Research Technical Completion Report

ANALYSIS AND GENERATION OF LOW-FLOW SEQUENCES FOR IDAHO STREAMS USING DISAGGREGATION MODELING

by

Dennis R. Horn
Donna Dieziger-Kim

Department of Civil Engineering

Idaho Water Resources Research Institute
University of Idaho
Moscow, Idaho 83843

September, 1985

The research on which this report is based was financed in part by the United States Department of the Interior as authorized by the Water Research and Development Act of 1978 (P.L. 95-467).

Research Technical Completion Report
G903-03

ANALYSIS AND GENERATION OF LOW FLOW SEQUENCES FOR IDAHO STREAMS USING DISAGGREGATION MODELING

by

Dennis R. Horn
Donna Dieziger-Kim

Department of Civil Engineering

Submitted to
U.S. Geological Survey
United States Department of the Interior
Washington, D.C. 20242

Idaho Water Resources Research Institute
University of Idaho
Moscow, Idaho 83843

September, 1985

TABLE OF CONTENTS

ABSTRACT

Stochastic models of streamflow were developed for two rivers in Idaho, and the results analyzed to assess model performance and the characteristics of droughts. Multivariate modeling methods were applied to both historical records to extend their length, based on nearby longer-term records, and the unextended and extended data then used to determine subsequent model parameters. Annual flow models, coupled with condensed parameter disaggregation models, were applied to generate 40,000 years of annual/monthly streamflow records. The statistics and probability distributions of the annual and monthly flows comprising drought sequences are presented, and the theory of runs is used to estimate return periods of historical drought events. It is concluded that the assignment of probabilities to droughts based on historical record length yields inconsistent results when compared to the long-term stochastic process, and that data extension has a significant effect on critical model and run-definition parameters, providing improved estimates of population statistics. Procedures are suggested for using the modeling results for storage reservoir design, and for developing regionalized drought characteristics for Idaho streams.

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

BACKGROUND: Hydrologists and engineers involved in the planning and design of surface water projects have always had to deal with the problem of hydrologic uncertainty. This problem arises from either a total lack of critical streamflow information at or near the proposed development site, or a streamflow record which is far too short to adequately characterize the hydrologic regime. Even in cases where a relatively long streamflow record exists, a design based purely on the extreme event (such as a critical drought or low-flow sequence) in that historical record may imply an unquantifiable risk level associated with design failure. The use of historical records as the sole basis for design also introduces a element of inconsistency into the design approach, since different projects will have very different assumed risk levels. This may result in under- or over-sized projects with economics far from the desired optimum.

The traditional approaches to minimizing hydrologic uncertainty (although, perhaps, not truly quantifying it) have generally been deterministic in nature. Streamflow or river basin models, using algorithms based on the physical processes involved, can be used to synthetically extend or generate new streamflow data, which are then evaluated as the basis for design. These approaches invariably rely on some form of stochastic time series such as precipitation as the primary input data, and lead to a set of flow data which has a unique correspondence to the input. Accordingly, most purely deterministic modeling efforts tend to reduce the random variability actually observed in natural streamflow. Also, as the complexity of these models increases, with a corresponding requirement for more input data at short time intervals, the cost of model application becomes prohibitively expensive for smaller water resources projects.

Researchers have long recognized the stochastic nature of the streamflow process, and have attempted to overcome some of the drawbacks of deterministic modeling by statistical or stochastic simulation of the process. Much of the work in recent years has concentrated on enhancing existing stochastic methods, or developing new methods which have improved capabilities. With the introduction of an approach classified as "disaggregation" modeling in the mid- to late-1970s, hydrologists now appear to have a powerful tool for characterizing, extending, and forecasting streamflow, with the preservation of all of the historical properties including annual and periodic variability. Unfortunately, the model development work has not yet been translated into many practical applications, and the applicability of these models to different flow regimes and different length data sets has not been tested.

As the need for streamflow storage, regulation, and low-flow management continues to grow in Idaho, there is a corresponding need to better define the risk levels associated with critical low-flow sequences on streams throughout the state. This is especially true for situations where over-year regulation is necessary to supply water during critical drought periods, since historical records and deterministic approaches have not been entirely successful in quantifying risk levels. It is also true for low-flow sequence estimation and prediction for a variety of other purposes, including fish and wildlife management, maintenance of aesthetics and environmental quality, and recreation. Therefore, improved estimates of critical low-flow conditions, whether for water supply, hydropower, or any other project objective, will lead to more reliable predictions of the risk of project failure and better project economics.

STUDY OBJECTIVES: This research study examines the application of stochastic disaggreagtion modeling techniques to two rivers in Idaho, both of which have the potential for future storage development or other regulation projects. These rivers are the Coeur d'Alene and the South Fork of the Boise.

The modeling applications and analysis of model results include the following specific research objectives:

1) To test the validity and effectiveness of using multivariate data extension techniques on short streamflow records, prior to the development of stochastic disaggregation model parameter.

2) To select, from among competing model types, appropriate annual and monthly disaggregation models, and use these models to generate long sequences of monthly streamflow data.

3) To test and evaluate the performance of the selected models in preserving the statistical characteristics and relationships observed in the historical and extended streamflow records.

4) To use the model results for evaluating the nature and properties of 1 determining estimates of the probabilities of historical critical drought sequences on the study streams.

5) To compare the study results for both rivers to ascertain whether or not the same procedures, model forms, and research conclusions are applicable to both hydrologic regimes.

6) To assess the possibility of establishing regionalized stochastic streamflow parameters for low-flow sequences in Idaho.

REPORT CONTENTS: The study report includes, in the following chapters and appendices, a thorough review of all research methodologies, findings, and conclusions. It has been organized in a time-sequential study task manner, beginning with a chapter devoted to the initial selection of study streams and ending with a summary of findings and conclusions from throughout the study sequence.

To reduce the length of the main body of the report, many of the graphs, tables, and computer output results have been placed in appendices, and referenced appropriately in the text.

CHAPTER 1

SELECTION OF STUDY STREAMS

Described in this chapter are the criteria established for reviewing the candidate streamflow records for modeling, a brief discussion of the two streamflow records that were finally selected, and a review of the streamflow records that were originally envisioned to be used.

1.1 Criteria for the Selection of Streamflow Records

This section presents a brief discussion of the "ideal" characteristics which selected streamflow records should possess, followed by the constraints or criteria resulting from these characteristics. A more detailed examination of these constraints and criteria can be found in Appendix A.

It was desired that the selected records be homogeneous and long enough to provide reasonable estimates of the model parameters. In addition, the records should be of the best possible quality.

1) Each record should consist of at least 30 years of data in order to help reduce the uncertainty associated with the model parameter estimates.

2) The streamflow records should represent natural conditions. In other words, there should be a minimal amount of regulation and/or diversion.

3) The streamflow records should be described as at least "fair" over their entire length.

How the hydrologic regime of a basin affects the disaggregation model parameters and drought characteristics was to be examined. Consequently, other variables' effects on these factors had to be minimized or eliminated. In an attempt to lessen the effect of other variables, the following criteria were established for the two selected records:

4) The lengths of each record should not differ by more than 50%.

5) The drainage areas of each basin should not differ by more than 100%.

To assess the possibility of establishing regionalized stochastic streamflow parameters for low-flow sequences in the state of Idaho was another objective of this study.

6) The two selected streamflow stations should be as far apart geographically as possible. This will give a better representation of two different parts of the state.

7) If possible the selected streamflow stations should be in areas which have the greatest potential for storage development or other regulation projects.

Data extension was another factor whose effect on disaggregation model parameters and drought charateristics was to be observed. A simple cross correlation with another streamflow record plus a stochastic component was the envisioned model to be used for data extension (13). Hence, a secondary streamflow record was needed which could be used for data extension.

8) The secondary streamflow record should temporally overlap the original streamflow record by as much as possible to provide the most reliable estimate of the cross correlation between the two records. Approximately 20 years of overlap was considered to be the minimum.

9) The secondary streamflow record should extend beyond, or precede, the orginal by as much as possible. The minimum extension length was considered to be 15 years.

10) The hydrologic reponses of the secondary stream should be as similar as possible to those of the original stream. This helps provide a strong cross correlation between the two records. Frequently, similar hydrologic responses are observed between streams that are geographically close together and have similar physiographic and topographic characteristics.

Since the state of Idaho was chosen for this study, the Water Resource Data publications for Idaho (36) were reviewed, keeping in mind the established criteria and constraints. Also, nearby streamflow records in the states of Montana, Utah, Nevada, and Washington were considered. This review resulted in the selection of station 12413000 (Coeur d'Alene River at Enaville) which would be extended by using the record at station 12413500 (Coeur d'Alene River at Cataldo); and station 13186000 (South Fork Boise River near Featherville) which would be extended by using the record at station 13185000 (Boise River near Twin Springs).

1.2 Description of Selected Streamflow Basins and Records

The streams which were selected for use in this study generally reach their annual peaks during the spring snowmelt, but warm rains and thawing conditions from Pacific storms may cause extreme floods during the winter months as well. The low flows usually occur during the late summer.

Table 1.1 summarizes some of the physiographic, climatic, and geographic characteristics of each chosen streamflow record. Most of this information was taken from two series of maps (4,16). Also presented in Table 1.1 are remarks about the quality of each record taken from the Water Resource Data for Idaho (36). Figure 1.1 illustrates the location of the streamflow stations.

If a comparison is made between the basin characteristics listed in Table 1.1 of stations 12413000 and 13186000, it appears that station 12143000 possesses a somewhat milder, wetter climate than station 13186000. This climatic difference is due to the lower average elevation and geographic location of station 12413000 relative to station 13186000. However, both stations occur within the same landform province: the Northern Rocky Mountain Province which is characterized by high mountains and deep intermountain valleys.

Table 1.1

Description of Streamflow Records and Basins

Station	12413000	12413500	13186000	13185000
Latitude	47°34'20"	47°33'50"	48°29'40"	43°39'33"
Longitude	116°15'10"	116°18'25"	115°18'20"	115°43'34"
Avg Ann precip(in)	30-60	30-60	20-50	20-40
Snowfall (in)	80-112	64-96	64-96	32-64
Mean altitude (ft)	4000[1]	4500[1]	6840	6350
Gage evelation (ft)	2100	2100	4220	3256
Drain area (sq mi)	895	1220	635	830
Avg min Jan (°F)	17-24	17-24	4-13	4-13
Avg max Jan (°F)	27-35	32-35	32-35	32-35
Avg min July (°F)	45-52	45-52	39-48	39-48
Avg max July (°F)	76-83	76-83	79-87	79-87
Landform Province	Northern Rocky Mt		Northern Rocky Mt	
Record Length (yr)	44	52	38	72
Diversion/regulation	none	none	diversion[2]	none
Record Quality	good	fair-good	fair-good	good
Gage Moved	no	no	no	.3 miles

1. Estimated from topographic map

2. In the Water Resources Data for Idaho (36) an estimate of the amount of land irrigated above station 13186000 was given as 450 acres (1966 determination). The lowest monthly historical flow recorded at Station 13186000 was 4254 cfs days which occurred in September 1977 (determined by examining monthly record listing). To estimate the quantity of water diverted during this month it was assumed that the 450 acres were irrigated with 3 feet of water. This would be a very high irrigation rate, but it was made conservative in case other diversions have taken place since 1966 and for the inefficiency of the irrigation system. Three feet of water over 450 acres in one month equals 1350 acre-feet or 22 cfs days which is .52% of 4254 cfs days. Consequently, the effect of the diversion was considered small enough to be ignored.

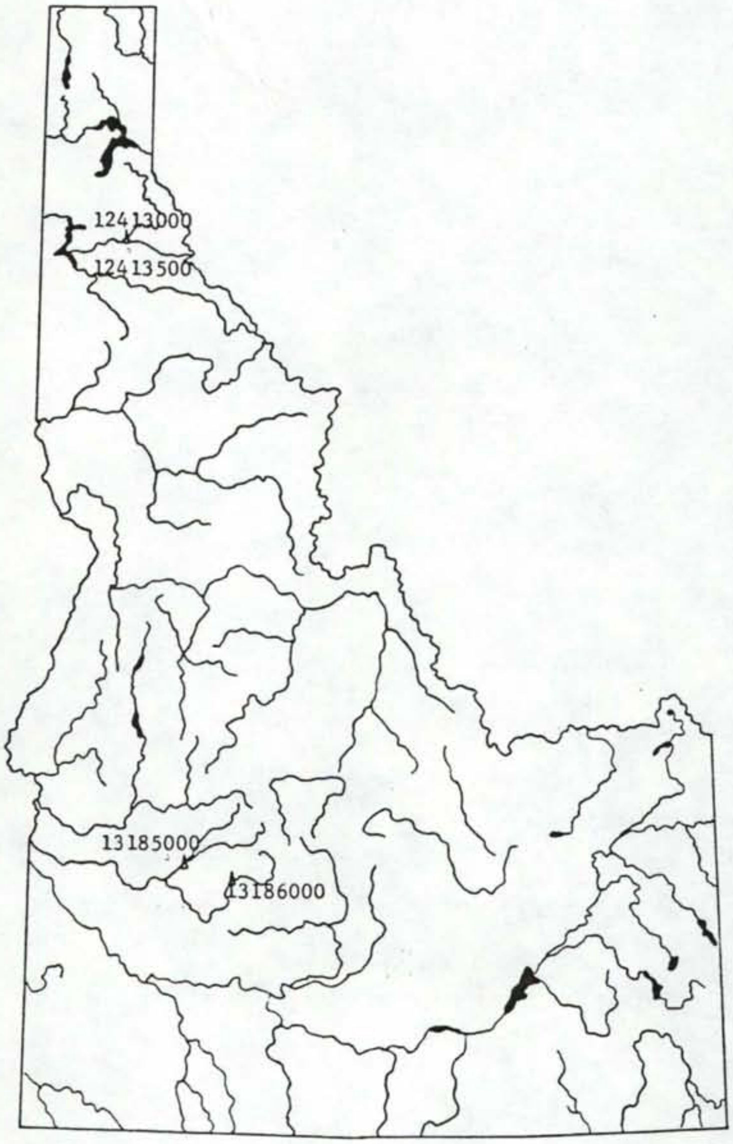


Figure 1.1 Location of Streamflow Stations

1.3 Streamflow Records Mentioned in Original Proposal

The Palouse and Teton Rivers were mentioned in the orginal proposal for this study as the probable streamflow records that would be used (13). However, due to the length of record, diversions, or regulation the records on these rivers were not selected. The streamflow records available on these rivers are listed in Table 1.2. The items that prevented their use are marked with an "*".

