Prepared in cooperation with the Ohio Department of Transportation

Flood-Frequency Estimates for Ohio Streamgages Based on Data through Water Year 2015 and Techniques for Estimating Flood-Frequency Characteristics of Rural, Unregulated Ohio Streams

Scientific Investigations Report 2019–5018

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

Cover. January 1937 flooding from the Ohio River near Cincinnati, Ohio. (Rendering based on original photograph by the U.S. Army Corps of Engineers.)

Flood-Frequency Estimates for Ohio Streamgages Based on Data through Water Year 2015 and Techniques for Estimating Flood-Frequency Characteristics of Rural, Unregulated Ohio Streams

By G.F. Koltun

Prepared in cooperation with the Ohio Department of Transportation

Scientific Investigations Report 2019–5018

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: 2019

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

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

Suggested citation:
Koltun, G.F., 2019, Flood-frequency estimates for Ohio streamgages based on data through water year 2015 and techniques for estimating flood-frequency characteristics of rural, unregulated Ohio streams: U.S. Geological Survey Scientific Investigations Report 2019–5018, 25 p., https://doi.org/10.3133/sir20195018.

ISSN 2328-0328 (online)

Acknowledgments

The author would like to acknowledge Jeffrey Syar, Kyle Brandon, Matthew Cozzoli, and Kelly Nye, with the Ohio Department of Transportation, for their assistance and support for this study.

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Length	
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
	Area	
square mile (mi^2)	2.590	square kilometer (km^2)
	Volume	
acre-foot (acre-ft)	4,3559.9	cubic foot (ft^3)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
	Hydraulic gradient	
foot per mile (ft/mi)	0.1894	meter per kilometer (m/km)

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

°F = (1.8 × °C) + 32.

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

°C = (°F – 32) / 1.8.

Abbreviations

AEP	annual exceedance probability
DA	drainage area, in square miles (referred to as DRNAREA in StreamStats)
GIS	geographical information system
GLS	generalized least-squares
HUC	hydrologic unit code
LPIII	log-Pearson Type III
NLCD	National Land Cover Dataset
OLS	ordinary least-squares
PeakFQ	U.S. Geological Survey peak-flow frequency analysis program
PILF	potentially influential low flood
$SL_{10–85}$	channel slope, in feet per mile, determined by the 10–85 method (referred to as CSL1085LFP in StreamStats)
USGS	U.S. Geological Survey
W	percentage of the basin classified in the 1992 National Land Cover Dataset as open water and wetlands (referred to as LC92STOR in StreamStats)
WREG	weighted-multiple-linear regression

Flood-Frequency Estimates for Ohio Streamgages Based on Data through Water Year 2015 and Techniques for Estimating Flood-Frequency Characteristics of Rural, Unregulated Ohio Streams

By G.F. Koltun

Abstract

Estimates of the magnitudes of annual peak streamflows with annual exceedance probabilities of 0.5, 0.2, 0.1, 0.04, 0.02, 0.01, and 0.002 (equivalent to recurrence intervals of 2-, 5-, 10-, 25-, 50-, 100-, and 500-years, respectively) were computed for 391 streamgages in Ohio and adjacent states based on data collected through the 2015 water year. The flood-frequency estimates were computed following guidance outlined in Bulletin 17C, developed by the Advisory Committee on Water Information. The Bulletin 17C guidelines retain the basic statistical framework of the superseded Bulletin 17B guidelines; however, the Bulletin 17C guidelines add several enhancements including an improved method of moments approach for fitting the log-Pearson Type III (LPIII) distribution to the flood peaks (called the expected moments algorithm), a generalization of the Grubbs Beck low-outlier test called the Multiple Grubbs Beck test) that permits identification of multiple potentially influential low floods, and new methods for estimating regional skew and uncertainty.

Equations for estimating flood-frequency characteristics at ungaged sites on rural, unregulated streams in Ohio were developed with a two-step process involving ordinary least-squares and generalized least-squares regression techniques. Data from 333 streamgages with 10 or more years of unregulated record were screened for redundancy and a regression dataset was selected that was composed of flood-frequency and basin-characteristic data for 275 streamgages in Ohio and adjacent states. Two sets of equations were developed—one set, referred to as the "simple model," uses regression region and drainage area as regressor variables, and a second set, referred to as the "full model," uses regression region, drainage area, main-channel slope, and the percentage of the watershed covered by water and wetlands as regressor variables.

The average standard errors of prediction ranged from about 40.5 to 46.5 percent for the simple-model equations and from about 37.2 to 40.3 percent for the full-model equations. For sites meeting the rural, unregulated criteria, flood-frequency estimates determined by means of LPIII analyses are reported along with weighted flood-frequency estimates, computed as a function of the LPIII estimates and the regression estimates. For sites with homogenous periods of regulation, flood-frequency estimates determined by means of LPIII analyses are reported. Ninety-five percent confidence limits are reported for all estimates.

Values of regressor variables were determined from digital spatial datasets by means of a geographic information system (GIS). The GIS datasets and the new full-model equations have been incorporated into Ohio's StreamStats application, a web-based, GIS-backed system designed to facilitate the estimation of streamflow statistics at ungaged locations on streams.

Seasonal patterns in peak flows were assessed for 295 streamgages in Ohio. Annual peak flows occurred most frequently between January and April, with March having the highest frequency of occurrence. The month with the fewest number of annual peaks was October. Peak-of-record flows occurred most frequently in March, followed by January (months in which two of Ohio's most severe widespread floods in recent history occurred). None of the peak-of-record flows occurred in October and only two occurred in November.

Temporal trend in annual peak flows were assessed for 133 streamgages on unregulated streams in Ohio with 30 or more years of systematic record. Trends were assessed by computing the rank correlation (as measured with the two-sided Kendall's tau statistic) between time and annual peak flows. Weak but statistically significant trends were indicated at 15 of the 133 streamgages. Of the 15 streamgages with significant trend in annual peak flows, 12 had an upward trend (positive tau) and 3 had a downward trend (negative tau). All 12 streamgages with positive tau values were at latitudes north of 40°33', and streamgages with negative tau values were at latitudes south of 40°33'.

Introduction

Data on the magnitudes of flood-peak discharges with selected annual exceedance probabilities (AEPs) are commonly referred to as flood-frequency data. The use of the term “frequency” results from a common interpretation of the reciprocal of the AEP as an average frequency of recurrence during a long period of time. For example, a flood-peak discharge that has a 0.01 (1-percent) AEP is said to be equaled or exceeded, on average, once every 100 years (and consequently is called a “100-year flood”). Although many people consider the frequency concept easier to grasp than the probability concept, to understand that the frequency concept is based on a long-term average is important. Therefore, the occurrence of a “100-year flood” in a given year does not preclude the occurrence of a flood of equal or greater magnitude in the next 100 years or even in the next year). Because of misconceptions associated with reported N-year (where N is a number) recurrence intervals, recent literature has migrated toward reporting AEPs instead. Consequently, what were formerly reported as 2-, 5-, 10-, 25-, 50-, 100-, and 500-year recurrence intervals are now reported as 0.5, 0.2, 0.1, 0.04, 0.02, 0.01, and 0.002 AEPs, respectively.

Flood-frequency data have many uses. For example, the data are used in the design of bridges, culverts, dams, and spillways to ensure that those structures contain or convey design flow conditions without failure or unnecessary flooding. Flood-frequency data are also used in flood-insurance studies to determine the elevation and boundaries of the water surface associated with prescribed peak-flow conditions. Ultimately, decisions made on the basis of these data can have significant monetary impact on government agencies and private citizens and, at times, can make the difference between life and death.

The most recent report presenting flood-frequency data and estimation techniques applicable to Ohio (fig. 1) streams was published by Koltun and others (2006) and was based on streamflow data collected through the 2001 water year a water year is the period from October 1 through September 30 and is designated by the year in which it ends). Since water year 2001, new methods for estimating flood-frequency statistics and regional skew have been developed, more than 10 years of additional peak-flow data have been measured at some previously described locations, and sufficient data have become available to compute flood-frequency characteristics for other locations for which flood-frequency characteristics have not been previously determined. Given the new methodology, the availability of the additional peak-flow data and considering the economic and safety-related importance of flood-frequency information and estimation techniques, the U.S. Geological Survey (USGS), in cooperation with the Ohio Department of Transportation, have cooperated to develop new flood-frequency estimates and methods for estimating flood-peak discharges of rural, unregulated streams in Ohio.

Description of Study Area

The study area includes the State of Ohio and small portions of Indiana, Michigan, and Pennsylvania (fig. 1). Ohio has a total area of approximately 44,826 square miles (mi^2 with approximately 3,965 mi^2 of that area covered by perennial water (U.S. Census Bureau, 2018a). Based on a classification of land cover from Landsat satellite data circa 2011 (Homer and others, 2015), agriculture was the dominant land cover in Ohio comprising more than 46 percent of the State followed by forest (28.5 percent) and developed (13.5 percent) land covers. According to Sanders (2001), Ohio has more than 60,000 miles of streams.

About two-thirds of Ohio was glaciated during the last great Ice Age (Ohio Department of Natural Resources, 2019). Overall topographic relief across the State of Ohio is benign with relatively shallow elevation gradients (Tomlinson and others, 2013). Tomlinson and others (2013) concluded that orographic enhancement or depletion of rainfall is not a major factor in Ohio.

Ohio has a humid continental climate characterized by large seasonal temperature changes and generally ample precipitation derived from frontal and convective storms. Between 2001 and 2015, statewide average annual temperature and rainfall totals were 51.6 degrees Fahrenheit and 41.7 inches, respectively (National Oceanic and Atmospheric Administration, National Centers for Environmental Information, 2018). By comparison, the average annual temperatures and rainfall totals for Ohio for the 30-year period from 1985 to 2015 were 51.4 degrees Fahrenheit and 40.2 inches, respectively (National Oceanic and Atmospheric Administration, National Centers for Environmental Information, 2018). The reader is referred to Tomlinson and others (2013) for a relatively detailed discussion of the weather and climate of Ohio, particularly as the weather and climate relate to the production of extreme rainfall.

As of July 1, 2017, the U.S. Census Bureau (2018b) estimated that the resident population of Ohio was 11,658,609, making Ohio the seventh most populated State in the United States. According to the U.S. Census Bureau (2018c), (as of July 1, 2017) Columbus was the most populated city in Ohio, followed by Cleveland and then Cincinnati (fig. 1).

Purpose and Scope

The purpose of this report is to describe the results of a study to (1) estimate flood-frequency characteristics for selected streamgages in Ohio and adjacent states based on data collected through water year 2015; (2) develop and present techniques for estimating flood-frequency characteristics of rural, unregulated streams in Ohio; and (3) implement the estimation techniques in the Ohio StreamStats application. This report supersedes USGS Water-Resources Investigations Report 2006–5312 (Koltun and others, 2006) in that it provides revised flood-frequency estimates for streamgages and

