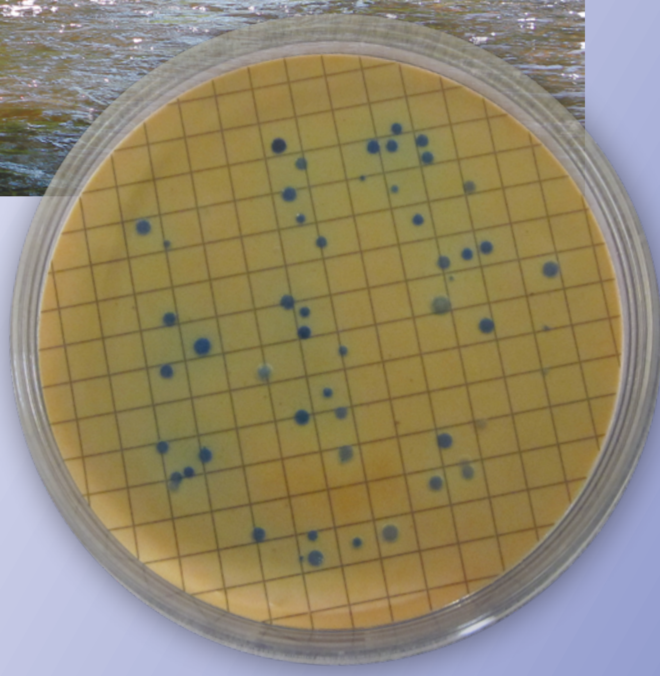
Prepared in cooperation with the Chester County Health Department and Chester County Water Resources Authority

Estimated Fecal Coliform Bacteria Concentrations Using Near Real-Time Continuous Water-Quality and Streamflow Data From Five Stream Sites in Chester County, Pennsylvania, 2007–16



Scientific Investigations Report 2017–5075
Version 1.2, March 2024

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

Cover. Brandywine Creek looking downstream towards U.S. Geological Survey gaging station 01481000 at Chadds Ford, Pennsylvania, on June 21, 2017. Inset shows growth on filter plate (after 24 hours of incubation) of fecal coliform bacteria colonies in 25 milliliters (mL) of stream water collected from Brandywine Creek at Chadds Ford, Pennsylvania, on June 21, 2017; resulting bacteria count is 204 colony forming units per 100 mL. Photographs by Andrew Reif, U.S. Geological Survey.

Estimated Fecal Coliform Bacteria Concentrations Using Near Real-Time Continuous Water-Quality and Streamflow Data From Five Stream Sites in Chester County, Pennsylvania, 2007–16

By Lisa A. Senior

Prepared in cooperation with the
Chester County Health Department and
Chester County Water Resources Authority

Scientific Investigations Report 2017–5075
Version 1.2, March 2024

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

U.S. Department of the Interior
RYAN K. ZINKE, Secretary

U.S. Geological Survey
William H. Werkheiser, Acting Director

U.S. Geological Survey, Reston, Virginia: 2017
Revised: April 2023 (ver. 1.1)
Revised: March 2024 (ver. 1.2)

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

For an overview of USGS information products, including maps, imagery, and publications, visit https://store.usgs.gov/.

Suggested citation:
Senior, L.A., 2017, Estimated fecal coliform bacteria concentrations using near real-time continuous water-quality and streamflow data from five stream sites in Chester County, Pennsylvania, 2007–16 (ver. 1.2, March 2024): U.S. Geological Survey Scientific Investigations Report 2017–5075, 46 p., https://doi.org/10.3133/sir20175075.

ISSN 2328-0328 (online)

Acknowledgments

The Chester County Water Resources Authority (CCWRA) and Chester County Health Department (CCHD) have supported collection of samples to determine fecal bacteria concentrations in Brandywine Creek in cooperation with U.S. Geological Survey (USGS) for more than 30 years as part of concern for stream water quality in Chester County, Pennsylvania. The contribution of Janet Bowers, Craig Thomas, Dani-Ella Betz, and Mark Chegwidden of CCWRA and Jeanne Casner, Kimberly Liao, and Matthew Skiljo of CCHD to the current study was essential and is gratefully acknowledged. Andrew Reif of USGS collected most of the fecal coliform bacteria data since 2005. Continuous measurements of turbidity and other water-quality data were made by numerous USGS personnel, including Patrick Anzman, David Galeone, Matthew Gyves, and Leif Olson.

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Length	
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
	Area	
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
	Volume	
cubic foot (ft^3)	28.32	cubic decimeter (dm^3)
cubic foot (ft^3)	0.02832	cubic meter (m^3)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$°F = (1.8 \times °C) + 32.$$

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

$$°C = (°F - 32) / 1.8.$$

Datum

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83) and the North American Datum of 1927 (NAD 27).

Supplemental Information

Concentrations of bacteria in water are given in number of colonies per 100 milliliters (col/100 mL) which are equivalent to units of colony forming units per 100 milliliters (cfu/100 mL).

Values for turbidity in water are given in Formazin Nephelometric Units (FNU).

Specific conductance is given in microsiemens per centimeter at 25 degrees Celsius (μS/cm at 25 °C).

Concentrations of chemical constituents in water are given in either milligrams per liter (mg/L) or micrograms per liter (μg/L).

Abbreviations

CFU	Colony forming unit
CCHD	Chester County Health Department
CCWRA	Chester County Water Resources Authority
EPA	U.S. Environmental Protection Agency
FNU	Formazin Nephelometric Units
NWIS	National Water Information System
PADEP	Pennsylvania Department of Enviromental Protection
USGS	U.S. Geological Survey

Estimated Fecal Coliform Bacteria Concentrations Using Near Real-Time Continuous Water-Quality and Streamflow Data from Five Stream Sites in Chester County, Pennsylvania, 2007–16

By Lisa A. Senior

Abstract

Several streams used for recreational activities, such as fishing, swimming, and boating, in Chester County, Pennsylvania, are known to have periodic elevated concentrations of fecal coliform bacteria, a type of bacteria used to indicate the potential presence of fecally related pathogens that may pose health risks to humans exposed through water contact. The availability of near real-time continuous stream discharge, turbidity, and other water-quality data for some streams in the county presents an opportunity to use surrogates to estimate near real-time concentrations of fecal coliform (FC) bacteria and thus provide some information about associated potential health risks during recreational use of streams.

The U.S. Geological Survey (USGS), in cooperation with the Chester County Health Department (CCHD) and the Chester County Water Resources Authority (CCWRA), has collected discrete stream samples for analysis of FC concentrations during March–October annually at or near five gaging stations where near real-time continuous data on stream discharge, turbidity, and water temperature have been collected since 2007 (or since 2012 at 2 of the 5 stations). In 2014, the USGS, in cooperation with the CCWRA and CCHD, began to develop regression equations to estimate FC concentrations using available near real-time continuous data. Regression equations included possible explanatory variables of stream discharge, turbidity, water temperature, and seasonal factors calculated using Julian Day with base-10 logarithmic (log) transformations of selected variables.

The regression equations were developed using the data from 2007 to 2015 (101–106 discrete bacteria samples per site) for three gaging stations on Brandywine Creek (West Branch Brandywine Creek at Modena, East Branch Brandywine Creek below Downingtown, and Brandywine Creek at Chadds Ford) and from 2012 to 2015 (37–38 discrete bacteria samples per site) for one station each on French Creek near Phoenixville and White Clay Creek near Strickersville. Fecal coliform bacteria data collected by USGS in 2016 (about nine samples per site) were used to validate the equations. The best-fit regression equations included log turbidity and seasonality factors computed using Julian Day as explanatory variables to estimate log FC concentrations at all five stream sites. The adjusted coefficient of determination for the equations ranged from 0.61 to 0.76, with the strength of the regression equations likely affected in part by the limited amount and variability of FC bacteria data. During summer months, the estimated and measured FC concentrations commonly were greater than the Pennsylvania Department of Environmental Protection established standards of 200 and 400 colonies per 100 milliliters for water contact from May through September at the 5 stream sites, with concentrations typically higher at 2 sites (White Clay Creek and West Branch Brandywine Creek at Modena) than at the other 3 sites. The estimated concentrations of FC bacteria during the summer months commonly were higher than measured concentrations and therefore could be considered cautious estimates of potential human-health risk. Additional water-quality data are needed to maintain and (or) improve the ability of regression equations to estimate FC concentrations by use of surrogate data.

Introduction

In Chester County in southeastern Pennsylvania, many streams provide opportunities for recreational activities, such as swimming, fishing, and boating. However, the periodic presence of elevated levels of fecal coliform (FC) bacteria in some streams indicates possible elevated human-health risks associated with exposure to potential pathogens in these waters. Fecal coliform bacteria are a group of bacteria found in the digestive systems of warm-blooded animals and are one of several types of fecal-indicator bacteria (FIB) that are indicative of the potential presence of feces and associated pathogens (disease-causing organisms). In order to protect human health during recreational use of streams and other waterways, standards for the presence of FC and other indicator bacteria have been established. In Pennsylvania, the FIB water-contact standards set by the Pennsylvania Department of Environmental Protection (PADEP) apply to FC concentrations in the water column, with standards specifying lower concentrations for the swimming season (May 1–September 30) than for the remainder of the year. One of the most common methods to determine FC concentrations involves membrane filtration of a grab sample and subsequent incubation for 24 hours (National Environmental Methods Index, 2016). However, this method introduces at least a 1-day delay in assessing bacteria concentrations. In addition, because of the type of sample (grab), the method is limited to determining bacteria concentrations at a single specific point in time at the sampling location. Interest in developing near real-time estimates of FIB concentrations has resulted in studies that evaluated bacteria surrogates, such as turbidity, that can be measured near real time in situ. Near real-time estimates of FIB concentrations can be used to improve protection of human health from exposure to potential pathogens during recreational use of surface waters. Regression equations that incorporate turbidity as a variable have been used to estimate bacteria concentrations for beaches along Lake Erie (Francy and others, 2013; Zimmerman, 2008) and in streams in Kansas (Rasmussen and Ziegler, 2003) and Georgia (Lawrence, 2012).

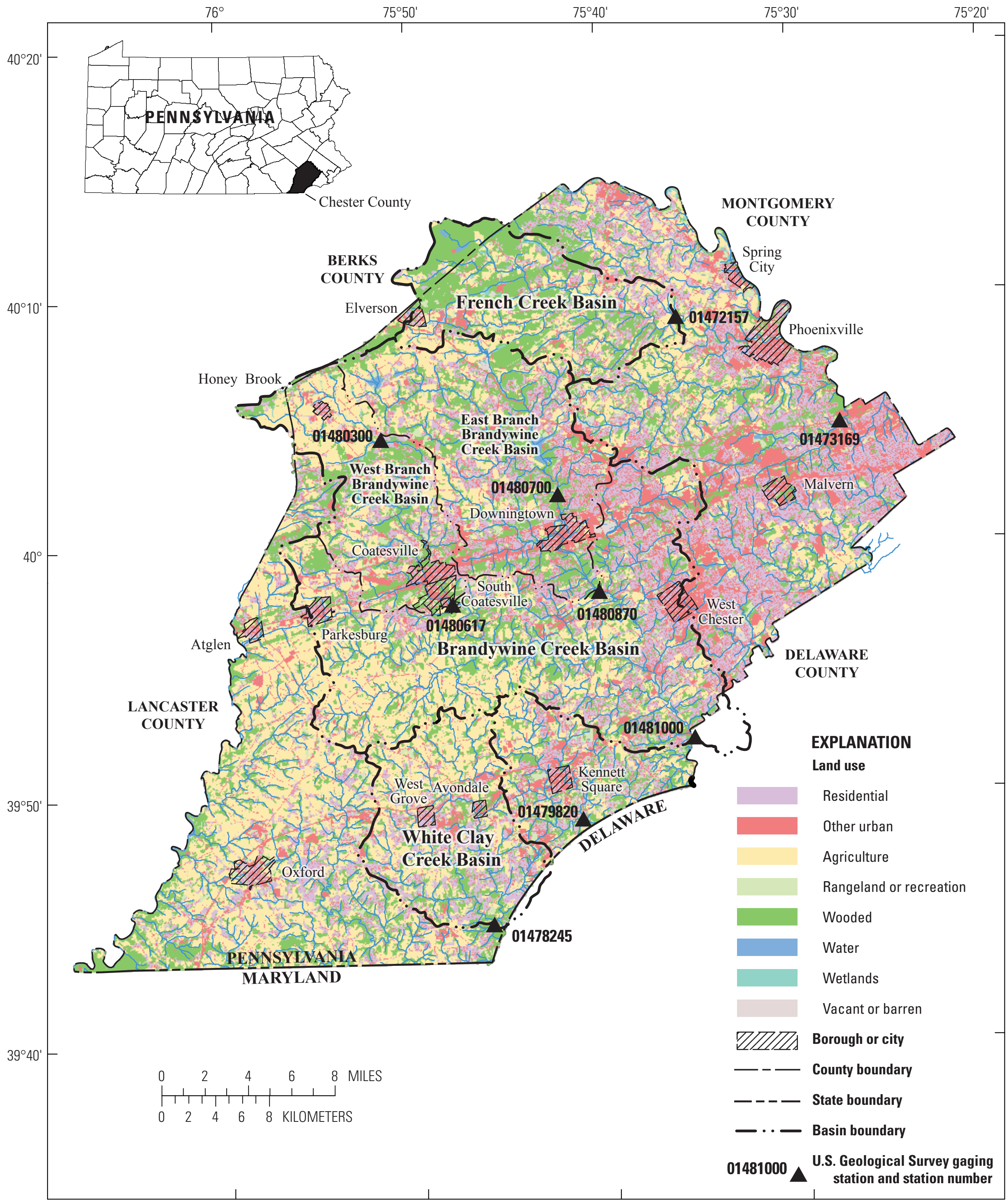
In Chester County, recent data are available on FC bacteria concentrations in discrete grab samples and near real-time continuous turbidity at 5 sites with gaging stations on streams used for recreational activities, including 3 sites on Brandywine Creek, 1 site on French Creek, and 1 site on White Clay Creek (fig. 1). Discrete grab samples for FC bacteria analysis at these five stream sites have been collected by the U.S. Geological Survey (USGS), in cooperation with the Chester County Water Resource Authority (CCWRA) and Chester County Health Department (CCHD). Collection of the earliest discrete bacteria data began in the 1970s at the three Brandywine Creek sites and in 2012 for the other two stream sites (table 1). Near real-time continuous turbidity and other water-quality data currently (2016) are collected by USGS, in cooperation with the CCWRA, CCHD, and City of Wilmington, Delaware, using sensors installed at these 5 and 4 other gaging stations. Thus, as of 2016, sufficient data are available to allow development of regression equations to estimate near real-time bacteria concentrations at 5 sites but with additional FC data collection, regression equations could be developed to estimate near real-time FC concentrations at the other 4 gaging stations. The earliest continuous turbidity record for the Brandywine Creek sites began in 2005, and the record for the White Creek and French Creek sites began in 2011 and 2006, respectively (table 1).

The availability of data and interest in improving health-related guidance for recreational use of the streams in Chester County made possible the development of regression equations for the estimation of near real-time FC concentrations so that the public might be better informed about possible actions to take to reduce chances of illness or infection. A preliminary unpublished evaluation of the use of surrogates, such as turbidity, to estimate FC concentrations for the three Brandywine Creek sites for the period 2007–10 was conducted by USGS for CCHD and CCWRA in 2012. Results of the preliminary evaluation indicate that turbidity and water temperature are better estimators of FC bacteria than turbidity alone. On the basis of these preliminary results and availability of data, it appeared that turbidity along with other water-quality properties might be used to estimate near real-time concentrations of FC bacteria and, thus, provide the public and health officials with an immediate indicator of potential health risk associated with recreational use of the selected streams in Chester County. In 2014, the USGS, in cooperation with the CCHD and CCWRA, initiated a more formal study to develop regression equations for the estimation of near real-time FC concentrations from available near real-time continuous data (collected at 15- to 30-minute intervals and transmitted hourly). On further analysis, inclusion of variables other than turbidity and water temperature appeared likely to improve the strength of the regression-based approach to estimating bacteria concentrations. However, additional data beyond that collected through 2014, especially for the two stream sites with the discrete bacteria record beginning in 2012, are thought to be needed in order to develop statistically significant regression equations.

Recreational Water Standards and Criteria for Fecal Indicator Bacteria

In Pennsylvania, the water-contact standards for FC concentrations in the water column established by the PADEP are 200 coliforms per 100 milliliters (coliforms/100 mL) during the swimming season (May 1–September 30) and 2,000 coliforms/100 mL during the remainder of the year (Pennsylvania Department of Environmental Protection, 025 Pa. Code § 93.7, accessed May 31, 2012, at http://www.pacode.com/secure/data/025/chapter93/s93.7.html). The PADEP standards of 200 or 2,000 coliforms/100 mL are defined as the geometric mean of FC concentrations in a minimum of five consecutive samples collected on different days within a 30-day period with no more than 10 percent of the FC concentrations exceeding 400 coliforms/100 mL in a 30-day period during the swimming season (May 1–September 30) (Pennsylvania Department of Environmental Protection, 025 Pa. Code § 93.7). The units of coliforms/100 mL used by PADEP are equivalent to units of colony forming units per 100 milliliters (cfu/100 mL) used by USGS and the U.S. Environmental Protection Agency (EPA).

The EPA has recommended the use of enterococci or *Escherichia coli (E. coli)* bacteria as indicators of fecal contamination in fresh water (U.S. Environmental Protection Agency, 2012) and has defined recommended criteria that differ somewhat from the regulatory standards set by PADEP. The EPA recommended recreational water-quality



Base from Pennsylvania Department of Transportation, 2007
Hydrology from National Hydrography Dataset, 2016

Figure 1. Location of gaging stations with continuous turbidity data collected as early as 2005 and through 2016 by the U.S. Geological Survey and 2005 land use in Chester County, Pennsylvania.

Table 1. Drainage area and periods of record for continuous streamflow, turbidity, and other water-quality data and discrete-sample fecal coliform bacteria concentration data collected by the U.S. Geological Survey at or near gaging stations on selected streams in Chester County, Pennsylvania, as of September 2016.

[Location of gaging stations shown in figure 1 identified by U.S. Geological Survey (USGS) station number; mi^2, square miles; SC, specific conductance; DO, dissolved oxygen; nd, no data]

USGS station number	Station name (stream site)	Drainage area (mi^2)	Period of continuous streamflow and water-quality data						Period of bacteria data (discrete samples)
			Streamflow	Turbidity	Water temperature	pH	SC	DO	
	Sites used for regression analysis to estimate bacteria concentrations								
01480617	West Branch Brandywine Creek at Modena, Pa.	55.0	1/1970–current	9/2005–current[1]	5/1971–10/1977; 8/1981–current	5/1971–10/1977; 8/1981–current	5/1971–10/1977; 8/1981–current	5/1971–10/1977; 8/1981–current	1981–2010; 3/2012–current
01480870	East Branch Brandywine Creek below Downingtown, Pa.	89.9	2/1972–current	10/2005–current[2]	2/1972–current	2/1972–current	2/1972–current	2/1972–current	1978–2010; 3/2012–current
01481000	Brandywine Creek at Chadds Ford, Pa.	287.0	8/1911–9/1953; 10/1962–current	10/2005–current[3]	10/1964–current	10/1965–9/1966; 12/1971–current	10/1965–current	10/1971–current	1976–2010; 3/2012–current
01472157	French Creek near Phoenixville, Pa.	59.1	10/1968–current	11/2006–current[4]	11/1998–4/1999; 6/2000–9/2001	nd	nd	nd	3/2012–current
01478245	White Clay Creek near Strickersville, Pa.	59.2	8/1996–current	11/2011–current[5]	nd	nd	nd	nd	3/2012–current
	Selected other sites with partial data for turbidity, water temperature, or bacteria								
01479820	Red Clay Creek near Kennett Square, Pa.	28.3	1/1988–current	12/2015–current	nd	nd	nd	nd	3/2012–10/2012
01473169	Valley Creek at Pa. turnpike near Valley Forge, Pa.	20.8	10/1982–current	10/2006–current	nd	nd	nd	nd	nd
01480300	West Branch Brandywine Creek near Honey Brook, Pa.	18.7	6/1960–current	11/2006–current	nd	nd	nd	nd	7/1998–9/1999
01480700	East Branch Brandywine Creek near Downingtown, Pa.	60.6	10/1965–current	11/2010–current	nd	nd	nd	nd	7/1998–9/1999

[1]Turbidity sensors operated at West Branch Brandywine Creek at Modena, Pa.: Analite NEP365G from September 22, 2005 to October 1, 2006; another Analite NEP395G from October 1 to December 12, 2006; YSI 6136 from February 28, 2007 to current.

[2]Turbidity sensors operated at East Branch Brandywine Creek below Downingtown, Pa.: Analite NEP365G from October 12, 2005 to August 7, 2008; YSI 6136 from August 7, 2008 to current.

[3]Turbidity s`ensors operated at Brandywine Creek at Chadds Ford, Pa.: Analite NEP365G from October 12, 2005 to October 18, 2008; *In-Situ* Troll RDO from October 18, 2008 to April, 22, 2009; YSI 6136 from April 22, 2009, to current.

[4]Turbidity sensors operated at French Creek near Phoenixville, Pa.: YSI 6136 from November 7, 2006, to current.

[5]Turbidity sensors operated at White Clay Creek near Strickersville, Pa.: YSI 6136 from November 15, 2011, to current.

criteria for the protection of human health at a rate of 32 illnesses per 1,000 exposures are geometric mean values of 30 and 100 cfu/100 mL for enterococci and *E. coli*, respectively, with statistical threshold values of no more than 10 percent of the samples having concentrations greater than 110 and 320 cfu/100 mL, respectively. At this same level of risk, the EPA has set single sample Beach Action Values of 60 and 190 cfu/100 mL for enterococci and *E. coli*, respectively (U.S. Environmental Protection Agency, 2014).

For two stream sites in Chester County, Pa., some data are available that allow comparison of enterococci and FC concentrations to each other and to respective standards and criteria (fig. 2). The water samples at these two sites, Brandywine Creek at Chadds Ford, Pa., and White Clay Creek near Strickersville, Pa., have been collected bimonthly by USGS in cooperation with PADEP since 2005 as part of the State's Water Quality Network (WQN) monitoring program. The samples are analyzed by the PADEP laboratory within 24 hours of collection, and the results are downloaded to, and are available from, the USGS National Water Information System (NWIS). Nearly all the bimonthly samples were analyzed for FC bacteria, and about one-half were analyzed for enterococci bacteria. At both stream sites, FC and enterococci concentrations are correlated, and the relations indicate that at FC concentrations equal to or greater than 200 cfu/100 mL (the PADEP standard for the swimming season of May through September), the enterococci concentrations are equal to or greater than 110 cfu/100 mL, the EPA recommended recreational water-quality statistical threshold value (criteria) for enterococci.

Purpose and Scope

This report presents regression equations developed using available near real-time continuous stream discharge, water-quality data (collected at 15- to 30-minute intervals and transmitted hourly), including turbidity and water temperature, and discrete sample bacteria concentrations (fecal coliform) to estimate continuous FC concentrations at five streams sites at or near gaging stations in Chester County, Pennsylvania, 2007–15. The data and methods used to develop the regression equations are described. Regressions were developed using data for 2007–15 for 3 stream sites on Brandywine Creek (West Branch Brandywine Creek at Modena, East Branch Brandywine Creek below Downingtown, and Brandywine Creek at Chadds Ford) and data for 2012–15 for 2 other stream sites (French Creek near Phoenixville and White Clay Creek near Strickersville). Data collected in 2016 were used to validate regression models.

The results of the regression equations, estimated continuous (15- to 30-minute interval) FC concentrations, are compared to established water-quality standards for recreational use of streams to demonstrate how the near real-time regression results may be used by water managers, health officials, and the public to gain more information about potential human-health risks associated with recreational use of those waters.

Description of Study Area

The study area, which includes the Brandywine Creek, French Creek, and White Clay Creek Basins in Chester County, is a 760-square mile (mi^2) area in southeastern Pennsylvania (fig.1). The county lies in the Piedmont Physiographic Province and is characterized by rolling hills principally underlain by deeply weathered crystalline rocks (such as schist, gneiss, and metasediments) and, to a lesser extent, sedimentary rocks. Precipitation locally recharges the fractured-rock aquifers that supply base flow to the streams, and although precipitation falls approximately evenly throughout the year, recharge (and base flow) tends to be lowest in late spring through mid-autumn when evapotranspiration rates are greatest. The region has a humid continental climate with warm to hot summers and cool to cold winters (Peel and others, 2007).

