WATER RESOURCES RESEARCH, VOL. 16, NO. 5, PAGES 863-870, OCTOBER 1980

# Bivariate Gamma Distributions for Extending Annual Streamflow Records From Precipitation: Some Large-Sample Results

R. T. CLARKE

*Instituto de Pesquisas Hidraulicas, Porto Alegre, RS Brasil*

This paper discusses the extension of streamflow records (commonly for annual time intervals) by correlation with longer records of precipitation for the purpose of estimating $\mu_Y$, the mean of the extended streamflow record. It is assumed that streamflow and precipitation have a bivariate gamma distribution, incorporating physical constraints on both variables, with $\mu_Y$ estimated by maximum likelihood, by a ratio estimate, and by a regression estimate. Large-sample variances of estimates given by these three estimation procedures are compared, and the effect on the ratio estimate of serial correlation in streamflow is calculated.

## INTRODUCTION

An earlier paper [*Clarke*, 1979] considered the effects, on the estimate $\hat{\mu}_Y$ of mean annual streamflow obtained when $n$ years of streamflow record are extended for a further $m$ years by regression on annual precipitation, of failure to take account of changes in the density of the rain gauge network used to calculate mean areal precipitation. These effects were explored for two alternative assumptions: (1) that annual mean areal precipitation $x$ and annual streamflow $y$ have a bivariate normal distribution and (2) that $x$ and $y$ have a bivariate gamma distribution, incorporating the common physical constraint that $x \geq y$. A referee of this earlier paper commented that it should be possible to derive simpler estimates of $\mu_Y$ (mean annual streamflow) than the maximum likelihood estimates discussed in it and pointed out the need for consideration of the effects of serial correlation in annual streamflow, assumed negligible in the paper. The purpose of the present work is to explore these suggestions for the case where $x$ and $y$ have the particular gamma distribution

$$p(x, y) = \frac{a^{p+q}}{\Gamma(p)\Gamma(q)} y^{p-1}(x - y)^{q-1} e^{-ax}\, dx\, dy \qquad x \geq y \geq 0$$

with three parameters $a$, $p$, $q$. Characteristics of this distribution, which is only one of a family of bivariate gammas, include the following:

1. The marginal distribution of annual precipitation $x$ and of annual streamflow $y$ are both gamma distributions of the form $(a^\lambda/\Gamma(\lambda))z^{\lambda-1}e^{-az}\, dz$; for annual precipitation, the parameter $\lambda$ is $p + q$, while for annual streamflow it is $p$.

2. The value of $y$ is restricted to be positive or zero and less than or equal to $x$. If attention is restricted to basins that are impermeable and have negligible year-to-year storage, this corresponds to the physical constraint that streamflow is less than precipitation.

3. The difference $x - y$ has a gamma distribution with $\lambda = q$. Again, for watertight basins with little overyear storage the difference $x - y$ estimates annual actual evaporation.

4. The regression of $y$ on $x$ is linear, with slope $p/(p + q)$, and passes through the origin, a physically sensible characteristic for watertight basins with negligible storage. The variance of residuals about the regression is also proportional to $x^2$.

5. The distribution $p(x, y)$ can easily be generalized to other hydrologically meaningful situations in which long records of precipitation and streamflow are to be used to extend shorter ones. Suppose, for example, that it were necessary to extend the record of streamflow $x$ at an upstream gauging site in a basin, given a longer streamflow record downstream $y$ and a longer record of precipitation $z$ in the basin above the upstream gauging post; if basin geology is such that the constraints $x \leq z$ and $x \leq y$ are physically reasonable, then the joint distribution of $x$, $y$, and $z$ with four parameters could be taken as

$$\frac{a^{m+n+p}}{\Gamma(m)\Gamma(n)\Gamma(p)} x^{n-1}(y - x)^{m-1}(z - x)^{p-1} e^{-a(y+z-x)}\, dx\, dy\, dz$$

It is clear that bivariate and multivariate gamma distributions offer considerable flexibility where it is necessary to incorporate nonnegativity and inequality constraints and where positive skewness is a characteristic of the distributions of hydrological variables concerned. Since the extension of streamflow records by correlation is commonly a necessary preliminary for many hydrological studies, we proceed to examine the properties of alternative estimates of $\mu_Y$ (mean annual streamflow) for different degrees of correlation between precipitation and streamflow and for varying lengths of the additional precipitation record. The paper considers only the case where record lengths are sufficiently large for asymptotic variance formulae to be applicable. This is clearly a considerable limitation in view of the short records commonly available. Without extensive Monte Carlo simulation, however, small sample properties of the estimates considered later in the paper are difficult to derive. Further assumptions that are common throughout the paper are (1) that annual streamflow is a stationary process fluctuating about a mean value $\mu_Y$ with variance $\sigma_Y^2$, (2) that for a given rain gauge network of $g$ gauges the estimated annual mean areal precipitation is a stationary process with mean $\mu_X(g)$ and variance $\sigma_X^2(g)$, where the symbol ( ) takes account of the fact that these quantities will depend upon the density and disposition of gauges. For much of the paper, also, we shall consider that both annual streamflow and annual precipitation are serially independent. Again, although the discussion is presented in terms of annual totals, the methods discussed are generally applicable to shorter time periods (say, a month or less) as long as the purpose is to esti-

Paper number 80W0578.
0043-1397/80/080W-0578$01.00

mate mean streamflow for such periods: the method will not be applicable, necessarily, where the objective is to reconstruct the sequence of monthly streamflow with its serial correlational structure. Thus, provided that October (November . . .) streamflow and October (November . . .) precipitation can be assumed to be pairwise bivariate gamma, the results of this paper are applicable to the estimation of mean October (November . . .) streamflow, given a longer record of monthly precipitation for the basin. It would not be correct, however, to use the methods to derive a reconstructed series of monthly streamflow, since any serial correlation between the deviations of monthly streamflows about their respective means would not be present in the reconstructed sequence.

The assumption of stationarity of streamflow (assumption 1 above) has, in practice, significant impact in planning and design. In the present paper, it has been adopted because of the lack of a more satisfactory alternative for the essentially theoretical treatment developed herein. In hydrological practice, the assumption of stationarity is difficult, even impossible, to validate; even where the assumption is not unreasonable, there remains the possibility that the sequence of years for which streamflow records are available constitute a run of wet or dry years within the stationary sequence. For example, flow in the Colorado River of the United States was apportioned between users following an analysis of pre-1920 streamflow records; the twenty or so years of that record proved to be a markedly wetter sequence than any subsequently observed.

