Prepared in cooperation with the Federal Emergency Management Agency, the U.S. Army Corps of Engineers, and the U.S. Forest Service

Regional Skew for California, and Flood Frequency for Selected Sites in the Sacramento-San Joaquin River Basin, Based on Data through Water Year 2006

Scientific Investigations Report 2010–5260

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

Cover: Photogrph showing flooding on the South Yuba River on January 1, 2006 at the Old Highway 49 crossing near Nevada City, California. Photograph was taken by Ian O'Halloran.

Regional Skew for California, and Flood Frequency for Selected Sites in the Sacramento–San Joaquin River Basin, Based on Data through Water Year 2006

By Charles Parrett, U.S. Geological Survey, Andrea Veilleux and J.R. Stedinger, Cornell University; N. A. Barth, Donna L. Knifong, and J.C. Ferris, U.S. Geological Survey

Prepared in cooperation with Federal Emergency Management Agency, U.S. Army Corps of Engineers, and U.S. Forest Service

Scientific Investigations Report 2010–5260

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

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2011

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

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

To order this and other USGS information products, visit http://store.usgs.gov

Suggested citation:
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.

Contents

Figures

Tables

Conversion Factors, Datums, and Acronyms

Inch/Pound to SI

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

SI to Inch/Pound

Multiply	By	To obtain
	Length	
centimeter (cm)	0.3937	inch (in.)
millimeter (mm)	0.03937	inch (in.)
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
	Area	
square centimeter (cm^2)	0.001076	square foot (ft^2)
square meter (m^2)	10.76	square foot (ft^2)
square centimeter (cm^2)	0.1550	square inch (ft^2)
square kilometer (km^2)	0.3861	square mile (mi^2)
	Volume	
square centimeter (cm^2)	0.03531	cubic foot (ft^3)
square meter (m^2)	35.31	cubic foot (ft^3)
square centimeter (cm^2)	1.308	cubic yard (yd^3)
square kilometer (km^2)	0.2399	cubic mile (mi^3)
	Flow rate	
cubic meter per second (m^3/s)	35.31	cubic foot per second (ft^3/s)

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

$$°F=(1.8\times°C)+32$$

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

$$°C=(°F-32)/1.8$$

Conversion Factors, Datums, and Acronyms—Continued

Datums

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

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

Elevation refers to distance above or below NAVD 88.

Water year is the 12-month period, October 1 through September 30, and is designated by the calendar year in which it ends. Thus, the water year ending September 30, 2001 is called "water year 2001."

Acronyms

ACOE	Army Corps of Engineers
ANOVA	analysis of variance
ASEV	average sampling error variance
AVP	average variance of prediction
B-GLS	Bayesian generalized least squares
B-WLS	Bayesian weighted least squares
DAR	drainage area ratio
DWR	Department of Water Resources
ELEV	mean basin elevation
EMA	expected moments algorithm
ERL	effective record length
EVR	error variance ratio
FEMA	Federal Emergency Management Agency
GIS	geographic information system
GLS	generalized least squares
LP3	log Pearson Type 3
MBV	misrepresentation of the beta variance
MHDP	Multi-Hazards Demonstration Project
MM-WLS	method-of-moments weighted least squares
MOVE.1	maintenance of variance extension type I
MOVE.2	maintenance of variance extension type II
MOVE.3	maintenance of variance extension 3
MOVE.4	maintenance of variance extension 4
MSE	mean square error
ND	normalized distance
NHDPlus	National Hydrologic Dataset
NLCD	National Land-Cover Dataset
NL-Elev	nonlinear regional skew model
NWIS	National Water Information System
PRISM	Parameter-Elevation Regressions on Independent Slopes Model
USGS	U.S. Geological Survey
VP	variance of prediction
WLS	weighted least squares

Regional Skew for California and Flood Frequency for Selected Sites in the Sacramento–San Joaquin River Basin Based on Data through Water Year 2006

By Charles Parrett[1], Andrea Veilleux[2], J.R. Stedinger[2], N. A. Barth[1], Donna L. Knifong[1], and J.C. Ferris[1]

Abstract

Improved flood-frequency information is important throughout California in general and in the Sacramento–San Joaquin River Basin in particular, because of an extensive network of flood-control levees and the risk of catastrophic flooding. A key first step in updating flood-frequency information is determining regional skew. A Bayesian generalized least squares (GLS) regression method was used to derive a regional-skew model based on annual peak-discharge data for 158 long-term (30 or more years of record) stations throughout most of California. The desert areas in southeastern California had too few long-term stations to reliably determine regional skew for that hydrologically distinct region; therefore, the desert areas were excluded from the regional skew analysis for California. Of the 158 long-term stations used to determine regional skew, 145 have minimally regulated annual-peak discharges, and 13 stations are dam sites for which unregulated peak discharges were estimated from unregulated daily maximum discharge data furnished by the U.S. Army Corp of Engineers. Station skew was determined by using an expected moments algorithm (EMA) program for fitting the Pearson Type 3 flood-frequency distribution to the logarithms of annual peak-discharge data.

The Bayesian GLS regression method previously developed was modified because of the large cross correlations among concurrent recorded peak discharges in California and the use of censored data and historical flood information with the new expected moments algorithm. In particular, to properly account for these cross-correlation problems and develop a suitable regression model and regression diagnostics, a combination of Bayesian weighted least squares and generalized least squares regression was adopted. This new methodology identified a nonlinear function relating regional skew to mean basin elevation. The regional skew values ranged from -0.62 for a mean basin elevation of zero to 0.61 for a mean basin elevation of 11,000 feet. This relation between skew and elevation reflects the interaction of snow with rain, which increases with increased elevation. The equivalent record length for the new regional skew ranges from 52 to 65 years of record, depending upon mean basin elevation. The old regional skew map in Bulletin 17B, published by the Hydrology Subcommittee of the Interagency Advisory Committee on Water Data (1982), reported an equivalent record length of only 17 years.

The newly developed regional skew relation for California was used to update flood frequency for the 158 sites used in the regional skew analysis as well as 206 selected sites in the Sacramento–San Joaquin River Basin. For these sites, annual-peak discharges having recurrence intervals of 2, 5, 10, 25, 50, 100, 200, and 500 years were determined on the basis of data through water year 2006. The expected moments algorithm was used for determining the magnitude and frequency of floods at gaged sites by using regional skew values and using the basic approach outlined in Bulletin 17B.

Introduction

Reliable estimates of peak discharge for various exceedance frequencies, commonly referred to as flood-frequency estimates, are needed by engineers, land-use planners, resource managers, and scientists. Flood-frequency information is required for flood-hazard assessment, safe and cost-effective design of water conveyance and transportation structures in or near streams, and floodplain delineation for flood insurance and land-use management. Water Science Centers within the U.S. Geological Survey (USGS) commonly provide updated flood-frequency estimates for streamflow-gaging stations every 5 to 10 years based on cooperator needs and funding availability. The USGS provides methods for estimating flood-frequency at ungaged sites also. State-wide updates of flood-frequency at gaged sites and methods for estimation at ungaged sites in California were last completed

---

[1] U.S. Geological Survey, California Water Science Center, Placer Hall, 6000 J Street, Sacramento, California 95819

[2] Cornell University, School of Civil & Environmental Engineering, 220 Hollister Hall, Ithaca, New York 14853

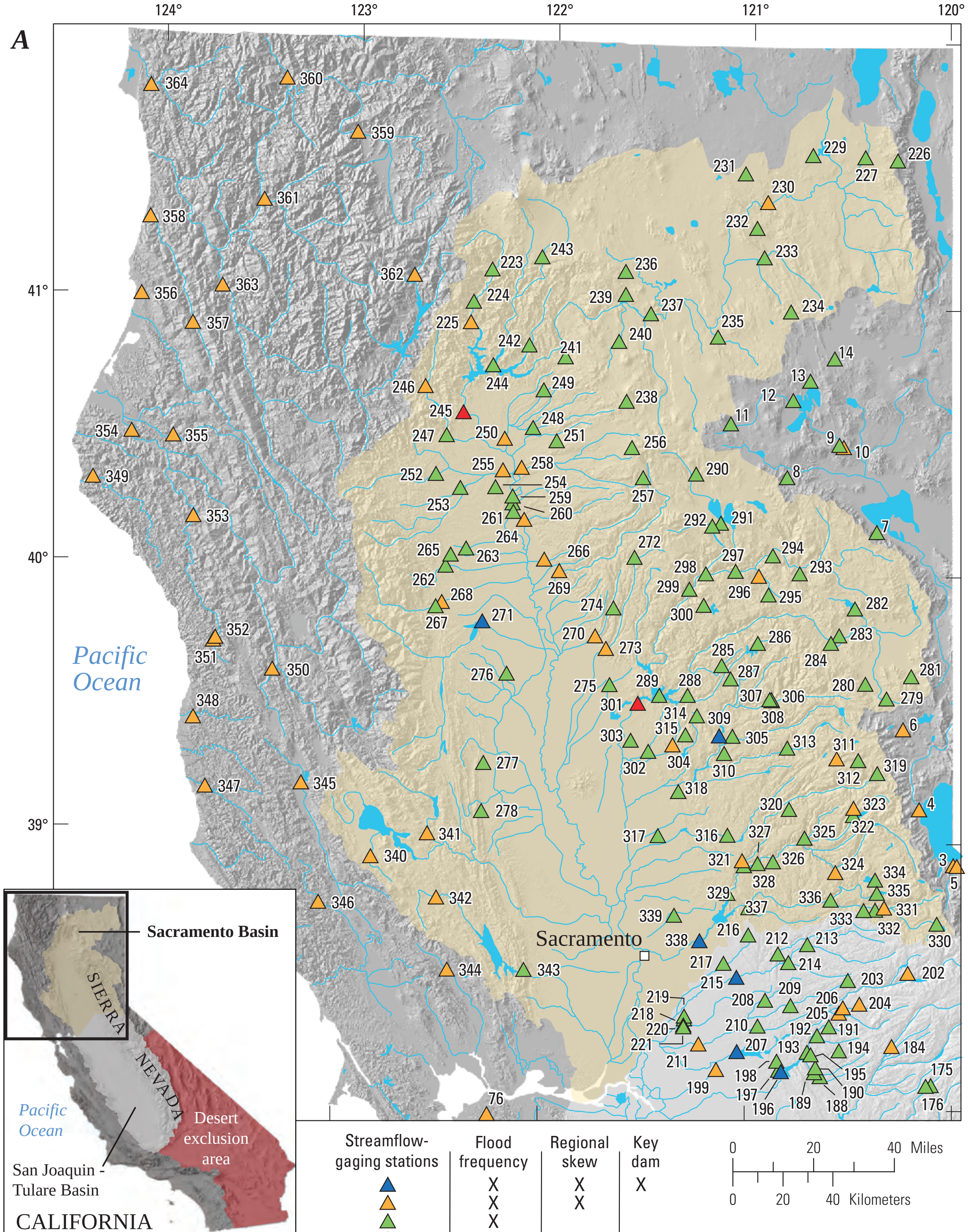
in 1977 (Waananen and Crippen, 1977). The additional 30 years of flow record and new analytical methods together with an increasing need for essential levee improvements and flood-frequency information due to increased population fully justify updating the frequency information. Accordingly, the USGS, in cooperation with the U.S. Forest Service, Federal Emergency Management Agency (FEMA), and the USGS Multi-Hazards Demonstration Project (MHDP), initiated a regional flood-frequency study for California in 2008. This comprehensive study consists of (1) updating flood-frequency information for all suitable USGS gages that have minimally regulated peak-discharge records for at least 10 years, (2) developing methodologies for estimating flood-frequency information for ungaged sites in California, and (3) implementing a StreamStats (Ries and others, 2008) application for California. The StreamStats web-page application will initially provide basin and climatic characteristics data and flood-estimation equations based on two earlier flood-frequency reports for California (Waananen and Crippen, 1977; Thomas and others, 1997). As new flood-estimation equations are developed for use throughout California, they will be incorporated into the California StreamStats web page.

Bulletin 17B of the Hydrology Subcommittee of the Interagency Advisory Committee on Water Data (1982), hereinafter referred to as Bulletin 17B, recommends that the log Pearson Type 3 probability distribution be used to estimate flood frequency at gaged sites. A key to accurately fitting this distribution to recorded flood data is to reliably estimate the shape or skewness of the distribution, which is often significantly affected by the presence of very small or very large discharges in the record (outliers). Accordingly, Bulletin 17B also recommends that at-site skew calculated from recorded data be weighted with regional skew determined from pooled data at nearby long-term sites. Recent studies described by Reis and others (2005), Weaver and others (2009), Feaster and others (2009), and Gotvald and others (2009) have shown that Bayesian generalized least squares (GLS) regression provides an effective statistical framework for estimating regional skew. The regional skewness estimators are more accurate and have smaller mean square errors than those attributed to the Bulletin 17B skew map. Thus, an important contribution of this study was the development of new regional skew relations for California using extensions to Bayesian GLS regression. A key first step in the regional skew analysis was determining at-site (station) skew values for selected long-term (30 or more years of peak-discharge record) stations. The desert areas in southeastern California had too few long-term stations to determine regional skew with reasonable reliability for that hydrologically distinct region, so the desert region shown on figure 1 was excluded from the regional skew analysis for California. Station skew was determined using a new expected moment algorithm (EMA) program for fitting the Pearson Type 3 flood-frequency distribution to the logarithms of annual peak-flow data developed by the USGS (Cohn and others, 1997, 2001; Griffis and others, 2004).

An area of particular interest and need for updated flood-frequency data is the combined Sacramento and San Joaquin River Basin in central California. Large population centers, including the capital city of Sacramento, Stockton, Modesto, and Fresno, and widespread suburban development between the cities are at least partially located on floodplains of the Sacramento and San Joaquin Rivers and their tributaries. An extensive network of levees provides flood protection, but many of the levees are in need of upgrades and rehabilitation. To ensure that levee upgrades are designed in accordance with the latest and most complete hydrologic information, the U.S. Army Corps of Engineers (ACOE), in cooperation with the California Department of Water Resources (DWR), has begun a hydrologic analysis of floodplain areas protected by the Federal-State levee system within the Sacramento and San Joaquin combined drainage basin. As part of their hydrologic analysis, the ACOE requires flood-frequency information for simulated unregulated peak discharges at stream sites where peak discharges are partially or completely regulated.

Because the USGS and ACOE share a common interest and need for updated flood-frequency information in central California, both agencies developed a secondary cooperative flood-frequency program. This program tasked the USGS with developing a method for estimating unregulated annual-peak discharge at 16 selected key dam sites in the Sacramento–San Joaquin River Basin and using the estimated annual-peak discharges at the sites as part of the regional skew analysis already begun by the USGS for all of California outside the southeastern desert region. Results from the regional skew analysis were to be used to update flood-frequency data also for the 16 key sites.

To provide updated flood-frequency in the timeliest manner possible for the area of special interest in central California, the USGS focused initially on updates at gaged sites in a study area that consists of the Sacramento and San Joaquin-Tulare Basins. This study area (fig. 1) generally conforms to the Sacramento–San Joaquin River Basin study area defined by the ACOE. For consistency, the study area for the flood-frequency determinations in this report will be hereinafter referred to as the Sacramento–San Joaquin River Basin.



Streamflow-gaging stations	Flood frequency	Regional skew	Key dam
Blue triangle	X	X	X
Orange triangle	X	X	
Green triangle	X		
Red triangle	X		X

Figure 1. Stations in northern (*A*), central (*B*), and southern California (*C*) selected for regional skew analysis outside the desert exclusion area and flood-frequency analysis in the Sacramento-San Joaquin River Basin, California.

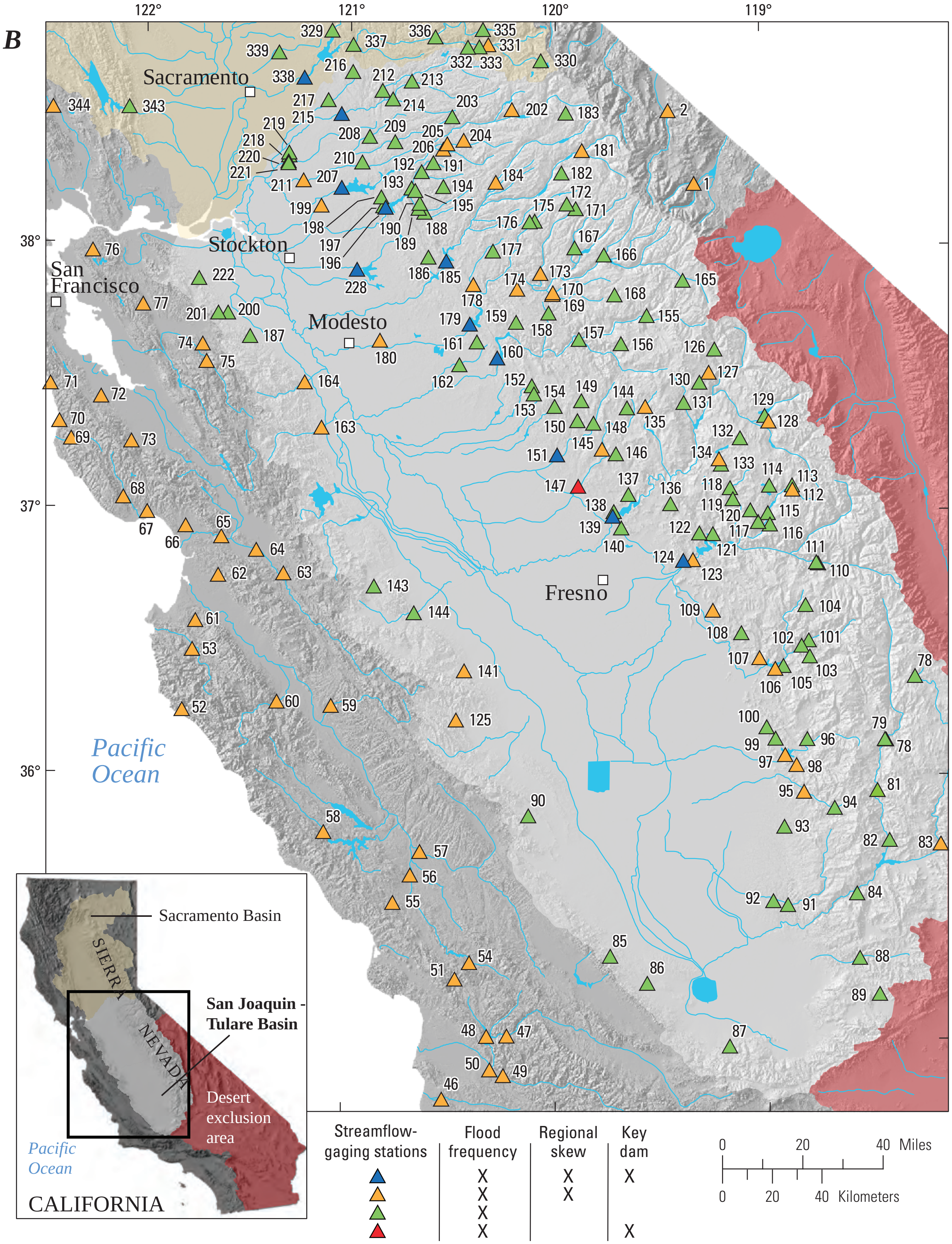


Figure 1.—Continued

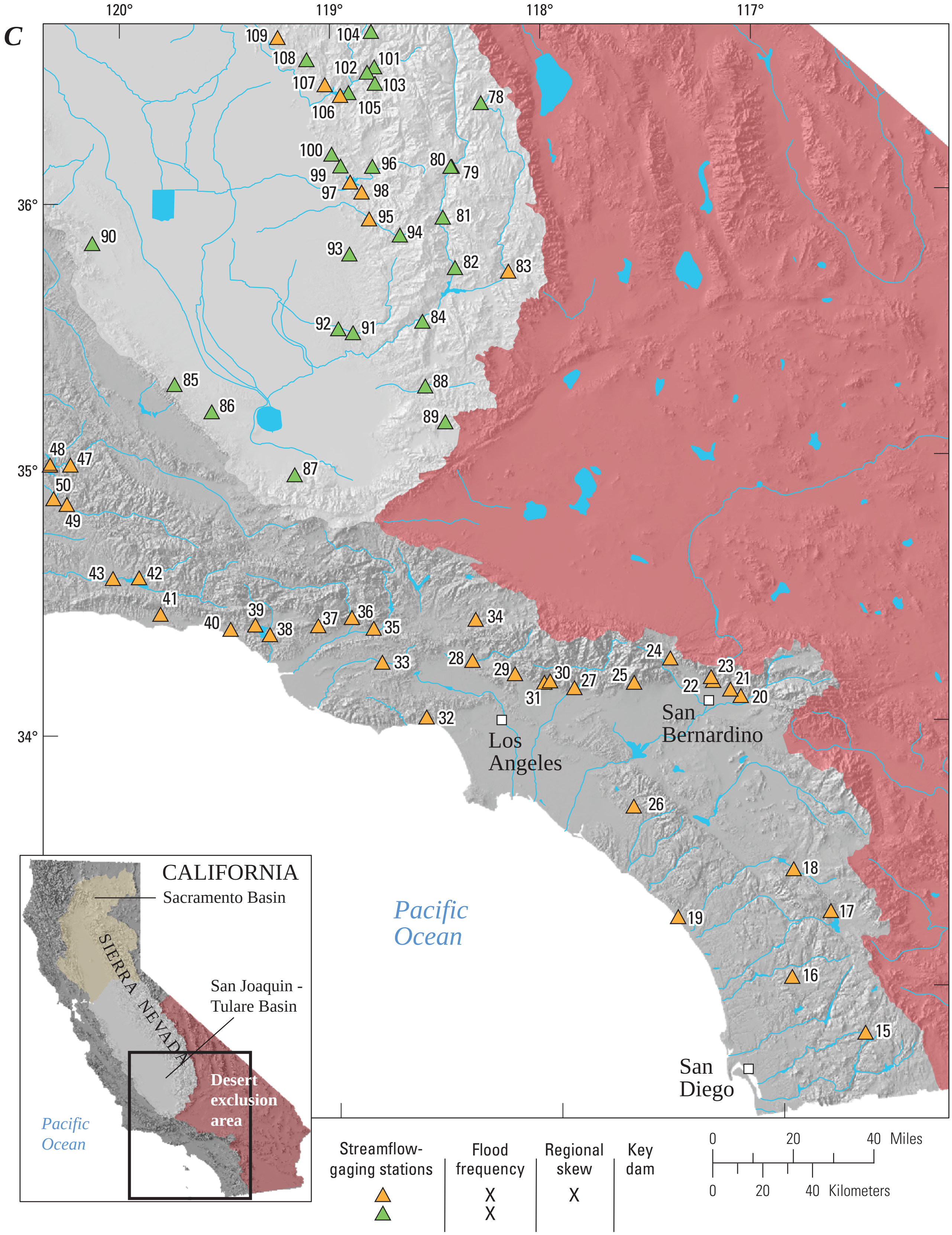


Figure 1.—Continued

Purpose and Scope

The primary purposes of this report are to (1) convey the results of the regional skew analysis for California, outside the southeastern desert region, and (2) present flood-frequency information for the 158 sites used in the regional skew analysis and for an additional 206 selected sites in the Sacramento-San Joaquin River Basin. The databases for annual-peak discharge used for the regional skew analysis and the determination of flood frequency at selected sites are described. Because key dam sites identified by the ACOE were included in the regional skew analysis, unregulated, annual-peak discharge data at those sites needed to be constructed from estimated unregulated annual-maximum-daily discharge data. The maintenance of variance extension (MOVE) method was used to estimate peak-discharge data at the key dam sites and is described in Appendix A. The new EMA method was used to compute moments of the logarithms for the LP3 distribution to determine a station skew at each site to use in the regional skew analysis and to subsequently update flood frequency at all selected sites. Finally, the Bayesian GLS regression method used for the regional skew analysis is described in some detail, and extensions to the method required for use in this California study and various diagnostics for analyzing regional skew results are presented in Appendixes B and C.

Updated flood-frequency data at selected sites in the study area, based on the new regional skew relations and applying the LP3 method, are presented in table format, and some example flood-frequency curves are shown. Specific flood-frequency information provided are peak discharges having annual exceedance probabilities (frequencies) of 0.50, 0.20, 0.10, 0.05, 0.02, 0.01, 0.005, and 0.002. Exceedance probabilities often are expressed in terms of their reciprocals as recurrence intervals. A peak discharge having an annual exceedance probability of 0.01, for example, has an associated recurrence interval of 100 years. Data in this report are presented both in terms of annual exceedance probability and recurrence interval.

Study Area Description

The study area for the regional skew analysis for California consists of the entire state outside the southeastern desert region (fig. 1). The excluded desert region largely conforms to the desert regions previously shown for California in a flood-frequency study for the southwestern desert region of the United States (Thomas and others, 1997). That previous study used data from several states to determine regional skew for the desert. The regional skew adopted by Thomas and others (1997) for the desert regions of the southwestern United States was zero.

Streamflow-gaging stations used in the regional skew analysis provided data for the broad range of hydrologic conditions throughout the study area. Along the California coast, streams drain the moderately rugged mountains of the Coastal Range, and annual-peak discharge most often results from large winter rainstorms. Annual-peak discharge from small streams, particularly those in drier areas of California may occasionally result from summer thunderstorms. Drainage within the flat valley floor of the Sacramento–San Joaquin River Basin is diffuse and often unpredictable because of the flat topography and agricultural land use, including extensive irrigation withdrawals and canal systems. Floods on the generally small streams that drain only the low-elevation foothills and valley floor areas commonly are the result of large winter rain storms, although floods may occasionally be a result of infrequent spring and summer rainstorms.

About a third of the stream sites selected for the regional skew analysis and many of the additional sites selected for the flood-frequency analysis in the Sacramento–San Joaquin River Basin are in the Sierra Nevada region near the eastern border of central California (fig. 1). This rugged, mountainous area has numerous streams that drain westward into the Sacramento and San Joaquin Rivers. The elevation of the northern part of the Sierra Nevada region is generally lower than the elevation of the southern part. Annual-peak discharges from streams draining the Sierra Nevada almost always occur during the winter and spring (November through June), and result from a complex interaction of rain and snow. A large winter storm might produce rain on the lower parts of a basin and snow on the colder, higher parts of the basin. Peak discharge from this kind of event would be less than if rain had fallen throughout the basin. Alternatively, the runoff from a large basin-wide rainstorm can be exacerbated if the higher-elevation part of the basin has a large volume of snowpack that is available for melting and subsequent runoff.

This complicated interaction of rain and snow on the production of annual-peak discharge is most prevalent for streams draining the Sierra Nevada, but it also occurs in other mountainous areas of California (Mount, 1995). The strength of the interaction depends largely upon the elevation of the basins above the gaged locations. The relation between average month of occurrence (timing) of the annual-peak discharge and mean basin elevation is shown in figure 2. The average month of occurrence of annual-peak discharge for each gaged site used in the regional skew analysis was computed, starting with October equal to 1, and plotted against mean basin elevation. Because relatively few annual peaks in California occur in July through September, those months were not used to compute the averages in figure 2. Annual-peak discharges from streams with mean basin elevations less than about 4,000 ft have an average timing clustered between mid-November and mid-January and thus most likely represent runoff from rainstorms. On the other

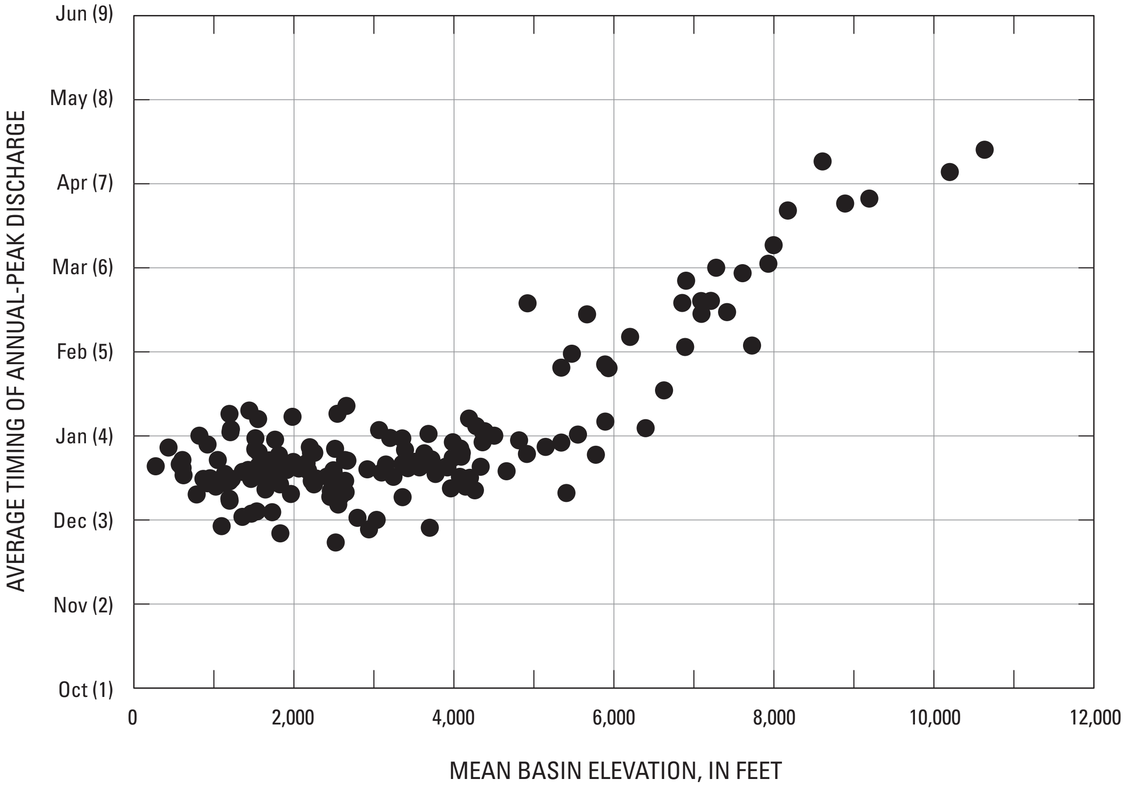


Figure 2. Relation between the average timing of peak discharge and mean basin elevation for 158 sites used for regional skew analysis in California. A small number of peaks in July, August and September were not used to determine average month of occurrence.

hand, the average timing of annual-peak discharge for mean basin elevations greater than 4,000 ft generally increases with increasing elevation up to about 8,000 ft, where the average timing generally flattens at about April. March through May is generally the snowmelt period, so that floods that occur during this period are generally the result of snowmelt, or rain and snowmelt.

The implication of this trend in annual peak-discharge timing is that rain is the main cause of peak discharge in basins with mean elevations lower than about 4,000 ft, and that the interaction of rain and snow increases as elevation increases above 4,000 ft. For basins with mean elevations above about 8,000 ft, the effects of snow on peak discharge are predominant. As discussed later in the section titled Methodology Adjustments for California, the rain-snow interaction significantly affects regional skew in California and, at some sites, the degree of interaction may be so great that the annual-peak discharges may require separation into two groups for flood-frequency analysis (mixed-population analysis) as described in Bulletin 17B. The need for a mixed-population flood-frequency analysis is expected to be most appropriate for sites with mean basin elevations greater than about 8,000 ft.

Data Used

Annual-peak discharge data were used for several specific purposes. The databases for each purpose are described in the following sections.

Regional Skew Database

New regional skew values were developed for California by using a database of 145 USGS gaged sites having essentially unregulated annual peak-flow records of at least 30 years through water year 2006. In addition, the regional skew database included 13 of the 16 stations at key dam sites for which unregulated peak flow data were estimated from unregulated daily maximum discharge data furnished by the ACOE. All sites used for the regional skew analysis are shown in figure 1. Station locations and names, together with information about the annual peak-discharge and skew data are shown in table 1. Table 1 includes information about stations in the Sacramento–San Joaquin River Basin for which flood frequency information was developed.

Although peak-discharge information for some of the 145 sites was recorded in the National Water Information System (NWIS) database with codes indicating that discharges were affected by regulation or diversion, the effects on peak discharge were considered negligible or the periods significantly affected by regulation or diversion were excluded from the analysis (table 1). Peak-discharge records at two sites on the San Benito River (stations 11158500 and 11158600) were combined into a single, longer record that was assigned an artificial station number of 11158699. Artificial station numbers ending in digits 99 were also assigned to the 16 stations at key dam sites for which unregulated peak flow data were estimated from unregulated daily maximum discharge data.

Elimination of Redundant Sites and Other Non-typical Sites

Redundancy results when the drainage basins of two gaged sites are nested, meaning that one is contained inside the other, and the sizes of the two basins are similar. Then, instead of providing two independent spatial observations depicting how drainage basin characteristics are related to skew (or flood quantiles), these two basins will likely have the same hydrologic response to a given storm and thus represent only one spatial observation. When sites are redundant, a statistical analysis using both gaged sites incorrectly represents the information in the regional data set (Gruber and Stedinger, 2008). To determine if two sites are redundant and thus represent the same hydrologic response, two pieces of information are considered: (1) whether their watersheds are nested and (2) the ratio of the basin drainage areas.

The first metric, normalized distance, is used to determine the likelihood the basins are nested. The normalized distance between two basin centroids, *ND*, is defined as (Veilleux, 2009)

$$ND = \frac{D_{ij}}{\sqrt[4]{DA_i DA_j}}, \tag{1}$$

where

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

DA_i and DA_j are the drainage areas at sites i and j.

The second measure, drainage area ratio (*DAR*), is used to determine if two nested basins are sufficiently similar in size to conclude that they essentially represent the same watershed for the purposes of developing a regional hydrologic model. The *DAR* is defined as

$$DAR = Max\left[\frac{DA_i}{DA_j}, \frac{DA_j}{DA_i}\right], \tag{2}$$

where DA_i and DA_j have already been defined.

Two basins might be expected to be redundant if they are close together and similar in size. A previous study in the southeastern United States (Veilleux, 2009) determined that site pairs having *ND* less than or equal to 0.5 and *DAR* less than or equal to 5 were likely to have redundancy problems for determining regional skew, and therefore one of each site pair was removed from the regional skew analysis. The same values for *ND* and *DAR* were used to screen and remove sites from the regional skew analysis in California. Following screening, basin boundaries of identified pairs were examined to determine if the two sites were really nested.

