred\_SD48}) + 1.1*(\text{WaveHtPred\_12})$$

Neshotah

$$\text{Log}_{10}\,(E.\ coli) = 2.12 + 0.35*\text{Log}_{10}\,(\text{Turb}) + 14.0* (\text{LakeHt\_1}) - 0.70*(\text{AirTempPred\_SD144}) - 0.29* (\text{WaterTempPred\_SD6}) - 7.0*(\text{CurrentO\_comp\_144}) + 1.13*(\text{WaveHtPred\_1}) + 0.26*(\text{CloudCover\_2})$$

Beach orientation is 61.3°.

Point Beach South, Central, and North

$$\text{Log}_{10}\,(E.\ coli) = 0.61 + 0.77*\text{Log}_{10}\,(\text{Turb}) - 57.8*(\text{LakeHt\_max24}) + 0.80*(\text{WaveHtPred\_2}) - 24.6*(\text{CurrentO\_comp\_120}) - 0.0029*(\text{Disch\_diff6}) + 0.050*(\text{WaterTempPred\_6}) - 0.53* (\text{WaveO\_comp\_24}) + 0.059*(\text{RadarRain\_48})^{1/2}$$

Source of stream-discharge data: Manitowoc River at Manitowoc, WI, USGS site 04085427
Beach orientations are 9.8 °(south), 1.8 °(central), 356.5°(north).

Fischer

$$\text{Log}_{10}\,(E.\ coli) = 0.75 + 0.60*\,\text{Log}_{10}\,(\text{Turb}) + 0.64*\,\text{Log}_{10} (\text{Turb\_stream}) + 7.0*(\text{CurrentA\_comp\_48}) + 0.29*(\text{Rain\_gage\_120})_{1/2} - 10.5*(\text{CurrentA\_comp\_D12})$$

Source of rain-gage data: Sheboygan River at Sheboygan, Wisconsin, USGS site 04086000
Beach orientation is 25.5°.

Hika

$$\text{Log}_{10}\,(E.\ coli) = 0.74 + 0.42*\,\text{Log}_{10}\,(\text{Turb}) + 0.0021*(\text{Conductance}) + 0.34*(\text{Rain\_gage\_48})1/2 + 0.24*(\text{Rain\_gage\_12})_{1/2}$$

Source of rain-gage data: Sheboygan River at Sheboygan, Wisconsin, USGS site 04086000

Thompson W. End Park

$$\text{Log}_{10}\,(E.\ coli) = -1.8 + 0.080*\,(\text{Turb}) + 0.87*(\text{CloudCover\_2}) + 0.15*(\text{RadarRain\_120})_{1/2} - 0.10*(\text{WindOPred\_comp\_6}) - 0.11*(\text{WindAPred\_comp\_24}) + 0.049*(\text{WaterTempPred\_C}) + 0.074*(\text{AirTempPred\_2}) + 0.55*(\text{Rain\_gage\_24})\,1/2$$

Source of rain-gage data: Whittlesey Creek near Ashland, Wisconsin, USGS site 040263205
Beach orientation is 78.4°.

Kreher Park

$$\text{Log}_{10}\,(E.\ coli) = 1.9 + 0.84*\,\text{Log}_{10}\,(\text{Turb}) + 0.078*(\text{AirTempPred\_48}) - 0.32* (\text{AirTempPred\_SD120}) + 12.8*(\text{LakeHt\_diff24}) + 0.76*(\text{Rain\_gage\_120})\,1/2 - 0.079*(\text{Disch\_min72}) + 13.0*(\text{LakeHt\_min24}) + 9.5*(\text{LakeHt\_1})$$

Source of rain-gage and stream-discharge data: Whittlesey Creek near Ashland, Wisconsin, USGS site 040263205

Maslowski

$$\text{Log}_{10}\,(E.\ coli) = 1.1 + 0.11*(\text{RadarRain\_48})_{1/2} - 0.0015*(\text{Conductance-stream}) + 0.052*(\text{AirTempPred\_2}) - 0.30*(\text{AirTempPred\_SD6}) + 0.46*\,\text{Log}_{10}\,(\text{Turb}) + 0.044*(\text{WindAPred\_comp\_2}) + 5.6*(\text{LakeHt\_diff24})$$

Beach orientation is 280.4°.

