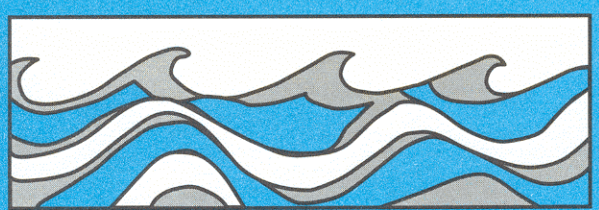
University of Washington
Department of Civil and Environmental Engineering

SERIAL CORRELATION IN ANNUAL STREAM RUNOFF

Kenneth W. Martig, Jr.
Thomas H. Campbell



Water Resources Series
Technical Report No. 24
June 1968

Seattle, Washington
98195

Department of Civil Engineering
University of Washington
Seattle, Washington 98195

SERIAL CORRELATION IN ANNUAL STREAM RUNOFF

Kenneth W. Martig, Jr.
Thomas H. Campbell

Water Resources Series
Technical Report No. 24

June 1968

SERIAL CORRELATION IN ANNUAL STREAM RUNOFF

A Progress Report of Project Number
161-34-10E-3992-3010
of the Office of Water Resources Research
Under Annual Allotment Agreement
Number 14-01-0001-1416
July 1, 1967 to June 30, 1968

University of Washington
Seattle, Washington, 98105

Kenneth W. Martig, Jr. and Thomas H. Campbell

June 1968

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Y_{t+1}	streamflow for the time t+1
Y_t	streamflow for the time t
b	regression coefficient {correlation coefficient}
$\overline{X}$	average of streamflow record sequence
S_{Y_{t+1}/Y_t}	standard error of estimate, Y_{t+1} on Y_t
R	random variable, normally distributed with zero mean and unit standard deviation
a	regression coefficient {Y_{t+1} intercept}
$\pi_1, \pi_2, \pi_3, \pi_4$	moments of frequency distributions
N	length of record {sequence}
D.F.	degrees of freedom
V	residual number being checked for randomness
σ	standard deviation
β_1	skewness coefficient
β_2	kurtosis coefficient
δ	difference between set average Y_{t+1} and the expected value which lies on the regression line

h	regression coefficient {skewness intercept}
i	regression coefficient {skewness on mid-band slope}
h'	regression coefficient {kurtosis intercept}
i'	regression coefficient {kurtosis on mid-band slope}
c_m	regression coefficient for multiple degree Markov chains

ACKNOWLEDGMENTS

The work upon which this report is based was supported by funds provided by the United States Department of the Interior, Office of Water Resources Research, as authorized under the Water Resources Research Act of 1964.

The computer facilities at the University of Washington were used in this study and the coefficients of the multiple Markov chain were computed according to BMD 03R, prepared by the Health Sciences Computing Facility, University of California at Los Angeles. All other programs used in this study were prepared by Mr. Kenneth Martig.

The authors wish to acknowledge the advice of Mr. Archie A. Harms in the pursuit of these studies, and of Dr. Eugene P. Richey and Dr. Joseph C. Kent in the preparation of this report.

ABSTRACT

Serial correlation in annual stream runoff is examined in the light of its application in simulation. For more than thirty streams with relatively long streamflow records it is found that the Thomas-Fiering linear algorithm can not be improved on with statistical significance by any of three similar non-linear expressions. It is also determined that historic deviations from this linear relation for each of twenty-four streams can not reliably be said to be non-normal in distribution. The use of a multiple degree Markov chain of serial correlation is found to give significant improvement over the one degree relation used in the Thomas-Fiering model. Evidence of possible cyclicity is found.

Key words: hydrology, streamflow, runoff, simulation, synthesis.

CHAPTER I
INTRODUCTION

Today's demands, created by an increasing population density and an advancing degree of technology, have brought numerous problems concerning both the quality and the quantity of water in our rivers and streams. Dilution of pollution; firm flow for hydroelectric power generation; determination of water storage capacity for agricultural, domestic and industrial uses; and proper water allocation are just a few of these problems. Objective engineering solutions to these problems require extensive knowledge of annual streamflow which frequently is not available in our historic records. As a result, several methods of generating adequate streamflow records were developed.

The earliest mathematical model of streamflow synthesis was developed by Hazen[1] in 1914. Later came contributions by Sudler[2] in 1927, Yule[3] in 1927, Barnes[4] in 1955, Brittan[5] in 1960, Julian[6] in 1961, Thomas and Fiering[7] in 1962, Maughan and Kawano[8] in 1963, Yagil[9] in 1963, and Yevdjevich[10] in 1964.

This report will concentrate on the annual streamflow synthesis of serially correlated data represented by Eq. 1.1

$$Y_{t+1} = \bar{X} + b(Y_t - \bar{X}) + S_{Y_{t+1}/Y_t} R \qquad (1.1)$$

where Y_{t+1} = non-historic streamflow for time t+1

Y_t = non-historic streamflow for time t

b = regression coefficient (correlation coefficient)

$\bar{X}$ = average of historic streamflow sequences

S_{Y_{t+1}/Y_t} = standard error of estimate, Y_{t+1} on Y_t

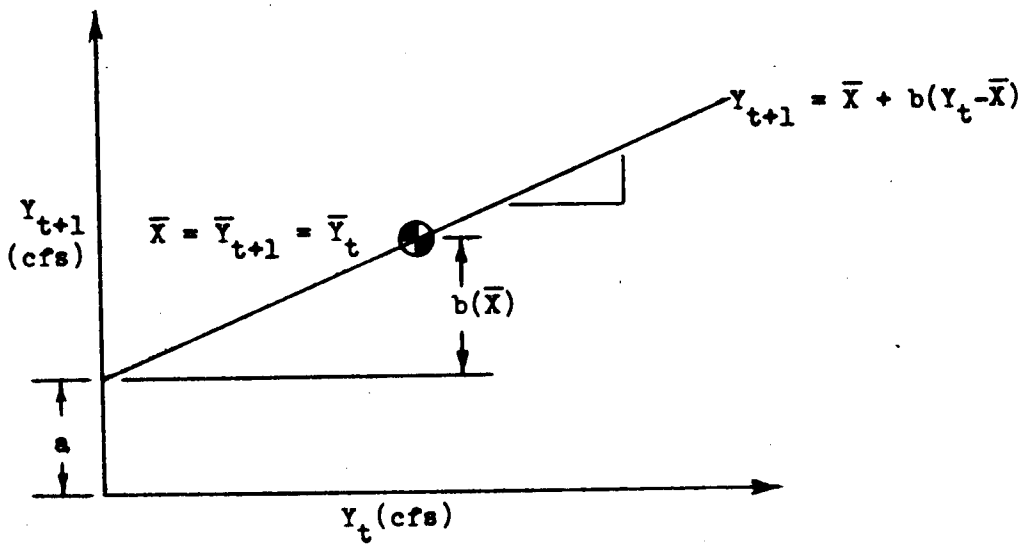
R = random variable, normally distributed with zero mean and unit standard deviation.

which was extensively examined by Thomas and Fiering, and referred to by Brittan, Julian and Yagil.

If we drop the unexplained error estimator term $\{(S_{Y_{t+1}/Y_t})R\}$ from Eq. 1.1 we get

$$Y_{t+1} = \bar{X} + b(Y_t - \bar{X}) \qquad (1.1\text{-a})$$

Figure 1.1: Graph of $Y_{t+1} = \bar{X} + b(Y_t - \bar{X})$



Looking at Figure 1.1 we see that

$$\bar{X} - b(\bar{X}) = a$$

Rearranging Eq. 1.1-a we get

$$Y_{t+1} = (\bar{X} - b\bar{X}) + bY_t \qquad (1.1\text{-b})$$

Substituting a for $\bar{X} - b(\bar{X})$ in Eq. 1.1-b we get

$$Y_{t+1} = a + bY_t \qquad (1.2)$$

which is the mathematical expression for the regression line through the plot of Y_{t+1} versus Y_t. Equation 1.2 can be mathematically classified as a one degree Markov chain.

> "A Markov chain may be defined as a stochastic process whose development may be treated as a series of transitions between certain values (called the "states" of the process) which have the property that the probability law of the future development of the process, once it is in a given state, depends only on the state and not on how the process arrived in that state, i.e., given the present, the future is independent of the past."[11]

Note that in the above reference stochastic process is defined as "a family of random variables [X(t), tϵT] where for each t in a set T, the observation X(t) is an observed value of a random variable."[11]

The algorithm {Eq. 1.1} presupposes that the historic data when arranged sequentially is normally distributed about the regression line {Eq. 1.2} and implies that normally distributed random numbers can be used to reproduce the unexplained variations.

