SELECTION OF TEST VARIABLE FOR MINIMAL TIME
DETECTION OF BASIN RESPONSE TO NATURAL OR INDUCED CHANGES

Completion Report

OWRR Project No. A-011-Colorado
Period July 1, 1969 - June 30, 1972

by

Dr. H.J. Morel-Seytoux

Department of Civil Engineering

Colorado State University

Submitted to

Office of Water Resources Research

U.S. Department of Interior
Washington D. C. 20240

December 1972

The work upon which this report is based was supported in part by funds provided by the U.S. Department of Interior, Office of Water Resources Research, as authorized under the Water Resources Research Act of 1964; and pursuant to Grant Agreement Nos. 14-31-0001-3006, 14-31-0001-3206, and 14-31-0001-3506.

ENVIRONMENTAL RESOURCES CENTER
Colorado State University
Fort Collins, Colorado

Norman A. Evans, Director

CER72-73HJM7

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

SELECTION OF TEST VARIABLE FOR MINIMAL TIME DETECTION OF BASIN RESPONSE TO NATURAL OR INDUCED CHANGES

In an age of increasing interference of man in nature it is important to detect rapidly the effects of purposeful management in a given sector as well as the inadvertent effects in a different sector. Due to the variability of nature and man's incomplete knowledge of its state it is difficult to assess whether an apparent change is the result of nature's caprice or of man's deliberate action. To answer this delicate question classical statistical tests have been utilized by the scientists and technicians. Often the physical variables have been reduced to mere statistical variables. While the statistical information about the variable has been somewhat utilized the physical information about it has often been neglected. As a result the tests used were far from being optimal. Detection tests which utilize both the physical and statistical information can be designed and they will have a superior power to the standard tests. The test variable to be used is a linear combination of several physical and random variables with weight factors determined by a minimization procedure. The minimization is restricted by several equality constraints of a physical origin. The test has proven to be more powerful than the classical (unconstrained) simpler tests. When applied to the Colorado River Basin Pilot Project area the power of the test, expressed in years needed for detection, is increased by a factor of two.

KEY WORDS--change, detection, minimum time, constrained optimization, statistical test, power.

RESEARCH OBJECTIVES

In the original proposal of January 10, 1969, it was stated: "The objectives are to answer the following two questions:

1. Given a region consisting of N basins in which changes are suspected and given that economic and financial constraints will limit to n the number of basins where measurements can be obtained, which n basins should be selected?

2. Furthermore, how should the measurements in individual basins be combined?

Both objectives have been achieved.

ACHIEVEMENT OF CONTRACTS

It is not desirable to repeat in this completion report all the results obtained over the past three years and the detailed procedures by which they were obtained. These results and procedures can be found in 1 dissertation [Saheli, 1972] and 2 submitted papers (both accepted for publication) [Morel-Seytoux and Saheli, 1972, 1973]. Additional papers are in preparation. Rather a brief review of the methods of attack and a sample of results will be given.

1. Introduction

In an age of ever increasing interference of man in nature it is very important to detect rapidly the effects of direct purposeful management in a given sector as well as the possibly undesirable inadvertent effects in a different sector. Due to the high variability of nature and man's incomplete knowledge of its state it is usually very difficult to assess whether an apparent change is the result of nature's caprice or of man's deliberate action.

In an effort to answer this delicate question classical statistical test procedures have been utilized by the scientists and the technicians. In so doing the physical variables have often been reduced to mere statistical variables. While the statistical information about the variable has been somewhat utilized, the physical information about it has often been neglected. As a result the tests used were not optimal. Furthermore, the more or less conscious realization on the part of the scientists or the technicians that indeed physical information has been lost in the process may have caused the sometimes encountered attitude of suspicion toward statistics, leading to a less than full utilization of even the statistical information.

Special detection tests which utilize both the physical and statistical information can be designed. The message of this report is that such special tests for a particular physical situation are likely to have a superior power to the standard tests. This new technique of detection will be discussed in the context of winter precipitation management. However, the purpose of this report is not to discuss the potential or the effectiveness of cloud seeding. For further discussion of these aspects see for example Hurley (1972) and Morel-Seytoux (1972). The purpose of this report is to discuss a technique of detection that has applicability whenever a hydrologic change and a cause for change are suspected.