Finally, the results of this paper will still apply where $x_i$ and $y_i$ are, say, annual maximum discharges, with the longer flow record at a downstream site giving rise to the sequence of annual maxima $\{x_i\}$ and with the shorter flow record at an upstream site giving rise to the sequence of annual maxima $\{y_i\}$. If the distributional assumptions are satisfied, the longer record $\{x_i\}$ may be used to assist in the estimation of the mean annual flood at the upstream gauging site; and for such an application it may well be reasonable to assume absence of serial correlation in either sequence.

## GAIN IN PRECISION OF ESTIMATED MEAN ANNUAL STREAMFLOW $\mu_Y$ ESTIMATED BY CORRELATION WITH LONGER RECORDS OF ANNUAL PRECIPITATION: MAXIMUM LIKELIHOOD METHOD

Using the bivariate gamma distribution given above, with annual precipitation denoted by $\{x_i\}$, $i = 1, \cdots, n + m$ and streamflow by $\{y_i\}$, $i = 1, \cdots, n$ and with the assumption that both $y_i$ and $x_i$ are serially independent, then the parameters $a$, $p$ (and hence the mean annual streamflow $\mu_Y = p/a$) may be estimated in two alternative ways: first, by using the available streamflow record $y_1, y_2, \cdots, y_n$ alone, and second, by using both the available streamflow record and the longer precipitation record $x_1, \cdots, x_{n+m}$, which is correlated with it.

If $p$, $a$ are estimated by maximum likelihood, we easily obtain for the first of these alternatives that $\mu_Y$ is estimated by $\hat{p}/\hat{a}$, the ratio of these estimates, and that the variance of this estimate is calculated as $\text{Var}\,\hat{\mu}_Y = \hat{p}/n\hat{a}^2$. For the second alternative we have

$$\begin{aligned}\log L = {} & n(p + q)\log a - n \log\Gamma(p) - n\log\Gamma(q) + (p - 1) \\ & \cdot \sum_{i=1}^{n} \log y_i + (q - 1) \sum_{i=1}^{n} \log (x_i - y_i) \\ & - a \sum_{i=1}^{n} x_i + m(p + q) \log a \\ & - m \log\Gamma(p + q) + (p + q - 1) \\ & \cdot \sum_{i=n+1}^{n+m} \log x_i - a \sum_{i=n+1}^{n+m} x_i \end{aligned} \qquad (1)$$

Maximum likelihood estimates of $a$, $p$, $q$ are obtained by solving the three equations $\partial \log L/\partial a = \partial \log L/\partial p = \partial \log L/\partial q = 0$; the variances of the maximum likelihood estimates satisfying these equations, together with the covariances between them, are given by calculation of the inverse matrix:

$$\begin{bmatrix} (n + m)(p + q)/a^2 & -(n + m)/a & -(n + m)/a \\ -(n + m)/a & n\psi_{pp}(p) + m\psi_{pp}(p + q) & m\psi_{pq}(p + q) \\ -(n + m)/a & m\psi_{pq}(p + q) & n\psi_{qq}(q) + m\psi_{qq}(p + q) \end{bmatrix}^{-1} \qquad (2)$$

where $\psi_{pp}(p + q)$, for example, is given by

$$\psi_{pp}(p + q) = \partial^2 \log\Gamma(p + q)/\partial p^2 \qquad (3)$$

and can be calculated from the asymptotic expansion:

$$\begin{aligned}\partial^2 \log\Gamma(x)/\partial x^2 = {} & (1/x) + 1/(2x^2) + 1/(6x^3) - 1/(30x^5) \\ & + 1/(42x^7) - 1/(30x^9) + 5/(66x^{11}) - \cdots \end{aligned} \qquad (4)$$

When the above inverse matrix is calculated, the first and second terms on the principal diagonal give Var $\hat{a}$ and Var $\hat{p}$, respectively, and the term in the first row, second column gives Cov $(\hat{a}, \hat{p})$. Using the relation

$$\text{Var}\,\hat{\mu}_Y \approx (\hat{p}^2/\hat{a}^2)\{\text{Var}\,\hat{p}/\hat{p}^2 + \text{Var}\,\hat{a}/\hat{a}^2 - 2\,\text{Cov}\,(\hat{a}, \hat{p})/(\hat{a}\hat{p})\} \qquad (5)$$

which is valid for large samples, we find that

$$\begin{aligned}\text{Var}\,\hat{\mu}_Y = {} & (n + m)^{-1}a^{-2}\{(n + m)(p + q)[n\psi_{qq}(q) + m\psi_{qq}(p + q)] \\ & - (n + m)^2 + p^2[n\psi_{pp}(p) \\ & + m\psi_{pp}(p + q)][n\psi_{qq}(q) + m\psi_{qq}(p + q)] \\ & - m^2\psi^2{}_{pq}(p + q) - 2p(n + m) \\ & \cdot [n\psi_{qq}(q) + m\psi_{qq}(p + q)] - 2mp(m + n)\psi_{pq}(p + q)\} \\ & \div \{(p + q)[[n\psi_{pp}(p) + m\psi_{pp}(p + q)] \\ & \cdot [n\psi_{qq}(q) + m\psi_{qq}(p + q)] \\ & - m^2\psi^2{}_{pq}(p + q)] - (n + m)[n\psi_{qq}(q) \\ & + m\psi_{qq}(p + q) - m\psi_{pp}(p + q)] \\ & + (n + m)[m\psi_{pq}(p + q) - n\psi_{pp}(p) - m\psi_{pp}(p + q)]\} \end{aligned} \qquad (6)$$

We note that when $m = 0$, so that records of both annual streamflow and precipitation are of the same length, this variance reduces to

$$\begin{aligned} & n^{-1}a^{-2}\{p^2\psi_{pp}(p)\psi_{qq}(q) + q\psi_{qq}(q) - p\psi_{qq}(q) - 1\} \\ & \qquad \div \{(p + q)\psi_{pp}(p)\psi_{qq}(q) - \psi_{qq} - \psi_{pp}\} \end{aligned} \qquad (7)$$

