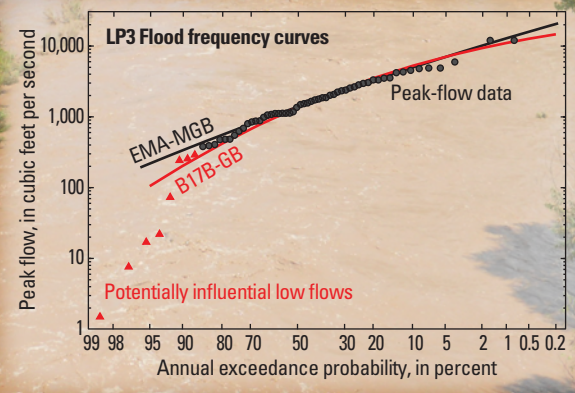
USGS
science for a changing world

Prepared in cooperation with the Flood Control District of Maricopa County

Evaluation of the Expected Moments Algorithm and a Multiple Low-Outlier Test for Flood Frequency Analysis at Streamgaging Stations in Arizona



Scientific Investigations Report 2014–5026

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

FRONT COVER
Flood frequency analysis plot comparing the EMA-MGB test to the B17B-GB test for the Santa Cruz River near Lochiel, Arizona. Background photographs show the San Pedro River at Charleston, Arizona, at low flow (top) and in flood (bottom). Photographs by Kurt Ehrenberg.

Evaluation of the Expected Moments Algorithm and a Multiple Low-Outlier Test for Flood Frequency Analysis at Streamgaging Stations in Arizona

By Nicholas V. Paretti, Jeffrey R. Kennedy, and Timothy A. Cohn

Prepared in cooperation with the Flood Control District of Maricopa County

Scientific Investigations Report 2014–5026

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

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

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

U.S. Geological Survey, Reston, Virginia: 2014

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

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

Suggested citation:
Paretti, N.V., Kennedy, J.R., and Cohn, T.A., 2014, Evaluation of the expected moments algorithm and a multiple low-outlier test for flood frequency analysis at streamgaging stations in Arizona: U.S. Geological Survey Scientific Investigations Report 2014–5026, 61 p., available online only at *http://dx.doi.org/10.3133/sir20145026*.

ISSN 2328–0328 (online)

Contents

Figures

Tables

Conversion Factors

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)
	Volume	
cubic foot (ft^3)	28.32	cubic decimeter (dm^3)
cubic foot (ft^3)	0.02832	cubic meter (m^3)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

Supplemental Information

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

°F=(1.8×°C)+32

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

°C=(°F-32)/1.8

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.”

Abbreviations

AEP	annual exceedance probability
AG	average gain
B17B	Bulletin 17B
B17B-GB	Bulletin 17B Grubbs-Beck test
BGLS	Bayesian generalized-least-squares
CPA	conditional probability adjustment
EMA	Expected Moments Algorithm
EMA-MGB	Expected Moments Algorithm multiple Grubbs-Beck test
ERL	effective record length
FFA	flood frequency analysis
HFAWG	Hydrologic Frequency Analysis Work Group
IQR	interquartile range
LP3	log-Pearson Type 3
MAPD	mean absolute percent difference
MSE	mean square error
NWIS	National Water Information System
PILF	potentially influential low-flow
RPD	relative percent difference
USACE	U.S. Army Corps of Engineers
USGS	U.S. Geological Survey

Evaluation of the Expected Moments Algorithm and a Multiple Low-Outlier Test for Flood Frequency Analysis at Streamgaging Stations in Arizona

By Nicholas V. Paretti, Jeffrey R. Kennedy, and Timothy A. Cohn

Abstract

Flooding is among the costliest natural disasters in terms of loss of life and property in Arizona, which is why the accurate estimation of flood frequency and magnitude is crucial for proper structural design and accurate floodplain mapping. Current guidelines for flood frequency analysis in the United States are described in Bulletin 17B (B17B), yet since B17B's publication in 1982 (Interagency Advisory Committee on Water Data, 1982), several improvements have been proposed as updates for future guidelines. Two proposed updates are the Expected Moments Algorithm (EMA) to accommodate historical and censored data, and a generalized multiple Grubbs-Beck (MGB) low-outlier test. The current guidelines use a standard Grubbs-Beck (GB) method to identify low outliers, changing the determination of the moment estimators because B17B uses a conditional probability adjustment to handle low outliers while EMA censors the low outliers. B17B and EMA estimates are identical if no historical information or censored or low outliers are present in the peak-flow data. EMA with MGB (EMA-MGB) test was compared to the standard B17B (B17B-GB) method for flood frequency analysis at 328 streamgaging stations in Arizona. The methods were compared using the relative percent difference (RPD) between annual exceedance probabilities (AEPs), goodness-of-fit assessments, random resampling procedures, and Monte Carlo simulations. The AEPs were calculated and compared using both station skew and weighted skew. Streamgaging stations were classified by U.S. Geological Survey (USGS) National Water Information System (NWIS) qualification codes, used to denote historical and censored peak-flow data, to better understand the effect that nonstandard flood information has on the flood frequency analysis for each method. Streamgaging stations were also grouped according to geographic flood regions and analyzed separately to better understand regional differences caused by physiography and climate.

The B17B-GB and EMA-MGB RPD-boxplot results showed that the median RPDs across all streamgaging stations for the 10-, 1-, and 0.2-percent AEPs, computed using station skew, were approximately zero. As the AEP flow estimates decreased (that is, from 10 to 0.2 percent AEP) the variability in the RPDs increased, indicating that the AEP flow estimate was greater for EMA-MGB when compared to B17B-GB. There was only one RPD greater than 100 percent for the 10- and 1-percent AEP estimates, whereas 19 RPDs exceeded 100 percent for the 0.2-percent AEP. At streamgaging stations with low-outlier data, historical peak-flow data, or both, RPDs ranged from −84 to 262 percent for the 0.2-percent AEP flow estimate. When streamgaging stations were separated by the presence of historical peak-flow data (that is, no low outliers or censored peaks) or by low outlier peak-flow data (no historical data), the results showed that RPD variability was greatest for the 0.2-AEP flow estimates, indicating that the treatment of historical and (or) low-outlier data was different between methods and that method differences were most influential when estimating the less probable AEP flows (1, 0.5, and 0.2 percent). When regional skew information was weighted with the station skew, B17B-GB estimates were generally higher than the EMA-MGB estimates for any given AEP. This was related to the different regional skews and mean square error used in the weighting procedure for each flood frequency analysis. The B17B-GB weighted skew analysis used a more positive regional skew determined in USGS Water Supply Paper 2433 (Thomas and others, 1997), while the EMA-MGB analysis used a more negative regional skew with a lower mean square error determined from a Bayesian generalized least squares analysis.

Regional groupings of streamgaging stations reflected differences in physiographic and climatic characteristics. Potentially influential low flows (PILFs) were more prevalent in arid regions of the State, and generally AEP flows were larger with EMA-MGB than with B17B-GB for gaging stations with PILFs. In most cases EMA-MGB curves would fit the largest floods more accurately than B17B-GB. In areas of the State with more baseflow, such as along the Mogollon Rim and the White Mountains, streamgaging stations generally had fewer PILFs and more positive skews, causing estimated AEP flows to be larger with B17B-GB than with EMA-MGB. The effect of including regional skew was similar for all regions, and the observed pattern was increasingly greater B17B-GB flows (more negative RPDs) with each decreasing AEP quantile.

A variation on a goodness-of-fit test statistic was used to describe each method's ability to fit the largest floods. The mean absolute percent difference between the measured peak flows and the log-Pearson Type 3 (LP3)-estimated flows, for each method, was averaged over the 90th, 75th, and 50th percentiles of peak-flow data at each site. In most percentile

subsets, EMA-MGB on average had smaller differences (1 to 3 percent) between the observed and fitted value, suggesting that the EMA-MGB-LP3 distribution is fitting the observed peak-flow data more precisely than B17B-GB. The smallest EMA-MGB percent differences occurred for the greatest 10 percent (90th percentile) of the peak-flow data. When stations were analyzed by USGS NWIS peak flow qualification code groups, the stations with historical peak flows and no low outliers had average percent differences as high as 11 percent greater for B17B-GB, indicating that EMA-MGB utilized the historical information to fit the largest observed floods more accurately.

A resampling procedure was used in which 1,000 random subsamples were drawn, each comprising one-half of the observed data. An LP3 distribution was fit to each subsample using B17B-GB and EMA-MGB methods, and the predicted 1-percent AEP flows were compared to those generated from distributions fit to the entire dataset. With station skew, the two methods were similar in the median percent difference, but with weighted skew EMA-MGB estimates were generally better. At two gages where B17B-GB appeared to perform better, a large number of peak flows were deemed to be PILFs by the MGB test, although they did not appear to depart significantly from the trend of the data (step or dogleg appearance). At two gages where EMA-MGB performed better, the MGB identified several PILFs that were affecting the fitted distribution of the B17B-GB method.

Monte Carlo simulations were run for the LP3 distribution using different skews and with different assumptions about the expected number of historical peaks. The primary benefit of running Monte Carlo simulations is that the underlying distribution statistics are known, meaning that the true 1-percent AEP is known. The results showed that EMA-MGB performed as well or better in situations where the LP3 distribution had a zero or positive skew and historical information. When the skew for the LP3 distribution was negative, EMA-MGB performed significantly better than B17B-GB and EMA-MGB estimates were less biased by more closely estimating the true 1-percent AEP for 1, 2, and 10 historical flood scenarios.

Introduction

The National Weather Service estimates that flooding caused approximately 50 billion dollars' worth of damage in the United States during the 1990s (National Weather Service-Hydrologic Information Center, 2001), and between 1955 and 2000 flood damage in Arizona exceeded 1.3 billion dollars (1995 dollars adjusted for inflation; Pielke and others, 2002). Reliable estimates of flood frequency and magnitude are necessary to effectively minimize the damage caused by floods and to accurately determine flood risk for the National Flood Insurance Program. Beginning in the 1960s, a national multiagency effort was organized to develop uniform and consistent methods for estimating flood frequency statistics in the United States. After several revisions, the Hydrology Subcommittee of the Interagency Advisory Committee on Water Data published Bulletin 17B (B17B), The "Guidelines for Determining Flood Flow Frequency" (Interagency Advisory Committee on Water Data, 1982). B17B recommends using the method-of-moments to fit a Pearson Type 3 distribution to the logarithms of the annual flood time series. The fitted-distribution approach provides a means for predicting large-magnitude and less probable floods beyond the observed data for the purpose of planning structural design and mapping floodplains. Predicted floods are commonly expressed in terms of the recurrence interval of a flood, such as the "100-year flood." The use of this terminology can be misleading to the public, because it implies a certain magnitude flood will occur once every 100 years, when in fact flood events are assumed to be random in B17B. Flood frequency estimates are now reported as annual exceedance probabilities (AEPs) to reinforce the fact that flood estimates are probabilistic.

A variety of probability distributions and parameter estimation methods have been tested as alternatives to the logarithm-Pearson Type 3 (LP3) method-of-moments and the B17B approach to flood-frequency analysis. Probability distribution types include normal, log-normal, Gumbel, log-Gumbel, and two- and three-parameter gamma distribution types (Beard, 1974; Interagency Advisory Committee on Water Data, 1982). The Extreme Value and Wakeby distributions are also commonly used in flood frequency analysis (Houghton, 1978; Rao and Hamed, 2000). Alternative methods for parameter estimation are the maximum likelihood, or probability weighted moments, and L-moments (Greenwood and others, 1979; Hosking, 1990). Several studies have tested the performance of several moment estimators using Monte Carlo experiments and compared the applicability of different distributions and moment estimation methods (Fill and Stedinger, 1995; Griffis and others, 2004; Potter and Lettenmaier, 1990). In the southwestern United States, Vogel and others (1993) found that the LP3, Generalized Extreme Value, and two- and three-parameter lognormal distributions all provided good approximations to flood-flow data. Despite extensive research investigating alternative distribution types and moment estimators, the LP3 distribution method-of-moments has proven to be a flexible, robust, and relatively straightforward approach to flood frequency analysis (Griffis and Stedinger, 2007).

Although B17B has served as the standard method in the United States for the past 30 years, it has several long-standing issues, which are discussed in B17B itself, concerning the treatment of low-outlier, historical, and censored flood information. Research by Cohn and others (2001), Stedinger and Griffis (2008), and Reis and others (2005) have shown that the computed confidence intervals fail to represent the correct uncertainty in the station skew coefficient, and that the recommended statistical procedures for computing a regional skew coefficient are not adequate for estimating the accuracy and precision of the skew estimate. To address several of these inconsistencies, the Expected Moments Algorithm with a

multiple Grubbs-Beck test (EMA-MGB) has been proposed as an alternative method to the traditional B17B-standard Grubbs-Beck (B17B-GB) test moment estimation methods (Cohn and others, 1997, 2001; England and others, 2003). As with B17B-GB, EMA-MGB assumes that the log-Pearson Type 3 distribution represents the probability distribution function of annual maximum peak flows, except when historical, low-outlier, or censored information exists (Cohn and others, 1997; Griffis and others, 2004). EMA–MGB permits the efficient use of interval and perception threshold data, which most accurately represent historical information, low outliers, and censored flood data (Cohn and others, 1997).

Purpose and Scope

The purpose of this report is to summarize the consequences of using Expected Moments Algorithm (EMA) with a multiple Grubbs-Beck (MGB) test instead of Bulletin 17B (B17B) with the standard Grubbs-Beck (GB) test to compute the flood frequency analysis (FFA) at streamgaging stations in Arizona. This report presents a comparison between the standard FFA method, B17B-GB low-outlier test, and EMA-MGB low-outlier test, a proposed moments-estimator alternative for the efficient use of low-outlier, historical, and censored data. To understand the differences and performance of each estimator method, FFA from 328 streamgaging stations within Arizona (fig. 1, table 1) are compared using (1) the relative percent difference (RPD) of the annual exceedance probabilities (AEPs), (2) a measure of how well the fitted distribution approximates the observed peak-flow data, (3) random resampling procedures, and (4) Monte Carlo simulations of known LP3 distributions to determine estimator efficiency and precision. Both station and weighted skew coefficients were used, and for each method a different regional skew coefficient was used to weight the station skew for the FFA. The 10-, 1-, and 0.2-percent AEPs were selected as a representative subset for the RPD comparison (all quantile and RPD statistics are available in appendixes 1 and 2). As part of the comparison, streamgaging stations were stratified in two ways; (1) geographically by six flood regions as defined in Water Supply Paper 2433 (Thomas and others, 1997; fig. 1) and (2) by peak-flow data types, consisting of four categories determined by the presence or absence of U.S. Geological Survey (USGS) National Water Information System (NWIS) qualification codes and potentially influential low flows (PILFs)—(1) stations that have no PILFs as identified by the MGB test, no historical data (NWIS code 7), and no other NWIS qualification codes (1, 4, and 8); (2) stations that exclusively contain data with a historical code 7 or additional historical information specifying a “highest flood since” a water year but with no magnitude is specified for the historical flood water year; (3) stations that exclusively have PILFs identified by the MGB test and no other NWIS code information; and (4) stations that have one or more NWIS qualification codes 1, 4, or 8 and contain no historical information or PILFs (table 1).

Background and Previous Studies

The Hydrologic Frequency Analysis Work Group (HFAWG) is a working group under the Subcommittee on Hydrology of the Advisory Committee on Water Information consisting of Federal agencies, academia, interest groups, and private citizens. HFAWG was formed in 1999 to recommend procedures to increase the effectiveness of the current guidelines for Hydrologic Frequency Analysis computations and to evaluate alternative procedures for the frequency analysis of floods. The transition from Bulletin 17B Grubbs-Beck (B17B-GB) methods to the Expected Moments Algorithm multiple Grubbs-Beck test (EMA-MGB) (or other possible alternatives) requires careful examination of how the methods differ, both from a theoretical viewpoint and as applied to streamgaging station records. Several studies and HFAWG investigations have addressed this issue both using Monte Carlo simulation of synthetic data, in which the true underlying distribution is known, and using measured flood-peak data, in which it is not. Other comparison methods include resampling procedures, relative percent difference statistics, and visual assessments.

Several studies have documented the limitations of B17B-GB and have proposed EMA-MGB as the logical replacement, primarily because (1) it is more efficient at using historical information; (2) it does as well as maximum likelihood estimators for estimating the moments when historical information is available; (3) it addresses low outliers better than the conditional probability adjustment (CPA) used by B17B; and (4) it provides essentially the same estimates as B17B-GB if no historical, low-outlier, or other censored flood data are present (Griffis and others, 2004; England and Cohn, 2008; Stedinger and Griffis, 2008).

England and others (2003) compared the performance of B17B-GB and EMA by testing the ability of each to utilize historical and paleoflood data. They showed that B17B-GB performed as well as EMA for estimating 1-percent AEP when the LP3 distribution had a positive coefficient of skew and the historical period was 200 years or less, although EMA performed much better than B17B when the skew was negative and generally did better in most of the other historical period simulations. A single-large-flood case demonstrated that on average B17B-GB had a sample mean square error (MSE) of 1.5 and 3.0 times greater than EMA for positive and negative skew coefficients, respectively. In the case of multiple large floods B17B-GB was not able to fully utilize the historical information when the historical period exceeded 200 years.

Griffis and others (2004) investigated how regional skew information and low outlier adjustments would affect the performance of seven Pearson type 3 (P3) parameter estimation methods that also included B17B-GB and EMA. They conducted several Monte Carlo simulations to compare estimator methods and the performance was measured by MSE and bias of the quantile estimates. Although differences in MSE among the estimation techniques were modest, the inclusion of regional information and proper treatment of low outliers significantly

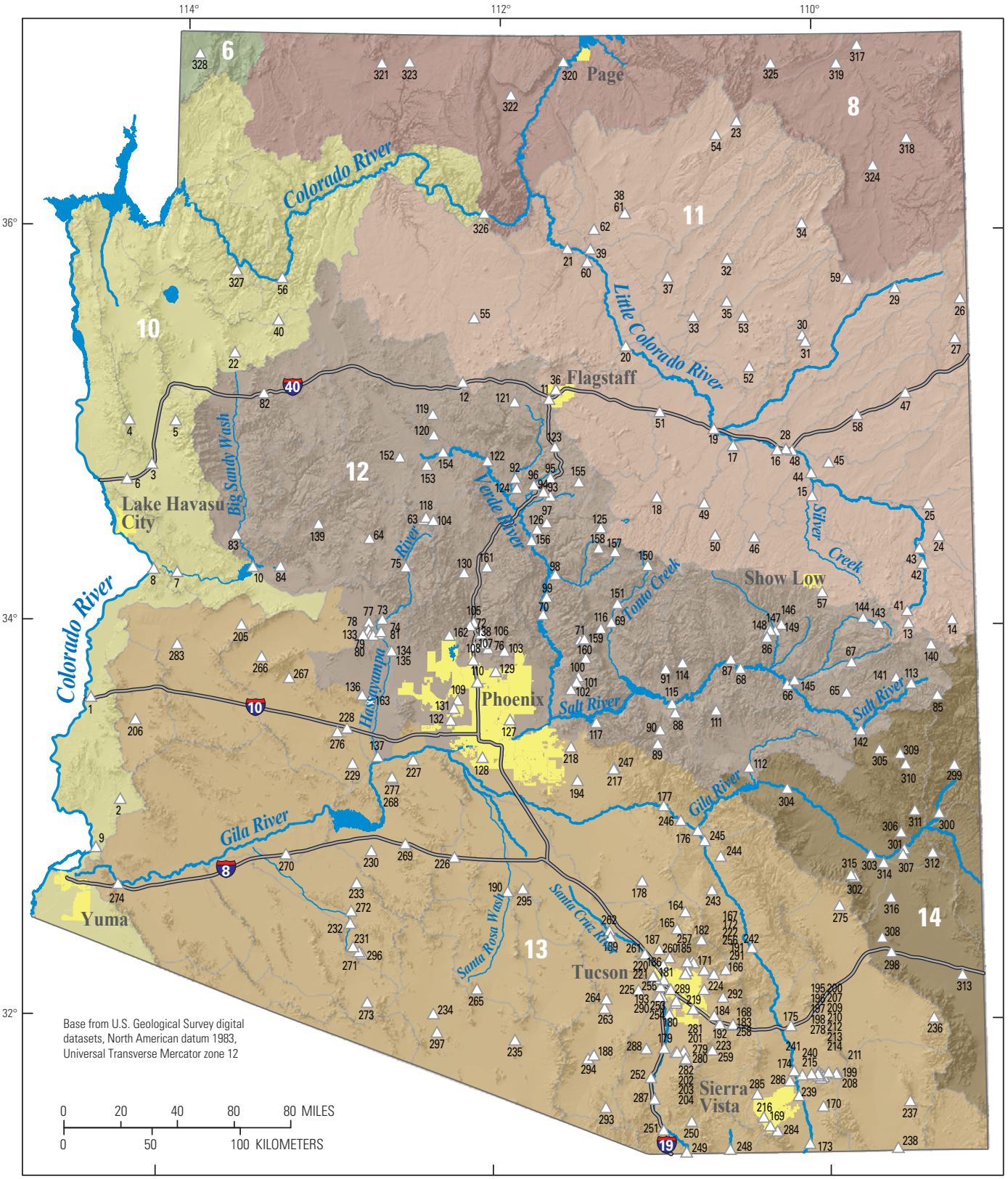


Figure 1. Map of Arizona showing locations of 328 streamflow-gaging stations (white triangles) used in the methods comparison analysis. Flood regions from Thomas and others (1997) are color coded and labeled by regression-equation numbers (white).

Table 1. Data for the 328 Arizona streamgaging stations used in the methods comparison analysis, with peak-flow data through water year 2010.

[Agencies: USGS, U.S. Geological Survey; Maricopa (Ma.), Maricopa County; Pima, Pima County; ARS, U.S. Department of Agriculture, Agricultural Research Service; NAD83,North American Datum, 1983; NWIS, National Water Information System; code 1, maximum mean daily discharge; code 4, less than a the reported discharge; code 8, greater than the reported discharge; code 7, historical discharge; NA, NWIS code category not applicable; record lengths given in years; WSP-2433, U.S. Geological Survey Water-Supply Paper 2433 (Thomas and others, 1997)]

Agency	Station		Decimal degrees, NAD83		Multiple Grubbs-Beck Test		Number of code				Record length		Number of		Code Group	WSP-2433 Region
	Identifier	Name	Latitude	Longitude	Threshold	No. of outlers	1	4	8	7	Systematic	Historical	Zero flows	Interval data		
USGS, Arizona	09429150	Creosote Wash nr Ehrenberg	33.620858	-114.49551	0	0	0	1	0	0	12	12	0	1	Code 1, 4, or 8 only	Region 10
	09429400	Indian Wash Trib nr Yuma	33.109205	-114.2955	0	0	0	0	0	1	15	18	0	3	Historical only	
	09423820	Sacramento Wash nr Yucca	34.811118	-114.1619	520	2	0	0	0	0	12	12	2	0	Low-outliers only (MGB test)	
	09423760	Little Meadow Creek nr Oatman	35.030558	-114.30912	20	5	0	1	0	0	12	12	3	1	NA	
	09423780	Walnut Creek nr Kingman	35.033338	-114.01884	152	3	0	1	0	0	12	12	1	1		
	09423900	Sacramento Wash Trib nr Topock	34.72973	-114.31329	10	3	0	1	0	0	14	14	2	1		
	09426500	Bill Williams River at Planet	34.266683	-113.98411	25,000	10	1	0	0	1	22	85	0	64		
	09427700	Monkeys Head Wash nr Parker	34.277793	-114.13022	5	4	0	1	0	0	14	14	3	1		
	09429510	Mittry Lake Trib nr Yuma	32.859766	-114.4355	22	5	0	1	0	0	12	12	2	1		
	09425500	Santa Maria River nr Alamo	34.30002	-113.51743	0	0	0	0	0	0	28	28	0	0	No low-outliers historical or other codes	
	09400650	Sinclair Wash At Flagstaff	35.163901	-111.68072	0	0	0	1	0	0	11	36	0	26	Code 1, 4, or 8 only	Region 11
	09403930	West Cataract Creek nr Williams	35.24779	-112.22517	0	0	0	1	0	0	13	13	0	1		
	09383400	Little Colorado River At Greer	34.016714	-109.45731	0	0	0	0	0	1	26	50	0	24	Historical only	
	09383500	Nutrioso Cr. above Nelson Res. nr Springerville	34.030327	-109.18647	0	0	0	0	0	1	20	44	0	24		
	09393500	Silver Creek nr Snowflake	34.6667	-110.04234	0	0	0	0	0	1	66	76	0	10		
	09397300	Little Colorado River nr Joseph City	34.901137	-110.2554	0	0	0	0	0	1	38	39	0	1		
	09398000	Chevelon Creek nr Winslow	34.926413	-110.53152	0	0	0	0	0	1	50	91	0	41		
	09398500	Clear Creek below Willow Creek, near Winslow	34.667526	-111.00763	0	0	0	0	0	1	44	54	0	10		
	09400350	Little Colorado River nr Winslow	35.011688	-110.65124	0	0	0	0	0	1	10	140	0	130		
	09401000	Little Colo. River at Grand Falls	35.433339	-111.2007	0	0	0	0	0	1	43	140	0	97		
	09402000	Little Colorado River nr Cameron	35.926382	-111.56737	0	0	0	0	0	1	72	140	0	74		
	09404343	Truxton Wash nr Valentine	35.384162	-113.65772	0	0	0	0	0	1	18	112	0	94		
	09379100	Long House Wash nr Kayenta	36.567221	-110.48875	480	7	0	0	0	0	15	15	1	0	Low-outliers only (MGB test)	
	09385800	Little Colorado R Trib nr St Johns	34.451152	-109.25704	16	1	0	0	0	0	14	14	1	0		
	09386250	Carrizo Wash nr St. Johns	34.614759	-109.31843	116	1	0	0	0	0	12	12	0	0		
	09395850	Black Creek Tributary nr Window Rock	35.654185	-109.08954	102	3	0	0	0	0	14	14	0	0		
	09395900	Black Creek nr Lupton	35.452522	-109.12648	830	1	0	0	0	0	19	19	0	0		
	09397100	Leroux Wash nr Holbrook	34.905026	-110.20151	3,730	3	0	0	0	0	14	14	0	0		
	09400100	Ganado Wash Trib nr Ganado	35.711125	-109.49788	150	5	0	0	0	0	14	14	1	0		
	09400290	Teshbito Wash Trib nr Holbrook	35.480567	-110.08818	600	6	0	0	0	0	15	17	0	3		
	09400300	Teshbito Wash nr Holbrook	35.448623	-110.06873	374	3	0	0	0	0	14	14	0	0		
	09400560	Oraibi Wash Trib nr Oraibi	35.872226	-110.55625	65	3	0	0	0	0	14	14	0	0		
	09400562	Oraibi Wash nr Tolani Lake	35.579729	-110.77403	219	1	0	0	0	0	15	15	0	0		
	09400565	Polacca Wash Trib nr Chinle	36.047225	-110.08123	180	3	0	0	0	0	13	13	0	0		
	09400568	Polacca Wash nr Second Mesa	35.65584	-110.56208	416	3	0	0	0	0	16	16	0	0		
	09400680	Switzer Canyon at Flagstaff	35.212234	-111.63988	51	5	0	0	0	0	12	12	0	0		
	09401110	Dinnebito Wash nr Sand Springs	35.781113	-110.9332	760	4	0	0	0	0	17	17	0	0		
	09401260	Moenkopi Wash at Moenkopi	36.104993	-111.20181	2,090	10	0	0	0	0	37	37	0	0		
	09401500	Moenkopi Wash nr Cameron	35.924994	-111.42153	2,470	3	0	0	0	0	13	13	0	0		
	09404310	Yampai Canyon Trib nr Peach Springs	35.551936	-113.38882	5	3	0	0	0	0	13	13	3	0		
	09383600	Fish Creek nr Eagar Ariz.	34.076435	-109.46315	45	4	0	1	0	0	13	13	1	1	NA	
	09384000	Little Colorado R above Lyman Lake nr St. Johns	34.314486	-109.36232	120	1	0	0	0	1	71	110	0	39		
	09384200	Lyman Reservoir Trib nr St Johns	34.391706	-109.38065	33	5	0	1	0	0	14	14	1	1		

Table 1. Data for the 328 Arizona streamgaging stations used in the methods comparison analysis, with peak-flow data through water year 2010.—Continued

Agency	Station		Decimal degrees, NAD83		Multiple Grubbs-Beck Test		Number of code				Record length		Number of		Code Group	WSP-2433 Region
	Identifier	Name	Latitude	Longitude	Thresh-old	No. of outlers	1	4	8	7	Sys-tem-atic	His-torical	Zero flows	Inter-val data		
USGS, Arizona	09394500	Little Colorado River at Woodruff	34.782808	-110.04428	0	0	1	0	0	1	88	105	0	20	NA	Region 11
	09395100	Carr Lake Draw Trib. nr Holbrook	34.834751	-109.934	35	6	0	1	0	0	13	13	3	1		
	09395200	Decker Wash nr Snowflake	34.461148	-110.40484	60	5	0	1	0	0	14	14	1	4		
	09396100	Puerco River nr Chambers	35.182246	-109.44705	1,550	3	0	0	0	1	39	40	0	1		
	09397000	Little Colorado River at Holbrook	34.897804	-110.16318	0	0	1	0	0	1	39	140	0	108		
	09397500	Chevelon Fork below Wildcat Canyon, nr Winslow	34.63642	-110.7143	170	3	0	0	0	1	53	81	3	28		
	09397800	Brookbank Canyon nr Heber	34.472258	-110.6479	78	4	0	1	0	0	13	13	0	1		
	09400530	Cow Canyon nr Winslow	35.100015	-110.9882	55	6	0	2	0	0	15	15	2	2		
	09400580	Castle Butte Wash nr Winslow	35.325013	-110.42291	58	5	0	1	0	0	13	13	2	1		
	09400583	Jeddito Wash Near Jeddito	35.577509	-110.46235	266	2	2	0	0	0	12	12	0	2		
	09401245	Klethla Valley Trib nr Kayenta	36.498053	-110.62153	80	4	0	1	0	0	15	15	0	1		
	09404050	Spring Valley Wash Trib nr Williams	35.574444	-112.15405	20	6	0	3	0	0	14	14	2	3		
	09404208	Diamond Creek Near Peach Springs	35.764988	-113.36827	361	1	1	0	0	0	17	17	0	1		
	09390500	Show Low Creek nr Lakeside	34.179488	-109.98789	0	0	0	0	0	0	57	57	0	0	No low-outliers historical or other codes	
	09396400	Dead Wash Tributary nr Holbrook	35.075023	-109.75067	0	0	0	0	0	0	13	13	0	0		
	09400200	Steamboat Wash Trib nr Ganado	35.763899	-109.80067	0	0	0	0	0	0	14	14	0	1		
	09401220	Cedar Wash nr Cameron	35.858607	-111.44292	0	0	0	0	0	0	10	10	0	0		
	09401280	Moenkopi Wash nr Tuba	36.104993	-111.20181	0	0	0	0	0	0	15	15	0	0		
	09401400	Moenkopi Wash Nr Tuba City	36.023604	-111.39736	0	0	0	0	0	0	27	38	0	11		
	09502960	Granite Creek At Prescott	34.551969	-112.46239	0	0	1	0	0	0	20	78	0	63	Code 1, 4, or 8 Only	Region 12
	09424480	Ash Creek Near Kirkland	34.453358	-112.79657	0	0	0	0	0	1	17	20	0	3	Historical only	
	09489700	Big Bonito C Nr Fort Apache	33.667274	-109.84676	0	0	0	0	0	1	23	29	0	6		
	09490500	Black River Near Fort Apache	33.712829	-110.21177	0	0	0	0	0	1	53	104	0	48		
	09492400	East Fork White River nr Fort Apache	33.822272	-109.81454	0	0	0	0	0	1	53	65	0	12		
	09497500	Salt River nr Chrysotile	33.798106	-110.49983	0	0	0	0	0	1	87	104	0	17		
	09498870	Rye Creek nr Gisela	34.033374	-111.29236	0	0	0	0	0	1	21	33	0	12		
	09508500	Verde R below Tangle Creek, aboveHorseshoe Dam	34.073092	-111.71626	0	0	0	0	0	1	92	1,011	0	919		
	09510080	West Fork Sycamore Cr. nr Sunflower	33.945874	-111.48541	0	0	0	0	0	1	15	20	0	5		
	09512800	Agua Fria River Near Rock Springs	34.015589	-112.16794	0	0	0	0	0	1	42	119	0	77		
	09515500	Hassayampa River At Box Damsite Nr Wickenburg	34.045029	-112.7099	0	0	0	0	0	1	48	92	0	44		
Maricopa	5228	Hassayampa River at US60	33.970306	-112.72703	611	3	0	0	0	0	17	17	3	0	Low-outliers only (MGB test)	
	5352	Hassayampa River at Wagoner	34.310139	-112.56867	95	3	0	0	0	0	19	19	0	0		
	5583	Cline Creek	33.901	-112.055	3	1	0	0	0	0	9	9	1	0		
	7013	Martinez Creek	34.029111	-112.79103	116	1	0	0	0	0	16	16	1	0		
	7043	Sols Wash nr Matthie	33.987417	-112.79297	119	2	0	0	0	0	16	16	2	0		
	7083	Flying E Wash	33.96225	-112.78289	51	2	0	0	0	0	17	17	1	0		
	7093	Casandro Wash	33.962056	-112.76525	14	1	0	0	0	0	17	17	1	0		
	7113	Powder House Wash	33.980833	-112.71731	195	7	0	0	0	0	16	16	2	0		
USGS, Arizona	09424200	Cottonwood Wash No. 1 nr Kingman	35.181116	-113.46966	2,620	4	0	0	0	0	15	15	0	0		
	09424450	Big Sandy River nr Wikieup	34.462517	-113.62438	4,950	15	0	0	0	0	45	45	0	0		
	09424900	Santa Maria River nr Bagdad.	34.305854	-113.34714	4,200	20	0	0	0	0	41	44	2	3		
	09489080	Hannagan Creek nr Hannagan Meadow	33.647276	-109.28952	10	2	0	0	0	0	13	13	0	0		
	09496800	Carrizo Creek Trib. nr Show Low	33.954493	-110.33205	100	1	0	0	0	0	14	14	1	0		
	09497800	Cibecue Creek nr Chrysotile	33.843105	-110.55761	1,200	4	0	0	0	0	52	52	0	0		
	09498400	Pinal Cr at Inspiration Dam, nr Globe	33.573107	-110.90123	406	2	0	0	0	0	31	31	0	0		
	09498501	Pinto Creek Blw Haunted Canyon nr Miami, Az	33.418665	-111.00956	203	2	0	0	0	0	15	15	0	0		
	09498502	Pinto Creek Near Miami	33.48783	-110.99539	540	2	0	0	0	0	16	16	0	0		
	09498503	South Fork Parker Creek nr Roosevelt	33.79727	-110.9604	1.4	3	0	0	0	0	23	24	0	1		
	09504500	Oak Creek Near Cornville, Az	34.764464	-111.89099	2,000	18	0	0	0	0	71	126	0	57		

Table 1. Data for the 328 Arizona streamgaging stations used in the methods comparison analysis, with peak-flow data through water year 2010.—Continued

