≋ **USGS**
science for a changing world

Prepared in cooperation with the National Park Service

Development and Implementation of a Regression Model for Predicting Recreational Water Quality in the Cuyahoga River, Cuyahoga Valley National Park, Ohio 2009–11

Scientific Investigations Report 2012–5074

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

Development and Implementation of a Regression Model for Predicting Recreational Water Quality in the Cuyahoga River, Cuyahoga Valley National Park, Ohio 2009–11

By Amie M.G. Brady and Meg B. Plona

Scientific Investigations Report 2012–5074

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2012

Suggested citation:
Brady, A.M.G., and Plona, M., 2012, Development and implementation of a regression model for predicting recreational water quality in the Cuyahoga River, Cuyahoga Valley National Park, Ohio 2009–11: U.S. Geological Survey Scientific Investigations Report 2012–5074, 14 p.

Contents

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)
Volume		
milliliter (mL)	0.06102	cubic inch (in^3)
Flow rate		
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

Bacteria concentrations are given in colony-forming units per 100 milliliters (CFU/100 mL).

Filter pore size refers to the actual opening size of the pores in a membrane filter and is reported in microns (μm).

Turbidity is reported in nephelometric turbidity ratio units (NTRU).

Distance along a river is measured along the course of the river and is reported in river miles.

Development and Implementation of a Regression Model for Predicting Recreational Water Quality in the Cuyahoga River, Cuyahoga Valley National Park, Ohio 2009–11

By Amie M.G. Brady[1] and Meg B. Plona[2]

Abstract

The Cuyahoga River within Cuyahoga Valley National Park (CVNP) is at times impaired for recreational use due to elevated concentrations of *Escherichia coli (E. coli)*, a fecal-indicator bacterium. During the recreational seasons of mid-May through September during 2009–11, samples were collected 4 days per week and analyzed for *E. coli* concentrations at two sites within CVNP. Other water-quality and environmental data, including turbidity, rainfall, and streamflow, were measured and (or) tabulated for analysis. Regression models developed to predict recreational water quality in the river were implemented during the recreational seasons of 2009–11 for one site within CVNP–Jaite. For the 2009 and 2010 seasons, the regression models were better at predicting exceedances of Ohio's single-sample standard for primary-contact recreation compared to the traditional method of using the previous day's *E. coli* concentration. During 2009, the regression model was based on data collected during 2005 through 2008, excluding available 2004 data. The resulting model for 2009 did not perform as well as expected (based on the calibration data set) and tended to overestimate concentrations (correct responses at 69 percent). During 2010, the regression model was based on data collected during 2004 through 2009, including all of the available data. The 2010 model performed well, correctly predicting 89 percent of the samples above or below the single-sample standard, even though the predictions tended to be lower than actual sample concentrations. During 2011, the regression model was based on data collected during 2004 through 2010 and tended to overestimate concentrations. The 2011 model did not perform as well as the traditional method or as expected, based on the calibration dataset (correct responses at 56 percent). At a second site—Lock 29, approximately 5 river miles upstream from Jaite, a regression model based on data collected at the site during the recreational seasons of 2008–10 also did not perform as well as the traditional method or as well as expected (correct responses at 60 percent). Above normal precipitation in the region and a delayed start to the 2011 sampling season (sampling began mid-June) may have affected how well the 2011 models performed. With these new data, however, updated regression models may be better able to predict recreational water quality conditions due to the increased amount of diverse water quality conditions included in the calibration data.

Daily recreational water-quality predictions for Jaite were made available on the Ohio Nowcast Web site at *www.ohionowcast.info*. Other public outreach included signage at trailheads in the park, articles in the park's quarterly-published schedule of events and volunteer newsletters. A U.S. Geological Survey Fact Sheet was also published to bring attention to water-quality issues in the park.

Introduction

Cuyahoga Valley National Park (CVNP), located in northeastern Ohio between the cities of Akron and Cleveland, is a popular destination for recreationists (fig. 1). Over 2.5 million people visit the park annually to hike, ride the scenic railroad, bike the Towpath Trail, or enjoy the many recreational opportunities provided by the Cuyahoga River and its tributaries. Recent studies have shown that the Cuyahoga River is at times impaired for recreational use due to elevated concentrations of fecal-indicator bacteria (Bushon and Koltun, 2003; Bushon and others, 2007; Brady, 2007; Brady and others, 2009; Brady and Plona, 2009). Because of uncertain water quality, park managers have discouraged canoeing, swimming, or wading, but would like to promote the use of the river when the water quality is considered acceptable.
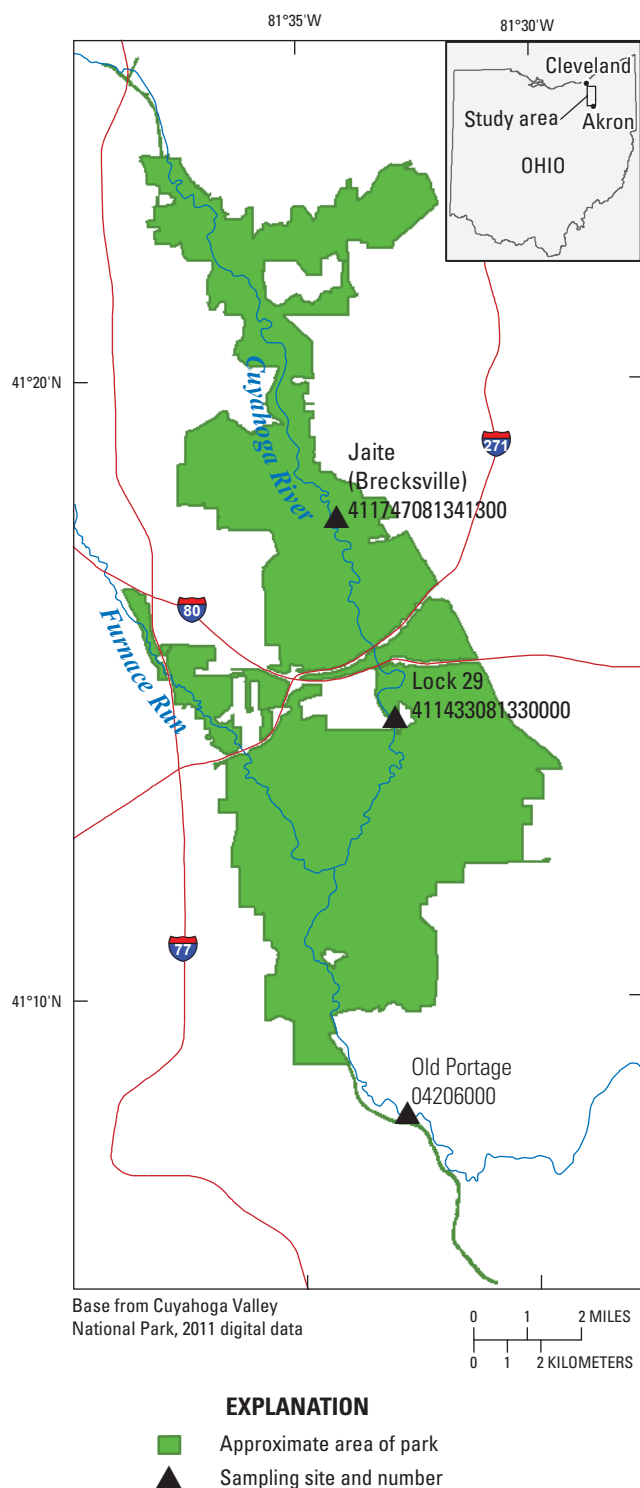
Figure 1.   Cuyahoga Valley National Park, Ohio, and water-quality sampling sites during the May through September recreational seasons of 2009–11. In this figure, the Cuyahoga River generally flows southeast to northwest.

