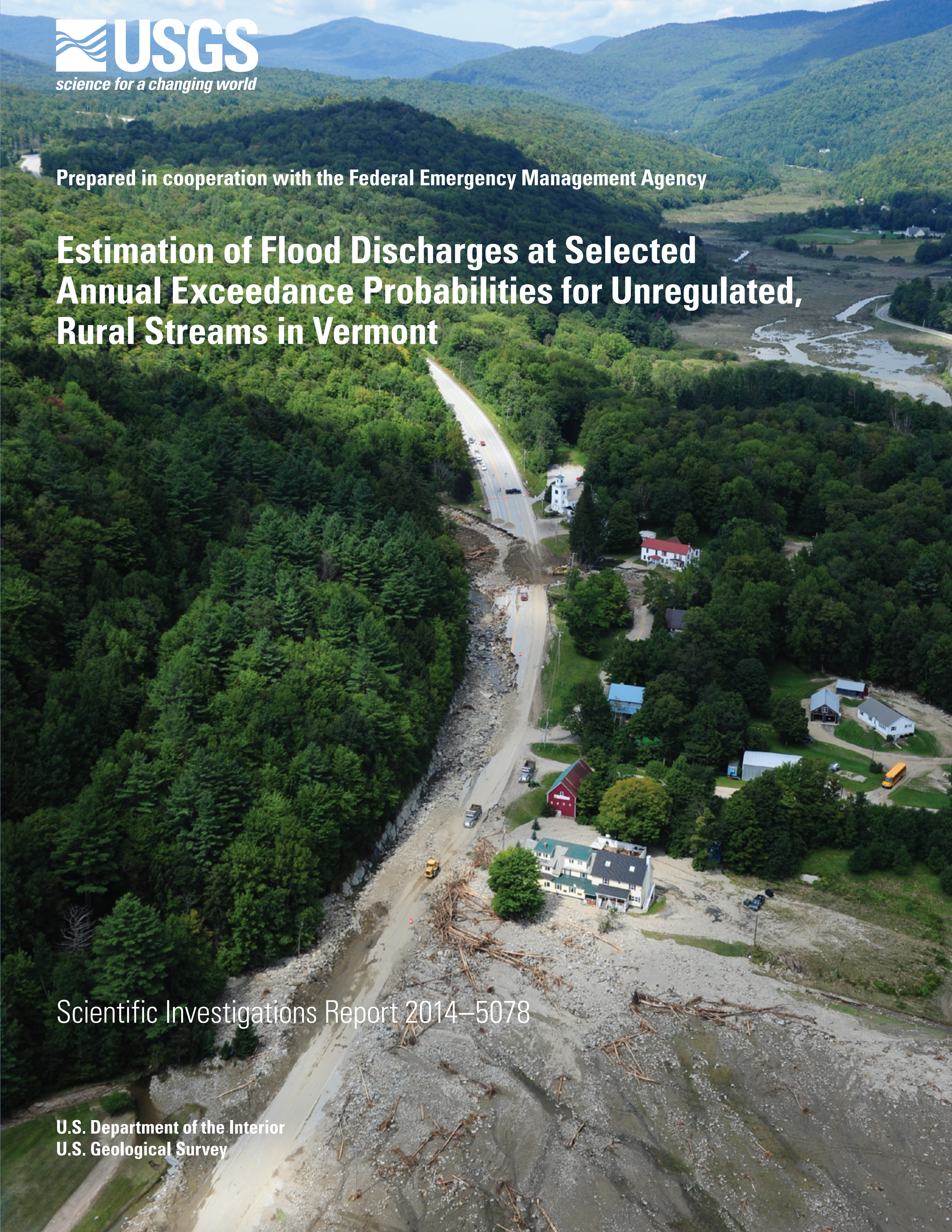


USGS
science for a changing world

Prepared in cooperation with the Federal Emergency Management Agency

Estimation of Flood Discharges at Selected Annual Exceedance Probabilities for Unregulated, Rural Streams in Vermont

Scientific Investigations Report 2014–5078

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

Cover. Aerial view of damage along U.S. Route 4 in Killington, Vermont, as a result of floodwaters from Roaring Brook during Tropical Storm Irene in August 2011. Photograph courtesy of Lars Gange, Mansfield Heliflight.

Estimation of Flood Discharges at Selected Annual Exceedance Probabilities for Unregulated, Rural Streams in Vermont

By Scott A. Olson

With a section on Vermont Regional Skew Regression
By Andrea G. Veilleux

Prepared in cooperation with the Federal Emergency Management Agency

Scientific Investigations Report 2014–5078

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

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2014

Suggested citation:
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 appendixes, *http://dx.doi.org/10.3133/sir20145078.*

ISSN 2328-0328 (online)

Contents

Figures

Tables

Conversion Factors and Datum

Multiply	By	To obtain
	Length	
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
	Area	
acre	4,047	square meter (m^2)
acre	0.4047	hectare (ha)
acre	0.004047	square kilometer (km^2)
square foot (ft^2)	929.0	square centimeter (cm^2)
square foot (ft^2)	0.09290	square meter (m^2)
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
	Volume	
cubic foot (ft^3)	28.32	cubic decimeter (dm^3)
cubic foot (ft^3)	0.02832	cubic meter (m^3)
acre-foot (acre-ft)	1,233	cubic meter (m^3)
acre-foot (acre-ft)	0.001233	cubic hectometer (hm^3)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
cubic foot per second per square mile [(ft^3/s)/mi^2]	0.01093	cubic meter per second per square kilometer [(m^3/s)/km^2]

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

°C=(°F–32)/1.8

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

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

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

Abbreviations

AEP	annual exceedance probability
AFPD	annual flood-probability discharge
ASEV	average sampling error variance
AVP_{new}	average variance of prediction at a new streamgage
B-GLS	Bayesian generalized least-squares regression
B-WLS	Bayesian weighted least-squares regression
EMA	Expected Moments Algorithm
EVR	Error Variance Ratio
FEMA	Federal Emergency Management Agency
GIS	geographic information system
GLS	generalized least-squares regression
gSSURGO	Gridded Soil Survey Geographic
MBV*	Misrepresentation of the Beta Variance
MSE	mean square error
MSEG	mean square error G
NED	National Elevation Dataset
OLS	ordinary least-squares analysis
PRISM	Parameter-elevation Regressions on Independent Slopes Model
Pseudo ANOVA	Pseudo Analysis of Variance
STATSGO	State soil geographic database
USGS	U.S. Geological Survey
VIF	variance inflation factor
Vs	variance of estimates
WLS	weighted least-squares analysis
WREG	Weighted-Multiple-Linear Regression Program
WY	water year

Estimation of Flood Discharges at Selected Annual Exceedance Probabilities for Unregulated, Rural Streams in Vermont

By Scott A. Olson

Abstract

This report provides estimates of flood discharges 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-, 200-, and 500-years, respectively) for ungaged, unregulated, rural streams in Vermont. The equations were developed using generalized least-squares regression. Flood-frequency and drainage-basin characteristics from 145 streamgages were used in developing the equations. The drainage-basin characteristics used as explanatory variables in the regression equations include drainage area, percentage of wetland area, and the basin-wide mean of the average annual precipitation. The average standard errors of prediction for estimating the flood discharges at the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent AEP with these equations are 34.9, 36.0, 38.7, 42.4, 44.9, 47.3, 50.7, and 55.1 percent, respectively.

Flood discharges at selected AEPs for streamgages were computed by using the Expected Moments Algorithm. To improve estimates of the flood discharges for given exceedance probabilities at streamgages in Vermont, a new generalized skew coefficient was developed. The new generalized skew for the region is a constant, 0.44. The mean square error of the generalized skew coefficient is 0.078. This report describes a technique for using results from the regression equations to adjust an AEP discharge computed from a streamgage record. This report also describes a technique for using a drainage-area adjustment to estimate flood discharge at a selected AEP for an ungaged site upstream or downstream from a streamgage.

The final regression equations and the flood-discharge frequency data used in this study will be available in StreamStats. StreamStats is a World Wide Web application providing automated regression-equation solutions for user-selected sites on streams.

Introduction

Flooding is the most costly natural hazard experienced in Vermont. Intense precipitation, a series of closely spaced major storms, springtime storms combined with snowmelt, tropical storms, and ice jams have all caused flooding in Vermont. 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 occurred in parts of the State in 1927, 1936, 1938, 1973, 1982, 1984 (Hammond, 1991), 1995, 1996, 1997, 1998, 2002, and 2011.

In response to extensive damage caused by the closely spaced floods of 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 to decrease damages caused by flooding of major rivers. However, flooding continues to be a threat. In 2011, maximum streamflows resulting from Tropical Storm Irene were the greatest ever recorded at 37 streamgages on rivers and streams in Vermont (Olson and Bent, 2013).

The U.S. Geological Survey (USGS) and other agencies have been measuring and recording discharge at numerous streamgages throughout Vermont for the past 100 years. One use of the data collected from these streamgages is the characterization of the magnitudes and frequencies of flood discharges for rivers in the State. In 2013, there were 55 continuously operating streamgages and 29 crest-stage gages (only maximum annual discharge is determined) on Vermont rivers and streams. There are also 10 gages that continuously monitor the stage of lakes and reservoirs in Vermont.

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 flood plains. Computation of flood-discharge magnitude and estimation of AEP require a statistical analysis of peak-discharge data collected at streamgages. However, estimates often are required for ungaged sites where no observed peak-discharge data are available. Several investigations that provide methods for estimating flood-discharge frequency at ungaged sites in Vermont have been published, including Benson (1962), Potter (1957a, b), Johnson and Tasker (1974), Dingman and Palaia (1999), Olson (2002) and Jacobs and Jardin (2010). Updated flood-discharge frequency estimates, benefiting from additional years of peak-discharge data and enhanced statistical procedures, can improve techniques for estimating flood-discharge frequency at ungaged sites. To address this, the USGS, in cooperation with the Federal Emergency Management Agency (FEMA), conducted this study to develop updated methods for estimating the flood discharges at selected recurrence intervals for unregulated and ungaged stream locations in and adjacent to Vermont.

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 for streamgages in and adjacent to Vermont and (2) describes methods, including the use of equations developed from regression analyses, for estimating flood discharges at selected AEPs on ungaged, unregulated Vermont streams. In addition, this report (3) presents a method for estimating the standard error of prediction for each estimate made with 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 (mi^2) of land area in the northeastern United States, nearly one-eighth of the total land area of New England. The State is approximately 155 miles long from north to south and ranges from about 35 miles wide (east to west) at the southern end of the State to nearly 90 miles wide at its northern end. Vermont is bordered on the northern half of its western boundary by Lake Champlain and on the east by the Connecticut River. Within the State of Vermont there are more than 7,000 miles of rivers and streams and more than 800 lakes and ponds (Vermont Fish and Wildlife Department, 2010).

Vermont is largely forested, with 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 190 feet (ft) along the Connecticut River at the southern boundary of Vermont and 100 ft along the shoreline of Lake Champlain to more than 3,000 ft at numerous peaks. The climate of the region is temperate and humid with four distinct seasons. Precipitation is distributed fairly evenly across the region, with averages ranging from about 40 to 50 inches (in.) per year, except in regions of high elevation, which can receive an additional 10 to 20 in. of precipitation annually. Annual snowfall also varies across the State. In the Connecticut River Valley and in the Lake Champlain Valley, normal snowfall ranges from 55 to 65 in. annually. Higher elevations receive substantially more snowfall. The annual mean temperature is 46 degrees Fahrenheit (°F) at Burlington, Vermont, and 44.6 °F in Rutland, Vt. The number of days per year the minimum temperature is less than or equal to 32 °F is, on average, 144 in Burlington, Vt., and 169 in Rutland, Vt. (National Weather Service, 2013).

