VARIABILITY FACTORS IN MOUNTAIN-WATERSHED HYDROMETEOROLOGY IN AN ARID REGION[1]

JAMES E. McDONALD
INSTITUTE OF ATMOSPHERIC PHYSICS, UNIVERSITY OF ARIZONA

INTRODUCTION. — The outstanding climatological characteristic of arid regions — precipitation characterized by low mean values and high relative variability — puts a distinctive stamp on the hydrology of such regions. Potential evapotranspiration exceeds precipitation (essentially by definition of an arid region) with the result that streamflow is ephemeral over most of such areas. For example, Schwalen (1942) notes that the Santa Cruz River, in its intermittent flow past Tucson, had an average annual discharge of about 13,000 acre-feet in the 19-year period 1923-41, whereas estimated annual basinwide precipitation amounted to about 1.9 million acre-feet for this period. Since net groundwater recharge in this period was believed to be essentially zero, these figures imply that the actual evapotranspiration over the generally low-lying Santa Cruz basin averaged 99.4 per cent of precipitation for the period, making the ephemeral character of the flow quite understandable. Only in the mountainous parts of an arid region does precipitation become large enough and potential evapotranspiration low enough that at least seasonally continuous streamflow may develop. Thus, throughout the world, inhabitants of arid regions come to depend upon these mountain-created "humid islands" for much of their water supply. The Tibesti, the Atlas, and the Ahaggar Mountains of North Africa, the Elburz of Iran, the Pamirs of southwest Asia, the Andes of Patagonian Argentina, and numerous other humid islands in or bordering upon the world's arid regions thus have much in common with the mountain watersheds of Arizona's Mogollon Rim area that will be the subject of the following discussion. In every one of these cases, relatively low precipitation amounts and high temporal variability aggravate the problems of the irrigation agriculture that typically evolves in areas dependent upon such water supplies. Hence there is need for greatly increased understanding of all those climatic, hydrometeorological, and ecologic factors that introduce variability into mountain-watershed hydrology in arid regions. Information gained in any one of those regions is certain to shed useful light on at least some of the hydrologic problems of the other regions.

In central Arizona, irrigation agriculture is now dependent for about one-third of its total water supply on the runoff of the Gila River basin (including the Salt and its tributaries). The fact that the remaining two-thirds (roughly 4×10^6 acre-feet) of Arizona's irrigation water now comes from deep wells whose water levels are falling at rates of the order of 10 ft/yr has focused increasing attention (e.g., Barr *et al.*, 1956) on prospects for deriving greater amounts of irrigation water from the mountain watersheds of the Mogollon Rim. However, present knowledge of *quantitative* relationships fundamental to intelligent manipulation of watershed runoff seems quite limited when measured against the great complexities of the phenomena involved. One of the many areas of our present ignorance is that of the "hydrometeorological climatology" of arid-region watersheds. In the following remarks I wish first to set out, for the use of other watershed investigators, a number of purely descriptive data which display some important hydrometeorological characteristics of several Arizona watersheds, and secondly to call attention to several simple statistical arguments bearing on important problems associated with watershed modification.

DEFINITIONS AND DATA. — Throughout the following, "winter" will mean the six-month period November through April, "summer," the remaining six months. Specifically, "winter, 1913" will mean the period from November 1, 1912 to April 30, 1913. For the purposes of Arizona precipitation climatology those season-definitions are preferred despite the fact that "winter" then begins one month after the generally accepted beginning of the "water year" in October.

When the total runoff derived from a given watershed (or portion of a watershed) in a specified time interval is divided by the total watershed area in question, a quantity termed the "unit yield" is obtained. It may be measured in inches, and represents the depth to which water would stand over the watershed if all of the runoff could be spread in uniform thickness over the watershed.

Whereas the term, "unit yield," enjoys some currency in hydrology, another equally fundamental and useful measure appears to be unnamed: the ratio of the unit yield of a watershed to the mean precipitation over the watershed for any given time interval. This ratio will be here called, for sake of brevity, the "runoff ratio." It will be expressed in per cent, and represents the quotient of actual runoff measured at a specified gauging point divided by the total precipitation falling over the watershed above that point in the given time interval. For example, in the cited case of the Santa Cruz basin above Tucson one

[1]Work done under the University of Arizona Arid Lands Research Program, supported by the Rockefeller Foundation.

would say that the mean annual runoff ratio is 0.6 per cent. For the entire United States, the annual runoff ratio (9 inches mean unit yield, 29 inches mean precipitation (Benton and Estoque, 1954)) is about 31 per cent, by comparison.

The frequently occurring unit of water volume, the acre-foot, will here be abbreviated "af," and a thousand acre-feet will be written "1 taf."

Runoff data will be used for four different streams, the primary source being various U. S. Geological Survey Water Supply Papers (chiefly No. 1313, Pt. 9). That designated here as "Verde" represents data obtained on the Verde River below Bartlett Dam prior to 1938, 16 mi above Bartlett from 1939 to 1945, and above Horseshoe Dam since 1946. The composite Verde record does not differ significantly from the storage-adjusted data published for "below Bartlett Dam" so long as *seasonal* totals are employed, as is the case here. The runoff data designated as "Tonto" are for Tonto Creek near Roosevelt before 1941, and for Tonto Creek above Gun Creek since 1941. That designated as "Salt" is for the Salt River above Roosevelt Dam for the full period employed. That designated as "Gila" is for the gauging point on the Gila River at the head of Safford Valley near Solomon for the full period employed. The approximate drainage areas for the above four gauging points are: Verde about 6200 mi², Tonto 700 mi², Salt 4300 mi², and Gila 8000 mi², subject to slight variations associated with shifts of gauging sites.

