Prepared in cooperation with Wisconsin Department of Transportation

Estimating Flood Magnitude and Frequency for Unregulated Streams in Wisconsin

Scientific Investigations Report 2022–5118
Supersedes Scientific Investigations Report 2016–5140

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

Front and back cover. Photo showing flooding on the Kickapoo River, August 2018 in Ontario, Wisconsin. Photograph by Zack Scott, U.S. Geological Survey.

Estimating Flood Magnitude and Frequency for Unregulated Streams in Wisconsin

By Sara B. Levin and Chris A. Sanocki

Prepared in cooperation with Wisconsin Department of Transportation

Scientific Investigations Report 2022–5118
Supersedes Scientific Investigations Report 2016–5140

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

U.S. Geological Survey, Reston, Virginia: 2023
Supersedes Scientific Investigations Report 2016–5140

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:
Levin, S.B., and Sanocki, C.A., 2023, Estimating flood magnitude and frequency for unregulated streams in Wisconsin: U.S. Geological Survey Scientific Investigations Report 2022–5118, 25 p., https://doi.org/10.3133/sir20225118. [Supersedes Scientific Investigations Report 2016–5140.]

Associated data for this publication:
Levin, S.B., 2022, PeakFQ inputs and outputs for 299 streamgages in Wisconsin through water year 2020: U.S. Geological Survey Data Release, https://doi.org/10.5066/P9BDHH6D.

Levin, S.B., 2023, Model archive—Regional regression models for estimating flood frequency characteristics of unregulated streams in Wisconsin: U.S. Geological Survey data release, https://doi.org/10.5066/P9JVO8QV.

ISSN 2328-0328 (online)

Contents

Abstract 1
Introduction....1
Purpose and Scope 1
Description of Data....2
Peak Flow Data and Site Selection....2
Basin Characteristics....2
Redundancy Analysis....4
Flood Frequency Analysis at Gaged Locations....4
Trend Analysis 5
EMA Methodology....5
Multiple Grubbs-Beck Test....6
Regional Flood Frequency Regression Equations....8
Flood Frequency Regions 8
Development of Regional Regression Equations 10
Accuracy and Limitations of Regression Equations 11
Prediction Intervals for Regression-based Flood Discharge Estimates....17
Application of Techniques for Estimating Flood Magnitudes at Gaged and Ungaged Locations....19
Estimating the Weighted Flood Discharge at a Streamgage....19
Example 1 20
Estimating the Flood Discharge at an Ungaged Location on a Gaged Stream 21
Example 2 21
Estimating the Flood Discharge at an Ungaged Location without a Nearby Streamgage....22
Example 3 22
Web Application for Solving Regional Regression Equations....23
Summary....23
References Cited....23

Figures

1. Map showing streamgages and flood frequency regions used to estimate peak flow frequencies and magnitudes in Wisconsin 3
2. Boxplots showing distributions of selected basin characteristics for four flood frequency regions in Wisconsin 9
3. Graph showing the relation between drainage area and 1-percent AEP flood discharge for four flood frequency regions in Wisconsin....10
4. Graphs showing flood discharges corresponding to the 0.5- and 1-percent annual exceedance probabilities estimated by the at-site log-Pearson type III distribution and regression equations for four hydrologic regions in Wisconsin 15
5. Boxplots showing flood frequency estimates for the 1-percent annual exceedance probability per unit drainage area for streamgages with different periods of record for four regions in Wisconsin 18
6. Graph showing annual peak streamflow at U.S. Geological Survey streamgage 05436500, Sugar River near Brodhead, WI, from water year 1940 through 2020 19

Tables

1. U.S. Geological Survey streamgages in Wisconsin that recorded extraordinary floods between water years 2011 and 2020 7
2. Regression equations for estimating discharges for the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probability floods for ungaged streams in Wisconsin....12
3. Ranges of basin characteristics used in regional flood frequency regression equations for four regions in Wisconsin....14

Conversion Factors

U.S. customary units to International System of Units

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

International System of Units to U.S. customary units

Multiply	By	To obtain
	Length	
centimeter (cm)	0.3937	inch (in.)
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
	Area	
square kilometer (km^2)	0.3861	square mile (mi^2)
	Flow rate	
cubic meter per second (m^3/s)	35.31	cubic foot per second (ft^3/s)

Datum

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

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

Supplemental Information

A water year is the 12-month period from October 1 through September 30 and is designated by the year in which it ends.

Abbreviations

AEP	annual exceedance probability
AVP	average variance of prediction
DAR	drainage area ratio
EMA	expected moments algorithm
GLS	generalized least squares
LP3	Log-Pearson Type III distribution
OLS	ordinary least squares
PeakFQ	U.S. Geological Survey peak-flow frequency analysis program
PILF	potentially influential low flow
PRESS	predicted residual sum of squares
R^2	coefficient of determination
SD	standardized distance
SME	standard model error
Sp	standard error of prediction
USGS	U.S. Geological Survey
WREG	weighted-multiple-linear-regression program

Estimating Flood Magnitude and Frequency for Unregulated Streams in Wisconsin

By Sara B. Levin and Chris A. Sanocki

Abstract

Flood frequency characteristics and estimated flood discharges for the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities were computed at 299 streamgaged locations in Wisconsin. The State was divided into four flood frequency regions using a cluster analysis to produce regions which are homogeneous with respect to physical basin characteristics. Regression equations relating flood discharges to basin characteristics within each region were developed and can be used to estimate flood discharges at ungaged locations in Wisconsin. Basin characteristics included in the final regression equations include drainage area, saturated hydraulic conductivity, percent forest, percent herbaceous upland, percent open water, and the maximum 24-hour precipitation with a 10-year recurrence interval. The standard error of prediction for regression equations ranges between 40 and 71 percent, and the pseudo coefficient of determination ranges between 0.8 and 0.95. Nonmonotonic trends in the annual peak flow time series in the southwest part of the State are producing bias in some flood discharge estimates at streamgages with shorter (less than 20 years) periods of record. This bias increases the uncertainty in regression equations in this flood frequency region.

Introduction

Flood frequency analysis refers to the statistical analysis used to estimate the magnitude and frequency of floods at gaged or ungaged locations. Flood frequency information is used in a variety of infrastructure and public safety projects such as the design of dams, culverts, bridges, and highways and is used in flood insurance and flood-plain management. Annual peak streamflow collected at U.S. Geological Survey (USGS) streamgages (U.S. Geological Survey, 2022) are used to estimate flood discharges at specific annual exceedance probabilities (AEPs). The estimated flood discharges at gaged locations are used to develop regression equations relating physical basin characteristics to flood magnitudes. These regression equations can then be used to estimate probable flood discharges at ungaged locations.

Periodic updates of flood frequency estimates and regional regression equations are necessary to incorporate new data and methods. There is a lengthy history of flood frequency studies in Wisconsin starting with an initial report in 1961 (Ericson, 1961) and with additional reports following roughly every 10 years with updated data and statistical techniques (Conger, 1971; Conger, 1981; Krug and others,1992; Walker and Krug, 2003; Walker and others, 2017). Since the publication of the previous flood frequency report by Walker and others (2017), the USGS has issued new guidelines for estimating flood frequency estimates at gaged locations (England and others, 2019). These new guidelines, hereafter referred to as Bulletin 17C, address several concerns with the statistical methodology used in previous reports and improve flood frequency estimates at streamgages with censored data or low outliers. This study, done in cooperation with the Wisconsin Department of Transportation, includes updated annual peak flow data through water year 2020 and implements the new statistical methodologies outlined in Bulletin 17C.

Regression equations are used to estimate flood discharges corresponding to selected AEPs for ungaged locations on streams. Regression equations in this study were developed using the updated flood frequency estimates and basin characteristics at streamgages without substantial regulation or urbanization and do not include the main stems of the Wisconsin River, St. Croix River, or the Mississippi River. In large and hydrologically diverse states such as Wisconsin, regression equations can be improved by grouping the available streamgages into hydrologically similar regions before the development of the equations. For this study, four flood frequency regions were developed for Wisconsin using a clustering algorithm. The additional data and updated statistical methodologies used in this report increase the confidence in the resulting flood frequency estimates and supersede the frequency analyses and regression equations in previous reports.

Purpose and Scope

The purpose of this report is to present methods for estimating the magnitude and frequency of floods for unregulated, rural streams in Wisconsin. This report (1) describes the statistical methods used to estimate the flood discharge magnitudes at gaged locations; (2) presents the estimated flood discharges for the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent AEPs at 299 streamgages in Wisconsin; (3) describes the methods used to develop regression equations for estimating the frequency and magnitude of floods at ungaged locations; and (4) presents the final regional flood frequency regression equations for Wisconsin with examples for their usage.

Description of Data

Flood frequency analysis and development of regression equations require two types of data: (1) annual peak streamflow and (2) physical basin characteristics. Annual peak streamflow, defined as the maximum instantaneous streamflow recorded during a water year, is needed to estimate flood discharge for selected AEPs at streamgaged locations. Regression equations relate flood discharges for selected AEPs to basin characteristics at each streamgage and are used to estimate probable flood discharges on ungaged streams.

Peak Flow Data and Site Selection

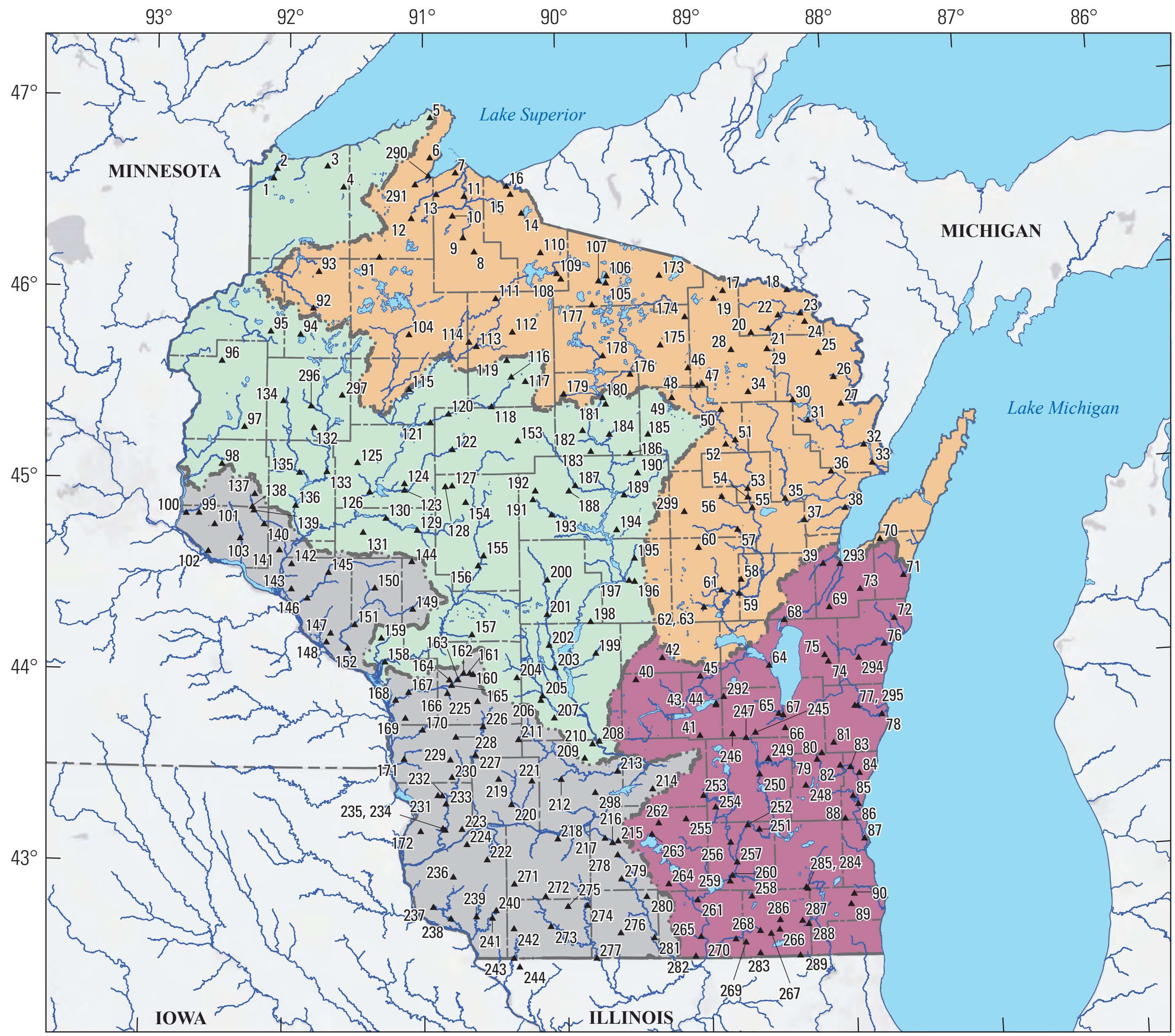
Annual peak streamflow data were obtained for all active and inactive streamgages in basins in Wisconsin with at least 10 years of annual peak streamflow observations through water year 2020 (U.S. Geological Survey, 2022). Annual peak streamflow can be affected by a variety of anthropogenic changes to the drainage basin including, but not limited to, dams and water diversions, urbanization, and agricultural practices such as tile drainage and ditching. Streamgages that were regulated or anthropogenically altered were removed from the analysis because of the different statistical properties of annual peak flows at these sites. Streams that are substantially affected by regulation from upstream dams were identified either through information within the National Water Information System data records (U.S. Geological Survey, 2022) or by knowledge of the flow systems.

