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

**Prepared in cooperation with the
U.S. Forest Service**

Two-Station Comparison of Peak Flows to Improve Flood-Frequency Estimates for Seven Streamflow-Gaging Stations in the Salmon and Clearwater River Basins, Central Idaho

Water-Resources Investigations Report 03–4001

USGS
science for a changing world

Two-Station Comparison of Peak Flows to Improve Flood-Frequency Estimates for Seven Streamflow-Gaging Stations in the Salmon and Clearwater River Basins, Central Idaho

By Charles Berenbrock

Water-Resources Investigations Report 03–4001

Prepared in cooperation with the
U.S. Forest Service

Boise, Idaho
2003

U.S. DEPARTMENT OF THE INTERIOR
GALE A. NORTON, Secretary

U.S. GEOLOGICAL SURVEY
CHARLES G. GROAT, Director

Additional information can be obtained from:

District Chief
U.S. Geological Survey
230 Collins Road
Boise, ID 83702-4520

http://idaho.usgs.gov

Copies of this report also are available in PDF format, which can be viewed using Adobe Acrobat Reader, at URL:

http://idaho.usgs.gov/public/reports.html

CONTENTS

FIGURES

TABLES

CONVERSION TABLE

Multiply	By	To obtain
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
mile (mi)	1.609	kilometer (km)
square mile (mi^2)	2.590	square kilometer (km^2)

Two-Station Comparison of Peak Flows to Improve Flood-Frequency Estimates for Seven Streamflow-Gaging Stations in the Salmon and Clearwater River Basins, Central Idaho

By Charles Berenbrock

Abstract

Improved flood-frequency estimates for short-term (10 or fewer years of record) streamflow-gaging stations were needed to support instream flow studies by the U.S. Forest Service, which are focused on quantifying water rights necessary to maintain or restore productive fish habitat. Because peak-flow data for short-term gaging stations can be biased by having been collected during an unusually wet, dry, or otherwise unrepresentative period of record, the data may not represent the full range of potential floods at a site. To test whether peak-flow estimates for short-term gaging stations could be improved, the two-station comparison method was used to adjust the logarithmic mean and logarithmic standard deviation of peak flows for seven short-term gaging stations in the Salmon and Clearwater River Basins, central Idaho.

Correlation coefficients determined from regression of peak flows for paired short-term and long-term (more than 10 years of record) gaging stations over a concurrent period of record indicated that the mean and standard deviation of peak flows for all short-term gaging stations would be improved. Flood-frequency estimates for seven short-term gaging stations were determined using the adjusted mean and standard deviation. The original (unadjusted) flood-frequency estimates for three of the seven short-term gaging stations differed from the adjusted estimates by less than 10 percent, probably because the data were collected during periods representing the full range of peak flows. Unadjusted flood-frequency estimates for four short-term gaging stations differed from the adjusted estimates by more than 10 percent; unadjusted estimates for Little Slate Creek and Salmon River near Obsidian differed from adjusted estimates by nearly 30 percent. These large differences probably are attributable to unrepresentative periods of peak-flow data collection.

INTRODUCTION

The State of Idaho has initiated an adjudication of all water rights in the Snake River Basin, including the Salmon and Clearwater River Basins. To protect its interest, the Federal Government is attempting to establish and quantify the State appropriative and Federal reserved water rights held by the United States on its own behalf and as a trustee for the American public. Some water rights claims in the Snake River Basin have been settled; however, the claims in the Salmon and Clearwater River Basins have yet to be resolved.

For the past several years, the Boise Adjudication Team of the U.S. Forest Service (USFS) has been focusing on the quantification of water rights necessary to maintain or restore productive fish habitat. Hundreds of streams within the Salmon and Clearwater River Basins either are, or historically have been, capable of providing habitat for large populations of resident and anadromous fish species. The USFS has made water rights claims designed to protect these fish species by ensuring adequate instream flows. These claims are part of the adjudication of the Snake River Basin by the State of Idaho. The upper limit of the USFS claim is the 25-year, or 4-percent probability, flood as defined by the Organic Act for channel maintenance of fish habitat. The 4-percent probability flood has a 1-in-25 chance of occurring in any given year and has a recurrence interval of 25 years. Similarly, a 1-percent probability flood has a 1-in-100 chance of occurring in any given year and has a recurrence interval of 100 years.

In 2000, the USFS entered into a cooperative agreement with the U.S. Geological Survey (USGS) to provide flood-frequency analyses in support of the instream flow studies. Three steps were planned: (1) Estimate flood frequency and magnitude for 13 streamflow-gaging stations in the Salmon and Clearwater River Basins (completed in 2000); (2) improve flood-frequency estimates for 7 short-term gaging stations in the Salmon and Clearwater River Basins by applying the two-station comparison method (described in this report); and (3) develop equations for estimating monthly streamflow exceedances at 80-, 50-, and 20-year recurrence intervals for ungaged sites statewide on the basis of relations between streamflow and various basin and climatic variables (published in a report by Hortness and Berenbrock, 2001).

Purpose and Scope

The purpose of this report is to describe the application and results of the two-station comparison method used to adjust the logarithmic mean and logarithmic standard deviation of peak flows for selected gaging stations in the Salmon and Clearwater River Basins. The objective of this study was to improve flood-frequency estimates for short-term gaging stations by comparing peak-flow data for short-term gaging stations to data for nearby long-term gaging stations and making inferences about the total population of peak flows at the short-term station.

Description of Study Area

The Salmon and Clearwater River Basins are located in central Idaho (fig. 1) and encompass an area of about 23,700 mi^2. The area includes parts of seven national forests, one national recreation area, four wilderness areas, and five designated wild and scenic rivers. Combined, the four wilderness areas compose more than 6,000 mi^2, or about 25 percent of the Salmon and Clearwater River Basins.

The Salmon River is 425 mi long and drains an area of about 14,000 mi^2. The Salmon River Basin comprises about 7,600 mi of streams. Principal tributaries are the East Fork Salmon, Pahsimeroi, Lemhi, North Fork Salmon, Middle Fork Salmon, South Fork Salmon, and Little Salmon Rivers (fig. 1). The mean annual streamflow of the Salmon River at the USGS gaging station nearest the mouth (Salmon River at White Bird, 13317000) is 13,600 ft^3/s; the 100-year recurrence interval flood is 130,000 ft^3/s, and the 25-year recurrence interval flood is 111,000 ft^3/s. The drainage area of the basin upstream from the Salmon River at White Bird gaging station is 13,550 mi^2, or 97 percent of the entire basin area.

The Clearwater River is about 175 mi long and drains an area of about 9,700 mi^2. The Clearwater River Basin comprises about 5,300 mi of streams. Principal tributaries are the North Fork, South Fork, and Middle Fork Clearwater Rivers and the Lochsa and Selway Rivers, the confluence of which forms the Middle Fork Clearwater River. The mean annual streamflow of the Clearwater River at the USGS gaging station nearest the mouth (Clearwater River at Spalding, 13342500) is 15,400 ft^3/s; the 100-year recurrence interval flood is 168,000 ft^3/s, and the 25-year recurrence interval flood is 141,000 ft^3/s. The drainage area of the basin upstream from the Clearwater River at Spalding gaging station is 9,570 mi^2, or 99 percent of the entire basin area.