All precipitation data are taken, as published, from U. S. Weather Bureau records, chiefly Bulletin W and the Supplement thereto. In certain cases, station records containing gaps are employed. The estimation of missing data was done using a method described elsewhere (McDonald, 1957).

In several tables, figures, and analyses, the composite precipitation records of six stations in the Mogollon Rim area will be used. It will be designated as "the 6-station mean" and represents a

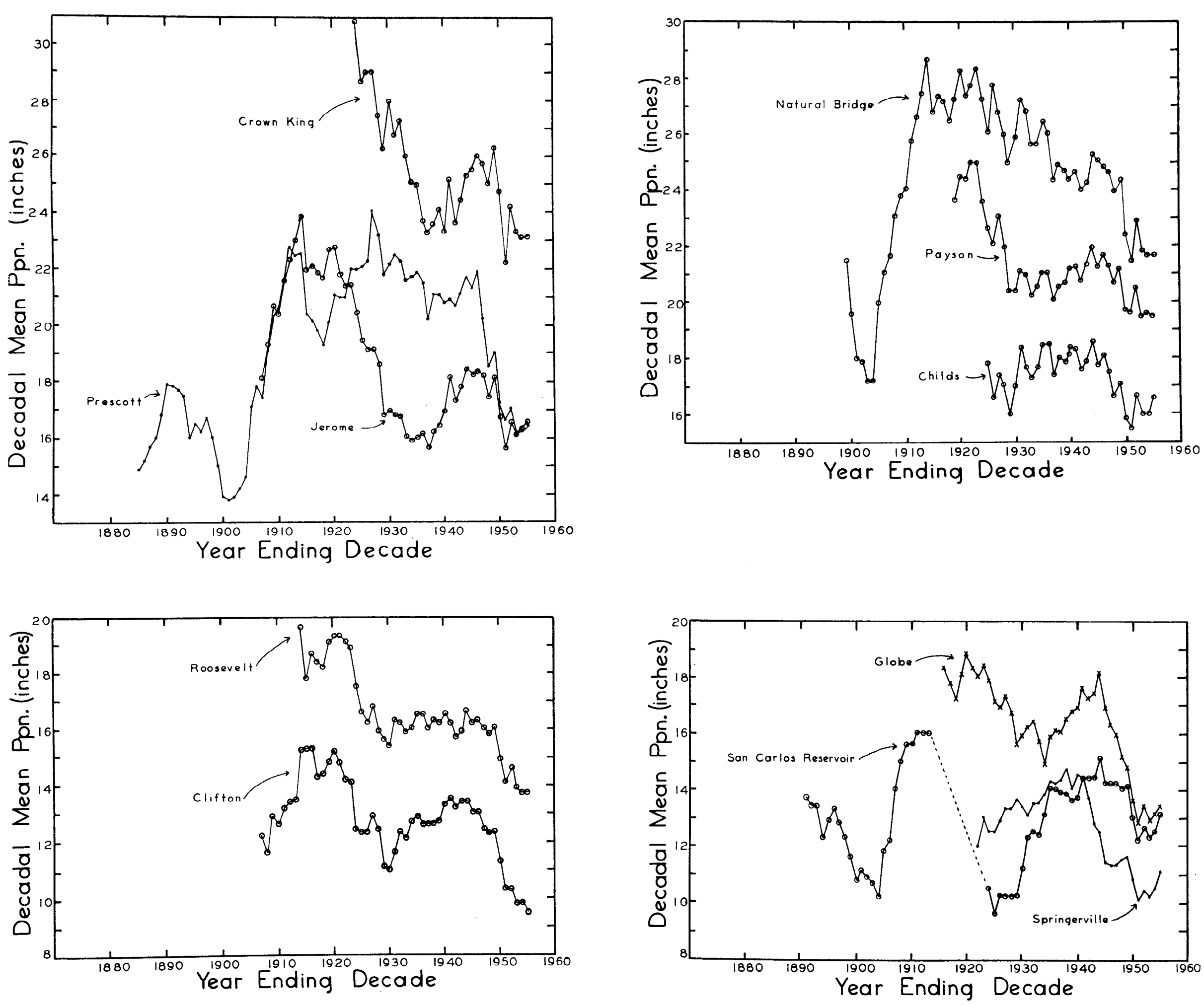

Figure 1.—10-year moving average annual precipitation amounts.

simple unweighted average of the precipitation reported at Flagstaff, Jerome, Walnut Grove, Prescott, Natural Bridge, and Whiteriver-Ft. Apache (the latter a composite of two nearby stations' records). As is well known to all who have tried to study the hydrology of this area, there is a serious dearth of long-record stations in the area of Mogollon Rim.

VARIABILITY OF PRECIPITATION AND RUNOFF. — The present section will summarize a number of data on temporal and spatial variability of precipitation and runoff.

Temporal variability. — In Fig. 1 are shown 10-yr moving average annual precipitation amounts for a number of stations in or near the Mogollon Rim area. The decadal averages are plotted above the year *ending* the successive 10-yr periods. Note carefully that, to save space, the ordinate scales are not set at zero. The process of taking moving averages over 10 years conceals the erratic year-to-year variability that is so important a feature of arid-region precipitation, but has the advantage that it permits the eye to discern more easily those longer-term trends that play so important a role in agriculture and other activities. In particular, Fig. 1 reveals the general downward trend in precipitation that has been characteristic of the Southwest during roughly the past two decades and that has stimulated increasing interest in the many hydrologic problems of the region.

Fig. 2 shows, for comparison, a plot of decadal moving mean annual runoff amounts for the combined Salt, Tonto, and Verde watersheds. Clearly, runoff has declined markedly during the predominantly dry years of the past two decades. But it has been equally low in the recent past, yet has recovered, so Fig. 2 has implications both pessimistic and optimistic. Figs. 1 and 2 together suggest that at least a major cause of decline of runoff has surely been the decline of precipitation itself. The crucial question of whether runoff has declined at a rate significantly different from that of precipitation is treated in a later section.

