Prepared in cooperation with the Flood Control District of Maricopa County

Methods for Estimating Magnitude and Frequency of 1-, 3-, 7-, 15-, and 30-day flood-duration flows in Arizona

Scientific Investigations Report 2014–5109

Version 1.1, April 2015

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

FRONT COVER

Photograph of December 30, 2004, flooding on the Agua Fria River downstream from U.S. Geological Survey streamgaging station 09512500, Agua Fria River near Mayer, Arizona (U.S. Geological Survey photograph by Jeffrey Kennedy).

Methods for Estimating Magnitude and Frequency of 1-, 3-, 7-, 15-, and 30-day flood-duration flows in Arizona

By Jeffrey R. Kennedy, Nicholas V. Paretti, and Andrea G. Veilleux

Prepared in cooperation with the Flood Control District of Maricopa County

Scientific Investigations Report 2014–5109
Version 1.1, April 2015

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

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

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

U.S. Geological Survey, Reston, Virginia: 2015
First release: 2014
Revised: April 2015 (ver. 1.1)

For product and ordering information: World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth,
its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Suggested citation:
Kennedy, J.R., Paretti, N.V., and Veilleux, A.G., 2015, Methods for estimating magnitude and frequency of 1-, 3-, 7-, 15-, and 30-day flood-duration flows in Arizona (ver. 1.1, April 2015): U.S. Geological Survey Scientific Investigations Report 2014–5109, 35 p., *http://dx.doi.org/10.3133/sir20145109*.

ISSN 2328-0328 (online)

Contents

Figures

Tables

Appendixes

[Available online only at http://pubs.usgs.gov/of/2014/5109]

1. Final flood-duration flow frequency estimates for station, regression, and weighted computation for streamgages in Arizona and western New Mexico.
2. Variance estimates for station, regression, and weighted flood-duration flow frequency statistics for streamgages in Arizona and western New Mexico.
3. Plots showing leverage and influence statistics for stations used in regional regression equations (annual exceedance possibilities: 0.2=5 year, 0.1=10 year, 0.04=25 year, 0.02=50 year, 0.01=100 year, 0.005=200 year, 0.002=500 year).

Acronyms and Abbreviations

AEP	annual exceedance probability
AVP	average variance of prediction
B-GLS	Bayesian generalized least squares
DRNAREA	drainage area/contributing area of the watershed
ELEV	average basin elevation
EMA	expected moments algorithm
AVP	average variance of prediction
GB	Grubbs-Beck
GIS	geographic information system
GLS	generalized least squares
IACWD	Interagency Advisory Committee on Water Data
LP3	log-Pearson type III
MGB	multiple Grubbs-Beck
OLS	ordinary least squares
PRISM	parameter-elevation regressions on independent slopes model
$SE_{p,ave}$	average standard error of prediction (also SEP)
STATSGO	State Soil Geographic (soil data)
USGS	U.S. Geological Survey
VIF	variance inflation factor
VP_r	variance of prediction, regression estimate
VP_s	variance of prediction, station estimate
VP_w	variance of prediction, weighted estimate
WLS	weighted least squares
WREG	USGS linear regression software

This page left intentionally blank.

Methods for Estimating Magnitude and Frequency of 1-, 3-, 7-, 15-, and 30-day flood-duration flows in Arizona

By Jeffrey R. Kennedy, Nicholas V. Paretti, and Andrea G. Veilleux

Abstract

Large floods have historically caused extensive damage in Arizona. Although peak-flow frequency estimates are required for managing the risk posed by floods, estimates of the frequency of sustained flood flow (flood-duration flow) are also useful for planning and assessing the adequacy of retention and conveyance structures and for water-resource planning. This report presents a flood-duration flow frequency analysis for selected durations (1 day, 3 day, 7 day, 15 day, and 30 day) at 173 streamgaging stations throughout Arizona and in western New Mexico. For each *n*-day duration, a log-Pearson type III distribution was fitted to the annual series of *n*-day flood-duration flows using the expected moments algorithm with a multiple Grubbs-Beck low-outlier test. Regional skews were developed independently for each *n*-day duration using a hybrid weighted least squares/generalized least squares method. No basin characteristics were found to adequately explain variation in skew among stations and a constant statewide skew model was used for all *n*-day durations. The regional skewness coefficient is negative for all *n*-day durations and becomes increasingly negative for longer *n*-day durations. Uncertainty associated with the skewness coefficient is estimated using a Bayesian generalized least squares technique.

Regression equations, which allow predictions of *n*-day flood-duration flows for selected annual exceedance probabilities at ungaged sites, were developed using generalized least-squares regression and flood-duration flow frequency estimates at 56 streamgaging stations within a single, relatively uniform physiographic region in the central part of Arizona, between the Colorado Plateau and Basin and Range Province, called the Transition Zone. Drainage area explained most of the variation in the *n*-day flood-duration annual exceedance probabilities, but mean annual precipitation and mean elevation were also significant variables in the regression models. Standard error of prediction for the regression equations varies from 28 to 53 percent and generally decreases with increasing *n*-day duration. Outside the Transition Zone there are insufficient streamgaging stations to develop regression equations, but flood-duration flow frequency estimates are presented at select streamgaging stations.

Introduction

Flood-frequency analyses are a common tool for assessing flood-hazard risk. Such analyses typically focus on the frequency of maximum instantaneous flow (peak flow), and use statistical methods to predict the annual peak flow for a specified probability, known as the annual exceedance probability (AEP). The estimated peak flow of large floods, with low AEP, are widely used to delineate flood-plain boundaries and predict potential property damage but also for designing structures designed to convey runoff at a sufficient rate, such as bridges, channels, and culverts. However, for detention and retention type structures, estimates of the frequency of a volume of flood flow over some duration of time (flood-duration flow) are also needed. Furthermore, flood-duration frequency estimates can be used for water-resources planning and management, particularly on river systems with water-storage reservoirs.

The log-Pearson type III (LP3) distribution has been adopted as the standard flood-frequency model throughout the United States. Methods for fitting the moments (mean, standard deviation, and skew) of the LP3 distribution are described in a report published by the Interagency Advisory Committee on Water Data (IACWD) and widely known as “Bulletin 17B” (IACWD, 1982). Although Bulletin 17B procedures are typically used for estimating the AEPs of flood peaks, the same procedures can also be used for flood-duration flows. Since publication, several improvements to Bulletin 17B have been suggested concerning the treatment of low-outlier, historical, and other censored flood information (Stedinger and Griffis, 2008). The expected moments algorithm (EMA), used with the multiple Grubbs-Beck (MGB) test, is a revision to the traditional Bulletin 17B estimation methods that explicitly addresses censored data (Cohn and others, 1997; Cohn and others, 2001; England and others, 2003). Of particular note for Arizona and other semiarid regions with large variability in annual maximum floods, the MGB test efficiently accounts for multiple potentially influential low-flows, which may otherwise have undue influence on the estimated magnitude of large, low-probability floods. An evaluation of the implications for replacing Bulletin 17B methods with EMA/MGB methods for Arizona streamgaging stations (Paretti and others, 2014a) found that although predicted peak flows using EMA/MGB were neither consistently larger nor smaller than Bulletin 17B predictions, goodness-of-fit criteria indicated EMA/MGB provided a better representation of the peak-flow data. Therefore, EMA/MBG methods are used to implement the flood-duration flow analysis in this report.

StreamStats is a national U.S. Geological Survey (USGS) map-based Web application that provides easy access to published flood-frequency and basin-characteristic statistics for user-selected watersheds. This interactive Web application allows the user to select a point on a stream channel (gaged or ungaged), delineate a watershed boundary, and retrieve flood-frequency estimates derived from the current regional regression equations and geographic information system (GIS) data within the basin selected. StreamStats provides consistent statistics, minimizes user error, and reduces the need for large datasets and costly standalone GIS software. Peak-flow frequency estimates (Paretti and others, 2014b) and *n*-day flood-duration flow frequency estimates (this report) are available online in the StreamStats Web application at http://streamstatsags.cr.usgs.gov/az_ss.

Physical Setting

Streamgaging stations used in the flood-duration flow frequency analysis are located throughout Arizona but are primarily concentrated in a region in the central part of the State, between the Basin and Range Province to the southwest and the Colorado Plateau to the northeast, called the Transition Zone (fig. 1). The Transition Zone region is characterized by high relief with small, relatively shallow aquifers. Land-surface elevations in this region range from about 2,000 feet near the confluence of the Salt and Verde Rivers to about 11,400 feet at the headwaters of the Salt River in eastern Arizona (fig. 2*A*). Most major Arizona streams and rivers, with the exception of the Colorado River, have their headwaters in this region, including the Gila, Salt, Verde, and Hassayampa Rivers. Smaller drainages in this region are mostly intermittent or ephemeral. Precipitation and air temperature are highly variable throughout the Transition Zone and are correlated with land-surface elevation; higher elevations experience lower average temperatures and greater precipitation amounts that do lower elevations. Average annual precipitation in the region ranges from 39 inches per year near the headwaters of the Salt River in the White Mountains to less than 10 inches per year in the lower deserts (fig. 2*B*).

Both the Basin and Range Province to the southwest of the Transition Zone and the Colorado Plateau to the northeast have very little perennial surface water, and streamflow, even in large drainages, often occurs only in response to discrete precipitation events. Notable exceptions are certain reaches of the Santa Cruz and San Pedro Rivers in the Basin and Range and the upper reaches of the Little Colorado River and its tributaries on the Colorado Plateau. Land-surface elevations in the Basin and Range vary from 100 feet along the Lower Colorado River, to a few thousand feet on basin floors, to more than 10,000 feet in some mountain ranges (fig. 2*A*). Less rainfall and higher temperatures are characteristic of the Basin and Range lowlands as compared to the Transition Zone. Mean annual precipitation in this region ranges from less than 4 inches per year in southwest Arizona to greater than 30 inches per year at high elevations toward the southeast corner of the State (fig. 2*B*). The Colorado Plateau covers roughly 45,000 square miles in northeast Arizona and is characterized by low relief punctuated by numerous canyon drainages, the most notable being the Grand Canyon of the Colorado River. The average elevation on the plateau is about 5,000 feet and average rainfall is about 10 inches per year (fig. 2 *A* and *B*). At higher altitudes on the plateau, annual peak flows can be influenced by snowmelt, but no streamgaging stations where flows are dominated by snowmelt are included in this flood-duration flow analysis. Streams are generally spring fed and, as with the Basin and Range, typically have high transmission losses, and streamflow quickly infiltrates downstream.

Purpose and Scope

The primary purposes of this report are to (1) present an application of newly developed flood-frequency methods, namely the expected moments algorithm and multiple Grubbs-Beck low-outlier test, and a hybrid Bayesian weighted least-squares/generalized least-squares method for estimating regional skewness coefficients and uncertainty; (2) present estimates of the annual maximum 1-, 3-, 7-, 15-, and 30-day flood-duration flows for the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities at 173 streamgaging stations in Arizona with 10 or more years of record; and (3) present regional regression equations for estimating the annual maximum 1-, 3-, 7-, 15-, and 30-day flood-duration flows for the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities at ungaged basins in the central part of the State.

Data Development

Flood-duration flow can be defined as the average mean daily flow over a specified duration, often referred to as the annual maximum *n*-day flood flow. Durations considered in this report are the 1-, 3-, 7-, 15-, and 30-day flood-duration flows. The 1-day flood-duration flow is simply the highest annual mean daily flow and most often occurs during the same event and on the same day as the annual instantaneous peak flow. The longer *n*-day intervals are determined as the period of *n* consecutive days with the highest average flow in a given water year. As the duration interval length increases, the probability that it encompasses the annual instantaneous peak flow decreases (fig. 3). At some stations, there may not be 15 or 30 days of continuous flow during the year, and high flows for these durations can include a period of zero flow.

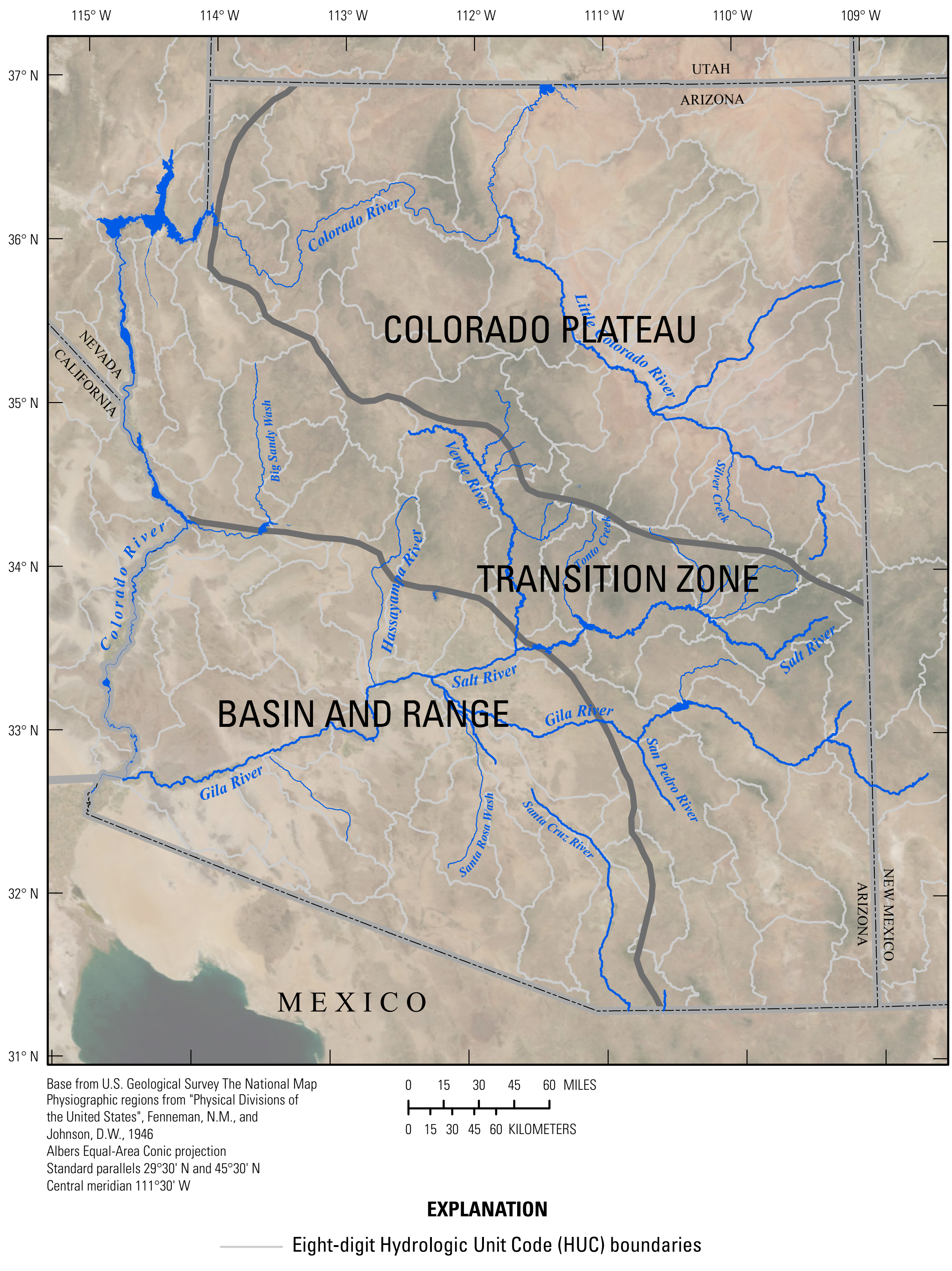


Figure 1. Map of Arizona showing major physiographic regions.

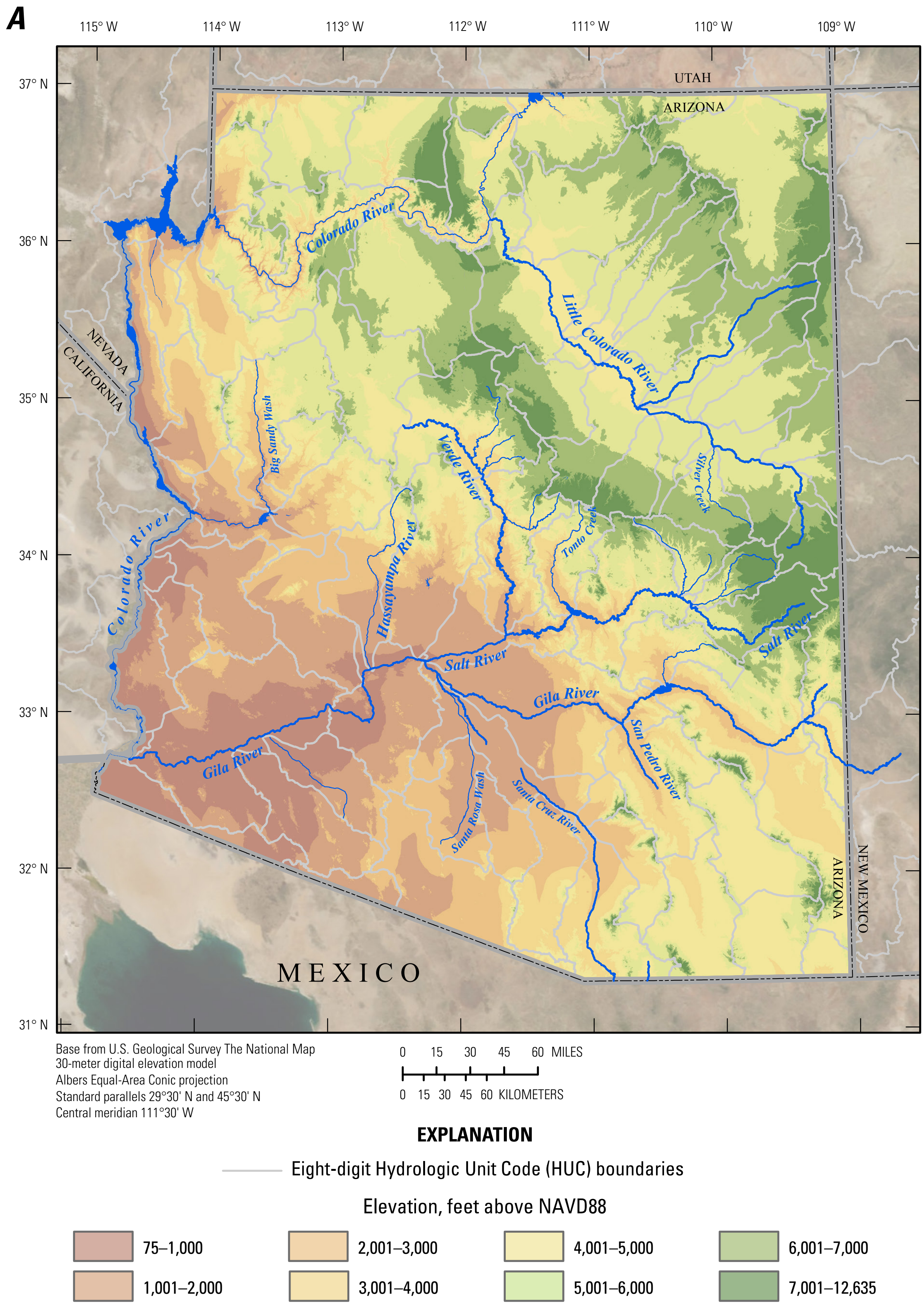


Figure 2. Maps of (*A*) elevation and (*B*) mean annual precipitation in Arizona.

