In cooperation with the
Ohio Water Development Authority, Northeast Ohio Regional Sewer District,
Ohio Lake Erie Office, Cuyahoga County Board of Health, Cuyahoga County Sanitary Engineers,
and Cuyahoga River Community Planning Organization

Escherichia coli at Ohio Bathing Beaches—Distribution, Sources, Wastewater Indicators, and Predictive Modeling

Water-Resources Investigations Report 02–4285

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

A

B

C

D

A. Geese at Mosquito Lake, Cortland, Ohio. (Photo by Ted Smith, Trumbull County Health Department)
B. USGS scientist digging holes for collection of swash-zone materials at Edgewater Park, Cleveland, Ohio.
C. Rain gage at Huntington Reservation, Bay Village, Ohio.
D. Sample locations 3 and 6 feet inland from the outer edge of the swash zone.

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

Escherichia coli at Ohio Bathing Beaches—Distribution, Sources, Wastewater Indicators, and Predictive Modeling

By Donna S. Francy, Amie M. Gifford, and Robert A. Darner

Water-Resources Investigations Report 02-4285

In cooperation with the Ohio Water Development Authority, Northeast Ohio Regional Sewer District, Ohio Lake Erie Office, Cuyahoga County Board of Health, Cuyahoga County Sanitary Engineers, and Cuyahoga River Community Planning Organization

U.S. Department of the Interior
GALE A. NORTON, Secretary

U.S. Geological Survey
Charles G. Groat, Director

For additional information write to

District Chief
U.S. Geological Survey
6480 Doubletree Avenue
Columbus, OH 43229-1111

Copies of this report can be purchased from
U.S.Geological Survey
Information Services
Box 25286
Denver, CO 80225-0286

Columbus, Ohio
2003

CONTENTS

ILLUSTRATIONS

TABLES

CONVERSION FACTORS AND ABBREVIATED WATER- AND SEDIMENT-QUALITY UNITS

Multiply	By	To obtain
micrometer (µm)	0.00003937	inch
millimeter (mm)	0.03937	inch
centimeter (cm)	0.3937	inch
foot (ft)	0.3048	meter
mile (mi)	1.609	kilometer
milliliter (mL)	0.06102	cubic inch
liter (L)	0.03531	cubic foot
gram (g)	0.03527	ounce

Temperature: Temperature is given in degrees Celsius (°C) which can be converted to degrees Fahrenheit (°F) by use of the following equation:

$$°F=1.8(°C) + 32$$

Abbreviated water- and sediment-quality units used in this report: Chemical concentrations in water are reported in milligrams per liter (mg/L) and micrograms per liter (µg/L). Both units express the concentration of chemical constituents as weight (milligrams or micrograms) of chemical per unit volume (liter) of water. Turbidity is reported in Nephelometric Turbidity Units (NTU).

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

Concentrations of bacteria in sediment are reported in colonies per gram of (dry weight) sediment (col/g_{dw}).

Escherichia coli at Ohio bathing beaches—Distribution, sources, wastewater indicators, and predictive modeling

By Donna S. Francy, Amie M. Gifford, and Robert A. Darner

Abstract

Results of studies during the recreational seasons of 2000 and 2001 strengthen the science that supports monitoring of our Nation's beaches. Water and sediment samples were collected and analyzed for concentrations of *Escherichia coli* (*E. coli*). Ancillary water-quality and environmental data were collected or compiled to determine their relation to *E. coli* concentrations. Data were collected at three Lake Erie urban beaches (Edgewater, Villa Angela, and Huntington), two Lake Erie beaches in a less populated area (Mentor Headlands and Fairport Harbor), and one inland-lake beach (Mosquito Lake).

The distribution of *E. coli* in water and sediments within the bathing area, outside the bathing area, and near the swash zone was investigated at the three Lake Erie urban beaches and at Mosquito Lake. (The swash zone is the zone that is alternately covered and exposed by waves.) Lake-bottom sediments from outside the bathing area were not significant deposition areas for *E. coli*. In contrast, interstitial water and subsurface sediments from near the swash zone were enriched with *E. coli*. For example, *E. coli* concentrations were as high as 100,000 colonies per 100 milliliters in some interstitial waters. Although there are no standards for *E. coli* in swash-zone materials, the high concentrations found at some locations warrant concern for public health.

Studies were done at Mosquito Lake to identify sources of fecal contamination to the lake and bathing beach. *Escherichia coli* concentrations decreased with distance from a suspected source of fecal contamination that is north of the beach but increased at the bathing beach. This evidence indicated that elevated *E. coli* concentrations at the bathing beach are of local origin rather than from transport of bacteria from sites to the north.

Samples collected from the three Lake Erie urban beaches and Mosquito Lake were analyzed to determine whether wastewater indicators could be used as surrogates for *E. coli* at bathing beaches. None of the concentrations of wastewater indicators of fecal contamination, including 3β-coprostanol and cholesterol, were significantly correlated (α=0.05) to concentrations of *E. coli*. Concentrations of the two compounds that were significantly correlated to *E. coli* were components of coal tar and asphalt, which are not necessarily indicative of fecal contamination.

Data were collected to build on an earlier 1997 study to develop and test multiple-linear-regression models to predict *E. coli* concentrations using water-quality and environmental variables as explanatory variables. The probability of exceeding the single-sample bathing-water standard for *E. coli* (235 colonies per 100 milliliters) was used as the model output variable. Threshold probabilities for each model were established. Computed probabilities that are less than a threshold probability indicate that bacterial water quality is most likely acceptable. Computed probabilities equal to or above the threshold probability indicate that the water quality is most likely not acceptable and that a water-quality advisory may be needed.

Models were developed at each beach, whenever possible, using combinations of 1997, 2000, and (or) 2001 data. The models developed and tested in this study were shown to be beach specific; that is, different explanatory variables were used to predict the probability of exceeding the standard at each beach. At Mentor Headlands and Fairport Harbor, models were not developed because water quality was generally good. At the three Lake Erie urban beaches, models were developed with variable lists that included the number of birds on the beach at the time of sampling, lake-current direction, wave height, turbidity, streamflow of a nearby river, and rainfall. The models for Huntington explained a larger percentage of the variability in *E. coli* concentrations than the models for Edgewater and Villa Angela. At Mosquito Lake, a model based on 2000 and 2001 data contained the explanatory variables rainfall,

number of dry days preceding a rainfall, date, wind direction, wind speed, and turbidity. Additional research could include testing the threshold probabilities assigned for these models in subsequent years and comparing the models' ability to predict recreational water quality to results from the current method—using antecedent *E. coli* concentrations. Each year the model is tested, new data can be added and model variables can be recalculated to determine whether the predictive ability improves with additional data.

Introduction

Water-resource managers and the scientific community have long recognized the need for improved monitoring methods to adequately protect public health at our Nation's beaches. Across the country, a total of 11,270 closings and advisories were issued during 2000 (Natural Resources Defense Council, 2001). Of these, 85 percent were based on monitoring that detected fecal-indicator bacteria levels exceeding water-quality standards. Even as more states begin to monitor their beaches, not all states have routine monitoring programs or have adopted the U.S. Environmental Protection Agency (USEPA) recommended fecal indicators for beach monitoring (Natural Resources Defense Council, 2001) —*Escherichia coli* (*E. coli*) or enterococci for fresh waters and enterococci for marine waters.

The USEPA, recognizing the problem and the inconsistency of monitoring methods, initiated the Beaches Environmental Assessment, Closure, and Health Program (BEACH), the goal of which is to reduce the risk of infection to users of the Nation's recreational waters (U.S. Environmental Protection Agency, 1998). As part of the BEACH program, USEPA published the "Beach Action Plan" to enable consistent management of recreational-water-quality programs and strengthen the science that supports such monitoring programs (U.S. Environmental Protection Agency, 1999a). The Beach Action Plan addresses three areas of scientific research: (1) water-quality indicators, (2) modeling and monitoring research, and (3) exposure and health-effects research.

A topic mentioned for exposure and health-effects research under the Beach Action Plan is to determine whether the swash zone may be conducive to the growth of bacterial pathogens and indicators. The swash zone is the zone that is alternately covered and exposed by waves and is an area where children commonly play in the sand. The water that occupies the spaces between the sand particles near or in this zone is often referred to as "interstitial water." Little is known about the concentrations of pathogens and indicators in interstitial waters or subsurface sediments collected from in or near the swash zones and potential effects on children and other sensitive populations.

Also mentioned in the Beach Action Plan is the goal to carry out research on the development of mathematical models to determine or predict recreational water quality. Current methods for assessing the recreational water quality are based on measured concentrations of fecal indicators. These methods take at least 24 hours to complete—too long a lapse between sampling and analytical results to be relevant to water-resource managers and the public. Water-quality conditions can change overnight, so a water-quality advisory may be issued when the recreational-water-quality standard is met or may not be issued when the standard is exceeded. Some beach managers post the beach with a water-quality advisory whereas other beach managers close the beach to swimming. Mathematical models based on water-quality and environmental surrogates or hydrodynamic processes may be able to provide an assessment of recreational water quality within a few hours.

The Beach Action Plan mentions that another way to identify risk before exposure takes place is to develop real-time or near-real-time (less than 2 hours) analytical methods. One possible method is the use of wastewater indicators of fecal contamination. These include such compounds as caffeine and coprostanol, which are present in human wastes and may be a suitable surrogate for the presence of *E. coli* and pathogens.

The U.S. Geological Survey (USGS), in cooperation with the Ohio Water Development Authority, Northeast Ohio Regional Sewer District, Ohio Lake Erie Office, Cuyahoga County Board of Health, Cuyahoga County Sanitary Engineers, and the Cuyahoga River Community Planning Organization, studied the use of predictive models and wastewater indicators as surrogates for *E. coli*. The USGS collaborated with The Ohio State University (OSU), Great Lakes Forecasting System (GLFS) in predictive model development. In addition, data were gathered on concentrations of *E. coli* in lake-bottom sediments and subsurface sediments collected near the swash zone and on fecal contaminant sources in an inland lake. This project addresses several of the USEPA research priorities described in the Beach Action Plan.

Purpose and scope

This report describes field studies done throughout the recreational seasons (May through August) of 2000 and 2001 at six public bathing beaches in Ohio. Four types of studies were done—distribution, source, spatial, and routine studies. During all studies, water and (or) sediment samples were analyzed for *E. coli* concentrations. Additional water samples were collected to determine whether wastewater indicators could be used as surrogates for *E. coli* concentrations at bathing beaches.

For the distribution, source, and spatial studies, the USGS collected water and sediment samples and analyzed them for *E. coli* during 1- to 4-day studies at three Lake Erie beaches and one inland beach. Ancillary data were collected to help understand patterns of *E. coli* concentrations. These data included particle-size distributions and organic carbon concentrations of sediments; and turbidity, specific conductance, and temperature of interstitial and bathing waters. Distribution studies were done at four beaches to compare *E. coli* concentrations in water and sediments from the bathing area to concentrations near the swash zone and outside the bathing area. Source studies were done to determine a possible source of *E. coli* contamination to the bathing beach at the inland lake. Spatial studies were done to aid the GLFS in the development of a hydrodynamic predictive model with two objectives: (1) collect data on the vertical distribution of *E. coli* in the water column outside the bathing area, and (2) provide a detailed characterization of the spatial and temporal distribution of *E. coli* in water and sediments within the bathing area during dry weather and after a significant rainfall.

Routine studies were designed to collect data with which to test an existing predictive *E. coli* model and to develop new models. For the routine studies, water samples were collected by local agencies at five Lake Erie beaches and at one inland lake 4 or 5 days a week throughout the season. Statistical methods were used to evaluate the relations between *E. coli* and environmental and water-quality variables. These variables were used to develop multiple-linear-regression models for predicting *E. coli* concentrations.

Related studies

Distribution of fecal-indicator bacteria in sediment. In an earlier study (Francy and Darner, 1998), the distribution of *E. coli* in lake-bottom sediments was investigated at three beaches in the Cleveland, Ohio, metropolitan area. Concentration patterns of *E. coli* indicated that short-term storage (less than 1 week) of *E. coli* in sediments may have occurred, although no evidence for long-term storage was found during the sampling period. The authors hypothesized that an increase in *E. coli* in bathing waters in the absence of rainfall may have been due to resuspension of *E. coli* from lake-bottom sediments in the bathing area or from deeper sediments outside the area.

In a study done to determine sources of enterococci contamination to California beaches (Grant and others, 2001), bottom-sediment samples were collected from marsh and surf zones. Nineteen percent of the sediment samples from the marsh were positive for enterococci compared to 2 percent from the surf zone, indicating that the marsh was a significant source of enterococci. In the marsh sediments, bacteria were concentrated in the top 1 cm of the sediment cores. In another study (Schiff and Kinney, 2000), sediment samples were collected at 20 shoreline monitoring sites in Mission Bay near San Diego, Calif. Concentrations of fecal coliforms and enterococci in bottom sediments were low during dry weather, increased substantially 1 day after a storm, and were only slightly higher than dry-weather levels 2 weeks after a storm. From these data, the authors concluded that levels of fecal-indicator bacteria in sediments did not appear to represent a long-lasting source of contamination to bay waters.

It is well known and has been repeatedly demonstrated that bacteria survive longer and are present in higher numbers in sediments than in water (Burton and others, 1987; LaLiberte and Grimes, 1982; Matson and others, 1978; Sherer and others, 1992). Bacteria that are attached to sediments may be protected from attack by predators and bactericidal factors such as sunlight. Fecal-indicator bacteria survival has been investigated in stream-bottom sediments, lake-bottom sediments, storm-drain sediments, and ocean-outfall sediments; however, there is a paucity of information on the presence and survival of pathogens or fecal-indicator bacteria in sediments along the shorelines of bathing beaches (swash zone).

Wastewater indicators. The use of wastewater indicators as surrogates for *E. coli* has received considerable attention. These compounds are typically analyzed by gas chromatography/mass spectrometry (GS/MS), and portable GS/MS instruments would facilitate onsite, rapid analysis (Mark Sandstrom, U.S. Geological Survey, oral commun., 1998).

Sterols, such as coprostanol and cholesterol, are present in municipal wastewaters (Nguyen and others, 1995) and have been used to determine the spatial distribution of contamination from sewage sludge (Kelly and Campbell, 1995) and to differentiate human and animal sources of fecal contamination (Leeming and others, 1996). Coprostanol is produced in the digestive tract of humans and other higher animals by the microbiological degradation of cholesterol. Leeming and Nichols (1996) found that water samples from an Australian estuary contaminated by sewage contained between 0.007 and 0.954 mg/L of coprostanol. A statistically significant relation was found between enterococci or fecal coliforms and coprostanol, although several outliers confounded the relation between fecal coliforms and coprostanol. In a study of the Mississippi River (Pereira and others, 1995), investigators found the highest concentrations of coprostanol in sediments near municipal-sewage outfalls. In another study of streams in the Puget Sound Basin, coprostanol or cholesterol were detected at the three sites with the highest fecal-indicator bacteria concentrations among the sites tested (Embrey, 2001).

Caffeine is a potential indicator of domestic wastewater because it is solely of anthropogenic origin (Seiler and others, 1999). Caffeine is ubiquitous in the human diet, is one of the most widely consumed drugs in the world, and is present in raw and treated sewage (Ogunseitan, 1996). Caffeine was shown to be a good tracer of domestic wastewater in a study of the fate of organic contaminants in the Mississippi River (Pereira and others, 1995). In the Mississippi River study, caffeine concentrations ranged from 0.010 to 0.070 µg/L, and elevated concentrations were found downstream from major metropolitan areas. In contrast, in the Puget Sound Basin study, caffeine was detected at sites with and without elevated concentrations of fecal-indicator bacteria (Embrey, 2001).

Other compounds are found in wastewaters. Surfactants are major ingredients of soaps and detergents and may be indicators of domestic wastewater. In the Mississippi River study, anionic detergent concentrations had peaks in the vicinity of major cities and decreased rapidly downstream (Pereira and others, 1995). In the Puget Sound study (Embrey, 2001), polynuclear aromatic hydrocarbons (PAHs) such as naphthalene, fluoranthene, acenaphthalene, and anthracene were found at sites with elevated fecal-indicator bacteria concentrations. Two chemicals known to be used as fumigants and deodorizers in lavatories, 1,3-dichlorobenzene and 1,4-dichlorobenzene, were detected in a sample from an urban stream (Embrey, 2001).

Predictive models. A survey done by the USEPA in 1998 revealed that few local agencies were using predictive models for assessing recreational water quality (U.S. Environmental Protection Agency, 1999b). Of the agencies using models, two approaches to the development of models to predict recreational water quality were used. One approach is a statistics-based model that was used to predict *E. coli* concentrations by means of simple or multiple-linear-regression analysis and environmental and water-quality variables measured onsite as predictors. The second approach was the use of deterministic or mechanics-based models based on complex modeling of the dominant mixing and transport processes (U.S. Environmental Protection Agency, 1999b).

The city of Stamford, Conn., and the state of Delaware are currently using statistics-based models in their beach programs. These models were developed in studies that established site-specific relations between rainfall and concentrations of fecal-indicator bacteria. The city of Stamford developed beach-closure guidelines on the basis of rainfall and enterococci data collected for 8 years at four estuary beaches on Long Island Sound (Joseph E. Kuntz, City of Stamford, Conn., Health Department, written commun., 2001). Investigators found that rainfall greater than 1 in. usually resulted in enterococci concentrations greater than the established standard of 61 colonies per 100 milliliters (col/100 mL), except during two relatively dry summers wherein the threshold was lower. At one of four beaches, however, high enterococci counts were found without any antecedent rainfall; the cause of contamination was found to be the presence of a marina nearby where boat operators were not properly disposing of wastes (Joseph E. Kuntz, City of Stamford, Conn., Health Department, oral commun., 2001). The state of Delaware developed rainfall thresholds to establish beach-closure guidance on the basis of exceedance of the enterococci standard at seven freshwater ponds. Thresholds were based on the linear relations between enterococci concentrations and 12-, 24-, and 48-hour antecedent rainfall amounts (Jack Pingree, State of Delaware, oral commun., 2001). After the first 2 years of threshold development, they have successfully applied pond-specific thresholds for 10 years, having correctly issued advisories based on thresholds 50-85 percent of the time, depending on the year.

In the Cleveland, Ohio, area during summer 1997, water-quality and environmental variables were measured and evaluated for possible inclusion in statistics-based predictive models for three Lake Erie beaches (Francy and Darner, 1998). Turbidity, antecedent rainfall, volumes of wastewater-treatment plant overflows and metered outfalls, a sediment resuspension index, and wave heights were found to be related to *E. coli* concentrations. Wind speed, wind direction, water temperature, and the presence of swimmers were shown to be statistically unrelated to *E. coli* concentrations. A single variable, turbidity, was found to be a poor predictor of *E. coli* and, as a result, investigators used multiple linear regression (MLR) to better explain the variability in *E. coli* concentrations. An MLR model that most reasonably represented the system and accounted for 58 percent of the variability in *E. coli* concentrations was chosen for further testing. The model included weighted categorical rainfall, beach-specific turbidity, and wave height as explanatory variables. Because 90-percent prediction intervals for *E. coli* concentrations were fairly wide, the model did a poor job in providing an accurate numerical estimate of *E. coli*. Instead, the model was used to determine the probability, given a set of input variables, that *E. coli* concentrations will be greater than 235 col/100 mL—the single-sample bathing-water standard in Ohio. Using this method, the model provided an assessment of recreational water quality as well as, and in some cases better than, the current method (that is, use of previous day's *E. coli* concentration).

In a similar study, work is ongoing to develop and test predictive equations for *E. coli* concentrations at two Milwaukee, Wis., beaches based on variables such as rainfall, combined-sewer-overflow volumes, wind vectors, water temperature, and turbidity. The investigators believe that more accurate predictions can be made when real-time measurements of hydrometeorological and water-quality parameters are used in the regression equations (Greg Olyphant, Indiana University, written commun., 2001).

The second approach to modeling recreational water quality is to develop deterministic prediction models. Those models are physics oriented, require accurate meteorological information as initial and boundary conditions, and involve intensive computations. Virginia, New York-New Jersey, Rhode Island, and Washington are using deterministic models in their beach programs (U.S. Environmental Protection Agency, 1999b). A preliminary combined-sewer overflow (CSO) study of fecal-coliform distributions near

Cleveland was done for the Northeast Ohio Regional Sewer District (NEORSD) (Podber and others, 1994) using a deterministic prediction model. The GLFS at OSU has been making real-time predictions of Lake Erie water levels, wave heights, and three-dimensional temperature and current fields based on a three-dimensional coastal ocean circulation model since 1994 (Schwab and Bedford, 1994). Development of a high-resolution three-dimensional model to predict *E. coli* concentrations at one Lake Erie beach also is underway at OSU.

Site descriptions

Six Ohio bathing beaches were selected for this investigation: three Lake Erie urban beaches—Edgewater Park, Villa Angela, and Huntington Reservation (fig. 1), two Lake Erie beaches a in less populated area—Mentor Headlands and Fairport Harbor (fig. 2), and one inland beach—Mosquito Lake State Park (fig. 3).

Edgewater Park (Edgewater), in Cleveland, Ohio, operated by the Ohio Department of Natural Resources, is used extensively during the recreational season. The bathing area is open to the lake, and the breakwall (east of the beach) impedes flowthrough of longshore currents (fig. 4A). The beach was divided into six sampling areas for this investigation based on the locations of markers (lifeguard stations and trees) on the beach. Sources of fecal contamination to the beach are stormwater runoff and combined-sewer overflows (CSOs) from the Edgewater outfall (fig. 4A), the 117th Street outfall, and the Rocky River (fig. 1). Potential sources from the east are the Westerly Wastewater Treatment Plant Combined-Sewer Overflow Treatment Facility (CSOTF) and the Cuyahoga River. The CSOTF discharges primary treated CSO after heavy rains into the inner harbor area east of the Edgewater Park breakwall. Wastewater from the Westerly Wastewater Treatment Plant is an unlikely source of fecal contamination because the effluent is disinfected throughout the recreational season.

Villa Angela, in Cleveland, Ohio, is another popular bathing beach operated by the Ohio Department of Natural Resources. At Villa Angela, four breakwalls were built to stabilize the beach area. The locations of these breakwalls were used to divide the beach into four sampling areas; two sampling areas are in front of breakwalls and two are between breakwalls (fig. 4B). Sources of fecal contamination include the East 156th Street outfall, Ninemile Creek, Dugway Brook, Doan Brook, and the Cuyahoga River to the west and Euclid Creek to the east (fig. 1). All of these outfalls and streams receive inputs from stormwater runoff and CSOs, and some receive inputs from sanitary sewer overflows (SSOs). The Euclid Creek Pump Station pumps sanitary sewage to a higher elevation and then to the wastewater treatment plant. The overflow from the Euclid Creek Pump Station is a minor source of fecal contamination. The Easterly Wastewater Treatment Plant discharges disinfected effluent into Lake Erie and thus is an unlikely source of fecal contamination to Villa Angela.

Huntington Reservation (Huntington) is in a suburb of Cleveland—Bay Village, Ohio—and is operated by the City of Cleveland Metroparks (fig. 1). Water quality in Cahoon and Porter Creeks, directly to the east, is generally good, and these streams are not major sources of fecal contamination to Huntington (Don Killinger, Cuyahoga County Board of Health, oral commun., 2001). The City of Bay Village operates the Rocky River Wastewater Treatment Plant, which discharges treated sewage effluent into Lake Erie to the east of Huntington. Two outfalls discharge storm runoff from the parking lot into Huntington. The sources of fecal contamination to Huntington are largely unidentified. Huntington was divided into four sampling areas for this investigation based on the locations of piers (fig. 4C).

