USGS
science for a changing world

In cooperation with the
Erie County Health Department

Monitoring and Modeling to Predict *Escherichia coli* at Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania

Scientific Investigations Report 2006-5159

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

Monitoring and Modeling to Predict *Escherichia coli* at Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania

By Tammy M. Zimmerman

In cooperation with the
Erie County Health Department

Scientific Investigations Report 2006-5159

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

U.S. Department of the Interior
DIRK KEMPTHORNE, Secretary

U.S. Geological Survey
P. Patrick Leahy, Acting Director

U.S. Geological Survey, Reston, Virginia: 2006

For sale by U.S. Geological Survey, Information Services
Box 25286, Denver Federal Center
Denver, CO 80225

For more information about the USGS and its products:
Telephone: 1-888-ASK-USGS
World Wide Web: http://www.usgs.gov/

Suggested citation:
Zimmerman, T.M., 2006, Monitoring and modeling to predict Escherichia coli at Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania: U.S. Geological Survey Scientific Investigation Report 2006-5159, 15 p.

Contents

Figures

Tables

Conversion Factors

Multiply	By	To obtain
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
milliliter (mL)	0.06102	cubic inch (in^3)

Concentrations of bacteria in water are reported in colonies per 100 milliliters (col/100 mL)

Monitoring and Modeling to Predict *Escherichia coli* at Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania

By Tammy M. Zimmerman

Abstract

The Lake Erie shoreline in Pennsylvania spans nearly 40 miles and is a valuable recreational resource for Erie County. Nearly 7 miles of the Lake Erie shoreline lies within Presque Isle State Park in Erie, Pa. Concentrations of *Escherichia coli* (*E. coli*) bacteria at permitted Presque Isle beaches occasionally exceed the single-sample bathing-water standard, resulting in unsafe swimming conditions and closure of the beaches.

E. coli concentrations and other water-quality and environmental data collected at Presque Isle Beach 2 during the 2004 and 2005 recreational seasons were used to develop models using tobit regression analyses to predict *E. coli* concentrations. All variables statistically related to *E. coli* concentrations were included in the initial regression analyses, and after several iterations, only those explanatory variables that made the models significantly better at predicting *E. coli* concentrations were included in the final models. Regression models were developed using data from 2004, 2005, and the combined 2-year dataset. Variables in the 2004 model and the combined 2004–2005 model were $\log_{10}$ turbidity, rain weight, wave height (calculated), and wind direction. Variables in the 2005 model were $\log_{10}$ turbidity and wind direction. Explanatory variables not included in the final models were water temperature, streamflow, wind speed, and current speed; model results indicated these variables did not meet significance criteria at the 95-percent confidence level (probabilities were greater than 0.05). The predicted *E. coli* concentrations produced by the models were used to develop probabilities that concentrations would exceed the single-sample bathing-water standard for *E. coli* of 235 colonies per 100 milliliters. Analysis of the exceedence probabilities helped determine a threshold probability for each model, chosen such that the correct number of exceedences and nonexceedences was maximized and the number of false positives and false negatives was minimized. Future samples with computed exceedence probabilities higher than the selected threshold probability, as determined by the model, will likely exceed the *E. coli* standard and a beach advisory or closing may need to be issued; computed exceedence probabilities lower than the threshold probability will likely indicate the standard will not be exceeded. Additional data collected each year can be used to test and possibly improve the model. This study will aid beach managers in more rapidly determining when waters are not safe for recreational use and, subsequently, when to issue beach advisories or closings.

Introduction

Pennsylvania has approximately 40 mi of Lake Erie shoreline in Erie County—a valuable recreational resource for the county—attracting visitors for activities that include boating, hiking, biking, bird watching, and swimming. Approximately 7 mi of that shoreline lies within Presque Isle State Park in Erie, Pa. (fig. 1). Most of the Lake Erie shoreline in Presque Isle State Park is unpermitted; permitted beaches for public bathing at Presque Isle make up less than 1 mi (Natural Resources Defense Council, 2005). In 2004, *Escherichia coli* (*E. coli*) bacteria concentrations at permitted Presque Isle beaches exceeded the single-sample bathing-water standard and resulted in 3 advisory/closing days (Natural Resources Defense Council, 2005).

Beach advisories or closings are issued on the basis of recreational water-quality standards for fecal-indicator bacteria. Most states have adopted recreational water-quality standards that include criteria for fecal-indicator bacteria. Tests for fecal-indicator bacteria are relatively easy and inexpensive to conduct. These bacteria are not usually harmful, but are indicative that fecal contamination and pathogenic (disease-causing) bacteria may be present. Sources of fecal contamination to recreational waters include combined-sewer overflows and sanitary-sewer overflows; incomplete treatment of sewage; fecal pollution from birds, swimmers, or boaters; and stormwater runoff.

The level of fecal contamination of recreational waters in Pennsylvania was assessed using concentrations of fecal coliform bacteria up to July 2004 when Pennsylvania adopted the U.S. Environmental Protection Agency (EPA)-recommended bacteriological criteria that used *E. coli* (U.S. Environmental Protection Agency, 1986) as the standard for determining the bacteriological quality of all public bathing beaches in the Commonwealth of Pennsylvania (Pennsylvania Bulletin, 2004; Pennsylvania Code, 28 PaCode § 18). Anticipating that the new regulations would be adopted in Pennsylvania, beach

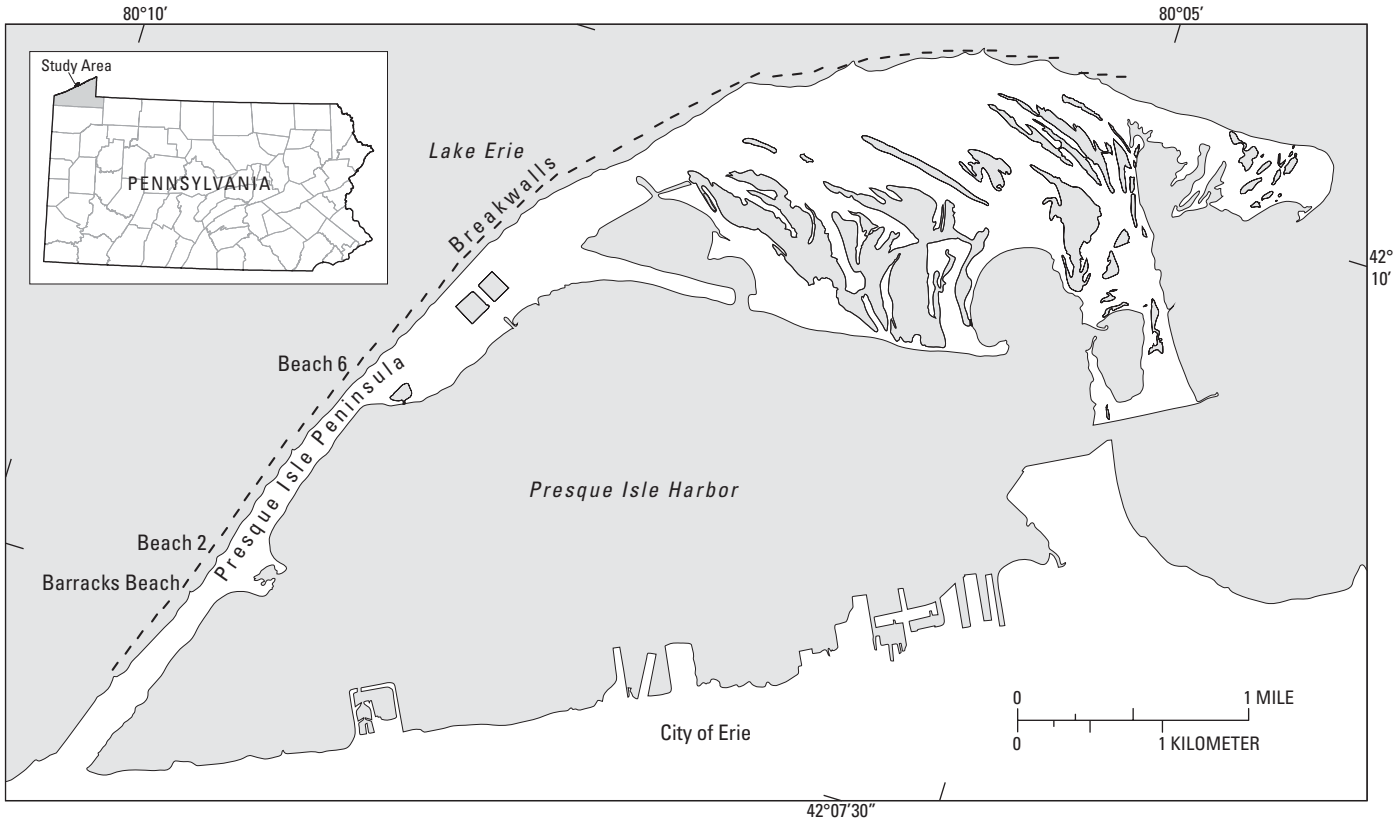

Figure 1. Location of Presque Isle Beach 2 and adjacent beaches at Presque Isle State Park, City of Erie, Erie County, Pennsylvania (modified from Pennsylvania Department of Conservation and Natural Resources, 2006).