Agency	Station		Decimal degrees, NAD83		Multiple Grubbs-Beck Test		Number of code				Record length		Number of		Code Group	WSP-2433 Region
	Identifier	Name	Latitude	Longitude	Thresh-old	No. of outlers	1	4	8	7	Sys-tem-atic	His-torical	Zero flows	Inter-val data		
USGS, Arizona	09505200	Wet Beaver Creek nr Rimrock	34.674744	-111.67209	2,740	22	0	0	0	0	49	49	0	0	Low-outliers only (MGB test)	Region 12
	09505250	Red Tank Draw nr Rimrock	34.695299	-111.71432	327	5	0	0	0	0	21	21	0	0		
	09505300	Rattlesnake Canyon nr Rimrock	34.766964	-111.67376	288	4	0	0	0	0	23	23	0	0		
	09505350	Dry Beaver Creek nr Rimrock	34.728631	-111.77571	2,510	20	0	0	0	0	50	50	0	0		
	09505800	West Clear Creek nr Camp Verde	34.538636	-111.69404	3,630	23	0	0	0	0	46	46	0	0		
	09507980	East Verde River nr Childs	34.276421	-111.63876	4,100	24	0	0	0	0	49	50	0	1		
	09508300	Wet Bottom Creek nr Childs	34.160868	-111.69292	1,260	21	0	0	0	0	43	43	0	0		
	09510170	Camp Creek nr Sunflower	33.759766	-111.49625	26	2	0	0	0	0	36	48	2	12		
	09510180	Rock Creek nr Sunflower	33.730321	-111.50847	916	13	0	0	0	0	30	48	1	18		
	09510200	Sycamore Creek nr Fort Mcdowell	33.69421	-111.5418	668	8	0	0	0	0	51	51	0	0		
	09512280	Cave Creek blw Cottonwood Cr nr Cave Creek	33.88726	-111.95404	73	2	0	0	0	0	31	31	0	0		
	09512420	Lynx Creek Trib nr Prescott	34.547525	-112.40017	130	3	0	0	0	0	10	10	1	0		
	09512700	Agua Fria R Trib #2 nr Rock Springs	34.033367	-112.14571	140	4	0	0	0	0	38	48	0	10		
	09513780	New River nr Rock Springs	33.974202	-112.09905	424	8	0	0	0	0	49	49	0	0		
	09513800	New River At New River	33.911424	-112.14127	805	4	0	0	0	0	22	22	0	0		
	09513820	Deadman Wash nr New River	33.841703	-112.14516	489	15	0	0	0	0	30	41	4	11		
	09513835	New River At Bell Road, nr Peoria	33.638372	-112.24016	876	6	0	0	0	0	24	41	3	21		
	09513860	Skunk Creek nr Phoenix	33.729205	-112.11988	175	7	0	0	0	0	42	42	3	0		
	09468300	Sevenmile Wash Trib nr Globe	33.586164	-110.65066	4	3	0	0	0	1	17	47	2	30	NA	
	09468500	San Carlos River Near Peridot	33.296447	-110.45149	565	1	0	0	0	1	82	95	0	13		
	09489100	Black River Near Maverick	33.707552	-109.44731	0	0	1	0	0	1	21	30	0	10		
	09497980	Cherry Creek Near Globe	33.827826	-110.85623	125	1	0	0	0	1	45	50	0	5		
	09498500	Salt River Near Roosevelt	33.619495	-110.9215	0	0	9	0	0	2	98	623	0	534		
	09498900	Gold Creek Near Payson	34.002818	-111.35902	50	2	0	1	0	1	15	17	0	3		
	09501300	Tortilla Creek At Tortilla Flat	33.52727	-111.38763	965	18	0	0	0	1	38	69	1	32		
	09503000	Granite Creek Near Prescott	34.56308	-112.44489	230	1	0	0	0	1	33	78	0	45		
	09503740	Hell Canyon Trib N Ash Fork	35.083905	-112.40851	4	2	0	0	0	1	10	12	2	2		
	09503750	Limestone Canyon Nr Paulden	34.980018	-112.40212	70	2	0	0	0	1	11	12	0	1		
	09503800	Volunteer Wash Near Bellemont	35.150568	-111.89905	447	6	0	0	0	1	14	15	2	1		
	09504000	Verde River Near Clarkdale	34.852242	-112.06599	0	0	1	0	0	1	49	104	0	56		
	09504400	Munds Canyon Trib nr Sedona	34.922241	-111.64515	90	7	0	1	0	1	16	17	0	2		
	09504800	Oak Creek Trib Near Cornville	34.712521	-111.88127	1	2	0	4	0	1	15	18	1	7		
	09505600	Dirty Neck Canyon nr Clints Well	34.512526	-111.35903	18	2	0	1	0	1	12	15	1	4		
	09505900	Cottonwood Wash nr Camp Verde	34.505581	-111.75348	0	0	0	1	0	1	14	16	0	3		
	09512100	Indian Bend Wash at Scottsdale	33.538654	-111.91653	76	2	0	0	0	2	11	48	2	37		
	09512200	Salt River Trib in South Mntn Park nr Phoenix	33.34699	-112.08487	7	23	3	0	0	0	50	50	19	3		
	09512300	Cave Creek Nr Cave Creek	33.783372	-112.00737	610	9	0	0	0	1	35	37	4	2		
	09512600	Turkey Creek Near Cleator	34.282252	-112.20766	530	1	0	0	0	1	14	23	0	9		
	09513890	New River At Peoria	33.595317	-112.26321	1,390	4	0	0	0	1	14	29	2	16		
	09513910	New River nr Glendale	33.536707	-112.28182	490	5	0	0	0	1	23	42	4	20		
	09515800	Hartman Wash nr Wickenburg	33.962807	-112.82851	140	11	0	0	0	1	36	47	4	11		
	09516500	Hassayampa R. nr Morristown	33.885032	-112.66212	594	5	0	1	0	1	57	95	1	42		
	09516600	Ox Wash nr Morristown	33.883365	-112.65073	150	9	0	0	0	1	37	51	4	14		
	09516800	Jack Rabbit Wash nr Tonopah	33.658924	-112.82851	612	14	0	0	0	1	37	47	0	10		
	09517000	Hassayampa River nr Arlington	33.347264	-112.72573	1,230	13	0	0	0	1	49	95	1	48		
Ma.	5588	Skunk Creek at New River	33.926139	-112.08267	0	0	0	0	0	0	16	16	0	0	No low-outliers historical or other codes	
USGS, Arizona	09424700	Iron Spring Wash Trib nr Bagdad	34.522243	-113.11269	0	0	0	0	0	0	15	16	0	1		
	09489070	North Fork Of East Fork Black R nr Alpine	33.903106	-109.32286	0	0	0	0	0	0	13	13	0	0		
	09489200	Pacheta Creek At Maverick	33.739773	-109.54064	0	0	0	0	0	0	23	23	0	0		
	09489500	Black River Blw Pumping Plant, nr Point of Pines	33.47672	-109.76398	0	0	0	0	0	0	57	57	0	0		
	09490800	North Fork White River nr Greer	34.013936	-109.64232	0	0	0	0	0	0	13	13	0	0		
	09491000	North Fork White R nr Mcnary	34.045879	-109.73816	0	0	0	0	0	0	39	40	0	1		
	09494000	White River nr Fort Apache	33.736441	-110.16677	0	0	0	0	0	0	53	53	0	0		
	09496000	Corduroy C nr Mouth nr Show Low	34.01838	-110.24233	0	0	0	0	0	0	26	54	0	28		
	09496500	Carrizo Creek nr Show Low	33.985881	-110.28094	0	0	0	0	0	0	60	60	0	0		

Table 1. Data for the 328 Arizona streamgaging stations used in the methods comparison analysis, with peak-flow data through water year 2010.—Continued

Agency	Station		Decimal degrees, NAD83		Multiple Grubbs-Beck Test		Number of code				Record length		Number of		Code Group	WSP-2433 Region
	Identifier	Name	Latitude	Longitude	Thresh-old	No. of outliers	1	4	8	7	Sys-tem-atic	His-torical	Zero flows	Inter-val data		
USGS, Arizona	09496600	Cibecue 1, Trib Carrizo Cr nr Show Low	33.991159	-110.32483	0	0	0	0	0	0	14	14	0	0	No low-outliers historical or other codes	Region 12
	09496700	Cibecue 2, Trib Carrizo Cr, nr Show Low	33.988103	-110.31122	0	0	0	0	0	0	14	14	0	0		
	09498600	Cristopher Creek Trib nr Kohl's Ranch	34.322258	-111.06735	0	0	0	0	0	0	11	11	0	0		
	09498800	Tonto Creek Near Gisela	34.128927	-111.25541	0	0	0	0	0	0	11	11	0	0		
	09502800	Williamson Valley Wash nr Paulden	34.866687	-112.61323	0	0	0	0	0	0	31	46	0	15		
	09502900	Del Rio Springs nr Chino Valley	34.825577	-112.44461	0	0	0	0	0	0	14	14	0	0		
	09503700	Verde River nr Paulden	34.89502	-112.34295	0	0	0	0	0	0	48	48	0	0		
	09505220	Rocky Gulch nr Rimrock	34.746966	-111.49459	0	0	0	0	0	0	30	33	0	3		
	09506000	Verde River nr Camp Verde	34.448361	-111.78987	0	0	0	0	0	0	41	77	0	43		
	09507600	East Verde River nr Pine	34.391697	-111.26875	0	0	0	0	0	0	13	13	0	0		
	09507700	Webber Cr above West Fork Webber Cr nr Pine	34.41114	-111.37292	0	0	0	0	0	0	16	16	0	0		
	09510100	East Fork Sycamore Creek nr Sun-flower	33.949485	-111.46152	0	0	0	0	0	0	26	26	0	0		
	09510150	Sycamore Creek nr Sunflower	33.851431	-111.45319	0	0	0	0	0	0	15	51	0	36		
	09512500	Agua Fria River nr Mayer	34.315307	-112.06405	0	0	0	0	0	0	71	71	0	0		
	09512860	Humbug Creek, nr Castle Hot Springs	33.967256	-112.2935	0	0	0	0	0	0	11	11	0	0		
	09516790	Star Wash nr Tonopah	33.633056	-112.77889	0	0	0	0	0	0	9	9	0	0		
Pima	1080	Canada Del Oro Wash Northeast of Saddlebrooke	32.564214	-110.84783	0	0	0	1	0	0	21	22	0	2	Code 1, 4, or 8 only	Region 13
	1100	Canada Del Oro Wash at Golder Ranch Road	32.478076	-110.89885	0	0	0	3	0	0	16	18	0	5		
	2070	Tanque Verde Wash 0.5 mi South of Chiva Tank	32.267897	-110.60698	0	0	0	2	0	0	16	19	0	5		
	2170	Ventana Canyon Wash at Sunrise Rd	32.308747	-110.83898	0	0	0	2	0	0	15	15	0	2		
	4280	Cienega Creek at Interstate 10	31.985961	-110.56798	0	0	0	2	0	0	21	23	0	4		
USGS, Arizona	09470800	Garden Canyon nr Fort Huachuca	31.472875	-110.34786	0	0	0	0	1	0	23	51	0	29		
	09470900	San Pedro River Trib nr Bisbee	31.570095	-110.0273	0	0	0	1	0	0	16	17	0	2		
	09483200	Agua Caliente Wash Trib nr Tucson	32.268687	-110.73814	0	0	0	1	0	0	16	16	0	1		
	09484000	Sabino Creek nr Tucson	32.316742	-110.81037	0	0	0	1	0	0	79	79	0	1		
	09470500	San Pedro River at Palominas	31.380101	-110.11119	0	0	0	0	0	1	73	104	0	32	Historical only	
	09471550	San Pedro River nr Tombstone	31.750922	-110.20119	0	0	0	0	0	1	38	95	0	61		
	09471700	Fenner Wash Near Benson	31.980361	-110.21646	0	0	0	0	0	1	16	20	0	4		
	09473500	San Pedro R at Winkelman	32.977288	-110.77038	0	0	0	0	0	1	27	104	0	77		
	09474000	Gila River at Kelvin	33.10284	-110.9765	0	0	0	0	0	1	19	62	0	43		
	09478200	Durham Wash nr Florence	32.722293	-111.109	0	0	0	0	0	1	19	27	0	8		
	09482000	Santa Cruz River at Continental	31.871473	-110.98009	0	0	0	0	0	1	67	118	0	51		
	09482200	Flato Wash nr Sahuarita	32.045358	-110.95065	0	0	0	0	0	1	19	56	0	37		
	09482500	Santa Cruz River at Tucson	32.221187	-110.98176	0	0	0	0	0	1	94	118	0	24		
	09483300	Sabino C nr Mt Lemmon	32.422298	-110.75204	0	0	0	0	0	1	11	20	0	10		
	09484560	Cienega Creek nr Pantano	31.985638	-110.56647	0	0	0	0	0	1	15	24	0	9		
	09484600	Pantano Wash nr Vail	32.035914	-110.67758	0	0	0	0	0	1	53	80	0	27		
	09485500	Pantano Wash nr Tucson	32.250076	-110.85064	0	0	0	0	0	1	18	54	0	36		
	09486000	Rillito Cr nr Tucson	32.294519	-110.98537	0	0	0	0	0	1	68	70	0	2		
	09486300	Canada Del Oro nr Tucson	32.37424	-111.00927	0	0	0	0	0	1	20	34	0	14		
	09486800	Altar Wash nr Three Points	31.838972	-111.40427	0	0	0	0	0	1	33	70	0	37		
	09487250	Los Robles Wash nr Marana	32.437849	-111.30427	0	0	0	0	0	1	15	99	0	84		
	09488500	Santa Rosa Wash nr. Vaiva Vo	32.667557	-111.92819	0	0	0	0	0	1	19	88	0	69		
Pima	2090	Tanque Verde Wash at Tanque Verde Guest Ranch	32.245796	-110.68277	1,422	9	0	0	0	0	19	19	0	0	Low-outliers Only (MGB test)	
	4310	Davidson Canyon Wash 0.25 miles south of Interstate 10	31.993577	-110.64517	1,420	5	0	0	0	0	22	23	0	1		
	6040	Santa Cruz River at Valencia Road	32.133055	-110.99309	610	1	0	0	0	0	17	18	0	1		
Ma.	6723	Queen Creek at Cap	33.232167	-111.50314	134	1	0	0	0	0	12	12	1	0		

Table 1. Data for the 328 Arizona streamgaging stations used in the methods comparison analysis, with peak-flow data through water year 2010.—Continued

Agency	Station		Decimal degrees, NAD83		Multiple Grubbs-Beck Test		Number of code				Record length		Number of		Code Group	WSP-2433 Region
	Identifier	Name	Latitude	Longitude	Thresh-old	No. of outlers	1	4	8	7	Sys-tem-atic	His-torical	Zero flows	Inter-val data		
ARS	67.102	Walnut Gulch Experimental Watershed USDA/ARS	31.742196	-110.052582	6.1	18	0	0	0	0	45	46	0	1	Low-outliers Only (MGB test)	Region 13
	67.104		31.741229	-110.052413	13.1	19	0	0	0	0	45	46	0	1		
	67.105		31.743467	-110.054881	0.3	1	0	0	0	0	41	46	0	5		
	67.106		31.742033	-110.053682	1.5	17	0	0	0	0	45	46	0	1		
	67.112		31.737136	-109.943714	1.6	10	0	0	0	0	42	46	0	4		
	67.125		31.724913	-110.052927	4.9	9	0	0	0	0	30	30	0	0		
	76.004	Santa Rita Experimental Range USDA/ARS	31.851268	-110.9045	2	1	0	0	0	0	36	36	0	0		
	76.006		31.813418	-110.854627	0	1	0	0	0	0	35	35	1	0		
	76.007		31.816447	-110.852773	0.2	1	0	0	0	0	35	35	0	0		
	76.008		31.816925	-110.852688	0.7	4	0	0	0	0	35	35	0	0		
USGS	09428545	Cunningham Wash Trib nr Wenden	34.006971	-113.57854	12	2	0	0	0	0	13	13	0	0		
	09428800	Tyson Wash Trib nr Quartzsite	33.512529	-114.21744	55	1	0	0	0	0	14	14	0	0		
ARS	09471080	Walnut Gulch 63.010 nr Tombstone	31.720368	-110.02563	120	4	0	0	0	0	43	44	0	1		
	09471087	Walnut Gulch 63.111 nr Tombstone	31.734535	-109.94841	42.3	3	0	0	0	0	20	20	1	0		
	09471090	Walnut Gulch 63.009 nr Tombstone	31.717868	-110.02508	81	1	0	0	0	0	44	44	0	0		
	09471110	Walnut Gulch 63.015 nr Tombstone	31.712868	-110.04091	165	14	0	0	0	0	54	56	0	2		
	09471120	Walnut Gulch 63.011 nr Tombstone	31.741201	-109.99508	132	12	0	0	0	0	47	48	0	1		
	09471130	Walnut Gulch 63.008 nr Tombstone	31.723146	-110.0448	134	1	0	0	0	0	27	27	0	0		
	09471140	Walnut Gulch 63.006 nr Tombstone	31.724535	-110.05535	200	2	0	0	0	0	48	49	0	1		
	09471180	Walnut Gulch 63.003 nr Tombstone	31.73259	-110.05758	22	4	0	0	0	0	56	57	0	1		
	09471195	Walnut Gulch 63.007 nr Tombstone	31.732868	-110.09813	104	13	0	0	0	0	44	45	0	1		
USGS, Arizona	09471310	Huachuca Canyon nr Fort Huachuca	31.518056	-110.38722	25	3	0	0	0	0	10	10	0	0		
	09478600	Queen Creek Trib 3 at Whitlow Dam	33.291721	-111.28124	35	3	0	0	0	0	14	14	1	0		
	09479200	Queen C Trib A Apache Junc	33.40366	-111.54152	19	5	0	0	0	0	19	19	4	0		
	09482330	Pumping Wash nr Vail	32.069524	-110.80703	30	1	0	0	0	0	16	16	0	0		
	09483030	Anklam Wash At Tucson	32.225076	-111.03121	2	1	0	0	0	0	17	17	1	0		
	09483040	West Speedway Wash nr Tucson	32.238965	-111.04593	104	8	0	0	0	0	17	17	0	0		
	09484510	Ventana Canyon Wash nr Tucson	32.309798	-110.83953	86	2	0	0	0	0	17	17	0	0		
	09484590	Davidson Canyon Wash nr Vail	31.993693	-110.64508	587	2	0	0	0	0	14	14	0	0		
	09485100	Saguaro Corners Wash nr Tucson	32.1698	-110.73814	23	3	0	0	0	0	10	10	1	0		
	09487140	San Joaquin Wash nr Tucson	32.168688	-111.13343	60	1	0	0	0	0	13	13	1	0		
	09488650	Vekol Wash nr Stanfield	32.841716	-112.25181	136	1	0	0	0	0	21	21	1	0		
	09514200	Waterman Wash nr Buckeye, Ariz.	33.330321	-112.50988	325	9	0	0	0	0	43	47	6	4		
	09517400	Winters Wash nr Tonopah, Ariz.	33.489483	-112.91879	390	7	0	0	0	0	30	49	5	19		
	09517490	Centennial Wash at Southern Pacific Railroad Bridge	33.31032	-112.88184	28	1	0	0	0	0	25	30	1	5		
	09519760	Sauceda Wash nr Gila Bend	32.870603	-112.75905	530	18	0	0	0	0	37	48	5	11		
	09520110	Hot Shot Arroyo nr Ajo,	32.347003	-112.80932	110	5	0	0	0	0	16	16	0	0		
	09520170	Rio Cornez nr Ajo,	32.499501	-112.88127	1,390	1	0	0	0	0	14	14	1	0		
	09520200	Black Gap Wash nr Ajo	32.70644	-112.846	280	5	0	0	0	0	18	18	3	0		
	09535100	San Simon Wash nr Pisinimo	32.044237	-112.37097	174	2	0	0	0	0	39	39	0	0		
	09535200	Sells Wash Trib at Sells	31.915355	-111.87901	1,500	4	0	0	0	0	15	15	0	0		
	09536350	Surprise Canyon nr Dos Cabezas	32.011198	-109.35395	30	5	0	0	0	0	14	14	3	0		
	09537200	Leslie Creek nr Mcneal	31.590096	-109.50896	250	11	0	0	0	0	37	41	0	4		
	09537500	Whitewater Draw nr Douglas	31.352325	-109.58507	1,370	42	0	0	0	0	84	95	0	11		
	09471000	San Pedro River at Charleston	31.625926	-110.17452	1,490	3	5	0	0	1	100	107	0	12	NA	
ARS	09471190	Walnut Gulch 63.002 nr Tombstone Az: USDA/SEA/AR	31.734812	-110.09841	114	1	0	0	0	1	56	120	0	64		
	09471200	Walnut Gulch 63.001 nr Tombstone Az: USDA/SEA/AR	31.729257	-110.15341	128	3	0	0	0	1	54	120	0	66		
USGS, Arizona	09472000	San Pedro River nr Redington	32.380628	-110.44647	3,460	16	0	0	0	1	68	104	0	36		
	09472400	Mammoth Wash nr Mammoth	32.676458	-110.68538	0	0	0	2	0	1	15	21	0	8		
	09473000	Aravaipa Creek nr Mammoth	32.844233	-110.6301	0	0	1	0	0	1	61	92	0	32		
	09473200	Green Lantern Wash nr Winkelman	32.925066	-110.72705	100	1	0	0	0	1	14	18	0	4		
	09473600	Tam O'shanter Wash nr Hayden	33.029509	-110.87344	180	2	0	0	0	1	15	19	1	4		
	09478500	Queen Creek blw Whitlow Dam nr Superior	33.299221	-111.27763	0	0	1	0	0	1	17	43	0	26		
	09480000	Santa Cruz River nr Lochiel	31.355378	-110.58953	380	8	0	0	0	1	62	84	0	22		
	09480500	Santa Cruz River nr Nogales	31.344544	-110.85147	0	0	4	0	0	1	86	118	0	37		

Table 1. Data for the 328 Arizona streamgaging stations used in the methods comparison analysis, with peak-flow data through water year 2010.—Continued

Agency	Station		Decimal degrees, NAD83		Multiple Grubbs-Beck Test		Number of code				Record length		Number of		Code Group	WSP-2433 Region
	Identifier	Name	Latitude	Longitude	Thresh-old	No. of outlers	1	4	8	7	Sys-tem-atic	His-torical	Zero flows	Inter-val data		
USGS, Arizona	09481500	Sonoita Creek nr Patagonia	31.499816	-110.81814	1,550	9	0	0	0	1	45	55	0	10	NA	Region 13
	09481700	Calabasas Canyon nr Nogales	31.457039	-110.98648	0	0	0	1	0	1	14	16	0	3		
	09481750	Sopori Wash at Amado	31.723699	-111.06176	1,500	7	0	0	0	1	20	31	0	11		
	09482350	South Fork Airport Wash nr Tucson	32.100079	-110.90898	59	3	0	0	0	1	15	24	1	9		
	09482370	North Fork Airport Wash nr Tucson	32.11119	-110.90898	30	2	0	0	0	1	17	24	1	7		
	09482480	Big Wash at Tucson	32.186188	-111.00259	20	5	0	0	0	1	17	61	4	44		
	09484200	Bear Creek nr Tucson	32.306187	-110.80148	192	3	0	0	0	1	16	20	0	4		
	09484500	Tanque Verde Creek at Tucson	32.265353	-110.8412	3,120	20	2	0	0	1	45	71	1	28		
	09484570	Mescal Arroyo nr Pantano	31.989804	-110.56508	130	1	0	0	0	1	18	51	0	33		
	09484580	Barrel Canyon nr Sonoita	31.861751	-110.69119	879	9	0	2	0	0	16	49	0	35		
	09485950	Geronimo Wash nr Tucson	32.332297	-110.94426	35	3	0	1	0	0	18	18	0	1		
	09486500	Santa Cruz River at Cortaro	32.351185	-111.09454	4,270	8	1	0	0	1	65	96	0	32		
	09486520	Santa Cruz R at Trico Road, nr Ana	32.471459	-111.30761	3,440	7	1	0	0	1	22	71	1	49		
	09487000	Brawley Wash Near Three Points	32.075634	-111.33872	1,150	2	0	0	0	1	37	71	0	34		
	09487100	Little Brawley Wash nr Three Points	32.123689	-111.32983	310	1	0	0	0	1	15	519	1	504		
	09487400	Quijotoa Wash Trib. nr Quijotoa	32.173681	-112.10902	127	5	0	2	0	0	13	13	2	2		
	09517200	Centennial Wash Trib nr Wenden	33.844475	-113.45076	90	16	0	0	0	1	36	48	3	12		
	09517280	Tiger Wash nr Aguila	33.7417	-113.27936	620	10	0	0	0	1	37	48	0	11		
	09519600	Rainbow Wash Trib nr Buckeye	33.2431	-112.63822	299	18	0	0	0	1	38	48	0	10		
	09519750	Bender Wash nr Gila Bend	32.90699	-112.5521	287	11	0	1	0	2	36	48	5	13		
	09520100	Military Wash nr Sentinel	32.845327	-113.27963	100	10	0	0	0	1	36	48	2	12		
	09520160	Gibson Arroyo at Ajo	32.380058	-112.86182	0	0	0	1	0	1	15	21	0	7		
	09520230	Crater Range Wash nr Ajo	32.562277	-112.87766	132	16	0	0	0	1	35	48	4	13		
	09520300	Alamo Wash Trib nr Ajo	32.100063	-112.77154	74	5	0	0	0	1	29	31	3	2		
	09520400	Ligurta Wash at Ligurta	32.675881	-114.29466	58	4	0	0	0	1	15	18	2	3		
	09536100	Pitchfork Canyon Trib nr Fort Grant	32.588958	-109.91174	120	5	0	1	0	0	14	14	2	1		
Ma.	5108	Delaney Wash	33.469806	-112.97714	0	0	0	0	0	0	11	11	0	0	No low-outliers historical or other codes	
	6953	Rainbow Wash	33.2356	-112.6392	0	0	0	0	0	0	10	10	0	0		
ARS	67121	Walnut Gulch Experimental Watershed 67.121 Usda/Ars	31.730217	-110.036016	0	0	0	0	0	0	38	39	0	1		
	76001	Santa Rita Experimental Range 76.001 Usda/Ars	31.857048	-110.862571	0	0	0	0	0	0	36	36	0	0		
	76002	Santa Rita Experimental Range 76.002 Usda/Ars	31.855001	-110.863293	0	0	0	0	0	0	36	36	0	0		
	76003	Santa Rita Experimental Range 76.003 Usda/Ars	31.853304	-110.913377	0	0	0	0	0	0	36	36	0	0		
	76005	Santa Rita Experimental Range 76.005 Usda/Ars	31.815409	-110.851866	0	0	0	0	0	0	36	36	0	0		
USGS, Arizona	09428550	Bouse Wash Trib nr Bouse	33.901413	-113.97439	0	0	0	0	0	0	14	14	0	0		
	09470750	Ramsey Canyon nr Sierra Vista	31.446667	-110.30583	0	0	0	0	0	0	11	11	0	0		
	09471380	Up. Babocomari R nr Huachuca City	31.635	-110.42472	0	0	0	0	0	0	11	11	0	0		
	09471400	Babocomari River nr Tombstone	31.700278	-110.22639	0	0	0	0	0	0	11	11	0	0		
	09481740	Santa Cruz River at Tubac	31.612868	-111.04148	0	0	0	0	0	0	15	15	0	0		
	09481800	Demetrie Wash Trib nr Continental	31.870917	-111.08815	0	0	0	0	0	0	14	14	0	0		
	09482420	Julian Wash At Tucson	32.170911	-110.94092	0	0	0	0	0	0	12	12	0	0		
	09482450	West Branch Santa Cruz R at Tucson	32.133411	-111.00898	0	0	0	0	0	0	16	16	0	0		
	09483100	Tanque Verde Creek nr Tucson	32.246742	-110.68008	0	0	0	0	0	0	26	26	0	0		
	09485000	Rincon Creek nr Tucson	32.129523	-110.62591	0	0	0	0	0	0	58	58	0	0		
	09486590	Arivaca Creek nr Arivaca	31.572312	-111.33232	0	0	0	0	0	0	10	10	0	0		
	09486700	Chiltepines Wash nr Sasabe	31.818973	-111.43844	0	0	0	0	0	0	13	13	0	0		
	09488600	Silver Reef Wash nr Casa Grande	32.682281	-111.83485	0	0	0	0	0	0	14	26	0	13		
	09520130	Darby Arroyo nr Ajo	32.355337	-112.82599	0	0	0	0	0	0	16	16	0	0		
	09535300	Vamori Wash at Kom Vo	31.951184	-112.34791	0	0	0	0	0	0	39	39	0	0		
	09456400	Gold Gulch Near Bowie, Ariz.	32.347852	-109.6034	0	0	0	1	0	0	14	14	0	1	Code 1, 4, or 8 only	Region 14
	09444200	Blue River Near Clifton	33.290895	-109.19618	0	0	0	0	0	1	44	125	0	81	Historical only	
	09444500	San Francisco River at Clifton	33.049508	-109.2959	0	0	0	0	0	1	104	140	0	36		
	09448500	Gila River at head of Safford Valley, nr Solomon,	32.868397	-109.51119	0	0	0	0	0	1	96	104	0	8		

Table 1. Data for the 328 Arizona streamgaging stations used in the methods comparison analysis, with peak-flow data through water year 2010.—Continued

Agency	Station		Decimal degrees, NAD83		Multiple Grubbs-Beck Test		Number of code				Record length		Number of		Code Group	WSP-2433 Region
	Identifier	Name	Latitude	Longitude	Thresh-old	No. of outlers	1	4	8	7	Sys-tem-atic	His-torical	Zero flows	Inter-val data		
USGS, Arizona	09458200	Deadman Creek nr Safford	32.733124	-109.81647	0	0	0	0	0	1	17	27	0	10	Historical Only	Region 14
	09458500	Gila River at Safford	32.847288	-109.71591	0	0	0	0	0	1	35	96	0	68		
	09466500	Gila River at Calva	33.185613	-110.22009	0	0	0	0	0	1	81	104	0	22		
	09445500	Willow C nr Point of Pines nr Morenci	33.379223	-109.65064	178	1	0	0	0	0	23	23	0	0	Low-outliers Only (MGB test)	
	09447800	Bonita Creek nr Morenci	32.955618	-109.53119	269	4	0	0	0	0	30	30	0	0		
	09451900	Agricul Resrch Serv Safford Wtrshed W-I Ariz	32.840898	-109.52202	93	14	0	0	0	0	31	31	0	0		
	09456680	Agricul Resrch Serv Safford Wtrshed W-V Ariz	32.422293	-109.65812	15	3	0	0	0	0	30	31	0	1		
	09446000	Willow C N Double Circle Rnch nr Morenci	33.354225	-109.52564	629	6	0	0	0	1	25	30	0	5	NA	
	09446500	Eagle C N Double Circle Rnch nr Morenci	33.30006	-109.4923	456	1	0	0	0	1	25	30	0	5		
	09447000	Eagle Creek Above Pumping Plant, nr Morenci	33.064506	-109.4423	1,550	24	0	0	0	1	69	95	0	26		
	09451800	Tollgate Wash Trib nr Clifton, Ariz.	32.850066	-109.33813	15	7	0	1	0	0	14	14	5	1		
	09456000	San Simon River nr San Simon Ariz.	32.22508	-109.17561	4,190	6	5	0	0	1	17	22	0	10		
	09457000	San Simon River nr Solomon, Ariz.	32.801733	-109.63924	1,520	2	0	0	0	1	53	104	0	51		
	09460150	Frye Creek nr Thatcher, Az.	32.743957	-109.83814	0	0	2	0	0	1	32	44	0	14		
	09456820	Agricul Resch Ser Safford Wtrshed W-Iv Ariz	32.625068	-109.60063	0	0	0	0	0	0	29	30	0	1	No low-outliers historical or other codes	
	09379200	Chinle Creek nr Mexican Water	36.943891	-109.71067	0	0	0	0	0	1	48	60	0	12	Historical Only	Region 8
	09379060	Lukachukai Cr Trib nr Lukachukai	36.469445	-109.40622	10	4	0	0	0	0	14	14	2	0	Low-outliers only (MGB test)	
	09379180	Laguna Creek at Dennehotso	36.853891	-109.84595	1,380	5	0	0	0	0	10	10	0	0		
	09382000	Paria River at Lees Ferry,	36.87221	-111.59461	998	9	0	0	0	0	87	87	0	0		
	09403800	Bitter Seeps Wash Trib nr Fredonia	36.85693	-112.75909	30	3	0	0	0	0	14	14	3	0		
	09383020	House Rock Wash Trib Nr Marble Canyon	36.701376	-111.92989	0	0	0	5	0	1	14	42	0	33	NA	
	09403780	Kanab Creek Nr Fredonia	36.863876	-112.57992	0	0	1	0	0	1	18	522	0	505		
	09379030	Black Mountain Wash nr Chinle	36.333335	-109.62428	0	0	0	0	0	0	15	15	0	0	No low-outliers historical or other codes	
	09379560	El Capitan Wash nr Kayenta	36.858889	-110.26597	0	0	0	0	0	0	14	14	0	0		
	09403000	Bright Angel Cr nr Grand Canyon	36.103038	-112.09628	0	0	0	0	0	0	50	50	0	0		
	09404222	Spencer Creek nr Peach Springs	35.800822	-113.65883	0	0	0	0	0	0	13	13	0	0		
	09415000	Virgin Rv at Littlefield	36.891644	-113.92441	0	0	0	0	0	0	80	81	0	1		

improved performance. The CPA used for low outlier data by B17B was one notable example. When low outliers were present in small sample sizes (less than 25 annual peaks), the CPA estimator had much larger MSE values for the 1-percent AEP when compared to EMA. EMA also had a lower MSE in the "excess censoring scenario" and was especially effective when skew was negative. This finding is relevant to Arizona where low-outlier censoring can be as high as 50 percent.

A Monte Carlo analysis by Griffis (2008) clearly demonstrated that EMA is more efficient than B17B-GB at simultaneously employing historical information, regional skew information, and adjustments for low outliers. Similar to the study by Griffis and others (2004), six P3 estimator methods were used in the Monte Carlo experiment with varying systematic and historical periods, nonexceendance thresholds, and population skews. The inclusion of regional and historical information greatly improved the flood frequency analysis of EMA and B17B-GB over the other method-of-moment estimators. When skews were less than −0.2, the MSE for EMA was much lower than B17B-GB, indicating that EMA was outperforming B17B-GB when skews were negative. When no historical information is available the MSE of the quantile estimators showed very minor differences, but as the historical period lengthened, MSE became increasingly lower for EMA. Conclusions were similar to the 2004 (Griffis and others) publication: EMA was more efficient than B17B-GB for incorporating historical flood information, and EMA performs as well as or better than the B17B-GB with its CPA for low outliers in the absence of historical information.

England and Cohn (2008) proposed several changes to the B17B methodology based on a comparison between

B17B-GB and EMA. Using real peak-flow datasets from 82 long-term USGS streamflow records across the Nation, they used RPD metrics, resampling procedures, and Monte Carlo simulation to compare the two estimators. For the single test gage (Big Sandy River at Bruceton, Tennessee 03606500) they demonstrated that when no historical or regional information is available the methods are mostly indistinguishable except for confidence intervals, which B17B-GB always underestimates (Cohn and others, 2001). The relative percent differences increase on the order of 1 to 3 percent for the 2-, 1-, and 0.2-percent AEP when historical and regional information are included, and again major differences were only observed with the confidence intervals. Differences were mostly negligible for this particular gaging-station example, likely due in part to positive station skew, no low outliers, and the similarity in magnitude of historical peak flows relative to the high-flows that occurred during the systematic record. From their other preliminary results and review of Monte Carlo results from other investigations they concluded that EMA performs as well or better when compared to the existing B17B-GB approach, EMA provides much more accurate uncertainty estimates, and EMA is able to incorporate flood information that is inconsistent with the B17B procedures.

EMA, without the MGB test, was compared to B17B-GB and other methods in a U.S. Army Corps of Engineers (USACE) investigation in the Upper Mississippi basin (U.S. Army Corps of Engineers, 2000). This application of EMA was somewhat different from the previous studies discussed because all data below the median value were coded as threshold data, rather than only years missing from the systematic record or low outliers. As a result, the regional EMA-estimated skew coefficient, 0.36, was much higher than that predicted using B17B-GB, −0.26. If the smallest annual floods are much smaller than the largest annual floods, as an increasing number of small floods are considered threshold data (just under 50 percent of all annual peaks, in the USACE study), they have diminishing influence on the distribution fit to the large floods. That is, if the appearance of the distribution curve is "concave-down" on the quantile-probability plot, it will become more linear if the small annual floods on the left-hand tail are given less weight by being considered threshold data. The 1-percent AEP flood at 23 stations as predicted using EMA with station skew was on average 4.8 percent higher than using B17B-GB; the maximum difference was 18 percent higher. The report concluded that all of the methods tested (B17B-GB, EMA, log-normal, and generalized extreme value) produced similar results, and that EMA may be preferable at sites where historical information exits.

Description of Study Area and Regional Flood Information

Arizona's geographic proximity to the Gulf of Mexico and its position in the "horse latitudes" (latitudes of high atmospheric pressure), promotes extreme weather events, both wet and dry (precipitation and drought), often resulting in very high or very low peak-flow discharges for a given water year. Precipitation events responsible for flooding can be broadly grouped into three storm types: convective, frontal, and tropical storms. Convective storms associated with the North American monsoon, especially in southern Arizona, can be unpredictable and intense, and a single localized storm can deliver a majority of the annual precipitation in a particular area. The largest floods often occur during winter frontal storms where antecedent conditions develop from multiple widespread rainfall events that saturate the soil, followed by periods of localized intense rainfall. Although less frequent, dissipating tropical storms from the Pacific Ocean can also cause extreme flooding. These events mostly occur in central and southern Arizona and are responsible for the largest recorded flood events at a number of streamgaging stations. Arizona is also prone to drought, often resulting in zero or low-flow discharge for the annual maximum peak flow for a given water year. In this investigation more than 12 percent of all the peak flows used in the FFA were identified as potentially influential low-flows (PILFs) using the multiple Grubbs-Beck test (fig. 2).

The significant influence of physiography on weather patterns requires regionalization when developing relations between flood frequency and the physical environment. Arizona has three prominent physiographic regions that influence climate conditions and related flood hydrology: the Colorado Plateau Province, Transition Zone, and Basin and Range Province (fig. 3). The Colorado Plateau covers roughly 45,000 square miles, or two-fifths of Arizona, and is characterized by moderate to considerable relief that follows numerous canyon drainages, the most notable being the Grand Canyon of the Colorado River. The average elevation on the Plateau is around 5,000 feet and average rainfall is around 10 inches per year (PRISM Climate Group, 2012). At high elevations peak flows can be influenced by snowmelt, but relatively few gages are affected. Regions 8 and 11 defined in USGS Water-Supply Paper 2433 (Thomas and others, 1997) are mostly located within the Plateau area. Peak flows at streamgaging stations in these regions are generally lesser in magnitude than gages in located in the Transition Zone and Basin and Range Province, with the exception of major drainages, such as the Colorado and Virgin Rivers.

The Transition Zone is characterized mostly by mountainous terrain with small, relatively shallow, intermontane basins. Land-surface elevations range from about 2,000 feet near the confluence of the Salt and Verde Rivers to about 11,400 feet on Mount Baldy in the White Mountains. The Transition Zone has physiographic characteristics of both of the other provinces but is unique in the fact that it is the source of much of the water that sustains streams and rivers in the central part of the State. Most major streams and rivers of the State, with the exception of the Colorado River, have their headwaters in this region including the Salt, Verde, Agua Fria, and Hassayampa Rivers. The area drains several major mountain ranges, including the Bradshaw, White, Mazatzal, Santa Maria, and Sierra Ancha. Precipitation and air temperature are highly variable throughout the Transition Zone and are dependent in large part on land-surface elevation. Average annual precipitation varies widely over the region, from as much as 38 inches per year in the White

Mountains to as little as 16 inches per year at the low elevations along the Salt River (PRISM climate group, 2012). Regions 12 and 14 (Thomas and others, 1997) mostly cover the area of the Transition Zone. Streamgaging stations in these regions generally have the largest peak flows in the State, but many of the stations can also have several PILFs. Because the Transition Zone area shares properties of both the Basin and Range and Colorado Plateau Provinces, streamgaging stations experience multiple flood-generating precipitation processes, such as snowmelt, rainfall, or rain-on-snow.

The physiography of the Basin and Range Province was formed by tectonic activities and can be described as broadly sloping valleys separated by abruptly rising mountain ranges (Anderson and others, 1992). Land-surface elevations range from 100 feet along the lower Colorado River to a few thousand feet in the basins and more than 10,000 feet in some mountain ranges (Wilson, 1962). Compared to the Transition Zone, the Basin and Range lowlands are characterized by less rainfall and higher temperatures because of the predominance of lower land-surface elevations. Mean annual precipitation ranges from less than 3 inches per year in the Yuma area to greater than 30 inches per year in the high elevations of the Chiricahua Mountain Range (PRISM Climate Group, 2012). Average annual precipitation over much of the Basin and Range lowlands is 5 to 15 inches per year. Extended periods of high temperatures above 100 ºF are common throughout the summer months at low elevations. Perennial rivers and streams in the Basin and Range are relatively rare because of the aridity of the region, dams and impoundments upstream that capture streamflow, and groundwater pumping. Many of the streams are intermittent or ephemeral but can have very high flow in response to intense convective thunderstorms. Regions 10 and 13 (Thomas and others, 1997) are located within the Basin and Range Province, and streamgaging stations within these regions have many zero and low-flow peak flows.