Fairport Harbor and Mentor Headlands are in Lake County to the east of Cleveland in a less-populated area (fig. 2). Fairport Harbor is a small, popular beach operated by Lake Metroparks. The beach area is protected by breakwalls and consists of a shallow swimming area with little wave action. Sources of fecal contamination include septic systems, wastewater-treatment-plant effluent, and stormwater runoff. The Grand River drains into Lake Erie directly west of Fairport Harbor. Mentor Headlands is a 1-mi-long, heavily used beach operated by Ohio Department of Natural Resources. Sources of fecal contamination to Mentor Headlands are similar to those at Fairport Harbor. At Fairport Harbor, one sample was collected from the center of the beach area for routine studies (fig. 2). At Mentor Headlands, two samples were collected for routine studies—one at the east end and one at the west end the beach (fig. 2).

Mosquito Lake State Park (Mosquito Lake), operated by Ohio Department of Natural Resources, is the only inland beach site in this study. The beach is on the southwest corner of an 11-mi-long reservoir from north to south that is 1 mi wide at most points (fig. 3). The dam, operated by the U.S. Army Corps of Engineers, is less than one-half mile southeast of the beach. The beach was divided into three areas for this investigation based on the locations of lifeguard stations (fig. 5). Ancillary sampling sites were established to investigate possible sources of fecal contamination to the lake (fig. 3). Possible sources of fecal contamination include discharges from septic systems from a subdivision north of the beach, runoff from parking lots and wooded areas, birds, and recreational users.

Acknowledgments

Individuals from many agencies helped to ensure the successful completion of this investigation. The authors thank Eva Hatvani, Ildiko Kubiak, and Oljeg Jamnicky, of the Northeast Ohio Regional Sewer District; Jill Lis of the Cuyahoga County Board of Health; Ann McCready-Gliha and Bob Martz of the Cuyahoga County Sanitary Engineers;

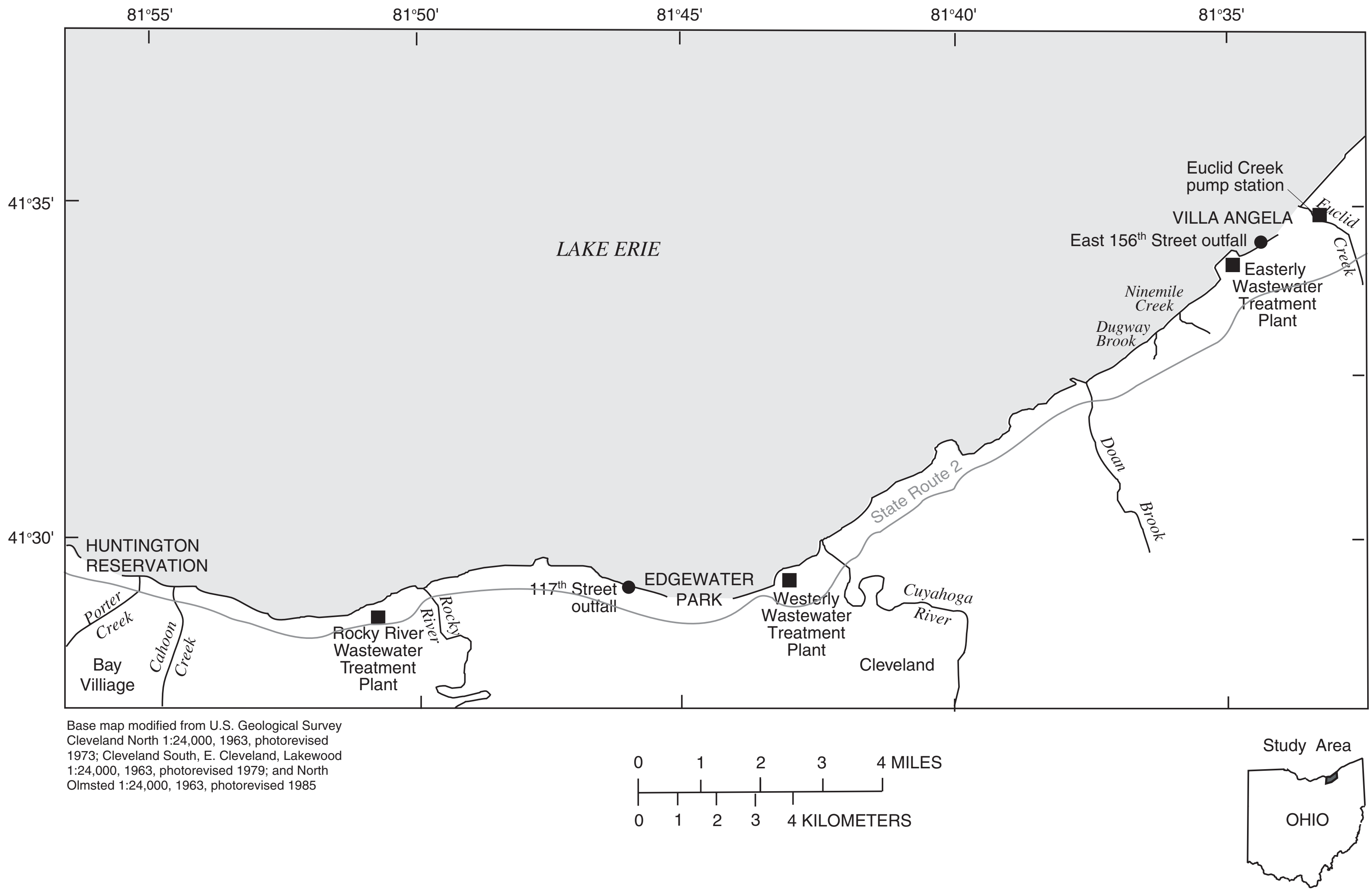


Figure 1. Locations of Lake Erie urban beaches–Huntington Reservation, Edgewater Park, and Villa Angela–in the Cleveland, Ohio, metropolitan area, 2000 and 2001.

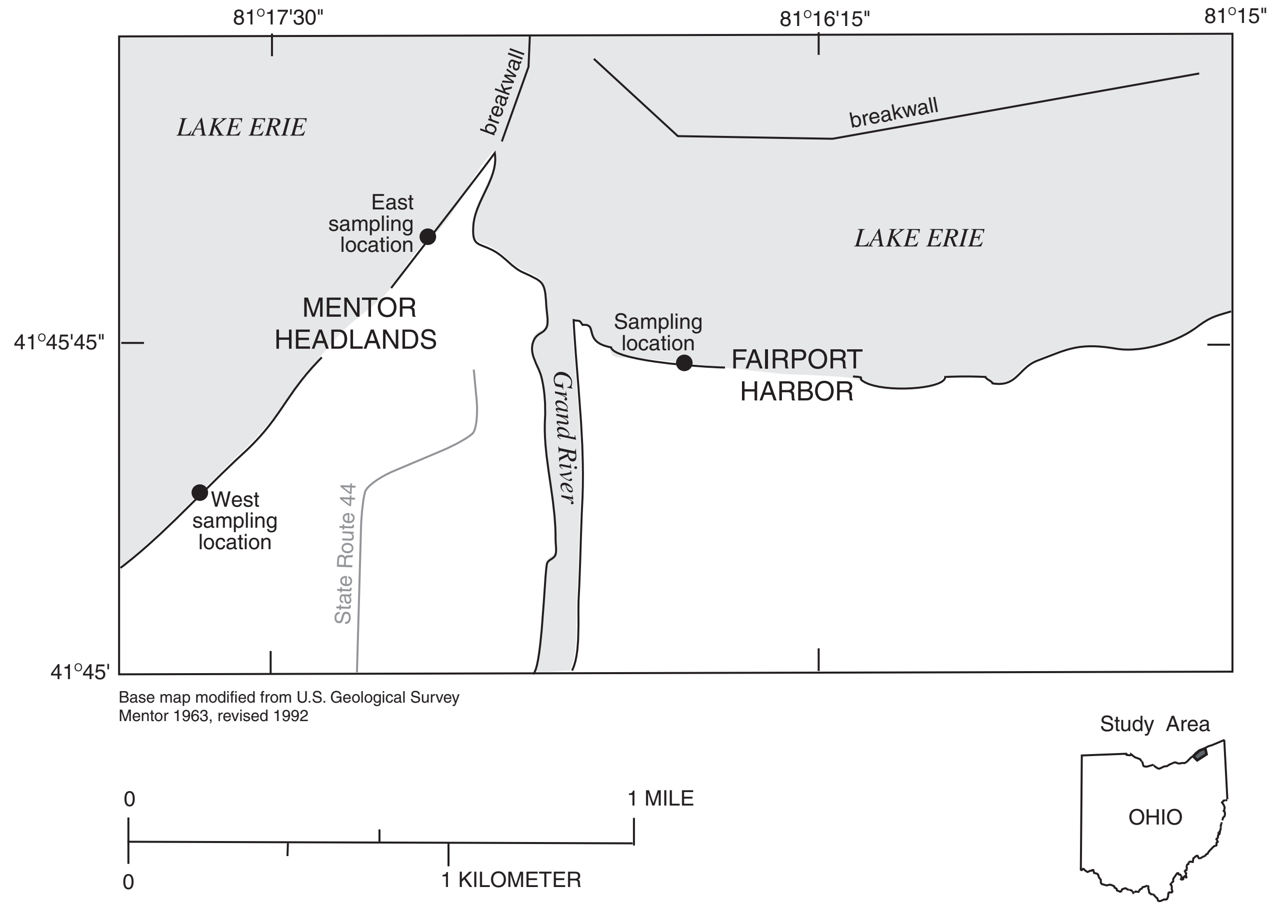


Figure 2. Locations of Lake Erie beaches in a less populated area–Mentor Headlands and Fairport Harbor–in Lake County, Ohio, 2000.

and Dr. Adam Leff and Marc Crissman of Kent State University, Trumbull Campus, for their assistance with project planning and sample collection and analysis. The authors also acknowledge the assistance of others in the planning and implementation of this project— Lester Stumpe, John Graves, and Dan Hudson of the Northeast Ohio Regional Sewer District; Wayne Holmes and Thomas Filbert of the Ohio Department of Natural Resources; Donald Killinger of the Cuyahoga County Health Department; Ted Smith of the Trumbull County Health Department; and Edward Binic of the Lake County Health Department. Thanks are also extended to Paul Anderson and Ted Conlin, at Ohio Environmental Protection Agency, and Robert Powers at Mosquito Lake State Park, for the operation and use of their boats in collection of samples. Special thanks are extended to Armah de la Cruz, at the U.S. Environmental Protection Agency, for analyzing many samples for caffeine, and to Gary Tasker, a retired USGS employee, for providing essential help on the statistical modeling.

Methods of study

Concentrations of *E. coli* were used to assess recreational water quality. This follows the recommendation in the Beach Action Plan, wherein USEPA stated that it intends to promulgate the exclusive use of *E. coli* or enterococci criteria by all states for analysis of freshwater samples by 2003 (U.S. Environmental Protection Agency, 1999a). The USEPA added that current monitoring approaches that use the geometric mean of five samples collected over a 30-day period are outdated and there is interest in developing monitoring requirements that use the results of single samples. For this report, therefore, the single-sample bathing-water standard is used as a point of reference to evaluate recreational water quality. For Ohio, the single-sample bathing water standard of 235 col/100 mL cannot be exceeded in more than 10 percent of samples collected during any 30-day period (Ohio Environmental Protection Agency, 2002).

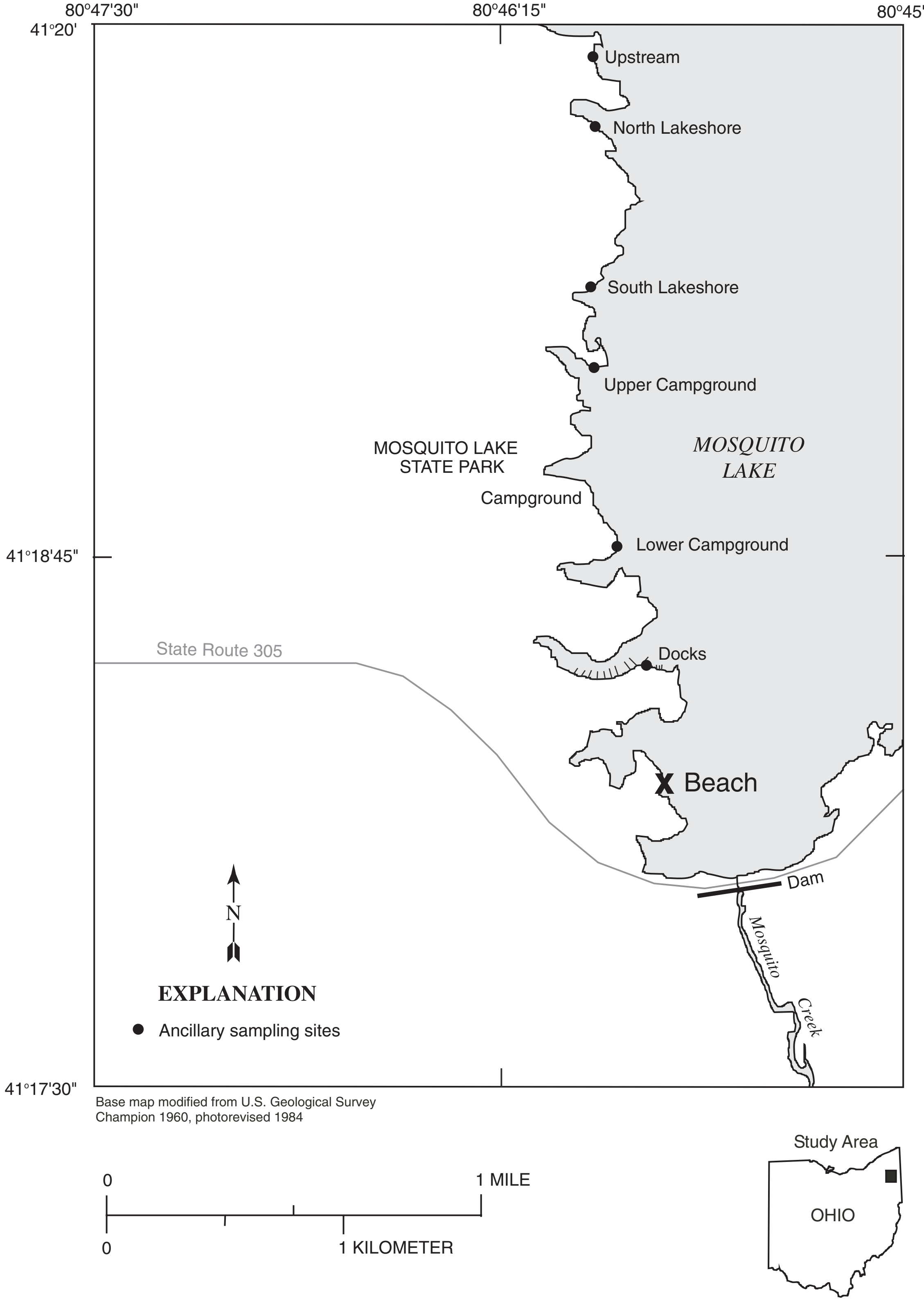


Figure 3. Location of inland beach–Mosquito Lake State Park–Cortland, Ohio, 2000 and 2001.

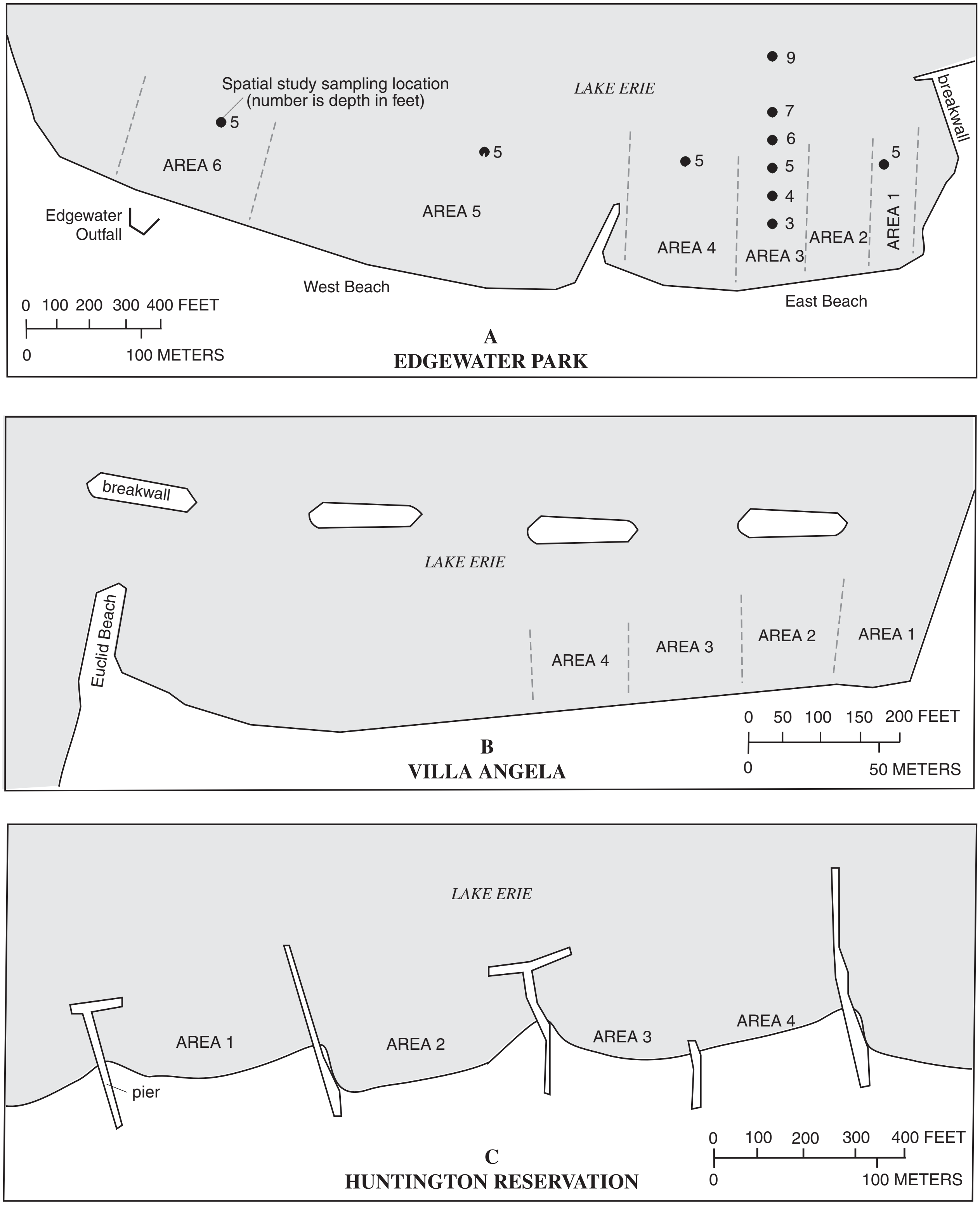


Figure 4. Sampling areas at Lake Erie urban beaches: (A) Edgewater Park, (B) Villa Angela, and (C) Huntington Reservation in the Cleveland, Ohio, metropolitan area, 2000 and 2001.

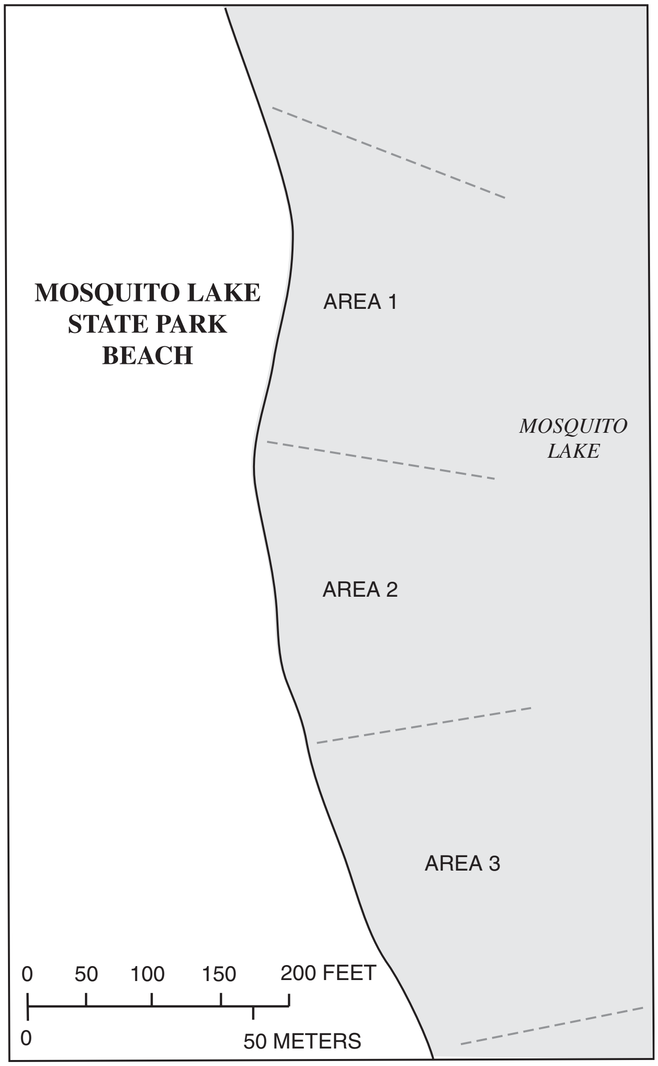


Figure 5. Sampling areas at Mosquito Lake State Park Beach in Cortland, Ohio, 2000 and 2001. (Beach area in relation to rest of lake is shown in fig.3.)

Sampling frequency and locations

Distribution, source, and spatial studies. During distribution, source, and spatial studies, water and sediment samples were collected between 7 and 11 a.m. at Edgewater, Villa Angela, Huntington, and Mosquito Lake by the USGS on selected days in 2000 and at Edgewater and Mosquito Lake in 2001 (table 1). For distribution studies, samples were collected from two or three areas at each beach (figs. 4 and 5). From each area, samples were collected from within the bathing area, outside the bathing area (about 100 ft offshore from buoys that demarcate the boundary of the bathing area), and (or) near the swash zone. For source studies, samples were collected at six ancillary sites at Mosquito Lake and at the beach (fig. 3). Samples were collected from ancillary sites at three locations: shoreline samples were collected less than 50 ft from the shore, nearshore samples were collected approximately 200–400 ft from the shore, and offshore samples were collected approximately one-fourth mile offshore in open water. For spatial studies during 2000, sampling was done from a boat at two randomly selected areas at Edgewater outside the bathing area. Water depths were from 7 to 10 ft. During 2001 spatial studies, samples were collected at Edgewater at 5-ft water depths from areas 1, 4, 5, and 6 and at six depths, ranging from 3 to 9 ft, from area 3 (fig. 4A).

