USGS
science for a changing world

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

Methods for Estimating Streamflow Characteristics at Ungaged Sites in Western Montana Based on Data through Water Year 2009

Chapter G of
Montana StreamStats

Scientific Investigations Report 2015–5019–G

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

Methods for Estimating Streamflow Characteristics at Ungaged Sites in Western Montana Based on Data through Water Year 2009

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

Chapter G of
Montana StreamStats

Scientific Investigations Report 2015–5019–G

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

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

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

U.S. Geological Survey, Reston, Virginia: 2016

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)
inch (in.)	25.4	millimeter (mm)
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)

International System of Units to U.S. customary units

Multiply	By	To obtain
Length		
meter (m)	3.281	foot (ft)

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

$$°F = (1.8 \times °C) + 32.$$

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

$$°C = (°F − 32) / 1.8.$$

Datums

Horizontal coordinate information is referenced to the 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 2009 is the period from October 1, 2008, through September 30, 2009.

Abbreviations

α	significance level
σ_{δ}^{2}	model error variance
A	contributing drainage area
a	regression coefficient
CI	confidence interval
CL	confidence level
exp	coefficient for drainage area adjustment
FLU	Final Land Unit classification
G	specific streamflow-gaging station of interest
g	gaged
GIS	geographic information system
GLS	generalized least squares
h	leverage
K	regression constant
L	lower
LCC2000	Northern Land Cover of Canada, circa 2000-vector
MODIS	Moderate Resolution Imaging Spectroradiometer
MOD16	Moderate Resolution Imaging Spectroradiometer global evapotranspiration product
MSE	mean square error
MT DEQ	Montana Department of Environmental Quality
MT DNRC	Montana Department of Natural Resources and Conservation
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 used in the regression equation
p-value	statistical probability level
PI	prediction interval
PRISM	Parameter-elevation Regression on Independent Slopes Model
Q	streamflow for a specified streamflow characteristic
Q_0	estimated streamflow for an ungaged site
Q_{AL7Q10}	Streamflow for which the mean of a consecutive 7-day annual low flow can be expected to be lower, on average, once every 10 years (also known as the nonexceedance probability of 10 percent) based on annual low-flow data
Q_{S214Q5}	Streamflow for which the mean of a consecutive 14-day seasonal (July–October) low flow can be expected to be lower, on average, once every 5 years (also known as the nonexceedance probability of 20 percent) based on July–October (season 2) low-flow data
R	hydrologic region
R^2	coefficient of determination
$RMSE$	root mean square error
SEM	mean standard error of the model for the regression equation
SEP	mean standard error of prediction for the regression equation
SEP_0	standard error of prediction of an ungaged site
SLP_{50}	percentage of drainage basin with slope greater than or equal to 50 percent
t	Student's t value
U	upper
u	ungaged
USGS	U.S. Geological Survey
VIF	variance inflation factor
WBD	Watershed Boundary Dataset
WREG	Weighted-Multiple-Linear Regression Program
x	value of the explanatory variable
x_0	row vector consisting of the value 1.0 in the first column followed by the log transformed values of the p explanatory variables for the ungaged site used in the regression equations
x_0^{T}	transpose of the vector x_0
$(X^{\mathrm{T}} \Lambda^{-1} X)^{-1}$	covariance matrix for the generalized least squares regional regression equation
$(X^{\mathrm{T}} X)^{-1}$	covariance matrix for the ordinary least squares regional regression equation

Definitions of Streamflow Terminology

annual mean streamflow	arithmetic mean of all daily mean streamflows for a single specified year
daily mean streamflow	mean streamflow for a single specified day
mean annual streamflow	arithmetic mean of all annual mean streamflows for the period of record or for a specific period of multiple years
mean monthly streamflow	arithmetic mean of all monthly mean streamflows for a specified month for the period of record or for a specific period of multiple years
monthly mean streamflow	arithmetic mean of all daily mean streamflows for a single specified month in a single specified year

Acknowledgments

Methods for Estimating Streamflow Characteristics at Ungaged Sites in Western Montana Based on Data through Water Year 2009

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

Abstract

The U.S. Geological Survey, in cooperation with the Montana Department of Environmental Quality and the Montana Department of Natural Resources and Conservation, developed regional regression equations based on basin and streamflow characteristics for streamflow-gaging stations through water year 2009 that can be used to estimate streamflow characteristics for ungaged sites in western Montana. The regression equations allow estimation of low-flow frequencies; mean annual and mean monthly streamflows; and the 20-, 50-, and 80-percent durations for annual and monthly duration streamflows for ungaged sites in western Montana that are unaffected by regulation.

Regression equations were developed for 54 streamflow characteristics in each of the 4 hydrologic regions generally located in western Montana. Exploratory data analyses determined that of the 40 basin characteristics used as explanatory variables, only 3 were significant in the final equations. These three basin characteristics are contributing drainage area, mean annual precipitation of the drainage basin, and percentage of basin with slopes greater than 50 percent. Low-flow frequency data and daily mean streamflow data from 152 streamflow-gaging stations were used to develop the regression equations. These 152 streamflow-gaging stations were considered to be unregulated (or only the unregulated period of record for regulated streamflow-gaging stations were used). For this study, a streamflow-gaging station is considered to be unregulated if cumulative drainage area upstream from all upstream dams is less than 20 percent of the drainage basin for the streamflow-gaging station, and if no large diversion canals are upstream from the streamflow-gaging station.

In addition to the regional regression equations, two methods for estimating streamflow characteristics based on drainage areas of a nearby streamflow-gaging station are provided. Also, spreadsheet tools are provided for calculating streamflow characteristics based on regression equations or the drainage-area ratio method.

All of the data used to calculate basin characteristics were derived from publicly available data sources and are available through the U.S. Geological Survey Streamstats program (http://water.usgs.gov/osw/streamstats/) for Montana. 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 have been loaded to the Montana StreamStats application and can be used to derive streamflow characteristics for ungaged sites.

Introduction

Information about streamflow characteristics is essential for development and management of surface-water resources. Water and land-use managers, planners, administrators, builders, engineers, recreationists, and the general public use information on all aspects of streamflow to assess water conditions and evaluate land-use alternatives.

Annual, seasonal, and monthly streamflow characteristics are particularly useful for characterizing streamflow variability. Streamflow characteristics, such as the annual 7-day, 10-year low flow (Q_{AL7Q10}; streamflow for which the mean of a consecutive 7-day annual low flow can be expected to be lower, on average, once every 10 years [also known as the nonexceedance probability of 10 percent] based on annual low-flow data), commonly are used when setting effluent limits and allowable pollutant loads to meet water-quality standards. Annual and seasonal low-flow characteristics used in conjunction with monthly duration streamflow characteristics also indicate the probable availability of water in streams during times when conflicts between water supply and demand are most prevalent.

The U.S. Geological Survey (USGS) has previously published reports that describe and document streamflow characteristics at streamflow-gaging stations (hereinafter referred to as "gaging stations") in Montana (Shields and White, 1981; Waltemeyer and Shields, 1982; Omang, 1984; McCarthy, 2005; McCarthy, 2016). Streamflow characteristics also are needed at ungaged locations in Montana. Thus, the USGS, in cooperation with the Montana Department of Environmental

Quality (MT DEQ) and the Montana Department of Natural Resources and Conservation (MT DNRC), initiated a study to develop regional regression equations in western Montana based on basin and streamflow characteristics for gaging stations through water year 2009 (water year is the 12-month period from October 1 through September 30 and is designated by the year in which it ends). The regression equations can be utilized to estimate selected annual and seasonal low-flow frequencies and selected monthly and annual flow durations for ungaged sites.