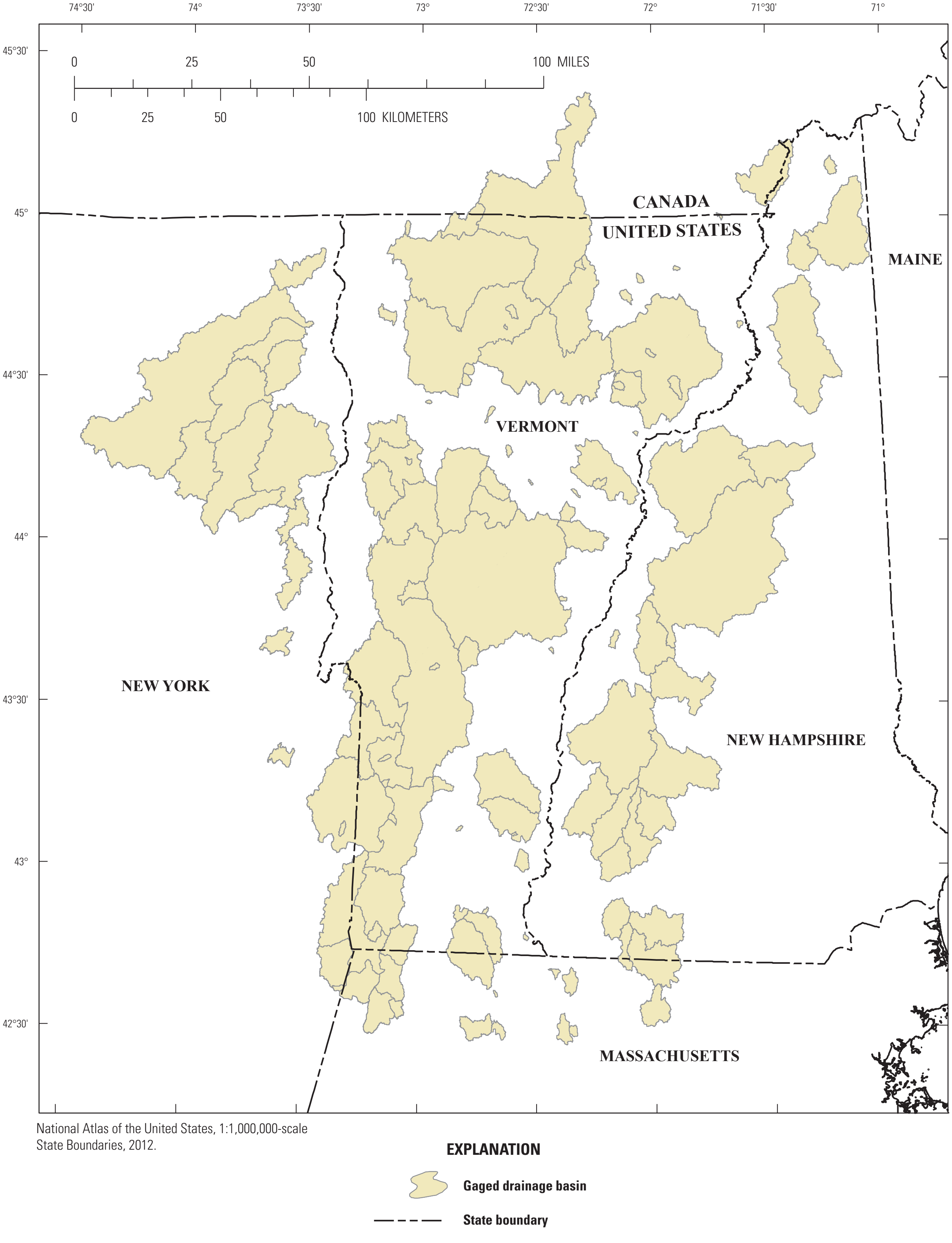
Flood Discharges at Selected Annual Exceedance Probabilities for Streamgages

To develop techniques for estimating flood-discharge magnitudes and frequencies for ungaged stream locations, flood-discharge magnitude and frequency are first computed at long-term streamgages for which sufficient 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 ungaged site by using drainage-basin characteristics.

Peak-Discharge Data Used in This Study

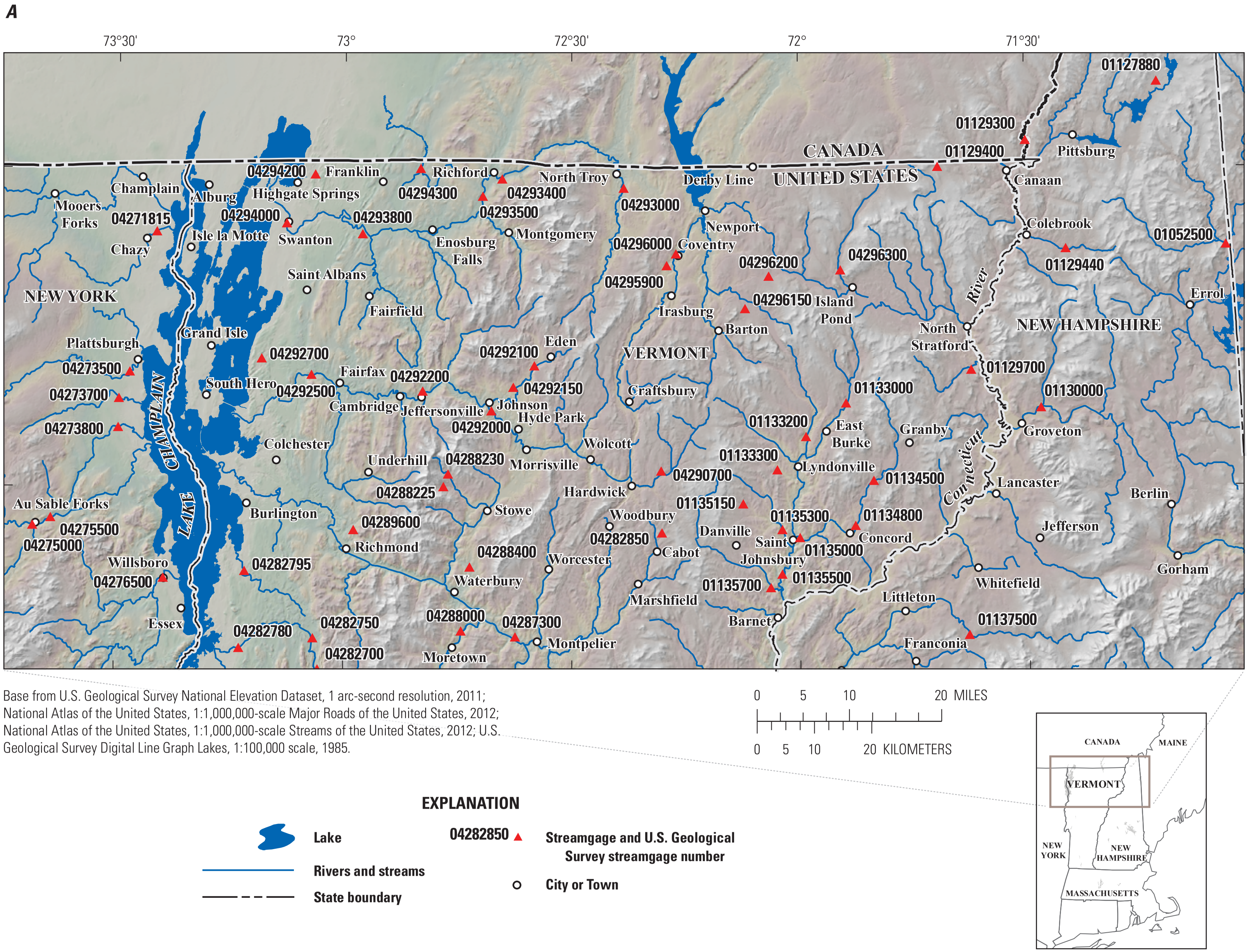
All available annual peak-discharge data for Vermont and adjacent, physiographically similar areas in New Hampshire, Massachusetts, and New York States, and Quebec, Canada, collected by the USGS, U.S. Forest Service, University of Vermont, and Environment Canada were considered for this study. These data include records from continuously recording streamgages and crest-stage streamgages (streamgages that record only the annual peak discharge), both current and discontinued. Current records for this report include records through water year 2011. The water year (WY) is designated by the calendar year in which it ends. It begins October 1 of the previous calendar year and ends September 30.

Of the sites considered, 153 streamgages were selected for use in this study (fig. 2; *appendix 1*). The selection criteria required the streamgage to have a minimum of 10 years of annual peak-discharge data that were free of substantial trends and 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. The streamgages selected are spatially well distributed in and adjacent to Vermont (fig. 2).



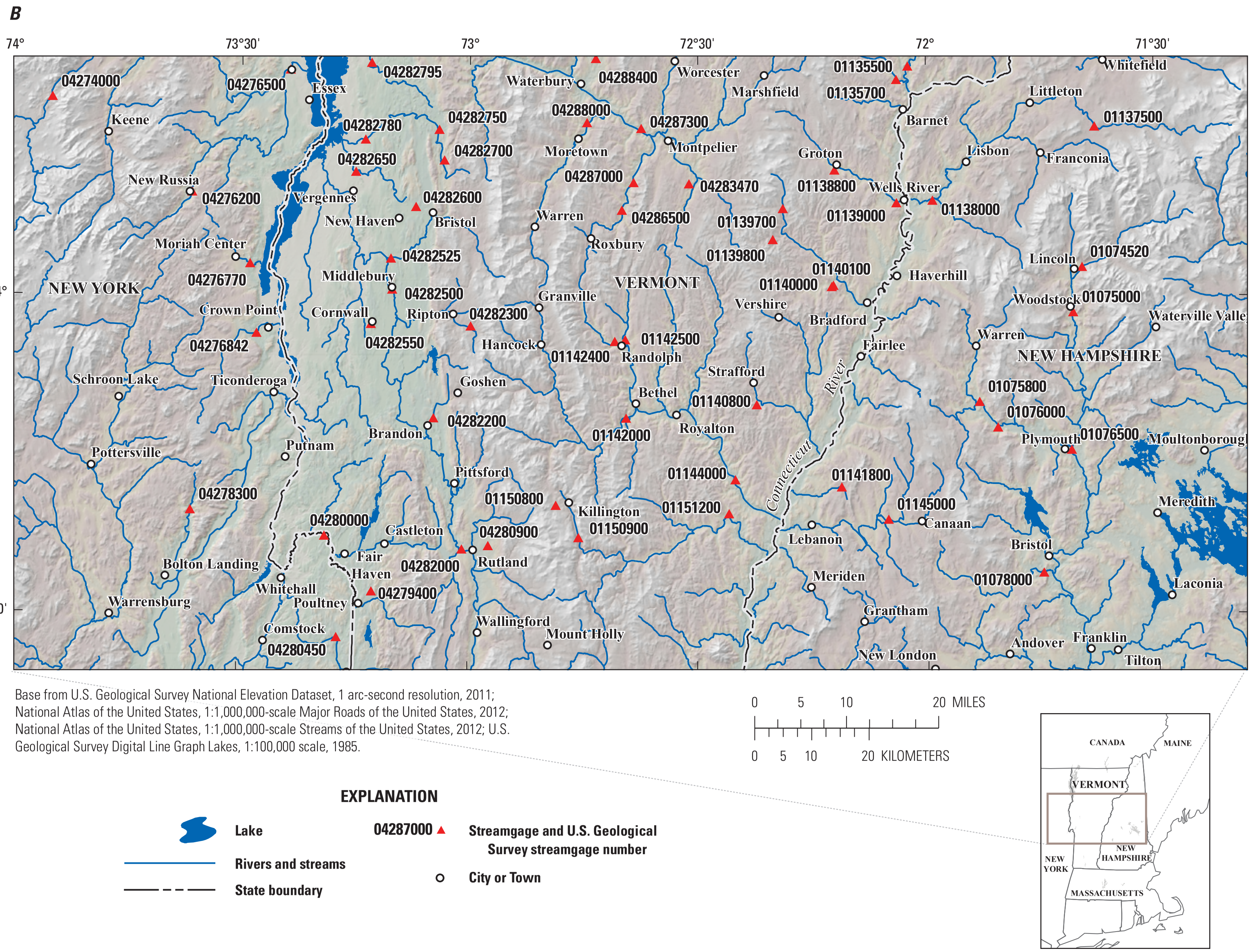
National Atlas of the United States, 1:1,000,000-scale State Boundaries, 2012.

Figure 1. Location of selected drainage basins with streamgages, Vermont and vicinity.



Base from U.S. Geological Survey National Elevation Dataset, 1 arc-second resolution, 2011;
National Atlas of the United States, 1:1,000,000-scale Major Roads of the United States, 2012;
National Atlas of the United States, 1:1,000,000-scale Streams of the United States, 2012; U.S.
Geological Survey Digital Line Graph Lakes, 1:100,000 scale, 1985.

