USGS
science for a changing world

Prepared in cooperation with the Montana Department of Natural Resources and Conservation

Methods for Estimating Peak-Flow Frequencies at Ungaged Sites in Montana Based on Data through Water Year 2011

Chapter F of
Montana StreamStats

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

Methods for Estimating Peak-Flow Frequencies at Ungaged Sites in Montana Based on Data through Water Year 2011

By Roy Sando, Steven K. Sando, Peter M. McCarthy, and DeAnn M. Dutton

Chapter F of
Montana StreamStats

Scientific Investigations Report 2015–5019–F
Version 1.1, February 2018

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

U.S. Department of the Interior
RYAN K. ZINKE, Secretary

U.S. Geological Survey
William H. Werkheiser, Deputy Director
    exercising the authority of the Director

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

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

Contents

Figures

Tables

Appendix Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
Length		
inch (in.)	2.54	centimeter (cm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
Area		
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
Flow rate		
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
inches per month	2.54	centimeters per month

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

$$°F=(1.8×°C)+32$$

Datum

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

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

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

Supplemental Information

Water year is the 12-month period from October 1 through September 30 of the following calendar year. The water year is designated by the calendar year in which it ends. For example, water year 2011 is the period from October 1, 2010, through September 30, 2011.

Abbreviations

±	plus or minus
α	significance level
A	contributing drainage area
AEP	annual exceedance probability
CI	confidence interval
DA	drainage area
E	mean basin elevation
E_{5000}	percent of drainage basin above 5,000 feet elevation
E_{6000}	percent of drainage basin above 6,000 feet elevation
ET	evapotranspiration
ET_{SPR}	mean spring (March–June) evapotranspiration
F	percent of drainage basin with forest land cover
G	specific streamflow-gaging station of interest
GIS	geographic information system
GLS	generalized least squares
i	specific site of interest
K	regression constant
LCC2000	Land Cover, circa 2000-vector
MEV	model error variance
MODIS	Moderate Resolution Imaging Spectroradiometer
MOD16	Moderate Resolution Imaging Spectroradiometer global evapotranspiration product

MVP	mean variance of prediction
n	number of streamflow-gaging stations
NED	National Elevation Dataset
NHD	National Hydrography Dataset
NLCD	National Land Cover Dataset
NWIS	National Water Information System
OLS	ordinary least squares
P	mean annual precipitation
p	number of explanatory variables
p-value	statistical probability level
PRISM	Parameter-elevation Regression on Independent Slopes Model
Q_{AEP}	peak-flow magnitude for indicated annual exceedance probability (AEP), where the AEP value is in percent
R^2	coefficient of determination
SEM	mean standard error of model
SEP	mean standard error of prediction
SLP_{30}	percentage of drainage basin with slope greater than or equal to 30 percent
U	specific ungaged site of interest
USGS	U.S. Geological Survey
VIF	variance inflation factor
WBD	Watershed Boundary Dataset
WLS	weighted least squares
WREG	Weighted-Multiple-Linear Regression Program
x	basin characteristic value or vector of basin characteristic values
$\left(X^T \Lambda^{-1} X\right)^{-1}$	covariance matrix for the generalized least squares regional regression equation

Acknowledgments

Methods for Estimating Peak-Flow Frequencies at Ungaged Sites in Montana Based on Data through Water Year 2011

By Roy Sando, Steven K. Sando, Peter M. McCarthy, and DeAnn M. Dutton

Abstract

The U.S. Geological Survey (USGS), in cooperation with the Montana Department of Natural Resources and Conservation, completed a study to update methods for estimating peak-flow frequencies at ungaged sites in Montana based on peak-flow data at streamflow-gaging stations through water year 2011. The methods allow estimation of peak-flow frequencies (that is, peak-flow magnitudes, in cubic feet per second, associated with annual exceedance probabilities of 66.7, 50, 42.9, 20, 10, 4, 2, 1, 0.5, and 0.2 percent) at ungaged sites. The annual exceedance probabilities correspond to 1.5-, 2-, 2.33-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence intervals, respectively.

Regional regression analysis is a primary focus of Chapter F of this Scientific Investigations Report, and regression equations for estimating peak-flow frequencies at ungaged sites in eight hydrologic regions in Montana are presented. The regression equations are based on analysis of peak-flow frequencies and basin characteristics at 537 streamflow-gaging stations in or near Montana and were developed using generalized least squares regression or weighted least squares regression.

All of the data used in calculating basin characteristics that were included as explanatory variables in the regression equations were developed for and are available through the USGS StreamStats application (http://water.usgs.gov/osw/streamstats/) for Montana. StreamStats is a Web-based geographic information system application that was created by the USGS to provide users with access to an assortment of analytical tools that are useful for water-resource planning and management. The primary purpose of the Montana StreamStats application is to provide estimates of basin characteristics and streamflow characteristics for user-selected ungaged sites on Montana streams. The regional regression equations presented in this report chapter can be conveniently solved using the Montana StreamStats application.

Selected results from this study were compared with results of previous studies. For most hydrologic regions, the regression equations reported for this study had lower mean standard errors of prediction (in percent) than the previously reported regression equations for Montana. The equations presented for this study are considered to be an improvement on the previously reported equations primarily because this study (1) included 13 more years of peak-flow data; (2) included 35 more streamflow-gaging stations than previous studies; (3) used a detailed geographic information system (GIS)-based definition of the regulation status of streamflow-gaging stations, which allowed better determination of the unregulated peak-flow records that are appropriate for use in the regional regression analysis; (4) included advancements in GIS and remote-sensing technologies, which allowed more convenient calculation of basin characteristics and investigation of many more candidate basin characteristics; and (5) included advancements in computational and analytical methods, which allowed more thorough and consistent data analysis.

This report chapter also presents other methods for estimating peak-flow frequencies at ungaged sites. Two methods for estimating peak-flow frequencies at ungaged sites located on the same streams as streamflow-gaging stations are described. Additionally, envelope curves relating maximum recorded annual peak flows to contributing drainage area for each of the eight hydrologic regions in Montana are presented and compared to a national envelope curve. In addition to providing general information on characteristics of large peak flows, the regional envelope curves can be used to assess the reasonableness of peak-flow frequency estimates determined using the regression equations.

Introduction

Reliable information on peak-flow characteristics at specific sites is essential for many water-resources applications including effective planning and management of water resources and flood plains, protection of lives and property in flood-prone areas, determination of actuarial flood-insurance rates, and design of highway infrastructure. Peak-flow data are readily available at sites that are monitored by streamflow-gaging stations (hereinafter referred to as gaging stations) and can be downloaded through the U.S. Geological Survey (USGS) National Water Information System (NWIS;

http://waterdata.usgs.gov/nwis; U.S. Geological Survey, 2015a). Streamflow data from gaging stations can be statistically analyzed to estimate peak-flow frequencies (that is, peak-flow magnitudes, in cubic feet per second, with annual exceedance probabilities (AEPs) of 66.7, 50, 42.9, 20, 10, 4, 2, 1, 0.5, and 0.2 percent). The AEPs correspond to 1.5-, 2-, 2.33-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence intervals, respectively. Sando, McCarthy, and Dutton (2016) reported peak-flow frequencies for 725 gaging stations in Montana based on data through water year 2011 (water year is the 12-month period from October 1 through September 30 and is designated by the year in which it ends). For many water-resources applications, the peak-flow frequencies also are needed at ungaged sites. Peak-flow frequencies can be estimated for ungaged sites using various methods, including regional regression analysis. Regional regression analysis involves standard multivariate regression techniques that analyze relations between peak-flow frequencies and physical basin characteristics (such as contributing drainage area and mean basin elevation), as well as climatic basin characteristics (such as mean annual precipitation).

Previous reports of methods for estimating peak-flow frequencies at ungaged sites in Montana include Berwick (1958), Parrett and Omang (1981), Omang and others (1986), Omang (1992), and Parrett and Johnson (2004). The most recent report (Parrett and Johnson, 2004) was based on data through water year 1998. Changing climatic conditions, increasing periods of data collection, new gaging stations, and improved analytical methods necessitate periodic updates of the regional regression equations. Thus, the USGS, in cooperation with the Montana Department of Natural Resources and Conservation, completed a study to update methods for estimating peak-flow frequencies at ungaged sites in Montana based on peak-flow data at gaging stations through water year 2011.

Purpose and Scope

The study described in Chapter F of this Scientific Investigations Report is part of a larger study to develop a StreamStats application for Montana, compute streamflow characteristics at gaging stations, and develop regional regression equations to estimate streamflow characteristics at ungaged sites (as described fully in Chapters A through G of this Scientific Investigations Report). The purpose of Chapter F is to describe methods for estimating peak-flow frequencies in Montana, with emphasis on estimating peak-flow frequencies at ungaged sites. Regional regression analysis is a primary focus of this report chapter, which documents the development of regression equations (for eight hydrologic regions in Montana) that are based on peak-flow frequencies (Sando, McCarthy, and Dutton, 2016) and basin characteristics at 537 gaging stations (fig. 1, table 1–1 in appendix 1 at the back of this report chapter [available at https://doi.org/10.3133/sir20155019F]); map numbers assigned according to McCarthy and others [2016]) in or near

Montana. The regression equations were developed using generalized least squares (GLS) regression (Tasker and Stedinger, 1989) or weighted least squares (WLS) regression (Tasker, 1980) and can be used to estimate peak-flow frequencies at ungaged sites in eight hydrologic regions (fig. 1) in Montana.

This report chapter also presents other methods for estimating peak-flow frequencies at ungaged sites. Two methods for estimating peak-flow frequencies at ungaged sites located on the same streams as gaging stations are described. Additionally, envelope curves relating maximum recorded annual peak flows to contributing drainage area for each of eight hydrologic regions in Montana are presented and compared to a national envelope curve and regional regression lines for 1-percent AEP peak flows (Q_1).

General Flood Characteristics in Montana

Montana is a large (approximately 147,000 square miles [mi^2]) State with diverse topographic and climatic conditions creating highly variable hydrologic characteristics. The western part of Montana generally consists of rugged, mountainous terrain sometimes separated by large, intermontane valleys, whereas the eastern part of Montana is characterized by rolling or flat plains, interspersed with areas of deeply incised streams and rugged relief referred to as "badlands" or "breaks." Most of the mountainous, western part of Montana is in the Canadian, Northern, and Middle Rockies ecoregions, whereas most of the nonmountainous, eastern part is in the Northwestern Glaciated Plains and Northwestern Great Plains ecoregions (Woods and others, 2002). Elevations in Montana range from about 12,800 feet (ft) above the North American Vertical Datum of 1988 (NAVD 88) in some mountain ranges to about 1,800 ft in eastern plains areas and in the Kootenai River Basin in extreme northwest Montana. The general elevation information was based on a geographic information system (GIS) analysis of the National Elevation Dataset (NED; Gesch and others, 2002). Mean annual precipitation also is highly variable and ranges from about 110 inches (in.) in some mountainous areas of western Montana to about 10 in. generally in low-altitude plains areas (PRISM Climate Group, 2004).

In this report chapter, the terms "flood" and "annual peak flows" are used in the discussion of high-streamflow characteristics. A flood is any high streamflow that overtops the natural or artificial banks of a river. An annual peak flow is the annual maximum instantaneous discharge recorded for each water year that an individual gaging station is operated. A given annual peak flow might not overtop the river banks and thus might not qualify as a flood. "Peak flow" is used in reference to high-streamflow characteristics at gaging stations; "flood" or "flooding" is used in more general reference to high-streamflow characteristics of an area or hydrologic region.
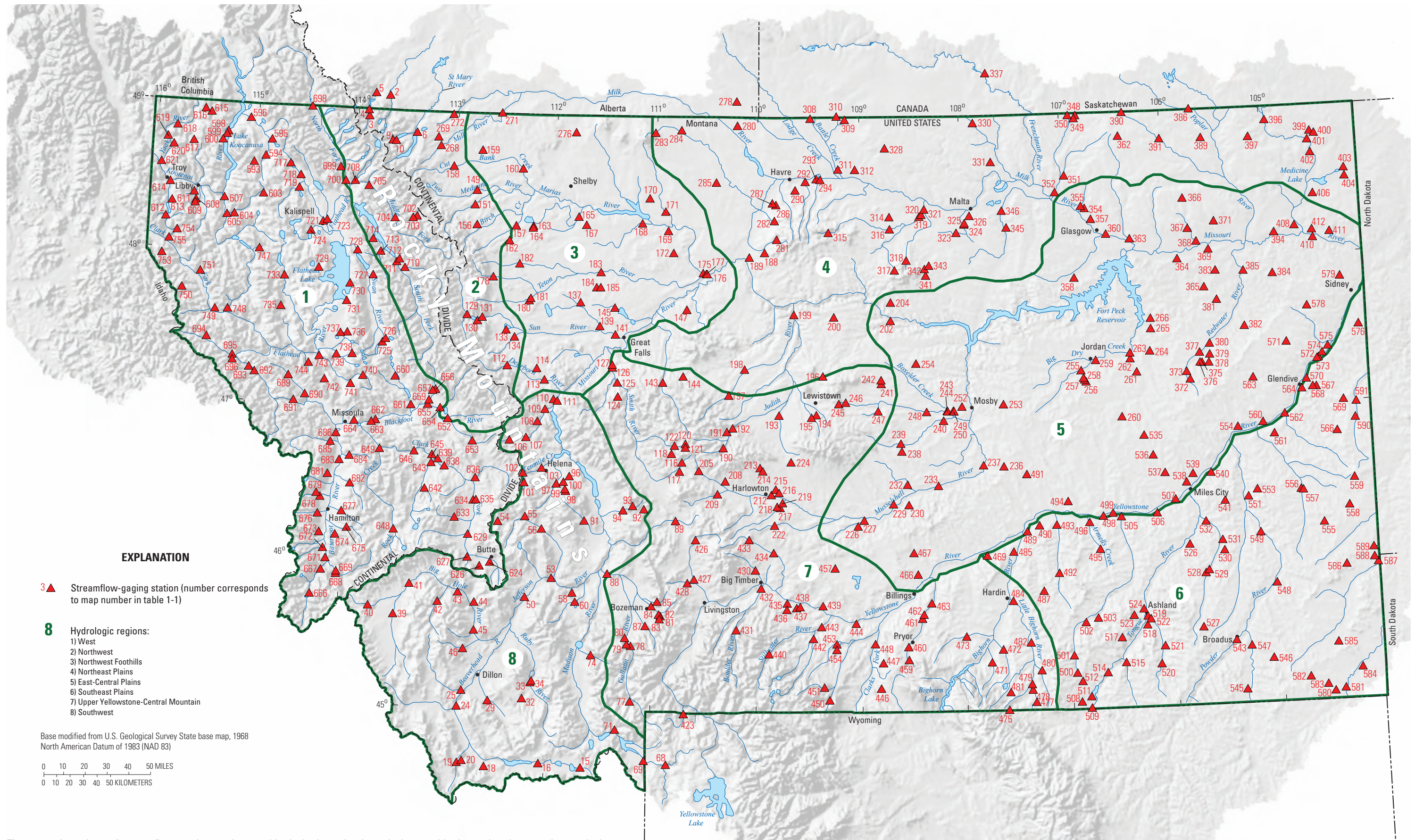
Figure 1. Locations of streamflow-gaging stations and hydrologic region boundaries used in the regional regression analysis.