When traditional methods of assessing water-quality conditions are used, concentrations of fecal-indicator bacteria are not available until the day after sample collection, and this period of time is too long to assess water quality and take adequate control measures. This is especially true in the river where concentrations can change greatly over short periods of time in response to decay, dilution, dispersion, and transport of bacteria (Myers and others, 1998). Predictive models have been shown to work well in other river and lake systems (Christensen and others, 2002; David and Haggard, 2011; Elaria and Vogel, 2005; Francy and Darner, 2007; Francy and others, 2009; Nevers and Whitman, 2006; Nevers and others, 2007; Rasmussen and Ziegler, 2003; Chattahoochee Riverway Project, 2011), and may be able to provide reliable results of the current day's fecal-indicator bacteria concentrations at CVNP, thus allowing park managers to determine whether they might promote river use.

During 2009 through 2011, the U.S. Geological Survey (USGS) and the National Park Service (NPS) implemented the use of models to predict *E. coli* concentrations at a centralized location within CVNP. Beginning in the recreational season of 2009, results from the predictive model developed during a previous study (Brady and Plona, 2009) were made available to the public through the Ohio Nowcast Web site (*www. ohionowcast.info*). Signage was placed throughout the park to provide the public with information and the address of the Web site. Each subsequent year, the regression model was refined to include the data from the previous season in the calibration dataset. This report summarizes the performance of the predictive models during 2009–11, information on the usage of the CVNP page of the Ohio Nowcast Web site, and future directions for predictive models within CVNP.

Methods

Escherichia coli (E. coli) concentrations were used to assess recreational water quality. *E. coli* are fecal-indicator bacteria whose concentrations have been shown to be correlated with occurrences of swimming-related gastrointestinal illness in epidemiological studies (Dufour, 1984). The U.S. Environmental Protection Agency recommends the use of *E. coli* for monitoring recreational water quality of fresh waters (U.S. Environmental Protection Agency, 1986). The Ohio primary-contact single-sample maximum of 298 colony-forming units per 100 milliliters (CFU/100 mL) was used as a benchmark to evaluate water quality and will be henceforth called the "water-quality standard" (WQS). On days when measured concentrations of *E. coli* in river samples were below or above the WQS, the river was considered safe or unsafe for recreation, respectively.

Sample Collection and Analysis

Water samples were collected by USGS or NPS staff from the Cuyahoga River at Jaite (USGS site number 411747081341300), a site centrally located in the park, and at Lock 29 (USGS site number 411433081330000), a site approximately 5 river miles upstream from Jaite (fig. 1). Recreational season (mid-May–September) sampling at Jaite started in 2004 and has continued through the present (2011). Recreational season sampling at Lock 29 started in late July 2008 and also has continued through the present. This stretch of the river was targeted for study because it is currently being used (unofficially) by recreational paddlers. During 2011, actual sampling was delayed until mid-June due to student scheduling.

Samples were collected at one location near the center of flow by use of a weighted-bottle sampler fitted with a sterile 1-liter (L) bottle (Wilde, 2006). The weighted sampler was submerged and allowed to fill. The bottle was capped and placed on ice in a cooler until analyses were performed. Samples were collected 4 days per week, and a replicate sample was collected once per week.

Samples were processed within 6 hours of sample collection following the modified mTEC method (U.S. Environmental Protection Agency, 2006). Aliquots of water were filtered through a 0.45-micron (µm) filter and then placed onto an agar plate. Inverted plates were incubated at 35 degrees Celsius (°C) for 2 hours and then transferred to an incubator at 44.5 ± 0.5°C for 20–22 hours. Magenta-colored colonies were counted as *E. coli,* and results were reported as CFU/100 mL.

Turbidity provides an indication of the clarity of the water. A laboratory turbidimeter (Hach 2100P) was used to measure turbidity within 2 hours of sample collection so that results were available to run predictive models. The average of two consecutive measurements that agreed within 10 percent was reported for each sample in units of nephelometric turbidity ratio units (NTRU).

Environmental Data Collection

River gage height was collected by means of a weighted, steel tape measure. The distance to the water surface was measured with the tape from a designated point on the bridge with a known elevation. This tape-down measurement was recorded in feet (ft). In addition, instantaneous streamflow at time of sampling as measured at Old Portage, Ohio (USGS streamgage 04206000), upstream from Lock 29, was compiled. Instantaneous streamflow was reported in units of cubic feet per second.

Rainfall data were compiled from the National Weather Service local climatology data stations at Akron Fulton International Airport (call sign AKR) and Akron-Canton Airport (call sign CAK) (National Oceanic and Atmospheric Administration, 2011). The airports are located approximately 18 and 27 miles, respectively, south of the Jaite site and approximately 15 and 24 miles, respectively south of the Lock 29 site. Everyday during the recreational season, hourly precipitation totals were summed for the 24-hour period prior to 8:00 AM (Rain24). A weighted rainfall variable (Rw_{48}) was calculated by multiplying the sampling day's Rain24 by two and adding the previous day's Rain24 (provides more weight to the most recent rainfall). Rainfall was recorded in inches (in.).

Quality-Control and Quality-Assurance Procedures

Quality-control and quality-assurance (QA/QC) procedures are important to ensure the integrity of the data being collected. Therefore, strict QA/QC procedures were followed throughout this study. Written protocols were distributed to USGS and NPS staff prior to the onset of sample collection. Collection and (or) analysis that followed protocols established in Francy and others (2010) were done daily for filter blanks, weekly for replicate samples, and two or three times each season for field blanks. Quality-control checks were performed onsite monthly to ensure that sample collection and analyses were done according to protocols. No QA/QC problems were identified during quality-control checks. All filter and field blanks were negative for bacterial growth. Variability of *E. coli* concentrations, defined as the absolute value of the difference between $\log_{10}$-transformed replicate pair concentrations divided by the average of the $\log_{10}$-transformed concentrations, ranged from 0 to 11 percent. This variability was similar to what was reported during previous studies at the park (0 to 10 percent) (Brady and others, 2009; Brady and Plona, 2009).

Statistical Methods

Correlation analysis and scatter plots were used to examine the relations between *E. coli* concentrations and environmental and water-quality variables. A linear correlation coefficient (Pearson's r) was used to determine the degree to which parameters and (or) variables were related to covariates. The more the coefficient differed from 1 or -1, the weaker the relation. Linear regression analysis was used to estimate $\log_{10}$-transformed *E. coli* concentrations in the river based on one explanatory variable (simple linear regression or SLR) or more than one variable (multiple linear regression). For model selection, the adjusted coefficient of determination (adj. R^2) and the Mallows' Cp statistic were used to narrow the subset of all possible models. The root mean squared error (RMSE), also referred to as the standard error of the estimate (SEE), was determined to assess the variation in the predicted versus the actual concentrations. Assumptions of linear regression (homoscedasticity, autocorrelation, and normality) were tested by examining plots of the residuals against predicted values and over time and by using normal probability plots. For all models, heteroscedasticity was not observed. Some

autocorrelation (based on Durbin-Watson tests, D ranged from 1.5 to 1.6) and non-normality (based on Shapiro-Wilk tests, W ranged from 0.92 to 0.95) were observed and indicate that the models could be improved upon. The methods utilized in model development and (or) refinement and selection steps are further described in Francy and Darner (2006) and specifically for CVNP in Brady and others (2009). The level of significance for statistical tests was set at α = 0.05 for this report.