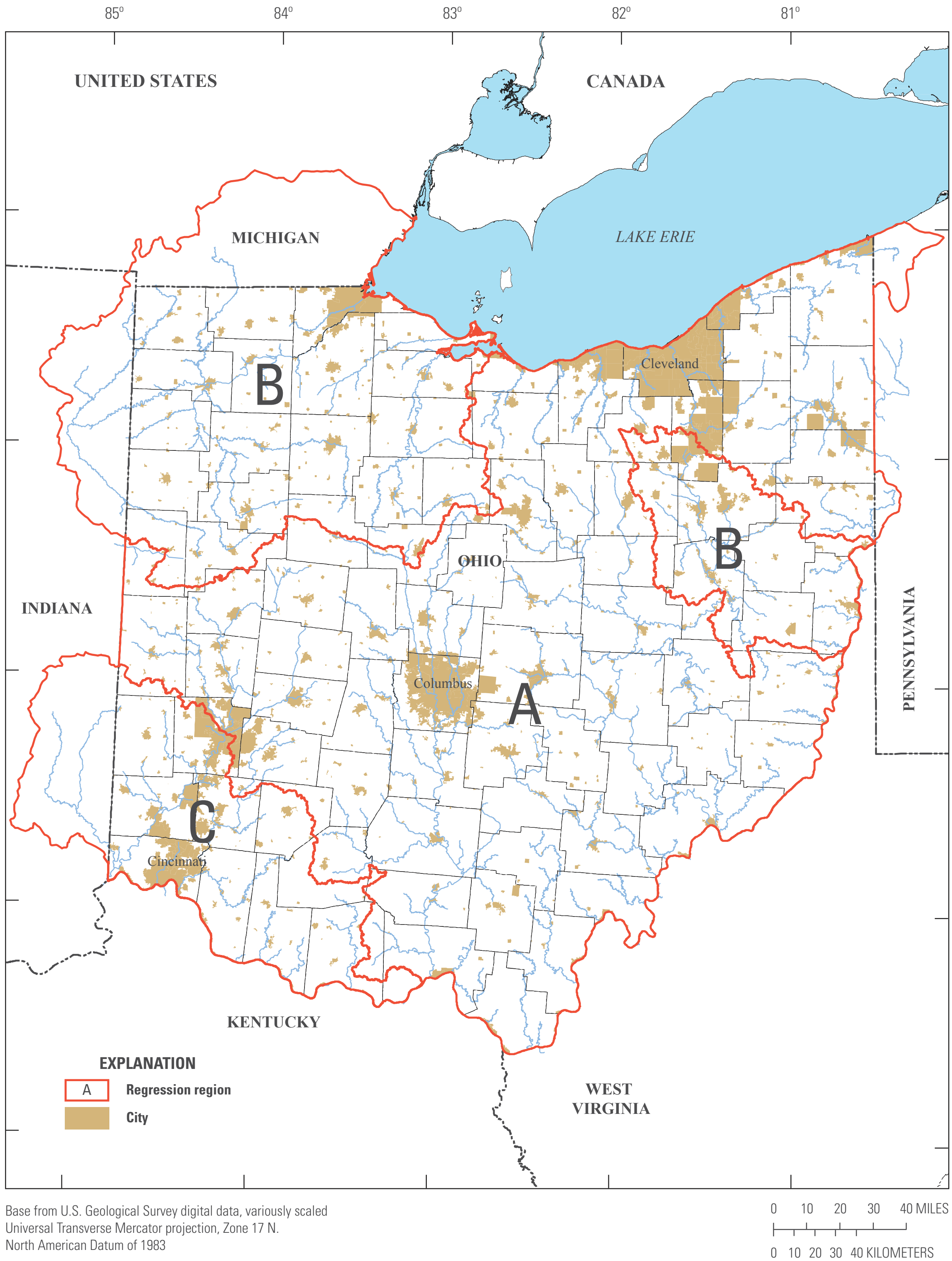

Figure 1. Study area showing regression regions.

presents new methods and equations for estimating flood-frequency characteristics of rural, unregulated streams in Ohio.

Previous Investigations

In several previous reports, flood-frequency data have been tabulated and methods have been presented for estimating flood-frequency characteristics of rural, unregulated streams in Ohio (Cross, 1946; Cross and Webber, 1959; Cross and Mayo, 1969; Webber and Bartlett, 1977; Koltun and Roberts, 1990; Koltun, 2003; and Koltun and others, 2006). The most recent of those reports (Koltun and others, 2006) marked the first use in Ohio of a web-based application called StreamStats to compute values of regressor variables and subsequent estimates of peak flows associated with selected AEPs.

Seasonal Patterns of Peak Flows

Annual peak flows can occur during any month on Ohio streams; however, the occurrence of annual peak flows is more common in some months than others. Based on a frequency analysis of more than 9,500 unregulated annual peak flows observed at 295 streamgages in Ohio through the 2015 water year (fig. 2), annual peak flows in Ohio occurred most often (59.7 percent of the observations) in the 4-month period between January and April, with March having the highest frequency of occurrence (17.3 percent of the observations). The month with the fewest annual peaks (1.1 percent) was October. Peak flows for which the month of occurrence was not listed were excluded.

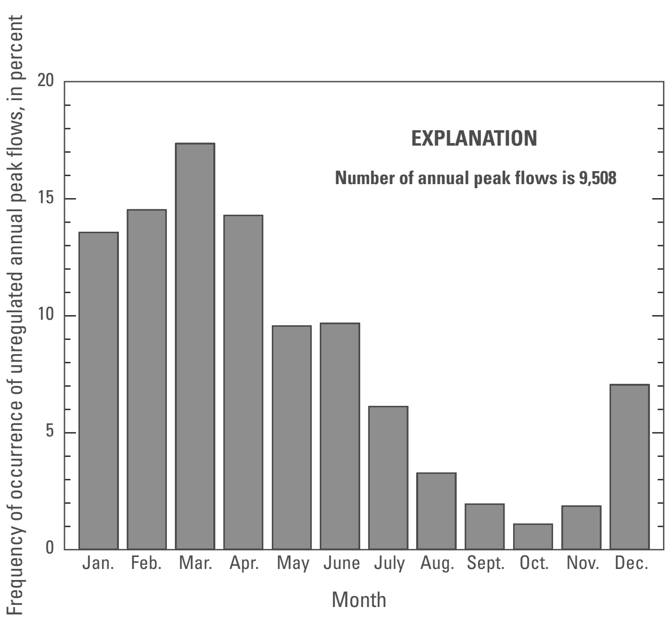

Figure 2. Frequency of occurrence of annual peak flows by month.

A frequency analysis of the largest recorded annual peak flow (peak-of-record flows) at each of the 295 streamgages indicated that the peak-of-record flows occurred most frequently in March (20.7 percent of the observations), followed by January (19.3 percent of the observations) (fig. 3). March and January coincide with months in which two of Ohio's most severe widespread floods in recent history (the March 1913 and January 1959 floods) occurred. July and June ranked third (13.6 percent of observations) and fourth (11.9 percent of observations) with respect to the frequency of peak-of-record flows. No peak-of-record flows occurred in October and only two peak-of-record flows (0.7 percent of the observations) occurred in November (fig. 3).

The monthly peak-of-record frequency characteristics separated into two classes as a function of drainage area is shown in figure 3, with one class representing sites with drainage areas less than 20 mi^2 and the other class representing sites with drainage areas greater than 20 mi^2. The smaller drainages peak-of-record flows occurred most frequently in June followed by July (fig. 3). That result likely reflects factors related to basin size and period of record. The peak flows of the smaller basins are likely to be influenced more than larger basins by summer-time convective storms that can produce high rainfall amounts with limited spatial extent. In addition, the smaller basins tend to be more recent additions to the gage network, with very few gages on smaller basins having record earlier than the 1940s (and so their records do not include the March 1913 flood).

Magnitude and Frequency of Floods at Gaged Sites

Flood-frequency analyses were completed using annual peak streamflow data measured at 391 streamgages. Although, this report is focused on Ohio, data from a total of 39 streamgages near the State borders with Michigan, Indiana, Kentucky, West Virginia, and Pennsylvania were used in this study (fig. 4).

Separate flood-frequency analyses were done for regulated and unregulated peak-flow records. For the purposes of this study, regulated records include any records where the peak streamflows are thought to be substantially altered from what would be expected for a "natural" stream. Regulation of peak flows can occur because of obvious forms of regulation such as operation of dams, but also can occur because of less obvious factors such as urbanization, channelization, and mining. Determination of whether peak flows were affected by regulation was based primarily on peak-flow qualifier codes in the USGS peak-flow database, but also on USGS streamgage descriptions and other supporting documentation contained in USGS files. Unless the streamgage description or other available information provided reasonable evidence otherwise, the codes in the peak-flow database were used to identify and classify peak-flow regulation. Streamgages with peak-flow

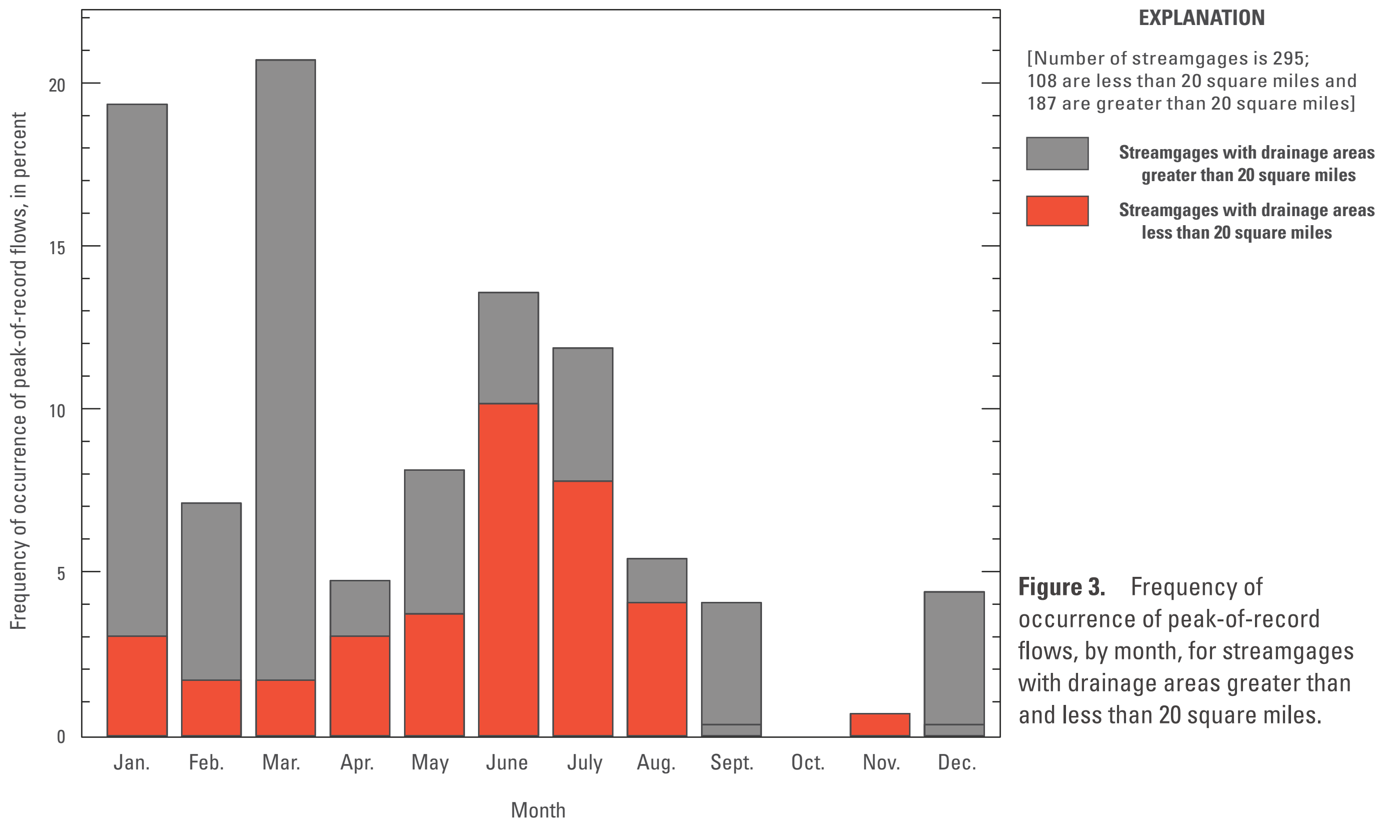

Figure 3. Frequency of occurrence of peak-of-record flows, by month, for streamgages with drainage areas greater than and less than 20 square miles.

codes of 5 (indicating discharge affected to unknown degree by regulation or diversion), 6 (indicating discharge affected by regulation or diversion), and C (indicating all or part of the record was affected by urbanization, mining, agricultural changes, channelization or other) were classified as regulated. Some peak flows were coded as estimated (code 2) or as maximum daily averages (code 1). These peak flows were included in the flood-frequency analyses without adjustment. Peak flows affected by dam failure (code 3) were not used in any of the flood-frequency analyses. The peakflow records are maintained in a database and are publicly available online at the USGS National Water Information System: Web Interface (U.S. Geological Survey, 2018).

