TA7
C6
CER 61-53 B
copy 2

A STUDY OF THE STATISTICAL PREDICTABILITY OF STREAM-RUNOFF IN THE UPPER COLORADO RIVER BASIN

By

Paul R. Julian

Research Staff
High Altitude Observatory
University of Colorado

PART II

ENGINEERING RESEARCH
AUG 11 '71
FOOTHILLS READING ROOM

October 1961

CER61RAS-VMY53B

ACKNOWLEDGEMENTS

I have written this report primarily for hydrologists who have had some mathematical and statistical training but are not versed in the particular statistical techniques used herein. Some of the more complicated procedures and explanations have been relegated to three appendices. It is hoped that the brief and incomplete start that this report represents will provide a stimulus to workers in hydrology to utilize further the techniques employed here. Many people have assisted the author in many ways. Computation and clerical assistance was rendered by Mr. Harold Petrie and Mr. Frank Weinhold. Consultation with Messrs. Eugene Peck, Norman Macdonald, Walter Langbein, and Nicholas Matalas proved invaluable. Cooperation and discussion with fellow workers on the overall project, Dr. Richard Schleusener, Mr. Loren Crow, Dr. V. Yevdjevich, Miss Margaret Brittan, and Director, Dr. Morris Garnsey, served to crystallize the objectives and procedures of the project. Dr. Max Woodbury of the College of Engineering, New York University, deserves special credit for his part in suggesting certain phases of the work, and for co-authoring two of the appendices.

U18401 0593110

TABLE OF CONTENTS

TABLES AND FIGURES

TABLES AND FIGURES (cont.)

TABLES AND FIGURES (cont.)

I. INTRODUCTION

A. The Problem

The research carried out by the author was primarily directed toward gaining knowledge of the basic hydrologic processes in an effort to draw conclusions concerning the amount of statistical predictability inherent in streamflow amounts. In carrying out this basic objective it was, of course, necessary to investigate certain other related problems.

The following itemized questions are those that the author feels were, with some degree of profit, investigated by this phase of the project.

(1) What are some of the basic statistical relationships of measured precipitation to stream runoff?
(2) Is there any statistical predictability in precipitation or streamflow amounts or are they indistinguishable from random numbers?
(3) What significance has an answer to (2) for practical hydrological purposes?
(4) What are the causes of the observed decrease of runoff efficiency of the Upper Colorado Basin in the past 50-60 years?
(5) Are there reasonably good relationships between large-scale (that is, hemispheric) atmospheric circulation parameters and specific Basin weather?

The question might legitimately be asked as to why the approach involving streamflow amounts and statistics was chosen rather than an effort toward solving some of the basic problems of why the atmosphere behaves as it does and attempting to make a long-term forecast of precipitation within the Basin. The answer to this question can be stated rather simply. The author's familiarity with the meteorological forecast problem and with the application of statistical methods to meteorological problems led him to the conclusion that the greatest promise of producing something of value lay in the statistical approach rather than with investigations into ordinary physical prediction problems of the atmosphere.

It should be remembered that the physical processes by which the atmosphere operates and by which the hydrologic cycle operates are extremely complex, and that only a major effort involving a great deal of expense, time, and scientific manpower is going to solve these problems. A limited research effort must therefore place its emphasis where, in the opinion of the scientists involved, the limited effort will produce the greatest results.

B. The Use of the Statistical Method

Since the ultimate goal of research into atmospheric and hydrologic problems is prediction, it will behoove us to examine briefly the prediction problem in meteorology. Similar remarks could be made concerning hydrology.

At the risk of oversimplification, the process of prediction of a physical system can be approached in two ways. The first of these, herein called the physical approach, is dependent upon the description of the physical system, in our case the atmosphere, in terms of mathematical equations. The appropriate equations for the atmosphere are non-linear differential equations describing the motions of the atmosphere using Newtonian mechanics which are, of course, time dependent. The prediction problem involves (1) determining which equations are appropriate, (2) determining the proper boundary conditions for the system, (3) determining the proper initial specification of the atmosphere (since the problem is a time-dependent one), and (4) determining how to solve the equations. Such a scheme might be called a deterministic one, since theoretically at least, the equations would be capable of specifying the future behavior of the atmosphere for an infinitely long period. It should be obvious that such a deterministic prediction scheme is, at present, an unrealistic one. Meteorologists are just beginning to have an idea how to handle the four problems outlines above and the determination of the future behavior of the atmosphere in any detailed sense by the physical approach is practical only over a few days at most.

The second approach is the statistical one. In this scheme, the actual physical processes occurring within the system are not necessarily important. A mathematical model is used, to be sure, but the prediction is done by using past (in time) information in a form which may have no connection with the actual physical system itself. Thus, a statistical prediction model similar in all respects could be used to predict future weather conditions, stock market prices, or the population of the United States -- totally different physical systems. Statistical prediction schemes work by considering the amount of information in the past record of a quantity -- that is, whether the past record has any characteristics which are systematic--so that the systematic component can be utilized for prediction. Examples of systematic characteristics in a time series might be cycles, trends, or linkages between adjacent values.

The systematic portion of the time-series may be called the deterministic component; the portion of the series which is statistically independent from year-to year (random in time) is called the non-deterministic or stochastic component. The series is assumed to be composed of the sum of these mutually independent components. The crux of any statistical prediction scheme is simply the determination of the relative importance or magnitude of these two components. For the purposes of this study the naturally-occuring time-series will be compared with a completely random series--that is, a series of numbers drawn independently from some frequency distribution. Any significant deviation from this hypothetical series will be considered as a deterministic component and a mathematical model used to reproduce this component.

The advantages of the statistical approach may be summed up as follows: It can be applied when no or very little knowledge of the physical situation is available. The resulting forecasts are made in terms of probability statements, and a confidence limit may be placed on each forecast. This foregoing advantage to many people might seem like a disadvantage since it demands that a person think in statistical terms. However, when we realize that statements involving future happenings at our present level of knowledge always involve a certain amount of indeterminacy whether explicitly recognized or not, the acceptance of probability statements as forecasts is largely a matter of understanding and education. For many operational problems, statistical forecasts may be efficiently utilized.

The hydrological profession has made use of one type of statistical model already. In designing equipment to cope with floods, engineers have come to depend on extreme value theory -- a part of statistics concerned with the probability of extremely unlikely events. Thus, design characteristics for a hundred-year flood, etc., are being utilized and these design characteristics are, it should be recognized, a form of statistical forecast. The engineers are gambling, in this case, that the likelihood of a larger than hundred-year flood is small enough that building costs to attempt to handle anything larger are prohibitive. This risk is inherent in any design problem where events detrimental to the structure under design have a small but finite chance of occurrence; and underlying the choice of design parameters is a statistical model -- in this case the model is based on the economic considerations of the cost of construction taking into account the likelihood of the meteorological event occurring, against the cost of loss of the system if that event occurs.

In other fields of the hydrological profession, statistical methods are also used. For example, multiple regression or multi-variate techniques are used to forecast next spring and summer's runoff using precipitation, snow-pack, etc. data from this winter. The forecasts made by this scheme have a certain probable error associated with them and a confidence level can thus be given.

Thus, although statistical methods are not new to hydrology, the approach taken in this study will apply statistical methods to precipitation and streamflow data in a fashion which has not appeared to date (to the author's knowledge). In the next section, some literature which has appeared in the last few years will be discussed which, along with this study, can be said to be laying the groundwork for the practical utilization of statistical methods for prediction of streamflow amounts.

C. Literature Pertinent to the Discussion

The following publications will be briefly discussed -- not because they throw light on the procedures utilized by the author, but because they are what he considers to be the most important sources in contributing to our knowledge of the variations in precipitation and streamflow.

The first of these is The Prediction of Long-Continuing Drought in South and Southwest Texas by Dr. Don G. Friedman of the Travelers Weather Research Group (11). This study is an excellent example of a statistical evaluation of predictability in rainfall amounts together with a particularly cogent review of problems connected with the definition of a drought and with prediction of rainfall using related data. Friedman uses some fairly sophisticated tests on rainfall amounts in southern and southwest Texas in an effort to ascertain whether any trends or cycles are present. By a "trend" we mean a systematic movement throughout a limited length of record; it can be thought of as a fluctuation with a period much greater than the length of record. Specifically, the following questions were asked of the data: (1) Is the climate of south and southwest Texas becoming progressively dryer? (2) Are there regularly recurring cycles of wet and dry periods in Texas climate? (3) Is there any year-to-year persistence in the wet and dry spells (climatic persistence from year to year)? From the statistical tests on ten rainfall stations, Friedman concluded that there is no statistical evidence for the existence of trends, cycles, or persistence in south Texas rainfall. Except for different, individual stations appearing significant in one test or another, as approximately 5% of them would do considering the significance level Friedman chose, the rainfall data resembled random numbers.

Friedman proceeds then to illustrate how the conclusion of independence of rainfall amounts together with their frequency distribution can be used to make useful statements about the probability of future rainfall amounts and how such studies are of value in the Farm Mortgage Loan Program in Texas.

The second reference to be discussed is Geological Survey Circular 410, Probability Analysis Applied to a Water-Supply Problem by Luna B. Leopold (22). This paper presents in a forthright and lucid manner some very important conclusions concerning streamflow variations and their estimation by the use of simple statistical analysis. The analysis happens to be of the flow of the Colorado River at Lee Ferry which makes the study of even greater importance here. The more important conclusions of this paper are: (1) Streamflow amounts by water-year are not independent of each other; some serial or sequential correlation is present. (2) The effect of this serial interdependence is to effectively reduce the length of record for the estimation of mean values over any length of time. Thus, for example, a 100-year actual record may have the same effective length as a 25-year record of independent, random values, and, thus, only 25 independent estimates of the long-term mean rather than 100. (3) An illustration is given calculating the probability of a mean over a given length of future record being higher or lower than the comparable mean value over the past record. For example, there is by Leopold's method of calculation a 9% chance that the next 10-year mean of flow at Lee Ferry will be less than 12.3 million acre feet.

These basic facts set out by Leopold are of fundamental importance in understanding the variations in streamflow and the use made of averages and variabilities in hydrologic problems. All of the important conclusions set out by him, as listed above, are reinforced by the results of the investigation reported on here.

The third piece of literature to be discussed is a small monograph by P.A.P. Moran, Professor of Statistics in the National University of Australia, Canberra, and is entitled The Theory of Storage (28). This book can be described largely as theoretical, but it is the type of theoretical treatise which provides the foundation for real, useful, practical advances. Moran considers situations in which a store exists; it may be a storage reservoir, a warehouse, etc., and thoroughly investigates the probability models such a storage system suggests. For example, given a storage reservoir with random imputs, i.e., water-year discharges, and a prescribed release rule, the probability of finding a given amount of water in the reservoir may be calculated. In more practical situations, that is, with more than one dam on a river, or with the release rule prescribed on a weekly or monthly basis, or with an empty reservoir to start, a direct analytic calculation is not possible. In these cases, Moran suggests the use of linear programming models and "Monte Carlo" methods, and gives some references to work already done along those lines. The use of such linear programming models and "Monte Carlo" methods to obtain the probability distribution of future reservoir contents is termed a "synthetic hydrology."

The final item to be discussed is a paper concerned with the practical application of the foregoing theory by Moran. It is entitled Queing Theory and Water Storage by W. B. Langbein and appeared in the Journal of the Hydrolics Division, Proceedings of the American Society of Civil Engineers (Proc. paper 1811) (19). Langbein gives examples of various release rules and how the probability distribution of reservoir contents can be calculated in specific instances. He also considers, and this is important to the present study, how the effect of non-randomness or serially correlated flows may be handled in certain simplified instances. Incidentally, he shows mathematically how the presence of non-randomness or sequential correlation serves to increase storage requirements over what a purely random sequence would demand.

The previous articles, although not complete by any means, are reviewed with the object of pointing up the emphasis that will be made in the ensuing report.

II. ANALYSIS OF BASIC DATA

A. Data Source

The raw data for the meteorological and hydrological studies carried out were obtained from three sources: The precipitation data came from Bulletin W and the Supplement to Bulletin W of the U.S. Weather Bureau and from the punch-card data prepared under this contract by Colorado State University. The streamflow data were taken from the USGS Water Supply Paper series. In all cases, gauged stream discharge measurements were corrected for all such trans-mountain diversions or artificial regulatory data as were published.

Throughout the remainder of this report, precipitation amounts refer to November through April totals unless otherwise indicated, and, when necessary, 'winter 1959', for example, is taken to mean 1 November 1958 to 30 April 1959. All streamflow data are by water-year unless otherwise indicated. Missing monthly precipitation data were estimated when needed by the method given by Paulhus and Kohler (31) and recently tested by McDonald (25).

B. Variability of Precipitation and Runoff

From what is presently known about the runoff process, the following generalizations may be made. Neglecting the many complicating factors, indigenous to a given drainage basin, that affect streamflow, the average relation between precipitation and streamflow is not apparently a linear one (20). The specific curve relating the two quantities is a function of climate -- or to use a specific but cruder measure -- of mean temperature. The efficiency of the runoff process is an important quantity defined rather simply as the percentage of precipitation falling within a basin that actually runs off. This runoff efficiency also varies with 'climate', or with mean temperature. The physical reason why this is so seems obviously to be the fact that evaporative and transpirative processes are more efficient at higher temperatures, thereby reducing the efficiency of the runoff process.

A simple model of the runoff process may be set up as in the following equation:

Runoff (effective precipitation ± natural storage) = Precipitation - Evapotranspiration - X factor.

All quantities above have their usual interpretation (see USGS Water Supply Paper #1541) except for the quantity termed X factor. Into this quantity are lumped all factors tending to produce inhomogeneity in the historic records we possess. For example, instrumental errors due to changing gauging methods or rain-gauge exposure, increasing losses to streamflow because of increasing water usage by man, changes in streamflow brought about by man-made changes in land cover, etc. Historic records of streamflow and point precipitation are available. Stream runoff can be obtained provided records of diversions, artificial storage, etc. are available. By using recession techniques, information on effective precipitation can be obtained by removing the storage term in the above equation. In the final report under this contract by Dr. V. Yevdjevich, Colorado State University, such a technique is carried out. Two terms are not known -- the amount of inhomogeniety in the record, the X-factor, and the evapotranspirative term. In a semi-arid region, the lack of information on the latter can be serious. For example, using very rough estimates of average precipitation over the upper Basin, the runoff efficiency of the Colorado above Lee Ferry can be said to be something less than 20%.

Thus, the runoff process may, crudely, be said to be a simple system in which the runoff represents the relatively small difference between rather larger amounts of precipitation and evapotranspiration, which difference may be very small in the case of semi-arid to arid basins to moderate to large for more humid basins. The important point, however, is that the relative variability of the streamflow will always be larger than the relative variability of either the precipitation or the evapotranspiration because the temporal variability of these latter quantities is at least to some degree independent. In statistical terms, the variance of streamflow is equal to the sum of the variances of precipitation and evapotranspiration reduced by a factor proportional to the correlation between precipitation and evapotranspiration. Only if this correlation were perfect, which, of course, it is not, would the variance of the streamflow be equal to the variance of either parameter.

In Table II-1 are presented the coefficients of variation (standard deviations divided by means) of precipitation and streamflow records in the upper Colorado Basin. It will be noted that the coefficients of variability of the runoff data are at least roughly correlated with altitude (temperature) and basin size. They vary from 0.23 and 0.24 for high mountain small basin streams (Blue, Fraser, Roaring Fork, etc.) to 0.32 and 0.45 (San Juan and Green) for the larger "main stem" tributaries. Lee Ferry's coefficient of variability is 0.31 (using Leopold's virgin flow figures, 0.28). The individual precipitation stations show variabilities in about the same range, but it is extremely important to notice that these are for <u>individual</u> stations. If all precipitation stations were combined within a basin so as to get a better estimate of 'actual' precipitation, the variability would be greatly reduced. For example, the coefficient of variation for the sum of seven Basin stations--Silverton, Shoshone, Dillon, Gunnison, Ignacio, Montrose, and Grand Junction-- is 0.21. If it were possible to measure total Basin precipitation this figure would, without any doubt, be lower yet.

To sum up, the temporal variability of runoff exceeds either that of precipitation or evapotranspiration because it represents the relatively small difference between the latter much larger quasi-independent quantities.

The spatial variability of precipitation and runoff also are important basic facts. It is an accepted fact that gauge measurements of precipitation are affected to some fairly large but unknown degree by very localized conditions. The degree to which gauge precipitation measurements are representative of true precipitation over a much wider area is a difficult question to answer. Many studies have been carried out correlating gauge precipitation measurements with streamflow nearby; the wide range of correlation coefficients obtained merely emphasizes the sampling problems inherent in making precipitation measurements.

From a physical point of view, we would expect runoff measurements to be better spatially correlated than precipitation measurements, and indeed this is borne out. Table II-2 contains a summary of a number of correlation coefficients [by rank correlation method. See, for example (29)] between precipitation stations, between precipitation and streamflow stations, and between streamflow stations. First of all, note that the correlation coefficients for stations within a few (less than 30) miles of each other -- even in extremely rugged terrain -- are well correlated. The Silverton-Trout Lakes-Ames threesome and the Dillon-Fraser and Delta-Montrose pairs illustrate this. As would be expected, as the distance separating the stations is increased, the correlation coefficient is reduced. No meaningful, spatial pattern in these correlations, however, was evident to the author. Some seemingly anomalous correlation coefficients are evident, most likely due to topographic features: Ignacio and Shoshone, over 170 miles apart are correlated +0.42 while Grand Junction and Dillon, about 140 miles apart are correlated -0.24. Although the variation of such spatial correlation coefficients is of interest in itself, the only conclusion drawn here is that such coefficients decrease with increasing distance such that very little or no correlation remains when the precipitation-observing stations are 150-200 miles apart.

Examining the runoff correlations, a much more coherent picture emerges. Adjacent basins are extremely well correlated; the lowest correlation obtained (0.52) was for the two most widely separated basins, the San Juan at Rosa, New Mexico, and the Main Stem at Hot Sulphur Springs. The runoff of the smaller basins, separated by 150-200 miles, is thus better correlated (on the order of 0.6) than are precipitation amounts. The coefficients thus bear out what we would intuitively expect.

An important point here, moreover, is just why the runoff data are better correlated. It does not follow that this better correlation is due solely to the sampling problems inherent in gauged precipitation measurements, although this must certainly be a large factor. If the meteorological conditions affecting evaporation and transpiration are better correlated spatially than precipitation amounts, as may well be the case, more coherent runoff values than precipitation values would result.

In one instance two nearly adjacent basins were compared by calculating all the possible correlation coefficients. The diagram below gives the results for precipitation-precipitation, precipitation-runoff, and runoff-runoff, for the Blue and the Fraser.

```
Dillon precipitation ——— .62 ——— Fraser precipitation
       |        \                  /         |
      .50       .52            .58          .59
       |                                     |
Blue runoff ——————————— .80 ——————— Fraser runoff
```

Again the highest coefficient involved is in the runoff values.

One final point of some interest is contained in the coefficients for the three main tributaries, the Green at Green River, Utah, the Main Stem at Cisco, and the San Juan at Bluff. The Green and the San Juan both correlate 0.85 with the sub-basin separating them. If the two were statistically independent of each other a correlation coefficient of $(0.85)^2 = 0.72$ between them would be expected. This 0.72 correlation would then be due to the fact they are both correlated 0.85 with a third basin. The actual correlation is 0.61, and the partial correlation coefficient r(Green - San Juan . Main Stem) = -0.39 which at least suggests that the Green and the San Juan are oppositely correlated over and above the tendency for the entire Basin to be correlated. In other words, although the whole upper Basin tends to be wet or dry at the same time (that is, in any given winter) the latitudinal extremes of the Basin, the Green and the San Juan, have a tendency to be slightly opposed so that, for example, one would be dry and the other not-so-dry or one wet and the other not-so-wet.

It is possible that there is a reasonable physical explanation for this behavior. In the final report of the project written by Schleusener and Crow (Colorado State University), there is some material to suggest that large storms (as arbitrarily defined by Crow) have exhibited a trend in their behavior such that in the early part of the century large storms had a tendency to occur in northern Colorado and in later years in southern Colorado. Such a fact, if significant, could be attributed to a shift in storm tracks over the century and could be classified as a climatic change.

A somewhat closer look at the precipitation-runoff relationship was carried out for four basins. In each basin less than 130 square miles (Fraser, Blue, Ashley Creek), only one station measuring precipitation occurred, and in the largest (Animas-Durango) three stations (Silverton + Trout Lake + Cascade) were used. The runoff amounts by water-year were expressed in terms of specific yield or equivalent inches of water over the entire watershed. In addition, a quantity termed the runoff ratio, the quotient of the specific yield and the November through April precipitation was also calculated. This quantity was introduced by McDonald (26) and follows his definition. It is a simple expression for the year-by-year efficiency of the runoff process. Those quantities for the four small basins are given in Table II-3.

Plotted in Figs. II-(1-4) are the April to July runoff against the November - April precipitation for the four basins. Subjective examination of these figures confirms what has been pointed out by many workers, namely, that the residuals from a regression line of precipitation on runoff are not random, independent quantities. For example, for Ashley Creek at Vernal every year from 1931 through 1937 is below the regression line and 1921-1926 above the line. Roughly, the same conclusion can be drawn about the Blue at Dillon. This property, however, does not seem to be as marked in the other two basins. Oltman and Tracy (30) and Peck (32) have noted the tendency for the regression line to shift, so to speak, with time and have suggested that low-flow (winter) streamflow values be used as an additional variable in the regression equation.

For our purposes, however, the fact that the residuals from a precipitation-runoff regression line are not random simply indicates that other hydrological forces are playing an important role in the precipitation-runoff process, and that the time-dependent structure of precipitation and stream runoff time-series is of utmost importance.

TABLE II-1. COEFFICIENT OF VARIATION
Upper Colorado Basin Precipitation

Station	Year's Record	Winter Coefficient of Variability		
		Nov-Apr S^2	Nov-Apr Mean	Coef. of Var.
Silverton	52	16.69	12.57	0.32
Trout Lake	43	14.27	14.38	0.26
Shoshone	49	9.92	10.70	0.29
Dillon	44	5.85	9.47	0.20
Fraser	42	7.526	10.92	0.25
Gunnison	54	2.864	4.55	0.37
Ignacio	46	6.14	6.91	0.36
Delta	52	1.166	3.42	0.32
Montrose	56	1.572	3.92	0.32
Grand Junction	61	1.419	3.88	0.31
Laketown, Utah	60	9.586	7.41	0.42
Silverton + Trout Lake + Ames	43	96.195	38.14	0.26

Seven station sum = 0.21

TABLE II-1. COEFFICIENT OF VARIATION
Upper Colorado Basin Runoff

Streamflow Station	Year's Record	Water-Year Coefficient of Variability Water-Year S^2	Water-Year Mean		Coef. of Var.
Lee Ferry	44	16.61	13.28	m.a.f.	0.31
Lee Ferry (virgin)	61	18.70	15.03	m.a.f.	0.28
Lee Ferry April-July	44	11.299	9.12	m.a.f.	0.37
San Juan (Bluff)	43	0.845	2.04	m.a.f.	0.45
Main Stem (Cisco)	46	3.355	5.92	m.a.f.	0.31
Green (Green River)	53	2.437	4.81	m.a.f.	0.32
Fraser	47	46.733	30.01	t.a.f.	0.23
Blue	47	411.04	85.00	t.a.f.	0.24
Gunnison (Gunnison)	54	.14703	1.322	m.a.f.	0.29
White (Meeker)	48	10054.	460.8	t.a.f.	0.24
Main Stem (Hot Sulphur Springs)	53	12465.	497.5	t.a.f.	0.23
Roaring Fork (Glenwood)	52	.074398	1.053	m.a.f.	0.26
Taylor	47	4675.	253.4	t.a.f.	0.27
San Juan (Rosa)	48	143380.	896.98	t.a.f.	0.42
Animas (Durango)	46	38703.	622.25	t.a.f.	0.32

TABLE II-2. RANK CORRELATION SUMMARY

Precipitation:	R	Approximate Distance Apart (miles)
Trout Lake - Ames	.81	5
Silverton - Trout Lake	.79	12
Silverton - Ames	.78	15
Delta - Montrose	.75	20
Dillon - Fraser	.62	27
Grand Junction - Delta	.49	35
Shoshone - Dillon	.49	60
Gunnison - Grand Junction	.48	95
Shoshone - Silverton	.45	125
Shoshone - Ignacio	.42	170
Gunnison - Silverton	.36	65
Grand Junction - Silverton	.34	100
Gunnison - Ignacio	.32	105
Dillon - Gunnison	.27	90
Silverton - Fraser	.10	180
Grand Junction - Fraser	.06	160
Ignacio - Dillon	-.12	190
Grand Junction - Dillon	-.24	140
Precipitation vs. Runoff:		
Animas (Durango) - Silverton + Cascade + Trout Lake		
Nov-Apr precip. - Apr-Jul runoff	.67	
Nov-Apr precip. - water-year runoff	.65	
Fraser - Fraser		
Nov-Apr precip. - Apr-Jul runoff	.64	
Nov-Apr precip. - water-year runoff	.63	
Blue - Dillon		
Nov-Apr precip. - Apr-Jul runoff	.56	
Nov-Apr precip. - water-year runoff	.51	
Gunnison - Gunnison		
water-year runoff	.32	
Ashley Creek - Vernal, Utah		
Nov-Apr precip. - Apr-Jul runoff	.71	
Nov-Apr precip. - water-year runoff	.67	
Runoff:		
Gunnison - Roaring Fork	.93	
Roaring Fork - Main Stem (Hot Sulphur Springs)	.82	
Main Stem - San Juan (Bluff)	.85	
Gunnison - Main Stem (Hot Sulphur Springs)	.82	
Fraser - Blue	.80	
Fraser - White	.80	
Fraser - Taylor	.76	
Taylor - White	.73	
White - Animas (Durango)	.68	
Green - San Juan (Bluff)	.61	
Main Stem (Hot Sulphur Springs) - San Juan (Rosa)	.52	

TABLE II-3

SILVERTON + TROUT LAKE + CASCADE PRECIPITATION ANIMAS RIVER AT DURANGO

Drainage area = 692 square miles

Water Year	Runoff in Inches	Apr-Jul Runoff in Inches	Nov-Apr Runoff in Inches	NA Precip. / Total Runoff	NA Precip. / AJ Runoff
1914	22.57	17.79	48.86	2.16	2.74
1915	18.59	14.05	37.40	2.01	2.66
1916	23.69	17.16	67.69	2.85	3.94
1917	26.77	19.08	34.40	1.28	1.80
1918	14.39	10.04	31.79	2.20	3.16
1919	18.94	14.51	44.80	2.36	3.08
1920	27.70	22.52	41.94	1.51	1.86
1921	24.82	18.26	36.49	1.47	1.99
1922	21.50	17.51	50.96	2.37	2.91
1923	18.14	13.44	40.14	2.21	2.98
1924	14.74	10.92	34.23	2.32	3.13
1925	14.84	9.70	47.91	3.22	4.93
1926	17.43	12.56	40.71	2.33	3.24
1927	22.03	14.05	46.24	2.09	3.29
1928	15.18	10.04	29.82	1.96	2.97
1929	20.89	13.77	46.51	2.22	3.37
1930	14.68	9.82	23.81	1.62	2.42
1931	7.89	4.98	19.77	2.50	3.96
1932	20.10	15.27	50.43	2.50	3.30
1933	11.68	8.67	26.66	2.28	3.07
1934	6.77	4.18	23.65	3.49	5.65
1935	15.37	11.84	43.12	2.80	3.64
1936	14.16	10.06	38.54	2.72	3.83
1937	14.65	11.34	39.12	2.67	3.44
1938	19.23	15.32	59.27	3.08	3.86
1939	11.55	7.21	35.10	3.03	4.86
1940	9.77	6.74	39.28	4.02	5.82
1941	25.71	20.18	59.00	2.29	2.92
1942	22.54	14.63	45.08	2.00	3.08
1943	14.59	10.46	49.78	3.41	4.75
1944	20.81	17.19	56.87	2.73	3.30
1945	14.83	11.19	43.94	2.96	3.92
1946	11.43	7.98	34.13	2.98	4.27
1947	16.96	11.90	42.71	2.51	3.58
1948	20.84	15.91	56.50	2.71	3.55
1949	21.00	17.56	51.11	2.43	2.91
1950	11.12	7.92	38.42	3.45	4.85
1951	8.79	6.09	37.87	4.30	6.21
1952	22.03	18.53	72.72	3.30	3.92
1953	10.62	7.56	36.24	3.41	4.79
1954	9.86	6.93	25.21	2.55	3.63
1955	11.10	7.34	31.21	2.81	4.25
1956	10.26	7.52	44.43	4.33	5.90

r = 0.67 April-July runoff vs. precipitation

r = 0.65 water-year runoff vs. precipitation

TABLE II-4

FRASER PRECIPITATION, FRASER RIVER NEAR WINTER PARK

Drainage Area = 27.6 Square Miles

Water Year	Runoff in in./sq. mi.	April-July Runoff in in./sq. mi.	Nov. - Apr. Precip. in Inches	NA Precip. / Total Runoff	NA Precip. / AJ Runoff
1916	20.94	15.35	12.41	.59	.80
1917	21.32	16.02	10.69	.50	.66
1918	29.47	24.65	13.62	.46	.55
1919	16.16	10.59	5.48	.33	.51
1920	20.57	15.14	11.36	.55	.75
1921	26.75	21.52	15.06	.56	.69
1922	17.11	12.90	11.25	.65	.87
1923	22.20	17.18	11.39	.51	.66
1924	20.57	16.02	9.32	.45	.58
1925	19.28	13.31	7.21	.37	.54
1926	26.62	20.64	10.11	.37	.48
1927	21.39	15.35	13.18	.61	.85
1928	27.70	21.39	11.79	.42	.55
1929	23.83	16.70	9.05	.37	.54
1930	22.00	15.71	13.14	.59	.83
1931	14.33	9.98	5.85	.40	.58
1932	17.59	13.17	13.26	.75	1.00
1933	22.88	18.40	13.40	.58	.72
1934	14.46	10.73	6.65	.45	.61
1935	18.33	13.92	7.15	.39	.51
1936	22.81	16.64	13.03	.57	.78
1937	16.43	11.48	7.81	.47	.68
1938	23.56	17.72	14.07	.59	.79
1939	17.31	13.04	11.85	.68	.90
1940	15.01	10.94	10.23	.68	.93
1941	19.22	14.37	13.02	.67	.90
1942	20.91	16.50	14.16	.67	.85
1943	17.79	13.56	10.75	.60	.79
1944	15.14	11.52	8.41	.55	.73
1945	17.04	11.71	10.64	.62	.90
1946	16.64	12.24	7.32	.43	.59
1947	23.15	17.57	11.57	.49	.65
1948	19.83	14.74	10.60	.53	.71
1949	21.80	17.19	11.47	.52	.66
1950	14.60	10.61	10.63	.72	1.00
1951	18.54	14.42	15.08	.81	1.04
1952	21.18	16.63	13.16	.62	.79
1953	15.14	11.13	7.23	.47	.64
1954	9.10	5.85	5.69	.62	.97
1955	14.33	9.54	10.39	.72	1.08
1956	16.98	13.64	14.23	.83	1.04
1957	25.73	20.02	15.73	.61	.78