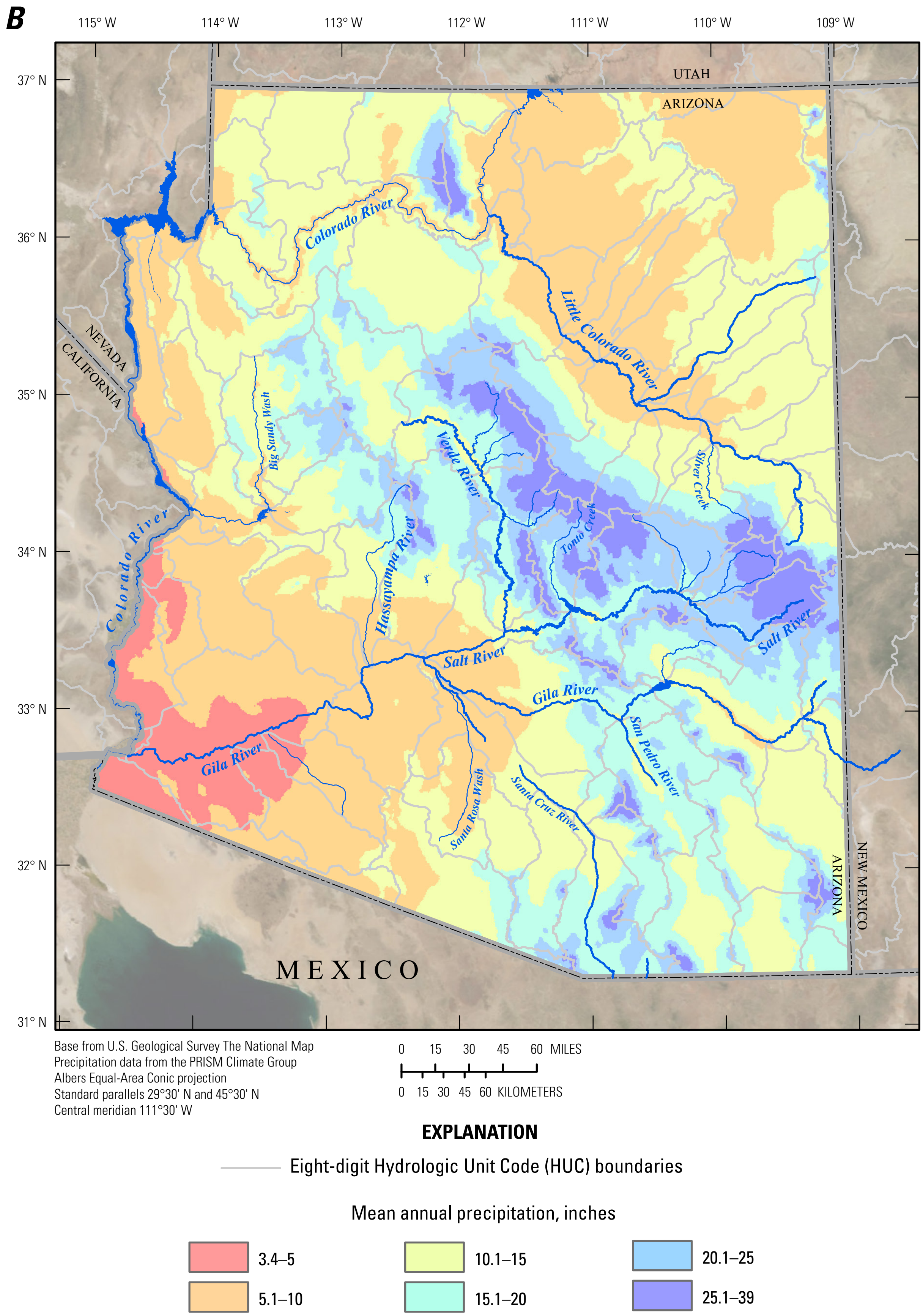


Figure 2.—Continued

Flood-duration flows are reported in dimensions of volume per time and units of cubic feet per second; to convert to total volume the flow rate is simply multiplied by the length of the duration interval considered.

Site Selection

Determination of *n*-day flood-duration flow requires continuous records of mean daily flow. Some gages operated by the USGS in Arizona are crest-stage gages, which record only the maximum stage between site visits; therefore, there are fewer gages available for flood-volume analysis than for flood-peak analysis. Annual maximum flood-duration flow data for 1-, 3-, 7-, 15-, and 30-day duration intervals were retrieved using SWSTAT software from the USGS National Water Information System for an initial dataset of 198 streamgaging stations with record lengths 10 years or longer. Stations are located primarily in central and southeastern Arizona, with significant gaps in northeast, northwest, and southwest Arizona (fig. 4). Stations with *n*-day flood-duration flows that were poorly represented by the LP3 distribution were excluded, as were stations significantly affected by impoundments, diversions, or urbanization. Therefore, no flood-duration flow frequency estimates are given for the Colorado River, Verde River below Horseshoe Dam, Salt River below Roosevelt Dam, and Gila River below Coolidge Dam. After removing unsuitable stations, 173 remaining stations were used in the analysis (table 1), of three types:

- Stations with greater than 20 years of record, well approximated by the LP3 distribution, used to determine the regional skewness coefficient of the LP3 distribution. Redundant stations (stations that are near another station with similar basin characteristics) were removed as described below in the section Regional Skew Analysis and Cross-Correlation Models.
- Stations with between 10 and 20 years of record, located in the Transition Zone, and well approximated by the LP3 distribution. These stations, combined with stations used in the regional skew analysis that are in the Transition Zone, were used to generate the regional regression equations.
- Stations with more than 10 years of record not used in the regional skew or regional regression analyses.

The Mann-Kendall trend test was used to test for trends in *n*-day flood-duration flows at stations with 30 or more years of record (Helsel and Hirsch, 2002). The null hypothesis (H_0, no trend in streamflow) was rejected at the 5-percent significance level at four streamgaging stations (table 2). At three stations, H_0 was rejected for all durations; at the fourth station, H_0 was rejected only for the 1-, 3-, and 7-day flood-duration flows. The trend in flood-duration flow was downward at all of these stations, and no stations had increasing trends in flood-duration flow. Despite the apparent trend at these stations, flood-duration flow frequency results are presented, and three of the four were used in the regional skew analysis because they have long records and represent watersheds in geographic and (or) physiographic regions where there are no alternative stations. None of the stations where H_0 was rejected were used in the regional regression equations. Possible reasons for downward trends at these stations include changes in watershed characteristics, such as vegetation and channel

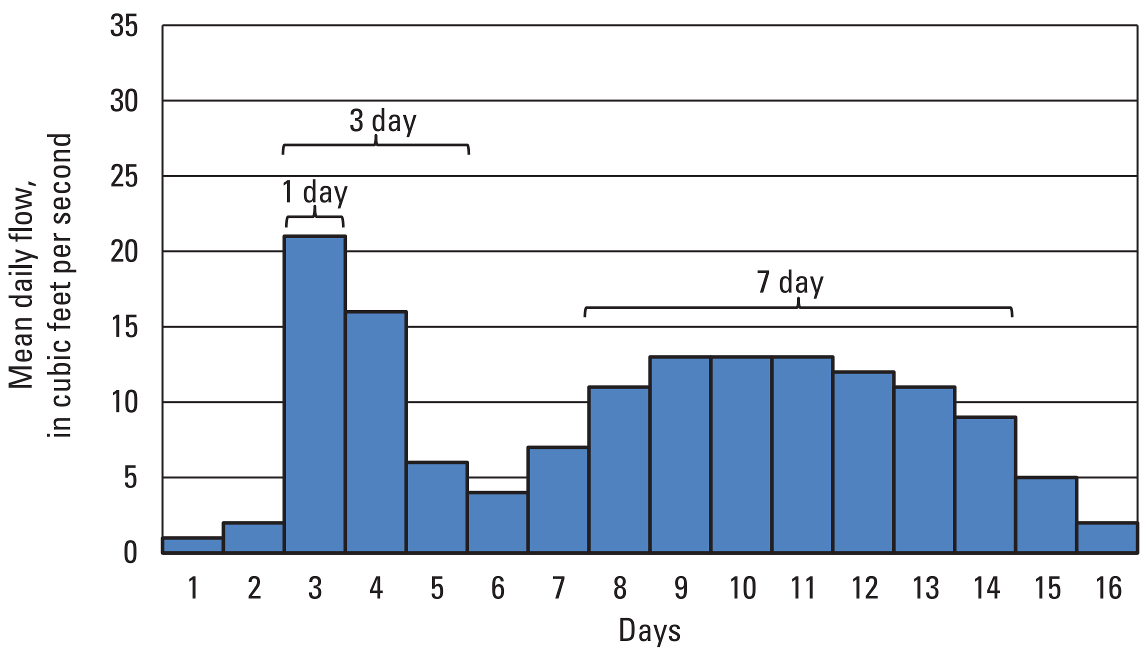


Figure 3. Graph of example data used to calculate 1-day, 3-day, and 7-day flood-duration flows and their relation to annual maximum peak flow.

morphology, human activities, and decreasing seasonal rainfall (Thomas and Pool, 2006; Kennedy and Gungle, 2010).

Basin Characteristics

As part of the larger Arizona StreamStats project (Ries and others, 2008; Paretti and others, 2014b), watershed boundaries for each streamgaging station were calculated using the Watershed Boundary Dataset, the 1:24,000 National Hydrography Dataset, and the 1/3 arc-second (10-meter) National Elevation Dataset. Within each watershed boundary, several characteristics were computed using the best available data (table 3). Elevation is calculated as the mean elevation throughout the watershed area. Precipitation metrics were identified using parameter-elevation regressions on independent slopes model (PRISM) monthly data (PRISM Climate Group, 2012). Two soil characteristics, permeability and available water capacity, were identified using the State Soil Geographic Database (STATSGO; Natural Resources Conservation Service, 2013). Full details of the basin characteristic identification process are in Paretti and others (2014b).

Flood-Volume Frequency Analysis

Several approaches to *n*-day flood-duration flow analysis have been presented in the hydrologic literature. The following discussion considers statistical approaches. Alternatively, rainfall-runoff models can be constructed to predict flood flow and duration as a function of the probability of a given rainfall event (for example, Bohman, 1990; Sherwood, 1994), but these are more suitable for small watersheds and do not fully use the historical runoff data collected by the USGS.

Statistical approaches to flood-duration flow frequency analysis, in which measured runoff volumes are used directly to estimate the probability of a given duration of flood flow, generally fall into one of three categories. First, individual distributions, such as the log-Pearson type III (LP3) distribution commonly used in analyzing instantaneous peaks (IACWD, 1982), can be fitted to the *n*-day flood-duration flows. The USGS Manual of Hydrology, discussing graphical methods of fitting curves to data on a probability plot states simply: “the frequency of flood volume can be determined by the same method as the frequency of flood peaks” (Dalrymple, 1960). The Bulletin 17B manual also recognizes “the same techniques could also be used to treat . . . flood volumes” but states such applications were not evaluated (IACWD, 1982). Devulapalli and Valdes (1996) used the LP3 distribution successfully to model 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, and 10-day volumes for rural watersheds in Texas. Sherwood (1994) modeled small urban watershed flood volumes with the LP3 distribution, but flood intervals were much shorter (1 to 32 hours) than those considered in the present study. Two approaches to distribution-fitting can be taken—either (1) unique distribution moments can be defined for each *n*-day interval or (2) an average distribution can be fitted to all of the data and a scaling parameter identified that controls the spacing between the different *n*-day intervals on a quantile-probability plot (Javelle and others, 2003; Cunderlik and Ouarda, 2006). The latter approach was investigated for streamgaging stations in Arizona, but large variation existed in the quantile-probability plots among different *n*-day durations, and it was found to be unsuitable.

A second approach to high-flow frequency analysis is the joint probability distribution approach, which treats flood peak, volume, and duration as random variables to be predicted concurrently (Yue and Rasmussen, 2002; Mediero and others, 2010). This method recognizes that flood peaks and volume are not independent; rather, for a flood peak with a given probability there may be a range of flood volumes that occur with varying probability. Finally, a third approach develops regression equations that relate flood volume to flood peak flow (Eychaner, 1976; Perry, 1984; Singh and Hossein, 1986). The instantaneous flood peak at a given probability is estimated by fitting a statistical distribution, and the corresponding flood volume at the same probability is determined from the regression. In effect, this approach is the opposite of the joint probability approach; it assumes a unique relation between a flood peak of a given probability and a corresponding flood volume. Both the joint probability and the peak-volume regression approaches require datasets that contain both the flood peak and corresponding flood volume for each year. Although the annual instantaneous flood peak often corresponds to the same event as the 1-day and 3-day flood volumes, at longer intervals this is often not the case (Balocki and Burges, 1994). Determining correspondence between flood peaks and volumes at these longer intervals is a significant task; therefore, the fitted-distribution approach is taken in the present study.

The LP3 distribution is defined by the first, second, and third moments (the mean, standard deviation, and skew, denoted by μ, σ, and γ, respectively). On a log-probability plot of annual flood peaks (fig. 5), the distribution can be represented as either a straight line (skew = 0) or one that curves. The distribution mean determines the position of the line along the y-axis and the standard deviation determines the slope of the line. The basic equation for determining flood frequency from the three moments is:

$$\log Q_p = \overline{X} + K_p S \, , \qquad (1)$$

where

- Q_p is the annual-peak flow (in this case, *n*-day flood-duration flow) for the exceedance probability, *P*,
- $\overline{X}$ is the mean of the logarithms of the annual-peak flow,
- K_p is a factor based on the weighted skew coefficient and the exceedance probability, *P*, which can be obtained from appendix 3 of Bulletin 17B (IACWD, 1982), and
- S is the standard deviation of the logarithms of the annual-peak flow, which is a measure of the degree of variation in the annual values about the mean value.

Table 1. Table of streamgaging stations in Arizona and western New Mexico with flood-duration flow frequency statistics.

[AZ, Arizona; NM, New Mexico; Y, yes; N, no]

Map ID	Station ID	Station Name	Period of Record	Number of years of record	Drainage area (square miles)	Mean annual precipitation (inches)	Mean elevation (feet above NAVD88)	Used in skew analysis	Used in regional regression
1	09379200	Chinle Creek near Mexican Water, AZ	1965–2010	46	3,612	10.1	6,244	Y	N
2[1]	09382000	Paria River at Lees Ferry, AZ	1924–2010	87	1,362	11.9	6,138	Y	N
3	09383400	Little Colorado River at Greer, AZ	1961–2010	24	28.9	32.8	9,437	Y	N
4	09383500	Nutrioso Creek above Nelson Res near Springerville, AZ	1968–2010	17	83.4	23.4	8,532	N	N
5	09384000	Little Colorado River above Lyman Lake near St. Johns, AZ	1941–2010	70	711	18.4	7,833	Y	N
6	09386250	Carrizo Wash near St. Johns, AZ	1999–2010	12	2,143	13.3	7,124	N	N
7[1]	09386950	Zuni River above Black Rock Reservoir, NM	1970–2010	41	829	15.7	7,363	N	N
8	09390500	Show Low Creek near Lakeside, AZ	1954–2010	57	68.0	27.6	7,290	Y	N
9	09392500	Show Low Creek at Show Low, AZ	1945–1954	10	92.6	25.9	7,107	N	N
10	09393500	Silver Creek near Snowflake, AZ	1951–1995	45	840	18.0	6,354	Y	N
11	09394500	Little Colorado River at Woodruff, AZ	1906–2010	80	7,652	14.4	7,039	Y	N
12	09395900	Black Creek near Lupton, AZ	1965–1982	15	494	14.5	7,407	N	N
13	09397000	Little Colorado River at Holbrook, AZ	1906–2010	31	10,836	13.9	6,797	N	N
14	09397500	Chevelon Fork below Wildcat Canyon near Winslow, AZ	1948–2010	38	272	24.8	7,070	N	N
15	09398000	Chevelon Creek near Winslow, AZ	1917–2006	45	759	19.5	6,554	Y	N
16	09398500	Clear Creek below Willow Creek near Winslow, AZ	1948–1991	44	318	27.0	7,172	Y	N
17	09399000	Clear Creek near Winslow, AZ	1930–2006	52	607	20.8	6,560	Y	N
18	09400562	Oraibi Wash near Tolani Lake, AZ	1996–2010	15	665	10.4	6,278	N	N
19	09400568	Polacca Wash near Second Mesa, AZ	1995–2010	16	912	10.1	6,408	N	N
20	09400583	Jeddito Wash near Jeddito, AZ	1994–2005	12	148	10.8	6,371	N	N
21	09401000	Little Colorado River at Grand Falls, AZ	1927–1994	35	20,003	13.4	5,885	N	N
22	09401110	Dinnebito Wash near Sand Springs, AZ	1994–2010	17	478	10.4	6,300	N	N
23	09401260	Moenkopi Wash at Moenkopi, AZ	1977–2010	34	1,231	9.7	6,115	N	N
24	09401280	Moenkopi Wash near Tuba, AZ	1927–1940	14	1,393	9.5	6,035	N	N
25	09401400	Moenkopi Wash near Tuba City, AZ	1941–1978	26	1,731	9.1	5,892	Y	N
26	09401500	Moenkopi Wash near Cameron, AZ	1954–1964	11	1,881	8.9	5,807	N	N
27	09402000	Little Colorado River near Cameron, AZ	1948–2010	63	24,350	12.7	6,170	Y	N
28	09403000	Bright Angel Creek near Grand Canyon, AZ	1924–1973	50	101	22.1	7,312	Y	N

Table 1. Table of streamgaging stations in Arizona and western New Mexico with flood-duration flow frequency statistics.—Continued