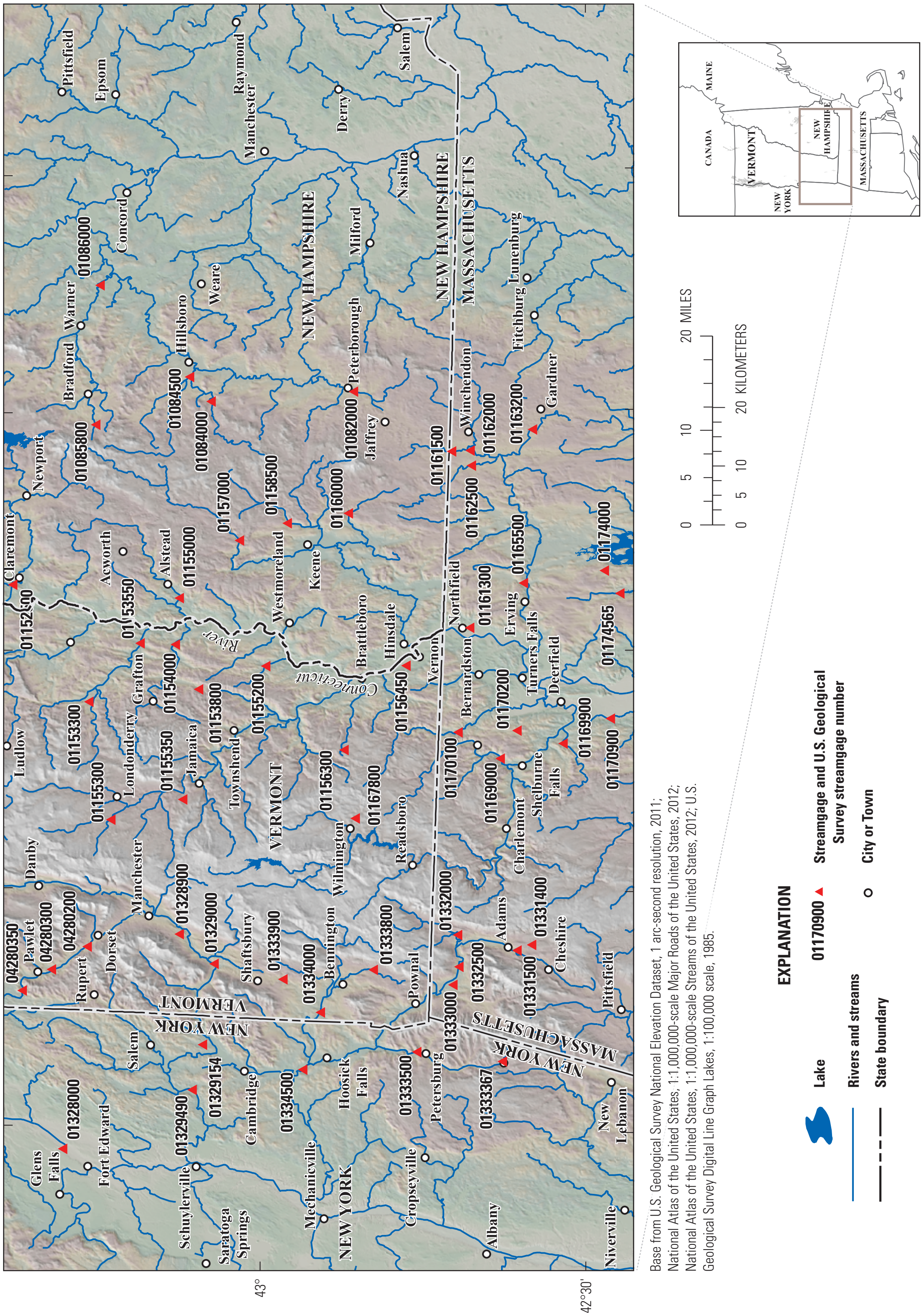
Figure 2. Location of selected streamgages with peak-discharge data in the *A*, northern section, *B*, middle section, and *C*, southern section of Vermont and vicinity.



Base from U.S. Geological Survey National Elevation Dataset, 1 arc-second resolution, 2011; National Atlas of the United States, 1:1,000,000-scale Major Roads of the United States, 2012; National Atlas of the United States, 1:1,000,000-scale Streams of the United States, 2012; U.S. Geological Survey Digital Line Graph Lakes, 1:100,000 scale, 1985.

Figure 2. Location of selected streamgages with peak-discharge data in the *A*, northern section, *B*, middle section, and *C*, southern section of Vermont and vicinity. —Continued

C



Base from U.S. Geological Survey National Elevation Dataset, 1 arc-second resolution, 2011;
National Atlas of the United States, 1:1,000,000-scale Major Roads of the United States, 2012;
National Atlas of the United States, 1:1,000,000-scale Streams of the United States, 2012; U.S.
Geological Survey Digital Line Graph Lakes, 1:100,000 scale, 1985.

Figure 2. Location of selected streamgages with peak-discharge data in the *A*, northern section, *B*, middle section, and *C*, southern section of Vermont and vicinity.—Continued

Of the 153 streamgages selected, data for each of 3 streamgages were combined with data from one of three other streamgages because the sites were close together on the same stream. This is discussed further in “Combined Records of Nearby Streamgages.” Combining the data resulted in 150 streamgages with data available for the regression analysis. In addition, during the regression analysis, it was found that data for the basin characteristics that were the best explanatory variables did not extend into Canada. For that reason, five streamgages that have drainage basins extending into Canada were eliminated from the development of the regression equations. A total of 145 streamgages were used to develop the regression equations.

None of the streamgages included in this investigation have drainage basins considered to be urbanized. The maximum percentage of land area in a streamgage drainage basin classified as developed land in the 2006 National Land Cover Data (Fry and others, 2011) is 24.8 percent; the average is 4.5 percent.

In recent years, there has been much speculation regarding stationarity (the assumption that the mean and variability of data from past observations will continue unchanged in the future) of annual peak-discharge data (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 completed. The trend test was done with software called PeakFQ developed by the USGS to analyze peak-flow data (U.S. Geological Survey, 2013a). Streamgage records with less than 30 years of annual peak-discharge data were not tested for trends because trends over a period of record this short cannot be distinguished from serial correlation.

No substantial trends were found by using the Kendall Tau test for the peak-discharge data used in this study. The Kendall Tau statistics indicated that an upward trend may exist (p-value less than or equal to 0.1) for 20 of the 70 streamgages with at least 30 years of record. Of these 20 streamgages, the trend was considered marginal (p-value less than or equal to 0.1 and greater than 0.05) at 6 streamgages.

For the streamgages indicating a possible trend, the trend could be explained as being the result of extreme climatic anomalies near the beginning or end of the peak-discharge record, such as the 2011 flooding or the drought of 1960–69 (Hammond, 1991). The evidence of trends did not exist or was statistically insignificant when extreme events, such as those, were eliminated from the Kendall Tau trend tests. Hence, the annual peak-discharge data used in this study are regarded as random, independent events that are homogeneous for a streamgage throughout the period of record.

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

The flood discharges with AEPs of 50-, 20-, 10-, 4-, 2-, 1-, 0.5- and 0.2-percent for the 150 streamgages (*appendix 1*) were computed by using the Expected Moments Algorithm (EMA) (Cohn and others, 1997 and 2001). The EMA methodology generally follows guidelines provided in Bulletin 17B of the U.S. Interagency Advisory Committee on Water Data (1982) and guidelines from the U.S. Geological Survey (2012). EMA uses the log-Pearson Type III distribution for estimating flow frequency; however, it provides updated procedures for incorporating historical peaks and censored peaks. A summary of input data for incorporating historical and censored peaks is provided in *appendix 2*. The Multiple Grubbs-Beck test was applied to detect and treat low outliers as recommended by the Hydrologic Frequency Analysis Work Group (*http://acwi.gov/hydrology/Frequency/minutes/Minutes_HFAWG_meeting_mar19_2012_040212.pdf*). Software developed by the USGS to analyze peak-discharge data (PeakFQ version 7.0) was used for these computations (Office of Surface Water, U.S. Geological Survey, written commun., April 3, 2013). The annual peak-flow data used as input to the PeakFQ program were retrieved from the National Water Information System (U.S. Geological Survey, 2013b). Peak discharges affected by dam failure, ice jam breach, or a similar event are not included in the frequency analyses. Annual peak discharges through WY 2011 were included in the frequency analysis.

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 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 in estimating flood discharges for selected recurrence intervals at a streamgage is weighted with a generalized skew estimated by pooling the skews from nearby streamgages. The generalized skew can be taken from the generalized skew map in Bulletin 17B (U.S. Interagency Advisory Committee on Water Data, 1982) or Olson (2002); however, these maps are considered outdated. For these reasons, a new method for obtaining generalized skew for Vermont was developed.

Bayesian weighted least-squares/generalized least-squares regression analysis was used to develop the new skew. The analysis was based upon peak-discharge data from 72 streamgages in Vermont and the surrounding states, each having 30 or more years of record. In addition to the peak-discharge data, 28 basin characteristics for each of the 72 sites were calculated as explanatory variables in the regional study. The basin characteristics available include land cover and climatic characteristics, as well as the more standard morphometric characteristics, such as location of the basin centroid, drainage area, main basin slope, and mean drainage basin elevation, among others. However, none of the basin characteristics were statistically significant in explaining the site-to-site variability in skewness. Thus, the best model was a region-wide constant model with a value of 0.44. The mean square error of the generalized skew is 0.078. Additional details on the development of the new generalized skew can be found in the section titled “Vermont Regional Skew Regression.”

Combined Records of Nearby Streamgages

Three discontinued streamgages included in this investigation were located a short distance upstream or downstream from new, active streamgages. In each case, the locations of the discontinued streamgage and the newer, active streamgage are considered to be proximate enough to have similar drainage-basin characteristics and represent a single streamgage peak-flow dataset. For this reason, *appendix 1* contains 153 streamgages, although the AEPs are reported for 150 streamgages. Before combining the peak-discharge record of the discontinued streamgage and peak-discharge record of the newer, active streamgage, the peak discharges from either the discontinued record or the active record were adjusted by the drainage area ratio of the two sites. In each of the three cases, the downstream streamgage was used in the regression analysis. Streamgages for which peak discharge records were adjusted for drainage area and combined with the record from a nearby streamgage 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 are listed in *appendix 3*. The discharges reported in *appendix 3* supercede discharges reported in Olson and Bent (2013) because of the updated generalized skew and regression equations.

Maximum Recorded Floods

The maximum recorded annual-peak discharges (*appendix 4*) plotted in relation to drainage area for each streamgage in this study are displayed in figure 3. Annual-peak discharges affected by dam failure, ice jam breach, or a similar event are not included. A New England regional envelope curve developed by Crippen and Bue (1977) also is shown in figure 3 along with a line developed by using generalized least-squares regression analysis showing the relation between drainage area and the peak discharge with a 1-percent annual exceedance probability. Figure 3 can be used to evaluate the reasonableness of flood estimates made by using techniques described in this report.

Table 1. Active and discontinued streamgages with peak-discharge records combined to create a longer period of record, Vermont and vicinity.

