

Prepared in cooperation with Idaho Transportation Department

# Estimating Peak-Flow Frequency Statistics for Selected Gaged and Ungaged Sites in Naturally Flowing Streams and Rivers in Idaho



Scientific Investigations Report 2016–5083

**Cover:** U.S. Geological Survey streamgage No. 12392155, Lightning Creek at Clark Fork, Idaho. Heavy rainfall and associated snowmelt increased the creek streamflow from about 200 cubic feet per second (ft<sup>3</sup>/s) to more than 7,000 ft<sup>3</sup>/s within 24 hours. Photograph by Ryan Smith, U.S. Geological Survey, December 9, 2015.

# **Estimating Peak-Flow Frequency Statistics for Selected Gaged and Ungaged Sites in Naturally Flowing Streams and Rivers in Idaho**

By Molly S. Wood, Ryan L. Fosness, Kenneth D. Skinner, and Andrea G. Veilleux

Prepared in cooperation with Idaho Transportation Department

Scientific Investigations Report 2016–5083

**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

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 <http://www.usgs.gov> or call 1–888–ASK–USGS.

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

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Wood, M.S., Fosness, R.L., Skinner, K.D., and Veilleux, A.G., 2016, Estimating peak-flow frequency statistics for selected gaged and ungaged sites in naturally flowing streams and rivers in Idaho: U.S. Geological Survey Scientific Investigations Report 2016–5083, 56 p., <http://dx.doi.org/10.3133/sir20165083>.

ISSN 2328-0328 (online)

# Contents

Abstract.....	1
Introduction.....	2
Purpose and Scope .....	2
Description of Study Area .....	3
Previous Studies .....	5
Data Screening and Compilation .....	6
Streamgage Selection .....	6
Basin and Climatic Characteristics .....	8
Estimating Peak-Flow Frequency Statistics at Selected Gaged Sites.....	13
Peak-Flow Frequency Analysis .....	13
Expected Moments Algorithm Frequency Analysis and Multiple Grubbs-Beck Test for Detecting Low Outliers.....	13
Regional Skew Analysis.....	16
Final Peak-Flow Frequency Statistics .....	16
Estimating Peak-Flow Frequency Statistics at Ungaged Sites Through a Regional Regression Analysis .....	17
Determination of Regions for Regional Regression Analysis.....	17
Exploratory Data Analysis and Ordinary Least Squares Regression .....	17
Generalized Least Squares Regression.....	18
Regional Peak-Flow Regression Equations.....	18
Uncertainty and Limitations .....	22
Alternative to Regional Regression Equations .....	24
Comparison of Results from Previous Studies.....	25
Estimating Flow Statistics Using StreamStats.....	27
Potential Areas for Further Study .....	27
Summary .....	28
Acknowledgments.....	29
References Cited .....	30
Appendix A. Supplemental Information for the Development of Regional Peak-Flow Regression Equations in Idaho .....	35
Appendix B. Regional Skew Regression Analysis for the Pacific Northwest, including Idaho, Montana, Oregon, and Washington .....	37

## Figures

1. Map of study area showing physiographic and ecological divisions in Idaho and bordering States .....	4
2. Flow chart showing streamgage selection and screening process for the regional peak-flow regression analysis in Idaho .....	7
3. Map showing locations of streamgages in Idaho and bordering States selected for the regional peak-flow regression analysis .....	9
4. Graph showing comparison between Geographic Information System (GIS) estimated drainage area and NWIS-reported drainage area for streamgages in Idaho and bordering States .....	12
5. Graphs showing output from the U.S. Geological Survey (USGS) PeakFQ program showing Expected Moments Algorithm (EMA) treatment of historical peaks and gaps in systematic record using a perception threshold and fitted frequency curve with historical peaks, systematic peaks, confidence limits, potentially influential low-flow (PILF) threshold, and peaks identified as PILF outliers for USGS streamgage 13185000, Boise River near Twin Springs, Idaho.....	15

## Tables

1. Annual exceedance probabilities with corresponding recurrence intervals used in computing peak-flow statistics and regional regression equations.....	3
2. Basin and climatic characteristics estimated for selected streamgages and evaluated in regional regression equations.....	10
3. General perception threshold and flow interval settings applied to various scenarios in the Expected Moments Algorithm analysis to estimate peak-flow statistics at streamgages .....	14
4. Regional regression equations for estimating peak-flow frequency statistics at ungaged sites in Idaho.....	20
5. Ranges in values of basin and climatic characteristics used to develop regional peak-flow regression equations for Idaho .....	23
6. Regression coefficients for generalized least squares regressions relating AEP-percent peak flow to drainage area, used as exponents in the alternative drainage-area ratio exponent.....	24
7. Selected results of regional peak-flow regression equations compared with results in Berenbrock (2002) and Hortness and Berenbrock (2003).....	25
8. Average standard errors of prediction for selected peak-flow statistics estimated using regional regression equations in the current and selected previous studies in Idaho.....	26

## Conversion Factors

Inch/Pound 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 <sup>2</sup> )	259.0	hectare (ha)
square mile (mi <sup>2</sup> )	2.590	square kilometer (km <sup>2</sup> )
Flow rate		
cubic foot per second (ft <sup>3</sup> /s)	0.02832	cubic meter per second (m <sup>3</sup> /s)
inch per year (in/yr)	25.4	millimeter per year (mm/yr)

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

$$^{\circ}\text{F} = (1.8 \times ^{\circ}\text{C}) + 32.$$

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

$$^{\circ}\text{C} = (^{\circ}\text{F} - 32) / 1.8.$$

## Datums

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

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

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

## Abbreviations

AEP	Annual exceedance probability
AVP	Average variance of prediction
DAR	Drainage area ratio
GAGES II	Geospatial Attributes of Gages for Evaluating Streamflow, version II, database
GIS	Geographic information system
GLS	Generalized least squares
MEV	Model error variance
MGBT	Multiple Grubbs Beck test
MSE	Mean square error
NCSS	Number Crunching Statistical Software
NED	National elevation dataset
NWIS	National Water Information System
OLS	Ordinary least squares
$Q_{AEP}$	Peak-flow magnitude for annual exceedance probability (AEP)
ROI	Region of influence
SD	Standardized distance
SEM	Standard error of the model
SEP	Standard error of prediction
USGS	U.S. Geological Survey
WLS	Weighted least squares
WREG	Weighted Multiple Linear Regression program
WY	Water year

# Estimating Peak-Flow Frequency Statistics for Selected Gaged and Ungaged Sites in Naturally Flowing Streams and Rivers in Idaho

By Molly S. Wood, Ryan L. Fosness, Kenneth D. Skinner, and Andrea G. Veilleux

## Abstract

The U.S. Geological Survey, in cooperation with the Idaho Transportation Department, updated regional regression equations to estimate peak-flow statistics at ungaged sites on Idaho streams using recent streamflow (flow) data and new statistical techniques. Peak-flow statistics with 80-, 67-, 50-, 43-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities (1.25-, 1.50-, 2.00-, 2.33-, 5.00-, 10.0-, 25.0-, 50.0-, 100-, 200-, and 500-year recurrence intervals, respectively) were estimated for 192 streamgages in Idaho and bordering States with at least 10 years of annual peak-flow record through water year 2013. The streamgages were selected from drainage basins with little or no flow diversion or regulation. The peak-flow statistics were estimated by fitting a log-Pearson type III distribution to records of annual peak flows and applying two additional statistical methods: (1) the Expected Moments Algorithm to help describe uncertainty in annual peak flows and to better represent missing and historical record; and (2) the generalized Multiple Grubbs Beck Test to screen out potentially influential low outliers and to better fit the upper end of the peak-flow distribution. Additionally, a new regional skew was estimated for the Pacific Northwest and used to weight at-station skew at most streamgages. The streamgages were grouped into six regions (numbered 1\_2, 3, 4, 5, 6\_8, and 7, to maintain consistency in region numbering with a previous study), and the estimated peak-flow statistics were related to basin and climatic characteristics to develop regional regression equations using a generalized least squares procedure. Four out of 24 evaluated basin and climatic characteristics were selected for use in the final regional peak-flow regression equations.

Overall, the standard error of prediction for the regional peak-flow regression equations ranged from 22 to 132 percent. Among all regions, regression model fit was best for region 4 in west-central Idaho (average standard error of prediction=46.4 percent; pseudo- $R^2$ >92 percent) and region 5 in central Idaho (average standard error of prediction=30.3 percent; pseudo- $R^2$ >95 percent). Regression model fit was poor for region 7 in southern Idaho (average standard error of prediction=103 percent;

pseudo- $R^2$ <78 percent) compared to other regions because few streamgages in region 7 met the criteria for inclusion in the study, and the region's semi-arid climate and associated variability in precipitation patterns causes substantial variability in peak flows.

A drainage area ratio-adjustment method, using ratio exponents estimated using generalized least-squares regression, was presented as an alternative to the regional regression equations if peak-flow estimates are desired at an ungaged site that is close to a streamgage selected for inclusion in this study. The alternative drainage area ratio-adjustment method is appropriate for use when the drainage area ratio between the ungaged and gaged sites is between 0.5 and 1.5.

The updated regional peak-flow regression equations had lower total error (standard error of prediction) than all regression equations presented in a 1982 study and in four of six regions presented in 2002 and 2003 studies in Idaho. A more extensive streamgage screening process used in the current study resulted in fewer streamgages used in the current study than in the 1982, 2002, and 2003 studies. Fewer streamgages used and the selection of different explanatory variables were likely causes of increased error in some regions compared to previous studies, but overall, regional peak-flow regression model fit was generally improved for Idaho. The revised statistical procedures and increased streamgage screening applied in the current study most likely resulted in a more accurate representation of natural peak-flow conditions.

The updated, regional peak-flow regression equations will be integrated in the U.S. Geological Survey StreamStats program to allow users to estimate basin and climatic characteristics and peak-flow statistics at ungaged locations of interest. StreamStats estimates peak-flow statistics with quantifiable certainty only when used at sites with basin and climatic characteristics within the range of input variables used to develop the regional regression equations. Both the regional regression equations and StreamStats should be used to estimate peak-flow statistics only in naturally flowing, relatively unregulated streams without substantial local influences to flow, such as large seeps, springs, or other groundwater-surface water interactions that are not widespread or characteristic of the respective region.

## Introduction

Reliable estimates of the magnitude and frequency of floods are needed by Federal, regional, State, and local infrastructure designers and water-resource managers for the design of highway, road, and other bridge crossings of rivers, delineation of flood plains, flood inundation mapping, design of water-control structures and culverts, and management of water supplies. Additionally, there is an increasing interest in peak flows with high annual exceedance probabilities (AEPs) and more frequent recurrence intervals, such as those with 80–40 percent AEPs (1.25- to 2.50-year recurrence intervals), because of their role in maintaining aquatic and riparian ecosystems. Peak-flow frequency statistics (hereinafter peak-flow statistics) typically are estimated for streamgages that have at least 10 years of annual peak-flow record, historically using statistical methods described in Bulletin 17B of the Interagency Advisory Committee on Water Data (1982). Peak-flow statistics can then be estimated for ungaged sites through the development of regional regression equations between basin and climatic characteristics and peak-flow statistics at streamgages. These peak-flow statistics and regional regression equations typically are made available to water-resource managers and the public through the U.S. Geological Survey (USGS) StreamStats program (<http://streamstats.usgs.gov>). USGS StreamStats is a web-based Geographic Information System tool that is useful for water-resources planning and management and infrastructure design. StreamStats estimates peak-flow statistics based on underlying basin and climatic characteristic datasets and regional regression equations, and allows users to obtain peak-flow statistics and other information for gaged and ungaged sites on streams. The underlying data layers, flow statistics, and regression equations used in StreamStats must be routinely re-evaluated and revised to ensure accuracy of the information provided to the public.

Hydrologists in the USGS Idaho Water Science Center last estimated peak-flow statistics for USGS streamgages on naturally flowing (relatively unregulated) streams in Idaho and developed regional regression equations to estimate peak-flow statistics at ungaged sites in 2002. The statistics were generated using flow data only through the mid-1990s. Results were published in Berenbrock (2002) and Hortness and Berenbrock (2003). The USGS recognized the need to update peak-flow statistics and regional regression equations in Idaho because of: (1) the occurrence of several flood events, (2) the availability of several new streamgages with at least 10 years of record, and (3) the development of new statistical techniques after the Berenbrock (2002) and Hortness and Berenbrock (2003) studies were completed. The USGS recently (2013) released new internal guidance (Tim Cohn, Julie Kiang, and Andrea Veilleux, U.S. Geological Survey, written commun., 2013) and improved statistical tools for calculating peak-flow statistics and regional regression equations. In the past, USGS personnel primarily followed

procedures outlined in Bulletin 17B of the Interagency Advisory Committee on Water Data (1982) when estimating peak-flow statistics, but the use of historical flood information was limited. A new technique called the Expected Moments Algorithm (EMA) allows historical information to be represented as intervals, which has been shown to increase the accuracy of peak-flow statistics, as discussed in Cohn and others (2001), England and others (2003), and England and Cohn (2008). Additionally, the accuracy of peak-flow statistics can be increased when low, highly variable peak flows, called low outliers, are removed from the analysis. The generalized Multiple Grubbs Beck Test (MGBT), described in Cohn and others (2013), was developed as a way to objectively and systematically detect and remove low outliers, and can be used in concert with EMA techniques. Additionally, the generalized regional skew map of logarithms (base 10; hereinafter “log”) of annual peak flow developed by Hardison (1974) and included in Bulletin 17B of the Interagency Advisory Committee on Water Data (1982) was based on a limited dataset and did not account for low outliers or historical information. The Bayesian-Generalized Least Squares method, described in Veilleux (2011) and Veilleux and others (2012), was developed to estimate regional skew with greater precision than the skew values in Bulletin 17B of the Interagency Advisory Committee on Water Data (1982). Numerous U.S. government agencies have advocated a revision to the guidelines for flood frequency analysis contained in Bulletin 17B of the Interagency Advisory Committee on Water Data (1982) to include these methods (Stedinger and Griffis, 2008).

The USGS, in cooperation with the Idaho Transportation Department, updated peak-flow statistics for selected streamgages and developed revised, regional regression equations to estimate peak-flow statistics in naturally flowing streams and rivers in Idaho. The aforementioned EMA and MGBT procedures, as well as the new regional skew coefficients, were used in the analysis, which included peak-flow data through water year 2013 (October 1, 2012, through September 30, 2013).

## Purpose and Scope

This report documents peak-flow statistics for selected streamgages and development of regional regression equations that can be used to estimate peak-flow statistics at ungaged sites in naturally flowing Idaho streams and rivers. The regional regression equations were developed using a generalized least squares (GLS) analysis between peak-flow statistics and basin and climatic characteristics for selected streamgages. This report updates or refines the regional peak-flow regression equations published in Berenbrock (2002) and Hortness and Berenbrock (2003) by including more recent flow data through water year 2013 and by implementing advances in statistical techniques developed after those reports were published.

Values of peak-flow statistics for selected streamgages as well as regional regression equations to estimate peak-flow statistics at un-gaged sites are presented for 11 statistics with AEPs ranging from 80 to 0.2 percent (1.25- to 500-year recurrence intervals; table 1). Streamgages were selected for inclusion in the development of the regional peak-flow regression equations if they had at least 10 years of record of annual peak flows, through at least 1980 if the record started before 1980, and were on streams considered naturally flowing. “Naturally flowing” is a relative term in this analysis because streams and rivers with flows completely unaffected by flow diversions are rare in Idaho. As a result, “naturally flowing” was defined as a stream not affected by direct regulation (dams, major diversions) that controlled flow from more than 20 percent of the drainage basin for a particular streamgage. Screening of streamgages for use in this analysis resulted in 192 streamgages that met the criteria. The regional regression equations presented in this report should be used to estimate peak-flow statistics only in “naturally flowing” streams and rivers in Idaho.

## Description of Study Area

The study area was generally located in the State of Idaho including parts of the bordering States and Canadian province, British Columbia (fig. 1). The study area included parts of two physiographic divisions: the Rocky Mountain System and the Intermontane Plateaus (Fenneman, 1946; fig. 1). Ecological divisions and sections in the Rocky Mountain System include the Middle Rocky Mountain Steppe, extending through central and east-central Idaho and Montana; the Northern Rocky Mountain Forest-Steppe in northern Idaho along the Washington-British Columbia-Montana border; and the Southern Rocky Mountain Steppe in eastern Idaho along the Wyoming border (McNab and others, 2007; fig. 1). The terrain throughout the Rocky Mountain System generally includes steep, rugged, medium-to-high elevation mountains. The Intermontane Plateaus physiographic division primarily includes a single intermountain semidesert ecological division with highly irregular terrain, ranging from high elevation mountains, to expansive basalt flats, to deep vertical-walled canyons, to rolling prairie. Elevation extremes in the study area range from about 300 ft at the confluence of the Snake River with the Columbia River near Kennewick, Washington (fig. 1) to greater than 13,000 ft in the Snake River headwaters in western Wyoming. The elevation is highest in the southeastern part of the study area and generally decreases along the path of the Snake River. About 61 percent of the terrain throughout the study area contains moderate- to high-elevation mountains with steep slopes. Plains, plateaus, and basin and range topography consist of about 35 percent of the terrain; the remaining 4 percent of the terrain is low-relief hills (McNab and others, 2007).

**Table 1.** Annual exceedance probabilities with corresponding recurrence intervals used in computing peak-flow statistics and regional regression equations.

[Abbreviations:  $Q_{AEP}$ , peak-flow magnitude, in cubic feet per second, for the indicated annual exceedance probability (AEP)]

Annual exceedance probability (percent)	$Q_{AEP}$ abbreviation	Recurrence interval (years)
80	$Q_{80}$	1.25
67	$Q_{67}$	1.50
50	$Q_{50}$	2.00
43	$Q_{43}$	2.33
20	$Q_{20}$	5.00
10	$Q_{10}$	10.0
4	$Q_4$	25.0
2	$Q_2$	50.0
1	$Q_1$	100
0.5	$Q_{0.5}$	200
0.2	$Q_{0.2}$	500

Geology and soils in the Rocky Mountain System are highly variable ranging from volcanic rock in the Southern Rocky Mountain Steppe; weathered granites (or granitic rocks) and sedimentary rocks/deposits in the Middle Rocky Mountain Steppe; glacial till in the Northern Rocky Mountain Forest-Steppe; and glacial outwash silt on the Palouse Dry Steppe. Highly permeable soils (predominantly sand or gravel) in the Middle and Northern Rocky Mountain Steppes allow for high infiltration rates and minimal runoff. Soils in the Intermontane Plateaus are predominantly derived from volcanic rock, namely basalt and clay (McNab and others, 2007). Soils in the vicinity of the Owyhee River in the Intermontane Semidesert are largely clay and therefore have high runoff potential (McNab and others, 2007; Natural Resources Conservation Service, 2007).

Climate variability tends to align with the terrain in that the high elevations are generally cool with high relative precipitation, and low elevations are generally warm with low relative precipitation. Mean annual temperature generally aligns with the elevation and physiographic divisions, where the Rocky Mountain System is typically cooler than the Intermontane Plateau, with a few exceptions. Annual peak-flow patterns typically correlate with the magnitude of snowpack and the timing of the snowmelt runoff. Peak-flow runoff generally occurs in low-elevation areas between February and May, and as late as July in high-elevation areas. Annual peak flows generally occur during the snowmelt runoff; however, peak annual flows resulting from localized high-intensity and (or) -duration rainfall events are not uncommon.

4 Estimating Peak-Flow Frequency Statistics for Selected Gaged and Ungaged Sites in Naturally Flowing Streams and Rivers in Idaho

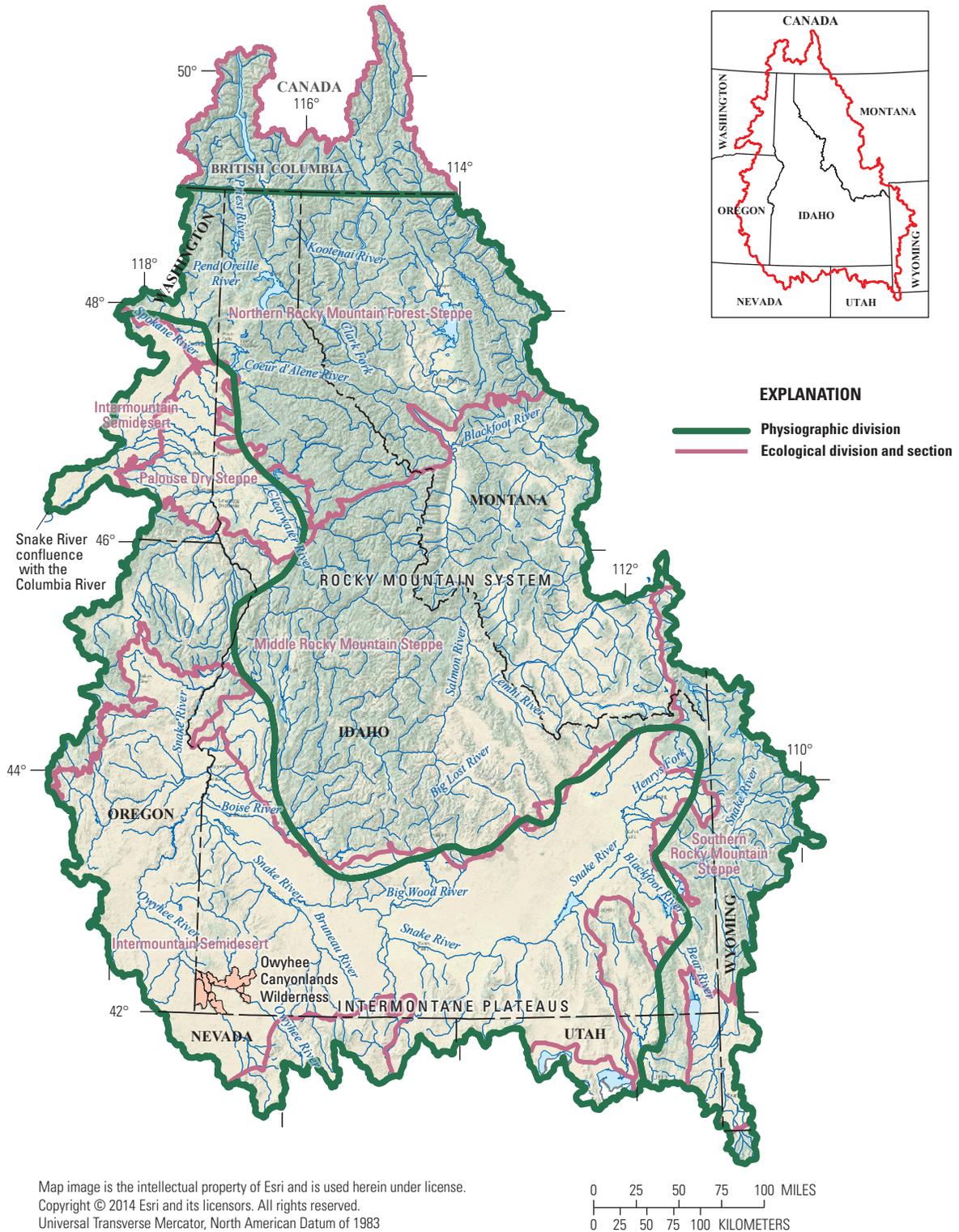


Figure 1. Study area showing physiographic and ecological divisions in Idaho and bordering States.

Extending from the headwaters in high-elevation mountains in Wyoming to the confluence with the Columbia River in Washington State, the Snake River is the largest river basin in the study area containing about 61 percent of the total study area (56,900 mi<sup>2</sup>). Major tributaries to the Snake River include the Salmon and Clearwater Rivers. Located in the north-central part of the study area and originating in Montana, the Pend Oreille River Basin accounts for about 16 percent of the study area (26,430 mi<sup>2</sup>). The Kootenai River Basin, located in northern Idaho, Montana, and British Columbia accounts for about 7 percent of the study area (30,650 mi<sup>2</sup>). The Spokane River Basin contains about 4 percent of the study area (8,000 mi<sup>2</sup>). Although collectively smaller than one-half the size of the Snake River, the Pend Oreille, Kootenai, and Spokane Rivers collectively contribute more total flow to the Columbia River than the entire Snake River.

River regulation is common throughout much of the study area and affects the magnitude and timing of flow runoff to all downstream reaches. Regulation ranges from dams (constructed for irrigation, power generation, and flood risk management) to irrigation diversions including canals, ditches, and aqueducts. All major river basins previously listed are affected to a certain extent by regulation, even the Salmon River, which is widely considered unregulated but, in fact, contains numerous minor diversions. Streamgages and annual peak flows substantially affected by regulation were not included in this study because of the site-specific alteration of flows which invalidates regional efforts to estimate peak-flow statistics in naturally flowing streams.

## Previous Studies

Prior to the study described in this report, the most recent studies to estimate peak-flow statistics at ungaged sites in Idaho were Berenbrock (2002) for the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent AEPs and Hortness and Berenbrock (2003) for the 67- and 43-percent AEPs. Hortness and Berenbrock (2003) focused on peak-flow statistics that were thought to describe “bankfull” conditions, or flows associated with the mobilization of streambed sediments to maintain long-term aquatic habitat and channel capacity. Various studies described in Hortness and Berenbrock (2003), including Emmett (1975), Leopold (1994), and Castro and Jackson (2001), have defined bankfull flows as those with AEPs in the range of 80–40 percent. A number of other studies have been completed to estimate flood-frequency characteristics for Idaho streams and are fully described in Berenbrock (2002), including Thomas and others (1973), Harenberg (1980), Kjelstrom and Moffatt (1981), Hedman and Osterkamp (1982), Quillian and Harenberg (1982), and Thomas and others (1994). The most recent, comprehensive, and comparable studies to the current study described in this report are Quillian and Harenberg (1982), Berenbrock (2002), and Hortness and Berenbrock (2003).

Quillian and Harenberg (1982) completed a network and cost-effectiveness analysis of the streamgage network in Idaho. They completed a regional regression analysis by dividing the State into nine regions and relating peak (50-, 10-, 2-, and 1-percent AEPs) and mean annual flow statistics to basin and climatic characteristics. Streamgages were included in their analysis if they had 5 or more years of flow record that was minimally affected by regulation or diversion. Overall, 315 streamgages were used by Quillian and Harenberg (1982) to develop the peak-flow regression equations. Drainage area or both drainage area and mean annual precipitation, depending on the region and statistic, were selected for use in the equations. Berenbrock (2002) noted that the standard errors of prediction (SEPs) of the regression equations generated by Quillian and Harenberg (1982) were larger than previous regional peak-flow regression studies in Idaho.

Berenbrock (2002) completed an updated regional peak-flow study from Quillian and Harenberg (1982) and presented various methods for estimating peak-flow statistics at ungaged sites in naturally flowing rivers, including the Region of Influence (ROI) method and regional regression equations. Procedures followed by Hortness and Berenbrock (2003) were identical to Berenbrock (2002), except that the ROI method was not evaluated and different peak-flow statistics were estimated. Annual peak-flow records for 333 streamgages were used in the analyses. Peak-flow statistics were estimated for the streamgages using Bulletin 17B of Interagency Advisory Committee on Water Data (1982) methods, and regional regression equations were developed in nine regions (identified 1, 2, 3, 4, 5, 6, 7a, 7b, and 8) using GLS regression techniques between the peak-flow statistics and basin and climatic characteristics. The ROI method, described in Tasker and Slade (1994), was tested by Berenbrock (2002) as an alternative to the regional regression method. The ROI method also was based on developing regressions between peak-flow statistics at streamgages and basin and climatic characteristics, but the regressions were estimated using only streamgages within a given distance of the ungaged location of interest. Berenbrock (2002) noted that the ROI method did not produce more accurate results and required much more complex calculations than the regional regression method.

Berenbrock (2002) and Hortness and Berenbrock (2003) selected streamgages for the analysis that had at least 10 years of record through 1997. Streamgages were screened out if their annual peak flows were considered substantially affected by regulation or diversion. No additional streamgage screening evaluations were described, such as whether streamgages were removed because they were too close together in the same drainage basin (called nested or redundant basin analysis) or because unusual trends were present in annual peak-flow records. Basin and climatic characteristics selected for the final regression equations varied by region and statistic but included drainage area, mean basin elevation, percentage of forest cover, mean annual precipitation, percentage of north-facing slopes greater than 30 percent, percentage of slopes greater than 30 percent, and average basin slope.

Recently, regional peak-flow regression studies completed by the USGS have migrated away from strict use of Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) methods and toward use of revised, more accurate methods, including the EMA and MGBT procedures, previously described and discussed in more detail in section, “Estimating Peak-Flow Frequency Statistics at Selected Gaged Sites.” USGS studies using the revised procedures to estimate peak-flow statistics have been completed in Arizona (Paretti and others, 2014), California (Parrett and others, 2011; Gotvald and others, 2012), and Iowa (Eash and others, 2013) and are underway in Washington (Mark Mastin, U.S. Geological Survey, oral commun., 2016), among other States.

## Data Screening and Compilation

The preparation of the datasets used to develop the regional regression equations to estimate peak-flow statistics at ungaged sites required several steps to screen out streamgages not appropriate for use. After finalizing the list of streamgages to be used in the analysis and reviewing their records of annual peak flows, basin and climatic characteristics for each streamgage were retrieved and compiled from various sources.

### Streamgage Selection

The selection of streamgages for use in the regional peak-flow regression analysis was completed in seven steps (fig. 2): (1) starting with Idaho streamgages in the Geospatial Attributes of Gages for Evaluating Streamflow, version II, database (GAGES II; Falcone, 2011); (2) adding streamgages used by Berenbrock (2002) and Hortness and Berenbrock (2003); (3) adding streamgages in a 50-mi buffer in adjacent States; (4) removing streamgages with less than 10 years of record or with records starting before but not extending through at least 1980; (5) removing streamgages with greater than 20 percent of their basin affected by regulation or diversion; (6) evaluating streamgages that provided redundant information on a stream or were nested close together in a basin to determine which should be removed, and (7) removing streamgages with a statistically significant ( $p < 0.05$ ) time trend inconsistent with climate patterns (deemed anthropogenic influence).