Routine studies. During routine studies, water samples were collected 4 or 5 days a week by local agencies at all six beach study sites throughout summer 2000 and at four beaches (excluding Fairport Harbor and Mentor Headlands) during 2001. Sampling frequencies, times, and areas sampled were different among the six beach study sites (table 2). At Huntington and Mentor Headlands during 2000, two samples were collected daily because of the extended length of these beaches. During 2001, it was decided to collect two samples daily at all four beaches to reduce sampling and analytical variability. An average of the two values was used for data analysis.

Sampling methods

Field personnel used established sampling methods, described below, to collect water and (or) sediment samples for *E. coli* during source and routine studies and during most distribution and spatial studies. Special sampling methods were employed to collect some samples during distribution and spatial studies. For chemical constituents and suspended-sediment analyses, separate sample bottles and different sampling techniques were required, as described below.

Established sampling methods for all studies. Lake-water samples were collected by means of a grab-sampling technique that minimized contamination of sterile sampling containers (Myers and Sylvester, 1997). After wading or swimming to the area and depth designated, a sterile polypropylene bottle was opened about 18 in. below the water surface and filled. For spatial studies during 2001, at some locations and depths, an additional bottle was opened at about 18 in. above the lake bottom and filled.

Lake-bottom sediments were collected into autoclaved wide-mouth 250-mL polypropylene jars. A diver swam out (or was transported by boat) to the designated sampling point, secured the lid on the sampling jar, opened the lid upon reaching the lake bottom, and scooped the bottom sediments to obtain a sample. The diver closed the lid of the jar before surfacing to minimize contamination by the overlying water. Because of spatial heterogeneity of bacteria concentrations in sediment, three sediment jars were collected from each sampling point and composited before analysis.

Table 1. Dates of sampling and locations of water and sediment samples collected at Ohio beaches for distribution, source, and spatial studies, 2000 and 2001

[Samples from within the bathing area were collected on all dates.]

		Type of study			Samples collected				
Beach	**Date(s)**	**Distribution**	**Source**	**Spatial**	**Outside bathing**	**Near swash zone**	**Shoreline**	**Nearshore**	**Offshore**
Edgewater	June 19 & 21, 2000	x				x			
(Lake Erie)	June 20 & 22, 2000	x		x	x				
	July 26, 2000	x				x			
	August 7, 2000	x				x			
	August 8, 2000	x							
	August 17, 2000	x				x			
	August 28, 2000	x							
	August 29, 2000	x		x	x				
	February 27, 2001	x				x			
	June 27 - 28, 2001	x		x	x	x			
	August 4 - 6, 2001	x		x	x	x			
Huntington	June 19 & 21, 2000	x				x			
(Lake Erie)	June 20 & 22, 2000	x			x				
	July 26, 2000	x				x			
	August 7, 2000	x				x			
	August 8, 2000	x							
	August 17, 2000	x				x			
	August 28, 2000	x							
Mosquito Lake	June 5 & 7, 2000	x	x			x	x		
	June 6 & 8, 2000	x	x		x				x
	July 31 & August 2, 2000	x	x			x		x	
	August 1 and 3, 2000	x	x		x				x
	August 30, 2000	x	x			x	x	x	
	February 27, 2001	x				x			
	June 12 - 13, 2001	x				x			
	August 28, 2001	x			x			x	x
Villa Angela	June 19 & 21, 2000	x				x			
(Lake Erie)	June 20 & 22, 2000	x			x				
	July 26, 2000	x				x			
	August 7, 2000	x				x			
	August 8, 2000	x							
	August 17, 2000	x				x			
	August 28, 2000	x							
	August 29, 2000	x			x				

Table 2. Beach study sites and sampling information for routine studies at Ohio beaches, May through August 2000 and 2001

Beach	Year and area(s) sampled	Sampling frequency / time	Collecting agency / analyzing agency[a]
Edgewater	2000 - Area 3 2001 - Areas 2 and 3	Monday to Friday, 8 - 9 a.m.	Northeast Ohio Regional Sewer District
Villa Angela	2000 - Areas 3 2001 - Areas 2 and 3	Monday to Friday, 7 - 8 a.m.	Northeast Ohio Regional Sewer District
Huntington	2000 - Areas 1 and 2 2001 - Areas 1 and 2	Monday to Thursday, 7 - 9 a.m.	Cuyahoga County Board of Health/ Cuyahoga County Sanitary Engineers
Fairport Harbor	2000 - Central	Monday and Wednesday, 8 - 10 a.m. Tuesday and Thursday, 9 a.m. - 1 p.m.	Lake County General Health District Ohio Department of Health/ Cuyahoga County Sanitary Engineers
Mentor Headlands	2000 - West and East	Monday and Wednesday, 8 - 11 a.m. Tuesday and Thursday, 11 a.m. - 12 p.m.	Lake County General Health District Ohio Department of Health/ Cuyahoga County Sanitary Engineers
Mosquito Lake	2000 - Area 2 2001 - Areas 2 and 3	Monday to Thursday, 8 - 9 a.m.	U.S. Geological Survey contractor

[a] Collecting and analyzing agency are the same unless indicated otherwise.

Special sampling methods for samples from near the swash zone. Interstitial-water and sediment samples from near the swash zone were collected and analyzed for *E. coli* during distribution studies. Sample sites were determined by first locating the landward edge of the swash zone, which is the zone alternatively covered and exposed by waves. Sample locations were chosen 3 and 6 ft inland from the outer edge of the swash zone. Three subsample points were then marked in a row, each 1.5 ft apart, parallel to the shore at both the 3 and 6 ft inland locations. The depth of dry sand was measured and removed with a shovel from each subsample point. Using a sterile post-hole digger for the subsample holes, a field technician made a 6-in.-diameter hole in the moist sand. The post-hole digger was sterilized by applying household bleach for 2 minutes and rinsing it with sterile sodium thiosulfate and sterile deionized water. The hole was dug deep enough to allow free interstitial water to accumulate. Once the interstitial water entered the hole, a sterilized well casing was inserted to prevent the hole from collapsing. During initial studies, ethanol was used to sterilize the well casings; however, ethanol disinfection was found to be inadequate, so bleach was used in later studies. In order to protect *E. coli* from the effects of ultraviolet radiation that they are not normally exposed to, a tarp was placed over the holes as they filled with interstitial water. Once stabilized, the interstitial water was collected with a sterile 25-mL pipet (fig. 6). Approximately 100 mL of interstitial water was collected from each of the three subsample holes and composited into a sterile 1-L polypropylene bottle. After collecting water, the casings were removed, and the distances from the top of moist sand to the interstitial water line (depth to the water table) were measured with a tape measure; an average depth of three holes was calculated. Subsurface sediment was removed from each subsample hole with a sterile spatula and was placed in a 125-mL sterile plastic jar. The sediments collected from the three holes were composited before analysis.

Special sampling methods for discrete vertical sampling. For spatial studies during 2000, discrete vertical water samples for *E. coli* were collected outside the bathing area where water depths ranged from 7 to 9 ft. Samples were collected from a boat by means of a subsurface grab sampler (Wheaton Science Products, Millville, N.J.); the pole of the sampler was marked to the nearest 0.5 ft. The approximate depth of the water at the sampling point was measured with a sonar depth finder. Taking into account depth, the sampler operator identified three evenly spaced vertical locations for sampling, starting 1 ft from the bottom and ending at least 2 ft below the surface. The operator screwed a sterile bottle onto the end of the sampler, lowered the sampler through the water column to the designated depth, and pulled a rope through a pulley to expose a sampling port. After the bottle was almost full, the operator closed the sampling port using a spring-operated cap and then retrieved, capped, and placed the sample on ice. The sampling port was sterilized and cleaned between samples with dilute bleach, sodium thiosulfate, and sterile deionized water.

Sampling methods for chemical constituents and suspended sediment. For some studies, samples were collected in separate bottles for analyses of chemical constituents and suspended sediment. For distribution and source studies during 2000 and for routine studies during 2001, water samples were collected and analyzed for wastewater indicators and (or) caffeine concentrations. These samples were collected in the same manner that the general samples were collected. The 1-L glass amber bottles used were previously baked at 450°C to remove all organic contaminants. During distribution studies, interstitial-water samples for total nitrogen concentrations were removed by use of 25 mL pipets, and approximately 30 mL of water from each hole was composited into 125-mL brown polypropylene bottles. After collection, the total nitrogen samples were preserved with 1 mL of 4.5 *N* sulfuric acid. For spatial studies, two glass 500-mL bottles were filled with water to determine suspended-sediment concentrations. For spatial studies during 2000, the discrete vertical sampling technique was used to fill the bottles. For spatial studies during 2001, the grab sampling technique was used.

Analysis of water and sediment samples

All water and sediment samples for *E. coli* were placed on ice and transported to the laboratory. In the laboratory, samples were analyzed for *E. coli* and processed for other constituents within 6 hours for water and 24 hours for sediment samples. Samples collected for wastewater indicators, caffeine, and total nitrogen and suspended-sediment concentrations were sent to the appropriate analyzing laboratory within 3 days of sampling.

Figure 6. Sampling of interstitial water from subsample holes, 3 feet and 6 feet inland from the swash zone.

Distribution, source, and spatial studies. For distribution, source, and spatial studies, USGS personnel analyzed the water samples at a local laboratory (the Cuyahoga County Sanitary Engineers Laboratory in Valley View, Ohio, or Kent State University, Trumbull Campus, in Warren, Ohio). Sediment samples were sent for analysis to the USGS Ohio District Microbiology Laboratory in Columbus, Ohio, by overnight mail or courier. The constituents included and frequencies of analysis were different among the types of studies; standard methods were followed for most tests (table 3).

All water and sediment samples were analyzed for concentrations of *E. coli* by use of the mTEC membrane-filtration method (U.S. Environmental Protection Agency, 1985). Approximately 10 percent of sediment samples collected during source studies were also analyzed by use of the most-probable-number Colilert Quantitray method (Idexx Laboratories, Westbrook, Maine). This was done because of sediment interferences in the growth of bacteria on some agar plates; however, poor recoveries of *E. coli* on Colilert precluded use of Colilert as a substitute for membrane filtration for sediment samples. For sediment samples, additional sample-processing steps, developed during an earlier study (Francy and Darner, 1998), were required before plating. Briefly, 50 g of sediment was aseptically removed from each of three replicate sample jars and composited into a sterile 1-L jar. Twenty grams of the mixed sediment was then placed into a bottle containing 200 mL of saline buffer; a second aliquot of mixed sediment was removed to determine percent dry weight. The analyst placed the bottle on a shaker for 45 minutes, removed the bottle, allowed suspended materials to settle for 30 seconds, and decanted the liquid phase for plating. Calculations were made as described in Francy and Darner (1998) to convert colony counts to colonies per gram of dry weight sediment (col/g_{DW}).

After processing water and sediment samples for *E. coli*, some additional analyses were done. Turbidity was determined in all water samples (Hach Company, 1989). Sediment samples were randomly selected for particle-size distributions and total organic carbon concentrations. For these analyses, after *E. coli* samples were processed, the remaining sediment in the three jars was composited. For particle-size analysis, the composited sediment was dried for several days at room temperature and processed through a series of sieves to weigh the percentage of sediment finer than 63, 250, and 1,000 µm (Guy, 1969). For total organic carbon analysis, the composited sediment was processed through a 2-mm sieve, and the < 2-mm fraction was analyzed at the USGS National Water Quality Laboratory (NWQL) in Lakewood, Colo.

Lake-water samples were collected for wastewater indicators and sent to two laboratories. The USGS NWQL analyzed samples for a wastewater indicator method that was under development at the time of the study (Michael Schroeder, U.S. Geological Survey, written commun., 2000). Water samples were extracted with methylene chloride by means of continuous liquid-liquid extraction under acidic and then basic conditions. The extracts were concentrated and analyzed by selected ion monitor gas chromatography/mass spectrometry (SIM GC/MS) for the determination of 45 compounds. The compounds included, but were not limited to, plasticizers, fumigants, detergents, polyaromatic hydrocarbons, an analgesic, fecal indicators, and a stimulant (caffeine), many of which are common in the environment. The USEPA in Cincinnati, Ohio, analyzed samples for caffeine using a method of solid-phase extraction and high-performance liquid chromatography with fluoresence detection (Piocos and de la Cruz, 2000). During 2000, sample volumes of 100 mL were subjected to the preconcentration and extraction step; however, this proved to be too small a volume to detect caffeine. During 2001, the analytical volume was increased to 200 mL to improve the sensitivity of the analysis (Armah de la Cruz, U.S. Environmental Protection Agency, oral commun., 2001).

Other water-quality properties and constituents were measured. Field measurements of specific conductance, pH, temperature, and (or) concentrations of dissolved oxygen in lake and interstitial waters were measured according to standard USGS methods (Wilde and Radtke, 1998). For interstitial waters, measurements of specific conductance and temperature were made in each hole; measurements from the three holes were averaged. Interstitial waters were sent to the USGS NWQL for total nitrogen analysis, and lake-water samples were sent to Heidelberg Water Quality Laboratory, Tiffin, Ohio, for determination of suspended-sediment concentrations.

Routine studies. The methods of analyses during routine studies, when applicable, were the same methods as those described above and used for distribution, source, and spatial studies (table 3). During routine studies, field personnel measured water temperature and specific conductance onsite; all water samples were analyzed for turbidity and concentrations of *E. coli* in local laboratories. (See table 2 for collecting and analyzing agencies.)

Collection and compilation of ancillary information

Ancillary environmental data were collected by field crews or compiled from a variety of sources. These data were used to develop predictive models. At all beaches, personnel from the USGS or cooperating agencies estimated the number of birds and wave heights at the time of sample collection. Data from four USGS-operated gaging stations were used to estimate streamflow—Cuyahoga River at Independence (04020800) and Rocky River near Berea (04201500) for three Lake Erie urban beaches, Grand River near Painesville (04212100) for two Lake Erie beaches in a less populated area, and Eagle Creek at Phalanx Station (03093000) for Mosquito Lake (Shindel and others, 2001 and 2002). Wind speed and direction were measured at a weather station at the Youngstown Regional Airport (National Oceanic and Atmospheric Administration, 2001), 6 mi. southeast of Mosquito Lake. Data on the average elevation of the lake pool and the average flow through the dam outlet at Mosquito Lake were obtained from the U.S. Army Corps of Engineers (George Kusko, U.S. Army Corps of Engineers, written commun., 2000 and 2001). The ultraviolet index (UV) is a measure of the intensity of UV radiation. Daily UV index data were measured at noon during 2000 (Craig Long, National Oceanic and Atmospheric Administration, written commun., 2000). Data for 2001 were an average of two daily values (Accuweather, 2001); one measurement obtained between 11 a.m. to 1 p.m. and the second from 4 to 6 p.m. Calculated wave heights and current directions were obtained from the GLFS at OSU for three Lake Erie locations—the Rocky River (used for Huntington), Edgewater Park (used for Edgewater), and the Cuyahoga River (used for Villa Angela) (Yifei Philip Chu, The Ohio State University, written commun., 2000 and 2001). Personnel at the GLFS used a three-dimensional hydrodynamic prediction model based on surface meteorological observations and forecasts from numerical weather prediction models to forecast currents and waves (Schwab and Bedford, 1994).

Daily rainfall amounts were compiled from several agencies. The Northeast Ohio Regional Sewer District (NEORSD) operates a network of rain gages in the Cleveland, Ohio, area. Data from two of these gages (John Graves, Northeast Ohio Regional Sewer District, written commun., 1998) were used to develop models in 1997 for Edgewater and Villa Angela during an earlier study (Francy and Darner, 1998) and are used in this report to test the 1997 model. These NEORSD data were not used in new model development because the data were not readily available and often required interpretation and qualification. Instead, rain data used in new model development were obtained from a gage at Hopkins International Airport (National Oceanic and Atmospheric Administration, 2001) for Edgewater, Villa Angela, and Huntington; a gage at the Painesville Municipal Water Plant for Mentor Headlands and Fairport Harbor (Ed Binic, Lake County Health Department, written commun., 2000); and a gage at the Youngstown Airport (National Oceanic and Atmospheric Administration, 2001) for Mosquito Lake. These gages represent rainfall amounts in nearby watersheds. To obtain rainfall amounts for a specific beach area, the USGS installed a rain gage on the roof of a structure at Huntington Beach during 2000.

Table 3. Constituents and properties determined on water and sediment samples collected at Ohio beaches during distribution, source, spatial, and routine studies, 2000 and 2001

[mL, milliliters; g_{DW}, gram per dry weight sediment; MPN, most-probable number; NTU, nephelometric turbidity unit; g/kg, grams per kilogram; µS/cm, microsiemens per centimeter; µg/L, micrograms per liter; mg/ L, milligrams per liter; USEPA, U.S. Environmental Protection Agency]

Constituent or property	Type of Study: Distribution	Source	Spatial	Routine	Frequency or number of samples	Method (reference)	Detection limit
Escherichia coli in water	x	x	x	x	Every sample	mTEC method (USEPA, 1985)	1 colony / 100mL
Escherichia coli in sediment	x	x	x		Every sample	mTEC method (modified from USEPA, 1985)	1 colony / g_{DW}
Escherichia coli in water				x	Every sample[a]	Colilert (Idexx Laboratories, Westbrook, Maine)	1 MPN/100 mL
Escherichia coli in sediment		x			10 percent of samples	Colilert (Idexx Laboratories, Westbrook, Maine)	1 MPN/100 mL
Percent dry weight of sediment	x	x	x		Every sample	(American Society of Agronomy, 1982, p. 790-791	Not applicable
Turbidity of water	x		x	x	Every sample	Nephelometric method (Hach Company, 1989)	0.01 NTU
Particle size of sediment	x x	x			70 samples, 2000 41 samples, 2001	(Guy, 1969, p. 47-51)	Not applicable
Total organic carbon of sediment	x				6 subsurface samples from near the swash zone, 2000	USGS O-5101-83 (Wershaw and others, 1987; Brandt and others, 1990)	0.1 g/kg
Wastewater indicators in water	x	x		x	13 samples, 2000	(LeRoy Schroeder, U.S. Geological Survey, 2000,written commun.)	0.03 - 6.4 µg/L
Caffeine in water	x	x			55 samples, 2000	(Piocos and de la Cruz, 2000)	0.008 µg/L
				x	84 samples, 2001		
Temperature of water	x			x	Every sample	Thermistor thermometer (Wilde and Radtke, 1998, chapter 6.1)	Not applicable
Specific conductance of water	 x	 x		x x	Every sample, 2000[b] Every sample, 2001[c] Every sample	Conductivity sensor (Wilde and Radtke, 1998, chapter 6.3)	1 µS/cm

Table 3. Constituents and properties determined on water and sediment samples collected at Ohio beaches during distribution, source, spatial, and routine studies, 2000 and 2001 —Continued

[mL, milliliters; g_{DW}, gram per dry weight sediment; MPN, most-probable number; NTU, nephelometric turbidity unit; g/kg, grams per kilogram; µS/cm, microsiemens per centimeter; µg/L, micrograms per liter; mg/ L, milligrams per liter; USEPA, U.S. Environmental Protection Agency]

Constituent or property	Type of Study: Distribution	Source	Spatial	Routine	Frequency or number of samples	Method (reference)	Detection limit
Dissolved oxygen of water			x		Every sample	Amperometric method (Wilde and Radtke, 1998, chapter 6.2)	1 mg/L
pH of water			x		Every sample	Hydrogen ion electrode (Wilde and Radtke, 1998, chapter 6.4)	Not applicable
Total nitrogen in water	x				12 subsurface samples from near the swash zone, 2001	(Patton and Truitt, 2000)	0.08 mg/L
Suspended sediment in water			x		Every sample	(Guy, 1969, p. 11-13)	1.0 mg/L

[a] Fairport Harbor and Mentor Headlands only.
[b] Mosquito Lake only.
[c] Edgewater Park and Villa Angela only.

Statistical methods

Correlation analysis and x/y scatterplots were used as exploratory tools to examine the relations between *E. coli* concentrations and environmental or water-quality variables. This is the first step in identifying explanatory variables that are related to *E. coli* and could be used in regression models. Pearson's *r* is a correlation coefficient that measures the linear association between two variables. Spearman's *rho*, another correlation coefficient, measures the monotonic relation (nonlinear or linear) between two variables. Because the Pearson's *r* values were similar to the Spearman's *rho* values for most relations in this study, only the Pearson's *r* values were used to identify variables related to *E. coli* concentrations. For Pearson's *r*, if the data lie exactly along a straight line with positive slope, then the correlation coefficient is equal to 1 (Helsel and Hirsch, 1992, p. 209-218). The more the correlation coefficient deviates from 1 or –1 and approaches zero, the weaker the relation. The level of significance was set at α=0.05, unless specified otherwise.

The rank transform test, a nonparametric analysis of variance (ANOVA), was used to compare more than two groups of data. In the rank transform test, all data are combined and ranked from lowest to highest value, and an ANOVA is computed on the ranks. This test determines whether the median differs between any of the groups. If the rank transform test showed differences among groups, the Tukey-Kramer multiple comparison test was used to determine which groups differed from each other (Helsel and Hirsch, 1992, p. 198-200).

Multiple linear regression was used to develop models to predict *E. coli* concentrations from explanatory variables. Models were chosen among all possible variable combinations to maximize the coefficient of determination (R^2) and minimize the Mallows' *Cp* (Mallows, 1973). The R^2 of the model is the fraction of the variation in the *E. coli* concentration that can be explained by a combination of explanatory variables. The adjusted R^2 was used and reported for all statistical analyses; the adjusted R^2 differs from the R^2 in that the former has been adjusted for degrees of freedom. The *Cp* statistic is a measure of the error variance and the bias introduced by not including important variables in a model. When several models had nearly equal R^2 and *Cp* values, a set of models was chosen to reduce multi-collinearity (where at least one explanatory variable is related to one or more other explanatory variables). Further evaluation of these models was based on their having significant parameter estimates and acceptable partial residual plots. In a partial residual plot, log *E. coli* is regressed against all explanatory variables except for one, and the residuals are plotted against the omitted explanatory variable. These plots provide information on how much influence the omitted variable has on the regression by eliminating the effects from other variables; they are also useful in evaluating whether the relation between *E. coli* and explanatory variables is specified correctly. After the partial plots were examined, a beach model (or more than one model) was selected on the basis of having a set of explanatory variables that seemed reasonable and having data that were easy to collect or compile.

Even with the best models, prediction intervals were shown to be too wide to offer an accurate prediction of *E. coli* concentrations (Francy and Darner, 1998). Consequently, the probability of exceeding the single-sample bathing-water standard of 235 col/100 mL was used as the model output variable. This variable was calculated as the probability that Student's t with n-p degrees of freedom is greater than or equal to X, where $X = (\log(235) - y)/s$ and y is the regression estimate of the log of *E. coli*, s is the standard error of prediction of y, n is the number of observations used in the regression, and p is the number of regression coefficients estimated in the regression equation. For example, a set of explanatory variables may result in an output value of 45 percent—this means that, given this set of explanatory variables, there is a 45 percent chance that the *E. coli* concentration will exceed 235 col/100 mL. Threshold probabilities were set by taking the data set used to develop the model and finding the probability that provided the highest number of correct responses and lowest number of false negative responses. Output values below the threshold probability would result in the assumption by the beach manager that water quality is acceptable; output values above the threshold probability would result in the posting of the beach with a water-quality advisory.

