![USGS logo] USGS
science for a changing world

Magnitude, Frequency, and Trends of Floods at Gaged and Ungaged Sites in Washington, Based on Data through Water Year 2014

Scientific Investigations Report 2016–5118
Version 1.2, November 2017

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

Magnitude, Frequency, and Trends of Floods at Gaged and Ungaged Sites in Washington, Based on Data through Water Year 2014

By Mark C. Mastin, Christopher P. Konrad, Andrea G. Veilleux, and Alison E. Tecca

Scientific Investigations Report 2016–5118
Version 1.2, November 2017

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

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.

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
Length		
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
Area		
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
Volume		
cubic foot (ft^3)	0.02832	cubic meter (m^3)
Flow rate		
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
cubic foot per second per square mile [(ft^3/s)/mi^2]	0.01093	cubic meter per second per square kilometer [(m^3/s)/km^2]
cubic foot per day (ft^3/d)	0.02832	cubic meter per day (m^3/d)
inch per year (in/yr)	25.4	millimeter per year (mm/yr)

Datums

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

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

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

Abbreviations

AEP	annual exceedance probability
AR	atmospheric river
EMA	Expected Moment Algorithm
FEMA	Federal Emergency Management Agency
GIS	Geographical Information System
GLS	generalized-least-squares
HUCs	hydrologic unit codes
IACWD	Interagency Advisory Committee on Water Data
MGB	Multiple Grubbs-Beck test for low outliers
MSE	mean square error
NFF	National Flood Frequency
NHD	National Hydrologic Dataset
NWIS	USGS National Water Information System
PDO	Pacific Decadal Oscillation
PeakFQ	USGS software Peak flow FreQuency analysis program
PILF	potentially influential low flow
SOI	Southern Oscillation Index
USGS	U.S. Geological Survey
WIE	Weighted Independent Estimates
WREG	Weighted-multiple-linear REGression

Magnitude, Frequency, and Trends of Floods at Gaged and Ungaged Sites in Washington, Based on Data through Water Year 2014

By Mark C. Mastin, Christopher P. Konrad, Andrea G. Veilleux, and Alison E. Tecca

Abstract

An investigation into the magnitude and frequency of floods in Washington State computed the annual exceedance probability (AEP) statistics for 649 U.S. Geological Survey unregulated streamgages in and near the borders of Washington using the recorded annual peak flows through water year 2014. This is an updated report from a previous report published in 1998 that used annual peak flows through the water year 1996. New in this report, a regional skew coefficient was developed for the Pacific Northwest region that includes areas in Oregon, Washington, Idaho and western Montana within the Columbia River drainage basin south of the United States-Canada border, the coastal areas of Oregon and western Washington, and watersheds draining into Puget Sound, Washington. The skew coefficient is an important term in the Log Pearson Type III equation used to define the distribution of the log-transformed annual peaks. The Expected Moments Algorithm was used to fit historical and censored peak-flow data to the log Pearson Type III distribution. A Multiple Grubb-Beck test was employed to censor low outliers of annual peak flows to improve on the frequency distribution. This investigation also includes a section on observed trends in annual peak flows that showed significant trends (p-value < 0.05) in 21 of 83 long-term sites, but with small magnitude Kendall *tau* values suggesting a limited monotonic trend in the time series of annual peaks. Most of the sites with a significant trend in western Washington were positive and all the sites with significant trends (three sites) in eastern Washington were negative.

Multivariate regression analysis with measured basin characteristics and the AEP statistics at long-term, unregulated, and un-urbanized (defined as drainage basins with less than 5 percent impervious land cover for this investigation) streamgages within Washington and some in Idaho and Oregon that are near the Washington border was used to develop equations to estimate AEP statistics at ungaged basins. Washington was divided into four regions to improve the accuracy of the regression equations; a set of equations for eight selected AEPs and for each region were constructed. Selected AEP statistics included the annual peak flows that equaled or exceeded 50, 20, 10, 4, 2, 1, 0.5 and 0.2 percent of the time equivalent to peak flows for peaks with a 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence intervals, respectively. Annual precipitation and drainage area were the significant basin characteristics in the regression equations for all four regression regions in Washington and forest cover was significant for the two regression regions in eastern Washington. Average standard error of prediction for the regional regression equations ranged from 69.12 to 119.59 percent for Regression Regions 1 and 2 on the eastern side of the Cascade Mountains and from 43.22 to 58.04 percent for Regression Regions 3 and 4 on the western side of the Cascade Mountains. The pseudo coefficient of determination (where a value of 100 signifies a perfect regression model) ranged from 68.66 to 90.66 for Regression Regions 1 and 2, and 92.35 to 95.44 for Regions 3 and 4.

The calculated AEP statistics for the streamgages and the regional regression equations are expected to be incorporated into StreamStats after the publication of this report. StreamStats is the interactive Web-based map tool created by the U.S. Geological Survey to allow the user to choose a streamgage and obtain published statistics or choose ungaged locations where the program automatically applies the regional regression equations and computes the estimates of the AEP statistics.

Introduction

The planning of infrastructure or developments near water courses such as bridges, homes, culverts, levees and dams require the knowledge of the magnitude and frequency of large floods that may interact with the planning site. An accurate estimate of the magnitude of the flood for a particular frequency, sometimes referred to as a design flood, is needed to balance safety in the design with cost. For example, the Federal Emergency Management Agency (FEMA) through their National Flood Insurance Program generally requires developments to be located above the water level of a flood that occurs on average once every 100 years or has a 1 percent chance to be equaled or exceeded during any 1 year. Mapping these areas inundated by this 1-percent flood, commonly referred to as floodways or flood zones, requires an estimate of the 1-percent flood magnitude. Planners, engineers, and developers need accurate and easily applicable methods to estimate flood magnitudes for their sites of interest.

Based on the history of peak flows recorded at long-term streamgages, an estimate of the magnitude of floods for various frequencies can be made and used for the design flood for a particular floodway analysis. When a project area is not near a long-term streamgage and an estimate of the design flood is needed, regression equations based on basin characteristics can provide a reasonably accurate estimate.

The U.S. Geological Survey (USGS) has a long history of measuring streamflow and maintaining a public database of streamflow data for streamgages, which includes the instantaneous annual maximum streamflow (hereinafter called peak flow) for each water year[1]. This USGS network of streamgages in Washington and the surrounding States is the sole sources of annual peak-flow information used in this report.

The terminology for expressing a flood frequency in this report is annual exceedance probability (AEP). A flood discharge with a 0.02 AEP or 2 percent AEP for a location on a river would equal or exceed the discharge on average twice every 100 years or have a 2 percent chance of equaling or exceeding the discharge in any given year. Often these design floods are described by its recurrence interval or return interval, *T*, such as the *T*-year flood or the 50-year flood. The AEPs are the inverse of the return interval, that is, the 100-year flood is the same as the 0.01 AEP flood or 1 percent AEP flood and the 10-year flood is equivalent to the 0.1 AEP or 10 percent AEP flood (table 1). Flood magnitudes for various frequencies or AEPs also are collectively termed flood quantiles.

Table 1.   Annual exceedance probabilities and corresponding return intervals for flood frequencies discussed in this report.

Annual exceedance probability (percent)	Return interval (years)
50	2
20	5
10	10
4	25
2	50
1	100
0.5	200
0.2	500

Purpose and Scope

The flood frequency analyses in this report were compiled for 649 USGS streamgages in and near Washington State with at least 10 years of non-zero annual peak-flow data that were considered unregulated. Many streamgages have small upstream diversions that are usually used for irrigation during the low-flow season. For this report, these streamgages with small diversions were not considered regulated. Only when the regulation was believed to affect the magnitude of the peak flows was the streamgage categorized as regulated and not used in the analyses. For example, the streamgage on the Elwha River at McDonald Bridge near Port Angeles, Washington (USGS streamgage 12045500, map ID 69, table 2) is located downstream of the former Glines Canyon Dam, which was removed in 2012. Glines Canyon Dam was operated as a run-of-the-river dam which means that outflows from the dam were typically equal to the inflows and thus the dam had minimal effect on peak flows, so the streamgage was considered unregulated and the record was used without adjustment for flood frequency analysis. Most streamgages with dams located upstream were considered regulated and not used in the flood-frequency analyses. Generally, the complete record of peak flows at a streamgage was used for the frequency analysis.

Table 2.   Expected Moments Algorithm input summary for all U.S. Geological Survey streamgages analyzed for this investigation.

[Table 2 is an Excel® file available for download at http://dx.doi.org/10.3133/sir20165118]

---

[1]A water year is the 12-month period from October 1 through September 30 and is designated by the year in which it ends (for example, water year 2014 begins October 1, 2013, and ends September 30, 2014).

Frequency analyses to compute flood quantiles were performed using guidelines published by the Interagency Advisory Committee on Water Data (1982) which suggests that the frequency distribution for a streamgage is improved when the skew coefficient determined from the peak-flow data is weighted with a regional skew coefficient. The Interagency Advisory Committee on Water Data provides a regional skew coefficient; however, a new regional skew coefficient was calculated as part of this study for the Pacific Northwest region, a much expanded study area over that of the rest of the report which includes streamgages from Washington, Oregon, Idaho, and western Montana. The development of the regional skew coefficient is described in the appendix A.

Regression equations that relate basin characteristics to flood quantiles were developed for hydrologic regions within Washington to provide estimates of flood quantiles at ungaged sites. These regression equations are valid for unregulated basins within the range of the basin parameters used to develop the equation. Regression equations were developed for eight selected AEPs (50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent) and for the four regression regions (Regression Regions 1–4) resulting in a total of 32 regression equations. The equations and methodology to develop them are discussed in this report.

The flood-frequency analysis includes the assumption of stationarity or the idea that the hydrologic conditions that generated the annual peak flows in the recent past will be the same in the future. With the understanding that the climate is changing, this assumption of stationarity is not completely valid. However, our understanding of how climate change will affect peak flows is not well understood and difficult to apply to a flood-frequency analysis. No adjustments were made for climate change in the flood-frequency analysis as described in this report, but some observations of the trends of annual peak flows as seen in the observed record are provided as an indication of change in peak flows over the last 100 years.

This flood-frequency information is expected to be made available through a user-interactive, Web-based program called StreamStats. Users can click on any point on a drainage feature and the program will compute the drainage area, basin characteristics, and the estimated flood quantiles based on the regional regression equations developed as part of this investigation. Access the StreamStats program for more information at http://water.usgs.gov/osw/streamstats/index.html.

Previous U.S. Geological Survey Flood Reports for Washington

Previous flood reports for Washington include descriptive as well as quantitative reports that compute or estimate flood quantiles. Those reports also vary by various spatial levels of coverage from the Columbia Basin to state-wide studies, to individual drainage basins, and down to individual floods on a river. Reports for specific drainage basins include studies by Anderson (1948) on the floods of the Puyallup and Chehalis River Basins, Anderson and Bodhaine (1957) on the flood of 1956 in the Esquatzel Coulee area, Bailey (1960) on floods of the Nooksack River, Richardson (1965) on the effect of logging on flooding in the upper Green River Basin, and Mastin (1998) on flooding of South Prairie Creek. The Skagit River flood history was described in detail by Stewart and Bodhaine (1961).

The floods of January 1974 in Washington are described by Longfield (1974). Hubbard (1991) described the January 1990 floods in northwestern Oregon and southwestern Washington and again Hubbard (1994) described the November 1990 floods in western Washington.

A series of USGS fact sheets have been produced for the general public in Washington on understanding flood data and causes of flooding. Jones (2011) described the "popular myths" about flooding in western Washington, and Knowles and Sumioka (2001) described the National Flood-Frequency Program and how to estimate flood magnitude and frequency in rural Washington. Konrad and Booth (2002) described hydrologic trends associated with urban development in western Washington followed by Konrad (2003) on the effects of urban development on floods. Kresch and Dinicola (1997) produced a fact sheet on "What Causes Floods in Washington State?"

Studies of regional or state-wide flood frequency in Washington include an early report by Rantz and Riggs (1949) for the Columbia River Basin. Bodhaine and Robinson (1952) computed flood frequencies and magnitudes for western Washington and derived a regression equation for the mean annual flood from four basin characteristics—drainage area, mean altitude of drainage basin, geographic factor, and area of lakes and ponds. Haushild (1979) created regression equations for estimating flood quantiles for small ephemeral streams in eastern Washington. An early state-wide flood frequency and magnitude report was completed by Bodhaine and Thomas (1960) and they again published a flood frequency and magnitude report for the Pacific slope basins in Washington and upper Columbia River Basin in 1964 (Bodhaine and Thomas, 1964). Cummans and others (1975) completed a state-wide flood frequency and magnitude study that defined a set of equations for estimating flood quantiles at ungaged sites for the western and eastern Washington. From these two regions, they plotted the residuals from the regression equation to redefine Washington into 12 regions, each with a separate set of regression equations to estimate flood quantiles based on basin characteristics of annual precipitation, drainage area, and for eastern Washington, forest cover. A similar report by Sumioka and others (1998) divided the State into nine geographic regions based on hydrologic-unit code boundaries and created sets of regression equations to estimate flood quantiles for each of the regions. The set of regression equations from this 1998 study was incorporated into StreamStats.

Study Area

The study area for this report is primarily within the State of Washington, but includes the border regions in Idaho and Oregon. The regional skew study discussed in the appendix A extends beyond this area to include western Montana, and most of Idaho and Oregon.

The topography of Washington is quite diverse and can be divided into seven physiographic provinces (Livingston, 1969). These include (1) the Blue Mountains in the extreme southeast, the northeastern end of a mountain range that extends from Oregon; (2) the Cascade Mountains, the north-south mountain spine that divides Washington into two distinct regions, eastern and western Washington, consisting of several prominent active volcanoes; (3) the Columbia Basin (sometimes called the Columbia Plateau) in east-central Washington and consisting of gently rolling terrain bounded by the Snake River to the south and the Columbia River to the north and west; (4) the Okanogan Highlands in the northeastern corner of the State with elevations above 7,000 ft and relatively gentle terrain that slopes gradually to the south; (5) the Olympic Mountains in the northwest, with elevations of about 8,000 ft and rivers draining from the center of the range to all directions; (6) the Puget Lowlands, extending from the Canadian border south with a narrow area farther extending to the Columbia River and shaped by Pleistocene continental glaciation that left a layer of glacial till and outwash gravels with little bedrock exposed; and (7) the Willapa Hills in the southwest, with steep hills at about 2,000 ft surrounded by flat ocean beaches to the west, the lower Chehalis River valley to the north, and the Columbia River to the south (fig. 1).

Land use varies dramatically from eastern to western Washington. Eastern Washington consists of irrigated agriculture and sparsely populated areas with the exception of Spokane near the eastern border, Yakima in the center of the State, and the Tri-cities of Kennewick, Pasco, and Richland near the southern border. In western Washington, lush forest support timber harvesting and major population centers including the Olympia-Tacoma-Seattle-Everett corridor near Puget Sound, Bellingham to the north, and Vancouver on the southern border across the Columbia River from Portland, Oregon.

The Pacific Ocean creates a marine Mediterranean climate of wet winters and dry summers in western Washington. The Cascade Mountains are a significant barrier from the wet weather systems that generally come from the west. Orographic precipitation caused by these mountains results high precipitation amounts more than 100 in/yr in the high elevations (fig. 2). In the orographic shadow of the Cascade Mountains, much drier conditions exist in eastern Washington with annual precipitation less than 10 in/yr in much of the eastern interior.

Flood Hydrology

Floods in Washington are generated from several sources. Most of the floods in western Washington are generated by fall and winter rainstorms that generally come from the west. With warm air temperatures and a pre-existing snowpack, the floods may be augmented with rain-on-snow snowmelt. High elevation basins often receive precipitation as snow that accumulates during the winter and may not contribute to flooding until warmer temperatures in spring or summer rapidly melts the snow. In eastern Washington, intense thunderstorms that usually occur in the summer can produce extremely heavy but localized precipitation that generates relatively large floods in small basins. Thunderstorm-derived floods in small basins on the eastern side of the State have resulted in the highest recorded unit runoff (peak discharge divided by the drainage area) of floods in Washington (Crippen and Bue, 1977).

Floods can occur any time of the year, but a regular seasonal pattern can be seen for several regions of the State (fig. 3). Rainfall-generated peaks in the winter dominate the pattern for western Washington except for small, high-elevation basins that receive most of their precipitation as snow. On the eastern side of the State, snowmelt-generated peaks are most common with basins whose headwaters lie in the high elevations of the Cascade Mountains. These peak flows are delayed until summer whereas low-elevation basins more commonly experience peaks in the spring.

In western Washington, most of the largest floods are associated with a weather phenomenon known as atmospheric rivers (ARs) that make landfall in Washington. These ARs are characterized by long, narrow plumes of warm, moist air originating in the tropical regions of the Pacific. Neiman and others (2011) analyzed the annual peak daily runoff for four major watersheds in western Washington and compared the timing of the flooding with ARs that made landfall in Washington between water years 1998 and 2009 when satellite imagery was available to classify cyclones as ARs by their integrated water volume. Nieman and others (2011) determined that 46 of the 48 annual peak-flow events coincided with landfall ARs.

Land cover and soil conditions can influence the amount of runoff that occurs with inputs of precipitation. Urbanization can transform formerly porous land cover into impervious land covers with the construction of roads and rooftops leading to a greater percentage of precipitation becoming runoff which is funneled quickly to the river resulting in flashier, higher-magnitude floods than would have resulted in undisturbed watersheds. Antecedent soil-moisture content also can affect the percentage of precipitation becoming runoff. Early fall storms with heavy precipitation often do not result in large floods because the soil-moisture deficit resulting from the preceding dry summer needs to be replenished before significant runoff can be generated. If a

fall or winter storm quickly follows a previous storm, it is likely that the soil-moisture storage is near capacity, and a greater percentage of the rainfall becomes runoff and floods are generated. Mastin (2009) showed how frozen ground can generate large floods with relatively little rainfall in the Crab Creek Basin in eastern Washington. The frozen soils restrict infiltration of water at the surface, and therefore, most of the rain falling on the surface quickly becomes surface runoff that flows to the creeks and rivers. Vegetation cover is important in flood hydrology because as much as 0.2 in. of precipitation can be stored in the foliage during an individual storm, and accumulated over a year. As a result, 10–50 percent of annual precipitation can be intercepted that is later evaporated and lost to contributing to runoff (Kresch and Dinicola, 1997).
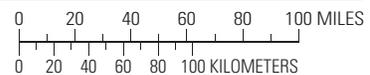
Figure 1.   Physiographic regions in Washington.
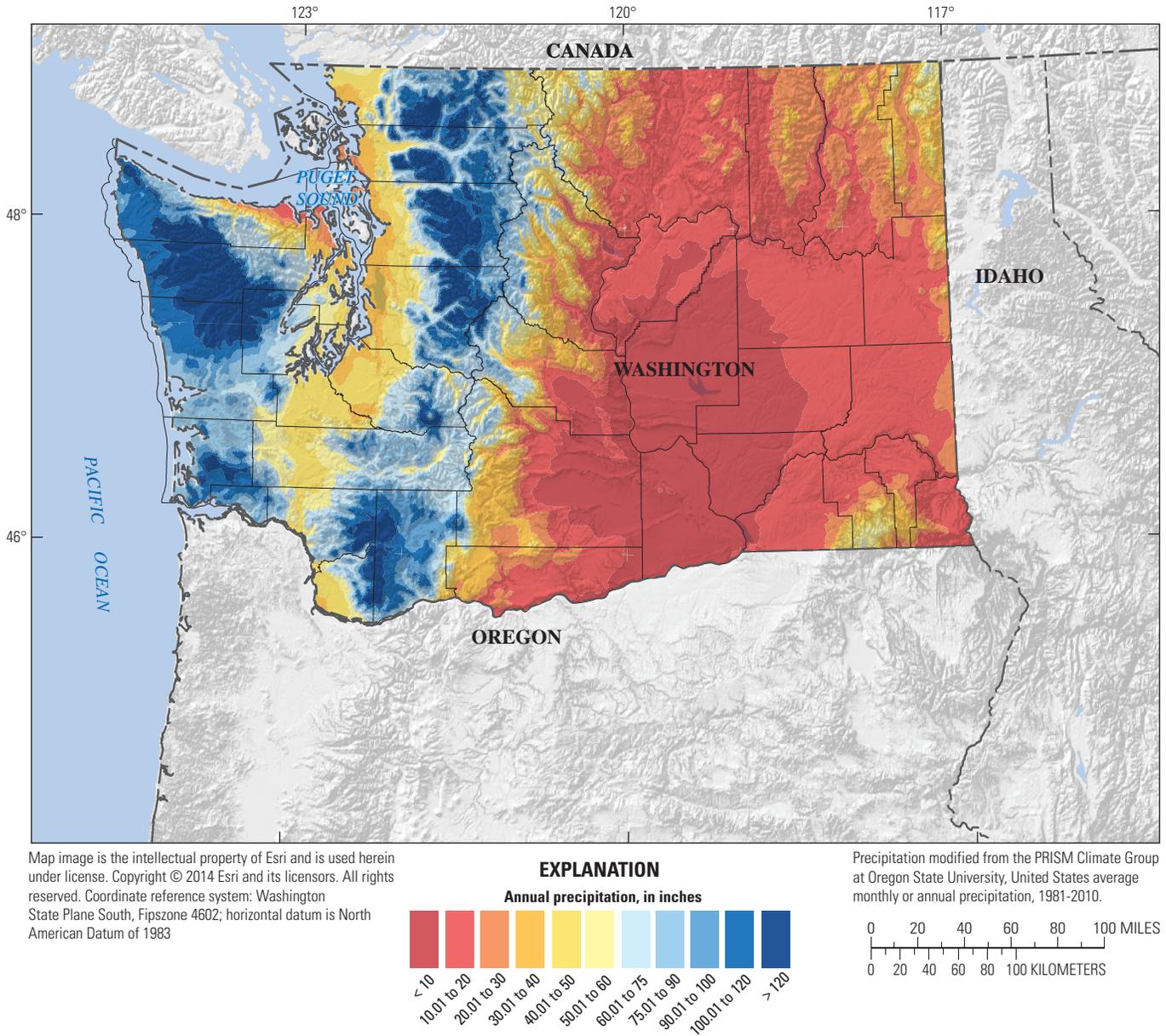
Figure 2.   Annual precipitation in Washington.
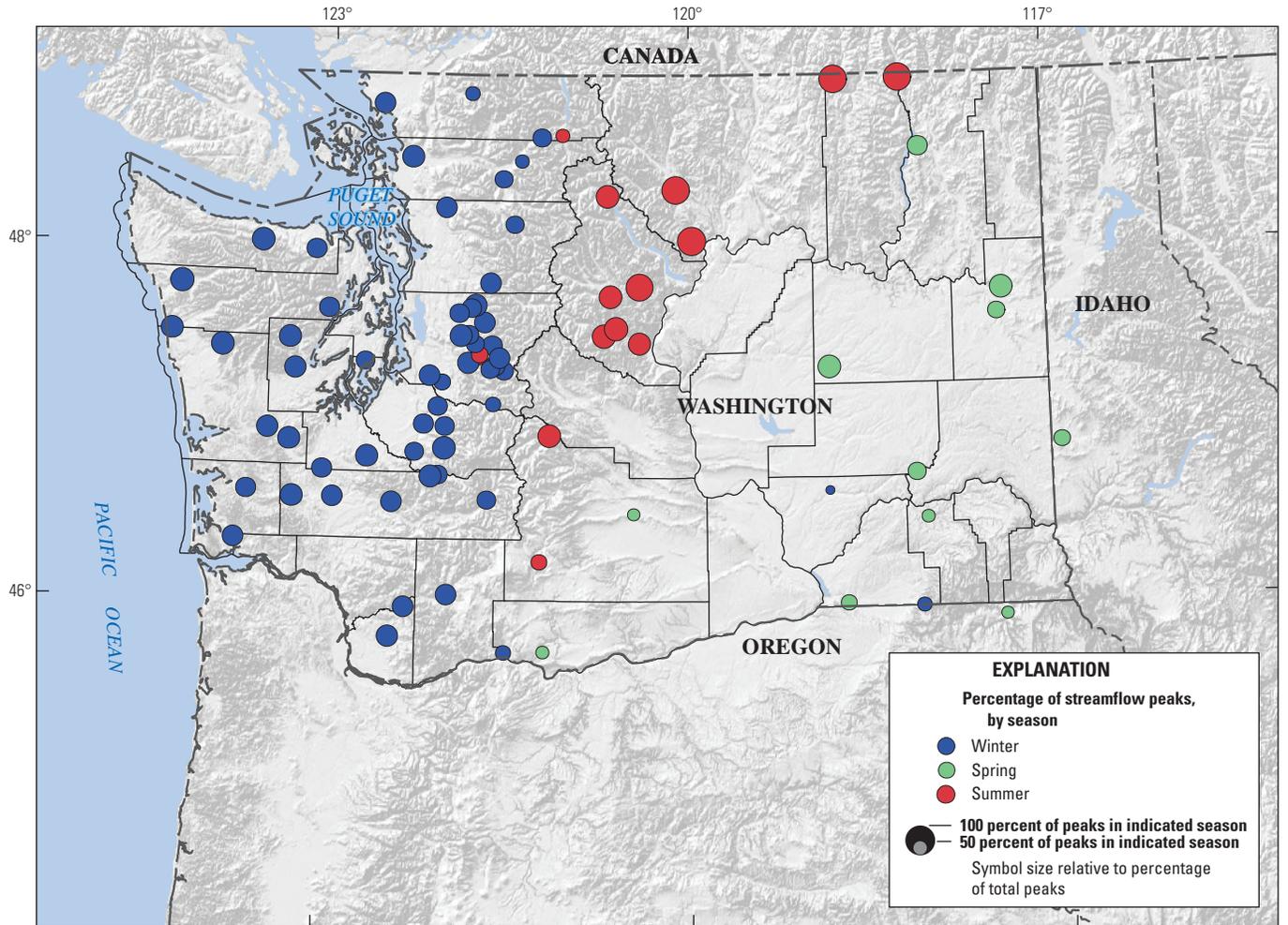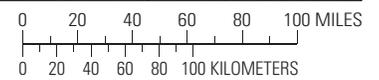
Figure 3. Seasonality of the annual peak streamflows for non-urban, unregulated streamgage sites with more than 50 years of annual peak-flow records through water year 2012, Washington. Season is the period when most peaks occur for unregulated, non-urban gages with 50 or more years of record. Spring is February through April; summer is May through July; and winter is November through January.