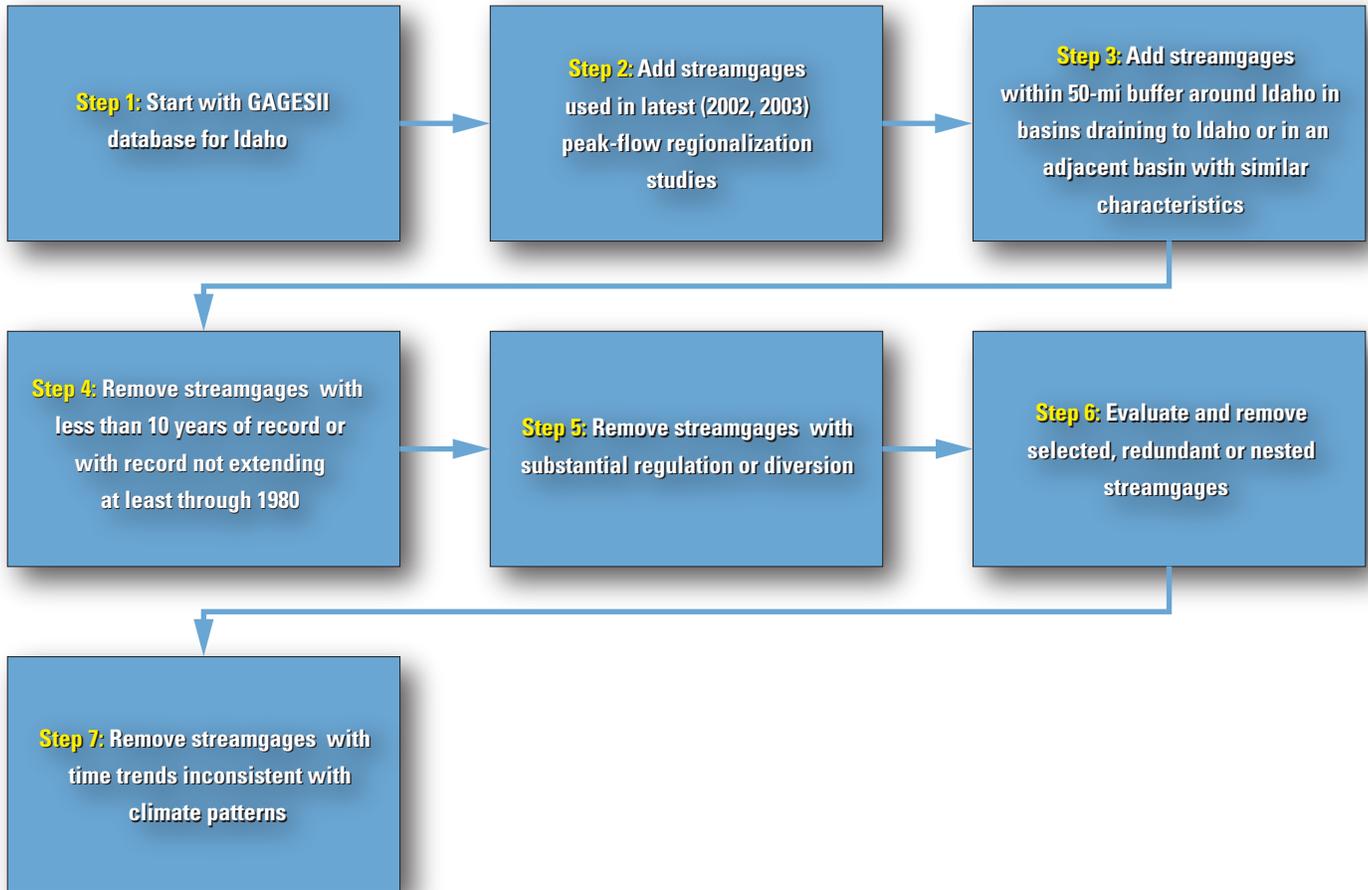
The streamgage selection process started (fig. 2, step 1) with the GAGES II database, which provided geospatial data and classifications for 9,322 streamgages maintained by the USGS (Falcone, 2011). Only streamgages in the United States and operated by the USGS were used in the analysis for consistency and quality assurance. The GAGES II database consisted of streamgages which had either at least 20 complete but not necessarily continuous years of streamflow record from 1950, or were actively operated as of water year 2009,

and whose watersheds lay in the United States. The geospatial data included several hundred basin characteristics compiled from national data sources, including environmental features and metrics of anthropogenic influence. The database also included comments from local USGS Water Science Centers, based on Annual Data Reports, pertinent to hydrologic modifications and influences.

The GAGES II dataset was then compared to the list of streamgages used in the Berenbrock (2002) and Hortness and Berenbrock (2003) studies. Streamgages used in those studies but excluded from the GAGES II database for record lengths between 10 and 20 years or for other, unknown reasons were added to the list of streamgages considered for use in this study (fig. 2, step 2). The USGS National Water Information System (NWIS) Mapper (<http://maps.waterdata.usgs.gov/mapper/index.html?state=id>) and other GIS tools were then used to identify streamgages in adjacent States with at least 10 years of annual peak flow record (fig. 2, step 3). Such streamgages were included if they were within about 50 mi of Idaho and were in a basin draining to (or from) Idaho or in a basin with very similar basin and climatic characteristics as nearby basins in Idaho. Steps 1–3 resulted in a list of 1,215 streamgages evaluated for use.

Streamgages in the list were then checked for record length (number of annual peaks) and end date (last year of record). Streamgages with fewer than 10 years of record or with records not extending through at least 1980 (fig. 2, step 4) were excluded. Clark (2010) evaluated changes in patterns and timing of flow in unregulated basins in Idaho and parts of Wyoming and Nevada and noted statistically significant ( $p < 0.05$ ) decreasing trends in flow statistics at several streamgages with long-term record. The largest downward trends have occurred from dry climatic conditions since the early 1970s, which followed a wet period with above-average flows in the 1950s through early 1970s. Thus, streamgages with records ending before 1980 were excluded to maintain consistency relative to this regional climatic trend and ensure relatively recent climatic conditions were represented in the analysis. If a streamgage meeting other screening criteria had records extending after 1980, the full period of record was used in the analysis. Streamgages were initially removed from the analysis if their record did not extend through at least 1984, which was 30 years prior to the last year of record used in this analysis (2013). However, 39 streamgages operated on unregulated streams in Idaho were discontinued in 1980–82, so the threshold was changed to 1980 to allow the consideration of these streamgages.

The list of streamgages was further screened to exclude streamgages substantially affected by regulation or diversion (fig. 2, step 5). All streamgages with annual peak flows with codes of “6,” meaning, “known or planned regulation” (Ryberg, 2008, p. 5) were removed. In some cases, the regulation started after at least 10 years of unregulated flow record was collected. If a particular streamgage had 10 years



**Figure 2.** Streamgauge selection and screening process for the regional peak-flow regression analysis in Idaho.

of unregulated flow record extending through at least 1980, the pre-regulation part of the record was used in the analysis. Additional streamgages were excluded if site descriptions indicated more than 20 percent of the total drainage area was affected by regulation or diversion or if other evidence (knowledge of the site, anecdotal information) indicated that annual peak flows were not representative of natural conditions. The streamgages located beyond Idaho borders and remaining after steps 1–5 were reviewed by USGS Hydrologists in those States to ensure that streamgages were representative of naturally flowing conditions and appropriate for a regional analysis.

The next screening step involved evaluating streamgages for redundancy. An assumption in the regional peak-flow regression analysis was that each streamgauge provided an independent spatial observation depicting how basin and climatic characteristics were related to flow. Streamgages located close together in the same basin had the potential to provide redundant, spatially-dependent information, which violates the assumption. When streamgages in basins (streamgauge-pairs) are redundant, a statistical analysis using

both streamgages incorrectly represents the information in the regional dataset (Gruber and Stedinger, 2008). To determine if two streamgages are redundant and thus represent the same hydrologic conditions, two types of information are considered: (1) whether their basins are nested, and (2) whether they are located close together on the same stream reach.

The standardized distance (SD), is used to determine the likelihood that the basins are nested. The standardized distance between two basin centroids, SD is defined as:

$$SD_{ij} = \frac{D_{ij}}{\sqrt{0.5(DRNAREA_i + DRNAREA_j)}} \quad (1)$$

where

$D_{ij}$  is the distance between centroids of basin  $i$  and basin  $j$ ; and  
 $DRNAREA_i$  is the drainage area at streamgauge  $i$ ; and  
 $DRNAREA_j$  is the drainage area at streamgauge  $j$ .

The drainage area ratio (DAR) is used to determine if two streamgages are located close together on the same stream reach or if the streamgages are in nested basins are sufficiently similar in size to conclude that they are, or are at least in large part, the same watershed for the purposes of developing a regional hydrologic model. The DAR is defined as (Veilleux, 2011):

$$\text{DAR} = \text{Max} \left[ \frac{\text{DRNAREA}_i}{\text{DRNAREA}_j}, \frac{\text{DRNAREA}_j}{\text{DRNAREA}_i} \right] \quad (2)$$

where

DAR is the Max (maximum) of the two values in brackets;

$\text{DRNAREA}_i$  is the drainage area at streamgage  $i$ ; and

$\text{DRNAREA}_j$  is the drainage area at streamgage  $j$ .

The list of streamgages was input into a script written in the R programming language (R Core Team, 2015) that identified streamgage pairs with SD less than or equal to 0.50 and DAR less than or equal to 5, which were identified in Veilleux (2011) as desirable thresholds. These streamgages were considered to be redundant or nested and were evaluated to determine which should be retained (fig. 2, step 6). The factors considered when deciding which streamgages to remove from the analysis included: (1) the length of records, (2) recentness of records, and (3) whether two streamgages on tributaries draining to a river could be used instead of just one streamgage on the river main stem.

The final screening step (fig. 2, step 7) was to evaluate the time trends in the annual peak-flow records, using Kendall's  $\tau$ , described in Helsel and Hirsch (2002). Streamgages with statistically significant ( $p < 0.05$ ) Kendall's  $\tau$  values were investigated to determine possible causes. Streamgages were excluded if the trends (1) were inconsistent with climate trends or with trends in data from other streamgages in the same or nearby basins or (2) could be attributed to an anthropogenic change in a basin, such as increased shallow groundwater withdrawals for an industrial process. Three streamgages were excluded because of significant Kendall's  $\tau$  and strong evidence of anthropogenic influence.

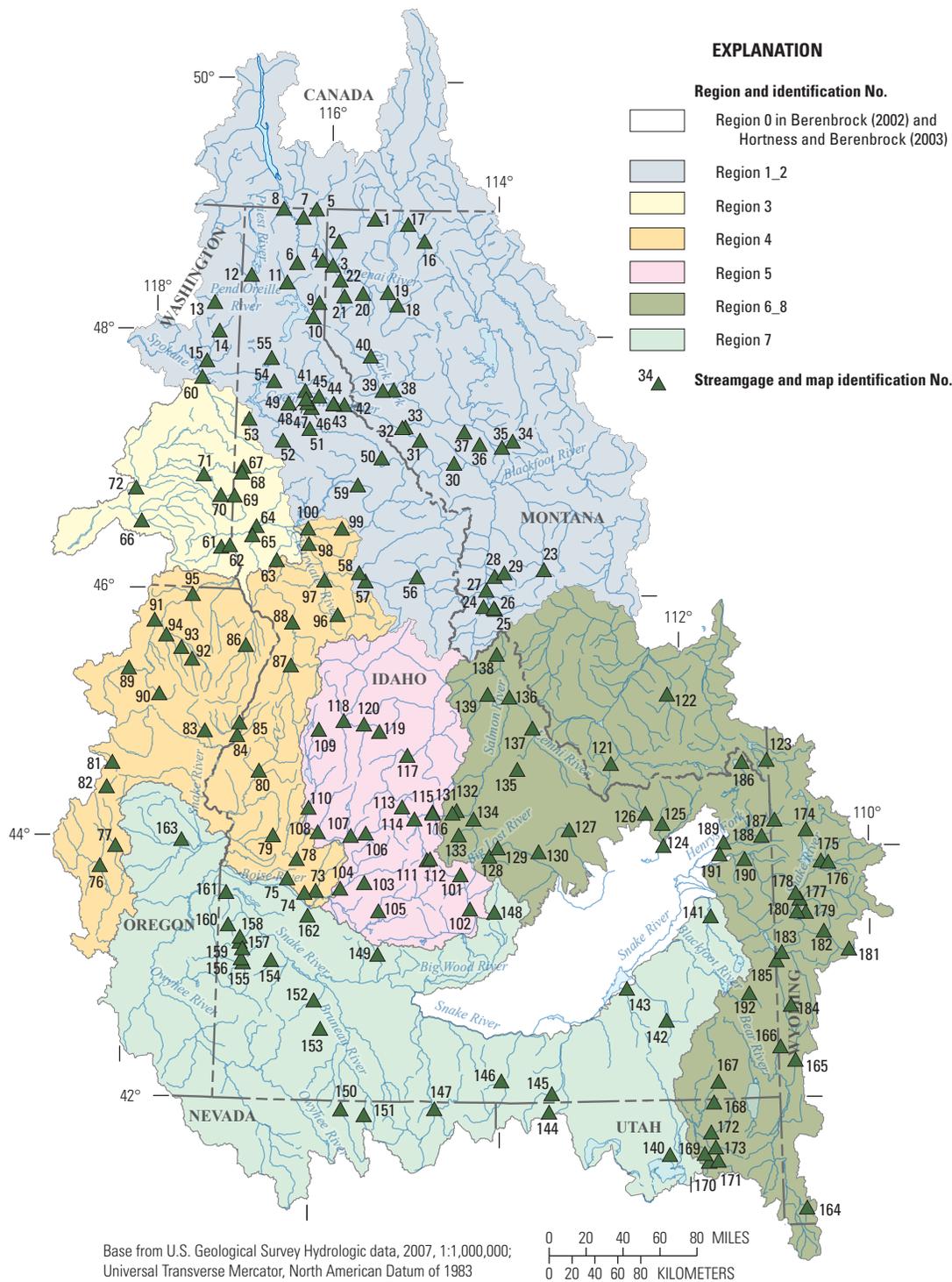
The screening process (fig. 2) resulted in a final list of 192 streamgages, with a combined total of 8,779 years of peak-flow record, for use in the regional peak-flow regression analysis. The full list of streamgages, along with information on their peak-flow records, is provided in appendix A (table A1). The selected streamgages were fairly well distributed throughout Idaho and bordering States (fig. 3), except for southern Idaho, where streamgages located on relatively unregulated, naturally flowing streams were sparse.

## Basin and Climatic Characteristics

Basin and climatic characteristics were estimated from various datasets for the 192 streamgages selected for the regional peak-flow regression analysis. For streamgages in Idaho, Oregon, Utah, and Washington, drainage basins were delineated using the USGS StreamStats program (U.S. Geological Survey, 2012). For streamgages in Montana, Nevada, and Wyoming, where the StreamStats application was not yet fully implemented, drainage basins were delineated using the Hydrology Toolset in ArcGIS (Environmental Systems Resource Institute, 2013) using 10-m digital elevation models (DEMs) from the 1/3-arc-second resolution National Elevation Dataset (Gesch and others, 2002; U.S. Geological Survey, 2014). The methodology in the ArcGIS Hydrology Toolset was nearly identical to the methodology in the StreamStats online application. The drainage basin boundaries were then manually reviewed to check for errors and corrected if required.

The StreamStats/ArcGIS-derived drainage areas (table A2) were compared with the areas reported for each streamgage in the USGS National Water Information System (NWIS; U.S. Geological Survey, 2015) (fig. 4). Some of the NWIS-reported drainage areas were historically derived by manually planimetrying basin boundaries on 1:24,000-scale maps (USGS 7.5-minute quadrangle maps) and were subject to human error. Other drainage areas reported in NWIS likely have been re-measured using GIS-based methods and updated through the years to meet specific needs, although no comprehensive updates have occurred. For the current study, any discrepancies between the StreamStats/ArcGIS- and NWIS-reported drainage area values were further investigated. In nearly all cases, the drainage area determined using StreamStats/ArcGIS was noted to be more accurate and reasonable than the value reported in NWIS, was used in the regional peak-flow regression analysis, and will be updated in NWIS. The correct drainage area was difficult to determine for a few streamgages. For example, the drainage basin for streamgage 12427000, Little Spokane River near Elk, Washington, consists of hummocky terrain, making it very difficult to define surface-water flow direction and the drainage basin boundary. In these cases, the StreamStats/ArcGIS-derived drainage areas were used to be consistent with other streamgages.

Basin and climatic characteristic selection focused initially on updating datasets for the characteristics used in the previous studies by Berenbrock (2002) and Hortness and Berenbrock (2003), including higher resolution DEMs (10-m cell size instead of a previously used 30-m cell size) and their derived datasets such as basin slope characteristics (table 2).



**Figure 3.** Locations of streamgages in Idaho and bordering States selected for the regional peak-flow regression analysis (area in southeastern Idaho shown in white was excluded because of extent of regulation, groundwater-surface water interactions, and infiltration).

**Table 2.** Basin and climatic characteristics estimated for selected streamgages and evaluated in regional regression equations.

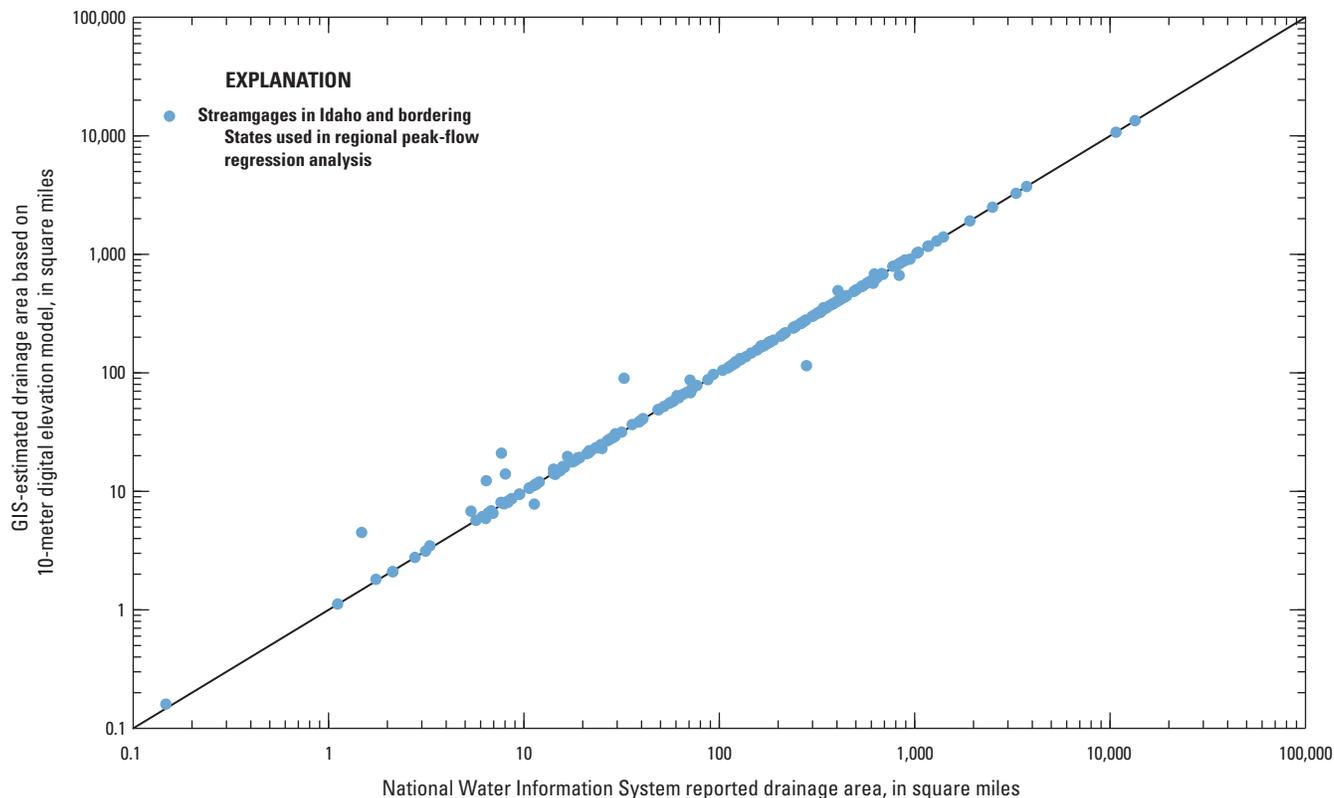
[Characteristics in bold (in addition to hydrologic unit code, not presented here) were evaluated for use in the regional skew regression analysis. StreamStats abbreviation: From [http://streamstatsags.er.usgs.gov/ss\\_defs/basin\\_char\\_defs.aspx](http://streamstatsags.er.usgs.gov/ss_defs/basin_char_defs.aspx). Abbreviations: USGS, U.S. Geological Survey; DEM, digital elevation model; NLCD, National Land Cover Database; PRISM, Parameter-elevation Regressions on Independent Slopes Model; NOAA, National Oceanic and Atmospheric Administration; NRCS, Natural Resources Conservation Service]

Basin and climatic characteristic	StreamStats abbreviation	Units	Datasource
Perimeter of the basin	BASINPERIM	Miles	Gesch and others, 2002, USGS National Elevation Dataset, 10-meter DEM, <a href="http://nationalmap.gov/elevation.html">http://nationalmap.gov/elevation.html</a>
<b>Drainage area of the basin</b>	<b>DRNAREA</b>	Square miles	Gesch and others, 2002, USGS National Elevation Dataset, 10-meter DEM, <a href="http://nationalmap.gov/elevation.html">http://nationalmap.gov/elevation.html</a>
Percentage of basin area covered by forest	FOREST	Percent	Homer and others, 2015, National Land Cover Database 2011, NLCD 2011, <a href="http://www.mrlc.gov">http://www.mrlc.gov</a>
<b>Mean annual precipitation</b>	<b>PRECIP</b>	Inches	PRISM Climate Group, 2007, PRISM climate datasets, 800-meter PRISM, <a href="http://www.prism.oregonstate.edu/">http://www.prism.oregonstate.edu/</a> and Natural Resources Canada, 2014, <a href="http://cfs.nrcan.gc.ca/projects/3?lang=en_CA">http://cfs.nrcan.gc.ca/projects/3?lang=en_CA</a>
Mean basin slope computed from 10m DEM	BSLDEM10M	Percent	Gesch and others, 2002, USGS National Elevation Dataset, 10-meter DEM, <a href="http://nationalmap.gov/elevation.html">http://nationalmap.gov/elevation.html</a>
Percentage of basin with slopes greater than 30 percent	SLOP30	Percent	Gesch and others, 2002, USGS National Elevation Dataset, 10-meter DEM, <a href="http://nationalmap.gov/elevation.html">http://nationalmap.gov/elevation.html</a>
Percentage of basin with north-facing slopes greater than 30 percent	NFSL30	Percent	Gesch and others, 2002, USGS National Elevation Dataset, 10-meter DEM, <a href="http://nationalmap.gov/elevation.html">http://nationalmap.gov/elevation.html</a>
<b>Minimum basin elevation</b>	<b>ELEVMIN</b>	Feet	Gesch and others, 2002, USGS National Elevation Dataset, 10-meter DEM, <a href="http://nationalmap.gov/elevation.html">http://nationalmap.gov/elevation.html</a>
<b>Maximum basin elevation</b>	<b>ELEVMAX</b>	Feet	Gesch and others, 2002, USGS National Elevation Dataset, 10-meter DEM, <a href="http://nationalmap.gov/elevation.html">http://nationalmap.gov/elevation.html</a>
<b>Mean basin elevation</b>	<b>ELEV</b>	Feet	Gesch and others, 2002, USGS National Elevation Dataset, 10-meter DEM, <a href="http://nationalmap.gov/elevation.html">http://nationalmap.gov/elevation.html</a>
Percentage of basin area above an elevation of 6,000 feet	EL6000	Percent	Gesch and others, 2002, USGS National Elevation Dataset, 10-meter DEM, <a href="http://nationalmap.gov/elevation.html">http://nationalmap.gov/elevation.html</a>
<b>The difference between the maximum and minimum basin elevations</b>	<b>RELIEF</b>	Feet	Gesch and others, 2002, USGS National Elevation Dataset, 10-meter DEM, <a href="http://nationalmap.gov/elevation.html">http://nationalmap.gov/elevation.html</a>
Mean maximum January temperature	JANMAXTMP	Degrees Celsius	PRISM Climate Group, 2007, PRISM climate datasets, 800-meter PRISM, <a href="http://www.prism.oregonstate.edu/">http://www.prism.oregonstate.edu/</a> and Natural Resources Canada, 2014, <a href="http://cfs.nrcan.gc.ca/projects/3?lang=en_CA">http://cfs.nrcan.gc.ca/projects/3?lang=en_CA</a>
<b>Mean minimum January temperature</b>	<b>JANMINTMP</b>	Degrees Celsius	PRISM Climate Group, 2007, PRISM climate datasets, 800-meter PRISM, <a href="http://www.prism.oregonstate.edu/">http://www.prism.oregonstate.edu/</a> and Natural Resources Canada, 2014, <a href="http://cfs.nrcan.gc.ca/projects/3?lang=en_CA">http://cfs.nrcan.gc.ca/projects/3?lang=en_CA</a>

**Table 2.** Basin and climatic characteristics estimated for selected streamgages and evaluated in regional regression equations.—Continued

Basin and climatic characteristic	StreamStats abbreviation	Units	Datasource
Mean 6-hour precipitation that occurs on average once in 2 years	I6H2Y	Inches	National Oceanic and Atmospheric Administration, 2007, National Oceanic and Atmospheric Administration rainfall intensity and climate dataset, 400-meter grid, <a href="http://www.nws.noaa.gov/oh/hdsc/">http://www.nws.noaa.gov/oh/hdsc/</a>
Mean 24-hour precipitation that occurs on average once in 2 years —Equivalent to precipitation intensity index	I24H2Y	Inches	National Oceanic and Atmospheric Administration, 2007, National Oceanic and Atmospheric Administration rainfall intensity and climate dataset, 400-meter grid, <a href="http://www.nws.noaa.gov/oh/hdsc/">http://www.nws.noaa.gov/oh/hdsc/</a>
<b>Mean 6-hour precipitation that occurs on average once in 100 years</b>	<b>I6H100Y</b>	Inches	National Oceanic and Atmospheric Administration, 2007, National Oceanic and Atmospheric Administration rainfall intensity and climate dataset, 400-meter grid, <a href="http://www.nws.noaa.gov/oh/hdsc/">http://www.nws.noaa.gov/oh/hdsc/</a>
<b>Mean 24-hour precipitation that occurs on average once in 100 years</b>	<b>I24H100Y</b>	Inches	National Oceanic and Atmospheric Administration, 2007, National Oceanic and Atmospheric Administration rainfall intensity and climate dataset, 400-meter grid, <a href="http://www.nws.noaa.gov/oh/hdsc/">http://www.nws.noaa.gov/oh/hdsc/</a>
Mean January–February–March precipitation	MJFM_P <sup>1</sup>	Inches	PRISM Climate Group, 2007, PRISM climate datasets, 800-meter PRISM, <a href="http://www.prism.oregonstate.edu/">http://www.prism.oregonstate.edu/</a> and Natural Resources Canada, 2014, <a href="http://cfs.nrcan.gc.ca/projects/3?lang=en_CA">http://cfs.nrcan.gc.ca/projects/3?lang=en_CA</a>
Mean February–March–April precipitation	MFMA_P <sup>1</sup>	Inches	PRISM Climate Group, 2007, PRISM climate datasets, 800-meter PRISM, <a href="http://www.prism.oregonstate.edu/">http://www.prism.oregonstate.edu/</a> and Natural Resources Canada, 2014, <a href="http://cfs.nrcan.gc.ca/projects/3?lang=en_CA">http://cfs.nrcan.gc.ca/projects/3?lang=en_CA</a>
Watershed compactness ratio, a measure of basin shape related to basin perimeter and drainage area	COMPRAT	Unitless	Gesch and others, 2002, USGS National Elevation Dataset, 10-meter DEM, <a href="http://nationalmap.gov/elevation.html">http://nationalmap.gov/elevation.html</a>
Mean annual air temperature	TEMP	Degrees Celsius	PRISM Climate Group, 2007, PRISM climate datasets, 800-meter PRISM, <a href="http://www.prism.oregonstate.edu/">http://www.prism.oregonstate.edu/</a> and Natural Resources Canada, 2014, <a href="http://cfs.nrcan.gc.ca/projects/3?lang=en_CA">http://cfs.nrcan.gc.ca/projects/3?lang=en_CA</a>
Percentage of snow in the total precipitation estimate	PCT_SNOWPRCP <sup>1</sup>	Percent	U.S. Department of Commerce, 1999, NOAA National Climatic Data Center, <a href="http://www.ncdc.noaa.gov/">http://www.ncdc.noaa.gov/</a> , and Marshall and others, 1999, Canada National Ecological Framework, <a href="http://sis.agr.gc.ca/cansis/nsdb/eostrat/1999report/data_tables.html">http://sis.agr.gc.ca/cansis/nsdb/eostrat/1999report/data_tables.html</a>
Average soil permeability	SOILPERM	Inches/hour	U.S. Department of Agriculture, 2014, NRCS soil statistics dataset, 100-meter NRCS USDA STATSGO soil data, <a href="http://websoilsurvey.nrcs.usda.gov/">http://websoilsurvey.nrcs.usda.gov/</a> , and British Columbia Ministry of Environment, 2014, <a href="http://www.env.gov.bc.ca/soils/">http://www.env.gov.bc.ca/soils/</a>

<sup>1</sup>New characteristic not currently defined in described form within StreamStats.



**Figure 4.** Comparison between Geographic Information System (GIS) estimated drainage area and NWIS-reported drainage area for streamgages in Idaho and bordering States.

A smaller rather than larger DEM cell size results in more accurate basin delineations, as well as slope and elevation estimates. These improved accuracies are most beneficial in basins with small drainage areas, complex terrain, or small relief. Updated land cover datasets were used where possible to better represent current (2013) conditions in the study area. Other basin and climatic characteristics utilized in similar regional peak-flow regression studies (Austin and others, 2011, Gotvald and others, 2012) also were considered (table 2). A noted difference from the Berenbrock (2002) and Hortness and Berenbrock (2003) was the availability of some Canadian geospatial datasets for use in the current study, which allowed the use of streamgages with drainage basins crossing into Canada. Some basin and climatic characteristics that were initially considered were not available in Canada and were removed from further consideration.