2. Classical Procedures

Classically the problem of detection is solved by performing a significance test (Mood and Graybill, 1963) on a sample of observations. The null hypothesis (Mood and Graybill, 1963) is either accepted or rejected depending upon whether the calculated test variable falls within the critical region or outside of it (Mood and Graybill, 1963). Of course it is not possible to perform the test until observations have been collected, which may take years.

Before embarking on experiments over several years it is wise to check ahead of time what the probability is that a significance test will provide a positive result. This probability is the power (Brownlee, 1961) of the test. Equivalently one can calculate the number of years (or sample size) necessary to detect a given percentage increase, e.g., 10% of say the population mean at the α^{th} significance level and β^{th} power. Certainly one would reject a design of an experiment which after five years would have a power of only one in four. If the duration of the experiment cannot be extended then one would look for a better experimental design.

a. <u>The two-sample u-test</u> - The two-sample u-test is a test of the hypothesis that assumes that the population mean is equal to a given value while the population standard deviation is known and stationary (Li, 1964). The statistic used in testing this hypothesis is:

$$u = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \tag{1}$$

in which $\bar{x}$ = the sample mean; μ = the population mean; σ = the standard deviation; and n = the sample size. Here the critical region is $|u| > 1.96$ if the 5% significance level is used. The significance of an increase, say in spring runoff, will barely be achieved if the observed statistic u just equals 1.96, i.e., if:

$$u = \frac{\overline{\Delta Q}}{\frac{\sigma_Q}{\sqrt{N}}} = 1.96 \tag{2}$$

in which $\overline{\Delta Q}$ = the suspected increase in spring runoff; N = the number of years necessary to establish the significance of the increase with a 50% power; and σ_Q = the standard deviation of the natural spring runoff. This last equation can be used to calculate N :

$$N = \frac{3.84\ \sigma_Q^2}{(\overline{\Delta Q})^2} \tag{3}$$

Let κ be the suspected percentage change in the mean of the hydrologic variable of concern (e.g., spring runoff) and define C_v as the coefficient of variation (ratio of standard deviation over mean) in %. Then the expression for N can be rewritten as:

$$N = 3.84 \left(\frac{C_v}{\kappa}\right)^2 \simeq 4 \left(\frac{C_v}{\kappa}\right)^2 \quad . \tag{4}$$

Suppose now that in 1970 one were to start winter cloud seeding experiments in the Upper Colorado River Basin. Suppose one would expect, due to cloud seeding, an increase in spring (i.e., snowmelt) runoff of the order of 10%. How many years of experiments would be required to insure an even chance of detection at the 95% significance level? Since in the Upper Colorado River Basin the coefficient of variation is rarely less than 30%, the number of years would be of the order of 36 years. Thus by the year 2006 the u-test might indicate that a statistically significant change in seasonal runoff has occurred.

b. <u>The target-control chi-square test</u> - Clearly the two-sample u-test is not very powerful. A more sophisticated test is needed. Roughly speaking the use of a control makes the target look as though it has an effective coefficient of variation much smaller than the actual one. The larger the coefficient of correlation between the target and the control, the smaller the apparent coefficient of variation of the target. As a result the detectability of a change in the target behavior is increased. The analog formula to Eq. (4) for N (Morel-Seytoux, 1968) is:

$$N = 3.84 \; (1 - \rho^2) \; \frac{C^2_{v,t}}{\kappa^2} \simeq 4 \; (1 - \rho^2) \; \frac{C^2_{v,T}}{\kappa^2} \quad . \tag{5}$$

Suppose again that in 1970 cloud seeding experiments were to be initiated in the Upper Colorado River Basin, say in the San Juan

Mountains areas east of the Los Pinos River Basin. The candidates for the test in the target area are the East San Juan, the San Juan, the Rio Blanco and the Piedra rivers. Outside the target area the runoff station with the highest correlation coefficient with some of the target stations is Hermosa Creek near Hermosa. With the Piedra River near Piedra its coefficient of correlation is 96%. From Eq. (5) the value of N is calculated, yielding a value of 7.4 years. There is a significant improvement over the u-test analog value of 36. Nevertheless this number of years is still quite high.

