STATISTICAL DISCRIMINATION OF CHANGE IN DAILY RUNOFF

by

Andre J. Dumas and Hubert J. Morel-Seytoux

August 1969

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

34

STATISTICAL DISCRIMINATION OF CHANGE IN DAILY RUNOFF

by

Andre J. Dumas

and

Hubert J. Morel-Seytoux

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80521

August 1969

No. 34

ACKNOWLEDGMENTS

The present paper is based primarily upon Mr. Andre J. Dumas' Master of Science Thesis in the Department of Civil Engineering, Colorado State University entitled, "Detection of a Change in Watershed Response by a Stochastic Analysis of Daily Streamflows." The work was supported by the U.S. Bureau of Reclamation, Contract numbered BR 14-06-D-6597, whose help is gratefully acknowledged.

The U.S. Geological Survey, Water Resources Division, Automatic Data Processing Unit, supplied the daily flow data. Their gracious cooperation is acknowledged.

RELATION OF HYDROLOGY PAPER NO. 34 TO RESEARCH PROGRAM: "HYDROLOGY OF WEATHER MODIFICATION"

The present study is part of a more comprehensive project which has as one of its objectives the development of methods of evaluation of atmospheric water resources programs. Correlatively the application of the methods to a variety of basins forms a basis for selection of suitable watersheds, basins or regions.

Several approaches are possible and are pursued. This report discusses one of them. Two other approaches will be discussed in forthcoming papers with tentative titles, "Suitability of the Upper Colorado River Basin for Precipitation Management" and "Multivariate Discrimination of Change in Seasonal Runoff."

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ABSTRACT

The purpose of this study was the development of a technique for rapid detection of the occurrence of a suspected hydrologic change in high mountain watersheds. A method has been developed that uses a sequence of independent daily flows.

This procedure is superior to previous ones based on seasonal or yearly flows. The results of this investigation show the use of daily, instead of seasonal flow, data in a Student t-test reduces the number of necessary years of data for detection by an average of five in 14 out of the 20 cases studied, or by an average of three for the 20 cases. All of the cases come from the Upper Colorado River Basin. The study is particularly relevant to the planned cloud seeding operations of the Bureau of Reclamation in high elevation areas of the Colorado Rocky Mountains.

The statistical procedure of detection relies on the Target Control concept and the application of a conditional Student t-test, a test of the difference between the adjusted means obtained by the regression lines between Target and Control for the seeded and non-seeded periods.

STATISTICAL DISCRIMINATION OF CHANGE IN DAILY RUNOFF

by

Andre J. Dumas* and Hubert J. Morel-Seytoux**

Chapter I

INTRODUCTION

1.1 Water resources planning. The increasing demand, and in some parts of the world the desperate need for water, has almost inevitably led men in positions of responsibility to be concerned with the problem of water shortage in particular and of water resources in general [1]. Planning of water resources had, until the relatively recent past, been confined primarily to the task of redistribution in space and time of the naturally available water, or to the task of better utilization and reutilization. It is only recently that the idea [2] of increasing the water supply beyond the natural yield of the hydrologic cycle has started to be realized. At present at least two engineered means of increasing the water supply seem to hold promise for the near future: ocean water desalination [3] and precipitation management [4].

The water situation is particularly critical in the Colorado River Basin. The Colorado River system is the largest in the United States that flows mainly through lands with a chronic water deficiency for cultivation of crops [5]. The average specific (or unit) yield of the Lower Colorado River Basin is only 0.3 inches, the lowest yield in the United States for a drainage area of this size [5]. (Unit yield is the depth, in inches, of the cumulative volume of flow during a given period, in this instance a year, when volume is spread uniformly over the whole watershed.) The Upper Colorado River Basin does not yield much better, 2.2 inches. It outranks only a few basins, the Rio Grande and the Missouri basins, but it is far below the Mississippi's 10 inches and the Columbia's 16 inches. Since the 1940's, the basin's population has increased rapidly with an accompanying growth in demand on the region's water resources for irrigation, industrial and domestic uses [6]. Over the decade from 1951 to 1960, the population of the five states comprising the Upper Colorado River Basin has increased by 40 percent, while over the same period the population of the nation as a whole has increased by only 20 percent [7]. Population projections and the associated water demands indicate a need for actual importation of approximately 3 million acre-feet annually by the year 2080 [8]. Development of the vast oil-shale resources alone would require an additional 1 million acre-feet by the year 2000, assuming a daily oil production of four million barrels [5,8]. "This amount of water simply is not there now." [8] Although "the Colorado Basin is closer than most other basins in the United States to utilizing the last drop of available water for man's needs."[5]

Of course there are alternatives to importation to meet these demands: better utilization, reutilization, desalination and precipitation management. Prohibition by Congress to undertake studies of importation schemes for the next ten years emphasizes the serious need for considering the alternatives. Desalination in the Upper Colorado River Basin appears largely unfeasible at present. The lowest quoted cost estimate suggests water in southern California may cost $35 per acre-foot at the source, with storage, transport, and delivery costs additional [5], and of course it is uphill all the way! Within 400 miles from the source it is estimated the cost would have risen to $120 [8]. On the other hand the cost of water produced by cloud seeding winter storms, from ground-based silver iodide nuclei generators, is estimated at roughly $2 per acre-foot, and under full scale operations it is estimated an average additional 1.9 million acre-feet would appear annually in the rivers [9]. The potential economic and quantitative significance of precipitation management is now reasonably well established.

1.2 Evaluation of atmospheric water resources attainments. Successful water resources management in this field requires techniques for detection and measurement of the increase in water yield induced by weather modification. The main difficulties in this evaluation are caused by (a) the natural variability of hydrologic variables which exceeds the expected range of the increase induced by man, and (b) the inaccuracy of the discharge measurements. Simple statistical tests have been developed [10]. They have not proven very sensitive and, as a result, require long periods of observations, prior to and during seeding operations, in order to give satisfactory test results. Furthermore, these tests are insensitive when experiments are performed during a dry period of annual stream flow sequences. Therefore, more sophisticated techniques were needed. The target control concept was introduced, and different tests were devised [10], including a Chi-square test and a Student-t test. In a recent study [11], a target-control Chi-square test was applied to the mean annual or mean seasonal flows of some rivers and it was shown the number of years M(or sample size) necessary to detect, at the 95% level of significance and 50% power a given percentage h of increase in the yearly or seasonal flows was:

$$M = 4(1-\rho^2)\,\frac{C^2_{v,T}}{h^2} \qquad (1)$$

*Former M.S. Graduate of Colorado State University, Civil Engineering Department, Fort Collins, Colorado, presently with Ministere des Richesses Naturelles, Quebec, Canada

**Associate Professor of Civil Engineering, Colorado State University, Fort Collins, Colorado.

where ρ is the correlation coefficient between the target and the control watersheds, and $C_{v,T}$ is the coefficient of variation of the target watershed. Calculations were performed for a few stations in the Upper Colorado Basin to get an idea of what could be expected if seeding operations were conducted in the area. In particular the expected number of years to detect a 10% increase was calculated [11]. The results are shown in Table 1. The results are encouraging though still too high. The best results, 4 and 6, have to be discounted largely because of the proximity of the target and control and the resulting quasi-impossibility to prevent contamination. What then can be done to reduce the number of years needed to obtain significance?

TABLE 1

EXPECTED NUMBER OF YEARS TO DETECT A 10% INCREASE AT THE 95% LEVEL OF SIGNIFICANCE FOR A FEW PAIRS OF TARGET-CONTROL STATIONS IN THE UPPER COLORADO RIVER BASIN, BASED ON SEASONAL RUNOFF

TARGET					CONTROL			TARGET-CONTROL PAIR		
CSU Number	USGS Station Name	Drainage (sq mi)	Elevation (ft)	Coefficient of Variation (%)	CSU Number	Station Name	Drainage	Years of Common Record	Coefficient of Correlation (%)	Years Needed for Significance at 95% confidence level
1073440	Junction Creek near Durango, Colorado	26	7045	36	1073448	Hermosa Creek near Hermosa, Colorado	172	5	85	14
1073480	Animas River at Howardsville, Colorado	56	9617	27	1073448	Hermosa Creek near Hermosa, Colorado	172	25	90	6
1278800	Dolores River below Rico, Colorado	105	8422	45	1073448	Hermosa Creek near Hermosa, Colorado	172	13	98	4
1590000	Roaring Fork at Glenwood Springs, Colorado	1460	5720	33	1600000	Colorado River at Glenwood Springs, Colorado	4560	58	89	9
1594236	North Fork Fryingpan near Norrie, Colorado	41	8400	30	1594260	Fryingpan River at Norrie, Colorado	90	23	91	7

There are several avenues open to answer this fundamental question. One avenue is to improve the test to which the data are subjected. It was not promising. Another avenue consists of grouping observations in some favorable manner for several targets, or better, for several targets and controls. Both avenues are presently being pursued. The last avenue, which is the subject of this study, looks for an optimal test variable, given the test, i.e., a single target-control conditional Student's t-test [11].

First one must answer the following question: which variable, annual, seasonal, monthly or daily runoff, is a better detector? Theoretically this question has been answered, in general and the daily runoff variable is the most promising. The basic underlying idea is that the shorter the time interval --by which the time series of river streamflow is divided into a discrete time series--the more information one will derive. (Daily flow is defined in this study as the average daily runoff at a section of river, the averaging being done either from a continuous record of an automatic recorder or from river stage measurements taken at representative time intervals to make interpolation and averaging consistent.) From a practical point of view, however, it is not so clear cut because the power of the detection procedure depends not only on the sample size, but also on the variability of the runoff (which increases as the unit of time decreases), the magnitude of the measurement error, the degree of correlation between the variable in the watershed of interest and a control watershed, the physical nature of the suspected cause of the change in runoff, and the magnitude of the resulting effect. The purpose of this study was to initiate a preliminary investigation of the practical value of daily runoff for evaluation. The qualified conclusion of the study is that, indeed, it has practical merit.

Chapter II

THE TARGET-CONTROL CONDITIONAL STUDENT'S t-TEST

2.1 An optimization problem in detection. The problem of early detection of a change in watershed runoff received impetus as controversy characterized the field of weather modification. Early weather modification experiments were conducted without much care for the statistical design of the experiments. In an early stage of a new science this oversight is understandable. What purpose is it to draw tables of the number of years for significance at a given level versus all possible hypothetical percentage increases, if even the order of magnitude of that increase is totally unknown? The availability of the table would not have affected the decision to proceed with the experiments. On the other hand, once the order of magnitude of the increase is known, the table becomes crucial. It is crucial because the percentage increase in runoff turns out to be small, on the order of 10%. Careful inspection of the table becomes a requirement in the design of new experiments. It may lead to a variety of questions; e.g., will it be possible to show significance at say the 90% level within the contemplated five years of experiments? If not, can significance be attained by shifting the experiments to a different location? If not...well, how good was the table in the first place?

At this point it is necessary to state clearly the objective of a method of detection. For different objectives different methods will be required. Ideally one wants to find the technique that will permit one to ascertain, in the minimum amount of time, that an identified cause, e.g., cloud seeding, has affected a selected measure of watershed response at a chosen significance level. Once that technique has been found, it becomes possible to calculate the number of years needed for significance at a given power. (The power is the probability that significance will be attained within this number of years.) This number of years depends on several parameters, the chosen significance level, the chosen power, the degree of certainty of identification of the cause (i.e., is cloud seeding really responsible for the detected change?), the selected response (e.g., hourly precipitation, monthly runoff), the characteristics of the watershed (i.e., the nature of the transfer function between cause and effect), and the magnitude of the change in watershed response. Ideally one would like to find the technique for which the calculated number of years is minimum for all possible values of the previously listed parameters....It cannot be done....Even less ambitious optimization problems cannot receive a general solution. A technique will be optimal for a certain range of parameters but not for others [12]. One is therefore forced to limit the original ambition to a more realizable level. Besides, the optimization problem will not present itself usually in this unconstrained form. The detection scheme must be compatible with a variety of restraints of diverse nature. For example, from a statistical point of view the target-control pair Dolores-Hermosa (line 3 of Table 1) would be ideal. However, the accuracy of targeting with ground-based generator is not sufficiently developed to permit such a close control.

Short of overall optimization one must settle for suboptimization. Of course once this step is taken, and there is no other choice, there is an infinite variety of possible options. As discussed in the Introduction there are several avenues for research. In the present study the following suboptimization problem was considered. Given that the cause of a suspected change has been identified (be it cloud seeding, timber cuts, etc.)--that its effect can be measured as runoff, that the statistical technique to which the data will be subjected is the single target-control conditional Student's t-test--what is the optimal test variable, seasonal or daily runoff? This is the problem.

It is a much restricted problem in appearance but an important practical one. This assertion is validated by the conclusion of the study. Without the benefit of the conclusion it could nevertheless be inferred a priori from the following heuristic reasoning. In the limited number of cases for which formulae are actually available to calculate the number of years, this number is inversely proportional to the number of data per year. Using daily flow versus a four-months seasonal flow could therefore bring a reduction by two orders of magnitude. One expects a greater variance for daily flow. Because the number of years is proportional to this variance, one expects a reduction in the potential gain from using daily flow. Similarly the expected decrease in the coefficient of correlation between target and control will further limit the gain. It is difficult to believe these effects could completely wipe out a gain of 100! However, the most severe limitation will come from the choice of the test itself. It is therefore important to discuss this test and the assumptions underlying its derivation. This is the purpose of the next section.

One might ask, "Why not use a better test?" The answer to this question is two-fold: if there is one, it is well hidden in the literature, and second it is fairly evident, from experience, that sophistication in statistical techniques reaches rapidly a point of diminishing returns unless paralleled with judicious selection of variables to be tested and a thorough knowledge of the particular local hydrologic conditions. Again this point is justified by the conclusion of the present study.

2.2 Target-control conditional Student's t-test. The goal of weather modification experiments is to increase the runoff in the watershed, and it is logical to postulate the null [12] hypothesis:

H_o: There is no change in mean runoff due to the weather modification experiments. This will be tested against the alternative [12] hypothesis

H_a: There is a change in mean runoff caused by man's weather modification experiments. If the art of weather modification is advanced enough the possibility of a decrease need not be considered and a one-tailed [12] test is implied. If not, a two-tailed [12] test is implied.

The level of significance α which is the probability of rejecting a true hypothesis, will be either 5% or 1%.

The target-control concept uses the relationship existing between the streamflows, from a treated or target watershed, to those from an adjacent and untreated watershed; the latter serving as a control to the previous watershed, since its flows are not affected by the cloud seeding operations. Additional information from the control watershed can be used to discriminate a change in the target watershed behavior. In other words, it makes the target look as though it has an effective coefficient of variation much smaller than its actual one. The larger the coefficient of correlation between target and control the smaller the apparent coefficient of variation of the target. This concept assumes:

(a) The target and control streamflows are highly correlated.

(b) The control watershed is sufficiently far from the target watershed to preclude contamination, but close enough to provide a high correlation.

(c) The target streamflow observations are independent.

With (x) being the series of independent flows for the control and (y) the corresponding series for the target, a bivariate normal distribution is assumed for the joint series (x,y) for the non-seeded period. The seeded period will provide two new sets of observations (ξ) and (η); (ξ) and (η) being the sets of independent flow values, respectively, for the control and the target. It is assumed that the coefficient of correlation ρ between target and control has not changed during the seeded period, and that the joint series (ξ,η) has also a bivariate normal distribution.

When the above conditions are satisfied, any significant difference in streamflow, taking into account the relation between the two watersheds, beyond that associated with a natural variation can be attributed to cloud seeding effects.

Because variances of the target and control variables and their coefficient of correlation are assumed unaffected by seeding, the two regression lines, one for the sample before seeding, one for the sample after seeding, are parallel (see Fig. 1). Then, the null hypothesis is that the two populations have the same regression line, that is, the difference in ordinates at the origin AB is not significantly different from zero. It should be noted that whether or not the control mean has changed under seeded conditions will not affect the test.

The null hypothesis can be formulated in this way: the adjusted means of the two populations, $\bar{y}_{x_o}$ and $\bar{\eta}_{x_o}$ at $x = \xi = x_o$, are equal, whatever the value of x_o. The adjusted means are:

for the non-seeded period, $\bar{y}_{x_o} = \bar{y} - \bar{b}(x-\bar{x})$, and

for the seeded period, $\bar{\eta}_{x_o} = \bar{\eta} - \bar{b}(\xi-\bar{\xi})$.