Temporal variability data of several types have been presented for a number of individual Arizona precipitation stations elsewhere (McDonald, 1956), but it is of present interest to examine the dispersion of the 6-station mean data that will be employed in several analyses below. Coefficients of variation (standard deviations divided by means) are given, along with the means themselves, for the winter, summer, and annual periods in the first row of Table 1. As is to be expected on simple statistical grounds, averaging six stations' records tends to suppress variability below that typical of *single-station* records. Thus the long-term coefficients of variability for winter and summer precipitation, respectively, are found to be 0.43 and 0.33 at Natural Bridge, 0.37 and 0.29 at Flagstaff, and 0.54 and 0.32 at Prescott, but 0.39 and 0.26 for the 6-station composite record.

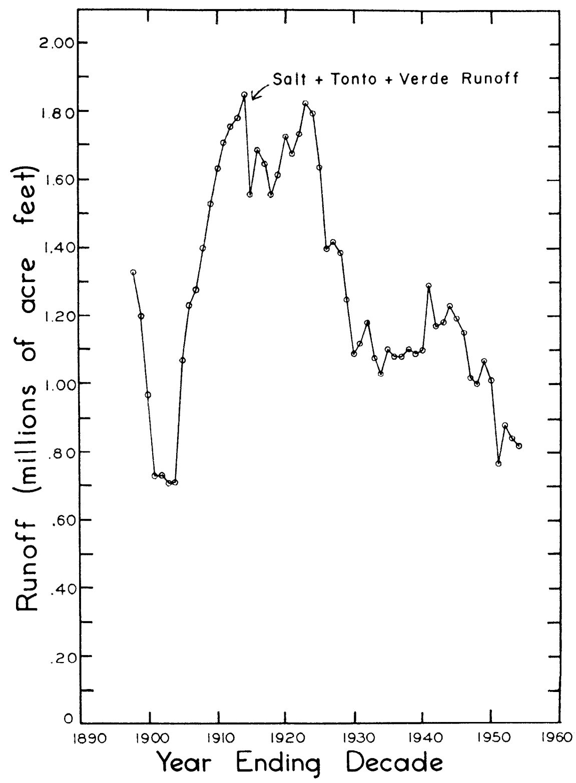

Figure 2.—10-year moving average runoff for combined Salt, Tonto, and Verde watersheds.

In the lower part of Table 1 are displayed coefficients of variation and means for the runoff of the four streams here considered. The periods of record are long enough to make these reasonably representative values. The most striking feature of the runoff variability data is the fact that in this arid region of typically high precipitation variability *coefficients of variation for runoff are distinctly higher*

Table 1. Means and coefficients of variation.

Data	Period of Record	Means			Coefficients of Variation		
		Winter	Summer	Annual	Winter	Summer	Annual
6-station precipitation	1898-1954	10.1 in	9.6 in	19.6 in	0.39	0.26	0.25
Verde runoff	1898-1954	377 taf	111 taf	488 taf	0.79	0.44	0.64
Tonto runoff	1914-1954	82	16	98	0.88	0.83	0.78
Salt runoff	1914-1954	453	198	651	0.92	0.56	0.75
Gila runoff	1914-1954	214	118	332	1.22	0.68	0.95

than for precipitation. One might have anticipated *lower* relative variability for runoff from an entire watershed than for precipitation means for just six raingauges scattered over the areas. That the actual situation is the opposite reflects the counteracting influence of the large amounts of variance introduced into runoff time-series by numerous complex hydrologic factors operating once the precipitation is on the ground. Another way of putting this important point is that runoff represents the rather *small difference* between precipitation and evaportranspirational losses. Since both of the latter terms are temporally variable and to at least some extent independently variable, the time-series of their differences must be expected to show even greater *relative* variability.

Table 1 shows that another significant feature of coefficients of variation of runoff in the Mogollon Rim watersheds is the way in which the greater relative variability of winter as compared with summer *precipitation* (cf. McDonald, 1956) carries over into seasonal differences in *streamflow* variability. In some respects this feature seems even more contrary to expectation than the point noted in the preceding paragraph. Since it is well known that thunderstorm rainfall of the type dominating the summer precipitation in Arizona produces short-lived flash-floods distributed in spatially erratic fashion, one might have anticipated that summer runoff would be *more* rather than less variable than winter runoff, and all the more so because summer runoff averages generally half or less the value of the winter runoff, as shown by the means of Table 1. Perhaps this tends to occur *because* of rather than in spite of the flash-flood character of the summer flow. Perhaps the very rapidity with which intensely delivered thunderstorm precipitation moves into and down the watercourses to gauging points precludes the randomizing operation of many of those hydrologic processes influencing variability of yields from winter precipitation.

On simple statistical grounds, one would expect that coefficients of variation of runoff should tend to rise as one goes to successively smaller watersheds or parts of watersheds. However, Table 1 does not support this prediction very well; again presumably because of effects of numerous hydrologic variables other than mere integrating area itself. The same hydrologic complexity may be seen operating in interesting fashion on a smaller scale by examining two streamflow records for stations near Tucson. The 42-year record ending in 1950 for annual flows gauged near the mouth of Rillito Creek near Tucson (916 mi^2 drainage area) yields an annual mean of 13.4 taf and a coefficient of variation of 1.41. But the 25-year record, ending in 1950, for annual flows of Sabino Creek at the point where the stream debouches from the Santa Catalina Mountains exhibits a mean of 8.5 taf and a coefficient of variation of 0.99, noticeably *smaller* than that for Rillito Creek despite the facts that (1) Sabino Creek drains only an area of 35 mi^2 vs. Rillito's 916 mi^2, and (2) Sabino Creek is actually *tributary* to the Rillito! Clearly statistical intuition based on the idea of watersheds as integrating-areas is entirely misleading in hydrologic problems, particularly in an arid region. As a single example of the type of hydrologic processes capable of nullifying such simple integrating-area inferences it is of interest to note Schwalen and Shaw, 1957) that the entire Sabino spring flow of 10 taf in 1952 was completely consumed by percolation into the Rillito streambed sands along a course of only half a dozen miles just above the Rillito gauging point.