Map ID	Station ID	Station Name	Period of Record	Number of years of record	Drainage area (square miles)	Mean annual precipitation (inches)	Mean elevation (feet above NAVD88)	Used in skew analysis	Used in regional regression
29	09403780	Kanab Creek near Fredonia, AZ	1964–1980	17	1,124	15.0	6,000	N	N
30	09404110	Havasu Creek at Supai, AZ	1996–2010	15	2,428	14.3	6,054	N	N
31	09404208	Diamond Creek near Peach Springs, AZ	1994–2010	17	276	12.0	4,925	N	N
32	09404222	Spencer Creek near Peach Springs, AZ	1999–2010	12	257	12.8	4,781	N	N
33	09404343	Truxton Wash near Valentine, AZ	1994–2010	17	375	14.9	5,112	N	N
34	09415000	Virgin River at Littlefield, AZ	1930–2000	58	4,858	14.4	5,170	Y	N
35	09424200	Cottonwood Wash No. 1 near Kingman, AZ	1965–1978	14	135	19.3	5,363	N	Y
36	09424447	Burro Creek at Old US 93 Bridge near Bagdad, AZ	1981–2010	16	611	18.4	4,658	N	N
37	09424450	Big Sandy River near Wikieup, AZ	1967–2010	44	2,562	15.7	4,326	Y	Y
38	09424900	Santa Maria River near Bagdad, AZ	1967–2010	41	1,130	16.9	3,992	Y	Y
39	09425500	Santa Maria River near Alamo, AZ	1941–1965	25	1,433	15.9	3,725	N	N
40	09426500	Bill Williams River at Planet, AZ	1915–1946	19	5,307	14.9	3,718	N	N
41	09430500	Gila River near Gila, NM	1929–2010	82	1,856	20.4	7,451	Y	Y
42	09431500	Gila River near Redrock, NM	1931–2010	73	2,828	19.8	6,896	N	N
43	09442000	Gila River near Clifton, AZ	1912–2010	78	4,007	18.1	6,227	Y	Y
44	09442680	San Francisco River near Reserve, NM	1961–2010	50	333	21.0	7,800	Y	Y
45	09444000	San Francisco River near Glenwood, NM	1928–2010	83	1,653	20.8	7,231	N	N
46	09444200	Blue River near Clifton, AZ	1969–2010	38	505	23.1	6,852	Y	Y
47	09444500	San Francisco River at Clifton, AZ	1914–2010	84	2,765	20.9	6,811	Y	Y
48	09445500	Willow Creek near Point Of Pines near Morenci, AZ	1945–1967	23	107	21.0	6,295	N	N
49	09446000	Willow Creek near Double Circle Ranch near Morenci, AZ	1945–1967	23	155	21.0	6,239	Y	Y
50	09446500	Eagle Creek near Double Circle Ranch near Morenci, AZ	1945–1967	23	383	21.3	6,281	N	N
51	09447000	Eagle Creek above Pumping Plant near Morenci, AZ	1945–2010	66	621	20.5	6,009	Y	Y
52	09447800	Bonita Creek near Morenci, AZ	1982–2010	29	302	17.4	5,247	Y	Y
53	09448500	Gila River at Head of Safford Valley near Solomon, AZ	1921–2010	88	7,888	19.2	6,329	Y	Y
54	09456000	San Simon River near San Simon, AZ	1920–1940	13	823	16.2	4,881	N	N
55	09457000	San Simon River near Solomon, AZ	1932–1982	48	2,243	14.1	4,334	Y	N
56	09458200	Deadman Creek near Safford, AZ	1968–1993	15	4.7	29.1	7,361	N	N
57	09458500	Gila River at Safford, AZ	1941–1965	15	10,483.0	17.9	5,828	N	N
58	09460150	Frye Creek near Thatcher, AZ	1968–2010	30	4.0	33.6	8,127	Y	Y

Table 1. Table of streamgaging stations in Arizona and western New Mexico with flood-duration flow frequency statistics.—Continued

Map ID	Station ID	Station Name	Period of Record	Number of years of record	Drainage area (square miles)	Mean annual precipitation (inches)	Mean elevation (feet above NAVD88)	Used in skew analysis	Used in regional regression
59	09468500	San Carlos River near Peridot, AZ	1930–2010	81	1,026.0	19.1	4,443	Y	Y
60	09470500	San Pedro River at Palominas, AZ	1931–2010	54	738.0	19.2	5,033	N	N
61	09470750	Ramsey Canyon near Sierra Vista, AZ	2001–2010	10	4.2	28.5	7,325	N	N
62	09470800	Garden Canyon near Fort Huachuca, AZ	1960–2010	21	8.6	25.8	6,707	Y	N
63[1]	09471000	San Pedro River at Charleston, AZ	1905–2010	95	1,216	18.1	4,938	Y	N
64	09471310	Huachuca Canyon near Fort Huachuca, AZ	2001–2010	10	4.1	26.4	6,811	N	N
65	09471380	Upper Babocomari River near Huachuca City, AZ	2001–2010	10	156	18.6	5,138	N	N
66	09471400	Babocomari River near Tombstone, AZ	2001–2010	10	303	17.8	5,005	N	N
67	09471550	San Pedro River near Tombstone, AZ	1968–2010	33	1,729	17.7	4,898	N	N
68	09471800	San Pedro River near Benson, AZ	1967–2010	14	2,487	17.1	4,746	N	N
69	09472000	San Pedro River near Redington, AZ	1944–1997	50	2,925	17.1	4,681	Y	N
70	09472050	San Pedro River at Redington Bridge near Redington, AZ	1999–2010	12	2,925	17.1	4,681	N	N
71	09473000	Aravaipa Creek near Mammoth, AZ	1932–2010	54	538	18.6	4,572	Y	Y
72	09473500	San Pedro River at Winkelman, AZ	1967–1978	12	4,451	17.4	4,444	N	N
73	09480000	Santa Cruz River near Lochiel, AZ	1950–2010	61	82.0	19.7	5,093	Y	N
74	09480500	Santa Cruz River near Nogales, AZ	1914–2010	82	532	19.8	4,891	Y	N
75	09481500	Sonoita Creek near Patagonia, AZ	1931–1972	40	209	21.2	4,919	Y	N
76	09481740	Santa Cruz River at Tubac, AZ	1996–2010	14	1,213	20.1	4,617	N	N
77	09482000	Santa Cruz River at Continental, AZ	1941–2010	58	1,673	19.6	4,391	Y[2]	N
78	09482400	Airport Wash at Tucson, AZ	1966–1981	16	29.6	13.7	2,848	N	N
79	9482500	Santa Cruz River at Tucson, AZ	1906–2010	84	2,192	18.5	4,095	Y	N
80	09483000	Tucson Arroyo at Vine Ave at Tucson, AZ	1945–1981	37	7.6	12.4	2,516	N	N
81	09483010	High School Wash at Tucson, AZ	1974–1983	10	1.0	12.2	2,464	N	N
82	09483100	Tanque Verde Creek near Tucson, AZ	1960–1974	15	43.1	21.3	4,858	N	N
83	09484000	Sabino Creek near Tucson, AZ	1933–2010	63	35.2	30.2	6,077	Y	N
84	09484200	Bear Creek near Tucson, AZ	1960–1974	15	16.9	27.9	5,781	N	N
85	09484500	Tanque Verde Creek at Tucson, AZ	1941–2010	25	220.0	21.1	4,372	Y	N
86	09484600	Pantano Wash near Vail, AZ	1960–2010	36	456.0	19.1	4,618	Y	N
87	09485000	Rincon Creek near Tucson, AZ	1953–2010	43	44.7	21.4	5,104	Y	N
88	09485450	Pantano Wash at Broadway Blvd. at Tucson, AZ	1989–2010	21	598	18.7	4,434	Y[2]	N

Table 1. Table of streamgaging stations in Arizona and western New Mexico with flood-duration flow frequency statistics.—Continued

Map ID	Station ID	Station Name	Period of Record	Number of years of record	Drain-age area (square miles)	Mean annual precipitation (inches)	Mean eleva-tion (feet above NAVD88)	Used in skew analysis	Used in regional regression
89	09485700	Rillito Creek at Dodge Boulevard at Tucson, AZ	1991–2010	19	868	19.2	4,348	N	N
90	09486055	Rillito Creek at La Cholla Blvd near Tucson, AZ	1996–2010	15	912	19.0	4,283	N	N
91	09486300	Canada Del Oro near Tucson, AZ	1966–1978	13	250	19.0	3,926	N	N
92	09486350	Canada Del Oro below Ina Road near Tucson, AZ	1996–2010	15	255	18.6	3,894	N	N
93	09486500	Santa Cruz River at Cortaro, AZ	1940–2010	59	3,461	18.5	4,084	N	N
94	09486800	Altar Wash near Three Points, AZ	1967–2010	27	466	18.6	3,741	Y[2]	N
95	09487000	Brawley Wash near Three Points, AZ	1993–2010	18	785	17.5	3,622	N	N
96	09488500	Santa Rosa Wash near Vaiva Vo, AZ	1955–1980	26	1,734	11.1	2,217	Y	N
97	09489000	Santa Cruz River near Laveen, AZ	1941–2010	68	8,568	14.5	3,019	N	N
98	09489070	North Fork of East Fork Black River near Alpine, AZ	1966–1978	13	38.4	28.9	9,052	N	Y
99	09489100	Black River near Maverick, AZ	1963–1982	20	314	28.7	8,538	N	N
100	09489200	Pacheta Creek at Maverick, AZ	1958–1980	23	16.3	32.9	8,604	N	N
101	09489500	Black River below Pumping Plant near Point of Pines, AZ	1954–2010	57	556	27.7	8,058	Y	Y
102	09489700	Big Bonito Creek near Fort Apache, AZ	1958–1980	23	114	31.3	8,077	Y	Y
103	09490500	Black River near Fort Apache, AZ	1915–2010	54	1,224	25.5	7,222	N	N
104	09490800	North Fork White River near Greer, AZ	1966–1978	13	40	36.4	9,520	N	Y
105	09492400	East Fork White River near Fort Apache, AZ	1958–2010	53	47	35.5	8,425	Y	Y
106	09494000	White River near Fort Apache, AZ	1958–2010	53	628	29.1	7,241	Y	Y
107	09496000	Corduroy Creek near Mouth near Show Low, AZ	1952–2005	26	206	22.3	6,372	N	N
108	09496500	Carrizo Creek near Show Low, AZ	1952–2010	50	441	22.2	6,329	Y	Y
109	09496600	Cibecue 1 Trib. Carrizo Creek near Show Low, AZ	1959–1971	13	0.1	21.0	5,428	N	N
110	09496700	Cibecue 2 Trib. Carrizo Cr, AZ	1959–1971	13	0.1	20.6	5,221	N	N
111	09497500	Salt River near Chrysotile, AZ	1925–2010	86	2,831	25.1	6,755	Y	Y
112	09497800	Cibecue Creek near Chrysotile, AZ	1960–2010	51	290	23.2	5,743	Y	Y
113	09497900	Cherry Creek near Young, AZ	1964–1977	14	62.2	30.2	5,992	N	N
114	09497980	Cherry Creek near Globe, AZ	1966–2010	44	200	26.8	5,543	Y	Y
115	09498400	Pinal Creek at Inspiration Dam near Globe, AZ	1981–2010	30	195	21.0	4,172	Y	Y
116	09498500	Salt River near Roosevelt, AZ	1914–2010	97	4,289	24.5	6,183	N	N
117	09498501	Pinto Creek below Haunted Canyon near Miami, AZ	1996–2010	15	36.4	24.4	4,416	N	N
118	09498502	Pinto Creek near Miami, AZ	1995–2010	16	102	23.0	4,216	N	Y

Table 1. Table of streamgaging stations in Arizona and western New Mexico with flood-duration flow frequency statistics.—Continued

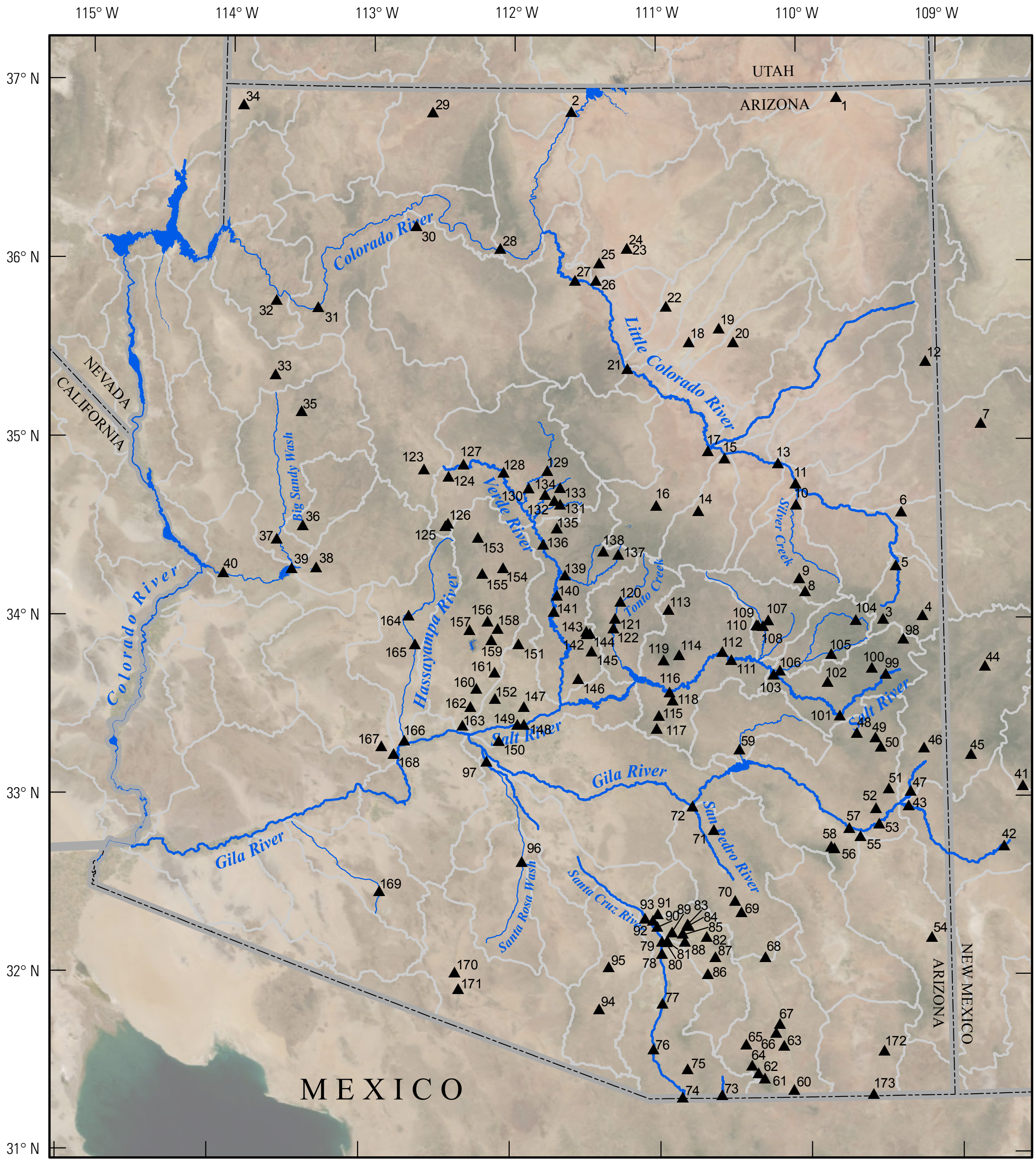
Map ID	Station ID	Station Name	Period of Record	Number of years of record	Drainage area (square miles)	Mean annual precipitation (inches)	Mean elevation (feet above NAVD88)	Used in skew analysis	Used in regional regression
119	09498503	South Fork Parker Creek near Roosevelt, AZ	1987–2010	22	1.1	33.5	6,647	Y	Y
120	09498800	Tonto Creek near Gisela, AZ	1966–1975	10	433	27.9	5,536	N	Y
121	09498870	Rye Creek near Gisela, AZ	1967–1985	19	123	22.9	4,294	N	Y
122	09499000	Tonto Creek above Gun Creek near Roosevelt, AZ	1942–2010	69	672	25.9	5,083	Y	Y
123	09502800	Williamson Valley Wash near Paulden, AZ	1966–2010	29	255	16.6	5,136	Y	Y
124	09502900	Del Rio Springs near Chino Valley, AZ	1997–2010	14	39.9	13.0	4,762	N	N
125	09502960	Granite Creek at Prescott, AZ	1996–2010	15	30.2	22.6	5,952	N	Y
126	09503000	Granite Creek near Prescott, AZ	1933–2010	31	39.4	22.4	5,906	Y	Y
127	09503700	Verde River near Paulden, AZ	1964–2010	47	2,149	16.2	5,434	Y	Y
128	09504000	Verde River near Clarkdale, AZ	1916–2010	49	3,143	17.5	5,719	N	N
129	09504420	Oak Creek near Sedona, AZ	1982–2010	29	233	27.1	6,727	N	N
130	09504500	Oak Creek near Cornville, AZ	1941–2010	67	355	24.8	6,108	Y	Y
131	09505200	Wet Beaver Creek near Rimrock, AZ	1962–2010	42	109	25.0	6,549	Y	Y
132	09505250	Red Tank Draw near Rimrock, AZ	1958–1978	21	51.0	24.3	6,065	Y	Y
133	09505300	Rattlesnake Canyon near Rimrock, AZ	1958–1980	23	25.1	25.8	6,451	Y	Y
134	09505350	Dry Beaver Creek near Rimrock, AZ	1961–2010	50	142	25.1	6,191	Y	Y
135	09505800	West Clear Creek near Camp Verde, AZ	1966–2010	45	241	26.1	6,635	Y	Y
136	09506000	Verde River near Camp Verde, AZ	1935–2010	33	4,650	18.9	5,573	N	N
137	09507600	East Verde River near Pine, AZ	1962–1971	10	6.4	31.7	6,396	N	N
138	09507700	Webber Creek above West Fork Webber Creek near Pine, AZ	1960–1974	15	4.8	33.3	7,026	N	Y
139	09507980	East Verde River near Childs, AZ	1962–2010	47	326	26.5	5,246	Y	Y
140	09508300	Wet Bottom Creek near Childs, AZ	1968–2010	43	36.3	24.3	4,918	Y	Y
141	09508500	Verde River below Tangle Creek above Horseshoe Dam, AZ	1946–2010	60	5,499	19.6	5,573	Y	Y
142	09510070	West Fork Sycamore Creek above Mcfarland Canyon near Sunflower, AZ	1966–1985	12	4.6	31.9	5,443	N	N
143	09510080	West Fork Sycamore Creek near Sunflower, AZ	1962–1974	13	9.8	31.5	5,335	N	Y
144	09510100	East Fork Sycamore Creek near Sunflower, AZ	1962–1985	24	4.5	30.5	5,228	Y	Y
145	09510150	Sycamore Creek near Sunflower, AZ	1962–1976	15	52.4	27.9	4,560	N	Y
146	09510200	Sycamore Creek near Fort McDowell, AZ	1961–2010	50	164	24.3	3,803	Y	Y
147	09512100	Indian Bend Wash at Scottsdale, AZ	1961–1984	23	59.8	10.1	1,432	Y	N

Table 1. Table of streamgaging stations in Arizona and western New Mexico with flood-duration flow frequency statistics.—Continued

