WATER RESOURCES RESEARCH, VOL. 21, NO. 5, PAGES 715–723, MAY 1985

Minimum Variance Streamflow Record Augmentation Procedures

RICHARD M. VOGEL

Department of Civil Engineering, Tufts University, Medford, Massachusetts

JERY R. STEDINGER

Department of Environmental Engineering, Cornell University, Ithaca, New York

Traditional approaches for augmentation of short-streamflow records have exploited the cross-correlation among flows at two or more gages to obtain maximum likelihood estimates of the mean and variance of the flows at the short-record gage. While such estimators asymptotically have minimum variance among all asymptotically unbiased estimators, they are not necessarily optimal for the small samples of interest in hydrology. Improved estimators of the mean and variance of the flows at the short-record gage are developed. The gains in information transfer associated with the use of both the traditional and improved estimators of the mean and variance are documented. Monte-Carlo results illustrate the performance of these estimators when the true cross-correlation of the flows must be estimated. The information transfer gains when the true cross-correlation must be estimated are comparable to the gains achieved when the cross-correlation is known. The potential advantages and limitations of these new estimators are discussed within the framework of augmentation and/or extension of both peak annual floods and also monthly streamflow records.

INTRODUCTION

Since their introduction into the water resources literature by *Fiering* [1963] and *Matalas and Jacobs* [1964], augmentation procedures have been recognized as a useful approach for estimating the mean and variance of short hydrologic records. By employing the cross-correlation between a long x record and a short y record, *Fiering* [1962, 1963] made use of regression equations to estimate the mean μ_y and variance σ_y^2 of the flows at the short-record gage when the observations are independent across time and arise from a joint normal distribution. Those estimators, which are also the maximum likelihood estimators, were originally developed by *Wilks* [1932] and extended by *Cochran* [1953] and *Anderson* [1957]. *Morrison* [1971] later derived the expectations and the complete covariance matrix of the estimators. An important contribution was made by *Matalas and Jacobs* [1964], as well as *Moran* [1974], when they derived the unbiasing factor for the estimator of the variance of the flows at the short-record gage as well as an expression for Var $[\hat{\sigma}_y^2]$.

When more than one long-streamflow record is available for use in augmenting a short record, trivariate [*Fiering*, 1962, 1963] and multivariate record extension procedures [*Gilroy*, 1970; *Moran*, 1974] have been developed; unfortunately, Gilroy's results were not entirely correct [see *Moran*, 1974, pp. 83–85]. All of these studies assume that the individual series are without serial correlation. When hydrologic data exhibit serial correlation, the sampling properties of the estimators of the mean and variance at the short-record site will depend upon the parameters and the form of the underlying stochastic process [see *Frost and Clark*, 1973; *Natural Environment Research Council*, chapter 3, 1975; *Matalas and Langbein*, 1962; *Salas-La Cruz*, 1972; *Wallis and Matalas*, 1972].

The cited studies are concerned primarily with efficient estimation of the mean and variance of flows at the short-record gage. This is termed record augmentation. Another goal would be to actually extend monthly, weekly, or daily streamflow records for use in sequential analyses such as reservoir design and reservoir operations studies. For these purposes it is not sufficient to obtain efficient estimates of the mean and variance of the flows at the short record gage. The problem of actually extending streamflow records has been addressed by *Hirsch* [1979, 1982] and by *Alley and Burns* [1983]. This issue is discussed later.

This study develops unbiased estimators $\hat{\mu}_y{}^*$ of the mean and $\hat{\sigma}_y{}^{*2}$ of the variance of the short-record y series which can have lower variance in the small samples of interest in hydrology than the corresponding unbiased maximum likelihood estimators. These estimators also circumvent the standard tests needed to determine if the variance of the at-site (i.e., based solely on the y series) estimators $\bar{y}$ and s_y^2 of the mean and variance are greater or less than the variance of the corresponding maximum likelihood estimators (MLE's) $\hat{\mu}_y$ and $\hat{\sigma}_y^2$. A Monte-Carlo experiment is performed to evaluate the performance of these new estimators relative to the MLE's as well as the at-site sample estimators. The advantages and limitations of the improved estimators are discussed within the framework of extension of both annual flood peaks and monthly streamflows.

UNBIASED MAXIMUM LIKELIHOOD ESTIMATORS OF THE MEAN AND VARIANCE

This section reviews the unbiased maximum likelihood estimators for the mean and variance of flows at the short-record gage when the two series arise from a bivariate normal population. The observed events are denoted by

$$x_1, \cdots, x_{n_1}, x_{n_1+1}, \cdots, x_{n_1+n_2}$$

$$y_1, \cdots, y_{n_1}$$

where n_1 is the length of the short record, and $n_1 + n_2$ is the length of the long record. The n_1 concurrent observations need not correspond to the first n_1 observations nor do they need to be consecutive; however, no loss of generality arises from the above representation. The original series need not be normally distributed; however, we assume that a suitable

Paper number 5W0064.
0043-1397/85/005W-0064$05.00

transformation to normality is available. The advantages of using a logarithmic transformation, for example, have been documented elsewhere [*Stedinger*, 1980, 1981, 1983]. Both series are assumed to be serially independent.

A historical review of the development of the unbiased MLE of μ_y and the (biased) MLE of σ_y^2 is given by *Fiering* [1963]. *Matalas and Jacobs* [1964] developed a procedure for obtaining the unbiased MLE of both μ_y and σ_y^2. Their approach is based upon the relationship

$$y_i = \mu_y + \beta(x_i - \mu_x) + (1 - \rho^2)^{1/2}\sigma_y\varepsilon_i \tag{1}$$

between the x and y values, where ε_i is a random variable with zero mean and unit variance. Here $\rho = \beta\sigma_x/\sigma_y$. Use of least squares estimators leads to the estimated relationship between possible values $\tilde{y}_i$ of y_i for a given x_i:

$$\tilde{y}_i = \bar{y}_1 + \hat{\beta}(x_i - \bar{x}_1) + \alpha(1 - \hat{\rho}^2)^{1/2}s_{y_1}e_i \tag{2}$$

where

$$\bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i$$

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i$$

$$s_{y_1}^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(y_i - \bar{y}_1)^2$$

$$s_{x_1}^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n_1}(x_i - \bar{x}_1)(y_i - \bar{y}_1)}{\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2}$$

$$\hat{\rho} = \hat{\beta}\frac{s_{x_1}}{s_{y_1}}$$

with the e_i independent standard normal variates, and α is a constant used to make the expected sample variance of the $\tilde{y}_i$ equal its population value, which is the purpose of the noise term in (2). Use of (2) to derive estimators for μ_y and σ_y^2 results in terms which include the sample mean and variance of the generated e_i as well as the sample covariance between x_i and e_i. *Matalas and Jacobs* [1964] substituted the expectation of those terms for their sample moments to obtain the unbiased estimators of μ_y and σ_y^2. The unbiased estimator of the mean of the complete extended record is then

$$\hat{\mu}_y = \bar{y}_1 + \frac{n_2}{n_1 + n_2}\hat{\beta}(\bar{x}_2 - \bar{x}_1) \tag{3}$$

where

$$\bar{x}_2 = \frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2} x_i$$

This is also the MLE given by *Wilks* [1932], *Cochran* [1953], *Anderson* [1957], and *Moran* [1974].

Matalas and Jacobs [1964] derived the unbiased estimator of the variance:

$$\hat{\sigma}_y^2 = \frac{1}{n_1 + n_2 - 1}\left\{(n_1 - 1)s_{y_1}^2 + (n_2 - 1)\hat{\beta}^2 s_{x_2}^2 + (n_2 - 1)\alpha^2(1 - \hat{\rho}^2)s_{y_1}^2 + \frac{n_1 n_2}{(n_1 + n_2)}\hat{\beta}^2(\bar{x}_2 - \bar{x}_1)^2\right\} \tag{4}$$

where

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

and

$$s_{x_2}^2 = \frac{1}{n_2 - 1}\sum_{i=n_1+1}^{n_1+n_2}(x_i - \bar{x}_2)^2$$

Equation (4) is identical to the unbiased MLE developed by *Moran* [1974].

