AUGMENTING ANNUAL RUNOFF RECORDS USING TREE-RING DATA

Charles W. Stockton and Harold C. Fritts

Laboratory of Tree-Ring Research
The University of Arizona, Tucson, Arizona 85721

INTRODUCTION

Any statistical work involving hydrologic records is handicapped when the records are of relatively short duration, as are most such records in the Southwestern United States. This is because the short records are not necessarily a random sample of the infinite population of events, and consequently any statistical descriptions are likely to be in error to some extent.

Work recently completed at the Laboratory of Tree-Ring Research [*Stockton*, 1971] has shown that tree ring data can be used to extend available runoff records backward in time, thereby providing a longer record from which to more accurately estimate the three most common statistics used in hydrology: the mean, the variance, and the first order autocorrelation.

STATISTICAL PARAMETERS

In statistical analysis of hydrologic phenomena, it is usually assumed that a record of events that is of finite length represents a random sample from an infinite population, the occurrence of each event being governed by some probability distribution. Any change in the hydrologic regime with which a given record of events is associated results in a change in the probability distribution.

For practical purposes, a probability distribution is described by the mean (a measure of central tendency), the variance (a measure of the average spread of the events about the mean), and the skewness (a measure of the asymmetry of the distribution of the events about the mean). In some cases these three parameters uniquely define a probability distribution and are useful for describing hydrologic phenomena. For most annual runoff and tree-ring index series, the variables are normally distributed (skewness equals zero) and the probability distribution is completely described by the mean and variance. In almost every mathematical model of runoff time series, the first order autocorrelation (a measure of persistence in a series of events) is used along with the mean and variance.

Reprinted from
HYDROLOGY AND WATER RESOURCES IN ARIZONA AND THE SOUTHWEST, Vol. 1.
Proceedings of the 1971 meetings of The Arizona Section - American Water Resources Association and the Hydrology Section-Arizona Academy of Science, April 22-23, 1971, Tempe, Arizona.

The population values of these statistics are usually unknown and therefore must be estimated from the existing record of observations. Consequently, the reliability of the estimates depends primarily upon the length of record of the observations--in other words, the total number of observations.

If there are errors in the estimates of the population parameters owing to shortness of observed records, these errors are preserved in any synthetic series that is generated from the available data. Recently, *Rodríguez-Iturbe* [1969] showed that if the length of an annual runoff record is 40 years or less, there may be an error of 2% to 20% in estimation of the mean, from 15% to 60% in the estimation of the variance, and as much as 200% in the estimation of the first order autocorrelation. The high error in the autocorrelation is probably related to the inadequacy of short records for estimation of the low-frequency persistence in climatologic data, which *Mandelbrot and Wallis* [1968] have dubbed the "Noah and Joseph effects" after the well-known Biblical calamities.*

Fiering [1962], *Matalas and Jacobs* [1964], and *Julian and Fritts* [1968] have demonstrated the use of the correlation technique for augmenting hydrologic records. In each of these cases, a *single* record was used to augment another. *Fiering* [1963] also approached the problem using multiple linear regression--that is, using several independent variables to predict a dependent variable. He showed that a better estimate of the mean can be obtained in the multivariate case if $R^2 \geqslant q_i/(n_1 - q_i)$, where R is the combined correlation coefficient, q_i is the number of variables included in the prediction equation, and n_1 is the length of the record to be extended. In the case of the variance, the variance of the reconstructed record is a better estimate if the relative information ratio I (the ratio of the variance to that estimated from the original record) exceeds 1. When I exceeds unity, it implies that the variance of the estimate of a moment made from the original record is longer than that of the estimate made from the combined record, and therefore a more precise estimate is computed from the combined data. As a general rule the estimate from the longer series is more reliable if R exceeds 0.80 [Table 3 of *Fiering,* 1963; p. 2 of *Matalas and Jacobs,* 1964]. However, *Matalas and Jacobs* [1964] point out that these requirements can be reduced and that the parameters estimated from the longer series are an unbiased estimate if a noise factor is added to the estimated values.