$r = 0.64$ April-July runoff; precipitation

$r = 0.63$ water-year runoff; precipitation

TABLE II-5

DILLON PRECIPITATION, BLUE RIVER AT DILLON

Drainage Area = 129 Square Miles

Water Year	Runoff in in./sq. mi.	April-July Runoff in in./sq. mi.	Nov.-Apr. Precip. in Inches	N.-A Precip. / Total Runoff	NA Precip. / AJ Runoff
1914	19.56	14.99	10.25	.52	.68
1915	12.08	8.59	5.98	.49	.69
1916	12.48	8.21	10.93	.87	1.33
1917	15.31	11.45	9.12	.59	.79
1918	16.33	13.10	10.85	.66	.82
1919	10.11	7.19	5.46	.54	.75
1920	12.80	9.27	9.50	.74	1.02
1921	17.77	13.11	12.03	.67	.91
1922	10.57	6.63	8.27	.78	1.24
1923	16.27	11.76	8.44	.51	.71
1924	12.97	9.06	7.30	.56	.80
1925	10.30	6.34	6.42	.62	1.01
1926	17.47	13.37	13.54	.77	1.01
1927	13.34	9.69	11.30	.84	1.16
1928	15.24	11.45	9.92	.65	.86
1929	12.15	7.71	8.34	.68	1.08
1930	11.99	7.63	5.58	.46	.73
1931	9.45	6.58	5.44	.57	.82
1932	11.01	7.99	9.52	.86	1.19
1933	10.22	7.60	11.71	1.14	1.54
1934	7.90	5.39	9.37	1.18	1.73
1935	9.48	6.58	9.85	1.03	1.49
1936	15.87	11.66	17.56	1.10	1.50
1937	8.18	5.64	6.05	.73	1.07
1938	12.86	9.52	10.59	.82	1.11
1939	11.27	8.50	10.94	.97	1.28
1940	7.13	5.03	7.80	1.09	1.55
1941	10.25	7.52	10.11	.98	1.34
1942	11.32	8.40	10.44	.92	1.24
1943	11.22	8.52	10.92	.97	1.28
1944	9.09	7.08	8.99	.98	1.26
1945	10.88	7.19	10.71	.98	1.48
1946	10.70	7.75	9.28	.86	1.19
1947	15.75	12.09	11.67	.74	.96
1948	13.57	10.35	8.88	.65	.85
1949	13.78	10.52	8.24	.59	.78
1950	10.35	7.55	8.00	.77	1.05
1951	14.82	11.35	12.68	.85	1.11
1952	12.84	9.78	11.78	.91	1.20
1953	12.17	9.26	9.89	.81	1.06
1954	5.77	3.69	5.25	.90	1.42
1955	8.90	5.20	7.09	.79	1.36
1956	11.62	8.84	11.97	1.03	1.35
1957	15.95	11.74	8.93	.55	.76

r = 0.56 April-July runoff; precipitation

r = 0.51 water-year runoff; precipitation

TABLE II-6

VERNAL, UTAH, PRECIPITATION, ASHLEY CREEK AT VERNAL, UTAH

Drainage Area = 101 Square Miles

Water Year	Runoff in in./sq. mi.	April-July Runoff in in./sq. mi.	Nov.-Apr. Precip. in Inches	NA Precip. / Total Runoff	NA Precip. / AJ Runoff
1920	16.93	12.78	6.49	.38	.50
1921	23.94	18.58	4.65	.19	.25
1922	23.57	17.81	6.94	.29	.38
1923	18.75	13.68	5.31	.28	.38
1924	10.58	6.68	2.15	.20	.32
1925	10.86	6.71	.90	.08	.13
1926	13.49	8.23	2.09	.15	.25
1927	16.02	10.14	4.60	.28	.45
1928	16.26	10.61	3.23	.19	.30
1929	18.04	13.08	5.67	.31	.43
1930	15.61	9.62	2.53	.16	.26
1931	7.91	3.76	2.82	.35	.75
1932	13.72	9.85	4.65	.33	.47
1933	9.00	5.71	3.44	.38	.60
1934	5.81	3.26	2.72	.46	.83
1935	11.87	8.76	6.52	.54	.74
1936	7.74	4.56	1.91	.24	.41
1937	14.59	10.78	6.67	.45	.61
1938	14.30	10.42	4.49	.31	.43
1939	12.32	6.85	4.92	.39	.71
1940	10.01	6.15	4.35	.43	.70
1941	17.15	11.77	5.10	.29	.43
1942	18.75	11.75	3.81	.20	.32
1943	11.74	8.20	2.85	.24	.34
1944	17.41	13.44	4.99	.28	.37
1945	11.60	7.55	3.47	.29	.45
1946	8.78	5.38	2.86	.32	.53
1947	17.11	12.26	4.17	.24	.34
1948	12.59	8.69	2.59	.20	.29
1949	14.93	11.56	4.12	.27	.35
1950	16.22	11.77	5.20	.32	.44
1951	11.06	6.94	2.13	.19	.30
1952	19.02	13.74	5.63	.29	.40
1953	10.78	6.82	3.09	.28	.45
1954	9.91	7.00	2.61	.26	.37
1955	9.18	6.28	1.81	.19	.28
1956	10.83	8.07	2.82	.26	.34
1957	13.21	9.37	2.73	.20	.29
1958	12.38	8.87	4.15	.33	.46
1959	7.72	5.23	2.38	.30	.45

r = 0.71 April-July runoff; precipitation

r = 0.67 water-year runoff; precipitation

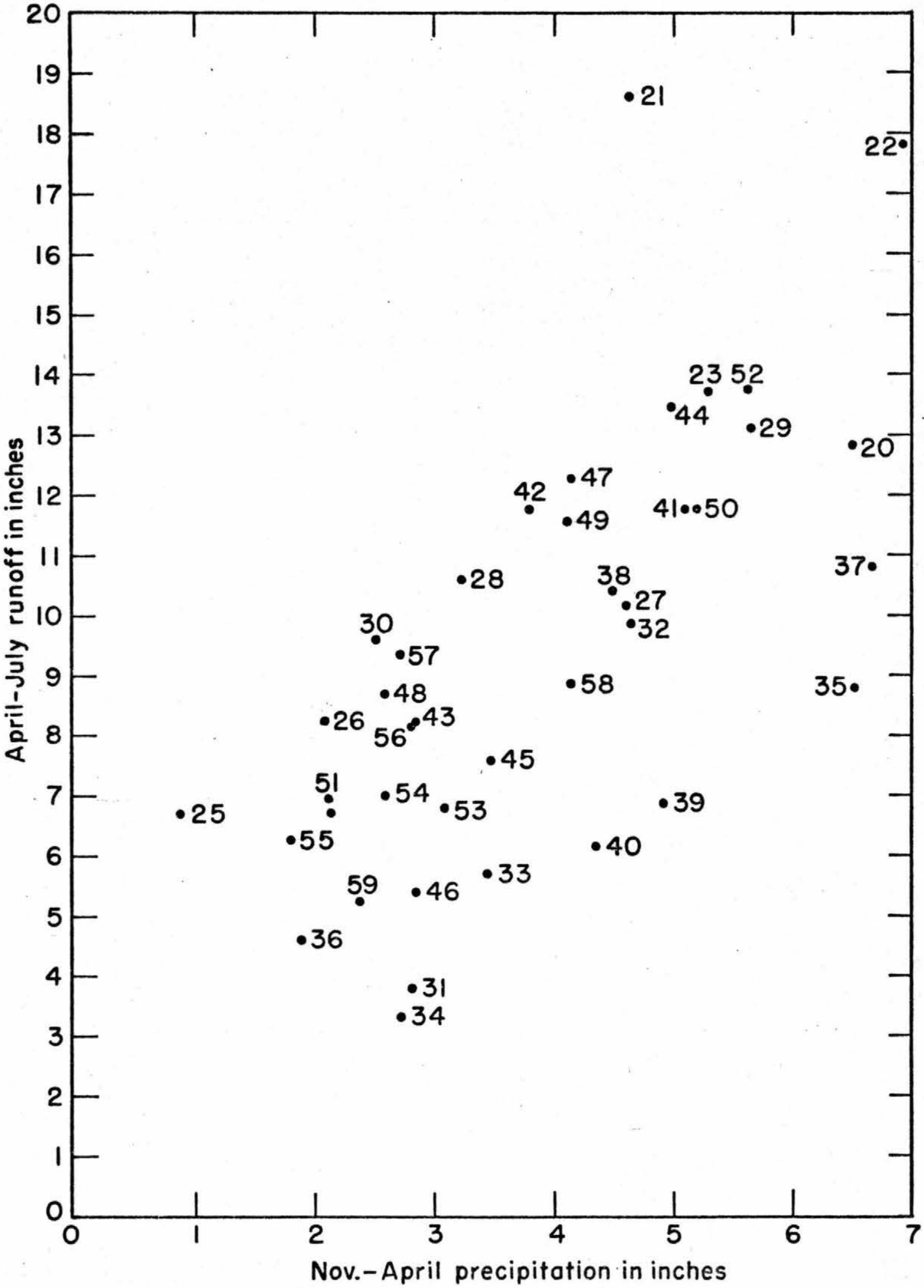


Figure II-1. Ashley Creek at Vernal, Utah - Vernal, Utah precipitation.

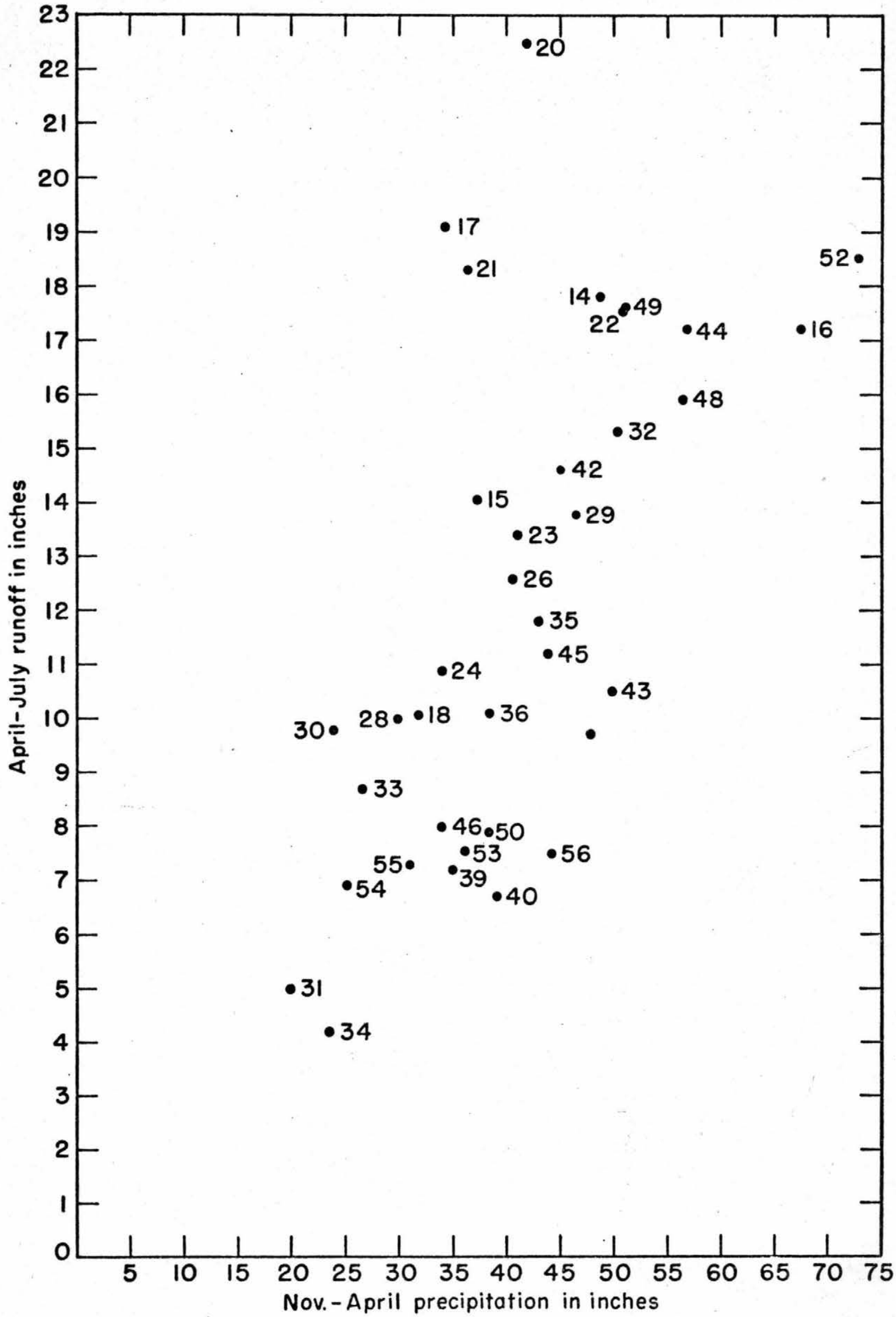


Figure II-2. Animas at Durango - Trout Lake + Silverton + Cascade precipitation.

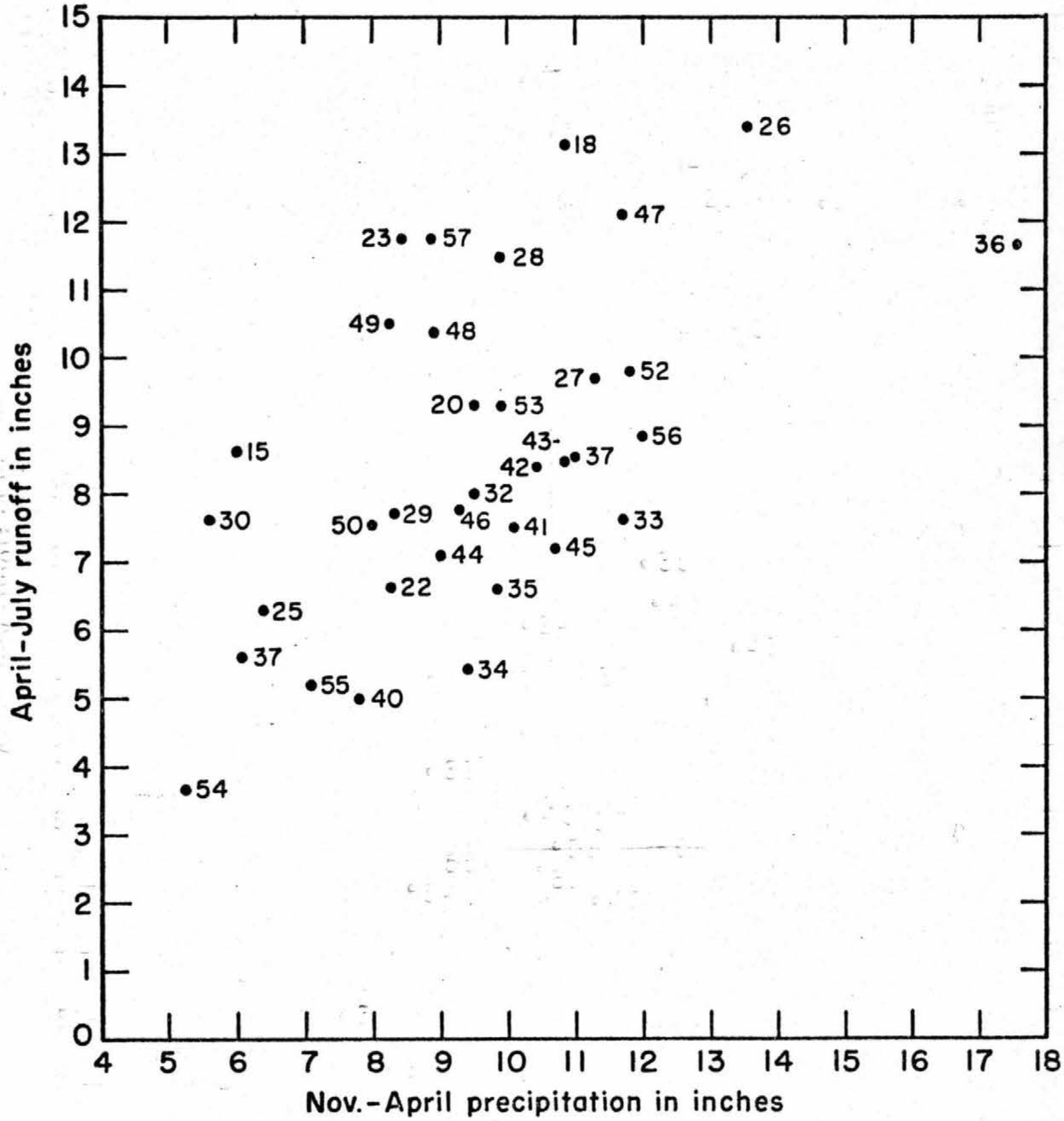


Figure II-3. Blue River at Dillon—Dillon precipitation.

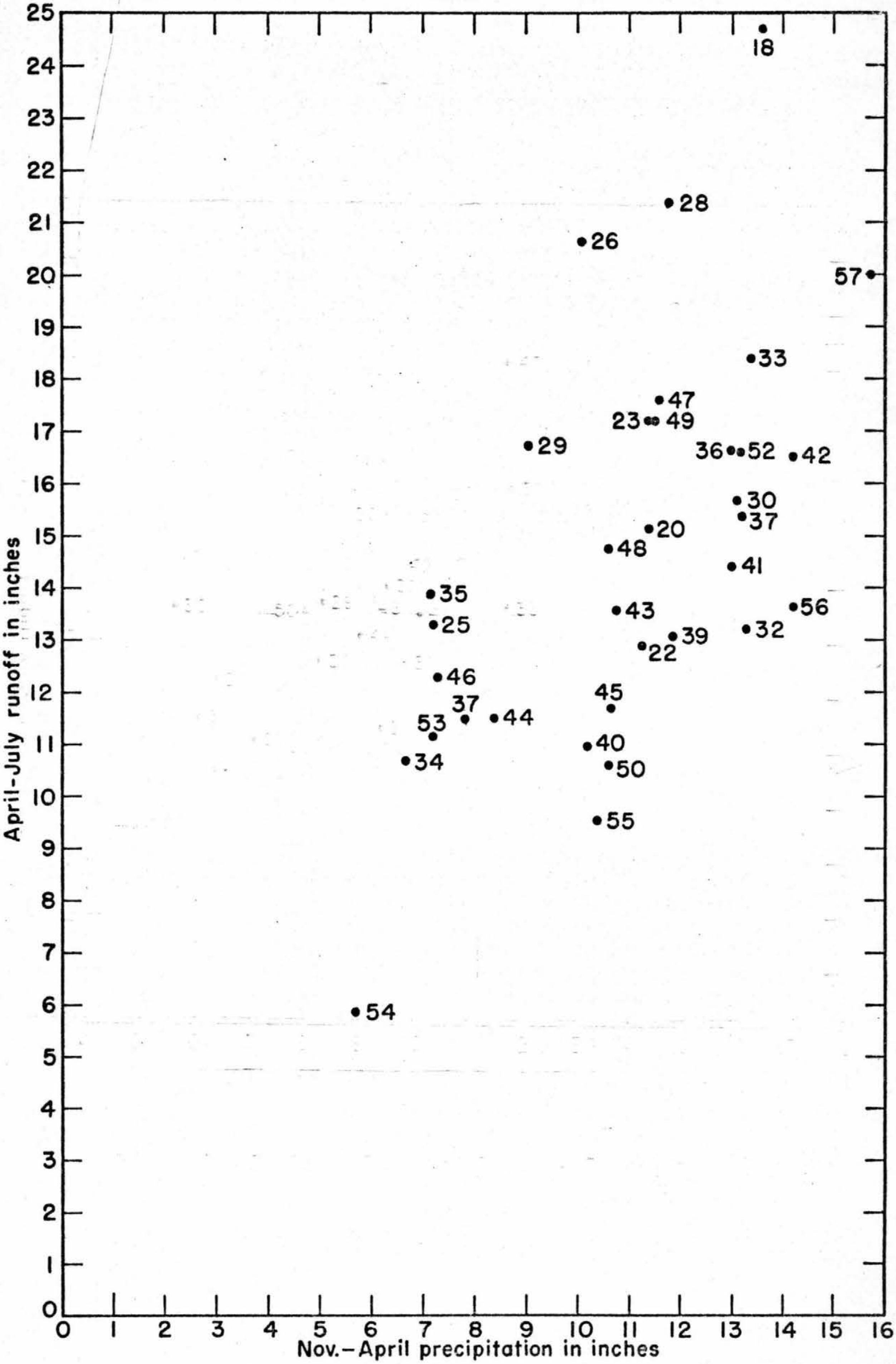


Figure II-4. Fraser River - Fraser precipitation.

III. STATISTICAL ANALYSIS OF PRECIPITATION AND STREAMFLOW DATA

A. Introduction

The primary purpose of this and the next few sections is to attempt to answer question two stated in the Introduction: Is there any statistical predictability in precipitation and runoff series or are they indistinguishable from random numbers? Before beginning a discussion of the approach taken in this report a few facts concerning statistical analysis of time-series should be presented.

B. Hypothesis Testing

The first major subject to be discussed is that of hypothesis testing in time-series analysis. Only a brief discussion of certain applicable points will be made here as this subject is a complicated and involved one. The primary object in analyzing natural time-series is, of course, to be able to make statements concerning the presence or absence of regularities or uniformities in the series which then determines the degree of dominance of the random component in the series. There are many tests for randomness in time-series, and in a study such as this all are not appropriate and all those that are cannot be performed. The particular tests which are presented here are those the author chose to use on the basis of his knowledge. It is an extremely difficult problem to determine the 'best' test of randomness in a particular situation. This problem reaches into rather involved statistical theory and will not be gone into here. However, it is important to note all tests of randomness applied to the data are shown here, and not just those giving results favorable to one position or the other.

Friedman (11) in his paper examining the randomness of precipitation data used a series of non-parametric tests; these tests are called non-parametric because no assumption need be made, or information had, about the form of the frequency distribution from which the sample was drawn. This factor is a decided advantage as we need not worry about what the true frequency distribution might be. One of the three tests used by Friedman, that of Wald and Wolfowitz (40), is used in this report. A great many of the tests for randomness are tests on the serial correlation coefficients--that is, the series correlated with itself displaced in time. The serial correlation coefficients for a random series would by definition be zero. It can be shown, however, that the first serial correlation computed from a sample of data is not a good estimate of the true population serial correlation coefficient.

The more sophisticated tests such as that described by Wald and Wolfowitz (40) attempt to modify the definition of the serial correlation coefficient slightly to produce a statistic which does give a good estimate of the true serial correlation coefficient. This test is used and explained in Section VIII.

The application of the randomness tests will be fully explained in the appropriate section. In all cases, however, the null hypothesis will be tested at the 5% level. This means that we will accept the hypothesis that the time-series tested consists of numbers sequentially drawn at random from some frequency distribution, and will accept the risk inherent in the testing of statistical data that five times out of a hundred, on the average, a false conclusion concerning the null hypothesis will be made, namely that the null hypothesis is rejected when it is actually true. We are thus testing essentially how often a series of random numbers resembles the numbers making up our time-series. If the series in hand is sufficiently unlike a series of random numbers the null hypothesis will be rejected, and we can say with some degree of confidence how often a mistake will be made in doing this. Note, then, that the testing of a single statistical hypothesis involves (1) the proving of a hypothesis false, and (2) a certain possiblilty that the conclusion reached is the wrong one. Such is the nature of statistical reasoning.

C. The Definition of 'Time'

The second major point to be discussed is that involving time. We must ask the question, "What will be the nature of the statistical model to be used to forecast the variable quantity?" The reason for this question appropos the book by Moran cited in the Introduction is that in the case of a dam-reservoir system the time scale or time base used in the prediction scheme is of utmost importance. Specifically, will time be considered as discrete or continuous? If a model is chosen, which only approximates reality, in which the flow into the dam occurs all at once and the release from the reservoir also occurs instantaneously, the time scale is considered discrete. Such a discrete time scale may be applied to weekly, monthly, or annual data, but it stipulates that all inflow and drafts on storage occur instantanously once each week, month, or year. A model using a continuous time base acknowledges that in nature the flow into a reservoir is continuous and so in all probability are the outflows. Why, then, is a discrete time scale used at all? There are a number of reasons: First of all, the mathematics for the discrete model are much simpler and solutions can be obtained in a wider variety of cases and situations. Secondly, some things of interest which occur in nature by their very nature do not occur on a continuous time scale. Winter (or seasonal) precipitation values, for example, represent quantities which do not occur on a continuous time base. There is obviously a six-month gap between the six-month (October - April) precipitation totals used in this report, and these quantities, then, are on a discontinuous or discrete time scale.

Thus, it is imperative to know the structure of the time-scale that the statistical model will assume. And the answer to the question of the randomness of the series will depend also on whether time is to be considered discrete or continuous. The runoff amounts at Lee

Ferry by water year can be looked upon as discrete numbers; that is, each year's runoff can be assumed to occur all at once, say, on each September 30th. If this discrete time scale is used the numbers giving the water-year totals can be subjected to statistical tests for randomness as they stand. The conclusions resulting from such tests however only apply to a situation involving discrete time. Conversely, if the flows at Lee Ferry are to be characterized on a continuous time scale, then appropriate measures must be taken to insure that the averaging effects (which are necessitated in expressing a continuous function by discrete values) are corrected for. If such corrections can be made, randomness tests may be run on the numbers, which now represent a continuously varying quantity.

The point of this discussion is to clarify what is meant by randomness when the question, "Are serial stream-runoff values at Lee Ferry indistinguishable from random numbers?" is asked; randomness in time necessitates a definition of the time scale.

IV. VARIANCE SPECTRA AS A TEST OF RANDOMNESS

A. Introduction

The principal point of departure in this report in the examination of the time-series of precipitation and stream-runoff is the use of variance spectra. The term 'variance spectra' is applied to a special form of harmonic analysis of the time-series; the term spectra comes from the fact that the total variance of the series is depicted as a function of frequency of oscillation. The advantages of this representation are: (1) Any periodic movement or cycle contained in the record will show up as a narrow spike in the spectrum occuring at the particular frequency or period of the cycle. (2) Any deviation from randomness, and this includes the periodic movement just mentioned, will be determined by the distribution of variance over a frequency range. A random series, by definition, would have no more variance associated with any one frequency that with any other. A spectrum of a series of random numbers would thus be a constant value over the frequency scale. A third advantage for rather sophisticated analysis is that the shape of the spectrum is rather easily related to various time-series models that have been proposed, mainly the auto-regressive and moving average model.

The auto-regressive scheme can be mathematically represented by $X_t = aX_{t-1} + bX_{t-2} + \ldots + \epsilon_t$, where $X_{\text{subscript}}$ is the value of the variable at the time given by the subscript, and the ϵ's are a series of random uncorrelated numbers, and a,b,c, etc., are constants. The moving average scheme can be represented by

$$X_t = a\epsilon_t + b\epsilon_{t-1} + c\epsilon_{t-2} + \ldots.$$

Both of these schemes involve a random component, the ϵ's, together with a component which depends upon either the past history of the series or upon some specified combination of the random component. It should be noted that either of these schemes allow a statistical prediction to be made. A great deal of previous work on the statistical properties of these model series has been carried out. See, for example, Wold (43).

The use of variance spectra in this report will thus be to examine the questions: (1) Are time-series of naturally occuring precipitation and streamflow random, uncorrelated, or is some statistical predictability present? (2) Can either the auto-regressive or moving-average scheme be used to simulate the flow of the Colorado River or its tributaries?

B. Mathematical Basis of the Spectrum

A brief mathematical description of spectrum analysis will be given below. For more complete details the reader is referred to Lee (21), Hanna (12), and to Blackman and Tukey (6). Ordinary harmonic analysis should be familiar to most engineers and those having some advanced mathematics courses. A continuous periodic function f(t), that is, a function which repeats itself exactly every T units of time, can be expressed by an infinite number of sine and cosine terms of increasing frequency. Using complex notation for simplicity,

$$f(t) = \sum_{n=-\infty}^{\infty} F(n)e^{in\omega_1 t}$$

in which $F(n) = \frac{1}{T_1} \int_{-T/2}^{T/2} f(t)e^{-in\omega_1 t}dt \qquad n = 0, \pm 1, \pm 2, \ldots \; \omega_1 = \frac{2\pi}{T}$

and is called the complex spectrum of f(t). Note that the harmonic order n assumes only discrete values; this comes about because of the strict periodicity of the function f(t) assumed. Such a discrete spectrum is called a line spectrum. It is important to note that when f(t) is not specified as a continuous function of time, as in practice it rarely is, the complex line spectrum is restricted to a finite number of harmonics; namely, a number equal to the number of discrete values used to approximate the function f(t) within the period of record T.

