USGS
science for a changing world

Prepared in cooperation with the Federal Emergency Management Agency

Estimating Flood Discharges at Selected Annual Exceedance Probabilities for Unregulated, Rural Streams in Vermont, 2023

Scientific Investigations Report 2025–5088

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

A U.S. Geological Survey hydrologic technician takes a high-water mark measurement in knee-deep water in the Black River in Newport, Vermont, in early July 2023.

Estimating Flood Discharges at Selected Annual Exceedance Probabilities for Unregulated, Rural Streams in Vermont, 2023

By Scott A. Olson

Scientific Investigations Report 2025–5088

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

U.S. Geological Survey, Reston, Virginia: 2025

Acknowledgments

Contents

Figures

Tables

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)
Volume		
cubic foot (ft^3)	28.32	cubic decimeter (dm^3)
cubic foot (ft^3)	0.02832	cubic meter (m^3)
Flow rate		
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

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

$$°C = (°F − 32) / 1.8.$$

Datums

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

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

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

Supplemental Information

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

Abbreviations

AEP   annual exceedance probability

USGS   U.S. Geological Survey

Estimating Flood Discharges at Selected Annual Exceedance Probabilities for Unregulated, Rural Streams in Vermont, 2023

By Scott A. Olson

Abstract

This report provides estimates of flood discharge at selected annual exceedance probabilities (AEPs) for streamgages in and adjacent to Vermont and equations for estimating flood discharges at AEPs of 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent (recurrence intervals of 2-, 5-, 10-, 25-, 50-, 100-, and 500-years, respectively) for ungaged, unregulated, rural streams in Vermont with drainage areas between 0.47 and 851 square miles. The equations were developed using generalized least-squares regression and flood-frequency and drainage-basin characteristics from 156 streamgages. Flood-frequency analyses were completed using data through the 2023 water year. The drainage-basin characteristics used as explanatory variables in the regression equations are drainage area, percentage of wetland area, and basin-wide mean of the average annual precipitation. The average standard errors of prediction used to estimate flood discharges at the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent AEP with these equations are 34.9, 37.1, 38.2, 41.6, 43.8, 46.0, 49.1, and 53.2 percent, respectively.

Flood discharges at selected AEPs for streamgages were computed using the Expected Moments Algorithm. Techniques used to adjust an AEP discharge computed from a streamgage record with results from the regression equations and to estimate flood discharge at a selected AEP for an ungaged site upstream or downstream from a streamgage using a drainage-area adjustment are both described.

The final regression equations and the flood-discharge frequency data used in this study will be available in StreamStats. StreamStats is an internet-based application that provides automated regression-equation solutions for user-selected sites on streams.

Introduction

Flooding is the costliest natural hazard experienced in Vermont. Flooding in Vermont has been caused by a variety of factors such as intense precipitation, a series of closely spaced major storms, springtime storms combined with snowmelt, tropical storms, and ice jams. Floods in Vermont tend to be localized; rarely do floods in Vermont have the same severity statewide. Since systematic monitoring of Vermont streams and their floods began in the early 1900s, flood discharges with an annual exceedance probability (AEP) of less than 2 percent, which is equivalent to having a recurrence interval greater than 50 years, have happened in parts of the State in 1927, 1936, 1938, 1973, 1982, 1984 (Hammond, 1991), 1995, 1996, 1997, 1998, 2002, 2011, and 2023.

In response to extensive damage caused by a series of major floods in 1927, 1936, and 1938, flood-control dams and reservoirs were built by the U.S. Army Corps of Engineers in the Winooski and Connecticut River Basins in Vermont to decrease damage caused by the flooding of major rivers. However, flooding continues to be a threat statewide. In 2011, the maximum discharges resulting from Tropical Storm Irene were the greatest recorded at the 26 Vermont streamgages that had more than 5 years of record (Olson and Bent, 2013). Flooding in 2023 resulted in maximum discharges that were the greatest ever recorded at 11 streamgages on rivers and streams in Vermont (Smith and others, 2025).

The U.S. Geological Survey (USGS) and other agencies have been measuring and recording discharge at many streamgages throughout Vermont for over 100 years. One use of the data collected from these streamgages is the ability to characterize the magnitudes and frequencies of flood discharges for rivers in the State. In 2023, 52 continuously operating streamgages and 27 crest-stage gages (only maximum annual discharge is determined) were available on Vermont rivers and streams.

Estimates of the magnitude and frequency of flood discharges are needed to design safe and economical bridges, culverts, and other structures in or near streams; identify flood-hazard areas; and manage floodplains. Computing flood-discharge magnitude and estimating AEP require a statistical analysis of peak-discharge data collected at streamgages. However, estimates of peak-discharge data are often required for ungaged sites where no data are available. Several investigations that provide methods for estimating flood-discharge frequency at ungaged sites have been published, including Potter (1957a, 1957b), Benson (1962), Johnson and Tasker (1974), Dingman and Palaia (1999), Olson (2002; 2014) and Jacobs and Jardin (2010). Updated flood-discharge frequency estimates that benefit from additional years of peak-discharge data can improve techniques used to estimate flood-discharge frequency at ungaged sites. To get these data, the USGS, in cooperation with the Federal Emergency Management Agency, carried out this study to develop updated methods to estimate the flood discharges at selected recurrence intervals for unregulated and ungaged stream locations in and adjacent to Vermont. The methods and equations provided in this report supersede those in previous publications.

Purpose and Scope

This report (1) provides estimates of flood discharges at AEPs of 50-, 20-, 10-, 4-, 2-, 1-, 0.5- and 0.2-percent (recurrence intervals of 2-, 5-, 10-, 25-, 50-, 100-, and 500-years, respectively) for streamgages in and adjacent to Vermont and (2) describes methods, including the use of equations developed from regression analyses, to estimate flood discharges at selected AEPs for rural, ungaged, unregulated Vermont streams. Additionally, this report (3) presents a method to estimate the standard error of prediction for each estimate made using the regression equations and (4) describes methods for transferring a flood-discharge estimate for a selected AEP at a streamgage to a site upstream or downstream on the basis of drainage area.

Description of Study Area

Vermont encompasses 9,250 square miles (mi²) of land area in the northeastern United States, nearly one-eighth of the total land area of New England. The State is about 155 miles long from north to south and ranges from being about 35 miles wide (east to west) at the southern end of the State to being nearly 90 miles wide at its northern end. Vermont is bordered on the northern half of its western boundary by Lake Champlain and is bordered on its eastern border by the Connecticut River. The state of Vermont has more than 7,000 miles of rivers and streams and more than 800 lakes and ponds (Vermont Fish and Wildlife Department, 2025).

Vermont is largely forested and has rolling hills and the more mountainous terrain of the Appalachian Mountains running north-south through much of the center of the State. Land-surface elevations range from approximately 100 feet at the shoreline of Lake Champlain along the northwestern border of Vermont to more than 3,000 feet at numerous peaks. The climate of the region is temperate and humid with four distinct seasons. Precipitation is well distributed throughout the year with averages ranging from about 38 to 45 inches (in.) per year; however, the mountainous character of the State results in a great amount of variability in these ranges of precipitation. High elevation regions can receive over 70 in. of precipitation annually. Annual snowfall also varies across the State. Normal snowfall amounts range from 35 to 60 in. annually, and high elevation locations in the State receive over 100 in. of snowfall annually. The annual mean temperature ranges from 43 to 46 degrees Fahrenheit (°F) throughout the State; however, elevation and aspect can affect these ranges. The average number of days per year the minimum temperature is less than or equal to 0 °F ranges from 10 to 60 days across the State (Office of the Vermont State Climatologist, 1996).

Estimating Flood Discharges at Selected Annual Exceedance Probabilities for Streamgages

To develop techniques to estimate flood-discharge magnitudes and frequencies for ungaged stream locations, flood-discharge magnitude and frequency are first computed at long-term streamgages for which sufficient unregulated annual peak-discharge information is available. The magnitude and frequency of floods at these streamgages then can be statistically related to the physical and climatic characteristics of the contributing drainage basin (fig. 1) upstream from the streamgage (drainage-basin characteristics). The statistical relations that are established at the streamgages then can be used to estimate the magnitudes and frequencies of floods at an unregulated, ungaged site using drainage-basin characteristics.