ArcGIS (Environmental Systems Resource Institute, 2013) was used to calculate all basin and climatic

characteristics (except for the aforementioned drainage areas determined in StreamStats for streamgages in States where StreamStats was available) because of the large number of streamgages and the need for a consistent methodology in measuring and calculating characteristics. Twenty-four separate basin characteristics were obtained for each of the 192 streamgages (table 2) and evaluated as explanatory variables in the regional peak-flow regression equations. Of the 24 characteristics initially evaluated, 4 were included as explanatory variables in at least one of the equations. Values of basin and climatic characteristic estimated for the 192 streamgages and used in the peak-flow regression equations are provided in appendix A (table A2). Of the four characteristics used in the final equations, two are updated versions of datasets used in Berenbrock (2002) and Hortness and Berenbrock (2003)—percentage of basin covered by forest and mean annual precipitation.

## Estimating Peak-Flow Frequency Statistics at Selected Gaged Sites

A peak-flow frequency analysis is based on a statistical evaluation of a record of annual peak flows, typically maximum instantaneous flows but sometimes mean daily flows, collected at USGS streamgages. Peak-flow frequency refers to the statistical probability that a flow will meet or exceed a certain value during any given year, referred to as the annual exceedance probability or AEP. Previous peak-flow studies in Idaho and elsewhere have commonly described peak-flow statistics relative to recurrence intervals, which are the inverse of AEPs. Describing flood magnitude estimates in terms of recurrence intervals has caused confusion. For example, use of the term “100-year flood,” where 100 years is the recurrence interval, gives the false impression that a flow will occur only once every 100 years. In fact, the flow has a 1-percent chance of occurring in any given year, or AEP. As a result, the USGS and other agencies have encouraged the usage of AEP instead of recurrence interval (Holmes and Dinicola, 2010).

### Peak-Flow Frequency Analysis

Eleven peak-flow statistics, ranging from the 80- to 0.2-percent AEP, were estimated for this study (table 1). The Bulletin 17B method (Interagency Advisory Committee on Water Data, 1982), has long been the standard methodology in the United States for fitting peak flows to a log Pearson Type III probability distribution and performing frequency analyses. The general process for fitting measured annual peak-flows to a log-Pearson Type III distribution is discussed in detail in Berenbrock (2002). The basic equation for fitting the log-Pearson Type III distribution to a measured series of annual peak flows is:

$$\log Q_{\text{AEP}} = \bar{X} + K_p S \tag{3}$$

where

- $Q_{\text{AEP}}$  is the annual exceedance probability flow, in cubic feet per second (ft<sup>3</sup>/s);
- $\bar{X}$  is the mean of the logarithms of the annual peak flows;
- $K_p$  is a frequency factor based on the skew coefficient and the given annual exceedance probability; and
- $S$  is the standard deviation of the logarithms of the annual peak flows.

The mean, standard deviation, and skew coefficient can be used to describe the mid-point, slope, and asymmetry or curvature of the frequency distribution of annual peak flows,

respectively. The skew coefficient, which dictates term  $K_p$  in equation 3, can be estimated based only on a streamgage’s peak-flow record, called “station skew,” or on a station skew weighted with a generalized or regional skew value to compensate for uncertainties associated with short record lengths. For the study described in this report, a log-Pearson type III distribution initially was fit to every streamgage’s annual peak-flow record using the USGS PeakFQ program, version 7.1 (Flynn and others, 2006; Veilleux and others, 2014).

### Expected Moments Algorithm Frequency Analysis and Multiple Grubbs-Beck Test for Detecting Low Outliers

Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) procedures allow limited options for detecting outliers and representing historical peak flows to improve the accuracy of peak-flow statistics. The USGS recently released and incorporated two new peak-flow statistical techniques into the USGS PeakFQ program: EMA and generalized MGBT, for potentially influential low peak-flow outliers. As initially described in the section, “Introduction,” both EMA and MGBT techniques were used in PeakFQ in this study because the combined techniques have been shown to provide more efficient, accurate estimates than those generated using the standard Bulletin 17B process when a peak-flow record contains gaps, historical flood measurements, censored data, or low outliers (Cohn and others, 1997, 2013; Paretti and others, 2014).

Many streamgages have two types of data in the annual peak-flow record: (1) systematic, with a peak-flow value recorded for each year; and (2) historical, or isolated measurements made outside the systematic period of record (typically during extreme hydrologic conditions) or estimated using paleoflood techniques. In these two general types of peak-flow data, some peaks can be identified as “censored,” which means that the actual peak flow is uncertain and is documented as greater than or less than some value. Censored peaks typically are noted when the water level during a flood rises above the maximum elevation or below the minimum elevation that can be recorded by the streamgage equipment. The latter scenario can occur at streamgages with crest stage gages, which are passive measurement devices that document only the highest water level between servicing visits. Crest stage gages typically are installed at an elevation above the average water level and have a minimum measurable elevation and associated flow. In a given year, if the peak water level does not rise above the minimum elevation of the crest stage gage, the annual peak is documented as censored, or less than the minimum measurable flow value. The EMA methods allow improved handling of all of these data types than was previously possible using strict Bulletin 17B procedures.

Prior to EMA, historical flood measurements that occurred outside the systematic period could be used in the analysis, but potentially useful information on conditions during any gaps in the systematic period was not considered. For example, during a gap in the systematic record at a particular streamgage, the knowledge that a particular flow would have been noticed and measured if it had occurred can provide valuable information for the peak-flow frequency analysis. The EMA method allows the use of intervals to describe conditions during gaps between the historical measurements and the systematic record. The EMA method requires setting two sets of intervals for each peak and year in the analysis, starting with the first measured or recorded peak flow: a perception threshold interval and a flow interval (table 3). Perception thresholds describe the minimum and maximum peak flows that would have been measured if they had occurred (Veilleux and others, 2014). Flow intervals describe the uncertainty associated with a peak flow. Defining the perception thresholds often is subjective because of a lack of historical documentation and anecdotal information. For this study, perception thresholds generally were set to (0, infinity) for the systematic record, (peak, infinity) for any historical peak measurements, and (infinity, infinity) for any gaps in the systematic record if no additional information was available (table 3). In several cases, the availability of historical measurements during gaps in the systematic record provided some useful information about what flows might

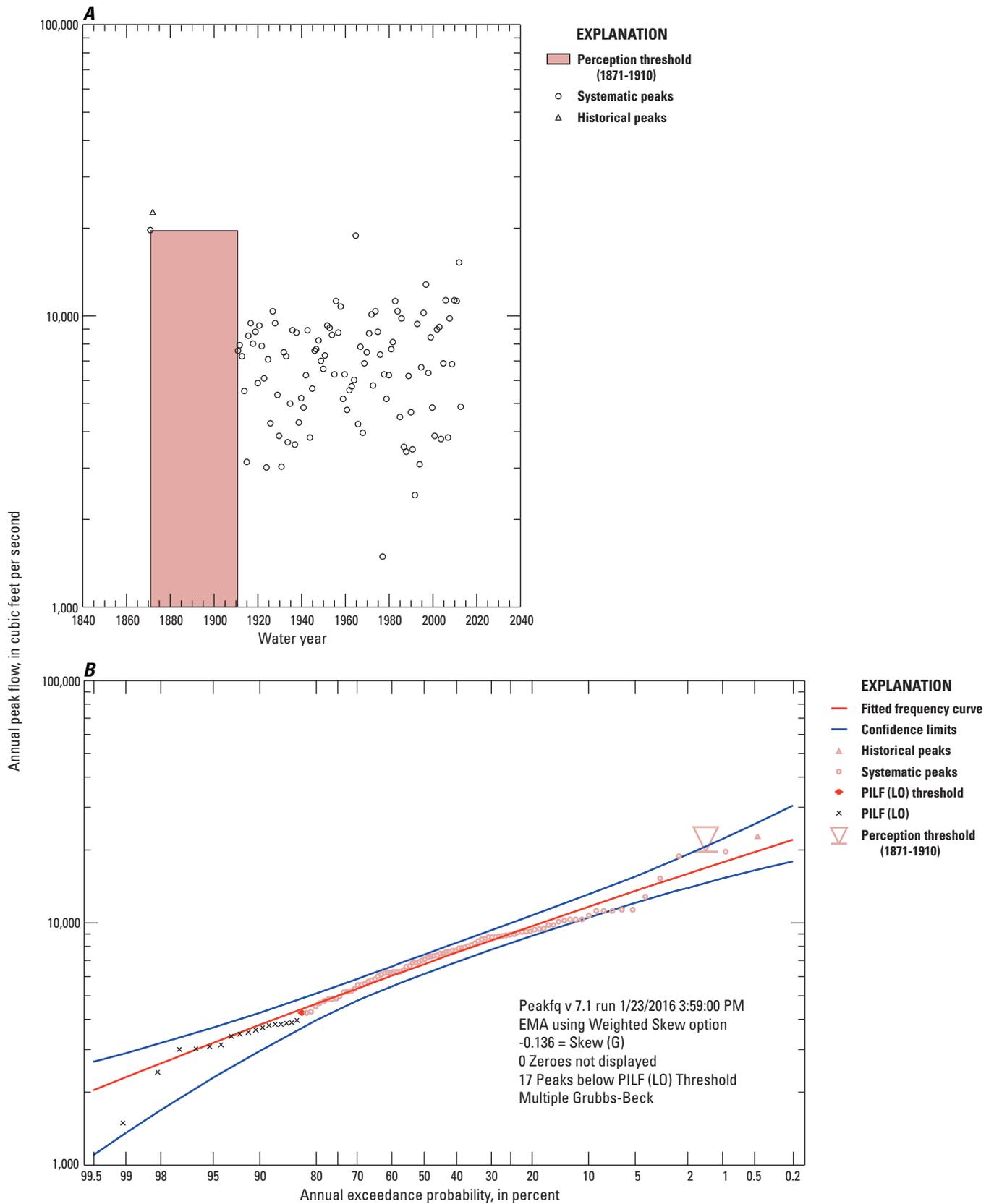
have been observed and measured if they had occurred. Peaks can be categorized as truly “historical” if they: (1) occurred outside the systematic period, and (2) are confidently believed to be larger than any other peak flow that would have occurred in the gap in the systematic record. Several streamgage records contained peak flows that were coded as historical but were considerably smaller in magnitude than other historical or systematic peaks. In these cases, the peak was considered an “opportunistic” peak rather than a truly “historical” peak and was omitted from the EMA analysis by setting the perception threshold to (infinity, infinity) and the flow interval to (0, infinity) (table 3). If peaks were categorized as truly “historical,” this knowledge was used to infer that peak flows higher than the historical measurement would likely have been measured if they had occurred during gaps in the systematic record around the time of the historical peak. In this case, the perception threshold was set to (historical peak, infinity) during the gap in systematic record. For example, the record of annual peak flows for USGS streamgage 13185000 included systematic peaks in 1911–2014 and two peaks coded as historical in 1871 and 1872 (fig. 5A). Using the EMA method, 1871–1910 was assigned a perception threshold equal to the 1871 peak flow (the smaller of the two historical peaks, which is labeled as systematic in fig. 5A), assuming that a flow larger than the 1871 peak would have been measured if it had occurred in 1871–1910. In fact, such a flow did occur and was measured as the 1872 historical peak.

**Table 3.** General perception threshold and flow interval settings applied to various scenarios in the Expected Moments Algorithm analysis to estimate peak-flow statistics at streamgages.

Annual scenario or peak-flow type	Perception thresholds		Flow intervals	
	Minimum	Maximum	Minimum	Maximum
Systematic record peak, known with confidence	0	Infinity	Peak	Peak
Systematic record peak, peak greater than stated value	0	Peak	Peak	Infinity
Historical peak <sup>1</sup>	Historical peak	Infinity	Historical peak	Historical peak
Crest stage gage peak within measurable range	Lowest flow measureable by gage	Infinity	Peak	Peak
Crest stage gage peak below measurable range, peak less than stated value	Lowest flow measureable by gage	Infinity	0	Lowest flow measureable by gage
Gaps in systematic record, no other available information	Infinity	Infinity	0	Infinity
Gaps in systematic record, additional information available from historical peak(s) <sup>2</sup>	Historical peak	Infinity	0	Historical peak

<sup>1</sup>For streamgages with multiple historical peaks, the lowest historical peak generally was used to define the perception thresholds.

<sup>2</sup>The selection and application of perception thresholds and flow intervals for gaps in systematic record was subjective and site dependent. See appendix A, tables A3 and A4 for site-specific information.



**Figure 5.** Output from the U.S. Geological Survey (USGS) PeakFQ program showing (A) Expected Moments Algorithm (EMA) treatment of historical peaks and gaps in systematic record using a perception threshold and (B) fitted frequency curve with historical peaks, systematic peaks, confidence limits, potentially influential low-flow (PILF) threshold, and peaks identified as PILF outliers for USGS streamgage 13185000, Boise River near Twin Springs, Idaho.

Flow intervals were set for each year in a streamgage's annual peak-flow record to represent uncertainty in the available information (table 3). Most systematic and historical peaks were considered known with confidence, so the flow intervals were set as (peak, peak). Flow intervals for other peak-flow types were set to (0, censoring level), (0, infinity), (0, peak), or (peak, infinity) depending on the scenario (table 3). A flow interval of (0, infinity) means that there is no available information to be able to assess uncertainty and was applied to years with gaps in the systematic record with no additional historical or anecdotal information. Perception thresholds and flow intervals for annual peak-flow records and various peak types for streamgages used in the analysis are provided in appendix A (tables A3, A4). Additional information on the EMA method's treatment of peak-flow types and on setting intervals is provided in Paretti and others (2014).

Many of the evaluated annual peak-flow records contained one or more low peak flows, referred to as low outliers. These low outliers often had high leverage or influence in fitting the frequency curve to the entire record of peak flows, which resulted in a poor fit of the curve at high flows with low AEPs, which are the peak-flow statistics most frequently used for flood protection and infrastructure design. Additionally, low outliers often are considered to reflect physical processes that are not necessarily related to the processes associated with large flood events, and their use in the frequency analysis should be limited (Cohn and others, 2013). Bulletin 17B procedures state that low outliers should be subjectively removed based on hydrologic judgment using the Grubbs-Beck test described in Grubbs and Beck (1972). The generalized MGBT, described in Cohn and others (2013), objectively and systematically detects and removes potentially influential low-flow (PILF) outliers below a PILF threshold and can be used in concert with EMA methods in the USGS PeakFQ program. An example of a dataset with multiple PILF outliers is provided for USGS streamgage 13185000 (fig. 5B). The PILF threshold for the dataset was 4,200 ft<sup>3</sup>/s, so all annual peak flows less than the PILF threshold were categorized as PILF outliers. For this and other streamgages used in the study, removing the PILFs generally produced a better fit of the frequency curve for low AEP statistics (large peak flows).

## Regional Skew Analysis

The at-station skew coefficient described in section, "Peak-Flow Frequency Analysis," is sensitive to extreme peak flows and might not provide an accurate measure of the true skew of peak flows that occur at a site, particularly for streamgages with short periods of record. The use of a regional skew coefficient, estimated from long-term streamgages in a large area and representative of regional peak-flow characteristics, can be used to weight the at-station skew to provide a more accurate measure of skew for calculating peak-flow statistics. The regional skew map of

logarithms of annual peak flow developed by Hardison (1974) and included in Bulletin 17B has traditionally been used for peak-flow frequency analysis in Idaho, including most recently by Berenbrock (2002) and Hortness and Berenbrock (2003). However, this map was based on a limited dataset and did not account for low outliers or historical information. The Bayesian-Generalized Least Squares method, described in Veilleux and others (2012), was developed to estimate regional skew with greater precision and accuracy than the Bulletin 17B skew map from Hardison (1974). The Bayesian-Generalized Least Squares method was used to estimate skew values for a regional-scale area in the Pacific Northwest that included Idaho, Montana, Oregon, and Washington and was based on 290 streamgages using annual peak-flow records through water year 2012. The process to determine regional skew is described in detail in appendix B.

A constant regional skew of -0.07 was determined to be the best fit of the datasets evaluated in the Bayesian-Generalized Least Squares analysis. The regional skew model produced for this study had a higher precision than the Bulletin 17B skew map and contained 2.4 times the information content (as measured by pseudo- or effective record length) of that claimed by the Bulletin 17B skew map. The constant regional skew of -0.07 was used to weight the at-station skew coefficients in the EMA analysis for all streamgages used in the regional peak-flow regression analysis except for six streamgages in Montana. The annual peak-flow records from these Montana streamgages were shown in Parrett and Johnson (2004) to have a mixed population; that is, the hydrologic cause of the peaks varied from year to year. Some peaks were caused by spring rain-on-snow events but others were caused by summer monsoonal events, which resulted in high uncertainty when fitting a single frequency curve to all peaks in the dataset. Parrett and Johnson (2004) ran separate frequency analyses for the two types of peaks and integrated the analyses using a joint-probability approach. Subsequent work by USGS Hydrologists in Montana showed that simply using the at-station skew at these streamgages resulted in peak-flow statistics that were similar to those generated by the more complex, joint-probability approach (Sando and others, 2016). After consultation with USGS Hydrologists in Montana, it was decided to include the six Montana streamgages with mixed-population peak-flow records but only use the at-station skew coefficients in the peak-flow frequency analysis. The skew option used for each streamgage is identified in appendix A (table A1).

## Final Peak-Flow Frequency Statistics

The final peak-flow statistics estimated for each streamgage, using EMA and MGBT techniques and appropriate regional skew values, are provided in appendix A (table A2). The statistics for streamgages in States surrounding Idaho were submitted for review by USGS Hydrologists in those States, and were revised in some cases based on review comments, prior to use in the regional peak-flow regression analysis.

## Estimating Peak-Flow Frequency Statistics at Ungaged Sites Through a Regional Regression Analysis

After the calculation of peak-flow statistics at selected streamgages (table A2), the process to develop the regional regression equations to estimate statistics at ungaged sites involved multiple steps. The steps included (1) determination of regions with similar hydrologic and basin characteristics, (2) performing an exploratory data analysis and ordinary least squares (OLS) regression procedure to determine which basin and climatic characteristics were most closely correlated with peak-flow statistics, and (3) creating and selecting the final regression equations using a GLS regression procedure. The selected regression equations should be applied with an understanding of the inherent limitations and uncertainties. An alternative method to the regional regression equations also was developed for use in estimating peak-flow statistics at ungaged sites close to streamgages used in the analysis.

### Determination of Regions for Regional Regression Analysis

Berenbrock (2002) and Hortness and Berenbrock (2003) initially divided Idaho into nine regions based on (1) grouping of similar basin and climatic characteristics, completed with the aid of a statistical cluster analysis; (2) geographic features, such as mountain ranges, vegetation cover, and breaks from mountains to plains; and (3) professional judgment based on knowledge of the flow characteristics of each region. Berenbrock (2002) and Hortness and Berenbrock (2003) later separated the Snake River Plain (region 7) into two regions (7a and 7b), the western and eastern Snake River Plains, based on observations during the cluster analysis showing some diversity in basin and climatic characteristics between the two areas, which also was supported by observations described in Thomas and others (1973, 1994). Berenbrock (2002) and Hortness and Berenbrock (2003) also excluded from the regional peak-flow regression analysis a part of the eastern Snake River Plain (called region 0 in those reports; fig. 3) because of the extent of regulation, groundwater-surface water interactions, and high infiltration rates in the area. The effects of these features were considered too great to allow the accurate characterization of peak-flow statistics using regional methods.

The same boundaries of these regions used in the Berenbrock (2002) and Hortness and Berenbrock (2003) analysis were retained for the current study (fig. 3), except that regions 1 and 2, and 6 and 8 were combined because of similar basin and climatic characteristics and because the number of streamgages available in the separate regions was small. The streamgage screening process described in section, “Streamgage Selection,” resulted in a low number of streamgages available for the analysis, in region 7

in particular. The Berenbrock (2002) and Hortness and Berenbrock (2003) regions 7a and 7b were combined into one region (7) for the current study to benefit from a larger number of streamgages and wider range of basin and climatic characteristics available for the analysis than if the region was split. The combination of the regions resulted in improved relations observed through exploratory data analysis, described in the section, “Exploratory Data Analysis and Ordinary Least Squares Regression.” The original region labels were retained (for example, 1\_2) to maintain consistency with and to compare with the results in Berenbrock (2002) and Hortness and Berenbrock (2003). The blank area (region 0, fig. 3) in the eastern Snake River Plain was excluded from the analysis in the current study for the same reasons identified in Berenbrock (2002) and Hortness and Berenbrock (2003).

### Exploratory Data Analysis and Ordinary Least Squares Regression

An exploratory data analysis was completed using a combination of statistical techniques supported by the R (R Core Team, 2015) and Number Crunching Statistical Software (NCSS; 2015) packages. Matrices of scatter plots between estimated peak-flow statistics and basin and climatic characteristics were created to evaluate which characteristics had the best visual, linear relation for each statistic in each region. Some patterns indicated transformations of variables would be appropriate to develop more linear relations. Various transformations were evaluated, including log, natural logarithm, square root, and reciprocal square root; the log transformation resulted in the best linear relations and most constant variance about the regression line compared to the other transformations.

Next, OLS regression was completed in R or NCSS to evaluate the best statistical relations between basin and climatic characteristics and peak-flow statistics in each region. An “all possible” regression was first completed in R or NCSS to determine the best models using simple linear regression and multiple linear regression with up to five explanatory variables. Next, a “best subset selection” linear regression technique called “hierarchical forward selection with switching” (Number Crunching Statistical Software, 2015) was completed in NCSS to determine the best OLS fit for regression models with one, two, three, and four explanatory variables. The selection of the final, “best fit” OLS model was made after reviewing several statistical metrics, including the coefficient of determination ( $R^2$ ) for simple linear regression models or adjusted- $R^2$  for multiple linear regression models; standard error; normal distribution evaluated using the Shapiro Wilk normality test; random patterns in residuals; multicollinearity evaluated using variance inflation factors; and serial correlation evaluated using the Durbin-Watson test. In general, the number of explanatory variables evaluated in each OLS regression was limited to about 1 for every 10 streamgages according to guidance in Harrell (2015).

Variance inflation factors greater than 3 can indicate possible multicollinearity, which means that explanatory variables may be linearly dependent and correlated, and factors greater than 10 indicate serious problems could occur by including the explanatory variable (Helsel and Hirsch, 2002). Multicollinearity was unlikely in the current study because variance inflation factors were less than 2 for all explanatory variables selected for the “best fit” OLS regression models.

Drainage area was selected as a possible explanatory variable for all regions during the exploratory data analysis phase. In several regions, however, basin perimeter was better correlated than drainage area with peak-flow statistics. The authors consulted with USGS statistical experts regarding potential concern over the use of basin perimeter, because it is not always derived or measured in a consistent way depending on the resolution of the underlying DEM datasets. Additionally, basins with similar drainage areas but large rather than small relief or terrain complexity may have drainage divides that are more jagged, resulting in larger basin perimeters. In such cases, the ratio of perimeter to area might be considered to be useful for representing differences in hydrologic processes and peak-flow statistics in basins with low compared to high terrain complexity; however, other explanatory variables evaluated in this study that represented terrain complexity (such as basin slope and relief) did not have statistically significant correlations with peak-flow statistics. No physical explanation emerged why basin perimeter would be better correlated with peak-flow statistics than drainage area in this study. As a result, drainage area was selected over basin perimeter as an explanatory variable during the exploratory data analysis phase.

## Generalized Least Squares Regression

The results of the exploratory data analysis and OLS regression were used to inform and narrow the combinations of basin and climatic characteristics and peak-flow statistics to evaluate in the final GLS regression analysis. The USGS Weighted Multiple Linear Regression (WREG) program, version 1.05 (Eng and others, 2009), was used to develop final regional peak-flow regression equations. Three least squares regression techniques are provided by WREG to relate each peak-flow to selected basin and climatic characteristics: OLS, Weighted Least Squares (WLS), and GLS. The GLS technique is preferred for peak-flow frequency analysis when peak-flow records used in the analysis are of varying lengths and are possibly correlated for concurrent years at nearby streamgages (Griffis and Stedinger, 2007; Gotvald and others, 2012). The GLS technique weights each peak-flow statistic based on record length and also accounts for cross-correlation caused by an overlapping period of record at nearby gages and variable record length among sites. The GLS technique, assumptions, and limitations are described in more detail in Stedinger and Tasker (1985), Tasker and Stedinger (1989), Griffis and Stedinger (2007), and Eng and others (2009).

All explanatory variables were log transformed prior to evaluation in the GLS regression models based on observations during the exploratory data analysis phase. A log transformation on the dependent variable,  $Q_{AEP}$ , was required by the WREG program. Some basin and climatic characteristics were normalized or adjusted before evaluation using GLS regression techniques in WREG. Characteristics associated with elevation were divided by 1,000 before log transformation so that the regression coefficients would be on a similar scale with the regression coefficients for other explanatory variables. A value of 1 percent was added to all characteristics in percent units before log transformation because zero values, if they occur in a basin, cannot be log transformed and would result in an error in the regression model output. Explanatory variables were considered statistically significant if their p-values were less than the threshold significance level of 0.05.

Output of WREG provides various measures of the reliability of the GLS regression equations: the average variance of prediction (AVP, in log units), model error variance (MEV, in log units), the average standard error of prediction (SEP, in percent), the average standard error of model (SEM, in percent), the pseudo- $R^2$ , and leverage and influence of individual observations on the regression. Equations for calculating these metrics are available in Eng and others (2009) and Gotvald and others (2012). The SEP is considered the total error for a regression model. AVP represents the average accuracy of prediction for all the streamgages used in the regression analysis. AVP and SEP are measures that indicate how well the equation performs at predicting  $Q_{AEP}$  for ungaged sites and were considered the most useful metrics for evaluating the regression models. The pseudo- $R^2$  and SEM are metrics that indicate how well the regression model performs at predicting  $Q_{AEP}$  for the streamgages used in the analysis.

## Regional Peak-Flow Regression Equations

The final regional peak-flow regression equations (table 4), based on peak-flow statistics and basin and climatic characteristics from 192 streamgages, had the following log-linear form:

$$\log Q_{AEP} = \log K + a_1 \log x_1 + a_2 \log x_2 + \dots + a_p \log x_p \quad (4)$$

where

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

Equation 4 was then written in terms of the actual variable values rather than in log form as:

$$Q_{\text{AEP}} = K x_1^{a_1} x_2^{a_2} \dots x_p^{a_p} \quad (5)$$

where

$K$  is the antilog ( $10^K$ ) of the regression constant or intercept and all other terms are as previously described.

The final regression equations were selected by evaluating regression metrics output by WREG, weighed with consideration of which explanatory variables were logical for estimating peak-flow statistics in each region. Desirable regression equations resulted in: (1) low error in prediction and model, expressed as the average SEP and SEM; (2) high pseudo- $R^2$ ; (3) random distribution of regression residuals around the zero line in residuals plots; and (4) low number of data points with high leverage or influence. Leverage and influence were used as regression metrics to identify outliers in datasets. Data points with high leverage have values of independent variables that are unusual or substantially different than the corresponding values at other sites and have the potential to influence a regression. Data points with high influence are outliers that do have an influence on the regression coefficients. Leverage and influence statistics were output by the WREG program. About 33 data points with high leverage or influence were investigated for possible causes. Four data points with high influence were noted to be from streamgages with annual peaks that were measured using crest stage gages, which was previously unknown during the peak-flow frequency analysis. As a result, the peak-flow statistics and regression equations were re-estimated after changing the perception thresholds and flow intervals in the EMA analysis. No errors were noted with the other investigated data points, so they were left in the regression analysis.

Drainage area was a statistically significant explanatory variable ( $p < 0.05$ ) in the final regression equations for all statistics in all regions, and mean annual precipitation was statistically significant in four out of six regions (table 4). The regression equations in region 3 were limited to one explanatory variable (drainage area) because only 13 streamgages were available in that region for the analysis. Use of more than one explanatory variable did not substantially improve the regressions and would have violated the general statistical rule of thumb to limit explanatory variables to about 1 per 10 observations (Harrell, 2015). One exception to the statistical rule of thumb presented in Harrell (2015) was made in region 4. The final regression equations for region 4 were built using three explanatory variables (drainage area, forested area, and mean annual precipitation), although only 28 streamgages were available for the analysis. The decision was made to use three explanatory variables

because (1) the selection of these three variables as the best fit for a three-variable model was consistent across all statistics, but the best variables for a two-variable model differed slightly across statistics, and (2) the regression model metrics were substantially improved by using a three-variable model, particularly for the large peak flow, low AEP statistics.

The use of percentage forested area in region 4 initially was questioned because of work documented in Rea and Skinner (2009) and Wood and others (2009), which showed that percentage-based variables such as forested area tended to produce unreasonable estimates of flow statistics when the regression equation was used in extrapolation. In very small drainage areas, for example, percentage-based variables tended to have extreme values of either zero or 100 percent. These extremes are rarely present in the large basins upstream of streamgages. For the current study described in this report, a sensitivity analysis was completed to test the performance of the selected regression equations in region 4 in extrapolation if forested area was used as an explanatory variable. None of the tested scenarios produced statistics that appeared unreasonable for the input variables. A high percentage of forest cover has the potential to slow overland runoff during rainfall and snowmelt runoff events, thus reducing the magnitude of peak flows compared to a similar basin with low forest cover. Forested area was considered a logical explanatory variable in the region 4 regression because of the wide range of forest land cover in this region and significant negative correlation with peak-flow statistics.

Another consideration when selecting the final regression equations was ensuring consistency across the range of peak-flow statistics evaluated for this report. A consistent regression model form with the same explanatory variables across all statistics was preferred over regression models optimized for each individual statistic. Using the absolute “best fit” regression model for each statistic in a region, often requiring use of different explanatory variables for each statistic, resulted in a discontinuity in estimated statistics for some regions. In other words, regression models optimized for each statistic sometimes resulted in a 10-percent AEP flow that was smaller than the 20-percent AEP flow if different explanatory variables were used, which was not logical. The final regression equations for high AEP (small peak flows) statistics in regions 4 and 7 included an explanatory variable that had a p-value greater than the threshold used to determine statistical significance (0.05; table 4) to avoid the discontinuity in statistics. The regression models for regions 4 and 7 were the best fit for the low AEP (large peak flows) statistics, and the same model form was retained for all statistics in the region. For region 4, the forested area term was not significant ( $p = 0.099$ ) for the 80-percent AEP but was significant for all other statistics. For region 7, the minimum basin elevation term was not significant ( $p = 0.064$  to 0.883) for the 80–10-percent AEPs but was significant for 4–0.2-percent AEPs.

