In cooperation with the National Park Service and the Ohio Lake Erie Commission

Predicting Recreational Water Quality Using Turbidity in the Cuyahoga River, Cuyahoga Valley National Park, Ohio, 2004–7

Scientific Investigations Report 2009–5192

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

Predicting Recreational Water Quality Using Turbidity in the Cuyahoga River, Cuyahoga Valley National Park, Ohio, 2004–7

By Amie M.G. Brady, Rebecca N. Bushon, and Meg B. Plona

In cooperation with the National Park Service and the Ohio Lake Erie Commission

Scientific Investigations Report 2009–5192

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

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

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

U.S. Geological Survey, Reston, Virginia: 2009

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit http://www.usgs.gov or call 1-888-ASK-USGS

For an overview of USGS information products, including maps, imagery, and publications, visit http://www.usgs.gov/pubprod

To order this and other USGS information products, visit http://store.usgs.gov

Suggested citation:
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.

Cover photo:
The Cuyahoga River at the State Route 82 (Chippewa Rd.) bridge. (Photo courtesy of the National Park Service.)

Contents

Figures

Tables

Conversion Factors

Multiply	By	To obtain
	Length	
inch (in.)	2.54	centimeter (cm)
mile (mi)	1.609	kilometer (km)
	Volume	
ounce, fluid (fl. oz)	0.02957	liter (L)
	Area	
acre	4,047	square meter (m^2)
	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 millilliters (CFU/100 mL).

Predicting Recreational Water Quality Using Turbidity in the Cuyahoga River, Cuyahoga Valley National Park, Ohio, 2004–7

By Amie M.G. Brady[1], Rebecca N. Bushon[1], and Meg B. Plona[2]

Abstract

The Cuyahoga River within Cuyahoga Valley National Park (CVNP) in Ohio is often impaired for recreational use because of elevated concentrations of bacteria, which are indicators of fecal contamination. During the recreational seasons (May through August) of 2004 through 2007, samples were collected at two river sites, one upstream of and one centrally-located within CVNP. Bacterial concentrations and turbidity were determined, and streamflow at time of sampling and rainfall amounts over the previous 24 hours prior to sampling were ascertained. Statistical models to predict *Escherichia coli* (*E. coli*) concentrations were developed for each site (with data from 2004 through 2006) and tested during an independent year (2007). At Jaite, a sampling site near the center of CVNP, the predictive model performed better than the traditional method of determining the current day's water quality using the previous day's *E. coli* concentration. During 2007, the Jaite model, based on turbidity, produced more correct responses (81 percent) and fewer false negatives (3.2 percent) than the traditional method (68 and 26 percent, respectively). At Old Portage, a sampling site just upstream from CVNP, a predictive model with turbidity and rainfall as explanatory variables did not perform as well as the traditional method. The Jaite model was used to estimate water quality at three other sites in the park; although it did not perform as well as the traditional method, it performed well—yielding between 68 and 91 percent correct responses. Further research would be necessary to determine whether using the Jaite model to predict recreational water quality elsewhere on the river would provide accurate results.

[1] U.S. Geological Survey.
[2] Cuyahoga Valley National Park.

Introduction

The Cuyahoga Valley National Park (CVNP) is a valuable national resource, encompassing more than 33,000 acres of green space between the cities of Akron and Cleveland, Ohio (fig. 1). Nearly 3 million people visit the park annually, making it a major destination and recreational resource within the Lake Erie watershed. In all, 22 mi of the Cuyahoga River and more than 190 mi of tributary streams are within the park boundaries. The Cuyahoga River drains approximately 400 mi^2 prior to entering the park, and another 300 mi^2 of drainage area are added before the river leaves the park. However, because of variable water quality, recreation in the river, including canoeing, swimming, and wading, currently (2009) is discouraged.

The Ohio Environmental Protection Agency (Ohio EPA) has established minimum water-quality requirements for all state surface waters based on assigned use designations. These requirements are imposed to "protect public health and welfare; and to enhance, improve and maintain water quality" (Ohio Environmental Protection Agency, 2008). The Cuyahoga River within the park has been assigned a use designation of primary-contact recreation, meaning the river is suitable for full-body contact such as swimming and canoeing. However, previous studies within the park have found concentrations of *Escherichia coli* (*E. coli*) and fecal coliforms in the river and its tributaries above the Ohio EPA standards (Myers and others, 1998; Bushon and Koltun, 2003; Brady, 2007). Potential sources of contamination to the river have been identified and include combined sewer overflows and other upstream sources such as poorly functioning septic systems, farm runoff, and urban stormwater runoff.

Park managers want to promote the use of the river when the water quality is acceptable. Traditional methods used to monitor the concentrations of fecal-indicator bacteria in the water take at least 18 hours to obtain results. The elapsed time between the occurrence of elevated fecal-indicator bacteria (such as *E. coli* and fecal coliforms) concentrations in recreational waters and their detection is too long to assess water quality and take adequate control measures.

This is especially true in the river, where decay, dilution, dispersion, and transport of fecal-indicator bacteria can cause concentrations to change greatly over short periods of time (Myers and others, 1998). Predictive models have been shown to work well in other systems (USGS BacteriALERT Program, *http://ga2.er.usgs.gov/bacteria/default.cfm*; Francy and Darner, 2007; Francy and others, 2003; Nevers and Whitman, 2006; Nevers and others, 2007), and may be able to provide reliable results of the current day's fecal-indicator bacteria concentrations in the Cuyahoga River, allowing park managers to determine whether they might encourage river use.

During 2004 through 2007, the U.S. Geological Survey (USGS), in cooperation with the National Park Service (NPS), conducted a series of studies within the CVNP to develop predictive models to provide near-real-time (within a couple hours of sample collection) estimates of *E. coli* concentrations in the Cuyahoga River. Water-quality and environmental data were collected to facilitate model development. This report describes how the predictive models were developed and how well the models predicted *E. coli* concentrations in the river.

Methods

Samples were collected 3 to 7 days per week, and a replicate sample was collected once per week during the recreational seasons (May through August) of 2004 through 2007. On the basis of previous water-quality sampling within the park, sites at Old Portage and Jaite were selected for further investigation and as potential locations for development of predictive models (fig. 1). Both sites are accessible to pedestrians and have a bridge over the river from which to sample. Old Portage is just upstream from the park (USGS station number 04206000), and Jaite is centrally located within the park (USGS station number 411747081341300). During the recreational seasons of 2004 and 2005, periodic samples were collected at two other sites along the river—Bath and Independence. Bath (USGS station number 04206200) is downstream from Old Portage and the Akron Water Pollution Control Station but upstream from Jaite and Independence. Independence (USGS station number 04208000) is at the downstream edge of the park and was sampled only during 2005. At each sampling site, grab samples were collected from the center of the river with a weighted sampler containing a sterile 1-L bottle. Samples were kept in the dark at 4°C until they were analyzed.