Where $\bar{b}$ is the weighted average regression coefficient:

$$\bar{b} = \frac{\sum_{i=1}^{N} (x_i-\bar{x})(y_i-\bar{y}) + \sum_{j=1}^{M} (\xi_j-\bar{\xi})(\eta_j-\bar{\eta})}{\sum_{i=1}^{N} (x_i-\bar{x})^2 + \sum_{j=1}^{M} (\xi_j-\bar{\xi})^2} \ .$$

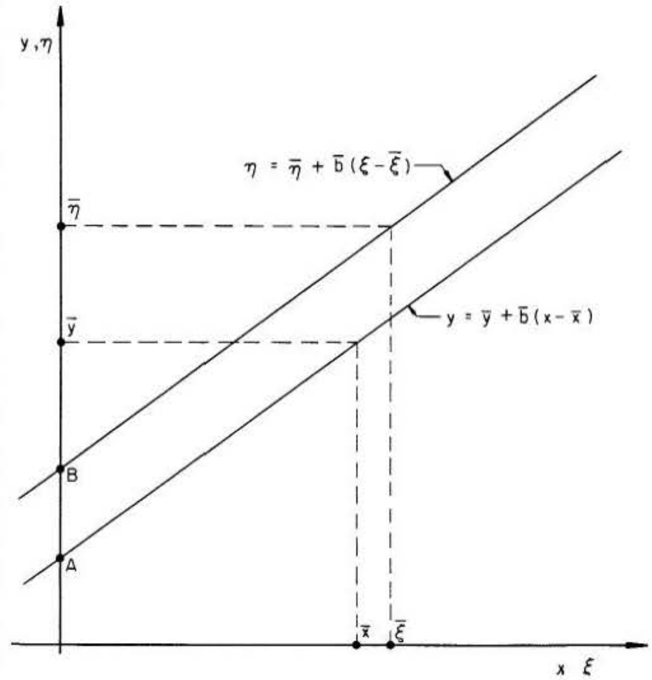


Fig. 1 Target control regression lines before and after seeding operations

The difference AB is:

$$\bar{y}_{x_o} - \bar{\eta}_{x_o} = \bar{y} - \bar{\eta} - \bar{b}(\bar{x}-\bar{\xi}) \ .$$

AB is a linear combination of three independent observations $\bar{y}$, $\bar{\eta}$, $\bar{b}$ with population means μ_y, μ_η, β and variances $\frac{\sigma^2}{N}$, $\frac{\sigma^2}{M}$, $\frac{\sigma^2}{\sum_{i=1}^{N} (x_i-\bar{x})^2 + \sum_{i=1}^{M} (\xi_i-\bar{\xi})^2}$ respectively. Then AB has a normal distribution with mean $\mu_y - \mu_\eta - \beta(\bar{x}-\bar{\xi})$ and variance

$$\sigma^2 \left[\frac{1}{N} + \frac{1}{M} + \frac{(\bar{x} - \bar{\xi})^2}{\sum_{i=1}^{N} (x_i-\bar{x})^2 + \sum_{i=1}^{M} (\xi_i-\bar{\xi})^2}\right]$$

where σ^2 is the common variance of the arrays.

Under the null hypothesis, $H_o: \mu_y - \mu_\eta = \beta(\bar{x}-\bar{\xi})$, the statistic

$$t_o = \frac{\bar{y} - \bar{\eta} - \bar{b}(\bar{x} - \bar{\xi})}{s\left[\frac{1}{N} + \frac{1}{M} + \frac{(\bar{x} - \bar{\xi})^2}{\sum_{i=1}^{N} (x_i-\bar{x})^2 + \sum_{j=1}^{M} (\xi_j-\bar{\xi})^2}\right]^{1/2}} \qquad (2)$$

where s^2 is the unbiased estimate of the common variance of the arrays:

$$s^2 = (1-r\ ) \frac{\left[\sum_{i=1}^{N} (y_i-\bar{y})^2 + \sum_{j=1}^{M} (\eta_j-\bar{\eta})^2\right]}{N + M - 3}$$

$$r^2 = \frac{\left[\sum_{i=1}^{N} (x_1-\bar{x})(y_i-\bar{y}) + \sum_{j=1}^{M} (\xi_j-\bar{\xi})(\eta_j-\bar{\eta})\right]^2}{\left[\sum_{i=1}^{N} (x_i-\bar{x})^2 + \sum_{j=1}^{M} (\xi_j-\bar{\xi})^2\right]\left[\sum_{i=1}^{N} (y_i-\bar{y})^2 + \sum_{j=1}^{M} (\eta_j-\bar{\eta})^2\right]}$$

follows Student's t distribution with $(N + M - 3)$ degrees of freedom [13].

On the basis of the data, t_o can be computed; a subroutine has been written for this purpose [14], and it performs a one- or two-tailed test by comparison of t_o with a table of the Student's t distribution as a function of the number of degrees of freedom.

Chapter III

STREAMFLOW DATA USED FOR STUDY

All streamgage stations used in this study are located in the Upper Colorado River Basin within the State of Colorado. The target and control watersheds must satisfy some criteria as closely as possible. These conditions, which form the basis for the selection of the watersheds, are now discussed.

3.1 Physiography and location. The statistical investigation of weather modification attainments as presented in this paper were undertaken in connection with a project of the Bureau of Reclamation, Office of Atmospheric Water Resources. A pilot project to increase winter precipitation over high elevation watersheds in two areas of the Upper Colorado River Basin [15] is to be initiated in 1969.

The watersheds selected for this study are located in the Upper Colorado River Basin and have elevations as near as possible to the 9,000 feet level--a level determined [9] as a requirement to start a nucleation process in cloud seeding experiments. The majority of the selected stations are about 7,000 feet high.

No restriction was imposed on the size of the drainage area. Watersheds of more than 100 square miles are preferred because they are more likely to provide a more representative response to a man-made increase in precipitation.

3.2 Availability of records. A rather sizable number of data is required when working with daily flows; therefore, the computations were handled by the CDC 6400 computer at Colorado State University. Because better and fast processing of data can be done on magnetic tapes, watersheds with available data on these tapes were selected. Selection of thirty-one stations in the Upper Colorado River Basin from a U.S. Geological Survey tape was based on the accuracy of historical records.

3.3 Virginity of the flows and accuracy of the measurements. Most of the rivers of the Colorado River Basin have been subjected at one time or another, to some kind of human intervention, regulation or diversion. For the purpose of detection of an increase due to artificial precipitation, virginity of the flow is strongly required because man-made diversions or regulations by dams often far exceed the range of the expected increase due to cloud seeding and are not often consistent in time and in quantities from year to year.

Streamflows affected by Transmountain Diversions were excluded because such diversions generally involve important quantities of water, and the data required for corrections were not available.

Streamflows with upstream regulation or transbasin diversions were excluded except where the dams causing the regulation are small or the diversions are made for irrigation of very small acreages. Streamflows with intrabasin diversion for irrigation were accepted if the size of the irrigated area was small.

For the spring season the United States Geological Survey considers the accuracy of the discharge measurements as good.

3.4 Correlation target-control. A high correlation between target and control watersheds daily flows is desirable for the purpose of this study. To discriminate among the stations before starting the study of the daily flows, the correlation between target an control was estimated using seasonal flow, i.e., water yield from March to August.

On the basis of these criteria, 10 stations were selected (see Table 2, 3 and 4 and Fig. 2.

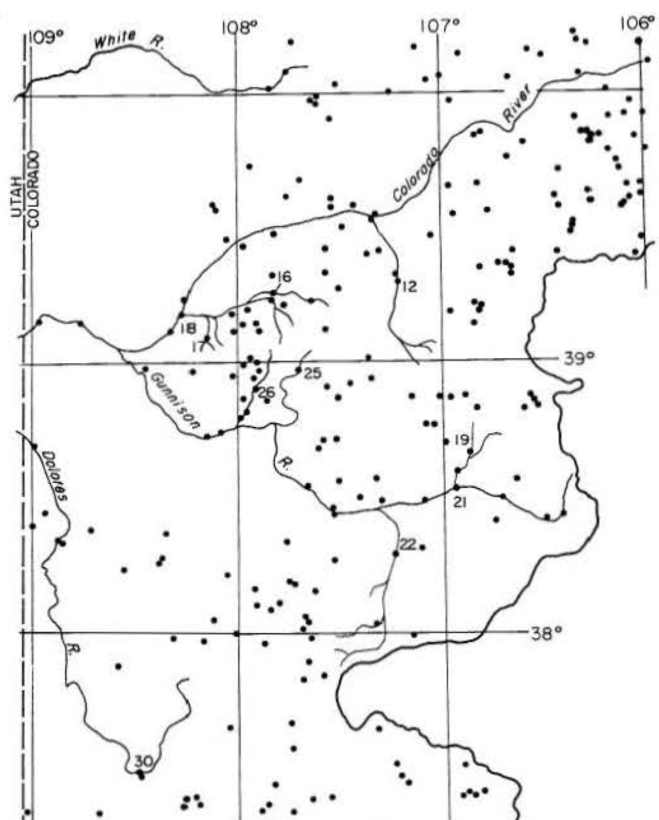


Fig. 2 Location of the selected stations in the Upper Colorado River Basin

TABLE 2

DESCRIPTION OF THE STATIONS SELECTED

Identification Tape no.	USGS no.	CSU no.	Elevation ft.	Drainage area sq. mi.	Length of record year	Trans-mountain diversion	Upstream regulation	Trans-basin diversion	Intra-basin diversion*
12	9.0825	1592140	6400	225	25	None	None	None	irrig. for 2050 ac.b.
16	9.0975	1425625	6920	139	39	None	None	to irrig. 280 ac.	irrig. for 1300 ac. a.
17	9.1045	1420800	7400	7	20	None	small dam	None	None
18	9.1050	1420000	4800	604	21	None	small dams	None	irrig. for 25000 ac. a.
19	9.1125	1378100	8008	295	38	None	None	None	irrig. for 7400 ac. a.
21	9.1190	1377200	7628	1020	20	None	None	None	irrig. for 24000 ac. a.
22	9.1245	1375400	7827	338	23	None	None	None	irrig. for 24000 ac. a.
25	9.1345	1373020	7160	35	20	None	small dam	None	small irrig. no data
26	9.1435	1371810	6500	39	40	None	small dam	small-no data	small irrig. no data
30	9.1665	1277200	6924	556	48	None	None	None	irrig. 2000 ac. b.

*irrig. means irrigation; ac. a. means acres above station; ac. b. means acres below station.

STATION DESCRIPTIONS

Identification Tape no.	USGS no.	CSU no.	Name
12	9.0825	1592140	Crystal River near Redstone, Colorado
16	0.0975	1425625	Buzzard Creek near Collbran, Colorado
17	9.1045	1420800	Mesa Creek near Mesa, Colorado
18	9.1050	1420000	Plateau Creek near Cameo, Colorado
19	9.1125	1378100	East River at Almont, Colorado
21	9.1190	1377200	Tomichi Creek at Gunnison, Colorado
22	9.1245	1375400	Lake Fork at Gateview, Colorado
25	9.1345	1373020	Leroux Creek near Cedaredge, Colorado
26	9.1435	1371810	Surface Creek at Cedaredge, Colorado
30	9.1665	1277200	Dolores River at Dolores, Colorado

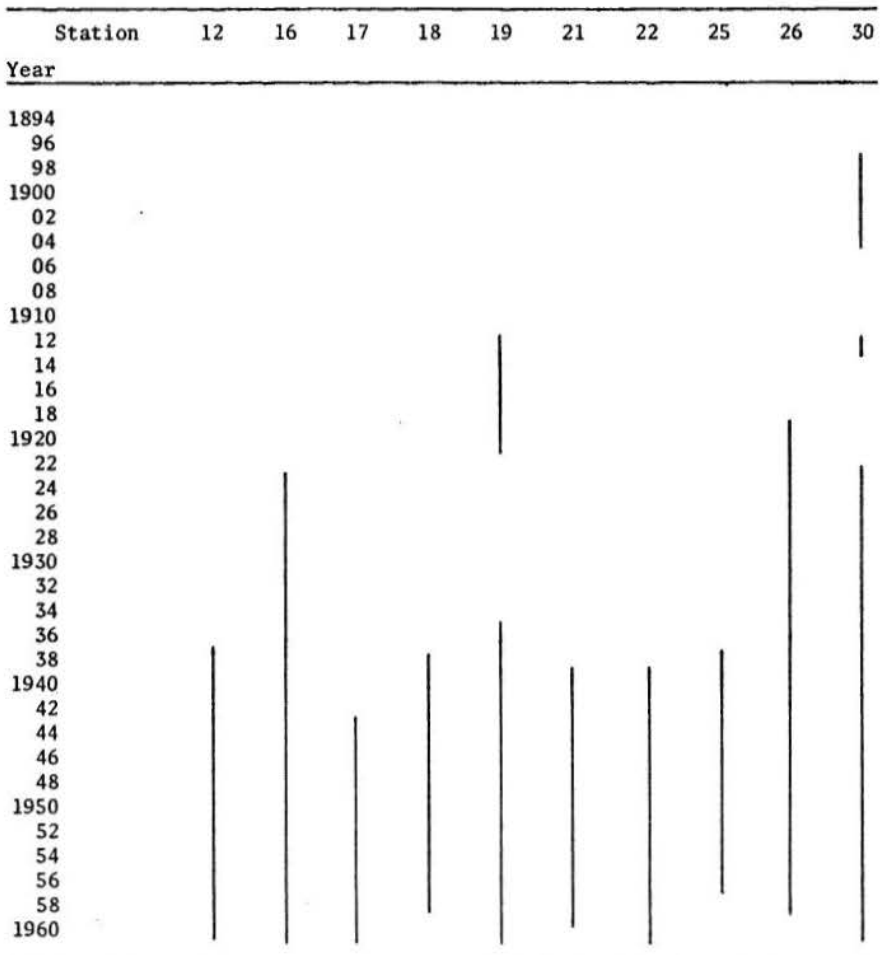
TABLE 3

TARGET-CONTROL CORRELATION ON THE BASIS OF SEASONAL FLOWS

Identi-fication	12	16	17	18	19	21	22	25	26
12									
16	0.771								
17	0.625	0.889							
18	0.728	0.969	0.892						
19	0.94	0.629	0.515	0.618					
21	0.825	0.829	0.715	0.811	0.862				
22	0.807	0.866	0.736	0.832	0.792	0.878			
25	0.88	0.852	0.822	0.838	0.771	0.766	0.795		
26	0.776	0.876	0.836	0.833	0.659	0.765	0.827	0.92	
30	0.785	0.854	0.889	0.877	0.694	0.848	0.914	0.803	0.872

TABLE 4

LENGTH AND AVAILABILITY OF HISTORICAL RECORD FOR DAILY FLOWS



Chapter IV

THE STOCHASTIC STRUCTURE OF DAILY FLOW

4.1 The naive approach. It might be summarily inferred that the use of daily runoff instead of seasonal runoff in the application of the test would only entail a larger amount of data processing. However, this quick extrapolation is erroneous for two reasons:

(1) The daily flow observations for different days of the year come from different statistical populations, and

(2) From day to day the flow values are highly correlated.

For these two reasons the straight application of the test to daily runoff for every day of the season and on face value would violate the assumptions of the derivation of the test and invalidate the results of the test. Assertion (1) is demonstrated in Fig. 3. The expected value $P(t)$, or more rigorously its estimate, $\hat{P}(t)$, of the daily flow, $Q(t)$, varies from one date to another. In this study the time variable t takes only discrete integer values, with $t = 1$ corresponding to the first day of the water year, i.e. October 1st, and $t = 365$ to September 30. For convenience a table of correspondence between calendar dates and values of t is given (Table 5). The sets of Fig. 3 show that the standard deviation also varies considerably from day to day. In these figures, the coefficient of variation (ratio of standard deviation over mean) is also given.

Assertion (2) is also clearly supported in Fig. 4 which shows the autocorrelation values, $r(k)$, for all dates of the year and for various lags.