EXPLANATION

3 ▲ Streamflow-gaging station (number corresponds to map number in table 1-1)

8 Hydrologic regions:
1) West
2) Northwest
3) Northwest Foothills
4) Northeast Plains
5) East-Central Plains
6) Southeast Plains
7) Upper Yellowstone-Central Mountain
8) Southwest

Base modified from U.S. Geological Survey State base map, 1968
North American Datum of 1983 (NAD 83)

0   10   20   30   40   50 MILES
0   10   20   30   40   50 KILOMETERS

Flooding in Montana primarily is affected by topography and the source and timing of precipitation events and snowmelt (Parrett and Johnson, 2004). General flood characteristics for selected hydrologic regions in Montana are presented in table 1. Frequency distributions of proportions of annual peak flows in each month (that is, the monthly timing of peak flows) for all gaging stations in each hydrologic region are shown in figure 2. In much of western Montana, most of the annual precipitation falls as snow in the winter and comes from moist air masses that originate over the Pacific Ocean. Thus, flooding generally is the result of mountain snowmelt runoff, frequently combined with rainfall runoff, in May and June. Winter rains or rain onto melting snow in western Montana valleys can occasionally cause substantial flooding, and intense summer thunderstorms can occasionally cause flooding. On the eastern slopes of the Continental Divide, severe flooding sometimes results from large May or June rains that originate from moist air masses from the Gulf of Mexico. Although these rains generally dissipate as the moist air is uplifted over the crest of the Continental Divide, the largest storms have crossed the divide and caused severe flooding on the western slopes as well as the eastern slopes (Boner and Stermitz, 1967).

Flooding in the plains and breaks of eastern Montana is less predictable (Parrett and Johnson, 2004). Large storms that result in flooding might come from the Pacific Ocean or Gulf of Mexico. In some years, flooding in this area might result from snowmelt runoff in the spring or snowmelt combined with rain over the plains. Intense summer thunderstorms can sometimes cause flooding on the plains. Flooding in eastern Montana tends to be more variable, both spatially and temporally, than in western Montana because precipitation from large storms is more variable. Thunderstorms are more prevalent in eastern Montana than in western Montana, and thunderstorms are highly variable in terms of extent, location, and precipitation amounts and intensities.

Peak-Flow Frequencies at Streamflow-Gaging Stations

The USGS has been collecting and publishing annual peak-flow records at gaging stations in Montana for more than 100 years (U.S. Geological Survey, 2015; table 1–1). Sando, McCarthy, and Dutton (2016) determined peak-flow frequencies for 725 gaging stations in and near Montana that had at least 10 years of systematic record based on data through water year 2011. Methods of data compilation and analysis are described by Sando, McCarthy, and Dutton (2016). These methods relate to determination of the regulation status of gaging stations, data compilation and pre-analysis manipulation, and peak-flow frequency analysis.

Methods for Estimating Peak-Flow Frequencies at Ungaged Sites in Montana

The USGS, in cooperation with the Montana Department of Natural Resources and Conservation, updated methods for estimating peak-flow frequencies at ungaged sites in Montana based on peak-flow data at gaging stations through water year 2011, which is the focus of this report chapter. The development and results of the updated methods are described in the following sections.

Regional Regression Analysis and Results

Regional regression analysis involves determining relations between peak-flow frequencies and basin characteristics at gaging stations to estimate peak-flow frequencies at ungaged sites. Various procedures used in the regional regression analysis are described in the following subsections.

Selection of Streamflow-Gaging Stations Used in the Regional Regression Analysis

Sando, McCarthy, and Dutton (2016) determined peak-flow frequencies for 725 gaging stations in or near Montana that had at least 10 years of systematic record using methods described by the U.S. Interagency Advisory Council on Water Data (1982), commonly referred to as Bulletin 17B. The 725 gaging stations were screened for suitability for inclusion in the regional regression analysis for the study described in this report chapter based on the following criteria: (1) contributing drainage area less than about 2,750 mi^2, (2) peak-flow records unaffected by major regulation, (3) small redundancy with nearby gaging stations, and (4) representation of peak-flow frequencies at sites within Montana.

The criterion of contributing drainage area less than about 2,750 mi^2 serves to restrict the regional regression analysis to smaller streams that might not be represented by data from gaging stations. Typically, most streams with contributing drainage areas larger than about 2,750 mi^2 have one or more gaging stations on the stream channel, and the gaged records can be used to provide peak-flow frequency estimates at ungaged locations on those streams. Thus, only gaging stations with contributing drainage areas less than about 2,750 mi^2 were included in the regional regression analysis.

Reservoir storage and operations have the potential to substantially affect streamflow characteristics, and peak-flow data affected by regulation is unsuitable for the regional regression analysis. The USGS maintains a geospatial database of dams in Montana (McCarthy and others, 2016) that was used to define the regulation status for Montana gaging stations. The

Table 1.   Hydrologic regions and flood characteristics in Montana (modified from Parrett and Johnson, 2004).

Hydrologic region (ordered clockwise from northwestern Montana)	Hydrologic region number in figure 1	General description and extent	Flood characteristics
West	1	Mountains and valleys west of Continental Divide; parts of Flathead and Blackfoot River Basins	Most floods caused by snowmelt or snowmelt mixed with rain. Annual peak flows less variable than in other regions.
Northwest	2	Eastern parts of Flathead and Blackfoot River Basins; mountains and foothills east of the Continental Divide and northeast of Missoula, Montana	Largest floods caused by runoff from rain associated with moist air masses from the Gulf of Mexico. Most annual peak flows are from snowmelt or snowmelt mixed with rain.
Northwest Foothills	3	Foothills and plains of the Marias, Teton, Sun, and Dearborn River Basins near Great Falls, Montana	Floods caused by snowmelt, large amounts of rain, or thunderstorms. Annual peak flows are more variable than those from similar-sized streams in the mountainous regions.
Northeast Plains	4	Rolling plains of the Milk River Basin upstream from Glasgow; foothills and plains part of the Judith River Basin	Floods on larger streams caused by prairie snowmelt or snowmelt mixed with rain. Most floods on smaller streams caused by thunderstorms. Annual peak flows are more variable than those from streams in the Northwest Foothills region.
East-Central Plains	5	Plains and badlands of the lower parts of Musselshell, Missouri, Milk, and Poplar River Basins; northern part of Yellowstone River Basin east of Billings, Montana	Floods on larger streams caused by prairie snowmelt or snowmelt mixed with rain. Most floods on smaller streams caused by thunderstorms. Thunderstorms are more prevalent and intense than in any other region. Annual peak flows are more variable than in any other region.
Southeast Plains	6	Rolling plains of southern part of Yellowstone River Basin east of Billings, Montana	Floods on larger streams caused by prairie snowmelt or snowmelt mixed with rain. Most floods on smaller streams caused by thunderstorms. Annual peak flows are somewhat less variable and smaller than those from similar-sized streams in the East-Central Plains region.
Upper Yellowstone-Central Mountain	7	Mountains and valleys of the upper Yellowstone River Basin; mountains and valleys of the Smith River Basin; parts of the Judith and Musselshell River Basins	Floods caused by snowmelt or snowmelt mixed with rain on larger streams and snowmelt or thunderstorms on smaller streams. Annual peak flows are similar to, though more variable than, those in the West region.
Southwest	8	Mountains and valleys of the Missouri River Basin upstream from the Dearborn River	Floods caused by snowmelt or snowmelt mixed with rain on larger streams and snowmelt or thunderstorms on smaller streams. Annual peak flows generally are smaller and more variable than those from similar-sized streams in other mountainous regions.
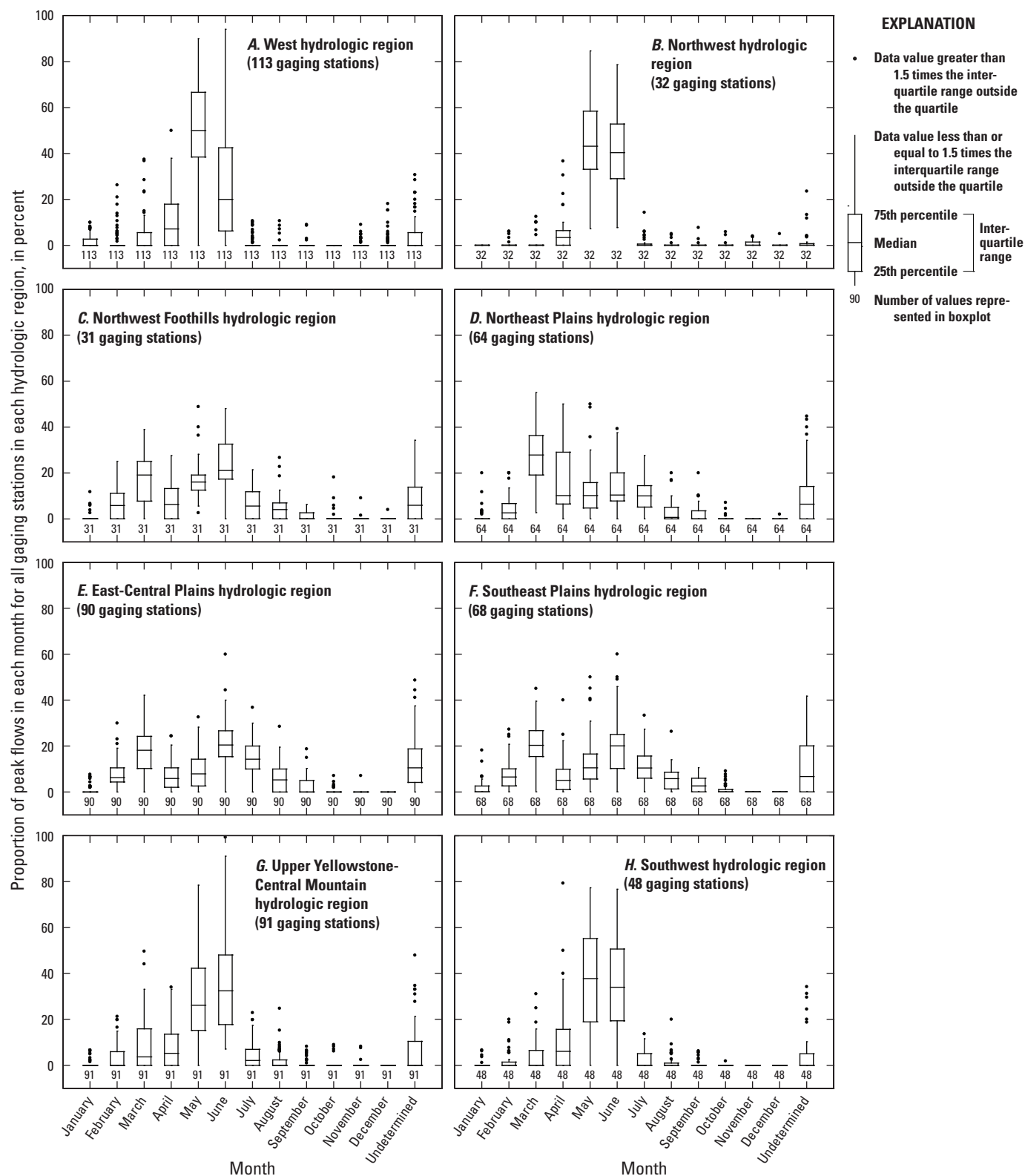
Figure 2. Statistical distributions of proportions of peak flows in each month for all streamflow-gaging stations in each hydrologic region. *A*, West hydrologic region; *B*, Northwest hydrologic region; *C*, Northwest Foothills hydrologic region; *D*, Northeast Plains hydrologic region; *E*, East-Central Plains hydrologic region; *F*, Southeast Plains hydrologic region; *G*, Upper Yellowstone-Central Mountain hydrologic region; and *H*, Southwest hydrologic region.

specific methods used for this study to determine the regulation classification of gaging stations in Montana are described by McCarthy and others (2016). Based on the USGS regulation-classification criteria used for this study, a gaging station is considered to be unregulated if the cumulative drainage area of all upstream dams is less than 20 percent of the drainage area of the gaging station and no large diversion canals are upstream from the gaging station. A gaging station is considered to be regulated if the cumulative drainage area of all upstream dams exceeds 20 percent of the drainage area of the given gaging station. If the drainage area of a single upstream dam exceeds 20 percent of the drainage area of a given gaging station, the regulation is classified as major. If no single upstream dam has a drainage area that exceeds 20 percent of the drainage area of a given gaging station, the regulation is classified as minor. In the regional regression analysis, peak-flow frequency estimates affected by major regulation were excluded. For this study, in cases where a large diversion canal was known to be located on the channel upstream from a gaging station, the gaging station also was considered to have major regulation, and affected peak-flow frequency estimates were excluded from the regional regression analysis. In some cases, a gaging station had peak-flow records before and after the construction of major regulation structures; peak-flow frequency estimates for the unregulated period were included in the regional regression analysis. Gaging stations classified as having minor regulation also were included in the regional regression analysis.

A redundant gaging-station analysis was conducted to account for spatial autocorrelation in peak-flow records of gaging stations located on the same stream channel. In cases where a gaging station was located on a large tributary upstream from a gaging station on a primary stream channel, the two gaging stations were considered to be on the same stream channel in the redundant gaging-station analysis. If there were multiple gaging stations on the same stream channel, the drainage areas of the gaging stations were examined. If two adjacent gaging stations on the same stream channel had drainage areas that were within about 0.5–2.0 times the other gaging station, the gaging station with the shortest period of record was usually excluded from the regional regression analysis; however, if excluding the gaging station with the longer period of record allowed for the inclusion of an additional gaging station because another instance of redundancy was eliminated, the gaging station with the longer period of record was excluded.

The drainage basins of some of the gaging stations included in Sando, McCarthy, and Dutton (2016) are largely or entirely outside of Montana. Some of those gaging stations were excluded from the regional regression analysis if their drainage basins were considered to provide poor representation of peak-flow frequencies in Montana or if there were potential effects from undocumented regulation in the basin.

Of the 725 gaging stations with peak-flow frequencies reported by Sando, McCarthy, and Dutton (2016), 537 gaging stations met the screening criteria and were selected for inclusion in the regional regression analysis. Information on the 537 selected gaging stations is presented in table 1–1 in appendix 1 at the back of this report chapter.