Trends in Peak Flows

The current methods for performing flood-frequency analyses assume that the past climate will be similar to future climates (stationarity). This assumption allows researchers to estimate the flood quantiles from past streamflow records and apply them to the future without adjustments. Human disturbance (for example, reservoirs, urbanization, and greenhouse gases) as well as natural climate changes and low-frequency variability (for example, Pacific Decadal Oscillation, Mantua and others [1997]) can alter the hydrology of a river and its frequency and magnitude of flooding to something different than the probability defined by record of observations (Jain and Lall, 2001). Current climate science tells us that substantial anthropogenic change to the Earth climate has altered the hydrology of rivers (Milly and others, 2008) and will continue to alter hydrology such that stationarity cannot be assumed. Future trends in flooding are difficult to estimate and apply to traditional flood-frequency analysis, and it is beyond the scope of this report to incorporate those uncertain trends into the frequency analysis. For this report, peak flows are assumed to be stationary for our estimation of flood quantiles; however, peak-flow trends were examined as part of this investigation.

Some regional climate model simulations for future scenarios show "substantial" increases in flood risk in the Pacific Northwest by mid-21st century (Salathe and others, 2014). The climate scenarios project future warming conditions and transitional basins, basins such as those in the Cascade Mountains, where a small change in air temperature significantly effects the amount of basin area receiving snow or rainfall during a storm, have the highest increase in flooding for the simulated future climate scenario (Tohver and others, 2014). Global Circulation Models (GCMs) also are simulating future atmospheric rivers to have higher integrated water vapor resulting in more days of extreme precipitation along the West Coast of North America and implying a future of more extreme flooding in western Washington (Warner and others, 2014). Although the GCMs suggest an increasing trend in flooding, the trend may not be evident in the observed record. Floods vary significantly from year to year, and trends are difficult to discern in a short record. Additionally, various long-term persistence trends may be occurring that are poorly understood, which may cloud the significance of a trend calculated for the short observed record (Cohn and Lins, 2005).

Peak flows are extremely variable and peak-flow trends can vary substantially from site to site based on record length and period of record, and can be affected by decadal changes in atmospheric rivers and land use. Peak flows may have decadal trends superimposed on long-term trends and active or current trends that may extend back several decades; or, may only be significant trends for the past decade.

Long-Term Trends in Floods

Streamflow data were analyzed for long-term trends in the magnitude and frequency of peaks for streamgages on unregulated streams in Washington having 50 or more years of peak-flow data. Peak-flow data were analyzed for increasing or decreasing magnitudes and the number of independent peaks above a base level was analyzed for changes in the frequency of floods.

The trend in the magnitude of peak flows was examined by compiling two-sided Kendall's *tau* and *p*-values (the probability that the null hypothesis of no trend is not rejected) for systematic peaks for 83 streamgages. A Kendall *tau* of 1 indicates a perfectly increasing trend and a -1 indicates a perfectly decreasing trend. Because of the highly variable nature of annual peak flows, no strong trends were noted and only one site had a Kendall *tau* value greater than 0.3 (USGS streamgage 12167000, *tau* = 0.41) and the lowest value was -0.28. Of the 83 sites, 21 sites had a *p*-value equal or less than 0.05. Of those 21 sites with a *p*-value equal or less than 0.05, 5 sites indicated that they have negative trends and 16 had positive trends. In terms of the spatial context, most sites (3 of 5) with significant negative trends were on the eastern side of the Cascade Mountains and all sites with increasing trends were on the western side of the Cascade Mountains (fig. 4).

Long-term trends in the frequency of floods were analyzed by compiling the number of times an independent peak exceeded a base discharge each water year for 65 unregulated streamgages having 50 or more years of record in Washington. The independent peaks were determined using the daily mean discharge record; an independent peak is defined by a mean daily discharge that exceeds the base discharge and there is as least one day before and after that peak when the daily mean discharge is at least 75 percent lower. The base discharge for this analysis is defined as the computed discharge for the 0.5 AEP, using a log Pearson Type III distribution that used the station skew coefficient. The definitions for independent peaks used in this study are similar to definition used by the USGS Water Mission Area to determine independent peaks above a base discharge for instantaneous peaks (Novak, 1985, p. 93). Examination of the records shows that 26 of the 65 streamgages had data as far back as water year 1930. Before 1930, the number of stations with data decreases sharply; thus, water years 1930–2014 was established as the period for this analysis. The total number of peaks above base discharge for all 65 streamgages for each water year was summed and then divided by the number of stations with data for the particular year to determine the average number of peaks above base discharge for all sites (fig. 5).
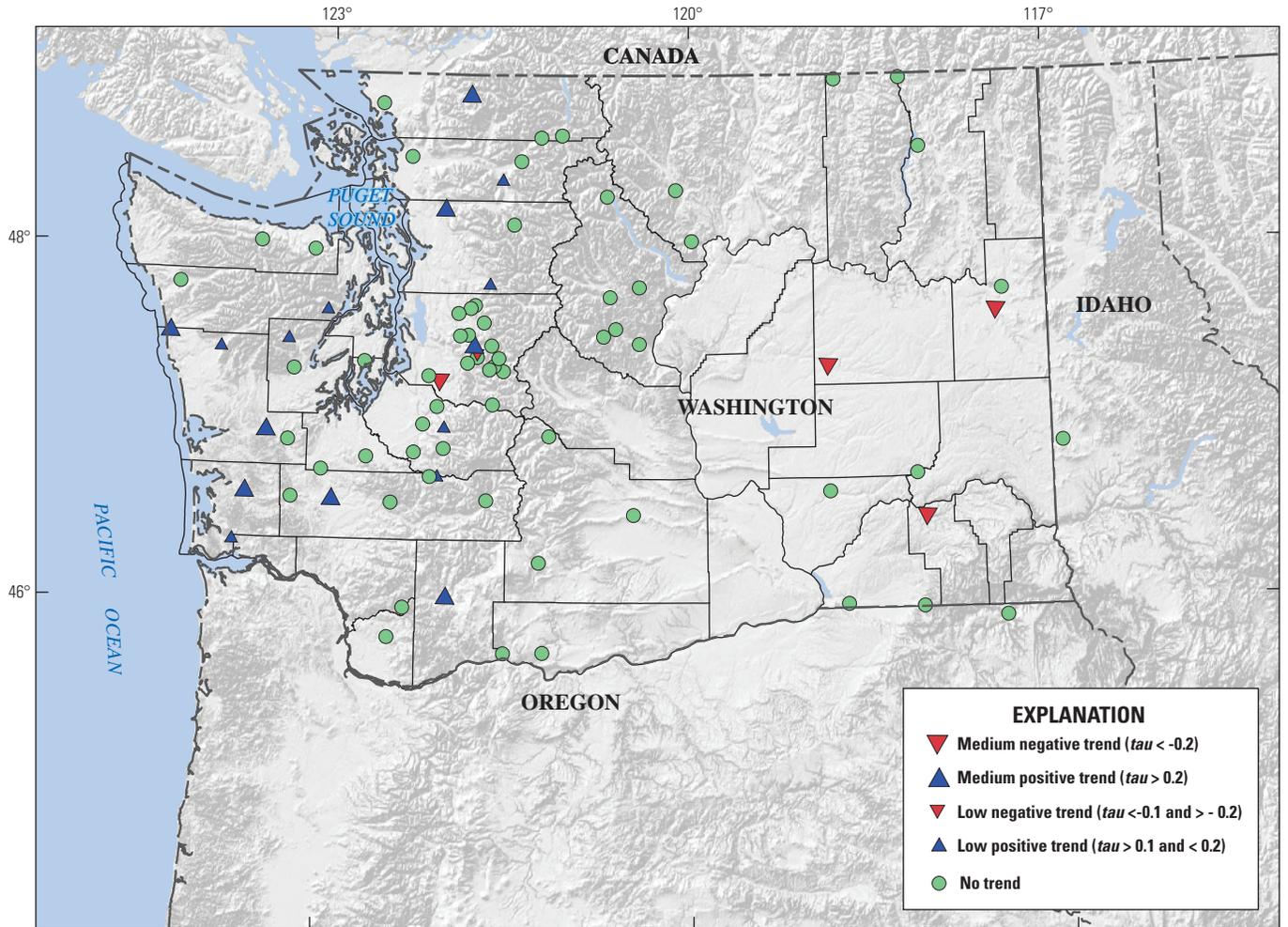
Figure 4.   Trends in the magnitude of annual peak flows for streamgages with 50 or more years of record through water year 2012, Washington. Sites defined as having no trend had less than 0.05 *p*-value in a Kendall's *tau* analysis.
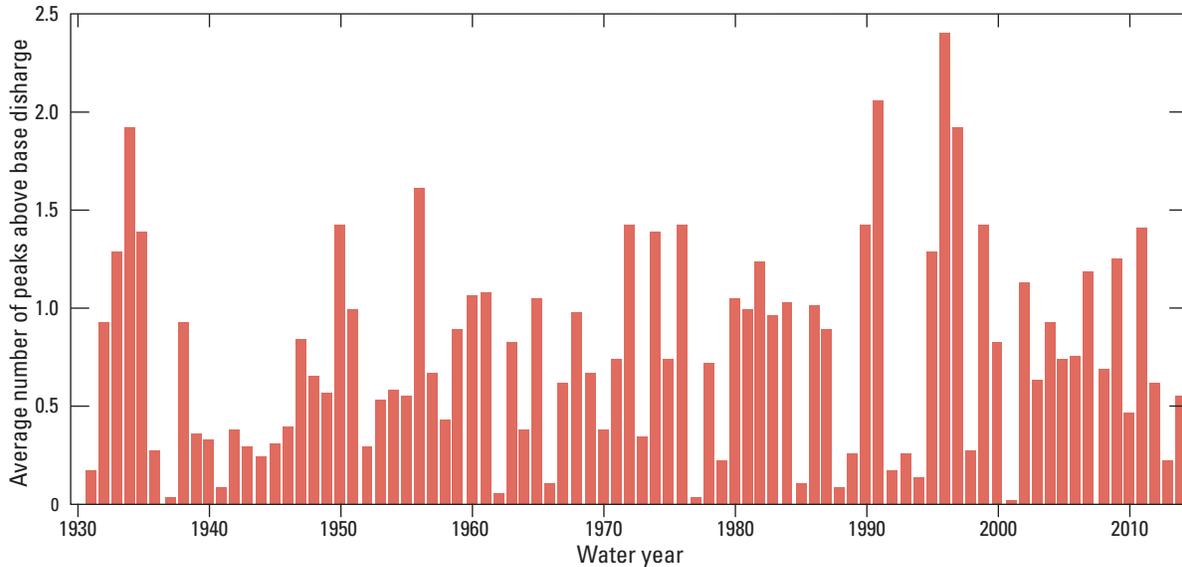
Figure 5.   Average number of peaks above base discharge for 65 long-term, unregulated streamgages, Washington, water years 1930–2014.

A Kendall's rank correlation two-sided test with the data shown in figure 5 did show a slight increasing trend in the number of floods above the base discharge, but it was not significant at the 0.05 level (tau=0.125, p-value=0.0897). The Pacific Decadal Oscillation (PDO) and the Southern Oscillation Index (SOI) are two patterns of low frequency variability that have teleconnections (climate anomalies in separate locations sometimes thousands of miles apart that are related to each other) to the climate in the Pacific Northwest (Mantua and others, 1997; Ward and others, 2014). Monthly values of PDO (Mantua, 2015) and SOI (National Center for Atmospheric Research, 2015) were averaged for each water year from 1930 to 2014. Monthly SOI values were processed as described by Trenberth (1984). A multiple linear regression model was used to test if these PDO and SOI values can explain the regional pattern of peaks above base discharge. The adjusted R-square value was 0.03 and none of the explanatory variables were significant at the 0.1 level. The spatial distribution of the trend values for peaks above base discharge for the individual sites (fig. 6) shows that most sites did not have a significant trend, 2 sites had a negative trend, and 10 sites, all on the western side of the State, had a positive trend.
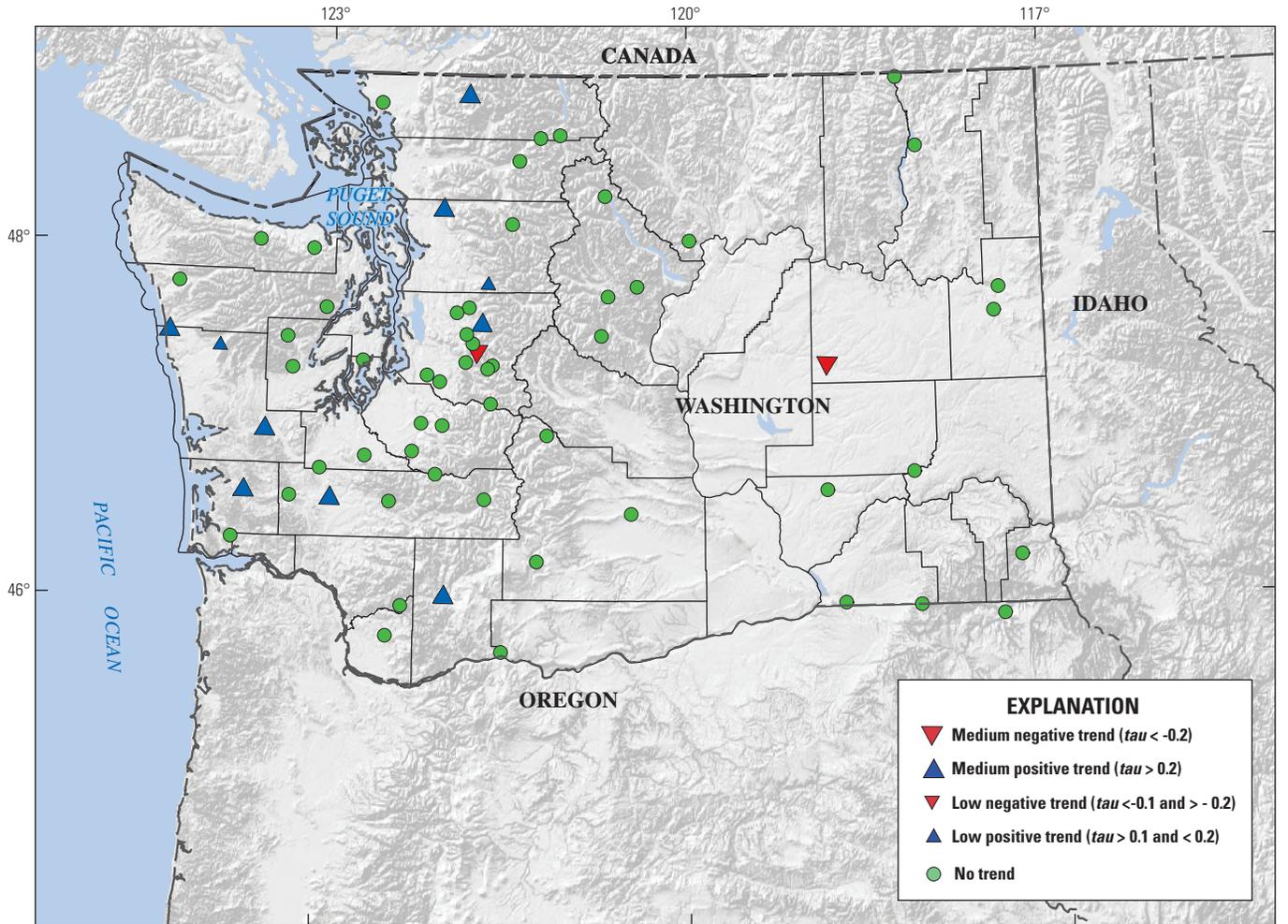
Figure 6.   Trends in the annual peaks above base discharge for streamgages with 50 or more years of daily discharge record through water year 2014, Washington. Sites defined as having no trend had less than 0.05 *p*-value in a Kendall's *tau* analysis.

Decadal Trends in Annual Peak Streamflow

Trends in peak flows were examined at 616 unregulated, non-urbanized sites (defined as sites with drainage basins with less than 5 percent impervious cover) in Washington, Idaho, and Oregon that had at least 10 recorded annual peaks. Mann-Kendall rank correlation coefficients (τ) were calculated between year and the magnitude of peak streamflow for every combination of starting and ending year spanning at least a 10-year period. Results are reported for trends that were significant at $p<0.05$.

Regional trends in annual peak flows can be expected as a response to climatic shifts between dry and wet conditions. In the Pacific Northwest region, decadal shifts in precipitation are linked to atmospheric circulation and sea surface temperatures (Cayan and others, 1998). As result, decadal trends in annual peak flows are evident at many sites (fig. 7). Decadal trends generally represent a period that started or ended with years having extreme values (low or high) for peak flows. Decades with increasing trends generally spanned from dry to wet years. Decades with decreasing trends spanned from wet to dry years. Increasing peak-flow magnitudes were common during the decades centered around 1945, 1956, 1970, 1992, and 2004 and decreasing peak flows during the decades centered around 1936, 1990, 2000, and 2010. Despite these dominant patterns, both upward and downward trends occurred in the State during many 10-year periods. Typically, 10-year trends at a site are cyclic—a period of increasing annual peak flows is followed by a period of decreasing annual peak flows.
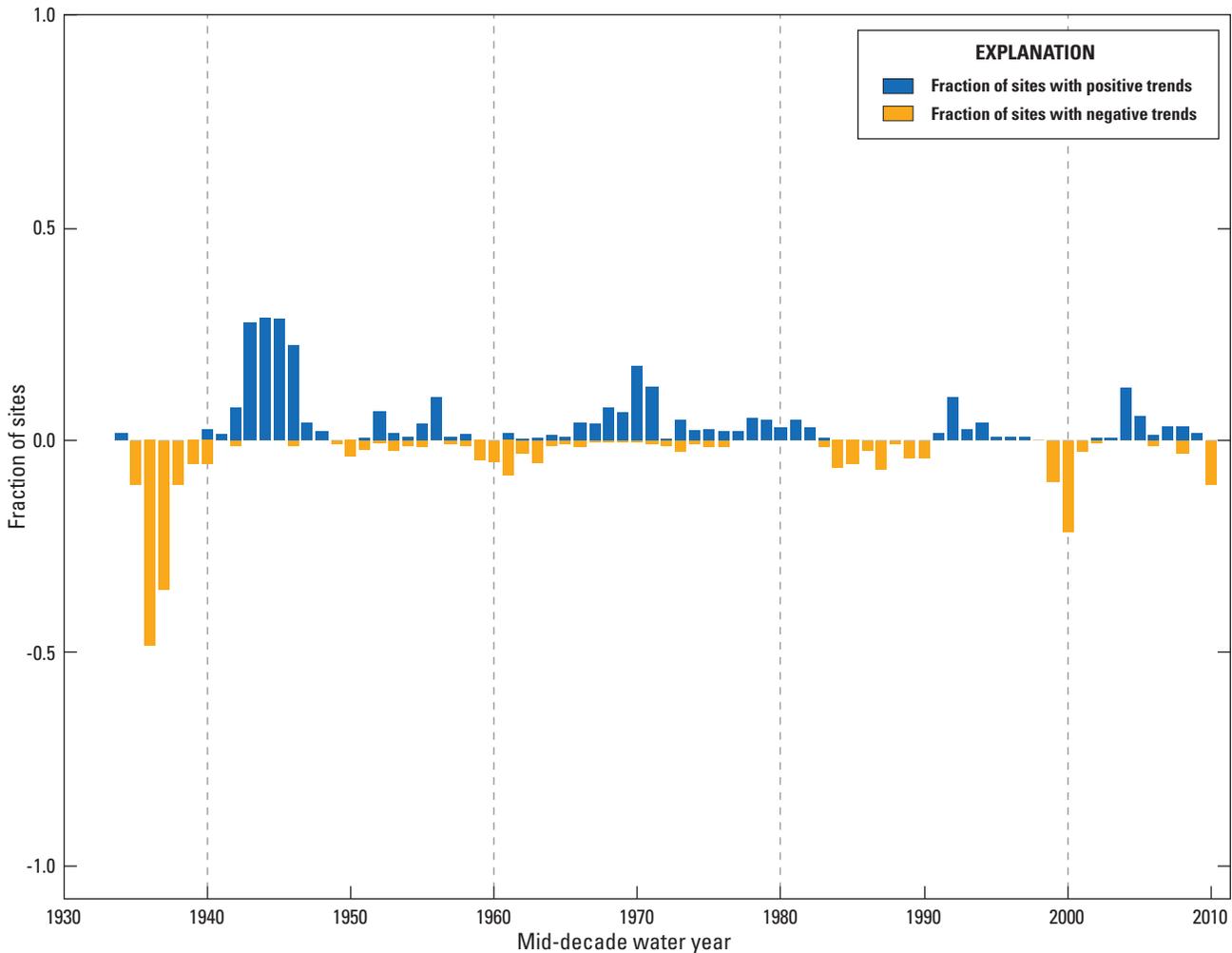
Figure 7.   Fraction of sites with significant 10-year trends in annual peak flows plotted at the water year in middle of trend period (for example, the trend from water years 1996 through 2005 is plotted at water year 2000), Washington.

A decadal trend that was not reversed or was not repeated in subsequent decades represents a persistent step trend. A step trend occurs when annual peak flows shift up or down over a relatively brief period of time. Flood series have different frequency distributions before and after significant steps. Step trends can be identified from trend tests when there is a significant trend for any period starting before the step and ending after the step, but no trends for periods that both start and end before the step or after the step. In contrast to step trends, continuous trends represent a gradual change in flood frequency that can be identified by significant trends with the same direction starting in each of a series of years.

Active Trends in Annual Peak Streamflow

Streamgages with a significant trend for at least 10 years including water year 2014 were classified as having active trends. In Washington, 38 sites have positive active trends in peak streamflow while 21 sites had negative active trends (fig. 8). Active trends were analyzed to distinguish sites with persistent and on-going changes in peak flows from sites where recent changes in peak flows may only be a short-term response to climatic variability. The earliest year when an active trend appeared at a site was identified and designated as the beginning of the active trend. One-half of the positive active trends began after 1965 while one-half of the negative active trends began after 1972.

Active and step trends were analyzed in this investigation with a graphical tool call flag plots. Flag plots (McCabe and Wolock, 2002) provide a way to visualize peak-flow trends associated with starting and ending years and can be used to determine whether trends are cyclic or persistent and whether they are step or continuous. A flag plot for a site is constructed by plotting the values of a trend test (Kendall *tau*) for all possible pairs of starting and ending years.

Periods when trends in peak flow were common are shown in a flag plot of the fraction of sites (fig. 9). For Washington, the large fraction of sites with positive trends starting about 1940 and continuing to the present and the lack of positive trends starting between 1945 and 1965 indicate the likelihood of a positive-step trend between about 1940 and 1950 at many sites (dark blue band on upper, left quadrant of fig. 9*B*). This step persists through 2014 at most of the 50 sites with peak-flow records that extend back to 1940. Likewise, another positive step likely occurred in the late 1960s (dark blue band on upper, center of fig. 9*B*) although the strength of this trend appears to have decreased since 2000. Strong negative trends started from about 1910 to 1925 and continued through the 1940s. Another period of negative trends started around 1945 and continued through to 2014.

Flood frequencies computed before and after a step trend will have different magnitudes, so the periods of records of sites were assessed in relation to the positive step trends around 1940 and late 1960s to identify potential bias. Only 50 sites have flood records prior to 1940 and those sites generally contain records of many more post-1940 floods, so these sites are unlikely to have much of a downward bias in their computed flood frequencies. Sites with records beginning well before the step trends in the late 1960s may have lower computed flood frequencies than would be computed if records were censored to exclude pre-1970 floods. Although these step trends were not considered significant enough across Washington to censor peak-flow records in general, recent flood magnitudes should be examined when using flood estimates at sites with long periods of record.

Examples of Strong Active Trends in Annual Peak Streamflow

USGS streamgage 12091200, Leach Creek near Fircrest, Washington, had a steep increase in peak flows around 1980–90 that persists as the strongest ($\tau = 0.43$) active trend among all sites with continuous records of annual peaks since the beginning of the trend (fig. 10). This trend is likely a result of urban development (Konrad and Booth, 2002). There was a brief period in the mid-1980s without any large peaks that started around 1986 that is consistent with a statewide downward trend (fig. 7), but the positive trend continued after 1995.

The magnitude of peak flows in the North Fork of the Stillaguamish River (USGS streamgage 12167000) has increased continuously since 1936 with only two periods of short-term negative trends. This is one of the longest and strongest ($\tau = 0.37$) positive active trends in Washington. Regional climate may be a factor given that many others sites in western Washington have increasing trends beginning around 1940 and continuing through 2014 and also beginning in the late 1960s (fig. 11). Hall and others (2014) suggest that a shift in precipitation from snow to rain may also contribute to the increasing trend.