The variability data of Table 1 are perhaps of greatest interest in that they reveal the important point that attempts to influence runoff through watershed modification programs will encounter major difficulties of disentangling treatment effects from *natural* runoff variability. This difficulty is exactly analogous to that which has for years blocked progress in evaluating treatment effects in the field of cloud modification where the effects of seeding clouds cannot easily be separated from the background of high variability of natural precipitation. In fact, Table 1 suggests that *in arid regions it will be considerably more difficult to assess any economic benefits derivable from watershed modification efforts than to assesss the economic value of cloud modification efforts because the relative variability of runoff data is even greater than that of precipitation.* This

Table 2. Seasonal precipitation correlations for Natural Bridge.

Station	Years of Record	Winter	Summer
Flagstaff	50	0.74	0.75
Walnut Grove	58	0.92	0.61
Jerome	58	0.86	0.58
Prescott	65	0.87	0.60
Whiteriver	65	0.84	0.53
Roosevelt	50	0.89	0.70
Globe	50	0.87	0.66
San Carlos	50	0.81	0.64
Clifton	50	0.65	0.49
Holbrook	50	0.72	0.57
Ft. Bayard, N. M.	50	0.61	0.29

Table 3. Interbasin runoff correlations.

		r		
Basin pair	Period of Record	Winter	Summer	Annual
Verde - Tonto	1914-1954	0.90	0.53	0.91
Verde - Salt	1902-1954	0.85	0.63	0.86
Verde - Gila	1915-1954	0.56	0.64	0.79
Tonto - Salt	1914-1954	0.84	0.47	0.85
Salt - Gila	1915-1954	0.82	0.80	0.86

important point will be discussed further below.

Spatial variability. — In general, spatial variability is of less importance than temporal variability in arid region hydrometeorology. Nevertheless there do arise questions of spatial homogeneity of watershed characteristics in a variety of research-planning efforts (e.g., one often wishes to know whether he can safely lump together two watersheds or parts of a watershed in a given study); so a few such data will be presented here.

As a rough measure of spatial homogeneity of precipitation amounts, correlation coefficients r for winter and summer precipitation totals for a number of long-record stations in or near the Mogollon Rim area are shown in Table 2 for a single base-station, Natural Bridge, located in the Rim area near the center of the state. Double-mass analysis reveals (McDonald, 1956) that the record at Natural Bridge is itself temporally homogeneous. As was noted earlier, r is seen in Table 2 to run systematically higher for winter than for summer precipitation in Arizona, despite winter's greater degree of temporal variability. The greater spatial homogeneity of winter precipitation results from the fact that widespread cyclonic and frontal systems govern winter precipitation whereas isolated orographic and instability thunderstorms account for most summer precipitation.

The fact that the Natural Bridge winter r-values stay near 0.8 over so much of the Rim area of present interest implies that runoff, chiefly dependent upon winter precipitation, ought to be tolerably homogeneous over the same area. Even for Ft. Bayard, N. M., about 200 miles east-southeast of Natural Bridge at the extremity of the Rim ranges, r falls only to about 0.6 in winter. Furthermore, the fact that r exceeds 0.7 for such stations outside of the Verde and Salt watersheds as Walnut Grove, Flagstaff, and Holbrook suggests that records at such peripheral sttaions will be of some prediction value in precipitation-runoff studies for the Verde or Salt basins. This is relevant information inasmuch as there are so few longrecord precipitation stations on the watersheds proper that it is always tempting to use such nearby stations' records in statistical analyses.

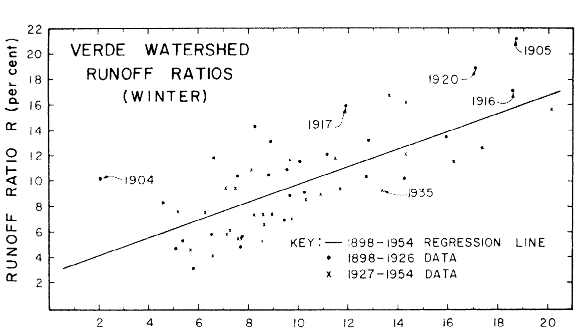

Figure 3.—Winter runoff ratios *vs.* precipitation for Verde basin.

Turning from precipitation homogeneity to runoff homogeneity, we may consider interbasin runoff correlations. Table 3 shows such figures for various watershed-pairs. Similar to, and basically because of, the greater winter than summer spatial homogeneity, runoff is spatially more homogeneous in winter than in summer. However, whereas winter precipitation correlations in Arizona typically exceed corresponding annual precipitation correlations, Table 3 shows the somewhat surprising fact that runoff in every watershed-pair is slightly better correlated on an annual basis than for winter alone. It is not clear to the writer why summer runoff should exert the type of compensating action implied in this result. Since an analogous process does not operate on precipitation itself, the compensatory action is presumably hydrologic and not meteorological in nature.

The fact that r for the summer runoffs of the Salt and Tonto is 0.47 whereas r for the Salt and Gila is 0.80 is interesting. The surprising point is not so much the low value for the Salt-Tonto correlation as it is the quite high value for the Salt-Gila correlation. An adequate explanation of this curious relationship could only be given after a detailed study of the contribution of various tributaries of the Salt and the Gila during the summer thunderstorm season. This has not been done in the course of the work reported here, so the problem can only be cited as one more warranting further analysis.

