USGS
science for a changing world

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

Factors Affecting *Escherichia coli* Concentrations at Lake Erie Public Bathing Beaches

Water-Resources Investigations Report 98-4241

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

COVER: USGS scientists collect and record turbidity and wave height data at Sims Park, Euclid, Ohio.

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

Factors Affecting *Escherichia coli* Concentrations at Lake Erie Public Bathing Beaches

By Donna S. Francy and Robert A. Darner

Water-Resources Investigations Report 98-4241

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
Bruce Babbitt, Secretary

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

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

For additional information write to:

District Chief
U.S. Geological Survey
975 West Third Avenue
Columbus, OH 43212-3192

Copies of this report can be purchased from:

U.S.Geological Survey
Branch of Information Survices
Box 25286
Denver, CO 80225-0286

CONTENTS

Abstract ... 1
Introduction ... 2
    Purpose and scope ... 3
    Previous studies ... 3
    Acknowledgments ... 4
Methods of study ... 4
    Site selection and sampling frequency ... 4
    Collection of water and sediment samples ... 7
    Analysis of water and sediment samples ... 8
    Collection and compilation of ancillary information ... 9
    Statistical methods ... 10
    Quality-assurance and quality-control practices ... 11
Factors affecting *Escherichia coli* concentrations in water and lake-bottom sediments ... 12
    Relations between *Escherichia coli* concentrations and environmental or water-quality variables ... 20
    Prediction of *Escherichia coli* concentrations from environmental and water-quality variables ... 24
        Prediction of *Escherichia coli* concentrations from turbidity—simple linear regression ... 24
        Prediction of *Escherichia coli* concentrations from several environmental and water-quality factors—multiple linear regression ... 26
    Comparison of multiple-linear-regression model to current methods for evaluating beach water quality ... 28
Summary and conclusions ... 30
References cited ... 32
Appendix A—Evaluation of methods for enumeration of fecal-indicator bacteria in lake-bottom sediment ... 34
Appendix B—Variability of *Escherichia coli* concentrations in water and sediment ... 36

FIGURES

1. Map showing locations of beach study sites—Edgewater Park, Villa Angela, and Sims Park—in the Cleveland, Ohio, metropolitan area, 1997 ... 5
2. Map showing sampling areas at (A) Edgewater Park and (B) Villa Angela, Cleveland, Ohio, and (C) Sims Park, Euclid, Ohio ... 6
3-5. Graphs showing mean concentrations of *Escherichia coli* in water and lake-bottom sediments, rainfall amounts, overflow indicators, and wave heights for the recreational season of 1997:
    3. Edgewater Park ... 14
    4. Villa Angela ... 15
    5. Sims Park
        a. Area 1 ... 16
        b. Area 2 ... 17
        c. Area 3 ... 18
        d. Area 4 ... 19
6-7. Graphs showing *Escherichia coli* concentrations in water at Lake Erie beaches, May-September 1997:
    6. By wind direction ... 23
    7. By wave height ... 23
8. Graphs showing regression relations between *Escherichia coli* concentrations and turbidity at (A) Edgewater Park, (B) Villa Angela, Cleveland, Ohio, and (C) Sims Park, Euclid, Ohio, and (D) for all data combined, 1997 ... 25

FIGURES -continued

A1-2. Graphs showing concentrations of fecal coliforms recovered from Lake Erie bottom sediments:
A1. Sims Park, both recovery methods 35
A2. Standard method, various beaches 36
B1-2. Graphs showing differences between quality-control replicate samples for concentrations of *Escherichia coli* in (A) water and (B) sediment:
B1. Percent differences 38
B2. Absolute value $\log_{10}$ differences 39

TABLES

1. Constituents determined on water and lake-bottom sediment samples collected during field studies from May through September 1997 at three Lake Erie beaches 8
2. Concentrations of *Escherichia coli* in water and lake-bottom sediments collected at three Lake Erie beaches on 41 selected mornings from May through September 1997 12
3. Particle-size analysis and total organic carbon concentrations in lake-bottom sediments collected at three Lake Erie beaches, 1997 13
4. Summary of correlations between $\log_{10}$ *Escherichia coli* concentrations in water and environmental or water-quality factors at three Lake Erie beaches, May-September 1997 22
5. Temporal difference in *Escherichia coli* concentrations in water and the relation to number of swimmers on 10 selected days, June-September 1997, at three Lake Erie beaches 24
6. Regression statistics for $\log_{10}$ *Escherichia coli* concentrations and $\log_{10}$ turbidities 26
7. Prediction of *Escherichia coli* concentrations using the multiple-linear-regression model and different combinations of randomly selected explanatory variables 28
8. Classification table for predictions of current recreational water-quality conditions using antecedent *Escherichia coli* concentrations for Edgewater Park, Cleveland, Ohio, May-September 1997 29
9. Results of classification tables comparing the proportions of correct and incorrect predicted recreational water-quality conditions using antecedent *Escherichia coli* concentrations and a multiple-linear-regression model for Lake Erie beaches, May-September 1997 30
B1. Summary statistics for within-bottle and between-bottle differences of replicate quality-control samples for concentrations of *Escherichia coli* 40
B2. Statistical analysis of between-bottle and within-bottle differences of replicate quality-control samples grouped by magnitude of concentration and by beach 41

CONVERSION FACTORS AND ABBREVIATED WATER-QUALITY UNITS

Multiply	By	To obtain
micrometer (µm)	0.00003937	inch
millimeter (mm)	0.03937	inch
foot (ft)	0.3048	meter
mile (mi)	1.609	kilometer
milliliter (mL)	0.06102	cubic inch
liter (L)	0.03531	cubic foot
gallon per minute (gal/min)	3.785	liter per minute
million gallons per day (Mgal/d)	3,785	cubic meter per day
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-quality units used in this report: Chemical concentrations in water are reported in milligrams per liter (mg/L). Milligrams per liter is a unit expressing the concentration of chemical constituents as weight (milligrams) of chemical per unit volume (liter) of water.

Abbreviated sediment-quality units used in this report: Chemical concentrations in sediment are reported in milligrams per kilogram (mg/kg).

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}).

Turbidity is reported in Nephelometric Turbidity Units (NTU's).

Factors Affecting *Escherichia coli* Concentrations at Lake Erie Public Bathing Beaches

By Donna S. Francy and Robert A. Darner

Abstract

The environmental and water-quality factors that affect concentrations of *Escherichia coli* (*E. coli*) in water and sediment were investigated at three public bathing beaches—Edgewater Park, Villa Angela, and Sims Park—in the Cleveland, Ohio metropolitan area. This study was done to aid in the determination of safe recreational use and to help water-resource managers assess more quickly and accurately the degradation of recreational water quality.

Water and lake-bottom sediments were collected and ancillary environmental data were compiled for 41 days from May through September 1997. Water samples were analyzed for *E. coli* concentrations, suspended sediment concentrations, and turbidity. Lake-bottom sediment samples from the beach area were analyzed for *E. coli* concentrations and percent dry weight. Concentrations of *E. coli* were higher and more variable at Sims Park than at Villa Angela or Edgewater Park; concentrations were lowest at Edgewater Park. Time-series plots showed that short-term storage (less than one week) of *E. coli* in lake-bottom sediments may have occurred, although no evidence for long-term storage was found during the sampling period. *E. coli* concentrations in water were found to increase with increasing wave height, but the resuspension of *E. coli* from lake-bottom sediments by wave action could not be adequately assessed; higher wave heights were often associated with the discharge of sewage containing *E. coli* during or after a rainfall and wastewater-treatment plant overflow.

Multiple linear regression (MLR) was used to develop models to predict recreational water quality at the three beaches using the variables shown to be related to *E. coli* concentrations in water. The related variables included turbidity, antecedent rainfall, antecedent weighted rainfall, volumes of wastewater-treatment plant overflows and metered outfalls (composed of storm-water runoff and combined-sewer overflows), a resuspension index, and wave heights. For the beaches in this study, wind speed, wind direction, water temperature, and the presence of swimmers were not included in the model because they were shown to be statistically unrelated to *E. coli* concentrations.

From the several models developed, one model was chosen that accounted for 58 percent of the variability in *E. coli* concentrations. The chosen MLR model contained weighted categorical rainfall, beach-specific turbidity, wave height, and terms to correct for the different magnitudes of *E. coli* concentrations among the three beaches. For 1997, the MLR model predicted the recreational water quality as well as, and in some cases better than, antecedent *E. coli* concentrations (the current method). The MLR model improved the

sensitivity of the prediction and the percentage of correct predictions over the current method; however, the MLR model predictions still erred to a similar degree as the current method with regard to false negatives. A false negative would allow swimming when, in fact, the bathing water standard was exceeded.

More work needs to be done to validate the MLR model with data collected during other recreational seasons, especially during a season with a greater frequency and intensity of summer rains. Studies could focus on adding to the MLR model other environmental and water-quality variables that improve the predictive ability of the model. These variables might include concentrations of *E. coli* in deeper sediments outside the bathing area, the direction of lake currents, site-specific-rainfall amounts, time-of-day information on overflows and metered outfalls, concentrations of *E. coli* in treated wastewater-treatment plant effluents, and occurrences of sewage-line breaks. Rapid biological or chemical methods for determination of recreational water quality could also be used as variables in model refinements. Possible methods include the use of experimental rapid assay methods for determination of *E. coli* concentrations or other fecal indicators and the use of chemical tracers for fecal contamination, such as coprostanol (a degradation product of cholesterol) or caffeine.

Introduction

Lake Erie is a valuable resource for the people of Ohio and nearby states for swimming, boating, and fishing. Water-quality advisories and beach closings because of sewage contamination are common at Lake Erie beaches in Ohio and other beaches in the United States. During 1995, United States ocean, bay, and Great Lakes beaches were closed or advisories were issued against swimming on more than 3,522 occasions (Natural Resources Defense Council, 1996). Fecal-coliform concentrations in excess of Ohio's bathing-water standard (geometric mean of 200 colonies per 100 milliliters) resulted in 65 beach advisories in 1990 and 34 in 1991 at state-park beaches along Lake Erie and at inland reservoirs and lakes (Ohio Environmental Protection Agency, 1992).

Most states have adopted recreational water-quality standards based on concentrations of fecal-indicator bacteria to protect citizens from the risk of contracting waterborne disease from exposure to sewage-contaminated waters. Fecal-indicator bacteria are not typically disease causing (pathogenic), but they indicate the possible presence of pathogenic organisms. In the United States, the U.S. Environmental Protection Agency (USEPA) recommends the use of *Escherichia coli (E. coli)* or enterococci as the preferred and most useful fecal indicators of the quality of freshwater recreational waters for body contact. A direct relation has been demonstrated in freshwater between the rate of gastroenteritis among swimmers and the concentrations of *E. coli* or enterococci, but not fecal coliforms (Dufour, 1984).

In Ohio, water-resource managers have the choice of using *E. coli* or fecal coliforms as the basis for recreational water-quality standards (Ohio Environmental Protection Agency, 1992). If concentrations of either *E. coli* or fecal-coliform bacteria exceed the state standard, then beach managers, at their discretion, may post the beach with a water-quality advisory. A factor complicating the assessment of recreational water quality is that standard tests for determining concentrations of fecal indicators, including *E. coli*, take at least 24 hours to complete. The elapsed time between the occurrence and detection of elevated fecal-indicator concentrations is too long to take adequate control measures in a timely manner; consequently, concentrations may change dramatically between the time of sampling and the reporting of results.

One alternative to waiting 24 hours for results of fecal-indicator concentrations is to use water-quality and environmental surrogates to predict recreational water quality. For example, waters at public bathing beaches frequently contain concentrations of fecal-indicator bacteria that exceed bathing-water standards during periods of rainfall and runoff. In Ohio, sources of these bacteria include street refuse, animal waste, sanitary sewer overflows, and combined-sewer overflows (CSO's) that occur when the capacity of the sewage-collection system is exceeded (Myers, 1992). In the United States, officials cited rain as the cause of 371 beach closings during 1995; however, 510 beach closings during the same year were not associated with rain, and the specific source of pollution was reported as unknown (Natural Resources Defense Council, 1996). Therefore, other factors that affect fecal-indicator concentrations need to be examined. There is speculation that one mechanism of contamination is the resuspension of bacteria previously deposited into lake-bottom sediments from CSO's and other sources.

To better understand the water-quality and environmental factors that affect the degradation of recreational water quality, the U.S. Geological Survey (USGS), in cooperation with the Ohio Water Development Authority, Northeast Ohio Regional Sewer District, Cuyahoga County Board of Health, Cuyahoga County Sanitary Engineers, Cuyahoga River Community Planning Organization, and the Ohio Lake Erie Office, studied the occurrence of *E. coli* in water and lake-bottom sediments and the environmental and water-quality factors that affected *E. coli* concentrations at Lake Erie public bathing beaches in Ohio.

Purpose and scope

This report describes field studies done throughout the recreational season of 1997 (May through September) at three public bathing beaches in the Cleveland, Ohio, metropolitan area. Concentrations of *E. coli* were determined in water and lake-bottom sediments collected during a variety of environmental conditions—during dry, calm weather; before, during, and after rainfall; for various increased wave heights; and before and after heavy recreational use. The concentrations of *E. coli* in water and lake-bottom sediments were plotted as a function of time along with wave heights, occurrence of wastewater-treatment plant overflows, and rainfall amounts. These plots were examined qualitatively to determine if these factors affected *E. coli* concentrations in water and sediment. Statistical methods were used to evaluate quantitatively the relations between *E. coli* concentrations in water and several measured variables—rainfall amount, volumes of wastewater-treatment plant overflows and metered outfalls (composed of storm-water runoff and CSO's), wind speed and direction, a resuspension indicator, turbidity, suspended-sediment concentration, water temperature, wave height, and number of swimmers. Regression techniques were then used to develop a predictive model for *E. coli* concentrations in water at the beaches studied; the best predictive model contained terms for turbidity, antecedent rainfall, and wave height.

This report provides water-resource managers with information on the water-quality and environmental factors that affect fecal-indicator concentrations at three Lake Erie beaches. It also provides evidence that predictive models may be developed to help water-resource managers more quickly and accurately assess the degradation of recreational water quality to protect the public health.

Previous studies

Environmental factors have been shown to be related to concentrations of fecal-indicator bacteria in recreational waters. Several studies showed that a reservoir of sediment-stored fecal-indicator bacteria may be returned to the water column by physical disturbances of bottom sediments; this included dredging (Grimes, 1980), wind and wave actions (Lehman and Fogel, 1976), and disturbance of sediments by swimmers or boaters (Bromel and others, 1978). In two studies (Sherer and others, 1988; Stephenson and Rychert, 1982), investigators found that by disturbing the streambed sediments with a rake, fecal coliforms could be resuspended in the water column and detected downstream. Aldom and others (1998) found a relation between *E. coli* concentrations and wind speed, wind direction, and wave height at Lake Huron bathing beaches. They suggested that these factors may be used to develop models to predict *E. coli* concentrations.

Several investigators examined the relation between fecal-indicator concentrations and water-quality variables. Grimes (1980) found fecal-indicator concentrations to be highly correlated with turbidity at a Mississippi River site downstream from dredging operations. In contrast, in a study of canals along the Texas Coast (Goyal and others, 1977), investigators found no statistically significant relation between fecal-indicator concentrations and temperature, pH, turbidity, or suspended solids concentrations. Tunnicliff and Brickler (1984) found a statistically significant correlation between turbidities and fecal-coliform concentrations in samples collected during storm events but not for samples collected during base flow.

Because bacteria survive longer in sediments than in water, a process affecting concentrations of fecal indicators may be the resuspension of accumulated bacteria from bottom sediments (Marino and Gannon, 1991). The large surface area for attachment and the nutrient-rich environment that sediments provide have been shown to promote survival of bacteria (Burton and others, 1987; LaLiberte and Grimes, 1982; Matson and others, 1978; Sherer and others, 1992). Gerba and McLeod (1976) attributed the longer survival of *E. coli* in sediment than in water to the higher content of organic matter in the sediment. In addition, bacteria sorbed to sediments may be protected from attack by predators and bactericidal factors such as ultraviolet radiation (Davies and others, 1995; Pommepuy and others, 1992). LaLiberte and Grimes (1982) investigated the survival of *E. coli* in bottom sediments in dialysis bags in a Wisconsin lake and found that sand and mud sediments supported the survival of *E. coli* for the length of the study, 4 days. In another study, Davies and others (1995) used membrane diffusion chambers placed at a river site in Australia. Fecal coliforms survived in freshwater sediments for up to 60 days, although their numbers decreased 2-3 orders of magnitude after 29 days and then stabilized. In laboratory experiments (Sherer and others, 1992), half-lives of fecal coliforms ranged from 11 to 30 days in fine and coarse sediments; the half-life of fecal coliforms in the overlying water was only 2.8 days. According to Marino and Gannon (1991), storm-drain sediments were acting as reservoirs of fecal indicators during warm, dry weather periods for up to 6 days.

Because of the extended survival of bacteria in sediments, LaLiberte and Grimes (1982) suggested that the enumeration of sediment-associated fecal bacteria in recreational areas is as important as the enumeration of fecal bacteria in the water column. However, there is no available information on the occurrence of stored fecal-indicator bacteria in lake-bottom sediments at Lake Erie beaches. More information is needed about the role of physical disturbances on the resuspension of sediment-stored bacteria and about the importance of other environmental and water-quality factors on the concentrations of fecal-indicator bacteria in relation to degradation of recreational waters.

Acknowledgments

The authors thank Eva Roller, Lester Stumpe, Keith Linn, and Frank Foley of the Northeast Ohio Regional Sewer District for their assistance during data-collection and report-review phases of the project. The authors also acknowledge the assistance of others in the planning and implementation of this project—Wayne Holmes and Thomas Filbert of the Ohio Department of Natural Resources, Robert Gall of the City of Euclid, and Donald Killinger of the Cuyahoga County Health Department. The authors thank Donna Childs and Timothy Gallagher of the Cuyahoga County Health Department for help with sampling and analysis and Gary Tasker of the USGS for assistance with the statistical modeling. A special thanks is extended to the management and staff of the Cuyahoga County Sanitary Engineers—Ruth Langsner, John Campbell, Ann McCready-Gliha, and Suzanne Oravec—for the use of their laboratory facilities and assistance with laboratory activities.

Methods of study

Data were collected during eight field studies throughout the 1997 recreational season—May through September. The duration of each study ranged from 3 to 14 days, totaling 41 days of data collection. The field studies were done during a range of conditions—during dry, calm weather; before, during, and after rainfall; during increased wave heights; and before and after heavy recreational use.

In this investigation, *E. coli* concentrations were used to monitor recreational water quality because *E. coli* is better than fecal coliforms as an indicator of the risk of swimming in fecal-contaminated waters. For *E. coli*, the Ohio geometric-mean bathing-water standard is 126 colonies per 100 mL (col/100 mL); the single-sample bathing water standard is 235 col/100 mL. The geometric mean is based on a minimum of five samples collected in a 30-day period, and it is used in this report to evaluate median *E. coli* concentrations in terms of recreational water quality. The single-sample bathing-water standard is used in this report to evaluate recreational water quality on any given day and cannot be exceeded in more than 10 percent of the samples collected in a 30-day period.

Site selection and sampling frequency

Water and lake-bottom sediment samples were collected and ancillary environmental data were compiled during field studies at three public bathing beaches in the Cleveland, Ohio, metropolitan area (fig. 1): Edgewater Park, Villa Angela, and Sims Park.