Some streamgages had peak-flow records for water years prior to the onset of regulation. The records from those streamgages were divided into unregulated and regulated periods, with separate flood-frequency analyses completed on each period with 10 or more years of record. There were 333 streamgages with 10 or more years of unregulated record (fig. 4; table 1.1) and 82 streamgages with 10 or more years of regulated (or potentially regulated) record (fig. 5; table 1.2). Twenty-four of the streamgages had 10 or more years of regulated and unregulated record. Of the 82 streamgages, 5 did not have codes in the peak-flow database indicating regulation; however, land-cover characteristics indicate the potential that peak flows may have been affected by urbanization

Food-frequency analyses were completed using version 7.1 of the USGS peak-flow frequency analysis program (PeakFQ) (Veilleux and others, 2014) following the methodology described in Bulletin 17C (England and others, 2018). The Bulletin 17C guidelines retain the basic statistical framework of the superseded Bulletin 17B guidelines (Interagency Advisory Committee on Water Data, 1982); however, Bulletin 17C guidelines include the following enhancements:

- an improved method-of-moments approach for fitting the log-Pearson Type III (LPIII) distribution to the flood peaks (called the expected moments algorithm [Cohn and others, 1997]) that can accommodate interval estimates of peak flow, censored estimates of peak flow, and multiple thresholds of observation;
- a generalization of the Grubbs Beck low-outlier test (called the Multiple Grubbs Beck test [Cohn and others, 2013]) that permits identification of multiple potentially influential low floods (PILFs); and
- new methods for estimating regional skew and uncertainty (Veilleux and others, 2011).

Systematic data are distinguished from historical data in the flood-frequency analyses. Annual peak-flow data collected as part of the systematic operation of a streamgage are called “systematic data.” “Historical data” can take on several forms including (1) observations of large flows that occurred outside of the period of systematic record, (2) knowledge that one or more floods within the period of systematic record are the largest in a longer period, and (3) knowledge that flood magnitudes did not exceed a given value during a period outside of the period of systematic record. The period of systematic

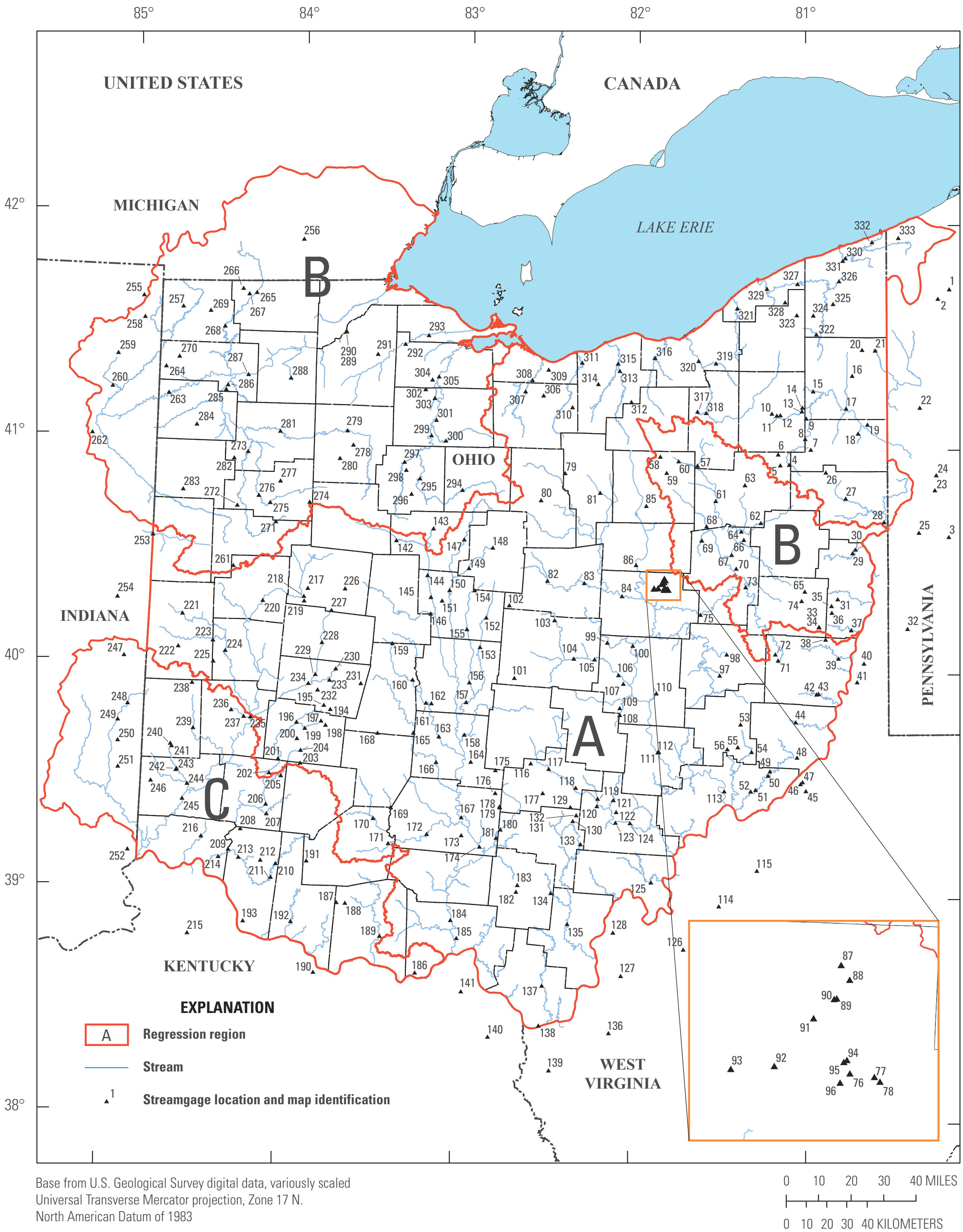

Figure 4. Locations of unregulated streamgages for which flood-frequency characteristics were determined in this study.

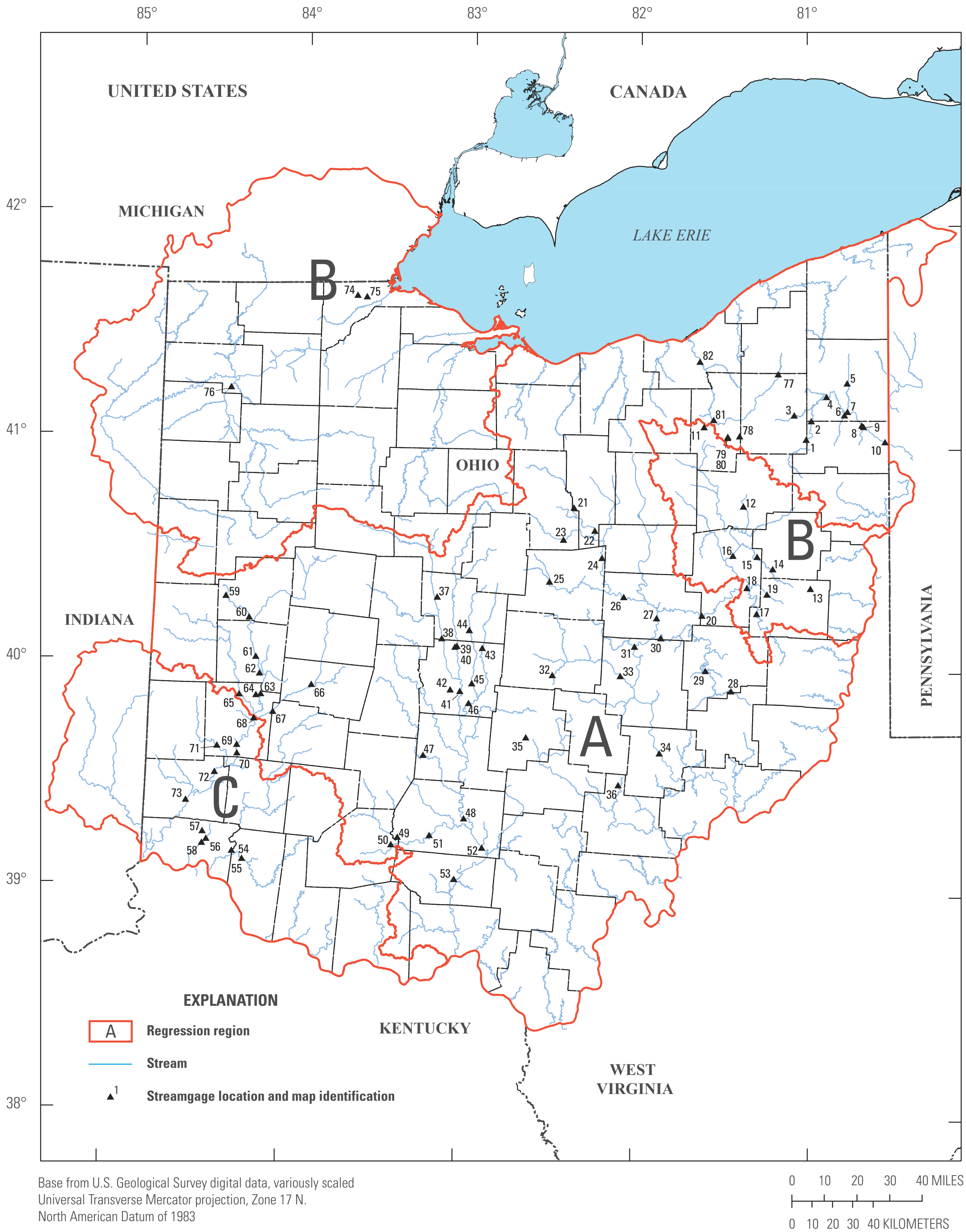

Figure 5. Locations of regulated streamgages for which flood-frequency characteristics were determined in this study.

record together with the intervening years between the systematic and historical peak flows define the “historical period” of the streamgage.

As was true with Bulletin 17B, the Bulletin 17C methodology prescribes the use of a LPIII distribution to fit the logarithms of annual peak-flow data as a function of the mean, standard deviation, and coefficient of skewness of the logarithms of annual peak flows using the following equation:

$$\hat{Q}_P = \bar{X} + K_P S \quad (1)$$

where,

$\hat{Q}_P$ is estimated P-percent AEP flow, in cubic foot per second (ft^3/s), where P equals the probability that the annual peak flow will be equaled or exceeded in any given year;

$\bar{X}$ is the mean of the logarithms of the annual peak flows;

K_P is a frequency factor dependent on the AEP and the coefficient of skewness of the log-transformed annual peak-flow series; and

S is the standard deviation of the log-transformed annual peak-flow series.

Regional Skew

Regional skew information should be considered and weighted appropriately with the station skew coefficient to provide an improved estimate of skewness (England and others, 2018). The station and regional skew coefficients are weighted inversely proportional to their individual mean square errors by the following equation:

$$G_W = \frac{MSE_{\bar{G}}(G) + MSE_G(\bar{G})}{MSE_{\bar{G}} + MSE_G} \quad (2)$$

where

G_W is the weighted skew coefficient,

$MSE_{\bar{G}}$ is the mean square error of the regional skew,

G is the station skew coefficient,

MSE_G is the mean square error of the station skew, and

$\bar{G}$ is the regional skew coefficient.

Under Bulletin 17B guidelines (Interagency Advisory Committee on Water Data, 1982), a generalized skew coefficient, determined from a map of skew coefficients prepared for the United States, could be used to compute the weighted skew coefficient. For this study, the regional skew coefficient that was used was developed for a region covering most of hydrologic unit code (HUC) region 04 and part of HUC region 05 (including parts of Illinois, Indiana, Kentucky, Michigan, Minnesota, New York, Ohio, Pennsylvania, Vermont, West Virginia, and Wisconsin) (U.S. Geological Survey, 2019). The regional skew was determined from a Bayesian-regression analysis of station skew coefficients and basin characteristics for 368 long-term streamgages in HUC regions 04 and 05, and the regional skew development is described in detail in Veilleux and Wagner (in press). None of the basin characteristics tested were statistically significant and explained an appreciable amount of the spatial variation in skewness; consequently, a constant skew coefficient of 0.086 was determined to be the best model for the region. The mean square error of the constant skew model is 0.13, which is less than one-half the 0.302 mean square error of the Bulletin 17B generalized skew map and is approximately equal to a generalized standard error of 0.3606.

For unregulated peak-flow records in this study, the regional skew coefficient of 0.086 and station skew coefficients from the streamgage records were used to compute weighted skews, which in turn were used to compute the flood-frequency estimates. Flood-frequency estimates for regulated peak-flow records were based only on station skews because of the possibility that the skewness of annual peaks at regulated streamgages could differ substantially from those at unregulated streamgages. Summaries of flood-frequency-analysis-related characteristics of the unregulated and regulated streamgage records are listed in table 1.3 and table 1.4, respectively. Flood-frequency estimates for unregulated and regulated records are reported in table 1 and table 2 (available for download at https://doi.org/10.3133/sir20195018), respectively. Flood-frequency estimates are reported for regulated records only in those cases where the LPIII curve was judged to provide a reasonably good fit to the observed peaks. The reader should be mindful that changes in regulation subsequent to the period of analysis could appreciably change the flood-frequency characteristics of regulated streams.