Map ID	Station ID	Station Name	Period of Record	Number of years of record	Drain-age area (square miles)	Mean annual precipitation (inches)	Mean eleva-tion (feet above NAVD88)	Used in skew analysis	Used in regional regression
148	09512162	Indian Bend Wash at Curry Road Tempe, AZ	1993–2010	18	221	11.6	1,733	N	N
149	09512165	Salt River at Priest Drive near Phoenix, AZ	1995–2010	16	13,285	21.3	5,223	N	N
150	09512200	Salt River Trib. in South Mountain Park Phoenix, AZ	1961–1996	36	1.7	8.8	1,801	N	N
151	09512280	Cave Creek below Cottonwood Creek near Cave Creek, AZ	1981–2010	30	72.8	20.0	3,766	Y	Y
152	09512400	Cave Creek at Phoenix, AZ	1958–1991	34	229	15.1	2,644	N	N
153	09512450	Agua Fria River near Humboldt, AZ	2001–2010	10	175	18.8	5,425	N	N
154	09512500	Agua Fria River near Mayer, AZ	1941–2010	70	585	19.2	4,938	Y	Y
155	09512600	Turkey Creek near Cleator, AZ	1980–1992	13	89.3	21.8	5,267	N	Y
156	09512800	Agua Fria River near Rock Springs, AZ	1971–2010	39	1,111	19.4	4,528	N	N
157	09512860	Humbug Creek near Castle Hot Springs	1984–1994	11	59.9	22.2	3,964	N	Y
158	09513780	New River near Rock Springs, AZ	1966–2010	44	68.4	20.8	3,967	Y	Y
159	09513800	New River at New River, AZ	1961–1982	22	84.7	19.6	3,642	N	N
160	09513835	New River at Bell Road near Peoria	1968–1993	20	186	15.4	2,604	Y	Y
161	09513860	Skunk Creek near Phoenix, AZ	1968–2010	43	65.0	13.9	2,241	Y	Y
162	09513910	New River near Glendale, AZ	1965–1998	14	623	13.8	2,293	N	N
163	09513970	Agua Fria River at Avondale, AZ	1968–1982	14	2,403	16.5	3,309	N	N
164	09515500	Hassayampa River at Box Damsite near Wickenburg, AZ	1947–1982	36	416	19.8	4,535	Y	Y
165	09516500	Hassayampa River near Morristown, AZ	1939–2010	27	796	17.0	3,747	N	N
166	09517000	Hassayampa River near Arlington, AZ	1991–2010	20	1,423	14.0	2,901	Y	N
167	09517490	Centennial Wash at Southern Pacific Railroad Bridge, AZ	1981–2010	25	1,681	8.9	1,862	Y	N
168	09517500	Centennial Wash near Arlington, AZ	1961–1979	19	1,769	8.8	1,817	N	N
169	09520170	Rio Cornez near Ajo, AZ	1968–1978	11	244	8.4	1,928	N	N
170	09535100	San Simon Wash near Pisinimo, AZ	1973–2010	38	579	10.2	2,231	Y	N
171	09535300	Vamori Wash at Kom Vo, AZ	1973–2010	37	1,290	14.5	2,664	Y	N
172	09537200	Leslie Creek near McNeal, AZ	1970–2010	36	78.8	18.2	5,332	Y	N
173[1]	09537500	Whitewater Draw near Douglas, AZ	1919–2010	56	1,231	15.8	4,745	Y	N

[1]Downward trend in streamflow.

[2]Station used in flood volume frequency skew analysis but not in the flood peak skew analysis.



Base from U.S. Geological Survey The National Map
Albers Equal-Area Conic projection
Standard parallels 29°30' N and 45°30' N
Central meridian 111°30' W

0 15 30 45 60 MILES

0 15 30 45 60 KILOMETERS

EXPLANATION

▲ Streamgaging stations with flood-duration flow frequency statistics

Eight-digit Hydrologic Unit Code (HUC) boundaries

Figure 4. Map of streamgaging stations in Arizona and western New Mexico used in the flood-duration flow frequency analysis.

Table 2. Streamgaging stations in Arizona and western New Mexico with significant trends in 1-, 3-, 7-, 15-, and 30-day flood-duration flows.

[AZ, Arizona; NM, New Mexico; Y, yes; N, no]

Map ID	Station ID	Station name	Flood duration					Used in skew analysis?
			1 day	3 day	7 day	15 day	30 day	
2	09382000	Paria River at Lees Ferry, AZ	x	x	x	x	x	Y
7	09386950	Zuni River above Black Rock Reservoir, NM	x	x	x	x	x	N
63	09471000	San Pedro River at Charleston, AZ	x	x	x	x	x	Y
173	09537500	Whitewater Draw near Douglas, AZ	x	x	x			Y

Table 3. Basin characteristics and data sources considered in the regionalization analysis of flood-duration flow for streamgaging stations in Arizona.

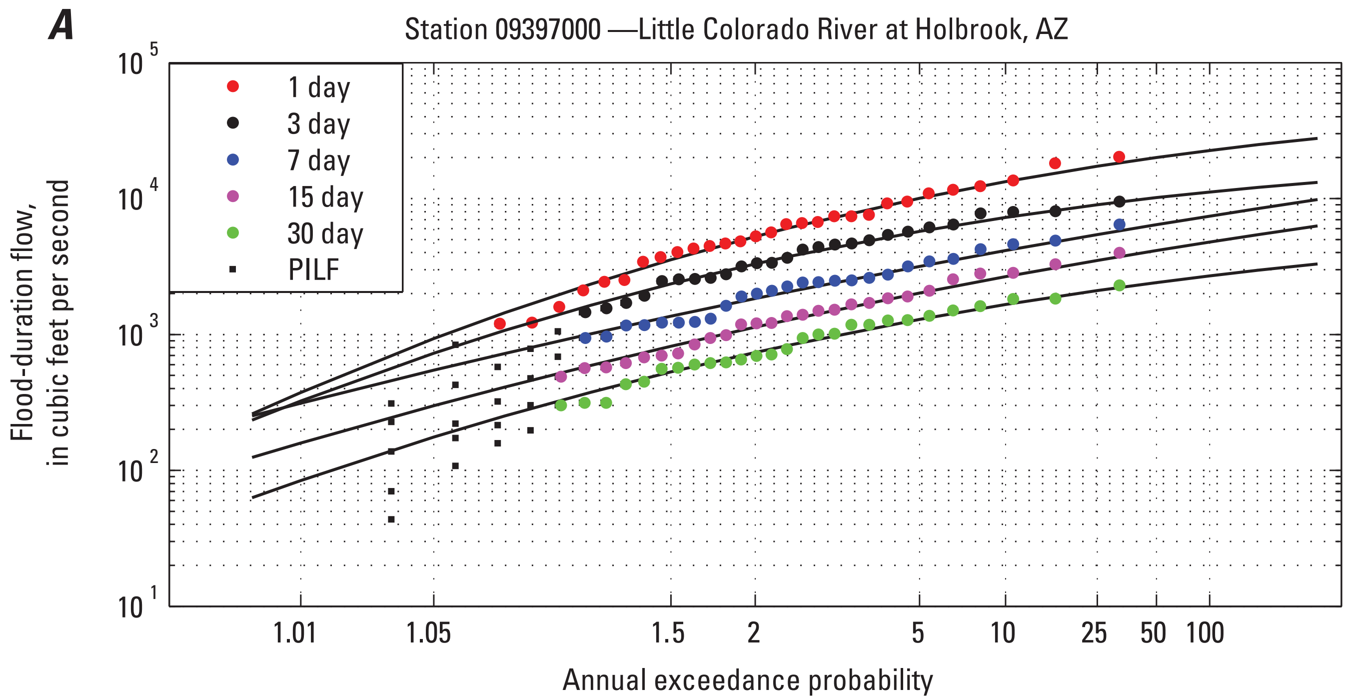
[STATSGO, State Soil Geographic; PRISM, parameter-elevation regressions on independent slopes model; DEM, digital elevation model elevation]

Basin characteristic ID	Basin characteristic description	Data source
DRNAREA	Drainage area/contributing area of the watershed	Calculated from 10-meter DEM
ELEV	Mean basin elevation	10-meter DEM
PRECIP	Mean annual precipitation	PRISM
AUGAVPRE	Mean August precipitation	PRISM
I24H100Y	100-year, 24-hour rainfall intensity	PRISM
SOILPERM	Soil permeability	STATSGO
WATCAP	Soil water capacity	STATSGO

Methods for fitting LP3 moments are described in Bulletin 17B (IACWD, 1982). Research has shown that the computed quantile confidence intervals using Bulletin 17B methods fail to represent the correct uncertainty in the skewness coefficient and that the recommended statistical procedures for computing a regional skewness coefficient are not adequate for estimating the accuracy and precision of the skewness coefficient error (Cohn and others, 2001; Reis and others, 2005). EMA, used with the MGB test, is a revision to the traditional Bulletin 17B estimation methods that explicitly accounts for that method's shortcomings (Cohn and others, 1997; Cohn and others, 2001; England and others, 2003). As with Bulletin 17B, EMA assumes that the LP3 distribution represents the probability distribution function of annual maximum peak flows, except when historical, low-outlier, or censored information exists (Cohn and others, 1997; Griffis and others, 2004). EMA permits the efficient use of interval and threshold data, which most accurately represents historical information, low outliers, and censored flood data (Cohn and others, 1997). Although historical flood records, from first-hand accounts such as newspaper articles or geomorphic evidence such as slackwater deposits, are useful evidence of large floods prior to the systematic record at a particular gage, they provide no information about flood volume. Therefore, the primary benefit of EMA/MBG is to accurately identify and incorporate potentially influential low flows.

Visual inspection of the quantile-probability plots shows that at many streamgaging stations in Arizona, a distinct "dogleg," or shift, exists between a few small events and the remaining data (fig. 5*A*). These small events in the left-hand tail of the distribution, termed potentially influential low flows, can have significant influence on the fit of the distribution to the right-hand tail (that is, the largest flood events with lower AEPs). Therefore, a statistical test is useful to determine if these observations are unusually small compared to the rest of the population. Bulletin 17B allows for the identification and removal (by truncation) of potentially influential low flows using the Grubbs-Beck (GB) test (Grubbs and Beck, 1972), but as implemented in the USGS software PeakFQ version 5.2 (Flynn and others, 2006), typically only a single potentially influential low flow is identified. For the streamgaging stations in this analysis, visual inspection of quantile-probability plots suggest that often several low-flow data points depart from the trend of the data and multiple potentially influential low flows should be considered for censoring. The MGB test, a generalization of the GB test, was developed to address this situation (Gotvald and others, 2012). The MGB test differs from an iterative GB test in that it tests a group of potentially influential low flows against the remaining population simultaneously, rather than removing low flows one at a time. Furthermore, in an EMA analysis, these potentially influential low flows are not completely removed from the analysis as in the B17B-GB procedure but instead are recoded as censored data with reduced influence for determining the LP3 moments.

For this report, the annual series of *n*-day flood-duration flows are assumed independent for each duration, and the MGB test is applied individually to each duration at each station. Nearly half of the streamgaging stations used in the analysis have one or more potentially influential low flows identified using the MGB test (table 4). At some stations, the



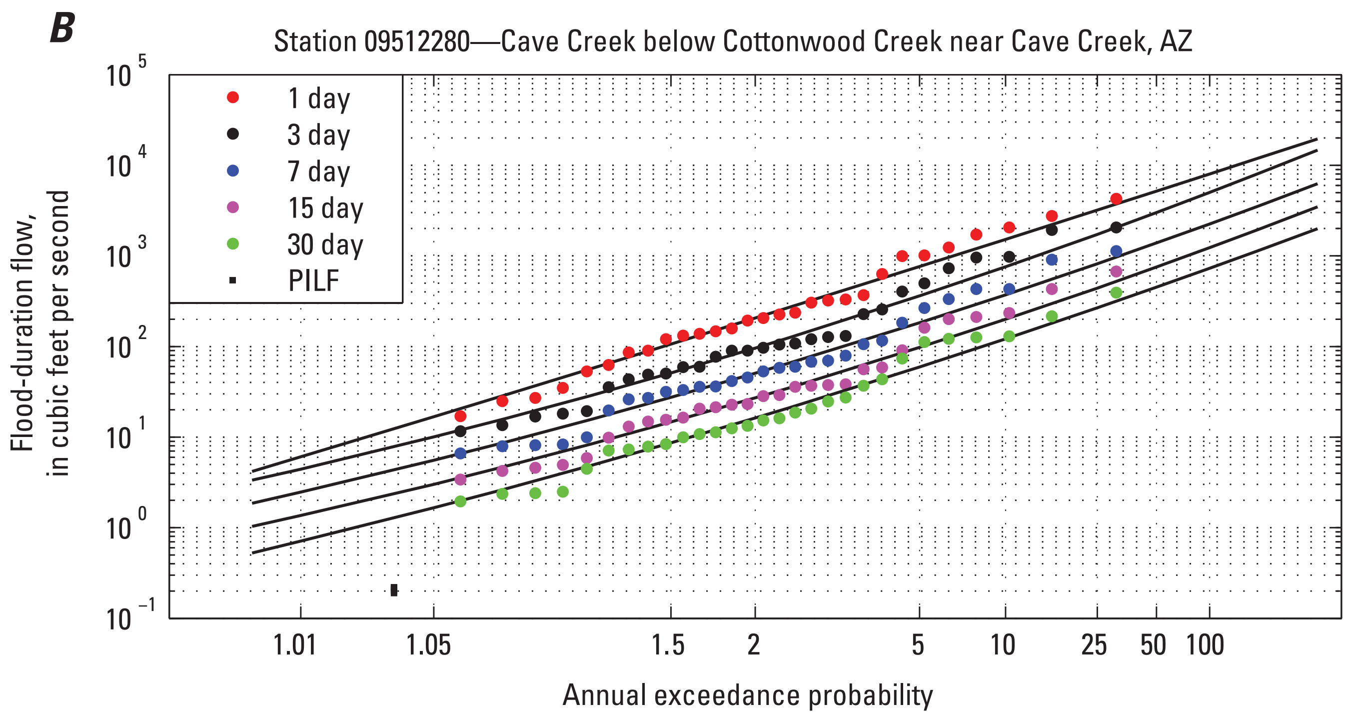


Figure 5. Example quantile-probablility plots of LP3 distributions fit to different flood-duration flow data from streamgaging stations in Arizona. *A*, station 09397000, where different numbers of potentially influential low flows and skewness coefficients were identified for each duration; *B*, station 09512280, where no potentially influential low flows were identified and skewness coefficients are similar for all durations. PILF, potentially influential low flow.

number of potentially influential low flows identified is similar for all durations, and these stations tend to have similar skewness coefficients and goodness-of-fit for all durations (fig. 5*B*). At other stations the number of potentially influential low flows varies differs for the different durations, and the resulting LP3 moments can vary significantly (fig. 5*A*). However, for each duration, the most statistically probable number of potentially influential low flows are identified and treated as censored data, and therefore each duration was not forced to have the same number of potentially influential low flows.

Regional Skew Analysis and Cross-Correlation Models

The third moment of the LP3 distribution, skew, determines the curvature of the fitted distribution on quantile-probability plots. Negative skewness coefficients result in a concave-down profile and relatively low estimates of low-AEP events; positive skews result in a concave-up profile and estimates of low-AEP events are relatively high. To decrease variability from station to station, most studies using the LP3 distribution combine the at-site skewness coefficient estimate, determined from data at a single streamgaging station, with a regional skewness coefficient estimate. For peak-flow studies, a map of regional skew is presented in Bulletin 17B. As an alternative, many recent USGS studies have used a Bayesian generalized least squares (B-GLS) method to relate regional skew to basin characteristics (Reis and others, 2005; Weaver and others, 2009; Parrett and others, 2011). B-GLS regression considers the precision of the regional skew model, differences in record length between stations, and cross correlation of skewness coefficients between stations. The Bayesian aspect of the B-GLS regression provides an estimate of the precision of the estimated model error variance, a pseudo analysis of variance, and enhanced diagnostic statistics (Griffis and Stedinger, 2007).

An important part of B-GLS regression is estimating the cross-correlation of skew between gages, which can be estimated from the cross-correlation of the annual time series between gages (Martins and Stedinger, 2002; Lamontagne and others, 2012):

$$\rho(\gamma_i,\gamma_j)=Sign(\rho_{ij})cf_{ij}|\rho_{ij}|^{\kappa}, \qquad (2)$$

where ρ_{ij} is the cross-correlation of concurrent annual *n*-day flood-duration flows for two streamgaging stations, $Sign(p_{ij})$ denotes the sign (positive or negative) of the cross-correlation, 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, defined as:

$$cf_{ij}=n_{ij}/\sqrt{(n_{ij}+n_i)(n_{ij}+n_j)}, \qquad (3)$$

where n_{ij} is the length of the period of concurrent record, and n_i and n_j are the number of nonconcurrent observations corresponding to sites *i* and *j*, respectively.

As part of the B-GLS skew analysis for peak flows (Paretti and others, 2014b), streamgaging stations suitable for skew analysis were identified as those with record lengths greater than 20 years and adequate LP3 flood-frequency fits. Not all of these stations represent unique watershed characteristics; stations may be identified as redundant if one is nested entirely within another and the streamflow data are highly correlated. The drainage-area ratio of a nested station and the nearest downstream station was used to screen for redundancy; in general, a ratio less than or equal to 5 was used to identify redundant station pairs. When redundant pairs were identified, the station with a longer period of record was retained unless the other station was determined to be better represented by the LP3 distribution using goodness-of-fit criteria (Paretti and others, 2014b). Seventy-nine nonredundant stations were identified for the flood-duration flow skew analysis (table 1, fig. 6). Redundant stations not used in the skew analysis are identified in table 1.

Although the cross-correlation of the concurrent annual flood-duration flows between two sites, ρ_{ij}, has high variability, there is a downward trend with increasing distance (fig. 7). Various models relating cross-correlation to various basin characteristics were considered. A logit model using the Fisher *z*-transform ($Z = \log[(1+r)/(1-r)]$) provided a convenient transformation of the sample correlations r_{ij} from the (−1, +1) range to the (−∞,+∞) range. The adopted model for estimating the cross–correlations of concurrent annual peak flow at two stations, which used the distance between basin centroids, D_{ij}, as the only explanatory variable, is

$$Z_{ij}=b_1+ exp\ (b_2+b_3\times D_{ij})\ , \qquad (4)$$

which is the same form as the cross-correlation model used in the peak-flow analysis (Paretti and others, 2014b). The coefficients b_1, b_2, and b_3 vary for each *n*-day duration (table 5).

The cross-correlation models for *n*-day flood-duration flows show increasing correlation between gages with longer durations of flood flow, and all durations show greater cross-correlation than the time series of annual peak flows (fig. 8). The greater cross-correlation for longer duration flood events can be explained by Arizona's hydroclimatology. The largest flood peaks in watersheds throughout the State are generally caused by summer convective thunderstorms, which are relatively small in spatial extent and of short duration (Sheppard and others, 2002), affecting only one or a few streamgaging stations. In contrast, long-duration, high-volume flood events are often frontal or tropical storms (Sheppard and others, 2002) that cause widespread runoff across many streamgaging stations.