An Improved Estimator of the Mean

The variance of the unbiased MLE of the mean derived by *Cochran* [1953], H. A. Thomas, Jr. (unpublished manuscript, 1958), and *Morrison* [1971], is

$$\text{Var}[\hat{\mu}_y] = \frac{\sigma_y^2}{n_1}\left[1 - \frac{n_2}{(n_1 + n_2)}\left(\rho^2 - \frac{(1 - \rho^2)}{(n_1 - 3)}\right)\right] \tag{5}$$

Hence $\hat{\mu}_y$ in (3) has a larger variance than $\bar{y}_1$, the short-sample mean, if

$$\rho^2 < (n_1 - 2)^{-1} \tag{6}$$

Only asymptotically as n_1 and n_2 become large has one any assurance that $\hat{\mu}_y$ has minimum variance among all unbiased estimators. In fact, when (6) holds, $\bar{y}_1$ is both unbiased and has smaller variance.

Instead of $\hat{\mu}_y$ we consider the minimum variance linear estimator $\hat{\mu}_y^*$ of the form

$$\hat{\mu}_y^* = (1 - \theta_1)\bar{y}_1 + \theta_1\hat{\mu}_y \tag{7a}$$

$$= \bar{y}_1 + \theta_1\left(\frac{n_2}{n_1 + n_2}\right)\hat{\beta}(\bar{x}_2 - \bar{x}_1) \tag{7b}$$

for some θ_1. Note the second term in (7*b*) has expectation zero and is negatively correlated with $\bar{y}_1$ for $\rho > 0$.

The value of θ_1 which minimizes the variance of $\hat{\mu}_y^*$ is derived in the appendix. The required value is

$$\theta_1^* = \frac{(n_1 - 3)\rho^2}{(n_1 - 4)\rho^2 + 1} \tag{8}$$

which can be used for $n_1 \geq 4$ and which depends only on our estimate of the cross-correlation ρ of the concurrent flows. Figure 1 illustrates the value of θ_1^* as a function of n_1 and ρ; θ_1^* is often substantially less than unity. Even for $\rho = 0.7$ and $n_1 = 6$, θ_1^* is only 0.74. Clearly, $\hat{\mu}_y^*$ may differ appreciably from $\hat{\mu}_y$ which always employs $\theta_1 = 1$.

Of primary interest here are the information transfer gains which may be obtained by the two augmentation procedures.

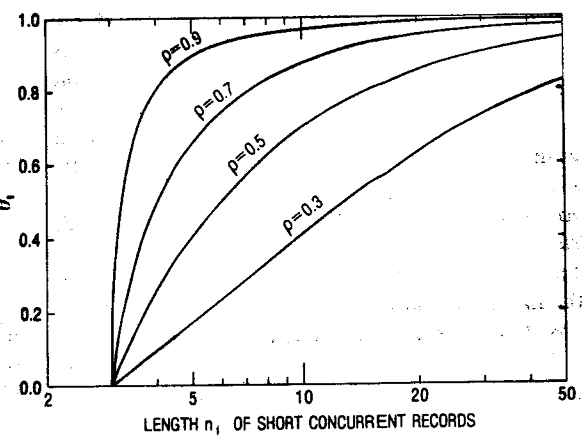

Fig. 1. Value of θ_1^* is shown for various values of the cross-correlation ρ of concurrent flows at the two sites and as a function of the length n_1 of the record at the short-record site.

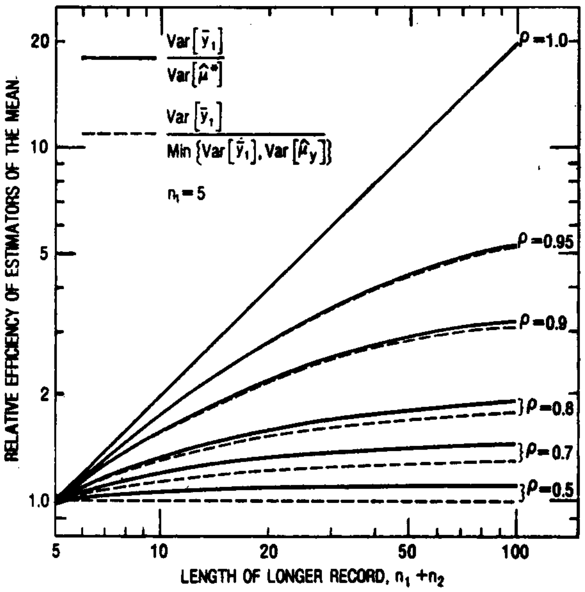

Fig. 2. Efficiency of $\hat{\mu}_y^*$ relative to $\bar{y}_1$ and the efficiency of the better of $\hat{\mu}_y$ or $\bar{y}_1$ relative to $\bar{y}_1$ are shown for various cross-correlations ρ when the length n_1 of the short-record is equal to 5.

We compare the efficiency relative to $\bar{y}_1$ of two estimators: $\hat{\mu}_y$ or $\bar{y}_1$ (whichever has smaller variance), and $\hat{\mu}_y^*$. Thus we document the information transfer gains which may be obtained by both the Matalas-Jacobs procedure [*Matalas and Jacobs*, 1964] and our estimator in (7). For this comparison we assume that a good regional estimate of ρ can be obtained so that for the moment we can ignore the error with which θ_1^* is estimated. Then the variance of $\hat{\mu}_y^*$ is obtained by substitution of (8) into (A1), which yields

$$\begin{aligned}\text{Var}\,[\hat{\mu}_y^*] &= \frac{\sigma_y^2}{n_1}\left[1 - \frac{n_2}{n_1+n_2}\left(\frac{(n_1-3)\rho^4}{(n_1-4)\rho^2+1}\right)\right] \\ &= \frac{\sigma_y^2}{n_1}\left[1 - \frac{n_2\theta_1^*\rho^2}{n_1+n_2}\right]\end{aligned} \qquad (9)$$

Figure 2 illustrates the information transfer gains that may be obtained with the two augmentation procedures when the length of the short record is fixed at $n_1 = 5$. The information transfer gains for both augmentation procedures are substantial when ρ and n_2 are large. Figure 3 illustrates the information transfer gains that may be obtained by both augmentation procedures when the longer record length is $n_1 + 60$. In both figures one can observe that use of $\hat{\mu}_y^*$ is always as good or yields a slight improvement in terms of information transfer over use of either $\hat{\mu}_y$ or $\bar{y}_1$.

The results reported thus far assume that the cross-correlation coefficient ρ needed to estimate θ_1^* is accurate. It can be based on the observed cross-correlation between streamflows which bear a relationship similar to those denoted here by x and y but for which there are reasonable records to estimate the cross-correlation. We emphasize that the traditional procedure also requires an estimate of ρ so that one may choose between $\hat{\mu}_y$ and $\bar{y}_1$ corresponding to $\theta_1 = 1$ or 0; an intermediate value of θ_1 seems more reasonable in most cases.

AN IMPROVED ESTIMATOR OF THE VARIANCE

The variance of the unbiased MLE of the variance derived by *Matalas and Jacobs* [1964] and by *Moran* [1974] is

$$\text{Var}\,[\hat{\sigma}_y^2] = \frac{2\sigma_y^4}{n_1-1} + \frac{n_2\sigma_y^4}{(n_1+n_2-1)^2(n_1-3)}(A\rho^4 + B\rho^2 + C) \qquad (10)$$

where

$$A = \frac{(n_2+2)(n_1-6)(n_1-8)}{(n_1-5)} + (n_1-4)\left(\frac{n_1n_2(n_1-4)}{(n_1-3)(n_1-2)} - \frac{2n_2(n_1-4)}{(n_1-3)} - 4\right)$$

$$B = \frac{6(n_2+2)(n_1-6)}{(n_1-5)} + 2(n_1^2 - n_1 - 14) + (n_1-4)\left(\frac{2n_2(n_1-5)}{(n_1-3)} - 2(n_1+3) - \frac{2n_1n_2(n_1-4)}{(n_1-3)(n_1-2)}\right)$$

$$C = 2(n_1+1) + \frac{3(n_2+2)}{(n_1-5)} - \frac{(n_1+1)(2n_1+n_2-2)(n_1-3)}{(n_1-1)} + (n_1-4)\left(\frac{2n_2}{(n_1-3)} + 2(n_1+1) + \frac{n_1n_2(n_1-4)}{(n_1-3)(n_1-2)}\right)$$

The variance of the sample variance at the short-record gage is simply

$$\text{Var}\,[s_{y_1}^2] = \frac{2\sigma_y^4}{n_1-1} \qquad (11)$$

The Water Resources Council [*Water Resources Council*, 1981] recommends the use of the unbiased MLE $\hat{\sigma}_y^2$ if Var $[\hat{\sigma}_y^2] <$ Var $[s_{y_1}^2]$ which occurs when

$$\rho^2 > \frac{-B \pm (B^2 - 4AC)^{1/2}}{2A} \qquad (12)$$

Again the unbiased MLE has minimum variance asymptotically among all asymptotically unbiased estimators (i.e., as n_1 and n_2 become large). The sample estimator $s_{y_1}^2$ has a smaller variance when (12) does not hold.