Flow Intervals and Perception Thresholds

The expected moments algorithm can account for various forms of uncertainty in peak-flow estimates. For every year in the historical period (including gaps in the systematic record), a flow interval (defined by an upper and lower bound for the peak flow) must be provided. For most peak flows during the systematic period, the default upper and lower bounds equal the observed peak flow; and, for most water years when no information has been recorded, the default upper and lower bounds are infinity and zero, respectively. If a peak flow has definable uncertainty, the range of plausible flows for that peak can be entered as the upper and lower bounds for the flow interval.

Along with flow intervals, the expected moments algorithm technique also requires that a perception threshold be entered for every year. Perception thresholds identify the range of flows that would have been measured or recorded if they had occurred. Generally, for the systematic record, the range of perception thresholds is from zero to infinity (because all annual peak flows are assumed to have been measured during the period of systematic data collection). At some streamgages for example, crest-stage gages) flows can be determined only when water in the stream reaches a certain minimum measurable level. In some years, the water may not reach that minimum level, consequently the lower perception threshold is the flow associated with the minimum measurable water level. Perception thresholds also are set for the ungaged period within the historical period. Generally, the lower threshold is set to the minimum flow that the analyst thinks would be measured or recorded if the minimum flow had occurred (even though the streamgage was not being operated), and the upper threshold is set to infinity.

Some peak flows may be classified as “opportunistic.” Opportunistic peak flows are peak flows measured outside of the systematic record based on factors other than exceedance of a perception threshold. Because the observation of opportunistic peak flows is not truly random, their sampling properties are unknown. Consequently, opportunistic peaks were not included in the flood-frequency analyses because of the potential that those peaks will bias the sample.

Nondefault flow intervals used in the analyses of unregulated streamgage records are reported in table 1.5. No nondefault flow intervals were used in the analyses of regulated streamgage records. Perception thresholds used in the analyses of unregulated and regulated streamgage records are reported in table 1.6 and table 1.7, respectively. Inputs to PeakFQ, which include all flow and perceptions intervals used in this study, are presented in Koltun (2019a). The report files from the PeakFQ analyses also are available in Koltun (2019a).

Tests for Potentially Influential Low Floods

In typical flood-frequency analyses, most of the interest is in large floods (floods with small AEPs), which are near the upper end of the peak-flow distribution. Sometimes peak-flow records contain low-magnitude peaks that depart significantly from the trend of the higher-flow values (low outliers). These PILFs can have a large influence on frequency estimates in the upper end of the peak-flow distribution. The Multiple Grubbs Beck test (Cohn and others, 2013) was used to identify PILFs that were then censored from the frequency distribution. Censoring the PILFs typically results in improved agreement between the high end (where AEPs are small) of the observed frequency distribution and the high end of the estimated frequency distribution. In some instances, censoring the PILFs may degrade the fit at the low end (where AEPs are large) of the frequency distribution. Low outlier thresholds used in the analyses of unregulated and regulated streamgage records are reported in table 1.3 and table 1.4.

Tests for Temporal Trends in Flood Magnitudes

Prior to completing the flood-frequency analyses, unregulated data from 333 streamgages in Ohio and adjacent states with 10 or more years of unregulated record were tested for temporal trends by computing Kendall’s tau, a nonparametric measure of correlation. These tests were done because a necessary assumption for the LPIII analysis is that the peak-flow data are a reliable and representative sample of random homogeneous events (Interagency Advisory Committee on Water Data, 1982). The trends were assessed by computing the two-sided Kendall’s tau statistic and associated *p*-values. Kendall’s tau is a rank correlation coefficient that, in this case, provides a measure of the correlation between the rank order of time and the rank order of peak flows. The *p*-values indicate the probability that tau is equal to zero. A Kendall’s tau of 1 indicates a perfect monotonically upward trend and a -1 indicates a perfect monotonically downward trend; a tau of 0 indicates no trend.

An alpha level of 0.05 (5 percent) was used for tests of significance in this study. Statistically significant (*p*-value less than or equal to 0.05) trends in annual peak flows were determined for 37 (about 11 percent) of the streamgages with data from approximately two-thirds of the 37 streamgages exhibiting positive trends. The median absolute tau value for the streamgages with significant trend was 0.32, indicating weak trend (Mukaka, 2012), and the 10 streamgages with the largest absolute tau values all had short records (averaging less than 13 years). The 37 streamgages (30 streamgages in Ohio and 7 in surrounding states) with significant trends were researched to determine if a cause for trends could be identified. The cause could not be determined with reasonable certainty, so the trends were assumed to be due to chance or to result from a short time sample that was not representative of a long-term trend. Although some of the identified trends possibly could reflect a changing climate, data from all 37 streamflow streamgages with statistically significant trend were retained in the analysis. The data were all retained because of uncertainty about causation and a lack of guidance on how to account for climate change on flood frequency. Kendall’s tau values and their associated *p*-values are reported in table 1.3 and table 1.4 for the unregulated and regulated streamgage records, respectively.

Long-Term Trends in Annual Peak Flows

Annual peak-flow data for 133 streamgages on unregulated streams in Ohio with 30 or more years of systematic record through the 2015 water year were analyzed for long-term trends by computing Kendall’s tau (as previously described). Thirty years of record was chosen as a minimum for this analysis in an attempt to minimize identification of spurious trends resulting from an unrepresentative sampling period. Fifteen of the 133 tau values (about 11 percent) had associated *p*-values less than or equal to 0.05, indicating statistically significant trends. The median number of systematic peaks for the 15 streamgages was 75 (average of about 68) and all except 4 of the streamgages had records extending through the 2015 water year. The tau values for those 15 sites ranged from -0.317 to 0.246, indicating the trends (although statistically significant) were weak. Of the 15 streamgages with significant trend, 12 had positive tau values (indicating an upward trend in annual peak flows) and 3 sites had negative tau values (indicating a downward trend). All 12 streamgages with positive tau values were at latitudes north of 40°33', and streamgages with negative tau values were at latitudes south of 40°33' (fig. 6).

Development of Regional Regression Equations

Regional regression equations for estimating flood-peak discharges at ungaged sites on rural, unregulated streams were developed with a two-step process involving ordinary least-squares (OLS) regression and generalized least-squares (GLS) regression techniques. GLS regression equations were determined because Stedinger and Tasker (1985) determined that, compared to OLS regression, GLS regression provides more accurate parameter (regression coefficient) estimates, better estimates of the accuracy with which the regression model's coefficients are estimated, and almost unbiased estimates of the model error.

Because of the long computer execution time associated with GLS regression analyses, OLS regression techniques were used in the first step to determine the "best" models relating basin characteristics to peak discharge estimates for each AEP. In the second step, the regressor variables chosen on the basis of OLS regression were used in GLS regression analyses to develop equations that will be used for predictive purposes. Prior to completing regression analyses, tests were done to identify streamgages that provided potentially redundant information. A model archive of the regression analyses is available in Koltun (2019b).

Tests for Redundancy

When two or more streamgages effectively provide nearly the same information to a regression model, the streamgages are said to provide redundant information. Instead of providing independent spatial observations that depict how drainage basin characteristics are related to flood magnitudes with a given AEP, these streamgages exhibit nearly the same hydrologic response to a given storm and, thus, effectively represent only one spatial observation. Consequently, when some streamgages provide redundant information, a statistical analysis using the redundant observations incorrectly represents the information content in the regional dataset (Gruber and Stedinger, 2008). To determine if two streamgages provided potentially redundant information, the following two types of information were considered: (1) the distance between basins centroids and (2) the ratios of the basin drainage areas.

A standardized distance was used in part to determine the likelihood that the streamgages provide redundant information. The standardized distance between two streamgages (*i* and *j*), SD_{ij}, is defined as follows:

$$SD_{ij} = \frac{D_{ij}}{\sqrt{0.5(DA_i + DA_j)}} \tag{3}$$

where

D_{ij} is the distance, in miles, between centroids of basin *i* and basin *j*;

DA_i is the drainage area in square miles at site *i*; and

DA_j is the drainage area in square miles at site *j*.

Along with the standardized distance, a drainage area ratio was used to determine if the drainages associated with streamgages were sufficiently similar in location and size to conclude that the streamgages may provide redundant information for the purposes of developing a regional hydrologic model.

The drainage area ratio, *DAR*, is defined as follows:

$$DAR \quad e^{(|\log(DA_i/DA_j)|)} \tag{4}$$

where

DA_i is the drainage area at site *i*, and

DA_j is the drainage area at site *j*.

Recent studies (Veilleux, 2009; Mastin and others, 2016) indicate that screening thresholds of standardized distance less than or equal to 0.50 combined with drainage area ratio less than or equal to 5 are appropriate for identifying sites with potentially redundant information; consequently, those thresholds were adopted for this study. All possible combinations of streamgage pairs from the 333 streamgages with 10 or more years of unregulated record were considered in the redundancy analysis. All streamgage pairs meeting the threshold criteria were investigated and, if deemed potentially redundant, one streamgage from the pair was removed from the regression dataset. Of the 333 streamgages with 10 or more years of unregulated record, 56 (about 17 percent) were removed from the regression dataset because of potential redundancy. Another 2 streamgages were not included in the regression dataset for other reasons, leaving a regression dataset composed of data for 275 streamgages. The number of annual peaks on which flood-frequency results for the 275 streamgages was based ranged from 10 to 103, with a median value of 26 annual peaks.

Ordinary Least-Square Regression

Koltun (2003) previously tested 20 basin characteristics as explanatory variables and determined 5 to be most useful for predicting flood-frequency characteristics in Ohio. Those five characteristics included drainage area (DA), main-channel slope determined by the 10–85 method (SL_{10-85}) (described in Benson, 1962), the percentage of the basin area classified in the 1992 National Land Cover Dataset (NLCD) (Vogelmann and others, 2001) as open water and wetlands (W), and binary regression region indicator variables (OHREGA and OHREGC). In the Ohio StreamStats application,

Figure 6. Locations of streamgages with 30 or more years of peak-flow record that had statistically significant correlations between the magnitudes of peak flow and time.

drainage area, main-channel slope determined by the 10–85 method, and the percentage of the basin area classified in the 1992 NLCD as open water are referred to as DRNAREA, CSL1085LFP, and LC92STOR, respectively. As described in Koltun and others (2006), the drainage into and within Ohio was divided into three regression regions (A, B, and C) as shown in figure 1.

The binary regression region indicator variables were assigned the combined values listed in the following table based on the regression region in which the site is located.

Region	OHREGA	OHREGC
A	1	0
B	0	0
C	0	1

Streamgages in border states that are on streams that do not drain into Ohio were assigned regression region codes based on the regression region in the closest proximity.

For streamgages in Ohio, the drainage areas and main-channel slopes were computed with the Ohio StreamStats application (Koltun and others, 2006). For streamgages outside Ohio, the drainage areas and main-channel slopes were computed, when possible, with the applicable state StreamStats applications (Ries and others, 2017); otherwise, the drainage areas and main-channel slopes were obtained from other USGS sources such as the National Water Information System: Web Interface, the GAGES–II dataset (Falcone, 2017), or publications.

The potential explanatory variables that were previously tested by Koltun and others (2006) and excluded from the regression equations were not retested for this study; however, the explanatory variables that were previously tested and used in the regression equations were retested. Consideration was given to updating *W* (the percentage of the basin classified in the 1992 National Land Cover Dataset as open water and wetlands) based on a more recent land cover dataset; however, because the bulk of the record at long-term gaging stations predates 1992, the decision was made that values of *W* computed based on the 1992 NLCD are still relevant (in that values representative of the early 1990s would be more central to the period of record at long-term gaging stations than a later measure of *W*).

The following new explanatory variables were added for testing: (1) *LC11IMP*, the percentage of impervious area determined from the 2011 NLCD impervious dataset (Homer and others, 2015) and (2) *LC11DEV*, the percentage of developed land covered by land cover classes 21–24 in the 2011 NLCD. All explanatory variables, except for the regression region indicator variables, were $\log_{10}$ transformed and 1.0 was added to the value of *LC11IMP*, *LC11DEV*, and *W* before log transformation. Applicable basin characteristics of all sites with 10 or more years of unregulated record are listed in table 1.1.