TABLE 1. Gain in Information (= Var $\hat{\mu}_Y$ $(n)$/Var $\hat{\mu}_Y$ $(m + n)$ for Values of Parameter $p$, Correlation Between Annual Precipitation and Streamflow, and Ratio $m/n$ of Additional Length of Precipitation Record to Length of Streamflow Record: Large Samples, Bivariate Gamma Distribution

| | m/n | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ρ | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 2.0 | 5.0 |
| | | | | $p = 2$ | | | | |
| 0.9 | 1.13 | 1.34 | 1.55 | 1.76 | 1.97 | 2.17 | 3.19 | 6.18 |
| 0.8 | 1.13 | 1.27 | 1.41 | 1.54 | 1.65 | 1.76 | 2.16 | 2.83 |
| 0.7 | 1.13 | 1.24 | 1.34 | 1.42 | 1.49 | 1.56 | 1.78 | 2.09 |
| 0.6 | 1.13 | 1.21 | 1.28 | 1.34 | 1.38 | 1.42 | 1.56 | 1.73 |
| 0.5 | 1.13 | 1.19 | 1.24 | 1.28 | 1.31 | 1.34 | 1.42 | 1.53 |
| 0.4 | 1.13 | 1.18 | 1.21 | 1.23 | 1.26 | 1.27 | 1.33 | 1.40 |
| | | | | $p = 5$ | | | | |
| 0.9 | 1.05 | 1.22 | 1.39 | 1.54 | 1.69 | 1.83 | 2.45 | 3.71 |
| 0.8 | 1.05 | 1.18 | 1.30 | 1.40 | 1.50 | 1.58 | 1.91 | 2.40 |
| 0.7 | 1.05 | 1.15 | 1.23 | 1.30 | 1.36 | 1.41 | 1.60 | 1.84 |
| 0.6 | 1.05 | 1.12 | 1.18 | 1.23 | 1.26 | 1.30 | 1.41 | 1.54 |
| 0.5 | 1.05 | 1.10 | 1.14 | 1.17 | 1.20 | 1.22 | 1.28 | 1.36 |
| 0.4 | 1.05 | 1.08 | 1.11 | 1.13 | 1.14 | 1.16 | 1.20 | 1.25 |
| | | | | $p = 10$ | | | | |
| 0.9 | 1.02 | 1.19 | 1.34 | 1.49 | 1.63 | 1.76 | 2.32 | 3.39 |
| 0.8 | 1.02 | 1.15 | 1.26 | 1.36 | 1.45 | 1.53 | 1.82 | 2.27 |
| 0.7 | 1.02 | 1.12 | 1.20 | 1.26 | 1.32 | 1.37 | 1.54 | 1.76 |
| 0.6 | 1.02 | 1.09 | 1.15 | 1.19 | 1.23 | 1.26 | 1.36 | 1.48 |
| 0.5 | 1.02 | 1.07 | 1.11 | 1.14 | 1.16 | 1.18 | 1.24 | 1.31 |
| 0.4 | 1.02 | 1.06 | 1.08 | 1.10 | 1.11 | 1.12 | 1.16 | 1.20 |
| | | | | $p = 15$ | | | | |
| 0.9 | 1.02 | 1.18 | 1.33 | 1.47 | 1.61 | 1.73 | 2.27 | 3.28 |
| 0.8 | 1.02 | 1.14 | 1.25 | 1.34 | 1.43 | 1.51 | 1.80 | 2.22 |
| 0.7 | 1.02 | 1.11 | 1.19 | 1.25 | 1.31 | 1.35 | 1.52 | 1.74 |
| 0.6 | 1.02 | 1.08 | 1.14 | 1.18 | 1.22 | 1.24 | 1.35 | 1.46 |
| 0.5 | 1.02 | 1.06 | 1.10 | 1.12 | 1.15 | 1.17 | 1.23 | 1.30 |
| 0.4 | 1.02 | 1.05 | 1.07 | 1.08 | 1.10 | 1.11 | 1.15 | 1.18 |
| | | | | $p = 20$ | | | | |
| 0.9 | 1.01 | 1.17 | 1.32 | 1.46 | 1.60 | 1.72 | 2.24 | 3.23 |
| 0.8 | 1.01 | 1.14 | 1.24 | 1.34 | 1.42 | 1.50 | 1.78 | 2.20 |
| 0.7 | 1.01 | 1.10 | 1.18 | 1.24 | 1.30 | 1.35 | 1.51 | 1.73 |
| 0.6 | 1.01 | 1.08 | 1.13 | 1.17 | 1.21 | 1.24 | 1.34 | 1.46 |
| 0.5 | 1.01 | 1.06 | 1.09 | 1.12 | 1.14 | 1.16 | 1.22 | 1.29 |
| 0.4 | 1.01 | 1.04 | 1.06 | 1.08 | 1.09 | 1.10 | 1.14 | 1.18 |
| | | | | $p = 50$ | | | | |
| 0.9 | 1.00 | 1.16 | 1.31 | 1.45 | 1.58 | 1.70 | 2.20 | 3.14 |
| 0.8 | 1.00 | 1.13 | 1.23 | 1.32 | 1.41 | 1.48 | 1.76 | 2.17 |
| 0.7 | 1.00 | 1.10 | 1.17 | 1.23 | 1.29 | 1.33 | 1.50 | 1.70 |
| 0.6 | 1.00 | 1.07 | 1.12 | 1.16 | 1.20 | 1.23 | 1.32 | 1.44 |
| 0.5 | 1.00 | 1.05 | 1.08 | 1.11 | 1.13 | 1.15 | 1.21 | 1.27 |
| 0.4 | 1.00 | 1.03 | 1.05 | 1.07 | 1.08 | 1.09 | 1.13 | 1.16 |

This is smaller than $p/(a^2n)$, the variance of the estimated mean annual streamflow $\mu_Y$ calculated when only the $n$ years of streamflow record are used, although the two expressions tend to equality when $p$ becomes large. Thus for long records of precipitation and streamflow for which the bivariate gamma assumption is justified, use of both records to estimate $p$ and $a$ gives smaller variance for the quantity $\hat{\mu}_Y$ that is of interest, even when these records of precipitation and streamflow are the same length.

If we write (1) Var $\hat{\mu}_Y(n)$ for the variance of mean annual streamflow calculated from the $n$ years of streamflow record alone and (2) Var $\hat{\mu}_Y(n + m)$ for the variance of mean annual streamflow calculated from both the $n$ years of common streamflow and precipitation record and also the additional $m$ years of precipitation measurements, then we can calculate

$$I(ML) = \text{Var } \hat{\mu}_Y(n)/\text{Var } \hat{\mu}_Y(n + m) \qquad (8)$$

which can be expected to be greater than unity because the precision of $\hat{\mu}_Y$ calculated from all the data should be greater than the precision calculated from the use of the short streamflow record alone. The extent to which $I(ML)$ is greater than unity is therefore a measure of the gain in information (reduction in sampling variance) that has resulted from 'extending' the streamflow record.