E. coli concentrations were used to monitor recreational water quality. The Ohio primary-contact single-sample standard for *E. coli* of 298 colony-forming units per 100 milliliters (CFU/100 mL) was used as a benchmark to evaluate water quality in this study and will be herein referred to as the "water-quality standard" (WQS). In order to meet the recreational-use designation, this standard cannot be exceeded in more than 10 percent of the water samples collected during any 30-day period (Ohio Environmental Protection Agency, 2008).

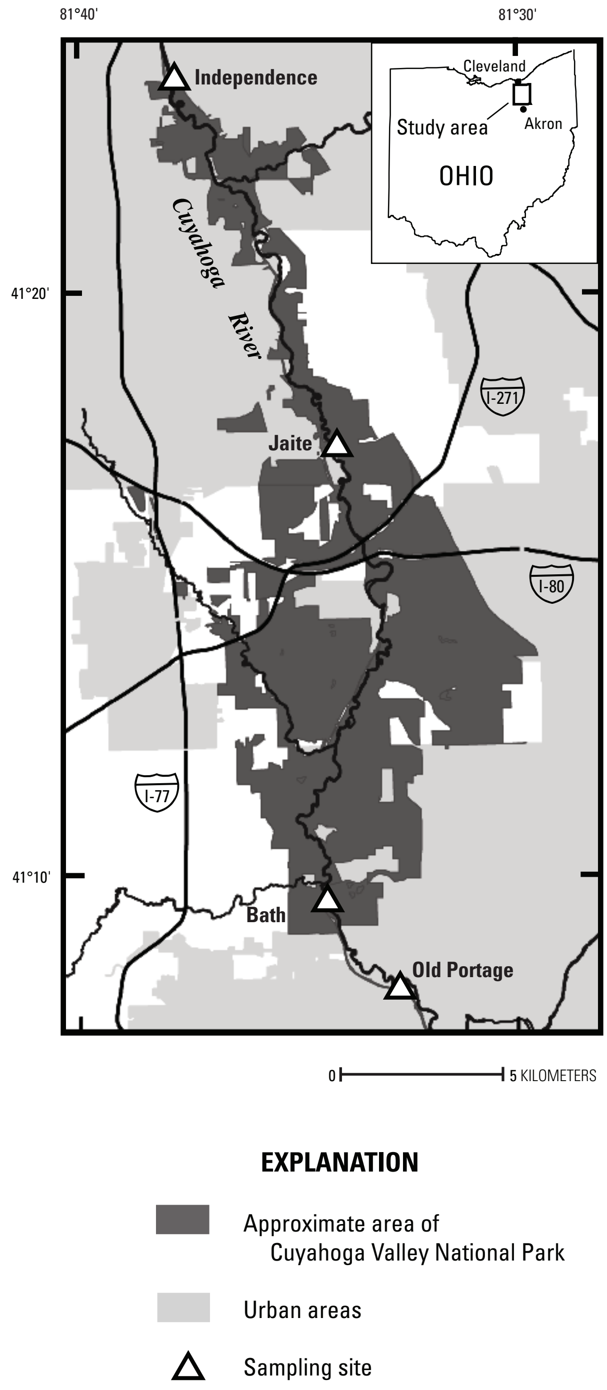

Figure 1. Location of study sites, Cuyahoga Valley National Park, northeastern Ohio. The Cuyahoga River flows north through the park.

Water-Quality Sample Analyses

Turbidity was determined in the laboratory with a Hach Model 2100AN turbidimeter (Hach Company, Loveland, Colo.) and was reported in nephelometric turbidity ratio units (NTRU). Standard membrane-filtration techniques were used to determine *E. coli* concentrations by using modified mTEC agar (U.S. Environmental Protection Agency, 2006). An aliquot of sample was run through a 0.45-µm filter onto which bacteria were concentrated. The filter was transferred to a modified mTEC agar plate and incubated at 35°C for 2 hours and then transferred to a 44.5°C incubator for another 22 hours. Magenta-colored colonies were counted as *E. coli*.

Environmental Data Collection

Streamflow data were collected for the Old Portage site from the nearby streamgage. Because there is no active streamgage near the Jaite site, streamflow was determined by using standard USGS techniques based on river stage (Rantz and others 1982). Rainfall information was obtained from the Automated Flood Warning System (AFWS), a network of rain gages supported by numerous communities and State and Federal agencies. Data are available online at (*www.afws.net*). On the day of sample collection, the total rainfall over the previous 24 hours at AFWS sites closest to the sampling sites was recorded. For the Jaite sampling site, the Macedonia AFWS site, approximately 3 mi northeast of Jaite, was used (AFWS site ID 8431). For the Old Portage and Bath sampling sites, the Bath AFWS site, approximately 4.5 mi northwest of Old Portage and 2.5 mi west of Bath, was used (AFWS site ID 8433). The Walton Hills AFWS site, approximately 3 mi southeast of Independence, (AFWS site ID 8314) was used for the Independence sampling site.

Quality-Control and Quality-Assurance Practices

Quality-control and quality-assurance (QA/QC) procedures were adopted as set forth by the American Public Health Association and others (1998), Britton and Greeson (1989), and Francy and others (2007). Several QA/QC procedures were performed for the membrane-filtration method. These included field, filter, and procedure blanks and replicate samples as described in Francy and others (2007). Field blanks—sterile buffered water transferred in the field and treated as a sample—were collected once during every recreational season. Every day that samples were processed, a filter blank—which was an aliquot of sterile buffered water filtered prior to sample analysis—was completed. Each week that samples were analyzed, a procedure blank—which was an aliquot of sterile buffered water filtered after sample analysis—and a replicate sample were analyzed. If either a filter blank or procedure blank showed bacterial growth, immediate action was taken to resolve the issue. A new buffer lot or improved rinsing between filtered aliquots reconciled these issues. All replicate samples agreed within 10 percent, indicating that 10 percent or less of the variability of the sample results was due to analytical and(or) sampling variability. (The agreement, or percent difference, was calculated as the difference of the log-transformed replicate pairs divided by the log of the average of the two pairs and multiplied by 100.)