Edgewater Park (Edgewater), operated by the Ohio Department of Natural Resources, is midway between Rocky River and Cuyahoga River. The East Beach of Edgewater includes 900 ft of guarded beach that is used heavily during the recreational season, and the West Beach is unguarded and used primarily by boaters (fig. 2). Lake currents are generally west to east, so that two metered outfalls (composed of

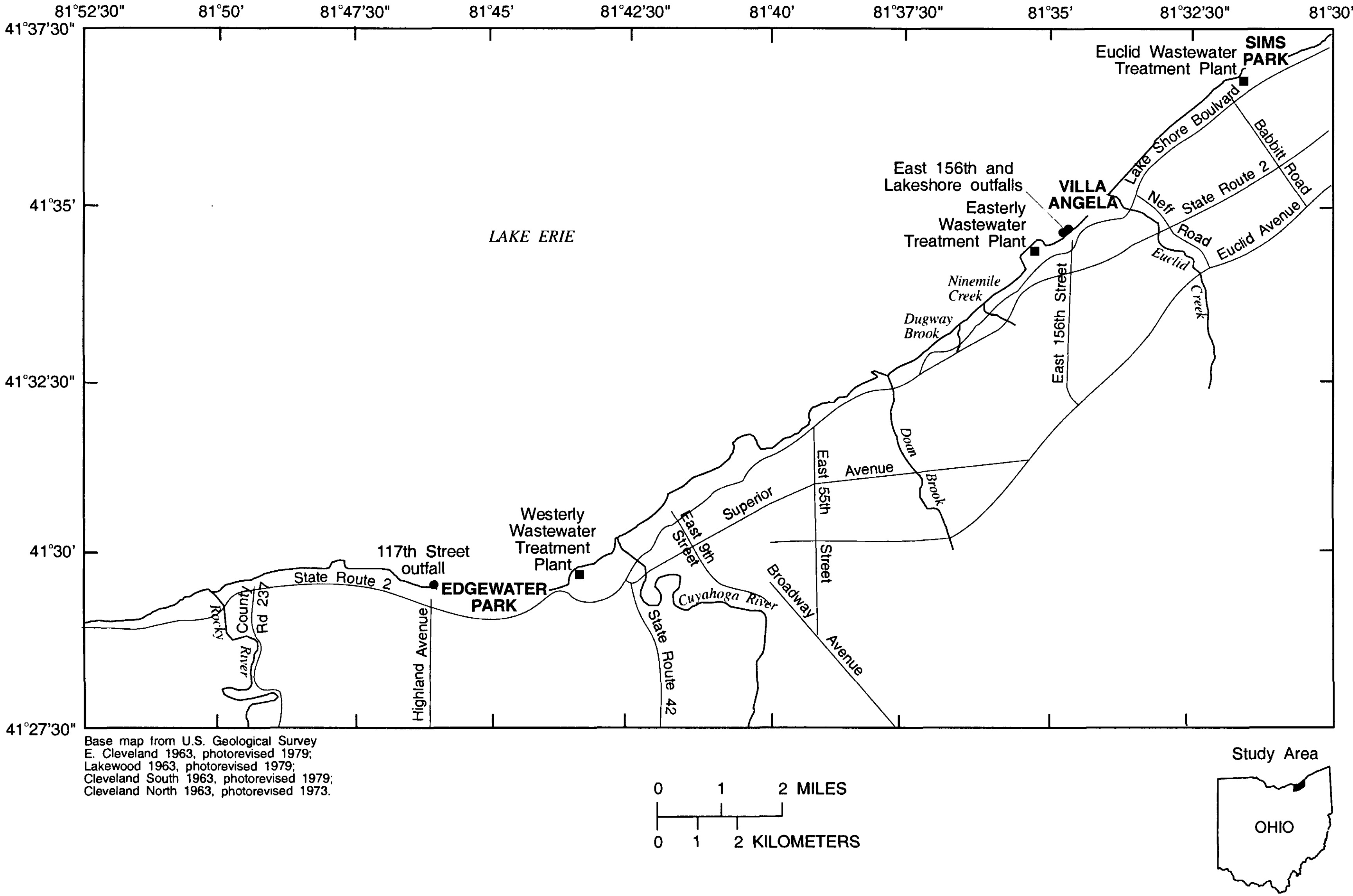

Figure 1. Locations of beach study sites—Edgewater Park, Villa Angela, and Sims Park—in the Cleveland, Ohio, metropolitan area, 1997.

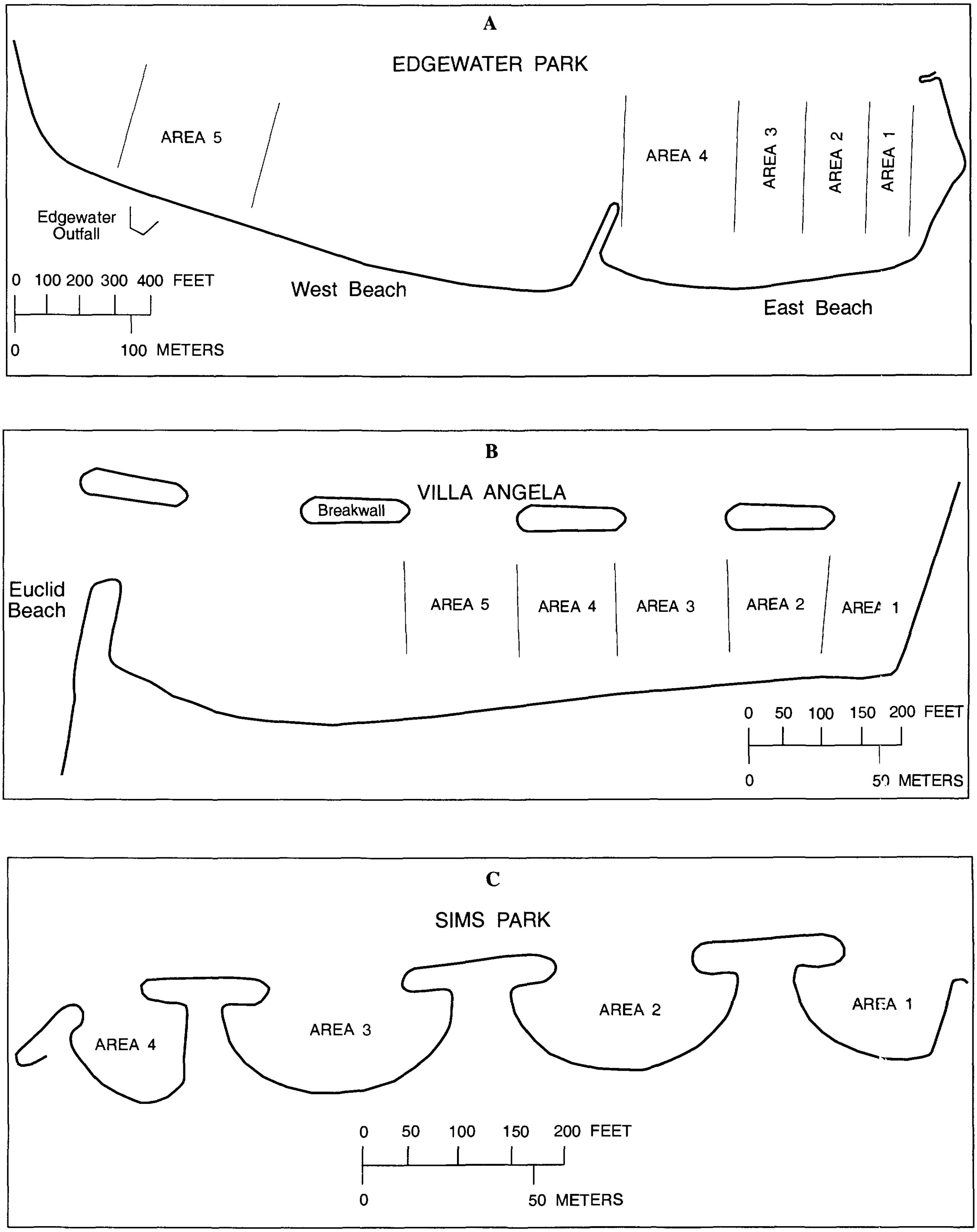

Figure 2. Sampling areas at (A) Edgewater Park and (B) Villa Angela, Cleveland, Ohio,and (C) Sims Park, Euclid, Ohio.

storm-water runoff and CSO's) have the potential to affect water quality at the East and West Beaches. The Edgewater outfall discharges into the West Beach (fig. 2) and the 117th Street outfall discharges into Lake Erie at Highland Avenue (fig. 1). The Westerly Wastewater Treatment Plant is one mile east of Edgewater Park, and the discharge pipe is several thousand feet from the shore into the open lake. Edgewater was divided into five sampling areas numbered from east to west (fig. 2) for data collection. The East Beach was divided into Areas 1 through 4, which were demarcated by metal rods installed to support lifeguard stations. Area 5 was located at the western edge of the West Beach near the Edgewater outfall.

Villa Angela is a 900-ft bathing beach constructed in 1995 and operated by the Ohio Department of Natural Resources. It is east of Euclid Beach (fig. 2), a popular bathing beach that was not investigated in this study because of the presence of large boulders in the swimming area and a lack of sediment particles smaller than gravel. The Easterly Wastewater Treatment Plant is 1.5 mi west of Villa Angela. Doan Brook, Dugway Brook, and Ninemile Creek are within 4.5 mi to the west of Villa Angela, and Euclid Creek is directly east of the beach (fig. 1). All of these tributaries receive inputs from storm-water runoff and CSO's. The East 156th and Lakeshore outfalls both discharge into Lake Erie near 156th Street and may also affect the water quality at Villa Angela. Villa Angela was divided into five sampling areas numbered from east to west for data collection (fig. 2). The areas were divided on the basis of the locations of breakwalls built to stabilize the beach area. Areas 2 and 4 were behind breakwalls and areas 1, 3, and 5 were open to the lake.

Sims Park (Sims) is a lakefront recreational area operated by the city of Euclid (fig. 1). Although the beach area at Sims is not a designated bathing beach, sunbathers and swimmers have easy access to it and frequently use it. The City of Euclid Wastewater Treatment Plant is directly west of Sims. The beach area at Sims was divided into four sampling areas numbered from east to west (fig. 2). Each area is an isolated cove separated from adjacent areas by sand and rocks.

During field studies, water and lake-bottom sediments were collected at beach study sites once every morning between 6:30 and 9:30 a.m. On 10 selected days, water samples were also collected after 1 p.m., when swimmers were present at the beaches. Water samples were collected from two randomly selected areas each at Edgewater and Villa Angela, and from the four areas at Sims (due to the separate nature of the coves). Therefore, a total of eight water samples were collected every morning or afternoon. Lake-bottom sediments were collected from each of the five areas at Edgewater and Villa Angela and the four areas at Sims. A total of 14 sediment samples were collected every morning.

A sampling location was identified in an area by randomly selecting a water depth using computer-generated integers from 2 to 6 ft. The field crew waded to a point in the area at the selected water depth and collected the sample. At each sampling point, one water sample and (or) three jars of lake-bottom sediments were collected. For the morning or afternoon sampling at Edgewater and Villa Angela, the mean concentrations of *E. coli* in water and sediment were determined by taking the average concentration of the four samples collected. Because of the configuration of the bathing area at Sims, each area was treated as a separate beach in determination of *E. coli* concentrations, and mean concentrations for all four areas at Sims were not determined.

Collection of water and sediment samples

Water samples were collected using a grab sampling technique in a manner that minimized contamination of the sterile collection containers (Myers and Sylvester, 1997). After wading or swimming to the randomly selected depth in each area, a sterile 1-L polypropylene bottle was opened and plunged downward, about 18 in. below the water surface. The bottle was filled, allowing about 2 in. of headspace, and the lid replaced.

Lake-bottom sediments were collected by scooping the lake bottom with an autoclaved wide-mouth 250-mL polypropylene jar. To minimize contamination by the overlying water, the jars remained covered until they touched the lake bottom and the sediment was collected. This was done using a specially designed sediment sampler or by diving. The sediment sampler was a polyvinyl chloride barrel secured on the end of a 6-ft metal sampling rod. The sample jar was placed in the barrel, and the opening of the jar and barrel were covered securely by a plastic lid with a rubber gasket. The sampler was lowered through the water column and upon touching the lake bottom, the operator raised the lid with a spring-loaded pulley. The operator would then scoop the bottom sediments into the 250-mL jar to obtain a sample and close the lid before raising the sampler to the surface. This method proved difficult to use, especially when waves were high. Alternatively, sediment samples were collected by diving and using the same principle as the sediment sampler. The diver secured the lid on the sampling jar, opened the lid upon reaching the lake bottom, and scooped the bottom sediments to obtain a sample. As with the sediment sampler, the lid of the jar was closed before the diver surfaced. Because of the spatial heterogeneity of bacteria concentrations in sediment, three sediment samples were collected from each area at the same depth and composited.

Water and lake-bottom sediment samples were placed on ice and transported to the laboratory for processing within 6 hours of sample collection.

Analysis of water and sediment samples

At the time of sample collection, a four-parameter water-quality meter was used to make field measurements of specific conductance, pH, temperature, and concentrations of dissolved oxygen. The meter was lowered to a point about 18 in. below the water's surface at a water depth of 3 ft in the middle area of each beach. These measurements are reported in Shindel and others (1998), and only the temperature data were used in the data analysis for this report. Water and sediment samples were processed by USGS employees at the Cuyahoga County Sanitary Engineers Laboratory in Valley View, Ohio. The constituents and methods of analysis, along with analyzing laboratory, frequency of analysis, and minimum detection limits are listed in table 1.

All water samples were analyzed for turbidity and *E. coli* concentrations within 6 hours of sample collection; 60 percent of water samples were analyzed for suspended-sediment concentrations. After processing water samples for *E. coli*, turbidity was measured by use of a Hach Model 2100P portable turbidimeter (Hach Company, Loveland, Colo.). The remaining water sample was carefully poured into a disposable polypropylene bottle and shipped to the USGS Iowa District Sediment Laboratory, Iowa City, Iowa, for determination of suspended-sediment concentration. Suspended-sediment concentrations were determined by use of the filtration method described in Guy (1969, p. 11-13)

Table 1. Constituents determined on water and lake-bottom sediment samples collected during field studies from May through September 1997 at three Lake Erie beaches

[mL, milliliters; g_{DW}, gram-dry weight of sediment; mg/L, milligrams per liter; NTU, Nephelometric Turbidity Unit; g/kg, grams per kilogram; USGS, U.S. Geological Survey; USEPA, U.S. Environmental Protection Agency; NEORSD, Northeast Ohio Regional Sewer District]

Constituent or determination	Analyzing agency	Frequency of analysis	Method and (reference)	Detection limit
Escherichia coli in water	USGS, Ohio District	Every sample	USEPA 1103.1 (USEPA, 1985)	1 colony/100 mL
Escherichia coli in sediment	USGS, Ohio District	Every sample	Modified from USEPA 1103.1 (USEPA, 1985)	1 colony/ g_{DW}
Suspended-sediment concentration in water	USGS, Iowa District	Every other day or when environmental conditions changed	(Guy, 1969, p. 11-13).	1.0 mg/L
Turbidity in water	USGS, Ohio District	Every sample	Hach Company, Loveland, Colorado	0.01 NTU
Total organic carbon of sediment	NEORSD	1 sample/study[a] or when environmental conditions changed	USEPA, 9060A (USEPA, 1986)	0.1 g/kg
Percent dry weight of sediment	USGS, Ohio District	Every sample	(American Society of Agronomy, 1982, p. 790-791)	Not applicable
Particle size of sediment	USGS, Ohio District	1 sample/study[a] or when environmental conditions change	(Guy, 1969, p. 47-51)	Not applicable

[a]Eight studies were done, ranging from 3 to 14 days for each study.

Water samples were analyzed for concentrations of *E. coli* by use of the mTEC agar membrane-filtration (MF) method (U.S. Environmental ProtectionAgency, 1985). In this method, plates were incubated on mTEC agar for 2 hours at 35°C and then for 20-22 hours at 44.5°C. After incubation, the membranes containing yellow colonies were placed in a urea broth for 15 to 20 minutes. The colonies remaining yellow, indicating a negative test for the enzyme, urease, were counted as *E. coli*. Concentrations of *E. coli* were calculated as described in Myers and Sylvester (1997) and expressed as colonies per 100 milliliters (col/100 mL).

Unlike water, standard methods for enumeration of fecal-indicator bacteria in sediments are not well established. Several treatments for the separation of bacteria from Lake Erie bottom sediments were tested for use in field studies. After a series of experiments, described in Appendix A, a protocol for determination of *E. coli* concentrations in sediments was established for field studies. Fifty grams of sediment were 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 (U.S. Environmental Protection Agency, 1985). The bottle containing the sediment/buffer mixture was placed on a wrist-action shaker for 45 minutes. After shaking, the bottle was allowed to settle for 30 seconds, the liquid phase was poured into a second sterile bottle, and the remaining sediment was discarded. Concentrations of *E. coli* were determined in the liquid phase by use of the mTEC agar method, described previously for analysis of water samples.

Concentrations of *E. coli* in sediment were reported as colonies per gram of dry weight sediment (col/g_{dw}). Several values were determined to complete this calculation. Percent dry weights were determined by placing about 25 g of composited sediment in a tared metal dish. After drying for 24 hours at 105°C, the sediment weight was recorded and percent dry weights were calculated. A sediment dilution factor was also determined. Because 20 g of dry or wet sediment displaces approximately 10 mL of buffer, the total volume of the sediment/buffer mixture was 210 mL. The dilution factor of the sediment sample was therefore, 10.5 mL/g (210 mL / 20 g).

The equation used to determine concentrations of *E. coli* in lake-bottom sediment was modified from American Society of Agronomy, Inc., and Soil Science Society of America, Inc. (1982, p. 790-791), as follows:

$$col/gdw = ((pc \times 10.5) \div (volume \times dw)),$$

where col/g_{dw} is colonies/gram of (dry weight) sediment, pc is plate count, 10.5 is sediment dilution factor, volume is volume of sample plated, and dw is percent dry weight of sediment.

Some sediment samples were analyzed for total organic carbon and particle-size distribution. For these analyses, six replicate jars (instead of three jars) were collected at each location to ensure a sufficient amount of sediment for analysis. For total organic carbon analysis, equal amounts of sample from each replicate jar were composited to obtain more than 500 mL of sediment. The composited sediment was then processed through a 2 mm stainless steel sieve using lake water as a rinse. The less-than-2-mm fraction was transported in a cooler to the Northeast Ohio Regional Sewer District laboratory, Cuyahoga Heights, Ohio, for determination of total organic carbon. The remainder of the sediment in the six jars was composited and particle-size analysis was done at the USGS Ohio District Laboratory, Columbus, Ohio.

Collection and compilation of ancillary information

Ancillary environmental data were collected by the USGS or compiled from a variety of sources. Wind speed and direction were measured at the National Oceanic and Atmospheric Administration (NOAA) Surface Airways Station at Burke Lakefront Airport, Cleveland, Ohio, and archived by the Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration Coast Watch Program (George Leshkevich, NOAA, written commun., 1997). The number of swimmers and wave heights were estimated by USGS personnel at the time of sample collection. Information on flow and duration of wastewater treatment plant overflows and metered outfalls was obtained from the Northeast Ohio Regional Sewer District (Eva Roller, Northeast Ohio Regional Sewer District, written commun., 1997). This included sewage-treatment plant overflows at the Easterly and Westerly Wastewater Treatment Plants and overflows from the Edgewater, West 117th Street, East 156th, and Lakeshore metered outfalls. Information on flow and duration of effluent that was diverted to the City of Euclid's wet-weather treatment facility was obtained from the City of Euclid (Robert Gall, City of Euclid, written commun., 1997). Daily rainfall amounts were measured by NEORSD at the Easterly and Westerly Wastewater Treatment Plants (Ruth Crowl, Northeast Ohio Regional Sewer District, written commun., 1997). Westerly rainfall data were used at Edgewater Park, and Easterly rainfall data were used at Villa Angela and Sims Park.