Peak-Discharge Data Used in this Study

All available annual peak-discharge data for Vermont and adjacent physiographically similar areas in the States of New Hampshire, Massachusetts, and New York and Quebec, Canada, were considered for this study (Olson, 2025). Streamgages within 30 miles of the Vermont border that recorded annual peak-discharge data were considered adjacent. These data include records from continuously recording streamgages and crest-stage streamgages (streamgages that record only the annual peak discharge), both current (2023) and discontinued. This report includes records through water year (WY) 2023. The WY is designated by the calendar year in which it ends. It begins October 1 of the previous calendar year and ends September 30.
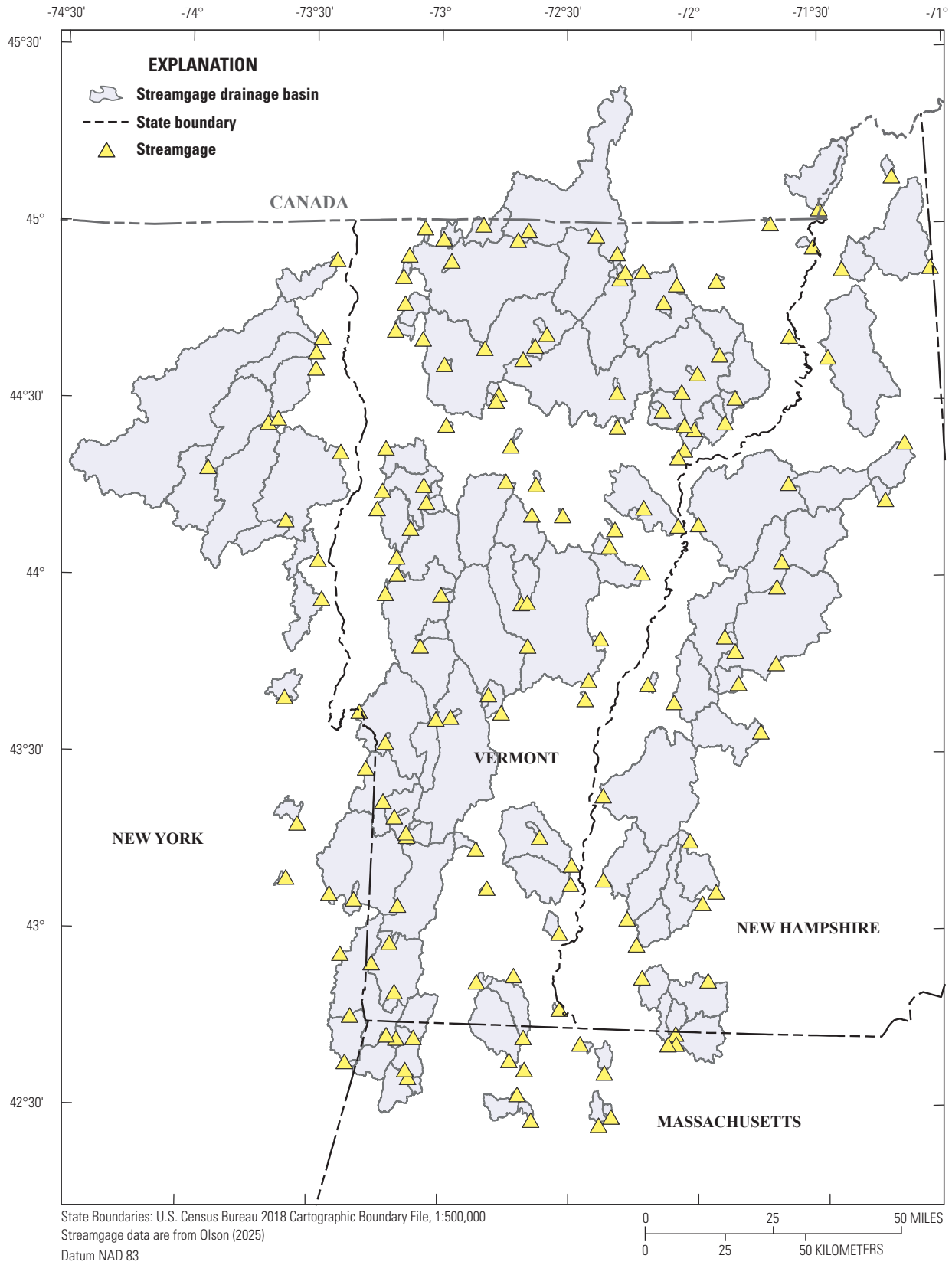
Figure 1.   Map showing the location of selected drainage basins with streamgages used in this investigation in Vermont and its vicinity.

Of the sites considered, 156 streamgages were selected for use in this study (fig. 1; app. 1). The selection criteria required the streamgage to have a minimum of 10 years of annual peak-discharge data that were unaffected by regulation or urbanization. Regulation was assumed to have a negligible effect on peak discharges if the usable storage in the basin was less than 4.5 million cubic feet per square mile of drainage area (Benson, 1962). Peak-discharge data from sites that had greater than 4.5 million cubic feet of usable storage per square mile of drainage area were not used. None of the streamgages used for this investigation have drainage basins considered to be urbanized. The average percentage of urbanized land for basins in this study is 5.5 percent (Commission for Environmental Cooperation, 2023). The streamgages selected for this study are spatially well distributed in and adjacent to Vermont (fig. 1).

Peak-Discharge Data Stationarity

In recent decades, stationarity (the assumption that the mean and variability of data from past observations will continue unchanged in the future) of annual peak-discharge data has been challenged (Milly and others, 2008) in light of climatic and land-use changes. To determine whether trends in the annual peak-discharge data exist, a two-sided Kendall Tau trend test (Helsel and Hirsch, 1992) was done. The trend test was done with software called PeakFQ developed by the USGS to analyze peak-flow data (Flynn and others, 2006). Streamgage records with less than 30 years of annual peak-discharge data were not tested for trends because of the uncertainty in defining a trend in a period of record of this length.

Some trends in the peak discharge data from this study were found using the Kendall Tau test. The Kendall Tau statistics indicated an upward trend in annual peak-discharge data (p-value less than or equal to 0.05) for 22 of the 104 streamgages with at least 30 years of record. For an additional two streamgages, the upward trend was considered marginal (p-value less than or equal to 0.1 and greater than 0.05).

For the streamgaging data that indicate a trend, the trend can often be explained as being the result of extreme climatic anomalies near the beginning or end of the peak-discharge record. For instance, for 12 of the 22 streamgages indicating an upward trend, the trend began during the drought of 1960–69 (Hammond, 1991). The evidence of trends did not exist or was statistically insignificant when extreme events near the beginning or end of the record were eliminated. Although some streamgage data indicate trends, the annual peak-discharge data used in this study are analyzed as random, independent events that are homogeneous for a streamgage throughout the period of record. Furthermore, a consensus on best practices to adjust for with trends has not yet emerged (England and others, 2018).

Determining the Magnitude and Annual Exceedance Probabilities of Flood Discharges for Streamgages

Flood discharges with AEPs of 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent (app. 1) were computed for the 156 streamgages using methods outlined in Guidelines for Determining Flood Flow Frequency, Bulletin 17C (England and others, 2018). The Bulletin 17C guidelines incorporate the Expected Moments Algorithm, apply a Multiple Grubbs-Beck test to identify low outliers, provide procedures that accommodate interval and censored data, and compute accurate confidence intervals that account for historical data and regional skew. A summary of the historical, interval, and censored discharge data that was input is provided in appendix 2. Software developed by the USGS to analyze peak-discharge data (PeakFQ version 7.5.1) was used for these computations (U.S. Geological Survey, 2024). The annual peak-flow data that was input to the PeakFQ program were retrieved from the National Water Information System (U.S. Geological Survey, undated). Peak discharges affected by dam failure, ice jam breach, or a similar event are not included in the frequency analyses. There is evidence of mixed populations of annual peak discharges—tropical and non-tropical storm events— at streamgages in Vermont and its vicinity; however, even at streamgages with many decades of annual peak-discharge data insufficient tropical storm events were available to do mixed population analyses. Annual peak discharges through WY 2023 were included in the frequency analysis (Olson, 2025).

Generalized Skew

Estimates of the magnitude and frequency of flood discharges are sensitive to skew—the measure of the lack of symmetry in the probability distribution of annual peak-flow data. Extreme flood events often affect skews computed from a streamgage's peak-discharge record, and the impact of an extreme flood on skew is greater the shorter the length of streamgage record. To compensate for this effect, the skew used to estimate flood discharges for selected recurrence intervals at a streamgage is weighted with a generalized skew estimated by pooling the skews from nearby streamgages. This generalized skew was developed specifically for Vermont (Olson, 2014) and was 0.445 with a standard error of 0.279.

Combined Records of Nearby Streamgages

Three active streamgages in this investigation were a short distance upstream or downstream from a discontinued streamgage. In each case, the location of the discontinued streamgage and the newer, active streamgage are proximate enough to have similar drainage-basin characteristics and to represent a single streamgage peak-flow dataset, warranting the combination of their discharge records. Before combining the peak-discharge record of the discontinued streamgage and peak-discharge record of the newer, active streamgage, the peak discharges from the discontinued streamgage record were adjusted by the drainage area ratio of the two sites. Streamgages that had their peak-discharge records adjusted for drainage area and combined to create a longer period of record are shown in table 1.

Magnitude and Annual Exceedance Probabilities of Flood Discharges for Streamgages

The magnitudes of flood discharges at selected annual exceedance probabilities for streamgages used in this study (Olson, 2025) are listed in appendix 3. The variance of each streamgage's estimated discharge magnitude is reported in appendix 4. The discharges reported in appendix 3 supersede discharges reported in Olson (2014) and Smith and others (2025) because of additional available data and the updated regression equations.

Maximum Recorded Floods

The maximum recorded annual-peak discharges (app. 1) plotted in relation to drainage area for each streamgage in this study are displayed on figure 2. Annual-peak discharges that are affected by dam failure, ice jam breach, or a similar event are not included. A New England regional envelope curve of potential maximum flood discharges developed by Crippen and Bue (1977) is also shown on figure 2 along with a line developed using generalized least-squares regression analysis showing the relation between drainage area and the peak discharge with a 1-percent annual exceedance probability. Figure 2 can be used to evaluate the reasonableness of flood estimates made using techniques described in this report.

Table 1.   Peak-discharge records from discontinued and active streamgages combined to create a longer period of record, Vermont and vicinity.

[Data are from Olson (2025). USGS, U.S. Geological Survey; mi^2, square mile; NH, New Hampshire; VT, Vermont]

Discontinued streamgage				Active streamgage				[a]Drainage area ratio
USGS streamgage number	Streamgage name	Drainage area (mi^2)	Period of record	USGS streamgage number	Streamgage name	Drainage area (mi^2)	Period of record	
01074500	East Branch Pemigewasset River near Lincoln, NH	104	1929–53, 1960, 1968–72	[a]01074520	East Branch Pemigewasset River at Lincoln, NH	115	1993–2023	1.058
01153500	Williams River at Brockways Mills, VT	103	1938, 1941–84	[a]01153550	Williams River near Rockingham, VT	112	1987–2023	1.087
[a]01155000	Cold River at Drewsville, NH	83.5	1941–78	01154950	Cold River at High Street, at Alstead, NH	74.6	2010–23	0.893

[a]Discharges were adjusted to the drainage area of this streamgage, and basin characteristics from this streamgage were used to develop regression equations.

Characteristics of Streamgage Drainage Basins

In flood-frequency regression analysis, the variations in the magnitude of flood discharges at a selected AEP are related to variations in basin characteristics. The flood discharges are the dependent variables, and the basin characteristics are the independent or explanatory variables. For this study, 92 basin characteristics were determined for each streamgage, including physical properties, such as drainage area, channel slope, elevation, forest cover, lake area, soil permeability, and climatic characteristics, such as precipitation and temperature.

Boundaries for each streamgage drainage basin were obtained from StreamStats (Ries and others, 2017). StreamStats is a USGS web application that can delineate drainage basins and determine selected basin characteristics for user selected locations on a stream or river. Dimensional properties of the drainage basins were computed with the geographic information system (GIS) software, ArcGIS Pro (Environmental Systems Research Institute, Inc., undated).

With appropriate GIS datasets, other basin characteristics were also computed with the GIS software. Digital elevation models from the 3D Elevation Program (USGS, 2023) were used to compute basin characteristics associated with elevation. For parts of basins that extended into Canada beyond the coverage of the 3D Elevation Program datasets, digital elevation models from the high-resolution Digital Elevation Model (Natural Resources Canada, 2024) were used. The National Hydrography Dataset (USGS, 2023), the 2020 Land Cover of North America (Commission for Environmental Cooperation, 2023), the 2021 National Land Cover Database (Dewitz, 2023), the State Soil Geographic (STATSGO) database (Schwarz and Alexander, 1995), and the gridded National Soil Survey Geographic (gNATSGO) database (Boiko and others, 2021) were the source GIS datasets for land-surface properties.