managers for Presque Isle beaches were prepared and began using *E. coli* as the standard for closing/advisory determinations in July 2004. The new (2004) regulations established the single-sample bathing-water standard for *E. coli* as 235 col/100 mL. The new regulations also stated that *E. coli* concentrations in all water samples collected in any 30-day period during the recreational season can not exceed a geometric mean of 126 col/100 mL (Pennsylvania Code, 28 PaCode § 18.28). Presque Isle beaches are posted for closure if either the single-sample bathing-water standard or the geometric mean is exceeded.

The use of *E. coli* as an indicator of recreational water quality has been largely effective in determining when fecal contamination is present; however, there are drawbacks with using it as the *only* indicator. Concentrations of *E. coli* may change significantly between the time of sample collection and the reporting of results (anywhere from 18-24 hours). A more rapid method that some managers of recreational waters have adopted is the use of water-quality and environmental variables as surrogates for fecal-indicator bacteria that include, for example, precipitation, wind speed and direction, streamflow, and turbidity to predict, or forecast, when concentrations of fecal-indicator bacteria will exceed recreational standards. These emerging techniques may supplement the use of *E. coli* as an indicator of fecal contamination.

The U.S. Geological Survey (USGS), in cooperation with the Erie County Health Department (ECHD), studied the use of water-quality and environmental variables in beach-specific predictive models as surrogates for *E. coli* in forecasting the bacteriological health of one Presque Isle beach near Erie, Pa.

Purpose and Scope

This report describes monitoring conducted and statistical methods used to develop regression equations to predict *E. coli* concentrations. The regression equations were developed from data collected and compiled during the 2004 and 2005 recreational seasons (May through September each year) at Presque Isle Beach 2 in Erie, Pa. (fig. 1). Statistical relations between *E. coli* and selected water-quality and environmental variables were determined and are presented. Regression techniques were used and regression equations are presented to provide a predictive model for *E. coli* concentrations at Beach 2. These predicted concentrations were then used to develop probabilities that the single-sample bathing-water standard for *E. coli* would be exceeded. The predictive model may be used by beach managers to more rapidly assess exceedences of *E. coli* bacteria standards and subsequently alert the public with water-quality advisories and beach closings.

Previous Studies

Previous studies have helped resource managers of recreational waters gain insights to understanding the various complex processes that contribute to exceedences of recreational water-quality standards. Results from studies similar to the study on which this report is based demonstrate how water-quality and environmental data can be used to develop predictive models for determining exceedences of bacteria standards. Statistical methods for developing and testing predictive models using regression techniques have been documented in several previous studies. Examples of more recent studies are summarized below.

Studies by the USGS in the Cleveland, Ohio, area since 1997 have documented the use of bacteria predictive modeling as a tool to aid resource managers in determining when to issue advisories and closings. A study conducted during the 1997 recreational season (May – September) at three Lake Erie public bathing beaches in Ohio used environmental and water-quality data to determine what factors affect *E. coli* concentrations (Francy and Darner, 1998). Simple- and multiple-linear-regression techniques were used to develop bacteria predictive models based on the environmental and water-quality factors statistically correlated to *E. coli*. Francy and Darner (1998) found that using a single factor of turbidity in models using simple linear regression did not explain as much of the variation in *E. coli* concentrations as using multiple factors in models using multiple linear regression. The variables used in the models were beach specific and included combinations of turbidity, antecedent rainfall, volumes of wastewater-treatment plant overflows, metered outfalls made up of storm-water runoff and combined-sewer overflows, a resuspension index, and wave heights. The best model with the explanatory variables turbidity, weighted rainfall, and wave height explained 58 percent of the variability in $\log_{10}$ *E. coli* concentrations. The models were originally intended to predict concentrations of *E. coli*, but because the prediction intervals were too wide to provide confident predictions, the models instead were used to provide probabilities of exceeding the Ohio single-sample bathing-water standard for *E. coli* of 235 col/100 mL (Francy and Darner, 1998). More recent studies reported in Francy and Darner (2002) and Francy and others (2003) looked at additional beaches and at improving models with additional data at beaches where models had already been developed. Additional data did not always result in improved models, and explanatory variables were found to be beach specific.

A similar study was conducted during the 2000 recreational season at 63rd Street Beach in Chicago, Ill., to develop a regression model to predict when bacteria concentrations met or exceeded full-body contact recreational water quality (Olyphant and Whitman, 2004). Water-quality and environmental variables were collected or compiled and considered for inclusion in the model. The best model included the explanatory variables wind speed, wind direction, rainfall, insolation (incoming solar radiation), lake stage, water temperature, and turbidity and accounted for 71 percent of the variability in *E. coli* concentrations.

Another study conducted by the Stamford Connecticut Department of Health used the single explanatory variable of rainfall to predict when enterococcus concentrations at Stamford beaches would exceed recreational water-quality standards (Kuntz and Murray, 2000). The study analyzed 8 years of data (1989 to 1996) to develop statistical models for Stamford beaches. The best models for the beaches used categories of rainfall (less than 1 in. and 1 in. or more) to predict exceedences to the enterococcus standard of 61 col/100 mL. It was determined that rainfall greater than 1 in. in 24 hours typically resulted in exceedences to the enterococcus standard.

Methods of Study

Data were collected during the 2004 and 2005 summer recreational seasons (May through September) at one Presque Isle beach—Presque Isle Beach 2 (fig. 1). Sampling each week involved the collection of daily water samples Sunday through Tuesday (2004 recreational season) and Sunday through Wednesday (2005 recreational season) from two locations, one near the east and one near the west end of the beach. Field personnel (Presque Isle State Park, ECHD, or USGS staff) collected the water samples using grab-sample techniques described in Myers (2003) to maintain sterile sampling conditions. Grab samples were collected in sterile 125-mL polypropylene bottles in at least 3 ft of water that were opened approximately 12 in. below the water surface. Field characteristics of water temperature, specific conductance, pH, dissolved oxygen, and turbidity were determined at the time of sampling using a Hydrolab Quanta water-quality monitoring system (Hydrolab Corporation, 2002). To ensure samples were collected consistently at the same east (hereafter referred to as Beach 2-East) and west (hereafter referred to as Beach 2-West) locations, latitudes and longitudes of the sample locations were recorded using a Global Positioning System (GPS) unit at the beginning of the recreational season, and the locations were marked with wooden stakes on shore.