PRECIPITATION AND RUNOFF ON THE VERDE WATERSHED. — Chronic water deficiencies in agricultural areas of the lowlands bordering any moutainous "humid island" in an arid region must sooner or later direct attention to the possibility of man's modifying the nearby mountain watersheds in some way aimed at increasing the fraction of precipitated water available for lowland agriculture. Arizona has reached that stage within recent years. It is certainly to be hoped that knowledge gained in current and future Arizona studies can later be used in those now less well-developed arid regions where the same turning point has yet to be reached.

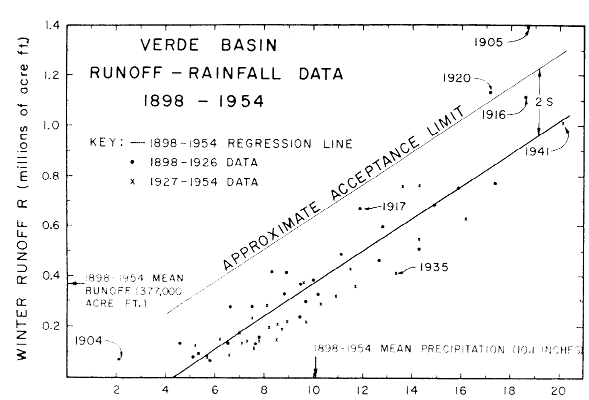

Figure 4.—Winter runoff *vs.* precipitation for Verde basin.

The present section attempts to derive useful insights into watershed modification prospects by simple statistical methods applied to such scanty hydrometeorological data as now exist.

Two such efforts have been made previously. Cooperrider and Sykes (1938) studied the relation of the flow of the Salt River above Roosevelt Dam to certain precipitation records for a 35-year period ending in 1936, directing much of their discussion towards what one infers must have been contemporary debate over plans to modify watershed vegetal cover in the hope of deriving a larger runoff ratio. Many of the findings of Cooperrider and Sykes and many of their general discussions remain today very relevant to an understanding of modification problems. More recently, Anderson (Barr *et al.*, 1956) examined precipitation-runoff relationships for the Salt basin, obtaining many useful descriptive statistics based on longer precipitation and streamflow records. Neither Cooperrider and Sykes nor Anderson employed statistical methods yielding quantitative estimates of difficulties inherent in treatment evaluation problems associated with watershed modification. Also, in each of these earlier studies, *annual* precipitation and streamflow data were employed in those parts of the analyses aimed at discerning downward trends in runoff ratio, which is less than optimal procedure, for reasons to be given below. Most of the precipitation stations Anderson employed lay in the Verde basin whereas his streamflow data, in his more important analyses, were for the combined Verde, Tonto, and Salt Rivers, which creates certain difficulties of interpretation of his results. And Cooperrider and Sykes' precipitation stations included a number of stations at rather too low elevations to be the best of predictors of runoff yield. In all, the problem seems far from exhaustively analyzed as far as past studies are concerned.

Here only the runoff for the Verde River above Bartlett Dam will be used, for the reason that the long-term precipitation records available to the writer were more complete for the Verde basin than for any of the other basins of interest. These data, the source and nature of which have been discussed earlier here, include precipitation records for three stations within the Verde basin (Jerome, Prescott, and Natural Bridge), two just outside it (Flagstaff and Walnut Grove), and one in the upper Salt basin (Whiteriver-Ft. Apache). Table 2 shows that for the winter season of chief runoff importance, Natural Bridge in the upper Verde basin, is correlated with the other five stations at a level above 0.8 for all but Flagstaff, for which r in winter is 0.74. It is felt on rather general climatological grounds that inclusion of the three non-basin records yields improved overall accuracy of specification of mean watershed precipitation.

Precipitation-runoff correlations.—The unweighted 6-station mean precipitation values were linearly correlated against Verde runoff for the 57-year period 1898-1954. The annual totals gave a correlation coefficient of 0.77, summer values gave 0.61, and winter gave 0.89. Splitting the record into two nearly equal parts (29 and 28 years, respectively) gave winter correlations of 0.88 for 1898-1926, and 0.90 for 1927-54, the corresponding summer values being 0.63 and 0.66. Thus there was no significant shift in r between the early and late halves of the 57-year period, since the standard errors of even the half-period r-values are about 0.04 in winter, and 0.10 in summer.

As stressed by Cooperrider and Sykes and by Anderson, and as is readily seen from Table 1 above, winter precipitation accounts for only a bit over half the annual total, yet winter runoff accounts for over three-fourths of the annual flow. It is thus understandable why the correlation of 57 years of *annual* Verde runoff against *winter* precipitation for the 6-station data gave a correlation coefficient of 0.88, appreciably greater than annual vs. annual, and almost as high as winter runoff against winter precipitation.

Although precipitation-runoff correlations as high as the preceding (and even higher for more elaborate statistical prediction schemes) are fairly common, no one appears to have stressed the point that such high degrees of concordance between precipitation and runoff are really quite remarkable in view of the fact that, especially in the arid Southwest, runoff is a rather small residual, the difference between widely varying precipitation and evapotranspiration. That values of r near 0.9 do characterize precipitation-runoff correlations must be construed as revealing a very strong tendency for evapotranspiration processes to operate with intensity controlled much more by sheer amounts of available watershed precipitation than by any and all other hydrometeorological or ecologic factors. This influence is obviously very relevant in any considerations of plans to enhance runoff through watershed modification efforts.