The sources for climatic data were the Precipitation Frequency for Northeastern states datasets (National Oceanic and Atmospheric Administration, 2015) and the 1991–2020 PRISM (Parameter-elevation Regressions on Independent Slopes Model; PRISM Climate Group, 2022a and b; 2024a and b) 30-year normals of precipitation and temperature. For parts of basins that extended into Canada beyond the coverage of the PRISM datasets, precipitation data from the 1991–2020 Derived Normal Climate Data provided by Agriculture and Agri-food Canada (2023) supplemented the PRISM data. A complete list of basin characteristics that were tested as explanatory variables in the regression analysis is in appendix 5.

Regression Equations Used to Estimate Flood Discharges at Selected Annual Exceedance Probabilities for Ungaged Stream Sites

Multiple-regression techniques, using generalized least-squares regression (Stedinger and Tasker, 1985), were used to define relations between the flood discharges determined for the streamgages at the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent AEP (dependent variables) and the basin characteristics (independent variables) of those streamgages. Using generalized least-squares regression allows streamgage data to be weighted to compensate for differences in record length and the cross-correlation of concurrent records among streamgages. Furthermore, Stedinger and Tasker (1985) showed that generalized least-squares regression equations are more accurate and provide better estimates of model error than ordinary least-squares regression equations when working with flood frequency.

The regression results provide equations to estimate the values of dependent variables from one or more independent variables. The regression equations take the general form

$$Y_P = b_o + b_1 X_1 + b_2 X_2 + \ldots + b_j X_j, \tag{1}$$

where

Y_P is the magnitude of the flood discharge that has an annual exceedance probability of P percent,

X_1 to X_j are the basin characteristics, and

b_0 to b_j are coefficients developed from the regression analysis.

When transformations to the explanatory and response variables are logarithmic, equation 1 can be manipulated to take the form

$$Y_P = 10^{bo} X_1^{b1} X_2^{b2} \ldots X_j^{bj}, \tag{2}$$

The limitations, sensitivity, and accuracy of the regression equations are reported following the final regression equations. In addition, techniques are discussed in this report that can be used to determine the accuracy and confidence intervals of each individual estimate from the regression equations. Methods of weighting regression equation estimates with streamgage data when the regression equations are used for a site near or at a streamgage are also discussed.

Regression Analysis and Final Regression Equations

A total of 92 basin characteristics were determined for each streamgage and tested for potential use as explanatory variables in the regression equations (app. 5). Mathematical transformations of logarithms and squares were also applied to each basin characteristic and flood-discharge statistic to obtain the most linear relations. Correlation (R Core Team, 2024a) and ordinary least squares regression (Hebbali, 2024) data computed in R (R Core Team, 2024b)—a software environment for statistical computing—were used together to evaluate which basin characteristics, transformed or untransformed, were the most significant explanatory variables while not being correlated with another explanatory variable. This initial evaluation determined that drainage area was a significant explanatory variable when discharge and drainage area were logarithmically transformed (fig. 2). This evaluation also eliminated numerous basin characteristics that were correlated with drainage area.

Next, all combinations of the remaining explanatory variables, using up to four regressors, were evaluated using ordinary least squares (Hebbali, 2024). The evaluations were based upon the R-square, adjusted R-square, root-mean-square error, Mallow's Cp, and the Akaike information criteria of each regression model. This evaluation left 20 potential explanatory variables to be analyzed for the final regression equations.

Finally, generalized least-squares regression techniques were used to analyze and determine the final significant basin characteristics and to compute the final regression equations (Olson, 2025). The generalized least-squares regression analysis was done using the Weighted-Multiple-Linear Regression program (WREG; Farmer, 2021), a hydrologic regression package that works within the R software. The WREG package uses the generalized least-squares regression procedures outlined in Farmer and others (2019). The three basin characteristics used to develop the final regression equations—drainage area, percentage of the basin that is wetlands, and mean annual precipitation—are listed in appendix 6, by streamgage. The final regression equations

EXPLANATION

- - - - $Q_1 = 224A^{0.807}$ (Olson, 2025)
——— Regional envelope curve for maximum flood discharges (Crippen and Bue, 1977)
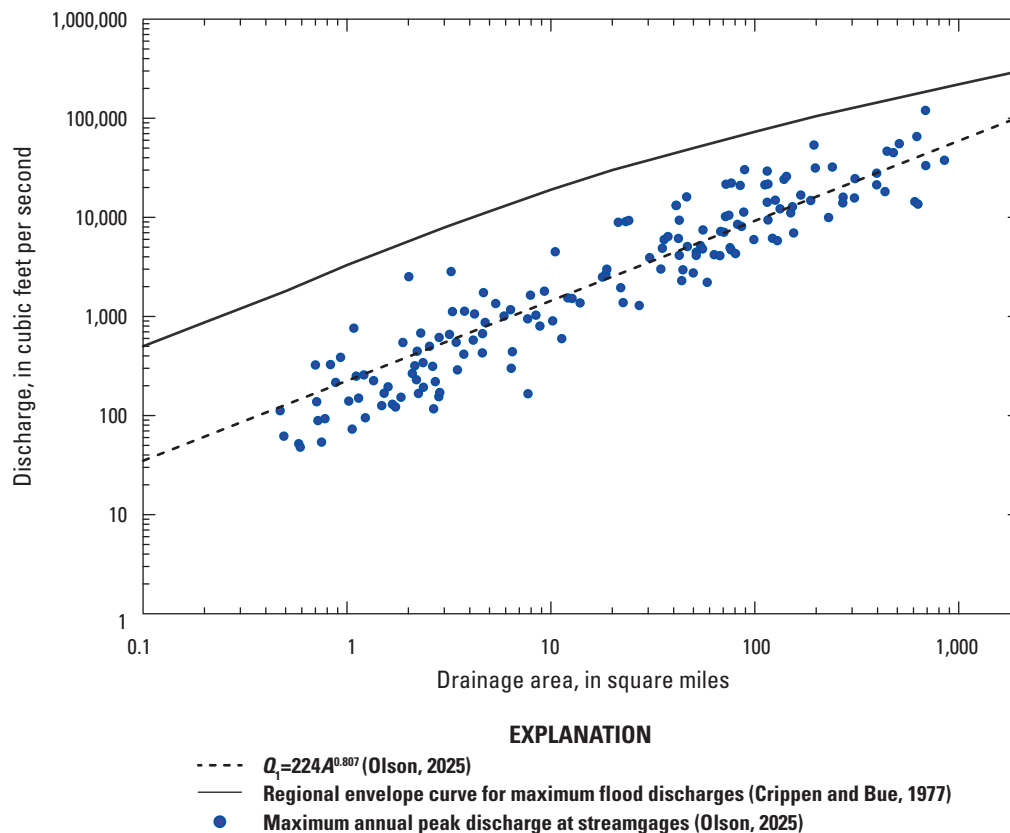● Maximum annual peak discharge at streamgages (Olson, 2025)

Figure 2. Graph showing the maximum recorded annual-peak discharges at streamgages in Vermont and its vicinity in water year 2023 in relation to drainage area, and a regional envelope curve and a regression line relating the 1-percent annual exceedance probability flood discharge (Q_1) to drainage area.

(eqs. 3–10) used to estimate flood discharges on ungaged, unregulated streams in rural drainage basins in Vermont are as follows:

$$Q_{50} = 0.0127A^{0.849}10^{-0.0114W}P^{2.181}, \qquad (3)$$

$$Q_{20} = 0.0157A^{0.830}10^{-0.0110W}P^{2.252}, \qquad (4)$$

$$Q_{10} = 0.0176A^{0.820}10^{-0.0109W}P^{2.293}, \qquad (5)$$

$$Q_{4} = 0.0205A^{0.809}10^{-0.0107W}P^{2.331}, \qquad (6)$$

$$Q_{2} = 0.0229A^{0.802}10^{-0.0107W}P^{2.354}, \qquad (7)$$

$$Q_{1} = 0.0254A^{0.796}10^{-0.0106W}P^{2.375}, \qquad (8)$$

$$Q_{0.5} = 0.0281A^{0.790}10^{-0.0106W}P^{2.393}, \text{ and} \qquad (9)$$

$$Q_{0.2} = 0.0320A^{0.784}10^{-0.0106W}P^{2.414}, \qquad (10)$$

where

Q_p is the estimated flood discharge, in cubic feet per second, at the p-percent annual exceedance probability;

A is the drainage area of the basin, in square miles;

W is the percentage of the basin with land cover categorized as wetlands or open water from the 2020 Land Cover of North America (Commission for Environmental Cooperation, 2023); and

P is the basin-wide mean of the average annual precipitation, in inches, determined with the PRISM 1991–2020 average annual precipitation dataset (PRISM Climate Group, 2022a) supplemented by the 1991–2020 Derived Normal Climate Data (Agriculture and Agri-food Canada, 2023) in basins that extend into Canada.

The procedure for estimating the flood discharge at a selected recurrence interval for an ungaged stream site using the regression equations is described in appendix 7.

Attempts were made to group streamgages with similar geographic or drainage-basin characteristics into subregions to reduce the standard error of the regression equations. This grouping would have resulted in a set of regression equations for each subregion. To evaluate whether subregions should be generated, residuals, the difference between the flood discharges estimated from the frequency analysis and the flood discharges predicted from the regression equations, were determined for each streamgage and for each regression equation. These residuals were plotted spatially at the centroid of the drainage basin of the streamgage and in relation to drainage-basin characteristics. Residuals of the 10- and 1-percent annual exceedance probability regression equations

in relation to drainage-basin characteristics are shown as examples in figures 3A and 3B. No apparent trends or patterns were observed in any of the plots. Thus, the streamgages were not grouped into subregions, and the equations in this report are intended for statewide use and not for subregions.

The residual plots shown in figures 3A and 3B were used as a diagnostic tool for the regression equations. The random scatter of the points above and below the zero-reference line verifies that the model is satisfactorily meeting the assumptions of multiple-linear-regression techniques. Other diagnostics are the evaluation leverage and influence metrics of each observation—the data used to develop the regression equations. Leverage is a measure of the distance an observation is from the centroid of all the other observations giving an indication of whether the observation is unusual compared to the other observations. An influence metric, such as Cook's D (Helsel and Hirsch, 1992), is a measure of the sensitivity of regression parameters to a particular observation (Farmer and others, 2019). The combination of high leverage and influence metrics indicate an observation may be an outlier or require additional analysis. However, after evaluating an observations' leverage and influence metrics, the data were determined to be accurate and there was no justification for excluding any observations from the regression analysis.