Table 1.2

Streamflow Records on the Teton and Palouse Rivers

Station Number	Years of Record	Remarks
		Teton River Stations
13052200	* 23	* 42,000 acres irrigated
13055000	50	* 58,000 acres irrigated
13055198	* 7	* Paritally regulated and diversions
13055340	* 4	* Records fair, diversions
		Palouse River Stations
13414000	* 24	Records good except ice - fair
13345000	* 18	* Low- and medium-flow regulated by millpond
13340000	* 8	* Small diversion, low-flow regulated
13348000	* 25	Minor diversions for domestic use, regulation by dam and sewage plant
13346100	* 9	* Small diversion, regulation by millpond and sewage disposal

* Item which prevented use of record in this study

CHAPTER 2

STATISTICAL PROPERTIES OF SELECTED FLOW DATA

Once the streamflow records had been selected, various properties of each time series were determined. Discussed in this chapter are the homogeneity, consistency and statistics of the selected streamflow records.

2.1 Monthly Streamflow Listings

A listing of the monthly streamflow records for stations 12413000, 12413500, 13186000, and 13185000 were obtained from the HISARS (28) system at the University of Idaho. HISARS is an acronym for Hyrologic Information Storage and Retrieval System. This system stores hydrologic information on disk which can be retrieved by a data base program.

After examining the listing for each station, it was discovered that the monthly records for water years 1982 and 1983 were missing. In addition, data for water year 1978 was missing for station 12413000. The missing data was obtained from the Water Resources Data for Idaho (36) and added to the HISARS listing of the monthly flows. A listing of the monthly streamflow records as used in this study can be found in Appendix B as Tables B.1 through B.4.

2.2 Consistency and Homogeneity

Before the streamflow records could be used they had to be checked for consistency and homogeneity. Inconsistencies are systematic errors, while nonhomogeneity results from changes in the watershed or climate caused either by humans or natural processes. Inconsistency and nonhomogeneity change the population from which the streamflow measurements (random variables) are taken. Therefore, any inconsistencies or nonhomogeneity must be identified and removed, if possible, from the record before properties of the population are estimated from the sample (streamflow record).

ANNUAL HYDROGRAPHS: Inconsistency and nonhomogeneity often can be identified by a trend or jump in the streamflow record. As a result, several properties of the record such as the mean and standard deviation may be affected (42, 43). A preliminary assessment of the consistency and homogeneity of the records was made by a visual inspection of the annual hydrographs of each stream (Figures 2.1 through 2.4). From this visual inspection, the records did not appear to have any significant trends or jumps, but rather to fluctuate randomly about a constant mean.

HYPOTHESIS TESTS: As a further gross examination to test the observation of no significant trends or jumps, the

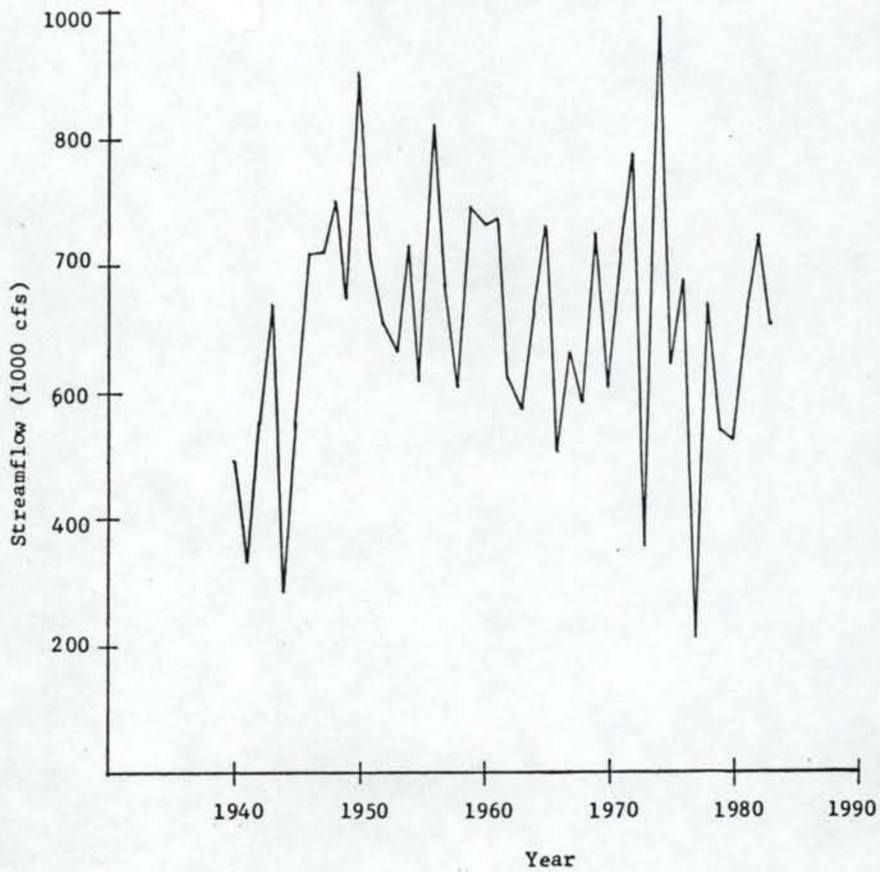


Figure 2.1 Annual Hydrograph at Station 12413000

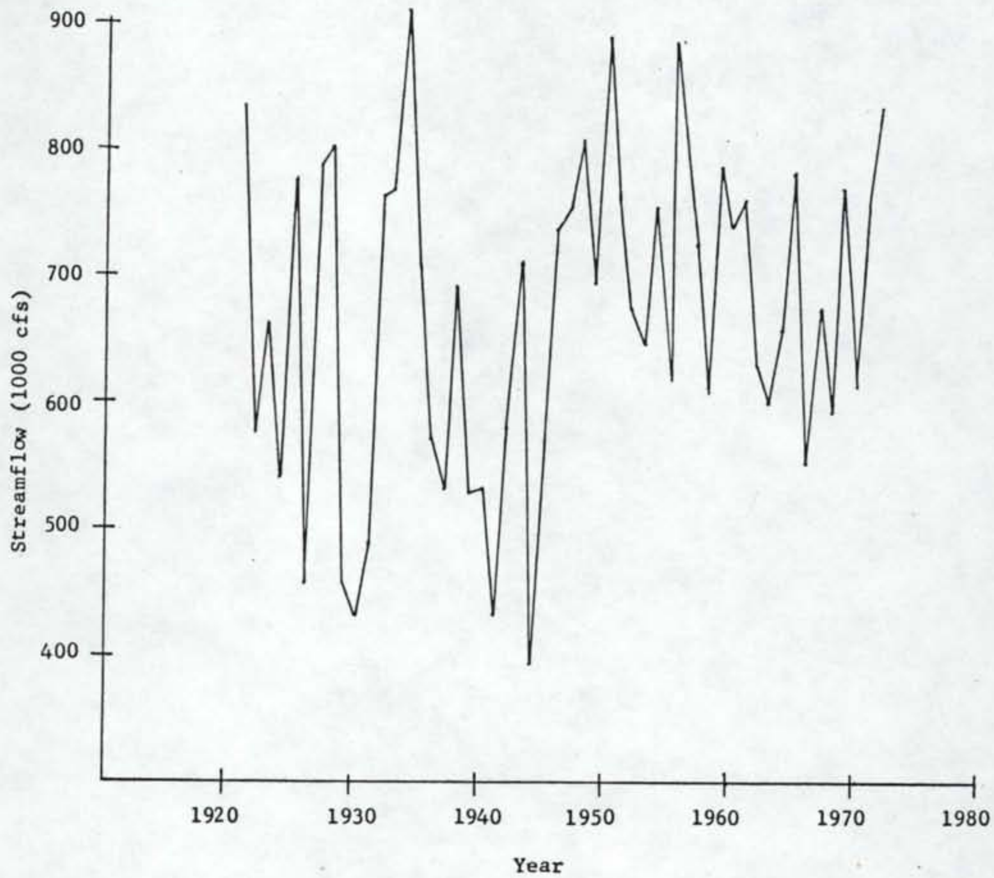


Figure 2.2 Annual Hydrograph at Station 12413500

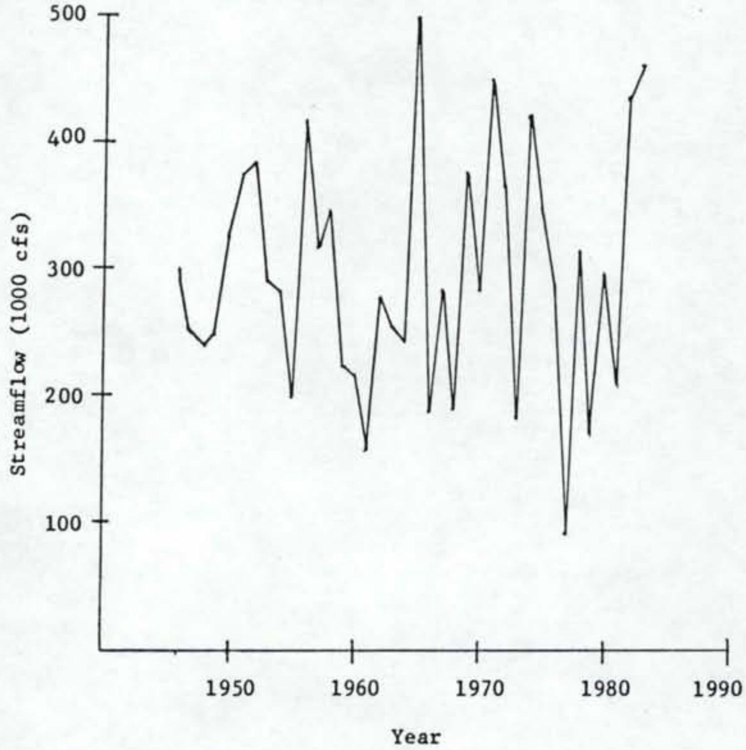


Figure 2.3 Annual Hydrograph at Station 13186000

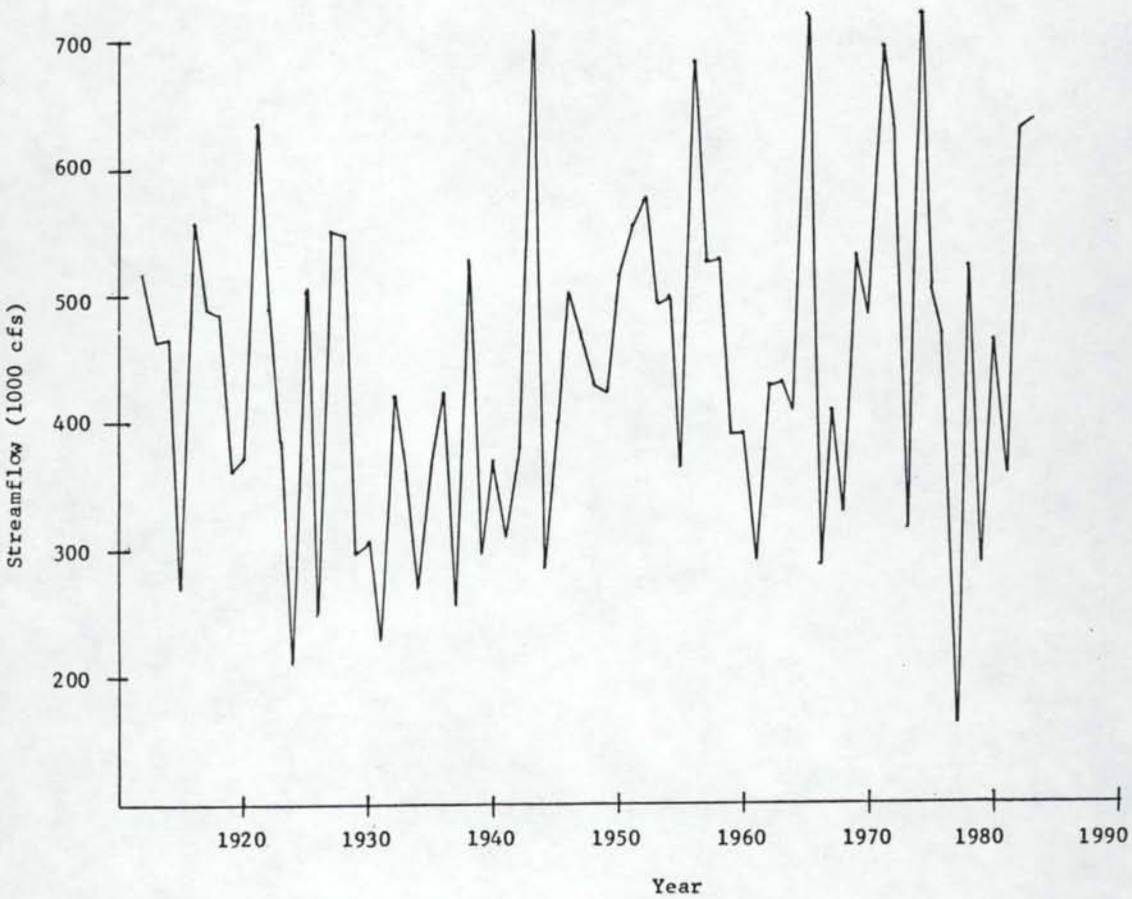


Figure 2.4 Annual Hydrograph at Station 13185000

records were divided in half; thus forming two equal-length subseries. Next, the mean and standard deviation were estimated for each subseries using the method of moments (equations 2.1 and 2.2). The statistics of each subseries are listed in Table 2.1. These statistics were then compared at the 95 percent significance level in order to determine if they were statistically different.

Table 2.1

Annual Statistics of Subseries used to Test for Homogeneity and Consistency of the Streamflow Records

Station	Record Length	Mean cfsd	Standard Deviation cfsd	Record Length	Mean cfsd	Standard Deviation cfsd
12413000	1940-61	723959	193289	1962-83	689749	207594
12413500	1921-46	842218	289517	1947-72	1025536	185774
13186000	1946-64	281212	67339	1965-83	307826	115195
13185000	1912-46	409484	120029	1947-83	476144	134219

In general, annual streamflow series are distributed almost normally (32). This can partly be explained by the central limit theorem since the annual values are the sum of 365 daily values. Consequently, the methods used to compare the means and standard deviations assumed normality of the samples.

Hypothesis tests were used to compare the statistics of the two subseries from a record. The means of the subseries were tested using a t-statistic which varied depending on whether or not the standard deviations of the two subseries were equal (Table 2.2). Therefore, the equality of the standard deviations had to first be tested using an F-statistic (Table 2.2). The results of these tests are presented in Tables 2.3 and 2.4.