Crab Creek near Moses Lake, Washington (USGS streamgage 12467000), has the strongest ($\tau = -0.43$) active negative trend of any streamgage with continuous record of peaks during the trend period (fig. 12). The trend started in 1969. There are numerous diversions upstream of the streamgage as well as declining regional groundwater that may be associated with this trend. The Nisqually River at La Grande, Washington (USGS streamgage 12086500) has the longest record with a negative active trend ($\tau = -0.21$) (fig. 13). Two reservoirs began operating in water year 1945, but trends spanning the pre- and post-reservoir periods only appeared around 2001.
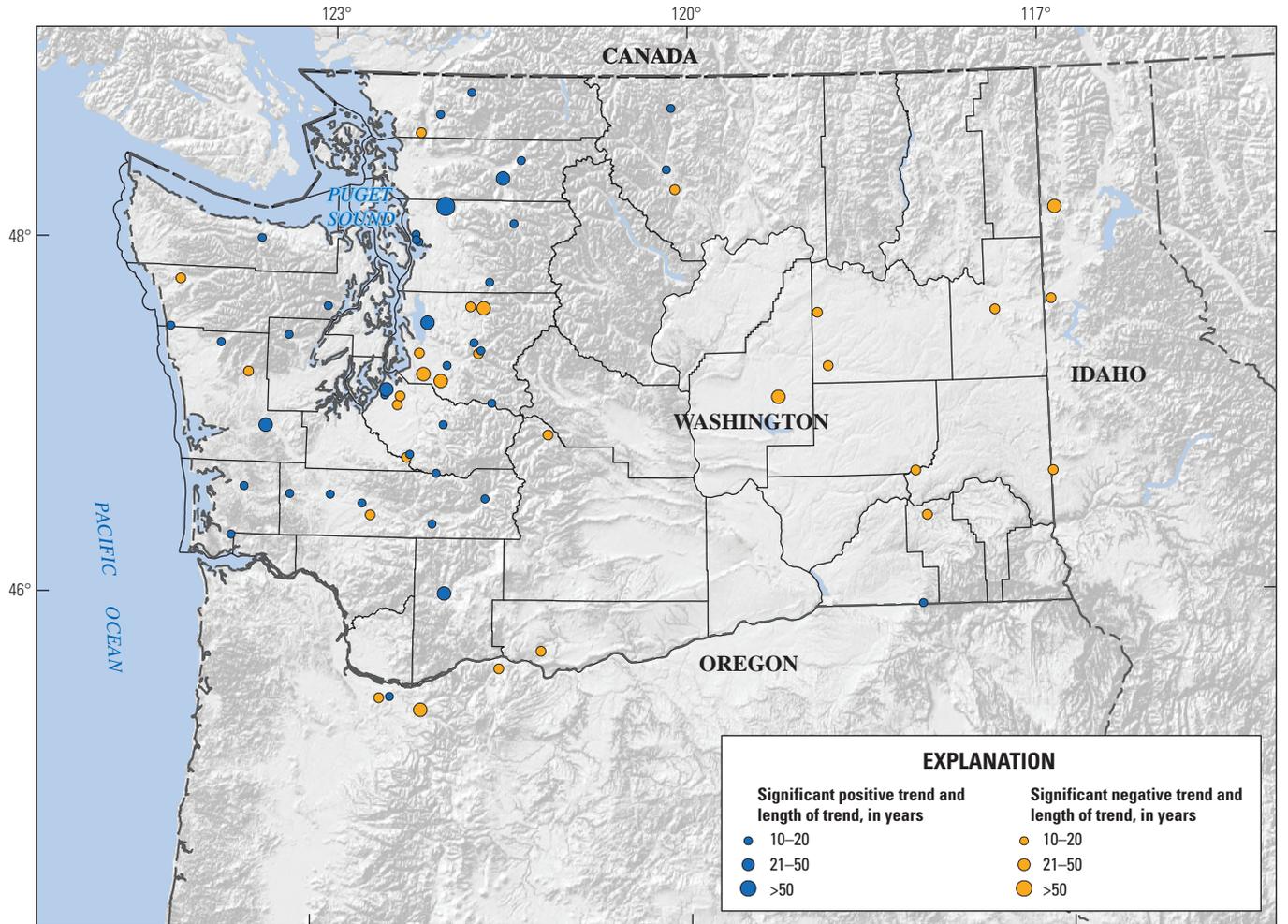
Figure 8.   Sites with active trends through water year 2014, Washington. These trends have persisted continuously for the length of time indicated by the size of the circle.
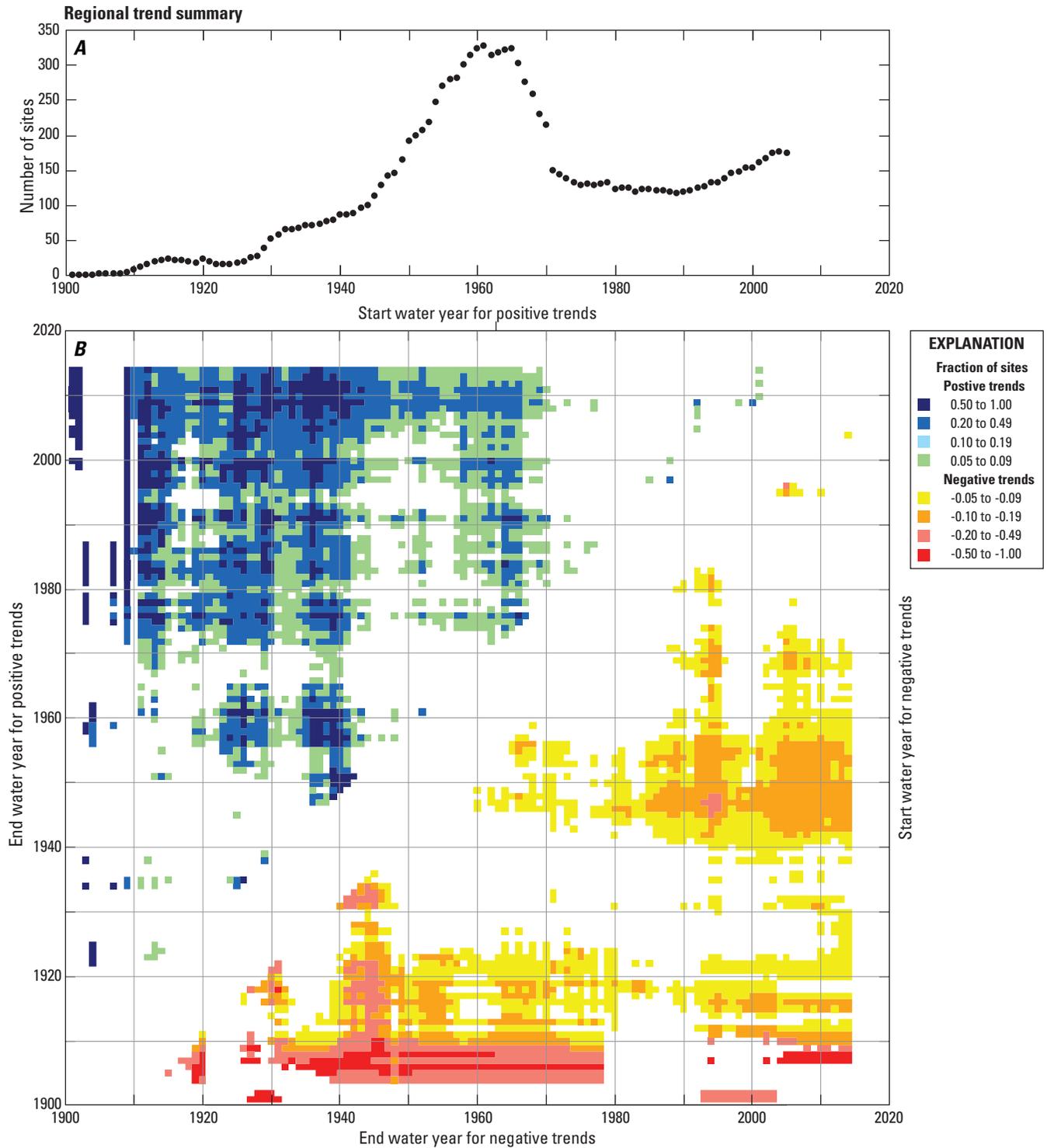
Figure 9.   Regional trend summary of (*A*) number of sites and (*B*) flag plot showing fraction of sites with positive and negative trends in annual peak streamflow, Washington.
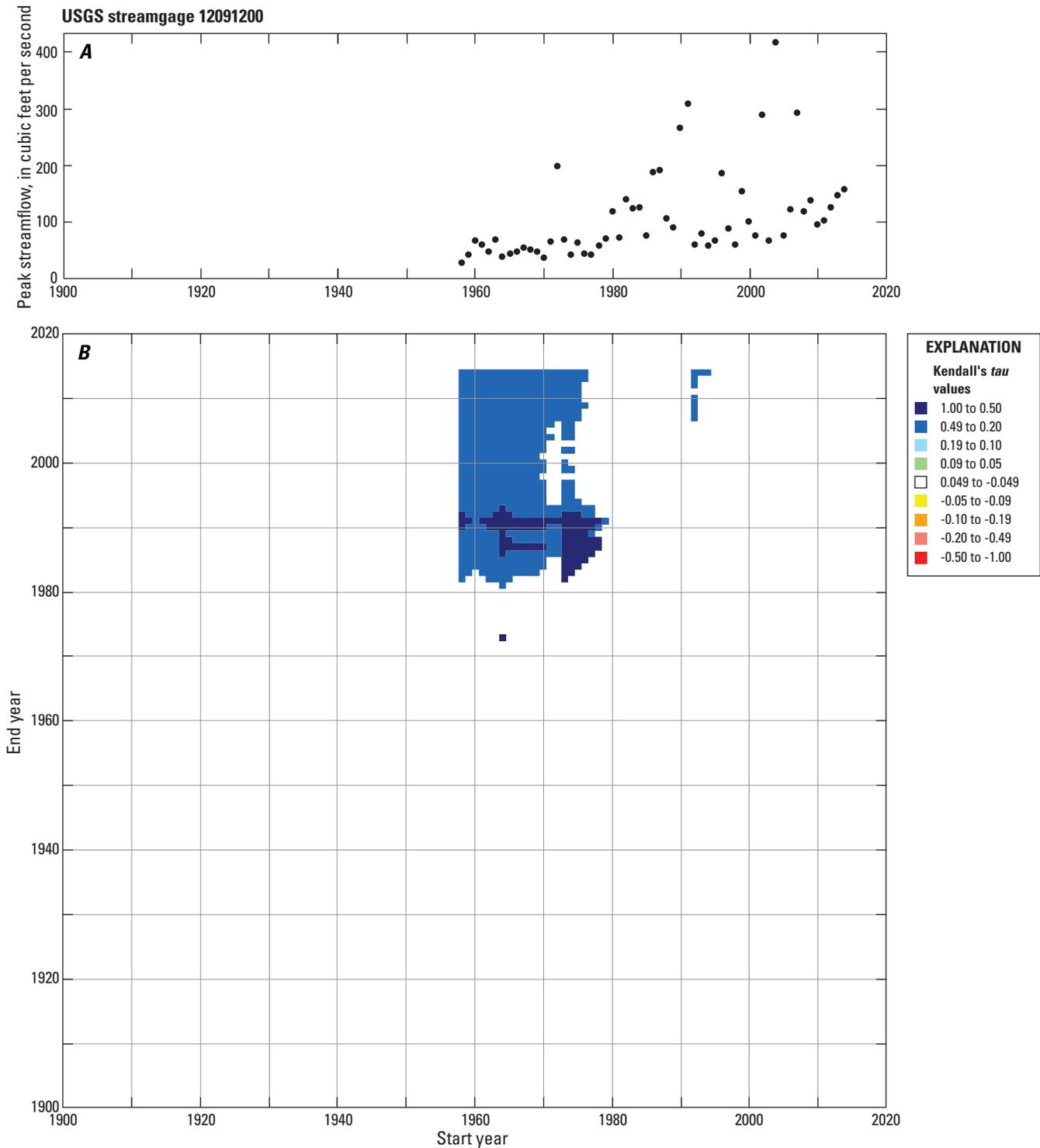
Figure 10.   Annual series of (*A*) peak streamflow and (*B*) flag plot of trends for Leach Creek near Fircrest, Washington (U.S. Geological Survey streamgage 12091200), 1958–2014.
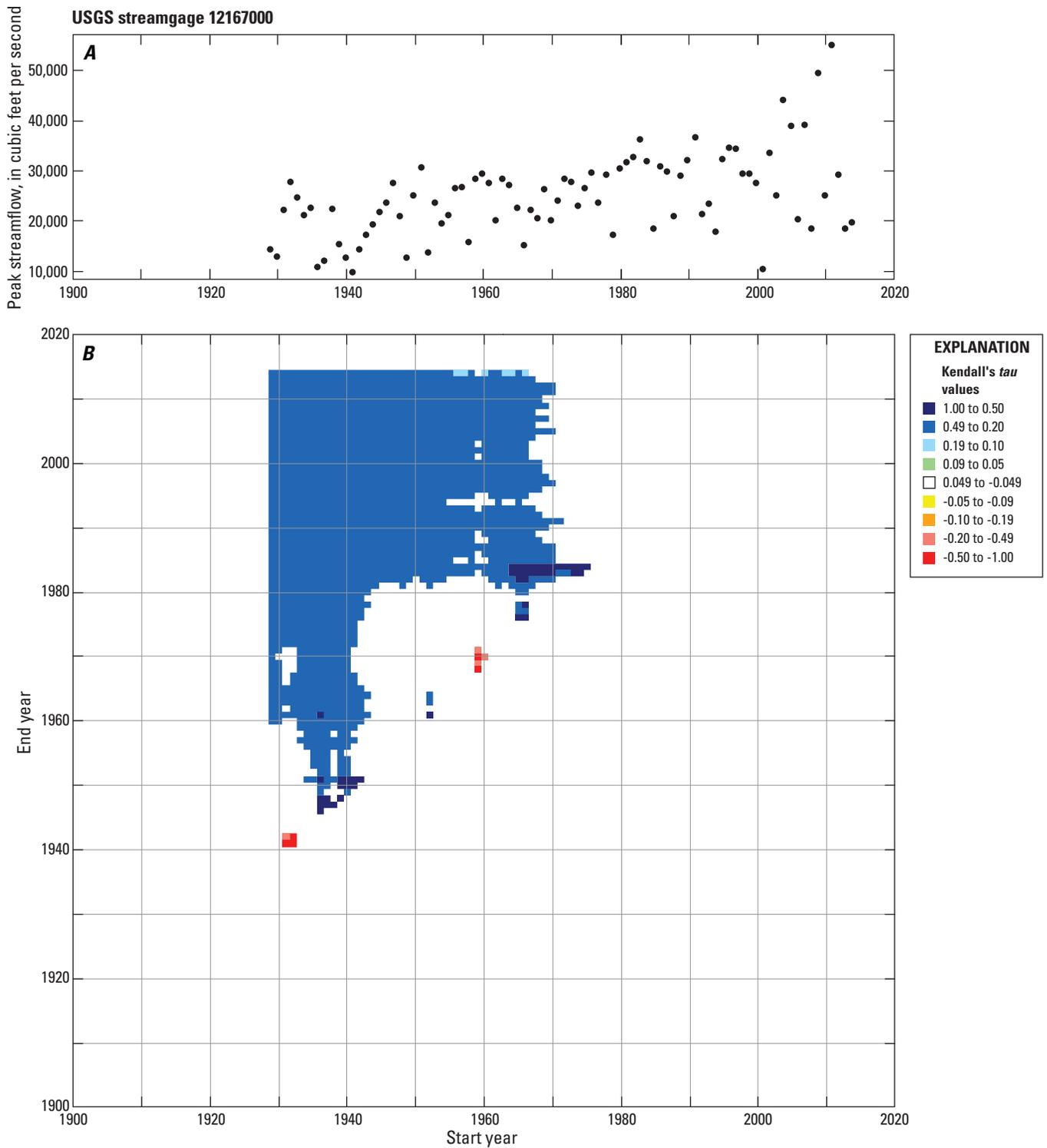
Figure 11.   Annual series of (*A*) peak streamflow and (*B*) flag plot of trends for North Fork Stillaguamish River near Arlington, Washington (U.S. Geological Survey streamgage 12167000), 1929–2014.
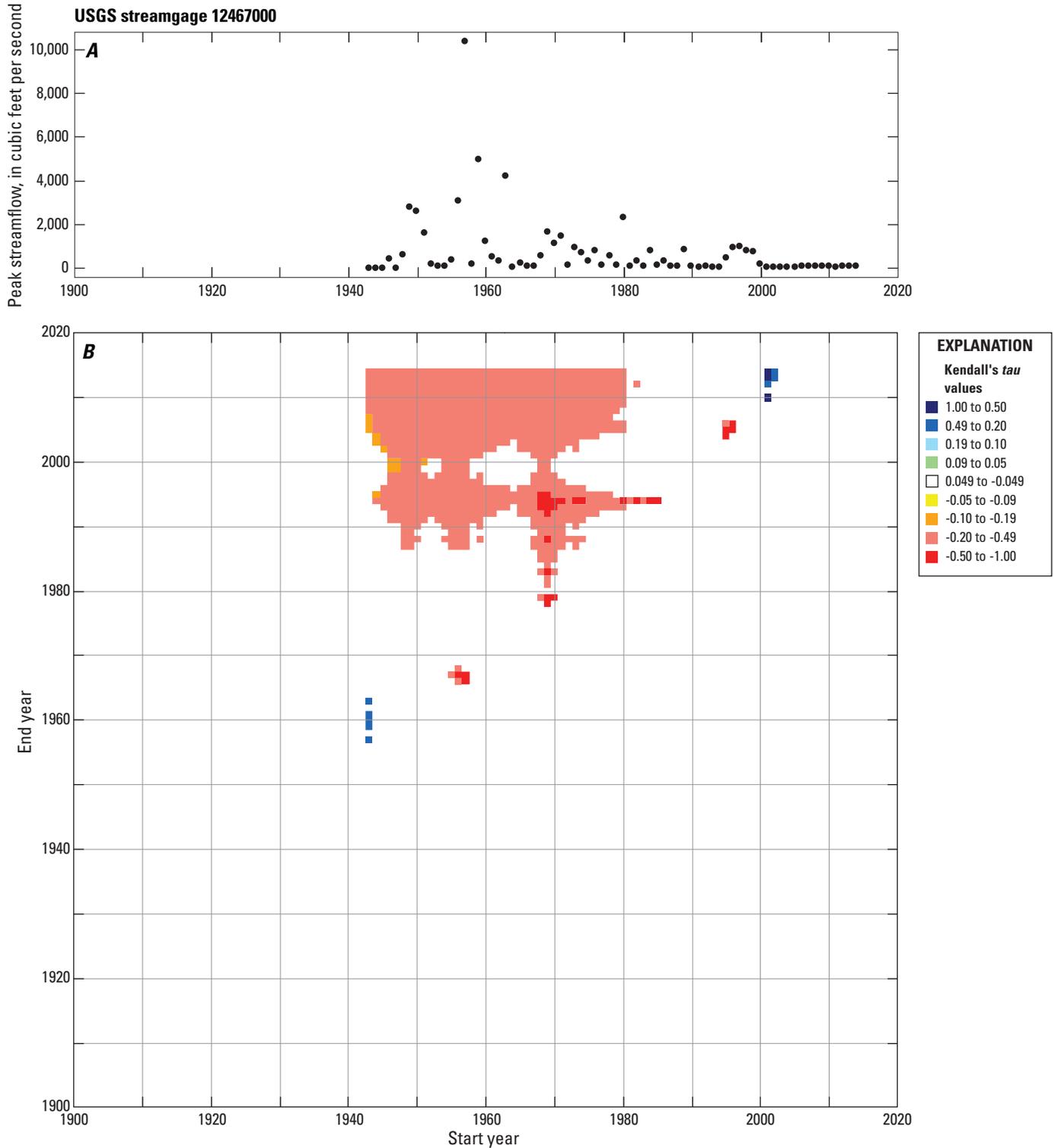
Figure 12.   Annual series of (*A*) peak streamflow and (*B*) flag plot of trends for Crab Creek near Moses Lake, Washington (U.S. Geological Survey streamgage 12467000), 1943–2014.
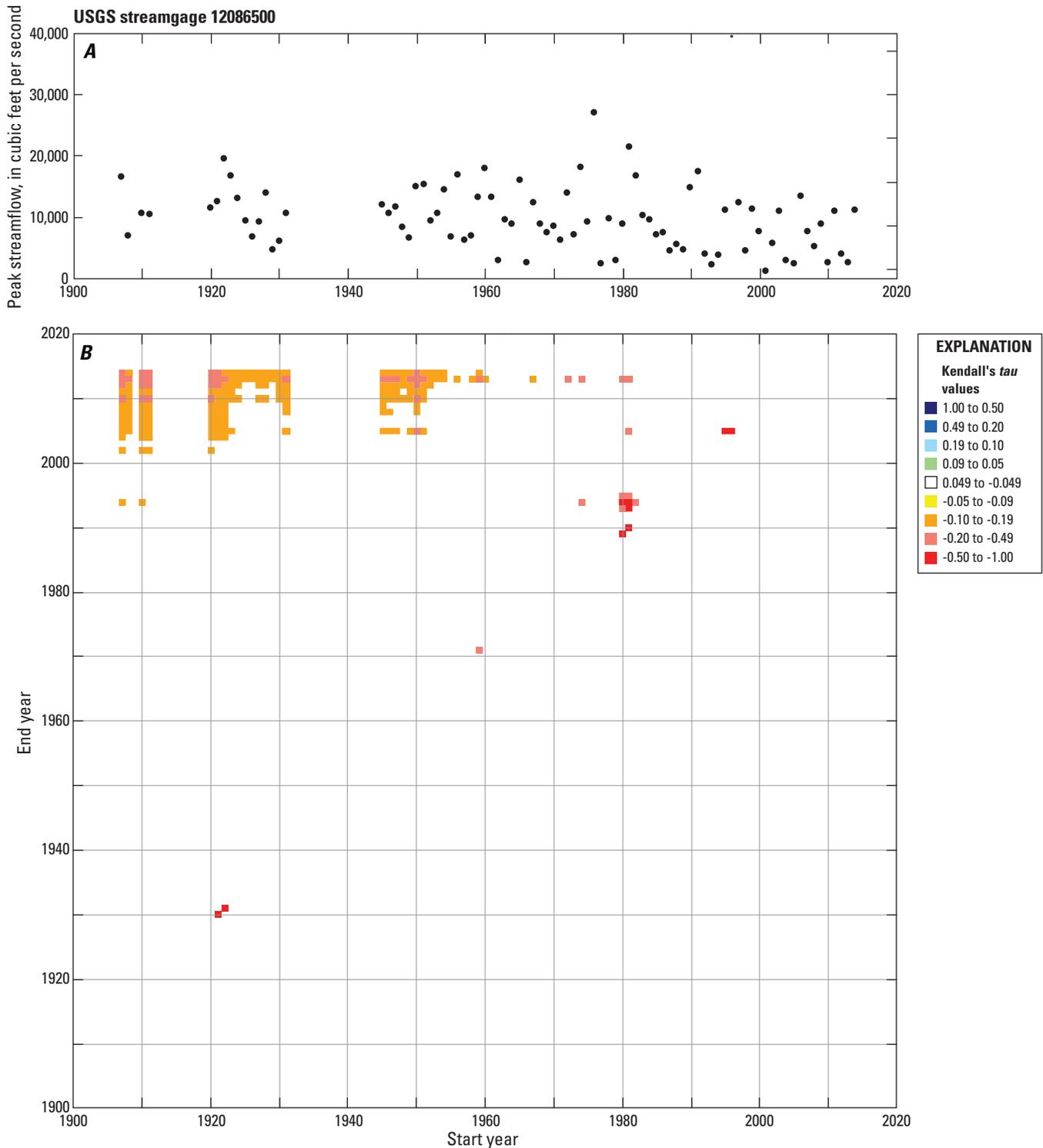
Figure 13.   Annual series of (*A*) peak streamflow and (*B*) flag plot of trends for Nisqually River at La Grande, Washington (U.S. Geological Survey streamgage 12086500), 1907–2014.

Estimating Magnitude and Frequency of Floods

Four methods of determining the magnitude and frequency of floods in Washington are described here because the best estimate depends on the available data at the site of interest. These methods of estimating flood magnitudes for various frequencies (flood quantiles) are discussed in this section and include:

1. If a long-term streamgage exists at the site of interest, the best method is to use the station data available to develop the statistical distribution that defines the flood frequency at a site. Once this is done, the flood quantiles can be easily determined.

2. If the site of interest is ungaged, regional regression equations based on basin characteristics may be used to estimate the flood quantiles.

3. If a streamgage exists at or near a site of interest but does not have a long record, the statistical distribution is determined from the streamgage data, but is improved by weighting it with an independent estimate such as those determined from the regional regression equations.

4. If a site of interest is ungaged, but on a gaged stream, the flood quantiles may be estimated using a drainage area ratio method to weight the results from the statistical distribution obtained for the gaged site or sites.

After publication of this report, the flood quantiles at the long-term, unregulated streamgages are expected to be made available in StreamStats (http://water.usgs.gov/osw/streamstats/index.html), the interactive USGS Web-based program. The regional regression equations will also be available in StreamStats program for Washington with the tools and GIS datasets of basin characteristics that will allow users to compute the flood quantiles for any location on the river network and lies within the range of basin characteristics used to develop the regional regression equations. StreamStats is networked to the National Flood Frequency (NFF) (Knowles and Sumioka, 2001) program which is a database for regional regression equations. The regional regression equations will be included in the NFF database.

The following sections describe the four methods of determining flood quantiles and provide the data for the unregulated streamgage sites in Washington with at least 10 years of non-zero annual peak information. Examples are given to show how to apply the methods.

Magnitude and Frequency of Floods at Gaged Sites

Flood-frequency analyses are performed using peak flows at streamgages and provide an estimate of the magnitude of flows that are expected to be equaled or exceeded for a specified probability for any one year (AEP or recurrence interval). Flood-frequency analyses in this report uses the annual instantaneous peak discharges for each water year and was performed for 649 streamgage sites. All the streamgages are or were operated by the USGS and have at least 10 years of non-zero, unregulated peak-flow records. The peak-flow records are maintained in a database and are publicly available online on the USGS National Water Information System (NWIS, access http://waterdata.usgs.gov/wa/nwis/sw for surface-water data for Washington). Although, this report is focused on Washington, 36 sites in Oregon and 24 sites in Idaho, near the State borders with Washington, also were included in this study. Determinations of whether a river (where a streamgage is located) is regulated and whether it would be included in the flood-frequency analyses were based on the streamgage descriptions and the peak-flow codes in the peak-flow database in NWIS. Unless the station description or knowledge of the streamgage by the authors could identify the streamgage with a reasonable amount of certainty about the nature of the regulation in the basin, the decision to classify peak flows as regulated or unregulated was made using the discharge codes found in the tables of annual peak flows in the NWIS database. Streamgages with code 5 peak flows "discharge affected to unknown degree by regulation or diversion," were classified as unregulated, and they were included in the flood-frequency analyses with the assumption that the regulation for these streamgages likely occurs during low-flow season for irrigation and does not influence peak flows. Streamgages with code 6 peak flows, "discharge affected by regulation or diversion," were considered regulated and not used in the flood-frequency analyses. Many streamgages had peak-flow records for water years prior to the establishment of a dam. These streamgages were included in the flood-frequency analyses, but only the unregulated record of peak flows prior to the construction of the dam were used. Some peak flows were coded as an estimate (code 2) or maximum daily average (code 1). Of the 13,996 recorded peaks for the Washington sites, 63 peaks (0.45 percent) were code 1 and 174 peaks were code 2 (1.2 percent). These peak flows were included in the flood-frequency analyses without any adjustment to the peak flow. Peak flows affected by dam failure (code 3) were not used in the flood-frequency

analyses. The time series of peak flows were combined for two pairs of streamgages that were located so close to one another on the same stream that they could be considered the same site. The time series of peak flows for Skookum Creek near Wickersham, Washington (USGS streamgage 12209500, drainage area = 22.46 mi²), were combined with the peak flows for Skookum Creek above Diversion near Wickersham, Washington (USGS streamgage 12209490, drainage area = 22.48 mi²) and the flood quantiles are listed for USGS streamgage 12209490. Additionally, the peak flows for the Humptulips River near Humptulips, Washington (USGS streamgage 12039000, drainage area = 130.2 mi²), were combined with the peak flows for Humptulips River below Highway 101 near Humptulips, Washington (USGS streamgage 12039005, drainage area = 132.8 mi²), and the flood quantiles are listed for USGS streamgage 12039005.