[USGS, U.S. Geological Survey; mi^2, square miles; N.H., New Hampshire; Vt., Vermont]

Discontinued streamgage				Active streamgage				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, N.H.	106	1929–52, 1960, 1968–70	[1]01074520	East Branch Pemigewasset River at Lincoln, N.H.	115	1993–2011	1.085
01153500	Williams River at Brockways Mills, Vt.	102	1941–84	[1]01153550	Williams River near Rockingham, Vt.	112	1987–2011	1.098
[1]01155000	Cold River at Drewsville, N.H.	83.5	1941–78, 2006	01154950	Cold River at High Street, at Alstead, N.H.	74.6	2010–11	1.119

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

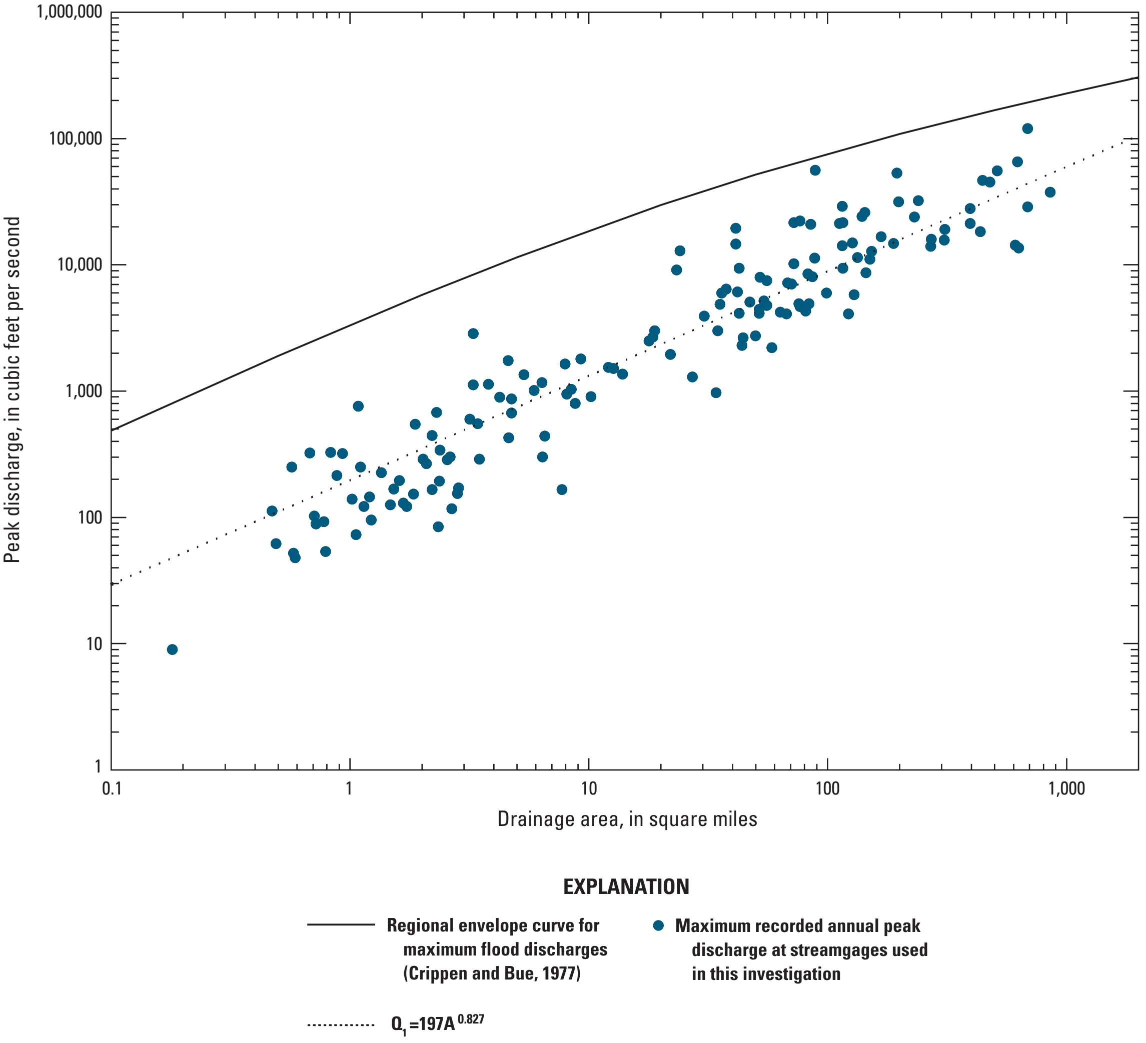


Figure 3. Maximum recorded annual-peak discharges at streamgages in Vermont and vicinity in relation to drainage area, and a regional envelope line and a regression line relating the 1-percent annual exceedance probability flood discharge (Q_1) to drainage area.

Characteristics of Streamgage Drainage Basins

In flood-frequency regression analysis, the variations in the magnitude of flood discharges at a selected AEP for streamgages used in the study 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, 120 basin characteristics were determined for each streamgage, including physical properties, such as drainage area, channel slope, elevation, forest cover, lake area, and soil permeability, and climatic characteristics, such as precipitation and temperature.

Boundaries for the streamgage drainage basins were determined by using a digital elevation dataset derived from the National Elevation Dataset (NED) (U.S. Geological Survey, 2004a) resampled to a 10-meter resolution. Prior to being used for basin delineation, the NED was hydrologically corrected by using the National Hydrography Dataset (U.S. Geological Survey, 2004b) to ensure the correct location of stream centerlines and the Watershed Boundary Dataset (National Resources Conservation Service, 2001) to ensure the correct location of selected basin boundaries. Some boundaries were corrected manually by using 1:24,000 digital raster graphs (U.S. Geological Survey, 2001). Additional dimensional properties of the drainage basins and waterways were computed with the ArcHydro software (Environmental Systems Research Institute, Inc., 2008).

With appropriate geographic information system (GIS) datasets, other basin characteristics also were delineated with the ArcHydro software. The National Hydrography Dataset, the 2006 National Land Cover Data (Fry and others, 2011), the State soil geographic (STATSGO) database (U.S. Geological Survey, 1995), the Gridded Soil Survey Geographic (gSSURGO) by state (National Resources Conservation Service, 2012), and the National Wetlands Inventory (U.S. Fish and Wildlife Service, 2009) were the source GIS datasets for land-surface properties. The sources for climatic data were PRISM (Parameter-elevation Regressions on Independent Slopes Model) (PRISM Group, Oregon State University, 2012a–c) and Extreme Precipitation in New York & New England (Northeast Regional Climate Center, 2013). A complete list of basin characteristics determined for potential use as explanatory variables in the regression analysis is presented in *appendix 5*.

Regression Equations for Estimation of Flood Discharges at Selected Annual Exceedance Probabilities for Ungaged Stream Sites

Multiple-regression techniques, employing 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. The use of generalized least-squares regression allows for weighting of streamgage data to compensate for the 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 for estimating 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 having 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

$$Y_P = 10^{b_0} X_1^{\,b_1} X_2^{\,b_2} \ldots X_j^{\,b_j} \, . \tag{2}$$

The limitations, sensitivity, and accuracy of the regression equations are reported following the final regression equations. In addition, techniques are discussed for determining 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 also are discussed.

Regression Analysis and Final Regression Equations

A total of 120 basin characteristics were determined for each streamgage and used in the regression analysis. Mathematical transformations were applied to each basin characteristic and flood discharge statistic to obtain the most linear relations. The transformations used were logarithms, square roots, squares, and raising the values to the -0.125 power. Correlation data and stepwise linear regression (SAS Institute, Inc., 1990) were used to evaluate which basin characteristics, transformed or untransformed, were the most significant explanatory variables.

Next, generalized least-squares regression techniques were used to determine the final significant basin characteristics and to compute the final regression equations. The generalized least-squares regression analysis was done by using the Weighted-Multiple-Linear Regression Program (WREG), a hydrologic regression program that uses the generalized least-squares regression procedure (Eng and others, 2009). The basin characteristics used in the development of the final regression equations are listed in *appendix 6*, by streamgage. Logarithmic base-10 transformations were made on all final variables in the equations. The final regression equations (equations 3–10) for estimating flood discharges on ungaged, unregulated streams in rural drainage basins in Vermont are as follows:

$$Q_{50} = 0.145 A^{0.900} W^{-0.274} P^{1.569} \, , \tag{3}$$

$$Q_{20} = 0.179 A^{0.884} W^{-0.277} P^{1.642} \, , \tag{4}$$

$$Q_{10} = 0.199 A^{0.875} W^{-0.280} P^{1.685} \, , \tag{5}$$

$$Q_{4} = 0.219 A^{0.866} W^{-0.286} P^{1.740} \, , \tag{6}$$

$$Q_2 = 0.237A^{0.860}W^{-0.291}P^{1.774}, \tag{7}$$

$$Q_1 = 0.251A^{0.854}W^{-0.297}P^{1.809}, \tag{8}$$

$$Q_{0.5} = 0.266A^{0.849}W^{-0.301}P^{1.840}, \text{ and} \tag{9}$$

$$Q_{0.2} = 0.289A^{0.844}W^{-0.309}P^{1.876}, \tag{10}$$

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, plus 1.0, from the National Land Cover Data (Fry and others, 2011) using a GIS; and

P is the basin-wide mean of the average annual precipitation, in inches, determined with the PRISM 1981–2010 annual precipitation dataset (PRISM Group, Oregon State University, 2012a) resampled to a 800-meter-cell resolution by using bilinear interpolation.

Because of a lack of comparable wetland and precipitation GIS databases available for Canada, streamgages with a drainage basin extending into Canada were not used to develop regression equations 3–10. These gages include Halls Stream near East Hereford, Quebec (01129300); Black Brook at Averill, Vt. (01129400); Missisquoi River near East Berkshire, Vt. (04293500); Missisquoi River at Swanton, Vt. (04294000); and Pike River at East Franklin, near Enosburg Falls, Vt. (04294300). 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 4A and B. No apparent trends or patterns were observed in any of the plots. Thus, the streamgages were not grouped into subregions, and the equations presented in this report are intended for regionwide use.

The residual plots shown in figures 4A and B also were used as a diagnostic tool for the regression equations. The random scatter of the points above and below the zero reference line provides verification that the model is satisfactorily meeting the assumptions of multiple-linear-regression techniques. Other diagnostic tools included the evaluation of Cook's D and the variance inflation factor (VIF) (Helsel and Hirsch, 1992). Cook's D is a value that is computed for each observation—the data used for developing the regression equations. It is a measure of the influence of each observation on the regression equations and can be used to assist in the identification of outliers. The magnitude of Cook's D flagged some observations as potential outliers; however, it was concluded that the potential outliers were sound data and there was no justification for excluding them from the regression analysis.

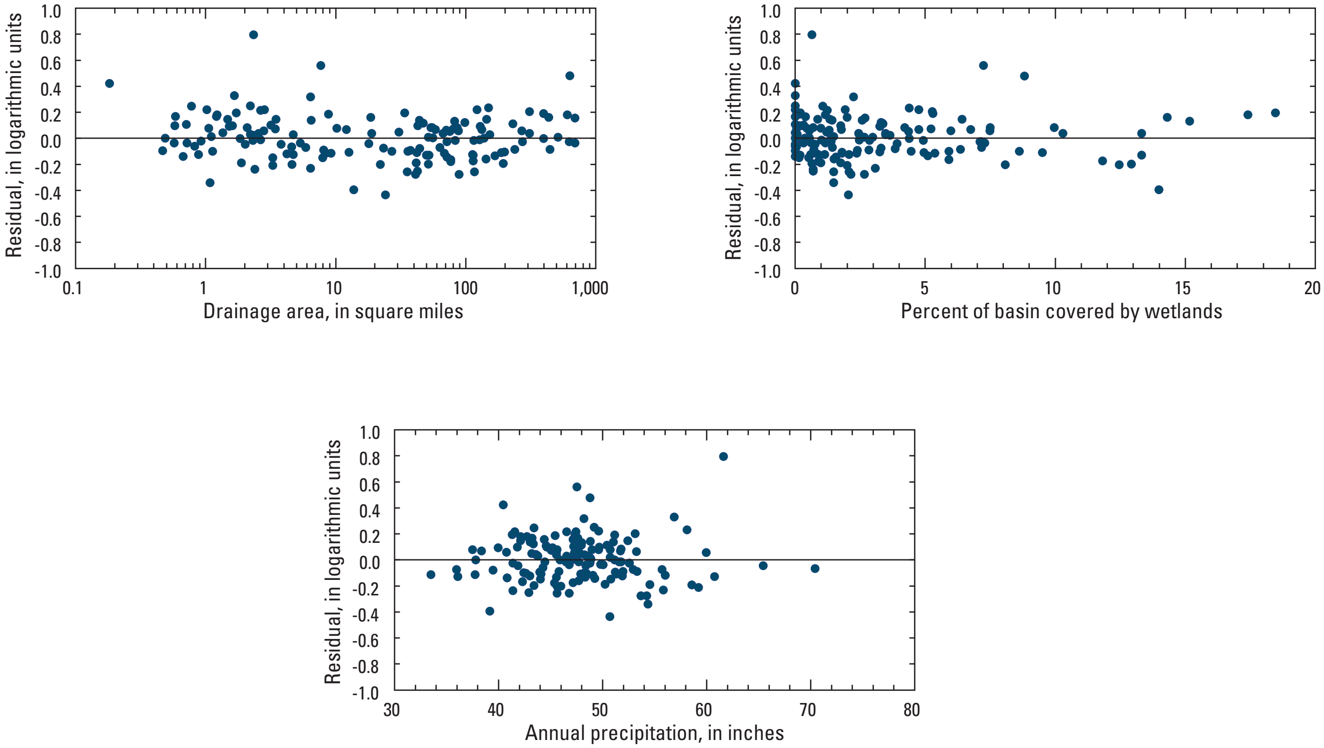
The VIF is a diagnostic tool that may be used to evaluate collinearity of explanatory variables. There are no formal criteria for VIF, although some authors suggest that a VIF exceeding 10 may be cause for concern (Freund and Littell, 2000), indicating that explanatory variables may be correlated. The greatest VIF computed for a variable used in the final regression equations was 1.9.