It is important to remember that $I(ML)$ is a measure only of the 'large-sample' gain in precision where both $n$ and $m$ can be assumed large enough for asymptotic variance formulae to apply. For short records, such as are commonly all that are available for water resource management studies, the gain in information need not always be positive; if $y_t$ and $x_t$ have a bivariate normal distribution, for example, with little correlation between them, it has been shown analytically [*Fiering*, 1962; *Moran*, 1974] that information may be lost by including the longer record $x_1, \cdots, x_{n+m}$ in the estimation procedure for

TABLE 2. Gain in Information ($I$ ($RAT$)) of Ratio Estimate, as Function of Correlation $\rho$ Between Annual Precipitation and Streamflow, and of Ratio $m/n$ of Additional Length of Precipitation Record to Length of Streamflow Record: Large Samples, Bivariate Gamma Distribution

| | $m/n$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\rho$ | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 2.0 | 5.0 |
| 0.9 | 1.00 | 1.16 | 1.30 | 1.44 | 1.56 | 1.68 | 2.17 | 3.08 |
| 0.8 | 1.00 | 1.12 | 1.22 | 1.32 | 1.40 | 1.47 | 1.74 | 2.14 |
| 0.7 | 1.00 | 1.09 | 1.16 | 1.22 | 1.27 | 1.32 | 1.48 | 1.69 |
| 0.6 | 1.00 | 1.06 | 1.11 | 1.16 | 1.19 | 1.22 | 1.32 | 1.43 |
| 0.5 | 1.00 | 1.04 | 1.08 | 1.10 | 1.12 | 1.14 | 1.20 | 1.26 |
| 0.4 | 1.00 | 1.03 | 1.05 | 1.06 | 1.08 | 1.09 | 1.12 | 1.15 |

$\hat{\mu}_Y$. The corresponding small-sample distribution of the estimate $\hat{\mu}_Y = \hat{p}/\hat{a}$ when $x$ and $y$ are gamma-distributed has not been found, however, and we are therefore unable to determine whether information can be lost by record extension when samples are small. In the absence of knowledge about the small-sample distribution of $\hat{\mu}_Y$ the possible existence of such losses of information can only be established by Monte Carlo simulation.

Although the quantity $I(ML)$ in (8) is no more than an asymptotic measure of the information gain resulting from extending the streamflow record $\{y_i\}$, it is of interest to tabulate it for selected values of the ratio $m/n$, for selected values of the correlation $\rho$ between annual precipitation and streamflow, and for a range of values of the parameter $p$. The values of $I(ML)$ are shown in Table 1.

Inspection of the table gives several results of interest. First, as the parameter $p$ increases, the gains in information decrease; remembering that the mean annual streamflow $\mu_Y = p/a$ increases with $p$ and that for large $p$ the gamma distribution will also tend to normality, we see that the gain in information increases with the skewness of the distribution of annual streamflow. Second, the gain in information increases with the length of the precipitation record, as is to be expected. Third, the gain in information is greater where the correlation between annual streamflow and precipitation is greater. Fourth, it can be seen that even where the precipitation record is of the same length as the record of annual streamflow ($m/n = 0$), there is a gain in information for the use of the bivariate gamma model as compared with the use of a univariate gamma model applied to the streamflow record alone. When $m/n = 0$, the gain in information is greater for smaller values of $p$ (that is to say, where the skewness of annual streamflow is greater). As $p$ becomes large and the distributions of both annual streamflow and annual precipitation tend to normality, the gain that results from using the precipitation record in a bivariate gamma model, when $m/n = 0$, decreases. This, again, is to be expected; with $m/n = 0$ and a bivariate normal model, all information on $\mu_Y$ is contained in the streamflow record itself.

## ALTERNATIVE ESTIMATES OF $\mu_Y$

The above calculation of $I(ML)$ assumed that the estimates $\hat{p}$, $\hat{a}$, and hence the estimate $\hat{\mu}_Y = \hat{p}/\hat{a}$, were obtained by the method of maximum likelihood; this requires iterative solutions for $a$, $p$, $q$ of the three equations obtained by setting the derivatives of log $L$ equal to zero. Simpler estimates, free from the necessity for iterative calculation, can also be obtained, one of the simplest being the ratio estimate

$$\hat{\mu}_Y(RAT) = \bar{y}_{(n)}\bar{x}_{(n+m)}/\bar{x}_{(n)} \qquad (9)$$

where $\bar{y}_{(n)}$ is the mean of the $n$ years of streamflow; $\bar{x}_{(n)}$, $\bar{x}_{(n+m)}$ are the mean annual precipitation for the $n$ years and $(n + m)$ years, respectively. Recalling that the expected values of $\bar{y}_{(n)}$, $\bar{x}_{(n)}$, and $\bar{x}_{(n+m)}$ are $p/a$, $(p + q)/a$, and $(p + q)/a$, respectively, it is clear that the ratio estimate $\hat{\mu}_Y(RAT)$ is an unbiassed estimate of mean annual streamflow when the available records are long. Its variance, again by use of the large-sample formula, is found to be

$$\text{Var } \hat{\mu}_Y(RAT) = p[(1 + m/n)q + p]/[na^2(1 + m/n)(p + q)] \qquad (10)$$

A quantity $I(RAT)$, defined as $(p/na^2)$/Var $\hat{\mu}_Y(RAT)$, can then be defined, giving a measure of the increase in information resulting from the use of the ratio estimate $\hat{\mu}_Y(RAT)$ of mean annual streamflow instead of from the use of the $n$ years of streamflow record alone; this quantity $I(RAT)$, which is independent of $p$, is given in Table 2 for values of the ratio $m/n$ and the correlation coefficient $\rho$ between precipitation and streamflow.

Inspection of Table 2 shows that the increase in precision of the estimate $\hat{\mu}_Y$ of mean annual streamflow obtained by the ratio estimate, using a longer record of annual precipitation, (1) increases with the additional length of precipitation record

TABLE 3. Efficiency (≡ Var $\hat{\mu}_Y$ ($RAT$, $m$, $n$)/Var $\hat{\mu}_Y$ ($ML$, $m$, $n$)) of Ratio Estimate ($\bar{y}_{(n)}\bar{x}(n + m)/\bar{x}(n)$) Relative to Maximum Likelihood Estimate of Mean Annual Streamflow $\mu_Y$, Found Through the Combined Use of $n$ Years of Streamflow Record and $(n + m)$ Years of Precipitation Record: Case $p = 2$ (Skewness of Streamflow Distribution = $\sqrt{2}$)

| | $m/n$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\rho$ | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 2.0 | 5.0 |
| 0.9 | 1.13 | 1.16 | 1.19 | 1.22 | 1.26 | 1.29 | 1.47 | 2.00 |
| 0.8 | 1.13 | 1.14 | 1.16 | 1.17 | 1.18 | 1.20 | 1.24 | 1.32 |
| 0.7 | 1.13 | 1.14 | 1.15 | 1.16 | 1.17 | 1.18 | 1.20 | 1.24 |
| 0.6 | 1.13 | 1.14 | 1.15 | 1.16 | 1.17 | 1.18 | 1.18 | 1.21 |
| 0.5 | 1.13 | 1.14 | 1.15 | 1.16 | 1.17 | 1.18 | 1.18 | 1.21 |
| 0.4 | 1.13 | 1.14 | 1.15 | 1.16 | 1.17 | 1.17 | 1.18 | 1.21 |