Norbert Weiner, in a work published a number of years ago (41), showed how the concept of harmonic analysis could be extended to functions which were not assumed to be periodic outside the length of record by use of the Fourier integral. Instead of (1) and (2), we now have

$$f(t) = \int_{-\infty}^{\infty} F(\omega)\, e^{i\omega t}d\omega, \qquad F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt$$

Here $F(\omega)$ is a continuous function of frequency and is called the complex continuous spectrum of the function f(t). Instead of synthesizing the original record f(t) by an infinite number of finite harmonics at particular frequencies as in the periodic case, the continuous spectrum represents an aggregate of infinitesimal sinusoids of all frequencies from $-\infty$ to ∞. The same remarks concerning the representation of f(t) by a finite number of equally-spaced values as applied in the periodic case also apply here. In other words, in the case that time is made discrete,

$$f(t) = \int_{-\pi}^{\pi} F(\omega)e^{i\omega t}d\omega,$$

and the frequency domain is restricted to $-\pi$ to π, using an angular frequency structure such that

$$T_1 = \frac{2\pi}{\omega_1} \quad .$$

Thus, when a discrete time scale is used, oscillations with frequencies outside the range $-\pi$ to π are indistinguishable from oscillations with frequency inside the range according to $\cos(\omega + 2k\pi)t = \cos \omega t$, where t are positive integers and $k = 0, \pm 1, \pm 2, \ldots$ etc. This property of confusion of frequencies in discrete time is called 'aliasing' by Blackman and Tukey (6). If the continuous record being analyzed is approximated by values in discrete time some attention must be given to this problem.

If we now define a new quantity $\Phi(\omega) = 2\pi \left| F(\omega) \right|^2$ and call it the power (or variance) density spectrum, then the total 'energy' or variance of the original aperiodic function f(t) is $\int_{-\infty}^{\infty} \Phi(\omega)\,d\omega$. The use of negative frequencies in the foregoing theory is simply for mathematical convenience since, of course, a negative frequency has no physical meaning. In actual practice the spectrum is depicted over a frequency range from 0 to ω', where ω' is called the Nyquist frequency and is equal to $\frac{1}{2\Delta t}$, Δt being the interval between successive observations. This 'folding' of the negative frequency scale onto the positive scale results in the plotting of a quantity

$$2\int_{\omega a}^{\omega b} \Phi(\omega)\,d\omega \text{ which comes from } \left[\int_{-\omega b}^{-\omega a} + \int_{\omega a}^{\omega b}\right] \Phi(\omega)\,d\omega$$

which then gives the total variance or 'energy' of the series between any two frequencies ωa and ωb. Representation of the variance on such a frequency scale, it should be kept in mind, requires that while the amount of variance in any frequency band is finite, the variance at any particular frequency is zero.

The details of estimating the spectra from a data sample will not be discussed in detail here since they are available from a wide source of the literature (6) (12) (24). Suffice it to say that the theory regarding the estimation of variance spectrum was outlined by Bartlett (5), Tukey (39), and others apparently more or less independently. The basic idea is a harmonic analysis of the autocorrelation or serial correlation function, and the mechanics of the computation can be shown to be related to the calculation of a smoothed periodogram. The periodogram, originally developed by Schuster, (34) is simply a harmonic decomposition of the original record by a finite number of sines and cosines as in (1). Plotted against the order of the harmonic are the coefficients of the Fourier terms. Thus,

$$f(t) = \bar{f} + \sum_{i=1}^{n/2} \left[A_i \sin \frac{it}{T} + B_i \cos \frac{it}{T}\right] \quad \text{and}$$

$$P_e = \frac{A_i^2 + B_i^2}{2},$$

where T equals the total length of the series consisting of N observations or data points and P_e are the periodogram estimates.

The periodogram enjoyed a wide usage in natural time-series for a good many years. Bartlett, however, showed that if the length of data, T, is not to be considered as strictly periodic and therefore a continuous spectrum is inferred, the periodogram is not a satisfactory tool for the estimation of the continuous spectrum. This undoubtedly is the explanation for the fact that the application of the sampling theory of the periodogram seldom met with success--that is, independent data, especially of economic and meteorological series, failed to confirm the existence of supposedly significant periodicities.

Bartlett (5) and others apparently independently suggested that a 'smoothed' periodogram would be a satisfactory estimate of the spectrum and showed how such a smoother harmonic analysis could be carried out using the autocorrelation function. The crux of the matter, however, is that the continuous variance spectrum is estimated by computing the first few terms of the ordinary harmonic analysis which would normally result in the 'line' spectrum mentioned above and then smoothing adjacent values on the frequency scale. Such an estimate is essentially a harmonic analysis using a broad 'window' for analysis on the frequency scale. This window, using the Tukey method, has an equivalent width of

$$\frac{1}{m\Delta t}$$

where Δt is the sampling interval or time between observations and m is the number of serial correlation coefficients used. The choice of m in the analysis is critical. It must be kept small in proportion to N, the total number of observations in the series, in order to maintain some degree of statistical reliability, but it should be large enough to resolve significant characteristics on the frequency scale. The greater m, the less is the statistical reliability of the estimates; they will fluctuate greatly from one adjacent frequency band to another.

The sampling theory of the spectral estimates has been the subject of some work by Tukey (39) and also Lominicki and Zaremba (24). In this paper the Tukey analysis will be used. The purpose of the sampling theory is to obtain confidence limits on the population or universe spectrum against which the estimated spectrum obtained from a truncated sample of data can be compared. Assuming that the numbers of the time-series are drawn from random from a normal distribution, Tukey has reasoned that the sample spectrum estimates should be distributed about the population spectrum according to the chi-squared distribution divided by the number of degrees of freedom. The actual number of degrees of freedom is somewhat difficult to determine but Tukey has given evidence that a good approximation is given by

$$\text{Number of degrees of freedom} = \frac{2N - m/2}{m}$$

where N equals the total number of observations in the time-series and m the number of serial correlation coefficients used. The preceding remarks concerning the statistical reliability of the spectral estimates as dependent on m should now be clear to the reader.

The sampling theory just presented makes one very important assumption about the nature of the time-series being examined. This assumption, which can be stated precisely in mathematical terms, can be qualitatively interpreted as stipulating that the statistics used in estimating the spectrum, namely the serial correlation coefficients, should be dependent only on differences in time, not upon the time-scale itself. That is, the probability laws controlling the course of the time-series through time should not change with time. A precise discussion of the physical implications of the mathematical-statistical requirements for stationarity, as the assumption is called, is a bit difficult at our present state of knowledge. It seems, however, that since we are unable to decide the validity of stationarity in meteorological time-series from physical grounds the best protection against non-stationarity is to have independent data with which to test the results. In the case of hydrometeorological records, the relatively brief span of time covered prevents independent-in-time-series being used for this purpose. Instead, in this paper it will be assumed that the series of precipitation and stream-runoff in the upper Colorado Basin can be said to be stationary time-series, and independent-in-space series can be used to check this assumption. For example, examination of winter precipitation series from other parts of the United States sufficiently far removed from the upper Basin as to be uncorrelated can be considered as independent data.

The author readily admits that the problem of stationarity may prove to be a serious one with regard to the analysis presented here. Much work of a theoretical nature needs yet to be done on this question.

V. RESULTS OF THE SPECTRAL ANALYSIS

A. Analysis of Winter (Nov. - April) Precipitation

Examination of the precipitation records for the stations located in western Colorado indicated that ten of them had sufficiently homogeneous records to be of some use. Each station was checked by means of independent double-mass techniques. Special credit should go to Mr. Eugene Peck of the U. S. Weather Burear Water Forecast Unit in Salt Lake City for supplying data to check with the author's own analysis. These ten stations were corrected for gauge location changes if the results of the double mass-curves agreed with historical moves. Table V-1 lists the stations, their lengths of records, elevation, and other pertinent data. Table V-2 lists the first six serial correlation coefficients for each station. These are used in the computation of the variance spectra and are shown for the benefit of those readers who are accustomed to using correlation analysis of time-series data. An interesting but probably not significant fact is that the majority of the coefficients for lags of one and two years are negative. None of the correlations shown are very large; none of the six average serial correlations given are greater than 0.1.

The spectral estimates for all ten stations are plotted in Fig. V-1. No effort is made to distinguish between stations in this Figure. The ordinate portrays normalized variance (total variance equal to unity) per frequency interval or band and the abscissa is marked as a frequency and period scale. For reasons too complicated to cover here, this frequency scale is only approximate: However, this fact in no way vitiates the conclusions to be drawn. For most spectra shown in this report seven frequency bands are used: a half-band at the zero frequency and Nyquist frequency (0.5 per year) ends of the scale, and six full bands in between. It is important to realize that the plotted points represent the estimate of the variance contained in the frequency band about which the point is plotted and not just at the particular frequency represented by the point.

A number of points need to be made clear at once. A 'white-noise' or constant spectrum of a series of random numbers would be at constant at 1.00/7 or about 0.143 ordinate value. The resolution or discrimination of adjacent frequencies of the spectra shown here is 0.0833 cycles per year which is rather broad. In other words, all the variance is being distributed over a finite frequency range, 0 per year to 0.5 per year, in just seven increments. It can be argued that such poor resolution is not sufficient to detect cycles or fluctuations with a given, constant period in the record. This objection is certainly true and can be answered as follows: in spite of various claims, no statistically significant cycle with a period greater than two years has been demonstrated to exist in any meteorological or hydrological record by any means whatever. The lengths of annual records, especially in this country, are simply too short to allow satisfactory determination of such cycles, if they exist at all. Thus, the analysis used

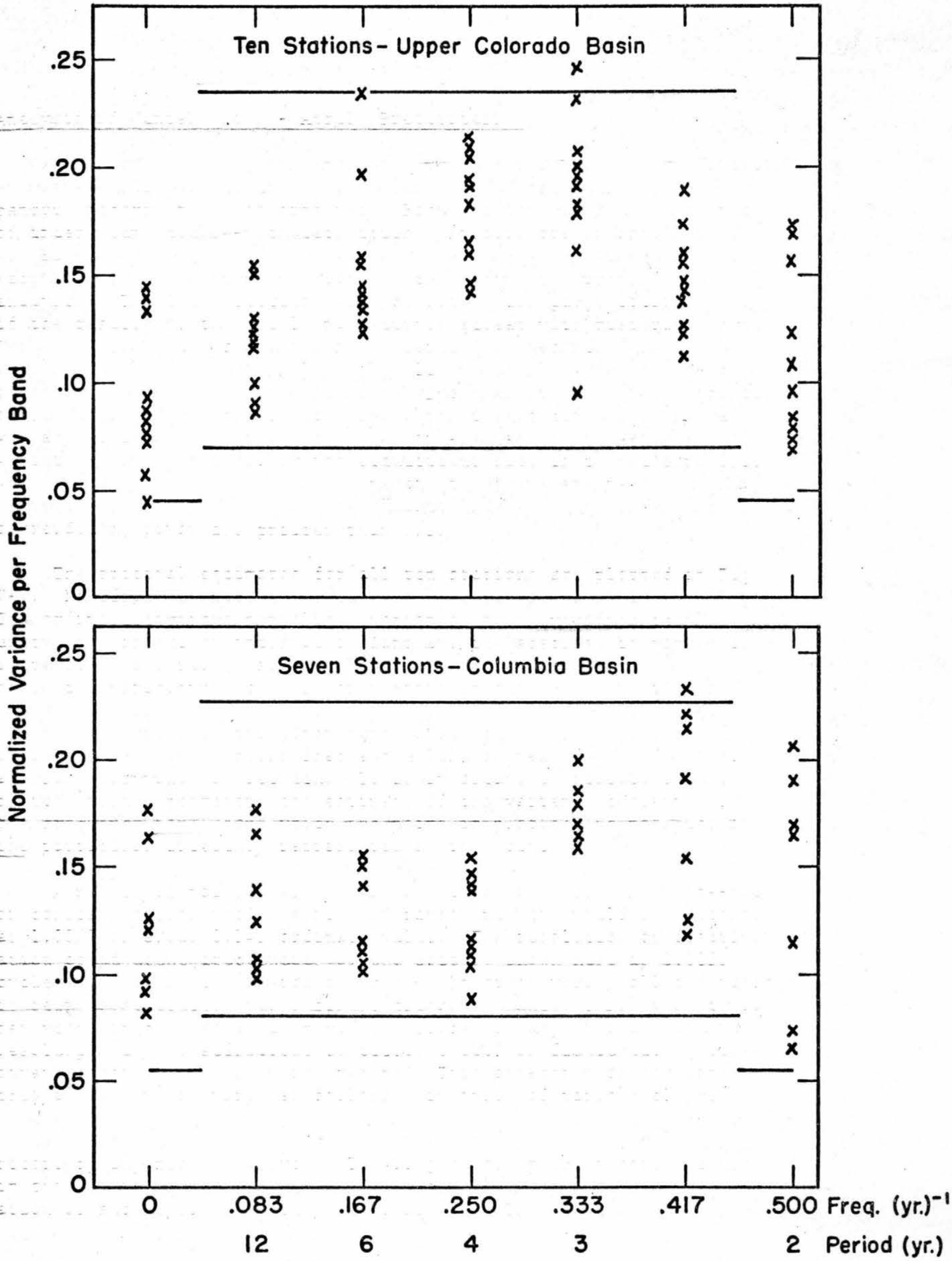


Figure V-1. Variance Spectra, Nov.-Apr. Precipitation, Discrete Time.

here recognizes the random appearance of precipitation and runoff data and analyzes the time-series as possessing fluctuations over a continuous distribution of frequencies.

Referring again to Fig. V-1, the following features are of some interest. The points of the spectra of the ten precipitation stations scatter about the ordinate value 0.143. Only two points in all of the ten spectra exceed the 5% confidence limits shown. These limits are for a random process with the uncorrelated elements drawn from a normal distribution according to Tukey's theory. The limits shown are based on an <u>average</u> length of record for the ten stations; when each station is considered individually only one of the points exceeds the 5% limits. We can safely conclude, then, that winter precipitation amounts over periods of time less than roughly 60-90 years are indistinguishable from a series of random numbers. Also of note is the fact that the spectral estimates in the lowest frequency band, that is the variance in the band 0 to 0.0833 per year, is not greater than we would expect from a random series. There is even a suggestion (not statistically significent) that there is systematically less variance here than we would expect from a random series.

This conclusion on the randomness of precipitation is in good agreement with other workers (11), (38), (18).

B. <u>Analysis of Water-year Stream-runoff Spectra - (Discrete time)</u>

Figs. V-2 through V-6 show the variance spectra for the streams in the upper Colorado Basin. Table V-3 gives the first six serial correlation coefficients for six of the smaller basins. The values used in all instances except where noted are those for water years, and time is taken as discrete.

Cofidence limits analagous to those shown for the precipitation spectra are given, in each case based on the length of the record for the stream shown.

First of all note Fig. V-5; these graphs contain spectra of various combinations of Lee Ferry data. Runoff amounts by water year, calendar year, and April-July runoff are shown for the gauged plus corrections for diversions and storage data for 1914-1957. Also shown is the spectra for the virgin flow data given by Leopold (22).

The main conclusions to be drawn are: (1) There is not much difference between the Lee Ferry spectra for any of the various data used except for the April-July totals. (2) the April-July totals contain about the same low frequency variance as the water-year spectrum and more medium frequency variance, with less of the high frequency component (about 2 years).

Figs. V-2 through V-4 giving the variance spectra of the smaller basins (discrete time) indicate that none of the spectra differ from a random number spectrum by Tukey's significance criteria on the 5% level. An interesting (and reassuring) feature is that nearly all of these spectra show a 'hump' at frequencies on the order of 0.33 year^{-1} which is also apparent in the precipitation spectra. Neither the runoff or precipitation spectra individually suggest that this hump is significant, but collectively the implication is that a bit more variance exists in this frequency band than elsewhere. Nothing will be claimed for this hump, however, as a universal property of hydrologic series, as examination of precipitation and stream-runoff records in other Basins fails to indicate and such 'hump'.

The spectra for the three major basins--the Colorado Main Stem, the Green, and the San Juan--are shown in Fig. V-4. The similarities are obvious and it should be noted that the peak at three years is the greatest in the San Juan and least in the Green. None of the three depart from a random number spectrum using the Tukey significance test.

C. Stream-runoff Spectra--(Continuous Time)

The results of the previous section apply only if time is discrete. The spectra shown, then, do not represent the true spectrum of a continuously varying runoff over the frequency range 0 to 0.5 per year. In order that they do represent the true spectrum in continuous time some notion must be gotten of the effects that aliasing and averaging (into water year totals, in this case) have upon the estimated spectrum. The considerations that apply here are rather complicated and are expanded upon in Appendix Two. The problem might simply be stated thusly: In continuous time the spectrum will be distributed over a continuous frequency range and will not be limited as in the case of discrete time to, for example, 0 to 0.5 per year. The true spectrum of continuously varying stream-runoff will have some variance (small, but nevertheless finite) associated with fluctuations on the order of seconds, minutes, hours, etc., on up to the range of years just considered. If we estimate the 'true' spectrum (for continuous time) using water-year totals then we must know the effect that averaging and aliasing have on our estimate; effects brought about because we are truncating the spectrum at 0.5 per year whereas the true spectrum contains variance at higher frequencies.

As brought out in Appendix Two there is no satisfactory method of making corrections for these effects unless the true spectrum is known. One practical solution, however, is to utilize data taken more frequently in time--in this case, monthly runoff data is used to estimate the spectrum from 0 to 0.5 per month. The aliasing and averaging effects are thereby greatly reduced since the frequency range 0 to 0.5 per year is but a small portion of the frequency range covered. This small protion is at the low frequency end of the spectrum where such effects are the smallest (see Appendix Two). This procedure has two advantages: (1) It allows a crude estimate of the continuous spectrum over the lower frequencies to be made, and (2) it is of interest in itself by allowing any peaks, or periodicities, in the spectrum to be revealed if they are present.

For this analysis monthly gauged flows at Lee Ferry from 1914-1957 were first subjected to a normalizing procedure to remove the annual variation and to create as near a stationary time-series as possible. The 44 years were stratified by month of the year, and each month's data were normalized so that the mean was zero and the variance unity. The spectrum of these normalized monthly data was then computed. Fig. V-6 presents the estimated spectra. No peak at a frequency of 1/12 months (period equal to one year) is noted since the normalizing precedure effectively removed the annual variation. The resulting spectrum is smooth; that is, no significant peaks are present. The variance exhibited by the normalized monthly series is strongly concentrated in the very low frequencies, corresponding to periods on the order of years. The spectra presented in the previous Figures, of course, covered a frequency range which is spanned in this Figure by the region to the left of the dotted line. The general decreasing variance from low to high frequencies in the range 0 to 0.5 per year indicated in Fig. V-5 is seen in this Figure. According to this wider frequency range the decrease continues into the higher frequencies but changes slope abruptly at about 1/12 per month (annual period).

Thus, the following conclusions seem warranted: (1) the continuous spectrum of the runoff process at Lee Ferry in the frequency range 0 to 0.5 per year is roughly given by the discrete spectrum, Fig. V-5 The effects of aliasing, which adds variance to the high frequency end of the spectrum estimated from annual data are apparently offset by the effects of averaging, which removes variance from the high frequency end. For a more complete discussion of this point, see Appendix Two. (2) Most of the serial correlation present between monthly runoff totals (considering the annual variation removed) is due to the very long-period movements (with periods on the order of years) in the record. (3) There are no significant periodicities (except, of course, the annual period) with periods between one year and two months.

The results and conclusions from the spectra shown can be summarized as follows: (1) The random component of streamflow is large and dominates all of the spectra. Only one set of data tested, namely the gauged plus published corrections for trans-mountain diversions and regulation water-year totals for Lee Ferry 1914-1957, differed significantly from a set of random numbers. This conclusion applies only to the numbers representing the water-year runoff amounts in discrete time. When considering the spectra between 0 frequency and 0.5 per year as a limited frequency range spectra of a process in continuous time extending over a much wider range of frequencies no such conclusion concerning statistical significance could be drawn. The spectrum of normalized monthly flows of gauged Lee Ferry data 1914-1957 also differed significantly from a random number spectrum. From this spectrum, moreover, the conclusion was drawn that the spectrum of annual or water-year flows in continuous time was not greatly different from that spectrum in discrete time.

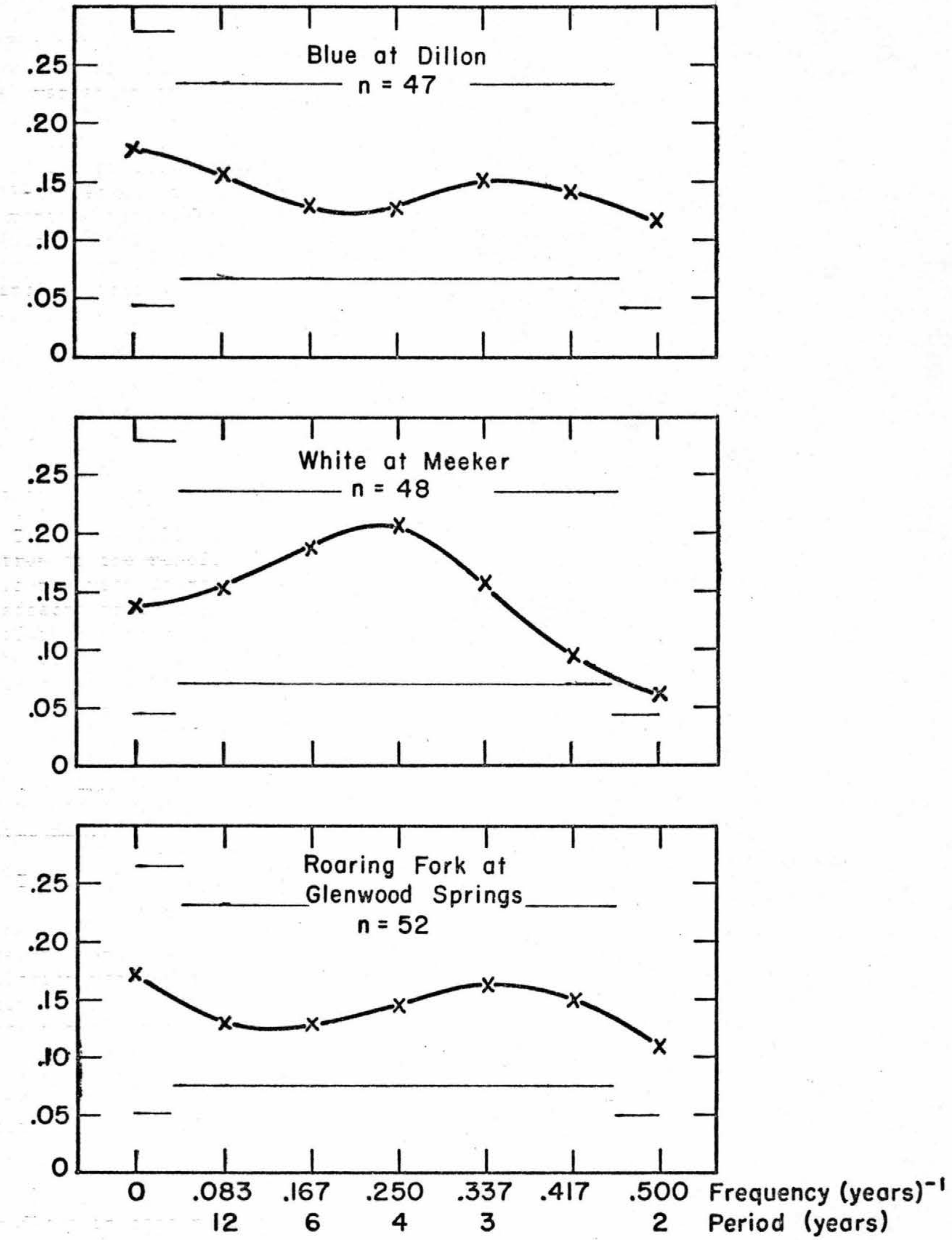


Figure V-2 Small Basins, Discrete Time

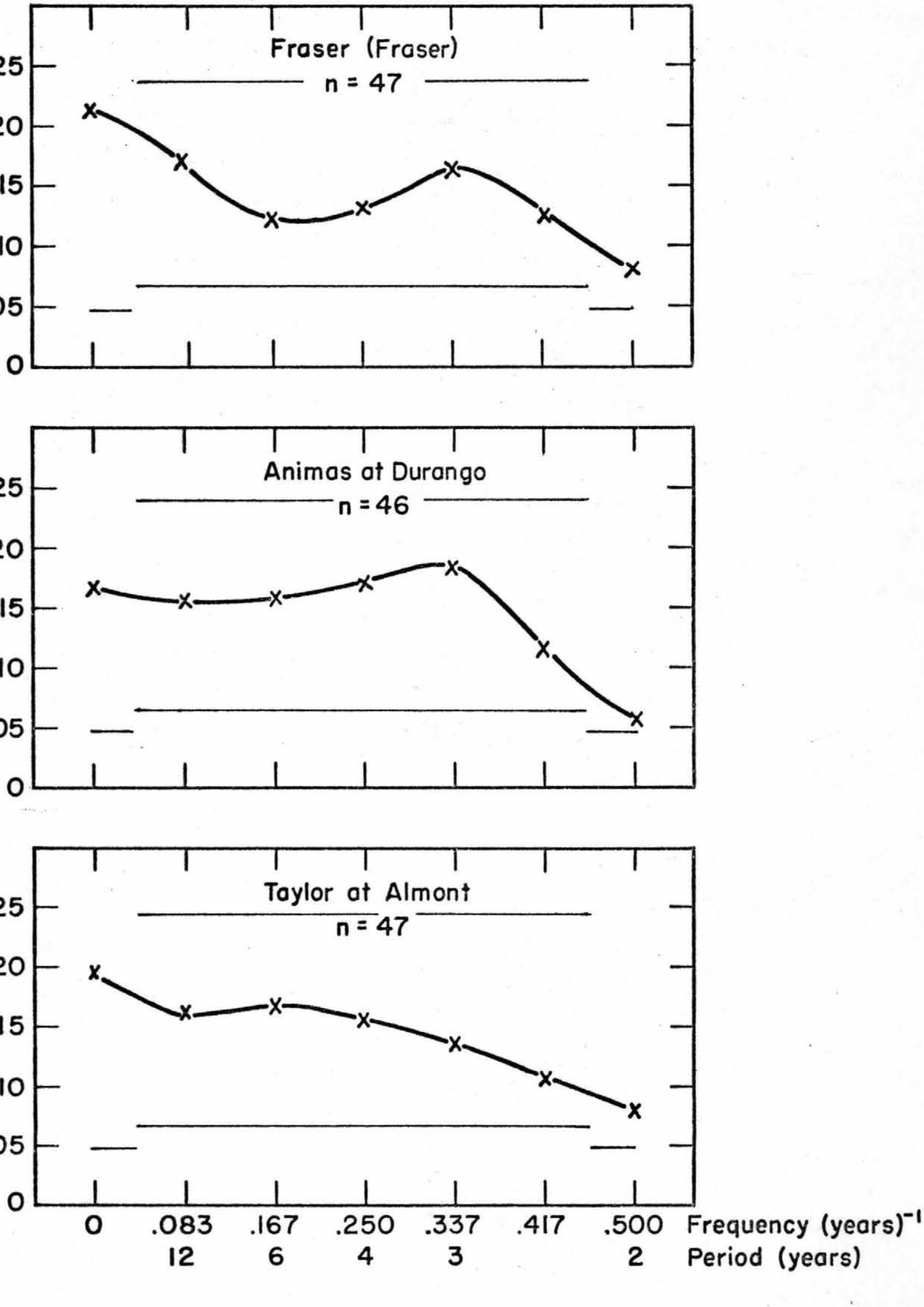


Figure V-3. Small Basins, Discrete Time.

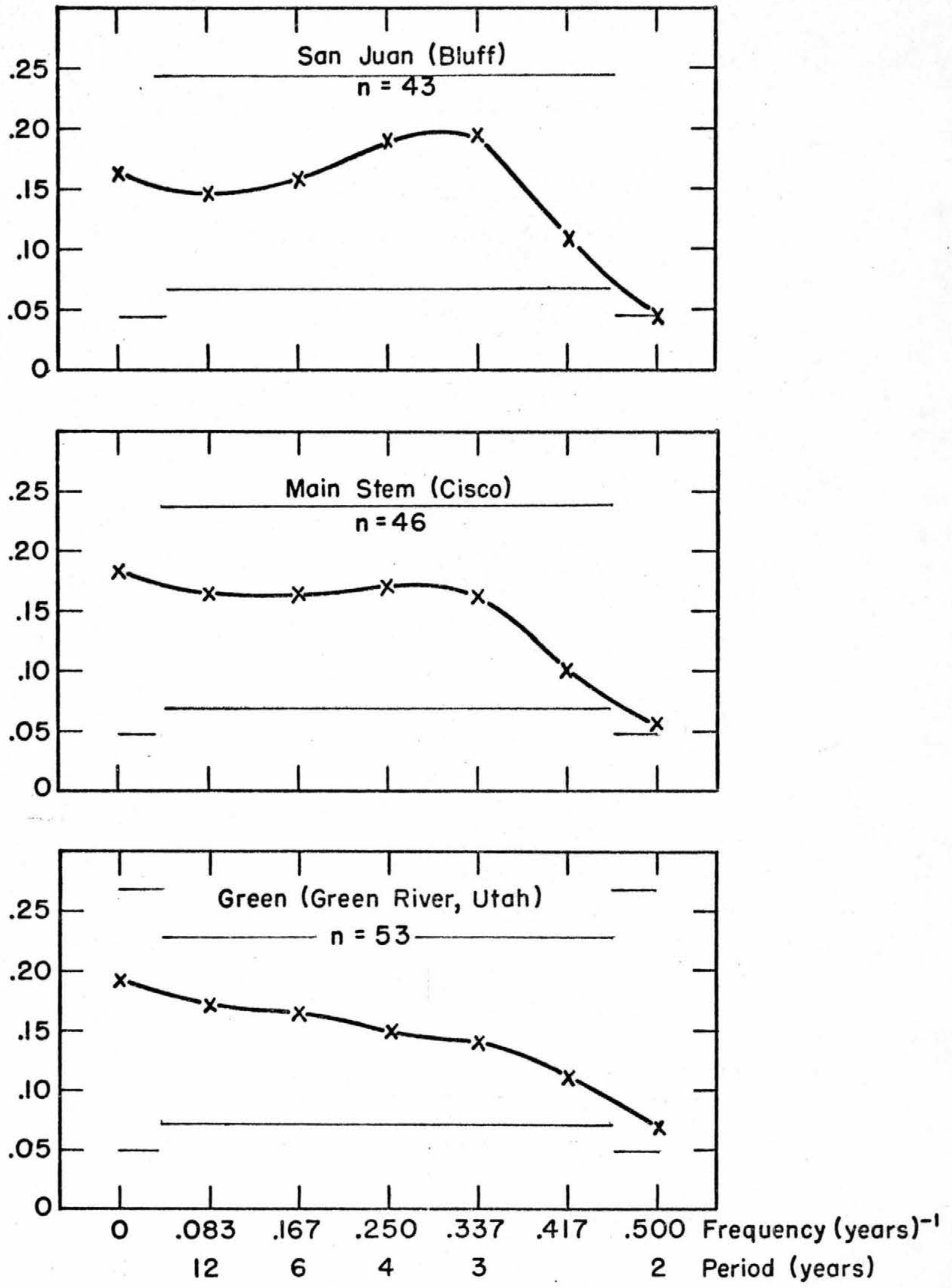


Figure V-4. Main Basins, Discrete Time.