For model development and (or) refinement and comparisons between models and the traditional method (using the previous day's *E. coli* concentrations), the method's effectiveness for determining recreational water quality was based on several response statistics. The predicted *E. coli* concentration for each day was compared to the measured concentration as determined by the membrane-filtration technique. Rates of correct predictions above or below the WQS, false positives, and false negatives were calculated. False positives were defined as model predictions above the WQS when the measured concentrations were below the WQS (nonexceedance). False negatives were defined as model predictions below the WQS when the measured concentrations exceeded the WQS. Model sensitivity and specificity were also determined. Sensitivity is the proportion of correctly predicted exceedances of the WQS out of all exceedances. Specificity is the proportion of correctly predicted nonexceedances out of all nonexceedances. In terms of public health protection, the frequency of false negatives and the sensitivity of the model were the principle statistics considered; however, false positives

and the specificity of the model are important statistics that could cause a decrease in river use, because the public may be unnecessarily discouraged from recreating in the river.

Water-Quality Model for Jaite

Model Development and Validation, 2009

During the recreational season of 2009, a regression model with $\log_{10}$ turbidity as the sole explanatory variable was used to predict recreational water quality of the Cuyahoga River at Jaite (table 1). This model was based on data collected during the recreational seasons of 2005 through 2008. During model development, this model, which excluded the 2004 data from the calibration dataset, was determined to provide more accurate results compared to the model including the 2004 data (81 and 77 percent correct for the calibration set that excluded and included the 2004 data, respectively). Precipitation during the 2004 season (as observed at CAK), was greater than 7 in. above normal, and storms during the season resulted in record flooding in the park. The greatest river stage recorded in over 80 years was 13.74 ft on May 22, 2004 as observed at the Old Portage USGS streamgage. The extreme conditions resulted in unusual observations, which may explain why the model performed better during model development when data from 2004 were not included.

Table 1. Variables and regression statistics for Cuyahoga River models, Cuyahoga Valley National Park, Ohio.

[RMSE, root mean square error; GH, gage height; Rw_{48}, rainfall weighted for 48-hour period prior to sampling]

Site and test year	Calibration data seasons (number of samples)	Adjusted R^2	RMSE	Variables in Model	Parameter estimate	Significance of variable
Jaite 2009	2005–8 (151)	0.75	0.33218	y-intercept	1.361	<0.0001
				$\log_{10}$ turbidity	1.056	<0.0001
Jaite 2010	2004 –9 (287)	0.68	0.35168	y-intercept	1.608	<0.0001
				$\log_{10}$ turbidity * $\log_{10}$ GH	1.046	<0.0001
Jaite 2011	2004–10 (335)	0.69	0.34405	y-intercept	1.277	<0.0001
				$\log_{10}$ turbidity	1.125	<0.0001
Lock 29 2011	2008–10 (100)	0.76	0.27396	y-intercept	1.295	<0.0001
				$\log_{10}$ turbidity	1.147	<0.0001
				Rw_{48}	0.201	0.0022

Forty-nine samples were collected at Jaite between May 26 and August 20, 2009. Twenty-six of the samples, or 53 percent, exceeded the WQS. The Jaite 2009 model correctly predicted exceedances or nonexceedances for 34 samples, or 69 percent (table 2). This percentage is lower than expected—81 percent of the calibration data sample results were predicted correctly (n = 151, data not shown), and similar percentages were expected for the validation year. Precipitation during the 2009 season was a little above normal (approximately 1 in.), which may help explain why the Jaite 2009 model did not perform as well as expected. An additional contributor to the unexpected performance of the Jaite 2009 model was the exclusion of the 2004 data, which included wet-weather sampling, from the calibration dataset.

The higher error rate for the Jaite 2009 model resulted from a greater percentage of false positives, indicating that the model was overestimating concentrations in the river as can be seen in figure 2. Compared to the traditional method of determining advisories by using the previous day's *E. coli* concentration, the Jaite 2009 model had fewer false negatives and more false positives (table 2). The Jaite 2009 model had a higher sensitivity but a lower specificity than that of the traditional method, which indicates that the model provides better protection in terms of public health, but could be deterring park visitors from using the river when the water is safe for recreation.

Model Refinement and Validation, 2010

In keeping with conventional procedures for maintaining operational Nowcast models (Francy and Darner, 2006; Francy and others, 2009; Stidson and others, 2012) the entirety of the data, including data collected during the validation year, was tabulated for model refinement for the upcoming season. Scatter plots and correlation analyses were done to examine the data in detail. Due to the decrease in predictive capabilities observed during the 2009 validation year, subsets of data based on year of collection were not explored during model refinement for the 2010 recreational season.

Statistically significant correlations were observed between $\log_{10}$-transformed *E. coli* concentrations and $\log_{10}$ turbidity (r = 0.83), the mathematical product of $\log_{10}$ turbidity and $\log_{10}$ gage height (r = 0.82), and Rw_{48} at AKR (r = 0.72). These variables and selected others were included as potential explanatory variables during model refinement. The new model prepared for 2010 (Jaite 2010) was a SLR model, but was based on the mathematical product of $\log_{10}$ turbidity and $\log_{10}$ gage height (table 1). This cross-product model was chosen over the SLR model with $\log_{10}$ turbidity as the sole explanatory variable, because the cross-product model provided more correct responses and higher sensitivity and specificity compared to the $\log_{10}$ turbidity model for the calibration data set.

Table 2. Comparison of the percentage of correct and false positive and negative responses of recreational water-quality conditions using the Jaite 2009 model and the traditional method for determining water quality during model development and validation, Cuyahoga Valley National Park, Ohio.

[Jaite 2009 model based on $\log_{10}$ turbidity; traditional method uses the *Escherichia coli* concentration from the previous day; WQS, water quality standard; %, percent]