**Table 4.** Regional regression equations for estimating peak-flow frequency statistics at ungaged sites in Idaho.

[Variables in red font were not statistically significant at a significance level of 0.05 but were retained in the regression model to ensure a logical increase in peak-flow statistics with a decrease in annual exceedance probability (AEP). **Abbreviations:**  $Q_{AEP}$ , peak-flow magnitude, in cubic feet per second, for specified AEP; **MEV**, model error variance; **AVP**, average variance of prediction; **SEM**, average standard error of model; **SEP**, average standard error of prediction; **Pseudo-R<sup>2</sup>**, pseudo coefficient of determination; **n**, number of gaging stations used in developing regression equations for indicated hydrologic region; **A**, contributing drainage area, in square miles; **P**, mean annual precipitation, in inches; **F**, percent of drainage basin with forest land cover;  $E_{min}$ , minimum elevation in basin, in feet]

Regression equation for indicated $Q_{AEP}$	MEV (log units)	AVP (log units)	SEM (percent)	SEP (percent)	Pseudo-R <sup>2</sup> (percent)
Region 1_2; n = 59					
$Q_{80} = 0.000815 A^{0.960} P^{2.49}$	0.065	0.069	64.3	66.7	91.6
$Q_{67} = 0.00141 A^{0.945} P^{2.41}$	0.063	0.067	63.0	65.3	91.6
$Q_{50} = 0.00238 A^{0.930} P^{2.34}$	0.063	0.066	62.8	65.0	91.4
$Q_{43} = 0.00294 A^{0.923} P^{2.32}$	0.062	0.066	62.3	64.5	91.4
$Q_{20} = 0.00616 A^{0.899} P^{2.23}$	0.065	0.069	64.3	66.7	90.4
$Q_{10} = 0.00962 A^{0.882} P^{2.18}$	0.070	0.074	66.8	69.4	89.4
$Q_4 = 0.0148 A^{0.865} P^{2.14}$	0.075	0.080	70.0	72.9	88.2
$Q_2 = 0.0191 A^{0.854} P^{2.12}$	0.080	0.086	72.7	75.8	87.2
$Q_1 = 0.0239 A^{0.844} P^{2.11}$	0.086	0.093	76.2	79.7	85.9
$Q_{0.5} = 0.0288 A^{0.834} P^{2.10}$	0.093	0.100	79.7	83.5	84.7
$Q_{0.2} = 0.0355 A^{0.823} P^{2.09}$	0.101	0.108	84.0	88.2	83.1
Region 3; n = 13					
$Q_{80} = 8.15 A^{0.818}$	0.135	0.158	102	114	78.3
$Q_{67} = 11.7 A^{0.807}$	0.117	0.137	92.6	103	80.3
$Q_{50} = 17.2 A^{0.796}$	0.097	0.114	82.2	91.2	82.7
$Q_{43} = 20.1 A^{0.792}$	0.090	0.106	78.2	86.7	83.6
$Q_{20} = 35.6 A^{0.777}$	0.064	0.076	63.5	70.4	87.5
$Q_{10} = 51.8 A^{0.769}$	0.048	0.059	54.0	60.4	90.1
$Q_4 = 75.7 A^{0.764}$	0.032	0.040	42.7	48.8	93.2
$Q_2 = 95.9 A^{0.762}$	0.022	0.030	35.2	41.5	95.1
$Q_1 = 117 A^{0.761}$	0.014	0.021	27.8	34.5	96.8
$Q_{0.5} = 140 A^{0.762}$	0.009	0.015	21.5	29.1	98.0
$Q_{0.2} = 171 A^{0.764}$	0.003	0.010	13.0	22.8	99.2
Region 4; n = 28					
$Q_{80} = 0.000592 A^{0.981} P^{2.80} (F + 1)^{-1.52}$	0.059	0.069	60.8	66.4	92.1
$Q_{67} = 0.00126 A^{0.972} P^{2.69} (F + 1)^{-1.72}$	0.051	0.060	55.9	61.0	92.8
$Q_{50} = 0.00272 A^{0.962} P^{2.57} (F + 1)^{-1.93}$	0.044	0.051	51.2	55.8	93.5
$Q_{43} = 0.00370 A^{0.957} P^{2.52} (F + 1)^{-2.01}$	0.040	0.047	48.5	52.9	94.0
$Q_{20} = 0.0123 A^{0.939} P^{2.34} (F + 1)^{-2.31}$	0.030	0.035	41.2	45.2	95.2
$Q_{10} = 0.0279 A^{0.926} P^{2.21} (F + 1)^{-2.51}$	0.025	0.031	38.0	42.0	95.6
$Q_4 = 0.0773 A^{0.908} P^{2.04} (F + 1)^{-2.70}$	0.022	0.027	34.8	38.9	96.1
$Q_2 = 0.149 A^{0.897} P^{1.92} (F + 1)^{-2.82}$	0.020	0.025	33.1	37.5	96.3
$Q_1 = 0.292 A^{0.884} P^{1.80} (F + 1)^{-2.92}$	0.019	0.025	32.7	37.3	96.3
$Q_{0.5} = 0.534 A^{0.872} P^{1.69} (F + 1)^{-3.00}$	0.019	0.024	32.3	37.2	96.3
$Q_{0.2} = 1.09 A^{0.859} P^{1.56} (F + 1)^{-3.11}$	0.019	0.026	32.9	38.2	96.0

**Table 4.** Regional regression equations for estimating peak-flow frequency statistics at ungaged sites in Idaho.—Continued

Regression equation for indicated $Q_{AEP}$	MEV (log units)	AVP (log units)	SEM (percent)	SEP (percent)	Pseudo-R <sup>2</sup> (percent)
Region 5; n = 20					
$Q_{80} = 0.000316 A^{0.934} P^{2.84}$	0.009	0.012	22.6	25.1	97.8
$Q_{67} = 0.000506 A^{0.939} P^{2.75}$	0.010	0.011	22.7	25.0	97.8
$Q_{50} = 0.000802 A^{0.943} P^{2.67}$	0.010	0.012	22.8	25.1	97.8
$Q_{43} = 0.000984 A^{0.944} P^{2.63}$	0.011	0.013	24.1	26.5	97.6
$Q_{20} = 0.00189 A^{0.945} P^{2.53}$	0.012	0.014	25.5	28.0	97.4
$Q_{10} = 0.00279 A^{0.944} P^{2.47}$	0.013	0.016	26.8	29.5	97.1
$Q_4 = 0.00420 A^{0.941} P^{2.42}$	0.014	0.017	28.0	31.1	96.9
$Q_2 = 0.00526 A^{0.939} P^{2.39}$	0.016	0.020	30.2	33.6	96.4
$Q_1 = 0.00644 A^{0.936} P^{2.37}$	0.018	0.022	31.3	34.9	96.1
$Q_{0.5} = 0.00758 A^{0.934} P^{2.35}$	0.019	0.023	32.4	36.2	95.9
$Q_{0.2} = 0.00912 A^{0.931} P^{2.34}$	0.021	0.026	34.4	38.6	95.4
Region 6_8; n = 48					
$Q_{80} = 0.00115 A^{0.783} P^{2.64}$	0.065	0.070	64.2	66.9	81.7
$Q_{67} = 0.00252 A^{0.768} P^{2.49}$	0.060	0.064	61.2	63.7	82.2
$Q_{50} = 0.00557 A^{0.754} P^{2.34}$	0.056	0.060	58.9	61.3	82.4
$Q_{43} = 0.00773 A^{0.748} P^{2.28}$	0.056	0.060	58.7	61.1	82.2
$Q_{20} = 0.0250 A^{0.726} P^{2.05}$	0.057	0.061	59.2	61.7	80.7
$Q_{10} = 0.0534 A^{0.713} P^{1.91}$	0.060	0.064	61.1	63.8	78.9
$Q_4 = 0.118 A^{0.698} P^{1.76}$	0.067	0.072	65.2	68.1	75.8
$Q_2 = 0.198 A^{0.688} P^{1.66}$	0.072	0.078	68.4	71.5	73.6
$Q_1 = 0.314 A^{0.679} P^{1.58}$	0.078	0.084	71.7	75.1	71.2
$Q_{0.5} = 0.474 A^{0.670} P^{1.50}$	0.084	0.091	75.1	78.8	69.0
$Q_{0.2} = 0.789 A^{0.659} P^{1.41}$	0.093	0.101	79.9	84.0	65.8
Region 7; n = 24					
$Q_{80} = 1.64 A^{0.563} (E_{min}/1,000)^{1.09}$	0.164	0.190	118	132	63.7
$Q_{67} = 5.20 A^{0.547} (E_{min}/1,000)^{0.602}$	0.144	0.166	107	119	66.0
$Q_{50} = 17.2 A^{0.529} (E_{min}/1,000)^{0.104}$	0.127	0.146	97.8	108	68.4
$Q_{43} = 27.9 A^{0.523} (E_{min}/1,000)^{-0.098}$	0.120	0.138	94.2	104	69.7
$Q_{20} = 155 A^{0.500} (E_{min}/1,000)^{-0.808}$	0.103	0.120	85.2	94.2	73.3
$Q_{10} = 442 A^{0.489} (E_{min}/1,000)^{-1.24}$	0.097	0.114	82.1	91.3	75.2
$Q_4 = 1,240 A^{0.483} (E_{min}/1,000)^{-1.65}$	0.095	0.113	80.9	90.7	76.9
$Q_2 = 2,320 A^{0.480} (E_{min}/1,000)^{-1.91}$	0.097	0.117	82.2	92.8	77.2
$Q_1 = 4,000 A^{0.478} (E_{min}/1,000)^{-2.13}$	0.100	0.121	83.6	94.9	77.6
$Q_{0.5} = 6,500 A^{0.477} (E_{min}/1,000)^{-2.33}$	0.105	0.128	86.4	98.6	77.5
$Q_{0.2} = 11,600 A^{0.476} (E_{min}/1,000)^{-2.56}$	0.112	0.138	90.2	104	77.4

Region 7 presented the greatest challenges for fitting a regional regression model among all other regions. Region 7 is semi-arid, with high variability in precipitation, generally low relief topography, and few naturally flowing (unregulated) streams. The best fit regression models in region 7 varied by statistic: a regression equation with drainage area alone was the best fit for the high AEP statistics, and a regression equation with relief and minimum elevation was the best fit for the low AEP statistics. The use of drainage area alone produced a very poor regression fit for the low AEP statistics, and the use of different variables among statistics resulted in a discontinuity or illogical progression in flow statistics through the range of AEPs. Additionally, the use of relief and minimum elevation produced unreasonable estimates of flow statistics in extrapolation. The use of drainage area and minimum elevation across all statistics in region 7 produced reasonable regression equations and regression metrics that were similar to the metrics for “best fit” regression equations for each individual statistic.

## Uncertainty and Limitations

The accuracy of the regression models presented in this report depended on model fit and sampling errors. Model fit error (hereinafter model error) is a measure of how well the explanatory variables used in the regression explain the peak-flow statistics estimated from the annual peak-flow records used in the analysis. Sampling error is a measure of how well a select group of streamgages with finite record lengths can be used to estimate the true peak-flow statistic for a particular streamgage and depends on the number of streamgages and length of annual peak-flow records. The various regression metrics output by WREG (table 4) provided some insight into model and sampling errors for the selected regression models. The MEV and SEM are measures of the model error, in units of log and percentage, respectively. The AVP is a measure of the spread of the flow characteristics being estimated and represents the average accuracy of prediction for streamgages used in the regression. The SEP is the AVP expressed as a percentage of observed values and is a measure of the total model error, or combined model and sampling errors. The AVP and SEP provide some indication of how well the regression model will predict the peak-flow statistic at ungaged sites. Regression models with small SEMs and SEPs have greater accuracy than those with large SEMs and SEPs. The SEP provides a better overall measure of a model’s predictive power and reliability than does the SEM (Hortness and Berenbrock, 2003), in particular because the principal purpose of the regression equations is to estimate

peak-flow statistics at ungaged locations. The pseudo- $R^2$  is another measure of model fit and reliability. Pseudo- $R^2$  explains the degree to which the regression model and model error explain the variability in the data used to develop the regressions. Pseudo- $R^2$  is similar to the standard coefficient of determination,  $R^2$ , except that it removes the effect of the sampling error. Pseudo- $R^2$  can range from 0 to 100 percent; higher rather than lower values of pseudo- $R^2$  generally indicate that more variability in the measured data is explained by the regression model.

Among all regions and statistics, the final regression equations resulted in SEPs (and pseudo- $R^2$ ) that ranged from 22.8 (and 99.2) percent (region 3, 0.2-percent AEP) to 132 (and 63.7) percent (region 7, 80-percent AEP) (table 4). Overall, regression model fit was best (pseudo- $R^2 > 90$  percent; low relative SEM and SEP) for regions 4 and 5 compared to other regions. Regression model fit was poorest for region 7 compared to other regions because of the low number of available streamgages that met the criteria for inclusion in the study and poor correlations between basin and climatic characteristics and annual peak flows. Regression model fit also was relatively poor for the low AEP statistics in region 6\_8 and the high AEP statistics in region 3 (table 4).

Peak-flow estimates should not be estimated for data that are outside the range of values used to develop the regression model, where the limits of statistical confidence are unknown. Common statistical principles dictate that regression equations should not be used in extrapolation, but rather should be used only for interpolation (that is, in the range of measured values of the explanatory variables used in the regression analysis). The regression equations can be expected to have the average accuracy stated in table 4 if applied to ungaged sites that have basin or climatic characteristics in the range of explanatory variables used to develop the equations (table 5). Accuracy of statistics is unknown and the uncertainty is likely much greater than what is shown in table 4 when the equations are applied outside the range of explanatory variables. Few streamgages were available for small, headwater streams with drainage areas less than 10 mi<sup>2</sup>. As a result, uncertainty for peak-flow statistics in small basins and headwater streams generated using the regression equations is likely high. Additionally, the regression equations presented in table 4 should be applied only to naturally flowing, relatively unregulated streams in Idaho to calculate peak-flow statistics. The regression equations are intended to produce an estimate representative of average, regional conditions. Any localized effects such as small-scale, groundwater-surface water interactions like seeps, springs, and sinks are not represented by the regional peak-flow regression equations and will provide inaccurate estimates in such cases.

**Table 5.** Ranges in values of basin and climatic characteristics used to develop regional peak-flow regression equations for Idaho.

[**Abbreviations:** *A*, contributing drainage area, in square miles; *F*, percent of drainage basin with forest land cover; *P*, mean annual precipitation, in inches; *E<sub>min</sub>*, minimum elevation in basin, in feet; –, not applicable/variable not used in region]

Region	<i>A</i>		<i>F</i>		<i>P</i>		<i>E<sub>min</sub></i>	
	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
1_2	1.11	10,700	–	–	21.3	66.8	–	–
3	2.13	2,500	–	–	–	–	–	–
4	5.36	13,400	2.35	94.4	18.7	46.8	–	–
5	8.63	1,040	–	–	30.6	49.7	–	–
6_8	2.77	3,740	–	–	18.9	54.6	–	–
7	0.15	1,400	–	–	–	–	2,230	7,020

The SEPs presented in table 4 are measures of the average uncertainty of a regression equation in a region, but the SEP also can be used to estimate uncertainty for individual peak-flow estimates. The SEP for a particular peak-flow statistic can be estimated for a particular site using equation 6:

$$SEP_i = \left[ MEV_i + \left( x_i^T (X^T \Lambda^{-1} X)^{-1} x_i \right) \right]^{0.5} \quad (6)$$

where

$SEP_i$  is the standard error of prediction, in log units (table 4), for an estimate of  $Q_{AEP}$  at site  $i$ ;

$MEV_i$  is the model error variance, in log units (table 4), for the hydrologic region and  $Q_{AEP}$ ;

$x_i^T$  is the transpose of the vector  $x_i$ ;

$x_i$  is a row vector consisting of the value 1.0 in the first column followed by log transformed values of the  $p$  explanatory variables (basin and climate characteristics) used in the regression equation; and

$(X^T \Lambda^{-1} X)^{-1}$  is the covariance matrix (provided in appendix A, table A5) of the GLS-based regional peak-flow regression equation.

After the  $SEP_i$  has been estimated for a particular peak-flow statistic, it can be used to estimate a confidence interval around the statistic using equation 7:

$$CI_{i,\alpha} = \pm t_{\left(\frac{\alpha}{2}, n-p+1\right)} (SEP_i) \quad (7)$$

where

$CI_{i,\alpha}$  is the confidence interval, in log units, for a statistic at site  $i$  with a confidence level of  $\alpha$ ;

$t_{\left(\frac{\alpha}{2}, n-p+1\right)}$  is the Student's  $t$  value for a confidence level of  $100(1-\alpha)$  percent and  $(n-p+1)$  degrees of freedom, which can be obtained in tabular form from National Institute of Standards and Technology (2012); and

$n$  and  $p$  are the number of streamgages and explanatory variables, respectively, used to generate the GLS-based regression equation.

The confidence interval estimated in equation 7 can be converted to normal units of flow using equation 8:

$$10^{(\log Q_{AEP,i} - CI_{i,\alpha})} \leq \text{true } Q_{AEP,i} \leq 10^{(\log Q_{AEP,i} + CI_{i,\alpha})} \quad (8)$$

where

$\text{true } Q_{AEP,i}$  is the true AEP peak flow at site  $i$ .

### Alternative to Regional Regression Equations

If peak-flow statistics are desired at an ungaged site that is close to a streamgage on the same stream, more accurate results might be obtained by using a drainage-area ratio adjustment on the estimated statistics for the streamgage (table A2) rather than by using the regional regression equations. Additionally, the use of a drainage-area ratio adjustment method might produce more reasonable estimates than the regional regression equations when peak-flow statistics are desired at an ungaged site with basin characteristics outside the range of values used to generate the regional regression equations (table 5). The drainage-area ratio adjustment method can be used to estimate an AEP-percentage peak flow at the ungaged site ( $Q_{AEP,U}$ ) using equation 9 (Berenbrock, 2002):

$$Q_{AEP,U} = Q_{AEP,G} \left( \frac{DA_U}{DA_G} \right)^{exp_{AEP}} \tag{9}$$

where  $Q_{AEP,G}$  is the AEP peak flow for the streamgage, in cubic feet per second;

- $DA_U$  is the drainage area at the ungaged site, in square miles;
- $DA_G$  is the drainage area at the streamgage, in square miles; and
- $exp_{AEP}$  is the regression coefficient or slope for a GLS regression between the log of the AEP peak flow and the log of the drainage area.

Equation 9 should only be used in cases when the drainage-area ratio,  $DA_U/DA_G$  is between 0.5 and 1.5. For ungaged sites where the value of  $DA_U/DA_G$  is outside the range 0.5–1.5, the regional peak-flow regression equations (table 4) likely provide more accurate and reliable statistics than equation 9. The  $exp_{AEP}$  term was generated for each AEP statistic in each region (table 6), based on a GLS regression analysis between log of the AEP peak flow to the log of the drainage area in each region. Only streamgages that met the screening criteria and were used to generate the regression equations described in this report (table A1) should be used in the drainage-area ratio adjustment method to generate peak-flow statistics at ungaged locations.

**Table 6.** Regression coefficients for generalized least squares regressions relating AEP-percent peak flow to drainage area, used as exponents in the alternative drainage-area ratio exponent.

[Abbreviations:  $Q_{AEP}$  peak-flow magnitude, in cubic feet per second, for the indicated annual exceedance probability (AEP)]

$Q_{AEP}$	Regression coefficient/drainage-area ratio method relating $Q_{AEP}$ to drainage area for indicated region					
	1_2	3	4	5	6_8	7
$Q_{80}$	0.924	0.818	0.932	0.908	0.679	0.528
$Q_{67}$	0.911	0.807	0.918	0.913	0.671	0.530
$Q_{50}$	0.896	0.796	0.903	0.918	0.663	0.530
$Q_{43}$	0.890	0.792	0.897	0.919	0.659	0.531
$Q_{20}$	0.866	0.777	0.872	0.922	0.644	0.537
$Q_{10}$	0.850	0.769	0.855	0.922	0.634	0.543
$Q_4$	0.833	0.764	0.836	0.921	0.623	0.551
$Q_2$	0.822	0.762	0.824	0.919	0.616	0.556
$Q_1$	0.812	0.761	0.813	0.916	0.610	0.559
$Q_{0.5}$	0.803	0.762	0.803	0.914	0.603	0.563
$Q_{0.2}$	0.792	0.764	0.791	0.912	0.596	0.567

## Comparison of Results from Previous Studies

The regional peak-flow regression equations produced for this study provided more accurate estimates in some cases and less accurate estimates in other cases as compared to the equations previously published in Berenbrock (2002) and Hortness and Berenbrock (2003). Regression model fit was poor in regions 6\_8 and 7 and was good in regions 4 and 5 relative to other regions, both for the current study (table 4) and Berenbrock (2002) and Hortness and Berenbrock (2003). The difficulties in fitting a good regression model in regions 6\_8 and 7 were likely because of the semi-arid climate and high variability in precipitation and flows in these areas. Although an attempt was made to screen out streamgages on streams with regulation and diversion, the area encompassed by regions 6\_8 and 7 is heavily regulated, and the presence of minor diversions might have affected some annual peaks and contributed uncertainty to the regional peak-flow regression analysis. An analysis beyond the scope of this report would

be required to examine seasonal diversion records to screen out individual peaks that might have been affected, in an attempt to reduce regression model uncertainty in these areas. Similarly, localized groundwater-surface water interactions might have had an effect on peaks in these regions but are not fully understood or quantified.

The average regression model SEPs were improved (reduced) in the current study compared to Berenbrock (2002) and Hortness and Berenbrock (2003) in regions 3, 4, 5, and 6\_8 across all statistics (table 7). The SEPs for the 1-percent AEP, which is arguably the peak-flow statistic of greatest interest for infrastructure design for flood protection, were improved in the current study in regions 3, 4, and 5 and were similar (within 4–7 percent) for regions 2 and 6\_8 compared to Berenbrock (2002). Additionally, the minimum and maximum of the average SEPs were reduced in the current study for the 10-, 2-, and 1-percent AEPs compared to Quillian and Harenberg (1982) across all regions (table 8). The minimum of the average SEP for the 10-, 4-, 2-, and 1-percent AEPs across all regions was lower, but the maximum of the average SEP was higher, for the current study than in Berenbrock (2002).

**Table 7.** Selected results of regional peak-flow regression equations compared with results in Berenbrock (2002) and Hortness and Berenbrock (2003).

[**Region:** Hydrologic regions 1 and 2 and 6 and 8 were combined in the current study, so explanatory variables and SEP are repeated for the individual subregions in the combined region. **Abbreviations:** SEP, standard error of prediction, in percent; *A*, contributing drainage area, in square miles; *P*, mean annual precipitation, in inches; *E*, mean basin elevation, in feet; *F*, percent of drainage basin with forest land cover; *NF*<sub>30</sub>, percent of north-facing slopes greater than 30 percent; *E*<sub>min</sub>, minimum basin elevation, in feet; *BS*, average basin slope, in percent; *S*<sub>30</sub>, percent of slopes greater than 30 percent]

Region	Number of streamgages used in regional regression analysis		Explanatory variables used in regression equations		Average SEP of regression equations for all annual exceedance probabilities		SEP of regression equation for 1 percent annual exceedance probability	
	Current study	Berenbrock (2002) and Hortness and Berenbrock (2003)	Current study	Berenbrock (2002) and Hortness and Berenbrock (2003)	Current study	Berenbrock (2002) and Hortness and Berenbrock (2003)	Current study	Berenbrock (2002)
1	15	21	<i>A, P</i>	<i>A, E, F</i>	72.5	69.0	79.7	64.8
2	44	44	<i>A, P</i>	<sup>1</sup> <i>A, E, P</i>	72.5	69.6	79.7	74.1
3	13	26	<i>A</i>	<i>A, E</i>	63.8	66.6	34.5	55.1
4	28	60	<i>A, F, P</i>	<i>A, E</i>	46.4	67.6	37.3	56.9
5	20	46	<i>A, P</i>	<i>A, NF</i> <sub>30</sub> , <i>P</i>	30.3	47.4	34.9	48.4
6	19	31	<i>A, P</i>	<i>A, P</i>	68.7	72.7	75.1	71.8
<sup>2</sup> 7	24	28; 17	<i>A, E</i> <sub>min</sub>	<i>A, E; A</i>	103	70.8; 97.9	94.9	63.3; 66.1
8	29	60	<i>A, P</i>	<i>A, BS, S</i> <sub>30</sub>	68.7	82.4	75.1	79.9

<sup>1</sup>Berenbrock (2002) used *A* and *P* as explanatory variables for the *Q*<sub>50</sub> statistic but *A, E, and P* as explanatory variables for all other calculated statistics in region 2.

<sup>2</sup>Berenbrock (2002) and Hortness and Berenbrock (2003) separated region 7 into two regions, 7a and 7b. The first values shown are for region 7a; the second values shown are for region 7b.

**Table 8.** Average standard errors of prediction for selected peak-flow statistics estimated using regional regression equations in the current and selected previous studies in Idaho.

[**Abbreviations:** SEP, standard error of prediction, in percent;  $Q_{10}$ , peak-flow magnitude, in cubic feet per second, for an annual exceedance probability (AEP) of 10 percent or a recurrence interval of 10 years;  $Q_4$ , peak-flow magnitude, in cubic feet per second, for an AEP of 4 percent or a recurrence interval of 25 years;  $Q_2$ , peak-flow magnitude, in cubic feet per second, for an AEP of 2 percent or a recurrence interval of 50 years;  $Q_1$ , peak-flow magnitude, in cubic feet per second, for an AEP of 1 percent or a recurrence interval of 100 years; min, minimum; max, maximum]

Peak-flow frequency	Average SEP			
		Quillian and Harenberg (1982) <sup>1</sup>	Berenbrock (2002)	Current study
$Q_{10}$	min	49	45	30
	max	107	87	91
$Q_4$	min	–	46	31
	max	–	78	91
$Q_2$	min	46	47	34
	max	118	79	93
$Q_1$	min	49	48	34
	max	123	80	95

<sup>1</sup>As reported in Berenbrock (2002).

Berenbrock (2002) used 333 streamgages in his regional peak-flow regression analysis, compared to 192 streamgages used in the current study. Several streamgages used in the Berenbrock (2002) had annual peak-flow records that ended more than 30 years prior to the last year of data used in the analysis, had later been noted to be affected by diversion or regulation, or were on the same stream and provided redundant information. Overall, the more extensive streamgage screening completed in the current study compared to Berenbrock (2002) resulted in higher uncertainty in the regression equations in some regions, particularly in region 7, but the revised procedures and more stringent screening most likely resulted in a more accurate representation of naturally flowing conditions than previously published regression equations.

The explanatory variables selected for the final regression equations differed between Berenbrock (2002) and Hortness and Berenbrock (2003) and the current study for all regions except the part of region 6\_8 that was defined as region 6 in Berenbrock (2002; table 7). Drainage area was selected as an explanatory variable for all regions in all studies. Mean basin elevation frequently was selected as an explanatory variable in Berenbrock (2002) and Hortness and

Berenbrock (2003), but only the elevation-based variable “minimum basin elevation” was noted as a good fit with peak-flow statistics in one region in the current study. Quillian and Harenberg (1982) used only drainage area or drainage area and mean annual precipitation in all of their peak-flow regression equations, which more closely matched selections for explanatory variables used in the current study than in Berenbrock (2002) and Hortness and Berenbrock (2003). Regression equations developed for the current study in five regions were limited to fewer explanatory variables than in Berenbrock (2002) and Hortness and Berenbrock (2003), primarily because fewer streamgages were considered for the regression analysis. As previously discussed in section, “Regional Peak-Flow Regression Equations,” some explanatory variables used in previous regional regression studies were noted to produce unreasonable flow estimates when the equations were used in extrapolation, as described in Rea and Skinner (2009) and Wood and others (2009). Although extrapolation of regression equations is undesirable because accuracy of the resulting estimates is unknown, in practice, users of the equations often do not have other viable alternatives. Wood and others (2009) showed that simpler regression equations with fewer explanatory variables rather than regression equations with several explanatory variables, including categorical and percent-based variables, are sometimes preferred because they produce more reasonable results if the equations are applied in extrapolation.

Berenbrock (2002) also presented the method for drainage-area ratio adjustments for estimating peak-flow statistics along gaged streams, but presented only a single, average exponent for all statistics in a region, likely generated using an OLS regression analysis. For the study described in this report, GLS regression techniques were used in the drainage-ratio adjustment method (as well as in the regional regression equations) to attempt to account for correlation among streamgage records. Additionally, separate exponents were generated and provided for each AEP, based on findings in Farmer and others (2015) that using a different, best-fit exponent for each AEP improved peak-flow statistics. The exponents shown in table 6 were within about 8 percent of the average exponents presented in Berenbrock (2002), except for regions 6\_8 and 7. The exponents for regions 6\_8 and 7 in table 6 were about 18–34 percent lower than the corresponding exponents in Berenbrock (2002), likely because of different streamgages used in the analysis, different groupings of regions, the exclusion of some older annual peak-flow records, and the inclusion of more recent annual peak-flow records.

## Estimating Flow Statistics Using StreamStats

The basin and climatic characteristics, regional peak-flow regression equations (table 4), and drainage-area ratio adjustment exponents (table 6) described in this report will be integrated in the USGS StreamStats program (<http://streamstats.usgs.gov>) to allow estimation of peak-flow statistics at ungaged locations on Idaho streams. StreamStats is a Web-based GIS application that provides users an assortment of analytical tools useful for water resources planning and engineering design. StreamStats makes the process of calculating flow statistics for ungaged sites faster, more accurate, and more consistent than using manual calculation methods. StreamStats users choose locations of interest from an interactive map and easily obtain flow statistics, basin characteristics, and descriptive information. If a user selects the location of a USGS streamgage, the user can obtain available, published flow statistics for the streamgage. If a user selects an ungaged location, StreamStats will delineate the drainage-basin boundary, measure basin characteristics, and estimate flow statistics for the site based on available, published regional regression equations. If the ungaged location has basin and climatic characteristics within the range of characteristics used to develop the regional regression equations, StreamStats also will output confidence intervals and SEP (described in section, “[Uncertainty and Limitations](#)”). Ries and others (2008) provide a detailed description of the StreamStats application.