From Table 2.2, it can be seen that at the 95% significance level the standard deviations were statistically different at stations 124135000 and 13186000. Likewise, from Table 2.3 it was found that the means at stations 12413500 and 13185000 were statistically different. These results suggested that the statistics were changing with time.

Frequently the differences present in record statistics are due to sampling fluctuations and do not represent true population characteristics. Furthermore, trends or jumps should be supported by physical evidence such as a land use change, flow regulation, diversions or a change in gage location. The records chosen for this study were purposely selected such that they had a minimal amount of diversions and/or regulations above their gages. Also, from the Water Resources Data for Idaho (36) it can be found that the gage locations have, for all practical purposes, remained unchanged over the period of record studied. Therefore, some other physical justification was sought which might explain the statistical differences in the means and standard deviations.

Table 2.2

Hypothesis Tests using t- and F-statistics

F-Statistic

Null Hypothesis: $\sigma_m^2 = \sigma_n^2$
Alternative Hypothesis: $\sigma_m^2 \neq \sigma_n^2$

Test Statistic: $F = s_m^2/s_n^2$

Degrees of Freedom: m-1, n-1 (m = sample size of s_m^2)
(n = sample size of s_n^2)

t-statistic

Null Hypothesis: $\mu_m = \mu_n$
Alternative Hypothesis: $\mu_m \neq \mu_n$

If $\sigma_m^2 = \sigma_n^2$ as determined by F-statistic, then

Test Statistic:

$$t = \frac{\overline{y}_m - \overline{y}_n}{Sp\ (1/m + 1/n)^{1/2}}$$

$$Sp^2 = \frac{(m-1)s_m^2 + (n-1)s_n^2}{m + n - 2}$$

Degrees of Freedom: m + n - 2

If $\sigma_m^2 \neq \sigma_n^2$ as determined by F-statistic, then

Test Statistic:

$$t = \frac{\overline{y}_m - \overline{y}_n}{(s_m^2/m + s_n^2/n)^{1/2}}$$

Degrees of Freedom:

$$\frac{(s_m^2/m + s_n^2/n)^2}{\frac{(s_m^2/m)^2}{m - 1} + \frac{(s_n^2/n)^2}{n - 1}}$$

where:
The subscripts m and n represent two series of lengths m and n, respectively. The Greek letters represent the population statistics while the lowercase letters represent the sample estimates of these statistics.

Table 2.3

Hypothesis Test for Equality of Standard Deviations from Subseries

Station	s_m cfsd	m yrs	s_n cfsd	n yrs	Sample F	95% F	Null Hypoth
12413000	207594	22	193289	22	1.15	2.57	accept
12413500	289517	26	185774	26	2.43	2.36	reject
13186000	115195	19	67339	19	2.90	2.80	reject
13185000	134219	36	120029	36	1.25	2.07	accept

Table 2.4

Hypothesis Tests for Equality of Means from Subseries

Station	$\overline{Y}_m$ cfsd	m yrs	$\overline{Y}_n$ cfsd	n yrs	Sample t	95% t	Null Hypoth
12413000	723958	22	689749	22	0.566	2.02	accept
12413500	842218	26	1025536	26	2.717	2.02	reject
13186000	307826	19	281212	19	0.869	2.04	accept
13185000	409484	36	476144	36	2.221	2.00	reject

By using the longest records available in the Columbia Basin, it has been suggested that the low-flows during the 1930's were the most severe in the last 100 years (17). As a result, the means of the subseries from stations 124135000 and 13185000 which include the 1930's records are statistically different, probably because of the drought conditions that existed during this time. Therefore, this difference was assumed to be caused by sampling fluctuation and not by nonhomogeneity or inconsistency. The differences in the standard deviations were also assumed to be due to sample fluctuation since no physical reason was found to explain their apparent change.

Usually, when a hydrologic series is homogeneous and consistant with respect to its annual mean and standard deviation, the entire series may be considered to be homogeneous and consistent (43). As a result, no tests were considered necessary to determine whether differences existed between other parameters such as the skew, serial correlation coefficients, and the individual monthly means and standard deviations.

2.3 Statistics

Next, the statistics of the records at stations 12413000, 12413500, 13186000, and 13185000 were determined. These statistics are later used to help determine the distribution of the flows, and the type, order and parameters of the needed stochastic streamflow models.

For each monthly and annual series at stations 12413000, 12413500, 13186000, and 13185000 the mean, variance, standard deviation, skew, coefficient of variation, coefficient of skew, and serial correlation coefficients for one to twelve lags were computed by the method of moments. The resulting statistics are summarized in Appendix B as Tables B.5 through B.8. In addition, the correlgrams for stations 12413000 and 13186000 are presented as Figures B.1 through B.26 in Appendix B (A correlogram is a plot of the lag-k serial correlation coefficient versus k). Equations 2.1 through 2.14 provide the basis for estimating the population statistics (Greek letters) from the sample statistics (lowercase letters), along with a brief definition of each statistic.

Annual Statistics

Mean: measure of central tendency

$$\mu = \bar{y} = \frac{1}{n} \sum_{t=1}^{n} y_t \tag{2.1}$$

Variance: measure of spread about mean

$$\sigma^2 = s^2 = \frac{1}{(n-1)} \sum_{t=1}^{n} (y_t - \bar{y})^2 \tag{2.2}$$

Standard Deviation: measure of spread about mean

$$\sigma = s = (s^2)^{1/2} \tag{2.3}$$

Coefficient of Variation: dimensionless measure of spread about mean

$$CV = s/\bar{y} \tag{2.4}$$

Skew: measure of symmetry

$$\alpha = a = \frac{1}{(n-1)(n-2)} \sum_{t=1}^{n} (y_t - \bar{y})^3 \tag{2.5}$$

Coefficient of Skew: dimensionless measure of symmetry

$$\gamma = g = a/s^3 \tag{2.6}$$

Serial Correlation Coefficient: measure of linear dependence between streamflow values separated by k years. The limits of r are 1 to -1 representing perfect linear dependence, with r = 0 representing no linear dependence at all.

$$\rho(k) = r(k) = \frac{1}{(n - k)} \frac{\sum_{t=2}^{n} (y_t - y_t)(y_{t-1} - y_{t-1})}{(s_t)(s_{t-1})} \qquad (2.7)$$

where:

y_t = annual streamflow value at time t
n = number of years of streamflow record

<u>Monthly Statistics (for v=1 to 12)</u>

Mean:

$$\mu_v = \overline{x}_v = \frac{1}{n} \sum_{t=1}^{n} x_{v,t} \qquad (2.8)$$

Variance:

$$\sigma_v^2 = s_v^2 = \frac{1}{(n - 1)} \sum_{t=1}^{n} (x_{v,t} - \overline{x}_v)^2 \qquad (2.9)$$

Standard Deviation:

$$\sigma_v = s_v = (s_v^2)^{1/2} \qquad (2.10)$$

Coefficient of Variation:

$$CV_v = s_v / \overline{x}_v \qquad (2.11)$$

Skew:

$$\alpha_v = a_v = \frac{1}{(n - 1)(n - 2)} \sum_{t=1}^{n} (x_{v,t} - \overline{x}_v)^3 \qquad (2.12)$$

Coefficient of Skew:

$$\gamma_v = g_v = a_v / s_v^3 \qquad (2.13)$$

Serial Correlation Coefficient:

$$\rho(k)_v = r_v(k) = \frac{1}{(n - k)} \frac{\sum_{t=1}^{n} (x_{v,t} - x_v)(x_{v-1,t} - x_{v-1})}{(s_v)(s_{v-1})} \qquad (2.14)$$

where:

x_v = monthly streamflow value during month v
t = year
v = month (when v=1, then v-1 = 12 and t = t-1)
k = time lag (months)

A summary of the annual and extreme monthly statistics of the four selected records is presented in Table 2.4. At this point it would be helpful to remember that stations 12413000 and 13186000 were to be extended by the records at stations 124135000 and 13185000, respectively. Consequently, the record characteristics of the shorter and longer record pairs should be similar to help assure a strong cross correlation between the two records. The statistics in Table 2.4 indicate that these record pairs are similar.

For the most part, Table 2.4 also shows that the streams at stations 12413000 and 13186000 have their extreme events during the same time of the year. The summer flows are the lowest with the least variabilty and highest serial correlation, suggesting that base flow is the major contributor to flow. This seems reasonable since most precipitation on these basins occurs during the winter and spring months producing flows with a larger variability and lower correlation. The largest flows at both stations occur in the spring. Also, it can be noted that overall the monthly flows at station 13186000 are less variable and have higher skews than the flows at station 12413000.

Table 2.4

Annual and Extreme Monthly Statistic

	12413000	12413500	13186000	13185000
Min $\bar{x}$	Sept(8944)	Sept(12610)	Sept(7115)	Sept(11093)
Max $\bar{x}$ (cfs days)	May(173468)	May(228944)	May(85770)	May(118829)
Min CV	Sept(.219)	Aug(.268)	Oct(.185)	Sept(.221)
Max CV	Dec(.857)	Dec(1.073)	July(.558)	July(.540)
Min g	May(-.108)	May(-.179)	May(.221)	June(.066)
Max g	Jan(2.671)	Jan(3.005)	Dec(3.411)	Dec(2.931)
Min r(1)	April(.003)	April(.094)	Dec(.220)	Dec(.327)
Max r(1)	Aug(.852)	Aug(.909)	Aug(.911)	Aug(.907)
Ann $\bar{y}$ (cfs days)	706854	933877	294519	442814
Ann CV	.281	.276	.319	.295
Ann g	-.273	-.262	.229	.227
Ann r(1)	-.056	.178	-.048	-.026

CHAPTER 3

MODELING CONCEPTS AND DATA EXTENSION

One objective of this study was to examine the effect that data extension has on disaggregation model parameters and drought characteristics generated by the models. Therefore, the monthly streamflow records at stations 12413000 and 13186000 were extended to the same length as the records at stations 12413500 and 13185000, respectively. The period of record available at each station is shown in Table 3.1.

Table 3.1

Length of Historical Streamflow Record

Station	Shorter Record	Station	Longer Record
12413000	10/1939 - 9/1983	12413500	8/1920 - 9/1972
13186000	5/1945 - 9/1983	13185000	4/1911 - 9/1983

This chapter reviews several modeling principles that were used throughout this study, discusses the constraints associated with four models which were considered for data extension, examines the residuals from each model, describes the methodology actually selected to extend the records at stations 12413000 and 13186000, and then finally discusses the results.

3.1 Preliminary Modeling Concepts and Analysis

CORRELATION COEFFICIENTS: Two types of correlation coefficients were examined while trying to extend the shorter records of stations 12413000 and 13186000: serial and cross. A schematic illustration of serial and cross correlation is shown in Figure 3.1 along with the symbols that are used throughout this chapter.

The serial correlation coefficient measures the degree of linear dependence between sequential streamflow values at the same station separated by a time lag of k. The lag-one serial correlation coefficients of each monthly flow series for the period of overlapping record were determined by equation 2.12 and are presented in Appendix C (Table C.1).

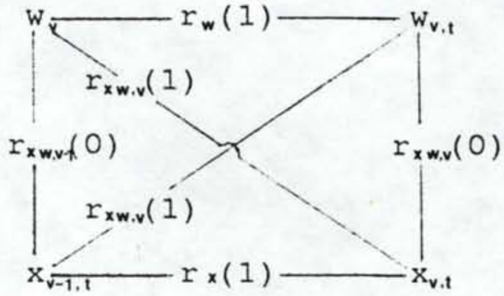
The cross correlation coefficient measures the degree of linear dependence between streamflow values at two different stations separated by a time lag of k. For the purposes of data extension, only the lag-zero (k=0) cross correlation coefficients were needed. The lag-zero cross correlation coefficients between corresponding monthly values for each set of stations were computed using equation 3.1 and are presented in Appendix C (Table C.1).

Figure 3.1

Illustration of Serial and Cross Correlation

Key:

Subordinate:



where:

- w = streamflow value from station with longer record (key)
- x = streamflow value from station with shorter record (subordinate)
- $r_w(1)$ = lag-one serial correlation coefficient at station w.
- $r_x(1)$ = lag-one serial correlation coefficient at station x
- $r_{xw}(0)$ = lag-zero cross correlation coefficient
- $r_{xw}(1)$ = lag-one cross correlation coefficient
- v = month (if v=1 then v-1 = 12 and t = t - 1)
- t = year

NOTE: Hereafter, for clarity the lag number corresponding to the correlation coefficients will be dropped. In this chapter it will be assumed that all serial correlation coefficients have a time lag of one and all cross correlation coefficients have a lag of zero, unless specifically stated otherwise.

$$r_{xw,v} = \frac{\sum_{t=1}^{n_o} (x_{v,t} - \bar{x}_v)(w_{v,t} - \bar{w}_v)}{(n_o - 1) s_{x,v} \; s_{w,v}} \tag{3.1}$$

where:

r_{xw} = monthly lag-zero cross correlation coefficient between stations x and w

x, w = monthly streamflow value at station x and w

$\bar{x}, \bar{w}$ = mean monthly streamflow at station x and w (only for period of overlapping record)

s_x, s_w = standard deviation of monthly streamflow at station x and w (only for period of overlapping record)

n_o = number of overlapping record years

STOCHASTIC STREAMFLOW MODELS: Generally, a streamflow series is considered to be composed of two distinct parts: deterministic relationships and a random component. Streamflow models can be developed which contain either one or both of these components.

Streamflow models based only on deterministic relationships will always predict the same streamflow value when given a particular set of independent variable values. In other words, deterministic relationships predict one unique streamflow value for every set of unique input variables. The deterministic relationships which often form a part of stochastic streamflow models are usually based on serial and/or cross correlations.

If deterministic relationships are tested by using historical streamflow values, the streamflow value predicted by the relationships will often differ from the actual historical streamflow values. This implies that a pure deterministic model does not completely describe the streamflow process; there still remains a part of the process which is not being accounted for by the model. The numeric difference between a streamflow value predicted by deterministic relationships and the corresponding historical streamflow value is known as a residual. Because of these residual values, often a pure deterministic model will underestimate the variablity found in the historical record. Consequently, stochastic streamflow models include a random component, in addition to a deterministic component, which models the residuals, and thereby preserves the total variablity of the historical record.

A basic concept in stochastic modeling is to structure the deterministic component is such a manner that the resulting residual series is temporally independent, homoscedastic (constant variance), and, if possible, normally distributed. A residual series with these properties can be modeled easily as a random process, and included as a random component in the stochastic model. Therefore, the above assumptions must be checked to ensure the adequacy of the models. Following is a discussion of two simple methods that are used in this study to test the normality and independence of the residuals. (It is assumed that the residuals are homoscedastic).