Key to Acronyms and Variables

nr, near

USGS, U.S. Geological Survey

USACE, U.S. Army Corps of Engineers

field, source of data is a field observation or measurement

lab, source of data is a lab measurement

airport, source of data is the National Weather Service nearest airport site

radar, source of data is the National Weather Service radar data

USGS gage, source of data is a nearby USGS gage

NOAA measurement, source of data is a measurement provided by the National Oceanic and Atmospheric Administration (NOAA)

NOAA prediction, source of data is a prediction provided by the NOAA Great Lakes Coastal Forecasting System (GLCFS)

Local, source of data is a local weather station or gage

AirTemp: (field) Air temperature in degrees Celsius (_C) or Fahrenheit (_F) at the time of sampling (_INST) or the mean for the 4-hour period (4) or 24-hour period (24) prior to sampling

AirTemp_mean24: (airport) Mean air temperature for previous 24 hours, in degrees Celsius

AirTempPred: (NOAA prediction) Hourly air temperature from the nearest GLCFS grid point used to compute mean for the 2-hour (_2) or 48-hour (_48) period before sampling or standard deviation for the 6-hour (_SD6), 120-hour (_SD120), or 144-hour (_SD144) period before sampling

Algae_cat: (field) Category of quantity of algae present at the beach at time of sampling—none, some, many

BaroPress: (local) Mean barometric pressure, in millimeters of mercury at the time of sampling (_INST) or for the 6-hour period prior to sampling (_6)

Beach orientation: If the beach lies straight along the north/south line, beach orientation is 0 degrees with land to the west and water to the east or 180 degrees with water to the west and land to the east. Beach orientations from 1 to 179 degrees deviate from the north/south line accordingly, with land to the west or north and water to the east or south. Beach orientations from −1 to −179 degrees deviate from the north/south line accordingly, with land to the east or south and water to the west or north.

Birds_no: (field) Number of birds on the beach at the time of sampling

Cloud_cat: (field) Categories for sunny (1), partly cloudy (2), or cloudy (3) at the time of sampling

Chloro-A: (lab) Chlorophyll a, in relative fluorescence units, using a laboratory fluorometric sensor

CloudCover: (NOAA prediction) Hourly cloud cover from the nearest GLCFS grid point, mean over 2 (_2) or 24 (_24) hours prior to sampling

Conductance: (field) Conductivity of the water at the beach, measured at the time of sampling

Conductance-stream: (field) Conductivity of the water in a nearby stream, measured near the time of sampling

CurrentO_comp: (NOAA prediction) Measure of the component of the surface current velocity moving perpendicular to the shoreline at the nearest GLCFS grid point using mean for the 1-hour (_1), 24-hour (_24), 120-hour (_120), or 144-hour (_144) period before sampling

CurrentA_comp: (NOAA prediction) Measure of the component of the surface or depth-averaged current velocity moving parallel to the shoreline at the nearest GLCFS grid point using the mean depth-averaged current for the 12-hour period before sampling (_D12) or the mean for the 24-hour (_24) or 48-hour (_48) period before sampling

Ct: (lab) mean sample cycle threshold value (Ct) from quantitative polymerase chain reaction for *E. coli*

Dayofyear: Number representing the date beginning with 1 for January 1 and 365 or 366 for December 31 (the latter being a leap year)

Debris_cat: (field) Debris categorical variable (1 for no debris, 2 for some, 3 for visible floating, 4 for thick layer, 5 for multiple layers)

Disch: (USGS gage) Discharge, in cubic feet per second:

- at a specific time (_*time*)

