*

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 be computed 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.

*

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

$$\mathrm{Var}(\overline{Y}) = \frac{\left(S_{y_1}\right)^2}{N_1}\left[1 - \frac{N_2}{N_1 + N_2}\left(r^2 - \frac{(1-r^2)}{(N_1-3)}\right)\right] \quad (7\text{-}3)$$

Since $(S_{y_1})^2/N_1$ is the variance of $\overline{Y}_1$, the short-record mean, $\overline{Y}$ will be a better estimate of the true mean than $\overline{Y}_1$ if the term $r^2 - \frac{1-r^2}{N_1-3}$ in equation 7-3 is positive. 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.

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

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

Equation 7-5b saves recomputing a new $\overline{X}_2$ 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 mean (equation 7-3) to the variance of the mean ($\overline{Y}_3$) for the entire short record period (N_3). Compute the variance of the mean $\overline{Y}_3$ using equation 7-6:

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

*

*

where S_{y_3} is the standard deviation of the logarithms 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 ($r=0$), 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 ($r=1$), 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 limits 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:

*

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

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 $S_{y_1}{}^2$. 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:

$$S_y^{\,2} = \frac{1}{(N_1+N_2-1)} \Big[(N_1-1)S_{y_1}^{\,2} + (N_2-1)b^2 S_{x_2}^{\,2}$$

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

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 flow 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 variance ($S_{y_3}^{\,2}$) for the entire period (N_3). Compute the variance of the short-record variance ($S_{y_3}^{\,2}$) using equation 7-11.

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

*

*

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_3} 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

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

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

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

Y_1 = Logarithms of flows from short record during concurrent period

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

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

$\overline{Y}_2$ = not used

*

*

$\overline{Y}_3$ = 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

*

*

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

*