Limitations and Sensitivity

It is important to note that basin characteristics used to develop equations 3 through 10 were determined with the ArcHydro (Environmental Systems Research Institute, Inc., 2008) software by using datasets described in "Regression Analysis and Final Regression Equations." Determining the basin characteristics for use in the regression equations with alternate data sources or by using different computational methods than those of the ArcHydro software may produce statistics that are different from those reported here and may introduce bias and yield discharge estimates that have unknown error.

The regression equations are applicable to sites on ungaged, unregulated streams in rural basins that are within the region covered by the drainage basins used in this investigation. Use of the equations is appropriate to sites with drainage-basin characteristics that are within the range of drainage-basin characteristics used in the development of the equations. The ranges of drainage-basin characteristics used in the analysis are shown in table 2 and figure 5. If 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 ranges, a simplified equation that uses drainage area as the only explanatory variable is provided in the section titled "Drainage-Area-Only Regression Equations." For sites that are considered urban, Moglen and Shivers (2006) describe techniques for transforming rural flood-discharge frequency estimates to estimates for urban watersheds.

***A.* 10-Percent Annual Exceedance Probability**



***B.* 1-Percent Annual Exceedance Probability**

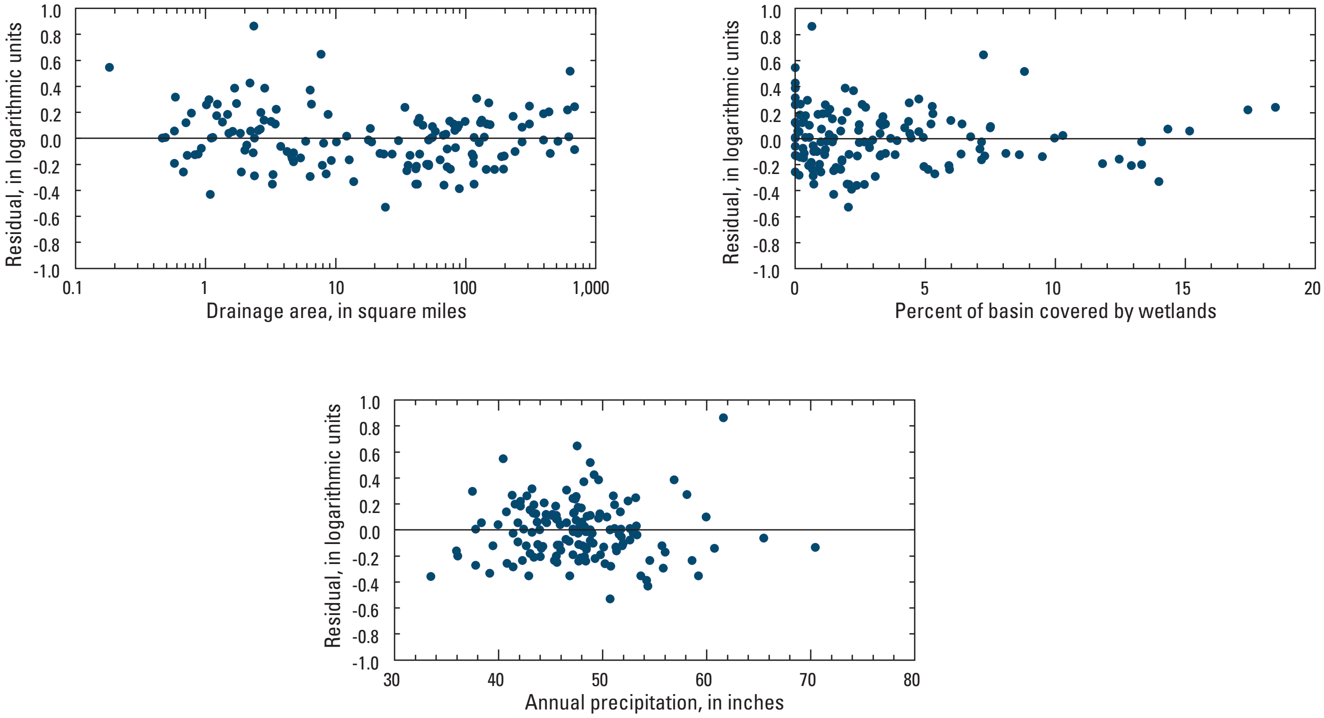


Figure 4. Residuals of the regression equations for estimating the magnitude of a discharge with a *A*, 10-percent and *B*, 1-percent exceedance probability in relation to basin characteristics in Vermont and vicinity.

Table 2. Ranges of explanatory variables used in the development of the regression equations for estimating flood discharges at selected annual exceedance probabilities for ungaged, unregulated streams in Vermont and vicinity.

Explanatory variable	Minimum	Maximum	Mean
Drainage area, in square miles	0.18	689	84.2
Percent of basin covered by wetlands	0	18.5	3.42
Basin-wide mean of the average annual precipitation, in inches	33.5	70.4	47.6

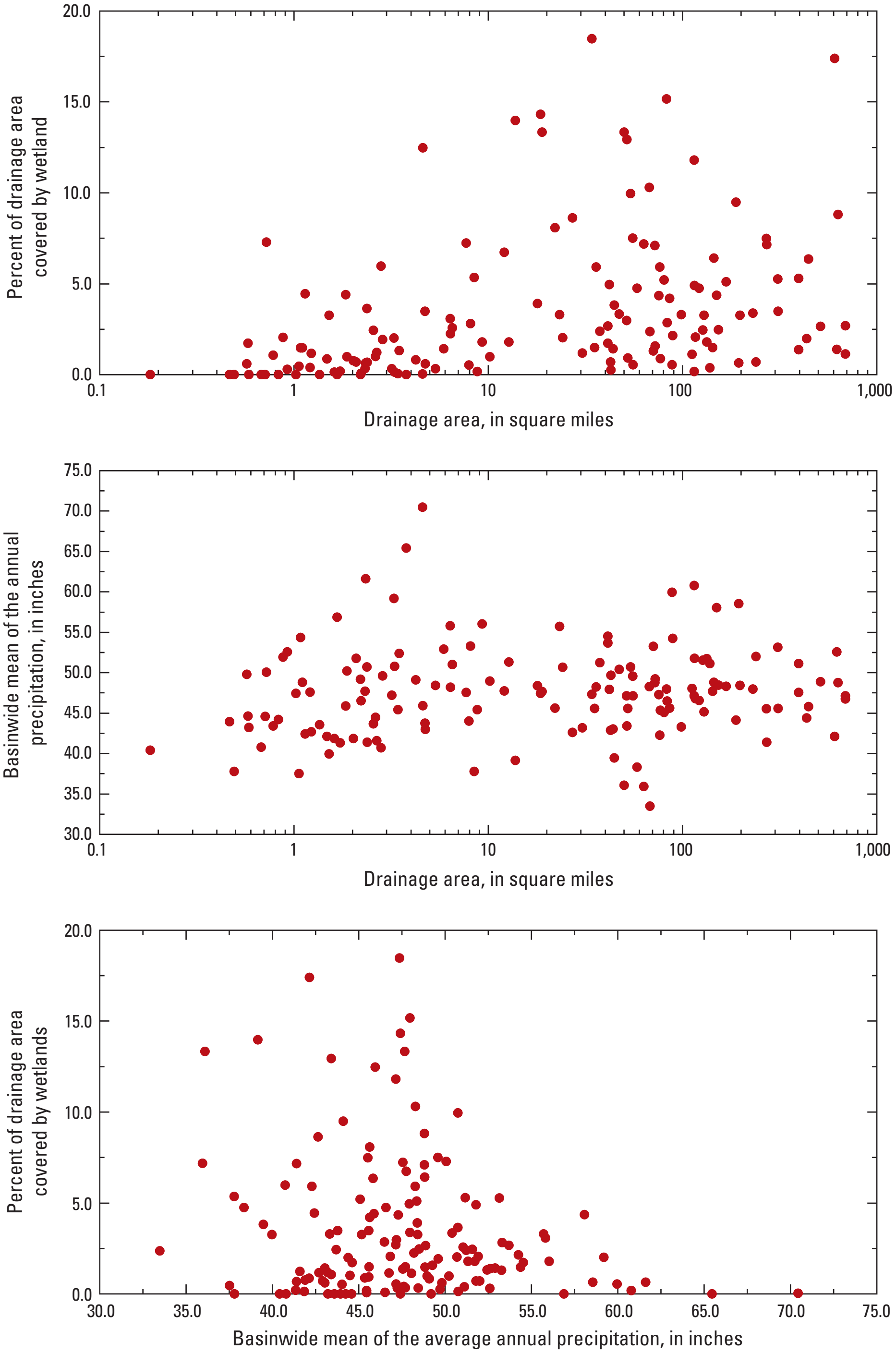


Figure 5. Two-dimensional ranges of explanatory variables used to develop the regression equations for estimating flood discharges at selected annual exceedance probabilities for streams in Vermont and vicinity.

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 conducted by adjusting a basin characteristic by plus or minus 10 percent while holding the other basin characteristics constant at their respective mean magnitudes. The results of the sensitivity analysis listed in table 3 indicate that the regression equations are most sensitive to changes in drainage area and 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 percentage of its computed flood discharge range (table 2) than does a 10-percent change in drainage area.

Accuracy of the Regression Equations

There are several measures of the accuracy of a regression equation. The pseudo coefficient of determination, or R^2_{pseudo}, indicates the variability observed in the dependent variable that is accounted for by the regression model after removing the effect of the time-sampling error (Eng and others, 2009). The closer the adjusted coefficient of determination is to 1, the better the regression explains the variation in the dependent variables. The pseudo coefficient of determination for each regression equation is presented in table 4.

One of the most common measures of accuracy is the root mean square error (table 4). The standard error is a

Table 3. Results of sensitivity analysis of regression equations for Vermont and vicinity presented as percent change in computed flood discharge as a result of a 10-percent change of the input basin characteristic.

Percent change in basin characteristic	Percent change in 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	0.1 percent	0.2 percent
	Drainage area							
-10	-9.05	-8.89	-8.81	-8.72	-8.66	-8.60	-8.56	-8.51
+10	8.96	8.79	8.70	8.60	8.54	8.48	8.43	8.38
	Percent of basin covered by wetlands							
-10	2.23	2.26	2.28	2.33	2.37	2.42	2.45	2.52
+10	-2.02	-2.04	-2.07	-2.11	-2.15	-2.19	-2.22	-2.28
	Basin-wide mean of the average annual precipitation							
-10	-15.2	-15.9	-16.3	-16.8	-17.0	-17.4	-17.6	-17.9
+10	16.1	16.9	17.4	18.0	18.4	18.8	19.2	19.6

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 vicinity.

[R^2, coefficient of determination; Log, logarithmic]

Flood discharge with an annual exceedance probability of:	Pseudo R^2	Root mean square error		Average standard error of prediction	
		Log units	Percent	Log units	Percent
50 percent	0.968	0.155	-29.9 to 42.8	0.147	-28.7 to 40.3
20 percent	0.965	0.164	-31.5 to 46.1	0.152	-29.5 to 41.8
10 percent	0.959	0.176	-33.3 to 49.9	0.162	-31.2 to 45.3
4 percent	0.950	0.193	-35.9 to 56.0	0.177	-33.4 to 50.2
2 percent	0.944	0.207	-37.9 to 61.1	0.186	-34.8 to 53.5
1 percent	0.937	0.221	-39.9 to 66.5	0.195	-36.2 to 56.6
0.5 percent	0.928	0.236	-41.9 to 72.1	0.208	-38.0 to 61.3
0.2 percent	0.914	0.255	-44.4 to 79.9	0.224	-40.3 to 67.4

measure of how much the regression results deviate from the observed data. The standard error is computed from the variance of the regression results with *n*-4 degrees of freedom (*n*=number of streamgages in analysis).

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 development of model parameters from samples of the population. Thus, the average standard error of prediction is a measure of the expected accuracy of a regression model applied at an ungaged location with basin characteristics similar to those used to develop the regression equation. This measure of accuracy is needed because regression equations typically are 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 pseudo coefficient of determination, the root mean square error, and the average standard error of prediction (table 4) are computed from 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. The measures of accuracy for an individual estimate include standard error of prediction and prediction intervals.

Standard Error of Prediction

Hodge and Tasker (1995) describe the mathematical formulation for computing the variance of prediction, V_{pred}, of a flood-discharge frequency estimate as

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

where

V_{pred} is the standard error of prediction;

γ^2 is the model error variance (see table 5);

x_i is a row vector containing 1, $\log_{10}(A)$, $\log_{10}(W)$, 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 a 1 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 in weighting a result from a regression equation with the AEP from the analysis of the streamgage record and is discussed in “Use of Regression Equations at Streamgages” and “Use of Regression Equations Near Streamgages.” A weighted AEP result at a streamgage location will have reduced uncertainty (U.S. Interagency Advisory Committee on Water Data, 1982).