*By "Noah effect" is meant that extreme precipitation tends to be very extreme, the archetype being the 40-day rains that resulted in inundation of the entire earth (Gen. 7:11-21). By "Joseph effect" is meant that a period of unusually high (or low) precipitation is commonly an extended one, so named after the widespread famine of seven years' duration that Joseph had predicted from Pharaoh's dream (Gen. 41:51-57).

CLIMATIC INPUTS

The basis for comparing annual runoff series with tree-ring series is the hypothesis that the two series respond to a common climatic signal or signals that permit prediction of annual runoff from the annual ring-width index.* A schematic diagram of the climatic variables influencing both of the series and the resultant predictability is shown in Figure 1.

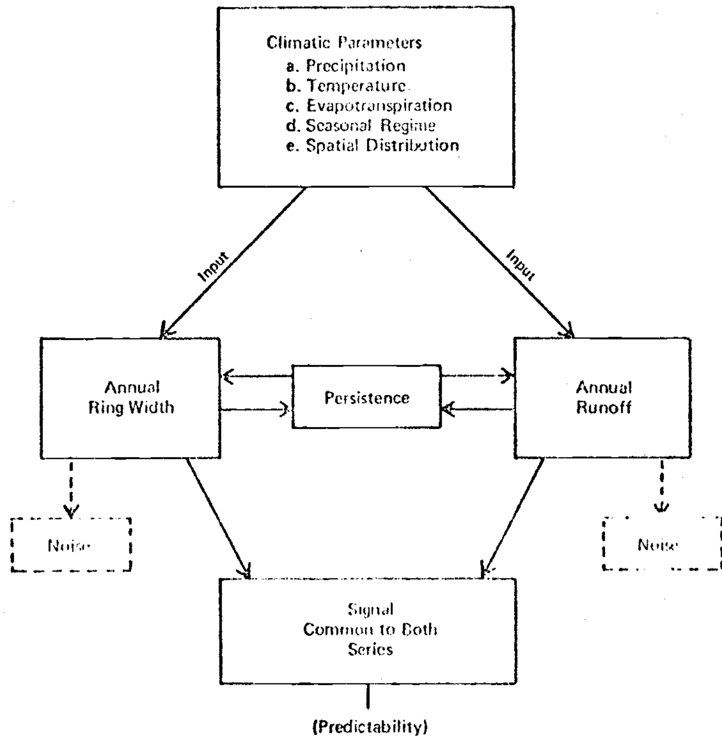

Fig. 1. Schematic diagram of relationship between ring-width series and annual runoff series for medium and large watersheds.

Precipitation (a), temperature (b), and evapotranspiration (c) influence the water balance of both runoff and tree growth. However, in the case of tree growth, these variables, and especially temperature, have physiological influences not directly related to the water balance; these influences are diagrammed in *Fritts et al.* [1970]. The seasonal distribution of the variables (d) influences both runoff and tree growth, and in the case of tree growth the influence of the monthly distribution extends to at least a 14-month period--from the July prior to the growing season in which the ring is

*"Indexing" (standardization) is necessary to convert the nonstationary ring-width series to a stationary time series [*Stokes and Smiley*, 1968].

formed to the July concurrent with the growing season [*Fritts et al.*, 1970]. Spatial distribution of precipitation and temperature (c) within large watersheds may influence both the annual runoff regime and the variability in growth of trees from site to site.

The noise component in Figure 1 represents both the model's inability to adequately describe the two series and the differences in the way the two series respond to climatic inputs.