Turbidity was analyzed in duplicate for each sample. If results for subsamples did not agree within 10 percent of the average of the subsamples, samples were reanalyzed until results agreed for two consecutive subsamples. Reported sample turbidities were an average of these two results.

Statistical Methods

Correlation analysis was performed to determine the strength of the relation between two variables. Pearson's correlation coefficient (r) was calculated. Linear regression analysis was used to estimate *E. coli* concentrations in the river based on one explanatory variable (simple linear regression, or SLR) or more variables (multiple linear regression, or MLR). For model selection, the coefficient of determination (R^2) and the Mallows' Cp statistic were used to narrow the subset of all possible models. The methods utilized in model development and selection steps are further described in Francy and Darner (2006). The alpha (α) level for all tests of significance in this report was set at 0.05.

The effectiveness of the models was based on several response statistics. The predicted *E. coli* concentration for each day was examined and compared to the actual concentration as determined by the membrane-filtration technique. Rates of correct predictions above or below the WQS (298 CFU/100 mL), false positives, and false negatives were calculated. False positives are defined as model predictions above the WQS when the actual concentrations were below the WQS; false negatives are defined as model predictions below the WQS when the actual concentrations were above the WQS. Model sensitivity and specificity were also determined. Sensitivity is the proportion of correctly predicted exceedances out of all exceedances. Specificity is the proportion of correctly predicted nonexceedances out of all nonexceedances.

The traditional method for determining the current day's water quality is to use the previous day's *E. coli* concentration. By examining this data in comparison to the actual concentrations, response statistics as described above could be calculated. Because samples were not collected every day throughout the study period, this dataset was smaller than the dataset for the models.

To determine whether the Jaite model could predict recreational water quality at other sites along the river, another method to determine exceedances of the WQS—the threshold-probability method—was used. Briefly, the t-distribution was used to test the hypothesis that the model-predicted value is greater than the WQS. The t-value is determined as:

$$t=[Y' - \log_{10}(298)]/SEP$$

where Y′ is the predicted concentration in $\log_{10}$ units and SEP is the standard error of the prediction. The computed t-value is used along with the number of degrees of freedom for the regression model to determine the one-tailed probability that Y′ is greater than the WQS (298 CFU/100 mL). This method is described in more detail elsewhere (Francy and others, 2003; Francy and Darner, 2006).

A scatterplot of the actual *E. coli* concentrations (x-axis) in comparison to the probability of exceeding the WQS (y-axis) is examined. A vertical line representing the WQS is also plotted. The analyst can then use a horizontal line to determine a threshold probability, above which model predictions are considered exceedances of the WQS. The threshold probability is decided upon at a probability that maximizes the number of correct responses and (or) minimizes the number of false negative responses. Response statistics were calculated as described previously.

In terms of public health protection, the frequency of false negatives and the sensitivity of the model were the principal 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 deterred from recreating on the river. Therefore, when deciding which models to implement at the park, all of the response statistics were examined and compared.

Relation of *Escherichia coli* Concentrations to Environmental Variables

During the recreational seasons of 2004 through 2007, 211 samples were collected at Jaite, and 180 samples were collected at Old Portage. Concentrations of *E. coli* ranged from 67 to 46,000 CFU/100 mL at Jaite and from 110 to 49,000 CFU/100 mL at Old Portage. Summary statistics for *E. coli* concentrations in samples, rainfall in the 24-hour period prior to sample collection, streamflow at time of sampling, and sample turbidity are listed for both sites in table 1.

Log_{10} *E. coli* concentrations at both sites were significantly related to rainfall (table 2). Log_{10} turbidities were strongly correlated to *E. coli* concentrations at Jaite (fig. 2A). Although the correlation between turbidities and *E. coli* concentrations was significant at Old Portage, the relation was weaker than at Jaite (fig. 2B). Streamflow was related to *E. coli* concentrations at Jaite ($r = 0.57$ for all years), but not at Old Portage ($r = 0.12$ for all years). The highest *E. coli* concentrations at Old Portage, however, were observed with relatively low corresponding streamflow at time of sampling (fig. 3), implying that there may be sources of *E. coli* that affect concentrations at this site that are not related to hydrologic factors.

The relation between the current day's and the previous day's measured $\log_{10}$ *E. coli* concentrations were significant at both sites ($r = 0.43$ for Jaite and $r = 0.38$ for Old Portage for all years) (table 2). For the predictive models, this variable was not considered for inclusion because ease-of-use was a major priority for an implemented model. Ideal models would be those that could be used without or with limited site visits; that is, model parameters that could be assessed via online sources, for example. By including this variable, daily samples analyzed for *E. coli* concentrations would be necessary to predict current conditions and may not be readily available every day; for example, daily samples may be difficult to obtain during a weekend or holiday when staff are limited or occupied with other duties.

Table 1. Summary statistics for results of sample analyses and measured environmental parameters at time of sampling at Jaite and Old Portage, Cuyahoga Valley National Park, Ohio, for the recreational seasons (May through September) of 2004 through 2007. Rainfall previous 24 hrs. is defined as the rainfall during the 24 hours prior to sample collection.

[*E. coli*, *Escherichia coli*; CFU/100 mL, colony-forming units per 100 milliliters; ft^3/s, cubic feet per second; NTRU, nephelometric turbidity ratio units]

Variable	Number of samples	Minimum	Median	Maximum	Mean
			Jaite		
E. coli, in CFU/100 mL	211	67	420	46,000	2,300
Rainfall previous 24 hours, in inches	210	0	0	2.5	0.20
Streamflow, in ft^3/s	195	171	412	1,900	563
Turbidity, in NTRU	211	4.1	18	1,200	56
			Old Portage		
E. coli, in CFU/100 mL	180	110	500	49,000	1,800
Rainfall previous 24 hours, in inches	172	0	0	1.48	0.12
Streamflow, in ft^3/s	180	50	186	3,510	387
Turbidity, in NTRU	180	3.9	9.2	100	12

Table 2. Pearson's r correlation coefficients of environmental variables to $\log_{10}$-transformed *Escherichia coli* (*E. coli*) concentrations by recreational seasons (May through September) as indicated by year. [Relations significant at $\alpha = 0.05$ are indicated in bold. The number of samples for each relation are indicated in parentheses. Rainfall previous 24 hours is defined as the rainfall during the 24 hours prior to sample collection]

[NTRU, nephelometric turbidity ratio units; ft^3/s, cubic feet per second; *E. coli*, *Escherichia coli*; CFU/100 mL, colony-forming units per 100 milliliters]