Quality-assurance and quality-control practices

Quality-assurance and quality-control (QA/QC) practices are considered an integral part of all data-collection activities. Field and laboratory protocols for all studies were written and distributed to ensure that procedures were followed correctly and consistently by USGS and cooperative agencies. Detailed QA/QC practices for collection of water-quality data at the USGS, Ohio District, are described in Francy and others (1998). For the USGS Ohio District Microbiology Laboratory, procedures for laboratory operation and equipment maintenance are described in Francy and others (2001).

Quality-control samples were collected to measure sampling and analytical variability or contamination and to ensure that data satisfied project objectives. To ensure that membrane-filtration equipment was clean and sterile and that reagents were uncontaminated, a filter blank—a 50-mL aliquot of sterile buffered water plated before the sample—was included with each new bottle of buffer. All filter blanks were negative. To measure sampling and analytical variability, duplicate samples were collected and analyzed for several constituents. For turbidity, duplicate aliquots were measured from the same bottle, and measurements that did not agree within 10 percent were repeated; an average turbidity of two aliquots that agreed within 10 percent was reported for each sample. For percent dry weight, duplicate jars were analyzed for approximately 10 percent of the samples; for some samples, split samples were also analyzed (Appendix A1). Splits are subsamples taken from the same jar or bottle. For *E. coli* analysis, approximately 4 percent of the samples from routine studies (Appendix A2), as well as distribution, source, and spatial studies (Appendix A3), were collected in duplicate and (or) analyzed as split samples. For caffeine, approximately 10 percent of the samples collected were quality-control samples—replicates, spikes, and blanks (Appendix A4). Results of quality-control samples for percent dry weights, *E. coli*, and caffeine were examined qualitatively to ensure that procedures were correctly followed. For suspended-sediment analysis, two duplicate bottles from each location were collected. Results from both bottles were reported because a comparison of suspended-sediment concentrations between duplicate bottles found large differences in some cases (Appendix B).

Distribution, sources, and wastewater indicators for *Escherichia coli* at bathing beaches

Data collected during distribution, source, and spatial studies at three Lake Erie beaches and one inland lake for water, sediment, and concentrations of wastewater indicators are listed in Appendixes B, C**,** and D, respectively. Data collected during routine studies for concentrations of caffeine in water samples are listed in Appendix E. Samples were collected during extended dry periods and after rainfalls ranging from trace amounts to 1.5 in.

To characterize the size class of sediment at each beach and aid in data interpretation, particle-size analyses were done on a subset of lake-bottom sediments and subsurface sediments from near the swash zone (Appendix C). More detailed summary statistical information is shown (table 4) for the percent finer than 250 µm, because these values showed the greatest variability among sites and sediment types. For sediment samples collected within the bathing area, the greatest percentage of sediments coarser than 1,000 µm and finer than 63 µm were found at Mosquito Lake. Among Lake Erie beaches, the coarsest sediments were found at Villa Angela in all three types of sediment. At Edgewater and Huntington, the subsurface sediments from near the swash zone varied considerably in

Table 4. Summary of particle-size analysis of lake-bottom and subsurface sediments collected at Ohio beaches, 2000 and 2001

[Bathing, samples collected within the bathing area; Outside, samples collected outside the bathing area; Swash zone, subsurface sediments collected from near the swash zone]

Beach	Type of sediment	Number of samples	Percent finer than (micrometers)				
			1000,	250			63,
			average	Average	Minimum	Maximum	average
Edgewater	Bathing	19	99.3	94.0	78.9	98.0	5.4
	Outside	8	97.5	95.3	89.2	98.6	2.0
	Swash zone	20	96.1	75.2	24.9	93.9	2.9
Huntington	Bathing	2	95.8	59.2	48.1	70.2	0.7
	Outside	1	99.9[a]	98.6[a]			1.1[a]
	Swash zone	6	92.7	45.7	13.3	71.7	1.3
Villa Angela	Bathing	2	64.6	3.2	0.2	6.1	0.1
	Outside	4	94.9	59.4	33.6	81.9	1.2
	Swash zone	6	67.6	7.6	0.1	24.4	0.2
Mosquito	Bathing	12	50.9	28.7	11.1	83.8	11.3
	Outside	1	43.6[a]	23.1[a]			3.4[a]
	Swash zone	14	52.9	19.1	13.2	32.1	1.6

[a] Not an average value because only one sample was collected.

particle size and were generally classified as fine to coarse sands (63 to 1,000 µm). At Villa Angela and Mosquito Lake, most subsurface sediments from near the swash zone were classified as medium sand to gravel (250 to >1,000 µm).

Distribution studies

Distribution of *Escherichia coli* within and outside the bathing area. The distributions of *E. coli* concentrations in water (fig. 7) and sediments (fig. 8) within the bathing area were compared to concentrations outside the bathing area to determine whether sediments outside the bathing area were a significant deposition area for *E. coli*. Results from samples collected from within the bathing area are shown in boxplots, and results from samples collected outside the bathing area are superimposed on the boxplots. This provides a comparison between the two groups of data based on the 5th to 95th percentiles. For Lake Erie beaches (Edgewater, Villa Angela, and Huntington), samples were collected within the bathing areas at water depths of 3–5 ft and outside the bathing areas at water depths of 8–11 ft. At Mosquito Lake, samples within the bathing area were collected at 3- to 6-ft water depths, and samples outside the bathing area were collected at depths ranging from 3 to 11 ft.

For Lake Erie water samples (fig. 7A), concentrations of *E. coli* outside the bathing area followed a similar distribution or were less than those within the bathing area. This same pattern was seen in water samples collected at Mosquito Lake except on June 2000 (fig. 7B). For Lake Erie and Mosquito Lake sediment samples (fig. 8), concentrations of *E. coli* outside the bathing area were generally in the same range or were less than those within the bathing area except in June 2000. The two highest sediment *E. coli* concentrations (820 and 3,000 col/g_{DW}) were found in June 2000 at Edgewater within the bathing area of area 1; this area is the eastern section of the beach closest to the break wall (fig. 4A). For most time periods, therefore, concentrations of *E. coli* in water and sediments outside the bathing area were in the same range or less than concentrations within the bathing area.

Concentrations of *E. coli* in water were compared to their concentrations in associated lake-bottom sediments by means of correlation analysis. A combined data set from within and outside the bathing area was used for this analysis (fig. 9). Statistically significant correlations ($\alpha = 0.05$) were found between water *E. coli* concentrations and sediment *E. coli* concentrations at Edgewater during 2000 but not at Edgewater during 2001 (fig. 9A). The relation between concentrations of *E. coli* in water and sediment was significant at Huntington for 2000. This may be because *E. coli* concentrations at Huntington were above 50 col/100 mL in the water samples collected (fig. 9B). This was not the case at the other beaches and may be why, for example, the correlation was poor at Edgewater in 2001. At Villa Angela, the coarseness of the sediments from the bathing

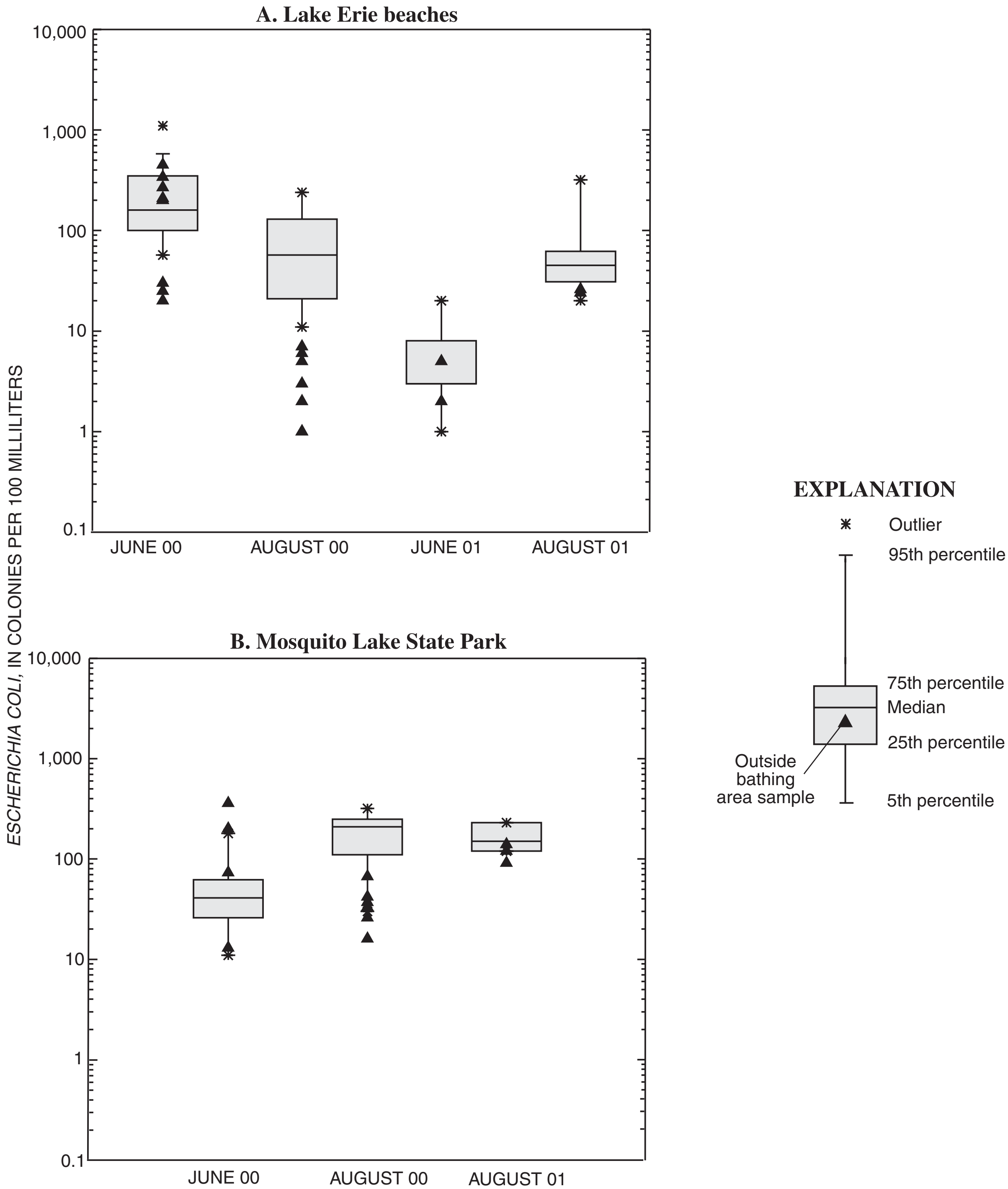


Figure 7. Distribution of *Escherichia coli* concentrations in water samples collected within the bathing area and comparison to samples collected outside the bathing area for (A) Lake Erie beaches and (B) Mosquito Lake State Park.

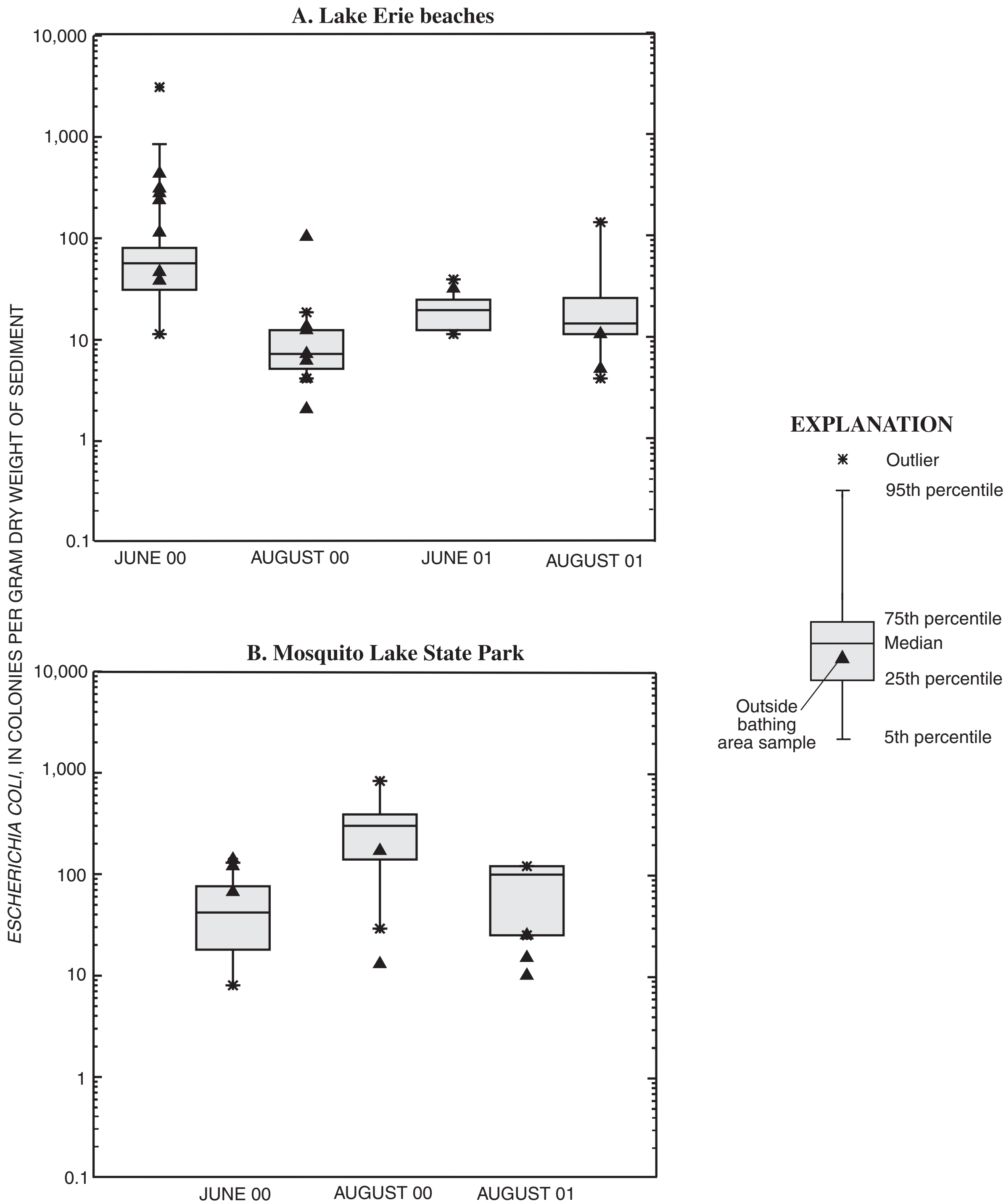


Figure 8. Distribution of *Escherichia coli* concentrations in sediment samples collected within the bathing area and comparison to samples collected outside the bathing area for (A) Lake Erie beaches and (B) Mosquito Lake State Park.

area may be why the relation was weak (fig. 9C and table 4); coarse sediments have less surface area for bacterial attachment than fine sediments do. At Mosquito Lake, a statistically significant relation was found between water and sediment *E. coli* concentrations when 2000 and 2001 data were combined (r=0.346, p=0.036), but not when each year was examined separately (fig. 9D). Overall, the data indicate that water and sediment *E. coli* concentrations were related at Huntington to a greater extent than the other beaches.

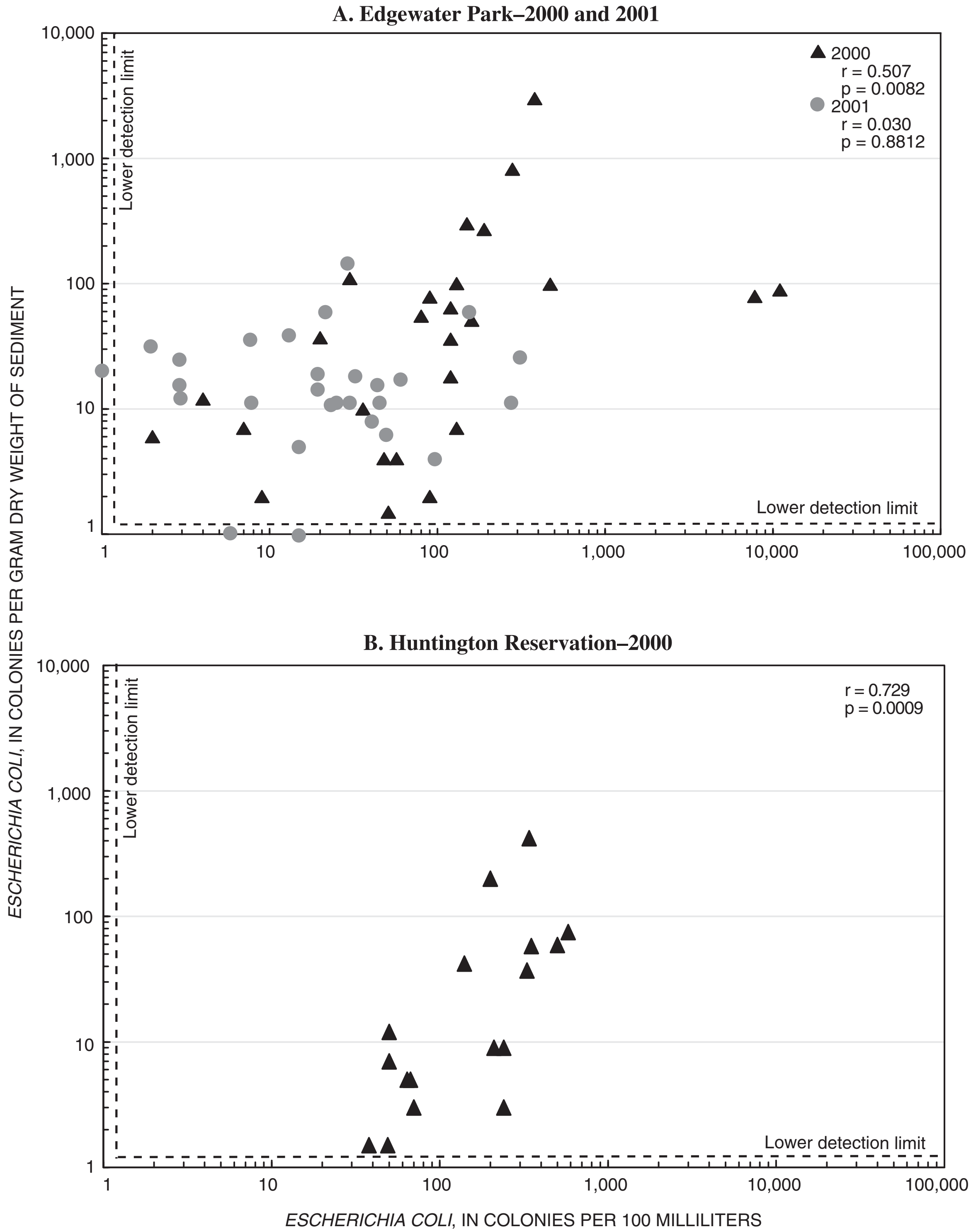


Figure 9. Comparison of *Escherichia coli* concentrations in bathing water to *Escherichia coli* concentrations in lake-bottom sediments at (A) Edgewater Park in 2000 and 2001, (B) Huntington Reservation in 2000, (C) Villa Angela in 2000, and (D) Mosquito Lake State Park in 2000 and 2001. (*r* is the Pearson's correlation coefficient, and *p* is the significance of the correlation; samples were collected within and outside the bathing area.)

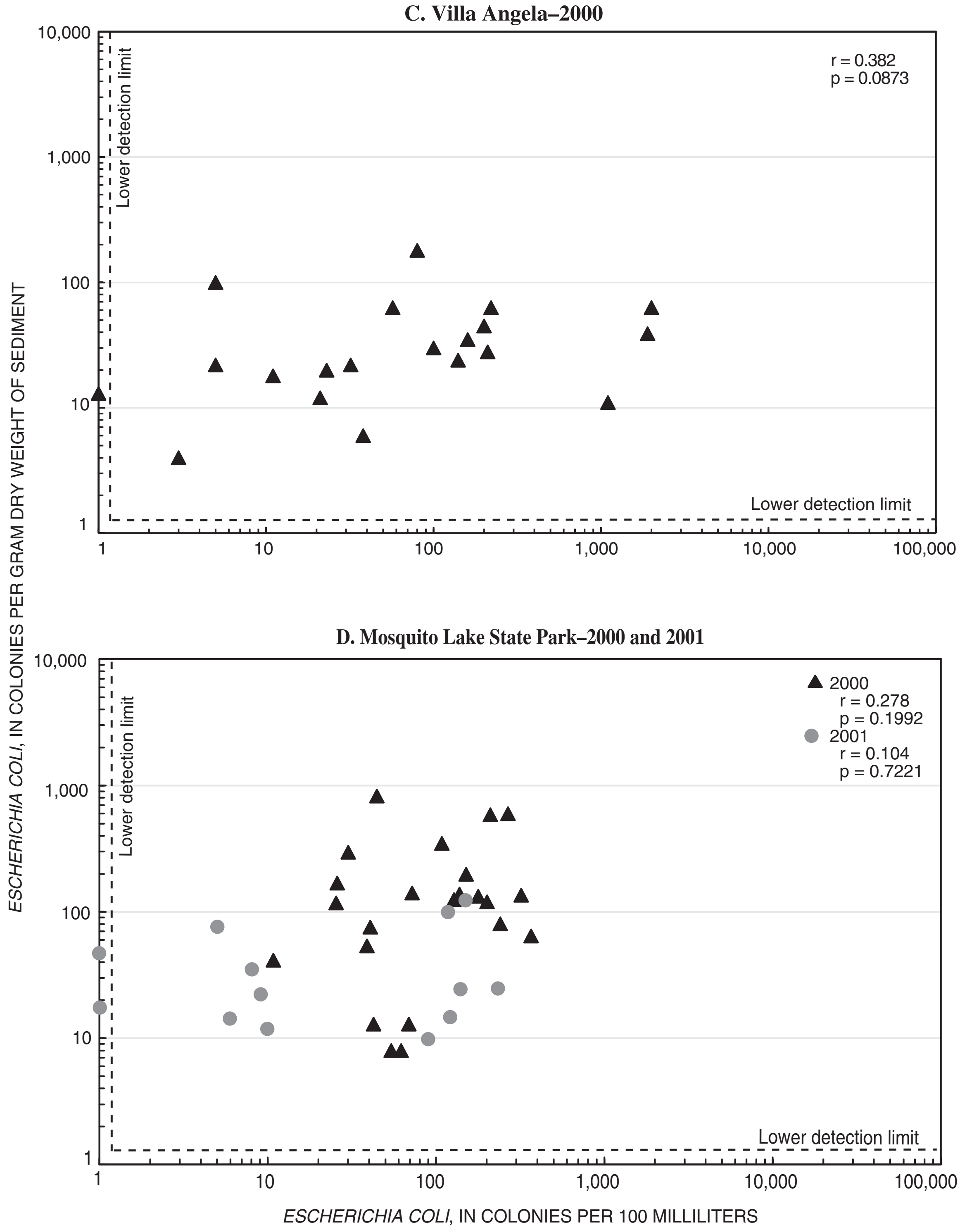


Figure 9. Comparison of *Escherichia coli* concentrations in bathing water to *Escherichia coli* concentrations in lake-bottom sediments at (A) Edgewater Park in 2000 and 2001, (B) Huntington Reservation in 2000, (C) Villa Angela in 2000, and (D) Mosquito Lake State Park in 2000 and 2001. (*r* is the Pearson's correlation coefficient, and *p* is the significance of the correlation; samples were collected within and outside the bathing area.)