Another diagnostic tool used was the variance inflation factor (VIF) (Helsel and Hirsch, 1992). The VIF may be used to evaluate collinearity of explanatory variables. VIF does not have any formal criteria, although some authors suggest that a VIF exceeding 10 may be cause for concern (Freund and Littell, 2000) because explanatory variables may then be correlated. None of the VIF values were close to this level; the greatest VIF computed for a variable used in the final regression equations was 1.2.

Limitations and Sensitivity

Basin characteristics used to develop equations 3 through 9 were determined with ArcGIS Pro (Environmental Systems Research Institute, Inc., undated) software using datasets described in "Regression Analysis and Results" section of this report. This is important because determining the basin characteristics to use in the regression equations with alternate data sources or using different computational methods than those of a GIS software may produce statistics that are different from those reported in this report and may introduce bias and yield discharge estimates that have unknown error.

In addition, although wetland area and annual precipitation values were used in the equations to estimate flood discharges in Vermont in Olson (2014) and in this investigation, the source of these basin characteristics is not the same. Basin characteristics used in the equations published in 2014 cannot be used in these new set of equations, nor vice versa. For instance, the annual precipitation values used in the equations published in 2014—which were 30-year
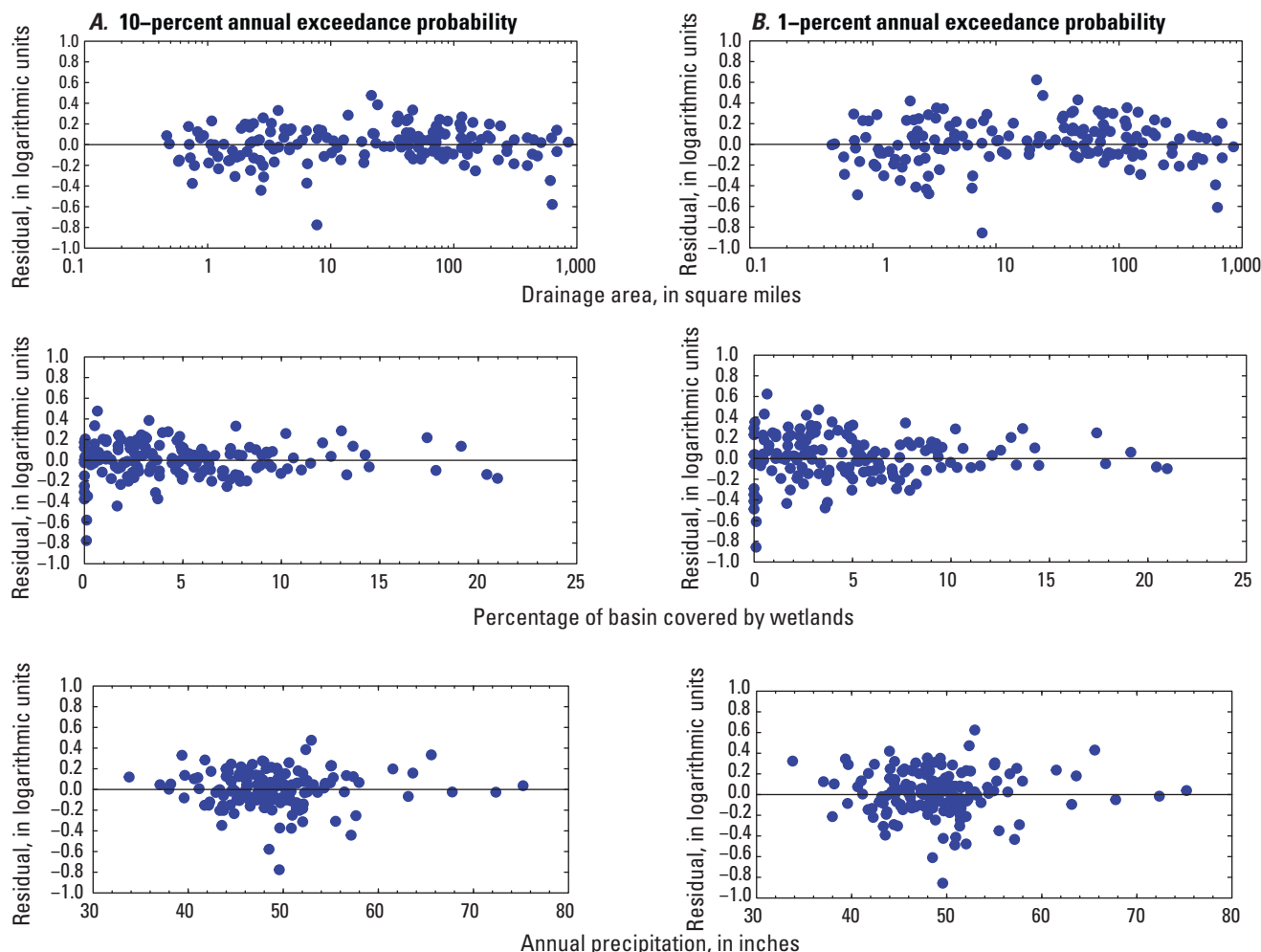
Figure 3.   Graphs showing the residuals of the regression equations used to estimate the magnitude of a discharge with *A,* 10-percent and *B,* 1-percent exceedance probability in relation to basin characteristics in Vermont and its vicinity.

precipitation normals from 1981–2010—should not be applied to the equations developed for this study—which use the 30-year precipitation normals from 1991–2020. Use of basin characteristics from the wrong source will produce unreliable results with unknown error.

The regression equations are applicable to sites on ungaged, unregulated streams in rural basins that are within the region encompassed by the drainage basins used for this investigation. Use of these equations is appropriate for sites with drainage-basin characteristics that are within the range of drainage-basin characteristics used to develop the equations. In general, the uncertainty in the regression equation results will be greater when the explanatory variables used in the equations are further from their mean. The mean and ranges of drainage-basin characteristics used for the analysis are shown in table 2 and figure 4. If the independent variables used in the regression equations are outside of these ranges, the results of the equations are considered extrapolations, and the accuracy of the predictions is unknown. For sites that have drainage-basin characteristics outside the acceptable

Table 2.   Ranges of values for explanatory variables used to develop regression equations to estimate flood discharges at selected annual exceedance probabilities for ungaged, unregulated streams in Vermont and its vicinity.

[Data are from Olson (2025).]

Explanatory variable	Minimum	Maximum	Mean
Drainage area, in square miles	0.47	851	87.7
Percentage of basin covered by wetlands	0	21.0	5.20
Basin-wide mean of the average annual precipitation in inches	33.8	75.2	49.0
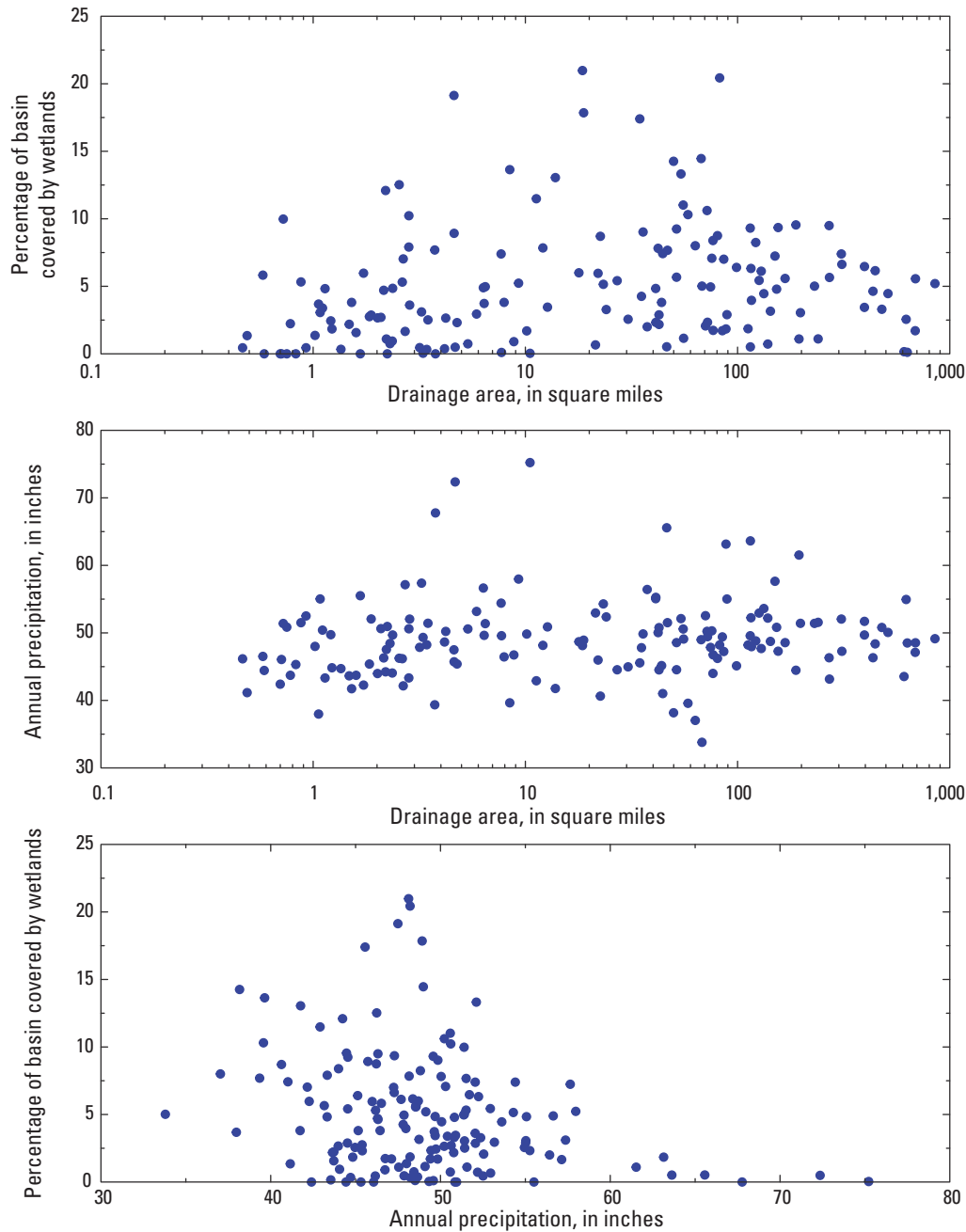
Figure 4.   Graphs showing two-dimensional ranges of explanatory variables used to develop the regression equations that can be used to estimate flood discharges at selected annual exceedance probabilities for streams in Vermont and its vicinity.

ranges, a simplified equation that uses drainage area as the only explanatory variable is used for this study and is provided in this report in the section titled "Drainage-Area-Only Regression Equations." For sites that are considered urban, Moglen and Shivers (2006) describe techniques that transform rural flood-discharge frequency estimates to estimates for urban watersheds.