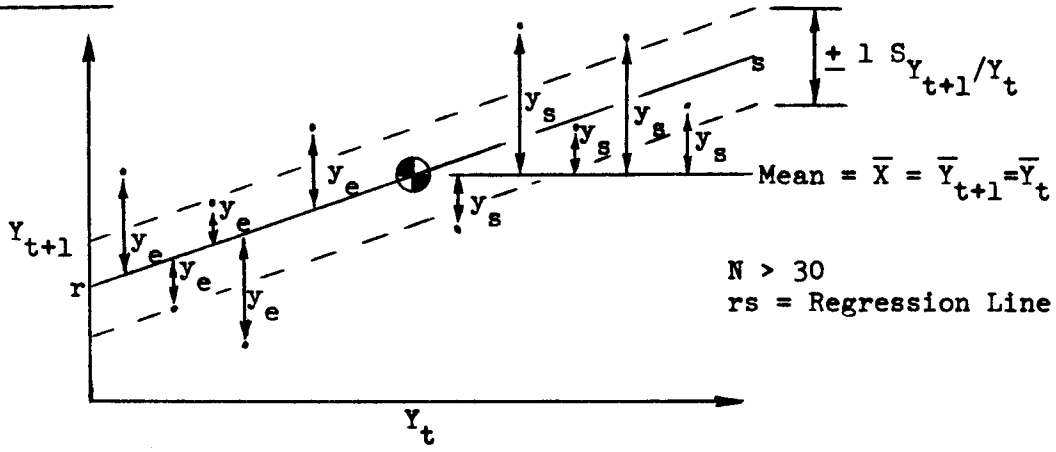
Because low flows cannot become negative, but instead level off to some minimum asymptote and because we can intuitively imagine an upper limit to high flows for a given watershed, the sequentially arranged historic data might not be normally distributed about the regression line described by Eq. 1.2 except in the vicinity of the average flow position.

Chapter II examines sequentially arranged historic streamflow data with Eq. 1.2 as the regression line to see if, in fact, the values of Y_{t+1} on Y_t are normally distributed about the entire length of the regression line.

Review of Statistics[13]

Before we start the examination of sequentially arranged historic data, let us review the parameters used as indicators of statistical properties.

Figure 1.2: Regression Statistics Illustration



Relationships Derived From Illustrations of Figure 1.2

a) The four moments

$$\Pi_1 = \text{Moment \#1} = \frac{\sum (y_s)^1}{\text{N-D.F.}} = 0.0 \text{ when least squares regression analysis is employed.}$$

$$\Pi_2 = \text{Moment \#2} = \frac{\sum (y_s)^2}{\text{N-D.F.}}$$

$$\Pi_3 = \text{Moment \#3} = \frac{\sum (y_s)^3}{\text{N-D.F.}}$$

$$\Pi_4 = \text{Moment \#4} = \frac{\sum (y_s)^4}{\text{N-D.F.}}$$

where N = length of record (years)

D.F. = degrees of freedom

b) The standard deviation

$\sigma = \sqrt{\text{Moment \#2}}$

$\sigma = 1.0$ for normal random numbers

c) The skewness coefficient

$$\beta_1 = \frac{(\text{Moment \#3})^2}{(\text{Moment \#2})^3}$$

$\beta_1 = 0.0$ for normal distributions

or Moment #3 = "-" for negatively skewed distributions
Moment #3 = "+" for positively skewed distributions

d) The kurtosis coefficient

$$\beta_2 = \frac{(\text{Moment \#4})_2}{(\text{Moment \#2})}$$

$\beta_2 = 3.0$ for normal distributions

$\beta_2 < 3.0$ for flat distributions

$\beta_2 > 3.0$ for peaked distributions

3) The standard error of estimate

$$S_{Y_{t+1}/Y_t} = \left(\frac{\sum (y_e)^2}{\text{N-D.F.}} \right)^{\frac{1}{2}}$$

S_{Y_{t+1}/Y_t} is small when the regression equation found by least squares techniques represents the data closely. The larger the S_{Y_{t+1}/Y_t} value the poorer the regression representation.

CHAPTER II

EXAMINATION OF EQUATION 1.2 AS BEST REGRESSION FIT THROUGH SEQUENTIALLY ARRANGED DATA

If the historic data points represented by the regression line {Eq. 1.2} are normally distributed about the line at all points along the line, Eq. 1.1 cannot be improved and is indeed the best possible algorithm available to generate stochastic annual streamflow.

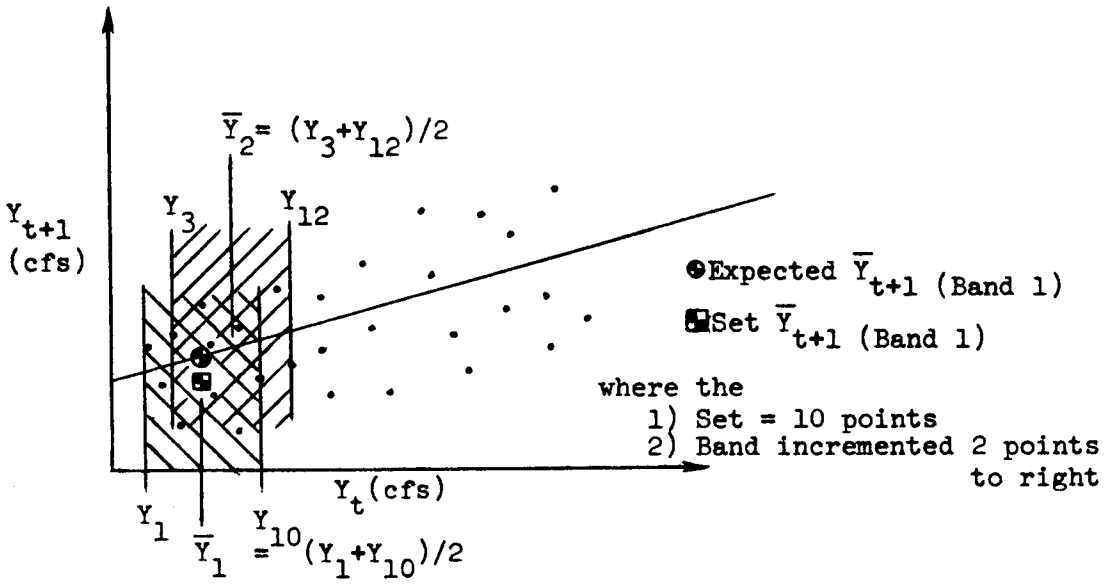
To test this, an examination of Y_{t+1} on Y_t is made placing Eq. 1.2 through this autocorrelated data using the least squares regression technique. It is now possible to examine the scatter of points about the regression line. A moving band procedure is employed for this examination.

2.1 Moving Band Procedure

A moving band is simply a moving set which contains a constant given number of points out of the total population of points. The total population is equal to the historic record length minus one {N-1}. {For a one degree Markov chain, as is Eq. 1, one year of record is lost at the end of the series because "t+1" cannot exceed N, hence, the maximum "t" value is given by "N-1"}. This moving band starts at the left of the plot and moves to the right. Its initial position at the extreme left takes the first set of specified size and examines it for σ, Π_1, Π_2, Π_3, Π_4, β_1, β_2 and a deviation $\{\delta\}$. The preceding symbols are commonly recognized as representing the standard deviation, the four moments, the skewness coefficient, and the kurtosis coefficient respectively. The term deviation $\{\delta\}$ represents the difference between $\bar{Y}_{t+1}$, derived using

the points within the set, and the expected $\overline{Y}_{t+1}$ value which lies on the regression line where a vertical line through the middle of the set intersects the regression line. {See Figure 2.1}

Figure 2.1: Moving Set Boundaries and Middle Point



The band is then moved to the right a specified number of points and the same calculations are made for the set in its new location. The amount that the band moves {the number of points passing out of the set through the left boundary as the set moves along the Y_t axis} should be less than the number of points within the set so that some overlap occurs between successive positions of the band. This procedure continues until the band is positioned such that the right boundary of the set coincides with the largest Y_t value.

Because all calculated terms for a given band are representative of a position at the middle of the band as shown in Figure 2.1 it is now possible to tabulate the various calculated parameters with their respective mid-band values and examine the tabulation for meaningful trends.

The above procedure was employed using band widths or sets containing from five to forty-five points. The purpose of varying the set size was to substantiate the fact that for statistical calculations a minimum set size of thirty points is required before stable results are obtainable.[15] Stable results were obtained when the set size was thirty or greater.

Because a minimum of thirty points are required for a set to be representative, only streams of long historic virgin annual streamflow record can be examined in the manner described above. There are few streams on record with the length and quality of record which satisfy this requirement. For example, a stream with N = 60 being examined as described above with a band containing thirty-five points will give only six checks on the statistics of the scatter of points about the regression line through the plot of Y_{t+1} versus Y_t when the band is incremented by four points to the right.

2.2 Examination of Columbia River Data

There is for the Columbia River at The Dalles, Oregon a record of eighty-five years of historic annual streamflow which has been adjusted back to represent virgin conditions as well as possible. These data were analyzed by the moving band technique. The band contained a set of

thirty-five points out of a population of eighty-four {N-1} points, and was moved an increment of four points each time until the entire scatter about the regression line {Eq. 1.2} was examined.

It was found that the standard deviations calculated for each band position were very nearly equal, which means that each set of thirty-five points was in fact representative of the entire population of eighty-four points.