Basin Characteristics

Basin characteristics investigated as potential explanatory variables in the regional regression analyses were selected based on previous studies (Berwick, 1958; Parrett and Omang, 1981; Omang and others, 1986; Omang, 1992; and Parrett and Johnson, 2004), theoretical relations with peak flows, and the ability to generate the characteristics using GIS analysis and digital datasets. In previous regional regression studies for Montana, basin characteristics were manually estimated using paper topographic maps and overlaying transparent gridded cells on the maps. In previous studies, the number of candidate basin characteristics has ranged from 2 (Berwick, 1958) to 12 (Parrett and Omang, 1981). For this study, 28 basin characteristics were selected as candidate variables in the regression analyses and are presented in table 2. Because of the nonlinear relation between streamflow and the explanatory variables, all data were log-transformed prior to analysis. Additionally, the basin characteristics of mean basin elevation (E), maximum basin elevation, minimum basin elevation, and relief (maximum minus minimum elevation of drainage basin) were divided by 1,000 prior to analysis to get coefficients that are comparable in magnitude to other basin characteristics. Also, a value of one was added to basin characteristics that are presented as a percentage of the basin (that is, percentage of drainage basin above 5,000 ft elevation [E_{5000}], 5,500 ft elevation, 6,000 ft elevation [E_{6000}], 6,500 ft elevation, and 7,000 ft elevation; percentage of drainage basin with forest land cover [F], urban land cover, and wetland land cover; percentage of drainage basin in lakes, ponds, and reservoirs; percentage of drainage basin with north-facing slopes greater than or equal to 30 percent; and percentage of drainage basin with slopes greater than 30 percent [SLP_{30}] and 50 percent) to allow for log-transformation of basin characteristic values that were previously zero.

Of the 28 candidate basin characteristics, 7 were determined to have significant (p-value less than 0.05) relations with peak-flow characteristics (table 2) and were used in the final regression equations. The most consistently important basin characteristic was contributing drainage area (A), which was used in all of the regression equations. Other basin characteristics determined to be significant and used in the final regression equations of one or more of the hydrologic regions include E_{5000}, E_{6000}, mean spring (March–June) evapotranspiration (ET_{SPR}), F, mean (1971–2000) annual precipitation (P), and SLP_{30}.

Drainage basins were delineated using a combination of 30-meter digital elevation data from the NED (Gesch and others, 2002) and the Watershed Boundary Dataset (WBD) obtained from the National Hydrography Dataset (NHD) version 2 (Horizon Systems Corporation, 2013). The data for each candidate basin characteristic were converted into a digital

Table 2. Selected basin characteristics considered as potential candidate explanatory variables in the regional regression equations.

[*, denotes statistically significant (*p*-value less than 0.05) coefficient in the regional regression equations of at least one of the hydrologic regions]

Basin characteristic	StreamStats abbreviation	Abbreviation for basin characteristics discussed in this report	Description
Percent agricultural land	AG_OF_DA	--	Percentage of drainage area with agricultural land cover[1].
Mean basin slope	BSLDEM30M	--	Mean basin slope computed as the first derivative of the 30-meter elevation dataset[2].
Compactness ratio	COMPRAT	--	A measure of basin shape related to basin perimeter and drainage area.
Contributing drainage area*	CONTDA	A	Area that contributes flow to a point on a stream, in square miles, delineated using 30-meter elevation data[2].
Percent above 5,000 feet*	EL5000	E_{5000}	Percentage of drainage basin above 5,000 feet elevation[2].
Percent above 5,500 feet	EL5500	--	Percentage of drainage basin above 5,500 feet elevation[2].
Percent above 6,000 feet*	EL6000	E_{6000}	Percentage of drainage basin above 6,000 feet elevation[2].
Percent above 6,500 feet	EL6500	--	Percentage of drainage basin above 6,500 feet elevation[2].
Percent above 7,000 feet	EL7000	--	Percentage of drainage basin above 7,000 feet elevation[2].
Mean basin elevation	ELEV	E	Mean elevation of drainage basin, in feet[2].
Maximum basin elevation	ELEVMAX	--	Maximum elevation of drainage basin, in feet[2].
Mean spring evapotranspiration*	ET0306MOD	ET_{SPR}	Mean spring (March–June) evapotranspiration, in inches per month[3].
Mean summer evapotrasnpiration	ET0710MOD	--	Mean summer (July–October) evapotranspiration, in inches per month[3].
Percent forest*	FOREST	F	Percentage of drainage basin with forest land cover[1].
Percent lakes and ponds	LAKEAREA	--	Percentage of drainage basin in lakes, ponds, and reservoirs[4].
Mean March temperature	MARAVTMP	--	Mean (1971–2000) March air temperature, in degrees Fahrenheit[5].
Mean maximum March temperature	MARMAXTMP	--	Mean (1971–2000) March maximum air temperature, in degrees Fahrenheit[5].
Mean annual maximum temperature	MAXTEMP	--	Mean (1971–2000) annual maximum air temperature, in degrees Fahrenheit[5].
Minimum basin elevation	MINBELEV	--	Minimum drainage basin elevation, in feet[2].
Mean annual minimum temperature	MINTEMP	--	Mean (1971–2000) annual minimum air temperature, in degrees Fahrenheit[5].
North facing slopes greater than 30 percent	NFSL30_30M	--	Percentage of drainage basin with north-facing slopes greater than or equal to 30 percent computed from 30-meter elevations data[2].
Basin perimeter	PERIMMI	--	Basin perimeter, in miles.
Mean annual precipitation*	PRECIP	P	Mean (1971–2000) annual precipitation, in inches[5].
Relief	RELIEF	--	Maximum minus minimum elevation of drainage basin, in feet[2].
Slopes greater than 30 percent*	SLOP30_30M	SLP_{30}	Percentage of drainage basin with slopes greater than or equal to 30 percent, computed from the 30-meter elevation data[2].
Slopes greater than 50 percent	SLOP50_30M	--	Percentage of drainage basin with slopes greater than or equal to 50 percent computed from the 30-meter elevation data[2].
Percent urban area	URBAN	--	Percentage of drainage basin with urban land cover[1].
Percent wetlands	WETLAND	--	Percentage of drainage basin with wetland land cover[1].

[1] Land cover variables determined from the 2001 National Land Cover Dataset (NLCD; Homer and others, 2007) and Land Cover, circa 2000-vector (LCC2000; Natural Resources Canada, 2009).

[2] Elevation and related variables determined or calculated from the National Elevation Dataset (NED; Gesch and others, 2002). Elevation refers to distance above North American Vertical Datum of 1988.

[3] Evapotranspiration determined from the Moderate Resolution Imaging Spectroradiometer (MODIS) global evapotranspiration product (MOD16) data (Mu and others, 2007).

[4] Percentage of drainage basin in lakes, ponds, or reservoirs determined from the National Hydrography Dataset (NHD) version 2 high resolution dataset (Horizon Systems Corporation, 2013).

[5] Precipitation and air temperature variables determined from climatic datasets obtained from Parameter-elevation Regression on Independent Slopes Model (PRISM) data (PRISM Climate Group, 2004) and Long Term Mean Climate Grids for Canada (Natural Resources Canada, 2015) for 1971–2000.

grid or raster format and overlaid on the basin boundaries for each gaging station using standard tools available in ArcMap (Esri, Inc., 2014). The data could then be summarized for each gaging station and its associated basin. All of the data used in calculating basin characteristics that were used as explanatory variables in the final regression equations are available through the USGS StreamStats Program (http://water.usgs.gov/osw/streamstats/; U.S. Geological Survey, 2015b) application for Montana. Basin characteristics for ungaged basins can be calculated using the StreamStats tool described in the following paragraph.

StreamStats is a Web-based GIS application that was created by the USGS to provide users with access to an assortment of analytical tools that are useful for water-resource planning and management (U.S. Geological Survey, 2015a). StreamStats was designed for national application, with local USGS water science centers responsible for developing and processing the necessary geospatial data, computing streamflow characteristics, and developing regional regression equations to be deployed within StreamStats. StreamStats is accessed through a map-based user interface to make GIS-based estimation of streamflow characteristics easier, faster, and more consistent than previously used manual techniques. Also, GIS-based calculation of basin characteristics allows consideration of many more basin characteristics potentially affecting streamflow characteristics than had previously been possible. The primary purpose of the Montana StreamStats application is to provide estimates of basin characteristics and streamflow characteristics for user-selected ungaged sites on Montana streams (McCarthy and others, 2016). Additional information about StreamStats usage and limitations can be accessed at the StreamStats Web site (http://water.usgs.gov/osw/streamstats/).

To estimate the peak-flow frequencies for 18 gaging stations used in the regression analyses, the peak-flow records were augmented by combining peak-flow records from two or more closely located gaging stations (typically with drainage areas within about 5 percent) on the same channel (Sando, McCarthy, and Dutton, 2016). To determine the basin characteristics for an individual augmented gaging station, the basin characteristics of the closely located gaging stations were combined by applying a weighted mean of the basin characteristic values on the basis of peak-flow record length that was contributed to the augmented dataset.

Definition of Hydrologic Region Boundaries for Montana

Definition of the hydrologic region boundaries for Montana was based on exploratory analysis in conjunction with consideration of the regional boundaries from the previous reporting of methods for estimating peak-flow frequencies at ungaged sites in Montana (Parrett and Johnson, 2004). Initially, peak-flow frequencies and basin characteristics relations were investigated on a statewide basis. A type of

all-subsets ordinary least squares (OLS) regression was done (on a dataset that included all of the 537 selected gaging stations and the 28 candidate basin characteristics [table 2]) by using the Exploratory Regression tool in ArcGIS Desktop 10.2 (Esri, Inc., 2014). The exploratory OLS regression analysis determined that three basin characteristics (A, relief, and mean [1971–2000] annual precipitation) provided the best multivariate regression equation, as determined by comparison of pseudo coefficient of determination (R^2) values, and combined to account for about 60 percent of the variability in peak-flow frequencies on a statewide basis in Montana. Peak-flow magnitudes for the 2-, 1-, and 0.5-percent AEPs were then predicted with an OLS regression equation using the three explanatory variables. The 537 gaging stations were then separated into eight groups based on iterative K nearest-neighbor (Altman, 1992) spatially constrained cluster analyses of the standardized residuals from the OLS analyses of the 2- and 1-percent AEPs using the Grouping Analysis tool in ArcGIS Desktop 10.2 (Esri, Inc., 2014). The groups were then plotted in conjunction with the hydrologic region boundaries defined by Parrett and Johnson (2004). Spatial patterns in the groups generally were well represented by the hydrologic region boundaries; however, in some cases, the residuals for an individual gaging station located near the boundary of two adjacent hydrologic regions were larger than typical, which indicated that minor adjustments to the hydrologic region boundaries might provide improvements in the regional regression equations. Thus, during the final stages of regression equation development, minor adjustments were made by moving a few gaging stations to adjacent hydrologic regions and appropriately redefining the hydrologic region boundaries. The eight hydrologic regions used in the regional regression analysis are (ordered clockwise from northwestern Montana) (1) West hydrologic region, (2) Northwest hydrologic region, (3) Northwest Foothills hydrologic region, (4) Northeast Plains hydrologic region, (5) East-Central Plains hydrologic region, (6) Southeast Plains hydrologic region, (7) Upper Yellowstone-Central Mountain hydrologic region, and (8) Southwest hydrologic region (fig. 1).

Exploratory Data Analysis

Initially, for each hydrologic region, relations of peak-flow frequencies and basin characteristics were investigated using a type of all-subsets OLS regression analysis in the Exploratory Regression tool in ArcGIS Desktop 10.2 (Esri, Inc., 2014). The all subsets regression analysis in the Exploratory Regression tool incorporates several statistical diagnostic methods. In the selection of best-fit regression equations, the analysis considered (1) the adjusted R^2, (2) the statistical significance of the coefficients of the explanatory variables (as determined by a p-value less than 0.05), (3) the cross-correlation of explanatory variables (as determined by the variance inflation factor [VIF; Helsel and Hirsch, 2002]), (4) the normality of the residuals (as determined by the Jarque-Bera test; Jarque and Bera, 1987), and (5) the

spatial autocorrelation of the residuals (as determined by the global Moran's I Index value; Moran, 1950). A nonparametric random forest analysis (Breiman, 2001), with all 28 candidate basin characteristics (table 2) included, also was done to further assess multivariate and univariate importance of explanatory variables. The exploratory analyses were done on all AEPs, but when evaluating the results, emphasis was placed on the 2- and 1-percent AEPs. Results of the exploratory analyses were used to identify a best-fit OLS regression equation with the most important and consistent combination of candidate basin characteristics for each hydrologic region. Selection of the best-fit OLS regression equation for each hydrologic region primarily was based on the regression equation with the largest adjusted R^2 while also having (1) explanatory variables with significant coefficients (p-value less than 0.05) for the regression equations for either the 2- or 1-percent AEP peak flows, (2) a VIF value less than 2, and (3) residuals with a nonsignificant (p-value greater than 0.05) global Moran's I Index value. Other considerations in selection of the best-fit OLS regression equation included investigation of (1) the explanatory variables used in regional regression equations from previous studies (Omang, 1992; Parrett and Johnson, 2004), (2) the normality of the explanatory variables, (3) the Akaike Information Criterion (Akaike, 1973) of each regression equation, and (4) the hydrologic basis for relations between the peak-flow frequencies and the explanatory variables.

Regional Regression Analysis

After selection of best-fit OLS regression equations for each hydrologic region, final regression equations for seven hydrologic regions (the West, Northwest Foothills, Northeast Plains, East-Central Plains, Southeast Plains, Upper Yellowstone-Central Mountain, and Southwest hydrologic regions) were developed with GLS regression (Tasker and Stedinger, 1989). Final regression equations for one hydrologic region (the Northwest hydrologic region) were developed with WLS regression. All GLS and WLS regression analyses were conducted using the Weighted-Multiple-Linear Regression Program (WREG; Eng and others, 2009). Because differences between OLS regression and GLS or WLS regression can potentially affect relative importance among explanatory variables that might be spatially autocorrelated, the basin characteristics in the best-fit OLS regression equations initially were verified as also representing best-fit GLS or WLS regression equations. Emphasis was placed on the best-fit GLS or WLS regression equations for 2- or 1-percent AEP peak flows in each hydrologic region.

Generalized Least Squares Regression Analysis

GLS regression, unlike OLS regression, considers the time-sampling error and the interstation correlation of the dependent variable (that is, a peak-flow magnitude for the indicated annual exceedance probability [Q_{AEP}]). Two

assumptions of OLS regression that commonly are violated in regional regression analyses are that annual peak flows have constant variance, or homoscedasticity, and are independent from site to site, or no spatial autocorrelation. The assumption of homoscedasticity typically is violated because the variance is somewhat dependent on the length and timing of the systematic record, which often varies between gaging stations. The assumption of no spatial autocorrelation commonly is violated because of cross correlation between concurrent peak flows for different gaging stations. The GLS regression procedure takes into consideration the time-sampling error in the peak-flow series (heteroscedasticity) and the interstation correlation (spatial autocorrelation) between sites, and thus overcomes the violation of assumptions that can happen when applying OLS regression to regional streamflow studies. The GLS regression procedure also provides better estimates of the predictive accuracy of peak-flow frequencies that are computed by the regression equations and also provides almost unbiased estimates of the variance of the underlying regression equation error (Tasker and Stedinger, 1989). Thus, GLS regression generally results in equations that are more reliable than those developed by OLS regression for this purpose.