Concentrations of *Escherichia coli* near the swash zone. Concentrations of *E. coli* in interstitial water and subsurface sediment samples collected from near the swash zone for each beach by year are summarized in table 5. The number of water-sample results is less than the number of sediment-sample results because results for water samples collected from June 19 through August 16, 2000, were discarded because of contamination from casings. Interstitial water comes in contact with the casings, whereas subsurface sediments do not. In interstitial water samples, the highest concentrations of *E. coli* were found at Mosquito Lake during 2000 and at Edgewater during 2001 (table 5). The high concentrations of *E. coli* in interstitial water samples from Edgewater during 2001 were not reflected in sediment samples; sediment *E. coli* concentrations at Edgewater were in the same range during 2000 and 2001. During 2000 when three Lake Erie beaches were studied, median concentrations of *E. coli* in interstitial water and subsurface sediments were lowest at Huntington and highest at Villa Angela. Although there are no standards for *E. coli* in interstitial waters, the high concentrations found in sediment and water samples alike at some locations warrant some concern for public health.

At Edgewater and Mosquito Lake, swash-zone samples were collected during 2000 and 2001 to facilitate a more in-depth spatial and temporal analysis. Concentrations of *E. coli* in interstitial water and subsurface sediments for individual dates are shown in figures 10 and 11, respectively. At Edgewater, concentrations of *E. coli* in interstitial waters (fig. 10A) collected on the same date differed among areas by about 1 to more than 2 orders of magnitude. Interstitial-water samples collected 6 ft inland showed a spatial pattern on August 4, 5, and 6 at Edgewater; concentrations were highest in area 3 and lower in areas 1 and 4 on all three dates. This pattern was not seen in the samples collected 3 ft inland on these dates. At Mosquito Lake, concentrations of *E. coli* in interstitial waters (fig. 10B) were very high during 2000, remained somewhat elevated in February 2001, and dropped off significantly in June 2001. Concentrations of *E. coli* in interstitial waters at Mosquito Lake collected on the same date differed among areas by only 1 to 1.5 orders of magnitude or less. No spatial patterns of *E. coli* concentrations were observed in the subsurface sediments at Edgewater (fig. 11A). Concentrations of *E. coli* in

Table 5. Concentrations of *Escherichia coli* in interstitial water and subsurface sediment samples collected from near the swash zone, 2000 and 2001

Beach	Year	Number of samples	Minimum	Maximum	Geometric mean
Interstitial water [a]					
Edgewater	2000	4	<33	4,100	580
	2001	34	<3	110,000	1,800
Villa Angela	2000	4	290	3,500	1,500
Huntington	2000	4	67	670	270
Mosquito Lake	2000	7	4,800	400,000	38,000
	2001	15	<3	19,000	33
Subsurface sediment [b]					
Edgewater	2000	22	<1	220	19
	2001	34	3	300	30
Villa Angela	2000	12	6	2,100	26
Huntington	2000	14	<1	90	5
Mosquito Lake	2000	23	22	30,000	1,600
	2001	15	2	9,200	23

[a] Colonies per 100 milliliters.
[b] Colonies per gram dry weight of sediment.

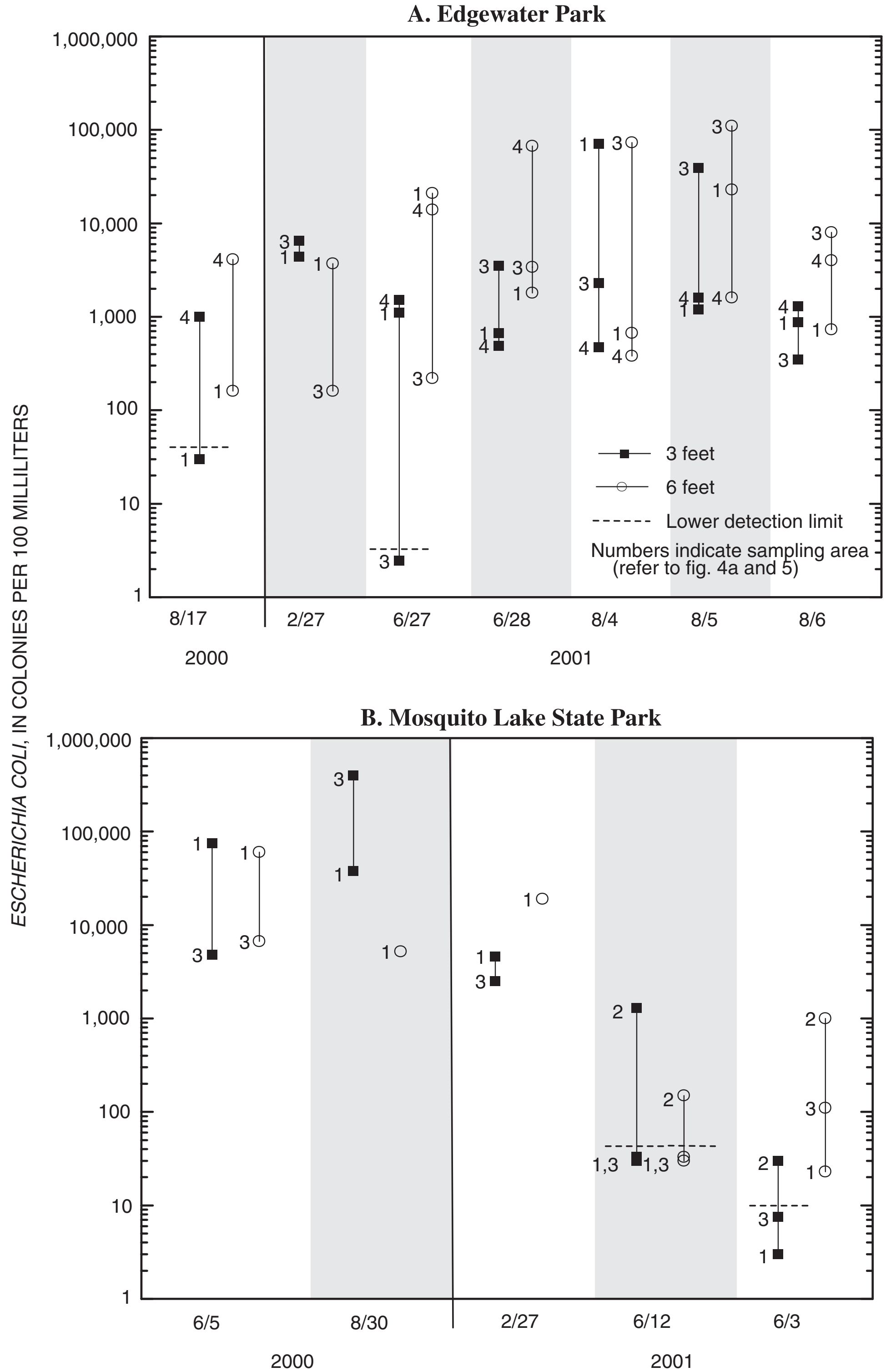


Figure 10. Concentrations of *Escherichia coli* in interstitial waters collected 3 feet and 6 feet inland from near the swach zone during 2000 and 2001 at (A) Edgewater Park and (B) Mosquito Lake State Park.

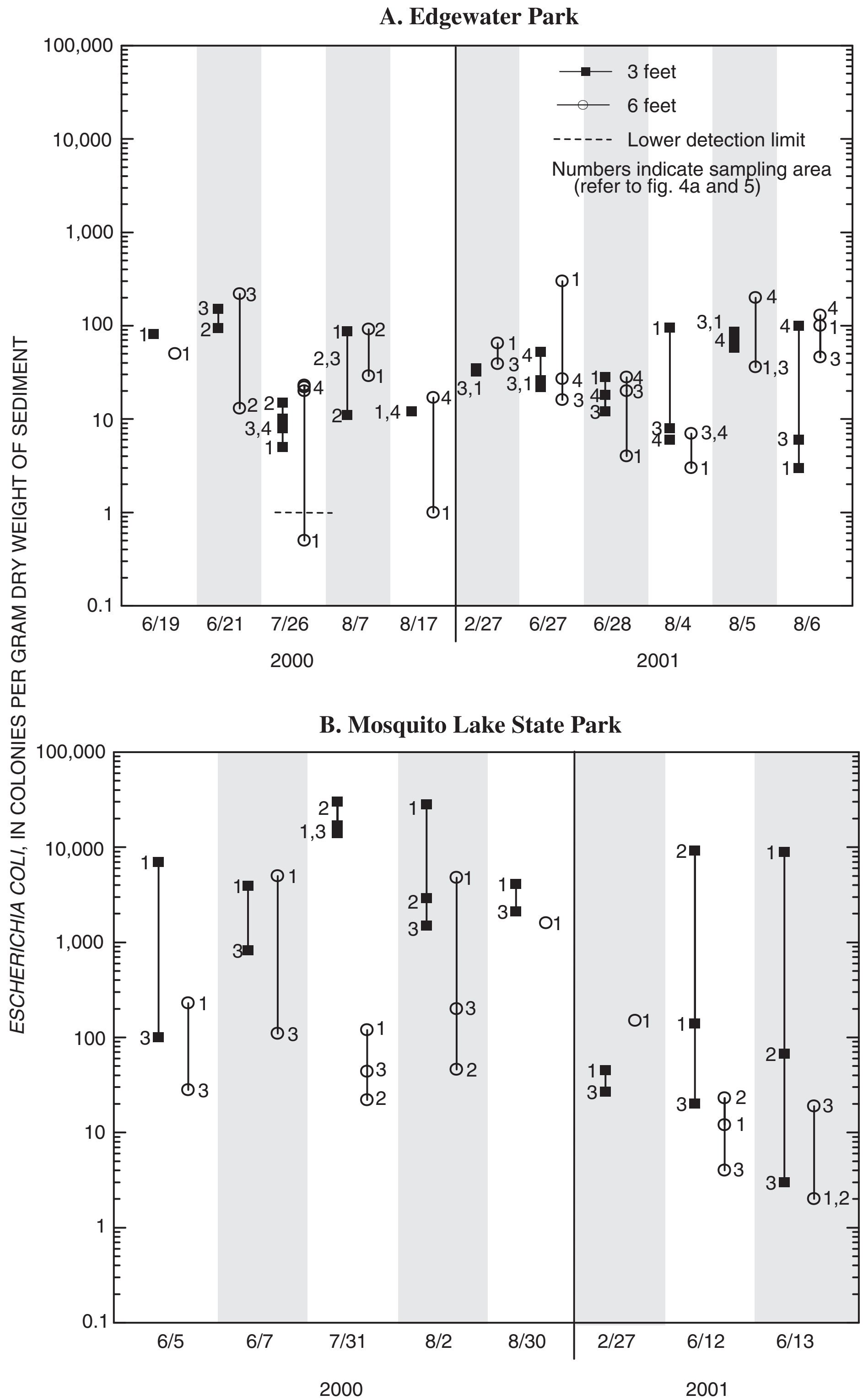


Figure 11. Concentrations of *Escherichia coli* in subsurface sediments collected 3 feet and 6 feet inland from near the swash zone during 2000 and 2001 at (A) Edgewater Park and (B) Mosquito Lake State Park.

subsurface sediments at Mosquito Lake (fig. 11B) were more variable between areas on the same date than those found at Edgewater. At Mosquito Lake, subsurface sediment samples collected 3 ft inland were generally higher than those collected concurrently at 6 ft inland; the highest concentrations were usually found in samples from area 1.

For Edgewater and Mosquito Lake during 2000 and 2001, the relations between *E. coli* concentrations found in bathing waters to those found in interstitial waters or subsurface sediments collected from near the swash zone were examined (table 6). At Edgewater, no significant correlations were found. At Mosquito Lake, in contrast, significant correlations were found between bathing waters and interstitial waters collected 3 ft inland; the relation between bathing waters and interstitial waters collected 6 ft inland was not significant, owing to the influence of an outlier sample collected during February 2001.

Source studies at Mosquito Lake

Source studies were done at Mosquito Lake to investigate possible sources of fecal contamination to the lake and the bathing beach. Water and sediment samples were collected at six sites north of the bathing beach (fig. 3), referred to as "ancillary sites," and at the three areas at the bathing beach (fig. 5). Samples collected outside the bathing area at areas 1, 2, and 3, were considered offshore samples for this data analysis.

Because shoreline, nearshore, and offshore samples were not all sampled on one date (table 1), data analysis is complicated; however, by combining the data one can gain insight into the patterns of *E. coli* concentrations in samples collected from the southwest part of Mosquito Lake. Concentrations of *E. coli* were elevated in shoreline water samples collected from North Lakeshore and South Lakeshore but were considerably lower in the nearshore and offshore water samples from these sites (fig. 12). In fact, *E. coli* concentrations showed a decreasing pattern as one moved further away from the North Lakeshore shoreline until reaching the bathing areas. Concentrations of *E. coli* in water in shoreline, nearshore, and offshore bathing area sites were higher than those found at the docks, lower campground, and upper campground sites.

Lake-bottom sediment samples were not collected from shoreline ancillary locations. The predominance of near-detection and below-detection values found in samples from nearshore and offshore ancillary locations suggests that *E. coli* is not accumulating in the bottom sediments at these sites (fig. 13). In contrast, at the bathing beach, *E. coli* was detected in sediments collected from the shoreline and nearshore areas, indicating some storage of *E. coli* in lake-bottom sediments. Concentrations of *E. coli* in the offshore bathing-area sites were highly variable.

Spatial studies at Edgewater

Spatial studies were done during 2000 and 2001 to aid the GLFS in development of deterministic models to predict *E. coli* concentrations. Data on the spatial and temporal variability of *E. coli* concentrations and water-quality measurements were made at Edgewater. Comparisons among suspended-sediment concentrations could not be made because sample variability (as measured by duplicate samples) was larger than the vertical and spatial differences that were being measured (Appendix B).

Table 6. Correlations between $\log_{10}$ *Escherichia coli* concentrations in bathing waters and those in interstitial water or subsurface-sediment samples collected from near the swash zone at Ohio beaches, 2000 and 2001

[Pearson's *r* correlations significant at a=0.05 are indicated in bold]

Bathing water samples	Samples collected from near the swash zone			
	Interstitial water		Subsurface sediment	
	3 feet inland	6 feet inland	3 feet inland	6 feet inland
Edgewater	0.184	-0.255	0.243	0.179
Mosquito Lake	**0.632**	0.433	-0.303	-0.162

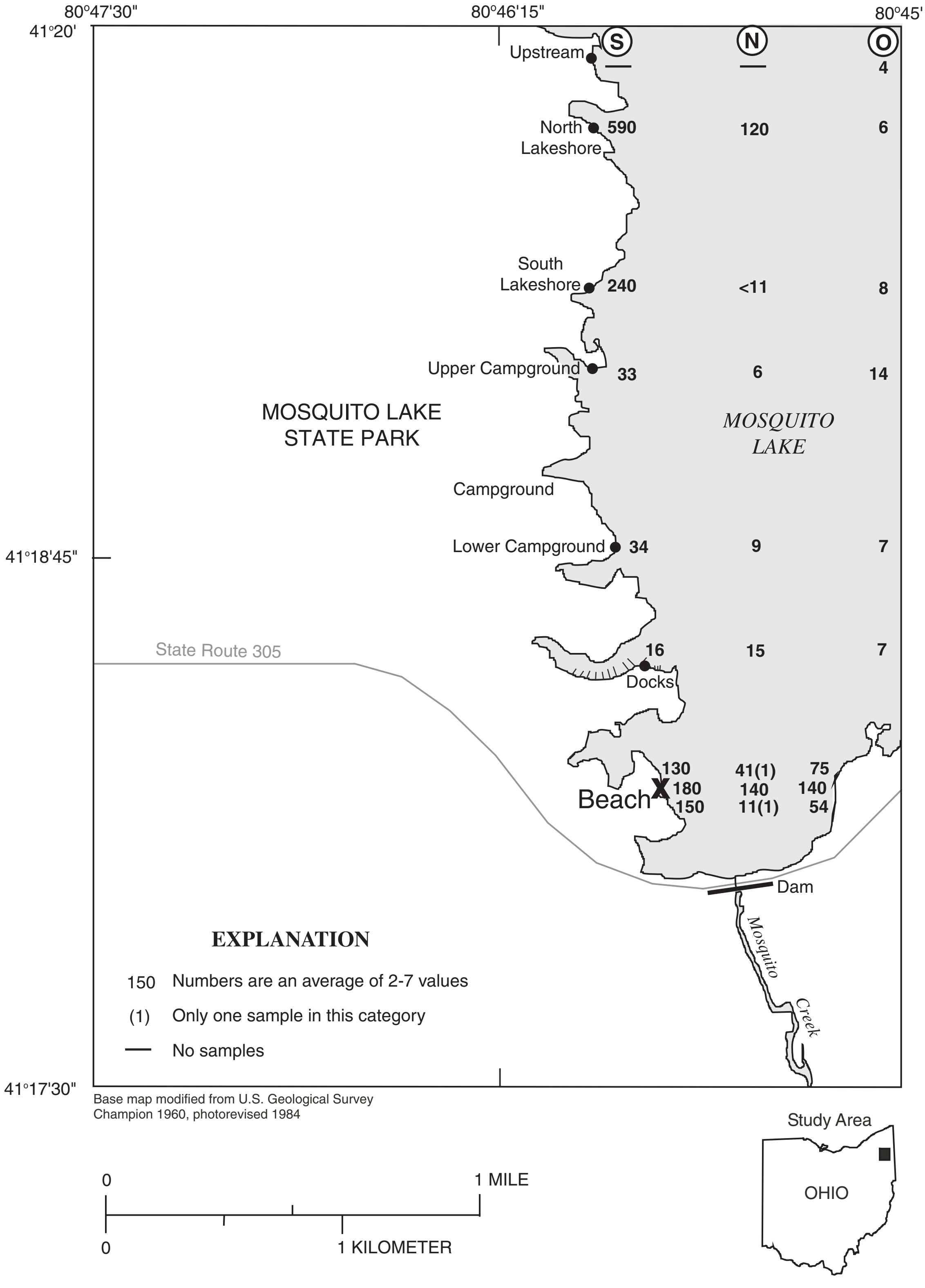


Figure 12. Average concentrations of *Escherichia coli* in (S) shoreline, (N) nearshore, and (O) offshore water samples collected at Mosquito Lake State Park, June-August 2000 and August 2001. (Concentrations are in colonies per 100 milliliters.)

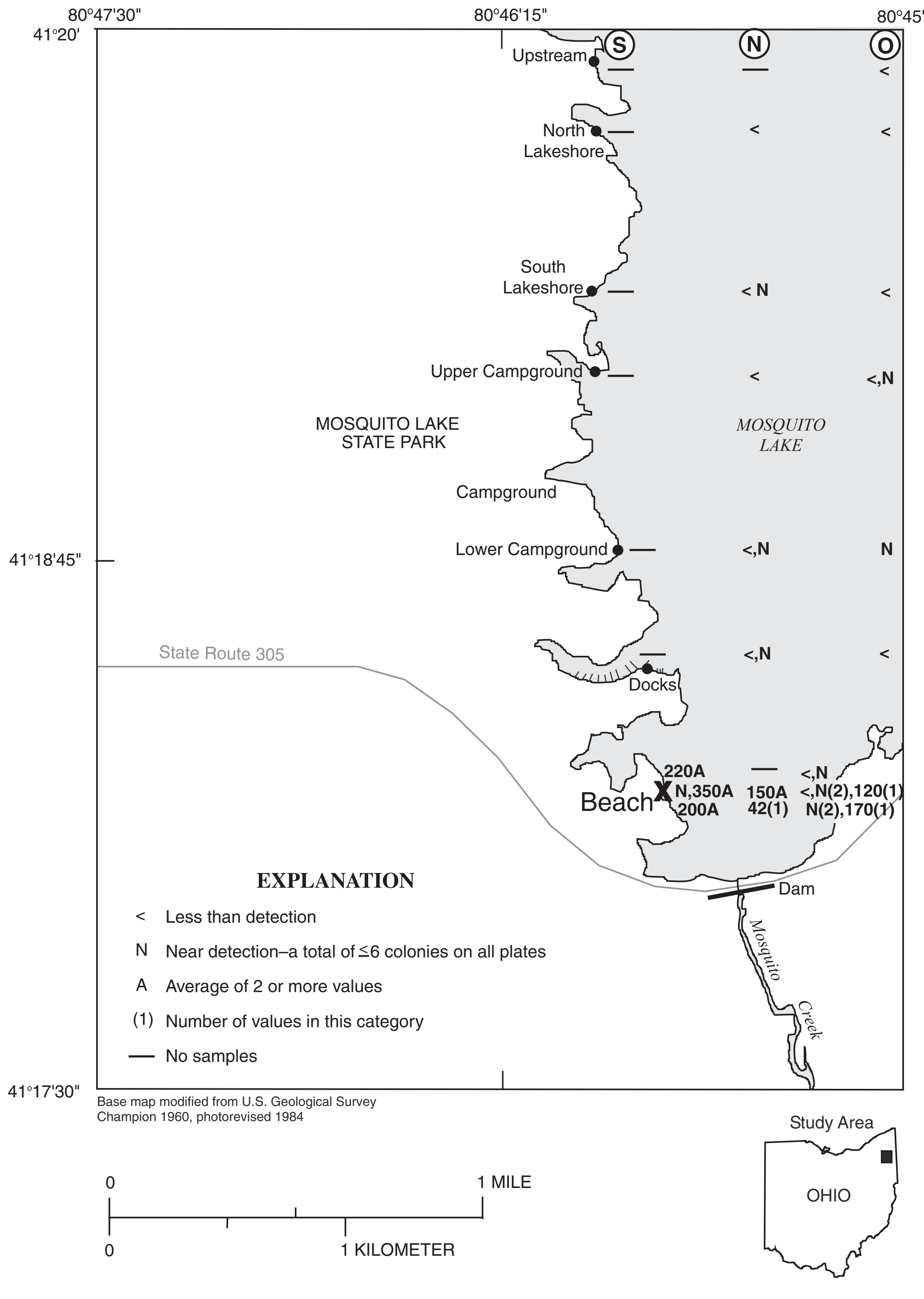


Figure 13. Concentrations and detections of *Escherichia coli* in (S) shoreline, (N) nearshore, and (O) offshore sediment samples collected at Mosquito Lake State Park, June-August 2000 and August 2001. (Concentrations are in colonies per gram dry weight of sediment.)

During 2000, spatial studies consisted of vertical profiles of measurements at locations outside the bathing area (table 7). Measurements were made and samples were collected at 3-, 6-, and 9-ft points when water depths were 10 to11 ft in June and at 2-, 4-, and 7-ft points when water depths were 8 to 9 ft in August. The range of values for *E. coli* concentrations and the maximum differences among values in each profile for temperature, specific conductance, turbidity, and dissolved oxygen were fairly small and precluded any comparisons among vertical points (table 7).

During 2001, spatial studies were done within the bathing area during dry weather and after a significant rainfall. Water and lake-bottom sediment samples were collected on June 27 and 28 after 5 days of dry weather and on August 4 through 6 after 1.4 in. of rainfall fell on August 3. Sampling locations were at 5-ft water depths from areas 1, 4, 5, and 6 and at six different water depths, ranging from 3 to 9 ft, from area 3 (fig. 4A). To gain information on vertical distributions, some of the deeper sampling locations included two sampling points—one approximately 18 in.