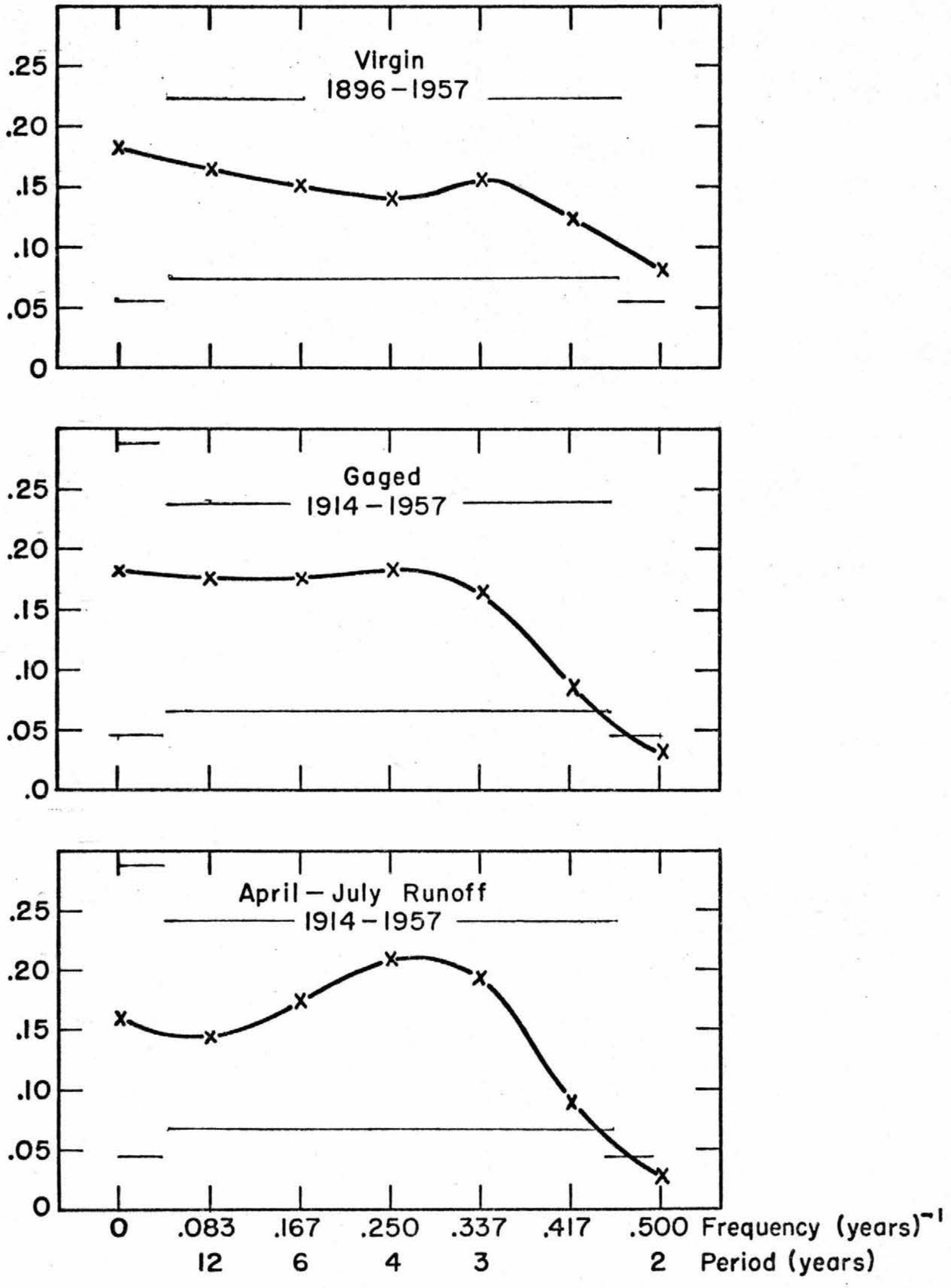


Figure V - 5 Lee Ferry Spectra, Discrete Time

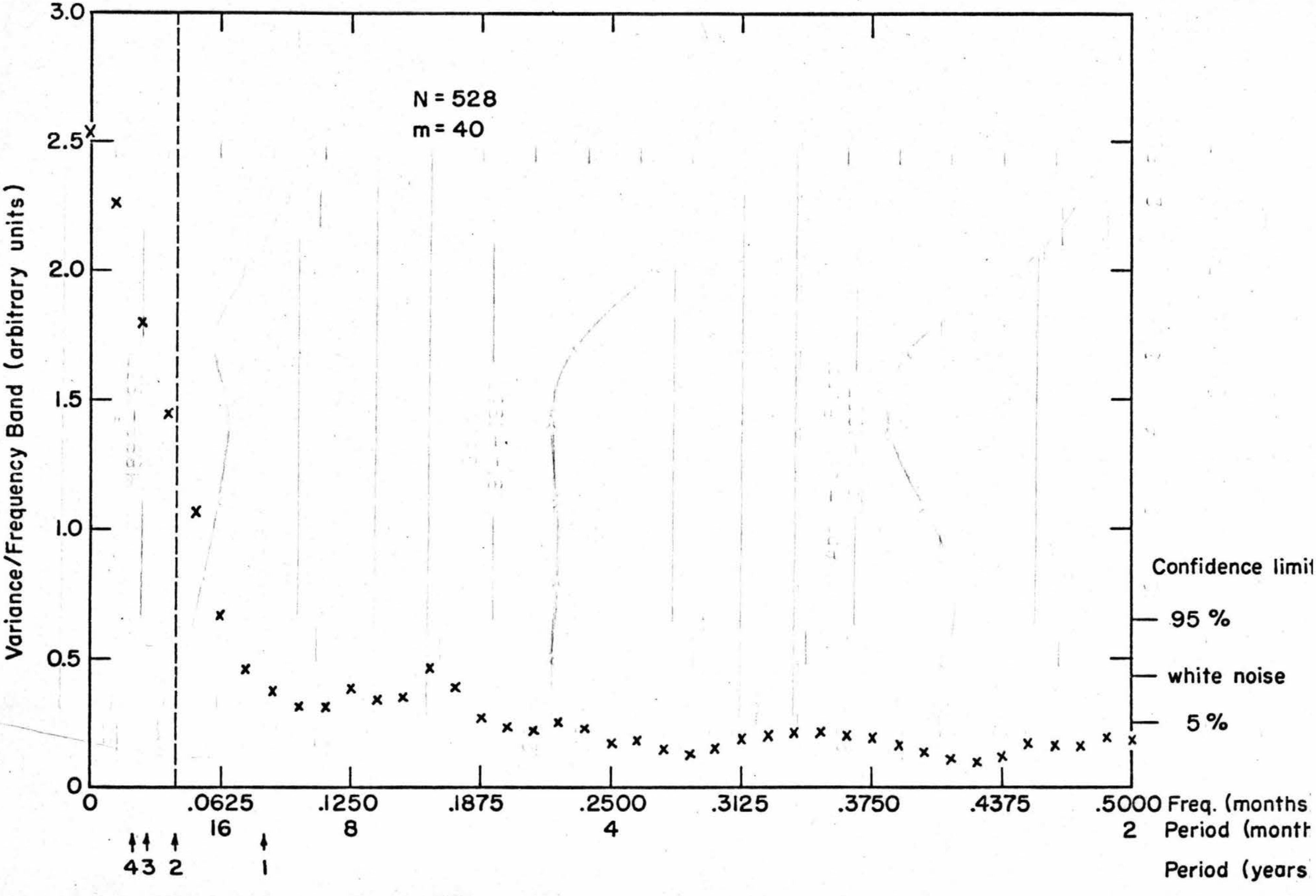


Figure V-6. Lee Ferry gauged runoff spectrum, 1914-1957.

(2) The shape of the Lee Ferry spectrum is apparently relatively insensitive to what particular data are used. Leopold's virgin flow figures 1896-1957 give a spectrum differing only slightly from the uncorrected gauged data. However, the difference is sufficient to reduce the significance level of the virgin flow spectrum to slightly below the 5% level.

(3) The spectra of the smaller basins show no significant differences from those of the three main sub-basins. All spectra indicate approximately the same amount of variance in the zero-frequency band (the very low frequencies) and further this variance is no more or less than could be expected in a series of random numbers.

(4) The Lee Ferry spectrum indicates a general decrease of variance from low to high frequency. Such a spectrum is consistent with a varying quantity with a slight amount of persistence or serial correlation from one value to the next. (this point is more fully discussed in Appendix I.)

Such an interpretation of the spectra of streamflow records is consistent with the analysis by V. Yevdjevich (CSU report) who utilized serial correlation analysis. The physical reasons why streamflow values exhibit persistence may be summarized as follows: First, carryover of runoff from one water-year to the next because of all forms of storage; second, inhomogenieties in the records (see Section IIB); third, a true persistence or serial correlation in either precipitation, evapotranspiration, or both. Although Yevdjevich's report considers these factors in some detail, it must be remarked here that the spectra of point precipitation totals suggest that no persistence is present in precipitation; the observed persistence in streamflow must then be accounted for by the remaining factors.

TABLE V-1

	Station Name	Elevation	Length of Record		Double-Mass Corrections
(1)	Trout Lake	9700	1914-1956	(43)	none
(2)	Silverton	9400	1906-1957	(52)	1936
(3)	Dillon	8900	1914-1957	(44)	1919
(4)	Fraser	8560	1916-1957	(42)	1936, 1951
(5)	Gunnison	7694	1904-1957	(54)	1950
(6)	Ignacio	6424	1914-1959	(46)	none
(7)	Shoshone	5918	1911-1959	(49)	1933
(8)	Montrose #2	5830	1902-1957	(56)	none
(9)	Delta	5115	1906-1957	(52)	1946
(10)	Grand Junction	4849	1897-1957	(61)	none

TABLE V-2

COLORADO BASIN

	Gunnison	Delta	Montrose	Grand Junction	Silverton
	1.00	1.00	1.00	1.00	1.00
1	+0.00	-0.10	-0.11	-0.18	+0.07
2	+0.18	-0.05	-0.19	+0.07	-0.06
3	-0.17	+0.05	+0.19	-0.11	+0.19
4	+0.24	-0.07	+0.06	-0.00	+0.10
5	+0.02	-0.09	+0.00	-0.19	+0.09
6	-0.15	-0.04	-0.05	-0.03	+0.18

	Trout Lake	Ignacio	Fraser	Shoeshone	Dillon
	1.00	1.00	1.00	1.00	1.00
1	0.04	-0.11	-0.01	-0.19	-0.07
2	-0.06	-0.20	-0.26	-0.03	-0.18
3	0.17	0.20	-0.15	0.05	-0.10
4	0.04	-0.01	-0.19	0.06	-0.12
5	-0.19	-0.12	0.00	-0.00	0.01
6	0.09	0.12	0.30	0.19	0.10

Lag	No. Positive	No. Negative
1	3	7
2	2	8
3	6	4
4	5	5
5	5	5
6	6	4

	Average
	1.00
1	-0.07
2	-0.08
3	+0.03
4	+0.01
5	-0.05
6	+0.07

TABLE V-3

STREAM-RUNOFF SPECTRA

	Fraser	Blue	White	R. F.	Taylor	Animas
	1.00	1.00	1.00	1.00	1.00	1.00
1	.23	.09	.16	.10	.24	.21
2	.26	.17	-.13	.14	.12	-.05
3	.34	.15	.06	.22	.16	.28
4	.20	.11	.11	.16	.18	-.00
5	.18	.10	.06	.31	.28	.16
6	.21	.18	-.06	.27	.37	.25

Lag	No. Positive	No. Negative
1	6	0
2	4	2
3	6	0
4	5	1
5	6	0
6	5	1

VI. CONFIDENCE IN ESTIMATES OF FUTURE VARIABILITY

A. Comparison of Leopold Approach with that of the Present Report

The material in this Section bears directly on Leopold's paper (22) in which he essentially assumes a model of the persistence existing in the Lee Ferry flow figures and proceeds to calculate the probable variations of future means of different lengths. Briefly Leopold's approach was to estimate the degree of persistence in the Lee Ferry data by showing that the variance S^2 of the means over 5, 10, 15, and 20 (N), years did not show a decrease by a factor 1/N as would be expected from a series of random numbers but instead decreased less rapidly. By drawing a smooth curve through the

$$\sqrt{\frac{S_N^2}{S^2}}$$

values, Leopold was in actual fact assuming a persistence model for the Colorado at Lee Ferry.

The following section will examine a number of points associated with Leopold's treatment. (1) Leopold's estimates of the ratio of

$$\frac{S_N^2}{S^2}$$

contained sampling errors, and the question of the magnitude of these errors might be legitimately asked. In point of fact, how good was his estimate of the degree of the persistence using this method?

(2) it can be shown that an analysis of the variability of future means can conveniently be calculated once knowledge of the spectrum is obtained. If S_N^2 represents the variance of means of N years and

$$\Phi(\omega)$$

is the spectrum of the runoff process, then

$$S_N^2 = \int_{-\pi}^{\pi} \left[\frac{\sin \frac{N\omega}{2}}{N \sin \frac{\omega}{2}}\right]^2 \Phi(\omega)\, d\omega \qquad \text{VI-1}$$

If

$$\Phi(\omega)\, d\omega$$

is a constant as it is for a random process, then

$$S_N^2 = \frac{S^2}{N}$$

which is the property of a random series that Leopold tested against. With the estimates that we have of the spectrum of the Lee Ferry flow, we are in a position to (1) calculate the S_N^2's using the estimated spectrum and (2) calculate the spectrum Leopold assumed by comparing his S_N^2's with those estimated in VI-1.

It is convenient at this point to ask what the relationships above would reduce to in the case of a particular autoregressive scheme known as a Markov process. These processes have been extensively studied and may be written as

$$X_{t+1} = \rho X_t + \bar{\epsilon}, \quad \rho = \text{constant} < 1.$$

Interpreted, this auto-regressive scheme of order one simply says that the succeeding value of the variable X depends only upon its present value plus a random component. It is an open question as to whether such a process can represent observed streamflow series, although it has been suggested by various investigators, including the author (19), (10), (14). The limitation imposed by the length of the records we possess plus the uncertainties in the degree of the homogeneity of our records require that it remain an open question.

If, to return to the original argument, we assume that streamflow may be represented by a Markov process the above equations reduce to

$$\frac{S_N^2}{S_1^2} = \frac{1}{N}\frac{(1+\rho)}{(1-\rho)} - \frac{2\rho(1-\rho^N)}{N^2(1-\rho)^2} \qquad \underline{\text{VI-2}}$$

This follows because the spectrum of a Markov process may be written as

$$\left[\Phi(\omega)\,d\omega\right]_\rho = \frac{1-\rho^2}{\pi(1+\rho^2-2\rho\cos\omega)}\left[0<\omega<\pi\right]$$

Assuming a value of the 'carry-over' coefficient we can then calculate

$$\frac{S_N^2}{S_1^2}.$$

Assuming the Lee Ferry spectrum to be a Markov spectrum taking into account sampling variations a value of $\rho = 0.25$ can be used. That is, we will assume that

$$X_t = 0.25X_{t-1} + \epsilon$$

can represent the series of flows at Lee Ferry. The spectrum of such a model is shown in Fig. VI-1 compared with the estimated spectrum.

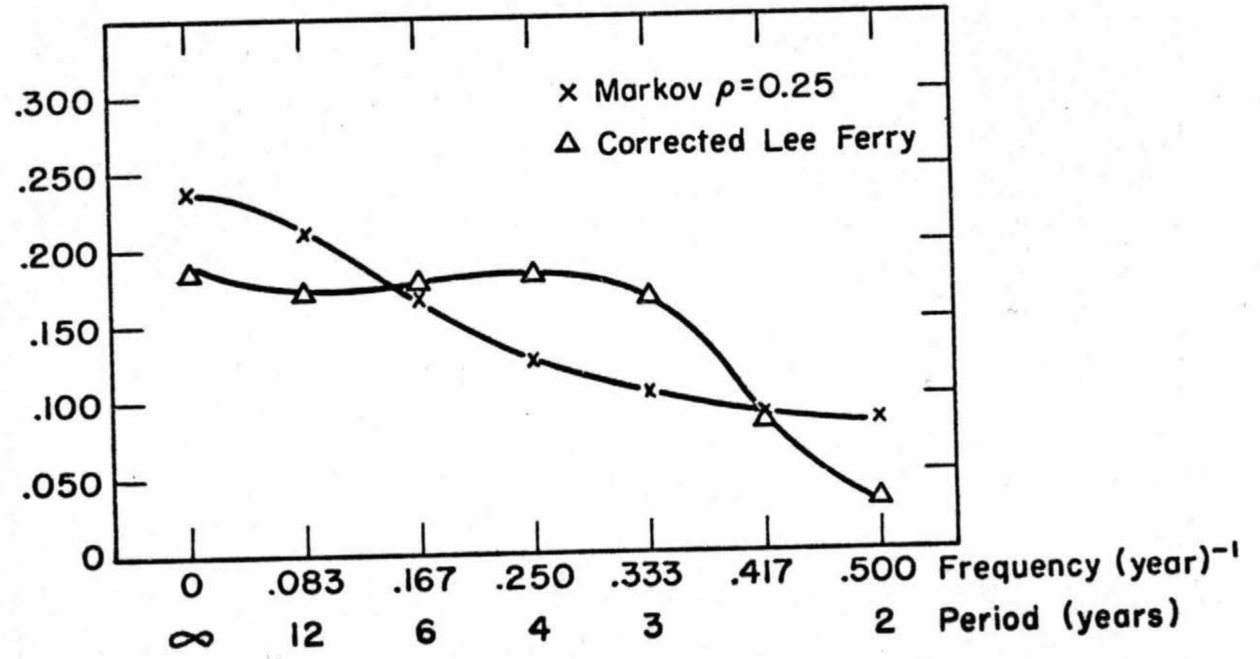


Figure VI-1. Lee Ferry Runoff Spectrum vs. Markov process Spectrum ρ = 0.25.

Using a simple Chi-squared test of fit indicates that the differences are not significant on the 5% level; the length of record is too short to be able to distinguish the actual estimated spectrum from a Markov (ρ = 0.25) process. This same comment also holds for more sophisticated auto-regressive or moving average models which might be suggested, i.e. higher order models.

Table VI-1 contrasts the variance of five and ten year means estimated by three different methods:

(1) Leopold, using observed

$$\frac{S_N^2}{S_1^2}$$

(2) Markov process, ρ = 0.25 from equation VI-2.

(3) A purely random series.

Examination of the comparison indicates that the estimated variance of the averages is, of course, in excess of that calculated for a random series and that they are in fair agreement with Leopold's values.

The question of why Leopold's values differ from the author's can be answered by examining just what parameters were used to estimate

$$\frac{S_N^2}{S_1^2}.$$

This report considers a simple Markov process with ρ = 0.25, an adequate model to use from comparison of the two spectra (fig. VI-1). Leopold plotted points for

$$\frac{\hat{S}_N^2}{\hat{S}_1^2}$$

where $\hat{S}_N^2$ was calculated from contiguous values of X_N, that is, X's averaged over non-overlapping lengths of N years. The S_{20}^2 calculated by him was thus based on only three values (61 years total length of record). Leopold gives no consideration to sampling fluctuations of his $\hat{S}_N^2$ and is open to criticism on that account. On the other hand the analysis used in this report is incapable of distinguishing between a Markov process with ρ = 0.25 and the estimated spectrum, and the resulting uncertainty in the variance of future means is therefore quite large. (Refer to Table VI-1.) The analysis <u>is</u> capable, however, of saying that the ratio

$$\frac{S_N^2}{S_1^2}$$

is significantly greater than 1/N on the 5% level.

TABLE VI-1

$N = 10$ Variance of Ten-year-means	$\frac{s_{10}^2}{s_1^2}$
(1) Leopold	.193
(2) Markov, $\rho = 0.25$	.159
(3) Random	.100
$N = 5$ Variance of Five-year-means	$\frac{s_5^2}{s_1^2}$
(1) Leopold	.345
(2) Markov, $\rho = 0.25$	.298
(3) Random	.200
$N = 20$ Variance of Twenty-year-means	$\frac{s_{20}^2}{s_1^2}$
(1) Leopold	.122
(2) Markov, $\rho = 0.25$	.081
(3) Random	.050

Since Leopold's estimates of

$$\frac{s_N^2}{s^2}$$

were rather conveniently calculated, it is of interest to determine just what spectral estimates correspond to his values of

$$\frac{s_N^2}{s^2}.$$

This would give an idea of the spectrum he assumed which could then be compared with the Tukey estimate of the spectrum. Appendix III gives the calculation of spectrum assumed by using the

$$\frac{\hat{s}_N^2}{\hat{s}^2}$$

calculated by Leopold. The solution to this problem is roughly the inverse of the analysis just presented. We use the

$$\frac{s_N^2}{s^2}$$

to estimate $\Phi(\omega)$.

Fig. VI-2 gives the results of the calculation. The spectrum implicitly assumed by Leopold in his analysis contains an excess of very low frequency variance compared with high frequency variance whereas the Tukey

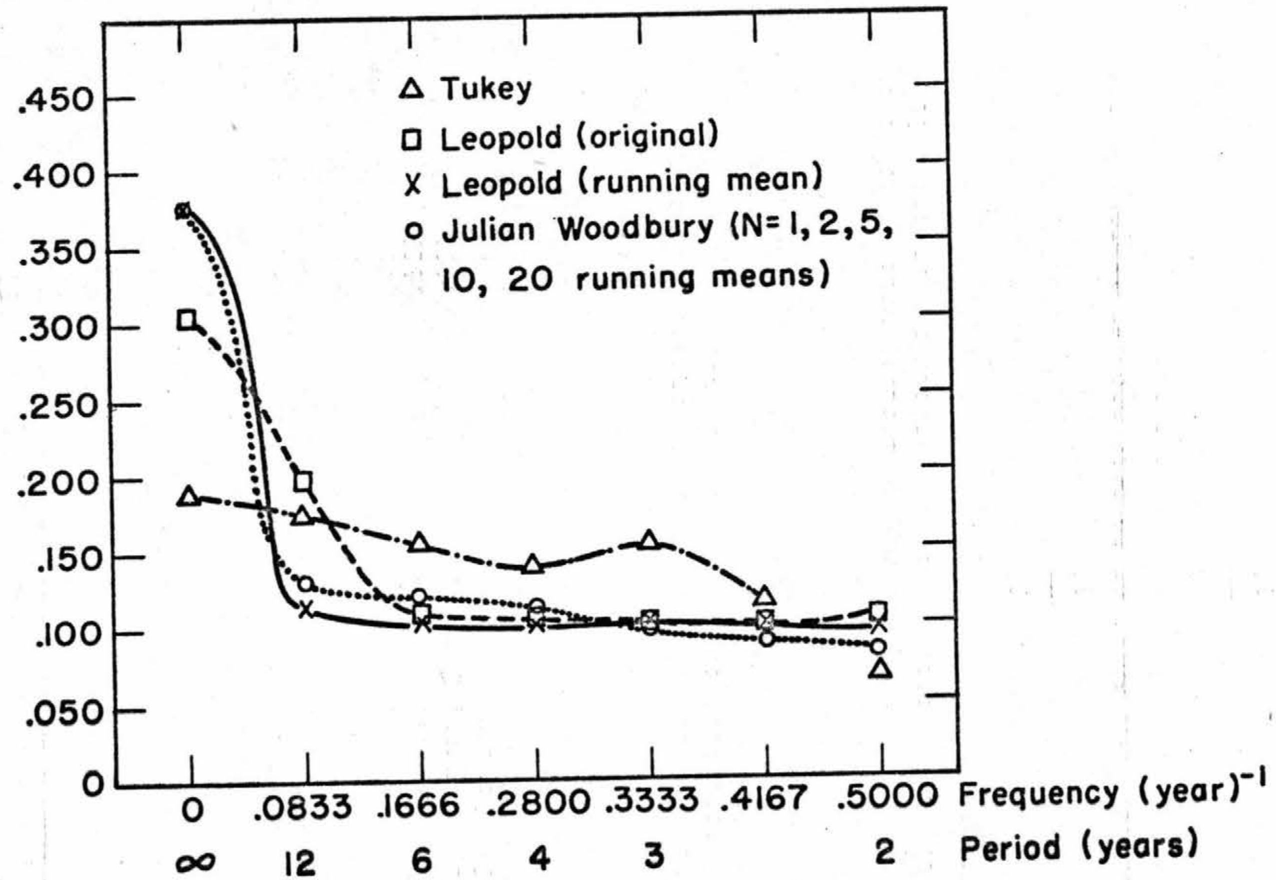


Figure VI-2. Spectra Estimated from Variances of Lee Ferry Flows Treated with Averages of N years, and Tukey Estimated Spectrum.

spectrum is much more 'flat' over the medium to low frequencies.

From statistical theory it can be argued that the procedure by way of the Tukey spectrum is to be preferred since this class of methods of estimation of the spectrum has received much theoretical attention in the last few years. Leopold's method should not be criticized too heavily, however, because it is easy for the statistical layman to understand, and with the length of record available, gives results comparable to the more sophisticated methods.

B. Probability Analysis Using Hypothesized Markov Model

A probability analysis similar to Leopold's using the results of Table VI-1 can be made. For example, the variance of five or ten-year means of a Markov process with $\rho = 0.25$ gives a distribution as shown in fig. VI-3. To this must be added the error associated with assuming the 1914-1959 mean to be the true mean. In such a Markov process, it can be shown that there are $46 \times (1 - 0.25)/(1 + 0.25) = 29$ independent observations of the true mean. The standard error of the mean is thus = $4.07/\sqrt{29}$ or 0.75 million acre feet.

To obtain the overall variation in future 10-year periods, the standard error of the mean must be combined with the standard error of the 10-year means. These can be considered independent quantities and add in their squares.

Variability of future 10-year means is equal to

$$\sqrt{(0.75)^2 + (1.62)^2} = 1.78 \text{ m.a.f.}$$

Thus the probability is 68% that a future 10-year mean will be between 13.19 ± 1.78 m.a.f. at Lee Ferry and thus there is a probability of roughly 16% that such a mean will be below 11.4 m.a.f.

Fig. IV-4 shows graphically that the effect of the non-randomness is the positioning of the confidence limits. The dashed line indicates the probability that a 10-year mean would fall within the limits delineated while the solid indicates the same limits taking into account the non-randomness of the data.

This analysis differs from that of Leopold's because, (1) the sample mean and variance were different (he used the so-called virgin flow, 1896-1956) and (2) a different model of the non-random component was used; the

$$\frac{s_N^2}{s^2}$$

ratios for a Markov process from table VI-1. It should be emphasized that those limits are for stream-runoff at Lee Ferry, not stream discharge. No depletion model has been used or other such considerations made.

