USGS
science for a changing world

Prepared in cooperation with the Connecticut Department of Transportation

Estimating Flood Magnitude and Frequency on Streams and Rivers in Connecticut, Based on Data Through Water Year 2015

Scientific Investigations Report 2020–5054

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

Cover. Road closure in Connecticut due to flooding. Photograph by the Connecticut Department of Transportation.

Estimating Flood Magnitude and Frequency on Streams and Rivers in Connecticut, Based on Data Through Water Year 2015

By Elizabeth A. Ahearn and Glenn A. Hodgkins

Prepared in cooperation with the Connecticut Department of Transportation

Scientific Investigations Report 2020–5054

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

U.S. Department of the Interior
DAVID BERNHARDT, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2020

Suggested citation:
Ahearn, E.A., and Hodgkins, G.A., 2020, Estimating flood magnitude and frequency on streams and rivers in Connecticut, based on data through water year 2015: U.S. Geological Survey Scientific Investigations Report 2020–5054, 42 p., https://doi.org/10.3133/sir20205054.

Associated data for this publication:
Ahearn, E.A., 2020, Flood frequency and source data used in regional regression analysis of annual peak flows in Connecticut: U.S. Geological Survey data release, https://doi.org/10.5066/P9S4F751.

Ahearn, E.A., and Veilleux, A.G., 2020, Worksheet for computing annual exceedance probability flood discharges and prediction intervals at stream sites in Connecticut: U.S. Geological Survey data release, https://doi.org/10.5066/P9EWHAYW.

ISSN 2328-0328 (online)

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
inch (in.)	2.54	centimeter (cm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
square mile (mi^2)	2.590	square kilometer (km^2)
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

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

$$°C = (°F - 32) / 1.8.$$

Datum

Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD 88).

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

Elevation, as used in this report, refers to distance above the vertical datum.

Abbreviations

AEP	annual exceedance probability
AVP	average variance of prediction
B–WLS/B–GLS	Bayesian weighted-least squares/Bayesian generalized least squares
DRNAREA	drainage area, in square miles
EMA	expected moments algorithm
GIS	geographic information system
GLS	generalized least squares
I24HxxY	maximum 24-hour precipitation that occurs on average once in xx years
LTP	long-term persistence
LP3	log-Pearson type III
MGBT	multiple Grubbs-Beck test
NOAA	National Oceanic and Atmospheric Administration
NWS	National Weather Service [NOAA]
OLS	ordinary least squares
PILF	potentially influential low flow
pseudo-R^2	pseudo coefficient of determination
R^2	coefficient of determination
SOILCorD	percentage of basin area classified as hydrologic soil groups C, C/D, or D
SSURGO	Soil Survey Geographic [database]
STP	short-term persistence
USGS	U.S. Geological Survey
VIF	variance inflation factor
WREG	weighted-multiple-linear regression [program]

Estimating Flood Magnitude and Frequency on Streams and Rivers in Connecticut, Based on Data Through Water Year 2015

By Elizabeth A. Ahearn and Glenn A. Hodgkins

Abstract

The U.S. Geological Survey, in cooperation with the Connecticut Department of Transportation, updated flood-frequency estimates with 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities (2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence intervals, respectively) for 141 streamgages in Connecticut and 11 streamgages in adjacent States using annual peak-flow data through water year 2015. Peak-flow regression equations were derived for estimating flows at ungaged stream sites with annual exceedance probabilities from 50 to 0.2 percent. Methods for estimating prediction intervals for the peak-flow regression equations are presented. The regression equations are applicable for basins in Connecticut with drainage areas ranging from 0.69 to 325 square miles that are not affected by flood-control regulation or flow diversions.

The flood discharges for select annual exceedance probabilities were estimated following new (2018) national guidelines for flood-frequency analyses. New guidelines have improved statistical methods for flood-frequency analysis including (1) the expected moments algorithm to help describe uncertainty in annual peak flows and to better represent missing and historical record and (2) the generalized multiple Grubbs-Beck test to screen out potentially influential low outliers and to better fit the upper end of the peak-flow distribution. Additionally, a new regional skew (0.37) derived for New England was used in the flood-frequency analysis for the streamgages.

Annual peak flows were analyzed for trends for four time periods (30, 50, 70, and 90 years) through 2015. Trend results show some statistical evidence of increasing peak flows in each of the time periods analyzed; however, multidecadal climate cycles may be influencing the number and magnitude of the trends. Historical peak-flow trends in and near Connecticut do not offer clear and convincing evidence for incorporating trends into flood-frequency analyses. For this study, the traditional assumption of stationarity is used with no adjustment for trends.

Generalized least squares regression techniques were used to develop the final set of multivariable regression equations for estimating flood discharges with 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities. The standard error of prediction for the regional regression equations ranged from 26.3 to 45.0 percent. The standard error of prediction was slightly smaller in the current study compared to the 2004 study, indicating an improvement in the predictive ability of the equations (6 percent smaller at the 50-percent annual exceedance probability to about 1 percent smaller at the 1-percent annual exceedance probability). Generalized least squares regression techniques also were used to develop a one-variable (drainage-area-only) equation. Drainage-area-only equations can be used as an alternative to the multiexplanatory variable statewide regression equations if decreased accuracy is acceptable.

The revised statistical procedures and additional streamgage data applied in the current study result in a more accurate representation of peak-flow conditions in Connecticut than was previously available. The regional regression equations will be integrated in the U.S. Geological Survey StreamStats program, which estimates basin and climatic characteristics and streamflow statistics at user-selected ungaged stream sites.

Introduction

Flooding in Connecticut has caused millions of dollars in damage to towns and cities and has disrupted major transportation systems (Bogart, 1960; L.R. Johnston Associates, 1983). Additionally, riverine infrastructure inadequately designed for flood discharges can result in a multitude of problems resulting in danger to human life, extensive loss and damage to public and private property, and damage to the environment (impaired aquatic biota and aquatic habitat).

To minimize the damage from floods, protect human health and safety, and conserve wildlife habitat, reliable estimates of the magnitude and frequency of floods are essential. Federal, State, regional, and local agencies rely on accurate estimates of the magnitude and frequency of flood

discharges to effectively plan and manage land use and water resources, protect lives and property, and administer flood insurance programs. Flood-frequency statistics compiled from records maintained at streamgages provide a basis for the design of infrastructure, such as roads and bridges; the management of flood risk; the protection of aquatic species; and other engineering and environmental analyses. Flood-frequency statistics can be computed directly for a particular streamgage with a suitable length of record. For short-record streamgages or ungaged sites, regression equations developed from flood-frequency statistics and basin characteristics for a regional group of streamgages can provide estimates of flood-frequency statistics.

The U.S. Geological Survey (USGS) has published reports that provide methods for estimating flood-frequency statistics at gaged and ungaged sites in States across the United States, including Connecticut. This report is the fourth in a series of reports presenting techniques for estimating flood discharges for ungaged stream sites in Connecticut. Previous flood regionalization studies for Connecticut were conducted by Bigwood and Thomas (1955), Weiss (1975, 1983), and Ahearn (2004).

These studies benefit from periodic updates to incorporate new streamflow information, improved measurements of basin characteristics, and improved computational techniques. Recently (2018), the national guidelines for flood-frequency analysis (Bulletin 17C; England and others, 2018) were updated with improved statistical techniques for flood-frequency analysis. This present flood-frequency study for Connecticut was undertaken by the USGS, in cooperation with the Connecticut Department of Transportation, to incorporate (1) 14 or more additional years of streamflow data at most sites, (2) updated basin characteristics using new digital geospatial datasets, and (3) improved statistical techniques for flood-frequency analysis including the expected moments algorithm (EMA) and the multiple Grubbs-Beck test (MGBT). This study also incorporates the most current regional skew estimate.

Purpose and Scope

The purpose of this report is to (1) update the flood discharges at gaged locations for the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities (AEPs) using data through water year 2015, (2) update regional regression equations for estimating flood discharges at ungaged stream sites in Connecticut, and (3) provide a method for estimating the 70-, 80-, 90-, 95- , and 99-percent prediction intervals estimated from the regression equations. Results are integrated into StreamStats, a web-based USGS map tool that provides a graphical means of interactively selecting a location and automatically calculating the associated streamflow statistics.

The information presented herein supersedes previous USGS data and reports on estimation of peak-flow magnitude and frequency on streams in Connecticut. The new regression equations are based on an additional 14 or more years of flow data and improved statistical techniques for flood frequency (EMA, MGBT, regional skew, corrected confidence intervals described in the national guidelines for flood-frequency analyses, denoted as Bulletin 17C [England and others, 2018]).

Description of Study Area

The area considered for flood-frequency analysis in this study, referred to in this report as the "study area," includes the State of Connecticut, an area of about 5,000 square miles (mi^2), and a 5-mile (mi) buffer extending into Massachusetts, New York, and Rhode Island (fig. 1). Streamgages outside of Connecticut but near the State border can provide additional information representative of the varied hydrologic response in the State. Three streamgages in Rhode Island are outside the 5-mi buffer but parts of their basins are within the 5-mi buffer and were considered representative regional peak flows for the regression. Connecticut extends approximately 60 mi from north to south and 90 mi from west to east. Connecticut has four physiographic regions: Northwestern Uplands, Eastern Uplands, Lower Connecticut River Valley, and Southern New England Coastal Lowlands. Both uplands regions are characterized by steep hills and are heavily forested. The Northwestern Uplands generally have the steepest topography; land-surface elevations range from about 500 to 2,300 feet (ft) above the North American Vertical Datum of 1988 (NAVD 88) with average slopes of about 11 percent. The Eastern Uplands have land-surface elevations ranging from about 500 to 1,300 ft above NAVD 88 with average slopes of about 8 percent. The Lower Connecticut River Valley traverses the center of the State and averages about 30 mi wide. Most of the land is gently to moderately sloping, except for the narrow trap rock ridges that run from Long Island Sound to Massachusetts. The trap rock ridges rise to more than 1,000 ft above sea level. The Southern New England Coastal Lowlands extend along the southern shore of the State, accounting for about 460 mi of actual coastline forming a narrow strip of land ranging from about 6 to 16 mi wide. Densely populated urban areas are prominent along the southwestern-coastal and central-valley regions.

Connecticut's climate is characterized by cold, snowy winters and warm, humid summers (Runkle and others, 2017). The weather pattern is highly variable with abundant precipitation that is normally evenly distributed throughout the year. The annual average temperature is 49 degrees Fahrenheit (°F) with average temperatures of 26 °F in January and 72 °F in July. The maximum annual precipitation during the century was 64 inches (in.), and the minimum was 31 in. with a standard deviation of 7.3 in. The average annual precipitation is 44.2 in. for Hartford, Connecticut, and 42.8 in. for Bridgeport, Conn. (U.S. Climate Data, 2019). Precipitation averages from 3 to 4 in. per month with occasional large storms in any month. The average snowfall is about 30 in. along the coast, 40 in. inland, and 50 in. in the northwestern corner of the

Figure 1. Locations of the U.S. Geological Survey streamgages in Connecticut and adjacent States used in the flood-frequency and regression analyses of peak flows.

State (Miller and others, 2002). Extreme precipitation events, nor'easters, winter storms, tornadoes, and hurricanes are part of Connecticut's climate. During hurricane season, tropical cyclones often affect the region. Thunderstorms are most frequent during the summer, occurring on average 30 times annually. These storms can be severe and contribute substantial rainfall amounts to the region (Paulson and others, 1940).

Hurricanes, remnants of hurricanes, and storms that never developed to hurricane strength are major causes of floods in Connecticut. The National Oceanic and Atmospheric Administration (NOAA) indicates that 44 hurricanes, tropical storms, tropical depressions, and extratropical storms have passed within a 75-mi radius of New Haven, Conn., since 1851 (NOAA, 2019; appendix 1). These storms typically originate in the central Atlantic Ocean and often follow a track along the eastern United States up through New England. The Great New England Hurricane of 1938 was the first catastrophic hurricane to impact New England since 1869 (early climate records) and holds the record for the worst natural disaster in the State's 350-year history (Runkle and others, 2017). Historical flooding in Connecticut has been well chronicled (Thomson and others, 1964). The most severe and memorable floods in Connecticut include September 1938, August 1955, October 1955, and June 1982.

Data Compilation

Peak-flow records were retrieved from the USGS National Water Information System database (U.S. Geological Survey, 2016). Selection considerations for flood-frequency analysis and regionalization of peak flows included record length and the effect of any streamflow regulation or diversion, urbanization, or natural damming of water affecting peak flow. Peak-flow data for the selected sites were reviewed to assure the quality of the records and tested for homogeneity or presence of trends over time, which could invalidate the assumptions of the analyses. Once the peak-flow records were compiled and reviewed, then physical and climatic characteristics of gaged basins were derived from various geospatial datasets for regionalization of peak flows.

Peak-Flow Records

Annual peak-flow data collected through September 2015 were obtained for 152 gaging stations that had at least 15 consecutive years of record (table 1) (U.S. Geological Survey, 2016a). Of these sites, 141 are in Connecticut, and 11 are in adjacent States. Although this study is focused on Connecticut, many more than 11 records from gaging stations in adjacent States were investigated for use in development of regression equations. Peak-flow records from adjacent States used in the final analysis included 3 stations in Massachusetts, 5 stations in Rhode Island, and 3 stations in New York.

Peak-flow records were inspected for anomalous values, and qualification codes associated with the peaks were reviewed for accuracy. Revisions were applied prior to an analysis of streamflow data for this study. Peak flows for which a discharge qualification code was published in the peak-flow file were included or omitted in the flood-frequency analysis according to procedures described for USGS flood-frequency analysis program PeakFQ version 7.1 (Flynn and others, 2006). Peak-flow records from regulated streams were reviewed individually to determine the effect of the unknown (code 5) or known (code 6) degree of regulation and whether the peak-flow data were suitable for regional regression analysis.

Trends in Peak Flows

The traditional assumption underlying flood-frequency analysis is stationarity in time. The assumption allows researchers to estimate the flood magnitude and frequency from past records and apply them to the future without adjustments. Milly and others (2008) called the assumption of climate-related stationarity into question and advocated for new methods to replace models based on stationarity. Several studies have documented increases in low and median flows across the United States (McCabe and Wolock, 2002; Lins and Slack, 2005; Small and others, 2006), but trends in peak flows are less evident in the literature. In New England, Walter and Vogel (2010) found increasing high flows in urbanizing basins, and Hodgkins and Dudley (2005), Collins (2009), and Huntington and others (2009) found increasing high flows in basins minimally affected by urbanization. If peak-flow increases continue in the future, flood magnitude and frequency estimates based on historical data may underestimate flood risk.

Increasing and decreasing trends in the magnitude and frequency of peak flows are difficult to estimate and apply to traditional flood-frequency analysis. Peak flows can be extremely variable; trends depend on the time period analyzed and can also vary substantially from site to site. Peak flows can be affected by decadal changes or may have decadal trends superimposed on long-term trends.

For the trend analysis, peak-flow data were analyzed for long-term trends in the magnitude of peaks for streamgages on unregulated streams in Connecticut and surrounding

Table 1. Descriptions of U.S. Geological Survey streamgages in Connecticut and adjacent States used in the flood-frequency analysis and regionalization of peaks flows in Connecticut.

[Table available for download at https://doi.org/10.3133/sir20205054]

States extending about 50 mi from the Connecticut border. Streamgages in adjacent States can provide additional information on regional streamflow trends. Subsets of streamgages with long records were created to evaluate trends during the past 30, 50, 70, and 90 years through 2015. All 10-year blocks within each time period analyzed were required to be at least 80 percent complete so that no part of the time series would have substantial missing data. These length and completeness criteria resulted in 79 streamgages for the 30-year period, 64 streamgages for the 50-year period, 44 streamgages for the 70-year period, and 9 streamgages for the 90-year period.

The magnitudes of the trends were computed with Sen slope (also known as the Kendall-Theil robust line), the median of all possible pairwise slopes in each time series (Helsel and Hirsch, 2002). The Sen slope is multiplied by the number of years of peaks to obtain the magnitude of the trend or total change in the annual peak flows over the period analyzed. For example, a Sen slope of 23.18 cubic feet per second (ft^3/s) multiplied by 70 for the 70-year period results in a trend magnitude of 1,623 ft^3/s for the Salmon River near East Hampton (station 01193500).

The trends were computed with methods that consider the possibility of short- and long-term persistence in the temporal data. This is an important issue that is often ignored in trend studies. Trends over time are sensitive to assumptions of whether underlying hydroclimatic data are independent, have short-term persistence, or have long-term persistence (Cohn and Lins, 2005; Koutsoyiannis and Montanari, 2007; Hamed, 2008; Khaliq and others, 2009; Kumar and others, 2009). Short- and long-term persistence may represent the occurrence of wet or dry conditions that tend to cluster from year to year (Koutsoyiannis and Montanari, 2007; Hodgkins and others, 2017). For further discussion and references on persistence, see Hodgkins and Dudley (2011).

Because the long-term time-series structure of peak-flow data is not well understood, temporal trend significance with three different null hypotheses of the serial structure of the data are reported: independence, short-term persistence (STP), and long-term persistence (LTP) (Hamed and Rao, 1998; Hamed, 2008). The serial structure of data referred to as "independence" means annual peaks from year to year are independent from each other (ignores any short or long clusters of wet and dry years). Trends were considered statistically significant at $p \leq 0.05$; this level represents a 5-percent probability that a trend is due to random chance. Results from the trend analysis for 30-, 50-, 70- and 90-year time periods under the three serial correlation structures, magnitudes of Sen slopes, and *p*-values are shown in tables 2 through 5. Peak-flow trend results depend on the period of record analyzed and assumptions about the serial correlation structure of the annual peak flows.

Increasing trends were detected in each of the four time periods (30, 50, 70, and 90 years) analyzed. For the 30-year period (1986–2015), 19.0 percent of streamgages (15 of 79) have increasing trends if independence of annual peak flows is assumed (table 2). If STP is assumed, 10.1 percent of streamgages (8 of 79) have increasing trends, and if LTP is assumed, no streamgages have trends. There are no decreasing trends for any assumption of the serial correlation structure.

For the 50-year period (1966–2015), there are considerably fewer trends detected than for the three other time periods analyzed (table 3); 6.2 percent of streamgages (4 of 64) have increasing trends if independence of annual peak flows is assumed. If STP is assumed, 1.6 percent (1 of 64) of streamgages have increasing trends, and none have trends if LTP is assumed. As with the 30-year trends, there were no significant decreasing trends.

There was a relatively high percentage of increasing trends detected for the 70-year period (1946–2015) with the assumption of independence; 47.7 percent of streamgages had increasing trends (21 of 44) (table 4). Assuming STP, this percentage decreased to 36.4 percent of streamgages (16 of 44) and decreased to 9.1 percent of streamgages (4 of 44) if LTP is assumed. For the limited number of streamgages with adequate data for 90 years (1926–2015), 33.3 percent of streamgages (3 of 9) had increasing trends if independence or STP is assumed. If LTP is assumed, 11.1 percent of streamgages (1 of 9) had increasing trends (table 5). No streamgages showed significant decreasing trends for the 70- or 90-year periods.