The two key dam sites removed because of redundancy (Sacramento River at Keswick [station 11370599] and Feather River at Oroville [station 11407099]) had the two largest basins considered for the regional skew analysis. These large basins included several subbasins used in the regional skew analysis and thus truly were nested even though computed *ND* and *DAR* were greater than 0.5 and 5.0, respectively.

One key dam site, Fresno River below Hidden Dam near Daulton (station 11258099), was removed from the regional skew database because preliminary analysis indicated that an LP3 distribution provided only a very poor fit to the estimated annual peak-discharge record for both the lower and upper tail. Thus, station skew for this site likely did not represent the general variation in skew throughout the study area.

Basin and Climatic Characteristics

Various basin characteristics for each of the 158 sites in the regional skew analysis were derived from various national geographic information system (GIS) databases, including the National Hydrologic Dataset (NHDPlus), the National Land-Cover Dataset (NLCD), and the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) climatic dataset based on data from 1970 to 2000. Table 2 gives the basin-characteristic names, descriptions, units, and sources of information. Table 3 shows basin characteristics for the 158 sites used in the regional skew analysis. At most of the sites shown in table 3, the drainage area determined from the NHDPlus GIS dataset closely matched the drainage area manually determined from topographic maps and reported in the NWIS peak-flow database. At two sites, station 11063000 and the key dam site at station 11259099, the drainage area determined from the GIS dataset differed from the drainage area reported in NWIS by more than 10 percent. For these two sites, the only basin characteristics considered to be reliable were those relating to basin elevation, and no other basin characteristics are given in table 3. For the two sites on the San Benito River where peak-discharge records were combined (station 11158500 and station 11158600), the basin characteristics measured for station 11158600 were used for the artificial combined-record station 11158699.

Test for Trends in Long-Term Data Used for Regional Skew Analysis

Flood-frequency analysis requires annual peak-flow data at each site that are random, independent, and generated by a process that is invariant (stationary) over time. Peak-flow data that indicate trends over time may reflect watershed or climatic changes that can significantly change flood characteristics and make flood-frequency estimates difficult to interpret and unreliable. To determine whether annual peak-discharge data are showing trends in California, 69 sites used in the regional skew analysis that had complete annual discharge records from 1977 to 2006 (30 years) were tested for monotonic trends using Kendall's tau, a non-parametric test for trends described by Helsel and Hirsch (1992). The locations of the 69 sites represented the locations of all 158 sites used for the regional skew analysis. The two primary outputs from the test were the tau value and the p-value. The tau value measures the strength of the correlation between the annual peak-flow values and time. Positive values of tau indicate increasing trends and negative values indicate decreasing trends. Trends generally are considered to be significant when the p-value is less than or equal to 0.05. A p-value of 0.05 indicates that there is a 5 percent probability that the test will identify a trend when there is no actual trend present.

Of the 69 sites tested for a trend in annual-peak discharge from 1977 to 2006, none had tests with p-values less than or equal to 0.05 Table 4 lists the long-term sites used for the trend test and the data from that test. On the basis of the trend-test results, monotonic trends in annual-peak discharge are not considered to be a factor anywhere in California and thus do not affect the interpretation or overall reliability of flood-frequency results.

Database for Stations at Key Dam Sites

Unregulated peak-discharge data were estimated for 16 key dam sites selected by the ACOE. Ten of the 16 selected sites had concurrent recorded unregulated, annual-peak discharge and annual-maximum-daily discharge data obtained before dams were constructed. In addition to the concurrent unregulated peak-discharge and maximum-daily-discharge data, all sites had longer records of estimated, unregulated annual-maximum-daily discharge data that were developed and provided by the ACOE (John High, Chief, Hydrology Section, Sacramento District, U.S. Army Corps of Engineers, written commun., March 2009). These longer records of estimated, unregulated annual-maximum-daily discharge were used to estimate long-term records of annual-peak discharge that were subsequently used in the regional skew analysis for California. The 16 sites for which annual-peak discharges were estimated are shown in table 5 with a brief indication of the periods of discharge records. The estimated, unregulated annual-maximum-daily discharge data provided by the ACOE generally are the same values as those synthesized for a previous hydrologic study of streams in the Sacramento–San Joaquin River Basin (U.S. Army Corps of Engineers, 2002). The unregulated annual-maximum-daily discharge data provided by the ACOE for the 16 selected sites are shown in table 6.

Database for Flood-Frequency Analysis in the Sacramento–San Joaquin River Basin

Flood-frequency statistics were calculated for 256 sites in the Sacramento–San Joaquin River Basin in California. Included in the 256 sites were 50 of the 158 sites used in analyzing regional skew for all of California. Flood-frequency statistics were calculated for all 16 key dam sites also, even though only 13 of the 16 sites were used in the regional skew analysis. All sites for which flood-frequency was analyzed had minimally regulated peak-discharge records for at least 10 years. Periods of regulated peak-discharge record were excluded from analysis for sites that had at least 10 years of unregulated record. Some otherwise eligible sites also were excluded from flood-frequency analysis if 25 percent or more of the recorded peak discharges were zero, or if the number of peak discharges other than zero were less than 10. Finally, some otherwise eligible sites also were excluded if the LP3 distribution provided only a very poor fit to the recorded data. The very poor fits were most often the result of one or more large outliers in a short record period. Peak-discharge records for two sites on Panoche Creek (stations 11255500 and 11255575) were combined into a single, longer record that was assigned the artificial station number of 11255599. All sites that were analyzed for flood frequency are shown in figure 1 and listed in table 1.

Analytical Methods

Various methods were used to analyze annual-peak discharge data in order to determine flood frequency at gaged sites. Those methods are described in the following sections.

Flood Frequency Based on LP3 Distribution

Flood-frequency estimates for gaged sites are computed by fitting a mathematical probability distribution to the series of annual-peak discharges as described in Bulletin 17B. The LP3 distribution, the Pearson Type 3 distribution applied to the logarithms (base 10) of annual-peak discharge data, is commonly used to estimate flood frequency in the United States and was used for the current California study.

The LP3 distribution is a three-parameter distribution that requires estimates of the mean, the standard deviation, and the skew coefficient of the population of logarithms of annual-peak discharge at each gaged site. The basic equation for determining flood frequency from the three parameters is the following:

$$\log Q_P = \overline{X} + K_p S, \tag{3}$$

where

Q_p is the annual-peak discharge for the exceedance probability, P,

$\overline{X}$ is the mean of the logarithms of the annual-peak discharge,

K_p is a factor based on the weighted skew coefficient and the exceedance probability, P, which can be obtained from Appendix 3 in Bulletin 17B, and

S is the standard deviation of the logarithms of the annual-peak discharge, which is a measure of the degree of variation in the annual values about the mean value.

The mean, the standard deviation, and the skew coefficient can be estimated from the available sample data (recorded annual-peak discharges). However, a skew coefficient calculated from a small sample tends to be an unreliable estimator of the population skew coefficient. Accordingly, the guidelines in Bulletin 17B indicate that the skew coefficient calculated from at-site sample data (station skew) needs to be weighted with a generalized, or regional, skew determined from an analysis of selected long-term gaged sites in the study region. The value of the skew coefficient used in equation 3 is this weighted skew that is based on station skew and regional skew. As previously described, Bayesian generalized least squares (GLS) regression, a newly developed method for determining regional skew, was used for the current study. The regional skew analysis is described in detail in a later section of the report titled Statistical Analysis of Regional Skew. Some of the more technical details of the mathematics involved are more fully presented in Appendixes B and C.

Equation 3 forms the basis for calculating flood frequency at gaged sites, but Bulletin 17B also provides methods for adjusting the results for zero flows, testing and adjusting for low outliers, and adjusting for historical floods that occur outside the period of systematic peak-discharge data collection. While these adjustments generally improve flood-frequency estimates, the new expected moments algorithm (EMA) incorporates historical discharges more efficiently and allows peak discharges that are known only to be within some range of plausible values (interval or bounded discharges) to be used in flood-frequency analysis. Consequently, the EMA was used in the current study.

Expected Moments Algorithm (EMA)

The EMA method was used for an initial LP3 frequency analysis in order to determine station skew for all sites used in the regional skew analysis. For sites that have systematic annual-peak discharge records for complete periods, no low outliers, and no historical flood information, the EMA method calculates identical values of the LP3 parameters (mean, standard deviation, and station skew) as the conventional method-of-moments described in Bulletin 17B. The EMA method, however, can incorporate into the analysis censored peak-discharge data. Censored data may be expressed in terms of discharge perception thresholds during historical periods outside the period of systematic data collection. For example, a site may have some historical information that indicates that a large recorded peak discharge of Q_{hist} was the largest since 1900, before systematic data collection was started in 1930. Each annual peak from 1900 to 1930 can thus be characterized as a censored discharge whose value is known not to have exceeded the perception threshold, Q_{hist}, and estimates of those bounded discharges between 0 and Q_{hist} can be used in the LP3 flood-frequency analysis. In the same way, the EMA method allows use of bounded discharges to characterize any missing data during periods of systematic data collection. These missing peak discharges can be described by perception thresholds or, if we have more knowledge about the likely range of missing discharge, by interval discharges that have specific upper and lower bounds. For example, if a peak was not recorded because the peak stage did not reach the elevation of the gage, the missing peak might be characterized as an interval discharge with a range that is bounded by zero and the peak discharge associated with the elevation of the gage. Missing peaks during periods of systematic data collection typically are ignored when the conventional LP3 method is used.

Censored data also can be low outliers in the systematic record. Low outliers are peak discharges that are significantly smaller than other recorded peak discharges and consequently often have a large effect on the LP3 distribution fit to all the recorded data. The primary focus of flood-frequency studies is the upper tail of the distribution (larger, rarer peak discharges), so closely fitting the upper tail is more important than fitting all data points, particularly abnormally low peak discharges. The LP3 distribution only has three adjustable parameters and thus may not be able to always fit the smallest and the largest flood flows. Accordingly, the conventional LP3 method described in Bulletin 17B incorporates a Grubbs-Beck statistical test to determine when a recorded peak discharge is unusually small compared with all other recorded peaks and should be treated as a low outlier. Bulletin 17B further describes a conditional probability adjustment that is made when low outliers are identified. The EMA also makes use of the Grubbs-Beck test to identify low outliers. In this case, the test is iterated to determine if censoring an outlier causes any of the remaining peaks to be identified as an outlier. The EMA computation used to fit an LP3 distribution when low outliers are censored is different from that described in Bulletin 17B. Although the Grubbs-Beck test provides a reasonable way to identify low outliers that may result in fitting problems with the LP3 distribution, sometimes not all low peak discharges that cause fitting problems are identified in either the conventional LP3 method or in the EMA program. Thus, when either method is applied, sometimes a user-specified low-outlier threshold that is larger than the values identified by the Grubbs-Beck test is used. Individual flood-frequency curves were visually inspected to determine whether one or more peak discharges in the lower tail (discharges with an annual exceedance probability of 0.50 or greater) of the distribution might be adversely affecting the fit of the upper tail of the frequency curve and thus require censoring. For a few sites, the curve fit for the upper tail was substantially improved by censoring the complete lower tail of the distribution. However, the substantial improvement came at the expense of a large increase in the mean square error (MSE) of the station skew as computed by the EMA program. All sites for which a user-specified low-outlier threshold was used are noted in table 1.

Although the EMA allows the use of censored data that can significantly improve flood-frequency analyses, establishing reasonable bounds on the discharges can require considerable judgment on the part of the analyst. Fortunately, results from the EMA program generally are not sensitive to small changes in the perception thresholds or bounds used for interval discharges.

In practice, the EMA provides estimates of missing, but bounded, discharge in a 5-step iterative process described by England (2003b):

1. Estimate an initial set of the three sample statistics (mean, standard deviation, and skew) from the logarithms of peak-discharge data with known magnitudes. These discharges are typically recorded peaks from the gaging station records and possibly some historical discharges. At this step, interval (bounded) discharges are not included.
2. Use the initial sample statistics from step (1) to estimate a set of LP3 distribution parameters .
3. Use the set of LP3 parameters from step (2) to estimate a new set of sample statistics based on the complete data set, including unknown discharges less than a threshold, unknown discharges that exceed a threshold, and unknown discharges with specific lower and upper bounds. The threshold values and lower and upper bounds are used as the initial estimates of the unknown discharges.
4. Use this new set of moments to estimate a new set of LP3 parameters. These estimates are based on expected values given that the unknown discharges are less than the upper thresholds and bounds and greater than the lower thresholds and bounds.
5. Compare the parameters from step (4) with those computed from step (2). Repeat steps (3) and (4) until the parameter estimates converge. The main equations used by EMA to make the estimates in the iterative process are listed by Cohn and others (1997), England (1999), and England and others (2003a,b). Cohn and others (2001) describe the EMA computation for evaluating the sampling variance of parameters and quantiles.

Statistical Analysis of Regional Skew

Tasker and Stedinger (1986) developed a weighted least squares (WLS) procedure for estimating regional skew coefficients that is based on sample skew coefficients corresponding to the logarithms of annual peak-discharge data. Their method of regional analysis of skewness estimators accounts for the precision of the skewness estimator for each station, which depends on the length of record for each station and the accuracy of the regional skew model. More recently, Reis and others (2005), Gruber and others (2007), and Gruber and Stedinger (2008) developed a Bayesian generalized least squares (GLS) regression model for regional skew. While WLS regression accounts for the precision of the regional model and the effect of the record length on the variance of skewness estimators, GLS regression considers the cross correlations among the skewness estimators also. As explained later in the report section titled Methodology Adjustments for California, the cross correlations among the skewness estimators were important for the California regional skew study. The new Bayesian GLS regression procedures describe the precision of the estimated model error variance, a pseudo analysis of variance and enhanced diagnostic statistics (Griffis and Stedinger, 2007). A Bayesian GLS regional skew analysis was used in recently completed flood-frequency studies for the Southeastern United States (Feaster and others, 2009; Gotvald and others, 2009; Weaver and others, 2009).

The California regional skew study described here is based on use of Bayesian GLS regression procedures. However, the statistical procedures used in the Southeastern United States regional flood-frequency study had to be extended because of two problems that arose in the analysis of the California data set. The first problem was the difficulty in estimating the cross correlation of at-site skew estimators that were determined from the EMA analysis of the California regional-skew data set. This difficulty arose because EMA allows for censoring of low outliers and the use of estimated interval discharges for missing recorded data, and computing cross correlations when peak discharges are not represented by single values is difficult. The second problem was the extensive cross-correlation among concurrent recorded peak discharges in California. This extensive cross-correlation was not present in previous regional skew studies using Bayesian GLS regression procedures and required special attention in the analysis. To properly account for the cross-correlation problems and develop a suitable regression model and regression diagnostics, Bayesian WLS and GLS regressions were combined. In essence, the regression parameters of the regional skew model for California were determined using Bayesian WLS regression procedures, and the accuracy of the regression parameters and the regression models were determined using a special Bayesian GLS regression procedure. Those procedures are described in Appendixes B and C.

Regional Regression Models

The basic model for a regional (or generalized) skew analysis when there are k explanatory variables and n stations is

$$\hat{\boldsymbol{\gamma}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (4)$$

where

$\hat{\boldsymbol{\gamma}}$ is an $(n \times 1)$ vector of the unbiased estimated at-site skew coefficients for every station (see Appendix B for more discussion about unbiased skew estimators),

$\mathbf{X}$ is an $(n \times k)$ matrix of k basin characteristics with a column of ones corresponding to a constant in the model,

$\boldsymbol{\beta}$ is a $(k \times 1)$ vector of model coefficients,

$\boldsymbol{\varepsilon}$ is the $(n \times 1)$ vector of total errors, including both model and sampling errors where $E[\boldsymbol{\varepsilon}] = 0$ and $\boldsymbol{\Lambda}$ is the covariance matrix that represents $E\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right]$.

The matrix $\boldsymbol{\Lambda}$ is computed as the sum of two covariance matrices (Reis and others, 2005) $\sigma_\delta^2 \mathbf{I} + \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}})$, where σ_δ^2 is the model error variance describing the precision with which the proposed model $\mathbf{X}\boldsymbol{\beta}$ can predict the true skews, which are denoted γ_i, and the matrix $\boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}})$ represents the sampling variances and covariances of the skewness estimators $\hat{\gamma}_i$. The value of $\boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}})$ is determined by the length of record at each station, the regional skew, and the cross-correlation of the concurrent flows.

The standard WLS or GLS estimator of $\boldsymbol{\beta}$, which for given $\boldsymbol{\Lambda}$ is unbiased with minimum variance, is

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \boldsymbol{\Lambda}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \boldsymbol{\Lambda}^{-1} \hat{\boldsymbol{\gamma}}. \quad (5)$$

In WLS, the $\boldsymbol{\Lambda}$ matrix has non-zero elements only on the diagonal. In GLS, the $\boldsymbol{\Lambda}$ matrix nominally has the same diagonal elements, but the off-diagonal elements are also non-zero to reflect the cross-correlation among the at-site skewness estimators $\hat{\gamma}_i$.

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

$$\hat{\rho}\left(\hat{\gamma}_i, \hat{\gamma}_j\right) = Sign\left(\rho_{ij}\right) cf_{ij} \left|\rho_{ij}\right|^{\kappa}, \quad (6)$$

where

ρ_{ij} is the cross-correlation of concurrent annual-peak discharges for two gaged stations,

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 and is defined as follows:

$$cf_{ij} = n_{ij} / \sqrt{\left(n_{ij} + n_i\right)\left(n_{ij} + n_j\right)}, \quad (7)$$

where

n_{ij} is the length of the period of concurrent record,

n_i and n_j are the number of nonconcurrent observations corresponding to sites i and j, respectively.

Cross-Correlation Model of Concurrent Annual-Peak Discharge

A cross-correlation model for the annual- peak discharges in California was developed using 21 sites with more than 65 years of concurrent records containing no censored peaks. None of the key dam sites identified by the ACOE were used in this analysis because annual-peak discharge at those sites was estimated. Various models relating the cross-correlation of the concurrent annual-peak discharge at two sites, $\rho_{ij,}$ to various basin characteristics were considered. In general, 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 $(-\infty,+\infty)$ range. The adopted model for estimating the cross–correlations of concurrent annual-peak discharge at two stations, which used the distance between basin centroids, D_{ij}, as the only explanatory variable, is

$$\rho_{ij} = \frac{\exp\left(2Z_{ij}\right)-1}{\exp\left(2Z_{ij}\right)+1}, \qquad (8)$$

where

$$Z_{ij} = \exp\left(0.27 - 0.0037 D_{ij}\right). \qquad (9)$$

An ordinary least squares regression analysis based on 159 station-pairs indicated that this model is as accurate as one having 52 years of concurrent annual peaks from which to calculate a cross-correlation. Figure 3 shows the fitted relation between Z and the distance between the basin centroids together with the plotted sample data from the 159 station-pairs of data. Figure 4 shows the functional relation between the untransformed cross correlation and the distance between the basin centroids in the California study and the Southeastern United States study. The cross correlations decrease more gradually with increasing distance between the basin centroids in California than they do in the Southeastern United States. This difference between California and the Southeastern United States indicates that large flood-producing storms cover more area in California than in the Southeastern United States.

The cross-correlation model was used to estimate site-to-site cross correlations for concurrent annual-peak discharges at all pairs of sites. Figure 5 is a histogram of the relative frequency of the distribution of the estimated cross correlations among the 158 sites in the California data set and, for comparison, the distribution of the cross correlations among 342 sites in the Southeastern United States.

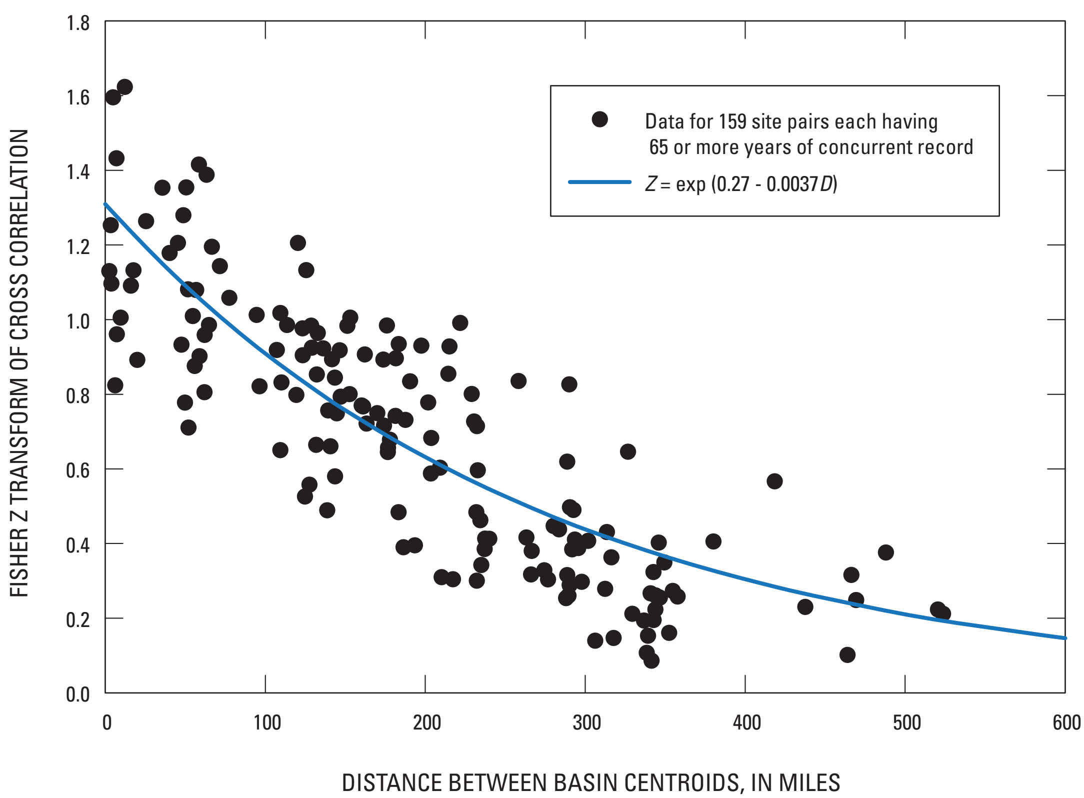


Figure 3. Relation between the Fisher Z transform (Z) of logs of annual-peak discharge and distance between basin centroids for 159 station-pairs in California.

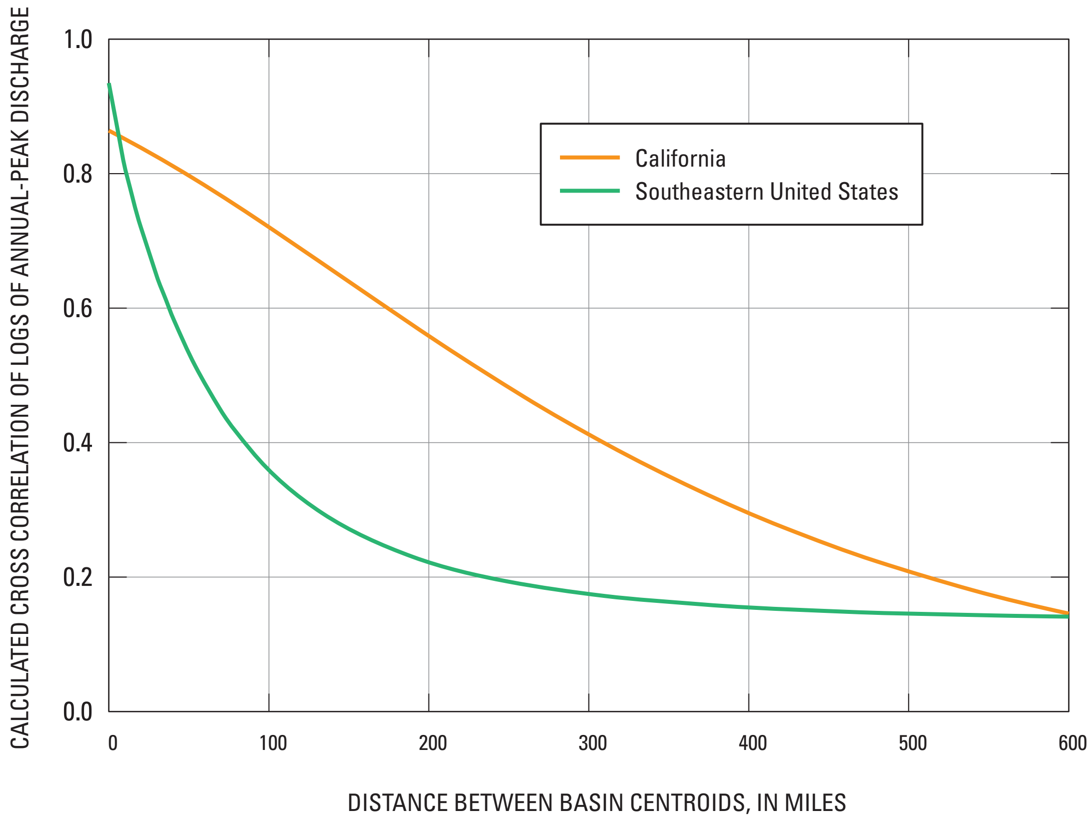


Figure 4. Relation between the cross-correlation of logs of annual-peak discharge and the distance between basin centroids based on data from 159 station-pairs in California and 1,317 station-pairs in the Southeastern United States.

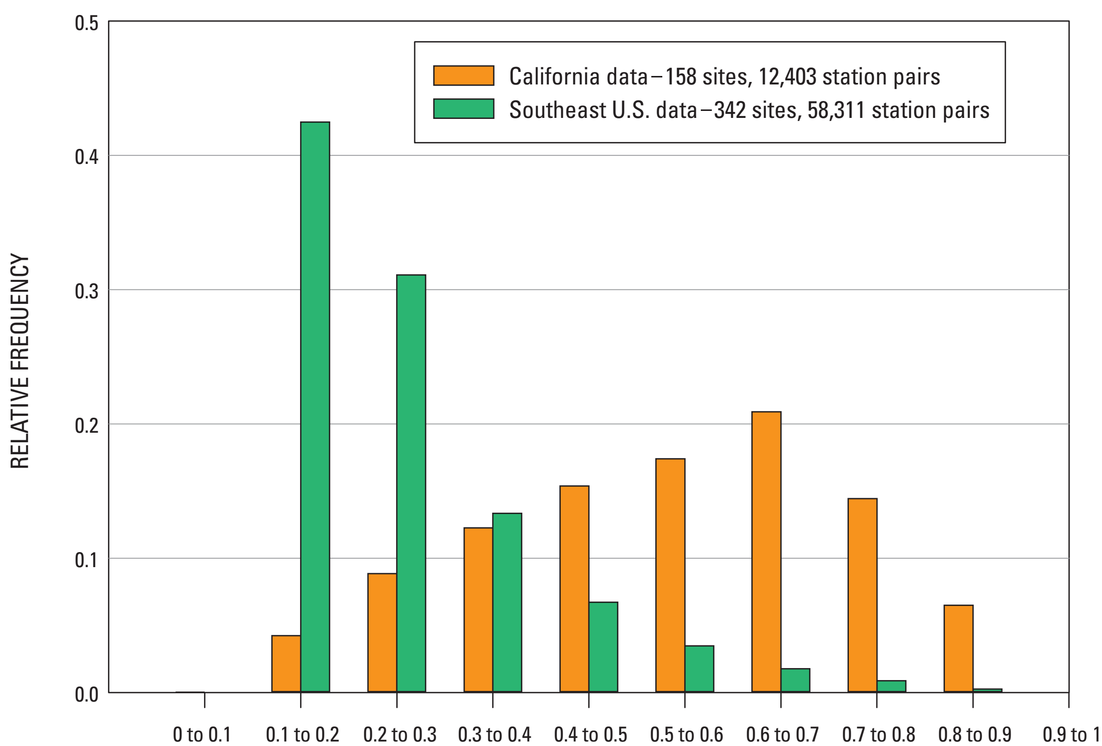


Figure 5. Histogram of relative frequency of calculated cross-correlation values in California (158 sites) and in the Southeastern United States (342 sites).

Methodology Adjustments for California

The Southeastern United States regional skew analysis illustrates how a Bayesian GLS analysis would generally proceed (Feaster and others, 2009; Gotvald and others, 2009; Veilleux, 2009; and Weaver and others, 2009). However, when a Bayesian GLS analysis of the California data set was attempted, reliable results were not obtained because of the large cross correlations. Thus, an alternative procedure that uses a combination of Bayesian WLS and GLS was developed so that the regional skew analysis would provide more stable and defensible results. The need for the alternative procedure and the specific computational steps for the procedure are described in Appendix B. The results of the California regional skew regression using the alternative procedure are provided below.

All of the available basin characteristics were initially considered as explanatory variables in the regression analysis for regional skew. The one key basin characteristic that was statistically significant in explaining the site-to-site variability in skew was the mean basin elevation (ELEV). Table 7 gives the final results for three models: a constant skew denoted “Constant,” a model that uses a linear relation between skew and mean basin elevation denoted “Elev,” and a model that uses a nonlinear relation between skew and mean basin elevation denoted “NL-Elev.”

As shown in table 7, the linear Elev model has a Pseudo R^2_δ of 41 percent, while the nonlinear NL-Elev model has a larger Pseudo R^2_δ of 48 percent and a slightly smaller AVP_{new}. The Pseudo R^2_δ values describe the fraction of the variability in the true skews explained by each model (Gruber and others, 2007). A Constant model does not explain any variability, so the Pseudo R^2_δ is equal to 0 percent. Also, the posterior mean of the model error variance, σ^2_δ, for the NL-Elev model is 0.10, which is smaller than that for the linear Elev model (σ^2_δ = 0.12) and substantially smaller than that for the Constant model (σ^2_δ = 0.20). The average sampling error variance (ASEV) in table 7 is the average error in the regional skewness estimator at the sites in the data set.

The average variance of prediction at a new site (AVP_{new}) corresponds to the mean square error (MSE) used in Bulletin 17B to describe the precision of the generalized skew. In table 7, the NL-Elev model has the lowest AVP_{new}, equal to 0.14. However, this AVP_{new} is an average value computed by averaging the variance of prediction at a new site (VP_{new}) for all of the 158 sites in the California study. Just as generalized skew varies from site to site, depending upon mean basin elevation, so too do the values of VP_{new}. Table 8 gives values of the variance of prediction for the regional skew, VP_{new}, and effective record length (ERL) for the NL-Elev model for values of mean basin elevation between 0 and 11,000 ft.

Thus, the NL-Elev regional skew model for California has effective record lengths ranging from 52 years to 65 years, depending upon the mean basin elevation. A VP_{new} ranging from about 0.13 to 0.17 is a marked improvement over the Bulletin 17B skew map, whose MSE is 0.302 (Interagency Advisory Committee on Water Data, 1982) for a corresponding effective record length of only 17 years.

The nonlinear elevation model provides a reasonable fit for the California regional skew data (fig. 6). While the more complicated nonlinear model is not that different from the simpler linear elevation model, the nonlinear model provides smaller values of positive skew at high elevations and less negative values of skew for low elevations. For example, when a mean basin elevation is zero at sea level, the nonlinear model provides a regional skew of -0.62, while the linear elevation model provides a regional skew of -0.76. Conversely, when a mean basin elevation is 11,000 ft, the nonlinear model provides a regional skew of 0.61, while the linear model provides a regional skew of 0.79. These differences, though subtle, are significant, and the nonlinear model indicates that regional skew flattens out in the tails instead of continually increasing in absolute value. This flattening of skew at both low and high elevations is consistent with the relation between the timing of annual-peak discharge and the elevation, which is largely reflective of the degree of rain-snow interaction affecting peak discharge. Annual peak-discharges from basins that have mean elevations less than about 4,000 ft have little rain-snow interaction (fig. 2) and thus, might be expected to have constant or near-constant regional skews. Likewise, at the other extreme, basins at very high elevations tend to have annual-peak discharges that are predominantly the result of spring snowmelt events. Thus, beyond some point, higher elevation has less effect on the distribution of annual maxima because few, if any, of the flood peaks are caused by winter rainfall events.

Only six sites that have a mean basin elevation greater than about 8,000 ft were used in the regional skew analysis (fig. 2). Because of the scarcity of such high-elevation sites, the calculated regional skew values for high-elevation sites may be less reliable than those for lower-elevation sites. In addition, combining a few large, winter-rain caused peaks with many more smaller, spring snowmelt peaks often results in fitted frequency curves from the LP3 distribution with a sharp upward curvature that may poorly represent the true frequency of the largest floods. Peak-discharge data for sites that have mean basin elevations above about 8,000 ft need to be examined to determine if a mixed-population analysis for determining flood frequency described in Bulletin 17B might be more appropriate than the standard LP3 method. When a mixed-population analysis is used the rain-caused floods and snowmelt floods are analyzed separately, and the separate frequency curves are combined to represent the joint

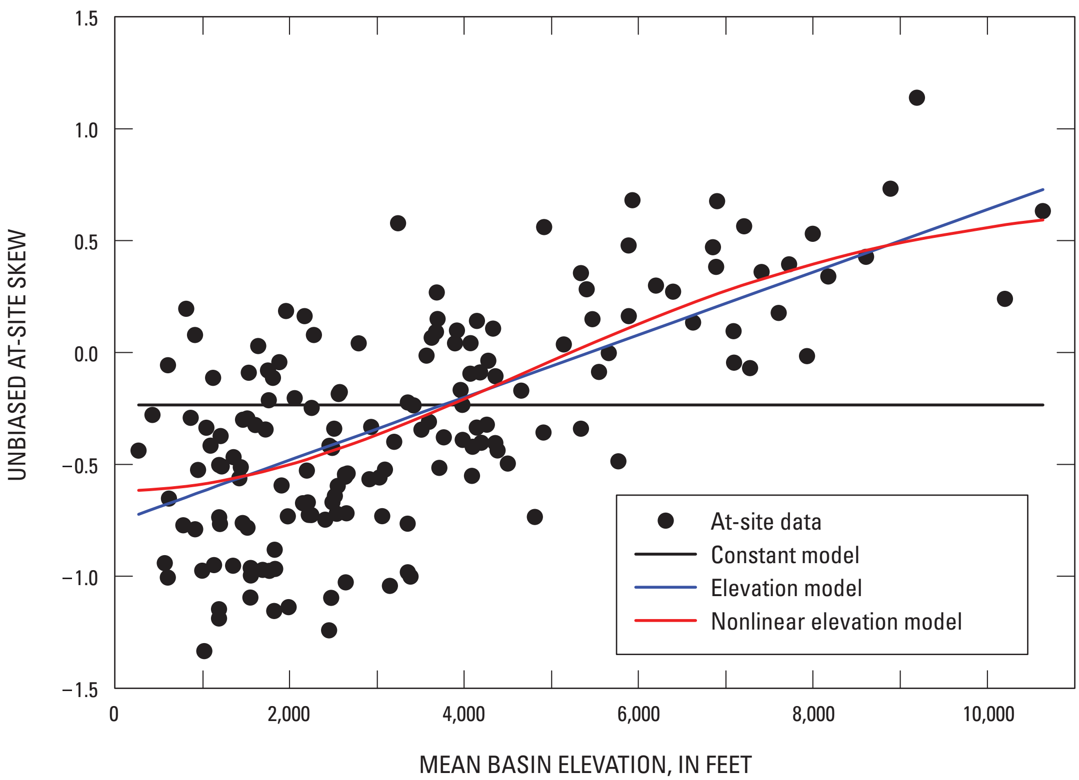


Figure 6. Relations between the unbiased at-site skew and the mean basin elevation for 158 sites in California. The lines represent a model based on a constant skew (Constant), a model with a linear relation between skew and mean basin elevation (Elev), and a model with a nonlinear relation between skew and mean basin elevation (NL-Elev). The models were developed from Bayesian weighted least squares and generalized least squares (WLS-GLS) analyses.

probability of flooding from any cause (Murphy, 2001). The Sierra Nevada in California has previously been indicated as an area having a mixture of rain and snowmelt events (Interagency Advisory Committee on Water Data, 1982, p. 16), and the ACOE commonly analyzes rain-caused floods separately from snowmelt-caused floods in this area (U.S. Army Corps of Engineers, Sacramento District, 2002).