Collection of Ancillary Data

Ancillary data were collected by field personnel at the time of sampling or compiled from other sources by the USGS. Detailed information on local activity, such as number of birds, beach debris, and boat activity, was recorded by field personnel at the time of sample collection. An estimated wave height also was determined by field personnel. Wave heights were categorized into one of five groups by visual inspection (0 to 2 ft, 1 to 3 ft, 2 to 4 ft, 3 to 5 ft, and 4 to 6 ft). Data from the nearest USGS streamflow-gaging station at Brandy Run near Girard, Pa. (04213075), were used to determine instantaneous streamflow to the nearest 30 minutes of when water samples were collected (Siwicki, 2005, 2006). Brandy Run does not discharge directly

to Lake Erie but is a tributary to Elk Creek. The Brandy Run gage is approximately 5 mi from where Elk Creek discharges into Lake Erie, which is approximately 15 mi from the Presque Isle beaches. Current speed and direction (instantaneous) were from the National Oceanic and Atmospheric Administration (NOAA) Great Lakes Coastal Forecasting System (Gregory Lang, National Oceanic and Atmospheric Administration, written commun., 2005). Wind speed, wind direction, and calculated wave height (all instantaneous) were from a weather station buoy (45005) owned and maintained by the NOAA National Data Buoy Center (National Oceanic and Atmospheric Administration, 2005a). Rainfall data were from a weather station at the Erie International Airport, Erie, Pa. (National Oceanic and Atmospheric Administration, 2005b).

Water samples were analyzed for *E. coli* bacteria using modified mTEC membrane-filtration techniques (U.S. Environmental Protection Agency, 2002) (fig. 2). Water samples were processed by ECHD staff in their laboratory within 6 hours of sample collection.

Statistical Analysis

Scientists from the USGS Pennsylvania Water Science Center compiled and statistically analyzed all water-quality and ancillary data. The nonparametric Spearman's rho statistical test and x/y scatterplots were used as general screening tools to determine if correlations were present between *E. coli* and other continuous water-quality or environmental variables. To compute Spearman's correlation coefficients on the dataset for this study that included multiple detection limits for *E. coli* concentrations (two detection limits: less than (<) 4 and <10 col/100 mL), the data were censored at the higher detection limit (<10 col/100 mL). After the data were censored to a common detection limit, the data for each variable were ranked separately and Spearman's correlation coefficient (nonparametric test) was computed by calculating Pearson's correlation coefficient (parametric test) on the ranks (Helsel, 2005). Correlations at the 95-percent confidence interval (probabilities less than 0.05) as determined by Spearman's rho were considered statistically significant.

The Kruskal-Wallis test was used as a screening tool to compare three or more groups of data. The Kruskal-Wallis test is a rank-transform test. In this study, the $\log_{10}$ *E. coli* data were ranked from lowest to highest; all censored data (all observations <10 col/100 mL) were tied at the lowest rank. The test compared the ranked $\log_{10}$ *E. coli* data to environmental categorical data—estimated wave height, wind direction, and current direction—to test for statistical differences in the medians between the groups of data. If results of the Kruskal-Wallis test showed differences, the Tukey-Kramer multiple-comparison test was performed to determine which groups differed from each other. The groups with the highest medians were assigned a letter "A," the groups with the next highest medians were assigned a letter "B" or an "AB" combination, and so on. Any groups assigned the same letter (or combination of letters) designations were not statistically different from each other. For example, a group assigned a letter "A" would not be statistically different from another group assigned a letter "A" or a group assigned a combination of letters "AB." On the other hand, a group assigned a letter "A" would be statistically different from another group assigned any other letter or combination of letters that does not include the letter "A".

Tobit regression was used in the development of models to predict *E. coli* because of the capability it has to handle multiple-censored data that linear regression does not. Tobit regressions were performed using the LIFEREG procedure with SAS statistical software (SAS Institute, 1990). According to Allison (1995), the LIFEREG procedure uses maximum-likelihood

Photo by Donald Williams, U.S. Geological Survey

Figure 2. Erie County Health Department staff processing *Escherichia coli* samples using membrane-filtration techniques.

methodology to produce estimates for parametric regression models. The LIFEREG procedure has the capability to incorporate different types of censored data into the regression analyses—right-censored, left-censored, and interval-censored data. For this study, the data were left-censored and interval-censored. Left-censored data are data less than a detection limit (the nondetects). For example, the *E. coli* concentration reported for the sample collected at Beach 2-West on June 26, 2005, was <4 col/100 mL. It can be stated definitively that the concentration was "less than" 4 col/100 mL, but the actual concentration could be anywhere from 3.999999 to 0. In contrast, interval-censored data would be data from water samples that had *E. coli* concentrations between the two detections limits (<4 and <10 col/100 mL). For example, on June 26, 2005, the *E. coli* concentration reported for the Beach 2-East sample was 8 col/100 mL. This is a reported concentration that falls between the <4 and <10 detection limits. By using interval censoring, the data between the two detections limits for this study are not lumped in with the nondetects or left-censored data (samples reported as either <4 or <10 col/100 mL). Instead, the data are assigned to an interval higher in value than the nondetects. Although a definitive value is not assigned to the data in either group (left-censored or interval-censored), hierarchy of the data is preserved.

Regression models were developed using the 2004 data, the 2005 data, and the combined 2004–2005 data. All variables found to be statistically related to *E. coli* concentrations using Spearman's correlation or the Kruskal-Wallis test were used in the initial tobit regression model. Model-variable estimates having probabilities greater than 0.05 were eliminated one by one using backwards elimination techniques—in each iteration, the variable with the highest probability was eliminated, and the model was run again. When all model-variable estimates were statistically significant at the 95-percent confidence level (probabilities of 0.05 or less), those variables were looked at individually in models with a single variable. The variable with the lowest probability was analyzed first; then one by one, variables were added back into the models by forward selection techniques to see if the same final model was produced as with backwards elimination techniques, ultimately ensuring the best model was selected.

Tobit regression models were evaluated using likelihood-ratio chi-square statistics and generalized R^2 values. Models were chosen such that the combination and number of variables explaining the concentrations of *E. coli* in the model were significantly better, as determined using likelihood-ratio chi-square statistics, than simpler, nested models with fewer variables. For example, to test whether or not $model_{simple}$ (with one variable) is significantly better than $model_{complex}$ (with three variables), with $model_{simple}$ nested within $model_{complex}$, the likelihood-ratio chi-square statistic (G^2) can be calculated as twice the positive difference in the log likelihoods of the two models (eq. 1).

$$G^2 = 2\,(LL_{simple} - LL_{complex}), \quad (1)$$

where LL_{simple} is the log likelihood of the simpler, nested model ($model_{simple}$), and
$LL_{complex}$ is the log likelihood of the model with more variables than the nested model ($model_{complex}$).

Once G^2 was calculated, a critical value for chi-square at the 95-percent confidence level (probability = 0.05) was obtained from standard chi-square distribution tables. The critical chi-square value was obtained for a distribution with degrees of freedom equal to the difference in the number of variables in the complex model compared to the number of variables in the simpler, nested model. If G^2 was greater than the critical value, the model with more variables was significantly better than the simpler model.

Models were also selected to maximize the generalized R^2 (generalized coefficient of determination). Unlike the R^2 for linear regression models, the generalized R^2 value for the tobit regression model can not be interpreted as the proportion of variation in the dependent variable (in this case, $\log_{10}$ *E. coli* concentrations) that can be explained by the explanatory variables in the model; rather, it is a number between 0 and 1 that is larger when the explanatory variables are more strongly associated with the dependent variable (Allison, 1995). A value for generalized R^2 was calculated using the following equation from Allison (1995)(eq. 2).

$$\text{Generalized } R^2 = 1 - \exp\{-G^2 / n\}, \quad (2)$$

where G^2 is the likelihood-ratio chi-square statistic, and
n is the number of samples.

Evaluation criteria for the best models also included ensuring model-variable estimates were statistically significant at the 95-percent confidence level (probabilities = 0.05 or less) and graphically evaluating model fit by creating plots of the residuals from the tobit regression models.

A threshold probability of exceeding Pennsylvania's single-sample bathing-water standard for *E. coli* (235 col/100 mL) was determined using the following methodology. First, the tobit regression model computed a predicted *E. coli* concentration for each observation used in the regression. Next, these predicted values for *E. coli* were used in the following equation (eq. 3) modified from Allison (1995, p. 264) to determine the probabilities (exceedence probabilities) that the *E. coli* standard would be exceeded by these predicted concentrations.

$$\text{Prob} = 1 - \text{Probnorm}\left(\frac{\log(235) - \text{lp}}{\text{sep}}\right) \quad (3)$$

where

Probnorm is a SAS function (SAS Institute, 1990) that computes the probability that an observation from a standard normal distribution falls below the given value for x—in this case x = ((log(235)-lp)/sep),

lp is the predicted *E. coli* concentration, and

sep is the standard error of the prediction.