In the 70- and 90-year periods, increasing trends were detected under all three serial correlation structures, 4 streamgages in the 70-year period and 1 streamgage in the 90-year period. The Quinnipiac River (station 01196500) showed an increasing trend in all three serial correlation structures in the 70-year period. Based on a trend magnitude of 1,548 ft^3/s for the 70-year period, the annual peak flows on the Quinnipiac River (station 01196500; fig. 1) increased by about 22 ft^3/s per year (or 220 ft^3/s every decade).

In summary, there is no evidence for decreasing annual peak flows over time and some evidence of increasing peak flows over time. The difference in the percentage of statistically significant ($p \leq 0.05$) increasing trends based on the period analyzed indicates that multidecadal cycles may be present and influencing the magnitude of the trends. For example, many of the streamgages with increasing peak-flow trends for the 70-year period (1946–2015) had many low-magnitude peaks prior to 1965 or 1970. There are fewer trends in the 50-year period (1966–2015); this period largely comes after the period that had many low-magnitude peaks. Flood-poor and flood-rich periods may be related to multidecadal climate cycles. The number of major peaks flows (flows with annual exceedance probabilities less than 0.4 percent [25-year recurrence interval]) have been shown to be related to the phase of the Atlantic Multidecadal Oscillation for relatively large basins (>386 mi^2) in parts of North America that include the Northeast United States (Hodgkins and others, 2017).

Table 2. Trend results for the 30-year period (1986–2015) assuming serial correlation structure of the annual peak flows and Sen slopes for 79 streamgages in Connecticut and adjacent States.

[Drainage area from U.S. Geological Survey National Water Information System database. Sen slope indicates rate of change over period analyzed. Table cells are shaded for *p*-values less than or equal to 0.05, which indicate statistically significant trends. mi², square mile; ft³/s, cubic foot per second; CT, Connecticut; MA, Massachusetts; NJ, New Jersey; NY, New York; RI, Rhode Island]

Station identifier	Station name	Drainage area (mi²)	Sen slope magnitude ([ft³/s]/period)	Independence		Short-term persistence		Long-term persistence	
				p-value	Trend	*p*-value	Trend	*p*-value	Trend
01118300	PENDLETON HILL BROOK NEAR CLARKS FALLS, CT	4.02	75	0.10	Increase	0.11	Increase	0.29	Increase
01119500	WILLIMANTIC RIVER NEAR COVENTRY, CT	121	663	0.56	Increase	0.57	Increase	0.71	Increase
01121000	MOUNT HOPE RIVER NEAR WARRENVILLE, CT	28.6	305	0.28	Increase	0.23	Increase	0.48	Increase
01123000	LITTLE RIVER NEAR HANOVER, CT	30	498	0.21	Increase	0.22	Increase	0.42	Increase
01127500	YANTIC RIVER AT YANTIC, CT	89.3	1,375	0.15	Increase	0.20	Increase	0.35	Increase
01184000	CONNECTICUT RIVER AT THOMPSONVILLE, CT	9,660	−4,688	0.72	Decrease	0.66	Decrease	0.82	Decrease
01184100	STONY BROOK NEAR WEST SUFFIELD, CT	10.4	45	0.79	Increase	0.76	Increase	0.86	Increase
01184490	BROAD BROOK AT BROAD BROOK, CT	15.5	68	0.67	Increase	0.63	Increase	0.78	Increase
01187300	HUBBARD RIVER NEAR WEST HARTLAND, CT	19.9	29	0.89	Increase	0.87	Increase	0.93	Increase
01188000	BUNNELL BROOK NEAR BURLINGTON, CT	4.1	184	0.07	Increase	0.09	Increase	0.25	Increase
01192500	HOCKANUM RIVER NEAR EAST HARTFORD, CT	73.4	900	0.01	Increase	0.02	Increase	0.08	Increase
01192883	COGINCHAUG RIVER AT MIDDLEFIELD, CT	29.8	434	0.07	Increase	0.08	Increase	0.24	Increase
01193500	SALMON RIVER NEAR EAST HAMPTON, CT	100	3,600	0.00	Increase	0.07	Increase	0.19	Increase
01195100	INDIAN RIVER NEAR CLINTON, CT	5.68	158	0.01	Increase	0.02	Increase	0.10	Increase
01196500	QUINNIPIAC RIVER AT WALLINGFORD, CT	115	25	0.94	Increase	0.94	Increase	0.96	Increase
01196620	MILL RIVER NEAR HAMDEN, CT	24.5	600	0.07	Increase	0.04	Increase	0.24	Increase
01199000	HOUSATONIC RIVER AT FALLS VILLAGE, CT	634	1,917	0.10	Increase	0.14	Increase	0.30	Increase
01199050	SALMON CREEK AT LIME ROCK, CT	29.4	515	0.02	Increase	0.05	Increase	0.13	Increase
01200500	HOUSATONIC RIVER AT GAYLORDSVILLE, CT	996	4,650	0.10	Increase	0.15	Increase	0.31	Increase
01204000	POMPERAUG RIVER AT SOUTHBURY, CT	75.1	1,140	0.27	Increase	0.24	Increase	0.47	Increase
01208925	MILL RIVER NEAR FAIRFIELD, CT	28.6	329	0.13	Increase	0.23	Increase	0.33	Increase
01208950	SASCO BROOK NEAR SOUTHPORT, CT	7.38	185	0.09	Increase	0.17	Increase	0.27	Increase
01208990	SAUGATUCK RIVER NEAR REDDING, CT	21	234	0.24	Increase	0.27	Increase	0.51	Increase
01209700	NORWALK RIVER AT SOUTH WILTON, CT	30	334	0.43	Increase	0.49	Increase	0.61	Increase
01097000	ASSABET RIVER AT MAYNARD, MA	116	255	0.35	Increase	0.36	Increase	0.55	Increase
01097300	NASHOBA BROOK NEAR ACTON, MA	12.8	68	0.28	Increase	0.33	Increase	0.48	Increase
01099500	CONCORD RIVER BELOW RIVER MEADOW BROOK, AT LOWELL, MA	400	800	0.23	Increase	0.25	Increase	0.44	Increase
01101000	PARKER RIVER AT BYFIELD, MA	21.3	130	0.09	Increase	0.06	Increase	0.27	Increase

Table 2. Trend results for the 30-year period (1986–2015) assuming serial correlation structure of the annual peak flows and Sen slopes for 79 streamgages in Connecticut and adjacent States.—Continued

[Drainage area from U.S. Geological Survey National Water Information System database. Sen slope indicates rate of change over period analyzed. Table cells are shaded for *p*-values less than or equal to 0.05, which indicate statistically significant trends. mi², square mile; ft³/s, cubic foot per second; CT, Connecticut; MA, Massachusetts; NJ, New Jersey; NY, New York; RI, Rhode Island]

Station identifier	Station name	Drainage area (mi²)	Sen slope magnitude ([ft³/s]/period)	Independence		Short-term persistence		Long-term persistence	
				p-value	Trend	*p*-value	Trend	*p*-value	Trend
01101500	IPSWICH RIVER AT SOUTH MIDDLETON, MA	44.5	280	0.10	Increase	0.09	Increase	0.29	Increase
01102000	IPSWICH RIVER NEAR IPSWICH, MA	125	740	0.26	Increase	0.24	Increase	0.46	Increase
01102500	ABERJONA RIVER AT WINCHESTER, MA	24.5	33	0.75	Increase	0.71	Increase	0.84	Increase
01103500	CHARLES RIVER AT DOVER, MA	183	409	0.13	Increase	0.09	Increase	0.32	Increase
01104500	CHARLES RIVER AT WALTHAM, MA	251	570	0.09	Increase	0.09	Increase	0.27	Increase
01105000	NEPONSET RIVER AT NORWOOD, MA	34.7	177	0.07	Increase	0.10	Increase	0.31	Increase
01105500	EAST BRANCH NEPONSET RIVER AT CANTON, MA	27.2	222	0.04	Increase	0.07	Increase	0.21	Increase
01105600	OLD SWAMP RIVER NEAR SOUTH WEYMOUTH, MA	4.5	42	0.35	Increase	0.47	Increase	0.71	Increase
01109000	WADING RIVER NEAR NORTON, MA	43.3	221	0.13	Increase	0.15	Increase	0.33	Increase
01109060	THREEMILE RIVER AT NORTH DIGHTON, MA	84.3	514	0.09	Increase	0.11	Increase	0.27	Increase
01109070	SEGREGANSET RIVER NEAR DIGHTON, MA	10.6	65	0.40	Increase	0.41	Increase	0.58	Increase
01110000	QUINSIGAMOND RIVER AT NORTH GRAFTON, MA	25.6	101	0.24	Increase	0.27	Increase	0.44	Increase
01171500	MILL RIVER AT NORTHAMPTON, MA	52.6	1,400	0.03	Increase	0.06	Increase	0.16	Increase
01174500	EAST BRANCH SWIFT RIVER NEAR HARDWICK, MA	43.7	228	0.30	Increase	0.36	Increase	0.53	Increase
01175670	SEVENMILE RIVER NEAR SPENCER, MA	8.81	161	0.02	Increase	0.08	Increase	0.13	Increase
01176000	QUABOAG RIVER AT WEST BRIMFIELD, MA	150	764	0.02	Increase	0.05	Increase	0.14	Increase
01181000	WEST BRANCH WESTFIELD RIVER AT HUNTINGTON, MA	94	317	0.94	Increase	0.95	Increase	0.97	Increase
01185500	WEST BRANCH FARMINGTON RIVER NEAR NEW BOSTON, MA	91.7	375	0.54	Increase	0.51	Increase	0.69	Increase
01197000	EAST BRANCH HOUSATONIC RIVER AT COLTSVILLE, MA	57.6	180	0.73	Increase	0.70	Increase	0.83	Increase
01197500	HOUSATONIC RIVER NEAR GREAT BARRINGTON, MA	282	1,538	0.15	Increase	0.18	Increase	0.35	Increase
01377000	HACKENSACK RIVER AT RIVERVALE, NJ	58	1,088	0.05	Increase	0.06	Increase	0.21	Increase
01384500	RINGWOOD CREEK NEAR WANAQUE, NJ	17.9	321	0.04	Increase	0.07	Increase	0.20	Increase
01387400	RAMAPO RIVER AT RAMAPO NY	86.9	420	0.59	Increase	0.63	Increase	0.76	Increase
01387420	RAMAPO RIVER AT SUFFERN NY	93	321	0.79	Increase	0.80	Increase	0.88	Increase
01387450	MAHWAH RIVER NEAR SUFFERN, NJ	12.3	167	0.48	Increase	0.55	Increase	0.67	Increase
01387500	RAMAPO RIVER NEAR MAHWAH, NJ	120	1,243	0.28	Increase	0.35	Increase	0.48	Increase
01199477	STONY BK NR DOVER PLAINS, NY	1.93	60	0.18	Increase	0.15	Increase	0.38	Increase
01361000	KINDERHOOK CREEK AT ROSSMAN, NY	329	1,905	0.27	Increase	0.28	Increase	0.48	Increase

Table 2. Trend results for the 30-year period (1986–2015) assuming serial correlation structure of the annual peak flows and Sen slopes for 79 streamgages in Connecticut and adjacent States.—Continued

[Drainage area from U.S. Geological Survey National Water Information System database. Sen slope indicates rate of change over period analyzed. Table cells are shaded for *p*-values less than or equal to 0.05, which indicate statistically significant trends. mi², square mile; ft³/s, cubic foot per second; CT, Connecticut; MA, Massachusetts; NJ, New Jersey; NY, New York; RI, Rhode Island]

Station identifier	Station name	Drainage area (mi²)	Sen slope magnitude ([ft³/s]/period)	Independence		Short-term persistence		Long-term persistence	
				p-value	Trend	*p*-value	Trend	*p*-value	Trend
01361500	CATSKILL CREEK AT OAK HILL, NY	95.8	11	1.00	Increase	1.00	Increase	1.00	Increase
01362100	ROELIFF JANSEN KILL NR HILLSDALE, NY	27.5	339	0.33	Increase	0.39	Increase	0.55	Increase
01362200	ESOPUS CREEK AT ALLABEN, NY	63.7	1,020	0.56	Increase	0.57	Increase	0.73	Increase
01362500	ESOPUS CREEK AT COLDBROOK, NY	192	4,174	0.41	Increase	0.41	Increase	0.59	Increase
01365000	RONDOUT CREEK NEAR LOWES CORNERS, NY	38.3	111	0.90	Increase	0.90	Increase	0.94	Increase
01367500	RONDOUT CREEK AT ROSENDALE, NY	383	1,385	0.56	Increase	0.55	Increase	0.71	Increase
01371500	WALLKILL RIVER AT GARDINER, NY	695	2,465	0.24	Increase	0.33	Increase	0.55	Increase
01372500	WAPPINGER CREEK NEAR WAPPINGERS FALLS, NY	181	1,181	0.22	Increase	0.29	Increase	0.47	Increase
01374250	PEEKSKILL HOLLOW CR AT TOMPKINS CORNER, NY	14.9	258	0.12	Increase	0.12	Increase	0.40	Increase
01111300	NIPMUC RIVER NEAR HARRISVILLE, RI	16	387	0.11	Increase	0.14	Increase	0.31	Increase
01111500	BRANCH RIVER AT FORESTDALE, RI	91.2	469	0.44	Increase	0.46	Increase	0.62	Increase
01112500	BLACKSTONE RIVER AT WOONSOCKET, RI	416	1,821	0.26	Increase	0.35	Increase	0.52	Increase
01114000	MOSHASSUCK RIVER AT PROVIDENCE, RI	23.1	375	0.03	Increase	0.17	Increase	0.44	Increase
01114500	WOONASQUATUCKET RIVER AT CENTERDALE, RI	38.3	173	0.43	Increase	0.50	Increase	0.67	Increase
01116000	SOUTH BRANCH PAWTUXET RIVER AT WASHINGTON, RI	62.8	386	0.09	Increase	0.09	Increase	0.33	Increase
01117000	HUNT RIVER NEAR EAST GREENWICH, RI	22.9	196	0.06	Increase	0.06	Increase	0.21	Increase
01117350	CHIPUXET RIVER AT WEST KINGSTON, RI	9.59	66	0.05	Increase	0.05	Increase	0.19	Increase
01117420	USQUEPAUG RIVER NEAR USQUEPAUG, RI	36.1	217	0.04	Increase	0.04	Increase	0.20	Increase
01117468	BEAVER RIVER NEAR USQUEPAUG, RI	8.87	72	0.02	Increase	0.02	Increase	0.12	Increase
01117500	PAWCATUCK RIVER AT WOOD RIVER JUNCTION, RI	100	342	0.04	Increase	0.04	Increase	0.19	Increase
01117800	WOOD RIVER NEAR ARCADIA, RI	35.2	175	0.08	Increase	0.09	Increase	0.25	Increase
01118000	WOOD RIVER AT HOPE VALLEY, RI	72.4	500	0.06	Increase	0.05	Increase	0.22	Increase
01118500	PAWCATUCK RIVER AT WESTERLY, RI	295	1,300	0.01	Increase	0.02	Increase	0.10	Increase

Table 3. Trend results for the 50-year period (1966–2015) assuming serial correlation structure of the annual peak flows and Sen slopes for 64 streamgages in Connecticut and adjacent States.

[Drainage area from U.S. Geological Survey National Water Information System database. Sen slope indicates rate of change over period analyzed. Table cells are shaded for *p*-values less than or equal to 0.05, which indicate statistically significant trends.; mi^2, square mile; ft^3/s, cubic foot per second; CT, Connecticut; MA, Massachusetts; NJ, New Jersey; NY, New York; RI, Rhode Island]

Station identifier	Station name	Drainage area (mi^2)	Sen slope magnitude ([ft^3/s]/period)	Independence		Short-term persistence		Long-term persistence	
				p-value	Trend	*p*-value	Trend	*p*-value	Trend
01118300	PENDLETON HILL BROOK NEAR CLARKS FALLS, CT	4.02	40	0.25	Increase	0.26	Increase	0.40	Increase
01119500	WILLIMANTIC RIVER NEAR COVENTRY, CT	121	0	1.00	No change	1.00	No change	1.00	No change
01121000	MOUNT HOPE RIVER NEAR WARRENVILLE, CT	28.6	258	0.32	Increase	0.30	Increase	0.46	Increase
01123000	LITTLE RIVER NEAR HANOVER, CT	30	121	0.74	Increase	0.74	Increase	0.81	Increase
01127500	YANTIC RIVER AT YANTIC, CT	89.3	452	0.63	Increase	0.68	Increase	0.77	Increase
01184000	CONNECTICUT RIVER AT THOMPSONVILLE, CT	9,660	357	0.95	Increase	0.93	Increase	0.96	Increase
01184100	STONY BROOK NEAR WEST SUFFIELD, CT	10.4	250	0.07	Increase	0.06	Increase	0.19	Increase
01187300	HUBBARD RIVER NEAR WEST HARTLAND, CT	19.9	91	0.73	Increase	0.71	Increase	0.80	Increase
01188000	BUNNELL BROOK NEAR BURLINGTON, CT	4.1	161	0.07	Increase	0.11	Increase	0.26	Increase
01192500	HOCKANUM RIVER NEAR EAST HARTFORD, CT	73.4	400	0.23	Increase	0.34	Increase	0.56	Increase
01193500	SALMON RIVER NEAR EAST HAMPTON, CT	100	1,014	0.24	Increase	0.32	Increase	0.63	Increase
01196500	QUINNIPIAC RIVER AT WALLINGFORD, CT	115	465	0.47	Increase	0.48	Increase	0.61	Increase
01199000	HOUSATONIC RIVER AT FALLS VILLAGE, CT	634	1,022	0.38	Increase	0.37	Increase	0.54	Increase
01199050	SALMON CREEK AT LIME ROCK, CT	29.4	207	0.22	Increase	0.21	Increase	0.42	Increase
01200500	HOUSATONIC RIVER AT GAYLORDSVILLE, CT	996	1,694	0.48	Increase	0.49	Increase	0.65	Increase
01204000	POMPERAUG RIVER AT SOUTHBURY, CT	75.1	690	0.51	Increase	0.51	Increase	0.69	Increase
01208950	SASCO BROOK NEAR SOUTHPORT, CT	7.38	68	0.49	Increase	0.59	Increase	0.67	Increase
01208990	SAUGATUCK RIVER NEAR REDDING, CT	21	−219	0.19	Decrease	0.24	Decrease	0.56	Decrease
01209700	NORWALK RIVER AT SOUTH WILTON, CT	30	−106	0.80	Decrease	0.83	Decrease	0.89	Decrease
01097000	ASSABET RIVER AT MAYNARD, MA	116	125	0.62	Increase	0.61	Increase	0.72	Increase
01097300	NASHOBA BROOK NEAR ACTON, MA	12.8	54	0.33	Increase	0.43	Increase	0.50	Increase
01099500	CONCORD RIVER BELOW RIVER MEADOW BROOK, AT LOWELL, MA	400	586	0.29	Increase	0.30	Increase	0.44	Increase
01101000	PARKER RIVER AT BYFIELD, MA	21.3	98	0.13	Increase	0.11	Increase	0.27	Increase
01101500	IPSWICH RIVER AT SOUTH MIDDLETON, MA	44.5	162	0.10	Increase	0.10	Increase	0.23	Increase
01102000	IPSWICH RIVER NEAR IPSWICH, MA	125	485	0.19	Increase	0.20	Increase	0.33	Increase
01102500	ABERJONA RIVER AT WINCHESTER, MA	24.5	160	0.22	Increase	0.17	Increase	0.37	Increase
01103500	CHARLES RIVER AT DOVER, MA	183	141	0.59	Increase	0.57	Increase	0.69	Increase
01104500	CHARLES RIVER AT WALTHAM, MA	251	−24	0.90	Decrease	0.89	Decrease	0.93	Decrease