	Number of samples	Exceedances of WQS	Correct responses	False positive	False negative	Sensitivity	Specificity
			Value (number of samples)				
Development—calibration data							
Model	151	56% (85)	81% (122)	11% (17)	7.9% (12)	86%	74%
Traditional method	101	58% (59)	68% (69)	13% (13)	19% (19)	68%	69%
Validation—2009 data							
Model	49	53% (26)	69% (34)	28% (14)	2.0% (1)	96%	39%
Traditional method	35	54% (19)	71% (25)	14% (5)	14% (5)	74%	69%
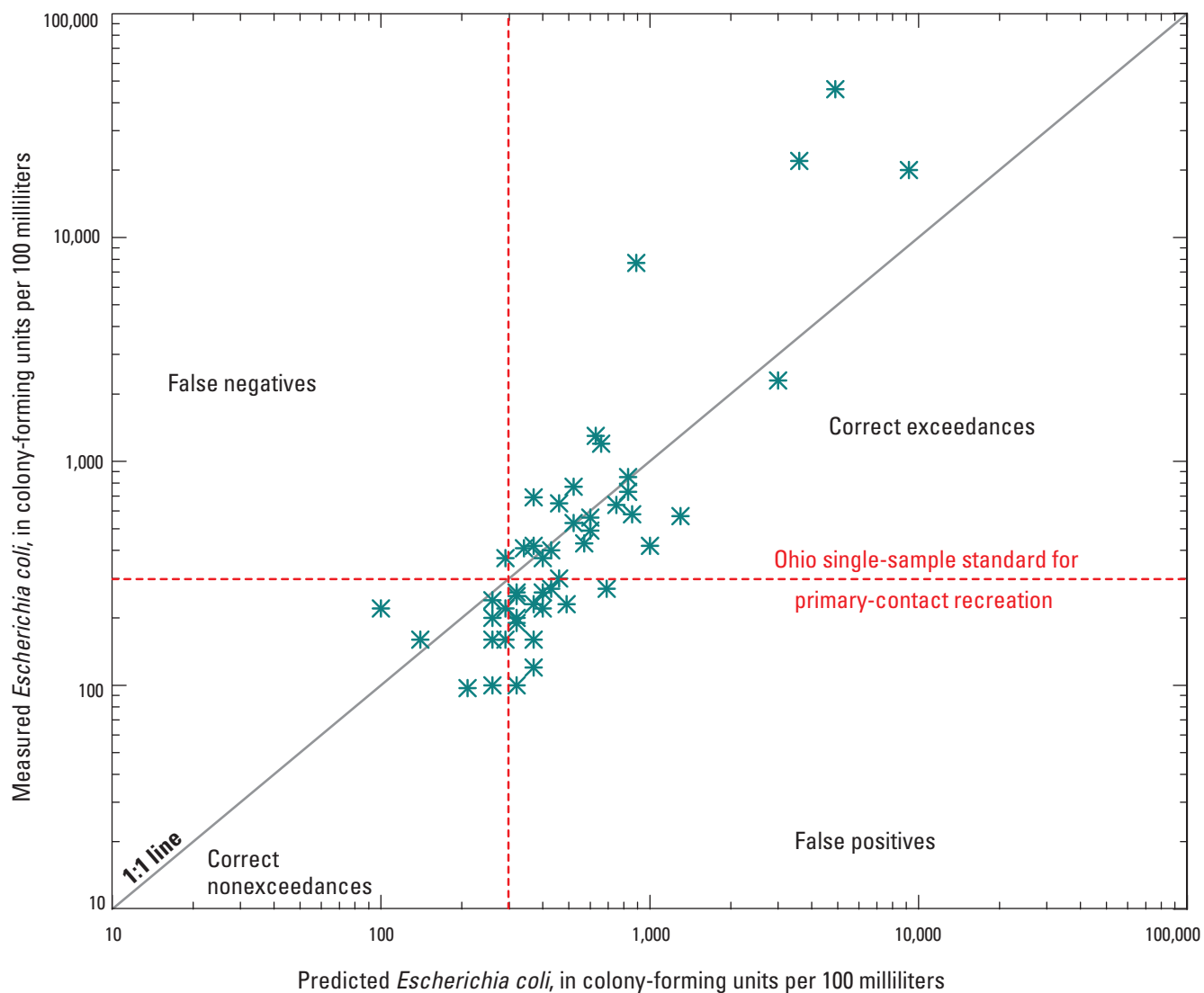
Figure 2.   Predicted *Escherichia coli (E. coli)* concentrations from the regression model developed using 2005 through 2008 recreational-season data compared to the measured *E. coli* concentrations at Jaite, Cuyahoga Valley National Park, Ohio, May through August 2009.

Forty-four samples were collected from May 27 through August 25, 2010. Twenty-seven of the samples, or 61 percent, exceeded the WQS. The Jaite 2010 model correctly predicted exceedances or nonexceedances for 39 samples, or 89 percent (table 3). The 2010 model tended to underestimate *E. coli* concentrations when measured concentrations were greater than the WQS (fig. 3), but generally the predicted concentrations were still above the WQS, resulting in correct classification of the samples as exceedances. Compared to the traditional method, the Jaite 2010 model predicted correct exceedances and nonexceedances more often. The model statistics indicate that, compared to the traditional method, the Jaite 2010 model is better able to provide protection in terms of public health as well as less often discourage park visitors from using the river when the water is safe for recreation.

Table 3. Comparison of the percentage of correct and false positive and negative responses of recreational water-quality conditions using the Jaite 2010 model and the traditional method for determining water quality during model refinement and validation, Cuyahoga Valley National Park, Ohio.

[Jaite 2010 model based on the mathematical product of $\log_{10}$ turbidity and $\log_{10}$ gage height; traditional method uses the previous day's *Escherichia coli* concentration; WQS, water quality standard; %, percent]

	Number of samples	Exceedances of WQS	Correct responses	False positive	False negative	Sensitivity	Specificity
			Value (number of samples)				
			Refinement—calibration data				
Model	287	60% (173)	78% (224)	16% (46)	5.9% (17)	90%	60%
Traditional method	221	62% (137)	72% (158)	13% (29)	15% (34)	75%	65%
			Validation—2010 data				
Model	44	61% (27)	89% (39)	4.5% (2)	6.8% (3)	89%	88%
Traditional method	32	59% (19)	78% (25)	9.4% (3)	12% (4)	79%	77%

Figure 3. Predicted *Escherichia coli (E. coli)* concentrations from the regression model developed using 2004 through 2009 recreational-season data compared to the measured *E. coli* concentrations at Jaite, Cuyahoga Valley National Park, Ohio, May through August 2010.
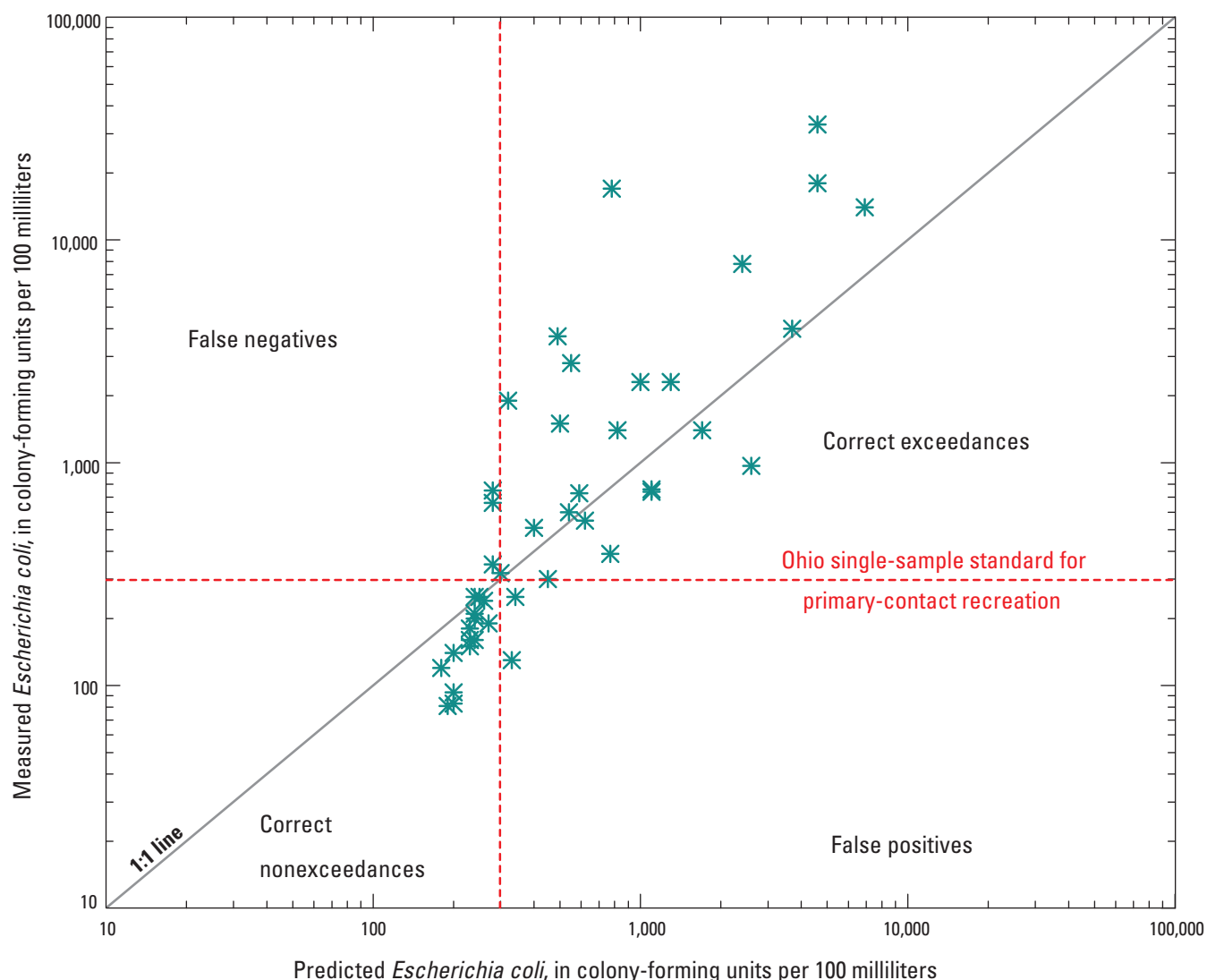
Model Refinement and Validation, 2011