METHOD

Because peak-flow data for a short-term (10 or fewer years of record) gaging station may have been collected during an unusually dry, wet, or otherwise unrepresentative period, the data may not represent the full range of potential floods at the site. The two-station comparison method is a way to adjust the logarithmic mean and standard deviation of peak flows for a short-term gaging station by regression with peak flows for a long-term (more than 10 years of record) gaging station over a concurrent period, that is, a number of years when peak flows were concurrently recorded at both gaging stations. The regression equation and the long-term logarithmic mean and standard deviation can be used to adjust the short-term statistics. If the correlation is high enough, the adjustment based on the longer, more representative, record may reduce

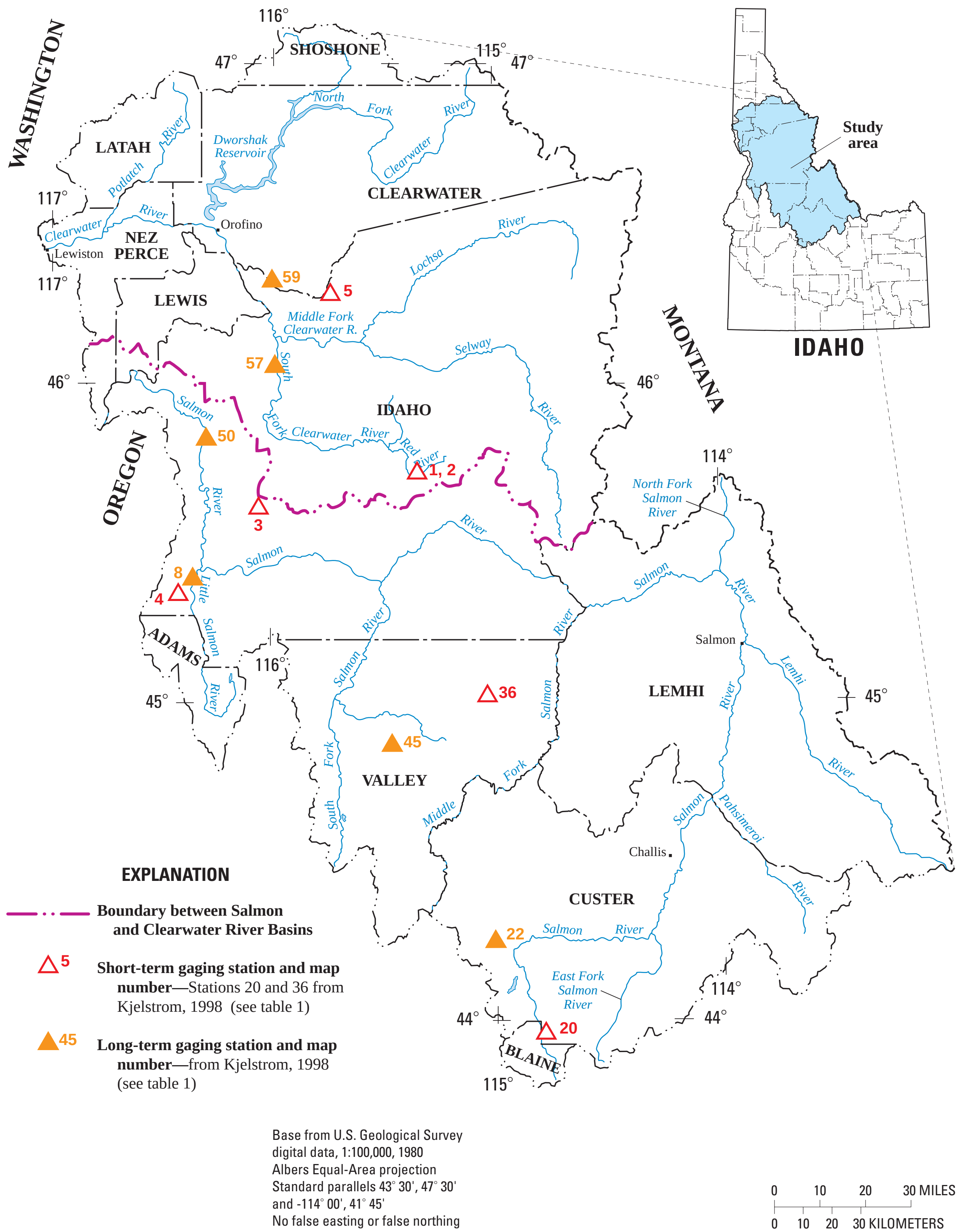

Figure 1. Locations of short-term and long-term streamflow-gaging stations in the Salmon and Clearwater River Basins, central Idaho, used for application of the two-station comparison method.

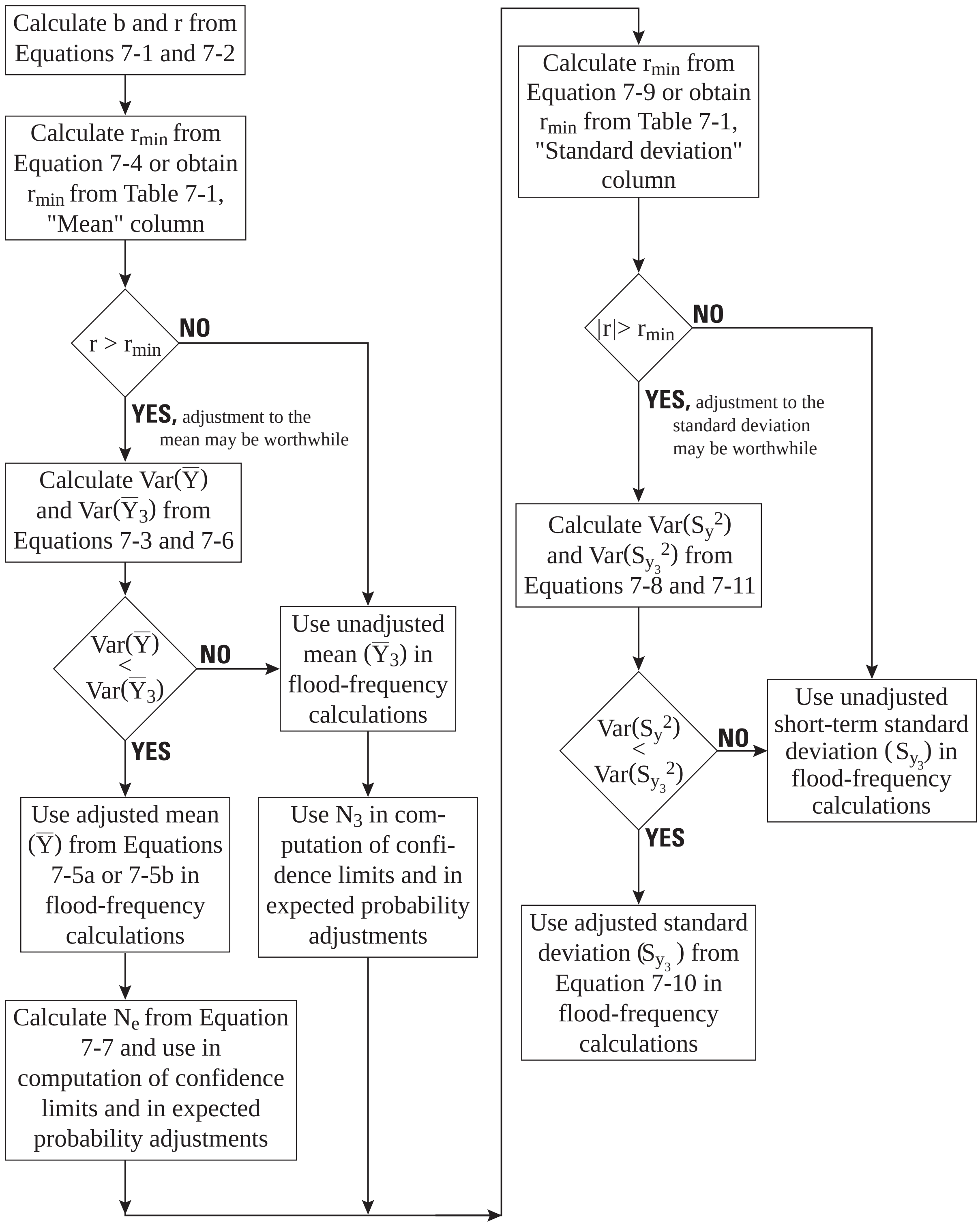

Figure 2. Flow diagram of procedure for using the two-station comparison method to adjust the logarithmic mean and logarithmic standard deviation of peak flows for flood-frequency calculations. [Equations and tables on which this method is based are shown in attachment B, back of report (appendix 7 of Bulletin 17B, Interagency Advisory Committee on Water Data, 1982)]

the possibility of bias as a result of unrepresentative peak-flow data at the short-term gaging station.

A detailed description of the procedure for using the two-station comparison method is provided in attachment B, back of report (appendix 7 of Bulletin 17B, published by the Interagency Advisory Committee on Water Data, 1982). A flow diagram outlining the procedure is shown in figure 2, and a brief description is provided in the following paragraphs. All equations and tables referenced in the following paragraphs are from Bulletin 17B.