Table 3. Trend results for the 50-year period (1966–2015) assuming serial correlation structure of the annual peak flows and Sen slopes for 64 streamgages in Connecticut and adjacent States.—Continued

[Drainage area from U.S. Geological Survey National Water Information System database. Sen slope indicates rate of change over period analyzed. Table cells are shaded for *p*-values less than or equal to 0.05, which indicate statistically significant trends.; mi², square mile; ft³/s, cubic foot per second; CT, Connecticut; MA, Massachusetts; NJ, New Jersey; NY, New York; RI, Rhode Island]

Station identifier	Station name	Drainage area (mi²)	Sen slope magnitude ([ft³/s]/period)	Independence		Short-term persistence		Long-term persistence	
				p-value	Trend	*p*-value	Trend	*p*-value	Trend
01105000	NEPONSET RIVER AT NORWOOD, MA	34.7	82	0.31	Increase	0.33	Increase	0.46	Increase
01105500	EAST BRANCH NEPONSET RIVER AT CANTON, MA	27.2	60	0.44	Increase	0.50	Increase	0.64	Increase
01105600	OLD SWAMP RIVER NEAR SOUTH WEYMOUTH, MA	4.5	−16	0.76	Decrease	0.79	Decrease	0.86	Decrease
01109000	WADING RIVER NEAR NORTON, MA	43.3	53	0.79	Increase	0.81	Increase	0.85	Increase
01109060	THREEMILE RIVER AT NORTH DIGHTON, MA	84.3	−39	0.90	Decrease	0.91	Decrease	0.93	Decrease
01109070	SEGREGANSET RIVER NEAR DIGHTON, MA	10.6	−50	0.61	Decrease	0.64	Decrease	0.71	Decrease
01110000	QUINSIGAMOND RIVER AT NORTH GRAFTON, MA	25.6	65	0.22	Increase	0.22	Increase	0.37	Increase
01171500	MILL RIVER AT NORTHAMPTON, MA	52.6	1,107	0.04	Increase	0.09	Increase	0.29	Increase
01174500	EAST BRANCH SWIFT RIVER NEAR HARDWICK, MA	43.7	157	0.36	Increase	0.35	Increase	0.52	Increase
01175670	SEVENMILE RIVER NEAR SPENCER, MA	8.81	100	0.04	Increase	0.11	Increase	0.18	Increase
01176000	QUABOAG RIVER AT WEST BRIMFIELD, MA	150	250	0.31	Increase	0.39	Increase	0.57	Increase
01181000	WEST BRANCH WESTFIELD RIVER AT HUNTINGTON, MA	94	−1,489	0.44	Decrease	0.49	Decrease	0.74	Decrease
01185500	WEST BRANCH FARMINGTON RIVER NEAR NEW BOSTON, MA	91.7	250	0.62	Increase	0.60	Increase	0.76	Increase
01197000	EAST BRANCH HOUSATONIC RIVER AT COLTSVILLE, MA	57.6	77	0.89	Increase	0.88	Increase	0.92	Increase
01197500	HOUSATONIC RIVER NEAR GREAT BARRINGTON, MA	282	1,000	0.15	Increase	0.14	Increase	0.29	Increase
01377000	HACKENSACK RIVER AT RIVERVALE, NJ	58	927	0.02	Increase	0.02	Increase	0.08	Increase
01387450	MAHWAH RIVER NEAR SUFFERN, NJ	12.3	0	1.00	No change	1.00	No change	1.00	No change
01387500	RAMAPO RIVER NEAR MAHWAH, NJ	120	186	0.82	Increase	0.83	Increase	0.89	Increase
01362100	ROELIFF JANSEN KILL NR HILLSDALE, NY	27.5	74	0.70	Increase	0.73	Increase	0.79	Increase
01362200	ESOPUS CREEK AT ALLABEN, NY	63.7	520	0.72	Increase	0.73	Increase	0.82	Increase
01362500	ESOPUS CREEK AT COLDBROOK, NY	192	5,400	0.26	Increase	0.27	Increase	0.41	Increase
01365000	RONDOUT CREEK NEAR LOWES CORNERS, NY	38.3	91	0.89	Increase	0.89	Increase	0.92	Increase
01367500	RONDOUT CREEK AT ROSENDALE, NY	383	2,500	0.22	Increase	0.21	Increase	0.36	Increase
01371500	WALLKILL RIVER AT GARDINER, NY	695	1,875	0.25	Increase	0.34	Increase	0.52	Increase
01372500	WAPPINGER CREEK NEAR WAPPINGERS FALLS, NY	181	222	0.80	Increase	0.82	Increase	0.89	Increase
01111300	NIPMUC RIVER NEAR HARRISVILLE, RI	16	253	0.14	Increase	0.18	Increase	0.29	Increase
01111500	BRANCH RIVER AT FORESTDALE, RI	91.2	0	1.00	No change	1.00	No change	1.00	No change
01112500	BLACKSTONE RIVER AT WOONSOCKET, RI	416	268	0.85	Increase	0.86	Increase	0.89	Increase

Table 3. Trend results for the 50-year period (1966–2015) assuming serial correlation structure of the annual peak flows and Sen slopes for 64 streamgages in Connecticut and adjacent States.—Continued

[Drainage area from U.S. Geological Survey National Water Information System database. Sen slope indicates rate of change over period analyzed. Table cells are shaded for *p*-values less than or equal to 0.05, which indicate statistically significant trends.; mi^2, square mile; ft^3/s, cubic foot per second; CT, Connecticut; MA, Massachusetts; NJ, New Jersey; NY, New York; RI, Rhode Island]

Station identifier	Station name	Drainage area (mi^2)	Sen slope magnitude ([ft^3/s]/period)	Independence		Short-term persistence		Long-term persistence	
				p-value	Trend	*p*-value	Trend	*p*-value	Trend
01114000	MOSHASSUCK RIVER AT PROVIDENCE, RI	23.1	382	0.00	Increase	0.05	Increase	0.22	Increase
01114500	WOONASQUATUCKET RIVER AT CENTERDALE, RI	38.3	−63	0.83	Decrease	0.84	Decrease	0.87	Decrease
01116000	SOUTH BRANCH PAWTUXET RIVER AT WASHINGTON, RI	62.8	−29	0.85	Decrease	0.86	Decrease	0.92	Decrease
01117000	HUNT RIVER NEAR EAST GREENWICH, RI	22.9	97	0.29	Increase	0.31	Increase	0.44	Increase
01117500	PAWCATUCK RIVER AT WOOD RIVER JUNCTION, RI	100	139	0.46	Increase	0.45	Increase	0.58	Increase
01117800	WOOD RIVER NEAR ARCADIA, RI	35.2	24	0.84	Increase	0.85	Increase	0.89	Increase
01118000	WOOD RIVER AT HOPE VALLEY, RI	72.4	188	0.39	Increase	0.38	Increase	0.53	Increase
01118500	PAWCATUCK RIVER AT WESTERLY, RI	295	458	0.34	Increase	0.35	Increase	0.51	Increase

Table 4. Trend results for the 70-year period (1946–2015) assuming serial correlation structure of the annual peak flows and Sen slopes for 44 streamgages in Connecticut and adjacent States.

[Drainage area from U.S. Geological Survey National Water Information System database. Sen slope indicates rate of change over period analyzed. Table cells are shaded for *p*-values less than or equal to 0.05, which indicate statistically significant trends. mi², square mile; ft³/s, cubic foot per second; CT, Connecticut; MA, Massachusetts; NJ, New Jersey; NY, New York; RI, Rhode Island]

Station identifier	Station name	Drainage area (mi²)	Sen slope magnitude ([ft³/s]/ period)	Independence		Short-term persistence		Long-term persistence	
				p-value	Trend	*p*-value	Trend	*p*-value	Trend
Streamflow-gaging stations in Connecticut									
01119500	WILLIMANTIC RIVER NEAR COVENTRY, CT	121	1,176	0.04	Increase	0.05	Increase	0.21	Increase
01121000	MOUNT HOPE RIVER NEAR WARRENVILLE, CT	28.6	578	0.01	Increase	0.03	Increase	0.11	Increase
01127500	YANTIC RIVER AT YANTIC, CT	89.3	1,700	0.01	Increase	0.03	Increase	0.09	Increase
01184000	CONNECTICUT RIVER AT THOMPSONVILLE, CT	9,660	−8,167	0.35	Decrease	0.29	Decrease	0.47	Decrease
01188000	BUNNELL BROOK NEAR BURLINGTON, CT	4.1	175	0.01	Increase	0.03	Increase	0.19	Increase
01192500	HOCKANUM RIVER NEAR EAST HARTFORD, CT	73.4	754	0.00	Increase	0.02	Increase	0.13	Increase
01193500	SALMON RIVER NEAR EAST HAMPTON, CT	100	1,623	0.02	Increase	0.06	Increase	0.33	Increase
01196500	QUINNIPIAC RIVER AT WALLINGFORD, CT	115	1,548	0.00	Increase	0.01	Increase	0.03	Increase
01199000	HOUSATONIC RIVER AT FALLS VILLAGE, CT	634	521	0.58	Increase	0.57	Increase	0.69	Increase
01200500	HOUSATONIC RIVER AT GAYLORDSVILLE, CT	996	500	0.78	Increase	0.79	Increase	0.85	Increase
01204000	POMPERAUG RIVER AT SOUTHBURY, CT	75.1	1,171	0.09	Increase	0.12	Increase	0.33	Increase
Streamflow-gaging stations in adjacent States									
01097000	ASSABET RIVER AT MAYNARD, MA	116	361	0.13	Increase	0.14	Increase	0.23	Increase
01099500	CONCORD RIVER BELOW RIVER MEADOW BROOK, AT LOWELL, MA	400	652	0.10	Increase	0.14	Increase	0.22	Increase
01101000	PARKER RIVER AT BYFIELD, MA	21.3	70	0.11	Increase	0.10	Increase	0.22	Increase
01101500	IPSWICH RIVER AT SOUTH MIDDLETON, MA	44.5	174	0.04	Increase	0.04	Increase	0.11	Increase
01102000	IPSWICH RIVER NEAR IPSWICH, MA	125	543	0.05	Increase	0.06	Increase	0.13	Increase
01102500	ABERJONA RIVER AT WINCHESTER, MA	24.5	275	0.00	Increase	0.00	Increase	0.02	Increase
01103500	CHARLES RIVER AT DOVER, MA	183	350	0.08	Increase	0.08	Increase	0.18	Increase
01104500	CHARLES RIVER AT WALTHAM, MA	251	219	0.41	Increase	0.42	Increase	0.54	Increase
01105000	NEPONSET RIVER AT NORWOOD, MA	34.7	188	0.00	Increase	0.01	Increase	0.06	Increase
01109000	WADING RIVER NEAR NORTON, MA	43.3	149	0.14	Increase	0.21	Increase	0.31	Increase
01110000	QUINSIGAMOND RIVER AT NORTH GRAFTON, MA	25.6	60	0.16	Increase	0.21	Increase	0.32	Increase
01171500	MILL RIVER AT NORTHAMPTON, MA	52.6	1,128	0.03	Increase	0.09	Increase	0.30	Increase
01174500	EAST BRANCH SWIFT RIVER NEAR HARDWICK, MA	43.7	286	0.04	Increase	0.05	Increase	0.16	Increase
01176000	QUABOAG RIVER AT WEST BRIMFIELD, MA	150	486	0.02	Increase	0.02	Increase	0.17	Increase
01181000	WEST BRANCH WESTFIELD RIVER AT HUNTINGTON, MA	94	945	0.42	Increase	0.49	Increase	0.76	Increase

Table 4. Trend results for the 70-year period (1946–2015) assuming serial correlation structure of the annual peak flows and Sen slopes for 44 streamgages in Connecticut and adjacent States.—Continued

[Drainage area from U.S. Geological Survey National Water Information System database. Sen slope indicates rate of change over period analyzed. Table cells are shaded for *p*-values less than or equal to 0.05, which indicate statistically significant trends. mi², square mile; ft³/s, cubic foot per second; CT, Connecticut; MA, Massachusetts; NJ, New Jersey; NY, New York; RI, Rhode Island]

Station identifier	Station name	Drainage area (mi²)	Sen slope magnitude ([ft³/s]/period)	Independence		Short-term persistence		Long-term persistence	
				p-value	Trend	*p*-value	Trend	*p*-value	Trend
Streamflow-gaging stations in adjacent States—Continued									
01185500	WEST BRANCH FARMINGTON RIVER NEAR NEW BOSTON, MA	91.7	111	0.84	Increase	0.85	Increase	0.92	Increase
01197000	EAST BRANCH HOUSATONIC RIVER AT COLTSVILLE, MA	57.6	285	0.37	Increase	0.35	Increase	0.57	Increase
01197500	HOUSATONIC RIVER NEAR GREAT BARRINGTON, MA	282	700	0.23	Increase	0.24	Increase	0.38	Increase
01377000	HACKENSACK RIVER AT RIVERVALE, NJ	58	1,019	0.00	Increase	0.00	Increase	0.00	Increase
01387500	RAMAPO RIVER NEAR MAHWAH, NJ	120	1,197	0.10	Increase	0.14	Increase	0.30	Increase
01362500	ESOPUS CREEK AT COLDBROOK, NY	192	5,335	0.15	Increase	0.14	Increase	0.28	Increase
01365000	RONDOUT CREEK NEAR LOWES CORNERS, NY	38.3	583	0.39	Increase	0.43	Increase	0.55	Increase
01367500	RONDOUT CREEK AT ROSENDALE, NY	383	3,328	0.15	Increase	0.17	Increase	0.32	Increase
01371500	WALLKILL RIVER AT GARDINER, NY	695	3,030	0.07	Increase	0.14	Increase	0.33	Increase
01372500	WAPPINGER CREEK NEAR WAPPINGERS FALLS, NY	181	310	0.70	Increase	0.72	Increase	0.79	Increase
01111500	BRANCH RIVER AT FOREST DALE, RI	91.2	875	0.02	Increase	0.05	Increase	0.07	Increase
01112500	BLACKSTONE RIVER AT WOONSOCKET, RI	416	2482	0.02	Increase	0.06	Increase	0.12	Increase
01114500	WOONASQUATUCKET RIVER AT CENTERDALE, RI	38.3	267	0.04	Increase	0.09	Increase	0.17	Increase
01116000	SOUTH BRANCH PAWTUXET RIVER AT WASHINGTON, RI	62.8	108	0.43	Increase	0.46	Increase	0.65	Increase
01117000	HUNT RIVER NEAR EAST GREENWICH, RI	22.9	187	0.00	Increase	0.00	Increase	0.01	Increase
01117500	PAWCATUCK RIVER AT WOOD RIVER JUNCTION, RI	100	242	0.03	Increase	0.04	Increase	0.09	Increase
01118000	WOOD RIVER AT HOPE VALLEY, RI	72.4	367	0.02	Increase	0.02	Increase	0.07	Increase
01118500	PAWCATUCK RIVER AT WESTERLY, RI	295	819	0.02	Increase	0.03	Increase	0.09	Increase

Table 5. Trend results for the 90-year period (1926–2015) assuming serial correlation structure of the annual peak flows and Sen slopes for nine streamgages in Connecticut and adjacent States.

[Drainage area from U.S. Geological Survey National Water Information System database. Sen slope indicates rate of change over period analyzed. Table cells are shaded for *p*-values less than or equal to 0.05, which indicate statistically significant trends. mi², square mile; ft³/s, cubic foot per second CT, Connecticut; MA, Massachusetts; NJ, New Jersey; NY, New York]

Station identifier	Station name	Drainage area (mi²)	Sen slope magnitude ([ft³/s]/period)	Independence		Short-term persistence		Long-term persistence	
				p-value	Trend	*p*-value	Trend	*p*-value	Trend
Streamflow-gaging stations in Connecticut									
01199000	HOUSATONIC RIVER AT FALLS VILLAGE, CT	634	1,350	0.11	Increase	0.11	Increase	0.23	Increase
01200500	HOUSATONIC RIVER AT GAYLORDSVILLE, CT	996	2,034	0.28	Increase	0.29	Increase	0.45	Increase
Streamflow-gaging stations in adjacent States									
01109000	WADING RIVER NEAR NORTON, MA	43.3	125	0.10	Increase	0.16	Increase	0.28	Increase
01176000	QUABOAG RIVER AT WEST BRIMFIELD, MA	150	527	0.00	Increase	0.00	Increase	0.04	Increase
01185500	WEST BRANCH FARMINGTON RIVER NEAR NEW BOSTON, MA	91.7	277	0.48	Increase	0.51	Increase	0.71	Increase
01197500	HOUSATONIC RIVER NEAR GREAT BARRINGTON, MA	282	466	0.34	Increase	0.34	Increase	0.47	Increase
01387500	RAMAPO RIVER NEAR MAHWAH, NJ	120	1,673	0.01	Increase	0.02	Increase	0.08	Increase
01367500	RONDOUT CREEK AT ROSENDALE, NY	383	2,236	0.23	Increase	0.25	Increase	0.39	Increase
01371500	WALLKILL RIVER AT GARDINER, NY	695	4,227	0.00	Increase	0.02	Increase	0.06	Increase

Bulletin 17C (England and others, 2018) acknowledges concern about changes in flood risk associated with climate variability but does not include methods to account for such changes in flood-frequency analysis. Vogel and others (2011) recommend use of magnification and recurrence reduction factors to examine how a linear trend would affect flood magnitudes and recurrence intervals at a future time. Zarriello (2017) applied the magnification factor to two streamgages with long-term records that showed increasing trends (independence was assumed) in peak flows—Ipswich River at Ipswich, Massachusetts (station 01102000), and Mill River at Northampton, Mass. (station 01171500). The application of the magnification factor method indicated a flood with a given annual exceedance probability will, on average, be 2, 4, and 7 percent greater in magnitude in 10, 20, and 30 years, respectively. The trends observed in the data used in this study and the effects on flood frequency will require further work as the science evolves and new data are obtained.

Historical peak-flow trends in and near Connecticut do not offer clear and convincing evidence of the need to incorporate trends into flood-frequency analyses. If the evidence was clear, a well-defined deterministic mechanism should be identified prior to incorporating trends (Salas and others, 2018). For this study, the traditional assumption of stationarity is used with no adjustment for historical trends.