The sensitivity of each regression equation to changes in the magnitude of the independent variables was tested to evaluate the amount of error that can be introduced if basin characteristics are incorrectly computed. The sensitivity

analysis was done by adjusting basin characteristic values by plus or minus 10 percent while holding the other basin characteristics constant at their respective mean magnitudes (Olson, 2025). The results of the sensitivity analysis listed in table 3 indicate that the regression equations are most sensitive to changes in drainage area and changes in the basin-wide mean of the average annual precipitation. However, a 10-percent change in the basin-wide mean of the average annual precipitation value accounts for a greater change in the computed flood discharge than does a 10-percent change in drainage area.

Accuracy of the Regression Equations

The accuracy of regression equations can be measured in several ways. The coefficient of determination, R^2, indicates the proportion of variability of the dependent variable and is explained by combining the explanatory variable in the regression model. The pseudo coefficient of determination, R^2_{pseudo}, can be described in the same way but accounts for the effect of the time-sampling error (Farmer and others, 2019). The closer the R^2 or R^2_{pseudo} is to 1, the better the regression explains the variation in the dependent variables. The R^2 and R^2_{pseudo} for each regression equation (Olson, 2025) is in table 4.

One of the most common measures of accuracy is the root mean square error (table 4); it measures how much the regression results deviate from the observed data. The root mean square error represents the standard deviation of the residuals.

Another measure of accuracy is the average standard error of prediction (table 4). The average standard error of prediction has a model error component—the error resulting from the model—and a sampling error component—the error that results from model parameters that are developed from samples of the population. Thus, the average standard error of prediction is a measure of the expected accuracy of a regression model when applied at an ungaged location with basin characteristics like those used to develop the regression equation. This measure of accuracy is needed because regression equations are typically used for ungaged locations. About two-thirds of the estimates made using a regression equation for ungaged locations will have errors less than the average standard error of prediction for that equation.

Accuracy Analysis of Individual Estimates from the Regression Equations

Because the coefficient of determination, the pseudo coefficient of determination, the root mean square error, and the average standard error of prediction (table 4) are computed using available streamgage data, they are approximations of the overall accuracy of the regression equations for ungaged sites. Techniques for computing the accuracy of individual regression equation estimates for ungaged sites are available and are discussed in this section. Individual estimates are measured for accuracy by means of the standard error of prediction and prediction intervals.

Standard Error of Prediction

Hodge and Tasker (1995) describe how to compute the variance of prediction (V_{pred}) of a flood-discharge frequency estimate as

$$V_{pred} = \gamma^2 + x_i (X^{tr} \Lambda^{-1} X)^{-1} x_i^{tr}, \qquad (11)$$

where

V_{pred}	is the standard error of prediction;
γ^2	is the model error variance (refer to table 5);
x_i	is a row vector containing 1, $\log_{10}(A)$, W, and $\log_{10}(P)$ for the study site i;
tr	is the matrix algebra symbol for transposing a matrix; and
$(X^{tr}\Lambda^{-1}X)^{-1}$	is the ($p \times p$) matrix with X being a ($n \times p$) matrix that has rows of logarithmically transformed basin characteristics augmented by one, and Λ being the ($n \times n$) covariance matrix used for weighting sample data in the generalized least-squares regression; n is the number of streamgages used in the regression analysis, and p is the number of basin characteristics plus 1. The $(X^{tr}\Lambda^{-1}X)^{-1}$ matrices for selected recurrence intervals are shown in table 5.

The variance of prediction can be used to weight a result from a regression equation with the AEP from the analysis of the streamgage record and is discussed in the "Use of Regression Equations at or near Streamgages" section of this report. A streamgage's weighted AEP result will have reduced uncertainty (U.S. Interagency Advisory Committee on Water Data, 1982).

The standard error of prediction is computed with the following formula:

$$SE_{pred} = (V_{pred})^{1/2}, \qquad (12)$$

where

SE_{pred}	is the standard error of prediction, and
V_{pred}	is the variance of prediction.

The standard error of prediction for an estimate can be converted to positive and negative percent errors with the following formulas:

$$S_{pos} = 100(10^{SEpred} - 1), \text{ and} \qquad (13)$$

$$S_{neg} = 100(10^{-SEpred} - 1), \qquad (14)$$

where

S_{pos}	is the positive percent error of prediction,
SE_{pred}	is the standard error of prediction in logarithmic units, and
S_{neg}	is the negative percent error of prediction.

The probability that the true value of flood discharge at a given frequency is between the positive- and negative-percent standard error of prediction is approximately 68 percent. For example, if S_{neg} is −27.1 percent and S_{pos} is 37.1 percent, there is a 68-percent chance that the true AEP discharge at a site ranges from −27.1 percent to +37.1 percent of the estimated AEP discharge.

Table 3.   Results of the sensitivity analysis calculated from regression equations for Vermont and its vicinity as percentage change in computed flood discharge because of a 10-percent change in the input basin characteristic.

[Data are from Olson (2025).]

Percent change of basin characteristic	Percent change of computed flood discharge by the regression equation for estimated floods with an annual exceedance probability of							
	50 percent	20 percent	10 percent	4 percent	2 percent	1 percent	0.5 percent	0.2 percent
Drainage area								
−10	−8.56	−8.37	−8.28	−8.17	−8.10	−8.04	−7.99	−7.93
+10	8.43	8.23	8.13	8.02	7.94	7.88	7.82	7.76
Percent of basin covered by wetlands								
−10	1.38	1.33	1.31	1.29	1.29	1.28	1.28	1.28
+10	−1.36	−1.31	−1.30	−1.27	−1.27	−1.26	−1.26	−1.26
Basinwide mean of the average annual precipitation								
−10	−20.5	−21.1	−21.5	−21.8	−22.0	−22.1	−22.3	−22.5
+10	23.1	23.9	24.4	24.9	25.2	25.4	25.6	25.9

Table 4.   Measures of accuracy of the regression equations for estimating flood discharges at selected annual exceedance probabilities for ungaged, unregulated streams in rural drainage basins in Vermont and its vicinity.

[Data are from Olson (2025). R^2, coefficient of determination; log, logarithmic]

Flood discharge with an annual exceedance probability of	Coefficient of determination (R^2)	Pseudo-R^2	Root mean square error		Average standard error of prediction	
			Log units	Percent	Log units	Percent
50 percent	0.961	0.966	0.157	37.3	0.147	34.9
20 percent	0.956	0.961	0.166	39.6	0.156	37.1
10 percent	0.950	0.958	0.176	42.3	0.160	38.2
4 percent	0.939	0.950	0.193	46.6	0.173	41.6
2 percent	0.931	0.944	0.206	50.1	0.182	43.8
1 percent	0.921	0.938	0.219	53.9	0.190	46.0
0.5 percent	0.911	0.930	0.233	57.9	0.202	49.1
0.2 percent	0.896	0.918	0.252	63.4	0.217	53.2

Table 5.   Flood-frequency characteristic, model error variance, and the basin characteristics matrices for the regression equations.

[Data are from Olson (2025). Numbers in matrices are in scientific notation. Annual precipitation refers to the basin-wide mean of the average annual precipitation. AEP, annual exceedance probability]

Flood-frequency characteristic	Model error variance (γ^2)	Basin characteristics	$(X^{tr}\Lambda^{-1}X)^{-1}$ matrix			
			Intercept	Drainage area	Percent wetland	Annual precipitation
Flood discharge with a 50-percent AEP	0.0208	Intercept	1.94983E-01	7.30781E-04	−4.41280E-04	−1.14432E-01
		Drainage area	7.30781E-04	2.51035E-04	−8.85562E-06	−6.35880E-04
		Percent wetland	−4.41280E-04	−8.85562E-06	8.91120E-06	2.40918E-04
		Annual precipitation	−1.14432E-01	−6.35880E-04	2.40918E-04	6.75237E-02
Flood discharge with a 20-percent AEP	0.0232	Intercept	2.30064E-01	8.06082E-04	−5.20878E-04	−1.34917E-01
		Drainage area	8.06082E-04	3.02311E-04	−9.49728E-06	−7.34861E-04
		Percent wetland	−5.20878E-04	−9.49728E-06	1.04110E-05	2.83456E-04
		Annual precipitation	−1.34917E-01	−7.34861E-04	2.83456E-04	7.95976E-02
Flood discharge with a 10-percent AEP	0.0244	Intercept	2.56046E-01	8.60343E-04	−5.77774E-04	−1.50080E-01
		Drainage area	8.60343E-04	3.39409E-04	−9.79067E-06	−8.09035E-04
		Percent wetland	−5.77774E-04	−9.79067E-06	1.14738E-05	3.13703E-04
		Annual precipitation	−1.50080E-01	−8.09035E-04	3.13703E-04	8.85363E-02
Flood discharge with a 4-percent AEP	0.0285	Intercept	3.14624E-01	1.04050E-03	−7.08351E-04	−1.84372E-01
		Drainage area	1.04050E-03	4.16219E-04	−1.13144E-05	−9.91203E-04
		Percent wetland	−7.08351E-04	−1.13144E-05	1.39974E-05	3.83963E-04
		Annual precipitation	−1.84372E-01	−9.91203E-04	3.83963E-04	1.08765E-01
Flood discharge with a 2-percent AEP	0.0313	Intercept	3.57565E-01	1.17197E-03	−8.03846E-04	−2.09505E-01
		Drainage area	1.17197E-03	4.72399E-04	−1.23400E-05	−1.12540E-03
		Percent wetland	−8.03846E-04	−1.23400E-05	1.58338E-05	4.35272E-04
		Annual precipitation	−2.09505E-01	−1.12540E-03	4.35272E-04	1.23592E-01
Flood discharge with a 1-percent AEP	0.0341	Intercept	4.01967E-01	1.31020E-03	−9.02765E-04	−2.35494E-01
		Drainage area	1.31020E-03	5.30134E-04	−1.33984E-05	−1.26507E-03
		Percent wetland	−9.02765E-04	−1.33984E-05	1.77339E-05	4.88428E-04
		Annual precipitation	−2.35494E-01	−1.26507E-03	4.88428E-04	1.38923E-01
Flood discharge with a 0.5-percent AEP	0.0384	Intercept	4.60446E-01	1.50556E-03	−1.03438E-03	−2.69754E-01
		Drainage area	1.50556E-03	6.04730E-04	−1.50844E-05	−1.45116E-03
		Percent wetland	−1.03438E-03	−1.50844E-05	2.02861E-05	5.59438E-04
		Annual precipitation	−2.69754E-01	−1.45116E-03	5.59438E-04	1.59136E-01
Flood discharge with a 0.2-percent AEP	0.0443	Intercept	5.40529E-01	1.77457E-03	−1.21500E-03	−3.16671E-01
		Drainage area	1.77457E-03	7.06616E-04	−1.73758E-05	−1.70659E-03
		Percent wetland	−1.21500E-03	−1.73758E-05	2.37838E-05	6.56876E-04
		Annual precipitation	−3.16671E-01	−1.70659E-03	6.56876E-04	1.86815E-01