Purpose and Scope

The study described in Chapter G of this Scientific Investigations Report is part of a larger study to develop a Stream-Stats 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). Chapter G describes methods for estimating selected streamflow characteristics in western Montana for ungaged sites based on basin and streamflow characteristics for gaging stations through water year 2009. Regional regression analyses are the primary focus of this report chapter. The development of regression equations and drainage-area ratio methods for estimating streamflow characteristics in western Montana is documented in this report chapter. The regression equations are based on selected low-flow frequencies, on selected annual and monthly streamflow characteristics from McCarthy (2016), and on basin characteristics at 152 gaging stations (fig. 1; table 1–1 in appendix 1 at the back of this report chapter [available at http://dx.doi.org/10.3133/sir20155019G]; 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) for the low-flow frequencies and ordinary least squares (OLS) regression (Eng and others, 2009) for the monthly and annual streamflow characteristics. Low-flow frequencies for which GLS regression was used to develop regional regression equations include Q_{AL7Q10} and the July–October, 14-day, 5-year low flow (Q_{S214Q5}; streamflow for which the mean of a consecutive 14-day seasonal low flow can be expected to be lower, on average, once every 5 years [also known as the non-exceedance probability of 20 percent] based on July–October [season 2] low-flow data). Selected streamflow characteristics estimated using OLS regression include the mean annual streamflow; mean monthly streamflow; and the 20-, 50-, and 80-percent durations for annual and monthly duration streamflows.

All of the data used in calculating basin and streamflow characteristics are available through the USGS StreamStats program (http://water.usgs.gov/osw/streamstats/; U.S. Geological Survey, 2015a). StreamStats is a Web-based geographic information system (GIS) application designed for national application, with each State, territory, or group of States responsible for creating unique geospatial datasets and regression equations to compute streamflow characteristics. Stream-Stats was created by the USGS to make GIS-based estimates of streamflow statistics easier, faster, and more consistent than previously used manual techniques. Also, StreamStats allows the user to calculate basin characteristics and delineate a drainage basin for any user-specified point located on a stream. The GIS-based calculation of basin characteristics allows consideration of many more basin characteristics potentially affecting streamflow characteristics than had previously been possible. The regression equations used to estimate streamflow characteristics in StreamStats are the same as those documented in this report. Additional information about StreamStats usage and limitations are available at http://water.usgs.gov/osw/streamstats/ (U.S. Geological Survey, 2015a), specific procedures for estimating streamflow characteristics at ungaged sites by using StreamStats are available at http://water.usgs.gov/osw/streamstats/instructions1.html, and the Montana StreamStats application is described by McCarthy and others (2016).

Previous Studies

Regression equations were developed in previous studies to estimate streamflow characteristics in various regions of Montana; however, only regression equations to estimate peak-flow frequencies have been developed for a statewide application (Sando, S.K., and others, 2016; Parrett and Johnson, 2004). No previous studies in Montana have developed regression equations to estimate low-flow frequencies. The following studies were completed in selected areas of Montana and include methods for estimating streamflow characteristics that are comparable to those developed for this study.

Sando and others (2009) developed methods for estimating streamflow characteristics at ungaged sites for streams in, and near, the Charles M. Russell National Wildlife Refuge in northeastern Montana (not shown in fig. 1). Methods were derived to relate short-term streamflow characteristics at gaging stations in the Charles M. Russell National Wildlife Refuge to long-term streamflow characteristics in proximity to the Charles M. Russell National Wildlife Refuge. The estimated streamflow characteristics included mean annual and mean monthly streamflows, selected annual and monthly duration values, and selected peak-flow frequency values. The mean annual and annual duration values were estimated using basin characteristics (drainage area and relief), and monthly values were estimated using coefficients and the long-term mean annual streamflow estimate (Sando and others, 2009). The mean standard error of the estimates for annual streamflow characteristics ranged from 49 to 65.1 percent, and the mean standard error of estimate for monthly streamflow characteristics ranged from 29.6 to 122.6 percent.
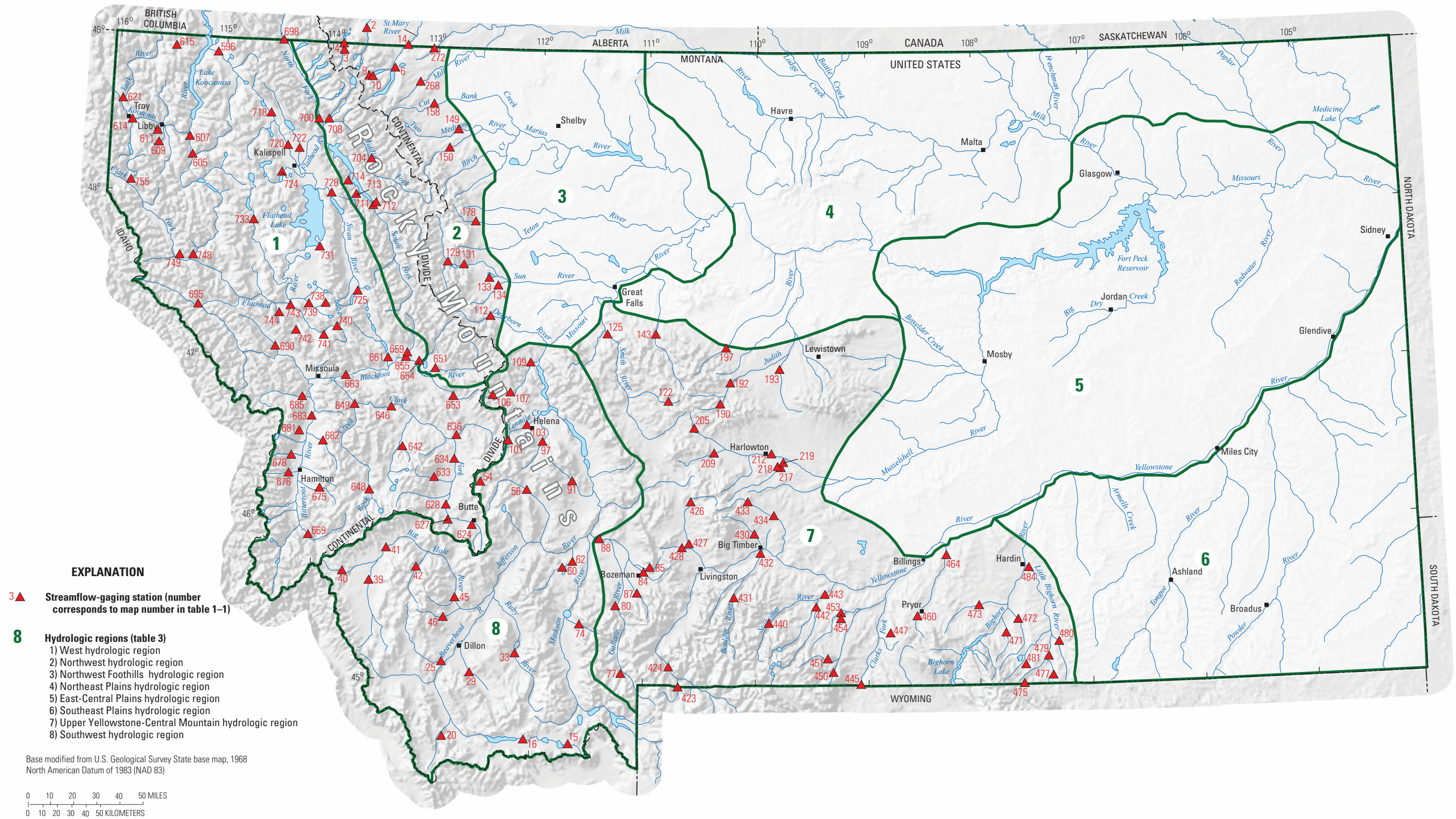
Figure 1.   Study area, hydrologic regional boundaries, and selected streamflow-gaging stations used in this study.