Flood Frequency Methods

The Bulletin 17B Grubbs-Beck (B17B-GB) and Expected Moments Algorithm-multiple Grubbs-Beck (EMA-MGB) methods both determine the first, second, and third moments (mean, standard deviation, and skew) of the LP3 distribution, but in a different manner. The following sections provide a brief overview of B17B-GB methods, an explanation of the EMA method, and the MGB test.

Bulletin 17B and the Grubbs-Beck Low Outlier Test

The log-Pearson Type 3 (LP3) distribution has shape, scale, and location parameters (α, β, and τ) and the parameters of the LP3 distribution are functions of the first three population moments (the mean, standard deviation, and skew, denoted by μ, σ, and γ, respectively). The B17B methodology uses a method of moments approach to calculate the first three moments of the LP3 distribution;

$$\alpha = \frac{4}{\gamma^2} \tag{1}$$

$$\beta = sign(\gamma)\left(\frac{\sigma^2}{\alpha}\right)^{1/2} \tag{2}$$

$$\tau = \mu - \alpha\beta \tag{3}$$

which are then used to estimate the flood quantiles. The moments are calculated from the logarithms of the annual

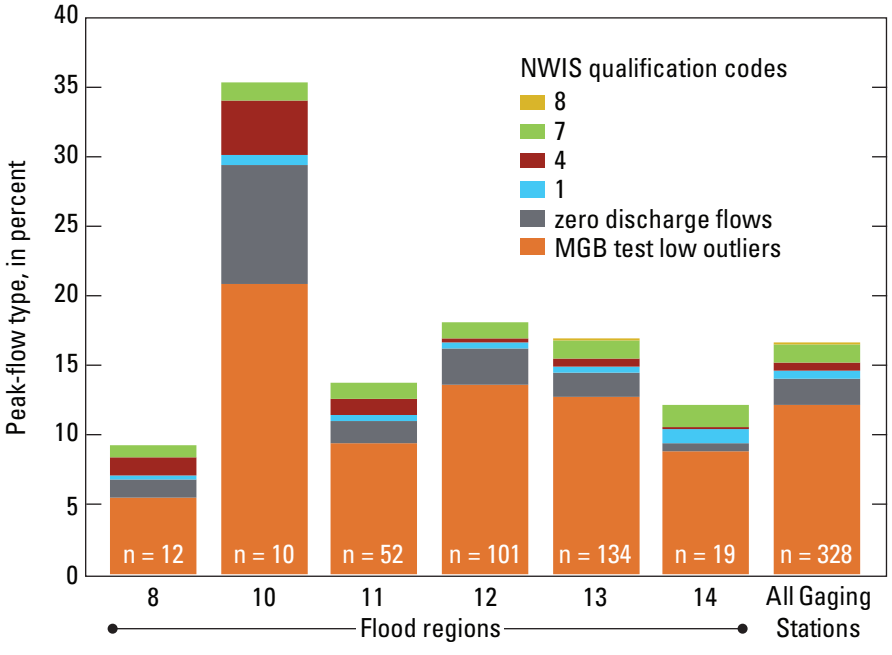


Figure 2. Bar graph showing the percentage of peak flows that have USGS National Water Information System qualification codes 8, 7, 4, 1, equal zero discharge, and identified as low outliers by the multiple Grubbs-Beck (MGB) test. Sites are divided by flood regions as defined in Thomas and others (1997). The number of stations in each region is labeled (n). Remaining percentage of peak flows in each region not qualified.

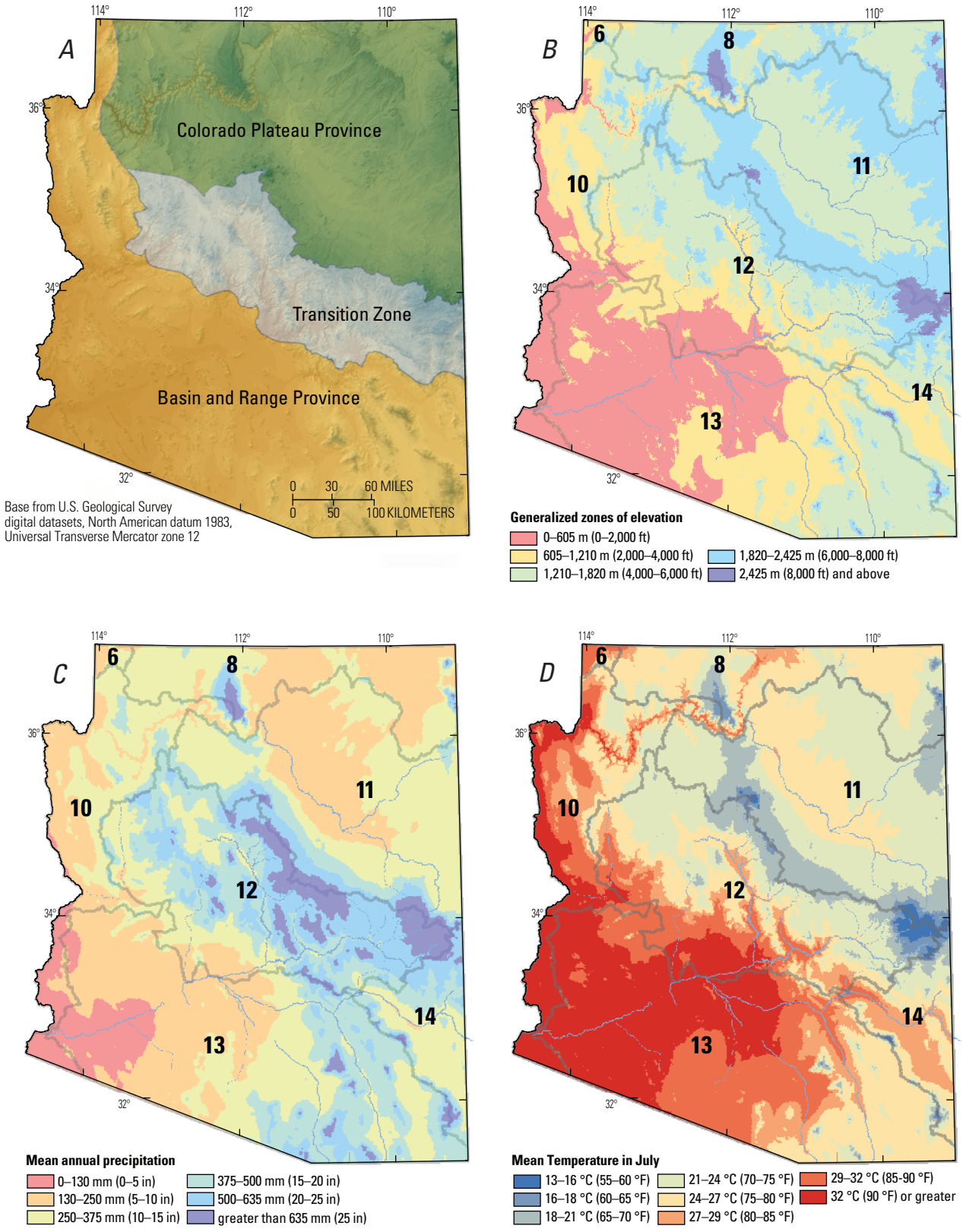


Figure 3. Maps of Arizona showing (*A*) physiographic provinces (modified from Trapp and Reynolds, 1995); (*B*) zones of generalized elevation (created from National Elevation Dataset, Gesh and others, 2002); (*C*) mean annual precipitation; and (*D*) mean temperature in July (grids created with parameter-elevation regressions on independent slopes model data, accessed 2012). Boundaries of numbered flood regions from Thomas and others (1997) shown by heavy grey lines.

peak-flow record and a "*K*" value is determined for a given probability and skew, and the predicted discharge is calculated from the log-linear equation

$$log\ Q = \overline{X} + KS \quad (4)$$

where log *Q* is the base 10 logarithm of the discharge, $\overline{X}$ and *S* are the mean and standard deviation, respectively, of the logarithms of the peak streamflow record, and *K* is a function of the skew and the selected exceedance probability.

Skew is a measure of the symmetry of the shape of the distribution and is sensitive to outliers, especially when sample sizes are small (less than 15). B17B recommends weighting the station skew with a generalized or regional skew to improve the accuracy of the station skew estimator. The regional skew map (plate I) in B17B is more than 30 years old, and several studies have documented that the map is a poor estimator of regional skew (Veilleux, 2009; Stedinger and Griffis, 2008; Griffis and Stedinger, 2007). Recent studies described by Reis and others (2005) have shown that Bayesian generalized least squares (BGLS) regression provides an effective statistical framework for estimating regional skew. A new regional skew analysis using BGLS was conducted for Arizona (Paretti and others, 2013) as part of the update of the regional regression equations for estimating flood frequency and magnitude for Arizona. A constant regional skew of −0.09 and mean square error (MSE) of the skew of 0.08 were used for weighting the station skew for the EMA flood frequency analysis (FFA), and a regional skew of 0.0 and MSE of 0.31 were used for B17B-GB weighted FFA for the comparison (Thomas and others, 1997).

After the LP3 moments are determined and before the flood quantiles are calculated, adjustments are made for historical data, low or high outliers, and zero-flows. In practice high outliers are rarely removed because of the importance of large peak flows, but the other adjustments are common. The order in which these are performed depends on the value of station skew. The historical adjustment is calculated if historical data exist and the systematic record is deemed representative of the period between historical peaks (if more than one exist), and the period between historic peaks and the start of the systematic record. The historical adjustment is calculated by applying a weighting factor based on the relative lengths of the historical and systematic records, either by modifying the logarithms of the peak streamflow record directly or by adjusting the moments, if they have previously been calculated. One limitation of the historical adjustment in B17B is that a discharge value must be specified for historical events. EMA, on the other hand, allows the user to specify that flood peaks during the period between the historical event(s) and the systematic record were below some threshold, without requiring a discharge value for the historical events themselves. In many cases peak-flow records indicate that the flood peak in a particular year was the "highest flood since" a specified water year, but no discharge value or other information is available about the historic peak.

Using the B17B method, outliers are identified using a one-sided t-test, known as the Grubbs-Beck test (Grubbs and Beck, 1972), by determining a threshold using the equation

$$X_T = \overline{X} + K_N S \quad (5)$$

where X_T is the outlier threshold, $\overline{X}$ is the mean logarithm of the systematic peaks (excluding historical data and zero flows), *S* is the standard deviation of the logarithm of the systematic peaks, and K_N is dependent on sample size and is presented in a table of values in B17B. K_N is a critical value at the 10-percent significance level (one-sided) of the lognormal distribution. Peaks above X_T (for the high-outlier test) or below X_T (for the low-outlier test) are removed, along with zero-flows, and the moments recalculated. A conditional probability adjustment (Jennings and Benson, 1969) is then performed, by which the exceedance probabilities are multiplied by a weighting factor based on the proportion of omitted flow (in effect, shifting the quantile-probability curve to the right), and "synthetic" moments are calculated to define the distribution that fits the adjusted probabilities.

Expected Moments Algorithm (EMA)

The Expected Moments Algorithm (EMA) is a moments-based approach that expands on the B17B framework. When gaging stations have a systematic peak-flow record with no historical information or outliers, EMA calculates the same LP3 parameter estimates as the conventional method of moments described in B17B. However, peak-flow data that is nonsystematic, such as historical and uncertain peak-flow information, is treated very differently between the two methods. Peak-flow data that is unaffected by diversions or urbanization is generally reported in the USGS National Water Information System (NWIS) in several ways (fig. 4): peak flows measured or estimated at a continuous or crest-stage streamgaging station during the operation of the systematic record; peak flows documented during a historical period recorded before the start of the systematic record, typically estimated from high-water marks, historical account, or other evidence of paleofloods, and recorded with a NWIS code 7; a peak-flow perception threshold in which a discharge was known to not be exceeded over a period outside the systematic record, and documented with a "highest flood since" a specified water year; peak flows during the systematic record of unknown discharge, but known to be above some threshold, such as a maximum mean daily value (NWIS code 1) or data collected from a streamgaging station destroyed prior to the peak discharge during a major flood event that is recorded with a NWIS code 8; and peak flows during the systematic record of unknown discharge, but known to be below some minimum recordable gage base-flow, and recorded with a NWIS code 4. While infrequent, sometimes a gage height will be reported without an associated peak discharge, and in some cases

statistical methods can be used to conservatively estimate an interval that would capture the flow that occurred. EMA utilizes these nonstandard data types more effectively than B17B, because it can incorporate censored and interval peak-flow data into the analysis. Censored data are expressed in terms of a discharge perception threshold ($T_{thresh\ lower,}$ $T_{thresh\ upper}$) where a range of potential discharges are represented independently of the actual peak flows. Interval discharges ($Q_{int\ lower,}$ $Q_{int\ upper,}$) can be used to characterize peaks known to be greater or less than a specific discharge or that can be reliably estimated with a specific range of discharge.

Historical information is the most common and one of the most important types of information used in flood frequency analysis. The B17B method for addressing historical peaks, by which the systematic record is weighted to represent unobserved values, doesn't completely utilize the information that the historic discharge was not exceeded during the historic period. The EMA was developed specifically to incorporate this nonexceedance, or "threshold" information (Cohn and others, 1997). Furthermore, EMA incorporates censored low outliers and regional skew information simultaneously with historical information. B17B-GB incorporates these sequentially, and the predicted flood quantiles are dependent on the order in which they are performed (Griffis, 2008). A significant difference between the B17B and EMA methods arises from the assumptions made about nonexceedance observations during the historical period. The B17B method assumes that the moments of the nonexceedance observations are equal to the moments of the systematic record, calculated for all peaks below the nonexceedance threshold. The EMA method, in contrast, assumes that the moments of the nonexceedance observations are equal to the moments of the overall distribution, truncated at the nonexceedance threshold. That is, the EMA method identifies the moments of one distribution that fits both nonexceedance observations and the systematic record, whereas the B17B method determines individual distributions for each and combines them in a weighted average.

The EMA method calculates moments using an iterative procedure that adjusts the initial moment estimators based on threshold information (Cohn and others, 1997; Griffis and others, 2004). First, initial sample moments are estimated from the systematic gage record using the method of moments. Next, LP3 parameters are estimated using equations 1–3. These parameters are then used to update the moment estimates. The updated mean is given by

$$\mu_{i+1} = \frac{\Sigma X_S^{<} + \Sigma X^{>} + N_H^{<} E\,[X_H^{<}]}{N} \qquad (6)$$

where the subscript $i+1$ indicates the current step in the iterative procedure, $X_S^{<}$ denotes the logarithms of systematic peaks below some (log) threshold T, $X^{>}$, is the logarithms of systematic and historical peaks above T, $N_H^{<}$ is the number of historical peaks below T, and $E\,[X_H^{<}]$ is the conditional expectation of X, given that $X<T$, which can be calculated from the incomplete Gamma function using the LP3 parameters from the current iteration. The second and third moments are calculated at each iteration using similar equations, and bias-correction factors are included so that the EMA method coincides with the B17B method when there is no historical information available.

EMA and the Multiple Grubbs-Beck (MGB) Test

Outliers are recognized in B17B as "data points which depart significantly from the trend of the remaining data" (Interagency Advisory Committee on Water Data, 1982. Low outliers, in the left-hand tail of the distribution, can have significant influence on the fit of the distribution to the right-hand tail (that is, fitting the largest flood events or estimating the lower AEPs), an effect that increases with increasingly negative skew. Therefore, a statistical test can be applied to determine if the smallest observation deviates from the remaining population. The B17B method allows for the identification and removal of low outliers using the Grubbs-Beck (GB) test. The GB test calculates a one-sided, 10-percent significance-level critical value for a normally distributed sample, but rarely is more than a single low outlier removed. Quantile-probability plots, however, often have a distinct "dogleg" (Thomas and others, 1997), or break in slope, between a few small events and the remaining data. Furthermore, zero-flow years are common throughout Arizona and log (streamflow) is undefined for the LP3 distribution (fig. 5). One reason for multiple low outliers in the statistical analysis is that in arid to semiarid regions, the distribution of annual flood peaks is influenced not only by the distribution of flood-generating rainfall, but also by basin characteristics. In particular, channel-infiltration losses can greatly attenuate or even eliminate an annual peak between two stream gages. Or, rainfall in a given year may be low and evapotranspiration demand high, so that no measurable runoff occurs. The result is that the series of annual peaks appear to be generated from a mixed distribution, whereby small and large flood events are generated by different processes.

To prevent zero- and low-flow events from influencing the distribution fit for large events, multiple potentially influential low outliers should be censored. The B17B method suggests that "procedures for treating outliers ultimately require judgment involving both mathematical and hydrologic considerations" (Interagency Advisory Committee on Water Data, 1982) and allows multiple low outliers to be removed subjectively, but no quantitative method for removing multiple low outliers is presented. A generalization of the GB test was developed by Cohn and others to address this shortcoming (Cohn and others, 2013) by systematically testing the hypothesis that k samples in the left-hand tail are from the same sample of normally distributed observations in the remaining population. The number of samples k is increased until the test statistic indicates the largest possible group of low outliers has been identified.

The multiple Grubbs-Beck (MGB) test involves sequentially evaluating the statistic

$$\tilde{\omega}_{[k;N]} = \frac{X_{[k;N]} - \hat{\mu}_k}{\hat{\sigma}_k} \qquad (7)$$

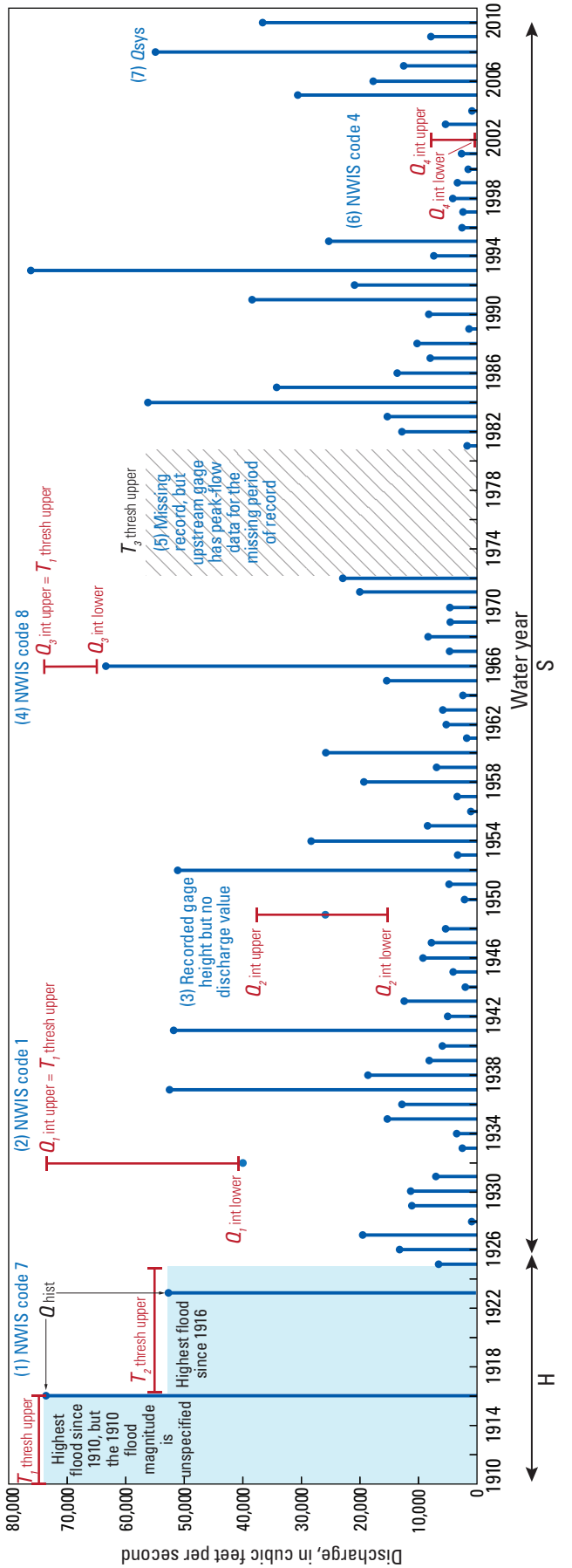


Figure 4. Example peak-discharge time series and the treatment of National Water Information System (NWIS) qualification codes using the Expected Moments Algorithm. Shaded area represents historical period (H) and nonshaded area is the systematic record (S). The seven types of peak streamflow data are numbered: (1) historical discharge (NWIS code 7) (Q_{hist}) and historical perception threshold specified with an upper perception threshold (T_1 and $T_{2\ thresh\ upper}$); (2) mean daily maximum (NWIS code 1) is described with an upper discharge interval bound ($Q_{1\ int\ upper}$) equal to a perception threshold and a known discharge as the lower interval bound ($Q_{1\ int\ lower}$); (3) a recorded gage height, missing the corresponding discharge, could be used to estimate a discharge interval ($Q_{2\ int}$) where both interval bounds are uncertain; (4) discharge greater than known maximum discharge (NWIS code 8) and used as the lower interval bound ($Q_{3\ int\ lower}$) and an upper interval bound equal to a perception threshold ($Q_{3\ int\ upper}$); (5), missing peak-flow data for a portion of the period of record but a correlated nearby gage could be used to set an upper perception threshold ($T_{3\ thresh\ upper}$), (6) a discharge less than a known recordable gage base flow (NWIS code 4) is describe with a lower interval bound ($Q_{4\ int\ lower}$ = zero) and an upper interval bound ($Q_{4\ int\ upper}$) equal to minimum gage base discharge; and (7) systematic peak-flow discharge (Q_{sys}).

for each flood peak. Here, $X_{[k;N]}$ denotes the logarithm of the kth smallest flood peak in an ordered sample, and $\hat{\mu}_{Z;k}$ and $\hat{\sigma}_{Z;k}$ are the partial mean and partial standard deviation, respectively, for all flood peaks larger than $X_{[k;N]}$. The null hypothesis, that all peaks are drawn from a sample of independent, normally distributed random variates, is tested by comparing the test statistic, $\tilde{\omega}_{[k;N]}$, to a critical value η. Noting that

$$P[\tilde{\omega}_{[k;N]} < \eta] = P\left[\frac{Z_{[k;N]} - \hat{\mu}_{Z,k}}{\hat{\sigma}_{Z,k}} < \eta\right] \quad (8)$$

where $Z_{[k;N]}$ is the kth-order statistic (equal to $Z_{[k;N]} - \mu) / \sigma$ in a standard normal sample of size N, corresponding to the rank of $X_{[k;N]}$, and $\hat{\mu}_{Z;k}$ and $\hat{\sigma}_{Z;k}$ are the partial mean and partial standard deviation for Z-statistics larger than $Z_{[k;N]}$, therefore η can be computed by integrating the distributions of , $Z_{[k;N]}$, $\hat{\mu}_{Z;k}$ and $\hat{\sigma}_{Z;k}$. A quasi-analytical solution of the three-dimensional integral can be derived by approximating the joint distribution of $\hat{\mu}_{Z;k}$ and $\hat{\sigma}_{Z;k}$. The set of low outliers is then identified as those flood peaks equal to or smaller than that peak where $P[\tilde{\omega}_{[k;N]} < \eta]$ is less than the threshold value of 10 percent, and the null hypothesis is rejected.

A newer version of the MGB test, available in the PeakfqSA software (version 0.995), was released for inclusion in an updated B17B Guidelines and for implementation in an updated version of the USGS PeakFQ software (version 7.0). The newer iteration of the MGB test operates in a two-step procedure to identify potentially influential low flows (PILFs). First, starting at the median and sweeping outward towards the smallest observation, each observation is tested and is identified as an outlier if $p\,(i;n) \leq 0.5$ percent. If the kth largest observation is identified as a low outlier, the outward sweep stops and the kth and all smaller observations (that is, for all $i \leq k$) are also identified as low outliers. In the second step as with the current GB procedure in B17B, an inward sweep starts at the smallest observation and moves towards the median, where the ith observation is identified as an outlier if $p(i;n) \leq 10$ percent. If an observation $m \leq 1$ fails to be identified as outlier by the inward sweep, the inward sweep stops. The number of low-outliers identified by the procedure is then the larger of k and $m-1$. PeakfqSA version 0.995 of the MGB test more closely mimics B17B by using 10-percent significance test with its single inward sweep.

PeakfqSA version 0.995 of MGB test uses a less restrictive probability threshold in the testing procedure than the version 0.974, meaning that the test identifies fewer peaks as PILFs. The highly variable flooding regime in Arizona can result in a highly skewed flood frequency distribution where the lowest 50 percent of the peaks can have the greatest influence on the right-hand tail of flood frequency distribution, and an MGB test that is more aggressive in censoring potentially influential portions of the peak flow population is necessary for certain gaging-station peak-flow records. A comparison was made between the effects of using the two MGB tests in the PeakfqSA versions 0.974 and 0.995. A majority of the 328 stations showed no difference, but about 20 percent of the stations had a 1-percent AEP difference ranging from 0.5 to 106 percent and the median skew of −0.437 increased to −0.934 at those same stations. Inspection of LP3 flood frequency fits at the stations with the greatest differences consistently showed better fits of the right-hand tail of the distribution with the MGB test in PeakfqSA version 0.974. Although there are many advantages to retaining more peak-flow data using a more conservative MGB test, there are regions such as the Southwest, where additional censoring must be done to ensure that the peak-flow data function within LP3 distribution parameters. Results from the comparison influenced the decision to maintain the same version of the MGB test that was used in Gotvald and others (2012) as a means to better fit the peak-flow data observed in gaging stations in Arizona.

B17B and EMA Comparison Methods

Software

Two FFA software packages were used in the method comparison. The USGS software PeakFQWin version 5.2.0 (http://water.usgs.gov/software/PeakFQ/; Flynn and others, 2006) was used to compute the B17B-GB flood frequency analysis, and the EMA-MGB test was implemented using PeakfqSA versions 0.974 and 0.995 (http://www.timcohn.com/TAC_Software/PeakfqSA/; Cohn, 2011).

Peak-Flow Data

All streamgaging stations with 10 or more annual peak-flow records were retrieved from the USGS National Water Information System (NWIS) database at http://nwis.waterdata.usgs.gov/usa/nwis/peak. Three additional peak-flow data sources were used in this investigation. Two of the cooperating Flood Control Districts (FCD), Maricopa County and Pima County, supplied peak-flow information from County-operated streamgaging stations. In certain instances these streamgaging locations are co-occurring or discontinued USGS streamflow gages, which provide the Counties with preliminary or supporting discharge and channel dimension information. When available, concurrent data were checked and compared for quality assurance. The FCD of Maricopa County data are available at: http://www.fcd.maricopa.gov/Rainfall/Streamflow/streamflow.aspx, and Pima County FCD data were acquired via electronic request. Data from Walnut Gulch Experimental Watershed and the Santa Rita Experimental Range, both in southeast Arizona, are available at the U.S. Department of Agriculture, Agricultural Research Service, Southwest Watershed Research Center webpage http://www.tucson.ars.ag.gov/dap/Default.htm.

The peak-flow data were reviewed to ensure quality of the records and identify the presence of trends. A standardized data-check review process was completed using the PFReports

computer program described by Ryberg (2008) to identify any peak-flow file anomalies or errors within the file. Streamgaging stations that were highly urbanized or regulated were not used in the analysis. Streamgaging stations with unregulated peak-flow record prior to regulation or significant urbanization were used in the analysis with the regulated or urbanized peaks excluded.

USGS NWIS Qualification Codes and Software Utilization

USGS (NWIS) qualification codes are assigned to peak flows where the magnitude and frequency have been affected by watershed conditions that may have increased or decreased the magnitude of a flood (for example, dam failure, urbanization, or regulation). Qualification codes are also assigned to historic peaks or situations where measurement conditions may have affected the accuracy of the recorded value. By default the PeakFQ software will recognize several of these qualification codes and use them differently to control the statistical computation, whereas EMA does not recognize these codes, requiring the user to supply the information in the form of a nonexceedance threshold or as interval data (fig. 4). This investigation will focus on the USGS NWIS qualification codes 1, 4, 7, and 8. Approximately 80 percent of the 328 streamgaging stations used in this study have one, all, or some combination of historical, low-outlier, or censored peak-flow data (fig. 2).

NWIS code 1 indicates that the instantaneous peak discharge is a maximum daily average. A code 1 implies that the peak estimate is likely biased low. PeakFQ treats peaks qualified with a code 1 the same as all other peak flows in the record. Because EMA allows the user to include data as an interval or a threshold, code 1 data can be entered as a peak-flow interval between the reported daily mean maximum and the largest flood that could be expected to result in a reliable discharge estimate or, if no additional information is available for the upper bound, then it can be coded as infinity indicating it is unknown.

NWIS code 4 identifies discharge less than the minimum recordable discharge at a site, often a result of an unmeasured and nonzero flow that occurred at some point during the water year. For example, a crest stage gage may be set higher than the highest peak flows during a dry year. Lacking other evidence collected in the field, the discharge corresponding to the pin elevation (the lowest recordable flow) could be entered with a code 4. Peaks less than or equal to the highest peak flagged with code 4 are automatically set to zero by the PeakFQ program and

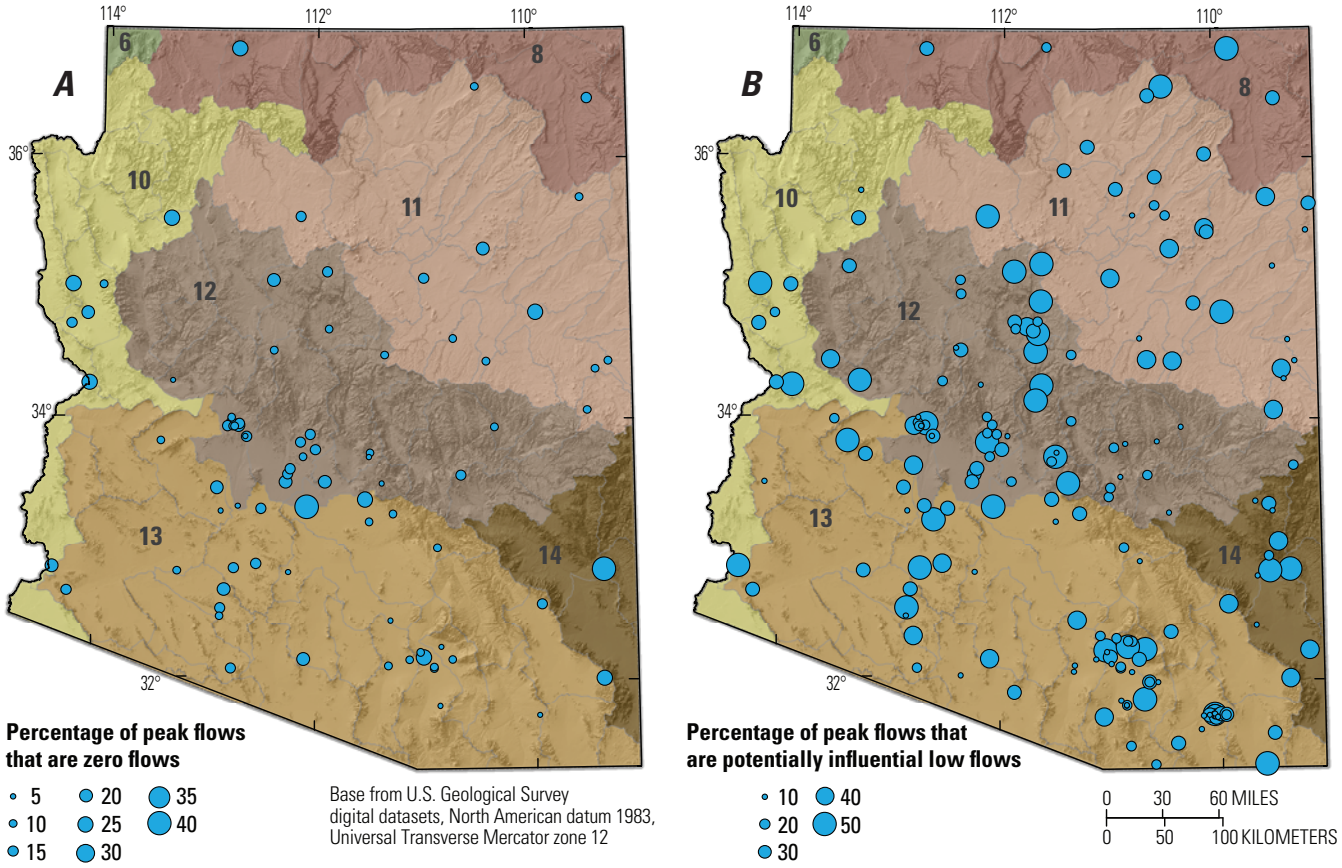


Figure 5. Maps of Arizona showing the percentage of (*A*) zero-flow annual peak flows and (*B*) potentially influential low flows as identified by the multiple Grubbs-Beck test for streamgaging stations used in the comparison study. Color-coded and numbered flood regions from Thomas and others (1997).

a conditional-probability adjustment is applied. As with Code 1, the flexibility of EMA allows the user to supply peak flow information in the form of an interval. In the situation of a code 4 the upper bound is the minimum recordable discharge and the lower bound is a smaller or zero discharge.

NWIS code 7 indicates a historic peak flow, defined as occurring either before or after the systematic record, or during extended breaks within the systematic streamgaging record, and that would not have been observed or recorded except for evidence indicating that they were of relatively large magnitude. Peaks with code 7 form a biased subsample and have to be treated separately from systematic records in statistical analyses. PeakFQ excludes code 7 peaks unless the user specifies the historical period associated with that recorded peak, in which case the program applies a weighted historical adjustment to the flood frequency curve. PeakFQ sets an historic threshold at a single historic discharge value for the user-specified historical period, thus omitting magnitude information provided by other historic peaks that occurred over the period. EMA employs a more general description of peak-flow data and eliminates the need to distinguish between historical and systematic data. With EMA, both historic peaks and each missing year between the historic and systematic records can be described with interval or nonexceedance information, effectively providing the equivalent of decades to hundreds of years' worth of systematic streamgage data (Cohn and others, 1997).

NWIS code 8 indicates the discharge was actually greater than the indicated value—for example, if a gage was overtopped during a high-flow event and high water marks were not available to support more exact estimates. By default, codes 8s are always excluded from PeakFQ statistical analyses because they may grossly underestimate the actual peak flow. Instead of eliminating this potentially useful information, EMA can more easily incorporate this peak-flow information as an interval between the indicated value and the largest flood that could be expected to result in a reliable discharge estimate. If no additional information is available for the upper bound, then it is entered as infinity, indicating that it is unknown.

Relative Percent Difference

The relative percent difference (RPD) was used as the metric to measure similarity between EMA-MGB and B17B-GB AEP flow estimates. For each site the RPD is computed between EMA-MGB and B17B-GB for the specified AEPs:

$$RPD(Q)_{g,P} = \left(\frac{\hat{Q}_{g,P}^{EMA\text{-}MGB} - \hat{Q}_{g,P}^{B17B\text{-}GB}}{\hat{Q}_{g,P}^{B17B\text{-}GB}} \right) * 100 \qquad (9)$$

where P = 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent AEPs and g is the skew coefficient estimate. Two values for g are used: the at-site or station skew coefficient $g = \hat{\gamma}$ and a weighted skew coefficient $g = \hat{G}$. The B17B guidelines recommend using a weighted estimate of the skew by weighting the station skew2 and a generalized, or regional, skew to provide better estimates. For this investigation EMA-MGB was used with a regional skew of −0.086 and a mean square error (MSE) of the skew of 0.079 and B17B-GB used a generalized skew of 0 and MSE of 0.31. $RPD(\hat{Q})_{g,P}$ values close to zero indicate the two methods are providing similar quantile estimates, negative values indicate B17B-GB is predicting larger peak flows at a particular AEP, and positive values indicate EMA-MGB is predicting larger peak-flow estimates. The $RPD(\hat{Q})_{g,P}$ statistic does not indicate which method is a better model to the observed data.

Goodness-of-Fit

For several long-term streamgaging stations a visual, qualitative goodness-of-fit assessment was made to determine how well B17B-GB and EMA-MGB were fitting the observed peak-flow data. The frequency curves were superimposed, and the observed data were graphed with each method's respective plotting positions (B17B uses a Weibull and EMA-MGB uses a Hirsch-Stedinger). A more quantitative approach was also used, consisting of selecting the peak- flow data at a streamgaging station that exceeded the 90th, 75th, and 50th percentile of the data distribution (greatest 10, 25, and 50 percent of the data, respectively) and computing the percent difference of the observed discharge value to the fitted discharge value. The absolute values of these percent differences were then averaged, individually for each percentile, as a measure of total discrepancy from the fitted frequency curve (fig. 6). The intent was to maintain a normalized scale of percent difference rather than a typical goodness-of-fit statistic or sum of squares residuals that try to estimate individual fit error. The goodness-of-fit statistic was computed using the following equation:

$$MAPD_{g,Px} = \frac{1}{n} \sum_{i=0}^{n} \left| \left(\frac{\hat{Q}_{ig,Px}^{Obs} - \hat{Q}_{ig,Px}^{Fitted}}{\hat{Q}_{ig,Px}^{Fitted}} \right) \right| * 100 \qquad (10)$$

where $MAPD$ is the mean absolute percent difference, g denotes station skew at an individual streamgaging station, Px equals the 90th, 75th, and 50th percentile of the observed peak flow-data, n is the number of peak flows within that percentile, and $\hat{Q}_i^{Obs}$ and $\hat{Q}_i^{Fitted}$ are the measured and LP3-fitted peak flows, respectively. Differences in the $MAPD$ distributions for each method were examined using a nonparametric Wilcoxon rank-sum test to assess whether one method had lower percent differences than the other.

Resampling Procedures

The true occurrence probability of a given peak flow is always unknown. One method to address this problem is random resampling, in which some subsample of the data at a particular gage is used to predict flood quantiles, which can be compared to the quantiles predicted using all of the data (representing the best estimate of the true occurrence probability). If many random

subsamples are taken, inferences can be made regarding the uncertainty of a particular method. Resampling can be carried out using either measured peak flows or synthetic data generated from assumed statistical distributions. The latter instance, in which the true occurrence probabilities are known, is useful for determining whether a flood-frequency method is effective at recovering the underlying distributions, and it is presented in the "Monte Carlo Simulations" section. With measured peak flows, resampling is useful for comparing the behavior of different fitting methods. The following discussion concerns resampling using measured discharge data.