In addition to regulation from dams, candidate streamgages were screened for urbanization. The percentage of developed land computed by the Wisconsin StreamStats web application (U.S. Geological Survey, 2016) was used to identify streamgages in urban basins. Streamgages in basins with more than 20 percent developed land were removed from the study because of the potential for streamflow alteration from impervious surface or changes in channelization. After screening for regulation and urbanization, data from a total of 299 streamgages remained for flood frequency analysis including 168 continuous and 131 crest-stage gages (fig 1; map numbers and streamgage numbers are available in Levin, 2023, table 1).

Basin Characteristics

Basin characteristics were used in cluster analysis for identifying homogeneous regions for regression equation development and as explanatory variables in the regional regression equations. A suite of 24 basin characteristics consisting of geophysical, climatic, and land-use characteristics were determined for each streamgage in the study. Basin characteristics that were considered for this study included basin characteristics developed and used in previous flood frequency studies in Wisconsin (Walker and others, 2017). All the basin characteristics that were used in this study, except for mean basin slope, were previously described by Walker and others (2017) and were computed using the Wisconsin StreamStats web application (U.S. Geological Survey, 2016). The full suite of basin characteristics are as follows, and those that were used as explanatory variables in the final regional regressions are listed in Levin (2023, table 2).

- Drainage area (*DRNAREA*), in square miles, is the area of surface runoff contributing to each streamgage. Watershed boundaries were delineated for each streamgage using the StreamStats web application (U.S. Geological Survey, 2016). Before delineation, streamgage locations were adjusted if necessary to assure the streamgage was coincident with the stream grid used by StreamStats for watershed delineation. No adjustments to drainage areas were made for potential noncontributing areas.
- Channel slope (*CSL10_85*), in feet per mile, is the change in elevation divided by length between points 10 and 85 percent of distance along the main channel to the basin divide.
- Mean basin slope (*BSLDEM*), measured in degrees. Mean basin slope was computed in a Geographic Information System using the slope tool, which uses a moving window to estimate slope across each basin and returns the average slope value.
- Mean annual precipitation (*PRECIP*), in inches, computed for the years 1971–2000.
- Mean annual snowfall (*SNOWFALL*), in inches, for the years 1971–2000.
- Saturated hydraulic conductivity (*SSURGOKSAT*), in micrometers per second, is the ease in which water can move through a medium.
- 24-hour precipitation indices (*I24H100Y* etc.), in inches, is the maximum 24-hour precipitation that happens on average once every 2, 5, 10, 25, 50, or 100 years.
- Climate-factors (*CLIFAC100Y*, *CLIFAC25Y*, *CLIMFAC2YR*), dimensionless, are regression-based indices developed by Lichty and Karlinger (1990) that



Base U.S. Geological Survey 1:2,000,000 Digital Line Graph and ESRI Online CANVEC various scales
Base map from Esri and its licensors, copyright 2022
Universal Transverse Mercator projection, Zone 15
Horizontal coordinate information is referenced to the North American Datum 1983

EXPLANATION

Flood frequency regions

- 1
- 2
- 3
- 4

Region boundary

U.S. Geological Survey streamgage station and map number—See Levin (2023, table 1) for map numbers and associated streamgages

Figure 1. Streamgages and flood frequency regions used to estimate peak flow frequencies and magnitudes in Wisconsin.

identify regional trends in small-basin flood frequency based on long-term rainfall and pan evaporation information.

- Land-use categories (land use categories from the 2001 National Land Cover Dataset) were computed within the StreamStats web application as a percentage of total basin area. Land uses in this study included Forest (*FOREST*), Wetland (*WETLAND*), Developed (*DEVNLCD01*), cultivated crops and hay (*LC01CRPHAY*), herbaceous upland (*LC01HERB*), and open water (*LC01WATER*). Additionally, the updated 2011 National Land Cover Dataset land use categories emergent herbaceous wetlands (*LC11EMWET*) and wooded wetlands (*LC11WDWET*) were computed.

Redundancy Analysis

One assumption in regression analysis is that the data are spatially independent. Redundancy happens when two streamgages of similar drainage areas are nested (meaning that one basin is contained within the other) and have similar basin characteristics and peak flow magnitudes. This can happen when two gages are on the same stream and there are no large confluences or incoming tributaries between them. In these cases, the basins will likely have the same response to a given storm and thus would represent only a single spatial observation.

A redundancy analysis outlined by Veilleux (2009) and Parrett and others (2011) was used to identify redundant streamgages in Wisconsin. To be considered redundant, the basins must be nested and have similar drainage areas. Two indicators were computed for each possible pair of streamgages: the standardized distance (SD_{ij}) and the maximum drainage area ratio (DAR). The SD_{ij}, defined below, is used to identify potentially nested basins.

$$SD_{ij} = \frac{D_{ij}}{\sqrt{0.5\left(DA_i + DA_j\right)}}, \tag{1}$$

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.

Previous studies have determined that an SD_{ij} equal to or less than 0.5 indicates a high likelihood that two basins are nested (Veilleux, 2009; Parrett and others, 2011).

The DAR is computed as the ratio of drainage area of the larger basin to the drainage area of the smaller basin and was used to determine if two nested basins are similar in size. Previous studies have considered nested pairs of gages with a DAR less than or equal to five to be redundant (Veilleux, 2009; Eash and others, 2013; Koltun, 2019). If the DAR is greater than five, even if basins are nested, the basin characteristics and response to storm systems are likely different; therefore, the streamgages are not considered redundant.

The SD_{ij} and DAR were computed for every possible pair of streamgages within the initial list of 299 unregulated, rural streamgages. For all pairs of streamgages with SD_{ij} less than or equal to 0.5 and DAR less than 5, streamgage locations were visually inspected to confirm that they were nested and that the streamgages were on the same stream reach with no confluences of large tributaries between them. When a pair of redundant streamgages was identified, the streamgage with the longer period of record was retained for use in the regression equations and the other streamgage was removed from the regression analysis. Of the candidate streamgages, 31 were removed because of redundancy. Streamgages that were excluded from regression analyses are noted in Levin (2023, table 1).

Flood Frequency Analysis at Gaged Locations

Flood frequency analysis uses statistical techniques to estimate flood discharges associated with specific AEPs or recurrence intervals. An AEP is the probability that a flood of a specific magnitude will happen in a given year. Formerly, AEPs have been reported as recurrence intervals where the recurrence interval, in years, is the reciprocal of the AEP; for example, the flood corresponding to the 1-percent AEP is commonly referred to as the 100-year flood. The recurrence interval terminology is now discouraged because its interpretation can cause confusion; therefore, the AEP terminology is used in this report.

Flood frequency analyses were performed using the Expected Moments Algorithm (EMA) method with the Multiple Grubbs-Beck test for potentially influential low floods (PILFs), as recommended in Bulletin 17C (England and others, 2019). Bulletin 17C is an update to previously used guidelines outlined by the Interagency Advisory Committee on Water Data (1982; hereafter referred to as Bulletin 17B). Flood frequency estimates at a streamgage are computed by fitting a Pearson type III distribution (LP3) to the logarithms of annual peak streamflows. The mean, standard deviation, and skew of the annual peak streamflow data describe the midpoint, slope, and curvature of the fitted distribution. Because skew values computed from datasets with few peak streamflow observations are unreliable, skews computed from the annual peak flows are weighted with a regional skew determined from an analysis of selected long-term streamgages in the study region. Regional skews used for streamgages in this study were computed and published by Walker and others (2017).

Trend Analysis

A primary assumption of flood frequency analysis is that the mean and variability of annual peak flows at a streamgage are not changing with time. Previous studies have identified trends in annual peak flows in Wisconsin streams (Juckem and others, 2008; Splinter and others, 2015; Gebert and others, 2016). Trends are most prevalent in the southwest area of Wisconsin (flood frequency region 4, fig. 1). In this region, decreasing trends in annual peak flows, accompanied by increasing trends in low flow or daily streamflow, have been attributed to a combination of changes in precipitation and changes in agricultural land management (Juckem and others, 2008; Gyawali and others, 2015; Gebert and others, 2016). Trends in peak flows in other areas of Wisconsin are weaker and less widespread than in the southwest part of the State.

Annual peak flow data were assessed for the presence of monotonic trends using the Mann-Kendall test (Helsel and Hirsch, 2002). Kendall's τ is the test statistic of the Mann-Kendall test. Kendall's τ measures the correlation between the rank order of annual peak flows and time. A Kendall's τ of 0 indicates there is no trend or correlation in peak flows through time. A correlation of 1 indicates a perfectly monotonically upward trend and a τ of -1 indicates a perfectly monotonically downward trend. A *p*-value is used to test the null hypothesis that the Kendall's τ value equals zero. A *p*-value equal to or less than 0.05 indicates a statistically significant upward or downward trend. Trend detection in streamflow is sensitive to the length of the period of record at the streamgage. Natural interdecadal climate fluctuations can affect streamflow and cause multiyear periods of higher or lower flows which may look like a monotonic trend for data with a short period of record. A period of record of 30 years or longer is typically used for detection of trends to avoid attributing trends in streamflow to natural climatic variability.

The results of Mann-Kendall tests of trends for all streamgages are in Levin (2023, table 1). Trends were identified in 28 streamgages having a period of record of 30 years or more, with upward trends at 10 streamgages and downward trends at 18 streamgages. Downward trends were found primarily in the southwest part of the state (flood frequency region 4, fig. 1) with several other downward trends in the northeast part of the state with drainage into Lake Michigan (flood frequency region 2, fig. 1). Upward trends were primarily in the southeastern part of the State (flood frequency region 3, fig. 1) surrounding Milwaukee. The Mann-Kendall test can be sensitive to multiyear sequences of high or low values at the beginning or end of the period of record. Many of the streamgages that had an upward trend also had two or more large floods at the end of the period of record which could have affected the detection of a trend and may not represent a long-term trend. Streamgages that were identified as having statistically significant trends were examined to determine if there was any apparent regulation, although no evidence of regulation was found.

Although there is evidence that trends in annual peak streamflow happen at a regional scale in Wisconsin, a detailed analysis of the causal factors of these trends is beyond the scope of this report. Consistent with previous flood frequency reports in Wisconsin (Walker and others, 2017), data from streamgages with a statistically significant trend were retained in the flood frequency analysis because of uncertainty about the causation and lack of guidance on how to account for trends in flood frequency. Flood frequency analyses at streamgages with trends may lead to greater uncertainty in regional flood frequency equations. Methods for estimating flood frequency in the presence of trends is an active area of research and may be incorporated into future flood frequency reports in Wisconsin to decrease the uncertainty in the flood frequency estimates and regional regression equations.

EMA Methodology

Flood frequency analyses for 299 streamgages in this study were computed with USGS software PeakFQ version 7.3 (Flynn and others, 2006) using the EMA method and Multiple Grubbs-Beck test for PILFs. Flood discharges corresponding to selected AEPs are listed in Levin (2023, table 1). Additional information about the analyses (such as perception thresholds, flow intervals, low outlier thresholds used by the Multiple Grubbs-Beck test, and flood discharges corresponding to additional AEPs) can be found in Levin (2022).

The EMA method was used to estimate the LP3 distribution for all streamgages used in this study. This method is described in Bulleting 17C (England and others, 2019). The EMA method addresses several methodological concerns identified in Bulletin 17B (Interagency Advisory Committee on Water Data, 1982), including the ability to incorporate censored data, historical floods, and improved treatment of low outliers. For streamgages that have peak flow records for complete periods (no data gaps) and no low outliers, censored values, or historical flood estimates, the EMA method will produce the same fit to the LP3 distribution as the previously used method-of-moments described in Bulletin 17B.

The EMA represents each peak flow as an interval with a lower and upper bound which enables the method to incorporate data with various forms of uncertainty, such as censored data or historical floods. A flow interval represents the uncertainty associated with a peak flow. For most annual peak flows, the upper and lower bound are set at the reported peak flow value. Censored data occur when an annual peak flow is only known to be above or below some threshold value. This can happen at crest-stage gages when the annual peak flow is known to be below a minimum recordable value. Previously used flood frequency methods described in Bulletin 17B omitted such values because of their lack of precision and high uncertainty; however, the EMA method can incorporate this information by representing the censored peak streamflow as an interval that is bounded by zero and the streamflow associated with the elevation of the gage's minimum recording threshold. By using interval data, the EMA method can use more years of data to fit the distribution while also accounting for the greater uncertainty in these censored values.