The flood-frequency analyses in this report follows the methodology set by Bulletin 17B (International Advisory Committee on Water Data, 1982) except for the use of the Expected Moments Algorithm (EMA) and the Multiple Grubbs-Beck (MGB) low-outlier test discussed later in this section. Bulletin 17B was written by the Interagency Advisory Committee on Water Data and is a standard methodology for conducting flood-frequency analyses in the United States. The Bulletin 17B methodology prescribes the use of a log-Pearson Type III distribution to fit the logarithms of annual peak-flow data using the mean, standard deviation, and skew coefficient of the peak-flow data to compute the flood discharge for a particular AEP with the general equation:

$$\log Q_P = \bar{X} + K_P S \qquad (1)$$

where

$\quad Q_P \quad$ is the flood discharge for AEP, P, in ft³/s;

$\quad \bar{X} \quad$ is the mean of the logarithms of the annual peak flows, in ft³/s; and

$\quad K_P \quad$ is value obtained from appendix 3 in Bulletin 17B given the AEP and the skew coefficient, G:

$$G = \frac{N \sum (X - \bar{X})^3}{(N-1)(N-2) S^3} \qquad (2)$$

where

$\quad N \quad$ is the number of annual peak flows; and

$\quad S \quad$ is the standard deviation of the logarithms of the annual peak flows, in ft³/s.

Under recommendation of Bulletin 17B, some changes have been proposed by the Subcommittee on Hydrology, Hydrologic Frequency Analysis Work Group that were used in these analyses. These proposed changes include the use of the Expected Moments Algorithm (Cohn and others, 1997) and the MGB test for low outliers (Cohn and others, 2013). All flood-frequency analyses were performed using the USGS software program, Peak flow FreQuency (PeakFQ, ver. 7.1; Veilleux and others, 2014).

Annual peak flows collected as part of the systematic operation of the streamgage are called systematic data. Historical peak flows are peaks recorded outside the systematic period and are known to exceed some threshold discharge that is greater than any peak flows that occurred during the period between the historical peak flow or peak flows and the start of the systematic record. The period of systematic data together with the intervening years between the systematic and historical peak flows and the years of historical peak flows defines the historical period of the streamgage. The EMA accommodates flow-interval data for censored peaks where the peak is known to only be above or below some value, but the exact value is unknown. For example, a crest-stage gage (CSG) is a simple vertical pipe with a stick and cork inside which is used to record the peak stage of a flood. If there is no mark on the CSG for a given water year, then it is known any peak that may have occurred during the water year was below the elevation of the bolt used to support the stick—this would be a censored peak. EMA also accommodates historical data and uncertainty in data points. For every year in the historical period, including gaps in the systematic record, a flow interval must be provided. For most peak flows during the systematic period, the upper and lower intervals equal the peak flows, and for most water years when no information has been recorded the intervals are 0 and infinity. If there is uncertainty in a peak flow, the range of plausible flows for that peak can be entered as the lower and upper flow intervals. Along with flow interval, EMA also requires a perception threshold for every year. Perception thresholds are the range of flows that could have been measured if they occurred. Generally, for the systematic record, the range of perception thresholds is from 0 to infinity. At a CSG where flows below the base bolt or bottom of the gage and above the gage are unknown, the lower perception threshold would be the flow when the water surface equals the base bolt elevation and the upper threshold would be the flow when the water surface is above the gage. Perception thresholds also are set for the ungaged period within the historical period and generally the lower threshold is equal to the peak flow for the historical peak, and the upper threshold

is infinity. An opportunistic peak flow is one that was recorded outside of the systematic record and collected based on factors other than the exceedance of a perception threshold. These peak flows are not random and thus the sampling properties are unknown. Opportunistic peaks were not used in the flood-frequency analyses. The flow interval and perception thresholds that are different than the default setting are shown for each streamgage in the flood frequency analyses in table 2.

In flood frequency analyses, most of the interest is in AEPs with low probabilities (large floods) which are near the upper end of the peak-flow distribution. Sometimes peak flows that are smaller than any of the flood quantiles of interest may have a large influence on the upper end of the peak-flow

distribution. These potentially influential low flows (PILFs) or low outliers that have a large influence on the upper end of the peak-flow distribution may reflect different physical processes than those processes that generate the large floods. The MGB low-outlier test censors the PILFs from the frequency distribution and typically results in improved estimation of the flood quantiles. The frequency curve for the streamgage on the Dungeness River near Sequim, Washington (USGS streamgage 12048000), shows the results after the MGB test identified 21 PILFs below a threshold value of 2,170 ft^3/s that were censored (removed) from the computations of the mean, standard deviation, and skew coefficient which determine the shape of the frequency curve (fig. 14).
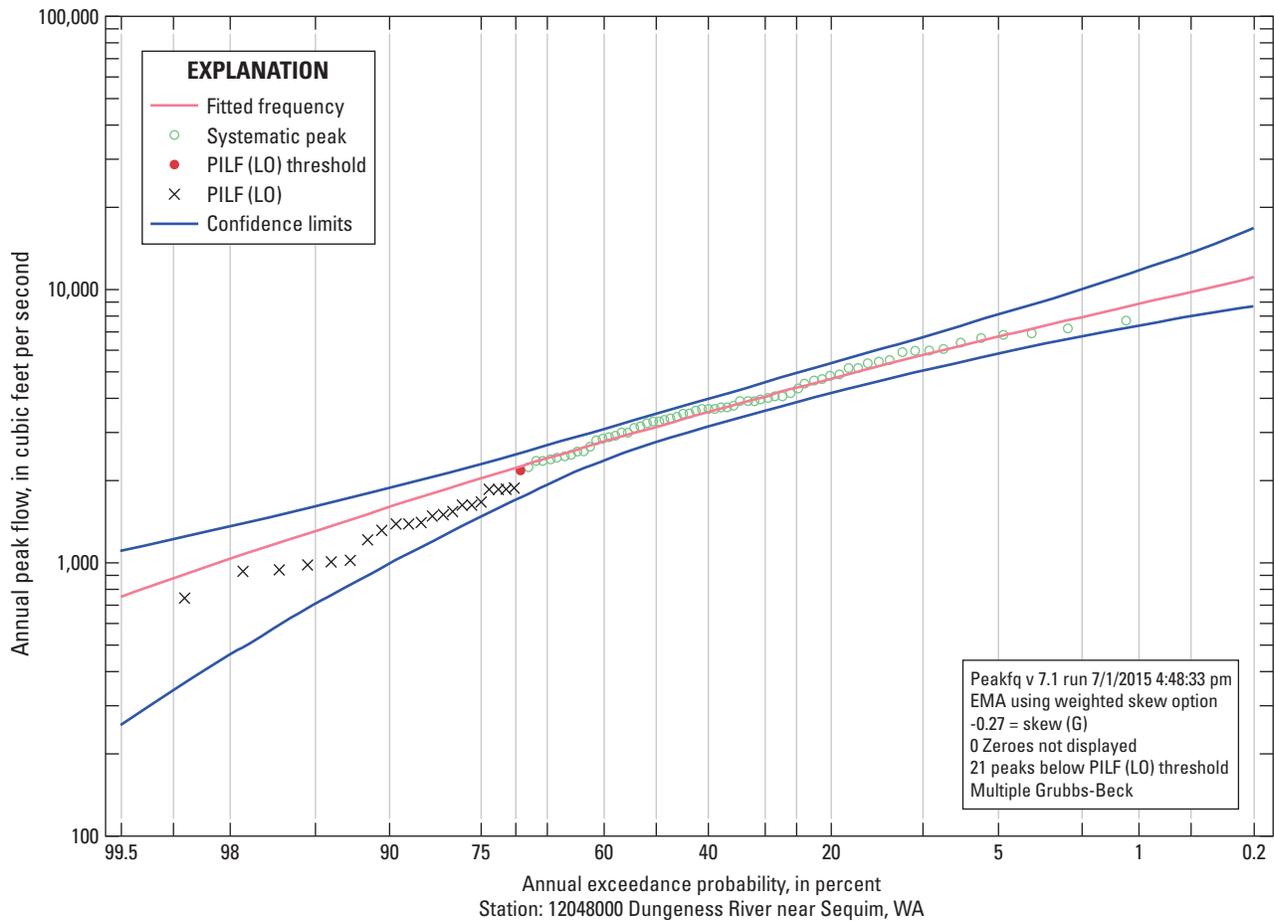
Figure 14.   Flood frequency curve and the 95-percent confidence limits for the annual peak flows recorded on the Dungeness River near Sequim, Washington (U.S. Geological Survey streamgage 12048000). The multiple Grubbs-Beck low-outlier test identified 21 potentially influential low flows (PILFs) below a threshold of 2,170 cubic feet per second that were removed for the computation of the parameters of the fitted frequency curve.

Regional Skew

The International Advisory Committee on Water Data (1982) states that a better estimate of the analysis skew coefficient can be obtained by combining the regional (or "generalized") skew coefficient with the station skew coefficient by weighting the skew coefficients inversely proportional to their individual mean-square errors (MSE) by the following equation:

$$G_W = \frac{MSE_{\bar{G}}(G) + MSE_G(\bar{G})}{MSE_{\bar{G}} + MSE_G} \qquad (3)$$

where

G_W	is the weighted skew coefficient,
G	is the station skew coefficient,
$\bar{G}$	is the generalized skew coefficient,
$MSE_{\bar{G}}$	is the mean-square error of generalized skew, and
MSE_G	is the mean-square error of the station skew coefficient.

In the past, the generalized skew coefficient could be taken from a national map of skew coefficients (International Advisory Committee on Water Data, 1982) that had a MSE of 0.302. For this investigation, a regional skew coefficient was developed for the Pacific Northwest region which was defined as the Columbia River Basin south of the U.S.-Canada border, the Oregon and Washington coastal areas, and the remainder of Washington. This region included most of Idaho and Oregon, the entire state of Washington and western Montana. The development of the Pacific Northwest regional skew coefficient was based on station skew coefficients and basin characteristics for 290 long-term gages, and it is described in detail in the appendix A. A constant statistical model provided the best model for the region resulting in a single regional skew coefficient of -0.07 that is applicable everywhere in the Pacific Northwest region. The MSE was reduced significantly from the national model (0.302) to 0.18, and is equal to a generalized standard error of 0.4243.

The flood frequency analyses for this study used a regional skew coefficient of -0.07 and station skew coefficients from the streamgage records to compute a weighted skew coefficient. The weighted skew coefficients were used to compute the final flood-quantile values except for the streamgages draining urbanized basins (defined as basins with more than 5 percent land cover classified as impervious). Increasing percentages of impervious land cover in a basin, which is commonly the result of urbanization, can increase the magnitude of floods when compared to the same basin prior to urbanization. Streamgages draining urbanized basins were not used in the development of the regional skew coefficient so it would be inappropriate to weight the skew coefficient for these basins. The threshold value of 5-percent impervious land cover is a conservative number in the sense that it would screen out all basins from the regression analysis that may be have an increasing trend in the magnitude and frequency of flood due to urbanization. Konrad and Booth (2002) demonstrated that sites in the Puget Sound Lowlands with more than 5 percent impervious land cover showed no statistical significant trend in the mean annual flood. If the weighted skew coefficient was not used, the station skew coefficient based solely on the station data was used.

Consideration of Mixed Population Analysis

Many of the streamgages with high-elevation basins have peak flows that occur in spring or summer due to snowmelt, as well as peak flows that result from winter rain events. Streamgages with two or more different peak-generation processes are classified as having mixed population of peak flows. At streamgages with mixed populations of peak flows, it is possible to have separate distributions of the peak flows that do not fit well to one log-Pearson Type III distribution and may require separate frequency analysis for each population to improve the estimates of the flood quantiles. Although many of the streamgages in the flood frequency analyses do have a mixed population of peak flows, the single log-Pearson Type III distribution seemed to capture both populations reasonably well, such that separate distributions were not necessary. The peak-flow record for the Stehekin River at Stehekin, Washington (USGS streamgage 12451000), characterizes this point. Out of the 93 annual peaks recorded at this streamgage, 17 peaks occurred in winter (October–February), 1 peak in early spring (March–April), and 75 peaks in summer (May–August). Although the summer peaks far outnumber the winter peaks, 4 of the top 5 in magnitude and 6 of the top 10 peaks occurred in winter. Figure 15 shows the one log-Pearson Type III distribution used to fit all of the peaks at USGS streamgage 12451000. Even though there is some divergence of the top seven peaks from the fitted distribution, the divergent peaks include both winter and summer peaks and the divergence is minimal. Comparing the fitted distribution with the plotting position of the peaks at this site shows a reasonable estimation of the distribution and no reason to initiate a mixed-population analysis. Frequency plots for all of the streamgages in this study were inspected for divergence of the fitted distribution that might indicate the presence of mixed populations which might require a mixed population analysis; and, no streamgages had substantially diverging distributions that required a mixed-population analysis.
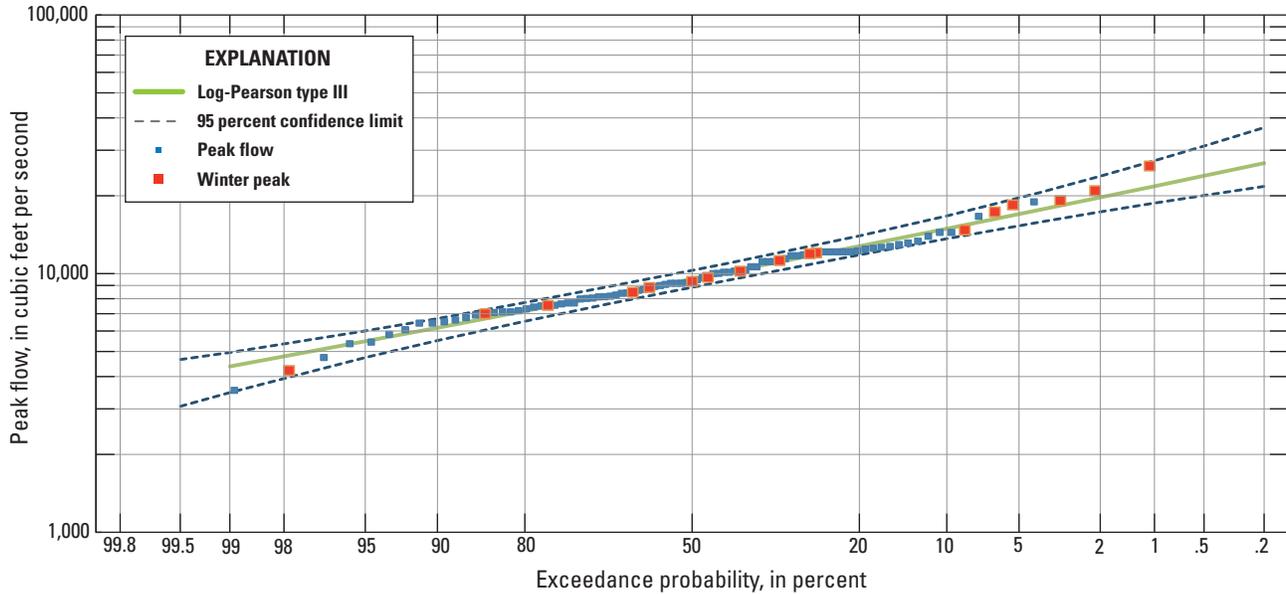
Figure 15.   Fitted frequency curve for the annual peaks on the Stehekin River at Stehekin, Washington (U.S. Geological Survey streamgage 12451000) showing the 95-percent confidence limits and winter peaks (October through February).

Regression Equations for Estimating Flood Magnitudes in Ungaged Basins

Hydrologic analysis involving flood magnitudes are often needed at stream reaches with no nearby long-term streamgages, and therefore, a hydrologist must make an estimate of the flood magnitude for a particular AEP without a record of peak flows. Regression equations, which can aid in the analysis at ungaged location, were developed to estimate specific flood frequencies based on basin characteristics. Basin characteristics are observable measures of the drainage basin upstream from the streamgage and can include both climatic and physical characteristics, such as mean annual precipitation and drainage area.

The regional regression equations use a multiple-linear regression that relates basin characteristics to a particular flood quantile. The general equation can be given by:

$$Q = aA^b B^c ... N^n \qquad (4)$$

where
Q	is the flood quantile, in ft^3/s,
$A, B, ..., N$	are the basin characteristics,
a	is the regression constant, and
$b, c, ..., n$	are the regression coefficients.

Base 10 logarithms are used to convert equation 2 to a linear form by transforming the variables to:

$$\log Q = \log a + b(\log A) + c(\log B) + ... + n(\log N) \qquad (5)$$

Exploratory Regression Analysis

Ordinary Least Squares Analysis, a form of multiple linear regression (Helsel and Hirsh, 1992), was used initially to investigate which basin characteristics may be significant. It also was use to investigate how Washington could be divided into hydrologically similar regions that grouped stations with similar statistical response to the explanatory variables (basin characteristics) and improve the accuracy of the regression equations.

Twenty-four basin characteristics were generated from various sources (table 3) as potential candidates of explanatory variables in the region regression equations for 641 sites in Washington and in Oregon and Idaho near the Washington border. When possible, the Washington, Oregon, and Idaho StreamStats programs were used to generate the basin characteristics with the exception of annual precipitation which was compiled from a newer source. The drainage area as computed from StreamStats was used for this investigation rather than the published values for the streamgage which was often slightly different; the average absolute difference was 3.4 mi^2 or 3.0 percent. Data not available from the StreamStats program was acquired from various sources (table 3) and processed with the GIS program ArcMap® for each drainage basin associated with a streamgage analyzed in this investigation. A GIS shapefile for all the drainage basins was obtained from StreamStats; these shapefiles were converted to GIS raster representations and then used to compile basin characteristics from other sources.

Table 3.   Basin characteristics compiled and data sources for each drainage basin for selected streamgages.

[**StreamStats abbreviations:** From http://streamstatsags.cr.usgs.gov/ss_defs/basin_char_defs.aspx. StreamStats, a web-based geographic information systems (GIS) application for computing streamflow statistics, available at: http://water.usgs.gov/osw/streamstats/index.html. **Description:** NAD 83, North American Datum of 1983. **Source:** 30 m DEM, Digital Elevation Model at a 30 meter cell size; NHD, National Hydrologic Dataset]

Basin characteristic	StreamStats abbreviation	Description	Source
Latitude of the centroid	LAT_CENT	Latitude, in decimal degrees at NAD 83 datum	Computed from 30 m DEM NHDPlus gridded data (http://www.horizon-systems.com/NHDPlus) using ArcMap© GIS commands
Longitude of the centroid	LONG_CENT	Longitude, in decimal degrees at NAD 83 datum	Computed from 30 m DEM NHDPlus gridded data using ArcMap© GIS commands
Latitude of streamgage	LAT_GAGE	Latitude, in decimal degrees at NAD 83 datum	Computed from StreamStats using a 30 m DEM NHDPlus gridded data using ArcMap© GIS commands
Longitude of streamgage	LNG_GAGE	Longitude, in decimal degrees at NAD 83 datum	Computed from 30 m DEM NHDPlus gridded data using ArcMap© GIS commands
Drainage area	DRNAREA	Total upstream area of the streamgage, in square miles, that drains to a point on a stream	Computed from StreamStats using the watershed delineation tool
Basin elevation	ELEV	Area-weighted mean elevation of the drainage basin, in feet above sea level	Computed from StreamStats using a 30 m DEM NHDPlus gridded data
Minimum basin elevation	ELEVMIN	Minimum elevation of the drainage basin, in feet above sea level	Computed from StreamStats using a 30 m DEM NHDPlus gridded data
Maximum basin elevation	ELEVMAX	Maximum elevation of the drainage basin, in feet above sea level	Computed from StreamStats using a 30 m DEM NHDPlus gridded data
Relief	RELIEF	Maximum minus the minimum elevation of the drainage basin, in feet above sea level	Computed from StreamStats using a 30 m DEM NHDPlus gridded data
Main basin slope	SLPPCT	Area-weighted mean slope of the drainage basin, in percent	Computed from StreamStats using a 30 m DEM NHDPlus gridded data
Main basin slope > 30	SLPPCT30	Percentage of the drainage basin area where the slope is greater than 30 percent	Computed from StreamStats using the mean basin slope gridded data
Main basin slope > 30 North	SLPPCT30N	Percentage of the drainage basin area where the slope is greater than 30 percent and facing north. In this instance, north equates to an aspect greater than 315 degrees and less than 45 degrees	Computed from StreamStats using the mean basin slope gridded data
Forest cover	CANOPY_PCT	Area-weighted mean percentage of the drainage area that is covered by forests	Computed from StreamStats using a 30-meter cell size gridded data of percent canopy from the National Land Cover Database (2001) canopy dataset
Annual precipitation	PRECIP	Area-weighted mean annual precipitation of the drainage basin, in inches	Computed using 30-year normal precipitation gridded data for the years 1981–2010 available from the PRISM Climate Group, Oregon State University (http://www.prism.oregonstate.edu/). The precipitation grid was resampled to 30-meter cell size

Table 3.   Basin characteristics compiled and data sources for each drainage basin for selected streamgages.—Continued

Basin characteristic	StreamStats abbreviation	Description	Source
Basin length	BASLENAH	Basin length is computed from the basin outlet to basin divide using the method in the ArcHydro© Toolset. First, the longest flow path is computed to determine the endpoints for the length computation. Second, a least-cost path is defined where minimum costs are in the center of the basin and maximum costs at the boundary. This forces the flow path through the approximate centroid of the basin. The length of this cost path is the basin length, in miles	Computed from 30 m DEM NHDPlus gridded data using ArcMap© GIS commands
Minimum January temperature	JANMINTMP	Mean minimum January air temperature, in degrees Fahrenheit	Computed using 30-year normal temperature gridded data for the years 1981–2010 available from the PRISM climate group, Oregon State University (http://www.prism.oregonstate.edu/)
24 hour-2 year precipitation intensity	I24H2Y	Mean 24-hour rainfall, in inches, that can be expected to be equaled or exceeded on average every 2 years within the drainage basin	Computed using precipitation frequency gridded data available from MGS Engineering Consultants, Inc. (http://www.mgsengr.com/Precipitation.html)
24 hour-50 year precipitation intensity	I24H50Y	Mean 24-hour rainfall, in inches, that can be expected to be equaled or exceeded on average every 50 years within the drainage basin	Computed using precipitation frequency gridded data available from MGS Engineering Consultants, Inc.
24 hour-100 year precipitation intensity	I24H100Y	Mean 24-hour rainfall, in inches, that can be expected to be equaled or exceeded on average every 100 years within the drainage basin	Computed using precipitation frequency gridded data available from MGS Engineering Consultants, Inc.
Imperviousness	IMPERV	Percentage of imperviousness of the drainage basin	Computed using the 30 m National Land Cover Database (2001) Percent Developed Imperviousness gridded data (http://www.mrlc.gov/nlcd11_data.php)
Total stream length	STRMTOT	Total length of stream segment in the drainage basin, in miles	Compute using the National Hydrography Dataset flowline shapefile (http://nhd.usgs.gov/data.html).
Drainage density	DRNDENSITY	Measure of the total stream segment length within the drainage basin divided by the contributing drainage area, in square miles	Total stream length/ Drainage area
Soils	SOILINDEX	Measure of the underlying soil type within each drainage basin	Computed using SSURGO soil maps from the USDA (http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm)—see table 4
Basin shape	BSHAPE	Length of the basin, in square miles, divided by the watershed drainage area, in square miles	(Basin Length)^2 / Drainage area

The soils basin characteristic was developed from the National Resources Conservation Service Hydrologic Soils Group (U.S. Department of Agriculture, 2007) that assigned letters A–D, or combinations, based on the water transmitting ability of the soil with A soils having the least potential for runoff and D soils having the highest runoff potential. To process a basin with several different letter assignments, a numerical score was given to each letter and combination of letters and called the Hydrologic Group Index for this investigation (table 4). Table 3 lists the basin characteristic names and the data source for all the streamgages in the analysis and the StreamStats database label for the basin characteristic. Table 5 lists all the values for the basin characteristics for all the streamgages used in the investigation. Basin characteristics could not be computed for streamgages for watersheds extending into Canada because of the lack of GIS data coverage. Additionally, the precipitation intensity GIS grids only extended partway into Idaho from the Washington border; therefore, streamgages with basins in Idaho that did not have complete GIS coverage were not included in the analysis.

A state-wide, multiple-linear regression using OLS regression determined which basin characteristics of the 24 (table 3) were statistically significant for further analysis in the development of the regional regression equations. Explanatory variables that were highly correlated to other explanatory variables in the regression model were not included to avoid a condition of multicollinearity. Correlation metrics and correlation plots of the variables were used to identify variable pairs with significant multicollinearity. Additionally, streamgages draining urbanized basins (more than 5 percent impervious land cover) were not used in the regression. The percentage of land cover classified as impervious is a measure of urbanization; as urbanization and imperious land cover increases, the runoff response to rainfall typically increases along with increased magnitude of peaks over basins with minimal amount of impervious land cover with similar inputs of rainfall (Konrad, 2003). Of the 641 potential basins for use in the regression analysis, 58 basins had impervious land cover of more than 5 percent, 44 basin had more than 10 percent, and 30 basins had more than 20 percent.