The basis of this method is the comparison of the variances of the estimated short-term statistics (logarithmic mean and standard deviation) with the variances of the corresponding adjusted statistics. The estimates with the smaller variances are the better estimates and are the ones that should be used in the flood-frequency curve. The procedure is as follows: First, the slope of the regression line, b, and the correlation coefficient, r, are calculated from logarithmic peak-flow data for a concurrent period of record for paired short-term and long-term gaging stations. Next, the calculated r value is compared with the minimum r value (r_{min}), which is calculated from equation 7–4 or obtained from table 7–1 using the “Mean” column. If $r \leq r_{min}$, then the variance of the adjusted mean will be greater than the short-term mean (from equation 7–3). The unadjusted short-term mean is used as the estimated mean in flood-frequency calculations, and the procedure continues to the second step (standard deviation). If $r >_{rmin}$, then adjustment to the mean may be worthwhile and the procedure continues by calculating the variance of the adjusted mean [$Var(\overline{Y})$] and the variance of the unadjusted short-term mean [$Var(\overline{Y}_3)$]. The mean with the smaller variance is used as the estimated mean in subsequent flood-frequency calculations.

After the adjusted mean calculations are completed, the number of equivalent years of peak-flow data of the adjusted mean (N_e) is calculated. N_e is based on the variance of the mean and is inversely proportional to the length of record for the short-term gaging station. If r = 1 (perfect correlation), then N_e is equal to the length of record for the long-term gaging station ($N_e = N_{long\text{-}term}$). If $r \leq r_{min}$, then N_e is less than or equal to the length of the concurrent record period. N_e is used in computing the confidence limits and expected probability adjustment. For this study, N_e was not needed in any subsequent calculations but still was computed.

A second step is done to determine whether the logarithmic standard deviation of peak flow might be improved by a similar procedure. Compare r (previously calculated) with r_{min} from equation 7–9 or table 7–1 using the “Standard deviation” column. If $r \leq r_{min}$, then the variance of the adjusted standard deviation will be greater than the short-term standard deviation (from equation 7–8). The unadjusted short-term standard deviation is used as the estimated standard deviation in flood-frequency calculations. If $r > r_{min}$, adjustment to the standard deviation may be worthwhile and the procedure continues by calculating the variance of the adjusted variance (square of the standard deviation) [$Var(Sy^2)$] and variance of the unadjusted variance [$Var(Sy_3{}^2)$] for the short-term gaging station for the entire period of record. The smaller of the two variances is used as the estimated standard deviation in subsequent flood-frequency calculations.

The assumption made when using this procedure is that the logarithmic peak-flow data for the short-term and long-term gaging stations have a joint normal probability distribution. This distribution has a skewness of 0. If this assumption is not valid, then the equations used in this procedure (Interagency Advisory Committee on Water Data, appendix 7, 1982) will not be exact and discretion will be needed in application of the results. Also, the precision of r depends on the number of years of concurrent peak-flow data for the short-term and long-term gaging stations (N_1). The Interagency Advisory Committee on Water Data (1982) recommends that the concurrent period be at least 10 years.

RESULTS OF TWO-STATION COMPARISON

Seven short-term gaging stations (fig.1) were analyzed by the two-station comparison method for potential improvement of flood-frequency estimates. In an earlier analysis of flow duration and flood frequency in the Salmon and Clearwater River Basins (Kjelstrom, 1998), six hydrologic regions were delineated. For the present study, each of the seven short-term gaging stations was paired with a long-term gaging station in one of the same regions as delineated by Kjelstrom (1998).

Short-term and long-term gaging stations were paired to be consistent with pairings used in an earlier USFS flood-frequency analysis (Jack King, USFS, written commun., 2000). The one exception for the present study was the pairing of the Valley Creek at

Stanley gaging station (13295000) with the Salmon River near Obsidian gaging station (13292500). The long-term Valley Creek gaging station, instead of the Salmon River below Yankee Fork near Clayton gaging station (13296500), was paired with the short-term Obsidian gaging station for the following reasons: (1) Valley Creek peak-flow data are more strongly correlated (r) with Obsidian data than with Yankee Fork data, (2) the variance of the standard deviation for Valley Creek is lower than that for Yankee Fork, (3) the drainage area of Valley Creek is more similar to that of Obsidian than to that of Yankee Fork, and (4) the equivalent years of record for the mean are greater for Valley Creek than for Yankee Fork. The paired short-term and long-term gaging stations and the associated periods of record of annual peak-flow data are shown in table 1.

The procedure outlined in the "Method" section and shown on figure 2 was used to determine whether adjustments to the logarithmic mean and logarithmic standard deviation were needed and, if so, to perform the adjustment calculations. Calculations for one pair of gaging stations—long-term Little Salmon River at Riggins (13316500) and short-term Little Slate Creek —are shown in attachment A (back of report). Calculations for the other six paired gaging stations were performed in a similar manner. The correlations between annual peak flows for seven paired short-term and long-term gaging stations are shown in figure 3. An r value (correlation coefficient) is shown in the upper left-hand corner of the graph for each pair of stations. The correlation coefficient is a measure of strength of the linear relation between two variables (Zar, 1998). An r value of 0 indicates that there is no linear association between the two variables, whereas an r value of 1 or −1 indicates a strong linear association. The correlation coefficient of annual peak flows for the short-term South Fork Red River gaging station with those for the long-term South Fork Clearwater

Table 1. Paired short-term and long-term streamflow-gaging stations in the Salmon and Clearwater River Basins, central Idaho

[No., number; —, no streamflow-gaging station No.]

Short-term gaging station (Stations 20 and 36 from Kjelstrom, 1998)				Long-term gaging station (from Kjelstrom, 1998)			
Map No. (fig. 1)	Gaging-station name	Gaging-station No.	Period of record	Map No. (fig. 1)	Gaging-station name	Gaging-station No.	Period of record
1	South Fork Red River	—	1986, 1988–2001	57	South Fork Clearwater River at Stites	13338500	1964–2001
2	Red River.	—	1986, 1988–2001	57	South Fork Clearwater River at Stites	13338500	1964–2001
3	Little Slate Creek.	—	1986–96, 1998–2001	8	Little Salmon River at Riggins	13316500	1948, 1951–54, 1957–2001
4	Rapid River	—	1986–2001	50	Salmon River at White Bird.	13317000	1911–2001
5	Lolo Creek.	—	1989, 1991–2001	59	Lolo Creek near Greer.	13339500	1980–2001
36	Middle Fork Salmon River at Middle Fork Lodge near Yellow Pine	13309220	1973–81, 1999–2001	45	Johnson Creek at Yellow Pine	13313000	1929–2001
20	Salmon River near Obsidian	13292500	1941–52	22	Valley Creek at Stanley	13295000	1911–13, 1921–72, 1974, 1993–2001