Flood-Frequency Results

Flood-frequency estimates for 158 stations used in the regional skew analysis and 206 additional stations in the Sacramento–San Joaquin River Basin are shown in table 1 at the back of the report. Table 1 includes information about peak-discharge record lengths, historical record periods, censored data and thresholds, and skew coefficients also. All flood-frequency estimates in table 1 were developed by using the EMA program.

The flood-frequency estimates were calculated by applying the LP3 probability distribution, with a weighted skew as described in Bulletin 17B, to the annual peak-discharge data at the stations. The weighted skew is determined by weighting the station skew and the regional skew inversely proportional to their respective mean square errors, as shown in the following equation:

$$G_w = \frac{MSE_R\left(G_s\right) + MSE_s\left(G_R\right)}{MSE_R + MSE_s}, \qquad (10)$$

where

G_w is the weighted skew,
G_s is the station skew,
G_R is the regional skew, and
MSE_R and MSE_s are the mean square error of the regional and station skew, respectively.

The MSE_R is equivalent to the variance of prediction for a new site (VP_{new}) described in the previous section. Bulletin 17B provides equations for calculating MSE_S, but these equations may not be reliable when peak-flow data are heavily censored. The EMA program, which can use heavily censored data, uses a first-order approximation for MSE_S, developed by Cohn and others (2001).

Flood-frequency curves show the LP3 distribution fitted to the recorded annual peak-discharge data for selected sites in California (figs. 7–10). Each figure shows the fitted curves based on station skew and weighted skew with the 90-percent confidence interval for the true flood-frequency distribution based on use of the weighted skew. The confidence interval determined by the EMA program defines a confidence band (difference between the upper and lower confidence limits) that generally is wider than the confidence band calculated using the conventional LP3 analysis, because the EMA results include the uncertainty in the estimated skew. As described by Cohn and others (2001), the EMA program produces more realistic confidence intervals than does the simple method used in the conventional LP3 analysis.

Figures 7 and 8 contain typical flood-frequency curves for stations that have no censored peak-flow data (no low outliers or historical periods) and that have mean basin elevations below 4,000 ft. These EMA-developed curves are identical to those that would be produced by a conventional LP3 frequency analysis and also represent flood-frequency curves for stream sites with little or no snowmelt runoff. The flood-frequency curves in figure 7 are for a station that has a relatively long period of record,73 years (Saratoga Creek, station 11169500), whereas the flood-frequency curves in figure 8 are for a station that has a short flow record, 11 years (Kingsbury Creek, station 11402700). The fitted curves based on station skew and on weighted skew are different in figure 7, indicating a substantial difference between regional skew and station skew for this long-record site. Also, the confidence interval for the long-record site in figure 7 is narrower than the confidence interval for the short-record site in figure 8. Many of the sites for which flood-frequency estimates were developed are for sites that had little or no censored data (table 1). Thus, the flood-frequency curves shown in figures 7 and 8 are typical—with varying degrees of scatter, widths of confidence intervals, and record lengths—of those for many of the sites where snow has little or no effect on peak discharge.

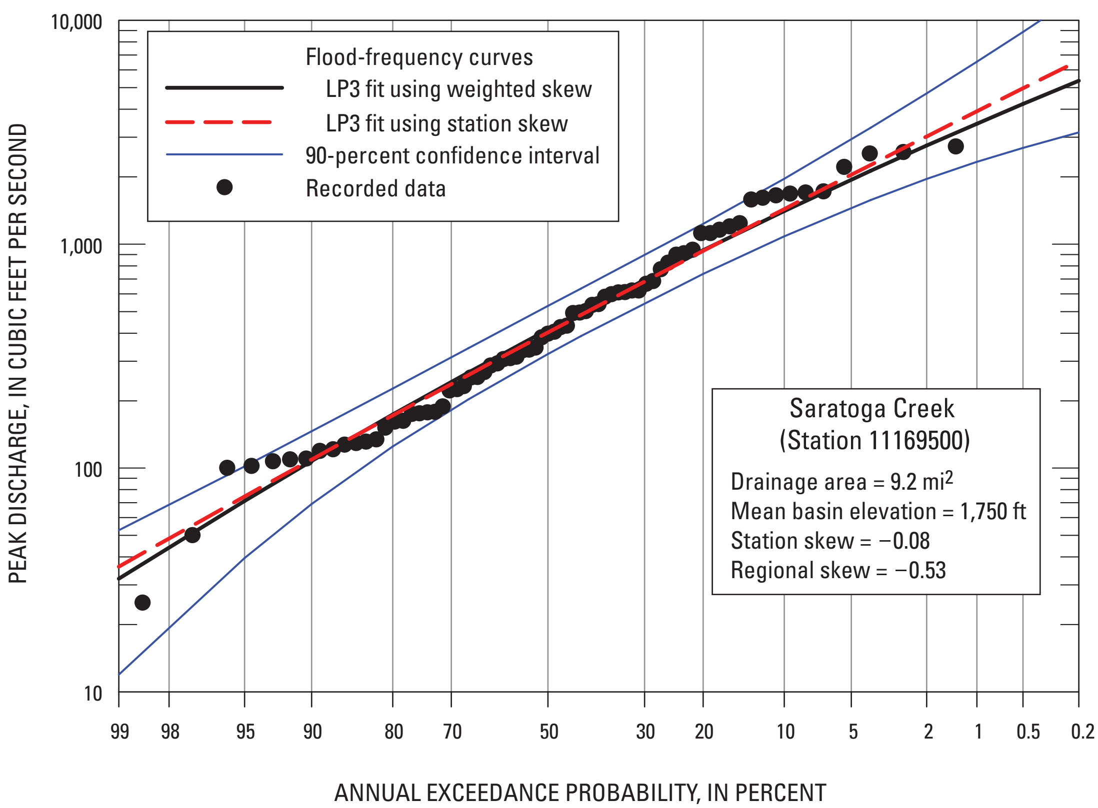


Figure 7. Flood-frequency curves for Saratoga Creek, California, (station 11169500) based on 73 years of recorded data with no censoring of annual-peak discharge. LP3, log Pearson Type 3; mi^2, square mile; ft, foot.

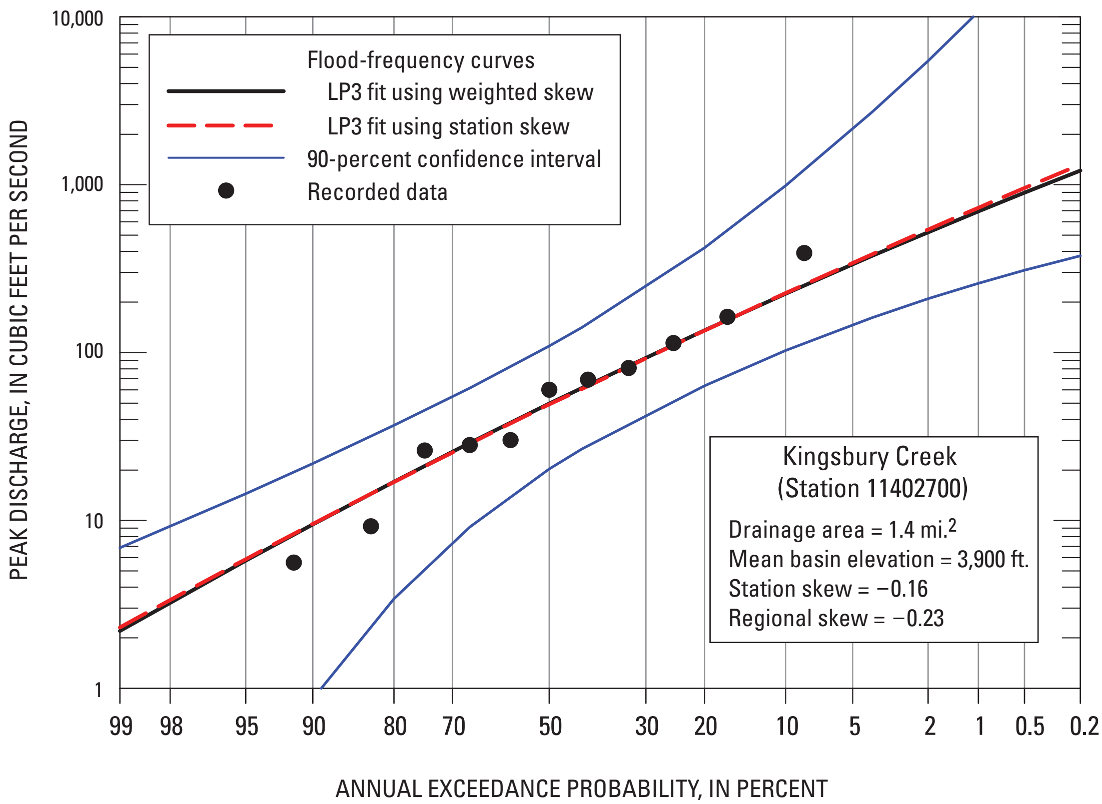


Figure 8. Flood-frequency curves for Kingsbury Creek, California, (station 11402700) based on 11 years of recorded data with no censoring of annual-peak discharge. LP3, log Pearson Type 3; mi^2, square mile; ft, foot.

Flood-frequency curves in figure 9 are for a high-elevation (mean basin elevation is 8,610 ft) station (West Walker River, station 10296500) that had censored data (an historical period with a perception threshold discharge equal to the largest recorded discharge in the systematic record). The largest recorded discharge plots above the fitted LP3 curves, but it is well within the relatively narrow 90-percent confidence interval for the true distribution. The fitted curves based on station skew and weighted skew are almost identical for this site. Only three of the 81 recorded peak discharges at station 10296500 occurred during the winter-storm rainy season (generally November through March), but the two largest annual-peak discharges were winter-storm rainy season peaks (fig. 9). In contrast, almost all of the annual floods on Kingsbury Creek (station 11402700) were from the November–March period. Thus, figures 8 and 9 represent the shift from a low-elevation flood hydrology, dominated by winter rainfall events, to a high-elevation flood hydrology, dominated by snow-melt events with only a few—usually the largest—annual peaks resulting from winter rainfall.

Figure 9 generally represents sites where rain-snow interaction affected the annual-peak discharges. Winter-storm rain season peaks tend to be the largest recorded peaks at these sites, and the differences between the large rain peaks and the other peaks are large enough to result in positively skewed LP3 distributions. The differences between winter-storm season rain peaks and the other peaks may be so large that an LP3 distribution, limited by three parameters, cannot provide a reasonable fit to the data, and a mixed-population analysis may be the only way to reasonably determine flood frequency.

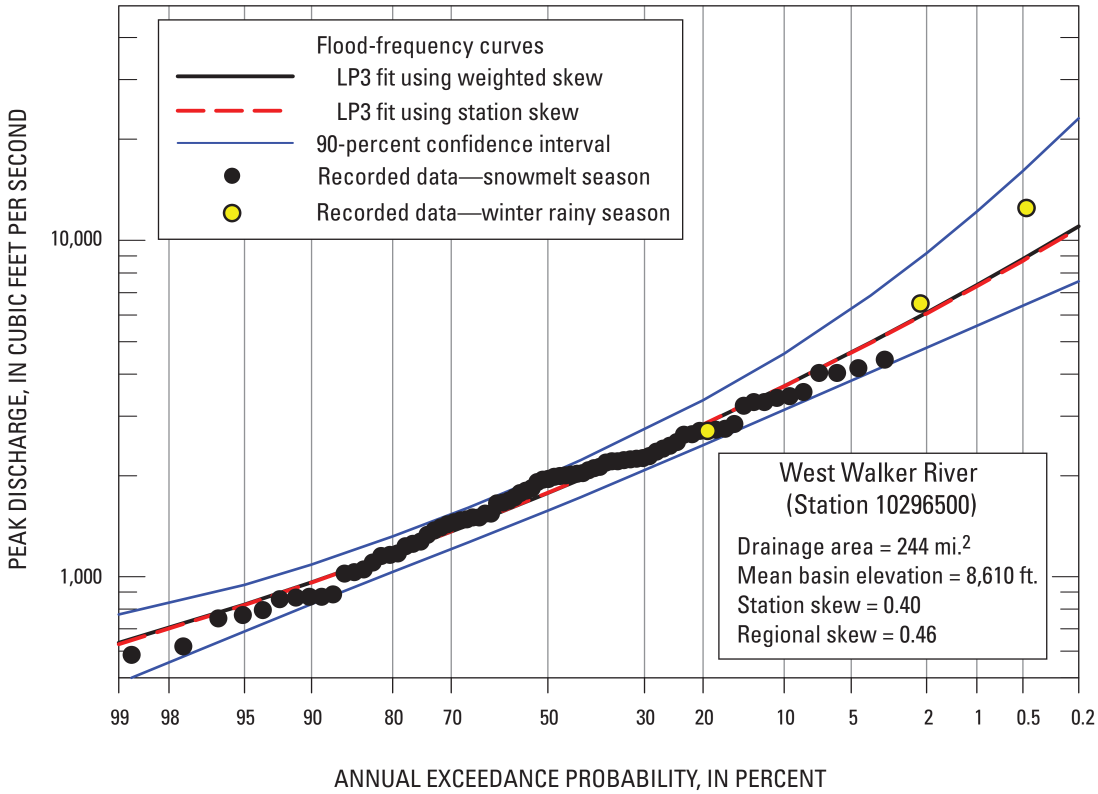


Figure 9. Flood-frequency curves for West Walker River (station 10296500), California, based on 81 years of data recorded during an historical period of 106 years with a perception threshold discharge equal to the largest discharge during the 81-year record period. LP3, log Pearson Type 3; mi^2, square mile; ft, foot.

Figure 10 shows frequency curves for a site (Cantua Creek, station 11253310) that had heavily censored data (the complete lower tail consisting of 24 recorded peaks, including one zero value, were considered to be low outliers). While the frequency curves generally fit the upper tail of the data, the confidence intervals are especially wide owing to the large number of censored peaks.

The flood-frequency curve examples in figures 7 through 10 illustrate how well the LP3 distribution fit the recorded flood data for a variety of different record lengths, data-censoring conditions, and mean basin elevations. Overall, using the newly developed regional skew function and applying the EMA for fitting the LP3 distribution provide reasonable estimates of flood frequency at gaged sites in the Sacramento–San Joaquin River Basin in California. Flood-frequency estimates at some higher-elevation sites might be improved by using a mixed-populations analysis. Although mixed-population analyses were beyond the scope of the current report, the flood-frequency study is ongoing in California and mixed-population analysis will be considered for some sites tied to poor-fitting LP3 frequency curves.

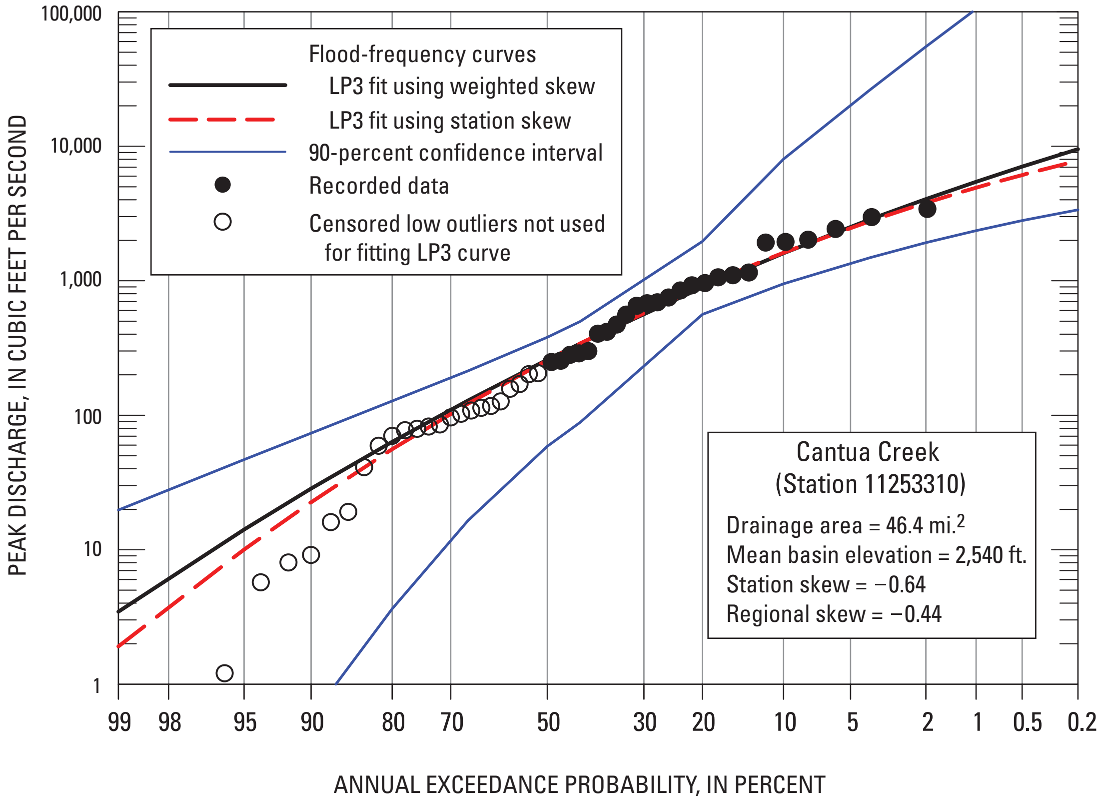


Figure 10. Flood-frequency curves for Cantua Creek (station 11253310), California, based on 49 years of recorded data with all peaks smaller than the 50-percent exceedance probability censored as low outliers.

Summary

Reliable estimates of peak discharge for various exceedance probabilities, commonly referred to as flood-frequency estimates, are needed by engineers, land-use planners, resource managers, and scientists. Accordingly, the U.S. Geological Survey (USGS), in cooperation with the U.S. Forest Service, Federal Emergency Management Agency (FEMA), and the USGS Multi-Hazards Demonstration Project (MHDP), initiated a regional flood-frequency study for California in 2008. An important aspect of the comprehensive USGS flood-frequency study for California was developing new regional skew relations for California.

Because of the common interest and the need for updated flood-frequency information in the Sacramento and San Joaquin River Basin by the USGS and the U.S. Army Corps of Engineers (ACOE), both agencies developed a secondary cooperative program for estimating unregulated annual-peak discharge at 16 selected key dam sites in the basin and using these estimated discharges as part of the regional skew analysis. A method using maintenance of variance extension (MOVE) techniques was used to estimate annual-peak discharge at the key dam sites.

Annual-peak discharges on streams draining the Sierra Nevada almost always occur during the winter and spring (November through June) and result from a complex interaction of rain and snow. Rain is the predominant cause of peak discharge in basins that have mean elevations below about 4,000 ft, and the interaction of rain and snow increases with increasing elevation above that elevation. For basins that have mean elevations above about 8,000 ft, snowmelt is the predominant cause of peak discharges. The rain-snow interaction significantly affects regional skew in California.

To determine whether annual peak flow data indicate trends in California, 69 sites that were used in the regional skew analysis and had complete annual discharge records from 1977 to 2006 (30 years) were tested for monotonic trends using Kendall's tau, a non-parametric test for trends. The locations of the 69 sites represented the locations of all 158 sites used for the regional skew analysis. Of the 69 sites tested for trends in annual-peak discharge over the 30-year period, none had p-values less than or equal to 0.05. On the basis of the trend-test results, monotonic trends in annual-peak discharge were not considered to be a factor in California and thus do not affect the interpretation or overall reliability of flood-frequency results.

Flood-frequency estimates for gaged sites are computed by fitting a mathematical probability distribution to the series of annual-peak discharges. The LP3 distribution, which is the Pearson Type 3 distribution applied to the logarithms (base 10) of annual-peak discharge data, commonly is used to estimate flood frequency in the United States and was used for the current California study. The expected moment algorithm (EMA) was used for an initial LP3 frequency analysis in order to determine station skew for all sites used in the regional skew analysis, with an adjustment for zero flows, floods identified as low outliers, and historical flood information.

The California regional skew study was based on Bayesian regression procedures. To properly account for problems caused by large cross correlations among annual-peak discharges, a combination of Bayesian weighted least squares (WLS) and generalized least squares (GLS) regression was adopted to ensure that the regression model and the diagnostics for the regression would be reliable.

Various basin characteristics were considered as possible explanatory variables in the regression analysis for regional skew. The characteristic that best explained the site-to-site variability in skew was the mean basin elevation (ELEV). Three models were developed: (1) a constant skew denoted “Constant,” (2) a model that used a linear relation between skew and mean basin elevation denoted “Elev,” and (3) a model that used a nonlinear relation between skew and mean basin elevation denoted “NL-Elev.” The average variance of prediction at a new site (AVP_{new}) corresponds to the mean square error (MSE) to describe the precision of the generalized skew. AVP_{new} was lowest for the NL-Elev model (0.14). Just as generalized skew varies from site-to-site depending upon mean basin elevation, so too do the values of variance of prediction at a new site, VP_{new} . The NL-Elev regional skew model for California has VP_{new} values ranging from about 0.13 to 0.17 and effective record lengths between 52 years and 65 years, depending upon the value of mean basin elevation. A VP_{new} between 0.13 and 0.17 is a marked improvement over the Bulletin 17B skew map, whose reported MSE is 0.302 with a corresponding effective record length of only 17 years.

Flood-frequency estimates for 158 sites used in the regional skew analysis and 206 additional sites in the Sacramento–San Joaquin River Basin were developed using the EMA program and applying the LP3 probability distribution with a weighted skew as described in Bulletin 17B to the annual-peak discharge data at the sites. Overall, using the newly developed regional skew function and applying the EMA program for fitting the LP3 distribution provide the best available estimates of flood frequency at gaged sites in the Sacramento–San Joaquin River Basin in California.

References Cited

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. 39, no. 9, p. 2089–2096.

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

England, J.F., Jr., 1999, Draft user’s manual for program EMA, at-site flood frequency analysis with historical/paleohydrologic data: Bureau of Reclamation, 52 p.

England, J.F., Jr., Jarrett, R.D., and Salas, J.D., 2003a, Data-based comparisons of moments estimators that use historical and paleoflood data: Journal of Hydrology, v. 278, no. 1–4, p. 170–194.

England, J.F., Jr., Salas, J.D., and Jarrett, R.D., 2003b, Comparisons of two moments-based estimators that utilize historical and paleoflood data for the log-Pearson type III distribution: Water Resources Research, v. 39, no. 9, p. 1243, doi:10.1029/2002WR001791.

Feaster, T.D., Gotvald, A.J., and Weaver, J.C., 2009, Magnitude and frequency of rural floods in the southeastern United States, 2006: Volume 3, South Carolina: U.S. Geological Survey Scientific Investigations Report 2009-5156, 226 p.

Gotvald, A.J., Feaster, T.D., and Weaver, J.C., 2009, Magnitude and frequency of rural floods in the southeastern United States, 2006: Volume 1, Georgia: U.S. Geological Survey Scientific Investigations Report 2009–5043, 120 p.

Griffis, V.W. , Stedinger, J.R., and Cohn, T.A. , 2004, LP3 quantile estimators with regional skew information and low outlier adjustments: Water Resources Research, v. 40, W07503, doi:1029/2003WR002697, 17 p.

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.

Griffis, V.W., and Stedinger, J.R., 2009, Log-Pearson type 3 distribution and its application in flood frequency analysis, III: sample skew and weighted skew estimators: Journal of Hydrology, v. 14, no. 2, p. 121–130.

Gruber, A.M., Reis, D.S., Jr., and Stedinger, J.R., 2007, Models of regional skew based on Bayesian GLS regression, Paper 40927-3285, *in* Kabbes, K.C., ed., World Environmental and Water Resources Congress 2007 - Restoring our Natural Habitat, American Society of Civil Engineers, Tampa, Florida, May 15–19, 2007.

Gruber, A.M., and Stedinger, J.R., 2008, Models of LP3 regional skew, data selection, and Bayesian GLS regression, Paper 596, *in* Babcock, R., and Walton, R., eds., World Environmental & Water Resources Congress 2008 - Ahupua‘a, American Society of Civil Engineers, Honolulu, Hawaii, May 12–16, 2008.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: New York, Elsevier, 522 p.

Hirsch, R.M., 1982, A comparison of four streamflow record extension techniques: Water Resources Research, v. 18, no. 4, p. 1081–1088.

Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood-flow frequency, Bulletin #17B of the Hydrology Subcommittee: Office of Water Data Coordination, U.S. U.S. Geological Survey, 183 p. Available at http://water.usgs.gov/osw/bulletin17b/dl_flow.pdf.

Martins, E.S., and Stedinger, J.R., 2002, Cross-correlation among estimators of shape: Water Resources Research, v. 38, no. 11, 7p.

Mount, J. F., 1995, California rivers and streams - The conflict between fluvial process and land use: University of California Press, p. 94–100.

Murphy, P.J., 2001, Evaluation of mixed-population flood frequency evaluation: Journal of Hydrologic Engineering, v. 6, p. 62–70.

Reis, D.S., Jr., Stedinger, J.R., and Martins, E.S., 2005, Bayesian generalized least squares regression with application to the log Pearson type III regional skew estimation: Water Resources Research, v. 41, W10419, doi:10.1029/2004WR003445, 14 p.

Ries, K.G., 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.

Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis—Ordinary, weighted and generalized least squares compared: Water Resources Research, v. 21, no. 9, p. 1421–1432. [with correction, Water Resources Research, v. 22, no. 5, p. 844, 1986.]

Tasker, G.D., and Stedinger, J.R., 1986, Regional skew with weighted LS regression: Journal of Water Resources Planning and Management, v.112, no. 2, p. 225–237.

Thomas, B.E., Hjalmarson, H.W., and Waltemeyer, S.D., 1997, Methods for estimating magnitude and frequency of floods in the southwestern United States: U.S. Geological Survey Water-Supply Paper 1997-2433, 205 p.

U.S. Army Corps of Engineers, 2002, Sacramento and San Joaquin River Basins comprehensive study: Technical studies documentation Appendix B - Synthetic hydrology technical documentation, accessed May 10, 2010, at http://www.compstudy.net/docs/techstudies/app_b_synthetichydrology_001.pdf.

Veilleux, A.G., 2009, Bayesian GLS regression for regionalization of hydrologic statistics, floods, and Bulletin 17B skew: Ithaca, New York, Cornell University, M.S. Thesis, August, 170 p.

Vogel, R.M., and J.R., Stedinger, 1985, Minimum variance streamflow record augmentation procedures: Water Resources Research, v. 21, no. 5, p. 715–723.

Waananen, A.O., and Crippen, J.R., 1977, Magnitude and frequency of floods in California: U.S. Geological Survey Water-Resources Investigations Report 77-21, 96 p..

Weaver, J.C., Feaster, T.D., and Gotvald, A.J., 2009, Magnitude and frequency of rural floods in the southeastern United States, 2006—Volume 2, North Carolina: U.S. Geological Survey Scientific Investigations Report 2009–5158, 113 p. Available at http://pubs.usgs.gov/sir/2009/5158/.

Table 1. Streamflow-gaging stations and statistical data used to analyze regional skew in California and to determine flood frequency for the Sacramento-San Joaquin River Basin, California.

Table 1 is available in a Microsoft© Excel spreadsheet and can be accessed and downloaded at URL http://pubs.usgs.gov/sir/2010/5260.

Table 2. Basin characteristics for analyzing regional skew in California.

[**Abbreviations**: DRNAREA is the drainage area of the basin. BASINPERIM is the perimeter of the basin. DEM is digital elevation model. NHDPlus is the National Hydrography Dataset. RELIEF is the difference between the maximum and minimum elevations in the basin. ELEVMAX and ELEVMIN are the maximum and minimum elevations in the basin, respectively. LAKEAREA is the percentage of the basin drainage area covered by lakes and ponds. EL6000 is the percentage of the basin above 6,000 feet in elevation. OUTLETELEV is the basin elevation at the gage. RELRELF is the basin RELIEF divided by the BASINPERIM. DIST2COAST is the distance from the gage to the Pacific Ocean measured perpendicular to the eastern border of California. ELEV is the mean basin elevation. BSLDEM30 is the average basin slope computed from a 30-meter digital elevation model (DEM). FOREST is the percentage of the basin covered by forest. IMPERV is the percentage of the basin covered by impervious area. PRECIP is the basin averaged mean annual precipitation. PRISM is the Parameter-Elevation Regressions on Independent Slopes Model. JANMAX is the basin averaged January maximum temperature. JANMIN is the basin averaged January minimum temperature. LAT_CENT is the latitude of the basin centroid. LONG_CENT is the longitude of the basin centroid. m, meter; >, greater than]

Name	Description	Data source
DRNAREA	Area, in square miles	30-m DEM, NHDPlus elev_cm grid http://www.horizon-systems.com/NHDPlus/
BASINPERIM	Distance, in miles	30-m DEM, NHDPlus elev_cm grid http://www.horizon-systems.com/NHDPlus/
RELIEF	Relief, in feet	30-m DEM, NHDPlus elev_cm grid http://www.horizon-systems.com/NHDPlus/
ELEVMAX	Maximum elevation, in feet	30-m DEM, NHDPlus elev_cm grid http://www.horizon-systems.com/NHDPlus/
ELEVMIN	Minimum elevation, in feet	30-m DEM, NHDPlus elev_cm grid http://www.horizon-systems.com/NHDPlus/
LAKEAREA	Percentage of area covered by lakes and ponds	2001 National Land Cover Database (NLCD) - Land Cover http://www.mrlc.gov/nlcd_multizone_map.php
EL6000	High Elevation Index - Percentage of area with elevation >6,000 feet	30-m DEM, NHDPlus elev_cm grid http://www.horizon-systems.com/NHDPlus/
OUTLETELEV	Elevation at outlet, in feet	30-m DEM, NHDPlus elev_cm grid http://www.horizon-systems.com/NHDPlus/
RELRELF	Relative relief, in feet per mile	30-m DEM, NHDPlus elev_cm grid http://www.horizon-systems.com/NHDPlus/
DIST2COAST	Distance, in miles, from basin centroid to coast along a line perpendicular to the eastern California border	30-m DEM, NHDPlus elev_cm grid http://www.horizon-systems.com/NHDPlus/
ELEV	Average basin elevation, in feet	30-m DEM, NHDPlus elev_cm grid http://www.horizon-systems.com/NHDPlus/
BSLDEM30M	Average basin slope, in percent	30-m DEM, NHDPlus elev_cm grid http://www.horizon-systems.com/NHDPlus/
FOREST	Percentage of basin covered by forest	2001 National Land Cover Database (NLCD) - Percent Canopy http://www.mrlc.gov/nlcd_multizone_map.php
IMPERV	Percentage of basin covered by impervious surface	2001 National Land Cover Database (NLCD) - Percent Impervious http://www.mrlc.gov/nlcd_multizone_map.php
PRECIP	Mean annual precipitation, in inches	800M resolution PRISM 1971-2000 data http://www.prism.oregonstate.edu/products/
JANMAX	Average maximum January temperature, in degrees Fahrenheit	800M resolution PRISM 1971-2000 data http://www.prism.oregonstate.edu/products/
JANMIN	Average minimum January temperature, in degrees Fahrenheit	800M resolution PRISM 1971-2000 data http://www.prism.oregonstate.edu/products/
LONG_CENT	Longitude of the basin centroid, in degrees	30-m DEM, NHDPlus elev_cm grid http://www.horizon-systems.com/NHDPlus/
LAT_CENT	Latitude of the basin centroid, in degrees	30-m DEM, NHDPlus elev_cm grid http://www.horizon-systems.com/NHDPlus/

Table 3. Basin characteristics for sites used in the regional skew analysis, California.

Table 3 is available in a Microsoft Excel spreadsheet and can be accessed and downloaded at URL http://pubs.usgs.gov/sir/2010/5260.

Table 4. Results of trend tests for annual peak discharge at selected sites in California.