Table 4. Description of the Hydrologic Group Index (HGI) used for this investigation based on U.S. Department of Agriculture SSURGO soil map Hydrologic Soil Groups.

Original SSURGO Hydrologic Soil Group	Description	HGI numerical value
A	Soils have low runoff potential when thoroughly wet. Water is transmitted freely through soil.	1
B	Soils have moderately low runoff potential when thoroughly wet. Water transmission through soil is unimpeded.	2
C	Soils have moderately high runoff potential when thoroughly wet. Water transmission through soil is somewhat restricted.	3
D	Soils have high runoff potential when thoroughly wet. Water movement through the soil is restricted or very restricted.	4
A/D	Soils have low runoff potential when drained and high runoff potential when undrained.	2.5
B/D	Soils have moderately low runoff potential when drained and high runoff potential when undrained.	3
C/D	Soils have moderately high runoff potential when drained and high runoff potential when undrained.	3.5
No data	No data were available from the SSURGO soil maps for areas of mountainous terrain, Native American reservation land, Hanford, etc.	2.5

Table 5. Basin characteristics for streamgage sites used in the investigation.

[Table 5 is an Excel® file available for download at http://dx.doi.org/10.3133/sir20165118]

Determining Regression Regions

The statewide OLS analysis identified nine basin characteristics as possible significant explanatory variables (drainage area, basin elevation, forest cover, annual precipitation, minimum January temperature, 24 hour-2 year precipitation intensity, 24 hour-50 year precipitation intensity, 24 hour-100 year precipitation intensity, and imperviousness). These nine basin characteristics were used to develop regression equations and regions for estimating the magnitude of peak flows at ungaged sites. Separate hydrologic regions were developed using a four-step approach to reduce systematic errors and regional biases:

1. A statewide model for mean annual peak streamflow was developed using multiple linear regression applied to 452 streamgages with relatively unimpaired peak flows (no large reservoirs, no urbanized basins). The final statewide model had a residual standard error of 0.65 (log ft^3/s) and an adjusted R-squared of 0.92. Precipitation, drainage area, elevation, canopy, and January minimum temperature were statistically significant explanatory variables.

2. Binary variables were created to group sites by 4-digit hydrologic unit codes (HUCs, Seaber and others, 1987), National Flood Frequency regions (Sumioka and others, 1998), basin area, and basin elevation. Sites were assigned a value of 1 (representing membership in a group) or 0 for each variable. The binary variables were added to the statewide model and tested for significance (that is, probability that its coefficient was equal to zero). 4-digit HUCs were selected as the best grouping variable because membership can be determined easily from a map and estimated flood quantiles will not abruptly change along a river other than at confluences with tributaries.

3. Four regional models for mean annual peak flows were developed representing the:

   a. Columbia Plateau, lower Snake River, and other areas of the State east of the Columbia River (HUCs 1701, 1706, 1707);

   b. eastern Cascade Range (HUCs 1702, 1703);

   c. Puget Sound Basin (HUC 1711); and

   d. Southwest Coastal (HUC 1710) and lower Columbia River valley (HUC 1708).

4. Regional model errors were calculated using streamgages in each 8-digit HUC to verify that the 8-digit HUC had been assigned to region with the lowest model error or to reassign 8-digit HUCs to an adjacent region where the regional model estimate had a lower error. This method was used to reassign Priest

(HUC 17010215) to the East Cascades (Region 2); Chief Joseph (HUC 17020005), Moses Coulee (HUC 17020012), Upper Crab (HUC 17020013), and Banks Lake (HUC 17020014) to the Columbia Plateau (Regression Region 1); and Middle Columbia-Hood (HUC 17070105) and Klickitat (HUC 17070106) to the Southwest Coastal/Lower Columbia (Region 4).

Regression Region 1 has 93 sites for the development of the regional regression equations; Regression Region 2 has 89 sites; Regression Region 3 has 142 sites; and Regression Region 4 has 139 sites (table 5). Figures 16–20 show the location of all streamgages used in this investigation and the Regression Regions.

Development of Regression Equations

The generalized-least-squares (GLS; Stedinger and Tasker, 1985) procedure was used for the multiple-linear regression analysis for each Regression Region. The GLS procedure is preferred for developing flood frequency regression equations because it weights streamgages by the length of record and accounts for cross-correlation between streamgages. Streamgages with longer peak-flow records commonly have better flood frequency estimates and thus are more heavily weighted using a weighting matrix in GLS. Additionally, when observed peaks for one streamgage temporally overlap the peaks from a nearby streamgage, cross-correlation will occur. The weighting matrix also is weighed against cross-correlation such that a group of streamgage sites that are highly correlated will have a low weight. The Weighted-Multiple-Linear Regression Program (WREG; Eng and others, 2009) was used to apply the GLS method for developing the regression equations.

The regression analyses used flood frequency analyses from 463 of 649 streamgages in and near Washington that were computed for this investigation. The flood frequency analyses from the 186 streamgages were not used because they had basins classified as urbanized, had basins extending into Canada, did not have precipitation intensity data, or were considered nested. Nested basins are defined as two or more streamgages on the same stream. If basins are nested near one another, it is expected that there would be large cross-correlation between concurrent peak flows and the observations would not be independent (Veilleux, 2009). In GLS analysis, it is best to avoid nested and highly correlated streamgages by removing one of the streamgages when two streamgages are close to each other in the same basin. In conjunction with the NHDplus, a digital vector dataset used by a GIS that incorporates the National Hydrography Dataset (NHD), the National Elevation Dataset, and the Watershed Boundary Dataset (Horizon Systems Corporation, 2015); a script was written to identify nested streamgages.
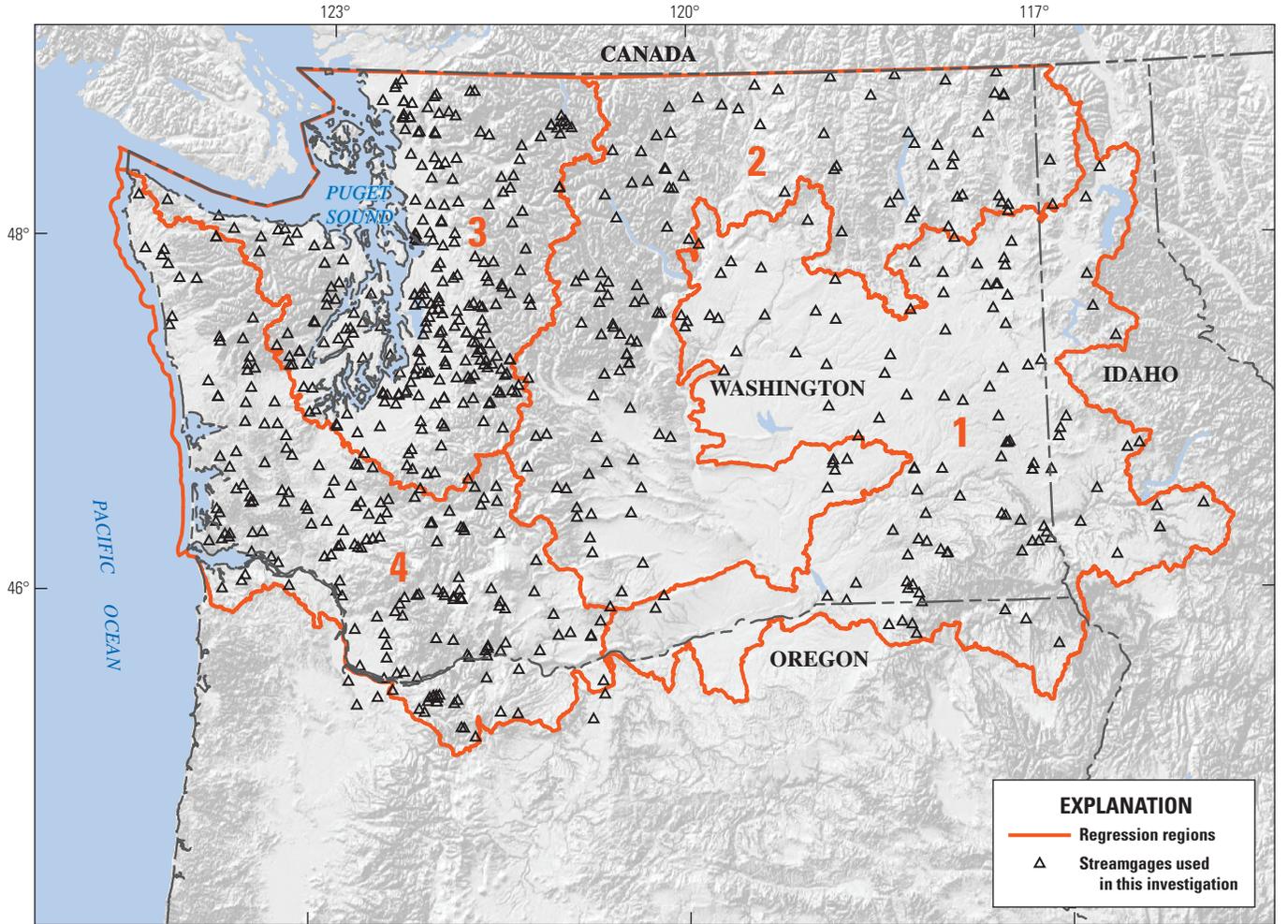
Figure 16.   Four regression regions and location of the U.S. Geological Survey streamgages used in this investigation.
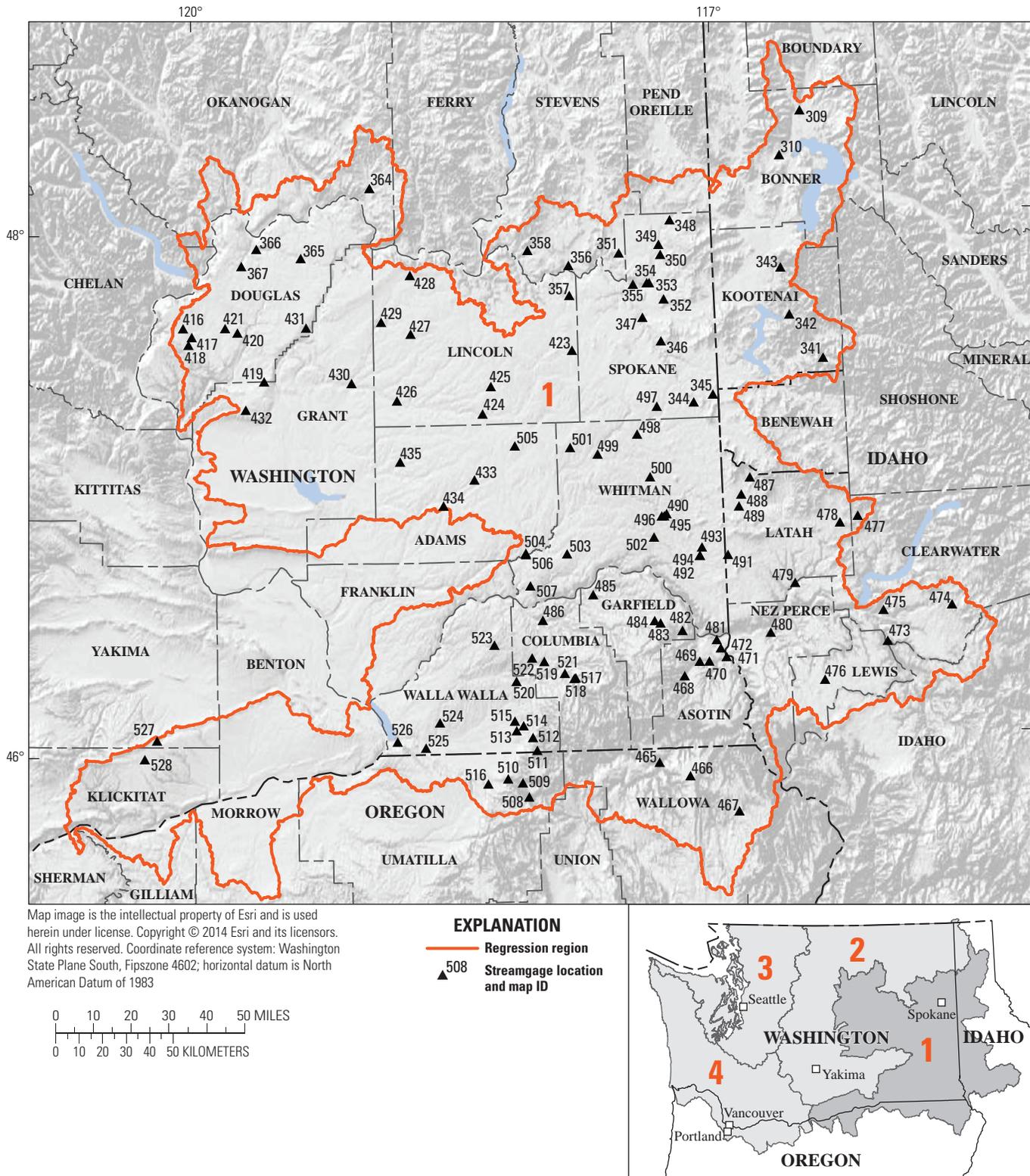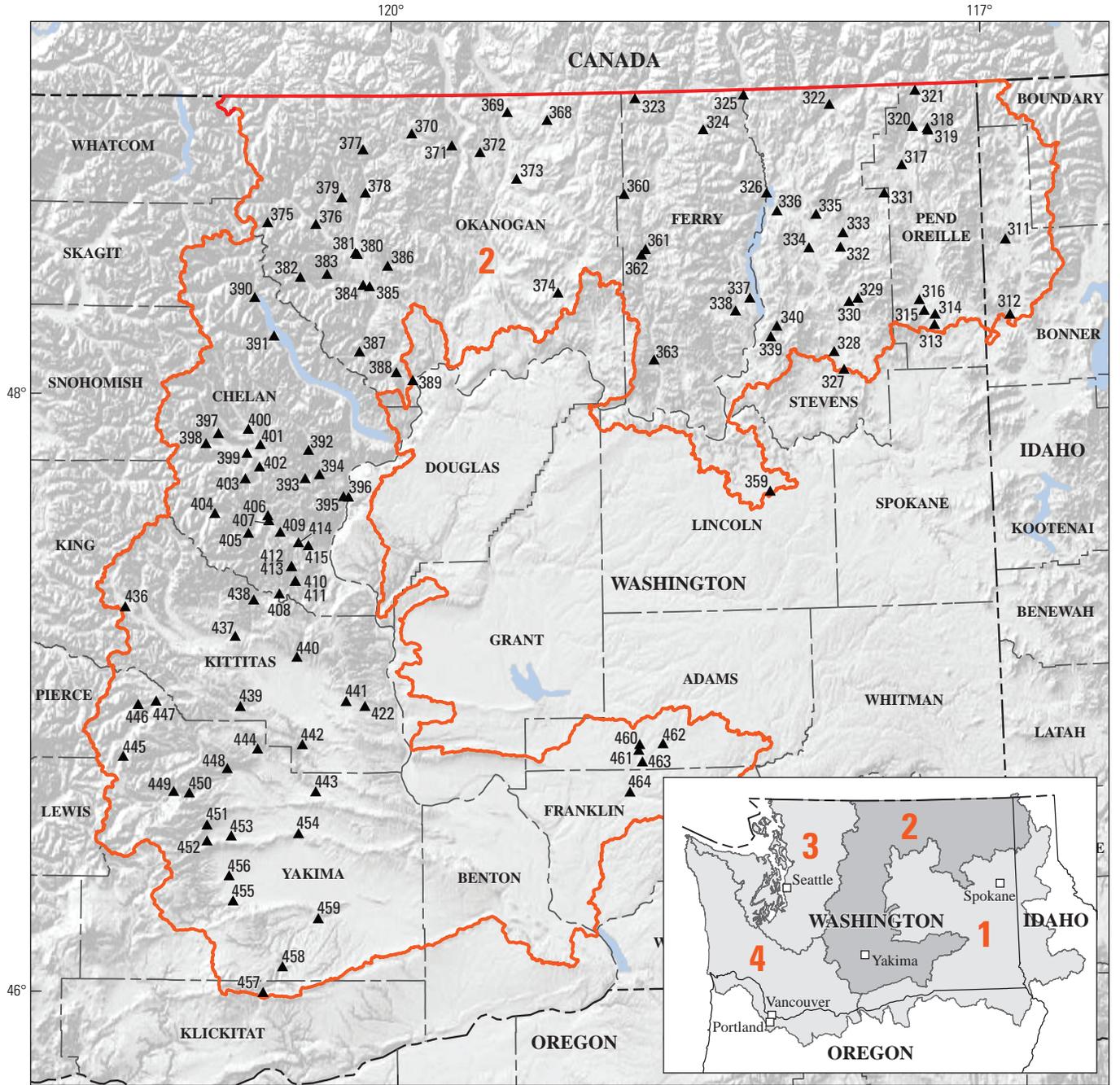
Figure 17.   Locations of U.S. Geological Survey streamgages in Regression Region 1, Washington. Streamgages are identified by map identification (ID) numbers and defined in table 5.
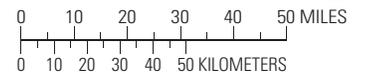
Figure 18. Locations of U.S. Geological Survey streamgages in Regression Region 2, Washington. Streamgages are identified by map identification (ID) numbers and defined in table 5.
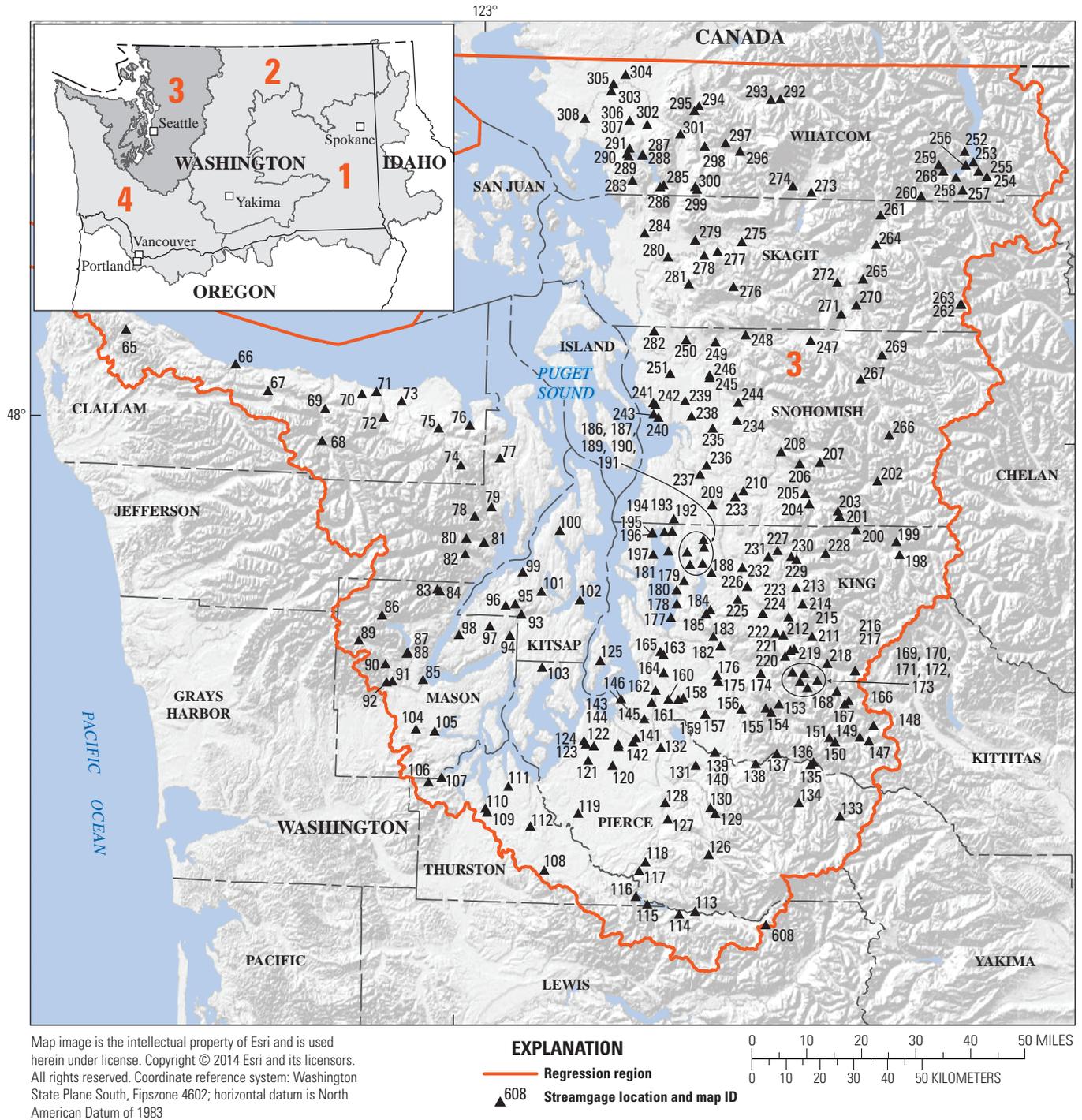
Figure 19.   Location of U.S. Geological Survey streamgages in Regression Region 3, Washington. Streamgages are identified by map identification (ID) numbers and defined in table 5.
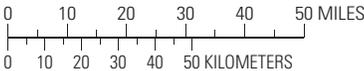
Figure 20. Location of U.S. Geological Survey streamgages in Regression Region 4, Washington. Streamgages are identified by map identification (ID) numbers and defined in table 5.

The script loops through each streamgage and determines if other streamgages are on the same stream using the NHD routing table. A second loop through determines the drainage area ratio for all pairs of streamgages that lie on the same streams. A drainage ratio of 5 was used to screen the pairs of streamgages; pairs of streamgages with a drainage ratio less than 5 were identified as a nested pair and pairs with a drainage ratio greater than 5 were not considered nested for this analysis. The identified nested pairs were reviewed and usually the streamgage with the longest record was kept in the analysis; if both had approximately the same length, usually the streamgage with the least representative drainage area was kept as a better representation of basin drainage areas.

The final distribution of the 463 streamgages used in the regression analysis shows a relatively high density of streamgages on the western side of the State, although there were no streamgages in the 8-digit HUC representing San Juan Islands in the northwestern corner of the State and some of the 8-digit HUCs along the flanks of the Olympic Mountains had less than five streamgages (fig. 21). In eastern Washington, the distribution of streamgages is fairly uniform except in the HUCs in south-central Washington. This is the driest area of Washington (fig. 2) which has few perennial streams. The hydrology in south-central Washington is similar to surrounding areas with better distributions of streamgages and thus the regression equations should be representative of the area. However, the minimum precipitation is 9.82 in/yr for Regression Region 1 and 8.86 in/yr for Regression Region 2, so streamgages in this desert region could be below these precipitation minimums and the regression equations would not be valid.

Input files were compiled for each regression region for analysis in the WREG program. The input files include flood quantiles, basin characteristics, peak-flow data, and variables used in the flood frequency analyses for each streamgage. Various combinations of the dependent variables, flood discharges (log-transformed), and independent variables were analyzed using the GLS method in WREG to develop various regression models. For each model, the transformation of the variables and the correlation smoothing function parameters are defined prior to analysis. The correlation smoothing function relates the correlation between streamflow time series at two streamgages to the geographic distance between the two streamgages. For the correlation smoothing function, minimum number of concurrent years between streamgage

is specified before they are plotted showing geographic distance against sample cross correlation. Two curve-shape parameters, alpha and theta, are then specified to define a curve that matches the general pattern of data points on the graph. The number of concurrent years that was used for the regression models was either 30 or 40 years. Using larger values of concurrent years emphasizes the use of the streamgages with the longer records, but limits the number of data points. The choice of 30 or 40 years seemed to be a reasonable balance between emphasizing the long-term streamgage and having enough data points to define a curve. The alpha parameter ranged from 0.004 to 0.01 and theta ranged from 0.985 to 0.99. After each GLS analysis of a particular combination of independent variables, the performance metrics of the regression model were recorded. Three metrics are provided by the WREG program. The first metric is the average standard error of prediction, Sp, reported as a percentage of the predicted value. About two-thirds of the estimates of flood quantiles for ungaged basins should be within the bounds of the regression equation as defined by Sp. The second metric is the pseudo coefficient of determination, R^2_{pseudo}, in percent, and it is based on the variability in the dependent variable explained by the regression. A R^2_{pseudo} value of 100 indicates a perfect regression model. The third metric is the standard model error reported as a percentage of the observed value. Eng and others (2009) provides detailed explanations and equations for these performance metrics. In Regression Regions 1 and 2, the average standard error of prediction for the final regression equations ranged from 69.12 to 119.59 percent. On the western side of the State, Regression Regions 3 and 4, the standard error was much lower and ranged from 43.22 to 58.04 percent (table 6). The R^2_{pseudo} values in the final models ranged from 68.66 to 90.66 for Regression Regions 1 and 2 and from 92.35 to 95.44 for Regression Regions 3 and 4 (table 6). The WREG program also provides leverage and influence statistics that identify unusual individual sites and provides the level of significance of regression parameters as "p-values" or probability values from a Student's t-distribution. Those p-values that exceeded 0.05 were flagged as not significant. Only one equation from the final selection of equations, Regression Region 2 equation for the 0.5 AEP flood discharge, had a parameter, forest canopy, which had a p-value of greater than 0.05. The final regression equations for each region are shown in table 6.
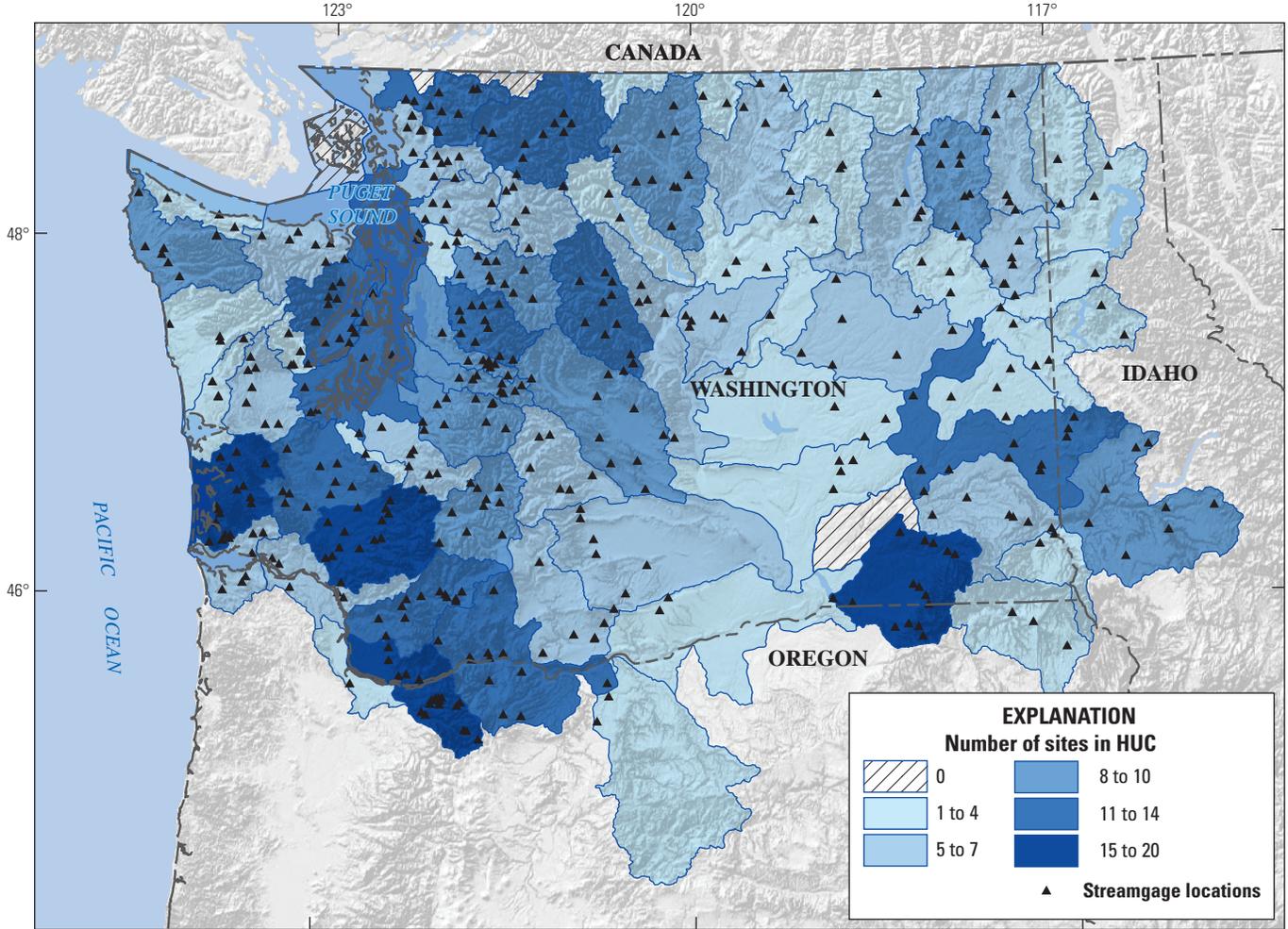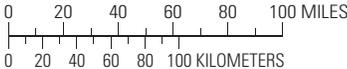
Figure 21. Streamgages used in the regression analysis by hydrologic unit codes (HUCs), Washington.