The deviation was quite small showing that the expected $\overline{Y}_{t+1}$ was very nearly obtained. However, there was an indication that at low flows {flows below average, $\overline{X}$} the average in each set was slightly higher than that expected and at high flows {flows above average $\overline{X}$} the average in each set was slightly lower than expected. This might suggest a tendency for the linear regression being examined to try to become "S" shaped.

The skewness showed an interesting trend also. It was non-zero and was negative at low flows and positive at higher flows. {β_1 = 0.0 for normal distributions}. This too might show a tendency for the regression line to become "S" shaped.

Kurtosis values showed an increasing trend from $\beta_2 < 3.0$ at low flows to $\beta_2 = 3.0$ at mean flow to $\beta_2 > 3.0$ at high flows with a "β_2" value equal to 3.0 for normal distributions.

These statistical parameters are summarized in Table 2.1.

Table 2.1: Summary of the Statistical Characteristics of Scatter about the Linear Regression Line {Eq. 1.2}.

Mid-Band Value	Deviation	σ	β_1	β_2
.803	.01	.166	-.072	2.02
.839	-.009	.169	-.014	1.86
.871	.003	.164	-.031	1.94
.916	-.008	.162	-.180	2.32
.947	-.016	.171	-.055	2.29
.985	.010	.162	-.329	2.80
1.010	.001	.154	-.202	2.99
1.030	.001	.174	.218	4.72
1.055	.008	.178	.095	4.24
1.080	.009	.175	.070	4.38
1.101	.024	.172	.047	4.49
1.119	.009	.179	.027	4.14
1.178	.011	.163	.135	4.58
1.305	-.017	.161	.132	4.74

Note: The above values were made dimensionless by dividing both Y_{t+1} and Y_t values by $\overline{X}$.

The trends visible in Table 2.1 suggest a non-normal distribution of the scatter of points about the linear regression designated by a one degree Markov chain {Eq. 1.2}.

It is possible that a regression line that curves as shown in Figure 2.2 might better represent the values of Y_{t+1} on Y_t and produce a scatter closer to normal because:

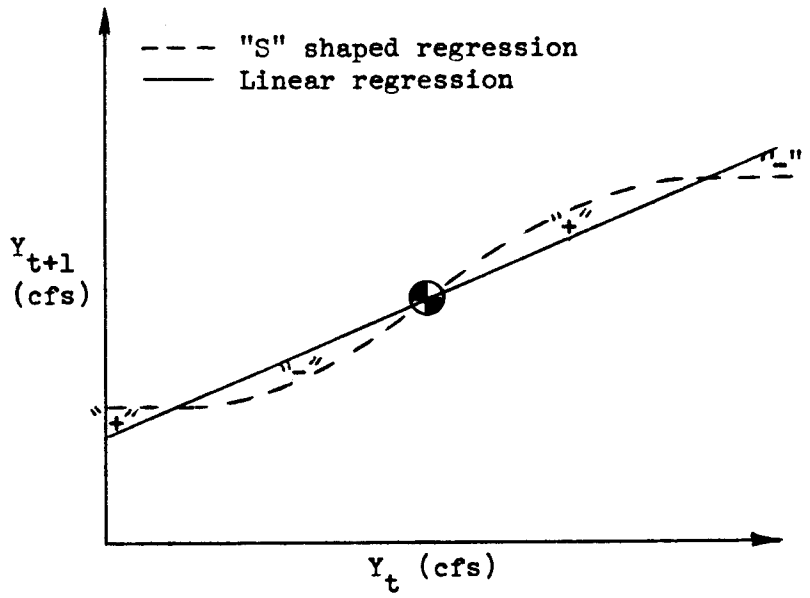
1) Intuitive limits on low flows and high flows for a given basin tend to raise averages in sets at low flows above that expected by linear regression analysis and lower averages in sets at high flows below that expected by linear regression analysis,

2) The Columbia River record which is among the best available showed that the

   a) Skewness is negative at low flows and positive at high flows,

   b) Kurtosis increases from $\beta_2 < 3.0$ at low flows to $\beta_2 > 3.0$ at high flows.

Figure 2.2: Proposed "S" Shaped Regression



Chapter III examines this possibility and in particular examines several mathematical expressions that represent "S" shaped curves.

CHAPTER III

EXAMINATION OF NON-LINEAR REGRESSION EQUATIONS AS POSSIBLE BETTER FITS

Two equations that develop "S" shapes when their regression constants are calculated using the least squares regression technique were tried with thirty-one different autocorrelations of Y_{t+1} on Y_t developed from thirty-one sequences of historically recorded annual streamflow of various record lengths from rivers in the Pacific Northwest Region {See Appendix A}. The values for the standard error of estimate $\{S_{Y_{t+1}/Y_t}\}$ were calculated for each of the proposed equations and compared with the S_{Y_{t+1}/Y_t} value obtained when Eq. 1.2 was used. The two equations which were examined in this manner are: (1) a three degree polynomial and (2) the sinh function converted to a natural log function.

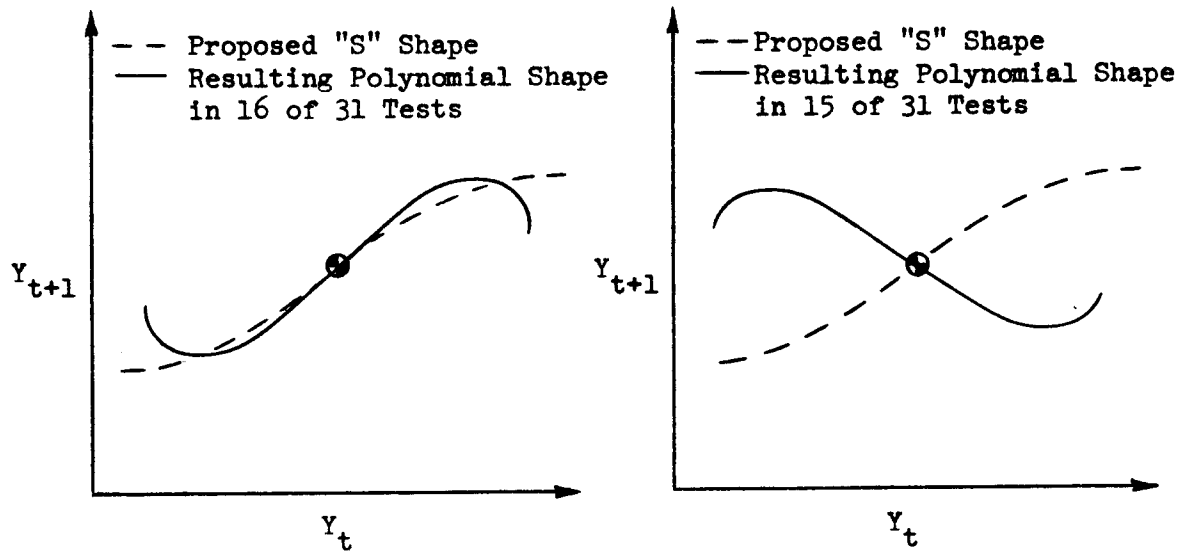
3.1 The Three Degree Polynomial

The three degree polynomial is written mathematically as

$$Y_{t+1} = C_1 + C_2(Y_t) + C_3(Y_t)^2 + C_4(Y_t)^3 \qquad 3.1$$

Instead of developing the desired "S" shaped portion, the proposed polynomial for the most part tended to become fully developed with sixteen of the thirty-one examinations possessing a positive internal limb slope and fifteen of thirty-one possessing a negative internal limb slope as shown in Figure 3.1.

Figure 3.1: Results of Polynomial Test



a) Positive Internal Limb Slope b) Negative Internal Limb Slope

However, a few did approach the desired shape and all polynomial {Eq. 3.1} tests showed standard error of estimate terms very close to those values obtained using Eq. 1.2. Six of the thirty-one tests actually developed smaller standard error of estimate terms and hence represented better fits to the plot of Y_{t+1} vs. Y_t. Table 3.1 compares the standard error of estimate values obtained using the one degree Markov chain {Eq. 1.2} and the polynomial {Eq. 3.1} as regression equations.

Because only 20% of the tests showed improvement over Eq. 1.2 with the maximum improvement being 4%, it can be concluded that the polynomial proposal is not a good one.