The Brandywine, White Clay, and French Creeks are used for fishing, swimming, and boating, most commonly from spring to fall. The Brandywine and White Clay Creeks are also used for water supply, and disposal and assimilation of treated sewage effluent. The USGS operates gaging stations and water-quality monitors at various stream sites in Chester County, including the five with data evaluated for this study: West Branch Brandywine Creek at Modena, Pa. (USGS station 01480617), East Branch Brandywine Creek below Downingtown, Pa. (USGS station 01480870), Brandywine Creek at Chadds Ford, Pa (USGS station 01481000; site on main stem downstream from confluence of the East and West Branches), French Creek at Phoenixville, Pa. (USGS station 01472157), and White Clay Creek near Strickersville, Pa. (USGS station 01478245) (fig. 1; table 1). Drainage areas above these five gaging stations range from 55 to 287 mi^2 (table 1). The streams are largely free-flowing riffle-pool sequences of reaches upstream from the gaging stations, although the upper East Branch Brandywine Creek and two of its tributaries (Marsh Creek and Beaver Creek) and two tributaries to the West Branch Brandywine Creek (Birch Run and Rock Run) have been dammed.

The county is undergoing urbanization with increases in residential and commercial land use in some areas, especially in the central and eastern parts of the county, but is rural in character elsewhere, especially in the western parts of the county. Major crops are hay and corn; in selected areas, mushroom farming is common. Dairy and horse farms are also common in parts of the county.

Land use varies in each of the study basins (fig. 1; table 2). The French and White Clay Creek Basins are the least urbanized of the five basins but differ in amounts of other land uses. The French Creek Basin has the most forested area; development is mostly in the eastern part of this basin. The White Clay Creek Basin has the most agricultural area. Much of the upper part of White Clay Creek is agricultural. Mushroom growing and the related activity of composting are widespread in the eastern and central parts of the White Clay Creek Basin. Land use in the Brandywine Creek Basin is more

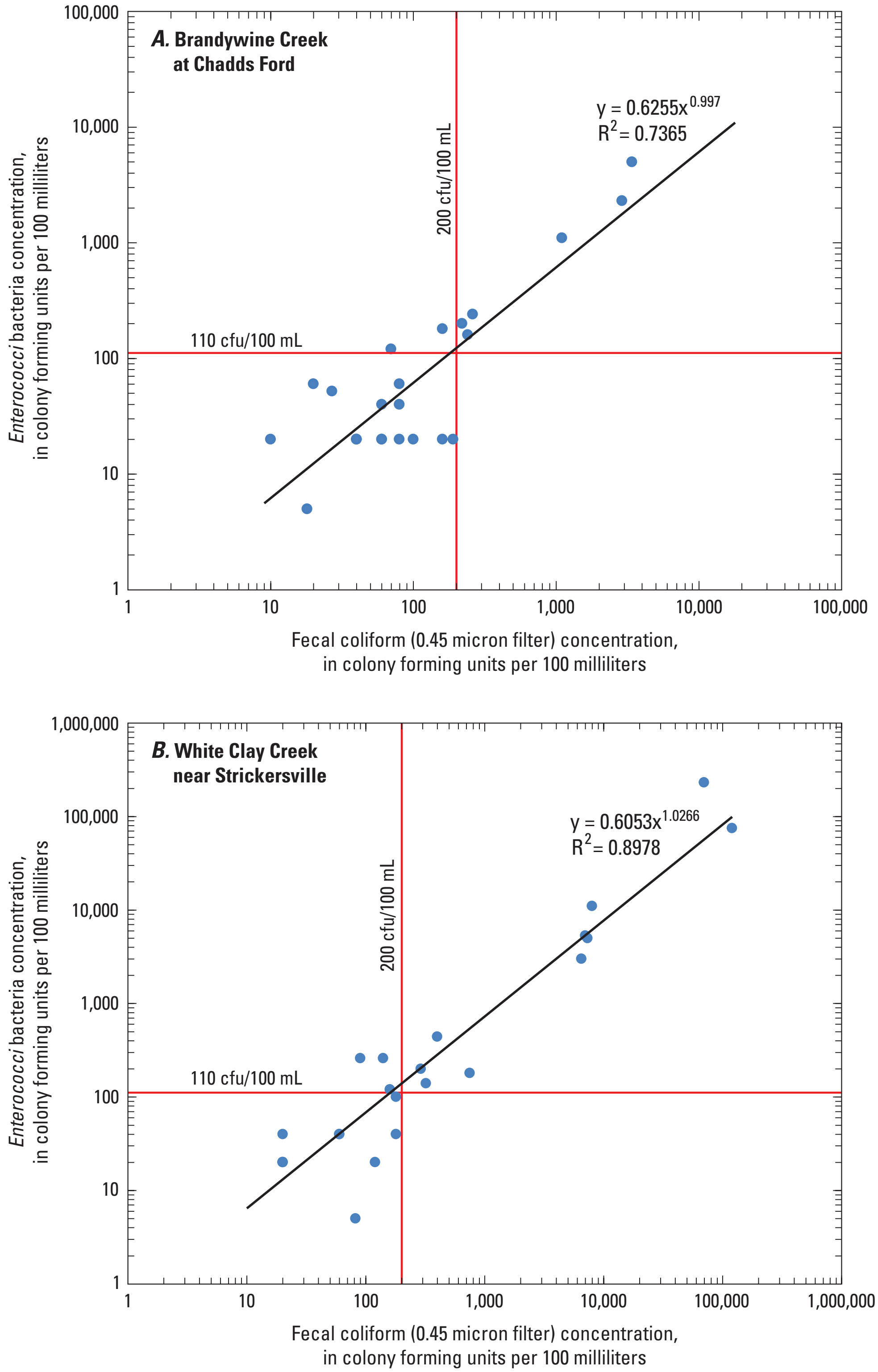


Figure 2. Relation between concentrations of fecal coliform and enterococci bacteria in discrete stream samples collected bimonthly by the U.S. Geological Survey near gaging stations at *A*, Brandywine Creek at Chadds Ford, Pa., and *B*, White Clay Creek near Strickersville, Pa., 2005–15. Bacteria concentrations were determined at the Pennsylvania Department of Environmental Protection laboratory in Harrisburg, Pa. (cfu/100 mL, colony forming units per 100 milliliters)

Table 2. Land use in drainage areas upstream from five gaging stations on Brandywine, French, and White Clay Creeks in Chester County, Pennsylvania where fecal coliform bacteria, turbidity, and other water-quality data were collected by the U.S. Geological Survey, as of September 2016. Location of stream sites shown in figure 1 identified by USGS station number.

[Location of gaging stations shown on figure 1; USGS, U.S. Geological Survey; mi^2, square miles]

USGS station number	Station name (stream site)	Drainage area (mi^2)	Land use (percent of basin area upstream from gaging station)					Date of and reference for land use
			Forested and open	Agriculture and pasture	Urbanized[1]	Water	Total[2]	
01480617	West Branch Brandwyine Creek at Modena, Pa.	55.0	27.0	37.4	25.6	3.0	93.0	2005, Sloto and Olson (2011)
01480870	East Branch Brandywine Creek below Downingtown, Pa.	89.9	27.7	29.8	31.9	4.5	93.9	2005, Sloto and Olson (2011)
01481000	Brandywine Creek at Chadds Ford, Pa.	287.0	33.7	41.7	22.1	1.1	98.6	1995, Senior and Koerkle (2003a)
01472157	French Creek near Phoenixville, Pa.	59.1	44.6	33.0	16.1	3.6	97.3	2005, Sloto and Olson (2011)
01478245	White Clay Creek near Strickersville, Pa.	59.2	27.8	54.6	16.1	0.5	99.0	1995, Senior and Koerkle (2003b)

[1] Includes residential, commercial, industrial, and transportation (roads) land uses.

[2] Percentage of land uses may not sum to 100 because small areas of other land uses were excluded, in addition to differences due to rounding.

mixed than in French and White Clay Creek Basins. The upper parts of the East Branch and West Branch of Brandywine Creek include relatively large amounts of agricultural land, with the West Branch having more agricultural land than the East Branch. Both the East and West Branches of Brandywine Creek flow south to traverse the highly urbanized Great Valley (underlain by carbonate rocks) and population centers of Coatesville and Downingtown (fig. 1). Downstream from the confluence of the East and West Branches, the land use along Brandywine Creek is less urbanized and includes mixed rural residential, forested, and agricultural areas. Numerous municipal and privately operated wastewater treatment plants of various sizes discharge to the streams in the White Clay Creek and Brandywine Creek Basins. As noted by Town (2001) and Cinotto (2005), FC in streams potentially may be derived from both point and non-point sources in addition to wildlife and stored sediments within the stream channel.

Previous Investigations

A review of historical (1973–99) FC data collected by USGS in the Brandywine Creek Basin indicated downward trends in bacteria concentrations from 1980 to 1987 as a probable result of wastewater-treatment plant upgrades, decreases in point-source discharges, and decreases in agricultural land use (Town, 2001). Evaluation of the spatial and temporal distribution of FC bacteria concentrations during 1998–99 in the Brandywine Creek Basin indicated that bacteria concentrations generally were higher during elevated streamflow conditions than during base-flow conditions, higher during the warm summer months than during the cooler spring and fall months, and lower in streams downstream from reservoirs than in the streams feeding the reservoirs (Town, 2001). A 2005 USGS study of bacteria occurrence in the West Branch Brandywine Creek near the City of Coatesville in Chester County showed that bacteria were strongly associated with stream sediments and discussed the possible use of turbidity (as a surrogate for suspended sediment) to estimate bacteria concentrations (Cinotto, 2005). Use of continuous turbidity data to estimate suspended sediment concentrations through relations determined by linear regression has been done recently for selected stream sites in Chester County, including two on Brandywine Creek (Sloto and Olson, 2011). A study of potentially pathogenic bacteria and microbial source tracking markers done in the Brandywine Creek Basin during 2009–10 found that increases in frequency of pathogenic bacteria and markers linked to bovine sources were associated with high flows in the West Branch and main stem of Brandywine Creek but not in the East Branch Brandywine Creek (Duris and others, 2011). This study further indicated that FIB and fecally derived pathogens were present in both the East Branch and West Branch Brandywine Creek, and microbial markers indicated human sources for some bacteria samples collected throughout the Brandywine Creek Basin.

Methods

Site Selection

Five sites on streams used for recreation [fishing, boating, and (or) swimming] that have continuous streamflow and turbidity data and a new or continuing record of discrete bacteria data were selected for regression analysis. All five sites are at or near existing USGS gaging stations, including 3 sites on Brandywine Creek, 1 site on French Creek and 1 site on White Clay Creek (fig. 1; table 1). Four sites on other streams in Chester County that currently have near real-time continuous turbidity data but limited or no discrete bacteria data (table 1) are potential candidates for regression analysis should additional bacteria data become available. Of the 5 stream sites selected for analysis, regression relations are expected to be more representative of the range of possible hydrologic conditions at the 3 sites on Brandywine Creek than at the sites on French Creek and White Clay Creek because the Brandywine Creek sites have longer and denser periods of record for continuous water-quality and discrete bacteria data (table 1).

Data Collection

Data on both continuous turbidity and discrete bacteria concentrations have been available only since 2005 when turbidity sensors were installed at the three Brandywine Creek gaging stations. Turbidity sensors were installed at the French Creek and White Clay Creek gaging stations in November 2006 and 2011, respectively (fig. 1; table 1). However, because of changes in instrumentation, which is discussed in the section “Continuous Turbidity, Other Water-Quality, and Stream Discharge Data,” turbidity data collected since 2007 are more consistent and complete than earlier data and were selected for regression analysis.

USGS has determined FC concentrations in discrete grab samples of stream water at the three gaging stations on Brandywine Creek (fig. 1) from the 1970s until fall 2010 (table 1) in cooperation with the CCHD and CCWRA. During the first half of this period through 1996, samples for FC analysis were collected by USGS about once a week from at least March through November each year. From 1997 to 2010, USGS collected samples 2 to 4 times a month from March through September, when both recreational use of streams and bacteria concentrations tend to be highest.

In spring 2012, the USGS resumed seasonal (March–October) monthly collection of FC bacteria samples at the three gaging stations on Brandywine Creek and, to provide some data where none were previously available, started seasonal monthly collection at or near three other gaging stations on streams with recreational uses: White Clay Creek, Red Clay Creek, and French Creek (fig.1; table 1). Sample collection for bacteria analysis was discontinued at Red Clay Creek after 2012. Since spring 2013, limited seasonal sampling of about 9 samples per year (generally 1 sample per month plus up to 3 additional samples from March to October, with sampling ending in September during 2015–16) has continued at the five stream sites (3 on Brandywine Creek and 1 each on White Clay and French Creeks) through 2016.

Continuous Turbidity, Other Water-Quality, and Stream Discharge Data

The USGS operates turbidity sensors at gaging stations on Brandywine Creek (USGS stations 01480617, 01480870, and 01481000), French Creek (USGS station 01472157), and White Clay Creek (USGS station 01478245), where continuous data (15- to 30-minute intervals) on gage height (stage) and computed stream discharge also are collected using standard methods (Sauer and Turnipseed, 2010; Turnipseed and Sauer, 2010). In addition, the USGS operates multi-parameter sondes at the three gaging stations on Brandywine Creek to provide continuous data on water temperature, specific conductance, pH, and dissolved oxygen concentrations collected using standard methods (Wagner and others, 2006). At the French and White Clay Creek gaging stations, water temperature is measured only as part of the turbidity instrumentation. These water temperature data were not formally reviewed before October 2016 and therefore should be considered provisional for the period 2012–16. Checks on the provisional water temperature data collected at French and White Clay Creeks are discussed in the section “Quality Assurance and Quality Control.” All gage height, computed discharge, turbidity, and water-quality data, except for the French Creek and White Clay Creek provisional water temperature data, are available in NWIS. The French Creek and White Clay Creek water temperature data for period starting October 1, 2016, will be available in NWIS.

The continuous turbidity data were collected by USGS in accordance with the maintenance and calibration protocols described by Wagner and others (2006). The Yellow Spring Instrument (YSI) Optical Monitoring System 600 series (YSI 6136) turbidity sensors and YSI 6920V2 sondes have been used since April 2009 at all five sites (Sloto and Olson, 2011). Initially, McVan Analite NEP395 turbidity sensors were installed in autumn 2005 at the three Brandywine Creek stations and operated until replaced by other sensors during 2007–09 (table 1). At East Branch Brandywine Creek below Downingtown, the Analite sensor was replaced on August 7, 2008, by the YSI 6136 turbidity sensor. At West Branch Brandywine Creek at Modena, the Analite sensor was replaced on October 1, 2006, by an NEP395 sensor, which operated until December 12, 2006, and was replaced on February 28, 2007, by the YSI 6136 turbidity sensor. At Brandywine Creek at Chadds Ford, the Analite sensor was replaced on October 18, 2008, by an In-situ Troll turbidity sensor, which operated until April 22, 2009, when it was replaced by the YSI 6136 turbidity sensor. YSI turbidity sensors were installed on November 7, 2006, at French Creek near Phoenixville and on November 15, 2011, at White Clay Creek near Strickersville. Detailed descriptions of turbidity instrumentation, calibration, monitoring, and quality-assurance procedures for stations operated by USGS in Chester County are given in Sloto and Olson (2011).

For the period of data evaluated in this study, March 2007 through September 2016 (regression data 2007 through 2015 and validation data from 2016), potential differences caused by the use of different instrumentation to measure turbidity at two sites—East Branch Brandywine Creek below Downingtown and Brandywine Creek at Chadds Ford—for part of the time (1.5 and 2 years of the 8-year period, respectively) may introduce additional variability in any apparent relations between measured turbidity and bacteria concentrations, although this additional variability likely is small. Potential differences between turbidity measured using the Analite and YSI sensors at gaging stations in Chester County were evaluated by Sloto and Olson (2011) who found that the change in sensors introduced no bias to the turbidity data relative to use in regressions. The Analite NEP395G sensor has a range of 1 to 1,000 Formazin Nephelometric Units (FNU) with an accuracy of +/- 0.2 FNU, and turbidity data from the Analite sensors were recorded in whole numbers (Sloto and Olson, 2011). The YSI 6136 sensor has a range of 1 to 1,000 FNU with an accuracy of the greater of +/- 0.3 FNU or 2 percent of measurement, and turbidity data from the YSI sensors were recorded to 0.1 FNU. Turbidity data were collected using only one type of instrumentation (YSI 6136 turbidity sensors) for the other three sites for their entire respective periods of evaluation (March 2007–September 2016 for West Branch Brandywine Creek at Modena and March 2012–September 2016 for French Creek near Phoenixville and White Clay Creek near Strickersville).

Prior to October 2012, turbidity data were recorded at 30-minute intervals. Since October 2012, turbidity data have been recorded at 15-minute intervals at all of the gaging stations, except at USGS station 01480617 West Branch Creek at Modena, where turbidity has been recorded at 30-minute intervals from 2007 through September 2016. To obtain the closest turbidity measurement at the time of discrete sample collection (which commonly was recorded as occurring to the nearest 15 minutes but sometimes to the nearest 5 or 10 minutes), turbidity values were in some cases estimated by linear interpolation between recorded values.

Discrete Bacteria Sample Collection and Analysis

Discrete (grab) samples for bacteria analysis are collected by USGS personnel by wading at the midpoint of stream sections near the gaging stations at each site. Samples are collected from the stream adjacent to the gaging stations at three sites (West Branch Brandywine Creek at Modena, East Branch Brandywine Creek below Downingtown, and French Creek near Phoenixville) and at distances up to about 1,100 feet (ft) from the gaging stations at two sites (about 600 ft upstream from the gaging station on Brandywine Creek at Chadds Ford and about 1,100 ft downstream from the gaging station on White Clay Creek near Strickersville). At the time of discrete sample collection, water temperature and other stream chemical and physical properties (pH, specific conductance, dissolved oxygen concentration) were measured at the mid-channel location where samples were collected.

Stream depths at mid-channel at all five sites generally are less than 2 ft at the time of sample collection and commonly less than 1 ft. Generally, the streams at sample collection sites are considered to be vertically well mixed because all sites are downstream from rifles. Four of the five streams are relatively laterally well mixed at sample collection sites, as documented by water-quality data measured across the stream channel at the time of sample collection for bacteria analysis, periodically as part of quality-assurance procedures to support water-quality monitors at the gaging stations, and as part of other sampling programs. The stream at the sample collection site on White Clay Creek is less well mixed than at the other four stream sites. Stream water at the White Clay Creek sampling site on the main stem about 1,700 ft downstream from the confluence of the two branches of the creek has specific conductance that consistently increases by as much as 15 percent across the stream from the right to left banks, likely reflecting incomplete mixing of water from the two branches (that drain 33.5 and 25.5 mi^2, respectively). However, the specific conductance measured at mid-channel at the White Clay Creek site is typically within 10 percent or less of the specific conductance of composite equal-width increment samples collected under standard protocols, indicating that mid-channel samples where bacteria are collected are representative of composite samples, and thus, are collected in a relatively well mixed section of the stream channel.

The water samples for FC bacteria analysis are chilled until processed in the laboratory at the USGS office in Exton, Pa., within 6 hours of sample collection. The number of FC colonies in a sample is determined by membrane filtration using a 0.7-micron filter and subsequent plating and incubation, as described by Myers and others (2014) and Town (2001). At the laboratory, a range of dilutions is plated for each stream-water sample to obtain optimal counts (20–60 colonies) on at least one plate. Aliquots of 1 to 20 milliliters (mL) of sample water added to sterile phosphate-buffered water for filtration and plating typically yield optimal counts for most samples, except those collected at high flows or in midsummer, especially at two sites (West Branch Brandywine Creek at Modena and White Clay Creek near Strickersville) for which aliquots as small as 0.1 mL yield optimal counts (Andrew Reif, USGS, oral commun., 2016). During the cooler months of spring and fall, optimal counts occur on sample plates with less dilution than during the warmer summer months when FC bacteria concentrations typically are highest. Plates are counted within 22–24 hours after plating; results from plates with optimal counts are recorded in units of cfu/100 mL and entered into the USGS NWIS database. If no plate has an optimal count, the FC concentration is determined by summing counts from plates with less than the ideal number of colonies and dividing by the total sample volume for those plates.

Prior to 2011, samples were collected 2–4 (usually 3) times per month at the three Brandywine Creek stations at fixed time intervals from March to October each year, regardless of hydrologic conditions. No bacteria data were collected in 2011. Collection of bacteria data resumed on a limited basis in March 2012 at the three Brandywine Creek stations, and collection began at French and White Clay Creeks, with about 9 samples collected from March through October each year during 2012–14 and from March to September each year during 2015–16. The 9 samples included 6 to 7 samples collected in a fixed time interval (1 sample per month), regardless of hydrologic conditions, and as many as 3 additional samples collected under targeted specific flow conditions during the 6- to 8-month sampling period. From 2012 through 2014, the 2 or 3 additional targeted samples were collected under relatively high-flow conditions associated with rainfall events during the 6- to 8-month sampling period each year. Following a review of data collected through 2014, the USGS determined that bacteria concentrations during some hydrologic conditions (post-storm recessions) were not well represented in the dataset; consequently, the collection scheme was modified to target post-storm recession flows during 2015–16 instead of any relatively high flow for the two to three targeted samples.

Quality Assurance and Quality Control

Stream sections were typically well mixed, as documented by periodic stream water-quality (pH, temperature, specific conductance, and dissolved oxygen) cross sections, and turbidity cross-section data indicated that turbidity measured at the sensor locations was representative of the stream section, as described by Sloto and Olson (2011). Cross-section water-quality data are archived with other USGS field notes for the gaging stations.

Bacteria concentrations are verified by comparing counts in various dilutions of the same stream-water sample. Blanks run on filtration equipment between samples indicate no cross contamination related to sample processing in the laboratory.

Water temperatures measured in mid-channel using a multi-parameter hand-held YSI sonde at the time of discrete sample collection were compared to water temperatures measured continuously with instrumentation at the gaging stations to evaluate their relative accuracy for use in developing and applying regression equations to estimate bacteria concentrations. The discrete water-temperature data were used for the bacteria regression analysis, and the continuous water-temperature data were used to estimate bacteria from the regression relations. The hand-held sonde and the instrumentation at the three gaging stations are checked periodically against a National Institute of Standards and Technology calibrated thermistor. At all five study sites, the discrete and continuous water-temperature data were strongly correlated and generally agreed within less than 1 degree Celsius (figs. 2 and 3), indicating that little to no error is introduced from use of both types of data in developing and applying regression equations to estimate bacteria counts.

Relations between the discrete water temperatures measured using the hand-held sonde and continuous water temperatures measured with a thermistor either on a multi-parameter sonde (three sites on Brandywine Creek, temperature data formally checked and reviewed) or a turbidity sensor (two sites, French Creek and White Clay Creeks, raw provisional temperature data) were similar (figs. 3 and 4), indicating no apparent bias or difference between continuous temperatures measured using different instrumentation. Thus, the provisional continuous water-temperature data recorded by turbidity sensors at French Creek and White Clay Creek gaging stations appear generally similar in accuracy to the formally reviewed and checked water-temperature data collected at the three Brandywine Creek sites.

The slope of the linear regression line between discrete and continuous data was nearly equal to 1 (about 0.995–1.03) for all five sites (figs. 3 and 4), indicating a 1 to 1 correlation with the slope most different from 1 for data at French Creek (slope of about 1.03). The slight difference between discrete and continuous temperature data at French Creek indicated by the slope of about 1.03 may be due to the field setting of the turbidity sensor where periodically shallow stream depths may result in water temperatures that are higher than in midchannel. Differences between temperatures measured by a hand-held water-quality sonde and the turbidity sensor were greater at French Creek than at White Clay Creek, especially at higher temperatures, partly because the turbidity sensor, as installed at French Creek through 2016 was in relatively shallow water during the low flows common in late summer. The median and average differences in temperature measured by a hand-held water-quality sonde and a turbidity sensor were -0.1 and -0.15 degrees Celsius, respectively, at White Clay Creek (48 values) and -0.9 and -0.6 degrees Celsius, respectively, at French Creek (35 values), indicating that water temperatures measured by the turbidity sensors were slightly higher one-half of the time and on average were higher than those measured by hand-held water-quality sonde.