3. The Concept of Grouping of Observations

Looking back at Eq. (4) one can see that a possibility to decrease N is to reduce the coefficient of variation. If only one station is involved C_v is a fixed quantity and no reduction in C_v is possible. If on the other hand the region over which a hydrologic change is suspected to occur includes several gauge sites, the observations can be grouped in a linear combination say:

$$Q^* = \sum_{i=1}^{n} x_i Q_i \quad . \tag{6}$$

The weights, x_i, are fixed but unknown parameters. In terms of Q^* and of the Q_i, Eq. (4) can be rewritten:

$$N = 4 \frac{\mathrm{Var}(Q^*)}{(\overline{\Delta Q^*})^2} = 4 \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \mathrm{Cov}(Q_i, Q_j)}{\left(\sum_{i=1}^{n} x_i \overline{\Delta Q_i} \right)^2} \tag{7}$$

where $\overline{\Delta Q_i}$ is the suspected change of the mean of Q_i. Due to the indeterminate form of Eq. (7) when all $x_i = 0$, it is convenient to introduce a normalization condition on the x_i, which is not a constraint at all on the minimization because Eq. (7) is homogeneous in the x_i. The normalization constraint could be:

$$\sum_{i=1}^{n} x_i = n \quad . \tag{8}$$

That normalization constraint has however no physical meaning. The Q_i may be random variables but are nonetheless physical variables (e.g., basin seasonal runoff). On the other hand Q^* is an artificial variable, random without immediate apparent physical significance. Meaning can be attached to the mean of Q^* by requiring that:

$$\overline{Q^*} = \sum_{i=1}^{n} x_i \, \overline{Q}_i = \sum_{i=1}^{n} \overline{Q}_i \quad . \tag{9}$$

In this case the mean of Q^* is the global mean of the runoff for the region of interest. The constraint defined by Eq. (9) has thus a physical connotation whereas the normalization constraint by Eq. (8) did not. If the suspected changes, the $\overline{\Delta Q_i}$, were known an optimal test would be obtained by minimizing the objective function, N, defined by Eq. (7) subject to constraint Eq. (9). In general, however, the $\overline{\Delta Q_i}$ are not known and this is why one wants to apply a detection test. It may be possible to estimate the $\overline{\Delta Q_i}$. For example in the cloud seeding experiments one may not know exactly what the increase in snowfall will be, but one may estimate that the mean increase will be about 10% of the historical mean snowfall with minor variations in the percentage from basin to basin. Based on this assumption (a uniform 10% increase in mean winter precipitation) one can estimate the increase in mean seasonal runoff for all individual basins (Nakamichi and Morel-Seytoux, 1971). A new constraint is then introduced:

$$\overline{\Delta Q^*} = \sum_{i=1}^{n} x_i \, \overline{\Delta Q_i} = \sum_{i=1}^{n} \overline{\Delta Q_i} \quad . \tag{10}$$

In other words the increase in the mean of Q^* is the increase in the global runoff mean for the region. The minimization problem for the x_i with objective function N defined by Eq. (7) subject to constraint Eqs. (9) and (10) can be solved easily by the Lagrange multipliers technique (Nakamichi and Morel-Seytoux, 1971). The power of the test was greatly increased over that obtained from the u-test, even if applied to the total runoff, i.e., Q^* for the particular case when all x_i are set equal to 1. However, the optimal x_i , denoted x_i^* , can be positive or negative. If the $\overline{\Delta Q_i}$ are poorly estimated and in usual cases of interest they are likely to be, the negativity of some of the x_i^* may result in very small and possibly negative actual observed $\overline{\Delta Q^*}$ as distinct from the estimated $\overline{\Delta Q^*}$ from Eq. (10) and used to derive the optimal x_i^*. For this reason an additional constraint is added for safety to Eqs. (9) and (10), namely the non-negativity condition. Finally the concept of grouping of observations can be combined with the target-control concept to maximize the efficiency of the test.