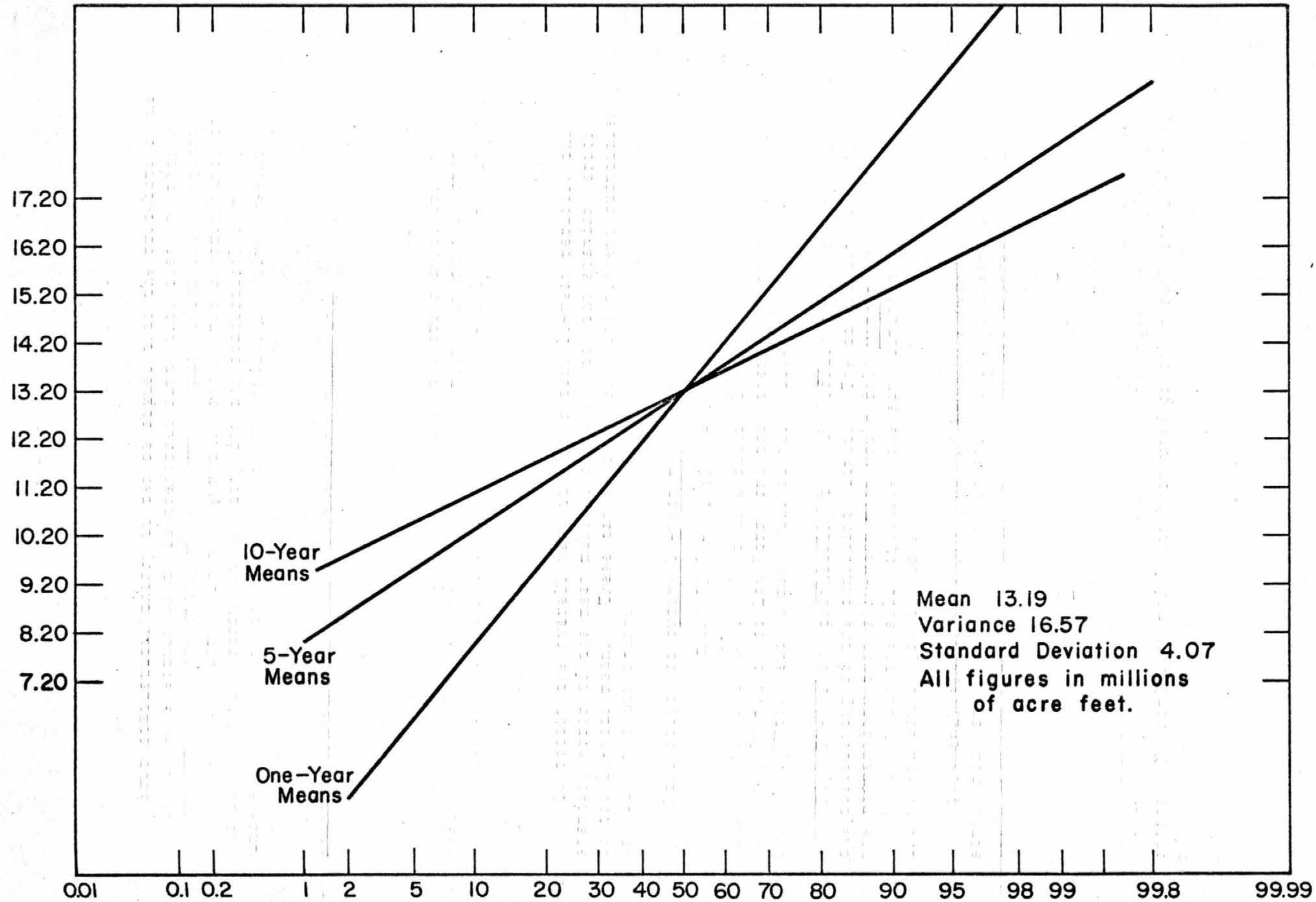


Figure VI-3 Lee Ferry, gaged and corrected, 1914—1959

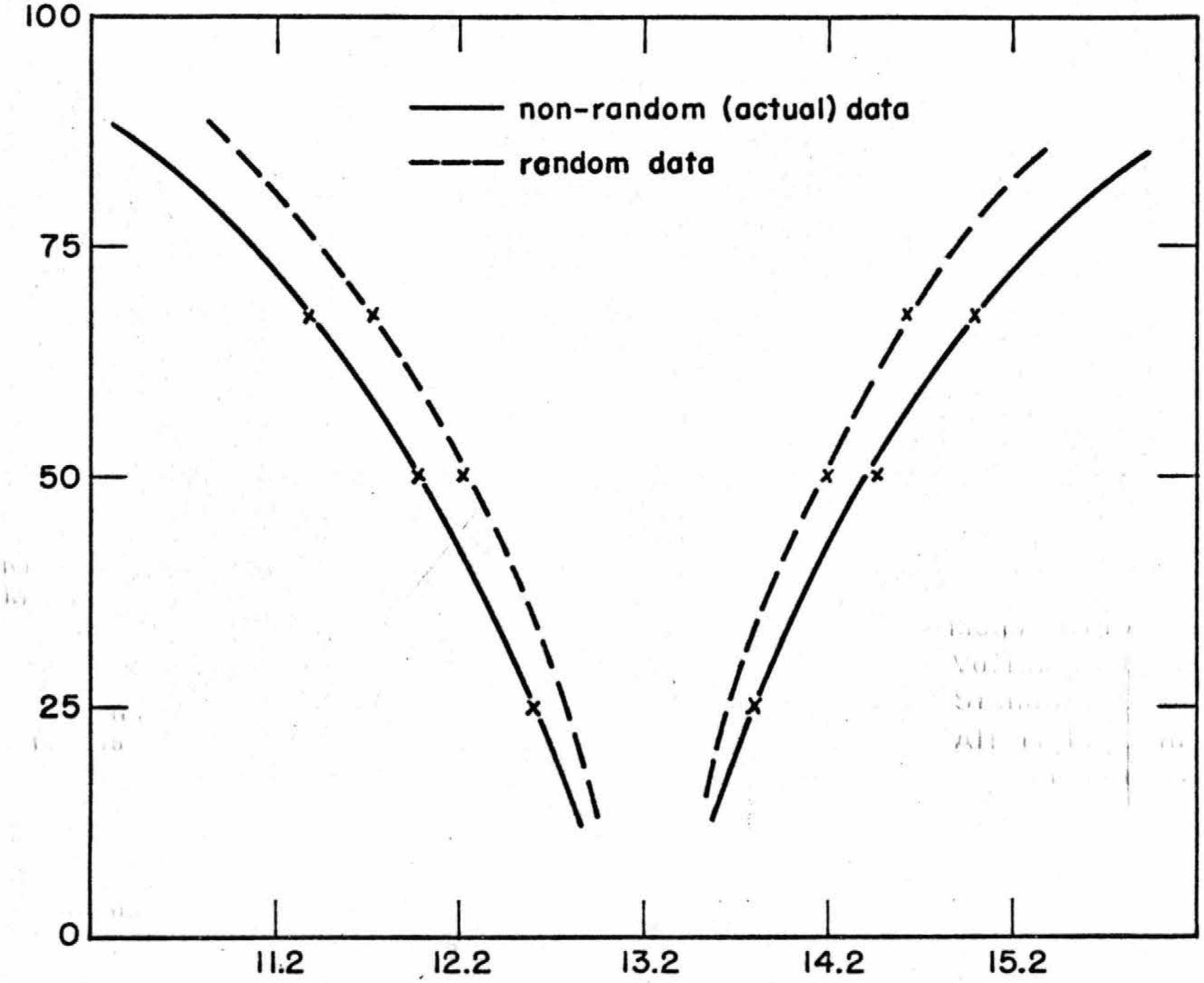


Figure VI-4. Probability of 10-year Lee Ferry flow falling within the limits shown.

VII. THE APPLICATION OF STATISTICAL FILTERING TECHNIQUES TO THE HYDROLOGIC TIME-SERIES

A. Introduction

The variance spectrum technique depicts the structure of a time-series by distributing the total variance of the quantity over a frequency scale in a number of finitely-wide frequency bands. Such a representation is capable of revealing strict periodicities in the series if the band-width of the analysis is sufficiently narrow. Records of precipitation and streamflow in the basin are too short to reliably (from a statistical point of view) determine the presence, or absence, of any periodicities in these records, as pointed out in a previous section.

There is, however, a method by which such periodicities can be investigated in short lengths of record provided the hypothesis is made that the periodicity retains a constant phase throughout the length of record. This method, using weighted running averages operating on the time-series, has been described as a band-pass filtering technique. The weights of the running averages are determined so as to filter out or retain a particular portion of the spectrum.

Let $x_1, x_2, \ldots x_n$ represent the time-series of interest and $w_o, w_1, w_2, \ldots 2_k, (k \ll n)$ the weights of a meaning operation on the x-series defined as

$$y_n = \sum_{j=0}^{k} \omega_j (x_{n-j} + x_{n+j}) \qquad \text{VII-1}$$

This running average is thus symmetric with ω_o being the central weight and the total length of the running mean 2k + 1 values of the series. The new series, y_n, that is generated by the meaning process is thus shorter than the x-series by 2k items (k items on each end) and has a different appearance since certain of the frequencies in the x-series have been suppressed and certain others retained.

If we consider a periodic function

$$x = A \cos (2\pi ft + \emptyset), = A \cos(\omega t + \emptyset)$$

which in case of discrete time can be written

$$x_n = A \cos(\omega n \alpha + \emptyset),$$

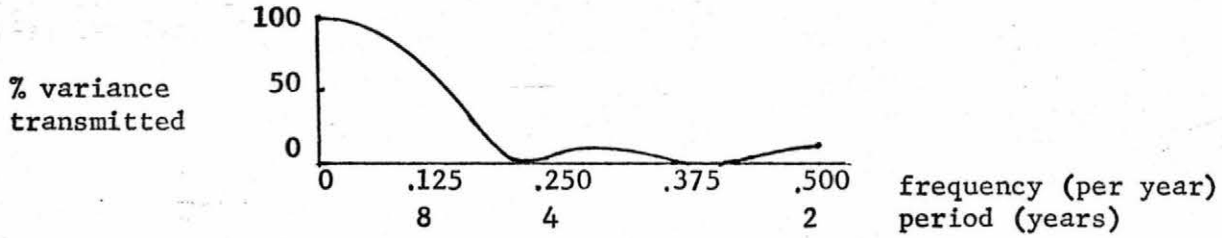
α = interval between each of the n observations. Operating on this, with the symmetric weighted running mean filter, it can be shown that

$$y_n = A \cos (\omega n \alpha + \phi) \sum_{j=0}^{k} \omega_j \cos \omega \alpha j$$

The effect of filtering is to thus preserve the period and phase of the original periodic function and to change its amplitude by a factor

$$Y(\omega) = \sum_{j=0}^{k} \omega_j \cos \omega \alpha j \qquad \text{VII-2}$$

Thus, by taking α equal to unity and specifying the particular weights used, the amplitude response function of the running average may be calculated. Taking the square of the amplitude response function results in a quantity termed the admittance function or filter factor. This function gives the portion of the spectrum "passed" by the filter and the portion suppressed. For example, consider a simple unweighted running average; all the weights, ω_j, in this case, are equal and equal to $1/(2k + 1)$. Suppose the x_n series consists of yearly values and a symmetrical moving average of five terms $(2k + 1 = 5)$ is selected. The filter function of this averaging process would look like this:



The filter transmits no variance associated with 5-year periodicities and an increasing amount at longer periods (lower frequencies). This filter also transmits a small amount of variance associated with periods less than five years as shown. This rather undesirable feature can be suppressed by using weights with higher values at the central part of the filter $(j = 0)$ and decreasing values toward the "ends" of the filter $(j = k)$.

A complete description of these techniques has been given by the Labroustes (16), Carruthers (8), Craddock (9), Brier (7), and others. Appendix I also covers a few of the subjects under this heading.

Because the running mean filter will modify the spectrum of the original series, oscillations will result in the y_n series which may not have been present in the original record. Consider, for example, a random number spectrum. If the random numbers are treated with the five-year running average used as an example just previously, the spectrum of the resulting series will be given by the shape of the filter function curve. It can be shown, moreover, that the spectral value of the filtered series at any given frequency, ω, will be the product of the filter function and the spectral value of the original series both at the given frequency.

$$\Phi_y(\omega) = \left| Y(\omega) \right|^2 \Phi_x(\omega)$$

A limited length of random record filtered with a five-year running average will thus emphasize oscillation with periods greater than five years, whereas in the original series these oscillations were no more marked than those with any other periods.

A possible method used to insure that the oscillations remaining after filtering are actually in the record and are not the result of the filtering itself, is to specify that the resulting oscillations retain a constant phase throughout the record.

The band-pass filtering technique will be used in this report for two different purposes. Because the filtering preserves the phase information of any periodicity which might be in the record, it will be used to examine the series for a 'hidden periodicity' of constant phase as suggested. In particular the technique will be used to examine the precipitation and streamflow records in the Upper Colorado Basin for a direct constant phase relationship with the sun-spot number. Such relationships, although lacking any physical basis, have been suggested by many workers, for example, (3), (37), (42). The particular running average used for this investigation and its filter function are shown in Fig. VII-1. This particular filter was taken directly from Craddock (9).

Another use of the band-pass filters will be to investigate the very long periods (low frequencies) in the precipitation and streamflow records by means of a low-frequency band-pass filter. The particular filter used for this is also taken from Craddock and is shown in Fig VII-1. Note that this filter has zero transmission at a frequency 1/12 year^{-1}, or where the previously mentioned filter had a maximum transmission. The low-frequency filter is thus preserving mainly the oscillations with periods on the order of or larger than the length of the record. A study of the so-called 'trends' in the records can thus be conveniently made.

B. Results

The two filters were applied to the majority of the ten precipitation stations chosen for analysis in this report and all of the streamflow stations. The following discussion surveys the results of the filter emphasizing the 11-year period band as shown in the solid curve of Fig. VII-1. To assist in drawing firm quantitative conclusions the stipulation was made that the phase of the 11-year oscillations remaining after filtering as measured from the years of maximum and minimum values, must remain constant within three years with respect to the

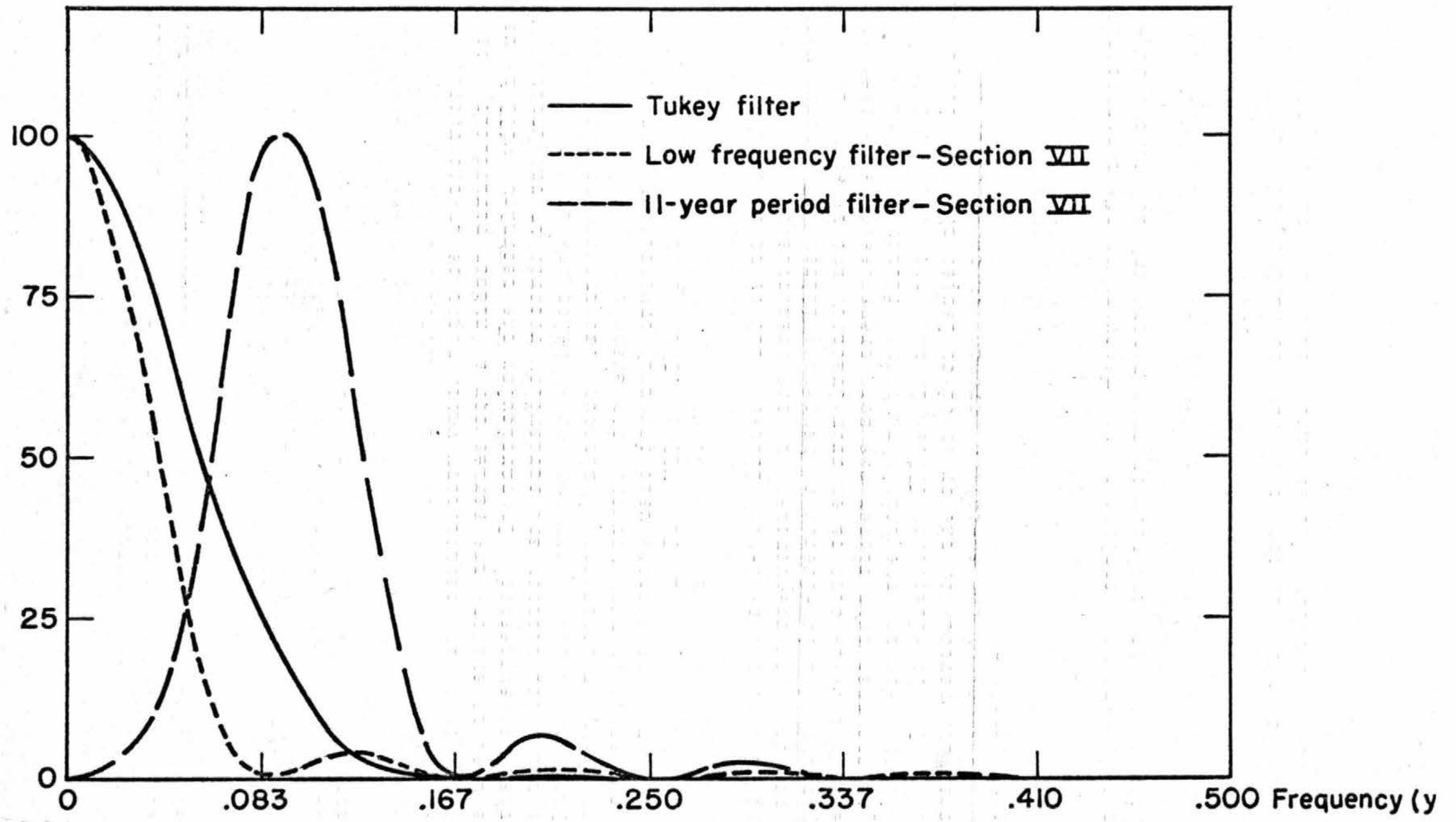

Tukey filter
Low frequency filter - Section VII
11-year period filter - Section VII
100
75
50
25
0
0
.083
.167
.250
.337
.410
.500 Frequency (y

Figure VII-1.

11-year sunspot cycle throughout the entire record. Fig. VII-2 gives the results at the Lee Ferry streamflow and Silverton winter precipitation. The vertical lines extending upward and downward from the abscissa represent the years of maximum and minimum sunspot number respectively. Inspection of these curves suggests that a constant phase relationship exists for Lee Ferry. For example, the maxima in streamflow occurred with a lag of 1, 3, and 3 years following sunspot maxima and lags of 2, 0, 3 following sunspot minima. Such a relationship is within the limitation imposed above. Silverton's precipitation, however, does not so qualify as a glance at the curve will indicate.

In Fig. VII-3 we show the winter precipitation at Grand Junction and the annual precipitation at Yellowstone Park. Grand Junction does not qualify because of a changing phase relationship in the beginning of the record. Yellowstone's record of annual precipitation also is not indicative of any relationship. (This record was included because it represents a long record of good precipitation measurements; it is, of course, not within the Upper Colorado Basin.) Fig. VII-4, Shoshone and Gunnison precipitation, and Fig. VII-5, Animas and Roaring Fork streamflow, also indicate that no consistent phase relationship is present.

As the Lee Ferry record is the only record meeting the imposed restrictions favorable to a sunspot-runoff relationship, the three major sub-basins are shown in Fig. VII-6 to check on the individual behavior of the major sub-basins. In this figure, the Green, the data for which extend back to the early part of the century, fails to meet the criterion, as the minimum in 1912 comes one year before the 1913 sunspot minimum. In addition, the Main Stem data show a secondary maximum in 1945 which is as strong as the 1940 maximum. Although the three basins are reasonably well correlated in this portion of the spectrum, particularly between 1920 and 1940, the individual behavior shown here and the failure of the smaller basins and precipitation records to indicate a phase relationship which meets the imposed criterion, suggest that the Lee Ferry record's success was likely fortuitous.

If, however, the Lee Ferry record would in fact ultimately show a consistent phase relationship with the sunspot cycle in the 11-year frequency band, the knowledge of this fact will not benefit us any in actual forecasting practice. First of all, no relationship between the amplitudes is present so that all that would be forecastable would be the date of maxima and minima. In addition, the fraction of the total variance represented by this frequency band is so small that even if such a date could be forecast only a very small degree of predictability would be gained.

Another filter was used to investigate the very low frequency oscillations in the series under consideration. This filter has a filter function curve as shown in Fig. VII-1 by the dashed curve.

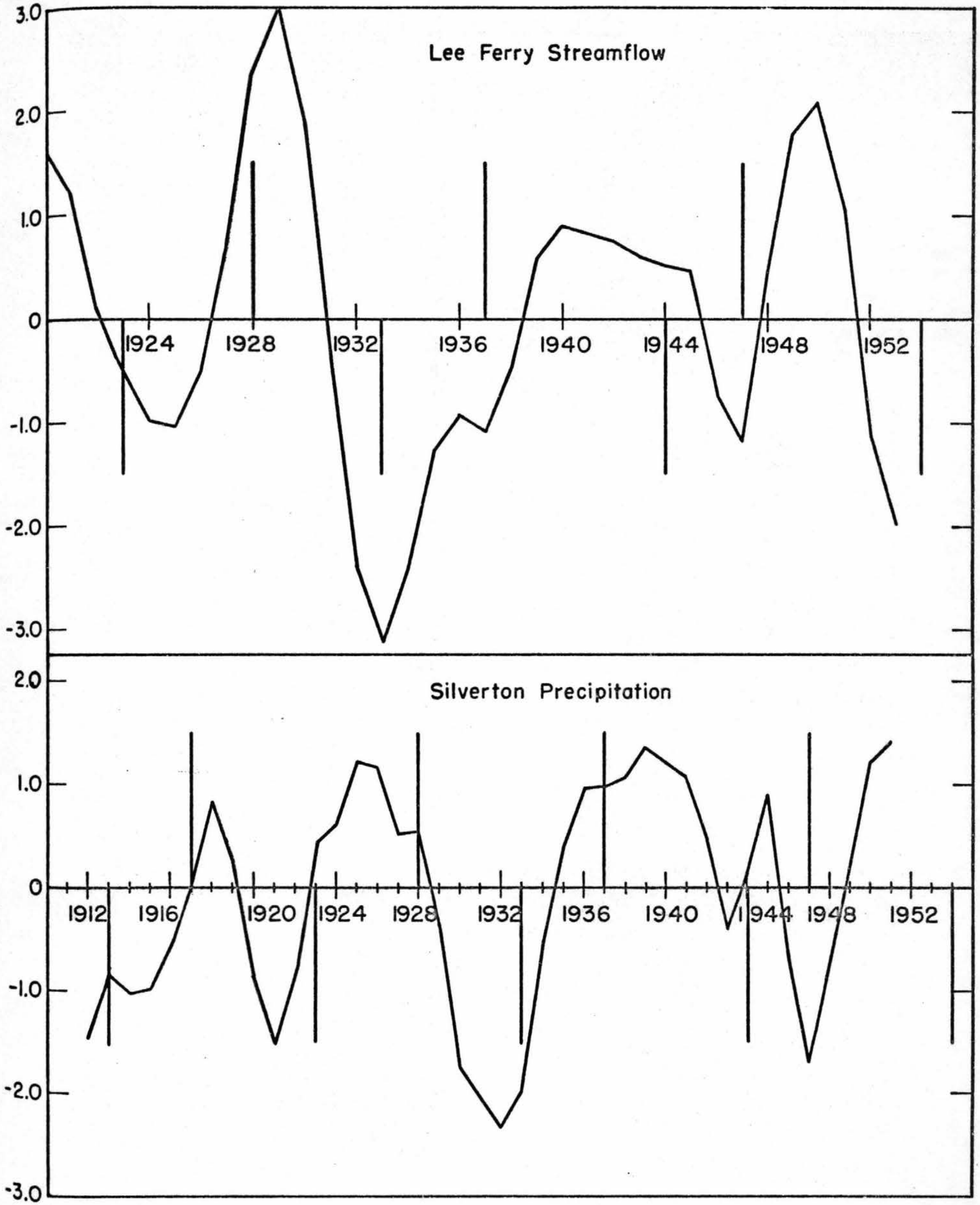


Figure VII-2. Streamflow-precipitation data treated with 11-year band-pass filter. Vertical lines denote years of sunspot maxima and minima.

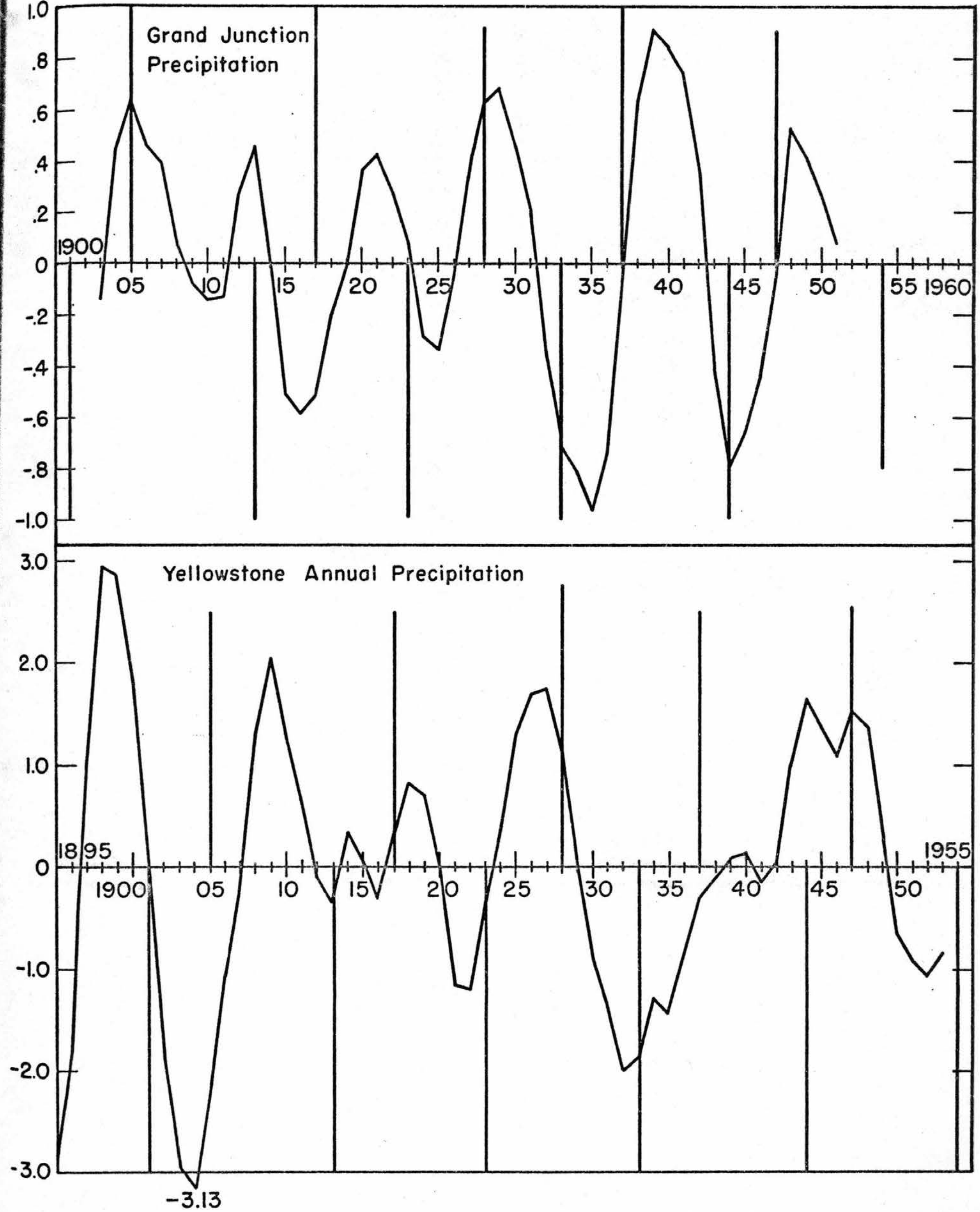


Figure VII-3. Precipitation data treated with 11-year band-pass filter. Vertical lines denote years of sunspot maxima and minima.

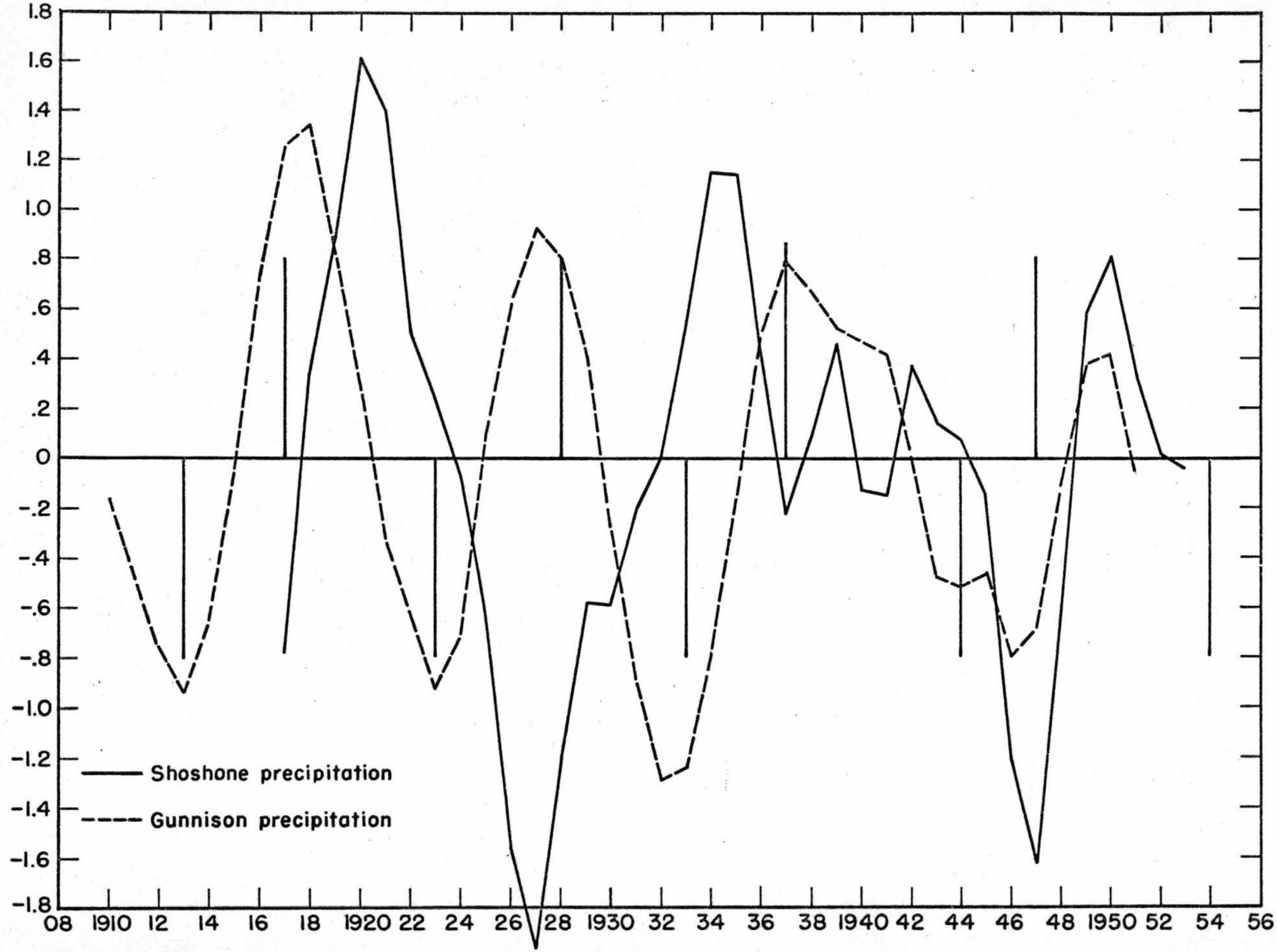


Figure VII-4. Precipitation data treated with 11-year band-pass filter. Vertical lines denote years of sun-

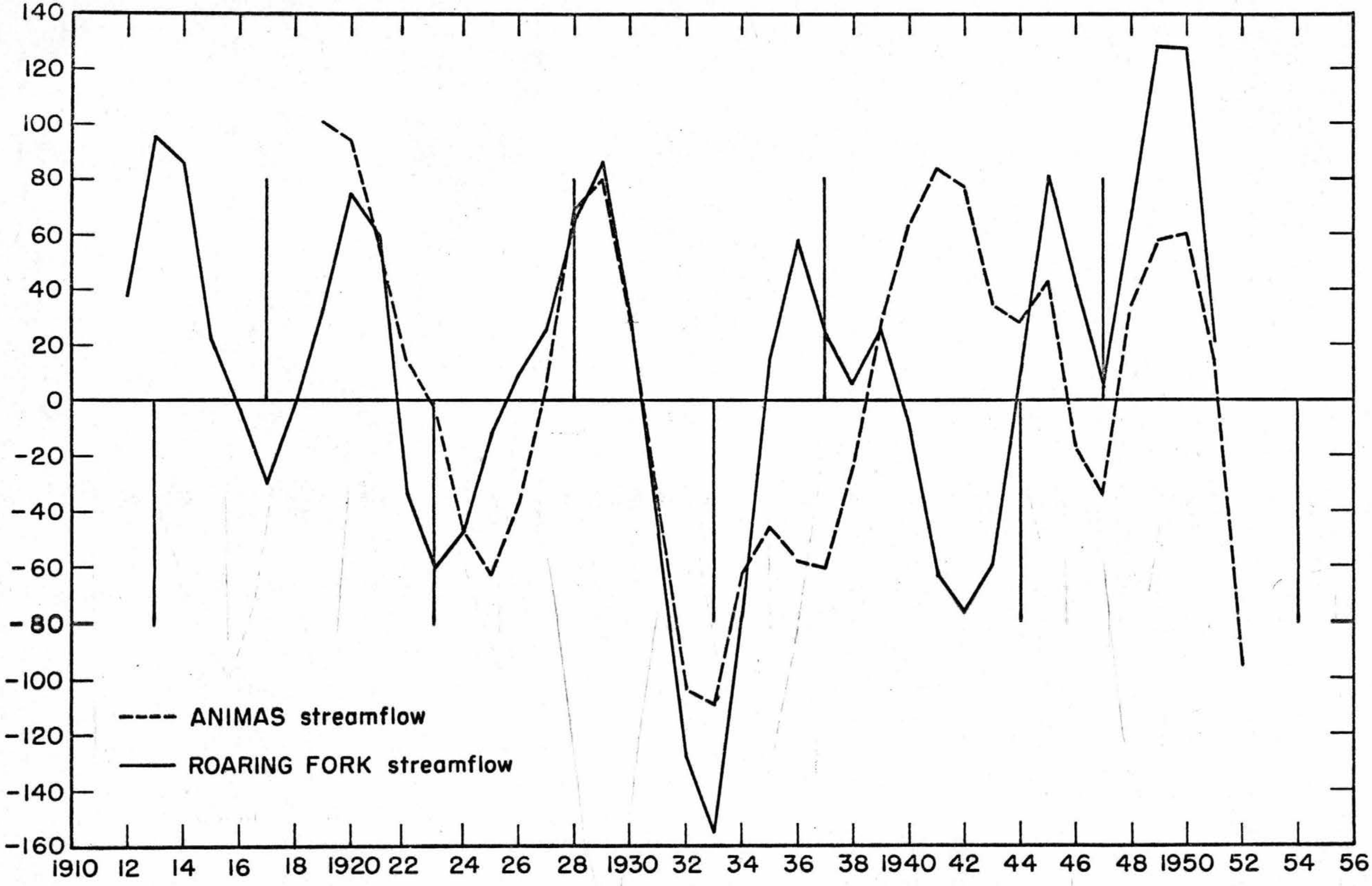


Figure VII-5. Streamflow data treated with 11-year band-pass filter. Vertical lines denote years of sunspot maxima and minima.