The WREG procedures for GLS regression allow the fitting of a smoothing function that describes the general relations between peak-flow series and geographic distance among the streamgages to assist in compensating for spatial autocorrelation. Appropriate smoothing functions were able to be developed for the seven hydrologic regions (the West, Northwest Foothills, Northeast Plains, East-Central Plains, Southeast Plains, Upper Yellowstone-Central Mountain, and Southwest hydrologic regions) for which GLS regression was applied; however, an appropriate smoothing function could not be developed for the Northwest hydrologic region because it had strong spatial autocorrelation among a large proportion of the streamgages that largely was independent of spatial distance between individual streamgages. Thus, regression equations for the Northwest hydrologic region were developed using WLS regression analysis, as described in the following section "Weighted Least Squares Regression Analysis."

Weighted Least Squares Regression Analysis

The final regression equations for the Northwest hydrologic region were developed with WLS regression (Tasker, 1980) using the WREG Program (Eng and others, 2009). In WLS regression, weights are assigned such that streamgages that have more reliable estimates of peak-flow frequencies (typically a function of the length of the period of record) have larger weights. Thus, WLS regression considers time-sampling error, but unlike GLS regression, it does not consider interstation correlation.

Regional Regression Equations

For each gaging station, the data for the dependent variables (Q_{AEP}) and explanatory variables (basin characteristics) that were used in developing the final GLS or WLS regression equations are presented in tables 1–2 and 1–3, respectively, in appendix 1 at the back of this report chapter (available at https://doi.org/10.3133/sir20155019F). Because of nonlinear relations between Q_{AEP} and the explanatory variables, all variables were transformed before analysis. Thus, the regression equations are of the following log-linear form:

$$\log Q_{AEP} = \log K + a_1 \log x_1 + a_2 \log x_2 + \ldots a_p \log x_p, \quad (1)$$

where

Q_{AEP}	is the peak flow, in cubic feet per second, with an annual exceedance probability (AEP) in percent;
K	is a regression constant;
p	is the number of explanatory variables (basin characteristics);
a_1 through a_p	are regression coefficients; and
x_1 through x_p	are values of the explanatory variables (basin characteristics).

Equation 1 can be expressed in terms of the actual variable values rather than logarithms as

$$Q_{AEP} = K' x_1^{a_1} x_2^{a_2} \ldots x_p^{a_p}, \quad (2)$$

where K' is the antilog (10^K) of the linear regression constant and all other terms are as previously described.

For each hydrologic region, the final GLS or WLS regression equations for estimating peak-flow frequencies at ungaged sites in Montana are presented in table 1–4 in appendix 1 at the back of this report chapter (available at https://doi.org/10.3133/sir20155019F). Included in table 1–4 are measures of reliability of the equations, including the model error variance (σ_δ^2, in log units), the mean variance of prediction (*MVP*, in log units), the mean standard error of prediction (*SEP*, in percent), the mean standard error of model (*SEM*, in percent), and the pseudo R^2 (in percent). The *SEP* is the sum of the model error and the sampling error. The *MVP* represents the mean accuracy of prediction for all the gaging stations used in the regression analysis. The *MVP* and *SEP* are measures that indicate how well the equation will predict Q_{AEP} for ungaged sites. The pseudo R^2 and *SEM* are metrics that indicate how well the equation predicted Q_{AEP} for the gaging stations used in the analysis.

Although the *SEP* provides an indication of the mean reliability of a regression equation within a region, the *SEP* should be calculated for individual estimates if reliability of a particular estimate is required. The following equation can be used to calculate the standard error of prediction for a particular estimate (SEP_0):

$$SEP_0 = \sqrt{\sigma_\delta^2 + x_0 \left(X^T \Lambda^{-1} X\right)^{-1} x_0^T} \quad (3)$$

where

SEP_0	is the standard error of prediction, in log units, for an estimate of Q_{AEP} at an ungaged site;
σ_δ^2	is the model error variance, in log units (table 1–4), for the appropriate regression equation for the hydrologic region of the ungaged site;
x_0	is a row vector consisting of the value 1.0 in the first column followed by the log transformed values of the p explanatory variables (basin characteristics) for the ungaged site used in the regression equation;
x_0^T	is the transpose of the vector x_0; and
$(X^T \Lambda^{-1} X)^{-1}$	is the covariance matrix for the GLS or WLS regional regression equation (table 1–5 in appendix 1 at the back of this report chapter [available at https://doi.org/10.3133/sir20155019F]).

Once the SEP_0 has been calculated for a particular estimate, the SEP_0 can be used to calculate a confidence interval for the same estimate using the following equation:

$$CI_{0,\alpha} = \pm t_{\left(\frac{\alpha}{2},\, n-(p+1)\right)} \left(SEP_0\right) \quad (4)$$

where

$CI_{0,\alpha}$	is the confidence interval, in log units, for an estimate at site 0 with a significance level of α;
$t_{\left(\frac{\alpha}{2},\, n-(p+1)\right)}$	is the Student's t value for a confidence level of $100(1-\alpha)$ percent and $(n-[p+1])$ degrees of freedom; and
n and p	are the number of gaging stations used in the regression equation and the number of explanatory variables (basin characteristics) used in the regression equation, respectively.

If the peak-flow frequency at site 0 ($Q_{AEP,\,0}$) is converted to a logarithm (log $Q_{AEP,\,0}$) the confidence interval can be expressed in units of discharge using the following equation:

$$10^{\left(\log Q_{AEP,0} - CI_{0,\alpha}\right)} \leq true\ Q_{AEP,0} \leq 10^{\left(\log Q_{AEP,0} + CI_{0,\alpha}\right)} \quad (5)$$

where

$CI_{0,\alpha}$	is the confidence interval, in log units, for an estimate at site 0 with a confidence level of α;
Q_{AEP}	is the peak flow, in cubic feet per second, with an annual exceedance probability (AEP) in percent; and
$true\ Q_{AEP,0}$	is the true AEP peak flow at site 0.

For example, assume the Q_{AEP} for site 0 is estimated to be 400 cubic feet per second (ft³/s) and has an SEP_0 of 0.225 log units. Also assume that site 0 is located in a region that used 75 gaging stations and 2 basin characteristics (explanatory variables) in the regression analysis. The 90-percent confidence interval for this estimate would be computed as follows. First, the one-tailed Student's t value would be determined for $t_{\left(\frac{\alpha}{2},\, n-(p+1)\right)}$ or $t_{0.05,72}$, with the $\frac{\alpha}{2}$ term of 0.05 calculated by $\frac{(1-0.90)}{2}$ and the n-(p+1) degrees of freedom term of 72 calculated by 75-(2+1). The one-tailed Student's t value for $t_{0.05,72}$ is 1.67 (StatSoft, Inc., 2013). Then, using equation 4, the calculation would be

$$CI_{0,\alpha} = \pm t_{\left(\frac{\alpha}{2},\, n-(p+1)\right)}\left(SEP_0\right),$$

$$CI_{0,\alpha} = \pm 1.67(0.225),$$

$$CI_{0,\alpha} = \pm 0.375.$$

Then, using equation 5, the calculation would be

$$10^{\left(logQ_{AEP,0} - CI_{0,\alpha}\right)} \leq true\ Q_{AEP,0} \leq 10^{\left(logQ_{AEP,0} + CI_{0,\alpha}\right)}$$

$$10^{\left(log\,400 - 0.375\right)} \leq true\ Q_{AEP,0} \leq 10^{\left(log\,400 + 0.375\right)}$$

$$10^{(2.227)} \leq true\ Q_{AEP,0} \leq 10^{(2.977)}$$

$$169\ ft^3/s \leq true\ Q_{AEP,0} \leq 948\ ft^3/s.$$

The confidence interval about the estimate does not mean that there is a 90-percent probability that the *true* $Q_{AEP,0}$ is greater than 169 ft³/s and less than 948 ft³/s. Rather, it should be interpreted such that all values within the confidence interval are not significantly different from the *true* $Q_{AEP,0}$ at the 10-percent level.

Limitations of Regional Regression Equations

The regression equations (table 1–4 in appendix 1 at the back of this report chapter) might not be reliable for an ungaged site if the values of any explanatory variables (basin characteristics) for that site are outside the range of values used to develop the equations (table 3). Also, the regression equations might not be reliable if the values of the basin characteristics at a particular ungaged site do not fall within the joint probability distribution of all values of the explanatory variables used for that region. In other words, the regression equations might not be reliable if the values of the basin characteristics at a particular ungaged site are anomalously high or low compared to the values of all basin characteristics for gaging stations in that region. Solving the covariance matrix, $\left(X^T\Lambda^{-1}X\right)^{-1}$, for a given site can be used to determine if the joint distribution of the explanatory variables (basin characteristics) at that site is unreliably far from the center of the joint distribution of all of the values of the explanatory variables (basin characteristics) for that region. An example of solving the covariance matrix for a given site is presented in the section "Case 1—Ungaged Site with No Nearby Gaging Stations on the Same Stream." If the solution to the covariance matrix is greater than about $3p/n$ (where p is the number of explanatory variables used in the regional regression equation,

Table 3.   Ranges of values of basin characteristics used to develop regional regression equations.

[A, contributing drainage area, in square miles; E_{5000}, percentage of basin above 5,000 feet elevation[1]; E_{6000}, percentage of basin above 6,000 feet elevation[1]; ET_{SPR}, mean spring (March through June) evapotranspiration, in inches per month; F, percentage of basin with forest land cover; P, mean annual precipitation, in inches; SLP_{30}, percentage of basin with slopes greater than 30 percent; Min, minimum; Max, maximum; --, not used in regional regression equation]

Hydrologic region (ordered clockwise from northwestern Montana) (fig. 1)	A		E_{5000}		E_{6000}		ET_{SPR}		F		P		SLP_{30}	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
West	0.60	2,465.66	--	--	--	--	--	--	20.40	99.04	14.62	62.02	--	--
Northwest	2.43	1,556.17	--	--	--	--	--	--	--	--	--	--	--	--
Northwest Foothills	0.19	1,238.09	--	--	--	--	--	--	--	--	10.13	23.36	--	--
Northeast Plains	0.18	2,747.31	0.00	30.52	--	--	--	--	--	--	--	--	--	--
East-Central Plains	0.11	2,550.96	--	--	--	--	0.90	1.57	--	--	--	--	0.00	31.87
Southeast Plains	0.10	1,962.05	--	--	--	--	0.96	1.67	0.00	57.64	--	--	--	--
Upper Yellowstone-Central Mountain	0.39	2,039.76	--	--	0.00	100.00	--	--	--	--	--	--	--	--
Southwest	0.42	2,472.17	--	--	0.00	100.00	--	--	--	--	--	--	--	--

[1]Elevation refers to distance above North American Vertical Datum of 1988.

and n is the number of gaging stations used to develop the regional regression equation), the regression result might not be reliable.

The regression equations were developed on streams that are considered to be unaffected or minimally affected by regulation or issues related to urbanization. Thus, the equations might not be reliable for estimating Q_{AEP} for sites on regulated streams or sites that are affected by urbanization.

Although an effort was made to decrease the effect of regional bias on transregional streams, regression equations might not be reliable if an ungaged site of interest is located in a different region from the region where the stream originates. For streams that cross regional boundaries, the regression equation for each region can be applied separately, using basin characteristics at the site. The separate results then can be weighted in accordance with the proportion of drainage area in each region. For example, if 40 percent of the drainage area of an ungaged site is in the upstream region and 60 percent is in the downstream region, the estimate based on the equation for the upstream region can be multiplied by 0.4 and added to 0.6 times the estimate based on the equation for the downstream region. The *SEP* for such a weighted estimate also can be approximated by using the same weighting procedure based on drainage area. When the upstream part of an ungaged drainage basin is in the Northwest hydrologic region and the downstream part of the ungaged drainage basin is in the Northwest Foothills hydrologic region, weighting the separately calculated peak-flow frequencies in proportion to drainage area in each region is appropriate only for peak flows with AEPs greater than 4 percent (Parrett and Johnson, 2004). Historically, some large peak flows on some streams in the Northwest Foothills hydrologic region that originated in the Northwest hydrologic region attenuate from upstream to downstream. Estimating peak-flow frequencies at ungaged sites in the Northwest Foothills hydrologic region that originate in the Northwest hydrologic region requires careful consideration of the characteristics of the specific ungaged drainage basin and the hydrologic complexities of the two regions.

Comparison of Regional Regression Equations with Results of Previous Studies

Selected results from this study are compared with results of previous studies (Parrett and Johnson, 2004; Omang, 1992) in table 4 and figure 3. In discussion of the comparisons, emphasis is placed on comparison of the results from this study with the results of Parrett and Johnson (2004), with less emphasis placed on comparison of the results from this study with the results of Omang (1992). In general, the regression equations reported by Omang (1992; based on data through water year 1988) have lower *SEP*s than the regression equations reported for this study and Parrett and Johnson (2004). Since about the mid-1970s, variability in climatic conditions and peak-flow characteristics in Montana has generally increased (Pederson and others, 2010; Sando, McCarthy, and others, 2016). Thus, the extended periods of record used

in this study and Parrett and Johnson (2004) in relation to Omang (1992) might contribute to generally larger *SEP*s of the regression equations reported for this study and Parrett and Johnson (2004) in relation to regression equations reported in Omang (1992).

The equations presented for this study are considered to be an improvement on the previously reported equations (Parrett and Johnson, 2004) primarily because this study (1) included 13 more years of peak-flow data; (2) included 35 more gaging stations; (3) used a detailed GIS-based definition of the regulation status of gaging stations, which allowed better determination of the unregulated peak-flow records that are appropriate for use in the regional regression analysis; (4) included advancements in GIS and remote sensing technologies, which allowed more convenient calculation of basin characteristics and investigation of many more candidate basin characteristics; and (5) included advancements in computational and analytical methods, which allowed more thorough and consistent data analysis.

For most hydrologic regions, the explanatory variables (basin characteristics) in the regression equations of this study were similar to or the same as the explanatory variables (basin characteristics) used in Parrett and Johnson (2004; table 4); however, the GIS methods for computing the basin characteristics in this study (table 2; McCarthy and others, 2016) strongly differed from the methods of Parrett and Johnson (2004), which were based on manual analysis of paper topographic maps. Also, this study includes mean spring (March–June) evapotranspiration (ET_{SPR}, table 2) in the regression equations for the East-Central Plains and Southeast Plains hydrologic regions.