Site number	Station number	Station name	Kendall's tau	p-value
2	10296500	West Walker River near Coleville, California	0.20	0.12
4	10336676	Ward Creek at Hwy 89 near Tahoe Pines, California	0.10	0.43
6	10343500	Sagehen Creek near Truckee, California	0.03	0.86
15	11015000	Sweetwater River near Descanso, California	–0.17	0.20
16	11028500	Santa Maria Creek near Ramona, California	–0.16	0.21
18	11042400	Temecula Creek near Aguanga, California	–0.23	0.08
20	11055500	Plunge Creek near East Highlands, California	–0.12	0.35
21	11055800	City Creek near Highland, California	–0.03	0.84
26	11075800	Santiago Creek at Modjeska, California	–0.18	0.16
29	11098000	Arroyo Seco near Pasadena, California	–0.04	0.76
42	11124500	Santa Cruz Creek near Santa Ynez, California	0.12	0.37
44	11132500	Salsipuedes Creek near Lompoc, California	0.09	0.48
47	11136800	Cuyama River below Buckhorn Canyon near Santa Maria, California	–0.03	0.86
51	11141280	Lopez Creek near Arroyo Grande, California	0.12	0.35
52	11143000	Big Sur River near Big Sur, California	0.08	0.53
53	11143200	Carmel River at Robles del Rio, California	0.10	0.46
56	11147500	Salinas River at Paso Robles, California	0.10	0.46
57	11148500	Estrella River near Estrella, California	–0.11	0.38
58	11148900	Nacimiento River below Sapaque Creek near Bryson, California	0.08	0.54
59	11151300	San Lorenzo Creek below Bitterwater Creek near King City, California	0.05	0.72
60	11152000	Arroyo Seco near Soledad, California	0.13	0.34
62	11152600	Gabilan Creek near Salinas, California	0.17	0.20
66	11159200	Corralitos Creek at Freedom, California	0.24	0.06
67	11160000	Soquel Creek at Soquel, California	0.21	0.10
68	11160500	San Lorenzo River at Big Trees, California	0.13	0.34
69	11162500	Pescadero Creek near Pescadero, California	0.18	0.18
71	11162630	Pilarcitos Creek at Half Moon Bay, California	0.13	0.34
72	11164500	San Francisquito Creek at Stanford University, California	0.14	0.27
73	11169500	Saratoga Creek at Saratoga, California	0.03	0.83
75	11176400	Arroyo Valle below Lang Canyon near Livermore, California	0.11	0.41
77	11182500	San Ramon Creek at San Ramon, California	0.07	0.59
96	11200800	Deer Creek near Fountain Springs, California	0.08	0.53
126	11224500	Los Gatos Creek above Nunez Canyon near Coalinga, California	0.00	1.00
135	11237500	Pitman Creek below Tamarack Creek, California	0.05	0.69
136	11242400	North Fork Willow Creek near Sugar Pine, California	0.00	0.99
142	11253310	Cantua Creek near Cantua Creek, California	–0.04	0.78
164	11274500	Orestimba Creek near Newman, California	0.09	0.47
165	11274630	Del Puerto Creek near Patterson, California	0.12	0.34
175	11284400	Big Creek above Whites Gulch near Groveland, California	0.02	0.90
185	11294500	North Fork Stanislaus River near Avery, California	–0.07	0.58
205	11316800	Forest Creek near Wilseyville, California	0.00	0.99
206	11317000	Middle Fork Mokelumne River at West Point, California	0.06	0.67
207	11318500	South Fork Mokelumne River near West Point, California	0.03	0.83
226	11342000	Sacramento River at Delta, California	0.13	0.33
231	11348500	Pit River near Canby, California	0.10	0.43
251	11374000	Cow Creek near Millville, California	0.00	0.99
256	11376000	Cottonwood Creek near Cottonwood, California	0.11	0.41
267	11381500	Mill Creek near Los Molinos, California	–0.04	0.79
270	11383500	Deer Creek near Vina, California	0.04	0.75
274	11390000	Butte Creek near Chico, California	0.10	0.45

Table 4. Results of trend tests for annual peak discharge at selected sites in California.—Continued

Site number	Station number	Station name	Kendall's tau	p-value
297	11402000	Spanish Creek above Blackhawk Creek at Keddie, California	0.05	0.71
324	11427700	Duncan Canyon Creek near French Meadows, California	0.05	0.71
325	11431800	Pilot Creek above Stumpy Meadows Reservoir, California	0.03	0.84
332	11439500	South Fork American River near Kyburz (river only), California	0.02	0.90
341	11449500	Kelsey Creek near Kelseyville California	0.14	0.29
346	11461000	Russian River near Ukiah, California	–0.02	0.89
348	11468000	Navarro River near Navarro, California	0.07	0.59
349	11468500	Noyo River near Fort Bragg, California	0.11	0.40
350	11469000	Mattole River near Petrolia, California	–0.02	0.89
353	11475560	Elder Creek near Branscomb, California	0.15	0.26
354	11476500	South Fork Eel River near Miranda, California	–0.11	0.38
355	11477000	Eel River at Scotia, California	0.06	0.63
356	11478500	Van Duzen River near Bridgeville, California	0.18	0.18
357	11481200	Little River near Trinidad, California	0.04	0.78
359	11482500	Redwood Creek at Orick, California	0.09	0.49
360	11519500	Scott River near Fort Jones, California	–0.03	0.80
362	11522500	Salmon River at Somes Bar, California	–0.03	0.80
363	11523200	Trinity River above Coffee Creek near Trinity Center, California	0.10	0.45
365	11532500	Smith River near Crescent City, California	0.04	0.78

Table 5. Key dam sites and drainage areas, periods of estimated unregulated annual-maximum-daily discharge, and periods of concurrent unregulated annual-maximum-daily and peak-discharge data.

[ACOE, U.S. Army Corps of Engineers; USGS, U.S. Geological Survey]

Station number [1]	Station name as shown in USGS or ACOE records	Drainage area (square mile)	Period of estimated unregulated, annual-maximum-daily-discharge record	Period of concurrent unregulated annual-maximum-daily and peak-discharge record
11222099	Kings River at Piedra, California	1,681	1896–1999	1900–02, 1904–07, 1909–50
11251099	San Joaquin River below Friant, California	1,678	1911–99	
11258099	Fresno River below Hidden Dam, California	258	1942–99	1942–75
11259099	Chowchilla River below Buchanan Dam, California	235	1922–23, 1932–99	1932–72
11270099	Merced River at Exchequer, California	1,038	1902–14, 1916–99	1902–13, 1916–25
11288099	Tuolumne River above La Grange Dam, California	1,532	1897–1999	
11299599	Stanislaus River below Melones Dam, California	904	1932–99	
11308999	Calaveras River below New Hogan Dam, California	373	1964–98	
11323599	Mokelumne River below Camanche Dam, California	628	1905–97	1905–28
11335099	Cosumnes River at Michigan Bar, California	535	1908–97	1908–54
11344099	Littlejohns Creek below Farmington Reservoir, California [2]	208	1951–99	
11370599	Sacramento River at Keswick, California [3]	6,468	1932–98	1932–43
11388099	Stony Creek below Black Butte Dam, California	742	1964–98	
11407099	Feather River at Oroville, California	3,624	1902–97	1902–10
11413599	North Yuba River below Bullards Bar Dam, California	487	1941–66, 1970–97	1941–66
11446599	American River at Fair Oaks, California	1,888	1905–98	1905–54

[1] Station number is the USGS station number with the last two digits (usually zeros) replaced by 99.

[2] All available flow data collected by ACOE.

[3] Peak-flow records available at this site from 1939-P. Peak-flow records from upstream site at Kennett (11369500—drainage area = 6,355 square miles) for 1926–1938 were also used at this site.

Table 6. Unregulated, annual-maximum-daily discharge for key dam sites, California.

[Discharge is in cubic feet per second]

Kings River		San Joaquin River		Fresno River		Chowchilla River	
11222099		11251099		11258099		11259099	
Date	Discharge	Date	Discharge	Date	Discharge	Date	Discharge
1896	9,400						
1897	17,800						
1898	8,140						
1899	24,000						
1/3/1900	13,400						
1/7/1901	33,200						
4/7/1902	20,800						
5/13/1903	15,000						
5/16/1904	13,800						
6/13/1905	9,780						
6/20/1906	24,900						
6/3/1907	16,200						
4/29/1908	6,460						
6/4/1909	20,300						
1/1/1910	14,700						
1/31/1911	20,500	1/31/1911	41,069				
5/30/1912	12,400	6/5/1912	15,481				
5/23/1913	7,210	5/28/1913	6,865				
1/26/1914	30,400	1/26/1914	27,426				
6/1/1915	16,300	6/9/1915	15,249				
6/9/1916	16,300	5/6/1916	13,559				
6/9/1917	13,200	6/10/1917	13,812				
6/13/1918	12,800	6/11/1918	11,618				
5/29/1919	11,200	10/2/1918	11,217				
5/20/1920	14,900	5/21/1920	12,867				
6/8/1921	12,800	6/11/1921	12,445				
6/5/1922	17,100	6/5/1922	18,184			2/11/1922	3,620
5/16/1923	11,500	5/17/1923	11,653			4/10/1923	1,990
5/8/1924	3,930	5/9/1924	4,199				
5/26/1925	9,240	5/27/1925	9,928				
5/5/1926	9,490	5/5/1926	10,095				
5/17/1927	14,000	5/17/1927	13,778				
5/15/1928	6,750	3/25/1928	11,047				
6/16/1929	9,560	6/16/1929	8,669				
5/28/1930	7,070	6/13/1930	6,594				
5/7/1931	5,030	5/7/1931	4,358				
5/17/1932	12,800	2/7/1932	14,642			12/28/1931	4,520
6/14/1933	11,900	6/15/1933	11,290			1/30/1933	331
12/13/1933	4,690	4/16/1934	3,588			2/23/1934	1,010

Table 6. Unregulated, annual-maximum-daily discharge or key dam sites, California.—Continued

[Discharge is in cubic feet per second]

Kings River		San Joaquin River		Fresno River		Chowchilla River	
11222099		11251099		11258099		11259099	
Date	Discharge	Date	Discharge	Date	Discharge	Date	Discharge
6/4/1935	12,500	6/5/1935	13,637			4/8/1935	2,980
5/14/1936	11,600	5/14/1936	10,690			2/23/1936	3,530
2/6/1937	25,600	2/6/1937	18,639			2/6/1937	8,890
12/11/1937	37,800	12/11/1937	40,497			2/11/1938	7,760
4/22/1939	5,780	4/22/1939	5,341			3/10/1939	525
5/15/1940	12,000	5/13/1940	11,205			1/26/1940	3,340
6/6/1941	15,300	5/24/1941	16,145			2/12/1941	5,080
5/25/1942	14,000	5/26/1942	14,652	12/29/1941	1,680	12/29/1941	2,730
1/22/1943	16,200	1/22/1943	17,074	1/22/1943	2,300	3/10/1943	2,760
5/9/1944	8,490	5/9/1944	8,250	3/4/1944	1,190	3/4/1944	1,260
2/2/1945	32,200	2/2/1945	34,376	2/2/1945	4,610	2/2/1945	5,200
5/7/1946	11,600	5/7/1946	10,959	3/30/1946	1,840	3/30/1946	1,990
5/6/1947	9,580	5/3/1947	8,716	12/27/1946	794	11/23/1946	649
5/16/1948	10,400	5/27/1948	10,615	4/10/1948	1,450	4/10/1948	2,490
5/27/1949	9,240	5/27/1949	9,864	3/4/1949	942	3/4/1949	1,760
5/31/1950	10,500	5/31/1950	10,664	2/6/1950	807	2/6/1950	1,720
11/19/1950	51,600	11/19/1950	42,352	11/19/1950	5,130	11/19/1950	6,000
6/6/1952	15,500	5/28/1952	18,149	1/25/1952	3,760	1/25/1952	4,950
6/16/1953	7,390	4/27/1953	8,460	12/31/1952	819	1/14/1953	1,090
5/19/1954	11,596	5/19/1954	10,744	1/25/1954	455	2/14/1954	914
6/8/1955	11,445	6/9/1955	11,572	5/8/1955	312	1/2/1955	448
12/23/1955	77,955	12/23/1955	74,984	12/23/1955	10,400	12/23/1955	18,400
5/19/1957	14,493	5/19/1957	16,847	5/19/1957	672	2/25/1957	600
6/19/1958	14,610	5/19/1958	17,541	4/3/1958	6,700	4/3/1958	7,250
2/16/1959	5,663	2/16/1959	7,928	2/16/1959	630	2/16/1959	914
5/12/1960	5,076	5/12/1960	6,612	2/9/1960	430	2/10/1960	1,100
5/24/1961	3,350	5/24/1961	3,850	12/2/1960	122	12/2/1960	122
5/6/1962	12,746	2/10/1962	13,828	2/11/1962	4,430	2/10/1962	4,620
2/1/1963	35,491	2/1/1963	40,982	2/1/1963	3,540	2/1/1963	4,190
5/21/1964	7,152	5/20/1964	6,655	11/21/1963	200	11/20/1963	397
12/24/1964	15,640	12/23/1964	25,531	12/23/1964	1,970	12/23/1964	3,130
5/6/1966	7,837	5/7/1966	7,445	12/30/1965	625	12/30/1965	1,400
12/6/1966	71,711	12/6/1966	42,394	4/18/1967	4,030	12/6/1966	3,920
5/29/1968	5,689	5/28/1968	5,454	2/21/1968	233	2/18/1968	207
1/25/1969	48,816	1/25/1969	29,324	2/24/1969	7250	2/24/1969	7,010
1/16/1970	16,166	1/16/1970	15,961	1/16/1970	2,330	1/16/1970	4,230
5/16/1971	7,309	6/13/1971	8,583	12/2/1970	252	12/22/1970	348
6/7/1972	5,382	6/8/1972	7,122	12/26/1971	300	2/6/1972	367
5/18/1973	17,501	5/18/1973	17,091	2/11/1973	4,500	2/11/1973	5,160
6/7/1974	14,528	5/28/1974	13,622	4/2/1974	3,240	4/2/1974	3,960
6/1/1975	15,813	6/1/1975	16,941	3/26/1975	740	2/10/1975	1,320

Table 6. Unregulated, annual-maximum-daily discharge or key dam sites, California.—Continued

[Discharge is in cubic feet per second]

Kings River		San Joaquin River		Fresno River		Chowchilla River	
11222099		11251099		11258099		11259099	
Date	Discharge	Date	Discharge	Date	Discharge	Date	Discharge
9/11/1976	5,141	5/14/1976	4,864	3/1/1976	176	3/1/1976	216
6/9/1977	4,427	6/9/1977	5,208	1/3/1977	43	1/3/1977	27
6/9/1978	20,898	6/8/1978	19,472	3/4/1978	4,403	2/9/1978	4,784
5/21/1979	14,574	5/22/1979	14,271	3/28/1979	2,104	3/28/1979	2,410
1/13/1980	36,296	1/14/1980	32,001	1/14/1980	3,524	1/14/1980	3,009
5/2/1981	9,161	5/2/1981	8,314	3/20/1981	473	1/29/1981	1,084
4/11/1982	52,007	4/11/1982	59,295	4/11/1982	5,698	1/5/1982	7,514
12/22/1982	28,808	12/22/1982	26,580	1/27/1983	5,665	12/22/1982	7,167
12/25/1983	13,573	12/25/1983	18,185	12/27/1983	1,910	12/25/1983	2,571
4/15/1985	7,461	4/15/1985	6,985	3/28/1985	519	2/9/1985	822
2/18/1986	28,060	2/19/1986	33,515	2/18/1986	5,817	2/18/1986	6,786
5/16/1987	6,014	5/16/1987	7,978	2/13/1987	343	2/13/1987	450
1/5/1988	7,791	5/16/1988	5,672	4/20/1988	203	3/2/1988	132
4/15/1989	6,024	5/9/1989	6,671	3/3/1989	353	3/26/1989	472
5/7/1990	4,666	5/6/1990	4,883	1/14/1990	78	1/14/1990	63
3/4/1991	13,564	6/4/1991	8,989	3/19/1991	1,161	3/19/1991	1,408
4/30/1992	5,830	5/8/1992	6,002	2/13/1992	647	2/15/1992	1,392
1/14/1993	16,489	5/24/1993	15,323	1/14/1993	7,203	1/14/1993	7,574
5/13/1994	7,830	5/12/1994	6,952	2/18/1994	170	2/19/1994	141
3/10/1995	29,959	3/10/1995	39,333	3/11/1995	8,611	3/11/1995	7,982
5/16/1996	28,742	5/16/1996	32,217	2/20/1996	1,888	2/5/1996	2,442
1/2/1997	53,390	1/2/1997	77,467	1/2/1997	7,718	1/2/1997	7,957
6/16/1998	20,884	6/16/1998	20,199	3/25/1998	3,327	3/25/1998	3,920
5/26/1999	8,927	5/27/1999	10,800	2/9/1999	722	2/9/1999	1,100

Table 6. Unregulated, annual-maximum-daily discharge for key dam sites, California.—Continued

[Discharge is in cubic feet per second]

Merced River		Tuolumne River		Stanislaus River		Calaveras River	
11270099		11288099		11299599		11308999	
Date	Discharge	Date	Discharge	Date	Discharge	Date	Discharge
		5/25/1897	16,200				
		4/24/1898	7,600				
		3/25/1899	26,800				
		1/3/1900	14,200				
		2/19/1901	22,560				
5/28/1902	5,420	4/7/1902	12,088				
4/1/1903	12,000	4/1/1903	19,740				
5/14/1904	10,100	3/20/1904	16,365				
3/19/1905	8,500	10/11/1904	14,551				
1/19/1906	19,800	3/24/1906	26,200				
3/19/1907	24,400	3/19/1907	50,425				
4/29/1908	3,900	4/30/1908	6,478				
1/14/1909	20,400	1/14/1909	26,719				
12/9/1909	14,800	12/9/1909	20,900				
1/30/1911	37,200	1/30/1911	52,560				
6/3/1912	6,100	6/5/1912	13,800				
5/23/1913	3,130	5/16/1913	7,590				
11/20/1913	204	1/25/1914	31,300				
		5/13/1915	15,300				
3/5/1916	12,600	3/20/1916	17,100				
2/22/1917	18,500	2/21/1917	23,000				
3/19/1918	14,300	3/12/1918	15,200				
5/29/1919	8,740	5/29/1919	13,801				
5/20/1920	8,320	5/21/1920	12,971				
1/18/1921	13,000	6/8/1921	12,424				
2/11/1922	13,300	6/5/1922	18,376				
4/6/1923	8,980	5/15/1923	14,326				
5/3/1924	2,490	5/3/1924	8,762				
2/6/1925	9,280	2/7/1925	17,998				
2/14/1926	6,360	4/26/1926	14,437				
5/17/1927	8,640	2/19/1927	16,305				
3/25/1928	15,973	3/25/1928	43,351				
6/16/1929	5,819	6/17/1929	15,279				
5/21/1930	3,962	6/12/1930	8,407				
5/7/1931	2,435	5/7/1931	5,624				
2/7/1932	11,430	2/7/1932	22,955	2/7/1932	9,053		
5/31/1933	6,039	6/15/1933	13,144	5/29/1933	8,590		
1/1/1934	4,632	1/2/1934	6,227	3/30/1934	2,681		
4/8/1935	21,991	4/8/1935	23,450	4/8/1935	10,433		
2/23/1936	14,712	2/23/1936	21,865	2/22/1936	14,475		
2/6/1937	25,203	2/7/1937	24,389	5/14/1937	9,584		

Table 6. Unregulated, annual-maximum-daily discharge for key dam sites, California.—Continued

[Discharge is in cubic feet per second]

Merced River		Tuolumne River		Stanislaus River		Calaveras River	
11270099		11288099		11299599		11308999	
Date	Discharge	Date	Discharge	Date	Discharge	Date	Discharge
12/11/1937	33,964	12/12/1937	74,424	12/12/1937	36,983		
4/12/1939	3,469	5/1/1939	6,185	4/9/1939	3,842		
2/27/1940	11,344	3/31/1940	29,396	3/31/1940	20,741		
12/27/1940	13,575	12/28/1940	17,575	5/12/1941	10,555		
5/26/1942	8,470	12/3/1941	22,728	5/26/1942	10,671		
1/23/1943	13,475	1/22/1943	23,155	3/10/1943	20,796		
5/9/1944	5,482	5/9/1944	9,772	5/10/1944	6,636		
2/2/1945	33,046	2/2/1945	45,472	2/3/1945	18,982		
12/22/1945	11,592	12/22/1945	18,914	5/6/1946	7,562		
5/3/1947	5,290	5/3/1947	9,972	5/4/1947	4,688		
5/27/1948	6,761	5/27/1948	12,613	5/27/1948	9,480		
5/14/1949	6,275	5/14/1949	11,734	5/14/1949	8,633		
5/28/1950	6,061	5/31/1950	12,056	5/22/1950	7,706		
11/19/1950	46,545	11/19/1950	67,047	11/19/1950	58,648		
1/25/1952	13,320	6/5/1952	16,997	5/28/1952	11,889		
6/6/1953	5,385	4/27/1953	14,511	4/28/1953	10,543		
5/9/1954	5,769	3/9/1954	17,015	3/10/1954	11,428		
5/23/1955	5,340	6/12/1955	10,974	5/23/1955	5,448		
12/23/1955	74,838	12/23/1955	121,555	12/23/1955	80,805		
5/19/1957	9,274	5/19/1957	18,042	5/19/1957	12,588		
4/3/1958	19,784	4/3/1958	18,867	4/3/1958	13,127		
2/16/1959	6,701	2/16/1959	10,985	5/14/1959	2,933		
2/9/1960	6,182	2/9/1960	10,533	2/9/1960	7,866		
4/7/1961	2,828	5/23/1961	5,097	4/4/1961	2,342		
2/10/1962	12,894	2/10/1962	16,487	5/6/1962	6,745		
2/1/1963	38,354	2/1/1963	70,087	2/1/1963	38,248		
5/20/1964	3,384	11/15/1963	10,602	5/20/1964	3,872	1/22/1964	2,623
12/24/1964	33,093	12/23/1964	72,700	12/24/1964	43,062	12/23/1964	12,789
11/24/1965	6,548	11/24/1965	10,075	4/2/1966	3,794	12/30/1965	2,020
12/7/1966	17,197	12/6/1966	30,014	5/24/1967	16,162	1/22/1967	6,738
5/1/1968	2,856	2/20/1968	9,032	2/21/1968	4,963	2/21/1968	1,647
1/21/1969	33,467	1/21/1969	49,822	1/21/1969	23,881	1/21/1969	14,674
1/16/1970	14,724	1/16/1970	30,681	1/22/1970	28,084	1/21/1970	7,200
5/16/1971	4,940	6/8/1971	9,896	6/27/1971	9,492	12/2/1970	2,983
6/8/1972	3,808	5/30/1972	8,402	5/16/1972	4,200	12/25/1971	4,922
2/11/1973	12,763	5/19/1973	16,736	5/20/1973	12,498	1/16/1973	7,695
4/1/1974	11,061	11/12/1973	23,383	4/2/1974	9,734	3/2/1974	9,124
6/1/1975	10,329	6/12/1975	62,744	6/2/1975	18,115	3/25/1975	5,783
5/13/1976	2,380	10/27/1975	5,851	10/27/1975	1,831	3/2/1976	240
6/10/1977	2,238	9/8/1977	5,778	6/4/1977	940	3/16/1977	112
2/9/1978	15,068	3/4/1978	18,298	3/5/1978	22,572	3/5/1978	5,770

Table 6. Unregulated, annual-maximum-daily discharge for key dam sites, California.—Continued

[Discharge is in cubic feet per second]

Merced River		Tuolumne River		Stanislaus River		Calaveras River	
11270099		11288099		11299599		11308999	
Date	Discharge	Date	Discharge	Date	Discharge	Date	Discharge
1/11/1979	13,504	1/11/1979	20,027	5/22/1979	9,164	2/22/1979	5,388
1/13/1980	31,413	1/14/1980	60,099	1/13/1980	45,520	1/14/1980	8,648
4/30/1981	4,738	5/1/1981	9,017	4/24/1981	5,590	1/29/1981	3,160
4/11/1982	39,147	2/16/1982	49,812	2/15/1982	42,930	1/5/1982	12,321
12/22/1982	21,355	3/1/1983	26,349	3/13/1983	18,340	3/13/1983	10,433
12/25/1983	17,797	12/26/1983	28,544	12/26/1983	18,700	12/25/1983	8,029
4/14/1985	4,255	4/15/1985	9,011	4/15/1985	5,130	2/8/1985	3,769
2/17/1986	31,917	2/19/1986	57,956	2/19/1986	45,320	2/17/1986	23,494
5/16/1987	3,075	2/13/1987	7,268	2/13/1987	3,460	3/6/1987	1,761
5/16/1988	2,726	5/16/1988	5,798	4/14/1988	1,582	1/17/1988	403
4/10/1989	4,085	3/8/1989	12,046	3/8/1989	6,570	3/25/1989	927
4/28/1990	2,707	10/24/1989	6,213	4/23/1990	2,263	2/17/1990	695
5/25/1991	4,746	3/5/1991	10,952	6/4/1991	5,580	3/26/1991	3,939
4/30/1992	4,608	4/18/1992	6,388	2/15/1992	3,150	2/15/1992	5,114
1/14/1993	13,758	1/14/1993	15,188	1/22/1993	8,479	1/13/1993	5,317
5/12/1994	3,153	4/19/1994	7,476	5/31/1994	2,394	2/20/1994	909
3/10/1995	38,212	3/10/1995	46,631	3/10/1995	24,027	3/11/1995	10,146
5/16/1996	16,251	5/16/1996	38,549	5/16/1996	21,081	2/21/1996	5,653
1/2/1997	67,040	1/2/1997	117,709	1/2/1997	72,865	1/2/1997	16,801
2/3/1998	16,813	3/25/1998	28,439	3/24/1998	11,753	2/3/1998	16,919
2/9/1999	8,560	2/9/1999	16,629	2/9/1999	11,017		

Table 6. Unregulated, annual-maximum-daily discharge for key dam sites, California.—Continued

[Discharge is in cubic feet per second]

Mokelumne River		Cosumnes River		Little Johns Creek		Sacramento River	
11323599		11335099		11344099		11370599	
Date	Discharge	Date	Discharge	Date	Discharge	Date	Discharge
5/17/1905	4,940						
6/12/1906	9,000						
3/19/1907	23,000						
4/30/1908	3,020	1/21/1908	2,020				
1/14/1909	12,600	1/14/1909	20,800				
11/21/1909	7,200	3/21/1910	7,200				
1/30/1911	16,700	1/31/1911	22,400				
6/3/1912	4,920	3/7/1912	1,100				
5/18/1913	3,840	1/19/1913	1,220				
1/26/1914	11,100	1/22/1914	13,900				
6/1/1915	7,750	2/2/1915	5,920				
3/20/1916	8,040	3/20/1916	8,920				
6/10/1917	7,550	2/22/1917	13,500				
3/12/1918	6,940	3/12/1918	10,800				
2/11/1919	7,060	2/11/1919	13,100				
5/20/1920	5,500	3/21/1920	3,210				
1/18/1921	7,350	1/18/1921	11,500				
6/3/1922	7,970	2/9/1922	7,970				
5/16/1923	5,430	12/13/1922	9,570				
5/2/1924	1,770	2/8/1924	910				
2/6/1925	9,700	2/6/1925	15,200				
4/8/1926	3,100	2/12/1926	2,950				
5/17/1927	6,160	4/3/1927	8,630				
3/26/1928	20,300	3/25/1928	17,400				
6/16/1929	3,530	2/4/1929	2,800				
5/21/1930	3,319	3/5/1930	4,360				
5/6/1931	2,022	2/19/1931	879				
5/14/1932	5,616	2/6/1932	7,340			12/27/1931	34,921
5/30/1933	5,105	5/30/1933	783			3/28/1933	19,476
3/29/1934	2,858	1/1/1934	4,920			1/2/1934	24,437
4/8/1935	6,214	4/8/1935	11,300			4/8/1935	44,755
2/22/1936	15,034	2/22/1936	15,600			2/22/1936	53,145
5/15/1937	6,194	2/6/1937	7,800			3/13/1937	29,107
12/11/1937	22,970	2/11/1938	15,500			12/11/1937	102,046
4/8/1939	2,570	3/9/1939	1,500			3/13/1939	37,858
3/31/1940	10,740	3/31/1940	16,700			2/28/1940	161,435
5/12/1941	6,214	3/2/1941	4,600			4/4/1941	71,408
1/27/1942	10,397	1/27/1942	14,100			2/6/1942	79,872
3/10/1943	11,380	3/10/1943	18,700			1/23/1943	44,393
5/9/1944	3,964	3/4/1944	4,660			11/21/1943	6,771
2/2/1945	12,891	2/2/1945	13,100			7/30/1945	9,519

Table 6. Unregulated, annual-maximum-daily discharge for key dam sites, California.—Continued

[Discharge is in cubic feet per second]

Mokelumne River		Cosumnes River		Little Johns Creek		Sacramento River	
11323599		11335099		11344099		11370599	
Date	Discharge	Date	Discharge	Date	Discharge	Date	Discharge
12/22/1945	6,133	12/23/1945	8,510			1/4/1946	28,300
5/4/1947	4,046	3/10/1947	2,610			2/12/1947	33,936
5/27/1948	5,415	3/24/1948	3,140			1/7/1948	62,689
5/13/1949	5,108	3/3/1949	9,010			3/19/1949	42,520
6/1/1950	5,227	2/6/1950	5,410			1/23/1950	29,204
11/21/1950	30,862	11/21/1950	16,700	12/8/1950	5,284	10/29/1950	51,803
5/28/1952	7,447	1/12/1952	8,300	3/15/1952	5,019	12/27/1951	77,184
4/27/1953	5,342	4/28/1953	2,630	1/14/1953	725	1/9/1953	81,549
3/10/1954	5,491	3/10/1954	3,020	3/17/1954	723	1/17/1954	61,599
5/23/1955	4,191	1/1/1955	3,160	1/1/1955	3,556	12/6/1954	31,992
12/23/1955	34,657	12/23/1955	32,789	12/24/1955	8,497	12/22/1955	140,880
5/19/1957	7,874	3/5/1957	6,544	3/5/1957	2,232	2/24/1957	77,415
4/3/1958	9,679	4/3/1958	20,090	4/3/1958	7,272	2/24/1958	83,075
2/16/1959	2,719	2/17/1959	2,847	2/16/1959	1,419	1/12/1959	64,496
2/8/1960	5,426	2/8/1960	6,833	2/10/1960	1,402	2/8/1960	65,202
5/23/1961	2,143	3/25/1961	470	2/2/1961	102	1/31/1961	43,389
2/10/1962	5,241	2/15/1962	6,163	2/15/1962	5,086	2/13/1962	70,179
2/1/1963	29,861	2/1/1963	27,560	2/13/1963	3,205	4/14/1963	61,885
5/15/1964	3,321	1/22/1964	2,959	1/22/1964	898	1/20/1964	62,888
12/23/1964	36,173	12/23/1964	29,883	12/26/1964	8,760	12/22/1964	169,171
5/7/1966	2,845	12/31/1965	1,999	1/30/1966	2,071	1/4/1966	37,155
5/23/1967	8,651	1/22/1967	7,148	1/22/1967	4,324	1/29/1967	62,545
2/21/1968	3,352	2/20/1968	3,482	2/21/1968	1,241	2/23/1968	48,582
1/21/1969	15,415	1/21/1969	19,384	1/21/1969	3,707	1/21/1969	91,765
1/21/1970	14,756	1/21/1970	11,475	1/21/1970	3,953	1/23/1970	164,653
3/26/1971	5,335	3/26/1971	5,786	11/29/1970	2,624	3/26/1971	62,285
5/14/1972	3,818	12/25/1971	3,143	12/25/1971	1,267	1/22/1972	38,269
5/18/1973	5,987	1/12/1973	9,692	2/11/1973	5,368	1/16/1973	74,532
11/12/1973	7,905	3/2/1974	6,657	3/2/1974	4,749	1/16/1974	190,847
6/7/1975	6,733	3/25/1975	7,361	3/22/1975	2,742	3/19/1975	56,295
10/27/1975	2,355	3/2/1976	347	9/11/1976	10	2/29/1976	22,709
5/23/1977	1,122	2/23/1977	170	10/1/1976	0	9/29/1977	8,340
5/15/1978	6,211	3/5/1978	6,472	2/9/1978	3,447	1/16/1978	95,610
5/22/1979	6,292	3/1/1979	4,015	2/21/1979	5,080	2/13/1979	30,570
1/13/1980	31,924	1/14/1980	20,311	1/12/1980	4,921	2/18/1980	89,540
4/30/1981	4,873	3/26/1981	3,262	1/29/1981	3,890	1/28/1981	30,480
4/11/1982	24,642	2/16/1982	25,608	3/31/1982	6,522	12/19/1981	81,790
5/26/1983	12,304	3/13/1983	18,455	11/30/1982	6,620	3/13/1983	92,570
12/25/1983	13,559	12/26/1983	14,218	12/25/1983	5,755	12/11/1983	66,160
4/15/1985	3,766	2/8/1985	3,062	2/8/1985	2,411	11/12/1984	23,980
2/17/1986	27,878	2/17/1986	35,933	2/19/1986	9,555	2/17/1986	126,980

Table 6. Unregulated, annual-maximum-daily discharge for key dam sites, California.—Continued

[Discharge is in cubic feet per second]

Mokelumne River		Cosumnes River		Little Johns Creek		Sacramento River	
11323599		11335099		11344099		11370599	
Date	Discharge	Date	Discharge	Date	Discharge	Date	Discharge
4/30/1987	2,689	3/13/1987	1,531	3/6/1987	2,891	3/13/1987	39,673
4/29/1988	2,255	1/17/1988	991	1/18/1988	63	12/6/1987	32,897
3/8/1989	5,492	3/25/1989	5,839	3/4/1989	45	3/9/1989	72,974
5/31/1990	2,273	3/5/1990	1,044	4/16/1990	25	5/27/1990	31,487
6/4/1991	4,654	3/25/1991	3,631	3/26/1991	2,718	3/4/1991	28,971
4/22/1992	3,298	2/15/1992	3,083	2/15/1992	4,517	2/20/1992	35,598
5/17/1993	6,262	1/21/1993	7,560	1/13/1993	2,697	3/17/1993	82,188
5/11/1994	2,790	2/18/1994	929	2/20/1994	281	1/24/1994	17,942
5/1/1995	15,637	3/11/1995	18,236	1/27/1995	4,854	1/9/1995	111,630
5/16/1996	18,015	3/5/1996	7,917	2/21/1996	3,941	2/21/1996	68,733
1/2/1997	76,137	1/2/1997	61,822	1/2/1997	7,777	1/1/1997	215,623
				2/3/1998	11,270	2/3/1998	78,535
				2/9/1999	4,517		

Table 6. Unregulated, annual-maximum-daily discharge for key dam sites, California.—Continued

[Discharge is in cubic feet per second]