Of major concern in the reconstruction of annual runoff series from tree-ring records is the difference in persistence within each of the two series—that is, how much do events of the previous year or years influence the current year? During this study, differences in persistence were resolved by using lagged dependent variables on the right-hand side of the reconstruction equation, as described by *Johnston* [1963]. Unfortunately, this causes the residuals to be dependent upon residuals of prior reconstructed values. Also, the regression coefficients tend to be biased although they have the properties of consistency and efficiency [*Johnston*, 1963] if the residuals are normally distributed. Another remedy would be to use a matrix of the tree-ring data, lagged up to three times, and extract principal components from this supplemental matrix. The covariation in this matrix can be decomposed by extracting the eigenvectors. A new set of uncorrelated variables is obtained from the amplitudes of the eigenvectors [*Fritts et al.*, 1970]. These amplitudes may be lagged in certain ways with the runoff data, and multiple regression may be used to weight the respective series so that the differences in persistence are accounted for.

EFFECTS ON TREE GROWTH AND RUNOFF

It is now necessary to determine how both tree growth and runoff respond to the climatic inputs. *Fritts et al.* [1970] described a method for modeling the response of trees to different climatic variables. Their method, which provides a means of determining the importance of monthly temperature and precipitation throughout the 14-month period prior to actual growth, uses multiple linear regression to predict ring-width indices from the amplitudes of eigenvectors of monthly precipitation and temperature along with variables representing the persistence within the ring-width series. That is, the tree-ring indices are fit to the model

$$y_t = \theta_1 \xi_{1t} + \theta_2 \xi_{2t} + \ldots + \theta_p \xi_{pt} + \phi_1 y_{t-1} + \ldots + \phi_3 y_{t-3} + e_t, \qquad (1)$$

where

y_t = normalized ring-width index in year t
θ_p = least squares coefficient for variable ξ_p
ξ_{pt} = amplitude for year t of p eigenvectors extracted from a correlation matrix of climatic variables
ϕ_{t-n} = least squares coefficient for variable y_{t-n}
y_{t-n} = the normalized ring-width index at time $t-n$
e_t = error component.

Because of the transformations performed on the climatic data, i.e., the derivation of the amplitudes of the eigenvectors, the climatic variables are orthogonal and fulfill one major assumption—that of independence of the "independent" variables. Additionally, use of the principal components reduces the number of variables, thereby reducing the dimension of the problem. The use of these transformations, however, somewhat obscures the physical relationship of the effects of climate upon ring width. *Fritts et al.* [1970] suggest that a solution to this undesirable effect is in the "response function," which transforms the principal components back to the original variables. If the components are expressed in terms of the original variables, $x_1, x_2, \ldots x_n$, Eq. (1) is transformed to a linear equation in x. Each additional component changes the coefficients attached to the several x_i terms, these changes being proportional to the elements of the eigenvector (corresponding to the amplitude) newly added. Thus,

$$y_t = \theta_0 + \theta_1(a_{11}x_1 + a_{21}x_2 + \ldots) \qquad \text{for } \xi_1 \tag{2}$$

or

$$\begin{aligned} y_t &= \theta_0 + \theta_1 a_{11}x_1 + \theta_1 a_{12}x_2 + \ldots + \theta_2 a_{21}x_1 + \theta_2 a_{22}x_2 + \ldots \\ &= \theta_0 + x_1(\theta_1 a_{11} + \theta_2 a_{21}) + x_2(\theta_1 a_{12} + \theta_2 a_{22}) + \ldots \qquad \text{for } \xi_1 \text{ and } \xi_2, \end{aligned} \tag{3}$$

where the a's are elements of the respective eigenvectors and the x's are observed values of the climatic variables. Thus, if the variable x_i is factored out of any term, the resultant term is the sum of the regression coefficient times the eigenvector elements. Since these regression coefficients and eigenvector elements are determined in an unbiased manner from the observed values of the variables, the result should be a way to compare the response of the dependent variable y against the respective independent variables x. By plotting these sums of regression coefficients times eigenvector elements for the same independent variables but different dependent variables, one can compare "response functions" for various dependent variables.