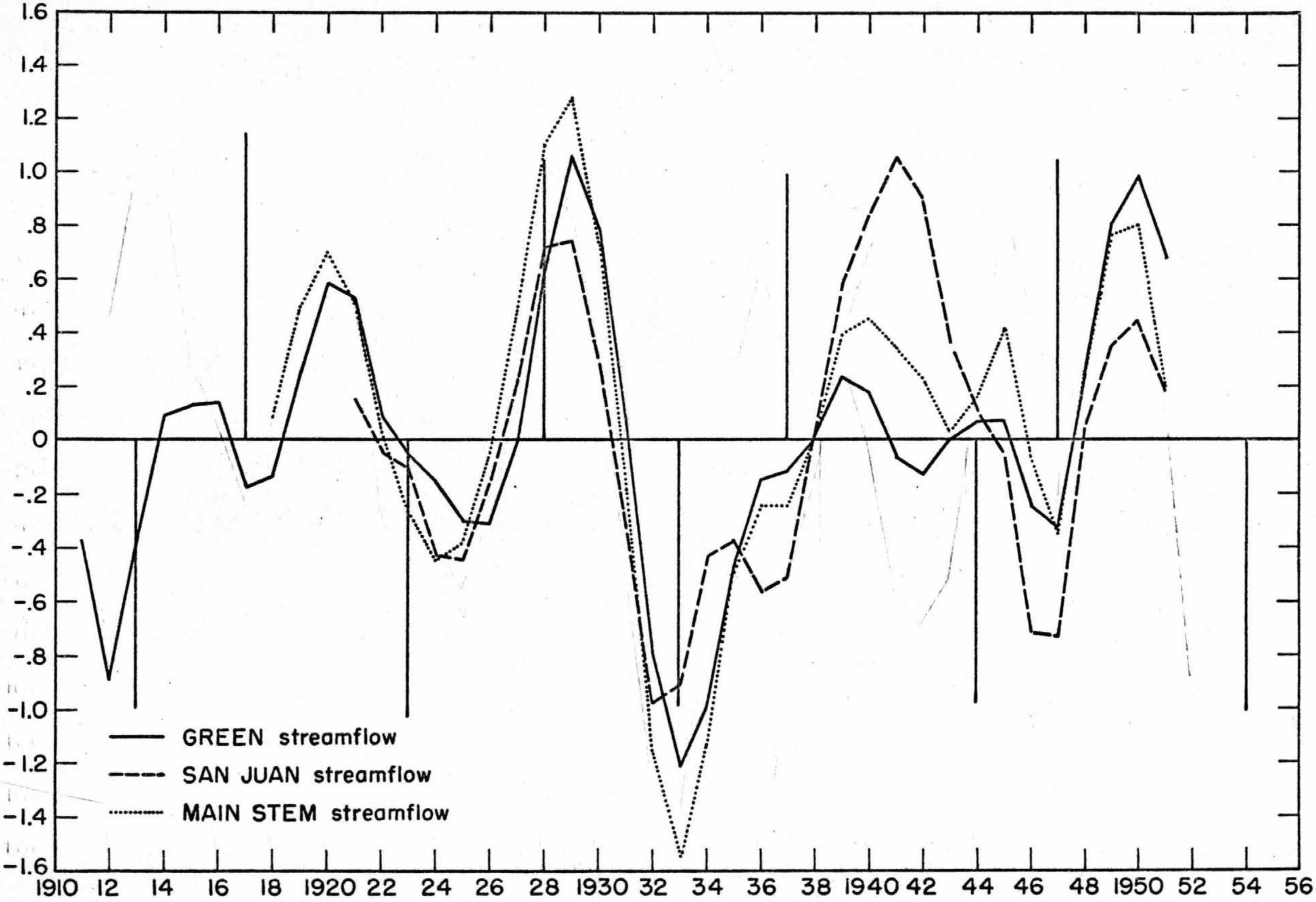


Figure VII-6. Streamflow data treated with 11-year band-pass filter. Vertical lines denote years of sunspot maxima and minima.

Note that the filter effectively blocks everything with a period less than 11 years and that the 50 per cent transmission level is about 25 years. This filter thus places emphasis on the movement of the series which have periods equal to, or greater than, the length of the record, and it should be pointed out that this is a much narrower filter than that involved in the Tukey method of estimating the spectrum used in Section V. The filter functions inherent in the Tukey method is shown as the dotted line in Fig. VII-1.

Because the filter function for this particular filter is known, the effect on the variance of the filtered series by equation VII-2 and an estimate of the amount of variance within this frequency band in the original series, or in other words, an estimate of the spectrum, can be obtained. In the following analysis this spectral estimate will be normalized so that a value of unity would correspond to the amount of variance in the frequency band expected from a random series.

Table VII-1 presents the normalized low-frequency band variance for the ten precipitation stations and for the runoff series used in Section V.

TABLE VII-1

NORMALIZED ZERO-FREQUENCY-BAND FILTER VARIANCE

Precipitation (Nov.-Apr.):

Trout Lake	1.22	Delta	0.51
Silverton	0.90	Dillon	0.50
Gunnison	0.75	Ignacio	0.35
Montrose #2	0.68	Grand Junction	0.32
Shoshone	0.68	Fraser	0.19

Stream- runoff:

Taylor	3.16	Green (Green River, U.)	2.96
Roaring Fork	2.89	Main Stem (Cisco)	2.07
Blue	2.28	San Juan (Bluff)	1.24
White	1.78	Lee Ferry	2.11
Animas	1.66		

The immediate conclusions to be drawn from this table are, (1) All but a single precipitation station have less low-frequency variance than would be expected from a random series (1.0). (2) All stream-runoff stations have more variance, most by a factor two, than a random series would be expected to have. It is precisely the latter fact that suggests that the runoff data is non-random in nature. It should be emphasized again, that the differences in Table VII-1 and the Tukey spectra concerning the amount of zero-frequency variance, arise because of the difference in band-width of the two analyses.

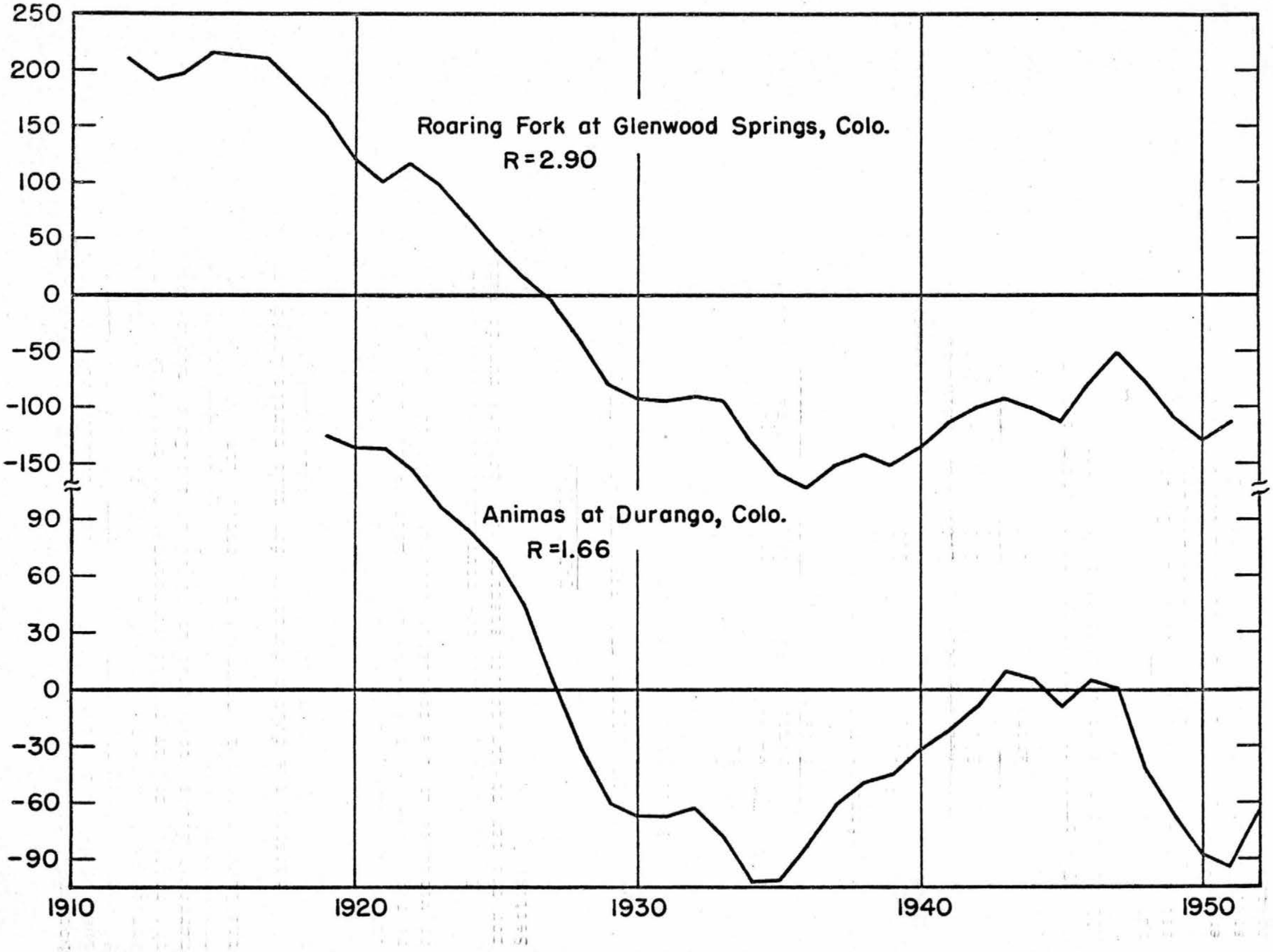


Figure VII-7. Streamflow treated with Low-frequency Band-pass filter.

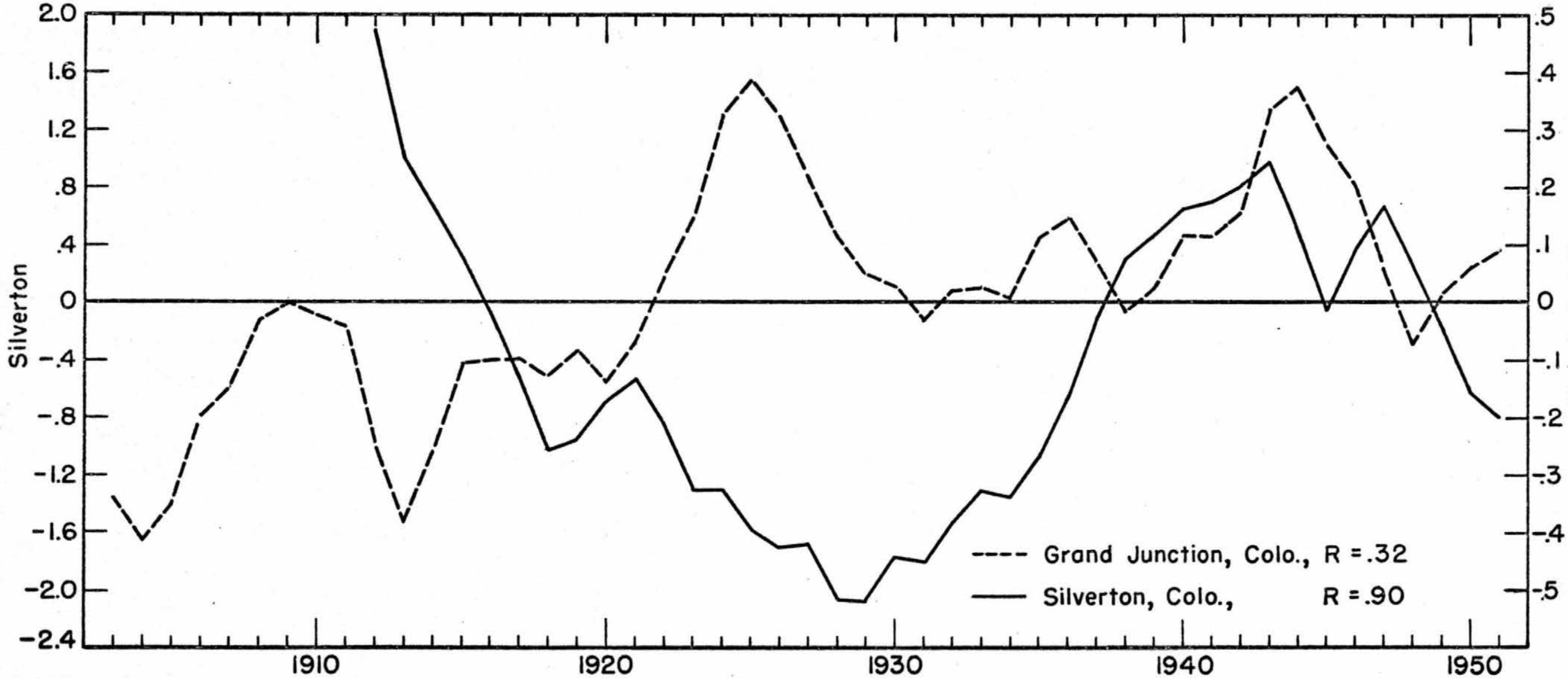


Figure VII - 8. Winter Precipitation totals treated with low-frequency band-pass filter.

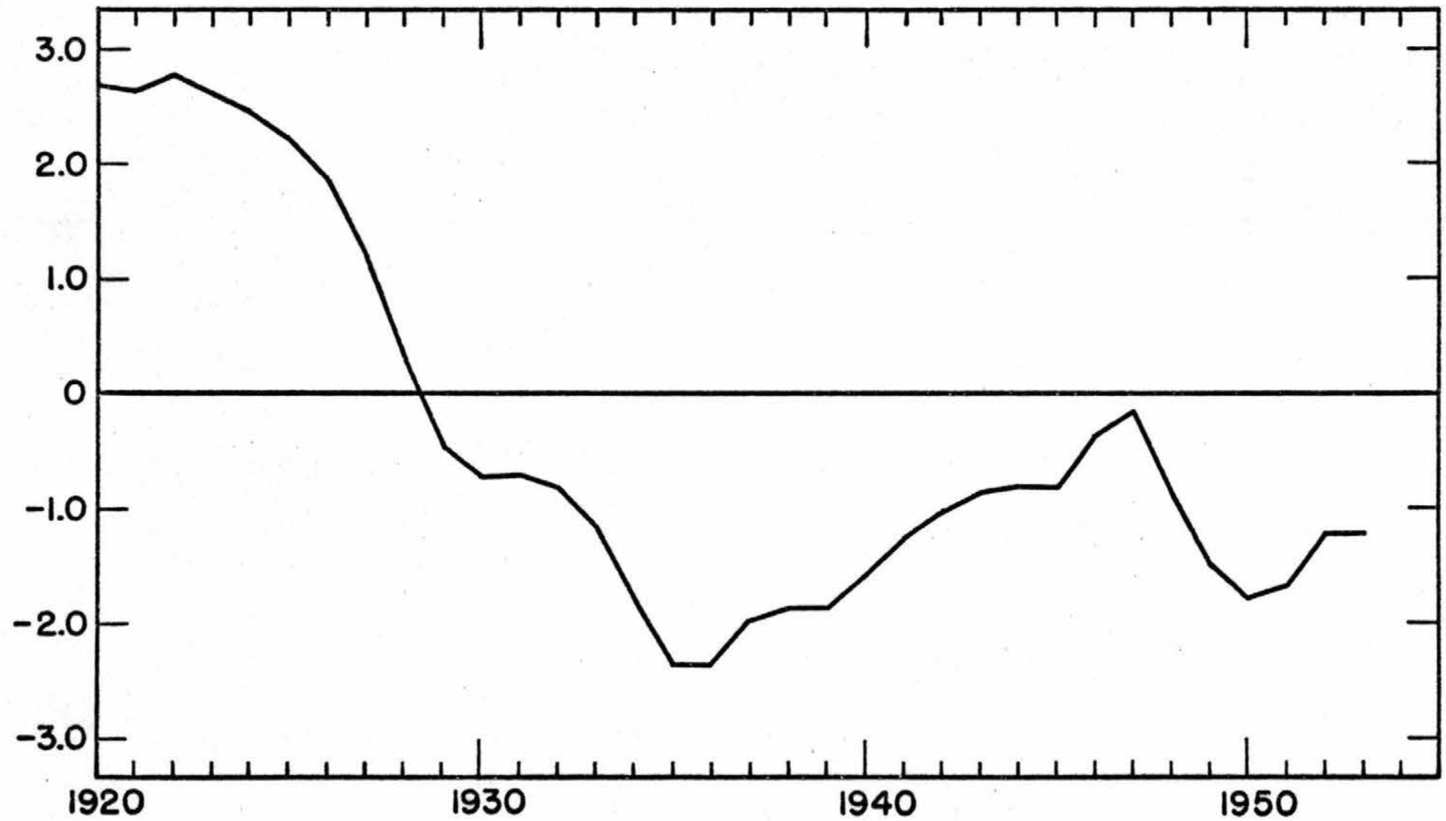


Figure VII-9. Lee Ferry Water-Year Stream Flow treated with Low-Frequency Band-Pass Filter. R = 2.11.

It is also of interest to plot each of the series after filtering as was done in the previous discussion on correlation with solar activity. This has been done in Figs. VII-7 to VII-9 for selected stations only, in the interest of saving space. The precipitation data, first two figures, indicate that no general trend has occurred in the precipitaion measured in the basin since the turn of the century. The precipitaion stations shown were not picked because they illustrated this point; all ten stations agreed with respect to this conclusion. The runoff data, however, all show a pronounced downward trend. This fact need not be pointed out to most readers and the author makes no claim as to the originality of the illustrated precipitation-runoff ratio trend. Kohler and Linsley made the point very well in 1949 (15).

The possible reasons that Kohler and Linsley gave for the decreasing efficiency of the upper Colorado runoff process are still valid today: (a) sampling variations of precipitation data, (b) sampling variations in streamflow data, (c) increase of natural evaporative losses, (d) increase in consumptive loss due to man-made control of the river. The interesting fact about these conclusions is that all of them could very possibly have worked in the same direction, namely, to decrease the basin efficiency. Reason (d) is certainly valid; the critical question is one of magnitude. Reason (c) is also very probably valid since an increase in temperature over the last 60 years seems a legitimate statistical conclusion. Much less likely, but still possible, are the sampling errors inherent in both precipitation and streamflow measurements. As reason (c) is a factor over which man has no control, an investigation into secular temperature changes over the last century might be of some value.

VIII. AN ADDITIONAL STATISTICAL TEST

A. Introduction

An additional test for persistence in winter precipitation and streamflow was carried out: This test due to Wald and Wolfowitz (40) is identical with the one Friedman applied to south-Texas rainfall data. The test involves a measure of the correlation between successive observations (the serial correlation coefficient of lag one). It is a non-parametric test which means it is not sensitive to the parent frequency distribution from which the data were drawn. The statistic tested is

$$R = \sum_{t=1}^{n-1} X_t X_{t+1} + X_n X_1 \qquad t = 1, 2, 3, \ldots n$$

The magnitude of R gives a measure of the year-to-year persistence present in the precipitation or streamflow data. Wald and Wolfowitz have shown that the magnitude of R may be tested for significance if N exceeds about 20. Computing the mean of expected value of R from

$$M(R) = \frac{S_1^2 - S_2}{N - 1}$$

and its variance from

$$\mathrm{Var}\ (R) = \frac{S_2^2 - S_4}{N - 1} + \frac{S_1^4 - 4S_1^2 S_2 + 4S_1 S_3 + S_2^2 - 2S_4}{(N - 1)\ (N - 2)} + M(R)^2$$

where

$$S_k = \sum_{1}^{N} X_t^k.$$

When M(R) and Var(R) are computed we may specify the level of significance by requiring that the mean of R exceed by a given amount its standard deviation. In the case of the 5% level,

$$\frac{R - M(R)}{\sqrt{\mathrm{Var}(R)}} > +1.64$$

for the null hypothesis to be rejected, that is, for the data at hand to exhibit more persistence than a series of random numbers.

B. Results

Table VIII-1 below gives the stations tested, the magnitude of

$$\frac{R - M(R)}{\sqrt{Var(R)}}$$

and the decision concerning the null hypothesis. From this test, we conclude that the Lee Ferry runoff data contain an element of persistence which, with a risk of five out of one hundred, can be said to be too great for a series of random numbers. For the precipitation stations, on the other hand, we accept the null hypothesis that winter precipitation totals are random numbers. Agreement with the randomness tests of the spectral analysis is good. The tests agree on the precipitation data and disagree on the Lee Ferry Virgin Flow data with the level of significance of the 5% value. Note that with the spectral analysis test the corrected flow was non-random on the 5% level while the virgin flow figures had a spectrum which did not reach the 5% level of significance. (Disagreement is also noted on Taylor at Almont and Fraser streamflow.) It is with some assurance, then, that we can conclude that the stream-runoff values at Lee Ferry exhibit significant year-to-year persistence, whereas the precipitation figures do not. The smaller, higher streams, also, appear to exhibit random flows.

TABLE VIII-1

Station	N	$\frac{R-M(R)}{Var(R)}$	Null Hypothesis
November-April precipitation			
Silverton	52	+0.64	accept
Montrose	56	-1.00	accept
Ames + Trout Lake + Silverton	43	+0.03	accept
Ignacio	46	-1.08	accept
Delta	52	-0.42	accept
Dillon	44	-0.30	accept
Fraser	42	+0.19	accept
Shoshone	49	-1.15	accept
Gunnison	54	+0.19	accept
Grand Junction	61	-1.13	accept
Streamflow			
Lee Ferry gaged	46	+1.65	reject
Lee Ferry gaged plus corrections	46	+1.69	reject
Lee Ferry virgin (Leopold)	61	+1.96	reject
Lee Ferry, April-July gaged	46	+1.32	accept
Animas (Durango)	46	+1.56	accept
San Juan (Rosa)	48	+1.00	accept
Roaring Fork (Glenwood Springs)	52	+1.17	accept
Gunnison (below tunnel)	54	+1.18	accept
Taylor (Almont)	47	+2.22	reject
Blue (Dillon)	47	+0.76	accept
White (Meeker)	47	+1.09	accept
Fraser	47	+1.82	reject
Colorado (Hot Sulphur Springs)	52	-0.16	accept

IX. HEMISPHERIC UPPER-AIR CIRCULATION PATTERNS AND PRECIPITATION IN WESTERN COLORADO

A. Introduction

It was suggested to the author by Mr. N. MacDonald, that a follow-up to an earlier study by N. LeSeur (23) might be of value for the overall objectives of the project. Therefore, as a side-study, an objective test of a specific suggestion made by LeSeur was undertaken.

Relations between large-scale (i.e. hemispheric) perameters which describe the general circulation of the atmosphere and regional and local meteorological parameters are of extreme importance because of the fact that dynamical prediction schemes will involve the former, whereas the latter are of practical interest. For some time, the tropospheric circulation of the northern hemisphere has been recognized to be an eccentric one; that is, the center of circulation of the circumpolar westerly winds quite often does not coincide with the geographic pole

An excellent discussion of this feature together with a treatment of some of the problems in characterizing hemispheric circulation resulting from such an eccentricity was given in 1954 by N. E. LeSeur (23). LeSeur suggested, in addition to other things, that during periods when the degree of eccentricity of the circumpolar westerlies was marked, preferred climatic anomalies occurred over the western United States. Specifically, he suggested a negative correlation between vortex eccentricity and precipitation anomaly (that is, deviations from long-term expected value) in the southwest. It is the purpose of this inquiry to examine LeSeur's suggestion further by using a large amount of data.

B. Treatment of Basic Data

For this study, ordinary harmonic analyses of five-day average 500-mb height about 50° N. Latitude were used. These analyses were kindly supplied to Mr. Norman J. MacDonald by Mr. Y. Arai of the Meteorological Agency of Japan. The data covered roughly a ten-year interval of non-overlapping five-day intervals, extending from 1 January 1947 to 6 March, 1957 for the winter months of December, January, and February only. The period of time between 6 March and 1 April each year was analyzed in like fashion by Mr. Frank Weinhold so that the months of December through March for the ten-year period were completely covered. For use as monthly means, the five-day periods were so combined as to give a thirty-day average most closely coinciding with the actual calendar month. In addition, monthly average 500-mb charts for the same winter months for the winters 57-58 and 58-59 were analyzed to give additional data. The height data were read at 10° longitude intervals and 18 harmonics computed. Only the amplitude and phase of the first harmonic (A_1) will concern us here. As LeSeur points out, a mean of the positions of the circulation pole determined from a number of contours is to be preferred in order to best approximate the circulation pole. With Arai's data, only one position of the circulation pole is available, namely that determined by the contour height variation along 50° N. latitude. This estimate, however, should be very well correlated with the circulation pole determined by a number of contours.

In Fig. IX-1 are plotted the amplitude of the first harmonic against the phase angle. In agreement with LeSeur's conclusions, the marked preference for the eccentricity to extend down the 180th meridian is quite noticeable, particularly when the amplitude is large. In an effort to ascertain if a large amplitude of the wave number one implied a preferred positioning of the higher wave numbers, phase angles of the 2nd, 3rd, and 4th harmonics were plotted against the amplitude, A_1. No correlation was detected in all instances.

C. Investigation of Precipitation Anomalies

To check LeSeur's suggestion that specific anomalies occur over the southwestern United States during periods of large eccentric circulation, precipitation records for a number of Colorado western slope stations were tabulated. Attempts to formulate a precipitation index to correlate with the A_1 values were hampered by the fact that a large number of five-day periods were precipitationless. Averaging over longer periods than five days is, of course, a method of obviating the foregoing difficulty. Therefore monthly averages of A_1 were obtained and plotted against a number of different precipitation indices. The first (and only one presented here) was a simple average of fifteen Colorado western slope stations; these stations were chosen on the basis of geographic distribution, elevation, and lack of station moves during the ten-year period. These stations are listed in Table IX-1.

Fig. IX-2 shows a plot of average monthly amplitude of the first harmonic against the 15-station monthly precipitation total. No noticeable relationship is suggested and what little correlation is present must be due to the two points representing very large monthly precipitation totals at relatively small amplitudes. Of course, if more points representing large amounts of precipitation were available, the question could be better answered.