Statistical methods

Analysis of variance (ANOVA) was used to compare more than two groups of data. Histograms and the Shapiro-Wilk test (Wilk and Shapiro, 1968) were used to determine the normality of the distribution of $\log_{10}$-transformed bacteria concentrations. For small data sets or if the data were still not normally distributed after a $\log_{10}$-transformation, the nonparametric rank transform test was done instead of the parametric ANOVA. In the rank transform test, all data are combined and ranked from lowest to highest value, and an ANOVA is computed on the ranks. The parametric and nonparametric ANOVA determines whether the mean or median, respectively, differs between groups. The null hypothesis is that each group mean or median is the same; the alternative hypothesis is that at least one is different. If ANOVA 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). The level of significance for ANOVA was set at α=0.05, unless specified otherwise.

Correlation coefficients were calculated to determine the strength of association between two continuous variables. Correlation coefficients are a measure of the strength of the monotonic relation—y generally increases or decreases as x increases. Pearson's r is a correlation coefficient that measures the linear association between two variables and is computed using means and standard deviations directly from the observed data. Spearman's rho, another correlation coefficient, measures the monotonic relation (nonlinear or linear) between two variables and is computed on the ranks of the data. If the data lie exactly along a straight line with positive slope, then the correlation coefficient is equal to one (Helsel and Hirsch, 1992, p. 209-218). The more the correlation coefficient deviates from 1 or -1 and approaches zero, the weaker the relation. Correlation coefficients were considered statistically significant if the p-value was less than 0.05.

Linear regression analysis was used to predict *E. coli* concentrations in water from one or more explanatory variables. Simple linear regression (SLR) was used to describe the relation between *E. coli* concentrations and one explanatory variable. Multiple linear regression (MLR) is the extension of SLR to the case of multiple explanatory variables (Helsel and Hirsch, 1992, p. 295).

For MLR, models were chosen among all possible variable combinations to maximize the coefficient of determination (R^2) and minimize the Mallows' Cp statistic (Mallows, 1973). The R^2 of the model is the fraction of the variation in the dependent variable (*E. coli* concentrations) that can be explained by a combination of explanatory variables. The Cp statistic is a measure of the standard error and the bias introduced by not including important variables in a model. The Cp statistic is designed to achieve a workable compromise between the desire to explain as much variance in *E. coli* concentrations as possible (minimizing bias) by including all relevant variables and to minimize the standard error by keeping the number of variables small (Helsel and Hirsch, 1992, p. 312-313). When several models had nearly equal R^2 and Cp values, a model was chosen on the basis of reduced multicollinearity (where at least one explanatory variable is related to one or more other explanatory variables) and cost of data collection.

To evaluate how well *E. coli* concentrations can be predicted from MLR equations, prediction intervals were determined from a randomly chosen set of values for explanatory variables. Prediction intervals were used to estimate the range of predicted *E. coli* concentrations that result given a particular level of uncertainty. Given a single set of explanatory variables, a 90-percent prediction interval represents the range of values a single *E. coli* concentration is expected to assume that includes the true *E. coli* concentration 90 percent of the time.

Because prediction intervals are used to predict a single *E. coli* concentration, they are generally too wide to offer a reasonable prediction of recreational water quality. Alternatively, the probability of exceeding a threshold value—in this case, the single-sample bathing-water standard of 235 col/100 mL—was used to assess usefulness of the MLR model. The error associated with a predicted *E. coli* concentration for a given set of explanatory variables was used to estimate the probability that the true *E. coli* concentration would exceed 235 col/100 mL.

Another way to examine a model's ability to accurately predict *E. coli* concentrations is to use classification tables. Classification tables compare the proportions of correct and incorrect predictions. A classification table is a 2 by 2 frequency table of observed and predicted events and nonevent responses. An event was defined as a sample with an *E. coli* concentration equal to or exceeding the single-sample bathing-water standard; a nonevent was defined as a sample with an *E. coli* concentration less than the single-sample bathing-water standard. Sensitivity was the proportion of event responses that were predicted correctly as events. Specificity was the proportion of nonevent responses that were correctly predicted as nonevents. The false positive rate was the proportion of predicted events that were observed as nonevents, and the false negative rate was the proportion of predicted nonevents that were observed as events (SAS Institute, 1990).

One way the chosen MLR model was tested was by developing and comparing two classification tables. The first classification table was generated from all of the data used to develop the MLR model without removing any portion of the data set. For the second classification table, the data were ranked by wave height and were randomly divided into three data sets. The data were sorted by wave height to ensure that each data set contained a reasonable representation of the overall data. The MLR model was then run three times with two data sets each, omitting a different data set each time. A classification table was obtained by summing the results of the three runs. The model was considered to be reasonable if the second classification table was similar to the classification table produced from all of the data used to develop the MLR model (the first classification table).

Quality-assurance and quality-control practices

Quality-assurance and quality-control (QA/QC) practices were followed for all phases of data collection, analysis, and data validation. Field and laboratory protocols were written and distributed to ensure that procedures were performed according to established methods and in a uniform manner by all personnel.

The spatial heterogeneity of *E. coli* concentrations in lake-bottom sediments and water was investigated before field studies. The results of these investigations were considered while designing sampling protocols and identifying sampling points in field studies. These protocols are described in a previous section of this report.

Spatial heterogeneity of *E. coli* concentrations in water were found to be less than in sediment at Edgewater and Villa Angela, but not at Sims. For sediments collected at the three beaches, 10-fold to 15-fold differences in *E. coli* concentrations were often found between samples collected from different areas at the same beach, from below different water depths in the same area, and from the same area at the same water depth. At Edgewater and Villa Angela, 2-fold differences in *E. coli* concentrations were found in water samples collected from different areas at the same beach; most differences were considerably less than 2-fold. At Sims, however, the differences in *E. coli* concentrations between water samples collected from different areas were considerably greater than at the other two beaches; for example, in one morning sampling, a 20-fold difference was found.

Quality-control samples were collected to measure sampling variability and analytical bias and to ensure that data satisfied the project objectives. Variability is the degree of variation in independent measurements as the result of repeated application of the measurement process under specified conditions. Bias is a systematic error inherent in a method or caused by some artifact or property of the measurement system. Bias may be either positive (from contamination) or negative (from loss, degradation, or poor method recovery) (Schertz and others, 1998). The following quality-control samples were collected:

1. Turbidity—turbidity measurements were taken in duplicate by measuring two aliquots of water from the sample bottle. Measurements that did not agree within 10 percent were repeated.
2. Suspended-sediment concentrations—approximately 5 percent of the suspended-sediment samples collected were analyzed in duplicate.
3. Percent dry weights—approximately 10 percent of percent-dry-weight determinations were done in duplicate.
4. *E. coli* concentrations—eleven percent of the water samples and 7 percent of the sediment samples were collected in a nested replicate design. Two water samples were collected in two different bottles or two sediment samples were mixed with saline buffer in two different bottles (replicate sample bottles). Each water or sediment replicate bottle was then plated in duplicate (split samples) for concentrations of *E. coli*. Results of quality-control replicate samples were used to

determine sampling and analytical variability and are described in Appendix B.

In the laboratory, equipment and supplies were regularly checked to ensure proper performance. The incubators were monitored throughout experiments and field studies to ensure that temperatures were 35°C ± 0.5°C or 44.5°C ± 0.2°C. The sterility of the buffer water and media were tested by use of blanks—aliquots of buffer water filtered before each sample. Sample results were rejected if incubator temperatures were outside acceptable ranges or quality-control testing showed contaminated blanks. The autoclave operating temperature and pressure were checked for each run, and heat-indicating tape was used to identify supplies that had been sterilized. Other standard laboratory practices—cleanliness, safety practices, procedures for media preparation, specifications for reagent water quality—were adopted by USGS employees as set forth by American Public Health Association and others (1995, section 9020) and Britton and Greeson (1989).

Factors affecting *Escherichia coli* concentrations in water and lake-bottom sediments

Summary statistics of *E. coli* concentrations in water and lake-bottom sediments collected during morning sampling events on 41 days throughout the 1997 recreational season are shown in table 2. Concentrations of *E. coli* were lowest at Edgewater among the three beaches, and bacterial water quality was generally good at Edgewater, exceeding the single-sample bathing water standard on only 7 of the 41 days sampled. Concentrations of *E. coli* in the four areas of Sims were higher and more variable than at either

Table 2. Concentrations of *Escherichia coli* in water and lake-bottom sediments collected at three Lake Erie beaches on 41 selected mornings from May through September 1997

[For Edgewater Park and Villa Angela, the daily concentration was determined by calculating the mean of two water samples or five lake-bottom sediment samples; NA, not applicable]

Beach	Median	Minimum	Maximum	Number of days bathing-water standard[a] was exceeded
		Water[b]		
Edgewater	86	9	830	7
Villa Angela	150	13	8,100	17
Sims 1	400	20	16,000	23
Sims 2	450	21	19,000	24
Sims 3	390	10	36,000	27
Sims 4	400	13	29,000	27
		Sediment[c]		
Edgewater	7	1	38	NA
Villa Angela	35	5	170	NA
Sims 1	150	2	8,000	NA
Sims 2	130	4	2,600	NA
Sims 3	72	2	7,200	NA
Sims 4	34	4	750	NA

[a]Number of days the concentration of *Escherichia coli* in water exceeded the single-sample bathing-water standard of 235 colonies per 100 milliliters, out of 41 days sampled.

[b]Colonies per 100 milliliters.

[c]Colonies per gram dry weight of sediment.

Edgewater or Villa Angela, except for concentrations of *E. coli* in sediments at Sims area 4.

To characterize the sediment quality at each beach and aid in data interpretation, particle-size analysis and total organic carbon concentrations were determined in lake-bottom sediments collected on selected sampling days (table 3). Finer sediments were found at Edgewater than at Villa Angela or Sims; at Edgewater, most of the sediments were classified as medium to fine sands (63 to 250 μm). The sediments collected at the beaches at Villa Angela and Sims varied considerably in particle size; most were classified as medium sands to gravels (250 to greater than 1,000 μm). Total organic carbon concentrations in sediment ranged from 0.5 to 3.2 mg/kg; not enough data were collected to examine total organic carbon concentration as a factor affecting *E. coli* concentrations in sediment.

Data collected each morning at the beaches at Edgewater, Villa Angela, and Sims are shown in figures 3 through 5. Concentrations of *E. coli* in water and lake-bottom sediments and the physical disturbances that were expected to affect those concentrations—rainfall, wastewater-treatment plant overflows, and wave heights—are included.

Table 3. Particle-size analysis and total organic carbon concentrations in lake-bottom sediments collected at three Lake Erie beaches, 1997

[ND, not determined]

Dates	Sampling area	Sampling depth (feet)	Percent finer than (micrometers)			Total organic carbon (milligrams per kilogram)
			1,000	250	63	
Edgewater Park						
6-18	4	4	98.8	88.3	0.1	2.1
6-20	3	5	99.6	96.6	0.5	ND
6-20	4	2	99.2	89.0	0.2	ND
7-09	3	6	99.5	96.4	0.5	0.8
8-11	2	3	99.6	98.3	1.0	3.2
8-18	3	4	99.8	98.3	0.5	0.8
Villa Angela						
6-18	3	4	55.1	0.7	0.1	1.4
6-20	3	6	99.9	59.7	0.3	ND
6-20	5	2	47.6	1.2	0.2	ND
7-9	5	6	62.6	10.2	0.1	1.4
8-12	4	2	29.0	0.2	0.1	2.4
8-18	4	5	99.5	10.0	0.2	0.6
Sims Park						
6-18	1	4	26.7	5.9	3.4	0.5
6-20	2	6	86.7	13.2	6.4	ND
6-20	3	2	77.5	4.3	0.1	ND
7-09	2	5	39.5	4.2	0.1	3.2
8-12	1	5	18.3	6.4	3.8	1.0
8-18	1	4	8.6	1.2	0.3	1.9

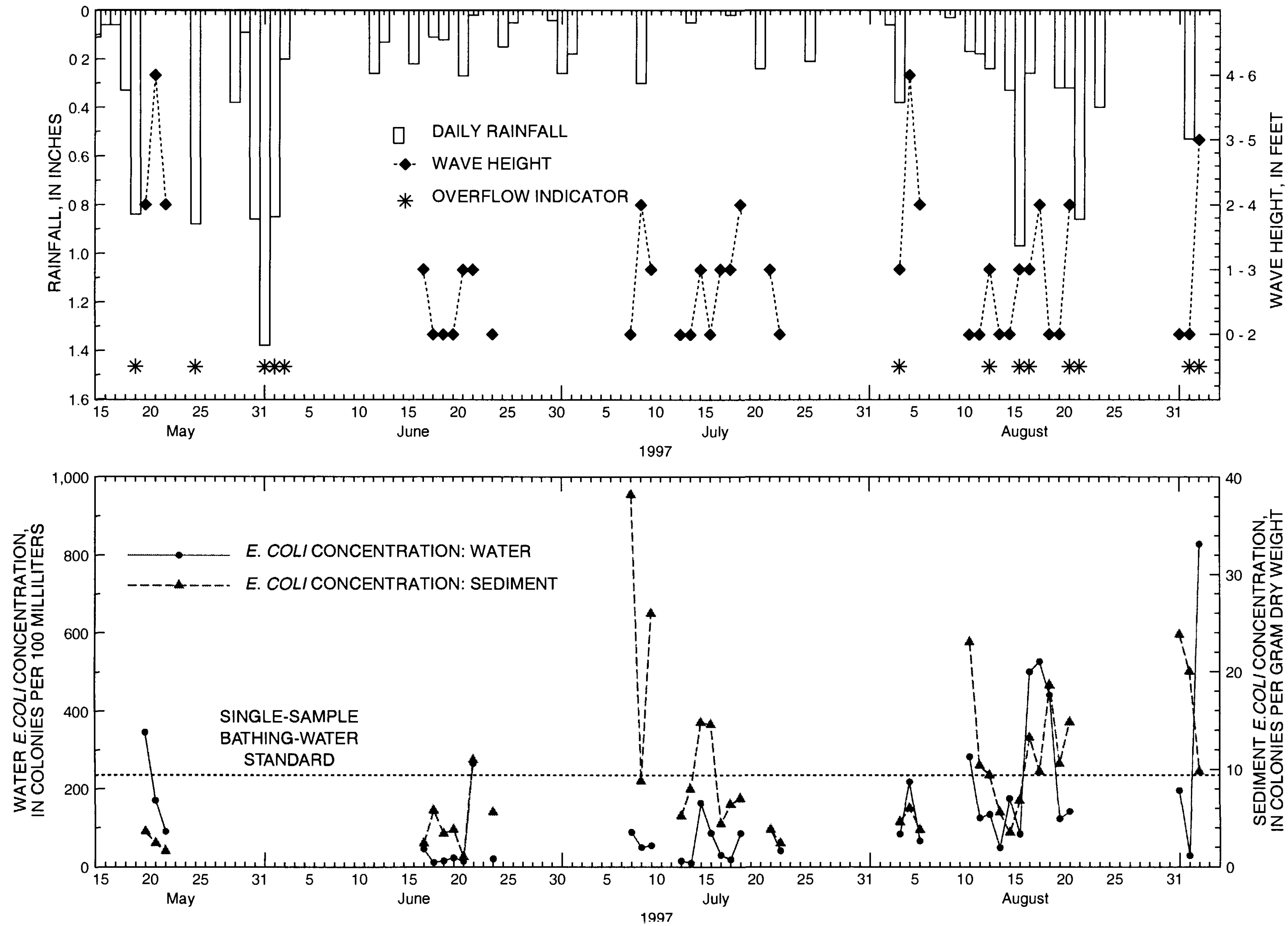

Figure 3. Mean concentration of *Escherichia coli (E. coli)* in water and lake-bottom sediments, rainfall amounts, overflow indicators, and wave heights for the recreational season of 1997 at Edgewater Park, Cleveland, Ohio.

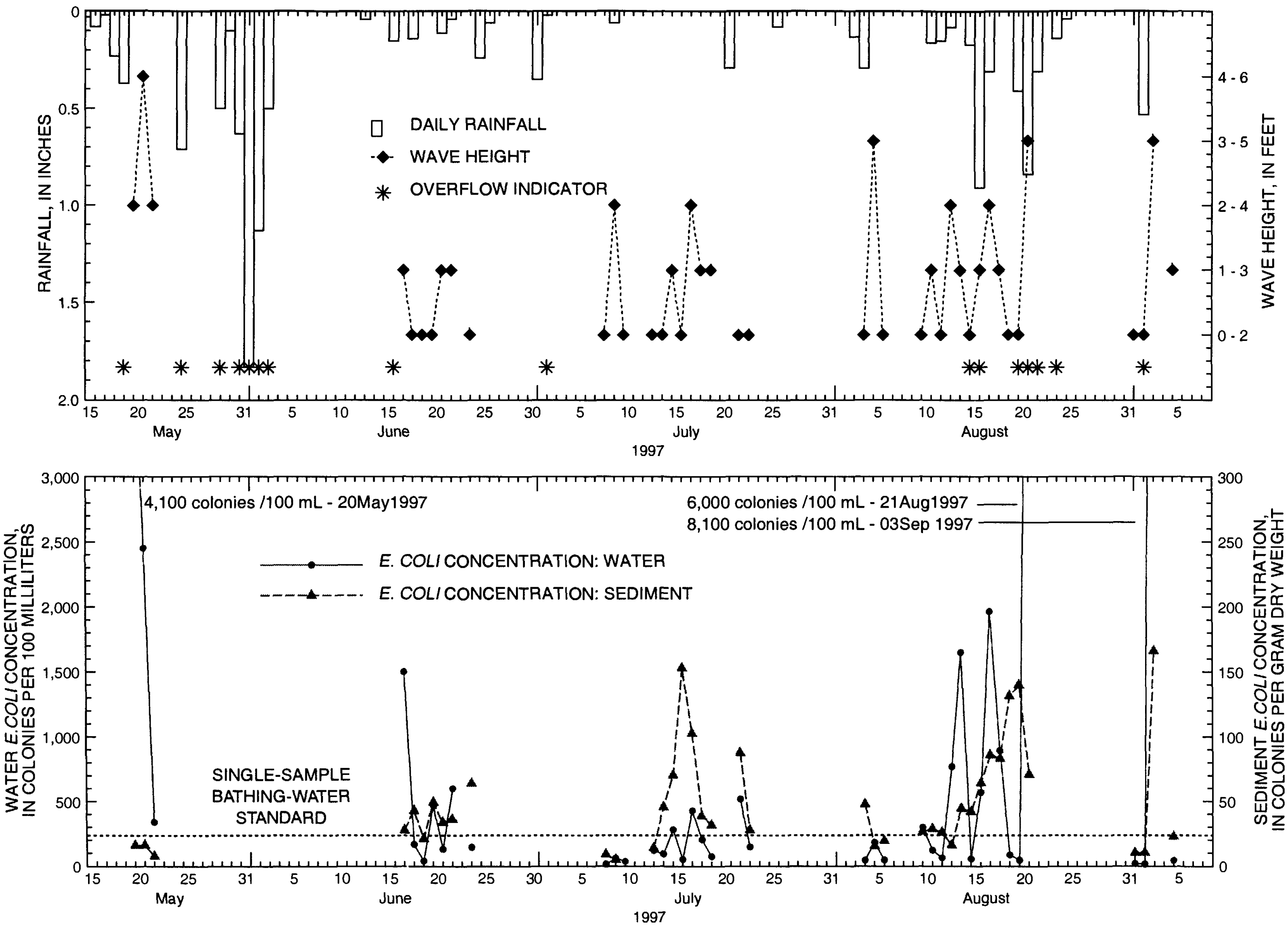

Figure 4. Mean concentration of *Escherichia coli* (*E.coli*) in water and lake-bottom sediments, rainfall amounts, overflow indicators, and wave heights for the recreational season of 1997 at Villa Angela, Cleveland, Ohio.