NORMALITY: Phien, Sunchindah, and Patnaik (30) reviewed three techniques for testing the hypothesis of normality. They found that though some methods were statistically more sound than others, the methods did not differ greatly in their results. One of the tests studied was to determine if the coefficient of skew for any sample is statistically equal to zero. A normal distribution is symmetrical with zero skew, and consequently, if a sample or data set coefficient of skew is statistically equal to zero, the hypothesis of normality is accepted.

The 95% confidence limits for the coefficient of skew equal to zero can be calculated from equation 3.2. If the sample skew coefficient exceeds this calculated value, then the series is not considered to be normally distributed.

$$g(95\%) = \pm 1.96(6/n)^{1/2} \tag{3.2}$$

where:

$g(95\%)$ = 95% critical skew coefficient for hypothesis that $g = 0$.

n = sample size

TIME INDEPENDENCE: If the serial correlation coefficients are not statistically different from zero, then the sequential values in a series can be considered to be independent. The critical 95% serial correlation coefficients can be calculated from an equation given by Anderson (2) for the probability limits of the serial correlation coefficients of an independent series. Anderson's equation is presented as equation 3.3.

$$r(k,95\%) = \frac{-1 \pm 1.96(n - k - 1)^{1/2}}{n - k} \qquad (3.3)$$

where:

$r(k,95\%)$ = 95% critical serial correlation coefficient for hypothesis that $r(k) = 0$
k = lag time
n = sample size

PARAMETER PARSIMONY: As the number of correlation relationships increases, so does the number of model parameters which must be estimated from the historical series. The historical series is considered to be only one sample from a larger population. As a result, if too many parameters are used, the sampling variabiltiy of the historical series is modeled rather than the actual population characteristics. Consequently, a principle known as parameter parsimony has evolved. As one quantitative measure of this concept, the following equation is often used (32):

$$\delta = n/n_p \qquad (3.4)$$

where:

δ = index of parameter parsimony
n = sample size
n_p = number of parameters estimated from the sample

with:

$1 \leq \delta < 3$	foolish
$3 \leq \delta < 5$	poor
$5 \leq \delta < 10$	fair
$10 \leq \delta < 20$	good
$20 \leq \delta$	excellent

The combined monthly sample size available at stations 12413000 and 12413500 (overlapping periods of record only) was 792, while the combined monthly sample size at stations 13186000 and 13185000 was 922. Since monthly records were used, parameters were estimated for each month individually. Most of the models examined required seven parameters for each set of monthly values: mean, standard deviation, lag-one serial correlation coefficient for each record and the lag-zero cross correlation between the two records. Hence, an index of parameter parsimony of 9.4 and 11 resulted for the two sets of stations. Though lower than 15, these values were considered to be acceptable since all of the estimated parameters were judged to be significant in terms of describing each monthly streamflow series.

STATISTICS: Besides the correlation coefficients, the mean and standard deviation of each monthly series were needed for data extension. Therefore, these statistics along with the coefficient of skew were calculated from equations 2.8, 2.10, and 2.13, respectively, for the period of overlapping record at each station. The results of these calcuations are presented in Appendix C as Table C.2.

3.2 Multivariate Model Constraints

Following is a discussion of the parameter constraints associated with four multivariate models. Multivariate models consider not only the serial correlation of records, but also the cross correlations between two or more different records. The constraints of each model were considered first, in order to determine if the model could be used for each set of streamflow records.

MODEL 1: Simple linear regression between the two streamflow records was the first model considered. Essentially this model considers only the lag-zero cross correlation coefficient, ignoring any serial correlation coefficients. No special constraints are associated with this model.

MODEL 2: In 1964 Fiering (9) presented a model which was designed to preserve the lag-one serial and lag-zero cross correlation coefficients of two stations. However, subsequent examination of this model by others (18, 20) showed that these correlation coefficients were not preserved unless one the of the following conditions was true (20).

$$r_w = 0$$
$$r_{xw} = 0 \qquad (3.5)$$
$$r_x = r_w r_{xw}^2$$

Both pairs of streamflow records had monthly values of r_w and r_{xw} significantly different from zero. Thus, the third constraint had to be considered for the monthly streamflow records. The results of examining the third constraint are shown in Table 3.2.

Based on the constraint: $r_x = r_w r_{xw}^2$, it was decided not to use the Fiering model for stations 13186000 and 13185000 since discrepencies of up to 50 percent were present (Table 3.2). On the other hand, stations 12413000 and 124135000 seemed to meet this constraint.

MODEL 3: In 1977 Lawrance (20) presented a modification of the Fiering model which preserves the lag-one serial and lag-zero cross correlation coefficients. In order for this model to be valid, the following constraint must be met:

$$s_e^2 > 0$$

$$s_e^2 = 1 - A_L^2 - B_L^2 - 2A_L B_L r_w r_{xw} \qquad (3.6)$$

$$A_L = \frac{r_x - r_w r_{xw}^2}{1 - r_w^2 \, r_{xw}^2} \qquad (3.7)$$

$$B_L = \frac{r_{xw}(1 - r_w r_x)}{1 - r_w^2 \, r_{xw}^2} \qquad (3.8)$$

Table 3.2

Evaluation of Constraint for Fiering Model

Period	12413000 & 13135000			13186000 & 13185000		
	$r_w r_{xw}^2$	r_x	Δr_x	$r_w r^2_w$	r_x	Δr_x
October	.5776	.5502	.0274	.5180	.7856	.2676
November	.6911	.6870	.0041	.2689	.5381	.2692
December	.6335	.6096	.0239	.2253	.2203	.0050
January	.3476	.3472	.0004	.5040	.7972	.2932
Feburary	.2665	.2787	.0122	.3082	.5853	.2771
March	.2789	.2827	.0038	.2552	.3600	.1048
April	-.0013	-.0513	.0500	.3584	.3458	.0126
May	.3673	.3809	.0136	.5335	.6365	.1030
June	.6847	.7239	.0392	.5657	.6853	.1196
July	.8209	.8127	.0082	.8189	.9055	.0866
August	.8284	.8616	.0332	.8165	.9110	.0945
September	.5142	.5106	.0036	.7873	.8991	.1118

This constraint was checked by calculating s_e^2 for each month and the results of these calculations are shown in Table 3.3. As can be seen from Table 3.3, s_e^2 was greater than zero for all the months from both sets of records.

Table 3.3

Evaluation of Constraint for Lawrance Model

Period	12413000 & 13135000			13186000 & 13185000		
	A_L	B_L	s_e^2	A_L	B_L	s_e^2
October	-.0416	1.0151	.01698	.3903	.7051	.04194
November	-.0081	.9965	.01808	.2970	.7894	.14547
December	-.0401	1.0230	.00403	-.0053	.9311	.13545
January	-.0004	.9942	.01177	.4308	.6482	.07857
Feburary	.0131	.9924	.00802	.3125	.8111	.07397
March	.0042	.9939	.00476	.1131	.9147	.09474
April	-.0500	.9951	.00708	-.0145	.9839	.09216
May	.0158	.9885	.01115	.1510	.8607	.08972
June	.0747	.9415	.01102	.1853	.8396	.07554
July	-.0270	1.0055	.03349	.3016	.7152	.03317
August	.1256	.8580	.06306	.3519	.6537	.05529
September	-.0050	.9795	.04565	.3407	.6812	.03935

Table 3.4

Evaluation of Constraint for Yevjevich Model

Period	12413000 & 13135000 constriant $\geq$	r_{xw}^2	13186000 & 13185000 constraint $\geq$	r_{xw}^2
October	0.9968	0.9819	0.8833	0.8536
November	0.9989	0.9819	0.9448	0.7746
December	0.9980	0.9950	0.9982	0.8645
January	1.0000	0.9882	0.8910	0.7951
Feburary	1.9999	0.9918	0.9228	0.8394
March	1.0000	0.9902	0.9931	0.8934
April	0.9975	0.9904	0.9989	0.9576
May	1.9999	0.9886	0.9958	0.8947
June	0.9965	0.9860	0.9896	0.9023
July	0.9859	0.9663	0.9727	0.9407
August	0.9873	0.9328	0.9932	0.9114
September	0.9985	0.9543	0.9614	0.9226

MODEL 4: The fourth model examined was developed by Yevjevich in 1973 (43). His model was also designed to preserve the lag-zero cross and lag-one serial correlation coefficients of two records. The correlation constraint of his model is:

$$r_{xw}^2 \leq \frac{(1 - r_x^2)(1 - r_w^2)}{(1 - r_x r_w)^2} \tag{3.9}$$

This constraint was examined and the results are presented in Table 3.4. From Table 3.4, it can be seen that all of the monthly streamflow data sets meet the constraint of the Yevjevich model.

3.3 Residuals of Multivariate Models

As mentioned earlier, the residuals from a stochastic model should be temporally independent, homoscedastic, and, if possible, normally distributed. Generally, the independence of the residuals is the most critical of these properties, because if the residuals are not independent, it suggests that the deterministic component of the model is inadequate.

Less critical is the property of normality, since there are several ways to handle the residuals if they are not normally distributed. Formulas are available which relate the skewness of the residuals to the skewness of the historical series. There is a disadvantage to this method however: it requires the estimation of another parameter, the skew coefficient, which requires large sample sizes (>70) for accurate estimation. Another method that can be used is to reduce the skew of the residuals by transforming the original data. The problem with this method is that sometimes modeling the transformed data will not preserve the statistics of the original data.

Consequently, the residuals from each model were calculated in order to determine their characteristics. These calculations were performed for the period of common record between stations 12413000 and 12413500 by determining the difference between the historical flows and those predicted by the model. Likewise, the same was done for the overlapping record at stations 13186000 and 13185000.

MODEL 1: Simple linear regression was the first model tried on both sets of stations. The reasons for this were: 1) it is the simplest model to apply, 2) of the four models examined, it required the least number of parameters, 3) it was hoped that since the lag-one cross correlation coefficients were so high, the serial correlation coefficients would be indirectly preserved, and 4) the results of this model could serve as a base from which the other models could be compared, to determine if substantially better results were obtained by adding more parameters.

Simple linear regression can be expressed as equation 3.10. The parameters (A_R and B_R) were calculated for each month and are listed in Appendix C as Table C.3.

$$x_{v,t} = A_{R,v} + B_{R,v} \, w_{v,t} + e_{v,t} \tag{3.10}$$

$$B_{R,v} = \frac{\mathrm{cov}(x_v, w_v)}{s^2_{w,v}} \tag{3.11}$$

$$A_{R,v} = \overline{x}_v - B_{R,v}\overline{w}_v \tag{3.12}$$

where:

w = monthly streamflow value at key station
x = monthly streamflow value at subordinate station
e = residual series
n_o = number of overlapping record years

$$\mathrm{cov}(x_v, w_v) = (x_{v,t} - x_v)(w_{v,t} - w_v)/(n_o - 1)$$

Using the parameter estimates as calculated from equations 3.11 and 3.12, the residuals of the linear regression model were determined using equation 3.13. The statistics of the residuals from this model are summarized in Table 3.5, along with the 95% significance levels for the skew coefficient and lag-one serial correlation coefficient equaling zero (equations 3.2 and 3.3, respectively).

$$e_{v,t} = x_{v,t} - B_{R,v}\, w_{v,t} - A_{R,v} \qquad (3.13)$$

After examining the residuals from the linear regression model, it was discovered that the residuals, for the most part, did not have a lag-one serial correlation coefficient statistically equal to zero. Meanwhile, the skew coefficients of the residuals were only significant for less than half of the months. In order to try and improve the statistics of the residuals the other multivariate models were considered.

The other three multivariate models were developed for the simultaneous generation of streamflow values at more than one station. They all assume that the key station record is generated first by an AR(1) model (AR(1) models are discussed in Chapter 5), and then simultaneous streamflow values are generated at the subordinate station using the generated values at the key station. The

Table 3.5

Statistics of Residuals from Linear Regression

Stations 12413000 and 12413500

Period	Mean (cfsd)	Standard Deviation (cfsd)	Coefficient of Skew calc.	Coefficient of Skew 95%	Lag one serial corelation coef calc.	Lag one serial corelation coef 95%
October	.002	1215	.018	.701	**.428	.310
November	-.001	2924	* -.779	-.701	**.316	.310
December	.022	2984	.093	.701	**.334	.310
January	-.009	2599	-.685	-.701	.218	.310
Feburary	.005	3838	-.165	-.701	**.364	.310
March	-.047	4358	.194	.701	**.369	.310
April	-.016	5973	.051	.701	.077	.310
May	.071	7633	* .764	.701	.246	.310
June	.044	3543	* 1.269	.701	**.478	.310
July	.001	1388	-.148	-.701	**.491	.310
August	-.005	714	.471	.701	**.487	.310
September	.003	443	-.212	-.701	**.692	.310

Stations 13186000 and 13185000

Period	Mean (cfsd)	Standard Deviation (cfsd)	Coefficient of Skew calc.	Coefficient of Skew 95%	Lag one serial corelation coef calc.	Lag one serial corelation coef 95%
October	.003	537	.272	.654	**.447	.291
November	.002	680	*1.278	.654	**.743	.291
December	.003	1005	-.214	-.654	**.622	.291
January	.001	795	*1.190	.654	**.809	.291
Feburary	-.001	661	* .586	.654	**.579	.291
March	-.008	1248	*1.212	.654	**.359	.291
April	-.012	3760	.465	.654	**.303	.291
May	-.032	10719	.306	.654	**.476	.291
June	-.007	10115	.041	.645	**.764	.287
July	.005	3563	-.113	-.645	**.854	.287
August	.002	983	-.230	-.645	**.711	.287
September	.002	491	.142	.645	**.804	.287

* Skew coefficients $\neq 0$
** Lag one serial correlation coefficient > 0

streamflow records at stations 12413000 and 13186000 (subordinate) were to be extended only to the same period of time as represented by the longer historical record at the nearby stations (key). As a result, the key station streamflow values would not be generated from the AR(1) model but instead the actual historical values would be used.

The next model tried was MODEL 3. MODEL 2 was not used because it is only an approximate model, and MODEL 3 is a modification of MODEL 2 designed to take care of deficiencies in MODEL 2.