4.2 Standardization of daily streamflows. To overcome difficulty (1), i.e., the fact that daily flow observations for different dates of the year come from different statistical populations, it is necessary to perform a transformation on the daily flow values. Hopefully the transformed data will belong to the same population. If $Q(t)$ denotes the daily flow for date t, $P(t)$ its expected value, $\hat{P}(t)$ the estimate of $P(t)$, $S(t)$ and $\hat{S}(t)$ the standard deviation and its estimate, then the annual observation of $Q(t)$, $Q_i(t)$ can be standardized by the transformation:

$$q_i(t) = \frac{Q_i(t) - \hat{P}(t)}{\hat{S}(t)} \tag{3}$$

with i being an index referring to the year,

$$\hat{P}(t) = \frac{1}{n} \sum_{i=1}^{n} Q_i(t) \quad \text{for any given } t,$$

where n is the number of years with available records, and

$$\hat{S}^2(t) = \frac{1}{n-1} \sum_{i=1}^{n} [Q_i(t) - \hat{P}(t)]^2 \text{ for any given } t.$$

The standardized daily runoff variable:

$$q(t) = \frac{Q(t) - \hat{P}(t)}{\hat{S}(t)} \tag{4}$$

is approximately normal if $Q(t)$ is normally distributed, with expected value approximately zero and variance approximately unity. For the historical period of record the sample estimate of the expected value of $q(t)$ is exactly zero and the estimate of the variance is exactly one, from the very definition of $q(t)$.

To pool together and use the daily flows for different t, as elements of one and the same population, the series must be "stationary." In hydrologic investigation, it is generally considered sufficient to have wide-sense stationarity. Wide-sense stationarity is defined by the following two equations where E[ ] denotes the expected value:

$$E[q(t)] = \text{Constant}$$

$$\text{Cov}[q(t_1)q(t_2)] = C(t_2-t_1): \text{ a function of } (t_2-t_1)=k \text{ only.}$$

From the very definition of $q(t)$ the first condition is met and the second condition is met for $t_1 = t_2$. It remains to verify that the second condition is met for various lag values. The dependence of a given day t_1 with another day t_2 can be measured by the correlation coefficient r, computed over the two samples of n elements of the populations of the daily flow for these two given days:

$$\hat{r}(k) = \frac{\hat{C}\text{ov}[q(t_1),q(t_2)]}{\left[\hat{V}\text{ar}[q(t_1)]\hat{V}\text{ar}[q(t_2)]\right]^{1/2}}$$

with $k = (t_2-t_1)$.

By the nature of the standardization procedure this expression reduces [16] to the simpler form:

$$\hat{r}(k) = \frac{1}{n} \sum_{i=1}^{n} q_i(t_1)q_i(t_2) \ . \tag{5}$$

The computation of $r(k)$ was performed for different values of t_2 and k; t_2 varying from 1 to 365 and k from 1 to 37. Analysis of the results points to the following:

(a) For a given value of k, $r(k)$ varies significantly for different t_2, that is, from day to day, and the assumption, $r(k)$ depends only on $k = t_2-t_1$, cannot be considered as valid throughout the whole year. In other words, the standardization did not yield stationarity in the wide sense.

(b) For a given day (t_2), $r(k)$ decreases and tends toward zero, as k increases.

However, it is possible to consider that the coefficient of correlation, depends only on k for some period of the year (see Fig. 4). This period is the spring season, more precisely it extends from March to June.

For the spring season it is legitimate to consider that the conditions of stationarity in the wide sense are met. It is then possible to consider, as is usually done [17], that the mathematical expectation of both $q(t_1)$ and $q(t_1)q(t_2)$--obtained by averaging over an ensemble of realizations of the time series--can be replaced by the time averages of the same quantities over one realization. The advantage of this procedure is to permit the use of a sample of larger size. Proper application also requires that correlation between ordinates of the random function $q(t)$, taken at different instants of time, should decrease with sufficient rapidity, since it is only in this case that one realization with respect to time can be approximately considered as a set of several independent realizations, and that the difference between means obtained by these two methods vanish. This latter condition is accepted on the basis of the results found for $r(k)$.

The serial correlation coefficient $R_i(k)$ for a given realization i, that is for a given year i can be computed. Again by the nature of the standardization procedure [16] the expression is simple:

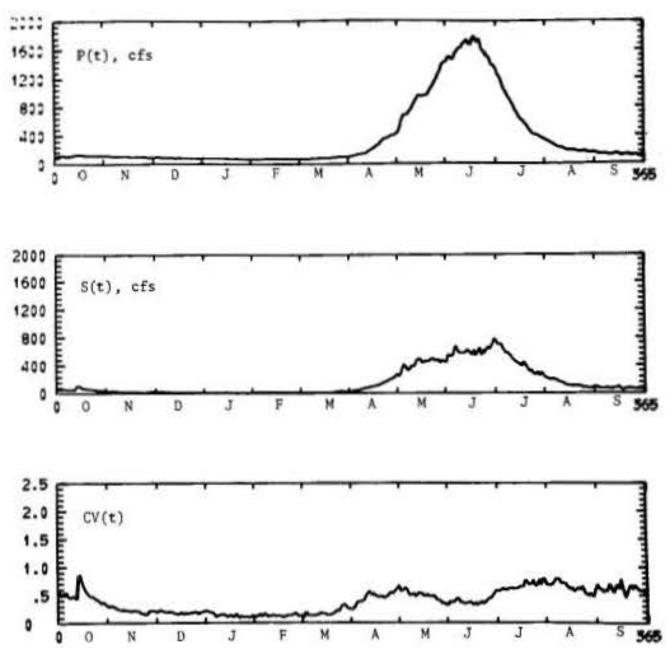
$$R_i(k) = \frac{1}{\beta-\alpha} \sum_{t=\alpha}^{\beta} q_i(t)\, q_i(t-k) \tag{6}$$

In eq. (6) α and β are the indexes of the days which respectively begin and end the considered spring period. For a station with n years of historical records, n values for $R(k)$ can be computed for every value of k. If all the realizations have been obtained under identical conditions, it is suggested [18] that each of them should be analyzed by the method indicated above. Then the estimated values of the mathematical expectations and correlation functions should be averaged over all the realizations.

The average of the $R_i(k)$ over all realizations i is:

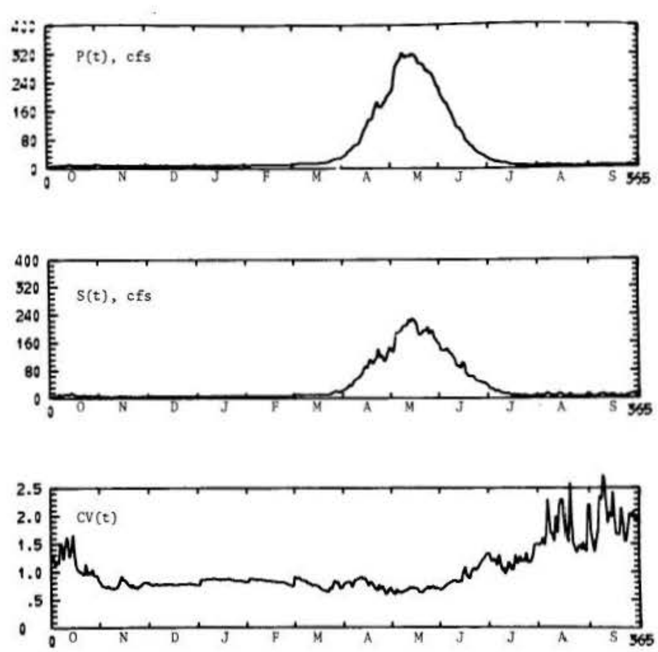
$$\overline{R}(k) = \frac{1}{n} \sum_{i=1}^{n} R_i(k) \ . \tag{7}$$

Based on the correlograms, i.e., graphs of $\overline{R}(k)$ versus k, it is possible to determine a minimum lag beyond which the standardized daily flows can be considered as independent. The resulting series of spaced standardized daily flows then satisfies the conditions of applicability of the target-control test.



Characteristics of the daily flow random function $Q(t)$, for Station 12 -- Crystal River near Redstone, Colorado

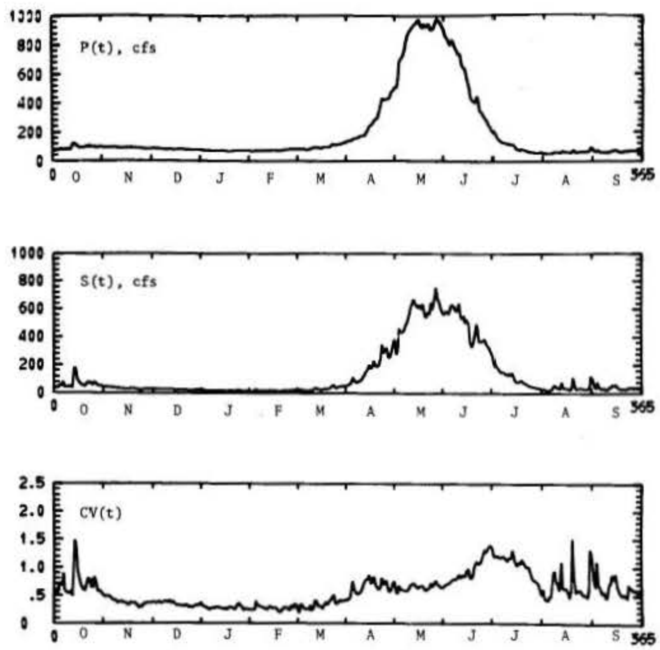
$P(t)$: Expectation of $Q(t)$
$S(t)$: Standard deviation of $Q(t)$
$CV(t)$: Coefficient of variation of $Q(t)$



Characteristics of the daily flow random function $Q(t)$, for Station 16 -- Buzzard Creek near Collbran, Colorado

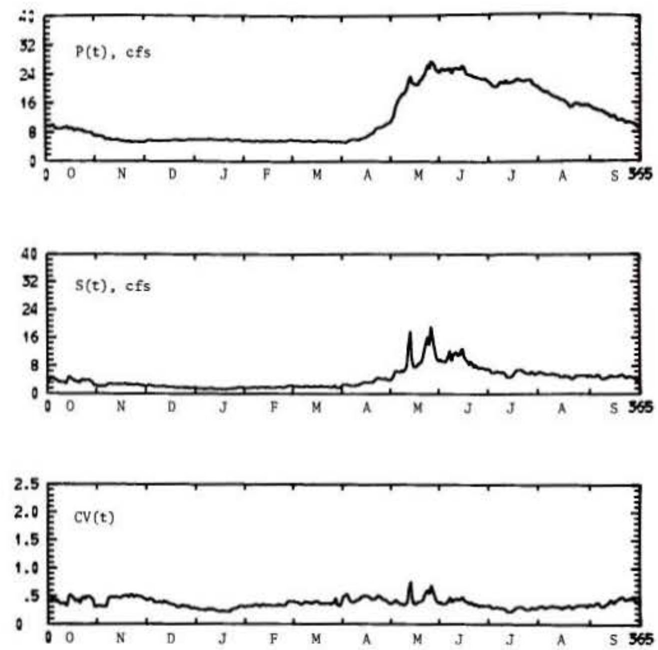
$P(t)$: Expectation of $Q(t)$
$S(t)$: Standard deviation of $Q(t)$
$CV(t)$: Coefficient of variation of $Q(t)$

Figure 3



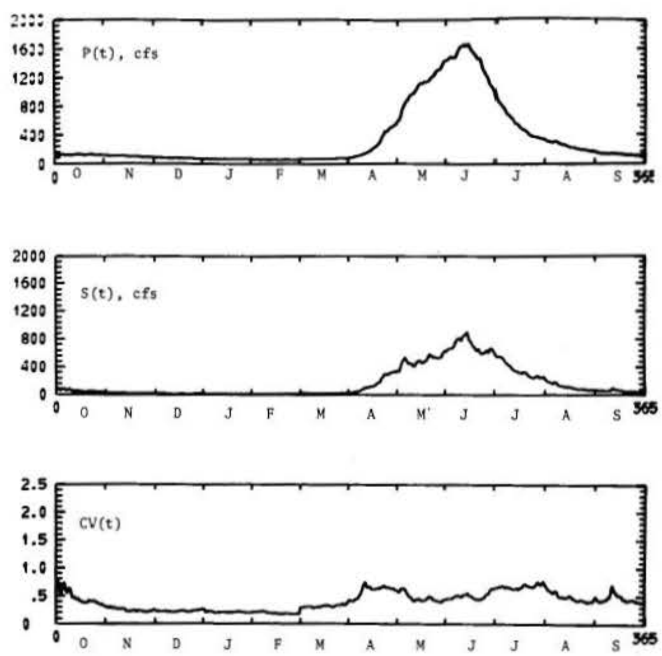
Characteristics of the daily flow random function Q(t), for Station 17 -- Mesa Creek near Mesa, Colorado

P(t): Expectation of Q(t)
S(t): Standard deviation of Q(t)
CV(t): Coefficient of variation of Q(t)



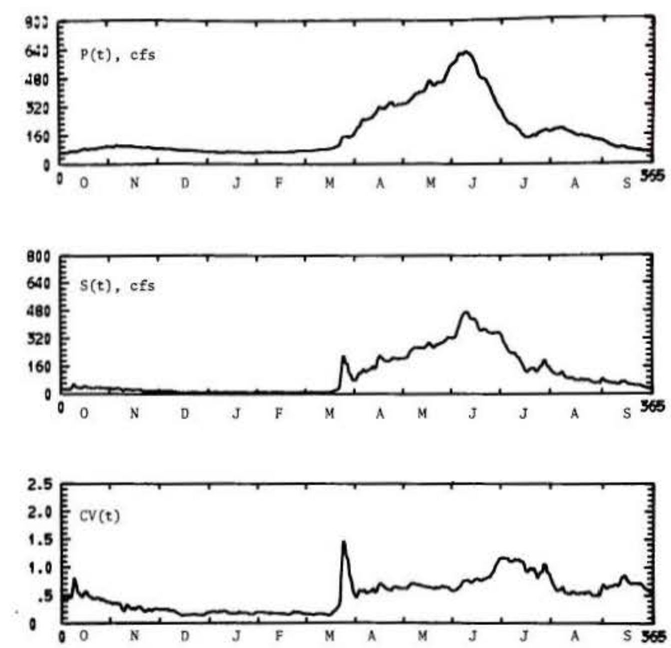
Characteristics of the daily flow random function Q(t), for Station 18 -- Plateau Creek near Cameo, Colorado

P(t): Expectation of Q(t)
S(t): Standard deviation of Q(t)
CV(t): Coefficient of variation of Q(t)



Characteristics of the daily flow random function Q(t), for Station 19 -- East River at Almont, Colorado

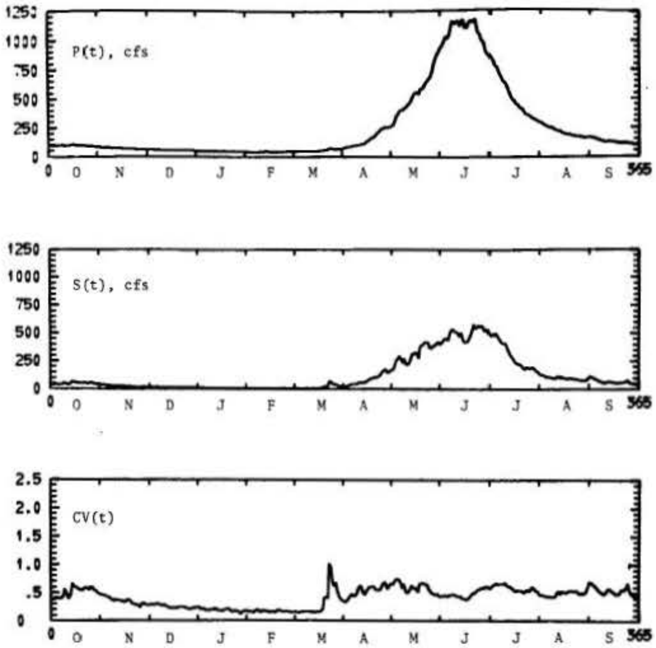
P(t): Expectation of Q(t)
S(t): Standard deviation of Q(t)
CV(t): Coefficient of variation of Q(t)



Characteristics of the daily flow random function Q(t) for Station 21 -- Tomichi Creek at Gunnison, Colorado