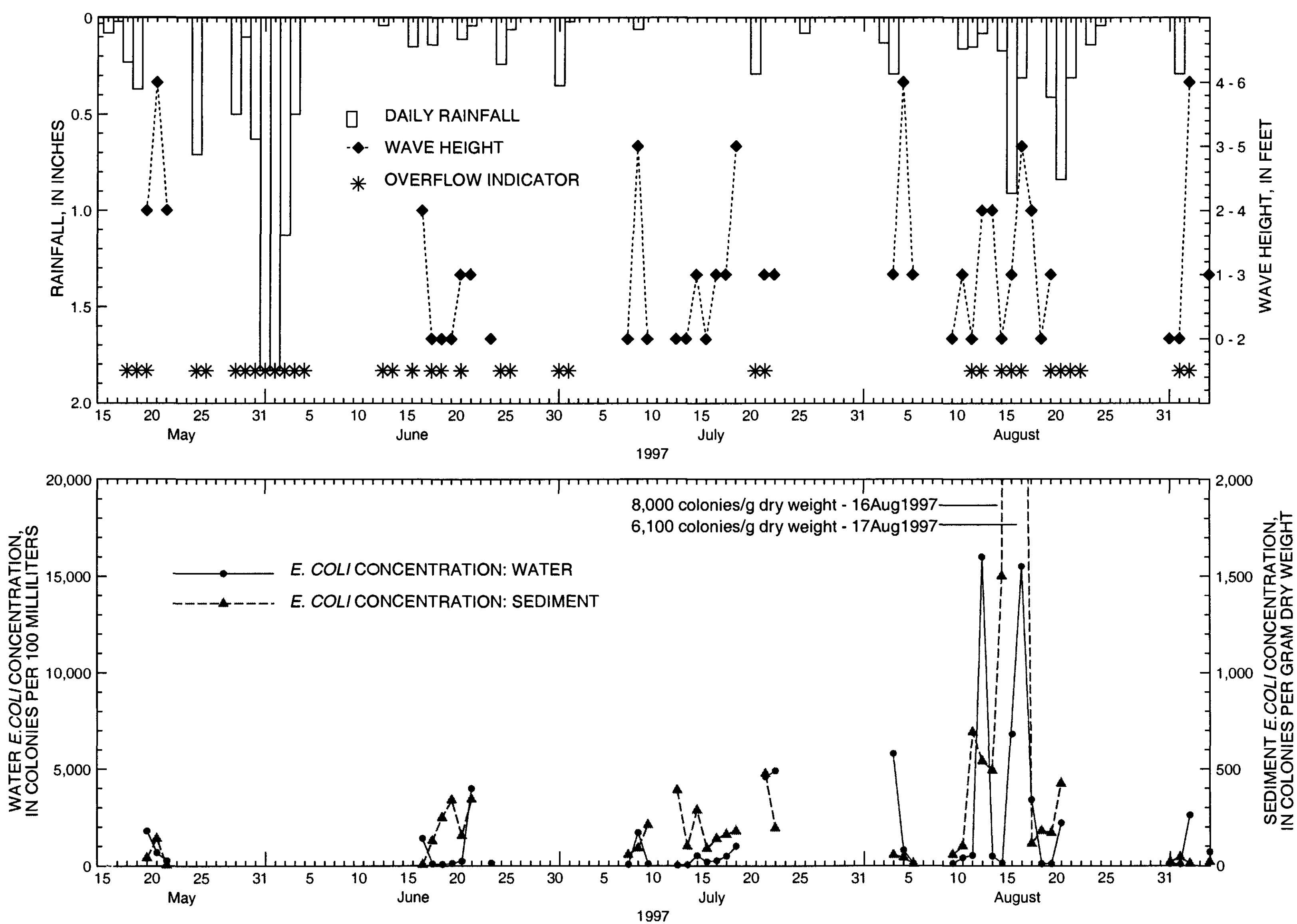

Figure 5a. Concentration of *Escherichia coli* (*E.coli*) in water and lake-bottom sediments, rainfall amounts, overflow indicators, and wave heights for the recreational season of 1997 at Sims Park, area 1, Euclid, Ohio.

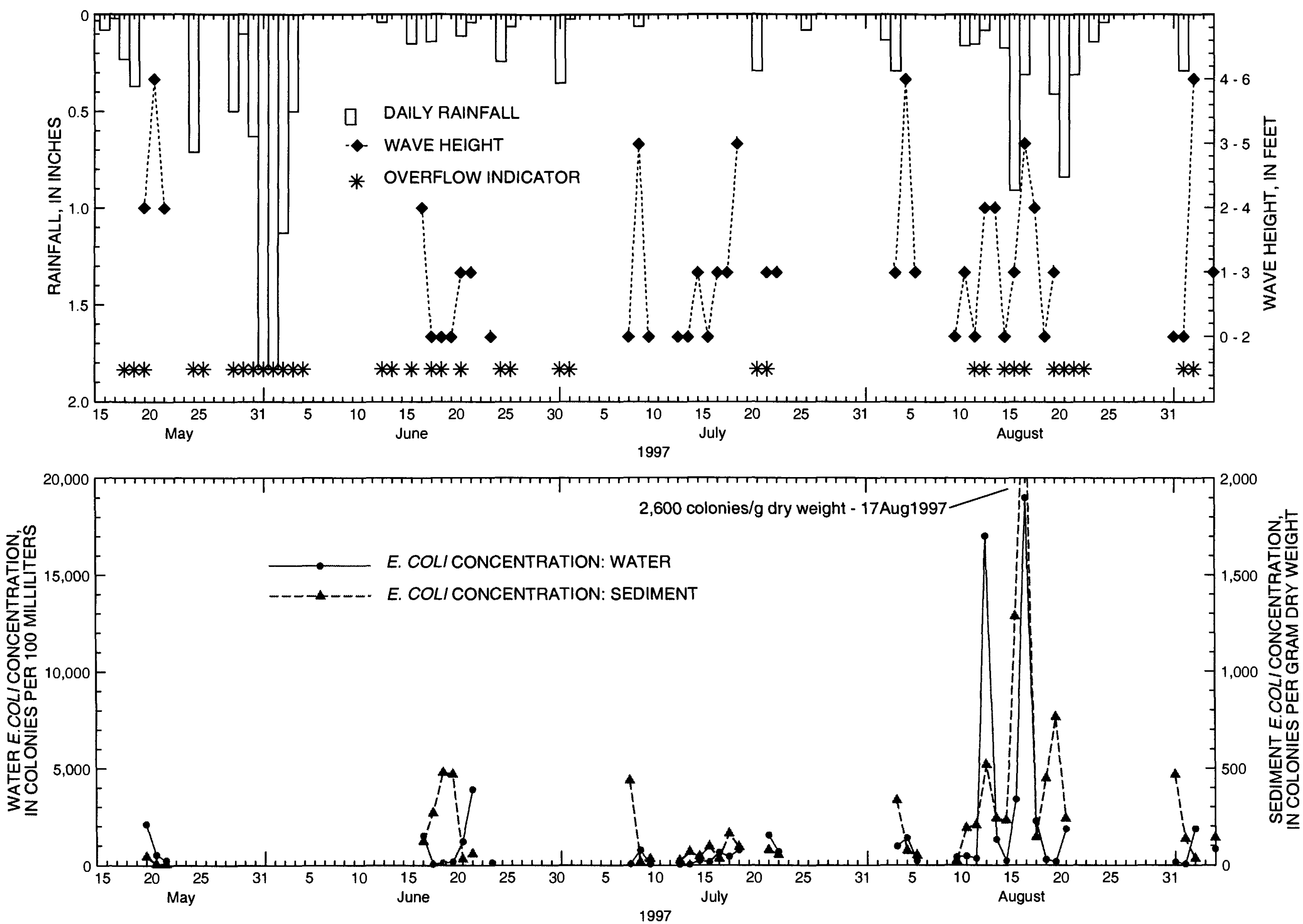

Figure 5b. Concentration of *Escherichia coli* (*E.coli*) in water and lake-bottom sediments, rainfall amounts, overflow indicators, and wave heights for the recreational season of 1997 at Sims Park, area 2, Euclid, Ohio.

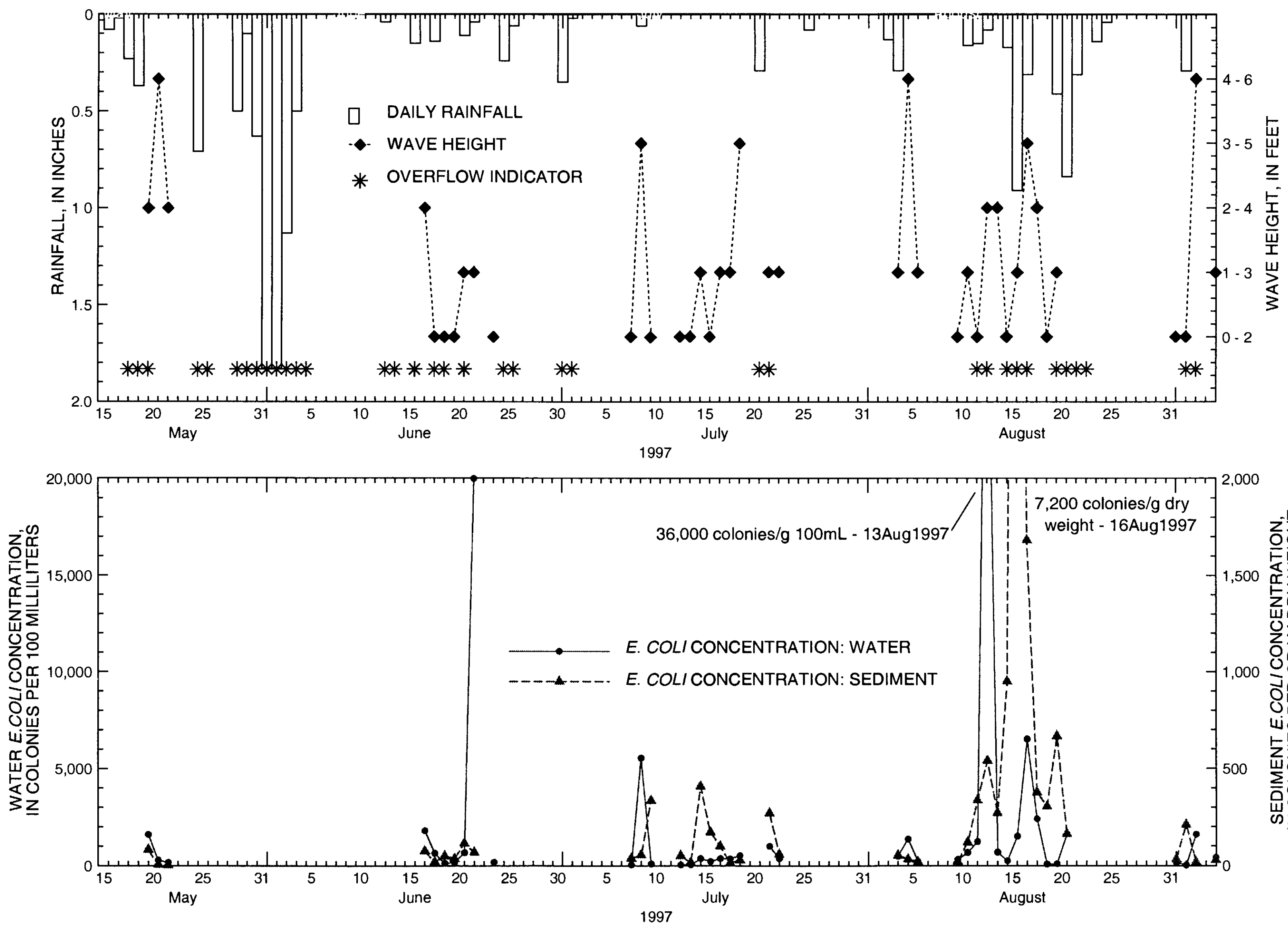

Figure 5c. Concentration of *Escherichia coli* (*E.coli*) in water and lake-bottom sediments, rainfall amounts, overflow indicators, and wave heights for the recreational season of 1997 at Sims Park, area 3, Euclid, Ohio.

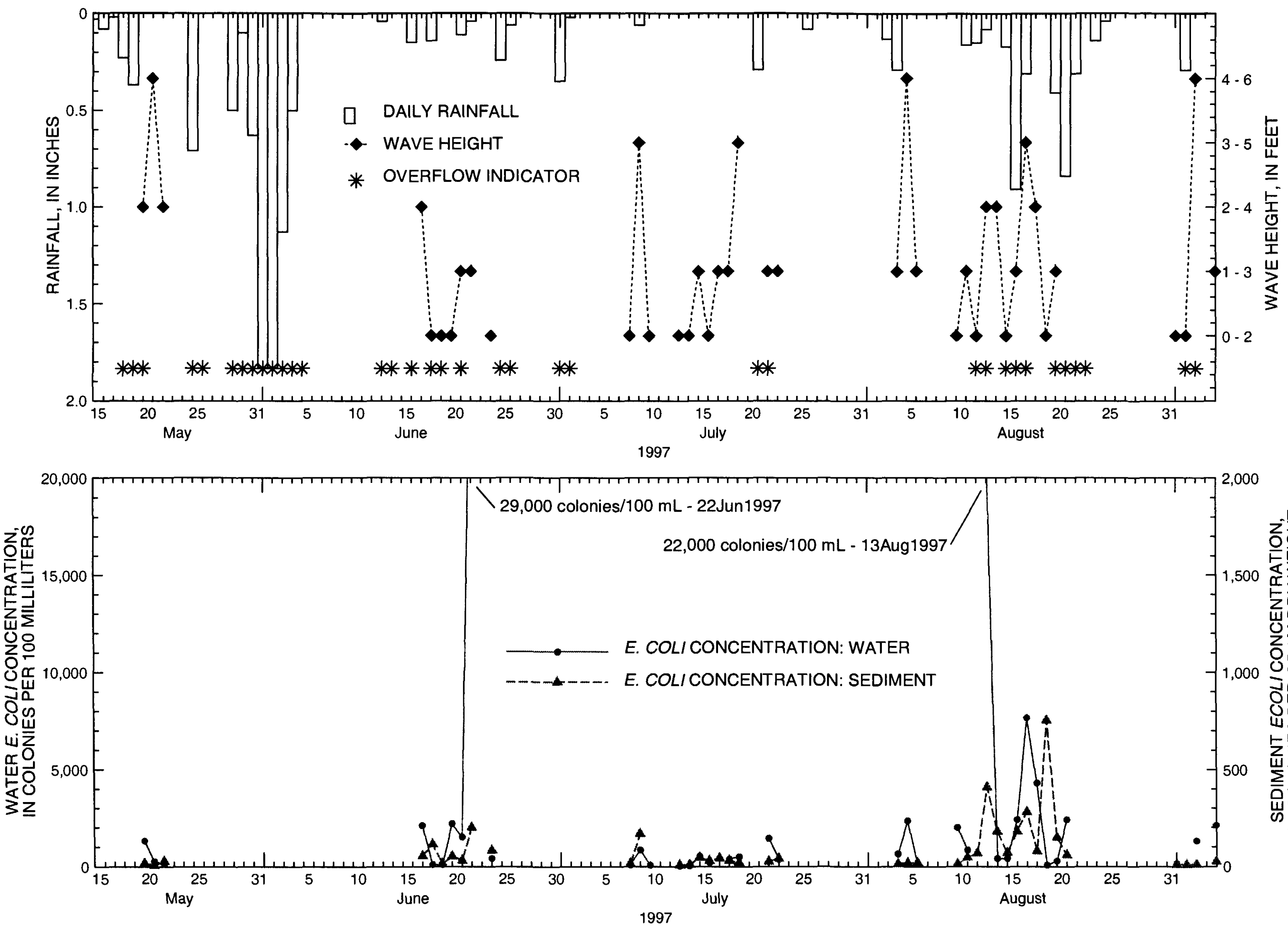

Figure 5d. Concentration of *Escherichia coli* (*E.coli*) in water and lake-bottom sediments, rainfall amounts, overflow indicators, and wave heights for the recreational season of 1997 at Sims Park, area 4, Euclid, Ohio.

These plots suggest some general patterns found at all three beaches (figs. 3-5). Peak *E. coli* concentrations in water samples occurred either during or immediately following rainfall or wastewater-treatment plant overflows, with a few exceptions. Higher concentrations of *E. coli* in water were also found when wave heights were 2 to 4 ft or higher; these higher wave heights usually occurred during or after rainfall. Sediment *E. coli* concentrations generally increased at the same time or after a peak in water *E. coli* concentrations.

Exceptions to these generalities can be found upon closer examination of the plots for each beach. These exceptions were examined to provide some insight into the extent of storage of *E. coli* in lake-bottom sediments during this sampling period. Periods of storage in sediments occur when sediment concentrations increase as water concentrations decrease or fluctuate.

At Edgewater, from August 15 through 21, sediment *E. coli* concentrations increased gradually while water concentrations fluctuated (fig. 3). During this period, there may have been short-term storage of *E. coli* in sediments; water and sediment concentrations were still somewhat elevated on September 1. Earlier that summer on July 8 and 10, sediment *E. coli* concentrations were the highest among the samples collected at Edgewater; this occurred when water *E. coli* concentrations remained low. A closer examination of the data revealed that sediment *E. coli* concentrations were considerably higher in samples collected at only one of the five areas of the beach on each of those two days, and concentrations were not higher at the same area on both days. These high unexplained *E. coli* concentrations in sediments may be due to temporary storage of *E. coli* in isolated locations.

At Villa Angela, *E. coli* concentrations in water were elevated on May 20 and 21, but concentrations in the sediment remained low (fig. 4). This sampling was early in the season, and storage of *E. coli* in sediments was not apparent. Storage of *E. coli* in sediments may have occurred from July 13 through 16 and again from August 13 through 20 at Villa Angela. In addition, from July 13 through 16, concentrations of *E. coli* in water generally increased when wave heights increased; this occurred in the absence of rainfall or overflows. This increase may have been due to resuspension of *E. coli* from lake-bottom sediments in the nearshore area or recirculation of *E. coli* from deeper sediments outside the bathing area, perhaps from a shift in the direction of lake currents. (The directions of lake currents were not investigated during this study). On July 16, *E. coli* concentrations in the sediments peaked due to extremely high concentrations found in individual samples from areas 1 and 2.

At Sims (figs. 5a-5d), temporary storage of *E. coli* in sediments may have occurred in some areas in June and July, but sediment concentrations followed the same patterns as water concentrations during August. On June 17-20 in areas 1 and 2, sediment *E. coli* concentrations increased while water concentrations and wave heights decreased (figs. 5a and 5b). From July 14 through 19, various degrees of *E. coli* storage in sediments may have occurred in areas 1, 2, and 3, but not in area 4, even though this period was dry and no wastewater-treatment-plant overflows occurred. Perhaps the elevated *E. coli* concentration in water at area 1 on July 19 was due to resuspension from lake-bottom sediments; wave heights were 3-5 ft on that day. On July 15, an unexplained peak *E. coli* concentration in sediment occurred in area 3 at Sims; similar peaks were found at the other two beaches, and may be due to temporary storage of bacteria in sediments at isolated locations.

Data collected at the three beaches indicate that short-term storage (less than one week) of *E. coli* in lake-bottom sediments may have occurred. Although there is no direct evidence for long-term storage of *E. coli,* concentrations in water were found to increase with increasing wave height. Unfortunately, the resuspension of *E. coli* from lake-bottom sediments by wave actions could not be adequately assessed because higher wave heights usually occurred during or after a rainfall or wastewater-treatment-plant overflow. At Edgewater, sediment particle sizes were the smallest among the three beaches, which makes more surface area available for bacterial attachment, settling, and resuspension. The bacterial water quality at Edgewater was generally good during the sampling period, so it was difficult to investigate thoroughly the storage and resuspension of *E. coli* in the sediments most likely to have attached bacteria.

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 factors and *E. coli* concentrations in water.