Pseudo-R^2 was used as a diagnostic statistic for the cross-correlation models. Pseudo-R^2 is a measure of the percent of the variability in the dependent variable (*n*-day flood-duration

Table 4. Number of potentially influential low flows and at-site skewness coefficients for streamgaging stations in Arizona and western New Mexico used in the flood-duration flow frequency analysis.

Map ID	Station ID	Number of censored potentially influential low flows for indicated flood-duration flow					Sample log-space skewness coefficient for indicated flood-duration flow				
		1 day	3 day	7 day	15 day	30 day	1 day	3 day	7 day	15 day	30 day
1	09379200	0	0	0	0	0	0.063	0.116	0.049	-0.122	-0.044
2	09382000	0	0	0	0	0	0.073	0.130	0.150	0.150	0.134
3	09383400	12	12	0	0	0	-0.132	-0.200	-0.165	-0.153	-0.204
4	09383500	0	0	0	0	0	-0.121	-0.136	-0.102	-0.084	-0.125
5	09384000	6	8	12	12	0	-0.155	-0.223	-0.195	-0.200	-0.294
6	09386250	3	3	3	3	3	-0.052	-0.145	-0.152	-0.149	-0.094
7	09386950	0	0	0	0	0	-0.228	-0.166	-0.039	0.078	0.147
8	09390500	0	22	24	22	28	-0.039	-0.230	-0.214	-0.250	-0.339
9	09392500	0	0	0	0	0	-0.042	-0.058	-0.112	-0.130	-0.240
10	09393500	0	0	0	0	0	-0.075	0.021	0.005	0.018	-0.035
11	09394500	32	32	0	36	40	-0.146	-0.262	-0.198	-0.243	-0.341
12	09395900	2	0	0	5	5	-0.144	-0.236	-0.210	-0.109	-0.131
13	09397000	2	5	5	4	4	-0.208	-0.269	-0.172	-0.203	-0.346
14	09397500	3	3	7	18	19	-0.117	-0.267	-0.246	-0.194	-0.296
15	09398000	0	0	15	19	22	0.016	-0.104	-0.180	-0.198	-0.298
16	09398500	10	11	18	21	22	-0.179	-0.262	-0.141	-0.156	-0.314
17	09399000	1	7	16	22	24	-0.069	-0.208	-0.170	-0.200	-0.260
18	09400562	1	1	4	5	3	-0.131	-0.264	-0.118	-0.124	-0.266
19	09400568	1	1	1	1	1	-0.168	-0.160	-0.135	-0.119	-0.228
20	09400583	2	2	2	2	2	-0.078	-0.070	-0.073	-0.087	-0.139
21	09401000	0	0	2	2	2	-0.061	-0.120	0.009	0.081	0.046
22	09401110	4	5	8	5	4	-0.048	-0.123	-0.179	-0.151	-0.165
23	09401260	0	0	0	0	0	-0.139	-0.255	-0.258	-0.247	-0.326
24	09401280	0	0	0	0	0	-0.116	-0.111	-0.125	-0.118	-0.080
25	09401400	0	0	0	0	0	-0.169	-0.184	-0.192	-0.192	-0.284
26	09401500	1	1	1	1	1	-0.095	-0.138	-0.060	0.047	0.114
27	09402000	0	0	0	0	0	-0.054	-0.045	-0.145	-0.211	-0.257
28	09403000	0	0	0	0	0	0.079	-0.023	-0.102	-0.109	-0.112
29	09403780	5	0	6	7	5	-0.029	-0.269	-0.189	-0.191	-0.275
30	09404110	0	0	0	0	0	-0.018	-0.002	-0.015	-0.035	-0.086
31	09404208	7	0	0	0	0	-0.107	-0.074	-0.028	0.028	0.049
32	09404222	0	0	0	0	0	-0.070	-0.090	-0.075	-0.011	-0.008
33	09404343	1	0	0	1	1	-0.159	-0.240	-0.213	-0.163	-0.246
34	09415000	0	0	0	0	0	0.125	0.204	0.171	0.202	0.237

Table 4. Number of potentially influential low flows and at-site skewness coefficients for streamgaging stations in Arizona and western New Mexico used in the flood-duration flow frequency analysis.—Continued

Map ID	Station ID	Number of censored potentially influential low flows for indicated flood-duration flow					Sample log-space skewness coefficient for indicated flood-duration flow				
		1 day	3 day	7 day	15 day	30 day	1 day	3 day	7 day	15 day	30 day
35	09424200	0	0	0	0	0	-0.123	-0.223	-0.210	-0.236	-0.342
36	09424447	0	0	0	0	0	-0.095	-0.150	-0.145	-0.130	-0.189
37	09424450	18	18	20	20	20	-0.138	-0.154	-0.141	-0.154	-0.139
38	09424900	20	20	20	20	20	-0.153	-0.233	-0.201	-0.207	-0.320
39	09425500	0	0	0	0	0	-0.122	-0.144	-0.101	-0.075	-0.063
40	09426500	0	0	7	0	0	-0.151	-0.220	-0.174	-0.201	-0.275
41	09430500	0	0	0	0	0	0.064	0.056	-0.024	-0.028	-0.034
42	09431500	31	0	0	0	0	-0.169	-0.035	-0.110	-0.078	-0.069
43	09442000	0	0	0	28	29	0.048	-0.013	-0.094	-0.181	-0.310
44	09442680	0	0	0	21	22	0.088	0.026	-0.059	-0.206	-0.111
45	09444000	0	0	41	0	0	0.121	0.131	-0.175	0.082	0.049
46	09444200	0	0	0	0	0	-0.021	-0.017	-0.050	-0.070	-0.098
47	09444500	0	0	0	0	0	0.021	0.043	0.048	0.064	0.042
48	09445500	0	0	0	0	0	0.017	0.060	0.049	0.077	0.098
49	09446000	0	0	0	0	0	-0.005	0.052	0.049	0.086	0.100
50	09446500	0	0	0	0	0	0.024	0.054	0.043	0.088	0.102
51	09447000	0	0	0	0	0	0.085	0.183	0.192	0.270	0.317
52	09447800	0	0	0	0	0	-0.094	-0.027	0.002	0.094	0.144
53	09448500	0	0	0	0	0	0.051	0.066	0.043	0.075	0.069
54	09456000	0	0	0	0	0	-0.132	-0.188	-0.145	-0.120	-0.142
55	09457000	1	0	0	0	0	0.013	-0.123	-0.050	-0.091	-0.099
56	09458200	4	5	6	6	5	-0.059	-0.204	-0.147	-0.052	-0.157
57	09458500	0	0	0	0	0	-0.047	-0.080	-0.129	-0.122	-0.122
58	09460150	3	3	4	11	11	-0.189	-0.279	-0.200	-0.154	-0.189
59	09468500	0	40	40	0	0	0.026	-0.278	-0.236	0.018	0.064
60	09470500	1	1	1	1	1	0.034	0.117	0.017	-0.040	-0.196
61	09470750	0	0	0	0	0	-0.149	-0.226	-0.197	-0.210	-0.313
62	09470800	0	3	3	8	8	-0.144	-0.203	-0.195	-0.123	-0.182
63	09471000	1	1	1	1	1	-0.019	0.034	0.013	-0.016	-0.162
64	09471310	0	0	0	0	0	-0.145	-0.218	-0.181	-0.181	-0.251
65	09471380	1	1	0	0	0	-0.038	-0.095	-0.155	-0.128	-0.070
66	09471400	0	0	0	0	0	-0.084	-0.192	-0.188	-0.200	-0.278
67	09471550	1	1	1	8	8	0.013	-0.030	-0.076	-0.172	-0.250
68	09471800	0	1	1	1	1	-0.122	-0.177	-0.100	-0.121	-0.215

Table 4. Number of potentially influential low flows and at-site skewness coefficients for streamgaging stations in Arizona and western New Mexico used in the flood-duration flow frequency analysis.—Continued

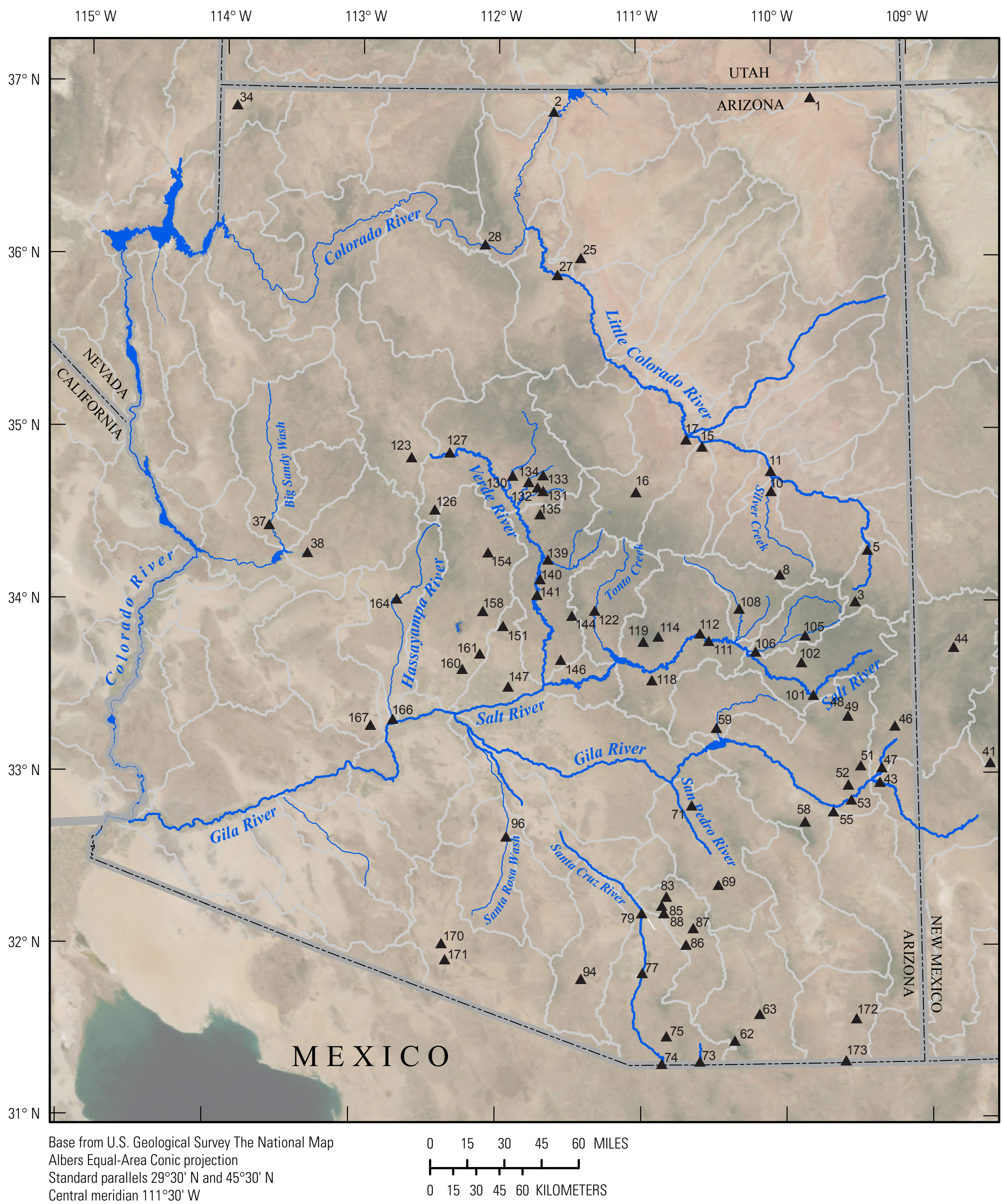
Map ID	Station ID	Number of censored potentially influential low flows for indicated flood-duration flow					Sample log-space skewness coefficient for indicated flood-duration flow				
		1 day	3 day	7 day	15 day	30 day	1 day	3 day	7 day	15 day	30 day
69	09472000	6	5	4	2	2	-0.140	-0.254	-0.242	-0.304	-0.282
70	09472050	1	1	1	1	1	-0.160	-0.207	-0.191	-0.161	-0.323
71	09473000	0	0	0	0	0	0.088	0.147	0.128	0.178	0.222
72	09473500	0	0	0	0	0	-0.038	-0.035	-0.034	-0.033	-0.133
73	09480000	6	6	7	7	0	-0.146	-0.242	-0.178	-0.222	-0.383
74	09480500	0	0	0	0	0	-0.009	0.015	-0.114	-0.183	-0.304
75	09481500	0	0	0	0	0	-0.142	-0.220	-0.164	-0.016	-0.014
76	09481740	0	0	0	0	0	-0.024	-0.055	-0.042	0.001	0.005
77	09482000	0	0	0	0	0	0.036	0.043	0.058	0.070	0.007
78	09482400	0	0	0	0	0	-0.061	-0.084	-0.100	-0.102	-0.173
79	09482500	0	1	0	0	0	-0.140	-0.045	-0.144	-0.138	-0.161
80	09483000	0	0	0	0	0	-0.029	-0.066	-0.092	-0.128	-0.298
81	09483010	0	0	0	0	0	-0.049	-0.067	-0.111	-0.090	-0.034
82	09483100	0	0	0	0	0	-0.055	-0.072	-0.112	-0.057	-0.023
83	09484000	1	1	0	0	0	0.039	-0.020	-0.171	-0.126	-0.226
84	09484200	4	4	4	4	4	-0.109	-0.149	-0.150	-0.028	-0.050
85	09484500	12	8	8	12	7	-0.132	-0.225	-0.187	-0.073	-0.300
86	09484600	2	1	1	1	9	-0.128	-0.160	-0.146	-0.153	-0.312
87	09485000	19	13	15	14	7	-0.138	-0.210	-0.138	-0.179	-0.342
88	09485450	1	0	0	1	1	-0.097	-0.220	-0.199	-0.142	-0.219
89	09485700	2	2	2	2	2	-0.056	-0.076	-0.051	-0.056	-0.103
90	09486055	2	2	2	3	3	-0.092	-0.103	-0.136	-0.149	-0.233
91	09486300	0	0	0	0	0	-0.104	-0.169	-0.161	-0.165	-0.264
92	09486350	0	0	0	0	0	-0.161	-0.181	-0.159	-0.124	-0.208
93	09486500	0	0	0	0	0	-0.057	0.004	-0.026	-0.002	-0.081
94	09486800	0	0	0	0	3	-0.136	-0.222	-0.188	-0.222	0.008
95	09487000	1	1	1	1	1	-0.191	-0.267	-0.273	-0.290	-0.386
96	09488500	0	7	7	7	7	-0.174	-0.174	-0.093	-0.087	-0.079
97	09489000	0	0	0	0	0	0.076	0.109	0.092	0.108	0.062
98	09489070	0	0	0	0	0	-0.147	-0.199	-0.175	-0.176	-0.230
99	09489100	0	0	10	0	0	-0.034	-0.100	-0.169	-0.153	-0.179
100	09489200	10	11	11	11	11	-0.115	-0.208	-0.182	-0.174	-0.227
101	09489500	0	10	11	20	7	-0.204	-0.147	-0.278	-0.208	-0.427
102	09489700	0	0	0	11	11	-0.086	-0.123	-0.159	-0.200	-0.282

Table 4. Number of potentially influential low flows and at-site skewness coefficients for streamgaging stations in Arizona and western New Mexico used in the flood-duration flow frequency analysis.—Continued

Map ID	Station ID	Number of censored potentially influential low flows for indicated flood-duration flow					Sample log-space skewness coefficient for indicated flood-duration flow				
		1 day	3 day	7 day	15 day	30 day	1 day	3 day	7 day	15 day	30 day
103	09490500	0	0	0	10	25	-0.118	-0.173	-0.218	-0.261	-0.246
104	09490800	0	0	0	0	0	-0.024	-0.021	-0.028	-0.025	-0.056
105	09492400	1	1	2	11	19	0.105	0.009	-0.173	-0.269	-0.223
106	09494000	0	1	24	26	26	-0.066	-0.133	-0.197	-0.224	-0.318
107	09496000	0	9	13	12	13	-0.099	-0.224	-0.174	-0.182	-0.259
108	09496500	0	19	19	19	20	-0.093	-0.212	-0.126	-0.169	-0.329
109	09496600	0	0	0	0	0	-0.076	-0.162	-0.111	-0.014	0.005
110	09496700	0	0	0	0	0	-0.026	-0.069	-0.041	-0.063	-0.023
111	09497500	0	0	0	39	41	0.020	0.048	-0.008	-0.133	-0.164
112	09497800	0	0	0	0	0	0.004	0.104	0.118	0.213	0.254
113	09497900	0	0	5	5	0	-0.092	-0.174	-0.098	-0.091	-0.195
114	09497980	9	10	11	11	11	-0.138	-0.268	-0.238	-0.196	-0.260
115	09498400	0	0	0	0	0	-0.053	0.012	0.083	0.123	0.143
116	09498500	0	0	0	0	0	-0.001	0.047	0.049	0.057	-0.038
117	09498501	1	0	0	0	0	-0.077	-0.221	-0.182	-0.164	-0.180
118	09498502	0	0	0	0	0	-0.172	-0.224	-0.181	-0.163	-0.156
119	09498503	2	2	2	2	3	-0.117	-0.145	-0.209	-0.249	-0.285
120	09498800	0	0	0	0	0	-0.132	-0.191	-0.178	-0.111	-0.091
121	09498870	0	0	0	0	0	-0.133	-0.215	-0.201	-0.204	-0.297
122	09499000	30	1	32	17	17	-0.176	-0.288	-0.133	-0.140	-0.245
123	09502800	0	0	0	0	0	-0.161	-0.226	-0.161	-0.139	-0.143
124	09502900	0	0	0	0	0	-0.069	-0.099	-0.064	-0.060	-0.157
125	09502960	0	0	0	0	0	-0.014	-0.012	-0.047	-0.024	-0.108
126	09503000	0	0	0	0	15	-0.044	-0.081	-0.089	-0.047	-0.285
127	09503700	0	0	0	0	0	0.044	0.115	0.157	0.234	0.292
128	09504000	0	0	0	0	0	-0.074	-0.082	-0.069	-0.033	-0.016
129	09504420	0	5	9	10	0	-0.189	-0.234	-0.120	-0.097	-0.226
130	09504500	0	0	18	22	21	-0.083	-0.213	-0.209	-0.240	-0.323
131	09505200	21	15	16	15	21	-0.148	-0.220	-0.115	-0.186	-0.291
132	09505250	5	5	8	7	8	-0.132	-0.166	-0.071	-0.185	-0.064
133	09505300	0	6	6	5	5	-0.194	-0.190	-0.161	-0.166	-0.265
134	09505350	0	11	11	13	16	-0.178	-0.223	-0.212	-0.231	-0.333
135	09505800	18	18	18	20	15	-0.152	-0.216	-0.164	-0.176	-0.346
136	09506000	0	0	0	0	0	-0.064	-0.075	-0.071	-0.038	-0.100