Table 7. Summary of the range of *Escherichia coli* concentrations and the maximum differences for physical water-quality properties in spatial studies at Edgewater Park, 2000 and 2001

[ft, feet; col/100 mL, colonies per 100 millimeters; °C, degrees Celsius; µS/cm, microsiemens per centimeter at 25°C; NTU, nephelometric turbidity units; mg/L, milligrams per liter; nd, not determined; --, not applicable]

Date	Area	Sampling/ measurement depths (ft)	*Escherichia coli* concentrations, range (col/100 mL)	Maximum difference				
				Water temperature (°C)	Specific conductance (µS/cm)	pH (standard units)	Turbidity (NTU)	Oxygen, dissolved (mg/L)
Vertical profiles - water samples, 2000								
20-Jun-00	1	3, 6, 9	150-240	0.3	4	nd	4	nd
20-Jun-00	4	3, 6, 9	180-220	0.2	7	nd	5	nd
22-Jun-00	1	3, 6, 9	26-38	0	1	nd	1	nd
22-Jun-00	3	3, 6, 9	20-30	0	0	nd	0	nd
29-Aug-00	1	2, 4, 7	3-9	0.9	7	0.1	0.7	0.7
29-Aug-00	4	2, 4, 7	4-7	0.5	0	0.1	0.8	0.6
Spatial and temporal distributions - water samples, 2001								
27-Jun-01	1,3,4,5,6	3-9	3-21	0.8	27	0.2	2.4	1.8
28-Jun-01	1,3,4,5,6	3-9	1-13	1.5	5	0.1	0.4	1.7
4-Aug-01	1,3,4,5,6	3-7	30-320	0*	50*	nd	9.1	nd
5-Aug-01	1,3,4,5,6	3-9	15-62	1.4	10	0.6	3.2	1.0
6-Aug-01	1,3,4,5,6	3-9	20-56	0.4	6	0.1	0.4	1.0
Spatial and temporal distributions - sediment samples, 2001								
27-Jun-01	1, 3, 4, 6	3-7	11-40	--	--	--	--	--
28-Jun-01	1, 3, 4, 6	3-9	3-75	--	--	--	--	--
4-Aug-01	1, 3, 4, 5, 6	3-7	3-59	--	--	--	--	--
5-Aug-01	1, 3, 4, 5, 6	3-9	2-57	--	--	--	--	--
6-Aug-01	1, 3, 4, 5, 6	3-9	2-140	--	--	--	--	--

* Difference of only two samples.

above the lake bottom (AB) and one 18 in. below the water surface (BW).

The range of *E. coli* concentrations in water found on any given sampling day were fairly narrow except on August 4, 2001, after significant rainfall (table 7). On August 4, the highest *E. coli* concentrations were found in water samples collected from areas 5 and 6 (320 and 280 col/100 mL, respectively), followed by the sample collected from area 1 (160 col/100 mL) (Appendix B). On August 4, 5, and 6, 2001, the highest *E. coli* concentrations in sediment were found in area 1 (Appendix C). As for vertical distributions, *E. coli* concentrations were fairly similar in those water samples collected BW and AB at each location (Appendix B). Values for water temperature, specific conductance, pH, and dissolved oxygen showed little spatial variation on the days where data were available. Because of meter problems, the spatial variation of water-quality measurements could not be assessed on August 4, although for the two specific-conductance measurements made on that date, the difference was the largest found in this study (50 μS/cm). The highest turbidity value (12 NTU) was found at area 1 on August 4.

Wastewater indicators as surrogates for *Escherichia coli*

During 2000, the USGS collected 13 water samples for a suite of wastewater indicators from a variety of locations and beaches (Appendix D). To serve as positive controls, two lake-water samples were spiked with primary-treated wastewater collected from the Westerly Wastewater Treatment Plant in Cleveland, Ohio.

Of the 45 target compounds in the wastewater indicator analysis, 16 were not detected in any of the samples (Appendix D). Information on the other 29 wastewater indicators, including the number of detections for each compound, is listed in table 8. Of the 29 compounds detected in environmental samples, 13 were also detected in both of the spiked samples. These were 3β-coprostanol; caffeine; cholesterol; cotinine; ethanol, 2-butoxy-phosphate; fluoranthene; N,N-diethyltoluamide; NPEO1; phenol; para-nonylphenol; pyrene; tri (2-chloroethyl) phosphate; and triclosan.

If a compound was detected in four or more samples, correlation analysis was done to determine the relation between the compound and *E. coli* concentrations (table 8). Two fecal indicators, 3β-coprostanol and cholesterol, were found in one sample or had no relation to *E. coli* concentrations, respectively. Some other general indicators of wastewater contamination—caffeine, NPEO1, and 1,4,-Dichlorobenzene—were poorly correlated to *E. coli* or were not detected enough to evaluate the relation. The concentrations of two compounds, however, showed a significant relation to *E. coli* concentrations at $\alpha=0.1$— fluoranthene and pyrene. (An $\alpha = 0.1$ was used as the level of significance for this analysis, because no concentrations of any wastewater indicators were related to concentrations of *E. coli* at a significance level of $\alpha = 0.05$.) Fluoranthene and pyrene are components of coal tar and asphalt and not necessarily indicative of fecal contamination. Most likely, these compounds are delivered to storm sewers through runoff across asphalt roads and parkings lots. It seems reasonable, therefore, that the delivery of these compounds would correlate well with the delivery of bacteria from CSOs and storm sewers.

For many wastewater compounds, poor or variable method performance was indicated; hence, detected values did not have the precision that one might expect for other compounds. In addition, at the time of sampling and analysis, this method was under development, and improved techniques or new wastewater indicators have since been incorporated into the method. During this study, samples were collected from a variety of study sites having different contaminant sources; this variety makes establishing a significant relation between *E. coli* and wastewater indicators difficult.

In addition to the samples for the wastewater indicators described above, investigators collected samples that were specifically analyzed for caffeine at the USEPA laboratory. The USGS collected 55 samples for caffeine during 2000 distribution and source studies at a variety of locations, including ancillary sites at Mosquito Lake (Appendix E). Local cooperators collected 84 samples for caffeine during 2001 routine studies at bathing-water locations (Appendix E).

An examination of the data from quality-control samples—replicates and blanks—places limitations on the interpretation of caffeine results (Appendix A4). The quality-control data were unacceptable because of two reasons. For one, the range of differences between bottles of replicate samples was 0.004 to 0.431 μg/L of caffeine with an average difference of 0.082 μg/L and an average percent difference of 33.6 percent. Some of these differences were rather large and were the same value as concentrations found in many samples. A second problem involved detections of caffeine in two of the blank samples, indicating the potential for contamination during collection and handling of samples. Because the samples were extracted and frozen for several months before analysis, quality-control data were not available to take corrective actions while sampling was in progress.

A large number of nondetections of caffeine in samples collected during 2000 (Appendix E) necessitated changes to the analytical method to increase sensitivity for samples collected during 2001. As a result, caffeine was detected in only 25 percent of the samples collected in 2000 but in 99 percent of the samples collected in 2001. Of the 14 detections in 2000, the five highest concentrations of caffeine were found in samples collected at Mosquito Lake, and four of these were from North or South Lakeshore

Table 8. Detections of wastewater indicators in samples collected at Lake Erie beaches and Mosquito Lake, 2000, and their relations to $\log_{10}$ *Escherichia coli* concentrations

[µg/L, micrograms per liter of sample; nd, not determined; correlations that are significant at α = 0.1 are in bold text.]

Analyte[a]	Number of detections for		Pearson's r[b]	Method reporting limit (µg/L)	Possible uses or sources[c]
	Samples (n=13)	Spikes (n=2)			
1,4-Dichlorobenzene	2	0	nd	0.03	Moth repellant; lavatory fumigant; deodorant
3β-Coprostanol	1	2	nd	0.60	Found in feces of man and carniverous animals
Bis(2-ethylhexyl) adipate	3	1	nd	2.0	Plasticizer
Bis(2-ethylhexyl) phthalate	2	0	nd	2.5	Plasticizer
Bisphenol A	6	1	0.322	0.09	Manufacturer of epoxy resins and polycarbonates; fungicide; plasticizer; flame retardant
Caffeine	8	2	0.349	0.08	Stimulant and diuretic
Chloropyrifos	1	0	nd	0.02	Domestic pest and termite control
Cholesterol	12	2	0.019	1.50	Found in all body tissues and in animal fats and oils
Codeine	2	0	nd	0.10	Narcotic
Cotinine	6	2	0.306	0.04	Nicotine metabolite
Diazinon	3	0	nd	0.03	pesticide; >40 percent nonagricultural usage
Diethyphthalate	3	1	nd	0.25	Plasticizer
Ethanol,2-butoxy-,phosphate	4	2	0.396	0.2	Plasticizer
Fluoranthene	6	2	0.482	0.03	Coal tar and asphalt
N,N-Diethyltoluamide	11	2	0.268	0.04	Insect repellent; poisonous
Naphthalene	4	1	- 0.225	0.02	From coal tar; toxic to humans; used in manufucture of synthetic resins, celluloid, lubricants, and motor fuels; moth repellent and insecticide; antiseptic
NPEO1-total	5	2	-0.090	1.00	General indicators of wastewater contamination; directly associated with nonionic detergent use
OPEO1	7	1	0.036	0.10	Nonionic detergent metabolite
OPEO2	2	0	nd	0.20	Nonionic detergent metabolite
Para-cresol	3	1	nd	0.04	Wood preservative
Para-Nonylphenol-total	3	2	nd	0.50	In preparation of lubricating-oil additives, resins, surface active agents, and plasticizers
Phenol	7	2	0.136	0.25	From coal tar; general disinfectant for toilets,stables, cesspools, floors, drains, etc.; antiseptic
Phthalic anhydride	4	1	0.168	0.25	Prepared from naphthalene; used in the manufacture of synthetic indigo, artificial resins
Pyrene	6	2	0.494	0.03	Occurs in coal tar; fuel combustion
Tetrachloroethylene	3	0	nd	0.03	Solvent; degreaser; veterinary anthelmintic
Tri (dichloroisopropyl) phosphate	2	1	nd	0.10	Flame retardant

Table 8. Detections of wastewater indicators in samples collected at Lake Erie beaches and Mosquito Lake, 2000, and their relations to $\log_{10}$ *Escherichia coli* concentrations —Continued

[µg/L, micrograms per liter of sample; nd, not determined; correlations that are significant at α = 0.1 are in bold text.]

Analyte[a]	Number of detections for		Pearson's r[b]	Method reporting limit (µg/L)	Possible uses or sources[c]
	Samples (n=13)	Spikes (n=2)			
Tri(2-chloroethyl)phosphate	4	2	0.057	0.04	Hypnotic; sedative; plasticizer; fire retardant
Triclosan	8	2	0.265	0.04	Bacteriostat and preservative for cosmetic and detergent preparations; disinfectant
Triphenyl phosphate	3	0	nd	0.10	Plasticizer; resin; wax; finish; roofing paper

[a] Analytes were included if one or more detections were found.
[b] Correlations were done if four or more detections were found.
[c] Zaugg and others, 2001.

ancillary locations. Samples for caffeine analyses were not collected at Mosquito Lake during 2001. Because sample variability was large and contamination potential was unknown, statistical analysis cannot be done on the caffeine data.

Predictive models for *Escherichia coli*

Summary statistics of *E. coli* concentrations in water collected during the recreational seasons of 2000 and 2001 during routine studies are shown in table 9. Concentrations of *E. coli* were lowest at Fairport Harbor, Mentor Headlands, and during 2001 at Mosquito Lake. Bacterial water quality was generally good at these beaches, exceeding the single-sample bathing-water standard 10 percent or less of the days sampled. At Edgewater, Huntington, and Mosquito Lake, higher *E. coli* concentrations were found in 2000 than in 2001. Villa Angela had the poorest water quality with respect to *E. coli* concentrations among the beaches sampled, exceeding the single-sample bathing-water standard in about 40 percent of the samples collected during 2000 and 2001.

Relations between *Escherichia coli* concentrations and environmental or water-quality variables

Statistical tests were done to evaluate quantitatively the relations between environmental or water-quality variables and *E. coli* concentrations in water. This is the first step in development of predictive models; that is, identifying those factors that show a relation to *E. coli* and may be used in the models. Continuous variables are listed for the three Lake Erie urban beaches (table 10), Mosquito Lake (table 11), and the two Lake Erie beaches in a less populated area (table 12).

Because of the importance of rainfall as a predictor of *E. coli* concentrations, several rainfall variables were developed. For Edgewater, Villa Angela, Huntington, and Mosquito Lake, “rainfall 24” was defined as the amount of rain, in inches, that fell in the 24-hour period preceding the 9 a.m. sampling. “Rainfall 48” and “rainfall 72” were the amounts of rain that fell in the 24-hour periods 2 days and 3 days, respectively, before the 9 a.m. sampling. These two variables were used to determine whether there was a lag between rainfall in the watershed and elevated *E. coli* concentrations. The variables “rainfall weighted 72” and “rainfall local weighted 72” (USGS-installed rain gage) were computed from the rainfall amounts in the 72-hour period preceding the 9:00 a.m. sampling, with the most recent rainfall receiving the highest weight. These variables were computed as the sum of following three values:

- 3 × rainfall amount for the 0- to 24-hour antecedent period
- 2 × rainfall amount for the greater than 24- to 48-hour antecedent period
- 1 × rainfall amount for the greater than 48- to 72-hour antecedent period

“Rainfall weighted 72 × dry days” was a variable developed for Mosquito Lake (table 11). It is the “rainfall weighted 72” times the number of preceding dry days; dry days are days in which the daily rainfall was less than or equal to 0.2 in. For Fairport Harbor and Mentor Headlands, the beaches in a less populated area, rainfall was measured at 5 p.m. daily. The rainfall variables (table 12) were defined as rainfall in the 24-hour period up to 5 p.m. on the previous day (“rainfall antecedent”) and rainfall in the 72-hour period up

Table 9. Summary statistics of *Escherichia coli* concentrations in water collected at five Lake Erie beaches and one inland lake during routine studies, 2000 and 2001

[Concentrations are in colonies per 100 mililiters]

Beach	Number of samples	*Escherichia coli* concentrations			Number (percent) of days bathing-water standard[a] was exceeded
		Median	Minimum	Maximum	
Edgewater					
2000	72	140	4	12,000	24 (33.3)
2001[b]	70	73	10	4,000	15 (21.4)
Villa Angela					
2000	72	170	1	4,500	29 (40.3)
2001[b]	69	130	6	6,400	28 (40.6)
Huntington					
2000[b]	53	130	8	6,600	14 (26.4)
2001[b]	50	43	3	1,200	10 (20.0)
Fairport Harbor					
2000	49	16	2	390	3 (6.1)
Mentor Headlands					
2000[b]	50	14	2	950	5 (10.0)
Mosquito Lake					
2000	53	110	6	5,400	15 (28.3)
2001[b]	52	25	3	320	2 (3.8)

[a] Numbers of days the concentration of *Escherichia coli* in water exceeded the single-sample bathing-water standard of 235 colonies per 100 milliliters.
[b] The daily concentrations were determined by calculating the mean of two water samples

to 5 p.m. on the day of sampling, with the most recent rainfall receiving the highest weight ("rainfall weighted 72").

Other variables need further descriptions. "Date" is based on the chronological day of year and the hypothesis that, in some situations, *E. coli* may accumulate over the course of the summer. "*Escherichia coli* previous day" is the concentration of *E. coli* determined by analyzing a water sample collected on the previous day. "Streamflow 7am" is the instantaneous streamflow of a stream in a nearby watershed at 7 a.m. on the day of sampling. This variable was developed to provide a measure of streamflow that couldeasily be obtained by beach managers for the Lake Erie urban beaches (table 10). At Mosquito Lake, "streamflow previous day" was the daily mean streamflow for the day before the day of sampling (table 11). For Mosquito Lake, the "wind direction" was the instantaneous wind direction at 9 a.m. on the day of sampling; "wind speed" was similarly determined. The cosine and sine of the wind direction at 9 a.m. on the day of sampling also were used as variables. By using the cosine and sine of the wind direction, one is able to assign weighted measures on wind direction; for example, an easterly wind direction would have a sine of 1 and westerly wind direction a sine of –1 with all other directions having intermediate values. This places more weight on winds from an easterly direction.

Table 10. Summary of Pearson's *r* correlations between $\log_{10}$ *Escherichia coli* concentrations in water and environmental or water-quality factors at three Lake Erie urban beaches, 2000 and 2001

[NS, not statistically significant at α = 0.05; all other realations were statistically significant; GLFS, Great Lakes Forecasting System at The Ohio State University; --, no data]

Factor	Edgewater			Villa Angela			Huntington		
	2000	2001	Combined	2000	2001	Combined	2000	2001	Combined
Birds, number on beach at time of sampling	0.293	0.376	0.354	NS	NS	NS	-0.280	NS	NS
Date	NS	0.273	NS	NS	NS	NS	-0.374	0.361	NS
Escherichia coli previous day	NS	NS	0.266	0.342	NS	0.199	NS	0.395	0.370
Rainfall 24 [a]	0.471	NS	0.282	0.466	0.329	0.390	0.468	0.276	0.399
Rainfall 48 [b]	NS	NS	NS	NS	NS	NS	0.277	NS	0.236
Rainfall 72 [c]	NS	NS	NS	NS	NS	NS	NS	NS	NS
Rainfall weighted 72 [d]	0.395	NS	0.256	0.343	0.298	0.320	0.459	NS	0.394
Rainfall local weighted 72 [e]	--	--	--	--	--	--	NS	--	--
Streamflow 7 a.m. [f]	0.269	NS	0.306	0.296	NS	0.234	0.386	NS	0.331
Turbidity	NS	NS	NS	0.516	0.278	0.408	0.673	0.600	0.500
Turbidity, log	NS	NS	NS	0.497	0.381	0.439	0.627	0.511	0.443
Ultraviolet intensity, previous 24 hours	NS	NS	NS	NS	NS	NS	NS	NS	0.253
Water temperature	NS	NS	NS	NS	NS	NS	-0.356	0.543	0.212
Wave height, predicted by GLFS	0.451	NS	0.274	0.573	NS	0.366	0.570	0.349	0.430

[a] Rainfall 24 was the amount, in inches, at Hopkins Airport, Cleveland, Ohio, in the 24-hour period preceding the 9 a.m. sampling.
[b] Rainfall 48 was the amount, in inches, at Hopkins Airport, Cleveland, Ohio, in the 24-hour period two days before the 9 a.m. sampling.
[c] Rainfall 72 was the amount, in inches, at Hopkins Airport, Cleveland, Ohio, in the 24-hour period three days before the 9 a.m. sampling.
[d] Rainfall weighted 72 was the amount, in inches, at Hopkins Airport, Cleveland, Ohio, in the 72-hour period preceding the 9 a.m. sampling, with the most recent rainfall receiving the greatest weight.
[e] Rainfall local weighted 72 was the amount, in inches, at the USGS rainfall gage at Huntington Beach in the 72-hour period preceding the 9 a.m. sampling, with the most recent rainfall receiving the greatest weight.
[f] Streamflow 7 a.m. was the streamflow at the Cuyahoga River at Independence, Ohio, at 7 a.m. on the date of sampling.

One may want to place more weight on winds from an easterly direction, for example, if the source of contamination is suspected to be from a source to the east. “Wind direction, sum sine” and “wind direction, sum cosine” were the sums of the sines and cosines, respectively, of the wind direction at 9 a.m. on the day of sampling, on the previous day, and 2 days prior.

Pearson’s *r* correlation coefficients were computed to assess the linear relation between *E. coli* concentrations and the continuous variables described above. “*E. coli* previous day,” “rainfall 24,” “rainfall weighted 72,” “streamflow 7am,” and “wave height” were statistically related to *E. coli* concentrations for at least one data set at each of the three Lake Erie urban beaches (table 10). “Turbidity” was

Table 11. Summary of Pearson's *r* correlations between $\log_{10}$ *Escherichia coli* concentrations in water and environmental or water-quality factors at Mosquito Lake State Park, 2000 and 2001

[NS, not statistically significant at $\alpha = 0.05$; all other relations were statistically significant; --, no data]

Factor	2000	2001	2000-2001
Birds, number on beach at time of sampling	NS	NS	0.250
Date	0.542	0.535	0.448
Escherichia coli previous day	0.368	0.550	0.610
Flow (average) through dam outlet, previous	NS	0.298	NS
Pool elevation (average), previous day	-0.518	-0.528	NS
Rainfall 24 [a]	NS	0.397	0.203
Rainfall 48 [b]	NS	NS	0.188
Rainfall 72 [c]	NS	NS	NS
Rainfall weighted 72 [d]	NS	0.406	0.282
Rainfall weighted 72 × dry days[e]	NS	0.402	0.232
Specific conductance	NS	--	--
Streamflow previous day[f]	NS	NS	NS
Turbidity	0.495	0.527	0.471
Turbidity, log	0.482	0.576	0.502
Water temperature	NS	NS	NS
Wind direction 9 a.m. [g]	NS	NS	NS
Wind direction 9 a.m., cosine	0.307	NS	NS
Wind direction 9 a.m., sine	NS	NS	NS
Wind direction, sum cosine [h]	0.313	NS	0.174
Wind direction, sum sine [i]	0.358	NS	0.235
Wind speed 9 a.m.	NS	NS	0.231

[a] Rainfall 24 was the amount, in inches, at Youngstown Regional Airport, Youngstown, Ohio, in the 24-hour period preceding the 9 a.m. sampling.
[b] Rainfall 48 was the amount, in inches, at Youngstown Regional Airport, Youngstown, Ohio, in the 24-hour period two days before the 9 a.m. sampling.
[c] Rainfall 72 was the amount, in inches, at Youngstown Regional Airport, Youngstown, Ohio, in the 24-hour period three days before the 9 a.m. sampling.
[d] Rainfall weighted 72 was the amount, in inches, at Youngstown Regional Airport, Youngstown, Ohio, in the 72-hour period preceding the 9 a.m. sampling, with the most recent rainfall receiving the greatest weight.
[e] Rainfall weighted 72 × dry days was the "rainfall weighted 72" times the number of preceding dry days; dry days are days in which the daily rainfall was < 0.2 inch.
[f] Daily mean streamflow at Eagle Creek at Phalanx Station, Ohio, the day before sampling.
[g] Wind direction was instantaneous wind direction at 9 a.m. the day of sampling, measured at Youngstown Regional Airport, Youngstown, Ohio.
[h] Wind direction, sum cosine was the sums of the cosines of the instantaneous wind direction at 9 a.m. on the day of sampling and on the two previous days.
[i] Wind direction, sum sine was the sums of the sines of the instantaneous wind direction at 9 a.m. on the day of sampling and on the two previous days.

Table 12. Summary of Pearson's *r* correlations between $\log_{10}$ *Escherichia coli* concentrations in water and environmental or water-quality factors at Ohio beaches in a less populated area, 2000

[NS, not statistically significant at $\alpha = 0.05$; all other relations were statistically significant]

Factor	Mentor Headlands	Fairport Harbor
Birds, time of sampling	NS	NS
Date	NS	NS
Escherichia coli previous day	NS	NS
Rainfall antecedent [a]	NS	NS
Rainfall weighted 72 [b]	NS	NS
Streamflow 7 a.m.[c]	NS	NS
Streamflow antecedent [d]	NS	NS
Turbidity	0.327	NS
Turbidity, log	0.490	NS
Water temperature	NS	NS

[a] Rainfall antecedent was the amount, in inches, that accumulated in the 24-hour period up to 5 p.m. on the previous day at Painesville Municipal Water Plant, Painesville, Ohio.