Finally, a scatterplot of actual *E. coli* concentrations and predicted exceedence probabilities for the corresponding predicted *E. coli* concentrations was used to select a threshold probability that maximized the correct number of exceedences and nonexceedences and minimized the number of false positive and false negative results. Determinations of whether beach closings/advisories should be issued can be made if values exceed the established threshold probability.

Quality Control

Quality-control measures were practiced in this study to ensure data quality met project objectives. Quality control in the field included the collection of a field blank approximately once each month during the 2004 and 2005 recreational seasons to ensure sterile sampling techniques, to maintain sterile conditions, and to assess any field contamination of samples. No bacteria colonies were detected in any of the field blanks. Another quality-control measure in the field was the collection of duplicate field measurements for all field characteristics (turbidity, specific conductance, pH, dissolved oxygen, total dissolved solids, and water temperature). Measurements that did not agree within 10 percent were repeated and an average (of samples that agreed within 10 percent) was used as the final measurement.

Quality-control measures in the ECHD laboratory included processing blank samples using sterile buffered water and processing duplicate samples. Filter blanks were processed with every sample run to ensure sterile equipment and supplies were being used. Procedure blanks also were processed with every sample run to ensure that proper rinsing techniques were being used such that there was no carryover between samples. No bacteria colonies were detected in any of the filter blanks or procedure blanks. Duplicate water samples of varying volumes (25 mL, 10 mL, or 1 mL) were processed and analyzed for *E. coli* by ECHD on every sample.

Replicate samples were processed and analyzed by the USGS Ohio Water Microbiology Laboratory (OWML) scientists on approximately 10 percent of all *E. coli* samples collected to test analytical variability between labs. A comparison of ECHD routine sample data to OWML replicate sample data is shown in table 1. One thing to note is sample-processing times. All water samples were processed by the ECHD within 6 hours of sample collection, whereas water samples were processed by the OWML within 24 hours of sample collection. The results from both laboratories showed agreement in when sample concentrations would be above or below the standard (there were no instances where one laboratory reported bacteria concentrations below the detection limit and the other laboratory reported bacteria concentrations above the detection limit, or vice versa). The results were similar between laboratories when concentrations reported from one or both laboratories were within the ideal colony count range (20 – 80 col/100 mL) or higher (greater than 80 col/100 mL). Replicate samples agreed within 20 percent for 7 out of 12 sample pairs that had concentrations of 20 col/100 mL or higher reported from one or both laboratories. Replicate samples agreed within 5 percent for the two replicate sample pairs that had concentrations greater than 80 col/100 mL. Results were least similar between laboratories when concentrations were near the detection limit. In one instance, *E. coli* bacteria was detected in the ECHD sample and not detected in the OWML sample. This could possibly be explained by the difference in sample volumes analyzed by the two laboratories or the lag time in processing by the OWML. For example, ECHD analyzed a 1-mL sample volume and a 10-mL sample volume at Beach 2–West on July 12, 2004. The 1-mL sample had zero colonies and the 10-mL sample had 1 colony. The number of colonies per 100 mL was calculated to be 10. The OWML analyzed a 100-mL sample volume on July 13, 2004, from a sample collected on July 12, 2004, and did not find any bacteria. Thus, the concentration is reported as

Table 1. Erie County Health Department routine sample data and U.S. Geological Survey Ohio Water Microbiology Laboratory replicate sample data.

[E. coli, *Escherichia coli*; ECHD, Erie County Health Department; OWML, Ohio Water Microbiology Laboratory; <, less than]

Location	Sample date	*E. coli*, colonies per 100 milliliters	
		ECHD routine sample	OWML replicate sample
Beach 2-West	7/12/04	10	<1
Beach 2-East	7/12/04	20	12
Beach 2-East	8/9/04	10	5
Beach 2-West	8/23/04	40	30
Beach 2-East	8/24/04	<10	7
Beach 2-West	6/29/05	12	20
Beach 2-East	7/20/05	32	21
Beach 2-West	7/27/05	110	110
Beach 2-East	8/3/05	25	14
Beach 2-West	8/17/05	4	9
Beach 2-East	8/24/05	4	2
Beach 2-East	8/31/05	680	940

<1 col/100 mL. In another sample, collected on August 24, 2004, at Beach 2-East, *E. coli* was detected by the OWML, but not detected by the ECHD. Because of the sample volume analyzed by the ECHD (highest sample volume analyzed was 10 mL on August 24, 2004), however, the concentration was reported as <10 col/100 mL. The OWML detected 7 colonies in the 100-mL sample analyzed. Thus, the concentrations reported by both agencies were in agreement. The results of this study are similar to those in studies conducted in 2001 at 24 sites across the United States where it was found that about 70 percent of the surface-water samples collected showed no significant difference between concentrations of *E. coli* processed within 8 hours and 48 hours of sample collection (Pope and others, 2003).

Another quality-control measure in the ECHD lab was the processing of positive control cultures obtained from the OWML. Positive control cultures were run once each recreational season (2004 and 2005) for modified mTEC agar to test the integrity of the agar and ensure proper execution of sampling methodology. ECHD and OWML results agreed within 10 percent for both control cultures. A verification step was also done by the ECHD using Enterotube II verification system for Enterobacteriaceae (Becton Dickinson and Company, 1999) to ensure the mTEC media was enumerating true *E. coli* colonies. Enterotube tests were done at least twice a month during the recreational seasons on routine samples and on control cultures obtained from the OWML. All results using the Enterotube II system were positive for *E. coli*. Negative controls also were run once a month on routine samples using Enterotube II, and all results were negative for *E. coli*.

Monitoring *Escherichia coli*

The bacteriological quality of waters at Presque Isle Beach 2 in Erie, Pa., was generally good as determined by analyzing *E. coli* concentrations in 178 water samples collected during the 2004 and 2005 recreational seasons. *E. coli* was not detected in 35 of 178 samples (20 percent)—20 samples at a detection limit of <4 col/100 mL and 15 samples at a detection limit of <10 col/100 mL. The maximum concentration of *E. coli* was 845 col/100 mL following a storm in late August 2005. The single-sample bathing-water standard of 235 col/100 mL was exceeded 5 of 31 days sampled in 2004 and 7 of 57 days sampled in 2005 (7 percent of the total days sampled).

Continuous Variable Analysis

Statistical tests were used to determine if correlations exist between *E. coli* concentrations in water and selected water-quality and environmental variables. Spearman's rho correlation coefficient was used to calculate correlations between $\log_{10}$ *E. coli* and other continuous variables (table 2).

Some of the variables used in the statistical analysis (table 2) need to be defined. The "birds" variable included the number of birds on the beach, in the water, and in the air at the time of sampling. "Rain24," "rain48," and "rain72" were the amounts of rainfall that fell in the 0 to 24-hour, 24- to 48-hour, and 48- to 72-hour period, respectively, preceding collection of the sample. "Rain weight" was the sum of weighted rainfall amounts from the 72-hour period preceding sampling, giving the most weight to amounts closest to sampling (eq. 4).

$$\text{"Rain weight"} = (3 \times \text{"rain24"}) + (2 \times \text{"rain48"}) + (1 \times \text{"rain72"}) \quad (4)$$

"Q_inst log" is the $\log_{10}$ of the instantaneous streamflow from the USGS streamflow-gaging station at Brandy Run near Girard, Pa. (04213075). A streamflow value is determined every 30 minutes and the value closest to the time of sampling was used for the "Q_inst log" variable. "Q_prev log" is the $\log_{10}$ of the mean daily streamflow at Brandy Run near Girard, Pa., the day prior to sampling.