Year	Rainfall previous 24 hours, in inches	Log_{10} turbidity, in NTRU	Streamflow, in ft^3/s	Previous day's Log_{10} *E. coli*, in CFU/100 mL
		Jaite		
2004–6	**0.59** (179)	**0.82** (180)	**0.56** (164)	**0.44** (147)
2007	**0.61** (31)	**0.87** (31)	**0.77** (31)	0.36 (19)
2004–7	**0.59** (210)	**0.83** (211)	**0.57** (195)	**0.43** (166)
		Old Portage		
2004–6	**0.52** (162)	**0.47** (170)	0.13 (170)	**0.37** (137)
2007	0.47 (10)	**0.71** (10)	0.03 (10)	0.68 (6)
2004–7	**0.51** (172)	**0.49** (180)	0.12 (180)	**0.38** (143)

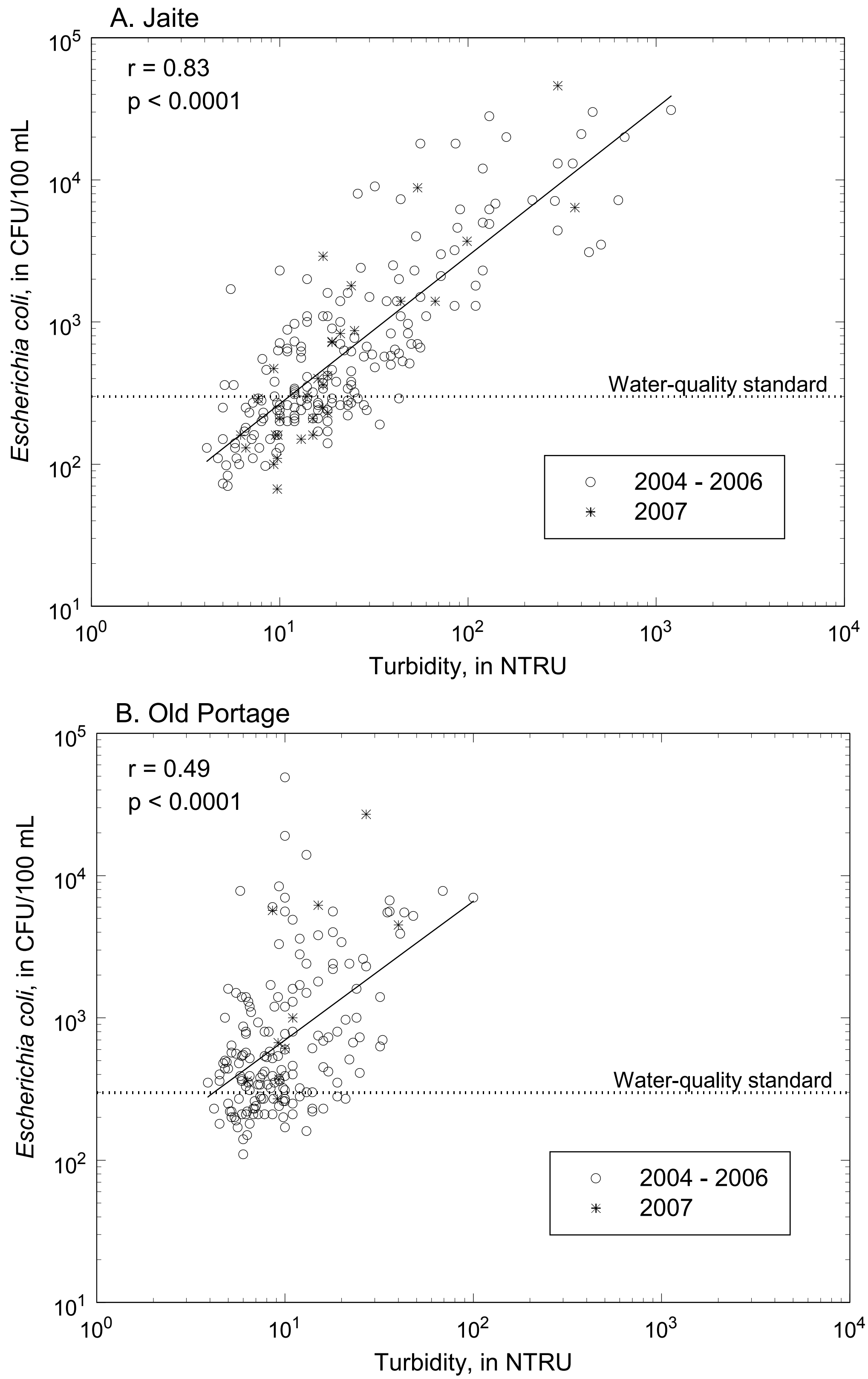

Figure 2. Relation between *Escherichia coli* concentrations and turbidity at A) Jaite and B) Old Portage, Cuyahoga Valley National Park, Ohio, during the recreational seasons (May through August) of 2004 through 2007. The solid line indicates the best-fit line. [r, Pearson's correlation coefficient; p, significance of the relation; CFU/100 mL, colony-forming units per 100 milliliters; NTRU, nephelometric turbidity ratio units]