Multiple Grubbs-Beck Test

PILFs are values in the peak flow data that have lower magnitudes than other peak flows at the same location and which exert a high influence on the fit of the LP3 distribution. The physical processes that result in PILFs are often different than those that result in large floods, and including PILFs in the data when fitting the LP3 distribution can bias estimation of the largest floods (the lowest AEPs). Because these large floods are typically of more interest, it is recommended by Bulletin 17C that PILFs be identified and removed from the data to improve the fit of the distribution to the larger floods that correspond to smaller AEPs.

Previous flood frequency analyses in Wisconsin used the standard Grubbs-Beck test recommended by Bulletin 17B to identify PILFs. This test is adept at identifying a single PILF within a dataset but is unreliable when there are two or more PILFs. Bulletin 17C recommends the use of the Multiple Grubbs-Beck test, which is a generalization of the former test that is more sensitive to the presence of several PILFs.

Differences between estimated flood discharges in this report and the previously published report (Walker and others, 2017) may happen because of additional years of peak flow data or methodological changes. Of the streamgages that were used in both reports, roughly 90 percent had changes in flood frequency estimates that were within ±20 percent of those published previously by Walker and others (2017); however, there were isolated cases where the updated flood frequency estimates differed substantially from previous estimates. Many of the increases in estimated flood discharges were because of large floods that happened between 2011 and 2020 across much of the state, after the analysis period of Walker and others (2017). Bulletin 17C defines an extraordinary flood as a peak streamflow whose magnitude exceeds the second largest peak streamflow at a streamgage by a factor of two or more (England and others, 2019). There were 31 active streamgages with at least 20 years of data that recorded an annual peak of record between 2011 and 2020. Six of those streamgages had one or more annual peak flows since 2010 that were twice as large or larger than any prior peak (table 1). These extraordinary floods exerted a large influence on the fitted LP3 distribution and resulted in increases in estimates of flood discharges that were more than twice as large as previous estimates. Changes in methodology also contributed to changes in flood frequency estimates. Because of differences in how the EMA methodology handles PILFs and censored values, streamgages with many of these types of data may have differences in their estimated flood frequencies compared to previously published estimates.

Table 1. U.S. Geological Survey streamgages in Wisconsin that recorded extraordinary floods between water years 2011 and 2020.

[USGS, U.S. Geological Survey; ft³/s, cubic foot per second]

USGS streamgage number	Station name	Period of record (water years)	Maximum recorded annual peak streamflow through water year 2010		Maximum recorded annual peak streamflow from 2011 to 2020	
			Water year	Peak streamflow (ft^3/s)	Water year	Peak streamflow (ft^3/s)
04024430	NEMADJI RIVER NEAR SOUTH SUPERIOR, WI	1974–2020	2001	15,800	2011	33,000
04027200	TWENTYMILE CREEK AT GRAND VIEW, WI	1960–2020	1992	1,920	2016	9,000
04074850	LILY RIVER NEAR LILY, WI	1970–2020	2005	190	2020	419
05331833	NAMEKAGON RIVER AT LEONARDS, WI	1996–2020	2001	952	2016	2,680
05379288	BRUCE VALLEY CREEK NEAR PLEASANTVILLE, WI	1996–2017	2010	560	2016	2,080
05382200	FRENCH CREEK NEAR ETTRICK, WI	1960–2020	2001	2,950	2017	7,100

Regional Flood Frequency Regression Equations

Regional regression equations were developed to estimate the magnitude of flood discharges for selected exceedance probabilities at ungaged streams in Wisconsin. Before regression equation development, flood frequency regions were delineated by which had similar basin characteristics. Regression equations were developed using multiple linear regression, which relates streamflows corresponding to various AEPs to basin characteristics by region.

Flood Frequency Regions

Climatic and physiographic characteristics that affect the flood responses of streams vary widely across Wisconsin. Dividing the State into hydrologically similar regions can help increase the predictive accuracy of regression equations. Hydrologic similarity refers to the tendency for streamflow in two or more basins to respond similarly to a given rainfall event. Streamgages in regions that exhibit hydrologic similarity typically have similar basin characteristics and produce regression equations with higher accuracy. Previous studies have divided Wisconsin into five regions (Conger, 1971; Conger, 1981; Krug and others;1992; Walker and Krug, 2003) based on the patterns of residuals from a statewide regression using flood frequency and basin characteristics data from 1971 or earlier. Walker and others (2017) updated Wisconsin flood frequency regions using ecoregion boundaries to divide the State into eight regions. Flood frequency regions were redelineated for this study to optimize regional homogeneity while also ensuring that flood frequency regions contained enough streamgages to produce reliable regression equations.

A clustering analysis was used to divide the State into homogeneous regions with respect to basin characteristics using a process outlined in Rao and Srinivas (2008). Cluster analysis is a statistical method that classifies multidimensional data into distinct groups. Groups are chosen such that the similarity between members in a group is maximized while the similarity of members within different groups is minimized. In defining flood frequency regions, the goal is to produce spatially nonoverlapping regions that are homogeneous with respect to basin characteristics and that have an adequate number of streamgages (preferably, a minimum of 40 gages per region) on which to base the regression equations. A subset of basin characteristics that were independent and showed variability across the state were selected for the cluster analysis. Basin characteristics that were considered in the clustering analysis include *BSLDEM*, *SSURGOKSAT*, *I24H10Y*, *SNOWFALL*, *WETLAND*, *FOREST*, *LC01WATER*, *LC01HERB*. Basin characteristics such as drainage area were not used in the cluster analysis because the distribution of drainage areas does not vary across the state.

Clustering was performed with the K-means clustering algorithm using the factoextra package in R (Kassambara and Mundt, 2020; R Core Team, 2021). Following a procedure outlined in Rao and Srinivas (2008), the optimum number of clusters was chosen based on visual interpretation of maps and validity metrics, which measure the similarity of streamgages within each group and the difference in basin characteristics between different groups. Because cluster analyses can be sensitive to the starting set of basin characteristics, the analysis was repeated using several different starting sets of basin characteristics. The final clustering arrangement was chosen from among the optimal clustering from each analysis based on validity metrics and tests of homogeneity.

After selecting the final cluster results, regional boundaries were manually adjusted to prevent overlapping regions or to maintain region boundaries consistent with 8-digit or 12-digit hydrologic unit boundaries. Although it may be unavoidable for a large drainage basin to cross into more than one region, it is desirable to have regions that generally follow drainage basin boundaries to avoid a situation where a stream crosses in and out of the same region multiple times, which could result in inconsistent flood frequency estimates at different points along the stream.

Four final flood frequency regions were defined for Wisconsin (fig 1). The flood frequency regions developed from the clustering analysis were affected by north-south gradients of precipitation (*PRECIP*), snowfall (*SNOWFALL*), and land cover patterns such as percent forest (*FOREST*), wetlands (*WETLAND*), and open water (*LC01WATER*). Differences in regional basin characteristics and the relation between drainage area and the 1-percent AEP flood frequency estimate are shown in figures 2 and 3. Regions 1 and 2 cover the northern half of the State including drainage into Lakes Superior and Michigan and the headwaters of larger south- flowing rivers. These two regions are characterized by lower annual precipitation (*PRECIP*) and maximum 24-hour, precipitation with a 10-year recurrence interval *(I24H10Y)* and greater snowfall (*SNOWFALL*), soil permeability (*SSURGOKSAT*), and percentage of forested (*FOREST*) area than the southern regions (fig. 2). Streamflows in region 2 corresponding to various AEPs were generally lower than other areas of the state (fig. 3). The southern-most regions, 3 and 4, are characterized by greater precipitation (*PRECIP*) and lesser snowfall (*SNOWFALL*) and percentages of forest (*FOREST*) and wetlands (*WETLAND*) than the northern regions. Region 4 covers much of the driftless area of Wisconsin. This region was not glaciated and is geomorphologically different than the rest of the State. Region 4 has greater basin slopes (*BSLDEM*) and fewer lakes (*LC01WATER*) and wetlands (*WETLAND*) than the other three regions and also is characterized by greater magnitude flood discharges than the other regions (figs. 2 and 3).

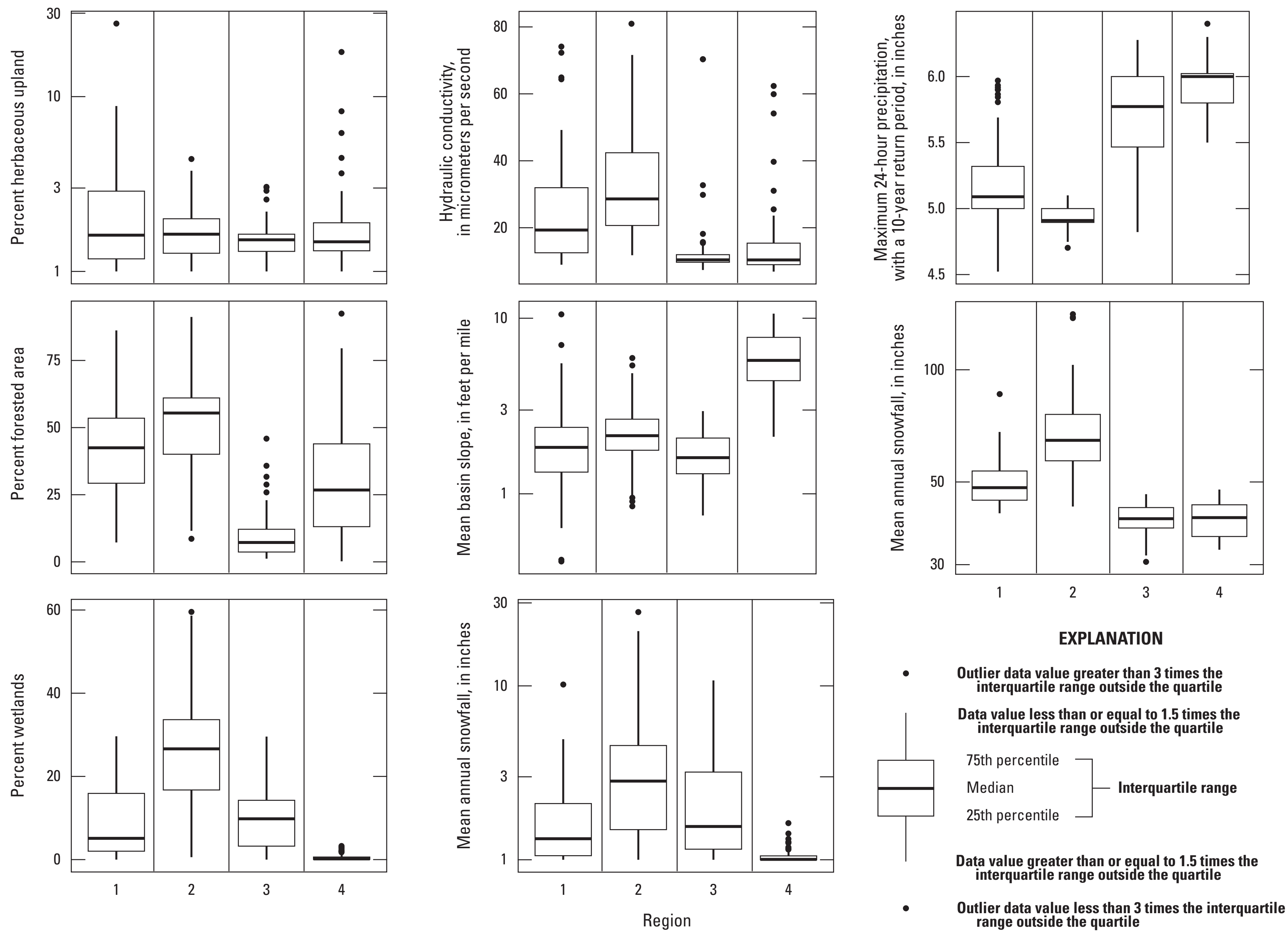


Figure 2. Distributions of selected basin characteristics for four flood frequency regions in Wisconsin.