StreamStats version 4 is planned to include an “Estimate Flows Based on Similar Streamgaging Stations” tool, which will apply the drainage-area ratio adjustment method described in section, “[Alternative to Regional Regression Equations](#)”. After an ungaged point of interest is selected on a stream, the tool will search upstream and downstream along the stream network to locate nearby streamgages that have drainage areas within 0.5–1.5 times the drainage area for the ungaged site. The flow statistics and the drainage area for the streamgage with the drainage-area ratio closest to one will be retrieved. Equation 6 and the drainage-area ratio exponents provided in table 6 will then be used by the program to estimate peak-flow statistics. The user also will be able to customize the upstream and downstream streamgage selections to select any streamgage with a drainage area within the specified limits (Kernell Ries, U.S. Geological Survey, written commun., October 15, 2015). If both an upstream and a downstream site are available and selected for use in estimating flow statistics at an ungaged location, the final estimates will be determined by weighting the upstream and downstream statistics using a process explained in Ries and Dillow (2006).

## Potential Areas for Further Study

The regional regression equations presented in this report and their applicability for estimating peak-flow statistics at ungaged locations in Idaho would be improved by including more peak-flow records on headwater streams with small drainage areas (for example, smaller than 10 mi<sup>2</sup>). Peak-flow statistics often are needed in headwater streams for evaluations of aquatic habitat, flood protection for small communities, and the design of small-scale infrastructure such as culverts and bridges. Streamgages are most often installed on large streams and rivers to balance multiple needs and cost; few are installed on small, headwater streams. As an alternative, networks of crest stage gages have been installed by the USGS on headwater streams in many States to fill data gaps in peak-flow records and regional regression datasets. Crest stage gages cost much less to install and operate than a continuously recording streamgage and passively document the peak stage in a stream between servicing visits (Sauer and Turnipseed, 2010). Flow estimates associated with peak stages can be obtained from stage-discharge relations based on flow measurements at the site. The annual peak flows collected from crest stage gages can then augment records from continuously recording streamgages used in the regional peak-flow regression analysis.

Efforts have been made to extend flow records and estimate flow statistics in streams in remote areas of the Owyhee Canyonlands Wilderness (fig. 1) based on an indexing technique, which relates discrete flow measurements to recorded flows at nearby streamgages to create a long-term synthetic flow record at a site of interest. The initial work was documented in Wood and Fosness (2013) and Wood (2014); the data collection and statistical analysis for the project is ongoing (as of 2016). After enough measurements have been made at project sites over a variety of hydrologic conditions, the synthetic annual peak-flow records from the measurement sites could be included in a future regional regression update for Idaho. The inclusion of these sites would likely improve the regression equations in region 7, particularly the Intermountain Semidesert (fig. 1) part of region 7, which is sparsely gaged (fig. 3).

The study described in this report focuses on regional regression equations for estimating peak-flow statistics, but the USGS StreamStats program also includes regression equations to estimate monthly, annual, and low-flow statistics. The latter flow statistics were based on work documented in Hortness and Berenbrock (2001) and Hortness (2006) using flow data through 1999 and 1990, respectively. Seasonal drought conditions have been frequent in Idaho after the Hortness and Berenbrock (2001) and Hortness (2006) studies, particularly in

the last decade (2005–15). In fact, drought conditions in 2015 resulted in record low flows at several USGS streamgages in the Pacific Northwest (U.S. Geological Survey, 2015). An update to monthly, annual, and low-flow statistics and associated regional regression equations in StreamStats, similar to the update for peak-flow statistics described in this report, would beneficially inform water-resource management decisions in Idaho.

Recent research has shown that climate variations have caused an increase in the frequency and intensity of wildfire in some Pacific Northwest watersheds (Trouet and others, 2006; Westerling and others, 2006). Similarly, climate models predict changes in the magnitude and timing of peak flow because of changes in snowpack (Mote and others, 2005; Pederson and others, 2013). Research has not yet linked changes in wildfire with associated changes in runoff and peak annual flow in Idaho, particularly in mid- and low-elevation watersheds where snowmelt is not the major contributor to flow. In forested watersheds substantially affected by wildfire, loss of vegetation can lead to increased runoff because of decreased interception of precipitation and surface roughness (Neary and others, 2008). Additionally, wildfires can cause the upper soil layer to become hydrophobic, further reducing infiltration and watershed time of concentration (Neary and others, 2008). As a result, runoff and peak flow might occur earlier and might be of higher magnitude than in a similar watershed not affected by wildfire. A trend analysis could be completed to determine whether wildfire has a statistically significant, long-lasting effect on the timing and magnitude of peak flow, particularly at mid- and low-elevation drainage basins. Regression equations to estimate peak-flow statistics in drainage basins with frequent wildfires could be potentially improved by the inclusion of variables such as percentage of burned area in recent years, associated average burn intensity, wildfire probability, soil type, and empirical vegetation recovery rates.

## Summary

Estimates of peak-flow frequency statistics (peak-flow statistics) are needed to support water-resource management decisions for flood protection, infrastructure design, and riparian and aquatic habitat protection. Annual peak-flow records from U.S. Geological Survey streamgages can be related to basin and climate characteristics to develop regional regression equations to estimate peak-flow statistics at ungaged locations. The peak-flow statistics and regional regression equations periodically should be updated to account for new climate and flow information, improved explanatory datasets, and improved statistical techniques. The most recent, comparable studies in Idaho to estimate peak-flow statistics at ungaged locations were performed in 1982, 2002, and 2003. In 2013, the USGS, in cooperation with the Idaho Transportation Department, developed updated peak-flow statistics for

selected streamgages and regional regression equations to estimate peak-flow statistics at naturally flowing, relatively unregulated streams and rivers in Idaho.

A total of 1,215 streamgages in Idaho and surrounding areas with at least 10 years of annual peak-flow record were initially considered for use in the regional regression analysis described in this report. That number was reduced to 192 streamgages after screening for record end date, regulation and diversion, redundant or nested streamgages, and time trends inconsistent with climate patterns. Peak-flow statistics were estimated for each streamgage by fitting the record of annual peak flows to a log-Pearson type III distribution in the USGS PeakFQ program using the Expected Moments Algorithm, which allows the use of perception thresholds and flow intervals to provide more information on systematic and historical peaks as well as periods of missing data than was possible in previous studies. Additionally, the generalized Multiple Grubbs-Beck Test was used to detect and screen out potentially influential low outliers in the peak-flow frequency distribution. One of the inputs to the calculations of peak-flow statistics was the skew, which is a measure of the asymmetry and curvature of the peak-flow data distribution. For most streamgages used in the analysis, the peak-flow statistics were estimated by weighting the skew estimated from the streamgage's annual peak-flow record, called at-station skew, with a regional skew. A constant regional skew of -0.07 was estimated for the Pacific Northwest using a Bayesian-Generalized Least Squares analysis using streamgages with at least 30 years of annual peak-flow record.

The State of Idaho was divided into six regions with similar basin, climatic, and flow characteristics, based on professional judgment and previous work completed during the 2002 study. Four of the regions identified in the 2002 study were combined into two regions for the study described in this report because of a limited number of streamgages available for the analysis and because of similarities in basin characteristics in the regions which were ultimately combined. The peak-flow statistics estimated from selected streamgages initially were related to 24 basin and climatic characteristics to determine which characteristics had the best visual and statistical correlations. The final regional peak-flow regression equations were developed using a generalized least squares procedure in the USGS Weighted Multiple Linear Regression (WREG) program to estimate 11 peak-flow statistics with 80-, 67-, 50-, 43-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities (AEPs) (1.25-, 1.50-, 2.00-, 2.33-, 5.00-, 10.0-, 25.0-, 50.0-, 100-, 200-, and 500-year recurrence intervals, respectively). Of the 24 basin and climatic characteristics evaluated, only four (drainage area, mean annual precipitation, percentage of basin covered by forest, and minimum basin elevation) were used in the final equations. The final regional peak-flow regression equations were selected based on low total and model error, expressed as standard error of prediction (SEP) and standard error of the model (SEM), respectively; amount of variability in peak-flow

statistics explained by explanatory variables, expressed as the pseudo- $R^2$ ; random patterns in residuals plots; low number of data points with high leverage or influence; and consistency in model form across all estimated statistics. Regression equations for some peak-flow statistics in regions 4 and 7 included an explanatory variable that was not statistically significant in order to maintain a consistent regression model form and avoid having an illogical progression in the magnitude of flow statistics across the range of AEPs.

Overall, the SEPs for the regional peak-flow regression equations ranged from 22.8 to 132 percent. Regression model fit was best (pseudo- $R^2 > 90$  percent; low relative SEM and SEP) for regions 4 and 5 compared to other regions. Model fit was poor for region 7 compared to other regions because of the lack of available streamgages that met the criteria for inclusion in the study and the semi-arid climate, which likely caused extreme precipitation and flow variability, poor correlations between basin and climatic characteristics and annual peak flows; and as a result, poor predictions of peak-flow statistics. A drainage-area ratio adjustment method, using ratio exponents estimated using GLS regressions between peak-flow statistics and drainage area, was presented as an alternative to the regional peak-flow regression equations if estimates are desired at an ungaged location that is close to a streamgage selected for inclusion in this study. The alternative drainage-area ratio adjustment method is appropriate when the drainage-area ratio between the ungaged and gaged sites is between 0.5 and 1.5.

The regional peak-flow regression equations developed for this study had improved (reduced) SEPs compared to the 2002 and 2003 studies in regions 3, 4, 5, and 6\_8. The SEPs for the 1-percent AEP, which is arguably the peak-flow statistic of greatest interest for flood protection and infrastructure design, were improved in the current study compared to the 2002 study in regions 3, 4, and 5 and were similar (within 4–7 percent) for regions 2 and 6\_8. Additionally, the SEPs were improved (reduced) in the current study compared to the 1982 study across all regions and statistics estimated in both studies. Overall, the more extensive streamgage screening completed in the current study and different explanatory variables used as compared to the 2002 and 2003 studies resulted in higher uncertainty in the regression equations in some regions, particularly in regions 1\_2 and 7. However, the revised procedures, increased screening, and higher resolution datasets used in the current study most likely resulted in a more statistically accurate representation of naturally flowing conditions in Idaho streams.

The regression equations developed for this study will be made available in the USGS StreamStats program, a GIS-based program that allows a user to select a point of interest on a stream, delineate a drainage basin, estimate basin and climate characteristics, and estimate peak-flow statistics. The StreamStats program also allows the use of flow records and statistics from nearby streamgages to estimate peak-flow statistics at ungaged locations using the

alternative drainage-area ratio adjustment method described in this report. Regardless of method used, the StreamStats program will produce estimates of peak-flow statistics with quantifiable certainty only when used at locations with basin and climatic characteristics in the range of input variables in the regional regression equations. Additionally, both the regional peak-flow regression equations and the StreamStats program should be used to estimate flow statistics only in naturally flowing, relatively unregulated streams without substantial local influences, such as seeps, springs, or other groundwater-surface water interactions that are not characteristic of the region.

The regional regression equations presented in this report and their applicability to estimate peak-flow statistics at ungaged locations in Idaho might be improved by including more peak-flow records on headwater streams with small drainage areas. Peak-flow statistics often are needed in headwater streams for evaluations of aquatic habitat, flood protection for small communities, and the design of small-scale infrastructure such as culverts and bridges. The utility of the regression equations in headwater streams might be improved in the future through the installation of a crest-stage-gage network, which can be used to passively and cost effectively collect annual peak-flow records. Additionally, the regression equations and resulting peak-flow statistics might be improved in the future through the inclusion of synthetic flow records generated using indexing techniques, which currently are being used to estimate flow statistics in the remote Owyhee Canyonlands Wilderness. The inclusion of other considerations, such as climate trends and the effect of wildfire on the timing and magnitude of peak flows, might benefit the prediction of peak-flow statistics at ungaged locations. The peak-flow statistics presented in this report, as well as other flow statistics previously estimated for Idaho using regional regression techniques, should be re-evaluated and updated about every 10 years to account for new information and to incorporate improved statistical techniques and professional judgment.

## Acknowledgments

The authors wish to thank USGS employees Mark Mastin, Cory Angeroth, Kirk Miller, John Risley, Steven Sando, and Toby Welborn for their assistance in reviewing assumptions and peak-flow statistics generated for streamgages in States bordering Idaho. Special thanks go to Lotwick Reese, Ned Parrish, and Bryan Martin of the Idaho Transportation Department for support of the study. The authors also are grateful to Julie Kiang, Andrea Veilleux, William Farmer, and Timothy Cohn of the USGS Office of Surface Water for their leadership in statistical hydrology and their valuable guidance throughout the study.

## References Cited

- Austin, S.H., Krstolic, J.L., and Wiegand, Ute, 2011, Peak-flow characteristics of Virginia streams: U.S. Geological Survey Scientific Investigations Report 2011–5144, 106 p., + 3 tables and 2 appendixes on CD, accessed June 3, 2015, at <http://pubs.usgs.gov/sir/2011/5144/>.
- Berenbrock, Charles, 2002, Estimating the magnitude of peak flows at selected recurrence intervals for streams in Idaho: U.S. Geological Survey Water-Resources Investigations Report 02–4170, 59 p., accessed October 2, 2013, at <http://pubs.usgs.gov/wri/2002/4170/>.
- British Columbia Ministry of Environment, 2014, Ecosystems branch—Soils: British Columbia Ministry of Environment soils Web page, accessed April 14, 2014, at <http://www.env.gov.bc.ca/soils/>.
- Castro, J.M., and Jackson, P.L., 2001, Bankfull discharge recurrence intervals and regional hydraulic geometry relationships, patterns in the Pacific Northwest, USA: *Journal of the American Water Resources Association*, v. 37, no. 5., accessed January 22, 2015, at <http://dx.doi.org/10.1111/j.1752-1688.2001.tb03636.x>.
- Clark, G.M., 2010, Changes in patterns of streamflow from unregulated watersheds in Idaho, Western Wyoming, and Northern Nevada: *Journal of the American Water Resources Association*, v. 46, p. 486–497, accessed September 10, 2014, at <http://dx.doi.org/10.1111/j.1752-1688.2009.00416.x>.
- Cohn, T.A., England, J.F., Berenbrock, C.E., Mason, R.R., Stedinger, J.R., and Lamontagne, J.R., 2013, A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series: *Water Resources Research*, v. 49, p. 5,047–5,058, accessed October 13, 2014, at <http://dx.doi.org/10.1002/wrcr.20392>.
- Cohn, T.A., Lane, W.L., and Baier, W.G., 1997, An algorithm for computing moments-based flood quantile estimates when historical flood information is available: *Water Resources Research*, v. 33, no. 9, p. 2,089–2,096, accessed February 26, 2015, at <http://dx.doi.org/10.1029/97WR01640>.
- Cohn, T.A., Lane, W.L., and Stedinger, J.R., 2001, Confidence intervals for expected moments algorithm flood quantile estimates: *Water Resources Research*, v. 37, no. 6, p. 1,695–1,706, accessed October 20, 2014, at <http://dx.doi.org/10.1029/2001WR900016>.
- Eash, D.A., Barnes, K.K., and Veilleux, A.G., 2013, Methods for estimating annual exceedance probability discharges for streams in Iowa, based on data through water year 2010: U.S. Geological Survey Scientific Investigations Report 2013–5086, 63 p. with appendix, accessed January 13, 2016, at <http://pubs.usgs.gov/sir/2013/5086/>.
- Emmett, W.W., 1975, The channels and waters of the upper Salmon River area, Idaho: U.S. Geological Survey Professional Paper 870-A, 115 p., accessed October 20, 2014, at <http://pubs.usgs.gov/pp/0870a/report.pdf>.
- Eng, Ken, Chen, Yin-Yu, and Kiang, J.E., 2009, User's guide to the weighted-multiple-linear-regression program (WREG version 1.0): U.S. Geological Survey Techniques and Methods, book 4, chap. A8, 21 p., accessed January 15, 2014, at <http://pubs.usgs.gov/tm/tm4a8>.
- England, J.F., and Cohn, T.A., 2008, Bulletin 17B flood frequency revisions—Practical software and test comparison results: World Environmental and Water Resources Congress 2008, Ahupua'a, Honolulu, Hawai'i, May 12–16, 2008, 11 p., accessed January 17, 2014, at [http://dx.doi.org/10.1061/40976\(316\)559](http://dx.doi.org/10.1061/40976(316)559).
- England, J.F., Salas, J.D., and Jarrett, R.D., 2003, Comparisons of two moments-based estimators that utilize historical and paleoflood data for the log Pearson type III distribution: *Water Resources Research*, v. 39, no. 9, p. 1,243, accessed January 17, 2014, at <http://dx.doi.org/10.1029/2002WR001791>.
- Environmental Systems Resource Institute (ESRI), 2013, ArcGIS Desktop, Release 10: Redlands, California, <http://www.esri.com/software/arcgis>.
- Falcone, J.A., 2011, GAGES-II—Geospatial attributes of gages for evaluation streamflow: U.S. Geological Survey [digital spatial dataset], accessed October 15, 2014, at <https://pubs.er.usgs.gov/publication/70046617>.
- Farmer, W.H., Over, T.M., and Vogel, R.M., 2015, Multiple regression and inverse moments improve the characterization of the spatial scaling behavior of daily streamflows in the Southeast United States: *Water Resources Research*, v. 51, no. 3, p. 1,775–1,796, accessed December 13, 2015, at <http://dx.doi.org/10.1002/2014WR015924>.
- Fenneman, N.M., 1946, Physical divisions of the United States: Washington, D.C., U.S. Geological Survey special map, scale 1:7,000,000.
- Flynn, K.M., Kirby, W.H., and Hummel, P.R., 2006, User's manual for program PeakFQ annual flood-frequency analysis using Bulletin 17B guidelines: U.S. Geological Survey Techniques and Methods, book 4, chap. B4, 42 p., accessed June 10, 2015, at <http://pubs.usgs.gov/tm/2006/tm4b4/>.
- Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., and Tyler, D., 2002, The National elevation dataset: *Photogrammetric Engineering & Remote Sensing*, v. 68, no. 1, p. 5–11, accessed May 15, 2015, at <http://www.asprs.org/a/publications/pers/2002journal/january/highlight.html>.

- Gotvald, A.J., Barth, N.A., Veilleux, A.G., and Parrett, Charles, 2012, Methods for determining magnitude and frequency of floods in California, based on data through water year 2006: U.S. Geological Survey Scientific Investigations Report 2012–5113, 38 p., 1 pl., accessed April 14, 2015, at <http://pubs.usgs.gov/sir/2012/5113/>.
- Griffis, V.W., and Stedinger, J.R., 2007, The use of GLS regression in regional hydrologic analyses: *Journal of Hydrology*, v. 344, p. 82–95, accessed July 16, 2015, at <http://dx.doi.org/10.1016/j.jhydrol.2007.06.023>.
- Grubbs, F.E., and Beck, Glenn, 1972, Extension of sample sizes and percentage points for significance tests of outlying observations: *Technometrics*, v. 14, no. 4, p. 847–854.
- Gruber, A.M., and Stedinger, J.R., 2008, Models of LP3 regional skew, data selection and Bayesian GLS regression, Paper 596, in Babcock, R.W., and Watson, R., eds., *World Environmental and Water Resources Congress—Ahupua‘a*, Honolulu, Hawai‘i, May 12–16, 2008: American Society of Civil Engineers, paper 563, 10 p., accessed July 25, 2015, at [http://dx.doi.org/10.1061/40976\(316\)563](http://dx.doi.org/10.1061/40976(316)563).
- Hardison, C.H., 1974, Generalized skew coefficients of annual floods in the United States and their application: *Water Resources Research*, v. 10, no. 4, p. 745–752, accessed October 13, 2014, at <http://dx.doi.org/10.1029/WR010i004p00745>.
- Harenberg, W.A., 1980, Using channel geometry to estimate flood flows at ungauged sites in Idaho: U.S. Geological Survey Water-Resources Investigations Report 80–32, 39 p., accessed June 17, 2015, at <https://pubs.er.usgs.gov/publication/wri8032>.
- Harrell, Frank, 2015, *Regression modeling strategies* (2d ed.): Springer International Publishing, 2,582 p.
- Hedman, E.R., and Osterkamp, W.R., 1982, Streamflow characteristics related to channel geometry of streams in Western United States: U.S. Geological Survey Water-Supply Paper 2193, 17 p., accessed January 13, 2015, at <https://pubs.er.usgs.gov/publication/wsp2193>.
- Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 522 p., accessed June 12, 2015, at <http://pubs.usgs.gov/twri/twri4a3/>.
- Holmes, R.R., Jr., and Dinicola, K., 2010, 100-Year flood—It’s all about chance: U.S. Geological Survey General Information Product 106, 1 p., accessed January 7, 2015, at <http://pubs.usgs.gov/gip/106/>.
- Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., 2015, Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information: *Photogrammetric Engineering and Remote Sensing*, v. 81, no. 5, p. 345–354, accessed April 10, 2015, at [http://www.asprs.org/a/publications/pers/2015journals/PERS\\_May\\_2015/HTML/index.html#345/z](http://www.asprs.org/a/publications/pers/2015journals/PERS_May_2015/HTML/index.html#345/z).
- Hortness, J.E., 2006, Estimating low-flow frequency statistics for unregulated streams in Idaho: U.S. Geological Survey Scientific Investigations Report 2006–5035, 40 p., accessed October 3, 2014, at <http://pubs.usgs.gov/sir/2006/5035/>.
- Hortness, J.E., and Berenbrock, Charles, 2001, Estimating monthly and annual streamflow statistics at ungauged sites in Idaho: U.S. Geological Survey Water-Resources Investigations Report 2001–4093, 36 p., accessed October 3, 2014, at <https://pubs.er.usgs.gov/publication/wri014093>.
- Hortness, J.E., and Berenbrock, Charles, 2003, Estimating the magnitude of bankfull flows for streams in Idaho: U.S. Geological Survey Water-Resources Investigations Report 03–4261, 37 p., accessed October 3, 2014, at <http://pubs.er.usgs.gov/publication/wri034261>.
- Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency: *Hydrology Subcommittee Bulletin 17B*, 28 p., 14 app., 1 pl., accessed October 3, 2014, at [http://water.usgs.gov/osw/bulletin17b/bulletin\\_17B.html](http://water.usgs.gov/osw/bulletin17b/bulletin_17B.html).
- Kjelstrom, L.C., and Moffatt, R.L., 1981, Method of estimating flood-frequency parameters for streams in Idaho: U.S. Geological Survey Open-File Report 81–909, 99 p., accessed October 3, 2014, at <https://pubs.er.usgs.gov/publication/ofr81909>.
- Leopold, L.B., 1994, *A view of the river*: Cambridge, Mass., Harvard University Press, 298 p.
- Marshall, I.B., Schut, P.H., and Ballard, M., 1999, A national ecological framework for Canada—Attribute data, agriculture and Agri-Food Canada: Research Branch, Centre for Land and Biological Resources Research, and Environment Canada, State of the Environment Directorate, Ecozone Analysis Branch, Ottawa/Hull, accessed April 22, 2014, at [http://sis.agr.gc.ca/cansis/nsdb/ecostrat/1999report/data\\_tables.html](http://sis.agr.gc.ca/cansis/nsdb/ecostrat/1999report/data_tables.html).
- McNab, W.H., Cleland, D.T., Freeouf, J.A., Keys, Jr., J.E., Nowacki, G.J., and Carpenter, C.A., 2007, Description of ecological subregions—Sections of the conterminous United States: General Technical Report WO-76B, Washington, D.C., U.S. Department of Agriculture, Forest Service, 80 p., accessed May 15, 2015, at [http://www.edc.uri.edu/ATMT-DSS/report\\_forecast/landscape\\_dynamics/SectionDescriptions.pdf](http://www.edc.uri.edu/ATMT-DSS/report_forecast/landscape_dynamics/SectionDescriptions.pdf).

- Mote, P.W., Hamlet, A.F., Clark, M.P., and Lettenmaier, D.P., 2005, Declining mountain snowpack in western North America: *Bulletin of the American Meteorological Society*, v. 86, no. 1, p. 39–49, accessed January 6, 2016, at <http://dx.doi.org/10.1175/BAMS-86-1-39>.
- National Institute of Standards and Technology, 2012, e-Handbook of statistical methods, April 2012 update: Web page, accessed July 20, 2015, at <http://itl.nist.gov/div898/handbook/eda/section3/eda3672.htm>.
- National Oceanic and Atmospheric Administration, 2007, National Weather Service Hydrometeorological Design Studies Center, precipitation frequency data server: Web page, accessed April 3, 2014, at <http://www.nws.noaa.gov/oh/hdsc/>.
- Natural Resources Canada, 2014, Regional, national and international climate modeling: Climatic datasets Web page, accessed April 22, 2014, at [http://cfs.nrcan.gc.ca/projects/3?lang=en\\_CA](http://cfs.nrcan.gc.ca/projects/3?lang=en_CA).
- Natural Resources Conservation Service, 2007, Hydrologic soil groups, chapter 7 of part 630 hydrology, national engineering handbook: U.S. Department of Agriculture, accessed January 6, 2016, at <http://directives.sc.egov.usda.gov/OpenNonWebContent.aspx?content=22526.wba>.
- Nearby, D.G., Ryan, K.C., DeBano, L.F., eds., 2008, Wildland fire in ecosystems—Effects on fire on soils and water: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, General Technical Report RMRS-GTR-42, v. 4, 250 p., accessed January 10, 2016, at [http://www.fs.fed.us/rm/pubs/rmrs\\_gtr042\\_4.pdf](http://www.fs.fed.us/rm/pubs/rmrs_gtr042_4.pdf).
- Number Crunching Statistical Software (NCSS), 2015, NCSS version 10: NCSS, LLC, Kaysville, Utah, [www.ncss.com/software/ncss](http://www.ncss.com/software/ncss).
- Parrett, C., 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 2003–4308, 101 p., accessed June 22, 2015, at <http://pubs.usgs.gov/wri/wri03-4308/>.
- Parrett, C., Veilleux, A., Stedinger, J.R., Barth, N.A., Knifong, D.L., and Ferris, J.C., 2011, Regional skew for California, and flood frequency for selected sites in the Sacramento-San Joaquin River Basin, based on data through water year 2006: U.S. Geological Survey Scientific Investigations Report 2010–5260, 94 p., accessed July 14, 2015, at <http://pubs.usgs.gov/sir/2010/5260/>.
- Paretti, N.V., Kennedy, J.R., and Cohn, T.A., 2014, Evaluation of the expected moments algorithm and a multiple low-outlier test for flood frequency analysis at streamgaging stations in Arizona: U.S. Geological Survey Scientific Investigations Report 2014–5026, 61 p., accessed July 14, 2015, at <http://dx.doi.org/10.3133/sir20145026>.
- Pederson, G.T., Betancourt, J.L., and McCabe, G.J., 2013, Regional patterns and proximal causes of the recent snowpack decline in the Rocky Mountains, U.S.: *Geophysical Research Letters*, v. 40, p. 1,811–1,816, accessed January 7, 2016, at <http://dx.doi.org/10.1002/grl.50424>.
- PRISM Climate Group, 2007, PRISM climate data: Oregon State University database, accessed April 22, 2014, at <http://prism.oregonstate.edu/products/matrix.phtml>.
- Quillian, E.W., and Harenberg, W.A., 1982, An evaluation of Idaho stream-gaging networks: U.S. Geological Survey Open-File Report 82–865, 61 p., accessed October 20, 2015, at <http://pubs.usgs.gov/of/1982/0865/report.pdfhttps://pubs.er.usgs.gov/publication/ofr82865>.
- R Core Team, 2015, R—A language and environment for statistical computing, version 3.2.2: R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>.
- Rea, A., and Skinner, K.D., 2009, Estimating perennial streams of Idaho and related geospatial datasets: U.S. Geological Survey Data Series 412, 32 p., accessed October 3, 2014, at <http://pubs.usgs.gov/ds/412/>.
- Ries K.G., III, and Dillow, J.J.A., 2006, Magnitude and frequency of floods on nontidal streams in Delaware: U.S. Geological Survey Scientific Investigations Report 2006–5146, p. 31, accessed July 27, 2015, at <http://pubs.usgs.gov/sir/2006/5146/>.
- Ries, K.G., III, Guthrie, J.G., Rea, A.H., Steeves, P.A., and Stewart, D.W., 2008, StreamStats—A water resources Web application: U.S. Geological Survey Fact Sheet 2008–3067, 6 p., accessed October 14, 2015, at <http://pubs.usgs.gov/fs/2008/3067/>.
- Ryberg, K.R., 2008, PFRReports—A program for systematic checking of annual peaks in NWISWeb: U.S. Geological Survey Open-File Report 2008–1284, 18 p., accessed January 12, 2015, at <http://pubs.usgs.gov/of/2008/1284/>.
- 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., accessed November 14, 2015, at <http://dx.doi.org/10.3133/sir20155019C>.
- Sauer, V.B., and Turnipseed, D.P., 2010, Stage measurement at gaging stations: U.S. Geological Survey Techniques and Methods, book 3, chap. A7, 45 p., accessed April 14, 2015, at <http://pubs.usgs.gov/tm/tm3-a7/>.