It is unlikely that evapotranspiration (ET) would have a direct substantial effect on streamflow during peak-flow conditions; however, it is likely that ET might serve as a surrogate for several hydrologic and land-surface characteristics that affect peak-flow potential in a drainage basin. The following discussion of possible indirect relations between ET and peak-flow potential is not intended to be a detailed analysis of the relevant issues but rather is intended to provide some possible explanations for the observed strong statistical relations between ET_{SPR} and peak-flow frequencies in eastern Montana. Evapotranspiration might be affected by many factors, including land-surface temperature, vegetation cover, available moisture, and surface-to-atmosphere convective and advective processes. The regression coefficients of ET_{SPR} for both regions are negative (table 1–4), which indicates an inverse relation between ET_{SPR} and peak-flow frequencies. That is, in drainage basins where ET_{SPR} is higher, Q_{AEP} tends to be lower; conversely, in drainage basins where ET_{SPR} is lower, Q_{AEP} tends to be higher. ET might be higher in drainage basins with higher vegetation covers. Increasing vegetation cover might disrupt the delivery of precipitation to the land surface, disrupt overland flow and attenuate runoff, and increase delivery of moisture from the land surface to the atmosphere. ET might be lower in drainage basins with lower soil temperatures. Lower soil temperatures might contribute to increased frequency

Table 4. Selected results of regional regression analyses of this study compared with previous studies (Parrett and Johnson, 2004; Omang, 1992).

[SEP, mean standard error of prediction; $Q_{66.7}$, peak-flow magnitude for annual exceedance probability of 66.7 percent; A, contributing drainage area; P, mean annual precipitation; F, percentage of basin with forest land cover; E_{5000}, percentage of drainage basin above 5,000 feet elevation[2]; E, mean basin elevation; SLP_{30}, percentage of basin with slopes greater than 30 percent; ET_{SPR} spring mean evapotranspiration; E_{6000}, percentage of drainage basin above 6,000 feet elevation[2]]

Hydrologic region (ordered clockwise from northwestern Montana) (fig. 1)	Number of gaging stations used in regional regression analysis			Explanatory variables			Mean SEP of regression equations[1]			SEP of regression equation for 1-percent annual exceedance probability		
	This report ($Q_{66.7}$ in parentheses)	Parrett and Johnson (2004)	Omang (1992)	This report	Parrett and Johnson (2004)	Omang (1992)	This report	Parrett and Johnson (2004)	Omang (1992)	This report	Parrett and Johnson (2004)	Omang (1992)
West	113 (113)	96	85	A, P, F	A, P, F	A, P	55.4	58.2	48.3	56.0	58.5	48.0
Northwest	32 (31)	35	31	A	A, P	A, P	36.7	42.8	37.0	13.6	40.2	38.0
Northwest Foothills	31 (28)	24	26	A, P	A	A, E	68.9	65.2	61.3	65.8	61.0	62.0
Northeast Plains	64 (59)	57	80	A, E_{5000}	A, E	A, E	61.9	94.0	62.0	54.5	101.4	56.0
East-Central Plains	90 (85)	85	85	A, SLP_{30}, ET_{SPR}	A, E	A, E	73.9	84.7	72.0	73.5	85.7	65.0
Southeast Plains	68 (67)	69	75	A, F, ET_{SPR}	A, F	A, E	90.9	101.1	78.0	71.1	91.6	62.0
Upper Yellowstone–Central Mountain	91 (87)	92	79	A, E_{6000}	A, E_{6000}	A, E_{6000}	78.3	66.6	53.4	69.0	56.8	50.0
Southwest	48 (47)	44	61	A, E_{6000}	A, E_{6000}	A, E_{6000}	77.5	81.8	69.9	73.8	80.3	66.0

[1]Determined by averaging the 50-, 20-, 10-, 4-, 2-, 1-, and 0.2-percent annual exceedance probabilities for each of the studies. The selected annual exceedance probabilities were reported in all of the studies.

[2]Elevation refers to distance above North American Vertical Datum of 1988.
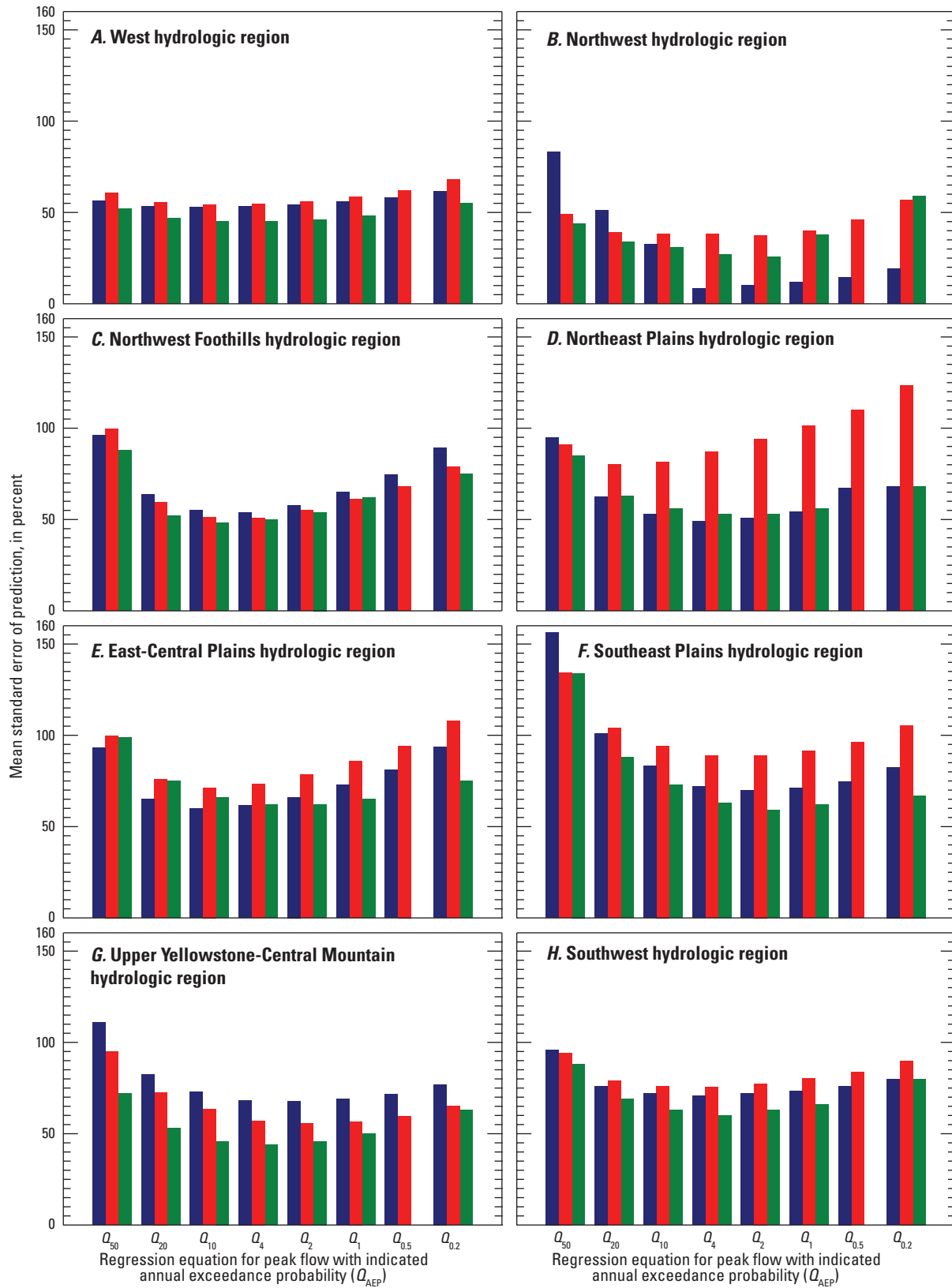
Figure 3. Comparison of mean standard error of prediction (*SEP*) from this study with SEPs from previous studies (Parrett and Johnson, 2004; Omang, 1992).

of frozen soil conditions during runoff events and thereby increase runoff potential. Thus, the observed strong statistical relations between ET_{SPR} and peak-flow frequencies in eastern Montana were considered to be reasonable on the basis of theoretical hydrologic relations.

The *SEP* provides information on the reliability of the regression equations (that is, generally how well the equations can predict Q_{AEP} at ungaged sites). Lower *SEP*s indicate lower errors and generally higher confidence in the results. Direct comparison of *SEP*s for this study with those of Parrett and Johnson (2004) is difficult because of fundamental differences in the datasets used in the regional regression analyses. The differences in the datasets include the period of record on which the peak-flow frequencies (Q_{AEP}, the dependent variable) were determined, the specific gaging stations included, and the specific values for basin characteristics that were determined using different methods; however, consideration of general patterns in *SEP*s from regression equations included in both this study and Parrett and Johnson (2004; that is, the 50-, 20-, 10-, 4-, 2-, 1-, 0.1-, and 0.2-percent AEPs) might provide useful information on the progression of estimating peak-flow frequencies at ungaged sites in Montana.

For most hydrologic regions, the regression equations reported for this study had lower *SEP*s than the regression equations reported by Parrett and Johnson (2004); however, for two hydrologic regions (Northwest Foothills and Upper Yellowstone-Central Mountain), the regression equations reported for this study had slightly to moderately higher *SEP*s than the regression equations reported by Parrett and Johnson (2004). In the remainder of this section of this report chapter, specific differences between this study and Parrett and Johnson (2004) are discussed by hydrologic region. Emphasis is placed on differences in the number of gaging stations, the explanatory variables, and the *SEP*s.

West Hydrologic Region

For the West hydrologic region (fig. 1), the results of this study generally were similar to the results of Parrett and Johnson (2004). This study included 17 more gaging stations than Parrett and Johnson (2004; table 4); the net change generally was because of the additional data collected after 1998. The regression equations reported for this study used the same explanatory variables as Parrett and Johnson (2004) but had slightly lower *SEP*s (a mean of about 3 percent lower and ranging from about 1 to 7 percent lower for an individual Q_{AEP}) than the regression equations reported by Parrett and Johnson (2004; fig. 3).

Northwest Hydrologic Region

For the Northwest hydrologic region (fig. 1), the results of this study were somewhat different from the results of Parrett and Johnson (2004). This study included three fewer gaging stations than Parrett and Johnson (2004; table 4); the net change generally was because of more discriminant screening of redundant gaging stations in this study, the use of

WLS regression analysis rather than OLS regression analysis, and the use of different explanatory variables. The regression equations reported for this study used only *A* as an explanatory variable; however, Parrett and Johnson (2004) used *A* and *P*. The *SEP*s for this study were higher for the 50- and 20-percent AEP equations (a mean of about 27 percent higher), but lower for all other equations (a mean of about 26 percent lower and ranging from about 4 percent lower to about 38 percent lower for an individual Q_{AEP}) than the regression equations reported by Parrett and Johnson (2004; fig. 3). Parrett and Johnson (2004) used OLS regression analysis for the Northwest hydrologic region because, for most of the gaging stations, a mixed-population analysis was used to determine peak-flow frequencies. The mixed-population analysis did not allow calculation of the distributional statistics needed for WLS regression. In contrast, the peak-flow frequencies for mixed-population peak-flow records in this study were determined using an alternative procedure (Sando, McCarthy, and Dutton, 2016) that allowed calculation of the distributional statistics needed for WLS regression.

Northwest Foothills Hydrologic Region

For the Northwest Foothills hydrologic region (fig. 1), the results of this study generally were similar to the results of Parrett and Johnson (2004). This study included seven more gaging stations than Parrett and Johnson (2004; table 4); the net change generally was because of discretionary considerations related to uncertainty of regulation or general nonconformity of flood frequency analyses at particular sites. The regression equations reported for this study for this region used *A* and *P* as explanatory variables; however, Parrett and Johnson used only *A* as an explanatory variable. The regression equations reported for this study had slightly higher *SEP*s (a mean of about 4 percent higher and ranging from about 7 percent lower to 10 percent higher for an individual Q_{AEP}) than the regression equations reported by Parrett and Johnson (2004; fig. 3). Inclusion of the additional gaging stations might have contributed to the slightly higher *SEP*s but is considered to provide accurate representation of the large variability in peak-flow characteristics in the hydrologically complex Northwest Foothills hydrologic region.

Northeast Plains Hydrologic Region

For the Northeast Plains hydrologic region (fig. 1), the results of this study were somewhat different from the results of Parrett and Johnson (2004). This study included seven more gaging stations than Parrett and Johnson (2004; table 4); the net change generally was because of the additional data collected after 1998 and because of discretionary considerations related to uncertainty of regulation or general nonconformity of flood frequency analyses at particular sites. The regression equations reported for this study used the explanatory variables of *A* and E_{5000} (table 4); however, the regression equations reported by Parrett and Johnson (2004) used the explanatory variables of *A* and mean basin elevation (*E*, table 4). In

this study, E_{5000} was able to describe more variability in Q_{AEP} than E. The regression equations reported for this study generally have substantially lower SEPs (a mean of about 32 percent lower and ranging from about 18 to 56 percent lower for an individual Q_{AEP}) than the regression equations reported by Parrett and Johnson (2004); however, the SEP for the Q_{50} regression equation reported for this study was about 5 percent higher than the SEP for the Q_{50} regression equation reported by Parrett and Johnson (2004; fig. 3). Definitive explanations for the large differences between the SEPs of this study and the SEPs of Parrett and Johnson (2004) are uncertain. Possible explanations might relate to effects of (1) including the additional data collected after 1998 and (2) use of a detailed GIS-based definition of the regulation status of gaging stations, which allowed better determination of the unregulated peak-flow records that are appropriate for use in the regional regression analysis. Notably, the regression equations reported for this study have SEPs generally similar to those reported by Omang (1992).

East-Central Plains Hydrologic Region

For the East-Central Plains hydrologic region (fig. 1), the results of this study were somewhat different from the results of Parrett and Johnson (2004). This study included five more gaging stations than Parrett and Johnson (2004; table 4); the net change generally was because of the additional data collected after 1998 and because of discretionary considerations related to uncertainty of regulation or general nonconformity of flood frequency analyses at particular sites. The regression equations reported for this study used the explanatory variables of A, SLP_{30}, and ET_{SPR}; however, the regression equations reported by Parrett and Johnson (2004) used the explanatory variables of A and E (table 4). The SLP_{30} variable is strongly based on theoretical hydrologic relations; larger land-surface slopes generally have greater runoff potential than smaller land-surface slopes. In this study, SLP_{30} was able to describe more variability in Q_{AEP} than E. Inclusion of ET_{SPR} was based on strong statistical relations between ET_{SPR} and Q_{AEP} and was considered to be reasonable on the basis of theoretical hydrologic relations. The regression equations reported for this study have moderately lower SEPs (a mean of about 11 percent lower and ranging from about 6 to 15 percent lower for an individual Q_{AEP}) than the regression equations reported by Parrett and Johnson (2004; fig. 3).