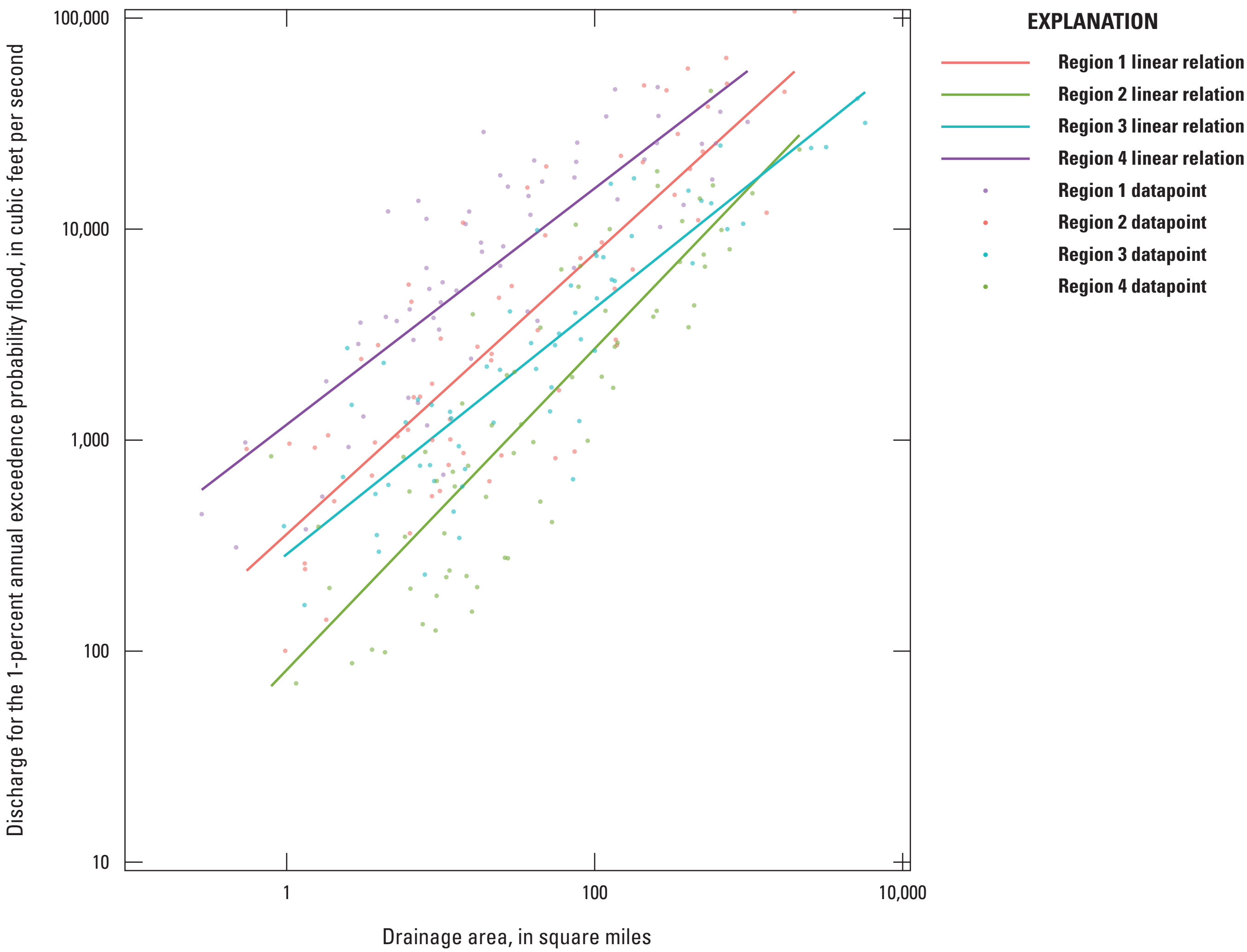


Figure 3. Relation between drainage area and 1-percent AEP flood discharge for four flood frequency regions in Wisconsin.

Development of Regional Regression Equations

Regional regression equations were developed for estimating flood discharges corresponding to selected AEPs at ungaged locations on streams in Wisconsin. The development of regional regression equations includes two steps: (1) exploratory analysis, in which variables were transformed and the pool of potential explanatory variables was reduced; and (2) final model selection, in which the final models are selected and fit for all AEPs. During exploratory analyses, ordinary least squares (OLS) regression was used because of the ease of use and the availability of variable selection techniques for this regression method. For selection and fitting of the final models, generalized least squares (GLS) regression was used. GLS regression accounts for unequal record lengths as well as spatial correlation of concurrent flows at different streamgages and provides better estimates of the predictive accuracy of the regression equations (Stedinger and Tasker, 1985). All GLS equations were fit using the WREG package in R (Farmer, 2017; R Core Team, 2021). Documentation of all GLS model parameters and WREG outputs is documented in Levin (2023). For further detailed explanations about OLS and GLS regression techniques, refer to the WREG user's guide and other related publications (Eng and others, 2009; Stedinger and Tasker, 1985; Tasker and Stedinger, 1989).

During exploratory analysis, the relation between explanatory variables and the flood discharge corresponding to the 1-percent AEP was examined for linearity and variables were examined for potential multicollinearity. Scatterplot matrices of the log-transformed (base 10) flood discharges, log-transformed (base 10) explanatory variables, and untransformed explanatory variables were generated to evaluate whether log-transformation of the explanatory variables was needed and to check for correlation of the explanatory variables with flood discharge. Multicollinearity happens when explanatory variables used in a regression model are highly correlated with each other. Regression models that include variables with multicollinearity are unreliable because the regression equation coefficients and standard errors may be biased. The potential set of explanatory variables was reduced before the variable selection process such that no two explanatory variables had a Pearson's R correlation coefficient greater than 0.6.

The variable selection process identifies the best subset of explanatory characteristics to use in a regression model. To minimize predictive inconsistencies between flood frequency estimates among different AEPs, variable selection analyses were performed using the 1-percent AEP flood discharge. After a final set of explanatory variables was chosen for a region, equations for the other AEPs were fit with GLS regression using the same set of explanatory variables. The best subsets method (from the 'leaps' R package; Lumley 2020) was used to identify the best potential subsets of explanatory variables. This method fits regression models for all possible combinations of explanatory variables and returns the best three 1- through 5-variable models based on the coefficient of determination (R^2) values. Candidate regression models were then evaluated based on maximizing the R^2, while minimizing the predicted residual sum of squares (PRESS) and Mallow's Cp. Additionally, explanatory variables for each candidate model were assessed for statistical significance and multicollinearity. Multicollinearity was evaluated by computing the variance inflation factors. For this study, candidate models were eliminated from consideration if variance inflation factors were greater than 2 or if coefficients for any explanatory variables had p-values greater than 0.05.

The best three equations, as suggested by the best subsets OLS analysis, were refit and examined using GLS. Final GLS regional regression equations were selected based on minimizing the standard model error (SME), standard error of prediction (Sp), average variance of prediction (AVP), and the PRESS statistic while maximizing pseudo coefficient of variation (pseudo R^2) as well as visual assessments of fit and residuals. The performance metrics pseudo R^2 and SME indicate how well the equations perform on the streamgages used in the regression analyses. The Sp, AVP, and PRESS statistic are measures of the accuracy with which GLS regression models can predict streamflows corresponding to various AEPs at ungaged sites. Regression models contain sampling error and model error. Sampling error refers to uncertainty in the flood frequency data used to derive the regression equation. Model error refers to errors stemming from uncertainty in the coefficients of the model equation. SME measures the error of the model itself and does not include sampling error. The Sp represents the sum of the model error and the sampling error. The AVP is a measure of the average accuracy of prediction for all sites used in the development of the regression model and assumes that the explanatory variables for the streamgages included in the regression analysis are representative of all streamgages in the region. The pseudo R^2 is a measure of the percentage of the variation in annual peak streamflow explained by the variables included in the model.

Streamgages that were flagged by the WREG program as having large influence or leverage were further examined for elimination. Leverage is a measure of how much the values of explanatory variables at a streamgage vary from values of those variables at all other streamgages. Influence is a measure of how strongly the values for a streamgage affect the estimated regression parameters. Residual scatterplots were compared to fitted values and explanatory variables were examined to determine if flagged streamgages with large influence and leverage were isolated hydrologic outliers and could be removed from the analysis.

Streamgages that had high leverage, influence, or substantial lack of fit within the selected regression and had fewer than 15 years of annual peak streamflow records were removed from the dataset because of the high level of uncertainty in the estimates of the selected AEPs. Although a period of record of at least 10 years of peak flow data is recommended for flood frequency analysis (England and others, 2019), flood frequency estimates for streamgages having periods of record less than 20 years are highly uncertain and may be biased by short term climate variability or unreliable LP3 parameter estimation (Douglas and others, 2000; Hu and others, 2020). In these cases, the lack of fit at a particular streamgage is likely caused by the short period of record and not representative of flood frequency characteristics of the region. Eleven streamgages were removed from the regression analysis for these reasons and are noted in Levin (2023, table 1).

Accuracy and Limitations of Regression Equations

The final regression equations and performance metrics are shown in table 2. Drainage area (*DRNAREA*) was a statistically significant basin characteristic in all flood frequency regions. Other basin characteristics that were statistically significant in the regional regression equations include *SSURGOKSAT*, *LC01WATER*, *LC01HERB*, *FOREST*, *WETLAND*, and *I24H10Y*. Overall, regression-based estimates of flood discharges had good agreement with those estimated by the at-site LP3 analysis (fig. 4). The Sp for regression equations in all regions ranged from 40.0 to 71.2 percent. The pseudo R^2 ranged from 80.0 to 95.0 percent, and the SME ranged from 38.1 to 66.9 percent. The regression equations presented here are valid for estimating the magnitude and frequency of floods at ungaged locations on streams in Wisconsin for which (1) the streamflow is not substantially altered because of urbanization or regulation and (2) the basin characteristics at the ungaged location are within the range of those used to develop the equations (table 3),

The accuracy and uncertainty of the regression equations is affected by imprecision, inaccuracies, or incomplete data in the basin characteristics and the estimates of selected AEPs at gaged sites on which the equations are based. Regression equations are a simplification of actual physical processes and may not adequately represent the physical flood dynamics in all cases; for example, the effect of lakes and wetlands on streamflow in regions 1–3 depends on their size and their location within the drainage network. A simple percentage of

Table 2. Regression equations for estimating discharges for the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probability floods for ungaged streams in Wisconsin.

[RR, regression region; AEP, annual exceedance probability; R^2, coefficient of determination; Sp, standard error of prediction; SME, standard model error; AVP, average variance of prediction; %, percent; *QP*, estimated flood discharge for the *p*-percent annual exceedance probability; *DRNAREA*, drainage area, in square miles; *SSURGOKSAT*, hydraulic conductivity, in micrometers per second; *LC01HERB*, percentage of LC01HERBaceous upland; *LC01WATER*, percentage of open LC01WATER in drainage basin; *I24H10Y*, 24-hour maximum precipitation with a 10-year return period; *FOREST*, percent forest in drainage basin; *WETLAND*, percent wetland in drainage basin; -, not applicable]

RR	AEP	Equation	*a*	*b*	*c*	*d*	*e*	*f*	Pseudo R^2	Sp	SME	AVP
1	50%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT) + d \times \log(LC01HERB+1) + e \times \log(LC01WATER+1) + f \times I24H10Y$	0.936	0.840	−0.668	−0.374	−0.514	0.423	0.92	53.58	50.83	0.05
1	20%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT) + d \times \log(LC01HERB+1) + e \times \log(LC01WATER+1) + f \times I24H10Y$	0.437	0.834	−0.654	−0.440	−0.583	0.610	0.91	55.98	53.05	0.05
1	10%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT) + d \times \log(LC01HERB+1) + e \times \log(LC01WATER+1) + f \times I24H10Y$	0.205	0.830	−0.638	−0.473	−0.621	0.698	0.90	58.47	55.35	0.06
1	4%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT) + d \times \log(LC01HERB+1) + e \times \log(LC01WATER+1) + f \times I24H10Y$	−0.018	0.825	−0.616	−0.507	−0.663	0.784	0.89	61.00	57.64	0.06
1	2%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT) + d \times \log(LC01HERB+1) + e \times \log(LC01WATER+1) + f \times I24H10Y$	−0.148	0.822	−0.599	−0.529	−0.691	0.835	0.89	62.72	59.19	0.06
1	1%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT) + d \times \log(LC01HERB+1) + e \times \log(LC01WATER+1) + f \times I24H10Y$	−0.252	0.819	−0.582	−0.547	−0.717	0.877	0.88	65.24	61.50	0.07
1	0.50%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT) + d \times \log(LC01HERB+1) + e \times \log(LC01WATER+1) + f \times I24H10Y$	−0.339	0.816	−0.566	−0.564	−0.741	0.913	0.87	67.77	63.80	0.07
1	0.20%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT) + d \times \log(LC01HERB+1) + e \times \log(LC01WATER+1) + f \times I24H10Y$	−0.433	0.812	−0.544	−0.583	−0.770	0.953	0.86	71.16	66.90	0.08
2	50%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(LC01WATER+1) + d \times I24H10Y + e \times FOREST + f \times WETLAND$	−1.980	0.945	−0.720	1.002	−0.007	−0.007	0.95	40.04	38.08	0.03
2	20%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(LC01WATER+1) + d \times I24H10Y + e \times FOREST + f \times WETLAND$	−2.172	0.939	−0.770	1.121	−0.007	−0.008	0.94	42.88	40.74	0.03
2	10%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(LC01WATER+1) + d \times I24H10Y + e \times FOREST + f \times WETLAND$	−2.283	0.936	−0.794	1.183	−0.007	−0.009	0.94	44.83	42.54	0.03

Table 2. Regression equations for estimating discharges for the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probability floods for ungaged streams in Wisconsin.—Continued

[RR, regression region; AEP, annual exceedance probability; R^2, coefficient of determination; Sp, standard error of prediction; SME, standard model error; AVP, average variance of prediction; %, percent; *QP*, estimated flood discharge for the *p*-percent annual exceedance probability; *DRNAREA*, drainage area, in square miles; *SSURGOKSAT*, hydraulic conductivity, in micrometers per second; *LC01HERB*, percentage of LC01HERBaceous upland; *LC01WATER*, percentage of open LC01WATER in drainage basin; *I24H10Y*, 24-hour maximum precipitation with a 10-year return period; *FOREST*, percent forest in drainage basin; *WETLAND*, percent wetland in drainage basin; -, not applicable]