Prediction Intervals

Prediction intervals indicate the uncertainty in the equation results. For example, there is a 90 percent probability that the true value of a flood-discharge estimate lies within the 90-percent prediction interval. Prediction intervals for selected percentages can be computed as follows:

$$PI_{upper} = Q_{pred}10^{\left(t_{\alpha/2,n-p}SE_{pred}\right)} \tag{15}$$

$$PI_{lower} = \frac{Q_{pred}}{10^{\left(t_{\alpha/2,n-p}SE_{pred}\right)}} \tag{16}$$

where

PI_{upper} is the upper prediction interval, in cubic feet per second;

PI_{lower} is the lower prediction interval, in cubic feet per second;

Q_{pred} is the computed discharge at a selected frequency from the regression equation, in cubic feet per second;

$t_{\alpha/2,n-p}$ is the critical value from a Student's-t distribution at alpha level α (α=0.10 for a 90-percent prediction interval; α=0.05 for a 95-percent prediction interval) with *n-p* degrees of freedom; *n*=156, the number of stations used in the regression analysis; and *p*=4, the number of basin characteristics in the regression equation, plus 1 (for the 95- and 90-percent prediction interval $t_{\alpha/2,n-p}$ is 1.98 and 1.66, respectively); and

SE_{pred} is the standard error of prediction of a flood-discharge frequency estimate.

Use of Regression Equations for Streamgages

An estimate of flood discharge at a selected annual exceedance probability made at a streamgage can be improved by combining regression equation results with the frequency curve computed from the streamgage record. The procedure recommended by Cohn and others (2012) is to compute a flood discharge using the regression equation estimate and the result of the frequency analysis of the streamgage record for a given annual exceedance probability weighted by the inverse of the variance of each of the discharge estimates. The procedure was applied to all the streamgages used in this study, and the weighted flood-discharge results are in appendix 3. Generally, the weighted flood discharge estimate provides better estimates of the true discharges than those determined from either the flood-frequency analysis or the regression analysis alone. The weighted discharges were computed using the following equation:

$$\log_{10}Q_w = \frac{\log_{10}Q_s\left(V_{pred}\right)+\log_{10}Q_{r(g)}\left(V_s\right)}{V_{pred}+V_s} \tag{17}$$

where

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

Q_s is the flood discharge for the selected annual exceedance probability computed from the streamgage record, in cubic feet per second,

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

V_{pred} is the variance of prediction for the result of the regression equation ($Q_{r(g)}$) in logarithmic units computed using equation 11, and

V_s is the variance in the estimate of annual exceedance probability discharge given in logarithmic units computed from the streamgage record (Q_s) (see app. 4).

Prediction intervals for the weighted discharge (Q_w) can be computed with equations 15 and 16; however, the standard error of prediction for the weighted discharge (SE_w) is to be substituted for the standard error of prediction for the regression estimate (SE_{pred}). The standard error of prediction for the weighted discharge can be computed with the following formula:

$$SE_w = \sqrt{\frac{V_sV_{pred}}{V_s+V_{pred}}} \tag{18}$$

Use of Regression Equations for Areas near Streamgages

Estimates of flood-discharge magnitude at selected annual exceedance probabilities for ungaged sites that are not at, but are near, a streamgage and are on the same unregulated stream can be improved by combined use of the regression equations and the nearby streamgage data. A method to adjust the weighted discharge for a streamgage (Q_w) developed from equation 17 to a site of interest upstream or downstream from the streamgage is given in this report. This method increases the weight of a discharge estimate derived from a regression-equation over the weight of a discharge estimate derived from streamgage data the farther upstream or downstream the site of interest is from the streamgage. This method is modified from a technique described in Ries (2007) that adjusts the weight linearly in logarithmic units, satisfying the logarithmic relation between flood discharge and drainage area. The logarithmic slope (*c*) of a line that goes through the weighted estimate of the discharge at the streamgage and converges on a location upstream or downstream where full weight will be given to the regression equations at a selected annual exceedance probability is computed as follows:

$$c = \frac{\log_{10}\left(\frac{Q_{r(u)}}{Q_{r(g)}}\right)}{\log_{10}\left(\frac{A_u}{A_g}\right)} + \frac{\log_{10}\left(\frac{Q_{r(g)}}{Q_w}\right)}{\log_{10}(a)},$$ (19)

where

$Q_{r(u)}$ is the flood-discharge estimate generated using the regression equation for the ungaged site, in cubic feet per second;

$Q_{r(g)}$ is the flood-discharge estimate generated using the regression equation for the streamgage, in cubic feet per second;

A_u is the drainage area of the ungaged site, in square miles;

A_g is the drainage area at the streamgage, in square miles;

Q_w is the weighted flood-discharge estimate at the streamgage location computed using equation 17, in cubic feet per second; and

a is the percentage of the gaged drainage area, where full weight is given to the regression equation results. As a rule, $a=0.5$ for A_u less than A_g, and $a=1.5$ for A_u greater than A_g.

The value of a determines where, as a percentage of the gaged drainage area, full weight will be given to the regression equation. Hence, when the site of interest is upstream from the streamgage, $a=0.5$, the site is required to have a drainage area no smaller than 50 percent of the streamgage drainage area. When the site of interest is downstream from the gage, $a=1.5$, the site is required to have a drainage area no larger than 150 percent of the streamgage drainage area.

The final step is to compute the weighted flood-frequency estimate for the ungaged site, Q_u, using

$$Q_u = Q_w \left(\frac{A_u}{A_g}\right)^c$$ (20)

An example that uses this technique is in appendix 7. As with any technique used to compute a weighted flood-discharge estimate, unexpected results could happen if there is a substantial difference between the discharges being weighted. If the difference in discharges being weighted is substantial, c could become negative, indicating discharge and drainage area are inversely related and making the procedure invalid

Drainage-Area-Only Regression Equations

For some ungaged sites, the percentage of the basin covered by wetlands or the basin-wide mean of the average annual precipitation may be outside the acceptable ranges required for the full regression equations (eqs. 3–10). The acceptable basin characteristic ranges are described in this report in the section titled "Limitations and Sensitivity." In addition, some users of the equations may not have access to basin characteristics beyond drainage area. Because of this, a set of simplified regression equations (Olson, 2025) that incorporate drainage area as the only independent variable was developed. Generalized least-squares regression techniques were used to compute the coefficients in the equations. The simplified regression equations (eqs. 21–28) used to estimate flood discharges from ungaged, unregulated streams in rural drainage basins in Vermont are as follows:

$$Q_{50}=52.6A^{0.854},$$ (21)

$$Q_{20}=85.7A^{0.838},$$ (22)

$$Q_{10}=113A^{0.829},$$ (23)

$$Q_4=152A^{0.819},$$ (24)

$$Q_2=187A^{0.813},$$ (25)

$$Q_1=224A^{0.807},$$ (26)

$$Q_{0.5}=267A^{0.802}, \text{ and}$$ (27)

$$Q_{0.2}=330A^{0.795},$$ (28)

where

Q_P is the estimated flood discharge, in cubic feet per second, at the P-percent exceedance probability; and

A is the drainage area of the basin, in square miles.

The same 156 streamgages used to develop the previous regression equations were used to develop the simplified regression equations. The simplified regression equations are still applicable to sites with drainage areas ranging from 0.47 to 851 mi². Because the equations have only one explanatory variable, they are less accurate than the full regression equations in this report. The coefficient of determination, the pseudo coefficient of determination, the root mean square error, and the average standard error of prediction of the simplified equations are in table 6.

Although drainage-area-only regression equations are less accurate than the multivariable equations, these simplified equations are still valuable. The exponent in each of the drainage-area-only regression equations is the slope of the average linear logarithmic relation between drainage area and flood discharge for a selected AEP. Hence, the exponent can be used as an alternate method to adjust flood-frequency data from a streamgage to flood-frequency data from streamgages upstream and downstream. This use of the method is to be limited to streamgages within 50- to 150-percent of the streamgage drainage area (Wandle, 1983). To use this method, one would use equation 19 with the exponent from the simplified regression equation at a selected recurrence interval substituted for c.

Table 6.   Accuracy measures for the simplified regression equations (drainage area only) that are used to estimate flood discharges at selected annual exceedance probabilities for ungaged, unregulated streams in rural drainage basins in Vermont and its vicinity.

[Data are from Olson (2025). R^2, coefficient of determination; log, logarithmic]

Flood discharge with an annual exceedance probability of	Coefficient of determination (R^2)	Pseudo-R^2	Root mean square error		Average standard error of prediction	
			Log units	Percent	Log units	Percent
50 percent	0.935	0.937	0.202	49.2	0.198	48.2
20 percent	0.928	0.934	0.211	51.5	0.200	48.7
10 percent	0.921	0.928	0.220	54.0	0.207	50.5
4 percent	0.910	0.920	0.233	57.9	0.217	53.2
2 percent	0.900	0.914	0.245	61.2	0.223	55.0
1 percent	0.890	0.905	0.257	64.7	0.233	57.8
0.5 percent	0.880	0.899	0.269	68.4	0.240	59.7
0.2 percent	0.864	0.887	0.286	73.7	0.252	63.3

Vermont StreamStats

StreamStats, an internet-based application (https://www.usgs.gov/streamstats/), allows users to obtain drainage-basin characteristics, discharge statistics, and other information for user-selected sites on streams. StreamStats users choose stream sites of interest from an interactive map. If a user selects the location of a USGS streamgage, they can get previously published information for the site from a database. If a user selects an ungaged site, the StreamStats application will determine the boundary of the drainage basin upstream from the site and measure the basin characteristics required by the regression equations to estimate discharge statistics for the site. The application then solves the equations. The results are then in a table along with a map showing the basin outline.