Pearson and Spearman's correlation coefficients were computed to assess the linear and monotonic relation, respectively, between *E. coli* concentrations and other continuous variables. Continuous variables describing environmental or water-quality factors were (1) turbidity, (2) suspended-sediment concentration, (3) 24-hour antecedent rainfall (rainfall), (4) weighted antecedent rainfall (weighted rainfall), (5) 24-hour antecedent wastewater-treatment plant overflow, (6) wind speed, (7) water temperature, and (8) a resuspension index. Rainfall was defined as the amount of rain in inches that fell in the 24-hour period preceding the 6:30 a.m. sampling. The weighted rainfall was computed from the rainfall amounts that occurred in the 72-hour period preceding the 6:30 a.m. sampling, with the most recent rainfall receiving the highest weight. Rainfall amounts for the 24-hour antecedent period were multiplied by a factor of three. Rainfall amounts in the greater than 24- to 48-hour antecedent period were multiplied by two, and rainfall amounts in the greater than 48- to 72-hour antecedent period were multiplied by one. The three weighted terms were then summed to provide weighted rainfall for the time of sampling. The overflow was defined as the amount of wastewater-treatment-plant overflow, in million gallons, that occurred on the previous day. The resuspension index was computed by multiplying the mean concentration of *E. coli* in sediment found on the previous day by the wave height on the current day.

Pearson's r and Spearman's rho correlation coefficients are shown in table 4 for all data, for data grouped by beach, and for data grouped by area at Sims. Statistically significant correlations were found for all relations between *E. coli* concentrations and turbidity, rainfall, and weighted rainfall. Statistically significant correlations were found for the relations between *E. coli* concentrations and wastewater-treatment-plant overflows, except for the linear relation (Pearson's r) at Villa Angela, and the resuspension index except for linear relation at Edgewater and two areas at Sims. Statistically significant correlations ranged from 0.341 to 0.708, and the highest correlation coefficient was found for *E. coli* concentrations and rainfall at Sims area 3. Correlation coefficients by area for Sims were not much different from those found for all Sims areas combined. Relations between *E. coli* concentrations and wind speed or water temperature were weak or not statistically significant.

Analysis of variance was used to assess the relations between categorical environmental variables (wind direction or wave height) and *E. coli* concentrations in water. For wind direction, data on $\log_{10}$ *E. coli* concentrations were placed into four groups based on wind direction in degrees clockwise from north (NE, 1-90°; SE, 91-180°; SW, 181-270°; and NW, 271-360°) (fig. 6). When winds were from the SE at the time of sample collection, significantly lower mean *E. coli* concentrations were found than when winds were from the NE or NW. However, no statistically significant differences were found in *E. coli* concentrations when wind directions were from the SW (when most samples were collected) than when winds were from the other three categories.

Patterns of waves on any body of water exposed to winds generally contain waves of different periods and amplitudes (U.S. Army Corps of Engineers, 1984). These different patterns of waves are known as a wave train. The minimum and maximum wave heights within each wave train were used to categorize wave height into the following categories: (1) a minimum of 0 to a maximum of 2 ft, (2) a minimum of 1 to a maximum of 3 ft, (3) a minimum of 2 to a maximum of 4 ft, (4) a minimum of 3 to a maximum of 5 ft, and (5) a minimum of 4 to a maximum of 6 ft.

Log_{10} *E. coli* concentrations were placed into five groups based on wave heights, and an analysis of variance was done to determine the relation between wave height and *E. coli* concentrations (fig. 7). With the exception of wave-height category 1, median *E. coli* concentrations for wave height categories exceeded the geometric mean bathing-water standard of 126 col/100 mL. Except for wave height category 5, median *E. coli* concentrations generally increased with increasing wave height. Statistically significant differences were found among some of the wave-height categories. Because categories 4 and 5 had relatively small data sets and were not statistically different, category 5 was combined with category 4. This new wave-height category 4 contained waves with a minimum height of 3 ft to a maximum height of 6 ft.

The effect of recreational use on *E. coli* concentrations in water was investigated by collecting morning and afternoon samples on 10 days at three Lake Erie beaches. The difference in *E. coli* concentration found in the afternoon sample (when swimmers were present) and the concentration found in the morning sample (when no swimmers were present) was calculated (table 5). On days when few (less than 20) or many (greater than 20) swimmers were present during the afternoon sampling, median *E. coli* concentrations were higher in the morning than in the afternoon;

Table 4. Summary of correlations between $\log_{10}$ *Escherichia coli* concentrations in water and environmental or water-quality factors at three Lake Erie beaches, May-September 1997

[NS, is not statistically significant at $\alpha = 0.05$; all other relations were statistically significant; n is the number of samples]

	Turbidity	Suspended sediment concentration	Rainfall[a]	Weighted rainfall[b]	Overflow[c]	Wind speed	Water temperature	Resuspension index[d]
All data								
Pearson's r	.443	NS	.391	.393	.385	.182	NS	.599
Spearman's rho	.571	.586	.358	.368	.447	.243	NS	.540
n	121	82	121	121	121	119	116	29
Edgewater Park								
Pearson's r	.392	.175	.420	.552	.355	NS	NS	NS
Spearman's rho	.427	NS	.321	.516	.387	NS	NS	.330
n	39	27	39	39	39	39	39	39
Villa Angela								
Pearson's r	.645	.485	.477	.510	NS	.320	NS	.421
Spearman's rho	.591	.761	.463	.522	.314	.388	NS	.433
n	41	27	41	41	41	40	39	41
Sims Park- all areas								
Pearson's r	.354	NS	.576	.542	.463	NS	NS	.400
Spearman's rho	.540	.522	.697	.607	.511	.341	NS	.679
n	41	28	41	41	41	40	38	41
Sims Park area 1								
Pearson's r	.405	NS	.646	.605	.515	NS	NS	.477
Spearman's rho	.676	.502	.656	.613	.448	NS	NS	.611
n	41	27	41	41	41	39	39	40
Sims Park area 2								
Pearson's r	.346	NS	.601	.589	.508	NS	NS	.480
Spearman's rho	.505	.528	.607	.586	.475	NS	NS	.444
n	41	27	41	41	41	40	39	40
Sims Park area 3								
Pearson's r	.341	NS	.529	.482	.445	NS	NS	NS
Spearman's rho	.577	.657	.708	.558	.491	NS	NS	.498
n	41	28	41	41	41	40	39	40
Sims Park area 4								
Pearson's r	.368	.362	.489	.453	.399	NS	NS	NS
Spearman's rho	.493	.505	.545	.473	.421	.389	NS	NS
n	39	25	39	39	39	38	37	38

[a]Rainfall was the amount in inches that occurred in the 24-hour period preceding the 6:30 a.m. sampling.

[b]Weighted rainfall was the amount in inches that occurred in the 72-hour period preceding the 6:30 a.m. sampling, with the most recent rainfall receiving the highest weight.

[c]Overflow was the amount in million gallons of wastwater-treament overflow that occurred on the previous day.

[d]Resuspension index was computed by miltiplying the mean concentration of *Escherichia coli* in sediment found on the previous day by the wave height on the current day.

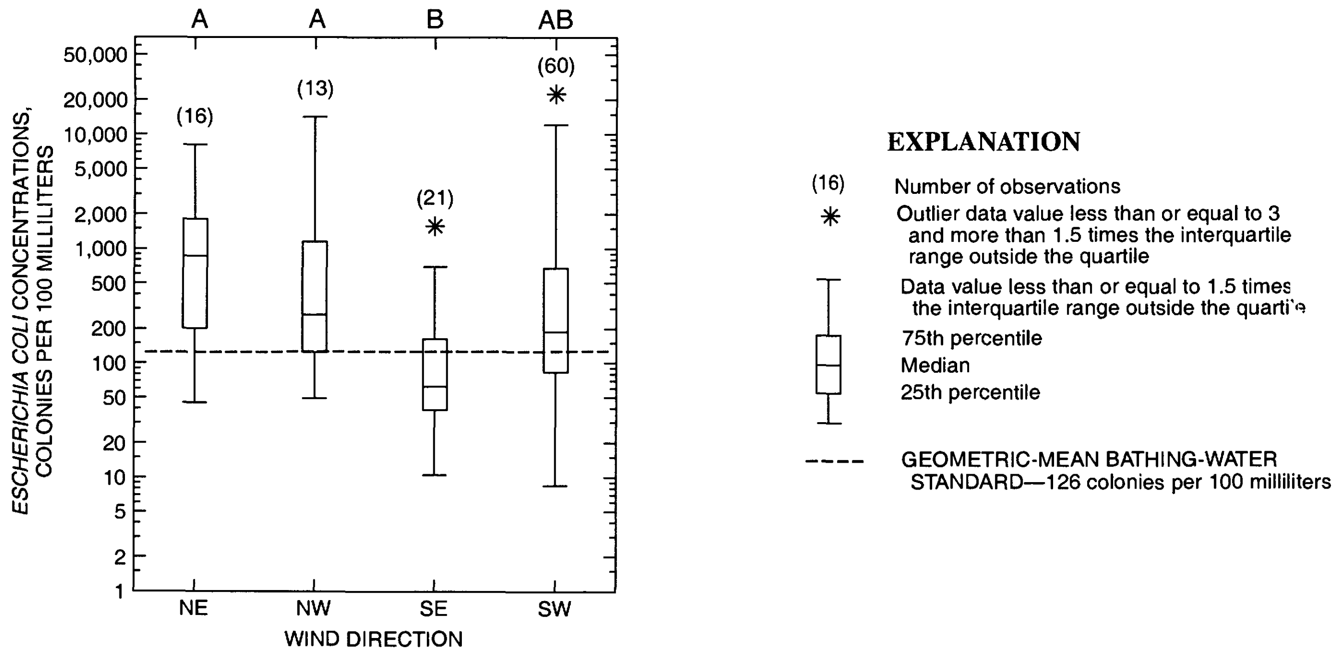

Figure 6. *Escherichia coli* concentrations in water by wind direction at Lake Erie beaches, May-September 1997. (Results of Tukey's test are presented as letters, and concentrations with at least one letter in common do not differ significantly.)

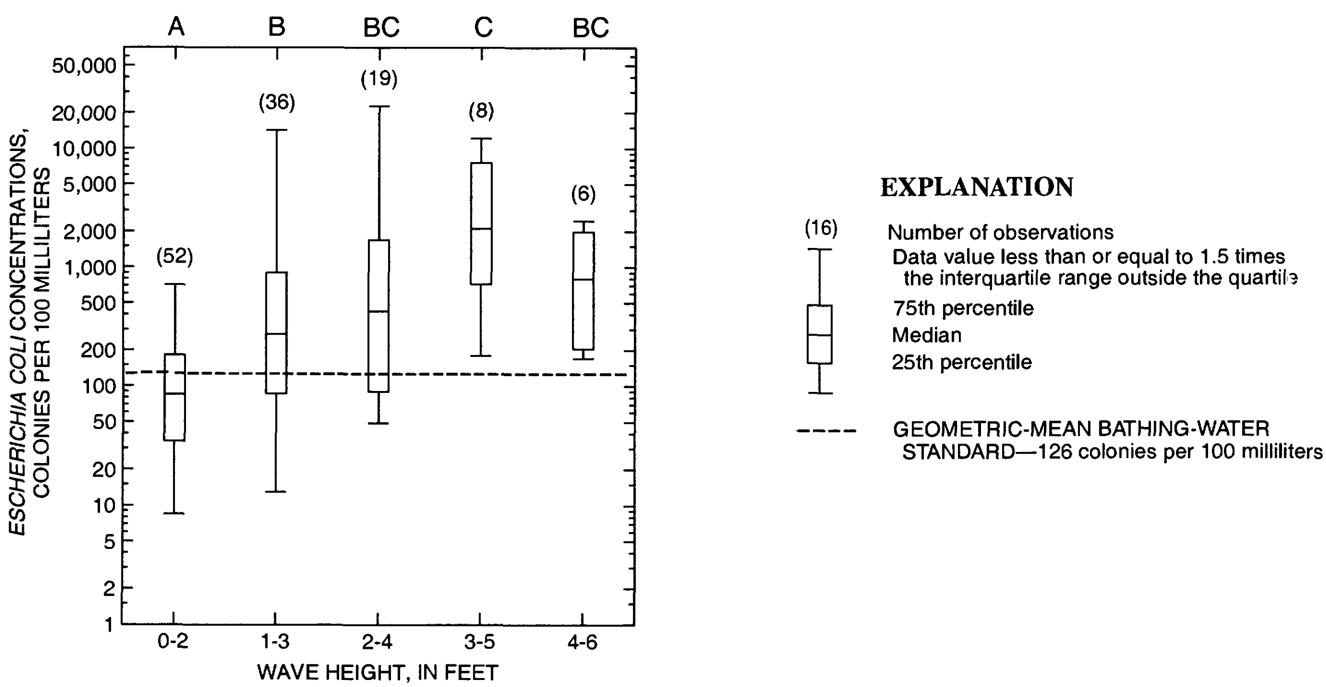

Figure 7. *Escherichia coli* concentrations in water by wave height at Lake Erie beaches, May-September 1997. (Resuts of Tukey's test are presented as letters, and concentrations with at least one letter in common do not differ significantly.)

Table 5. Temporal difference in *Escherichia coli* concentrations in water and the relation to number of swimmers on 10 selected days, June-September 1997, at three Lake Erie beaches

Swimmers[a]	Number of samples	Difference[b]		
		Median	Minimum	Maximum
Few (<20)	33	-0.176	-0.845	1.26
Many (>20)	39	-0.123	-1.43	1.22

[a]Swimmers present during afternoon sampling.
[b]Difference $\log_{10}$ *Escherichia coli* concentration found in the afternoon sample (when swimmers were present) and $\log_{10}$ *Escherichia coli* concentration found in the morning sample (when no swimmers were present).

increased recreation was not generally associated with increased *E. coli* concentrations in the water. A closer examination of the data revealed that on only one day, August 10, *E. coli* concentrations at Villa Angela and Sims were noticeably higher in the afternoon than in the morning (data not shown). Washed-up debris was reported at Villa Angela and Sims at the time of the afternoon sampling. Samples were not collected on August 10 at Edgewater because the beach was closed due to a large public event. Because samples were collected only during the afternoon on 10 days during 1997, more data are needed to thoroughly determine the effect of recreational use on bacterial water quality.

Prediction of *Escherichia coli* concentrations from environmental and water-quality variables

Environmental and water-quality variables found to be related to *E. coli* concentrations were used to develop statistical models that may be used to predict *E. coli* concentrations more quickly and easily than the bacteria-culture methods currently used.

Prediction of *Escherichia coli* concentrations from turbidity—simple linear regression. Turbidity is an easily measured water-quality characteristic. Turbidity values can be obtained within 5 minutes, whereas determining *E. coli* concentrations with current methods takes 24 hours. Simple linear-regression analysis was done to determine the possibility of using turbidity as a predictor of *E. coli* concentrations at the beaches studied.

The data representing the linear-regression relation between *E. coli* concentrations and turbidity by beach and for all data combined are shown in figure 8. The $\log_{10}$- transformed *E. coli* concentrations are linearly related to $\log_{10}$- transformed turbidities; however, much scatter is present in the relations.

Regression diagnostics were computed and plots of residuals against predicted values were examined for curvature and heteroscedasticity. The analysis did not indicate any problems with the transformation of the variables.

Regression statistics for *E. coli* concentrations and turbidity are shown in table 6. The standard error of the regression (S) measures the degree of deviation of observed values from the regression line and is an indicator of the level of uncertainty associated with a prediction, expressed as a percentage of the predicted mean. Standard errors ranged from 46.2 percent at Edgewater to 62.7 percent at Sims.

The slope of the regression line is a measure of the rate of change in $\log_{10}$ *E. coli* with change in $\log_{10}$ turbidity. Although the slopes were all positive, the magnitude of the slope values were quite different among the four data sets (table 6). For Villa Angela data, the rate of change of *E. coli* concentration with changes in turbidity was nearly 1.0, whereas for Sims data, the rate of change was considerably less than 1.0.

The y-intercept is the value for $\log_{10}$ *E. coli* that corresponds to a zero value for $\log_{10}$ turbidity. The y-intercepts from all data sets were positive values, and t-tests on the y-intercepts indicated that they were not significantly different from zero ($p<0.0001$).

The coefficient of determination (R^2) is the fraction of the variation in $\log_{10}$ *E. coli* concentrations that can be explained by $\log_{10}$ turbidity concentrations. For example, an R^2 of 0.37 for all data indicates that 37 percent of the variation in the $\log_{10}$ *E. coli* concentrations can be explained by $\log_{10}$ turbidity (table 6). Different amounts of the variation in *E. coli* concentrations at each beach could be explained by turbidity, with Villa Angela having the highest R^2 and Edgewater having the lowest R^2. This variation could be caused by the different physical characteristics of each

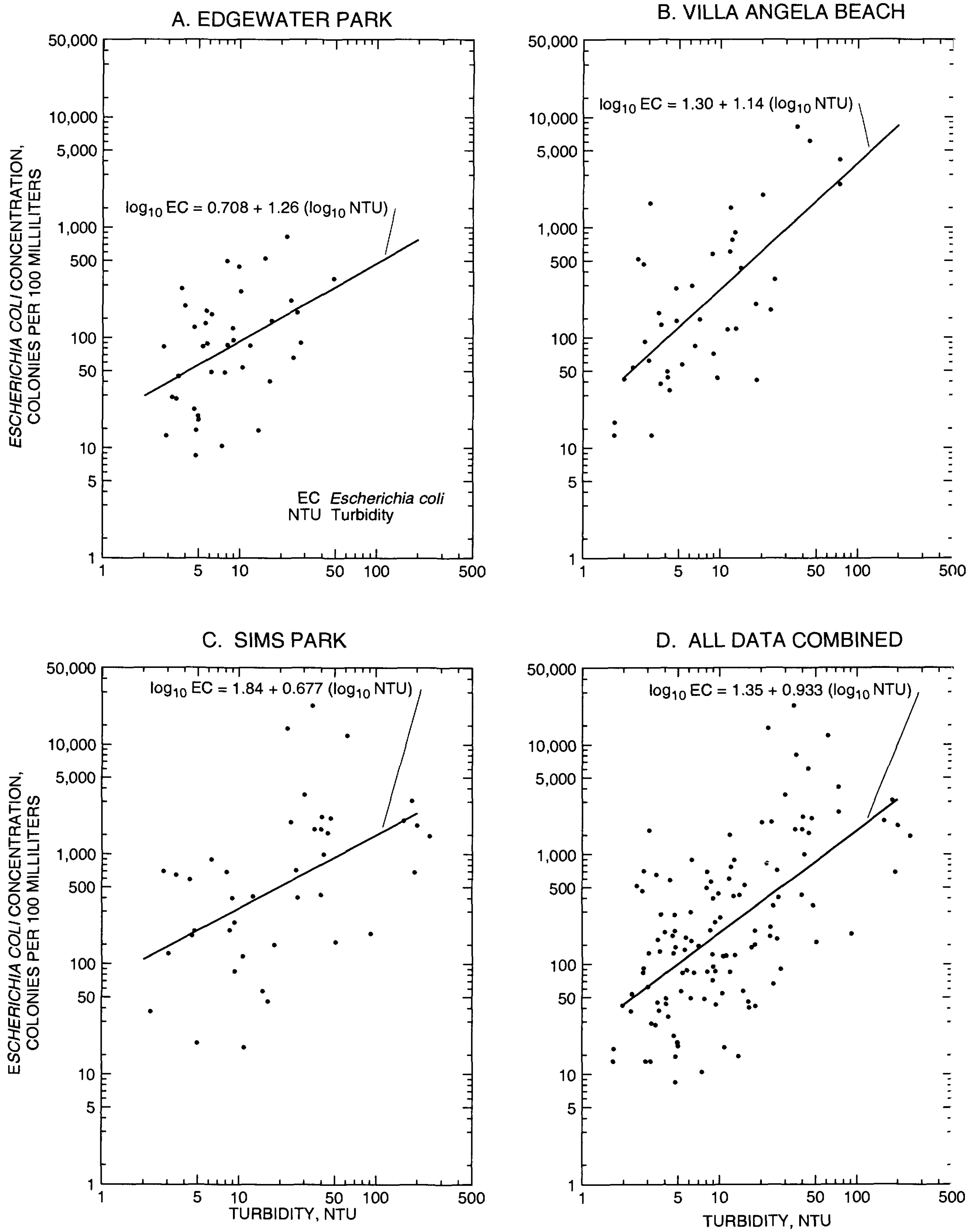

Figure 8. Regression relations between *Escherichia coli* concentrations and turbidity at (A) Edgewater Park, (B) Villa Angela, Cleveland, Ohio, and (C) Sims Park, Euclid, Ohio, and (D) for all data combined, 1997.

beach. In addition, the low R^2 values found for the relations between *E. coli* concentrations and turbidity indicate that there may be other factors that can help explain the variability in *E. coli* concentrations.