RR	AEP	Equation	a	b	c	d	e	f	Pseudo R^2	Sp	SME	AVP
2	4%	log QP = a + b × log(DRNAREA) + c × log(LC01WATER+1) + d × I24H10Y + e × FOREST + f × WETLAND	−2.405	0.933	−0.819	1.249	−0.007	−0.009	0.93	47.73	45.21	0.04
2	2%	log QP = a + b × log(DRNAREA) + c × log(LC01WATER+1) + d × I24H10Y + e × FOREST + f × WETLAND	−2.490	0.931	−0.834	1.293	−0.007	−0.010	0.92	50.55	47.83	0.04
2	1%	log QP = a + b × log(DRNAREA) + c × log(LC01WATER+1) + d × I24H10Y + e × FOREST + f × WETLAND	−2.567	0.929	−0.848	1.332	−0.007	−0.010	0.92	53.35	50.42	0.05
2	0.50%	log QP = a + b × log(DRNAREA) + c × log(LC01WATER+1) + d × I24H10Y + e × FOREST + f × WETLAND	−2.641	0.927	−0.860	1.368	−0.007	−0.010	0.91	56.15	53.00	0.05
2	0.20%	log QP = a + b × log(DRNAREA) + c × log(LC01WATER+1) + d × I24H10Y + e × FOREST + f × WETLAND	−2.732	0.925	−0.874	1.412	−0.007	−0.011	0.90	59.88	56.45	0.06
3	50%	log QP = a + b × log(DRNAREA) + c × log(LC01WATER+1) + d × WETLAND	1.625	0.785	−0.476	−0.013	-	-	0.92	44.60	42.75	0.03
3	20%	log QP = a + b × log(DRNAREA) + c × log(LC01WATER+1) + d × WETLAND	1.882	0.772	−0.536	−0.014	-	-	0.91	46.05	44.10	0.04
3	10%	log QP = a + b × log(DRNAREA) + c × log(LC01WATER+1) + d × WETLAND	2.012	0.766	−0.565	−0.015	-	-	0.90	48.19	46.10	0.04
3	4%	log QP = a + b × log(DRNAREA) + c × log(LC01WATER+1) + d × WETLAND	2.148	0.762	−0.595	−0.016	-	-	0.88	51.60	49.28	0.04
3	2%	log QP = a + b × log(DRNAREA) + c × log(LC01WATER+1) + d × WETLAND	2.234	0.759	−0.614	−0.017	-	-	0.87	54.49	51.98	0.05
3	1%	log QP = a + b × log(DRNAREA) + c × log(LC01WATER+1) + d × WETLAND	2.310	0.757	−0.630	−0.017	-	-	0.86	57.39	54.69	0.05
3	0.50%	log QP = a + b × log(DRNAREA) + c × log(LC01WATER+1) + d × WETLAND	2.379	0.756	−0.644	−0.018	-	-	0.85	60.33	57.43	0.06
3	0.20%	log QP = a + b × log(DRNAREA) + c × log(LC01WATER+1) + d × WETLAND	2.461	0.754	−0.661	−0.019	-	-	0.83	65.00	61.80	0.07
4	50%	log QP = a + b × log(DRNAREA) + c × log(SSURGOKSAT)	2.365	0.623	−0.420	-	-	-	0.86	54.80	52.90	0.05

Table 2. Regression equations for estimating discharges for the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probability floods for ungaged streams in Wisconsin.—Continued

[RR, regression region; AEP, annual exceedance probability; R^2, coefficient of determination; Sp, standard error of prediction; SME, standard model error; AVP, average variance of prediction; %, percent; *QP*, estimated flood discharge for the *p*-percent annual exceedance probability; *DRNAREA*, drainage area, in square miles; *SSURGOKSAT*, hydraulic conductivity, in micrometers per second; *LC01HERB*, percentage of LC01HERBaceous upland; *LC01WATER*, percentage of open LC01WATER in drainage basin; *I24H10Y*, 24-hour maximum precipitation with a 10-year return period; *FOREST*, percent forest in drainage basin; *WETLAND*, percent wetland in drainage basin; -, not applicable]

RR	AEP	Equation	*a*	*b*	*c*	*d*	*e*	*f*	Pseudo R^2	Sp	SME	AVP
4	20%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT)$	2.892	0.585	−0.560	-	-	-	0.87	49.92	48.08	0.04
4	10%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT)$	3.168	0.565	−0.637	-	-	-	0.87	49.48	47.55	0.04
4	4%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT)$	3.459	0.546	−0.720	-	-	-	0.86	51.56	49.44	0.04
4	2%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT)$	3.645	0.534	−0.773	-	-	-	0.85	53.41	51.13	0.05
4	1%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT)$	3.810	0.525	−0.822	-	-	-	0.83	56.44	53.96	0.05
4	0.50%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT)$	3.959	0.517	−0.866	-	-	-	0.82	59.15	56.49	0.06
4	0.20%	$\log QP = a + b \times \log(DRNAREA) + c \times \log(SSURGOKSAT)$	4.138	0.508	−0.920	-	-	-	0.80	63.83	60.90	0.06

Table 3. Ranges of basin characteristics used in regional flood frequency regression equations for four regions in Wisconsin.

[*n*, number of streamgages in region; -, basin characteristic was not used in region]

Basin characteristic	Region 1, *n*=66		Region 2, *n*=68		Region 3, *n*=61		Region 4, *n*=62	
	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
Drainage area, in square miles	0.58	2,085.76	0.83	2,243.41	1.01	5,994.64	0.30	1,034.01
Percent forest	-	-	8.60	91.20	-	-	-	-
Maximum 24-hour precipitation with a 10-year recurrance interval, in inches	3.61	4.17	3.40	4.00	-	-	-	-
Percent herbaceous upland	0	25.16	-	-	-	-	-	-
Percent open water	0	9.20	0	25.62	0	9.75	-	-
Hydraulic conductivity, in micrometers per second	8.95	74.16	-	-	-	-	6.90	62.34
Percent wetland	-	-	0.58	59.58	0	29.57	-	-

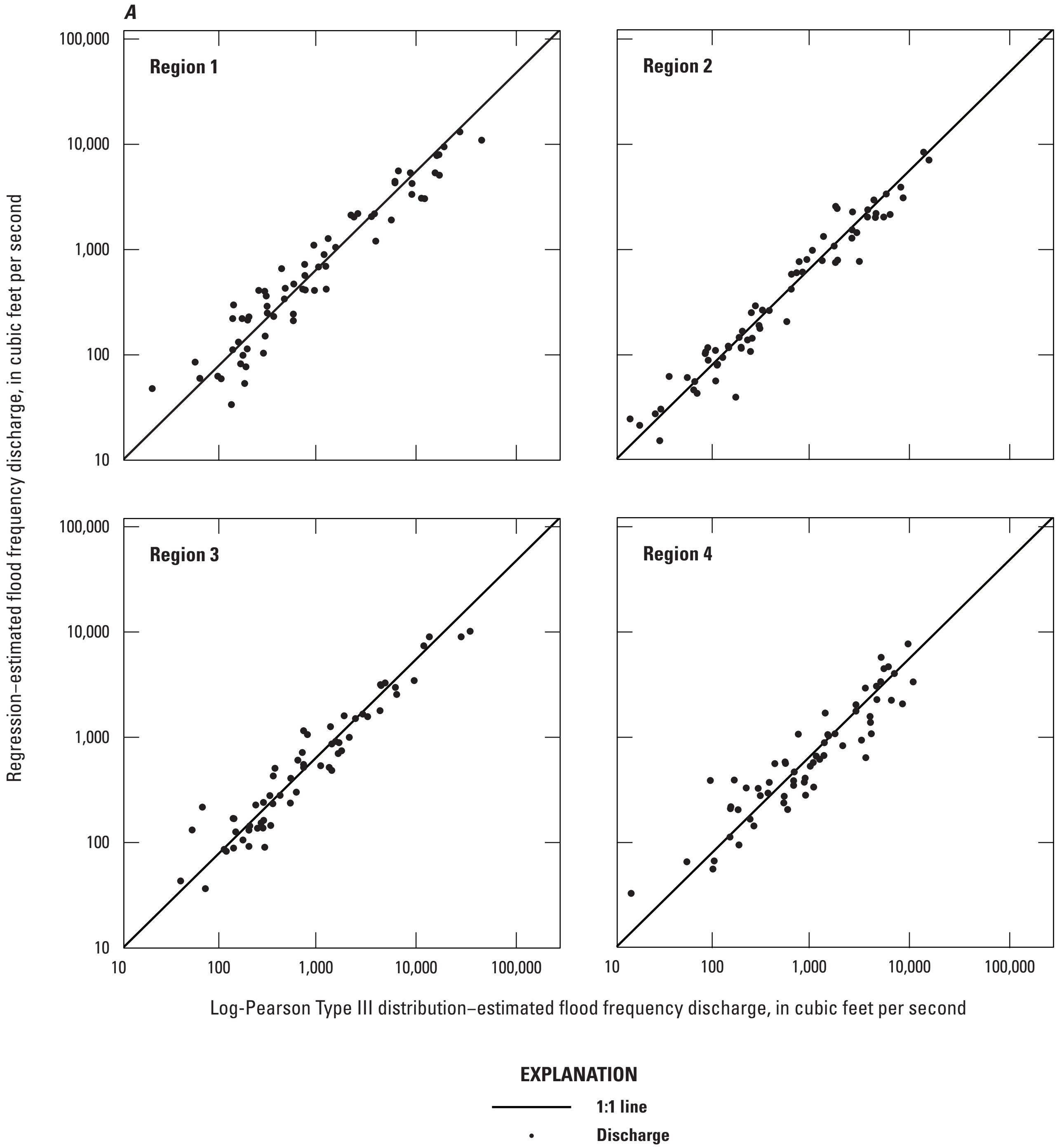


Figure 4. Flood discharges corresponding to the, *A*, 0.5- and, *B*, 1-percent annual exceedance probabilities estimated by the at-site log-Pearson type III distribution and regression equations for four hydrologic regions in Wisconsin.

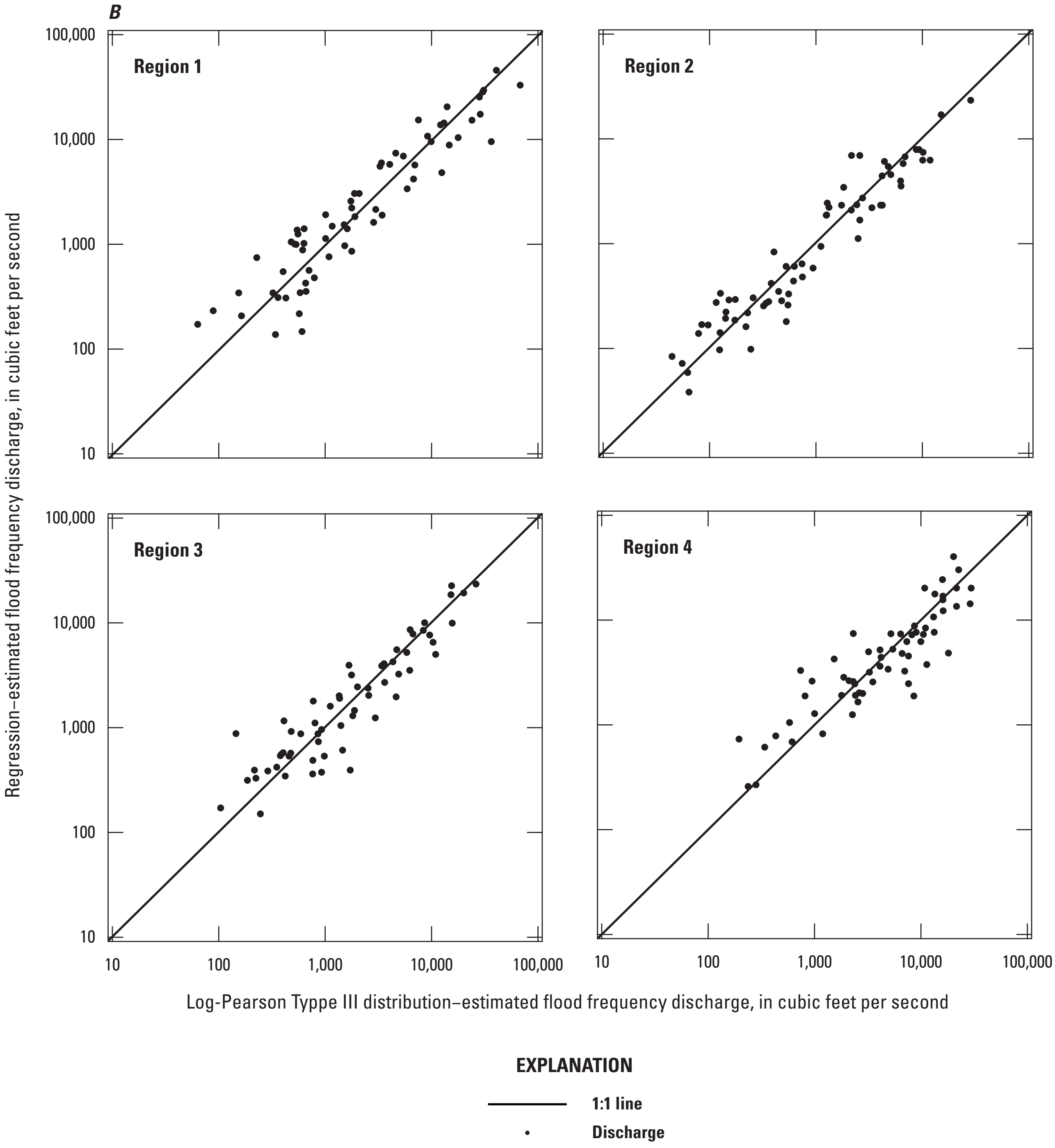


Figure 4. Flood discharges corresponding to the, *A*, 0.5- and, *B*, 1-percent annual exceedance probabilities estimated by the at-site log-Pearson type III distribution and regression equations for four hydrologic regions in Wisconsin.—Continued

drainage area that is open water used in these regression equations may not always adequately account for the spatial and hydrologic connectivity of these basin characteristics.