Stony Creek		Feather River		North Yuba River		American River	
11388099		11407099		11413599		11446599	
Date	Discharge	Date	Discharge	Date	Discharge	Date	Discharge
		4/6/1902	38,100				
		3/30/1903	93,000				
		2/24/1904	106,000				
		12/30/1904	68,400			3/19/1905	21,200
		1/18/1906	96,300			1/19/1906	44,500
		3/19/1907	187,000			3/19/1907	105,000
		1/21/1908	16,300			12/27/1907	8,460
		1/16/1909	137,000			1/14/1909	98,000
		12/9/1909	31,000			12/2/1909	47,000
		1/31/1911	75,400			1/31/1911	69,100
		1/26/1912	16,400			6/2/1912	11,300
		5/9/1913	14,200			5/10/1913	11,600
		12/31/1913	88,110			1/1/1914	57,700
		5/11/1915	69,049			5/12/1915	41,800
		3/20/1916	43,093			3/20/1916	33,200
		2/25/1917	73,106			2/25/1917	37,600
		3/26/1918	28,566			4/10/1918	12,400
		2/11/1919	46,335			2/11/1919	45,000
		4/16/1920	21,383			4/16/1920	18,800
		11/19/1920	51,792			1/18/1921	32,800
		5/20/1922	35,090			5/18/1922	23,200
		4/6/1923	20,890			12/13/1922	29,800
		2/8/1924	32,786			2/8/1924	10,600
		2/6/1925	51,084			2/6/1925	68,200
		4/8/1926	46,655			4/6/1926	22,700
		2/21/1927	82,287			2/21/1927	48,200
		3/26/1928	125,168			3/25/1928	119,000
		2/4/1929	12,046			2/4/1929	14,800
		12/15/1929	77,702			3/5/1930	18,800
		3/19/1931	9,731			3/19/1931	7,920
		3/20/1932	18,570			2/7/1932	18,900
		5/31/1933	9,235			5/30/1933	12,700
		3/29/1934	16,971			1/2/1934	13,300
		4/8/1935	53,306			4/8/1935	49,300
		2/22/1936	57,064			2/22/1936	46,400
		4/15/1937	19,379			2/14/1937	22,500
		12/11/1937	158,984			12/11/1937	81,100
		3/27/1939	8,309			3/9/1939	8,500
		3/30/1940	134,761			3/30/1940	69,600
		2/11/1941	73,315	2/10/1941	14,400	2/11/1941	26,900
		2/6/1942	89,118	2/6/1942	21,800	1/27/1942	54,600

Table 6. Unregulated, annual-maximum-daily discharge for key dam sites, California.—Continued

[Discharge is in cubic feet per second]

Stony Creek		Feather River		North Yuba River		American River	
11388099		11407099		11413599		11446599	
Date	Discharge	Date	Discharge	Date	Discharge	Date	Discharge
		1/23/1943	65,063	1/21/1943	22,200	1/22/1943	73,800
		3/4/1944	18,745	5/8/1944	4,610	3/4/1944	12,400
		2/2/1945	47,628	2/2/1945	18,600	2/2/1945	70,900
		12/29/1945	46,425	12/29/1945	15,300	12/22/1945	32,400
		2/12/1947	32,263	2/12/1947	10,100	2/13/1947	20,100
		4/17/1948	33,318	4/17/1948	11,500	4/18/1948	17,600
		4/23/1949	14,223	4/23/1949	4,690	3/3/1949	25,500
		2/6/1950	40,459	2/6/1950	10,200	2/6/1950	22,800
		11/21/1950	69,883	11/21/1950	29,900	11/21/1950	132,000
		2/2/1952	47,154	2/2/1952	15,400	2/2/1952	30,500
		1/9/1953	98,847	1/9/1953	25,600	4/28/1953	27,600
		3/10/1954	48,137	3/9/1954	18,100	3/10/1954	36,500
		5/9/1955	11,869	5/9/1955	4,650	5/9/1955	10,528
		12/23/1955	181,528	12/23/1955	57,000	12/23/1955	189,073
		2/24/1957	63,111	2/24/1957	17,200	5/19/1957	36,924
		2/25/1958	76,631	2/25/1958	18,800	4/3/1958	42,302
		2/17/1959	28,718	2/17/1959	6,170	2/17/1959	15,394
		2/8/1960	99,125	2/8/1960	32,600	2/8/1960	63,014
		1/31/1961	15,718	2/10/1961	3,050	4/4/1961	6,914
		2/10/1962	36,020	2/10/1962	11,800	2/10/1962	35,216
		2/1/1963	136,203	2/1/1963	42,000	2/1/1963	152,614
1/20/1964	3,487	1/21/1964	20,465	11/15/1963	5,850	11/15/1963	17,002
12/22/1964	36,993	12/23/1964	178,544	12/22/1964	63,700	12/23/1964	183,242
1/4/1966	15,562	4/10/1966	17,029	4/11/1966	4,040	4/2/1966	8,659
1/21/1967	12,288	1/30/1967	54,276			3/17/1967	36,197
1/14/1968	8,642	2/21/1968	40,151			2/21/1968	24,697
1/20/1969	14,195	1/21/1969	137,082			1/21/1969	83,526
1/24/1970	23,530	1/24/1970	117,684	1/23/1970	39,008	1/22/1970	88,316

Table 6. Unregulated, annual-maximum-daily discharge for key dam sites, California.—Continued

[Discharge is in cubic feet per second]

Stony Creek		Feather River		North Yuba River		American River	
11388099		11407099		11413599		11446599	
Date	Discharge	Date	Discharge	Date	Discharge	Date	Discharge
1/16/1971	12,177	3/26/1971	64,381	3/26/1971	14,839	3/26/1971	34,047
1/23/1972	3,694	2/29/1972	19,994	1/23/1972	7,722	3/4/1972	10,046
1/18/1973	12,304	1/16/1973	48,339	1/16/1973	13,955	1/12/1973	49,291
1/16/1974	26,751	3/30/1974	108,249	3/30/1974	25,317	1/17/1974	40,631
3/7/1975	10,523	2/13/1975	31,924	3/25/1975	11,352	3/25/1975	30,037
2/27/1976	1,560	2/29/1976	12,079	2/29/1976	3,761	10/27/1975	10,389
3/16/1977	376	2/21/1977	4,289	2/22/1977	1,234	6/10/1977	2,359
1/16/1978	23,978	1/16/1978	54,954	1/15/1978	13,573	1/17/1978	31,169
3/27/1979	6,164	2/14/1979	23,413	1/11/1979	7,203	1/12/1979	18,301
1/13/1980	20,968	1/13/1980	137,623	1/13/1980	54,716	1/14/1980	124,915
1/27/1981	9,586	2/14/1981	18,864	1/28/1981	6,519	3/26/1981	15,531
12/19/1981	13,599	12/20/1981	98,900	12/20/1981	40,045	2/16/1982	113,126
3/1/1983	30,136	3/13/1983	98,773	3/13/1983	29,853	3/13/1983	68,791
12/25/1983	18,068	12/25/1983	74,706	12/25/1983	24,655	12/26/1983	65,182
2/8/1985	4,140	2/8/1985	17,547	2/8/1985	6,024	2/8/1985	13,473
2/17/1986	36,446	2/17/1986	217,024	2/17/1986	69,649	2/18/1986	170,960
3/13/1987	3,268	2/13/1987	30,978	2/13/1987	11,885	2/14/1987	11,690
1/4/1988	9,539	12/2/1987	18,756	12/10/1987	4,491	1/17/1988	5,447
3/11/1989	4,376	3/10/1989	86,723	3/10/1989	23,377	3/25/1989	33,949
1/13/1990	2,076	1/13/1990	14,726	5/31/1990	6,092	5/31/1990	7,606
3/4/1991	7,080	3/4/1991	49,728	3/4/1991	21,656	3/5/1991	27,362
2/12/1992	7,192	2/20/1992	24,208	2/20/1992	9,439	2/20/1992	13,266
1/20/1993	21,429	3/18/1993	59,057	1/22/1993	14,230	1/22/1993	34,244
2/7/1994	3,057	3/6/1994	9,457	12/8/1993	2,566	5/11/1994	5,009
1/9/1995	44,984	3/10/1995	134,188	1/14/1995	29,405	3/11/1995	68,260
2/4/1996	9,981	2/5/1996	57,809	2/5/1996	21,784	5/16/1996	54,315
1/1/1997	29,246	1/1/1997	312,893	1/1/1997	87,988	1/2/1997	252,431
2/3/1998	29,811					2/3/1998	41,819

Table 7. Regional skew models for California.

[Constant is the linear regression model with a constant skew. β_0 is the regression model constant, and β_1 and β_2 are regression model coefficients. σ_δ^2 is the model error variance. ASEV is the average sampling error variance. AVP_{new} is the average variance of prediction for a new site. Pseudo R_δ^2 describes the fraction of the variability in the true skews explained by each model (Gruber and others, 2007). Standard deviations are in parentheses. Elev is the linear regression model relating skew to ELEV. NL-Elev is the non-linear model relating skew to ELEV. %, percent]

Model	Model equation	β_0	β_1	β_2	σ_δ^2	ASEV	AVP_{new}	Pseudo R_δ^2
Constant:	$\hat{\gamma} = \beta_0$	–0.23	—	—	0.20	0.03	0.23	0%
		(0.17)			(0.06)			
Elev:	$\hat{\gamma} = \beta_0 + \beta_1(\text{ELEV})$	–0.76	1.4E-04	—	0.12	0.03	0.15	41%
		(0.22)	(3.4E-05)		(0.04)			
NL-Elev:	$\hat{\gamma} = \beta_0 + \beta_2\left\{1 - \exp\left[-(\text{ELEV}/6{,}500)^2\right]\right\}$	–0.62	—	1.3	0.10	0.03	0.14	48%
		(0.19)		(0.31)	(0.04)			

Table 8. Average regional skew, variance of prediction (VP_{new}) and equivalent record length (ERL) for nonlinear regional skew model NL-Elev for various values of mean basin elevation (ELEV), California.

Elevation, in feet	Average regional skew	VP_{new}	ERL
0	–0.62	0.14	65
1,000	–0.59	0.14	65
2,000	–0.50	0.14	62
3,000	–0.37	0.13	58
4,000	–0.21	0.13	55
5,000	–0.04	0.13	53
6,000	0.13	0.14	52
7,000	0.28	0.14	52
8,000	0.40	0.15	53
9,000	0.49	0.16	54
10,000	0.56	0.16	55
11,000	0.61	0.17	55

Glossary

Annual Exceedance Probability The probability, often expressed as a decimal fraction less than 1.0, that an annual peak discharge will be exceeded in a 1-year period. Exceedance probabilities can be expressed in terms of their reciprocals as recurrence intervals or return periods in years.

Annual-Maximum-Daily Discharge The maximum daily mean discharge occurring during a water year.

Annual Peak The maximum instantaneous discharge occurring during a water year.

Basin Centroid The center of a basin area, which is represented by an irregular polygon. For irregularly shaped polygons, the centroid is derived mathematically and represents an approximate "center of gravity."

Bayesian Generalized Least Squares Regression (B-GLS) A form of GLS regression that uses a Bayesian statistical framework to estimate the regression model parameters and the model error variance and the precision of those estimators.

Effective Record Length (ERL) A representation of the precision of an estimator of a streamflow statistic, or in a regression analysis, the at-site variance of prediction. The effective record length represents the length of gaged record required to estimate a flow statistic at a site with the same accuracy as a regional regression for that region.

Mean Basin Elevation (ELEV) The area-weighted average height of the basin above a vertical datum, usually the North American Vertical Datum of 1988 (NAVD 88).

EL6000 The percentage of a basin that is more than 6,000 feet in elevation.

Expected Moments Algorithm (EMA) A method for fitting a probability distribution to annual peak-discharge data using a generalized method of moments that uses different types of censored data. The censored data often occur with historical information, where we have only limited knowledge about floods outside the period of systematic data collection at a gage. For example, we may have information that no floods have overtopped a road built near a gage 50 years before the gage was established. Censored data also may be in the form of interval discharges, where we know that an annual peak discharge was between some lower bound and some upper bound, such as zero and the smallest recordable discharge for a crest-stage gage.

Generalized Least Squares Regression (GLS) A regression method that accounts for differences in the variances and cross correlations of the errors associated with different recorded discharges. Differences in variances can result from differences in the length of record for each site, whereas cross correlations among concurrent annual peak discharges results in cross-correlation between estimated flood statistics, such as quantiles and skewness coefficients.

Historical Flood Information Information about the magnitude of flood flows, including recorded or estimated annual peak discharges, outside the systematic period of record.

Maintenance of Variation Extension (MOVE) A linear regression technique used for filling in missing streamflow data measurements or producing a unique extended streamflow sequence that maintains the mean and variance for the sample.

Mean Square Error (MSE) The average of the squares of the differences between the estimated values and the measured values. This metric represents how closely, on average, an estimated value matches a measured value. Of particular concern in this report is the MSE of of the regional skewness estimator.

Mixed Population Analysis A method for analyzing flood frequency at a site whose annual flood series is a mixture of two populations of flood discharges caused by different hydroclimatic events. In these situations, the frequency curve of annual events can best be described by computing separate curves for each type of event and then statistically combining the separate curves to derive a distribution applicable to the entire annual flood series.

Outlier A data point that departs from the trend of the rest of a data set as described by a distribution or other mathematical relationship.

Skewness Coefficient (γ) A statistical measure of the lack of symmetry in a flood-frequency distribution. Station skew generally is computed from the logarithms of annual peak discharge at a streamflow-gaging station. Because station skew is sensitive to outliers, it may be an unreliable estimate of the true population skew, especially for small samples. For that reason, Bulletin 17B (Interagency Committee on Water Data, 1982) recommends that station skew be weighted with a regional, or generalized, skew that is based on data from many long-term stations to produce at-site flood-frequency estimates.

Standard Error (SE) A measure of the precision of an estimator, equal to the square root of the variance of the sampling error.

Variance of Prediction (VP) A measure of the likely difference between the prediction provided by a regression model and the actual value of the variable.

Variance of Prediction, new site (VPnew) A measure of the likely difference between the prediction provided by a regression model and the actual value of the variable for a new site that was not used to calibrate a model or estimate its parameters.

Weighted Least Squares Regression (WLS) A regression method that accounts for the variation in the errors due to unequal record lengths at gaging stations used to estimate the flood characteristics of interest. WLS incorporates weights associated with each data point into the fitting criterion. The size of the weights corresponds to the precision of the information contained in the record.

100-Year Flood (Q_{100}) An annual peak discharge having an average recurrence interval of 100 years, corresponding to an annual exceedance probability of 0.01.

Appendix A. Move Methods for Estimating Unregulated Annual-Peak-Flow Data at Key Dam Sites

A method based on a linear relation between the logarithms of annual-peak discharge and the logarithms of annual-maximum-daily discharge was used to estimate unregulated, annual-peak discharge for the 16 key dam sites selected by the U.S. Army Corps of Engineers (ACOE). At ten sites, the log-linear relation was used to extend the period of unregulated annual-peak discharge record already available, and at six sites that had no unregulated annual-peak discharge record, a log-linear relation based on data from a nearby site was used to estimate unregulated annual-peak discharge values.

For streamflow record extension, a linear relation between concurrent flows or their logarithms is often used to estimate missing flow values or to extend the record at a short-record station using flow values at a site with a longer record. Hirsch (1982) showed that using simple ordinary least squares (OLS) regression estimators to extend a short record results in a variance for the extended record that is on average smaller than that for the short record. Because of this variance reduction problem, Hirsch (1982) developed maintenance of variance extension methods denoted type I (MOVE.1) and type II (MOVE.2) for extending a short record to preserve the expected variance of the flows at the short-record station. The MOVE.1 procedure estimates the coefficients of the linear transformation for extending the short record using only the period of concurrent record, whereas MOVE.2 uses the entire record at the long-record site to estimate the coefficients. Hirsch (1982) showed that the MOVE.2 method was slightly better and less biased than the simpler MOVE.1 method. Vogel and Stedinger (1985) developed two additional MOVE methods (MOVE.3 and MOVE.4), which also use information from both the shorter concurrent records and the longer record; MOVE.3 and MOVE.4 are intended to ensure that the original and generated flow values together have a sample mean and variance that are identical to the best estimates of the mean and the variance of the flows at the short-record site.

To estimate periods of unregulated annual-peak discharge record from concurrent periods of unregulated annual-maximum-daily discharge in this study, MOVE.1 and MOVE.3 were considered. The general equation for estimating missing values of discharge at a short-record site is

$$Y_i = a + b\left(X_i - \overline{X}_1\right), \tag{A1}$$

where

Y_i is the logarithm of estimated flow for year i at the short-record site,

a is the linear equation constant,

b is the linear equation coefficient,

X_i is the logarithm of flow for year i at the long-record site, and

$\overline{X}_1$ is the mean of the logarithms of flow at the long-record site for the concurrent record (whose length is n_1).

The MOVE.1 equation for estimating missing data is

$$Y_i = \overline{Y}_1 + \frac{S_{Y_1}}{S_{X_1}}\left(X_i - \overline{X}_1\right), \tag{A2}$$

where

$\overline{Y}_1$ is the mean of the logarithms of flow values at the short-record site,

S_{Y_1} is the standard deviation of the logarithms of flow values at the short-record site,

S_{X_1} is the standard deviation of the logarithms of flow values at the long-record site for the period of concurrent record, and the other terms are as defined above.

Thus, for MOVE.1, the linear equation constant and coefficient in equation A1 can be written as

$$a = \overline{Y}_1, \tag{A3}$$

and

$$b = \frac{S_{Y_1}}{S_{X_1}} \tag{A4}$$

where all terms are defined as above.

The MOVE.3 equation can be written in a form similar to the general linear equation A1 as follows:

$$Y_i = a + b\left(X_i - \overline{X}_2\right), \tag{A5}$$

where $\overline{X}_2$ is the mean of the logarithms for the long-record site for the nonconcurrent record period (whose length is n_2) and the other terms are as defined above. For MOVE.3 described by Vogel and Stedinger (1985), the estimates of the constant and the slope in equation A5 are

$$a = \frac{(n_1 + n_2)\hat{\mu}_y - n_1\overline{Y}_1}{n_2}, \tag{A6}$$

and

$$b^2 = \frac{\left[\begin{array}{l}(n_1 + n_2 - 1)\hat{\sigma}_y^{\,2} - (n_1 - 1)S_{Y_1}^{\,2} \\ \quad - n_1(\overline{Y}_1 - \hat{\mu}_y)^2 - n_2(a - \hat{\mu}_y)^2\end{array}\right]}{(n_2 - 1)S_{X_2}^{\,2}}, \tag{A7}$$

where $\hat{\mu}_y$ and $\hat{\sigma}_y$ are the Matalas-Jacobs estimators of the mean and the variance at the short record site. The Matalas-Jacobs estimator for the mean is

$$\hat{\mu}_y = \overline{Y}_1 + \frac{n_2}{n_1 + n_2}\hat{\beta}(\overline{X}_2 - \overline{X}_1), \tag{A8}$$

where $\overline{X}_2$ is defined above for A5 and $\hat{\beta}$ is calculated from

$$\hat{\beta} = \frac{\sum_{i=1}^{n_1}(X_i - \overline{X}_1)(Y_i - \overline{Y}_1)}{\sum_{i=1}^{n_1}(X_i - \overline{X}_1)^2}. \tag{A9}$$

The Matalas-Jacobs estimator for the variance is

$$\hat{\sigma}_y^{\,2} = \frac{1}{n_1 + n_2 - 1}\left\{\begin{array}{l}(n_1 - 1)S_{y_1}^{\,2} + (n_2 - 1)\hat{\beta}^2 S_{x_2}^{\,2} + (n_2 - 1)\alpha^2\left(1 - \hat{\rho}^2\right)S_{y_1}^{\,2} \\ \quad + \dfrac{n_1 n_2}{(n_1 + n_2)}\hat{\beta}^2\left(\overline{X}_2 - \overline{X}_1\right)^2\end{array}\right\}, \tag{A10}$$

where

$$\alpha^2 = \frac{n_2(n_1 - 4)(n_1 - 1)}{(n_2 - 1)(n_1 - 3)(n_1 - 2)}, \tag{A11}$$

and

$$\hat{\rho} = \hat{\beta}\frac{S_{X_1}}{S_{Y_1}}. \tag{A12}$$

To determine whether the simpler MOVE.1 method would produce peak flow estimates that are as reliable as those produced from the MOVE.3 method, the two procedures were compared using data from stations that had long concurrent records of both annual-peak and annual-maximum-daily discharge. The four selected sites, in the Sacramento–San Joaquin River Basin, are listed in table A1 along with their concurrent record lengths and the Pearson cross-correlation between the logs of annual-peak discharge and annual-maximum-daily discharge. For each site, six samples of a 10-year short record period (n_{s1}) were selected and assumed to be the only concurrent records available.

The complete record periods available ($n_1 + n_2$) varied from 76 to 95 years (table A1); for each sample, $n_{s1} = 10$ and n_{s2} varied from 66 to 85 years. The six 10-year samples of concurrent record at each site were determined by using (1) the first 10 years of actual concurrent record, (2) the second 10 years of actual concurrent record, (3) the third 10 years of actual concurrent record, (4) the last 10 years of actual concurrent record, (5) the next to last 10 years of actual concurrent record, and (6) the third to last 10 years of actual concurrent record. For each sample at each site, the MOVE.1 and MOVE.3 methods were used to estimate annual-peak discharges for the n_{s2} years assumed to have no actual peak-discharge record. The means and the standard deviations of the logarithms of the annual-peak discharge for the complete record period (n_{s2} years of estimated peak discharge plus n_{s1} = 10 years of concurrent and recorded peak discharge) were then compared with the means and the standard deviations of the logarithms of the actual recorded peak discharge for the complete record period. The results of the comparisons of the 6 samples from each of the 4 test sites (24 comparisons) are shown as box plots in figure A1. The box plots indicate that both MOVE.1 and MOVE.3 produced estimated means that were slightly biased on the low side for the 24 replicates (median log residuals less than zero and the vertical boxes centered below zero). Both methods produced estimates of standard deviation with little or no bias (median log residuals were very close to zero, and the vertical boxes were centered near zero. Differences between MOVE.1 and MOVE.3 were negligible for both the mean and the standard deviation. Figures A2 and A3 show how well MOVE.1 and MOVE.3 functioned for each of the 24 samples. Figure A2 shows the residual from the actual mean for each sample for both methods, and figure A3 shows the residual from the actual standard deviation for each sample. Both MOVE methods produced positive and negative residuals for the same samples, and the differences between the MOVE.1 and the MOVE.3 residuals for each sample were very small. On the basis of these results, the simpler MOVE.1 method was adopted to estimate annual-peak discharge from annual-maximum-daily discharge in California. Thus, the MOVE.1 method was used for the ten sites that had some concurrent records of unregulated annual-peak discharge and annual-daily-maximum discharge before dam construction.

Table A1. Four long-record sites used to test MOVE.1 and MOVE.3 methods in the Sacramento–San Joaquin River Basin, California.

Station number	Abbreviated names of stations	Period of concurrent annual maximum-daily and peak discharge record	Number of years of concurrent record ($n_1 + n_2$)	Pearson correlation coefficient (*r*) relating concurrent annual maximum-daily discharge and peak discharge
11230500	Bear Creek near Lake Thomas	1922–88, 1992–2006	82	0.96
11317000	MF Mokelumne River at West Point	1912–2006	95	0.98
11338500	Deer Creek near Vina	1913–15, 1921–2006	89	0.95
11390000	Butte Creek near Chico	1931–2006	76	0.98

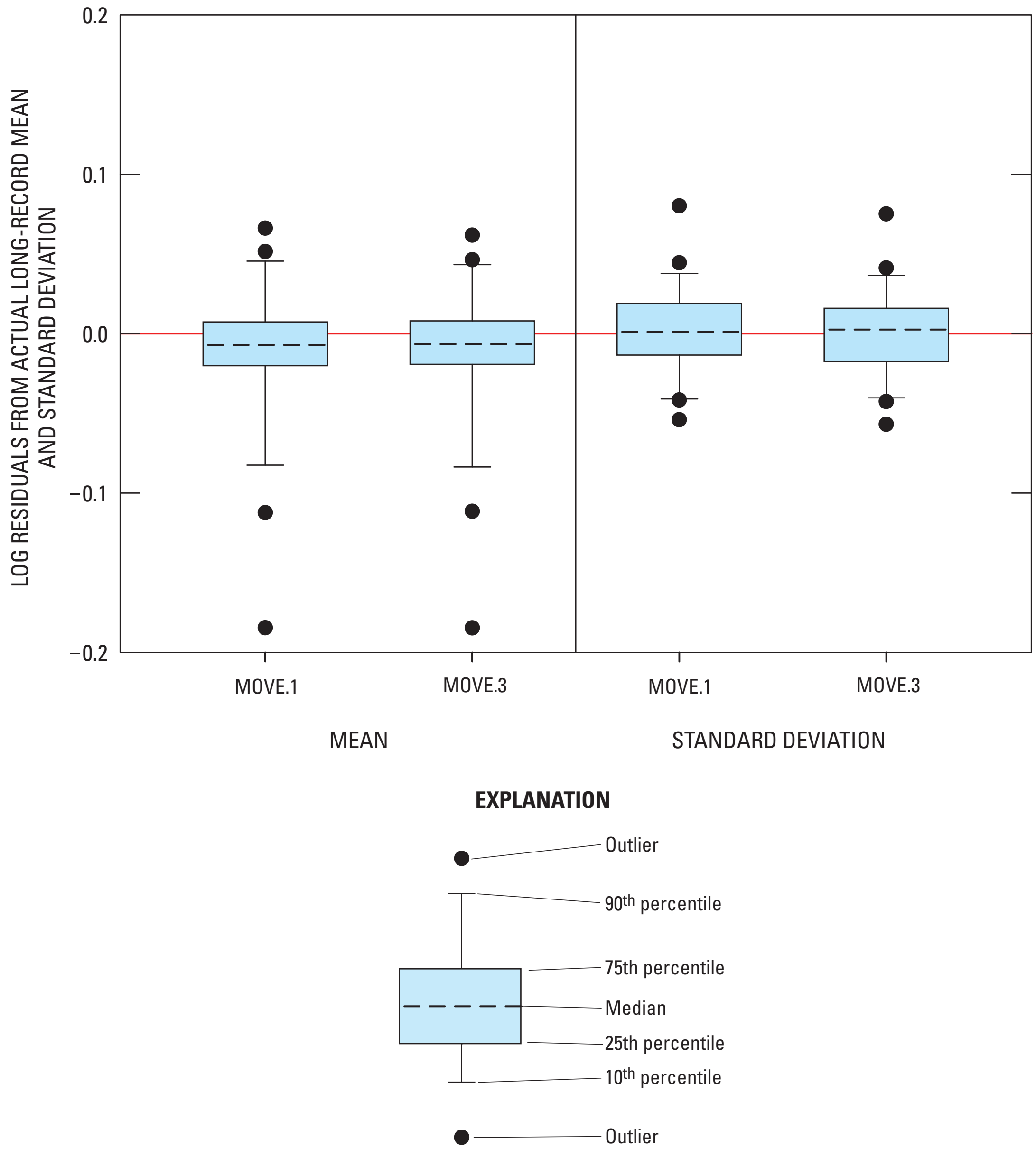


Figure A1. Comparing the MOVE.1 method with the MOVE.3 method to estimate the mean and standard deviations of the logs of annual-peak discharges from the logs of annual-maximum-daily discharge based on data from four sites (24 samples) in California. Each sample had a short, concurrent record of 10 years and an estimation period varying from 66 to 85 years.

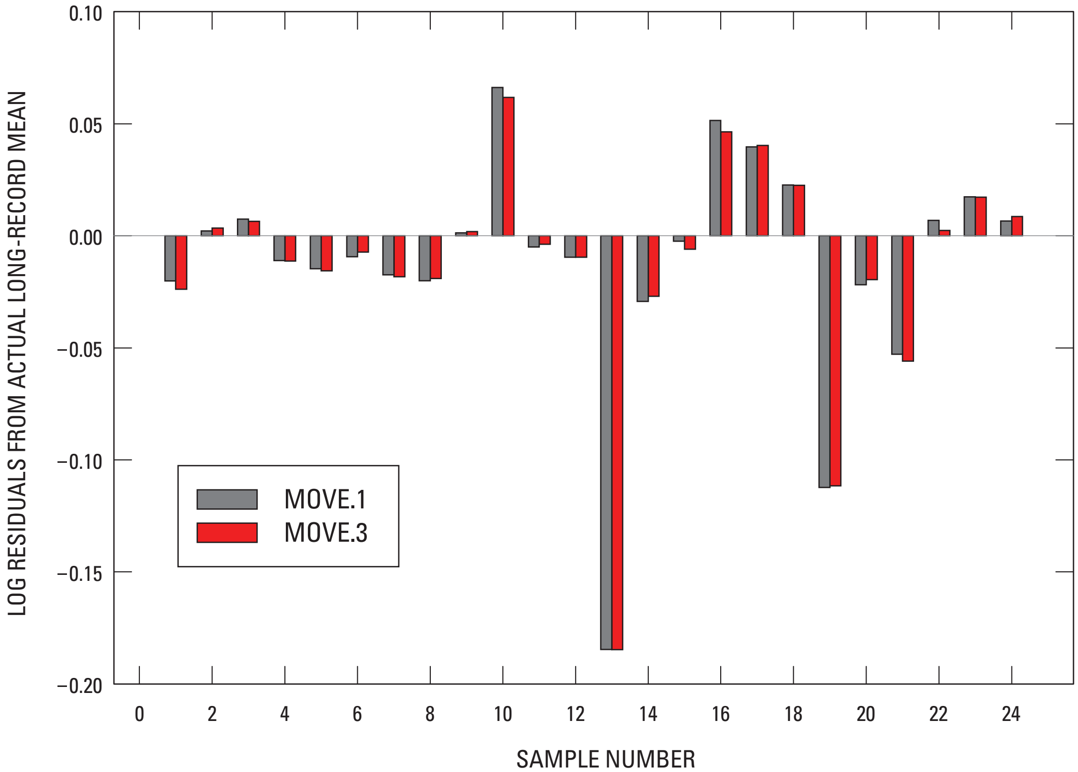


Figure A2. Sample by sample comparison of MOVE.1 and MOVE.3 to estimate the mean of the logs of annual-peak discharges from the logs of annual-maximum-daily discharge based on data from four sites (24 samples) in California.

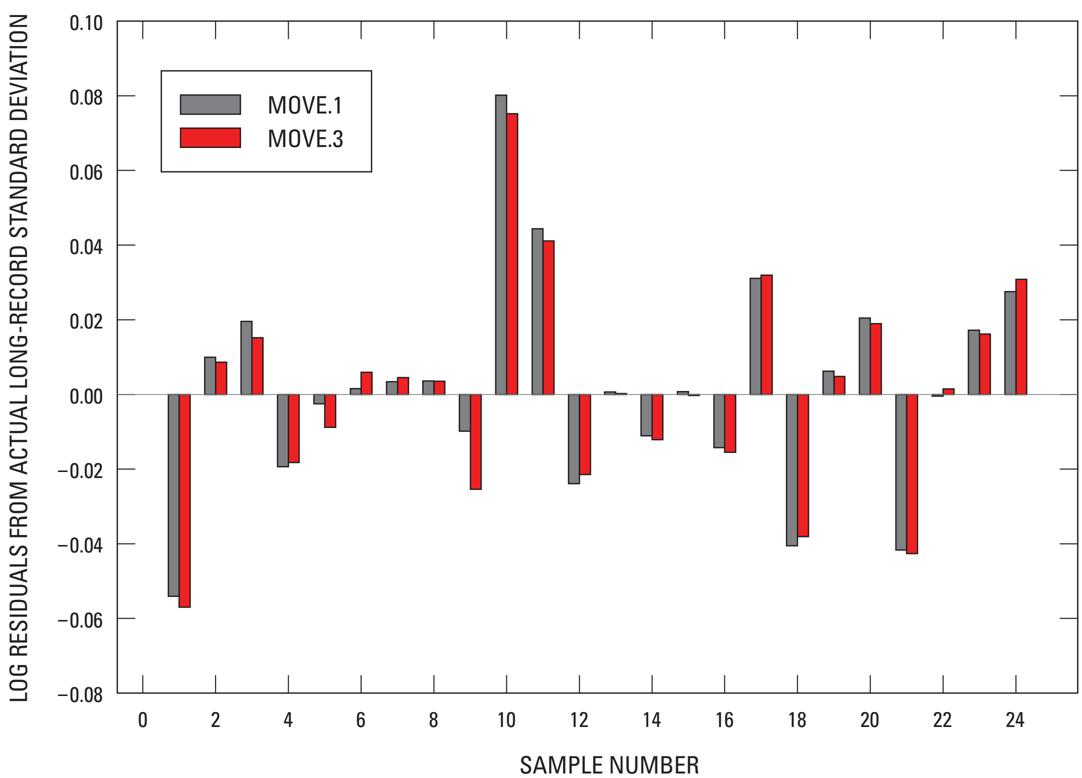


Figure A3. Sample by sample comparison of MOVE.1 and MOVE.3 to estimate the standard deviation of the logs of annual-peak discharges from the logs of annual-maximum-daily discharge based on data from four sites (24 samples) in California.

The site-specific MOVE.1 equations developed for the ten sites were converted to power relations based on discharge rather than the logarithms of discharge to generate unregulated annual-peak discharge estimates for each year after dam construction for which the ACOE had previously estimated unregulated, annual-maximum-daily discharge. The power equations are of the form

$$y_i = Ax_i^b, \tag{A13}$$

where

y_i are the calculated annual-peak discharges,
x_i are the unregulated, annual- maximum-daily-discharges, and
b is the power equation exponent already defined in equation (A4).

The power equation coefficient can be expressed using terms already defined in the basic MOVE.1 equations (A1 through A4) as

$$A = 10^{\left(a - b\overline{X}_1\right)}. \tag{A14}$$

Table A2 shows the ten sites for which a MOVE.1 equation was developed and pertinent data for the MOVE.1 relation, including record lengths n_1 and n_2, the power equation coefficient and exponent, and the Pearson correlation coefficient. The correlations between the concurrent unregulated, annual-maximum-daily-discharges and the annual-peak discharges exceeded 0.90 for every site. Thus, the MOVE analysis provided a reliable extension of the annual peak-discharge records. All but one of the exponents were greater than one, indicating that the annual-peak discharge values varied more than the annual-maximum-daily discharge values.