Table 3.1: Standard Error of Estimate Comparisons**

Sequence Number*	One Degree Markov chain (Eq. 1.2)	Three Degree Polynomial (Eq. 3.1)	Sinh Function (Eq. 3.3)	Natural Log vs. Natural Log function (Eq. 3.4)
1	.3188	.3078	.3188	.3290
2	.1587	.1619	.1587	.1634
3	.1549	.1574	.1549	.1538
4	.1702	.1709	.1701	.1774
5	.1737	.1776	.1737	.1836
6	.2086	.2162	.2087	.2206
7	.2149	.2233	.2149	.2353
8	.2420	.2484	.2419	.2693
9	.1414	.1426	.1414	.1450
10	.1656	.1584	.1657	.1661
11	.1517	.1551	.1517	.1532
12	.2010	.1959	.2013	.2038
13	.1538	.1570	.1538	.1558
14	.1664	.1676	.1664	.1737
15	.1813	.1814	.1813	.1841
16	.1687	.1715	.1687	.1699
17	.1814	.1830	.1815	.1830
18	.1979	.1978	.1980	.2105
19	.1814	.1757	.1814	.1866
20	.1959	.2009	.1960	.2010
21	.2021	.2041	.2022	.2121
22	.1869	.1906	.1869	.1914
23	.1905	.1903	.1906	.1894
24	.1712	.1766	.1712	.1774
25	.1853	.1896	.1855	.1884
26	.1764	.1794	.1765	.1787
27	.1798	.1843	.1799	.1786
28	.1277	.1296	.1278	.1241
29	.2257	.2280	.2258	.2372
30	.1740	.1754	.1739	.1747
31	.1695	.1722	.1695	.1782

**The above standard error of estimate terms are dimensionless values obtained by dividing the S_{Y_{t+1}/Y_t} term by the mean $\{\bar{X}\}$

*See Appendix A for streams represented by sequence numbers 1 thru 31.

3.2 The Sinh Function

The sinh function is written mathematically as

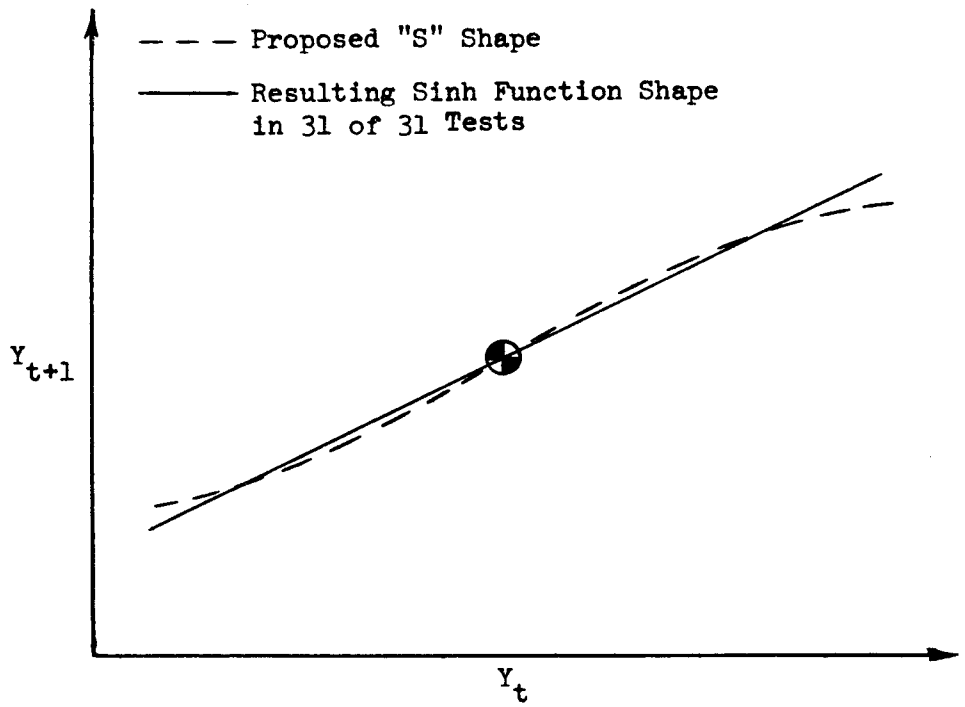
$$Y_{t+1} = C_1 \ (\sinh^{-1}(Y_t)) \qquad 3.2$$

It can be transformed into a natural log function written mathematically as

$$Y_{t+1} = C_1 \ \ln(Y_t + \sqrt{(Y_t)^2 + 1}) \qquad 3.3$$

This form of regression when tested, in all cases produced essentially a straight line within plotting accuracy instead of the proposed "S" shaped curve as shown in Figure 3.2.

Figure 3.2: Results of Sinh Function Test



Therefore, the limited historic data present in a plot of Y_{t+1} versus Y_t represents only the interior limb of the "S" shaped curve.

However, again the standard error of estimate terms calculated for the thirty-one tests using Eq. 3.3 are all nearly equal to those values obtained when Eq. 1.2 was used. In fact, eighteen of the thirty-one resulting S_{Y_{t+1}/Y_t} values are equal to or less than 1% better than the one degree Markov chain values of S_{Y_{t+1}/Y_t}. See Table 3.1 for a comparison between the values of the standard error of estimate determined using Eq. 1.1 and Eq. 3.3.

Although the sinh equation could be used for the Columbia River and other Pacific Northwest streams, no significant improvement or simplification over the linear model {Eq. 1.2} has been made by either of the proposed non-linear models, Eq. 3.1 or Eq. 3.3.

One more test was made to see if a non-linear model might better represent the plot of Y_{t+1} versus Y_t.

It has been widely accepted by hydrologists that a plot of the natural log of flow versus the probability of occurrence will sometimes approximate a straight line more closely than will a plot of arithmetic flow versus probability. Therefore, as a final check let us try taking the natural log of both Y_{t+1} and Y_t in Eq. 1.2 to see if the resulting equation would better represent the values of Y_{t+1} on Y_t. If a non-linear fit is characteristic of the values of Y_{t+1} on Y_t, an equation of this type would tend to linearize it.

3.3 The Natural Log of Y_{t+1} and Y_t

If the natural log is taken of the flow terms $\{Y_{t+1}$ and $Y_t\}$ in Eq. 1.2 it then becomes

$$\ln (Y_{t+1}) = C_1 + C_2 \ln(Y_t) \qquad 3.4$$

When the curve of Eq. 3.4 was fitted to the values of Y_{t+1} on Y_t using the least squares regression technique a straight line resulted which had very nearly the same slope and standard error of estimate values as did Eq. 1.2. Four of these thirty-one tests showed an improvement over the arithmetic one degree Markov chain, although the improvement was only as high as 3% and therefore insignificant. See Table 3.1 for a comparison of the standard error of estimate values determined using Eq. 1.2 and Eq. 3.4.

In summary, this investigation has found no grounds for justifying a non-linear regression equation to represent the values of Y_{t+1} on Y_t. This means that a linear model {Eq. 1.1} is the best mathematical representation of the sequential values of annual streamflow $\{Y_{t+1}$ on $Y_t\}$ and suggests that the reason for the non-normal scatter of points about the regression line represented by Eq. 1.2 as shown by Table 2.1 might lie in the unexplained variance term $\{(S_{Y_{t+1}/Y_t})R\}$ of Eq. 1.1. If this is indeed the case, then the variable {R} might not be random as assumed.

Chapter IV examines the actual historic variations required to reproduce the exact scatter of points about the regression line {Eq. 1.2} to see if they do indeed satisfy the required properties of a random number.

CHAPTER IV

EXAMINATION FOR THE PRESENCE OF RANDOM VARIATIONS

A random number {R} was previously introduced as a random variable, normally distributed with a mean of zero and a standard deviation {σ} equal to one. An hypothesis is presumed that these numbers, when multiplied by the standard error of estimate, simulate the historic scatter about the regression line placed through the values of Y_{t+1} on Y_t.

Let us henceforth call R a residual number {V} until it has been shown that V should be defined as a random number {R}.

Rewriting Eq. 1.2 with the error estimator term from Eq. 1.1 added

$$Y_{t+1} = a + b(Y_t) + S_{Y_{t+1}/Y_t} R$$

and solving for the residual {V = R} we get

$$V = \frac{Y_{t+1} - a - b(Y_t)}{S_{Y_{t+1}/Y_t}} \qquad 4.1$$

4.1 Examination of Residual Mean and Standard Deviation

Twenty-four sequences of historical annual streamflow are examined with Eq. 1.2 as the regression line about which the scatter of points is to be examined {see Appendix B}. Once the regression constants {a and b} and the standard error of estimate {S_{Y_{t+1}/Y_t}} are evaluated it is possible to use Eq. 4.1 to calculate all of the residual numbers necessary to reproduce the exact scatter developed by nature.

These values of "V" are then grouped versus their respective Y_t values so that the statistical properties of these residual numbers can be examined. This examination is conducted by placing a straight line regression $R = C + d(Y_t)$ through the values of R on Y_t and employing the moving band technique described earlier.

Both of the regression coefficients {c & d} were found to be zero in each of the twenty-four tests, proving that the mean value $\{\bar{V}\}$ of the historic residuals for both high and low flows is equal to zero.

Again, the standard deviation values $\{\sigma\}$ for each band position were essentially equal, illustrating that each set of thirty-five points incremented to the right across the values of R on Y_t each time by four points was equally representative of the total population of points.

Also, there were visible trends in either or both the skewness coefficient $\{\beta_1\}$ and the kurtosis coefficient $\{\beta_2\}$ for each of the twenty-four tests.

Three of these tests are summarized in Table 4.1.