Uncertainty in the regression equations also is affected by uncertainty in the LP3 flood discharge estimates at the streamgages used in the regression analysis. Bias in flood discharge estimates can result from short periods of record that do not adequately cover the full range of long-term climatic conditions at a stream or periods of record in which there is a monotonic trend. Uncertainty in estimated flood discharges resulting from the presence of a trend in peak streamflow propagates into the regression equations and adds additional uncertainty and, potentially, bias. Although trends in annual peak flows were detected at some locations, the causal attribution of those trends and development of an appropriate method for adjusting the estimated AEP discharge at locations with trends was beyond the scope of this study.

The equations in region 4 show some bias toward overestimating 1-percent AEP flood frequencies for smaller drainage basins with predicted flood discharge of 1,000 cubic feet per second (ft^3/s) or less and underestimating moderately sized basins with predicted flood discharge around 10,000 ft^3/s (fig 4*B*). There are two factors that are potentially affecting the fit of the regression equations in this region: (1) LP3 estimated flood discharges per unit of drainage area in region 4 have higher variability by record length than in other regions, particularly at gages with short (less than 30 years) periods of record (fig. 5); and (2) annual peak flow time series at many long-term gages (greater than 70 years of record) in region 4 have a distinct "U" shape with downward trends until the 1980s and subsequent upward trends after 2000 (fig. 6). Most of these nonmonotonic trends were not detected by standard trend tests such as the Mann-Kendall test. Figure 6 shows an example of this pattern at USGS streamgage 05436500, Sugar River near Brodhead, Wis. (map number 281, fig 1), with a loess-smoothed line representing the trend in the median annual peak flow. Depending on the starting and ending years of recorded data, the LP3 flood frequency estimates at a streamgage with a shorter period of record may only represent a part of this "U" shape and may not reflect the full range of annual peak streamflows at that location resulting in a biased estimate of flood discharge. This nonmonotonic pattern in peak streamflow causes a high amount of variability in flood frequency estimates throughout the region as streamgages that only cover the middle part of the period may have LP3 flood frequency estimates that are substantially lower than streamgages whose period of record only cover the earliest or latest 20–30 years; for example, the LP3 estimated flood discharge for the 1-percent AEP for the data shown in figure 6 is 15,970 ft^3/s; however, if there were only 20 years of data available at this stream, the resulting 1-percent LP3 estimated flood discharge could be as low as 8,700 ft^3/s (using data from 1975 through 1995) or as high as 24,400 ft^3/s (using data from 1914 through 1934). This issue is most prevalent in small- and medium-sized basins in region 4, which also have the shortest record lengths.

Prediction Intervals for Regression-based Flood Discharge Estimates

The goodness-of-fit metrics reported in table 2 are measures of average model uncertainty based on all streamgages used in the model, but they are not representative of the uncertainty for a single estimated flood discharge. Users of the regression equations may be interested in the uncertainty associated with a specific flood discharge estimate at an ungaged location. One such measure of site-specific uncertainty is a prediction interval. A prediction interval is a range of values that will encompass the true value with some nominal probability; for example, a 90-percent prediction interval for an estimated flood discharge has a 90 percent probability that the true value of the flood discharge is within the interval. While prediction intervals for OLS regressions can be easily computed from the standard error of the regression equation, prediction intervals for the GLS regressions used in this report must account for the cross-correlations between peak flow time series at all streamgages and the differing lengths of peak flow record.

Tasker and Driver (1988) developed a method for estimating the prediction interval of a GLS estimate:

$$\frac{Q}{C} < Q < QC \quad , \tag{2}$$

where

Q is the estimated flood discharge for a given AEP at an ungaged location predicted from a regression equation; and

C is computed as:

$$C = 10^{t\left(\frac{\alpha}{2}, n-p\right) SE_{p,i}} \quad , \tag{3}$$

where

$t\left(\frac{\alpha}{2}, n-p\right)$ is the critical value from a student's t-distribution for an alpha level (α), and degrees of freedom (n−p). Critical values for 90-percent (α=0.1) prediction intervals for each equation are available in Levin (2023, table 4); and

$SE_{p,i}$ is the standard error of prediction for ungaged site i, computed as:

$$SE_{p,i} = [MEV + \mathbf{X_i U X_i^T}]^{0.5} \quad , \tag{4}$$

where

MEV is the model error variance,

$\mathbf{X_i}$ is a row vector of basin characteristics, starting with 1 as a placeholder for the intercept term, for ungaged site i,

$\mathbf{U}$ is the covariance matrix for the regression coefficients, and

$\mathbf{X_i^T}$ is the matrix transpose of $\mathbf{X_i}$.

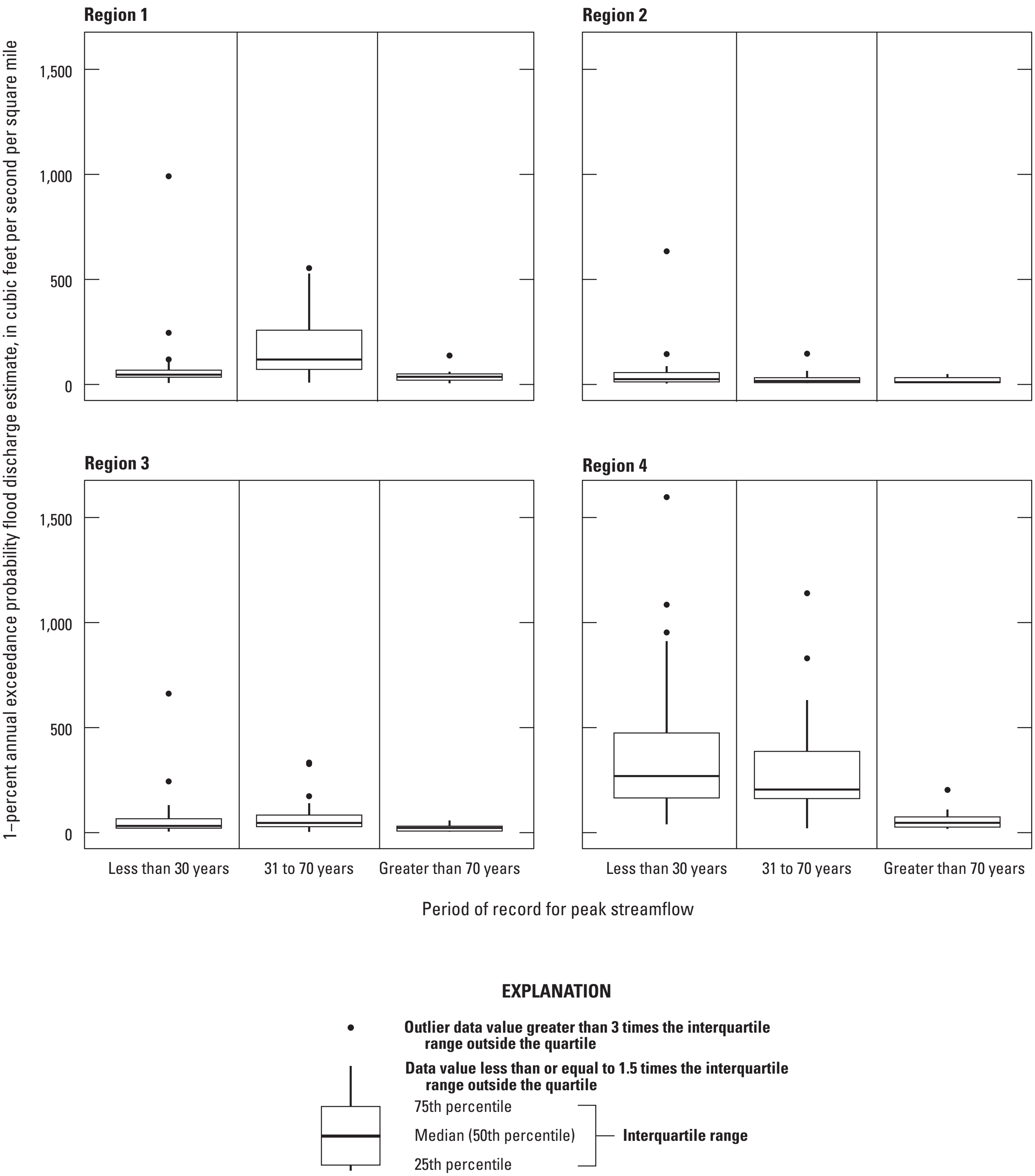


Figure 5. Flood frequency estimates for the 1-percent annual exceedance probability per unit drainage area for streamgages with different periods of record for four regions in Wisconsin.

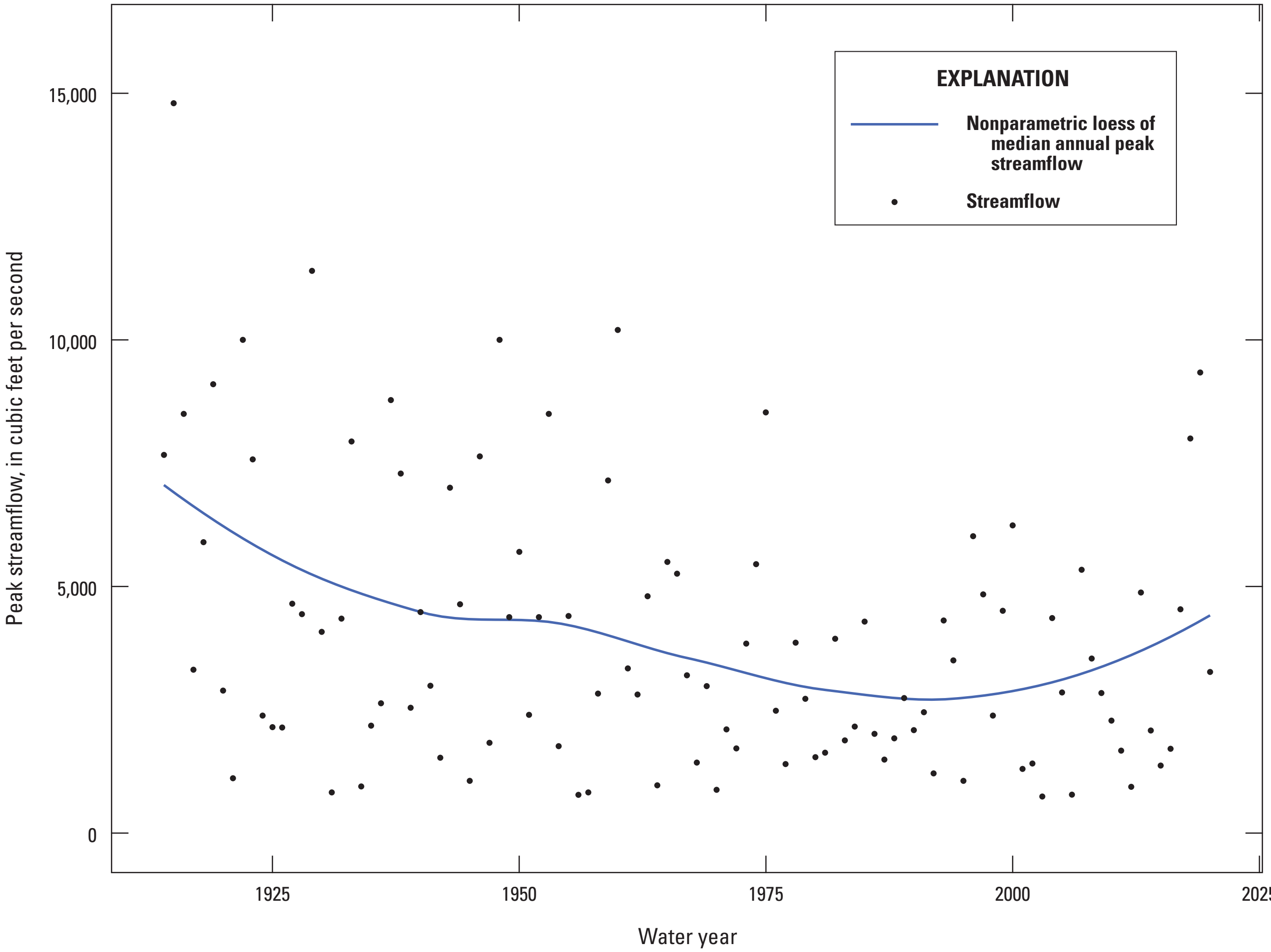


Figure 6. Annual peak streamflow at U.S. Geological Survey streamgage 05436500, Sugar River near Brodhead, WI, from water year 1940 through 2020.