For 6 of the 16 sites that had no earlier periods of concurrent unregulated, annual-peak-discharge and annual-maximum-daily discharge record, a MOVE.1 equation developed for a nearby and similar site that had long concurrent records for the two series was adjusted for scale and used to generate unregulated, annual-peak discharges for the entire period for which the ACOE had developed unregulated, annual maximum-daily-discharge estimates. The general MOVE equation (eq. A1) for the nearby site is written $Y_i = a + b(X_i - \overline{X}_1)$, where Y and X refer to the concurrent records of logarithms of annual-peak discharge and annual-maximum-daily discharge data for each nearby site. Because the six "estimation sites" do not have concurrent records of annual-peak and daily-maximum discharge of length n_1 and separate extension periods, there is no MOVE.1 versus MOVE.3 problem, and the method can be described as a general MOVE procedure.

Table A2. Ten sites for which MOVE.1 was used to extend unregulated peak flow record in the Sacramento–San Joaquin River Basin, California.

[n_1 is the length of concurrent daily-maximum and peak-flow record, in years. n_2 is the length of the non-concurrent maximum-daily-discharge record, in years. N, north]

Station number [1]	Names of stations	Period of unregulated, annual-maximum flow record		Power equation		Pearson correlation coefficient (r)
		Daily ($n_1 + n_2$) in parentheses	Peak [2] (n_1) in parentheses	Coefficient	Exponent	
11222099	Kings River at Piedra	1896–1999 (104)	1900–02, 1904–07, 1909–53 (52)	0.179	1.215	0.91
11258099	Fresno River below Hidden Dam	1942–99 (59)	1942–75 (34)	0.961	1.075	0.98
11259099	Chowchilla River below Buchanan Dam	1922–23,1932–99 (70)	1932–72 (41)	1.34	1.056	0.97
11270099	Merced River at Exchequer	1902–14,1916–99 (96)	1902–13,1916–25 (22)	0.737	1.059	0.97
11323599	Mokelumne River below Camanche Dam	1905–97 (93)	1905–28 (24)	1.115	1.012	0.97
11335099	Cosumnes River at Michigan Bar	1908–97 (90)	1908–54 (47)	1.099	1.031	0.98
11370599	Sacramento River at Keswick	1932–98 (67)	1932–43 (12)	3.40	0.908	0.92
11407099	Feather River at Oroville	1902–97 (96)	1902–10 (9)	0.472	1.077	0.99
11413599	N Yuba River below Bullards Bar Dam	1941–66,1970–97 (54)	1941–66 (26)	0.486	1.116	0.98
11446599	American River at Fair Oaks	1905–98 (94)	1905–09, 1911–17, 1919–54 (48)	0.835	1.045	0.99

[1] The station number is the U.S. Geological Survey station number with the last two digits replaced by "99."

[2] Peaks recorded during years when no daily maxima were available are not included in n_1.

Using a MOVE relation developed for a nearby site is based on the assumption that flows at the estimation site are *k* times larger or smaller than the corresponding flows at the nearby site. Thus, for each of the six estimation sites, the logarithms of annual-peak discharge are $U_i = Y_i + \log(k)$ and the logarithms of annual-maximum-daily discharge are $V_i = X_i + \log(k)$. Substituting these relations in the basic MOVE equation (eq. A1) yields

$$U_i - \log(k) = a + b\left[V_i - \log(k) - \overline{X}\right], \qquad \text{(A15)}$$

which can be rewritten to show the relation between the values that have been measured at the estimation site (V_i) and the values that are to be generated (U_i):

$$U_i = \left[\left(a - b\overline{X}\right) + (1-b)\log(k)\right] + bV_i \, . \qquad \text{(A16)}$$

Finally, log*(k)* is estimated as the mean logarithm of annual-maximum-daily discharge at the estimation site ($\overline{V}$) minus the mean logarithm of annual-maximum-daily discharge at the nearby site ($\overline{X}$). With that estimator for log(*k*), the MOVE equation for annual-peak discharge at the estimation site can be expressed simply in terms of the annual-maximum-daily discharge at the estimation site and the MOVE constant and coefficient developed for concurrent discharges at the nearby site so that

$$U_i = \left[\left(a + \overline{V} - \overline{X}\right) + b\left(V_i - \overline{V}\right)\right]. \qquad \text{(A17)}$$

Thus, the term log(*k*), which equals $\overline{V} - \overline{X}$ becomes a scaling factor that is added to the MOVE constant for the nearby site to correct for differences in discharge magnitudes at the two sites.

To help determine which nearby site was most suitable for each of the six estimation sites, a correlation analysis of the logs of concurrent annual-maximum-daily discharges at the estimation site and at several nearby sites was used. In general, the nearby site for which the Pearson correlation coefficient was the greatest was selected; however, some sites that had a smaller Pearson correlation coefficient but a longer period of concurrent annual-maximum-daily discharges or that had a MOVE relation based on a longer period of concurrent annual-peak discharge and annual-maximum-daily discharge were selected. Table A3 shows the selected nearby site for each of the six estimation sites, the Pearson correlation coefficient, and the period of concurrent annual-maximum-daily discharges. The nearby sites had reasonably similar hydrologic responses to the estimation sites based on the correlation coefficients and the periods of concurrent discharge. All values of the Pearson correlation coefficient were greater than 0.75 and several exceeded 0.90 (table A3).

The final MOVE equations were converted to a power form and used to calculate annual-peak discharges at each of the six estimation sites from the record of annual-maximum-daily discharge. Table A3 also includes the scaling (adjustment) factor applied to the MOVE constant for the nearby site [$\log(k) = \overline{V} - \overline{X}$] and values of the adjusted power equation coefficient (*A*) and exponent (*b*) for each MOVE equation. The adjusted power equation coefficient (*A*) was calculated after adding the needed scaling factor.

Table A3. Six sites that had no unregulated peak flow record, selected nearby sites, Pearson correlation coefficients, and concurrent records for paired sites, scaling factors, and power equation coefficients and exponents for the Sacramento–San Joaquin River Basin, California.

[USGS, U.S. Geological Survey]

Station numbers [1]	Names of stations that had no unregulated peak-flow record	Names and USGS numbers for correlation stations selected nearby	Pearson correlation coefficient (*r*)	Period of concurrent record	Scaling factors [log (*k*)]	Power equation	
						Adjusted coefficient (A)	Exponent (b)
11251099	San Joaquin River	Kings River (11222099)	0.92	1911–99	0.0097	0.179	1.215
11288099	Tuolumne River	Mokelumne River (11323599)	0.90	1905–97	0.3977	1.103	1.011
11299599	Stanislaus River	Mokelumne River (11323599)	0.91	1932–97	0.1496	1.003	1.011
11308999	Calaveras River	Cosumnes River (11335099)	0.86	1964–97	–0.1281	1.109	1.031
11344099	Littlejohns Creek	Cosumnes River (11335099)	0.76	1951–97	–0.5834	1.145	1.031
11388099	Stony Creek	Fresno River (11258099)	0.88	1982–98	0.9484	0.817	1.075

[1] Station number is the USGS station number with the last two digits replaced by “99.”

All exponents are greater than one, indicating that the annual-peak discharges have more variability than the annual-maximum-daily discharges.

The MOVE relations for each of the ten sites for which concurrent periods of recorded unregulated annual-peak discharges and annual-maximum-daily discharges were available are shown in figures A4 through A13. Correlations between annual-peak discharge and annual-maximum-daily discharge were compared with correlations between annual-peak discharge and annual-maximum 3-day discharge for each of the ten sites to test whether annual-maximum 3-day discharge might predict annual-peak discharge better than annual-maximum-daily discharge. For all ten sites, the correlations between annual-peak discharge and annual-maximum-daily discharge were larger than those between annual-peak discharge and annual-maximum 3-day discharge. On the basis of the tests at the ten sites having concurrent records of recorded unregulated annual-peak discharges and annual-maximum-daily discharges, annual-maximum-daily discharge was selected as the best predictor for annual-peak discharge for all 16 key dam sites.

Unregulated, annual-peak discharges for the 16 key dam sites are shown in table A4. These data are displayed in a form similar to that used by the version of the EMA program used for flood-frequency analyses in California. Most of the tabulated annual-peak discharges were calculated from the MOVE.1 relations described above, although periods of recorded, unregulated annual-peak discharges at the ten sites are also included and noted in the table.

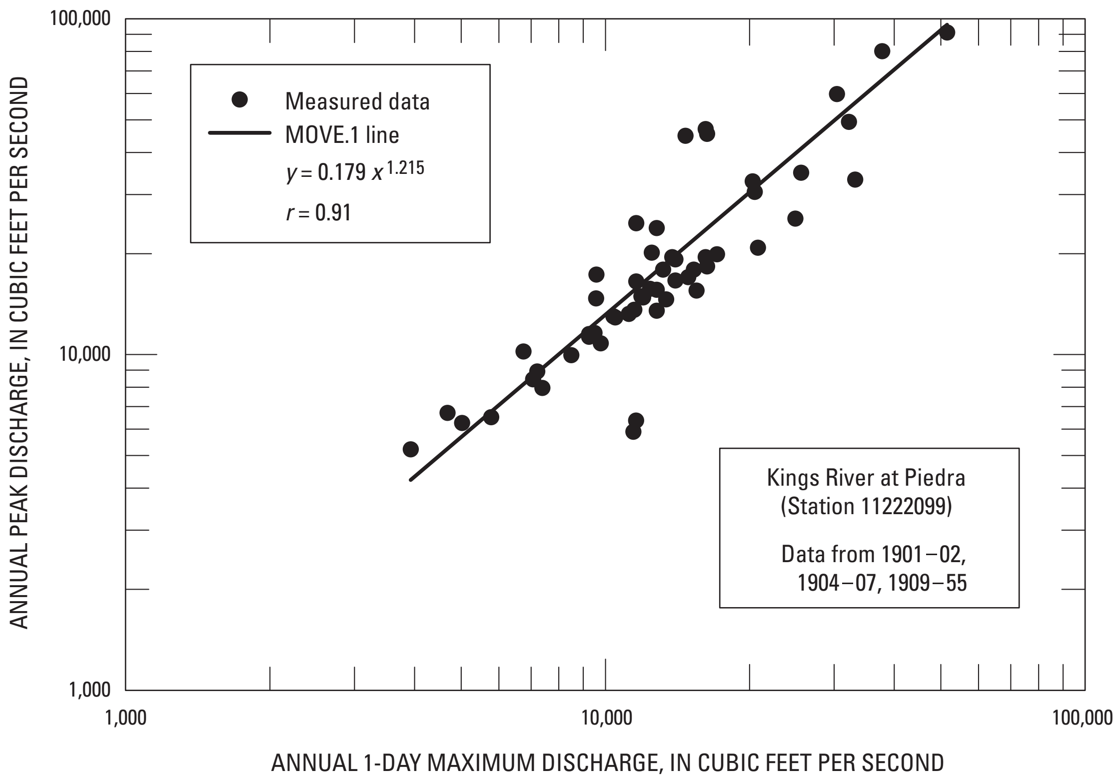


Figure A4. MOVE.1 line relating annual-peak discharge to annual-maximum-daily discharge for Kings River at Piedra, California, (station 11222099) based on data from 1901–02, 1904–07, and 1909–55. *r*, Pearson correlation coefficient.

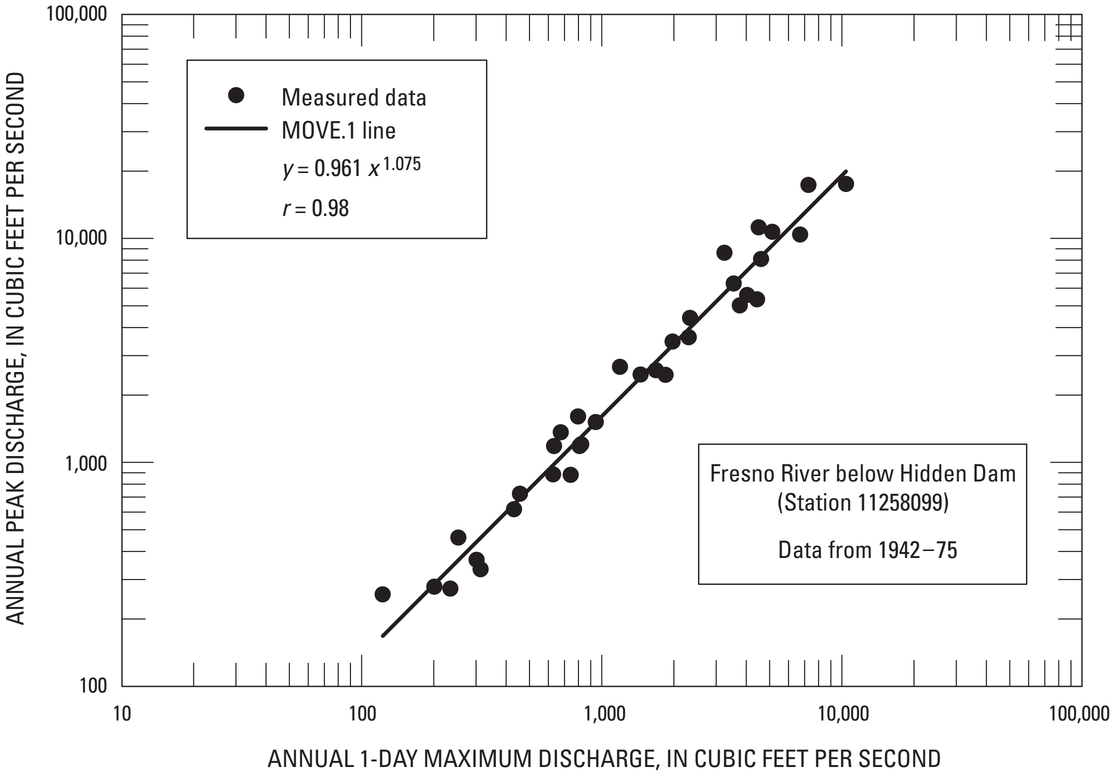


Figure A5. MOVE.1 line relating annual-peak discharge to annual-maximum-daily discharge for Fresno River below Hidden Dam, California, (station 11258099) based on data from 1942–75. *r*, Pearson correlation coefficient.

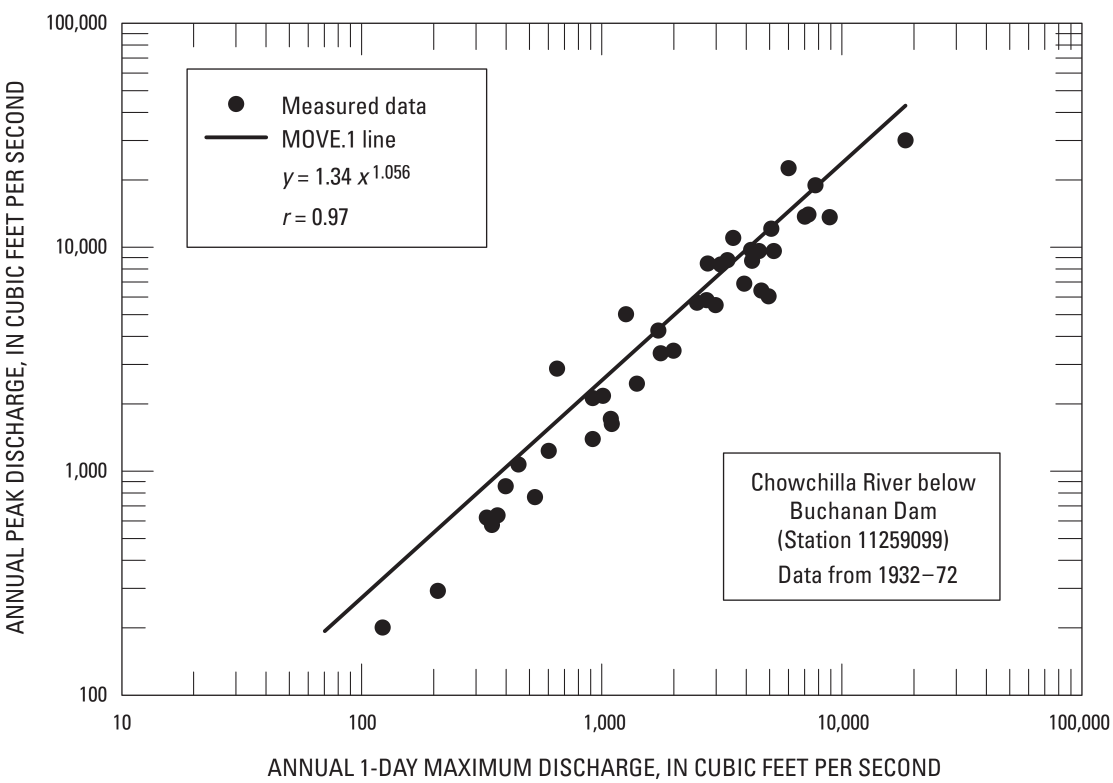


Figure A6. MOVE.1 line relating annual-peak discharge to annual annual-maximum-daily discharge for Chowchilla River below Buchanan Dam, California (station 11259099) based on data from 1932–72. *r*, Pearson correlation coefficient.

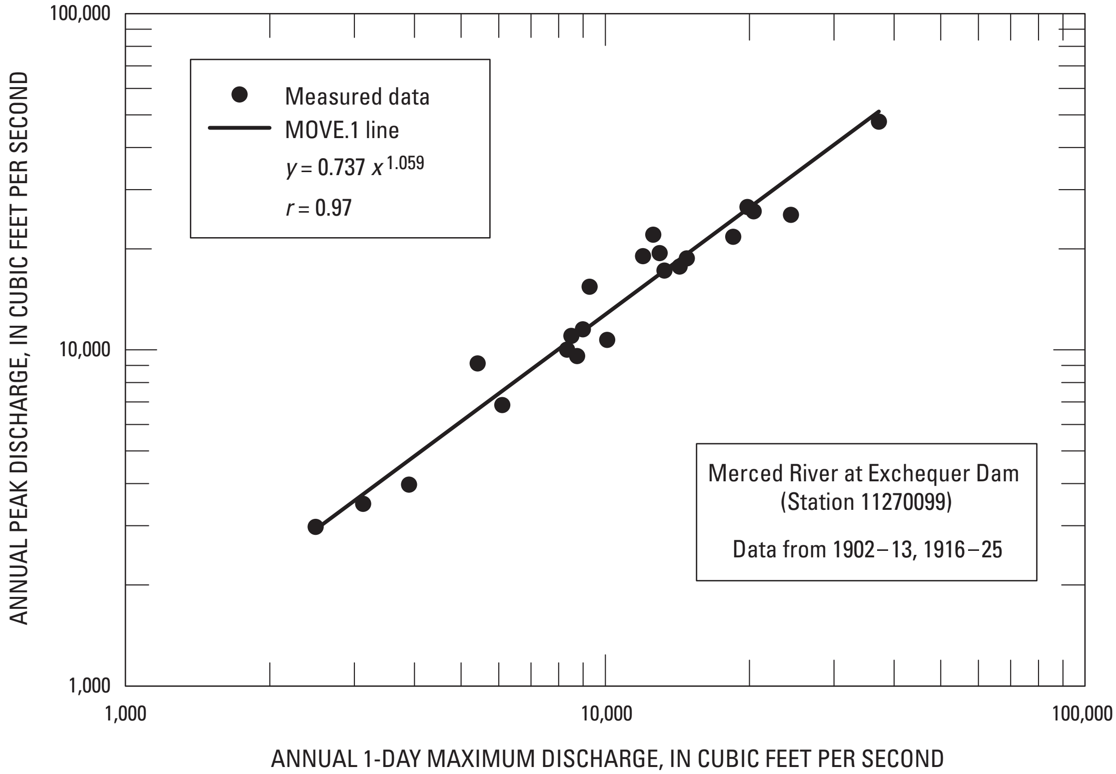


Figure A7. MOVE.1 line relating annual-peak discharge to annual-maximum-daily discharge for Merced River at Exchequer Dam, California, (station 11270099) based on data from 1902–13, 1916–25. *r*, Pearson correlation coefficient.

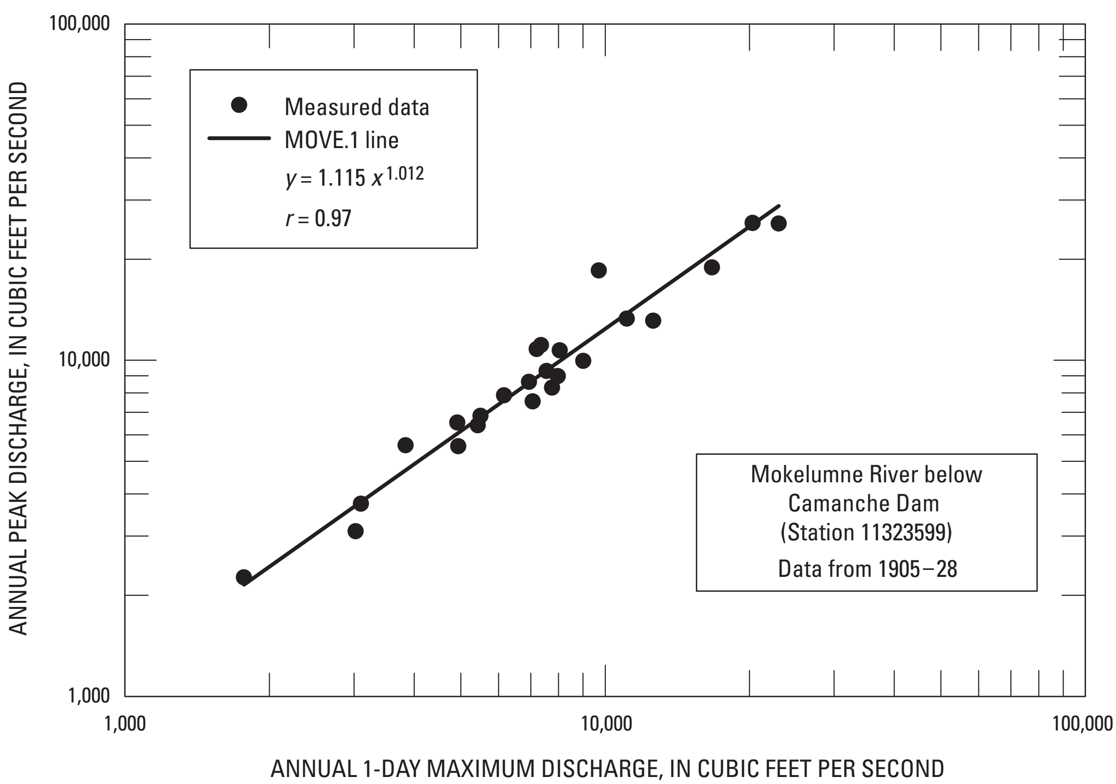


Figure A8. MOVE.1 line relating annual-peak discharge to annual-maximum-daily discharge for Mokelumne River below Camanche Dam, California, (station 11323599) based on data from 1905–28. *r*, Pearson correlation coefficient.

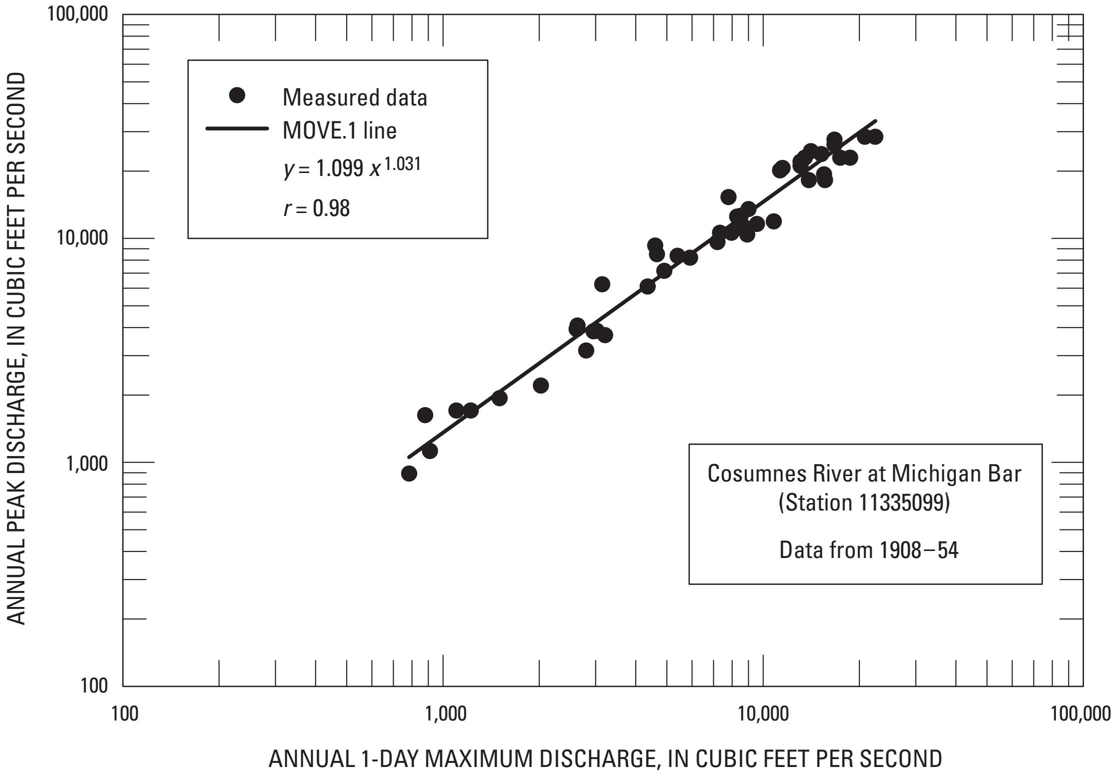


Figure A9. MOVE.1 line relating annual-peak discharge to annual-maximum-daily discharge for Cosumnes River at Michigan Bar, California, (station 11335099) based on data from 1908–54. *r*, Pearson correlation coefficient.

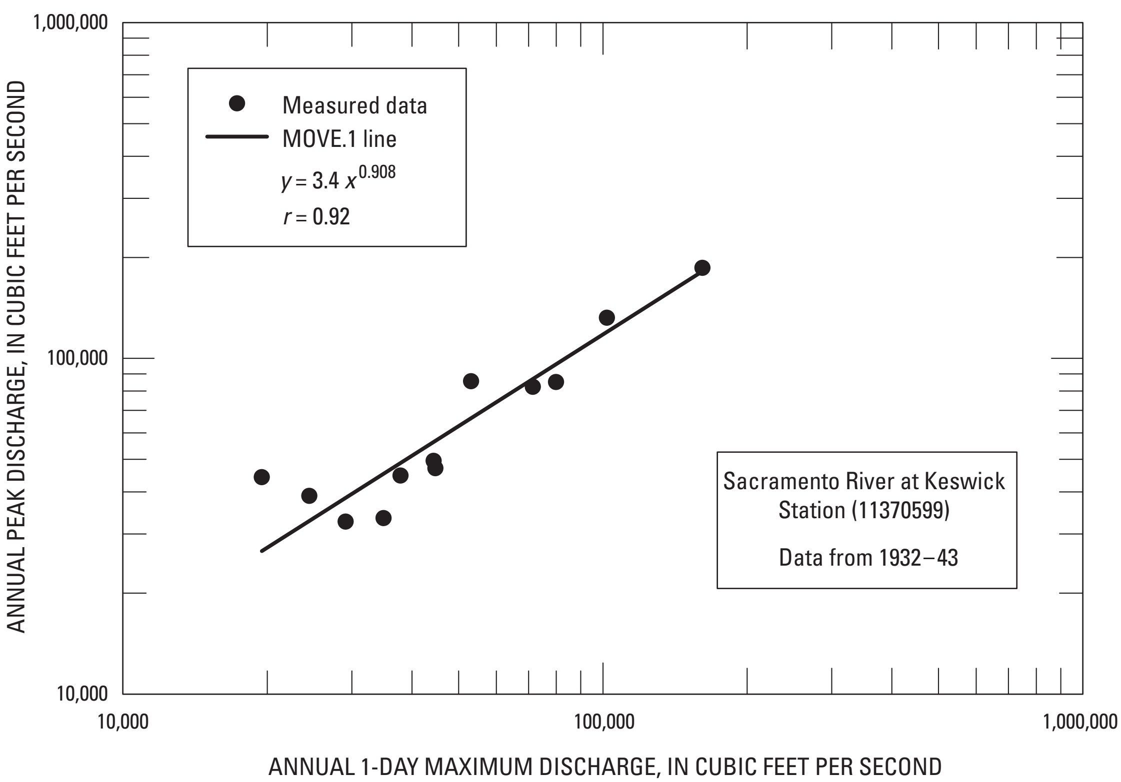


Figure A10. MOVE.1 line relating annual-peak discharge to annual-maximum-daily discharge for Sacramento River at Keswick, California (station 11370599) based on data from 1932-43. *r*, Pearson correlation coefficient.

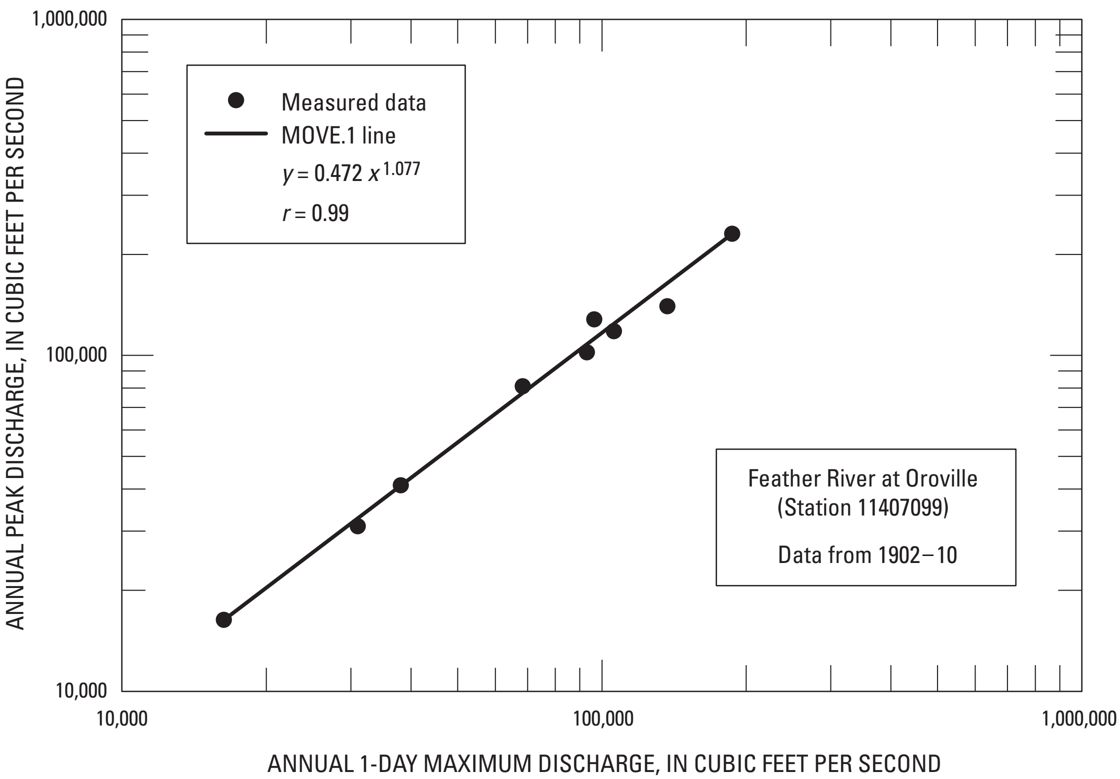


Figure A11. MOVE.1 line relating annual-peak discharge to annual-maximum-daily discharge for Feather River at Oroville, California, (station 11407099) based on data from 1902–10. *r*, Pearson correlation coefficient.

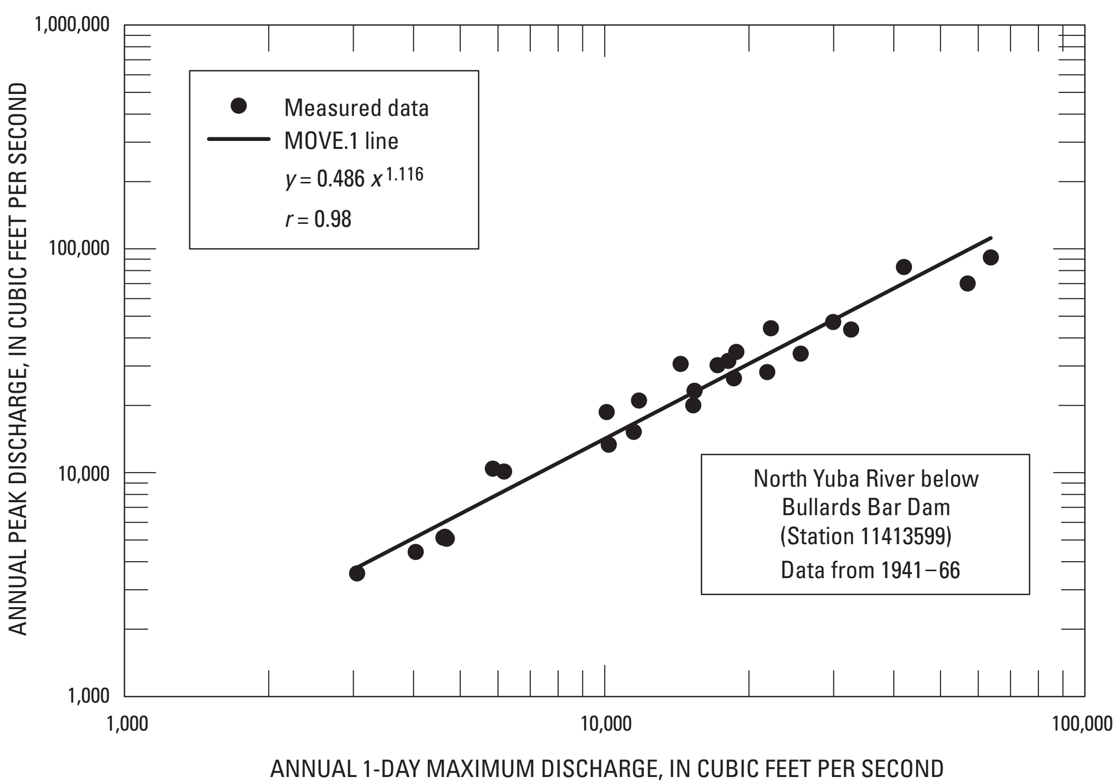


Figure A12. MOVE.1 line relating annual-peak discharge to annual-maximum-daily discharge for North Yuba River below Bullards Bar Dam, California, (station 11413599) based on data from 1941–66. *r*, Pearson correlation coefficient.