The data collected during 2010 were added to the dataset, and scatter plots and correlation analyses were done to examine the 2004–10 data in detail. Similar to what was observed during the previous year's data analysis, statistically significant relations were observed between log_{10}-transformed *E. coli* concentrations and log_{10} turbidity (r = 0.83), the mathematical product of log_{10} turbidity and log_{10} gage height (r = 0.82), and Rw_{48} at AKR (r = 0.70). These variables and selected others were included as potential explanatory variables during model refinement. The new model prepared for 2011 (Jaite 2011) was a SLR model, but it was based solely on log_{10} turbidity (table 1). Other models that included additional variables were explored, but there was little to no improvement in the model's predictive performance compared to the chosen model.

Precipitation for the 2011 season was above normal, and the year was ranked as the second highest for precipitation (National Weather Service Forecast Office, 2012). Previous years with above normal precipitation for the season tended to have percentages of WQS exceedance greater than those closer to normal or below normal precipitation (data not shown). Thirty-nine samples were collected from June 13 through September 19, 2011. Fifteen samples, or 38 percent, exceeded the WQS. The delayed season may help explain why percentages of WQS exceedances were lower (38 percent) than in previous years (60 percent) (table 4). Further, the Jaite 2011 model correctly predicted exceedances or nonexceedances of the WQS for 56 percent of the samples (fig. 4), which is much lower than expected based on the calibration dataset (78 percent) (table 4). Together, these general observations imply that the Jaite 2011 model needs to be improved upon. With the addition of the 2011 data to the calibration dataset, an updated regression model may be able to better predict recreational water quality during more diverse weather conditions, such as those observed during the 2011 season.

Table 4. Comparison of the percentage of correct and false positive and negative responses of recreational water-quality conditions using the Jaite 2011 model and the traditional method for determining water quality during model refinement and validation, Cuyahoga Valley National Park, Ohio.

[Jaite 2011 model based on log_{10} turbidity; traditional method uses the previous day's *Escherichia coli* concentration; WQS, water quality standard; %, percent]

	Number of samples	Exceedances of WQS	Correct responses	False positive	False negative	Sensitivity	Specificity
			Value (number of samples)				
Refinement—calibration data							
Model	335	60% (202)	78% (262)	16% (53)	6.0% (20)	90%	60%
Traditional method	253	62% (156)	72% (183)	13% (32)	15% (38)	76%	67%
Validation—2011 data							
Model	39	38% (15)	56% (22)	41% (16)	2.6% (1)	88%	33%
Traditional method	25	28% (7)	72% (18)	20% (5)	8.0% (2)	71%	72%
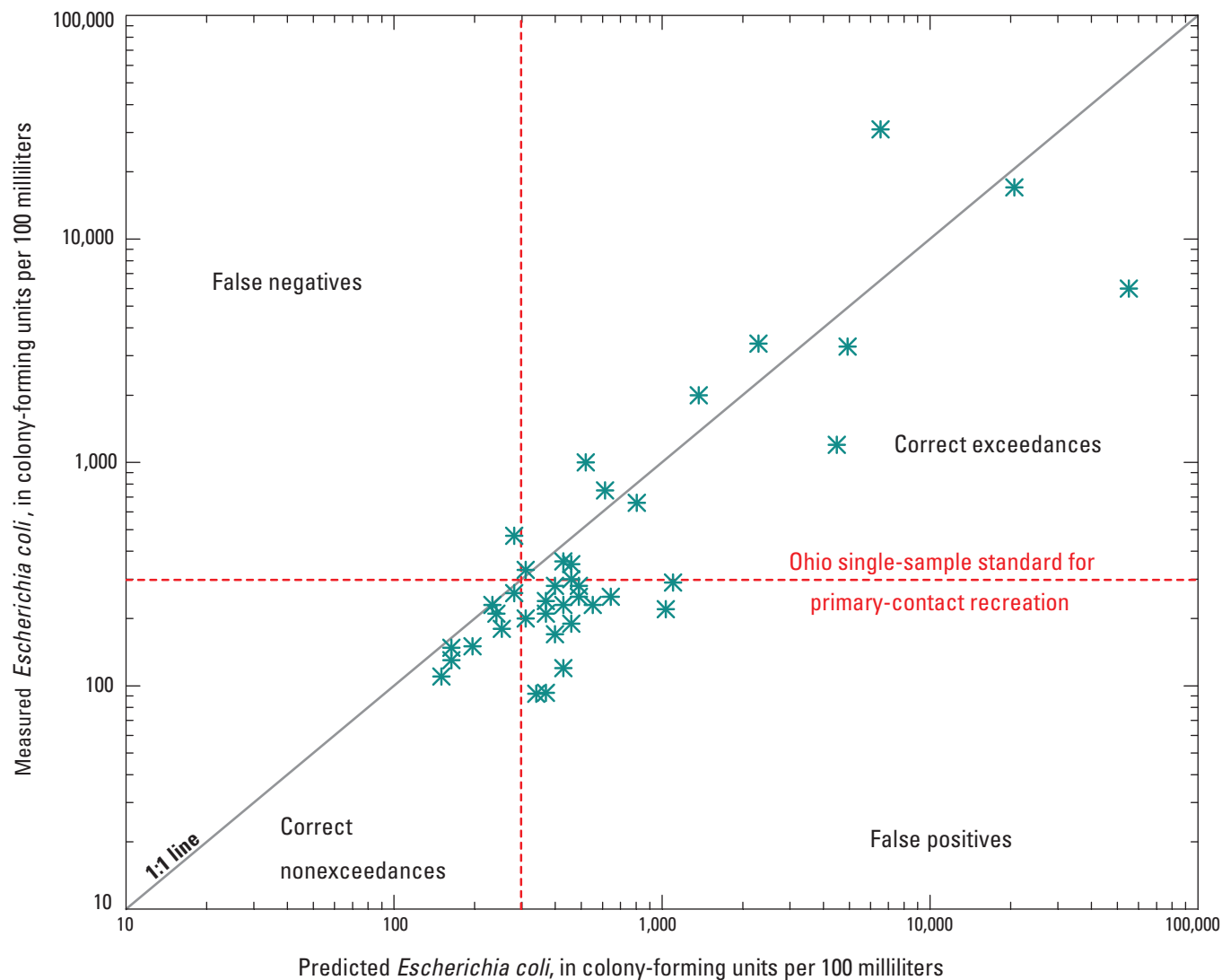
Figure 4. Predicted *Escherichia coli (E. coli)* concentrations from the regression model developed using 2004 through 2010 recreational-season data compared to the measured *E. coli* concentrations at Jaite, Cuyahoga Valley National Park, Ohio, June through September 2011.