Statistically significant correlations (95-percent confidence level) were found between $\log_{10}$ *E. coli* and turbidity log, wave height (calculated), wind speed, and current speed in all cases (2004 data, 2005 data, and combined 2004–2005 data) (table 2). The strongest correlations were between $\log_{10}$ *E. coli* and turbidity log in the 2005 dataset and the combined 2004–2005 dataset; Spearman's rho correlation coefficients were 0.715 and 0.662, respectively. Weaker correlations exist between $\log_{10}$ *E. coli* concentrations and other variables as indicated by smaller values of Spearman's rho. For example, statistically significant correlations were found between $\log_{10}$ *E. coli* and rain24 in the 2005 dataset and combined 2004–2005 dataset; Spearman's rho correlation coefficients were 0.450 and 0.369, respectively. None of the rain variables were statistically correlated to $\log_{10}$ *E. coli* in the 2004 dataset. Another example is streamflow that was weakly correlated to *E. coli* in the 2004 dataset (Q_prev log only with Spearman's rho of 0.273) and in the combined 2004–2005 dataset (Q_inst log and Q_prev log with Spearman's rho values of 0.228 and 0.212, respectively). None of the streamflow variables were statistically correlated to $\log_{10}$ *E. coli* in the 2005 dataset. In general, fewer and weaker correlations were observed in the 2004 dataset, which had the smallest number of samples, compared to the 2005 dataset. Variables statistically correlated to *E. coli*, as indicated by boldfaced type in table 2, were used in model development.

Categorical Variable Analysis

The Kruskal-Wallis test and Tukey-Kramer multiple-comparison test were used to determine if there were any relations between *E. coli* concentrations and categorical environmental variables. The categorical variables were wind direction, current direction, and estimated wave height. Statistical relations were found between $\log_{10}$ *E. coli* and wind direction. The $\log_{10}$ *E. coli* data were grouped according to wind direction,

Table 2. Summary of Spearman's rho correlations between *Escherichia coli* concentrations in water and selected water-quality or environmental variables at Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania, 2004–2005.

[variable names that include "log" are $\log_{10}$ of the variable; **bold type** denotes variables statistically correlated to $\log_{10}$ *Escherichia coli* at the 95-percent confidence level (probabilities of 0.05 or less); <, less than]

Variable	Spearman's rho correlation coefficient (probability) Number of samples		
	2004	2005	2004–2005
Water temperature	0.210 (.113) 58	0.123 (.187) 116	0.085 (.264) 174
Dissolved oxygen	.010 (.9416) 58	-.108 (.2659) 108	-.041 (.6024) 166
Turbidity log	.462 **(.0003)** 58	.715 **(<.0001)** 108	.662 **(<.0001)** 166
Birds	.142 (.2865) 58	-.033 (.7429) 102	.060 (.4479) 160
Wave height (calculated)[1]	.351 **(.0051)** 62	.407 **(<.0001)** 116	.398 **(<.0001)** 178
Rain24[2]	.165 (.2004) 62	.450 **(<.0001)** 116	.369 **(<.0001)** 178
Rain48[2]	-.0380 (.7694) 62	.105 (.2627) 116	.077 (.3066) 178
Rain72[2]	-.24373 (.0563) 62	.05164 (.5820) 116	-.040 (.5967) 178
Rain weight[3]	-.084 (.5148) 62	.360 **(<.0001)** 116	.245 **(.0010)** 178
Q_inst log[4]	.208 (.1040) 62	.150 (.1080) 116	.228 **(.0022)** 178
Q_prev log[5]	.273 **(.0320)** 62	.135 (.1484) 116	.212 **(.0045)** 178
Wind speed[1]	.297 **(.0191)** 62	.367 **(<.0001)** 112	.374 **(<.0001)** 174
Current speed[6]	.477 **(<.0001)** 62	.407 **(<.0001)** 116	.343 **(<.0001)** 178

[1]Wind speed (instantaneous) and wave height (calculated) were from a weather-station buoy (45005) owned and maintained by the National Oceanic and Atmospheric Administration National Data Buoy Center (National Oceanic and Atmospheric Administration, 2005a).

[2]Rain24, rain48, and rain72 were the amounts of rain that fell at Erie International Airport, Erie, Pa., in the 0-24, 24-48, and 48-72 hour period, respectively, before the sample was collected at Presque Isle Beach 2.

[3]Rain weight is the sum of weighted rainfall amounts from the 24-hour periods 1, 2, and 3 days prior to sampling giving the most weight to amounts closest to sampling.

[4] Q_inst log is the $\log_{10}$ of the instantaneous streamflow measurement (nearest 30-minute measurement to time of sampling) from the USGS streamflow-gaging station at Brandy Run near Girard, Pa. (04213075).

[5]Q_prev log is the $\log_{10}$ of the mean daily streamflow at Brandy Run near Girard, Pa., the day prior to sampling.

[6]Current speed (instantaneous) was from the National Oceanic and Atmospheric Administration Great Lakes Coastal Forecasting System (Gregory Lang, National Oceanic and Atmospheric Administration, written commun., 2005).

with fewer than five observations for any particular wind direction being combined with a logically similar wind direction (east and northeast, for example). Significantly higher median concentrations of $\log_{10}$ *E. coli* were observed when north (north, northeast, northwest) or southwest winds were blowing than when winds were from the south, southeast, or east (fig. 3). Winds were from the north (north, northeast, or northwest) or southwest when concentrations of *E. coli* exceeded the single-sample bathing-water standard of 235 col/100 mL (fig. 3). High concentrations of *E. coli* were usually observed when winds were parallel with the shore line (north, northeast, or southwest), keeping the contamination in the swimming area. This finding implies that the source of bacterial contamination is near-shore.

Results from the Kruskal-Wallis test indicate no statistically significant differences between $\log_{10}$ *E. coli* concentrations and current direction categories. Similar to wind direction, *E. coli* data were grouped according to current direction with fewer than five observations for any particular current direction being combined with a logically similar current direction. Even though most currents (approximately 70 percent) were from the northeast or east-northeast (northeast currents would bring contamination into the beach area), the median concentrations of *E. coli* when currents were from those directions were similar to when they were from any other direction. Current direction, therefore, was not used in model development.

Wave-height categories estimated by field personnel at the time of sampling were analyzed using the Kruskal-Wallis test and Tukey-Kramer multiple-comparison test to determine if there were relations to *E. coli* concentrations. Wave heights were categorized into one of five groups by visual inspection (0 to 2 ft, 1 to 3 ft, 2 to 4 ft, 3 to 5 ft, and 4 to 6 ft) and *E. coli* concentrations were grouped by estimated wave-height category. No estimated wave heights fell in the 4- to 6-ft category during the study period. Only six observations had an estimated wave height of from 3 to 5 ft. Significantly higher median concentrations of $\log_{10}$ *E. coli* concentrations were observed when waves were from 3 to 5 ft than when estimated wave height was from 0 to 2 and 1 to 3 ft (fig. 4). *E. coli* concentrations did not exceed the standard when estimated wave heights were from 2 to 4 ft, but the standard was exceeded when wave heights were in the other observed groups (fig. 4). Past studies have found that median *E. coli* concentrations generally increased as wave height increased (Francy and Darner, 1998; Francy and others 2003).