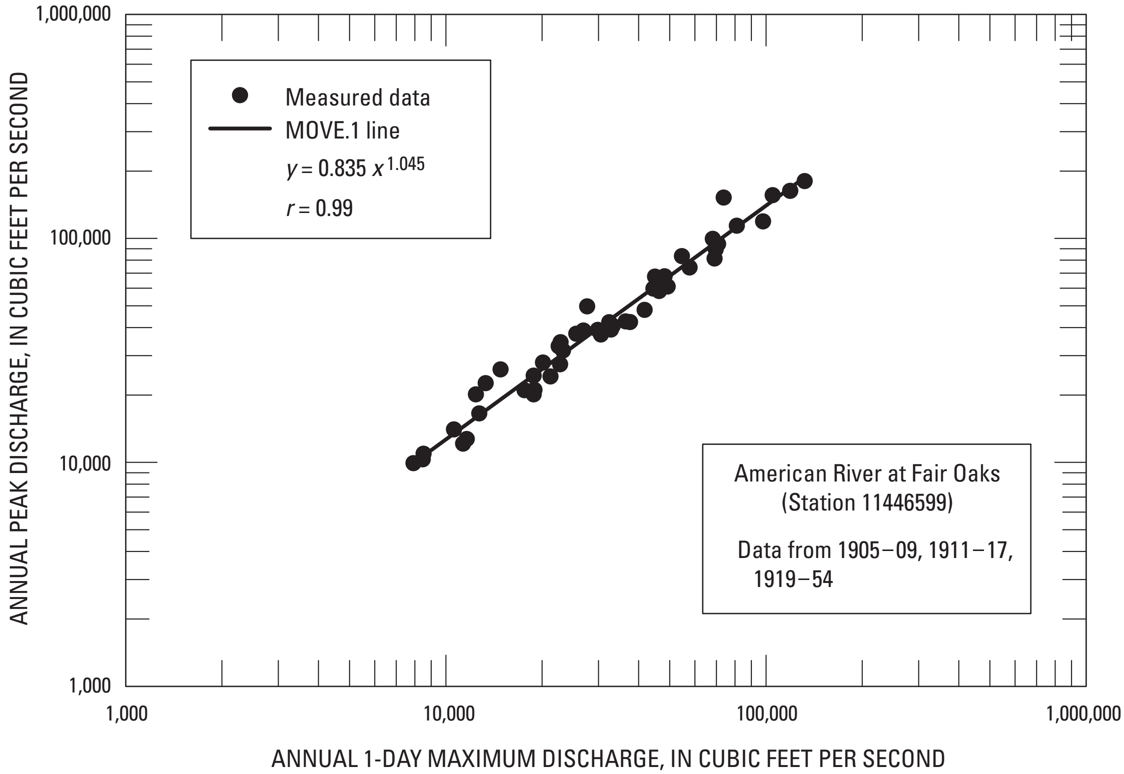


Figure A13. MOVE.1 line relating annual-peak discharge to annual-maximum-daily discharge for American River at Fair Oaks, California, (station 11446599) based on data from 1905–09, 1911–17, 1919–54. *r*, Pearson correlation coefficient.

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11222099 Kings River at Piedra, California (simulated unregulated peaks)			
Q	1896	12,000	
Q	1897	26,100	
Q	1898	10,100	
Q	1899	37,600	
Q	1900	14,600	1
Q	1901	33,200	1
Q	1902	20,800	1
Q	1903	21,200	
Q	1904	19,500	1
Q	1905	10,800	1
Q	1906	25,400	1
Q	1907	19,500	1
Q	1908	7,630	
Q	1909	32,800	1
Q	1910	44,800	1
Q	1911	30,500	1
Q	1912	15,700	1
Q	1913	8,900	1
Q	1914	59,700	1
Q	1915	18,300	1
Q	1916	45,400	1
Q	1917	17,900	1
Q	1918	13,500	1
Q	1919	13,200	1
Q	1920	17,000	1
Q	1921	15,600	1
Q	1922	19900	1
Q	1923	13,600	1
Q	1924	5,210	1
Q	1925	11,300	1
Q	1926	11,600	1
Q	1927	1,9200	1
Q	1928	10,200	1
Q	1929	14,700	1
Q	1930	8,430	1
Q	1931	6,250	1
Q	1932	23,800	1
Q	1933	14,900	1
Q	1934	6,690	1

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11222099 Kings River at Piedra, California (simulated unregulated peaks)—Continued			
Q	1935	20,100	1
Q	1936	24,600	1
Q	1937	34,800	1
Q	1938	80,000	1
Q	1939	6,500	1
Q	1940	14,800	1
Q	1941	17,900	1
Q	1942	16,600	1
Q	1943	46,900	1
Q	1944	9,950	1
Q	1945	49,300	1
Q	1946	16,500	1
Q	1947	17,300	1
Q	1948	13,000	1
Q	1949	11,500	1
Q	1950	12,900	1
Q	1951	91,000	1
Q	1952	15,500	
Q	1953	7,950	
Q	1954	15,500	
Q	1955	15,300	
Q	1956	157,000	
Q	1957	20,400	
Q	1958	20,600	
Q	1959	6,500	
Q	1960	5,690	
Q	1961	3,430	
Q	1962	17,400	
Q	1963	60,400	
Q	1964	8,630	
Q	1965	22,300	
Q	1966	9,640	
Q	1967	142,000	
Q	1968	6,640	
Q	1969	89,000	
Q	1970	23,200	
Q	1971	8,860	
Q	1972	6,110	
Q	1973	25,600	

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11222099 Kings River at Piedra, California (simulated unregulated peaks)—Continued		
Q	1974	20,400
Q	1975	22,600
Q	1976	780
Q	1977	4,820
Q	1978	31,800
Q	1979	20,500
Q	1980	62,100
Q	1981	11,700
Q	1982	96,100
Q	1983	46,900
Q	1984	18,800
Q	1985	9,080
Q	1986	45,400
Q	1987	6,990
Q	1988	9,570
Q	1989	7,010
Q	1990	5,140
Q	1991	18,800
Q	1992	6,730
Q	1993	23,800
Q	1994	9,630
Q	1995	49,200
Q	1996	46,800
Q	1997	99,200
Q	1998	31,700
Q	1999	11,300
Station - 11251099 San Joaquin River below Friant, California (simulated unregulated peaks)		
Q	1911	1,800
Q	1912	21,900
Q	1913	8,170
Q	1914	43,900
Q	1915	21,500
Q	1916	18,700
Q	1917	19,100
Q	1918	15,500
Q	1919	14,800
Q	1920	17,500
Q	1921	16,800

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11251099 San Joaquin River below Friant, California (simulated unregulated peaks)—Continued		
Q	1922	26,700
Q	1923	15,500
Q	1924	4,500
Q	1925	12,800
Q	1926	13,100
Q	1927	19,000
Q	1928	14,600
Q	1929	10,800
Q	1930	7,780
Q	1931	4,700
Q	1932	20,500
Q	1933	15,000
Q	1934	3,720
Q	1935	18,800
Q	1936	14,000
Q	1937	27,500
Q	1938	70,600
Q	1939	6,020
Q	1940	14,800
Q	1941	23,100
Q	1942	20,500
Q	1943	24,700
Q	1944	10,200
Q	1945	57,800
Q	1946	14,400
Q	1947	10,900
Q	1948	13,900
Q	1949	12,700
Q	1950	14,000
Q	1951	74,500
Q	1952	26,600
Q	1953	10,500
Q	1954	14,100
Q	1955	15,400
Q	1956	149,000
Q	1957	24,300
Q	1958	25,500
Q	1959	9,730
Q	1960	7,810

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11251099 San Joaquin River below Friant, California (simulated unregulated peaks)—Continued		
Q	1961	4,050
Q	1962	19,100
Q	1963	71,600
Q	1964	7,870
Q	1965	40,300
Q	1966	9,020
Q	1967	74,600
Q	1968	6,180
Q	1969	47,700
Q	1970	22,800
Q	1971	10,700
Q	1972	8,540
Q	1973	24,700
Q	1974	18,800
Q	1975	24,500
Q	1976	5,380
Q	1977	5,840
Q	1978	29,000
Q	1979	19,900
Q	1980	53,000
Q	1981	10,300
Q	1982	112000
Q	1983	42,300
Q	1984	26,700
Q	1985	8,340
Q	1986	56,100
Q	1987	9,810
Q	1988	6,480
Q	1989	7,890
Q	1990	5,400
Q	1991	11,300
Q	1992	6,940
Q	1993	21,700
Q	1994	8,300
Q	1995	68,100
Q	1996	53,400
Q	1997	155,000
Q	1998	30,300
Q	1999	14,200

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11258099 Fresno River below Hidden Dam near Daulton, California (simulated unregulated peaks)

Threshold	1939–1941	17,500 -INF	
Q	1938	15,000	1
Q	1942	2,580	1
Q	1943	3,620	1
Q	1944	2,670	1
Q	1945	8,090	1
Q	1946	2,460	1
Q	1947	1,600	1
Q	1948	2,470	1
Q	1949	1,510	1
Q	1950	1,180	1
Q	1951	10,700	1
Q	1952	5,020	1
Q	1953	1,200	1
Q	1954	723	1
Q	1955	333	1
Q	1956	17,500	1
Q	1957	1,360	1
Q	1958	10,400	1
Q	1959	1,180	1
Q	1960	616	1
Q	1961	257	1
Q	1962	5,340	1
Q	1963	6,290	1
Q	1964	278	1
Q	1965	3,460	1
Q	1966	882	1
Q	1967	5,590	1
Q	1968	273	1
Q	1969	17,300	1
Q	1970	4,420	1
Q	1971	461	1
Q	1972	367	1
Q	1973	11,200	1
Q	1974	8,620	1
Q	1975	877	1
Q	1976	249	

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11258099 Fresno River below Hidden Dam near Daulton, California (simulated unregulated peaks)—Continued			
Q	1977	54.8	
Q	1978	7,940	
Q	1979	3,590	
Q	1980	6,250	
Q	1981	721	
Q	1982	10,500	
Q	1983	10,400	
Q	1984	3,230	
Q	1985	797	
Q	1986	10,700	
Q	1987	511	
Q	1988	291	
Q	1989	527	
Q	1990	104	
Q	1991	1,890	
Q	1992	1,010	
Q	1993	13,500	
Q	1994	240	
Q	1995	16,300	
Q	1996	3,190	
Q	1997	14,500	
Q	1998	5,870	
Q	1999	1,140	
Station - 11259099 Chowchilla River below Buchanan Dam near Raymond, California (simulated unregulated peaks)			
Q	1931	156	1
Q	1932	9,610	1
Q	1933	620	1
Q	1934	2,170	1
Q	1935	5,510	1
Q	1936	11,000	1
Q	1937	13,600	1
Q	1938	18,900	1
Q	1939	765	1
Q	1940	8,750	1
Q	1941	12,100	1
Q	1942	5,800	1

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11259099 Chowchilla River below Buchanan Dam near Raymond, California (simulated unregulated peaks)—Continued

Q	1943	8,460	1
Q	1944	5,020	1
Q	1945	9,610	1
Q	1946	3,450	1
Q	1947	2,870	1
Q	1948	5,660	1
Q	1949	3,360	1
Q	1950	4,250	1
Q	1951	22,500	1
Q	1952	6,040	1
Q	1953	1,710	1
Q	1954	1,390	1
Q	1955	1,070	1
Q	1956	30,000	1
Q	1957	1,230	1
Q	1958	14,000	1
Q	1959	2,120	1
Q	1960	1,620	1
Q	1961	200	1
Q	1962	6400	1
Q	1963	9,740	1
Q	1964	855	1
Q	1965	8,380	1
Q	1966	2,460	1
Q	1967	6,880	1
Q	1968	292	1
Q	1969	13,700	1
Q	1970	8,700	1
Q	1971	574	1
Q	1972	634	1
Q	1973	11,200	
Q	1974	8,440	
Q	1975	2,650	
Q	1976	391	
Q	1977	43.5	
Q	1978	10,300	
Q	1979	4,990	
Q	1980	6,310	

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11259099 Chowchilla River below Buchanan Dam near Raymond, California (simulated unregulated peaks)—Continued			
Q	1981	2,150	
Q	1982	16,600	
Q	1983	15,800	
Q	1984	5,350	
Q	1985	1,600	
Q	1986	14,900	
Q	1987	849	
Q	1988	233	
Q	1989	893	
Q	1990	106	
Q	1991	2,830	
Q	1992	2,800	
Q	1993	16,700	
Q	1994	249	
Q	1995	17,700	
Q	1996	5,060	
Q	1997	17,600	
Q	1998	8,350	
Q	1999	2,180	
Station - 11270099 Merced River at Exchequer, California (simulated unregulated peaks)			
Threshold	1915–1915	107,000–INF	
Q	1902	9,120	1
Q	1903	19,000	1
Q	1904	10,700	1
Q	1905	11,000	1
Q	1906	26,600	1
Q	1907	25,200	1
Q	1908	3,970	1
Q	1909	25,800	1
Q	1910	18,700	1
Q	1911	47,700	1
Q	1912	6,840	1
Q	1913	3,480	1
Q	1914	206	1
Q	1916	22,000	1
Q	1917	21,700	1
Q	1918	17,700	1

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11270099 Merced River at Exchequer, California (simulated unregulated peaks)—Cont.			
Q	1919	9,580	1
Q	1920	10,000	1
Q	1921	19,400	1
Q	1922	17,200	1
Q	1923	11,500	1
Q	1924	2,970	1
Q	1925	15,400	1
Q	1926	11,100	2
Q	1927	10,900	1
Q	1928	20,800	1
Q	1929	7,150	1
Q	1930	4,760	1
Q	1931	2,840	1
Q	1932	14,600	1
Q	1933	7,440	1
Q	1934	5,620	1
Q	1935	29,200	1
Q	1936	19,100	1
Q	1937	33,800	1
Q	1938	46,300	1
Q	1939	4,140	1
Q	1940	14,500	1
Q	1941	17,500	1
Q	1942	10,600	1
Q	1943	17,400	1
Q	1944	6,710	1
Q	1945	45,000	1
Q	1946	14,800	1
Q	1947	6,470	1
Q	1948	8,380	1
Q	1949	7,750	1
Q	1950	7,470	1
Q	1951	64,700	1
Q	1952	17,200	1
Q	1953	6,590	1
Q	1954	7,090	1
Q	1955	6,530	1
Q	1956	107,000	1
Q	1957	11,700	1

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11270099 Merced River at Exchequer, California (simulated unregulated peaks)—Cont.			
Q	1958	26,100	1
Q	1959	8,310	1
Q	1960	7,630	1
Q	1961	3,330	1
Q	1962	16,600	1
Q	1963	52,700	1
Q	1964	4,030	1
Q	1965	45,100	1
Q	1966	8,100	1
Q	1967	22,500	1
Q	1968	3,370	1
Q	1969	45,600	1
Q	1970	19,100	1
Q	1971	6,010	1
Q	1972	4,560	1
Q	1973	16,400	
Q	1974	14,100	
Q	1975	13,100	
Q	1976	2,770	
Q	1977	2,600	
Q	1978	19,600	
Q	1979	17,400	
Q	1980	42,600	
Q	1981	5,750	
Q	1982	53,800	
Q	1983	28,300	
Q	1984	23,400	
Q	1985	5,130	
Q	1986	43,400	
Q	1987	3,640	
Q	1988	3,200	
Q	1989	4,920	
Q	1990	3,180	
Q	1991	5,760	
Q	1992	5,590	
Q	1993	17,800	
Q	1994	3,740	
Q	1995	52,500	
Q	1996	21,200	

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11270099 Merced River at Exchequer, California (simulated unregulated peaks)—Cont.			
Q	1997	95,200	
Q	1998	22,000	
Q	1999	10,800	
Station - 11288099 Tuolumne River above La Grange Dam near La Grange, California (simulated unregulated peaks)			
Threshold	1863–1896	130,000–INF	
Threshold	2000–2006	153,000–INF	
Q	1862	130,000	1
Q	1897	20,000	
Q	1898	9,290	
Q	1899	33,200	
Q	1900	17,500	
Q	1901	27,900	
Q	1902	14,900	
Q	1903	24,400	
Q	1904	20,200	
Q	1905	17,900	
Q	1906	32,500	
Q	1907	63,000	
Q	1908	7,900	
Q	1909	33,100	
Q	1910	25,800	
Q	1911	65,700	
Q	1912	17,000	
Q	1913	9,280	
Q	1914	38,900	
Q	1915	18,800	
Q	1916	21,100	
Q	1917	28,500	
Q	1918	18,700	
Q	1919	17,000	
Q	1920	15,900	
Q	1921	15,300	
Q	1922	22,700	
Q	1923	17,600	
Q	1924	10,700	
Q	1925	22,200	
Q	1928	54,100	
Q	1929	18,800	

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11288099 Tuolumne River above La Grange Dam near La Grange, California (simulated unregulated peaks)—Continued

Q	1930	10,300
Q	1931	6,850
Q	1932	28,400
Q	1933	16,200
Q	1934	7,590
Q	1935	29,000
Q	1936	27,000
Q	1937	30,200
Q	1938	93,400
Q	1939	7,540
Q	1940	36,500
Q	1941	21,700
Q	1942	28,100
Q	1943	28,700
Q	1944	12,000
Q	1945	56,700
Q	1946	23,400
Q	1947	12,200
Q	1948	15,500
Q	1949	14,400
Q	1950	14,800
Q	1951	84,000
Q	1952	21,000
Q	1953	17,900
Q	1954	21,000
Q	1955	13,500
Q	1956	153,000
Q	1957	22,300
Q	1958	23,300
Q	1959	13,500
Q	1960	12,900
Q	1961	6,200
Q	1962	20,300
Q	1963	87,900
Q	1964	13,000
Q	1965	91,200

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11288099 Tuolumne River above La Grange Dam near La Grange, California (simulated unregulated peaks)—Continued

Q	1966	12,400
Q	1967	37,300
Q	1968	11,100
Q	1969	62,200
Q	1970	38,100
Q	1971	12,100
Q	1972	10,300
Q	1973	20,600
Q	1974	28,900
Q	1975	78,600
Q	1976	7,130
Q	1977	7,040
Q	1978	22,600
Q	1979	24,700
Q	1980	75,200
Q	1981	11,000
Q	1982	62,200
Q	1983	32,700
Q	1984	35,400
Q	1985	11,000
Q	1986	72,500
Q	1987	8,880
Q	1988	7,060
Q	1989	14,800
Q	1990	7,580
Q	1991	13,400
Q	1992	7,790
Q	1993	18,700
Q	1994	9,130
Q	1995	58,200
Q	1996	48,000
Q	1997	148,000
Q	1998	35,300
Q	1999	20,500

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11299599 Stanislaus River below Melones powerhouse near Sonora, California (simulated unregulated peaks)

Threshold	2000–2006	102,000–INF
Q	1932	11,200
Q	1933	10,600
Q	1934	3,260
Q	1935	12,900
Q	1936	17,900
Q	1937	11,800
Q	1938	46,400
Q	1939	4,690
Q	1940	25,800
Q	1941	13,000
Q	1942	13,200
Q	1943	25,900
Q	1944	8,150
Q	1945	23,600
Q	1946	9,310
Q	1947	5,740
Q	1948	11,700
Q	1949	10,600
Q	1950	9,490
Q	1951	73,900
Q	1952	14,700
Q	1953	13,000
Q	1954	14,100
Q	1955	6,680
Q	1956	102,000
Q	1957	15,600
Q	1958	16,300
Q	1959	3,570
Q	1960	9,680
Q	1961	2,840
Q	1962	8,290
Q	1963	48,000
Q	1964	4,730
Q	1965	54,100
Q	1966	4,630
Q	1967	20,100
Q	1968	6,080
Q	1969	29,800
Q	1970	35,100
Q	1971	11,700

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11299599 Stanislaus River below Melones powerhouse near Sonora, California (simulated unregulated peaks)—Continued		
Q	1972	5,130
Q	1973	15,500
Q	1974	12,000
Q	1975	22,500
Q	1976	2,220
Q	1977	1,130
Q	1978	28,100
Q	1979	11,300
Q	1980	57,200
Q	1981	6,860
Q	1982	53,900
Q	1983	22,800
Q	1984	23,300
Q	1985	6,290
Q	1986	56,900
Q	1987	4,220
Q	1988	1,910
Q	1989	8,070
Q	1990	2,750
Q	1991	6,840
Q	1992	3,840
Q	1993	10,400
Q	1994	2,910
Q	1995	30,000
Q	1996	26,300
Q	1997	92,000
Q	1998	14,500
Q	1999	13,600
Station - 11308999 Calaveras River below New Hogan Dam near Valley Springs, California (simulated unregulated peaks)		
Q	1964	3,710
Q	1965	19,000
Q	1966	2,830
Q	1967	9,800
Q	1968	2,290
Q	1969	21,900
Q	1970	10,500

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11308999 Calaveras River below New Hogan Dam near Valley Springs, California (simulated unregulated peaks)—Continued			
Q	1971	4,230	
Q	1972	7,090	
Q	1973	11,200	
Q	1974	13,400	
Q	1975	8,370	
Q	1976	315	
Q	1977	144	
Q	1978	8,350	
Q	1979	7,780	
Q	1980	12,700	
Q	1981	4,490	
Q	1982	18,300	
Q	1983	15,400	
Q	1984	11700	
Q	1985	5,390	
Q	1986	35,500	
Q	1987	2,460	
Q	1988	538	
Q	1989	1,270	
Q	1990	943	
Q	1991	5,640	
Q	1992	7,380	
Q	1993	7,680	
Q	1994	1,240	
Q	1995	14,900	
Q	1996	8,180	
Q	1997	25,100	
Q	1998	25,300	
Station - 11323599 Mokelumne River below Camanche Dam, California (simulated unregulated peaks)			
Threshold	1998–2006	97,200–INF	
Q	1905	5,540	1
Q	1906	9,960	1
Q	1907	25,500	1
Q	1908	3,100	1
Q	1909	13,100	1
Q	1910	10,800	1
Q	1911	18,900	1

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11323599 Mokelumne River below Camanche Dam, California (simulated unregulated peaks)—Continued

Q	1912	6,520	1
Q	1913	5,580	1
Q	1914	13,300	1
Q	1915	8,290	1
Q	1916	10,700	1
Q	1917	9,300	1
Q	1918	8,620	1
Q	1919	7,540	1
Q	1920	6,820	1
Q	1921	11,100	1
Q	1922	8,970	1
Q	1923	6,400	1
Q	1924	2,260	1
Q	1925	18,500	1
Q	1926	3,740	1
Q	1927	7,870	1
Q	1928	25,600	1
Q	1929	4,340	
Q	1930	4,080	
Q	1931	2,470	
Q	1932	6,950	
Q	1933	6,310	
Q	1934	3,510	
Q	1935	7,690	
Q	1936	18,800	
Q	1937	7,670	
Q	1938	28,900	
Q	1939	3,150	
Q	1940	13,400	
Q	1941	7,690	
Q	1942	13,000	
Q	1943	14,200	
Q	1944	4,880	
Q	1945	16,100	
Q	1946	7,590	
Q	1947	4,980	
Q	1948	6,690	
Q	1949	6,310	

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11323599 Mokelumne River below Camanche Dam, California (simulated unregulated peaks)—Continued

Q	1950	6,460
Q	1951	39,000
Q	1952	9,240
Q	1953	6,600
Q	1954	6,790
Q	1955	5,160
Q	1956	43,800
Q	1957	9,780
Q	1958	12,000
Q	1959	3,330
Q	1960	6,710
Q	1961	2,620
Q	1962	6,480
Q	1963	37,700
Q	1964	4,080
Q	1965	45,700
Q	1966	3,490
Q	1967	10,800
Q	1968	4,120
Q	1969	19,300
Q	1970	18,500
Q	1971	6,590
Q	1972	4,700
Q	1973	7,410
Q	1974	9,820
Q	1975	8,340
Q	1976	2,880
Q	1977	1,360
Q	1978	7,690
Q	1979	7,790
Q	1980	40,300
Q	1981	6,020
Q	1982	31,000
Q	1983	15,400
Q	1984	16,900
Q	1985	4,640
Q	1986	35,100
Q	1987	3,300

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11323599 Mokelumne River below Camanche Dam, California (simulated unregulated peaks)—Continued			
Q	1988	2,760	
Q	1989	6,790	
Q	1990	2,780	
Q	1991	5,740	
Q	1992	4,050	
Q	1993	7,750	
Q	1994	3,420	
Q	1995	19,600	
Q	1996	22,600	
Q	1997	97,200	
Station - 11335099 Cosumnes River at Michigan Bar, California (simulated unregulated peaks)			
Threshold	1998–2006	95,600–INF	
Q	1907	71,000	1
Q	1908	2,200	1
Q	1909	28,400	1
Q	1910	9,640	1
Q	1911	28,400	1
Q	1912	1,700	1
Q	1913	1,700	1
Q	1914	18,200	1
Q	1915	8,200	1
Q	1916	10,400	1
Q	1917	22,900	1
Q	1918	11,900	1
Q	1919	22,000	1
Q	1920	3,700	1
Q	1921	20,600	1
Q	1922	10,600	1
Q	1923	11,600	1
Q	1924	1,120	1
Q	1925	23,800	1
Q	1926	3,850	1
Q	1927	11,400	1
Q	1928	22,900	1
Q	1929	3,160	1
Q	1930	6,090	1
Q	1931	1,620	1

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11335099 Cosumnes River at Michigan Bar, California (simulated unregulated peaks)—Continued

Q	1932	10,600	1
Q	1933	890	1
Q	1934	7,170	1
Q	1935	20,100	1
Q	1936	18,200	1
Q	1937	15,300	1
Q	1938	19,300	1
Q	1939	1,930	1
Q	1940	26,200	1
Q	1941	9,280	1
Q	1942	24,500	1
Q	1943	22,900	1
Q	1944	8,490	1
Q	1945	21,100	1
Q	1946	12,600	1
Q	1947	3,930	1
Q	1948	6,240	1
Q	1949	13,500	1
Q	1950	8,360	1
Q	1951	27,600	1
Q	1952	12,500	1
Q	1953	4,080	1
Q	1954	3,860	1
Q	1955	4,460	
Q	1956	49,700	
Q	1957	9,440	
Q	1958	30,000	
Q	1959	4,340	2
Q	1960	1,1200	2
Q	1961	625	
Q	1962	8,880	
Q	1963	41,600	
Q	1964	4,170	
Q	1965	45,200	
Q	1966	2,880	2
Q	1967	15,900	2
Q	1968	4,930	
Q	1969	28,900	
Q	1970	16,900	

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11335099 Cosumnes River at Michigan Bar, California (simulated unregulated peaks)—Continued			
Q	1971	8,590	2
Q	1972	4,430	
Q	1973	15,000	2
Q	1974	9,610	
Q	1975	11,000	2
Q	1976	457	
Q	1977	219	
Q	1978	9,340	
Q	1979	6,990	2
Q	1980	34,200	2
Q	1981	5,890	2
Q	1982	38,600	
Q	1983	27,500	
Q	1984	21,000	
Q	1985	6,290	2
Q	1986	54,700	
Q	1987	2,110	
Q	1988	1,350	
Q	1989	8,400	
Q	1990	1,420	
Q	1991	6670	2
Q	1992	5340	2
Q	1993	11,000	
Q	1994	1,260	
Q	1995	27,200	
Q	1996	11,500	
Q	1997	95,600	
Station - 11344099 Littlejohns Creek below Farmington Reservoir (simulated unregulated peaks; U.S. Army Corps of Engineers station)			
Q	1951	7,870	
Q	1952	7,460	
Q	1953	1,020	
Q	1954	1,010	
Q	1955	5,230	
Q	1956	12,800	
Q	1957	3,240	
Q	1958	10,900	
Q	1959	2,030	
Q	1960	2,000	

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11344099 Littlejohns Creek below Farmington Reservoir (simulated unregulated peaks; U.S. Army Corps of Engineers station)—Continued		
Q	1961	135
Q	1962	7,560
Q	1963	4,700
Q	1964	1,270
Q	1965	13,200
Q	1966	3,000
Q	1967	6,400
Q	1968	1,770
Q	1969	5,460
Q	1970	5,830
Q	1971	3,820
Q	1972	1,810
Q	1973	8,000
Q	1974	7,050
Q	1975	4,000
Q	1976	12.3
Q	1977	0
Q	1978	5,070
Q	1979	7,550
Q	1980	7,310
Q	1981	5,740
Q	1982	9,770
Q	1983	9,920
Q	1984	8,590
Q	1985	3,500
Q	1986	14,500
Q	1987	4,230
Q	1988	81.9
Q	1989	57.9
Q	1990	31.6
Q	1991	3,970
Q	1992	6,690
Q	1993	3,930
Q	1994	382
Q	1995	7,210
Q	1996	5,810
Q	1997	1,1700
Q	1998	1,7200
Q	1999	6,690

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11370599 Sacramento River at Keswick, California (simulated unregulated peaks)			
Threshold	1999–2006	237,000–INF	
Q	1926	66,000	1
Q	1927	92,800	1
Q	1928	94,900	1
Q	1929	25,800	1
Q	1930	64,000	1
Q	1931	14,300	1
Q	1932	45,400	1
Q	1933	44,200	1
Q	1934	38,900	1
Q	1935	56,800	1
Q	1936	85,500	1
Q	1937	38,400	1
Q	1938	132,000	1
Q	1939	48,800	1
Q	1940	186,000	1
Q	1941	86,800	1
Q	1942	96,100	1
Q	1943	56,400	1
Q	1944	10,200	
Q	1945	13,900	
Q	1946	37,500	
Q	1947	44,200	
Q	1948	77,100	
Q	1949	54,200	
Q	1950	38,600	
Q	1951	64,900	
Q	1952	93,200	
Q	1953	98,000	
Q	1954	75,900	
Q	1955	41,900	
Q	1956	161,000	
Q	1957	93,400	
Q	1958	99,600	
Q	1959	79,200	
Q	1960	80,000	
Q	1961	55,200	
Q	1962	85,500	
Q	1963	76,200	

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11370599 Sacramento River at Keswick, California (simulated unregulated peaks)—Continued		
Q	1964	77,400
Q	1965	190,000
Q	1966	48,000
Q	1967	77,000
Q	1968	61,200
Q	1969	109,000
Q	1970	185,000
Q	1971	76,700
Q	1972	49,300
Q	1973	90,300
Q	1974	212,000
Q	1975	70,000
Q	1976	30,700
Q	1977	12,400
Q	1978	113,000
Q	1979	40,200
Q	1980	107,000
Q	1981	40,100
Q	1982	98,200
Q	1983	110,000
Q	1984	81,000
Q	1985	32,200
Q	1986	146,000
Q	1987	50,900
Q	1988	43,000
Q	1989	8,860
Q	1990	41,300
Q	1991	38,300
Q	1992	46,200
Q	1993	98,700
Q	1994	24,800
Q	1995	130,000
Q	1996	83,900
Q	1997	237,000
Q	1998	947,00

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11388099 Stony Creek below Black Butte Dam near Orland, California (simulated unregulated peaks)		
Q	1964	5,230
Q	1965	66,200
Q	1966	26,100
Q	1967	20,300
Q	1968	13,900
Q	1969	23,700
Q	1970	40,700
Q	1971	20,100
Q	1972	5,570
Q	1973	20,300
Q	1974	46,700
Q	1975	17,100
Q	1976	2,200
Q	1977	478
Q	1978	41,500
Q	1979	9,650
Q	1980	36,000
Q	1981	15,500
Q	1982	22,600
Q	1983	53,100
Q	1984	30,700
Q	1985	6,290
Q	1986	65,200
Q	1987	4,880
Q	1988	15,400
Q	1989	6,680
Q	1990	3,000
Q	1991	11,200
Q	1992	11,400
Q	1993	36,800
Q	1994	4,540
Q	1995	81,700
Q	1996	16,200
Q	1997	51,400
Q	1998	52,500

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11407099 Feather River at Oroville, California (simulated unregulated peaks)			
Threshold	1998–2006	391,000–INF	
Q	1902	41,000	1
Q	1903	102,000	1
Q	1904	118,000	1
Q	1905	81,000	1
Q	1906	128,000	1
Q	1907	230,000	1
Q	1908	16,300	1
Q	1909	140,000	1
Q	1910	31,000	1
Q	1911	84,500	
Q	1912	16,300	
Q	1913	14,000	
Q	1914	99,900	
Q	1915	76,900	
Q	1916	46,300	
Q	1917	81,700	
Q	1918	29,700	
Q	1919	50,000	
Q	1920	21,700	
Q	1921	56,400	
Q	1922	37,100	
Q	1923	21,200	
Q	1924	34,500	
Q	1925	55,600	
Q	1926	50,400	
Q	1927	92,800	
Q	1928	146,000	
Q	1929	11,700	
Q	1930	8,7300	
Q	1931	9,310	
Q	1932	18,700	
Q	1933	8,800	
Q	1934	17,000	
Q	1935	58,200	
Q	1936	62,600	
Q	1937	19,600	
Q	1938	189,000	

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11407099 Feather River at Oroville, California (simulated unregulated peaks)—Cont.		
Q	1939	7,860
Q	1940	158,000
Q	1941	82,000
Q	1942	101,000
Q	1943	72,100
Q	1944	18,900
Q	1945	51,500
Q	1946	50,100
Q	1947	33,900
Q	1948	35,100
Q	1949	14,000
Q	1950	43,200
Q	1951	77,900
Q	1952	51,000
Q	1953	113,000
Q	1954	52,100
Q	1955	11,500
Q	1956	218,000
Q	1957	69,800
Q	1958	86,000
Q	1959	29,900
Q	1960	113,000
Q	1961	15,600
Q	1962	38,100
Q	1963	160,000
Q	1964	20,700
Q	1965	214,000
Q	1966	17,000
Q	1967	59,300
Q	1968	42,900
Q	1969	161,000
Q	1970	136,000
Q	1971	71,300
Q	1972	20,200
Q	1973	52,300
Q	1974	125,000
Q	1975	33,500