Figure 2 shows the response functions to regional temperature and precipitation for (1) tree growth at a site within Upper San Francisco River basin, and (2) total annual runoff at Glenwood, New Mexico. In both cases, temperatures are based on monthly averages and precipitation on monthly totals.

The response function for tree growth shows that above-average growth results when precipitation is above normal in November, December, and February-July and below normal in August, coupled with below-normal temperatures in November-February, April-July, and September, and above-normal temperatures in March and August.

Above-normal annual runoff occurs when precipitation is above normal especially in November, January, February, April, May, and July-September, coupled with temperatures below normal in November, January, March, July, and September and above normal in December, April, and May.

The similarities between the response functions for tree growth and those for runoff represent climatic signals present in both series; the disparities represent the part of the signals lost as noise. One noticeable difference in the responses to precipitation is the consistently positive response of runoff, especially in November, January, April, and July, whereas the effect of precipitation on tree growth is less pronounced, noticeably in August. The responses for average monthly temperatures show major disparities in December and April-May (below normal for maximum growth, above normal for maximum runoff), and in March and August (above normal for maximum growth, below normal or normal, respectively, for maximum runoff). From the above, it is not hard to imagine conditions under which high runoff would occur but maximum growth would not occur. For example, high precipitation in November and January with high temperatures in December would lead to high runoff but would not contribute as markedly to tree growth.

RECONSTRUCTION OF RUNOFF SERIES FROM TREE-RING INDICES

With the above limitations in mind, it is possible to develop an equation for reconstructing a pattern of past annual runoff from tree-ring indices.

If the tree-ring data are sampled at widely dispersed sites over a moderately large watershed, say 2000 square miles, a means is needed to incorporate into the model the spatial distribution of the