4. <u>An Optimal Target-Control Test</u>

Using the subscript t for target and c for control, an optimal test is the target-control conditional student's t-test (Dumas and Morel-Seytoux, 1971) applied to the random variable Q_t^* defined by

$$Q_t^* = \sum_{i=1}^{n} x_i^* \, Q_{ti} \tag{11}$$

conditioned upon the random variable Q_c^* defined by:

$$Q_c^* = \sum_{i=n+1}^{n+m} x_i^* \, Q_{ci} \tag{12}$$

where the x_i^* are the solution of an optimization problem, namely (in vector notation), minimize:

$$\left\{ [\underline{x}_t' \, V_{tt} \, \underline{x}_t - (\underline{x}_t' \, V_{tc} \, \underline{x}_c)^2 \,/\, (\underline{x}_c' \, V_{cc} \, \underline{x}_c)] \,/\, (\underline{x}_t' \, \underline{\overline{\Delta Q}})^2 \right\}$$

with respect to $\underline{x}_t$ and $\underline{x}_c$, subject to the constraints:

$$\sum_{i=1}^{n} x_i \, \overline{Q}_{ti} = \sum_{i=1}^{n} \overline{Q}_{ti} \tag{13}$$

$$\sum_{i=1}^{n} x_i \, \overline{\Delta Q}_{ti} = \sum_{i=1}^{n} \overline{\Delta Q}_{ti} \tag{14}$$

$$\sum_{i=n+1}^{n+m} x_i \, \overline{Q}_{ci} = \sum_{i=n+1}^{n+m} \overline{Q}_{ci} \tag{15}$$

and

$$x_i \geq 0 \quad i = 1, 2 \ldots n \tag{16}$$

where the V are covariance matrices.

The optimal solution was obtained by a non-linear programming algorithm (Morel-Seytoux and Saheli, 1973) which is a combination of a modified General Differential Algorithm (Wilde and Beightler, 1967) and a search technique about a point where the Kuhn-Tucker conditions are satisfied (Saheli, 1972). The technique was applied to develop a test of the effect of winter precipitation management on seasonal runoff in the San Juan Mountains area (Hurley, 1972).

5. Case Study

The target area is shown in Fig. 1 (for a more complete discussion of the Bureau of Reclamation's project, see Hurley (1972)). The effect of increasing the snowpack by cloud seeding is expected to result in an increased snowmelt runoff. An analysis of the mean hydrographs for stations in the target area (for example, see Fig. 2) showed that with the choice of the 6 months period, March-August, this seasonal runoff will be a good index of a snowpact increase. It is seen on Fig. 2 that the rise of the hydrograph starts at the beginning of March while the recession of the surface runoff ends in August. Using the March-August seasonal runoff as the basic hydrologic variable an optimal combination was obtained by the minimization procedure described earlier. The results are shown in Table 1.

6. Conclusions

A new test was developed with a better power than the standard tests. The minimum number of years with the optimal weights is significantly lower than the number obtained with uniform weight of unity. Though the use of the technique was illustrated in the case of a hydrologic change by precipitation management, the technique can be utilized for the detection of changes due to suspected other causes (e.g., timber cutting, pollution, urbanization). It is necessary that a cause for the change be suspected so that physically meaningful constraints analogous to Eqs. (9) and (10) can be imposed on the optimization procedure.