Table 6. Regression equations for estimating flood discharge at ungaged sites in Washington for various annual exceedance probabilities.

[**Form of regression equation:** Q, flood discharge in cubic feet per second for indicated annual exceedance probability. S_p: Average standard error of prediction. R^2_{pseudo}: Pseudo coefficient of determination. **SEM:** Average standard error of model. **Range of values:** A, drainage area; P, mean annual precipitation, from PRISM Climate Group (2015), Oregon State University for the 1981–2010 normal; CAN, area-weighted forest canopy, computed from National Land Cover Database (2001) canopy dataset]

Regression region	Number of stations used in analysis	Form of the regression equation	Annual exceedance probability	Constant a	Coefficents b	Coefficents c	Coefficents d	S_p (percent)	R^2_{pseudo}	SEM (percent)	Range of values A	Range of values P	Range of values CAN
1	93	$Q = aA^b10^{cP}/10^{dCAN}$	0.5	3.846	0.745	0.032	0.0078	95.04	85.1	90.72	0.25–3,304	9.82–52.45	0.0–77.4
			0.2	12.106	0.713	0.028	0.0098	71.93	87.66	68.33			
			0.1	22.080	0.695	0.026	0.0107	70.67	86.72	66.91			
			0.04	42.170	0.674	0.024	0.0117	77.42	83.38	73.12			
			0.02	63.826	0.661	0.023	0.0124	84.76	80.32	79.92			
			0.01	92.470	0.649	0.022	0.0130	93.55	77.01	88.08			
			0.005	129.42	0.637	0.021	0.0136	103.98	73.32	97.75			
			0.002	193.20	0.624	0.020	0.0143	118.69	68.66	111.33			
2	89	$Q = aA^b10^{cP}/10^{dCAN}$	0.5	1.445	0.877	0.0253	0.0020	77.18	90.59	73.81	0.42–1,322	8.86–84.20	0.00–81.8
			0.2	5.768	0.860	0.0214	0.0061	69.12	90.66	66.09			
			0.1	12.106	0.847	0.0194	0.0082	72.24	89.29	68.92			
			0.04	26.485	0.833	0.0174	0.0105	81.23	86.56	77.31			
			0.02	43.853	0.823	0.0162	0.0119	89.22	84.33	84.76			
			0.01	68.549	0.814	0.0151	0.0132	96.94	82.32	91.92			
			0.005	103.514	0.806	0.0141	0.0145	106.38	80.05	100.69			
			0.002	169.82	0.796	0.0130	0.0159	119.59	77.09	112.09			
3	142	$Q = aA^bP^c$	0.5	0.0294	0.899	1.750	–	43.22	95.44	42.5	0.08–2,605	33.29–168.0	
			0.2	0.0565	0.895	1.700	–	44.35	95.18	43.56			
			0.1	0.0794	0.895	1.670	–	45.6	94.93	44.74			
			0.04	0.115	0.896	1.634	–	48.07	94.4	47.11			
			0.02	0.146	0.897	1.609	–	50.48	93.9	49.44			
			0.01	0.182	0.898	1.587	–	51.77	93.63	50.66			
			0.005	0.221	0.899	1.567	–	54.16	93.12	52.97			
			0.002	0.281	0.900	1.542	–	57.66	92.35	56.36			
4	139	$Q = aA^b10^{cP}$	0.5	15.346	0.911	0.0073	–	52.45	94.24	51.6	0.18–2,230	11.94–186.6	
			0.2	29.785	0.913	0.0060	–	50.61	94.47	49.72			
			0.1	42.073	0.915	0.0054	–	50.48	94.46	49.54			
			0.04	60.117	0.918	0.0047	–	51.67	94.19	50.64			
			0.02	75.162	0.920	0.0043	–	52.93	93.93	51.83			
			0.01	91.411	0.921	0.0040	–	54.21	93.67	53.04			
			0.005	108.89	0.923	0.0036	–	55.51	93.41	54.28			
			0.002	134.28	0.926	0.0033	–	58.04	92.89	56.7			

Uncertainty in the Regression Estimates

The regression equations only provide an estimate of the true flood discharge for an ungaged site, so it is important to know the uncertainty associated with the estimate. This uncertainty can be visualized in a plot of the observed values against the predicted values from the regression models (fig. 22). For individual estimates at an ungaged site, the uncertainty of the regressions can be defined by the prediction interval or the range of values for a given confidence level that the true value lies within. For this analysis and the example that follows, a 90-percent prediction probability is used. Tasker and Driver (1988) show the following equation for the computation of the prediction interval:

$$\frac{Q}{T} < Q < QT \qquad (6)$$

where

Q is the flood discharge estimated from one of the regression equations in table 6, and

$$T = 10^{\left[t_{\left(\frac{\alpha}{2}, n-p\right)} S_i \right]} \qquad (7)$$

where

$t_{\left(\frac{\alpha}{2}, n-p\right)}$ is the critical value, t, from the Student's t-distribution, $\alpha = 0.10$ for 90-percent prediction intervals with $n-p$ degrees of freedom, where n is the number of sites used in the regression equation and p is the number of variables plus 1 (values are listed in table 7); and

S_i is the standard error of prediction for site i, and is computed with the following equation:

$$S_i = \left[\sigma_\delta^2 + X_i \left(X^T \Lambda^{-1} X \right)^{-1} X_i^T \right]^{0.5} \qquad (8)$$

where

σ_δ^2 is the model error variance computed by WREG (table 7);

X_i is a row vector of the explanatory variables for site i, starting with the number 1 and followed by the values of the basin characteristic as used in the regression equation;

$\left(X^T \Lambda^{-1} X \right)^{-1}$ is the covariance matrix for the regression coefficients computed by WREG (table 7); and

X_i^T is the transpose of X_i

Table 7 provides the model error variance, Student's t critical value, and the covariance matrix, for each of the regional regression equations. These values are needed to compute the 90-percent prediction interval for any flood-quantile estimation from the regression equations.

Application of Regression

The application of the regression to an ungaged basin can be accomplished in three different ways:

1. The application of the equations can be done by hand as explained below with the knowledge of the basin characteristics for the ungaged site.

2. The StreamStats online program for Washington State (water.usgs.gov/osw/streamstats/Washington.html) will compute the basin characteristics and apply the equations to determine the flood quantiles for the ungaged site. It is expected that StreamStats will incorporate the equations outlined in this investigation after this report has been published.

3. The flood discharges and prediction intervals can be estimated using the Flood Q Tools (Excel® spreadsheet, sir20165118_floodqtools.xlsm, available for download at http://dx.doi.org/10.3133/sir20165118).
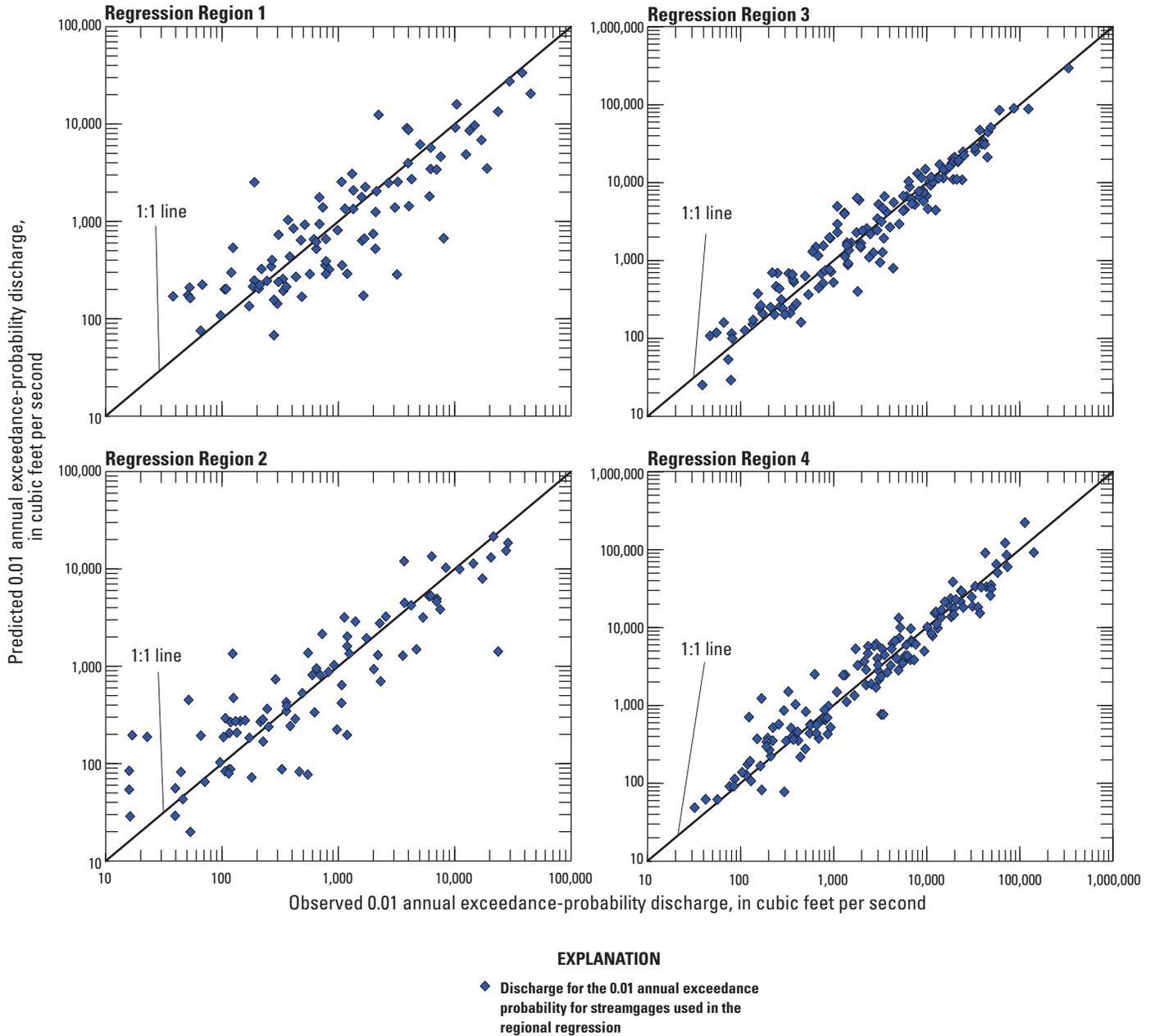
Figure 22.   Comparison of the computed discharges for the 0.01 annual exceedance probability using observed peak-flow data at streamgages and predicted discharges for the 0.01 annual exceedance probability from the regional regression equations for the four regression regions in Washington.

Table 7.   Regression model error variance and covariance matrix, values used to determine prediction intervals for the regional flood frequency regression equations for Washington.

[**AEP:** Annual exceedance probability. σ_δ^2: Regression model error variance, in log units. $t_{(\alpha/2,\ n-p)}$: Critical value from Student's t distribution for the 90-percent probability with $n-p$ degrees of freedom where n is the number of sites used in the regression equation and p is the number of variables plus 1. $(\mathbf{X^T\Lambda^{-1}X})^{-1}$: Covariance matrix. Intercept, y-axis intercept of regression equation; A, drainage area, in square miles; P, mean annual precipitation; CAN, canopy, in percent of basin]

				Regression Region 1			
				$(\mathbf{X^T\Lambda^{-1}X})^{-1}$			
AEP	σ_δ^2	$t_{(\sigma/2,\ n-p)}$		**Intercept**	**A** (m²)	**P** (inches)	**CAN** (percent)
0.5	0.113	1.662	Intercept	1.916×10^{-2}	-2.309×10^{-3}	-6.548×10^{-4}	1.923×10^{-4}
			A	-2.309×10^{-3}	1.615×10^{-3}	7.910×10^{-6}	-1.545×10^{-5}
			P	-6.548×10^{-4}	7.910×10^{-6}	4.033×10^{-5}	-1.659×10^{-5}
			CAN	1.923×10^{-4}	-1.545×10^{-5}	-1.659×10^{-5}	1.005×10^{-5}
0.2	0.072	1.662	Intercept	1.523×10^{-2}	-1.937×10^{-3}	-4.681×10^{-4}	1.250×10^{-4}
			A	-1.937×10^{-3}	1.153×10^{-3}	8.381×10^{-6}	-8.763×10^{-6}
			P	-4.681×10^{-4}	8.381×10^{-6}	2.821×10^{-5}	-1.157×10^{-5}
			CAN	1.250×10^{-4}	-8.763×10^{-6}	-1.157×10^{-5}	7.067×10^{-6}
0.1	0.070	1.662	Intercept	1.579×10^{-2}	-2.021×10^{-3}	-4.793×10^{-4}	1.265×10^{-4}
			A	-2.021×10^{-3}	1.167×10^{-3}	9.371×10^{-6}	-8.558×10^{-6}
			P	-4.793×10^{-4}	9.371×10^{-6}	2.888×10^{-5}	-1.191×10^{-5}
			CAN	1.265×10^{-4}	-8.558×10^{-6}	-1.191×10^{-5}	7.282×10^{-6}
0.04	0.081	1.662	Intercept	1.867×10^{-2}	-2.387×10^{-3}	-5.734×10^{-4}	1.536×10^{-4}
			A	-2.387×10^{-3}	1.379×10^{-3}	1.169×10^{-5}	-1.056×10^{-5}
			P	-5.734×10^{-4}	1.169×10^{-5}	3.468×10^{-5}	-1.438×10^{-5}
			CAN	1.536×10^{-4}	-1.056×10^{-5}	-1.438×10^{-5}	8.792×10^{-6}
0.02	0.093	1.662	Intercept	2.149×10^{-2}	-2.739×10^{-3}	-6.679×10^{-4}	1.816×10^{-4}
			A	-2.739×10^{-3}	1.596×10^{-3}	1.387×10^{-5}	-1.280×10^{-5}
			P	-6.679×10^{-4}	1.387×10^{-5}	4.049×10^{-5}	-1.684×10^{-5}
			CAN	1.816×10^{-4}	-1.280×10^{-5}	-1.684×10^{-5}	1.029×10^{-5}
0.01	0.108	1.662	Intercept	2.469×10^{-2}	-3.136×10^{-3}	-7.782×10^{-4}	2.151×10^{-4}
			A	-3.136×10^{-3}	1.849×10^{-3}	1.634×10^{-5}	-1.560×10^{-5}
			P	-7.782×10^{-4}	1.634×10^{-5}	4.727×10^{-5}	-1.970×10^{-5}
			CAN	2.151×10^{-4}	1.560×10^{-5}	-1.970×10^{-5}	1.202×10^{-5}
0.005	0.126	1.662	Intercept	2.828×10^{-2}	-3.577×10^{-3}	-9.054×10^{-4}	2.548×10^{-4}
			A	-3.577×10^{-3}	2.142×10^{-3}	1.909×10^{-5}	-1.905×10^{-5}
			P	-9.054×10^{-4}	1.909×10^{-5}	5.509×10^{-5}	-2.299×10^{-5}
			CAN	2.548×10^{-4}	-1.905×10^{-5}	-2.299×10^{-5}	1.401×10^{-5}
0.002	0.152	1.662	Intercept	3.319×10^{-2}	-4.177×10^{-3}	-1.082×10^{-3}	3.106×10^{-4}
			A	-4.177×10^{-3}	2.547×10^{-3}	2.292×10^{-5}	-2.404×10^{-5}
			P	-1.082×10^{-3}	2.292×10^{-5}	6.594×10^{-5}	-2.756×10^{-5}
			CAN	3.106×10^{-4}	-2.404×10^{-5}	-2.756×10^{-5}	1.677×10^{-5}

Table 7. Regression model error variance and covariance matrix, values used to determine prediction intervals for the regional flood frequency regression equations for Washington.—Continued

				Regression Region 2			
				$(X^T \Lambda^{-1} X)^{-1}$			
AEP	σ_δ^2	$t_{(\sigma/2,\, n-p)}$		Intercept	A (m²)	P (inches)	CAN (percent)
0.5	0.082	1.6630	Intercept	1.635×10^{-2}	-2.438×10^{-3}	-1.267×10^{-4}	-1.153×10^{-4}
			A	-2.438×10^{-3}	1.546×10^{-3}	1.063×10^{-6}	5.239×10^{-7}
			P	-1.276×10^{-4}	1.063×10^{-6}	5.535×10^{-6}	-1.668×10^{-6}
			CAN	-1.153×10^{-4}	5.239×10^{-7}	-1.668×10^{-6}	3.360×10^{-6}
0.2	0.068	1.6630	Intercept	1.541×10^{-2}	-2.310×10^{-3}	-1.148×10^{-4}	-1.072×10^{-4}
			A	-2.310×10^{-3}	1.387×10^{-3}	2.157×10^{-6}	7.600×10^{-7}
			P	-1.148×10^{-4}	2.157×10^{-6}	4.809×10^{-6}	-1.449×10^{-6}
			CAN	-1.072×10^{-4}	7.600×10^{-7}	-1.449×10^{-6}	3.027×10^{-6}
0.1	0.073	1.6630	Intercept	1.714×10^{-2}	-2.556×10^{-3}	-1.265×10^{-4}	-1.198×10^{-4}
			A	-2.556×10^{-3}	1.533×10^{-3}	2.759×10^{-6}	3.661×10^{-7}
			P	-1.265×10^{-4}	2.759×10^{-6}	5.287×10^{-6}	-1.614×10^{-6}
			CAN	-1.198×10^{-4}	3.661×10^{-7}	-1.614×10^{-6}	3.404×10^{-6}
0.04	0.088	1.6630	Intercept	2.068×10^{-2}	-3.076×10^{-3}	-1.523×10^{-4}	-1.458×10^{-4}
			A	-3.076×10^{-3}	1.877×10^{-3}	3.440×10^{-6}	-5.506×10^{-7}
			P	-1.523×10^{-4}	3.440×10^{-6}	6.459×10^{-6}	-2.027×10^{-6}
			CAN	-1.458×10^{-4}	-5.506×10^{-7}	-2.027×10^{-6}	4.240×10^{-6}
0.02	0.102	1.6630	Intercept	2.370×10^{-2}	-3.523×10^{-3}	-1.747×10^{-4}	-1.678×10^{-4}
			A	-3.523×10^{-3}	2.180×10^{-3}	3.947×10^{-6}	-1.405×10^{-6}
			P	-1.747×10^{-4}	3.947×10^{-6}	7.508×10^{-6}	-2.397×10^{-6}
			CAN	-1.678×10^{-4}	-1.405×10^{-6}	-2.397×10^{-6}	4.968×10^{-6}
0.01	0.116	1.6630	Intercept	2.669×10^{-2}	-3.964×10^{-3}	-1.968×10^{-4}	-1.897×10^{-4}
			A	-3.964×10^{-3}	2.479×10^{-3}	4.496×10^{-6}	-2.281×10^{-6}
			P	-1.968×10^{-4}	4.496×10^{-6}	8.543×10^{-6}	-2.764×10^{-6}
			CAN	-1.897×10^{-4}	-2.281×10^{-6}	-2.764×10^{-6}	5.691×10^{-6}
0.005	0.132	1.6630	Intercept	3.001×10^{-2}	-4.465×10^{-3}	-2.219×10^{-4}	-2.135×10^{-4}
			A	-4.465×10^{-3}	2.832×10^{-3}	4.958×10^{-6}	-3.449×10^{-6}
			P	-2.219×10^{-4}	4.958×10^{-6}	9.776×10^{-6}	-3.210×10^{-6}
			CAN	-2.135×10^{-4}	-3.449×10^{-6}	-3.210×10^{-6}	6.523×10^{-6}
0.002	0.155	1.6630	Intercept	3.446×10^{-2}	-5.143×10^{-3}	-2.559×10^{-4}	-2.453×10^{-4}
			A	-5.143×10^{-3}	3.318×10^{-3}	5.549×10^{-6}	-5.195×10^{-6}
			P	-2.559×10^{-4}	5.549×10^{-6}	1.148×10^{-5}	-3.834×10^{-6}
			CAN	-2.453×10^{-4}	-5.195×10^{-6}	-3.834×10^{-6}	7.665×10^{-6}

Table 7. Regression model error variance and covariance matrix, values used to determine prediction intervals for the regional flood frequency regression equations for Washington.—Continued

				Regression Region 3		
				$(X^T \Lambda^{-1} X)^{-1}$		
AEP	σ_δ^2	$t_{(\sigma/2,\, n-p)}$		**Intercept**	**A** (m²)	**P** (inches)
0.5	0.031	1.6559	Intercept	3.976×10^{-2}	9.135×10^{-5}	-2.031×10^{-2}
			A	9.135×10^{-5}	4.064×10^{-4}	-3.276×10^{-4}
			P	-2.031×10^{-2}	-3.276×10^{-4}	1.070×10^{-2}
0.2	0.033	1.6559	Intercept	4.405×10^{-2}	7.119×10^{-5}	-2.241×10^{-2}
			A	7.119×10^{-5}	4.346×10^{-4}	-3.400×10^{-4}
			P	-2.241×10^{-2}	-3.400×10^{-4}	1.176×10^{-2}
0.1	0.034	1.6559	Intercept	4.871×10^{-2}	7.815×10^{-5}	-2.475×10^{-2}
			A	7.815×10^{-5}	4.697×10^{-4}	-3.709×10^{-4}
			P	-2.475×10^{-2}	-3.709×10^{-4}	1.297×10^{-2}
0.04	0.038	1.6559	Intercept	5.639×10^{-2}	1.106×10^{-4}	-2.867×10^{-2}
			A	1.106×10^{-4}	5.327×10^{-4}	-4.341×10^{-4}
			P	-2.867×10^{-2}	-4.341×10^{-4}	1.503×10^{-2}
0.02	0.041	1.6559	Intercept	6.330×10^{-2}	1.363×10^{-4}	-3.219×10^{-2}
			A	1.363×10^{-4}	5.914×10^{-4}	-4.901×10^{-4}
			P	-3.219×10^{-2}	-4.901×10^{-4}	1.688×10^{-2}
0.01	0.043	1.6559	Intercept	6.826×10^{-2}	1.469×10^{-4}	-3.471×10^{-2}
			A	1.469×10^{-4}	6.309×10^{-4}	-5.263×10^{-4}
			P	-3.471×10^{-2}	-5.263×10^{-4}	1.819×10^{-2}
0.005	0.047	1.6559	Intercept	7.530×10^{-2}	1.649×10^{-4}	-3.828×10^{-2}
			A	1.649×10^{-4}	6.911×10^{-4}	-5.797×10^{-4}
			P	-3.828×10^{-2}	-5.797×10^{-4}	2.007×10^{-2}
0.002	0.052	1.6559	Intercept	8.546×10^{-2}	1.904×10^{-4}	-4.345×10^{-2}
			A	1.904×10^{-4}	7.792×10^{-4}	-6.569×10^{-4}
			P	-4.345×10^{-2}	-6.569×10^{-4}	2.277×10^{-2}

Table 7. Regression model error variance and covariance matrix, values used to determine prediction intervals for the regional flood frequency regression equations for Washington.—Continued

			Regression Region 4			
				$(X^T \Lambda^{-1} X)^{-1}$		
AEP	σ_δ^2	$t_{(\sigma/2,\ n-p)}$		**Intercept**	**A** (m²)	**P** (inches)
0.5	0.045	1.656	Intercept	4.633×10^{-3}	-5.756×10^{-4}	-3.314×10^{-4}
			A	-5.756×10^{-4}	4.384×10^{-4}	1.945×10^{-6}
			P	-3.314×10^{-4}	1.945×10^{-6}	3.270×10^{-5}
0.2	0.042	1.656	Intercept	4.694×10^{-3}	-5.551×10^{-4}	-3.316×10^{-4}
			A	-5.551×10^{-4}	4.210×10^{-4}	1.681×10^{-6}
			P	-3.316×10^{-4}	1.681×10^{-6}	3.227×10^{-5}
0.1	0.041	1.656	Intercept	4.976×10^{-3}	-5.679×10^{-4}	-3.492×10^{-4}
			A	-5.679×10^{-4}	4.296×10^{-4}	1.524×10^{-6}
			P	-3.492×10^{-4}	1.524×10^{-6}	3.368×10^{-5}
0.04	0.043	1.656	Intercept	5.531×10^{-3}	-6.098×10^{-4}	-3.864×10^{-4}
			A	-6.098×10^{-4}	4.616×10^{-4}	1.296×10^{-6}
			P	-3.864×10^{-4}	1.296×10^{-6}	3.698×10^{-5}
0.02	0.045	1.656	Intercept	5.998×10^{-3}	-6.482×10^{-4}	-4.180×10^{-4}
			A	-6.482×10^{-4}	4.915×10^{-4}	1.106×10^{-6}
			P	-4.180×10^{-4}	1.106×10^{-6}	3.985×10^{-5}
0.01	0.047	1.656	Intercept	6.467×10^{-3}	-6.873×10^{-4}	-4.498×10^{-4}
			A	-6.873×10^{-4}	5.221×10^{-4}	8.927×10^{-7}
			P	-4.498×10^{-4}	8.927×10^{-7}	4.273×10^{-5}
0.005	0.049	1.656	Intercept	6.935×10^{-3}	-7.268×10^{-4}	-4.816×10^{-4}
			A	-7.268×10^{-4}	5.534×10^{-4}	6.506×10^{-7}
			P	-4.816×10^{-4}	6.506×10^{-7}	4.563×10^{-5}
0.002	0.053	1.656	Intercept	7.663×10^{-3}	-7.937×10^{-4}	-5.319×10^{-4}
			A	-7.937×10^{-4}	6.073×10^{-4}	2.921×10^{-7}
			P	-5.319×10^{-4}	2.921×10^{-7}	5.033×10^{-5}

The Flood Q Regression Tool is patterned after a similar spreadsheet by McCarthy and others (2016) and includes directions in the spreadsheet; and contains the regression equations for each region to estimate the flood discharge at an ungaged site and the covariance matrices needed to estimate the prediction intervals. The user selects the appropriate regression region and then enters the drainage area of the ungaged basin, the mean annual precipitation, and, for Regression Region 1 or 2 only, the percent canopy. The computed results include flood discharges and prediction intervals for the eight AEPs.