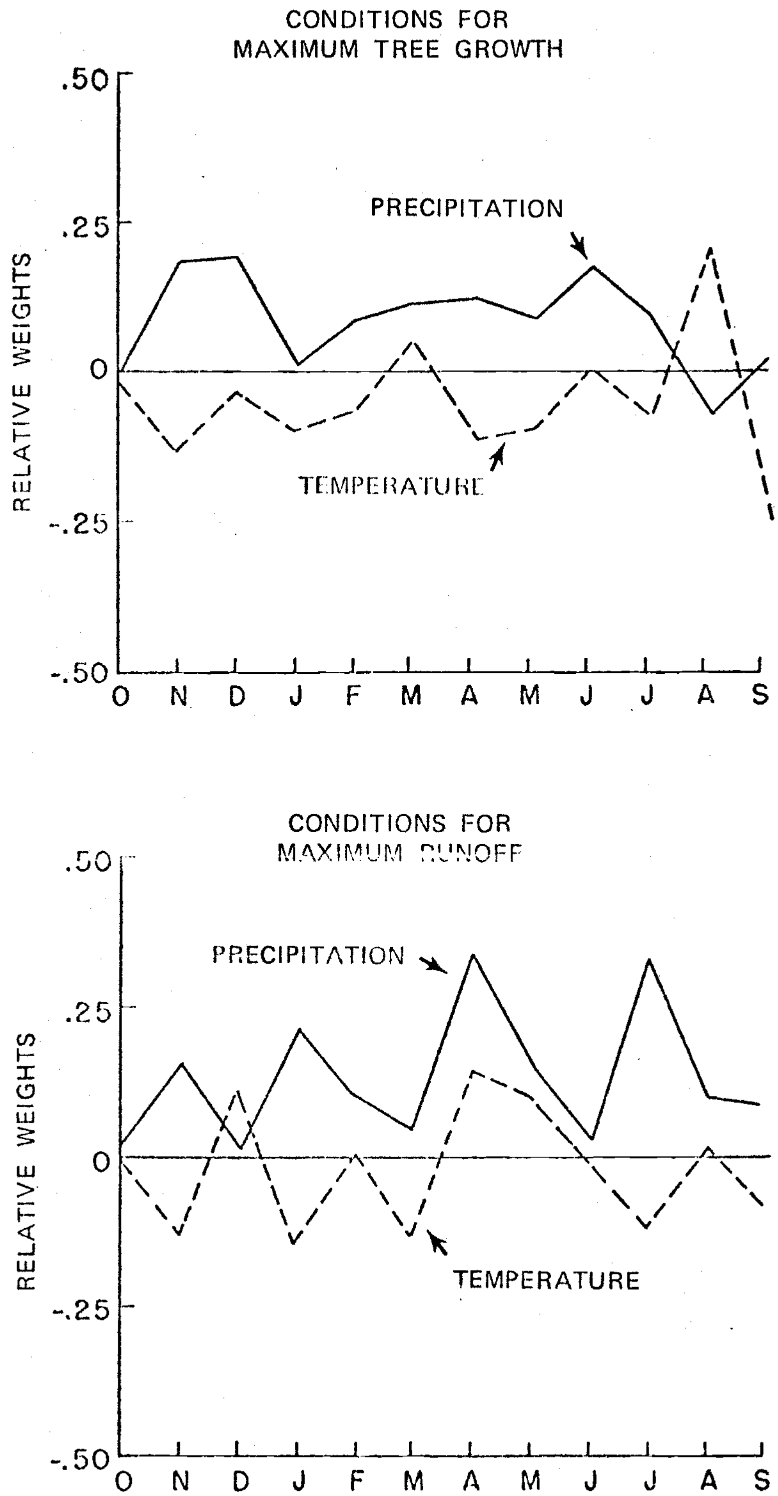

Fig. 2. Comparison of response functions of monthly climatic variables (water year) for tree growth and annual runoff in Upper San Francisco River basin.

tree-ring data at any time t. In addition, the persistence with time may vary among tree-ring sites, and a way is needed to compensate for this difference in persistence and at the same time model the generating mechanism within these data so that it can be compensated for in the reconstruction equation. By lagging the matrix of tree-ring series in time up to $t-3$ and extracting the eigenvectors from this combined matrix, a space-time distribution of the tree-ring series is accomplished. Thus, a least squares reconstruction equation is obtained:

$$f_t = \beta_0 + \beta_1 \xi_{1t} + \beta_2 \xi_{2t} + \ldots + f_{t-1} + f_{t-2} + f_{t-3} + e_t, \qquad (4)$$

where

f_t is runoff at time t,

β's are least square regression coefficients,

ξ_t's are the amplitudes of the eigenvectors extracted from the correlation matrix of the combined, lagged matrix,

f_{t-n} are previous-year runoff values, and

e_t is the error resulting from inadequacy of the model itself.

It was found, subsequent to the work of *Stockton* [1971], that by lagging the values of runoff with respect to the tree-ring series, one can compensate for the generating mechanism differences in the two series. That is, the runoff at time t is a function of the tree-ring data x at times $x_{t+1}, x_t, x_{t-1}, x_{t-2}$. This provides an expression of the mixed moving average-autoregressive model established by *Stockton* [1971] as typical for Douglas fir series in the Upper San Francisco River basin. In this case, prior runoff was not included as an independent variable.

In using a reconstruction equation like Eq. (4), five basic assumptions are made:

1. The climatic interaction between runoff and tree growth is constant and does not change with time.
2. A linear relationship exists between the tree-ring series and the annual runoff series.
3. The variables are multivariate normal distributed.
4. The residuals are independent (i.e., the cross product of the residuals is zero).
5. Expected value of residuals, e_t, is zero.

APPLICATION TO TWO WATERSHEDS

Tree-ring samples of a single species, Douglas fir, were taken in two watersheds of diverse hydrologic character, one in Arizona and one in New Mexico.