Prediction of *Escherichia coli* concentrations from several environmental and water-quality factors—multiple linear regression. Multiple linear regression (MLR) expands simple linear regression by adding multiple explanatory variables. This allows the model to explain more of the variation in *E. coli* concentrations, leaving as little variation as possible to unexplained "noise" (Helsel and Hirsch, 1992, p. 295).

Environmental and water-quality variables found in this study to be related to *E. coli* concentrations were considered for inclusion in the MLR model. Variables were added to the model in two steps: (1) easy-to-measure variables that are determined routinely by local agencies or can be determined without much additional work—turbidity, rainfall, weighted rainfall, wastewater-treatment plant overflow, and wave height; and (2) a hard-to-measure variable that is not routinely determined and requires additional time and quality-control checks—the resuspension index. Wind speed and direction, water temperature, and number of swimmers were shown to be weakly related or not related to *E. coli* concentrations and were not considered for inclusion in the model. Suspended-sediment concentration was not considered for inclusion in the model because it takes longer to determine than turbidity and was found to correlate with turbidity (Pearson's rho=0.425). Different combinations of these variables were used for model development to try to explain the maximum amount of variation in *E. coli* concentrations. Categorical rainfall variables were developed and included in MLR models to account for the days when rainfall amounts were zero.

The simple-linear regression model with turbidity as the independent variable indicated that the slopes, y-intercepts, and R^2 values differed among the three beaches (table 6). To account for the effects that turbidity has on the *E. coli* concentrations at the different sites, beach-specific turbidity values were added to the model.

Rainfall was added to the model using four different types of variables: (1) rainfall, (2) weighted rainfall, (3) categorical rainfall, and (4) categorical weighted rainfall. The rainfall and weighted rainfall were used as continuous variables in the model and are described in a previous section of this report. From the rainfall, a categorical variable was defined (categorical rainfall) containing the following groups; RAINCAT1 was antecedent rainfall greater than 0 to less than 0.2 in. and RAINCAT2 was antecedent rainfall greater than or equal to 0.2 in. An implied third group would be antecedent rainfall of zero. From the weighted rainfall, a categorical variable was defined (categorical weighted rainfall) containing the following groups: R4CAT1 was weighted antecedent rainfall greater than 0.0 and less than 0.5 in., R4CAT2 was weighted antecedent rainfall greater than or equal to 0.5 in., and an implied third group of weighted antecedent rainfall of zero.

Overflow data on metered outfalls were only available for Villa Angela. Wastewater-treatment plant overflows and metered outfalls were combined into a new variable, "overflows and metered outfalls." Because the start times and duration of the flows were unavailable, the flow for the entire previous day was

Table 6. Regression statistics for $\log_{10}$ *Escherichia coli* (*E. coli*) concentrations and $\log_{10}$ turbidities

[The standard error of the regression (S) is the degree of uncertainty associated with a prediction of *E. coli* concentrations from turbidities. The p-value of t-tests on the slopes of the regresssion lines is the probability that the null hypothesis (the slope is equal to zero) is true (α=0.05). The p-value of t-tests on the y intercepts of the regression lines is the probability that the null hypothesis (the y-intercept is equal to zero) is true (α=0.05). The coefficient of determination (R^2) is the fraction of the variation in *E. coli* concentration that can be explained by turbidity]

Data set	Sample size	S (percent)	Slope	p-value for t-tests on slope	y-intercept	p-value for t-tests on y intercept	R^2
Edgewater Park	39	46.2	0.708	0.0063	1.26	< 0.0001	0.185
Villa Angela	41	52.7	1.14	< .0001	1.30	< .0001	.466
Sims Park	41	62.7	.677	.0006	1.84	< .0001	.261
All data	121	58.7	.933	< .0001	1.35	< .0001	.370

used as the overflow and metered outfall variable in the MLR analysis.

In the first step of model construction, easy-to-measure variables were entered into the model and regression analysis was done using Mallows' Cp to help determine the best set of explanatory variables. From the 50 models with the lowest Cp, 10 that had the highest R^2 values and contained variables that were not colinear were chosen for further analysis. Because there were four related rainfall variables, the regression was done four times with a single but different rainfall variable each time, using Mallows' Cp to determine which rainfall variable produced the best-fit models. At the conclusion of the first step of the MLR, one model was chosen on the basis of a low Mallows' Cp, high R^2 values, reduced multicolinearity, and cost of data collection. This model best represented the system and accounted for the most variability in *E. coli* concentrations. In the second step, resuspension index, the hard-to-measure variable, was added. The resuspension index did not significantly improve the model and was eliminated.

The chosen model contained three explanatory variables: weighted categorical rainfall, beach-specific turbidity, and wave height. Also included were y-intercept terms to correct for the different magnitudes of *E. coli* concentrations among the three beaches. This model explained 58 percent of the variability (R^2 = 0.58) in $\log_{10}$ *E. coli* concentrations. This is an improvement over turbidity as the sole explanatory variable, wherein the R^2 was only 0.37 percent for all beaches combined. Weighted categorical rainfall was used because it produced the lowest Mallows' Cp and highest R^2 values among the four different rainfall variables. Overflows and metered outfalls for the 1997 recreational season were not statistically significant and therefore were not included in the model. This variable may be better represented by including time-of-day information, which was not available for this study. The model could also be improved by expanding the data set to include concentrations of *E. coli* in treated wastewater-treatment-plant effluents and occurrences of sewage-line breaks.

The equations to predict *E. coli* concentrations using the chosen model for each beach are as follows:

For Edgewater Park,

$$\log_{10}EC = 0.922 + 0.310(R4CAT1) + 0.523(R4CAT2) + 0.214(WAVEHT) + 0.168(A1TURB) \quad (1)$$

or

$$EC = 8.36 \times 2.04^{R4CAT1} \times 3.33^{R4CAT2} \times 1.64^{WAVEHT} \times 1.47^{A1TURB}. \quad (2)$$

For Villa Angela,

$$\log_{10} EC = 1.09 + 0.310(R4CAT1) + 0.523(R4CAT2) + 0.214(WAVEHT) + 0.584(A2TURB) \quad (3)$$

or

$$EC = 12.3 \times 2.04^{R4CAT1} \times 3.33^{R4CAT2} \times 1.64^{WAVEHT} \times 3.84^{A2TURB}. \quad (4)$$

For Sims Park,

$$\log_{10}EC = 1.83 + 0.310(R4CAT1) + 0.523(R4CAT2) + 0.214(WAVEHT) + 0.111(A3TURB) \quad (5)$$

or

$$EC = 67.6 \times 2.04^{R4CAT1} \times 3.33^{R4CAT2} \times 1.64^{WAVEHT} \times 1.29^{A3TURB}. \quad (6)$$

In all these equations,

EC is the *Escherichia coli* concentration, in colonies per 100 mL,

R4CAT1 is 1 if the 3-day weighted antecedent rainfall was greater than 0 in. but less than 0.5 in. and 0 if the 3-day weighted antecedent rainfall was greater than 0.5 in.

R4CAT2 is 0 if the 3-day weighted antecedent rainfall was greater than 0 in. but less than 0.5 in. and 1 if the 3-day weighted antecedent rainfall was greater than 0.5 in.

WAVEHT is assigned a value of 1, 2, 3, or 4 according to the following characteristics:

1 when wave heights within a series vary between a minimum 0 and a maximum of 2 ft.
2 when wave heights within a series vary between a minimum 1 and a maximum of 3 ft.
3 when wave heights within a series vary between a minimum 2 and a maximum of 4 ft.
4 when wave heights within a series vary between a minimum 3 and a maximum of 6 ft.

A1TURB is the $\log_{10}$ of the turbidity in NTU's at Edgewater Park.

A2TURB is the $\log_{10}$ of the turbidity in NTU's at Villa Angela.

A3TURB is the $\log_{10}$ of the turbidity in NTU's at Sims Park.

Table 7. Prediction of *Escherichia coli* (*E. coli*) concentrations using the multiple linear regression model and different combinations of randomly selected explanatory variables

[NTU, Nephelometric Turbidity Unit; mL is milliliters; col/100 mL is colonies per 100 milliliters]

Model-input variables			Predicted *E. coli* concentration (col/100 mL)	90-percent prediction interval (col/100 mL)	Probability that true concentration is greater than 235 col/100 mL (percent)
Wave height (feet)	Weighted categorical rainfall (inches)	Turbidity (NTU)			
Edgewater Park					
2 - 4	>0 to 0.5	25	130	9 - 1,870	36
0 - 2	>0.5	10	68	10 - 470	14
1 - 3	0	30	40	6 - 260	6
Sims Park					
3 - 5	>0 to 0.5	45	1,500	67 - 34,000	84
1 - 3	>0 to 0.5	17	510	45 - 5,800	70
0 - 2	0	5	130	16 - 1,100	33
2 - 4	>0.5	30	1,500	100 - 20,000	87
Villa Angela					
3 - 5	>0 to 0.5	40	1,600	52 - 48,000	82
1 - 3	>0.5	10	430	40 - 4,500	66
1 - 3	0	7	100	23 - 2,200	49

The MLR model was used to compute 90-percent prediction intervals for *E. coli* concentrations from different combinations of randomly selected explanatory variables (table 7). Ninety-percent prediction intervals were fairly wide because they represent the range of values that the true *E. coli* concentration is expected to assume 90 percent of the time, given a single instance of predictor-variable attributes. More important for assessing recreational water quality are the probabilities of *E. coli* concentrations being greater than 235 col/100 mL, the single-sample bathing-water standard. Knowledge of these probabilities may aid the beach manager in deciding when to post a beach by quantifying the level of uncertainty associated with a prediction. For example, at Edgewater, if wave heights are 2-4 ft, weighted categorical rainfall is > 0 to 0.5 in., and turbidity is 25 NTU, the predicted *E. coli* concentration is 130 col/100 mL. Although the 90-percent prediction interval is quite wide, there is a 36 percent chance that the true *E. coli* concentration exceeds the single-sample bathing-water standard. Conversely, at Sims, if wave heights are 3-5 ft, weighted categorical rainfall is > 0 to 0.5 in., and turbidity is 45 NTU, the predicted *E. coli* concentration is 1,500 col/100mL; there is an 84 percent chance that the true *E. coli* concentration exceeds the single-sample bathing water standard. With experience and a well-tested model, this could be an effective tool to aid beach managers.

Comparison of multiple-linear-regression model to current methods for evaluating beach 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* concentration). If antecedent *E. coli* concentration is greater than the single-sample bathing-water standard (the standard), the beach is posted with a water-quality advisory; if antecedent *E. coli* concentration is less than the standard, the beach is not posted. This method of determining current recreational water-quality conditions from antecedent *E. coli* concentrations does not take into account changes in water quality that occur between the time of sampling on the previous day and the time sample results are complete.

The accuracy of predicting current recreational water-quality conditions using antecedent *E. coli* concentrations was examined for Edgewater using data determined by NEORSD during the 1997 recreational season (Eva Roller, Northeast Ohio Regional Sewer District, written commun., 1997). Table 8 shows a classification table that compares the observed beach condition (current) to the predicted beach condition (based on antecedent *E. coli* concentrations). Using this method, Edgewater Beach would have been correctly posted on 5 of the 21 days (24 percent) when the standard was actually exceeded; this is the sensitivity of the prediction. The beach would not have been posted on 63 of the 79 days (80 percent) when the standard was not exceeded; this is the specificity. The false positive rate indicates that on 16 of the 21 days (76 percent) that the model predicted an exceedance of the standard, the observed *E. coli* concentration did not exceed the standard. The false negative rate indicates that on 16 of the 79 days (20 percent) the model predicted that the standard was not exceeded; the observed *E. coli* concentration exceeded the standard. These percentages are summarized in the first column of table 9.

The accuracy of predicting recreational water-quality conditions was determined for all three beaches using the MLR model and USGS-collected data. Two summaries of classification tables are presented in the second and third columns of table 9; (1) the percentages based on all data used to develop the MLR model without removing any portion of the data set, and (2) the percentages based on the test data set. The test data set was created by splitting the data set into three smaller data sets, ensuring that each smaller set contained a balance of each wave-height category. The MLR model was used on each of the smaller data sets, and the results were combined to form a second classification table. The percentages in these two classification tables were similar, indicating that the model was not biased to a portion of the overall data. This is not a validation of the model, but a test to ensure that one portion of the data does not control the regression. More work needs to be done to validate the MLR, especially during seasons that more typically represent meteorological conditions in the area—for example, a greater frequency and intensity of summer rains.

An accurate model is one that maximizes sensitivity and specificity while minimizing the false positive and negative rates. Table 9 shows that during 1997, the MLR model would have performed as well as, and in some cases better than, the use of antecedent *E. coli* concentrations (the current method) to predict recreational water quality. The use of the MLR model

Table 8. Classification table for predictions of current recreational water-quality conditions using antecedent *Escherichia coli* (*E. coli*) concentrations for Edgewater Park, Cleveland, Ohio, May-September 1997

[Antecendent *E. coli* concentrations were determined by the Northeast Ohio Regional Sewer District, Cuyahoga Heights, Ohio (Eva Roller, written commun., 1997)]

		Predicted		Total
		Event[a]	Nonevent[b]	
Observed	Event	5	16	21
	Nonevent	16	63	79
	Total	21	79	100

Sensitivity[c] = 24%
Specificity[d] = 80%
Correct = 68%
False positive[e] = 76%
False negative[f] = 20%

[a]An event occurs when the *E. coli* concentration exceeds the single-standard bathing-water standard of 235 colonies per 100 milliliters.
[b]A nonevent occurs when the *E. coli* concentration is less than the single-standard bathing-water standard of 235 colonies per 100 milliliters.
[c]Sensitivity was the proportion of event responses that were predicted correctly as events.
[d]Specificity was the proportion of nonevent responses that were correctly predicted as nonevents.
[e]The false positive rate was the proportion of predicted events that were observed as nonevents.
[f]The false negative rate was the proportion of predicted nonevents that were observed as events.

Table 9. Results of classification tables comparing the proportions of correct and incorrect predicted recreational water-quality conditions using antecedent *Escherichia coli* concentrations and a multiple-linear-regression (MLR) model for Lake Erie beaches, May-September 1997

[NEORSD is Northeast Ohio Regional Sewer District, Cuyahoga Heights, Ohio (Eva Roller, written commun., 1997); USGS is U.S. Geological Survey]

Model evaluation parameters	Predictions based on		
	Edgewater Park - NEORSD data (percentage)	All beaches - USGS data (percentage)	
	Antecedent *Escherichia coli* concentrations	MLR all data	MLR test data[a]
Sensitivity[b]	24	75	71
Specificity[c]	80	81	81
Correct	68	85	76
False positive[d]	76	25	27
False negative[e]	20	19	21

[a]Data sets were developed so that data used to develop the multiple linear-regression model were not used to test the model.
[b]Sensitivity was the proportion of event responses that were correctly predicted as events.
[c]Specificity was the proportion of nonevents responses that were correctly predicted as nonevents.
[d]The false positive rate was the proportion of predicted events that were observed as nonevents.
[e]The false negative rate was the proportion of predicted nonevents that were observed as events.

improved the sensitivity of the prediction and the percentage of correct predictions over using antecedent *E. coli* concentrations. The MLR predictions still erred to a similar degree as the current method with regard to false negatives and specificity. A false negative would allow swimming when, in fact, the bathing-water standard was exceeded. Nevertheless, this investigation shows that the use of a MLR model based on water-quality and environmental factors to predict *E. coli* concentrations is feasible and fairly accurate in its predictive ability.

Future research could focus on improving and validating MLR models to predict recreational water quality. To improve the predictive ability of the MLR models and, in particular, to improve the specificity, studies should focus on adding to the model other environmental and water-quality variables that may affect *E. coli* concentrations and better represent the variables on a beach-specific basis. Other environmental and water-quality factors that should be investigated or for which additional data are needed include (1) the effect of *E. coli* in sediments from the nearshore bathing area on the rate of improvement of recreational water quality after wet-weather events, (2) recirculation of *E. coli* from deeper sediments outside the bathing area (to help explain increases in *E. coli* concentrations in the absence of rainfall or wastewater-treatment-plant overflows), (3) recreational use, (4) direction of lake currents, (5) site-specific rainfall amounts, and (6) time-of-day information on overflows and metered outfalls. To help improve the specificity of MLR models, one may consider the inclusion of variables for concentrations of *E. coli* in treated wastewater-treatment-plant effluents and occurrences of sewage-line breaks. In addition, the use of several rapid methods for determination of recreational water quality could be used as additional variables in model refinements or as a substitute for current bacterial-culture assays. Possible methods include the use of experimental rapid assay methods for determination of *E. coli* concentrations or other fecal indicators and the use of chemical tracers for fecal contamination, such as coprostanol (a degradation product of cholesterol) or caffeine.

Summary and conclusions

Water-quality advisories and beach closings because of sewage contamination are common at Lake Erie beaches in Ohio and other beaches in the United

States. These advisories are issued on the basis of established state recreational water-quality standards for concentrations of fecal-indicator bacteria, such as *Escherichia coli*. Little is known, however, about the environmental and water-quality factors that affect concentrations of fecal indicators at beaches. The study described here was done to obtain more information on the effect of these factors so that water-resource managers may be able to assess recreational water quality more quickly and accurately than is provided by current bacterial-culture methods.

Data were collected during eight field studies throughout the 1997 recreational season (May through September) at three public bathing beaches in the Cleveland, Ohio, metropolitan area—Edgewater Park, Villa Angela, and Sims Park. Water and lake-bottom sediments were collected each morning during a variety of environmental conditions—during dry, calm weather; before, during, and after rainfall; during increased wave heights; and before and after heavy recreational use. On 10 selected days, water samples were also collected after 1 p.m., when swimmers were present at the beaches. Water samples were analyzed for *E. coli*, suspended-sediment concentrations, and turbidity. Ancillary environmental data compiled or collected included water temperature, wind speed and direction, number of swimmers, wave heights, volumes of effluent from wastewater-treatment-plant overflows and metered outfalls, and daily rainfall amounts. Lake-bottom sediment samples were analyzed for *E. coli* concentrations and percent dry weight. A special protocol was tested and developed for removal of *E. coli* from sediment before determining *E. coli* concentrations. Some sediment samples were also analyzed for total organic carbon and particle-size distribution.

Concentrations of *E. coli* in water and lake-bottom sediments were lowest at Edgewater Park among the three beaches. Concentrations of *E. coli* in the four areas of Sims Park were higher and more variable than at either Edgewater Park or Villa Angela. Of the 41 days sampled, concentrations of *E. coli* in water exceeded the single-sample bathing-water standard of 235 col/100 mL at Edgewater Park on 7 days, at Villa Angela on 17 days, and at Sims Park on 23 to 27 days.