Furthermore, the application ensures that the basin characteristics input into the regression equations are determined using the same data and methodologies as the basin characteristics used to develop the equations. Doing this avoids bias that could be introduced by improperly estimating basin characteristics.

The equations published in this report will be available online in StreamStats after the publication of this report. StreamStats will provide flood-discharge frequency data for streamgages used in this study and compute flood-discharge frequency estimates for ungaged locations using the final regression equations (eqs. 3–10). StreamStats does not determine flood-discharge frequency estimates for streamgages that are weighted with the regression equations results.

Summary

This report, prepared by the U.S. Geological Survey in cooperation with the Federal Emergency Management Agency, documents how regression equations were developed and used to estimate flood-discharge magnitudes for rural, unregulated streams in Vermont at annual exceedance probabilities of 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent using flood-discharge data through the 2023 water year. Regression techniques were used to determine relations between the flood-discharge magnitudes and selected basin characteristics at 156 streamgages in and adjacent to Vermont.

The flood-discharge magnitudes at selected recurrence intervals for the 156 streamgages were determined following the Guidelines for Determining Flood Flow Frequency, Bulletin 17C. A generalized skew coefficient of 0.445 with a standard error of 0.279 was used in the frequency analysis.

A total of 92 basin characteristics for each streamgage were determined using a geographic information system. Using correlation data, ordinary least-squares regression techniques, and generalized least-squares regression techniques, the 92 basin characteristics were narrowed down to the three variables that best explained the magnitude and variability of flood discharges: the drainage area, the percentage of the basin covered by wetlands, and the basin-wide mean of the average annual precipitation. The final regression equations were developed using generalized least-squares regression techniques. The average standard error of prediction to estimate peak discharges with 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probability with these equations are 34.9, 37.1, 38.2, 41.6, 43.8, 46.0, 49.1, and 53.2 percent, respectively.

The regression equations developed from these relations can be used as a method to estimate flood discharges at selected recurrence intervals for ungaged, unregulated, rural streams. This report also presents methods of how to adjust a flood-discharge frequency curve computed from a streamgage record with results from the regression equations. In addition, a technique is described of how to estimate flood discharge at a selected recurrence interval for an ungaged site upstream or downstream from a streamgage using a drainage-area adjustment.

The equations and flood-discharge frequency data used in this study are available in StreamStats, an internet-based application (https://www.usgs.gov/streamstats/) that provides statistics, drainage-basin characteristics, and other information for user-selected sites on streams.

References Cited

Agriculture and Agri-food Canada, 2023, Derived normal climate data—30-year average precipitation: Natural Resources Canada: Government of Canada web page, accessed January 27, 2025, at https://open.canada.ca/data/en/dataset/acd4c9f2-0598-47e8-aa4d-0a8a0964ce55.

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.

Boiko, O., Kagone, S., and Senay, G.B., 2021, Soil properties dataset in the United States: U.S. Geological Survey data release, accessed July 18, 2024, at https://doi.org/10.5066/P9TI3IS8.

Cohn, T.A., Berenbrock, C., Kiang, J.E., and Mason, R.R., Jr., 2012, Calculating weighted estimates of peak streamflow statistics: U.S. Geological Survey Fact Sheet 2012–3038, 4 p. [Also available at https://pubs.usgs.gov/fs/2012/3038/.]

Commission for Environmental Cooperation, 2023, North America Land Cover, 2020, North American Land Change Monitoring System: Canada Centre for Remote Sensing, U.S. Geological Survey, Ed. 1.0, Raster digital data. [Also available at http://www.cec.org/north-american-environmental-atlas/land-cover-30m-2020/S.]

Crippen, J.R., and Bue, C.D., 1977, Maximum floodflows in the conterminous United States: U.S. Geological Survey Water-Supply Paper 1887, 52 p.

Dewitz, J., 2023, National land cover database (NLCD) 2021 products: U.S. Geological Survey data release, accessed July 18, 2024, at https://doi.org/10.5066/P9JZ7AO3.

Dingman, S.L., and Palaia, K.J., 1999, Comparison of models for estimating flood quantiles in New Hampshire and Vermont: Journal of the American Water Resources Association, v. 35, no. 5, p. 1233–1243. [Also available at https://doi.org/10.1111/j.1752-1688.1999.tb04210.x.]

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, 148p. [Also available at https://doi.org/10.3133/tm4B5.]

Environmental Systems Research Institute, Inc., [undated], ArcGIS Pro: Redlands, Calif., Environmental Systems Research Institute, Inc. web page, accessed February 10, 2025, at https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview.

Farmer, W.H., Kiang, J.E., Feaster, T.D., and Eng, K., 2019, Regionalization of surface-water statistics using multiple linear regression: U.S. Geological Survey Techniques and Methods, book 4, chap. A12, 40 p., accessed February 12, 2025, at https://doi.org/10.3133/tm4A12.

Farmer, W.H., 2021, WREG—Weighted least squares regression for streamflow frequency statistics (ver. 3.0): Reston, Va., U.S. Geological Survey software release, accessed January 31, 2025, at https://doi.org/10.5066/P9ZCGLI1.

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. 4, sec. B, 42 p.

Freund, R., and Littell, R., 2000, SAS system for regression (3rd ed.): Cary, N.C., SAS Institute, Inc., 236 p.

Hammond, R.E., 1991, New Hampshire floods and droughts, in Paulson, R.W., Chase, E.B., Roberts, R.S., Moody, D.W., comp., National water summary 1988–89—Hydrologic events and floods and droughts: U.S. Geological Survey Water-Supply Paper 2375, p. 393–400.

Hebbali, A., 2024, olsrr—Tools for building OLS regression models: R package version 0.6.0, accessed July 18, 2024, at https://olsrr.rsquaredacademy.com.

Helsel, D.R., Hirsch, R.M., 1992, Statistical methods in water resources: New York, Elsevier Science Publishing Co., 522 p.

Hodge, S.A., and Tasker, G.D., 1995, Magnitude and frequency of floods in Arkansas: U.S. Geological Survey Water-Resources Investigation Report 95–4224, 52 p.

Jacobs, J., and Jardin, P., 2010, Estimating the magnitude of peak flows for steep gradient streams in New England: New England Transportation Consortium Report NETCR81, 42 p.

Johnson, C.G., and Tasker, G.D., 1974, Progress report on flood magnitude and frequency of Vermont streams: U.S. Geological Survey Open-File Report 74–130, 37 p. https://doi.org/10.3133/ofr74130.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.H., Kundzewicz, Z.W., Lettenmaier, D.P., and Stouffer, R.J., 2008, Stationarity is dead—Whither water management?: American Association for the Advancement of Science, accessed August 19, 2008, at www.sciencemag.org/.

Moglen, G.E., and Shivers, D.E., 2006, Methods for adjusting U.S. Geological Survey rural regression peak discharges in an urban setting: U.S. Geological Survey Scientific Investigations Report 2006–5270, 55 p. [Also available at https://doi.org/10.3133/sir20065270.]

National Oceanic and Atmospheric Administration, 2015, Precipitation frequency for Northeastern states, USA—NOAA Atlas 14 Volume 10: National Oceanic and Atmospheric Administration web site, accessed July 17, 2024, at https://hdsc.nws.noaa.gov/pfds/pfds_gis.html.

Natural Resources Canada, 2024, High resolution digital elevation model (HRDEM)—CanElevation series: Natural Resources Canada web page, accessed July 18, 2024, at https://open.canada.ca/data/en/dataset/957782bf-847c-4644-a757-e383c0057995.

Office of the Vermont State Climatologist, 1996, Climate of Vermont—Office of the Vermont State Climatologist: University of Vermont website, accessed February 7, 2025, at https://www.uvm.edu/~statclim/.

Olson, S.A., 2002, Flow-frequency characteristics of Vermont streams: U.S. Geological Survey Water-Resources Investigations Report 02–4238, 46 p.

Olson, S.A., 2014, Estimation of flood discharges at selected annual exceedance probabilities for unregulated, rural streams in Vermont, *with a section on* Vermont regional skew regression, by Veilleux, A.G.: U.S. Geological Survey Scientific Investigations Report 2014–5078, 27 p., plus app., https://doi.org/10.3133/sir20145078.

Olson, S.A., 2025, Data for the estimation of flood discharges at selected annual exceedance probabilities for unregulated, rural streams in Vermont, 2023: U.S. Geological Survey data release, https://doi.org/10.5066/P1MGCBK9.

Olson, S.A., and Bent, G.C., 2013, Annual exceedance probabilities of the peak discharges of 2011 at streamgages in Vermont and selected streamgages in New Hampshire, western Massachusetts, and northeastern New York: U.S. Geological Survey Scientific Investigations Report 2013–5187, 17 p., https://pubs.usgs.gov/sir/2013/5187/.

Potter, W.D., 1957a, Peak rates of runoff in the Adirondack, White Mountains, and Maine Woods area: Washington D.C., Bureau of Public Roads,

Potter, W.D., 1957b, Peaks rates of runoff in the New England hills and lowland area: Bureau of Public Roads.

PRISM Climate Group, 2022a, United States average total precipitation, 1991–2020, 800 m grid cell resolution, created December 5, 2022: Corvallis, Oreg., Oregon State University, accessed July 18, 2024, at http://www.prism.oregonstate.edu/.

PRISM Climate Group, 2022b, United States average monthly total precipitation, 1991–2020, 800 m grid cell resolution, created December 5, 2022: Corvallis, Oreg., Oregon State University, accessed July 18, 2024, at http://www.prism.oregonstate.edu/.

PRISM Climate Group, 2024a, United States average daily mean temperature, 1991–2020, 800 m grid cell resolution, created February 1, 2024: Corvallis, Oreg., Oregon State University, accessed July 18, 2024, at http://www.prism.oregonstate.edu/.

PRISM Group, 2024b, United States average monthly daily mean temperature, 1991–2020, 800 m grid cell resolution, created February 1, 2024: Corvallis, Oreg., Oregon State University, accessed July 18, 2024, at http://www.prism.oregonstate.edu/.

R Core Team, 2024a, The R stats package version 4.4.0: Vienna, Austria, R Foundation for Statistical Computing, accessed July 18, 2024, at https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html.