Evidence for a decline in mean runoff ratio.— Because there have been a number of fairly well-documented changes in vegetal cover on the Mogollon Rim watersheds during the present century (cf. Barr *et al.*, 1956), there has arisen both scientific and popular interest in the question of whether those changes have produced changes in the *runoff ratio* itself. Cooperrider and Sykes found, in their analysis of runoff of the Salt above Roosevelt Dam, that from 1905 to 1936 the runoff ratio declined from about 19 per cent to about 16 per cent as computed for the particular precipitation records they used. Anderson, using different streamflow data (Salt plus Tonto plus Verde) and different precipitation data, found that the mean runoff ratio for the period 1914-30 was 11.7 per cent, while that for the period 1931-53 was 9.4 per cent; but, doubting the statistical significance of this difference, he suggested that it might have resulted simply from the non-linear dependence of runoff on rainfall for the low rainfall amounts characteristic of the more recent period.

A difficulty inherent in Anderson's and in Cooperrider and Sykes' analyses was their use of *annual* precipitation data. One of the characteristics of the precipitation trends of the past decade or two in Arizona is the rise in *relative* contribution of summer as compared with winter precipitation (McDonald, 1956). On the other hand, summer precipitation is, as noted above, an inefficient producer of runoff. Hence to look for trend effects in the ratio of annual runoff to annual precipitation in a period marked by a distinct rise in *relative* importance of summer as contrasted with winter Arizona precipitation is a questionable procedure.

A preferable alternative is to look for trends in only the *winter* runoff ratios, i.e., considering only ratio of winter runoff to winter precipitation. Hence Verde runoff ratios have here been computed for each winter in the period 1898-1954. The watershed drainage areas used in calculating unit yields were those associated with the several gauge-locations as reported by the U. S. Geological Survey. A scatter diagram of runoff ratio R is shown in Fig. 3 for the Verde winter data. The 1898-1954 regression relation is plotted as the solid line, and a few points of special interest are dated for the reader's information. It is interesting that the consecutive years 1904 and 1905, which brought to Arizona greatly different precipitation amounts, exhibit rather similar residuals with respect to the regression relation. The 57-year mean runoff ratio for the unweighted precipitation data here employed was 9.8 per cent with a standard deviation of 3.9 percentage units. (The reader must note that in any analyses where unweighted precipitation data are used for only a small number of stations not ideally located relative to the watershed yield pattern, the *absolute* value of the runoff ratio cannot be regarded as very reliable. But *changes* in the runoff ratio are subject to much less objection, and it is chiefly the question of change that we consider here.)

As a first crude test, the number of winters for which the runoff ratio, R, lay in the range 5.0-8.0 per cent was counted separately for the first and second halves of the total period, yielding 8 and 10 cases for earlier and later periods, respectively. Doing the same for the R-interval from 12 to 15 per cent gave 4 cases each for the two sub-periods. This simple tally does not indicate any marked shift in runoff ratio for the Verde watershed during the entire 57-year period.

However, in order to be able to place some confidence limits on the reality of any change in R, the significance of the difference in mean R for the two subperiods was tested. First, because the R's are ratios whose distribution properties might be questioned, a chi-square test of normality was applied to the full set of R values, which yielded the result that the distribution of R's does not depart significantly from normality. Next the mean R and standard deviation of R were computed for each of the two subperiods, and from the latter the standard error of the differences between the two subperiod R's was found. The mean R for 1898-1926 was 10.6 per cent, that for 1927-54 was 8.9 per cent, the standard error of their difference being 1.0 percentage units. Using a two-tailed t-test, it is found that the difference between these two subperiod means is not statistically significant indication of the reality of a change in R between the two subperiods. (However, it could not be said to rule this out as a possibility, in view of the 0.10 significance level so obtained. If the vegetal changes could be regarded as indubitably contributing to a *decline*, a one-tailed test would be appropriate, yielding a 0.05 significance level for the difference in the two subsample means. Whether the magnitude of the observed decline is more than must be expected from the tendency for non-linearity under low precipitation amounts is a point deserving further study.)

Since the runoff ratio is lower over the Verde watershed than over the Salt watershed (cf. Anderson's data on unit yields), it would be of interest to do a comparable test of decline in R for the Salt. The precipitation data used here are, of course, not ideal for such a test; however, they are essentially as applicable as those that Anderson applied to the combined Salt, Tonto, and Verde watersheds. Therefore, as a crude test, runoff ratios were computed using the 6-station mean winter precipitation data employed above along with Anderson's already-computed *annual* unit yields for the combined Salt, Tonto, Verde. Due to (1) the controlling influence of winter precipitation, (2) the dominance of winter runoff in the annual runoff total, and (3) the fairly high precipitation correlations throughout the Rim area, this procedure is not as crude as it might at first seem. In any event, this procedure gave a 40-year mean R estimate of 19.0 per cent for the 1914-53 period, 21.6 per cent for the 1914-33 period, and 16.4 per cent for the 1934-53 period for the combined three watersheds. A two-tailed t-test then implied that the difference of mean value of R between the earlier and later 20-year subperiods was significant at the 0.01 level, a markedly different result from that found for the Verde alone. Because of the potential importance of this somewhat crude indication of a possibly significant decline in runoff ratio for the Salt (presuming that it would have to be vegetation changes on the *Salt* that could cause R for the combined Salt, Tonto, and Verde basins to show a significant decline when that for the Verde alone did not) it is highly desirable to apply more sophisticated trend tests to the Salt watershed and to do so with the best precipitation data that can be secured. Further analysis may show that the latter result is misleading, caused by unjustified application of Verde basin precipitation data to too large an area of the greater Salt basin. Dr. W. D. Sellers of the Institute of Atmospheric Physics is currently undertaking such analyses.