Development of Regression Models to Estimate Fecal Coliform Bacteria Concentrations

The approach to developing relations between measured bacteria concentration, turbidity, and other available stream data is similar to that used in studies done by USGS for streams in Wisconsin, Kansas, and Georgia, as documented by Baldwin and others (2012), Christensen and others (2000), and Lawrence (2012) and for beaches on the Great Lakes (Francy and others, 2013). Linear regressions are used to relate point measurements of fecal bacteria to instantaneous water-quality values or other data collected on a real-time continuous basis; the resulting regression equations can be used to estimate real-time continuous bacteria concentrations using the real-time continuous data.

Regression equations included turbidity, water temperature, and stream discharge as possible explanatory variables

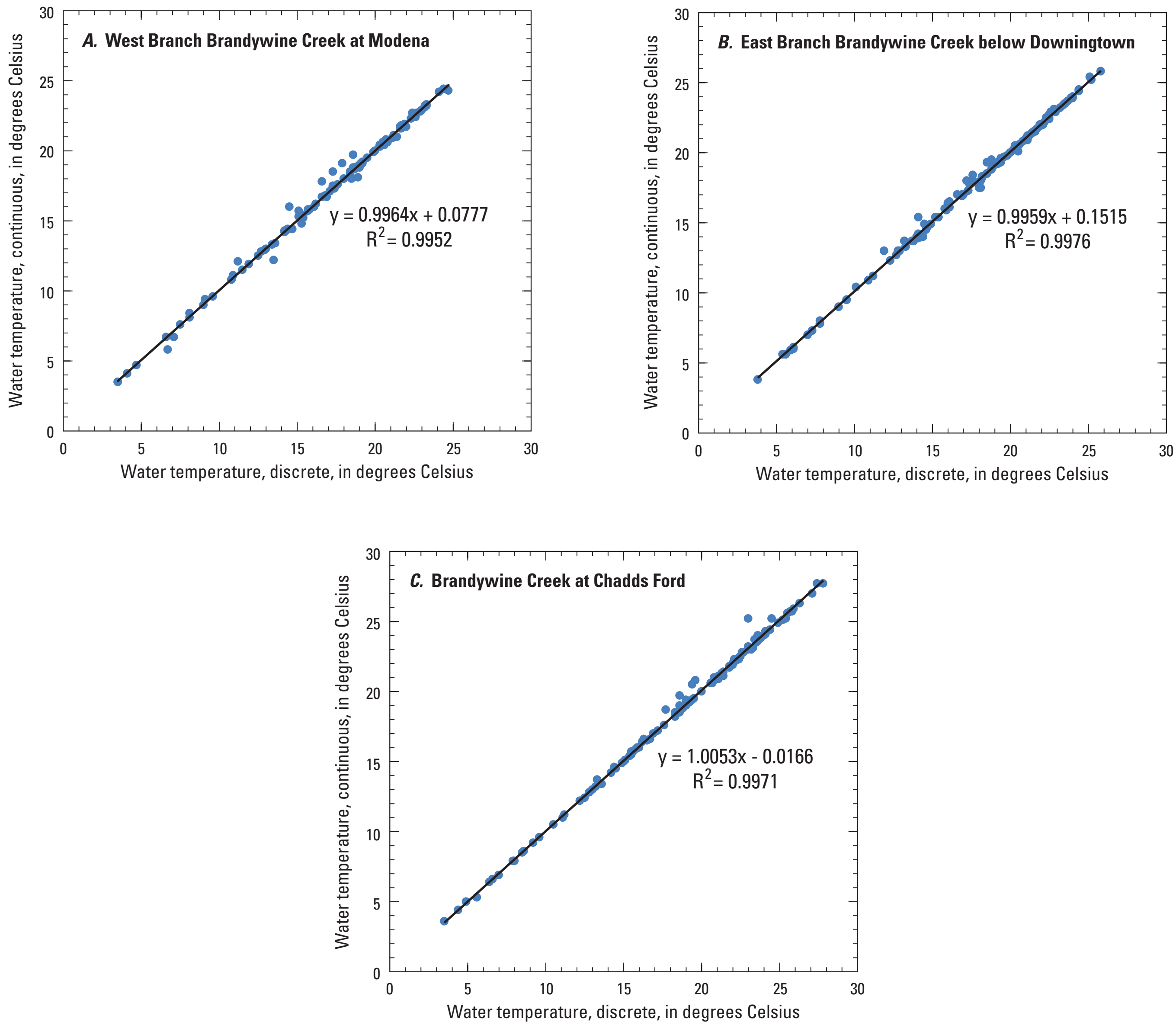


Figure 3. Relation between discrete water temperature measured by hand-held water quality sonde and continuous water temperature measured by thermistors on multi-parameter sondes, March 2012–October 2016 at *A*, East Branch Brandywine Creek at Modena, Pa., *B*, East Branch Brandywine Creek below Downingtown, Pa., and *C*, Brandywine Creek at Chadds Ford, Pa.

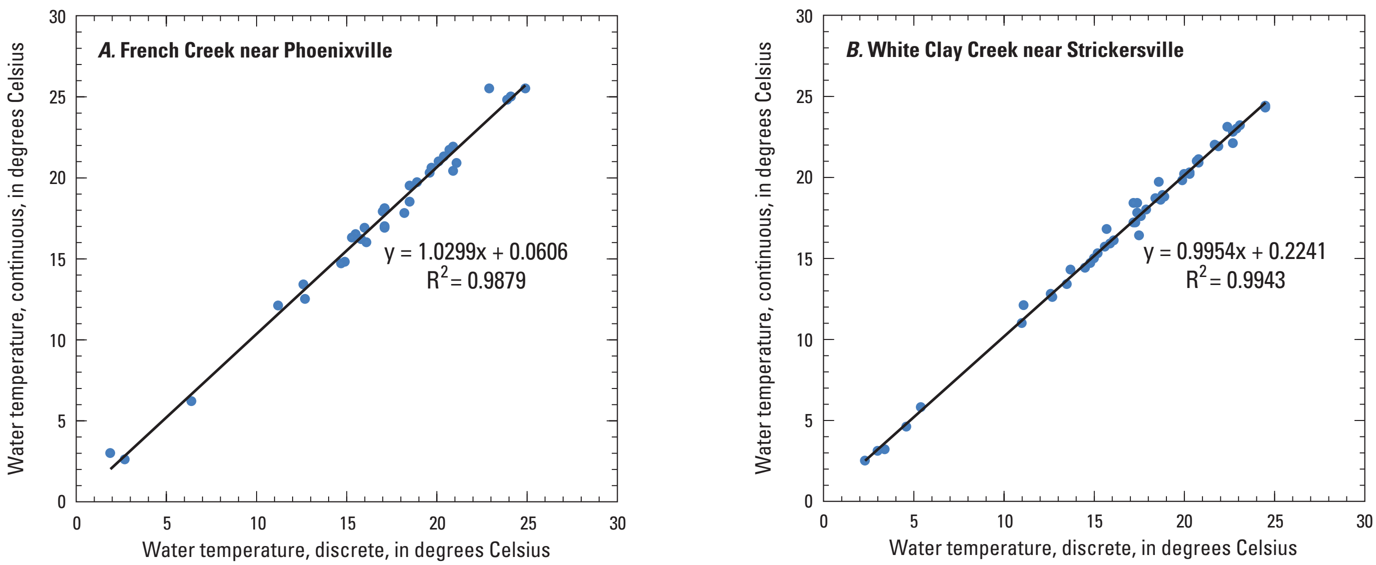


Figure 4. Relation between discrete water temperature measured by hand-held water quality sonde and continuous water temperature measured by thermistor on turbidity sensors at *A*, French Creek near Phoenixville, Pa., 2013–16 and *B*, White Clay Creek near Strickersville, Pa., 2012–16.

because of data availability and statistically significant regression results of other investigations in stream settings elsewhere (Baldwin and others, 2012; Foster and Graham, 2016). Other available continuous water-quality data, including pH, dissolved oxygen concentrations, and specific conductance, were considered as explanatory variables in the preliminary unpublished evaluation of Chester County data done by USGS in 2012, but none were statistically significant or stronger than the physical variables of turbidity, water temperature, and streamflow. In addition, the variables computed using the Julian Day (JD) as a fraction of the 365-day year to account for seasonal variability (warmest temperatures midsummer), sin(2πJD/365) and cos(2πJD/365) as used by Foster and Graham (2016) and hereafter referred to as seasonality variables, were also considered in regressions.

Computation of Statistical Model Fit and Uncertainty

Regression equations that incorporate various combinations of untransformed and transformed variables were considered and evaluated statistically for best fit, using methods described by Helsel and Hirsch (2002), including use of F-test to evaluate inclusion of variable. Statistical measures of fit were coefficient of determination (R^2), adjusted R^2, residual standard error, and model standard percentage error. These statistics are appropriate only for comparison of models with the same response variable units. In this study, for example, models with a response variable of FC in log-transformed units (logFC) need to be compared separately from models with a response variable of untransformed FC concentrations.

The coefficient of determination, R^2, is the fraction of the variance explained by the regression model. The adjusted R^2 is adjusted for the number of explanatory variables in the equation and degrees of freedom to allow comparison of models with differing numbers of explanatory variables. Higher values of R^2 and adjusted R^2 indicate better model fits, with a $R^2 = 1$ indicating a perfect fit. The residual standard error (SE; also known as mean root square error or RMSE) is a measure of the average distance (error), in units of the response variable, between measured response variable values and the regression line. The lower the residual standard error, generally, the better the model fit. The model standard percentage error (MSPE) also can be used to compare regression models; the model with the least uncertainty has the lowest MSPE. For models using variables in log-transformed units, the reported MSPE is the average of the upper and lower MSPE, where the upper MSPE = $100*(10^{RMSE} - 1)$ and the lower MSPE = $100*(1-10^{-RMSE})$.

Statistics were computed using TIBCO Spotfire S+® 8.1 for Windows® (November 2008) and also an R-script developed by Patrick Eslick and others at the USGS Kansas Water Science Center (P. Eslick, USGS, written commun., 2016). Information about the regression model selected on the basis of best-fit metrics and review of residual distributions as calculated using the R-script, as well as the dataset used to develop the regression, is presented in appendixes 1–5 and Senior (2017) for the five stream sites evaluated in this study.

Estimated Fecal Coliform Bacteria Concentrations

Regression equations relating measured discrete FC bacteria concentrations to available continuous explanatory variables were developed to estimate continuous FC concentrations for the five stream sites for the March–October period of the year. The range of discrete-sample bacteria concentrations and associated instantaneous discharge for the three Brandywine Creek sites for 2005–16 and for French Creek and White Clay Creek sites for 2012–16 are shown in figures 5 and 6, respectively. Regression equations were developed using data through 2015 and validated using data from 2016. Streamflows tend to be lower and FC bacteria concentrations to be higher during the summer months than during the spring months of the annual March through September (or March through October for 2012–14) sampling period (figs. 5 and 6).

Summary statistics for continuous discharge, water temperature, and turbidity, and discrete data for discharge, water temperature, turbidity, and FC bacteria concentrations, listed in table 3, show that discrete sample conditions are representative of the range of most (<90th percentile) hydrologic conditions at each of the five gaging stations. Plots of streamflow measured at the time of bacteria sample collection in relation to computed statistics for mean daily streamflow show that flow conditions at the time of sampling fall largely within the 10th to 90th percentile range of mean daily streamflow (figs. 7 and 8), and that mean daily flows vary seasonally, with highest flows in the spring and lowest flows in late summer. Few discrete data were collected under the highest instantaneous or daily mean flow conditions (>90th percentile), although under the highest flow conditions, as shown by Town (2001), bacteria concentrations likely are very high (>10,000 cfu/100 mL) and are likely to exceed Pennsylvania's primary contact recreational standards of 200 cfu/100 mL (geometric mean of 5 samples over 30 days) and 400 cfu/100 mL (no more than 10 percent of samples) for the swimming season May–September.

Of the three Brandywine Creek sites, median FC concentrations in grab samples were highest at West Branch Brandywine Creek at Modena and lowest at Brandywine Creek at Chadds Ford (table 3). Median FC concentrations 2007–15 at the three Brandywine Creek sites were, in decreasing order of magnitude, 705 cfu/100 mL (West Branch Brandywine Creek at Modena), 250 cfu/100 mL (East Branch Brandywine Creek below Downingtown), and 140 cfu/100 mL (Brandywine Creek at Chadds Ford). Fecal coliform concentrations exceeded the PADEP standard of 200 cfu/100 mL in most samples (nearly 90 percent) at West Branch Brandywine Creek at Modena but less at East Branch Brandywine Creek below Downingtown and Brandywine Creek at Chadds Ford frequently (about 50 and 40 percent of the samples, respectively) (fig. 5; table 3). Fecal coliform concentrations at White Clay Creek near Strickersville generally were higher and exceeded the PADEP standards of 200 cfu/100 mL and 2,000 cfu/100 mL more frequently than those at French Creek near Phoenixville (fig. 6; table 3). Median FC concentrations during 2012–15 at these two sites were 265 cfu/100 mL (White Clay Creek near Strickersville) and 120 cfu/100 mL (French Creek near Phonexville). In 2016, the FC concentrations generally were highest in stream-water samples from West Branch Brandywine Creek at Modena and White Clay Creek near Strickersville (median values of 990 and 580 cfu/100 mL, respectively), intermediate in stream samples from East Branch Brandywine Creek below Downingtown and French Creek near Phoenixville (median values of 230 and 205 cfu/100 mL, respectively), and lowest in stream-water samples from Brandywine Creek at Chadds Ford (median value of 150 cfu/100 mL) (table 3).

The results of the regression analysis for the 5 sites are shown in the following section "Site-Specific Regressions" in scatter and time-series plots of measured and estimated (computed from regression equations) values relative to fecal coliform standards for recreational waters to indicate predictive accuracy for the development of public health recommendations. The regression equations are considered cautious predictors of human-health risk if the resulting estimated FC concentrations are greater than measured FC concentrations.

Site-Specific Regressions

Fecal coliform concentrations were estimated using the linear regression equation with the best statistical fit and predictive accuracy relative to PADEP standards of 200 and 400 cfu/100 mL for each of the five sites evaluated. These best-fit relations determined by linear regression between measured FC concentrations and statistically significant ($p<0.05$) explanatory variables for the five sites are listed in table 4. Measures of best fit include R^2, adjusted R^2, root mean square error (RMSE), average model standard percentage error (MSPE), and the accuracy of the model to predict FC concentrations that exceed or do not exceed the 200 cfu/100 mL and 400 cfu/100 mL standards. The accuracy of the regression equation to estimate FC concentrations relative to a recreational water quality standard is computed as accuracy = (number of true positives + number of true negatives) / number of total observations, where a "true positive" indicates that both the estimated and measured FC concentrations exceed a standard and a "true negative" indicates that both the estimated and measured FC concentrations do not exceed a standard. The coefficients for terms in the best-fit regression equations selected to estimate FC concentrations at the five stream sites are listed in table 5, as are the accuracies of each equation relative to the standards of 200, 400, and 2,000 cfu/mL.

The number of bacteria samples available for regression varied among the sites, potentially affecting the power of the regression model fits. The 3 sites on Brandywine Creek have the most data of the 5 sites evaluated; the data were collected over the widest range of hydrologic conditions, especially for the period from 2007 through 2010 when bacteria samples were collected almost weekly from March to September each year (fig. 5).

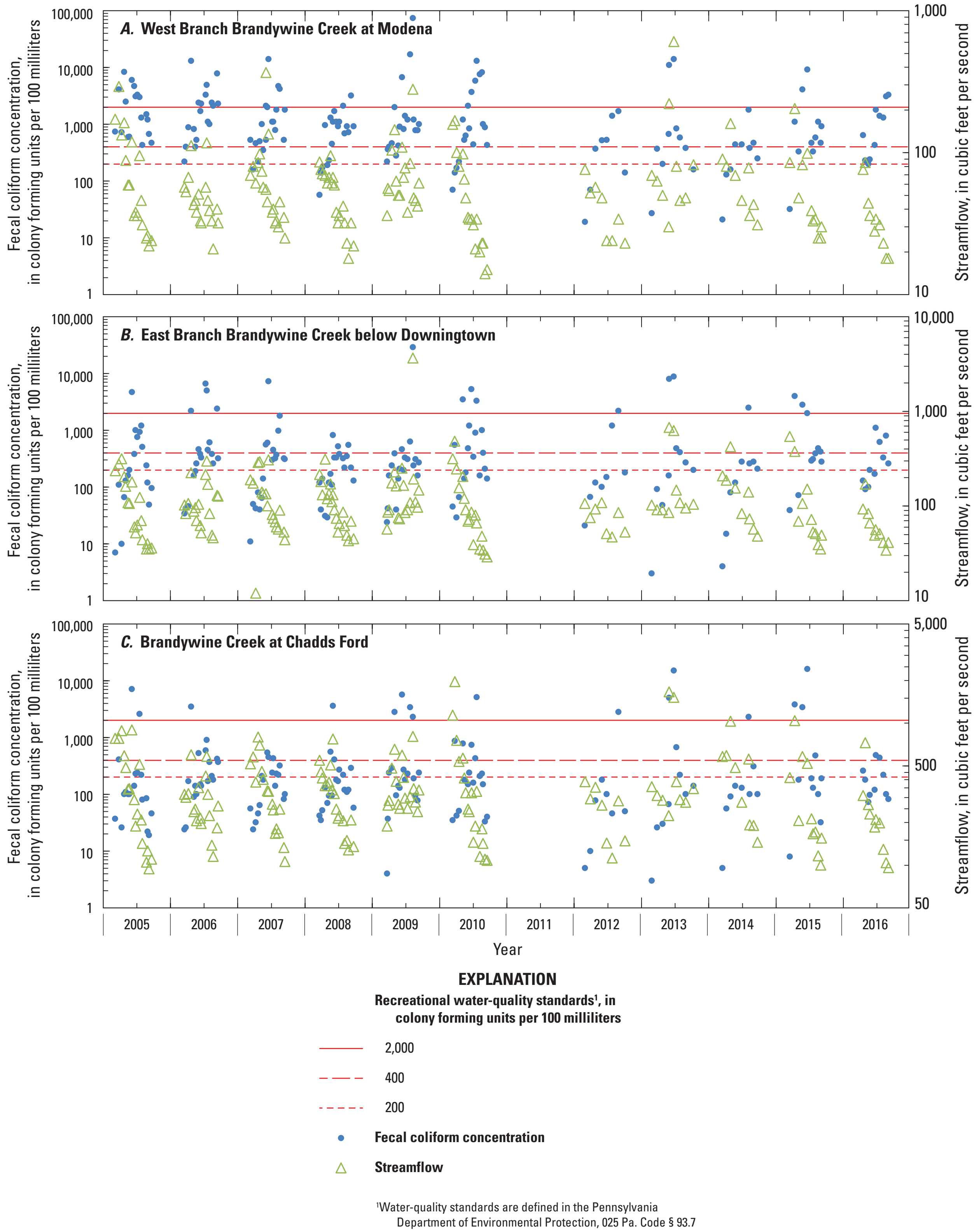


Figure 5. Fecal coliform concentrations in grab samples and associated instantaneous streamflow for the period 2005–16 at *A*, West Branch Brandywine Creek at Modena, Pa., *B*, East Branch Brandywine Creek below Downingtown, Pa., and *C*, Brandywine Creek at Chadds Ford, Pa.

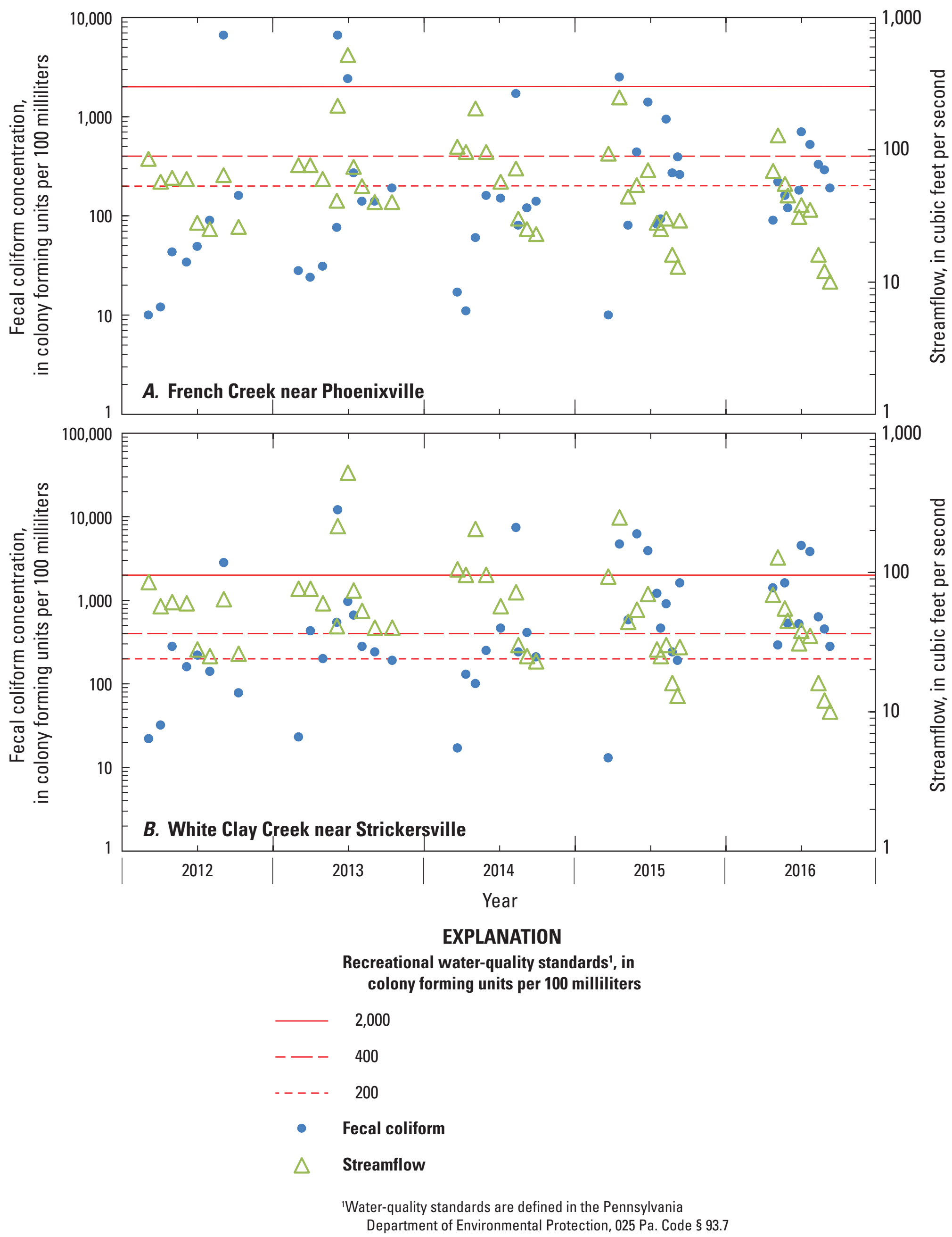


Figure 6. Fecal coliform concentrations in grab samples and associated instantaneous strreamflow for the period 2012–16 at *A*, French Creek near Phoenixville, Pa., and *B*, White Clay Creek near Strickersville, Pa.

Table 3. Summary statistics for continuous or instantaneous streamflow, turbidity, and temperature data and discrete-sample fecal coliform concentration data at five gaging stations on Brandywine, French, and White Clay Creeks in Chester County, Pennsylvania, for the periods of record for regression model, March–October 2007–15, and model validation, May–September 2016.