Table 4. Number of potentially influential low flows and at-site skewness coefficients for streamgaging stations in Arizona and western New Mexico used in the flood-duration flow frequency analysis.—Continued

Map ID	Station ID	Number of censored potentially influential low flows for indicated flood-duration flow					Sample log-space skewness coefficient for indicated flood-duration flow				
		1 day	3 day	7 day	15 day	30 day	1 day	3 day	7 day	15 day	30 day
137	09507600	0	0	3	4	4	-0.138	-0.228	-0.105	-0.144	-0.082
138	09507700	0	0	0	0	0	-0.138	-0.221	-0.223	-0.240	-0.267
139	09507980	23	23	0	0	0	-0.160	-0.238	-0.285	-0.233	-0.248
140	09508300	3	3	2	7	20	-0.207	-0.321	-0.324	-0.296	-0.231
141	09508500	0	29	0	22	0	-0.129	-0.201	-0.128	-0.165	-0.060
142	09510070	2	6	6	2	0	-0.140	-0.133	-0.108	-0.174	-0.287
143	09510080	0	0	0	0	0	-0.163	-0.247	-0.214	-0.231	-0.324
144	09510100	0	5	0	7	9	-0.190	-0.259	-0.255	-0.193	-0.273
145	09510150	0	0	0	0	0	-0.069	-0.155	-0.181	-0.196	-0.262
146	09510200	4	10	0	13	4	-0.213	-0.234	-0.335	-0.229	-0.462
147	09512100	3	3	3	3	3	-0.113	-0.208	-0.204	-0.208	-0.301
148	09512162	0	0	0	0	0	-0.103	-0.082	-0.018	0.022	-0.014
149	09512165	2	2	2	2	2	-0.089	-0.133	-0.109	-0.071	-0.075
150	09512200	16	16	16	16	16	-0.131	-0.203	-0.180	-0.189	-0.283
151	09512280	1	1	1	1	1	-0.057	0.041	0.025	0.060	0.040
152	09512400	8	11	8	8	6	-0.123	-0.207	-0.224	-0.235	-0.357
153	09512450	0	0	0	0	0	-0.125	-0.207	-0.178	-0.187	-0.228
154	09512500	0	0	0	0	0	0.080	0.091	0.024	0.079	0.044
155	09512600	1	1	1	1	1	-0.115	-0.141	-0.079	-0.005	0.000
156	09512800	0	0	0	0	0	0.024	0.004	-0.043	-0.051	-0.059
157	09512860	0	0	0	0	0	-0.039	-0.040	-0.038	-0.009	-0.021
158	09513780	10	3	3	3	22	-0.161	-0.331	-0.299	-0.319	-0.318
159	09513800	2	2	2	0	0	-0.199	-0.291	-0.265	-0.286	-0.400
160	09513835	7	7	8	8	8	-0.137	-0.227	-0.170	-0.172	-0.229
161	09513860	14	14	14	14	15	-0.124	-0.225	-0.208	-0.235	-0.285
162	09513910	2	2	2	2	2	-0.053	-0.083	-0.079	-0.050	-0.143
163	09513970	7	7	7	7	7	-0.115	-0.168	-0.141	-0.137	-0.213
164	09515500	11	11	0	0	0	-0.082	-0.192	-0.057	-0.012	0.035
165	09516500	1	5	1	1	1	-0.206	-0.208	-0.255	-0.193	-0.253
166	09517000	0	0	0	0	0	-0.030	0.038	0.017	0.047	0.085
167	09517490	1	12	1	12	6	-0.193	-0.050	-0.295	-0.045	-0.337
168	09517500	4	4	4	4	4	-0.160	-0.230	-0.206	-0.219	-0.326
169	09520170	0	1	0	1	0	-0.154	-0.135	-0.180	-0.099	-0.162
170	09535100	2	2	2	2	2	-0.025	-0.028	-0.075	-0.082	0.001
171	09535300	0	0	0	0	0	0.068	0.069	0.033	-0.059	-0.129
172	09537200	15	15	15	8	0	-0.130	-0.203	-0.183	-0.266	-0.392
173	09537500	17	20	5	7	6	-0.150	-0.174	-0.237	-0.220	-0.231



EXPLANATION

▲ Streamgaging stations used in regional skew analysis

Eight-digit Hydrologic Unit Code (HUC) boundaries

Figure 6. Map showing streamgaging stations in Arizona and western New Mexico used in the regional skew analysis.

Table 5. Model coefficients in equation 4 and pseudo-R^2 *for* the cross-correlation models of annual time series of *n*-day flood-duration flow for Arizona.

Flood duration	Beta parameters			Pseudo-R^2
	b_1	b_2	b_3	
1 day	0.081	0.120	-0.0078	60 percent
3 day	0.087	0.165	-0.0078	59 percent
7 day	0.128	0.187	-0.0084	57 percent
15 day	0.159	0.194	-0.0085	54 percent
30 day	0.153	0.254	-0.0081	52 percent
Annual peaks	0.11	-0.67	-0.0094	35 percent

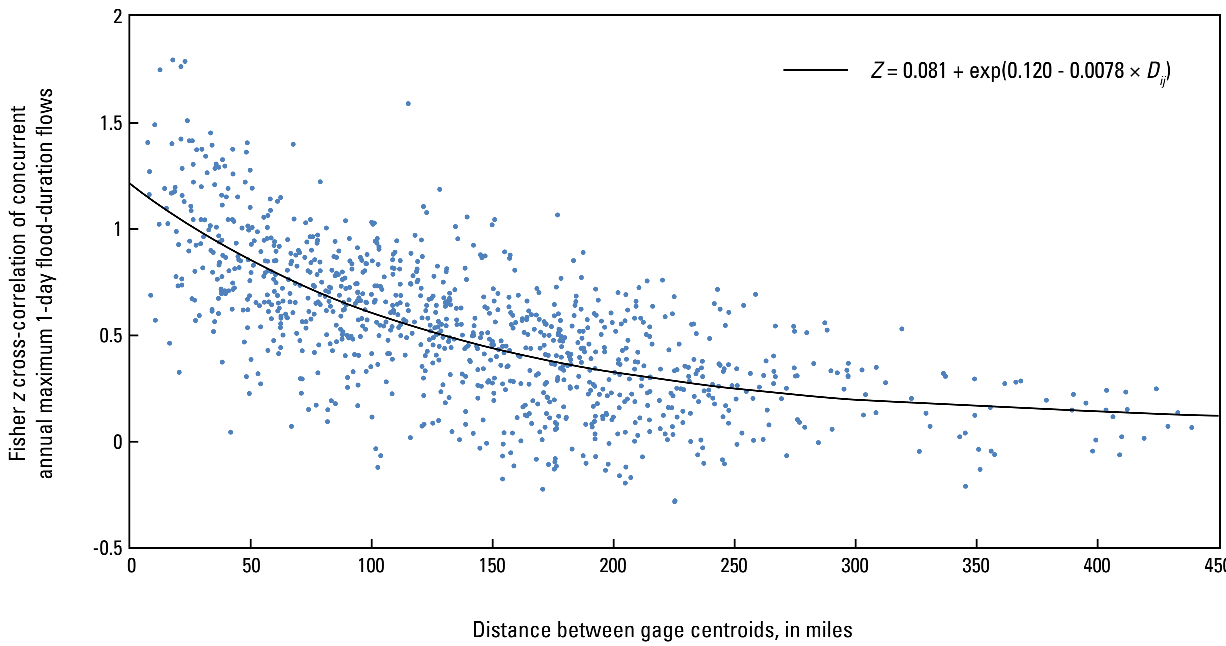


Figure 7. Scatterplot showing the cross-correlation model for 1-day flood-duration flows for Arizona. Each point represents the correlation of the annual time series between two streamgaging stations. *Z*, Fisher *z* cross-correlation; exp, exponential function; D_{ij}, distance between basin centroids.

flow) explained by the regression after removing the effect of time-sampling error, calculated as:

$$R^2_{pseudo} = 1 - \frac{\sigma^2_\delta(k)}{\sigma^2_\delta(0)}, \quad (5)$$

where $\sigma^2_\delta(k)$ is the model error variance from a regression analysis with k independent variables and $\sigma^2_\delta(0)$ is the model error variance from a regression analysis with no independent variables. Pseudo-R^2 (table 5) decreases with increasing length of flood-duration flow but is consistently higher for all durations than for the cross-correlation model developed for annual peaks (Paretti and others, 2014b).

The significant cross-correlation between stations complicates the GLS regression, but the relatively low precision of the cross-correlation model doesn't justify the sophisticated weighting matrix generated by B-GLS. Therefore, an alternative procedure was used (presented in detail in Lamontagne and others, 2012). First, an ordinary least squares (OLS) analysis is used to develop an initial regional-skew model, which is then used to generate a regional-skew estimate for each site. That OLS skewness coefficient estimate is used to compute the sampling variance of each skew estimator for use in a WLS analysis. Then, WLS is used to generate the estimator of the regional-skew model parameters. Finally, B-GLS is used to estimate the precision of that parameter estimator and to estimate the model-error variance. The three-step procedure was repeated to develop a regional-skew model and the associated error analysis for each flood duration.

The at-site skew for nearly all of the streamgaging stations in the study is negative; for all durations, less than 20 percent of stations have positive skew (table 4). Several basin characteristics were tested as explanatory variables in the B-GLS skew regression, including location (latitude and longitude), drainage area, mean elevation, mean annual precipitation, August mean precipitation (representative of summer

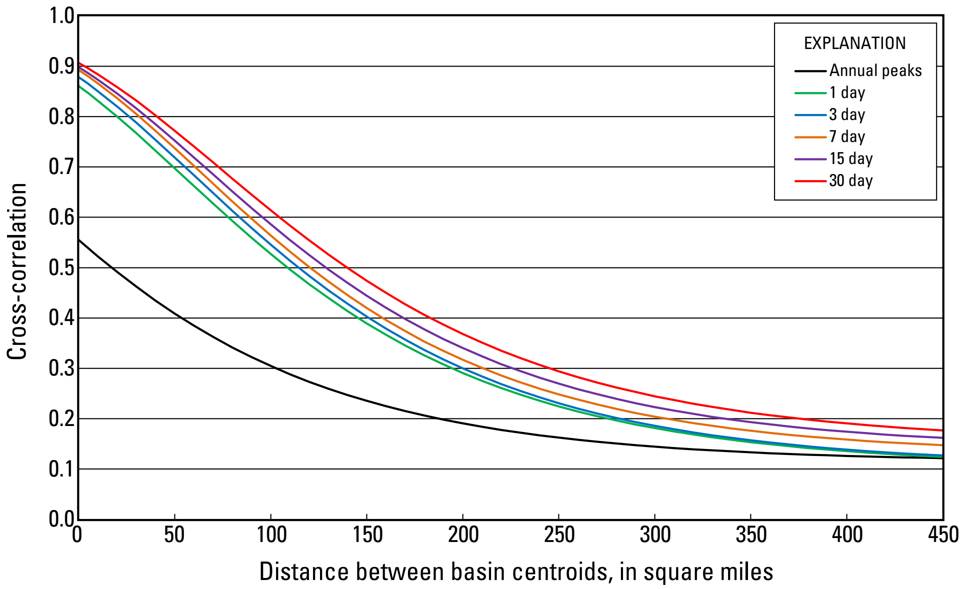


Figure 8. Graph showing cross-correlation models used in the regional skew and regional regression analyses for flood-duration flow for Arizona. The cross-correlation model for annual peak flows (Paretti and others, 2014b) is shown for reference.

convective thunderstorm runoff), 24-hour 100-year precipitation intensity, soil permeability, and soil water capacity. Maps of station skewness coefficients were also created to evaluate spatial patterns. None of these explanatory variables significantly improved skew estimates as compared to a constant model (the weighted average skew at all stations). Regional skew is highest for 1-day flood-duration flow, most negative for 30-day flood-duration flow, and the intervening 3-, 7-, and 15-day flood-duration flow skewness coefficients are intermediate, although not in a systematic manner (table 6). For comparison, the constant skewness coefficient for annual flood peaks is −0.09 (Paretti and others, 2014b) and the average variance of prediction (*AVP*) is 0.079. *AVP* is a diagnostic statistic that reflects both the underlying model error variance, σ_δ^2 , and the sampling variance:

$$AVP = \sigma_\delta^2 + \frac{1}{n}\sum_{p=1}^{n} \mathbf{x}_p (\mathbf{X}^T \boldsymbol{\Lambda}^{-1} \mathbf{X})^{-1} \mathbf{x}_p^T \qquad (6)$$

where $\mathbf{x}_p$ is a vector of independent variables at the *p*th gage, and $\mathbf{X}$ and $\boldsymbol{\Lambda}$ are the design matrix and covariance matrix from the regression analysis, respectively. *AVP* is used to weight the regional skewness coefficient when combined with station skewness coefficient to determine a weighted skewness coefficient (Reis and others, 2005):

$$\gamma_w = \frac{AVP_{new}\gamma_s + MSE_s\gamma_r}{AVP_{new} + MSE_s} \qquad (7)$$

where γ_w, γ_r, and γ_a are the weighted, regional, and station skewness coefficients, respectively, and MSE_s is the estimated mean square error of the station skewness coefficient. Further details of the skew analysis in Arizona are found in Paretti and others (2014b).

Table 6. Regional skewness coefficients and their average variance of prediction, by duration, for flood-duration flows in Arizona.

Flood duration	Skewness coefficient	Average variance of prediction
1 day	-0.103	0.091
3 day	-0.155	0.158
7 day	-0.133	0.133
15 day	-0.130	0.169
30 day	-0.209	0.259
Annual peaks	-0.090	0.079

Regionalization

Definition of Regions

The spatial extent and density of streamgaging stations with adequate data for the *n*-day flood-duration flow frequency analysis is smaller than that of the stations used to analyze peak flows (Paretti and others, 2014b), and, if the same regions were used (high elevation, Colorado Plateau, western Basin and Range, central highlands, and southeastern Basin and Range) there would be an inadequate number of gages in each region to define regression equations. Therefore, three alternative regions were tested:

- A single statewide region using all of the streamgaging stations with adequate data.
- A single statewide region but with the furthest outlying gages removed. This region includes streamgaging stations in the Colorado Plateau, central highlands, and

southeastern Basin and Range regions (Paretti and others, 2014b).

- A single region comprising the central highland region only (Paretti and others, 2014b) with the addition of streamgaging stations lying just outside this region to the southwest.

Two statistics were used to evaluate the alternative region definitions—(1) average standard error of prediction ($SE_{p,ave}$) and (2) pseudo-R^2 (equation 5). $SE_{p,ave}$ is an alternative way to express *AVP* as a percent of the predicted flood volume and is simply a transformation of units:

$$SE_{p,ave}=100\{e^{(ln10)^2 \mathrm{AVP}}-1\}^{1/2} . \quad (8)$$

Regression diagnostic statistics were significantly better using the third alternative, which is comprised of 85 stations (table 7). Therefore, regression equations were only developed for the central highland region. Outside this region, of the stations with record lengths and statistical fits adequate for regression analysis, only 30 stations are located to the south and 22 stations to the north. These were considered insufficient to be standalone regions, and regression equations in these areas are not presented. Flood-volume frequency estimates at streamgaging stations outside of the central highland region, where the regression equations are not applicable, are presented in appendix 1 (note that appendixes 1–3 are available online only at http://pubs.usgs.gov/of/2014/5109).

The basin characteristics of streamgaging stations used in the regression analysis for the central highland region vary widely as a result of the diverse topography and climatology in this area (table 8). The drainage areas of streamgaging stations used in the regression analysis vary over four orders of magnitude (fig. 10), but most are between 25 and 2,500 square miles. Basin centroid elevations also vary widely, from 2,240 foot elevation (Skunk Creek near Phoenix, station 09513860) to 9,520 foot elevation (North Fork White River near Greer, station 09490800). Mean basin precipitation varies from 13.9 inches in the low desert to 36.7 inches at high elevations, part of which is typically from snowfall (fig. 2*B*). The regression equations are only valid within the parameter ranges of these basin characteristics (table 8) and the region of the applicable streamgaging stations (fig. 9).

Model Development

Regression analysis was used to relate the quantile AEP estimates at each streamgaging station to basin characteristics, such as drainage area or mean annual rainfall. If an explanatory relation exists, the regression equations can then be used to make predictions at ungaged sites where no flood-duration flow data are available. Initially, seven basin characteristics considered to be most closely related to flood-duration flow (table 3) were tested for significance using OLS regression and the weighted-multiple-linear regression program WREG (Eng and others, 2009) by evaluating the *T value* statistic for each characteristic,

$$T\ value = \frac{\beta_k}{(Var\beta_k)^{1/2}} , \quad (9)$$

where β_k is the predicted coefficient of the *k*th basin characteristic and $Var\beta_k$ is the covariance of taken from the covariance matrix of the regression parameters. The *T value* statistic is assumed to follow a Student's *t*-distribution, and the probability, or *p*-value, that the null hypothesis (H_0, the model parameter is equal to zero) should be rejected can be calculated. Regression parameters with *p*-values less than 0.05 are deemed to be significant and are included in the final regression equations.