[b] Rainfall weighted 72 was the amount, in inches, that accumulated in the 72-hour period up to 5 p.m. on the previous day at Painesville Municipal Water Plant, Painesville, Ohio, with the most recent rainfall receiving the highest weight.

[c] Streamflow, 7 a.m. was the streamflow measured at the USGS gaging station, Grand River near Painesville, Ohio, at 7 a.m. on the date of sampling.

[d] Streamflow, antecedent was the daily mean streamflow for the previous day at the Grand River near Painesville, Ohio.

significantly related to *E. coli* at Villa Angela and Huntington, but not at Edgewater. "Rainfall 48" and "rainfall 72" were not significantly related to *E. coli* concentrations except for a weak correlation for "rainfall 48" at Huntington. At Huntington, the USGS-installed rain gage at the beach ("rainfall local weighted 72") was shown to provide rainfall data that were not related to *E. coli* concentrations, and this gage was discontinued after 2000. At Mosquito Lake (table 11), the factors "date," "*E. coli* previous day," and "turbidity" were related to *E. coli* concentrations for 2000, 2001, and the combined years. The sum of the sines or cosines of the wind directions for the previous 3 days were found to be related to *E. coli* concentrations for 2000 and the combined years, but not for 2001. At Mentor Headlands, only turbidity was related to *E. coli* concentrations (table 12). No statistically significant relations were found between *E. coli* concentrations and any variables at Fairport Harbor (table 12).

Analysis of variance and Tukey's test were used to assess the relations between categorical environmental variables (current direction and estimated wave height) and *E. coli* concentrations in water. Data on $\log_{10}$ *E. coli* concentrations were placed into groups based on the direction of currents the previous day, as predicted by the GLFS (fig. 14). If a current-direction category had less than five observations, categories were combined on the basis of the most reasonable combination of similar wind directions (for example, west and northwest). At Edgewater during 2000 and 2001, significantly higher *E. coli* concentrations were found for southeast currents than for northeast or east currents (fig. 14A). This seems reasonable, because southeast currents would bring contamination into the beach area where it would accumulate behind the Edgewater breakwall that lies to the east of the beach. In contrast, northeast or east currents would divert contamination away from the beach and into the open lake. (See fig. 4.) For modeling at Edgewater, currents were placed into three categories in ascending order: (1) currents to the east and northeast, (2) currents to the west, northwest, southwest, and north, and (3) currents to the southeast. At Villa Angela, no statistically significant differences among any of the current-direction categories were found (fig. 14B). Current direction was not used as a variable for modeling at Villa Angela. At Huntington, significantly higher *E. coli* concentrations were found for southwest currents than for east and southeast currents and west currents (fig. 14C). For modeling at Huntington, currents were placed into categories in ascending order: (1) west, northeast, east, and southeast currents and (2) southwest currents.

Wave heights were placed into four categories based on minimum and maximum heights in each wave train: (1) 0 to 2 ft, (2) 1 to 3 ft, (3) 2 to 4 ft, and (4) 3 to 5 ft or greater. During 2000 and 2001, as was found in the earlier study (Francy and Darner, 1998), median *E. coli* concentrations generally increased with increasing wave height, except for category 4 at Huntington (data not shown). Statistically significant differences of *E. coli* concentrations were found among some of the wave-height categories (data not shown); similar results were also found in the earlier study. As in the earlier study, wave height was found to be important variable affecting *E. coli* concentrations in 2000 and 2001.

Development and evaluation of beach-specific models

Environmental and water-quality variables that were related to *E. coli* concentrations were used to develop statistical models by use of MLR techniques. For each beach, models based on data collected during 1997, 2000, and (or) 2001 (table 13) were developed. The probability of exceeding the single-sample bathing-water standard of 235 col/100 mL was used as the model output variable (hereinafter "probability"). Of the models listed in table 13, those that were determined to be useful to beach managers (in terms of predictive ability and ease of use) were evaluated for their abilities to correctly predict a probability above or below a threshold probability. Some models were tested in subsequent years, if data were available.

The 1997 model for Edgewater Park and Villa Angela and model evaluation steps. A model developed based on 1997 data for Edgewater and Villa Angela during a previous study (Francy and Darner, 1998) was evaluated on 2000 data collected during routine studies. The 1997 model contained weighted categorical rainfall, beach-specific turbidity, and wave height and explained 58 percent of the variability in *E. coli* concentrations at Edgewater and Villa Angela (table 13, lines 1 and 11) and a third beach not evaluated during the current study.

The first step in model evaluation is to determine a threshold probability—that is, the lowest (most conservative) probability that produces the most correct responses and (or) fewest false negative responses. This concept can best be explained by examining the plot for the 1997 model at Villa Angela (fig. 15). Using the 1997 data, a threshold-probability of 44 percent was established for Villa Angela. The plot is divided into four quadrants by drawing a vertical line through 2.37 on the x-axis (represents the $\log_{10}$ of 235 colonies/100 mL) and a horizontal line through the threshold probability of 44. The four quadrants in figure 15 are

1. Correct nonexceedance. *E. coli* concentration met the standard (was less than 235 col/100 mL), and the predicted probability of the exceedance was below the threshold.
2. False positive. *E. coli* concentration met the standard, but the predicted probability of the exceedance was above the threshold.
3. Correct exceedance. Actual *E. coli* concentration exceeded the standard (was greater than 235 col/100 mL), and the predicted probability of exceedance was above the threshold.
4. False negative. Actual *E. coli* concentration exceeded the standard, but the predicted probability of the exceedance was below the threshold.

The threshold probability of 44 gives a total of 33 correct responses (quadrants 1 and 3), 3 false positives (quadrant 2), and 5 false negatives (quadrant 4). These numbers are summarized along with model information in table 13, line 11. Moving the threshold to 43 would result in fewer correct responses; moving the threshold to 45 would not improve on the number of correct responses (fig. 15).

A threshold probability was determined for the 1997 model at Edgewater (fig. 16 and table 13, line 1) with different results and a more complicated determination process. A threshold of 49 would produce the highest number of correct responses but would also produce a high number of false negatives and no false positives. False negative responses are especially troubling because the recreational water-quality is determined to be acceptable when in fact the standard was exceeded. The threshold at 49 would overemphasize the influence of four outlier values and establish a threshold that is too high to be realistic. Instead, establishing a threshold of 26 would still provide a reasonably high number of correct responses, minimize false negative responses, and better represent the system.

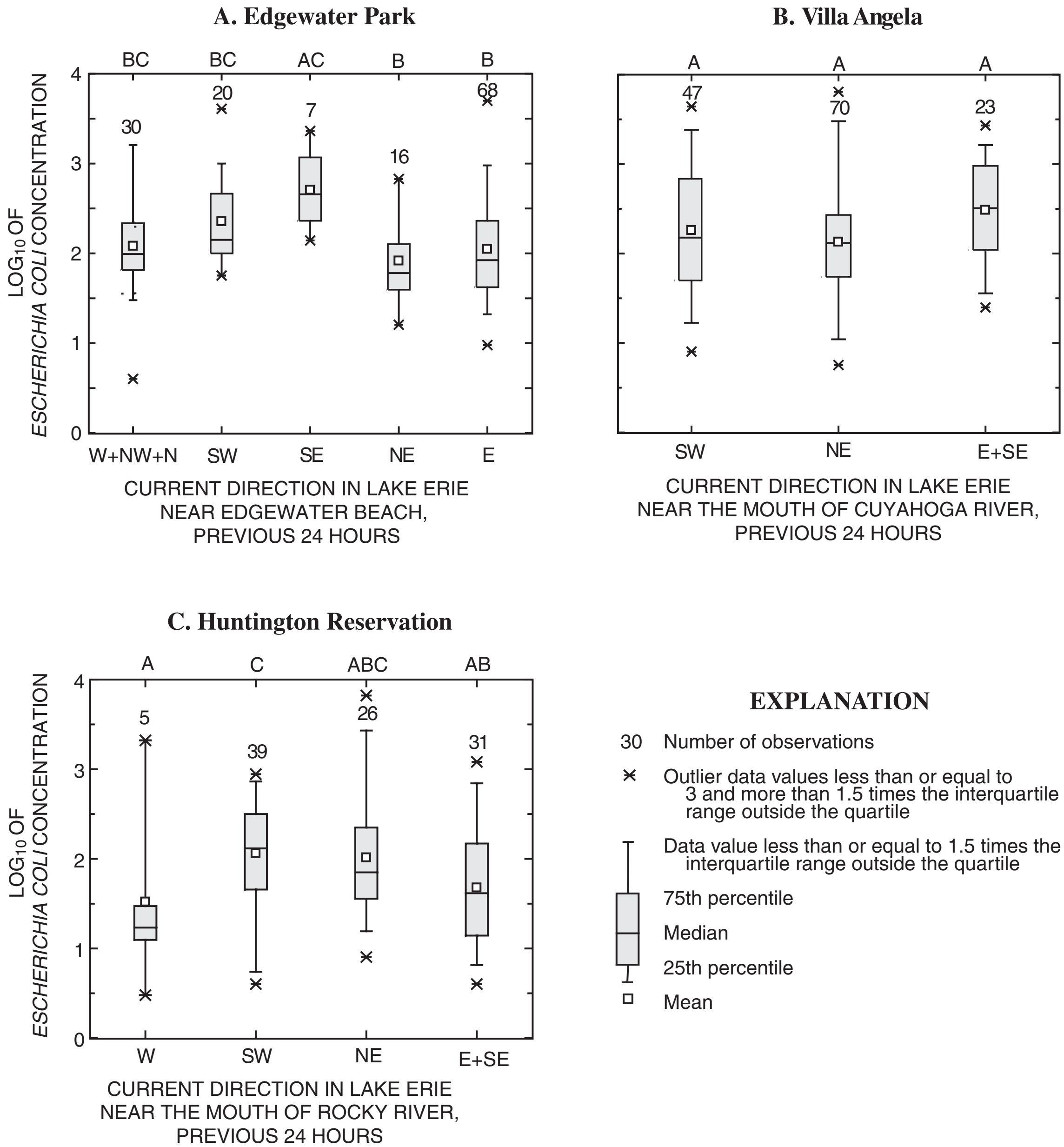


Figure 14. Distribution of *Escherichia coli* concentrations in water by current direction, 2000 and 2001, at (A) Edgewater Park, (B) Villa Angela, and (C) Huntington Reservation. (Results of Tukey's test are presented as letters, and concentrations with at least one letter in common do not differ significantly.)

Table 13. Variables, regression statistics, and threshold probablilities for beach models, 1997, 2000, and 2001

[*E. coli, Escherichia coli*; ND, not determined; R^2, fraction of the variation in the dependent variable that is explained by the model; threshold probability is based on meeting and exceeding the single-sample bathing standard for *Escherichia coli*]

	Beach	Time period of data for model development	Variables in model	R^2 of model	Threshold probability	Number of model responses by category		
						Correct (percent)	False positive (percent)	False negative (percent)
1	Edgewater	1997	Wave height, rainfall weighted 72[a], turbidity	0.58	26	26 (78.8)	5 (15.1)	2 (6.1)
2	Edgewater	1997 and 2000	Wave height, rainfall weighted 72	0.17	ND	ND	ND	ND
3	Edgewater	2000 and 2001	Number of birds[b], current direction[c], wave height, rainfall 24[d]	0.32	45	109 (79.5)	8 (5.8)	20 (14.6)
4	Huntington	2000	Wave height, rainfall weighted 72, turbidity	0.58	43	46 (90.2)	3 (5.9)	2 (3.9)
5	Huntington	2000 and 2001	Wave height, rainfall 24, turbidity, current direction	0.41	31	88 (88.9)	7 (7.1)	4 (4.0)
6	Huntington	2000 and 2001	Wave height, streamflow 7 a.m.[e], turbidity, number of birds	0.40	30	83 (83.8)	7 (7.1)	9 (9.1)
7	Mentor Headlands	2000	Wave height, rainfall weighted 72, log turbidity	0.26	ND	ND	ND	ND
8	Mosquito Lake	2000	Date, sum sine wind direction[f], turbidity, rainfall weighted 72	0.44	47	42 (80.8)	4 (7.7)	6 (11.5)
9	Mosquito Lake	2000 and 2001	Date, sum sine wind direction, rainfall weighted 72, wind speed	0.39	29	90 (90.9)	4 (4)	5 (5.1)
10	Mosquito Lake	2000 and 2001	Rainfall weighted 72 x dry days[g], previous day's *E. coli*, log turbidity	0.48	22	60 (81.1)	7 (9.5)	7 (9.5)
11	Villa Angela	1997	Wave height, rainfall weighted 72, turbidity	0.58	44	33 (80.5)	3 (7.3)	5 (12.2)
12	Villa Angela	1997 and 2000	Wave height, rainfall weighted 72, log turbidity	0.42	43	78 (74.3)	13 (12.4)	14 (13.3)
13	Villa Angela	2000 and 2001	Wave height, rainfall 24, turbidity	0.29	34	93 (73.2)	18 (14.2)	16 (12.6)
14	Villa Angela	2000 and 2001	Wave height, rainfall 24, turbidity, previous day's *E. coli*	0.31	ND	ND	ND	ND

Table 13. Variables, regression statistics, and threshold probablilities for beach models, 1997, 2000, and 2001 —Continued

[*E. coli, Escherichia coli*; ND, not determined; R^2, fraction of the variation in the dependent variable that is explained by the model; threshold probability is based on meeting and exceeding the single-sample bathing standard for *Escherichia coli*]

	Beach	Time period of data for model development	Variables in model	R^2 of model	Threshold probability	Number of model responses by category		
						Correct (percent)	False positive (percent)	False negative (percent)
15	Villa Angela	1997, 2000, and 2001	Wave height, rainfall weighted 72, log turbidity	0.32	39	119 (71.2)	24 (14.4)	24 (14.4)

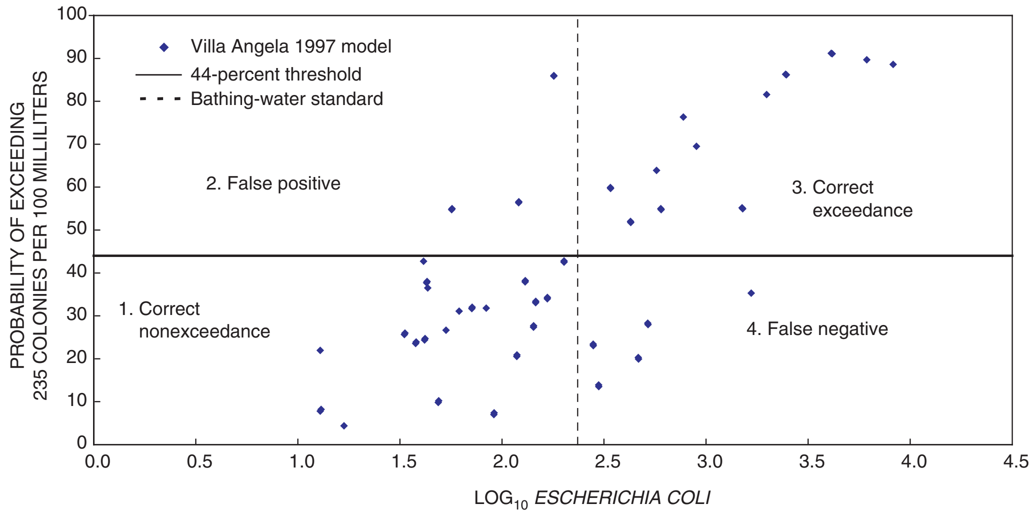
[a] Rainfall weighted 72 was the amount, in inches, at a nearby rain gage in the 72-hour period preceding the 9 a.m. sampling, with the most recent rainfall receiving the greatest weight.
[b] Number of birds on the beach at the time of sampling.
[c] Lake Erie current directions, as predicted by The Ohio State University, Great Lakes Forecasting System.
[d] Rainfall 24 was the amount, in inches, at Hopkins International Airport, Cleveland, Ohio, in the 24-hour period preceding the 9 a.m. sampling.
[e] Streamflow 7 a.m. was the streamflow at the Cuyahoga River at Independence, Ohio, at 7 a.m. on the date of sampling.
[f] Wind direction, sum sine was the sum of the sines of the instantaneous wind direction at 9 a.m. on the day of sampling and on the two previous days.
[g] Rainfall weighted 72 × dry days was the "rainfall weighted 72" times the number of preceding dry days; dry days are days in which the daily rainfall was ≤ 0.2 inch.

Models perform fairly well in predicting responses on the data sets used for model development. A true test of a model is to test responses that result during a subsequent year. The 1997 model was tested on the data collected during routine studies in 2000 by applying the threshold probabilities to the 2000 data sets. Analyzing the 2000 data with the 1997 model produced similar model responses at Edgewater and Villa Angela (table 14, lines 1 and 7). The false negative percentages of 27.1 and 25.0 at Edgewater and Villa Angela, respectively, were high and would need to be reduced for the predictive ability of the model to offer an acceptable level of protection to the public.

As a further test of the model, the predictive ability of the 1997 model can be compared to the current method used to assess recreational water quality. A commonly used method for determining whether to post a beach is to examine the *E. coli* concentration determined from samples collected on the previous day (antecedent *E. coli*). If antecedent *E. coli* is greater than the single-sample bathing-water standard of 235 col/100 mL, the beach is posted with a water-quality advisory; if it is less than the standard, the beach is not posted. The accuracy of predicting current recreational water-quality conditions by use of antecedent *E. coli* concentrations was compared to the accuracy of predicting water quality with the threshold probability and 1997 model. For Edgewater (table 14, lines 1 and 2) and Villa Angela (table 14, lines 7 and 8) during 2000, the 1997 model provided a higher percentage of correct responses and lower percentage of false positive responses than use of antecedent *E. coli* concentrations, but the model erred to a greater degree in the percentage of false negative responses.

Edgewater Park—new models to improve predictions of water quality. Because of the high numbers of false negatives obtained using the 1997 model at Edgewater during 2000, new beach-specific models were developed using 2 years of data and (or) additional variables. First, data for 1997 and 2000 were combined into a new model; this model is worth developing because it can be tested on data available from a subsequent year—2001. However, the 1997 and 2001 model for Edgewater did not improve the predictive ability over the 1997 model; in fact, the R^2 was so low that the model was not considered for further use and testing (table 13, line 2).

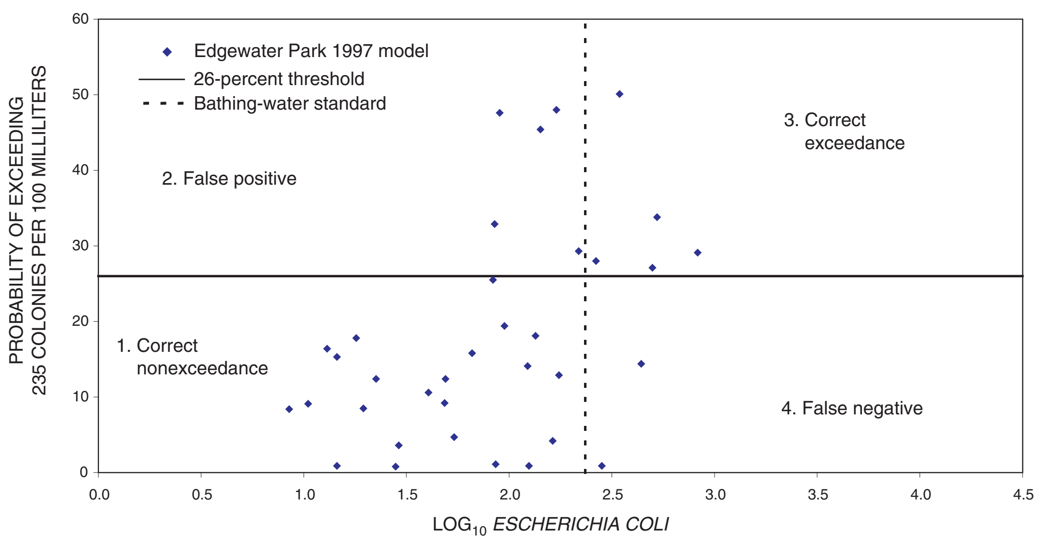
The next step was to attempt to develop an improved model for Edgewater using data for 2000 and 2001. Because the 1997 data lacked some important variables measured in 2000 and 2001 (current direction and number of birds), data from 1997 were not used in further model development. On the basis of data collected during 2000 and 2001, a model was chosen for Edgewater for future use and testing (table 13, line 3 and Appendix F). The model variables included the number of birds, current direction, wave height, and rainfall in the previous 24 hours.



Probability	Total correct	False+	False-
42	31	5	5
43	32	4	5
44	**33**	**3**	**5**
45	33	3	5
46	33	3	5

44-percent threshold

Figure 15. Establishment of the threshold probability based on the single-sample bathing water standard of 235 colonies per 100 milliliters and the 1997 model for Villa Angela, Cleveland, Ohio. (Samples were collected from May through August 1997.)



Probability	Total correct	False+	False-
25	25	6	2
26	**26**	**5**	**2**
27	26	5	2
49	28	0	5
50	28	0	5

26-percent threshold

Figure 16. Establishment of the threshold probability based on the single-sample bathing water standard of 235 colonies per 100 milliliters and the 1997 model for Edgewater Park, Cleveland, Ohio. (Samples were collected from May through August 1997.)

Table 14. Comparison of the numbers of correct and false negative and positive predictions of recreational water-quality conditions using regression models with indicated threshold probabilities and antecedent *Escherchia coli (E. coli)* concentrations

	Beach	Year tested	Model developed	Threshold probability	Model response for year tested			
					Number of samples	Correct predictions (percent)	False negative predictions (percent)	False positive predictions (percent)
1	Edgewater	2000	1997	26	59	41 (69.5)	16 (27.1)	2 (3.4)
2			Antecedent *E. coli*		57	29 (50.9)	13 (22.8)	15 (26.3)
3	Huntington	2001	2000	43	49	40 (81.6)	7 (14.3)	2 (4.1)
4			Antecedent *E. coli*		35	28 (80.0)	5 (14.3)	2 (5.7)
5	Mosquito Lake	2001	2000	47	51	41 (80.4)	0 (0.0)	10 (19.6)
6			Antecedent *E. coli*		38	36 (94.7)	1 (2.6)	1 (2.6)
7	Villa Angela	2000	1997	44	56	39 (69.6)	14 (25.0)	3 (5.4)
8			Antecedent *E. coli*		56	37 (66.1)	9 (16.1)	10 (17.8)
9	Villa Angela	2001	1997 and 2000	43	62	39 (62.9)	17 (27.4)	6 (9.7)
10			Antecedent *E. coli*		51	31 (60.8)	11 (21.6)	9 (17.6)

The threshold probability of 45 resulted in correct responses 79.5 percent of the time, but it produced false negative responses 14.6 percent of the time. The model explained only 32 percent of the variability in *E. coli* concentrations, leaving a large percentage unexplained. Indeed, there were several days in 2000 and 2001 when elevated *E. coli* concentrations were not explained by expected changes in environmental conditions. Perhaps a source of fecal contamination at Edgewater is from localized activity that cannot be measured or predicted, such as the discharge of wastes from boats. This was found to be true in a study in Connecticut where the beach was located near a marina, as is Edgewater.