Parrett and Cartier (1990) developed three methods for estimating mean monthly streamflow and selected monthly duration streamflows for parts of western Montana. The first method used basin characteristics (drainage area, relief, basin slope, mean basin elevation, and mean annual precipitation) and regression equations to estimate streamflow characteristics; the standard errors of these methods ranged from 43 to 107 percent. The second method used channel-width characteristics to estimate streamflow characteristics; the standard errors ranged from 41 to 111 percent. The third method used 12 once-monthly streamflow measurements at an ungaged site correlated with measurements at a nearby gaging station; the standard errors ranged from 19 to 92 percent.

Parrett and others (1983) developed methods to estimate mean annual streamflow at sites in western and northeastern Montana using channel-width characteristics. The standard errors ranged from 61 to 77 percent in northeastern Montana and from 38 to 42 percent in western Montana. Utilizing additional basin characteristics did not significantly improve the regression equations for estimating mean annual streamflow in Parrett and others (1983).

Parrett and others (1982) developed methods for estimating mean annual streamflow in southeastern Montana based on channel geometry characteristics. Separate equations were developed for ephemeral streams and intermittent streams in southeastern Montana. The standard error of these equations ranged from 38 to 79 percent.

Description of Study Area

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 generally flat plains, interspersed with areas of deeply incised streams and rugged relief generally termed badlands or breaks. Most of western Montana is in the Canadian, Northern, and Middle Rockies ecoregions, whereas most of eastern Montana is in the Northwestern Glaciated Plains and Northwestern Great Plains ecoregions (Woods and others, 2002). Elevations range from about 1,800 feet (ft) above the North American Vertical Datum of 1988 (NAVD 88) in the eastern plains areas and in the Kootenai River Basin in extreme northwest Montana to about 12,800 ft above NAVD 88 in some mountain ranges. The elevation information was derived from a GIS analysis of the National Elevation Dataset (NED; Gesch and others, 2002). Mean annual precipitation also is highly variable throughout Montana and ranges from about 10 inches (in.), generally in low-altitude plains areas to 110 in. in some mountainous areas of western Montana (PRISM Climate Group, 2004).

Methods for Estimating Streamflow Characteristics at Ungaged Sites in Montana

Streamflow characteristics are frequently needed for ungaged sites for which streamflow data are not reported. Regression equations are the most commonly used method for estimating streamflow characteristics at ungaged sites that are considered unregulated and are the primary focus of this study; however, when the ungaged site is located near a gaging station on the same stream, drainage-area ratio methods may produce more reliable estimates. Two drainage-area ratio methods are provided and described in this report.

Regional Regression Method

Regional regression analyses involve determining relations between streamflow characteristics and basin characteristics at gaging stations to estimate streamflow characteristics at ungaged sites. As described in the following sections, developing regional regression equations requires the identification of suitable gaging stations, and determination of selected streamflow and basin characteristics for those gaging stations. Additional analyses of these gaging stations determined that regression equations could not reliably be developed for the selected streamflow characteristics for parts of central and eastern Montana; therefore, the development of regression equations was restricted to western Montana.

Selection of Gaging Stations

The USGS has been collecting and publishing streamflow data at gaging stations in Montana for more than 100 years (U.S. Geological Survey, 2015b; table 1–1). McCarthy (2016) reported streamflow characteristics for 408 gaging stations in and near Montana that had at least 10 years of record through water year 2009. Methods to determine the regulation status of gaging stations are described in McCarthy and others (2016). Methods to compile and manipulate the data, and to analyze the streamflow characteristics for gaging stations are described in McCarthy (2016).

The 408 gaging stations (McCarthy, 2016) were screened for suitability for inclusion in the regional regression analyses based on the following criteria: (1) contributing drainage area less than about 2,750 mi^2; (2) streamflow records unaffected by regulation; (3) small redundancy with nearby gaging stations; and (4) representation of streamflow characteristics at sites within Montana.

The criterion of drainage area less than about 2,750 mi^2 serves to restrict the regional regression analyses to smaller streams that might not be represented by gaging stations. Typically, gaging stations with contributing drainage areas greater than 2,750 mi^2 were outliers in the regression equation

development, and streams with contributing drainage areas greater than 2,750 mi^2 commonly have one or more gaging stations; thus, the gaging stations can be used to estimate streamflow characteristics at ungaged sites on streams with contributing drainage areas greater than 2,750 mi^2 using the drainage-area ratio methods.

Reservoir storage and operations have the potential to substantially affect streamflow characteristics; therefore, gaging stations affected by regulation are unsuitable for the regional regression analyses. The USGS maintains a geospatial database of dams in Montana that was used to define the regulation status for Montana gaging stations as described in McCarthy and others (2016). In McCarthy and others (2016) and in this study, a gaging station is considered to be unregulated if the cumulative drainage area upstream from all 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 contributing drainage area of all upstream dams equals or exceeds 20 percent of the drainage area of the gaging station (McCarthy and others, 2016; McCarthy, 2005; Parrett and Johnson, 2004), or if a large diversion canal is upstream from the gaging station. For those regulated gaging stations, if 10 or more years of streamflow data were available for a period prior to regulation, the streamflow data for the unregulated period were included in the regional regression analyses. Data from regulated gaging stations for regulated periods were not included in the regression analyses.

A redundant gaging-station analysis was used to account for spatial-autocorrelation in streamflow characteristics 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. For streams with two or more gaging stations on a stream, the contributing drainage area of each gaging station was examined and determined to be redundant when the area of one gaging station was within 0.5–2.0 times the area of the other gaging station. The redundant gaging stations were reviewed and those with the shortest period of record were 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 gaging stations included in McCarthy (2016) are largely, or entirely, outside of Montana. These gaging stations were excluded from the regional regression analyses if their drainage basins were considered to provide poor representation of streamflow characteristics in Montana, or if there were potential effects from undocumented regulation in the basin.

Of the 408 gaging stations with streamflow characteristics reported by McCarthy (2016), 213 gaging stations were included in the initial regional regression analyses. As a result of further refinement of the selected gaging stations (described in the section "Exploratory Data Analyses"), only 152 gaging stations were used for the final regional regression equations in this report chapter. Information and streamflow characteristics for the 152 gaging stations are presented in table 1–1.

Selection of Streamflow Characteristics

McCarthy (2016) reported 424 streamflow characteristics for each gaging station in or near Montana; however, only 54 streamflow characteristics that are most commonly used by MT DEQ and MT DNRC were selected for development of regression equations and inclusion in this study. The selected streamflow characteristics are two low-flow frequencies (Q_{AL7Q10}, Q_{S214Q5}); mean annual and mean monthly streamflow; and the 20-, 50-, and 80-percent durations for annual and monthly duration streamflows. The 54 streamflow characteristics used in this study are listed and described in table 1.

Basin Characteristics

Basin characteristics investigated as potential explanatory variables in the regional regression analyses were selected based on previous studies performed in Montana (Sando and others, 2009; Parrett and Johnson, 2004), theoretical relations with streamflow characteristics, and the ability to generate the characteristics using GIS analyses and digital datasets. In previous regional regression studies from Montana, basin characteristics were manually measured or estimated using topographic maps, planimeters, and overlaying transparent gridded cells on the maps. The number of candidate basin characteristics used in the regional regression analyses in these studies ranged from 2 (Berwick, 1958) to 12 (Parrett and Omang, 1981). For this study, 40 basin characteristics were selected as candidate explanatory variables (table 2). Contributing drainage basins were delineated for each gaging station using a combination of 30-meter digital elevation data from the NED (Gesch and other, 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 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 were then summarized for each gaging station and its associated basin.