Fig. IX-3 presents a month-by-month plot of the same two parameters for the thirteen-year period. There is a noticeable trend in the annual averages of A_1, with low values in 1947 and 1959 and high values in 1952 and 1953. Such a trend is very well negatively correlated with the last sunspot cycle, but any claim of a relationship would be premature and unjustified. Incidentally, the trend just mentioned can be shown to be the explanation for the sunspot-500 mb circulation pattern

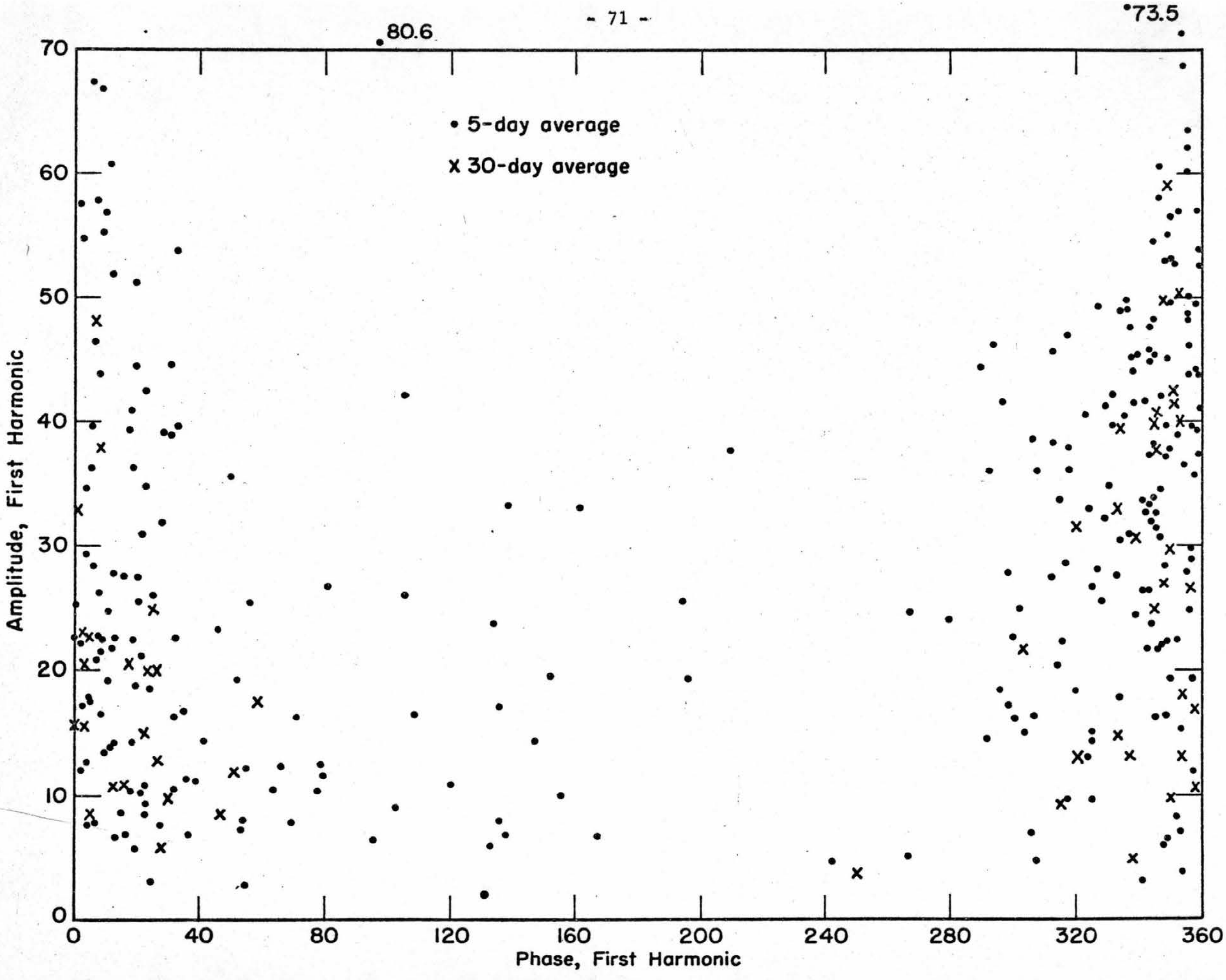


Figure IX -1. Amplitude vs. Phase, First Harmonic.

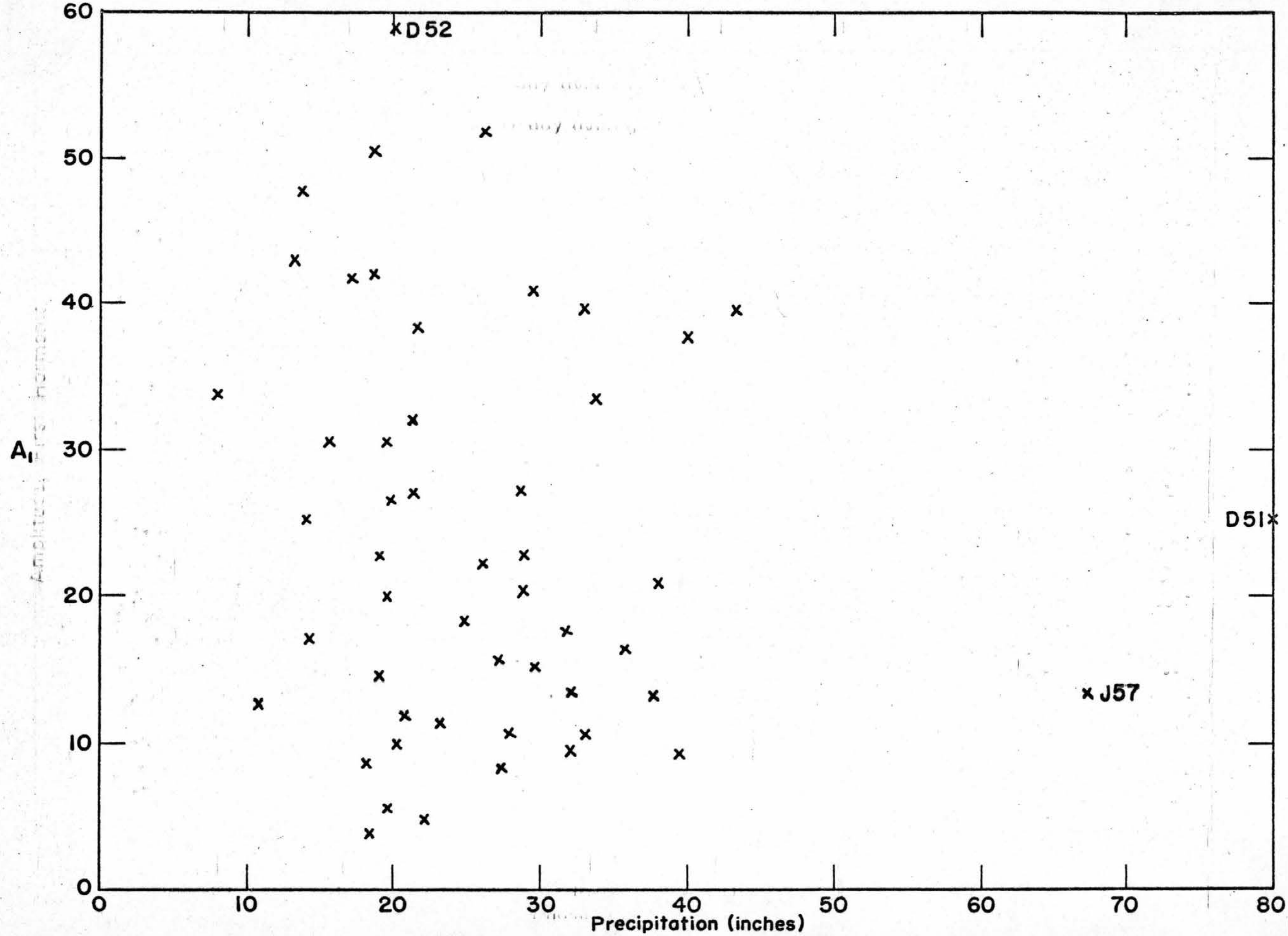


Figure IX - 2. Amplitude of First Harmonic (monthly mean) vs. Fifteen Station Monthly Precipitation.

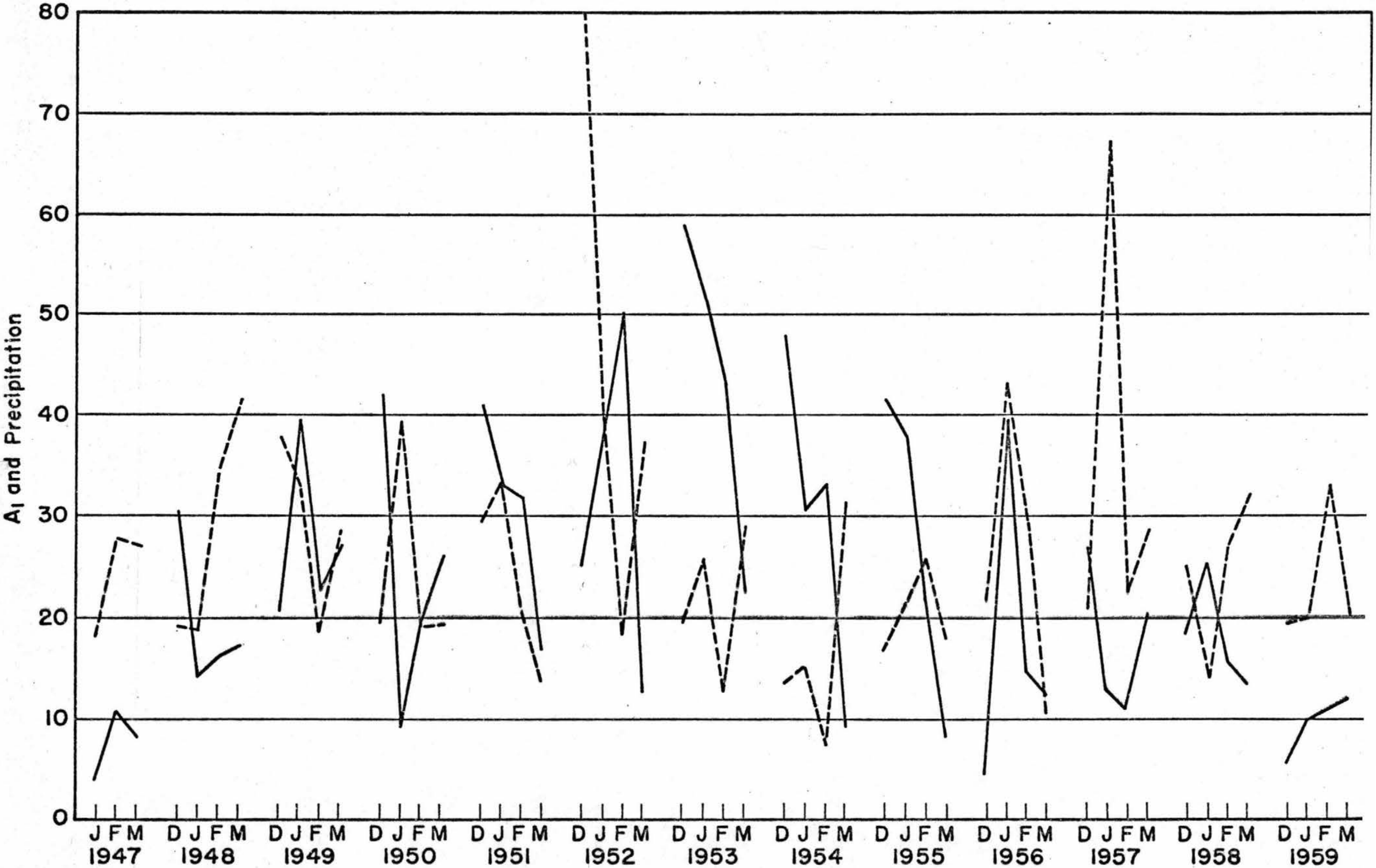


Figure IX-3. Month-by-month $\overline{A}_1$ vs. Monthly 15-station Precipitation, winter 1947-1959.

connection claimed by Arai (4). From the figure, a good negative correlation can be noticed in some years, namely 1950, 1952, 1957, and 1958. In 1947, 1956, and 1951, however, a positive correlation is suggested.

We can conclude from the two figures that no good relationship exists between the average eccentricity of the circumpolar westerly vortex and monthly precipitation amounts in western Colorado.

Rejecting the suggestion of a relationship between the eccentricity of the circulation and precipitation in the upper Colorado Basin on a monthly basis, the next problem is to determine if such a relationship holds on a shorter-term basis. The difficulty of specifying a precipitation index in the case of five-day periods has already been mentioned. A plot of A_1 over the five-day periods against total precipitation in the five-day period (not reproduced) shows a congestion of points near zero precipitation, as is to be expected. To circumvent this difficulty, the problem was redefined slightly. Twenty five-day periods were chosen from the ten-year 1947-1957 analysis on the basis of the magnitude of the five-day A_1 value. The periods with the ten highest and ten lowest amplitudes were chosen. The five-day total precipitation from twelve of the fifteen stations were obtained for each period. No five-day period of those twenty chosen was completely precipitationless. Three stations did not have daily precipitation records during the desired periods and were therefore omitted. Table IX-2 lists the five-day periods with the highest and lowest A_1 values together with the total precipitation and the rank (from 1 to 20) of that precipitation total. To avoid complications introduced by using estimated five-day precipitation distributions and sampling problems, a simple, ranking, non-parametric test was used. This test asks the question of whether the observed ranking of the precipitation amounts of the ten most eccentric and the ten least eccentric five-day periods could have occurred by chance. If any linear correlation between A_1 and five-day precipitation totals exists, it would be detectable by this test. These ranking tests are completely described by Moroney (29) or Seigel (35) for example.

Briefly the test determines the probability of getting a rank total of 85 (in this case) from the lowest of the two sums of ten ranks, when the expected sum is 105. Using the null hypothesis at the 5% level, the rank sum of 85 is not significant; a rank sum of 78 would be needed. Thus, although the relationship is definitely in the right direction, it is not significant at the 5% level and we must conclude that the relationship is not proven. The fact that all twenty cases may not be independent only serves to reduce what significance was found.

D. Discussion of the Physical Interpretation of LeSeur's Correlations

LeSeur makes the following conclusions about the eccentric circumpolar vortex. The circumpolar westerly flow is strongest about the vortex center at times of the greatest eccentricity values; and futher, at such times the average "meridional" flow, that is, flow normal to the asymmetric vortex, is the least. Although not very explicit about the explanation of the temperature and precipitation anomalies in terms of these conclusions, apparently LeSeur felt that at times of great eccentricity, similar 500 mb flow patterns tended to occur over the United States, in turn producing similar temperature and precipitation patterns. The data analyzed here, however, suggest that there is no good relationship between the amplitude of the first harmonic and the phase of any of the higher harmonics. Lacking any additional quantitative measures of circulation, we may conclude that insofar as the phase angles of the higher harmonics indicate the preference for ridge or trough conditions over the southwestern United States, no relation exists between the eccentricity of the vortex and the preference for ridging or troughing over the western United States.

A subjective evaluation of the 500 mb charts for the ten five-day periods with the greatest eccentricity resulted in the feeling that a relationship might be present but that quantitative measures need to be worked out. The evaluation revealed that these periods were marked by very strong zonal flow in the Pacific displaced equatorward of its "usual" position (this is in good agreement with LeSeur's quantitative conclusions) and that ridging was definitely present either slightly off the west coast of the United States or over the western part of the country. Such a configuration has the attributes of the so-called blocking situations Rex, (33) and if this is to be the preferred situation, the ridging might result in lower than normal precipitation in the Upper Colorado Basin.

From the foregoing quantitative study, however, we can conclude that any relationship between the degree of eccentricity of the spheric circumpolar vortex and precipitation anomaly along the western slopes of the Colorado Rockies must be a weak one.

TABLE IX-1

STATIONS USED IN COMPUTING PRECIPITATION TOTALS

1. Aspen
2. Cascade
3. Crested Butte
4. Dillon
5. Grand Lake, 1 N
6. Green Mountain Dam
7. Ignacio
8. Meeker
9. Montrose #2
10. Rifle
11. Shoshone
12. Silverton
13. Steamboat Springs
14. Taylor Park
15. Winter Park

TABLE IX-2

Ten Periods with Greatest A_1-Precipitation-Rank

March 11-15, 1948	3.54	7
March 2-6, 1950	1.89	13
Nov. 27-Dec. 1, 1951	0.03	20
Dec. 2-6, 1952	1.95	12
Dec. 22-26, 1952	0.84	16
Dec. 27-31, 1952	0.43	18
Jan. 1-5, 1953	2.61	9
Jan. 16-20, 1953	7.22	3
Dec. 2-6, 1953	3.15	8
Dec. 17-21, 1954	0.26	19
	Sum	125

Ten Periods with Smallest A_1-Precipitation-Rank

Feb. 25-March 1, 1947	12.45	1
March 16-20, 1948	8.67	2
March 26-30, 1948	5.21	5
Feb. 10-14, 1949	2.21	11
Jan. 6-10, 1950	2.51	10
March 12-16, 1951	1.07	14
March 21-25, 1952	4.90	6
Feb. 15-19, 1956	7.01	4
March 16-20, 1956	0.53	17
Jan. 16-20, 1957	0.85	15
	Sum	85

X. CONCLUSIONS

In this section the author will attempt to sum up the salient points of the investigation. The conclusions pertaining to each of the five questions outlined in the Introduction will be given separately.

(1) What are some of the basic statistical relationships of measured precipitation to stream-runoff? The runoff process can be represented by a simple system in which the actual runoff is a variable but generally small residual of the larger quantities, precipitation and evapotranspiration. The radiative variability of runoff values will always be larger than the relative variabilities of either precipitation or evaporative losses because precipitation and evaporation averaged over seasons are to some degree statistically independent.

Stream-runoff amounts are better areally correlated than point precipitation amounts. There are two reasons for this fact which are obvious from physical considerations but there is no statistical method whereby the two might be separated with present data. The first is, that point precipitation data give only an imperfect estimate of the true areal 'coherence' of precipitation. This is, of course, due to the fact that such data include to an unkown degree precipitation variability due to small-scale gauge effects and very small-scale local terrain effects.

The second reason why runoff values may be better spatially correlated than precipitation data is because evaporative and transpirative may be better correlated than the precipitation data. Because, as mentioned above, stream-runoff is in semi-arid climate a small residual, its areal correlation will be influenced, that is, increased, if evapotranspiration is more areally coherent than precipitation. Again, there seems to be no way at present of assigning quantitative numbers to the magnitude of each of these two causes.

A simple plot of specific yield against point precipitation amounts or runoff amounts vs. precipitation amounts indicates that ordinary linear regression analysis cannot be applied because the residuals from a regression line are not random, but serially linked. This result has been well known for some time and present analyses (13), (36), circumventing this difficulty are being utilized. The point is made here to emphasize the fact that the serial correlation structure of streamflow data, whatever it may be, pervades all types of statistical analyses of the data and must, in some manner or other, be taken into account.

(2) Is there any statistical predictability in precipitation and streamflow amounts, or are they indistinguishable from a set of random numbers?

Two different statistical tests were used to test the null hypothesis that winter precipitation and water-year stream-runoff amounts from the upper Colorado Basin are random numbers. On the 5% level the null hypothesis was accepted in all cases involving precipitation and rejected only once in the case of runoff. This rejection was for the gauged water-year plus published corrections (USGS Water Supply Papers) for Lee Ferry 1914-1957. The type of non-randomness exhibited by the stream-runoff data was that of an auto-regressive or moving average nature, with the stream-runoff amounts exhibiting various degrees of persistence from water-year to water-year. This means that a small measure of statistical predictability is present and can be utilized. By making use of the estimated variance spectrum of the Lee Ferry flows, some statistical estimates of future flows may be obtained. In this regard, Leopold's paper (22) was shown, by these independent approaches to have given reasonable estimates of the variability of future means of the Lee Ferry flow. With the limited data sample available, the confidence with which the variability of five or ten year means, say, can be estimated is not very great, but rejecting the null hypothesis results in the fact that this variability must be greater than would be expected if the Lee Ferry data were random in time.

Attempts to derive a statistical model to represent the Lee Ferry flow resulted in a very crude model because of the limited length of the data and the resulting crudeness of the estimated spectrum. For purposes of the probability analysis a simple Markov model $X_t = 0.25\ X_{t-1} + \epsilon$ was adopted.

Various other important aspects of the question of predictability were investigated. No direct linear correlation (or relationship) between Basin precipitation and streamflow with sunspot number was detected. This negative result means that no predictive element can be gained by using the quasi-periodicity of the sunspot numbers.

The pertinent discussion to question (3), "what significance has an answer to (2) for practical hydrological purposes?" was not explicitly included in the previous sections. However, the general discussion of Moran's The Theory of Storage, and the mention of a synthetic hydrology in Section I, apply. The question of the time-series structure of streamflow has relevance not only because it may allow some predictive scheme to be used, but because ordinary 'cook-book' type statistics used by many people in decision making positions are affected. For example, the emphasis laid on means or averages of streamflow over different periods of time, and the attempts by legal processes to define the long-term mean of a river's discharge, are cases in point. When dealing with woefully short historical records of a highly variable quantity such as streamflow, it should not surprise anyone that the so-called long-term means prove to be unstable ones.

But the critical point is in deciding quantitatively just how unstable such means are. If the streamflow were serially random (thus, each year independent of the adjacent years,) sixty years of data would give us sixty, independent estimates of the true long-term mean. In actual fact the data are frequently not random, resulting in a decrease in the number of independent estimates of the mean, and for that matter, any statistic computed from the data. The non-random component of the series need not be very large to make an appreciable difference in this regard. (See Leopold Fig. 5.) Of practical importance, then, is the fact that our historical data are deceiving us into thinking we have more insight into averages, low flows, etc., etc., than we really do.

The magnitude of the random component in a hydrologic system will be large because according to the findings herein, the precipitation which initiates the process is random. Any non-randomness in the streamflow data is thus reduced by physical processes, such as evapotranspiration, regulation, etc., operating upon the fallen precipitation. In addition, inhomogenieties in streamflow data introduced by changes in streamflow regimen or gauging methods and by artificial regulation and usage could introduce a non-random component. If it were possible to decide upon a statistical model representing the effects of these processes upon the initially random data, then future flows could be simulated by use of this model and random numbers. As pointed out above, this is the idea behind a synthetic hydrology.

Although one of the principal objectives of this investigation was the development of such a statistical model, the attempt must be said to have been only partially successful. The Markov process is the simplest and easiest to use and Langbein has already shown (19) how Moran's theory can be modified to a streamflow model using the Markov process. Our historical data are of such duration, however, that we cannot reliably distinguish between the Markov model and others which might be mentioned. As an example, the question of the order of the statistical model is particularly important for the filling of Powell Reservoir. If a Markov model is truly appropriate then Leopold's or the author's estimates of future five or ten year means may be used to adjudge statistical variations to be encountered during the critical filling period. This is because a Markov process is of first order: that is, dependent only on one previous value. Suppose now, that the true state of affairs is such that stream-runoff is a higher order process--say tenth. This means that values as much as ten (years) ago are still "affecting" the current (years') value. Then the question that must be asked of Leopold's probability approach in questions involving the filling of Powell Reservoir is "What is the variability of <u>the next</u> five-year or ten-year flows?" not "What is the variability of five-year mean flows <u>taken at random</u>?" which is what Leopold and Section VI actually are concerned with. The very practical answer as to what to expect of Lee Ferry in the next ten years thus depends upon a property of the data which we cannot at present learn much about.

To sum up the foregoing discussion, the determination of the serial correlation structure, or non-random, element in streamflow series is practically important because (1) significant serial correlation increases the amount of data necessary to gain knowledge of means, ranges, and other statistics. (2) The variability of means of a given length, and hence the uncertainty involved in such numbers, is increased if serial correlation is present. (3) A significant amount of such correlation makes a certain amount of statistical prediction of future flows feasible, and if sufficient knowledge of this correlation exists, the knowledge may be made use of in linear programming models, etc., for purposes of optimum basin operations.

Question (4), "What are the causes of the observed decrease in runoff efficiency in the past 50-60 years?" is one question which did not receive the attention during the course of the project that it deserved. In Section II the runoff ratios were shown for four small basins, all of which upon subjective examination indicate a decrease in runoff efficiency since the records began. The filtering process described in Section VII and the resulting filtered predipitation and streamflow records, Figs. VII 7 to 9, also show the effect as the streamflow data all indicate a downward movement whereas the precipitation data do not. Such results, however, merely confirm that the decrease has taken place, as indeed, many people have pointed out. The author would like to venture the opinion that attributing such a decrease to sampling errors in precipitation and streamflow as included in the suggestions by Kohler and Linsley in 1949 (15) does not now seem very likely although it cannot be dismissed altogether. Two other reasons, an increase in man-made losses and an increase in natural evaporative losses, seem much more probable. The first has certainly taken place; the question of the magnitude of these losses has not been resolved. The second, an effect not mentioned by Kohler and Linsley, has quite probably also been in operation. There is good evidence (17), (27) for a secular temperature rise having taken place in the Colorado Basin (and elsewhere) covering the period of our records. To the degree that increased evaporation and transpiration was brought about by this temperature rise, the observed loss in runoff efficiency can be explained by natural processes. Both of these mechanisms should have worked in the same direction, namely, to increase evaporative losses and thereby decrease the runoff efficiency. It should be emphasized that the latter cause is beyond the ability of mankind to control and simply reflects the natural variability of the runoff process (the remarks of Section II and in the first part of these Conclusions are appropriate here).

It might be appropriate to add that the direction of the man-made effect upon future flows is, of course, predictable. The influence natural evaporative losses will have on future flows is not predictable since long range meteorological forecasts are not now possible. The point, once again, is simply that streamflow is a much more highly variable quantity than our short, historical records indicate to us.

APPENDIX I

Running Averages, Spectra, and Time Series

by Max Woodbury* and Paul Julian

A model of the annual series of water flow at Lee's Ferry is that it is a sample from a stationary time-series. Luna Leopold (22) has made use of this feature in his report on probability analysis of Colorado streamflow data. Leopold notes that the average streamflow over 5, 10, 15, and 20 year periods shows a variance in excess of that calculated for a random series and correctly attributes it to correlation between the streamflow in different years. For a number of streams the greatest deviation from randomness as measured by the calculated variances occurs in the Niagara River flow at Buffalo, while the Mississippi at St. Louis shows very little departure from a random series for 5, 10, 15, and 20 years.

Leopold aggregates streamflow data from a number of rivers and uses the modal value of the variance of various averages as a correction in his further work on the Colorado at Lee Ferry. Here we will make better use of the information available to get at approximate confidence intervals for the variance of various averages.

Section 1. Moving Averages in Time-Series

Elementary concepts will be presented briefly; the reader is referred to these references (6), (12), (41) for a more thorough background. Here we need the notion of a transfer (or system) function.

If a (stationary) time-series is "smoothed" by a (uniform) moving average we find on inspection that it "removes corners" and short period variations but allows long period variations to pass with little change in amplitude and phase. It, of course, obliterates variations whose period is equal to the length of the (uniform) moving average. To examine the effect we will apply a moving average to various periodic functions to see what effect they have.

We consider uniform centered moving averages of the form

$$M_{2n+1}\left[X_t\right] = \frac{X_{t-n} + X_{t-n+1} + \ldots + X_{t+n}}{2N + 1} \qquad (1)$$

where n can be any non-negative half integer.

We first derive the result of smoothing a complex exponential; from that we derive the effect on sines and cosines.

---

* College of Engineering, New York University.

$$M_{2n+1}\left[e^{i\omega t}\right] = \frac{e^{i\omega(t+n)} + \dots + e^{i\omega(t-n)}}{2n+1}$$

$$= e^{i\omega t}\left(\frac{e^{i\omega n} + e^{i\omega(n-1)} + \dots + e^{-i\omega(n-1)} + e^{-i\omega n}}{2n+1}\right)$$

$$= e^{i\omega t}\left(\frac{1 + 2\cos\omega + 2\cos 2\omega + \dots + 2\cos n\omega}{1+2n}\right) \qquad (2)$$

$$= e^{i\omega t}Y_{2n+1}(\omega) \qquad (3)$$

where also,

$$Y_{2n+1}(\omega) = \frac{\sin\left(\frac{2n+1}{2}\right)\omega}{(2n+1)\sin\frac{\omega}{2}}.$$

Since

$$M_{2n+1}\left[e^{-i\omega t}\right] = e^{-i\omega t}Y_{2n+1}(-\omega) = e^{-i\omega t}Y_{2n+1}(\omega)$$

it is readily seen that

$$M_{2n+1}\left[\sin(\omega t)\right] = Y_{2n+1}(\omega)\sin(\omega t)$$

$$M_{2n+1}\left[\cos(\omega t)\right] = Y_{2n+1}(\omega)\cos(\omega t)\ .$$

Or more generally, if the representation

$$X_t = \int_{-\pi}^{\pi} e^{i\omega t}dZ_x(\omega) \qquad (4)$$

is valid, as it is for stationary time-series (functions obtained by adding together periodic components, etc.,) then

$$M_{2n+1}\left[X_t\right] = \int_{-\pi}^{\pi} e^{i\omega t}Y_{2n+1}(\omega)\,dZ_x(\omega) \qquad (5)$$

where $Y(\omega)$ is, then, the transfer function or system function of the averaging process.

If the moving average is a weighted average then the transfer function is different. In particular, if

$$y_t = \frac{a_0X_t + a_1(X_{t-1} + X_{t+1}) + \dots + a_n(X_{t-n} + X_{t+n})}{a_0 + 2a_1 + 2a_2 + 2a_3 + \dots + 2a_n} \qquad (6)$$

then

$$y_t = \int_{-\pi}^{\pi} e^{i\omega t} Y_a(\omega)\, dZ_x(\omega) \tag{7}$$

where

$$Y_a(\omega) = \frac{a_0 + 2a_1 \cos\omega + 2a_2 \cos 2\omega + \dots + 2a_n \cos n\omega}{a_0 + 2a_1 + 2a_2 + \dots + 2a_n} \tag{8}$$

An interesting weighted average is one that uses binomial weights:

$$y_t = \frac{1}{2^N} \left[\binom{N}{0} X_{t-N/2} + \binom{N}{1} X_{t-N/2+1} + \dots + \binom{N}{N} X_{t+N/2} \right]$$

For this average,

$$Y^N(\omega) = \left[\cos\frac{\omega}{2}\right]^N$$

The $Y(\omega)$ graphs for some of these averages are shown in Fig. 1. Note that by operating on $X(t)$ with a weighted running average as in (6), the variance of the resulting series $y(t)$ may be easily obtained, and since the transfer function is known from (8) an estimate of the spectrum in a particular band may be obtained. That is,

$$S^2_{y(t)} = \int_{-\pi}^{\pi} \left| Y(\omega) \right|^2 \Phi(\omega)\, d\omega \tag{9}$$

follows directly from (7).