The first, Bright Angel Creek watershed, is an area of 100 sq. mi. on the north rim of the Grand Canyon in north-central Arizona. The annual precipitation regime (mean of 25 in.) shows two maxima, one in July-September and the other in December-January. The winter maximum, however, is dominant and results in an average annual snow accumulation of approximately 150 in. The runoff pattern reflects this tremendous snow accumulation, in that 97% of the annual runoff occurs during April and May as the result of melting snow.

The Douglas fir in Bright Angel Creek basin are characteristic of a forest interior site and are less sensitive to climate than would be those from either a lower or upper forest border site. This means that the ring widths yield less climatic information than would be desirable. This deficiency was known when this watershed was chosen, but it was chosen anyway in order to contrast results obtained under less than desirable conditions with those of conditions closer to ideal.

An equation of the type of Eq. (4) was developed to reconstruct the record of past annual runoff from the tree-ring indices. The criterion for including or excluding any given variable (ξ's and f's) was that its *F*-ratio must equal or exceed 4.0. This gave an equation that accounted for 51% of the variance in the actual record. Using the equation, the record was reconstructed for the period 1753-1966 (214 years) as shown in Figure 3. (Superimposed on the graph is the actual observed record for the period 1924-1966.) The low-frequency Noah and Joseph effects are quite noticeable in the reconstructed series. These results, although not highly useful for reconstructing the past record on a year-to-year basis, do provide an improved estimate of the mean according to the criterion of *Fiering* [1963]. That is, the long-term mean from the reconstructed series (5.81 in.) versus the mean from the observed series (4.73 in.) is considered to be closer to the true population mean. If a noise element were added to each estimate, as discussed by *Matalas and Jacobs* [1964], this reconstructed series would also yield a better estimate of the variance.

The second watershed was the Upper San Francisco River basin in west-central New Mexico. The hydrologic characteristics are quite different from those of Bright Angel Creek basin. As at Bright Angel Creek, the annual precipitation (mean of 15 in.) shows two maxima, one in July-August and the other in December-January. Here, however, the July-August maximum is dominant, but owing to

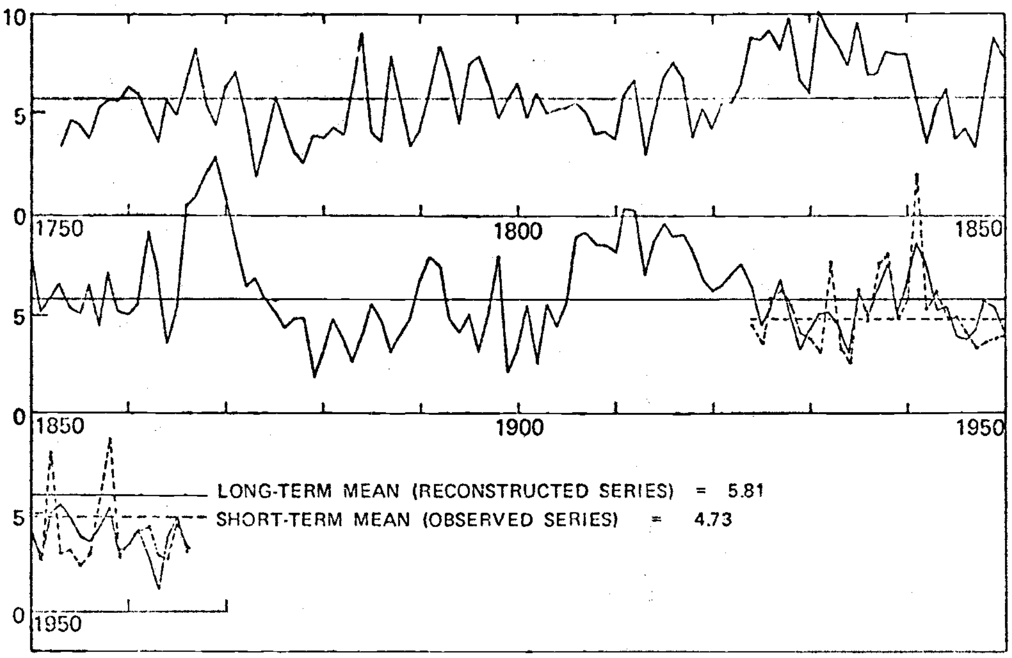

Fig. 3. Reconstructed hydrograph for Bright Angel Creek, 1753-1966 (214 years). Runoff data for this period have been predicted from tree-ring data for the same period. Observed runoff data for 1924-1966 are superimposed with dashed line.