R Core Team, 2024b, R—A language and environment for statistical computing version 4.4.0: Vienna, Austria, R Foundation for Statistical Computing, accessed July 18, 2024, at, https://www.R-project.org.

Ries, K.G., III, 2007, The national streamflow statistics program—A computer program for estimating streamflow statistics for ungaged sites: U.S. Geological Survey Techniques and Methods 4–A6, 37 p.

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., https://doi.org/10.3133/fs20173046.

Smith, T.L., Olson, S.A., LeNoir, J.M., Kalmon, R.D., and Ahearn, E.A., 2025, Flood of July 2023 in Vermont: U.S. Geological Survey Scientific Investigations Report 2025–5016, 21 p., http://dx.doi.org/10.3133/sir20255016.

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.]

Schwarz, G.E., and Alexander, R.B., 1995, Soils data for the conterminous United States derived from the NRCS state soil geographic (STATSGO) data base [Original title: State soil geographic (STATSGO) data base for the conterminous United States.]: U.S. Geological Survey data release, accessed July 17, 2024, at https://doi.org/10.5066/P94JAULO.

U.S. Geological Survey, 2023, National hydrography dataset (NHD)—USGS National map downloadable data collection: USGS—National Geospatial Technical Operations Center web page, accessed July 17, 2024, at https://www.sciencebase.gov/catalog/item/4f5545cce4b018de15819ca9.

U.S. Geological Survey, 2024, 1/3rd arc-second digital elevation models (DEMs)—USGS National map 3DEP downloadable data collection: U.S. Geological Survey data release, accessed July 17, 2024, at https://www.sciencebase.gov/catalog/item/4f70aa9fe4b058caae3f8de5.

U.S. Geological Survey, 2024, PeakFQ—U.S. Geological Survey data analysis tools: U.S. Geological Survey web page, accessed February 7, 2025, at https://www.usgs.gov/tools/peakfq.

U.S. Geological Survey, [undated], Peak streamflow for the Nation: National Water Information System website, accessed February 7, 2025, at https://nwis.waterdata.usgs.gov/usa/nwis/peak.

Vermont Fish & Wildlife Department, 2025, Fishing opportunities: Vermont Fish & Wildlife Department web page, accessed July 30, 2025, at https://www.vtfishandwildlife.com/doc/fish/fishing-opportunities.

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

Appendix 1.   Streamgages with Data Used in this Investigation and the Maximum Annual Peak Discharge Recorded at the Streamgages in Vermont and Vicinity

This appendix table can be downloaded at https://doi.org/10.3133/sir20255088.

Appendix 2.   Summary of Peak-Discharge Data Used in the Flood-Frequency Analyses at Streamgages in Vermont and Vicinity

This appendix table can be downloaded at https://doi.org/10.3133/sir20255088.

Appendix 3.   Flood Discharges for Selected Annual Exceedance Probabilities for Selected Streamgages in Vermont and Vicinity

This appendix table can be downloaded at https://doi.org/10.3133/sir20255088.

Appendix 4.   Variance of Estimate at Selected Annual Exceedance Probabilities for Streamgages in Vermont and Vicinity

This appendix table can be downloaded at https://doi.org/10.3133/sir20255088.

Appendix 5.   Basin Characteristics Tested for Use in the Regression Equations

This appendix table can be downloaded at https://doi.org/10.3133/sir20255088.

Appendix 6.   Basin Characteristics Used to Develop the Regression Equations

This appendix table can be downloaded at https://doi.org/10.3133/sir20255088.

Appendix 7.   Example Application

The general procedure used to estimate the magnitude of a flood discharge at a selected recurrence interval as described in this report is explained by the following example of the procedure done for a random location on Wells River, Vermont. The drainage area, the areal percent of wetlands in the basin, and the basin-wide mean of the average annual precipitation at this location were determined to be 71.8 square miles, 6.49 percent, and 45.8 inches, respectively.

An estimate of the 1-percent annual exceedance probability peak discharge (Q_1) was needed for this rural, ungaged stream in Vermont. To determine the magnitude of the peak discharge at a 1-percent annual exceedance probability in cubic feet per second, equation 8 would be solved as:

$$Q_1 = 0.0254A^{0.796}10^{-0.0106W}P^{2.375} \tag{8}$$

$$Q_1 = 0.0254(71.8)^{0.796}10^{-0.0106(6.49)}(45.8)^{2.375}$$

$$Q_1 = 5,730 \text{ cubic feet per second (ft}^3\text{/s).}$$

The variance of prediction for this estimate would be computed using equation 11. The row vector, x_i, in equation 11 is based on the basin characteristics and is as follows:

$$x_i = [1 \ \log_{10}(71.8) \ 6.49 \ \log_{10}(45.8)] \text{ or}$$

$$x_i = [1 \ 1.856 \ 6.49 \ 1.661].$$

From table 5, the model error variance, γ^2, is 0.0341, and the $(X^t \Lambda^{-1}X)^{-1}$ matrix is

$$(X^{tr}\Lambda^{-1}X)^{-1} =$$

$$\begin{bmatrix} 4.01967e-01 & 1.31020e-03 & -9.02765e-04 & -2.35494e-01 \\ 1.31020e-03 & 5.30134e-04 & -1.33984e-05 & -1.26507e-03 \\ -9.02765e-04 & -1.33984e-05 & 1.77339e-05 & 4.88428e-04 \\ -2.35494e-01 & -1.26507e-03 & 4.88428e-04 & 1.38923e-01 \end{bmatrix}$$

Inserting x_i, γ^2, and $(X^t\Lambda^{-1}X)^{-1}$ into equation 11, the variance of prediction is

$$V_{pred} = \gamma^2 + x_i(X^{tr}\Lambda^{-1}X)^{-1}x_i^{tr} \tag{11}$$

$$V_{pred} = 0.0341 + 0.00106 \text{ or}$$

$$V_{pred} = 0.0352.$$

The standard error of prediction is

$$SE_{pred} = V_{pred}^{1/2} \tag{12}$$

$$SE_{pred} = 0.0352^{1/2} \text{ or}$$

$$SE_{pred} = 0.1876.$$

The standard error of prediction for the estimate in percent is then computed using equations 13 and 14:

$$S_{pos} = 100(10^{SEpred} - 1) \text{ and} \tag{13}$$

$$S_{neg} = 100(10^{-SEpred} - 1) \tag{14}$$

or:

$$S_{pos} = 100(10^{0.1876} - 1) \text{ and}$$

$$S_{neg} = 100(10^{-0.1876} - 1).$$

Thus, $S_{pos} = 54.0$ percent, and $S_{neg} = -35.1$ percent.

The prediction intervals for this estimate can be computed using equations 15 and 16. For example, if the 90-percent prediction interval was required, the critical value from a Student's-t distribution for the appropriate confidence interval is in any standard statistics textbook. The value for $t_{0.05,152}$ is 1.66, and the standard error of prediction is 0.1876. Equations 15 and 16 are solved as follows:

$$PI_{upper} = Q_{pred}10^{(t_{\alpha/2,n-p}SE_{pred})} \tag{15}$$

$$PI_{upper} = 5,730*10^{(1.66*0.1876)} \quad \text{or}$$

$$PI_{upper} = 11,700 \ cubic \ feet \ per \ second(ft^3/s)$$

$$PI_{lower} = \frac{Q_{pred}}{10^{(t_{\alpha/2,n-p}SE_{pred})}} \tag{16}$$

$$PI_{lower} = \frac{5,730}{10^{(1.66*0.1876)}} \quad \text{or}$$

$$PI_{lower} = 2,800 \ ft^3/s$$

Because the site is close to the Wells River at Wells River, Vermont (01139000) streamgage, the 100-year discharge value as derived from the regression equation can be improved by weighting it with the 1-percent annual exceedance probability flood discharge determined from the analysis at the streamgage. To weight the result of the ungaged location, a regression-weighted result is first determined for the gaged location. A weighted average (Q_w) of the regression estimate ($Q_{r(g)}$) and the flood-discharge estimate from a frequency analysis of the streamgage record (Q_s) can be computed using equation 17.

$$\log_{10}Q_w = \frac{\log_{10}Q_s\left(V_{pred}\right) + \log_{10}Q_{r(g)}\left(V_s\right)}{V_{pred} + V_s} \tag{17}$$

The 1-percent annual exceedance probability discharge (Qs) derived from the frequency analysis of the streamgage data is 5,740 ft³/s (app. 3). The variance of the estimate, V_s, for this streamgage at the 1-percent annual exceedance probability is 0.0035 and is in appendix 4. At the streamgage location, the drainage area is 98.9 square miles, the percentage of the basin covered by wetlands is 6.40 percent, and the basinwide mean of the average annual precipitation is 45.1 inches (app. 6). Inputting these values into equation 8, the regression equation estimate of the 1-percent annual exceedance probability discharge at the gaged location ($Q_{r(g)}$) is 7,150 ft³/s. The variance of prediction (V_{pred}) is then computed using equation 11 with the basin characteristics from the gaged location and the appropriate model error variance, γ^2, and $(X^t \Lambda^{-1} X)^{-1}$ matrix from table 5. V_{pred} is 0.0353. These values are input into equation 17 as follows:

$$log_{10}Q_w = \frac{log_{10}5,740(0.0353) + log_{10}7,150(0.0035)}{0.0353 + 0.0035}$$

$$Q_w = 5,850 \, ft^3 / s.$$

The weighted flood discharge, Q_w, is for the gaged location. Because a weighted estimate of flood discharge is desired at an ungaged site near a streamgage, the weighted estimate, Q_u, can be determined using equations 19 and 20.

$$c = \frac{log_{10}\left(\dfrac{Q_{r(u)}}{Q_{r(g)}}\right)}{log_{10}\left(\dfrac{A_u}{A_g}\right)} + \frac{log_{10}\left(\dfrac{Q_{r(g)}}{Q_w}\right)}{log_{10}(a)} \quad (19)$$

$$c = \frac{log_{10}\left(\dfrac{5,730}{7,150}\right)}{log_{10}\left(\dfrac{71.8}{98.9}\right)} + \frac{log_{10}\left(\dfrac{7,150}{5,850}\right)}{log_{10}(0.5)}$$

$$c = 0.402$$

$$Q_u = Q_w \left(\frac{A_u}{A_g}\right)^c \quad (20)$$

$$Q_u = 5,850\left(\frac{71.8}{98.9}\right)^{0.402}$$

$$Q_u = 5,140 \, ft^3 / s$$

USGS