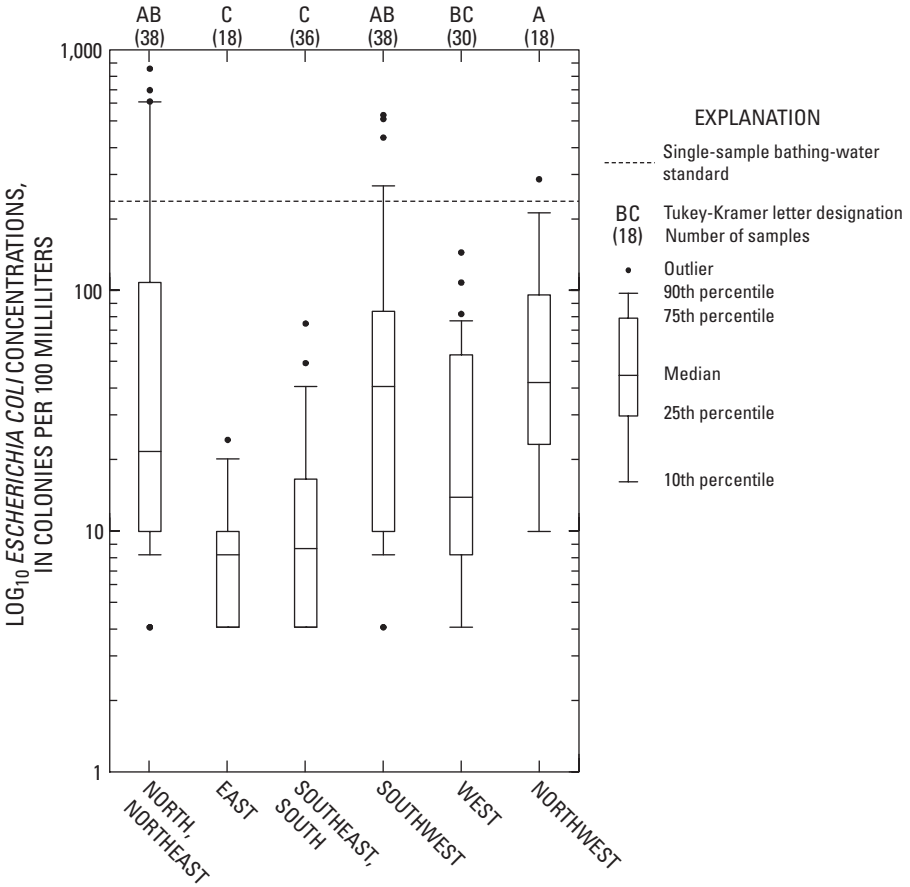

Figure 3. Distribution of $\log_{10}$ *Escherichia coli* concentrations by wind direction, Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania.

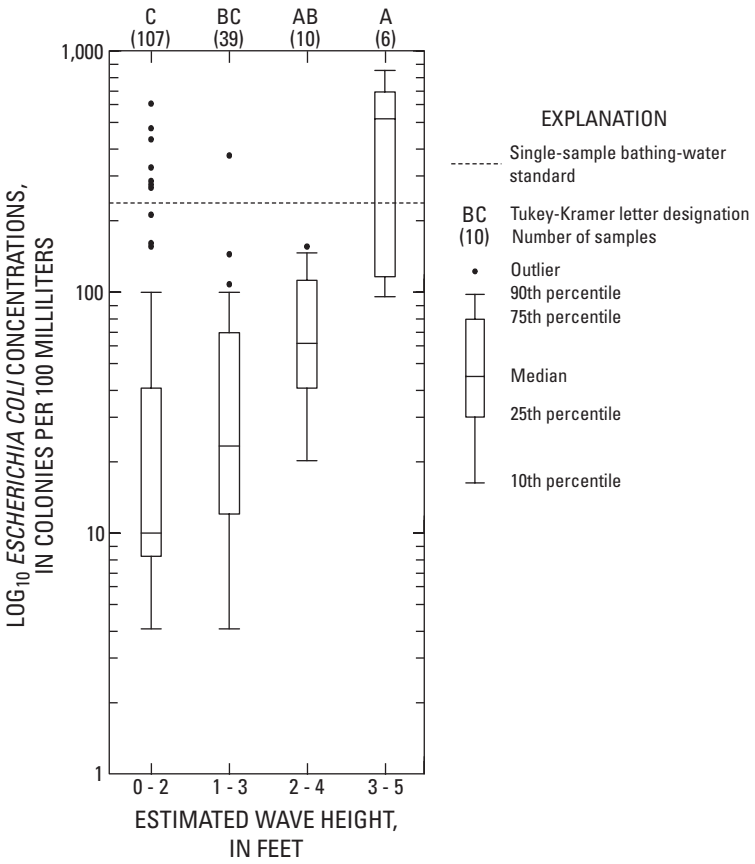

Figure 4. Distribution of $\log_{10}$ *Escherichia coli* concentrations by estimated wave height, Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania.

Modeling to Predict *Escherichia coli*

The water-quality and environmental variables related to *E. coli*, as determined by Spearman's test or the Kruskal-Wallis and Tukey-Kramer tests, were considered in the development of a model using tobit-regression-analysis techniques (table 2). Similar to previous studies (Francy and others, 2003), the model was used to predict the probability that the *E. coli* single-sample bathing-water standard of 235 col/100 mL would be exceeded. The predicted exceedence probabilities produced by the model were used to determine a threshold probability, which minimized the number of false positives and false negatives and maximized the number of correct exceedences and correct non-exceedences for *E. coli* concentrations.

Models were developed for each dataset—2004 only, 2005 only, and combined 2004–2005. Explanatory variables not included in the final models were water temperature, streamflow, wind speed, and current speed, because simulation results indicated these variables did not meet significance criteria at the 95-percent confidence level (probabilities were greater than 0.05). A summary of model output for the "best" models developed for each dataset analyzed is presented in table 3. The best 2004 model had four explanatory variables—turbidity log, rain weight, wave height (calculated), and wind direction. The best 2005 model had only two explanatory variables, turbidity log and wind direction. The fact that these same two variables are in both models seems to indicate that much of the variability in $\log_{10}$ *E. coli* concentrations at Presque Isle Beach 2 can be explained by them. Running the combined 2004–2005 dataset through the tobit regression analysis yields the same four explanatory variables as the 2004 model. Just as the 2004 model indicated, adding rain weight and wave height (calculated) yields a significantly better model for predicting *E. coli* exceedence probabilities when data from the 2 years are combined (table 3).

A scatterplot of actual *E. coli* concentrations and predicted exceedence probabilities for the corresponding predicted *E. coli* concentrations helped determine a threshold probability. The combined 2004–2005 results are shown in figure 5. The plot is divided into four sections using the *E. coli* single-sample bathing-water standard to divide the plot vertically (log of 235 col/100 mL in figure 5) and the chosen threshold probability (27 in figure 5) to divide the plot horizontally. Moving the threshold-probability line up or down on the plot changes the number of observations in each of the four sections. The goal in establishing a threshold probability is to maximize the number of correct observations, or responses, showing *E. coli* concentrations (1) above the *E. coli* standard of 235 col/100 mL and having a predicted probability of exceedence above the estab-

Table 3. Summary of tobit regression model explanatory variables, statistics, and performance with selected threshold probabilities in predicting exceedences to the *Escherichia coli* single-sample bathing-water standard of 235 colonies per 100 milliliters for Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania, 2004–2005.

[Generalized R^2, a number between 0 and 1 that is larger when the explanatory variables are more strongly associated with the dependent variable (Allison, 1995); log likelihood, a statistic produced in tobit regression that is used in model goodness-of-fit tests; threshold probability, exceedence probability chosen for a model on the basis of the *E. coli* standard of 235 colonies per 100 milliliters being met or exceeded; **bold type** highlights the explanatory variables in the final models]

Summary	2004 model	2005 model	Combined 2004–2005 model
Number of observations	58	108	166
Generalized R^2	0.54	0.71	0.64
Log likelihood	-73.33	-143.30	-224.31
Explanatory variables in model	**turbidity log rain weight wave height wind direction**	**turbidity log wind direction**	**turbidity log rain weight wave height wind direction**
Threshold probability	28	28	27
Number of correct exceedences	1	5	6
Number of correct nonexceedences	52	97	150
Number of false positives	1	4	4
Number of false negatives	4	2	6

lished threshold probability (correct exceedence), and (2) meeting the standard by being less than 235 col/100 mL and having a predicted probability of exceedence below the established threshold probability (correct nonexceedence). Minimizing the number of incorrect responses is another goal in establishing the threshold probability. False negatives result when the standard was, in fact, exceeded but the predicted probability of exceedence was below the established threshold probability. False positives result when the standard was not exceeded but the predicted probability of exceedence was above the established threshold probability.