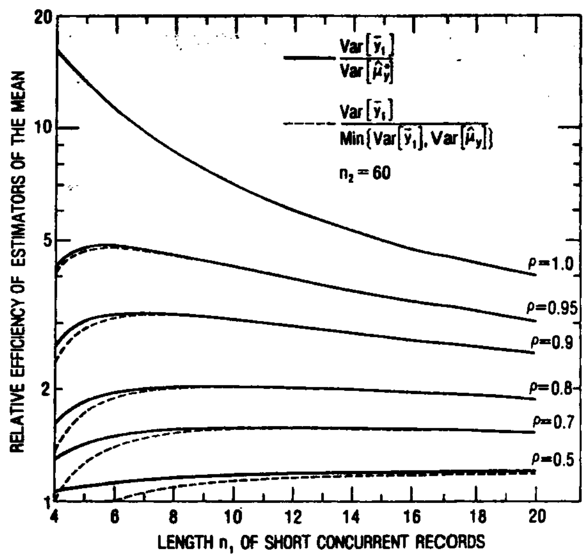

Fig. 3. Efficiency of $\hat{\mu}_y^*$ relative to $\bar{y}_1$ and the efficiency of the better of $\hat{\mu}_y$ or $\bar{y}_1$ relative to $\bar{y}_1$ are shown for various cross-correlations ρ when the length of the long record is equal to $n_1 + 60$.

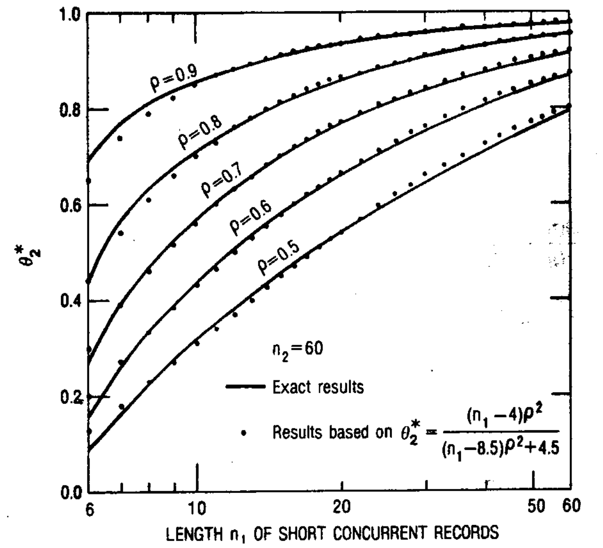

Fig. 4. Value of weight θ_2* as a function of ρ the cross-correlation of the flows and of the length of the record at the short-record station.

As an alternative, consider the estimator

$$\hat{\sigma}_y{}^{*2} = (1 - \theta_2)s_{y_1}{}^2 + \theta_2\hat{\sigma}_y{}^2 \tag{13}$$

which is a linear combination of $\hat{\sigma}_y{}^2$, given by (4), and $s_{y_1}{}^2$, the at-site sample estimator defined in (2).

The expression for the optimal value of θ_2, denoted $\theta_2{}^{\#}$, which minimizes the variance of $\sigma_y{}^{*2}$, derived in the second part of the appendix, is complex. Analytical efforts to simplify the expression in (A7) were unsuccessful. However, an approximation to $\theta_2{}^{\#}$ was derived. Our approximation is

$$\theta_2{}^* = \frac{(n_1 - 4)\rho^2}{(n_1 - 8.5)\rho^2 + 4.5} \tag{14}$$

Figure 4 provides a comparison of θ_2* and $\theta_2{}^{\#}$ and demonstrates the accuracy of the approximation. The optimal value $\theta_2{}^{\#}$ was insensitive to the value of n_2. The value of $\theta_2{}^{\#}$ is generally substantially less than unity. When $\rho = 0.8$ and $n_1 = 10$, the optimal value $\theta_2{}^{\#}$ is only 0.71. Clearly, $\hat{\sigma}_y{}^{*2}$ may differ appreciably from $\hat{\sigma}_y{}^2$ in (4) which always employs $\theta_2 = 1$.

Our interest here is in the information transfer gains which may be obtained by the two augmentation procedures. We compare the efficiency relative to $s_{y_1}{}^2$ of two estimators: $\hat{\sigma}_y{}^2$ or $s_{y_1}{}^2$ (whichever has smaller variance), and $\hat{\sigma}_y{}^{*2}$. We assume, for the moment, that a good regional estimate of ρ can be obtained so that we can ignore the error with which θ_2* is estimated. For these comparisons Var $[\hat{\sigma}_y{}^2]$ and Var $[s_{y_1}{}^2]$ are given by (10) and (11), while Var $[\hat{\sigma}_y{}^{*2}]$ is determined from (A4), (A6), and (A7).

Figures 5 and 6 illustrate the information transfer gains which may be obtained with the two augmentation procedures. When ρ and n_2 are large the information transfer gains for both augmentation procedures are substantial. Again, $\hat{\sigma}_y{}^{*2}$ always does as well or better than $\hat{\sigma}_y{}^2$ and $s_{y_1}{}^2$.

MONTE-CARLO EXPERIMENT

The sampling properties of our estimators of the mean and variance of the flows at the short-record site have been derived analytically assuming the value of ρ was available to determine θ_1* and θ_2*. A Monte-Carlo experiment allows for a comparison and evaluation of the estimators in a more realistic setting. In the following experiments the x and y flow sequences were generated from a bivariate normal distribution with cross-correlation ρ.

The Monte-Carlo experiments evaluate the consequences of using an estimate $\hat{\rho}$ instead of ρ to calculate θ_1* and θ_2* and to decide if Var $[\hat{\mu}_y]$ < Var $[\bar{y}_1]$ and Var $[\hat{\sigma}_y{}^2]$ < Var $[s_{y_1}{}^2]$. *Wallis and Matalas* [1972] also performed a Monte-Carlo experiment to evaluate the consequences of using an estimate $\hat{\rho}$ instead of ρ; however, their analysis quantified the probability of making the correct decision in terms of which estimator to use. Here we evaluate the consequences of using an estimate $\hat{\rho}$ instead of ρ on the average mean square error of the various estimators.

Estimators to be Compared

Traditionally, one chose between the Matalas-Jacobs estimators *Matalas and Jacobs* [1964] and the sample mean and variance of the flows at the short-record site. The resultant estimators are

$$\begin{aligned} \hat{\mu}_y' &= \hat{\mu}_y \qquad \hat{\rho}^2 > (n_1 - 2)^{-1} \\ \hat{\mu}_y' &= \bar{y}_1 \qquad \text{otherwise} \end{aligned} \tag{15}$$

$$\begin{aligned} \hat{\sigma}_y{}'^2 &= \hat{\sigma}_y{}^2 \qquad \hat{\rho}^2 > \frac{-B \pm (B^2 - 4AC)^{1/2}}{2A} \\ \sigma_y{}'^2 &= s_{y_1}{}^2 \qquad \text{otherwise} \end{aligned} \tag{16}$$

The prime signifies that these are strictly speaking neither the Matalas-Jacobs estimators nor the sample mean and variance. For the purposes of this Monte-Carlo experiment, the sample cross-correlation $\hat{\rho}$ is used in (8) and (14) to obtain the corresponding estimates of θ_1* and θ_2* for use in determining $\hat{\mu}_y$* and $\hat{\sigma}_y{}^{*2}$. We emphasize that in practice if n_1 is small, one should obtain a good regional estimate of ρ rather than using the sample estimator $\hat{\rho}$ to obtain θ_1* and θ_2*, as is done in these Monte-Carlo experiments.