[USGS, U.S. Geological Survey; Continuous data collected at 15-minute intervals unless otherwise noted; location of gaging stations shown in figure 1; N, number of observations; ft^3/s, cubic feet per second; FNU, Formazin Nephelometric Units; col/100 mL, colonies per 100 milliliters; --, no data; <, less than; unshaded rows indicate data from period of model validation; shaded rows indicate data from period of model calibration]

Type and period of data	Instantaneous streamflow (ft^3/s)							Turbidity (FNU)						
	N	Minimum	10th percentile	Median	90th percentile	Maximum	Average	N	Minimum	10th percentile	Median	90th percentile	Maximum	Average
USGS station 01480617 West Branch Brandywine Creek at Modena, Pa.[1]														
Continuous data 2007–15 (March–October)	211,091	12	24	54	142	6,500	90.3	93,310	<0.1	1.3	3.6	18	1,050	10.5
Discrete data 2007–15 (March–October)	106	14	24	53	107	605	71	106	0.4	1.3	3.4	10.4	110	7.3
Continuous data 2016 (April–September)	17,636	13	20	37	78	587	48.2	9,020	<0.1	1.2	2.8	10	630	6.6
Discrete data 2016 (April–September)	10	18	18	32.5	76.9	85	39.3	9	1.35	1.9	2.8	5	8	3.3
USGS station 01480870 East Branch Brandywine Creek below Downingtown, Pa.[2]														
Continuous data 2007–15 (March–October)	211,959	20	44	99	287	6,360	166	135,960	<0.1	1	2.6	11	1,080	7.7
Discrete data 2007–15 (March–October)	102	12	45	92	293	3,660	170	102	0.6	1	2.4	8.7	170	6.4
Continuous data 2016 (April–September)	17,558	17	34	59	140	1,720	82	16,961	0.2	1	1.9	4.7	480	3.9
Discrete data 2016 (April–September)	10	34	40	54	101	170	68.4	10	0.8	0.9	1.7	6.2	7	2.8
USGS station 01481000 Brandywine Creek at Chadds Ford, Pa.[3]														
Continuous data 2007–15 (March–October)	217,275	83	133	300	804	22,200	467	135,435	<0.1	1.5	4	17	980	10
Discrete data 2007–15 (March–October)	101	94	142	293	701	1,960	385	101	<0.1	1.6	3.9	15	120	8.5
Continuous data 2016 (April–September)	17,621	83	104	209	386	3,390	250	16,944	0.6	1.7	3.7	8.4	350	6.1
Discrete data 2016 (April–September)	10	96	103	204	351	728	245	9	1.7	2.2	2.5	4.5	5	3.1
USGS station 01472157 French Creek near Phoenixville, Pa.[4]														
Continuous data 2012–15 (March–October)	103,441	12	23	52	143	5,770	83.7	101,932	<0.1	1.3	3.5	11	1,200	7.3
Discrete data 2012–15 (March–October)	37	13	25	57	145	518	79.2	37	0.4	1.7	3.6	15.4	69	7.5
Continuous data 2013–15 (March–October)														
Discrete data 2013–15 (March–October)														
Continuous data 2016 (April–September)	17,379	8.4	12	31	88	790	47.6	16,882	0.3	1.8	3.3	6.4	180	4.9
Discrete data 2016 (April–September)[5]	10	10	11.8	36.5	74.9	128	43.9	10	2	2.4	3.5	5.1	5	3.5
USGS station 01478245 White Clay Creek near Strickersville, Pa.														
Continuous data 2012–15 (March–October)	91,458	15	27	54	118	7,700	83.8	90,923	<0.1	0.8	2	11	1,170	9.1
Discrete data 2012–15 (March–October)	38	22	28	54.5	117	946	89.4		0.5	0.9	2	10.7	260	12.3
Continuous data 2016 (April–September)	17,516	23	28	45	83	1,130	56.9		<0.1	0.8	1.8	6.5	460	4.6
Discrete data 2016 (April–September)	10	25	27.7	48.5	66.8	83	49.3	10	1.1	1.3	2	3.2	6	2.4

Table 3. Summary statistics for continuous or instantaneous streamflow, turbidity, and temperature data and discrete-sample fecal coliform concentration data at five gaging stations on Brandywine, French, and White Clay Creeks in Chester County, Pennsylvania, for the periods of record for regression model, March–October 2007–15, and model validation, May–September 2016.—Continued

[USGS, U.S. Geological Survey; Continuous data collected at 15-minute intervals unless otherwise noted; location of gaging stations shown in figure 1; N, number of observations; ft³/s, cubic feet per second; FNU, Formazin Nephelometric Units; col/100 mL, colonies per 100 milliliters; --, no data; <, less than; unshaded rows indicate data from period of model validation; shaded rows indicate data from period of model calibration]

Type and period of data	Fecal coliform bacteria (col/100 mL)							Water temperature (degrees Celsius)						
	N	Minimum	10th percentile	Median	90th percentile	Maximum	Average	N	Minimum	10th percentile	Median	90th percentile	Maximum	Average
USGS station 01480617 West Branch Brandwyine Creek at Modena, Pa.[1]														
Continuous data 2007–15 (March–October)	--	--	--	--	--	--	--	101,182	-0.1	8.9	18.3	23.3	29.7	17.1
Discrete data 2007–15 (March–October)	106	19	160	705	5,000	74,000	2,436	106	3.5	9.1	17.8	22.5	24.7	16.7
Continuous data 2016 (April–September)	--	--	--	--	--	--	--	9,467	5	12.8	20.9	24.8	28.5	19.6
Discrete data 2016 (April–September)	10	200	227	990	3,120	3,300	1,264	10	12.7	15	20.4	24.4	24.4	19.7
USGS station 01480870 East Branch Brandywine Creek below Downingtown, Pa.[2]														
Continuous data 2007–15 (March–October)	--	--	--	--	--	--	--	138,827	0	8.2	17.6	23.1	30.1	16.7
Discrete data 2007–15 (March–October)	102	3	40	250	2,180	29,000	1,020	102	3.8	7.9	17.9	22.9	25.1	17
Continuous data 2016 (April–September)	--	--	--	--	--	--	--	17,481	5.2	12.6	21	25.0	29.1	19.7
Discrete data 2016 (April–September)	10	91	99	230	830	1,100	380	10	14.1	15.3	21.5	25.3	25.8	20.7
USGS station 01481000 Brandywine Creek at Chadds Ford, Pa.[3]														
Continuous data 2007–15 (March–October)	--	--	--	--	--	--	--	142,946	-0.2	8.3	18.4	24.2	31.2	17.3
Discrete data 2007–15 (March–October)	101	3	33	140	2,800	16,000	861	101	3.5	10.5	19.2	24.1	27.8	18.3
Continuous data 2016 (April–September)	--	--	--	--	--	--	--	17,570	5.5	12.8	22.3	26.2	30.0	20.6
Discrete data 2016 (April–September)	10	73	81	150	445	490	206	10	14.2	16.5	23	27.2	27.8	22.2
USGS station 01472157 French Creek near Phoenixville, Pa.[4]														
Continuous data 2012–15 (March–October)	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Discrete data 2012–15 (March–October)	37	10	15	120	1,980	6,600	692	37	1.9	5.3	17.1	22.6	24.1	16.0
Continuous data 2013–15 (March–October)								*70,382*	*-0.1*	*5.6*	*17.9*	*24.6*	*31.8*	*16.1*
Discrete data 2013–15 (March–October)								*29*	*1.9*	*5.8*	*17.1*	*22.9*	*24.1*	*16.2*
Continuous data 2016 (April–September)	--	--	--	--	--	--	--	17,530	5.1	13	21.8	26.6	34.2	20.6
Discrete data 2016 (April–September)[5]	10	90	117	205	538	700	280	10	12.6	15	20.6	24.2	24.9	19.8
USGS station 01478245 White Clay Creek near Strickersville, Pa.														
Continuous data 2012–15 (March–October)		--	--	--	--	--	--	91,670	-0.1	7.9	17.5	22.6	29.3	16.3
Discrete data 2012–15 (March–October)	38	13	29	265	4,140	12,000	1,275	38	2.3	5.2	17.3	22.5	23.1	15.9
Continuous data 2016 (April–September)	--	--	--	--	--	--	--		4.1	11.8	20.2	24.3	28.2	19.0
Discrete data 2016 (April–September)	10	280	289	580	3,870	4,500	1,400	10	12.7	14.6	20.1	24.5	24.5	19.4

[1] West Branch Brandywine Creek at Modena, Pa.: Turbidity (30-minute data 2007–16).

[2] East Branch Brandywine Creek below Downingtown, Pa.: Turbidity (30-minute data 2007–12; 15-minute data 2013–16).

[3] Brandywine Creek at Chadds Ford, Pa.: Turbidity 144,625 (30-minute data 2007–12; 15-minute data 2013–16).

[4] Period of continuous water temperature data bst French Creek near Phoenixville, Pa. begins in December 2012, so summary statistics for water temperature are presented for March–October 2013–15 (*in italics*).

[5] Hand-held (hand-held can be 0.5 up to 1.0 colder than continuous).

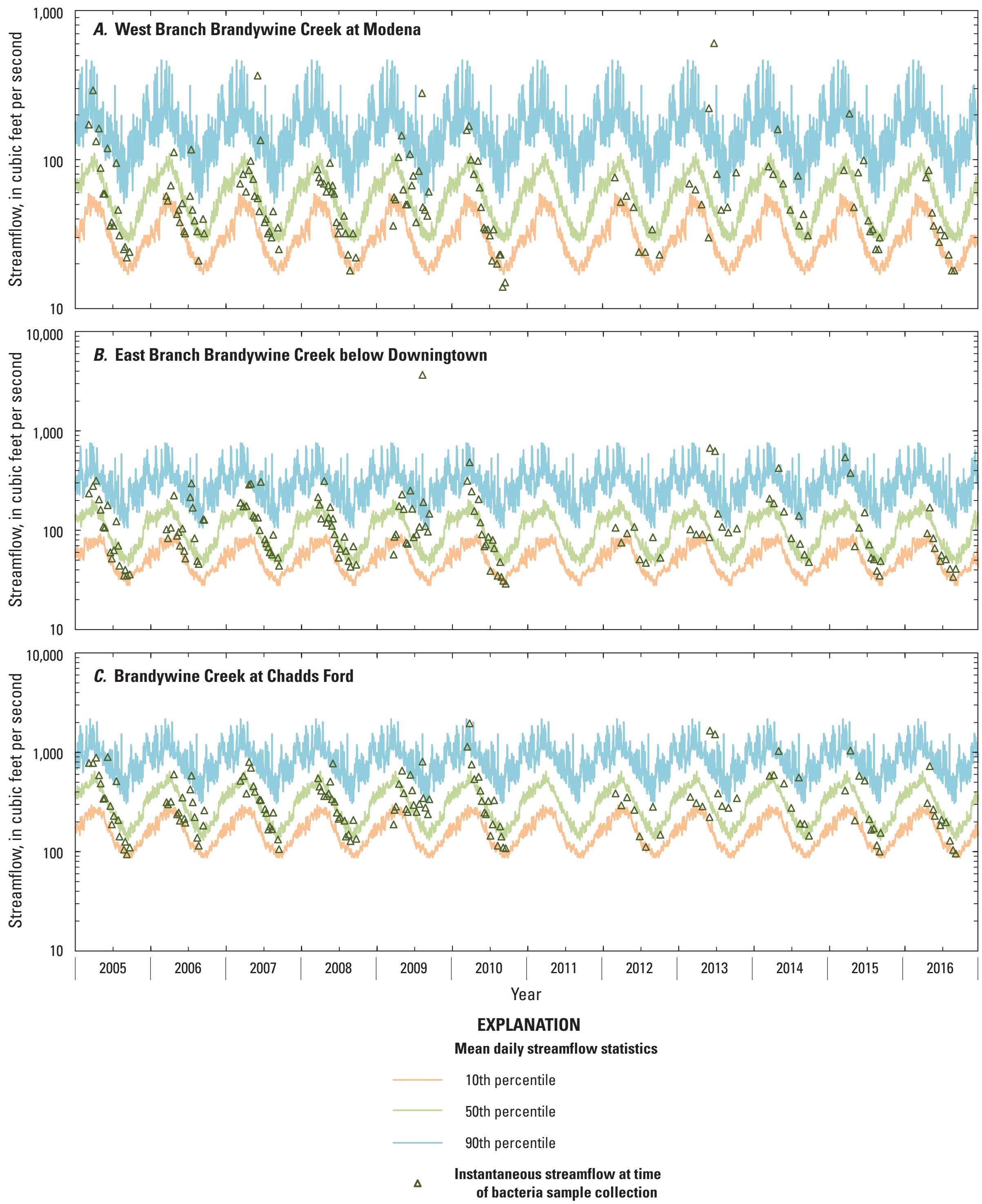


Figure 7. Instantaneous streamflow measured at the time of bacteria sample collection for the period 2005–16 and 10th, 50th, and 90th percentiles of mean daily streamflow at *A*, West Branch Brandywine Creek at Modena, Pa., *B*, East Branch Brandywine Creek below Downingtown, Pa., and *C*, Brandywine Creek at Chadds Ford, Pa.

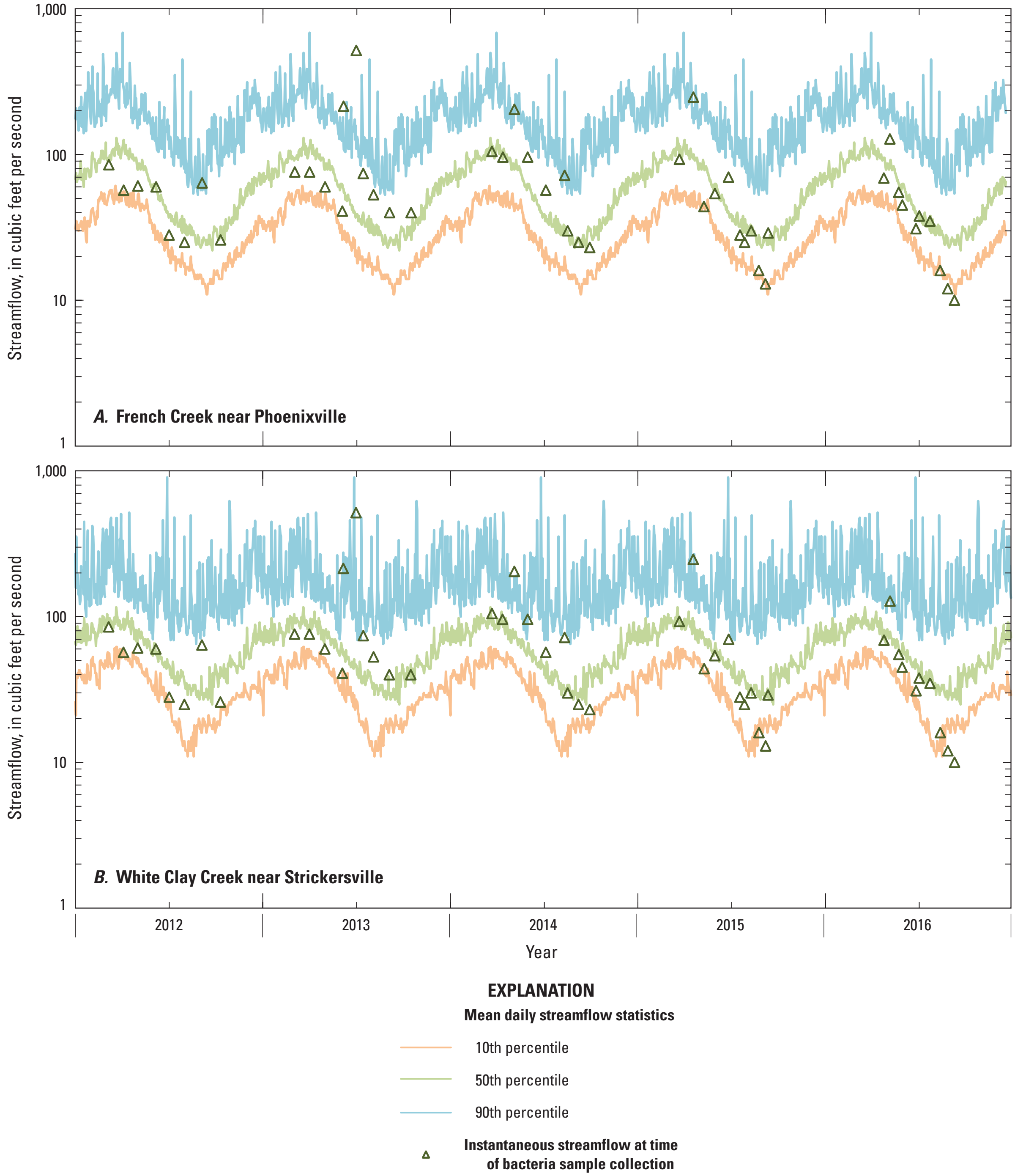


Figure 8. Instantaneous streamflow measured at the time of bacteria sample collection for the period 2012–16 and 10th, 50th, and 90th percentiles of mean daily streamflow at *A*, French Creek near Phoenixville, Pa., and *B*, White Clay Creek near Strickersville, Pa.

Table 4. Linear regression equations and related statistics showing statistically significant relations between fecal coliform bacteria concentrations and explanatory variables of turbidity, water temperature, streamflow, and seasonal factors at five gaging stations on Brandywine, French, and White Clay Creeks in Chester County, Pennsylvania, where bacteria, turbidity, and other water-quality data were collected by the U.S. Geological Survey during 2007–15. Explanatory variables are considered statistically significant in regression equations at p<0.05. Location of stream sites shown in figure 1 identified by USGS station number.

[USGS, U.S. Geological Survey; Period of record of data used in regression model for each station listed in table 3. Explanatory variables considered statistically significant in regression equations at $p < 0.05$; v, number of explanatory variables; n, number of observations; RMSE, root mean square error (residual standard error); MSPE, average model standard percentage error; R^2, coefficient of determination; adjR^2, adjusted coefficient of determination; BCF, bias correction factor; cfu/100 mL, colony forming units per 100 milliliters; nc, not calculated; FC, fecal coliform concentration (in cfu/100 mL); TURB, turbidity (in Formazin Nephelometric Units); TEMP, water temperature (in degrees Celsius); Q, instantaneous streamflow (in cubic feet per second); seasonal factors: sinπJD, sin(2*π*Julian Day/365); cos2πJD, cos(2*π*Julian Day/365); Accuracy of model to predict fecal coliform concentrations relative to a recreational water-quality standard is computed as accuracy = (number of true positives + number of true negatives / total number of observations)]

Linear regression equation	v	n	RMSE	MSPE	R^2	adjR^2	BCF	Accuracy relative to water-quality standard	
								200 cfu/ 100 mL	400 cfu/ 100 mL
USGS station 01480617 West Branch Brandywine Creek at Modena, Pa.									
logFC ~ logTURB + sin2πJD + cos2πJD	3	106	0.348	89	0.695	0.686	1.38	0.87	0.87
logFC ~ logTURB + TEMP	2	106	0.358	92.2	0.674	0.668	1.41	0.84	0.85
logFC ~ logTURB	1	106	0.446	129	0.442	0.436	1.66	nc	nc
USGS station 01480870 East Branch Brandywine Creek below Downingtown, Pa.									
logFC ~ logTURB + sin2πJD + cos2πJD	3	102	0.379	98.7	0.708	0.700	1.45	0.83	0.76
logFC ~ logTurb + TEMP + logQ	3	102	0.408	108	0.663	0.652	1.55	0.82	0.78
logFC ~ logTURB + TEMP	2	102	0.420	113	0.637	0.630	1.59	nc	nc
logFC ~ logTURB	1	102	0.596	184	0.265	0.257	2.07	nc	nc
USGS station 01481000 Brandywine Creek at Chadds Ford, Pa.									
logFC ~ logTURB + sin2πJD + cos2πJD	3	101	0.349	89.3	0.766	0.759	1.40	0.77	0.84
logFC ~ logTURB + TEMP	2	101	0.380	102	0.705	0.699	1.56	0.72	0.54
logFC ~ logTURB	1	101	0.524	152	0.462	0.457	1.86	nc	nc
USGS station 01472157 French Creek near Phoenixville, Pa.									
logFC ~ logTURB + sin2πJD + cos2πJD	3	37	0.428	115	0.710	0.684	1.57	0.84	0.87
logFC ~ logTURB + TEMP	2	37	0.510	147	0.575	0.551	1.78	0.73	0.84
logFC ~ logTURB	1	37	0.587	180	0.421	0.405	2.12	nc	nc
USGS station 01478245 White Clay Creek near Strickersville, Pa.									
logFC ~ logTURB + sin2πJD + cos2πJD	3	38	0.452	124	0.642	0.610	1.60	0.84	0.68
logFC ~ logTURB + TEMP + sin2πJD	3	38	0.404	107	0.714	0.689	1.49	0.79	0.63
logFC ~ logTURB + TEMP	2	38	0.442	120	0.648	0.628	1.70	nc	nc
logFC ~ logTURB	1	38	0.611	192	0.307	0.288	2.00	nc	nc

Table 5. Accuracy of and coefficients for terms in linear regression equations selected to estimate fecal coliform bacteria concentrations using explanatory variables of turbidity and seasonal factors based on Julian Day, and number (and fraction) of observed fecal coliform concentrations above and below water-quality standards at five gaging stations on Brandywine, French, and White Clay Creeks in Chester County, Pennsylvania, 2007–15. Period of record for each station listed in table 3 and statistics for equations listed in table 4. Accuracy of model to predict fecal coliform concentrations relative to a recreational water quality standard is computed as accuracy = (number of true positives + number of true negatives) / number of total observations.

[USGS, U.S. Geological Survey; Period of record used for regression analysis at each station listed in table 3 and statistics for regression equations listed in table 4. Accuracy of model to predict fecal coliform concentrations relative to a recreational water quality standard is computed as accuracy = (number of true positives + number of true negatives) / total number of observations; n, number of observations; false positive, model estimate is greater than standard but observed value is less than standard; false negative, model estimate is less than standard but observed value is greater than standard; fraction of false negatives, false positives, and observations with FC concentrations above and below standards calculated as number of those quantities / total number of observations; FC, fecal coliform; cfu/100 mL, colony forming units per 100 milliliters; Turb, turbidity (in Formazine Nephelometric Units); JD, Julian Day; <, less than; >, greater than]

Linear regression equation	n	Accuracy relative to water-quality standard		
		200 cfu/100 mL	400 cfu/100 mL	2,000 cfu/100 mL
USGS station 01480617 West Branch Brandywine Creek at Modena, Pa. (2007–15)				
logFC = 0.758*logTurb – 0.293*sin(2πJD/365) – 0.529*cos(2πJD/365) + 2.13	106			
Accuracy of model estimate		0.87	0.87	0.81
Number (and fraction) of model false positives		9 (0.08)	10 (0.09)	15 (0.14)
Number (and fraction) of model false negatives		5 (0.05)	4 (0.04)	5 (0.05)
Number (and fraction) of observations with FC concentrations < standard		*16 (0.15)*	*31 (0.29)*	*87 (0.82)*
Number (and fraction) of observations with FC concentrations > standard		*90 (0.85)*	*75 (0.71)*	*19 (0.18)*
USGS station 01480870 East Branch Brandywine Creek below Downingtown, Pa. (2007–15)				
logFC = 0.879*logTurb – 0.444*sin(2πJD/365) – 0.638*cos(2πJD/365) + 1.68	102			
Accuracy of model estimate		0.83	0.76	0.89
Number (and fraction) of model false positives		14 (0.14)	21 (0.21)	4 (0.04)
Number (and fraction) of model false negatives		3 (0.03)	4 (0.04)	7 (0.04)
Number (and fraction) of observations with FC concentrations < standard		*42 (0.41)*	*71 (0.70)*	*90 (0.88)*
Number (and fraction) of observations with FC concentrations > standard		*60 (0.59)*	*31 (0.30)*	*12 (0.12)*
USGS station 01481000 Brandywine Creek at Chadds Ford, Pa. (2007–15)				
logFC = 1.1*logTurb – 0.288*sin(2πJD/365) – 0.774*cos(2πJD/365) + 1.12	101			
Accuracy of model estimate		0.77	0.84	0.95
Number (and fraction) of model false positives		22 (0.22)	14 (0.14)	0 (0.0)
Number (and fraction) of model false negatives		1 (0.01)	2 (0.02)	5 (0.05)
Number (and fraction) of observations with FC concentrations < standard		*63 (0.62)*	*77 (0.76)*	*88 (0.87)*
Number (and fraction) of observations with FC concentrations > standard		*38 (0.38)*	*24 (0.24)*	*13 (0.13)*
USGS station 01472157 French Creek near Phoenixville, Pa. (2012–15)				
[1]logFC = 1.25*logTurb – 0.543*sin(2πJD/365) + 1.4	37			
Accuracy of model estimate		0.84	0.87	0.97
Number (and fraction) of model false positives		6 (0.16)	3 (0.08)	0 (0.0)
Number (and fraction) of model false negatives		0 (0.0)	2 (0.05)	1 (0.03)
Number (and fraction) of observations with FC concentrations < standard		*25 (0.68)*	*29 (0.78)*	*33 (0.89)*
Number (and fraction) of observations with FC concentrations > standard		*12 (0.32)*	*8 (0.22)*	*4 (0.11)*
USGS station 01478245 White Clay Creek near Strickersville, Pa.				
logFC = 0.653*logTurb – 0.281*sin(2πJD/365) – 0.747*cos(2πJD/365) + 1.87	38			
Accuracy of model estimate		0.84	0.68	0.92
Number (and fraction) of model false positives		5 (0.13)	9 (0.24)	1 (0.03)
Number (and fraction) of model false negatives		1 (0.03)	3 (0.08)	2 (0.05)
Number (and fraction) of observations with FC concentrations < standard		*12 (0.32)*	*21 (0.55)*	*32 (0.84)*
Number (and fraction) of observations with FC concentrations > standard		*26 (0.68)*	*17 (0.45)*	*6 (0.16)*

[1]The coefficient for the seasonality term cos(2πJD/365) in regression equation for French Creek was determined to be not statistically significantly different from zero.