Finer sediments were found at Edgewater than at Villa Angela or Sims; at Edgewater, most of the sediments were classified as medium to fine sands. The sediments collected at the beaches at Villa Angela and Sims Park varied considerably in particle sizes; most were classified as medium sands to gravels. Total organic carbon concentrations in sediment ranged from 0.8 to 3.2 mg/kg; however, not enough data were collected to examine total organic carbon concentration as a factor affecting *E. coli* concentrations in sediment.

Plots showing concentrations of *E. coli* in water and lake-bottom sediments and wave heights, volumes of wastewater-treatment-plant overflows, and rainfall amounts throughout the summer were examined to determine what factors affected *E. coli* concentrations in water and sediment. Peak *E. coli* concentrations in water samples occurred either during or immediately following rainfall or wastewater-treatment-plant overflows. Higher concentrations of *E. coli* in water were also found when wave heights were 2-4 ft or greater, although these same periods were usually during or after rainfall. Sediment *E. coli* concentrations generally increased at the same time or after a peak in water *E. coli* concentrations.

Data collected at the three beaches showed that short-term storage (less than one week) of *E. coli* in lake-bottom sediments may have occurred, although there was no direct evidence for long-term storage of *E. coli*. Unfortunately, the resuspension of *E. coli* from lake-bottom sediments by wave actions could not be adequately assessed because higher wave heights usually occurred during or after a rainfall or wastewater-treatment-plant overflow, and bacterial water quality at Edgewater was usually good during the sampling period.

Correlation analysis was used to measure the relations between *E. coli* concentrations and other continuous variables. Statistically significant correlations were found between *E. coli* concentrations and turbidity, rainfall, weighted rainfall, wastewater-treatment-plant overflows, and the resuspension index. Relations between *E. coli* concentrations and wind speed or water temperature were weak or not statistically significant.

Analysis of variance was used to measure the relations between *E. coli* concentrations and categorical variables. Wave height was found to be more useful in assessing recreational water quality than wind direction or the presence of swimmers. Median *E. coli* concentrations generally increased with increasing wave height, and statistically significant differences were found among some of the wave-height categories. With the exception of wave-height category 1 (minimum of 0 to maximum of 2 ft waves), median

E. coli concentrations for wave-height categories exceeded the geometric mean bathing-water standard of 126 col/100 mL.

Because turbidity results can be obtained within 5 minutes, simple linear-regression analysis was done to determine the possibility of using turbidity as a predictor of *E. coli* concentrations at the beaches studied. The regression relation was statistically significant, and the slopes and y-intercepts were significant and reasonable in sign and magnitude. However, small values for the coefficient of determinations for the relations by beach and for all data combined ($R^2 = 0.185$ to 0.466), indicated that there may be other factors that can help explain the variability in *E. coli* concentrations.

Environmental and water-quality variables that were shown to be related to *E. coli* concentrations were considered for inclusion in a multiple linear regression (MLR) model. One model was chosen that contained easy-to-measure variables, most reasonably represented the system, and accounted for the most variability in *E. coli* concentrations. This model contained y-intercept terms to correct for the different magnitudes of *E. coli* concentrations among the three beaches, weighted categorical rainfall, beach-specific turbidity, and wave height. These explanatory variables explained 58 percent of the variability ($R^2 = 0.58$) in *E. coli* concentrations. The MLR model improved the sensitivity of the prediction and the percentage of correct predictions over using antecedent *E. coli* concentrations (the current method); however, the MLR predictions still erred to a similar degree as the current method with regard to false negatives and specificity.

Future research could focus on improving and validating MLR models to predict recreational water quality. More work needs to be done to validate the MLR from data collected during other recreational seasons, especially during a season with a greater frequency and intensity of summer rains. To improve the predictive ability of the MLR models, studies should focus on adding to the model other environmental and water-quality variables that may affect *E. coli* concentrations and better represent the variables on a beach-specific basis. In addition, experimental rapid biological or chemical methods for determination of recreational water quality could be used as additional variables in model refinements or as a substitute for current bacterial-culture assays.

References cited

Aldom, J., Jamieson, E., Prout, T., Walsh, M., Van Bakel, D., Griffiths, R., and Palmateer, G., 1998, Rapid fecal coliform and *Escherichia coli* detection in the recreational waters of Lake Huron beaches and an inland beach in 1997: London, Ontario, Ausable Bayfield Conservation Authority, 32 p.

American Public Health Association, American Water Works Association, and Water Pollution Control Federation, 1995, Standard methods for the analysis of water and wastewater (19th ed.): Washington, D.C., American Public Health Association [variously paged].

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

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

Bromel, M., Saylor, G., Coles, F., Zimmerman, J., and Johnson, K., 1978, Bacteriological analyses of Lake Metigoshe water and sediments: Fargo, N.D., North Dakota State University, Agricultural Experiment Station Bulletin 507, 14 p.

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.

Davies, C.M., Long, J.A.H., Donald, M., and Ashbolt, N.J., 1995, Survival of fecal microorganisms in marine and freshwater sediments: Applied and Environmental Microbiology, v. 61, no. 5, p. 1888-1896.

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

Dutka, B.J., Bell, J.B., and Liu, L.S., 1972, Microbiological examination of offshore Lake Erie sediments: Journal Fisheries Research Board of Canada, v. 31, no. 3, p. 299-308.

Francy, D.S., Hart, T., and Virosteck, C.M., 1996, Effects of receiving-water quality and wastewater treatment on injury, survival, and regrowth of fecal-indicator bacteria and implications for assessment of recreational water quality: U.S. Geological Survey Water-Resources Investigations Report 96-4199, 42 p.

Gannon, J.J., Busse, M.K., and Schillinger, J.E., 1983, Fecal coliform disappearance in a river impoundment: Water Research, v. 17, no. 11, p. 1595-1601.

Geldreich, E.E., Nash, H.D., Spino, D.F., and Reasoner, D.J.,1980, Bacterial dynamics in a water supply fecal-indicator bacteria and implications for assessment of recreational water quality: U.S. Geological Survey Water-Resources Investigations Report 96-4199, 42 p.

Gannon, J.J., Busse, M.K., and Schillinger, J.E., 1983, Fecal coliform disappearance in a river impoundment: Water Research, v. 17, no. 11, p. 1595-1601.

Geldreich, E.E., Nash, H.D., Spino, D.F., and Reasoner, D.J.,1980, Bacterial dynamics in a water supply reservoir—A case study: Journal of American Water Works Association, v. 72, p. 31-40.

Gerba, C.P., and McLeod, J.S., 1976, Effect of sediments on the survival of *Escherichia coli* in marine waters: Applied and Environmental Microbiology, v. 32, no.1, p. 114-120.

Ghiorse, W.C., and Balkwill, D.L., 1983, Enumeration and morphological characterization of bacteria indigenous to subsurface environments: Developments in Industrial Microbiology, v. 24, p. 213-224.

Goyal, S.M., Gerba, C.P., and Melnick, J.L., 1977, Occurrence and distribution of bacterial indicators and pathogens in canal communities along the Texas coast: Applied and Environmental Microbiology, v. 34, no. 2, p. 139-149.

Grimes, D.J., 1980, Bacteriological water quality effects of hydraulically dredging contaminated upper Mississippi River bottom sediment: Applied and Environmental Microbiology, v. 39, no. 4, p. 782-789.

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].

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

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.

Lehman, G.S., and Fogel, M.M., 1976, Investigation of bacteriological pollution of recreational waters in Arizona: Tucson, Ariz., University of Arizona Department of Watershed Management and University of Arizona School of Renewable Natural Resources.

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

Marino, R.P. and Gannon, J.J., 1991, Survival of fecal coliforms and fecal streptococci in storm drain sediment: Water Research, v. 25, no. 9, p. 1089-1098.

Matson, E.A., Hornor, S.G., 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., 1992, Distribution and variability of fecal-indicator bacteria in Scioto and Olentangy Rivers in the Columbus, Ohio, area: U.S. Geological Survey Water-Resources Investigations Report 92-4130, 61 p.

Myers, D.N., and Sylvester, F.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.

Natural Resources Defense Council, 1996, Testing the waters—Volume VI: New York, 116 p.

Ohio Environmental Protection Agency, 1990, Numerical and narrative criteria for recreational use designations: Ohio Administrative Code, chap. 3745-1, May 1, 1990, p. 07-07 and 07-34.

——— 1992, Ohio water resource inventory: Columbus, Ohio, Division of Water Quality Planning & Assessment, Ecological Assessment Section, p. 121-122.

Pommepuy, M., Guillaud, J.F., Dupray, E., Derrien, A., Guyader, F.L., and Cornier, M., 1992, Enteric bacteria survival factors: Water Science Technology, v. 25, no. 12, p. 93-103.

SAS Institute, Inc., 1990, SAS/STAT User's guide, version 6, fourth edition, volume 2: Cary, N.C., p. 1071-1126.

Sherer, B.M., Miner, J.R., Moore, J.A., and Buckhouse, J.C., 1988, Resuspending organisms from a rangeland stream bottom: Transactions of the American Society of Agricultural Engineers, v. 31, no. 4, p. 1217-1222.

——— 1992, Indicator bacterial survival in stream sediments: Journal of Environmental Quality, v. 21, p. 591-595.

Schertz, T.L., Childress, C., Kelly, V., O'Brien, M., and Pederson, G. 1998, Workbook for preparing a district quality-assurance plan for water-quality activities: U.S. Geological Survey Open-File Report 98-98 [variously paged].

Shindel, H.L., Mangus, J.P., and Trimble, L.E., 1998, Water resources data, Ohio, water year 1997, Volume 2: Columbus, Ohio, U.S. Geological Survey Water Data Report OH-97-2, p. 396-411.

Stephenson, G.R. and Rychert, R.C., 1982, Bottom sediment—A reservoir of *Escherichia coli* in rangeland streams: Journal of Range Management, v. 35, no. 1, p. 119-123.

Tunnicliff, B., and Brickler, S., 1984, Recreational water quality analyses of the Colorado River corridor in Grand Canyon: Applied and Environmental Microbiology, v. 48, no. 5, p. 909-917.

U.S. Army Corps of Engineers, 1984, Shore protection manual, volume 1: Washington, D.C., p. 3-2.

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

——— 1986, Test methods for evaluating solid waste—physical/chemical methods, Laboratory Manual v., section C, SW-846.

APPENDIX A—Evaluation of methods for enumeration of fecal-indicator bacteria in lake-bottom sediment

Unlike water samples, standard methods for enumeration of fecal-indicator bacteria in sediment samples are not well established. The American Public Health Association (1995, section 9221 A) recommends the use of most-probable number (MPN) methods for solid or semisolid samples instead of membrane-filtration (MF) methods. However, MF methods are more precise than MPN methods and are easier to use on a large-scale study such as this one. In fact, the MF method has been used by several investigators in enumerating fecal-indicator bacteria in sediment (Gannon and others, 1983; Geldreich and others, 1980; Grimes, 1980; Van Donsel and Geldreich, 1971; Davies and others, 1995).

Because sediment can clog the membrane filter and interfere with colony differentiation, the bacteria must be separated from the solid phase before MF analysis. Investigators have used various methods to detach bacteria from sediments including agitation, surfactant application, sonication, and blending. In a study of bottom sediments in the Mississippi River, Grimes (1980) made a sediment/buffer mixture, which was mixed on a platform shaker for 30 minutes and allowed to settle for 1 hour. The resultant liquid phase was then used for MF analysis. In studying bacteria in subsurface sediments, mild surfactants were used to remove cells from sediment particles (Ghiorse and Balkwill, 1983). In a study on the survival of fecal coliforms in marine and freshwater sediments, sonication was shown to be a suitable method for separating bacteria from sediment particles for subsequent MF (Davies and others, 1995). Dutka and others (1972) combined Lake Erie bottom sediments and buffered water in a blender for 1 minute at moderate speed. The blended sample was used to enumerate several different groups of bacteria.

Methods. Several treatments for the separation of fecal-indicator bacteria from Lake Erie bottom sediments were tested to determine whether they were suitable for use in the field studies described in this report. A suitable method is relatively easy to use, has the highest and least variable recoveries of fecal-indicator bacteria from sediments among the treatments tested, and does not cause injury to cells. Fecal coliforms were used instead of *E. coli* because an enhanced-recovery method for injured fecal coliforms was previously tested, and determination of the proportion of injured cells was necessary to adequately evaluate the treatments. Because it was beyond the scope of this project to determine the actual numbers of bacteria separated from sediments, relative recoveries of bacteria were determined by enumerating fecal coliforms in the liquid phase after treatment.

Lake Erie bottom sediments were collected using the methods described in this report. Replicate aliquots of sediment were diluted in sterile saline buffer and the sediment/buffer mixtures received one or more of the following treatments: (1) shaking on a wrist-action shaker with and without Tween 80 (Sigma, St. Louis, Mo.), a commonly-used surfactant, (2) blending at 8,000 revolutions per minute for 30 seconds, and (3) sonicating at 100 watts for 30 seconds. After treatment, the sediment/buffer mixture was allowed to settle for 30 seconds, and the liquid phase was poured off into a second bottle. The liquid phase was plated by use of the standard mFC agar method for fecal coliforms (Britton and Greeson, 1989) and (or) the enhanced-recovery mFC agar method for recovery of injured fecal coliforms (Francy and others, 1996). In most experiments, each treatment was tested by determining mean (or median) recovery of fecal coliforms in at least three sediment/buffer replicate mixtures.

Results and conclusions. Two replicate jars of lake-bottom sediments were collected from the East Beach at Edgewater Park and composited to qualitatively evaluate the use of surfactants for recovery of fecal coliforms. The addition of two drops of Tween 80 to the sediment/buffer mixture before shaking was not more effective (mean=34 col/100 mL) than shaking alone (mean=33 col/100 mL) in recovery of fecal coliforms from sediment. The addition of five drops of Tween 80 resulted in lower recoveries of fecal coliforms by use of the standard method (mean=15 col/100 mL) than by use of the enhanced-recovery method (83 col/100 mL). Because enhanced-recovery methods detect both healthy and injured organisms and standard methods detect only healthy organisms, the results suggest that Tween 80 added in larger quantities may cause injury or death to cells or otherwise inhibit their growth. Based on these tests, it was decided to not use a surfactant to remove bacteria from lake-bottom sediments during the field studies.

The use of blending, sonicating, and shaking was evaluated statistically to determine their relative effectiveness in the recovery of fecal coliforms from lake-bottom sediments. In the first experiment, a lake-bottom sediment sample was collected from area 3 at Sims (fig. A1). Replicate sediment/buffer mixtures were prepared and treated by blending, sonicating, or shaking; concentrations of fecal coliforms were determined by use of both standard and enhanced-recovery methods. Overall, blending of lake-bottom sediments resulted in the lowest recoveries of fecal coliforms and sonicating resulted in the most variable recoveries. When the standard method was used, analysis of variance indicated that recovery of fecal coliforms by blending was significantly lower (α=0.05) than recovery by shaking or sonicating sediments. However, when the enhanced-recovery method was used, no statistically significant differences in recoveries of fecal coliforms were found between treatments. This suggests that blending may have

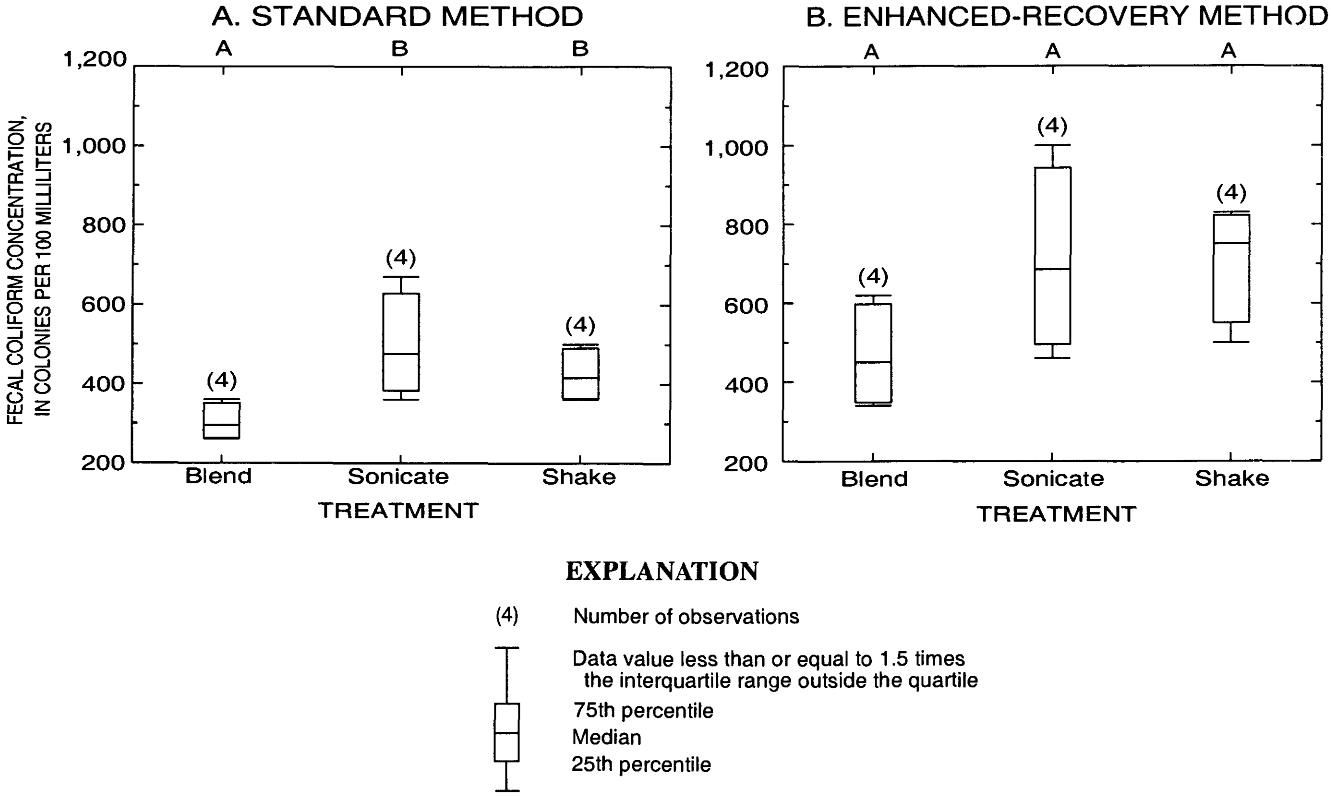

Figure A1. Concentrations of fecal coliforms recovered from Lake Erie bottom sediments collected at Sims Park, Euclid, Ohio, and determined by use of (A) standard and (B) enhanced-recovery methods. (Results of Tukey's test are presented as letters, and treatments with one letter in common were not significantly different.)

caused injury to cells so that they were able to be cultured by the enhanced-recovery method but not by the standard method.

In the second experiment, sediment/buffer mixtures were treated three ways, and recoveries of fecal coliforms from sediments collected at Edgewater (West Beach and East Beach) and Sims were determined by use of the standard method only (fig. A2). The only statistically significant difference found between median recoveries (α=0.05) was between shaking and blending of sediments collected at West Beach. Generally, blending resulted in the recovery of fewer fecal coliforms from sediments than shaking or sonicating. Sonicating resulted in more variable recoveries than shaking; that may be because it was often difficult to maintain a consistent output of sound waves throughout the sample. The motion of the water, the physical and magnetic barriers created by the stir bar, and the distribution of sediment in the sample affected the conduction of sound waves throughout the sample. Based on the results of these studies, shaking was selected as the treatment method for removal of bacteria from lake-bottom sediments during field studies.