The practicing engineer would be hesitant to use our estimators $\hat{\mu}_y$* and $\hat{\sigma}_y{}^{*2}$ if they were inconsistent with the at-site

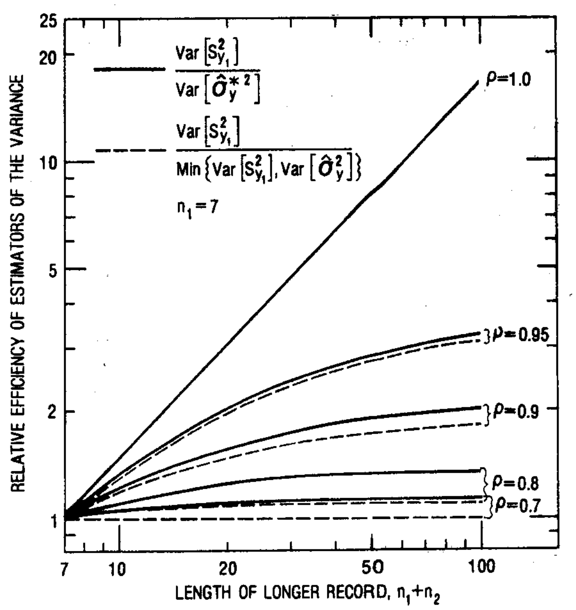

Fig. 5. Efficiency of $\hat{\sigma}_y{}^{*2}$ relative to $s_{y_1}{}^2$ and the efficiency of the better of $\hat{\sigma}_y{}^2$ or $s_{y_1}{}^2$ relative to $s_{y_1}{}^2$ are shown for various cross-correlations ρ when the length n_1 of the short-record is equal to 7.

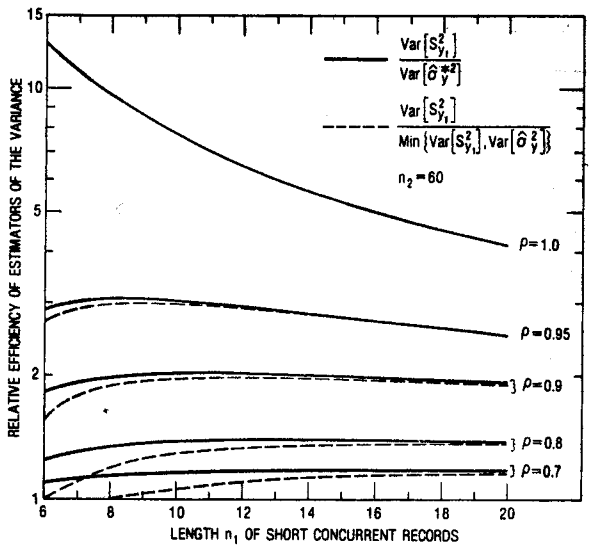

Fig. 6. Efficiency of $\hat{\sigma}_y{}^{*2}$ relative to $s_{y_1}{}^2$ and the efficiency of the better of $\hat{\sigma}_y{}^2$ or $s_{y_1}{}^2$ relative to $s_{y_1}{}^2$ are shown for various cross-correlations ρ when the length of the long record is equal to $n_1 + 60$.

sample estimates $\bar{y}_1$ and $s_{y_1}{}^2$. Thus we consider the following "clipped" estimators. They may prevent extreme errors in estimation of μ_y and $\sigma_y{}^2$. If the value of our estimate of the mean $\hat{\mu}_y{}^*$ falls outside the two-sided p% confidence interval for $\bar{y}_1$, the clipped estimator $\hat{\mu}_y{}^*(p)$ is set equal to the nearest limit of the constructed confidence interval. The clipped estimator of the mean of the flows at the short-record site becomes

$$\begin{aligned} \hat{\mu}_y{}^*(p) &= \bar{y}_1 + k \qquad && \hat{\mu}_y{}^* > \bar{y}_1 + k \\ \hat{\mu}_y{}^*(p) &= \hat{\mu}_y{}^* && \text{otherwise} \\ \hat{\mu}_y{}^*(p) &= \bar{y}_1 - k && \hat{\mu}_y{}^* < \bar{y}_1 - k \end{aligned} \tag{17}$$

where

$$k = \frac{s_{y_1} t(n_1 - 1, q/2)}{(n_1)^{1/2}}$$

$$q = 1 - 0.01p$$

and $t(n_1 - 1, q/2)$ is that quantile of Student's t distribution with $n_1 - 1$ degrees of freedom which is exceeded with probability $q/2$.

Similarly, when our estimate of the variance $\hat{\sigma}_y{}^{*2}$ falls above the one-sided p% confidence interval for $s_{y_1}{}^2$, then we set the clipped estimator $\hat{\sigma}_y{}^{*2}(p)$ equal to the upper limit of the constructed confidence interval. The clipped estimator of the variance of the flows at the short-record site becomes

$$\begin{aligned} \hat{\sigma}_y{}^{*2}(p) &= m \qquad && \hat{\sigma}_y{}^{*2} > m \\ \hat{\sigma}_y{}^{*2}(p) &= \hat{\sigma}_y{}^{*2} && \text{otherwise} \end{aligned} \tag{18}$$

$$m = \frac{(n_1 - 1)s_{y_1}{}^2}{\chi^2(q, n_1 - 1)}$$

$$q = 0.01p$$

and $\chi^2(q, n_1 - 1)$ is the chi-square quantile for $(n_1 - 1)$ degrees of freedom which is exceeded with probability q.

Results

A Monte-Carlo experiment determined the efficiencies of our unclipped estimators $\hat{\mu}_y{}^*$ and $\hat{\sigma}_y{}^{*2}$ relative to the "implemented" Matalas-Jacobs [*Matalas and Jacobs*, 1964] estimators $\hat{\mu}_y'$ and $\hat{\sigma}_y'^2$ when the sample cross-correlation $\hat{\rho}$ is used to decide if Var $[\hat{\mu}_y] <$ Var $[\bar{y}_1]$ or Var $[\hat{\sigma}_y{}^2] <$ Var $[s_{y_1}{}^2]$ and to estimate $\theta_1{}^*$ and $\theta_2{}^*$; the results are displayed in Table 1. Interestingly, the efficiency of $\hat{\sigma}_y{}^{*2}$ relative to $\hat{\sigma}_y'^2$ increases when $\hat{\rho}$ is used instead of ρ, while the efficiency of $\hat{\mu}_y{}^*$ relative to $\hat{\mu}_y'$ decreases when $\hat{\rho}$ is used instead of ρ. From Table 1 we conclude that in terms of mean square error $\hat{\sigma}_y{}^{*2}$ dominates $\hat{\sigma}_y'^2$, and $\hat{\mu}_y{}^*$ is generally an improvement over $\hat{\mu}_y'$, for the nine cases selected to capture the region of practical interest.

Figures 2, 3, 5, and 6 document the tremendous information transfer gains associated with the use of the Matalas-Jacobs estimators [*Matalas and Jacobs*, 1964] or our estimators when the value of ρ is known. Monte-Carlo experiments were performed to examine these information transfer gains when the sample cross-correlation $\hat{\rho}$ is used to estimate $\theta_1{}^*$ and $\theta_2{}^*$; the results are summarized in Table 2. Our estimators $\hat{\mu}_y{}^*$ and $\hat{\sigma}_y{}^{*2}$ are still dramatic improvements over the at-site sample estimators $\bar{y}_1$ and $s_{y_1}{}^2$ in terms of mean square error even when the sample estimator $\hat{\rho}$ is used instead of ρ to estimate

TABLE 1. Comparison of Estimates of Efficiencies of the Improved Estimators $\hat{\mu}_y{}^*$ and $\hat{\sigma}_y{}^{*2}$ Relative to $\hat{\mu}_y'$ and $\hat{\sigma}_y'^2$ When the True Value of ρ and the Sample Estimate $\hat{\rho}$ Are Used to Decide if Var $[\hat{\mu}_y] <$ Var $[\bar{y}_1]$ and Var $[\hat{\sigma}_y{}^2] <$ Var $[s_{y_1}{}^2]$ and to Estimate $\theta_1{}^*$ and $\theta_2{}^*$

		$\frac{\text{mse}[\hat{\mu}_y']}{\text{mse}[\hat{\mu}_y{}^*]}$		$\frac{\text{mse}[\hat{\sigma}_y'^2]}{\text{mse}[\hat{\sigma}_y{}^{*2}]}$	
ρ	n_1	ρ	$\hat{\rho}$	ρ	$\hat{\rho}$
0.5	6	1.128	1.08 (0.005)	1.010	1.12 (0.053)
0.5	10	1.032	1.02 (0.003)	1.025	1.06 (0.010)
0.5	25	1.003	1.00 (0.001)	1.014	1.02 (0.002)
0.7	6	1.060	1.05 (0.004)	1.089	1.15 (0.040)
0.7	10	1.013	1.00 (0.003)	1.090	1.09 (0.009)
0.7	25	1.001	1.00 (0.001)	1.009	1.01 (0.002)
0.9	6	1.013	0.99 (0.004)	1.166	1.21 (0.021)
0.9	10	1.002	0.99 (0.002)	1.029	1.05 (0.004)
0.9	25	1.000	1.00 (0.0002)	1.000	1.00 (0.001)

Here $n_2 = 60$; mse, mean square error. This table is based upon 50,000 replicates. A 95% confidence interval for each statistic is constructed by adding or subtracting the values in parentheses from the reported values.