Exploratory Data Analyses

Exploratory data analyses were used to define hydrologic region boundaries for Montana and to determine the best explanatory variables for final regression equation development in each of the eight hydrologic regions (fig. 1).

Table 1.   Definition of streamflow characteristics used in this study.

[--, StreamStats label has not been defined for this streamflow characteristic]

Streamflow characteristic	StreamStats abbreviation for streamflow characteristic	Description of streamflow characteristic
Q_{AL7Q10}	M7D10P	Streamflow for which the mean of a consecutive 7-day annual low-flow can be expected to be lower, on average, once every 10 years (also known as the non-exceedance probability of 10 percent) based on annual low-flow data
Q_{S214Q5}	--	Streamflow for which the mean of a consecutive 14-day seasonal low-flow can be expected to be lower, on average, once every 5 years (also known as the non-exceedance probability of 20 percent) based on July–October (season 2) low-flow data
$Q_{A0.2}$	--	Annual streamflow that is exceeded 20 percent of the time
$Q_{A0.5}$	MEDAN	Annual streamflow that is exceeded 50 percent of the time, also known as the median annual streamflow
$Q_{A0.8}$	--	Annual streamflow that is exceeded 80 percent of the time
Q_{AMEAN}	QA	Mean annual streamflow
$Q_{JAN0.2}$	JAND20	January streamflow that is exceeded 20 percent of the time
$Q_{JAN0.5}$	JAND50	January streamflow that is exceeded 50 percent of the time, also known as the median January streamflow
$Q_{JAN0.8}$	JAND80	January streamflow that is exceeded 80 percent of the time.
$Q_{JANMEAN}$	Q1	Mean January streamflow
$Q_{FEB0.2}$	FEBD20	February streamflow that is exceeded 20 percent of the time
$Q_{FEB0.5}$	FEBD50	February streamflow that is exceeded 50 percent of the time, also known as the median February streamflow
$Q_{FEB0.8}$	FEBD80	February streamflow that is exceeded 80 percent of the time
$Q_{FEBMEAN}$	Q2	Mean February streamflow
$Q_{MAR0.2}$	MARD20	March streamflow that is exceeded 20 percent of the time
$Q_{MAR0.5}$	MARD50	March streamflow that is exceeded 50 percent of the time, also known as the median March streamflow
$Q_{MAR0.8}$	MARD80	March streamflow that is exceeded 80 percent of the time
$Q_{MARMEAN}$	Q3	Mean March streamflow
$Q_{APR0.2}$	APRD20	April streamflow that is exceeded 20 percent of the time
$Q_{APR0.5}$	APRD50	April streamflow that is exceeded 50 percent of the time, also known as the median April streamflow
$Q_{APR0.8}$	APRD80	April streamflow that is exceeded 80 percent of the time
$Q_{APRMEAN}$	Q4	Mean April streamflow
$Q_{MAY0.2}$	MAYD20	May streamflow that is exceeded 20 percent of the time
$Q_{MAY0.5}$	MAYD50	May streamflow that is exceeded 50 percent of the time, also known as the median May streamflow
$Q_{MAY0.8}$	MAYD80	May streamflow that is exceeded 80 percent of the time
$Q_{MAYMEAN}$	Q5	Mean May streamflow
$Q_{JUN0.2}$	JUND20	June streamflow that is exceeded 20 percent of the time
$Q_{JUN0.5}$	JUND50	June streamflow that is exceeded 50 percent of the time, also known as the median June streamflow
$Q_{JUN0.8}$	JUND80	June streamflow that is exceeded 80 percent of the time
$Q_{JUNMEAN}$	Q6	Mean June streamflow.
$Q_{JUL0.2}$	JULD20	July streamflow that is exceeded 20 percent of the time.
$Q_{JUL0.5}$	JULD50	July streamflow that is exceeded 50 percent of the time, also known as the median July streamflow.
$Q_{JUL0.8}$	JULD80	July streamflow that is exceeded 80 percent of the time
$Q_{JULMEAN}$	Q7	Mean July streamflow
$Q_{AUG0.2}$	AUGD20	August streamflow that is exceeded 20 percent of the time

Table 1.   Definition of streamflow characteristics used in this study.—Continued

[--, StreamStats label has not been defined for this streamflow characteristic]

Streamflow characteristic	StreamStats abbreviation for streamflow characteristic	Description of streamflow characteristic
$Q_{AUG0.5}$	AUGD50	August streamflow that is exceeded 50 percent of the time, also known as the median August streamflow
$Q_{AUG0.8}$	AUGD80	August streamflow that is exceeded 80 percent of the time
$Q_{AUGMEAN}$	Q8	Mean August streamflow
$Q_{SEP0.2}$	SEPD20	September streamflow that is exceeded 20 percent of the time
$Q_{SEP0.5}$	SEPD50	September streamflow that is exceeded 50 percent of the time, also known as the median September streamflow
$Q_{SEP0.8}$	SEPD80	September streamflow that is exceeded 80 percent of the time
$Q_{SEPMEAN}$	Q9	Mean September streamflow
$Q_{OCT0.2}$	OCTD20	October streamflow that is exceeded 20 percent of the time
$Q_{OCT0.5}$	OCTD50	October streamflow that is exceeded 50 percent of the time, also known as the median October streamflow
$Q_{OCT0.8}$	OCTD80	October streamflow that is exceeded 80 percent of the time
$Q_{OCTMEAN}$	Q10	Mean October streamflow
$Q_{NOV0.2}$	NOVD20	November streamflow that is exceeded 20 percent of the time
$Q_{NOV0.5}$	NOVD50	November streamflow that is exceeded 50 percent of the time, also known as the median November streamflow
$Q_{NOV0.8}$	NOVD80	November streamflow that is exceeded 80 percent of the time
$Q_{NOVMEAN}$	Q11	Mean November streamflow
$Q_{DEC0.2}$	DECD20	December streamflow that is exceeded 20 percent of the time
$Q_{DEC0.5}$	DECD50	December streamflow that is exceeded 50 percent of the time, also known as the median December streamflow
$Q_{DEC0.8}$	DECD80	December streamflow that is exceeded 80 percent of the time
$Q_{DECMEAN}$	Q12	Mean December streamflow

Exploratory analyses were performed by using the all-subsets OLS regression in the Exploratory Regression tool in ArcGIS Desktop 10.2 (Esri, Inc., 2014). The all-subsets OLS regression performs analyses on all combinations of explanatory variables and provides several statistical diagnostics, which are (1) the adjusted coefficient of determination (R^2), (2) the statistical significance of the coefficients of the explanatory variables (as determined by a statistical probability level [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). The exploratory analyses were used to identify a best-fit OLS regression with the most important and consistent combination of candidate basin characteristics.

Determination of Hydrologic Regions

The hydrologic region boundaries for Montana were based on exploratory analyses in conjunction with consideration of the regional boundaries from previous studies of peak-flow frequencies and streamflow characteristics in Montana (Sando, S.K., and others, 2016; Parrett and Johnson, 2004). Initially, the relation between the streamflow characteristic and basin characteristics was investigated on a statewide basis using the all-subsets OLS regression. The statewide analysis determined that two basin characteristics (contributing drainage area and mean annual precipitation) provided the best multivariate regression as determined by the adjusted R^2 values. Streamflow characteristics were then predicted with an OLS regression using contributing drainage area and mean annual precipitation, and the standardized residuals were computed for each gaging station. The residuals were plotted in conjunction with the hydrologic region boundaries defined by Sando, Roy, and others (2016). These hydrologic regions

Table 2.   Selected basin characteristics evaluated as explanatory variables in the regional regression equations.