MODEL 3: The Lawrance model (20) as presented in 1977 is written as equations 3.14 and 3.15. Equation 3.14 represents the generation of the key station streamflow values by an AR(1) model.

$$W_{v,t} = r_{w,v} W_{v-1,t} + (1 - r_{w,v}^2)^{1/2} \lambda_{v,t1} \qquad (3.14)$$

$$X_{v,t} = A_{L,v} X_{v-1,t} + B_{L,v} W_{v,t} + s_{e,v} \lambda_{v,t2} \qquad (3.15)$$

where:

W = standardized flow at the key station $(w_{v,t} - \overline{w}_v)/s_{w,v}$

X = standardized flow at the subordinate station $(x_{v,t} - \overline{x}_v)/s_{x,v}$

λ = random deviate

The parameters (A_L, B_L, and s_e^2) were previously determined from equations 3.6, 3.7, and 3.8 respectively, and are listed in Table 3.3. By using these parameters, the residuals for the Lawrance model were found from equation

3.16, and the statistics of these residuals are summarized in Table 3.6.

$$E_{v,t} = X_{v,t} - A_{L,v} X_{v-1,t} - B_{L,v} W_{v,t} \qquad (3.16)$$

$$e_{v,t} = x_{v,t} + E_{v,t} \, s_{x,v}$$

where:

E = standardized residual
e = residual
X = historical standardized flow at subordinate station, $(x_{v,t} - \bar{x}_v)/s_{x,v}$
W = historical standardized flow at key station $(w_{v,t} - \bar{w}_v)/s_{w,v}$

If a comparision is made between the statistics of the residuals from the Lawrance model (Table 3.6) and those from the linear regression model (Table 3.5), it can be seen that little improvement is gained by the extra parameters of the Lawrance model for stations 12413000 and 124135000. However, the time dependency is reduced when the Lawrance model is used for stations 13186000 and 13185000. In an effort to improve upon the linear regression and the Lawrance model, MODEL 4 was tried.

Table 3.6

Statistics of Residuals from Lawrance Model

Stations 12413000 and 12413500

Period	Mean (cfsd)	Standard Deviation (cfsd)	Coefficient of Skew calc.	Coefficient of Skew 95%	Lag one serial corelation coef calc.	Lag one serial corelation coef 95%
October	-23.281	1170	-.035	-.701	**.504	.310
November	-.030	2938	* -.884	-.701	**.342	.310
December	.030	3564	.449	.701	**.419	.310
January	.000	2599	-.682	-.701	.241	.310
Feburary	.000	3747	-.398	-.701	**.353	.310
March	.091	4338	.157	.701	**.398	.310
April	.061	6667	.153	.701	.115	.310
May	-.030	7808	* .875	.701	.087	.310
June	.000	3918	* 1.343	.701	**.544	.310
July	.091	1398	-.065	-.701	**.450	.310
August	-.061	697	.541	.701	**.417	.310
September	.000	444	-.180	-.701	**.746	.310

Stations 13186000 and 13185000

Period	Mean (cfsd)	Standard Deviation (cfsd)	Coefficient of Skew calc.	Coefficient of Skew 95%	Lag one serial corelation coef calc.	Lag one serial corelation coef 95%
October	27.711	302	.171	.654	.054	.291
November	.053	542	* .944	.654	**.390	.291
December	-.026	1010	-.243	-.654	**.505	.291
January	.053	541	-.439	-.654	**.689	.291
Feburary	.000	522	* .702	.654	.245	.291
March	.026	1167	*1.528	.654	.017	.291
April	-.053	3770	.413	.654	**.357	.291
May	403.763	9531	.278	.654	**.365	.291
June	.051	8906	.250	.645	**.727	.287
July	.000	3440	*-.724	-.645	**.730	.287
August	-.026	903	-.335	-.645	**.479	.287
September	.051	359	-.036	-.645	**.531	.287

* Skew coefficients ≠ 0

** Lag one serial correlation coefficient > 0

MODEL 4: In 1973, Yevjevich (43) presented the following model:

$$W_{v,t} = r_{w,v} W_{v-1,t} + (1 - r_{w,v}^2)^{1/2} \lambda_{v,t1} \qquad (3.17)$$

$$X_{v,t} = r_{x,v} X_{v-1,t} + A_{y,v} \lambda_{v,t1} + B_{y,v} \lambda_{v,t2} \qquad (3.18)$$

$$A_{y,v} = \frac{r_{xw,v}(1 - r_{w,v} r_{x,v})}{(1 - r_{w,v}^2)^{1/2}} \qquad (3.19)$$

$$B_{y,v}^2 = 1 - r_{x,v}^2 - \frac{r_{xw,v}^2 (1 - r_{x,v} r_{w,v})^2}{(1 - r_{w,v}^2)} \qquad (3.20)$$

,

where:

W = standardized flow at the key station $(w_{v,t} - \bar{w}_v)/s_{w,v}$

X = standardized flow at the subordinate station $(x_{v,t} - \bar{x}_v)/s_{x,v}$

λ = random deviate

Equation 3.17 again represents an AR(1) model for the generation of streamflow values at the key station. The parmeters (A_y and B_y) for this model were estimated from equations 3.19 and 3.20 and the results are presented in Appendix C as Table C.4.

The residuals for the Yevjevich model were calculated using equations 3.21 and 3.22. The measured historical values at the key station were used in equation 3.21 in order to determine the value of $\lambda_{v,t1}$ needed in equation 3.22. The statistics of the resulting residuals are summarized in Table 3.7.

$$\lambda_{v,t1} = \frac{W_{v,t} - r_{w,v} W_{v-1,t}}{(1 - r^2_{w,v})^{1/2}} \quad (3.21)$$

$$E_{v,t} = X_{v,t} - r_{x,v} X_{v-1,t} - A_{y,v} \lambda_{v,t1} \quad (3.22)$$

$$e_{v,t} = \overline{x}_v + E_{v,t} s_{xv}$$

where:

E = standardized residual
e = residual
W = historical standardized flow at key station $(w_{vt} - \overline{w}_v)/s_{wv}$
X = historical standardized flow at subordinate station, $(x_{vt} - \overline{x}_v)/s_{xv}$

A considerable improvement in the statistics of the residuals is seen for the Yevjevich model. Both sets of records show a reduction in the time dependence and, for most of the months, a reduction in the skew coefficient and standard deviation of the residuals. Therefore, the Yevjevich model was chosen to extend the record at stations 12413000 and 13186000.

The better performance of the Yevjevich model seems to be due to the fact that in equation 3.18, the lag-one serial correlation coefficient at the subordinate station is used directly to relate successive monthly flows at the subordinate station, whereas the Lawrance model uses a parameter which is only a function of the lag-one serial correlation coefficient, and the linear regression model does not even consider serial correlation. The lag-one serial correlation coefficient is the best moment estimator

Table 3.7

Statistics of Residuals from Yevjevich Model

Stations 12413000 and 12413500

Period	Mean (cfsd)	Standard Deviation (cfsd)	Coefficient of Skew calc.	Coefficient of Skew 95%	Lag one serial corelation coef calc.	Lag one serial corelation coef 95%
October	4.812	1139	.067	.701	-.230	-.372
November	-.030	3005	.386	.701	-.327	-.372
December	-.061	4032	.620	.701	-.302	-.372
January	.030	2533	-.588	-.701	.192	.310
Feburary	-.091	3515	-.498	-.701	.032	.310
March	.091	4073	.280	.701	.152	.310
April	-.030	6667	.152	.701	.123	.310
May	.030	7550	* .828	.701	-.217	-.372
June	-.061	3235	.015	.701	.087	.310
July	.061	1182	.067	.701	.146	.310
August	.000	640	.100	.701	.157	.310
September	.000	316	-.137	-.701	.226	.310

Stations 13186000 and 13185000

Period	Mean (cfsd)	Standard Deviation (cfsd)	Coefficient of Skew calc.	Coefficient of Skew 95%	Lag one serial corelation coef calc.	Lag one serial corelation coef 95%
October	27.300	300	-.359	-.654	-.090	-.345
November	.053	454	*1.627	.654	.181	.291
December	.000	869	.118	.654	.280	.291
January	.000	360	.281	.654	**.328	.291
Feburary	-.026	472	* .729	.654	-.039	-.345
March	.026	1135	*1.358	.654	-.202	-.345
April	-.026	3697	.280	.654	-.201	-.345
May	388.158	8858	.078	.654	.036	.291
June	-.051	6283	-.642	-.645	**.294	.287
July	-.051	1944	-.630	-.645	-.074	-.340
August	-.026	674	*-1.850	-.645	.138	.287
September	-.051	251	.058	.645	-.207	-.340

* Skew coefficients ≠ 0

** Lag one serial correlation coefficient > 0

of the linear dependence that exists between sucessive elements in a time series, and thus, should best reduce the dependence of the resulting residuals at the subordinate station.

3.4 Data Extension

In order to use Yevjevich's model (equations 3.17 and 3.18), initial values of $W_{v-1,t}$, $W_{v,t}$, and $X_{v-1,t}$ are needed. Until this time, all of the analyses of the multivariate models were done by going forward in time. However, in order to take advantage of known starting values from the historical records, the model would have to be applied going backwards in time. Therefore, it was attempted to apply Yevjevich's model in a negative time sense.

However, a problem with parameter estimation arose, since the parameter "B_y^2" became negative for one month in each set of records. Apparently, this problem was caused by the fact that a strong cross correlation between two sucessive months forces the lag-one serial correlation coefficients of the corresponding months to not differ by more than a certain amount. This constraint was met going forward in time, but not going backwards. Therefore, the model was applied in the positive time sense after first estimating the intial value of $X_{v-1,t}$.

In order to determine a reasonable starting value, linear regression was used. The earliest key station monthly streamflow value available became the initial $W_{v-1,t}$. This $W_{v-1,t}$ value was then used in equation 3.10 to generate the initial $X_{v-1,t}$ value. A random normal deviate was used to preserve the variability of the series, since the residuals from the linear regression model for the beginning months had a skew coefficient which was statistically equal to zero. As a result, the earliest streamflow value in the extended series was generated by a different model (linear regression) than the rest of the extended record (Yevjevich model). However, the effects of this initial starting value soon become negligible after several months due to the inclusion of a random component and the cross correlation.

In summary, the equations used to extend the records at stations 12413000 and 13186000 are listed below along with a brief description of the proceedure used.

1) The initial $X_{v-1,t}$ streamflow value to be used in equation 3.18 was found from equation 3.10. (Then standardized result from equation 3.10).

$$x_{v-1,t} = A_{R,v-1} + B_{R,v-1} W_{v-1,t} + s_{v-1,t} (1 - r^2_{xw,v-1})^{1/2} \lambda_t \qquad (3.10)$$

2) The value of $\lambda_{v,t1}$ to be used in equation 3.18 was calculated from the observed historical record at the key station for each month.

$$\lambda_{v\,t1} = \frac{W_{v,t} - r_{w,v} W_{v-1,t}}{(1 - r^2_{w,v})^{1/2}} \qquad (3.21)$$

3) The skew coefficient of the residuals for each month was examined. If the skew was statistically equal to zero, then $\lambda_{v,t2}$ was taken as a random standard deviate from a normal distribution. However, if the skew was not statistically equal to zero, then the standard normal deviate was transformed

into a standard gamma deviate using the Wilson-Hilfery (32) transform which is shown as equation 3.23. The random deviates were then used in equation 3.18.

$$\lambda_{v,t2} = \frac{2}{g_e}\left[\left(1 + \frac{g_e(\lambda_n)}{6} - \frac{g_e^2}{36}\right)^3\right] - \frac{2}{g_e} \qquad (3.23)$$

where:

λ_n = random deviate from the standard normal distribution

g_e = skew coefficient of residuals

4) The extended record at the subordinate station was generated using equation 3.18 for all months except the first value which was calculated from equation 3.10.

$$X_{v,t} = r_{x,v} X_{v-1,t} + A_{y,v} \lambda_{v,t1} + B_{y,v} \lambda_{v,t2} \qquad (3.18)$$

The extended records as generated by equations 3.10, 3.18, 3.21 and 3.27 are presented in Appendix C as Tables C.5 and C.6. The statistics of the extended portion of record along with the statistics of the overlapping portion of record at stations 12413000 and 13186000 are summarized in Table 3.8. The annual streamflow values of the extended period were assumed to equal the sum of the generated monthly streamflow values for that year.

Table 3.8

Statistics of Subordinate Stations for Extended and Historical Portions of Record (Streamflow in cfs days)

Station 12413000

	Extended (8/20 - 9/39)				Historical (10/39 - 9/72)			
Mon	Mean	Stand Dev	Skew Coef	Serial Cor Cf	Mean	Stand Dev	Skew Coef	Serial Cor Cf
Oct	13066	9970	2.529	.918	13845	9049	1.684	.550
Nov	29040	34656	2.838	.884	29304	21760	1.011	.687
Dec	50268	69940	3.530	.336	48585	42032	1.397	.610
Jan	42183	48734	2.199	.827	41878	23937	.900	.347
Feb	43568	39300	1.237	.432	57816	42286	1.412	.279
Mar	77650	40238	.690	.624	74983	44228	2.052	.283
Apr	181900	56144	.500	.224	172436	60972	.174	-.051
May	154882	65354	-.347	.437	181806	71385	-.207	.381
Jun	47071	33596	1.480	.731	61055	29964	1.177	.724
Jul	15207	4894	.373	.908	20575	7557	.814	.813
Aug	8878	1952	.124	.884	10683	2752	.850	.862
Sep	8349	2693	2.253	.336	8994	2072	1.354	.511
Yr	671842	222967	.332	.092	721960	177501	-.519	.222

Station 13186000

	Extended (4/11 - 4/45)				Historical (5/45 - 9/83)			
Mon	Mean	Stand Dev	Skew Coef	Serial Cor Cf	Mean	Stand Dev	Skew Coef	Serial Cor Cf
Oct	6794	1373	.417	.945	7596	1403	.381	.786
Nov	7061	1903	2.954	.666	7629	1433	1.041	.538
Dec	7070	1704	.800	.369	7945	2731	3.411	.220
Jan	6933	1079	.651	.760	7755	1756	1.674	.797
Feb	6249	1043	.158	.713	7402	1651	1.361	.585
Mar	10053	3367	.718	.585	11606	3823	1.133	.360
Apr	33612	17714	2.441	.498	38289	18269	.565	.346
May	72897	31011	.706	.378	85770	33029	.221	.637
Jun	63215	39633	.098	.661	76265	32368	.351	.685
Jul	20235	14160	.726	.860	26255	14642	.990	.905
Aug	7622	3134	.408	.926	9488	3303	1.044	.911
Sep	5981	1753	.377	.933	7115	1767	.706	.899
Yr	247348	93653	.331	-.177	294519	94040	.229	-.048

Some of the statistics of the extended record appear to be substantially different from those of the historical record on the same station. However, this was to be expected since there was an extreme drought during the 1930's. Therefore, the extended periods of record which include this time of drought would reflect this event. Yet, to make sure that the differences were realistic, the statistics for the identical periods of time on the key (longer record) stations were compared to see if they followed the same pattern as seen in Table 3.8. The comparisions of the key station statistics are shown in Table 3.9.