TABLE 2. Comparison of Estimates of Efficiencies of the Estimators $\hat{\mu}_y{}^*$ and $\hat{\sigma}_y{}^{*2}$ Relative to $\bar{y}_1$ and $s_{y_1}{}^2$ When the True Value of ρ and the Sample Estimate $\hat{\rho}$ are Used to Estimate $\theta_1{}^*$ and $\theta_2{}^*$

		$\frac{\text{mse}[\bar{y}_1]}{\text{mse}[\hat{\mu}_y{}^*]}$		$\frac{\text{mse}[s_{y_1}{}^2]}{\text{mse}[\hat{\sigma}_y{}^{*2}]}$	
ρ	n_1	ρ	$\hat{\rho}$	ρ	$\hat{\rho}$
0.5	6	1.128	1.06 (0.009)	1.010	0.89 (0.04)
0.5	10	1.177	1.14 (0.01)	1.025	1.02 (0.010)
0.5	25	1.184	1.17 (0.008)	1.032	1.03 (0.006)
0.7	6	1.494	1.41 (0.015)	1.089	1.08 (0.04)
0.7	10	1.576	1.53 (0.02)	1.158	1.18 (0.02)
0.7	25	1.493	1.47 (0.001)	1.170	1.18 (0.01)
0.9	6	3.154	3.01 (0.04)	1.819	1.94 (0.06)
0.9	10	3.047	3.02 (0.05)	2.019	2.07 (0.03)
0.9	25	2.303	2.28 (0.03)	1.817	1.84 (0.02)

Here $n_2 = 60$. This table is based upon 50,000 replicates. A 95% confidence interval for each statistic is constructed by adding or subtracting the values in parentheses from the reported values.

TABLE 3. Estimates of the Root Mean Square Error of Various Estimators of μ_y for $\sigma_y^2 = 1$ and $n_2 = 60$

	$n_1 = 6$			$n_1 = 10$			$n_1 = 25$		
	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$
rmse[$\bar{y}_1$]	0.41	0.41	0.41	0.32	0.32	0.32	0.20	0.20	0.20
rmse[$\hat{\mu}_y'$]	0.41	0.35	0.23	0.30	0.26	0.18	0.18	0.16	0.13
rmse[$\hat{\mu}_y^*$]	0.40	0.34	0.23	0.30	0.26	0.18	0.18	0.16	0.13
rmse[$\hat{\mu}_y^*(50)$]	0.38	0.34	0.28	0.29	0.26	0.22	0.18	0.17	0.15
rmse[$\hat{\mu}_y^*(75)$]	0.39	0.34	0.24	0.30	0.26	0.19	0.18	0.16	0.14
rmse[$\hat{\mu}_y^*(95)$]	0.40	0.34	0.23	0.30	0.26	0.18	0.18	0.16	0.13

This table is based upon 50,000 replicates. Ninety-five percent confidence intervals are within $\pm 1\%$ of the reported values.

θ_1^* and θ_2^*. In general, use of $\hat{\rho}$ instead of ρ leads to approximately the same or in some cases only marginal increases or decreases in the efficiency of our estimators relative to the at-site estimators. Thus Figures 2–3 and 5 and 6 provide a good guide to the gains obtainable from record augmentation when the sample cross correlation is used to estimate θ_1^* and θ_2^*, respectively.

The results of a Monte-Carlo experiment to compare the performance of the estimators $\bar{y}_1$, $\hat{\mu}_y'$, and $\hat{\mu}_y^*$ and the clipped estimators $\hat{\mu}_y^*(p)$ for $p = 50$, 75, and 95 are summarized in Table 3. In terms of root mean square error, the estimators $\hat{\mu}_y^*$ and $\hat{\mu}_y'$ both dominate the at-site sample estimate $\bar{y}_1$, and the dominance is substantial for $\rho \geq 0.7$ as expected.

When the sample estimate $\hat{\rho}$ instead of ρ is used to decide if Var $[\hat{\mu}_y] <$ Var $[\bar{y}_1]$ and to estimate θ_1^* the estimators $\hat{\mu}_y'$ and $\hat{\mu}_y^*$ are unbiased. This is due to the symmetry of $\hat{\mu}_y$ and $\hat{\mu}_y^*$ about $\bar{y}_1$ and the independence of $\hat{\rho}$ and $\bar{y}_1$, $\bar{x}_1$, or $\bar{x}_2$. Furthermore, since the clipping of $\hat{\mu}_y^*$ to obtain $\hat{\mu}_y^*(p)$ is symmetric about $\hat{\mu}_y^*$, the clipped estimators of the mean are also unbiased.

In Table 3 the estimator $\hat{\mu}_y^*$ dominates $\hat{\mu}_y'$; however, the improvement is very small. Differences at two significant figures occur only when $n_1 = 6$. When $\rho = 0.5$, the performance of the clipped estimators of the mean of the flows at the short-record site is in some instances marginally superior to that of $\hat{\mu}_y^*$; however, in other cases ($\rho \geq 0.7$) the clipped estimators perform worse than $\hat{\mu}_y^*$. We conclude that the clipped estimators of the mean are not an attractive alternative.

A comparison of the performance of the estimators $s_{y_1}^2$, $\hat{\sigma}_y'^2$, and $\hat{\sigma}_y^{*2}$ to the clipped estimators $\hat{\sigma}_y^{*2}(p)$ for $p = 50$, 75, and 95 is given in Table 4. Here one observes the interesting result that $\hat{\sigma}_y'^2$ and $\hat{\sigma}_y^{*2}$ are biased estimators though the bias disappears as the length of the short record increases. Apparently the use of $\hat{\rho}$ instead of ρ to decide if Var $[s_{y_1}^2] >$ Var $[\hat{\sigma}_y^2]$ and to estimate θ_2^* introduces a bias which would otherwise not exist. This bias is negligible when compared to the extreme variability associated with the estimators which is illustrated using Box plots on a logarithmic scale in Figure 7. Figure 7 emphasizes the value of using clipped variance estimators which reduce the likelihood of extremely large estimation errors.

In terms of the root mean square error of estimators of σ_y^2, we conclude from Table 4 that, in general, the estimators $\hat{\sigma}_y^{*2}$ and $\hat{\sigma}_y^{*2}(95)$ dominate use of $\hat{\sigma}_y'^2$ for the cases considered. Furthermore, $\hat{\sigma}_y^{*2}(50)$ and $\hat{\sigma}_y^{*2}(75)$ are an improvement over both $\hat{\sigma}_y'^2$ and $\hat{\sigma}_y^{*2}$ when $n_1 = 6$; however, their performance approximates that of $\hat{\sigma}_y^{*2}$ and $\hat{\sigma}_y'^2$ when $n_1 = 25$. Our recommendation would be to use $\hat{\sigma}_y^{*2}(75)$ with small n_1 because it exhibits very little bias and has one of the smaller mean square errors.

The clipped estimators are, in general, downward biased and that bias does not disappear entirely even when $n_1 = 25$.

TABLE 4. Estimates of the Mean and Root Mean Square Error of various Estimators of σ_y^2 for $\sigma_y^2 = 1$ and $n_2 = 60$

	$n_1 = 6$			$n_1 = 10$			$n_1 = 25$		
	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$
$E[s_{y_1}^2]$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$E[\hat{\sigma}_y'^2]$	1.00	0.96	0.95	0.99	0.98	1.00	1.00	1.00	1.00
$E[\hat{\sigma}_y^{*2}]$	1.01	0.98	0.96	1.00	0.99	0.98	1.00	1.00	1.00
$E[\hat{\sigma}_y^{*2}(50)]$	0.97	0.94	0.87	0.98	0.95	0.90	0.99	0.97	0.94
$E[\hat{\sigma}_y^{*2}(75)]$	1.00	0.97	0.94	1.00	0.98	0.96	1.00	0.99	0.98
$E[\hat{\sigma}_y^{*2}(95)]$	1.00	0.98	0.95	1.00	0.99	0.98	1.00	1.00	1.00
rmse[$s_{y_1}^2$]	0.63	0.64	0.64	0.47	0.48	0.48	0.29	0.29	0.29
rmse[$\hat{\sigma}_y'^2$]	0.72‡	0.66†	0.51*	0.48	0.46	0.34	0.29	0.27	0.21
rmse[$\hat{\sigma}_y^{*2}$]	0.68†	0.62*	0.46	0.47	0.44	0.33	0.28	0.27	0.21
rmse[$\hat{\sigma}_y^{*2}(50)$]	0.60	0.55	0.44	0.45	0.41	0.34	0.28	0.26	0.23
rmse[$\hat{\sigma}_y^{*2}(75)$]	0.63	0.58	0.44	0.46	0.43	0.32	0.28	0.27	0.22
rmse[$\hat{\sigma}_y^{*2}(95)$]	0.66*	0.60	0.45	0.47	0.44	0.33	0.29	0.27	0.21

This table is based on 50,000 replicates. Ninety-five percent confidence intervals for the mean values of the various estimators are within $\pm 0.5\%$ of the reported values. Ninety-five percent confidence intervals for the rmse of the various estimators are within $\pm 1\%$ of the reported values except for the values marked with an asterisk, dagger, and double dagger whose 95% confidence intervals are within $\pm 2\%$, $\pm 4\%$ and $\pm 5\%$, respectively, of the reported values.