The underlying assumption of resampling as used in this study is that any subsample of the measured peak flows at a streamgaging station follows the same distribution as the entire population. This follows directly from B17B, which states that all peak flows in an FFA should be independent, random, and homogenous (Interagency Advisory Committee on Water Data, 1982). Resampling can thus be used to evaluate the "robustness" of a method—that is, whether variations in the input data and the number of low outliers will cause large variations in the predicted AEPs.

Resampling with measured peak flows is useful for evaluating the robustness of a method when multiple low outliers are present, but it is less effective for evaluating robustness in the presence of historical peaks. The reason is that a subsample of a population is likely to contain none, one, or many low outliers; that is, the number of low outliers varies between zero and the number of low outliers in the entire population. In contrast, gages with historical peaks usually have only one and subsamples drawn from this population have either one or zero historical peaks, resulting in bimodal behavior whereby one set of AEPs is most likely if the historical peak is included in the subsample, and a different set of AEPs is if the historical peak is not included. Therefore, the resampling method presented here focuses on the behavior of flood-frequency methods in the presence of low outliers. The influence of historical peaks on synthetic data is discussed in the "Monte Carlo Simulations" section.

Eight streamgaging stations with relatively long record lengths—necessary to adequately estimate the "true" occurrence probability—were considered in the resampling analysis (table 2). Record length varies from 49 to 67 years. These streamgaging stations have records with no low outliers, records with several PILFs where the low-occurrence-probability quantiles were different between EMA-MGB and B17B-GB, and records with several PILFs where the low-occurrence-probability quantiles were similar between EMA-MGB and B17B-GB. For each streamgaging station 1,000 subsamples were drawn, each comprising a random sample of size equal to one-half the

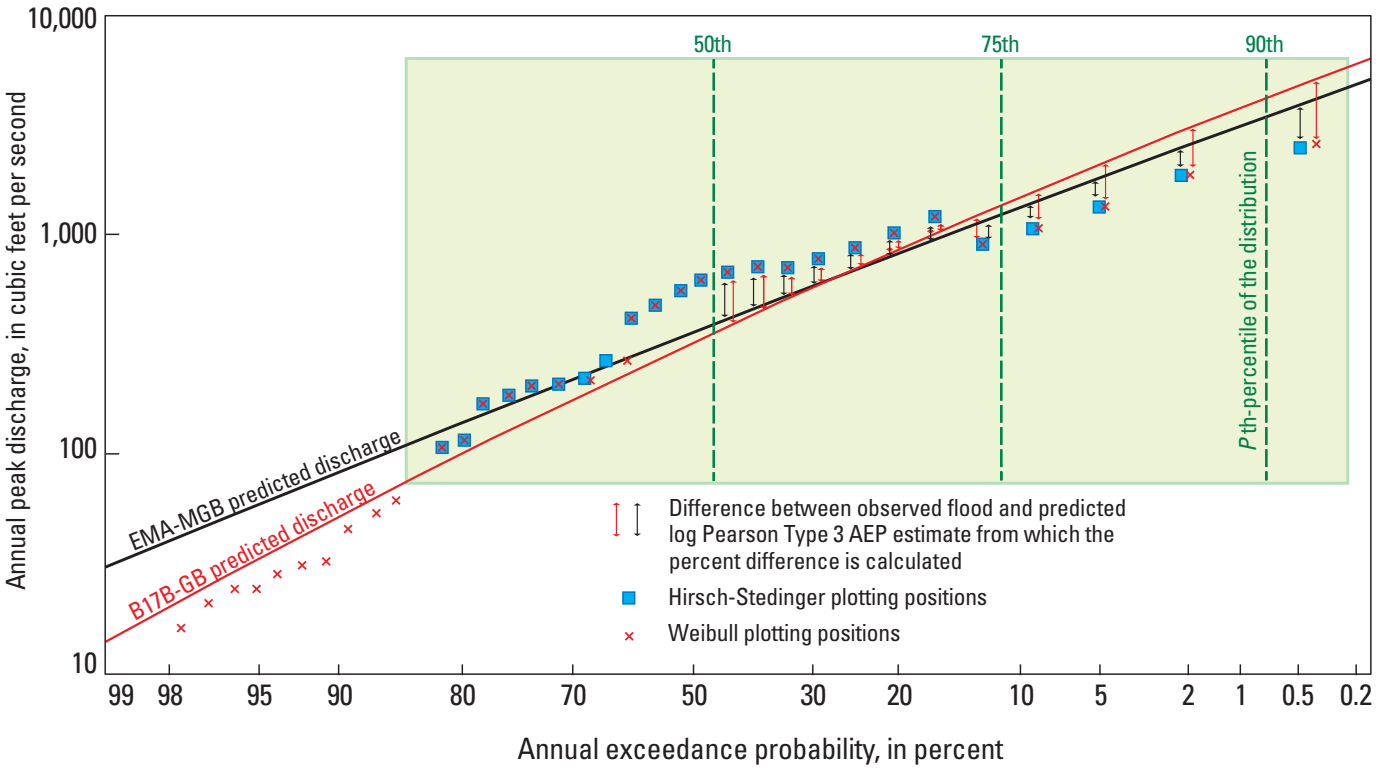


Figure 6. Schematic diagram demonstrating how goodness-of-fit statistics were calculated. From the population of nonpotentially influential low flow outliers, the difference between the measured flood and the predicted discharge of the annual exceedance probability (AEP) estimate equivalent to the respective method's plotting position was determined for 90th, 75th, and 50th percentiles of the peak-flow record. Because different plotting positions are associated with each method (B17B–GB and EMA–MGB), the AEP for a particular measured peak may not be identical.

Table 2. Streamgaging stations used in the resampling procedure.

[B17B-GB, Bulletin 17B Grubbs-Beck test; EMA-MGB, Expected Moments Algorithm multiple Grubbs-Beck test]

Station identifier	Station name, in Arizona	Period of record in years	Number of multiple Grubbs-Beck potentially influential low flows in complete record	Description
09480000	Santa Cruz River Near Lochiel	62	8	Large difference between B17B-GB and EMA-MGB
09482000	Santa Cruz River At Continental	67	0	Moderate difference between B17B-GB and EMA-MGB
09486500	Santa Cruz River At Cortaro	65	8	Small difference in quantiles; large difference in skew
09505200	Wet Beaver Creek Near Rimrock	49	22	Moderate difference between B17B-GB and EMA-MGB
09505350	Dry Beaver Creek Near Rimrock	50	20	Low outliers not from a distinctly different population
09513780	New River Near Rock Springs	49	8	Small difference in quantiles
09516500	Hassayampa River Near Morristown	58	5	Small difference in quantiles
09517000	Hassayampa River Near Arlington	49	13	Small difference in quantiles

number of annual peaks in the complete record. Each subsample was analyzed using both the B17B-GB and EMA-MGB methods. Historic peaks (NWIS code 7s) were included in the population of annual peaks from which samples were drawn but were not used to define a period of nonexceedance (that is, any historic peaks were included as part of the systematic record). Any threshold information was excluded from the EMA-MGB analysis. The entire resampling procedure was performed twice, once using station skew and once using a weighted regional skew.

Monte Carlo Simulations

The primary limitation of describing method performance with the RPD, correlations, and resampling procedures is that the true answer is unknown. Monte Carlo simulations provide a quantitative measure of performance by repeatedly drawing random observations from a single specified population where the true AEPs are known. Cohn and others (1997), Griffis (2008), and England and others (2003) have all done extensive Monte Carlo simulations to understand model performance for various distribution types. A portion of the work done by Cohn and others for the HFAWG is relevant to the Arizona method comparison and it is presented and summarized in this investigation. The following distributions were tested with the Monte Carlo simulation: LP3 ($G = 0$) without regional information; LP3 with positive skew ($G = 0.5$ and 1.0) without regional information; and LP3 with a negative skew ($G = -0.5$ and -1.0) without regional information. The results presented are based on 1,000 replicate samples drawn from each LP3 distribution variant. Each sample consists of 40 peaks representing the systematic record (S) within a 100-year historical period (H). For each sample-replicate, B17B-GB, and EMA-MGB were fitted and the true 1-percent AEP reported. To test the influence of historical flood information, simulations were run using three historical scenarios of 1, 2, and 10 historical floods occurring over the 100-year historical period. The effective record length (ERL) is the amount of equivalent systematic record that would provide the precision achieved with both systematic and historical information and it is used to quantify the efficiency of the different estimators. ERL is defined as the ratio:

$$ERL = 40 \left(\frac{Var\left[\log\left(\hat{Q}_{01\%}(S = 40, H = 0)\right)\right]}{Var\left[\log\left(\hat{Q}_{1\%}(S = 40, H = 100)\right)\right]} \right) \quad (11)$$

where Var is the variance of all the simulated 1-percent $AEP_{1\%}$ flow estimates for the systematic (S) record of 40 years and the historical period length (H) of 100 years. The analysis is further simplified by consideration of the average gain (AG) in ERL statistic, which is defined as:

$$AG = ERL - 40 \quad (12)$$

and the statistic expresses the benefit of each year of historical information in terms of an equivalent amount of systematic data.

Results and Discussion

Relative Percent Differences

The relative percent difference (RPD) between EMA-MGB and B17B-GB are plotted as boxplots for the 10-, 1-, and 0.2-percent AEPs in figure 7. For each boxplot, the three horizontal lines forming the box are the 25th quartile, median, and 75th quartile, the "whiskers" represent an outlier threshold at a distance 1.5 times the interquartile range (IQR; it is the 75th quartile minus the 25th quartile), and the open circles (data points graphed on top of one another were jittered or randomly offset along the x-axis for graphical visibility) are

Table 3. Summary statistics for the percent difference between the largest observed peak and the predicted annual exceedance probability (AEP) log-Pearson Type 3 (LP3) estimated flow for streamgaging stations in National Water Information System (NWIS) category 2 (historical information only) by Bulletin 17B (B17B-GB) and the Expected Moments Algorithm (EMA-MGB).

[The closer the minimum, 10th-, 25th-, median, 75th-, and 90th- percentile, maximum and mean are to zero, the closer the predicted peak-flow estimate is to the observed peak flow, suggesting a better frequency fit. The smaller the standard deviation and standard error of the mean the more precise the predicted peak-flow estimate is near the observed peak flow]

Summary Statistic, in percent	Bulletin 17B-Grubbs Beck	Expected Moments Algorithm-Multiple Grubbs Beck
Minimum	-229	-57
10th percentile	-122	-33
25th percentile	-58	-14
Median	-19	7
75th percentile	12	23
90th percentile	37	34
Maximum	69	50
Mean	-27	4
Standard Deviation	59	25
Standard Error of the Mean	9	4
Lower 95 confidence interval	-45	-4
Upper 95 confidence interval	-10	11
Streamgages with estimates closest to observed peak flow (Number of gages)	15	31

RPDs considered outliers. The IQR is a robust descriptive statistic that provides a measure of statistical dispersion, and it represents the middle 50 percent of the data within the distribution. Boxplots centered on zero indicate each method is providing similar AEP flow estimates. If the RPD boxplot quantile statistics are greater than zero then the magnitude of the EMA-MGB AEP flow estimate is larger than the B17B-GB estimate, and the opposite is true if the RPD boxplot quantile statistics are negative. Summary statistics for all flood quantiles computed using station and weighted skew can be found in appendixes 1 and 2.

The boxplots of all streamgaging station RPDs, regardless of USGS NWIS qualification codes or low-outlier information, showed the median RPD to be approximately zero for the 10-, 1-, and 0.2-percent AEPs (fig. 7*A*, table 4), indicating little difference for the middle of the RPD distribution. Although the median RPD was consistently zero for all AEPs, the IQR increased almost 4-fold from the 10-percent to the 0.2-percent AEP flow estimate. The increase in variability coincided with a positive shift in the 75th quartile and an increase in the number of positive outliers (a positive RPD indicates that the EMA-MGB flow estimate is greater than B17B-GB). The number of sites with RPDs exceeding 100 percent increased from 1 for the 10-percent AEP flow estimate to 19 for the 0.2-percent AEP flow estimate. RPDs ranged between -84 and 262 percent for the 0.2-percent AEP flow estimate. The different treatment of historical and low-outlier peak flow by each method was responsible for at least a third of the streamgaging stations exceeding an RPD of 20 percent. The results showed that the treatment of historical and low-outlier data will have a greater influence on the lower AEP flow estimates (1, 0.5, and 0.2 percent) than the higher AEP flows (50, 20, and 10 percent).

B17B recommends weighing a regional skew with the station skew in the flood frequency analysis (FFA) to improve the AEP flow estimates. This investigation used a different regional skew value for each method. The updated Arizona regional skew derived from the BGLS analysis was used in the EMA weighted skew and the regional skew and MSE calculated in USGS Water-Supply Paper 2433 (Thomas and others, 1997) was used in the B17B weighted skew. The median RPDs for the 10-, 1-, and 0.2-percent AEPs were all less than zero ((fig. 7*B*, table 4). The IQRs were similar to the station skew comparison, but the 75th quartile and the 25th quartile shifted below zero (negative direction). RPDs were more negative as a result of the more positive regional skew used for the B17B-GB analysis. Weighting with a slightly more positive regional skew (0.00 for B17B-GB, −0.09 for EMA-MGB) caused the curvature of the fitted distribution to become more concave, resulting in greater B17B-GB magnitude estimates. The combination of the value of the at-site and regional skew values as well as their respective MSEs gave considerably more weight to the slightly negative regional skew with a much lower MSE. The effect was visibly more pronounced in the right-hand tail of the distribution fit affecting the less probable AEP flow estimates. With the exception of category 4 this negative shift occurred for all NWIS qualification code categories and regional groups.

RPD by NWIS Qualification Code Categories

Sites were split into four categories based on NWIS qualification codes and the presence of potentially influential low-flow (PILF) peaks: (1) stations that have no low-outliers as identified by the MGB test, no historical data (code 7), and no other NWIS qualification codes (1, 4, and 8); (2) stations that exclusively contain data coded with a 7 or historical information specifying a water year preceding the historic flood in which an unquantified flood that magnitude or larger had been documented; (3) stations that exclusively have PILFs identified by the MGB test and no other NWIS code information; and (4) stations that have one or more NWIS qualification codes 1, 4, or 8 and contain no historical information or PILFs.

The RPDs for category 1 (n = 61) were almost all zero (fig. 8*A*, table 4). This is to be expected because EMA-MGB provides nearly the same AEP flow statistics when no NWIS qualification codes or PILFs are present in the data. The few outliers on the boxplot reflect streamgaging stations where censored flood information was added to the EMA-MGB input file in the form of interval data, a data type that cannot be used by B17B. Examples of these interval data include gaging stations for which a gage height was recorded but no

Table 4. Summary statistics for the relative percent difference between Bulletin 17B and the Expected Moments Algorithm using a station skew and a weighted regional skew, by National Water Information System category.

[The 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2- annual exceedance probability (AEP) statistics are grouped by U.S Geological Survey National Water Information System qualification code category. Number, number of stations; IQR, interquartile range; Std Dev, standard deviation; Std Err Mean, standard error of the mean; MGB, multiple Grubbs-Beck test; negative numbers indicate that Bulletin 17B Grubbs-Beck flow estimates are greater than Expected Moments Algorithm multiple Grubbs-Beck flow estimates; positive number indicate the opposite]

	All Gaging Stations-Station skew								All Gaging Stations-Weighted skew							
	Percent AEP								Percent AEP							
	50	20	10	4	2	1	0.5	0.2	50	20	10	4	2	1	0.5	0.2
Number	328	328	328	328	328	328	328	328	328	328	328	328	328	328	328	328
Minimum	-28	-34	-46	-62	-71	-77	-81	-84	-46	-34	-58	-76	-82	-87	-90	-93
10th	-8	-10	-11	-17	-22	-27	-32	-39	-9	-13	-20	-28	-32	-35	-37	-41
25th	-2	-3	-5	-6	-6	-8	-9	-12	-2	-5	-9	-13	-15	-17	-19	-22
Median	0	0	0	0	0	0	0	0	1	0	-1	-3	-4	-5	-5	-6
75th	6	3	3	5	7	11	12	19	10	2	2	3	4	5	7	9
90th	17	9	9	14	21	32	44	62	28	6	9	14	17	19	23	28
Maximum	259	111	138	102	77	112	156	263	297	45	69	121	153	178	205	276
IQR	9	6	8	10	13	18	21	30	12	6	11	16	20	23	26	31
Mean	6	0	-1	0	0	2	4	9	9	-1	-4	-5	-5	-5	-5	-4
Std Dev	23	11	13	15	19	25	33	48	31	9	15	21	25	28	32	37
Std Err Mean	1	1	1	1	1	1	2	3	2	1	1	1	1	2	2	2
Lower 95th	3	-1	-2	-2	-2	-1	1	4	6	-2	-5	-7	-8	-8	-8	-8
Upper 95th	8	2	1	1	2	5	8	14	13	0	-2	-3	-3	-2	-1	0
Qualification Code Categories—No low-outliers historical or other codes																
Number	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61
Minimum	-18	-34	-42	-51	-56	-61	-61	-69	-18	-34	-41	-50	-54	-59	-63	-67
10th	0	0	0	-1	-1	-1	-1	-2	-2	-1	-3	-8	-13	-17	-21	-27
25th	0	0	0	0	0	0	0	0	-1	0	-1	-4	-6	-9	-11	-14
Median	0	0	0	0	0	0	0	0	1	0	-1	-2	-3	-3	-4	-6
75th	0	0	0	0	0	0	0	0	2	1	0	1	2	2	2	2
90th	0	0	0	0	0	0	0	0	5	1	2	5	7	9	11	15
Maximum	28	28	23	15	9	3	3	0	27	27	24	18	14	14	18	23
IQR	0	0	0	0	0	0	0	0	3	1	2	5	8	11	13	16
Mean	1	0	0	-1	-2	-2	-2	-3	2	0	-1	-2	-3	-4	-5	-7
Std Dev	6	6	7	7	8	9	10	11	6	6	7	8	10	12	14	16
Std Err Mean	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2
Lower 95th	-1	-1	-2	-3	-4	-4	-5	-6	0	-1	-2	-4	-6	-7	-9	-11
Upper 95th	2	2	1	1	1	0	0	0	3	2	1	0	-1	-1	-2	-3
Historical Only																
Number	46	46	46	46	46	46	46	46	46	46	46	46	46	46	46	46
Minimum	-20	-25	-28	-34	-39	-44	-44	-54	-17	-26	-31	-36	-40	-46	-52	-59
10th	-10	-12	-15	-18	-20	-23	-23	-33	-10	-13	-17	-19	-22	-27	-33	-40
25th	-2	-4	-5	-10	-9	-11	-11	-14	-2	-2	-5	-8	-10	-11	-13	-19
Median	0	0	0	0	0	0	0	0	1	1	0	-1	-3	-5	-6	-8
75th	2	4	5	6	8	9	9	12	5	5	5	3	2	2	1	0
90th	6	14	19	20	25	33	33	47	11	19	25	24	18	20	25	31
Maximum	17	20	29	52	71	93	93	155	37	45	39	44	55	67	79	97
IQR	4	8	10	16	17	20	20	26	8	8	10	10	12	13	15	20
Mean	-1	0	0	1	1	2	3	4	1	2	1	-1	-2	-3	-5	-6
Std Dev	6	9	12	17	20	25	29	36	9	12	13	16	18	21	25	29
Std Err Mean	1	1	2	2	3	4	4	5	1	2	2	2	3	3	4	4
Lower 95th	-2	-3	-3	-4	-5	-5	-6	-7	-1	-2	-3	-6	-8	-10	-12	-15
Upper 95th	1	3	4	6	7	9	12	15	4	5	5	4	3	3	3	3

Table 4. Summary statistics for the relative percent difference between Bulletin 17B and the Expected Moments Algorithm using a station skew and a weighted regional skew, by National Water Information System category.—Continued

	All Gaging Stations-Station skew								All Gaging Stations-Weighted skew							
	Percent AEP								Percent AEP							
	50	20	10	4	2	1	0.5	0.2	50	20	10	4	2	1	0.5	0.2
Low-outliers Only (MGB test)																
Number	112	112	112	112	112	112	112	112	112	112	112	112	112	112	112	112
Minimum	-15	-29	-46	-62	-71	-77	-77	-84	-17	-27	-58	-76	-82	-87	-90	-93
10th	-7	-16	-15	-18	-24	-29	-29	-37	-5	-15	-28	-36	-43	-46	-49	-51
25th	-3	-7	-8	-9	-8	-9	-9	-11	-2	-8	-17	-22	-25	-26	-27	-28
Median	-1	0	-1	1	1	1	1	3	4	-2	-6	-7	-8	-8	-9	-9
75th	10	2	2	5	9	16	16	33	20	1	1	2	4	6	6	8
90th	28	5	4	12	24	38	38	92	39	2	5	9	12	14	17	22
Maximum	259	37	32	27	53	97	97	263	297	26	20	25	36	48	61	79
IQR	13	9	10	14	17	25	25	44	22	9	18	24	28	32	33	36
Mean	8	-2	-4	-2	0	4	9	17	15	-4	-9	-11	-12	-12	-12	-11
Std Dev	30	9	10	14	19	26	37	55	38	8	14	20	22	25	27	30
Std Err Mean	3	1	1	1	2	2	3	5	4	1	1	2	2	2	3	3
Lower 95th	3	-4	-6	-5	-3	-1	2	7	8	-5	-11	-15	-16	-17	-17	-17
Upper 95th	14	-1	-2	0	4	9	15	28	22	-3	-6	-8	-8	-7	-7	-5
Code 1, 4, or 8 Only																
Number	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Minimum	-17	-8	-8	-22	-42	-57	-57	-80	-46	-20	-8	-10	-13	-18	-23	-30
10th	-16	-4	-3	-21	-40	-54	-54	-76	-34	-13	-6	-9	-12	-16	-20	-24
25th	-11	5	3	-11	-24	-35	-35	-57	-20	2	2	6	6	6	5	4
Median	-1	9	7	-1	-7	-10	-10	-21	-13	5	14	27	34	40	45	45
75th	8	18	13	6	5	-1	-1	-3	-3	9	23	36	47	55	64	75
90th	15	27	16	15	13	13	13	11	3	14	28	43	55	70	85	107
Maximum	16	31	17	20	22	22	22	20	3	15	29	46	61	77	93	115
IQR	18	13	9	17	29	35	35	54	16	6	20	30	40	49	59	71
Mean	-1	11	7	-2	-9	-16	-22	-29	-12	4	12	21	27	32	36	42
Std Dev	10	10	7	12	18	23	27	31	13	9	12	19	24	30	36	45
Std Err Mean	3	3	2	3	5	6	7	8	3	2	3	5	6	8	10	12
Lower 95th	-7	5	3	-9	-19	-29	-38	-47	-20	-1	6	11	13	14	15	16
Upper 95th	5	17	11	5	1	-3	-6	-11	-5	9	19	32	41	49	57	68

associated peak discharge was reported, or where a more recent gaging station became operational on the same stream as a long-term gaging station and significant peak flows could be estimated or entered as nonexceedance data. These interval and nonexceedance data were used in the EMA-MGB FFA but not in the B17B-GB FFA. When weighted skew was included in the FFA the median RPD decreased slightly from zero to −0.61, −3.43, and −5.72 for the 10-, 1-, and 0.2-percent AEP, respectively (fig. 8*B*, table 4). The IQR increased from 0.04 to 15.60 for the 0.2-percent AEP demonstrating the influence that skew and the MSE have on the right side of the frequency fit and the pronounced difference between methods on the less probable AEP flow estimates.

The EMA-MGB frequency curve generally fit observed historical peaks better than the fitted B17B-GB frequency curve. For example, the percent differences between the observed historical peaks, at all gages, and the estimates from the LP3 frequency fit were on average larger with B17B-GB than EMA-MGB (fig. 9, table 3). The median percent difference between the EMA-MGB fitted curve and

the historical peaks was 6.6 percent and the frequency fit was −18.8 percent for the B17B-GB curve. The 46 stations that exclusively had historical information (category 2) showed four similar patterns to those for all streamgaging stations: (1) the median RPD for all three AEPs was approximately zero, (2) the variability increased with less probable AEP estimates, (3) the IQR was uniform about the zero line, and (4) the IQR was greater than 25 percent for the 0.2-percent AEP (fig. 10*A*, table 4). The RPD-boxplot statistics indicate that FFA methods were similar for the middle half of the RPD distribution, but RPDs exceeded ±20 percent at more than a third of the streamgaging stations for the 0.2-AEP flow estimate. Two Arizona streamgaging stations with historical peak flows, Truxton Wash near Valentine (09404343) and Verde River below Tangle Creek (09508500), had large RPDs for the 0.2-percent AEP, approximately 70 and −35 percent, respectively (fig. 11). The Truxton Wash example highlights the limitations of B17B-GB to use uncertain historical information and the effectiveness of the EMA-MGB utilizing the full 96 years of historical record in which the 1904 peak discharge of 49,000 cfs was the largest peak since 1898. The frequency curves for the Verde River gaging station did not adequately fit the observed data using either estimator method, but EMA-MGB more efficiently used the paleoflood information to produce a more accurate fit when compared to B17B-GB, which overestimated the paleoflood historical peak (180,000 cfs) almost twice that of EMA-MGB (percent differences between fitted and observed historical peak flows

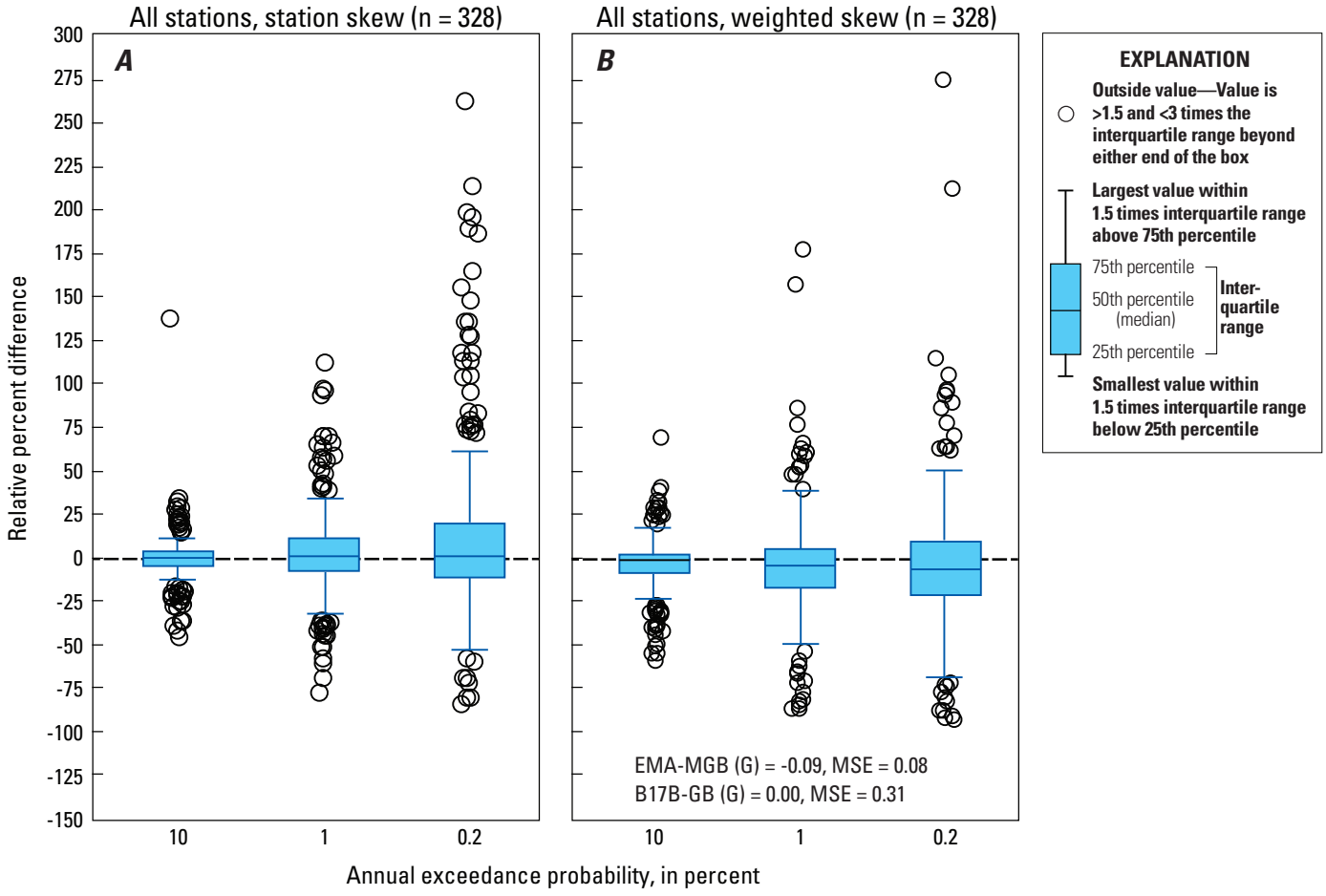


Figure 7. Boxplots of the relative percent difference (RPD) of 10-, 1-, and 0.2-percent annual exceedance probability for all stations using Bulletin 17B (B17B-GB) and Expected Moments Algorithm (EMA-MGB) with a multiple Grubbs-Beck test. *A*, Using station skew coefficient. *B*, Using weighted regional skew coefficient (G). The dashed line is plotted at an RPD of zero, signifying no difference between methods. RPDs greater than zero indicate that the EMA-MGB flow estimates are greater than B17B-GB, and the opposite is true for RPDs less than zero.

equal 69 and 36, respectively). The inclusion of a weighted skew had little effect on RPDs for the 10-percent AEP flow estimate, but the median and the 75th quartile decreased for the 1- and 0.2-percent AEP (fig. 10*B*, table 4). The negative shift in RPDs indicated that B17B-GB had greater AEP flow estimates than EMA-MGB.

There were 112 streamgaging stations where one or more PILFs were identified by the MGB-test and no historical information was present (category 3). The median RPD was −0.7 for the 10-percent AEP and 3.1 for the 0.2-percent AEP (fig. 12*A*, table 4.). The IQR increased by more than a factor of 4 from the 10-percent AEP to the 0.2-percent AEP and was related to the PILFs having high influence on the upper right-hand tail of the LP3 distribution fit, effectively causing greater RPDs between methods for the less probable AEP flow estimates. For the 0.2-percent AEP flow estimate, the 75th quartile was about 37 percent and there were 10 RPDs that exceeded 100 percent. Two Arizona streamgaging stations with multiple PILFs, Deadman Wash near New River (09513820) and East Verde River near Childs (09507980) (fig. 13) are examples showing the influence of PILFS on the right-hand tail of the frequency curve. The result was the overestimation of the less probable AEP flows by B17B-GB (station skew) for the frequency plots of these two gaging stations (RPD was 37.2 and 39.8 percent, respectively). The inclusion of a weighted regional skew caused the median RPD to become more negative, changing the mostly positive station median RPDs to −6.0, −8.2, and −9.0 for the 10-,

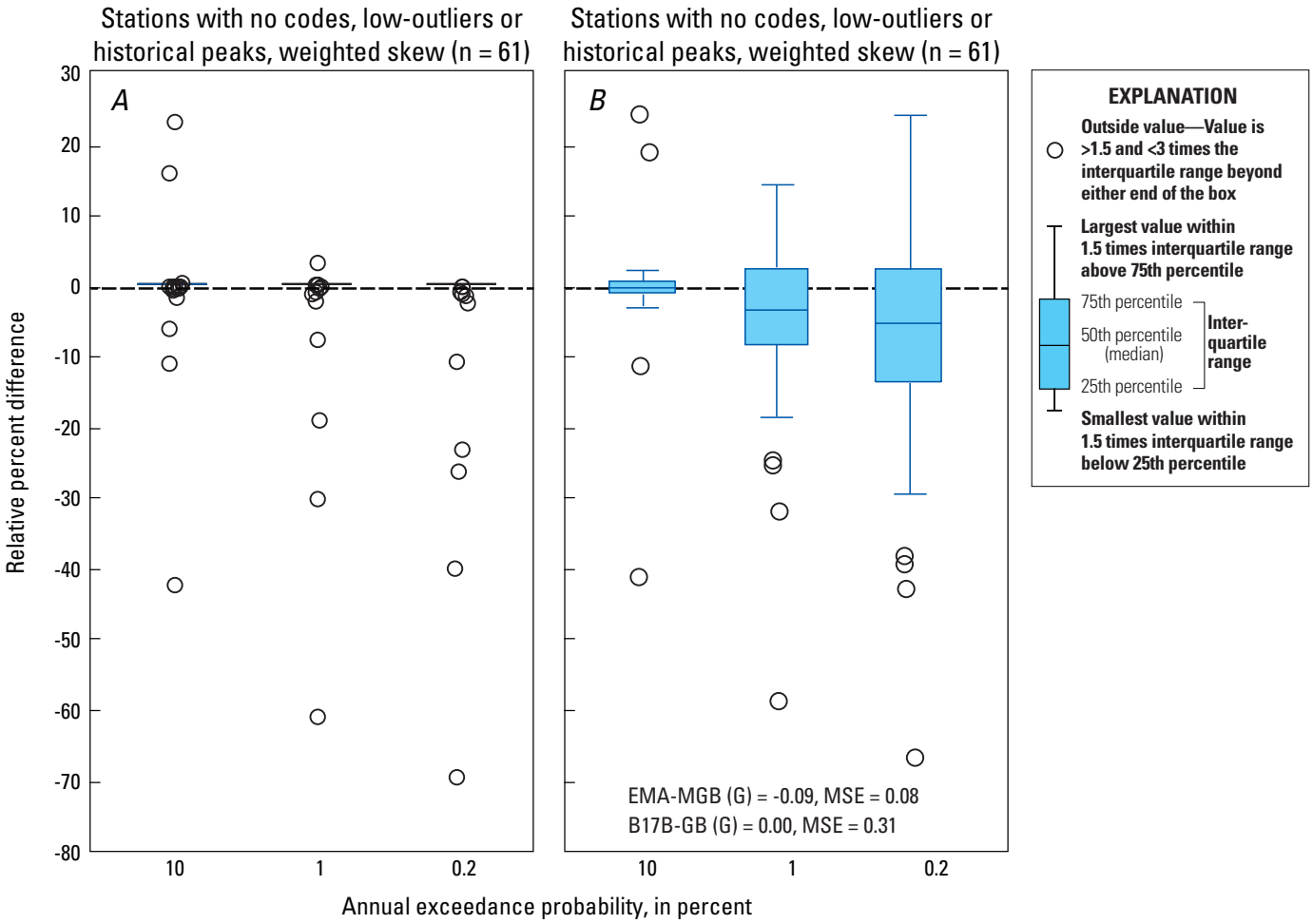


Figure 8. Boxplots for stations of category 1 showing the relative percent difference (RPD) of 10-, 1-, and 0.2-percent annual exceedance probability using Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB). *A*, Using station skew coefficient. *B*, Using weighted regional skew coefficient (G) with corresponding mean square error (MSE). The dashed line is plotted at an RPD of zero, signifying no difference between methods. RPDs greater than zero indicate that the EMA-MGB flow estimates are greater than B17B-GB, and the opposite is true for RPDs less than zero.

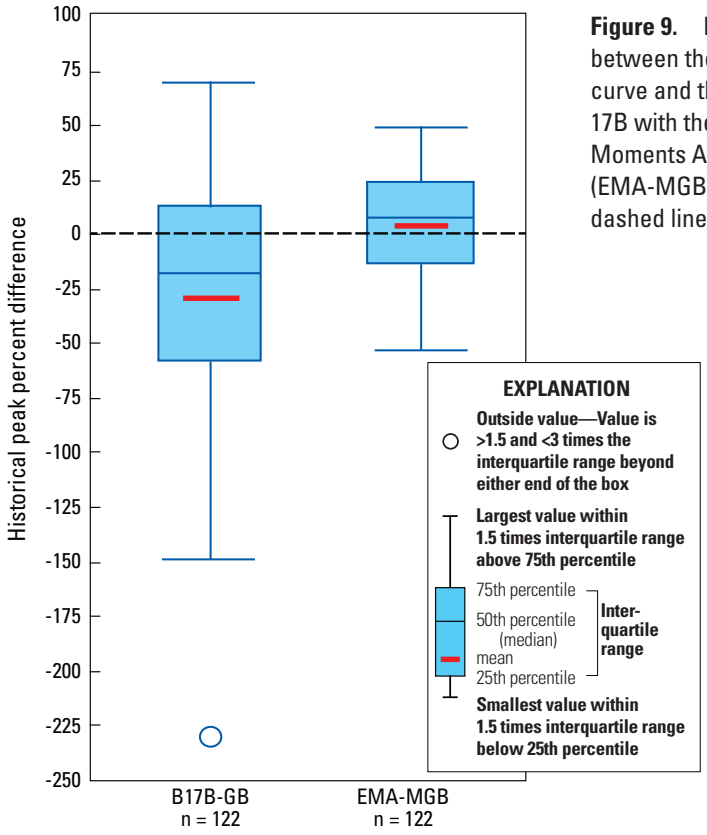
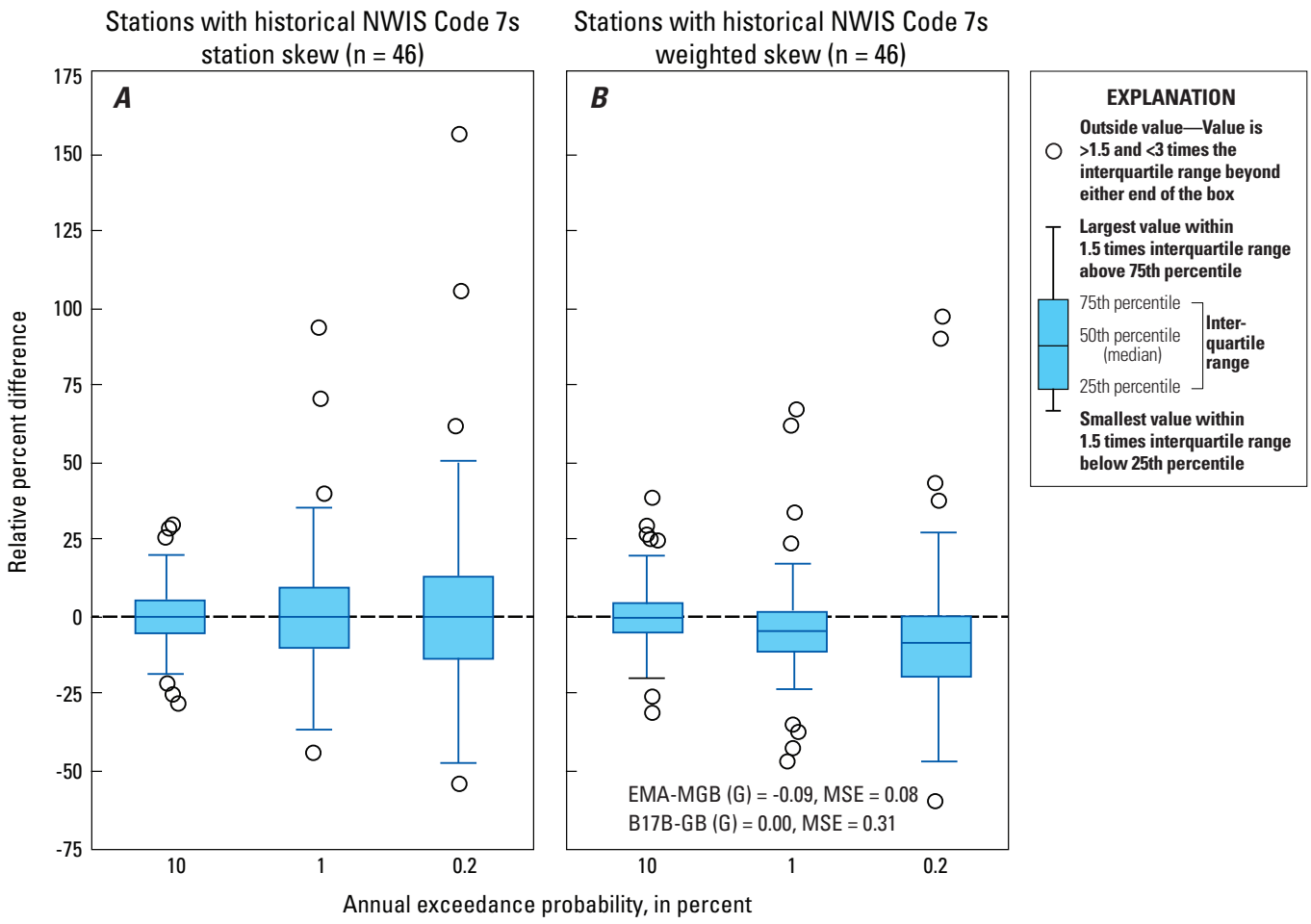


Figure 9. Boxplots (at left) of percent difference between the predicted log Pearson Type 3 frequency curve and the observed historical peak flow for Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB). The red line equals the mean, and the dashed line is plotted at zero (indicating no difference).