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11407099 Feather River at Oroville, California (simulated unregulated peaks)—Cont.			
Q	1976	11,800	
Q	1977	3,850	
Q	1978	60,100	
Q	1979	24,000	
Q	1980	162,000	
Q	1981	19,000	
Q	1982	113,000	
Q	1983	113,000	
Q	1984	83,700	
Q	1985	17,600	
Q	1986	264,000	
Q	1987	32,400	
Q	1988	18,900	
Q	1989	98,200	
Q	1990	14,600	
Q	1991	54,000	
Q	1992	24,900	
Q	1993	64,900	
Q	1994	9,030	
Q	1995	157,000	
Q	1996	63,500	
Q	1997	391,000	
Station - 11413599 North Yuba River below Bullards Bar Dam, California (simulated unregulated peaks)			
Threshold	1939–1940	160,000–INF	
Threshold	1967–1969	160,000–INF	
Threshold	1998–2006	160,000–INF	
Q	1938	69,800	1
Q	1941	30,600	1
Q	1942	28,100	1
Q	1943	44,200	1
Q	1944	5,110	1
Q	1945	26,400	1
Q	1946	20,000	1
Q	1947	18,600	1
Q	1948	15,200	1
Q	1949	5,060	1
Q	1950	13,300	1

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11413599 North Yuba River below Bullards Bar Dam, California (simulated unregulated peaks)—Continued			
Q	1951	47,100	1
Q	1952	23,200	1
Q	1953	34,000	1
Q	1954	31,600	1
Q	1955	5,160	1
Q	1956	70,000	1
Q	1957	30,200	1
Q	1958	34,600	1
Q	1959	10,100	1
Q	1960	43,600	1
Q	1961	3,540	1
Q	1962	21,000	1
Q	1963	83,000	1
Q	1964	10,400	1
Q	1965	91,600	1
Q	1966	4,420	1
Q	1970	6,4600	
Q	1971	22,000	
Q	1972	10,600	
Q	1973	20,500	
Q	1974	39,900	
Q	1975	16,300	
Q	1976	4,750	
Q	1977	1,370	
Q	1978	19,900	
Q	1979	9,810	
Q	1980	94,300	
Q	1981	8,780	
Q	1982	66,500	
Q	1983	47,900	
Q	1984	38,700	
Q	1985	8,040	
Q	1986	123,000	
Q	1987	17,200	
Q	1988	5,790	

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11413599 North Yuba River below Bullards Bar Dam, California (simulated unregulated peaks)—Continued			
Q	1989	36,500	
Q	1990	8,140	
Q	1991	33,500	
Q	1992	13,300	
Q	1993	21,000	
Q	1994	3,100	
Q	1995	47,100	
Q	1996	33,700	
Q	1997	160,000	
Station - 11446599 American River at Fair Oaks, California (simulated unregulated peaks)			
Threshold	1848–1861	300,000–INF	
Threshold	1863–1904	300,000–INF	
Threshold	1999–2006	300,000–INF	
QINT	1862	300,000–336,000	
Q	1905	24,200	1
Q	1906	59,700	1
Q	1907	156,000	1
Q	1908	10,300	1
Q	1909	119,000	1
Q	1910	47,000	1
Q	1911	81,300	1
Q	1912	12,100	1
Q	1913	12,700	1
Q	1914	74,100	1
Q	1915	47,900	1
Q	1916	40,700	1
Q	1917	42,300	1
Q	1918	15,800	
Q	1919	67,500	1
Q	1920	20,100	1
Q	1921	39,200	1
Q	1922	31,600	1
Q	1923	39,000	1
Q	1924	14,000	1
Q	1925	99,500	1
Q	1926	27,400	1

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11446599 American River at Fair Oaks, California (simulated unregulated peaks)—Cont.			
Q	1927	67,700	1
Q	1928	163,000	1
Q	1929	26,000	1
Q	1930	24,400	1
Q	1931	9,900	1
Q	1932	21,100	1
Q	1933	16,500	1
Q	1934	22,600	1
Q	1935	60,900	1
Q	1936	58,300	1
Q	1937	33,000	1
Q	1938	114,000	1
Q	1939	10,900	1
Q	1940	89,200	1
Q	1941	38,800	1
Q	1942	83,200	1
Q	1943	152,000	1
Q	1944	20,100	1
Q	1945	94,400	1
Q	1946	42,200	1
Q	1947	27,900	1
Q	1948	21,000	1
Q	1949	37,500	1
Q	1950	34,400	1
Q	1951	180,000	1
Q	1952	37,200	1
Q	1953	49,700	1
Q	1954	42,600	1
Q	1955	13,300	
Q	1956	273,000	
Q	1957	49,500	
Q	1958	57,000	
Q	1959	19,800	
Q	1960	86,500	
Q	1961	8,590	
Q	1962	47,100	
Q	1963	218,000	

Table A4. Estimated, unregulated annual peak discharge for key dam sites in the Sacramento–San Joaquin River Basin, California.—Continued

[For some stations, missing discharges are represented by interval discharge thresholds. These sites have a row of data after the STATION row with THRESHOLD indicated. The next two numbers in the THRESHOLD row show the first and last year of missing record. The next number represents a large discharge threshold that presumably was not reached during the missing years. The INF following the threshold discharge simply indicates the largest possible discharge during the missing years (infinitely large). The first column after the STATION and THRESHOLD rows indicates whether discharge is a single-valued discharge (Q) or an interval discharge (QINT). Second column is the water year for each discharge. For each single-valued discharge, the third column is the discharge in cubic feet per second, and for each interval discharge, the third and fourth columns represent the range in discharge in cubic feet per second. For each single-valued discharge, a value of 1 in the last column indicates that the discharge is a recorded, unregulated, discharge. A value of 2 in the last column indicates that the discharge was a recorded, regulated discharge]

Station - 11446599 American River at Fair Oaks, California (simulated unregulated peaks)—Cont.		
Q	1964	22,000
Q	1965	264,000
Q	1966	10,900
Q	1967	48,500
Q	1968	32,500
Q	1969	116,000
Q	1970	123,000
Q	1971	45,500
Q	1972	127,00
Q	1973	66,900
Q	1974	54,700
Q	1975	39,900
Q	1976	132,00
Q	1977	27,90
Q	1978	41,500
Q	1979	23,800
Q	1980	177,000
Q	1981	20,000
Q	1982	159,000
Q	1983	94,800
Q	1984	89,600
Q	1985	17,300
Q	1986	246,000
Q	1987	14,900
Q	1988	6,700
Q	1989	45,300
Q	1990	9,490
Q	1991	36,200
Q	1992	17,000
Q	1993	45,700
Q	1994	6,140
Q	1995	94,100
Q	1996	74,100
Q	1997	369,000
Q	1998	56,400

This page intentionally left blank.

Appendix B. Extended Bayesian GLS Regional Skew Analysis for California

The U.S. Geological Survey (USGS) Southeastern United States regional skew study shows how a Bayesian generalized least squares (B-GLS) analysis typically is conducted (Feaster and others, 2009; and Gotvald and others, 2009; Veilleux, 2009; Weaver and others, 2009). The cross correlations between annual-peak discharges in California were considerably larger than those in the Southeastern United States study (fig. 4). When a B-GLS analysis was attempted on the California data set, results were not consistent because of the high cross correlations. A Bayesian GLS analysis seeks to exploit the cross correlations among the sample skews to obtain the best possible estimates of model parameters. If the cross correlations are large, the GLS estimators can become more complicated as a result of the effort to find the most efficient estimator of the parameters. The accuracy of the model (eq. 9) used to calculate the cross correlations of concurrent annual-peak discharges in California was equivalent to 52 years of actual at-site data. While this model describes the overall structure of the California data set, the calculated cross correlation between any two sites is not precise enough to justify the sophisticated weights (both positive and negative) that the Bayesian GLS analysis generates. Thus, an alternative procedure was developed so that the regional skew analysis would provide reliable results.

To this end, a weighted least squares (WLS) analysis that does not use cross correlations was first used to develop estimators of the regression coefficients for each regional skew model. After the regression model coefficients were determined using WLS, the precision of the model and the precision of the regression coefficients were estimated using a modified GLS analysis. However, because of the extensive use of low-outlier censoring and historical information in the expected moments algorithm (EMA) analysis used in California, the simple formulas provided in Bulletin 17B and by Griffis and Stedinger (2009) do not reliably represent the variance of the sample skewness estimators. Thus, a Monte Carlo study was done to determine the actual sample variance of the skewness coefficient when a low outlier test is employed to identify samples for special treatment. Finally, a modified Bayesian GLS analysis using only data from pristine sites (that is, sites without low outliers, zero flows, reconstructed records at key dam sites, or historical information) provided the estimate of the model error variance (the precision of the model) and the precision of the estimated model parameters.

The specific computational steps used in the California regional skew analysis are described below.

1. A WLS analysis was used to derive the regression model parameter estimates using the complete set of records. The resultant model yielded an unbiased regional estimator of the skew at any site. The WLS analysis explicitly reflects variations in record length, but as previously described, does not consider cross correlations among the skewness estimators.

   The WLS analysis was done in two steps using unbiased at-site sample skewness estimators and variances of at-site unbiased skewness estimators when possible. A correction factor developed by Tasker and Stedinger (1986) and applied by Reis and others (2005) was used to unbias the at-site skews and the estimates of their variance. Equations to derive the unbiased at-site skews and their variances are given below.

$$\hat{\gamma}_i = \left[1 + \frac{6}{N_i}\right] G_i \,, \qquad \text{(B1)}$$

   and

$$Var\left[\hat{\gamma}_i\right] = \left[1 + \frac{6}{N_i}\right]^2 Var\left[G_i\right], \qquad \text{(B2)}$$

   where

   $\hat{\gamma}_i$ is the unbiased at-site skew for site i,
   G_i is the traditiional biased at-site skew estimator for site i, and
   N_i is the systematic record length at site i.

   When unbiasing the skew, only the number of systematic peaks was considered. Thus, any historical flood periods in the EMA analysis were not included in the calculation of the correction factors.

In the first step of the WLS analysis, BayesianWeighted Least Squares (B-WLS) was used to estimate the model error variance, denoted $\sigma^2_{\delta,B\text{-}WLS}$. Then, $\sigma^2_{\delta,B\text{-}WLS}$ was used in a method-of-moments WLS (MM-WLS) analysis to generate the weights, **W**, needed to estimate the regression parameters $\hat{\boldsymbol{\beta}}$. In order to compute the MM-WLS weights, a diagonal covariance matrix, $\boldsymbol{\Lambda}_{WLS}(\sigma^2_{\delta,B\text{-}WLS})$, was created. The diagonal elements of the covariance matrix are the sums of the estimated error variance, $\sigma^2_{\delta,B-WLS}$, and the variance of the unbiased at-site skew, $Var\left[\hat{\gamma}_i\right]$, which depends only on the at-site annual peak flow record length, and the estimated skew for that site. The at-site skew, G_i, and the variance of the at-site skew estimator, $Var[G_i]$, were calculated from the entire flow record, including historical and systematic peaks, using the EMA method (Cohn and others, 1997). These biased estimates of the skew and the variance of the skew produced by EMA were then unbiased using equations B1 and B2, respectively, which gave estimates of the unbiased at-site skew, $\hat{\gamma}_i$ and the at-site skew variance $Var[\hat{\gamma}_i]$. The off-diagonal elements of $\boldsymbol{\Lambda}_{WLS}(\sigma^2_{\delta,B\text{-}WLS})$ were zero, because cross correlations between gage sites were not considered in the WLS analysis. Thus, the ($n \times n$) covariance matrix, $\boldsymbol{\Lambda}_{WLS}(\sigma^2_{\delta,B\text{-}WLS})$, was

$$\boldsymbol{\Lambda}_{WLS}\left(\sigma^2_{\delta,B\text{-}WLS}\right) = \sigma^2_{\delta,B\text{-}WLS}\mathbf{I} + \mathbf{diag}\left(Var\left[\hat{\boldsymbol{\gamma}}\right]\right), \qquad \text{(B3)}$$

where

$\mathbf{I}$ is an ($n \times n$) identity matrix,

n is the number of gage sites in the study, and

$\mathbf{diag}(Var[\hat{\boldsymbol{\gamma}}])$ is an ($n \times n$) matrix containing the variances of the unbiased at-site sample skewness estimators, $Var[\hat{\boldsymbol{\gamma}}]$, on the diagonal and zeros on the off-diagonal.

Using that covariance matrix, the MM-WLS weights are calculated as

$$\mathbf{W} = \left[\mathbf{X}^T \boldsymbol{\Lambda}^{-1}{}_{WLS}\left(\sigma^2_{\delta,B\text{-}WLS}\right)\mathbf{X}\right]^{-1} \mathbf{X}^T \boldsymbol{\Lambda}^{-1}{}_{WLS}\left(\sigma^2_{\delta,B\text{-}WLS}\right), \qquad \text{(B4)}$$

where

$\mathbf{W}$ is ($k \times n$) matrix of weights,

$\mathbf{X}$ is an ($n \times k$) matrix of basin parameters, and

k is the number of basin characteristics.

These weights were used to compute the final estimates of the regression parameters $\hat{\boldsymbol{\beta}}$ as

$$\hat{\boldsymbol{\beta}} = \mathbf{W}\hat{\boldsymbol{\gamma}}, \qquad \text{(B5)}$$

where

$\hat{\boldsymbol{\beta}}$ is an ($k \times 1$) vector of regression parameters, and

$\hat{\boldsymbol{\gamma}}$ is an ($n \times 1$) vector of unbiased at-site sample skewness estimators.

Both the B-WLS and the MM-WLS analyses include sites that had historical information, zero flows, and low outliers.

2. After estimating the regression parameters, the true model error variance needs to be estimated. However, the extensive censoring of low outliers, the occurrence of zero flows, and the addition of regional historical flood information for some sites in the data set complicated the estimation of the model error variance. Thus, to estimate the true model error variance, a simpler, “pristine” data set was developed. The pristine data set is a subset of the larger data set used in the WLS estimation of the regression parameters, and it does not include sites that had zero flows, low outliers as determined by EMA, or any reconstructed flow records. Any historical flood information was ignored.

   Because the pristine data set excluded sites with low outliers, the formulas provided in Bulletin 17B and by Griffis and Stedinger (2009) misrepresented the variance of the sample skewness estimators. Thus, a new Monte Carlo study was done to determine the actual variance of the skewness coefficient when a low outlier test is used to exclude some data. The Monte Carlo analysis of sample skews from an LP3 distribution used only complete samples with no low outliers. Results from the Monte Carlo analysis were used to determine the bias associated with the sample skewness coefficient **G** when samples with low outliers were dropped from the analysis. Two functions were computed: the mean of the sample skew, denoted $m(\boldsymbol{\gamma}, \mathbf{N})$, and its variance, denoted $v(\boldsymbol{\gamma}, \mathbf{N})$,

$$m(\boldsymbol{\gamma}, \mathbf{N}) = E_{\{x_i|no-outliers\}}\left[\mathbf{G} \mid \boldsymbol{\gamma}, \mathbf{N}\right], \qquad \text{(B6)}$$

and

$$v(\boldsymbol{\gamma}, \mathbf{N}) = Var_{\{x_i|no-outliers\}}\left[\mathbf{G} \mid \boldsymbol{\gamma}, \mathbf{N}\right]. \qquad \text{(B7)}$$

These expectations were computed using only those LP3 samples, $\{x_i\}$, that did not contain low outliers, as determined by a 10 percent Grubbs-Beck test recommended by Bulletin 17B. Figure B1 summarizes the Monte Carlo results for 1,000 simulations of $N = 50$ years of at-site annual-peak discharges. The x-axis represents the population (true) skew, $\boldsymbol{\gamma}$, and the y-axis represents both the mean of the estimated skew $m(\boldsymbol{\gamma}, \mathbf{N})$ and the standard deviation of the estimated skew [the square root of the variance $v(\boldsymbol{\gamma}, \mathbf{N})$] for LP3 samples that had no low outliers. Figure B1 shows that when only samples without low outliers are considered, the mean of the estimated skews (the dashed line) can be significantly biased. Samples with low outliers are very likely to be negatively skewed, so analyzing data using an unbiased estimator and omitting samples with low outliers is expected to yield a regional skewness estimator that is positively biased if the bias is not corrected. When the true skew is highly positive, the bias is small and slightly negative, as few samples are omitted because of low outliers. When the true skew is highly negative, the standard deviation of the sample is greatly reduced, as shown by the dotted line in figure B1.

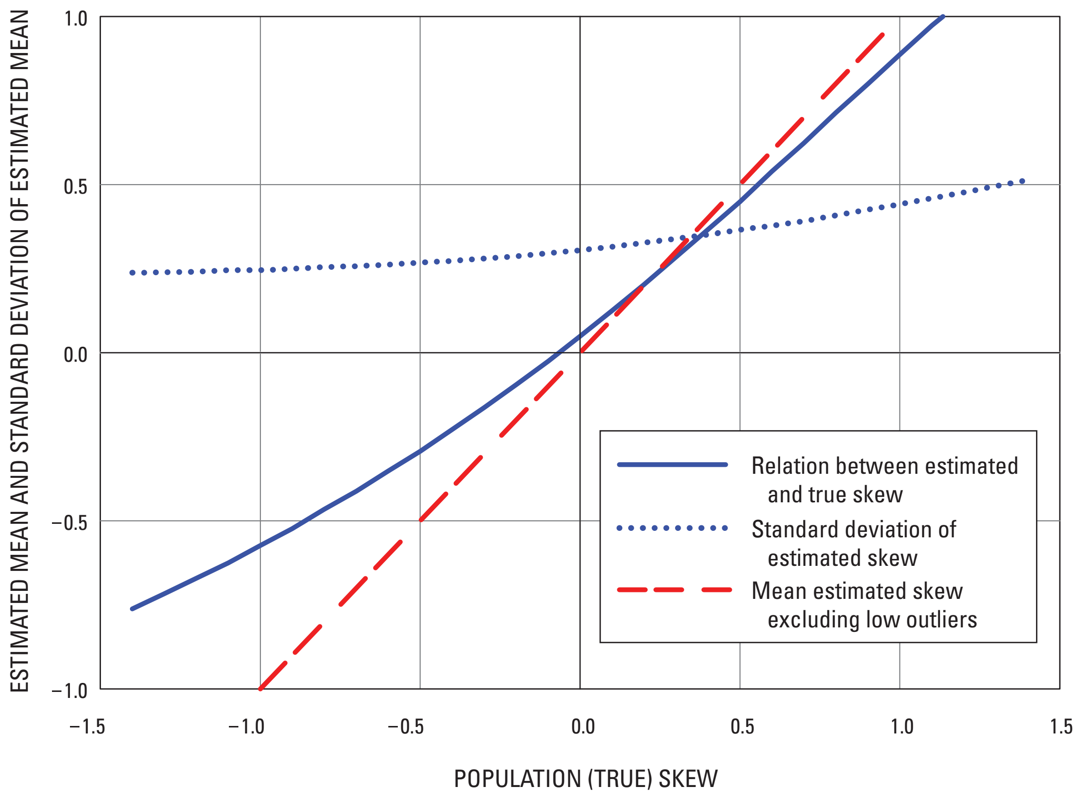


Figure B1. Relation between estimated skew and population (true) skew based on 1,000 Monte Carlo simulations of 50-year samples in California. The dashed line represents the mean, $m(\gamma,\mathbf{N})$, of the estimated skew across samples without outliers, and the dotted line represents the standard deviation [the square root of the variance, $v(\gamma,\mathbf{N})$] of that estimated skew across LP3 samples that do not contain low outliers.

Skew was more negative when the dataset had more samples with low outliers (table B1). For example, 32 percent of the samples with a true at-site skew of -0.5 had low outliers and were therefore dropped, but only 10 percent of the samples with a true at-site skew of 0.0 were dropped. On the other hand, when the true at-site skew was highly positive, few samples were rejected, resulting in very little impact on the sampling distribution of the estimated skew. In the Monte Carlo simulation, when the at-site record length was 50 years and the skew was 0.3, only 2 percent of the samples were rejected. When the skew was 0.5 or greater, none of the 50,000 samples generated contained low outliers (table B1).

The pattern evident in table B1 shows why the standard deviation of the estimated skew in samples without outliers increases as the true skew increases.

Table B1. Monte Carlo results showing the percent of samples, for different at-site skew values, dropped from the simulation due to the presence of low outliers

Skew:	-1.5	-1.0	-0.8	-0.5	-0.3	0.0	0.3	0.5	0.8	1.0	1.5
% of Samples Dropped:	71%	56%	47%	32%	22%	10%	2%	0%	0%	0%	0%

3. The Monte Carlo experiments provided the expected value and variance of the sample skewness estimator based on samples without outliers. These values can be used with the pristine data set in the computations of the model error variance using B-GLS. By using the relationship for regional skew generated by the WLS analysis, the WLS mean regional skew estimate, $m_R(i)$, can be calculated for each site i in the pristine data set as

$$\mathbf{m}_R = \mathbf{X}\hat{\boldsymbol{\beta}}, \quad \text{(B8)}$$

where

$\mathbf{m}_R$ is an $(n_p \times 1)$ vector of WLS regional skew estimates for each site in the pristine data set,

$\mathbf{X}$ is an $(n_p \times k)$ matrix of basin parameters,

$\hat{\boldsymbol{\beta}}$ is an $(k \times 1)$ vector of WLS regression parameters,

n_p is the number of gage sites in the pristine data set, and

k is the number of basin characteristics.

4. The last step (4) is to estimate the model error variance using the pristine data set. If the model error variance, $\sigma^2_{\delta,B-GLS}$ were zero, then all of the observed variability would be sampling error, and we would have

$$E\left[\left\{G_i - m(\gamma_i, N_i)\right\}^2\right] = v(\gamma_i, N_i). \quad \text{(B9)}$$

However, we anticipate that the model will not be perfect, and thus estimating the model error variance will be challenging. The derivative

$$d\left\{m(\boldsymbol{\gamma}, \mathbf{N})\right\} / d\boldsymbol{\gamma} \neq 1 \quad \text{(B10)}$$

will be used to correct the GLS analysis. Let the derivative in B10 be replaced by r,

$$r = d\left\{m(\boldsymbol{\gamma}, \mathbf{N})\right\} / d\boldsymbol{\gamma}, \quad \text{(B11)}$$

and let $\sigma^2_{\delta,B\text{-}GLS}$ be the model error variance. Then a first-order approximation of the expected value of regional skew variability would be

$$E\left[\left\{G_i - m(\gamma_i, N_i)\right\}^2\right] = r^2\sigma^2_{\delta,B\text{-}GLS} + v(\gamma_i, N_i). \quad \text{(B12)}$$

Thus, the GLS covariance matrix for the pristine data set is

$$\mathbf{\Lambda}_p\left(\sigma^2_{\delta,B-GLS}\right) = r^2\sigma^2_{\delta,B-GLS}\,\mathbf{I}_p + \mathbf{\Sigma}(\mathbf{G}), \tag{B13}$$

where

$\mathbf{\Lambda}_p\left(\sigma^2_{\delta,B-GLS}\right)$ is an $\left(n_p \times n_p\right)$ GLS covariance matrix,

$\mathbf{I}_p$ is an $\left(n_p \times n_p\right)$ identity matrix, and

$\mathbf{\Sigma}(\mathbf{G})$ is an $\left(n_p \times n_p\right)$ matrix containing the sampling variances of the biased skewness estimators $Var\left[G_i\right] = v\left(\gamma_i, N_i\right)^2$ and the covariances of the skewness estimators G_i in the pristine data set.

The values of $\mathbf{\Sigma}(\mathbf{G})$ were determined by the the cross-correlation of concurrent systematic annual peak flows (eq. 6) and the *cf* factor (eq. 7). In calculating the *cf* factor used to determine the ratio of the number of concurrent peak flows at a pair of sites to the total number of peak flows at both sites, only the systematic records were considered.

The covariance matrix for the skewness coefficients for the pristine data set, $\mathbf{\Lambda}_p\left(\sigma^2_{\delta,B\text{-}GLS}\right)$, and the conditional means of the sample skews, $m(\boldsymbol{\gamma},\mathbf{N})$ are used in a Bayesian framework to compute the posterior distribution of the model error variance, and in particular, the posterior mean of the true model error variance.

The B-GLS model error variance can then be used to compute the precision of the regression parameters, $\hat{\boldsymbol{\beta}}$, that were calculated with the WLS weights, $\mathbf{W}$, as $\hat{\boldsymbol{\beta}} = \mathbf{W}\hat{\boldsymbol{\gamma}}$. Using the posterior mean of the true model error variance in $\mathbf{\Lambda}$, the variance of $\hat{\boldsymbol{\beta}}$ is simply

$$Var\left[\hat{\boldsymbol{\beta}}\right] = \mathbf{W}\mathbf{\Lambda}\left(\sigma^2_{\delta,B-GLS}\right)\mathbf{W}^T, \tag{B14}$$

where $\mathbf{\Lambda}\left(\sigma^2_{\delta,B\text{-}GLS}\right)$ is an $(n \times n)$ covariance matrix that uses all of the sites, not just those listed in the pristine data set, and $\left(\sigma^2_{\delta,B\text{-}GLS}\right)$ is the posterior mean of the model error variance calculated from the B-GLS analysis described above. It is important to note that $\mathbf{\Lambda}\left(\sigma^2_{\delta,B\text{-}GLS}\right)$ is not the same as the covariance matrix $\mathbf{\Lambda}_{WLS}\left(\sigma^2_{\delta,B\text{-}WLS}\right)$ used in the MM-WLS analysis of all of the sites or the $\mathbf{\Lambda}_p\left(\sigma^2_{\delta,B\text{-}GLS}\right)$ used in the B-GLS analysis of just the pristine sites. Instead, $\mathbf{\Lambda}\left(\sigma^2_{\delta,B\text{-}GLS}\right)$ is

$$\mathbf{\Lambda}\left(\sigma^2_{\delta,B\text{-}GLS}\right) = \sigma^2_{\delta,B\text{-}GLS}\,\mathbf{I} + \mathbf{\Sigma}(\hat{\boldsymbol{\gamma}}), \tag{B15}$$

where

$\mathbf{\Lambda}\left(\sigma^2_{\delta,B\text{-}GLS}\right)$ is an $(n \times n)$ GLS covariance matrix,

$\mathbf{I}$ is an $(n \times n)$ identity matrix,

$\mathbf{\Sigma}(\hat{\boldsymbol{\gamma}})$ is an $(n \times n)$ matrix containing the sampling variances of the unbiased skewness estimators $Var\left[\hat{\gamma}_i\right]$ and the covariances of the skewness estimators $\hat{\gamma}_i$, based upon the size and the WLS estimator of the skew at each site.

Off-diagonal elements of $\mathbf{\Sigma}(\hat{\boldsymbol{\gamma}})$ were estimated by the the cross-correlation of concurrent systematic annual peak flows (eq. 6) and the *cf* factor (eq. 7), which depends on the record lengths and concurrent record lengths. In calculating the *cf* factor used to determine the ratio of the number of concurrent peak flows at a pair of sites to the total number of peak flows at both sites, only the systematic records were considered. Thus, any historical floods used in the EMA analysis were not used to calculate the cross-correlation of peak flows or the *cf* factor.

This page intentionally left blank.

Appendix C. Bayesian GLS Regression Diagnostics

To determine if a model is a good representation of the data and which regression parameters, if any, should be included in a regression model, diagnostic statistics have been developed to evaluate how well a model fits a regional hydrologic data set (Griffis and Stedinger, 2006; Gruber and others, 2008). In this report, the goal was to determine the set of possible explanatory variables that best fit the California peak-discharge data affording the most accurate skew prediction while also keeping the model as simple as possible. This appendix presents the diagnostic statistics for a Bayesian weighted least squares (WLS) or generalized least squares (GLS) analysis, and discusses the specific values obtained for the California regional skew study.

Table C1 presents a Pseudo analysis of variance (Pseudo ANOVA) table for the California regional skew analysis. The table contains regression diagnostics/goodness of fit statistics which are explained below.

In particular, the table provides equations for determining how much of the variation in the observations can be attributed to the regional model, and how much of the residual variation can be attributed to model error and sampling error, respectively. Difficulties arise in determining these quantities. The model errors cannot be resolved because the values of the sampling errors, η_i, for each site, i, are not known. However, the total sampling error sum of squares can be determined by $\sum_{i=1}^{n} Var[\hat{\gamma}_i]$. Because there are n equations, the total model error sum of squares due to the model error, δ, for a model with k parameters is $n\sigma_\delta^2(k)$. Thus, the residual variation attributed to the sampling error is $\sum_{i=1}^{n} Var[\hat{\gamma}_i]$, and the residual variation attributed to the model error is $n\sigma_\delta^2(k)$.

For a model with no parameters other than the mean (that is, the constant skew model), the estimated model error variance $\sigma_\delta^2(0)$ describes all of the anticipated variation in $\gamma_i = \mu + \delta_i$, where μ is the mean of the estimated at-site sample skews. Thus, the total expected sum of squares variation due to model error, δ_i, and due to sampling error, $\eta_i = \hat{\gamma}_i - \gamma_i$ should equal $n\sigma_\delta^2(0) + \sum_{i=1}^{n} Var\left(\hat{\gamma}_i\right)$. Therefore, the expected sum of squares attributed to a regional skew model with k parameters equals $n[\sigma_\delta^2(0) - \sigma_\delta^2(k)]$, because the sum of the model error variance $n\sigma_\delta^2(k)$ and the variance explained by the model must sum to $n\sigma_\delta^2(0)$. Table C1 shows results from models when k = 0 and 1.

Table C1. Pseudo ANOVA table for the Constant model and the NL-Elev model for regional skew in California.

[ANOVA, analysis of variance; NL-Elev model, nonlinear regional skew model; k, the number of parameters used in the regional regression; n, the number of stations used in the regional skew regression; EVR, error variance ratio; MBV*, misrepresentation of the beta variance; R_δ^2, Pseudo-R_δ^2, %, percent]

Source		Degrees of freedom		Equations	Sum of squares	
		Constant	NL-Elev		Constant	NL-Elev
Model	***k***	0	1	$n\left[\sigma_\delta^2(0) - \sigma_\delta^2(k)\right]$	0.0	15
Model error	***n*-*k*-1**	157	156	$n\sigma_\delta^2(k)$	32	17
Sampling error	***n***	158	158	$\sum_{i=1}^{n} Var\left[\hat{\gamma}_i\right]$	34	34
Total	**2*n*-1**	315	315	$n\sigma_\delta^2(0) + \sum_{i=1}^{n} Var\left[\hat{\gamma}_i\right]$	66	66
EVR					1.1	2.1
MBV*					13	16
R_δ^2					0%	48%

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

Table C1 compares the Pseudo ANOVA results for the Constant model and the NL-Elev model. Both models have the same sampling error because both use the same set of at-site skew data. Both have sampling error variances greater than their model error variances. However, the model error attributed to the NL-Elev model, $\sigma_\delta^2(1)$, is slightly more than half of the model error variance attributed to the Constant model, $\sigma_\delta^2(0)$. This difference is accounted for by the variation in the sample that the NL-Elev model appears to explain. Because the Constant model does not have any explanatory variables, the variation attributed to that model is 0.0. On the other hand, the NL-Elev model has one explanatory variable, which causes the variation attributed to the model to increase to 15. This reduces the model error variance from 32 with the Constant model to 17 with the NL-Elev model; thus, adding the nonlinear elevation explanatory variable in the NL-Elev model greatly improves the ability of the model to describe the observed skew coefficients. This impact is described by the Pseudo R_δ^2, which in this case is 48 percent because the NL-Elev model explains 48 percent of the estimated variation, $\sigma_\delta^2(0)$, in the true skew among the sites.

The Pseudo analysis of variance also provides the information needed to evaluate whether or not a sophisticated WLS or GLS analysis is needed to correctly interpret the data. In particular, the error variance ratio (EVR) is a modeling diagnostic used to evaluate whether a simple ordinary least squares (OLS) regression is sufficient or a more sophisticated WLS or GLS analysis is appropriate. EVR is the ratio of the average sampling error variance to the model error variance. Generally, an EVR greater than 20 percent indicates that the sampling variance is not significant when compared to the model error variance, suggesting the need for a WLS or GLS regression analysis. The EVR is calculated as

$$\mathrm{EVR} = \frac{SS(\text{sampling error})}{SS(\text{model error})} = \frac{\sum_{i=1}^{n} Var\left(\hat{\gamma}_i\right)}{n\sigma_\delta^2(k)}, \qquad \text{(C1)}$$

where

SS is the sum of squares.

The EVR was 1.1 for the constant model and 2.1 for the NL-Elev model. The sampling variability of skewness estimators was larger than model error variance of the regional model. Thus, given the variation of record lengths from site-to-site, using a WLS or GLS analysis is important for for evaluating the final precision of the model rather than using a simpler analysis that neglects the sampling error in the at-site skewness estimators.

The misrepresentation of the beta variance (MBV*) statistic is used to determine whether a WLS regression is sufficient or a GLS regression is appropriate to determine the precision of the estimated regression parameters (Griffis and Stedinger, 2006). The MBV* describes the error produced by a WLS regression analysis by evaluating the precision of b_0^{WLS}, which is the estimator of the WLS regression constant, β_0^{WLS}. The correlation among estimated at-site skews, which is ignored in WLS regression, generally has its greatest impact on the precision of the regression constant term (Stedinger and Tasker, 1985). If the MBV* is substantially greater than 1, a GLS error analysis should be used. The MBV* is calculated as

$$\mathrm{MBV}^* = \frac{Var\left[b_0^{WLS} \mid GLS\ analysis\right]}{Var\left[b_0^{WLS} \mid WLS\ analysis\right]} = \frac{\mathbf{w}^T \boldsymbol{\Lambda} \mathbf{w}}{\sum_{i=1}^{n} w_i}, \qquad \text{(C2)}$$

where

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

For the California regional skew study, the MBV* was equal to 16 for the NL-Elev model and 13 for the Constant model. This is a very large value indicating that the cross-correlation among the skewness estimators has had a major effect on the precision with which the regional average skew coefficient can be estimated; if a WLS precision analysis had been used for the estimated constant parameter in the NL-Elev model, the variance would have been underestimated by a factor of 16. Thus a WLS analysis would have seriously misrepresented the variance of the constant in the Constant model and in the NL-Elev model of regional skew. This would have caused the variance of prediction to be underestimated given that the sampling error in the constant term in both models was large enough to contribute appreciably to the average variance of prediction.

Publishing support provided by the U.S. Geological Survey
Publishing Network, Sacramento and Tacoma Publishing Service Centers

For more information concerning the research in this report, contact the
Director, California Water Science Center
Placer Hall
6000 J street
Sacramento, CA 95819
http://ca.water.usgs.gov/

Parrett and others—**Regional Skew for California, and Flood Frequency for Selected Sites in the Sacramento-San Joaquin River Basin, Based on Data through Water Year 2006**

SIR 2010–5260

Printed on recycled paper