The key and subordinate records did seem to follow the same trends in terms of their statistics. This observation was then checked using hypothesis tests at a 95% significance level. The standard deviations, and means of the extended and historical series were tested for equality using the F-, and t-statistic, respectively (Table 2.2). respectively. The serial correlation coefficients and skew coefficients were not tested statistically because an appropriate hypothesis test could not be found. Tables 3.10 and 3.11 show the results of the hypothesis tests for stations 12413000 and 124135000, while Tables 3.12 and 3.13 summarize the results for stations 13186000 and 13185000.

Table 3.9

Statistics of Key Stations for Periods of Record Corresponding to Extended and Historical Record of the Subordinate Stations (Streamflow in cfs days)

Station 12413500

	Historical (8/20 - 9/39)				Historical (10/39 - 9/72)			
	Mean	Stand Dev	Skew Coef	Serial Cor Cf	Mean	Stand Dev	Skew Coef	Serial Cor Cf
Oct	17364	13171	2.458	.951	19230	11474	1.508	.588
Nov	36988	44257	2.808	.889	38809	27046	.938	.704
Dec	64148	90697	3.376	.412	63731	53439	1.475	.637
Jan	55922	62574	2.202	.842	56752	30164	.726	.352
Feb	56565	51514	1.300	.406	76905	53591	1.402	.269
Mar	100338	50365	.787	.645	97789	55090	2.191	.282
Apr	227645	71622	.437	.290	217362	76735	.203	-.001
May	207681	80970	-.361	.430	241186	88188	-.196	.372
Jun	71414	46251	1.622	.692	89898	41269	.950	.694
Jul	23467	7559	.735	.950	30924	11112	.743	.850
Aug	13069	2814	.238	.935	15756	4234	.762	.888
Sep	11896	4127	2.202	.436	13043	3029	1.330	.539
Yr	886100	294889	.087	.092	961385	234624	-.519	.208

Station 13185000

	Historical (4/11 - 4/45)				Historical (5/45 - 9/83)			
	Mean	Stand Dev	Skew Coef	Serial Cor Cf	Mean	Stand Dev	Skew Coef	Serial Cor Cf
Oct	11309	2362	.612	.866	12823	2838	1.155	.607
Nov	12491	4543	2.918	.526	13750	3553	1.335	.347
Dec	13161	5255	1.522	.432	16647	9303	2.792	.260
Jan	12240	3757	1.671	.635	15711	5923	1.440	.634
Feb	12172	3393	.862	.629	16130	5613	1.112	.367
Mar	23165	9249	.767	.676	26278	11477	1.530	.286
Apr	61441	27049	1.994	.558	65950	27565	.267	.374
May	110198	37365	.473	.438	126353	39316	-.104	.596
Jun	93991	47131	.211	.608	111583	38434	.187	.627
Jul	33257	19931	1.280	.778	43079	20705	.574	.870
Aug	13020	4430	.854	.910	15982	4628	.701	.896
Sep	10194	2280	.357	.893	11877	2343	.427	.853
Yr	405005	123835	.442	-.149	476759	130619	.020	-.073

Table 3.10

Hypothesis Tests for Equality of Extended and Historical Variances

Subordinate Station 12413000

Period	s_m (cfsd)	s_n (cfsd)	m (yrs)	n (yrs)	Sample F	95% F
October	9970	9049	19	33	1.21	2.34
November	34656	21760	19	33	*2.54	2.34
December	69940	42032	19	33	*2.77	2.34
January	48734	23937	19	33	*4.14	2.34
February	42286	39300	33	19	1.16	2.90
March	44228	40238	33	19	1.21	2.90
April	60972	56144	33	19	1.18	2.90
May	71385	65354	33	19	1.19	2.90
June	33596	29964	19	33	1.26	2.34
July	7557	4894	33	19	2.38	2.90
August	2752	1952	33	20	1.99	2.54
September	2693	2072	20	33	1.69	2.31
Annual	222967	177501	19	33	1.58	2.34

Key Station 12413500

Period	s_m (cfsd)	s_n (cfsd)	m (yrs)	n (yrs)	Sample F	95% F
October	13171	11474	19	33	1.82	2.34
November	44257	27046	19	33	*2.68	2.34
December	90697	53439	19	33	*2.88	2.34
January	62574	30164	19	33	*4.30	2.34
February	53591	51514	33	19	1.08	2.90
March	55090	50365	33	19	1.20	2.90
April	76735	71622	33	19	1.15	2.90
May	88188	80970	33	19	1.19	2.90
June	46251	41269	19	33	1.26	2.34
July	11112	7559	33	19	2.16	2.90
August	4234	2814	33	20	2.26	2.54
September	4127	3029	20	33	1.86	2.31
Annual	294889	234624	19	33	1.58	2.34

* Standard deviations that were not statistically equal.

Table 3.11

Hypothesis Tests for Equality of Extended and Historical Means

Subordinate Station 12413000

Period	$\bar{x}_m$ (cfsd)	$\bar{x}_n$ (cfsd)	m (yrs)	n (yrs)	Sp (cfsd)	Sample t	95% t
October	13845	13066	33	19	9330	.304	2.01
November	29304	29040	33	19	-	.031	2.05
December	50268	48585	19	33	-	.098	2.06
January	42183	41878	19	33	-	.026	2.07
February	57816	43568	33	19	41427	1.253	2.01
March	77650	74983	19	33	43089	.225	2.01
April	181900	172436	19	33	59588	.579	2.01
May	181806	154882	33	19	69660	1.408	2.01
June	61055	47071	33	19	31080	1.639	2.01
July	20575	15207	33	19	6879	*2.843	2.01
August	10683	8878	33	20	2534	*2.641	2.01
September	8994	8349	33	20	2280	1.049	2.01
Annual	721960	671842	33	19	192040	.951	2.01

Key Station 12413500

Period	$\bar{x}_m$ (cfsd)	$\bar{x}_n$ (cfsd)	m (yrs)	n (yrs)	Sp (cfsd)	Sample t	95% t
October	19230	17364	33	19	12112	.535	2.01
November	38809	36988	33	19	-	.163	2.06
December	64148	63731	19	33	-	.018	2.06
January	56752	55922	19	33	-	.054	2.07
February	76905	56565	33	19	52853	1.336	2.01
March	100338	97789	19	33	53437	.166	2.01
April	227645	217362	19	33	74934	.476	2.01
May	241186	207681	33	19	85660	1.358	2.01
June	89898	71414	33	19	43129	1.488	2.01
July	30924	23467	33	19	9980	*2.595	2.01
August	15756	13069	33	20	3768	*2.516	2.01
September	13043	11896	33	20	3479	1.163	2.01
Annual	961385	886100	33	19	257946	1.014	2.01

* Means that were not statistically equal.

Table 3.12

Hypothesis Tests for Equality of Extended and Historical Variances

Subordinate Station 13186000

Period	s_m (cfsd)	s_n (cfsd)	m (yrs)	n (yrs)	Sample F	95% F
October	1403	1373	38	34	1.04	2.08
November	1903	1433	34	38	1.76	2.05
December	2731	1704	38	34	*2.57	2.08
January	1756	1079	38	34	*2.65	2.08
February	1651	1043	38	34	*2.50	2.08
March	3823	3367	38	34	1.29	2.08
April	18269	17714	38	35	1.06	2.07
May	33029	31011	39	35	1.13	2.06
June	39633	32368	35	39	1.50	2.03
July	14642	14160	39	35	1.07	2.06
August	3303	3134	39	34	1.11	2.08
September	1767	1753	39	34	1.02	2.08
Annual	94040	93653	38	34	1.01	2.08

Key Station 13185000

Period	s_m (cfsd)	s_n (cfsd)	m (yrs)	n (yrs)	Sample F	95% F
October	2838	2362	38	34	1.44	2.08
November	4543	3553	34	38	1.64	2.05
December	9303	5255	38	34	*3.13	2.08
January	5923	3757	38	34	*2.48	2.08
February	5613	3393	38	34	*2.74	2.08
March	11477	9249	38	34	1.54	2.08
April	27565	27049	38	35	1.04	2.07
May	39316	37365	39	35	1.11	2.06
June	47131	38434	35	39	1.50	2.03
July	20705	19931	39	35	1.08	2.06
August	4628	4430	39	34	1.09	2.08
September	2343	2280	39	34	1.06	2.08
Annual	130619	123835	38	34	1.11	2.08

* Standard deviations that were not statistically equal.

Table 3.13

Hypothesis Tests for Equality of Extended and Historical Means

Subordinate Station 13186000

Period	$\overline{x}_m$ (cfsd)	$\overline{x}_n$ (cfsd)	m (yrs)	n (yrs)	Sp (cfsd)	Sample t	95% t
October	7596	6794	38	34	1389	*2.444	2.00
November	7629	7061	38	34	1671	1.440	2.00
December	7945	7070	38	34	-	1.649	2.00
January	7755	6933	38	34	-	*2.420	2.00
February	7402	6249	38	34	-	*3.580	2.00
March	11606	10053	38	34	3615	1.820	2.00
April	38289	33612	38	35	18005	1.109	2.00
May	85770	72897	38	35	32092	1.723	2.00
June	76265	63215	39	35	35982	1.558	2.00
July	26255	20235	39	35	14416	1.794	2.00
August	9488	7622	39	34	3226	*2.465	2.00
September	7115	5981	39	34	1760	*2.746	2.00
Annual	294519	247348	39	34	93858	*2.129	2.00

Key Station 13185000

Period	$\overline{x}_m$ (cfsd)	$\overline{x}_n$ (cfsd)	m (yrs)	n (yrs)	Sp (cfsd)	Sample t	95% t
October	12823	11309	38	34	2624	*2.444	2.00
November	13750	12491	38	34	4050	1.317	2.00
December	16647	13161	38	34	-	1.980	2.00
January	15711	12240	38	34	-	*3.000	2.00
February	16130	12172	38	34	-	*3.663	2.00
March	26278	23165	38	34	10486	1.258	2.00
April	65950	61441	38	35	27319	.704	2.00
May	126353	110198	38	35	38407	1.806	2.00
June	111583	93991	39	35	42762	1.767	2.00
July	43079	33257	39	35	20343	*2.074	2.00
August	15982	13020	39	34	4537	*2.782	2.00
September	11877	10194	39	34	2314	*3.100	2.00
Annual	476759	405005	39	34	127466	*2.385	2.00

* Means that were not statistically equal.

After comparing the hypothesis tests (Tables 3.10 to 3.13) it was noted that, for all but one monthly mean, the same statistics were significantly different at both the subordinate and corresponding key station. Therefore, the deviations between the statisitics of the extended values and the historical values at the subordinate stations were considered to be reasonable. In addition, the following observations were made from the results of the hypothesis tests:

1) All of the standard deviations which were statistically different corresponded to winter months. As previously mentioned in section 2.3, the heaviest precipitation occurs during the winter and spring creating monthly flow series that have a larger variability than those corresponding to periods of little precipitation. It is more likely that subseries taken from a monthly series with a high variability (winter and spring), as opposed to one with a low variability would have standard deviations which are statistically different, because the extreme flows (high and/or low) creating the larger variability of the entire series may not be present in both subseries.

2) The means that were statistically different all occured during months with lower mean flows relative to most of the other months. Once again this result can be linked to the

variability of the records: the smaller the standard deviation, the narrower is the 95% significance band about the mean. The low flow months generally, possessed the lowest variability. Thus, since the extended subseries contained several severe low-flow sequences, the means of the low-flow months decreased, and would more easily fall outside of their respective 95% significance bands than means from a month with a larger variability.

3) The mean annual flows were statistically different at stations 13186000 and 13185000 while they were not at stations 12413000 and 124135000. The annual records at stations 12413500 and 13185000 (Tables B.2 and B.4) were used to generate the extended flow series at stations 12413000 and 13186000, respectively, and because of the strong cross correlation between the respective records, the flow characteristics seen in the key station records were generally seen in the subseries station records. The extended period of record at station 12413500 (1920-1939) did not possess as many severe low-flow sequences relative to its entire length as did the extended period of record at station 13185000 (1911-1945). Therefore, the means of the extended records at stations 13185000 and 13186000 were affected to a greater extent by these low flows (statistically different) than were the extended records at stations 12415000 and 12413000 (statistically equal).

3.5 Statistics of Extended Records

The extended portion of record at stations 12413000 and 13186000 was added to the already existing historical record of each respective station, and the statistics of the combined records (hereafter referred to as the extended record) were determined by equations 2.1 through 2.14. The results are listed in Appendix C along with the monthly and annual correlograms (Tables C.7 and C.8, and Figures C.1 to C.26, respectively). Table 3.14 summarizes how the statistics of the historical records changed after including data extension.

From Table 3.14 it can be seen that generally, the means decreased as a result of data extension, especially during the summer months. This is significant because the heaviest water use usually occurs during the summer, and these results suggest that a lower mean flow exists than defined by the unextended record.

The standard deviations (12413000, 13186000) and skew coefficients (12413000) generally increased during the fall and early winter months (Table 3.14), suggesting the addition of some more extreme flows. Most water resource related problems are the result of the extreme flows, thus making their properties of primary importance. However, the skew coefficient at station 13186000 decreased during the

Table 3.14

Effects of Data Extension upon
Monthly and Annual Statistics

Percent changes resulting in historical statistics as a result of adding in the extended subseries

	Station 12413000			Station 13186000		
Period	Mean	Stand Dev	Coeff Skew	Mean	Stand Dev	Coeff Skew
October	1.38	6.10	6.92	-4.99	2.42	-5.64
November	3.51	2.51	94.09	-3.51	17.44	102.07
December	-0.45	19.80	134.45	-5.20	-14.68	-7.11
January	-2.20	11.24	-10.72	-5.00	-13.27	4.32
February	-7.21	-3.17	-5.70	-7.35	-8.90	-0.53
March	0.62	-2.84	-16.05	-6.32	-3.92	-13.97
April	3.74	-1.76	-7.04	-5.86	-1.29	141.68
May	-3.23	-2.92	-8.80	-6.99	-1.50	93.17
June	-7.34	2.02	-4.26	-7.97	12.07	-78.00
July	-8.39	-3.40	35.83	-10.68	-0.05	-16.96
August	-5.92	-3.55	43.91	-9.16	1.03	-31.95
September	-2.08	12.81	27.79	-7.42	4.07	-33.23
Annual	-1.49	3.18	-31.65	-7.74	1.82	18.43

summer and early fall months and then varied more randomly throughout the rest of the year.