Figure 10. Boxplots (below) for stations of category 2 showing the relative percent difference (RPD) of 10-, 1-, and 0.2-percent annual exceedance probability using Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB). *A*, Using station skew coefficient. *B*, Using weighted regional skew coefficient (G) with corresponding mean square error (MSE). The dashed line is plotted at an RPD of zero, signifying no difference between methods. RPDs greater than zero indicate that the EMA-MGB flow estimates are greater than B17B-GB, and the opposite is true for RPDs less than zero.



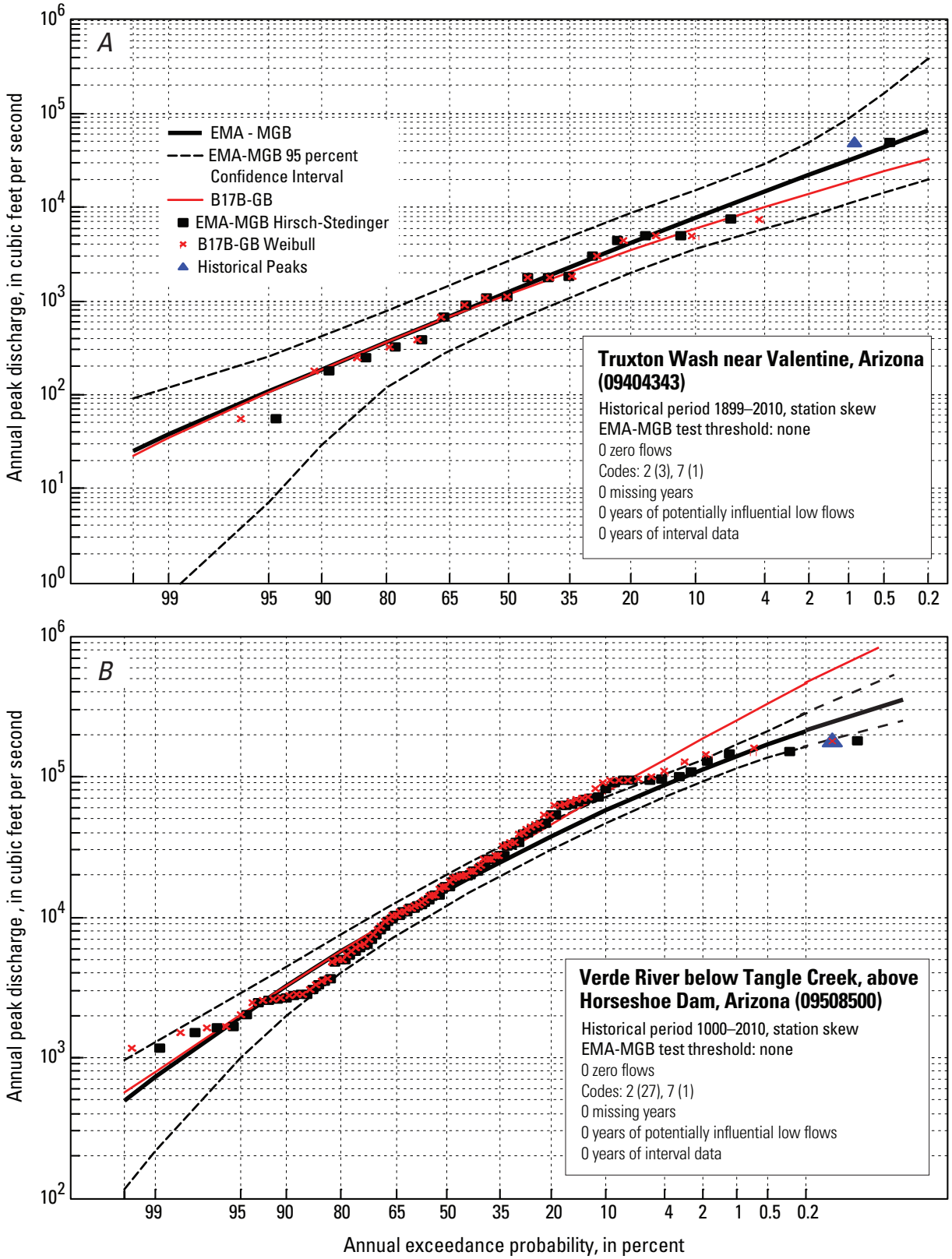


Figure 11. Annual exceedance probability curves derived using Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB) (black) and Bulletin 17B with the Grubbs-Beck test (B17B-GB) (red) for Arizona streamgaging stations (*A*) Truxton Wash near Valentine (09404343) and (*B*) Verde River Below Tangle Creek (09508500). The 1-percent annual exceedance probability relative percent difference for these two stations is 70 and −44 percent, respectively.

1-, and 0.2-percent AEP, respectively (fig. 12*B*, table 4). The IQR became more negative with the regional skew, the 75th quartile decreased to 8.2 percent, and RPDs no longer exceeded ± 100 percent. The changes observed between the station and weighted skew FFA can be attributed in part to an improvement from weighted skew in the B17B-GB method. The limited censoring of B17B-GB had a tendency to produce more pronounced skews than EMA-MGB and the frequency fits were less accurate, especially for sites with short record, meaning the addition of a data point to the distribution could dramatically change the skew. The weighting of a generalized skew helps to reduce a highly skewed distribution in order to produce more accurate flood estimates.

Peak-flow data with NWIS qualification codes of 1, 4, or 8 indicate uncertainty or censored information in the magnitude of a reported flow. These data could be viewed as a source of potential error in the flow estimate, yet if addressed appropriately, these types of information can improve flood frequency estimates. Only 14 stations were identified as having one or more codes 1, 4, or 8 and no PILFs or historical information (category 4). Though the sample size was small, this category had the greatest difference in median RPD

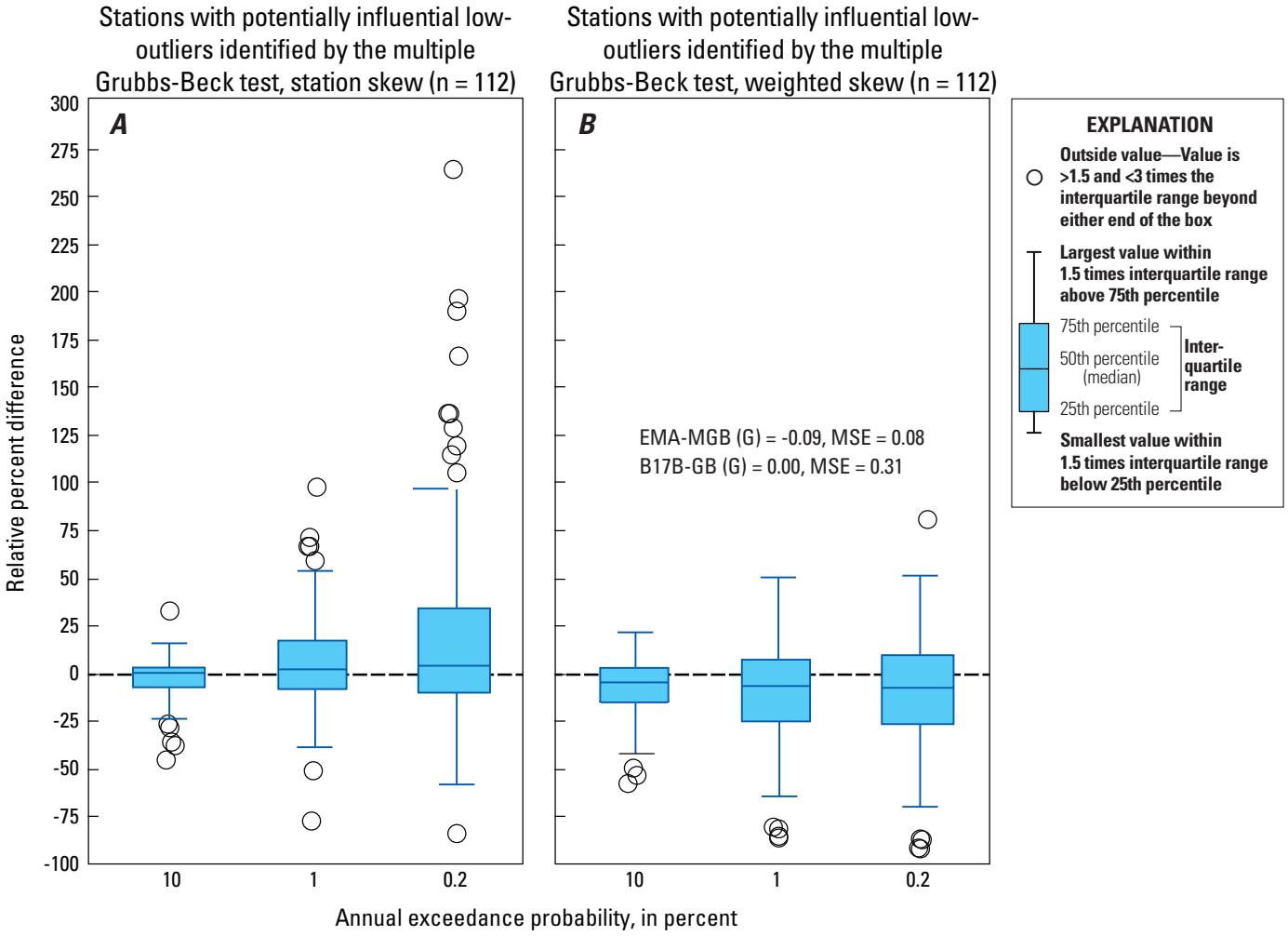


Figure 12. Boxplots for stations of category 3 showing the relative percent difference (RPD) of 10-, 1-, and 0.2-percent annual exceedance probability using Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB). *A*, Using station skew coefficient. *B*, Using weighted regional skew coefficient (G) with corresponding mean square error (MSE). The dashed line is plotted at an RPD of zero, signifying no difference between methods. RPDs greater than zero indicate that the EMA-MGB flow estimates are greater than B17B-GB, and the opposite is true for RPDs less than zero.

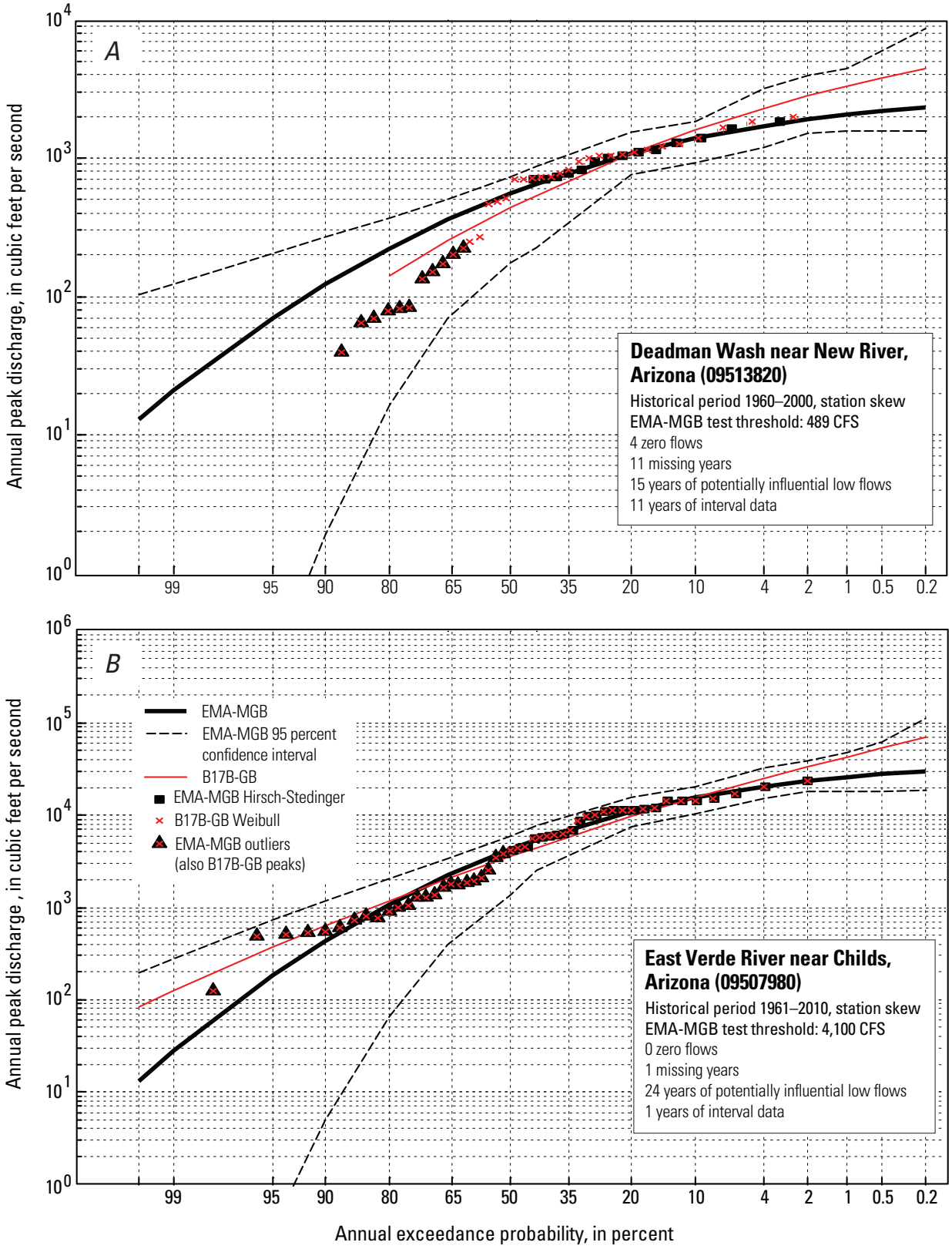


Figure 13. Annual exceedance probability curves derived using Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB) (black) and Bulletin 17B with the Grubbs-Beck test (B17B-GB) (red) for Arizona streamgaging stations (*A*) Deadman Wash near New River (09513820) and (*B*) East Verde River near Childs (09507980). The 1-percent annual exceedance probability relative percent difference for these two stations is -37 and -40 percent, respectively. CFS, cubic feet per second.

between the 10- and 0.2-percent AEP, shifting from 6.6 to −21.4, respectively (fig.14*A*, table 4). The variability was the largest among all categories, with IQRs of 35.0 and 54.3 for the 1- and 0.2-percent AEP flow estimate, respectively. Differences in RPDs can also be related to the short period of record at these stations, causing station skew to be less accurate, and therefore the use of the interval data by EMA will have a greater effect on the LP3 frequency distribution fit. This observation was supported when the estimates were weighted with regional skew and the resulting medians and IQRS for all three AEP estimates were greater than zero. The median RPD for the 1- and 0.2-percent AEP increased from −10.5 to 39.6 and −21.4 to 45.4, respectively (fig.14*B*, table 4). Of all the categories, 4 was the only one where RPDs became more positive once regional information was included.

Patterns were relatively consistent between categories, with the exception of category 1, which demonstrated that EMA-MGB and B17B-GB provide equal AEP flow estimates if no interval, historical, or PILF data are present. When comparing the other three categories, the PILF category (3) was most common for gaging stations, and the multiple identification of PILFs had the greatest influence on the less

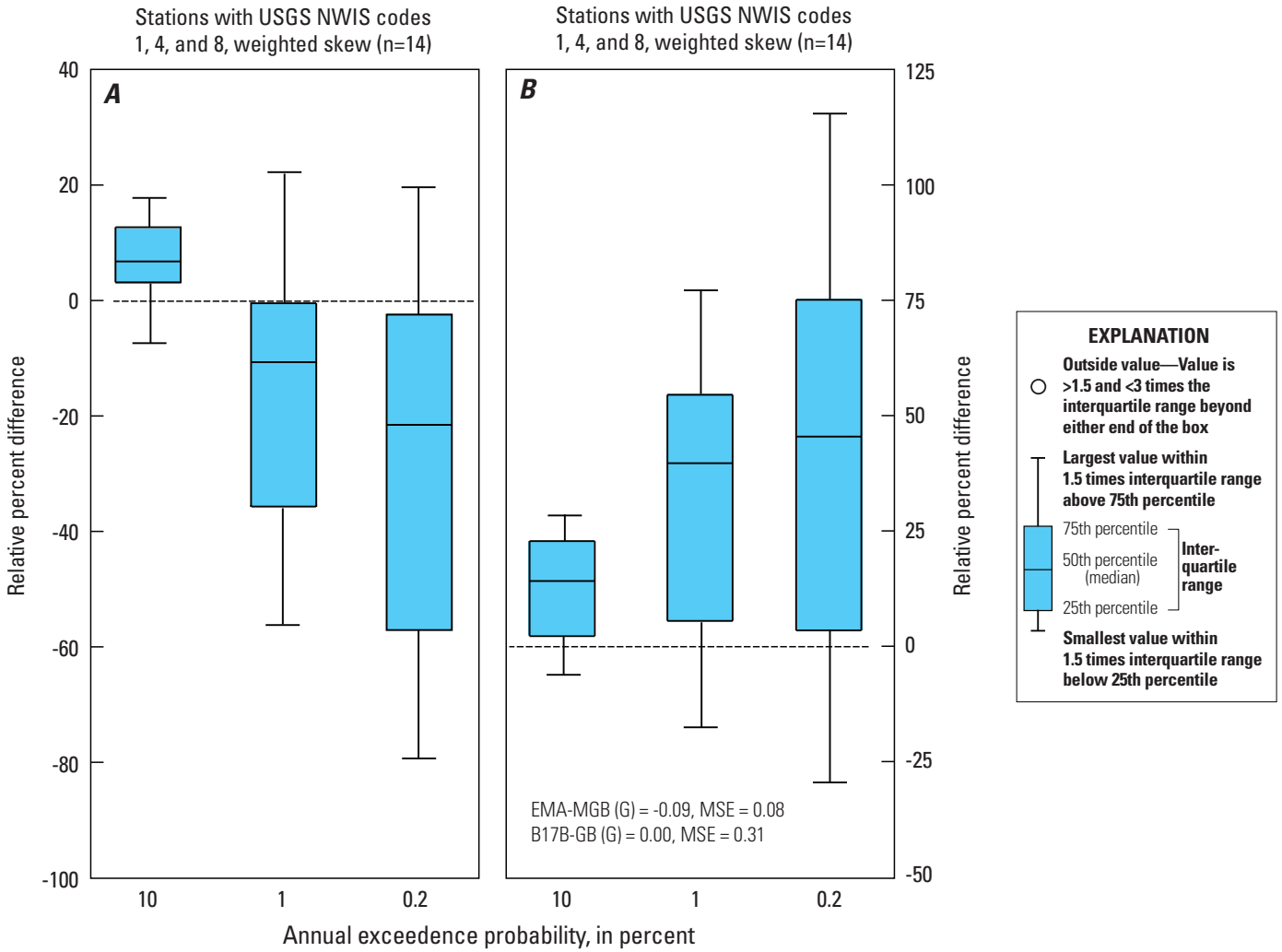


Figure 14. Boxplots for stations of category 4 showing the relative percent difference (RPD) of 10-, 1-, and 0.2-percent annual exceedance probability using Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB). *A*, Using station skew coefficient. *B*, Using weighted regional skew coefficient (G) with corresponding mean square error (MSE). The dashed line is plotted at an RPD of zero, signifying no difference between methods. RPDs greater than zero indicate that the EMA-MGB flow estimates are greater than B17B-GB, and the opposite is true for RPDs less than zero.

probable AEP estimates, as shown by the high variability and multiple RPDs exceeding 100 percent for the 0.2-percent AEP flow estimates (fig. 14*A*). Categories 2 and 4 also demonstrated through high variability and the presence of extreme RPDs that FFA was significantly affected by the way each method treated historical and other censored peak-flow data in the analysis. The weighted-skew RPD boxplots show fewer positive outliers, and categories 1, 2, and 3 all have negative medians, while category 4 was highly positive relative to the other categories.

RPD by Flood Regions

Streamgaging stations were analyzed by flood regions (Thomas and others, 1997) to understand how EMA-MGB and B17B-GB differ when flood-frequency results are analyzed by areas of similar physiographic characteristics. Regions 8 and 10 were not analyzed because sample sizes were small. There are regional differences in the qualification of certain types of peaks. Regional differences can be related to climatic conditions, streamgaging objectives (number of gages per

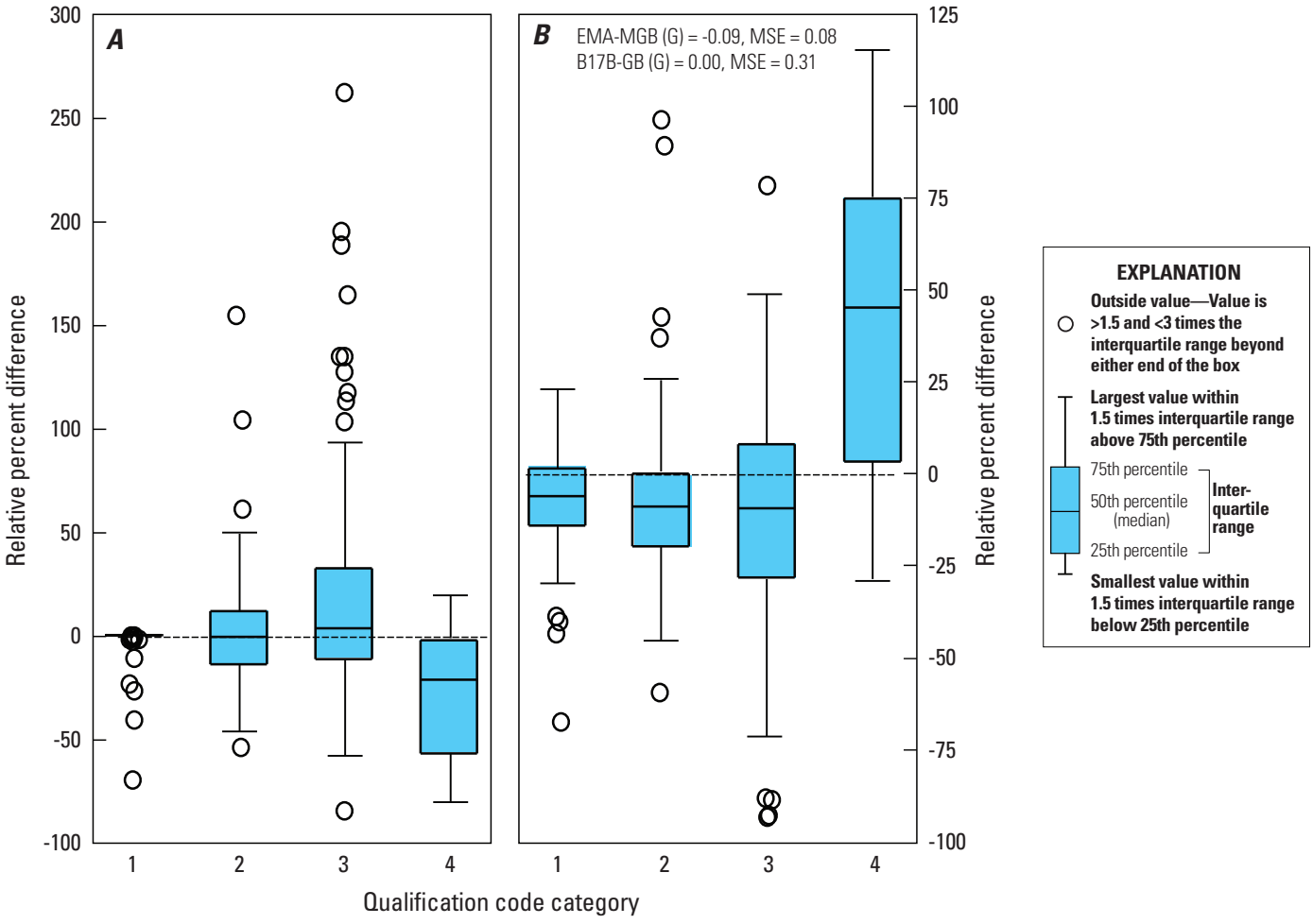


Figure 15. Boxplots for stations of all categories showing the relative percent difference (RPD) of the 0.2-percent annual exceedance probability using Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB). *A*, Using station skew coefficient. *B*, Using weighted regional skew coefficient (G) with corresponding mean square error (MSE). The dashed line is plotted at an RPD of zero, signifying no difference between methods. RPDs greater than zero indicate that the EMA-MGB flow estimates are greater than B17B-GB, and the opposite is true for RPDs less than zero.

region), or historical documentation related to flooding impacts on nearby populations. A mosaic plot (fig. 16) shows the proportion of stations in the four NWIS categories within each region. Regions 12 and 13 had a greater percentage of PILFs potentially reflecting the mixed populations within those regions (floods may originate from multiple types of precipitation events). Historical peak information within regions 14 and 11 was proportionally greater than in the other NWIS categories and is related to stations on the Gila and Little Colorado rivers that have long and well-documented periods of record. Region 13 had the greatest percentage of peaks with NWIS codes 1, 4 and 8, and this is related to the dry conditions, high infiltration rates, and intense precipitation events that occur throughout the Basin and Range Province.

In region 11 the median RPDs for the 10-, 1-, and 0.2-percent AEPs were near zero, with the greatest median equaling 3.1 percent for the 0.2-percent AEP flow station skew estimate (fig. 17*A*, table 5). The IQR for 10-percent AEP estimate was about 10 percent, and this increased 4-fold in the 0.2-percent AEP estimate along with a positive shift of roughly 36 percent in the 75th quartile of the RPD distribution. EMA-MGB and B17B-GB estimates were more similar for the larger AEP flows and several of the positive RPD outliers were more related to historical information (category 2), whereas negative RPD outliers were more related to sites with PILFs (category 3). The pattern was somewhat magnified for the smaller AEP flows, where many of the stations classified in categories 2 and 3 had greater increases in RPDs in both the positive and negative direction. The shift in the positive direction was related to the differences in flood frequency method treatment of historical information and PILFs in the frequency computation. Limited rainfall on the Colorado Plateau Province, combined with short record lengths (median systematic record length of 15 years) demonstrates the importance of the identification and treatment of PILFs, because this process will so significantly influence the frequency fit of the LP3 distribution. The inclusion of a regional skew shifted boxplots below zero for all AEP estimates, as shown by a decrease in the median RPDs and IQRs. The change was most prominent in the 0.2 AEP estimate with the median RPD changing from 3.1 to −11.7 and the 75th percentile decreasing more than 30 percent. The 1-percent AEP changed similarly, but the 10-percent AEP stayed roughly the same (fig. 17*B*, table 5). The RPDs of stations identified in category 3 (PILFs) were most affected by the incorporation of regional skew. A majority of those stations changed from greater than zero to less than zero. The weighting of a regional skew will have more influence at sites with a limited period of record. The more negative skew of the BGLS analysis with a lower MSE reduced many of the EMA-MGB lower AEP estimates, effectively shifting the boxplots more negative. Though B17B-GB was estimating greater AEP flow estimates, the graphical plots visually showed that B17B-GB may be overestimating the less probable AEP flow estimates because of the influence the PILFs had on the upper right-hand tail of the flood frequency fit.

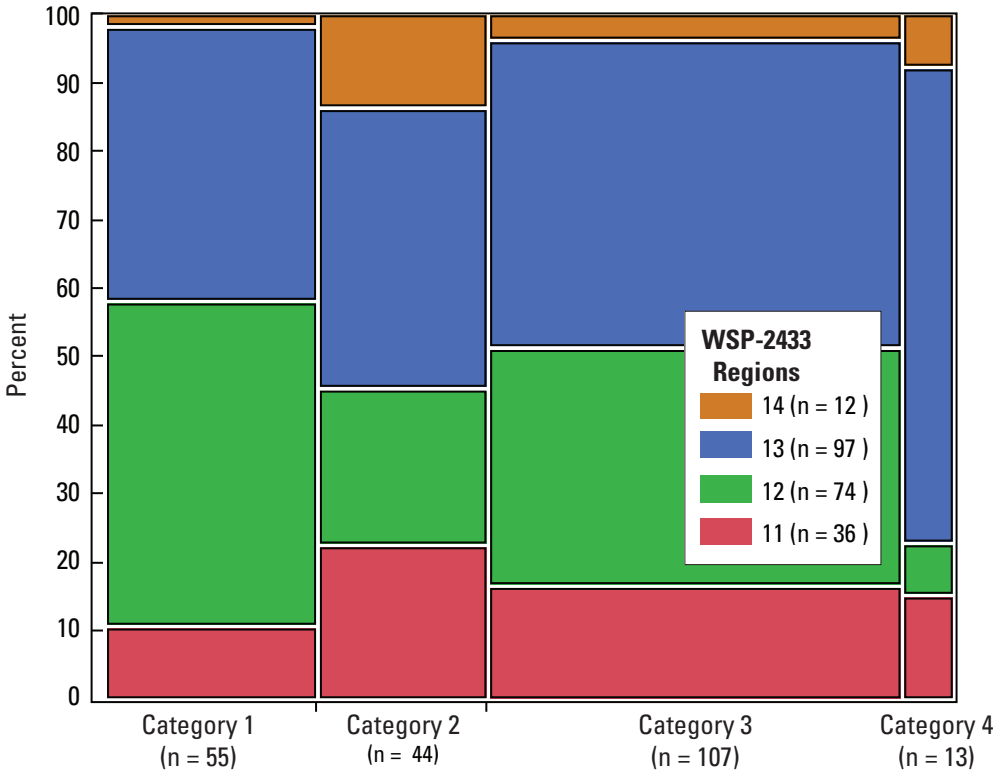


Figure 16. Bar graph showing proportion, in percent, of the Arizona streamgages within a USGS National Water Information System (NWIS) qualification codes category in the different regional flood regions (Thomas and others, 1997).

The extent of region 12 is mostly within the Transition Zone and flood populations can be mixed, meaning that peak flows within a record consist of multiple flood types generated from different precipitation events, such as snowmelt, convective rainfall, or rain-on-snow events. Peak-flow data that originate from different populations can be problematic when fitting one overall frequency curve, and the treatment of mixed datasets in FFA is less straightforward. The median RPDs were all approximately zero, and the IQRs were centered about zero for all AEP estimates. The variability increased significantly for the less probable AEP estimates, and the IQR for the 0.2-percent AEP was roughly four times the 10-percent AEP station skew estimates. The 0.2-percent AEP station skew estimates also had six RPDs greater than 100 percent, and (fig. 18*A*, table 5), as with region 11, the highly positive RPDs in the 0.2-percent AEP resulted from short streamgage records and presence of PILFs. The more negative RPDs (10th to 25th percentile) for the 0.2-percent AEP station skew estimates were mostly related to PILFS, historical peaks, or both. Stations that experience mixed flood populations may require extensive censoring of PILFs to adequately fit the LP3 distribution. The censoring of the MGB test enables EMA to fit the LP3 distribution without using a conditional probability adjustment and loses very little information in the censoring process. Because of the single outlier test employed by B17B-GB, the method is limited in fitting distributions that deviate from the LP3 distribution.

Examples of gaging stations likely experiencing mixed populations and the use of extensive censoring by the MGB

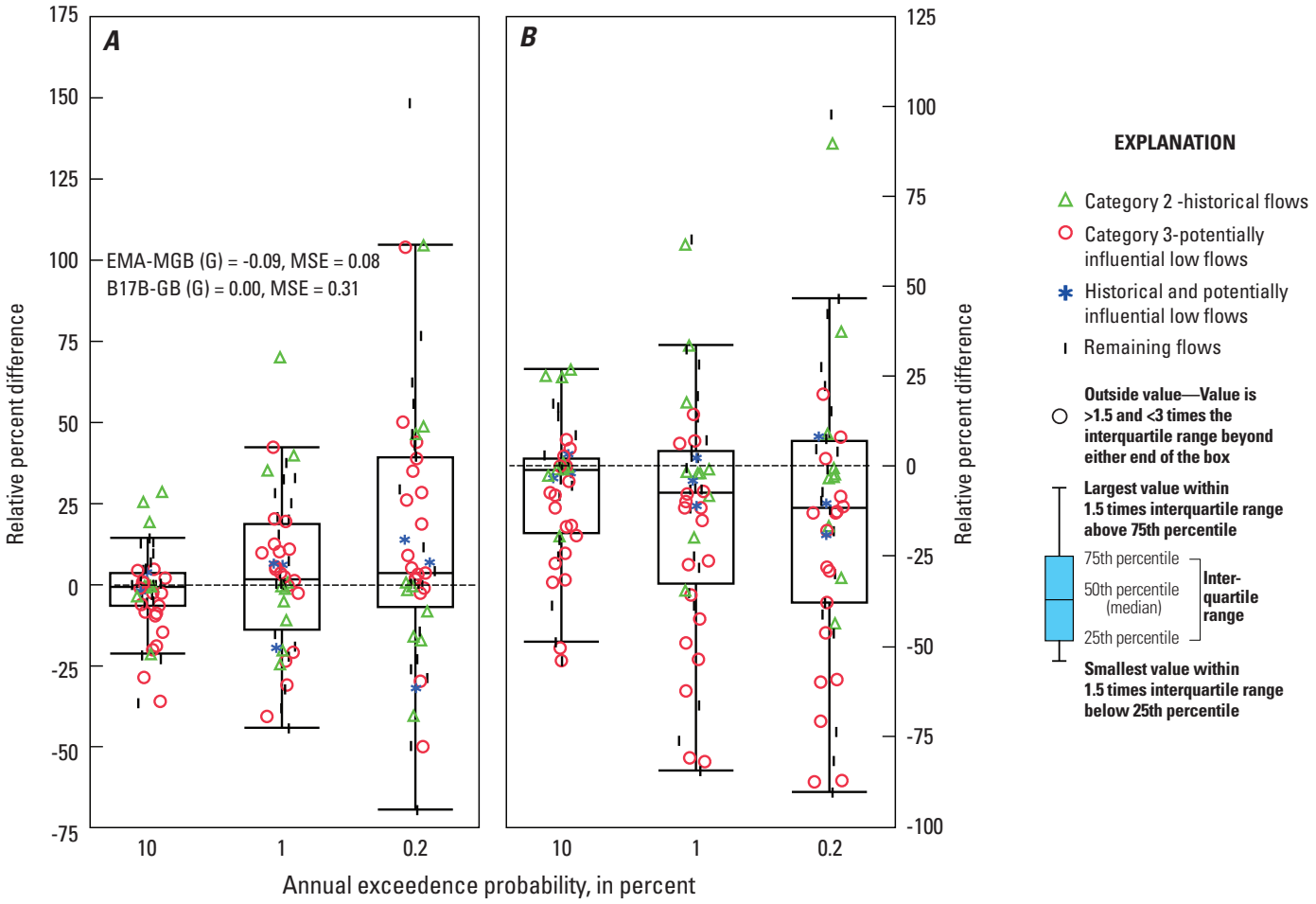


Figure 17. Boxplots for streamgaging stations in flood region 11 (Thomas and others, 1997) showing relative percent difference (RPD) of 10-, 1-, and 0.2-percent annual exceedance probability using Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB). *A*, Using station skew coefficient. *B*, Using weighted regional skew coefficient (G) with corresponding mean square error (MSE). The dashed line is plotted at an RPD of zero, signifying no difference between methods. RPDs greater than zero indicate that the EMA-MGB flow estimates are greater than B17B-Gb, and the opposite is true for RPDs less than zero.

test used by EMA to improve the flood frequency fit were observed at the three Verde River tributary gaging stations: West Clear Creek near Camp Verde (09505800), Oak Creek near Cornville (09504500), and Dry Beaver Creek near Rimrock (09505350). The three streamgages experience peak flows generated from several flood-generating storm types, and the different subpopulations influence the quality of the frequency fit. In these three examples the MGB test determined that 50, 25, and 40 percent of the data were PILFs, respectively. PILFs left unaddressed, such as those in the B17B example, influence the frequency analysis by not accurately fitting the largest observed peak flows (fig. 19). The B17B-GB AEP flow estimates were greater than EMA, but the LP3 frequency curve overestimated the observed largest flows. The inclusion of regional information shifted the median RPD below zero, meaning that B17B-GB was estimating larger AEP flow estimates than EMA-MGB for more of the streamgaging stations in region 12 and this was most pronounced for the 0.2-percent AEP flow (fig. 18*B*, table 5). Of all the regions the weighting of a regional skew changed station RPDs the least in region 12. The general pattern again showed that stations with PILFs (category 3) were most affected, and RPDs shifted below zero and became more similar between methods. Also, this was reflected in the decrease of the number of positive outliers (all positive RPDs were below 100 percent). The pooling of regional skew stabilized the station skew of streamgaging stations with shorter records and PILFs.