The explanatory variables considered in regression analysis are turbidity (Turb), in FNU; stream discharge (Q), in cubic feet per second; water temperature (Temp), in degrees Celsius; and the seasonality variables, sin(2πJD/365) and cos(2πJD/365), that are calculated using Julian Day (JD) as a fraction of a 365-day year and that account for seasonal factors, such as temperature, affecting bacteria concentrations. Turbidity and stream discharge values obtained from the available continuous 15-minute records (or 30 minutes for some turbidity records) for gaging stations, interpolated between recorded values when necessary, and water temperature values measured at the time of bacteria sample collection were used in the regressions. Regressions were computed using base 10 log transformations for the three variables with values that range over at least three orders of magnitude and likely have an approximately log-normal distribution, including FC concentrations (logFC), turbidity (logTurb), and discharge (logQ). Estimated log FC concentrations can be retransformed to original units in order to directly calculate or estimate FC concentrations, a procedure that introduces a bias in the calculated constituent. The bias may be corrected using Duan's Bias Correction Factor (BCF) computed for each model (table 4) using model residuals (Helsel and Hirsch, 2002). The BCF is calculated by determining the mean of model residuals after the residuals were transformed (from log) to original units.

Of the five explanatory variables considered, turbidity and the seasonality variables [sin(2πJD/365) and cos(2πJD/365)] resulted in the best fit of data by regression to estimate log FC at most stream sites (tables 4 and 5). Commonly, although FC concentrations tend to increase with increases in turbidity and stream discharge, turbidity was a better predictor of FC concentrations than stream discharge, and the variable logTurb was always included in regression models. However, because stream discharge is positively correlated with turbidity, stream discharge frequently was not statistically significant when both turbidity and stream discharge were considered in the regressions. Similarly, water temperature and the seasonality variables (sin2piJD and cos2piJD) are related, and commonly, either water temperature or the seasonality variables were statistically significant when considered in the regressions.

The inclusion of the seasonality variables, sin(2πJD/365) and cos(2πJD/365), rather than water temperature in the best-fit models indicates that seasonal factors other than, or in addition to, water temperature may affect FC concentrations. These seasonal factors may include changes in sources of bacteria related to wildlife and domestic animal populations and activities, length of day and intensity of sunlight, as well as human populations, activities, and land uses. Negative coefficients for the seasonality variables (sin2piJD and cos2piJD) in regression models estimating logFC account for measured increases in temperature and FC concentrations in summer. The magnitude of coefficients for the seasonality variables affects the timing of peak seasonal effects on FC concentrations. The regression analysis yielded coefficients for the seasonality variables that result in peaks occurring in late July to early August (Julian Day 203–219) for the White Clay Creek and three Brandywine Creek stream sites and in September (Julian Day 274) for the French Creek stream site (fig. 9).

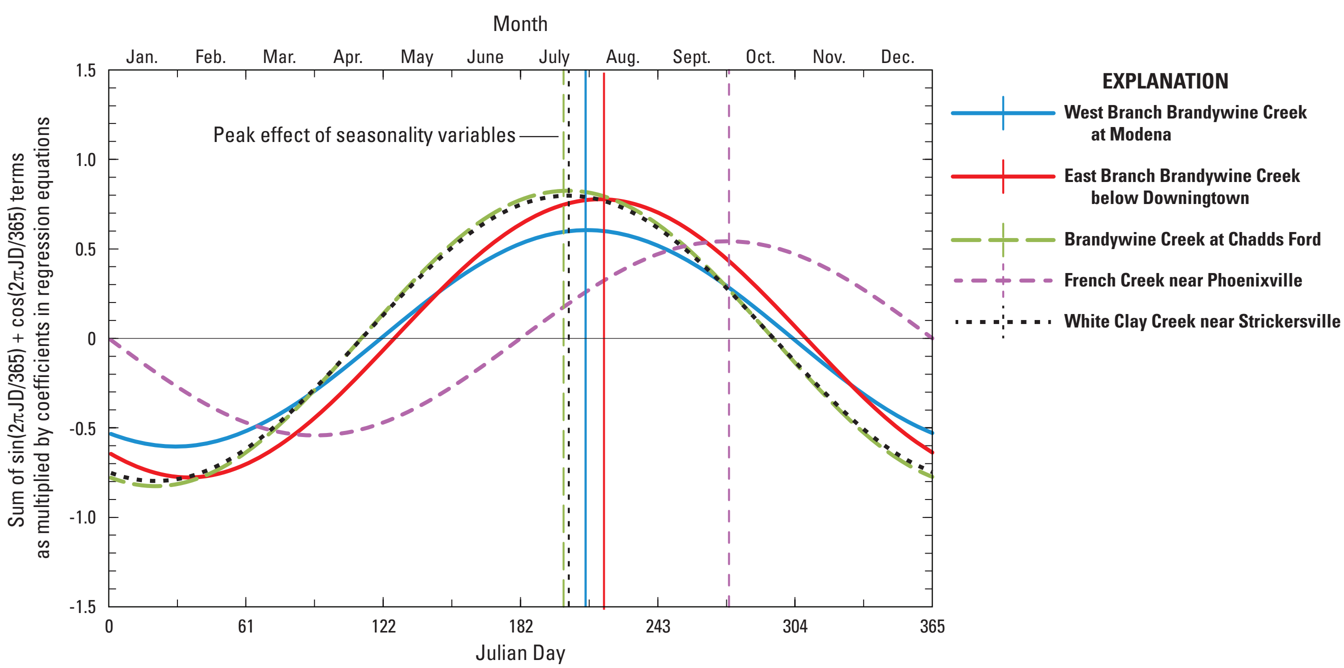


Figure 9. Computed sum of seasonal variable terms in relation to Julian Day (JD) for linear regression equations estimating log fecal coliform concentrations at five stream sites in Chester County, Pa. Sum of seasonality variable terms = coefficient*sin(2πJD/365) + coefficient*cos(2πJD/365), where coefficients for seasonality variables in regression equations are listed in table 5 for the five stream sites. Vertical lines indicate day of peak magnitude of seasonality variable affecting estimated fecal coliform concentrations at each site.

Regression results for individual sites are discussed in the following sections. Results include the accuracy of the final regression models relative to estimating FC concentrations above established standards of 200, 400, and 2,000 cfu/100 mL (table 5), with accuracy = (number of true positives + number of true negatives) / number of total observations, as described previously. Additionally, results include the number of “false negatives” for cases where the measured value exceeds, but the estimated value does not exceed (and thus underestimates), a standard, and the number of “false positives” for cases where the measured value does not exceed, but the estimated value does exceed (and thus overestimates), a standard. In most cases of “false negatives,” the underestimation is relative to only one standard, such that the estimated value might be less than a measured value above the standard of 2,000 cfu/100 mL but is still greater than the standards of 200 or 400 cfu/100 mL.

West Branch Brandywine Creek at Modena

Of the five explanatory variables considered, for models that include logTurb, only Temp and seasonality variables, sin(2πJD/365) and cos(2πJD/365), (table 4) were determined to be statistically significant ($p < 0.05$) for the West Branch Brandywine Creek at Modena model. The best fit was obtained by a regression model (table 5) where logFC = 0.758*logTurb - 0.293*sin(2πJD/365) - 0.529*cos(2πJD/365) + 2.13, which resulted in an adjusted R^2 of 0.686 (table 4). The 90th percentile prediction interval (fig. 10*A*) is a factor of about +/-5 for FC concentration in units of cfu/100 mL. Detailed information about data used in, and the statistics for, the best-fit regression model for West Branch Brandywine Creek at Modena is given in appendix 1. The retransformed regression model adjusted using the Duan’s BCF of 1.38 estimates FC concentrations in original units of cfu/100 mL as follows: 1.38×10^{logFC}, where logFC is the estimate from the regression equation.

Computed FC concentrations adjusted by applying Duan’s BCF tend to overestimate more than underestimate FC concentrations relative to the standards of 200, 400, and 2,000 colonies/100 mL (fig. 10*B*). The regression model underestimated about 14 percent of 106 measured FC concentrations relative to at least one, but typically not more than one, standard. For the “false negative” results, about 5 percent of underestimations were less than measured values above 200 cfu/100 mL, about 4 percent of underestimations were less than measured values above 400 cfu/100 mL, and about 5 percent of underestimations were less than measured values above 2,000 cfu/100 mL (see shaded areas on fig. 10*B* for underestimations relative to 200 and 2,000 cfu/100mL standards; table 5). The accuracy of the West Branch Brandywine Creek at Modena model to estimate values exceeding or not exceeding the standards of 200, 400, and 2,000 cfu/100 mL was 0.87 (87 percent), 0.87 (87 percent), and 0.81 (81 percent), respectively.

The regression model estimated that FC concentrations were higher during the warm summer months than during the cooler spring months, as indicated by measured concentrations for recent (2013–15; fig. 11) and prior periods (fig. 5; Town, 2001). Measured and computed FC concentrations were greater than 200 and 400 cfu/100 mL for most of the May through September period each year, and FC concentrations were greater than 200 cfu/100 mL into October (figs. 5*A* and 11*C*) at West Branch Brandywine Creek at Modena. Additionally, measured and computed FC concentrations commonly were greater than the highest standard of 2,000 cfu/100 mL during June and July each year at West Branch Brandywine Creek at Modena.

East Branch Brandywine Creek Below Downingtown

All of the five explanatory variables considered for models that include logTurb were determined to be statistically significant ($p < 0.05$) for the East Branch Brandywine Creek below Downingtown models, but the best fit included only logTurb and the seasonality variables (table 4). The best fit was obtained by a regression model (table 5) where

logFC =0.879*logTurb - 0.444*sin(2πJD/365) - 0.638*cos(2πJD/365) + 1.68 ,which resulted in an adjusted R^2 of 0.70 (table 4). The 90th percentile prediction interval (fig. 12*A*) is a factor of about +/-5 for FC concentration in units of cfu/100 mL. Detailed information about data used in, and the statistics for, the best-fit regression model for East Branch Brandywine Creek below Downingtown is given in appendix 2. The retransformed regression model adjusted using the Duan’s BCF of 1.45 estimates FC concentrations in original units of cfu/100 mL as follows: 1.45×10^{logFC} ,where logFC is the estimate from the regression equation.

Computed FC concentrations adjusted by applying Duan’s BCF tend to be overestimated more than underestimated relative to the standards of 200, 400, and 2,000 cfu/100 mL (fig. 12*B*). The regression model underestimated about 14 percent of 102 measured FC concentrations relative to at least one, but typically not more than one, standard. For these “false negative results,” about 3 percent of underestimations were less than measured values greater than 200 cfu/100 mL, about 4 percent of underestimations were less than measured values greater than 400 cfu/100 mL, and about 7 percent of underestimations were less than measured values greater than 2,000 cfu/100 mL (see shaded areas on fig. 10*B* for underestimations relative to 200 and 2,000 cfu/100 mL standards; table 5). The accuracy of the East Branch Brandywine Creek below Downingtown model to predict values exceeding or not exceeding the standards of 200, 400, and 2,000 cfu/100 mL was 0.83 (83 percent), 76 (76 percent), and 0.89 (89 percent), respectively.

The regression model estimated that FC concentrations were higher during the warm summer months than during the

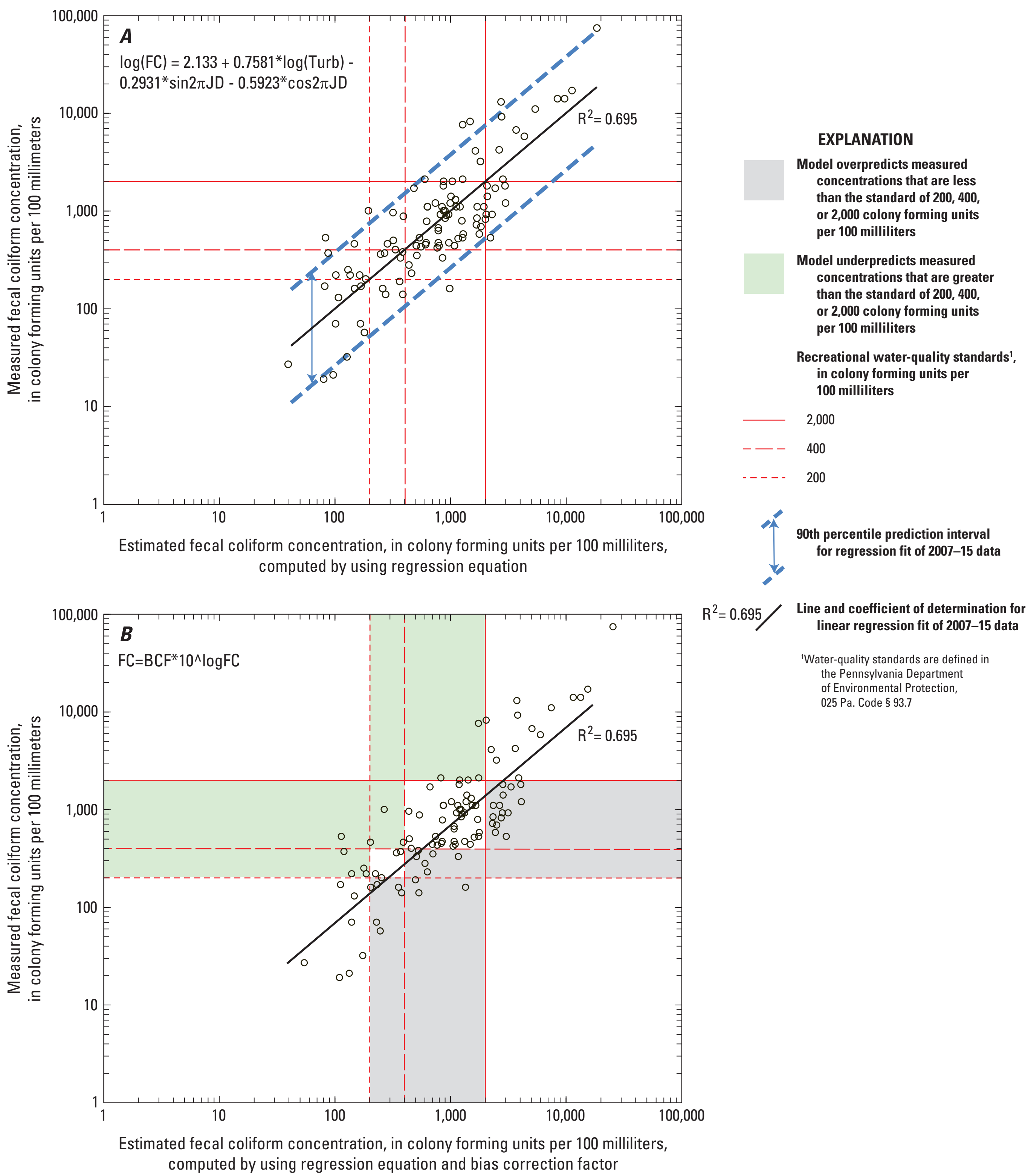


Figure 10. Relations between *A*, estimated and measured log fecal coliform concentrations and *B*, estimated (corrected for bias) and measured fecal coliform concentrations for West Branch Brandywine Creek at Modena, Pa., 2007–15. Estimated log fecal coliform (FC) concentrations were computed by using regression equation. A bias correction factor (BCF) of 1.376 was applied to the computed log fecal coliform concentrations after retransformation to original units as shown in *B*.

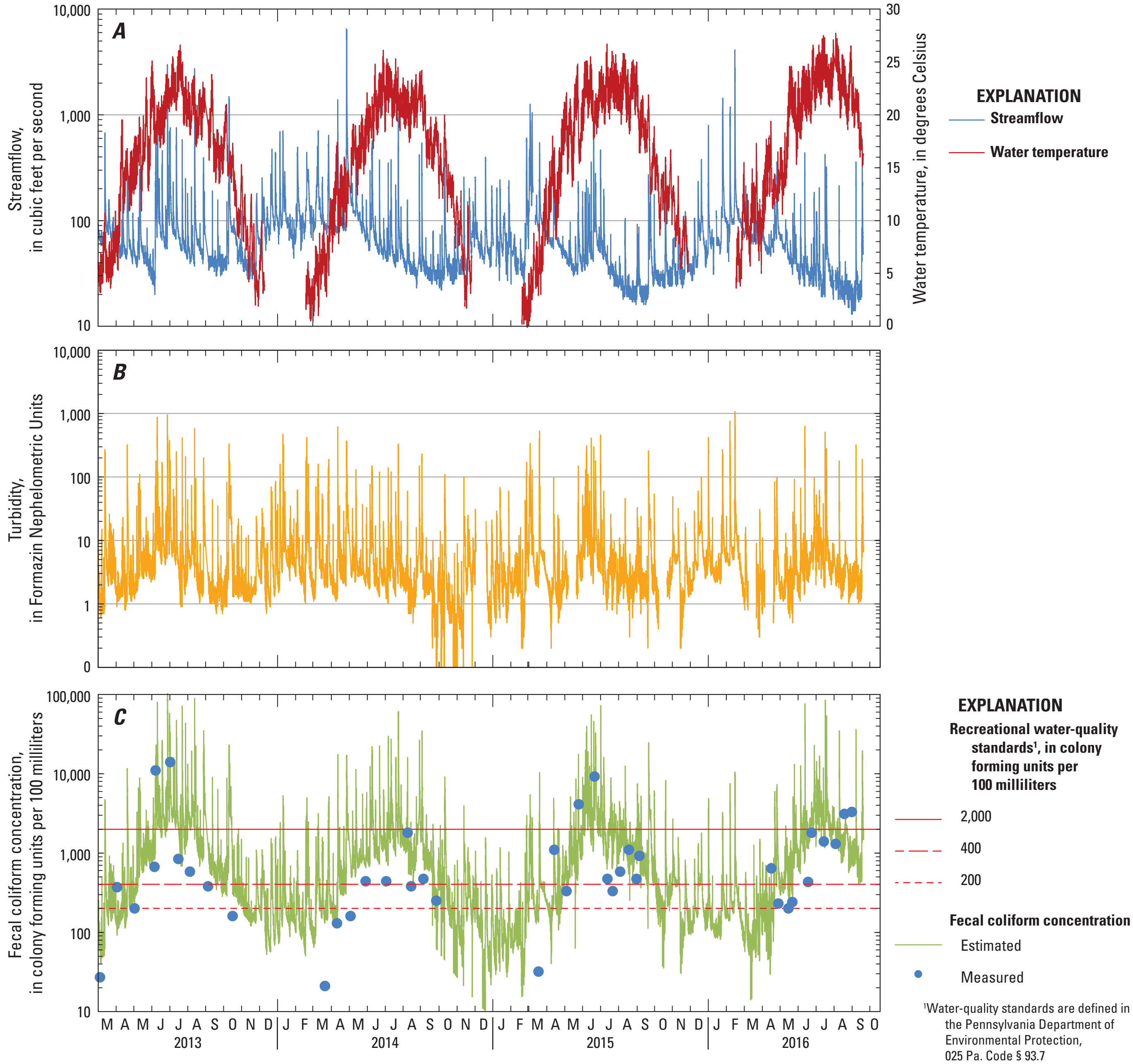


Figure 11. *A*, Measured stream temperature and streamflow, *B*, measured turbidity, and *C*, estimated (computed by using regression equation and bias correction factor) and measured fecal coliform concentrations for West Branch Brandywine Creek at Modena, Pa., 2013–16.

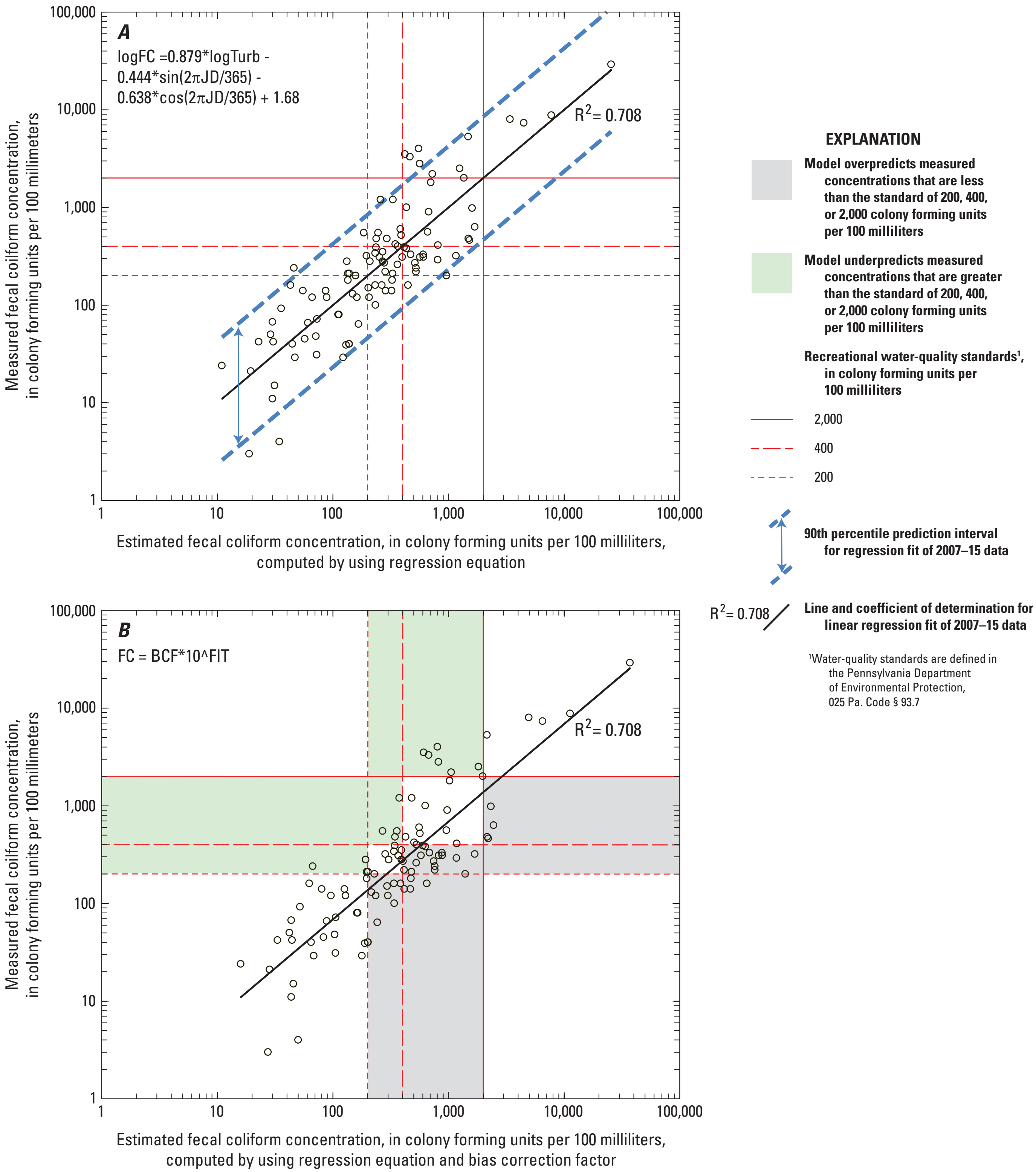


Figure 12. Relations between *A*, estimated and measured log fecal coliform concentrations and *B*, estimated (corrected for bias) and measured fecal coliform concentrations for East Branch Brandywine Creek below Downingtown, Pa., 2007–15. Estimated log fecal coliform (FC) concentrations were computed by using regression equation. A bias correction factor (BCF) of 1.45 was applied to the computed log FC concentrations after retransformation to original units as shown in *B*.

cooler spring months, as indicated by measured concentrations for recent (2013–16; fig. 13) and prior periods (Town, 2001), which is similar to the pattern observed at West Branch Brandywine Creek at Modena. Measured and computed FC concentrations were greater than 200 cfu/100 mL for most of the June through September period each year from 2013 to 2016 (figs. 5*B* and 13*C*) at East Branch Brandywine Creek below Downingtown. Additionally, measured and computed FC concentrations commonly were greater than the higher standards of 400 cfu/100 mL during June through August and 2,000 cfu/100 mL during July each year at East Branch Brandywine Creek below Downingtown.