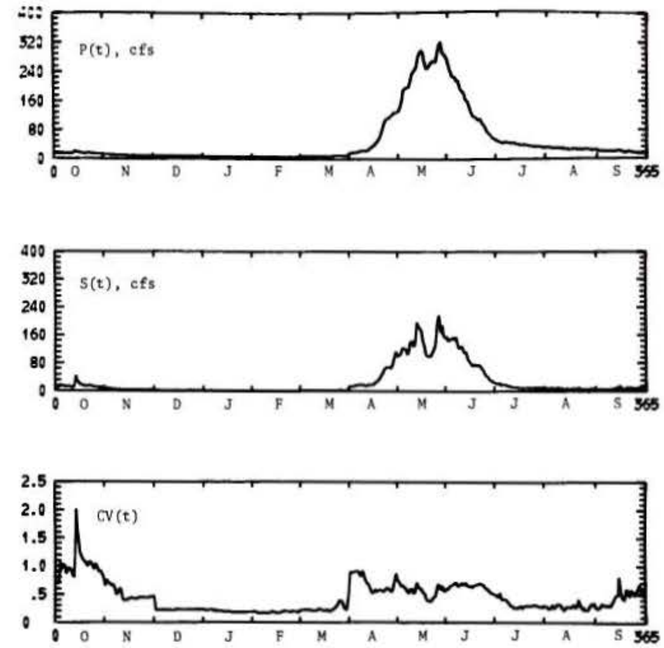
P(t): Expectation of Q(t)
S(t): Standard deviation of Q(t)
CV(t): Coefficient of variation of Q(t)

Figure 3 (continued)



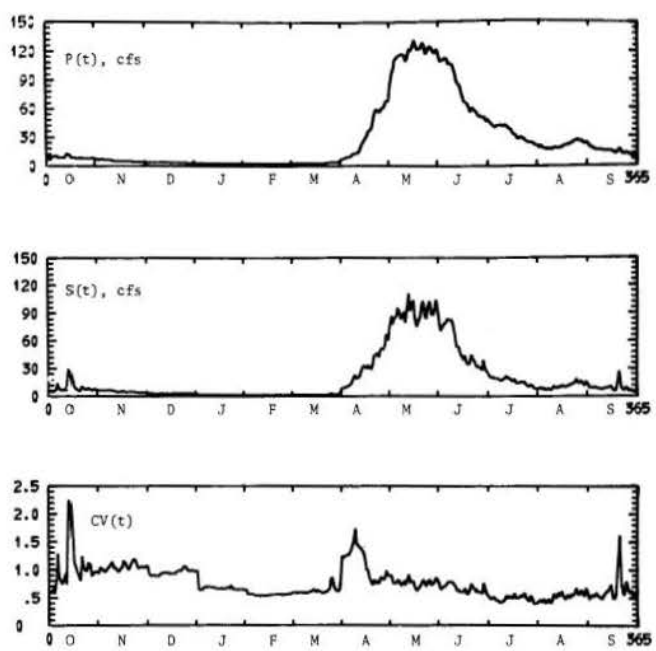
Characteristics of the daily flow random function Q(t), for Station 22 -- Lake Fork at Gateview, Colorado

P(t): Expectation of Q(t)
S(t): Standard deviation of Q(t)
CV(t): Coefficient of variation of Q(t)



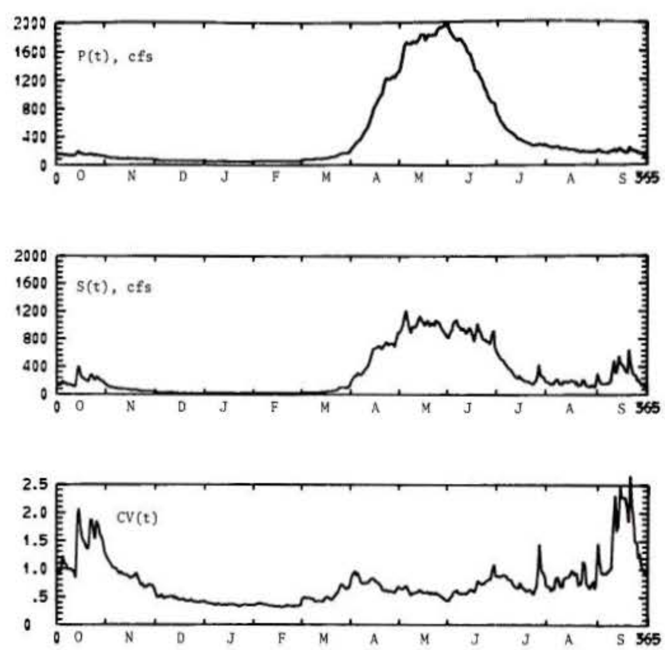
Characteristics of the daily flow random function Q(t), for Station 25 -- Leroux Creek near Cedaredge, Colorado

P(t): Expectation of Q(t)
S(t): Standard deviation of Q(t)
CV(t): Coefficient of variation of Q(t)



Characteristics of the daily flow random function Q(t), for Station 26 -- Surface Creek at Cedaredge, Colorado

P(t): Expectation of Q(t)
S(t): Standard deviation of Q(t)
CV(t): Coefficient of variation of Q(t)



Characteristics of the daily flow random function Q(t), for Station 30 --Dolores River at Dolores, Colorado

P(t): Expectation of Q(t)
S(t): Standard deviation of Q(t)
CV(t): Coefficient of variation of Q(t)

Figure 3 (continued)

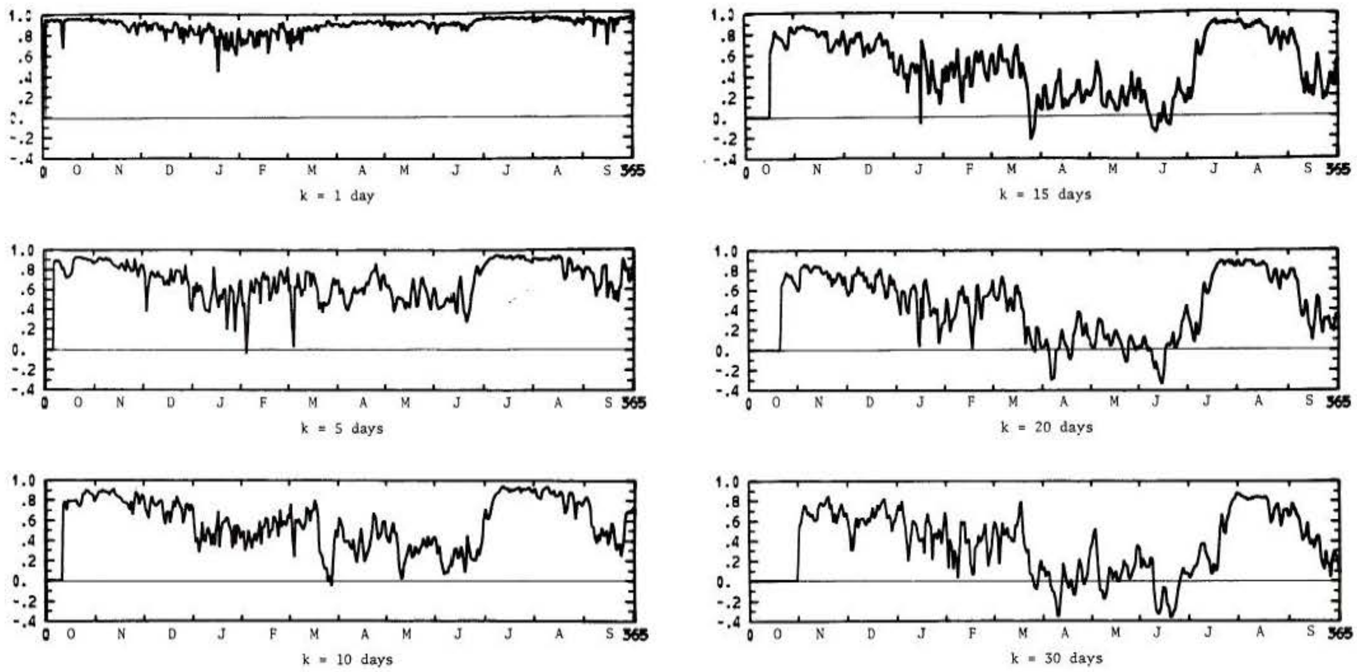


Figure 4 An illustration of r(k) versus time for Station 12 and different values of k

-r(k) has not been computed and has been set up equal to zero for the first k days of the water year

TABLE 5

CORRESPONDENCE BETWEEN CALENDAR YEAR DATE, WATER YEAR DATE AND DAY INDEX

Oct.		Nov.		Dec.		Jan.		Feb.		MARCH	
Water Year Date	Day Index	Water Year Date	Day Index	Water Year Date	Day Index	Water Year Date	Day Index	Water Year Date	Day Index	Water Year Date	Day Index
1-1	1	2-1	32	3-1	62	4-1	93	5-1	124	6-1	152
1-2	2	2-2	33	3-2	63	4-2	94	5-2	125	6-2	153
1-3	3	2-3	34	3-3	64	4-3	95	5-3	126	6-3	154
1-4	4	2-4	35	3-4	65	4-4	96	5-4	127	6-4	155
1-5	5	2-5	36	3-5	66	4-5	97	5-5	128	6-5	156
1-6	6	2-6	37	3-6	67	4-6	98	5-6	129	6-6	157
1-7	7	2-7	38	3-7	68	4-7	99	5-7	130	6-7	158
1-8	8	2-8	39	3-8	69	4-8	100	5-8	131	6-8	159
1-9	9	2-9	40	3-9	70	4-9	101	5-9	132	6-9	160
1-10	10	2-10	41	3-10	71	4-10	102	5-10	133	6-10	161
1-11	11	2-11	42	3-11	72	4-11	103	5-11	134	6-11	162
1-12	12	2-12	43	3-12	73	4-12	104	5-12	135	6-12	163
1-13	13	2-13	44	3-13	74	4-13	105	5-13	136	6-13	164
1-14	14	2-14	45	3-14	75	4-14	106	5-14	137	6-14	165
1-15	15	2-15	46	3-15	76	4-15	107	5-15	138	6-15	166
1-16	16	2-16	47	3-16	77	4-16	108	5-16	139	6-16	167
1-17	17	2-17	48	3-17	78	4-17	109	5-17	140	6-17	168
1-18	18	2-18	49	3-18	79	4-18	110	5-18	141	6-18	169
1-19	19	2-19	50	3-19	80	4-19	111	5-19	142	6-19	170
1-20	20	2-20	51	3-20	81	4-20	112	5-20	143	6-20	171
1-21	21	2-21	52	3-21	82	4-21	113	5-21	144	6-21	172
1-22	22	2-22	53	3-22	83	4-22	114	5-22	145	6-22	173
1-23	23	2-23	54	3-23	84	4-23	115	5-23	146	6-23	174
1-24	24	2-24	55	3-24	85	4-24	116	5-24	147	6-24	175
1-25	25	2-25	56	3-25	86	4-25	117	5-25	148	6-25	176
1-26	26	2-26	57	3-26	87	4-26	118	5-26	149	6-26	177
1-27	27	2-27	58	3-27	88	4-27	119	5-27	150	6-27	178
1-28	28	2-28	59	3-28	89	4-28	120	5-28	151	6-28	179
1-29	29	2-29	60	3-29	90	4-29	121	5-29	---	6-29	180
1-30	30	2-30	61	3-30	91	4-30	122	5-30	---	6-30	181
1-31	31	2-31	--	3-31	92	4-31	123	5-31	---	6-31	182

TABLE 5 (continued)

CORRESPONDENCE BETWEEN CALENDAR YEAR DATE, WATER YEAR DATE AND DAY INDEX

April		May		June		July		Aug.		Sept.	
Water Year	Day Index	Water Year Date	Day Index	Water Year Date	Day Index	Water Year Date	Day Index	Water Year Date	Day Index	Water Year Date	Day Index
7-1	183	8-1	213	9-1	244	10-1	274	11-1	305	12-1	336
7-2	184	8-2	214	9-2	245	10-2	275	11-2	306	12-2	337
7-3	185	8-3	215	9-3	246	10-3	276	11-3	307	12-3	338
7-4	186	8-4	216	9-4	247	10-4	277	11-4	308	12-4	339
7-5	187	8-5	217	9-5	248	10-5	278	11-5	309	12-5	340
7-6	188	8-6	218	9-6	249	10-6	279	11-6	310	12-6	341
7-7	189	8-7	219	9-7	250	10-7	280	11-7	311	12-7	342
7-8	190	8-8	220	9-8	251	10-8	281	11-8	312	12-8	343
7-9	191	8-9	221	9-9	252	10-9	282	11-9	313	12-9	344
7-10	192	8-10	222	9-10	253	10-10	283	11-10	314	12-10	345
7-11	193	8-11	223	9-11	254	10-11	284	11-11	315	12-11	346
7-12	194	8-12	224	9-12	255	10-12	285	11-12	316	12-12	347
7-13	195	8-13	225	9-13	256	10-13	286	11-13	317	12-13	348
7-14	196	8-14	226	9-14	257	10-14	287	11-14	318	12-14	349
7-15	197	8-15	227	9-15	258	10-15	288	11-15	319	12-15	350
7-16	198	8-16	228	9-16	259	10-16	289	11-16	320	12-16	351
7-17	199	8-17	229	9-17	260	10-17	290	11-17	321	12-17	352
7-18	200	8-18	230	9-18	261	10-18	291	11-18	322	12-18	353
7-19	201	8-19	231	9-19	262	10-19	292	11-19	323	12-19	354
7-20	202	8-20	232	9-20	263	10-20	293	11-20	324	12-20	355
7-21	203	8-21	233	9-21	264	10-21	294	11-21	325	12-21	356
7-22	204	8-22	234	9-22	265	10-22	295	11-22	326	12-22	357
7-23	205	8-23	235	9-23	266	10-23	296	11-23	327	12-23	358
7-24	206	8-24	236	9-24	267	10-24	297	11-24	328	12-24	359
7-25	207	8-25	237	9-25	268	10-25	298	11-25	329	12-25	360
7-26	208	8-26	238	9-26	269	10-26	299	11-26	330	12-26	361
7-27	209	8-27	239	9-27	270	10-27	300	11-27	331	12-27	362
7-28	210	8-28	240	9-28	271	10-28	301	11-28	332	12-28	363
7-29	211	8-29	241	9-29	272	10-29	302	11-29	333	12-29	364
7-30	212	8-30	242	9-30	273	10-30	303	11-30	334	12-30	365
7-31	---	8-31	243	9-31	---	10-31	304	11-31	335	-----	---

DATA PROCESSING AND ANALYSIS

The statistical techniques described in Chapter IV will be applied to the 10 watersheds selected in Chapter III.

5.1 Characteristics of the daily flow time series. The mean $\hat{P}(t)$ and the standard deviation $\hat{S}(t)$ for a given day (t) were computed for every day and for every station. Sets in Fig. 3 show the results for $\hat{P}(t)$ and $\hat{S}(t)$ plotted as a function of t for each station. All the watersheds have hydrographs very similar in shape. They show a rise in $\hat{P}(t)$ and $\hat{S}(t)$ during the spring season corresponding to the snowmelt with a decline beginning in June and ending in August which leads to a slowly decreasing or steady flow of small amplitude for the winter season. It corresponds to the time when the watershed is covered with snow and the stream is ice-packed. The coefficients of variation for a given day $C_v(t) = \frac{\hat{S}(t)}{\hat{P}(t)}$ were computed and plotted against t. They show a period of low values from January to June which coincides with the rising limb of the hydrograph. This period of the rising limb, which for other reasons will be selected as the period of study, is also the period with relatively smaller C_v. This constitutes a definite advantage for the purpose of detection.

5.2 Autocorrelation analysis. The autocorrelation r(k) for the 10 stations was computed for every day and for different values of k varying from 1 to 37. The results are shown on Figs. 4 and 5 for stations 12 and 30.