Water-Quality Model for Lock 29

Model Development and Validation, 2011

Data collected at Lock 29 during the recreational seasons of 2008 through 2010 were compiled. Scatterplots and correlation analyses were done to examine the data in detail. Statistically significant relations were observed between $\log_{10}$ *E. coli* concentrations and $\log_{10}$ turbidity (r = 0.86), the amount of rain that fell in 24-hour period prior to sample collection at AKR (r = 0.67), Rw_{48} at AKR (r = 0.73), and $\log_{10}$ instantaneous streamflow at Old Portage (r = 0.54). Several models were explored, including models that had streamflow as an explanatory variable. Multicollinearity between this and other explanatory variables tended to decrease the model's predictive capabilities and was therefore not included in the final model. The Lock 29 2011 model was a multiple linear regression model with $\log_{10}$ turbidity and Rw_{48} as explanatory variables (table 1).

From June 13 through September 19, 2011, 37 samples were collected. Sixteen of the samples, or 43 percent, exceeded the WQS. As observed for samples collected during the same season at the Jaite site, fewer WQS exceedances than expected were detected in samples collected at Lock 29 during the 2011 season (based on the calibration data). Correspondingly, the Lock 29 2011 model correctly predicted only 60 percent of the samples as nonexceedances or exceedances (table 5). Based on correct predictions and the number of false positives, the Lock 29 2011 model did not perform as well compared to the traditional method. The model did have 100 percent sensitivity, but this may be due to the model's tendency to overestimate concentrations (fig. 5). Above normal precipitation observed for the region resulted in the 2011 season being a poor year for validating the model. Inclusion of the 2011 data in the calibration dataset may allow an updated regression model to better predict recreational water quality.

Table 5. Comparison of the percentage of correct and false positive and negative responses of recreational water-quality conditions using the Lock 29 2011 model and the traditional method for determining water quality during model development and testing, Cuyahoga Valley National Park, Ohio.

[Lock 29 2011 model based on $\log_{10}$ turbidity and weighted 48-hour rainfall; traditional method uses the previous day's *Escherichia coli* concentration; WQS, water quality standard; %, percent]

	Number of samples	Exceedances of WQS	Correct responses	False positive	False negative	Sensitivity	Specificity
		Value (number of samples)					
Development—calibration data							
Model	100	58% (58)	77% (77)	14% (14)	9.0% (9)	84%	67%
Traditional method	70	56% (39)	69% (48)	14% (10)	17% (12)	71%	66%
Validation— 2011 data							
Model	37	43% (16)	60% (22)	41% (15)	0.0% (0)	100%	29%
Traditional method	22	41% (9)	64% (14)	18% (4)	18% (4)	56%	69%
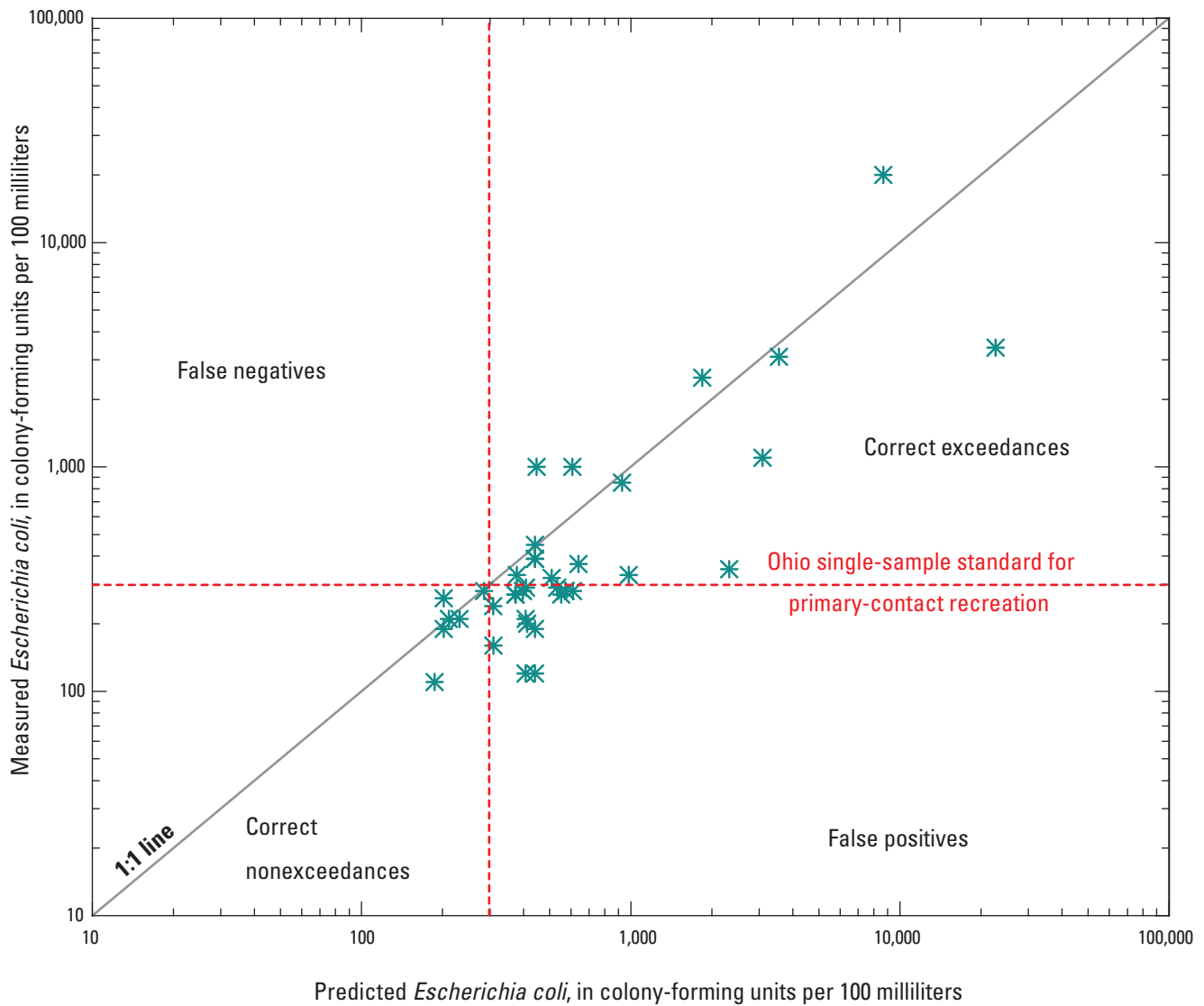
Figure 5.   Predicted *Escherichia coli (E. coli)* concentrations from the regression model developed using 2004 through 2010 recreational-season data compared to the measured *E. coli* concentrations at Lock 29, Cuyahoga Valley National Park, Ohio, June through September 2011.