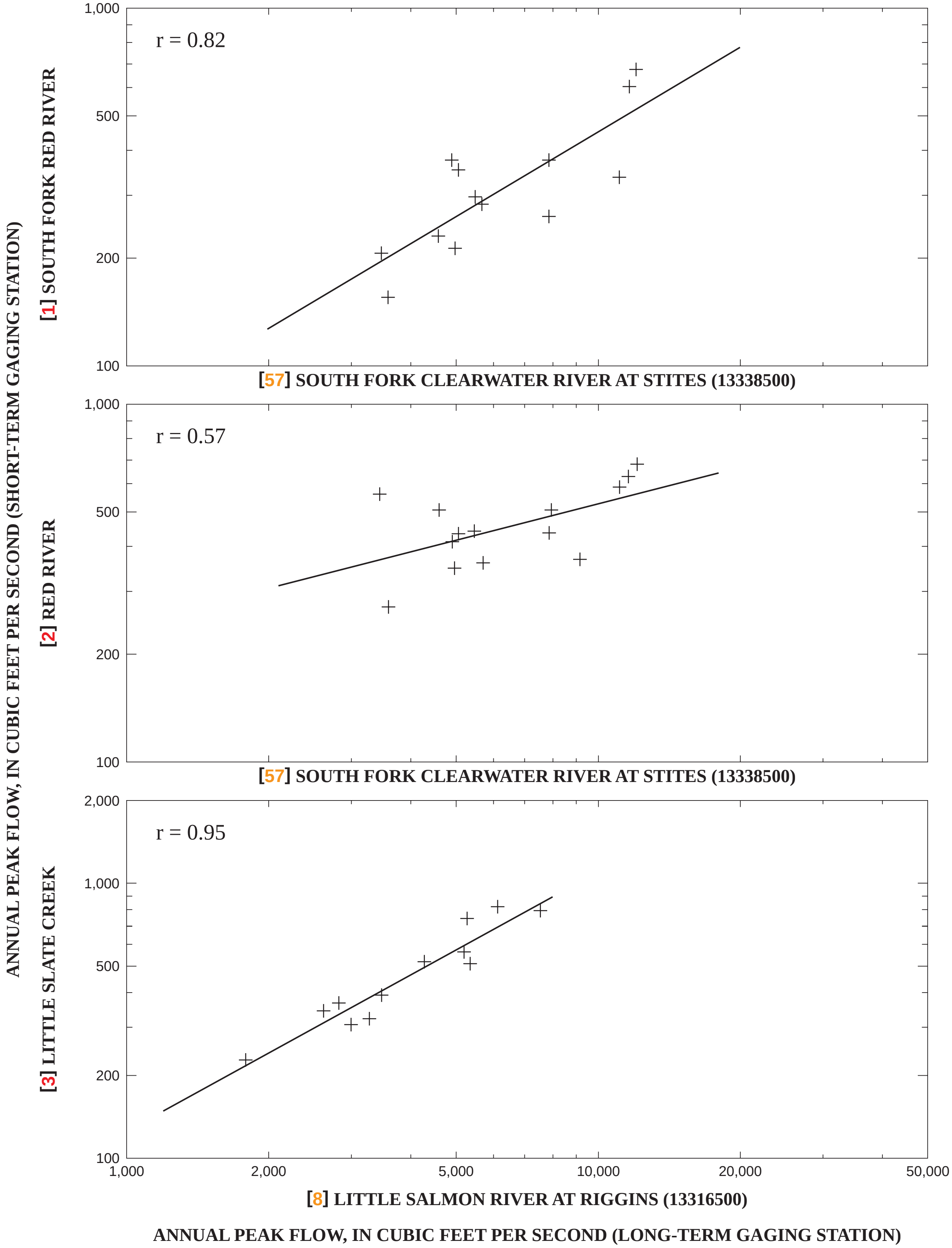

Figure 3. Correlations between annual peak flows for paired short-term and long-term streamflow-gaging stations in the Salmon and Clearwater River Basins, central Idaho. ([1] Site locations shown in figure1)

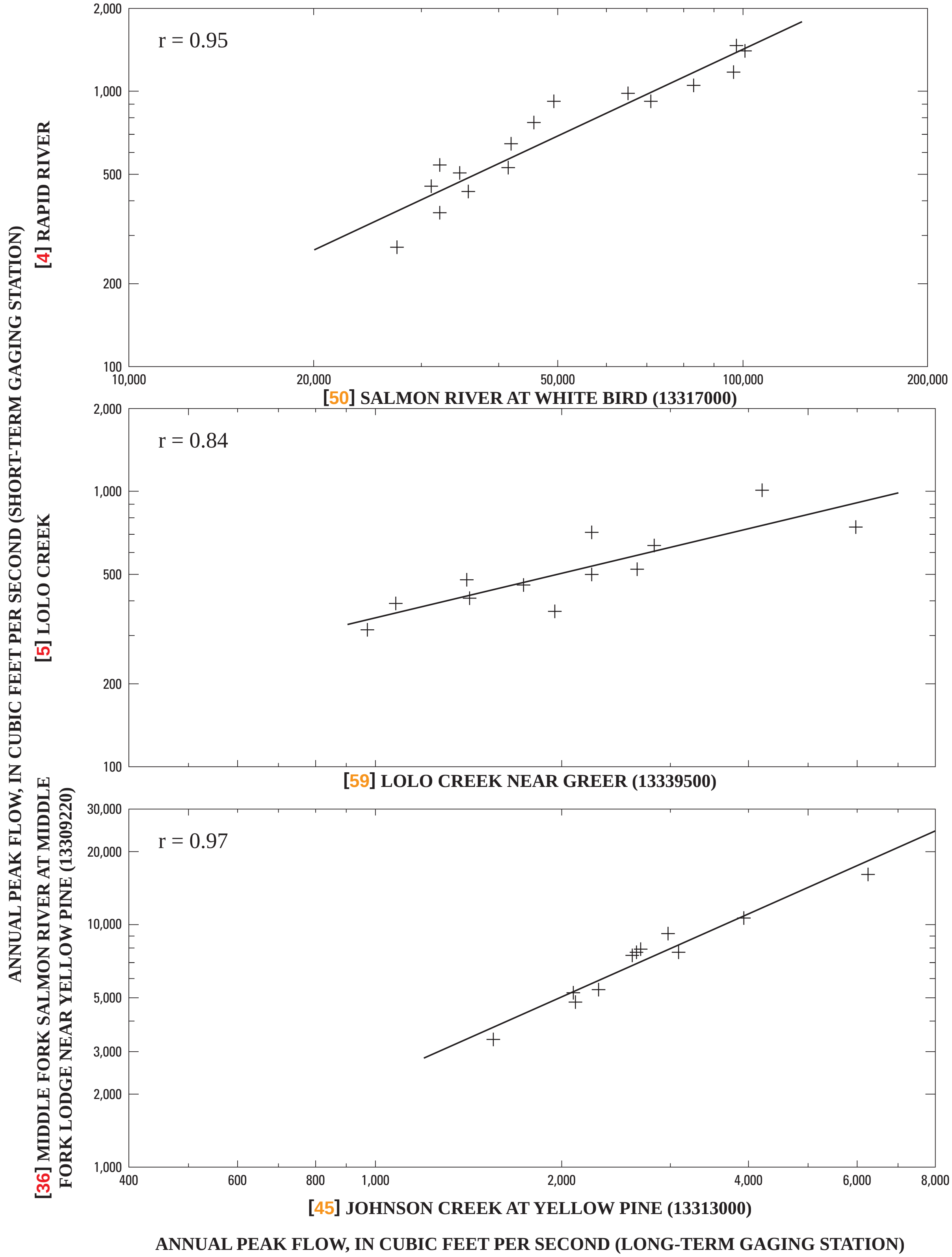

Figure 3. Correlations between annual peak flows for paired short-term and long-term streamflow-gaging stations in the Salmon and Clearwater River Basins, central Idaho—Continued. ([5] Site locations shown in figure1)

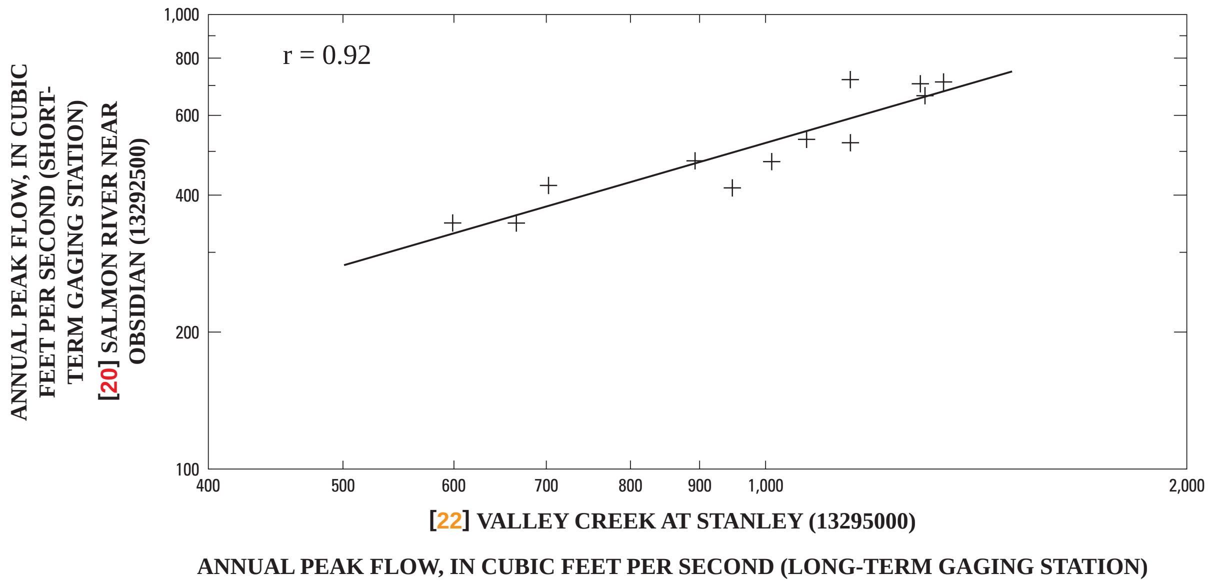

Figure 3. Correlations between annual peak flows for paired short-term and long-term streamflow-gaging stations in the Salmon and Clearwater River Basins, central Idaho—Continued. ([20] Site locations shown in figure1)

River at Stites gaging station (13338500) was 0.82. The correlation of peak flows for the other Red River gaging station with those for the South Fork Clearwater River at Stites gaging station was weaker (r = 0.57). The lower r value for this paired site is probably caused by the fair amount of scatter about the regression line. Correlations of peak flows for the other paired gaging stations were strong; the highest r value was 0.97 for the short-term Middle Fork Salmon River at Middle Fork Lodge near Yellow Pine gaging station (13309220) and the long-term Johnson Creek at Yellow Pine gaging station (13313000). Because r values were greater than their respective r_{min} values for the logarithmic mean and logarithmic standard deviation in all comparisons, adjustments were made to reduce the variance of the mean and standard deviation.