Table 4.1: Summary of Residual Number Statistics

Mid-Band Value (cfs)	σ	β_1	β_2
a) Columbia River at The Dalles, Ore., with N = 85 years and $\bar{X}$ = 202,242 cfs			
162,350	.954	-.064	1.95
169,700	.967	-.000	1.84
176,250	.967	-.018	2.02
185,200	.970	-.169	2.38
191,450	.989	-.070	2.34
199,300	.947	-.338	2.83
204,350	.906	-.185	3.01
208,350	1.007	.206	4.62
213,400	1.028	.098	4.22
218,450	1.009	.109	4.50
222,700	1.007	.040	4.52
226,250	1.057	.027	4.20
238,150	.960	.125	4.87
263,950	.967	.120	4.78
b) Merrimac River at Lawrence, Mass., with N = 71 years and $\bar{X}$ = 6,836 cfs			
5180	.982	.211	2.86
5922	1.024	.364	2.80
6122	1.081	.171	2.53
6452	1.081	.281	2.72
6603	.920	-.000	1.85
6934	.913	.001	1.92
7382	.950	-.030	1.75
7577	.945	-.011	1.79
7908	.997	-.112	1.96
c) Oostanaula River at Resaca, Ga., with N = 68 years and $\bar{X}$ = 2,748 cfs			
2011	.994	-.006	2.32
2200	1.047	-.046	2.09
2468	1.047	-.031	2.00
2632	1.107	-.011	1.91
2743	1.083	.002	2.07
2897	1.008	-.013	2.13
3025	1.068	.002	2.28
3262	1.000	-.007	2.59
3600	.964	.004	2.65

Looking at the values tabulated for the Columbia River {Table 4.1-a} it can be seen that the residual numbers are negatively skewed {β_1<1.0} flows below the mean {$\overline{X}$} and positively skewed {β_1>1.0} at flows above the mean {$\overline{X}$}. Also the kurtosis {β_2} varies from flat {Platykurtic} at flows below the mean to peaked {leptokurtic} at flows above the mean. The standard deviation {σ} is approximately equal to unity for all positions of the band although it is closer to unity at flows near the mean.

The Merrimac River has characteristics opposite to that of the Columbia River {Table 4.1-b}. It is positively skewed at lower flows and negatively skewed at higher flows. The kurtosis coefficient in general, decreases as the flow increases although it always remains flat or platykurtic. Note that again the standard deviation is very close to unity.

The third test using the historic data of the Oostanaula River shows still another trend {Table 4.1-c}. Here, the skewness is close to zero {on the negative side} at both high and low flows with a platykurtic kurtosis coefficient constant at a value of $\beta_2 \approx 2.25$. The value of the standard deviation is again essentially equal to unity.

In all of the twenty-four tests the mean value of the residual number {$\overline{V}$} was essentially equal to zero, and the standard deviation {σ} was essentially equal to one. However as has been shown, the skewness and kurtosis coefficients {β_1 and β_2} possess various visual trends which are not characteristic of a normal distribution in which β_1 and β_2 equal 0.0 and 3.0 respectively for any flow value.

The fact that all tests show the mean value $\{\bar{V}\}$ and the standard deviation $\{\sigma\}$ of the residuals to be respectively 0.0 and 1.0, satisfies two of the requirements of a normal random number but, because the skewness and kurtosis values were found to be non-normal, the validity of them as non-normal must be checked.

4.2 Validity Examination of Skewness and Kurtosis as Non-Normal

The main question to be answered here is, do we have a sufficient length of historic record to create stable predictions of either non-normal skewness form or non-normal kurtosis form for a given sequence?

To answer this question the eighty-five year record of the Columbia River at The Dalles, Oregon is again examined. If an eighty-five year record is sufficient, then the statistical properties of the residual numbers calculated when various small amounts of record are deleted from the full record should equal the statistical properties calculated when the full record is examined. {One, two, three, four or five years of record are eliminated in this test.} Also, the statistical properties of the residual numbers should remain constant when the property calculations are based upon either the first seventy-seven years of record or the last seventy-seven years of record.

To check the above hypothesis, the residual values $\{V\}$ are calculated for several lengths of record obtained by deleting various amounts from the total eighty-five year record. For each record length;

1) An array of the calculated residual number values $\{V\}$ on their respective Y_t values is made.

2) Through the array of these V on Y_t values, a straight line regression $\{V = c + dY_t\}$ is placed using the method of least squares.

3) Then skewness and kurtosis coefficients are calculated for the scatter of V values about the regression line $\{V = c + dY_t\}$ employing the moving band technique as was done in Section 4.1.

4) Finally, two new arrays are made;

   a) An array of the skewness coefficient values $\{\beta_1\}$ on their respective mid-band Y_t values.

   b) An array of the kurtosis coefficient values $\{\beta_2\}$ on their respective mid-band Y_t values.

5) Through these two new arrays a straight line regression is placed.

The first new array {step 4-a} is an array of skewness values versus their respective mid-band values for all band positions in the V on Y_t array. A straight line regression $\{\beta_1=h + i(Y_t)\}$ is placed through the positive skewness versus mid-band value portion and then through the negative skewness versus mid-band value portion of this total skewness versus mid-band value array. {See Appendix C for an explanation of why the negative and positive portions had separate regressions}. The slope and intersection of these straight line regressions should be constant for the various record lengths examined if the historic length of record is sufficient. The second new array {step 4-b} is an array of kurtosis versus its respective mid-band value for all band positions in the V on Y_t array. A straight line regression $\{\beta_2 = h' + i'(Y_t)\}$ is then placed through this total kurtosis versus mid-band value array. The slopes {i &i'} and intersections {h & h'} of these straight line regressions should be constant for the various record lengths examined if the historic record length is sufficient.

The comparison of the skewness and kurtosis regression formulations for the various record lengths found using the above procedure is tabulated in Table 4.2.

Table 4.2: Skewness and Kurtosis Trend Formulation Examination

a) Skewness predictions for various sequences of the 85-year historic Columbia River record. $\{\beta_1 = h + i(Y_t)\}$

Partial Sequence Examined from N = 85 years	Positive β_1 "versus" Mid-band (Y_t)			Negative β_1 "versus" Mid-band (Y_t)		
	h	i	S_{β_1/Y_t}	h	i	S_{β_1/Y_t}
First 77 years	.197	-2.0×10^{-7}	.042	0.91	-5.6×10^{-6}	.056
Last 77 years	.020	1.0×10^{-7}	.035	1.42	-8.6×10^{-6}	.062
Full 85 years	.198	-4.1×10^{-7}	.059	0.96	-5.9×10^{-6}	.080
First 84 years	.127	-1.7×10^{-7}	.050	0.58	-3.7×10^{-6}	.066
First 83 years	.790	-3.3×10^{-6}	.030	1.23	-7.5×10^{-6}	.072
First 82 years	-.090	8.0×10^{-7}	.031	1.02	-6.4×10^{-6}	.056
First 81 years	-.250	1.4×10^{-6}	.040	1.04	-6.5×10^{-6}	.057
84 year sequence with largest flow value absent	--	--	--	0.73	-4.6×10^{-6}	.064

b) Kurtosis predictions for various sequences of the 85-year historic Columbia River record. $\{\beta_2 = h' + i'(Y_t)\}$

Partial Sequence Examined from N = 85 years	h'	i'	S_{β_2/Y_t}
First 77 years	-5.19	4.20×10^{-5}	.55
Last 77 years	-3.02	3.10×10^{-5}	.58
Full 85 years	-4.46	3.80×10^{-5}	.52
First 84 years	-4.52	3.80×10^{-5}	.59
First 83 years	-6.17	4.80×10^{-5}	.44
First 82 years	-4.37	3.80×10^{-5}	.52
First 81 years	-3.77	3.44×10^{-5}	.45
84 year sequence with largest flow value absent	-2.36	2.56×10^{-5}	.19

It can be seen in Table 4.2-a that the skewness slope fluctuates markedly when various sequences of the total eighty-five year record are used. The slope of the positive skewness portion even reverses itself three times. For instance, when the first seventy-seven years of record are used the slope has a magnitude of twice that found when the last seventy-seven years of record are used and is a negative value; whereas, the last seventy-seven years of record has a slope value which is positive. Also, if the largest value of annual streamflow is removed from the total sequence and the resulting record is examined, the skewness normally found to be positive at higher flows becomes negative. The skewness intersection also fluctuates markedly, with a range from +0.20 to -0.25 which has a mean skewness intercept approximately equal to zero. Therefore, because the slopes are quite flat, with a mean intersection value of zero, and because the values of slope and intersection change so drastically for small changes in record length from a given sequence, it must be concluded that existing historical streamflow record does not have sufficient length to formulate residual number skewness trends which might be used to improve the predictability of Eq. 1.1.