- Stedinger, J.R., and Griffis, V.W., 2008, Flood frequency analysis in the United States—Time to update: *Journal of Hydrologic Engineering*, v. 13, no. 4, p. 199–204, accessed October 1, 2014, at [http://dx.doi.org/10.1061/\(ASCE\)1084-0699\(2008\)13:4\(199\)](http://dx.doi.org/10.1061/(ASCE)1084-0699(2008)13:4(199)).
- Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis 1—Ordinary, weighted, and generalized least squares compared: *Water Resources Research*, v. 21, no. 9, p. 1,421–1,432, accessed March 21, 2015, at <http://dx.doi.org/10.1029/WR021i009p01421>.
- Tasker, G.D., and Slade, R.M., Jr., 1994, An interactive regional regression approach to estimating flood quantiles, *in* Fontane, D.G., and Tuvel, H.N., *Water policy and management, solving the problems—Proceedings of the 21st Annual Conference of the Water Resources Planning and Management Division: American Society of Civil Engineers*, p. 782–785.
- Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: *Journal of Hydrology*, v. 111, p. 361–375, accessed March 21, 2015, at [http://dx.doi.org/10.1016/0022-1694\(89\)90268-0](http://dx.doi.org/10.1016/0022-1694(89)90268-0).
- Thomas, B.E., Hjalmarson, H.W., and Waltemeyer, S.D., 1994, Methods for estimating magnitude and frequency of floods in the Southwestern United States: U.S. Geological Survey Open-File Report 93–419, 211 p., accessed July 30, 2015, at <http://pubs.usgs.gov/of/1993/0419/report.pdf>.
- Thomas, C.A., Harenberg, W.A., and Anderson, J.M., 1973, Magnitude and frequency of floods in small drainage basins in Idaho: U.S. Geological Survey Water-Resources Investigations 7–73, 61 p., accessed July 30, 2015, at <http://pubs.usgs.gov/wri/1973/0007/report.pdf>.
- Trouet, V., Taylor, A.A., Carleton, A.M., and Skinner, C.M., 2006, Fire-climate interactions in forests of the American Pacific coast: *Geophysical Research Letters*, v. 33, L18704, accessed January 7, 2016, at <http://dx.doi.org/10.1029/2006GL027502>.
- U.S. Department of Agriculture, 2014, U.S. general soil map (STATSGO2): Web Soil Survey, Prepared by the soil survey staff of the Natural Resources Conservation Service, accessed April 13, 2014, at <http://soildatamart.nrcs.usda.gov>.
- U.S. Department of Commerce, 1999, U.S. snow climatology: Dataset documentation TD-9641M, Asheville, N.C., U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Climatic Data Center Web page, accessed April 22, 2014, at <http://www.ncdc.noaa.gov/>.
- U.S. Geological Survey, 2012, The StreamStats program: U.S. Geological Survey StreamStats Program Web application, accessed September 12, 2015, at <http://streamstats.usgs.gov>.
- U.S. Geological Survey, 2014, National Elevation Dataset: U.S. Geological Survey dataset, accessed October 3, 2014, at <http://ned.usgs.gov/>.
- U.S. Geological Survey, 2015, National Water Information System: U.S. Geological Survey database, accessed April 20, 2015, at <http://waterdata.usgs.gov/nwis/>.
- Veilleux, A.G., 2011, Bayesian GLS regression, leverage and influence for regionalization of hydrologic statistics: Ithaca, New York, Cornell University, Ph.D. dissertation, 184 p.
- Veilleux, A.G., Cohn, T.A., Flynn, K.M., Mason, R.R., Jr., and Hummel, P.R., 2014, Estimating magnitude and frequency of floods using the PeakFQ 7.0 program: U.S. Geological Survey Fact Sheet 2013–3108, 2 p., accessed November 2, 2014, at <http://dx.doi.org/10.3133/fs20133108>.
- Veilleux, A.G., Stedinger, J.R., and Eash, D.A., 2012, Bayesian WLS/GLS regression for regional skewness analysis for regions with large crest stage gage networks, *in* Loucks, E.D., ed., *Proceedings World Environmental and Water Resources Congress 2012—Crossing boundaries*, Albuquerque, N. Mex., May 20–24: American Society of Civil Engineers, p. 2,253–2,263, accessed August 5, 2015, at <http://ia.water.usgs.gov/media/pdf/report/Veilleux-Stedinger-Eash-EWRI-2012-227R.pdf>.
- Westerling, A.L., Hidalgo, H.G., Cayan, D.R., and Swetnam, T.W., 2006, Warming and earlier spring increase western U.S. forest wildfire activity: *Science*, v. 313, p. 940–943, accessed December 2, 2015, at <http://dx.doi.org/10.1126/science.1128834>.
- Wood, M.S., 2014, Streamflow statistics for development of water rights claims for the Jarbidge Wild and Scenic River, Owyhee Canyonlands Wilderness, Idaho, 2013–14—A supplement to Scientific Investigations Report 2013–5212: U.S. Geological Survey Scientific Investigations Report 2014–5143, 14 p., accessed October 14, 2015, at <http://dx.doi.org/10.3133/sir20145143>.
- Wood, M.S., and Fosness, R.L., 2013, Streamflow monitoring and statistics for development of water rights claims for Wild and Scenic Rivers, Owyhee Canyonlands Wilderness, Idaho, 2012: U.S. Geological Survey Scientific Investigations Report 2013–5212, 66 p., accessed October 14, 2015, at <http://dx.doi.org/10.3133/sir20135212>.
- Wood, M.S., Rea, A., Skinner, K.D., and Hortness, J.E., 2009, Estimating locations of perennial streams in Idaho using a generalized least-squares regression model of 7-day, 2-year low flows: U.S. Geological Survey Scientific Investigations Report 2009–5015, 27 p., accessed November 3, 2014, at <http://pubs.usgs.gov/sir/2009/5015/>.



## Appendix A. Supplemental Information for the Development of Regional Peak-Flow Regression Equations in Idaho

Appendix A tables A1–A5 are available for download at <http://dx.doi.org/10.3133/20165083>.

**Table A1.** Description of streamgages and peak-flow records used in developing regional peak-flow regression equations for Idaho.

**Table A2.** Peak-flow statistics and basin and climatic characteristics for streamgages used in developing regional peak-flow regression equations for Idaho.

**Table A3.** Expected Moments Algorithm perception threshold settings for streamgages used in developing regional peak-flow regression equations for Idaho.

**Table A4.** Expected Moments Algorithm flow interval settings for streamgages used in developing regional peak-flow regression equations for Idaho.

**Table A5.** Covariance matrices in matrix form  $([X^T \Lambda^{-1} X]^{-1})$  for generalized least squares regional peak-flow regression equations in Idaho.



# Appendix B. Regional Skew Regression Analysis for the Pacific Northwest, including Idaho, Montana, Oregon, and Washington

By Andrea G. Veilleux

## Introduction to Statistical Analysis of Regional Skew

A regional-scale skew study was conducted for an area that included Idaho, Oregon, Washington, and a part of Montana, known hereinafter as the Pacific Northwest study area or (PNW), and was based on a final list of 290 streamgages using annual peak-flow records through water year 2012. For the log-transformation of annual peak flows, Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) recommends using a weighted average of the at-station skew coefficient and a regional skew coefficient to help improve estimates of annual exceedance probability (AEP) flows (equation 3). Bulletin 17B supplies a national map, but also recommends hydrologists to develop more specific local relations. Since the first map was published in 1976, some 40 years of additional information has accumulated, and better spatial estimation procedures have been developed (Stedinger and Griffis, 2008).

Tasker and Stedinger (1986) developed a weighted least-squares (WLS) procedure for estimating regional skew coefficients based on sample skew coefficients for the logarithms (logs) of annual peak-flow data. Their method of regional analysis of skew estimators accounts for the precision of the skew-coefficient estimate for each streamgage or site, which depends on the length of record for each streamgage and the accuracy of an ordinary least-squares (OLS) regional mean skew. More recently, Reis and others (2005), Gruber and others (2007), and Gruber and Stedinger (2008) developed a Bayesian generalized least-squares (GLS) regression model for regional skew analyses. The Bayesian methodology allows for the computation of a posterior distribution of both the regression parameters and the model error variance. As shown in Reis and others (2005), for cases in which the model error variance is small compared to the sampling error of the site estimates, the Bayesian posterior distribution provides a more reasonable description of the model error variance than both the GLS method-of-moments and maximum likelihood point estimates (Veilleux, 2011). Whereas WLS regression accounts for the precision of the regional model and the effect of the record length on the variance of skew-coefficient estimators, GLS regression also considers the cross-correlations among the skew-coefficient estimators. In some studies the cross-correlations have had a large impact on the precision attributed to different parameter estimates (Feaster and others, 2009; Gotvald and others, 2009; Weaver and others, 2009; Parrett and others, 2011).

Owing to complications introduced by the use of the Expected Moments Algorithm (EMA) with Multiple Grubbs-Beck (MGB) censoring of low outliers (see Cohn and others, 1997) and large cross-correlations between annual peak flows at pairs of streamgages, an alternate regression procedure was developed to provide both stable and defensible results for regional skew (Veilleux, 2011; Lamontange and others, 2012; Veilleux and others, 2012). This alternate procedure is referred to as the Bayesian WLS (B-WLS)/Bayesian GLS (B-GLS) regression framework (Veilleux, 2011; Veilleux and others, 2011; Veilleux and others, 2012). The procedure uses an OLS analysis to fit an initial regional skew model; that OLS model is then used to generate a stable regional skew-coefficient estimate for each site. That stable regional estimate is the basis for computing the variance of each at-station skew-coefficient estimator used in the WLS analysis. Then B-WLS is used to generate estimators of the regional skew-coefficient model parameters. Finally, B-GLS is used to estimate the precision of those WLS parameter estimators, to estimate the model error variance and the precision of that variance estimator, and to estimate various diagnostic statistics.

The PNW regional skew study described here used the EMA with MGB (EMAw/MGB) to estimate the at-station skew and its mean square error. Because EMAw/MGB allows for the censoring of potentially influential low floods (PILFs), as well as the use of estimated interval flows for missing, censored, and historical data, it complicates the calculations of effective record length (and effective concurrent record length) used to describe the precision of sample estimators because the peak flows are no longer solely represented by single values. To properly account for these complications, the new B-WLS/B-GLS procedure was used.

## Methodology for Regional Skew Model

This section provides a brief description of the B-WLS/B-GLS methodology, as it appears in Veilleux (2011); and Veilleux and others (2011, 2012) provide a more detailed description.

## Ordinary Least-Squares Analysis

The first step in the B-WLS/B-GLS regional skew analysis is the estimation of a regional skew model using OLS. The OLS regional regression yields parameters  $\beta_{OLS}$  and a model that can be used to generate unbiased and relatively stable regional estimates of the skew for all streamgages:

$$\tilde{\mathbf{y}}_{OLS} = \mathbf{X}\hat{\boldsymbol{\beta}}_{OLS} \quad (\text{B1})$$

where

$\mathbf{X}$  is an  $(n \times k)$  matrix of basin characteristics;  
 $\tilde{\mathbf{y}}_{OLS}$  are the estimated regional skew values;  
 $n$  is the number of streamgages; and  
 $k$  is the number of basin parameters including a column of ones to estimate the constant.

These estimated regional skew values  $\tilde{\mathbf{y}}_{OLS}$  are then used to estimate unbiased station-regional skew variances using the equations reported in Griffis and Stedinger (2009). These station-regional skew variances are based on the regional OLS estimator of the skew coefficient instead of the at-station skew estimator, thus making the weights in the subsequent steps relatively independent of the at-station skew estimates.

## Weighted Least-Squares Analysis

A B-WLS analysis is used to develop estimators of the regression coefficients for each regional skew model (Veilleux, 2011; Veilleux and others, 2011). The WLS analysis explicitly reflects variations in record length, but intentionally neglects cross correlations thereby avoiding the problems experienced with GLS parameter estimators (Veilleux, 2011; Veilleux and others, 2011).

## Generalized Least-Squares Analysis

After the regression model coefficients,  $\hat{\boldsymbol{\beta}}_{WLS}$ , are determined with a WLS analysis, the precision of the fitted model and the precision of the regression coefficients are estimated using a B-GLS analysis (Veilleux, 2011; Veilleux and others, 2011). Precision metrics include the standard error of the regression parameters,  $SE(\hat{\boldsymbol{\beta}}_{WLS})$ , the model error variance,  $\sigma_{\delta, B-GLS}^2$ , pseudo coefficient of determination, pseudo- $R_{\delta}^2$ , and the average variance of prediction at a streamgage that is not used in the regional model,  $AVP_{new}$ .

## Data Analysis

This PNW regional skew study is based on annual peak-flow data from streamgages in Idaho, Oregon, and Washington, as well as from a part of the neighboring State of Montana. A list of 461 streamgages was originally considered for the analysis, but was pared down to a final list of 290 streamgages after excluding 171 because of insufficient record length or redundancy. The annual peak-flow data through September 2012 were downloaded from the USGS National Water Information System (NWIS; U.S. Geological Survey, 2015) database. In addition to the peak-flow data, 10 basin characteristics for each of the originally-considered 461 streamgages were evaluated as explanatory variables in

the regional study. The basin characteristics available include hydrologic unit codes (HUCs, hydrologic regions as defined in Seaber and others [1987]), as well as drainage area, mean basin elevation, minimum basin elevation, maximum basin elevation, basin relief, mean annual precipitation, mean minimum January temperature, mean 24-hour 100-year precipitation intensity, and mean 6-hour 100-year precipitation intensity. The part of Idaho omitted from the regional peak-flow regression analysis in the Snake River plain (region 0 in figure 3) also was omitted from the regional skew analysis because of the extent of regulation, groundwater-surface water interactions, and infiltration in the area.

## Station Skew Estimators

To estimate the log-transformed at-station skew coefficient,  $G$ , and its mean square error,  $MSE_G$ , the skew study used the results of the EMaW/MGB analysis described in section, “Expected Moments Algorithm Frequency Analysis and Multiple Grubbs-Beck Test for Detecting Low Outliers” in the body of this report (Cohn and others, 1997; Griffis and others, 2004). EMa provides a straightforward and efficient method for the incorporation of historical information and censored data, such as those from a crest-stage gage, contained in the record of annual peak flows for a streamgage. For this analysis PeakFQ version 7.1 (Veilleux and others, 2014), which combines EMa with MGB, was used. Documentation for PeakFQ is available at <http://water.usgs.gov/software/PeakFQ/>. PeakFQ was used to generate the site log estimates of  $G$  and the corresponding  $MSE_G$ , assuming a log-Pearson Type III distribution and generally using a Multiple Grubbs-Beck Test for PILF screening. EMa estimates, based on annual peak-flow data through September 30, 2012, of  $G$  and  $MSE_G$  are listed in table B1 for the 461 streamgages evaluated for the PNW regional skew study (see section “Expected Moments Algorithm Frequency Analysis and Multiple Grubbs-Beck Test for Detecting Low Outliers” in the main part of this report for more detail regarding EMa and MGB).

## Pseudo Record Length

Because the data set includes censored data and historical information, the effective record length used to estimate the precision of the skew estimators is no longer simply the number of annual peak flows at a streamgage. Instead, a more complex calculation was used to take into account the availability of historical information and censored values. Although historical information and censored peaks provide valuable information, they often provide less information than an equal number of years with systematically recorded peaks (Stedinger and Cohn, 1986). The following calculations provide a pseudo record length,  $P_{RL}$ , associated with skew, which appropriately accounts for all peak-flow data types available for a site.

**Table B1.** Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.

[Analysis based on data through water year 2012. **Abbreviations:** USGS, U.S. Geological Survey; No., number; MSE, mean-square error; no-R, streamgage not used in regional skew analysis due to redundancy; no- P, streamgage not used in regional skew analysis due to skew psuedo record length less than 35 years]

Regional skew index number	USGS site No.	Skew pseudo record length ( $P_{RL}$ )	Basin centroid latitude (decimal degrees)	Basin centroid longitude (decimal degrees)	Station skew coefficient (log units)	MSE of station skew coefficient (log units)	Streamgage used in regional skew study?
25	12039005	10	47.142	-123.892	-0.59	0.52	no-P
8	11495800	19	42.561	-120.870	-0.61	0.32	no-P
209	13046995	19	44.216	-110.925	0.52	0.31	no-P
211	13047600	19	44.209	-110.940	0.63	0.32	no-P
274	13310199	19	44.807	-114.963	-0.94	0.37	no-P
385	14188610	19	44.613	-122.451	0.66	0.32	no-P
394	14198400	19	44.964	-122.374	0.66	0.32	no-P
397	14201300	19	45.075	-122.764	0.28	0.28	no-P
91	12155300	20	48.053	-122.125	-0.34	0.28	no-P
124	12331500	20	46.403	-113.281	-0.73	0.31	no-P
125	12331800	20	46.363	-112.886	-0.02	0.25	no-P
180	12431500	20	47.975	-117.266	-0.82	0.33	no-P
223	13092747	20	42.351	-114.342	-0.28	0.27	no-P
268	13305310	20	44.769	-113.486	-0.33	0.27	no-P
353	14138720	20	45.457	-121.835	0.03	0.25	no-P
155	12388200	21	47.179	-114.028	-0.08	0.24	no-P
186	12448000	21	48.741	-120.120	-0.29	0.26	no-P
248	13200500	21	43.661	-116.049	1.05	0.40	no-P
269	13306385	21	45.232	-114.138	0.08	0.24	no-P
43	12080010	22	46.945	-122.742	-0.61	0.28	no-P
74	12137290	22	47.954	-121.911	0.41	0.25	no-P
184	12447383	22	48.694	-120.520	-0.40	0.26	no-P
289	13335050	23	46.231	-117.279	0.48	0.26	no-P
112	12323240	24	45.901	-112.507	-0.49	0.25	no-P
114	12323600	24	45.989	-112.623	0.12	0.22	no-P
117	12323800	24	46.055	-112.796	-0.88	0.29	no-P
160	12392155	24	48.274	-116.137	1.08	0.36	no-P
161	12392300	24	48.514	-116.612	0.64	0.27	no-P
1	10370000	25	42.240	-120.216	1.18	0.40	no-P
23	12035450	25	47.419	-123.599	-0.27	0.22	no-P
171	12413150	25	47.482	-115.829	-0.07	0.20	no-P
173	12413470	25	47.485	-116.028	-0.03	0.20	no-P
207	12512500	25	46.883	-118.751	-0.61	0.26	no-P
296	13340500	25	46.671	-115.210	-0.10	0.21	no-P
349	14127000	25	45.921	-121.934	-0.16	0.21	no-P
393	14197000	25	45.372	-123.347	0.77	0.27	no-P
40	12070000	26	47.573	-122.562	0.46	0.23	no-P
135	12350250	26	45.978	-114.136	-1.31	0.42	no-P
143	12361000	26	47.974	-113.727	0.59	0.24	no-P
172	12413210	26	47.491	-115.917	0.78	0.27	no-P
214	13055198	26	43.785	-111.169	-0.18	0.21	no-P
246	13196500	26	43.783	-115.758	0.00	0.19	no-P
280	13323500	26	45.255	-118.145	-0.41	0.23	no-P
281	13324300	26	45.793	-117.957	0.59	0.24	no-P
332	14090350	26	44.635	-121.741	0.20	0.21	no-P
336	14095500	26	44.981	-121.667	1.25	0.40	no-P
337	14096300	26	44.818	-121.728	-0.05	0.19	no-P
34	12050500	27	47.834	-122.939	-0.31	0.21	no-P
69	12120600	27	47.486	-121.995	-0.44	0.22	no-P
90	12153000	27	47.745	-121.623	-0.14	0.19	no-P
128	12334550	27	46.375	-113.109	-0.58	0.24	no-P

**Table B1.** Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index number	USGS site No.	Skew pseudo record length ( $P_{RL}$ )	Basin centroid latitude (decimal degrees)	Basin centroid longitude (decimal degrees)	Station skew coefficient (log units)	MSE of station skew coefficient (log units)	Streamgage used in regional skew study?
333	14090400	27	44.702	-121.725	0.55	0.23	no-P
411	14233400	27	46.525	-121.711	0.17	0.20	no-P
433	14308990	27	42.794	-123.039	-0.08	0.19	no-P
445	14324500	27	43.552	-123.956	-0.84	0.27	no-P
450	14337800	27	42.822	-122.645	-0.57	0.24	no-P
451	14337870	27	42.739	-122.780	0.56	0.24	no-P
42	12076500	28	47.078	-123.087	-0.13	0.19	no-P
66	12115700	28	47.416	-121.756	0.01	0.18	no-P
115	12323750	28	46.014	-112.693	-0.24	0.19	no-P
154	12387450	28	47.150	-114.275	0.12	0.19	no-P
156	12388400	28	47.230	-114.405	0.19	0.19	no-P
182	12433542	28	47.941	-118.062	0.23	0.20	no-P
265	13302005	28	44.426	-113.769	0.34	0.21	no-P
414	14237500	28	46.457	-122.426	0.34	0.21	no-P
427	14306400	28	44.278	-123.767	0.39	0.21	no-P
5	10403000	29	43.793	-119.612	-0.67	0.23	no-P
113	12323250	29	45.936	-112.508	-0.25	0.19	no-P
116	12323770	29	46.156	-113.111	-0.61	0.23	no-P
153	12383500	29	47.138	-113.923	-0.05	0.18	no-P
157	12388700	29	48.088	-114.048	-1.15	0.34	no-P
165	12396900	29	48.868	-117.155	0.09	0.18	no-P
166	12398000	29	48.831	-117.206	0.29	0.20	no-P
187	12448500	29	48.698	-120.294	-0.02	0.17	no-P
193	12454000	29	47.982	-120.973	2.62	1.03	no-P
202	12465500	29	47.709	-118.789	0.01	0.17	no-P
227	13116000	29	44.405	-112.671	0.81	0.25	no-P
233	13135500	29	43.796	-114.601	-0.15	0.18	no-P
239	13153500	29	43.347	-114.408	1.08	0.37	no-P
240	13159800	29	43.323	-115.625	0.87	0.26	no-P
241	13167500	29	42.201	-115.248	-0.21	0.19	no-P
314	14034480	29	45.284	-119.482	1.13	0.38	no-P
335	14093000	29	44.760	-121.483	-0.14	0.18	no-P
338	14096850	29	45.037	-121.463	-0.11	0.18	no-P
401	14206900	29	45.488	-122.713	0.14	0.18	no-P
412	14235500	29	46.641	-122.278	1.60	0.51	no-P
422	14303200	29	45.324	-123.531	-0.34	0.20	no-P
426	14306340	29	44.260	-123.603	0.60	0.23	no-P
456	14362250	29	42.166	-123.129	-0.49	0.22	no-P
458	14375100	29	42.099	-123.417	0.23	0.12	no-P
19	12030000	30	46.710	-123.037	-0.38	0.20	no-P
109	12303500	30	48.309	-115.899	0.48	0.20	no-P
149	12374250	30	47.801	-114.778	0.43	0.20	no-P
150	12375900	30	47.510	-113.956	-0.01	0.17	no-P
151	12377150	30	47.337	-113.925	-0.48	0.21	no-P
152	12381400	30	47.136	-113.797	0.18	0.18	no-P
183	12439300	30	48.924	-119.266	0.15	0.18	no-P
288	13334700	30	46.236	-117.374	0.69	0.23	no-P
310	14022200	30	45.544	-118.533	-0.24	0.19	no-P
315	14034800	30	45.188	-119.501	-0.43	0.20	no-P
388	14190000	30	44.731	-123.485	0.76	0.24	no-P
38	12068500	31	47.770	-122.634	0.03	0.16	no-P
59	12105710	31	47.234	-121.574	-0.70	0.23	no-P

**Table B1.** Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index number	USGS site No.	Skew pseudo record length ( $P_{RL}$ )	Basin centroid latitude (decimal degrees)	Basin centroid longitude (decimal degrees)	Station skew coefficient (log units)	MSE of station skew coefficient (log units)	Streamgage used in regional skew study?
80	12143900	31	47.521	-121.573	0.44	0.20	no-P
188	12448998	31	48.374	-120.414	0.32	0.19	no-P
215	13055340	31	43.785	-111.598	-0.83	0.24	no-P
304	14013500	31	46.045	-118.076	-0.20	0.18	no-P
417	14242580	31	46.305	-122.439	-1.18	0.33	no-P
430	14307700	31	42.953	-122.674	0.19	0.18	no-P
435	14311000	31	43.107	-123.171	-0.35	0.19	no-P
459	14375500	31	42.000	-123.737	0.66	0.22	no-P
57	12104000	32	47.230	-121.553	1.03	0.28	no-P
75	12141000	32	47.516	-121.482	-0.23	0.18	no-P
79	12143600	32	47.434	-121.760	-0.41	0.19	no-P
108	12303100	32	48.336	-115.678	0.40	0.19	no-P
137	12352500	32	46.181	-114.139	-0.04	0.16	no-P
218	13075983	32	43.111	-112.421	0.19	0.17	no-P
276	13313000	32	44.745	-115.506	0.23	0.18	no-P
392	14194300	32	45.372	-123.417	0.70	0.22	no-P
398	14201500	32	44.979	-122.556	-0.50	0.20	no-P
423	14303600	32	45.284	-123.653	-0.30	0.18	no-P
425	14306100	32	44.442	-123.578	0.77	0.23	no-P
84	12147000	33	47.705	-121.558	0.31	0.18	no-P
181	12433200	33	47.999	-117.837	-0.50	0.19	no-P
270	13306500	33	45.115	-114.290	-0.83	0.23	no-P
294	13339500	33	46.339	-115.792	0.14	0.16	no-P
302	13349210	33	46.886	-116.973	-0.11	0.16	no-P
330	14078000	33	44.149	-119.787	0.55	0.20	no-P
31	12043300	34	48.053	-123.807	0.30	0.17	no-P
37	12065500	34	47.544	-122.858	-0.92	0.24	no-P
55	12097850	34	47.050	-121.658	0.46	0.17	no-P
118	12324200	34	46.132	-112.810	-0.29	0.17	no-P
301	13346800	34	46.770	-116.972	-0.39	0.18	no-P
400	14203000	34	45.512	-123.265	1.11	0.29	no-P
405	14216000	34	46.150	-121.795	-0.42	0.18	no-P
120	12324680	35	46.306	-112.740	-0.23	0.16	no-R
225	13113000	38	44.466	-112.245	-0.05	0.14	no-R
358	14139800	38	45.444	-122.013	-0.40	0.16	no-R
194	12456500	41	47.999	-120.789	-0.08	0.13	no-R
136	12351200	43	46.068	-114.122	0.55	0.14	no-R
9	11497500	48	42.442	-120.970	-0.01	0.11	no-R
89	12150800	49	48.083	-122.054	0.15	0.11	no-R
198	12462500	50	47.718	-120.796	0.87	0.17	no-R
76	12141300	52	47.642	-121.589	-0.72	0.16	no-R
190	12449950	54	48.533	-120.250	-0.23	0.11	no-R
295	13340000	57	46.112	-115.331	-0.14	0.10	no-R
162	12394000	58	48.712	-116.914	-0.11	0.10	no-R
271	13307000	60	44.649	-114.092	-1.04	0.17	no-R
448	14330000	61	42.938	-122.356	-0.20	0.10	no-R
286	13332500	64	45.373	-117.852	-0.05	0.08	no-R
341	14107000	67	46.396	-121.244	0.15	0.09	no-R
415	14241500	69	46.235	-122.465	0.02	0.07	no-R
230	13119000	70	44.144	-113.325	-0.09	0.08	no-R
442	14319500	70	43.302	-122.654	-0.16	0.06	no-R
438	14315500	75	43.255	-122.194	0.40	0.08	no-R

**Table B1.** Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index number	USGS site No.	Skew pseudo record length ( $P_{RL}$ )	Basin centroid latitude (decimal degrees)	Basin centroid longitude (decimal degrees)	Station skew coefficient (log units)	MSE of station skew coefficient (log units)	Streamgage used in regional skew study?
87	12148500	77	47.717	-121.581	-0.15	0.08	no-R
437	14312000	82	42.955	-123.149	0.03	0.05	no-R
138	12353000	83	46.496	-113.451	-0.85	0.12	no-R
170	12413000	83	47.771	-116.152	0.29	0.08	no-R
7	11493500	84	42.971	-121.715	-1.03	0.13	no-R
197	12459000	84	47.794	-120.859	0.33	0.08	no-R
320	14046500	84	44.665	-119.166	-0.14	0.07	no-R
99	12189500	85	48.198	-121.305	0.14	0.07	no-R
235	13141000	85	43.647	-114.391	-0.79	0.11	no-R
131	12340500	87	46.631	-113.114	-0.35	0.08	no-R
10	11501000	92	42.564	-121.199	0.05	0.06	no-R
212	13049500	93	44.144	-111.032	0.51	0.08	no-R
266	13302500	100	44.388	-114.279	-0.83	0.10	no-R
144	12362500	102	47.788	-113.461	0.22	0.06	no-R
347	14120000	103	45.487	-121.663	-0.33	0.06	no-R
259	13266000	113	44.611	-116.534	-0.43	0.06	no-R
3	10371500	63	42.189	-120.129	0.05	0.09	Yes
56	10388000	35	43.020	-120.779	0.53	0.19	Yes
216	10396000	86	42.656	-118.738	-0.44	0.09	Yes
367	10406500	70	42.134	-118.353	0.03	0.08	Yes
454	11502500	96	42.748	-121.448	0.35	0.07	Yes
455	12010000	83	46.461	-123.676	-0.25	0.08	Yes
100	12013500	64	46.562	-123.553	-0.20	0.09	Yes
103	12020000	73	46.504	-123.308	0.76	0.12	Yes
273	12024000	40	46.641	-122.579	-0.12	0.14	Yes
364	12025000	72	46.632	-122.703	-0.34	0.09	Yes
308	12025700	45	46.730	-122.533	-0.27	0.13	Yes
343	12027500	84	46.685	-123.011	0.30	0.08	Yes
355	12031000	66	47.086	-123.309	0.43	0.11	Yes
360	12032500	41	47.217	-123.343	-0.17	0.13	Yes
386	12035000	83	47.424	-123.597	-0.25	0.08	Yes
201	12039000	39	47.138	-123.886	-0.31	0.15	Yes
238	12039500	102	47.643	-124.000	-0.87	0.10	Yes
255	12040500	75	47.817	-123.903	-0.05	0.07	Yes
298	12041200	52	48.051	-123.953	-0.51	0.14	Yes
443	12041500	52	47.865	-124.361	-0.08	0.10	Yes
24	12043000	78	48.232	-124.575	-1.08	0.13	Yes
52	12045500	110	48.076	-123.384	0.43	0.34	Yes
62	12048000	87	47.936	-122.963	-0.69	0.10	Yes
228	12056500	88	47.413	-123.233	-0.19	0.07	Yes
357	12060500	79	47.551	-122.768	-0.27	0.08	Yes
361	12069550	40	47.584	-122.745	0.56	0.16	Yes
372	12073500	58	47.052	-123.132	0.00	0.09	Yes
419	12082500	70	46.723	-122.006	-0.07	0.08	Yes
15	12083000	70	46.713	-122.312	-0.65	0.12	Yes
39	12087000	67	47.000	-122.380	-0.19	0.08	Yes
119	12088000	58	47.152	-122.403	0.41	0.12	Yes
260	12092000	76	46.911	-122.006	0.10	0.08	Yes
263	12093500	81	46.972	-121.869	-0.20	0.08	Yes
264	12094000	71	46.998	-122.008	-0.14	0.08	Yes
339	12095000	55	47.022	-121.509	-0.15	0.10	Yes