Selection of regressor variables for use in the final model was based on the following criteria:

- The choice of regressor variables, and the signs and magnitudes of their associated regression coefficients had to be hydrologically plausible in the context of peak flows. This criterion took precedence over all other criteria.
- All regressor variables had to be statistically significant at the 95-percent confidence level.
- The choice of regressor variables, with the constraints of the first two criteria, had to minimize the average variance of prediction and maximize the pseudo coefficient of determination (pseudo R^2, a measure of the proportion of the variation in the dependent variable accounted for by the regression equation after removing the effect of the time-sampling error [Griffis and Stedinger, 2007]).

The new explanatory variables (*LC11IMP* and *LC11DEV*) were not statistically significant. Evaluation of the significance of *LC11IMP* was based on data from only 261 of the 275 streamgages (including all of the Ohio streamgages) in the regression dataset, because that characteristic could be determined with relative ease for that subset. If significance tests for *LC11IMP* had indicated statistical significance (as indicated by a *p*-value less than or equal to 0.05) or near statistical significance (as indicated by a *p*-value less than or equal to 0.1), then that basin characteristic would have been determined for the remaining 14 streamgages and retested. Although neither *LC11IMP* nor *LC11DEV* were statistically significant, the values associated with each streamgage were reviewed to assess the potential for urban influence on peaks.

Because *LC11IMP* and *LC11DEV* reflect conditions circa 2011, they are not necessarily indicative of the level of urbanization associated with a streamgage's record. For example, a streamgage may have *LC11IMP* and *LC11DEV* values that indicate a high percentage of impervious area and development circa 2011; however, the peak-flow record for that streamgage may have ended in the 1980s. In that case, urbanization may not have been an issue during the period of record.

The records for five streamgages (USGS station numbers 03115973, 03226850, 03228000, 03262750, and 03263100) with peak flows that are not currently classified as regulated were treated as being regulated in this study because the streamgages had high *LC11IMP* (greater than 30 percent) and *LC11DEV* (greater than 80 percent) values, and no clear evidence indicated that the level of urbanization during the periods of record would not have been sufficient to affect at least some of the peak flows.

Ultimately, the regressor variables used in the equations reported by Koltun and others (2006) proved to be statistically significant in explaining the variation in the flood-frequency estimates determined for this study. Consequently, those same regressor variables were carried forward to the GLS regression analyses. The values of the regressor variables for unregulated streamgages discussed in this report are listed in table 1.1. Summary statistics of the regressor variables for the 275 streamgages in the regression dataset are listed in table 3.

Generalized Least-Squares Regression

The regressor variables selected in the OLS regression analyses were used to develop GLS regression equations for estimating flood-frequency characteristics. In addition, a second set of equations was developed for a model that used only the regression region codes and drainage area as regressor variables. To facilitate future discussion, the equations that use the full suite of regressor variables will be referred to as the "full model" equations, and the equations that use only the regression region code and drainage area variables will be referred to as the "simple model" equations.

The improvements afforded by the GLS regression technique (relative to the OLS regression technique) result from the fact that the GLS regression technique takes into consideration the variance and spatial correlation structure of the peak flows and weights each observation accordingly (Tasker and others, 1986). In addition, the time-sampling error in the estimated peak flow is accounted for in evaluating the accuracy of the regression equation. In contrast, OLS regression assumes equal reliability and variance and no cross-correlation between peak-flow records at the streamgages and, therefore, assigns equal weights to each of the peak-flow estimates.

Spatial correlation of annual peak flows for paired sites was estimated in the GLS analysis using the following correlation-smoothing function as originally described by Tasker and Stedinger (1989):

$$\hat{r}_{ij} = \theta^{\left[\frac{d_{ij}}{\alpha d_{ij}+1}\right]} \tag{5}$$

where

$\hat{r}_{ij}$ is the estimated linear cross-correlation of the time series of annual peak-flow values at sites i and j,

d_{ij} is the distance in miles between sites i and j, and

α and θ are dimensionless parameters.

The correlation-smoothing function is determined first by plotting distances between pairs of sites versus Pearson's r describing the linear correlation between the logarithms of annual peak flows at each site pair. The values that are plotted were determined by choosing data for site pairs with a minimum number of concurrent years of record (a minimum of 30 years of concurrent record was used for the evaluation in this study). The scatter plot was overlaid with a smooth line representing the correlation-smoothing function (eq. 5) and then the shape of that line was adjusted by changing two parameters, α and θ, until a "best fit" was obtained. The weighted-multiple-linear regression (WREG) program computes a Nash-Sutcliffe efficiency coefficient (Nash and Sutcliffe, 1970) for each parameter set to help evaluate the fit. An α of 0.002 and θ of 0.98 were determined to provide a good fit to the data and, therefore, were used for the smoothing function.

The GLS regression equations were determined by application of version 2.02 of the USGS program WREG (Farmer, 2018), which is an implementation written in the R statistical language (R Core Team, 2017) of the

Table 3. Summary statistics of regressor variables in the regression calibration dataset.

[*DA*, drainage area; mi^2, square mile; $SL_{10–85}$, channel slope determined by the 10–85 method; ft/mi, foot per mile; *W*, the percentage of the basin classified as water and wetlands in the 1992 National Landcover Dataset]

Region	Statistic	Regressor variable			Number of observations
		DA (mi^2)	$SL_{10–85}$ (ft/mi)	*W* (percent)	
A	Maximum	5,989	516	25.35	175
	Minimum	0.04	1.53	0.00	
	Mean	195	48.2	2.31	
	Median	39.2	13.2	0.79	
B	Maximum	6,309	457	7.10	68
	Minimum	0.04	1.21	0.00	
	Mean	242	25.2	1.44	
	Median	38.3	8.20	1.04	
C	Maximum	2,514	131	1.23	32
	Minimum	0.26	3.24	0.00	
	Mean	170	37.1	0.33	
	Median	19.2	18.6	0.25	

Weighted-Multiple-Linear-Regression Program (Eng and others, 2009). The equations and associated performance metrics for the simple and full models are reported in table 4 and table 5, respectively. The average standard errors of prediction, which provide a measure of the predictive ability of a model, ranged from about 40.5 to 46.5 percent for the simple equations (table 4) and from about 37.2 to 40.3 percent for the full-model equations (table 5).

Assessment of Fit

Visual assessments of fit of the regression models were done by examining scatter plots of observed and predicted values. The fit of the 0.5 AEP flood full-model equation (fig. 7) looked reasonable throughout the entire range of flows; however, the regression estimates have a pattern of being lower than the 0.01 AEP LPIII-based estimates of flow (fig. 8) for flows larger than about 50,000 ft^3/s. The set of seven streamgages associated with flows greater than 50,000 ft^3/s (USGS station numbers 03144500, 03234000, 03234500, 03237500, 03247500, 03270500, and 04193500) had drainage areas ranging from 387 to 6,309 mi^2, came from all three regression regions, and did not (with one exception) have other discernable unifying basin characteristics. What is remarkable about that set of streamgages is that four of the seven streamgages had drainage areas greater than 2,500 mi^2, and those four streamgages had the four largest drainage areas in the regression dataset. A review of full-model results for floods with other AEPs indicated that the regression equations underestimated flows at those same seven streamgages for floods with AEPs ranging from 0.2 to 0.002 with the percentage differences between the LPIII and regression estimates typically increasing with decreasing AEP. Scatterplots for the simple-model equations (not shown) were examined and indicated a bias similar to that observed with the full-model equations. These results suggest the potential that the regression equations underestimate floods with AEPs less than 0.5 for drainages more than 2,500 mi^2 and highlight the need to weight regression estimates with streamgage-based estimates whenever possible. The only currently unregulated river that has locations with drainage areas larger than 2,500 mi^2 is the Maumee River (in northwestern Ohio), and all points on that river with drainage areas larger than 2,500 mi^2 have a streamgage in close enough proximity to compute weighted flood-frequency estimates.

Residuals, the difference between observed and predicted values of the dependent variable, were examined to ensure that they did not meaningfully violate the assumptions of normality, independence, and constant variance. Examples of residual plots that were created and examined in that assessment are shown in figures 9–10. In addition, other tests were completed to look for conditions, such as moderate to strong collinearity in regressor variables that could negatively affect estimation of the regression parameters or their standard errors. No conditions were identified in the tests that would degrade or destabilize estimation or invalidate prediction intervals.

Table 4. Simple-model equations for estimating flood-frequency characteristics of rural, unregulated streams in Ohio.

[AEP, annual exceedance probability; pseudo R^2, pseudo coefficient of determination; SEP, standard error of prediction; Q, flood peak discharge, in cubic feet per second; *OHREGA* and *OHREGC* are binary regression region codes; *DA*, drainage area in square miles]

Recurrence interval (years)	AEP	Equation[1]	Constant a	Coefficient b	Coefficient c	Coefficient d	Pseudo R^2	Model error variance (σ^2_δ)	Average variance of prediction (VP_{avg})	Average SEP (percent)
2	0.5	$Q = [10^{(a+b[OHREGA]+c[OHREGC])}]DA^d$	1.894	0.159	0.397	0.705	95.39	0.031	0.031	42.40
5	0.2		2.124	0.182	0.410	0.672	94.43	0.028	0.029	40.49
10	0.1		2.243	0.197	0.418	0.655	94.83	0.030	0.030	41.84
25	0.04		2.368	0.214	0.429	0.639	94.26	0.031	0.032	43.10
50	0.02		2.448	0.226	0.438	0.629	93.69	0.033	0.034	44.61
100	0.01		2.518	0.237	0.446	0.620	93.37	0.034	0.035	45.20
500	0.002		2.657	0.261	0.466	0.604	92.61	0.035	0.037	46.51

[1]For manual computations, the terms in brackets in the equation above can be replaced by the following coefficients:

AEP	Regression region A	Regression region B	Regression region C
0.5	112.870	78.295	195.427
0.2	202.416	132.977	341.680
0.1	275.548	175.038	458.759
0.04	382.469	233.610	627.901
0.02	472.189	280.585	768.952
0.01	569.494	329.858	921.547
0.002	827.547	453.687	1,325.772

Table 5. Full-model equations for estimating flood-frequency characteristics of rural, unregulated streams in Ohio.

[AEP, annual exceedance probability; pseudo R^2, pseudo coefficient of determination; SEP, standard error of prediction; Q, flood peak discharge, in cubic feet per second; *OHREGA* and *OHREGC* are binary regression region codes; *DA*, drainage area in square miles; $SL_{10\text{–}85}$, main-channel slope determined by the 10–85 method, in foot per mile; *W*, percentage of the basin classified in the 1992 National Land Cover Dataset as open water and wetlands]

Recurrence interval (years)	AEP	Equation[1]	Constant	Coefficient					Pseudo R^2	Model error variance (σ^2_δ)	Average variance of prediction (VP_{avg})	Average SEP (percent)
			a	b	c	d	e	f				
2	0.5		1.668	0.126	0.326	0.781	0.162	−0.130	95.87	0.027	0.028	40.15
5	0.2		1.845	0.144	0.320	0.767	0.202	−0.174	96.09	0.024	0.024	37.16
10	0.1		1.939	0.157	0.319	0.760	0.222	−0.197	95.81	0.024	0.025	37.62
25	0.04	$Q = [10^{(a+b[OHREGA]+c[OHREGC])}]DA^{d}SL_{10\text{–}85}^{e}(W+1)^{f}$	2.040	0.172	0.321	0.752	0.240	−0.222	95.51	0.024	0.026	38.08
50	0.02		2.106	0.183	0.324	0.748	0.251	−0.238	95.20	0.025	0.027	37.79
100	0.01		2.165	0.194	0.328	0.743	0.259	−0.251	94.90	0.026	0.028	39.60
500	0.002		2.285	0.219	0.340	0.734	0.272	−0.277	94.46	0.027	0.028	40.27

[1]For manual computations, the terms in brackets in the equation above can be replaced by the following coefficients:

AEP	Regression region A	Regression region B	Regression region C
0.5	62.267	46.553	98.542
0.2	97.551	69.945	146.193
0.1	124.558	86.822	181.115
0.04	162.920	109.576	229.548
0.02	194.552	127.520	269.046
0.01	228.645	146.165	311.296
0.002	318.795	192.726	422.053