Brandywine Creek at Chadds Ford

Of the five explanatory variables considered, for models that include logTurb, only Temp, sin(2πJD/365), and cos(2πJD/365) (table 4) were determined to be statistically significant ($p < 0.05$). The best fit was obtained by a regression model using logTurb and the seasonality terms, sin(2πJD/365) and cos(2πJD/365), as explanatory variables (table 5) where logFC = 1.1*logTurb - 0.288*sin(2πJD/365) - 0.774*cos(2πJD/365) + 1.12, which resulted in an adjusted R^2 of 0.759 (table 4). The 90th percentile prediction interval (fig. 14*A*) is a factor of about +/-4 for FC concentration in units of cfu/100 mL. Detailed information about data used in, and the statistics for, the best-fit regression model for Brandywine Creek at Chadds Ford is given in appendix 3. The retransformed regression model adjusted using the Duan's BCF of 1.40 estimates FC concentrations in original units of cfu/100 mL as follows: $1.40 \times 10^{\text{logFC}}$, where logFC is the estimate from the regression equation.

Computed FC concentrations adjusted by applying Duan's BCF tend to be overestimated more than underestimated relative to the standards of 200, 400, and 2,000 cfu/100 mL (fig. 14*B*). The regression model underestimated about 8 percent of 101 measured FC concentrations relative to at least one, but typically not more than one, standard. For these "false negative" results, about 1 percent of underestimations were less than measured values greater than 200 cfu/100 mL, about 2 percent of underestimations were less than measured values greater than 400 cfu/100 mL, and about 5 percent of underestimations were less than measured values greater than 2,000 cfu/100 mL (see shaded areas on fig. 14*B*; table 5). The accuracy of the model to predict values exceeding or not exceeding the standards of 200, 400, and 2,000 cfu/100 mL was 0.77 (77 percent), 0.84 (84 percent), and 0.95 (95 percent), respectively.

The regression model estimated that FC concentrations were higher during the warm summer months than during the cooler spring months, as indicated by measured concentrations for recent (2013–16; fig. 15) and prior periods (Town, 2001), which is similar to the patterns observed at West Branch Brandywine Creek at Modena and East Branch Brandywine Creek below Downingtown. Computed and, to a lesser extent, measured FC concentrations were greater than 200 cfu/100 mL for most of the May through August months each year (figs. 5*C* and 13*C*) at Brandywine Creek at Chadds Ford. Additionally, measured and computed FC concentrations commonly were greater than 400 cfu/100 mL and occasionally were greater than 2,000 cfu/100 mL during July each year at Brandywine Creek at Chadds Ford. Computed and measured FC concentrations at Brandywine Creek at Chadds Ford exceed the standards of 200, 400, or 2,000 cfu/100 mL less frequently than at either of the upstream sites West Branch Brandywine Creek at Modena and East Branch Brandywine Creek below Downingtown (figs. 11*C*, 13*C*, 15*C*; table 5).

French Creek Near Phoenixville

Of the five explanatory variables considered, models that include logTurb, Temp, and sin(2πJD/365) were determined to be statistically significant ($p < 0.05$; table 4). The lack of statistical significance for the other seasonality variable, cos(2πJD/365), is equivalent to including the cosine term in the equation with a zero value coefficient, resulting in a shift in peak seasonal effects of increasing FC concentrations towards the end of September (fig. 9). The best fit was obtained by a regression model using logTurb and the seasonality term sin(2πJD/365) as explanatory variables (table 5) where logFC = 1.25*logTurb - 0.543*sin(2πJD/365) + 1.4, which resulted in an adjusted R^2 of 0.684, indicating a fit similar to those for the three Brandywine Creek sites despite fewer available measured values (table 4). The 90th percentile prediction interval (fig. 16*A*) is a factor of about +/-5 for FC concentration in units of cfu/100 mL, which is similar to the prediction intervals for the three Brandywine Creek stream sites. Detailed information about data used in, and the statistics for, the best-fit regression model for French Creek near Phoenixville is given in appendix 4. The retransformed regression model adjusted using the BCF of 1.57 estimates FC concentrations in original units of cfu/100 mL as follows: $1.57 \times 10^{\text{logFC}}$, where logFC is the estimate from the regression equation.

Computed FC concentrations adjusted by applying Duan's BCF tend to overestimate more than underestimate FC concentrations relative to the standards of 200, 400, and 2,000 cfu/100 mL (fig. 16*B*). The regression model underestimated about 8 percent of 37 measured FC concentrations relative to at least one of two standards (400 and 2,000 cfu/100 mL), but typically not more than one of these standards. For these "false negative" results, about 5 percent of underestimations were less than measured values greater than 400 cfu/100 mL and about 3 percent of underestimations were less than measured values greater than 2,000 cfu/100 mL (see shaded areas on fig. 16*B*; table 5). The accuracy of the French Creek near Phoenixville model to predict values exceeding or not exceeding the standards of 200, 400, and 2,000 cfu/100 mL was 0.84 (84 percent), 0.87 (87 percent), and 0.97 (97 percent), respectively. The model had a higher number of "false positive" results relative to the standards of 200 and 400 cfu/100 mL than to the standard of 2,000 cfu/100 mL (table 5).

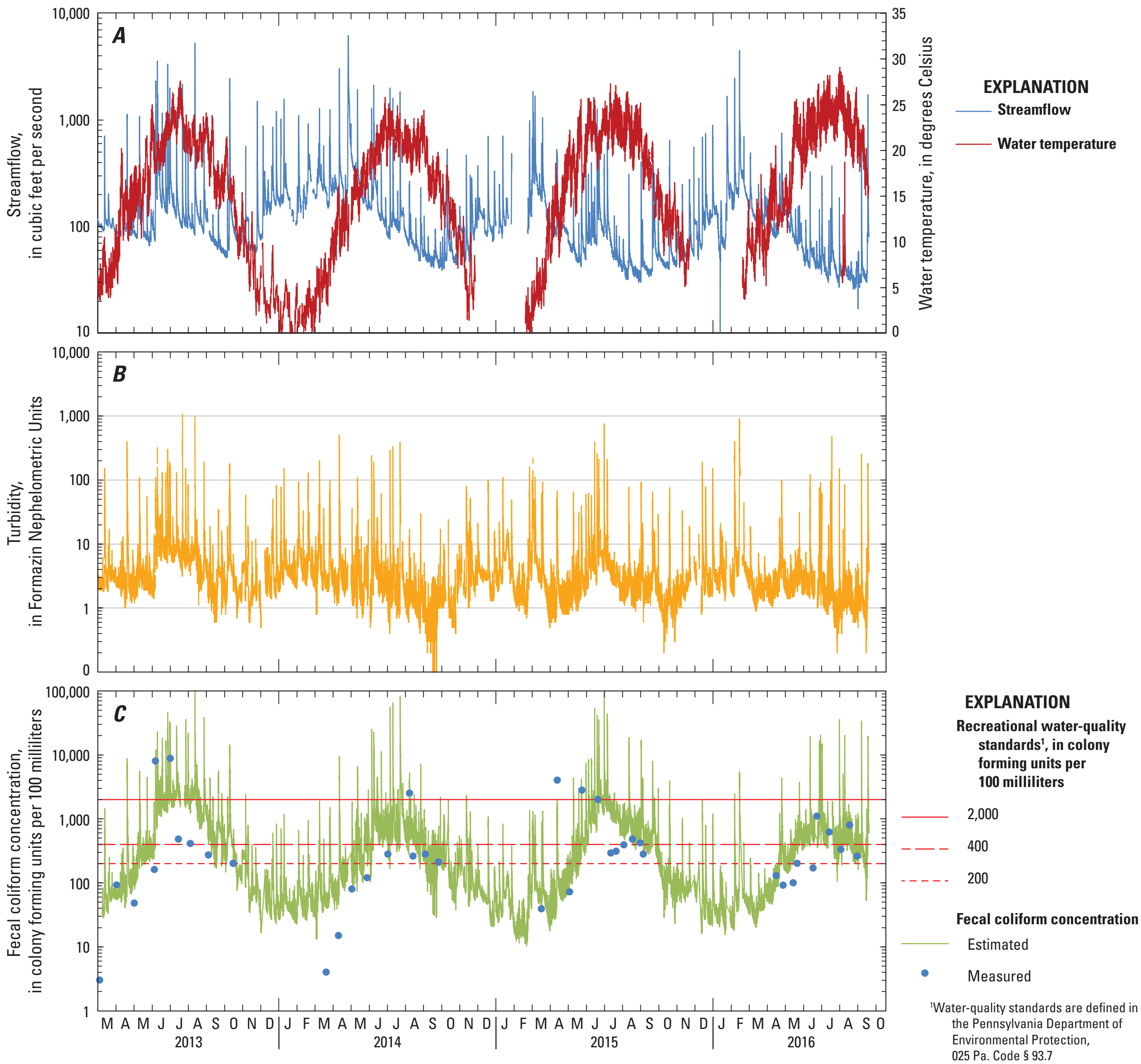


Figure 13. *A*, measured stream temperature and streamflow, *B*, measured turbidity, and *C*, estimated (computed by using regression equation and bias correction factor) and measured fecal coliform concentrations for East Branch Brandywine Creek below Downingtown, Pa., 2013–16.

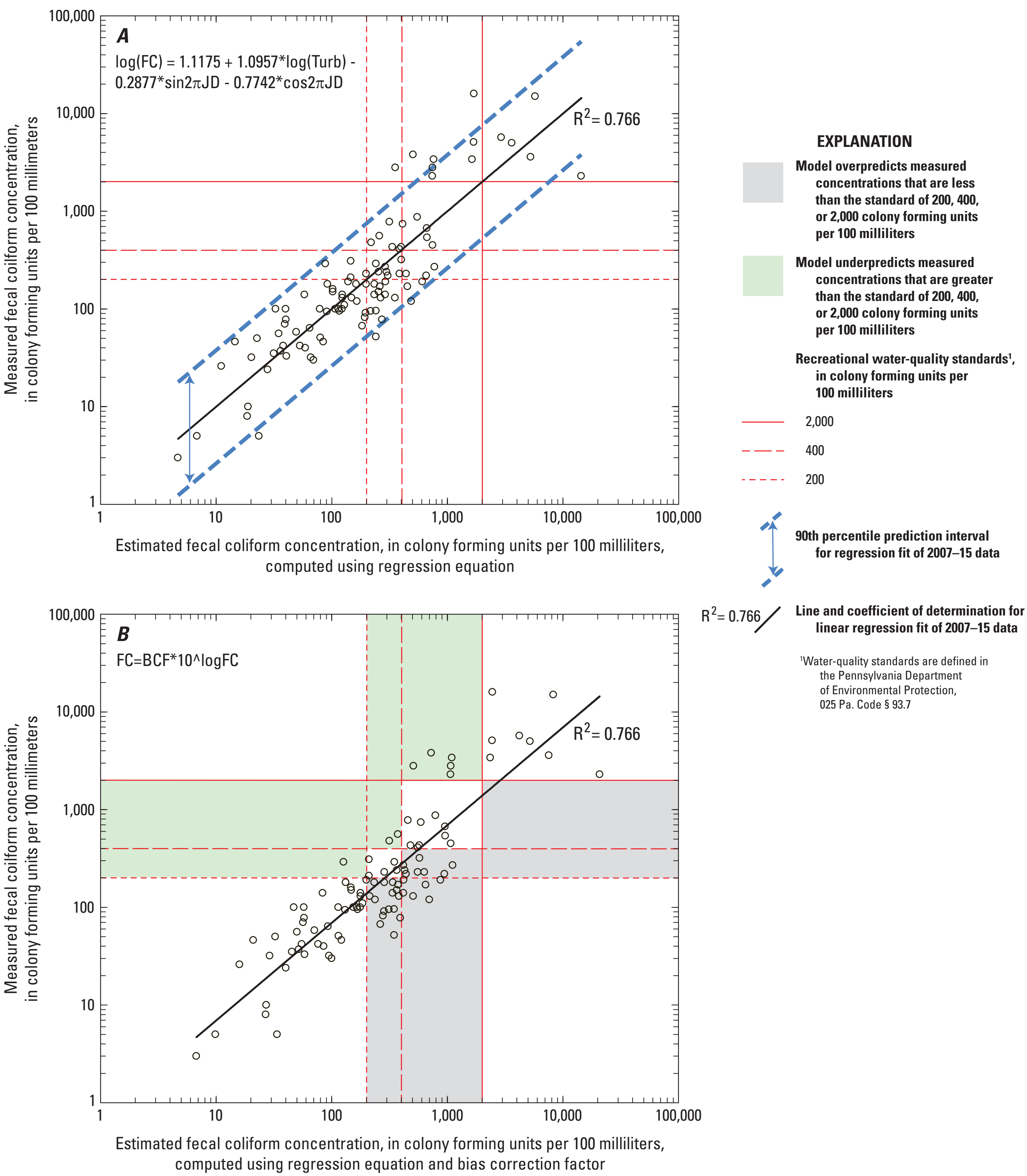


Figure 14. Relations between *A*, estimated and measured log fecal coliform concentrations and *B*, estimated (corrected for bias) and measured fecal coliform (FC) concentrations for Brandywine Creek at Chadds Ford, Pa., 2007–15. Estimated log fecal coliform concentrations were computed by using regression equation. A bias correction factor (BCF) of 1.40 was applied to the computed log fecal coliform concentrations after retransformation to original units as shown in *B*.

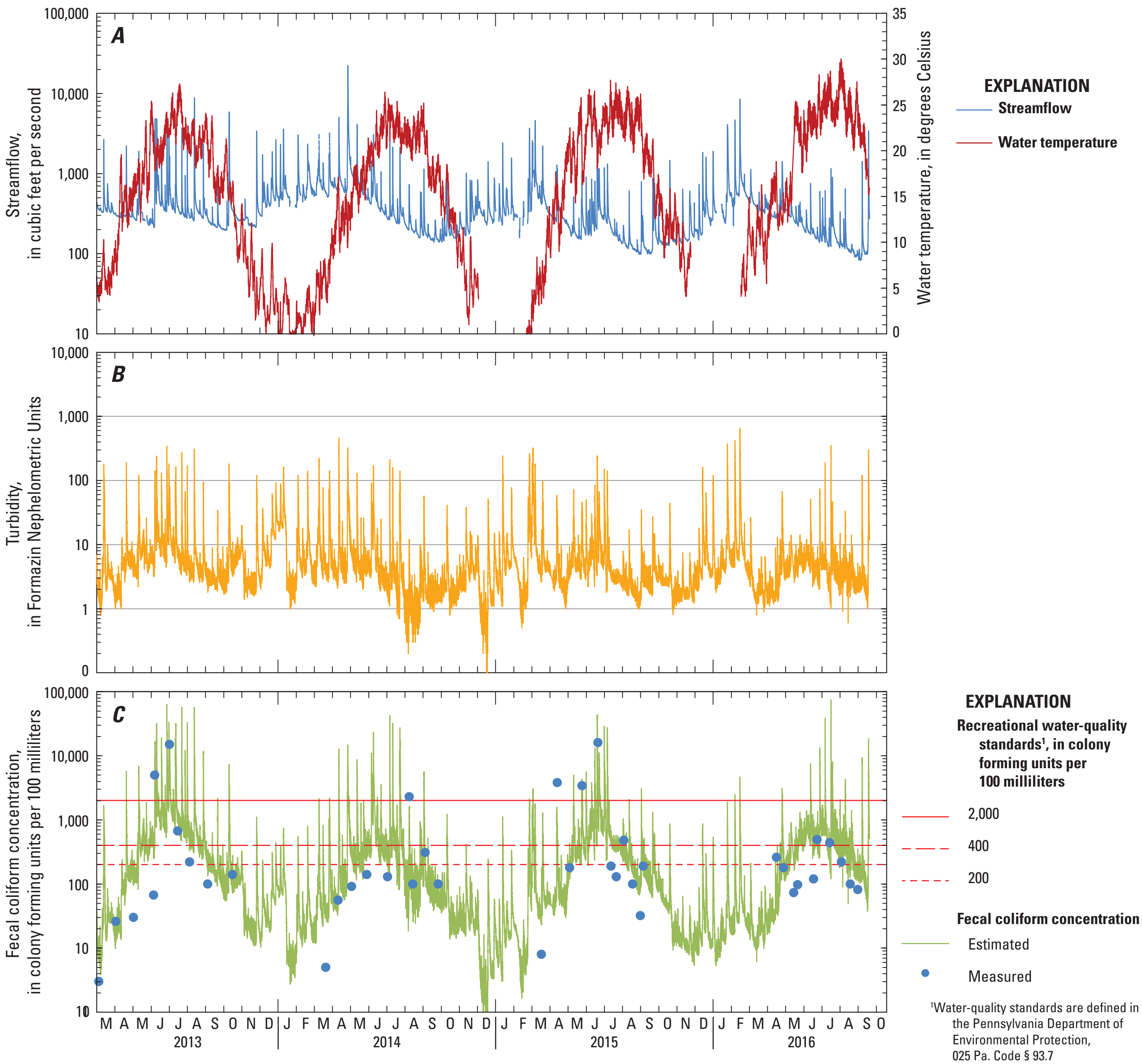


Figure 15. *A*, Measured stream temperature and streamflow, *B*, measured turbidity, and *C*, estimated (computed by using regression equation and bias correction factor) and measured fecal coliform concentrations for Brandywine Creek at Chadds Ford, Pa., 2013–16.

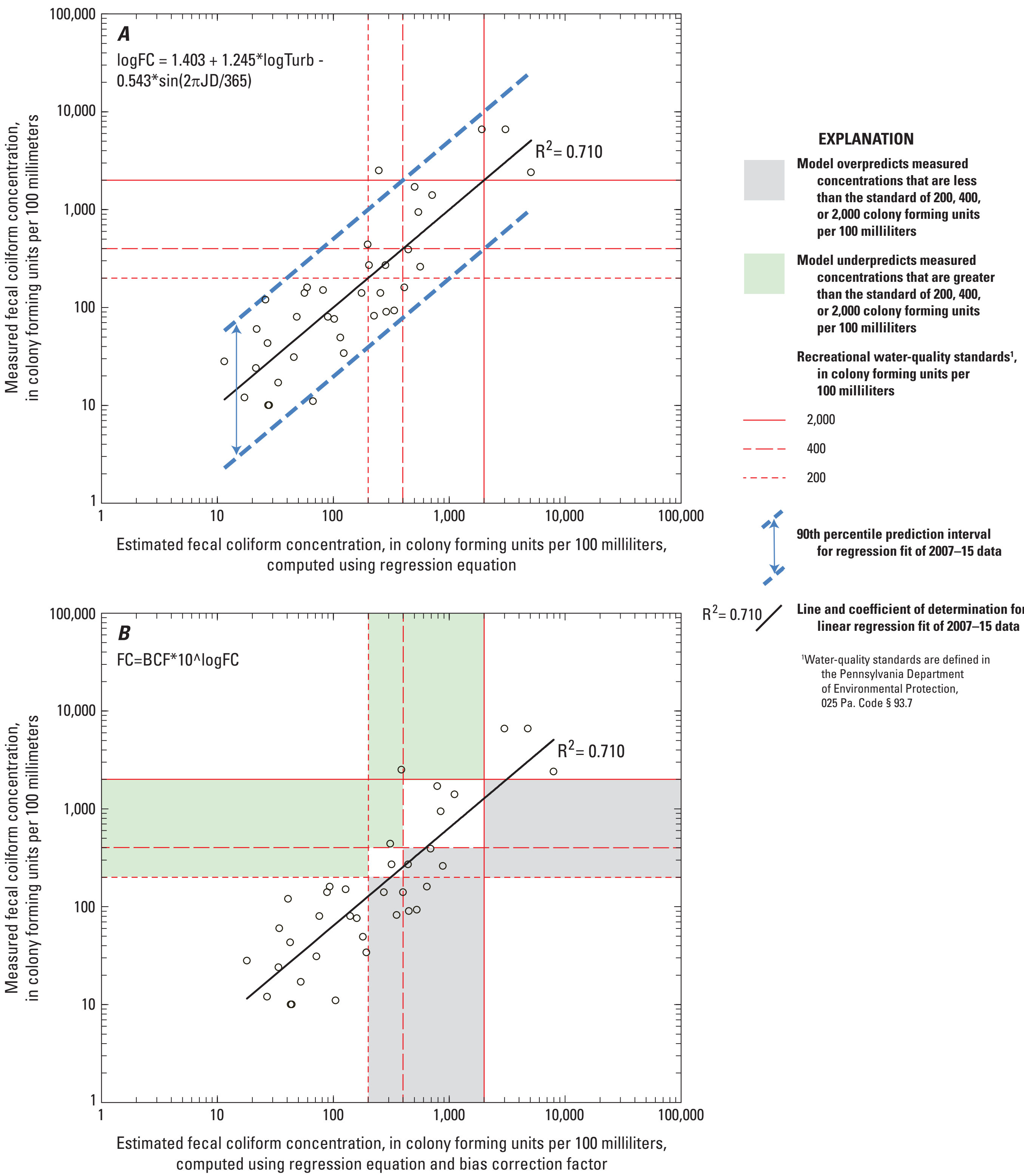


Figure 16. Relations between *A*, estimated and measured log fecal coliform concentrations and *B*, estimated (corrected for bias) and measured fecal coliform concentrations for French Creek near Phoenixville, Pa., 2012–15. Estimated log fecal coliform concentrations were computed by using regression equation. A bias correction factor (BCF) of 1.75 was applied to the computed log fecal coliform concentrations after retransformation to original units as shown in *B*.

The regression model estimated that FC concentrations were higher during the warm summer months than during the cooler spring months, as indicated by concentrations measured during March 2013–September 2016 (fig. 17), which is similar to the patterns observed at the three sites on Brandywine Creek but with a shift in higher bacteria concentrations to later in the summer. Computed and, to a lesser extent, measured FC concentrations were greater than 200 cfu/100 mL for most of July through September each year (figs. 6*A* and 17*C*) at French Creek near Phoenixville. Additionally, measured and computed FC concentrations commonly were greater than 400 cfu/100 mL and occasionally were greater than 2,000 cfu/100 mL during July through September each year.

White Clay Creek Near Strickersville

Of the five explanatory variables considered, models that include logTurb, Temp, sin(2πJD/365), and cos(2πJD/365) (table 4) were determined to be statistically significant ($p < 0.05$) in the White Clay Creek near Strickersville model. The best fit was obtained by a regression model using logTurb and Temp as measured by RMSE and R^2 values, but the accuracy in predicting exceedances relative to water-quality standards of 200 and 400 cfu/100 mL was best for the regression model using log Turb and the seasonality terms sin(2πJD/365) and cos(2πJD/365) as explanatory variables (table 4). Therefore, given the superior accuracy and the potential limitations of using provisional water temperature data for the White Clay Creek near Strickersville gaging station, the regression model selected to estimate bacteria used log Turb and the seasonality terms (table 5) where $logFC = 0.653*logTurb - 0.281*\sin(2\pi JD/365) - 0.747*\cos(2\pi JD/365) + 1.87$, which resulted in an adjusted R^2 of 0.642 (table 4). The 90th percentile prediction interval (fig. 18*A*) is a factor of about +/-5.6 for FC concentration in units of cfu/100 mL. Detailed information about data used in, and the statistics for, the best-fit regression model for White Clay Creek near Strickersville is given in appendix 5. The retransformed regression model adjusted by using the BCF of 1.60 estimates FC concentrations in original units of cfu/100 mL as follows: 1.60×10^{logFC}, where logFC is the estimate from the regression equation.

Computed FC concentrations adjusted by applying Duan’s BCF tend to overestimate more than underestimate FC concentrations relative to the standards of 200 and 2,000 cfu/100 mL (fig. 18*B*). The regression model underestimated about 16 percent of 38 measured FC concentrations relative to at least one standard. For these “false negative” results, about 3 percent of underestimations were less than measured values above 200 cfu/100 mL, about 8 percent of underestimations were less than measured values above 400 cfu/100 mL, and about 5 percent of underestimations were less than measured values above 2,000 cfu/100 mL (see shaded areas on fig. 18*B*; table 5). The accuracy of the model to predict values exceeding the standards of 200, 400, and 2,000 cfu/100 mL was 0.84 (84 percent), 0.68 (68 percent, and 0.92 (92 percent), respectively.