- the mean for the 24-hour period before sampling ($\_M_{d-1}$)

- the mean for the 24-hour period 2 days prior to sampling (_Md_$_{-2}$)

- the difference within 6-hour period before sampling (_diff6)

- the minimum within 72-hour period before sampling (_min72)

E. coli: (lab) Concentration of *Escherichia coli*, in most-probable number per 100 milliliters

Gulls_no: (field) Number of gulls on the beach at the time of sampling

Humidity: (local) Mean relative percent humidity for the 24 hours prior to sampling (_24)

LakeLevel: (NOAA measurement) Lake level, in feet at 8 a.m.,

- on the day of sampling

- the change in lake level (today – yesterday) (_24)

- the change in lake level (today – 2 days ago) (_48)

- the deviation from the spring average (_Spring)

LakeHt: (NOAA prediction) Difference of the water level at beach location compared to the mean of all GLCFS grid cells for the entire lake:

- 1 hour prior to sampling (_1)

- the maximum value within the 24-hour period before sampling (_max24)

- the minimum value within 24-hour period before sampling (_min24)

- difference between 1 hour and 24 hours prior to sampling (_diff24)

People_no: (field) Number of people at the beach (swimmers and on the sand) at the time of sampling

PreJuly24: Binary variable that equals 1 if the date is before July 24 and 0 if the date is July 24 or later

PostJuly24: Binary variable that equals 1 if the date is July 24 or later and 0 if the date is before July 24

RadarRain: (radar) Radar-indicated precipitation from the NWS National Mosaic Quantitative Precipitation Archive (4-kilometer grid for the United States). Grid cells relevant to the defined watershed were used to compute an area-weighted average by using Thiessen polygons. Total precipitation using the 24-, 48-, or 120-hour period before sampling: (_24), (_48), (_120)

Radarxcell-sum-Rw48: (radar) Sum from multiple cells for the 24-hour periods 1 and 2 days before sampling. The most recent rainfall receives the most weight. Calculated as: (2*Radarxcell-sum-Rd-1) + Radarxcell-sum-Rd-2

Rain_cat: (field) Rain occurred with 24 (category = 1), 48 (category =2), or 72 (category =3) hours

Rain_cat_Int: (field) If rain occurred, was it light (1), moderate/steady (2), or heavy (3)

Rain_gage: (USGS gage) Sum of rainfall, in inches: for the 12-hour (_12), 24-hour (_24), 48-hour (_48), or 120-hour (_120) period before sampling

R_{4hr}_cm: (airport) Total rainfall, in centimeters, for the 4-hour period prior to sampling

R_{6hr}_cm: (airport) Total rainfall, in centimeters, for the 6-hour period prior to sampling

R_{d-1}: (airport) Total rainfall, in inches, for the 24-hour period before sampling

R_{d-2}: (airport) Total rainfall, in inches, for the 24-hour period 2 days before sampling

R_{d-3}: (airport) Total rainfall, in inches, for the 24-hour period 3 days before sampling

R48: (airport) Amount of rainfall, in inches or centimeters, for the 48-hour period before sampling

R72: (airport) Amount of rainfall, in inches for the 72-hour period before sampling

Rw48: (airport) Amount of rainfall, in inches, for the 48-hour period before sampling, with the most recent rainfall receiving the most weight. Calculated as $(2*Rd-1)+Rd-2$

Rw72: (airport) Amount of rainfall, in inches, for the 72-hour period before sampling, with the most recent rainfall receiving the most weight. Calculated as $(3*R_{d-1}) + (2*_{Rd-2}) + R_{d-3}$

R_{30d}: (airport) Total rainfall, in inches, for the 30-day period before sampling

SolarRad: (local) Mean solar radiation, in watts per square meter, for the 4 hours (_4), 6 hours (_6), or the 24 hours (_24) prior to sampling

Subseason1: Equals 1 if the sampling date is June 15 or earlier and is 0 if the sampling date is after June 15