TABLE 4. Values of $I$ $(REG)$ for the Regression Estimate of Mean Annual Streamflow, for Varying Ratios of Record Length, and Various Correlations $\rho$ Between Annual Streamflow and Precipitation

| | $m/n$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\rho$ | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 2.0 | 5.0 |
| 0.9 | 1.00 | 0.92 | 0.91 | 0.93 | 0.96 | 1.00 | 1.22 | 1.82 |
| 0.8 | 1.00 | 0.93 | 0.93 | 0.94 | 0.97 | 1.00 | 1.16 | 1.55 |
| 0.7 | 1.00 | 0.95 | 0.94 | 0.96 | 0.98 | 1.00 | 1.12 | 1.37 |
| 0.6 | 1.00 | 0.96 | 0.96 | 0.97 | 0.98 | 1.00 | 1.09 | 1.25 |
| 0.5 | 1.00 | 0.97 | 0.97 | 0.98 | 0.99 | 1.00 | 1.06 | 1.16 |
| 0.4 | 1.00 | 0.98 | 0.98 | 0.99 | 0.99 | 1.00 | 1.04 | 1.10 |

available and (2) decreases with the correlation between annual precipitation and streamflow.

Comparison of Tables 1 and 2 shows that the behavior of the quantities $I(ML)$ and $I(RAT)$, regarded as functions of $\rho$ and $m/n$, is broadly similar for both maximum likelihood estimates (Table 1) and for the ratio estimate (Table 2). Moreover, by dividing the values in Table 1, for each $p$, by the values in Table 2, we can evaluate the quantity

$$\text{Var } \hat{\mu}_Y(RAT, m, n)/\text{Var } \hat{\mu}_Y(ML, m, n) \qquad (11)$$

which gives a measure of the precision of the ratio estimate of mean annual streamflow, relative to the precision of the maximum likelihood estimate; for large $p$, when the skewness of the distribution of both precipitation and streamflow is small, the ratio (11) is near to unity, particularly for small precipitation-streamflow correlations; for small $p$ and, consequently, large skewness, maximum likelihood estimates are considerably more efficient than ratio estimates, as Table 3 shows for the particular case $p = 2$. Thus for a precipitation record twice as long as the streamflow record ($m/n = 1$), ratio estimates are of the order of 20% less efficient than maximum likelihood estimates, where the correlation between precipitation and streamflow is of the order 0.8 or higher.

## A REGRESSION ESTIMATE OF $\mu_Y$

For a bivariate gamma distribution the regression of $y$ on $x$ is linear, passing through the origin, with variance of deviations about the regression proportional to $x^2$. If a least squares regression line were to be calculated by minimization of

$$\sum_{i=1}^{n} (y_i - \beta x_i)^2/x_i^2$$

the least squares estimate of $\beta$ is $\hat{\beta} = n^{-1} \sum(y_i/x_i)$, conditional on the set of annual precipitations $x_{n+1}, \cdots, x_{n+m}$ in the longer record; therefore, an estimate of $\mu_Y$ can be taken as

$$(n\bar{y}_{(n)} + m\beta\bar{x}_{(m)})/(n + m) \qquad (12)$$

This estimate is clearly a function of the particular precipitation values observed; however, if these independent variables were to be regarded as random variables, distributed independently about a mean $(p + q)/a$ with variance $(p + q)/a^2$, the above estimate (12) for unconditional $x$ values is an unbiased estimate of $\mu_Y$ when records are long. The (unconditional) variance of this estimate, again for large samples, can be shown to be

$$\text{Var } \hat{\mu}_Y(REG) = p/(n + m)^2 \{(np(1 + 3m) + q(m + n)^2)/(n(p + q)a^2)\} \qquad (13)$$

and, just as with the ratio estimate considered above, an information criterion can be defined as

$$I(REG) \equiv (p/a^2n)/\text{Var } \hat{\mu}_Y(REG)$$

or

$$I(REG) = (1 + m/n)^2(p + q) / \{p(1 + 3m/n) + q(1 + m/n)^2\} \qquad (14)$$

This quantity can be tabulated for varying proportions $m/n$ and correlation $\rho$ between annual precipitation and streamflow; values are given in Table 4.

Table 4 gives a number of results of interest. When the annual precipitation and streamflow records are of equal length ($m/n = 0$), the gain in information by using both records to estimate mean annual streamflow is clearly zero, since the regression estimate reduces to the mean of the $n$ available values of streamflow; what is less obvious, however, is why the variance ratio $I(REG)$ should equal unity also when $m/n = 1$ (that is, when the record of annual precipitation is twice the length of the streamflow record). A second point to be considered is the difference in behavior of the ratio $I(REG)$ for $(m/n) < 1$ and for $(m/n) > 1$. In the former case the precision of the regression estimate, relative to the precision of the estimate obtained from the $n$ years of streamflow record alone, increases as correlation between annual precipitation and streamflow decreases, while the converse is true for $(m/n) > 1$ (that is, for a precipitation record more than twice the length of the streamflow record).

The explanation for these results can be derived by consideration of the variance of $\hat{\mu}_Y$ conditional on the $(n + m)$ values of annual precipitation. When precipitation and streamflow values are bivariate-gamma distributed, the conditional variance of the estimate (12) is given by

$$\text{Var } (\hat{\mu}_Y|\text{x}, m, n) = p/(na^2(n + m)^2)[n^2 + m^2\bar{x}_{(m)}{}^2\sum(1/x_i^2) \div n + 2m\bar{x}_{(m)}\sum(1/x_i)] \qquad (15)$$

This variance will be less or greater than $p/na^2$, the variance of the estimate of $\mu_Y$ derived from the streamflow record alone, if

$$n^2 + 2m\bar{x}_{(m)}\sum(1/x_i) + m^2\bar{x}_{(m)}{}^2\sum(1/x_i^2)/n \lesseqgtr (n + m)^2 \qquad (16)$$

or

$$n^2 + 2mn[\bar{x}_{(m)} \cdot \sum(1/x_i)/n] + m^2[\bar{x}_{(m)}{}^2\sum(1/x_i^2)/n] \lesseqgtr (n + m)^2 \qquad (17)$$