Values for $t_{\left(\frac{\alpha}{2}, n-p\right)}$, *MEV*, and **U** for each regression are available in Levin (2023, table 4). An example of the application of regional regression equations and prediction intervals at an ungaged location is presented in the next section of this report.

Application of Techniques for Estimating Flood Magnitudes at Gaged and Ungaged Locations

The flood frequency estimation methods presented in this study can be applied to three types of rural, unregulated sites. The first case is at a streamgage location; for this case, flood discharge estimates from the LP3 distribution and the regression equations can be combined to for a weighted flood discharge estimate. The second case is at an ungaged location near a streamgage; in this case, the estimated flood discharge at the location of interest is weighted with the estimated flood discharge at the streamgage using the ratio of the two drainage areas. The third case is at an ungaged location that is not near a streamgage; in this case, the regression equation is used to estimate the flood discharge. For each of these three cases, a description of the appropriate method and an example are presented.

Estimating the Weighted Flood Discharge at a Streamgage

Two estimates of flood discharge for a streamgage are available: one from the at-site log-Pearson Type III frequency curve and the other from the appropriate regional regression equation developed in this study. A theoretically improved estimate of flood discharge can be calculated if the individual estimates are assumed to be independent and the variances of the individual estimates are known. If the independent estimates of flood discharge are weighted inversely proportional to their respective variances, then the variance of the weighted-average estimate will be less than the variances associated with each individual estimate (Tasker, 1975; Interagency Advisory Committee on Water Data, 1982).

For a particular AEP, the variance of prediction from the log-Pearson Type III analysis at a streamgage ($V_{P(g)s}$) is estimated by the EMA, as described in Bulletin 17C (England and others, 2019). The magnitude of the variance associated with the at-site LP3 estimate of flood discharge is dependent on the length of record; the mean, standard deviation, and skew of the fitted log-Pearson Type III frequency curve; and the accuracy of the method used to determine the generalized skew (Gotvald and others, 2009). Values for $V_{P(g)s}$ for all streamgages in this study can be found in Levin (2023, table 5) and computed using USGS PeakFQ software version 7.3. (Flynn and others, 2006)

The variances of prediction for flood discharges estimated using the regional regression equations ($V_{P(g)r}$) were computed during the regression fitting process and are dependent on the error covariance matrix and site-specific basin characteristics (Eng and others, 2009). Variances of prediction derived from the regional regression equations were computed using the WREG package in R (Farmer, 2017; Levin, 2023, table 5).

Using the variances from the two independent estimates of flood discharge, the weighted-average estimate of flood discharge is computed using the following equation (Gotvald and others, 2009):

$$\log Q_{P(g)w} = \frac{V_{P(g)r} \times \log Q_{P(g)s} + V_{P(g)s} \times \log Q_{P(g)r}}{V_{P(g)r} + V_{P(g)s}}, \quad (5)$$

where

$Q_{P(g)w}$ is the weighted flood discharge estimate for a *P*-percent AEP at a streamgage, *g*, in ft^3/s;

$Q_{P(g)s}$ is the flood discharge estimate for a *P*-percent AEP at a streamgage, *g*, computed from the at-site LP3 analysis (from Levin, 2023, table 1), in ft^3/s;

$Q_{P(g)r}$ is the flood discharge estimate for the *P*-percent AEP at a streamgage, *g*, computed from the appropriate regional regression equation (in table 2), in ft^3/s;

$V_{P(g)s}$ is the variance of prediction of a flood discharge estimate for the *P*-percent AEP at a streamgage, *g*, from at-site LP3 analysis (in Levin 2023, table 5), in logarithm units; and

$V_{P(g)r}$ is the variance of prediction of a flood discharge estimate for the *P*-percent AEP at a streamgage, *g*, associated with the appropriate regression equation (in Levin, 2023, table 5), in logarithm units.

Confidence intervals for a weighted flood discharge estimate are determined using a weighted variance of prediction computed using the following equations:

$$V_{P(g)w} = \frac{V_{P(g)s} \times V_{P(g)r}}{V_{P(g)s} + V_{P(g)r}}, \quad (6)$$

where

$V_{P(g)w}$ is the variance of prediction for a weighted flood discharge estimate for the *P*-percent AEP, and

$$CI_{90\%} = \left[10^{\left(\log Q_{P(g)w} - 1.65\sqrt{V_{P(g)w}}\right)}, 10^{\left(\log Q_{P(g)w} + 1.65\sqrt{V_{P(g)w}}\right)} \right], \quad (7)$$

where

$CI_{90\%}$ is the 90-percent confidence interval of a weighted flood discharge estimate for a *P*-percent AEP at a streamgage, *g*, in ft^3/s.

Example 1

This example illustrates the calculation of a weighted estimate of flood discharge corresponding to the 1-percent AEP ($Q_{1\%}$) for streamgage USGS 04086200, East Branch Milwaukee River at Kewaskum, WI (map number 80, fig. 1). This discontinued streamgage has 14 years of annual peak flow measurements, from water year 1968 through 1981 and is in flood frequency region 3. The flood discharge estimate from the at-site log-Pearson Type III analysis ($Q_{P(g)s}$) is 862 ft^3/s (Levin, 2023, table 1) and the variance of prediction for the at-site LP3 analysis $V_{P(g)s}$ is 0.0249 (Levin, 2023, table 5). The regression-based variance of prediction is 0.0568 ($V_{P(g)r}$; Levin, 2023, table 5) and the regression-based estimate of flood discharge ($Q_{P(g)r}$) can be computed using the equation in table 2 and basin characteristics for this location:

$$\begin{aligned} \log_{10} Q_{P(g)r} &= 2.310 + 0.757 \times \log_{10} DRNAREA - 0.630 \times \\ &\quad \log_{10}\left(LC01WATER + 1\right) - 0.017 \times WETLAND, \\ &= 2.310 + 0.757 \times 1.73 - 0.630 \times 0.536 - 0.017 \times 24.1 \\ &= 2.87 \\ Q_{P(g)r} &= 10^{2.87} \\ &= 740 \mathrm{ft}^3/\mathrm{s} \end{aligned}$$

A weighted estimate of the 1-percent AEP flood discharge at this streamgage can be computed using equation 5:

$$\log Q_{P(g)w} = \frac{0.0568 \times \log(862) + 0.0249 \times \log(740)}{0.0568 + 0.0249}$$
$$= 2.915$$
$$Q_{P(g)w} = 10^{2.915}$$
$$= 822 \text{ ft}^3/\text{s}$$

The weighted variance for this streamgage is computed using equation 6:

$$V_{P(g)w} = \frac{0.0249 \times 0.0568}{0.0249 + 0.0568}$$
$$= 0.0173$$

The 90-percent confidence interval for the weighted flood discharge estimate at this streamgages is calculated using equation 7:

$$CI_{90\%} = \left[10^{\left(\log Q_{P(g)w} - 1.65\sqrt{V_{P(g)w}}\right)}, 10^{\left(\log Q_{P(g)w} + 1.65\sqrt{V_{P(g)w}}\right)}\right]$$
$$= \left[10^{\left(2.915 - 1.65\sqrt{0.0173}\right)}, 10^{\left(2.915 + 1.65\sqrt{0.0173}\right)}\right]$$
$$= [498, 1355]$$

Estimating the Flood Discharge at an Ungaged Location on a Gaged Stream

For an ungaged location on a gaged stream with 10 or more years of annual peak flow record, the flood discharge estimate from the appropriate regional regression equation can be combined with the weighted-average flood discharge estimate, $Q_{P(g)w}$, from equation 5 and the regression-based flood discharge estimate, $Q_{P(g)r}$, from the nearby streamgage to produce an improved estimate. Sauer (1974) and Verdi and Dixon (2011) presented the following regression-weighted equation to improve the estimate of peak flow frequency for an ungaged location on a gaged stream:

$$Q_{P(u)w} = \left(\frac{2\left|A_g - A_u\right|}{A_g}\right) + \left(1 - \frac{2\left|A_g - A_u\right|}{A_g}\right)\left(\frac{Q_{P(g)w}}{Q_{P(g)r}}\right) \times Q_{P(u)r}, \quad (8)$$

where

- $Q_{P(u)w}$ is the regression-weighted estimate of flood discharge for the P-percent AEP at an ungaged location, u, in ft³/s;
- $Q_{P(g)w}$ is the weighted-average flood discharge estimate for the P-percent AEP at streamgage, g, (from eq. 5), in ft³/s;
- $Q_{P(g)r}$ is the flood discharge estimate for the P-percent AEP at a streamgage from the appropriate regression equation (from table 2), in ft³/s;
- $Q_{P(u)r}$ is the flood discharge estimate for the P-percent AEP at an ungaged location from the appropriate regression equation (table 2), in ft³/s;
- A_g is the drainage area associated with the streamgage, in square miles; and
- A_u is the drainage area associated with the ungaged location, in square miles.

If the drainage area associated with the ungaged location is between 50 and 150 percent of the drainage area associated with the streamgage, equation 8 is applicable. If the drainage area associated with the ungaged location is less than 50 or greater than 150 percent of the drainage area associated with the streamgage, then flood discharge at the ungaged location should be estimated using the appropriate regression equation from table 2 without weighting it with the streamgage estimate.

Example 2

This example illustrates the calculation of a regression-weighted estimate for the 1-percent AEP flood discharge for a stream in region 3 that is directly upstream of streamgage 04086200, East Branch Milwaukee River at Kewaskum, Wis. (map number 80, fig. 1). The following basin characteristics for this ungaged location were obtained from the StreamStats web application for Wisconsin (U.S. Geological Survey, 2016): drainage area = 35.1 mi², percentage of open water = 2.49 percent, and percentage of wetlands = 20.97 percent. The drainage area for the upstream gage (A_g) is 53.90 mi². The weighted average estimate of the 1-percent AEP flood discharge for the upstream gage ($Q_{P(g)w}$) and the regression-based estimate of the 1-percent AEP flood discharge ($Q_{P(g)r}$) were computed as 822 ft³/s and 740 ft³/s, respectively (see section "Example 1"). The regression estimate for the 1-percent AEP flood discharge for the ungaged location ($Q_{P(u)r}$), in ft³/s, can be computed using the basin characteristics at the ungaged location and the appropriate equation from table 2:

$$\log_{10} Q_{P(u)r} = 2.310 + 0.757 \times \log_{10} DRNAREA - 0.630 \times$$
$$\log_{10}(LC01WATER + 1) - 0.0174 \times WETLAND,$$
$$= 2.310 + 0.757 \times \log_{10} 35.1 - 0.630 \times$$
$$\log_{10}(2.49 + 1) - 0.0174 \times 20.97$$
$$= 2.77$$
$$Q_{P(u)r} = 10^{2.77}$$
$$= 588 \text{ ft}^3/\text{s}$$

Finally, the regression-weighted estimate of flood discharge for the 1-percent AEP at the ungaged location (in ft³/s) can be computed using equation 8:

$$Q_{P(u)w} = \left(\frac{2|53.9-35.1|}{53.9}\right) + \left(1-\frac{2|53.9-35.1|}{53.9}\right)\left(\frac{822}{740}\right)\times 588$$
$$= (0.697) + (1-0.697)(1.111)\times 588$$
$$= 608 \text{ ft}^3/\text{s}$$

Estimating the Flood Discharge at an Ungaged Location without a Nearby Streamgage

Flood discharge estimates at ungaged locations that are not near a streamgage are calculated using the appropriate regional regression equations from table 2. Confidence intervals for such estimates can be computed using equation 2.