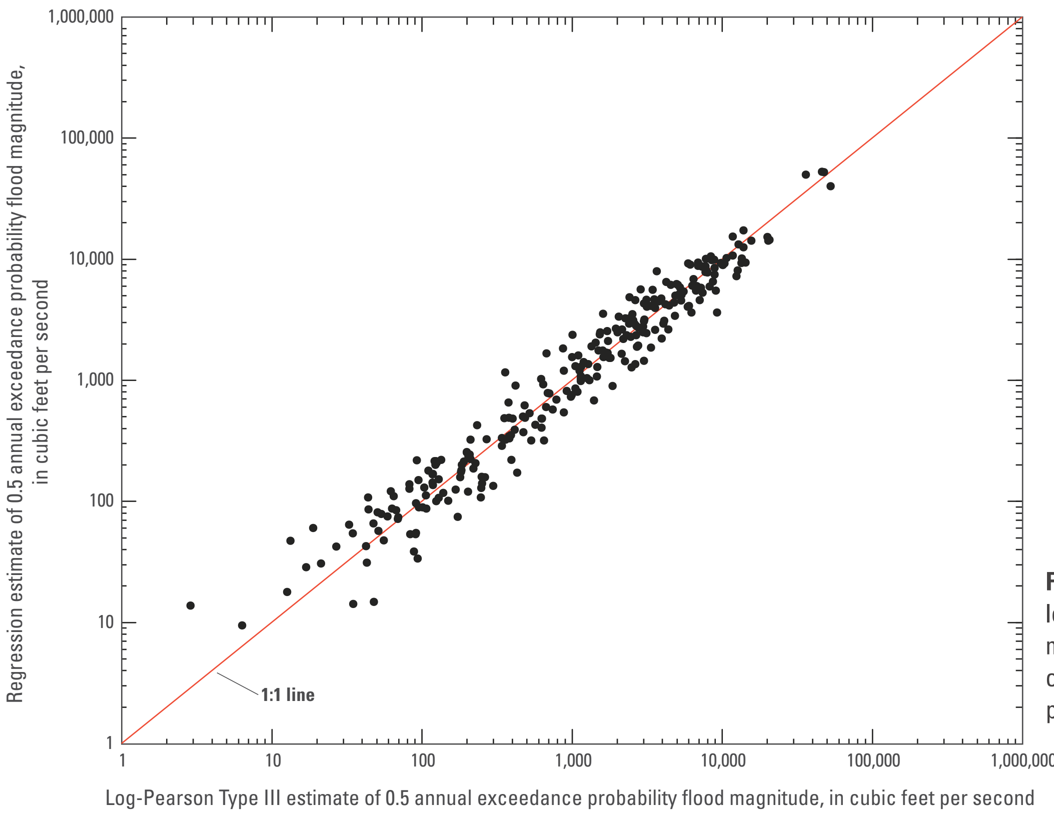

Figure 7. Scatter plot of log-Pearson Type III and full-model regression estimates of the 0.5 annual exceedance probability flood.

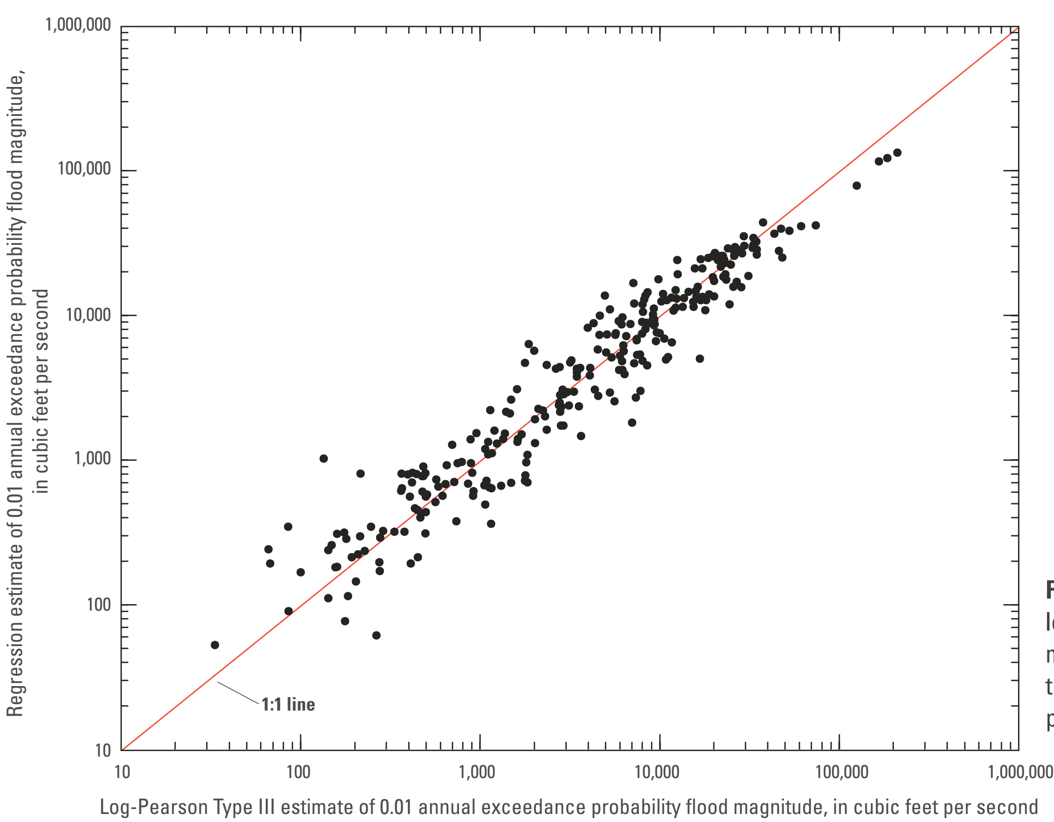

Figure 8. Scatter plot of log-Pearson Type III and full-model regression estimates of the 0.01 annual exceedance probability flood.

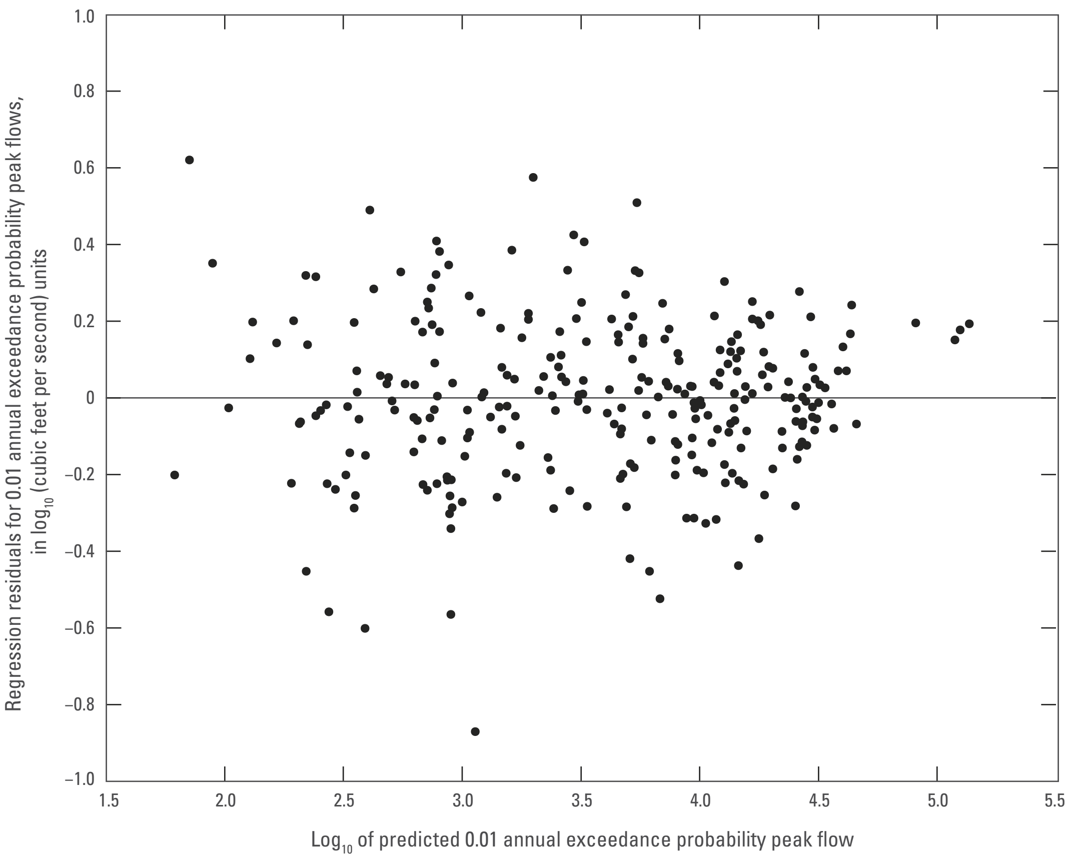

Figure 9. Scatter plot of $\log_{10}$ full-model predicted 0.01 annual exceedance probability peak flows and regression residuals.

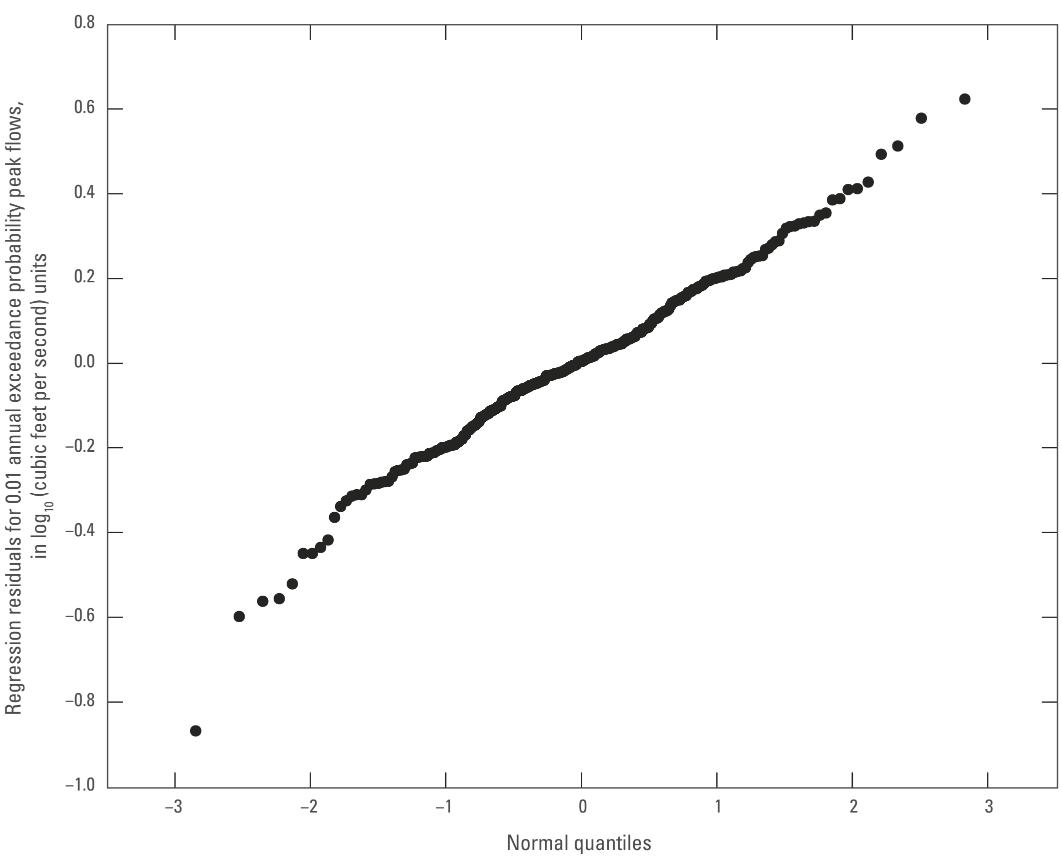

Figure 10. Normal quantile-quantile plot of full-model regression residuals for the 0.01 annual exceedance probability peak flows.

Variance of Prediction

WREG computes the variance of prediction (VP_{reg}) for each observation used in model development. The VP_{reg} for a given observation i, is computed as follows:

$$VP_{reg_i} = \sigma_\delta^2 + \sigma_{s_i}^2 \quad (6)$$

where

σ_δ^2 is the model error variance, and

$\sigma_{s_i}^2$ is the sampling error variance for the ith observation.