The regression model estimated that FC concentrations were higher during the warm summer months than during the cooler spring months, as indicated by concentrations measured during March 2012–September 2016 (fig. 19), which is similar to the patterns observed at the three sites on Brandywine Creek. Computed and, to a lesser extent, measured FC concentrations were greater than 200 cfu/100 mL for most of May through September each year (figs. 6*B* and 19*C*) at White Clay Creek near Strickersville. Additionally, measured and computed FC concentrations commonly were greater than 400 cfu/100 mL during June through September and often were greater than 2,000 cfu/100 mL during July and August each year.

Model Validation

Validation of regression equations developed using 2007–15 data for the three Brandywine sites and 2012–15 data for the French Creek and White Clay Creek sites was done by comparing estimated and measured FC concentrations using data collected from March through September 2016. Generally, the estimated FC concentrations computed using regression equations and the 2016 data fit within the 90th percentile prediction intervals for the regression models (figs. 20 and 21).

The time series plots of estimated and measured FC concentrations during 2016, where estimated values were computed by using regressions for available data prior to 2016 (2007–15 for the three Brandywine Creek sites and 2012–15 for French Creek and White Clay Creek sites), show that the estimated FC concentrations are greater than the recreational standard of 200 cfu/100 mL starting about mid-May (figs. 22 and 23) and remain elevated throughout the summer of 2016. Of the five sites, estimated FC concentrations generally were highest at West Branch Brandywine Creek at Modena, which is consistent with measured values. Although the estimated FC concentrations generally followed the temporal patterns indicated by measured concentrations, the estimated concentrations commonly were higher than the measured concentrations.

Because factors affecting bacteria mobilization may change through time upstream from a stream sampling site, regressions models should be evaluated periodically and potentially revised using new and recent data. Revision of models is done to improve model fit and predictive power. Therefore, to maintain the models, there is a need for continuing data collection that would consist of at least 8–10 samples collected for bacteria analysis at fixed time intervals per year and under a range of hydrologic conditions. More frequent sample collection, such as once per week, as had been done until 2010 at the three Brandywine Creek sites would provide a larger dataset for greater characterization of actual conditions and therefore likely would produce higher confidence in the regression results.

Discussion of Uncertainty and Other Limitations

Errors or uncertainty in measured values for explanatory or response variables increases uncertainty in the regression and resulting estimated values. The largest measurement errors and (or) uncertainty are associated with turbidity and FC concentrations. Point measurements, such as the discrete grab samples for bacteria or the fixed-time interval values for stream discharge, turbidity, and other continuously collected data, may not accurately represent temporal or spatial variability within the stream. Also, uncertainty or error may be introduced into model output because of factors affecting values entered into in the regression models, such as changes in or fouling of instrumentation (sensors) used to measure turbidity.

Although water temperature was not included in final regression equations for any of the five stream sites, the quality of some of the data may have introduced some additional uncertainty into preliminary regression analysis. Water temperature data collected for the purpose of turbidity sensor compensation at two of the gaging stations, White Clay Creek near Strickersville and French Creek near Phoenixville, are considered provisional and uncorrected for any errors or problems that occurred during 2012–16, which may affect regressions using measured values for this period. In particular, the temperature sensor was known to be or suspected of being out of the water or only partially submerged during extreme low-flow conditions that can occur for a few days or weeks during late summer at French Creek near Phoenixville. However, independent temperature measurements indicate that the provisional temperature data overall are relatively accurate (fig. 4). After October 2016, stream temperature data at White Clay Creek near Strickersville and French Creek near Phoenixville were reviewed and released as official record, which should reduce any problems in possible future evaluations that might have been associated with the use of past provisional data.

In application of the regression equation with log-transformed variables, the BCF can overcorrect the predicted value from the regression equation, as discussed by Rasmussen and others (2009). For the regression equations developed for the five stream sites in this study, the BCF ranged from about 1.38 to 1.75. However, because larger values for the BCF are related to larger residuals (and higher uncertainty) of the regression model, improving model fit and reducing uncertainty in the regression model, such as through inclusion of more data, may result in a smaller BCF.

The temporal distribution of the measured FC values from March to October each year limits the applicability of the regression equations to those months. The sparseness of FC data also contributes to the uncertainty of the regression equations. For example, only 8–10 samples have been collected during the sampling period March–October each year since 2012. The small number of samples collected each year is unlikely to represent the actual range of values that occur throughout hydrologic conditions. The relations among FC concentrations and explanatory variables may differ during base flow, rising flow, and falling flow or recession conditions. Inclusion of additional data that has been or will be collected after 2015 may improve model fits in potential subsequent regression analysis, especially at the two sites (French Creek near Phoenixville and White Clay Creek near Strickersville) with the fewest observations (measured FC concentrations).

Regression equations used to estimate FC concentrations at specific sites on a stream may not be representative of FC concentrations in reaches upstream or downstream from those sites. The spatial distribution of FC bacteria concentrations in the Brandywine Creek Basin during 1998–99 is described by Town (2001).

Application for Near Real-Time Estimates

The regression equations determined to best fit measured data can be used to provide near real-time estimates of FC concentrations within the range estimated by prediction intervals for March–October. For example, the near real-time estimates of FC concentrations and the relation of these concentrations to recreational water-quality standards or criteria can be updated hourly for display periods of days, weeks, or months. This approach is currently used by USGS for streams in numerous states through the on-line web site for National Real-Time Water Quality (nrtqw; https://nrtwq.usgs.gov/), which provides the real-time estimate and associated uncertainty for constituents, such as FC concentrations, determined using surrogate data, such as turbidity. However, because of the uncertainty associated with the regression model, one application may be the use of the model to predict whether bacteria concentrations are greater than or less than the PADEP standards or the EPA recreational water-quality criteria, rather than the determination of a precise value. The estimates from the regression equations presented in this report will be used to support the display on the web of near real-time bacteria concentrations in relation to PADEP recreational water-quality standards for the general public and cooperating agencies.

The regression models developed using available data from 2007 to 2015 for the three Brandywine Creek sites and from 2012 to 2015 for French Creek and White Clay Creek sites tend to overestimate concentrations that are greater than the PADEP recreational water-quality standards of 200, 400, and 2,000 cfu/100 mL more often than underestimate concentrations less than those standards, thus providing a cautious estimate of human-health risks associated with recreational use of those waters under certain conditions.

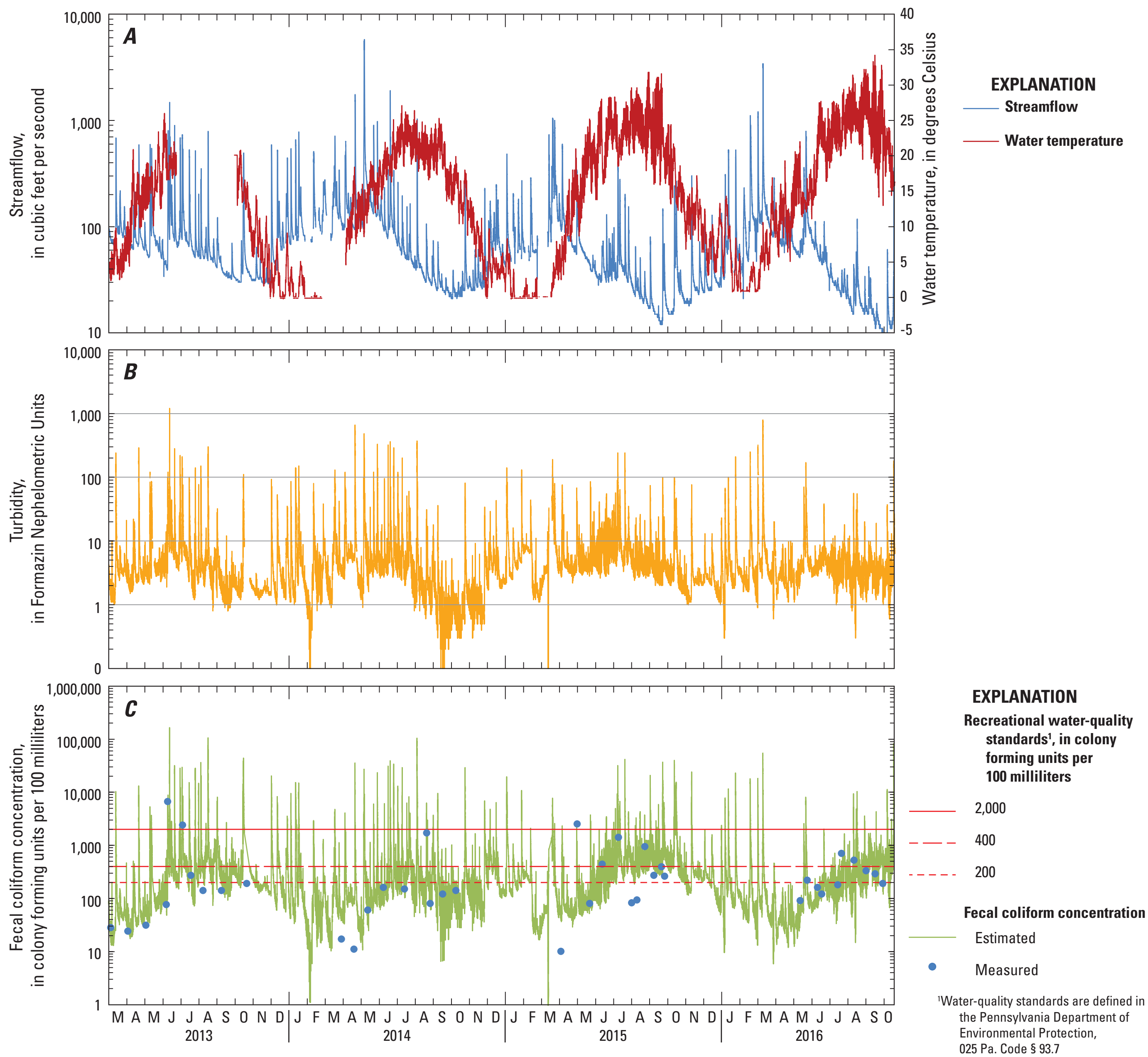


Figure 17. *A*, Measured stream temperature and streamflow, *B*, measured turbidity, and *C*, estimated (computed by using regression equation and bias correction factor) and measured fecal coliform concentrations for French Creek near Phoenixville, Pa., 2013–16.

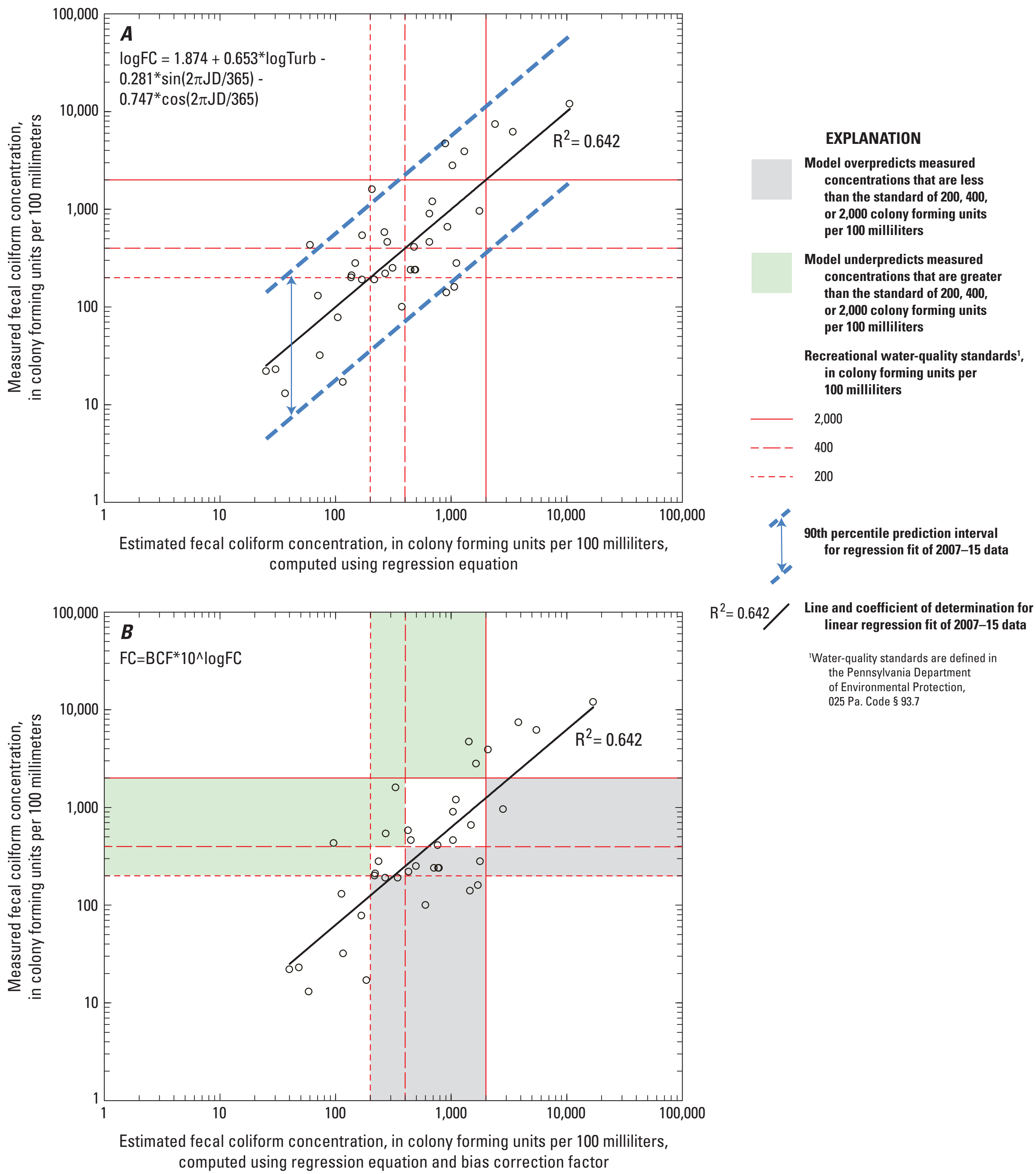


Figure 18. Relations between *A*, estimated and measured log fecal coliform concentrations and *B*, estimated (corrected for bias) and measured fecal coliform concentrations for White Clay Creek near Strickersville, Pa., 2012–15. Estimated log fecal coliform concentrations were computed by using regression equation. A bias correction factor (BCF) of 1.60 was applied to the computed log fecal coliform concentrations after retransformation to original units as shown in *B*.

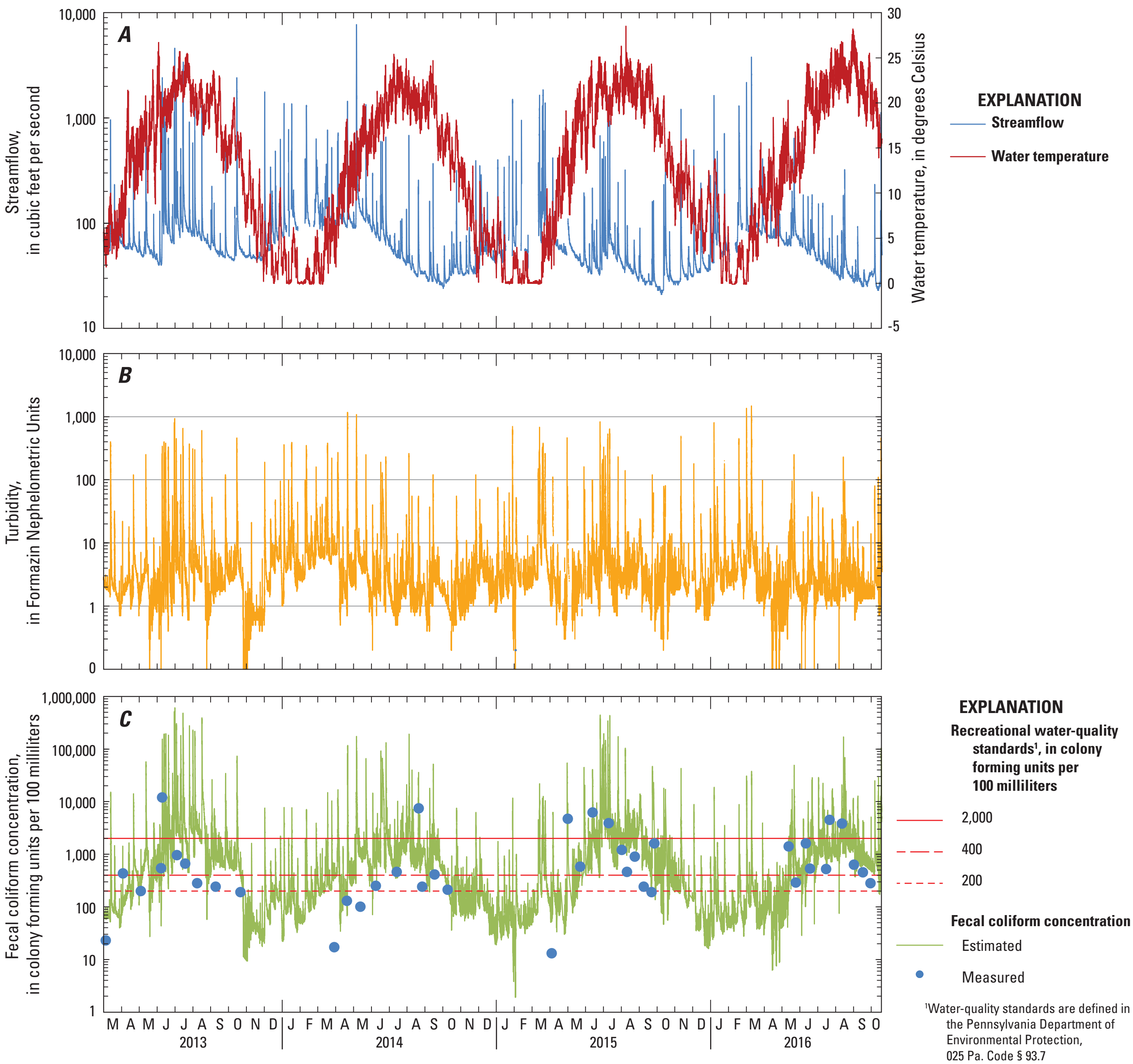


Figure 19. *A*, Measured stream temperature and streamflow, *B*, measured turbidity, and *C*, estimated (computed by using regression equation and bias correction factor) and measured fecal coliform concentrations for White Clay Creek near Strickersville, Pa., 2013–16.

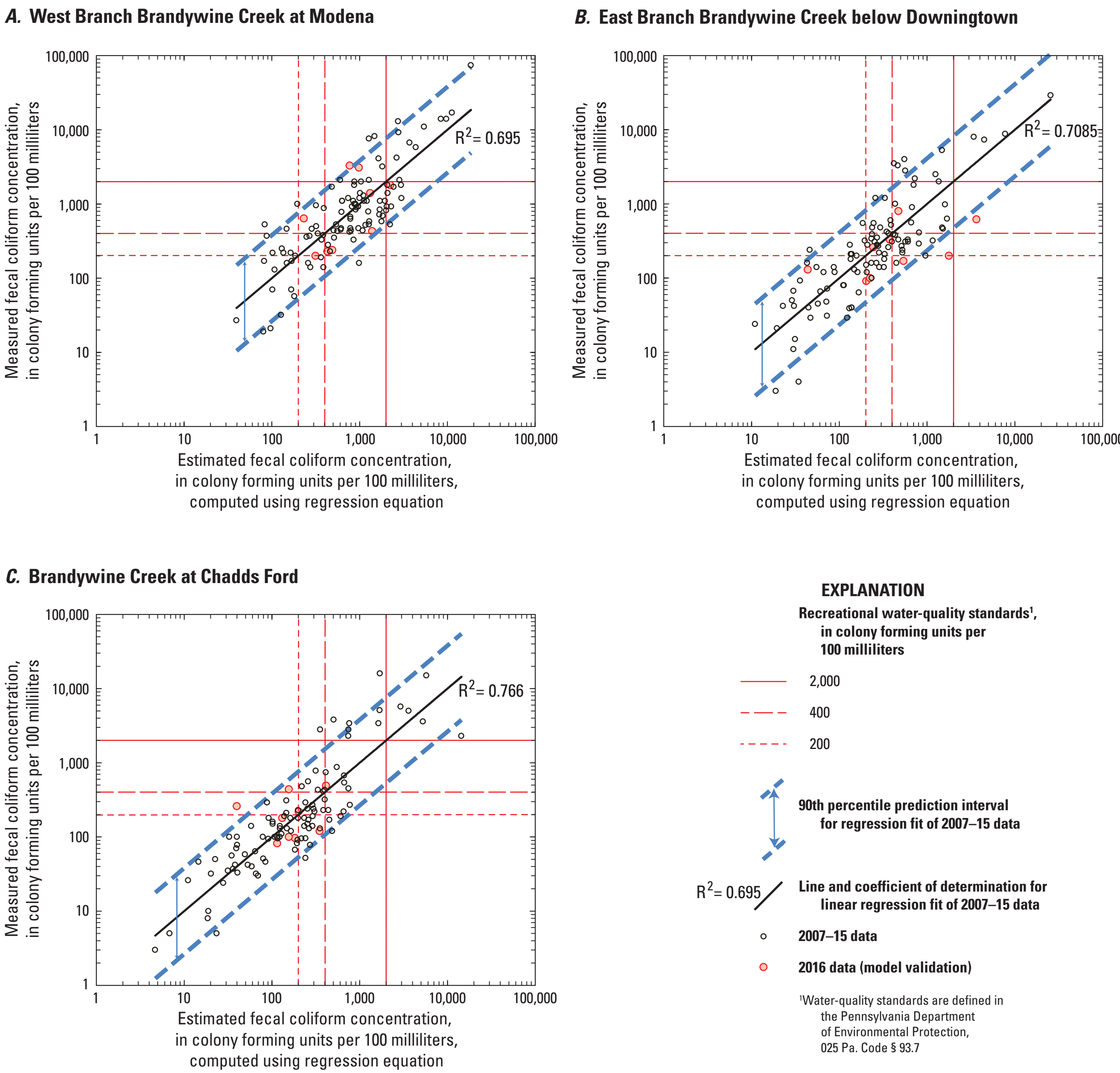


Figure 20. Relation between estimated and measured log fecal coliform concentrations for 2016 data, where estimated values were computed by using regression equations established for available 2007–15 data for *A*, West Branch Brandywine Creek at Modena, Pa., *B*, East Branch Brandywine Creek below Downingtown, Pa., and *C*, Brandywine Creek at Chadds Ford, Pa. The unadjusted coefficient of determination (R^2) shown is for linear regression of 2007–15 data.