Now consider the terms in square brackets; the expectation of $\bar{x}_{(m)}$ is $(p + q)/a$, and of $\sum(1/x_i)/n$, $a/(p + q)$. When $m$ and $n$ are both large and approximately equal ($m/n \approx 1$), therefore the terms in square brackets are approximately one, and the equality sign holds in (17). For values of $m$ smaller than $n$,

TABLE 5. Ratio of Variance of Regression Estimate $(n\bar{y}_{(n)} + m\hat{\beta}\bar{x}_{(m)})/(n + m)$ to Variance of Maximum Likelihood Estimate of Mean Annual Streamflow $\mu_Y$, Found Through the Combined Use of $n$ Years of Streamflow Record and $(n + m)$ Years of Precipitation Record: Case $p = 2$ (Skewness of Streamflow Distribution: $\sqrt{2}$)

| | $m/n$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\rho$ | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 2.0 | 5.0 |
| 0.9 | 1.13 | 1.45 | 1.70 | 1.89 | 2.05 | 2.17 | 2.61 | 3.40 |
| 0.8 | 1.13 | 1.36 | 1.52 | 1.64 | 1.70 | 1.76 | 1.86 | 1.82 |
| 0.7 | 1.13 | 1.30 | 1.42 | 1.48 | 1.52 | 1.56 | 1.59 | 1.52 |
| 0.6 | 1.13 | 1.26 | 1.33 | 1.38 | 1.41 | 1.42 | 1.43 | 1.38 |
| 0.5 | 1.13 | 1.23 | 1.32 | 1.31 | 1.32 | 1.34 | 1.34 | 1.32 |
| 0.4 | 1.13 | 1.20 | 1.23 | 1.24 | 1.27 | 1.28 | 1.28 | 1.27 |

however, the imprecision of $\bar{x}_{(m)}$ is likely to be greater than that of $\sum(1/x_i)/n$, (with similar arguments applied to $\bar{x}_{(m)}^2$, $\sum(1/x_i^2)/n$) so that the 'greater' than sign may be expected to hold in the inequality (17); the corresponding value of $I(REG)$ will therefore be less than unity. A similar argument applies for $m > n$. This therefore affords a qualitative explanation of why $I(REG)$ is less than, equal to, and greater than unity according as $m/n$ is less than, equal to, or greater than unity.

## LARGE-SAMPLE EFFICIENCY OF THE REGRESSION ESTIMATE OF MEAN ANNUAL STREAMFLOW, RELATIVE TO THE MAXIMUM LIKELIHOOD ESTIMATE

The values of $I(REG)$ in Table 4 show the value of using a longer record of annual precipitation to supplement the information on mean annual streamflow $\mu_Y$ given by a shorter streamflow record, when $\mu_Y$ is estimated by (12). By calculating the ratio of the entries from Tables 1 and 4 we obtain a measure of the efficiency of maximum likelihood estimation of $\mu_Y$ relative to the regression estimate. Table 5 gives the ratio of variances

$$R = \text{Var}\ \hat{\mu}_Y(REG)/\text{Var}\ \hat{\mu}_Y(ML) \qquad (18)$$

for the particular case $p = 2$.

A large value for an entry in Table 5 indicates that the regression estimate has much less precision (much greater variance) than the maximum likelihood estimate of $\mu_Y$, for the same values of $m/n$ and $\rho$. The same is true of the entries in Table 3, calculated for the ratio estimate; Tables 3 and 5 can therefore be compared directly to show how the precision of ratio estimates compares with the precision of regression estimates, both relative to maximum likelihood estimates. This comparison shows that where the skewness in annual streamflow is as large as $\sqrt{2}$, regression estimates are considerably less efficient than ratio estimates whatever $m/n > 0$ and whatever the value of $\rho$.

## THE EFFECT OF SERIAL CORRELATION ON THE RATIO ESTIMATE $\hat{\mu}_Y(RAT) = \bar{y}_{(n)}\bar{x}_{(n+m)}/\bar{x}_{(n)}$

Hitherto, we have assumed that both annual precipitation $x_t$ and annual streamflow $y_t$ were serially independent, a not unreasonable assumption where the river basin has little year-to-year storage. Where streamflow has a large base flow component, however, or where the stream is supplied from an upland marsh or lake, the assumption of serial independence will be invalidated, and it becomes necessary to consider how the simple model of cross-correlated $x$, $y$ may be extended to include serial correlation in $y$, or possibly in $x$ and $y$.

To see how this may be effected, it is necessary to reconsider the method by which the bivariate gamma distribution was originally developed. If $u_i$ and $v_i$ are independent random variables drawn from the same normal distribution with zero mean, then $U = u_1^2 + u_2^2 + \cdots + u_p^2$ and $V = v_1^2 + v_2^2 + \cdots + v_q^2$ have distributions

$$a^p U^{p-1} e^{-aU}/\Gamma(p) \qquad a^q V^{q-1} e^{-aV}/\Gamma(q)$$

respectively. A change of variable $x = U + V$, $y = U$ then gives the bivariate gamma distribution of $x$, $y$ $(x \geq y \geq 0)$ as

$$y^{p-1}(x - y)^{q-1} e^{-ax}/(\Gamma(p)\Gamma(q))$$

Suppose now that we consider the joint distribution of $z_1 = u_1^2 + \cdots + u_{p-r}$, $z_2 = v_1^2 + v_2^2 + \cdots + v_r^2$, $z_3 = w_1^2 + w_2^2 + \cdots + w_{p+q-r}^2$, and make the transformation

$$U = z_1 + z_2$$

$$W = z_2$$

$$V = z_2 + z_3$$

Then the joint distribution of $U$, $V$ is found, after integrating out the unwanted variable $W$, as

$$p(U, V) = \frac{a^{2p+q-r}}{\Gamma(p - r)\Gamma(r)\Gamma(p + q - r)} e^{-a(U+V)} \cdot \left| \int_{W=0}^{\min(U,V)} \phi\, W^{r-1}\, dW \right|$$

where $\phi = e^{aW}(U - W)^{p-r-1}(V - W)^{p+q-r-1}$.