An example of applying the regional regression equation is shown here for calculating the 0.01 AEP flood quantile (100-year recurrence-interval flood discharge) for USGS streamgage 12488500, American River near Nile, Washington. This station and its watershed is located in Regression Region 2 (Map ID 447, fig. 18). The drainage area is 79.22 mi², the percent canopy coverage is 68.1 percent, and the mean annual precipitation is 67.6 in. (table 5). The regional regression equation for the 0.01 AEP flood quantile, $Q_{0.01}$ in Regression Region 2 (table 6) is:

$$Q_{0.01} = \frac{a \times A^b 10^{cP}}{10^{dCAN}} \qquad (9)$$

where

a	is a regression constant;
b, c, and d	are regression coefficients;
A	is drainage area, in mi²;
P	is mean annual precipitation, in inches; and
CAN	is the percentage canopy coverage.

After entering the basin characteristics (table 5) and the regression constant and coefficients (table 6) into equation 9, the equation can now be written as follows:

$$
\begin{aligned}
Q_{0.01} &= \frac{68.549 \times 79.22^{0.814} 10^{0.0151 \times 67.6}}{10^{0.0132 \times 68.1}} \\
&= \frac{25,258.7981}{7.9235} \\
&= 3,187.8344 \\
&= 3,190\,\text{ft}^3/\text{s} \left(\text{rounded to three significant figures}\right).
\end{aligned}
$$

The 90-percent prediction interval for this flood-quantile estimate is calculated using equations 6–8. The critical value from the Student's t-distribution, $t_{\left(\frac{\alpha}{2}, n-p\right)}$, for the Regression Region 2 equations is 1.6630. The X_i vector for this example is:

$$X_i = \left[1, \log10(79.22), 67.6, 68.1\right].$$

The model error variance, σ_δ^2, from table 7 is 0.116, and the following table is the covariance matrix, $\left(X^T \Lambda^{-1} X\right)^{-1}$, from table 7 for the 0.01 AEP flood quantile for Regression Region 2:

	Intercept	A	P	CAN
Intercept	2.669×10^{-2}	-3.964×10^{-03}	-1.968×10^{-04}	-1.897×10^{-04}
A	-3.964×10^{-03}	2.479×10^{-03}	4.496×10^{-06}	-2.281×10^{-06}
P	-1.968×10^{-04}	4.496×10^{-06}	8.543×10^{-06}	-2.764×10^{-06}
CAN	-1.897×10^{-04}	-2.281×10^{-06}	-2.764×10^{-06}	5.691×10^{-06}

Matrix algebra is used to compute the product of $X_i \left(X^T \Lambda^{-1} X\right)^{-1} X_i^T$ by first multiplying X_i^T by the covariance matrix to obtain $\left(X^T \Lambda^{-1} X\right)^{-1} X_i^T$ and then multiplying this product by X_i. For this example,

$$X_i \left(X^T \Lambda^{-1} X\right)^{-1} X_i^T = 0.00868359.$$

The standard error of prediction, S_i, for the example site can now be determined by adding the model error variance with the last result:

$$S_i = \left[0.116 + 0.00868359\right]^{0.5} = 0.35311.$$

Therefore, T from equation 7 is:

$$T = 10^{(1.663)(0.35311)} = 3.8656.$$

Finally, the 90-percent prediction interval can be computed from equation 6.

$Q_{0.01}/T < Q_{0.01} < Q_{0.01} \times T$ or 825 ft³/s < $Q_{0.01}$ < 12,300 ft³/s; which is to say that given the assumptions of the study, one can be 90-percent confident that the true value of the estimate for this example is between 825 and 12,300 ft³/s.

Limitations of Regression

The regression equations were computed for unregulated streams, and therefore the equations are not valid for streams with dams or diversions that affect the peak flows. Impervious land cover associated with urbanization can increase the peak flows over similar areas without impervious land cover. The regression analysis for this investigation was restricted to basins with less than 5 percent impervious land cover so the equations should be used only for these types of basins. Sauer and others (1983) provide techniques for estimating peak flows on urban streams. The regression equations are only valid in the region for which they were developed and for basins where the basin characteristics inputs are within the range of values used to create the equations (table 6). Using basin characteristics obtained from sources different from those used in this investigation (table 3) may provide unpredictable results.

Weighted Estimates of Magnitude and Frequency of Floods at Gaged Sites

The accuracy of the statistical analysis outlined in section, "Magnitude and Frequency of Floods at Streamgaging Sites," is proportional to the length of the streamgage record and begins to degrade for estimates for the smaller (more rare) AEPs. The International Advisory Committee on Water Data (1982) suggests that refinements can be made to the frequency curve derived from the streamgage data when a second independent estimate of flood frequency is available. The regional regression equations as described in section, "Regression Equations for Estimating Flood Magnitudes in Ungaged Basins," provide a second independent estimate to the EMA statistical analysis that was applied to the streamgage data.

The MSE of the two independent estimates of the flood quantiles is used to compute a weighted estimate of the flood quantile, Q_{wtd}, for which the uncertainty of the weighted estimate will be less than the uncertainty of either independent estimate. Q_{wtd} can be computed as follows:

$$\log Q_{wtd} = \frac{MSE\left[\log\left(Q_{reg}\right)\right] \times \log\left(Q_{sta}\right) + MSE\left[\log\left(Q_{sta}\right)\right] \times \log\left(Q_{reg}\right)}{MSE\left[\log\left(Q_{sta}\right)\right] + MSE\left[\log\left(Q_{reg}\right)\right]} \tag{10}$$

where

Q_{sta} is the flood-quantile estimate from the EMA analysis of the station peak-flow data, in cubic feet per second; and

Q_{reg} is the flood-quantile estimate from the regional regression equations, in cubic feet per second.

The USGS Weighted Independent Estimates (WIE) program (Cohn and others, 2012) was designed to compute the weighted flood quantiles (Q_{wtd}) and confidence intervals using output data from the PeakFQ and WREG programs. The MSE for Q_{sta} is computed using the output from the PeakFQ program including the mean and standard deviation of the annual peaks, the weighted skew coefficient, and the number of years of systematic record. The MSE for Q_{reg} is computed using the regression equation estimate from inputs that include the standard error of the model, the number of explanatory variables, and the covariance matrix of the explanatory variables used in the regression equations (Cohn and others, 2012). Output from the WIE program includes the weighted value of the flood quantile (Q_{wtd}) in cubic feet per second and the 95-percent confidence interval of the weighed flood quantile.

The flood quantiles for the streamgage data, the regional regression equations, and the weighted values computed by the WIE program with the 95-percent confidence intervals for all the Washington streamgages (594 streamgages) were compiled in table 8. Generally, especially for streamgages with less than 50 years of peak-flow record (International Advisory Committee on Water Data, 1982), the best flood-quantile estimate for a streamgage site will be the weighted estimate from the WIE program. The weighted flood-quantile estimates were not computed for streamgages where the basin characteristics could not be computed or those classified as urbanized because the regression equations could not be computed or the equations were not applicable.

Table 8.   Flood quantiles for streamgages in Washington State.

[Table 8 is an Excel® file available for download at http://dx.doi.org/10.3133/sir20165118.]

Estimating Flood Magnitudes at Ungaged Sites near a Streamgage

If the site of interest is ungaged, but near a long-term (10 years or more of non-zero annual peak discharge) gaged site on the same stream, a more accurate estimate of the flood quantiles, than an estimate from just the regional regression equation, can be obtained by including the flood frequency information from the long-term streamgage (Ries, 2007). The ungaged site is considered near if the drainage area is between 0.5 and 1.5 times the drainage area of the long-term streamgage. The process requires two steps. First obtain the streamgage-based estimate for the ungaged site, $Q_{(u)g}$, from the equation:

$$Q_{(u)g} = \left[\frac{A_u}{A_g}\right]^b Q_{(g)wtd} \tag{11}$$

where

A_u is the drainage area of the ungaged site;

A_g is the drainage area of the streamgage on the same stream as the ungaged site;

b is the exponent of the drainage area variable in the regional regression equation (table 6); and

$Q_{(g)wtd}$ is the weighted estimate of the flood quantile for the streamgage (table 8).

Next, compute the weighted estimate for the ungaged site, $Q_{(u)wtd}$, using the result from equation 11 and the following equation:

$$Q_{(u)wtd} = \frac{2\Delta A}{A_g} Q_{(u)reg} + \left[1 - \frac{2\Delta A}{A_g}\right] Q_{(u)g} \qquad (12)$$

where

ΔA is the absolute value of the difference between the drainage areas of the streamgage and ungaged site; and

$Q_{(u)reg}$ is the flood-quantile estimate from the regional regression equation (table 6) for the ungaged site.

An example application of these two equations is provided below for an ungaged site on Mineral Creek just downstream from the confluence of the mainstem of Mineral Creek and the North Fork of Mineral Creek at the Mineral Creek Rd crossing. In this example, the 0.02 AEP flood discharge will be estimated using equations 11 and 12. Downstream from this site is the long-term streamgage, Mineral Creek near Mineral, WA (USGS streamgage 12083000). From StreamStats, the long-term streamgage has a drainage area (A_g) of 76.75 mi^2 and the ungaged site has a drainage area (A_u) of 56.96 mi^2, well within the $0.5 \times A_g$ limits of application of this method. The weighted 0.02 flood discharge for station 12083000, $Q_{(g)wtd}$, equals 10,600 ft^3/s (table 8). The mean annual precipitation for the ungaged site is 85.1 inches and both sites are within the Regression Region 3. Using the regional regression equation for Regression Region 3 (table 6), the 0.02 flood discharge equals 6,990 ft^3/s [$6,988.05 = 0.146 (56.96^{0.897})(85.1^{1.609})$]. The regional regression equation also provides b in equation 11 which equals 0.897.

Applying equation 11:

$$Q_{(u)g} = \left[\frac{56.96}{76.75}\right]^{0.897} (10,600)$$

$$= 8,110 \text{ ft}^3/\text{s}.$$

Applying equation 12 for a final answer:

$$Q_{(u)wtd} = \frac{39.58}{76.75} 6,990 + \left[1 - \frac{39.58}{76.75}\right] 8,110 = 3,604.7 + 3,927.7$$

$$= 7,532.4 \text{ or } 7,530 \text{ ft}^3/\text{s} \text{ (rounded to 3 digits).}$$

Estimates of flood discharge using this same drainage-ratio method can be determined with the "Flood Q Ratio Tool" (Excel® spreadsheet, sir20165118_floodqtools.xlsm, available for download at http://dx.doi.org/10.3133/sir20165188). A user selects the nearby streamgage from a drop-down list of long-term, non-regulated, and non-urbanized streamgages in Washington that is on the same stream as the ungaged site of interest. The user then enters the drainage area of the ungaged basin and an initial estimate is shown. Next, the user enters the other basin characteristics needed to calculate the regression estimates, and finally, the weighted estimate for the ungaged site is computed.

Summary

Statistically significant trends in peak flows were noted in some of the annual series of data for streamgages in unregulated, non-urbanized basins, but the occurrence and strength of the trends often depended on the start and end dates of the record. Trends of the observed systematic annual peaks were compiled using the two-sided Kendall *tau* statistic for 83 sites with over 50 years of record; and 21 streamgages had p-values less than or equal to 0.05. Of those 21 streamgages, 5 streamgages had negative trends and 16 had positive trends. Eastern Washington had three streamgages with significant trends that were all negative; and in western Washington, all but two of the streamgages with p-values equal or less than 0.05 showed an increasing trend. Evidence of decadal positive trends and negative trends were noted in the record by plotting the fraction of streamgages with positive and negative trends for every 10-year period. Decades centered around 1945, 1956, 1970, 1992, and 2004 showed positive trends and significantly decreasing peaks during the decades centered around 1936, 1990, 2000, and 2010. By plotting the trends for all combinations of starting and ending dates for all the streamgage records, a positive step trend was indicated between about 1940 and 1950 at many streamgages and a less persistent positive step trend in the late 1960s. Flood frequencies computed using data from a period of record that begins after a positive step may be higher than estimates based on records that include periods before the step. Conversely, negative step trends can result in flood frequencies that are lower than estimates based on records that include periods before the trend.

The observed annual peak flows for 594 streamgages in the State of Washington with at least 10 years of non-zero, unregulated flows were used to compute flood magnitudes for eight annual exceedance probabilities by fitting a log Pearson Type III distribution to the annual peak flows. The Expected Moments Algorithm was used to process historical, non-exceedance, and censored peak-flow information. The Multiple Grubbs Beck low-outlier test was used to censor potentially influential low flows from the computation of the parameters of the fitted frequency curve. Peak-flow data included all the recorded peaks through water year 2014. Flood magnitudes were computed for the 0.5, 0.2, 0.1, 0.04, 0.02, 0.01, 0.005, and 0.002 annual exceedance probabilities.

A new regional skew coefficient, -0.07, was determined for the Pacific Northwest Region using 290 streamgages with at least 35 years of annual peak-flow data located in Idaho, Oregon, Washington and western Montana. The calculated mean square error of the new regional skew coefficient is 0.180 which is much less than the mean square error of 0.3025 from the generalized skew coefficient from the national map. The reduction in the error of the regional skew coefficient translates into a reduction in the error of the final flood quantiles over previous reports that used the generalized skew coefficient from the national map. The regional skew coefficient was weighted with the station skew for all but 61 streamgages which had drainages with greater than 5 percent impervious land cover.

Regression equations were developed for four regions in Washington using the frequency data and basin characteristics for 463 streamgages including some Idaho and Oregon streamgages near the borders with Washington. The four regression regions were originally defined by the 4-digit hydrologic unit codes and refined based on the residuals of a state-wide multiple linear regression. Regression Regions 1 and 2 are east of the Cascade Mountains and use 93 and 89 sites, respectively, in their regressions. Of the 24 basin characteristics calculated and used in preliminary regression analysis, annual precipitation, drainage area, and percent of canopy coverage were used in the regression equations for Regression Regions 1 and 2. Regression Region 3 and 4 are west of the Cascade Mountains and used 142 and 139 sites, respectively, in their regressions. The basin characteristics of annual precipitation and drainage area were the two independent variables used in the regression equations for Regression Regions 3 and 4. Average standard error of prediction for the regional regression equations ranged from 69.12 to 119.59 percent for Regression Regions 1 and 2 and from 43.22 to 58.04 percent for Regression Regions 3 and 4. The pseudo coefficient of determination (where a value of 100 signifies a perfect regression model) ranged from 68.66 to 90.66 for Regions 1 and 2 and from 92.35 to 95.44 for Regions 3 and 4.

References Cited

Anderson, D.G., and Bodhaine, G.L., 1957, Flood of 1956 in the Esquatzel Coulee area, Washington: U.S. Geological Survey Open-File Report 57–3, 51 p.

Anderson, I.E., 1948, Floods of the Puyallup and Chehalis River basins, Washington: U.S. Geological Survey Water-Supply Paper 968-B, p. 61–124.

Bailey, E.G., 1960, Floods in the Nooksack River Basin, *in* Water resources of the Nooksack River Basin and certain adjacent streams: State of Washington, Department of Conservation, Division of Water Resources Water-Supply Bulletin No. 12, p. 95–98.

Bodhaine, G.L., and Robinson, W.H., 1952, Floods in western Washington, frequency and magnitude in relation to drainage basin characteristics: U.S. Geological Survey Circular 191, 124 p.

Bodhaine, G.L., and Thomas, D.M., 1960, Floods in Washington—Magnitude and frequency: U.S. Geological Survey Open-File Report 60–14, 25 p.

Bodhaine, G.L., and Thomas, D.M., 1964, Magnitude and frequency of floods in the United States—Part 12, Pacific slope basins in Washington and upper Columbia River basin: U.S. Geological Survey Water-Supply Paper 1687, 337 p.

Cayan, D.R., Dettinger, M.D., Diaz, H.F., and Graham, N.E., 1998, Decadal variability of precipitation over western North America, Journal of Climate, v. 11, p. 3,148–3,166.

Cohn, T.A., and Lins, H.F., 2005, Natures's style–Naturally trendy: Geophysical Research Letters, v. 32, L23402, 5 p.

Cohn, T.A., Berenbrock, Charles, Kiang, J.E., and Mason, R.R., Jr., 2012, Calculating weighted estimates of peak streamflow statistics: U.S. Geological Survey Fact Sheet 2012–3038, 2 p.

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, no. 8, p. 5,047–5,058.

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.

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

Cummans, J.E., Collings, M.R., and Nassar, E.G., 1975, Magnitude and frequency of floods in Washington: U.S. Geological Survey Open-File Report 74–336, 46 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.

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.

Hall, J.E., Beechie, T.J., and Pess, G.R., 2014, Influence of climate and land cover change on river discharge in the North Fork Stillaguamish River: Seattle, Washington, Final Report to the Stillaguamish Tribe of Indians, Northwest Fisheries Science Center, National Oceanic and Atmospheric Administration National Marine Fisheries Service, 41 p.

Haushild, W.L., 1979, Estimation of floods of various frequencies for the small ephemeral streams in eastern Washington: U.S. Geological Survey Water-Resources Investigations 79–81, 22 p.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources—Volume 49 of studies in environmental science: Amsterdam, Elsevier Science, 522 p.

Horizon Systems Corporation, 2015, NHDPlus home: Horizon Systems Web page, accessed August 6, 2015, at http://www.horizon-systems.com/NHDPlus/index.php.

Hubbard, L.L., 1991, Floods of January 9–11, 1990 in northwestern Oregon and southwestern Washington: U.S. Geological Survey Open-File Report 91-172, 10 p.

Hubbard, L.L., 1994, Floods of November 1990 western Washington: U.S. Geological Survey Open-File Report 93-631, 10 p.

Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency: Bulletin 17B of the Hydrology Subcommittee, Office of Water Data Coordination, U.S. Geological Survey, Reston, Va., 183 p.

Jain, Shaleen, and Lall, Upmanu, 2001, Floods in a changing climate—Does the past represent the future?: Water Resources Research, v. 37, no. 12, p. 3,193–3,205.

Jones, J.L., 2011, Popular myths about flooding in western Washington: U.S. Geological Survey Fact Sheet 2011–3146, 4 p.

Knowles, S.M., and Sumioka, S.S., 2001, The National Flood-Frequency program—Methods for estimating flood magnitude and frequency in Washington, 2001: U.S. Geological Survey Fact Sheet FS-016-01, 4 p.

Konrad, C.P., 2003, Effects of urban development on floods: U.S. Geological Survey Fact Sheet FS-076-03, 4 p.

Konrad, C.P., and Booth, D.B., 2002, Hydrologic trends associated with urban development for selected streams in the Puget Sound Basin, western Washington: U.S. Geological Survey Water-Resources Investigations Report 02–4040, 40 p.

Kresch, D.L., and Dinicola, Karen, 1997, What causes floods in Washington state?: U.S. Geological Survey Fact Sheet FS-228-96, 2 p.

Livingston, V.E., Jr., 1969, Geologic history and rocks and minerals of Washington: Olympia, Wash., State of Washington, Department of Natural Resources, Division of Mines and Geology, Information Circular No. 45, 42 p.

Longfield, R.J., 1974, Floods of January 1974 in Washington: U.S. Geological Survey Open-File Report, 13 p.

Mantua, Nate, 2015, PDO Index: Joint Institute for the Study of the Atmosphere and Ocean Web page, accessed June 22, 2015, at http://research.jisao.washington.edu/pdo/PDO.latest.

Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.M., and Francis, R.C., 1997, A Pacific interdecadal climate oscillation with impacts on salmon production: Bulletin of the American Meteorological Society, v. 78, p. 1,069–1,079.

Mastin, M.C., 1998, Flood potential of South Prairie Creek, Pierce County, Washington: U.S. Geological Survey Water-Resources Investigations Report 98–4009, 75 p.

Mastin, M.C., 2009, Watershed models for decision support for inflows to Potholes Reservoir, Washington: U.S. Geological Survey Scientific Investigation Report 2009–5081, 54 p.

McCabe, G.J., and Wolock, D.M., 2002, A step increase in streamflow in the conterminous United States, Geophysical Research Letters, v. 29, doi:10.1029/2002GL015999, 4 p.

McCarthy, P.M., Sando, Roy, Sando, S.K., and Dutton, D.M., 2016, Methods for estimating streamflow characteristics at ungaged sites in western Montana based on data through water year 2009: U.S. Geological Survey Scientific Investigations Report 2015–5019-G, 19 p., http:/dx.doi.org/10.3133/sir20155019G.

Milly, P.C.D., Betancourt, Julio, Falkenmark, Malin, Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., and Stouffer, R.J., 2008, Stationarity is dead—Whither water management?: Science, v. 319, p. 573–574.

National Center for Atmospheric Research, 2015, Southern Oscillation Index (SOI): National Center for Atmospheric Research, Climate Analysis Section Data Catalog Web page, accessed June 26, 2015, at http://www.cgd.ucar.edu/cas/catalog/climind/soi.html.

National Land Cover Database, 2015, National Land Cover: U.S. Geological Survey Earth Resources Observation and Science Center database, accessed August 19, 2015, at https://lta.cr.usgs.gov/NLC.

Neiman, P.J., Schick, L.J., Ralph, M.F., Hughes, Mimi, and Wick, G.A., 2011, Flooding in western Washington—the connection to atmospheric rivers: Journal of Hydrometeorology, v. 12, p. 1,337–1,358.

Novak, C.E., 1985, WRD data reports preparation guide: U.S. Geological Survey, Water Resources Division, 199 p.

PRISM Climate Group, 2015, 30-year normals: Northwest Alliance for Computational Science and Engineering, Oregon State University database, accessed June 4, 2015, at http://prism.oregonstate.edu/normals.

Rantz, S.E., and Riggs, H.C., 1949, Magnitude and frequency of floods in the Columbia River basin: U.S. Geological Survey Water-Supply Paper 1080, 472 p.

Richardson, Donald, 1965, Effect of logging on runoff in upper Green River basin, Washington—A progress report: U.S. Geological Survey Open-File Report 65–135, 45 p.

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

Salathe E.P., Jr., Hamlet, A.F., Mass, C.F., Lee, Se-Yeun, Stumbaugh, Matt, and Steed, Richard, 2014, Estimates of twenty-first-century flood risk in the Pacific Northwest based on regional climate model simulations: Journal of Hydrometeorology, v. 15, p. 1,881–1,898.

Sauer, V.B., Thomas, W.O., Jr., Sticker, V.A., and Wilson, K.V., 1983, Flood characteristics of urban watersheds in the United States: U.S. Geological Survey Water-Supply Paper 2207, 63 p.

Seaber, P.R., Kapinos, F.P., and Knapp, G.L., 1987, Hydrologic unit maps: U.S. Geological Survey Water-Supply Paper 2294, 63 p.

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.

Stewart, J.E., and Bodhaine, G.L., 1961, Floods in the Skagit River basin, Washington: U.S. Geological Survey Water-Supply Paper 1527, 66 p., accessed November 17, 2014, at http://pubs.er.usgs.gov/publication/wsp1527.

Sumioka, S.S., Kresch, D.L., and Kasnick, K.D., 1998, Magnitude and frequency of floods in Washington: U.S. Geological Survey Water-Resources Investigations Report 97–4277, 91 p.

Tasker, G.D., and Driver, N.E., 1988, Nationwide regression models for predicting urban runoff water quality at unmonitored sites: Water Resources Bulletin, v. 24, no. 5, p. 1,091–1,101.