Physical and Climatic Basin Characteristics

Peak-flow information can be estimated at ungaged sites through a multiple regression analysis that develops a relation between peak-flow characteristics (such as the 1-percent annual exceedance probability flow) and selected physical and climatic basin characteristics for gaged drainage basins. Thirty basin characteristics were selected as potential explanatory variables in the regression analyses on the basis of their theoretical relations to peak flows, results of previous peak-flow studies in similar hydrologic regions, and the ability to measure the basin characteristics using digital datasets and geographic information system (GIS) technology (table 6). The ability to measure the basin characteristics using GIS technology was important to facilitate automation of the process for measuring the basin characteristics and solving the regression equations in StreamStats, the USGS Streamflow Statistics and Spatial Analysis Tools for Water-Resources Applications (https://streamstats.usgs.gov). Basin characteristics were derived from various national geospatial datasets, including the National Land Cover Dataset (USGS, 2014a), the National Elevation Dataset (USGS, 2017a), the National Hydrologic Dataset Plus (USGS, 2017b), the Soil Survey Geographic (SSURGO) database (Soil Survey Staff, Natural Resources Conservation Service, U.S. Department of Agriculture, 2017), the National Quaternary Sediments in the Glaciated United States Dataset (USGS, 1970), the National Wetlands Dataset, the Parameter-elevation Regressions on Independent Slopes Model (PRISM) dataset, and the NOAA National Weather Service (NWS) precipitation frequency datasets (NOAA, 2015). Basin-characteristic names, descriptions, units of measure, and sources of information are listed in table 6. These variables can be broadly characterized by topography, climate, geology, soils, and land use type.

Magnitude and Frequency of Flood Discharges at Gaged Sites

Flood-frequency analysis is a statistical technique used to estimate the magnitude and frequency of streamflow at a streamgage site. The objective of frequency analysis is to relate the magnitude of streamflow to its frequency of occurrence through probability distribution that describes flood risk. The probabilities computed correspond to the AEP, the probability in any year that a flood threshold is exceeded.

The flood-frequency analyses in this report follow the methodology described in the current version of the national guidelines for flood-frequency analyses, Bulletin 17C (England and others, 2018). The guidelines for flood-frequency analyses have undergone several updates since they were first published in 1967. Recent (2018) updates include the following: adoption of a generalized representation of flood data that allows for interval and censored data types; a new method, called the EMA, which extends the method of moments so that it can accommodate interval data; a generalized approach to identification of low outliers in flood data using the MGBT low-outlier test; and an improved method for deriving regional skew coefficients and computing confidence intervals.

The Bulletin 17C methodology continues to prescribe the Pearson type III distribution with log transformation of the flood data (LP3 distribution) as the basic distribution for defining the annual flood series (U.S. Water Resources Council, 1967, 1976; Interagency Advisory Committee on Water Data, 1981). The LP3 distribution is a three-parameter distribution that requires estimates of the mean, the standard deviation, and the skew coefficient of the population of logarithms of annual peak discharge at each gaged site. The mean, the standard deviation, and the skew coefficient, which describe the midpoint, slope, and curvature of the peak-flow frequency curve, respectively, can be estimated from the available sample data (annual peak discharges). The basic equation for determining flood frequency from the three parameters is the following:

$$\log Q_p = X + K_p S, \tag{1}$$

Table 6. Basin and climatic characteristics considered for use as explanatory variables in the regional regression analysis for estimating flood discharges in Connecticut.

[NHD, National Hydrography Dataset; NAVD 88, North American Vertical Datum of 1988; PRISM, Parameter-elevation Regressions on Independent Slopes Model; NOAA, National Oceanic and Atmospheric Administration; SSURGO, Soil Survey Geographic; NLCD, National Land Cover Database; NGD, National Geospatial Data Asset; MRLC, Multi-Resolution Land Characteristics]

Characteristic	StreamStats variable name	Definition	Units	Data sources
			Topography	
Drainage area	DRNAREA	Area that drains to a point on a stream	Square miles	Basin boundaries coverage (internal)
Longitude	LNG_GAGE	Latitude of the streamgage along the NHD flowline	Decimal degrees	U.S. Geological Survey National Water Information System (U.S. Geological Survey, 2016b; https://waterdata.usgs.gov/nwis/)
Latitude	LAT_GAGE	Longitude of the streamgage along the NHD flowline	Decimal degrees	U.S. Geological Survey National Water Information System (U.S. Geological Survey, 2016b; https://waterdata.usgs.gov/nwis/)
Latitude of basin centroid	LAT_CENT	Latitude of the basin centroid	Decimal degrees	U.S. Geological Survey National Water Information System (U.S. Geological Survey, 2016b; https://waterdata.usgs.gov/nwis/)
Longitude of basin centroid	LONG_CENT	Longitude of the basin centroid	Decimal degrees	U.S. Geological Survey National Water Information System (U.S. Geological Survey, 2016b; https://waterdata.usgs.gov/nwis/)
Basin perimeter	PERI	Perimeter following the basin boundary	Miles	Basin boundaries coverage (internal)
Total length of streams	TOT_STR	Total length of perennial rivers in the basin	Miles	The National Map, National Hydrography Dataset, scale 1:24,000 (U.S. Geological Survey, 2017b; https://viewer.nationalmap.gov/viewer/).
Stream density	STRDEN	Total miles of rivers divided by drainage area	Miles/square miles	The National Map, National Hydrography Dataset, scale 1:24,000 (U.S. Geological Survey, 2017b; https://viewer.nationalmap.gov/viewer/)
Minimum basin elevation	MINBELEV	Minimum basin elevation	Feet	The National Map, National Elevation Dataset, 10-meter resolution, NAVD 88 (U.S. Geological Survey, 2017a; https://viewer.nationalmap.gov/viewer/)
Maximum basin elevation	ELEVMAX	Maximum basin elevation	Feet	The National Map, National Elevation Dataset, 10-meter resolution, NAVD 88 (U.S. Geological Survey, 2017a; https://viewer.nationalmap.gov/viewer/)
Mean basin elevation	ELEV	Mean basin elevation	Feet	The National Map, National Elevation Dataset, 10-meter resolution, NAVD 88 (U.S. Geological Survey, 2017a; https://viewer.nationalmap.gov/viewer/)
Basin relief	RELIEF	Maximum basin elevation minus the minimum basin elevation (at the streamflow-gaging station)	Feet	The National Map, National Elevation Dataset, 10-meter resolution, NAVD 88 (U.S. Geological Survey, 2017a; https://viewer.nationalmap.gov/viewer/)
Mean basin slope	SLOPE	Mean basin slope in the drainage basin derived from the intersection of basin polygon coverages and DEMs	Percent	The National Map, National Elevation Dataset, 10-meter resolution, NAVD 88 (U.S. Geological Survey, 2017a; https://viewer.nationalmap.gov/viewer/)
			Climate	
Mean annual precipitation PRISM 1971–2000	PRECPRIS90	Basin average mean annual precipitation for 1971 to 2000 from PRISM	Inches	30-year (1971–2000) normal data (PRISM Climate Group, 2012; http://www.prism.oregonstate.edu/normals/)

Table 6. Basin and climatic characteristics considered for use as explanatory variables in the regional regression analysis for estimating flood discharges in Connecticut.—Continued

[NHD, National Hydrography Dataset; NAVD 88, North American Vertical Datum of 1988; PRISM, Parameter-elevation Regressions on Independent Slopes Model; NOAA, National Oceanic and Atmospheric Administration; SSURGO, Soil Survey Geographic; NLCD, National Land Cover Database; NGD, National Geospatial Data Asset; MRLC, Multi-Resolution Land Characteristics]

Characteristic	StreamStats variable name	Definition	Units	Data sources
		Climate—Continued		
Mean annual precipitation PRISM 1981–2010	PRECPRIS90	Basin average mean annual precipitation for 1981 to 2010 from PRISM	Inches	30-year (1981–2010) normal data (PRISM Climate Group, 2012; http://www.prism.oregonstate.edu/normals/)
Mean annual temperature	AVETEMP	Basin average mean annual temperature from 1981 to 2000 from PRISM	Degrees	30-year (1981–2010) normal data (PRISM Climate Group, 2012; http://www.prism.oregonstate.edu/normals/)
24-hour 2-year precipitation	I24H2Y	Maximum 24-hour precipitation that occurs on average once in 2 years	Inches	National Weather Service Precipitation Frequency Data Server (NOAA, 2015)
24-hour 5-year precipitation	I24H5Y	Maximum 24-hour precipitation that occurs on average once in 5 years	Inches	National Weather Service Precipitation Frequency Data Server (NOAA, 2015)
24-hour 25-year precipitation	I24H10Y	Maximum 24-hour precipitation that occurs on average once in 10 years	Inches	National Weather Service Precipitation Frequency Data Server (NOAA, 2015)
24-hour 50-year precipitation	I24H25Y	Maximum 24-hour precipitation that occurs on average once in 25 years	Inches	National Weather Service Precipitation Frequency Data Server (NOAA, 2015)
24-hour 100-year precipitation	I24H50Y	Maximum 24-hour precipitation that occurs on average once in 50 years	Inches	National Weather Service Precipitation Frequency Data Server (NOAA, 2015)
24-hour 100-year precipitation	I24H100Y	Maximum 24-hour precipitation that occurs on average once in 100 years	Inches	National Weather Service Precipitation Frequency Data Server (NOAA, 2015)
24-hour 200-year precipitation	I24H200Y	Maximum 24-hour precipitation that occurs on average once in 200 years	Inches	National Weather Service Precipitation Frequency Data Server (NOAA, 2015)
24-hour 500-year precipitation	I24H500Y	Maximum 24-hour precipitation that occurs on average once in 500 years	Inches	National Weather Service Precipitation Frequency Data Server (NOAA, 2015)
		Soils		
Hydrologic soil group A	SOILA	Hydrologic soil group A	Percent	Soil Survey Geographic (SSURGO) database (Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture, 2017; https://sdmdataaccess.sc.egov.usda.gov)
Hydrologic soil group B	SOILB	Hydrologic soil group B	Percent	Soil Survey Geographic (SSURGO) database (Soil Survey Staff, Natural Resources Conservation Service, U.S. Department of Agriculture, 2017; https://sdmdataaccess.sc.egov.usda.gov)
Hydrologic soil group C	SOILC	Hydrologic soil group C	Percent	Soil Survey Geographic (SSURGO) database (Soil Survey Staff, Natural Resources Conservation Service, U.S. Department of Agriculture, 2017; https://sdmdataaccess.sc.egov.usda.gov)

Table 6. Basin and climatic characteristics considered for use as explanatory variables in the regional regression analysis for estimating flood discharges in Connecticut. —Continued

[NHD, National Hydrography Dataset; NAVD 88, North American Vertical Datum of 1988; PRISM, Parameter-elevation Regressions on Independent Slopes Model; NOAA, National Oceanic and Atmospheric Administration; SSURGO, Soil Survey Geographic; NLCD, National Land Cover Database; NGD, National Geospatial Data Asset; MRLC, Multi-Resolution Land Characteristics]

Characteristic	StreamStats variable name	Definition	Units	Data sources
		Soils—Continued		
Hydrologic soil group D	SOILD	Hydrologic soil group D	Percent	Soil Survey Geographic (SSURGO) database (Soil Survey Staff, Natural Resources Conservation Service, U.S. Department of Agriculture, 2017; https://sdmdataaccess.sc.egov.usda.gov)
Hydrologic soil groups C, C/D, or D	SOILCorD	Hydrologic soil groups C, C/D, or D	Percent	Soil Survey Geographic (SSURGO) database (Soil Survey Staff, Natural Resources Conservation Service, U.S. Department of Agriculture, 2017; https://sdmdataaccess.sc.egov.usda.gov)
		Geology		
Coarse-grained stratified glacial deposits	CGSD	Area of basin classified as coarse-grained (sand and gravel) stratified glacial deposits and coarse-grained stratified deposits over fine-grained (very fine sand, silt, and clay) stratified deposits	Percent	Soller and others (2012)
Glacial till	TILL	Area of basin classified as glacial till	Percent	Soller and others (2012)
		Land Cover		
Total forest area	Forest	Area of basin classified as forested (deciduous forest, deciduous shrubland, evergreen forest and mixed forest were aggregated into forest)	Percent	NLCD 2011 land cover—NGD land use land cover (2011 ed., amended Oct 10, 2014), MRLC consortium (U.S. Geological Survey, 2014a; http://viewer.nationalmap.gov/viewer/)
Water	Water	Area of basin classified as water	Percent	NLCD 2011 land cover— NGD land use land cover (2011 ed., amended Oct 10, 2014), MRLC consortium (U.S. Geological Survey, 2014a; http://viewer.nationalmap.gov/viewer/)
Total wetlands	Wetland	Area of basin classified as wetland (woody wetlands and emergent herbaceous were aggregated into wetlands)	Percent	NLCD 2011 land cover— NGD land use land cover (2011 ed., amended Oct 10, 2014), MRLC consortium (U.S. Geological Survey, 2014a; http://viewer.nationalmap.gov/viewer/)
Storage	StorNLCD	Area of lakes, ponds, reservoirs, and wetlands in a basin	Percent	NLCD 2011 land cover—NGD land use land cover (2011 ed., amended Oct 10, 2014), MRLC consortium (U.S. Geological Survey, 2014a; http://viewer.nationalmap.gov/viewer/)
Storage area of lakes, ponds, and wetlands	StorNHD	Area of rivers, lakes, ponds, and wetlands	Percent	NLCD 2011 land cover—(NGD land use land cover (2011 ed., amended Oct 10, 2014), MRLC consortium (U.S. Geological Survey, 2014a; http://viewer.nationalmap.gov/viewer/)
Area of wetlands	StorFWS	Area of total wetlands	Percent	National Wetlands Map (U.S. Fish and Wildlife Service, 201X; https://www.fws.gov/wetlands/data/mapper.html)

where

Q_p is the peak discharge for the annual exceedance probability,

X is the mean of the logarithms of the annual peak discharge,

K_p is a factor based on the weighted skew coefficient and the exceedance probability, P, which can be obtained from the appendix in Bulletin 17C, and

S is the standard deviation of the logarithms of the annual peak discharge, which is a measure of the degree of variation in the annual values about the mean value.

The USGS computer program PeakFQ version 7.1 (Flynn and others, 2006) was used to derive the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent AEP for gage sites. The program PeakFQ implements the Bulletin 17C procedures for flood-frequency analysis of streamflow records. The output from PeakFQ includes estimates of the parameters of the LP3 distribution, including the logarithmic mean, standard deviation, skew, and mean squared error of the skew. The output graph includes the fitted frequency curve, systematic peaks, low outliers, censored peaks, interval peaks, historical peaks, thresholds, and confidence limits.

Regional Skew Coefficient

A skew coefficient is used in defining the probability distribution of the annual peak flows at a streamgage to better reflect regional and long-term conditions. Bulletin 17C recommends the skew coefficient used in defining the probability distribution be a weighted average of the (at-site) station skew and a regional skew estimated from long-term streamgages and representative of regional peak-flow characteristics (Griffis and Stedinger, 2009). The at-site skew coefficient is sensitive to extreme peak flows and might not provide an accurate measure of the true skew of peak flows that occur at a site, particularly for streamgages with short periods of record. Since the mid-1970s, the published generalized skew map by the Interagency Advisory Committee on Water Data (1981, plate 1) has been used for deriving flood discharge estimates for gaged sites. In 2003, the skew map was superseded by a more accurate estimate of regional skew (0.34) for Connecticut developed by Ahearn (2003).

In Bulletin 17C, the recommended procedure for estimating regional skew is the Bayesian weighted-least squares/Bayesian generalized least squares (B-WLS/B-GLS) method (Veilleux and others, 2019). The B-WLS/B-GLS method has shown the ability to reduce the uncertainty in generalized skew estimates and allows the generalized skew estimate to take into account basin characteristics where applicable (Reis and others, 2005; Gruber and others, 2007; Gruber and Stedinger, 2008; Veilleux, 2011). Veilleux and others (2019) performed a regional skew analysis of streamgages in the New England using B-WLS/B-GLS methodology. The results of the regional skew analysis by Veilleux indicate a constant skew of 0.37 with an average variance of prediction of 0.14. The regional skew developed by Veilleux has the highest precision of all previous skew studies of streamgages in Connecticut and supersedes the regional skew coefficient by Ahearn (2003).

For stream sites with unregulated flow, the flood-frequency estimates were computed using a weighted average of the at-site station skew and a regional skew. For stream sites with regulated flow, the flood-frequency estimates were computed by fitting the annual peaks from the regulated flow record to the LP3 distribution using the at-site station skew.

Expected Moments Algorithm Frequency Analysis and Multiple Grubbs-Beck Test for Detecting Low Outliers

The new guidelines in Bulletin 17C include the EMA and MGBT techniques for flood-frequency determinations. Both EMA and MGBT have been shown to provide more efficient, accurate estimates of the magnitude and frequency of flood discharges at gaged sites than those generated using the previous Bulletin 17B (Interagency Advisory Committee on Water Data, 1981) guidelines when a peak-flow record contains gaps, historical flood measurements, censored data, or low outliers (Cohn and others, 1997, 2013; Paretti and others, 2014a, b).

Generally, peak-flow records of streamgages contain two types of data: (1) systematic, with a peak-flow value recorded for each year; and (2) historical or isolated measurements made outside the systematic period of record (typically during extreme hydrologic conditions). In these two general types of peak-flow data, some peaks can be identified as “censored,” which means that the actual peak flow is uncertain and is documented as greater than or less than some value. The EMA methods allow better handling of these data types than was previously possible using strict Bulletin 17B procedures.

The knowledge that a particular flow would have been noticed and measured if it had occurred provides valuable information for the peak-flow frequency analysis. The EMA method allows the use of perception thresholds and flow intervals to describe conditions outside the systematic record. Perception thresholds describe the minimum and maximum peak flows that would have been measured if they had occurred (Veilleux and others, 2014). Flow intervals describe the uncertainty associated with a peak flow. Defining the perception thresholds is based on historical documentation and anecdotal information. For this study, perception thresholds generally were set to (0, infinity) for the systematic record, (peak, infinity) for any historical peak measurements, and (infinity, infinity) for any gaps in the systematic record if no additional information was available (Ahearn, 2020). Additional information on the generalized representation of flood data that allows for interval and censored data types and the EMA technique to accommodate interval data is provided Bulletin 17C (England and others, 2018).

Several peak-flow records analyzed contained low outliers or peaks that depart significantly from the data population. Low outliers can have high leverage or influence in fitting the frequency curve to the entire record of peak flows, which results in a poor fit of the frequency curve at lower AEPs (1-percent AEP discharge). The peak-flow statistics most frequently used for flood protection and infrastructure design are the discharges with low AEPs. Additionally, low outliers often are considered to reflect physical processes that are not necessarily related to the processes associated with large flood events, and low outlier use in the frequency analysis should be limited (Cohn and others, 2013). The MGBT technique, described in Cohn and others (2013), objectively and systematically detects and removes potentially influential low-flow (PILF) outliers below a PILF threshold and can be used in concert with EMA methods in the USGS PeakFQ program. For streamgages used in this study, removing the PILFs generally produced a better fit of the frequency curve for low AEP flood discharges (large peak flows).

Flood-Frequency Estimates for Gaged Sites

Flood-frequency estimates for 141 streamgages in Connecticut and 11 streamgages in adjacent States—New York (3 streamgages), Massachusetts (3 streamgages), and Rhode Island (5 streamgages)—were derived by fitting LP3 distribution to the records of annual peak flows and applying the EMA and MGBT (Ahearn, 2020). The flood discharges from the frequency analyses for the 152 streamgages have AEPs of 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent (recurrence intervals of 2, 5, 10, 25, 50, 100, 200, and 500 years, respectively).

Flood-frequency estimates for regulated sites are presented for the pre-regulation and post-regulation periods, if 15 or more years of pre-regulation or 15 or more years of post-regulation record were available. No adjustments were applied to the annual peak flows or made to the flood-frequency estimates for available storage in the reservoirs before or during floods, nor for changes in regulation procedures during the period of regulation.