Regional skew coefficients, obtained from maps in a report by Kjelstrom and Moffatt (1981), were used to calculate a weighted skew coefficient for each of the seven short-term gaging stations. These weighted skew coefficients then were used to obtain values for the standardized variate, K (Interagency Advisory Committee on Water Data, appendix 3, 1982), for selected exceedance probabilities (attachment A). Using these K values and the final estimates of the logarithmic mean and logarithmic standard deviations, adjusted flood-frequency estimates at a range of recurrence intervals and associated exceedance probabilities for each of the seven short-term gaging stations were calculated (table 2).

Results show that all original (unadjusted) flood-frequency estimates for three short-term gaging stations (South Fork Red River [1], Red River [2], and Lolo Creek [5]) differed by less than 10 percent from the adjusted estimates (table 3). Differences for the Red River gaging station were equal to or less than 1 percent. One reason for this small difference is that the data probably were collected during periods representing the full range of flows. Unadjusted estimates for two short-term gaging stations (Rapid River [4] and Middle Fork Salmon River [36]) differed by less than 10 percent from the adjusted estimates for most exceedance probabilities (table 3). Unadjusted and adjusted flood-frequency estimates for Little Slate Creek [3] and Salmon River near Obsidian [20] differed by nearly 30 percent. For all exceedances, differences for Little Slate Creek were greater than 25.0 percent (table 3). The Salmon River near Obsidian had the largest range of differences, from 0 to 24.3 percent.

Flood-frequency curves for the long-term gaging station Johnson Creek at Yellow Pine [45] for the entire

Table 2. Adjusted peak flows at selected recurrence intervals and exceedance probabilities for short-term streamflow-gaging stations in the Salmon and Clearwater River Basins, central Idaho

[ft^3/s, cubic feet per second; No., number; %, percent]

Map No. (fig. 1)	Streamflow-gaging station name	Peak flow, in ft^3/s, for indicated recurrence interval, in years, and exceedance probability, in percent							
		2 50%	5 20%	10 10%	25 4%	50 2%	100 1%	200 0.5%	500 0.2%
1	South Fork Red River	310	449	550	686	796	911	1,030	1,210
2	Red River	455	565	632	714	771	827	880	950
3	Little Slate Creek	554	776	927	1,120	1,270	1,430	1,590	1,800
4	Rapid River	835	1,190	1,420	1,720	1,940	2,160	2,380	2,670
5	Lolo Creek	515	675	782	919	1,020	1,130	1,240	1,390
36	Middle Fork Salmon River at Middle Fork Lodge near Yellow Pine (13309220)	9,240	13,500	16,300	19,900	22,500	25,200	27,800	31,400
20	Salmon River near Obsidian (13292500)	517	688	793	918	1,010	1,090	1,170	1,280

Table 3. Percent differences between unadjusted and adjusted peak flows for short-term streamflow-gaging stations in the Salmon and Clearwater River Basins, central Idaho

[SF, South Fork; MF, Middle Fork]

Annual exceedance probability	Recurrence interval (years)	SF Red River [1]	Red River [2]	Little Slate Creek [3]	Rapid River [4]	Lolo Creek [5]	MF Salmon River (13309220) [36]	Salmon River near Obsidian (13292500) [20]
0.5	2	-3.3	0.5	25.1	19.8	-1.0	12.7	0
.2	5	-2.0	.5	26.6	12.3	1.2	9.4	7.2
.1	10	-3.8	.6	27.2	8.4	2.6	7.8	10.8
.04	25	-3.9	.8	27.6	5.5	4.0	6.2	14.8
.02	50	-4.7	.9	28.0	3.2	4.6	5.1	17.7
.01	100	-4.7	1.0	28.8	.9	6.2	4.2	19.4
.005	200	-5.4	.9	29.3	-.4	6.5	3.4	21.4
.002	500	-6.1	.8	29.5	-2.9	7.2	2.5	24.3
Average		-4.3	.7	27.7	5.8	3.9	6.4	14.4

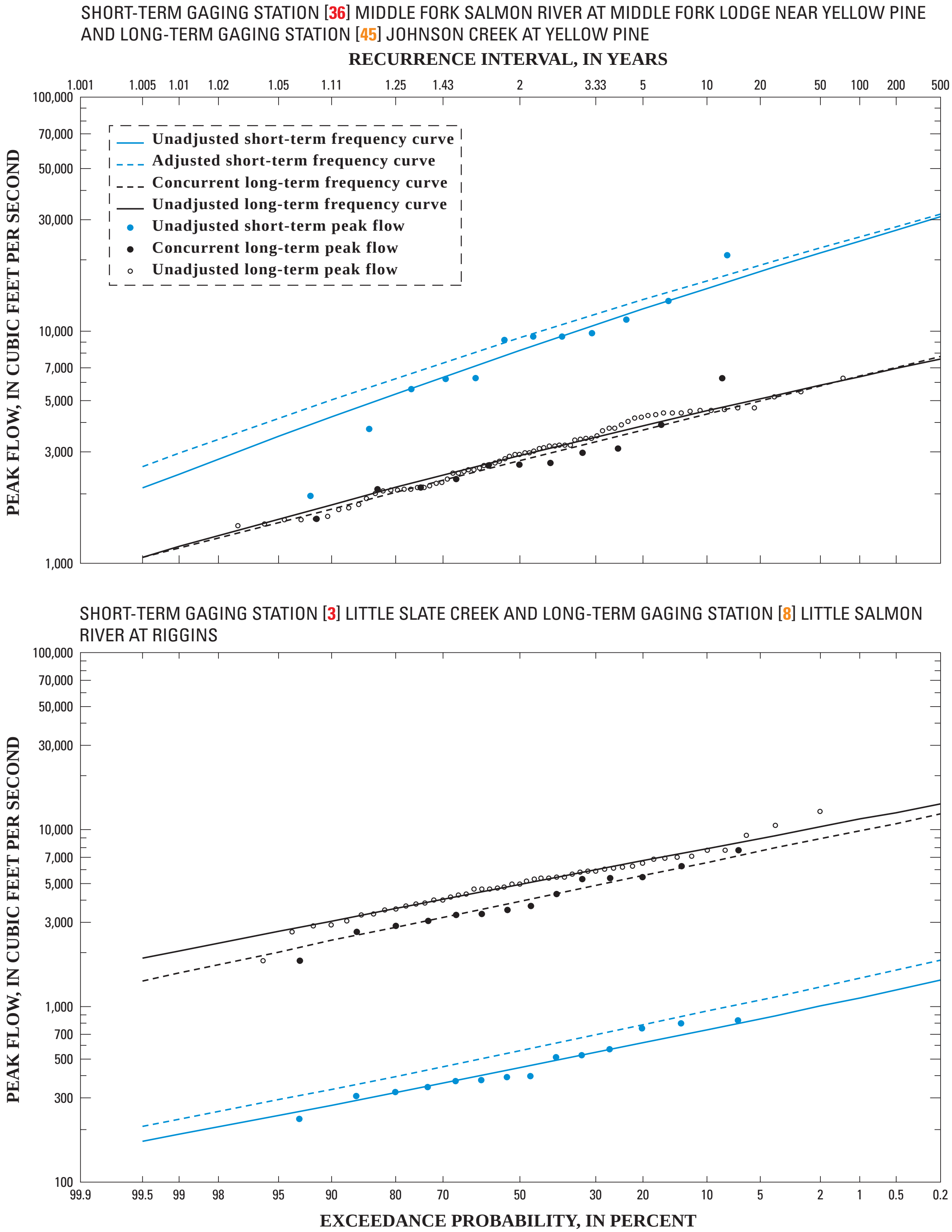

Figure 4. Unadjusted and adjusted frequency curves for two sets of paired short-term and long-term streamflow-gaging stations in the Salmon and Clearwater River Basins, central Idaho.