Tohver, I.M., Hamlet, A.F., and Lee, Se-Yeun, 2014, Impacts of 21st-century climate change on hydrologic extremes in the Pacific Northwest region of North America: Journal of the American Water Resources Association, v. 50, no. 6, p. 1,461–1,476.

Trenberth, K.E., 1984, Signal versus noise in the Southern Oscillation: Monthly Weather Review, v. 112, p. 326–332.

U.S. Department of Agriculture, Natural Resources Conservation Service, 2007, Hydrologic soil groups, Chapter 7: Natural Resources Conservation Service, Part 630 Hydrology National Engineering Handbook, accessed February 19, 2016, at www.nrcs.usda.gov/wps/portal/nrcs/detailfull/ny/technical/?cid=stelprdb1043063.

Veilleux, A.G., 2009, Bayesian GLS regression for regionalization of hydrologic statistics, floods and Bulletin 17 skew: Ithaca, New York, Cornell University, Master's thesis, 155 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., http://dx.doi.org/10.3133/fs20133108.

Ward, P.J., Jongman, Brenden, Kummu, Matti, Dettinger, M.D., Weiland, F.C. Sperna, and Winsemius, H.C., 2014, Strong influence of El Niño Southern Oscillation on flood risk around the world: Proceedings of the National Academy of Sciences of the United States of America, November 4, 2014, v. 111, no. 44, p. 15,659–15,664.

Warner, M.D., Mass, C.F., and Salathé, E.P., Jr., 2014, Changes in winter atmospheric rivers along the North American West Coast in CMIP5 climate models: Journal of Hydrometeorology, v. 16, p. 118–128.

Appendix A. Regional Skewness Regression Analysis for the Pacific Northwest, Including Idaho, Montana, Oregon, and Washington

By Andrea G. Veilleux

Introduction to Statistical Analysis of Regional Skew

For the log-transformation of annual peak discharges, Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) recommends using a weighted average of the station skew coefficient and a regional skew coefficient to help improve estimates of annual exceedance probability (AEP) discharges (eq. 2). 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 of annual peak-discharge data. Their method of regional analysis of skewness estimators accounts for the precision of the skew-coefficient estimate for each streamgage, which depends on the length of record for each streamgage and the accuracy of an ordinary least-squares (OLS) regional mean skewness. 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 skewness 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 station 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 discharges at pairs of streamgages, an alternate regression procedure was developed to provide both stable and defensible results for regional skewness (Veilleux, 2011; Lamontange and others, 2012; Veilleux and others, 2012). This alternate procedure is referred to as the Bayesian WLS/ Bayesian GLS (B-WLS/B-GLS) regression framework (Veilleux, 2011; Veilleux and others, 2011, 2012). It uses an OLS analysis to fit an initial regional skewness 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 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 compute various diagnostic statistics.

The Idaho-Montana-Oregon-Washington (hereinafter Pacific Northwest study area or PNW) regional skew study described here used the EMA with MGB (EMAw/MGB) to estimate the station skew and its mean square error. Because EMAw/MGB allows for the censoring of potentially influential low flows (PILFs), as well as the use of estimated interval discharges 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 discharges 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 Skewness Model

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

Ordinary Least-Squares Analysis

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

$$\tilde{\mathbf{y}}_{OLS} = \mathbf{X}\hat{\boldsymbol{\beta}}_{OLS} \tag{A1}$$

where

$\mathbf{X}$	is an $(n \times k)$ matrix of basin characteristics;
$\tilde{\mathbf{y}}_{OLS}$	are the estimated regional skewness 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 skewness values $\tilde{\mathbf{y}}_{OLS}$ are then used to calculate unbiased station-regional skewness variances using the equations reported in Griffis and Stedinger (2009). These station-regional skewness variances are based on the regional OLS estimator of the skewness coefficient instead of the station skewness estimator, thus making the weights in the subsequent steps relatively independent of the station skewness estimates.

Weighted Least-Squares Analysis

A B-WLS analysis is used to develop estimators of the regression coefficients for each regional skewness 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, $SE\left(\hat{\boldsymbol{\beta}}_{WLS}\right)$, the model error variance, $\sigma^2_{\delta,B-GLS}$, pseudo coefficient of determination, pseudo-R^2_δ, and the average variance of prediction at a streamgage that is not used in the regional model, AVP_{new}.

Data Analysis

This regional skew study is based on annual peak-discharge data from 461 streamgages in Washington, Idaho, and Oregon, as well as from the neighboring State of Montana. The annual peak-discharge data through September 2012 were downloaded from the USGS National Water Information System (NWIS) database. In addition to the peak-discharge data, 10 basin characteristics for each of the 461 streamgages were available 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, January mean minimum temperature, mean 24-hour 100-year precipitation intensity, and mean 6-hour 100-year precipitation intensity.

Part of Idaho in the Snake River plain was omitted from the regional skew analysis because of the extent of regulation, groundwater-surface water interactions, and infiltration in the area. The area is consistent with the boundaries omitted from a concurrent USGS investigation into regional peak-flow regression analysis for Idaho (Woods and others, 2016, region 0 in figure 3). The regional skew analysis also did not include streamgages draining urbanized basins; and therefore, the regional skew does not reflect these types of flood distributions. For the investigation of the Washington streamgages as described in this report, the percent of the basin classified as impervious was used to categorize urbanized basins. Those basins with more than 5 percent impervious land cover were considered urbanized. Sixty-one sites matched this classification. In table 8 of the main report, the flood quantiles for these sites are provided, but only the station skew was used in the calculation of the flood quantiles—no weighting of the regional skew was used.

Station Skewness Estimators

To estimate the station logarithm base10 (log) skew coefficient, G, and its mean square error, MSE_G, the skew study used the results of the EMAw/MGB analysis described in 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 CSG, contained in the record of annual peak discharges for a streamgage. For this analysis PeakFQ version 7.1 (Veilleux and others, 2014), which combines EMAw/MGB, was used. Documentation for PeakFQ is available at http://water.usgs.gov/software/PeakFQ/. PeakFQ was used to generate the station log estimates of G and the corresponding MSE_G, assuming a log-Pearson Type III distribution and generally employing a MGB test for PILF screening. EMA estimates, based on annual peak-discharge data through September 30, 2012, of G and MSE_G are listed in table A1 for the 461 streamgages evaluated for the PNW regional skew study (see section, "Magnitude and Frequency of Floods at Gaged Sites" 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 compute the precision of the skewness estimators is no longer simply the number of annual peak discharges at a streamgage. Instead, a more complex calculation was used to take into account the availability of historical information and censored values. Although historical information and censored peaks provide valuable information, they 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-discharge data types available for a site.

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 $\left(MSE(\hat{G})\right)$ 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 MSE of the at-site skew when only the systematic record is analyzed $\left(MSE(\hat{G}_S)\right)$ versus the MSE of the at-site skew when the all of the data, including historical and censored data, is analyzed $\left(MSE(\hat{G}_C)\right)$.

$$P_{RL} = \frac{P_s \times MSE\left(\hat{G}_S\right)}{MSE\left(\hat{G}_C\right)} \qquad (A2)$$

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;

$MSE\left(\hat{G}_S\right)$ is the estimated MSE of the skew when only the systematic record is analyzed; and

$MSE\left(\hat{G}_C\right)$ is the estimated MSE of the skew when all of the data, including historical and censored data, is 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, where the historical period is defined as the period which begins with the first water year for which peak-flow data is available and ends with the most recent water year for which peak-flow data is available), 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 skewness 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 the same hydrologic response to a given storm, 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 dataset (Gruber and Stedinger, 2008). To determine if two sites are redundant and thus represent the same hydrologic conditions, two types of information are considered: (1) whether their basins are nested, and (2) the ratio of the basin drainage areas.

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:

$$SD_{ij} = \frac{D_{ij}}{\sqrt{0.5\left(DRNAREA_i + DRNAREA_j\right)}} \qquad (A3)$$

where

D_{ij} is the distance between centroids of basin i and basin j;

$DRNAREA_i$ is the drainage area at site i; and

$DRNAREA_j$ is the drainage area at site j.

The drainage area ratio (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, 2009):

$$DAR = Max\left[\frac{DRNAREA_i}{DRNAREA_j}, \frac{DRNAREA_j}{DRNAREA_i}\right] \qquad (A4)$$

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 similar and the basins are nested. Previous studies 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 table A1.

From the 126 identified possible redundant streamgage-pairs, 113 were determined to be redundant but only 36 streamgages were actually 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 290 streamgages to use in the PNW regional skew study. Figure A1 shows the location of these 180 streamgages used in PNW regional skew study.

EXPLANATION

—— 2-Digit hydrologic unit boundary

▲ Streamgage site used in regional skew analysis

Figure A1.   Locations of streamgages selected for the regional skew study in the Pacific Northwest.

Table A1.   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 pseudo record length less than 35 years]

Regional skew index No.	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

Table A1. Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index No.	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?
128	12334550	27	46.375	-113.109	-0.58	0.24	no-P
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

Table A1.   Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index No.	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?
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
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

Table A1. Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index No.	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?
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
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

Table A1.   Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index No.	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?
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
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

Table A1.   Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index No.	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?
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
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

Table A1.   Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index No.	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?
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
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

Table A1.   Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index No.	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?
50	14021000	57	45.666	-118.404	0.19	0.10	Yes
122	14022500	63	45.473	-118.574	0.15	0.09	Yes
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

Table A1.   Description of streamgages that were evaluated for use in the regional skew analysis for the Pacific Northwest.—Continued

Regional skew index No.	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?
88	14172000	55	44.282	-122.578	-0.25	0.11	Yes
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

Unbiasing the Station Estimators

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

$$\hat{\gamma}_i = \left[1 + \frac{6}{P_{RL,i}}\right] G_i \qquad \text{(A5)}$$

where

$\hat{\gamma}_i$ is the unbiased station sample skewness estimate for site i;

$P_{RL,i}$ is the pseudo record length for site i as calculated in equation A2; and

G_i is the traditional biased station skewness estimator for site i from the flood frequency analysis.

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

$$Var\left[\hat{\gamma}_i\right] = \left[1 + \frac{6}{P_{RL,i}}\right]^2 Var\left[G_i\right] \qquad \text{(A6)}$$

where

$Var\left[G_i\right]$ is calculated using (Griffis and Stedinger, 2009):

$$Var\left(\hat{G}\right) = \left[\frac{6}{P_{RL}} + a\left(P_{RL}\right)\right]$$
$$\times\left[1 + \left(\frac{9}{6} + b\left(P_{RL}\right)\right)\hat{G}^2 + \left(\frac{15}{48} + c\left(P_{RL}\right)\right)\hat{G}^4\right] \qquad \text{(A7)}$$

where

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

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

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

Estimating the Mean Square Error of the Skewness Estimator

There are several possible ways to estimate MSE_G. The approach used by EMA (taken from Cohn and others [2001, eq. 55]) generates a first order estimate of the MSE_G, which should perform well when censored data are present. Another option is to use the Griffis and Stedinger (2009) formula in equation A7 (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 been 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 skewness model because it is more robust and relatively independent of the station skewness estimator. This methodology was used in previous regional skew studies (Eash and others, 2013; Southard and Veilleux, 2014).

Cross-Correlation Models

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

$$\hat{\rho}\left(\hat{\gamma}_i, \hat{\gamma}_j\right) = Sign\left(\hat{\rho}_{ij}\right) cf_{ij} \left|\hat{\rho}_{ij}\right|^k \qquad \text{(A8)}$$

where

$\hat{\rho}_{ij}$ is the cross-correlation of concurrent annual peak discharge for two streamgages;

$Sign\left(\hat{\rho}_{ij}\right)$ is plus or minus one depending on the sign of $\hat{\rho}_{ij}$;

κ is a constant between 2.8 and 3.3; and

cf_{ij} is a factor that accounts for the sample size difference between stations and their concurrent record length, is defined as follows:

$$cf_{ij} = CY_{ij} / \sqrt{\left(P_{RL,i}\right)\left(P_{RL,j}\right)} \qquad \text{(A9)}$$

where

CY_{ij} is the pseudo record length of the period of concurrent record; and

$P_{RL,i}$, and $P_{RL,j}$ are the pseudo record length corresponding to sites i and j, respectively (see eq. A2).

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 calculated. Because of the use of censored data and historical data, the effective concurrent record length calculation 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 calculate the concurrent years of record between site i and site j:

$$CY_{ij} = \left(YE_{ij} - YB_{ij} + 1\right)\left(\frac{P_{RL,i}}{P_{H,i}}\right)\left(\frac{P_{RL,j}}{P_{H,j}}\right). \qquad \text{(A10)}$$

The computed 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.

PNW Study Area Cross-Correlation Model of Concurrent Annual Peak Discharge

A cross-correlation model for the logarithm of the annual peak discharges in the PNW study area were 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 discharge at two sites, ρ_{ij}, to various basin characteristics were considered. A logit model, termed the Fisher Z Transformation $(Z = \log[(1+r)/(1-r)])$ (Kendall and Stuart, 1961), provided a convenient transformation of the sample correlations ρ_{ij} from the (-1, +1) range to the (-∞ +∞) range. The adopted models for estimating the cross-correlations of concurrent annual peak discharge at two stations, which used the distance between basin centroids, D_{ij}, as the only explanatory variable, are

$$\rho_{ij} = \frac{\exp\left(2Z_{ij}\right) - 1}{\exp\left(2Z_{ij}\right) + 1} \qquad \text{(A11)}$$

where

Z_{ij} is $0.21 + \exp\left(-0.17 - 0.0058 D_{ij}\right)$

The fitted relation between Z and distance between basin centroids together with the plotted sample data from the 411 streamgage-pairs of data are shown in figure A2. 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 A3. The cross correlation model was used to estimate streamgage-to-streamgage cross correlations for concurrent annual peak discharges at all pairs of streamgages in the regional skew study.
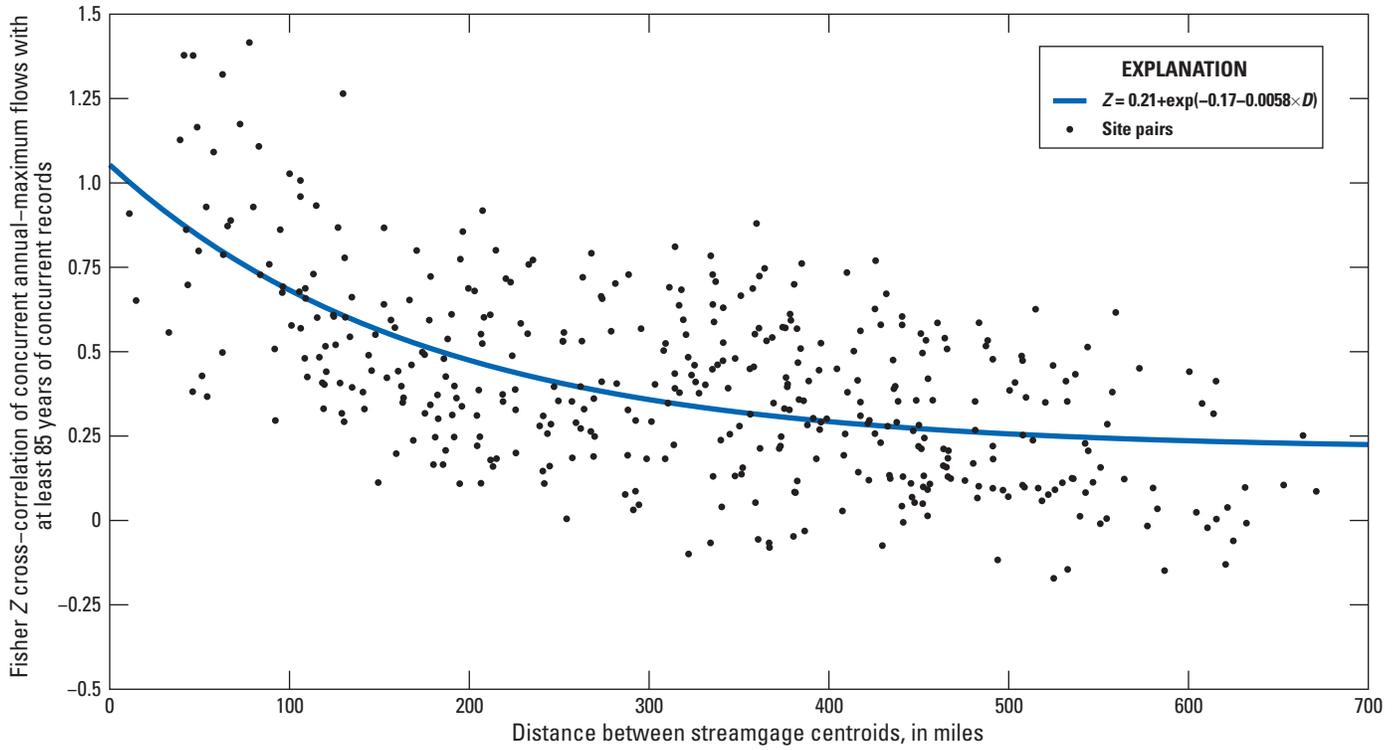
Figure A2. Relation between Fisher Z transformed cross-correlation of logs of annual peak discharge and distance between basin centroids for the regional skew study in the Pacific Northwest.
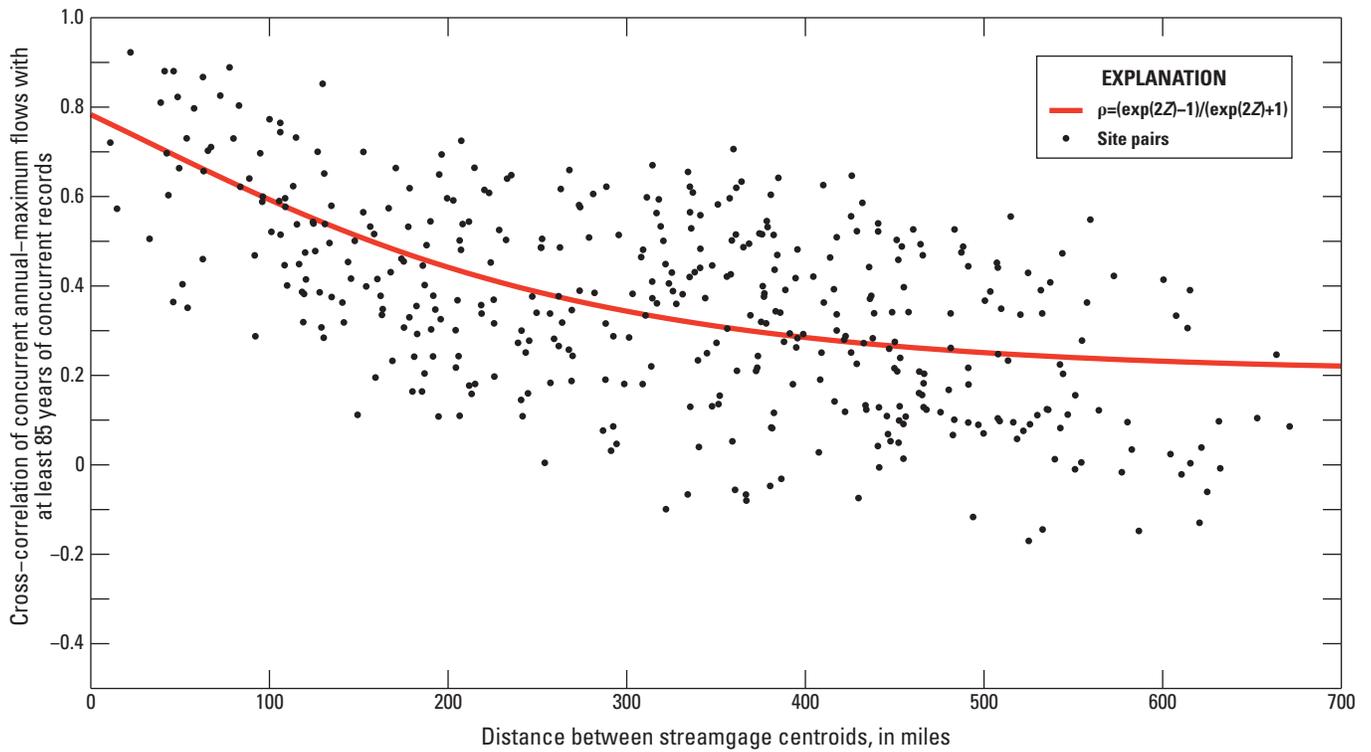

Figure A3. Relation between untransformed cross-correlation of logs of annual peak discharge and distance between basin centroids for the regional skew study in the Pacific Northwest.

Pacific Northwest Regional Skew Study Results

The results of the PNW regional skew study using the B-WLS/B-GLS regression methodology are provided here. All of the available basin characteristics (drainage area, measures of basin elevation, measures of basin precipitation, and HUC subregions) were initially considered as explanatory variables in the regression analysis for regional skew.

The best regional skew model is classified as having the smallest model error variance, σ_δ^2, and largest pseudo-R_δ^2. The pseudo R_δ^2 describes the estimated fraction of the variability in the true skewness from streamgage-to-streamgage explained by each model (Gruber and others, 2007; Parrett and others, 2011). The addition of any of the available basin characteristics were not statistically significant and did not produce a pseudo-R_δ^2 greater than 3 percent. This indicates that the inclusion of a basin characteristic as an explanatory variable in the regression did not help explain the variability in the true skewness. The addition of a basin characteristic is not warranted as the increased model complexity only provides a very small gain in model precision. Thus, the CONSTANT model is chosen as the best regional skewness model for the PNW study area. The final results for the constant skewness model, denoted CONSTANT, for the PNW study area using 290 streamgages with at least 35 years of pseudo record length are provided in table A2.

Thus, the CONSTANT model is chosen as the best regional skewness model for the PNW study area. A constant model does not explain any variability in the true skews, so the pseudo-R_δ^2 equals 0. The posterior mean of the model error variance, R_δ^2, for the CONSTANT model is $R_\delta^2 = 0.17$. The average sampling error variance (ASEV) in table A2 is the average error in the regional skewness estimator at the sites in the data set. The average variance of prediction at a new site (AVP_{new}) corresponds to MSE used in Bulletin 17B to describe the precision of the generalized skewness. The CONSTANT model has an AVP_{new}, equal to 0.18, which corresponds to an effective record length of 41 years.

Table A2. Regional skew model for the study area in the Pacific Northwest.

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

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

It is important to note that this regional skew model is not valid for an area in the Snake River Plain as described earlier and for urbanized basins of PNW.

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 discharges 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 A3. 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 η_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 δ 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 $n\sigma_\delta^2(k)$.

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 μ is the mean of the estimated station sample skews. Thus, the total expected sum of squares variation due to model error δ_i and sampling error $\eta_i = \hat{\gamma}_i - \gamma_i$ in expectation should equal: $n\sigma_\delta^2(0) + \sum_{i=1}^{n} Var(\hat{\gamma}_i)$.

Therefore, the expected sum of squares attributed to a regional skew model with k parameters equals $n\left[\sigma_\delta^2(0) - \sigma_\delta^2(k)\right]$, 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 A3 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 computed residual errors and the observed model predictions, while also ignoring the impact of correlation among the sampling errors.

Table A3 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 calculated as

$$EVR = \frac{\text{SS(sampling error)}}{\text{SS(model error)}} = \frac{\sum_{i=1}^{n} Var(\hat{\gamma}_i)}{n\sigma_\delta^2(k)} \quad \text{(A12)}$$

For the PNW regional skew study area, EVR had a value of 0.7 for the CONSTANT model. The sampling variability in the sample skewness estimators was larger than the error in the regional model. Thus an OLS model that neglects sampling

error in the station skewness 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 β_0^{WLS}, because the covariance among the estimated 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 employed. The MBV* is calculated as

$$MBV^* = \frac{Var\left[b_0^{WLS} \mid GLS\ analysis\right]}{Var\left[b_0^{WLS} \mid WLS\ analysis\right]} = \frac{w^T \Lambda w}{\sum_{i=1}^{n} w_i} \quad \text{(A13)}$$

where

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

Table A3.   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; $Var(\gamma_i)$, variance of the estimated sample skew at site i; EVR, error variance ratio; MBV*, misrepresentation of the beta variance; Pseudo-R_δ^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	$n\left[\sigma_\delta^2(0) - \sigma_\delta^2(k)\right]$	0
Model error	n-k-1	289	$n\left[\sigma_\delta^2(k)\right]$	50
Sampling error	n	290	$\sum_{i=1}^{n} Var(\hat{\gamma}_i)$	35
Total	$2n$-1	579	$n\left[\sigma_\delta^2(k)\right] + \sum_{i=1}^{n} Var(\hat{\gamma}_i)$	85
EVR				0.7
MBV*				10
Pseudo-R_δ^2				0%

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 skewness estimators had an effect on the precision with which the regional average skew coefficient can be estimated; if a WLS precision analysis were used for the estimated constant parameter in the CONSTANT model, the variance would be underestimated by a factor of 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 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 unusual 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 effect on the fitted (or predicted) values (Hoaglin and Welsch, 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 effect 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 and others (2011) and Veilleux (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 streamgages with high influence, in order of descending influence values, are USGS streamgage numbers:12354000, 12311000, 12358500, 12325500, 14353500, 14353000, 13336500, 14158790, 12043000, 12344000, 14091500, 13112000, 12342500, 12354500, 12039500, 12121600, 14032000, 14166500, 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.

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.

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: Cornell, 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, doi:1029/2003WR002697, 17 p.

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, Fla.: American Society of Civil Engineers, Reston, Va., 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: American Society of Civil Engineers, paper 563, p. 1–10 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.

Kendall, M.G. and Stuart, Alan, 1961, The advanced theory of statistics, Volume 2—Inference and relationship: Hafner Publishing Company, New York, 676 p.

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

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, 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., doi: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., 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., 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 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 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, nos. 1–4, p. 361–375.

Veilleux, A.G., 2009, Bayesian GLS regression for regionalization of hydrologic statistics, floods and Bulletin 17 skew: Ithaca, New York, Cornell University, Master's thesis, 155 p.

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., http://pubs.usgs.gov/fs/2013/3108/.

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, doi: 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. [Also available at http://pubs.usgs.gov/sir/2009/5158/.]

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 Investigation Report 2016–5083, 56 p., http://dx.doi.org/10.3133/sir20165083.