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

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

where

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

V_{pred} is the variance of prediction.

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

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

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

where

S_{pos} is the positive percent error of prediction,

S_{neg} is the negative percent error of prediction, and

SE_{pred} is the standard error of prediction in logarithmic units.

The probability that the true value of flood discharge at a given frequency is between the positive- and negative-percent standard errors 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.

Confidence Intervals

Confidence intervals indicate the uncertainty in the result of the equations. For example, one can be 90 percent confident that the true value of a flood-discharge estimate lies within the 90-percent prediction interval. Confidence intervals for selected percentages can be computed as follows:

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

Table 5. Model error variance and the $\left(X^{tr}\Lambda^{-1}X\right)^{-1}$ matrices for the regression equations.

[AEP, Annual Exceedance Probability; Numbers in matrices are in scientific notation; Annual precipitation refers to the basin-wide mean of the average annual precipitation, 1981–2010]

Flood-frequency characteristic	Model error variance, γ^2		$\left(X^{tr}\Lambda^{-1}X\right)^{-1}$ matrix			
			Intercept	Drainage area	Percent wetland	Annual precipitation
Flood discharge with a 50-percent AEP	0.0206	Intercept	2.09943e-01	1.36263e-03	-5.91487e-03	-1.24244e-01
		Drainage area	1.36263e-03	3.59243e-04	-3.39318e-04	-1.04633e-03
		Percent wetland	-5.91487e-03	-3.39318e-04	1.90930e-03	3.18722e-03
		Annual precipitation	-1.24244e-01	-1.04633e-03	3.18722e-03	7.40248e-02
Flood discharge with a 20-percent AEP	0.0217	Intercept	2.40878e-01	1.52635e-03	-6.85297e-03	-1.42489e-01
		Drainage area	1.52635e-03	4.18743e-04	-3.60256e-04	-1.20956e-03
		Percent wetland	-6.85297e-03	-3.60256e-04	2.12288e-03	3.68150e-03
		Annual precipitation	-1.42489e-01	-1.20956e-03	3.68150e-03	8.49375e-02
Flood discharge with a 10-percent AEP	0.0247	Intercept	2.90305e-01	1.82790e-03	-8.30275e-03	-1.71713e-01
		Drainage area	1.82790e-03	5.00880e-04	-4.13511e-04	-1.46169e-03
		Percent wetland	-8.30275e-03	-4.13511e-04	2.50730e-03	4.45355e-03
		Annual precipitation	-1.71713e-01	-1.46169e-03	4.45355e-03	1.02383e-01
Flood discharge with a 4-percent AEP	0.0291	Intercept	3.65998e-01	2.29571e-03	-1.05209e-02	-2.16476e-01
		Drainage area	2.29571e-03	6.24036e-04	-4.96771e-04	-1.84666e-03
		Percent wetland	-1.05209e-02	-4.96771e-04	3.09918e-03	5.63544e-03
		Annual precipitation	-2.16476e-01	-1.84666e-03	5.63544e-03	1.29103e-01
Flood discharge with a 2-percent AEP	0.0322	Intercept	4.22383e-01	2.64122e-03	-1.21807e-02	-2.49815e-01
		Drainage area	2.64122e-03	7.14871e-04	-5.55686e-04	-2.13313e-03
		Percent wetland	-1.21807e-02	-5.55686e-04	3.53172e-03	6.51880e-03
		Annual precipitation	-2.49815e-01	-2.13313e-03	6.51880e-03	1.49007e-01
Flood discharge with a 1-percent AEP	0.0352	Intercept	4.80981e-01	3.00081e-03	-1.39098e-02	-2.84463e-01
		Drainage area	3.00081e-03	8.08573e-04	-6.16775e-04	-2.43038e-03
		Percent wetland	-1.39098e-02	-6.16775e-04	3.98077e-03	7.43891e-03
		Annual precipitation	-2.84463e-01	-2.43038e-03	7.43891e-03	1.69693e-01
Flood discharge with a 0.5-percent AEP	0.0399	Intercept	5.56104e-01	3.47496e-03	-1.61160e-02	-3.28906e-01
		Drainage area	3.47496e-03	9.28384e-04	-7.03556e-04	-2.81157e-03
		Percent wetland	-1.61160e-02	-7.03556e-04	4.57896e-03	8.61560e-03
		Annual precipitation	-3.28906e-01	-2.81157e-03	8.61560e-03	1.96218e-01
Flood discharge with a 0.2-percent AEP	0.0462	Intercept	6.59500e-01	4.12862e-03	-1.91591e-02	-3.90078e-01
		Drainage area	4.12862e-03	1.09262e-03	-8.22279e-04	-3.33641e-03
		Percent wetland	-1.91591e-02	-8.22279e-04	5.40065e-03	1.02384e-02
		Annual precipitation	-3.90078e-01	-3.33641e-03	1.02384e-02	2.32727e-01

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

where

CI_{upper} is the upper confidence interval, in cubic feet per second;

CI_{lower} is the lower confidence 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 confidence interval of a prediction; α = 0.05 for a 95-percent confidence interval of a prediction) with $n–p$ degrees of freedom; n = 145, the number of stations used in the regression analysis; and p = 4, the number of basin characteristics in the regression equation, plus 1; and

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

Use of Regression Equations at 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 by 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 can be found in *appendix 3*. Generally, the weighted estimate of flood discharge 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 with the following equation:

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

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 variance of prediction of the regression equation result ($Q_{r(g)}$) in logarithmic units computed with equation 11; and

V_s is the variance of estimate of the annual exceedance probability discharge in logarithmic units computed from the streamgage record (Q_s) (see *appendix 8*).

Confidence 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:

$$SE_w = \sqrt{\frac{V_s V_{pred}}{V_s + V_{pred}}} \qquad (18)$$

Use of Regression Equations Near Streamgages

Estimates of the magnitude of flood discharges at selected annual exceedance probabilities for ungaged sites that are not at, but are relatively 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 for adjusting the weighted discharge at a streamgage, Q_w, from equation 17 to a site of interest upstream or downstream from the streamgage is provided. The method increases the weight of a regression-equation-derived discharge estimate over the weight of a streamgage-data-derived discharge estimate the farther upstream or downstream the site of interest is from the streamgage. The method improves upon of the technique described in Ries (2007) by adjusting 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:

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

where

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

$Q_{r(g)}$ is the flood-discharge estimate generated by 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 by using equation 17, in cubic feet per second; and
a is the percentage of the gaged drainage area, in decimal units, 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, $a = 0.5$, and 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, $a = 1.5$, and the site of interest 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

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

An example of the use of this technique is in appendix 7. As with any technique used to compute a weighted flood-discharge estimate, unexpected results could occur if there is a substantial difference between the discharges being weighted. If the difference is substantial, c could become negative, indicating discharge and drainage area are inversely related. This procedure is not valid if c is negative.

Drainage-Area-Only Regression Equations

For some ungaged sites, the percentage of 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 (equations 3–10). The acceptable ranges of the basin characteristics are described 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 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 (equations 21–28) for estimating flood discharges on ungaged, unregulated streams in rural drainage basins in Vermont are as follows:

$$Q_{50} = 48.2A^{0.869}, \qquad (21)$$

$$Q_{20} = 77.3A^{0.855}, \qquad (22)$$

$$Q_{10} = 101A^{0.847}, \qquad (23)$$

$$Q_{4} = 135A^{0.838}, \qquad (24)$$

$$Q_{2} = 164A^{0.833}, \qquad (25)$$

$$Q_{1} = 197A^{0.827}, \qquad (26)$$

$$Q_{0.5} = 234A^{0.822}, \text{ and} \qquad (27)$$

$$Q_{0.2} = 289A^{0.816}, \qquad (28)$$

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 145 streamgages used to develop the previously presented regression equations were used to develop the simplified equations. Also, five additional stations (see appendix 6, footnote 1) were available for this regression analysis. However, the equations are still applicable to sites with drainage areas of 0.18 to 851 mi^2. Because they have only one explanatory variable, the simplified regression equations are less accurate than the full regression equations presented in this report. The pseudo coefficient of determination, the root mean square error, and the average standard error of prediction of the simplified equations are presented in table 6.

Although the accuracy of the drainage-area-only regression equations is relatively poor, these simplified equations are 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 in an alternate method for adjusting flood-frequency data from a streamgage to locations upstream and downstream. This use of the method is to be limited to sites within 50- to 150-percent of the streamgage drainage area (Wandle, 1983). Using this approach, one would use equation 19 with the exponent from the simplified regression equation at a selected recurrence interval substituted for c.

Table 6. Measures of accuracy of the simplified drainage-area-only regression equations for estimating flood discharges at selected annual exceedance probabilities for ungaged, unregulated streams in rural drainage basins in Vermont and vicinity.

[R^2, coefficient of determination; Log, logarithmic]

Flood discharge with an annual exceedance probability of:	Pseudo R^2	Root mean square error		Average standard error of prediction	
		Log units	Percent	Log units	Percent
50 percent	0.948	0.195	-36.2 to 56.7	0.187	-35.0 to 53.8
20 percent	0.942	0.205	-37.6 to 60.2	0.195	-36.2 to 56.7
10 percent	0.936	0.215	-39.1 to 64.1	0.202	-37.2 to 59.2
4 percent	0.927	0.231	-41.3 to 70.3	0.214	-38.9 to 63.7
2 percent	0.917	0.244	-43.0 to 75.4	0.226	-40.6 to 68.3
1 percent	0.910	0.258	-44.7 to 81.0	0.232	-41.4 to 70.6
0.5 percent	0.900	0.271	-46.4 to 86.7	0.243	-42.9 to 75.0
0.2 percent	0.885	0.290	-48.7 to 94.8	0.257	-44.7 to 80.7

Vermont StreamStats

StreamStats, a World Wide Web application (*http://water.usgs.gov/osw/streamstats/*), allows users to obtain discharge statistics, drainage-basin characteristics, 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, the user will get previously published information for the site from a database. If a user selects an ungaged site, a GIS program 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 presented in a table along with a map showing the basin outline. Historically, determining the basin characteristics and solving the regression equations for an ungaged site could take an experienced person hours. StreamStats reduces the effort to only a few minutes.

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

The equations published in this report will be available online in Vermont StreamStats immediately following 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 by using the final regression equations (equations 3–10).

Summary

This report, prepared by the U.S. Geological Survey in cooperation with the Federal Emergency Management Agency, documents the development of regression equations for estimating flood-discharge magnitudes for rural, unregulated streams in Vermont and adjacent areas of New Hampshire, Massachusetts, and New York at annual exceedance probabilities of 50-, 20-, 10-, 4-, 2-, 1-, 0.2-, and 0.5-percent. Regression techniques were used to determine relations between the flood discharge magnitudes and selected basin characteristics at 145 streamgages in and adjacent to Vermont.

The flood discharge magnitudes at selected recurrence intervals for the 145 streamgages were determined by following guidelines in Bulletin 17B of the U.S. Interagency Advisory Committee on Water Data with the exceptions that the Expected Moments Algorithm (EMA) methods were used to incorporate historical and censored flood information and that a different low outlier test was used. A new generalized skew coefficient of 0.44 with a mean square error of 0.078 was developed for the frequency analysis.

A total of 120 basin characteristics for each streamgage were determined by using a geographic information system. By using correlation data, stepwise linear regression techniques, and generalized least-squares regression techniques, the 120 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 by using generalized least-squares regression techniques. The average standard error of prediction for estimating peak discharges with

50-, 20-, 10-, 4-, 2-, 1-, 0.2-, and 0.5-percent annual exceedance probability with these equations are 34.9, 36.0, 38.7, 42.4, 44.9, 47.3, 50.7, and 55.1 percent, respectively.

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

The equations and flood-discharge frequency data used in this study are available in StreamStats, a World Wide Web application (*http://water.usgs.gov/osw/streamstats/*) providing statistics, drainage-basin characteristics, and other information for user-selected sites on streams.