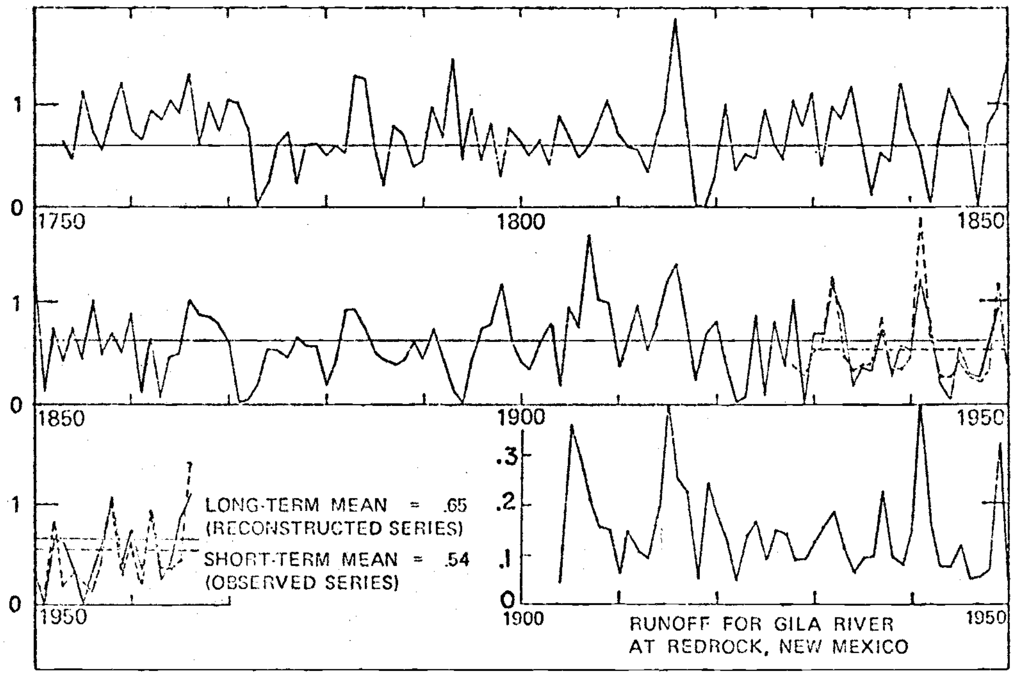

Fig. 4. Reconstructed hydrograph for Upper San Francisco River, 1753-1966 (214 years). Runoff data for this period have been predicted from tree-ring data for the same period. Observed runoff data for 1928-1966 are superimposed with dashed line. For comparison, the runoff record for the Gila River near Redrock (1904-1950) is also shown.

evaporation losses, most of the runoff, about 88%, occurs during the winter, November-May. Very little results from melting of accumulated snow.

In this basin, the tree-ring data are from a lower forest border of Douglas fir and thus they are far more sensitive to climate than at Bright Angel Creek. Consequently, a better correlation was expected between the ring-width series and the runoff series in this basin.

The reconstruction equation of the form of Eq. (4) for this basin accounted for 72% of the variance in the observed runoff record. (Again the criterion for inclusion of variables was $F \geqslant 4.0$.) The reconstructed hydrograph for the period 1753-1966 is shown in Figure 4 along with the observed record for the period 1928-1966. In this case the reconstructed record conforms with the observed record much better than in the first case. As in the first case, the long-term mean of the reconstructed series is higher than that of the observed series, 0.65 in. versus 0.54 in., which represents about 58,000 acre-feet per year versus about 47,000 acre-feet. Again, according to the criterion of *Fiering* [1963], this long-term record represents a better estimate of the true population mean. An improved estimate of the variance can also be gained from the reconstructed series.