Region 13 is entirely contained within the Basin and Range Province and encompasses much of southern Arizona, where rainfall and climate have high intra- and interannual

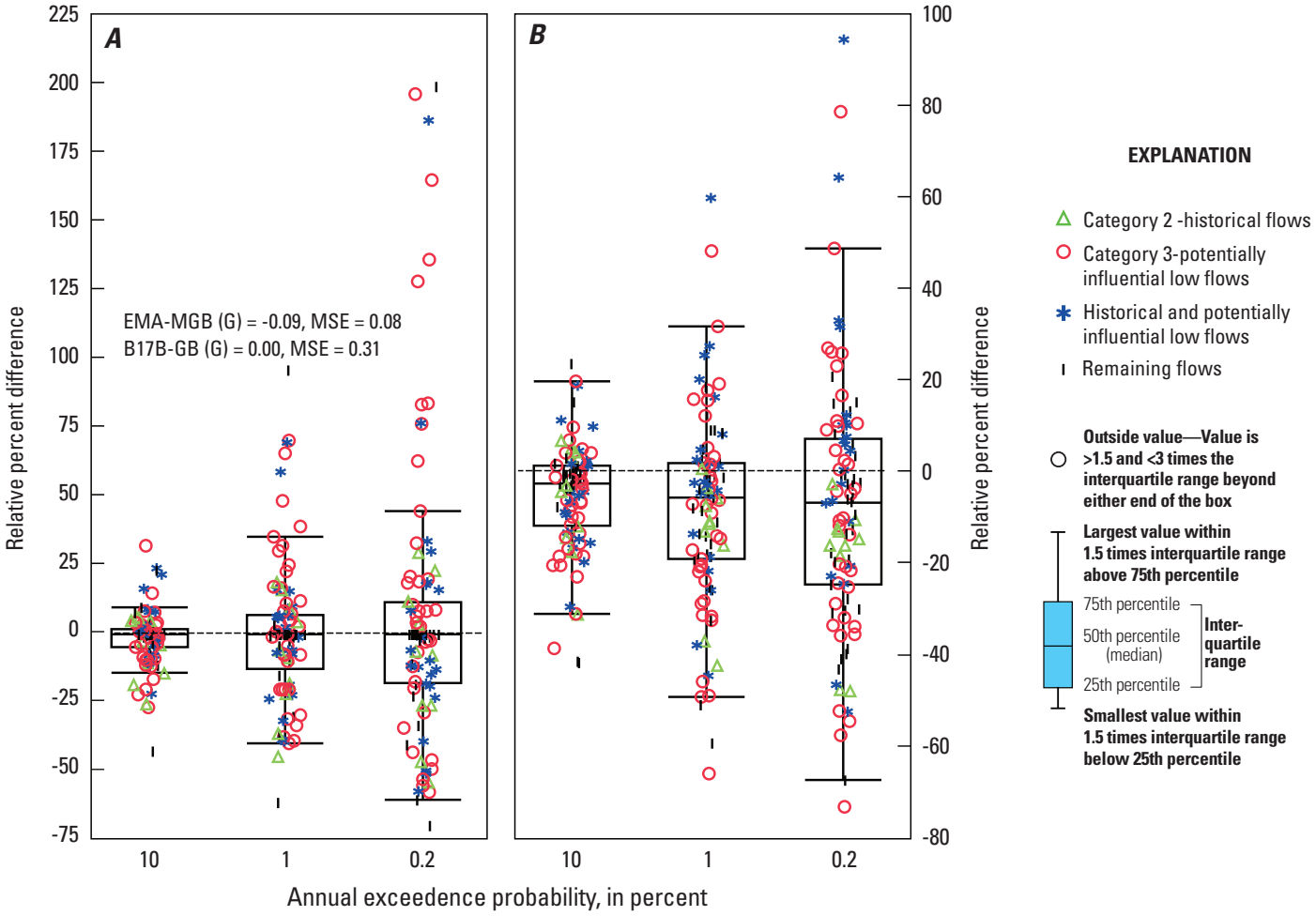


Figure 18. Boxplots for streamgaging stations in flood region 12 (Thomas and others, 1997) showing relative percent difference (RPD) of 10-, 1-, and 0.2-percent annual exceedance probability using Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB). *A*, Using station skew coefficient. *B*, Using weighted regional skew coefficient (G) with corresponding mean square error (MSE). The dashed line is plotted at an RPD of zero, signifying no difference between methods. RPDs greater than zero indicate that the EMA-MGB flow estimates are greater than B17B-GB, and the opposite is true for RPDs less than zero.

variability. This contributes to highly variable peak-discharge magnitudes in a given water year, and zero or low flows are often recorded as the annual maximum peak flows. Region 13 proportionally had the greatest number of streamgaging stations with low outliers, historical peaks, or both (fig. 16). Patterns were very similar to region 12 for the station skew RPD boxplots (fig. 20*A*, table 5). RPD boxplots for all three AEPs were symmetric about zero, IQRs ranged from 6.6 to 21.7 for the 10- and 0.2-percent AEP estimate, respectively, and 6 RPDs exceeded 100 percent. Twenty-eight streamgaging stations had both historical and low-outlier peak-flow data. The combined use of the conditional probability adjustment (CPA) and historical weighting adjustment used by the B17B method when multiple flood data-types are present caused the method to less accurately fit the largest observed floods. The Santa Cruz River near Lochiel (09480000) and Crater Range Wash Near Ajo (09520230) are two gaging station examples where the combination of PILFs and a historical adjustment affected the B17B-GB method's ability to adequately fit the largest peaks, and the RPDs for two streamgages were roughly −47 and 188 percent, respectively, for the 0.2-percent flow estimate (fig. 21). Region 13 was the only region where the weighted skew analysis reduced the spread or IQR of the RPD boxplots for the 10- and 1-percent AEP estimate and the medians all still remained close to zero (fig. 20*B*, table 5). Region 13 skews were among the most negative of all the regions, and as a result the regional skew weighting had less of an impact on the overall flow estimates. That is, weighting a regional skew similar to a station skew will change the AEPs less than if the station and regional were very different.

Region 14 is small in comparison to the other regions and has only 19 streamgaging stations (fig. 1). Most of region 14 is contained within the Transition Zone and characterized by small mountainous watersheds that

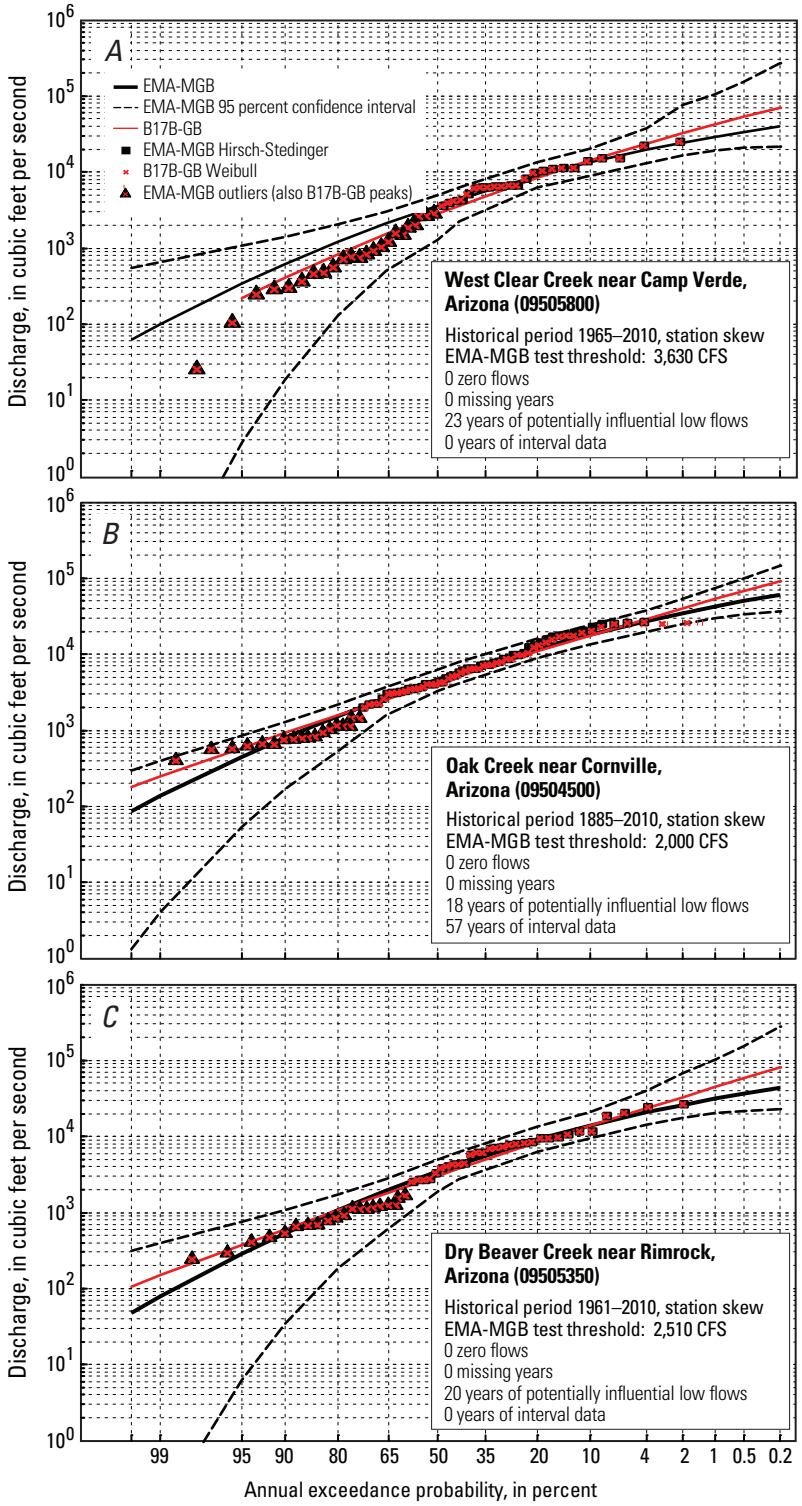


Figure 19. Annual exceedance probability curves derived using Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB) (black) and Bulletin 17B with the Grubbs-Beck test (B17B-GB) (red) for Arizona streamgaging stations (*A*) West Clear Creek near Camp Verde (09505800), (*B*) Oak Creek near Cornville (09504500), and (*C*) Dry Beaver Creek near Rimrock (09505350). The 1-percent annual exceedance probability relative percent difference (RPD) for these three stations is -31, -20, and -30 percent, respectively. CFS, cubic feet per second.

drain the White Mountains and headwaters of the upper Gila River watershed. Low outliers and zero annual peak flows were less common in region 14, and on average skews were more positive when compared to other regions. Of all the regions, the IQR station-skew RPD boxplots for region 14 were most shifted in the negative direction, with the 75th percentile less than 6 percent for all AEP flow estimates and the 25th percentile being −7.1, −20.7, and −28.4 for the 10-, 1-, and 0.2-percent AEP estimates, respectively (fig. 22*A*, table 5). The Gila River at Safford (09458500) and Bonita Creek near Morenci (09447800) are two streamgaging stations with some of the more extreme RPDs and two flood frequency examples that showed the effects of each method's treatment of PILFs, historical information, interval data in the form of NWIS code 1, and interval data as uncertain estimates from nearby gaging stations (fig. 23). EMA-MGB was able to incorporate this additional information at the Gila River at Safford location, which the B17B-GB was not, because EMA-MGB utilized the full extent of the historical information by using peak-flow information from nearby gages in the form of interval data from upstream and downstream streamgaging stations. Without the additional flood information from nearby gages the LP3 frequency fit would have been a less accurate fit to the observed data, which was shown by the B17B-GB frequency fit. The Bonita Creek example reiterates the influence that PILFs can have on the right-hand tail of

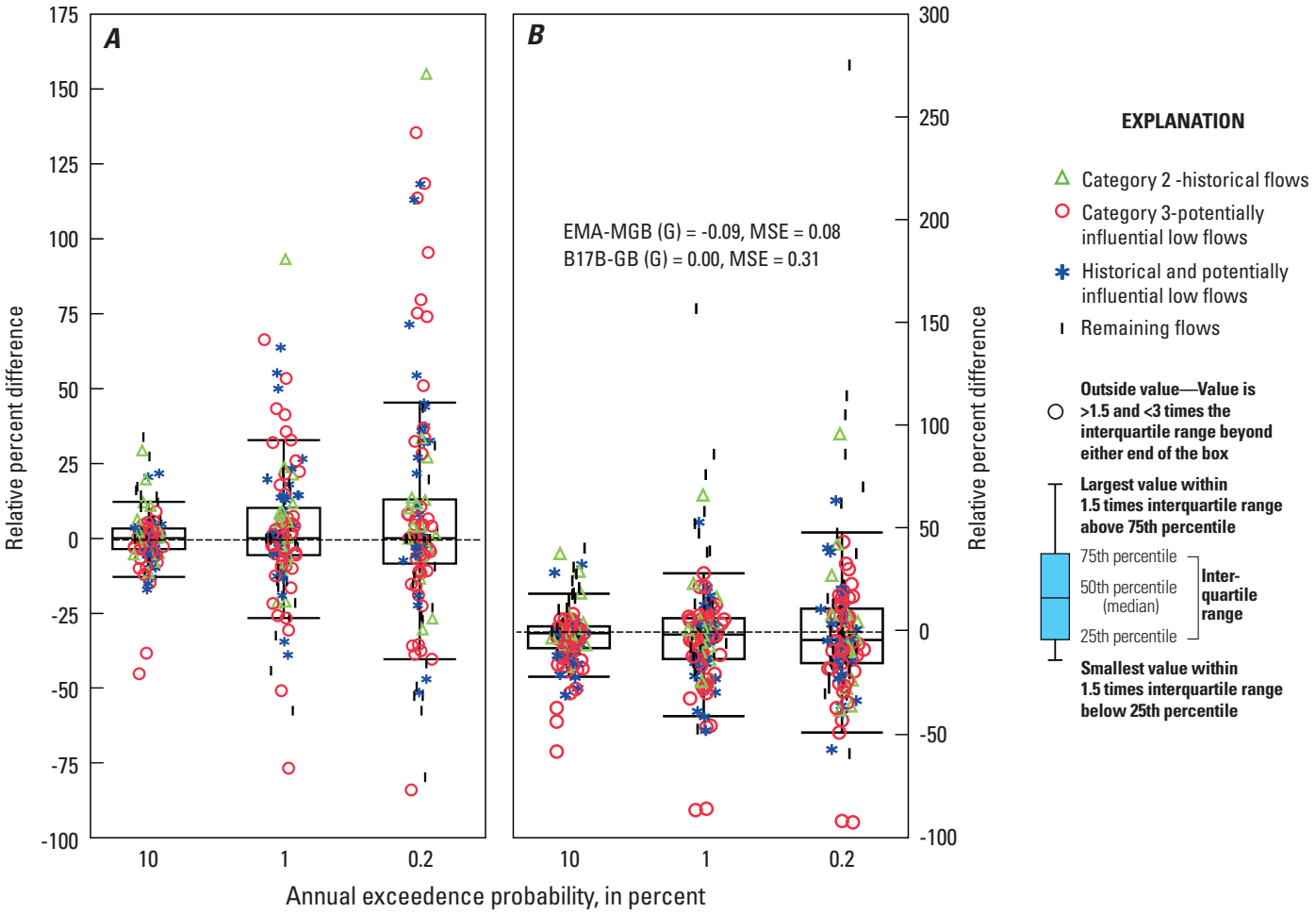


Figure 20. Boxplots for streamgaging stations in flood region 13 (Thomas and others, 1997) showing relative percent difference (RPD) of 10-, 1-, and 0.2-percent annual exceedance probability using Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB). *A*, Using station skew coefficient. *B*, Using weighted regional skew coefficient (G) with corresponding mean square error (MSE). The dashed line is plotted at an RPD of zero, signifying no difference between methods. RPDs greater than zero indicate that the EMA-MGB flow estimates are greater than B17B-GB, and the opposite is true for RPDs less than zero.

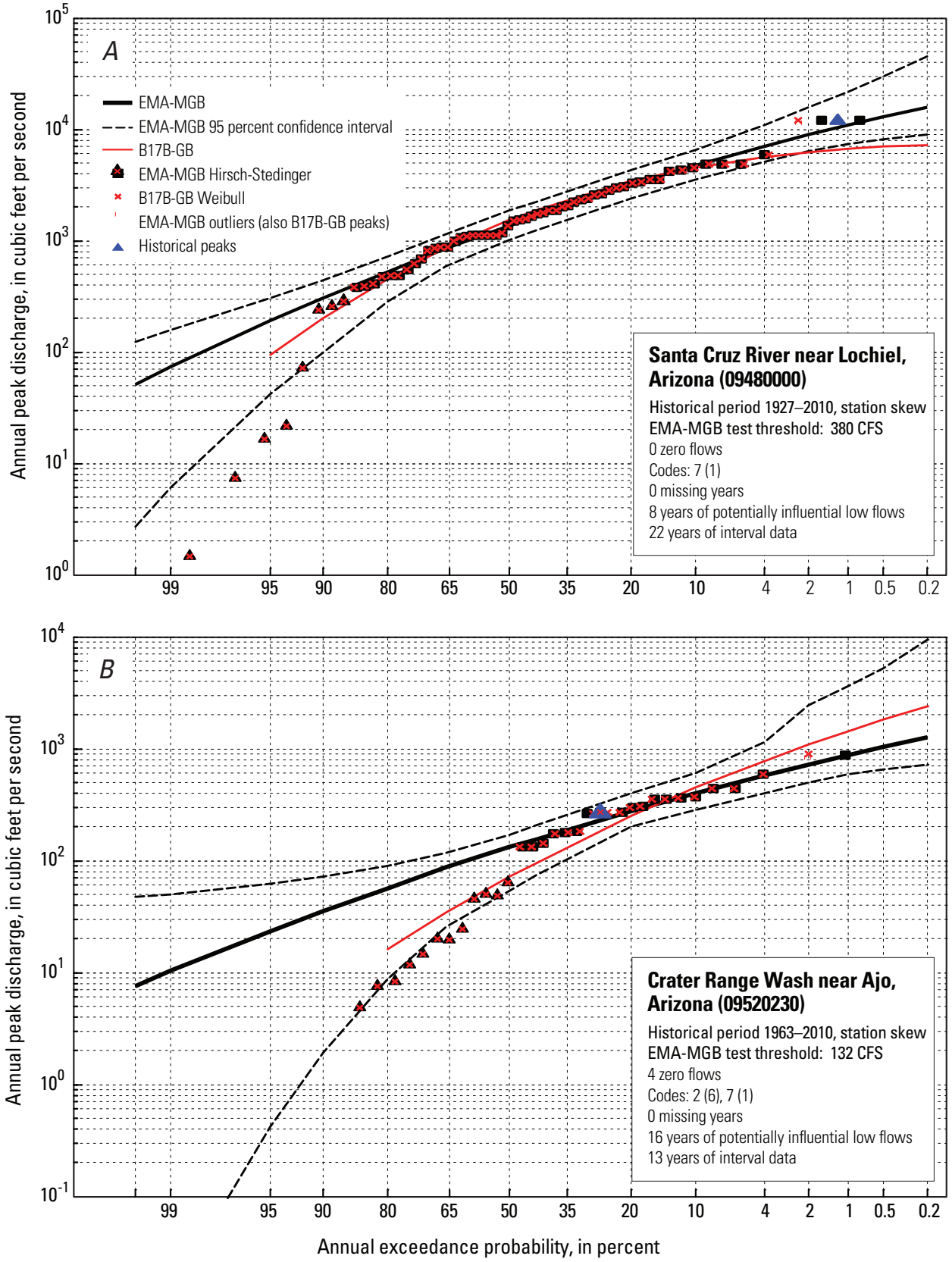


Figure 21. Annual exceedance probability curves derived using Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB) (black) and Bulletin 17B with the Grubbs-Beck test (B17B-GB) (red) for Arizona streamgaging stations (*A*) Santa Cruz River near Lochiel (09480000) and (*B*) Crater Range Wash near Ajo (09520230). The 1-percent annual exceedance probability relative percent difference (RPD) for these two stations is 64 and −39 percent, respectively. CFS, cubic feet per second.

the flood frequency curve. The Bonita Creek B17B-GB fit was noticeably influenced by four PILFs, and the influence of the peaks on the B17B-GB frequency fit resulted in an underestimation of the two largest peak flows. When regional information was included, the median RPDs decreased roughly 10 percent for the 1- and 0.2-percent AEP and the IQR increased from 25.6 and 33.7, to 48.7 and 71.4, respectively (fig. 22*B*, table 6). Of all the regions, the inclusion of a regional skew had the greatest effect on region 14. The increased variability and negative shift in boxplots was in part related to the overall more positive station skews weighted with the negative EMA-MGB regional skew and a more positive B17B-GB regional skew, resulting in even greater differences in the AEP flow estimates.

The differences in methods were greatest for the 0.2-percent AEP estimate because of sensitivity of predicting estimates for the right-hand tail of the LP3 distribution, and the differences in method procedures for fitting the distribution will be most pronounced in the tail or less frequent AEP estimates. To better understand differences between regions, the 0.2-percent AEP estimates were plotted together for the station-skew and weighted-skew estimates (fig. 24*A*, *B*). Of all the regions analyzed with station skew, region 11 was the most variable, meaning that methods of flood estimates were most different in this region. This was related to short peak-flow records, number of PILFs, and treatment of PILFs. EMA-MGB reduced the influence of the PILFs, but B17B-GB could

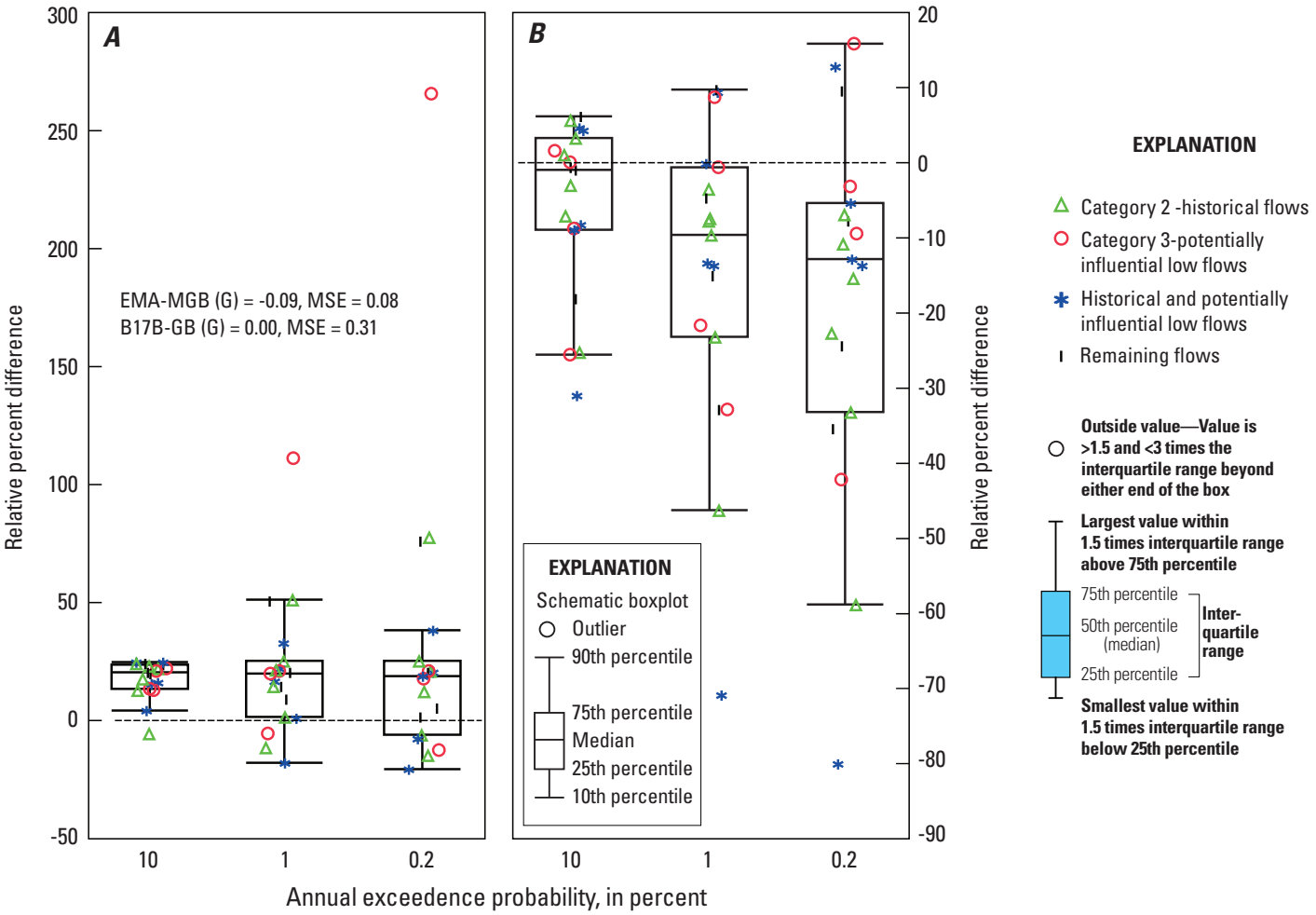


Figure 22. Boxplots for streamgaging stations in flood region 14 (Thomas and others, 1997) showing relative percent difference (RPD) of 10-, 1-, and 0.2-percent annual exceedance probability using Bulletin 17B with the Grubbs-Beck test (B17B) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB). *A*, Using station skew coefficient. *B*, Using weighted regional skew coefficient (G) with corresponding mean square error (MSE). The dashed line is plotted at an RPD of zero, signifying no difference between methods. RPDs greater than zero indicate that the EMA-MGB flow estimates are greater than B17B-GB, and the opposite is true for RPDs less than zero.

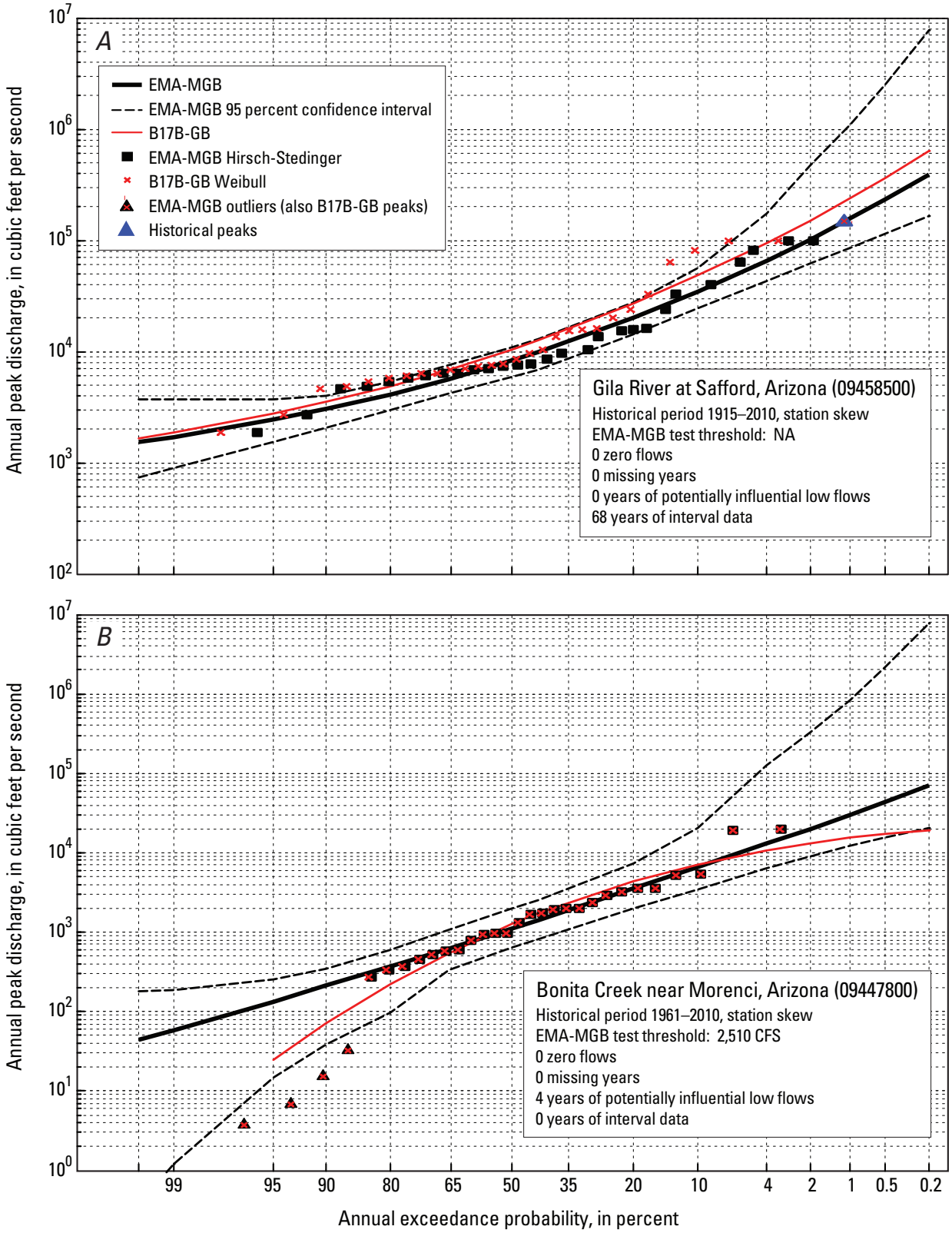


Figure 23. Annual exceedance probability curves derived using Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB) (black) and Bulletin 17B with the Grubbs-Beck test (B17B-GB) (red) for Arizona streamgaging stations (*A*) Gila River at Safford (09458500) and (*B*) Bonita Creek near Morenci (09447800). The 1-percent annual exceedance probability relative percent difference (RPD) for these two stations is −38 and 263 percent, respectively. CFS, cubic feet per second.

only address a single outlier, resulting in large RPDs for less probable AEP flows. Regions 12 and 13 were similar in that a large proportion of gaging stations had low outliers, historical peaks, or both. In these regions the middle 50 percent of the RPD distribution were similar between methods, but the remaining 50 percent of station RPDs were very different, with multiple RPDs exceeding 100 percent. Large differences were mostly due to the treatment of PILFs and use of historical information. Region 14 RPDs were mostly negative, indicating greater AEP flow estimates using B17B-GB, but in several instances these were related to the overestimation by the frequency fit of B17B-GB to the measured peak flows, usually a result of multiple PILFs. Weighting the FFA with a regional skew estimate shifted all 0.2-percent AEP RPDs less than zero as a result of the differences in regional skew and MSE. Overall differences and variability were reduced slightly, and IQRs were within the range from 15 to −40 percent.

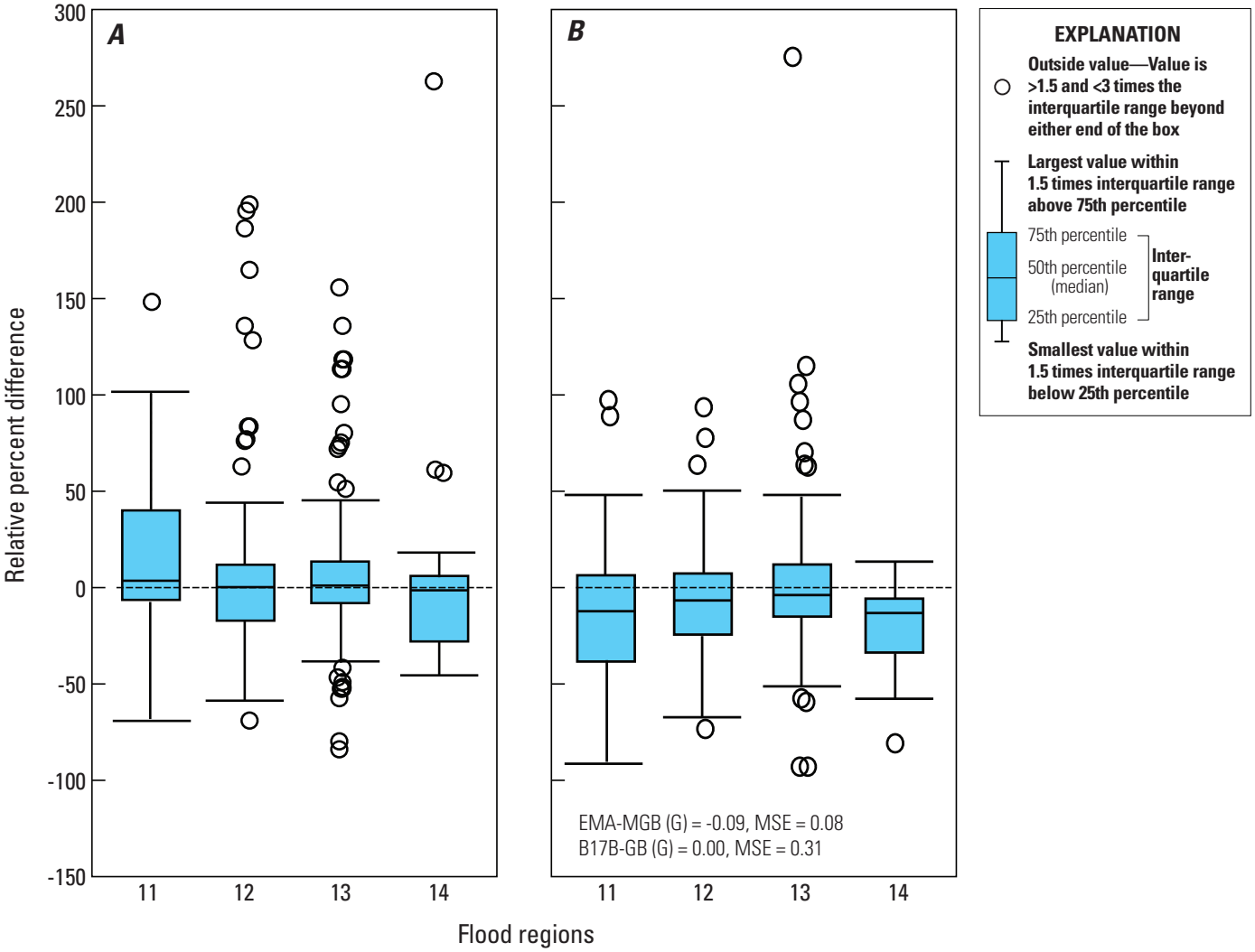


Figure 24. Boxplots for stations in flood regions 11, 12, 13, and 14 (Thomas and others, 1997) showing the relative percent difference (RPD) of the 0.2-percent annual exceedance probability using Bulletin 17B with the Grubbs-Beck test (B17B) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB). *A*, Using station skew coefficient. *B*, Using weighted regional skew coefficient (G) with corresponding mean square error (MSE). The dashed line is plotted at an RPD of zero, signifying no difference between methods. RPDs greater than zero indicate that the EMA-MGB flow estimates are greater than B17B-GB, and the opposite is true for RPDs less than zero.

Table 5. Summary statistics for the relative percent difference between Bulletin 17B and the Expected Moments Algorithm using a station skew and a weighted regional skew, by flood region.