The threshold probabilities for each model were very similar—28 for the 2004 and 2005 models and 27 for the combined 2004–2005 model. Overall, the performance of the 2004, 2005, and combined 2004–2005 models was good with 91 (53 of 58 samples), 94 (102 of 108 samples), and 94 (156 of 166 samples) percent correct predictions, respectively, indicating when the *E. coli* standard of 235 col/100 mL would be met or exceeded. Further breakdown of the performance of each model shows the 2004 and combined 2004–2005 models had more false negatives than the 2005 model. Of the five samples that exceeded the standard in the 2004 model, only one of five (20 percent) was correctly predicted as an exceedence. However, the 2004 model correctly predicted when the *E. coli* standard would not be exceeded in 52 of 53 samples (98 percent). The 2005 model did a much better job of predicting when the *E. coli* standard would be exceeded with five of seven samples (71 percent) correctly predicted as exceedences. The 2005 model also did a good job of predicting when *E. coli* concentrations would not exceed the standard with 97 of 101 (96 percent) correct nonexceedances. The combined 2004–2005 dataset produced a significantly better model with the explanatory variables shown in table 3 (than with turbidity log and wind speed alone) that correctly predicted when the *E. coli* standard would not be exceeded in 150 of 154 samples (97 percent) and correctly predicted when the standard would be exceeded in 6 of 12 samples (50 percent) (fig. 5). The following equations predict *E. coli* concentrations from the models developed with the 2004, 2005, and combined 2004–2005 datasets.

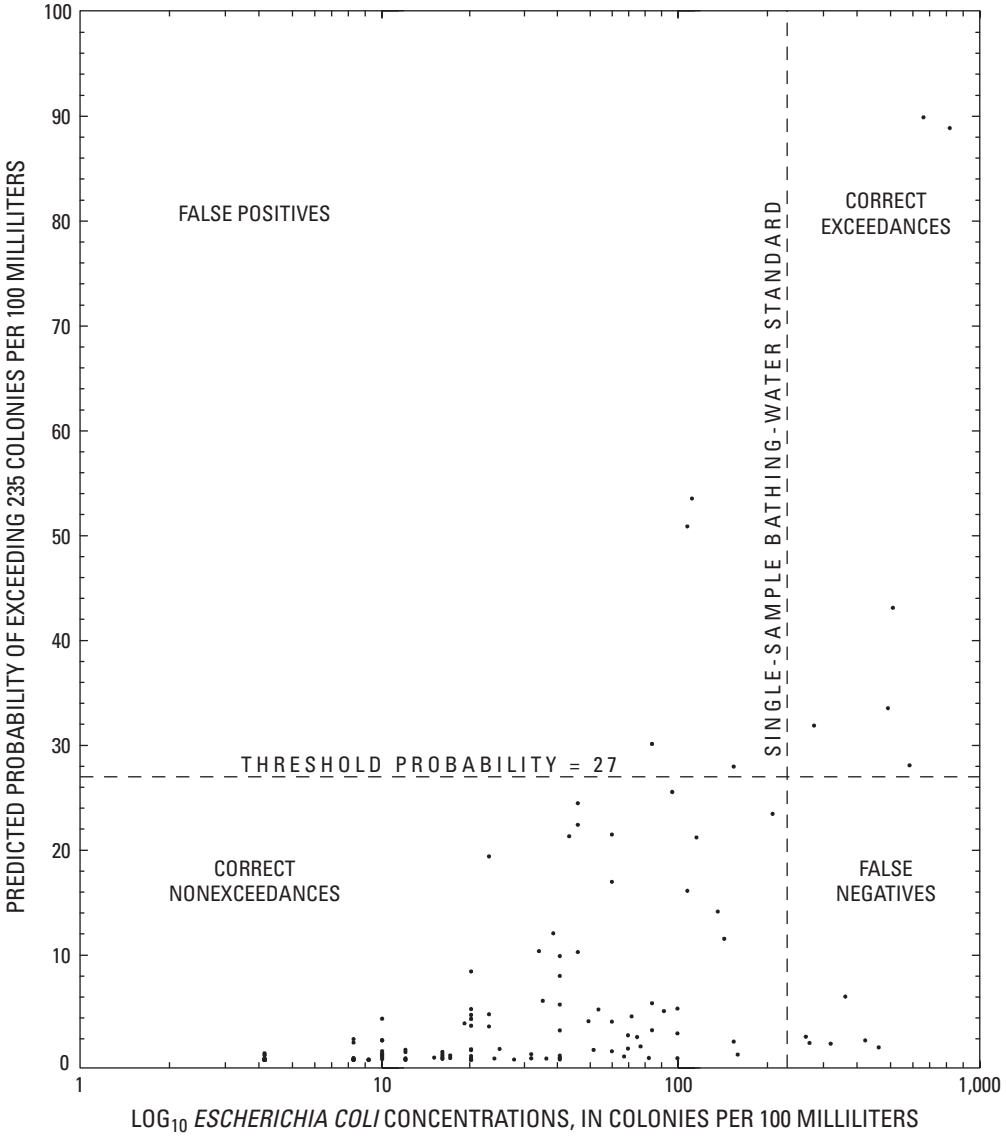

Figure 5. Log_{10} *Escherichia coli* concentrations and probability of exceeding 235 colonies per 100 milliliters to achieve optimum threshold probability for the combined 2004–2005 model for Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania.
[2004–2005 model explanatory variables for determining bacteria concentrations include turbidity log, rain weight, wave height (calculated), and wind direction].

2004 dataset

$$\text{Log}_{10}\, E.\ coli = 0.9551 + 0.434(\text{turbidity log}) + 0.147(\text{rain weight}) + 0.829(\text{wave height*}) - 0.208(\text{wind dir}) \quad (5)$$

$$\text{Or } E.\ coli = 9.018 + 2.716^{\text{turbidity log}} + 1.403^{\text{rain weight}} + 6.745^{\text{wave height*}} + 0.619^{\text{wind dir}} \quad (6)$$

2005 dataset

$$\text{Log}_{10}\, E.\ coli = 1.208 + 0.732(\text{turbidity log}) - 0.223(\text{wind dir}) \quad (7)$$

$$\text{Or } E.\ coli = 16.144 + 5.395^{\text{turbidity log}} + 0.598^{\text{wind dir}} \quad (8)$$

Combined 2004–2005 dataset

$$\text{Log}_{10}\, E.\ coli = 1.087 + 0.520(\text{turbidity log}) + 0.073(\text{rain weight}) + 0.365(\text{wave height*}) - 0.179(\text{wind dir}) \quad (9)$$

$$\text{Or} \quad E.\ coli = 12.218 + 3.311^{\text{turbidity log}} + 1.183^{\text{rain weight}} + 2.317^{\text{wave height*}} + 0.662^{\text{wind dir}} \quad (10)$$

where wave height* is the wave height (calculated) variable.

The combined 2004–2005 model is good at predicting when bacteria concentrations will be below the standard but not as good at predicting when bacteria concentrations will be above the standard. This is likely because Presque Isle Beach 2 is generally clean and concentrations typically do not exceed the standard. Data from additional sampling (more storm samples, for example) would provide information that might improve the model and would help better characterize the bacteriological quality of the beach—in particular, the predictive capability of the model in determining when the standard would be exceeded. For determining beach closures, the model is one tool that could be used in the decision-making process.

Summary and Conclusions

This report describes a study done during the 2004 and 2005 recreational seasons by the U.S. Geological Survey, in cooperation with the Erie County Health Department (ECHD), to develop regression models designed to predict *E. coli* concentrations used to determine the probability of exceeding the *E. coli* single-sample bathing-water standard at Presque Isle Beach 2 in Erie, Pa. Water-quality and environmental variables were compiled and analyzed to determine any statistical relation to *E. coli* concentrations. The results of this study provide a supplemental method of determining the recreational water quality of Presque Isle Beach 2 that will aid beach managers in more rapidly determining when waters are not safe for recreational use and when beach advisories or closings need to be issued.