Further, it can be seen from Table 4.2-b that the kurtosis slope fluctuates between 2.56×10^{-5} and 4.80×10^{-5} when various sequences of the total eighty-five year record are used and that the kurtosis intercept fluctuates between values of $|6.17|$ and $|2.36|$. This seemingly unstable formulation of a residual number kurtosis trend is supported by Fiering when he states that "the small sample instability of estimates of higher

moments increases astronomically, so that there is little utility in trying to preserve a parameter whose estimated value might easily be in error by several orders of magnitude."[13] Therefore, it must again be concluded that existing historical streamflow record does not have sufficient length to develop either residual number {R} skewness or kurtosis trend formulations that would be statistically valid.

In summary, the examination of the residual numbers {R} has found that the residual numbers can be classified as random variables, normally distributed with a mean of zero and a standard deviation of one because there appears to be insufficient record length upon which formulations of skewness and/or kurtosis can be developed. Therefore Eq. 1.1 appears to be a relatively good algorithm for predicting annual streamflow.

However, another avenue of exploration that might lead to an improvement in predictability of sequential annual streamflow remains unexplored. That avenue is the use of higher order Markov chains and is examined in Chapter V.

CHAPTER V

EXAMINATION OF MULTIPLE DEGREE MARKOV CHAINS

It has been shown in Chapters I thru IV that a one degree Markov chain {Eq. 1.2} best represents the least square regression line of Y_{t+1} on Y_t. Let us now examine this algorithm form to see if two, three or higher degree Markov chains might reduce the unexplained error and hence give us better predictability of Y_{t+1} than the one degree chain.

The general equation for a multiple degree Markov chain is given mathematically as

$$Y_{t+1} = a + bY_t + C_1 Y_{t-1} + C_2 Y_{t-2} \cdot \cdot \cdot \cdot + C_m Y_{t-m} \qquad (5.1)$$

5.1 Multiple Degree Chain Examination Procedure.

To test the predictability of Y_{t+1} using higher degree Markov chains we again use the standard error of estimate as an indicator and apply the following steps.

1) Calculate the standard error of estimate value for the one degree Markov chain least squares regression.

2) Calculate the standard error of estimate values for the two, three . . . . . (m+1) degree chain least squares regressions conserving the requirement that $[N-(m+2)] \geq 30$ {the minimum statistical set size}.

3) Subtract the higher degree standard error of estimate values from the one degree value, divide by the one degree value and multiply by 100%. If a resulting percentage is plus, the higher degree chain under examination is a better regression fit than the one degree chain and vice versa.

This procedure was employed in the multiple degree Markov chain examination. Twenty-five streams with various historic record lengths and hydrologic conditions were examined.

5.2 The 1 Thru 10 Degree Chain Examination

Using Eq. 5.1, varying the "m" value from "0" to "9", and employing the test procedure outlined in Sec. 5.1 the percentage change in the standard error of estimate value was calculated for the one through ten degree chains at twenty-five different river gaging stations. These percentage changes are tabulated in Table 5.1-a and 5.1-b. Note again that for a statistically valid set size a minimum of thirty points are required. Hence, for a six degree chain examination to be valid thirty-seven years $\{N \geq 30 + 6 + 1\}$ of record are required. For a ten degree chain examination forty-one years $\{N \geq 30 + 10 + 1\}$ of record are required. The percentage values listed for higher order chains which have set sizes less than thirty as defined only show trends. These trends are not valid statistically but are still good indicators of the standard error of estimate behavior for higher order chains.

Out of twenty-five streams tested, fifteen showed improvement {a plus percentage change in the standard error of estimate value as outlined above} when higher order chains were used. It can be seen in Table 5.1-a that the statistically valid improvements in the standard error of estimate value range from 0.4% to 15.2%. At the same time we see good trend indications of improvement possibly as high as 34%. Both trend improvements and statistically valid improvements group about the six degree Markov chain with this six degree chain as both median and mode.

Table 5.1: Standard Error of Estimate Improvement for Higher Order {2 thru 10 degree} Markov Chains.

a) Fifteen Streams Showing Significant Improvement.

Stream Name	N	% Change in Standard Error of Estimate Value for 2 thru 10 Degree Chain Examinations									
	(Yrs)	1	2	3	4	5	6	7	8	9	10
Duckabush R., Wn.	27	Base	-0.3	-1.2	-5.8	+10.0	+14.7	+34.3	+32.4	+26.1	+15.1
S. F. Skokomish, Wn.	34		-1.4	-0.8*	-1.2	+ 4.9	+ 8.2	+ 7.0	+ 4.3	+ 9.2	+ 3.7
N. F. Skokomish, Wn.	41		+1.8	+0.9	0	+ 3.7	+15.2	+12.5	+ 9.2	+10.5	+14.2*
Dungeness R., Wn.	28		+2.1	+1.8	-0.4	- 3.3	+ 1.5	+ 2.0	+ 4.0	- 4.8	-10.7
Wynoochee R., Wn.	40		-2.2	-4.1	-1.0	+ 5.3	+ 3.1	+ 3.6	+ 0.7	+ 3.5*	+ 2.5
Sauk R., Wn.	37		+0.3	-2.2	-5.2	- 5.0	+ 3.2*	- 1.1	- 3.8	0	- 3.1
White R., Wn.	36		-0.3	-3.4	-5.4	+ 0.6*	+ 3.2	+ 2.2	+ 9.3	+ 5.0	+ 1.7
Carbon R., Wn.	36		+4.2	+1.1	-1.5	+ 7.0*	+ 7.8	+ 5.9	+14.0	+ 9.4	+ 4.2
Puyallup R., Wn.	34		+0.1	+5.9*	+2.3	- 0.2	+ 2.4	+ 8.6	+ 5.3	- 0.9	- 2.9
Metolius R., Ore.	41		+3.3	+3.0	+0.9	0	+ 2.7	0	- 2.8	- 3.8	- 3.8*
Green R., Utah	56		-0.7	+0.4	+0.7	+16.9	+16.8	+17.3	+15.3	+14.7	+13.4*
Oostanaula R., Ga.	68		+1.4	- 0	+0.3	+ 4.8	+ 4.8	+ 4.8	+ 3.4	+ 3.8	+ 2.6*
Yellowstone R., Mont.	50		-2.1	+0.9	-0.7	- 0.3	+ 5.1	+ 9.6	+ 9.8	+ 8.4	+ 5.7*
Satsop R., Wn.	36		-2.8	+0.4	-1.4	+ 0.4*	+ 1.0	- 0	+ 9.7	+ 7.6	+15.6
Skykomish R., Wn.	37		+0.2	-2.0	-3.6	- 0.3	+ 4.5*	+ 0.3	- 2.4	+ 0.9	- 0.3

* Highest degree value within minimum statistical validity

Table 5.1: Continued

b) Ten Streams Showing No Significant Improvement

Stream Name	N (Yrs)	% Change in Standard Error of Estimate Value for 2 thru 10 Degree Chain Examinations									
		1	2	3	4	5	6	7	8	9	10
Hoh River, Wn.	38		-2.7	0	-0.3	+0.4	-2.4	-4.7*	- 3.9	- 1.4	- 4.4
S. F. Skykomish R., Wn	54		-1.7	-3.4	-1.8	-0.7	-1.2	-1.7	- 3.4	- 6.4	- 8.7*
N. F. Stillagaumish,Wn	37		+1.1	-2.0	-4.2	-1.2	-1.9*	-2.7	- 7.1	- 2.7	- 6.5
Quinalt R., Wn.	54		-1.6	-2.3	-2.7	-3.2	-2.1	-3.6	- 4.5	- 4.7	- 6.0*
Wenatchee R., Wn.	55		-1.9	-3.0	-4.9	-2.9	-1.2	-3.1	- 4.7	- 7.4	- 9.1*
S.F. Nooksack R., Wn.	33		-3.1*	-7.0	-7.4	-4.8	-1.8	-6.5	-10.8	-13.9	-17.7
Columbia R., Ore.	85		+0.6	+0.4	0	-0.9	-0.7	-2.0	- 2.5	- 2.4	- 3.8*
" (first 40 yrs.)	41	Base	+ 0	+0.1	-2.6	-4.0	-6.4	-10.2	- 9.2	- 8.3	-13.5*
" (first 30 yrs.)	31		-0.1	-2.5	-6.5	-8.2	-13.6	-19.8	-15.9	- 9.1	-18.4
Merrimack R., Mass.	72		-1.3	-0.6	+0.8	-0.5	-1.4	- 2.1	- 2.8	- 4.7	- 5.8*
" (first 35 yrs.)	36		+0.2	-0.7	+1.6	-1.1*	-4.6	- 3.9	- 3.4	- 2.3	- 1.0
Cascade R., Wn.	37		-0.9	-4.0	-6.6	-5.6	-1.6*	- 6.1	- 9.1	- 7.4	-11.3
S.F. Stillaguamish R., Wn.	37		+0.5	-2.8	-2.8	-2.0	-1.7*	- 6.1	- 7.7	- 3.8	- 1.9

* Highest degree value within minimum statistical validity

The remaining ten streams from the twenty-five tested did not show significant reduction in the standard error of estimate when higher order Markov chains were examined. Looking at the values tabulated in Table 5.1-b it can however, be seen that there is a tendency for the five, six and seven degree chain values to be substantially less negative than the higher order chain values on either side of these three. {This might indicate that the parameter causing the five, six and seven degree chains to have smaller standard error of estimate values than the one degree chain in the first fifteen drainage basins examined is not as strong an influence in the last ten drainage basins examined}. In fact, the six degree chain is again the mode and median for the grouping of these substantially less negative values. This suggests that the persistence represented by the isolated five, six, and seven year lags might have more influence upon the algorithm than the persistence represented by the isolated one through four year lags. A test for the validity of this suggestion was conducted by isolating the five, six, seven, eight and ten year lags from the N. F. Skokomish River data to form five different regression equations from which five different standard error of estimate values could be calculated.