If the extended flows at stations 12413000 and 13186000 could be compared to the actual flows occurring during their respective time periods, they would not exactly match. Yet, due to the strong cross correlation relationships used in estimating these flows, overall they would peak and fall in the same manner as the actual record. As a result, the extended series increases the data base from which the statistics of the series can be estimated, since they are felt to reasonably represent another period of the streamflow record.

The 95% confidence intervals about the annual means and standard deviations of the historical and extended records are listed in Table 3.15. From this table, it can be seen that the confidence intervals decreased considerably, thus making the extended records' statistics much more reliable, and therefore, a more accurate description of the streamflow distributions.

Table 3.15

Ninety-five Percent Confidence Intervals of Annual Means and Standard Deviations of Historical and Extended Records (cfs days)

95% Confidence Interval about Mean: $\bar{y} \pm 1.96\ s/(n)^{1/2}$
95% Confidence Interval about Standard Deviation:

$$\frac{(n-1)s^2}{a_L} \quad \text{to} \quad \frac{(n-1)s^2}{a_U}$$

where:

$$a_L = \alpha_{.05}/2 \quad \text{of} \quad \chi^2$$
$$a_U = 1 - \alpha_{.05}/2 \quad \text{of} \quad \chi^2$$

Station 12413000	Unextended	Extended
95% CI of Mean	648,060 to 765,648	645,595 to 746,995
95% CI of Stand Dev	164,394 to 247,479	189,125 to 249,105

Station 13186000	Unextended	Extended
95% CI of Mean	264,619 to 324,419	249,597 to 293,845
95% CI of Stand Dev	75,835 to 119,593	82,268 to 114,584

CHAPTER 4

NORMALITY OF THE ANNUAL FLOW SERIES

The stochastic models used to represent the annual streamflow series (described later in Chapter 5) differ in form depending on whether the series is normally or nonnormally distributed. Prior to model selection, therefore, the annual records at stations 12413000 and 13186000 had to be checked for normality. Three methods were used to test the normality of the annual records: 1) visual inspection of the annual histograms; 2) hypothesis tests that the coefficient of skew equaled zero and; 3) chi-squared goodness-of-fit test.

4.1 Histograms

A histogram is a graphical representation of a frequency distribution with ranges of values plotted against the number of times a sample value falls within each range (frequency). The frequency distribution for a normal distribution is a symmetrical bell shaped curve. Therefore, if the annual series are normally distributed, their histograms should roughly resemble this shape.

In order to obtain the histograms for the annual streamflow records, each record was divided into about twenty equal discharge ranges. The number of annual values which fell into each range (frequency) was counted and then plotted against that range. The resulting histograms are shown in Figures 4.1 through 4.4.

From these figures, it can be seen that all of the annual histograms roughly resembled a bell-shaped curve. This observation was then checked by statistical methods.

4.2 Test of the Coefficient of Skew

The coefficient of skew for each annual series was calculated from equation 2.6, and the null hypothesis that the coefficient of skew equaled zero was tested at the 95% probability level as described in section 3.1. The coefficient of skew corresponding to the 95% level may be found from the following equation (same as equation 3.2):

$$g(95\%) = \pm\ 1.96(6/n)^{1/2} \qquad (4.1)$$

where:

n = number of years of data

From the results of this test (Table 4.1), all of the annual records could be assumed to be normally distributed.

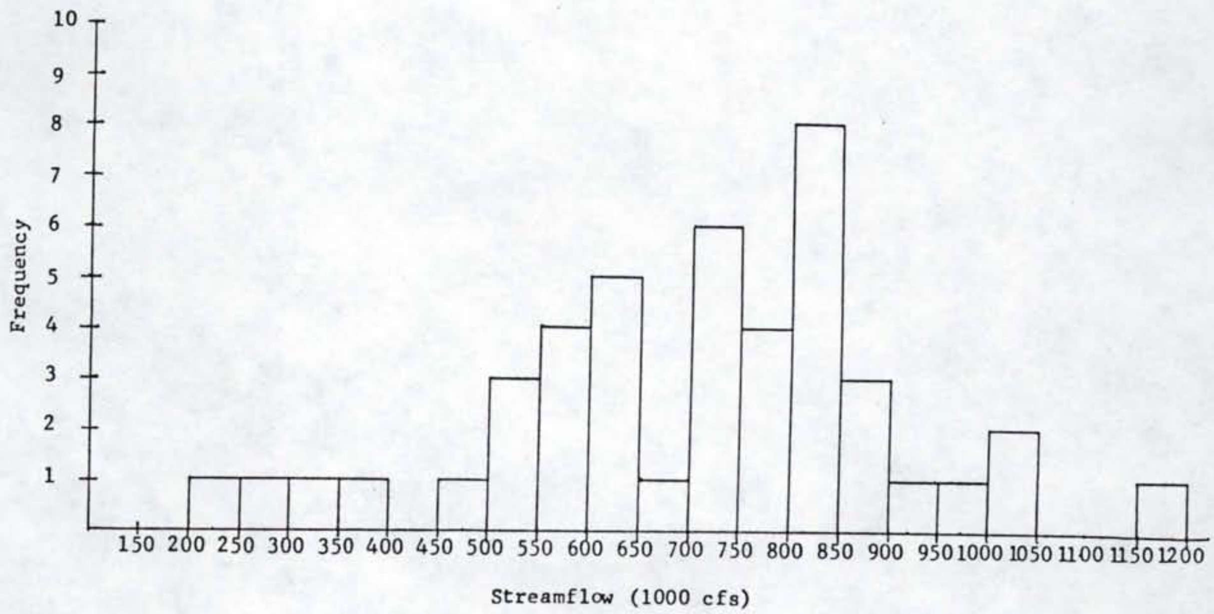


Figure 4.1 Histogram of Annual Flows from Unextended Record at Station 12423000

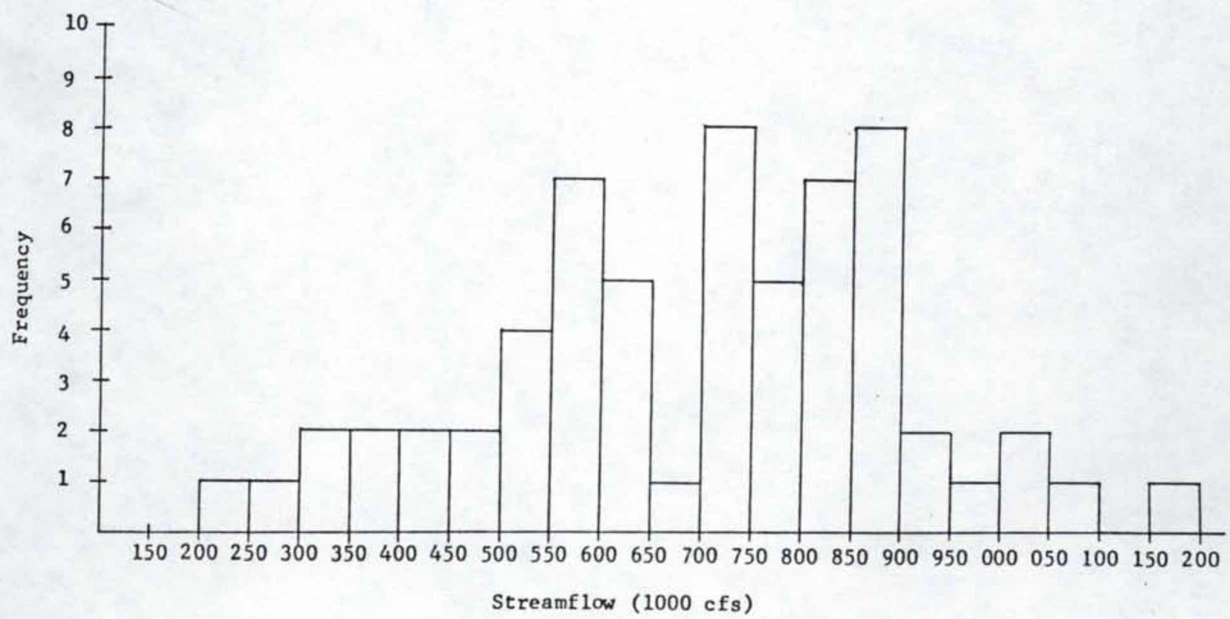


Figure 4.2 Histogram of Annual Flows from Extended Record at Station 12423000

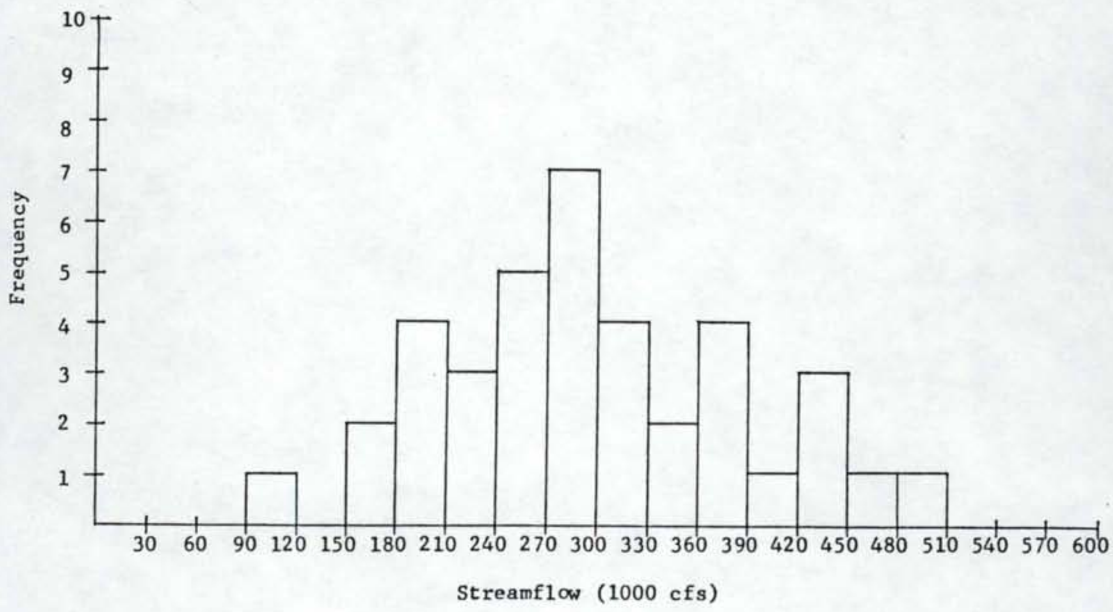


Figure 4.3 Histogram of Annual Flows from Unextended Record at Station 13186000

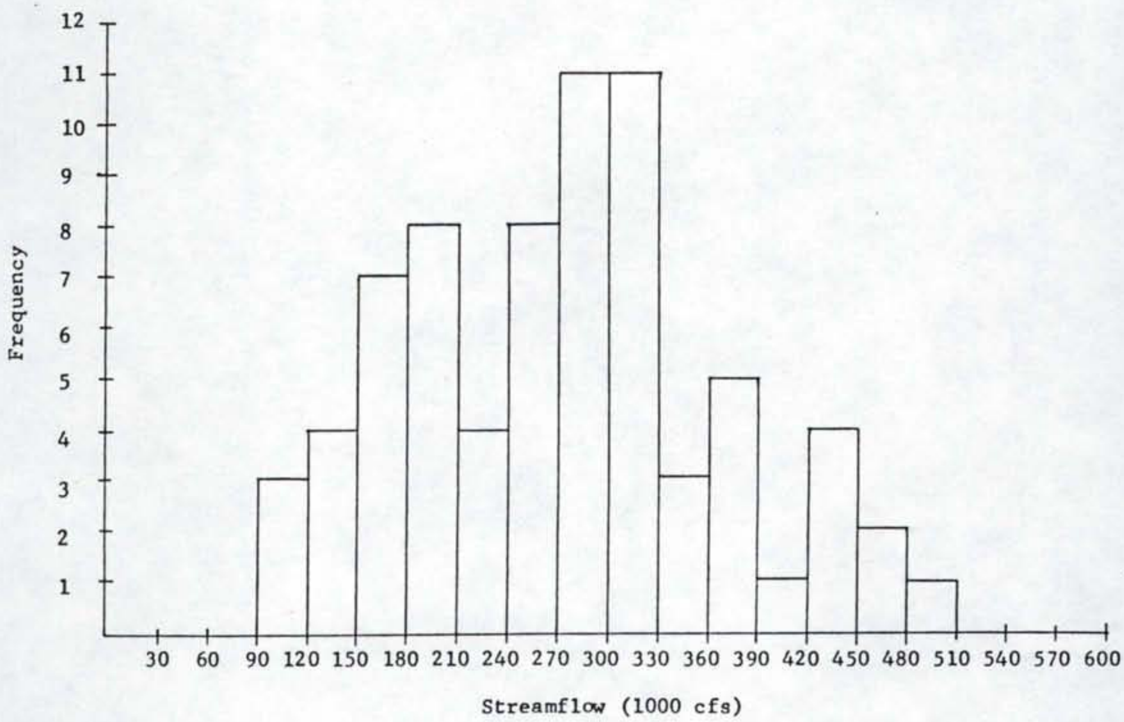


Figure 4.4 Histogram of Annual Flows from Extended Record at Station 13186000

Table 4.1

Coefficient of Skew Test of Normality

Null Hypothesis: $\gamma = 0$
Alternative Hypothesis: $\gamma \neq 0$

Station	Record	n yrs	Sample Skew Coef	95% Skew Coef	Null Hypothesis
12413000	Unextended	44	0.281	0.724	accept
12413000	Extended	63	-.187	-0.605	accept
13186000	Unextended	38	0.229	0.779	accept
13186000	Extended	72	0.271	0.566	accept

4.3 Chi-Squared Goodness-of-fit Test

The chi-squared test is based on a comparison of observed frequencies with expected frequencies, as determined by the use of any assumed probability distribution. In order to apply this test, the streamflow records were broken into discharge ranges and the number of streamflow values which fell into each range were counted. The actual number in each range represents the observed frequency.

The expected frequency was found by assuming a normal distribution, and multiplying the area under a standard normal curve, corresponding to the appropriate range, by the total number of streamflow values. This area corresponding to each range can be found by using the standardized variable

$$Z = (y - \bar{y})/s \qquad (4.2)$$

where $\bar{y}$ and s are the mean and standard deviation respectively, of the entire annual series. Z corresponds to all of the area under the standard normal curve which is to the left of it. Therefore, the Z's corresponding to the two limits of each range can be subtracted to find the area under the standard normal curve represented by each range. These calculations are shown in Table 4.2.