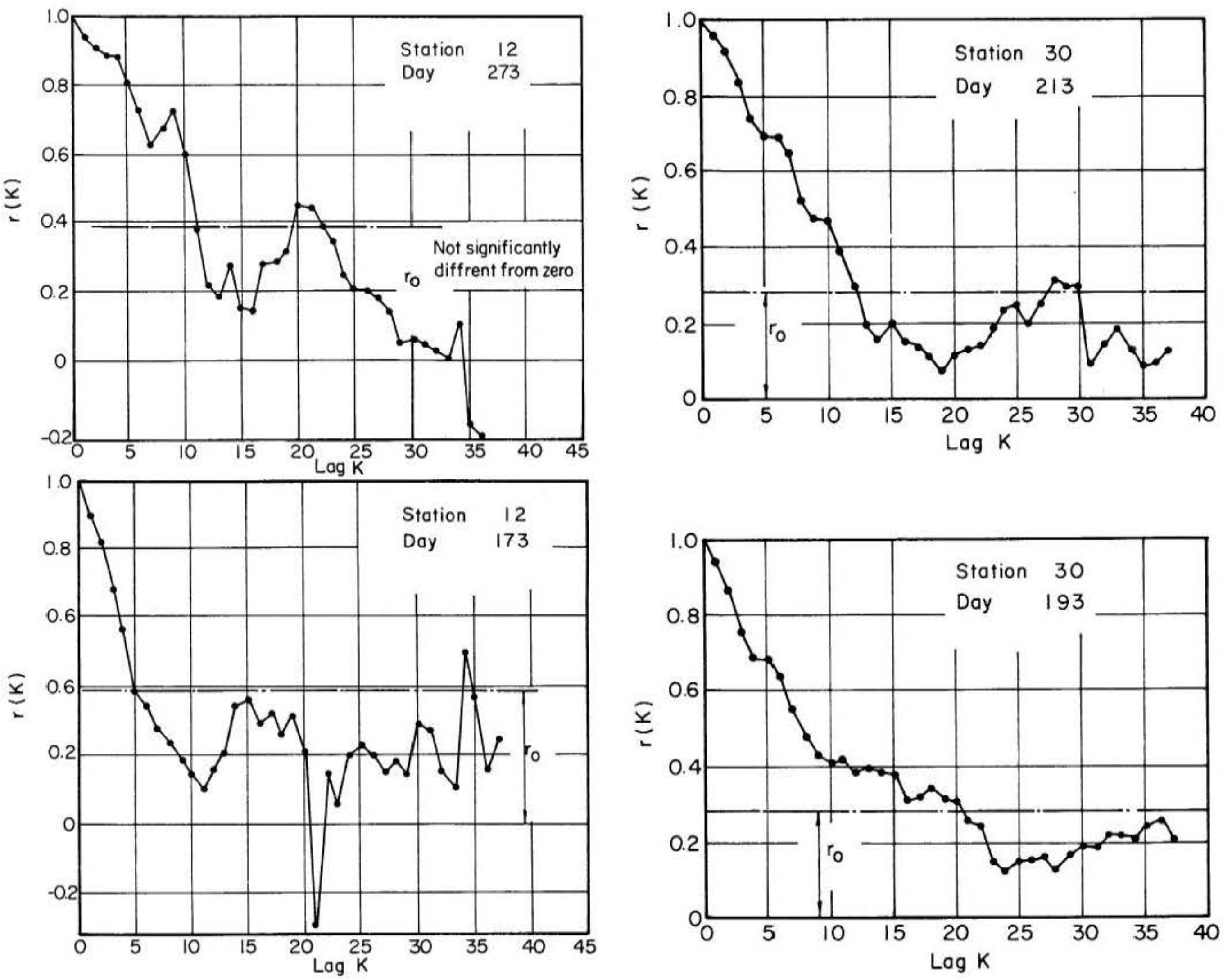


Figure 5 An illustration of r(k) versus k, for a given day and at a given station

Figure 4 shows that r(k) tends to be independent of the days t for the period: March 19th (t = 170) to June 30th (t = 273).

On the basis of this new period, during which the conditions of stationarity are satisfied R(k) was computed for every year (the results of this computation are shown for Station 18 in Table 6-a) then its mean $\overline{R(k)}$ and its variance var [R(k)] (see Table 6-b). This was done for k varying from 1 to 34 by using the formulas described in Chapter IV, where $\alpha = 170$ and $\beta = 273$. Sets in Fig. 6 show the plot of $\overline{R(k)}$ versus k.

An attempt to verify the assumption that consecutive years are independent was made by computing the correlation coefficient between two consecutive years, each day being paired with the same day for the following year, that is, R(k) was computed over two years with k = 365. It showed insignificant correlation.

5.3 Selection of a sequence of independent daily flows. On the basis of the various sets in Fig. 6, a lag common to the 10 stations was selected: K = 20 days. For this lag $\overline{R}(k)$ is considered as nonsignificantly different from zero.

Lag K = 20 yields a sample of seven independent daily flows values for the selected spring period, and for t equal respectively to: 153, 173, 193, 213, 233, 253, and 273 which correspond respectively to March 2, 22; April 11; May 1, 20; and June 10, 30. (As a side line it may be worthwhile to comment somewhat on this apparently highly wasteful procedure. Many daily runoff data are apparently not utilized. Appendix 1 shows that little or nothing would be gained by developing higher order models of the stochastic structure of daily flow for the purpose of detection.)

5.4 Target control correlation. The coefficient of correlation between the stations (one being considered as a target, the other as a control) was computed on the basis of the selected independent daily flow series (Table 7). It was also calculated with other independent daily flow series corresponding to the day-index:

t = 150, 170, 190, 210, 230, 250, 270

that is, for seven days each year corresponding to the dates: February 27; March 19; April 8, 28; May 18; June 7, 27.

As expected, the coefficients of correlation computed in these two manners were not found significantly different. These results are summarized in Table 7.

The results show that the correlations target-control computed with the daily sequences are consistently lower than those computed with the seasonal flows (see Table 3). This is natural because, as the time interval over which the flow is averaged becomes shorter, the watersheds must have very close behaviors to be correlated. In other words, the seasonal flows of two rivers may be correlated, not because the behavior or the patterns of their daily streamflows are exactly the same, but because compensations occur throughout the season, which make their seasonal flows vary in the same way.

It is also interesting to note that two watersheds may be located very far from each other and still have a relatively strong correlation between their streamflows. This is true of stations 30 and 18 (correlation: 0.80) or 12 and 22 (correlation: 0.73). This is an encouraging result for weather modification detection purposes, because in seeding operations one does not want the control watershed to be contaminated.

Table 7-a shows:

1 pair of stations with correlation higher than 0.8.

10 pairs of stations with correlation higher than 0.7.

25 pairs of stations with correlation higher than 0.6.

TABLE 6-a AN ILLUSTRATION OF THE RESULTS FOR R(k) VERSUS lag K, FOR THE 21 YEARS OF RECORDS (M = 1,2,...21) OF STATION 18
Results for $\overline{R(k)}$ versus K: (RBAR) and variance [R(k)]: VARR

Lag K:	1	2	3	4	5	6	7	8	9	10	11
M = 1	.89117	.79937	.72617	.67208	.62119	.52429	.39939	.32450	.23784	.18869	.10400
M = 2	.90524	.75891	.61607	.49990	.38523	.26144	.14073	.03643	-.03271	-.09594	-.11508
M = 3	.96603	.91948	.88975	.86393	.82367	.78994	.76262	.73106	.69323	.61876	.54268
M = 4	.92754	.81227	.71462	.62888	.55847	.49730	.44267	.39597	.36363	.31572	.24389
M = 5	.91464	.78240	.68440	.63731	.61105	.55869	.49321	.44984	.41670	.35221	.25898
M = 6	.93453	.84195	.75075	.66109	.59460	.51760	.43596	.36974	.32826	.28143	.23170
M = 7	.90636	.81200	.72884	.66020	.57004	.47908	.44629	.42808	.37459	.29908	.22691
M = 8	.96002	.88670	.81319	.75222	.71446	.68337	.65451	.61628	.56563	.51144	.46377
M = 9	.80001	.54861	.38941	.26301	.27640	.33259	.30671	.29953	.34472	.33942	.30106
M = 10	.89944	.74780	.62635	.55672	.55071	.56733	.57457	.54087	.51272	.51965	.52369
M = 11	.91035	.77985	.61044	.41922	.23392	.06886	-.05557	-.12137	-.15254	-.15905	-.13583
M = 12	.85875	.63606	.47568	.42163	.41935	.36431	.28297	.23562	.23465	.25437	.22512
M = 13	.81742	.57140	.48592	.45930	.40385	.34745	.30957	.27403	.26608	.29073	.30285
M = 14	.85671	.66483	.53293	.48257	.46408	.42905	.30189	.21606	.12727	.05396	-.02975
M = 15	.91943	.80444	.72939	.67484	.63542	.59070	.53643	.50372	.44925	.40649	.39105
M = 16	.88110	.68996	.50379	.35706	.26836	.24631	.27572	.28289	.25899	.18409	.08928
M = 17	.92126	.81891	.72869	.64028	.55918	.47548	.37984	.27726	.19115	.11774	.06340
M = 18	.91409	.79255	.69596	.62263	.58797	.53088	.43827	.33713	.22244	.15960	.15279
M = 19	.76101	.46935	.32371	.30880	.32115	.22494	.08901	-.02586	-.13716	-.23384	-.26185
M = 20	.74451	.38231	.11296	.04305	.05559	.05693	-.05124	-.18334	-.22566	-.20471	-.12980
M = 21	.98410	.96769	.95483	.94240	.92753	.90870	.88378	.85961	.84390	.81300	.79428
RBAR	.88922	.73747	.62352	.55082	.50392	.45025	.38321	.32610	.28014	.23871	.20205
VARR	.00401	.02160	.03814	.04297	.04219	.04529	.05600	.06693	.07239	.07219	.06696

TABLE 6-a (continued) AN ILLUSTRATION OF THE RESULTS FOR R(k) VERSUS lag K, FOR THE 21 YEARS OF RECORDS

Lag K:	12	13	14	15	16	17	18	19	20	21	22
M = 1	.02636	-.05298	-.11246	-.17555	-.21631	-.23673	-.21471	-.19177	-.17608	-.17118	-.14743
M = 2	-.14165	-.14931	-.12890	-.11556	-.08837	-.11351	-.15021	-.20035	-.22111	-.24113	-.24307
M = 3	.47701	.42241	.37246	.32038	.27593	.23933	.19057	.13539	.09473	.04852	.01030
M = 4	.17563	.11741	.07364	.06429	.06240	.04710	.04054	.05385	.04356	.00609	-.03149
M = 5	.16670	.11052	.07568	.03918	-.02344	-.08523	-.13617	-.17762	-.19595	-.20260	-.22377
M = 6	.17652	.10547	.02440	-.05971	-.14344	-.20901	-.28521	-.34905	-.41825	-.46985	-.50766
M = 7	.19218	.18114	.14147	.15510	.17573	.16488	.17566	.21600	.29016	.33224	.32103
M = 8	.42606	.40258	.39406	.39494	.38655	.35984	.31645	.27330	.23205	.21114	.21862
M = 9	.20128	.08378	.02543	.03021	.04309	.12469	.13488	.05396	.01611	-.02520	-.06084
M = 10	.50572	.47316	.41388	.38973	.39202	.40856	.40925	.40038	.39872	.40453	.43267
M = 11	-.11441	-.11157	-.11898	-.14492	-.16614	-.18693	-.22743	-.28529	-.34666	-.40023	-.44431
M = 12	.19880	.16743	.16700	.17591	.16386	.11321	.06829	.05003	.01593	-.01936	-.10552
M = 13	.35424	.37794	.34807	.26382	.20896	.18745	.22802	.28359	.28647	.27704	.30786
M = 14	-.08456	-.10815	-.07705	-.12534	-.17395	-.25583	-.27164	-.25830	-.23933	-.33090	-.36186
M = 15	.36108	.31680	.27877	.22587	.17981	.13642	.12044	.14705	.11530	.08606	.06845
M = 16	-.02538	-.11167	-.14545	-.14284	-.10607	-.05761	-.01444	.00593	-.02569	.03110	-.04483
M = 17	-.00086	-.05334	-.10133	-.15232	-.17031	-.18037	-.18620	-.20469	-.23685	-.25430	-.24794
M = 18	.16564	.16915	.13147	.05379	.02938	.02810	.04217	.06041	.06555	.02513	-.02745
M = 19	-.25736	-.28230	-.33315	-.37137	-.33717	-.28183	-.22606	-.17636	-.17232	-.15776	-.11343
M = 20	-.06783	-.03726	-.04060	.00874	.06611	.07374	.04236	.07578	.22489	.38648	.39604
M = 21	.77954	.75292	.73226	.71325	.70172	.66523	.64098	.61830	.58369	.54479	.49254
RBAR	.16737	.13210	.10099	.07370	.06002	.04483	.03322	.02526	.01839	.00384	-.01059
VARR	.06434	.06436	.06286	.06282	.06141	.06036	.05986	.06200	.06747	.07811	.08223

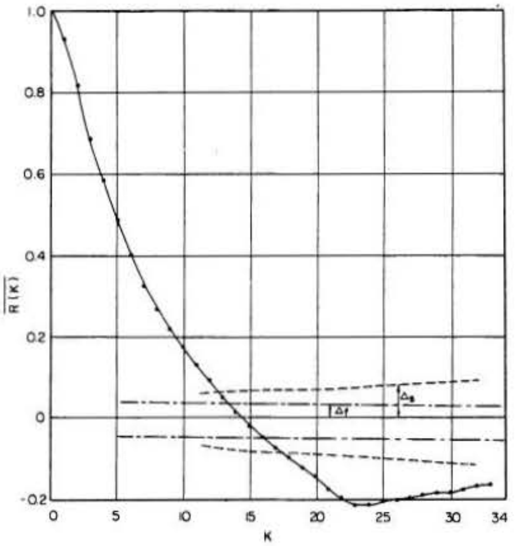
Lag K:	23	24	25	26	27	28	29	30	31	32	33	34
M = 1	-.10804	-.03683	-.02529	-.02356	-.02903	-.00436	.05969	.10778	.09356	.08006	.08370	.13270
M = 2	-.23624	-.19732	-.15065	-.13812	-.11788	-.09185	-.05961	-.01285	.05163	.12691	.20808	.26414
M = 3	-.01373	-.06279	-.10501	-.12037	-.13134	-.14957	-.17320	-.18814	-.20201	-.18903	-.16157	-.13156
M = 4	-.04119	-.02889	.01508	.07105	.11318	.13426	.14864	.12587	.04194	-.06188	-.15377	-.20057
M = 5	-.24257	-.28319	-.30709	-.27957	-.23696	-.22582	-.21834	-.20415	-.20454	-.20017	-.18905	-.18414
M = 6	-.53902	-.55129	-.55726	-.56453	-.56082	-.54343	-.51790	-.48057	-.45860	-.44648	-.41193	-.35784
M = 7	.34637	.42083	.48157	.47979	.46805	.48627	.47310	.41858	.30785	.24187	.22661	.22485
M = 8	.23459	.24496	.24556	.22116	.18515	.15648	.14418	.14746	.13393	.11851	.08649	.04930
M = 9	-.03173	-.01543	-.03343	-.09158	-.14039	-.19874	-.21377	-.17641	-.18764	-.18630	-.18250	-.16438
M = 10	.48194	.52955	.54831	.51110	.45794	.42513	.35282	.29284	.23573	.17705	.13359	.09802
M = 11	-.45682	-.47145	-.48069	-.45452	-.41760	-.37217	-.29981	-.22825	-.14346	-.07513	-.03699	-.00065
M = 12	-.15797	-.14357	-.06249	-.01775	-.03768	-.03846	-.05177	-.07547	-.09943	-.11423	-.12765	-.13546
M = 13	.27334	.21477	.18601	.12114	.10019	.06256	.00965	-.05939	.01626	.09864	.07083	.03877
M = 14	-.39652	-.37531	-.37206	-.41332	-.43387	-.37049	-.36523	-.32191	-.29357	-.23012	-.12634	-.01203
M = 15	.02032	.02047	.05243	.03950	.05149	.06863	.03259	.01678	.00255	-.03061	-.05424	-.12832
M = 16	.05825	.08370	.10762	.12920	.15651	.17173	.17819	.16264	.13891	.10123	.07441	.05590
M = 17	-.24362	-.23663	-.24090	-.24043	-.19223	-.14693	-.11568	-.13541	-.15880	-.18174	-.19746	-.21038
M = 18	-.06450	-.06133	-.03507	-.01071	.01300	-.03005	-.07322	-.08353	-.07979	-.10426	-.10644	-.13817
M = 19	-.06926	-.04528	-.00376	.03137	.12070	.21749	.32157	.41385	.43651	.39275	.38123	.36663
M = 20	.24779	.05905	-.03367	-.07535	-.07208	-.05850	-.03879	.04298	.03439	-.05870	-.14157	-.16265
M = 21	.42125	.31482	.24370	.17144	.08468	.00269	-.07125	-.11816	-.17381	-.20034	-.23270	-.24164
RBAR	-.02464	-.02958	-.02510	-.03115	-.02948	-.02405	-.02369	-.01693	-.02445	-.03533	-.04082	-.03988
VARR	.07954	.07651	.07812	.07345	.06726	.06264	.05829	.05238	.04434	.03741	.03487	.03426