Southeast Plains Hydrologic Region

For the Southeast Plains hydrologic region (fig. 1), the results of this study were somewhat different from the results of Parrett and Johnson (2004). This study included one less gaging station than Parrett and Johnson (2004; table 4). The regression equations reported for this study used the same explanatory variables of A and F (table 4) as the regression equations reported by Parrett and Johnson (2004); however, the regression equations reported for this study also included the explanatory variable of ET_{SPR}. Inclusion of ET_{SPR} was

based on strong statistical relations between ET_{SPR} and Q_{AEP} and was considered to be reasonable on the basis of theoretical hydrologic relations. The regression equations reported for this study generally have moderately lower SEPs (a mean of about 10 percent lower and ranging from about 3 to 23 percent lower for an individual Q_{AEP}) than the regression equations reported by Parrett and Johnson (2004); however, the SEP for the Q_{50} regression equation reported for this study was about 22 percent higher than the Q_{50} regression equation reported by Parrett and Johnson (2004; fig. 3). The SEP for the Q_{50} regression equation reported for this study (156.3 percent; table 1–4) is large and indicates substantial uncertainty in predictions. Notably, the SEPs for the Q_{50} regression equations reported by Parrett and Johnson (2004) and Omang (1992) also were substantially greater than 100 percent. The pattern of large SEPs for the Q_{50} regression equations indicates that peak-flow frequencies for high AEPs (greater than 50 percent) probably are complex and poorly defined by regional regression analysis for the Southeast Plains hydrologic region.

Upper Yellowstone-Central Mountain Hydrologic Region

For the Upper Yellowstone-Central Mountain hydrologic region (fig. 1), the results of this study generally were similar to the results of Parrett and Johnson (2004). This study included one less gaging station than Parrett and Johnson (2004; table 4); the net change generally was because of more discriminant screening of redundant gaging stations. The regression equations reported for this study used the same explanatory variables but had moderately higher SEPs (a mean of about 12 percent higher and ranging from about 10 to 16 percent higher for an individual Q_{AEP}) than the regression equations reported by Parrett and Johnson (2004; fig. 3); notably, there was little difference with respect to the included gaging stations and the explanatory variables. The moderate increases in SEPs for the Upper Yellowstone-Central Mountain hydrologic region might have been affected by larger variability in peak-flow data collected after 1998 in relation to earlier data.

Southwest Hydrologic Region

For the Southwest hydrologic region (fig. 1), the results of this study generally were similar to the results of Parrett and Johnson (2004). This study included four more gaging stations than Parrett and Johnson (2004; table 4); the net change generally was because of the additional data collected after 1998 and because of discretionary considerations related to uncertainty of regulation or general nonconformity of flood frequency analyses at particular sites. The regression equations reported for this study used the same explanatory variables as Parrett and Johnson (2004) but had slightly lower SEPs (a mean of about 5 percent lower and ranging from about 2 percent higher to 10 percent lower for an individual Q_{AEP}) than the regression equations reported by Parrett and Johnson (2004; fig. 3).
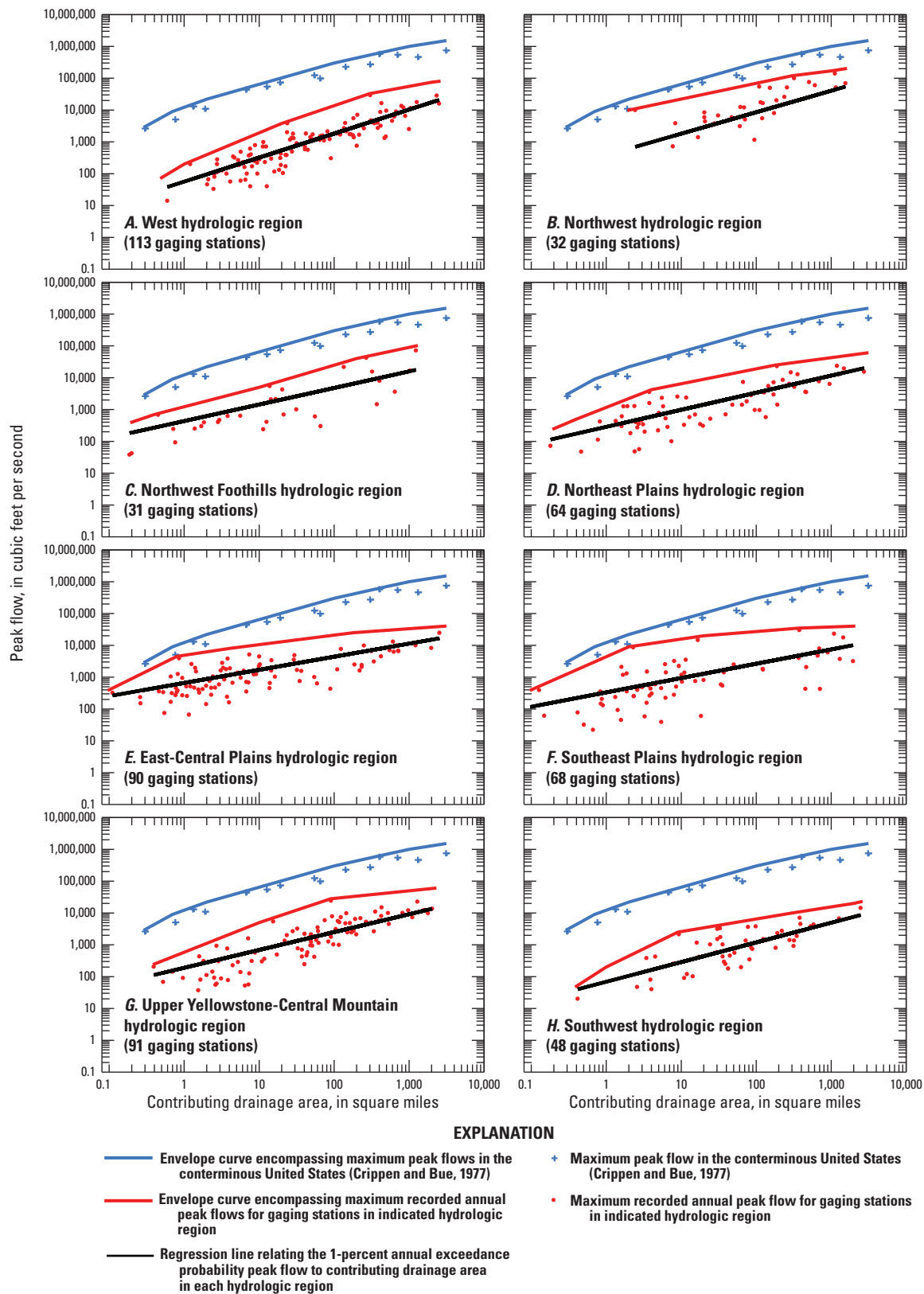
Figure 4. Maximum recorded annual peak flows, regional and national envelope curves, and ordinary least squares regression lines relating the 1-percent annual exceedance probability peak flows to contributing drainage area for hydrologic regions in Montana. *A*, West hydrologic region; *B*, Northwest hydrologic region; *C*, Northwest Foothills hydrologic region; *D*, Northeast Plains hydrologic region; *E*, East-Central Plains hydrologic region; *F*, Southeast Plains hydrologic region; *G*, Upper Yellowstone-Central Mountain hydrologic region; *H*, Southwest hydrologic region.

Envelope Curves Relating Largest Known Peak Flows to Contributing Drainage Area

Maximum recorded annual peak flows (table 1–2) for each gaging station used in the regression analysis are shown by hydrologic region in figure 4, plotted in relation to contributing drainage area. Additionally, envelope curves encompassing the maximum recorded annual peak flows for each region and selected maximum recorded annual peak flows for the conterminous United States (Crippen and Bue, 1977) are shown in figure 4. Furthermore, to compare the maximum recorded annual peak flows and the estimated Q_1 for each region, OLS regression lines relating Q_1 to contributing drainage area also are included in figure 4. Relations between the regional and national envelope curves indicate how the largest regional peak flows compare to the largest national peak flows in reference to drainage area. Relations between the regional envelope curves and the regional Q_1 OLS regression lines provide general information on the relative frequency of the maximum recorded annual peak flows in each region.

The West, Northeast Plains, Upper Yellowstone-Central Mountain, and Southwest hydrologic regions have generally similar relations between the maximum recorded annual peak flows, the envelope curves, and the Q_1 OLS regression lines. The separation between the regional and national envelope curves generally is large throughout the full ranges of drainage areas. The West, Upper Yellowstone-Central Mountain, and Southwest hydrologic regions predominantly are mountainous, and peak flows are primarily affected by snowmelt runoff, with smaller effects from intense precipitation events than the other hydrologic regions. The Northeast Plains hydrologic region is complex, and largely consists of low-elevation glaciated plains interspersed with occasional small mountain ranges; however, with respect to characteristics of large peak flows, the Northeast Plains hydrologic region is somewhat similar to the West, Upper Yellowstone-Central Mountain, and Southwest hydrologic regions.

For the Northwest and Northwest Foothills hydrologic regions, the upper parts of the regional envelope curves (at contributing drainage areas greater than about 100 mi²) generally are closer to the national envelope curves than is the case for the other hydrologic regions. The Northwest and Northwest Foothills hydrologic regions can be affected by large-scale regional spring rainfall events that typically are associated with snowmelt runoff.

For the East-Central Plains and the Southeast Plains hydrologic regions, the lower parts of the regional envelope curves (at contributing drainage areas less than about 20 mi²) generally are closer to the national envelope curves than is the case for most of the other hydrologic regions. The East-Central Plains and Southeast Plains hydrologic regions can be affected by intense local thunderstorms, which generally produce large flooding from small basins (Parrett and Johnson, 2004).

In addition to providing general information on characteristics of large peak flows, the regional envelope curves presented in figure 4 can be used to assess the reasonableness of Q_{AEP} estimates determined using the regression equations reported for this study or using other methods. For example, a Q_1 estimate that plots substantially above a regional envelope curve or substantially below the general trend of the data indicated by the OLS regional regression line might be unreasonable. In such cases, alternative methods for estimating Q_{AEP} might be considered.

Estimating Peak-Flow Frequencies at an Ungaged Site on a Gaged Stream

If an ungaged site is close to a gaging station on the same stream, peak-flow frequencies for the gaging station can be used to estimate peak-flow frequencies for the ungaged site using a drainage-area ratio adjustment. The AEP-percent peak flow at the ungaged site ($Q_{AEP,U}$) is calculated using the following equation:

$$Q_{AEP,U} = Q_{AEP,G} \left(\frac{DA_U}{DA_G} \right)^{exp_{AEP}} \qquad (6)$$

where

$\quad Q_{AEP,G}$ is the AEP-percent peak flow for gaging station G, in cubic feet per second;

$\quad DA_U$ is the drainage area at ungaged site U, in square miles;

$\quad DA_G$ is the drainage area at gaging station G, in square miles; and

$\quad exp_{AEP}$ is the regression coefficient for an OLS regression relating the log of the AEP-percent peak flow to the log of the drainage area within each region (table 5).

Equation 6 can be used to estimate peak-flow frequencies at ungaged sites on large streams, where the regression equations are not applicable because their large drainage areas fall outside of the range of values used to develop the equations (table 3). Equation 6 is considered unreliable if the value of DA_U/DA_G is less than 0.5 or greater than 1.5 (Parrett and Johnson, 2004). For ungaged sites where the values of DA_U/DA_G are outside the range of 0.5 to 1.5, the regression equations (table 1–4) might provide more reliable estimates of Q_{AEP} than equation 6.

If an ungaged site is between two gaging stations on the same stream, the logarithms of peak-flow frequencies at the ungaged site can be linearly interpolated between logarithms of peak-flow frequencies at the two gaging stations using the logarithms of drainage area as the basis for interpolation as follows:

$$\log Q_{AEP,U} = \log Q_{AEP,G1} + [(\log Q_{AEP,G2} - \log Q_{AEP,G1})/ \\ (\log DA_{G2} - \log DA_{G1})](\log DA_U - \log DA_{G1}) \qquad (7)$$

where

$\quad Q_{AEP,U}$ is the AEP-percent peak flow at ungaged site U, in cubic feet per second;

$\quad Q_{AEP,G1}$ is the AEP-percent peak flow for the upstream gaging station $G1$, in cubic feet per second;

$\quad Q_{AEP,G2}$ is the AEP-percent peak flow at the downstream gaging station $G2$, in cubic feet per second;

Table 5.   Regression coefficients for ordinary least squares regressions relating annual exceedance probability percent peak flow to contributing drainage area for use with ungaged sites on gaged streams.

AEP-percent peak flow (Q_{AEP})	Regression coefficient relating Q_{AEP} to drainage area for indicated region							
	West Region	Northwest Region	Northwest Foothills Region	Northeast Plains Region	East-Central Plains Region	Southeast Plains Region	Upper Yellowstone-Central Mountain Region	Southwest Region
$Q_{66.7}$	0.858	0.922	0.606	0.684	0.500	0.541	0.942	0.999
Q_{50}	0.843	0.904	0.575	0.690	0.488	0.527	0.896	0.939
$Q_{42.9}$	0.836	0.890	0.564	0.681	0.483	0.523	0.866	0.911
Q_{20}	0.813	0.832	0.534	0.639	0.463	0.502	0.761	0.818
Q_{10}	0.794	0.790	0.522	0.611	0.449	0.487	0.697	0.755
Q_4	0.777	0.741	0.516	0.579	0.434	0.471	0.634	0.690
Q_2	0.766	0.707	0.517	0.557	0.423	0.460	0.595	0.647
Q_1	0.755	0.675	0.521	0.537	0.414	0.450	0.561	0.609
$Q_{0.5}$	0.746	0.644	0.526	0.519	0.404	0.441	0.532	0.576
$Q_{0.2}$	0.735	0.605	0.536	0.496	0.393	0.430	0.498	0.533

DA_{G2} is the drainage area at the downstream gaging station $G2$, in square miles;

DA_{G1} is the drainage area at the upstream gaging station $G1$, in square miles; and

DA_U is the drainage area at ungaged site U, in square miles.

   Equation 7 also can be used to estimate peak-flow frequencies at ungaged sites on large streams, where the regression equations are not applicable because their large drainage areas fall outside of the range of values used to develop the equations (table 3). Application of equation 7 might provide unreliable results if the two gaging stations have notably different peak-flow characteristics caused by substantially different periods of record.

Estimating Peak-Flow Frequencies at Ungaged Sites Using StreamStats

   The StreamStats Web interface makes estimating peak-flow frequencies much easier and more consistent than manual calculation. Also, it allows the user to calculate basin characteristics and delineate a drainage basin for any user-specified point located along on a stream. In StreamStats, the user has the ability to calculate peak-flow frequencies for all four cases presented in the section "Examples of Estimating Peak-Flow Frequencies at Ungaged Sites." When computing peak-flow frequencies using regression equations in StreamStats, the regression equations used are the same as those presented in this report. Specific procedures for estimating peak-flow frequencies at ungaged sites with StreamStats are on the StreamStats Web site at http://water.usgs.gov/osw/streamstats/instructions1.html.