Vermont Regional Skew Regression

By Andrea G. Veilleux

For the log-transformation of annual peak discharges, Bulletin 17B (U.S. Interagency Advisory Committee on Water Data, 1982) recommends using a weighted average of the station skew coefficient and a regional skew coefficient to improve estimates of annual flood-probability discharges (AFPDs). Bulletin 17B supplies a national map but also encourages hydrologists to develop more specific local relations. Since the first map was published in 1976, some 36 years of additional information has accumulated, and better spatial estimation procedures have been developed (Stedinger and Griffis, 2008). For the Vermont study, a regression analysis was done to develop a regional skew.

Reis and others (2005), Gruber and others (2007), and Gruber and Stedinger (2008) developed a Bayesian generalized least-squares (GLS) regression model for regional skewness analyses. The method provides a more reasonable description of the model error variance than either the generalized least-squares method-of-moments or maximum likelihood point estimates (Veilleux, 2011). However, because of complications introduced by the use of the expected moments algorithm (EMA), with multiple Grubbs-Beck censoring of low outliers (Cohn and others, 1997) and large cross-correlations between annual peak discharges at pairs of streamgages, an alternate regression procedure was developed to provide stable and defensible results for regional skew regression (Veilleux and others, 2012; Veilleux, 2011; Lamontange and others, 2012). This alternate procedure is referred to as the Bayesian weighted least-squares/Bayesian generalized least-squares (B-WLS/B-GLS) regression framework (Veilleux and others, 2012; Veilleux, 2011; Veilleux and others, 2011).

The B-WLS/B-GLS regression analysis uses an ordinary least-squares (OLS) analysis to fit an initial regional skewness model; the OLS model is then used to generate a regional skew-coefficient estimate for each streamgage. This regional estimate is the basis for computing the variance of each station skew-coefficient estimator employed in the weighted least-squares (WLS) analysis. Then, B-WLS is used to generate estimators of the regional skew-coefficient model parameters. Finally, B-GLS is used to estimate the precision of those WLS parameter estimators, to estimate the model error variance and the precision of that variance estimator, and to compute various diagnostic statistics. The methodology for the regional skewness model is described in detail in Eash and others (2013).

This regional skew study is based on annual peak-discharge data from 112 streamgages in Vermont and the surrounding states that were included in the regional regression analysis described in the main report and additional streamgages just beyond the regional regression study area. The additional streamgages were added to increase the number of available streamgages in the skew analysis. In addition to following the criteria required for the peak-discharge frequency analysis, the initial streamgage selection for the skew analysis also limited the streamgages to those with at least 20 years of record. A streamgage drainage basin within a larger streamgage drainage basin, with the difference between the drainage areas less than 50 percent, was also not allowed.

Because the dataset includes censored data and historic information, the effective record length used to compute the precision of the skewness estimators is no longer simply the number of annual peak discharges at a streamgage. Instead, a more complex calculation is used to take into account the availability of historic information and censored values. While historic information and censored peaks provide valuable information, they often provide less information than an equal number of years with systematically recorded peaks (Stedinger and Cohn, 1986). The calculations made to compute the pseudo record length, P_{RL}, are described in Eash and others (2013). P_{RL} equals the systematic record length if such a complete record is all that is available for a site.

As stated in Bulletin 17B, the skew coefficient of the station is sensitive to extreme events, and more accurate estimates can be obtained from longer records. Thus, after ensuring adequate special and hydrologic coverage, those streamgages that do not have a minimum of 30 years of P_{RL} were removed from the regional skew study. Of the 112 streamgages, 40 were removed because of a P_{RL} less than 30 years. Thus, 72 streamgages remained from which to build a regional skewness model for the Vermont study area.

The station logarithmic skew coefficient, G, (*appendix 9*) and its mean square error, MSEG, were computed by using EMA (Cohn and others, 1997; Griffis and others, 2004). The streamgage skewness estimates are ensured to be unbiased by using the correction factor developed by Tasker and Stedinger (1986) and employed in Reis and others (2005). In addition to the skew data, 28 basin characteristics for each of the

112 streamgages were available as explanatory variables for the regional study. The basin characteristics available included land cover and climatic characteristics, as well as the more standard morphometric characteristics such as location of the basin centroid, drainage area, main basin slope, and mean drainage basin elevation, among others.

A cross-correlation model for the log annual peak discharges in the Vermont study area was developed by using streamgages with at least 55 years of concurrent systematic peaks. The model, termed the Fisher Z Transformation and described further in Eash and others (2013), provided an estimate of the cross-correlations of concurrent annual peak discharge at two streamgages, , using the distance between basin centroids, D_{ij}, as the explanatory variable:

$$\rho_{ij} = \frac{exp\left(2Z_{ij}\right)-1}{exp\left(2Z_{ij}\right)+1}, \tag{29}$$

where

$$Z_{ij} = exp\left(0.368 - 0.0695\left(\frac{D_{ij}^{0.503}-1}{0.503}\right)\right) \tag{30}$$

The cross-correlation model was used to estimate site-to-site cross correlations for concurrent annual peak discharges at all pairs of streamgages in the regional skew study.

Vermont Regional Skew Study Results

The Vermont generalized skew is a constant value of 0.44 over the study region. The results of the Vermont regional skew study using the B-WLS/B-GLS regression methodology are provided in table 7.

All of the available basin characteristics were initially considered as explanatory variables in the regression analysis for regional skew. Available basin characteristics include climate measures (mean annual precipitation,100-year/24-hour rainfall, 100-year/60-minute rainfall, mean annual maximum and minimum temperatures), soil properties (available water storage), basin measures (drainage area, mean basin elevation, mean basin slope, relief, percentage of the basin over 1,200 feet in elevation), and land cover (percentage characterized as forest, open water, wetland, field, shrubs, barren, or developed). None of the basin characteristics were statistically significant in explaining the site-to-site variability in skewness. Thus, the best model, as classified by having the smallest model error variance, σ_δ^2, and largest pseudo R_δ^2 is the constant model. Table 7 provides the final results for the constant skewness model denoted “Constant.”

Table 7. Statistics for the constant model of generalized skew statistics for Vermont and vicinity.

[Standard deviations are in parentheses; b_1, the constant model result for skew; σ_δ^2, the model error variance; ASEV, the average sampling error variance; AVP_new, the average variance of prediction for a new site; Pseudo R_δ^2 (%), the fraction of the variability in the true skews explained by each model in percent (%) (Gruber and others, 2007)]

Model	b_1	σ_δ^2	ASEV	AVP_new	Pseudo R_δ^2 (%)
Constant	0.44 (0.12)	0.06 (0.03)	0.015	0.078	0

Table 7 includes the pseudo R_δ^2, which describes the estimated fraction of the variability in the true skewness from site-to-site explained by each model (Gruber and others, 2007; Parrett and others, 2011). A constant model does not explain any variability, so the pseudo R_δ^2 equals 0. The posterior mean of the model error variance, σ_δ^2, for the Constant model is $\sigma_\delta^2 = 0.06$.

The addition of any of the available basin characteristics (none of which are statistically significant) did not produce a pseudo R_δ^2 greater than 13 percent or decrease the model error variance. This indicates that the inclusion of a basin characteristic as an explanatory variable in the regression did not help explain the variability in the true skewness. Thus, the addition of a basin characteristic is not warranted because the increased model complexity does not result in a gain in model precision. Thus, the Constant model is chosen as the best regional skewness model for the Vermont study area. The average sampling error variance (ASEV) in table 7 is the average error in the regional skewness estimator at the streamgages in the dataset. The average variance of prediction at a new streamgage (AVP_new) corresponds to the mean square error (MSE) used in Bulletin 17B to describe the precision of the generalized skewness. The Constant model has an AVP_new, equal to 0.078, which corresponds to an effective record length of 102 years. An AVP_new of 0.078 is a marked improvement over the Bulletin 17B national skew map, for which the reported MSE is 0.302 (U.S. Interagency Committee on Water Data, 1982) for a corresponding effective record length of only 17 years. Thus the new regional model has six times the information content (as measured by effective record length) of that calculated for the Bulletin 17B map.

Pseudo Analysis of Variance (Pseudo ANOVA) statistics for the Vermont regional skew analysis were determined as additional diagnostics for the Constant model. Explanations of how the statistics were computed can be found in Eash and others (2013). The Error Variance Ratio (EVR) is a modeling diagnostic used to evaluate whether a simple OLS regression is sufficient or a more sophisticated WLS or GLS analysis is appropriate. EVR is the ratio of the average sampling error variance to the model error variance. Generally, an EVR greater than 0.20 indicates that the sampling variance is not negligible when compared to the model error variance, indicating the need for a WLS or GLS regression analysis.

For the Vermont study-area data, EVR had a value of 2.2 for the constant model. The sampling variability in the sample skewness estimators was larger than the error in the regional model. Thus an OLS model that neglects sampling error in the streamgage skewness estimators may not provide a statistically reliable analysis of the data. Given the variation of record lengths from site to site, it is important to use a WLS or GLS analysis to evaluate the final precision of the model rather than using a simpler OLS analysis.

The Misrepresentation of the Beta Variance (MBV*) statistic is used to determine whether a WLS regression is sufficient or if a GLS regression is appropriate to determine the precision of the estimated regression parameters (Veilleux, 2011; Griffis, 2006). For the Vermont regional skew study, the MBV* is equal to 5.8 for the constant model. This is a large value, indicating the cross-correlation among the skewness estimators has had an effect on the precision with which the regional average skew coefficient can be estimated; if a WLS precision analysis were used for the estimated constant parameter in the constant model, the variance would be underestimated by a factor of 5.8. Thus, a WLS analysis would seriously misrepresent the variance of the constant in the Constant model. Moreover, a WLS model would result in underestimation of the variance of prediction, given that the sampling error in the constant term in both models was sufficiently large to make an appreciable contribution to the average variance of prediction.

Selected References

Belsley, D.A., Kuh, Edwin, and Welsch, R.E., 1980, Regression diagnostics—Identifying influential data and sources of collinearity: New York, Wiley, 292 p. [See chap. 2, p. 6–84.]

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

Cohn, T.A., 2011, PeakfqSA Version 0.960 (software), accessed March 25, 2014, at *http://www.timcohn.com/TAC_Software/PeakfqSA/*.

Cohn, T.A., Berenbrock, Charles, 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 *http://pubs.usgs.gov/fs/2012/3038/*.]

Cohn, T.A., Lane, W.L., and Baier, W.G., 1997, An algorithm for computing moments-based flood quantile estimates when historical flood information is available: Water Resources Research, v. 33, no. 9, p. 2089–2096.

Cohn, T.A., Lane, W.L., and Stedinger, J.R., 2001, Confidence intervals for expected moments algorithm flood quantile estimates: Water Resources Research, v. 37, no. 6, p. 1695–1706.

Cook, R.D., and Weisberg, Sanford, 1982, Residuals and influence in regression: New York, Chapman and Hall, 230 p.

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.

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.

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., *http://pubs.usgs.gov/sir/2013/5086/*.

Eng, Ken, Chen, Yin-Yu, and Kiang, J.E., 2009, User's guide to the weighted-multiple-linear-regression program (WREG version 1.0): U.S. Geological Survey Techniques and Methods, book 4, chap. A8, 21 p.

Environmental Systems Research Institute, Inc. (Esri), 2007, ArcGIS 9.2 desktop help: Environmental Systems Research Institute, Inc., Web page, *http://webhelp.esri.com/arcgisdesktop/9.2*.

Environmental Systems Research Institute, Inc. (Esri), 2008, GIS for water resources: Esri Web page, accessed April 1, 2008, at *http://www.esri.com/industries/water_resources/index.html/*.

Feaster, T.D., Gotvald, A.J., and Weaver, J.C., 2009, Magnitude and frequency of rural floods in the southeastern United States, 2006—Volume 3, South Carolina: U.S. Geological Survey Scientific Investigations Report 2009–5156, 226 p.

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. 4B, 42 p.

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

Fry, J.A., Xian, George, Jin, Suming, Dewitz, J.A., Homer, C.G., Yang, Limin, Barnes, C.A., Herold, N.D., and Wickham, J.D., 2011, Completion of the 2006 national land cover database for the conterminous United States, PE&RS, v. 77, no. 9, p. 858–864.

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.

Griffis, V.W., 2006, Flood frequency analysis—Bulletin 17, regional information, and climate change: Ithaca, N.Y., Cornell University, unpublished Ph.D. dissertation.

Griffis, V.W., and Stedinger, J.R., 2009, Log-Pearson type 3 distribution and its application in flood frequency analysis, III—Sample skew and weighted skew estimators: Journal of Hydrology, v. 14, no. 2, p. 121–130.

Griffis, V.W., Stedinger, J.R., and Cohn, T.A., 2004, LP3 quantile estimators with regional skew information and low outlier adjustments: Water Resources Research, v. 40, no. 7, [17 p.], *http://dx.doi.org/10.1029/2003WR002697*.