[-- not applicable]

StreamStats abbreviation	Abbreviations for basin characteristics discussed in this report	Definition
APRAVTMP	--	Mean April temperature, in degrees Fahrenheit[1]
AUGAVTMP	--	Mean August temperature, in degrees Fahrenheit[1]
BASINPERIM	--	Perimeter of the contributing drainage area, in miles[2]
BSLDEM30M	--	Mean basin slope for the contributing drainage area[2]
COMPRAT	--	Compactness ratio for the contributing drainage area, computed using the CONTDA and BASINPERIM basin characteristics[2]
CONTDA	*A*	Contributing drainage area, in square miles[2]
DECAVTMP	--	Mean December temperature, in degrees Fahrenheit[1]
EL5000	--	Percentage of contributing drainage basin above 5,000 feet elevation[2]
EL5500	--	Percentage of contributing drainage basin above 5,500 feet elevation[2]
EL6000	--	Percentage of contributing drainage basin above 6,000 feet elevation[2]
EL6500	--	Percentage of contributing drainage basin above 6,500 feet elevation[2]
EL7000	--	Percentage of contributing drainage basin above 7,000 feet elevation[2]
ELEV	--	Mean basin elevation, in feet[2]
ELEVMAX	--	Maximum basin elevation, in feet[2]
ET0306MOD	--	Spring (March - June) mean monthly evapotranspiration (2000 - 2012), in inches per month[3]
ET0710MOD	--	Summer (July - October) mean monthly evapotranspiration (2000 - 2012), in inches per month[3]
FEBAVTMP	--	Mean February temperature, in degrees Fahrenheit[1]
IRRIGAT_MT	--	Percent of contributing drainage basin under some irrigation regime[4]
JANAVTMP	--	Mean January temperature, in degrees Fahrenheit[1]
JULAVTMP	--	Mean July temperature, in degrees Fahrenheit[1]
JUNAVTMP	--	Mean June temperature, in degrees Fahrenheit[1]
LAKESNHDH	--	Percentage of contributing drainage basin in lakes, ponds, and reservoirs[5]
LC01CRPHAY	--	Percentage of contributing drainage basin covered by agriculture[6]
LC01DEV	--	Percentage of contributing drainage basin covered by urban land[6]
LC01FOREST	--	Percentage of contributing drainage basin covered by forest[6]
LC01WETLND	--	Percentage of contributing drainage basin covered by wetlands[6]
MARAVTMP	--	Mean March temperature, in degrees Fahrenheit[1]
MAXTEMP	--	Mean annual maximum temperature, in degrees Fahrenheit[1]
MAYAVTMP	--	Mean May temperature, in degrees Fahrenheit[1]
MINBELEV	--	Minimum basin elevation, in feet[2]
MINTEMP	--	Mean annual minimum temperature, in degrees Fahrenheit[1]
NFSL30_30M	--	Percentage of contributing drainage basin with north-facing slopes greater than or equal to 30 percent[2]
NOVAVTMP	--	Mean November temperature, in degrees Fahrenheit[1]
OCTAVTMP	--	Mean October temperature, in degrees Fahrenheit[1]
PRECIP	*P*	Mean annual precipitation, in inches[1]
RELIEF	--	Difference between the maximum and minimum elevations of the basin, in feet[2]

Table 2. Selected basin characteristics evaluated as explanatory variables in the regional regression equations.—Continued

[-- not applicable]

StreamStats abbreviation	Abbreviations for basin characteristics discussed in this report	Definition
SEPAVTMP	--	Mean September temperature, in degrees Fahrenheit[1]
SLOP30_30M	--	Percentage of contributing drainage basin with slopes greater than or equal to 30 percent[2]
SLOP50_30M	SLP_{50}	Percentae of contributing drainage basin with slopes greater than or equal to 50 percent[2]
TEMP	--	Mean annual temperature, in degrees Fahrenheit[1]

[1]Parameter-elevation Regressions on Independent Slopes Model (PRISM Climate Group, 2004) and Long Term Mean Climate Grids for Canada (Natural Resources Canada, 2015) for 1971–2000.

[2]30-meter National Elevation Dataset (NED; Gesch and others, 2002). Elevation refers to distance above North American Vertical Datum of 1988.

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

[4]Final Land Unit (FLU) classification (Montana Department of Revenue, 2014).

[5]National Hydrological Dataset (NHD) High-resolution waterbodies (Horizon Systems Corporation, 2013).

[6]National Land Cover Dataset 2001 (NLCD; Homer and others, 2007), Northern Land Cover of Canada Circa 2000 (LCC2000; Natural Resources Canada, 2009).

(table 3, fig.1) generally represented the spatial patterns in the residuals and thus were used in the final development of the regional regression equations for streamflow characteristics in this study.

Determination of Significant Basin Characteristics

A second round of exploratory data analyses was performed using the all-subsets OLS regression for each of the eight hydrologic regions to evaluate explanatory variables (table 2) for each of the 54 streamflow characteristics (table 1, table 1–1). The exploratory analyses were used to identify a best-fit OLS regression with the most important and consistent combination of candidate basin characteristics for each hydrologic region. The criteria for selecting the best-fit OLS regression for each hydrologic region primarily were 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, (2) a VIF values less than 2, (3) normality of the residuals, and (4) residuals with a nonsignificant (p-value equal to or greater than 0.05) global Moran's I Index value. Additional considerations in selection of the best-fit OLS regression included investigation of (1) the normality of the explanatory variables, (2) the Akaike Information Criterion (Akaike, 1973) of each regression equation, and (3) the hydrologic basis for relations between streamflow characteristics and the explanatory variables. These three additional criteria were generally only used to choose between two regression equations that produced similar results.

The exploratory analyses met the above criteria for four of the eight hydrologic regions, located generally in western

Montana. These four hydrologic regions are the West, Northwest, Upper Yellowstone-Central Mountain, and Southwest (table 3). Exploratory analyses for the four hydrologic regions located generally in eastern Montana (Northwest Foothills, Northeast Plains, East-Central Plains, and Southeast Plains) determined that regression equations could not reliably be developed for the 54 streamflow characteristics using the above criteria. Additionally, even using relaxed criteria, regression equations for many of the streamflow characteristics could not be developed due to the very high occurrence of streamflow characteristics being equal to zero. Non-zero streamflow characteristics are required to develop regression equations. A total of 152 gaging stations were located in the four hydrologic regions in western Montana (fig. 1, table 1–1).

Of the 40 candidate basin characteristics (table 2), 3 met the above criteria and were used to develop regional regression equations. The three basin characteristics are contributing drainage area (A), mean annual precipitation (P), and percentage of the contributing basin with slopes greater than or equal to 50 percent (SLP_{50}; table 1–2 in appendix 1 at the back of this report chapter [available at http://dx.doi.org/10.3133/sir20155019G]). The most consistently significant basin characteristic was contributing drainage area (A), which was used in all of the regression equations. One other basin characteristics was determined to be significant for one or more of the hydrologic regions, so the regression equations included either mean annual precipitation (P) or percentage of the contributing basin with slopes greater than 50 percent (SLP_{50}) as a second explanatory variable.

Table 3.   Hydrologic regions in Montana.