period of record (72 years) and concurrent period of record (11 years) are shown in figure 4. Despite the fact that the frequency curve for the concurrent period lacks 61 years of peak-flow data, the two long-term frequency curves are quite similar. The frequency curve for the concurrent period is representative of the full range of peak flows, as indicated by figure 4. Therefore, changes to the unadjusted short-term flood-frequency curve for the Middle Fork Salmon River [**36**] also will be small. Differences between the unadjusted and adjusted flood-frequency estimates for Middle Fork Salmon River (short term) ranged from 2.5 percent to 12.7 percent and averaged 6.4 percent (table 3).

Flood-frequency curves for the paired short-term gaging station Little Slate Creek [**3**] and the long-term gaging station Little Salmon River at Riggins [**8**] are shown in figure 4. Differences between the two flood-frequency curves (entire period of record and concurrent period of record) for the long-term station are quite large. In general, these large differences probably are attributable to the concurrent long-term data set having a predominance of lower peak flows (dry periods). Eight of the lowest peak flows in the concurrent long-term data set were ranked as part of the 10 lowest peak flows in the unadjusted long-term data set (fig. 4). Similar conditions exist for the long-term gaging station Johnson Creek at Yellow Pine [**45**], where peak flows in the concurrent data set were ranked in the lower half of the unadjusted data set. In fact, no peak flows in the concurrent data set were ranked in the upper 20 percent of the unadjusted data set. This indicates that data in the concurrent period probably were collected during a dry or otherwise unrepresentative period. Therefore, changes to the unadjusted flood-frequency curve for Little Slate Creek (short-term gaging station) also will be large. Differences between the unadjusted and adjusted flood-frequency estimates for Little Slate Creek ranged from 25.1 to 29.5 percent and averaged 27.7 percent (table 3).

REFERENCES CITED

Hortness, J.E., and Berenbrock, Charles, 2001, Estimating monthly and annual streamflow statistics at ungaged sites in Idaho: U.S. Geological Survey Water-Resources Investigations Report 01–4093, 36 p.

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

Kjelstrom, L.C., 1998, Methods for estimating selected flow-duration and flood-frequency characteristics at ungaged sites in central Idaho: U.S. Geological Survey Water-Resources Investigations Report 94–4120, 10 p.

Kjelstrom, L.C., and Moffatt, R.L., 1981, A method of estimating flood-frequency parameters for streams in Idaho: U.S. Geological Survey Open-File Report 81–909, 99 p.

Zar, J.H., 1998, Biostatistical analysis, 4th ed.: Englewood Cliffs, N.J., Prentice-Hall, Inc., 929 p.

Attachments

Attachment A. Calculation of peak flows using the two-station comparison method for long-term streamflow-gaging station Little Salmon River at Riggins [8] (13316500) and short-term streamflow-gaging station Little Slate Creek [3]

[Locations of gaging stations shown in figure 1; calculations are based on equations and tables in attachment B (appendix 7, Bulletin 17B, Interagency Advisory Committee on Water Data, 1982)]

List of Variables

N_1	Number of years when peak flows were concurrently recorded for long-term and short-term gaging stations (concurrent period)
N_2	Number of years when peak flows were recorded for the long-term gaging station but not recorded for the short-term gaging station (nonconcurrent period)
N_3	Number of years of peak-flow data for the short-term gaging station
N_e	Equivalent years of peak-flow data of the adjusted mean
Q_x	Peak flows for the long-term gaging station
X	Logarithm of peak flows for the long-term gaging station
$\overline{X}_1$	Mean logarithm of peak flows for the long-term gaging station for the concurrent period
$\overline{X}_2$	Mean logarithm of peak flows for the long-term gaging station for the nonconcurrent period
$\overline{X}_3$	Mean logarithm of peak flows for the long-term gaging station for the entire period
Q_y	Peak flows for the short-term gaging station
Y	Logarithm of peak flows for the short-term gaging station
$\overline{Y}$	Adjusted mean logarithm of peak flows for the short-term gaging station
$\overline{Y}_1$	Mean logarithm of peak flows for the short-term gaging station for the concurrent period
$\overline{Y}_3$	Mean logarithm of peak flows for the short-term gaging station for the entire period
Sx_1	Standard deviation of logarithm of peak flows for the long-term gaging station for the concurrent period
Sx_2	Standard deviation of logarithm of peak flows for the long-term gaging station for the nonconcurrent period
Sy	Adjusted standard deviation of logarithm of peak flows for the short-term gaging station
Sy_1	Standard deviation of logarithm of peak flows for the short-term gaging station for the concurrent period
Sy_3	Standard deviation of logarithm of peak flows for the short-term gaging station for the entire period
b	Slope of regression line
r	Correlation coefficient of the peak flows for the paired gaging stations for the concurrent period

Attachment A. Calculation of peak flows using the two-station comparison method for long-term streamflow-gaging station Little Salmon River at Riggins [8] (13316500) and short-term streamflow-gaging station Little Slate Creek [3]—Continued

Peak-flow records

[Peak flows in cubic feet per second; —, no data]

Long-term gaging station Little Salmon River at Riggins [8] (13316500)				Short-term gaging station Little Slate Creek [3]			
Water year	Peak flow (Q_x)	Logarithm of peak flow (X)	X^2	Peak flow (Q_y)	Logarithm of peak flow (Y)	Y^2	X*Y
1948	9,200	3.96379	15.71161				
1951	3,720	3.57054	12.74878				
1952	5,530	3.74273	14.00799				
1953	5,650	3.75205	14.07787				
1954	5,060	3.70415	13.72073				
1957	5,720	3.75740	14.11802				
1958	6,720	3.82737	14.64876				
1959	4,680	3.67025	13.47070				
1960	4,540	3.65706	13.37406				
1961	4,860	3.68664	13.59129				
1962	3,530	3.54777	12.58671				
1963	3,800	3.57978	12.81485				
1964	5,740	3.75891	14.12942				
1965	5,970	3.77597	14.25798				
1966	2,840	3.45332	11.92541	No data until 1986			
1967	6,080	3.78390	14.31793				
1968	3,960	3.59770	12.94341				
1969	5,280	3.72263	13.85800				
1970	6,960	3.84261	14.76565				
1971	7,500	3.87506	15.01610				
1972	6,760	3.82995	14.66849				
1973	4,190	3.62221	13.12043				
1974	12,600	4.10037	16.81304				
1975	4,600	3.66276	13.41579				
1976	4,870	3.68753	13.59787				
1977	1,510	3.17898	10.10589				
1978	4,570	3.65992	13.39499				
1979	4,100	3.61278	13.05221				
1980	3,960	3.59770	12.94341				
1981	5,290	3.72346	13.86412				
1982	6,400	3.80618	14.48701				
1983	6,850	3.83569	14.71252				
1984	5,860	3.76790	14.19705				
1985	4,540	3.65706	13.37406				

Attachment A. Calculation of peak flows using the two-station comparison method for long-term streamflow-gaging station Little Salmon River at Riggins [8] (13316500) and short-term streamflow-gaging station Little Slate Creek [3]—Continued