These five standard error of estimate values were all larger than the standard error of estimate value obtained when the full six degree Markov chain was used as can be seen in Table 5.2 Regression lines formed by combining lag separations {i.e. 1 & 5; 1 & 7; 1 & 10; and 5, 6, 7 & 8} were also examined and their resulting standard error of estimate values are tabulated in Table 5.2.

Table 5.2: Standard Error of Estimate Improvement Check for Isolated Lags and Lag Combinations Using the N. F. Skokomish River Data.

Lag	Lag Separation Equation*	Standard Error of Estimate Value
1 Year	$Y_{t+1} = a+b(Y_t)$	96.61 cfs.
5 Year	$Y_{t+1} = a+c_4(Y_{t-4})$	95.32 cfs.
6 Year	$Y_{t+1} = a+c_5(Y_{t-5})$	89.51 cfs.
7 Year	$Y_{t+1} = a+c_6(Y_{t-6})$	88.59 cfs.
8 Year	$Y_{t+1} = a+c_7(Y_{t-7})$	88.74 cfs.
10 Year	$Y_{t+1} = a+c_9(Y_{t-9})$	92.03 cfs.
1 & 5 Year	$Y_{t+1} = a+b(Y_t) + c_4(Y_{t-4})$	90.94 cfs.
1 & 7 Year	$Y_{t+1} = a+b(Y_t) + c_6(Y_{t-6})$	89.00 cfs.
1 & 10 Year	$Y_{t+1} = a+b(Y_t) + c_9(Y_{t-9})$	92.12 cfs.
5,6,7 & 8 Year	$Y_{t+1} = a+c_4(Y_{t-4}) + c_5(Y_{t-5}) + c_6(Y_{t-6}) + c_7(Y_{t-7})$	89.77 cfs.
1,2,3,4,5 & 6 Yr.	$Y_{t+1} = a+b(Y_t) + c_1(Y_{t-1}) + c_2(Y_{t-2}) \ldots c_5(Y_{t-5})$ {Eq. 5.1}	81.95 cfs.

*Equation 5.1 with some terms dropped, leaving isolated lags and lag combinations for predicting the Y_{t+1} value using the N. F. Skokomish River Data.

In no case did a lag separation examination or combined lag separation examination give a better reduction in the unexplained error than did the full uninterrupted six degree chain. We can therefore conclude that the one through four year lags are as necessary to the full six degree Markov chain as are the five and six year lags.

The fact that fifteen of the twenty-five streams examined above show reduction in the unexplained error when full higher order Markov chains are used as regression equations, indicates that there is better than a 50 percent chance that any stream examined might best be represented by a multiple degree Markov chain regression equation {Eq. 5.1} instead of the one degree equation {Eq. 1.2}.

Further examination of higher order Markov chains was conducted because of the high probability that they might give improved algorithms for many drainage basins. One such examination looks at the standard error of estimate values for the one degree through twenty-four degree Markov chains.

5.3 The 1 Thru 24 Degree Chain Examination

Four streams were examined using Markov chains from one degree to twenty-four degrees in size. The purpose of this test was to see how large a chain could be employed and still have it develop a smaller standard error of estimate value than that obtained using a one degree chain. Fiering notes a rationale for determining the number of lags that can be added to increase the multiple Markov chain.[13]

1) Lags generally increase with record length.

2) Truncation and round off error limits lag on strictly numerical (not statistical) grounds.

3) An arbitrary limit of 20 is imposed.

This rationale supplements the minimum statistical requirement that a minimum of thirty points be contained in the set for valid representation. For a twenty-four degree chain then to be statistically valid a minimum of fifty-five years of record $\{N \geq 30 + 24 + 1\}$ must be available.

The resulting standard error of estimate values obtained in this test are tabulated in Table 5.3. A close examination of this tabulation reveals signs of possible cyclicity. This cyclicity is represented by a substantial reduction in the standard error of estimate value when higher order chains of approximately one, six, twelve, eighteen and twenty-four degrees in size are employed {approximate multiples of six}. For instance, the North Fork of the Skokomish River has minimum standard error of estimate values when the one, six and thirteen degree chains are used. The Yellowstone River has minimum values at about the three, eight, twelve and eighteen degree chains. The standard error of estimate values minimize when the six, twelve, seventeen and twenty-four degree chains are used as regression equations for the Oostanaula River data. Finally, the minimum standard error of estimate values for the Columbia River at The Dalles, Oregon are developed when the two, six, eleven and sixteen degree chains are employed. The factors or parameters which control this action are not

Table 5.3: Tabulation of Standard Error of Estimate Values Displaying Possible Signs of Cyclicity.

	Standard Error of Estimate Values			
Chain Size	N.F. Skokomish R. N = 41	Yellowstone R. N = 50	Oostanaula R. N = 68	Columbia R. N = 85
1	96.6	641.0	740.5	34,302
2	94.8	654.6	730.3	34,113
3	95.7	635.0	741.1	34,150
4	96.6	645.6	738.5	34,220
5	93.0	643.2	705.0	34,606
6	81.9	608.3	704.9	34,546
7	84.6	579.0	704.4	34,994
8	87.7	578.1	715.6	35,143
9	86.5	586.8	712.7	35,113
10	83.0*	604.4	721.4	35,604
11	83.5	616.5	732.0	34,993
12	86.4	557.1	708.2	35,553
13	69.1	575.6	722.3	35,723
14	72.9	601.0	737.2	36.072
15	67.5	564.0	751.3	36.577
16		581.2	757.4	33,306
17		500.3	756.4	33,593
18		423.2	768.9	33,958
19		444.4*	791.3	34,028
20			804.4	34,504
21			826.2	34,904
22			837.5	35,696
23			865.8	36,529
24			854.0	37,508

* The highest degree value that satisfies minimum statistical validity.

understood at this point, but the fact that multiple degree Markov chains and chains which are approximate multiples of six give substantial reductions in the unexplained variation could be a key to better algorithms for generating stochastic annual streamflow.

CHAPTER VI

SUMMARY

The primary reason for examining the regression equation relating serially correlated annual streamflow data was the hope that an improved model for the simulation of annual stream runoff could be found.

Non-linear regression equations were examined for thirty-one historic streamflow sequences but those examined showed no significant improvement above the simulation capability of a linear equation.

Data arrays obtained by serial grouping $\{Q_{i+1}$ on $Q_i\}$ were examined to see if subsets from the total array at various positions along the regression equation had normal distributions of Q_{i+1} about the subset mean $\overline{Q}_{i+1}$. Indications that subsets at certain points along the regression were non-normal were later proved unstable and therefore the assumption that they be treated as normal is valid until larger sequences of data become available.

The hypothesis that the unexplained variation was normal and hence that random variables with zero mean and unit variance could be used in conjunction with the standard error of estimate to reproduce this unexplained variance was also examined using thirty-one streamflow sequences. This examination showed that with the virgin, historic, annual streamflow data available today, the hypothesis is a good one. However, as time goes by and more historic record becomes available, it might be possible to modify the linear regression to include a skewness or kurtosis trend determinable and stable for a given drainage basin.

Finally, an examination of serially correlated annual streamflow data was made using higher order {multilag} Markov chains as regression equations. These regression equations reduced the unexplained error below the value obtained when a one degree Markov chain {the linear algorithm} was employed and therefore become the best available algorithm for use in predicting or generating annual streamflow sequences. Unfortunately the physical reason why the higher order Markov chain gives better results is not known at this point. Some interesting events unexamined in this study due to lack of time are:

1) The higher order chains which are approximate multiples of six seem to give the best predictions, i.e., the six, twelve and eighteen degree chains.

2) The intercept coefficient for the Oostanaula River is greater than the average annual flow, suggesting a trend negative to that which initiated this model. However, this could be explained by the geology of the basin, i.e., broken shale and its reaction to moisture.

3) When the six degree chain was used to generate small sequences the average obtained seemed to increase slightly above the historic value. This increase seemed to be related to the values used as initiators of the short sequence and to the regression slope values.

If the degree of chain to be used and the regression coefficients associated with the chain can be correlated with some parameter common to all drainage basins that respond favorably to higher order Markov chains, then we will have an algorithm available to simulate annual streamflow sequences with less unexplained error than any algorithm in existence today. This method might even be extended to improve monthly streamflow generation as is indicated by one short study using a two degree Markov chain.