**Table B1.** Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index number	USGS site No.	Skew pseudo record length ( $P_{RL}$ )	Basin centroid latitude (decimal degrees)	Basin centroid longitude (decimal degrees)	Station skew coefficient (log units)	MSE of station skew coefficient (log units)	Streamgage used in regional skew study?
370	12096500	39	47.032	-121.666	1.00	0.26	Yes
21	12097000	47	47.189	-121.768	0.50	0.15	Yes
127	12097500	69	47.034	-121.625	0.63	0.12	Yes
251	12099600	35	47.040	-121.864	0.26	0.16	Yes
254	12104500	46	47.265	-121.523	0.21	0.13	Yes
292	12108500	68	47.360	-122.066	0.39	0.10	Yes
378	12112600	52	47.316	-122.281	0.54	0.14	Yes
429	12114000	39	47.326	-121.531	0.68	0.19	Yes
326	12114500	57	47.398	-121.670	0.07	0.10	Yes
68	12115000	63	47.322	-121.636	-0.30	0.10	Yes
461	12115500	66	47.381	-121.739	-0.35	0.10	Yes
185	12117000	55	47.365	-122.002	0.16	0.10	Yes
328	12118500	43	47.543	-122.139	-0.44	0.15	Yes
387	12121600	49	47.748	-122.057	-1.20	0.25	Yes
17	12133000	75	47.899	-121.365	0.00	0.07	Yes
199	12134500	84	47.938	-121.664	-0.14	0.07	Yes
272	12135000	66	47.927	-121.585	0.14	0.08	Yes
345	12142000	81	47.618	-121.618	-0.27	0.08	Yes
365	12143400	52	47.424	-121.742	-0.43	0.13	Yes
404	12144000	96	47.481	-121.882	-0.74	0.09	Yes
418	12144500	54	47.605	-121.976	-0.27	0.11	Yes
58	12145500	66	47.632	-121.827	0.10	0.09	Yes
111	12147500	57	47.700	-121.618	-0.16	0.10	Yes
258	12147600	49	47.709	-121.702	-0.51	0.14	Yes
275	12149000	83	48.030	-121.826	-0.11	0.07	Yes
297	12161000	52	48.074	-121.663	-0.32	0.12	Yes
354	12167000	84	48.305	-121.791	-0.25	0.08	Yes
356	12175500	82	48.576	-121.033	0.64	0.10	Yes
409	12177500	50	48.756	-121.202	0.73	0.16	Yes
53	12178100	52	48.612	-121.218	0.29	0.12	Yes
142	12182500	65	48.468	-121.205	0.65	0.12	Yes
167	12186000	89	48.052	-121.333	0.57	0.09	Yes
176	12196000	36	48.565	-121.924	-0.17	0.15	Yes
375	12201500	57	48.622	-122.289	0.07	0.10	Yes
350	12205000	75	48.877	-121.702	0.02	0.07	Yes
441	12208000	36	48.765	-121.963	0.08	0.15	Yes
70	12209000	63	48.637	-121.977	-0.25	0.10	Yes
86	12210500	69	48.804	-121.963	0.16	0.08	Yes
177	12301300	54	48.741	-114.923	-0.26	0.11	Yes
237	12302055	62	48.173	-115.155	-0.39	0.11	Yes
312	12305500	50	48.566	-116.178	0.22	0.12	Yes
383	12311000	46	48.570	-116.428	2.53	0.74	Yes
457	12324590	40	46.568	-112.500	-0.28	0.15	Yes
95	12325500	72	46.180	-113.272	-1.41	0.26	Yes
110	12329500	71	46.276	-113.310	0.04	0.08	Yes
285	12330000	73	46.389	-113.162	0.42	0.10	Yes
402	12332000	75	46.036	-113.538	-0.86	0.13	Yes
213	12334510	41	46.312	-113.608	-0.92	0.21	Yes
384	12335500	73	46.763	-112.670	-0.19	0.08	Yes
28	12340000	108	47.026	-113.101	-0.31	0.06	Yes
29	12342500	72	45.626	-114.293	-1.01	0.16	Yes

**Table B1.** Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index number	USGS site No.	Skew pseudo record length ( $P_{RL}$ )	Basin centroid latitude (decimal degrees)	Basin centroid longitude (decimal degrees)	Station skew coefficient (log units)	MSE of station skew coefficient (log units)	Streamgage used in regional skew study?
61	12344000	75	45.796	-114.116	-1.04	0.16	Yes
78	12346500	52	46.173	-113.842	-0.30	0.11	Yes
92	12354000	63	47.378	-115.392	2.27	0.40	Yes
96	12354500	100	46.594	-113.668	-0.89	0.11	Yes
134	12358500	73	48.331	-113.525	1.40	0.28	Yes
368	12359800	47	47.626	-113.327	0.32	0.13	Yes
399	12365000	54	48.461	-114.675	-0.43	0.13	Yes
219	12366000	63	48.533	-114.444	-0.26	0.10	Yes
222	12370000	91	47.640	-113.785	-0.06	0.06	Yes
407	12372000	118	48.242	-113.972	-0.05	0.05	Yes
460	12389500	60	47.822	-115.022	-0.20	0.10	Yes
82	12390700	57	47.547	-115.517	-0.08	0.10	Yes
106	12395000	105	48.593	-116.915	-0.30	0.06	Yes
145	12396000	56	48.250	-117.461	0.01	0.09	Yes
322	12408500	47	48.635	-117.734	-0.37	0.14	Yes
369	12409000	90	48.391	-117.774	-0.72	0.10	Yes
51	12411000	62	47.881	-116.202	-0.19	0.10	Yes
67	12413500	82	47.698	-116.124	0.30	0.08	Yes
192	12414500	94	47.202	-115.729	-0.06	0.06	Yes
229	12414900	47	47.056	-116.312	-0.21	0.12	Yes
324	12416000	49	47.841	-116.609	-0.24	0.12	Yes
377	12424000	65	47.373	-117.147	-0.13	0.09	Yes
406	12431000	70	47.982	-117.259	-0.50	0.11	Yes
449	12447390	44	48.875	-120.177	0.71	0.17	Yes
164	12449500	87	48.630	-120.310	-0.36	0.08	Yes
256	12451000	91	48.390	-120.860	0.06	0.06	Yes
282	12452800	55	47.992	-120.571	-0.01	0.09	Yes
327	12457000	90	47.909	-120.891	0.38	0.08	Yes
413	12458000	65	47.583	-120.936	0.65	0.13	Yes
63	12464800	45	47.504	-118.259	-0.28	0.12	Yes
85	12465000	70	47.454	-118.276	-0.93	0.15	Yes
101	12465400	38	47.750	-118.733	-0.66	0.17	Yes
159	12488500	73	46.919	-121.374	0.30	0.09	Yes
267	12500500	71	46.530	-121.065	0.03	0.07	Yes
299	12501000	58	46.476	-121.022	0.46	0.12	Yes
309	12502500	100	46.526	-120.905	-0.10	0.06	Yes
376	12513000	60	46.802	-118.770	-0.60	0.13	Yes
432	13047500	100	44.213	-110.930	0.07	0.06	Yes
434	13052200	51	43.663	-111.063	-0.16	0.11	Yes
436	13057940	35	43.218	-111.610	-0.05	0.15	Yes
440	13075000	58	42.476	-112.196	-0.06	0.09	Yes
41	13078000	53	41.993	-113.653	0.02	0.10	Yes
47	13082500	99	41.926	-114.120	0.56	0.08	Yes
205	13083000	100	42.138	-114.108	0.47	0.08	Yes
217	13092000	53	42.249	-114.268	-1.07	0.18	Yes
250	13112000	88	44.338	-111.952	-0.94	0.13	Yes
323	13113500	65	44.406	-112.189	-0.41	0.10	Yes
391	13116500	39	44.393	-112.586	0.67	0.20	Yes
395	13118700	55	44.214	-113.407	-0.16	0.10	Yes
313	13120000	69	43.892	-114.237	-0.96	0.15	Yes
390	13120500	105	43.857	-114.037	-0.59	0.08	Yes

**Table B1.** Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index number	USGS site No.	Skew pseudo record length ( $P_{RL}$ )	Basin centroid latitude (decimal degrees)	Basin centroid longitude (decimal degrees)	Station skew coefficient (log units)	MSE of station skew coefficient (log units)	Streamgage used in regional skew study?
158	13139500	96	43.710	-114.431	-0.35	0.07	Yes
208	13141500	89	43.350	-114.863	-0.71	0.10	Yes
305	13147900	49	43.587	-113.991	-0.40	0.13	Yes
316	13150430	38	43.328	-114.171	-0.09	0.14	Yes
371	13168500	74	42.153	-115.563	-0.14	0.08	Yes
453	13169500	64	42.533	-116.062	-0.27	0.10	Yes
107	13185000	137	43.850	-115.386	-0.23	0.05	Yes
169	13186000	69	43.646	-115.001	-0.81	0.13	Yes
247	13200000	62	43.846	-115.879	-0.42	0.11	Yes
306	13214000	73	43.915	-118.539	-0.29	0.09	Yes
2	13216500	58	44.170	-118.305	-0.36	0.11	Yes
64	13226500	41	44.055	-117.820	-0.58	0.18	Yes
104	13235000	72	44.142	-115.306	-0.16	0.08	Yes
139	13240000	67	44.951	-115.930	-0.10	0.08	Yes
146	13247500	41	44.400	-115.800	-1.20	0.28	Yes
311	13250600	38	44.127	-116.414	-0.21	0.14	Yes
439	13251500	56	45.015	-116.412	0.36	0.11	Yes
13	13258500	74	44.799	-116.458	-0.17	0.08	Yes
243	13261000	46	44.525	-116.262	0.13	0.12	Yes
317	13288200	40	45.025	-117.354	0.33	0.15	Yes
97	13292000	84	45.319	-116.960	0.60	0.10	Yes
178	13295000	75	44.267	-115.041	-0.21	0.08	Yes
196	13297330	40	44.321	-114.572	-1.08	0.26	Yes
226	13297355	40	44.318	-114.515	-0.36	0.15	Yes
277	13305000	57	44.674	-113.428	-0.33	0.11	Yes
20	13308500	45	44.403	-115.107	-0.28	0.13	Yes
65	13309220	36	44.596	-115.218	-0.51	0.17	Yes
73	13310700	46	44.749	-115.719	0.04	0.11	Yes
83	13316500	65	45.154	-116.323	-0.10	0.09	Yes
325	13317000	115	44.885	-114.774	-0.58	0.07	Yes
46	13320000	75	45.143	-117.631	-0.24	0.08	Yes
253	13329500	56	45.271	-117.311	-0.17	0.10	Yes
318	13330000	98	45.305	-117.397	-0.21	0.06	Yes
351	13330500	80	45.415	-117.519	-0.04	0.07	Yes
416	13331500	50	45.321	-117.578	-0.09	0.11	Yes
452	13333000	68	45.486	-117.797	0.21	0.09	Yes
60	13336500	95	45.978	-114.882	-1.09	0.14	Yes
287	13337000	86	46.412	-114.960	0.01	0.06	Yes
54	13337500	41	45.796	-115.385	-0.19	0.14	Yes
105	13338500	91	45.856	-115.783	-0.43	0.07	Yes
231	13340600	46	46.700	-115.272	0.17	0.12	Yes
245	13342450	38	46.290	-116.712	-0.12	0.14	Yes
410	13344500	57	46.366	-117.713	0.14	0.10	Yes
6	13345000	93	46.954	-116.754	-0.08	0.06	Yes
44	14013000	75	45.994	-118.019	0.43	0.10	Yes
45	14017000	61	46.230	-117.935	-0.31	0.10	Yes
179	14018500	62	46.129	-118.241	0.26	0.10	Yes
200	14020000	80	45.695	-118.188	0.40	0.09	Yes
342	14020300	37	45.530	-118.296	-0.18	0.15	Yes
50	14021000	57	45.666	-118.404	0.19	0.10	Yes
122	14022500	63	45.473	-118.574	0.15	0.09	Yes

**Table B1.** Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index number	USGS site No.	Skew pseudo record length ( $P_{RL}$ )	Basin centroid latitude (decimal degrees)	Basin centroid longitude (decimal degrees)	Station skew coefficient (log units)	MSE of station skew coefficient (log units)	Streamgage used in regional skew study?
204	14025000	49	45.410	-118.838	-0.04	0.11	Yes
16	14032000	59	45.375	-119.139	0.96	0.17	Yes
121	14037500	61	44.312	-118.690	0.25	0.10	Yes
132	14040500	64	44.335	-119.160	-0.10	0.09	Yes
252	14042500	67	45.167	-118.685	0.18	0.09	Yes
403	14044000	83	44.682	-118.720	-0.19	0.07	Yes
14	14048000	109	44.788	-119.524	-0.13	0.05	Yes
123	14050000	54	43.973	-121.754	-0.83	0.16	Yes
129	14050500	58	43.875	-121.832	-0.11	0.10	Yes
141	14051000	55	43.858	-121.903	-0.35	0.12	Yes
203	14052000	66	43.804	-121.919	-0.37	0.10	Yes
249	14053000	42	43.762	-121.922	-0.26	0.14	Yes
331	14054500	56	43.719	-121.884	-0.09	0.10	Yes
340	14055500	44	43.570	-122.034	0.86	0.19	Yes
420	14075000	79	44.163	-121.693	0.66	0.11	Yes
428	14088000	73	44.400	-121.800	0.18	0.08	Yes
242	14091500	92	44.541	-121.701	0.79	0.11	Yes
257	14097100	40	44.939	-121.514	0.12	0.14	Yes
366	14101500	73	45.231	-121.416	-0.06	0.08	Yes
27	14110000	70	46.311	-121.301	0.13	0.08	Yes
71	14112500	37	45.904	-120.864	-0.05	0.14	Yes
102	14113000	87	46.051	-121.116	0.16	0.07	Yes
126	14113200	45	45.584	-121.412	0.29	0.13	Yes
133	14118500	76	45.523	-121.754	-0.16	0.08	Yes
262	14123500	90	45.984	-121.539	0.62	0.10	Yes
279	14128500	48	45.871	-121.912	0.15	0.12	Yes
303	14134000	67	45.300	-121.708	0.36	0.10	Yes
373	14137000	101	45.308	-121.876	-0.04	0.05	Yes
48	14138850	46	45.483	-121.914	0.16	0.12	Yes
346	14138870	37	45.471	-121.984	0.09	0.14	Yes
374	14138900	46	45.535	-122.017	-0.06	0.11	Yes
382	14139700	39	45.447	-121.979	-0.48	0.17	Yes
447	14141500	94	45.415	-122.042	0.12	0.06	Yes
30	14143500	37	45.692	-122.169	0.30	0.16	Yes
389	14144800	39	43.483	-122.318	0.21	0.15	Yes
36	14146500	79	43.765	-122.207	-0.49	0.09	Yes
329	14147500	83	43.846	-122.223	-0.04	0.06	Yes
362	14150300	36	43.976	-122.476	-0.57	0.19	Yes
380	14150800	45	43.888	-122.573	-0.48	0.14	Yes
284	14152500	74	43.581	-123.022	-0.25	0.08	Yes
307	14156500	35	43.640	-122.915	-0.26	0.16	Yes
49	14158790	52	44.360	-122.075	1.31	0.28	Yes
77	14159200	54	43.968	-122.075	-0.05	0.10	Yes
381	14161100	40	44.279	-122.233	-0.22	0.14	Yes
396	14161500	61	44.233	-122.171	0.65	0.13	Yes
94	14163000	39	44.187	-122.494	-0.13	0.14	Yes
174	14165000	75	44.203	-122.816	-0.61	0.11	Yes
408	14166500	77	44.104	-123.452	-0.92	0.14	Yes
12	14167000	47	43.955	-123.235	-0.75	0.17	Yes
22	14171000	57	44.576	-123.498	-0.50	0.13	Yes
88	14172000	55	44.282	-122.578	-0.25	0.11	Yes

**Table B1.** Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index number	USGS site No.	Skew pseudo record length ( $P_{RL}$ )	Basin centroid latitude (decimal degrees)	Basin centroid longitude (decimal degrees)	Station skew coefficient (log units)	MSE of station skew coefficient (log units)	Streamgage used in regional skew study?
319	14173500	41	44.398	-122.876	-0.52	0.17	Yes
363	14178000	87	44.609	-121.939	-0.05	0.05	Yes
18	14179000	79	44.773	-121.960	-0.09	0.07	Yes
72	14182500	81	44.827	-122.336	-0.09	0.07	Yes
93	14185000	77	44.379	-122.318	-0.18	0.08	Yes
261	14185900	49	44.589	-122.311	0.46	0.14	Yes
4	14187000	51	44.328	-122.528	-0.11	0.11	Yes
291	14188800	37	44.688	-122.574	0.11	0.15	Yes
33	14189500	44	44.774	-123.574	-0.08	0.12	Yes
189	14190500	78	44.779	-123.444	0.04	0.05	Yes
344	14192500	59	45.070	-123.640	-0.23	0.10	Yes
379	14193000	58	45.201	-123.502	0.09	0.09	Yes
35	14198500	58	44.951	-122.352	0.49	0.12	Yes
224	14200000	81	45.074	-122.415	-0.09	0.07	Yes
98	14202000	52	45.003	-122.713	0.84	0.18	Yes
236	14208000	50	44.914	-121.861	0.00	0.10	Yes
168	14211500	72	45.466	-122.418	-0.66	0.12	Yes
195	14212000	45	45.779	-122.419	0.36	0.14	Yes
348	14216500	55	46.214	-122.016	0.26	0.11	Yes
147	14219800	53	46.039	-122.373	-0.41	0.13	Yes
191	14222500	82	45.808	-122.284	-0.14	0.07	Yes
293	14223500	46	46.080	-122.533	0.55	0.14	Yes
334	14232500	69	46.370	-121.659	0.03	0.08	Yes
300	14236200	56	46.605	-122.288	0.01	0.09	Yes
175	14242500	67	46.303	-122.423	0.53	0.11	Yes
359	14245000	45	46.157	-122.662	0.44	0.14	Yes
290	14247500	39	46.312	-123.274	-0.12	0.14	Yes
424	14301000	73	45.867	-123.392	-0.14	0.08	Yes
11	14301500	96	45.578	-123.520	-0.17	0.06	Yes
81	14305500	95	44.817	-123.720	-0.26	0.06	Yes
234	14306500	73	44.350	-123.675	-0.27	0.09	Yes
421	14307620	41	44.022	-123.570	-0.09	0.13	Yes
446	14308000	99	43.034	-122.701	-0.14	0.06	Yes
283	14308500	57	42.836	-122.855	0.39	0.11	Yes
220	14309500	57	42.830	-123.743	-0.41	0.12	Yes
431	14311500	57	43.079	-123.558	-0.83	0.15	Yes
140	14316500	63	43.249	-122.260	0.31	0.10	Yes
206	14316700	57	43.453	-122.647	0.09	0.10	Yes
210	14318000	48	43.195	-122.896	-0.46	0.12	Yes
221	14320700	38	43.464	-123.174	-0.05	0.12	Yes
352	14321000	106	43.161	-123.002	-0.03	0.05	Yes
26	14325000	96	42.792	-124.027	-0.17	0.06	Yes
163	14328000	77	42.962	-122.342	0.09	0.07	Yes
232	14333500	55	42.797	-122.277	0.31	0.11	Yes
444	14338000	67	42.793	-122.682	-0.03	0.08	Yes
130	14341500	61	42.319	-122.437	-0.14	0.09	Yes
321	14353000	35	42.120	-122.740	1.87	0.68	Yes
32	14353500	35	42.111	-122.703	2.01	0.76	Yes
278	14371500	49	42.693	-123.185	0.32	0.13	Yes
148	14377100	53	42.076	-123.595	0.14	0.07	Yes
244	14400000	43	42.233	-124.005	-0.50	0.16	Yes

The  $P_{RL}$  is defined in terms of the number of years of systematic record that would be required to yield the same mean square error of the skew ( $MSE(\hat{G})$ ) as the combination of historical and systematic record actually available at a streamgage. Thus, the  $P_{RL}$  of the skew is a ratio of the mean square error (MSE) of the at-station skew when only the systematic record is analyzed  $MSE(\hat{G}_S)$  as opposed to the MSE of the at-station skew when the all of the data, including historical and censored data, are analyzed  $MSE(\hat{G}_C)$ .

$$P_{RL} = \frac{P_s * MSE(\hat{G}_S)}{MSE(\hat{G}_C)} \quad (B2)$$

where

- $P_{RL}$  is the pseudo record length for the entire record at the streamgage;
- $P_s$  is the number of systematic peaks in the record;
- $MSE(\hat{G}_S)$  is the estimated MSE of the skew when only the systematic record is analyzed; and
- $MSE(\hat{G}_C)$  is the estimated MSE of the skew when all of the data, including historical and censored data, are analyzed.

As the  $P_{RL}$  is an estimate, the following conditions must also be met to ensure a valid approximation.  $P_{RL}$  must be non-negative. If  $P_{RL}$  is greater than  $P_H$  (the length of the historical period), then  $P_{RL}$  should be set to equal  $P_H$ . Additionally, if  $P_{RL}$  is less than  $P_s$ , then  $P_{RL}$  is set to  $P_s$ . This ensures that the pseudo record length will not be larger than the complete historical period or less than the number of systematic peaks.

As stated in Bulletin 17B, the skew coefficient of the station skew is sensitive to extreme events and more accurate estimates can be obtained from longer records. Thus, after ensuring adequate spatial and hydrologic coverage those streamgages that do not have a minimum of 35 years of pseudo record length were removed from the regional skew study. Of the 461 streamgages, 28 were removed because their  $P_{RL}$  was less than 25 years, 62 were removed because their  $P_{RL}$  was between 25 and 29 years, and 45 were removed because their  $P_{RL}$  was between 30 and 34 years. Thus, data from 326 streamgages remained from which to build a regional skew model for the PNW study area.

## Redundant Sites

Redundancy results when the drainage basins of two streamgages are nested, meaning that one basin is contained inside the other and the two basins are of similar size. Instead of providing two independent spatial observations that depict how drainage basin characteristics are related to skew (or AEPs), these two basins will have similar hydrologic responses to a given runoff event, and thus represent only

one spatial observation. When streamgages in basins (streamgage-pairs) are redundant, a statistical analysis using both streamgages incorrectly represents the information in the regional data set (Gruber and Stedinger, 2008). To determine if two sites might be redundant, two variables are considered: (1) standardized distance (SD), and (2) drainage-area ratio (DAR).

The SD is used to determine the likelihood that the basins are nested. The standardized distance between two basin centroids, SD is defined as:

$$SD_{ij} = \frac{D_{ij}}{\sqrt{0.5(DRNAREA_i + DRNAREA_j)}} \quad (B3)$$

where

- $D_{ij}$  is the distance between centroids of basin  $i$  and basin  $j$ ; and
- $DRNAREA_i$  is the drainage area at site  $i$ ; and
- $DRNAREA_j$  is the drainage area at site  $j$ .

The DAR is used to determine if two nested basins are sufficiently similar in size to conclude that they are, or are at least in large part, the same watershed for the purposes of developing a regional hydrologic model. The DAR is defined as (Veilleux, 2011):

$$DAR = \text{Max} \left[ \frac{DRNAREA_i}{DRNAREA_j}, \frac{DRNAREA_j}{DRNAREA_i} \right] \quad (B4)$$

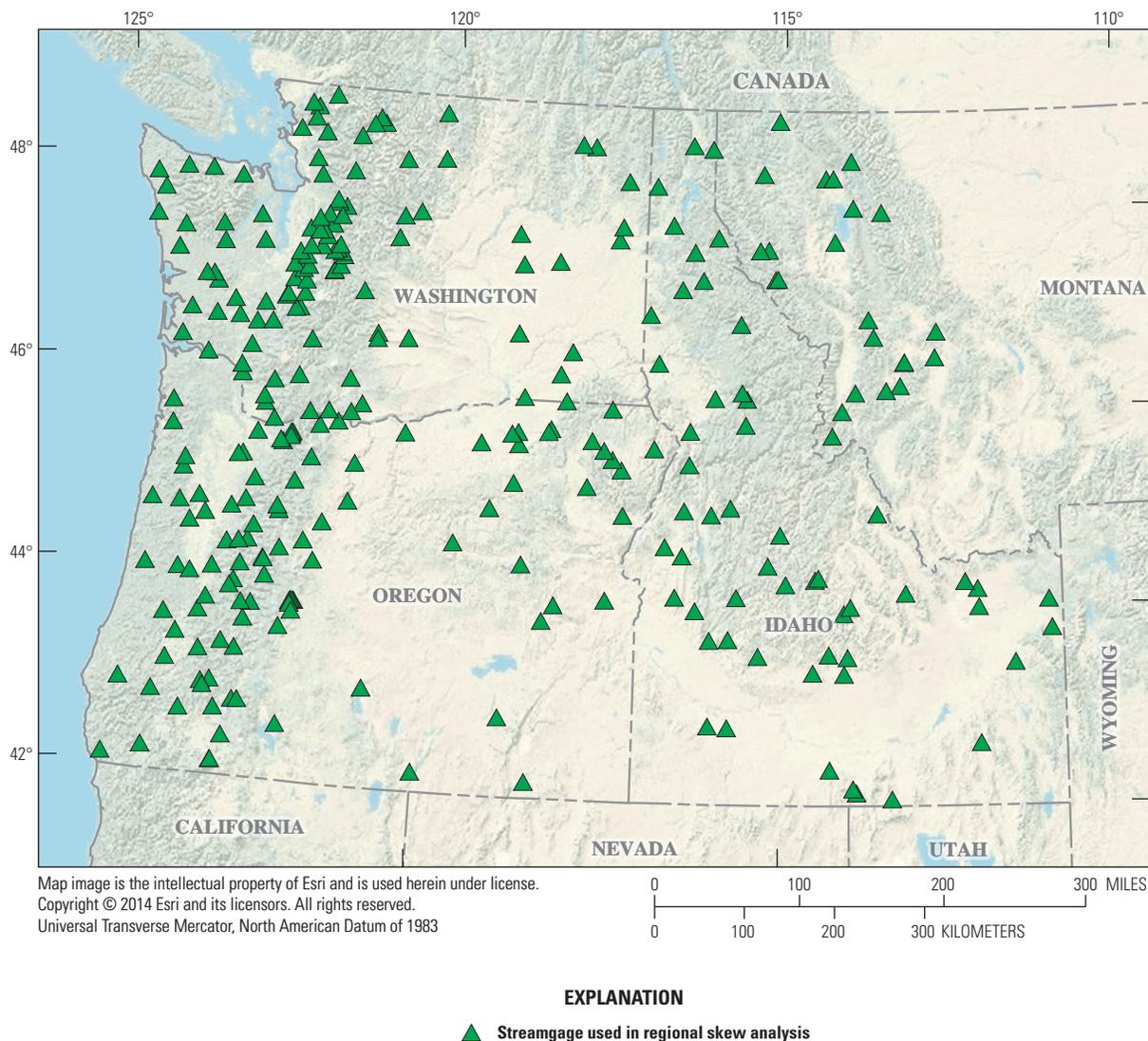
where

- DAR is the Max (maximum) of the two values in brackets;
- $DRNAREA_i$  is the drainage area at site  $i$ ; and
- $DRNAREA_j$  is the drainage area at site  $j$ .

Two basins might be expected to have possible redundancy if the basin sizes are fairly similar and the basins are nested. Previous studies (Southard and Veilleux, 2014) suggest that streamgage-pairs with SD less than or equal to 0.50 and DAR less than or equal to 5 were likely to have possible redundancy problems for purposes of determining regional skew. If DAR is large enough, even if the streamgage-pairs are nested, they will reflect different hydrologic responses because storms of different sizes and durations will affect each streamgage differently. All possible combinations of streamgage-pairs from the 326 streamgages were considered in the redundancy analysis. All streamgage-pairs identified as redundant were then investigated to determine if one streamgage of the pair is nested inside the other. For streamgage-pairs that are nested, one streamgage from the pair was removed from the regional skew analysis. Streamgages removed from the PNW regional skew study because of redundancy are identified in the last column of [table B1](#).

From 126 streamgage-pairs identified as possibly redundant, 113 were determined to be redundant, but only 36 streamgages were removed from the analyses as the same streamgages appeared in multiple streamgage-pairs. Thus, of the 326 streamgages, 36 were removed because of redundancy,

which left a final list of 290 streamgages to use in the PNW regional skew study (table B1). The location of these 290 streamgages used in PNW regional skew study is shown in figure B1.



**Figure B1.** Locations of streamgages selected for the regional skew analysis in the Pacific Northwest.

### Unbiasing the Site Estimators

The at-station skew estimates were unbiased by using the correction factor developed by Tasker and Stedinger (1986) and employed in Reis and others (2005). The unbiased at-station skew estimator using the pseudo record length is:

$$\hat{\gamma}_i = \left[ 1 + \frac{6}{P_{RL,i}} \right] G_i \tag{B5}$$

where

- $\hat{\gamma}_i$  is the unbiased at-station sample skew estimate for site  $i$ ;
- $P_{RL,i}$  is the pseudo record length for site  $i$  as estimated in equation B2; and
- $G_i$  is the traditional biased at-station skew estimator for site  $i$  from the flood frequency analysis.

The variance of the unbiased at-station skew includes the correction factor developed by Tasker and Stedinger (1986):

$$Var[\hat{\gamma}_i] = \left[ 1 + \frac{6}{P_{RL,i}} \right]^2 Var[G_i] \tag{B6}$$

where

$Var[G_i]$  is estimated using (Griffis and Stedinger, 2009)

$$Var(\hat{G}) = \left[ \frac{6}{P_{RL}} + a(P_{RL}) \right] \times \left[ 1 + \left( \frac{9}{6} + b(P_{RL}) \right) \hat{G}^2 + \left( \frac{15}{48} + c(P_{RL}) \right) \hat{G}^4 \right] \tag{B7}$$

where

$$a(P_{RL}) = -\frac{17.75}{P_{RL}^2} + \frac{50.06}{P_{RL}^3};$$

$$b(P_{RL}) = \frac{3.92}{P_{RL}^{0.3}} - \frac{31.10}{P_{RL}^{0.6}} + \frac{34.86}{P_{RL}^{0.9}}; \text{ and}$$

$$c(P_{RL}) = -\frac{7.31}{P_{RL}^{0.59}} + \frac{45.90}{P_{RL}^{1.18}} - \frac{86.50}{P_{RL}^{1.77}}.$$

### Estimating the Mean Square Error of the Skew Estimator

There are several ways to estimate  $MSE_G$ . The approach used by EMA (taken from Cohn and others [2001, equation 55]) generates a first order estimate of the  $MSE_G$ , which should perform well when interval data are present. Another option is to use the Griffis and Stedinger (2009) formula in

equation B7 (the variance is equated to the MSE), employing either the systematic record length or the length of the whole historical period. However, this method does not account for censored data, and thus can lead to inaccurate and underestimated  $MSE_G$ . This issue was addressed by using the pseudo record length instead of the length of the historical period; the pseudo record length reflects the impact of the censored data and the number of recorded systematic peaks. Thus, the unbiased Griffis and Stedinger (2009)  $MSE_G$  was used in the regional skew model because it is more stable and relatively independent of the at-station skew estimator. This methodology was used in previous regional skew studies (Southard and Veilleux, 2014; Eash and others, 2013).