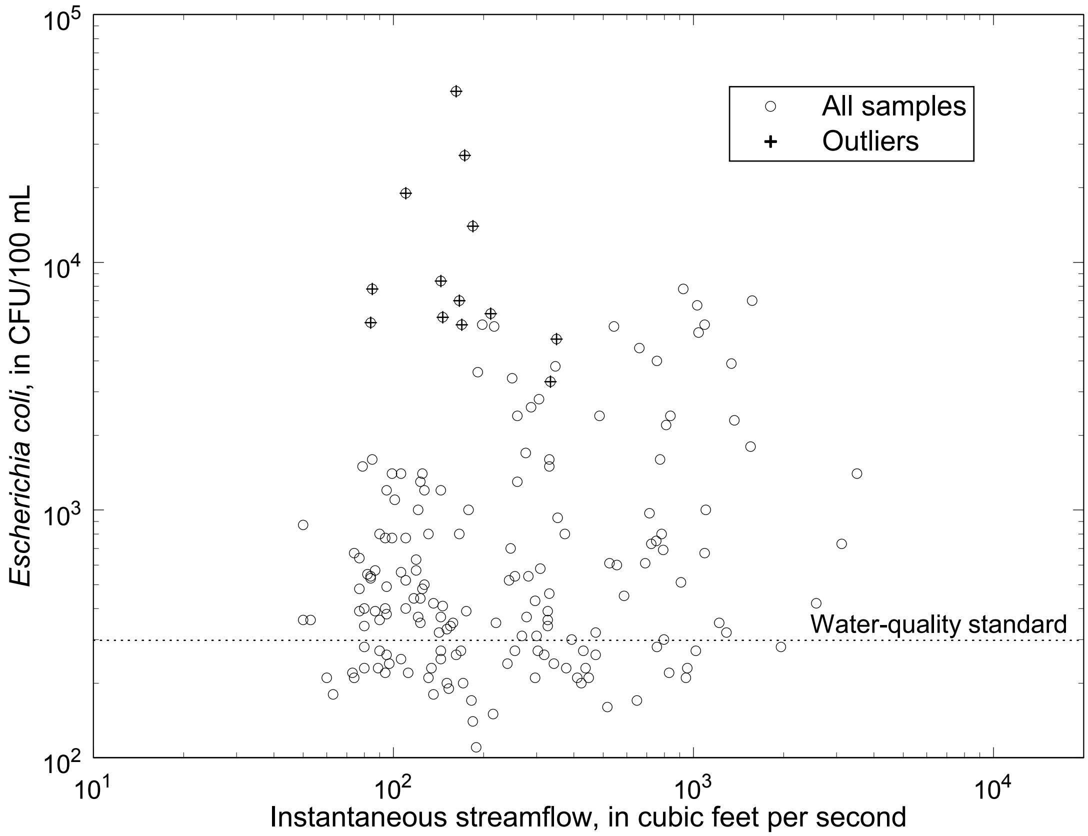

Figure 3. Relation between *Escherichia coli* concentrations and instantaneous streamflow at Old Portage, Cuyahoga Valley National Park, Ohio, during the recreational seasons (May through August) of 2004 through 2007. Samples where the actual concentrations were above the estimated upper confidence limit for the predicted concentrations based on the multiple linear regression model (outliers) are indicated. [CFU/100 mL, colony-forming units per 100 milliliters]

Predictive Models

Data collected during the recreational seasons of 2004 through 2006 were used to develop predictive models that were tested during the recreational season of 2007. Models were chosen on the basis of predictive capabilities and ease of use. Model statistics are listed in table 3.

A. Jaite

When examining all of the potential models using the model calibration data (data collected during the recreational seasons of 2004 through 2006), the SLR model based solely on $\log_{10}$ turbidity showed the most promise for use within the park. When rainfall and (or) discharge was included with turbidity, the parameter estimates for the additional variables were not statistically significant, indicating that these variables (singly or in combination) did not improve the model. As compared to the traditional method of determining the current day's conditions based on the previous day's bacterial count, the SLR turbidity model predicted conditions correctly more often: 77 percent compared to 74 percent of the time for the traditional method (table 4). Although the specificity of the model was lower than that of the traditional method, the sensitivity of the SLR model was greater, demonstrating that the model was better at predicting exceedances of the WQS than the traditional method.

Table 3. Statistics for models for the Jaite and Old Portage sites.

[<, less than; R^2, coefficient of determination; p, significance of the relation]

Site	Data used for model development	Adjusted R2 of model	Variables in model	Parameter estimate	p value
Jaite	2004–6	0.68	$\log_{10}$ turbidity	1.006	<0.0001
			y-intercept	1.433	<0.0001
Old Portage	2004–6	0.34	$\log_{10}$ turbidity	0.812	<0.0001
			rainfall	0.665	<0.0001
			discharge	-0.0002	0.0144
			y-intercept	2.006	<0.0001
Old Portage	2004–6	0.32	$\log_{10}$ turbidity	0.535	<0.0001
			rainfall	0.710	0.0003
			y-intercept	2.196	<0.0001
Jaite	2004–7	0.69	$\log_{10}$ turbidity	1.041	<0.0001
			y-intercept	1.382	<0.0001

During the recreational season of 2007, the previously developed model was tested. The SLR turbidity model performed better than the traditional method (table 4), correctly predicting water-quality conditions 81 percent of the time compared to 68 percent for the traditional method. The sensitivity was higher for the SLR model compared to the traditional method. These results support previous research that concentrations of bacteria in a river can change greatly over short periods of time (Myers and others, 1998) causing inaccurate predictions based on the previous day's concentration. The predicted *E. coli* concentrations from the SLR model compared to the actual *E. coli* concentrations are shown in Figure 4.

When all of the data collected during 2004 through 2007 were compiled, new models were created. Parameter estimates and response statistics for the SLR model (table 3 and table 4) changed only slightly with the additional data. Although the parameter estimate for rainfall was significant in the MLR model (with turbidity), the addition of this variable did not improve the response statistics compared to the SLR model. The parameter estimate for discharge was not significant in the MLR model (with turbidity). The SLR model yielded more correct responses than the traditional method. Although the traditional method had a somewhat higher specificity—68 percent compared to 58 percent for the SLR model—the SLR model had higher sensitivity: 89 percent compared to 77 percent for the traditional method.

Table 4. For the Jaite site, comparison of the percentage of correct and false positive and negative responses of recreational water-quality conditions using a regression model (with turbidity as the sole explanatory variable) and the traditional method for determining water quality (using the previous day's *E. coli* concentrations) during model development, testing, and revisions, Cuyahoga Valley National Park, Ohio.

[R^2, coefficient of determination; SLR, simple linear regression model; %, percent; —, not applicable]

	Number of samples	R^2	Correct responses	False positive	False negative	Specificity	Sensitivity
			Value (number of samples)				
Development—2004–6 data							
SLR model	180	0.68	77% (138)	18% (32)	5.6% (10)	54%	91%
Traditional method	147	—	74% (109)	13% (19)	13% (19)	66%	79%
Testing—results of 2007 data in above models							
SLR model	31	—	81% (25)	16% (5)	3.2% (1)	64%	94%
Traditional method	26	—	68% (13)	5.3% (8)	26% (5)	83%	62%
Revisions—2004–7 data							
SLR model	211	0.69	77% (162)	17% (35)	6.6% (14)	58%	89%
Traditional method	166	—	73% (122)	12% (20)	14% (24)	68%	77%