Also shown in Figure 4 is the observed record for a nearby station, on the Gila River, for the period 1904-1950, which allows visual comparison of the reconstructed record against one that was actually observed in the same region. Comparison of the Gila record at Redrock with the reconstructed series for the Upper San Francisco River illustrates one of the precautions that must be taken in using the tree-ring technique. As was pointed out in the section on response functions, certain monthly climatic regimes that result in maximum runoff are not conducive to maximum growth. One such regime occurred in 1904, when December and January were exceptionally wet and probably above average in temperature a condition for maximum runoff but less than maximum growth. The result is that the reconstructed value for 1904 is only about half of what actually occurred.

From Figure 4, one sees that the Noah and Joseph effects, although not as pronounced as in the case of Bright Angel Creek, are nonetheless quite evident in the reconstructed record. As shown by *Stockton* [1971], the long-term, low-frequency component (Joseph effect) results in a substantially different correlogram than does that of the observed record. Thus, the long-term reconstructed series should provide an improved estimate of the correlogram of the annual runoff series because the Joseph effect is included.

CONCLUSIONS

It has been shown that tree-ring data can be used to augment annual runoff records. Although the two examples cited differed substantially in the degree of conformance of the actual versus the reconstructed records, the conformance in both cases was still close enough that improved estimates of the mean and variance could be obtained. In interpreting runoff records reconstructed from tree-ring data, it must be borne in mind that there are certain monthly climatic regimes that result in high runoff but may not be as favorable to growth. An example of one such occurrence was illustrated. Fortunately, such occurrences are rare.

Acknowledgments. The research described herein was made possible by a grant from the Department of Interior, Office of Water Resources Research. The Computer Center at the University of Arizona provided some of the computer time used in the study.

REFERENCES CITED

Fiering, M. B., On the use of correlation to augment data, *J. Amer. Stat. Ass., 57,* 20-32, 1962.

Fiering, M. B., Use of correlation to improve estimates of the mean and variance, 9 pp., *U.S. Geol. Surv. Prof. Pap. 434-C,* 1965.

Fritts, H. C., T. J. Blasing, B. P. Hayden, and J. E. Kutzbach, Multivariate techniques for specifying tree-growth and climate relationships and for reconstructing anomalies in paleoclimate, 48 pp., *Final Report Prepared for Laboratory of Environmental Data Research, Environmental Science Services Administration, Contract E-41-70(N),* 1970.

Johnston, J., *Econometric Methods,* 300 pp., McGraw-Hill, New York, 1963.

Julian, P. R., and H. C. Fritts, On the possibility of quantitatively extending climatic records by means of dendrochronological analysis, pp. 76-82, *Proc. First Statistical Meteorol. Conf.,* Hartford, Conn., May 27-29, 1968.

Matalas, N. C., and B. Jacobs, A correlation procedure for augmenting hydrologic data, 7 pp., *U.S. Geol. Surv. Prof. Pap. 434-E,* 1964.

Mandelbrot, B. B., and J. R. Wallis, Noah, Joseph and operational hydrology, *Water Resour. Res., 4*(5), 909-918, 1968.

Rodríguez-Iturbe, Ignacio, Estimation of statistical parameters for annual river flows, *Water Resour. Res., 5*(6), 1418-1426, 1969.

Stockton, C. W., The feasibility of augmenting hydrologic records using tree-ring data, 172 pp., *Ph.D. Dissertation,* The University of Arizona, Tucson, 1971.

Stokes, M. A., and T. L. Smiley, *An Introduction to Tree-Ring Dating,* 73 pp., The University of Chicago Press, Chicago, 73 pp., 1968.