TABLE 6-b

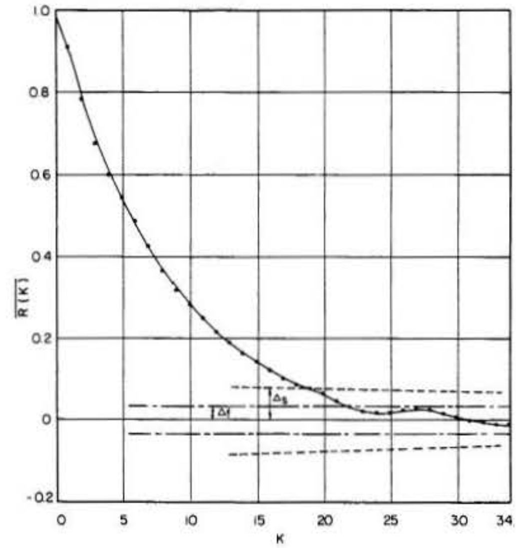
AUTOCORRELATION R(k)

Values of its Mean $\overline{R(k)}$ and of its Variance Var [R(k)] versus k

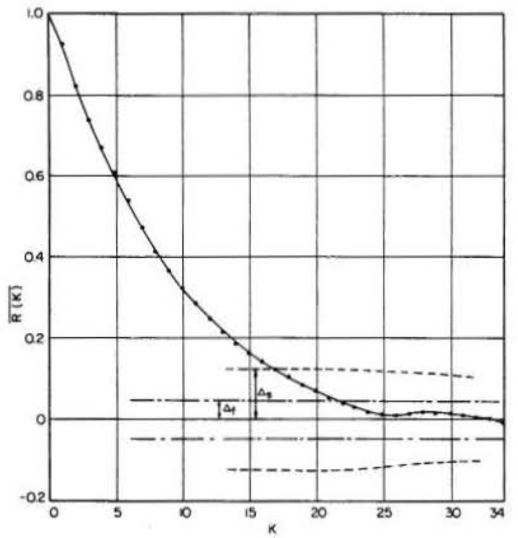
Station	12		16		17		18		19		21		22		25		26		30	
Lag K days	$\overline{R(k)}$	Var(R)	$\overline{R(k)}$	Var(R)	$\overline{R(k)}$	Var(R)	$\overline{R(k)}$	Var(R)	$\overline{R(k)}$	Var(R)	$\overline{R(k)}$	Var(R)	$\overline{R(k)}$	Var(R)	$\overline{R(k)}$	Var(R)	$\overline{R(k)}$	Var(R)	$\overline{R(k)}$	Var(R)
1	.931	.0006	.912	.002	.924	.0014	.889	.004	.940	.0008	.954	.0006	.927	.001	.916	.001	.897	.003	.923	.002
2	.809	.0037	.786	.013	.823	.008	.737	.021	.842	.005	.869	.004	.793	.009	.795	.008	.775	.008	.806	.009
3	.689	.008	.679	.026	.739	.017	.623	.038	.742	.012	.785	.011	.661	.019	.694	.018	.661	.018	.699	.020
4	.584	.011	.602	.033	.670	.025	.551	.043	.652	.020	.710	.019	.550	.025	.613	.027	.568	.027	.607	.030
5	.490	.001	.545	.034	.607	.033	.504	.042	.573	.028	.645	.026	.450	.030	.545	.036	.492	.037	.531	.037
6	.404	.015	.489	.038	.540	.042	.450	.046	.502	.035	.585	.032	.360	.035	.479	.047	.433	.046	.466	.042
7	.330	.020	.427	.046	.474	.049	.383	.056	.439	.040	.534	.035	.286	.039	.413	.058	.377	.058	.407	.046
8	.272	.024	.369	.053	.416	.053	.326	.067	.385	.045	.493	.037	.227	.046	.354	.067	.327	.069	.355	.049
9	.224	.027	.320	.059	.368	.059	.280	.072	.336	.048	.456	.036	.178	.051	.305	.072	.284	.076	.310	.052
10	.180	.029	.282	.064	.325	.061	.239	.072	.290	.054	.422	.037	.135	.051	.264	.070	.248	.080	.271	.054
11	.138	.028	.250	.065	.288	.066	.202	.067	.245	.059	.384	.041	.101	.048	.229	.063	.220	.081	.238	.055
12	.100	.028	.219	.064	.251	.070	.167	.064	.206	.065	.345	.047	.073	.047	.196	.060	.189	.085	.209	.055
13	.059	.031	.191	.063	.215	.072	.132	.064	.171	.069	.302	.056	.054	.046	.163	.059	.161	.090	.182	.057
14	.021	.034	.166	.063	.186	.072	.101	.063	.137	.069	.259	.065	.040	.046	.135	.059	.136	.099	.158	.059
15	-.011	.036	.146	.063	.163	.071	.074	.063	.107	.067	.220	.075	.029	.046	.113	.061	.117	.107	.136	.061
16	-.040	.033	.121	.064	.141	.071	.060	.061	.080	.063	.184	.085	.016	.047	.102	.063	.100	.113	.112	.063
17	-.064	.030	.100	.066	.123	.068	.045	.060	.053	.061	.152	.094	-.002	.046	.092	.066	.086	.115	.089	.064
18	-.088	.029	.087	.065	.106	.065	.033	.060	.030	.059	.125	.101	-.021	.046	.081	.068	.075	.112	.064	.066
19	-.110	.031	.079	.059	.089	.064	.025	.062	.010	.061	.093	.109	-.042	.047	.069	.070	.065	.108	.041	.070
20	-.134	.036	.066	.054	.071	.064	.018	.067	-.009	.061	.061	.113	-.063	.046	.056	.071	.054	.100	.019	.074
21	-.164	.039	.049	.052	.055	.066	.004	.078	-.030	.062	.030	.114	-.083	.043	.045	.073	.045	.092	-.001	.075
22	-.190	.042	.032	.050	.043	.068	-.010	.082	-.051	.061	.009	.114	-.098	.039	.034	.076	.038	.087	-.014	.074
23	-.202	.042	.020	.048	.031	.070	-.025	.079	-.067	.060	-.006	.117	-.108	.036	.027	.080	.031	.084	-.023	.072
24	-.201	.044	.017	.045	.020	.069	-.029	.076	-.077	.060	-.018	.124	-.116	.034	.035	.079	.031	.082	-.023	.071
25	-.196	.048	.019	.043	.015	.067	-.025	.078	-.084	.060	-.030	.130	-.121	.035	.044	.080	.031	.080	-.032	.071
26	-.193	.051	.022	.042	.013	.061	-.031	.073	-.091	.060	-.043	.134	-.125	.037	.050	.081	.031	.077	-.037	.071
27	-.187	.055	.024	.042	.016	.057	-.029	.067	-.100	.058	-.056	.137	-.127	.038	.053	.080	.029	.076	-.041	.070
28	-.181	.059	.024	.044	.020	.053	-.024	.062	-.111	.057	-.068	.139	-.125	.039	.057	.079	.025	.078	-.044	.069
29	-.178	.062	.016	.046	.019	.051	-.024	.058	-.121	.056	-.077	.139	-.119	.038	.062	.075	.024	.081	-.045	.068
30	-.174	.060	.007	.046	.018	.050	-.017	.052	-.129	.055	-.080	.137	-.113	.036	.061	.072	.027	.086	-.041	.068
31	-.171	.057	-.001	.048	.011	.050	-.024	.044	-.134	.052	-.077	.136	-.110	.034	.048	.065	.021	.088	-.032	.070
32	-.162	.053	-.008	.052	.006	.052	-.035	.037	-.141	.052	-.071	.131	-.114	.033	.031	.059	.019	.087	-.028	.070
33	-.155	.051	-.012	.056	.002	.055	-.041	.035	-.143	.053	-.063	.123	-.113	.033	.016	.054	.013	.085	-.022	.072
34	-.152	.048	-.016	.063	-.004	.059	-.040	.034	-.142	.054	-.060	.118	-.109	.034	-.002	.053	.002	.084	-.016	.074



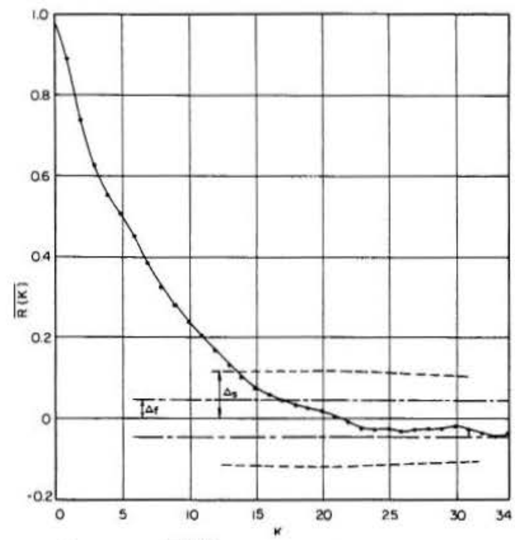
Correlogram: $\overline{R(k)}$ versus k for station 12



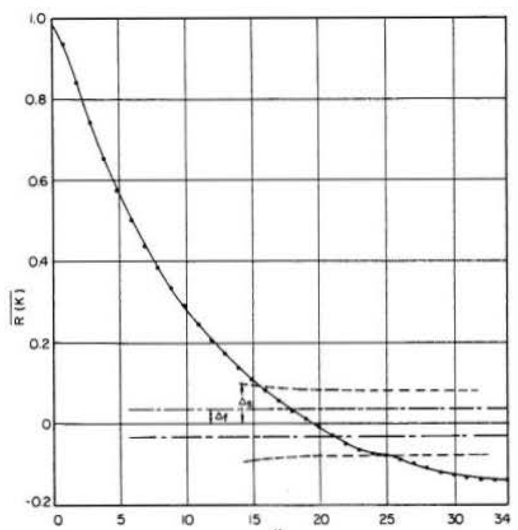
Correlogram: $\overline{R(k)}$ versus k for station 16



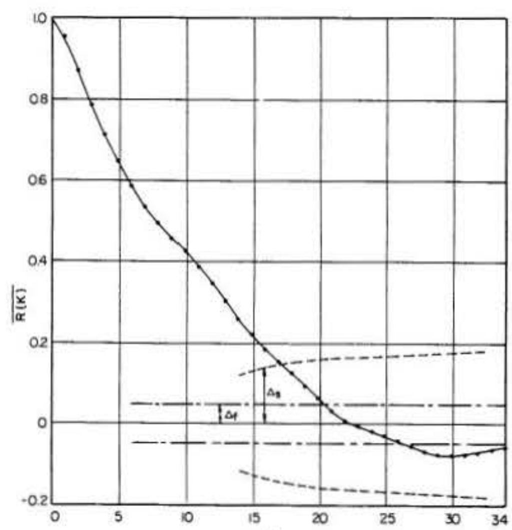
Correlogram: $\overline{R(k)}$ versus k for station 17



Correlogram: $\overline{R(k)}$ versus k for station 18

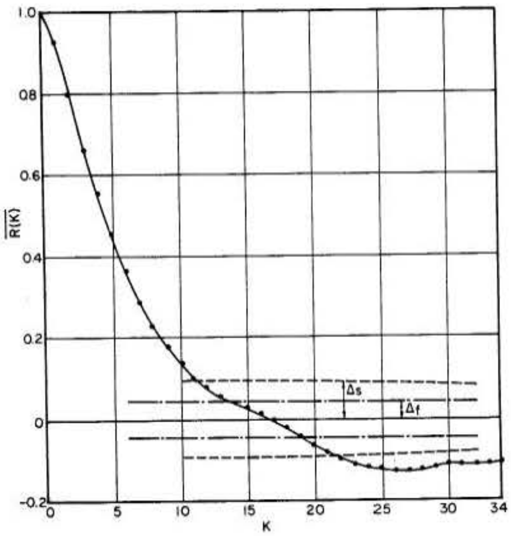


Correlogram: $\overline{R(k)}$ versus k for station 19

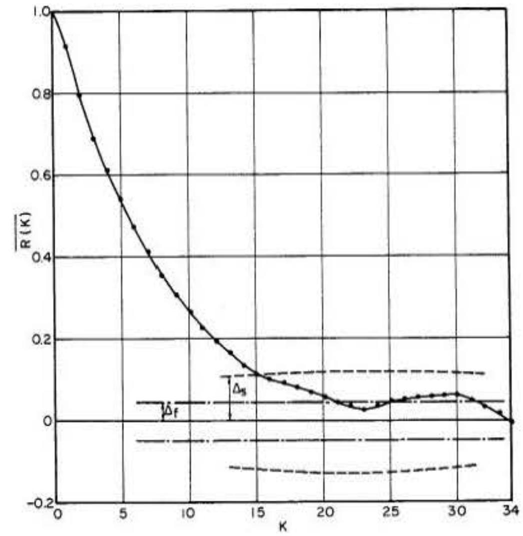


Correlogram: $\overline{R(k)}$ versus k for station 21

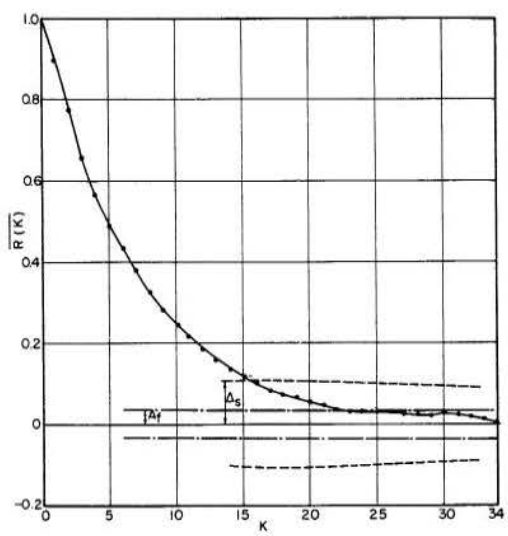
Figure 6



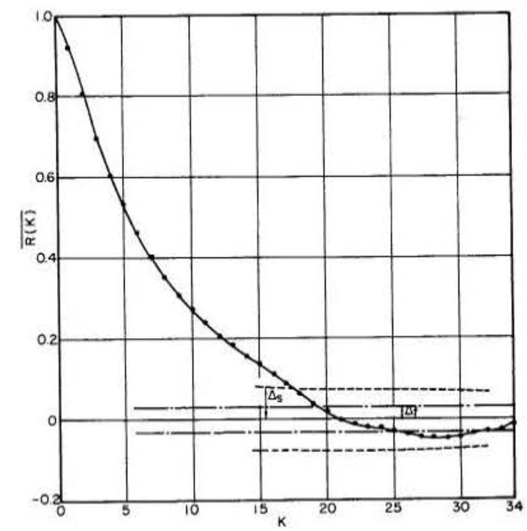
Correlogram: $\overline{R(k)}$ versus k for station 22



Correlogram: $\overline{R(k)}$ versus k for station 25



Correlogram: $\overline{R(k)}$ versus k for station 26



Correlogram: $\overline{R(k)}$ versus k for station 30

Figure 6 (continued)

TABLE 7

TARGET-CONTROL CORRELATION ON THE BASIS OF DAILY FLOWS

[a] Computed with the series t : 153, 173, 193, 213, 233, 253, 273

Identification	12	16	17	18	19	21	22	25	26	30
12										
16	.566									
17	.393	.523								
18	.710	.806	.549							
19	.798	.627	.431	.637						
21	.618	.674	.470	.732	.644					
22	.730	.553	.385	.576	.761	.676				
25	.632	.560	.550	.629	.614	.510	.489			
26	.495	.621	.506	.602	.502	.503	.433	.616		
30	.701	.666	.552	.796	.677	.722	.720	.667	.541	

TABLE 7 (continued)

TARGET-CONTROL CORRELATION ON THE BASIS OF DAILY FLOWS

[b] Computed with the series t : 150, 170, 190, 210, 230, 250, 270

Identification	12	16	17	18	19	21	22	25	26	30
12										
16	.615									
17	.463	.557								
18	.696	.808	.471							
19	.740	.625	.503	.598						
21	.563	.623	.408	.604	.557					
22	.712	.620	.463	.625	.737	.590				
25	.641	.656	.641	.658	.683	.436	.592			
26	.552	.641	.582	.646	.498	.495	.511	.618		
30	.668	.659	.599	.727	.604	.681	.697	.680	.547	

Chapter VI

DETECTION OF THE SUSPECTED CHANGE IN RUNOFF

In summary the theoretical analysis of the previous chapters and its application to actual records of daily runoffs at several gage stations have shown that by considering a set of days, whose adjacent elements are lagged by $K = 20$ days, ν independent standardized values of daily flow can be selected within each year, more precisely within each spring season. The lag of 20 days is common to all investigated stations. In other words a random function is selected whose ν ordinates are statistically independent; each ordinate being a random variable with zero mean and unit variance whose distribution is approximately normal.