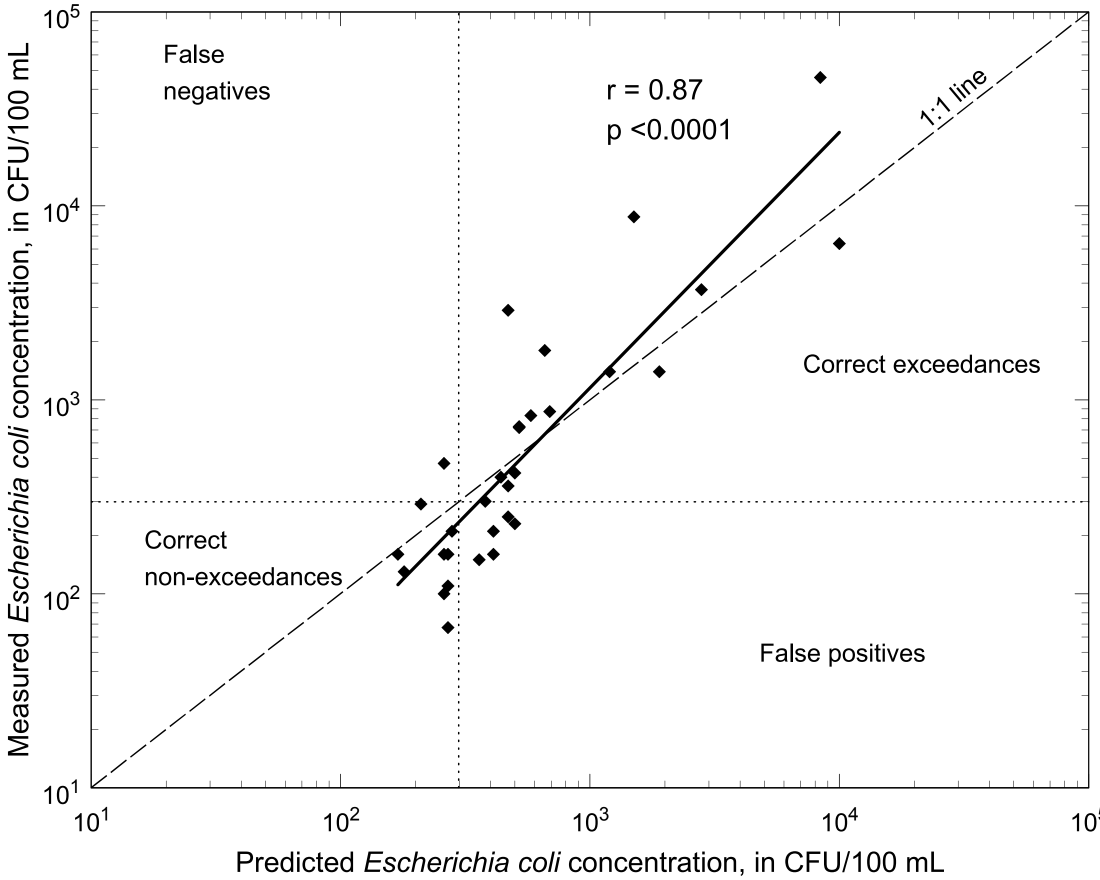

Figure 4. Predicted *Escherichia coli* concentrations based on the turbidity model developed using 2004 through 2006 recreational-season data compared to actual *Escherichia coli* concentrations at Jaite, Cuyahoga Valley National Park, Ohio, May through August, 2007. The dotted lines indicate the primary-contact recreation standard for a single sample (298 CFU/100 mL) as set by the Ohio Environmental Protection Agency (Ohio Environmental Protection Agency, 2008). The solid line indicates the best-fit line. [r, Pearson's correlation coefficient; p, significance of the relation; CFU/100 mL, colony-forming units per 100 milliliters]

B. Old Portage

The two models developed for the Old Portage site contained two variables (turbidity and rainfall) or three variables (turbidity, rainfall, and discharge) (table 3). For models developed and tested using the data from the recreational seasons of 2004 through 2006, there were fewer correct responses (72 percent) compared to the traditional method (81 percent) (table 5). Further, the sensitivities of the models were 100 percent because the models predicted all samples to be above the standard (WQS), whereas the traditional method had a lower sensitivity (87 percent) but a higher specificity (64 percent compared to 0 percent for both models). When the models were tested in 2007, the models again predicted all of the samples above the standard.

New models were developed using all of the data (2004–7), but very similar results were found. The models predicted the majority or all of the samples above the standard. For one model, two predictions were below the standard, but these were both false negatives. Although the traditional method had a higher percentage of false negatives, this method had higher correct responses and fewer false positives.

As shown in figure 3, the highest *E. coli* concentrations were observed when streamflow was relatively low. Outliers were defined as samples with actual *E. coli* concentrations above the upper 90-percent confidence limit of the predicted concentration produced by the MLR models. The rainfall amounts and turbidity values for these samples ranged widely but had relatively low medians (0 to 0.36 in. with a median of 0.08 in., and 5.8 to 27 NTRU with a median of 10 NTRU, respectively). This information supports the previously mentioned hypothesis that sources of *E. coli* to the river upstream of this site may not be affected by hydrologic factors; instead, they might be affected by industrial and domestic discharges or broken sewerlines. These sources may also have contributed bacteria to the river at other times throughout the study but bacteria may not have been observed when streamflow was greater because of dilution.

Table 5. For the Old Portage site, comparison of the percent of correct and false positive and negative responses of recreational water-quality conditions using regression models with indicated variables and the traditional method for determining water quality (using the previous day's *E. coli* concentrations) during model development, testing, and revisions, Cuyahoga Valley National Park, Ohio.

[R^2, coefficient of determination; MLR, multiple linear regression model; %, percent; —, not applicable]

	Number of samples	R^2	Correct responses	False positive	False negative	Specificity	Sensitivity
			Value (number of samples)				
Development — 2004–6 data							
MLR1 turbidity, rainfall, discharge	162	0.34	72% (117)	28% (45)	0% (0)	0%	100%
MLR2 turbidity, rainfall	162	0.32	72% (117)	28% (45)	0% (0)	0%	100%
Traditional method	137	—	81% (111)	9.5% (13)	9.5% (13)	64%	87%
Testing — results of 2007 data in above models							
MLR1 turbidity, rainfall, discharge	10	—	90% (9)	10% (1)	0% (0)	0%	100%
MLR2 turbidity, rainfall	10	—	90% (9)	10% (1)	0% (0)	0%	100%
Traditional method	8	—	75% (6)	12.5% (1)	12.5% (1)	0%	86%
Revisions — 2004–7 data							
MLR1 turbidity, rainfall, discharge	172	0.35	72% (124)	27% (46)	1.2% (2)	0%	98%
MLR2 turbidity, rainfall	172	0.32	73% (126)	27% (46)	0% (0)	0%	100%
Traditional method	145	—	81% (117)	9.6% (14)	9.6% (14)	62%	87%

C. Jaite Model Results at Other River Locations

Because the Jaite SLR turbidity model performed well at determining water-quality conditions at Jaite and the Old Portage models did not perform well, the Jaite SLR model was used to determine whether it could predict water-quality conditions elsewhere along the river. Overlapping sampling dates (days samples were collected at Jaite and at another site) provided a total of 171, 114, and 44 samples at Old Portage, Bath, and Independence, respectively (fig. 1). Model performance for these additional sites was determined in two ways. First, the predicted *E. coli* concentrations from the Jaite SLR model were compared to the actual concentrations at the additional sites as described previously (predicted-*E. coli* method). Second, the probability of exceeding the WQS was compared to a threshold probability (threshold-probability method).