Hydrologic region (ordered clockwise from northwestern Montana)	Hydrologic region number (fig. 1)	General description and extent[1]	Number of selected gaing stations	Regression equations developed
West	1	Mountains and valleys west of Continental Divide; parts of Flathead and Blackfoot River Basins	53	Yes
Northwest	2	Eastern parts of Flathead and Blackfoot River Basins; mountains and foothills east of the Continental Divide and northeast of Missoula, Montana	26	Yes
Northwest Foothills	3	Foothills and plains of the Marias, Teton, Sun, and Dear-born River Basins near Great Falls, Montana	10	No
Northeast Plains	4	Rolling plains of the Milk River Basin upstream from Glasgow, Montana; foothills and plains part of the Judith River Basin	26	No
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	12	No
Southeast Plains	6	Rolling plains of southern part of Yellowstone River Basin east of Billings, Montana	13	No
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	49	Yes
Southwest	8	Mountains and valleys of the Missouri River Basin upstream from the Dearborn River	24	Yes

[1] The general description and extent of the region as defned by Parrett and Johnson (2004).

Regional Regression Equations

After selection of the best-fit OLS regressions, final regression equations were performed using the Weighted-Multiple-Linear Regression Program (WREG; Eng and others, 2009). The GLS regression analyses were used for two streamflow characteristics that are frequency-based streamflow characteristics (Q_{AL7Q10}, Q_{S214Q5}). The OLS regression analyses were used for the remaining 52 streamflow characteristics because GLS regression analysis is not recommended for streamflow characteristics that are not frequency based (Eng and others, 2009). The 52 streamflow characteristics include the mean annual streamflow; mean monthly streamflow; and the 20-, 50-, and 80-percent durations for annual and monthly streamflows.

For each gaging station, the streamflow for a specified streamflow characteristic (Q) and explanatory variables (basin characteristics) that were used in developing the final regression equations are presented in table 1–1 and table 1–2, respectively. Because of nonlinear relations between Q and the explanatory variables, all variables were log-transformed before analysis. Basin characteristics that are computed as a percentage of the basin (percent forest, for example) could potentially have a value of zero; therefore, a value of one was added to these basin characteristics before performing the log-transformation. The regression equations are of the following log-linear form:

$$\log Q = \log K + a_1 \log x_1 + a_2 \log x_2 + \dots a_p \log x_p, \quad (1)$$

where

Q	is the streamflow, in cubic feet per second;
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

$$Q = K' x_1^{a_1} x_2^{a_2} \dots x_p^{a_p}, \quad (2)$$

where

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

Generalized Least Squares Regressions

Final regression equations for Q_{AL7Q10} and Q_{S214Q5} were developed with GLS regression (Tasker and Stedinger, 1989) using WREG (Eng and others, 2009) for each hydrologic region. The GLS regression is preferred for developing regression equations when gaging stations have varying lengths of record, and thus varying precisions, because the GLS

regression weights the variance of a gaging station according to its record length. The GLS regression also accounts for cross correlation of concurrent streamflows observed at different gaging stations. As a result, the GLS regression provides better estimates of the predictive accuracy of the regression equations and also provides relatively unbiased estimates of the variance of the underlying regression equation error (Tasker and Stedinger, 1989).

Final regression equations for Q_{AL7Q10} and Q_{S214Q5} are presented in table 1–3 in appendix 1 at the back of this report chapter (available at http://dx.doi.org/10.3133/sir20155019G). The WREG program reports multiple performance metrics that use the term "model" to reference to the regression equations. These performance metrics are reported in table 1–3 for GLS regressions and include the model error variance (, 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 the streamflow characteristic for ungaged sites. The pseudo R^2 and *SEM* are metrics that indicate how well the equation predicted the streamflow characteristic for the gaging stations used in the analysis.

Although the *SEP* provides a mean measure of how well the regression equation predicts streamflow characteristics within the region, the *SEP* should be calculated for individual estimates at the ungaged site of interest and subsequently used to determine the prediction intervals (*PI*) for the estimated streamflow at the ungaged site (Q_0). The following equation can be used to calculate the standard error of prediction for an ungaged site (SEP_0):

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

where
- SEP_0 is the standard error of prediction, in log units, for an estimate of Q at an ungaged site;
- σ_δ^2 is the model error variance, in log units (table 1–3), 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 ; and
- $(X^T\Lambda^{-1}X)^{-1}$ is the covariance matrix for the GLS regional regression equation (table 1–4 in appendix 1 at the back of this report chapter [available at http://dx.doi.org/10.3133/sir20155019G]).

The SEP_0 is used to calculate a confidence interval for a desired confidence level (*CL*) using the following equations:

$$\alpha = \frac{100-CL}{100} \tag{4}$$

where
- α is the significance level for the desired confidence level (*CL*), and
- *CL*   is the desired confidence level, in percent.

$$CI_{0,\alpha} = t_{\left(\frac{\alpha}{2}\,n-p-1\right)}(SEP_0) \tag{5}$$

where
- $CI_{0,\alpha}$ is the confidence interval, in log units, for the ungaged site;
- $t_{\left(\frac{\alpha}{2}\,n-p-1\right)}$ is the one-tailed Student's t value for a significance level, α, and (n-p-1) degrees of freedom;
- SEP_0 is the standard error of prediction, in log units, for an estimate of Q at an ungaged site; 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.

The estimated streamflow characteristic at the ungaged site (Q_0) is log transformed (log Q_0) and used with the confidence interval to determine the prediction interval using the following equation:

$$PI_{U_0,L_0} = 10^{(\log(Q_0)\pm CI_{0,\alpha})} \tag{6}$$

where
- PI_{U_0,L_0} is the prediction interval, in cubic feet per second, for the estimate Q_0 at the ungaged site with a significance level of α;
- U_0 is for the upper prediction interval for the ungaged site;
- L_0 is for the lower prediction interval for the ungaged site;
- Q_0 is the estimated streamflow characteristic, in cubic feet per second, for the ungaged site; and
- $CI_{0,\alpha}$ is the confidence interval, in log units, for the ungaged site.

Ordinary Least Squares Regressions

Final regression equations for the remaining 52 streamflow characteristics were developed for each hydrologic region with OLS regression using WREG (Eng and others, 2009). The streamflow characteristics included the mean annual and mean monthly streamflows, and the 20-, 50-, and 80-percent durations for annual and monthly streamflows. The GLS regression was not used for developing regression equations for these 52 characteristics because GLS regression analysis

is not recommended for streamflow characteristics that are not frequency based (Eng and others, 2009), and these streamflow characteristics were developed from daily mean streamflows and are not frequency based streamflow characteristics.

Final regression equations for the OLS regressions are presented in table 1–3. Also included in table 1–3 are measures of reliability of the equations, including the mean square error (MSE, in log units), root mean square error ($RMSE$, in log units and percent), and R^2 (in percent). The MSE and $RMSE$ are measures that indicate how well the equation will predict Q_0 at an ungaged site, and R^2 indicates how well the equation predicted Q for the gaging stations used in the analysis.

Although the MSE provides an indication of the average reliability of a regression equation within a region, the SEP should be calculated for individual estimates at the ungaged site of interest and subsequently used to determine the prediction intervals for Q_0 at the ungaged site. The following equation can be used to calculate the standard error of prediction for an ungaged site (SEP_0):

$$SEP_0 = \sqrt{MSE\,(1 + x_0(X^T X)^{-1} x_0^T)} \qquad (7)$$

where

SEP_0 is the standard error of prediction, in log units, for an estimate of Q at an ungaged site;

MSE is the mean square error, in log units (table 1–3), for the appropriate regression equation for 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 ; and

$(X^T X)^{-1}$ is the covariance matrix for the OLS regional regression equation (table 1–4).