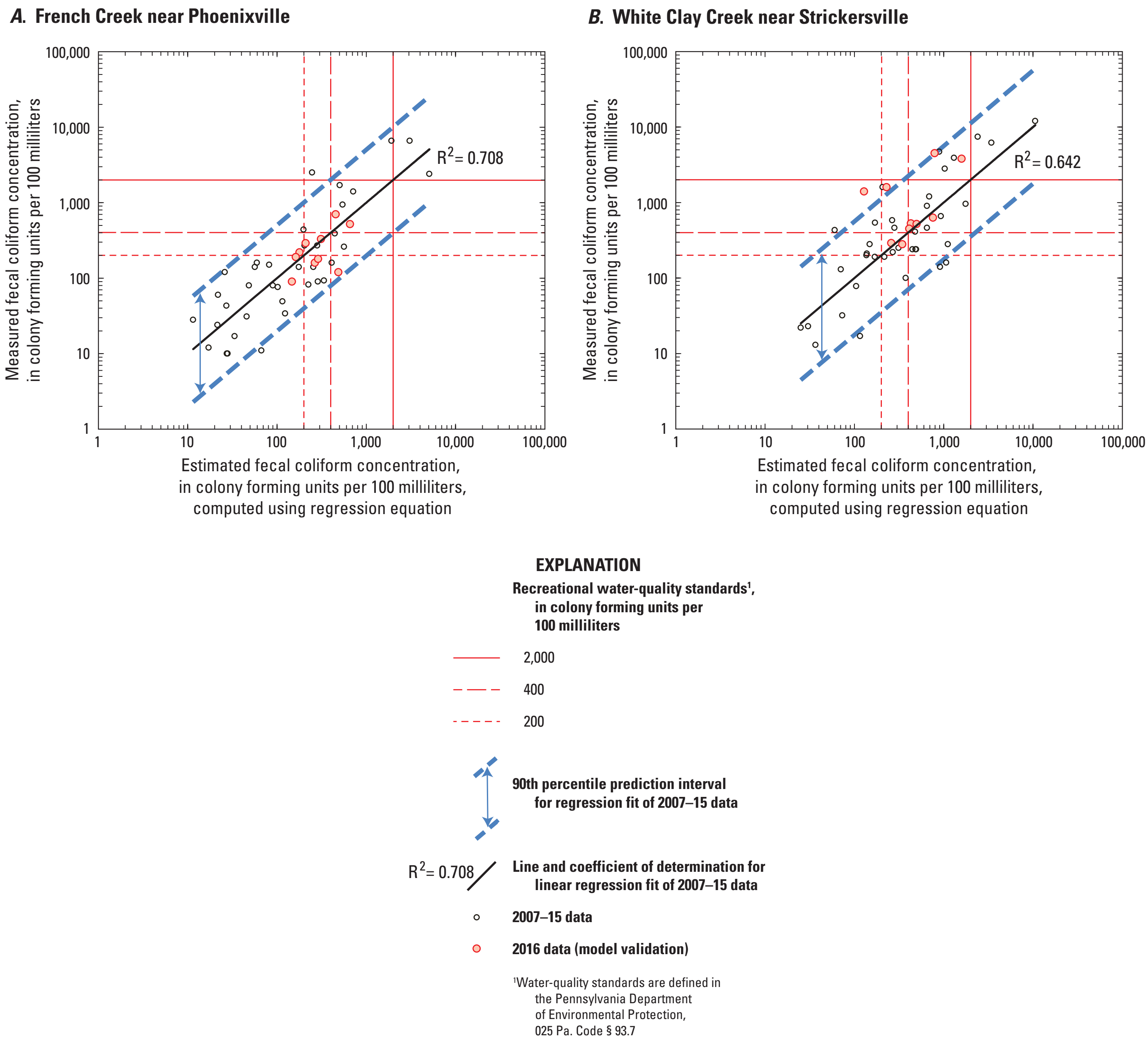


Figure 21. Relation between estimated and measured log fecal coliform concentrations for 2016 data, where estimated values were computed by using regression equations established from available 2012–15 data for *A*, French Creek near Phoenixville, Pa., and *B*, White Clay Creek near Strickersville, Pa.

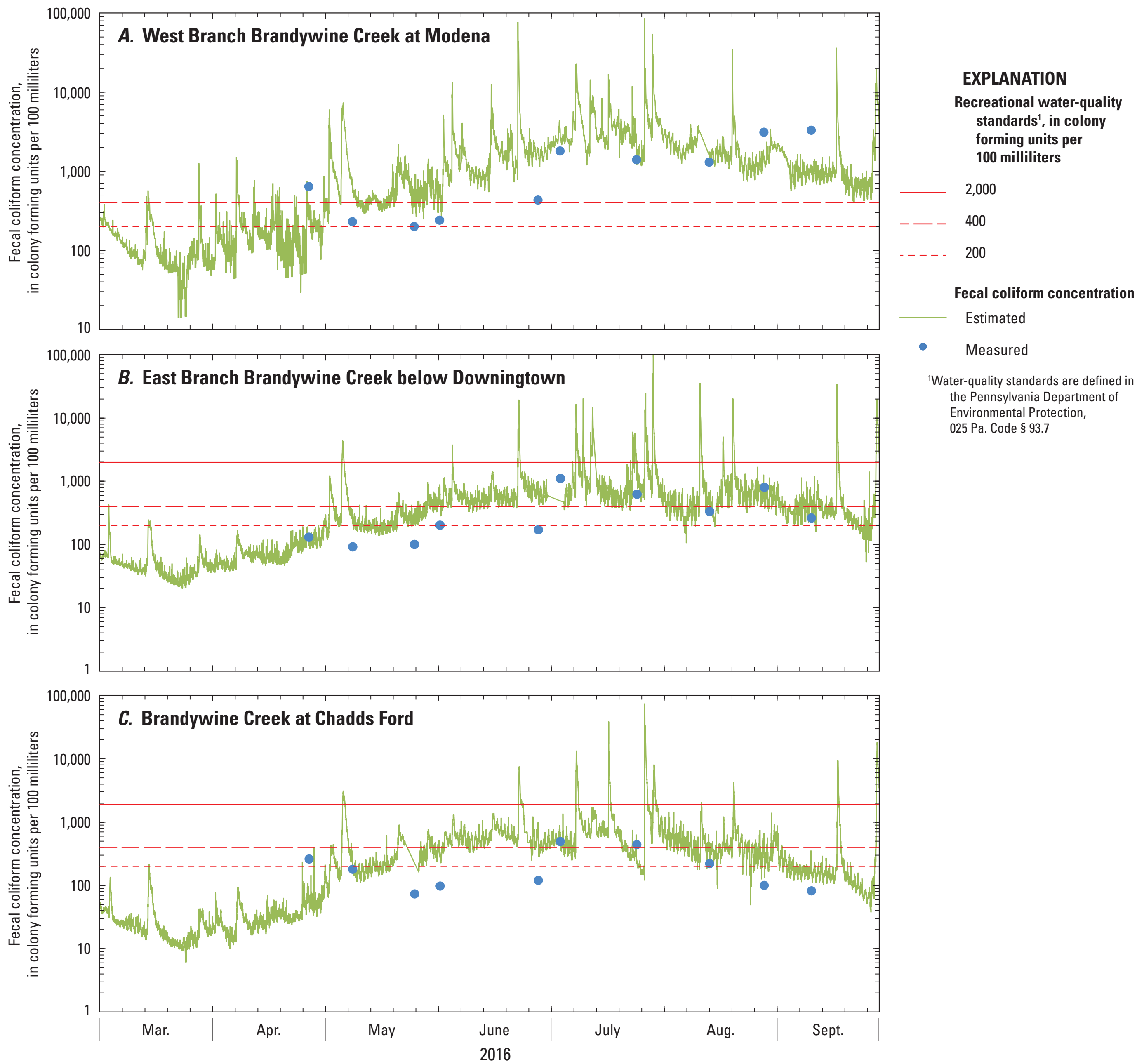


Figure 22. Estimated and measured log fecal coliform concentrations for 2016 data, where estimated values were computed by using bias correction factors and regression equations established from available 2007–15 data for *A*, West Branch Brandywine Creek at Modena, Pa., *B*, East Branch Brandywine Creek below Downingtown, Pa., and *C*, Brandywine Creek at Chadds Ford, Pa.

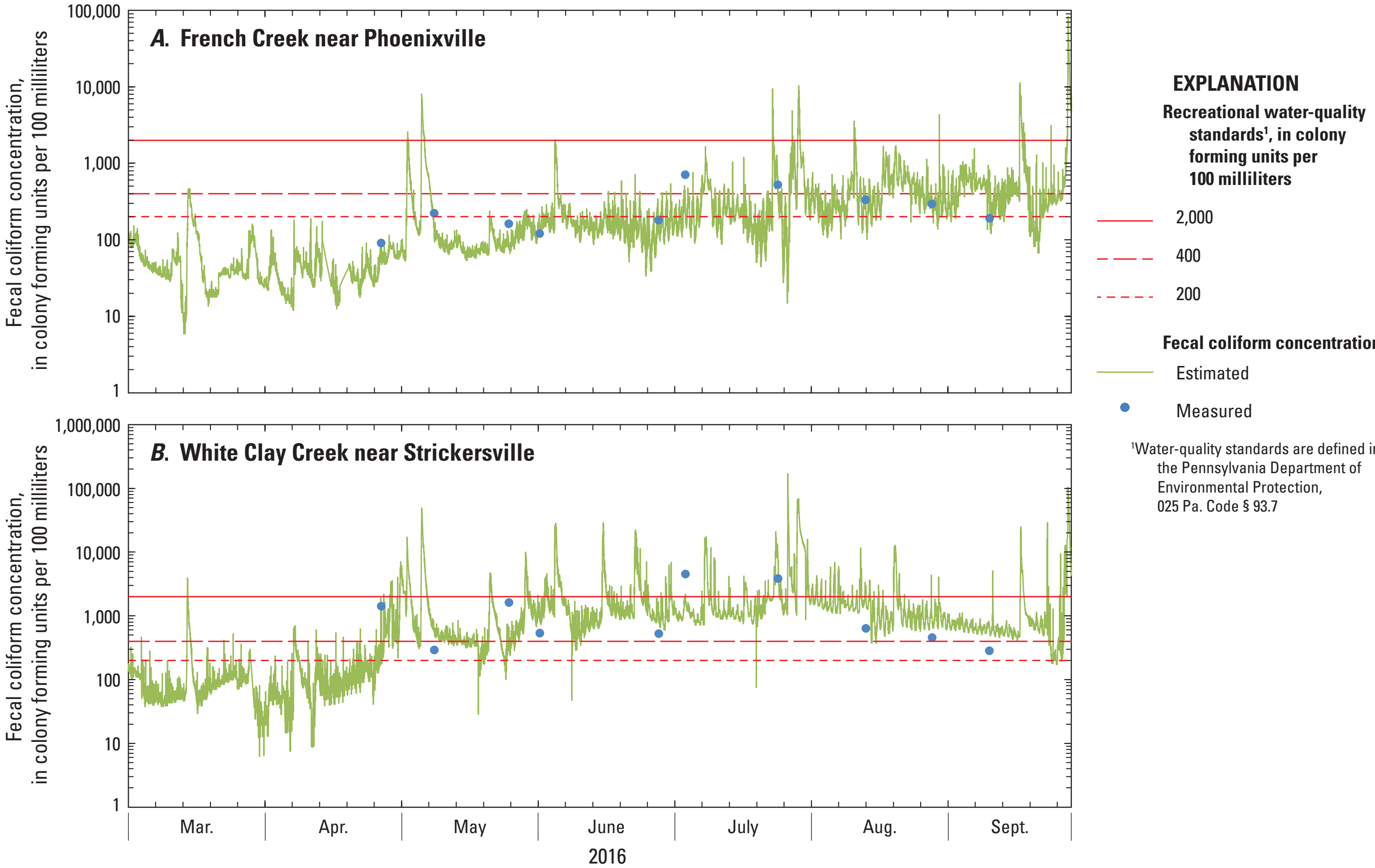


Figure 23. Estimated and measured log fecal coliform concentrations for 2016 data, where estimated values were computed by using bias correction factors regression equations established from available 2012–15 data for *A*, French Creek near Phoenixville, Pa., and *B*, White Clay Creek near Strickersville, Pa.

Summary and Conclusions

Several streams used for recreational activities, including fishing, swimming, and boating, in Chester County, Pennsylvania, are known to have periodic elevated concentrations of fecal coliform (FC) bacteria, a type of fecal-indicator bacteria. Fecal-indicator bacteria are not necessarily disease causing (pathogenic), but their presence in streams is indicative of the potential presence of feces and associated pathogens that may pose human-health risks to people exposed through water contact. One of the most common methods used to determine FC concentrations involves membrane filtration of a grab sample and subsequent incubation for 24 hours. This method introduces at least a 1-day delay in assessing bacteria concentrations, and because of the type of sample (grab), this method has the limitation of determining bacteria concentrations at a single specific point in time. Interest in developing near real-time estimates of FIB concentrations to improve protection of human health from exposure to potential fecally derived pathogens associated with FIB during recreational use of surface waters has resulted in studies that evaluated surrogates, such as turbidity, that can be measured near real-time in situ. Availability of near real-time continuous streamflow, turbidity, and other water-quality data presents an opportunity for the data to be used as surrogates to estimate near real-time concentrations of FC bacteria, thus, providing some information to the public about associated potential human-health risks.

The U.S. Geological Survey (USGS), in cooperation with the Chester County Health Department (CCHD) and Chester County Water Resources Authority (CCWRA), has collected discrete stream samples to determine FC concentrations at or near five gaging stations from March to October. Near real-time continuous data on stream discharge, turbidity, and water temperature have been collected since 2007 at 3 stations and since 2012 at 2 stations. The three gaging stations with the longest period of record (2007–16) are West Branch Brandywine Creek at Modena, East Branch Brandywine Creek below Downingtown, and Brandywine Creek at Chadds Ford. The two other stations with a shorter and more recent period of record (2012–16) are French Creek near Phoenixville and White Clay Creek near Strickersville. Land use in each of the drainage areas upstream from the gaging stations is mixed residential, urban, forested, and agricultural. Multiple possible sources of FC bacteria are present in each basin, although identification of these sources was not part of the study. French Creek has the most forested land, and White Clay Creek has the most agricultural land.

In 2012, the USGS, in cooperation with the CCWRA and CCHD, began to develop regression equations to estimate FC concentrations using available near real-time continuous data on a preliminary basis and, in 2014, planned continued development of regression relations for data collected through 2015 and model validation with subsequent data. Regression equations included possible explanatory variables of stream discharge, turbidity, water temperature, and seasonal factors calculated using Julian Day (JD) with log transformations of selected variables. The regression equations were developed using the available data from 2007 to 2015 for the three gaging stations on Brandywine Creek (West Branch Brandywine Creek at Modena, East Branch Brandywine Creek below Downingtown, and Brandywine Creek at Chadds Ford) and from 2012 to 2015 for stations on French Creek near Phoenixville and White Clay Creek near Strickersville. FC bacteria data collected in 2016 were used to validate the equations.

Of the three Brandywine Creek sites, FC concentrations in 2007–16 grab samples generally were highest at West Branch Brandywine Creek at Modena and lowest at Brandywine Creek at Chadds Ford. Fecal coliform concentrations exceeded the Pennsylvania Department of Environmental Protection (PADEP) swimming season (May 1–September 30) water-contact standard of 200 coliform forming units per 100 milliliters (cfu/100 mL) in most samples from West Branch Brandywine Creek at Modena, but the standard was less frequently exceeded in samples from the East Branch Brandywine Creek below Downingtown and Brandywine Creek at Chadds Ford. FC concentrations in samples collected during 2012–16 at White Clay Creek near Strickersville generally were the second highest among the five sites, and like the West Branch Brandywine Creek at Modena site, concentrations exceeded the PADEP standards of 200 cfu/100 mL (swimming season) and 2,000 cfu/100 mL (rest of year) more frequently than in samples from the other sites. FC concentrations generally were lowest at Brandywine Creek at Chadds Ford and French Creek near Phoenixville.

Regression analysis using log-transformed variables resulted in best-fit equations, which used the explanatory variables of log turbidity and seasonality terms calculated using JD as a fraction of the year [$\sin(2\pi JD/365)$ and $\cos(2\pi JD/365)$] to estimate log FC concentrations at all five stream sites. The adjusted coefficient of determination for the regression equations ranged from 0.61 to 0.76. The strength of the regression equations likely was affected by the limited amount and variability of data. The accuracy of the models to predict FC concentrations greater than or less than the standards at the five gaging stations ranged from 0.77 to 87 (77–87 percent) for the standard of 200 cfu/100 mL, 0.68 to 0.87 (68–87 percent) for the standard of 400 cfu/100 mL, and 0.81 to 0.97 (81–97 percent) for the standard of 2,000 cfu/100 mL. During summer months, the estimated and measured FC concentrations commonly were greater than established standards of 200 and 400 cfu/100 mL at the 5 stream sites, with concentrations typically higher at 2 sites (White Clay Creek and West Branch Brandywine Creek at Modena) than at the other 3 sites. The estimated concentrations of FC bacteria during the summer seasons commonly were higher than measured concentrations and therefore could be considered cautious estimates of potential health risk. The regression models for Brandywine Creek at Chadds Ford and French Creek near Phoenixville had relatively lower numbers of false negative values (about 8 percent each) relative to water-quality standards than the models for the other three stream sites (13–16 percent).

The regression models were developed using data on FC concentrations in samples collected from March to October each year and, thus, are limited in applicability to those months. Additional data are needed to maintain and (or) improve the use of surrogate data to estimate FC concentrations. Estimates of FC concentrations as computed from regression models for data collected at five specific stream sites do not represent FC concentrations elsewhere in streams.

Although regression equations do not estimate FC concentrations precisely, the equations can be used to estimate probable ranges of near real-time FC bacteria concentrations in selected streams in Chester County and thereby provide more timely information than results from prior-day grab samples to health officials and the public regarding potential human-health risks associated with exposure to FC bacteria in streams used for recreation during March–October. The estimates from regression equations will be used to support the display on the web of near real-time bacteria concentrations in relation to recreational water-quality standards for the general public and cooperating agencies.

References Cited

Baldwin, A.K, Robertson, D.M., Saad, D.A., and Magruder, Christopher, 2012, Refinement of regression models to estimate real-time concentrations of contaminants in the Menonomee River Drainage Basin, Southeastern Wisconsin, 2008–11: U.S. Geological Survey Scientific Investigations Report 2013–5174, 113 p.

Cinotto, P.J., 2005, Occurrence of fecal-indicator bacteria and protocols for identification of fecal contamination sources in selected reaches of the West Branch Brandywine Creek, Chester County, Pennsylvania: U.S. Geological Survey Water-Resources Investigations Report 2005–5039, 91 p.

Christensen, V.G., Jian, Xioadong, and Ziegler, A.C., 2000, Regression analysis and real-time water-quality monitoring to estimate constituent concentrations, loads, and yields in the Little Arkansas River, South-Central Kansas, 1995–99: U.S. Geological Survey Water-Resources Investigations Report 00–4126, 36 p.

Duris, J.W., Reif, A.G., Olson, L.E., and Johnson, H.E., 2011, Pathogenic bacteria and microbial-source tracking markers in Brandywine Creek Basin, Pennsylvania and Delaware, 2009–10: U.S. Geological Survey Scientific Investigations Report 2011–5165, 27 p.

Foster, G.M., and Graham, J.L., 2016, Logistic and linear regression model documentation for statistical relations between continuous real-time and discrete water-quality constituents in the Kansas River, Kansas, July 2012 through June 2015: U.S. Geological Survey Open-File Report 2016–1040, 27 p.

Francy, D.S., Brady, A.M.G., Carvin, R.B., Corsi, S.R., Fuller, L.M., Harrison, J.H., Hayhurst, B.A., Lant, Jeremiah, Nevers, M.B., Terrio, P.J., and Zimmerman, T.M., 2013, Developing and implementing predictive models for estimating recreational water quality at Great Lakes Beaches: U.S. Geological Survey Scientific Investigations Report 2013–5166, 68 p.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 510 p.

Lawrence, S.J., 2012, *Escherichia coli* bacteria density in relation to turbidity, streamflow characteristics, and season in the Chattahoochee River near Atlanta, Georgia, October 2000 through September 2008—Description, statistical analysis, and predictive modeling: U.S. Geological Survey Scientific Investigations Report 2012–5037, 81 p.

Myers, D.N., Stoeckel, D.M., Bushon, R.N., Francy, D.S., and Brady, A.M.G., 2014, Fecal indicator bacteria (ver. 2.1): U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A7, section 7.1, May 2014, accessed January 20, 2017, at http://pubs.water.usgs.gov/twri9A/.

National Environmental Methods Index, 2016, Standard Methods: 9222D: Membrane filtration test for fecal coliforms, accessed January 21, 2017, at https://www.nemi.gov/methods/method_summary/5587/.

Peel, M.C., Finlayson, B.L., and McMahon, T.L., 2007, Updated world map of the Koppen-Geiger climate classification: Hydrology and Earth Systems Sciences, v. 11, p. 1633–1644.

Pennsylvania Department of Environmental Protection, 2012, 025 Pa.Code 93.7. Specific water quality criteria, accessed May, 31 2012, at http://www.pacode.com/secure/data/025/chapter93/s93.7.html.

Rasmussen, P.P., and Ziegler, A.C., 2003, Comparison and continuous estimates of fecal coliform and *Escherichia coli* bacteria in selected Kansas streams, May 1999 through April 2002: Water-Resources Investigations Report 03–4056, 87 p.

Rasmussen, P.P., Gray, J.R., Glysson, G.D., and Ziegler, A.C., 2009, revised 2011, Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data: U.S. Geological Survey Techniques and Methods, book 3, chap. C4, 52 p.

Sauer, V.B., and Turnipseed, D.P., 2010, Stage measurement at gaging stations: U.S. Geological Survey Techniques and Methods, book 3, chap. A7, 45 p. [Also available at https://pubs.usgs.gov/tm/tm3-a7/.]

Senior, L.A., and Koerkle, E.H., 2003a, Simulation of streamflow and water quality in the Brandywine Creek Basin of the Christina River Basin, Pennsylvania and Delaware, 1994–98: U.S. Geological Survey Water-Resources Investigations Report 02–4279, 207 p.

Senior, L.A., and Koerkle, E.H., 2003b, Simulation of streamflow and water quality in the White Clay Creek Basin of the Christina River Basin, Pennsylvania and Delaware, 1994–98: U.S. Geological Survey Water-Resources Investigations Report 03–4031, 142 p.

Senior, L.A., 2017, Archival data for regression models developed to estimate fecal coliform concentrations at five stream sites, Chester County, Pennsylvania: U.S. Geological Survey data release, https://doi.org/10.5066/F79W0DC8.

Sloto, R.A., and Olson, L.E., 2011, Estimated suspended-sediment loads and yields in the French and Brandywine Creek Basins, Chester County, Pennsylvania, water years, 2008–09: U.S. Geological Survey Scientific Investigations Report 2011–5109, 31 p.

Town, D.A., 2001, Historical trends and concentrations of fecal coliform bacteria in the Brandywine Creek Basin, Chester County, Pennsylvania: U.S. Geological Survey Water-Resources Investigations Report 01–4026, 46 p.

Turnipseed, D.P., and Sauer, V.B., 2010, Discharge measurements at gaging stations: U.S. Geological Survey Techniques and Methods, book 3, chap. A8, 87 p. [Also available at https://pubs.usgs.gov/tm/tm3-a8/.]

U.S. Environmental Protection Agency, 2012, 2012 Recreational water quality criteria factsheet: EPA-820-F-12-061, December 2012, accessed January 12, 2017, at https://www.epa.gov/sites/production/files/2015-10/documents/rec-factsheet-2012.pdf.

U.S. Environmental Protection Agency, 2014, National beach guidance and performance criteria for grants, 2014 edition: EPA-823-B-14-001, accessed February 23, 2017, at https://www.epa.gov/sites/production/files/2014-07/documents/beach-guidance-final-2014.pdf.

U.S. Geological Survey, 2012, The national field manual for the collection of water-quality data (version 7): U.S. Geological Survey Techniques and Method, book 9, chaps. A1–A9, accessed April 5, 2013, at http://water.usgs.gov/owq/FieldManual/.

Wagner, R.J., Boulger, R.W., Jr., Oblinger, C.J., and Smith, B.A., 2006, Guidelines and standard procedures for continuous water-quality monitors—Station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1–D3, 51 p. + 8 attachments, accessed July 6, 2011, at http://pubs.water.usgs.gov/tm1d3.

Zimmerman, T.M., 2008, Modeling to predict *Escherichia coli* at Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania: U.S. Geological Survey Scientific Investigations Report 2008–5039, 13 p.

Appendixes 1–5.

Appendix 1. Model Archive Summary for Best-Fit Regression Developed to Estimate Fecal Coliform Concentration at Station 01480617; West Branch Brandywine Creek at Modena, Pennsylvania

Available for download at https://doi.org/10.3133/sir20175075.

Appendix 2. Model Archive Summary for Best-Fit Regression Developed to Estimate Fecal Coliform Concentration at Station 01480870; East Branch Brandywine Creek below Downingtown, Pennsylvania

Available for download at https://doi.org/10.3133/sir20175075.

Appendix 3. Model Archive Summary for Best-Fit Regression Developed to Estimate Fecal Coliform Concentration at Station 01481000; Brandywine Creek at Chadds Ford, Pennsylvania

Available for download at https://doi.org/10.3133/sir20175075.

Appendix 4. Model Archive Summary for Best-Fit Regression Developed to Estimate Fecal Coliform Concentration at Station 01472157; French Creek near Phoenixville, Pennsylvania

Available for download at https://doi.org/10.3133/sir20175075.

Appendix 5. Model Archive Summary for Best-Fit Regression Developed to Estimate Fecal Coliform Concentration at Station 01478245; White Clay Creek near Strickersville, Pennsylvania

Available for download at https://doi.org/10.3133/sir20175075.

For additional information, contact:
Director, Pennsylvania Water Science Center
U.S. Geological Survey
215 Limekiln Road
New Cumberland, PA 17070-2424

or visit our website at:
http://pa.water.usgs.gov

Prepared by USGS West Trenton Publishing Service Center

Senior—**Estimated Fecal Coliform Bacteria Concentrations Using Near Real-Time Continuous Data, Chester County, Pa., 2007–16**—Scientific Investigations Report 2017–5075, ver. 1.2

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20175075