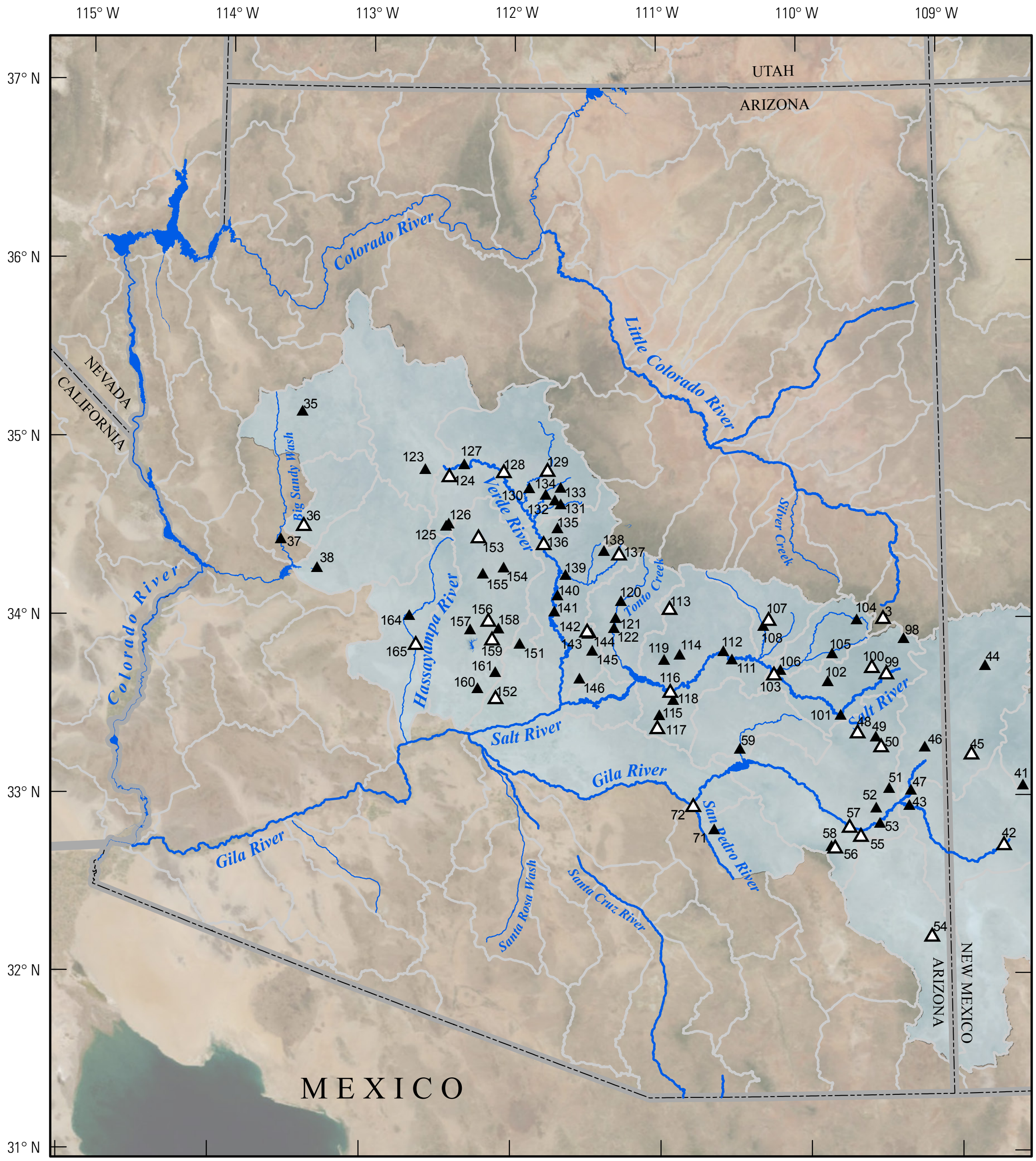
One assumption of regression analyses is that the explanatory variables are independent. If multicollinearity (correlation between variables) exists, model error may be underestimated. Variance inflation factor (VIF) was used to screen for multicollinearity (Johnston, 1972). Although increasing elevation generally corresponds to increasing precipitation in Arizona (fig. 10), the VIF for these two variables is 1.7, well below the commonly used threshold of 10 (Kroll and Song, 2013), and both were retained in the regression equations.

The final regression equations (table 9) use GLS regression to account for the cross-correlation between gages. As with the GLS regression used for the skew analysis, covariance matrices that account for cross-correlation of annual flood-duration flows are needed for each AEP for the regional regression analysis. The same cross-correlation models (table 5) were used for both.

For this study, the log of drainage area in square miles (DRNAREA), the log of mean annual precipitation in inches (PRECIP), and mean basin elevation in feet divided by 1,000 (ELEV) were determined to be significant in the final equations identified using GLS regression. ELEV was divided by 1,000 so that regression coefficients were smaller and more easily calculated (Eng and others, 2009). For most durations and AEPs, all three explanatory variables are included in the regression equations, but mean basin elevation is not included for some 50-percent AEP equations.

Model Diagnostics and Verification

Two statistics, leverage and influence, serve as regression model diagnostics. Leverage is calculated from the covariance matrix used in the GLS regression analysis. It represents the potential impact a single streamgaging station has on the regression and is primarily a factor of how "unique" a station is. If basin characteristics at a particular station are far from the mean of the remaining stations, it can potentially, but not necessarily, have a dominant effect on the regression. Such a station is said to have high leverage. Alternatively, influence measures the actual effect a particular station has on the



Base from U.S. Geological Survey The National Map
Albers Equal-Area Conic projection
Standard parallels 29°30' N and 45°30' N
Central meridian 111°30' W

0 15 30 45 60 MILES

0 15 30 45 60 KILOMETERS

EXPLANATION

▲ Streamgaging stations used in the regional regression analysis

△ Streamgaging stations with regression-weighted flood-duration flow estimates

Region over which volume regression equations apply

Eight-digit Hydrologic Unit Code (HUC) boundaries

Figure 9. Map showing the region in Arizona and western New Mexico for which regression equations are developed and streamgaging stations used in the analysis of flood-duration flow in Arizona.

Table 7. Regression diagnostic statistics for three regionalization schemes used for estimating flood duration-flows in Arizona.

[Pct. AEP, percent annual exceedance probability; Avg. SEP, average standard error of prediction]

	All gages		Single statewide region, select gages		Central highland region only	
1 day						
Pct. AEP	**Avg. SEP**	**Pseudo-R^2**	**Avg. SEP**	**Pseudo-R^2**	**Avg. SEP**	**Pseudo-R^2**
50	70.9	85.2	72.2	86.2	50.8	90.4
20	65.9	86.0	67.2	87.0	43.7	92.1
10	68.4	85.0	69.9	86.1	40.6	92.7
4	75.3	82.7	75.9	84.3	37.0	93.5
2	81.1	80.9	81.2	82.8	36.4	93.6
1	87.3	79.1	87.8	81.0	36.7	93.5
0.5	93.9	77.2	93.9	79.4	37.8	93.1
0.2	102.5	74.9	102.8	77.2	37.7	93.1
3 day						
Pct. AEP	**Avg. SEP**	**Pseudo-R^2**	**Avg. SEP**	**Pseudo-R^2**	**Avg. SEP**	**Pseudo-R^2**
50	71.7	85.6	70.9	87.2	52.3	90.0
20	65.8	86.8	66.7	87.8	43.6	92.4
10	66.5	86.4	68.3	87.2	40.4	93.3
4	69.4	85.3	71.6	86.2	40.3	93.2
2	72.6	84.3	75.0	85.2	39.3	93.6
1	76.8	83.0	78.7	84.2	39.9	93.5
0.5	81.1	81.8	82.6	83.2	40.6	93.4
0.2	86.3	80.4	88.1	81.8	35.3	95.2
7 day						
Pct. AEP	**Avg. SEP**	**Pseudo-R^2**	**Avg. SEP**	**Pseudo-R^2**	**Avg. SEP**	**Pseudo-R^2**
50	72.4	85.9	71.5	87.4	52.9	90.2
20	65.7	87.1	66.3	88.1	42.7	92.6
10	66.0	86.6	66.7	87.7	36.1	94.5
4	68.5	85.6	69.5	86.6	34.9	94.7
2	72.1	84.3	72.6	85.6	32.6	95.4
1	75.3	83.3	76.8	84.4	33.1	95.4
0.5	79.4	82.0	80.4	83.4	33.6	95.3
0.2	85.0	80.3	86.4	81.7	35.2	95.0
15 day						
Pct. AEP	**Avg. SEP**	**Pseudo-R^2**	**Avg. SEP**	**Pseudo-R^2**	**Avg. SEP**	**Pseudo-R^2**
50	75.0	85.7	73.1	87.4	51.3	91.2
20	67.0	86.7	66.6	88.0	40.4	93.6
10	65.9	86.9	66.4	87.9	36.4	94.5
4	68.2	85.8	69.6	86.6	34.9	94.7
2	71.6	84.6	72.4	85.6	32.6	95.4
1	74.5	83.6	75.5	84.6	32.0	95.6
0.5	78.4	82.2	79.7	83.3	31.6	95.8
0.2	75.9	82.3	78.1	83.2	28.7	96.9

Table 7. Regression diagnostic statistics for three regionalization schemes used for estimating flood duration-flows in Arizona.—Continued

	All gages		Single statewide region, select gages		Central highland region only	
			30 day			
Pct. AEP	Avg. SEP	Pseudo-R^2	Avg. SEP	Pseudo-R^2	Avg. SEP	Pseudo-R^2
50	76.9	85.7	72.9	88.2	49.5	92.0
20	68.4	86.8	68.0	88.0	42.7	92.9
10	68.1	86.7	70.4	86.7	38.2	94.1
4	70.0	85.6	70.4	86.6	36.6	94.2
2	71.8	84.9	72.3	85.9	34.2	94.9
1	74.4	83.8	75.0	84.8	32.6	95.4
0.5	77.3	82.8	78.0	83.8	31.2	95.9
0.2	81.4	81.3	81.5	82.6	27.1	97.1

estimated regression parameter values (Eng and others, 2009). Stations may have high leverage and high influence, high leverage and low influence, or low leverage and high influence. For this study, influence is calculated using a generalized Cook's D value (Eng and others, 2009). High influence may indicate an error in either the station record or basin characteristics at a station, but if no such errors exist, it alone is not sufficient justification for removing a station from the regression analysis.

Streamgaging stations used in the regression equations generally show uniform influence and leverage, although there are consistently a few stations with relatively high influence (appendix 3). The stations with high influence are not consistent across all flood durations or AEPs. Only two stations, Cibecue No. 1, Tributary to Carrizo Creek, near Show Low, Arizona (station 09496600), and Cibecue No. 2, Tributary to Carrizo Creek, near Show Low, Arizona (station 09496700), were completely removed from the regression analysis because they showed large influence. The watershed area for both of these stations, 0.1 square miles, is much smaller than any other stations in the regression, and therefore they have undue influence on the regression results for other small watersheds. The next largest watershed area, 1.1 square miles at South Fork Parker Creek near Roosevelt, Arizona (station 09498503), should be considered the lower limit of applicability of the regional regression equations. One or two other stations, as shown in appendix 3, were selectively removed from the regression for individual *n*-day flood-duration flows and (or) AEPs. These stations were determined to have poor quantile predictions, based on the quantile-probability plots, most often because of short record lengths.

The final regression equation exponents for each basin characteristic show that flood volume increases with increasing DRNAREA and PRECIP, and decreases with ELEV (table 9). The relative importance of drainage area and precipitation generally is higher for flood volumes with higher AEPs, and decreases with decreasing AEP. Conversely, the relative importance of elevation increases with decreasing AEP. *AVP* for the regression equations is generally low, ranging from 27 percent to 53 percent (table 10). The average for all *n*-day flood-duration flows and all AEPs is 38 percent. For comparison, *AVP* for the central highland region regression equations in the peak flow analysis (Paretti and others, 2014b) ranges from 57 to 91 percent. The lower *AVP* values for this study reflect decreased variability in *n*-day flood-duration flows as compared to instantaneous peak flow. Pseudo-R^2 can range from 0 to 100, with higher values indicating better performance. As with *AVP*, the Pseudo-R^2 values for the regression equations in this study, which range from 90.0 to 97.1 percent (table 10), indicate relatively good model performance. Model performance improves slightly for longer duration *n*-day flood-duration flows, as indicated by lower *AVP* and higher Pseudo-R^2.

The following limitations apply when using the final regional regression equations:

1. Applying the equations to sites on streams having explanatory variables outside the ranges of those used in this study (table 8) may result in prediction errors that are considerably greater than those indicated by the standard error of prediction percentages listed in table 10.

2. The methods are not appropriate (or applicable) for sites where flood-duration flows are affected substantially by flow regulation.

3. The methods are not appropriate (or applicable) for streams in urban areas with substantial impervious area unless the effects of urbanization are deemed insignificant.

Weighting Estimates at Streamgaging Stations

Flood-frequency estimates at a streamgaging station, particularly stations with short records, can be improved by taking the weighted average of the station estimate and the estimate from the regional regression equations (Cohn and others, 2012). The weighting is inversely proportional to the

Table 8. Statistics of the basin characteristics used for the regression equations in estimating flood-duration flow in Arizona.

Characteristic	Minimum	Maximum	Mean	Median
Drainage area, square miles	1.1	7,888.3	742.3	195.4
Basin centroid elevation, feet	2,240.5	9,520.3	5,750.3	5,543.0
Mean annual precipitation, inches	13.9	36.7	23.6	22.7

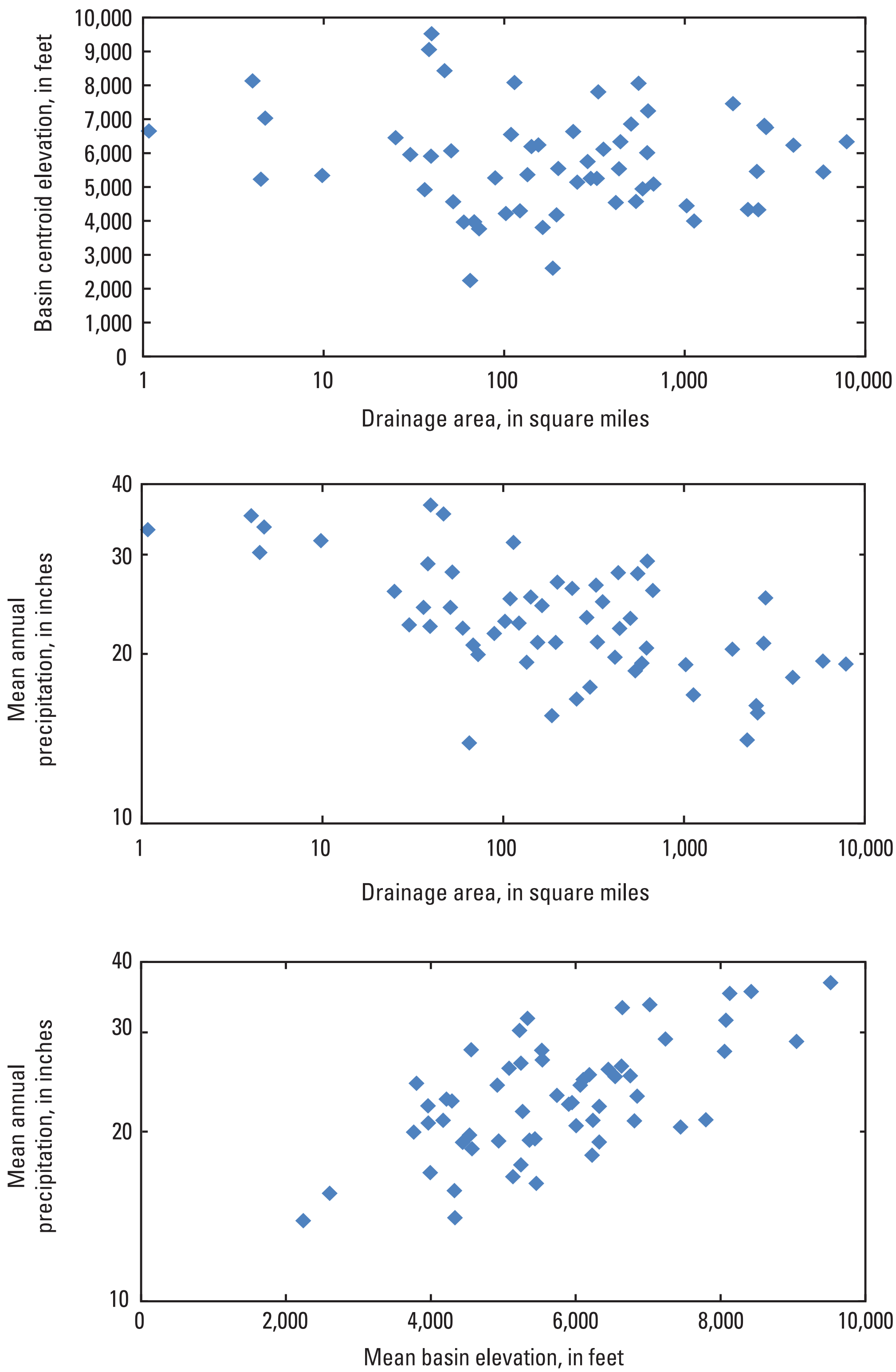


Figure 10. Scatterplots showing joint distributions of basin characteristics used for estimating flood-duration flow for Arizona.

Table 9. Regression equations for 1-, 3-, 7-, 15-, and 30-day flood duration flows for the central highland region in Arizona (see fig. 9).

[Pct. AEP, percent annual exceedance probability; DRNAREA, drainage areain square miles; PRECIP, mean annual precipitation in inches. ELEV, mean basin elevation in feet]

Pct. AEP	Regression equation
	1 day
50	$0.00759\ (DRNAREA)^{0.882}(PRECIP)^{2.454}10^{(-0.095*ELEV/1,000)}$
20	$0.0692\ (DRNAREA)^{0.836}(PRECIP)^{2.310}10^{(-0.128*ELEV/1,000)}$
10	$0.189\ (DRNAREA)^{0.808}(PRECIP)^{2.233}10^{(-0.131*ELEV/1,000)}$
4	$0.240\ (DRNAREA)^{0.781}(PRECIP)^{2.422}10^{(-0.136*ELEV/1,000)}$
2	$0.619\ (DRNAREA)^{0.765}(PRECIP)^{2.278}10^{(-0.138*ELEV/1,000)}$
1	$1.50\ (DRNAREA)^{0.751}(PRECIP)^{2.132}10^{(-0.139*ELEV/1,000)}$
0.5	$3.44\ (DRNAREA)^{0.739}(PRECIP)^{1.988}10^{(-0.140*ELEV/1,000)}$
0.2	$30.1\ (DRNAREA)^{0.700}(PRECIP)^{1.503}10^{(-0.144*ELEV/1,000)}$
	3 day
50	$0.00597\ (DRNAREA)^{0.875}(PRECIP)^{1.978}$
20	$0.0127\ (DRNAREA)^{0.868}(PRECIP)^{2.516}10^{(-0.101*ELEV/1,000)}$
10	$0.0524\ (DRNAREA)^{0.847}(PRECIP)^{2.360}10^{(-0.121*ELEV/1,000)}$
4	$0.173\ (DRNAREA)^{0.826}(PRECIP)^{2.285}10^{(-0.144*ELEV/1,000)}$
2	$0.568\ (DRNAREA)^{0.812}(PRECIP)^{2.081}10^{(-0.152*ELEV/1,000)}$
1	$1.68\ (DRNAREA)^{0.800}(PRECIP)^{1.882}10^{(-0.158*ELEV/1,000)}$
0.5	$4.61\ (DRNAREA)^{0.790}(PRECIP)^{1.688}10^{(-0.163*ELEV/1,000)}$
0.2	$23.6\ (DRNAREA)^{0.753}(PRECIP)^{1.365}10^{(-0.165*ELEV/1,000)}$
	7 day
50	$0.000538\ (DRNAREA)^{0.916}(PRECIP)^{2.527}$
20	$0.00314\ (DRNAREA)^{0.877}(PRECIP)^{2.669}10^{(-0.074*ELEV/1,000)}$
10	$0.00820\ (DRNAREA)^{0.871}(PRECIP)^{2.719}10^{(-0.118*ELEV/1,000)}$
4	$0.0267\ (DRNAREA)^{0.847}(PRECIP)^{2.672}10^{(-0.147*ELEV/1,000)}$
2	$0.180\ (DRNAREA)^{0.816}(PRECIP)2^{.288}10^{(-0.161*ELEV/1,000)}$
1	$0.298\ (DRNAREA)^{0.816}(PRECIP)^{2.246}10^{(-0.168*ELEV/1,000)}$
0.5	$0.877\ (DRNAREA)^{0.803}(PRECIP)^{2.041}10^{(-0.175*ELEV/1,000)}$
0.2	$3.24\ (DRNAREA)^{0.788}(PRECIP)^{1.787}10^{(-0.183*ELEV/1,000)}$
	15 day
50	$0.0000440\ (DRNAREA)^{0.958}(PRECIP)^{3.121}$
20	$0.000508\ (DRNAREA)^{0.908}(PRECIP)^{3.006}10(^{-0.065*ELEV/1,000)}$
10	$0.00209\ (DRNAREA)^{0.884}(PRECIP)^{2.880}10^{(-0.094*ELEV/1,000)}$
4	$0.00652\ (DRNAREA)^{0.860}(PRECIP)^{2.865}10^{(-0.129*ELEV/1,000)}$
2	$0.0217\ (DRNAREA)^{0.844}(PRECIP)^{2.678}10^{(-0.144*ELEV/1,000)}$
1	$0.0668\ (DRNAREA)^{0.829}(PRECIP)^{2.490}10^{(-0.157*ELEV/1,000)}$
0.5	$0.192\ (DRNAREA)^{0.816}(PRECIP)^{2.305}10^{(-0.168*ELEV/1,000)}$
0.2	$1.20\ (DRNAREA)^{0.808}(PRECIP)^{1.857}10^{(-0.172*ELEV/1,000)}$
	30 day
50	$0.00000789\ (DRNAREA)^{0.978}(PRECIP)^{3.519}$
20	$0.000512\ (DRNAREA)^{0.889}(PRECIP)^{2.637}$
10	$0.000361\ (DRNAREA)^{0.903}(PRECIP)^{3.208}10^{(-0.078*ELEV/1,000)}$
4	$0.000897\ (DRNAREA)^{0.882}(PRECIP)^{3.255}10^{(-0.113*ELEV/1,000)}$
2	$0.00261\ (DRNAREA)^{0.868}(PRECIP)^{3.103}10^{(-0.129*ELEV/1,000)}$
1	$0.00716\ (DRNAREA)^{0.855}(PRECIP)^{2.942}10^{(-0.141*ELEV/1,000)}$
0.5	$0.0187\ (DRNAREA)^{0.843}(PRECIP)^{2.778}10^{(-0.152*ELEV/1,000)}$
0.2	$0.111\ (DRNAREA)^{0.837}(PRECIP)^{2.327}10^{(-0.154*ELEV/1,000)}$

Table 10. Average variance of prediction, average standard error of prediction, and pseudo-R^2 for the regression equations used to predict 1-, 3-, 7-, 15-, and 30-day flood-duration flow in the central highland region of Arizona.