Peak-flow records—Continued

Long-term gaging station Little Salmon River at Riggins [8] (13316500)				Short-term gaging station Little Slate Creek [3]			
Water year	Peak flow (Q_x)	Logarithm of peak flow (X)	X^2	Peak flow (Q_y)	Logarithm of peak flow (Y)	Y^2	X*Y
1986	6,100	3.78533	14.32872	818	2.91275	8.48413	11.02573
1987	3,000	3.47712	12.09037	306	2.48572	6.17881	8.64315
1988	2,820	3.45025	11.90422	369	2.56703	6.58962	8.85688
1989	4,280	3.63144	13.18738	517	2.71349	7.36303	9.85389
1990	3,470	3.54033	12.53393	390	2.59106	6.71362	9.17322
1991	3,260	3.51322	12.34270	323	2.50920	6.29610	8.81537
1992	1,790	3.25285	10.58105	227	2.35603	5.55086	7.66381
1993	5,250	3.72016	13.83959	1,366	3.13545	9.83105	11.66438
1994	2,610	3.41664	11.67343	345	2.53782	6.44053	8.67082
1995	5,190	3.71517	13.80247	564	2.75128	7.56954	10.22146
1996	7,520	3.87622	15.02506	791	2.89818	8.39943	11.23396
1997	10,500	4.02119	16.16996	—	—	—	—
1998	5,350	3.72835	13.90062	509	2.70672	7.32632	10.09160
1999	5,280	3.72263	13.85800	740	2.86923	8.23249	10.68110
2000	3,660	3.56348	12.69840	397	2.59879	6.75371	9.26074
2001	3,300	3.51851	12.37994	376	2.57519	6.63159	9.06083

Descriptive statistics

(These values reflect censoring of peak flows for high and low outliers as recommended by Bulletin 17B)

N_1 = 14 (1986–92, 1994–96, 1998–2001)

N_2 = 34 (1948, 1951–54, 1957–76, 1978–85, 1997)

N_3 = 14 (1986–92, 1994–96, 1998–2001)

$\overline{X}_1 = 3.58511$

$\overline{X}_2 = 3.73101$

$\overline{X}_3 = 3.68846$

$\overline{Y}_1 = 2.64803$

$\overline{Y}_3 = 2.64803$

$Sx_1 = 0.16734$

$Sx_2 = 0.15951$

$Sy_1 = 0.16654$

$Sy_3 = 0.16654$

Calculate b and r

Equation 7–1 $b = 0.94388$

Equation 7–2 $r = 0.94843$

Criterion and adjustment procedure for the mean

Equation 7–4 $r = 0.94843 > 0.28867 = r_{min}$

Because $r > r_{min}$, adjustment to the mean is worthwhile.

Attachment A. Calculation of peak flows using the two-station comparison method for long-term streamflow-gaging station Little Salmon River at Riggins [8] (13316500) and short-term streamflow-gaging station Little Slate Creek [3]—Continued

Calculate Var($\overline{Y}$) and Var ($\overline{Y}_3$)

Equation 7–3 $\quad Var(\overline{Y}) = 0.00073158$

Equation 7–5b $\quad \overline{Y} = 2.74558$

Equation 7–6 $\quad Var(\overline{Y}_3) = 0.0019801$

If $Var(\overline{Y}) < Var(\overline{Y}_3)$, then use $\overline{Y}$. Otherwise, use $\overline{Y}_3$.

Final estimate of the mean

$\overline{Y} = 2.74558$ (from equation 7–5b)

Equivalent years of record for the mean

Equation 7–7 $\quad N_e = 37.9$

Criterion and adjustment procedure for the standard deviation

Equation 7–9 $\quad |r| = 0.94843 > 0.55500 = r_{min}$

Because $r > r_{min}$, adjustment to the standard deviation is worthwhile.

where A = -9.59779, B = 2.10468, C = 0.26234

Equation 7–8 $\quad Var(Sy^2) = 5.1912E\text{-}05$

Equation 7–10 $\quad Sy^2 = 0.029574$

$Sy = 0.17197$

Equation 7–11 $\quad Var(Sy_3^2) = 0.00011833$

If $Var(Sy^2) < Var(Sy_3^2)$, use Sy. Otherwise, use Sy_3.

Final estimate of standard deviation

Sy = 0.17197 (from equation 7–10)

Peak-flow estimates

[ft^3/s, cubic feet per second]

Exceedance probability	Recurrence interval (years)	Peak flow (ft^3/s)		Percent difference from unadjusted
		Unadjusted	Adjusted	
0.5	2	443	554	25.1
0.2	5	613	776	26.6
0.1	10	729	927	27.2
0.04	25	878	1,120	27.6
0.02	50	992	1,270	28.0
0.01	100	1,110	1,430	28.8
0.005	200	1,230	1,590	29.3
0.002	500	1,390	1,800	29.5

An original PDF file containing Appendix 7 is located at the FEMA website: http://www.fema.gov/mit/tsd/dl_flow.htm

APPENDIX 7

*

TWO STATION COMPARISON

INTRODUCTION

The procedure outlined herein is recommended for use in adjusting the logarithmic mean and standard deviation of a short record on the basis of a regression analysis with a nearby long-term record. The theoretical basis for the equations provided herein were developed by Matalas and Jacobs (29).

The first step of the procedure is to correlate observed peak flows for the short record with concurrent observed peak flows for the long record. The regression and correlation coefficients, respectively, can becomputed by the following two equations:

$$b = \frac{\Sigma X_1 Y_1 - \Sigma X_1 \Sigma Y_1 / N_1}{\Sigma X_1^2 - (\Sigma X_1)^2 / N_1} \qquad (7\text{-}1)$$

$$r = b \frac{S_{x_1}}{S_{y_1}} \qquad (7\text{-}2)$$

where the terms are defined at the end of this Appendix.

If the correlation coefficient defined by equation 7-2 meets certain criteria, then improved estimates of the short record mean and standard deviation can be made. Both of these statistics can be improved when the variance of that statistic is reduced. As each statistic is evaluated separately, only one adjustment may be worthwhile. The criterion and adjustment procedure for each statistic are discussed separately. In each discussion, two cases are considered: (1) entire short record contained in the long record, (2) only part of the short record contained in the long record. The steps for case 2 include all of those for case 1 plus an additional one.

*

*

CRITERION AND ADJUSTMENT PROCEDURE FOR MEAN

The variance of the adjusted mean ($\bar{Y}$) can be determined by equation 7-3:

$$Var(\bar{Y}) = \frac{(S_{Y_1})^2}{N_1} \left[1 - \frac{N_2}{N_1+N_2} \left(r^2 - \frac{(1-r^2)}{(N_1-3)} \right) \right] \quad (7\text{-}3)$$

Since $(S_{y_1})^2/N_1$ is the variance of $\bar{Y}_1$, the short-record mean, $\bar{Y}$ will be a better estimate of the true mean than $\bar{Y}_1$ if the term $r^2 - \frac{1-r^2}{N_1-3}$ in equation 7-3 is negative. Solving this relationship for r yields equation 7-4. If the correlation coefficient satisfies equation 7-4,

$$|r| > 1/(N_1-2)^{1/2} \quad (7\text{-}4)$$

then an adjustment to the mean is worthwhile. The right side of this inequality represents the minimum critical value of r. Table 7-1 contains minimum critical values of r for various values of N_1. The adjusted logarithmic mean can be computed using equation 7-5a or 7-5b.

$$\bar{Y} = \bar{Y}_1 + \frac{N_2}{N_1+N_2} \left[b \ (\bar{X}_2 - \bar{X}_1) \right] \quad (7\text{-}5a)$$

$$\bar{Y} = \bar{Y}_1 + b(\bar{X}_3 - \bar{X}_1) \quad (7\text{-}5b)$$

Equation 7-5b saves recomputing a new $\bar{X}_2$ at the long record station for each short record station that is being correlated with the long record station. While the adjusted mean from equation 7-5a or 7-5b may be an improved estimate of the mean obtained from the concurrent period, it may not be an improvement over the entire short record mean in case 2. It is necessary to compare the variance of the adjusted an (equation 7-3) to the variance of the mean ($\bar{Y}_3$) for the entire short record period (N_3). Compute the variance of the mean $\bar{Y}_3$ using equation 7-6:

$$Var(\bar{Y}_3) = \frac{(S_{y_3})^2}{N_3} \quad (7\text{-}6)$$

*

*

where S_{y_3} is the standard deviation of flows for the short record site for the period N_3. If the variance of equation 7-6 is smaller than the variance of $\bar{Y}$ given in equation 7-3, then use $\bar{Y}_3$ as the final estimate of the mean. Otherwise, use the value of $\bar{Y}$ computed in equation 7-5a or 7-5b.