Detection of treatment effects. — In any extensive field program designed to increase the runoff ratio of a watershed, it is indispensable to be able to state, following the modification of the watershed, whether or not the treatment has produced a significant change in runoff. This, as noted earlier, is exactly similar to the problem of evaluating cloud-seeding efforts and, as in the latter context, it is the inherent *natural variability* of the putatively modified phenomenon that poses the most serious obstacle to reaching an acceptable answer. Ironically, it is exactly in those arid regions where both cloud-seeding and watershed modification look most attractive that we find Nature making the task most difficult by virtue of the characteristically high relative variability of arid-region precipitation. And, as noted above in discussing coefficients of variation of precipitation and runoff, it appears that detecting treatment effects will be intrinsically more difficult in the case of arid-region watershed runoff than in the case of cloud precipitation. In just the above remarks one has very simple statistical inferences that must be clearly recognized by any workers planning modification experiments. Since it took over a decade before this simple point (concerning the way natural variability complicates treatment evaluation) was generally appreciated in the field of cloud-seeding, it would not be surprising if it took some time to be appreciated in the field of watershed modification. It is only a very simple point, nonetheless.

To secure some notion of the magnitude that watershed treatment effect would have to exhibit in order to be recognizable with statistical confidence at acceptable levels, the Verde runoff and precipitation data may be examined here in one final way. The equation of linear regression of winter runoff on winter 6-station mean precipitation was calculated. In Fig. 4 the regression line is superimposed on a scatter diagram for the 1898-1954 data. (In general curvilinear regression should be assumed in such cases, and the scattergram of Fig. 4 tends to support this contention. Indeed, *if* the runoff ratio *is* linearly related to precipittaion, runoff itself must vary roughly parabolically with precipitation. But for present purposes, a linear regression of runoff on precipitation will be adequate (see below)). The standard error of estimate S was found to be 131 taf with respect to the regression line. In general, one only accepts as statistically significant positive departures from regression that lie roughly 2 S above the regression line (i.e., roughly 0.025 significance-level on the one-tailed test that might be appropriate on an optimistic basis). Such a line, marking the limit of the region within which one could *not* reject the null hypothesis at the 0.025 level for any given winter following treatment, is plotted at a uniform distance of 2 S above the regression line in Fig. 4. Note carefully that two years out of these 57 *untreated* years gave winter runoffs on the Verde lying *above* this limiting line (a result tolerably close to probabilistic expectations). If one of these two years occurred immediately following completion of treatment, it would undoubtedly be misinterpreted as significant evidence of a treatment effect on runoff; but such a risk is intrinsic in settling for any finite acceptance limits, of course. (In practice, the way one could suppress this latter danger would be through various statistical controls relating runoff ratios on the treated watershed to concurrent runoff ratios on nearby untreated watersheds.)

The question of interest at this point in our analysis is that of the magnitude of runoff increase that might reasonably be expected from a large-scale watershed modification program. As an available answer to that question, we may use the results of an estimate which Croft (Barr *et al.*, pp. 172-174) made for the combined Verde and Salt and Tonto watersheds. Croft's data suggested to him the possibility of a 100 taf annual runoff increase. The details of his assumed watershed treatment cannot be recounted here but it is very pertinent to note Croft's emphasis in stating that the program "would involve maintaining approximately one-half million acres of watershed land in optimum condition for water yields as well as for soil stability. Since under ordinary conditions nature reacts to nullify changes in vegetal conditions which are brought about by artificial treatment, this would be a tremendous task and would require very substantial sums of money and a persistent, well planned and continuing operation." He notes, furthermore, that of the estimated 100 taf increase, "a large amount of this would be lost in transmission to the reservoirs."

Nevertheless, taking the 100 taf annual runoff increase as attainable despite the cited qualifications, we may compare this with results of our Verde winter regression analysis. Table 1 shows that the Verde *winter* runoff averaged 377 taf in the 1898-1954 period. By comparison, the combined Verde, Salt, and Tonto watersheds, from which Croft estimated an *annual* treatment yield of 100 taf, averaged 1237 taf *annually* during the same period. Hence the Verde *winter* runoff represents 377/1237 or 30 per cent of the overall annual total. Pro-rating to the Verde winter runoff its due share of Croft's estimate, we find that an increase of 30 taf per winter might be expected on the basis of Croft's analysis, ignoring the difficulties he stressed.

This estimate of treatment effect falls far short of the one-season acceptance limit (0.025 level) of about 260 taf suggested by the simple regression analysis of Fig. 4. Indeed, 30 taf amounts to just 0.23 S. One must expect that this small a positive departure from regression will, on the basis of the empirical probabilities derived from the historic record for the Verde, be exceeded purely as a result of chance *in about 41 per cent of all years.* An acceptable confidence level for rejection of the null hypothesis will be taken here and below as the 0.025 level. One would require over four successive years of experiment each exhibiting the full Croft winter treatment-estimate on the Verde to give statistically significant (0.025 level, one-tailed test) evidence of a real treatment effect in the fact of the inherently high natural variability of the Verde winter runoff. The above is based on the simplifying assumption that in *every one* of the experiment years, observed runoff is just 30 taf above the expectation value associated, through the regression relation, with concurrently observed precipitation. The simplification is acceptable for present purposes.

If, as Croft predicted, a large part of the treatment effect disappeared as transmission losses en route to the reservoir gauging points, the length of time required to carry out an experimental treatment that could be analyzed statistically for the reality of the treatment effects would be even longer, of course. Furthermore, the comparative coefficients of variation of runoff shown in Table 1 for the several basins suggest that the Verde watershed may well be the Mogollon Rim watershed on which treatment could *most easily* be disentangled from the background "noise" of natural variability. Therefore, we may conclude from the above simple statistical arguments that because of high runoff variability something like five years or more of controlled modification experiments would be required in order to extract a statistically acceptable answer as to whether watershed treatment was or was not yielding statistically significant results. Unlike the otherwise very similar case of cloud-seeding, *randomization* methods cannot here be involved unless we view the experiment as lasting for truly impractical lengths of time (centuries).