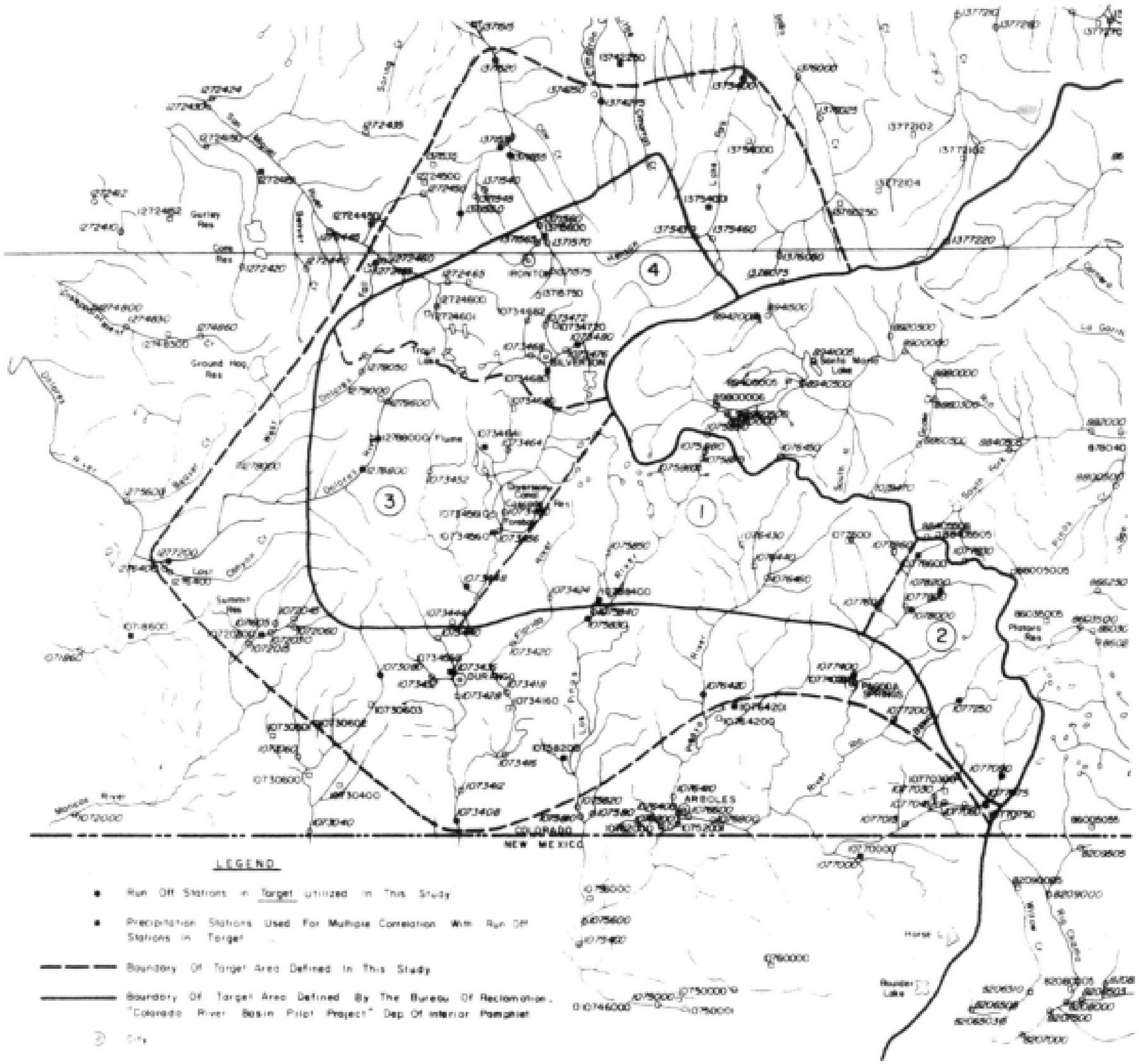

Figure 1. Map of target area.