Villa Angela— new models to improve predictions of water quality. As was done for Edgewater, new models were developed at Villa Angela. At Villa Angela, the combination of 1997 and 2000 data produced a model that explained 42 percent of the variability in *E. coli* concentrations (table 13, line 12). The 1997 and 2000 model tested on 2001 data (table 14, line 9) did not increase the number of correct responses or reduce the false negative percentage as compared to the 1997 model (table 14, line 7) or as compared to use of antecedent *E. coli* concentrations (table 14, line 10).

Steps were taken to improve the predictive ability by developing models for Villa Angela using additional variables measured in 2000 and 2001. Additional variables, however, were not found to be useful at Villa Angela. The two models based on 2000 and 2001 data for Villa Angela contained the same variable list as the 1997 model—wave height, rainfall, and turbidity (table 13, lines 13 and 14). Additional variables were not significant and were not included in the 2000 and 2001 models, except for the previous day's *E. coli*. Adding previous day's *E. coli* to the variable list improved the R^2 slightly (0.31) over not including previous day's *E. coli* (0.29). Because previous day's *E. coli* is not an easily measured variable and *E. coli* results are not usually available in the morning, the model with previous day's *E. coli* seems to be the least desirable from an operational standpoint and was not further evaluated. Furthermore, because the same variables were used in all the Villa Angela models, 3 years of data were combined into a new model (table 13, line 15 and Appendix F). The regression statistics and threshold information indicate that the 2-year and 3-year models were fairly similar in their abilities to assess recreational-water quality (table 13, lines 13 and 15). The 2-year and 3-year models both provided a similar percentage of correct (73.2 and 71.2 percent) and false negative (14.2 and 14.4 percent) responses. Both models left a large amount of the variability in *E. coli* concentrations as unexplained (R^2=0.29 and 0.32), as was found in the Edgewater 2000 and 2001 model.

Fairport Harbor and Mentor Headlands—predictive models. Using the only data available (2000 data), model development steps were taken for Fairport Harbor and Mentor Headlands. At Fairport Harbor, a model was not developed because no variables were found to be significantly related to *E. coli,* and the single-sample bathing-water standard was exceeded on only 3 days during 2000. At Mentor Headlands during 2000, because the standard was exceeded on only 5 days and the best model accounted for only 26 percent of the variability in *E. coli* concentrations, probability thresholds were not developed, and the model was not tested any further (table 13, line 7).

Huntington Reservation—predictive models. Models were developed for Huntington using 2000 or 2000 and 2001 data. The 2000 model contained wave height, rainfall, and turbidity as explanatory variables and established a threshold probability of 43 (table 13, line 4). An R^2 of 0.58 was the highest among any of the 2000 and (or) 2001 models. The 2000 model and associated threshold probabilities were then tested on the 2001 data. During 2001, the 2000 model for Huntington did a good job of predicting exceedance of the *E. coli* standard based on probabilities, and it provided the same percentage of correct responses as using antecedent *E. coli* concentrations (table 14, lines 3 and 4). This is an improvement over current methods, because the 2000 model for Huntington provided a greater number of predictions than by using antecedent *E. coli.* Data on antecedent *E. coli* concentrations are not always available.

Two models were developed and threshold probabilities were determined at Huntington on the basis of 2000 and 2001 data combined. Both models (Appendix F) contained wave height and turbidity with two additional variables—rainfall and currents (table 13, line 5) or streamflow and birds (table 13, line 6). Both models had moderate R^2 values (0.41 and 0.40), set similar threshold values (32 and 30), and produced less than 10 percent false negatives. The decision of which model to use can be based on the preferences of the beach manager; that is, whether the beach manager prefers to use rainfall and currents or use stream flow and birds to predict recreational water quality. Testing these models in subsequent years may reveal the superiority of one of the models in predicting the probability of exceeding the single-sample bathing water standard.

Mosquito Lake State Park—predictive models. Models were developed for Mosquito Lake using 2000 or 2000 and 2001 data. First, a model was developed using 2000 data only. Rainfall and turbidity were the only two variables common to models for Lake Erie beaches and Mosquito Lake. Two additional variables—sum of the sine of the wind direction and date—were also found to help describe the variability in *E. coli* concentrations at Mosquito Lake (table 13, line 8). Mosquito Lake is an inland lake with a long north-to-south fetch length, and the beach is along the southwest shoreline of the lake. It seems reasonable that a wind coming from the east would confine *E. coli* to the beach area. The date variable is less easily explained. It may be possible that in a more closed system such as a reservoir, *E. coli* concentrations tend to build up over the course of the summer. On the other hand, swimmer and bird use of the beach area may increase throughout the summer, leading to increased sources of contamination. Unfortunately, the model could not be adequately tested, because the standard was exceeded on only 2 days during 2001 (table 14, lines 5 and 6).

Threshold probabilities were determined for two models at Mosquito Lake based on 2000 and 2001 data. The variables included in these models were quite different than the variables used to develop models at the Lake Erie beaches. The first model was based on variables that are easily obtained—date, wind direction, rainfall, and wind speed (table 13, line 9 and Appendix F). This model explained 39 percent of the variability in *E. coli* concentrations. Adding previous day's *E. coli* to the second model (table 13, line 10) improved the predictive ability of the model, explaining 48 percent of the variability in *E. coli* concentrations. The second model was different from the first model in other ways; it contained log turbidity (instead of date), did not include variables for wind direction or speed, and included a variable with the number of antecedent dry days. Previous day's *E. coli*, however, is a variable that is not easily obtained nor is it always available. Furthermore, the 2000 and 2001 model without previous day's *E. coli* provided a higher threshold probability, a higher percentage of correct responses, and lower percentages of false negative and false positive responses than the model with previous day's *E. coli*.

Summary and conclusions

Four types of studies were done to strengthen the science that supports monitoring of our Nation's beaches—distribution, source, spatial, and routine studies. Data were collected during the recreational seasons of 2000 and 2001 at six public bathing beaches in Ohio—three urban Lake Erie beaches (Edgewater, Villa Angela, and Huntington), two Lake Erie beaches in a less populated area (Mentor Headlands and Fairport Harbor), and one inland-lake beach

(Mosquito Lake). For distribution, source, and spatial studies, the USGS collected water and sediment samples during 1- to 4-day studies at the three urban Lake Erie beaches and Mosquito Lake. For routine studies, local cooperators collected water samples 4 or 5 days a week at six beaches. During all studies, water and (or) sediment samples were analyzed for concentrations of *E. coli*, and water-quality and environmental data were measured or compiled.

Distribution studies were done to determine whether lake-bottom sediments outside the bathing area were significant deposition areas for *E. coli* and whether sediment and water *E. coli* concentrations were related. Generally, sediments outside the bathing area were not more bacteriologically enriched than those within the bathing area. Concentrations of *E. coli* in water were compared to their concentrations in associated sediments by means of correlation analysis. Sediment and water *E. coli* concentrations were related at Huntington to a greater extent than at Edgewater, Villa Angela, or Mosquito Lake.

Distribution studies were also done to determine concentrations of *E. coli* in interstitial water and subsurface sediments collected from near the swash zone. Concentrations of *E. coli* in interstitial waters ranged from <3 to 400,000 col/100 mL; in subsurface sediments, concentrations ranged from <1 to 30,000 col/g_{DW} sediment. Although there are no standards for *E. coli* in interstitial waters and subsurface sediments collected from near the swash zone, the high concentrations found in this study could be of some concern for public health. A spatial and temporal analysis of *E. coli* concentrations at Edgewater and Mosquito Lake indicated that the concentrations in water and sediment samples collected on the same date from different areas of the beach could vary by as much as 2.5 orders of magnitude. Further, concentrations of *E. coli* during February 2001 were in the same range as many samples collected during the spring or summer months. This finding indicates that there was a continuous source of *E. coli* during the winter and (or) an overwintering of *E. coli* in interstitial waters and subsurface sediments.

Source studies were done at Mosquito Lake to investigate possible sources of fecal contamination to the lake and bathing beach. It was suspected that fecal contamination from the north was affecting water quality at the bathing beach; however, investigators found decreasing *E. coli* concentrations with distance from the north shoreline and then an increase in *E. coli* concentrations at the bathing beach. The source of fecal contamination at the bathing beach is therefore, more likely of local origin. The large population of birds that reside on the beach and are fed by beachgoers has been suggested as a possible source of *E. coli*.

Spatial studies were done to aid The Ohio State University Great Lakes Forecasting System (GLFS) in development of deterministic models to predict *E. coli* concentrations at Edgewater. The data, however, were not very useful to the GLFS because values for the measured properties were fairly homogeneous and the resolution was too small to match the needs of deterministic modeling. Nevertheless, the greatest spatial variations among *E. coli* concentrations, specific conductance, and turbidity in water appeared to occur after a significant rainfall, although more work would be required to verify this finding.

The use of wastewater indicators as surrogates for *E. coli* was tested and evaluated during distribution, source, and routine studies. In a limited study in which 13 samples were collected from a variety of beach locations by the USGS during 2000, none of the compounds commonly found in domestic wastewater—such as 3β-coprostanol, cholesterol, caffeine, NPEO1, and p-nonylphenol—were related to concentrations of *E. coli*. The two compounds that were statistically related to *E. coli*—fluoranthene and pyrene—are components of coal tar and asphalt and not indicative of fecal contamination. Most likely, these compounds are delivered to storm sewers through runoff across asphalt roads and parkings lots. It seems reasonable, therefore, that the delivery of these compounds would correlate well with the delivery of bacteria from combined-sewer overflows and storm sewers. To add evidence to this delivery mechanism, future research could be done with improved analytical methods and a focus on the collection of samples at one beach or several beaches with a common fecal source, and not from a variety of sites, as was done in this study.

Routine studies were done to include research on the development of statistics-based models to predict concentrations of *E. coli*. The first step was to identify variables that showed a relation to *E. coli* and thus were possible explanatory variables in the multiple-linear regression (MLR) models. For continuous variables, correlation analysis was used; for categorical variables, analysis of variance was used. Variables based on the previous day's *E. coli* concentrations, rainfall in the last 24 hours, weighted rainfall in the last 72 hours, streamflow, and wave height were statistically related to *E. coli* concentrations at each of the three Lake Erie urban beaches. Turbidity, an important variable in an earlier study (Francy and Darner, 1998) was significantly related to *E. coli* at Villa Angela and Huntington, but not at Edgewater. At Mosquito Lake, previous day's *E. coli*, date, and turbidity showed the strongest relations to *E. coli* concentrations. Measurements of the wind speed and direction were significantly related to *E. coli* concentrations for some data sets at Mosquito Lake. Two Lake Erie beaches in a less populated area also were investigated. At Mentor Headlands, only rainfall and turbidity were related to *E. coli* concentrations; and at Fairport Harbor, no statistically significant relations were found between *E. coli* concentrations and any variables. Use of analysis of variance revealed significant differences for *E. coli* concentrations among current-direction categories at Edgewater and Huntington, but not at Villa Angela. As in the earlier study (Francy and

Darner, 1998), wave height was found to be important variable affecting *E. coli* concentrations in 2000 and 2001 at Lake Erie urban beaches.

Before any new models were developed, a model based on 1997 data for Edgewater and Villa Angela was tested on 2000 data. The 1997 model contained weighted categorical rainfall, beach-specific turbidity, and wave height. The model output variable is the probability that the single-sample bathing-water standard (235 col/100 mL) is exceeded. Threshold probabilities for the 1997 model were developed and tested on 2000 data. During 2000, the 1997 model provided approximately 70 percent correct responses at Edgewater and Villa Angela. This means that, for 2000, the 1997 model correctly predicted that the *E. coli* concentration would meet or exceed the bathing-water standard 70 percent of the time.

Work continued to develop improved predictive models for Edgewater. Combining the 2000 and 2001 data resulted in an improved model with a different variable list—number of birds, current direction, wave height, and rainfall in the previous 24 hours. This model explained 32 percent of the variability in *E. coli* concentrations, leaving a large percentage unexplained. The source of the unexplained fecal contamination at Edgewater may be from the discharge of wastes from boats en route to a nearby marina.

Attempts to develop new improved models at Villa Angela met with the same limited success as was found at Edgewater. At Villa Angela, unlike Edgewater, expanding the variable list did not prove to be useful. Investigators were left with the same list of explanatory variables for 2000 and 2001 data (wave height, rainfall, and turbidity) as was used for the 1997 model. It seemed reasonable, therefore, to combine 3 years of data—1997, 2000, and 2001—into a Villa Angela model. The 3-year model for Villa Angela explained 32 percent of the variability in *E. coli* concentrations, leaving a large percentage unexplained, as was found in the Edgewater 2000 and 2001 model. Three years of data for model development may be better than 2 years, because 3 years of data would provide a wider range of representative environmental and water-quality conditions. The true test of the three-year model will be its ability to provide accurate assessments of water quality in subsequent years.

Model development at Huntington using 2000 and (or) 2001 data was more successful than at Edgewater and Villa Angela. A model based on 2000 data and containing the variables, wave height, rainfall, and turbidity was tested on 2001 data. During 2001, the 2000 model for Huntington did a good job of predicting exceedance of the *E. coli* standard based on probabilities, and it resulted in the same number of correct and false negative responses as was found using antecedent *E. coli* concentrations. In addition, the 2000 model for Huntington provided a greater number of predictions than by using antecedent *E. coli.* Two models were developed at Huntington on the basis of 2000 and 2001 data. Both models contained wave height, turbidity, and two additional variables and explained approximately 40 percent of the variability in *E. coli* concentrations. The decision of which model to use can be based on the preferences of the beach manager and the performance of the models in predicting recreational water quality in subsequent years.

Mosquito Lake is an inland lake; therefore, the hydrologic processes affecting the beach are different from those affecting the Lake Erie beaches. It is not surprising that the variables included in two 2000 and 2001 models for Mosquito Lake were somewhat different from the variables used to develop models at the Lake Erie beaches. The two models for Mosquito Lake contained various combinations of easily obtained variables—rainfall, number of dry days preceding a rainfall, date, wind direction, wind speed, and turbidity. The first model did not include the previous day's *E. coli*, whereas the second one did; R^2 values were 0.39 and 0.49, respectively. In addition, adding previous day's *E. coli*, a hard to measure variable, did not improve the ability of the model to correctly predict probabilities. Both models need to be tested on data collected in subsequent years.

The variables that best explained the variability in *E. coli* concentrations have been identified for the beaches investigated during this study. The models are beach specific; that is, different variable lists were used to predict the probability of exceeding the standard at each beach. Additional research could include testing the models already developed for Edgewater, Villa Angela, Huntington, and Mosquito Lake in subsequent years and comparing each model's ability to predict recreational water quality with results obtained by use of antecedent *E. coli* concentrations. Each year the model is tested, new data can be added and model variables can be recalculated to determine whether the predictive ability improves with an added year of data. If the model is able to predict recreational water quality as well as or better than use of antecedent *E. coli* for several years in a row, beach managers may consider using the models to aid or direct decisions on posting beach advisories.

References

Accuweather, Inc., 2001, Fifteen-day forecast: accessed May through August 2001 at URL www.accuweather.com

American Society of Agronomy, and Soil Science Society of America, 1982, Methods of soil analysis, Part 2—Chemical and microbiological properties (2d ed.): Madison, Wis., American Society of Agronomy, p. 790-791.

Brandt, Elaine, Aruscavage, Phillip, and Papp, Clara, 1990, Determination of carbonate carbon in geologic materials by coulometric titration, *in* Arbogast, B.F., ed., 1990, Quality assurance manual for the Branch of Geochemistry: U.S. Geological Survey Open-File Report 90-668, p. 68-72.

Burton, G.A., Jr., Gunnison, D., and Lanza, G.R., 1987, Survival of pathogenic bacteria in various freshwater sediments: Applied and Environmental Microbiology, v. 53, no. 4, p. 633-638.

Embrey, S.S., 2001, Microbiological quality of Puget Sound Basin streams and identification of contaminant sources: Journal of the American Water Resources Association, v. 37, no. 2, p. 407-421.

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 Investigations Report 98-4241, 41 p.

Francy, D.S., Jones, A.L., Myers, D.N., Rowe, G.L., Eberle, Michael, and Sarver, K.M., 1998, Quality-assurance/quality-control manual for collection and analysis of water-quality data in the Ohio District, U.S. Geological Survey: U.S. Geological Survey Water-Resources Investigations Report 98-4057, 71 p.

Francy, D.S., Stoeckel, D.M., Bushon, R.N., and Kephart, C.M., 2001, Quality-assurance and quality-control manual for the Ohio District Microbiology Laboratory, accessed at http://oh.water.usgs.gov/micro/qcmanual/manual.html

Guy, H.P., 1969, Laboratory theory and methods for sediment analysis: U.S. Geological Survey Techniques of Water-Resources Investigations, book 5, chap. C1 [variously paged].

Grant, S.B., Sanders, B.F., Boehm. A.B., Redman, J.A., Kim, J.H., Mrse, R.D., Chu, A.K., Gouldin, M., McGee, C.D., Gardiner, N.A., Jones, B.H., Svejkovsky, J., Leipzig, G.V., and Brown, A., 2001, Generation of enterococci bacteria in a coastal saltwater marsh and its impact on surf zone water quality: Environmental Science & Technology, v. 35, no. 12, p. 2407-2416.

Hach Company, 1989, Water analysis handbook: Loveland, Colo., p. 589.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: New York, Elsevier Science Publishing Company [variously paged].

Kelly, A.G., and Campbell, L.A., 1995, Persistent organochlorine contaminants in the Firth of Clyde in relation to sewage sludge input: Marine Environmental Research, v. 41, no. 1, p. 99-132.

LaLiberte, P., and Grimes, D.J., 1982, Survival of *Escherichia coli* in lake bottom sediment: Applied and Environmental Microbiology, v. 43, no. 3, p. 623-628.

Leeming, R., Ball, A., Ashbolt, N., and Nichols, P., 1996, Using faecal sterols from humans and animals to distinguish faecal pollution in receiving waters: Water Research, v. 30, no. 12, p. 2893-2900.

Leeming, R., and Nichols, P.D., 1996, Concentrations of corpostanol that correspond to existing bacterial indicator guidelines limits: Water Research, v. 30, no. 12, p. 2997-3006.

Mallows, C.L., 1973, Some comments on Cp: Technometrics, v. 15, p. 661-675.

Matson, E.A., Hornor, S.B., and Buck, J.D., 1978, Pollution indicators and other microorganisms in river sediment: Journal of Water Pollution Control Federation, v. 50, p. 13-19.

Myers, D.N., and Sylvester, M.D., 1997, National field manual for the collection of water-quality data—Biological indicators: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A7, 38 p.

National Oceanic and Atmospheric Administration, 2001: National Virtual Data System, National Climatic Data Center, Asheville, N.C., accessed September 2001 at URL http://nndc.noaa.gov/

Natural Resources Defense Council, 2001, Testing the waters—A guide to water quality at vacation beaches: New York, Natural Resources Defense Council, 276 p.

Nguyen, D.-K., Bruchet, A., and Arpino, P., 1995, Determination of sterols in sewage sludge by combined *in situ* trimethylsilylation/supercritical fluid extraction and GC/MS: Environmental Science & Technology, v. 29, p. 1686-1690.

Ogunseitan, O.A., 1996, Removal of caffeine in sewage by *Pseudomonas putida*—Implications for water pollution index: World Journal of Microbiology and Biotechnology, v. 12, p. 251-256.

Ohio Environmental Protection Agency, 2002, Water use definitions and statewide criteria: Ohio Administrative Code, chap. 3745-1-07, July 21, 2002, p. 9 and 26.

Patton, C.J., and Truitt, E.P., 2000, Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory—Determination of ammonium plus organic nitrogen by a Kjeldahl digestion method and an automated photometric finish that includes digest cleanup by gas diffiusion: U.S. Geological Survey Open-File Report 00-170, 31 p.

Pereira, W.E., Moody, J.A., Hostettler, F.D., Rostad, C.E., and Leiker, T.J., 1995, Concentrations and mass transport of pesticides and organic contaminants in the Mississippi River and some of its tributaries, 1987-89 and 1991-92: U.S. Geological Survey Open-File Report 94-376, 169 p.

Piocos, E.A., and de la Cruz, A.A., 2000, Solid phase extraction and high performance liquid chromatography with photodiode array detection of chemical indicators of human fecal contamination in water: Journal of Liquid Chromatography and Relevant Technology, v. 23, p. 1281 and 1291.

Podber, D.P., Yen, C.C.J., Regenmorter, L., Bedford, K.W., and Wai, O.W.H., 1994, A nearshore model for Cleveland CSO remediation design selection, Proceedings of the National Hydraulic Conference, ASCE, Buffalo, N.Y.: p. 222-226.

Schiff, K.C., and Kinney, Patrick, 2000, Tracking sources of bacterial contamination in stormwater discharges from Mission Bay, California: Southern California Coastal Water Research Project, annual report 1999-2000, accessed October 9, 2001, at URL http://www.sccwrp.org/

Schwab, D.J. and Bedford, K.W., 1994, Initial implementation of the Great Lakes Forecasting System—A real-time system for predicting lake circulation and thermal structure: Water Pollution Research Journal of Canada, v. 29, nos. 2/3, p. 203-220.

Seiler, R.L., Zaugg, S.D., Thomas, J.M., and Howcroft, D.L., 1999, Caffeine and pharmaceuticals as indicators of waste water contamination in wells: Ground Water, v. 37, no. 3, p. 405-410.

Sherer, B.M., Miner, J.R., Moore, J.A., and Buckhouse, J.C., 1992, Indicator bacterial survival in stream sediments: Journal of Environmental Quality, v. 21, p. 591-595.

Shindel, H.L., Mangus, J.P., and Trimble, L.E., 2001, Water resources data, Ohio, water year 2000: U.S. Geological Survey Water-Data Report OH-00-2, v. 2, St. Lawrence River Basin and statewide project data, 440 p.

Shindel, H.L., Mangus, J.P., and Trimble, L.E., 2002, Water resources data, Ohio, water year 2001: U.S. Geological Survey Water-Data Report OH-01-2, v. 2., St. Lawrence River Basin and statewide project data, 405 p.

U.S. Environmental Protection Agency, 1985, Test methods for *Escherichia coli* and enterococci in water by the membrane-filter procedures: Cincinnati, Ohio, EPA-600/4-85-076, 24 p.

U.S. Environmental Protection Agency, 1998, Bacterial water quality standards for recreational waters—Status report: Washington, D.C., Office of Water, EPA-823-R-98-003.

U.S. Environmental Protection Agency, 1999a, EPA action plan for beaches and recreational waters: Washington, D.C., Office of Water, EPA-600-R-98-079.

U.S. Environmental Protection Agency, 1999b, Review of potential modeling tools and approaches to support the BEACH program: Washington, D.C., Office of Science and Technology, EPA-823-R-98-002.

Wershaw, R.L., Fishman, M.J., Grabbe, R.R., and Lowe L.E, 1987, Carbon, inorganic plus organic, total in bottom material, dry weight, induction furnace *in* Methods for the determination of organic substances in water and fluvial sediments: U.S. Geological Survey Techniques of Water-Resources Investigations, book 5, chap. A3, 80 p.

Wilde, F.D., and Radtke, D.B., 1998, National field manual for the collection of water-quality data—Field measurements: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A6.