Example 3

This example illustrates the calculation of the 1-percent AEP flood discharge and 90-percent confidence intervals at an ungaged stream in Wisconsin. For this example, an ungaged location was selected in flood frequency region 3 and basin characteristics were computed with the StreamStats web application (U.S. Geological Survey, 2016). This site has a drainage area of 23.7 mi², 0 percent open water, and 6.44 percent wetlands. First, the estimate of the 1-percent AEP flood discharge is estimated using the appropriate equation in table 2:

$$\log_{10}Q_{P(u)r} = 2.310 + 0.757\times\log_{10}DRNAREA - 0.630\times\log_{10}(LC01WATER+1) - 0.0174\times WETLAND,$$
$$= 2.310 + 0.757\times\log_{10}23.7 - 0.630\times\log_{10}(0+1) - 0.0174\times 6.44$$
$$= 3.23$$
$$Q_{P(u)w} = 1{,}698 \text{ ft}^3/\text{s}$$

Next, the 90-percent confidence interval for the estimate is computed using equations 2–4. This is done in 6 steps:

1. Compute the vector ($\mathbf{X_i}$) of log-transformed basin characteristics. The vector elements should be in the same order as they are listed in Levin (2023, table 4), using 1 for the intercept term:

$$\mathbf{X_i} = \{1,\ 6.44,\ \log_{10}(0+1),\ \log_{10}(23.7)\}$$
$$= \{1,\ 6.44,\ 0,\ 1.37)\}$$

2. Find the covariance matrix for the regression coefficients (**U**) from the appropriate equation in Levin (2023, table 4):

Variable	Intercept	*WETLAND*	*LC01WATER*	*DRNAREA*
Intercept	0.00549	−0.00008	−0.00065	−0.00167
WETLAND	−0.00008	0.00002	0.00004	−0.00010
LC01WATER	−0.00065	0.00004	0.01713	−0.00289
DRNAREA	−0.00167	−0.00010	−0.00289	0.00204

3. To compute the $\mathbf{X_i U X_i^T}$ term in equation 4, first perform matrix multiplication of $\mathbf{X_i}$ and **U** to get $\mathbf{X_i U}$ and then multiply $\mathbf{X_i U}$ and $\mathbf{X_i^T}$:

$$\mathbf{X_i U} = \{0.002692, -0.00006035, -0.004352, 0.0004679\}$$
$$\mathbf{X_i U X_i^T} = 0.00294$$

4. Obtain the model error variance (*MEV*) for the 1-percent AEP regression equation from Levin (2023, table 4) and compute $SE_{p,i}$ using equation 4:

$$SE_{p,1} = [MEV + \mathbf{X_i U X_i^T}]^{0.5}$$
$$= (0.0494 + 0.00294)^{0.5}$$
$$= 0.2288$$

5. Compute *C* using equation 3. The critical value can be obtained for each region in Levin (2023, table 4). In this case, the critical value is 1.6736:

$$C = 10^{t_{\left(\frac{\alpha}{2},n-p\right)} SE_{p,i}}$$
$$= 10^{(1.6736*0.2288)}$$
$$= 2.415$$

6. The 90-percent prediction interval is computed from equation 2 as:

$$(1{,}698/2.415) < Q < (1{,}698\times 2.415)$$
$$703 \text{ ft}^3/\text{s} < Q < 4{,}100 \text{ ft}^3/\text{s}$$

Web Application for Solving Regional Regression Equations

The USGS StreamStats web application incorporates the new peak flow frequency regression equations for Wisconsin and provides flood discharge estimates for unregulated streams in the basin (U.S. Geological Survey, 2016). The web application includes (1) a mapping tool to specify a location on a stream where peak flow statistics are desired; (2) a database that includes peak flow frequency statistics, hydrologic characteristics, location, and descriptive information for all USGS streamgages used in this study; and (3) an automated Geographic Information System procedure that measures the required basin characteristics and solves the regression equations to estimate flood frequency statistics for user-selected locations.

Summary

This study updates the regional regression equations that are used to estimate the magnitude of annual peak streamflows corresponding to the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities for nonurbanized, unregulated streams in Wisconsin. Estimates of flood discharge were computed at 299 streamgages in Wisconsin using the expected moments algorithm (EMA) to fit a log-Pearson type III frequency distribution and regional skew values that were developed previously. The EMA method addresses several methodological concerns identified with the previous procedures for determining flood frequency outlined in Bulletin 17B. Specifically, the EMA method can accommodate censored values, which are common at crest-stage gages, and has improved statistical treatment of potentially influential low floods.

A cluster analysis, using basin characteristics at streamgage locations, was used to delineate four new flood frequency regions in Wisconsin. The flood frequency regions developed from the clustering analysis were affected by north-south gradients of precipitation, snowfall, and patterns in land cover such as percent forest, wetlands, and open water and divide Wisconsin roughly into central, northern, southeastern, and southwestern regions. Regions were selected such that homogeneity of basin characteristics within the groups was maximized while retaining a minimum of at least 40 streamgages in each region.

Regression equations were developed for each flood frequency region by relating basin characteristics at streamgages in the region to the log-Pearson Type III distribution flood discharge estimates using generalized least-squares regression. Redundancy and trend analyses were performed to identify and remove streamages from the analysis that may violate assumptions of the GLS regression. Basin characteristics that were statistically significant in the equations included drainage area (*DRNAREA*), saturated hydraulic conductivity (*SSURGOKSAT*), percentage of open water (*LC01WATER*), percentage of wetlands (*WETLAND*), percentage of forest (*FOREST*), percentage of herbaceous upland area (*LC01HERB*), and the maximum 24-hour precipitation with a 10-year return period (*I24H10Y*). Resulting regression equations had standard errors of prediction ranging from 40.0 to 71.2 percent.

References Cited

Conger, D.H., 1971, Estimating magnitude and frequency of floods in Wisconsin: U.S. Geological Survey Open-File Report 91–2148, 200 p. [Also available at https://doi.org/10.3133/ofr7176.]

Conger, D.H., 1981, Techniques for estimating magnitude and frequency of floods for Wisconsin streams: U.S. Geological Survey Open-File Report 82–1214, 115 p. [Also available at https://doi.org/10.3133/ofr801214.]

Douglas, E.M., Vogel, R.M., and Kroll, C.N., 2000, Trends in floods and low flows in the United States—Impact of spatial correlation: Journal of Hydrology, v. 240, nos. 1–2, p. 90–105. [Also available at https://doi.org/10.1016/S0022-1694(00)00336-X.]

Eash, D.A., Barnes, K.K., and Veilleux, A.G., 2013, Methods for estimating annual exceedance-probability discharges for streams in Iowa, based on data through water year 2010: U.S. Geological Survey Scientific Investigations Report 2013–5086, 63 p. [Also available at https://pubs.usgs.gov/sir/2013/5086/.]

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., 2019, Guidelines for determining flood flow frequency—Bulletin 17C (ver. 1.1, May 2019): U.S. Geological Survey Techniques and Methods, book 4, chap. B5, 148 p. [Also available at https://doi.org/10.3133/tm4B5.]

Farmer, W., 2017, USGS–R/WREG (ver. 2.02): U.S. Geological Survey software, R package.[Also available at https://rdrr.io/github/USGS-R/WREG/.]

Flynn, K.M., Kirby, W.H., and Hummel, P.R., 2006, User's manual for program 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.]

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://pubs.usgs.gov/tm/tm4a8.]

Ericson, D.W., 1961, Floods in Wisconsin, magnitude and frequency, U.S. Geological Survey Open-File Report 61–44, 109 p. [Also available at https://doi.org/10.3133/ofr6144.]

Gebert, W.A., Garn, H.S., and Rose, W.J., 2016, Changes in streamflow characteristics in Wisconsin as related to precipitation and land use: U.S. Geological Survey Scientific Investigations Report 2015–5140, 23 p. [Also available at https://doi.org/10.3133/sir20155140.]

Gotvald, A.J., Feaster, T.D., and Weaver, J.C., 2009, Magnitude and frequency of rural floods in the southeastern United States, 2006—Volume 1, Georgia: U.S. Geological Survey Scientific Investigations Report 2009–5043, 120 p. [Also available at https://doi.org/10.3133/sir20095043.]

Gyawali, R., Greb, S., and Block, P., 2015, Temporal changes in streamflow and attribution of changes to climate and landuse in Wisconsin watersheds: Journal of the American Water Resources Association, v. 51, no. 4, p. 1138–1152. [Also available at https://doi.org/10.1111/jawr.12290.]

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources: U.S. Geological Survey Techniques of Water-Resource Investigations Report, book 4, chap. A3, 523 p. [Also available at https://doi.org/10.3133/tm4A3.]

Hu, L., Nikolopoulos, E.I., Marra, F., and Anagnostou, E.N., 2020, Sensitivity of flood frequency analysis to data record, statistical model, and parameter estimation methods—An evaluation over the contiguous United States: Journal of Flood Risk Management, v. 13, no. 1, p. e12580. [Also available at https://doi.org/10.1111/jfr3.12580.]

Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency: Reston, Va., Hydrology Subcommittee Bulletin 17B, 28 p. and appendixes. [Also available at http://water.usgs.gov/osw/bulletin17b/dl_flow.pdf.]

Juckem, P.F., Hunt, R.J., Anderson, M.P., and Robertson, D.M., 2008, Effects of climate and land management change on streamflow in the driftless area of Wisconsin: Journal of Hydrology, v. 355, no. 1–4, p. 123–130. [Also available at https://doi.org/10.1016/j.jhydrol.2008.03.010.]

Kassambara, A., and Mundt, M., 2020, factoextra—Extract and visualize the results of multivariate data analyses: R package version 1.0.7, accessed September 1, 2021, at https://CRAN.R-project.org/package=factoextra.

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. [Also available at https://doi.org/10.3133/sir20195018.]

Krug, W.R., Conger, D.H., and Gebert, W.A., 1992, Flood-frequency characteristics of Wisconsin streams: U.S. Geological Survey Water-Resources Investigations Report 91–4128, 185 p. [Also available at https://doi.org/10.3133/wri914128.]

Levin, S.B., 2022, PeakFQ inputs and outputs for 299 streamgages in Wisconsin through water year 2020: U.S. Geological Survey data release, https://doi.org/10.5066/P9BDHH6D.

Levin, S.B., 2023, Model archive—Regional regression models for estimating flood frequency characteristics of unregulated streams in Wisconsin: U.S. Geological Survey data release, https://doi.org/10.5066/P9JVO8QV.

Lichty, R.W., and Karlinger, M.R., 1990, Climate factors for small-basin flood frequency: Journal of the American Water Resources Association, v. 26, p. 577–586. [Also available at https://doi.org/10.1111/j.1752-1688.1990.tb01395.x.]

Lumley, T., 2020, leaps—Regression subset selection (ver. 3.1): R package, accessed September 1, 2021, at https://CRAN.R-project.org/package=leaps.

Parrett, C., Veilleux, A., Stedinger, J.R., Barth, N.A., Knifong, D.L., and Ferris, J.C., 2011, Regional skew for California, and flood frequency for selected sites in the Sacramento–San Joaquin River Basin, based on data through water year 2006: U.S. Geological Survey Scientific Investigations Report 2010–5260, 94 p. [Also available at https://doi.org/10.3133/sir20105260.]

R Core Team, 2021, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing software, accessed September 1, 2021, at https://www.R-project.org/.

Rao, A.R., and Srinivas, V.V., 2008, Regionalization of watersheds—An approach based on cluster analysis: Springer Science + Business Media B.V., accessed September 1, 2021, at https://doi.org/10.1007/978-1-4020-6852-2_3.

Sauer, V.B., 1974, Flood characteristics of Oklahoma streams: U.S. Geological Survey Water-Resources Investigations Report 52–73, 301 p. [Also available at https://doi.org/10.3133/wri7352.]

Splinter, D.K., Chester, R.C., and Dauwalter, D.C., 2015, Trends and change points in Wisconsin's streamflow: Sociological Imagination, v. 51, no. 2, p. 69–95.

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., 1975, Combining estimates of low-flow characteristics of streams in Massachusetts and Rhode Island: Journal of Research of the U.S. Geological Survey, v. 3, no. 1, p. 107–112.

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, v. 111, nos. 1–4, p. 361–375. [Also available at https://doi.org/10.1016/0022-1694(89)90268-0.]

U.S. Geological Survey, 2016, The StreamStats program for Wisconsin: U.S. Geological Survey web page, accessed September 1, 2021, at https://water.usgs.gov/osw/streamstats/wisconsin.html.

U.S. Geological Survey, 2022, Peak streamflow for the Nation: U.S. Geological Survey National Water Information System database, accessed September 1, 2021, at https://doi.org/10.5066/F7P55KJN. [Peak streamflow data directly accessible at https://nwis.waterdata.usgs.gov/usa/nwis/peak.]

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, accessed September 1, 2021, at https://hdl.handle.net/1813/13819.

Verdi, R.J., and Dixon, J.F., 2011, Magnitude and frequency of floods for rural streams in Florida, 2006: U.S. Geological Survey Scientific Investigations Report 2011–5034, 69 p. [Also available at https://doi.org/10.3133/sir20115034.]

Walker, J.F., and Krug, W.R., 2003, Flood-frequency characteristics of Wisconsin streams: U.S. Geological Survey Water-Resources Investigations Report 03–4250, 37 p.

Walker, J.F., Peppler, M.C., Danz, M.E., and Hubbard, L.E., 2017, Flood-frequency characteristics of Wisconsin streams (ver. 2.2, April 2020): U.S. Geological Survey Scientific Investigations Report 2016–5140, 33 p. [Also available at https://doi.org/10.3133/sir20165140.]

For more information about this publication, contact:
Director, USGS Upper Midwest Water Science Center
1 Gifford Pinchot Drive
Madison, WI 53726

For additional information, visit: https://www.usgs.gov/centers/upper-midwest-water-science-center

Publishing support provided by the Rolla Publishing Service Center

Levin and Sanocki—**Estimating Flood Magnitude and Frequency for Unregulated Streams in Wisconsin**—SIR 2022–5118

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20225118