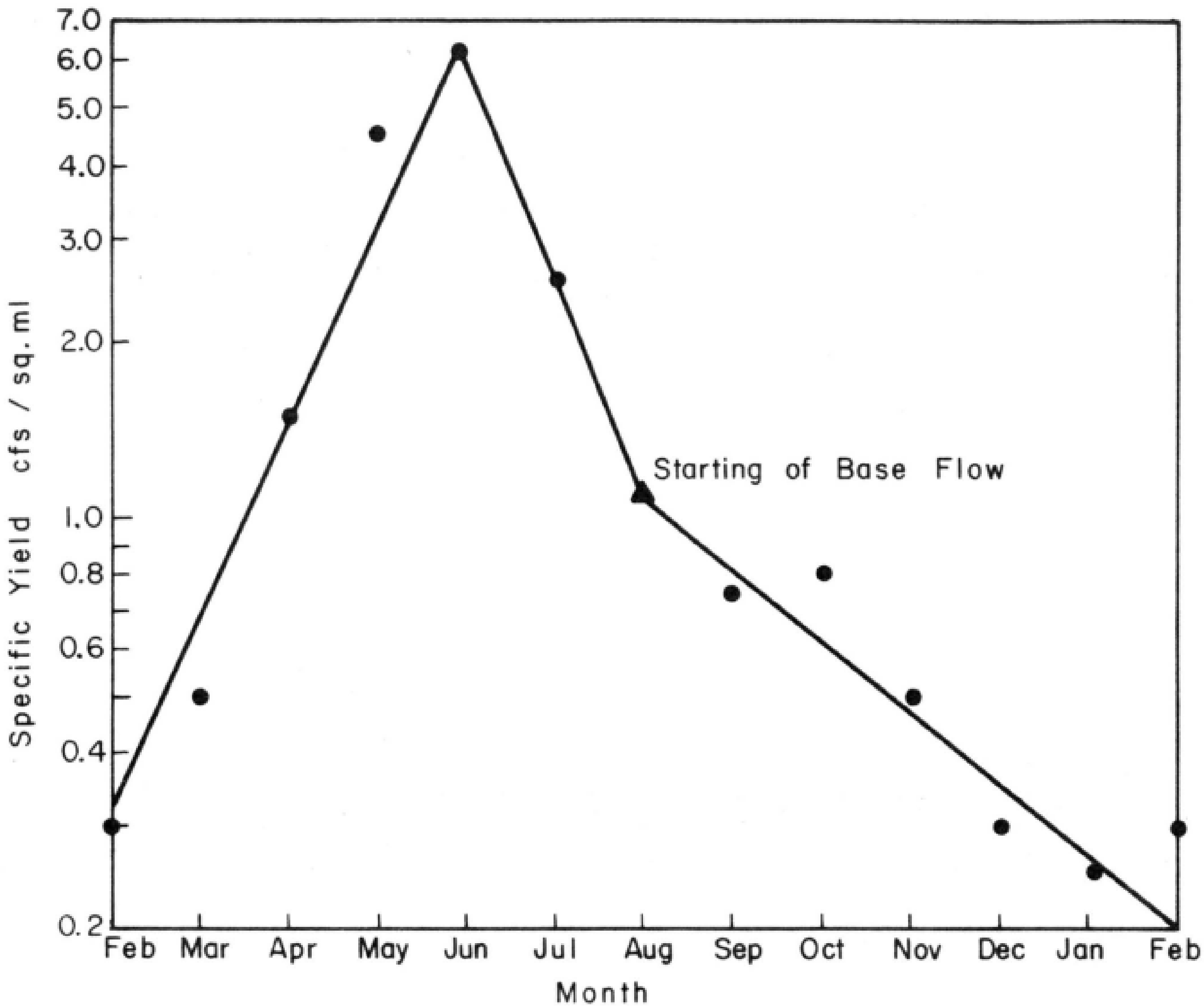

Figure 2. Separation of surface and base flow (Animas River at Durango).