The SEP_0 can be used to calculate a confidence interval for the same estimate using equation 5, and the prediction interval can be estimated using equation 6.

Limitations of Regional Regression Equations

The regional regression equations were developed to estimate streamflow characteristics for ungaged streams in western Montana (fig. 1) that are unaffected or minimally affected by regulation. The regression equations may be applicable to ungaged sites that are minimally affected by regulation if less than 20 percent of the drainage area for the ungaged site is upstream from dams and no large diversion canals are upstream from the ungaged site. Streamflow characteristics used to develop these regression equations were selected from gaging stations that are considered unregulated or from unregulated periods at gaging stations that are considered regulated. The regression equations (table 1–3) might not be reliable for

ungaged sites if the values of any explanatory variables (basin characteristics) for that site are outside the range of values used to develop the equations (table 4).

All of the regression equations developed in this study are multivariate; therefore, the regression equations might not be reliable if the explanatory variables used for an at-site estimate do not fall within the joint probability distribution 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, even though the values are within the ranges shown in table 4. A leverage metric can be used to identify such sites. The leverage for the at-site estimate (h_0) can be determined using equation 8 for regression equations developed using GLS and equation 9 for regression equations developed using OLS:

$$h_0 = x_0 (X^T \Lambda^{-1} X)^{-1} x_0^T \qquad (8)$$

where

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 , and

$(X^T \Lambda^{-1} X)^{-1}$ is the covariance matrix for the GLS regional regression equation (table 1–4).

$$h_0 = x_0 (X^{T-1} X)^{-1} x_0^T \qquad (9)$$

where

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

$(X^T X)^{-1}$ is the covariance matrix for the OLS regional regression equation (table 1–4).

If the at-site leverage (h_0) is greater than $3p/n$ (where p is the number of explanatory variables used in the regional regression equation and is equal to two for all equations in this study, and n is the number of gaging stations used to develop the regional regression equation) then the basin characteristics at that site are considered far from the center of the joint distribution of basin characteristics and thus the resulting estimates might be unreliable (Helsel and Hirsch, 2002; Eng and others, 2009).

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 where the stream originates. For

Table 4. Ranges of values for basin characteristics used to develop regional regression equations.

[A, contributing drainage area, in square miles; P, mean annual precipitation, in inches; SLP_{50}, percentage of basin with slope greater than 50 percent; Min, minimum; Max, maximum; --, not used in regression equations for indicated region]

Hydrologic region (ordered clockwise from northwestern Montana)	A		P		SLP_{50}	
	Min	**Max**	**Min**	**Max**	**Min**	**Max**
West	6.40	2,516	--	--	1.87	67.47
Northwest	7.74	1,556	20.74	83.16	0.06	65.94
Upper Yellowstone-Central Mountain	28.1	2,616	16.45	38.85	--	--
Southwest	24.4	2,472	16.74	37.02	0.96	22.55

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 contributing drainage area at 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.

Drainage-Area Ratio Methods

Streamflow characteristic estimates for ungaged sites that are on the same stream as a gaging station can be determined using the drainage-area ratio method. The drainage-area ratio method requires the computed streamflow characteristic at the gaging station and the contributing drainage area of both the ungaged site and the gaging station. A major assumption of the drainage-area ratio method is that the streamflow characteristics of the ungaged site are similar to streamflow characteristics of the gaging station and the sites are located on the same stream. Streamflow characteristics (table 1–1) and contributing drainage area (table 1–2) for the gaging station and the drainage area for the ungaged site are used in the following equation to estimate streamflow characteristics at the ungaged site:

$$Q_u = Q_g \left(\frac{A_u}{A_g}\right)^{exp_{Q,R}} \quad (10)$$

where

Q_u is the streamflow characteristic for the ungaged site,

Q_g is the streamflow characteristic for the gaging station (table 1–1),

A_u is the contributing drainage area for the ungaged site,

A_g is the contributing drainage area for the gaging station (table 1–2), and

exp is the coefficient for drainage area adjustment for the streamflow characteristic (Q) and region (R) of the gaging station (table 1–3).

The drainage-area ratio method is limited to sites that are within a range of $0.5A_g$–$1.5A_g$. For sites that are outside of this range, the regression equation (table 1–3) might provide more reliable estimates of Q than equation 10. The regression coefficient (*exp*) for the drainage area adjustment was computed for each streamflow characteristic in each region using OLS regression methods. If an ungaged site is between two gaging stations on the same stream, the logarithms (*log*) of Q and A for the gaging station can be linearly interpolated with the logarithm of A for the ungaged site to estimate Q at the ungaged site by using the following equation:

$$logQ_u = logQ_{g1} + \left(\frac{logQ_{g2}-logQ_{g1}}{logA_{g2}-logA_{g1}}\right)(logA_u-logA_{g1}) \quad (11)$$

where

Q_u is the streamflow characteristic for the ungaged site,

Q_g is the streamflow characteristic for the gaging station (table 1–1),

A_u is the contributing drainage area for the ungaged site,

A_g is the contributing drainage area for the gaging station (table 1–2), and

g_1 and g_2 are for gaging stations 1 and 2, respectively.

If the contributing drainage area of the ungaged site is outside of the range of $0.5A_g$–$1.5A_g$ of either gaging station, equation 11 might provide unreliable estimates. A major assumption for using equation 11 is that streamflow characteristics between the two gaging stations are similar and that the two gaging stations have similar periods of record.

Tools for Estimating Streamflow Characteristics at Ungaged Sites in Montana

StreamStats is a Web-based GIS application that provides users with access to an assortment of analytical tools, streamflow and basin characteristics, and descriptive information for USGS gaging stations and user-selected ungaged sites (McCarthy and others, 2016). The StreamStats application

(http://water.usgs.gov/osw/streamstats/; U.S. Geological Survey, 2015a) includes geospatial datasets, basin characteristics, and analytical processing tools to estimate streamflow characteristics at ungaged sites in western Montana based on regional regression equations developed for this study. Instructions and definitions of basin and streamflow characteristics are provided on the StreamStats Web page.

Additionally, tools for estimating streamflow characteristics at ungaged sites using regional regression equations (table 1–3) or drainage-area ratio method (equation 10 and table 1–3) are available in appendix 2 (available at http://dx.doi.org/10.3133/sir20155019G), which contains spreadsheet tools consisting of two Excel® tables with built-in functions to estimate streamflow characteristics based on user inputs. Estimates of streamflow characteristics, prediction intervals for a selected confidence interval, and test results for high leverage at an ungaged site can be computed in table 2–1 using regional regression equations (table 1–3) and basin characteristics, which are provided by the user after performing a basin delineation and computing basin characteristics in StreamStats. Estimates of streamflow characteristics at an ungaged site on the same stream as a gaging station with computed streamflow characteristics can be computed in table 2–2 using the drainage-area ratio method (equation 10), a user-selected index site (table 1–1), and contributing drainage area for the ungaged site, which is provided by the user and can be calculated in StreamStats. Instructions are provided in the respective tables.

Summary

The U.S. Geological Survey, in cooperation with the Montana Department of Environmental Quality and the Montana Department of Natural Resources and Conservation, developed regional regression equations based on basin and streamflow characteristics for streamflow-gaging stations through water year 2009 that can be used to estimate selected streamflow characteristics for ungaged sites in western Montana. The methods allow estimation of low-flow frequencies; mean annual and mean monthly streamflows; and 20-, 50-, and 80-percent durations for annual and monthly streamflows.