APPENDIX

APPENDIX A

LISTING OF THE THIRTY-ONE STREAMS USED IN NON-LINEAR REGRESSION EXAMINATION

Sequence Number	Stream
1	N.F. Ahatahum Cr. near Tampico, Wn. near Asotin, Wn.
2	Carbon R. near Fairfax, Wn.
3	Cascade R. at Marblemount, Wn.
4	Columbia R. at The Dalles, Ore.
5	Duckabush R. near Brinnon, Wn.
6	Dungeness R. near Sequim, Wn.
7	Green River near Palmer, Wn.
8	Greenwater R. at Greenwater, Wn.
9	Hoh R. near Spruce, Wn. near Forks, Wn.
10	Naselle R. near Naselle, Wn.
11	S.F. Nooksack near Wickersham, Wn.
12	North R. near Raymond, Wn.
13	Puyallup R. near Orting, Wn.
14	Quinalt R. at Quinalt Lake, Wn.
15	Satsop R. near Satsop, Wn.
16	Sauk R. near Sauk, Wn.
17	Sauk R. above Whitechuck R. near Darrington, Wn.
18	N.F. Skokomish R. below Staircase Rapids near Hoodsport, Wn.
19	S.F. Skokomish R. near Union, Wn.
20	Skykomish R. near Goldbar, Wn.
21	S.F. Skykomish R. near Index, Wn.
22	Soleduck R. near Fairholm, Wn.
23	Stehekin R. near Stehekin, Wn.
24	Stetattle Cr. near Newhalen, Wn.
25	N.F. Stillaguamish R. near Arlington, Wn.
26	S.F. Stillaguamish near Granite Falls, Wn.
27	Sultan R. near Startup, Wn.
28	Thunder Cr. at Newhalem, Wn.
29	Wenatchee R. near Plain, Wn.
30	White R. at Greenwater, Wn.
31	Wynoochee R. below Staircase Rapids near Hoodsport, Wn.

Note: Flows for above came from Water Supply Papers U. S. Geological Survey.

APPENDIX B

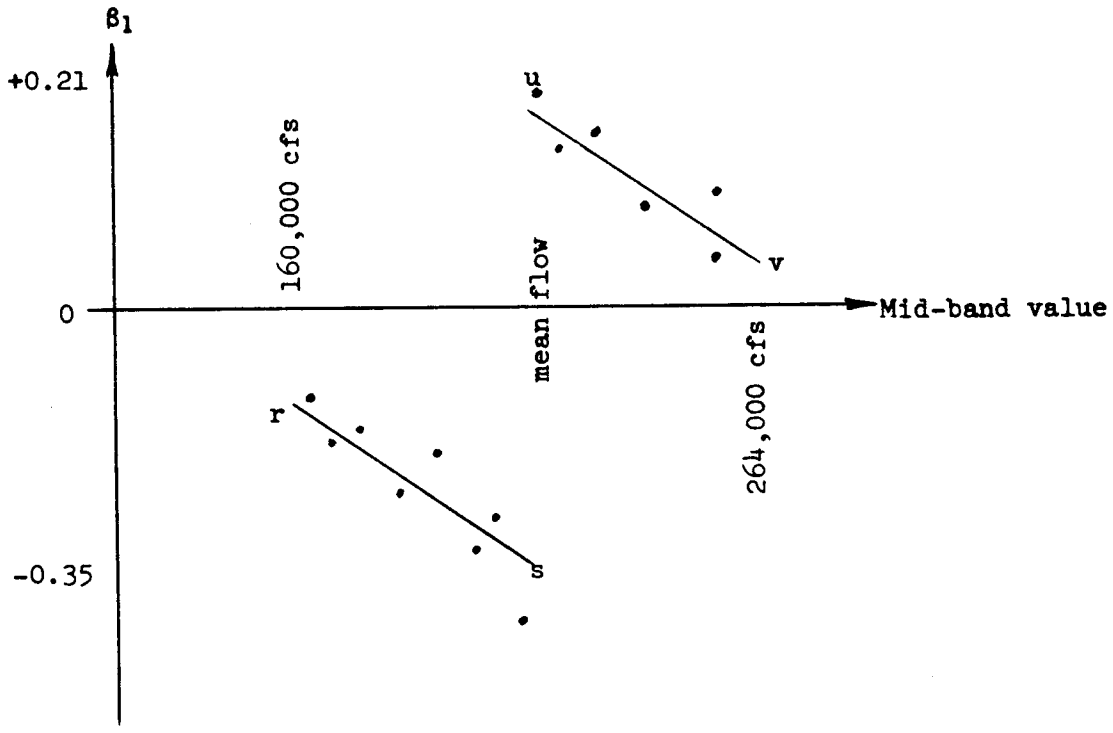
LISTING OF THE TWENTY-FOUR STREAMS USED IN THE EXAMINATION FOR THE PRESENCE OF RANDOM VARIATIONS

Sequence Number	Stream
1	Schoharie Creek at Prattsville, N.Y.
2	James River at Buchanan, Va.
3	Roanoke River at Roanoke, Va.
4	Chattahoochee River at West Point, Ga.
5	Greenbrier River at Alderson, W. Va.
	Allegheny River at Red House, N.Y.
	Wolf River at New London, Wis.
	Neches River near Rockland, Tex.
	Mill Creek near Salt Lake City, Utah
10	Kings River at Piedra, Calif.
	Arroyo Seed River near Soledad, Calif.
	South Branch Nashua River at Clinton, Mass.
	Penobscot River at Millinocket, Maine
	Presumpscot River at Outlet of Sebago Lake, Maine
15	Oostanaula River at Resaca, Ga.
	French Broad River at Asheville, N.C.
	Mississippi River at St. Paul, Minn.
	Red River of the North at Grands Forks, N. Dak.
	Yellowstone River at Corwin Springs, Mont.
20	Osage River near Bagnell, Mo.
21	Brazos River at Waco, Tex.
23	Green River at Green River, Utah
24	St. Lawrence River at Ogdensburg, N.Y.

APPENDIX C

EXPLANATION OF SKEWNESS VERSUS MID-BAND VALUE REGRESSION SEPARATION

If the skewness values were plotted versus their respective mid-band values for the Columbia River when the full historic record sequence is used the plot would look like that shown below.



The slope of a visual linear regression line through the negative skewness versus mid-band value portion of the plot {line rs} is approximately equal to the slope of a visual linear regression line through the positive skewness versus mid-band value portion of the plot {line uv}. For this reason, the total plot was broken into two portions as described in CHAPTER IV.

REFERENCES

1. Hazen, Allen, Storage to be Provided in Impounding Reservoirs for Municipal Water Supply," Transactions of the ASCE, 17, 1939, (1914).

2. Sudler, C. E. "Storage Required for the Regulation of Streamflow," Transactions of the ASCE, 91, 622, (1927).

3. Yule, G. U., "On The Method of Investigating Periodities in Disturbed Series With Special Reference to Wolfer's Sunspot Numbers," Transactions of the Royal Society, Series A, 226, 267, (1927).

4. Barnes, F. B. "Storage Required for a City Water Supply," Journal of the Institute of Engineers, 26, 198, Australia, (1954).

5. Brittan, M. R. "A Probability Model for Integration of Glen Canyon Dam Into the Colorado River System," Ph.D. Dissertation, University of Colorado, Boulder, Colorado, (August 1960).

6. Julian, P. R. "A Study of the Statistical Predictability of Stream Runoff in the Upper Colorado River Basin," Part II of Past and Probable Future Variations in Streamflow in the Upper Colorado River. University of Colorado Bureau of Economic Research, Boulder, Colorado, (Oct. 1961).

7. Thomas, H. A. and M. B. Fiering, "Mathematical Synthesis of Streamflow Sequences for the Analysis of River Basins by Simulation," Chapter 12 of Design of Water Resource Systems, Edited by A. Maas et al., Harvard University Press, Cambridge, Massachusetts, (1962).

8. Maughan, W. D., and R. Y. Kawans, "Project Yield by a Probability Method," Proceedings of the ASCE, Journal of the Hydraulics Division, (May 1963).

9. Yagil, S., "Generation of Input Data for Simulation," IBM System Journal, (Sept. - Dec. 1963).

10. Yevdjevich, V. Y., Fluctuation of Wet and Dry Years," Part II of Analysis by Serial Correlation. Fort Collins, Colo., Colorado State University, (June 1964).

11. Parazen, E., Stochastic Processes, San Francisco, Holden-Day, (1962).

12. Croxton & Cowden, Applied General Statistics, Second Edition, Englewood Cliffs, N.J., Prentice-Hall, Inc., (July 1958).

13. Fiering, M. B., Streamflow Synthesis, Harvard University Press, Cambridge, Mass. (1967).

14. Harms, A. A. and Campbell, T. H., An Extension to the Thomas-Fiering Model for the Sequential Generation of Streamflow, Water Resources Research, Vol. 3 Third Quarter 1967 Number 3.

15. Spiegel, M. R., Theory and Problems of Statistics, Schaum Publishing Co., New York (1961).