Turb: (field) Turbidity of the sample from the beach, measured with a turbidimeter, in nephelometric turbidity ratio units

Turb1hr: (field) Mean water turbidity for the 1-hour period prior to sampling from a nearshore buoy, in nephelometric turbidity units

Turb_stream: (field) Turbidity of the water in a nearby stream, measured with a turbidimeter, in nephelometric turbidity ratio units

Turb-stream_24: (USGS gage) Change in turbidity in a nearby stream (today – yesterday)

WaterTemp: (field) Water temperature at time of sampling, in degrees Celsius (_C) or in degrees Fahrenheit (_F)

WaterTempPred: (NOAA prediction) Hourly water temperature, in degrees Celsius, from the nearest GLCFS grid point used to compute mean for the 1-hour, 6-hour, or 24-hour period before sampling (_1, _6, _24) or standard deviation for the 6-hour (_SD6) or 48-hour (_SD48) period before sampling

WaveHt: (field) Wave height as measured with a graduated rod, in feet (_ft), inches (_in), or centimeters (_cm)

WaveHtPred: (NOAA prediction) Hourly wave height from the nearest GLCFS grid point used to compute mean for the 1-hour (_1), 2-hour (_2), 12-hour (_12), or 24-hour (_24) period before sampling

WaveO_comp: (NOAA prediction) Measure of the component of the wave velocity moving perpendicular to the shoreline at the nearest GLCFS grid point using the instantaneous value near 8 a.m. (INST) or the mean for the 1-hour (_1) or 24-hour (_24) period before sampling

WindA_comp: (airport) Measure of the component of the wind velocity moving parallel to the shoreline, for the instantaneous (INST) value near 8 a.m. or a 24-hour vector up to 8 a.m. (24), calculated as:

WindA = −wind speed * cosine ((wind direction – beach orientation) * π/180)
A positive value indicates winds are moving from right to left across the beach when looking toward the water from the shoreline (U.S. Environmental Protection Agency, 2012).

WindAPred_comp: (NOAA prediction) Measure of the component of the wind velocity moving parallel to the shoreline at the nearest GLCFS grid point using the mean for the 2-hour (_2) or 24-hour (_24) period before sampling

WindCode: (airport) Site-specific wind code calculated by assigning the most weight to the range of wind directions associated with the highest E. coli concentrations:

- at the time of sampling (_INST)

- summed each hour for the 4-hour period prior to sampling (_4)

WindDir: (airport) Wind direction, in degrees, at:

- the time of sampling (_INST)

- the 24-hour vector up to 8 a.m. (_24V)

WindO_comp: (airport) Measure of the component of the wind velocity moving perpendicular to the shoreline, using:

- the instantaneous (INST) value near 8 a.m.

- the mean wind speed and median wind direction for the previous 4 or 6 hours prior to sampling, or

- a 24-hour vector up to 8 a.m. (24)

  calculated as:
  WindO = wind speed * sine ((wind direction – beach orientation) * π/180)
  A positive value indicates winds are moving from the water toward shore (U.S. Environmental Protection Agency, 2012).

WindOPred_comp: (NOAA prediction) Measure of the component of the wind velocity moving perpendicular to the shoreline at the nearest GLCFS grid point using the mean for the 6-hour (_6) period before sampling

WindSp: (airport) Wind speed, in miles per hour, at:

- the time of sampling (_INST)

- the mean for the 4-hour period prior to sampling (_4)

- the 24-hour vector up to 8 a.m. (_24V)

Reference

U.S. Environmental Protection Agency, Center for Exposure Assessment Modeling, 2012, Exposure Assessment Models—Virtual Beach, accessed June 2013 at *http://www2.epa.gov/exposure-assessment-models/virtual-beach-vb*.

USGS

Francy and others—Developing and Implementing Predictive Models for Estimating Recreational Water Quality at Great Lakes Beaches —SIR 2013-5166