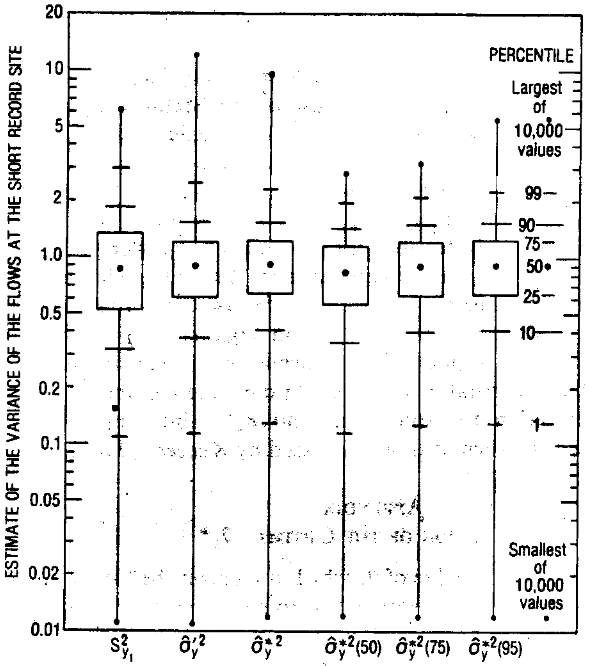

Fig. 7. Box plots illustrating the distribution of the estimators $s_{y_1}{}^2$, $\hat{\sigma}_y{}'^2$, $\hat{\sigma}_y{}^{*2}$, and $\hat{\sigma}_y{}^{*2}(p)$ for $p = 50$, 75, and 95 for a cross-correlation ρ equal to 0.9 and a short-record length n_1 equal to 6; each Box plot is based upon 10,000 estimates of $\sigma_y{}^2$.

As is often the case in practice, introduction of bias is accompanied by a reduction in the estimators' mean square error (see examples in the works by *Stedinger* [1980, 1981] and *Loucks et al.* [1981, pp. 104–106]). For estimators which exhibit such large variability as these, one should be willing to accept some bias to achieve a lower root mean square error.

Another interesting result, evident in Table 4, is that in terms of root mean square error the at-site sample estimator $s_{y_1}{}^2$ was not dominated by $\hat{\sigma}_y{}'^2$, $\hat{\sigma}_y{}^{*2}$, or $\hat{\sigma}_y{}^{*2}(95)$. However, $s_{y_1}{}^2$ was always dominated by $\hat{\sigma}_y{}^{*2}(50)$ and $\hat{\sigma}_y{}^{*2}(75)$ for the cases considered.

IMPROVED RECORD EXTENSION PROCEDURES

The procedures considered here are appropriate when one considers the problem of augmenting records of peak annual floods, mean annual flows, or an appropriate transformation thereof. Hence the estimators $\hat{\mu}_y{}^*$, $\hat{\sigma}_y{}^{*2}$, and $\hat{\sigma}_y{}^{*2}(p)$ developed in this study are improvements over those discussed by *Matalas and Jacobs* [1964] for estimating the first two moments of the distribution of peak annual flow series. However, the problem of actually extending available monthly streamflow records for use in sequential studies for reservoir design or operations poses other problems; development of optimal small sample estimators of the mean and variance of the streamflows solves only part of that problem.

Development of a Unique Extension

Hirsch [1982] observed that use of (2) to produce an extended monthly streamflow record fails to yield a unique extended record. Furthermore, (2) makes no effort to preserve the autocorrelation structure of the observed monthly flows; hence it fails to yield extended flow records with the appropriate serial correlation. For the actual extension of seasonal streamflow series, Hirsch suggests maintenance of variance extension (MOVE.1 and MOVE.2) techniques which use the linear equations

$$\hat{y}_i = a + bx_i \tag{19}$$

with the values of a and b chosen so as to generate a reasonable and unique extended record. These MOVE procedures are intended for use in situations in which the two streamflow populations do not differ substantially in terms of their distribution shape, serial correlation or seasonality. Since (19) simply represents a constant and linear transformation between the two sites, these procedures can be expected to produce extended sequences which exhibit properties much like the properties of the flows at the long-record site.

In MOVE.1, *Hirsch* [1982] chose the estimators of a and b so that if (19) were used to generate an entire sequence y_i, for $i = 1, \cdots, n_1 + n_2$, the historical sample moments $\bar{y}_1$ and $s_{y_1}{}^2$ would be reproduced. Similarly in MOVE.2, *Hirsch* [1982] chose a and b so that if (19) were used to generate an entire sequence $\hat{y}_i$ with $i = 1, \cdots, n_1 + n_2$ the unbiased MLE [*Matalas and Jacobs*, 1964] estimates $\hat{\mu}_y$ and $\hat{\sigma}_y{}^2$ would be reproduced. However, in practice, one uses (19) to generate the $\hat{y}_i$ only for $i = n_1 + 1, \cdots, n_1 + n_2$. This suggests that Hirsch used estimators of a and b which did not achieve what he intended. With the procedures suggested by Hirsch, the extended sequences have sample means and variances which fail to equal the short y record's moments, $\bar{y}_1$ and $s_{y_1}{}^2$, or the Matalas-Jacobs estimator's of the population's mean and variance, whichever were chosen to be the appropriate estimates of the moments of the y series.

A new approach (MOVE.3) for selection of a and b is motivated by a reconsideration of *Matalas and Jacob*'s [1964] paper. For fixed y_1 through y_{n_1}, Matalas and Jacobs obtained their estimators of μ_y and $\sigma_y{}^2$ by calculating the expectation of the sample moments of the extended sequences

$$\{y_1, \cdots, y_{n_1}, \tilde{y}_{n_1+1}, \cdots, \tilde{y}_{n_1+n_2}\}$$

Here $\tilde{y}_{n_1+1}$ through $\tilde{y}_{n_1+n_2}$ are generated using (2) with random e_i. Thus to obtain a unique extended record it would be reasonable to select a and b in (19) so that the resultant sequence of $n_1 + n_2$ values $\{y_1, \cdots, y_{n_1}, \hat{y}_{n_1+1}, \cdots, \hat{y}_{n_1+n_2}\}$ has mean and variance $\hat{\mu}_y$ and $\hat{\sigma}_y{}^2$ (the Matalas-Jacobs estimators). With this new approach the whole extended sequence would have a sample mean and variance equal to the Matalas-Jacobs estimators of the population values of those statistics. Moreover, using MOVE.3 with the Matalas-Jacobs estimators will force the sample mean and variance of the generated observations, $\hat{y}_{n_1+1}$ through $\hat{y}_{n_1+n_2}$, to equal the expected value of those two statistics given $y_1, \cdots, y_{n_1}$ as well as $x_1, \cdots, x_{n_1+n_2}$. Estimates of a and b for the MOVE.3 procedure may be obtained by rewriting (19) as

$$\hat{y}_i = a' + b(x_i - \bar{x}_2) \tag{20}$$

Then the MOVE.3 estimates of a' and b are obtained from

$$a' = \frac{(n_1 + n_2)\hat{\mu}_y - n_1\bar{y}_1}{n_2} \tag{21}$$

$$b^2 = [(n_1 + n_2 - 1)\hat{\sigma}_y{}^2 - (n_1 - 1)s_{y_1}{}^2 - n_1(\bar{y}_1 - \hat{\mu}_y)^2 - n_2(a' - \hat{\mu}_y)^2][(n_2 - 1)s_{x_2}{}^2]^{-1} \tag{22}$$

Monte-Carlo experiments indicated that MOVE.2 and the proposed MOVE.3 are nearly indistinguishable in the sense of mean square error of the estimators of the mean and variance of the complete extended record.