Development of Regional Regression Equations for Estimating Flood Discharges

Multiple-linear regression is used to develop the relation between two or more basin characteristics (called explanatory variables) and a streamflow statistic (called a response variable) by fitting a linear equation to the data. With the development of regression equations, measurements of basin characteristics at an ungaged stream site can be used to estimate flood discharges at that site. The general form of equations developed from multiple linear-regression analyses is shown in the following equation:

$$Y_i = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n + e_i \quad (2)$$

where

Y_i is the response variable (estimate of the streamflow statistic computed from observed streamflow) for site i,

X_1 to X_n are the n explanatory variables (basin characteristics) for site i,

b_0 to b_n are the n+1 regression model coefficient, and

e_i is the residual error (difference between the observed and predicted values of the response variable) for site i.

The basic assumptions of regression analyses are (1) the model adequately describes the linear relation between the response and explanatory variables, (2) the mean of e_i is zero, (3) the variance of e_i is constant and independent of the values of X_n, (4) the values of e_i are distributed normally, and (5) the values of e_i are independent of each other. Because streamflow data are naturally correlated spatially and temporally, the last assumption is not completely satisfied with the use of ordinary least squares (OLS) regression. As a result, generalized least squares (GLS) regression was used to develop the final equations for estimating AEP flood discharges. An overview of the OLS and GLS multiple-linear regression techniques used to develop the initial and final equations are presented in the following two sections.

Streamgages were selected for the regression analysis using the following criteria: (1) no substantial effects of flood-control regulation are observed in the basins; (2) no substantial effects of urbanization or other man-made influences, such as channel improvements, are observed in the basins; (3) the basin has less than 15 percent of the land cover designated as commercial, industrial, or medium or high-density development; (4) the station has a minimum of 15 years of record; and (5) the station is spatially independent. Regression analysis requires that data be as spatially independent as possible. When the drainage basins of two streamgages are nested, meaning that one is contained inside the other, and the sizes of the two basins are similar, the gages are considered redundant. Then, instead of providing two independent spatial observations depicting how basin characteristics are related to AEP flood discharges, these two basins will likely have the same hydrologic response to a given storm and thus represent only one spatial observation. Of the 152 streamgages with updated AEP flood discharges using data through 2015, 85 streamgages met the criteria and were used in the regional regression analysis.

Ordinary Least Squares Regression

OLS regression analyses were used to determine the best combinations of basin characteristics to use as explanatory variables in the multiple-linear regression equations for estimating AEP discharges. Logarithmic transformations (base 10) were performed for all response variables and for selected

explanatory variables used in the regression analyses. Data transformations were needed to obtain a more constant variance of the residuals about the regression line and to linearize the relation between the response variable and the explanatory variables. A constant of 1 was added to select explanatory variables expressed in percent.

OLS regression analyses were performed using Spotfire S+ statistical software (TIBCO Software Inc., 2008). The automated statistical methods "all-possible subsets" was used for selecting the explanatory variables. The selection method determined the statistical contribution of the explanatory variable, and variables were retained or deleted based on their statistical importance. In "all-possible subsets," all the equations created from all possible combinations of explanatory variables were examined and the coefficient of determination (R^2) was used to check for the best combination of variables. With this method, each explanatory variable can be included or excluded independently of the other explanatory variables.

Explanatory variables in the OLS models were selected to minimize the standard error of estimate (SEE), Mallows' Cp, and PRESS statistics and to maximize the adjusted coefficient of determination (adjusted-R^2; see glossary). The OLS models were evaluated to determine their adequacy, including graphical relations and residual plots, variance inflation factor (VIF), Cook's D statistic (Cook, 1977; Helsel and Hirsch, 2002), and high-leverage points. The selection of explanatory variables, and the signs and magnitudes of their respective regression coefficients, were each evaluated to ensure hydrologic validity in the context of AEP flood discharges. Correlation between explanatory variables and VIF was used to assess multicollinearity in the regression models. Multicollinearity problems were identified with a regression-diagnostics tool implemented in the USGS library version 4.0 (Lorenz and others, 2011) for Spotfire S+ statistical software (TIBCO Software Inc., 2008) by checking each explanatory variable for a VIF greater than 2.

Smaller regions were evaluated using OLS regression analysis. The physiographic regions of Connecticut and boundaries of the U.S. Environmental Protection Agency's Northeastern Coastal Zone and Northeastern Highland Level III ecoregions were used to subdivide the streamgages and investigate smaller flood regions (Omernik, 1995; U.S. Environmental Protection Agency, 2010). Results from the OLS analysis did not indicate that smaller regions improve the accuracy of the model enough to warrant separate models. Though the Northeastern Coastal Zone (ecoregion) shows a slight improvement in predictive accuracy compared to the statewide model, the Northeastern Highlands ecoregion has much lower predictive accuracy than the statewide models. The explanatory variables in the statewide model capture regional variation. A tradeoff in a statewide model is likely some loss of predictive power in the Northeastern Coastal Zone. The Northeastern Highlands ecoregion is a difficult region to model. It has the fewest gages, and the explanatory variables do not predict precise estimates of flow. Collecting more data in the Northeastern Highlands will improve the ability to estimate the magnitude and frequency of peak flows, particularly in that region of Connecticut.

From the OLS regression analysis, four potential explanatory variables (drainage area (DRNAREA), in square miles; maximum 24-hour precipitation that occurs on average once in xx years (I24HxxY), in inches; percentage of area of hydrologic Soil Type C or D from SSURGO (SOILCorD), in decimal percent; and storage, in percent of basin classified as open water or wetlands) were identified for further analysis using the more robust GLS regression analysis.

Generalized Least Squares Regression

GLS multiple-linear regression was used to develop the final set of regression equations for estimating AEP flood discharges for Connecticut. GLS regression, as described by Stedinger and Tasker (1985), Tasker and Stedinger (1989), and Griffis and Stedinger (2007), is a method that weights streamgages in the regression according to differences in streamflow reliability (record lengths) and variability (record variance) and according to spatial cross-correlations of concurrent streamflow among streamgages. The GLS regression techniques give less weight to streamgages that have shorter periods of record and more weight to streamgages with longer periods of record. Less weight is also given to streamgages where concurrent peak flows are correlated because of the geographic proximity to other streamgages. Compared to OLS regression, GLS regression provides improved estimates of AEP discharges and improved estimates of the predictive accuracy of the regression equations (Stedinger and Tasker, 1985).

GLS regression analyses were performed using the USGS weighted-multiple-linear regression program written in R programming language, WREG version 2.02 (USGS, 2014b). Output of WREG provides various measures of the reliability of the regression equations including the following: the average variance of prediction (AVP, in log units), the standard error of prediction (SEP, in percent), the standard error of estimate (SEE, in percent), the pseudo coefficient of determination (pseudo-R^2), the mean squared error (in log units), the root mean squared error (in percent), and leverage and influence of individual observations on the regression. Equations for calculating these metrics are available in Eng and others (2009) and Gotvald and others (2012).

Regression Equations for Estimating Flood Discharges at Ungaged Stream Sites

The final set of regional regression equations were selected based on the lowest model error, expressed as SEP and SEE; amount of variability in the flood estimates explained by explanatory variables, expressed as the pseudo-R^2; random patterns in residuals plots; low number of data points with high leverage or influence; and consistency in

model form across all estimated statistics. Residual plots of the simulated and observed values were used as a diagnostic tool for checking the validity of the models. Overall, the model appears to fit the data reasonably well. There was no evidence that any of the model assumptions have been violated. The plots show that the residuals are nearly equally distributed around zero for each of the AEP flood discharges. The model shows no curvature or changing variance. Furthermore, the residuals show no spatial pattern, indicating no geographical biases in the statewide models or need for additional explanatory variables (fig. 2).

Leverage and influence statistics for the GLS analysis from the WREG program were used to identify possible problem streamgages used in the regression. Streamgages that had leverage or influence metrics that exceeded the thresholds calculated by WREG, especially those that had both high leverage and high influence, were evaluated for potential erroneous data reporting or conditions that would make the streamgage ineligible for regression. Several streamgages were identified by the WREG program as having high influence or leverage but were not excluded in the final regression. A reasonable hydrologic justification for excluding the data could not be identified.

Multicollinearity problems were investigated using VIF. Multicollinearity problems can increase the variance of the coefficient estimates and make the estimates sensitive to minor changes in the model. VIF values greater than 5 to 10 indicate that an explanatory variable is highly correlated to other explanatory variables and inferences based on the model can be misleading or erroneous. None of the explanatory variables had a VIF value greater than 1, indicating no multicollinearity problems were found.DRNAREA, I24HxxY, and percentage of basin with SOILCorD were found to be the best explanatory variables for estimating peak-flow frequency for ungaged sites in Connecticut. The explanatory variables in the final regression equations were statistically significant at the 95-percent confidence level (p<0.05). The same explanatory variables were used to develop all seven regression equations to minimize the possibility of predictive inconsistencies between estimates of different probabilities. The final regional regression equations and select performance metrics for the 50- through 0.2-percent AEP flood discharges are listed in table 7.

The most significant explanatory variable (strongest predictor of flow), DRNAREA, is related positively to AEP flood discharges; ungaged basins with larger drainage areas will produce larger estimates of AEP flood discharges than ungaged basins with smaller drainage areas. The second most significant explanatory variable, I24HxxY, is also related positively to AEP flood discharges (fig. 3). Intuitively, the larger estimates of precipitation will produce larger estimates of AEP flood discharges. The third explanatory variable, SOILCorD, is related positively to AEP flood discharges (fig. 4). Hydrologic soil types C and D have higher runoff potential than hydrologic soil groups A and B due to low infiltration rates when wet. Basins with larger percentages of SOILCorD will produce larger estimates of AEP flood discharges than basins with smaller percentages of SOILCorD. The basin boundaries used to compute the drainage areas were obtained from existing basin boundary datasets or delineated from the 10-meter-resolution National Elevation Dataset (USGS, 2017a). The 24-hour, 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year maximum precipitation (in inches) was obtained from the NOAA NWS Precipitation Frequency Data Server database (NOAA, 2015; https://hdsc.nws.noaa.gov/hdsc/pfds/pfds_series.html). The explanatory variable—SOILCorD—was determined as the percentage of the basin area with hydrologic soil groups C, C/D, or D from the Soil Survey Geographic (SSURGO) database (Soil Survey Staff, Natural Resources Conservation Service, U.S. Department of Agriculture, 2017; https://sdmdataaccess.sc.egov.usda.gov).

Accuracy and Limitations of the Regression Equations

Several performance metrics from the WREG program further the understanding of the accuracy of the equations. Performance metrics of the final regional regression equations include variance of prediction, average variance of prediction, standard error of prediction, and pseudo-R^2. A measure of the uncertainty in a regression equation estimate for a site (i) is the variance of prediction ($V_{p,i}$). The $V_{p,i}$ is the sum of the model error variance and sampling error variance (Eng and others, 2009) and is computed using the following equation:

$$V_{p,i} = \gamma^2 + \mathrm{MSE}_{s,i} \quad (3)$$

where

γ^2 is the model error variance, and

$\mathrm{MSE}_{s,i}$ is the sampling mean squared error for site i.

Model error measures the ability of a set of explanatory variables to estimate the values of peak flows calculated from the streamgage records that were used to develop the equation. Sampling error measures the ability of a finite number of streamgages with a finite number of recorded annual peak flows to describe the true peak flows for a streamgage. Assuming the explanatory variables for the streamgages in a regression analysis are representative of all streamgages in the region, the average accuracy of prediction for a regression equation is determined by computing the AVP for n number of streamgages using the following equation:

$$\mathrm{AVP} = \gamma^2 + \mathrm{MSE}_{s,i} / (n) \quad (4)$$

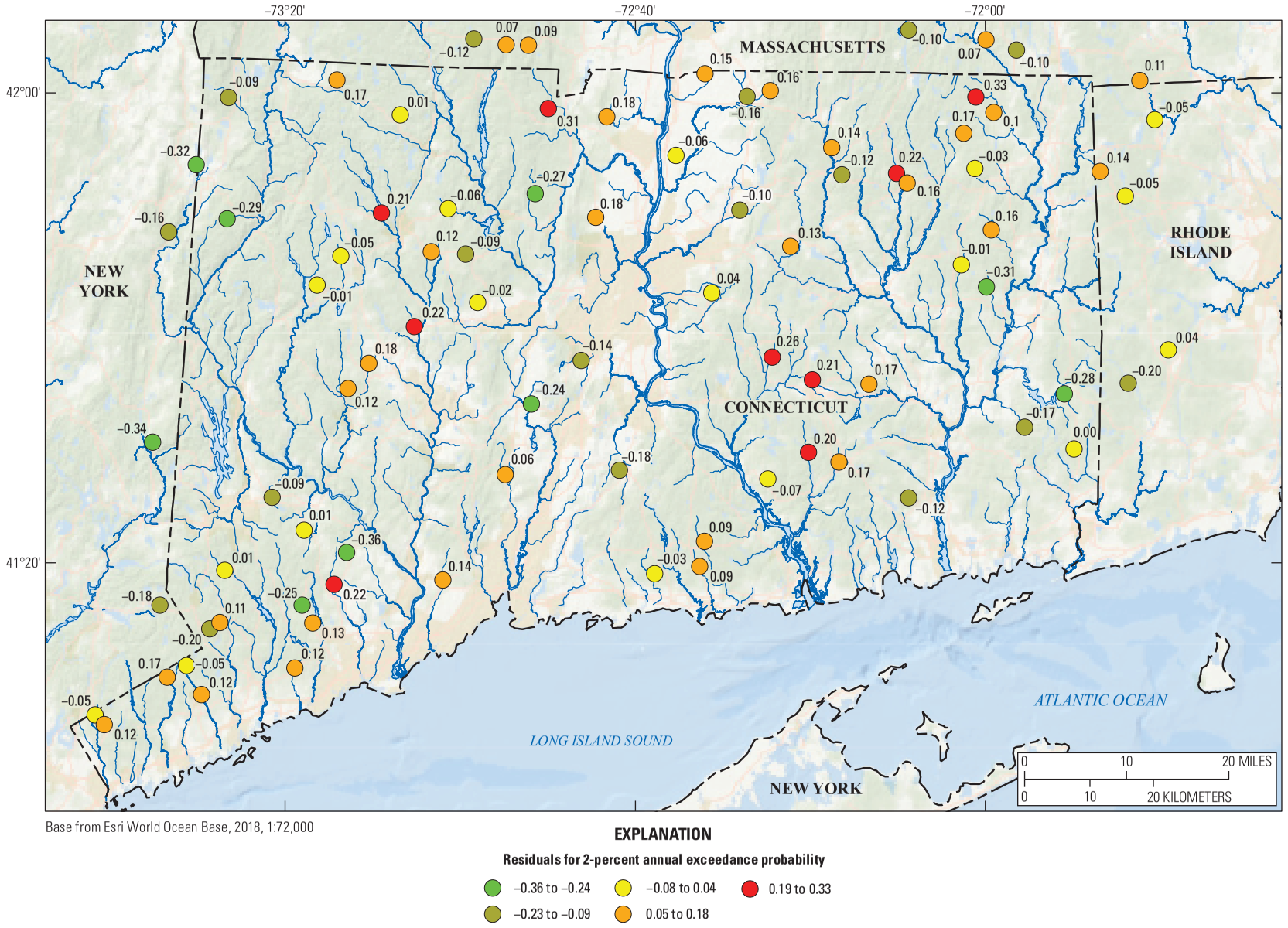


Figure 2. Generalized least squares regression residuals for the 2-percent annual exceedance probability discharge plotted at the basin centroids of streamgages in Connecticut.

Table 7. Regional regression equations and performance metrics for estimating flood discharges for select annual exceedance probabilities for ungaged streams in Connecticut.

[ft³/s, cubic foot per second; Q_p, peak discharge for the annual exceedance probability; DRNAREA, drainage area (square miles); I24HxxY, maximum 24-hour precipitation that occurs on average once in xx years (inches); and SOILCorD, percentage of drainage area classified as hydrologic soil groups C, C/D, and D (decimal percent)]

Annual exceedance probability (percent)	Regional regression equation (ft³/s)	Pseudo coefficient of determination (percent)	Mean squared error (log units)	Root mean squared error (percent)	Model error variance (log units)	Standard error of estimate (percent)	Average variance of prediction (log units)	Standard error of prediction (percent)
50	Q_{50}=10^−1.015*((DRNAREA)^0.819)*10^(0.618*I24H2Y)*10^(0.586*(SOILCorD+1)))	95.5	0.016	29.57	0.012	25.21	0.013	26.53
20	Q_{20}=10^−1.064*((DRNAREA)^0.803)*10^(0.541*I24H5Y)*10^(0.505*(SOILCorD+1)))	95.7	0.016	29.48	0.011	24.68	0.013	26.32
10	Q_{10}=10^−0.775*((DRNAREA)^0.795)*10^(0.426*I24H10Y)*10^(0.470*(SOILCorD+1)))	95.1	0.019	32.13	0.013	26.44	0.015	28.41
4	Q_4=10^−0.432*((DRNAREA)^0.790)*10^(0.322*I24H25Y)*10^(0.442*(SOILCorD+1)))	94.1	0.024	36.62	0.015	29.06	0.018	31.47
2	Q_2=10^−0.255*((DRNAREA)^0.789)*10^(0.274*I24H50Y)*10^(0.426*(SOILCorD+1)))	93.1	0.028	40.21	0.018	31.55	0.021	34.29
1	Q_1=10^−0.104*((DRNAREA)^0.789)*10^(0.238*I24H100Y)*10^(0.413*(SOILCorD+1)))	92.1	0.034	44.06	0.021	34.04	0.024	37.10
0.5	$Q_{0.5}$=10^0.372*((DRNAREA)^0.785)*10^(0.167*I24H200Y)*10^(0.390*(SOILCorD+1)))	90.7	0.040	48.31	0.024	37.16	0.029	40.56
0.2	$Q_{0.2}$=10^0.740*((DRNAREA)^0.784)*10^(0.120*I24H500Y)*10^(0.370*(SOILCorD+1)))	88.9	0.048	54.07	0.030	41.12	0.035	44.97