The variables $U$ and $V$ clearly have marginal distributions that are univariate gamma $a^p U^{p-1} e^{-aU}/\Gamma(p)$ and $a^{p+q} V^{p+q-1} e^{-aV}/\Gamma(p + q)$, respectively, so that their mean values are $p/a$ and $(p + q)/a$; their covariance, defined as

$$E(UV) - p(p + q)/a^2$$

can be shown to be

$$\frac{a^{2p+q-r}}{\Gamma(p - r)\Gamma(r)\Gamma(p + q - r)} \left\{ \int_{U=0}^{\infty} \int_{V=0}^{U} UVe^{-a(U+V)} \cdot \int_{W=0}^{V} \phi\, W^{r-1}\, dW + \int_{V=0}^{\infty} \int_{U=0}^{V} UVe^{-a(U+V)} \cdot \int_{w=0}^{U} \phi\, W^{r-1}\, dW \right\} - p(p + q)/a^2 \qquad (19)$$

The expression (19), divided by $(p(p + q)/a^2)^{1/2}$, gives the correlation between $U$ and $V$. For $p$, $q$, $r$ integers the triple integrals in (19) can readily be evaluated in terms of gamma functions. For the case $q = 0$, $U = y_t$, $V = y_{t-1}$ we then obtain an expression for the serial correlation between annual streamflow in one year and the year preceding it; for the case $q \neq 0$, $U = y_i$, $V = x_{i-1}$ the expression gives the cross correlation between streamflow in year $i$ and precipitation in year $i - 1$. We note that if the variables $U$, $V$, $W$ in the above development are all gamma-distributed, a correlation between annual streamflow $y_i$, $y_{i-1}$ implies the existence of a serial correlation in annual precipitation, although this will be small if $q$ is large relative to $r$. Finally, if we write $r = \lambda p$, we can use the expression (19) to calculate the serial correlation between $y_i$, $y_{i-j}$, when the serial correlation for annual streamflow follows a lag 1 autoregression by substituting $\lambda^j p$ instead of $r$. Likewise, the cross correlation of lag $j$ may also be determined. The appendix gives the series expansion for $E(U^r V^s)$ from which these correlations may be derived.

To examine the effect of neglecting to take account of serial correlation in annual streamflow, suppose that serial cross correlations of lag higher than 1 can be neglected. For simplicity we restrict attention to the ratio estimate of mean annual streamflow, $\hat{\mu}_Y = \bar{y}_{(n)}\bar{x}_{(n+m)}/\bar{x}_{(m)}$; this estimate is unbiased in large samples, whether or not serial correlations exist amongst annual streamflow, and serial correlation affects only the variance of this estimate. If serial correlation between streamflows $y_t$ and $y_{t-1}$ is denoted by $\rho_{yy}(1)$, etc. and serial correlation between $y_t$ and $y_{t-j}$ can be neglected for $j > 1$, we have (setting $a = 1$)

$$\operatorname{Var} \bar{y}_{(n)} = (p/n)\,(1 + 2\rho_{yy}(1))$$

Similarly, we have

$$\operatorname{Var} \bar{x}_{(n)} = ((p + q)/n)\,(1 + 2\rho_{xx}(1))$$

$$\operatorname{Var} \bar{x}_{(n+m)} = ((p + q)/(n + m))\,(1 + 2\rho_{xx}(1))$$

$$\operatorname{Cov}\,(\bar{y}_{(n)}, \bar{x}_{(n)}) = n^{-1}(\rho_{xy}(0) + 2\rho_{xy}(1))p^{\frac{1}{2}}(p + q)^{\frac{1}{2}}$$

$$\operatorname{Cov}\,(\bar{y}_{(n)}, \bar{x}_{(n+m)}) = (n + m)^{-1}(\rho_{xy}(0) + 2\rho_{xy}(1))p^{\frac{1}{2}}(p + q)^{\frac{1}{2}}$$

$$\operatorname{Cov}\,(\bar{x}_{(n)}, \bar{x}_{(n+m)}) = (n + m)^{-1}(1 + 2\rho_{xx}(1))\,(p + q)$$

giving

$$\operatorname{Var} \hat{\mu}_Y = \frac{p}{n}\left[\frac{(n + m)\,q + np}{(n + m)(p + q)} + 2\rho_{yy}(1) + \frac{2m\rho_{xx}(1)\,{\rho_{xy}}^2\,(0)}{(n + m)} - \frac{4m}{(n + m)}\,\rho_{xy}(0)\rho_{xy}(1)\right] \qquad (20)$$

It can be shown, either by the use of the formulae given in the appendix or by more intuitive methods, that $\rho_{yy}(1) = r/p$, $\rho_{xx}(1) = r/(p + q)$, and $\rho_{xy}(1) = r/\{p(p + q)\}^{1/2}$, where $0 \leqslant r \leqslant p$; substitution of these values in the expression (20) gives

$$\operatorname{Var} \hat{\mu}_Y = p\,\{(n + m)q + np\}/\{n(n + m)(p + q)\} + 2r\,\{p^2/(n + m) + 2pq/(n + m) + q^2/n\}/(p + q)^2$$

The second term in this expression measures the extent to which the variance of mean annual streamflow (obtained by the ratio estimate) is underestimated if no account is taken of serial correlation in streamflow when (1) $\mu_Y$ is estimated from the extended sequence of annual streamflow and (2) annual streamflow and precipitation are assumed to follow a bivariate gamma distribution. The percentage increase in Var $\hat{\mu}_Y$, which follows from the existence of serial correlation in streamflow, is therefore

$$200\lambda[1 + m(1 - \rho^2)^2/n]/[1 + m(1 - \rho^2)/n]$$

where $\rho$ is the zero lag cross correlation between the $x$ and $y$ series, and $\lambda$ is the lag 1 serial correlation in annual streamflow. The numerical value of this expression can be considerable; thus if $\lambda = 0.1$, $m/n = 1$, the increase in Var $\hat{\mu}_Y$ that results from the serial correlation in annual streamflow is of the order of 17%, changing little for different values of $\rho$.

## CONCLUSIONS

This paper has discussed the properties of three estimates of mean annual streamflow $\mu_Y$ derived from both (1) a sequence of $n$ years of streamflow record and (2) an additional $m$ years of precipitation record, under the assumptions that (1) annual streamflow and precipitation have a bivariate gamma distribution and (2) the record lengths $n$ years (streamflow) and ($n + m$) year (precipitation) are sufficiently long for large-sample formulae to be applicable when Var $\hat{\mu}_Y$ is calculated. With these restrictive assumptions the following results apply.

1. Where the precision of the maximum likelihood estimate, derived from both streamflow and precipitation records, is compared with the precision of the maximum likelihood estimate derived for only the $n$ years of streamflow record, the increase in precision that results from using the longer precipitation record is greater when the skewness of the annual streamflow distribution is greater. This conclusion applies for whatever values of $m$, $n$, and whatever the correlation $\rho$ between precipitation and streamflow.

2. Not unexpectedly, the increase in the precision of the maximum likelihood estimate $\hat{\mu}_Y$ which results from using both precipitation and streamflow records (instead of the streamflow record alone) increases as the ratio $m/n$ increases; the greater the correlation between streamflow and precipitation, the greater the gain in precision of the estimate $\hat{\mu}_Y$.