Therefore, these ν ordinates can be considered as ν independent values of the same normally distributed random variable y. Then in n years of historical records there are $N = \nu n$ independent values of a random variable y whose probability distribution is the standard normal distribution.

During the period of the suspected change the actual daily runoff observations for the corresponding dates provide a new sequence. One suspects that this sequence belongs to a different statistical population than the previous or historical one.

If this is the case and if the new set of data is standardized, according to eq. (4), where the estimates of $P(t)$ and $S(t)$ are the ones obtained based solely on the historical records, the new (historically) standardized daily flow sequence will no longer have zero mean and unit variance. The application of the target-control test will tell whether the change is significant or not. If data are available, grinding the answer from the programmed test subroutine is all that is left to do. On the other hand, if experiments are contemplated for the future and data therefore are not yet available, the required duration of the experiments can be inferred from a randomly generated sequence of daily flows. Of course, the data generation implies a model of what is likely to happen, based on an understanding of the physical phenomena and available experimental evidence.

6.1 Model for the effects of seeding. In the following it is assumed:

(a) Cloud seeding operations increase the values of the streamflows, and, more precisely, they increase the mean daily values $P(t)$.

(b) They do not affect the variance $S^2(t)$ of the daily flows.

(c) The relative increase, h, due to artificial precipitation is independent of time at least throughout the spring season (March to June).

These assumptions are more likely to be correct for cloud seeding operations taking place in winter above watersheds of high elevations--these operations attempt to increase the snowpack and only affect the streamflow during the melting season. The Bureau of Reclamation's pilot project in the Upper Colorado River Basin fits this category.

With the above assumptions, a value of daily flow $Q^*(t)$, affected by cloud seeding experiments, would have the form:

$$Q^*(t) = (1+h)P(t) + S(t)y(t)$$

where h is the relative increase in the mean daily value due to cloud seeding. The historical standardization of $Q^*(t)$ will give $\eta(t)$:

$$\eta(t) = \frac{Q^*(t) - \hat{P}(t)}{\hat{S}(t)} = h\,\frac{\hat{P}(t)}{\hat{S}(t)} + y(t)$$

where $y(t)$ would be the standardized value of $Q^*(t)$ if no increase h, due to cloud seeding, had happened. It follows that $y(t)$ is normally distributed with mean zero and variance unity and we have for a given t:

$E[\eta(t)] = E[h\,\frac{\hat{P}(t)}{\hat{S}(t)}]$, different from zero if h is different from zero.

It is assumed that artificial precipitation has not increased the statistical dependence between daily flows, and that adjacent daily flows, separated by the previously selected lag time k, can be considered as independent. Then, for m years of seeding experiments, $M = m\nu$ independent values of a random variable, η can be selected, whose distribution is assumed to be normal with mean: $E[h\,\frac{\hat{P}(t)}{\hat{S}(t)}]$, where t can take ν values. It should be noted that if $h\,\frac{\hat{P}(t)}{\hat{S}(t)}$ is constant for any of the t for the selected η values, then $\mathrm{var}\,[\eta] \approx \mathrm{var}\,[y] \approx 1$. The fluctuation of $h\,\frac{\hat{P}(t)}{\hat{S}(t)}$ with t, being small during the spring season, it can be assumed without much error that $\mathrm{var}\,[\eta] = \mathrm{var}\,[y]$.

6.2 Generation of seeded data. Monte Carlo Method. According to the general model for the seeded period, the variable for the control watersheds is $\xi(t)$ and is normally distributed with mean zero and variance unity. The variable for the target watershed is $\eta^*(t)$ such that:

$$\eta^*(t) = \eta(t) + h\,\frac{\hat{P}(t)}{\hat{S}(t)}$$

and $\eta(t)$ is correlated to $\xi(t)$ by the regression line obtained for the non-seeded period:

$$\eta(t) = b\xi(t) + \varepsilon(t) \;;$$

where b is the estimate of the slope of the regression line and $\varepsilon(t)$ is the random deviation of $\eta(t)$ about its estimate by the regression line. The joint distribution of (η,ξ) being assumed bivariate normal, $\varepsilon(t)$ is normally distributed around zero with variance $(1-\rho^2)$ $var[\eta(t)]$ where ρ is the correlation coefficient between ξ and η, then:

$$\eta^*(t) = b\xi(t) + \varepsilon(t) + h\,\frac{\hat{P}(t)}{\hat{S}(t)}$$

To generate data, ρ and b must be calculated for the pair of considered watersheds. Then independent random values are drawn

(a) For $\xi(t)$ from a normal population with mean zero and variance unity.

(b) For $\varepsilon(t)$ from a normal population with mean zero and variance $(1-\rho^2)$.

This was done with the computer CDC 6400 at the University. A subprogram, "Function Ranf" has been written by the University computer center to generate random numbers between 0 and 1, with a uniform density. (The procedure for transformation of this uniform density onto a normal one is described in Appendix 2.)

6.3 Results of the Student-t Test. According to the formulas and derivations given previously, the test was performed for the 10 pairs of stations with correlation coefficient larger than 0.70. For the seeded period the number of years was increased from 1 to 20 until significance at the 95% level (corresponding to a two-tailed test) was reached.

Results are also given for some stations at the 98% and 99% level for a two-tailed test, and at the 95% level for a one-tailed test. The results show (Table 8):

(a) Almost identical results are obtained by using a one-tailed test and a two-tailed test at the 95% level.

(b) Sometimes the same pair of watersheds shows very different results when their status of target and control is permuted. For example: Pair 16-18, with 16 as a target, required one year to show significance but pair 16-18, with 18 as a target, required more than 20 years. This could be because station 18 may not be suitable for a target. According to the way that data for the seeded period were generated, a watershed is suitable as a target if the ratios $\frac{P(t)}{S(t)}$ are large, for the seven selected days, in other words if the $C_v(t) = \frac{S(t)}{P(t)}$ are small. The coefficients of variation $C_v(t)$ were in fact smaller for station 16 than for station 18, but their ratios were much smaller than $\sqrt{20}$, which is the square root of the ratio of the required number of years for significance. Therefore, the differences in the coefficients of variation is not sufficient to explain the difference in the required number of years for significance. A more likely explanation lies in the paucity of the generated random data. In each case only one sequence of data was generated. Therefore, no power value can be attributed to the calculated number of years. It should be also noted from Table 2 that station 18 has probably the least reliable record of all.

(c) The consistency of the results for station 30, paired successively with a different control, is an encouraging result. It was somewhat expected, since the correlation coefficients between station 30 and these control watersheds are of the same order of magnitude. On the other hand, station 12, used as a target successively with a different control, shows great inconsistencies.

TABLE 8

RESULTS OF THE STUDENT t TEST FOR THE DETECTION OF A 10% INCREASE IN THE "DAILY MEANS"

Identification Target	Identification Control	Number of years in common for the non-seeded period	2-tailed 95% level N(yrs)	2-tailed 95% level t	2-tailed 98% level N(yrs)	2-tailed 98% level t	2-tailed 99% level N(yrs)	2-tailed 99% level t	1-tailed test 95% level N
12	18	21	5	1.96					
12	19	25	G.T.20						
12	22	23	3	2.74	3	2.74	3	2.74	2
12	30	25	10	1.95					
16	18	21	1	2.58	1	2.58	3	2.75	1
18	12	21	5	2.15					5
18	16	21	G.T.20						
18	21	20	G.T.20						
18	30	21	3	2.68	3	2.68	3	2.68	3
19	12	25	1	3.68	1	3.68	1	3.68	1
19	22	23	2	3.48	2	3.48	2	3.48	2
21	18	20	1	3.44	1	3.44	1	3.44	1
21	30	20	10	1.99					9
22	12	23	2	2.28	4	3.21	4	3.21	2
22	19	23	G.T.20						
22	30	23	8	2.43	8	2.43			
30	12	25	6	2.10	7	2.58	8	2.92	6
30	18	21	4	2.38	4	2.38			3
30	21	20	4	2.37	4	2.37			4
30	22	23	4	2.02					

G.T. means greater than

No attempt was made to transform the data prior to the application of the test, because the Student-t test has been shown [11] to be "robust." In other words, the fact that the joint bivariate distribution of the target and control population may not be normal does not affect the test significantly. To sum up the results:

Twelve stations among 20 required five years or less for detection at the 95% level and seven of them required five years or less for detection at the 99% level.

Only six stations among 20 required 10 years or more for detection at the 95% level.

On the basis of the following formula derived from a Chi-square test:

$$M = 4(1-\rho^2)\,\frac{C_{v,T}^{2}}{h^2} \quad \text{(already given in Chapter I),}$$

and using seasonal flows as variables, the number of years M required to detect a h = 10% increase in the mean seasonal flows at the 95% level and 50% power was computed. Computations were made by using the correlation coefficient ρ between target and control for the six month period (March-August), and the coefficient of variation of the target $C_{v,T}$ for the six month period, then for the four month period (April-July). Results are shown in Table 9.

For 15 stations among 20, the use of daily flows reduced the number of years required for detection in a very significant manner (by an average factor of five over 14 studied cases).

For only three stations out of 20 the use of daily flows was found to be a disadvantage.

TABLE 9

NUMBER OF YEARS REQUIRED FOR THE DETECTION OF A 10% INCREASE IN THE MEANS AT THE 95% LEVEL

Identification Target	Identification Control	Correlation coefficient with Daily flows	Correlation coefficient with Seasonal flows	Target coefficient of variation 4 months period	Target coefficient of variation 6 months period	Number of years for significance using 4 months Seasonal flows M4(yr)	Number of years for significance using 6 months Seasonal flows M6(yr)	Number of years for significance using Daily flows Md(yr)
12	18	.710	.728	.246	.255	11	12	5
12	19	.798	.940	.246	.255	3	3	G.T.20
12	22	.730	.807	.246	.255	8	9	3
12	30	.701	.785	.246	.255	8	10	10
16	18	.806	.969	.515	.504	6	6	1
18	12	.710	.728	.575	.537	62	54	5
18	16	.806	.969	.575	.537	8	7	G.T.20
18	21	.732	.811	.575	.537	45	39	G.T.20
18	30	.796	.877	.575	.537	30	27	3
19	12	.798	.940	.313	.312	4	4	1
19	22	.761	.792	.313	.312	15	14	2
21	18	.732	.811	.572	.510	45	35	1
21	30	.722	.848	.572	.510	37	29	10
22	12	.730	.807	.338	.326	16	15	2
22	19	.761	.792	.338	.326	17	16	G.T.20
22	30	.720	.914	.338	.326	8	7	8
30	12	.701	.785	.428	.413	28	26	6
30	18	.796	.877	.428	.413	17	16	4
30	21	.722	.848	.428	.413	20	19	4
30	22	.720	.914	.428	.413	12	11	4

6 months: March-August
4 months: April-July
G.T. means greater than.

Chapter VII

CONCLUSIONS

1. The standardization of the daily streamflows time-series did not provide stationarity in the wide sense, except for the spring period.

2. The watersheds under study had nearly identical hydrologic features, particularly the same hydrographs; as a result very similar correlograms were found for every station. The study was made possible because all the watersheds had the same stochastic structure, which made it possible to select sequences of independent daily flow values at dates and intervals common to every station.

3. The daily flow time-series show strong auto-correlation. Accordingly, only seven days with independent flow values per year could be selected.

4. The correlation between target and control watersheds, computed on the basis of the independent daily flow sequence, was found to be lower than the one computed on the basis of the corresponding seasonal flow. It was also found to be a good tool to select watersheds having the same hydrologic behavior.

5. The application of a Target-Control Student t-test shows that the use of daily flow as a variable instead of seasonal flow, by increasing the size of the sample, tends to reduce significantly the number of years required to detect a 10% increase in the mean flow at the 95% confidence level.

The most desirable properties that characterize the methods of statistical evaluation are applicability, generality, and power.

Applicability and generality remain open questions since this study considered only the case of winter seeding operations above high elevation watersheds. It is felt the method presented in this paper can be extended to different types of basins as long as they are hydrologically homogeneous, and can be used, not only for evaluation of weather modification, but also for the detection of changes in watershed responses, as long as the effects of such changes are changes in the mean flows, leaving the variance unchanged. For instance, effects of forest fires, land slides, and even urbanization could be investigated in a similar manner.

The sharp power of detection seems to be the best quality of the method. No conclusion can be reached for a particular pair of target-control watersheds as to the value of daily versus seasonal flow for minimal time evaluation because only one sequence of data was generated per pair. On the other hand the ensemble of the results shows rather clearly that the potential value of daily runoff as a detector of change in watershed response is high. The factor of reduction from its use averages three for the 20 cases studied. It is highly significant, and fully justifies additional more complete and more refined studies.

LIST OF SYMBOLS

Symbol	Description
$Q(t)$	Random function whose values are the daily flow values
$P(t)$	"Mean for a given day," i.e., mean daily value of $Q(t)$
$S(t)$	Standard deviation of $Q(t)$ for a given t
$q_i(t)$	Standardized daily flow values
t	Index referring to a day within a given year
i	Index referring to a year
n	Number of years of historical record for the non-seeded period
m	Number of years of record for the seeded period
N	Number of data or sample size for the non-seeded period
M	Number of data or sample size for the seeded period
y	Series of independent standardized daily flows for the non-seeded period of the target
η	Series of independent standardized daily flows for the seeded period of the target
x	Series of independent standardized daily flows for the non-seeded period of the control
ξ	Series of independent standardized daily flows for the seeded period of the control
$Q^*(t)$	Daily flow values affected by seeding operations
$r_k(t)$	Correlation coefficient between day t and day $t-k$
$R_i(k)$	Serial correlation coefficient for lag k and for year i
$\overline{R}(k)$	Average of $R_i(k)$ over n realizations
ρ	Target-control correlation coefficient
ν	Number of independent daily flow values for the spring season
b	Slope of the regression between target and control
$\hat{}$	The "hat" over a symbol means: we are considering the sample estimate of a given parameter
*	The "star" next to a symbol means: suspected to come from a different population than in the past.

REFERENCES

1. Plan to develop technology for increasing water yield from atmospheric sources: Office of Atmospheric Water Resources, Bureau of Reclamation, United States Department of the Interior, November 1966.

2. Gilman, Donald L.; Hibbs, James R.; and Paul L. Laskin; Weather and climate modification: U.S. Weather Bureau, July 10, 1965.

3. Saline water conversion, 1963 report: United Stated Department of the Interior, Office of Saline Water.

4. Weather and climate modification, 1966 report: The National Academy of Sciences - National Research Council.

5. Committee on Water of the National Research Council, "Water and Choice in the Colorado River Basin," National Academy of Sciences, Publication 1689, 107 pp., Washington, D.C., 1968.

6. Marlatt, W. and Riehl, H., "Precipitation Regimes Over the Upper Colorado River," Journal of Geophysical Research, Vol. 68, No. 24, pp. 6477-6458, December 1963.