Examples of Estimating Peak-Flow Frequencies at Ungaged Sites

   Methods for estimating peak-flow frequencies at ungaged sites are presented for four cases: (1) an ungaged site with no nearby gaging stations on the same stream, (2) an ungaged site on an ungaged stream that crosses hydrologic region boundaries, (3) an ungaged site with a single nearby gaging station on the same stream, and (4) an ungaged site between nearby gaging stations on the same stream.

   Although the process and general approach to manually solving the regression equations are fairly straightforward, some of the datasets and the ability to easily delineate the site drainage basin might not be readily available to a given individual in need of peak-flow information. Thus, use of the Web-based StreamStats Program (http://water.usgs.gov/osw/streamstats/; U.S. Geological Survey, 2015b) is recommended to estimate Q_{AEP} for ungaged sites on ungaged streams; however, examples for estimating Q_{AEP} without using StreamStats are provided in the following sections.

Case 1—Ungaged Site with No Nearby Gaging Stations on the Same Stream

   For case 1, an estimate of the 1-percent AEP peak flow (Q_1) is needed for a stream at ungaged site 0 in the Southeast Plains hydrologic region and no nearby gaging stations are located on the same stream. The contributing drainage area (A) for site 0 was delineated by GIS analysis of the NED dataset (Gesch and others, 2002) and determined to be 27.70 mi². The mean spring

(March–June) evapotranspiration (ET_{SPR}) was calculated by GIS analysis of the Moderate Resolution Imaging Spectroradiometer (MODIS) global evapotranspiration product (MOD16) data (Mu and others, 2007) and determined to be 1.34 inches per month. The percentage of the drainage basin with forest land cover (F) was calculated by GIS analysis of the 2001 National Land Cover Dataset (NLCD; Homer and others, 2007) and determined to be 24.80 percent. Using the Q_1 regression equation for the Southeast Plains hydrologic region (table 1–4), the peak-flow frequency was estimated as follows:

$$Q_1 = 649A^{0.531} \ (F+1)^{-0.039} \ ET_{SPR}^{\ -4.22},$$

$$= 649 \ (27.70)^{0.531} \ (25.80)^{-0.039} \ (1.34)^{-4.22},$$

$$= 649 \ (5.83) \ (0.881) \ (0.291),$$

$$= 970 \ \text{ft}^3/\text{s}$$

To calculate the *SEP*, in log units, for the example Q_1 estimate of 970 ft³/s, equation 3 is used as follows:

$$SEP_0 = \sqrt{\sigma_\delta^2 + x_0 \left(X^T \Lambda^{-1} X \right)^{-1} x_0^T}$$

The σ_δ^2 for Q_1 in the Southeast Plains hydrologic region is 0.067 (table 1–4) and the covariance matrix, $\left(X^T \Lambda^{-1} X \right)^{-1}$, for Q_1 in the Southeast Plains hydrologic region is presented in table 1–5. The x_0 row vector consists of the value 1.0 and the logarithms of the explanatory variables (basin characteristics) for site *0* and can be written for this example as follows:

$$x_0 = \begin{vmatrix} 1.0 & log\,27.70 & log\,25.80 & log\,1.34 \end{vmatrix}, \text{ or } \begin{vmatrix} 1.0 & 1.442 & 1.412 & 0.127 \end{vmatrix}$$

The transpose of x_0 ($x_0^{\ T}$) is written as:

$$x_0^{\ T} = \begin{vmatrix} 1.0 \\ 1.442 \\ 1.412 \\ 0.127 \end{vmatrix}$$

Substituting the appropriate σ_δ^2 (0.067; table 1–4), the $x_0 T$ transpose values, and the covariance matrix, $\left(X^T \Lambda^{-1} X \right)^{-1}$, from table 1–5 into equation 3 and solving by matrix algebra leads to the following solution:

$$SEP_0 = \sqrt{0.067 + \begin{vmatrix} 1.0 & 1.442 & 1.412 & 0.127 \end{vmatrix} \cdot \begin{vmatrix} 0.01061 & -0.00045 & -0.00306 & -0.04999 \\ -0.00045 & 0.00173 & -0.00074 & -0.01283 \\ -0.00306 & -0.00074 & 0.00724 & -0.00778 \\ -0.04999 & -0.01283 & -0.00778 & 0.85177 \end{vmatrix} \cdot \begin{vmatrix} 1.0 \\ 1.442 \\ 1.412 \\ 0.127 \end{vmatrix}}$$

$$SEP_0 = \sqrt{0.067 + \begin{vmatrix} -0.00071 & -0.00063 & 0.00510 & 0.02870 \end{vmatrix} \cdot \begin{vmatrix} 1.0 \\ 1.442 \\ 1.412 \\ 0.127 \end{vmatrix}}$$

$$SEP_0 = \sqrt{0.067 + 0.00923}$$

$$SEP_0 = 0.276$$

Because the *SEP* is commonly expressed in percent rather than log units, SEP_0 can be converted to percent using the following equation (Tasker, 1978):

$$SEP_{0,\text{percent}} = \sqrt{100\left(e^{\left(2.3026\,SEP_{0,\log}\right)^2} - 1\right)} \qquad (8)$$

where

e is the natural logarithmic base (approximately 2.71828), and

$SEP_{0,\log}$ is the SEP at site 0 in base 10 log units.

Substituting for $SEP_{0,\log}$ and solving the equation gives the following solution:

$$SEP_{0,\text{percent}} = \sqrt{100\left(e^{\left(2.3026*0.276\right)^2} - 1\right)}$$

$$SEP_{0,\text{percent}} = \sqrt{100\left(e^{0.404} - 1\right)}$$

$$SEP_{0,\text{percent}} = 70.5$$

A 90-percent confidence interval can be constructed using equation 4 as follows, with a significance level (α) of 0.10, number of gaging stations used in the regression equation (n) of 68 (table 1–4), and the number of explanatory variables (basin characteristics) used in the regression equation (p) of 3 (table 1–4):

$$CI_{0,\alpha} = \pm t_{\left(\frac{\alpha}{2}, n-(p+1)\right)}\left(SEP_0\right)$$

$$CI_{0,\alpha} = \pm t_{(0.05,64)}\left(0.276\right)$$

$$CI_{0,\alpha} = \pm 1.67\ \left(0.276\right)$$

$$CI_{0,\alpha} = \pm 0.461$$

The confidence interval can be expressed in cubic feet per second using equation 5 as follows:

$$10^{\left(\log Q_{AEP,0} - CI_{0,\alpha}\right)} \leq true\ Q_{AEP,0} \leq 10^{\left(\log Q_{AEP,0} + CI_{0,\alpha}\right)}$$

$$10^{\left(\log 970 - 0.461\right)} \leq true\ Q_{AEP,0} \leq 10^{\left(\log 970 + 0.461\right)}$$

$$10^{\left(2.99 - 0.461\right)} \leq true\ Q_{AEP,0} \leq 10^{\left(2.99 + 0.461\right)}$$

$$10^{2.53} \leq true\ Q_{AEP,0} \leq 10^{3.45}$$

$$338\ \text{ft}^3/\text{s} \leq true\ Q_{AEP,0} \leq 2{,}824\ \text{ft}^3/\text{s}$$

Thus, the Q_1 estimate of 970 ft³/s is not significantly different at the 10-percent level from any value between 338 ft³/s and 2,824 ft³/s.

An estimate might not be reliable if the joint distribution of the explanatory variables (basin characteristics) is far removed from the center of the joint distribution of all the explanatory variables (basin characteristics) in that region. To determine if this is the case, the solution to $x_0^T\left(X^T\Lambda^{-1}X\right)^{-1}x_0$ is compared with $3p/n$. For this example, $x_0^T\left(X^T\Lambda^{-1}X\right)^{-1}x_0$ is solved as

$$x_0^T\left(X^T\Lambda^{-1}X\right)^{-1}x_0 = \begin{vmatrix} 1.0 & 1.442 & 1.412 & 0.127 \end{vmatrix} \cdot \begin{vmatrix} 0.01061 & -0.00045 & -0.00306 & -0.04999 \\ -0.00045 & 0.00173 & -0.00074 & -0.01283 \\ -0.00306 & -0.00074 & 0.00724 & -0.00778 \\ -0.04999 & -0.01283 & -0.00778 & 0.85177 \end{vmatrix} \cdot \begin{vmatrix} 1.0 \\ 1.442 \\ 1.412 \\ 0.127 \end{vmatrix}$$

$$= \begin{vmatrix} -0.00071 & -0.00063 & 0.00510 & 0.02870 \end{vmatrix} \cdot \begin{vmatrix} 1.0 \\ 1.442 \\ 1.412 \\ 0.127 \end{vmatrix}$$

$$= 0.00924$$

Because the result of 0.00924 is considerably smaller than $3p/n$ for the Southeast Plains hydrologic region (3·[3 variables]/[68 gaging stations], or 0.132), the statistical location defined by the combination of values of explanatory variables for this site is not far from the center of the joint distribution of all the values of explanatory variables in the Southeast Plains hydrologic region. Furthermore, the Q_1 estimate of 970 ft³/s for a contributing drainage area of 27.70 mi² plots below the regional envelope curve and is reasonably consistent with the general trend of the data indicated by the OLS regional regression line (fig. 4F). Thus, the Q_1 estimate of 970 ft³/s can be considered reliable.

Case 2—Ungaged Site on an Ungaged Stream that Crosses Hydrologic Region Boundaries

For streams that cross regional boundaries, the regression equation for each region can be applied separately, using basin characteristics at the site. The separate results then can be weighted in accordance with the proportion of drainage area in each region. The *SEP* for such a weighted estimate also can be approximated by using the same weighting procedure based on drainage area.

For case 2, an estimate of the 10-percent AEP peak flow (Q_{10}) is needed for a small stream (at an ungaged site in the East-Central Plains hydrologic region) that originates in the Upper Yellowstone-Central Mountain hydrologic region and then flows into the East-Central Plains hydrologic region. The contributing drainage area for the site was delineated by GIS analysis of the NED dataset (Gesch and others, 2002) and determined to be 17.27 mi². For the purpose of solving the Q_{10} regression equation for the East-Central Plains hydrologic region, the site was determined (by GIS analysis of the NED and MOD16 [Mu and others, 2007] datasets) to have a SLP_{30} of 0.05 percent and an ET_{SPR} of 1.15 inches per month. For the purpose of solving the Q_{10} regression equation for the Upper Yellowstone-Central Mountain hydrologic region, the site was determined (by GIS analysis of the NED dataset) to have an E_{6000} of 0.00.

The Q_{10} regression equation for the East-Central Plains hydrologic region (table 1–4) is solved as follows:

$$Q_{10} = 178\ A^{0.489}\ (SLP_{30} + 1)^{0.214}\ ET_{SPR}^{-3.90}$$

$$= 178\ (17.27)^{0.489}\ (1.05)^{0.214}\ (1.15)^{-3.90},$$

$$= 178\ (4.03)\ (1.01)\ (0.580),$$

$$= 420\ \text{ft}^3/\text{s}.$$

The Q_{10} regression equation for the Upper Yellowstone-Central Mountain hydrologic region (table 1–4) is solved as follows:

$$Q_{10} = 41.1\ A^{0.741}\ (E_{6000} + 1)^{-0.052},$$

$$= 41.1\ (17.27)^{0.741}\ (1.00)^{-0.052},$$

$$= 41.1\ (8.26)\ (1.00),$$

$$= 339\ \text{ft}^3/\text{s}.$$

The example stream originates in the Upper Yellowstone-Central Mountain hydrologic region, and the drainage area upstream from where the stream crosses the regional boundary is 6.21 mi². Thus, the proportion of the drainage basin of the site that is in the Upper Yellowstone-Central Mountain hydrologic region (6.21/17.27) is 0.36 and the proportion of the drainage basin of site 0 that is in the East-Central Plains hydrologic region is 0.64. The weighted estimate of Q_{10} is calculated as follows:

$$Q_{10} = 0.36\ (339) + 0.64\ (420),$$

$$= 391\ \text{ft}^3/\text{s}.$$

Case 3—Ungaged Site with a Single Nearby Gaging Station on the Same Stream

For case 3, an estimate of the 2-percent AEP peak flow (Q_2) is needed for an ungaged site on Burns Creek near Savage, Montana, that has a contributing drainage area of 121.32 mi². The site is located upstream from the gaging station Burns Creek near Savage, Mont. (gaging station 06329200, map number 575, fig. 1) in the East-Central Plains hydrologic region. The contributing drainage area and Q_2 for gaging station 06329200 are 234.12 mi² and 4,480 ft³/s, respectively (tables 1–3 and 1–2, respectively). The OLS regression coefficient relating contributing drainage area to Q_2 for the East-Central Plains hydrologic region is 0.423 (table 5). The estimated Q_2 for the ungaged site on Burns Creek is calculated using equation 6 as follows:

$$Q_{AEP,U} = Q_{AEP,G} \left(\frac{DA_U}{DA_G} \right)^{exp_{AEP}},$$

$$Q_{2,U} = 4,480 \left(\frac{121.32}{234.12} \right)^{0.423},$$

$$= 4,480\ (0.518)^{0.423},$$

$$= 3,392\ \text{ft}^3/\text{s}.$$

The drainage-area ratio term (121.32/234.12 = 0.518) is only slightly larger than the suggested limiting value of 0.5. For comparison, the Q_2 regression equation for the East-Central Plains hydrologic region also might be used. For this purpose, the example ungaged site was determined (by GIS analysis of the NED and MOD16 datasets [Gesch and others, 2002); Mu and others, 2007] to have an SLP_{30} of 2.30 and an ET_{SPR} of 1.14 inches per month. The Q_2 regression equation for the East-Central Plains hydrologic region (table 1–4) is solved as follows:

$$Q_2 = 433\ A^{0.454}\ (SLP_{30} + 1)^{0.279}\ ET_{SPR}^{-3.48},$$

$$= 433\ (121.32)^{0.454}\ (3.30)^{0.279}\ (1.14)^{-3.48},$$

$$= 433\ (8.83)\ (1.40)\ (0.634),$$

$$= 3{,}394\ \text{ft}^3/\text{s}.$$

The case 3 example illustrates the use of the drainage-area ratio method for a site with a drainage area ratio near the suggested limit of applicability. In this example, the use of the Q_2 regression equation for the East-Central Plains hydrologic region provides an acceptable alternative method. Both methods produce estimates of Q_2 that are similar and probably of similar reliability. Selection of the most appropriate estimate might involve consideration of the purpose of the estimate. For design purposes, the more conservative (larger) estimate might be selected, whereas for other planning or regulatory purposes, the smaller estimate might be selected.