[Pct. AEP, percent annual exceedance probability; AVP, average variance of prediction; Avg. SEP, average standard error of prediction]

Pct. AEP	AVP (log units)	Avg. SEP (percent)	Pseudo-R^2 (percent)
1 day			
50	0.043	0.043	90.4
20	0.033	0.033	92.1
10	0.029	0.029	92.7
4	0.024	0.024	93.5
2	0.024	0.024	93.6
1	0.024	0.024	93.5
0.5	0.025	0.025	93.1
0.2	0.025	0.025	93.1
3 day			
50	0.046	52.3	90.0
20	0.033	43.6	92.4
10	0.028	40.4	93.3
4	0.028	40.3	93.2
2	0.027	39.3	93.6
1	0.028	39.9	93.5
0.5	0.029	40.6	93.4
0.2	0.022	35.3	95.2
7 day			
50	0.047	52.9	90.2
20	0.032	42.7	92.6
10	0.023	36.1	94.5
4	0.022	34.9	94.7
2	0.019	32.6	95.4
1	0.020	33.1	95.4
0.5	0.020	33.6	95.3
0.2	0.022	35.2	95.0
15 day			
50	0.044	51.3	91.2
20	0.029	40.4	93.6
10	0.023	36.4	94.5
4	0.022	34.9	94.7
2	0.019	32.6	95.4
1	0.018	32.0	95.6
0.5	0.018	31.6	95.8
0.2	0.015	28.7	96.9
30 day			
50	0.041	49.5	92.0
20	0.032	42.7	92.9
10	0.026	38.2	94.1
4	0.024	36.6	94.2
2	0.021	34.2	94.9
1	0.019	32.6	95.4
0.5	0.017	31.2	95.9
0.2	0.013	27.1	97.1

variance of the regression and station estimates (VP_r and VP_s, respectively, in log units) so that estimates with greater uncertainty have less weight in the weighted average. The variance of the regression estimate is calculated differently depending on if the site was used to develop the regional regression equations. For sites not used in the regional regression equations, VP_r is equivalent to AVP and calculated using equation 6. For individual sites that are used in the regional regression equations, VP_r is:

$$VP_r = \sigma_\delta^2 + \mathbf{x}_i\,(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\mathbf{X})^{-1}\mathbf{x}_i^T \qquad (10)$$

The matrix is provided as output from WREG. A first-order approximation of VP_s is output for each AEP from the PeakFQ software used to fit the LP3 distribution (Cohn and others, 2001). The weighted average for a particular AEP is then:

$$\log(\hat{Q}) = \frac{VP_r \times \log(Q_s) + VP_s \times \log(Q_r)}{VP_r + VP_s} \qquad (11)$$

For the stations within the central highland region for which the regional regression equations apply (fig. 9, table 8), the station estimate, regression estimate, and weighted estimates of discharge for varying AEPs are presented in appendix 1. The variance of prediction associated with the weighted estimate, VP_w, is computed as:

$$VP_w = \frac{VP_s VP_r}{VP_s + VP_r} \qquad (12)$$

VP estimates are given in appendix 2 for the station estimate, regression estimate, and weighted estimate of the predicted *n*-day flood-duration flows.

Estimates Near Streamgaging Stations on the Same Stream

Within the central highland region, if an ungaged site is near an existing streamgaging station for which flood-duration flow frequency statistics have been calculated, a weighted average flood-duration flow may be calculated that incorporates that station explicitly, rather than using only the regression equations (Ries and Crouse, 2002). Generally, "near" is defined as having a drainage area between 50 and 150 percent of that at the streamgaging station. First, the estimated flood-duration flow at the ungaged site is determined by comparing the drainage area to that at the streamgaging station:

$$Q_u = \left(\frac{\log(A_u)}{\log(A_g)}\right)^b \times Q_s \tag{13}$$

where Q_u is the area-weighted flood-duration flow estimate at the ungaged site, Q_s is the station estimate before weighting with the regional regression estimate (appendix 2), and A_u and A_g are the areas of the ungaged and gaged drainage areas, respectively. The exponent b is the exponent of the drainage area variable in the regional regression equation (DRNAREA; table 9).

After calculating the expected flood-duration flow at the ungaged site based on the drainage area ratio, it can be combined with the regional regression equation:

$$Q_{u(w)} = \left[\left(\frac{2\Delta A}{A_g}\right)Q_{u(r)} + \left(1 - \frac{2\Delta A}{A_g}\right)Q_u\right] \tag{14}$$

where $Q_{u(w)}$ is the weighted estimate of flood-duration flow at the ungaged site, ΔA is the absolute value of the difference between the drainage areas of the streamgaging station and the ungaged site, $|A_g - A_u|$, and $Q_{u(r)}$ is the flood-duration flow estimate for the ungaged site derived from the applicable regional equation (table 9).

Unlike the procedure for calculating the weighted average at streamgaging stations, the procedure for stations near a streamgaging station does not take into account the length of the streamgaging record. If the nearby streamgaging record is short (less than about 20 years) and the difference in drainage areas is large, the estimated flood-duration flow at streamgaging station may be excessively weighted in equation 14. In this case, the method for ungaged sites may produce better estimates of flood-duration flow.

If an ungaged site lies between two streamgaging stations on the same stream, the weighted average of the predicted flood volume at the two stations may be calculated, incorporating the relative distance of the ungaged site between the two stations. Major tributaries and (or) nonlinear variation in drainage area should be accounted for, and consideration given to the length of record, and therefore uncertainty, at each station. In areas of distributary flow, this method may not be appropriate.

Summary and Conclusions

This report presents n-day flood-duration flow frequency estimates at 173 streamgaging stations in Arizona and western New Mexico. These estimates are valuable for the design of runoff detention and retention structures and also for water-resources planning. For short duration flood flows (1-day and 3-day flood-duration flows), flood-duration flows and flood peaks are generally correlated (that is, a streamgaging station having a relatively large peak flow will also have high 1-day and 3-day flood-duration flows). However, longer duration flood flows, such as the 15-day and 30-day flood-duration flows, may have little or no correlation with peak flows. Ephemeral streams in particular, and especially those at which peak flows are generally caused by short-lived convective thunderstorms, may have large peak flows that pass quickly and have relatively low flood volume.

The expected moments algorithm with a multiple Grubbs-Beck low-outlier test was used to estimate annual maximum n-day flood-duration flow at the 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent annual exceedance probabilities for 1-, 3-, 7-, 15-, and 30-day durations at streamgaging stations throughout Arizona. A Bayesian generalized least squares regression analysis of regional skewness coefficients indicates that no basin characteristics were significant explanatory variables of skew, and a constant statewide model for each high-flow duration was used. Skew becomes increasingly negative with increasing duration of n-day flood-duration flow. Variance of the skewness coefficients also increases with increasing duration.

Flood-duration flow frequency estimates are presented at stations with 10 or more years of continuous streamflow data, unaffected by impoundments and urbanization, and adequately represented by the LP3 distribution. Not surprisingly, most stations are established near major population centers (often as flood-warning gages) and on major stream channels. As a result, gages with continuous streamgaging records, required for the flood-duration flow analysis, are most common in the central part of the State, especially on the streams that flow south and east from the Colorado Plateau, including the Verde, Salt, and Gila Rivers and their tributaries. This central highland region is the only part of the State where sufficient information exists to develop regression equations for ungaged sites. Outside this region, high-flow frequency estimates may still be calculated at streamgaging stations and for ungaged sites between two streamgaging stations.

The regression analysis in the central highland region indicates that drainage area, mean annual precipitation, and mean basin elevation are all determining factors when predicting the frequency of flood-duration flows. Regression verification statistics are generally good, with an average standard error of prediction that varies from 28 to 53 percent. Pseudo-R^2 varies from 90 to 97 percent. The flood-duration frequency regression equations developed in this study are available through the USGS StreamStats program, a Web-based application that provides streamflow statistics and basin characteristics for USGS streamgaging stations and ungaged sites of interest.

References Cited

Balocki, J.B., and Burges, S.J., 1994, Relationships between *n*-day flood volumes for infrequent large floods: Journal of Water Resources Planning and Management, v. 120, no. 6, p. 794–818.

Bohman, L.R., 1990, Determination of flood hydrographs for streams in South Carolina—Volume 1, simulation of flood hydrographs for rural watersheds in South Carolina: U.S. Geological Survey Water-Resources Investigations Report 89–4087, 53 p.

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

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

Cohn, T.A., Berenbrock, C., Kiang, J.E., and Mason Jr., R.R., 2012, Calculating weighted estimates of peak streamflow statistics: U.S. Geological Survey Fact Sheet 2012–3038, 4 p., available at http://pubs.usgs.gov/fs/2012/3038/.

Cunderlik, J.M., and Ouarda, T.B.M.J., 2006, Regional flood-duration–frequency modeling in the changing environment: Journal of Hydrology, v. 318, p. 276–291.

Dalrymple, T., 1960, Flood-frequency analyses, Manual of Hydrology, part 3: U.S. Geological Survey Water Supply Paper 1543–A, 80 p.

Devulapalli, R.S., and Valdes, J.N., 1996, Volume-duration-frequencies for ungagged catchments in Texas—Volume I, Calculation of regional regression equations: Technical Report No. 173, Texas Water Resources Institute, College Station, Texas, 332 p.

Eng, K., 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, chapter A8, 21 p., available at http://pubs.usgs.gov/tm/tm4a8/.

England J.F., Jr., Salas, J.D., and Jarret, R.D., 2003, Comparisons of two moments-based estimators that utilize historical and paleoflood data for the log Pearson type III distribution: Water Resources Research, v. 39, no. 9, 1243.

Eychaner, J.H., 1976, Estimating runoff volumes and flood hydrographs in the Colorado River Basin, southern Utah: U.S. Geological Survey Water-Resources Investigations 76–102, 18 p.

Fenneman, N.M., and Johnson, D.W., 1946, Physical divisions of the United States: U.S. Geological Survey, scale 1:7,000,000, 1 sheet.

Flynn, K.M., Kirby, W.H., and Hummel, P.R., 2006, User's manual for program PeakFQ, annual flood-frequency analysis using Bulletin 17B guidelines: U.S. Geological Survey, Techniques and Methods, book 4, chapter B4, 52 p.

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, no. 10, W07503.

Griffis, V.W., and Stedinger, J.R., 2007, The use of GLS regression in regional hydrologic analyses: Journal of Hydrology, v. 344, p. 82–95.

Gotvald, A.J., Barth, N.A., Veilleux, A.G., and Parrett, Charles, 2012, Methods for determining magnitude and frequency of floods in California, based on data through water year 2006: U.S. Geological Survey Scientific Investigations Report 2012–5113, 38 p., available at http://pubs.usgs.gov/sir/2012/5113/.

Grubbs, F.E., and Beck, G., 1972, Extension of sample sizes and percentage points for significance tests of outlying observations: Technometrics, v. 14, no. 4, p. 847–854.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources techniques of water resources investigations, book 4, chapter A3: U.S. Geological Survey, 522 p., available at http://pubs.usgs.gov/twri/twri4a3/.

Interagency Advisory Committee on Water Data [IACWD], 1982, Guidelines for determining flood-flow frequency, Bulletin 17B of the Hydrology subcommittee, Office of Water Data Coordination: U.S. Geological Survey, 183 p.

Javelle, P., Ouarda, T.B.M.J., and Bobée, B., 2003, Spring flood analysis using the flood-duration–frequency approach—Application to the provinces of Quebec and Ontario, Canada: Hydrological Processes, v. 17, p. 3717–3736.

Johnston, J., 1972, Econometric methods: McGraw-Hill, New York, 437 p.

Kennedy, J.R., and Gungle, B., 2010, Quantity and sources of base flow in the San Pedro River near Tombstone, Arizona: U.S. Geological Survey Scientific Investigations Report 2010–5200, 43 p., available at http://pubs.usgs.gov/sir/2010/5200/.

Kroll, C.N., and Song, P., 2013, Impact of multicollinearity on small sample hydrologic regression models: Water Resources Research, v. 49, doi:10.1002/wrcr.20315.

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., available at http://pubs.usgs.gov/sir/2012/5130/.

Martins, E.S., and Stedinger, J.R., 2002, Cross-correlation among estimators of shape: Water Resources Research, v. 38, no. 11, p. 1252–1259.

Mediero, L., Jiménez-Álvarez, A., and Garrote, L., 2010, Design flood hydrographs from the relationship between flood peak and volume: Hydrology and Earth System Sciences, v. 14, p. 2495–2505, doi:10.5194/hess-14-2495-2010.

Natural Resources Conservation Service, 2013, U.S. general soil map (STATSGO2) for Arizona: Natural Resources Conservation Service, United States Department of Agriculture, accessed April 22, 2013, available at http://soildatamart.nrcs.usda.gov.

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., available at http://pubs.usgs.gov/sir/2010/5260/.

Paretti, N.V., Kennedy, J.R., and Cohn, T.A., 2014a, Evaluation of the expected moments algorithm and a multiple low-outlier test for flood frequency analysis at streamgaging stations in Arizona: U.S. Geological Survey Scientific Investigations Report 2014–5026, 61 p., available at http://pubs.usgs.gov/sir/2014/5026/.

Paretti, N.V., Kennedy, J.R., Turney, L.A., and Veilleux, A.G., 2014b, Methods for estimating magnitude and frequency of floods in Arizona, developed with unregulated and rural peak-flow data through water year 2010: U.S. Geological Survey Scientific Investigations Report 2014–5211, 55 p., http://pubs.usgs.gov/sir/2014/5211/.

Perry, C.A., 1984, A method of estimating flood volumes in western Kansas: U.S. Geological Survey Water-Resources Investigations Report 84–4164, 22 p.

PRISM Climate Group, 2012, PRISM Climate Group, Northwest Alliance for Computational Science and Engineering: Oregon State University, Web site, accessed December 1, 2012, available at http://prism.oregonstate.edu.

Reis, D.S., Stedinger, J.R., and Martins, E.S., 2005, Bayesian generalized least squares regression with application to log Pearson type 3 regional skew estimation: Water Resources Research, v. 41, W10419, 14 p.

Ries, K.G., III, and Crouse, M.Y., 2002, The national flood frequency program, Version 3—A computer program for estimating magnitude and frequency of floods for ungaged sites: U.S. Geological Survey Water-Resources Investigations Report 02–4168, 42 p.

Ries, K.G., III, Guthrie, J.D., Rea, A.H., Steeves, P.A., and Stewart, D.W., 2008, StreamStats—A water resources web application: U.S. Geological Survey Fact Sheet 2008–3067, 6 p., available at http://pubs.usgs.gov/fs/2008/3067/.

Sheppard, P.R., Comrie, A.C., Packin, G.D., Angersbach, K., and Hughes, M.K., 2002, The climate of the US Southwest: Climate Research, v. 21, p. 219–238.

Sherwood, J.M., 1994, Estimation of peak-frequency relations, flood hydrographs, and volume-duration-frequency relations of ungagged small urban streams in Ohio: U.S. Geological Survey Water-Supply Paper 2432, 41 p.

Singh, V.P., and Hossein, A., 1986, An empirical relation between flood volume and peak of direct runoff: Water Resources Bulletin, v. 22, no. 5, p. 725–730.

Stedinger, J.R., and Griffis, V.W., 2008, Flood frequency analysis in the United States—Time to update: Journal of Hydrologic Engineering, v. 13, p. 199–204.

Thomas, B.E., and Pool, D.R., 2006, Trends in streamflow of the San Pedro river, southeastern Arizona, and regional trends in precipitation and streamflow in southeastern Arizona and southwestern New Mexico: U.S. Geological Survey Professional Paper 1712, 92 p., available at http://pubs.usgs.gov/pp/pp1712/.

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., available at http://pubs.usgs.gov/sir/2009/5158/.

Yue, S., and Rasmussen, P., 2002, Bivariate frequency analysis: discussion of some useful concepts in hydrological application: Hydrological Processes, v. 16, p. 2881–2898.

This page left intentionally blank.

Menlo Park Publishing Service Center, California
Manuscript approved for publication June 5, 2014
Edited by James W. Hendley II
Design and layout by Vivian Nguyen

USGS

Kennedy and others—**Estimating Magnitude and Frequency of 1-, 3-, 7-, 15-, and 30-day flood-duration flows in Arizona**—Scientific Investigations Report 2014–5109, ver. 1.1

ISSN 2328-0328 (online)
http://dx.doi.org/10.3133/sir20145109