### Cross-Correlation Models

A critical step for a GLS analysis is estimation of the cross correlation of the skew coefficient estimators. Martins and Stedinger (2002) used Monte Carlo experiments to derive a relation between the cross correlation of the skew estimators at two sites  $i$  and  $j$  as a function of the cross correlation of concurrent annual maximum flows,  $\rho_{ij}$ :

$$\hat{\rho}(\hat{\gamma}_i, \hat{\gamma}_j) = Sign(\hat{\rho}_{ij}) cf_{ij} |\hat{\rho}_{ij}|^\kappa \tag{B8}$$

where

- $\hat{\rho}_{ij}$  is the cross-correlation of concurrent annual peak flow for two streamgages;
- $\kappa$  is a constant between 2.8 and 3.3; and
- $cf_{ij}$  is a factor that accounts for the sample size difference between sites and their concurrent record length, is defined as follows:

$$cf_{ij} = CY_{ij} / \sqrt{(P_{RL,i})(P_{RL,j})} \tag{B9}$$

where

- $CY_{ij}$  is the pseudo concurrent record length (estimated in equation B10); and
- $P_{RL,i}$ , and  $P_{RL,j}$  are the pseudo record length corresponding to sites  $i$  and  $j$ , respectively (see equation B2).

### Pseudo Concurrent Record Length

After calculating the  $P_{RL}$  for each streamgage in the study, the pseudo concurrent record length between pairs of sites can be estimated. Because of the use of censored data and historical data, the calculation of effective concurrent record length is more complex than determining in which years the two streamgages both have recorded systematic peaks.

The years of historical record in common between the two streamgages is first determined. For the years in common, with beginning year,  $YB_{ij}$ , and ending year,  $YE_{ij}$ , the following equation is used to estimate the pseudo concurrent record length between site  $i$  and site  $j$ .

$$CY_{ij} = (YE_{ij} - YB_{ij} + 1) \left( \frac{P_{RL,i}}{P_{H,i}} \right) \left( \frac{P_{RL,j}}{P_{H,j}} \right). \quad (B10)$$

The estimated pseudo concurrent record length depends upon the years of historical record in common between the two streamgages, as well as the ratios of the pseudo record length to the historical record length for each of the two streamgages.

### Cross-Correlation Model of Concurrent Annual Peak Flow

A cross-correlation model for the logarithm of the annual peak flows in the PNW study area was developed using 33 sites with at least 85 years of concurrent systematic peaks (zero flows not included). Various models relating the cross correlation of the concurrent annual peak flow at two sites,  $r_{ij}$ , to various basin characteristics were considered. A logit model, termed the Fisher Z Transformation ( $Z = \log[(1+r)/(1-r)]$ ), provided a convenient transformation of the sample correlations  $r_{ij}$  from the (-1, +1) range to

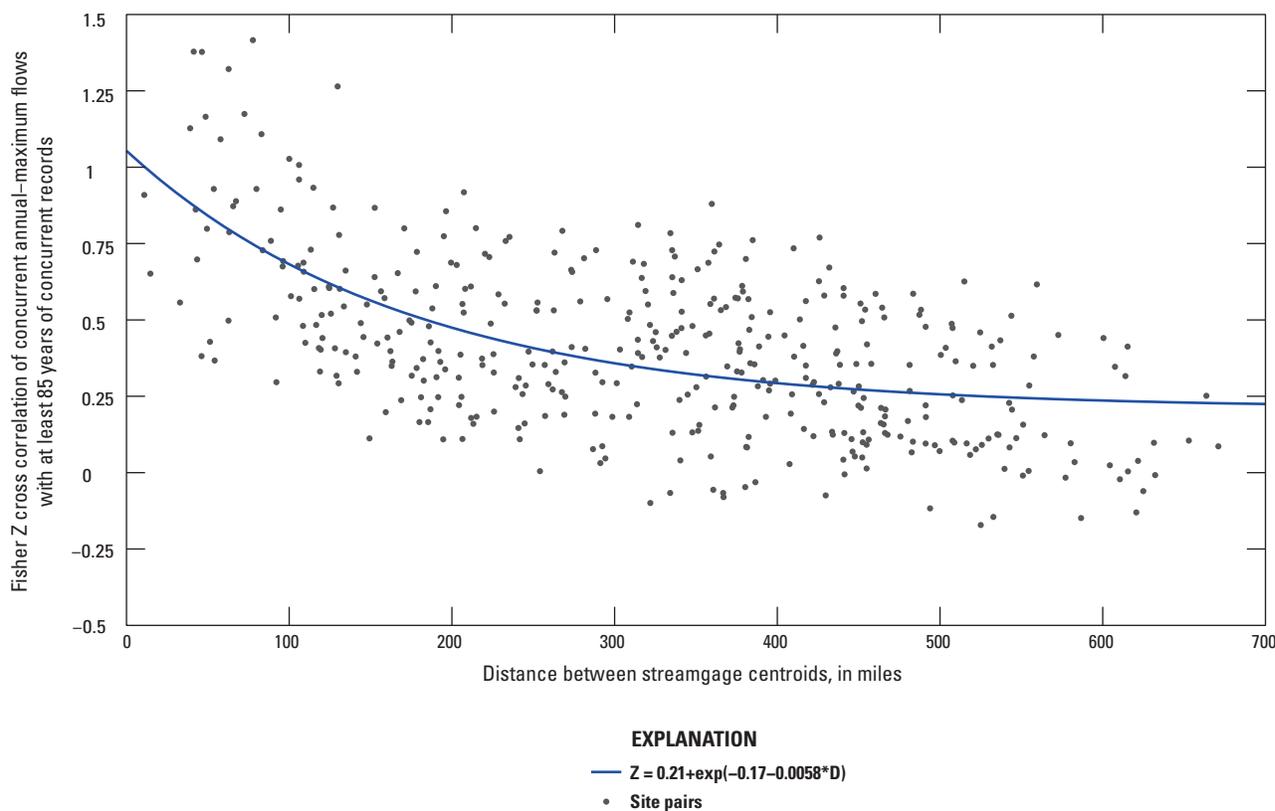
the  $(-\infty +\infty)$  range (Devore, 2004). The adopted model for estimating the cross correlations of concurrent annual peak flow at two sites, which used the distance between basin centroids,  $D_{ij}$ , as the only explanatory variable, is:

$$\rho_{ij} = \frac{\exp(2Z_{ij}) - 1}{\exp(2Z_{ij}) + 1} \quad (B11)$$

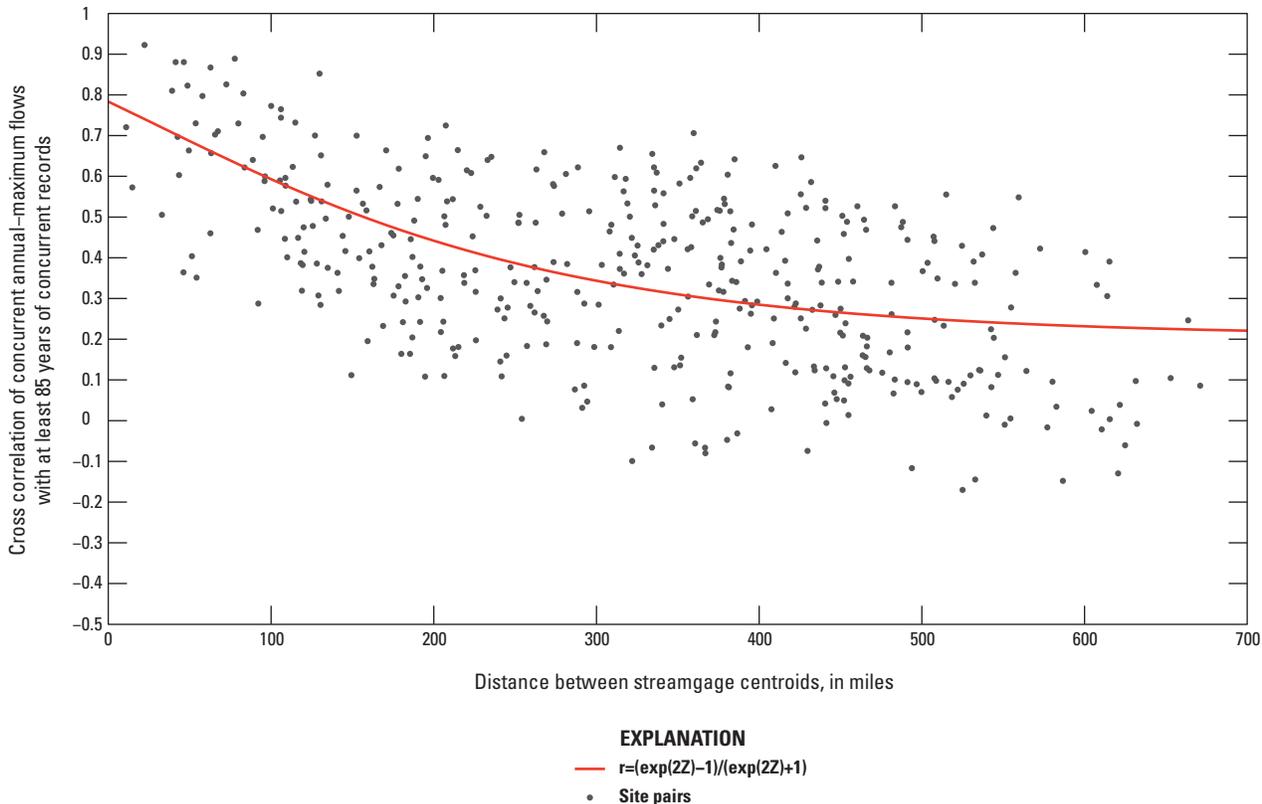
where

$$Z_{ij} = 0.21 + \exp(-0.17 - 0.0058D_{ij})$$

The fitted relation between Z and distance between basin centroids together with the plotted sample data from 411 streamgage-pairs of data are shown in figure B2. The functional relation between the untransformed cross correlation and distance between basin centroids together with the plotted sample data from the 411 streamgage-pairs of data are shown in figure B3. The cross-correlation model was used to estimate streamgage-to-streamgage cross correlations for concurrent annual peak flows at all pairs of streamgages in the regional skew study.



**Figure B2.** Relation between Fisher Z transformed cross correlation of logs of annual peak flow and distance between basin centroids for the regional skew study in the Pacific Northwest.



**Figure B3.** Relation between un-transformed cross correlation of logs of annual peak flow and distance between basin centroids for the regional skew study in the Pacific Northwest.

### Pacific Northwest Regional Skew Study Results

The best regional skew model is classified as having the smallest model error variance,  $\sigma_{\delta}^2$ , and largest pseudo- $R_{\delta}^2$ . The pseudo- $R_{\delta}^2$  describes the estimated fraction of the variability in the true skew from streamgauge-to-streamgauge explained by each model (Gruber and others, 2007; Parrett and others, 2011). None of the basin characteristics considered for the PNW regional skew study were statistically significant and none produced a pseudo- $R_{\delta}^2$  greater than 3 percent. This indicates that including basin characteristics as explanatory variables in the regression did not help explain the variability in the true skew and is not warranted, because the increased model complexity provides only a very small gain in model precision. Thus, the CONSTANT model is chosen as the best regional skew model for the PNW study area. The final results for the constant skew model, denoted CONSTANT, for the PNW study area using 290 streamgages with at least 35 years of pseudo record length are provided in table B3. The resulting generalized regional skew was -0.07.

A constant model does not explain any variability in the true skews, so the pseudo-  $R_{\delta}^2$  equals 0. The posterior mean of the model error variance,  $\sigma_{\delta}^2$ , for the CONSTANT

model is  $\sigma_{\delta}^2 = 0.17$ . The average sampling error variance (ASEV) in table B2 is the average error in the regional skew estimator at the sites in the data set. The average variance of prediction at a new site ( $AVP_{new}$ ) corresponds to the MSE used in Bulletin 17B to describe the precision of the generalized skew. The CONSTANT model has an  $AVP_{new}$  equal to 0.18, which corresponds to an effective record length of 41 years. An  $AVP_{new}$  of 0.18 is a marked improvement over the Bulletin 17B skew map, whose reported MSE is 0.303 (Interagency Advisory Committee on Water Data, 1982) for a corresponding pseudo or effective record length of only 17 years. Thus, the new regional model has 2.4 times the information content (as measured by pseudo or effective record length) of that claimed by the Bulletin 17B skew map.

It is important to note that this regional skew model is not valid in the portion of Idaho omitted from both the regional peak-flow regression and regional skew analysis in the Snake River Plain (region 0 in figure 3) because of extensive regulation, localized groundwater-surface water interactions, and substantial infiltration. The regional skew model should not be used to estimate a weighted skew for peak-flow frequency statistics in this area.

**Table B2.** Regional skew model for the study area in the Pacific Northwest.

[Standard deviations are in parentheses. **Abbreviations:**  $\sigma_{\delta}^2$ , model error variance; ASEV, average sampling error variance;  $AVP_{new}$ , average variance of prediction for a new site; Pseudo- $R_{\delta}^2$ , fraction of the variability in the true skews explained by each model (Gruber and others, 2007)]

Model	Regression parameter $\beta_1$	$\sigma_{\delta}^2$	ASEV	$AVP_{new}$	Pseudo $R_{\delta}^2$ (percent)
CONSTANT:	-0.07	0.17	0.010	0.18	0
$\hat{\gamma} = \beta_1$	(0.10)	(0.022)			

### Bayesian Weighted Least-Squares/Bayesian Generalized Least-Squares Regression Diagnostics

To determine if a model is a good representation of the data and which regression parameters, if any, should be included in a regression model, diagnostic statistics have been developed to evaluate how well a model fits a regional hydrologic data set (Griffis, 2006; Gruber and others, 2008).

In this study, the goal was to determine the set of possible explanatory variables that best fit annual peak flows for the PNW study area affording the most accurate skew predictions while also keeping the model as simple as possible. This section presents the diagnostic statistics for a B-WLS/B-GLS analysis, and discusses the specific values obtained for the PNW regional skew study.

A Pseudo Analysis of Variance (Pseudo ANOVA) table for the PNW regional skew analysis is shown in table B3. The table contains regression diagnostics/goodness of fit statistics. In particular, the table describes how much of the variation in the observations can be attributed to the regional model and how much of the residual variation can be attributed to model error and sampling error, respectively. Difficulties arise in determining these quantities. The model errors cannot be resolved because the values of the sampling errors  $\eta_i$  for each site  $i$ , are not known. However, the total sampling error sum of squares can be described by its mean value,  $\sum_{i=1}^n Var[\hat{\gamma}_i]$ .

Because there are  $n$  equations, the total variation due to the model error  $\delta$  for a model with  $k$  parameters has a mean equal to  $n\sigma_{\delta}^2(k)$ . Thus, the residual variation attributed to the sampling error is  $\sum_{i=1}^n Var[\hat{\gamma}_i]$ , and the residual variation attributed to the model error is  $n\sigma_{\delta}^2(k)$ .

**Table B3.** Pseudo analysis of variance (ANOVA) for the Pacific Northwest CONSTANT regional skew model.

[**Abbreviations:**  $k$ , number of estimated regression parameters not including the constant;  $n$ , number of observations (streamgages) used in regression;  $\sigma_{\delta}^2(0)$ , model error variance of a constant model;  $\sigma_{\delta}^2(k)$ , model error variance of a model with  $k$  regression parameters and a constant;  $Var(\hat{\gamma}_i)$ , variance of the estimated sample skew at site  $i$ ; EVR, error variance ratio; MBV\*, misrepresentation of the beta variance; Pseudo -  $R_{\delta}^2$ , fraction of variability in the true skews explained by each model (Gruber and others, 2007); %, percent]

Source	Degrees-of-freedom		Sum-of-squares	CONSTANT model
	Equations	CONSTANT model	Equations	
Model	$k$	0	$n[\sigma_{\delta}^2(0) - \sigma_{\delta}^2(k)]$	0
Model error	$n-k-1$	289	$n[\sigma_{\delta}^2(k)]$	50
Sampling error	$n$	290	$\sum_{i=1}^n Var(\hat{\gamma}_i)$	35
Total	$2n-1$	579	$n[\sigma_{\delta}^2(k)] + \sum_{i=1}^n Var(\hat{\gamma}_i)$	85
EVR				0.7
MBV*				10
Pseudo - $R_{\delta}^2$				0%

For a model with no parameters other than the mean (that is, the constant skew model), the estimated model error variance  $\sigma_\delta^2(0)$  describes all of the anticipated variation in  $\gamma_i = \mu + \delta_i$ , where  $\mu$  is the mean of the estimated at-station sample skews. Thus, the total expected sum of squares variation due to model error  $\delta_i$  and sampling error  $\eta_i = \hat{\gamma}_i - \gamma_i$  in expectation should equal  $n\sigma_\delta^2(0) + \sum_{i=1}^n \text{Var}(\hat{\gamma}_i)$ . Therefore, the expected sum of squares attributed to a regional skew model with  $k$  parameters equals  $n[\sigma_\delta^2(0) - \sigma_\delta^2(k)]$ , because the sum of the model error variance  $n\sigma_\delta^2(k)$  and the variance explained by the model must sum to  $n\sigma_\delta^2(0)$ . Table B3 considers a model with  $k = 0$  (a constant model).

This division of the variation in the observations is referred to as a Pseudo ANOVA because the contributions of the three sources of error are estimated or constructed, rather than being determined from the estimated residual errors and the observed model predictions, while also ignoring the impact of correlation among the sampling errors.

Table B3 contains the Pseudo ANOVA results for the PNW CONSTANT model. The CONSTANT model does not have any explanatory variables, thus the variation attributed to the models is 0.

The Error Variance Ratio (EVR) is a modeling diagnostic used to evaluate if a simple OLS regression is sufficient, or a more sophisticated WLS or GLS analysis is appropriate. EVR is the ratio of the average sampling error variance to the model error variance. Generally, an EVR greater than 0.20, indicates that the sampling variance is not negligible when compared to the model error variance, suggesting the need for a WLS or GLS regression analysis. The EVR is estimated as

$$\text{EVR} = \frac{\text{SS}(\text{sampling error})}{\text{SS}(\text{model error})} = \frac{\sum_{i=1}^n \text{Var}(\hat{\gamma}_i)}{n\sigma_\delta^2(k)}. \quad (\text{B12})$$

For the PNW regional skew study area, EVR had a value of 0.7 for the CONSTANT model. The sampling variability in the sample skew estimators was larger than the error in the regional model. Thus an OLS model that neglects sampling error in the at-station skew estimators may not provide a statistically reliable analysis of the data. Given the variation of record lengths from streamgage-to-streamgage, it is important to use a WLS or GLS analysis to evaluate the final precision of the model, rather than a simpler OLS analysis.

The Misrepresentation of the Beta Variance (MBV\*) statistic is used to determine whether a WLS regression is sufficient, or if a GLS regression is appropriate to determine the precision of the estimated regression parameters (Griffis, 2006; Veilleux, 2011). The MBV\* describes the error produced by a WLS regression analysis in its evaluation of the precision of  $b_0^{WLS}$ , which is the estimator of the constant  $\beta_0^{WLS}$ , because the covariance among the estimated at-station skews

$\hat{\gamma}_i$  generally has its greatest impact on the precision of the constant term (Stedinger and Tasker, 1985). If the MBV\* is substantially greater than 1, then a GLS error analysis should be used. The MBV\* is estimated as,

$$\text{MBV}^* = \frac{\text{Var}[b_0^{WLS} \mid \text{GLS analysis}]}{\text{Var}[b_0^{WLS} \mid \text{WLS analysis}]} = \frac{w^T \Lambda w}{\sum_{i=1}^n w_i} \quad (\text{B13})$$

where

$$w_i = \frac{1}{\sqrt{\Lambda_{ii}}}.$$

For the PNW regional skew study areas, MBV\* had a value of 10 for the CONSTANT model. This is a large value indicating that the cross correlation among the skew estimators had an impact on the precision with which the regional average skew coefficient can be estimated; if a WLS precision analysis were used for the estimated constant parameter in the CONSTANT model, the variance would be underestimated by a factor of 10. Thus, a WLS analysis would misrepresent the variance of the constant in the CONSTANT model. Moreover, a WLS model would have resulted in underestimation of the variance of prediction, given that the sampling error in the constant term in both models was sufficiently large enough to make an appreciable contribution to the average variance of prediction.

## Leverage and Influence

Leverage and influence diagnostics statistics can be used to identify rogue observations and to effectively address lack-of-fit when estimating skew coefficients. Leverage identifies those streamgages in the analysis where the observed values have a large impact on the fitted (or predicted) values (Hoaglin and Welsh, 1978). Generally, leverage considers whether an observation, or explanatory variable, is unusual, and thus likely to have a large effect on the estimated regression coefficients and predictions. Unlike leverage, which highlights points which have the ability or potential to affect the fit of the regression, influence attempts to describe those points which do have an unusual impact on the regression analysis (Belsley and others, 1980; Cook and Weisberg, 1982; Tasker and Stedinger, 1989). An influential observation is one with an unusually large residual that has a disproportionate effect on the fitted regression relations. Influential observations often have high leverage. For a detailed description of the equations used to determine leverage and influence for a B-WLS/B-GLS analysis see Veilleux (2011) and Veilleux and others (2011).

For the B-WLS/B-GLS CONSTANT regional skew models for PNW, no streamgages had high leverage. The differences in leverage values for the constant model reflect the variation in record lengths among sites.

Nineteen streamgages in the B-WLS/B-GLS CONSTANT regional skew models for PNW have high influence, and thus have an unusual impact on the fitted regression relation. The 19 USGS streamgages with high influence in order of descending influence values are: 12354000, 12311000, 12358500, 12325500, 14353500, 14353000, 13336500, 14158790, 12043000, 12344000, 14091500, 13112000, 12342500, 12354500, 12039500, 12121600, 14032000, 14166500, and 13120000. The streamgages with the 7 largest, in magnitude, residuals are among these 19 streamgages. These 19 streamgages with high influence also have 19 of the 32 largest, in magnitude, residuals among the 290 streamgages used in the regional skew study.

## References Cited

- Belsley, D.A., Kuh, E., and Welsch, R.E., 1980, Regression diagnostics—identifying influential data and sources of collinearity, chap. 2: New York, John Wiley and Sons, Inc., p. 6–84.
- Cohn, T.A., Lane, W.L., and Baier, W.G., 1997, An algorithm, for computing moments-based flood quantile estimates when historical flood information is available: *Water Resources*, v. 33, no. 9, p. 2,089–2,096.
- Cohn, T.A., Lane, W.L., and Stedinger, J.R., 2001, Confidence intervals for Expected Moments Algorithm flood quantile estimates: *Water Resources Research*, v. 37, no. 6, p. 1,695–1,706.
- Cook, R.D., and Weisberg, S., 1982, Residuals and influence in regression: New York, Chapman and Hall, 230 p.
- Devore, J.L., 2004, Probability and statistics for engineering and the sciences: Belmont, Calif., Brooks/Cole, 795 p.
- Eash, D.A., Barnes, K.K., and Veilleux, A.G., 2013, Methods for estimating annual exceedance-probability discharges for streams in Iowa, based on data through water year 2010: U.S. Geological Survey Scientific Investigations Report 2013–5086, 63 p. with appendix, accessed June 10, 2015, at <http://pubs.usgs.gov/sir/2013/5086/>.
- Feaster, T.D., Gotvald, A.J., and Weaver, J.C., 2009, Magnitude and frequency of rural floods in the Southeastern United States, 2006—Volume 3, South Carolina: U.S. Geological Survey Scientific Investigations Report 2009–5156, 226 p.
- Gotvald, A.J., Feaster, T.D., and Weaver, J.C., 2009, Magnitude and frequency of rural floods in the southeastern United States, 2006—Volume 1, Georgia: U.S. Geological Survey Scientific Investigations Report 2009–5043, 120 p.
- Griffis, V.W., 2006. Flood Frequency Analysis—Bulletin 17, regional information, and climate change: Ithaca, New York, Cornell University, Ph.D. dissertation, 246 p.
- Griffis, V.W., and Stedinger, J.R., 2009, Log-Pearson Type 3 distribution and its application in flood frequency analysis. III—Sample skew and weighted skew estimators: *Journal of Hydrologic Engineering*, v. 14, no. 2, p. 121–130.
- Griffis, V.W., Stedinger, J.R., and Cohn, T.A., 2004, Log Pearson type 3 Quantile estimators with regional skew information and low outlier adjustments: *Water Resources Research*, v. 40, W07503, 17 p., accessed June 10, 2014, at <http://dx.doi.org/1029/2003WR002697>.
- Gruber, A.M., Reis, D.S., Jr., and Stedinger, J.R., 2007, Models of regional skew based on Bayesian GLS regression, Paper 40927-3285, in Kabbes, K.C., ed., Restoring our natural habitat—Proceedings of the World Environmental and Water Resources Congress, May 15–18, 2007, Tampa, Florida: American Society of Civil Engineers, Reston, Virginia, 10 p.
- Gruber, A.M., and Stedinger, J.R., 2008, Models of LP3 regional skew, data selection and Bayesian GLS regression, Paper 596, in Babcock, R.W., and Watson, R., eds., World Environmental and Water Resources Congress: Ahupua‘a, Honolulu, Hawai‘i, May 12–16, 2008, 10 p.
- Hoaglin, D.C., and Welsch, R.E., 1978, The Hat Matrix in regression and ANOVA: *The American Statistician*, v. 32, no. 1, p. 17–22.
- Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency: Hydrology Subcommittee Bulletin 17B, 28 p., 14 appendices, 1 plate.
- Lamontagne, J.R., Stedinger, J.R., Berenbrock, Charles, Veilleux, A.G., Ferris, J.C., and Knifong, D.L., 2012, Development of regional skews for selected flood durations for the Central Valley Region, California, based on data through water year 2008: U.S. Geological Survey Scientific Investigations Report 2012–5130, 60 p.
- Martins, E.S., and Stedinger, J.R., 2002, Cross-correlation among estimators of shape: *Water Resources Research*, v. 38, no. 11, p. 34-1–34-7, accessed January 13, 2013, at doi: 10.1029/2002WR001589.
- Parrett, C., Veilleux, A., Stedinger, J.R., Barth, N.A., Knifong, D.L., and Ferris, J.C., 2011, Regional skew for California, and flood frequency for selected sites in the Sacramento–San Joaquin River Basin, based on data through water year 2006: U.S. Geological Survey Scientific Investigations Report 2010–5260, 94 p.

- Reis, D.S., Jr., Stedinger, J.R., and Martins, E.S., 2005, Bayesian generalized least squares regression with application to the log Pearson type III regional skew estimation: *Water Resources Research*, v. 41, W10419, 14 p., accessed February 14, 2013, at <http://dx.doi.org/10.1029/2004WR003445>.
- Seaber, P.R., Kapinos, P., Knapp, G.L., 1987, Hydrologic unit maps: U.S. Geological Survey Water-Supply Paper 2294, 66 p.
- Southard, R.E., and Veilleux, A.G., 2014, Methods for estimating annual exceedance-probability discharges and largest recorded floods for unregulated streams in rural Missouri: U.S. Geological Survey Scientific Investigations Report 2014–5165, 39 p., accessed November 14, 2014, at <http://dx.doi.org/10.3133/sir20145165>.
- Stedinger, J.R., and Cohn, T.A., 1986, Flood frequency analysis with historical and paleoflood information: *Water Resources Research*, v. 22, no. 5, p. 785–793.
- Stedinger, J.R., and Griffis, V., 2008, Flood frequency analysis in the United States—Time to update: *Journal of Hydrologic Engineering*, v. 13, no. 4, p. 199–204.
- Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis, 1. Ordinary, weighted and generalized least squares compared: *Water Resources Research*, v. 21, no. 9, p. 1,421–1,432.
- Tasker, G.D., and Stedinger, J.R., 1986, Regional skew with weighted LS regression: *Journal of Water Resources Planning and Management, ASCE*, v. 112, no. 2, p. 225–237.
- Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: *Journal of Hydrology*, v. 111, no. (1–4), p. 361–375.
- U.S. Geological Survey, 2015, National Water Information System database: U.S. Geological Survey National Water Information System, accessed April 20, 2015, at <http://waterdata.usgs.gov/nwis/>.
- Veilleux, A.G., 2011, Bayesian GLS regression, leverage and influence for regionalization of hydrologic statistics: Ithaca, New York, Cornell University, Ph.D. dissertation, 184 p.
- Veilleux, A.G., Cohn, T.A., Flynn, K.M., Mason, R.R., Jr., and Hummel, P.R., 2014, Estimating magnitude and frequency of floods using the PeakFQ 7.0 program: U.S. Geological Survey Fact Sheet 2013–3108, 2 p., accessed November 2, 2014, at <http://dx.doi.org/10.3133/fs20133108>.
- Veilleux, A.G., Stedinger, J.R., and Eash, D.A., 2012, Bayesian WLS/GLS regression for regional skewness analysis for regions with large crest stage gage networks *in* Loucks, E.D., ed., *Proceedings World Environmental and Water Resources Congress—Crossing Boundaries*, Albuquerque, N. Mex., May 20–24, 2012: American Society of Civil Engineers, p. 2,253–2,263, accessed November 14, 2013, at <http://dx.doi.org/10.1061/9780784412312.227>.
- Veilleux, A.G., Stedinger, J.R., and Lamontagne, J.R., 2011, Bayesian WLS/GLS regression for regional skewness analysis for regions with large cross-correlations among flood flows, paper 1303, *in* *World Environmental and Water Resources Congress 2011—Bearing knowledge for sustainability*, Palm Springs, Calif., May 22–26, 2011: American Society of Civil Engineers, p. 3,103–3,112.
- Weaver, J.C., Feaster, T.D., and Gotvald, A.J., 2009, Magnitude and frequency of rural floods in the Southeastern United States, through 2006—Volume 2, North Carolina: U.S. Geological Survey Scientific Investigations Report 2009–5158, 111 p., accessed November 13, 2013, at <http://pubs.usgs.gov/sir/2009/5158/>.

Publishing support provided by the U.S. Geological Survey  
Science Publishing Network, Tacoma Publishing Service Center  
For more information concerning the research in this report, contact the  
Director, Idaho Water Science Center  
U.S. Geological Survey  
230 Collins Road  
Boise, Idaho 83702  
<http://id.water.usgs.gov>