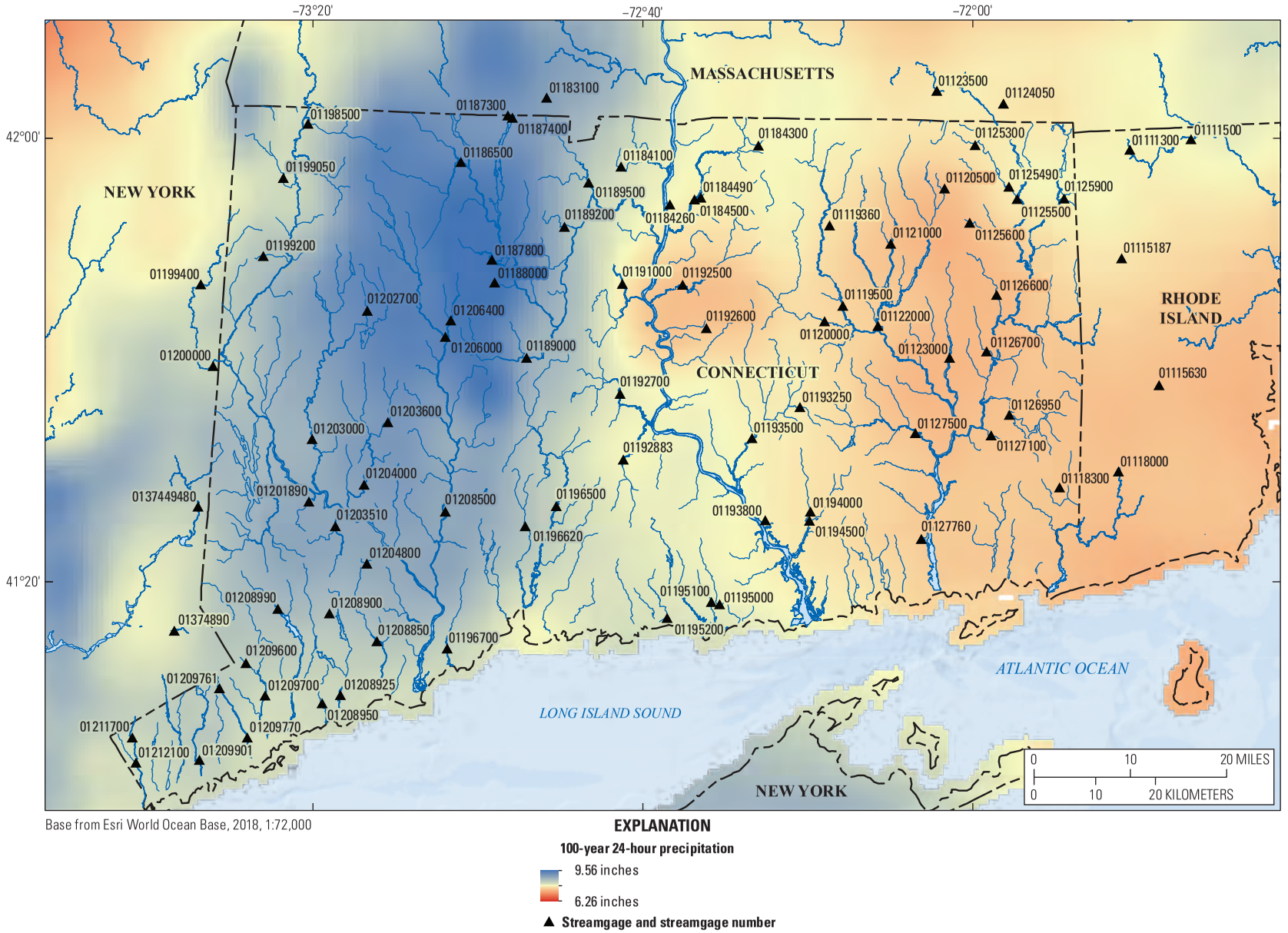


Figure 3. Distribution of the National Weather Service 24-hour, 1-percent annual exceedance probability precipitation with locations of the 85 streamgages in Connecticut and adjacent States used in development of the peak-flow regression equations. Precipitation frequency estimates are from the National Weather Service Precipitation Frequency Data Server (National Oceanic and Atmospheric Administration, 2015).

Base from Esri World Ocean Base, 2018, 1:72,000

EXPLANATION

Soil groups C, C/D, or D

Streamgage and streamgage number

Figure 4. Distribution of the Soil Survey Geographic (SSURGO) hydrologic soil groups C, C/D, and D with locations of the 85 streamgages in Connecticut and adjacent States used in the development of the peak-flow regression equations (Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture, 2017).

A more traditional measure of the accuracy is the standard error of prediction (SEP), which is simply the square root of the variance of prediction. Approximately two-thirds of the estimates obtained from a regression equation for ungaged sites have errors less than the SEP (Helsel and Hirsch, 2002). The SEP for a regression equation can be computed in percent using AVP (in log units) and the following transformation:

$$SEP = 100 \times [10^{2.3026(AVP)} - 1]^{0.5} \quad (5)$$

where

SEP is the standard error of prediction, in percent.

The SEP is a measure of the spread or dispersion of the predicted value from the observed value; hence, the lower the values, the less the expected spread of predictions around the true (unknown) value. The coefficient of determination (R^2) measures the proportion of the variation in the dependent variable explained by the independent variable that is predictable from the independent variable in OLS regressions. For GLS regressions, a more appropriate performance metric than R^2 is pseudo-R^2. Unlike the R^2 metric, pseudo-R^2 is based on the variability in the dependent variable explained by the regression after removing the effect of the time-sampling error.

Performance metrics for the final set of regional regression equations are given in table 7. The pseudo-R^2 for the regression equations ranged from 95.7 percent for the 20-percent AEP to 88.9 percent for the 0.2-percent AEP. The SEE and SEP values for the regression equations ranged from 24.7 and 26.3 percent for the 20-percent AEP to 41.1 and 45.0 percent for the 0.2-percent AEP, respectively. The SEP values were slightly improved in the current study compared to the 2004 study (from about 6-percent improvement at 50-percent AEP [2-year flood] to less than 1-percent improvement at the 1-percent AEP [100-year flood]).

The regression equations developed in this study apply to stream sites in Connecticut where peak discharges are not affected significantly by regulation, diversion, or backwater. The applicability and accuracy of the equations depend on whether the basin characteristics measured for an ungaged stream site are within the range of the characteristic values used to develop the regression equations (table 8). The equations also should be used with caution for ungaged stream sites with basin-characteristic values approaching the minimum or maximum limits, because the predictive errors of the equations increase with distance from the mean or median values of the explanatory variables and inconsistencies in the estimates may result. In addition, basin-characteristic measurements at ungaged sites should be computed using the same GIS datasets and measurement methods used in this study; the USGS StreamStats web-based GIS tool includes the same GIS data layers and measurement methods as used to develop the regression equations in this study.

Table 8. Ranges of explanatory variables used in the regional regression equations for estimating flood discharges in Connecticut.

Hydrologic characteristic (units)	Maximum	Minimum
Drainage area (square miles)	325	0.69
2-year, 24-hour rainfall (inches)	3.32	2.77
5-year, 24-hour rainfall (inches)	4.7	4
10-year, 24-hour rainfall (inches)	5.79	4.86
25-year, 24-hour rainfall (inches)	7.22	5.99
50-year, 24-hour rainfall (inches)	8.3	6.81
100-year, 24-hour rainfall (inches)	9.38	7.62
200-year, 24-hour rainfall (inches)	11.22	8.7
500-year, 24-hour rainfall (inches)	13.64	10.1
Hydrologic soil group C, C/D, or D (decimal percent)	0.945	0.118

Prediction Intervals of Regression Equations Estimates

Prediction intervals are useful indicators of the uncertainty inherent in the regression equations. Users of the regression equations may be interested in a measure of uncertainty for a flood discharge estimate at a particular site as opposed to the average uncertainty based on all streamgage data used to generate the regression equations. One such measure of uncertainty is the confidence interval of a prediction, or prediction interval. A prediction interval is given as the range in values of an estimated response variable over which the true value of the response variable occurs with some stated probability. The minimum and maximum values given in a 90-percent prediction interval for the 100-year peak flow for an ungaged site should be interpreted to mean that there is a 90-percent confidence that the true value of the 100-year peak flow is within the prediction interval. Tasker and Driver (1988) have shown that a 100 $(1-\alpha)$ prediction interval for a streamflow statistic estimated at an ungaged site from a regression equation can be computed as follows:

$$[Q/C] < Q < [QC] \quad (6)$$

where

Q is the flood magnitude for the ungaged site, in cubic feet per second; and

C is the confidence interval computed using the following equation:

$$C = 10\text{^}(T\text{*}SEP_{p,i}) \quad (7)$$

where

T is the Student's t-distribution value from a standard statistics table for a given confidence level and degree of freedom (for a 90-percent level $\alpha = 0.10$ and >100 degrees of freedom), and

$SEP_{p,i}$ is the standard error of prediction for site i; the value of $S_{p,i}$ is computed using the following equation:

$$SEP_{p,i} = [\gamma^2 + x_i U x_i']^{0.5} \quad (8)$$

where

γ^2 is the model error variance, in log units (computed using WREG and presented in table 9);

x_i is a row vector of the explanatory variables [DRNAREA, I24HxxY, and SOILCorD] for site i, augmented by a 1 as the first element;

U is the covariance matrix for the regression coefficients (table 9); and

x_i' is the transpose of x_i from Ludwig and Tasker (1993; see table 9 of this report).

The model error variance (γ^2) and the covariance matrix (U) were determined from the WREG GLS analysis and are reported in table 9. An example calculation of the 90-percent prediction interval is given for the Quinnipiac River for USGS streamgage 01196500 with the following characteristics: DRNAREA of 110 mi^2, I24H100Y of 8.53 in., and SOILCorD of 0.4207 decimal percent. The computed 1-percent AEP flood discharge from the regression equation is 13,300 ft^3/s. The procedure for computing the 90-percent prediction interval is as follows:

- Compute the standard error of prediction using matrix algebra to solve equation 8, where the model error γ^2 is retrieved from table 9, the x_i vector is in the form at x_i = {1, $\log_{10}$DRNAREA, I24HxxY, SOILCorD} and the covariance matrix (U) is retrieved from table 9,
- (γ^2 =0.021, x_i= {1, $\log_{10}$(110), I24H100Y (8.53), SOILCorD (0.4207+1)},$S_{p,i}$ = (0.021 + (0.00286))$^{0.5}$ =0.15446;
- Compute C from equation 7, ($C = 10^{(1.66*0.15446)}$ = 1.80718);
- Compute the 90-percent prediction interval from equation 6, (Q/C) <$Q_{1\%}$ <QC, or 7,360 ft^3/s < $Q_{1\%}$ < 24,000 ft^3/s, meaning that one can be 90 percent confident that the true value of the estimate for the site lies between 7,360 and 24,000 ft^3/s.

An application worksheet provided as a Microsoft Excel file is available for calculating flood discharges for a given AEP and user-selected prediction intervals (Ahearn and Veilleux, 2020). Appendix 2 includes a general description and screenshot of the worksheet. The worksheet is patterned after similar worksheets by Zarriello (2017) and Curran and others (2016).

Drainage-Area Only Regression Equations

Regression equations with one explanatory variable—drainage area—can provide quick estimates of flood discharges that are easier to calculate, although less accurate, than those computed by the regression equations with multiple explanatory variables shown in table 7. The drainage-area-only regression equations and their performance metrics are presented in table 10. The SEPs for the drainage-area-only equations are 8.5 to 3 percent higher than for the three-variable regression equations. The SEPs for the 10- and 0.1-percent AEP flood discharges are 36.3 and 42.4 percent for the drainage-area-only regression equations and 28.4 and 37.1 percent for the regression equations with multiple explanatory variables, respectively.

The regional exponent in each of the drainage-area only regression equations is the slope of the average linear logarithmic relation between drainage area and flood discharge for a selected AEP. The regional exponent can be used in an alternate method for adjusting flood-frequency data from a streamgage to locations upstream and downstream. This use of the method is to be limited to sites within 50 to 150 percent of the streamgage drainage area (Wandle, 1983). The regional exponents for selected AEP flood discharges ranging from 50 to 0.2 percent are in table 9.

Weighting of Streamgage Statistics and Regression Estimates

An estimate of AEP discharge at a streamgage can be improved by combining the regression equation estimate with the frequency curve computed from the streamgage record. A procedure recommended in Bulletin 17C (England and others, 2018) is to compute the AEP discharges from the regression equations and from the LP3 frequency analysis of the streamgage record and weight them by the inverse of their variances (eq. 9). The weighting procedure was only applied to streamgages that are unregulated and have limited urbanization. Discharges from regulated stations should not be weighted with regression equation estimates as these equations

Table 9. Model error variance and covariance values associated with selected annual exceedance probabilities used to determine prediction intervals for the Connecticut regional regression equations.

[Model error variance and covariance values were determined from the U.S. Geological Survey weighted-multiple-linear regression (WREG) program, version 2.02. Variance and covariance are in log units.; γ^2, regression model error variance (as used in eq. 8); GLSNet, Generalized Least Squares Network; U, covariance matrix (as used in eq. 8; the matrix horizontal and vertical variables are defined by the constant and the independent variables in the regression equations in the order they are given)]

Percent annual exceedance probability	Model error (γ^2, from GLSNet, in log units)	U (covariance matrix from WREG)			
50	0.012	0.150121456	−0.002881962	−0.014221043	−0.040368584
		−0.002881962	0.000594468	0.000298268	0.000517440
		−0.014221043	0.000298268	0.007274701	0.001057482
		−0.040368584	0.000517440	0.001057482	0.012321012
20	0.011	0.163813213	−0.002327223	−0.009032922	−0.034135235
		−0.002327223	0.000620062	0.000284794	0.000238519
		−0.009032922	0.000284794	0.007557628	−0.000543410
		−0.034135235	0.000238519	−0.000543410	0.008053048
10	0.013	0.177897628	−0.002426533	−0.009020613	−0.030816351
		−0.002426533	0.000744840	0.000359061	0.000161718
		−0.009020613	0.000359061	0.009230998	−0.000942228
		−0.030816351	0.000161718	−0.000942228	0.006142851
4	0.015	0.198520082	−0.002723581	−0.010275122	−0.027858473
		−0.002723581	0.000942556	0.000488384	0.000105127
		−0.010275122	0.000488384	0.011862813	−0.001202555
		−0.027858473	0.000105127	−0.001202555	0.004602554
2	0.018	0.219367699	−0.003067342	−0.012146441	−0.026767030
		−0.003067342	0.001127917	0.000608049	0.000079952
		−0.012146441	0.000608049	0.014321880	−0.001314005
		−0.026767030	0.000079952	−0.001314005	0.003908510
1	0.021	0.242310954	−0.003489404	−0.014440553	−0.026085650
		−0.003489404	0.001325347	0.000734776	0.000066694
		−0.014440553	0.000734776	0.016920388	−0.001375739
		−0.026085653	0.000066694	−0.001375739	0.003407436
0.5	0.024	0.208771318	−0.003554708	−0.017638995	−0.018389300
		−0.003554708	0.001569869	0.000924911	−0.000000373
		−0.017638995	0.000924911	0.020431633	−0.001417735
		−0.018389297	−0.000000373	−0.001417735	0.002141616
0.2	0.030	0.192440519	−0.003987762	−0.023226852	−0.013259300
		−0.003987762	0.001907400	0.001180929	−0.000036203
		−0.023226852	0.001180929	0.025123295	−0.001343871
		−0.013259301	−0.000036203	−0.001343871	0.001349216

Table 10. Simplified regional regression equations and performance metrics for estimating flood discharges for select annual exceedance probabilities for ungaged streams in Connecticut.

[ft^3/s, cubic foot per second; Q_p, peak discharge for the annual exceedance probability; DRNAREA, drainage area (square miles)]

Annual exceedance probability (percent)	Regional regression equation (ft^3/s)	Pseudo coefficient of determination (percent)	Mean squared error (log units)	Root mean squared error (percent)	Model error variance (log units)	Standard error of estimate (percent)	Average variance of prediction (log units)	Standard error of prediction (percent)	Regional exponent
50	Q_{50}=10^1.807*((DRNAREA)^0.778))	91.9	0.028	40.1	0.021	34.0	0.022	35.0	0.78
20	Q_{20}=10^2.053*((DRNAREA)^0.768))	92.0	0.032	43.0	0.020	33.8	0.022	35.0	0.77
10	Q_{10}=10^2.187*((DRNAREA)^0.764))	91.5	0.036	46.1	0.022	34.7	0.023	36.3	0.76
4	Q_4=10^2.331*((DRNAREA)^0.762))	90.9	0.043	50.6	0.023	36.0	0.025	37.8	0.76
2	Q_2=10^2.424*((DRNAREA)^0.761))	90.0	0.049	54.3	0.025	37.7	0.028	39.8	0.76
1	Q_1=10^2.508*((DRNAREA)^0.762))	88.9	0.055	58.3	0.028	40.1	0.031	42.4	0.76
0.5	$Q_{0.5}$=10^2.586*((DRNAREA)^0.763))	88.0	0.062	62.4	0.031	41.9	0.034	44.4	0.76
0.2	$Q_{0.2}$=10^2.681*((DRNAREA)^0.765))	86.4	0.072	68.3	0.035	45.2	0.039	48.0	0.76

do not apply to them. The weighted discharges can be found in the USGS data release associated with this study (Ahearn, 2020) and were computed with the following equation:

$$\log 10Qw = \frac{\log 10Qs(Vpred) + \log 10Qr(g)(Vs)}{Vpred + Vs} \quad (9)$$

where

Q_w is the weighted flood discharge, in cubic feet per second;

Q_s is the flood discharge for the selected AEP computed from the streamgage record, in cubic feet per second;

$Q_{r(g)}$ is the flood discharge for the selected AEP from the regression equation at the streamgage, in cubic feet per second;

V_{pred} is variance of prediction of the regression equation result ($Q_{r(g)}$) in logarithmic units; and

V_s is the variance of estimate of the AEP discharge in logarithmic units computed from the streamgage record (Q_s) in PeakFQ LP3 output.

Summary

Flood-frequency estimates are needed to support water-resource management decisions for flood protection, infrastructure design, and riparian and aquatic habitat protection. The flood-frequency estimates for streamgages and regional regression equations are periodically updated to account for new peak-flow information, improved explanatory datasets, and improved statistical techniques. This report presents the results of a cooperative study by the U.S. Geological Survey (USGS) and the Connecticut Department of Transportation to estimate the magnitude and frequency of peak flows at streamgages and develop regression equations for estimating peak flows at ungaged sites.

Flood-frequency estimates were calculated for 152 streamgages by fitting the record of annual peak flows to a log-Pearson type III distribution in the USGS PeakFQ program using the expected moments algorithm. The algorithm allows the use of perception thresholds and flow intervals to provide more information on systematic and historical peaks as well as periods of missing data than was possible in previous studies. Additionally, the generalized multiple Grubbs-Beck test was used to detect and screen out potentially influential low outliers in the peak-flow frequency distribution. Flood-frequency estimates for streamgages with unregulated flows were computed by weighting the skew estimated from the streamgage annual peak-flow record, called at-site skew, with a regional skew of 0.37.

A trend analysis of the annual peak flows for 79 long-term streamgages in Connecticut and in surrounding States extending 50 miles from the Connecticut border was performed. Trend results show some evidence of increasing peak flows over time and no evidence of decreasing trends. The number of statistically significant ($p \leq 0.05$) trends detected varies based on the assumptions of the serial correlation structure of the annual peak flows (independence, short-term persistence, or long-term persistence) and the time periods analyzed (30, 50, 70, or 90 years). The percentage of streamgages showing statistically significant trends in the four time periods analyzed ranged as follows: none to 11.1 percent, assuming long-term persistence; 1.6 to 33.3 percent, assuming short-term persistence; and 6.2 to 47.7 percent, assuming independence. Fewer increasing trends were detected with short-term and long-term persistence than with independence persistence indicating that multidecadal climate cycles may be present and influencing the magnitude of the trends. Clear evidence that the trends in peak flows in or near Connecticut are a normal oscillation or a true trend will require long-term monitoring and further analysis. No adjustments for trends were applied to the annual peak flows in the flood-frequency analysis.