3. Where the annual streamflow distribution has large skewness, inclusion of the precipitation record in the estimation of mean annual streamflow $\mu_Y$ improves the precision of the maximum likelihood estimate even when it is of the same length as the record of streamflow. Conversely, when skewness is small and both precipitation and streamflow records are approximately normally distributed (the gamma distribution of each tending to normality), a precipitation record of the same length as the streamflow record gives no improvement in the precision of the estimate $\hat{\mu}_Y$, since all information about mean annual streamflow is then contained in the streamflow record itself.

4. The ratio estimate of mean annual streamflow has greater precision than the estimate obtained from the streamflow record alone, the gain in precision being greater for larger values of the ratio $m/n$ and for larger correlation between precipitation and streamflow. Where skewness of the annual streamflow distribution is small, the ratio estimate differs but little in precision from the maximum likelihood estimate, but for large skewness and high correlation between streamflow and precipitation, the loss in information that results from using the computationally simpler ratio estimate can be considerable.

5. A regression estimate of $\mu_Y$, calculated using both $m + n$ years of precipitation record and $n$ years of streamflow record, has lower (large sample) precision than the estimate derived from the streamflow record alone, unless the precipitation record is more than twice as long as the streamflow record. The precision of the particular regression estimate used was always considerably less than the maximum likelihood estimate and also less than the ratio estimate giving conclusion 4 above.

6. If a ratio estimate is used to estimate mean annual streamflow, a formula is derived for its large-sample variance where serial correlation exists between annual streamflow totals and where streamflow and precipitation have bivariate gamma distribution. The correlational structure of both series was assumed to be such that both serial and cross correlations of lag greater than 1 were negligible. Where the record of annual precipitation was twice the length of the streamflow record, the percentage increase in Var $\hat{\mu}_Y$ was considerable even for serial correlation in streamflow as small as 0.1.

This paper must end on a cautionary note. The particular bivariate gamma model that is the subject of this paper is not in common hydrological use, and many questions remain to be answered before, and if ever, it is to be recommended for hydrological practice. These questions include the following: How would a user know if his data were bivariate gamma? What loss is incurred if a bivariate gamma were to be used when an alternative distribution would be more appropriate? No more than tentative answers to such questions can be suggested; in reply to the former it can be said that the small samples of hydrological record that are commonly available will seldom make possible, by statistical test, a clear discrimination in favor of the bivariate gamma (or, indeed, any other) distribution. The sole argument in favor of the bivariate gamma for the problem considered in this paper is that it describes some of the physical constraints on streamflow and precipitation sequences (or, equally, of certain flood sequences) where distributions in common use may not. To this extent, therefore, it could be said that the results of this paper add to

> the growing complexity of concepts [that] has necessitated the use of more complicated mathematical tools for their formulation ... and more data for model selection, parameter estimates, and testing

[*Klemes,* 1977] when

> the plain fact is that we usually do not know how much a more detailed model, a longer data base, more accurate parameter estimates, in short, an increase in information, have improved the engineering solution being sought.

The second question, set out above, is equally difficult; the results of *Slack et al.* [1975] raise serious doubts about the merits of modern sophisticated frequency analysis, and what is really required are distribution-free methods that are robust in practice. One such method, appropriate to the problem of this paper and suitable for estimating the quantiles of the unspecified distribution of the variable $y_i$, instead of its mean value $\mu_Y$, is the subject of another paper [*Cooper and Clarke,* 1980].

## APPENDIX

Denoting the moments about the origin of the distribution (19) by $E(U^xV^y)$, we have to evaluate this expectation for $(x, y)$ equal to (1, 0), (0, 1), (1, 1), (2, 0). For integral $p$, $q$, $r$, $x$, $y$ and with $\phi$ denoting $e^{aW}(U-W)^{p-r-1}(V-W)^{p+q-r-1}$, we have

$$\begin{aligned}E(U^xV^y) = \int_{U=0}^{\infty}\int_{V=0}^{\infty} U^xV^ye^{-a(U+V)}\\ \cdot \int_{W=0}^{\min(U,V)} \phi\, W^{r-1}\, dW\, dV\, dU/K\end{aligned}$$

with $K = (p-r-1)!(r-1)!(p+q-r-1)!$. If $\phi$ is expanded binomially and the result integrated by parts, it is found that

$$\begin{aligned}E(U^xV^y) = {} & (-1)^r/(r-1)! \sum_{i=0}^{p-r-1} \{i!(p-r-i-1)!\}^{-1}\\ & \cdot \sum_{j=0}^{p+q-r-1} (i+j+r-1)! \times \{j!(p+q-r-j-1)!\}^{-1}\\ & \cdot \{2(p+x-r-i-1)!(p+q+y-r-j-1)!\\ & + \sum_{k=0}^{i+j+r-1} (-1)(2p+q-2r+x+y+k-i-j\\ & -1)!/k! \times |(p+x+k-r-i)^{-1} + (p+q+y+k\\ & -r-j)^{-1}| - \sum_{1=0}^{p+q+y-r-j-1} (p+q+y-r-j-1)!(p\\ & +x+1-r-i-1)!/\{1!2^{p+x+1-r-i}\}\\ & - \sum_{1=0}^{p+x-r-i-1} (p+x-r-i-1)!(p+q-r+y\\ & +1-j-1)!/\{1!2^{p+q-r+y+1-j}\}^{-1}\}\end{aligned}$$

Substituting the appropriate values of $x$, $y$, we have that the correlation between $U$ and $V$ is given by

$$\begin{aligned}[E(UV)-E(U)E(V)]/[E(U^2)-E^2(U)]^{1/2}[E(V^2)\\ -E^2(V)]^{1/2} = r/(p(p+q))^{1/2}\end{aligned}$$

## REFERENCES

Clarke, R. T., Extension of annual streamflow record by correlation with precipitation subject to heterogeneous errors, *Water Resour. Res., 15*(5), 1081-1088, 1979.

Cooper, D. M., and R. T. Clarke, Distribution-free methods for estimating flood and streamflow exceedance probabilities by correlation, *Water Resour. Res., 16,* in press, 1980.

Fiering, M. B., On the use of correlation to augment data, *J. Amer. Statist. Ass., 57,* 20-28, 1962.

Klemes, V., Value of information in reservoir optimization, *Water Resour. Res., 13*(5), 837-850, 1977.

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

Slack, J. R., J. R. Wallis, and N. C. Matalas, On the value of information to flood frequency analysis, *Water Resour. Res., 11*(5), 629-647, 1975.

(Received October 10, 1979;
revised April 10, 1980;
accepted April 23, 1980.)