For the three additional sites, the regression-based prediction methods provided variable results (table 6). For Independence (fig. 5), the threshold probability (70 percent) was much higher than that at both Old Portage and Bath (30 percent and 40 percent, respectively). The response statistics for the probability method at Independence were better (more correct responses, fewer false positives, and higher specificity) compared to the predicted-*E. coli* method and the traditional method. Although the threshold-probability method provided more correct responses, the specificity was lower at Old Portage and Bath compared to the predicted-*E. coli* method. Further, the traditional method provided more correct responses at these two sites than either of the model-based prediction methods.

Table 6. For the additional sites (Old Portage, Bath, and Independence), comparison of the percent of correct and false positive and negative responses of recreational water-quality conditions using the results of the Jaite simple linear regression model based solely on turbidity and the traditional method for determining water quality (using the previous day's *E. coli* concentrations), Cuyahoga Valley National Park, Ohio. Two methods of determining how well the model performed were used: the predicted *Escherichia coli* (*E. coli*) concentration and the probability of exceeding the single-sample maximum in relation to a threshold probability were compared to the actual *E. coli* concentration.

[SLR, simple linear regression modell; %, percent; —, not applicable]

	Number of samples	R^2	Correct responses	False positive	False negative	Specificity	Sensitivity
			Value (number of samples)				
Old Portage — 2004–7 data compared to Jaite turbidity model (2004–7) results							
Jaite SLR—predicted *E. coli*	171	—	70% (120)	14% (24)	16% (27)	47%	78%
Jaite SLR—probability	171	30%	73% (125)	20% (35)	6.4% (11)	22%	91%
Traditional method at Old Portage	145	—	81% (117)	9.6% (14)	9.6% (14)	62%	87%
Bath — 2004–5 data compared to Jaite turbidity model (2004–7) results							
Jaite SLR—predicted *E. coli*	114	—	68% (77)	12% (14)	20% (23)	50%	73%
Jaite SLR—probability	114	40%	74% (85)	16% (18)	9.6% (11)	36%	87%
Traditional method at Bath	95	—	81% (77)	10% (9)	10% (9)	88%	59%
Independence — 2005 data compared to Jaite turbidity model (2004–7) results							
Jaite SLR—predicted *E. coli*	44	—	73% (32)	20% (9)	6.8% (3)	68%	81%
Jaite SLR—probability	44	70%	91% (40)	2.3% (1)	6.8% (3)	96%	81%
Traditional method at Independence	24	—	88% (21)	4.2% (1)	8.3% (2)	93%	80%

Model Applications and Suggestions for Future Work

Although the model results are not as reliable as results of the traditional method for Old Portage and Bath, the model is based on one site and one variable and therefore is an easier and quicker method than sampling several sites along the river. Further research and validation of this method would be needed before using it to estimate bacterial concentrations along the Cuyahoga River within CVNP. If this method is pursued and proves reliable, only Jaite would need to be sampled. A turbidity sensor could be used to provide real-time information that would further minimize the amount of staff hours required.

Implementation of the model for predicting recreational water quality at Jaite may begin during the recreational season of 2009. Model results are expected to be available daily via the World Wide Web. Future studies at the CVNP could include an examination of an in situ turbidity probe that could provide real-time water-quality estimates. However, a closer examination of the relation between model predictions at Jaite and actual concentrations of *E. coli* at other sites along the river would need to be completed prior to use of this method in the CVNP. Updates to water-treatment facilities and wastewater conveyance systems may affect water quality in the river and consequently the ability of the model to correctly predict conditions in the river. Therefore, continued water-quality monitoring of the river would be necessary to maintain the model's current predictive capabilities.