Implementation of Models at Cuyahoga Valley National Park

For the recreational season of 2009, the Ohio Nowcast Web site (*www.ohionowcast.info*) was adapted to include the CVNP Nowcast system. From the home page, viewers were able to see current predicted conditions at Jaite, and from the Cuyahoga River Web page, more detailed information about the sampling site and links to previous data (from 2009) were available (fig. 6).

The NPS managers posted signs at trailheads along the river to raise public awareness of water-quality issues and referenced the Ohio Nowcast Web site for further information. Other public outreach items included articles on the park's Web page (*www.nps.gov/cuva*) and in the park's quarterly published schedule of events and volunteer newsletters. A USGS Fact Sheet (Brady and Plona, 2010) was published and made available at multiple locations within the park. During 2011, to further communicate water-quality issues to park visitors, a student intern provided public outreach on the Nowcast system to over 400 park visitors on busy summer weekends. Business cards were prepared with a brief description of the Ohio Nowcast system and the Web page address, and the student intern was able to distribute the cards to interested park visitors.

During the recreational season of 2009, there were 1,342 external visits (excluding administrative checks and daily updates) to the CVNP Web page of the Ohio Nowcast Web site. During the recreational seasons of 2010 and 2011, there were 1,607 and 1,964 external visits to the Web page, respectively. For all years, there were more visits to the site during July than any other month (May through August). Because the Ohio Nowcast Web site is not specific to the Cuyahoga River, visits to the home page were not tallied. Therefore, the external visitor count demonstrates the number of visitors interested in more detailed information on the water quality of the river and may be an underestimate of the actual use of the Web site for this purpose.
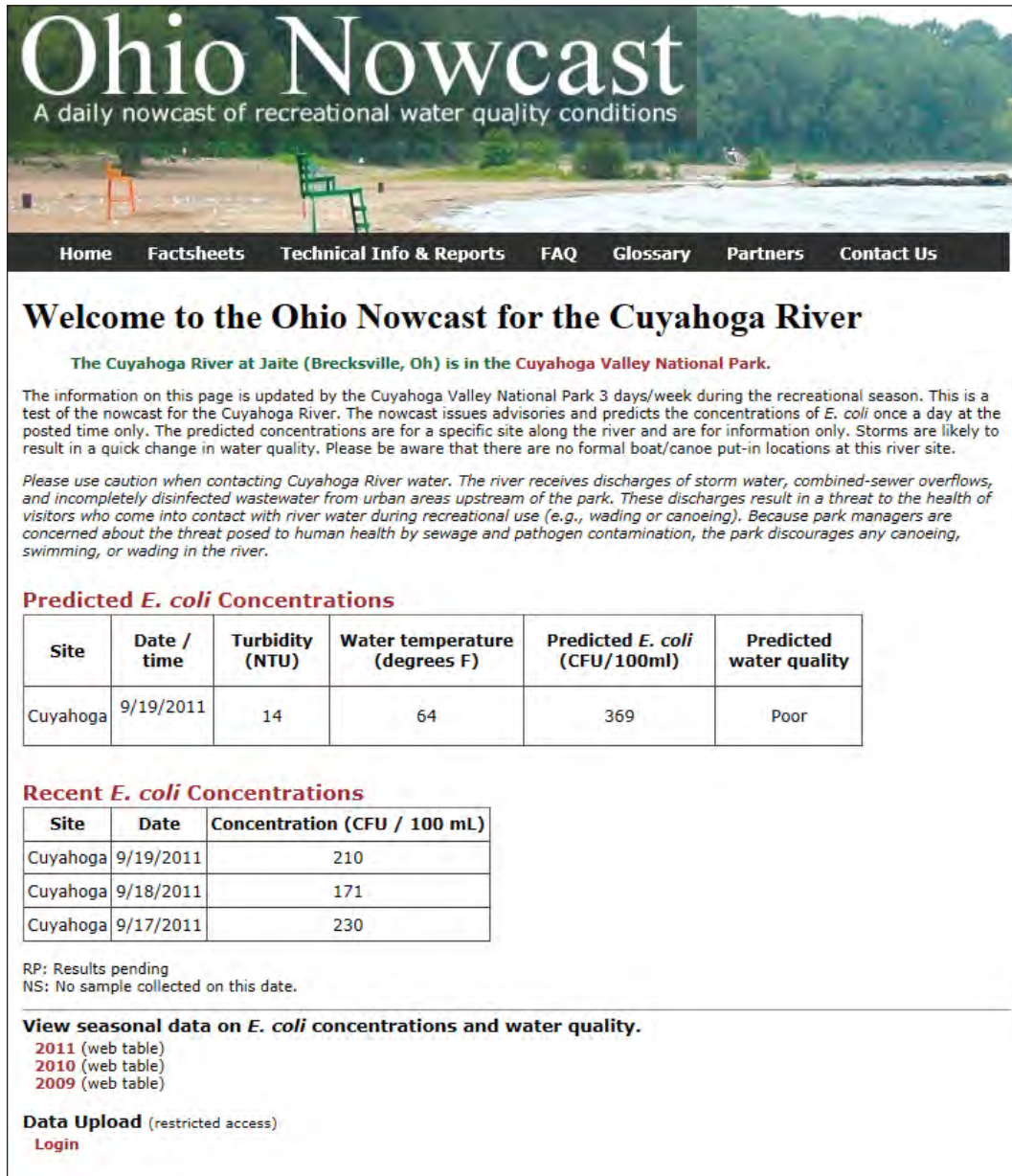
Figure 6.   Sample of the Cuyahoga River Web page of the Ohio Nowcast Web site (*www.ohionowcast.info*).

Summary and Future Directions

The Cuyahoga River within the Cuyahoga Valley National Park is often impaired for recreation due to elevated concentrations of *E. coli*. To assist park managers who would like to encourage the use of the river when it is safe to do so, the USGS in collaboration with the park, initiated a monitoring program that enabled the development and (or) refinement and implementation of regression models to predict recreational water quality of the river at two sites. At Jaite, a site centrally located in the park, water-quality data from 2004 were available to develop and (or) refine predictive models for the recreational seasons of 2009–11. Recreational water-quality predictions were posted daily on the Ohio Nowcast Web site during each year. Each year, the predictive models were refined to include the previous season's data in the calibration dataset. Compared to the traditional method for determining recreational water quality by using the previous day's *E. coli* concentrations, the regression models were better able to predict exceedances of the WQS and had fewer false negative responses during the 2009 and 2010 seasons. During the 2011 season, however, the model did not perform as well as the traditional method, possibly due to the unusual precipitation. At a second site on the river, Lock 29, approximately 5 river miles upstream of Jaite, monitoring was done during the 2008–10 seasons, and a regression model based on turbidity and rainfall over the previous 48 hours was developed. Similar to the Jaite model for the 2011 season, the Lock 29 model did not perform as well as the traditional method. Above normal precipitation for the region and a delayed sampling season (sample collection began mid-June 2011), which had not been included in the calibration datasets, may account for the decreased performance of both models during the 2011 season.