Correlation tests were conducted to determine the continuous variables that were related to $\log_{10}$ *E. coli* concentrations. In each of the datasets analyzed (2004 only, 2005 only, combined 2004–2005), turbidity log, wave height (calculated), wind speed, current speed, and wind direction were correlated to $\log_{10}$ *E. coli* concentrations. The strongest correlations were between $\log_{10}$ *E. coli* and turbidity log in the 2005 dataset and the combined 2004–2005 dataset with Spearman's rho correlation coefficients of 0.715 and 0.662, respectively. Weaker correlations (statistically significant at the 95-percent confidence level) exist between $\log_{10}$ *E. coli* concentrations and the streamflow variables and the rain variables as indicated by smaller values of Spearman's rho. None of the rain variables were statistically correlated to $\log_{10}$ *E. coli* in the 2004 dataset, and none of the streamflow variables were statistically correlated to $\log_{10}$ *E. coli* in the 2005 dataset. Overall, fewer and weaker correlations were observed in the 2004 dataset, which had the smallest number of samples, compared to the 2005 dataset. Variables statistically correlated to *E. coli* were considered in model development.

The Kruskal-Wallis test and Tukey-Kramer multiple-comparison test were used to determine if there were relations between *E. coli* concentrations and categorical environmental variables. No statistically significant relations were observed between $\log_{10}$ *E. coli* and current direction. Statistical relations were observed between $\log_{10}$ *E. coli* and wind direction. Significantly higher concentrations of $\log_{10}$ *E. coli* were observed when winds were from the north (north, northeast, northwest) or southwest compared to when winds were from the south, southeast, or east. The single-sample bathing-water standard was exceeded when winds were from the north, northeast, northwest, and southwest. Statistical relations also were observed between $\log_{10}$ *E. coli* and estimated wave height. Wave heights were categorized into one of five groups by visual inspection—0 to 2 ft, 1 to 3 ft, 2 to 4 ft, 3 to 5 ft, and 4 to 6 ft.

Log_{10} *E. coli* concentrations were significantly higher when the estimated wave height was from 3 to 5 ft than when estimated wave height was from 0 to 2 and 1 to 3 ft. Log_{10} *E. coli* concentrations exceeded the standard when estimated wave height was from 0 to 2 ft, 1 to 3 ft, and 3 to 5 ft. No estimated wave heights were in the 4- to 6-ft category during the study period. Log_{10} *E. coli* concentrations were statistically lower when estimated wave height was from 0 to 2 ft than when estimated wave height was from 2 to 4 and 3 to 5 ft. In general, median *E. coli* concentrations increased as wave heights increased.

Models were developed for Presque Isle Beach 2 using 2004 data, 2005 data, and combined 2004–2005 data. All water-quality and environmental variables related to log_{10} *E. coli* concentrations were considered in the tobit regression models. A model was developed for the 2004 data that included the explanatory variables turbidity log, rain weight, wave height (calculated), and wind direction. The model developed for the 2005 data had two explanatory variables, turbidity log and wind direction. Combining the 2004 and 2005 data produced a model with the same explanatory variables as the 2004 model. Just as the 2004 model indicated, adding rain weight and wave height (calculated) yielded a significantly better model for predicting *E. coli* exceedence probabilities when data from the 2 years are combined.

Further study could focus on improving the predictive ability of the model by adding data that were not collected as part of this study, such as information on combined-sewer overflows, outfall discharges, or tributary inputs. A study is underway between ECHD, Mercyhurst College, and the Regional Science Consortium at the Tom Ridge Center at Presque Isle State Park to determine the sources of *E. coli* contamination to Presque Isle beaches that will likely include collection of data to analyze effects of these additional factors on *E. coli* concentrations in Presque Isle beaches. Further study could also focus on collecting more samples following storms; data from storm samples likely would improve the predictive capability of the model in determining when the standard would be exceeded.

Acknowledgments

The author thanks Scott White from the Erie County Health Department for his assistance throughout all phases of the project. The author also acknowledges the assistance of others who helped with data collection and processing— Niel Bullock and Michael Rinkevich of the Erie County Health Department, Richard Bierbower, Raymond Bierbower, Seth Wilmore, and MuraliKumar Katta-Muddanna of Presque Isle State Park, and Donald Williams, Linda Zarr, Alyshia Inks, Eric Celebrezze, and Megan Rogers of the U.S. Geological Survey. A special thanks to Edward Koerkle of the U.S. Geological Survey for the statistical expertise and insight he provided during the data-analysis phase of the project.

References Cited

Allison, P.D., 1995, Survival analysis using SAS—A practical guide: Cary, N.C., SAS Institute, Inc., 292 p.

Becton Dickinson and Company, 1999, Enterotube II Identification System for Enterobacteriaceae: Sparks, Md., 13-06-43128-002, 7 p.

Francy, D.S., and Darner, R.A., 1998, Factors affecting *Escherichia coli* concentrations at Lake Erie public bathing beaches: U.S. Geological Survey Water-Resources Investigation Report 98-4241, 41 p.

Francy, D.S., and Darner, R.A., 2002, Forecasting bacteria levels at bathing beaches in Ohio: U.S. Geological Survey Fact Sheet FS-132-02, 4 p.

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.

Helsel, D.R., 2005, Nondetects and data analysis: New Jersey, John Wiley and Sons, Inc., 250 p.

Hydrolab Corporation, 2002, Hydrolab Quanta water quality monitoring system operating manual, February 2002: Austin, Tex., Revision C, 42 p.

Kuntz, J.E., and Murray R., 2000, Predictability of swimming prohibitions by observational parameters: Stamford, Connecticut Department of Health and Social Services, Laboratory Division, assessed February 24, 2006, at URL *http://www.cityofstamford.org/HealthDepartmentLab/*.

Myers, D.N., 2003, Fecal indicator bacteria: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A7 (3d ed.), section 7.1, accessed June 16, 2005, at URL *http://pubs.water.usgs.gov/twri9A/*.

National Oceanic and Atmospheric Administration, 2005a, National Data Buoy Center: Stennis Space Center, Miss., accessed September 26, 2005, at URL *http://seaboard.ndbc.noaa.gov/*.

National Oceanic and Atmospheric Administration, 2005b, National Virtual Data System: National Climatic Data Center, Asheville, N.C., accessed September 26, 2005, at URL *http://nndc.noaa.gov/*.

Natural Resources Defense Council, 2005, Testing the waters-A guide to water quality at vacation beaches, July 2005: accessed December 30, 2005, at URL *http://www.nrdc.org/water/oceans/ttw/sumpen.pdf*.

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. 35, no. 1-3, p. 175-190.

Pennsylvania Bulletin, 2004, Department of Health amendments to Pennsylvania Code Rules and regulations, Title 28. Health and safety, Chapter 18. Public Swimming and Bathing Places (34 Pa.B. 3695): Harrisburg, Pa., v. 34, no. 29, p. 3695-3698.

Pennsylvania Code, Title 28. Health and safety, Chapter 18. Public swimming and bathing places, (28PaCode § 18): accessed May 2, 2006, at *http://www.pacode.com/*.

Pennsylvania Department of Conservation and Natural Resources, 2006, State Parks: accessed May 2, 2006, at *http://www.dcnr.state.pa.us/stateparks/parks/maps/presqueisle_mini.pdf*.

Pope, M.L., Bussen, M., Feige, M.A., Shadix, L., Gonder, S., Rodgers, C., Chambers, Y., Pulz, J., Miller, K., Connell, K., and Standridge, J., 2003, Assessment of the effects of holding time and temperature on *Escherichia coli* densities in surface water samples: Applied and Environmental Microbiology, v. 69, no. 10, p. 6201 – 6207.

SAS Institute, 1990, SAS user's guide—The LIFEREG procedure (4^{th} ed.): Cary, N.C., SAS Institute, Inc., Version 6, Volume 2, p. 997 – 1,025.

Siwicki, R.W., 2005, Water resources data, Pennsylvania, water year 2004, vol. 3, Ohio and St. Lawrence River Basins: U.S. Geological Survey Water-Data Report PA-04-3, 337 p.

Siwicki, R.W., 2006, Water resources data, Pennsylvania, water year 2005, vol. 3, Ohio and St. Lawrence River Basins: U.S. Geological Survey Water-Data Report PA-05-3, 355 p.

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

U.S. Environmental Protection Agency, 2002, Method 1603—*Escherichia coli (E. coli)* in water by membrane filtration using modified membrane-thermotolerant *Escherichia coli* agar (modified mTEC): Washington, D.C., Office of Water, EPA-821-R-02-023, 9 p.