Further work was done to refine the shaking protocol for removal of bacteria from sediment. In a series of experiments, concentrations of fecal coliforms were determined in sediments collected at Edgewater, Villa Angela, and Sims after shake times of 15 or 45 minutes. Shaking for 45 minutes resulted in statistically higher recoveries than shaking for 15 minutes for most samples (data not shown). Experiments also were done to determine the effect of decanting or not decanting the liquid phase before plating for fecal coliforms. Using sediments collected at Edgewater, which were considerably finer than sediments from Villa Angela or Sims, not decanting the liquid phase resulted in interference of bacterial growth on culture plates by sediments. Decanting and plating the liquid phase resulted in distinct colonies and enhanced differentiation and enumeration of target colonies. In contrast, using sediments collected at Sims, not decanting resulted in higher recoveries than decanting in one study and no difference in recoveries in another. Based on the results of these experiments, a consistent protocol was established for recovery of *E. coli* from sediments collected from all three beaches for use during field studies. In this protocol, the sediment/buffer mixture was placed on a shaker for 45 minutes and the liquid phase was decanted before plating for concentrations of bacteria.

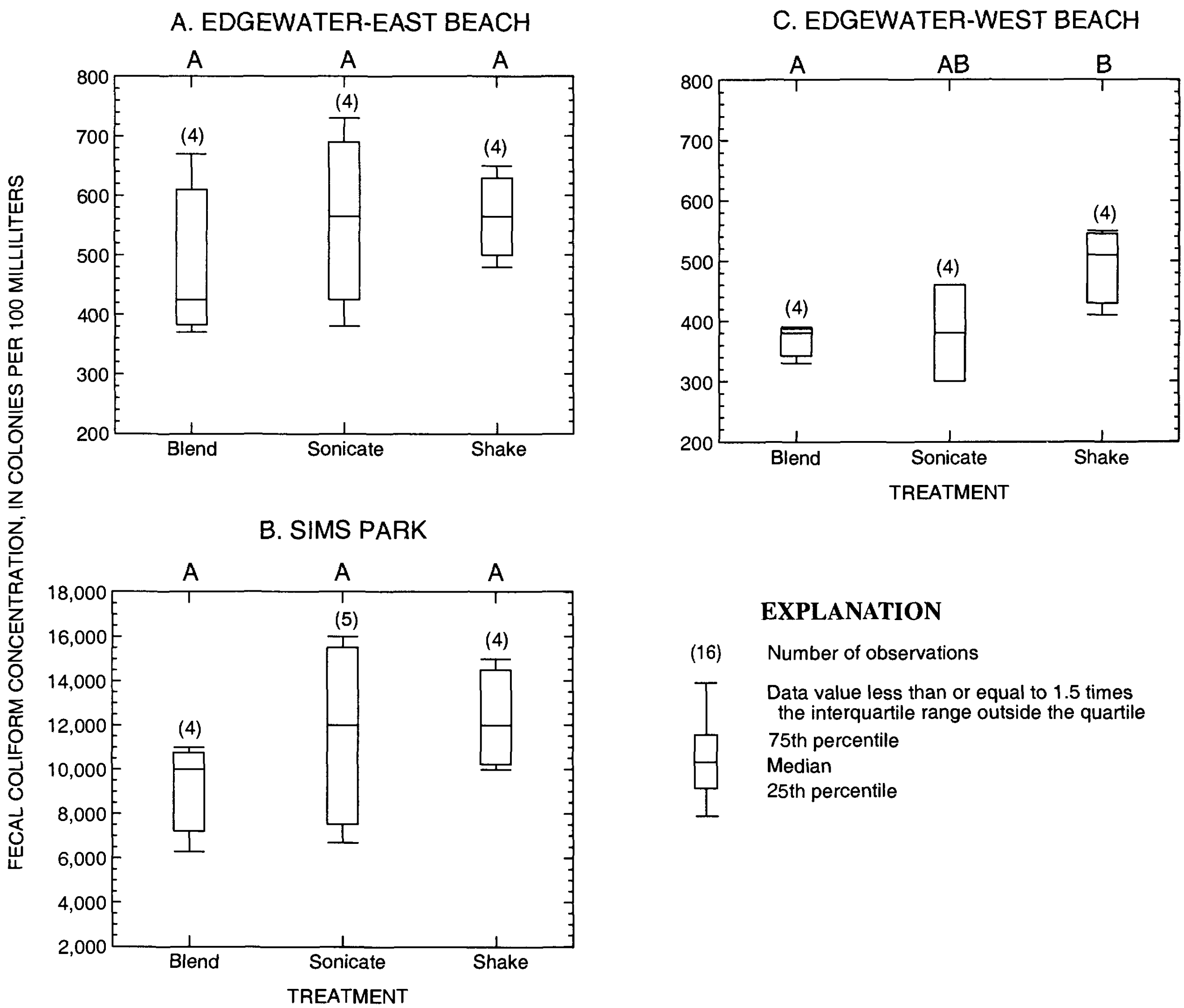

Figure A2. Concentrations of fecal coliforms obtained by use of the standard method recovered from Lake Erie bottom sediments collected at (A) Edgewater-East Beach, (B) Sims Park, and (C) Edgewater-West Beach. (Results of Tukey's test are presented as letters, and treatments with one letter in common were not significantly different.)

APPENDIX B.—Variability of *Escherichia coli* concentrations in water and sediment

Approximately 14 percent of water and 7 percent of sediment samples analyzed for *E. coli* concentrations in this study were quality-assurance replicates. Replicate samples were collected to gain insight into the variability of *E. coli* concentrations in water and sediment samples at recreational beaches, something for which there is little published data. Results from comparison of replicates will aid in the design of future projects and in the interpretation of bacterial water-quality differences in beach-monitoring projects.

Two replicate water or sediment bottles (bottles 1 and 2) were plated for *E. coli* concentrations in duplicate to assess sampling (water samples) or subsampling (sediment samples) and analytical variability. The resultant four replicate platings were called A (regular sample from bottle 1), B (split regular sample from bottle 1), C (replicate sample from bottle 2), and D (split replicate sample from bottle 2). Within-bottle differences were calculated for the following replicate pairs: A-B and C-D; between-bottle differences were calculated for these replicate pairs: A-C, A-D, B-C, B-D.

Between-bottle and within-bottle differences in *E. coli* concentrations for water and sediment samples were determined by using two calculated parameters: (1) the absolute value of the percent difference (PD) of raw data and (2) the absolute value of the $\log_{10}$ difference (AVLD). The PD was calculated by dividing the concentration difference between each replicate pair by the average concentration of the same replicate pair and taking the absolute value. To get the AVLD, the concentration data were $\log_{10}$ transformed and the absolute value of the difference between each replicate pair was taken. Absolute values were used because the A, B, C, and D designations were randomly assigned, and the amount of difference rather than the direction of change was of interest.

The within- and between-bottle PD's and AVLD's for water and sediment samples are shown in figs. B1 and B2. Generally, within-bottle PD's and AVLD's were less than between-bottle values. The PD's and AVLD's were evenly distributed among the range of *E. coli* concentrations in water; however, for *E. coli* concentrations greater than 50 col/g_{dw}, the PD's and AVLD's decreased as *E. coli* concentrations increased.

Summary statistics for PD's and AVLD's are shown in table B1. Median within- and between-bottle PD's and AVLD's for water samples were lower than for sediment samples. For both water and sediment samples, between-bottle PD's and AVLD's (which include sampling and analytical variability) were only slightly higher than within-bottle PD's and AVLD's (which include only analytical variability). When these data were examined by use of t-tests on the ranked data, statistically significant differences were found between within-bottle and between-bottle PD's and AVLD's for sediment samples, but not for water samples (α=0.05). This result indicates that most of the variability in water *E. coli* concentrations was due to analytical variability but that subsampling variability was a significant component of the variability in sediment *E. coli* concentrations. Therefore, in order to obtain the best estimate of *E. coli* concentrations in the water, it is more important to plate bottles in replicate than to collect bottles in replicate. For sediment samples, however, replicate subsamples and replicate platings should be done.

The 90-percent upper confidence limits for the 95-percent quantile is an important statistic in the design of future beach-monitoring projects. This statistic may be used to interpret differences between *E. coli* concentrations in two water samples to satisfy project objectives and to determine whether the differences are real or due to sampling and analytical variability. For example, the 95-percent quantiles (and 90-percent upper confidence limits) for between-bottle PD's and AVLD's for water samples were 77 percent (79 percent) and 0.35 $\log_{10}$ col/100 mL (0.36 $\log_{10}$ col/100 mL), respectively. This means that if the PD or AVDL between water sample 1 and water sample 2 was equal to or greater than 79 percent or 0.36 $\log_{10}$ col/100 mL, one can say with 90 percent confidence that there is only a 5 percent or less chance that the difference between water sample 1 and water sample 2 is due to sampling and analytical variability alone. In this case, one would be able to say with confidence that the concentration found in water sample 1 is different from the concentration found in water sample 2.

The water and sediment data were then grouped by two factors: (1) magnitude of *E. coli* concentration in replicate A and (2) beach, as shown in table B2. Separate ANOVA and Tukey's tests were done for within-bottle and between-bottle PD's and AVLD's on water and sediment samples to determine the effect of these factors on the PD's and AVLD's. No statistically significant differences were found for data on between- or within-bottle water or sediment samples grouped by magnitude of the *E. coli* concentration. For data grouped by beach, the only statistically significant difference among beaches was found for sediment data, where between-bottle differences were statistically lower at Edgewater Park than at Villa Angela or Sims Park. This difference suggests that in designing future projects, subsampling protocols may need to be changed so as to make between-bottle PD's and AVLD's indistinguishable among sites. For example, subsampling protocols may need to be different for sediments collected at Edgewater (fine to medium sand) than for Villa Angela Sims (coarse sand to fine gravel).

In summary, PD's and AVLD's can be used to examine quality-assurance data on the variability of *E. coli* concentrations in water and sediment. The results were the same whether examining the PD or AVLD, and either parameter may be used in data-analysis procedures or data interpretations. The choice of which one to use is based on project objectives and the personal preference of the investigator.

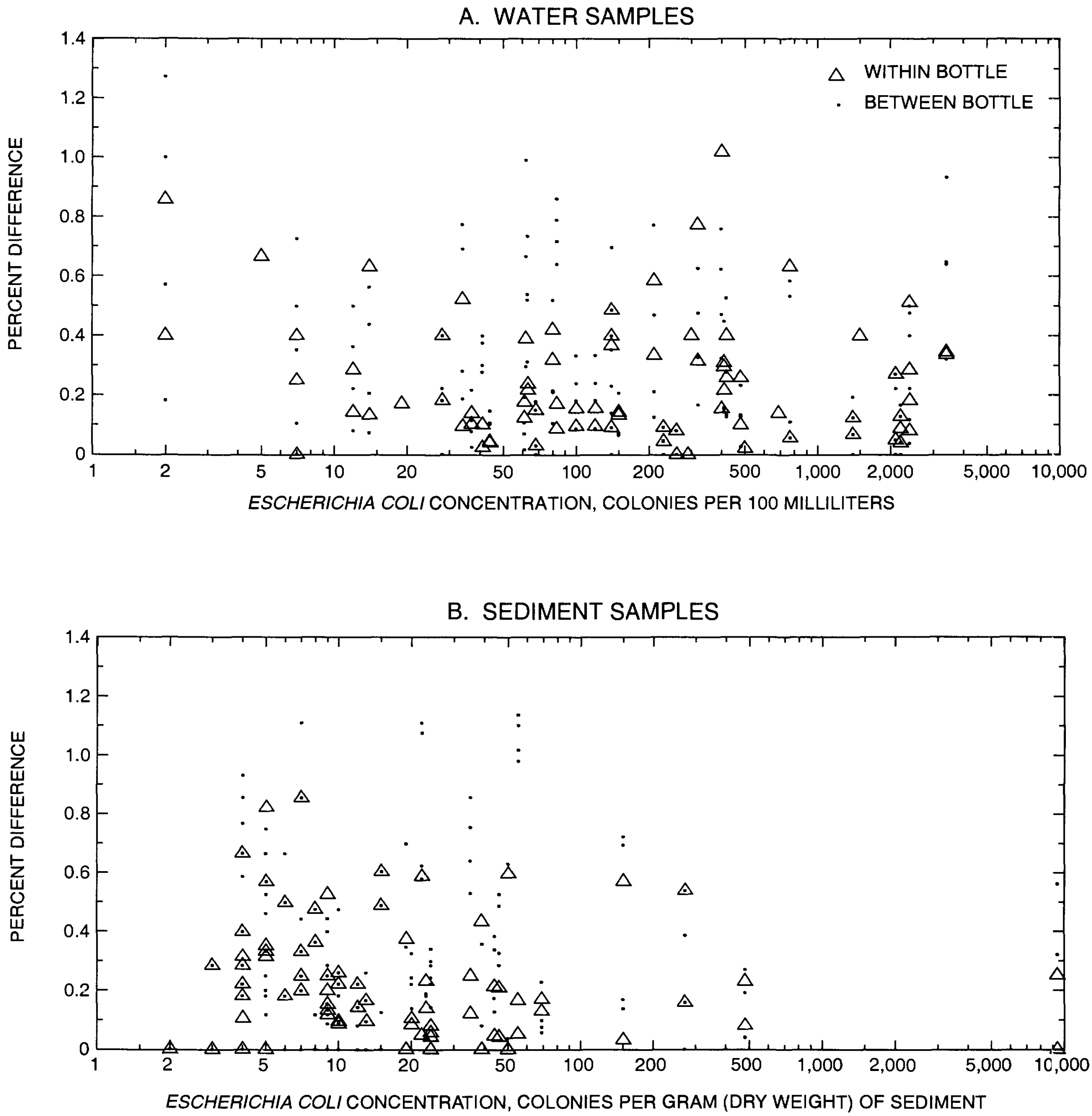

Figure B1. Percent differences (PD) between quality-control replicate samples for concentrations of *Escherichia coli* in (A) water and (B) sediment.

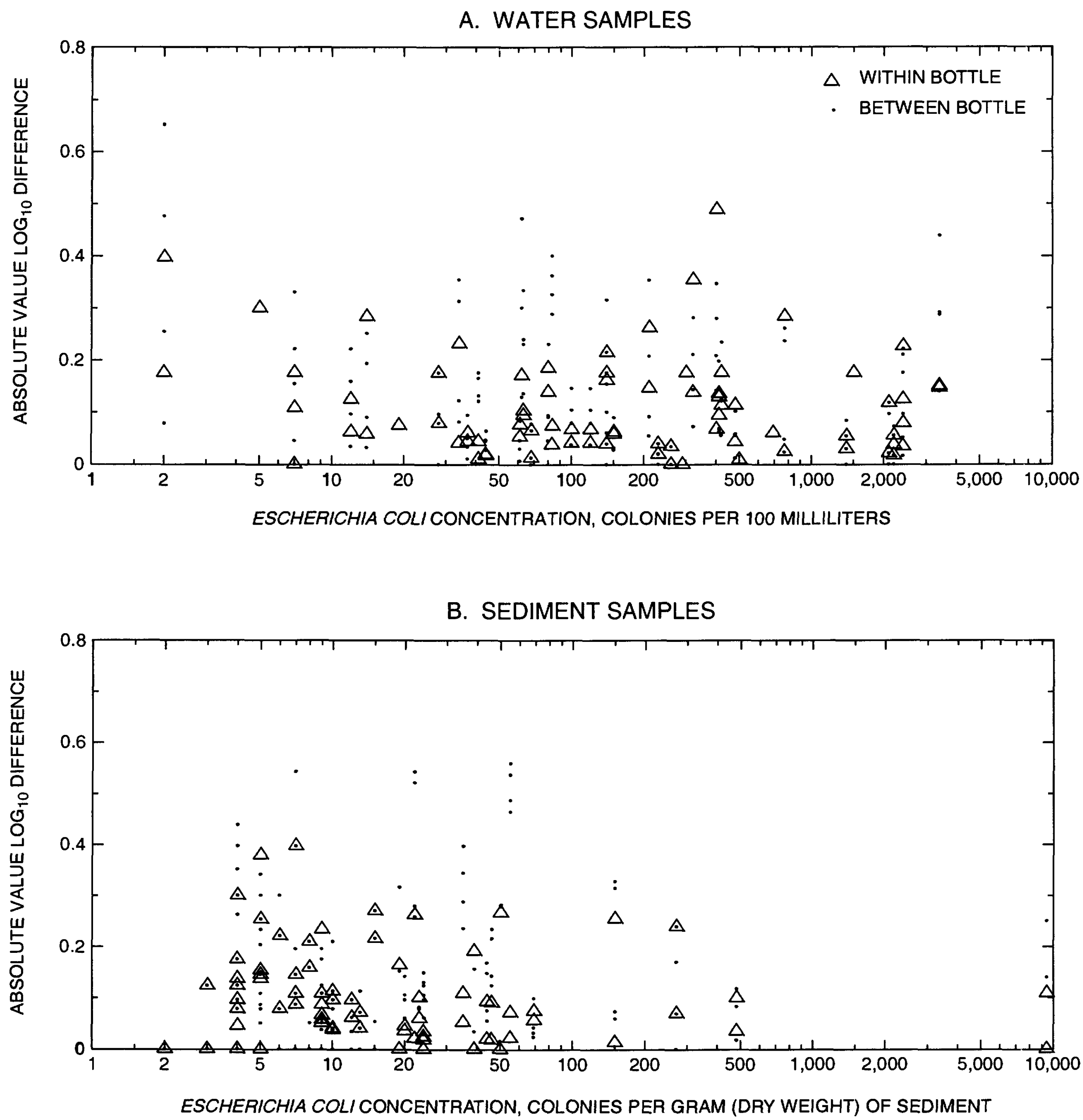

Figure B2. Absolute value $\log_{10}$ differences (AVLD) between quality-control replicate samples for concentrations of *Escherichia coli* in (A) water and (B) sediment.

Table B1. Summary statistics for within-bottle and between-bottle differences of replicate quality-control samples for concentrations of *Escherichia coli*. (Results of Tukey's test on the ranks are presented as letters, and medians with at least one letter in common do not differ significantly at $\alpha = 0.05$)

Variability type	Number of samples	Mean (standard deviation)	Median	Minimum	Maximum	95-percent quantile
Percent differences between sample pairs						
Water[a]						
Within bottle	81	24 (21)	17 A	0	100	63
Between bottle	142	29 (26)	21 A	0	130	77
Sediment[b]						
Within bottle	79	24 (21)	20 A	0	86	67
Between bottle	156	34 (28)	26 B	0	110	93
Absolute-value $\log_{10}$ differences between sample pairs						
Water[a]						
Within bottle	81	0.11 (.09)	0.07 A	0	0.49	0.28
Between bottle	142	0.13 (.12)	0.09 A	0	0.65	0.35
Sediment[b]						
Within bottle	79	0.11 (.09)	0.09 A	0	0.40	0.30
Between bottle	156	0.15 (.13)	0.12 B	0	0.56	0.44

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

Table B2. Statistical analysis of between-bottle and within-bottle differences of replicate quality-control samples grouped by magnitude of concentration and by beach. (Results of Tukey's test are presented as letters, and differences with at least one letter in common do not differ significantly at $\alpha = 0.05$)

Sample type	Data Set	Median percent differences between sample pairs		Median absolute-value $\log_{10}$ differences between sample pairs	
		Between bottle	Within bottle	Between bottle	Within bottle
		Magnitude of concentration			
Water[a]	<50	22 A	16 A	.10 A	.07 A
	50 - 235	21 A	16 A	.09 A	.07 A
	>235	17 A	22 A	.07 A	.09 A
Sediment[b]	<10	29 A	29 A	.12 A	.12 A
	10 - 50	24 A	13 A	.11 A	.06 A
	>50	30 A	16 A	.13 A	.07 A
		Beach			
Water[a]	Edgewater	18 A	15 A	.08 A	.07 A
	Villa Angela	25 A	22 A	.11 A	.09 A
	Sims	22 A	17 A	.10 A	.08 A
Sediment[b]	Edgewater	18 A	22 A	.08 A	.10 A
	Villa Angela	29 B	10 A	.12 B	.05 A
	Sims	39 B	22 A	.17 B	.10 A

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