Section 2. Variance and Spectra

The spectrum effectively gives an analysis of variance of a time-series. The variance is partitioned into components corresponding to various frequency bands. The variance of a smoothed series can be found from the original spectrum and the transfer function.

$$\sigma^2_x = \int_{-\pi}^{\pi} E \left| dZ_x(\omega) \right|^2 = \int_{-\pi}^{\pi} S(\omega)\, d\omega$$

and

$$\sigma^2_{M(x)} = \int_{-\pi}^{\pi} \left| Y_M(\omega) \right|^2 S(\omega)\, d\omega \tag{10}$$

where $Y_M(\omega)$ is the transfer function of the smoothing operation $M(X)$ as suggested in the previous section.

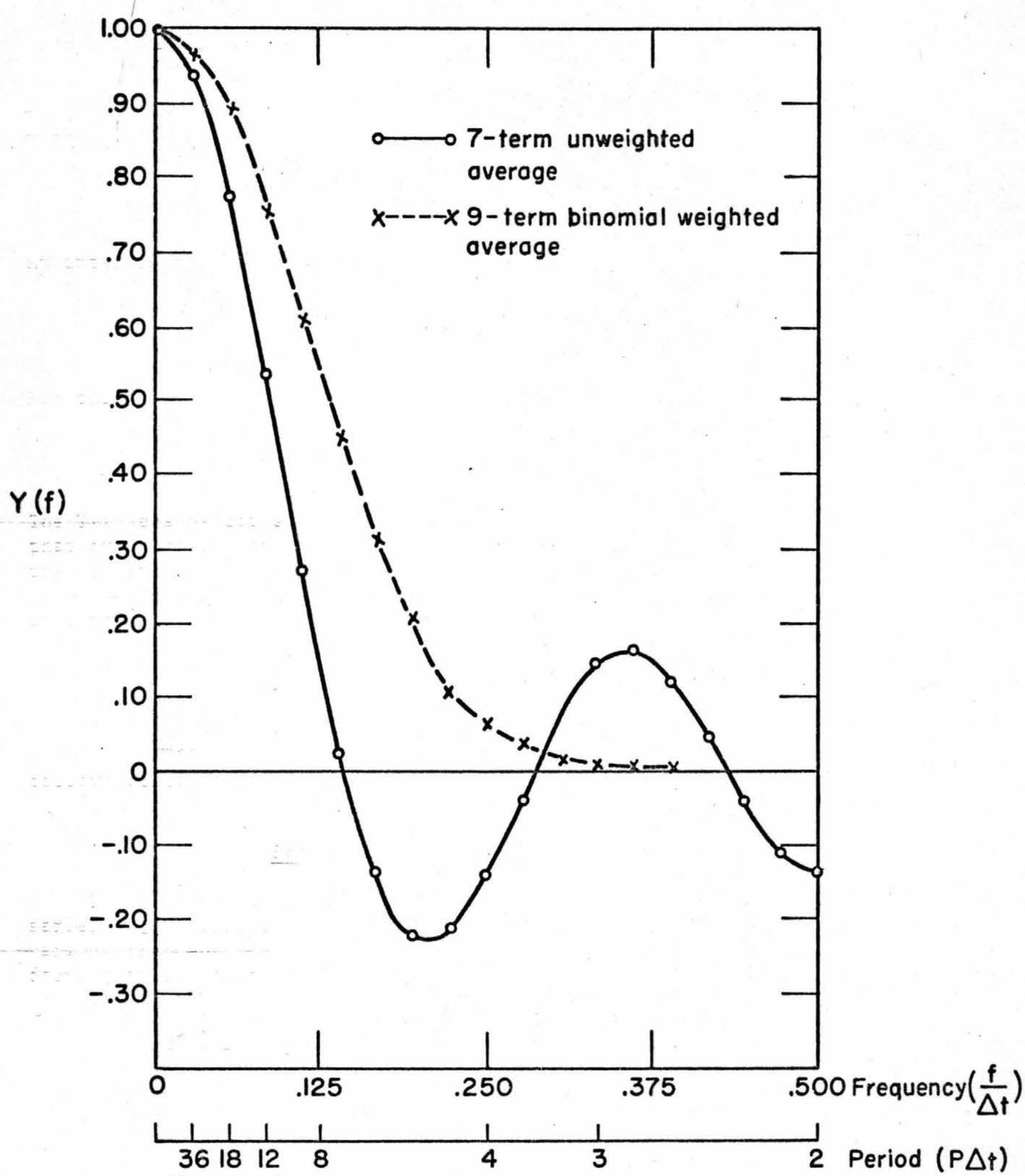


Figure 1. Transfer function curves.

If the original time-series is a series of independent terms, then the spectrum is "white," i.e.,

$$S(\omega) = \frac{\sigma_x^2}{2\pi} .$$

Then

$$\sigma_x^2 = \int_{-\pi}^{\pi} \frac{\sigma_x^2 \, d\omega}{2\pi} = \sigma_x^2$$

The variance of a moving average of N terms is

$$\sigma_N^2 = \frac{\sigma_x^2}{2\pi} \int_{-\pi}^{\pi} \left[\frac{\sin \frac{N\omega}{2}}{N \sin \frac{\omega}{2}} \right]^2 d\omega = \frac{\sigma_x^2}{N}$$

if the spectrum is 'white'.

If, however, the spectrum is 'colored', i.e., $S(\omega)$ is not constant, then

$$\sigma_N^2 = \int_{-\pi}^{\pi} \left[\frac{\sin \frac{N\omega}{2}}{N \sin \frac{\omega}{2}} \right]^2 S(\omega) \, d\omega \qquad (11)$$

so very rough estimates of certain functionals of the spectrum are available from estimates of σ_M^2

M = 5, 10, 15, 20, as provided by Leopold. Better estimates of the spectrum may be obtained by other means. See Blackman and Tukey (6).

Numerical estimates of the functions $Y_M(\omega)^2$ will be provided by M = 5(5)20 in Appendix III.

<u>Section 3. Relation Between Serial Dependence of Model Time-Series and Variance Spectra.</u>

Referring to (6) we note that a one-sided moving average

$$y(t) = M_N \left[X_t \right] = \frac{a_o X_t + a_1 (X_{t-1}) + a_2 (X_{t-2}) + \ldots + a_n X_{t-n}}{a_o + a_1 + a_2 + \ldots + a_n} \qquad (12)$$

would also have a transfer function $Y_a(\omega)$ as in (8), but

$$Y_a(\omega) = \frac{a_o + a_1 e^{-i\omega} + a_2 e^{-i2\omega} + \ldots + a_n e^{-in\omega}}{\sum_{j=0}^{N} a_j} = 1 + \alpha_1 e^{-i\omega} + \alpha_2 e^{-i2\omega} \ldots$$

It then follows from (4) that the spectra of y(t) and of x(t) are related very simply by

$$\frac{\Phi_y(\omega)}{\Phi_x(\omega)} = \left| Y_a(\omega) \right|^2 = \left| 1 + \alpha_1 e^{-i\omega} + \alpha_2 e^{-i2\omega} + \ldots \right|^2 \qquad (13)$$

Thus, knowledge of the spectrum of y(t) will give us exactly the structure of a moving average process operating on a random series

$$(\Phi_x(\omega) = \text{constant})$$

representing the y(t) series.

The inverse of the moving average scheme as (12) is called is the auto-regressive scheme.

$$y(t) = X(t) - b_1 y(t-1) - b_2 y(t-2) - \ldots - b_n y(t-n) \qquad (14)$$

In this case the relationship between the spectra is

$$\frac{\Phi_y(\omega)}{\Phi_x(\omega)} = \left(\left[1 + B_1 \beta_1^{-i\omega} + B_2 \beta_2^{-i2\omega} + \ldots \right]^2 \right)^{-1} \qquad (15)$$

or the inverse of (13).

A qualitative look at (13) and (15) indicates that if the spectrum of y(t) decreases from low to high frequencies ($\omega = 0$ to $\omega = \pi$) the 'weights' a_n of the moving average process decrease with increasing n. Also, however, the same statements can be made of the auto-regressive scheme, the weights increasing in a negative direction (since in (14) the weights are given as negative).

It is pointed out here that knowledge of the spectrum of a time-series enables the representation of the structure of the series through either a moving average or auto-regressive scheme.

APPENDIX II

Some Considerations of the Results of Averaging and Discrete Sampling Operations on the Estimation of Continuous Spectra

The material given here is taken largely from Blackman and Tukey, (6) and is intended only to supply to the interested reader of the main body of the report enough supplemental information to enable an understanding of the procedures used by the author.

The considerations here apply to the digital approximations made of a continuous record, $f(t)$, in order to estimate the spectrum; the continuous record is considered to have been instantaneously sampled, or a value taken, at equally spaced intervals of time, Δt, or the values in discrete time of the record $f(t)$ averaged over the interval Δt.

For mathematical convenience consider the following operator D, to be operating on the continuous record.

$$D_1(t,\Delta t) = \Delta t \sum_{q=-\infty}^{\infty} \delta(t - q\Delta t)$$

The $\delta(t - q\Delta t)$ represents the Dirac or so-called delta function which has the following defined properties:

$$\int_{-\infty}^{\infty} \delta(x - x_0)\,dx = 1 \quad \text{where} \quad \delta(x - x_0) = \begin{cases} \infty & \text{when } x = x_0 \\ 0 & \text{when } x \neq x_0 \end{cases}$$

The mathematical intricasies of this function need not concern us here. Suffice it to say that this operator, when operating on $f(t)$, causes the function to be evaluated at equi-spaced intervals of time, Δt. The record $f(t)$ now has the 'appearance' of a number of infinitely thin spikes spaced Δt time units apart. Such an operation simulates the instantaneous reading of a record, $f(t)$, at equally spaced intervals.

To consider the result this operator has upon the spectrum we merely have to take the Fourier transform of the operator, D. The transform of D defined as above is

$$\sum_{q=-\infty}^{\infty} \delta\left(f - \frac{q}{\Delta t}\right),$$

where the f denotes the frequency scale upon which the transform was made. If the spectrum is to be defined only on the positive frequency scale, and if the Nyquist frequency

$$f_n = \frac{1}{2\Delta t}$$

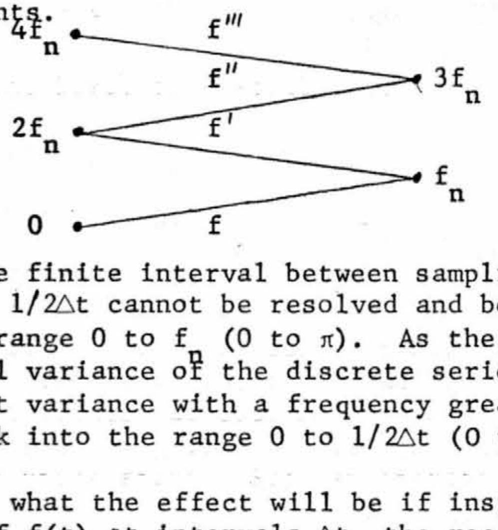
is used, then

$$\int(f) + \sum_{q=0}^{\infty}\int(f - 2qf_n) + \int(f + 2qf_n)$$

Physically this expression says that a delta-function operator operating on a continuous function has a spectrum which is defined as extending to infinite frequencies but where each frequency in the range 0 to f_n (0 to $\pi/\Delta t$, using circular frequency ω, since

$$2\pi f_n = \omega_n = \pi/\Delta t)$$

corresponds to frequency in the range $qf_n + (q + 1)f_n$ as given by the expression. Note that the frequency distribution is symmetrical about $2qf_n$. Blackman, and Tukey thus picture the confusion of frequencies or aliasing, as they term it, as the frequency scale folded upon itself about the qf_n points.



Thus, because of the finite interval between samplings, the Δt, frequencies greater than $1/2\Delta t$ cannot be resolved and become confused with frequencies in the range 0 to f_n (0 to π). As the Tukey estimation scheme contains the total variance of the discrete series within this range it is possible that variance with a frequency greater than $1/2\Delta t$ will be folded back into the range 0 to $1/2\Delta t$ (0 to f_n, 0 to π).

Now let us consider what the effect will be if instead of making instantaneous readings of f(t) at intervals Δt, the record is averaged over Δt, so that in general, $f(t) \neq f(q\Delta t)$. The operator to be considered now is the rectangular time function

$$D_o(t) = 1 \text{ for } \begin{cases} t > -\Delta t/2 \\ t < \Delta t/2 \end{cases}$$

and

$$D_o(t) = 0 \text{ for } \begin{cases} t > \Delta t/2 \\ t < -\Delta t/2. \end{cases}$$

Applying this operator to f(t) results in a series of steps of length Δt where f(t) is, of course, a constant over each Δt.

The Fourier transform of this operator is given by

$$Q_o(f) = \Delta t \frac{\sin(\pi f \Delta t)}{\pi f \Delta t} = \Delta t \frac{\sin(\frac{\omega}{2}\Delta t)}{\frac{\omega}{2}\cdot\Delta t} \; .$$

Squaring to obtain the response of this function on the spectrum the function can be plotted as shown in the following figure.

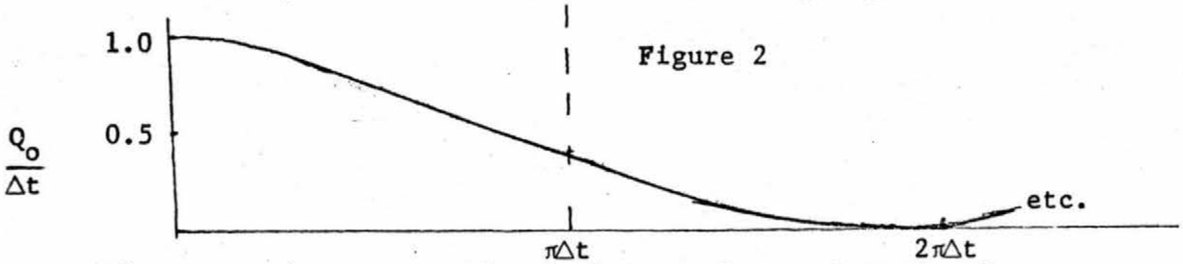


Figure 2

Thus averaging over an interval Δt, taken to be unity for convenience (say 1 sec., 1 day, 1 year, etc.) will have an effect on the spectrum of the original record f(t). At the Nyquist frequency $\omega_N = \pi$, for example, the spectrum will be reduced to about 40% of its actual value.

Combining the two operations just discussed, the aliasing and averaging effects on the spectrum may be written as

$$\Phi^{1}(\omega) = \left[\Phi(\omega) + \Phi(\omega - 2qf_n) =+ \Phi(\omega + q2f_n) \right] \left[\frac{\sin \omega/2}{\omega/2} \right]^2 \qquad \text{II-1}$$

for unit time interval, Δt.

Note that unless the behavior of the true spectrum

$$\Phi(\omega)$$

is known beyond $\omega_N = \pi$ that there is no way of recovering it. Generally enough must be known about the spectrum a priori in order to insure that aliasing of high frequency variance into the principal spectrum (0 to ω_N) is negligible. Note also that averaging over the interval Δt instead of using instantaneous f(t) protects somewhat against this aliasing problem as the function

$$\left[\frac{\sin \frac{\omega}{2}}{\frac{\omega}{2}} \right]^2$$

goes to zero at $\omega = 2\pi$ and is generally small in the region $\omega = \pi$ to 2π.

If a 'white-noise' spectrum is being considered, Fig 2 gives the result of the product of the random spectrum and the "averaging function,"

$$\left[\frac{\sin \frac{\omega}{2}}{\frac{\omega}{2}} \right]^2 .$$

At the Nyquist frequency $\omega = \pi\Delta t$, then the restored spectra would be twice the result of the estimated spectrum after treatment by the 'averaging function.' The 'folding' about $\omega = \pi\Delta t$ would just double the estimate at that frequency. Thus the resulting spectrum would be about 80% of the true white-noise spectrum there. For frequencies less than $\omega = \pi\Delta t$ the total effect of aliasing and averaging would approach 100% rapidly, as can be seen from Fig. 2. For a random spectrum, then, the two effects tend to cancel each other, aliasing adding energy and averaging removing energy in the spectrum near $\omega = \pi\Delta t$.

The spectrum of the streamflow and precipitation may be assumed to be random between 1/2 per year and 1/1 per year except, of course, for the rather sharp peak at the latter frequency arising because of the pronounced annual variation in these quantities. Since, however, the response of the averaging correction is zero at this frequency, no aliased energy is transmitted to the principal spectrum. By using observations made more frequently in time and estimating the spectrum over an expanded range of frequencies the effects of averaging and aliasing may be minimized. This comes about because the expanded scale spectrum, say from 0 to $3f_n$ or higher in terms of the original frequency scale, contains the portion 0 to f_n on the low frequency end. Here the effects of averaging are very small (see Fig. 2) and the aliased variance will also be very small because it is effectively multiplied by the averaging function before it is 'folded' into this part of the spectrum (see Equation II-1). Thus, these effects may be minimized by essentially using observations of a continuous record taken at, and averaged over varying time intervals and the spectrum estimated over varying frequency bands.

It is repeated here for emphasis that if f(t) is considered to have no existence between observations and therefore time is discrete, none of the foregoing considerations apply and the spectra stand as estimated.

APPENDIX III

Examination of Variance Spectra from Variance of Means Over N Years

by

Paul Julian and Max Woodbury

From the text it will be recalled that Leopold's statistic for measuring persistence was the ratio

$$\frac{s_N^2}{s^2}\text{, where } s_N^2$$

is the variance of means over N years. Although Leopold calculated these from contiguous blocks of data of length N, we will here consider these to be calculated from running means of N years.

It can be shown that the relation between s_N^2 and the spectrum is

$$(1) \qquad s_N^2 = \int_{-\pi}^{\pi} A_N(\omega)\, \Phi(\omega)\, d\omega$$

where A_N is the effect of the averaging process upon the distribution of variance. For running means of length N,

$$A_N(\omega) = \left[\frac{\sin \frac{N\omega}{2}}{N \sin \frac{\omega}{2}}\right]^2 .$$

As suggested in the text, the problem to be considered here is the inverse of equation (1), namely, the estimation of the spectrum from the

$$s_N^2$$

values in hand.

To begin, we will assume

$$(2) \qquad \Phi(\omega) = \sum_k \alpha_k A_k(\omega)$$

where a dummy index k (and later h) is substituted for N. Here the assumption is that the spectrum can be represented as a weighted sum of the power admittance functions for each k.

Substituting in (1)

$$S_h^2 = \sum_k \alpha_k \int_{-\pi}^{\pi} A_h A_x d\omega = \sum_k \underline{M}_{hk} \alpha_k$$

where $\underline{M}_{hk}$ is defined as the

$$\text{Matrix} = \int_{-\pi}^{\pi} A_h A_x d\omega.$$

Solving the system of equations by inverting the matrix $\underline{M}^{-1} = \underline{M}^{kh}$

$$\alpha_k = \sum_h M^{kh} S_h^2.$$

Therefore, by using (2) an estimate of $\Phi(\omega)$ can be obtained, since

$$\Phi(\omega) = \sum_k \alpha_k A_k$$

$$\hat{\Phi}(\omega) = \sum_h \sum_k A_k M^{kh} S_h^2$$

where ^ sign denotes an estimate. Letting

$$\hat{\Phi}(\omega) = \sum_h \sum_k A_k M^{kh} S_h^2$$

reduce to

$$\hat{\Phi}(\omega) = \sum_h F_h(\omega) S_h^2$$

(4) $$\hat{\Phi}(\omega) = S_1^2 F_1(\omega) + S_5^2 F_5(\omega) + S_{10}^2 F_{10}(\omega) + S_{15}^2 F_{15}(\omega) + S_{20}^2 F_{20}(\omega)$$

where the S_h^2 are the Leopold values of the variances of means of length h.

The problem now is to obtain values of

$$F_h(\omega) = \sum_k A_k M^{kh}.$$

To evaluate M_{hk}, it may first be recalled from work on time-series analysis (for example Bartlett (5), or Hannan (12)) that

$$X_t = \int_{-\pi}^{\pi} e^{i\omega t} dZ(\omega),$$

where $dZ(\omega)$ is the spectral process of the series X_t. Using (1) above

$$s_k^2 = \frac{1}{k^2} \sum_{-\frac{(k-1)}{2}}^{\frac{k-1}{2}} \sum_{-\frac{(k-1)}{2}}^{\frac{k-1}{2}} \int_{-\pi}^{\pi} e^{i\omega(r-s)} \Phi(\omega)\, d\omega$$

and

$$A_k(\omega) = \frac{1}{k^2} \sum_r \sum_s e^{i\omega(r-s)}$$

$$A_k(\omega) = \frac{1}{k^2} \sum_{-k}^{k} \left[k - |p|\right] e^{i\omega p}$$

since for $p = r - s \neq 0$, $\int_{-\pi}^{\pi} e^{i\omega p} d\omega = 0.$

Hence, $$\int_{-\pi}^{\pi} A_k(\omega) A_h(\omega)\, d\omega = \frac{1}{k^2 h^2} \sum_{-(k\wedge h)}^{(k\wedge h)} \left[k - |p|\right] \left[h - |p|\right]$$

where $(k\wedge h) = \min(k,h)$. With further manipulation the matrix

$$\underline{M}_{kh} = \int_{-\pi}^{\pi} A_h A_k\, d\omega$$

reduces to

$$\underline{M}_{kh} = \frac{h}{3h^2k^2} (1+3kh - h^2).$$

The matrix actually inverted was

$$kh\underline{M}_{kh} = h - \frac{h^2 - 1}{3kh}$$

The inverse of this matrix times the A_k function which was given under (1) above produces the $F_h(\omega)$ values. Note that it is possible to calculate the F_h values for any ω, that is, any frequency. Leopold's values for N = 1, 5, 10, 15, 20 were supplemented by the values N = 1, 2, 5, 10, 20 to get a better distribution on the linear frequency scale, and values for ω's comparable to those resulting from the Tukey approach calculated. This was done in order to have estimated points in the two instances at the same frequencies.

$\omega°$	F_1	F_5	F_{10}	F_{15}	F_{20}
0	-.0207	.0173	.0017	.0006	20.3398
30	.4921	5.2117	-5.6115	.2062	.2601
60	1.2184	-1.2376	.3377	.2770	-.4629
90	1.2173	-1.1255	.3347	-.1993	-.2471
120	1.2166	-1.0139	.0457	-.4522	.2345
150	1.2686	-1.4010	.0707	.0377	.0442
180	1.2160	-.9022	-.5227	.4360	-.2470

$\omega°$	F_1	F_2	F_5	F_{10}	F_{20}
0	-.0280	.0195	.0054	.0016	20.3400
30	-.1289	1.6435	4.2168	-5.5566	.3878
60	.2270	2.6241	-2.8260	.4078	-.2936
90	.9383	.7385	-1.5726	.2610	-.3832
120	1.6650	-1.1867	-.2956	-.0962	-.0587
150	2.1712	-2.3898	.0456	.1047	.0822
180	2.3623	-3.0347	.9348	-.3487	.0584

Multiplying these values by the corresponding S_N^2 values as indicated in (4) will give the spectrum estimates.

S_N^2

Leopold

N	(1)	(5)	(10)	(15)	(20)
S_N^2	17.64	7.29	4.00	3.24	1.96
NORM	1.000	0.413	0.227	0.184	0.111
S_N^2 (From running means)	17.46	5.49	3.87	3.11	2.71
NORM	1.000	0.314	0.222	0.178	0.155

Julian and Woodbury	(1)	(2)	(5)	(10)	(20)
	17.46	10.85	5.49	3.87	2.71
NORM	1.000	0.621	0.314	0.222	0.155

Spectra $\Phi(\omega)$	(ω)	$\Phi_1(\omega)$	$\Phi_2(\omega)$	$\Phi_3(\omega)$
Φ_1 Leopold (original)	0°	.303	.374	.376
	30°	.193	.114	.126
Φ_2 Leopold (running means)	60°	.105	.104	.121
	90°	.102	.102	.108
	120°	.100	.102	.096
Φ_3 Woodbury and Julian	150°	.096	.102	.089
	180°	.104	.101	.084

1. ____________, Bulletin W, and Supplement to Bulletin W, Published by the U. S. Weather Bureau, Washington, D. C.

2. ____________, U. S. Geological Survey Water Supply Papers, U. S. Geological Survey, Washington, D. C.

3. ABBOT, C. G., 'Long Range Weather Forecasting,' Smithsonian Miscellaneous Collections, 38 (1959).

4. ARAI, Y., 'Characteristics of Long Waves in the Westerlies Related to Solar Activity,' Journal of the Meteorological Society of Japan, 36 (1958) 46ff.

5. BARTLETT, M. S., 'Periodogram Analysis and Continuous Spectra,' Biometrika, 37 (1950) 1 - 16.

6. BLACKMAN, R. B. and TUKEY, J. W., The Measurement of Power Spectra, Dover, New York (1958) 190 pp.

7. BRIER, G., Unpublished manuscript of the U. S. Weather Bureau, 1961.

8. CARRUTHERS, N., 'A simple periodoscope for meteorological data,' Quarterly Journal of the Royal Meteorological Society, London, 70 (1944) 171ff.

9. CRADDOCK, J. M., 'An analysis of the slower temperature variations at Kew Observatory by means of mutually exclusive band-pass filters,' Journal of the Royal Statistical Society (series A), 120 (1957).

10. FIERING, M., Thesis, Department of Applied Engineering and Physics, Harvard University, 1960.

11. FRIEDMAN, D. G., The Prediction of Long-Continuing Drought in South and Southwest Texas, The Travelers Weather Research Center, Hartford, Conn., 1957.

12. HANNAN, E. J., Time Series Analysis, John Wiley and Sons, New York (1960) 152 pp.

13. Harris, B., et al, 'An improved statistical model for evaluating parameters affecting water yields of river basins,' paper presented at the April 1961 meetings of the American Geophysical Union; abstract to be published in forthcoming issue of the Journal of the American Geophysical Union.

14. JULIAN, P., 'A synthetic hydrology for the Colorado River,' Papers of the Western Resources Conference, University of Colorado Press, (1960) 168ff.

15. KOHLER, M. A. and LINSLEY, R. K., 'Recent developments in water supply forecasting from precipitation,' Transactions of the American Geophysical Union, 30 (1949) 434ff.

16. LABROUSTE, H. and LABROUSTE, Y., 'Harmonic analysis by means of linear combinations of ordinates,' Journal of Terrestrial Magnetism, 41 (1936) 15ff.

17. LANDSBERG, H., 'Note on the recent climatic fluctuation in the U.S.,' Journal of Geophysical Research, 65 (1960) 151ff.

18. LANDSBERG, H., et al, 'Power spectrum analysis of climatological data for Woodstock College, Maryland,' Monthly Weather Review, 87 (1959) 283ff.

19. LANGBEIN, W. B., 'Queuing theory and water storage,' Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers, Paper 181 (1958).

20. LANGBEIN, W. B., Annual Runoff in the United States, Geological Survey Circular No. 52, U. S. Geological Survey, Washington, D. C.

21. LEE, Y. W., Statistical Theory of Communications, Wiley and Sons, New York (1960) 509 pp.

22. LEOPOLD, L. B., Probability Analysis Applied to a Water-Supply Problem, Geological Survey Circular No. 410, U. S. Geological Survey, Washington, D. C. (1959).

23. LeSEUR, N. E., 'On the asymmetry of the middle-latitude circumpolar current,' Journal of Meteorology, 11 (1954) 43ff.

24. LOMINICKI, Z. A. and ZAREMBA, S. K., 'Spectral approach to time series,' Journal Royal Statistical Society (B), 19 (1957).

25. McDONALD, J. E., 'A note on the precision of estimation of missing precipitation data,' Transactions of the American Geophysical Union, 38 (1957) 657ff.

26. McDONALD, J. E., 'Variability factors in mountain-watershed hydrometeorology in an arid region,' Journal of the Arizona Academy of Science, 1 (1960) 89ff.

27. MITCHELL, J. M. Jr., 'On the causes of instrumentally observed secular temperature trends,' Journal of Meteorology, 10 (1953) 244ff.

28. MORAN, P. A. P., The Theory of Storage, Wiley and Sons, New York (1959) 110 pp.

29. MORONEY, M. J., Facts from Figures, Penguin Books, London (1951) 472 pp.

30. OLTMAN, R. E. and TRACY, H. J., Trends in Climate and in Precipitation Runoff Relation in the Missouri River Basin, Geological Survey Circular No. 98, U. S. Geological Survey, Washington, D. C. (1951).

31. PAULHUS, J. L. and KOHLER, M. A., 'Interpretation of missing precipitation records,' Monthly Weather Review, 80 (1952) 129ff.

32. PECK, E. L., Low Winter Streamflow as an Index to the Short and Long-Term Carry-Over Effects in Water Supply Forecasting, MS copy, Weather Bureau Water Supply Forecast Unit, Salt Lake City, Utah (1954).

33. REX, D., 'Blocking action in the middle troposphere and its effect upon regional climate,' Tellus, 2 (1950) 196ff, 275ff.

34. SCHUSTER, Sir A., 'The investigation of hidden periodicities,' Journal of Terrestrial Magnetism, 3 (1898) 13ff.

35. SIEGEL, S., Non-Parametric Methods for the Behavioral Sciences, McGraw-Hill, New York (1956).

36. SNYDER, W. M., 'Some possibilities for multivariate analysis in hydrologic studies,' paper presented at the April 1961 meetings of the American Geophysical Union, abstract to be published in forthcoming issue of the Journal of the American Geophysical Union.

37. TANNEHILL, I. R., Drought, Its Causes and Effects, Princeton University Press, (1947).

38. THOM, H. S., Report to the President's Advisory Committee on Weather Control, Washington, D. C. (1957).

39. TUKEY, J. W., 'The sampling theory of power spectrum estimates,' in Symposium on Applications of Autocorrelation Analysis to Physical Problems, Woods Hole, NAVEXOS-P-735, ONR, Washington, D. C. (1949).

40. WALD, A. and WOLFOWITZ, J., 'An exact test for randomness in the non-parametric case based on serial correlation,' Annals of Mathematical Statistics, 14 (1943) 378ff.

41. WEINER, N., The Fourier Integral and Certain of its Applications, Dover, New York (1933).

42. WILLET, H. C., 'Extrapolation of sunspot-climate relationships,' Journal of Meteorology, 8 (1951) 1ff.

43. WOLD, H., A Study in Analysis of Stationary Time-Series, Almquist and Wiksell, Stockholm (1954).