TABLE 1
OPTIMAL COMBINATION OF STATIONS IN THE TARGET
(SAN JUAN MOUNTAINS) AND THE CONTROL (MAROON PEAK)

CSU ID	Name	Location (Basin)	Number of Existing Years of Record Within (1939-1968)	Gage Elev (ft)	Drainage Area (sq mi)	Weight Factor X
		Stations in Target				
1371520	Uncompahgre River at (near) Colona, Colo.	Uncompahgre	30	6,318.80	437.0	5.142
1077250	Rio Blanco near Pagosa Springs, Colo.	Rio Blanco	30	7,950.00	58.0	2.585
1073480	Animas River at Howards-ville, Colo.	Animas	30	9,616.00	55.9	17.249
1073408	Animas River near Cedarhill, Colo.	Animas	30	5,960.00	1,090.0	1.217
		Stations in Control				
1590000	Roaring Fork at Glenwood Springs, Colo.	Crystal	30	5,720.70	1,460.0	0.298
1373055	North Fork Gunnison River near Crawford, Colo.	North Fork Gunnison	30	6,038.60	521.0	-2.608
1373360	Smith Fork near Crawford, Colo.	North Fork Gunnison	30	7,200.00	42.3	-13.927
1377200	Tomichi Creek at Sargents, Colo.	Tomichi Creek	30	5,720.70	1,020.0	11.920
1378100	East River at Almont, Colo.	East River	30	8,008.29	295.0	9.628
1379000	Taylor River below Taylor Park Reservoir, Colo.	Taylor	30	9,168.67	254.0	-3.597
1425625	Buzzard Creek near Collbran, Colo.	Plateau Creek	30	6,920.00	139.0	4.068
1370300	Kannah Creek near White River, Colo.	Kannah River	30		61.9	62.538
1740000	Blue River below Green Mountain Reservoir, Colo.	Blue River	30	7,510.00	623.0	-4.056

Minimum Number of Years with Optimal X's is 3.8 ≈ 4 Years. Minimum Number of Years with all X's =1 is 99 Years.

LIST OF PROJECT PUBLICATIONS

Morel-Seytoux, H. J. and F. P. Saheli. 1972. A system approach to minimal time detection of changes in watersheds. Proceedings of American Water Resources Association Symposium on Watersheds in Transition, June 19-23, Fort Collins, Colorado.

Morel-Seytoux, H. J. and F. P. Saheli. 1973. Test of runoff increase due to precipitation management for the Colorado River Basin Pilot Project. Submitted for publication to Jour. of Applied Meteorology, March 1972.

Saheli, F. P. 1972. Minimization of time of detection of the effect of weather modification. Ph.D. Dissertation, Colorado State University, May 1972. 101 p.

ADDITIONAL REFERENCES

Brownlee, K. A. 1961. Statistical theory and methodology in science and engineering. John Wiley and Sons, Inc. 570 p.

Cochran, W. G. and G. M. Cox. 1957. Experimental designs. John Wiley and Sons, Inc. 611 p.

Dumas, A. J. and H. J. Morel-Seytoux. 1971. Detection of a change in runoff by an analysis of daily flows. Bull. of the International Association of Scientific Hydrology, Vol. XVI, 1.3. pp. 69-87.

Graybill, F. A. 1961. Introduction to linear statistical models. McGraw-Hill, New York. 500 p.

Hurley, P. A. 1972. Colorado pilot project: design hydrometeorology. Jour. of the Hydraulics Div., ASCE, Vol. 98, No. HY5, May 1972. pp. 811-826.

Li, J. C. R. 1964. Statistical inference I. Edwards Brothers, Inc, Ann Harbor, Michigan.

Mood, A. M. and F. A. Graybill. 1963. Introduction to the theory of statistics. McGraw-Hill, New York. 443 p.

Morel-Seytoux, H. J. 1968. Suitability of basins to weather modification and statistical evaluation of attainment. Final report to the Bureau of Reclamation, July 1, 1968, Dept. of Civil Engineering, Colorado State University.

Morel-Seytoux, H. J. 1972. Techniques of statistical evaluation of runoff results of precipitation modification. Final report to the Bureau of Reclamation, Division of Atmospheric Water Resources Management, Project SKYWATER, March 1972. 45 p.

Nakamichi, H. and H. J. Morel-Seytoux. 1971. Runoff grouping for detection of a change in runoff. Jour. of the Irrigation and Drain Div., ASCE, Vol. 97, No. IR2, June 1971. pp. 275-290.

Wilde, D. J. and C. S. Beightler. 1967. Foundations of optimization. Prentice Hall, New York. 480 p.