Now, the above estimate of required duration of the evaluation experiment was based upon the relation of the estimated treatment effect (30 taf per winter) to the standard error of estimate for the particular regression technique employed (131 taf per winter). It is very important to stress here that one can develop statistical "prediction" techniques rather more sophisticated than those used above for purposes of illustration. If some form of multiple regression method, employing more predictors than the here-employed simple unweighted mean winter precipitation for six Rim stations, can account for an appreciably larger fraction of total Verde winter runoff variance (i.e., if we can obtain a higher effective correlation coefficient between runoff and some larger set of predictors), then treatment evaluation might be carried out in a shorter total number of years.

Statistical studies directed towards exactly this latter objective are currently under way in the Institute of Atmospheric Physics. Here it is only possible to call attention to the order of magnitude of the gain in evaluation efficiency one might hope to attain from such efforts. Experience on other watersheds suggests that multiple regression methods can raise prediction power to a level where perhaps 94 per cent of total runoff variance is explainable (i.e., a multiple correlation coefficient of about 0.97). Assuming, then, that such a technique has been developed for our hypothetically modified Verde basin, how is our evaluation task altered? The standard error of estimate S is lowered from 131 taf, as found here, to about 70 taf. The previously used value of 30 taf per center for the Croft estimate with respect to Verde runoff now amounts to 0.43S instead of only 0.23S as before. The probability of random sampling fluctuations yielding a positive departure from regression equal to or greater than 0.43S is about 0.33. If we imagine a succession of post-treatment years, each exhibiting a 30 taf excess over the value predicted by regression, we find that we will reach the 0.025 confidence level in roughly *three to four years*, slightly shorter duration of required experimentation than demanded for a scheme based simply on Fig. 4. Briefly, statistical refinements could save a year of test period.

If, however, as Croft's quoted remarks suggest, actual excess yields amounted to rather less than his original estimate, the required time rises again, but not by much. Suppose treatment was capable of delivering only *one-half* of Croft's estimated increase, i.e., 15 taf per winter on the Verde. Then assuming again that a multiple regression scheme characterized by 0.97 correlation coefficient has been developed we find, by repetition of the above arguments, that just over four years of observation subsequent to watershed treatment would be required before one could reject the null hypothesis at the 0.025 confidence level — that is, before one could assert that there was only a 2.5 per cent probability that the observed series of excess runoff yields might have risen solely due to "chance fluctuations."

Summary. — The general downward trend in Arizona precipitation of recent decades has led to a corresponding downward trend in runoff, but the *ratio* of runoff to precipitation has fallen off by mean amounts so small, as measured against the natural scale of variability of the runoff ratio, that they are barely statistically significant in the case of the Verde watershed, the basin for which the present study's data are most adequate. Specifically, the mean runoff ratio, R, for the 29-year period 1898-1926, was 10.6 per cent, that for the 28-year period 1927-54 was 8.9, this observed difference for the two subperiods being significant at the 0.05 level on a one-tailed test based on the dubious assumption that vegetal changes rule out possible *rise* in R. On a two-tailed test ignoring such claims, the difference is only significant at the 0.10 level. A very rough test applied to the combined Salt, Tonto, and Verde winter data suggests a more significant decline in runoff ratio, but better geographic distribution of precipitation stations is needed before this finding can be accepted.

To demonstrate, for the Verde winter runoff, the order of magnitude of the evaluation time required to obtain tcceptable statistical confidence in evaluating effects of watershed treatment aimed at increasing the runoff ratio, an estimate of yield-increase made recently by Croft is compared with the standard error of estimate of a simple linear regression of the historic Verde winter runoff against historic mean winter precipitation for six central Arizona stations. The so-estimated Verde winter increase and the regression standard error of estimate are 30 and 131 thousand acre-feet, respectively. About four successive years of evaluation would be required to establish reality of a positive treatment effect at the 0.025 confidence level. By developing more elaborate regression prediction methods equivalent to a correlation coefficient of 0.97 it would be possible to reduce the required time to perhaps three years. If actual increased water yields should be only half of Croft's upper limiting estimate, the experimental time would, even for 0.97 correlation coefficient, be four to five years. In all it seems rather reasonable to summarize the evaluation problem by predicting that something of the order of *five years* will be required to detect watershed modification effects at a confidence level of 0.025.

Acknowledgments. — The writer wishes to thank Dr. W. D. Sellers for several helpful discussions of topics treated here. Mr. R. B. DesJardins' aid in certain punchcard computing work is appreciated.

LITERATURE CITED

BARR, G. W., and others. 1956. Recovering rainfall, Pt. II. Tucson, Arizona Watershed Program, 218 pp.

BENTON, G. S., and M. A. ESTOQUE. 1954. Water-vapor transfer over the North American continent. J. Meteorol. *11*; 462-477.

COOPERRIDER, C. K., and G. G. SYKES. 1938. The relationship of stream flow to precipitation on the Salt River watershed above Roosevelt Dam. Tucson, Univ. Ariz., Agr. Expt. Sta. Tech. Bull. No. 76, 69 pp.

McDONALD, J. E. 1956. Variability of precipitation in an arid region: A survey of characteristics for Arizona. Tucson, Univ. Ariz., Inst. Atm. Phys. Tech. Rept. No. 1, 88 pp.

McDONALD, J. E. 1957. A note on the precision of estimation of missing precipitation data. Trans. Am. Geophys. Un., 38:657-661.

SCHWALEN, H. C. 1942. Rainfall and runoff in the upper Santa Cruz River drainage basin. Tucson, Univ. Ariz., Agr. Expt. Sta. Tech. Bull. No. 95. 421-472.

SCHWALEN, H. C., and R. J. SHAW. 1957. Ground water supplies of Santa Cruz Valley of southern Arizona between Rillito Station and the international boundary. Tucson, Univ. Ariz., Agr. Expt. Sta. Bull. No. 288, 119 pp.