7. Golze, A. R., "Reclamation in the United States," the Caxton Printers, Ltd., Caldwell, Idaho, 1961.

8. Smith, L.W., "Western States Water Augmentation Concept," 1968, Federation of Rocky Mountain States Inc. Suite 203/1390 Logan Street, Denver, Colorado 80203, 37 pp.

9. Hurley, Patrick A., "Augmenting Colorado River by Weather Modification," Journal of the Irrigation and Drainage Division, ASCE, Vol. 94, No. IR4, Proc. Paper 6271, December, 1968, pp. 363-380.

10. Markovic, R.D.,"Statistical Evaluation of Weather Modification Attainments," Hydrology Paper No. 16, Colorado State University, Fort Collins, Colorado, November 1966, 44 pp.

11. Morel-Seytoux, H.J., Final report for FY 1966 and 1967. Project SKYWATER. Hydrology Program Report No. CER68-69HJM5, Part 1, "Suitability of Basins to Weather Modification and Statistical Evaluation of Attainment," Colorado State University, July 1968.

12. Mood, Alexander M. and Graybill, Franklin A., "Introduction to the Theory of Statistics," McGraw-Hill, 1963, 443 pp.

13. Brownlee, K.A.,"Statistical Theory and Methodology in Science and Engineering," New York, John Wiley and Sons, Inc., 1961, 570 pp.

14. Morel-Seytoux, H.J., "Suitability of Basins to Weather Modification and Statistical Evaluation of Attainment," Final Report to the Bureau of Reclamation for FY 1966 and 1967, Part 4, 30 pp. Colorado State University, Fort Collins, Colorado, July 1, 1968.

15. "Reclamation to Conduct Project SKYWATER Program in Colorado Rockies this Winter," U.S. Department of Interior, Bureau of Reclamation, Office of Chief Engineer, Denver Federal Center, News Release September 22, 1968.

16. Dumas, Andre J., "Detection of a Change in Watershed Response by a Stochastic Analysis of Daily Streamflows," Master of Science Thesis, Colorado State University, December 1968, CET68-69AJD25, 79 pp.

17. Papoulis, A.,"Probability, Random Variables and Stochastic Processes,"McGraw-Hill, 1965, 583 pp.

18. Sveshnikov, A. A., "Applied Methods of the Theory of Random Functions," Pergamon Press, Chapter IV, pp. 225-242.

19. Quimpo, R. G., "Stochastic Model of Daily River Flow Sequences," Hydrology Paper No. 18, Colorado State University, Fort Collins, Colorado, February 1967.

20. Roesner, L. A., and Yevjevich, V. M., "Mathematical Models for Time Series of Monthly Precipitation and Monthly Runoff," Hydrology Paper No. 15, Colorado State University, Fort Collins, Colorado, October 1966.

21. World Meteorological Organization, Technical Note No. 71. W.M.O. No. 178TP88, pp. 62-63.

22. Box, G.E.P., and Muller, M.E., "A Note on the Generation of Random Normal Deviates," Annuals of Mathematical Statistics, Vol. 29, 1958, pp. 610-611.

APPENDIX 1

Fitting Markov I Model

The asymptotic behavior of the correlograms suggests the daily flow series could be fitted by a first order linear autoregressive scheme (or Markov first order linear model).

In this model, the correlogram of y(t) can be represented by

$$R(k) = a^k$$

and the autoregressive scheme is given by:

$$y(t) = ay(t-1) + \varepsilon(t)$$

where $\varepsilon(t)$ is independent of $y(t-1)$, $y(t-2)$... and of the other ε's. Then a could be estimated either by

(1) taking $a = R_1$ (empirical value of the first autocorrelation coefficient)

(2) fitting a function $R(k) = a^k$ to the empirical correlogram and estimating the value of a by the method of least square, which could be done by linearizing the exponential function before minimizing the sum of the squared differences. Criterion or a test for goodness of fit then can be used to determine how well the Markov I model will apply [19,20].

If such a model would fit well enough, it can be seen that using the series $\varepsilon(t)$ for the purpose of weather modification detection will yield a very large sample, since the ε's are independent. How would the series $\varepsilon(t)$ be affected by artificial precipitation? For the non-seeded period:

$$\varepsilon(t) = y(t) - ay(t-1)$$

and $\varepsilon(t)$ has a mean equal to zero and a variance:

$$\text{var}[\varepsilon(t)] = (1-a)^2\ \text{var}[y(t)] = 1-a^2 \quad .$$

For the seeded period:

$$\varepsilon^*(t) = y(t) - ay(t-1) + h\left(\frac{P(t)}{S(t)} - a\frac{P(t-1)}{S(t-1)}\right) \quad .$$

If we compare the series $\varepsilon(t)$ with the series y(t) involving seven data points per year, we find:

(a) The expectation of $\varepsilon^*(t)$ is roughly (1-a) times the expectation of $\eta(t)$,

(b) The standard deviation of $\varepsilon^*(t)$ is $\sqrt{1-a^2}$ times the deviation of y(t),

(c) For $\varepsilon(t)$ the number of data points per year is $g = 14.7$ times the one for the series y(t).

We shall assume that the correlation between target and control on the basis of the $\varepsilon(t)$ is the same as the one on the basis of the y(t). In fact it is likely to be much smaller. Denoting the number of necessary years for detection by the series $\varepsilon(t)$ and y(t) respectively by M_ε and M_y and ($\sim$) being the symbol of proportionality, we have

$$\frac{M_\varepsilon}{M_y} \sim \frac{t_{95,\varepsilon}}{t_{95,y}} \cdot \frac{\sqrt{1-a^2}}{(1-a)\sqrt{g}}$$

$$\frac{M_\varepsilon}{M_y} \sim \frac{t_{95,}}{t_{95,y}} \cdot \sqrt{\frac{1+a}{(1-a)g}} \ .$$

The number of degrees of freedom for $t_{95,\varepsilon}$ will be roughly g times the one for the $t_{95,y}$. But since the latter is already a large number, the ratio

$$\frac{t_{95,\varepsilon}}{t_{95,y}}$$

will be very close to one.

With two stations with n years of historical records and five years of seeded period we would have:

$$\frac{t_{95,\varepsilon}}{t_{95,y}} \approx .99 \quad .$$

Taking $a = 0.9$, we would find:

$$\frac{M_\varepsilon}{M_y} \approx .99 \times 1.14 = 1.13 \ .$$

Taking $a = 0.95$, we would find:

$$\frac{M_\varepsilon}{M_y} \approx .99 \times 1.64 = 1.62 \ .$$

Fitting the Markov I model would give values of (a) between 0.9 and 0.95, but in any case it would hardly improve the detection possibilities, whereas it would complicate and greatly expand the computations.

It is interesting to note that in the case where the Markov I model describes well our variable, ν, the number of independent days during a $G = 103$ days period, is given by [21]:

$$\nu = \frac{G}{1 + \frac{2a}{1-a}\left(1 - \frac{1}{G}\frac{1-a^G}{1-a}\right)} \ .$$

With $a = 0.9$ and $G = 103$ we find: $\nu = 6$ days, which is the number of independent days we selected for the 103 days period. (The seventh day is outside this interval for which the conditions of stationarity are met.)

The purpose of this Appendix is to describe a method [22] to generate random numbers from a normal population with high accuracy and favorable speed for the computer.

Let U_1 and U_2 be the independent random variables from the same rectangular density $f(U_1,U_2) = 1$, on the interval [0,1] and consider the random variables defined by:

$$X_1 = (-2 \, \mathrm{Log}_e \, U_1)^{\frac{1}{2}} \, \mathrm{Cos} \, 2\pi U_2 = g_1(U_1,U_2)$$

$$X_2 = (-2 \, \mathrm{Log}_e \, U_1)^{\frac{1}{2}} \, \mathrm{Sin} \, 2\pi U_2 = g_2(U_1,U_2) \quad .$$

We then have:

$$U_1 = e^{\frac{-(x_1^2 + x_2^2)}{2}} = h_1(X_1,X_2)$$

$$U_2 = -\frac{1}{2\pi} \arctan \frac{X_2}{X_1} = h_2(X_1,X_2) \quad .$$

And then we have:

$$P(a \le U_1 < b, \; c \le U_2 < d) = \int_a^b \int_c^d f(U_1,U_2) \, du_1 \, du_2$$

$$= \iint_{(S)} f[h_1(x_1,x_2), \, h_2(x_1,x_2)] \, |J| \, dx_1 \, dx_2$$

where (S) is the domain of the x_1, x_2 plane into which the rectangle ($a \le U_1 < b$, $c \le U_2 < d$) is mapped by the transformation, and

$$J = \begin{vmatrix} \partial h_1/\partial x_1 & \partial h_1/\partial x_2 \\ \partial h_2/\partial x_1 & \partial h_2/\partial x_2 \end{vmatrix} \quad .$$

The density function of the joint distribution for (x_1,x_2) is $\psi(x_1,x_2)$; $\psi(x_1,x_2) = f[h_1(x_1,x_2), h_2(x_1,x_2)]\,|J| = |J|$, since $f[\,] = 1$; and we find

$$\psi(x_1,x_2) = |J| = \frac{1}{2\pi} e^{\frac{-(x_1^2+x_2^2)}{2}} = \frac{1}{\sqrt{2\pi}} e^{\frac{-x_1^2}{2}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x_2^2}{2}}$$

$$= \psi(x_1)\psi(x_2) \; ;$$

x_1 and x_2 are a pair of independent random variables from the same normal population with mean zero and unit variance.

In this way we can draw a set of values of $\xi(t)$ and another set of values that once multiplied by $\sqrt{1-\rho^2}$ will give a set of values for $\varepsilon(t)$.

A fortran program was written for this purpose for different pairs of station and for h = 10%.

Key Words: Statistical discrimination, hydrologic change, daily runoff, precipitation management, evaluation

Abstract: The purpose of this study was the development of a technique for rapid detection of the occurrence of a suspected hydrologic change in high mountain watersheds. A method has been developed that uses a sequence of independent daily flows. This procedure is superior to previous ones based on seasonal or yearly flows. The results of this investigation show the use of daily, instead of seasonal flow, data in a Student t-test reduces the number of necessary years of data for detection by an average of five in 14 out of the 20 cases studied, or by an average of three for the 20 cases. All of the cases come from the Upper Colorado River Basin. The study is particularly relevant to the planned cloud seeding operations of the Bureau of Reclamation in high elevation areas of the Colorado Rocky Mountains. The statistical procedure of detection relies on the Target Control concept and the application of a conditional Student t-test, a test of the difference between the adjusted means obtained by the regression lines between Target and Control for the seeded and non-seeded periods.

References: Andre J. Dumas and Hubert J. Morel-Seytoux, Colorado State University Hydrology Paper No. 34 (August 1969) "Statistical Discrimination of Change in Daily Runoff."

Key Words: Statistical discrimination, hydrologic change, daily runoff, precipitation management, evaluation

Abstract: The purpose of this study was the development of a technique for rapid detection of the occurrence of a suspected hydrologic change in high mountain watersheds. A method has been developed that uses a sequence of independent daily flows. This procedure is superior to previous ones based on seasonal or yearly flows. The results of this investigation show the use of daily, instead of seasonal flow, data in a Student t-test reduces the number of necessary years of data for detection by an average of five in 14 out of the 20 cases studied, or by an average of three for the 20 cases. All of the cases come from the Upper Colorado River Basin. The study is particularly relevant to the planned cloud seeding operations of the Bureau of Reclamation in high elevation areas of the Colorado Rocky Mountains. The statistical procedure of detection relies on the Target Control concept and the application of a conditional Student t-test, a test of the difference between the adjusted means obtained by the regression lines between Target and Control for the seeded and non-seeded periods.

References: Andre J. Dumas and Hubert J. Morel-Seytoux, Colorado State University Hydrology Paper No. 34 (August 1969) "Statistical Discrimination of Change in Daily Runoff."

Key Words: Statistical discrimination, hydrologic change, daily runoff, precipitation management, evaluation

Abstract: The purpose of this study was the development of a technique for rapid detection of the occurrence of a suspected hydrologic change in high mountain watersheds. A method has been developed that uses a sequence of independent daily flows. This procedure is superior to previous ones based on seasonal or yearly flows. The results of this investigation show the use of daily, instead of seasonal flow, data in a Student t-test reduces the number of necessary years of data for detection by an average of five in 14 out of the 20 cases studied, or by an average of three for the 20 cases. All of the cases come from the Upper Colorado River Basin. The study is particularly relevant to the planned cloud seeding operations of the Bureau of Reclamation in high elevation areas of the Colorado Rocky Mountains. The statistical procedure of detection relies on the Target Control concept and the application of a conditional Student t-test, a test of the difference between the adjusted means obtained by the regression lines between Target and Control for the seeded and non-seeded periods.

References: Andre J. Dumas and Hubert J. Morel-Seytoux, Colorado State University Hydrology Paper No. 34 (August 1969) "Statistical Discrimination of Change in Daily Runoff."

Key Words: Statistical discrimination, hydrologic change, daily runoff, precipitation management, evaluation

Abstract: The purpose of this study was the development of a technique for rapid detection of the occurrence of a suspected hydrologic change in high mountain watersheds. A method has been developed that uses a sequence of independent daily flows. This procedure is superior to previous ones based on seasonal or yearly flows. The results of this investigation show the use of daily, instead of seasonal flow, data in a Student t-test reduces the number of necessary years of data for detection by an average of five in 14 out of the 20 cases studied, or by an average of three for the 20 cases. All of the cases come from the Upper Colorado River Basin. The study is particularly relevant to the planned cloud seeding operations of the Bureau of Reclamation in high elevation areas of the Colorado Rocky Mountains. The statistical procedure of detection relies on the Target Control concept and the application of a conditional Student t-test, a test of the difference between the adjusted means obtained by the regression lines between Target and Control for the seeded and non-seeded periods.

References: Andre J. Dumas and Hubert J. Morel-Seytoux, Colorado State University Hydrology Paper No. 34 (August 1969) "Statistical Discrimination of Change in Daily Runoff."

PREVIOUSLY PUBLISHED PAPERS

Colorado State University Hydrology Papers

No. 25 "An Experimental Rainfall-Runoff Facility," by W. T. Dickinson, M. E. Holland and G. L. Smith, September 1967.

No. 26 "The Investigation of Relationship Between Hydrologic Time Series and Sun Spot Numbers," by Ignacio Rodriguez-Iturbe and Vujica Yevjevich, April 1968.

No. 27 "Diffusion of Entrapped Gas From Porous Media," by Kenneth M. Adam and Arthur T. Corey, April 1968.

No. 28 "Sampling Bacteria in a Mountain Stream," by Samuel H. Kunkle and James R. Meimann, March 1968.

No. 29 "Estimating Design Floods from Extreme Rainfall," by Frederick C. Bell, July 1968.

No. 30 "Conservation of Ground Water by Gravel Mulches," by A. T. Corey and W. D. Kemper, May 1968.

No. 31 "Effects of Truncation on Dependence in Hydrologic Time Series," by Rezaul Karim Bhuiya and Vujica Yevjevich, November 1968.

No. 32 "Properties of Non-Homogeneous Hydrologic Series," by V. Yevjevich and R. I. Jeng, April 1969.

No. 33 "Runs of Precipitation Series," by Jose Llamas and M. M. Siddiqui, May 1969.

Colorado State University Fluid Mechanics Papers

No. 4 "Experiment on Wind Generated Waves on the Water Surface of a Laboratory Channel," by E. J. Plate and C. S. Yang, February 1966.

No. 5 "Investigations of the Thermally Stratified Boundary Layer," by E. J. Plate and C. W. Lin, February 1966.

No. 6 "Atmospheric Diffusion in the Earth's Boundary Layer--Diffusion in the Vertical Direction and Effects of the Thermal Stratification," by Shozo Ito, February 1966.

Colorado State University Hydraulics Papers

No. 1 "Design of Conveyance Channels in Alluvial Materials," by D. B. Simons, March 1966.

No. 2 "Diffusion of Slot Jets with Finite Orifice Length-Width Ratios," by V. Yevjevich, March 1966.

No. 3 "Dispersion of Mass in Open-Channel Flow," by William W. Sayre, February 1968.