The sampling error variance is computed as follows:

$$\sigma_{s_i}^2 = x_i (X^T \Lambda^{-1} X)^{-1} x_i^T \quad (7)$$

where

x_i is a row vector of the regressor variables associated with the ith observation, augmented by a value of 1.0 in the first column;

X is the (n x p) matrix consisting n rows (one for each observed peak discharge) of the p-1 regressor variables associated with the peak discharge augmented by a column of ones in the first column;

Λ is the (n by n) covariance matrix associated with the vector of logarithms of the n observed peak discharges;

$(X^T \Lambda^{-1} X)^{-1}$ is the covariance matrix for the regression coefficients; and

^{T} and $^{-1}$ are superscripts indicating the transpose and inverse of the matrices, respectively.

The $(X^T \Lambda^{-1} X)^{-1}$ matrices for the full model equations are reported in table 1.8.

Computation of Weighted Flood-Frequency Estimates and Confidence Limits at Gaged Sites

The uncertainty in a flood-frequency estimate can be reduced by computing a weighted average of the at-site estimate and the regional-regression estimate. Bulletin 17C England and others, 2018) describes a weighting method in which the base-10 logarithms of the regression and at-site LPIII estimates are weighted inversely proportional to their respective variances as follows:

$$Y_{w_i} = \frac{Y_{site_i} VP_{reg_i} + Y_{reg_i} VP_{site_i}}{VP_{site_i} + VP_{reg_i}} \quad (8)$$

where

Y_{w_i} is the base-10 logarithm of the weighted estimate for site i,

Y_{site_i} is the base-10 logarithm of the at-site LPIII estimate for site i,

VP_{reg_i} is the variance of prediction of the regression estimate for site i,

Y_{reg_i} is the base-10 logarithm of the regression estimate for site i, and

VP_{site_i} is the variance of prediction of the at-site estimate for site i.

Confidence limits for the weighted estimates can be computed as a function of the variance of the weighted estimate (V_{w_i} that is computed as follows:

$$V_{w_i} = \frac{VP_{site_i} VP_{reg_i}}{VP_{site_i} + VP_{reg_i}} \quad (9)$$

Once V_{w_i} has been determined, the upper and lower confidence limits can be computed as follows:

$$CL_{U_i} = 10^{\left[Y_{w_i} + t_{(\frac{\alpha}{2}, n-p)} \left(V_{w_i}\right)^{0.5}\right]} \quad (10)$$

$$CL_{L_i} = 10^{\left[Y_{w_i} - t_{(\frac{\alpha}{2}, n-p)} \left(V_{w_i}\right)^{0.5}\right]} \quad (11)$$

where

CL_{U_i}, CL_{L_i} are the upper and lower confidence limits, respectively, for site i; and

$t_{(\frac{\alpha}{2}, n-p)}$ is Student's t with a specified alpha (α) level and n–p degrees of freedom, where n is the number of sites used in the regression equation and p is the number of regressor variables plus 1.0. For 95-percent confidence limits, $t_{(\frac{\alpha}{2}, n-p)}$ equals approximately 1.969 for the full- and simple-model equations.

Because of the difficulty in computing the sampling error variance, nearly identical computations to those listed in equations 8–11 can be used (with reduced accuracy) to compute weighted estimates and confidence limits for estimates at sites not used in model development. The only difference is that the average variance of prediction for the models (VP_{avg}) (reported in table 4 and table 5) are substituted for VP_{reg}.

Weighting Flood-Frequency Estimates at Ungaged Sites with Data for a Nearby Gage

For unregulated streams, if the drainage area of an ungaged site is between 50 and 150 percent of the drainage area of a gaged site on the same stream and flood-frequency characteristics have been computed for the gaged site, then the following method of adjusting the estimated flood-peak discharge of the ungaged site is suggested:

$$Q_{p,a(u)} = Q_{p,r(u)} \left[R - \frac{2\left|\Delta DA\right|(R-1)}{DA_{(g)}} \right] \quad (12)$$

where

$$R = \left[\frac{Q_{p,w(g)}}{Q_{p,r(g)}} \right] \quad (13)$$

and where

$Q_{p,w(g)}$ is the weighted flood-peak discharge estimate with an AEP of p for the gaged site, reported in table 1;

$Q_{p,r(g)}$ is the regression estimate of flood-peak discharge with an AEP of p for the gaged site, reported in table 1;

$Q_{p,a(u)}$ is the adjusted flood-peak discharge estimate with an AEP of p for the ungaged site;

$Q_{p,r(u)}$ is the regression estimate of flood-peak discharge with an AEP of p for the ungaged site;

$\left|\Delta DA\right|$ is the absolute value of the difference in drainage areas of the gaged and ungaged sites; and

$DA_{(g)}$ is the drainage area of the gaged site.

This method gives more weight to the ungaged site's regression estimate as the difference in drainage areas between the gaged and ungaged sites increases, giving full weight to the regression estimate when the drainage area for the gaging station equals 0.5 or 1.5 times the drainage area of the ungaged site.

An example application of equations 12–13 is provided in equation 14 for an ungaged site on Mill Creek at Ostrander Road (in Delaware County, Ohio). In this example, the 0.01 AEP flood discharge is estimated using the 0.01 AEP regional regression equation listed in table 5 then weighted with the flood frequency estimates reported for streamflow gage 03220000 (Mill Creek near Bellepoint, Ohio), downstream from the ungaged site. The ungaged site is in regression region A (meaning *OHREGA*=1 and *OHREGC*=0 as indicated in the "Ordinary Least-Square Regression" section), and the site is determined with StreamStats to have a drainage area of 167 mi^2, a main-channel slope of 4.83 feet per mile, and 0.77 percent of the drainage covered by open water and wetlands as determined from the 1992 National Land Cover Dataset (Vogelmann and others, 2001). The gage near Bellepoint has a drainage area of 178 mi^2 (table 1.1) and so the drainage area at the ungaged site is about 94 percent of the drainage area at the gaged site, which meets the weighting criteria that the drainage area at the ungagged site must be between 50 and 150 percent of the gaged site.

Using the regression equation for the 0.01 AEP flood discharge in table 5 with the basin characteristics for the ungaged site, the 0.01 AEP flood estimate is determined to be 13,562 ft^3/s as follows:

$$Q_{0.01} = 10^{(2.1648+0.1943[OHREGA]+0.3283[OHREGC])} DA^{0.7431} SL_{10-85}{}^{0.2588} (W+1)^{-0.2508} \quad (14)$$

$$Q_{0.01} = 10^{(2.1648+0.1943[1]+0.3283[0])} 167^{0.7431} 4.83^{0.2588} (0.77+1)^{-0.2508} = 13{,}357 \frac{ft^3}{s}.$$

The 0.01 AEP flood regression and weighted estimates for gaging station 03220000 are 14,100 ft^3/s and 17,500 ft^3/s, respectively (table 1). All necessary data for computing the gage-weighted estimate are now available. Apply equation 13 to determine the value of R as follows:

$$R = \left[\frac{Q_{p,w(g)}}{Q_{p,r(g)}} \right] = \frac{17{,}500}{14{,}100} = 1.2411. \quad (15)$$

Then, substitute the value of R from equation 15 into equation 12 as follows:

$$Q_{p,a(u)} = Q_{p,r(u)} \left[R - \frac{2\left|\Delta DA\right|(R-1)}{DA_{(g)}} \right]$$

$$= 13{,}357 \left[1.2411 - \frac{2\left|178-167\right|(1.2411-1)}{178} \right]$$

$$Q_{p,a(u)} = 16{,}179 \frac{ft^3}{s}.$$

Rounding $Q_{p,a(u)}$ to three significant digits gives a gage-weighted estimate of 16,200 ft^3/s for the 0.01 AEP flood discharge at the ungaged site.

General Guidelines for Estimating Flood-Frequency Characteristics at Sites on Rural, Unregulated Streams

The best method for estimating flood-frequency characteristics for sites on Ohio streams depends on the data available for the site of interest as follows:

1. If a streamgage exists at the site of interest and has 10 or more years of record, techniques described in Bulletin 17C should be used to compute the flood quantiles. However, if peak streamflows are unregulated at the streamgage, the Bulletin 17C estimates should be used to compute a weighted average with estimates determined from regional regression equations.

2. If the site of interest is ungaged and is within 50 and 150 percent of the drainage area of a nearby streamgage on

the same unregulated stream for which flood-frequency characteristics have been or can be computed, then the flood-frequency estimates should be determined using the regional regression equations and then adjusted with equations 12–13 using data for the gaged site.

3. If the site of interest is ungaged and no streamgages are nearby on the same unregulated stream, then flood-frequency estimates should be determined with the regional regression equations.

4. If the site of interest does not fit into one of the categories described in items 1–3 for example, the site of interest is on a regulated stream with no nearby streamgage , then other methods must be used to estimate flood-frequency characteristics. The best method to use will depend on site-specific factors such as available data and resources and, therefore, cannot be generalized.

Flood-frequency estimates for many streamgages are available through StreamStats. If the station number is known for the streamgage, the most recent published characteristics (if available) can be obtained at https://streamstatsags.cr.usgs.gov/gagepages/html/00000000.htm, where 00000000 is substituted with the station number for the streamgage; for example, https://streamstatsags.cr.usgs.gov/gagepages/html/03220000.htm. Alternatively, streamgages can be identified through the StreamStats web application and their associated data accessed by left clicking on the streamgage's symbol and then left clicking on the link next to "StreamStats Gage page" in the window that pops up.

The full-model regression equations have been implemented in StreamStats (and supporting datasets populated) to facilitate computations at ungaged locations. For more information on StreamStats, see Ries and others (2017).

Limitations

The regression equations presented in this report can be used to estimate flood-frequency characteristics for streams in Ohio draining predominantly rural basins that are free of appreciable high-flow regulation. For the purposes of this study, regulated records include any records where the peak streamflows are thought to be substantially altered from what would be expected for a "natural" stream. In general, basins having usable storage of less than 103 acre-feet per square mile are considered to be unregulated; however, the flood-peak discharges for an ungaged site directly below a large reservoir could be considered regulated regardless of the usable storage criterion (Benson, 1962). In addition, basins with extensive channelization or land modifications, or both, associated with urbanization, mining, and some agricultural practices may be considered "regulated" when evaluating applicability of regression equations presented in this report.

The applicability of the regression equations is unknown when the basin-characteristic values associated with an ungaged site are outside a space defined by the basin characteristics associated with the regression dataset (table 3). Although the regression dataset includes data for streamgages with drainage areas greater than 2,500 mi^2, evidence indicates that the regression equations may underestimate flood magnitudes on drainages greater than 2,500 mi^2 for AEPs less than 0.5. Weighted averages of regression estimates and streamgage-based estimates should be used whenever possible, particularly for drainages greater than 2,500 mi^2.

In a previous report that presented equations for estimating flood-peak discharges of rural, unregulated streams in Ohio, Koltun and others (2006) identified a tendency for the full-model regression equations to overestimate peak discharges for basins with greater than 0.3 percent of their drainage classified as quarries, strip mines, or gravel pits. This same bias was not evident with the simple equations. The full-model bias associated with basins with greater than 0.3 percent of their drainage classified as quarries, strip mines, or gravel pits was not reevaluated in this study, but the bias is assumed to persist in this study because the same model form is used.

Flood-frequency characteristics were reported for selected regulated streamgages. In addition, some streamgages had sufficient record to compute flood-frequency characteristics for unregulated and regulated periods. The reader is cautioned that regulation can change at any time (thus changing the flood-frequency characteristics) and that, for streamgages with unregulated and regulated record, the regulated flood-frequency characteristics reflect the more recent of the two periods for that streamgage.

Flood-frequency estimates for streamgages outside of Ohio are reported for informational purposes and are not intended to supersede non-Ohio state-specific sources of flood-frequency information.

Summary

Estimates of the magnitudes of peak streamflows with annual exceedance probabilities of 0.5, 0.2, 0.1, 0.04, 0.02, 0.01, and 0.002 (equivalent to recurrence intervals of 2-, 5-, 10-, 25-, 50-, 100-, and 500-years, respectively) were computed for 391 streamgages in Ohio and adjacent states, based on data collected through the 2015 water year (ending September 30, 2015). Of the 391 streamgages, 333 had 10 or more years of unregulated record. The flood-frequency estimates were computed following guidance outlined in Bulletin 17C and made use of a new regional skew value to compute weighted measures of skewness. Separate analyses also were done to estimate flood-frequency characteristics of 82 streamgages with 10 or more years of regulated record collected through the 2015 water year. The analyses of regulated records were based on measures of skewness at each site.