Gruber, A.M., Reis, D.S., Jr., and Stedinger, J.R., 2007, Models of regional skew based on Bayesian GLS regression, Paper 40927–3285, *in* Kabbes, K.C., ed., Restoring our natural habitat—Proceedings of the 2007 World Environmental and Water Resources Congress—May 15–19, Tampa, Florida: Reston, Va., American Society of Civil Engineers [variously paged].

Gruber, A.M., and Stedinger, J.R., 2008, Models of LP3 regional skew, data selection and Bayesian GLS regression, Paper 596, *in* Babcock, R.W., and Walton, Raymond, eds., World Environmental and Water Resources Congress 2008–Ahupua'a—Proceedings of the congress—May 12–16, 2008, Honolulu, Hawaii: Reston, Va., American Society of Civil Engineers [variously paged].

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

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

Hoaglin, D.C., and Welsch, R.E., 1978, The hat matrix in regression and ANOVA: The American Statistician, v. 32, no. 1, p. 17–22.

Hodge, S.A., and Tasker, G.D., 1995, Magnitude and frequency of floods in Arkansas: U.S. Geological Survey Water-Resources Investigations Report 95–4224 [variously paged; see p. 37–39].

Jacobs, Jennifer, and Jardin, Patrick, 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.

Lamontagne, J.R., Stedinger, J.R., Berenbrock, Charles, Veilleux, A.G., Ferris, J.C., and Knifong, D.L., 2012, Development of regional skews for selected flood durations for the Central Valley Region, California, based on data through water year 2008: U.S. Geological Survey Scientific Investigations Report 2012–5130, 60 p.

Martin, R.O.R., and Hanson, R.L., 1966, Reservoirs in the United States: U.S. Geological Survey Water-Supply Paper 1838, 115 p.

Martins, E.S., and Stedinger, J.R., 2002, Cross-correlation among estimators of shape: Water Resources Research, v. 38, no. 11, p. 34–1–34–7, *http://dx.doi.org/10.1029/2002WR001589*.

Milly, P.C.D., Betancourt, Julio, Falkenmark, Malin, Hirsch, R.H., Kundzewicz, Z.W., Lettenmaier, D.P., and Stouffer, R.J., 2008, Stationarity is dead—Whither water management?: Science, v. 319, no. 5,863, p. 573–574, accessed August 19, 2008, at *http://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.

National Weather Service, 2013, Vermont climatology: National Oceanic and Atmospheric Administration, National Weather Service, accessed April 3, 2013, at *http://www.erh.noaa.gov/btv/climo/stations/*.

Natural Resources Conservation Service, 2001, Watershed boundary dataset, 1:24,000 scale: United States Department of Agriculture, National Resources Conversation Service, accessed October 18, 2013, at *http://datagateway.nrcs.usda.gov/*.

Natural Resources Conservation Service, 2012, Gridded soil survey geographic (gSSURGO) by state, 10 meter resolution: United States Department of Agriculture, National Resources Conversation Service, accessed October 18, 2013, at *http://datagateway.nrcs.usda.gov/*.

Northeast Regional Climate Center, 2013, Extreme precipitation in New York & New England, 30 arc-second gridded data: Northeast Regional Climate Center, accessed October 24, 2013, at *http://precip.eas.cornell.edu*.

Olson, S.A., 2002, Flow-frequency characteristics of Vermont streams: U.S. Geological Survey Water-Resources Investigations Report 02–4238, 47 p.

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., *http://dx.doi.org/10.3133/sir20135187/*.

Parrett, Charles, Veilleux, A.G., 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.

Potter, W.D., 1957a, Peak rates of runoff in the Adirondack, White Mountains, and Maine Woods area: U.S. Department of Commerce, Bureau of Public Roads [variously paged].

Potter, W.D., 1957b, Peak rates of runoff in the New England hill and lowland area: U.S. Department of Commerce, Bureau of Public Roads [variously paged].

PRISM Group, Oregon State University, 2012a, United States average monthly or annual precipitation, 1981–2010, 30 arc-second normal, created July 10, 2012: Oregon State University, PRISM Climate Group, accessed September 16, 2013, at *http://www.prism.oregonstate.edu/*.

PRISM Group, Oregon State University, 2012b, United States average monthly or annual maximum temperature, 1981–2010, 4 km grid cell resolution, created July 10, 2012: Oregon State University, PRISM Climate Group, accessed September 16, 2013, at *http://www.prism.oregonstate.edu/*.

PRISM Group, Oregon State University, 2012c, United States average monthly or annual minimum temperature, 1981–2010, 4 km grid cell resolution, created July 10, 2012: Oregon State University, PRISM Climate Group, accessed September 16, 2013, at *http://www.prism.oregonstate.edu/*.

Randall, A.D., 1996, Mean annual runoff, precipitation, and evapotranspiration in the glaciated northeastern United States, 1951–80: U.S. Geological Survey Open-File Report 96–395, 2 pls., accessed October 18, 2013, at *http://ny.water.usgs.gov/pubs/of/of96395/OF96-395.html*. [Geographic information system coverages added in 1998 by Andrew J. Cohen, accessed October 18, 2013, at *http://water.usgs.gov/GIS/metadata/usgswrd/XML/ofr96395_eva.xml*, *http://water.usgs.gov/GIS/metadata/usgswrd/XML/ofr96395_pre.xml*, and *http://water.usgs.gov/GIS/metadata/usgswrd/XML/ofr96395_run.xml*.]

Reis, D.S., Jr., Stedinger, J.R., and Martins, E.S., 2005, Bayesian generalized least squares regression with application to the log Pearson type 3 regional skew estimation: Water Resources Research, v. 41, no. 10, [14 p.], *http://dx.doi.org/10.1029/2004WR003445*.

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, book 4, chap. A6, 37 p.

Riggs, H.C., 1968a, Some statistical tools in hydrology: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A1, 39 p.

Riggs, H.C., 1968b, Frequency curves: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A2, 15 p.

Riggs, H.C., 1973, Regional analyses of streamflow characteristics: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. B3, 15 p.

SAS Institute, Inc., 1990, SAS/STAT User’s guide, version 6 (4th ed.) v. 1 and v. 2: Cary, N.C., SAS Institute, 1,686 p.

SAS Institute, Inc., 2000, Statistics I: Introduction to ANOVA, regression, and logistic regression, Course notes: Cary, N.C., SAS Institute, 504 p.

Stedinger, J.R., and Cohn, T.A., 1986, Flood frequency analysis with historical and paleoflood information: Water Resources Research, v. 22, no. 5, p. 785–793.

Stedinger, J.R., and Griffis, V.W., 2008, Flood frequency analysis in the United States—Time to update [editorial]: Journal of Hydrologic Engineering, v. 13, no. 4 [April], p. 199–204.

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.

Tasker, G.D., and Stedinger, J.R., 1986, Regional skew with weighted LS regression: Journal of Water-Resources Planning and Management, v. 112, no. 2, p. 225–237.

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

Thomson, M.T., Gannon, W.B., Thomas, M.P., and Hayes, G.S., 1964, Historical floods in New England: U.S. Geological Survey Water-Supply Paper 1779–M, 105 p.

U.S. Fish and Wildlife Service, 2009, National wetlands inventory: U.S. Fish and Wildlife Service, accessed October 18, 2013, at *http://www.fws.gov/wetlands/*.

U.S. Geological Survey, 1995, Soils data for the conterminous United States derived from the NRCS State soil geographic (STATSGO) data base: U.S. Geological Survey Water Resources National Spatial Data Infrastructure Node Web page, accessed October 18, 2013, at *http://water.usgs.gov/GIS/metadata/usgswrd/XML/ussoils.xml/*.

U.S. Geological Survey, 2001, USGS digital raster graphics: U.S. Geological Survey Topographic Maps Web page, accessed October 18, 2012, at *http://topomaps.usgs.gov/drg/*.

U.S. Geological Survey, 2004a, National elevation dataset, 1 arc-second: U.S. Geological Survey National Elevation Dataset, accessed October 18, 2013, at *http://ned.usgs.gov/*.

U.S. Geological Survey, 2004b, National hydrography dataset, high resolution: U.S. Geological Survey, The National Map, National Hydrography Dataset, accessed October 18, 2013, at *http://nhd.usgs.gov/*.

U.S. Geological Survey, 2012, Computation of annual exceedance probability (AEP) for characterization of observed flood peaks: U.S. Geological Survey Office of Surface Water Technical Memorandum 2013.01, 7 p., accessed April 8, 2013, at *http://water.usgs.gov/admin/memo/SW/sw13.01.pdf*.

U.S. Geological Survey, 2013a, Water resources application software, PeakFQ: U.S. Geological Survey Water Resources of the United States Web page, accessed April 3, 2013, at *http://water.usgs.gov/software/peakfq.html*.

U.S. Geological Survey, 2013b, Peak streamflow for the Nation: National Water Information System, accessed April 3, 2013, at *http://nwis.waterdata.usgs.gov/usa/nwis/peak*.

U.S. Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency—Bulletin #17B of the Hydrology Subcommittee (revised and corrected): Reston, Va., U.S. Geological Survey, Office of Water Data Coordination, 28 p. [plus appendixes].

Veilleux, A.G., 2009, Bayesian GLS regression for regionalization of hydrologic statistics, floods and Bulletin 17 skew: Ithaca, N.Y., Cornell University, unpublished M.S. Thesis.

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

Veilleux, A.G., Stedinger, J.R., and Eash, D.A., 2012, Bayesian WLS/GLS regression for regional skewness analysis for regions with large crest stage gage networks, Paper 227, *in* Loucks, E.D., ed., World Environmental and Water Resources Congress 2012—Crossing boundaries—Proceedings of the 2012 congress, May 20–24, Albuquerque, N.M.: Reston, Va., American Society of Civil Engineers, p. 2253–2263.

Veilleux, A.G., Stedinger, J.R., and Lamontagne, J.R., 2011, Bayesian WLS/GLS regression for regional skewness analysis for regions with large cross-correlations among flood flows, *in* Beighley, R.E., and Kilgore, M.W., eds., World Environmental and Water Resources Congress 2011—Bearing knowledge for sustainability—Proceedings of the 2011 World Environmental and Water Resources Congress, May 22–26, Palm Springs, Calif.: Reston, Va., American Society of Civil Engineers, p. 3103–3112.

Vermont Fish and Wildlife Department, 2010, The Vermont fishing experience—Good fishing comes naturally in Vermont: Vermont Fish and Wildlife Department Web page, accessed September 13, 2013, at *http://www.vtfishandwildlife.com/library/Reports_and_Documents/Vermont_Digest_of_Hunting_Fishing_and_Trapping_Laws/Archive/2010/__Sect%204%20Fishing%20Info%20pages%2039-53.pdf*.

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

Weaver, J.C., Feaster, T.D., and Gotvald, A.J., 2009, Magnitude and frequency of rural floods in the southeastern United States, through 2006—Volume 2, North Carolina: U.S. Geological Survey Scientific Investigations Report 2009–5158, 111 p.

Appendixes 1–9 available at *http://pubs.usgs.gov/sir/2014/5078/*

Appendix 1. Descriptions of Streamgages in Vermont and Vicinity Used to Develop the Regional Regression Equations

Appendix 2. Summary of Data Used in the Frequency Analysis of Annual Peak-Discharge Data

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

Appendix 4. Maximum Recorded Annual Peak Discharge at Streamgages in Vermont and Vicinity Used to Develop the Regression Equations

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

Appendix 6. Basin Characteristics Used to Develop the Regression Equations

Appendix 7. Example Application

Appendix 8. Variance of Estimates (Vs) at Selected Annual Exceedance Probabilities for Streamgages in Vermont and Vicinity

Appendix 9. Streamgages Evaluated for Development of the Generalized Skew Used in Estimating Flood Flow Frequency in Vermont and Surrounding Areas

Prepared by the Pembroke and West Trenton
Publishing Service Centers.

For more information concerning this report, contact:

Office Chief
U.S. Geological Survey
New England Water Science Center
New Hampshire-Vermont Office
331 Commerce Way, Suite 2
Pembroke, NH 03275
dc_nh@usgs.gov

or visit our Web site at:
http://nh.water.usgs.gov

USGS

Olson—**Estimation of Flood Discharges at Selected Annual Exceedance Probabilities for Unregulated, Rural Streams in Vermont**—Scientific Investigations Report 2014–5078

ISSN 2328-0328 (online)
http://dx.doi.org/10.3133/sir20145078