Case 4—Ungaged Site Between Nearby Gaging Stations on the Same Stream

For case 4, an estimate of the 4-percent AEP peak flow (Q_4) is needed for ungaged site U on the Missouri River between Fort Benton and Landusky, Mont., that has a contributing drainage area of 32,327 mi^2. The ungaged site U is located between the gaging stations on the Missouri River at Fort Benton, Mont. (gaging station 06090800; map number 146; fig. 1; contributing drainage area of 24,297 mi^2; Sando, McCarthy, and Dutton, 2016) and the Missouri River near Landusky, Mont. (gaging station 06115200; map number 203; fig. 1; contributing drainage area of 39,825 mi^2; Sando, McCarthy, and Dutton, 2016). The Q_4 values for gaging stations 06090800 and 06115200 are 55,100 and 84,500 ft^3/s, respectively (Sando, McCarthy, and Dutton, 2016). The Q_4 at U is calculated from equation 7 as follows:

$$\log Q_{AEP,U} = \log Q_{AEP,G1} + [(\log Q_{AEP,G2} - \log Q_{AEP,G1})/ (\log DA_{G2} - \log DA_{G1})](\log DA_U - \log DA_{G1})$$

$$\log Q_{4,U} = \log 55{,}100 + [(\log 84{,}500 - \log 55{,}100)/(\log 39{,}825 - \log 24{,}297)]\ (\log 32{,}327 - \log 24{,}297)$$

$$\log Q_{4,U} = 4.74 + [(4.93 - 4.74)/(4.60 - 4.39)]\ (4.51 - 4.39)$$

$$= 4.74 + [(0.19)/(0.21)]\ (0.12)$$

$$= 4.74 + [0.905]\ (0.12)$$

$$= 4.74 + 0.109$$

$$= 4.85$$

Thus, $Q_{4,U} = 10^{4.85}$

$$Q_{4,U} = 70{,}795\ \text{ft}^3/\text{s}.$$

Summary

The U.S. Geological Survey, in cooperation with the Montana Department of Natural Resources and Conservation, completed a study to update methods for estimating peak-flow frequencies at ungaged sites in Montana based on peak-flow data through water year 2011 (water year is the 12-month period from October 1 through September 30 and is designated by the year in which it ends). The methods allow estimation of peak-flow frequencies (that is, peak-flow magnitudes, in cubic feet per second, associated with annual exceedance probabilities [AEPs] of 66.7, 50, 42.9, 20, 10, 4, 2, 1, 0.5, and 0.2 percent) at ungaged sites. The AEPs correspond to 1.5-, 2-, 2.33-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence intervals, respectively.

Regional regression analysis is a primary focus of Chapter F of this Scientific Investigations Report, and regression equations for estimating peak-flow frequencies at ungaged sites in eight hydrologic regions in Montana are presented. The regression equations are based on analysis of peak-flow frequencies and basin characteristics at 537 streamflow-gaging stations in or near Montana and were developed using generalized least squares regression or weighted least squares regression.

For this study, 28 basin characteristics were selected as candidate variables in the regression analyses. Of the 28 candidate basin characteristics, 7 were determined to have significant relations with peak-flow characteristics and were used in the final regression equations. The most consistently significant basin characteristic was contributing drainage area, which was used in all of the regression equations. Other basin characteristics determined to be significant and used in the final regression equations of one or more of the hydrologic regions include percentage of drainage basin above 5,000 feet elevation, percentage of drainage basin above 6,000 feet elevation, mean spring (March–June) evapotranspiration, percentage of drainage basin with forest land cover, mean (1971–2000) annual precipitation, and percentage of drainage basin with slopes greater than or equal to 30 percent.

All of the data used in calculating basin characteristics were derived from and are available through the U.S. Geological Survey Streamstats application (http://water.usgs.gov/osw/streamstats/) for Montana. StreamStats is a Web-Based geographic information system application that was created by the USGS to provide users with access to an assortment of analytical tools that are useful for water-resource planning and management. Additional information about StreamStats usage and limitations can be found at the StreamStats home page at http://water.usgs.gov/osw/streamstats/. The primary purpose of the Montana StreamStats application is to provide estimates of basin characteristics and streamflow characteristics for user-selected ungaged sites on Montana streams. The regional regression equations presented in this report can be conveniently solved using the Montana StreamStats application.

For each hydrologic region, the final generalized least squares regression or weighted least squares regression equations for estimating peak-flow frequencies at ungaged sites in Montana are presented. Also presented are measures of reliability of the equations, including the model error variance, mean variance of prediction, the mean standard error of prediction (*SEP*, in percent), the mean standard error of model, and the pseudo coefficient of determination (R^2).

Selected results from this study were compared with results of previous studies. For most hydrologic regions, the regression equations reported for this study had lower *SEP*s than the previously reported regression equations for Montana. The equations presented for this study are considered to be an improvement on the previously reported equations primarily because this study (1) included 13 more years of peak-flow data; (2) included 35 more streamflow-gaging stations; (3) used a detailed GIS-based definition of the regulation status of streamflow-gaging stations, which allowed better determination of the unregulated peak-flow records that are appropriate for use in the regional regression analysis; (4) included advancements in GIS and remote sensing technologies, which allowed more convenient calculation of basin characteristics and investigation of many more candidate basin characteristics; and (5) included advancements in computational and analytical methods, which allowed more thorough and consistent data analysis.

This report chapter also includes other methods for estimating peak-flow characteristics at ungaged sites. Two methods for estimating flood frequency at ungaged sites located on the same streams as streamflow-gaging stations are described. Additionally, envelope curves relating maximum recorded annual peak flows to drainage area for each of the eight hydrologic regions in Montana are presented and compared to a national envelope curve. In addition to providing general information on characteristics of large peak flows, the regional envelope curves can be used to assess the reasonableness of peak-flow frequency estimates determined using the regression equations.

References Cited

Akaike, H., 1973, Information theory and an extension of the maximum likelihood principle, *in* Petrov, B.N., and Csaki, F., eds., 2nd International Symposium on Information Theory: Budapest, Akademiai Kiado, p. 267–281.

Altman, N.S., 1992, An introduction to kernel and nearest-neighbor nonparametric regression: The American Statistician, v. 46, p. 175–185. [Also available at http://dx.doi.org/10.2307/2685209.]

Berwick, V.K., 1958, Floods in eastern Montana—Magnitude and frequency: U.S. Geological Survey Open-File Report 58–15, 23 p.

Boner, F.C., and Stermintz, Frank, 1967, Floods of June 1964 in northwestern Montana: U.S. Geological Survey Water-Supply Paper 1840–B, 242 p. [Also available at http://pubs.er.usgs.gov/publication/wsp1840B.]

Breiman, Leo, 2001, Random forests: Machine Learning, v. 45, no. 1, p. 5–32. [Also available at http://dx.doi.org/10.1023/A:1010933404324.]

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.

Eng, K., Chen, Y., and Kiang, J., 2009, User's guide to the Weighted-Multiple-Linear-Regression Program (WREG version 1.05): U.S. Geological Survey Techniques and Methods, book 4, chap. A8, 21 p. [Also available at http://pubs.usgs.gov/tm/tm4a8.]

Esri, Inc., 2014, ArcGIS for desktop, Release 10.2: Redlands, Calif., Esri, Inc., accessed June 2014 at http://www.esri.com/software/arcgis/arcgis-for-desktop.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., and Tyler, D., 2002, The National Elevation Dataset: Photogrammetric Engineering and Remote Sensing, v. 68, p. 5–11.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources: Techniques of Water-Resources Investigations, book 4, chap. A3, 510 p. [Also available at http://pubs.usgs.gov/twri/twri4a3/.]

Homer, Collin; Dewitz, Jon; Fry, Joyce; Coan, Michael; Hossain, Nazmul; Larson, Charles; Herold, Nate; McKerrow, Alexa; VanDriel, J.N.; and Wickham, James, 2007, Completion of the 2001 National Land Cover Database for the conterminous United States: Photogrammetric Engineering and Remote Sensing, v. 73, p. 337–341.

Horizon Systems Corporation, 2013, NHDPlusHome NHDPlus: accessed June 2014 at http://www.horizon-systems.com/nhdplus.

Jarque, C.M., and Bera, A.K., 1987, A test for normality of observations and regression residuals: International Statistical Review, v. 55, no. 2, p. 163–172. [Also available at http://dx.doi.org/10.2307/1403192.]

McCarthy, P.M., Dutton, D.M., Sando, S.K., and Sando, Roy, 2016, Montana StreamStats—A method for retrieving basin and streamflow characteristics in Montana: U.S. Geological Survey Scientific Investigations Report 2015–5019–A, 16 p. [Also available at http://dx.doi.org/10.3133/sir20155019A.]

Moran, P.A.P., 1950, Notes on continuous stochastic phenomena: Biometrika, v. 37, p. 17–23. [Also available at http://dx.doi.org/10.2307/2332142.]

Mu, Q., Heinsch, F.A., Zhao, M., and Running, S.W., 2007, Development of a global evapotranspiration algorithm based on MODIS and global meteorology data: Remote Sensing of Environment, v. 111, p. 519–536. [Also available at http://dx.doi.org/10.1016/j.rse.2007.04.015.]

Natural Resources Canada, 2009, Land Cover, circa 2000-vector: Sherbrooke, Canada, Natural Resources Canada, Centre for Topographic Information, Earth Sciences Sector, accessed December 16, 2014, at http://www.geobase.ca/geobase/en/data/landcover/csc2000v/description.html.

Natural Resources Canada, 2015, Regional, national and international climate modeling: accessed October 26, 2015, at http://cfs-scf.nrcan-rncan.gc.ca/projects/3?lang=en_CA.

Omang, R.J., 1992, Analysis of the magnitude and frequency of floods and the peak-flow gaging network in Montana: U.S. Geological Survey Water-Resources Investigations Report 92–4048, 70 p. [Also available at http://pubs.er.usgs.gov/publication/wri924048.]

Omang, R.J., Parrett, Charles, and Hull, J.A., 1986, Methods for estimating magnitude and frequency of floods in Montana based on data through 1983: U.S. Geological Survey Water-Resources Investigations Report 86–4027, 85 p.

Parrett, Charles, and Johnson, D.R., 2004, Methods for estimating flood frequency in Montana based on data through water year 1998: U.S. Geological Survey Water-Resources Investigations Report 03–4308, 101 p. [Also available at http://pubs.usgs.gov/wri/wri03-4308/.]

Parrett, Charles, and Omang, R.J., 1981, Revised techniques for estimating magnitude and frequency of floods in Montana: U.S. Geological Survey Open-File Report 81–917, 66 p. [Also available at http://pubs.usgs.gov/wri/wri03-4308/.]

Pederson, G.T., Gray, S.T., Ault, Toby, Marsh, Wendy, Fagre, D.B., Bunn, A.G., Woodhouse, C.A., and Graumlich, L.J., 2010, Climatic controls on the snowmelt hydrology of the northern Rocky Mountains: Journal of Climate, v. 24, p. 1666–1687. [Also available at http://dx.doi.org/10.1175/2010JCLI3729.1.]

PRISM Climate Group, 2004, PRISM climate data: Oregon State University, accessed December 16, 2014, at http://prism.oregonstate.edu.

Sando, S.K., McCarthy, P.M., and Dutton, D.M., 2016, Peak-flow frequency analyses and results based on data through water year 2011 for selected streamflow-gaging stations in or near Montana: U.S. Geological Survey Scientific Investigations Report 2015–5019–C, 27 p. [Also available at http://dx.doi.org/10.3133/sir20155019C.]

Sando, S.K., McCarthy, P.M., Sando, Roy, and Dutton, D.M., 2016, Temporal trends and stationarity in annual peak flow and peak-flow timing for selected long-term streamflow-gaging stations in or near Montana through water year 2011: U.S. Geological Survey Scientific Investigations Report 2015–5019–B, 48 p. [Also available at http://dx.doi.org/10.3133/sir20155019B.]

StatSoft, Inc., 2013, Electronic statistics textbook: Tulsa, Oklahoma, StatSoft, at http://www.statsoft.com/textbook/.

Tasker, G.D., 1978, Relation between standard errors in log units and standard errors in percent: U.S. Geological Survey Water Resources Division Bulletin, January–March, p. 86–87.

Tasker, G.D., 1980, Hydrologic regression with weighted least squares: Water Resources Research, v. 16, no. 6, p. 1107–1113.

Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: Journal of Hydrology, v. 111, p. 361–375. [Also available at http://dx.doi.org/10.1016/0022-1694(89)90268-0.]

U.S. Geological Survey, 2015a, National Water Information System (NWISWeb): U.S. Geological Survey database, accessed June 10, 2015, at http://waterdata.usgs.gov/nwis.

U.S. Geological Survey, 2015b, The StreamStats Program: accessed March 5, 2015, at http://water.usgs.gov/osw/streamstats/.

U.S. Interagency Advisory Council on Water Data, 1982, Guidelines for determining flood flow frequency: Hydrology Subcommittee, Bulletin 17B, 28 p., appendixes 1–14, 28 p.

Woods, A.J., Omernik, J.M., Nesser, J.A., Shelden, James, Comstock, J.A., and Azevedo, S.H., 2002, Ecoregions of Montana (2d ed.): U.S. Environmental Protection Agency, Western Ecology Division, accessed December 16, 2014, at http://archive.epa.gov/wed/ecoregions/web/html/mt_eco.html.

Appendix 1

Appendix 1. Supplemental Information Relating to the Regional Regression Analysis

This appendix contains supplemental information relating to the regional regression analysis. For the 537 streamflow-gaging stations used to develop the regional regression equations, selected information is presented in table 1–1, peak-flow frequency data and maximum recorded annual peak flows are presented in table 1–2, and basin characteristics are presented in table 1–3. Regression equations used for estimating peak-flow frequencies at ungaged sites in Montana are presented in table 1–4. Covariance matrices for generalized least squares and weighted least squares regressions are presented in table 1–5.

An Excel file containing the tables is available at https://doi.org/10.3133/sir20155019F.

Table 1–1.   Information for selected streamflow-gaging stations used in the regional regression analysis.

Table 1–2.   Peak-flow frequency data and maximum recorded annual peak flows for streamflow-gaging stations used in developing the regional regression equations.

Table 1–3.   Basin-characteristics data for streamflow-gaging stations used in developing the regional regression equations.

Table 1–4.   Final generalized least squares (GLS) and weighted least squares (WLS) regression equations for estimating peak-flow frequencies at ungaged sites in Montana.

Table 1–5.   Covariance matrices, $[X^T \Lambda^{-1} X]^{-1}$, for generalized least squares and weighted least squares regression equations.

USGS

Sando and others—**Methods for Estimating Peak-Flow Frequencies at Ungaged Sites in Montana Based on Data through Water Year 2011**—SIR 2015–5019–F, ver. 1.1