Two sets of regression models, referred to as simple and full models, were developed for rural, unregulated streams to facilitate estimation of the magnitudes of annual peak streamflows with annual exceedance probabilities of 0.5, 0.2, 0.1, 0.04, 0.02, 0.01, and 0.002. Models were developed with a two-step process involving ordinary-least squares and generalized-least squares regression techniques. Both the simple and full models contain two binary indicator variables representing the regression region in which the basin is located. In addition, the simple model contains a drainage area variable; and the full model contains variables describing the drainage area, main-channel slope, and the percentage of the basin in three land-cover categories representing open-water and wetland areas. The average standard errors of prediction ranged from about 40.5 to 46.5 percent for the simple-model equations and from about 37.2 to 40.3 percent for the full-model equations.

An assessment of fit indicated that the regression equations tended to underestimate flood magnitudes with annual exceedance probabilities less than 0.5 for drainages larger than 2,500 mi^2. This information indicates that regression estimates should be weighted with streamgage-based estimates whenever possible. The full-model regression equations have been implemented in StreamStats to facilitate computations at ungaged locations within Ohio.

Seasonal patterns in annual peak and peak-of-record flows were assessed. A frequency analysis was done on unregulated annual peak flows observed at 295 streamgages in Ohio through the 2015 water year. Annual peak flows occurred most frequently in the 4-month period between January and April, with March having the highest frequency of occurrence. The month with the fewest number of annual peaks was October. The largest recorded annual peaks (peak-of-record flows) at the 295 streamgages occurred most frequently in March, followed by January. March and January coincide with months in which two of Ohio's most severe widespread floods in recent history (the March 1913 and January 1959 floods) occurred. June and July ranked third and fourth with respect to the frequency of peak-of-record flows. None of the peak-of-record flows occurred in October and only two occurred in November.

Annual peak-flow data for 133 streamgages on unregulated streams in Ohio with 30 or more years of systematic record through the 2015 water year were analyzed for long-term trends by computing the rank correlation (as measured with the two-sided Kendall's tau statistic) between time and annual peak flows. Weak but statistically significant trends were indicated at 15 (about 11 percent) of the 133 streamgages. Of the 15 streamgages with significant trend, 12 had positive tau values (indicating an upward trend in annual peak flows) and 3 had negative tau values (indicating a downward trend). All 12 streamgages with positive tau values were at latitudes north of 40°33', and streamgages with negative tau values were at latitudes south of 40°33'.

References Cited

Benson, M.A., 1962, Factors influencing the occurrence of floods in a humid region of diverse terrain: U.S. Geological Survey Water-Supply Paper 1580–B, 64 p.

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 statistic for detecting multiple potentially influential low outliers in flood series: Water Resources Research, v. 49, no. 8, p. 5047–5058.

Cohn, T.A., Lane, W.L., and Baier, W.G., 1997, An algorithm for computing moments-based flood quantile estimates when historical flood information is available: Water Resources Research, v. 3, no. 9, p. 2089–2096.

Cross, W.P., 1946, Floods in Ohio—Magnitude and frequency: Ohio Water Resources Board Bulletin 7, 154 p.

Cross, W.P., and Mayo, R.I., 1969, Floods in Ohio, magnitude and frequency—A supplement to Bulletin 32: Ohio Department of Natural Resources, Division of Water Bulletin 43, 230 p.

Cross, W.P., and Webber, E.E., 1959, Floods in Ohio, magnitude and frequency: Ohio Department of Natural Resources, Division of Water Bulletin 32, 325 p.

Eng, Ken, Chen, Yin-Yu, 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://pubs.usgs.gov/tm/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., accessed November 26, 2018, at https://doi.org/10.3133/tm4B5.

Falcone, J.A., 2017, U.S. Geological Survey GAGES–II time series data from consistent sources of land use, water use, agriculture, timber activities, dam removals, and other historical anthropogenic influences: U.S. Geological Survey data release, https://doi.org/10.5066/F7HQ3XS4.

Farmer, W.H., 2018, Weighted Multiple-Linear-REGression WREG) Program (version 2.02): GitHub web page, accessed June 8, 2018, at https://github.com/wfarmer-usgs/WREG.

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

Gruber, A.M., and Stedinger, J.R., 2008, Models of LP3 regional skew, data selection, and Bayesian GLS regression, *in*, Babcock, R.W. and Watson, R. eds, World Environmental and Water Resources Congress 2008 Ahupua'a, Proceedings, May 12–16, 2008: Honolulu, Hawaii, American Society of Civil Engineers, p. 1–10.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., 2015, Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information: Photogrammetric Engineering and Remote Sensing, v. 81, no. 5, p. 345–354.

Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency—Bulletin 17B of the Hydrology Subcommittee: Reston, Va., U.S. Geological Survey, Office of Water Data Coordination, 183 p.

Koltun, G.F., 2003, Techniques for estimating flood-peak discharges of rural, unregulated streams in Ohio: U.S. Geological Survey Water-Resources Investigations Report 2003–4164, 75 p., accessed November 26, 2018, at https://pubs.er.usgs.gov/publication/wri034164.

Koltun, G.F., 2019a, PeakFQ inputs and selected outputs for selected streamgages in Ohio and border areas of adjacent states (through water year 2015): U.S. Geological Survey data release, https://doi.org/10.5066/P9V3IG0P.

Koltun, G.F., 2019b, Model archive—Regional regression models for estimating flood-frequency characteristics of rural, unregulated Ohio streams: U.S. Geological Survey data release, https://doi.org/10.5066/P9PJ0O5W.

Koltun, G.F., Kula, S.P., and Puskas, B.M., 2006, A Streamflow Statistics (StreamStats) web application for Ohio: U.S. Geological Survey Scientific Investigations Report 2006–5312, 62 p., accessed November 26, 2018, at https://doi.org/10.3133/sir20065312.

Koltun, G.F., and Roberts, J.W., 1990, Techniques for estimating flood-peak discharges of rural, unregulated streams in Ohio: U.S. Geological Survey Water-Resources Investigations Report 89–4126, 68 p., accessed November 26, 2018, at https://pubs.er.usgs.gov/publication/wri894126.

Mastin, M.C., Konrad, C.P., Veilleux, A.G., and Tecca, A.E., 2016, Magnitude, frequency, and trends of floods at gaged and ungaged sites in Washington, based on data through water year 2014 (ver 1.2, November 2017): U.S. Geological Survey Scientific Investigations Report 2016–5118, 70 p., accessed November 28, 2018, at http://dx.doi.org/10.3133/sir20165118.

Mukaka, M.M., 2012, A guide to appropriate use of correlation coefficient in medical research: Malawi Medical Journal, v. 24, no. 3, p. 69–71, accessed November 30, 2018, at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3576830/.

Nash, J.E., and Sutcliffe, J.V., 1970, River flow forecasting through conceptual models part I—A discussion of principles: Journal of Hydrology, v. 10, no. 3, p. 282–290. [Also available at https://doi.org/10.1016/0022-1694(70)90255-6.]

National Oceanic and Atmospheric Administration, National Centers for Environmental Information, 2018, Climate at a Glance: National Oceanic Atmospheric Administration web page, accessed June 26, 2018, at https://www.ncdc.noaa.gov/cag/statewide/time-series/33.

Ohio Department of Natural Resources, 2019, Physiographic Regions, accessed February 5, 2019, at http://naturepreserves.ohiodnr.gov/natural-features-of-ohio/geologic-features/physiographic-regions.

R Core Team, 2017, The R Project for statistical computing: The R Foundation, Vienna, Austria, accessed November 30, 2018, at https://www.R-project.org/.

Ries, K.G., III, Newson, J.K., Smith, M.J., Guthrie, J.D., Steeves, P.A., Haluska, T.L., Kolb, K.R., Thompson, R.F., Santoro, R.D., and Vraga, H.W., 2017, StreamStats, version 4: U.S. Geological Survey Fact Sheet 2017–3046, 4 p., accessed November 26, 2018, at https://doi.org/10.3133/fs20173046.

Sanders, R., ed., 2001, A guide to Ohio streams: Columbus, Ohio, Ohio Chapter of the American Fisheries Society, Streams Committee, 55 p.

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., Eychaner, J.H., and Stedinger, J.R., 1986, Application of generalized least squares in regional hydrologic regression analysis, *in* Subitzky, S., ed., Selected papers in the Hydrologic Sciences 1986: U.S. Geological Survey Water-Supply Paper 2310, p. 107–115.

Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: Journal of Hydrology, v. 111, no. 1–4, p. 361–375.

Tomlinson, E., Kappel, B., Hulstrand, D., Muhlestein, G., and Lovisone, S., 2013, Probable maximum precipitation study for the State of Ohio: Monument, Colo., Applied Weather Associates, LLC, accessed November 26, 2018, at https://water.ohiodnr.gov/portals/soilwater/pdf/dam/1.0_Ohio-Statewide-PMP-Final-Report.pdf.

U.S. Census Bureau, 2018a, State area measurements and internal point coordinates: United States Census Bureau web page, accessed June 26, 2018, at https://www.census.gov/geo/reference/state-area.html.

U.S. Census Bureau, 2018b, Annual estimates of the resident population—April 1, 2010 to July 1, 2017: United States Census Bureau web page, accessed June 26, 2018, at https://factfinder.census.gov/bkmk/table/1.0/en/PEP/2017/PEPANNRES/0100000US|0100000US.04000|0200000US1|0200000US2|0200000US3|0200000US4.

U.S. Census Bureau, 2018c, Annual estimates of the resident population—April 1, 2010 to July 1, 2017: United States Census Bureau web page, accessed June 26, 2018, at https://factfinder.census.gov/bkmk/table/1.0/en/PEP/2017/PEPANNRES/0400000US39|0400000US39.05000.

U.S. Geological Survey, 2018, U.S. Geological Survey peak streamflow for the Nation: National Water Information System database, accessed December 20, 2018, at https://nwis.waterdata.usgs.gov/usa/nwis/peak.

U.S. Geological Survey, 2019, Hydrologic Unit Maps, accessed February 6, 2019, at https://water.usgs.gov/GIS/huc.html.

Veilleux, A.G., 2009, Bayesian GLS regression for regionalization of hydrologic statistics, floods and Bulletin 17 skew: Ithaca, N.Y., Cornell University, master's thesis, 155 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. [Also available at https://doi.org/10.3133/fs20133108.]

Veilleux, A.G., Stedinger, J.R., and Lamontagne, J.R., 2011, Bayesian WLS/GLS regression for regional skewness analysis for regions with large cross-correlations among flood flows, *in* World Environmental and Water Resources Congress 2011—Bearing Knowledge for Sustainability, Palm Springs, Calif., 2011, Proceedings: Palm Springs, Calif., American Society of Civil Engineers Environmental and Water Resources Institute, p. 2253–2263.

Veilleux, A.G., and Wagner, D.M, in press, Methods for estimating regional skewness of annual peak flows in parts of the Great Lakes and Ohio River Basins, based on data through water year 2013: U.S. Geological Survey Scientific Investigations Report 2019–5019.

Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K., and Van Driel, J.N., 2001, Completion of the 1990's National Land Cover Data Set for the conterminous United States: Photogrammetric Engineering and Remote Sensing, v. 67, p. 650–662.

Webber, E.E.; and Bartlett, W.P., Jr., 1977, Floods in Ohio, magnitude and frequency: Ohio Department of Natural Resources, Division of Water Bulletin 45, 76 p.

Appendix 1

These tables are available for download at https://doi.org/10.3133/sir20195018

Table 1.1 Basin characteristics of streamgages with 10 or more years of unregulated record.

Table 1.2 Basin characteristics of streamgages with 10 or more years of regulated record.

Table 1.3 Selected flood-frequency-analysis-related characteristics for unregulated streamgages.

Table 1.4 Selected flood-frequency-analysis-related characteristics for regulated streamgages.

Table 1.5 Nondefault flow intervals for unregulated streamgages.

Table 1.6 Perception thresholds for unregulated streamgages.

Table 1.7 Perception thresholds for regulated streamgages.

Table 1.8 Covariance matrices for regression coefficients $[(X^T\Lambda^{-1}X)^{-1}]$ in the full model equations.

For additional information contact:
Director, Ohio-Kentucky-Indiana Water Science Center
U.S. Geological Survey
6460 Busch Boulevard Ste 100
Columbus, OH 43229-1737

For additional information visit:
https://www.usgs.gov/centers/oki-water

Publishing support provided by the
Madison and Rolla Publishing Service Centers

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20195018