[The 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2- annual exceedance probability (AEP) statistics are grouped by flood regions defined in U.S Geological Survey Water-Supply Paper 2433 (Thomas and others, 1997). Number, number of stations; IQR, interquartile range; Std Dev, standard deviation; Std Err Mean, standard error of the mean; negative numbers indicate that Bulletin 17B Grubbs-Beck flow estimates are greater than Expected Moments Algorithm multiple Grubbs-Beck flow estimates; positive number indicate the opposite]

	WSP-2433 regional flood regions-Station Skew								WSP-2433 regional flood regions-Weighted Skew							
	Percent								Percent							
	50	20	10	4	2	1	0.5	0.2	50	20	10	4	2	1	0.5	0.2
							Region 8									
Number	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
Minimum	-24	-18	-14	-9	-12	-14	-21	-53	-42	-9	-20	-28	-32	-34	-36	-37
10th	-20	-13	-11	-7	-9	-10	-19	-42	-32	-7	-15	-23	-27	-30	-32	-34
25th	0	0	-1	-1	-1	-1	-3	-3	-1	-1	-1	-4	-7	-9	-12	-15
Median	0	0	0	0	0	0	0	0	1	0	-1	-3	-4	-5	-7	-9
75th	5	1	4	12	25	25	27	34	8	2	3	7	10	12	15	18
90th	16	80	100	78	48	51	86	150	22	18	51	89	114	134	150	165
Maximum	17	111	138	102	57	58	103	189	25	24	69	121	153	178	197	213
IQR	5	1	5	13	26	26	29	37	9	3	4	11	17	22	27	33
Mean	0	8	11	11	10	10	12	18	1	2	4	8	10	12	14	15
Std Dev	10	33	40	30	20	21	34	61	16	8	22	37	47	55	61	67
Std Err Mean	3	9	12	9	6	6	10	17	5	2	6	11	14	16	18	19
Lower 95th	-7	-12	-14	-8	-3	-3	-10	-20	-10	-3	-9	-16	-20	-23	-25	-27
Upper 95th	7	29	37	30	22	23	34	57	11	7	18	31	40	47	52	58
							Region 10									
Number	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
Minimum	-22	-7	-25	-49	-61	-69	-74	-80	-13	-16	-40	-58	-65	-70	-74	-77
10th	-21	-7	-23	-46	-58	-67	-73	-79	-13	-15	-37	-53	-60	-65	-68	-71
25th	-11	-4	-1	-6	-10	-14	-17	-20	-12	-6	-11	-11	-11	-12	-12	-12
Median	0	0	3	4	7	7	5	4	-4	0	1	1	2	6	10	12
75th	11	5	9	19	32	44	58	86	16	1	5	12	18	24	30	38
90th	83	21	13	44	73	107	146	205	100	12	25	37	44	50	55	60
Maximum	91	23	13	46	77	112	152	214	109	13	27	39	47	53	57	62
IQR	21	9	10	25	42	58	74	106	28	7	16	23	29	36	42	50
Mean	7	2	1	3	7	13	20	33	9	-2	-3	-2	0	2	4	8
Std Dev	32	8	11	25	38	52	67	89	37	8	17	25	29	33	36	40
Std Err Mean	10	3	3	8	12	16	21	28	12	3	5	8	9	10	11	13
Lower 95th	-16	-4	-7	-15	-20	-24	-27	-31	-17	-7	-15	-19	-21	-21	-21	-21
Upper 95th	30	8	9	22	34	50	68	97	36	4	10	16	21	26	30	36
							Region 11									
Number	52	52	52	52	52	52	52	52	52	52	52	52	52	52	52	52
Minimum	-16	-25	-36	-42	-41	-44	-57	-69	-46	-25	-54	-73	-80	-85	-88	-90
10th	-3	-17	-22	-27	-29	-29	-32	-32	-5	-18	-37	-53	-60	-65	-69	-73
25th	-1	-8	-7	-9	-9	-14	-12	-7	-1	-9	-19	-25	-29	-33	-37	-38
Median	1	0	-1	0	1	1	1	3	4	-1	-1	-3	-5	-8	-9	-12
75th	10	2	3	4	8	18	25	39	21	2	2	1	3	4	5	7
90th	51	13	13	13	21	34	43	60	82	8	16	19	21	26	31	35
Maximum	95	37	28	44	56	70	84	148	166	39	27	40	50	63	77	98
IQR	11	9	10	13	17	32	38	46	21	11	21	26	32	37	42	45
Mean	11	-1	-3	-2	0	3	7	13	20	-2	-7	-11	-12	-13	-14	-14
Std Dev	26	11	13	15	19	23	29	41	42	12	19	26	30	33	36	40
Std Err Mean	4	2	2	2	3	3	4	6	6	2	3	4	4	5	5	5
Lower 95th	4	-5	-7	-6	-5	-4	-2	2	8	-5	-13	-18	-20	-22	-23	-25
Upper 95th	19	2	1	2	5	9	15	25	32	1	-2	-4	-4	-4	-4	-3

Table 5. Summary statistics for the relative percent difference between Bulletin 17B and the Expected Moments Algorithm using a station skew and a weighted regional skew, by flood region.—Continued

	WSP-2433 regional flood regions-Station Skew								WSP-2433 regional flood regions-Weighted Skew							
	Percent								Percent							
	50	20	10	4	2	1	0.5	0.2	50	20	10	4	2	1	0.5	0.2
							Region 12									
Number	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
Minimum	-28	-34	-42	-51	-56	-61	-65	-69	-29	-34	-41	-54	-61	-66	-69	-73
10th	-9	-13	-12	-17	-24	-33	-40	-48	-10	-13	-19	-27	-31	-36	-41	-42
25th	-3	-3	-5	-10	-12	-13	-14	-18	-2	-5	-11	-13	-16	-19	-22	-24
Median	0	0	0	0	0	0	0	0	1	0	-2	-4	-5	-6	-6	-7
75th	9	3	2	3	4	7	10	11	9	1	2	2	2	2	4	7
90th	28	9	8	13	18	29	43	74	33	5	5	10	14	16	19	23
Maximum	58	37	32	39	62	96	136	199	122	27	24	31	45	60	75	95
IQR	12	7	7	13	17	19	24	29	11	6	13	15	18	21	27	32
Mean	4	0	-1	-2	-2	0	2	6	7	-2	-5	-7	-8	-8	-8	-8
Std Dev	15	10	10	13	18	25	34	51	21	9	12	15	18	21	24	27
Std Err Mean	1	1	1	1	2	2	3	5	2	1	1	2	2	2	2	3
Lower 95th	1	-2	-3	-5	-5	-5	-5	-4	3	-4	-7	-10	-11	-12	-13	-13
Upper 95th	6	2	1	1	2	5	9	16	11	-1	-3	-4	-4	-4	-4	-2
							Region 13									
Number	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
Minimum	-17	-29	-46	-62	-71	-77	-81	-84	-36	-27	-58	-76	-82	-87	-90	-93
10th	-6	-8	-10	-14	-16	-22	-28	-36	-7	-10	-17	-23	-26	-26	-29	-32
25th	-3	-3	-4	-3	-5	-6	-7	-9	-3	-5	-8	-10	-12	-13	-14	-15
Median	0	0	0	0	0	0	0	0	1	0	-1	-1	-1	-1	-2	-4
75th	5	4	3	5	7	10	10	13	8	2	3	5	5	6	8	12
90th	14	11	11	12	18	25	31	48	21	8	14	15	18	22	28	39
Maximum	259	38	34	52	71	93	118	155	297	45	41	77	115	158	205	276
IQR	8	7	7	8	11	16	17	22	11	6	11	15	17	19	23	27
Mean	5	1	0	0	1	2	4	6	7	-1	-2	-2	-2	-1	0	2
Std Dev	26	9	10	13	17	22	29	38	32	9	14	20	24	28	32	39
Std Err Mean	2	1	1	1	1	2	2	3	3	1	1	2	2	2	3	3
Lower 95th	0	-1	-1	-2	-2	-2	-1	0	2	-2	-4	-5	-6	-6	-5	-5
Upper 95th	9	2	2	2	4	6	9	13	12	1	1	1	3	4	6	8
							Region 14									
Number	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19
Minimum	-20	-25	-28	-33	-38	-42	-43	-45	-17	-20	-31	-55	-64	-71	-76	-80
10th	-11	-20	-18	-31	-33	-34	-36	-38	-5	-20	-26	-34	-40	-46	-52	-59
25th	-1	-2	-8	-12	-17	-21	-24	-28	0	-4	-9	-21	-25	-23	-28	-33
Median	0	0	0	1	0	-1	-2	-2	2	1	-1	-5	-6	-10	-14	-13
75th	8	3	4	4	5	5	5	5	9	5	3	1	0	-1	-2	-5
90th	36	5	4	13	22	33	44	61	38	6	6	6	8	9	11	13
Maximum	135	8	5	20	53	97	156	263	173	7	6	8	9	10	12	16
IQR	9	6	11	16	21	26	29	34	9	9	12	22	25	23	25	28
Mean	8	-1	-3	-3	-2	0	3	9	14	-1	-6	-10	-13	-15	-16	-18
Std Dev	33	8	8	14	21	30	43	68	41	8	11	16	19	20	22	24
Std Err Mean	7	2	2	3	5	7	10	16	9	2	3	4	4	5	5	5
Lower 95th	-7	-5	-7	-10	-12	-14	-18	-23	-5	-5	-11	-18	-21	-24	-27	-30
Upper 95th	24	3	1	4	8	15	24	42	34	3	0	-2	-4	-5	-6	-7

Goodness-of-Fit

Throughout the RPD analysis a qualitative visual assessment was used to assess goodness-of-fit, done by superimposing both frequency curves on the data plotted with the respective plotting positions (PeakFQ B17B-GB outputs Weibull plotting positions; EMA-MGB outputs Hirsch-Stedinger [1987] plotting positions). The quantitative goodness-of-fit measure, or mean of the absolute percent difference between the annual peak flows and the LP3 fitted values (*MAPD*) (appendix 4), showed that EMA-MGB was fitting the observed data slightly better than B17B-GB for each percentile of data selected, although the difference was less for lower percentiles (fig. 25, table 6). For the 90th percentile the median *MAPD* for all gages using EMA-MGB with station skew was roughly 3 percent less than B17B-GB (p = 0 .0065). For the 75th and 50th percentiles the median *MAPD* was roughly 2 percent less (p = 0 .0163 and 0.0278, respectively).

When stations were divided into qualification code categories, only the historical information category demonstrated a significant difference between the two methods at minimizing the *MAPD*. For the historical information category (2), the 90th-percentile median *MAPD* was about 5 percent less for EMA-MGB (p = 0.0260) (fig. 26*A*, table 7)

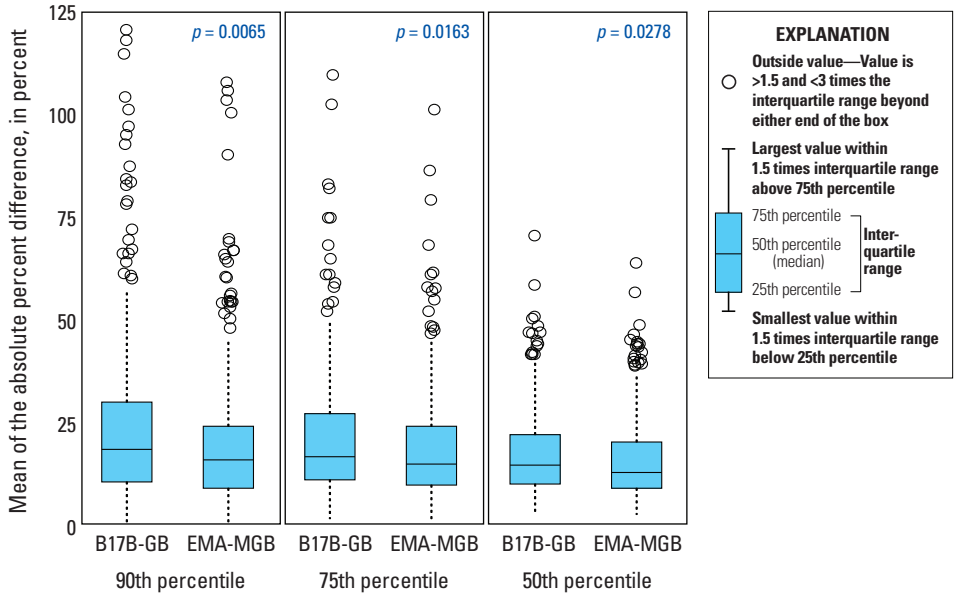


Figure 25. Boxplots showing the mean of the absolute percent difference (MAPD) between the annual peak-flows and the log Pearson Type 3 fitted values for Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB) for all stations in the study using a station skew. The 90th, 75th, and 50th percentiles of peak-flow data were selected for the comparison. A nonparametric Wilcoxon rank-sums test was used to identify significant differences between methods; *p*-values are labeled.

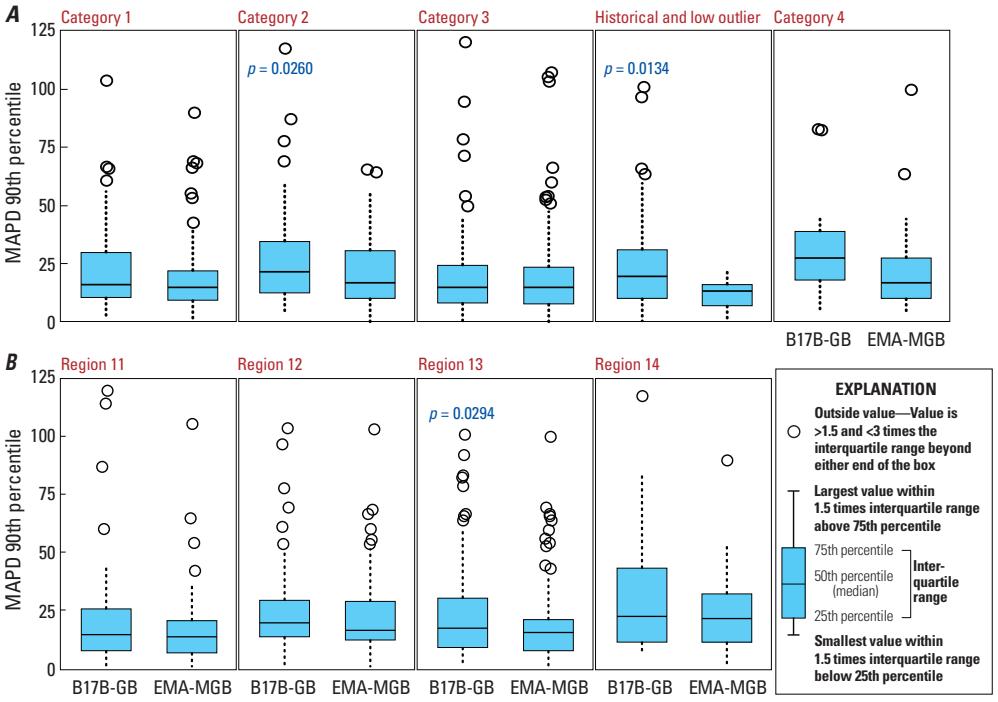


Figure 26. Boxplots comparing, across categories and regions, the mean of the absolute percent difference (MAPD) between the annual peak-flows and the log Pearson Type 3 fitted values for Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB) using a station skew. The 90th percentiles of peak-flow data were selected for the comparison, and data are grouped by (*A*) qualification code categories and (*B*) regional regression groups (Thomas and others, 1997). A nonparametric Wilcoxon rank-sums test was used to identify significant differences between methods; *p*-values are labeled for significantly different distributions.

Table 6. Mean of the absolute percent difference (MAPD) between the annual peak-flows and the predicted flows from the log-Pearson Type 3 (LP3) frequency curve of Bulletin 17B (B17B-GB) and the Expected Moments Algorithm (EMA-MGB) for 328 gaging stations; using station skew for the 50th, 75th, and 90th percentile of peak-flow data.

[The 50th-, 75th-, 90th-percentile of peak-flow data were used in the MAPD calculation. The mean is determined from the absolute (no sign) differences between the observed peak flow flows and the B17B-GB and EMA-MGB estimated values Smaller summary statistics suggest that the method is fitting the observed more closely]

Summary Statistic (in percent)	MAPD Percentile of the peak-flow distribution					
	B17B-GB 50th	EMA-MGB 50th	B17B-GB 75th	EMA-MGB 75th	B17B-GB 90th	EMA-MGB 90th
Minimum	2.22	2.13	1.13	1.04	0.04	0.06
10th percentile	6.65	5.92	6.87	6.29	4.84	3.83
25th percentile	9.45	8.46	10.59	9.15	9.95	8.46
Median	14.12	12.34	16.17	14.23	18.01	15.23
75th percentile	21.62	19.74	26.56	23.48	29.55	23.43
90th percentile	30.06	29.60	37.71	36.81	50.13	41.00
Maximum	70.21	63.57	109.39	100.77	201.42	107.36
Mean	16.90	15.58	20.48	18.22	24.22	19.56
Standard Deviation	10.43	10.21	15.32	13.89	23.36	17.47
Standard Error of Mean	0.58	0.56	0.85	0.77	1.29	0.96
Lower 95 percent	15.77	14.47	18.82	16.72	21.68	17.66
Upper 95 percent	18.04	16.69	22.15	19.73	26.76	21.46
Number of streamgaging stations where MAPD overestimated the actual peak-flow data	46	46	51	68	100	117
Number of streamgaging stations where MAPD underestimated the actual peak-flow data	282	282	277	260	228	211

compared to B17B-GB. In addition, the IQR of B17B-GB for category 2 was twice that of EMA-MGB. While the median *MAPD* for the low-outlier category (3) was about 2 percent less for EMA-MGB, the results were not statistically significant. For streamgaging stations that had both historical and low-outlier data, the results were similar to category 2, suggesting that EMA-MGB is more efficiently using the historical information to better fit the measured peak data. Category 4 results were similar between methods. Differences per category between the estimators were less noticeable for the 75th-percentile median *MAPD* and negligible for the 50th-percentile median *MAPD*.

Classifying sites by regions showed that EMA-MGB always did as well or slightly better at reducing the percent difference between the fitted LP3 distribution and the 90th percentile of measured peak-flow data (fig. 26*B*, table 8). Median MAPDs and IQRs were lower for EMA-MGB, but only region 13 was significantly lower than B17B (3 percent, $p = 0.0294$). B17B-GB and EMA-MGB were most similar in region 11; median *MAPD* differed by only 1 percent for all three percentiles. The two estimators were also similar in region 14, except that the IQR was greatly reduced for EMA using the 90th and 75th percentile datasets. This lower spread indicates the greatest observed floods were fitted more accurately at more stations using EMA-MGB.

Table 7. Mean of the absolute percent difference (MAPD) between the annual peak-flows and the predicted flows from the log-Pearson Type 3 frequency (LP3) curve of Bulletin 17B (B17B-GB) and the Expected Moments Algorithm (EMA-MGB) for 328 gaging stations; using station skew for the 50th, 75th, and 90th percentile of peak-flow data; grouped by NWIS code categories.

[NWIS, U.S. Geological Survey National Water Information System; NWIS code categories; category 1, stations containing no PILFs, historical information or other code; category 2, historical codes (7) only; category 3; PILFs only; and category 4, peak-flow records with one or more codes 1,4, or 8). The 50th-, 75th-, 90th-percentile of peak-flow data were used in the MAPD calculation. The mean is determined from the absolute (no sign) differences between the observed peak flow flows and the B17B-GB and EMA-MGB estimated values]

Qualification Category	Summary Statistic (in percent)	MAPD Percentile of the peak-flow distribution					
		B17B-GB 50th	EMA-MGB 50th	B17B-GB 75th	EMA-MGB 75th	B17B-GB 90th	EMA-MGB 90th
1	Minimum	5.29	5.28	3.92	3.89	1.11	0.92
	10th Percentile	7.40	7.41	7.99	7.85	3.82	3.67
	25th Percentile	9.44	9.31	11.21	11.23	10.75	10.42
	Median	14.49	14.34	17.04	16.62	16.25	16.07
	75th Percentile	22.07	23.39	28.72	28.19	29.71	31.19
	90th Percentile	37.83	37.84	42.06	42.02	54.34	54.09
	Maximum	48.08	48.37	64.55	60.57	103.90	103.21
	Mean	17.74	17.82	21.81	21.51	22.64	22.61
	Standard Deviation	10.77	11.01	14.38	14.16	19.51	19.55
	Standard Error Mean	1.38	1.41	1.84	1.81	2.50	2.50
	Lower 95th Confidence Interval	14.99	15.00	18.12	17.89	17.64	17.61
	Upper 95th Confidence Interval	20.50	20.64	25.49	25.14	27.63	27.62
2	Minimum	3.96	4.36	2.76	3.58	4.30	3.08
	10th Percentile	7.57	6.81	6.99	7.01	7.57	5.59
	25th Percentile	9.73	9.22	11.48	9.85	12.75	11.17
	Median	13.27	12.90	17.64	15.43	21.71	16.88
	75th Percentile	19.89	19.94	26.76	22.58	34.40	22.51
	90th Percentile	33.25	25.57	38.70	30.28	62.98	31.98
	Maximum	46.28	39.87	74.57	61.21	117.59	39.22
	Mean	16.44	14.88	21.18	17.29	28.51	17.56
	Standard Deviation	9.82	8.14	14.24	10.37	23.42	8.99
	Standard Error Mean	1.45	1.20	2.10	1.53	3.45	1.33
	Lower 95th Confidence Interval	13.52	12.47	16.95	14.21	21.55	14.89
	Upper 95th Confidence Interval	19.36	17.30	25.41	20.37	35.46	20.23
3	Minimum	2.22	2.13	1.13	1.04	0.04	0.13
	10th Percentile	5.27	5.10	5.93	4.94	3.09	3.17
	25th Percentile	8.82	7.19	9.27	7.44	8.19	5.96
	Median	12.93	10.95	13.57	13.12	15.02	13.18
	75th Percentile	20.74	18.90	21.84	20.03	24.32	21.25
	90th Percentile	29.74	29.12	35.97	35.98	40.95	43.05
	Maximum	70.21	63.57	109.39	100.77	120.23	107.36
	Mean	16.20	14.70	18.33	17.27	19.45	18.14
	Standard Deviation	11.01	10.90	15.83	16.11	18.56	18.84
	Standard Error Mean	1.04	1.03	1.50	1.52	1.75	1.78
	Lower 95th Confidence Interval	14.14	12.66	15.36	14.25	15.98	14.61
	Upper 95th Confidence Interval	18.26	16.74	21.29	20.28	22.93	21.67
4	Minimum	7.52	7.70	6.28	8.43	4.86	10.05
	10th Percentile	8.50	8.22	7.44	9.46	9.16	10.82
	25th Percentile	12.94	12.63	12.95	11.46	18.26	14.44
	Median	21.05	17.33	17.69	15.14	27.32	20.83
	75th Percentile	25.35	21.96	29.27	22.69	38.93	34.53
	90th Percentile	30.55	29.09	37.21	33.19	82.80	81.85
	Maximum	35.30	30.99	41.56	38.59	83.10	100.01
	Mean	19.50	17.67	20.48	17.84	32.79	29.82
	Standard Deviation	7.83	6.85	10.47	8.25	23.40	24.97
	Standard Error Mean	2.09	1.83	2.80	2.21	6.25	6.67
	Lower 95th Confidence Interval	14.98	13.72	14.44	13.07	19.27	15.40
	Upper 95th Confidence Interval	24.02	21.63	26.52	22.60	46.30	44.24

Table 8. Mean of the absolute percent difference (MAPD) between the annual peak-flows and the predicted flows from the log-Pearson Type 3 frequency (LP3) curve of Bulletin 17B (B17B-GB) and the Expected Moments Algorithm (EMA-MGB) for 328 gaging stations; using station skew for the 50th, 75th, and 90th percentile of peak-flow data; grouped by flood regions.

[Flood regions defined in U.S. Geological Survey Water-Supply Paper 2433 (Thomas and others, 1997), The 50th-, 75th-, 90th-percentile of peak-flow data were used in the MAPD calculation. The mean is determined from the absolute (no sign) differences between the observed peak flow flows and the B17B-GB and EMA-MGB estimated values]

WSP-2433 Regression Region	Summary Statistic (in percent)	MAPD Percentile of the peak-flow distribution					
		B17B-GB 50th	EMA-MGB 50th	B17B-GB 75th	EMA-MGB 75th	B17B-GB 90th	EMA-MGB 90th
11	Minimum	2	2	2	1	0	0
	10th Percentile	6	6	6	7	3	3
	25th Percentile	8	7	10	9	7	6
	Median	13	11	14	14	14	13
	75th Percentile	19	18	23	23	25	20
	90th Percentile	34	30	40	39	42	36
	Maximum	50	40	74	58	120	105
	Mean	16	15	19	18	21	17
	Standard Deviation	11	9	15	13	25	18
	Standard Error Mean	2	1	2	2	3	2
	Lower 95th Confidence Interval	13	12	15	14	14	12
	Upper 95th Confidence Interval	19	17	24	21	28	22
12	Minimum	5	3	5	2	0	0
	10th Percentile	8	7	9	7	8	7
	25th Percentile	12	10	12	11	13	12
	Median	18	16	20	18	19	16
	75th Percentile	26	25	32	31	29	28
	90th Percentile	35	36	43	42	46	44
	Maximum	58	56	82	86	104	103
	Mean	20	19	24	22	24	22
	Standard Deviation	10	11	15	16	18	17
	Standard Error Mean	1	1	1	2	2	2
	Lower 95th Confidence Interval	18	17	21	19	21	19
	Upper 95th Confidence Interval	22	21	27	25	28	25
13	Minimum	3	4	2	1	0	0
	10th Percentile	6	6	6	5	5	4
	25th Percentile	8	8	9	8	9	7
	Median	13	11	14	12	17	15
	75th Percentile	19	16	22	19	30	21
	90th Percentile	25	23	31	27	54	37
	Maximum	70	64	109	101	101	100
	Mean	15	13	17	15	23	18
	Standard Deviation	9	8	13	12	20	16
	Standard Error Mean	1	1	1	1	2	1
	Lower 95th Confidence Interval	13	12	15	13	19	15
	Upper 95th Confidence Interval	16	15	19	17	26	20
14	Minimum	5	2	7	2	7	0
	10th Percentile	6	6	8	7	8	5
	25th Percentile	10	7	9	10	11	11
	Median	12	11	16	16	22	21
	75th Percentile	22	18	27	22	43	32
	90th Percentile	30	24	47	33	84	54
	Maximum	46	30	75	46	118	90
	Mean	16	13	22	17	32	24
	Standard Deviation	10	7	17	11	30	20
	Standard Error Mean	2	2	4	2	7	5
	Lower 95th Confidence Interval	11	10	13	12	18	14
	Upper 95th Confidence Interval	21	17	30	22	46	34

Resampling Statistics

The effectiveness of the MGB test in identifying low outliers was quantified using resampling procedures at eight selected gaging stations. For each of 1,000 subsamples at each gage, the percent subsample difference,, was calculated as

$$\varepsilon_{ss} = \frac{(q_{ss} - q_{all})}{q_{all}} * 100 \quad (13)$$

where q_{ss} is the predicted discharge for a given subsample for a specific flood probability and q_{all} is the predicted discharge for that probability calculated using the complete record. In other words, ε_{ss} is the difference between the subsample-predicted flood and the "true" flood (as indicated by the complete record), and values closer to zero are more precise. The distributions of subsample differences at all stations are presented as boxplots where the horizontal dashed line represents the "true" flood for each method and AEP combination.

Desirable characteristics of the low-outlier method, as identified using subsampling, are that the subsample differences (are symmetric about zero (that is, the median is close to zero), and the difference between high and low percentiles is small (here, quantified as the inter-quartile range, IQR). By the former criteria, the methods were essentially equivalent when using station skew, but EMA performed better when using weighted skew (table 9). Median of B17B-GB resamples with weighted skew were generally similar to those with station skew, but with EMA-MGB, median was much less with weighted skew. This is likely because of the lower MSE of the regional skew used with EMA-MGB. The methods produced different values of IQR, varying up to 46 percent (table 10), and although IQR was smaller at more gages (5) when using B17B-GB, neither method was clearly performing better than the other. Positive values mean EMA-MGB was closer to the "true" value for the method, whereas negative values mean B17B-GB was closer to the "true" value. As the AEP decreased, the magnitude of the differences between the methods became larger, although the ways in which they differed remained similar.

The boxplots (figs. 27–30) showed that with station skew, B17B-GB performed better at two of the eight streamgaging stations (09505200 and 09505350), as evidenced by smaller values, median closer to zero, and differences better centered about zero. For three of the gages, the "true" value was not contained within the interquartile range of the B17B-GB boxplot. At two streamgaging stations (09480000 and 09516500) EMA-MGB performed better, and at the remaining stations neither method was clearly better.

Frequency curve plots are useful for visualizing why the resampling procedure indicates that one method is better than the other. For frequency fits that had several low peak flows

Table 9. The median percent difference for the resampling procedure for the selected streamgages using Bulletin 17B Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm multiple Grubbs-Beck test (EMA-MGB).

[The median percent difference is presented for the 1-, 0.5-, and 0.2-percent annual exceedance probability (AEP) for both station skew and weighted regional skew]

Skew Type	Station Identifier	Median percent difference, AEP estimate in percent					
		B17B-GB 1	EMA-MGB 1	B17B-GB 0.5	EMA-MGB 0.5	B17B-GB 0.2	EMA-MGB 0.2
Station skew	09480000	22.9	6.2	28.4	9.0	36.3	19.1
	09482000	2.5	2.6	2.9	3.8	4.6	10.4
	09486500	2.0	-8.4	2.3	-12.3	2.6	-28.1
	09505200	0.8	15.3	1.8	17.6	3.2	26.3
	09505350	1.1	26.2	1.3	36.1	1.2	75.5
	09513780	-0.2	-4.3	0.4	-5.4	0.1	-11.8
	09516500	17.6	3.2	23.7	2.8	29.8	4.6
	09517000	9.0	10.5	12.1	11.3	15.4	20.7
Weighted skew	09480000	18.1	-0.6	25.1	-0.3	36.2	0.4
	09482000	-1.1	-3.1	-1.7	-3.9	-1.1	-8.7
	09486500	3.7	3.1	4.4	2.8	5.5	3.3
	09505200	8.8	7.1	11.9	8.3	15.2	13.1
	09505350	3.4	3.8	3.9	6.3	5.3	15.9
	09513780	11.0	0.8	14.3	0.5	17.4	2.5
	09516500	15.8	-3.1	20.5	-3.9	28.4	-9.5
	09517000	10.4	6.2	13.0	6.4	17.3	12.2

Table 10. The interquartile range (IQR) for the resampling procedure for the selected streamgages using Bulletin 17B Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm multiple Grubbs-Beck test (EMA-MGB).

[The relative percent difference and IQR are presented for the 1-, 0.5-, and 0.2-percent annual exceedance probability (AEP) for both station skew and weighted regional skew. cfs, cubic feet per second]

Skew Type	Station Identifier	IQR 1-percent AEP (cfs)		Relative Percent Difference	IQR 0.5-percent AEP (cfs)		Relative Percent Difference	IQR 0.2-percent AEP (cfs)		Relative Percent Difference
		B17B-GB	EMA-MGB		B17B-GB	EMA-MGB		B17B-GB	EMA-MGB	
Station skew	09480000	5,181	4,705	9	7,396	6,706	9	10,733	9,939	7
	09482000	18,910	18,906	0	29,225	28,797	1	49,725	49,786	0
	09486500	21,695	21,483	1	30,485	30,331	1	44,310	44,121	0
	09505200	6,095	6,950	-14	7,860	9,310	-18	11,015	13,899	-26
	09505350	16,475	18,355	-11	25,130	28,612	-14	43,815	47,388	-8
	09513780	15,070	15,233	-1	22,300	23,192	-4	37,790	38,436	-2
	09516500	23,640	25,829	-9	33,890	37,801	-12	54,535	63,928	-17
	09517000	22,225	24,542	-10	32,485	35,858	-10	51,025	54,683	-7
Weighted skew	09480000	5,875	4,636	21	7,570	6,390	16	10,385	9,331	10
	09482000	16,230	13,643	16	23,890	19,128	20	37,830	28,775	24
	09486500	17,870	15,588	13	24,395	20,429	16	35,775	28,190	21
	09505200	7,500	8,619	-15	10,370	12,549	-21	14,070	18,502	-31
	09505350	14,700	16,284	-11	22,125	24,219	-9	35,590	37,498	-5
	09513780	21,135	25,689	-22	30,435	41,589	-37	47,980	70,007	-46
	09516500	24,080	24,753	-3	33,800	35,339	-5	52,455	53,893	-3
	09517000	23,205	28,185	-21	33,115	41,408	-25	53,465	65,071	-22

separated from the majority of the fitted peak flows and where B17B-GB appeared to be fitting the observed peak-flow data better than EMA-MGB, the separated low peak flows were visibly near the trajectory of the lower end of the frequency curve, meaning there was no step or dramatic departure of the peak flows from the fitted curve that would make those PILFs on the upper end on the frequency curve. Furthermore, the upper ends of these flood-frequency curves were relatively flat (reflecting a more homogenous peak-flow distribution), and removal of a single low-outlier when using B17B-GB had little effect on the frequency fit at the lower end of the frequency curve or influence on the smaller AEPs. In contrast, at the streamgaging stations where EMA-MGB performed better, the group of low peak flows plotted distinctly below the frequency curve, suggesting that the peak flows did not belong to the greater population of peak flows and departed from the LP3 distribution. At these stations the MGB test was more likely to consistently identify multiple PILFs; therefore EMA with the MGB test was less influenced by PILFs and the frequency curve visibly fit the observed peak flows better once they were properly identified. Because there were fewer PILFs at these stations, EMA-MGB varied less around the "true" value while B17B-GB varied more widely, depending on the number of PILFs in each subsample.

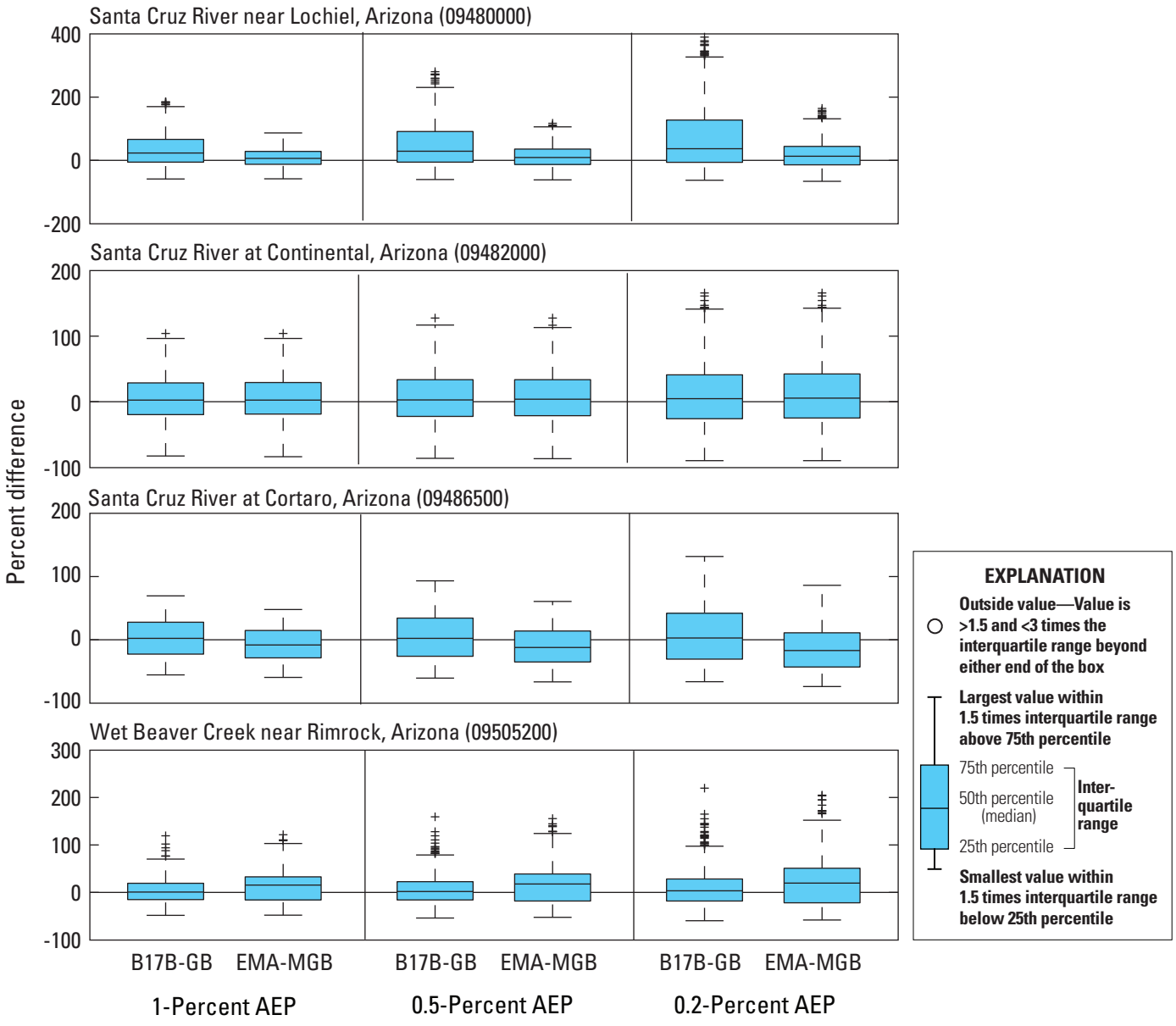


Figure 27. Boxplots showing the resampling-procedure percent difference at streamgaging stations 09480000, 09482000, 09486500, and 09505200 using Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB) with a station skew. The 1-, 0.5-, and 0.2-percent annual exceedance probabilities (AEP) are shown.

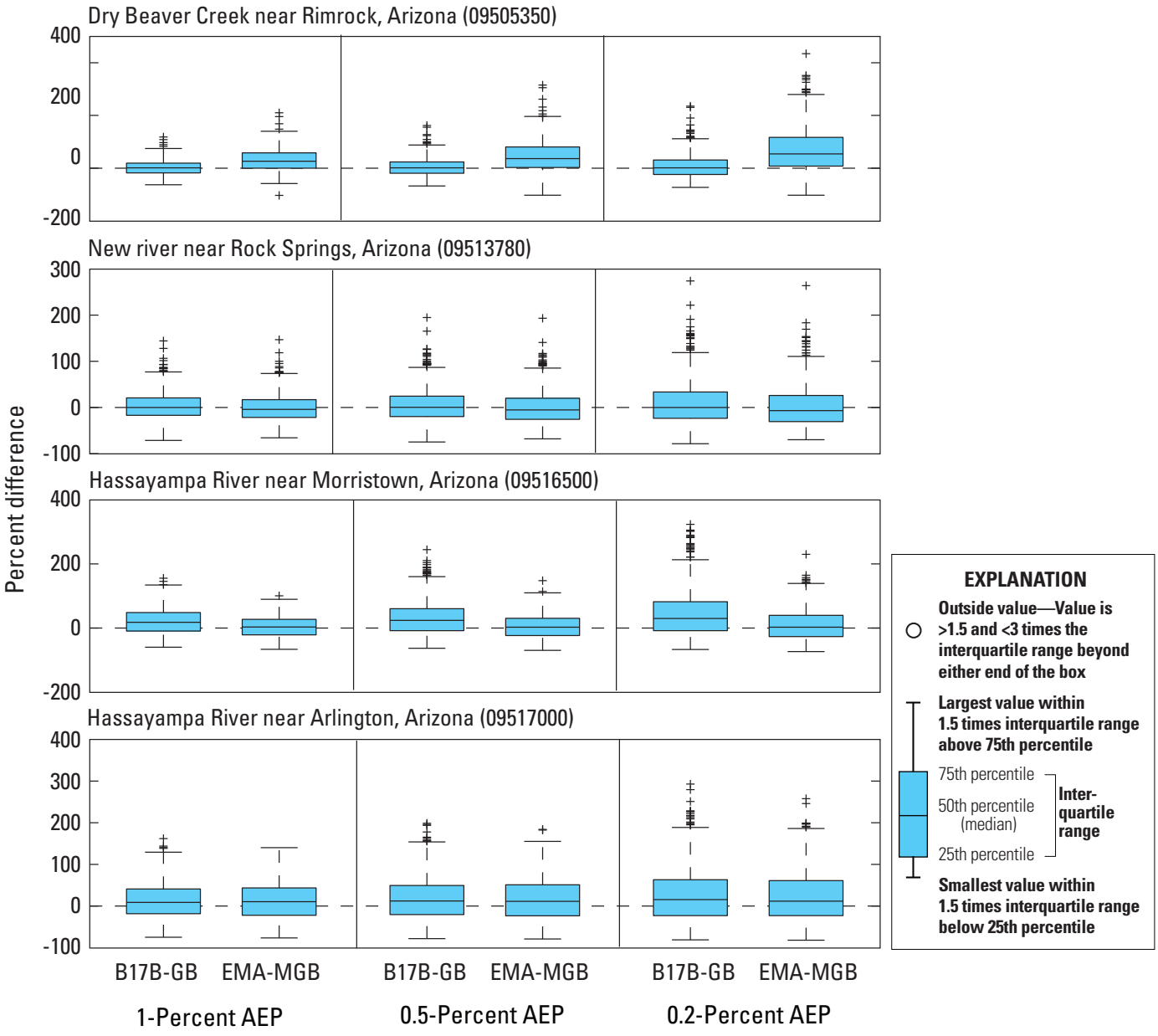


Figure 28. Boxplots showing the resampling-procedure percent difference at streamgaging stations 09505350, 09513780, 09516500, and 09517000 using Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB) with a station skew. The 1-, 0.5-, and 0.2-percent annual exceedance probabilities (AEP) are shown.