Consider also a fourth procedure (MOVE.4) which chooses

a and b in (19) so that the resultant sequence of $n_1 + n_2$ values $\{y_1, \cdots, y_{n_1}, \hat{y}_{n_1+1}, \cdots, \hat{y}_{n_1+n_2}\}$ has mean and variance $\hat{\mu}_y^*$ and $\hat{\sigma}_y^{*2}$. Estimates of a and b using the MOVE.4 procedure are obtained by replacing the Matalas-Jacobs estimators [*Matalas and Jacobs*, 1964] $\hat{\mu}_y$ and $\hat{\sigma}_y^2$ in (21) and (22) by our estimators $\hat{\mu}_y^*$ and $\hat{\sigma}_y^{*2}$. Clearly, MOVE.4 will have the same efficiency advantages over MOVE.3 that use of $\hat{\mu}_y^*$ and $\hat{\sigma}_y^{*2}$ had over use of $\hat{\mu}_y'$ and $\hat{\sigma}_y'^2$ because these are exactly the sample moments the MOVE.3 and MOVE.4 procedures reproduce.

New Opportunities With Monthly Flows

The range of techniques one can use for record extension and/or augmentation are greater with monthly flow records than with annual flood series. Given an $n_1 = 6$ year annual flood series, one can use the record augmentation procedures discussed here, regional regression [*Benson and Matalas*, 1967; *Thomas and Benson*, 1970], or empirical Bayes procedures for combining site and regional information to improve at-site estimators of the mean and variance of the flows or some transformation thereof [*Kuczera*, 1983]. However, when dealing with an $n_1 = 6$ year monthly flow record, one could assume that the relationship between flows at the two gages is independent of the month in which the flows occurred. For $n_1 = 6$, one then has $12n_1 = 72$ concurrent observations to derive a relationship between the flows at the two sites [*Alley and Burns*, 1983]. If the relationship between flows at the two sites exhibits seasonal differences, then one could develop different models for flows occuring in different months within distinct seasons (W. M. Alley, personal communication, 1983). With four seasons of equal length, n_1 concurrent years of record yields $3n_1$ concurrent (sometimes correlated) monthly flows. These longer concurrent monthly records greatly enhance the attractiveness of record augmentation procedures for estimating the statistics of flows occuring in each month because the parameters of the model relating the flows at the two sites can be estimated more precisely. However, the relationship between concurrent flows at two sites will in general exhibit some variation from month to month so that this increased precision is obtained by introducing some bias into the analysis.

CONCLUSIONS

Estimators of the mean and variance of the flows at a short-record site were derived which should have lower variance than the unbiased maximum likelihood or Matalas-Jacobs estimators [*Matalas and Jacobs*, 1964] for the small samples of interest in hydrology. Since the Matalas-Jacobs estimators as well as our estimators require an estimate of the flow's cross-correlation ρ, a Monte-Carlo experiment was performed to evaluate the impact of using the sample cross-correlation $\hat{\rho}$ on their root mean square error. In general, our estimators have essentially an equal or, as is more often the case, a lower rmse than the Matalas-Jacobs estimators. In addition, we note that use of $\hat{\sigma}_y^{*2}$ does not require evaluation of the complex expressions in (10) and (12) to determine if Var $[\hat{\sigma}_y^2] <$ Var $[s_{y_1}^2]$, as is required if one wants to choose between $\hat{\sigma}_y^2$ and $s_{y_1}^2$.

In this study we have examined the sampling properties of estimators of the mean and variance of the flows at the short-record site. Estimators of σ_y^2 are extremely unstable, especially when the length of the short record is small. In our experiment, clipped estimators decrease the likelihood of large estimation errors. In particular, $\hat{\sigma}_y^{*2}(50)$ and $\hat{\sigma}_y^{*2}(75)$, although downward biased, had appreciably lower root mean square errors when the length of the short record is less than 10.

Figures 2, 3, 5, and 6 document the tremendous information transfer gains associated with the use of the Matalas-Jacobs estimators [*Matalas and Jacobs*, 1964] or our estimators when the values of ρ and n_2 are large, as can be the case in practice. In practice, these procedures require an estimate of ρ and sample estimates are highly suspect when n_1 is small. Nevertheless, Table 2 indicates that the information transfer gains when $\hat{\rho}$ is used for ρ are comparable to the gains achieved when ρ is known. One should also consider use of multivariate regional regression equations based on regional hydrologic, meterologic, and topographic information [*Thomas and Benson*, 1970]. Alternatively, one could combine regional and at-site information to estimate μ_y and σ_y^2 using empirical Bayes procedures such as those discussed by *Kuczera* [1983].

APPENDIX: DERIVATION OF THE OPTIMAL θ_1^*

Here we derive the value of θ_1 which minimizes the variance of the estimator $\hat{\mu}_y^*$. The variance of this improved estimator of the mean is given by

$$\begin{aligned}\text{Var}\,[\hat{\mu}_y^*] &= E[(\hat{\mu}_y^* - \mu_y)^2] \\ &= E[(\bar{y}_1 - \mu_y)^2] + \theta_1^2\left(\frac{n_2}{n_1+n_2}\right)^2 E[\hat{\beta}^2]E[(\bar{x}_2 - \bar{x}_1)^2] \\ &\quad + 2\theta_1\beta\left(\frac{n_2}{n_1+n_2}\right)E[(\bar{x}_2 - \bar{x}_1)(\bar{y}_1 - \mu_y)] \end{aligned} \tag{A1}$$

where, upon taking expectations first over y and then over x one obtains

$$E[(\bar{y}_1 - \mu_y)^2] = \frac{\sigma_y^2}{n_1}$$

$$E[(\bar{x}_2 - \bar{x}_1)^2] = \frac{n_1 + n_2}{n_1 n_2}\sigma_x^2$$

$$E[(\bar{x}_2 - \bar{x}_1)(\bar{y}_1 - \mu_y)] = \frac{-1}{n_1}\rho\sigma_x\sigma_y$$

$$E[\hat{\beta}^2] = \frac{(1-\rho^2)\sigma_y^2}{(n_1 - 3)\sigma_x^2} + \beta^2$$

Here $\hat{\beta}$ and $(\bar{x}_2 - \bar{x}_1)$ are independently distributed, as are $\hat{\beta}^2$ and $(\bar{x}_2 - \bar{x}_1)^2$. Minimization of (A1) leads to the estimator

$$\theta_1^* = \frac{(n_1 - 3)\rho^2}{(n_1 - 4)\rho^2 + 1} \tag{A2}$$

for use in (7). As a check on our derivation of Var $[\hat{\mu}_y^*]$ in (A1) we note that substitution of $\theta_1^* = 1$ in (A1) yields the identity

$$\text{Var}\,[\hat{\mu}_y^*] = \text{Var}\,[\hat{\mu}_y] \tag{A3}$$

as expected.

DERIVATION OF THE OPTIMAL θ_2^*

The variance of the improved estimator of the variance is given by

$$\begin{aligned}\text{Var}\,[\hat{\sigma}_y^{*2}] &= E[(\hat{\sigma}_y^{*2} - \sigma_y^2)^2] \\ &= E[\hat{\sigma}_y^{*4}] - \sigma_y^4\end{aligned} \tag{A4}$$

since $\hat{\sigma}_y^{*2}$ is an unbiased estimator. By combining (4) and (13) our improved estimator may be rewritten as

$$\hat{\sigma}_y^{*2} = \frac{\theta_2}{n_1+n_2-1}\left[\left(\frac{n_1+(1-\theta_2)n_2-1}{\theta_2}\right)s_{y_1}^2+(n_2-1)\hat{\beta}^2 s_{x_2}^2 + (n_2-1)\alpha^2(1-\hat{\rho}^2)s_{y_1}^2 + \frac{n_1 n_2}{n_1+n_2}\hat{\beta}^2(\bar{x}_2-\bar{x}_1)^2\right] \quad \text{(A5)}$$