EQUIVALENT YEARS OF RECORD FOR THE MEAN

As illustrated in equations 7-3 and 7-6, the variance of the mean is inversely proportional to the record length at the site. Using equation 7-3 it can be shown that the equivalent years of record, N_e, for the adjusted mean is:

$$N_e = \frac{N_1}{1 - \frac{N_2}{N_1+N_2}\left(r^2 - \frac{(1-r^2)}{(N_1-3)}\right)} \qquad (7\text{-}7)$$

It may be seen from equation 7-7 that when there is no correlation (r=O), then N_e is less than N_1. This indicates that the correlation technique can actually decrease the equivalent years of record unless r satisfies equation 7-4. For perfect correlation (r=l), then $N_e = N_1 + N_2$, the total record length at the long record site.

Although N_e is actually the equivalent years of record for the mean, it is recommended that N_e be used as an estimate of the equivalent years of record for the various exceedance probability floods in the computation of confidence 1imits and in applying the expected probability adjustment.

CRITERION AND ADJUSTMENT PROCEDURE FOR THE STANDARD DEVIATION

The variance of the adjusted variance S_y^2 (square of the standard deviation) can be determined by equation 7-8:

*

*

$$\mathrm{Var}(S_y^2) = \frac{2(S_{y_1})^4}{N_1-1} + \frac{N_2 (S_{y_1})^4}{(N_1+N_2-1)^2}\left[Ar^4 + Br^2 + C\right] \qquad (7\text{-}8)$$

where A, B, and C are defined below and the other terms are defined at the end of the appendix. In equation 7-8, $2(S_{y_1})^4/N_1-1)$ is the variance of $S_{y_1}^2$ (the short-record variance). If the second term in equation 7-8 is negative, then the variance of S_y^2 will be less than the variance of $S_{y}^2{}_1$. Solving this relationship for r yields the following equation:

$$|r| > \left[\frac{-B \pm \sqrt{B^2-4AC}}{2A}\right]^{1/2} \qquad (7\text{-}9)$$

where

$$A = \frac{(N_2+2)(N_1-6)(N_1-8)}{(N_1-3)(N_1-5)} - \frac{8(N_1-4)}{(N_1-3)} - \frac{2N_2(N_1-4)^2}{(N_1-3)^2} + \frac{N_1N_2(N_1-4)^2}{(N_1-3)^2(N_1-2)} + \frac{4(N_1-4)}{(N_1-3)}$$

$$B = \frac{6(N_2+2)(N_1-6)}{(N_1-3)(N_1-5)} + \frac{2(N_1^2-N_1-14)}{(N_1-3)} + \frac{2N_2(N_1-4)(N_1-5)}{(N_1-3)^2} - \frac{2(N_1-4)(N_1+3)}{(N_1-3)} - \frac{2N_1N_2(N_1-4)^2}{(N_1-3)^2(N_1-2)}$$

$$C = \frac{2(N_1+1)}{N_1-3} + \frac{3(N_2+2)}{(N_1-3)(N_1-5)} - \frac{(N_1+1)(2N_1+N_2-2)}{N_1-1} + \frac{2N_2(N_1-4)}{(N_1-3)^2} + \frac{2(N_1-4)(N_1+1)}{(N_1-3)} + \frac{N_1N_2(N_1-4)^2}{(N_1-3)^2(N_1-2)}$$

*

*

The right side of the inequality (7-9) represents the minimum critical value of r. Table 7-1 gives approximate minimum critical values of r for various values of N_1. The table values are an approximation as they are solutions of equation 7-9 for a constant N_2. The variations in N_2 only affect the table values slightly.

If the correlation coefficient satisfies equation 7-9, then the adjusted variance can be computed by equation 7-10:

$$S_y^2 = \frac{1}{(N_1+N_2-1)} \Big[(N_1-1)S_{y_1}^2 + (N_2-1)b^2 S_{x_2}^2 + \frac{N_2(N_1-4)(N_1-1)}{(N_1-3)(N_1-2)} (1-r^2)S_{y_1}^2 + \frac{N_1 N_2}{N_1+N_2} b^2 (\bar{X}_2 - \bar{X}_1)^2 \Big] \quad (7\text{-}10)$$

The adjusted standard deviation S_y equals the square root of the adjusted variance in equation 7-10. The third term in brackets in equation 7-10 is an adjustment factor to give an unbiased estimate of S_y^2. This adjustment is equivalent to adding random noise to each estimated value of slow at the short-term site.

While the adjusted variance from equation 7-10 may be an improved estimate of the variance (standard deviation) obtained from the concurrent period. It may not be an improvement over the entire short record variance (standard deviation) in case 2. It is necessary to compare the variance of the adjusted variance (equation 7-8) to the variance of the variances ($S_{y_3}^2$) for the entire period (N_3). Compute the variance of the short-record variance ($S_{y_3}^2$) using equation 7-11.

$$\text{Var}\left(S_{y_3}\right)^2 = \frac{2\left(S_{y_3}\right)^4}{N_3-1} \quad (7\text{-}11)$$

*

*

where all terms are previously defined. If the variance of equation 7-11 is smaller than the variance of S_y^2 given in equation 7-8, then use S_{y_2} as the final estimate of the standard deviation. Otherwise, use the value of S_y determined from equation 7-10.

FURTHER CONSIDERATIONS

The above equations were developed under the assumption that the concurrent observations of flows at the short and long-term sites have a joint normal probability distribution with a skewness of zero. When this assumption is seriously violated, the above equations are not exact and this technique should be used with caution. In addition, the reliability of r depends on the length of the concurrent period, N_1. To obtain a reliable estimate of r, N_1 should be at least 10 years.

Notice that it is not necessary to estimate the actual annual peaks from the regression equation but only the adjusted logarithmic mean and standard deviation. The adjusted skew coefficient should be computed by weighting the generalized skew with the skew computed from the short record site as described in Section V.B.4.

*

NOTATION

*

N_1 = Number of years when flows were concurrently observed at the two sites

N_2 = Number of years when flows were observed at the longer record site but not observed at the short record site

N_3 = Number of years of flow at the short record site

N_e = Equivalent years of record of the adjusted mean

S_y = Standard deviation of the logarithm of flows for the extended period at the short record site

S_{x_1} = Standard deviation of logarithm of flows at the long record site during concurrent period

S_{x_2} = Standard deviation of logarithm of flows at the long record site for the period when flows were not observed at the short record site

S_{y_1} = Standard deviation of the logarithm of flows at the short record site for the concurrent period

S_{y_2} = not used

S_{y_3} = Standard deviation of logarithm of flows for the entire period at the short record site

X_1 = Logarithms of flows from long record during concurrent period

$\bar{X}_1$ = Mean logarithm of flows at the long record site for the concurrent period

$\bar{X}_2$ = Mean logarithm of flows at the long record site for the period when flow records are not available at the short record site

$\bar{X}_3$ = Mean logarithm of flows for the entire period at the long record site

Y_1 = Logarithms of flows from short record during concurrent period

$\bar{Y}$ = Mean logarithm of flows for the extended period at the short record site

$\bar{Y}_1$ = Mean logarithm of flows for the period of observed flow at the short record site (concurrent period)

$\bar{Y}_2$ = not used

*

*

$\bar{Y}_3$ = Mean logarithm of flows for the entire period at the short record site

b = Regression coefficient for Y_1 on X_1

r = Correlation coefficient of the flows at the two sites for concurrent periods

*

7-8

*

TABLE 7-1 MINIMUM r VALUES FOR IMPROVING

MEAN OR STANDARD DEVIATION ESTIMATES

CONCURRENT RECORD	MEAN	STANDARD DEVIATION
10	0.35	0.65
11	0.33	0.62
12	0.32	0.59
13	0.30	0.57
14	0.29	0.55
15	0.28	0.54
16	0.27	0.52
17	0.26	0.50
18	0.25	0.49
19	0.24	0.48
20	0.24	0.47
21	0.23	0.46
22	0.22	0.45
23	0.22	0.44
24	0.21	0.43
25	0.21	0.42
26	0.20	0.41
27	0.20	0.41
28	0.20	0.40
29	0.19	0.39
30	0.19	0.39
31	0.19	0.38
32	0.18	0.37
33	0.18	0.37
34	0.18	0.36
35	0.17	0.36

*

Berenbrock—Two-Station Comparison of Peak Flows, Salmon and Clearwater River Basins, Central Idaho—U.S. Geological Survey Water-Resources Investigations Report 03–4001

Printed on recycled paper