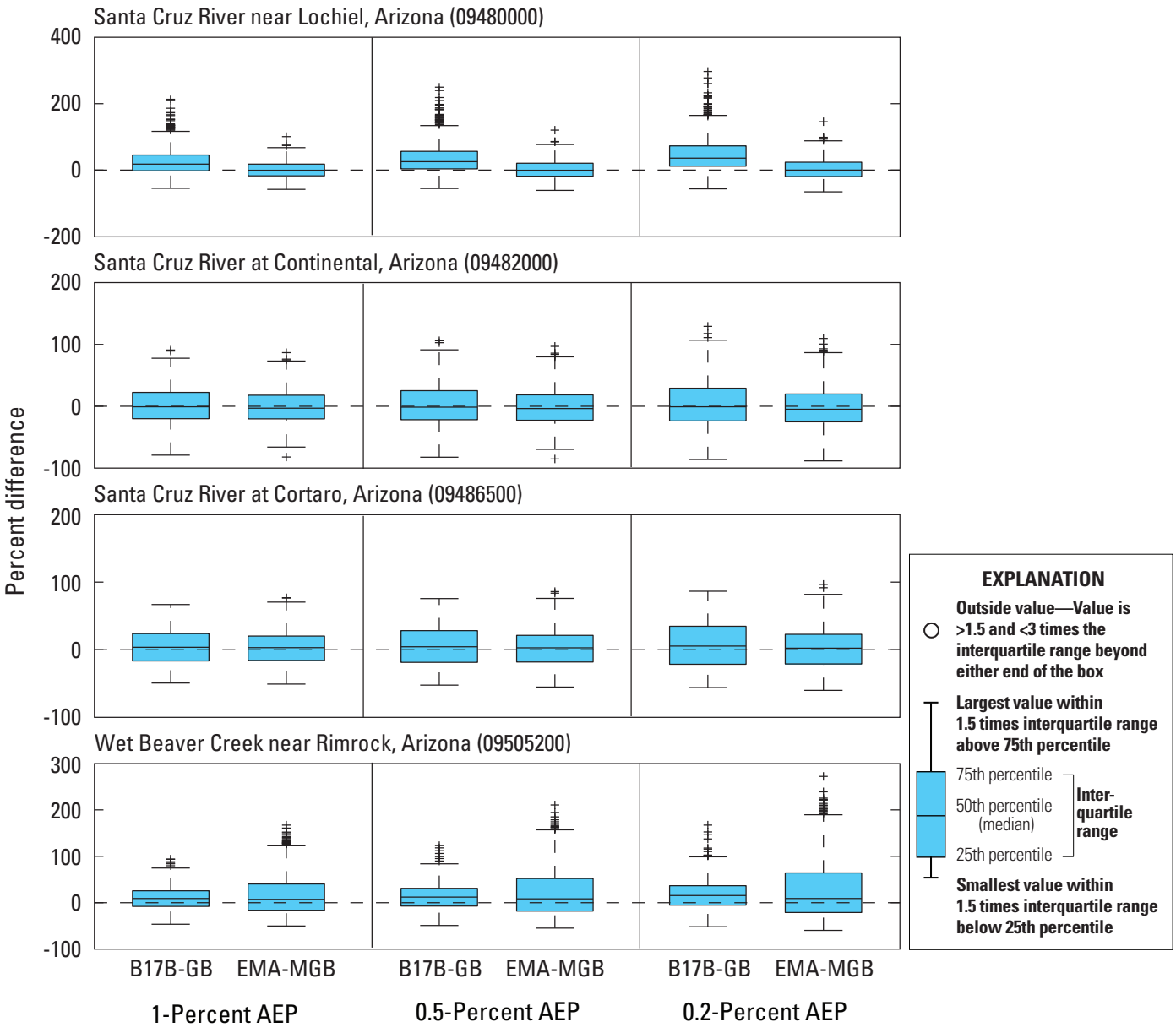


Figure 29. Boxplots showing the resampling-procedure difference at streamgaging stations 09480000, 09482000, 09486500, and 09505200 using Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB) with a weighted skew (B17B-GB, generalized skew = 0.0; mean square error (MSE) = 0.302. EMA-MGB, generalized skew = −0.09, MSE = 0.08). The 1-, 0.5-, and 0.2-percent annual exceedance probabilities (AEP) are shown.

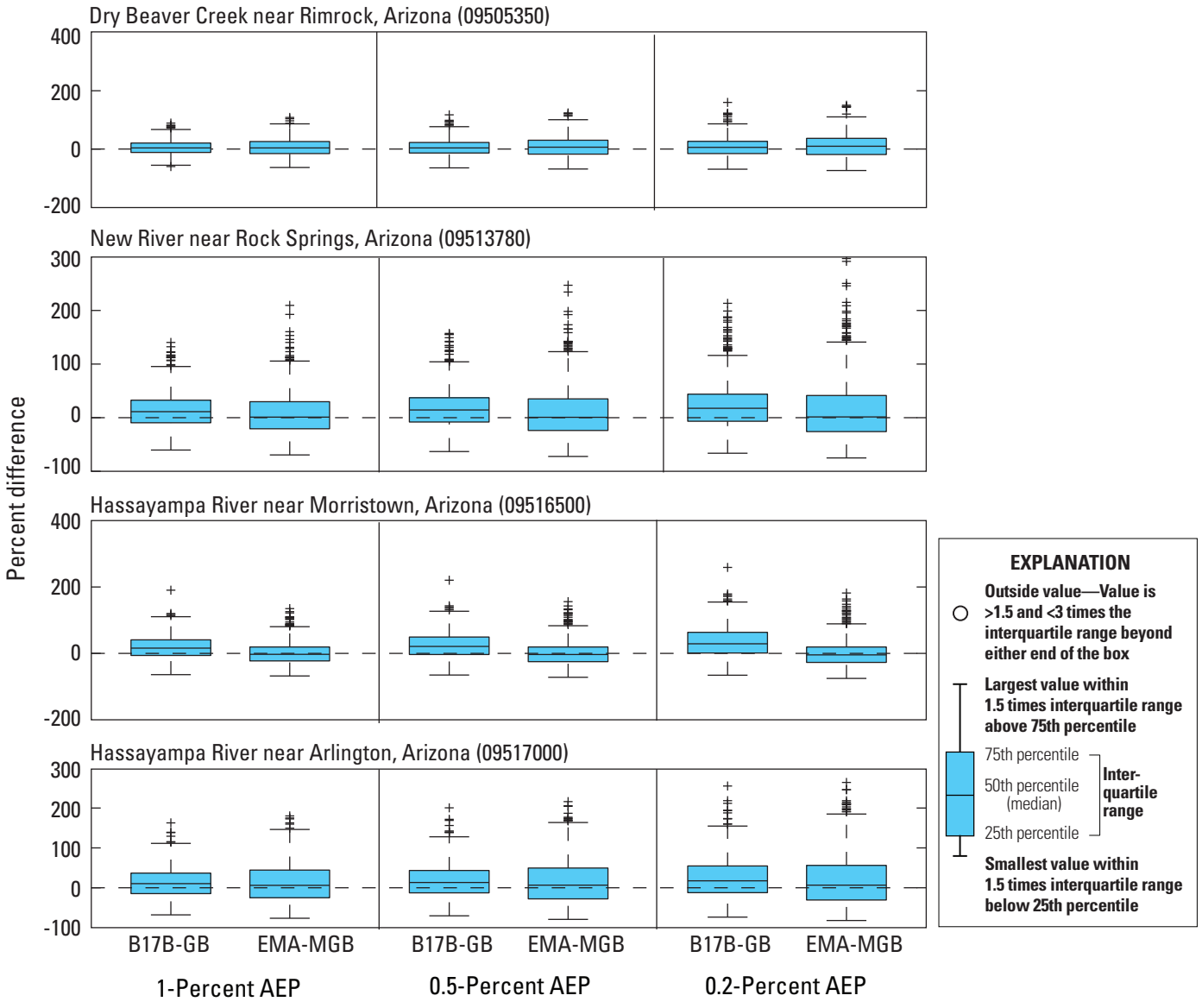


Figure 30. Boxplots showing the resampling-procedure percent difference at streamgaging stations 09505350, 09513780, 09516500, and 09517000 using Bulletin 17B with the Grubbs-Beck test (B17B-GB) and Expected Moments Algorithm with a multiple Grubbs-Beck test (EMA-MGB) with a weighted skew (B17B-GB, generalized skew = 0.0; mean square error (MSE) = 0.302. EMA-MGB, generalized skew = −0.09; MSE = 0.08). The 1-, 0.5-, and 0.2-percent annual exceedance probabilities (AEP) are shown.

Monte Carlo Simulation

Monte Carlo simulations were run for eight variations of LP3 distribution. The first five simulation examples assumed there was no regional information and only station skew was varied. Skew coefficients of −1.0, −0.5, 0, 0.5, and 1.0 were investigated. B17B-GB and EMA-MGB methods were compared, each using four historical flood scenarios (0, 1, 2, and 10 expected number of historical floods over $H = 100$ years). Performance was measured by assessing how close the median of the 1,000 replicate samples was to the actual 1-percent AEP and by the average gain (AG) of the effective record length (ERL), which effectively describes the method's efficiency in using historical information.

When the LP3 population skew was zero and no regional information was included, both EMA-MGB and B17B-GB did equally well when no historical floods were observed (AG was zero percent) (fig. 31). When 1 and 2 historical floods were included in the simulation, both B17B-GB and EMA-MGB methods were equally biased and close to predicting the actual 1-percent AEP, but the corresponding average gains were 30 and 41 for the 1-percent (1 expected flood) nonexceedance threshold, respectively, and 48 and 57 for the 2-percent (2 expected floods) nonexceedance threshold. When the 10-percent nonexceedance threshold (10 expected floods) was tested, both of the methods performed similarly, with AGs above 75.

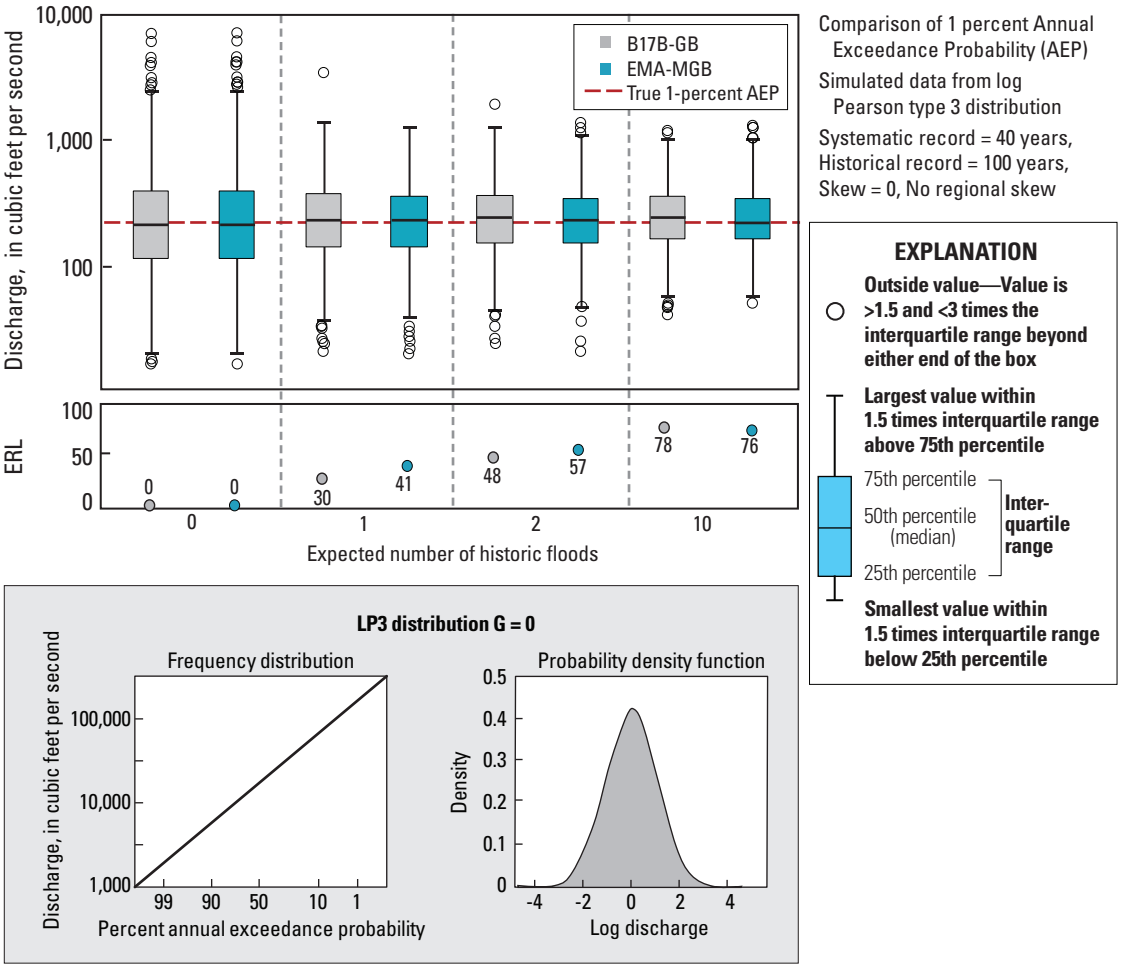


Figure 31. Graphs showing Monte Carlo simulations of a log Pearson type 3 (LP3) distribution model with a skew coefficient of 0.0 with no regional information. The boxplots represent comparison of the 1-percent quantile for the Bulletin 17B Grubbs-Beck (B17B-GB) low-outlier test (gray) and the Expected Moments Algorithm multiple Grubbs-Beck (EMA-MGB) test (blue), and the dashed red line is the true 1-percent flood exceedance. The systematic (S) record is 40 years and the historical (H) period of record equals 100 years. Simulations were run with the expectation that a gage may see 0, 1, 2, and 10 historical floods, and the effective record length (ERL) represents the number of additional years of record that would be gained by adding the historical information.

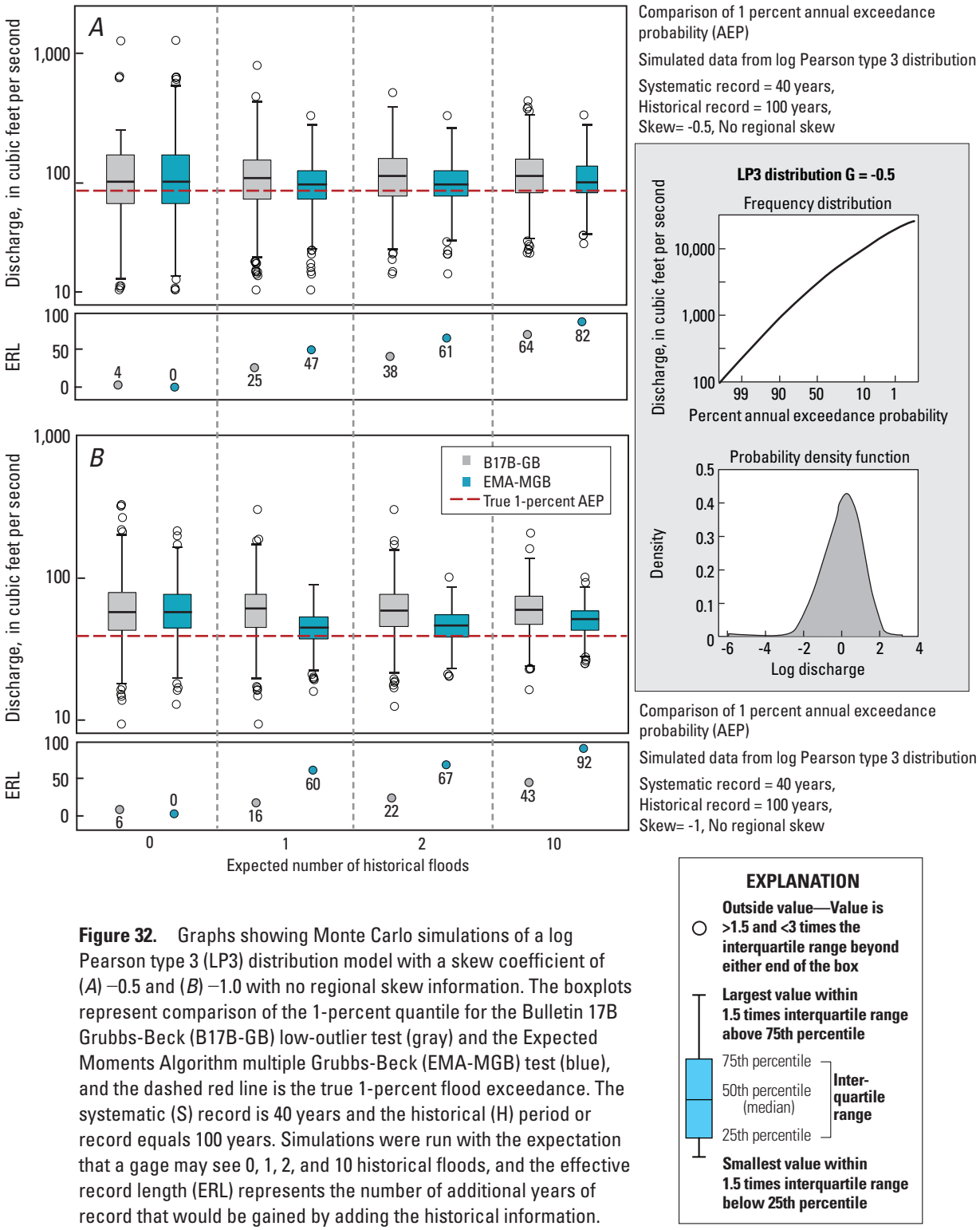
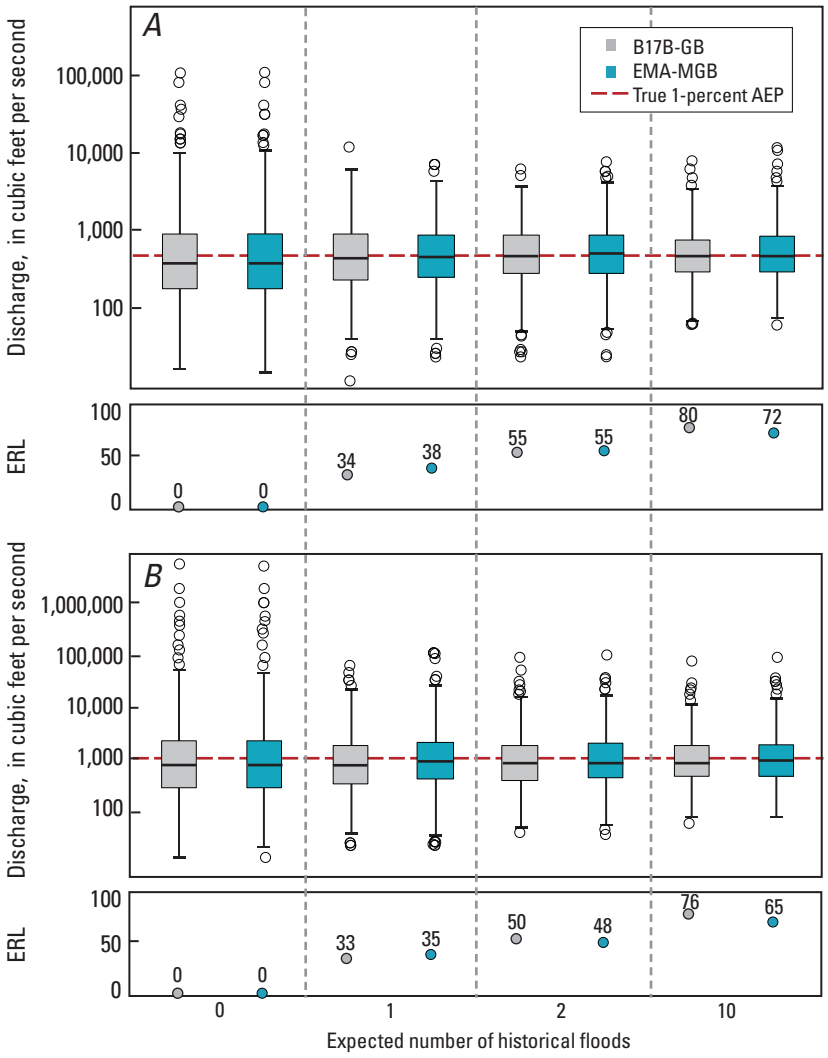
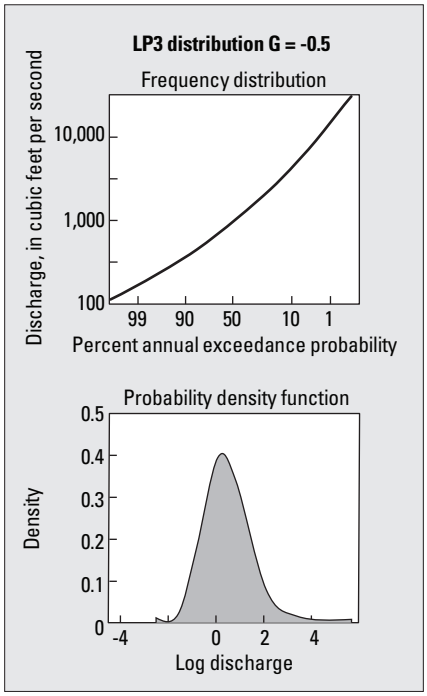


Figure 32. Graphs showing Monte Carlo simulations of a log Pearson type 3 (LP3) distribution model with a skew coefficient of (*A*) −0.5 and (*B*) −1.0 with no regional skew information. The boxplots represent comparison of the 1-percent quantile for the Bulletin 17B Grubbs-Beck (B17B-GB) low-outlier test (gray) and the Expected Moments Algorithm multiple Grubbs-Beck (EMA-MGB) test (blue), and the dashed red line is the true 1-percent flood exceedance. The systematic (S) record is 40 years and the historical (H) period or record equals 100 years. Simulations were run with the expectation that a gage may see 0, 1, 2, and 10 historical floods, and the effective record length (ERL) represents the number of additional years of record that would be gained by adding the historical information.



Comparison of 1 percent annual exceedance probability (AEP)
Simulated data from log Pearson type 3 distribution
Systematic record = 40 years,
Historical record = 100 years,
Skew= 0.5, No regional skew



Comparison of 1 percent annual exceedance probability (AEP)
Simulated data from log Pearson type 3 distribution
Systematic record = 40 years,
Historical record = 100 years,
Skew= 1.0, No regional skew

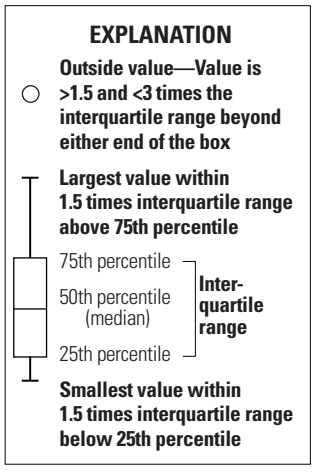


Figure 33. Graphs showing Monte Carlo Simulations of a log Pearson type 3 (LP3) distribution model with a skew coefficient of (*A*) 0.5 and (*B*) 1.0 and no regional skew information. The boxplots represent comparison of the 1-percent quantile for the Bulletin 17B Grubbs-Beck (B17B-GB) low-outlier test (gray) and the Expected Moments Algorithm multiple Grubbs-Beck (EMA-MGB) test (blue), and the dashed red line is the true 1-percent flood exceedance. The systematic (S) record is 40 years and the historical (H) period or record equals 100 years. Simulations were run with the expectation that a gage may see 0, 1, 2, and 10 historical floods, and the effective record length (ERL) represents the number of additional years of record that would be gained by adding the historical information.

For an LP3 population with skew equal to −0.5 (fig. 32*A*) and −1.0 (fig. 32*B*), the method estimator properties were much different compared to the populations that used zero and positive skews. A negative skew implies that more low outliers were expected, and with the MGB test EMA was expected to perform better than B17B, especially when historical information was considered. The medians were substantially greater than the true 1-percent AEP, indicating that both of the estimators were biased when only systematic data were employed. This observation has been documented in other investigations (Kirby, 1974; Stedinger and others, 1993), and the phenomenon can be explained by the method-of-moments estimator for the skew coefficient being biased toward zero—as a result, the method-of-moments quantile estimators are biased upwards for populations with negative skews and downwards for populations with positive skews. When historical information was included, EMA-MGB performed substantially better than B17B-GB. For all nonexceedance threshold scenarios (1-, 2-, and 10-percent), the AG for EMA-MGB was roughly 1.5 to 2 times greater than the AG of the B17B-GB estimator method using a skew of −0.5 and 2 to 6 times greater using a skew of −1.0. This result is consistent with the sensitivity of the method-of-moments to the smallest observations from negatively skewed populations. Thus EMA-MGB, which handles small observations through censoring, is more efficient at fitting large floods and performs much better than B17B-GB in low-outlier situations.

When the skew is a positive 0.5 (fig. 33*A*) or 1.0 (fig. 33*B*), then no low outliers are expected and as a result the opposite bias was observed. The systematic data-simulation boxplots for both skew scenarios clearly showed the median 1-percent AEP for both method estimators was greatly underestimating the true 1-percent AEP. When historical information was included, then both method estimators performed equally as well—the estimated 1-percent AEP was much closer to the true value and this was true for all the nonexceedance thresholds (1-, 2-, and 10-percent). Both of the methods made good use of the historical information, because for positively skewed populations the smallest values in the dataset have little leverage on the frequency fit and it makes little difference about how those data were treated.

Interactive Data Tools

The large number of gages considered in the comparison of flood-frequency analysis method precludes site-by-site analysis. Two interactive tools make statewide results more readily available (http://az.water.usgs.gov/projects/floodfreq). One is a kml file that displays all sites, with links to quantile-AEP plots and input and output files. Kml is the data format for Google Earth and can be displayed in other geospatial viewers, such as NASA's World Wind software. The second tool is a scatter-plot visualization that compares the AEPs and moments predicted by each method. Each point on the scatter plots is linked to the quantile-AEP plot for that gage, which in turn links to the input and output files. AEPs and moments can be plotted with a weighted skew option or as subsamples defined by region, drainage area, or hydrologic unit code.

Summary and Conclusions

This investigation analyzed the differences between two methods, B17B standard Grubbs-Beck test (B17B–GB) and EMA multiple Grubbs-Beck test (EMA–MGB), for estimating annual flood exceedance probabilities (AEPs) for 328 streamgaging stations in Arizona. Although it is difficult to make a direct comparison of method performance, because the true AEP flow estimate is never known, the results of the investigation support the inference that EMA-MGB performed as well or better than B17B-GB, especially when low outliers were present and (or) historical information was available. This finding was achieved by using several methods to compare the analyses—qualitative, quantitative, and robust model simulations.

The relative percent difference (RPD) metric was used to understand the magnitude and variability of differences in the AEP flow estimates determined using each method in a station-skew and weighted-skew analysis. RPD statistics were helpful in two ways: first, identifying patterns in the treatment of specific peak-flow data types and, second, understanding how flood frequency methods vary with regard to geoclimatic characteristics and related flood hydrology. The major findings from this analysis were: (1) the variability in RPDs increased with decreasing AEP estimates, indicating that that the number of low-outlier peaks identified in a flood series and historical information were very influential on the LP3 distribution fit for AEP estimates less than 2 percent (that is, 1, 0.5, and 0.2 percent); (2) records with less than 15 years affected the stability of the flood-frequency analyses (FFAs), and the presence of just one outlier would dramatically change the B17B-GB log-Pearson Type 3 (LP3) fit and increase the RPD significantly; (3) potentially influential low flows (PILFs) are present at more streamgaging stations and have more of an influence on the FFA than any other USGS National Water Information System (NWIS) qualification codes; (4) EMA-MGB consistently used historical information more effectively to fit the observed peak-flow data better than B17B-GB; (5) all previous flood regions (from Thomas and others, 1997) considered were significantly affected by PILFs and historical data, and this should be a significant consideration in future regionalization studies; and (6) the more positive B17B-GB regional skew mostly caused RPDs to be negative (B17B-GB estimate is greater than EMA-MGB), although visual inspection and the goodness-of-fit appeared to be more accurate with EMA-MGB. In the examples presented on a station-by-station basis the frequency fit comparison revealed that in most cases EMA-MGB was able to more properly identify PILFs and efficiently use historical data to effectively produce better visual fitted frequency curves, suggesting that the performance was better than B17B-GB.

A goodness-of-fit measure was modified slightly in order to quantitatively understand how the largest proportion of floods was modeled with an LP3-distribution fit using each method. The metric was calculated as the mean of the absolute percent differences (*MAPD*) for the 90th, 75th, and 50th percentile of the peak-flow data. While differences between the methods were on the order of 1 to 3 percent, the percent difference was significantly lower for EMA-MGB, meaning that on average the EMA-MGB-LP3 distribution is modeling the largest observed peaks better than B17B-GB. In addition, analyzing this information by category and region revealed that historical information significantly contributes to EMA-MGB's success in more precisely fitting the actual data. Although differences were not always statistically significant, results mostly showed smaller percent differences in the observed versus fitted data for EMA-MGB than for B17B-GB, by and large because the MGB test identifies those PILFS that contribute most to the leverage of low- and zero-flows by censoring PILFs in order to ensure that large floods are more correctly influencing the distribution fit.

Resampling indicated that EMA-MGB performed better when PILFs were from a distinct population. However, with station skew, B17B-GB median percent difference and IQR were generally smaller, especially when MGB-identified PILFs visually departed from the main population. With weighted skew, EMA-MGB median percent difference was nearly always lower, although IQR results were mixed. From the Monte Carlo simulation, it was determined that at stations with several PILFs, EMA-MGB outperformed B17B-GB and demonstrated less bias in the estimation of the 1-percent AEP. When historical information was available, the historical probability adjustment used by B17B-GB was less proficient at fitting historical floods. EMA-MGB treats the historical period as nonexceedance-interval data, effectively retaining more information for the historical period than is achieved by the B17B-GB historical adjustment. A major advantage of EMA-MGB over other estimators is that it preserves the B17B framework and provides nearly identical AEP flow estimates if no historical or low outlier information is present (Griffis and others, 2004; England and Cohn, 2008; Stedinger and Griffis, 2008).

Much of the analysis focused on analyzing differences between B17B-GB and EMA-MGB AEP estimates and trying to understand the causes for those differences. On a station basis, there were consistent patterns that suggested EMA-MGB maximizes the used of historical information by treating the information as a perception threshold rather than using a historical probability adjustment as with B17B. In addition, EMA-MGB properly addresses PILFs using the MGB test to ensure that zero and low-flow peaks that depart from the trend of the data have very little influence on the frequency fit. Finally, the findings in this study support those of previous studies that indicate the efficacy of EMA-MGB. For these reasons EMA-MGB appears to be a suitable successor to traditional B17B-GB methods.

References Cited

Anderson, T.W., Freethey, G.W., and Tucci, Patrick, 1992, Geohydrology and water resources of alluvial basins in south-central Arizona and adjacent states: U.S. Geological Survey Professional Paper 1406–D, 74 p.

Beard, L.R., 1974, Flood flow frequency techniques: Austin, University of Texas, Center for Research in Water Resources, 27 p.

Cohn, T.A., Lane, W.L., and Baier, W.G., 1997, An algorithm for computing moments-based flood quantile estimates when historical flood information is available: Water Resources Research, v. 33, p. 2089–2096.

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

Cohn, T.A., England, J.F., Barenbrock, C.E., Mason, R.R., Stedinger, J.R., and Lamontagne, J.R., 2013, A generalized Grubbs-Beck Test statistic for detecting multiple potentially influential low outliers in flood series: Water Resources Research, v. 49, p. 1–12.

England, J.F., Jr., and Cohn, T.A., 2008, Bulletin 17B flood frequency revisions; practical software and test comparison results: American Society of Civil Engineers, EWRI World Water & Environmental Resources Congress, 13–16 May, Honolulu, HI, 11 p.

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

Fill, H.D., and Stedinger, J.R. 1995, L-moment and probability plot correlation-coefficient goodness-of-fit tests for the Gumbel Distribution and impact of autocorrelation: Water Resources Research, v. 31, p. 225–229.

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

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., and Tyler, D., 2002, The National Elevation Dataset: Photogrammetric Engineering and Remote Sensing, v. 68, no. 1, p. 5–11.

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

Greenwood, J.A., Landwerh, J.M., Matalas, N.C., and Wallis, J.R., 1979, Probability weighted moments; definition and relation to parameters of distribution expressible in inverse form: Water Resources Research, v. 15, no. 5, p. 1049–1054.

Griffis, V.W., 2008, EMA with historical information, low outliers, and regional skew: American Society of Civil Engineers World Environmental and Water Resources Congress 2008, 10 p.

Griffis, V.W., and Stedinger, J.R., 2007, Evolution of flood frequency analysis with Bulletin 17: Journal of Hydrologic Engineering, v. 12, p. 283–297.

Griffis, V.W., Stedinger, J.R., and Cohn, T.A., 2004, Log Pearson type 3 quantile estimators with regional skew information and low outlier adjustments: Water Resources Research, v. 40, no. 10, W07503, doi:10.1029/2003WR002697.

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

Hirsch, R.M. and Stedinger, J.R., 1987, Plotting positions for historical floods and their precision: Water Resources Research, v. 23, no. 4, p. 715–727.

Hosking, J.R.M., 1990, L-moments; analysis and estimation of distributions using linear combinations of order statistics: Journal of the Royal Statistical Society B, v. 52, no. 2, p. 105–124.

Houghton, J.C., 1978, Birth of a parent—Wakeby Distribution for modeling flood flows: Water Resources Research, v. 14, p. 1105–1109.

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

Jennings, M.E., and Benson, M.A., 1969, Frequency curves for annual flood series with some zero events or incomplete data: Water Resources Research, v. 5, no. 1, p. 276–280.

Kirby, W., 1974, Algebraic boundedness of sample statistics: Water Resources Research, v. 10, no. 2, p. 220–222.

National Weather Service-Hydrologic Information Center, 2001, Flood Loss: National Weather Service-Hydrologic Information Center, http://www.nws.noaa.gov/oh/hic/flood_stats/ Flood_loss_time_series.htm (last accessed May 9, 2012).

Pielke, R.A., Downton, M.W., and Barnard Miller, J.Z, 2002, Flood damage in the United States, 1926–2000; a reanalysis of National Weather Service estimates: Boulder, Colo., National Oceanic and Atmospheric Administration, p. 1–86.

Paretti, N.V., Kennedy, J.R.., Turney, L.A., and Veilleux, A.G., in press, Methods for estimating magnitude and frequency of floods in Arizona, developed with unregulated and rural peak-flow data through water year 2010: U.S. Geological Survey Scientific Investigations Report 2014-XXXX, XX p.

Potter, K.W., and Lettenmaier, D.P., 1990, A comparison of regional flood frequency estimation methods using a resampling method: Water Resources Research, v. 26, no. 3, p. 415–424.

PRISM Climate Group, 2012, Parameter-elevation regressions on independent slopes model: Oregon State University, http://prism.oregonstate.edu (last accessed March 15, 2012).

Rao, A.R., and Hamed, K.H., 2000, Flood frequency analysis: Boca Raton, Florida, CRC Press, 350 p.

Reis, D.S., Stedinger, J.R., and Martins, E.S., 2005, Bayesian generalized least squares regression with application to log Pearson type 3 regional skew estimation: Water Resources Research, v. 41, p. 1–14.

Ryberg, K.R., 2008, PFReports—A program for systematic checking of annual peaks in NWISWeb: U.S. Geological Survey Open-File Report 2008–1284, 17 p.

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

Stedinger, J.R., Vogel, R.M., and Foufoula-Georgiou, E., 1993, Frequency analysis of extreme events, chap. 18 of Maidment, D., ed., Handbook of hydrology: New York, McGraw Hill, Inc.

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 2433, 205 p.

Trapp, R.A., and Reynolds, S.J., 1995, Map showing names and outlines of physiographic areas in Arizona used by the Arizona Geological Survey with base map showing township and range only: Arizona Geological Survey Open File Report OFR 95-2b, scale 1:1,000,000.

U.S. Army Corps of Engineers, 2000, An Investigation of flood frequency estimation methods for the Upper Mississippi Basin: U.S. Army Corps of Engineers, Upper Mississippi River System Flow Frequency Final Report, Appendix A, 162 p. Available at http://www.mvr.usace.army.mil/pdw/pdf/FlowFrequency/Documents/FinalReport/Reports/App. A HEC/App.AHydrology/App.AHydrology – AnInvestigationofFloodFreq.EstimationMethodsfortheUpperMississippiBasin.doc

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, 170 p.

Vogel, R.M., Thomas, W.O., and McMahon, T.A., 1993, Flood-flow frequency model selection in southwestern United States: Journal of Water Resources Planning and Management, v. 119, no. 3, p. 353–366.

Wilson, E.D., 1962, A resume of the geology of Arizona: Tucson,. University of Arizona Press, Arizona Bureau of Mines, Bulletin 171, 140 p.

Appendixes

Appendixes 1–4 are available online, and can be downloaded at http://pubs.usgs.gov/sir/2014/5026/.

Appendix 1. Station flood frequency analysis and relative percent difference statistics of the P-percent annual exceedance probability for 328 streamgaging stations estimated by Bulletin 17B Grubbs-Beck test and Expected Moments Algorithm multiple Grubbs-Beck test using station skew.

Appendix 2. Station flood frequency analysis and relative percent difference statistics of the P-percent annual exceedance probability for 328 streamgaging stations estimated by Bulletin 17B Grubbs-Beck test and Expected Moments Algorithm multiple Grubbs-Beck test using a weighted regional skew.

Appendix 3. Percent difference between the largest observed peak and the predicted annual exceedance probability (AEP) log-Pearson Type 3 estimated flow for streamgaging stations in National Water Information System category 2 by Bulletin 17B and the Expected Moments Algorithm.

Appendix 4. Mean of the absolute percent difference between the annual peak-flows and the predicted flows from the log-Pearson Type 3 frequency curve of Bulletin 17B and the Expected Moments Algorithm for 328 gaging stations; using station skew for the 50th, 75th, and 90th percentile of peak-flow data.

Menlo Park Publishing Service Center, California
Manuscript approved for publication February 11, 2014
Edited by Peter H. Stauffer
Design and layout by Jeanne S. DiLeo

ISSN 2328–0328 (online)
http://dx.doi.org/10.3133/sir20145026