Now we obtain

$$E[\hat{\sigma}_y^{*4}] = \frac{\theta_2^2}{(n_1+n_2-1)^2}\left[\left(\frac{n_1+(1-\theta_2)n_2-1}{\theta_2}\right)^2 E[s_{y_1}^4] + 2\left(\frac{n_1+(1-\theta_2)n_2-1}{\theta_2}\right)H_1 + H_2\right] \quad \text{(A6)}$$

where

$$H_1 = (n_2-1)\alpha^2 E[(1-\hat{\rho}^2)s_{y_1}^4] + (n_2-1)E[s_{x_2}^2]E[s_{y_1}^2\hat{\beta}^2] + \left(\frac{n_1 n_2}{n_1+n_2}\right)E[(\bar{x}_2-\bar{x}_1)^2]E[s_{y_1}^2\hat{\beta}^2]$$

$$\begin{aligned}H_2 = (n_2-1)^2[&\alpha^4 E[(1-\hat{\rho}^2)^2 s_{y_1}^4] + E[\hat{\beta}^4]E[s_{x_2}^4] \\ &+ 2\alpha^2 E[(1-\hat{\rho}^2)s_{y_1}^2\hat{\beta}^2]E[s_{x_2}^2]] \\ &+ \frac{2(n_2-1)n_1 n_2}{n_1+n_2}[\alpha^2 E[\hat{\beta}^2(1-\hat{\rho}^2)s_{y_1}^2]E[(\bar{x}_2-\bar{x}_1)^2] \\ &+ E[\hat{\beta}^4]E[(\bar{x}_2-\bar{x}_1)^2]E[s_{x_2}^2]] \\ &+ \left(\frac{n_1 n_2}{n_1+n_2}\right)^2 E[\hat{\beta}^4]E[(\bar{x}_2-\bar{x}_1)^4]\end{aligned}$$

$$E[s_{y_1}^4] = \frac{n_1+1}{n_1-1}\sigma_y^4$$

$$E[s_{x_2}^4] = \frac{n_2+1}{n_2-1}\sigma_x^4$$

$$E[\hat{\beta}^4] = \left[\rho^4 + \frac{6\rho^2(1-\rho^2)}{n_1-3} + \frac{3(1-\rho^2)^2}{(n_1-3)(n_1-5)}\right]\frac{\sigma_y^4}{\sigma_x^4}$$

$$E[(\bar{x}_2-\bar{x}_1)^2] = \frac{n_1+n_2}{n_1 n_2}\sigma_x^2$$

$$E[(\bar{x}_2-\bar{x}_1)^4] = \frac{3(n_1+n_2)^2}{n_1^2 n_2^2}\sigma_x^4$$

$$E[(1-\hat{\rho}^2)^2 s_{y_1}^4] = \frac{n_1(n_1-2)}{(n_1-1)^2}(1-\rho^2)^2\sigma_y^4$$

$$E[\hat{\beta}^2 s_{y_1}^2] = \left[\rho^4 + \frac{n_1+4}{n_1-1}\rho^2(1-\rho^2) + \frac{(n_1+1)(1-\rho^2)^2}{(n_1-1)(n_1-3)}\right]\frac{\sigma_y^4}{\sigma_x^2}$$

$$E[(1-\hat{\rho}^2)s_{y_1}^4] = \left[\frac{n_1+1}{n_1-1}(1-\rho^4) - \frac{n_1+4}{n_1-1}\rho^2(1-\rho^2) - \frac{(n_1+1)}{(n_1-1)^2}(1-\rho^2)^2\right]\sigma_y^4$$

$$E[\hat{\beta}^2(1-\hat{\rho}^2)s_{y_1}^2] = \left[\frac{n_1-2}{n_1-1}\rho^2(1-\rho^2) + \frac{n_1-2}{(n_1-1)(n_1-3)}(1-\rho^2)^2\right]\frac{\sigma_y^4}{\sigma_x^2}$$

Substitution of (A6) into (A4) yields the complete expression for Var $[\hat{\sigma}_y^{*2}]$. Minimization of (A4) leads to the estimator

$$\theta_2^* = \frac{(n_1+n_2-1)(n_2 E[s_{y_1}^4] - H_1)}{n_2^2 E[s_{y_1}^4] + H_2 - 2n_2 H_1} \quad \text{(A7)}$$

Acknowledgments. The authors appreciate the constructive comments provided by R. M. Hirsch, W. M. Alley, and N. C. Matalas. This work constitutes a portion of the first author's Ph.D. dissertation and was completed while the second author was on leave at the U.S. Geological Survey in Reston, Virginia. The assistance provided by C. Moss and others at the U.S. Geological Survey is gratefully acknowledged.

REFERENCES

Alley, W. M., and A. W. Burns, Mixed-station extension of monthly streamflow records, *J. Hydraul. Div. Am. Soc. Civ. Eng.*, *109*(10), 1271–1284, 1983.

Anderson, T. W., Maximum likelihood estimates for multivariate normal distribution parameters when some observations are missing, *J. Am. Stat. Assoc.*, *52*, 200–203, 1957.

Benson, M. A., and N. C. Matalas, Synthetic hydrology based on regional statistical parameters, *Water Resour. Res.*, *3*(4), 931–935, 1967.

Cochran, W. G., *Sampling Techniques*, John Wiley, New York, 1953.

Fiering, M. B., On the use of correlation to augment data, *J. Am. Stat. Assoc.*, *57*, 20–32, 1962.

Fiering, M. B., Use of correlation to improve estimates of the mean and variance, *U.S. Geol. Surv. Prof. Pap.*, *434-C*, C1–C9, 1963.

Frost, J., and R. T. Clarke, Use of cross correlation between hydrological time series to improve estimates of lag one autoregressive parameters, *Water Resour. Res.*, *9*(4), 906–917, 1973.

Gilroy, E. J., Reliability of a variance estimate obtained from a sample augmented by multivariate regression, *Water Resour. Res.*, *6*(6), 1595–1600, 1970.

Hirsch, R. M., An evaluation of some record reconstruction techniques, *Water Resour. Res.*, *15*(6), 1781–1790, 1979.

Hirsch, R. M., A comparison of four record extension techniques, *Water Resour. Res.*, *18*(4), 1081–1088, 1982.

Kuczera, G., Effect of sampling uncertainty and spatial correlation on an empirical Bayes procedure for combining site and regional information, *J. Hydrol.*, *65*, 373–398, 1983.

Loucks, D. P., J. R. Stedinger, and D. A. Haith, *Water Resource Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, N. J., 1981.

Matalas, N. C., and B. Jacobs, A correlation procedure for augmenting hydrologic data, *U.S. Geol. Surv. Prof. Pap.*, *434-E*, E1–E7, 1964.

Matalas, N. C., and W. B. Langbein, Information content of the mean, *J. Geophys. Res.*, *67*(9), 3441–3448, 1962.

Moran, M. A., On estimators obtained from a sample augmented by multiple regression, *Water Resour. Res.*, *10*(1), 81–85, 1974.

Morrison, D. F., Expectations and variances of maximum likelihood estimates of the multivariate normal distribution parameters with missing data, *J. Am. Stat. Assoc.*, *66*(335), 602–604, 1971.

Natural Environment Research Council, *Flood Studies Report*, vol. 1, London, 1975.

Salas-La Cruz, J. D., Information content of the regional mean, in *Proceedings of the International Symposium on Uncertainties in Hydrologic and Water Resources Systems*, pp. 646–660, vol. 2, University of Arizona, Tucson, 1972.

Stedinger, J. R., Fitting log normal distributions to hydrologic data, *Water Resour. Res.*, *16*(3), 481–490, 1980.

Stedinger, J. R., Estimating correlations in multivariate streamflow models, *Water Resour. Res.*, *17*(1), 200–208, 1981.

Stedinger, J. R., Estimating a regional flood frequency distribution, *Water Resour. Res.*, *19*(2), 503–510, 1983.

Thomas, D. M., and M. A. Benson, Generalization of streamflow characteristics from drainage-basin characteristics, *U.S. Geol. Surv. Water Supply Pap.*, *1975*, 1970.

Wallis, J. R., and N. C. Matalas, Information transfer via regression in Markovian worlds, *Rep. RC-4207*, IBM Res. Cent., Yorktown Heights, N. Y., 1972.

Water Resources Council, Guidelines for determining flood-flow frequency, *Bull. 17B*, Hydrol. Comm., Washington, D. C., 1981.

Wilks, S. S., Moments and distributions of estimates of population parameters from fragmentary samples, *Ann. Math. Stat.*, *3*, 163–195, 1932.

J. R. Stedinger, Department of Environmental Engineering, Cornell University, Ithaca, NY 14853.

R. M. Vogel, Department of Civil Engineering, Tufts University, Medford, MA 02155.

(Received July 5, 1984;
revised December 26, 1984;
accepted January 15, 1985.)