Continued monitoring during upcoming recreational seasons is planned. Utilization of the Ohio Nowcast Web site for serving daily predictions at Jaite, and possibly Lock 29, is anticipated. In addition, the use of a real-time turbidity sensor is being explored for the Jaite site, and this would involve establishing a semipermanent stream monitor at the site. Data would then be telemetered to the USGS, where it would be processed for display on the National Water Information System Web Interface (NWIS Web). Park managers could access the data as it becomes available and run the regression model 7 days per week, eliminating the need to manually collect daily samples. Continued manual sampling for bacteria is necessary to ensure that the regression model results continue to reflect environmental conditions; however, sampling could be reduced to 2 or 3 days per week.

Afternoon monitoring at Jaite is also planned. Those monitoring data would provide information needed to determine if relations between *E. coli* concentrations and environmental and water-quality variables previously developed, based on bacterial data collected in the morning, can also be applied to conditions present in the afternoon or if a new regression model would be needed. In either case, the ability to predict recreational water quality in the afternoon will be a benefit to recreationists planning on utilizing the river later in the day.

Additional public outreach, including hands-on training for park rangers, utilization of science communication interns for water-quality interpretation and education, additional signage, and presentations by researchers, is planned.

Acknowledgments

References Cited

Brady, A.M.G., 2007, Rapid method for *Escherichia coli* in the Cuyahoga River: U.S. Geological Survey Open-File Report 2007–1210, 5 p.

Brady, A.M.G., Bushon, R.N., and Plona, M.B., 2009, Predicting recreational water quality using turbidity in the Cuyahoga River, Cuyahoga Valley National Park, Ohio, 2004–7: U.S. Geological Survey Scientific Investigations Report 2009–5192, 16 p.

Brady, A.M.G., and Plona, M.B., 2009, Relations between environmental and water-quality variables and *Escherichia coli* in the Cuyahoga River with emphasis on turbidity as a predictor of recreational water quality, Cuyahoga Valley National Park, Ohio, 2008: U.S. Geological Survey Open-File Report 2009–1192, 6 p.

Brady, A.M.G., and Plona, M.B., 2010, Occurrence of *Escherichia coli* in the Cuyahoga River in the Cuyahoga Valley National Park, Ohio: U.S. Geological Survey Fact Sheet 2010–3068, 4 p.

Bushon, R.N., Brady, A.M.G., and Plona, M.B., 2007, Using a rapid method to predict recreational water quality at Cuyahoga Valley National Park, Ohio: Park Science, v. 24, no. 2, p. 89–93.

Bushon, R.N., and Koltun, G.F., 2003, Microbiological water-quality in relation to water-contact recreation, Cuyahoga River, Cuyahoga Valley National Park, Ohio, 2000 and 2002: U.S. Geological Survey Water-Resources Investigations Report 03–4333, 30 p.

Chattahoochee Riverway Project, 2011, BacteriALERT Program: Project Web page, accessed March 12, 2012, at *http://ga2.er.usgs.gov/bacteria/default.cfm*.

Christensen, V.G., Rasmussen, P.P., and Ziegler, A.C., 2002, Real-time water quality monitoring and regression analysis to estimate nutrient and bacteria concentrations in Kansas streams: Water Science and Technology, v. 45, no. 9, p. 205–219.

David, M.M., and Haggard, B.E., 2011, Development of regression-based models to predict fecal bacteria numbers at select sites within the Illinois River watershed, Arkansas and Oklahoma, USA: Water, Air, and Soil Pollution, v. 215, p. 525–547.

Dufour, A.P., 1984, Health effects criteria for fresh recreational waters: U.S. Environmental Protection Agency: Cincinnati, Ohio, EPA 600/1–84–004, 33 p.

Eleria, A., and Vogel, R.M., 2005, Predicting fecal coliform bacteria levels in the Charles River, Massachusetts, USA: Journal of the American Water Resources Association, v. 41, no. 5, p. 1195–1209.

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, 19 p.

Francy, D.S., Bushon, R.N., Brady, A.M.G., Kephart, C.M., Stelzer, E.A., and Milam, C.M., 2010, Quality assurance/quality control manual: U.S. Geological Survey, Ohio Water Microbiology Laboratory, accessed April 22, 2011, at *http://oh.water.usgs.gov/micro_qaqc.htm*.

Francy, D.S., and Darner, R.A., 2006, Procedures for developing models to predict exceedances of recreational water-quality standards at coastal beaches: U.S. Geological Survey Techniques and Methods book 6, chap. B5, 34 p.

Francy, D.S., and Darner, R.A., 2007, Nowcasting beach advisories at Ohio Lake Erie beaches: U.S. Geological Survey Open-File Report 2007–1427, 13 p.

Myers, D.N., Koltun, G.F., and Francy, D.S., 1998, Effects of hydrologic, biologic, and environmental processes on sources and concentrations of fecal bacteria in the Cuyahoga River, with implications for management of recreation waters in Summit and Cuyahoga Counties, Ohio: U.S. Geological Survey Water-Resources Investigations Report 98–4089, 45 p.

National Weather Service Forecast Office, 2012, Cleveland, OH: Akron/Canton, OH. Climate, 2011 climate summary: accessed March 13, 2012, *http://www.erh.noaa.gov/cle/climate/cak/climatecak.html*.

Nevers, M.B., and Whitman, R.L., 2006, Nowcast modeling of *Escherichia coli* concentrations at multiple urban beaches of southern Lake Michigan: Water Research, v. 39, no. 20, p. 5250–5260.

Nevers, M.B., Whitman, R.L., Frick, W.E., and Ge, Z., 2007, Interaction and influence of two creeks on *E. coli* concentrations of nearby beaches—Exploration of predictability and mechanisms: Journal of Environmental Quality, v. 36, no. 5, p. 1338–1345.

National Oceanic and Atmospheric Association, 2011, National Virtual Data System: Asheville, N.C., National Climatic Data Center, accessed August 2011 at *http://www.ncdc.noaa.gov/*.

Ohio Environmental Protection Agency, 2008, Water quality standards: Ohio Administrative Code chap. 3745–1, January 23, 2008, p. 01:1, 07:1–8, 07:24.

Rasmussen, P.P., and Ziegler, A.C., 2003, Comparison and continuous estimates of fecal coliform and *Escherichia coli* bacteria in selected Kansas streams, May 1999 through April 2002: U.S. Geological Survey Water-Resources Investigations Report 03–4056, 80 p.

Stidson, R.T., Gray, C.A., and McPhail, C.D., 2012, Development and use of modeling techniques for real-time bathing water quality predictions: Water and Environment Journal, v. 26, p. 7–18.

U.S. Environmental Protection Agency, 2006, Method 1603—*Escherichia coli* in water by membrane filtration using modified membrane-thermotolerant *Escherichia coli* agar: Washington, D.C., EPA 821–R–06–011 [variously paged].

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

Wilde, F.D., ed., 2006, Collection of water samples (ver. 2.0): U.S. Geological Survey Techniques of Water-Resources Investigations book 9, chap. A4, accessed March 31, 2011, at *http://pubs.water.usgs.gov/twri9A4/*.