Regional regression equations were developed using generalized least squares to estimate flood discharges with 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities (AEPs; 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence intervals, respectively). The three basin characteristics (drainage area in square miles; maximum 24-hour, x-year precipitation in inches; and percentage of area classified as hydrologic soil group C, C/D, or D in decimal percent) were found to be the best predictors of flood discharges. The final regional regression equations were selected based on the lowest model error; amount of variability explained by explanatory variables; random patterns in residuals plots; low number of data points with high leverage or influence; and consistency in model form across all estimated statistics. The standard error of prediction and standard error of estimate values ranged from 26.53 to 44.97 and 25.21 to 41.12, for the 50- to 0.2-percent AEPs, respectively. The pseudo coefficient of determination indicates that the explanatory variables explain 95.5 to 88.9 percent of the variance in the flood magnitude for 50- to 0.2-percent AEPs. The standard error of prediction values were slightly improved in the current study compared to previous values in 2004.

Final regression equations and their 90-percent prediction intervals will be made available in the USGS StreamStats program, a web-based tool that allows a user to select a point of interest on a stream, delineate a drainage basin, estimate basin and climate characteristics, and estimate peak-flow statistics. The StreamStats program will produce estimates of streamflow statistics, such as AEP flood discharges, with quantifiable certainty only when used at locations with basin and climatic characteristics in the range of input variables in the regional regression equations.

Acknowledgments

The authors are most grateful for help from several USGS colleagues. Caroline Mazo assembled GIS data and compiled basin characteristics. Pam Lombard and Amy McHugh graciously provided thorough and timely reviews of the report.

References Cited

Ahearn, E.A., 2003, Peak-flow frequency estimates for U.S. Geological Survey streamflow-gaging stations in Connecticut: U.S. Geological Survey Water-Resources Investigations Report 03–4196, 14 p.

Ahearn, E.A., 2004, Regression equations for estimating flood flows for the 2-, 10-, 25-, 50-, 100-, and 500-year recurrence intervals in Connecticut: U.S. Geological Survey Scientific Investigations Report 2004–5160, 68 p., https://doi.org/10.3133/sir20045160.

Ahearn, E.A., 2020, Flood frequency and source data used in regional regression analysis of annual peak flows in Connecticut: U.S. Geological Survey data release, https://doi.org/10.5066/P9S4F751.

Ahearn, E.A., and Veilleux, A.G., 2020, Worksheet for computing annual exceedance probability flood discharges and prediction intervals at stream sites in Connecticut: U.S. Geological Survey data release, https://doi.org/10.5066/P9EWHAYW.

Bigwood, B.L., and Thomas, M.P., 1955, A flood-flow formula for Connecticut: U.S. Geological Survey Circular 365, 15 p.

Bogart, D.B., 1960, Floods of August–October 1955, New England to North Carolina: U.S. Geological Survey Water-Supply Paper 1420, 854 p. [Also available at https://doi.org/10.3133/wsp1420.]

Cohn, T.A., England, J.F., Berenbrock, C.E., Mason, R.R., Stedinger, J.R., and Lamontagne, J.R., 2013, A generalized Grubbs-Beck test statistics for detecting multiple potentially-influential low outliers in flood series: Water Resources Research, v. 49, no. 8, p. 5047–5058. [Also available at https://doi.org/10.1002/wrcr.20392.]

Cohn, T.A., Lane, W.M., and Baier, W.G., 1997, An algorithm for computing moments-based flood quantile estimates when historical flood information is available: Water Resources Research, v. 33, no. 9, p. 2089–2096. [Also available at https://doi.org/10.1029/97WR01640.]

Cohn, T.A., and Lins, H.F., 2005, Nature's style–Naturally trendy: Geophysical Research Letters, v. 32, no. 23, 5 p., accessed August 30, 2017, at https://doi.org/10.1029/2005GL024476.

Collins, M.J., 2009, Evidence for changing flood risk in New England since the late 20th century: Journal of the American Water Resources Association, v. 45, no. 2, p. 279–290. [Also available at https://doi.org/10.1111/j.1752-1688.2008.00277.x.]

Cook, R.D., 1977, Detection of influential observations in linear regression: Technometrics, v. 19, no. 1, p. 15–18.

Curran, J.H., Barth, N.A., Veilleux, A.G., and Ourso, R.T., 2016, Estimating flood magnitude and frequency at gaged and ungaged sites on streams in Alaska and conterminous basins in Canada, based on data through water year 2012: U.S. Geological Survey Scientific Investigations Report 2016–5024, 47 p., accessed August 20, 2016, at https://doi.org/10.3133/sir20165024.

Eng, K., Chen, Y.-Y., and Kiang, J.E., 2009, User's guide to the weighted-multiple-linear-regression program (WREG version 1.0): U.S. Geological Survey Techniques and Methods, book 4, chap. A8, 21 p. [Also available at https://doi.org/10.3133/tm4A8.]

England, J.F., Jr., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas, W.O., Jr., Veilleux, A.G., Kiang, J.E., and Mason, R.R., Jr., 2018, Guidelines for determining flood flow frequency—Bulletin 17C: U.S. Geological Survey Techniques and Methods, book 4, chap. B5, 148 p. [Also available at https://doi.org/10.3133/tm4B5.]

Flynn, K.M., Kirby, W.H., and Hummel, P.R., 2006, User's manual for PeakFQ, annual flood-frequency analysis using bulletin 17B guidelines: U.S. Geological Survey Techniques and Methods, book 4, chap. B4, 42 p. [Also available at https://doi.org/10.3133/tm4B4.]

Gotvald, A.J., Barth, N.A., Veilleux, A.G., and Parrett, C., 2012, Methods for determining magnitude and frequency of floods in California, based on data through water year 2006: U.S. Geological Survey Scientific Investigations Report 2012–5113, 38 p., 1 pl., accessed July 24, 2012, at https://pubs.usgs.gov/sir/2012/5113/.

Griffis, V.W., and Stedinger, J.R., 2007, The use of GLS regression in regional hydrologic analyses: Journal of Hydrology (Amsterdam), v. 344, no. 1–2, p. 82–95. [Also available at https://doi.org/10.1016/j.jhydrol.2007.06.023.]

Griffis, V.W., and Stedinger, J.R., 2009, Log-Pearson type 3 distribution and its application in flood-frequency analysis. III—Sample skew and weighted skew estimators: Journal of Hydrologic Engineering, v. 14, no. 2, p. 121–130. [Also available at https://doi.org/10.1061/(ASCE)1084-0699(2009)14:2(121).]

Gruber, A.M., Reis, D.S., and Stedinger, J.R., 2007, Models of regional skew based on Bayesian GLS regression, *in* Kabbes, K.C., ed., Proceedings of the World Environmental and Water Resources Congress 2007, May 15–19, 2007: Tampa, Fla., American Society of Civil Engineers Environmental and Water Resources Institute. [Also available at https://doi.org/10.1061/40927(243)400.]

Gruber, A.M., and Stedinger, J.R., 2008, Models of LP3 regional skew, data selection, and Bayesian GLS regression, *in* Babcock, R.W., Jr., and Walton, R., eds., Proceedings of the World Environmental and Water Resources Congress 2008—Ahupua'a, May 12–16, 2008: Honolulu, Hawaii, American Society of Civil Engineers Environmental and Water Resources Institute. [Also available at https://doi.org/10.1061/40976(316)563.]

Hamed, K.H., 2008, Trend detection in hydrologic data—The Mann-Kendall trend test under the scaling hypothesis: Journal of Hydrology (Amsterdam), v. 349, no. 3–4, p. 350–363. [Also available at https://doi.org/10.1016/j.jhydrol.2007.11.009.]

Hamed, K.H., and Ramachandra Rao, A., 1998, A modified Mann-Kendall trend test for autocorrelated data: Journal of Hydrology (Amsterdam), v. 204, no. 1–4, p. 182–196. [Also available at https://doi.org/10.1016/S0022-1694(97)00125-X.]

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, 522 p.

Hodgkins, G.A., and Dudley, R.W., 2005, Changes in the magnitude of annual and monthly streamflows in New England, 1902–2002: U.S. Geological Survey Scientific Investigations Report 2005–5135, 37 p. [Also available at https://doi.org/10.3133/sir20055135.]

Hodgkins, G.A., and Dudley, R.W., 2011, Historical summer base flow and stormflow trends for New England rivers: Water Resources Research, v. 47, no. 7 [W07528], 16 p.

Hodgkins, G.A., Whitfield, P.H., Burn, D.H., Hannaford, J., Renard, B., Stahl, K., Fleig, A.K., Madsen, H., Mediero, L., Korhonen, J., Murphy, C., and Wilson, D., 2017, Climate-driven variability in the occurrence of major floods across North America and Europe: Journal of Hydrology (Amsterdam), v. 552, p. 704–717. [Also available at https://doi.org/10.1016/j.jhydrol.2017.07.027.]

Huntington, T.G., Richardson, A.D., McGuire, K.J., and Hayhoe, K., 2009, Climate and hydrological changes in the northeastern United States—Recent trends and implications for forested and aquatic ecosystems: Canadian Journal of Forest Research, v. 39, no. 2, p. 199–212. [Also available at https://doi.org/10.1139/X08-116.]

Interagency Advisory Committee on Water Data, 1981, Guidelines for determining flood-flow frequency: U.S. Geological Survey Bulletin 17B, 183 p. [Also available at https://water.usgs.gov/osw/bulletin17b/dl_flow.pdf.]

Khaliq, M.N., Ouarda, T.B.M.J., and Gachon, P., 2009, Identification of temporal trends in annual and seasonal low flows occurring in Canadian rivers—The effect of short- and long-term persistence: Journal of Hydrology (Amsterdam), v. 369, no. 1–2, p. 183–197. [Also available at https://doi.org/10.1016/j.jhydrol.2009.02.045.]

Koutsoyiannis, D., and Montanari, A., 2007, Statistical analysis of hydroclimatic time series—Uncertainty and insights: Water Resources Research, v. 43, no. 5. [Also available at https://doi.org/10.1029/2006WR005592.]

Kumar, S., Merwade, V., Kam, J., and Thurner, K., 2009, Streamflow trends in Indiana—Effects of long term persistence, precipitation and subsurface drains: Journal of Hydrology (Amsterdam), v. 374, no. 1–2, p. 171–183. [Also available at https://doi.org/10.1016/j.jhydrol.2009.06.012.]

L.R. Johnston Associates, 1983, Realizing the risk—A History of the June 1982 floods in Connecticut: Connecticut Department of Environmental Protection Water Planning Report 7, 142 p.

Lins, H.F., and Slack, J.R., 2005, Seasonal and regional characteristics of U.S. streamflow trends in the United States from 1940 to 1999: Physical Geography, v. 26, no. 6, p. 489–501. [Also available at https://doi.org/10.2747/0272-3646.26.6.489.]

Lorenz, D.L., and others, 2011, USGS library for S-PLUS for Windows—Release 4.0: U.S. Geological Survey Open-File Report 2011–1130, accessed Febuary 2018 at https://water.usgs.gov/software/S-PLUS/.

Ludwig, A.H., and Tasker, G.D., 1993, Regionalization of low-flow characteristics of Arkansas streams: U.S. Geological Survey Water-Resources Investigations Report 93–4013, 32 p. [Also available at https://doi.org/10.3133/wri934013.]

McCabe, G.J., and Wolock, D.M., 2002, A step increase in streamflow in the conterminous United States: Geophysical Research Letters, v. 29, no. 24, p. 2185–2189. [Also available at https://doi.org/10.1029/2002GL015999.]

Miller, D.R., Warner, G.S., Ogden, F.L., and DeGaetano, A.T., 2002, Precipitation in Connecticut: Special Reports 36, 65 p., accessed July 2018 at https://opencommons.uconn.edu/ctiwr_specreports/36.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., and Stouffer, R.J., 2008, Stationarity is dead—Whither water management?: Science, v. 319, no. 5863, p. 573–574. [Also available at https://doi.org/10.1126/science.1151915.]

National Oceanic and Atmospheric Administration [NOAA], 2015, NOAA Atlas 14 time series data: National Oceanic and Atmospheric Administration National Weather Service Precipitation Frequency Data Server, accessed June 2017 at https://hdsc.nws.noaa.gov/hdsc/pfds/pfds_series.html.

National Oceanic and Atmospheric Administration [NOAA], 2019, Historical hurricane tracks: National Oceanic and Atmospheric Administration Historical Hurricane Tracks mapping interface, accessed May 4, 2019, at https://coast.noaa.gov/hurricanes.

Omernik, J.M., 1995, Ecoregions—A spatial framework for environmental management, *in* Davis, W.S., and Simon, T.P., eds., Biological assessment and criteria—Tools for water resource planning and decision making: Boca Raton, Fla., Lewis Publishers, p. 49–62.

Paretti, N.V., Kennedy, J.R., and Cohn, T.A., 2014a, Evaluation of the expected moments algorithm and a multiple low-outlier test for flood frequency analysis at streamgaging stations in Arizona: U.S. Geological Survey Scientific Investigations Report 2014–5026, 61 p., accessed October 30, 2017, at https://pubs.er.usgs.gov/publication/sir20145026.

Paretti, N.V., Kennedy, J.R., Turney, L.A., and Veilleux, A.G., 2014b, Methods for estimating magnitude and frequency of floods in Arizona, developed with unregulated and rural peak-flow data through water year 2010: U.S. Geological Survey Scientific Investigations Report 2014–5211, 61 p., accessed October 30, 2017, at http://dx.doi.org/10.3133/sir20145211.

Paulson, C.C., Bigwood, B.L., Harrington, A.W., and others, 1940, Hurricane floods of September 1938: U.S. Geological Survey Water-Supply Paper 867, 562 p.

PRISM Climate Group, 2012, 30-year normals: Northwest Alliance for Computational Science and Engineering database, accessed June 2017 at http://www.prism.oregonstate.edu/normals/.

R Core Team, 2014, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, accessed August 3, 2016, at http://www.R-project.org.

Reis, D.S., Stedinger, J.R., and Martins, E.S., 2005, Bayesian generalized least squares regression with application to log Pearson type 3 regional skew estimation: Water Resources Research, v. 41, p. 1–14. [Also available at https://doi.org/10.1029/2004WR003445.]

Runkle, J., Kunkel, K., Champion, S., Easterling, D., Stewart, B., Frankson, R., and Sweet, W., 2017, Connecticut State climate summary: National Oceanic and Atmospheric Administration Technical Report NESDIS 149–CT, 4 p. [Also available at https://statesummaries.ncics.org/chapter/ct/#.]

Salas, J.D., Obeysekera, J., and Vogel, R.M., 2018, Techniques for assessing water infrastructure for nonstationary extreme events—A review: Hydrological Sciences Journal, v. 63, no. 3, p. 325–352. [Also available at https://doi.org/10.1080/02626667.2018.1426858.]

Small, D., Islam, S., and Vogel, R.M., 2006, Trends in precipitation and streamflow in the eastern U.S.—Paradox or perception?: Geophysical Research Letters, v. 33, no. 3 [L03403], 4 p.

Soil Survey Staff, Natural Resources Conservation Service, U.S. Department of Agriculture, 2017, Soil Survey Geographic (SSURGO) database: U.S. Department of Agriculture database, accessed August 14, 2017, at https://sdmdataaccess.sc.egov.usda.gov.

Soller, D.R., Packard, P.H., and Garrity, C.P., 2012, Database for USGS Map I-1970—Map showing the thickness and character of Quaternary sediments in the glaciated United States east of the Rocky Mountains: U.S. Geological Survey Data Series 656, accessed June 2017 at https://pubs.usgs.gov/ds/656/.

Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis—1. Ordinary, weighted, and generalized least squares compared: Water Resources Research, v. 21, no. 9, p. 1421–1432. [Also available at https://doi.org/10.1029/WR021i009p01421.]

Tasker, G.D., and Driver, N.E., 1988, Nationwide regression models for predicting urban runoff water quality at unmonitored sites: Journal of the American Water Resources Association, v. 24, no. 5, p. 1091–1101. [Also available at https://doi.org/10.1111/j.1752-1688.1988.tb03026.x.]

Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: Journal of Hydrology (Amsterdam), v. 111, no. 1–4, p. 361–375. [Also available at https://doi.org/10.1016/0022-1694(89)90268-0.]

Thomson, M.T., Cannon, W.B., Thomas, M.P., Hayes, S.S., and others, 1964, Historical floods in New England, chap. M *of* Contributions to the hydrology of the United States: U.S. Geological Survey Water-Supply Paper 1779, 105 p. [Also available at https://doi.org/10.3133/wsp1779M.]

TIBCO Software Inc, 2008, S–PLUS 6.1 modern statistics and advanced graphics: TIBCO Software Inc. software release.

U.S. Climate Data, 2019, Climate Connecticut: U.S. Climate Data database, accessed Febuary 2019 at https://www.usclimatedata.com/climate/hartford/connecticut/united-states/usct0094.

U.S. Environmental Protection Agency, 2010, U.S. level III and IV ecoregions: U.S. Environmental Protection Agency dataset, accessed February 5, 2018, at https://www.epa.gov/eco-research/ecoregions.

U.S. Fish and Wildlife Service, 2017, Wetlands Mapper: U.S. Fish and Wildlife Service web application, accessed August 2017 at https://www.fws.gov/wetlands/data/mapper.html.

U.S. Geological Survey, 1970, Quaternary sediments in the glaciated United States, Map I–1970, at https://pubs.usgs.gov/ds/656/.

U.S. Geological Survey, 2014a, NLCD 2011 land cover—National Geospatial Data Asset (NGDA) land use land cover (2011 ed., amended October 10, 2014): U.S. Geological Survey dataset, accessed February 2015 at http://www.mrlc.gov.

U.S. Geological Survey, 2014b, Weighted-multiple-linear REGression program, WREG version 2.02, at https://github.com/USGS-R/WREG.

U.S. Geological Survey, 2016a, Peak streamflow for the Nation, *in* USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed August 2, 2016, at https://doi.org/10.5066/F7P55KJN. [Peak streamflow information directly accessible at https://nwis.waterdata.usgs.gov/usa/nwis/peak.]

U.S. Geological Survey, 2016b, Site information for the Nation, *in* USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed August 2, 2016, at https://nwis.waterdata.usgs.gov/.

U.S. Geological Survey, 2017a, The National Map, National Elevation Dataset (NED) 10-meter resolution, NAVD 88: U.S. Geological Survey dataset, accessed June 2017 at https://viewer.nationalmap.gov/viewer/.

U.S. Geological Survey, 2017b, The National Map, National Hydrography Dataset (NHD), scale 1:24,000: U.S. Geological Survey dataset, accessed June 2017 at https://viewer.nationalmap.gov/viewer/.

U.S. Water Resources Council, 1967, A uniform technique for determining flood flow frequencies, Bulletin No. 15: Washington, D.C., U.S. Water Resources Council, Subcommittee on Hydrology, 15 p.

U.S. Water Resources Council, 1976, Guidelines for determining flood flow frequency, Bulletin No. 17: Washington, D.C., U.S. Water Resources Council, Subcommittee on Hydrology, variously paged.

Veilleux, A.G., 2011, Bayesian GLS regression, leverage and influence for regionalization of hydrologic statistics: Ithaca, N.Y., Cornell University Ph.D. dissertation, 184 p.