The regression equations are based on analyses of low-flow frequencies, daily streamflow data, and basin characteristics at 152 streamflow-gaging stations in four hydrologic regions in or near western Montana and were developed using generalized least squares and ordinary least squares regression methods. For this study, 40 basin characteristics were selected as candidate explanatory variables in the regression analyses, and 3 of these basin characteristics were determined to be significant for estimating streamflow characteristics and 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 for one or more of

the hydrologic regions included mean annual precipitation and percentage of the contributing basin with slopes greater than 50 percent.

In addition to the final regression equations by hydrologic region for estimating streamflow characteristics at ungaged sites in western Montana, the performance metrics for both the generalized least squares and ordinary least squares are presented. Methods also are provided for estimating prediction intervals and testing for high leverage for regional regression equation estimates. Additionally, two drainage-area ratio methods are provided for estimating streamflow characteristics at ungaged sites near streamflow-gaging stations located on the same stream as the ungaged site of interest.

The regional regression equations were developed using streamflow characteristics at unregulated streamflow-gaging stations having 10 or more years of record (or only the unregulated period of record for regulated streamflow-gaging stations were used); therefore, the regional regression equations only are applicable to ungaged sites that are considered unregulated. For this study, a streamflow-gaging station is considered to be unregulated if cumulative drainage area upstream from all upstream dams is less than 20 percent of the drainage basin for the streamflow-gaging station, and if no large diversion canals are upstream from the streamflow-gaging station.

All of the data used in calculating basin and streamflow characteristics were derived from publically available data and are available through the U.S. Geological Survey StreamStats program (http://water.usgs.gov/osw/streamstats/) for Montana. 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.

Spreadsheet tools for estimating streamflow characteristics at ungaged site using regression equations or the drainage-area ratio method are provided as spreadsheet tools. Basin characteristics used in the spreadsheet tools need to be provided by the user and can be computed using StreamStats. The spreadsheet tools compute estimates for streamflow characteristics, prediction intervals for the streamflow estimate, and test for high leverage for estimates using the regression equations. The spreadsheet tools provide the streamflow characteristics for streamflow-gaging stations and can be used to estimate streamflows using the drainage-area ratio method for a user-provided drainage area at an ungaged site.

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.

Berwick, V.K., 1958, Floods in eastern Montana—Magnitude and frequency: U.S. Geological Survey Open-File Report 58–15, 23 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., 2005, Statistical summaries of streamflow in Montana and adjacent areas, water years 1900 through 2002: U.S. Geological Survey Scientific Investigations Report 2004–5266, 317 p. [Also available at http://pubs.usgs.gov/sir/2004/5266/.]

McCarthy, P.M., 2016, Streamflow characteristics based on data through water year 2009 for selected streamflow-gaging stations in or near Montana: U.S. Geological Survey Scientific Investigations Report 2015–5019–E, 10 p. [Also available at http://dx.doi.org/10.3133/sir20155019E.]

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.]

Montana Department of Revenue, 2014, Revenue Final Land Unit (FLU) Classification: accessed December 16, 2014, at ftp://ftp.geoinfo.msl.mt.gov/Data/Spatial/NonMSDI/Geodatabases/revenue_flu.zip.

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, no. 4, 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., 1984, Streamflow characteristics of the Yellowstone River basin, Montana, through September 1982: U.S. Geological Survey Water-Resources Investigations Report 84–4063, 78 p. [Also available at http://pubs.er.usgs.gov/publication/wri844063.]

Parrett, Charles, and Cartier, K.D., 1990, Methods for estimating monthly streamflow characteristics at ungaged sites in western Montana: U.S. Geological Survey Water-Supply Paper 2365, 30 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.

Parrett, Charles, Omang, R.J., and Hull, J.A., 1982, Mean annual runoff and peak flow estimates based on channel geometry of streams in southeastern Montana: U.S. Geological Survey Water-Resources Investigations Report 82–4092, 33 p. [Also available at http://pubs.er.usgs.gov/publication/wri824092.]

Parrett, Charles, Omang, R.J., and Hull, J.A., 1983, Mean annual runoff and peak flow estimates based on channel geometry of streams in northeastern and western Montana: U.S. Geological Survey Water-Resources Investigations Report 83–4046, 53 p. [Also available at http://pubs.er.usgs.gov/publication/wri834046.]

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

Sando, Roy, Sando, S.K., McCarthy, P.M., and Dutton, D.M., 2016, Methods for estimating peak-flow frequencies at ungaged sites in Montana based on data through water year 2011: U.S. Geological Survey Scientific Investigations Report 2014–5019–F, 30 p. [Also available at http://dx.doi.org/10.3133/sir20155019F.]

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., Morgan, T.J., Dutton, D.M., and McCarthy, P.M., 2009, Estimation of streamflow characteristics for Charles M. Russell National Refuge, northeastern Montana: U.S. Geological Survey Scientific Investigations Report 2009–5009, 60 p. [Also available at http://pubs.usgs.gov/sir/2009/5009/.]

Shields, R.R., and White, M.K., 1981, Streamflow characteristics of the Hudson Bay and Upper Missouri River Basins, Montana, through 1979: U.S. Geological Survey Water-Resources Investigations Report 81–32, 144 p.

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, The StreamStats Program: accessed March 5, 2015, at http://water.usgs.gov/osw/streamstats/.

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

Waltemeyer, S.D., and Shields, R.R., 1982, Streamflow characteristics of the upper Columbia River Basin, Montana, through 1979: U.S. Geological Survey Water-Resources Investigations Report 81–82, 74 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.

Appendixes 1 and 2

Appendix 1. Supplemental Information Relating to the Regional Regression Analyses

   This appendix contains supplemental information relating to the regional regression analyses. For streamflow-gaging stations used to develop the regional regression equations, streamflow characteristics are presented in table 1–1 and basin characteristics are presented in table 1–2. Regression equations used for estimating streamflow characteristics at ungaged sites in western Montana are presented in table 1–3. Covariance matrices for generalized and ordinary least squares regressions are presented in table 1–4.

An Excel file containing the tables is available at http://dx.doi.org/10.3133/sir20155019G.

Table 1–1.   Streamflow characteristics for each streamflow-gaging station used to develop the regional regression equations.

Table 1–2.   Basin characteristics for each streamflow-gaging station used to develop the regional regression equations.

Table 1–3.   Regression equations for estimating streamflow characteristics at ungaged sites in western Montana.

Table 1–4.   Covariance matrices for the generalized least squares regressions, $[X^{\mathrm{T}}\Lambda^{-1}X]^{-1}$; and for the ordinary least squares regressions, $[X^{\mathrm{T}}X]^{-1}$.

Appendix 2. Estimates of Streamflow Characteristics at Ungaged Sites

This appendix contains spreadsheet tools consisting of two Excel® tables with built-in functions to estimate streamflow characteristics based on user inputs to estimate streamflow characteristics at ungaged sites. Estimates of streamflow characteristics, prediction intervals for a selected confidence interval, and test results for high leverage at an ungaged site can be computed in table 2–1 using regional regression equations (table 1–3) and basin characteristics, which are provided by the user after performing a basin delineation and computing basin characteristics in StreamStats. Estimates of streamflow characteristics at an ungaged site on the same stream as a streamflow-gaging station with computed streamflow characteristics can be computed in table 2–2 using the drainage-area ratio method (equation 10), a user-selected index site (table 1–1), and contributing drainage area for the ungaged site, which is provided by the user and can be calculated in StreamStats. Instructions are provided in the respective tables.

An Excel file containing the tables is available at http://dx.doi.org/10.3133/sir20155019G/.

Table 2–1.   Estimates of streamflow characteristics at an ungaged site based on regional regression equations and user-determined basin characteristics.

Table 2–2.   Estimates of streamflow characteristics at an ungaged site using the Drainage-Area Ratio Method.

USGS