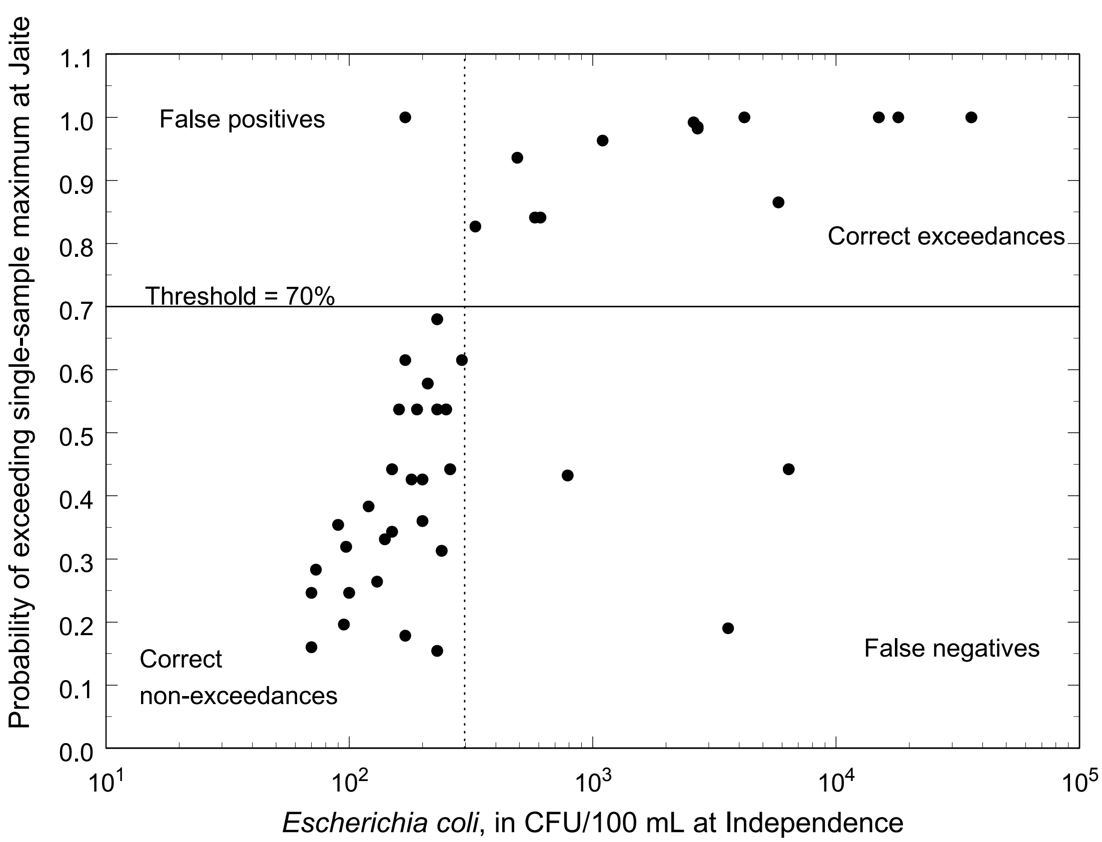

Figure 5. Relation between actual *Escherichia coli* concentrations at Independence and probability of exceeding the primary-contact recreation maximum at Jaite using the turbidity model developed using 2004 through 2007 recreational-season data, Cuyahoga Valley National Park, Ohio. The dotted line indicates the primary-contact recreation maximum for a single sample (298 CFU/100 mL) (WQS) as set by the Ohio Environmental Protection Agency (Ohio Environmental Protection Agency, 2008)

Summary and Conclusions

The Cuyahoga River within the Cuyahoga Valley National Park (CVNP) is often impaired for recreational use due to concentrations of *E. coli*, an fecal-indicator bacterium, above the Ohio primary-contact single-sample standard. During the recreational seasons (May through August) of 2004 through 2007, the USGS, in cooperation with the National Park Service and the Ohio Lake Erie Commission, conducted a series of studies within CVNP to develop predictive models to provide near-real-time estimates of *E. coli* concentrations in the Cuyahoga River.

Predictive models have been demonstrated to work well to determine same-day water quality at beaches (Francy and Darner, 2007; Francy and others, 2003; Nevers and Whitman, 2005; Nevers and others, 2007) and for river sites (USGS BacteriALERT Program, *http://ga2.er.usgs.gov/bacteria/default.cfm*). In this study, regression models were developed for Jaite and Old Portage, two sites on the Cuyahoga River within and just upstream of, respectively, CVNP to predict same-day recreational water quality. The model results were compared to sampling results from use of the membrane-filtration technique and to results from the traditional method for determining recreational water quality (using the membrane-filtration bacteria count from the previous day). The models were based on easily measured variables that can provide results in less than 1 hour.

The model developed for Jaite performed better than the traditional method and was able to provide same-day water-quality information. This model, based solely on turbidity, correctly predicted recreational water quality above or below the primary-contact single-sample standard set by the Ohio Environmental Protection Agency (WQS, 298 CFU/100 mL) in 81 percent of the samples collected during an independent year (2007). This correct prediction percentage is similar to what was found in earlier studies at two Lake Erie beaches—83 and 82 percent for Huntington and Edgewater beaches, respectively (Francy and Darner, 2007). Based on the model developed from all years of data collection at Jaite, a turbidity measurement greater than 11 NTRU will result in a model prediction greater than the WQS.

The two models developed for the second site just upstream of CVNP, Old Portage, did not perform as well as the traditional method. The models were able to correctly predict exceedances of the standard in 72 or 73 percent of the samples as compared to 81 percent for the traditional method. Elevated concentrations of *E. coli* were found when streamflow in the river was relatively low, indicating that some sources of bacteria may not be influenced by hydrologic factors. These sources could be point sources that are not directly associated with precipitation such as industrial and domestic discharges or broken sewerlines. Because of this unknown and possibly sporadic source(s) of bacteria, the models developed were unable to accurately predict water quality at this site.

The Jaite turbidity model was also used to predict water quality at other sites along the river. For the sites upstream from Jaite, Old Portage and Bath, the Jaite model did not predict water quality as well as the traditional method. However, the predictions were correct 73 percent and 74 percent of the time for Old Portage and Bath, respectively. For Independence, the site downstream from Jaite, the predictions were correct 91 percent of the time compared to 88 percent of the time for the traditional method. The correct percentages are relatively high for all three sites, and this method could be explored further to determine if predictions at one site can be used to predict water quality at other locations on the river.

Implementation of the model for predicting recreational water quality at Jaite may begin during the recreational season of 2009. If it does, model results may be available four days per week on the Ohio Nowcast website (*www.ohionowcast.info*). Future studies at the CVNP could include an examination of an *in situ* turbidity probe that could provide real-time water-quality estimates.

Acknowledgments

The authors acknowledge the assistance of all of the USGS and NPS student interns: James Cirreddu, Suzanna Dzejachok, Anthony Kinyo, Nickolas Fath, and Caitlin Splawski. This project was completed with support in part from the USGS/NPS Water-Quality Assessment and Monitoring Program and from the Ohio Lake Erie Commission.

References Cited

American Public Health Association, American Water Works Association, and Water Pollution Control Federation, 1998, Standard methods for the analysis of water and wastewater (20th ed.): Washington D.C., American Public Health Association [variously paged].

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.

Britton, L.J., and Greeson, P.E., eds., 1989, Methods for collection and analysis of aquatic biological and microbiological samples: U.S. Geological Survey Techniques of Water-Resources Investigations 5–A4, 363 p.

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.

Francy, D.S., Bushon, R.N., Bertke, E.E., Brady, A.M.G., Kephart, C.M., Likirdopulos, C.A., and Stoeckel, D.M., 2007, Quality assurance/quality control manual: Ohio Water Microbiology Laboratory, accessed July 2008 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 6–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.

Francy, D.S., Gifford, A.M., and Darner, R.A., 2003, *Escherichia coli* at Ohio beaches—Distribution, sources, wastewater indicators, and predictive modeling: U.S. Geological Survey Water-Resources Investigations Report 02–4285, 120 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..

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.

Ohio Environmental Protection Agency, 2008, Water Quality Standards: Ohio Administrative Code, chap. 3745–1, January 23, 2008, p. 01:1, 07:1–8, 07:23, and 26:6.

Rantz, S.E., and others, 1982, Measurement and computation of streamflow; vol. 1, Measurement of stage and streamflow: U.S. Geological Survey Water-Supply Paper 2175, 631 p.

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].

Publication support was provided by the USGS Columbus Publishing Service Center.

Brady and others—**Predicting Recreational Water Quality Using Turbidity in the Cuyahoga River, Cuyahoga Valley National Park, Ohio, 2004–7**—SIR 2009–5192