Veilleux, A.G., Cohn, T.A., Flynn, K.M., Mason, R.R., Jr., and Hummel, P.R., 2014, Estimating magnitude and frequency of floods using the PeakFQ 7.0 program: U.S. Geological Survey Fact Sheet 2013–3108, 2 p., accessed July 12, 2016, at https://pubs.usgs.gov/fs/2013/3108/.

Veilleux, A.G., Zariello, P.J., Hodgkins, G.A., Ahearn, E.A., Olson, S.A., and Cohn, T.A., 2019, Methods for estimating regional coefficient of skewness for unregulated streams in New England, based on data through water year 2011: U.S. Geological Survey Scientific Investigations Report 2017–5037, 29 p., accessed September 13, 2019, at https://doi.org/10.3133/sir20175037.

Vogel, R.M., Yaindl, C., and Walter, M., 2011, Nonstationarity—Flood magnification and recurrence reduction factors in the United States: Journal of the American Water Resources Association, v. 47, no. 3, p. 464–474. [Also available at https://doi.org/10.1111/j.1752-1688.2011.00541.x.]

Walter, M., and Vogel, R.M., 2010, Increasing trends in peak flows in the northeastern United States and their impacts on design, *in* Joint Federal Interagency Conference, 2d, Las Vegas, Nevada, June 27–July 1, 2010, Proceedings: Advisory Committee on Water Information, 16 p. [Also available at https://acwi.gov/sos/pubs/2ndJFIC/Contents/2F_Walter_03_01_10.pdf.]

Wandle, S.W., 1983, Estimating peak discharges of small, rural streams in Massachusetts: U.S. Geological Survey Water-Supply Paper 2214, 26 p.

Weiss, L.A., 1975, Flood flow formulas for urbanized and nonurbanized areas of Connecticut, *in* Watershed Management Symposium, Logan, Utah, August 11–13, 1975, Proceedings: American Society of Civil Engineers, Irrigation and Drainage Division, p. 658–675.

Weiss, L.A., 1983, Evaluation and design of a streamflow data network for Connecticut: Connecticut Water Resources Bulletin 36, 30 p.

Zarriello, P.J., 2017, Magnitude of flood flows at selected annual exceedance probabilities for streams in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2016–5156, 54 p., accessed September 5, 2017, at https://doi.org/10.3133/sir20165156.

Glossary

100-year flood An annual peak flow having an average recurrence interval of 100 years, corresponding to an annual exceedance probability of 1 percent.

adjusted coefficient of determination A modified version of coefficient of determination that has been adjusted for the number of predictors in the model.

annual exceedance probability The expected annual probability of a flood, previously referred to in terms of return period of a flood. It is the probability, often expressed as a decimal fraction less than 1.0, that an annual peak-flow discharge will be exceeded in a 1-year period. The reciprocal of the exceedance probability is referred to as the recurrence interval or return period in years.

average variance of prediction The average spread or dispersion of the predicted value from the observed mean.

covariance matrix A matrix whose element in the *i*, *j* position is the covariance between the *i*th and *j*th elements of a random vector. Covariance is a measure of how much two random variables change together. Positive values indicate that variables tend to show similar behavior, whereas negative values indicate that the greater value of one variable corresponds to the smaller value of the other variable. In multiple-variable regression, covariance is expressed in the form of a matrix sized according to the number of variables in the regression.

expected moments algorithm Method for fitting a probability distribution to annual peak-flow data using a generalized method of moments, similar to the standard log-Pearson type III method described in Interagency Advisory Committee on Water Data (1981), except the expected moments algorithm can also use interval data, whereas log-Pearson type III is restricted to point data. Interval data provide additional information that cannot be represented by point data, such as the potential range of annual peak flows outside of the systematic and historical record and the uncertainties around recorded peak flows used in the analysis.

generalized least squares A regression method that accounts for differences in the variances and cross correlations of the errors associated with different recorded discharges. Differences in variances can result from differences in the length of record for each site, whereas cross correlations among concurrent annual peaks result in cross correlation between estimated flood discharges.

leverage A metric used to identify those observations that are far away from corresponding average predictor values and may or may not have a large effect on the outcome of an analysis.

Mallows' Cp An estimate of the standardized mean squared error of prediction; this is a compromise between maximizing the explained variance by including all relevant variables and minimizing the standard error by keeping the number of variables as small as possible.

mean squared error The average of the squares of the differences between the estimated values and the measured values. This metric represents how closely, on average, an estimated value matches a measured value.

multicollinearity A statistical phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, in which case the regression coefficients may change erratically in response to small changes in the model or the data.

ordinary least squares Linear regression method that is the standard approach to the least squares solution of fitting an independent variable to one or more dependent variables.

outlier A data point that departs from the trend of the rest of a dataset as described by a distribution or other mathematical relation.

Pearson type III A frequency distribution determined from the statistical moments of the annual peak-flow mean, standard deviation, and skew.

pseudo coefficient of determination A statistic generated by the generalized least squares regression, the pseudo coefficient of

determination (or pseudo-R^2) is similar to the adjusted coefficient of determination in that it is a measure of the predictive strength of the regression model except that it removes the time sampling error.

root mean squared error The square root of the sum of the squares of the differences between estimated and the measured values divided by the number of observations minus 1. This metric represents the magnitude of the differences between the estimated and measured values.

skew A statistical measure of the data symmetry or lack thereof used to compute the flood-frequency distribution. The skew generally is computed from the logarithms of annual peak flows at the streamgage. Because the skew is sensitive to outliers, it may be an unreliable estimate of the true skew, especially for small samples; the guidelines in Interagency Advisory Committee on Water Data (1981) recommend that the skew is weighted with a regional, or generalized, skew that is based on data from many long-term streamgages to produce at-site flood-frequency estimates.

standard error of estimate Also referred to as the root mean squared error of the residuals, it is the standard deviation of observed values about the regression line. It is computed by dividing the unexplained variation or the error sum of squares by its degrees of freedom. In this study, the standard error is based on one standard deviation.

standard error of prediction The square root of the average spread or dispersion of the predicted value from the observed mean.

systematic record A period or periods of continuous annual peak-flow record.

variance A measure of the spread or dispersion of a set of values around their mean calculated by the mean of the squares of the deviation of the value from the mean, which is equal to the square of the standard deviation.

variance inflation factor Expresses the ratio of the actual variance of the coefficient of the explanatory variable to its variance if it were independent of the explanatory variables. A variance inflation factor greater than 5 to 10 generally indicates multicollinearity, a serious problem in the regression models.

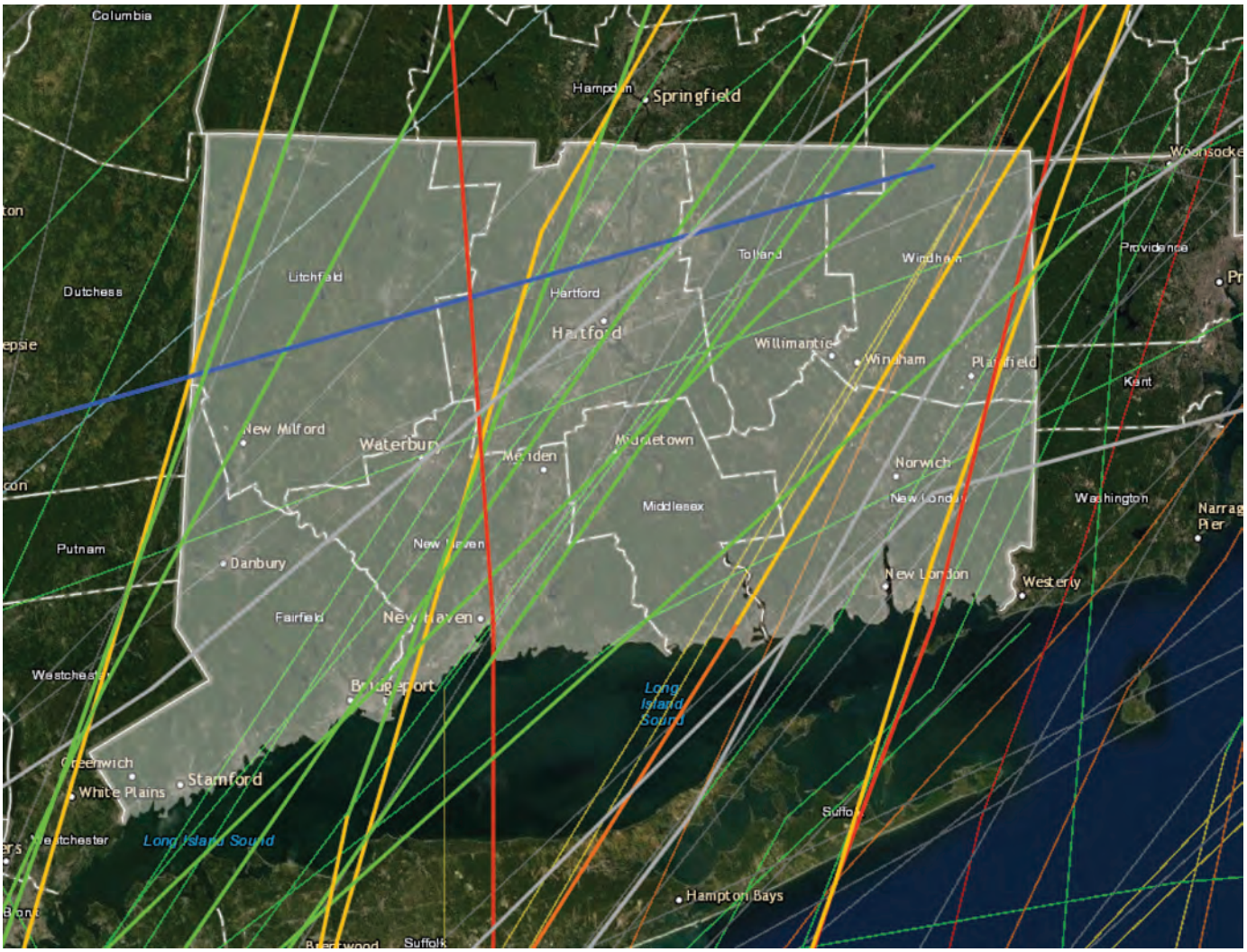
variance of prediction A measure of the likely difference between the prediction provided by a regression model and the actual value of the variable.

Appendix 1. Historical Hurricane Tracks

Figure 1.1 is a screenshot from the National Oceanic Atmospheric Administration (NOAA) summary search of hurricane and hurricane-related storm tracks from 1851 to 2016 in and near Connecticut (NOAA, 2019). NOAA's Historical Hurricane Tracks is a free online tool that allows users to track the paths of historical hurricanes. The site, developed by the NOAA Office for Coastal Management in partnership with NOAA's National Hurricane Center and National Centers for Environmental Information, offers data and information on coastal county hurricane strikes through 2016.

Reference Cited

National Oceanic and Atmospheric Administration [NOAA], 2019, Historical hurricane and hurricane-related storm tracks from 1851 to 2016 in and near Connecticut: National Oceanic and Atmospheric Administration Historical Hurricane Tracks mapping interface, accessed September 9, 2019, at https://oceanservice.noaa.gov/news/historical-hurricanes/.



EXPLANATION

Category

TS	Tropical storm	H1	Hurricane category 1
TD	Tropical depression	H2	Hurricane category 2
ET	Extratropical	H3	Hurricane category 3
N/A	Not applicable		

Figure 1.1. Screenshot showing hurricane and hurricane-related storm tracks from 1851 to 2016 in and near Connecticut (National Oceanic and Atmospheric Administration, 2019).

Appendix 2. Worksheet for Computing Annual Exceedance Probability Flood Discharges and Percent Prediction Intervals at Ungaged Sites

An application worksheet provided as a Microsoft Excel file is available for computing annual exceedance probability flood discharges and *x*-percent prediction intervals at stream sites (Ahearn and Veilleux, 2020; figure 2.1). The worksheet solves the peak-flow regression equations from user-specified explanatory values for the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities for an unregulated stream site in Connecticut. The worksheet also calculates the upper and lower discharge values for the 70-, 80-, 90-, 95- and 99-percent prediction intervals, the minus and plus standard error of prediction intervals, and the average standard error of prediction.

Reference Cited

Ahearn, E.A., and Veilleux, A.G., 2020, Worksheet for computing annual exceedance probability flood discharges and prediction intervals at stream sites in Connecticut: U.S. Geological Survey data release, https://doi.org/10.5066/P9EWHAYW.

Worksheet for computing annual exceedance probability flood discharges and prediction intervals for stream sites in Connecticut

This spreadsheet computes the regression estimate of the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities for an unregulated stream in Connecticut. The spreadsheet also computes the 70-, 80-, 90-, 95-, and 99-percent prediction intervals and the standard error of prediction of the regressions estimates. Note: The prediction interval provides upper and lower limits of the estimated flood discharge with a certain probability, or level of confidence in the accuracy of the estimate. The prediction interval is a range of discharges that is likely to contain the true discharge given specified settings of the predictors ranging from 70- to 99-percent. For example, for a 70 percent prediction interval, you can be 70% confident that the discharge for the site of interest from the regression is within the range of discharge defined by the lower and upper 70-percent prediction intervals. The standard error of prediction (SEP) indicates the overall prediction error for the regression estimate based on the data used to develop the regression equation.]

To use the spreadsheet, enter requested information in the yellow cells below.

Enter a site name:	Any river site in Connecticut
Enter a prediction interval, in percent (70, 80, 90, 95, or 99):	80

Table 1a: Enter explanatory variable data:

[DRNAREA, drainage area, in mi²; SOILCorD, area of hydrologic soil type C or D, in decimal percent; I24HRXYR, 24 hour rainfall that occurs in x year, in inches; mi², square miles]

Explanatory Variable		Value	Units	Applicable range
DRNAREA		110.00	mi²	0.69 to 325
SOILCorD		0.4207	decimal (percent of basin)	0.1176 to 0.9446
I24HRXYR	P2	3.10	inches	2.77 to 3.32
	P5	4.37	inches	4.0 to 4.70
	P10	5.34	inches	4.86 to 5.79
	P25	6.61	inches	5.99 to 7.22
	P50	7.57	inches	6.81 to 8.30
	P100	8.53	inches	7.62 to 9.38
	P200	10.01	inches	8.70 to 11.22
	P500	11.96	inches	10.10 to 13.64

Table 1b. Flood discharge estimates and prediction intervals for selected annual exceedance probabilities.

[SEP_{Pct}, standard error of prediction of the regression equation, reported as minus and positive units of percent and cubic feet per second; Average SEP_{Pct}, average standard error of prediction of the regression equation; ft^3/s, cubic feet per second. Note: The average SEP indicates the overall average prediction error for any estimated discharge based on the data used to develop the regression equation. See "Read.me about intervals" worksheet for documentation on converting the average SEPpct to a confidence level for the estimate, reported as -SEPpct and +SEPpct]

Annual exceedance probability (percent)	Estimated flood discharge (ft^3/s)	Lower percent prediction interval (ft^3/s)	Upper percent prediction interval (ft^3/s)	$-SEP_{Pct}$ (percent)	$+SEP_{Pct}$ (percent)	Average SEP_{Pct} (percent)	$-SEP_{Pct}$ (ft^3/s)	$+SEP_{Pct}$ (ft^3/s)
50	2,540	1,810	3,560	-23.0	29.8	26.5	1,960	3,300
20	4,540	3,270	6,300	-22.4	28.8	25.7	3,520	5,800
10	6,170	4,310	8,820	-24.2	31.9	28.2	4,680	8,140
4	8,650	5,910	12,700	-25.8	34.4	30.2	6,440	11,600
2	10,800	7,080	16,500	-28.0	38.9	33.8	7,770	15,000
1	13,300	8,400	21,100	-29.9	42.7	36.7	9,920	19,000
0.5	15,900	9,720	26,000	-31.7	46.4	39.5	10,900	23,300
0.2	20,000	11,500	34,600	-34.6	53.0	44.5	13,100	30,600

Table 1c. Regional regression equations for estimating flood discharges at stream sites in Connecticut.

[Q_p, flood discharge for x annual exceedance probability, in cubic feet per second; DRNAREA, drainage area, mi^2; I24HRxxYR, 24 hour rainfall that occurs in x years in, inches; SOILCorD, area of hydrologic soil group C or D, in decimal percent]

Annual exceedance probability (percent)	Recurrence interval (years)	Regression models
50	2	$Q_{50} = 10^{x-1.215} \times ((DRNAREA)^{x0.810}) \times 10^{x(0.5 * I24HRxxYR)} \times 10^{x(0.566 * (SOILCORD-1))}$
20	5	$Q_{20} = 10^{x-1.284} \times ((DRNAREA)^{x0.802}) \times 10^{x(0.541 * I24HRxxYR)} \times 10^{x(0.515 * (SOILCORD-1))}$
10	10	$Q_{10} = 10^{x-1.272} \times ((DRNAREA)^{x0.795}) \times 10^{x(0.529 * I24HRxxYR)} \times 10^{x(0.510 * (SOILCORD-1))}$
4	25	$Q_4 = 10^{x-1.237} \times ((DRNAREA)^{x0.788}) \times 10^{x(0.512 * I24HRxxYR)} \times 10^{x(0.542 * (SOILCORD-1))}$
2	50	$Q_2 = 10^{x-1.235} \times ((DRNAREA)^{x0.788}) \times 10^{x(0.527 * I24HRxxYR)} \times 10^{x(0.267 * (SOILCORD-1))}$
1	100	$Q_1 = 10^{x-1.234} \times ((DRNAREA)^{x0.789}) \times 10^{x(0.538 * I24HRxxYR)} \times 10^{x(0.4 * (SOILCORD-1))}$
0.5	200	$Q_{0.5} = 10^{x-1.272} \times ((DRNAREA)^{x0.786}) \times 10^{x(0.557 * I24HRxxYR)} \times 10^{x(0.260 * (SOILCORD-1))}$
0.2	500	$Q_{0.2} = 10^{x-1.340} \times ((DRNAREA)^{x0.784}) \times 10^{x(0.587 * I24HRxxYR)} \times 10^{x(0.378 * (SOILCORD-1))}$

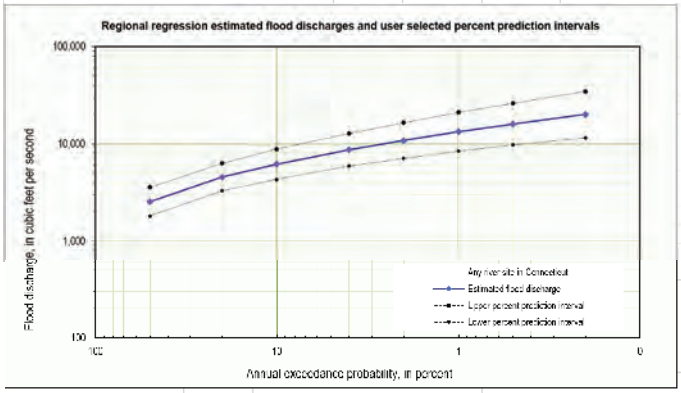


Figure 2.1. Screenshot showing worksheet for computing annual exceedance probability flood discharges and prediction intervals for ungaged stream sites in Connecticut.

For more information about this report, contact:
Director, New England Water Science Center
U.S. Geological Survey
10 Bearfoot Road
Northborough, MA 01532
dc_nweng@usgs.gov
or visit our website at
https://www.usgs.gov/centers/new-england-water

Publishing support provided by the
Pembroke and Madison Publishing Service Centers

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20205054