Table 4.2

Frequency Counts for Chi-squared Goodness-of-fit Test

Discharge Range (cfs days)	Z Range	Area of Norm Crv	Observ Freq	Expect Freq
	Station 12413000 - Unextended Record			
0 to 508195	to -.998	.159	5	7
508195 to 612778	-.998 to -.473	.159	9	7
612778 to 706854	-.473 to .000	.182	5	8
706854 to 800930	.000 to .473	.182	12	8
800930 to 905512	.473 to .998	.159	9	7
905512 to	.998 to	.159	4	7
	Station 12413000 - Extended Record			
0 to 445811	to-1.220	.111	8	7
445811 to 539229	-1.22 to -.765	.111	6	7
539229 to 607804	-.765 to -.431	.111	7	7
607804 to 667551	-.431 to -.140	.111	5	7
667551 to 725039	-.140 to .140	.111	6	7
725039 to 784786	.140 to .431	.111	7	7
784786 to 853361	.431 to .765	.111	9	7
853361 to 946779	.765 to 1.220	.111	10	7
946779 to	1.220 to	.111	5	7
	Station 13186000 - Unextended Record			
0 to 209883	to -.900	.184	7	7
209883 to 269410	-.900 to -.267	.211	8	8
269410 to 319627	-.267 to .267	.211	10	8
319627 to 379155	.267 to .900	.211	6	8
379155 to	.900 to	.184	7	7
	Station 13186000 - Extended Record			
0 to 154902	to-1.220	.111	8	8
154902 to 198472	-1.22 to -.765	.111	9	8
198472 to 230455	-.765 to -.431	.111	9	8
230455 to 258320	-.431 to -.140	.111	7	8
258320 to 285131	-.140 to .140	.111	8	8
285131 to 312997	.140 to .431	.111	6	8
312997 to 344980	.431 to .765	.111	10	8
344980 to 388549	.765 to 1.220	.111	7	8
388549 to	1.220 to	.111	8	8

After the observed and expected frequencies had been determined for each range, the following statistic was computed:

$$\chi^2 = \frac{\sum_{i=1}^{k} (f_{oi} - f_{ei})^2}{f_{ei}} \qquad (4.3)$$

where:

f_o = observed frequency
f_e = expected frequency
K = number of ranges

The above statistic follows the chi-squared distribution with $K-n_p-1$ degrees of freedom, where n_p is the number of parameters estimated from the series (in this case it equals two: the mean and standard deviation), and K is the number of discharge ranges. As a result, the 95% value of the chi-squared statistic can be obtained from chi-squared tables. If the computed chi-squared statistic is less than the 95% chi-squared statistic then the hypothesis of normality is accepted. In order for this test to be reliable the ranges have to be chosen such that the expected value in each range is at least five (5). The results of the chi-squared tests for each of the annual time series are shown in Table 4.3, and, as can be seen, all of the annual records passed the chi-squared test for normality.

Table 4.3

Chi-squared Goodness-of-fit Test for Normality

Null Hypothesis: annual series is normally distributed

Station	Record	Sample χ^2	95% χ^2	Null Hypothesis
12413000	Unextended	6.125	7.815	accept
12413000	Extended	3.428	12.592	accept
13186000	Unextended	1.000	5.992	accept
13186000	Extended	1.500	12.592	accept

4.4 Conclusions

From the results of the three tests for normality, it was assumed that all of the annual streamflow series were normally distributed. Also, it can be noted that neither data extension or station location affected the normality of the annual flow distributions.

CHAPTER 5

ANNUAL STREAMFLOW MODELS

Disaggregation modeling takes an existing time series and divides (disaggregates) it into smaller time intervals. In this study, annual time series were to be generated and then disaggregated into monthly streamflow values. Consequently, annual stochastic streamflow models had to be developed for the unextended and extended records at stations 12413000 and 13186000.

This chapter is divided into six main parts: 1) a description of ARMA(p,q) models, 2) a discussion of annual models with respect to the Hurst phenomenon, 3) methods used for model identification, 4) tests on the residuals from the fitted models in part 3, 5) the annual models selected for use, and 6) conclusions. Table 5.1 lists some symbols that will be used throughout this chapter.

Since all of the annual series were found to approximate the normal distribution (Chapter 4), no transformation was needed to normalize the annual series. Consequently, y_t equals the raw annual streamflow values.

Table 5.1

Definition of Symbols used in Annual Models

y_t = normalized annual stramflow value for year t
$\bar{y}$ = mean of normalized annual series
s_y = standard deviation of normalized annual series
k = time lag (years)
r_k = lag-k serial correlation coefficient
s_e^2 = variance of residual series
λ = standardized random deviate
n = number of years of data

5.1 ARMA(p,q) Models

ARMA(p,q) models are commonly known as "autoregressive moving average" models, consisting of two components: the autoregressive, or AR(p), component, and the moving average, or MA(q) component. ARMA(p,q) annual streamflow models utilize the serial correlation of streamflow values separated by 1 to p years, and the correlation that exists between sucessive residual values separated by 1 to q years, where the values of p and q define the "model order".

There is a physical basis for the use of such models in describing annual flows (39). Annual streamflow for a given year is the result of effective precipitation occuring in that year plus a contribution from the previous years' precipitation in the form of groundwater discharge. Also, added to this is the effect of surface storage. The autoregressive component of the ARMA(p,q) model can be used to represent the contribution of streamflow from groundwater discharge (base flow) and long-term surface storage (such as a lake), while the moving average component can be related to the precipitation from the previous q years that resulted in relatively rapid drainage (overland flow and interflow).

The effect of groundwater discharge usually results in an annual series with a positive time dependent structure; such that high flows tend to follow high flows and low flows tend to follow low flows ($r_k > 0$). However, this postive time dependence can be quite small, or negative due to sampling fluctuations.

5.2 Hurst Phenomenon

One major assumption behind stochastic modeling is that the historical time series is just one sample from a much larger population. Therefore, if the properties of the population can be estimated, other just as likely time series can be generated using a model designed to preserve the properties of the population. These generated samples can then be used to better assess the frequency of critical events, for the generated samples can be made as long as desired.

Stochastic models are built to reproduce the main statistical characteristics of an historical time series, assuming them to be the best estimates of the population characteristics. The main statistical characteristics are the mean, standard deviation, skewness and serial correlation structure. Also, when considering extreme events such as droughts, long-term persistence becomes an important characteristic.

Long term persistence has become a matter of concern in stochastic modeling ever since the studies of Hurst (14, 15) were performed. Hurst found that the rescaled range of many different time series could be related to the sample size by the following equation:

$$R_n = (.5n)^h \qquad (5.1)$$

where:

h = Hurst coefficient

R_n = rescaled range which is defined as:

$$R_n = \frac{R(max) - R(min)}{s_n}$$

$$R = \sum_{t=1}^{m} y_t - \frac{m}{n} \sum_{t=1}^{m} y_t$$

R = range

m = first m years

s_n = standard deviation of n years

In considering many different types of natural process time series, Hurst found that h had an average value of .73 with a standard deviation of .09. This is significant, since ARMA(p,q) models use random normal deviates for which the Hurst coefficient has been shown to equal .5 (7, 12). Therefore, the concern that ARMA(p,q) models do not preserve long term persistance prompted the introduction of models designed to preserve the Hurst coefficient for values of h greater than .5 (21, 24). However, studies such as those carried out by Yevjevich (40), O'Connell (29), Hipel and

McLeod (12), and Salas, et al (31), helped to demonstrate that simple ARMA(p,q) models are, for most hydrologic series, capable of reproducing the necessary statistics related to water resources planning problems, and therefore, ARMA(p,q) models were used.

5.3 Tools for Model Identification

AKAIKE INFORMATION CRITERION: Generally, as the number of model parameters estimated from the historical series increases, the preservation of the historical statistics improves and the unexplained variability (s_e^2) decreases. However, if too many parameters are used, the sampling variability of the historical series rather than the actual population characteristics are reproduced. Therefore, a stochastic model should reproduce the main statistical characteristics of the historical series with the minimum number of model parameters which must be estimated from the historical record (parameter parsimony, see section 3.1).

Akaike (1) proposed an equation which considers the number of model parameters and also the reduction in unexplained variability for different ARMA(p,q) models. His equation is known as the Akaike Information Criteria and is stated as:

$$AIC(p,q) = (n)\ln(s_e^2) + 2(p + q) \qquad (5.2)$$

where:

s_e^2 = maximum likelihood estimate of residual variance

Under this criterion, the model which gives the minimum AIC value is the one usually to be selected. Hipel et al. (11) suggests that the AIC criterion be used to aid in the selection of the model order of ARMA(p,q) models.

CORRELOGRAMS: As mentioned in earlier chapters, the correlogram is a plot of the lag-k serial correlation coefficient versus k. The serial correlation coefficients and correlograms for each annual series can be found in Appendix B (Tables B.5, B.7 and Figures B.13 and B.26) and in Appendix C (Tables C.7, C.8 and Figures C.13 and C.26). The correlogram of the sample should resemble the correlogram of the model used to represent it, in order preserve the correlation structure of the historical streamflow record. Hence, a visual comparision of the model and sample correlograms can be used to aid in the selection of a model.

PARTIAL AUTOCORRELATION FUNCTION: The partial autocorrelation function is another way of representing the time dependence of a series. It can be determined by solving recursively the following relations developed by Durbin (6) for $\phi_k(k)$.:

$$\phi_K(p) = \phi_{K-1}(p) - \phi_K(k)\phi_{K-1}(k-p) \quad (5.3)$$

$$\phi_K(k) = \frac{r_K - \sum_{p=1}^{K-1} \phi_{K-1}(p) r_{K-p}}{1 - \sum_{p=1}^{K-1} \phi_{K-1}(p) r_p} \quad (5.4)$$

where:

$\phi_K(p)$ = kth partial autocorrelation coefficient for an AR(p) model.

$\phi_K(k)$ = partial autocorrelation coefficient for k=p.

The partial autocorrelation coefficients as computed by equation 5.4 for the extended and unextended annual records at stations 12413000 and 13186000 are listed in Appendix D as Table D.1. A plot of $\phi_K(k)$ versus k is known as a partial correlogram (Figures D.1 through D.4 in Appendix D). Again, the sample and model partial correlograms should resemble each other in order to preserve the correlation structure of the historical record.

5.4 Selection of an ARMA(p,q) Model

ARMA(0,0): The simplest form of the ARMA(p,q) model was considered first: the ARMA(0,0) model. Then the order of p and q were increased by increments of one until the decrease in unexplained variability (s_e^2) was offset by the increased number of parameters as judged by the Akaike Information Criteria. In addition, the sample and model correlograms and partial correlograms values were compared. Hereafter, when referring to the sample and model correlograms, it will be assumed a reference is being made to both the correlogram and the partial correlogram, unless specifically stated otherwise.

ARMA(0,0): This type of model is commonly referred to as a pure probablistic model and can be expressed as:

$$y_t = \overline{y} + s_e \lambda_t \qquad (5.5)$$

where:

$$s_e = s_y$$

No type of linear dependence between successive streamflow values is assumed to exist. Consequently, the model correlograms can be expressed as follows:

$$r_K = 0 \quad \text{for all } k > 0 \qquad (5.6)$$

$$\phi_K(k) = 0 \quad \text{for all } k > 0 \qquad (5.7)$$

The sample correlogram values of the annual series at stations 12413000 and 13186000 are compared to the pure probablistic model correlogram values in Tables 5.2 through 5.5. Tables were used for comparison instead of graphical correlograms because the correlation coefficients for later models were so small that a scale which could illustrate the differences between these values was diffcult to draw.

The AIC(p,q) values for the pure probalistic model of each annual series were calculated by the following equation and are listed in Table 5.6.

$$AIC(0,0) = (n)\ln(s_e^2)$$

where:

$$s_e^2 = \frac{\sum_{t=1}^{n} (y_t - \overline{y})^2}{n}$$

Table 5.2

Comparision of Sample and Model Correlograms for Station 12413000 - Unextended

Serial Correlation Coefficients

Lag	Sample	ARMA(0,0)	AR(1)	MA(1)	ARMA(1,1)
1	-.056	.000	.0556	-.040	-.025
2	.204	.000	.0031	.000	-.018
3	.040	.000	$1.70\text{X}10^{-4}$	.000	-.013
4	.160	.000	$9.56\text{X}10^{-6}$	.000	-.0091
5	-.119	.000	$5.30\text{X}10^{-7}$	.000	-.0065
6	-.259	.000	$2.95\text{X}10^{-7}$	.000	-.0046
7	-.024	.000	$1.64\text{X}10^{-9}$	.000	-.0033
8	-.215	.000	$9.13\text{X}10^{-11}$	.000	-.0023
9	.001	.000	$5.08\text{X}10^{-12}$	.000	-.0017
10	-.045	.000	$2.82\text{X}10^{-13}$	.000	-.0012

Partial Correlation Coefficients

Lag	Sample	ARMA(0,0)	AR(1)	MA(1)	ARMA(1,1)
1	-.056	.000	-.056	-.040	-.025
2	.202	.000	.000	.0016	-.018
3	.063	.000	.000	$-6.37\text{X}10^{-5}$	-.014
4	.130	.000	.000	$-2.55\text{X}10^{-6}$	-.0010
5	-.131	.000	.000	$-1.02\text{X}10^{-7}$	-.0074
6	-.360	.000	.000	$-4.07\text{X}10^{-9}$	-.0055
7	-.049	.000	.000	$-1.62\text{X}10^{-10}$	-.0041
8	-.121	.000	.000	$-6.50\text{X}10^{-12}$	-.0030
9	.096	.000	.000	$-2.59\text{X}10^{-13}$	-.0022
10	.164	.000	.000	$-1.04\text{X}10^{-14}$	-.0016

Table 5.3

Comparision of Sample and Model Correlograms for Station 12413000 - Extended

Serial Correlation Coefficients

Lag	Sample	ARMA(0,0)	AR(1)	MA(1)	ARMA(1,1)
1	.017	.000	.0169	.013	-.0196
2	.141	.000	.00029	.000	.0145
3	-.085	.000	$4.83X10^{-6}$	.000	-.0108
4	.064	.000	$8.16X10^{-8}$	.000	.0080
5	-.070	.000	$1.38X10^{-9}$	.000	-.0059
6	-.125	.000	$2.33X10^{-11}$	.000	.0044
7	.053	.000	$3.94X10^{-13}$	.000	-.0032
8	-.161	.000	$6.65X10^{-15}$	.000	.0024
9	-.045	.000	$1.12X10^{-16}$	.000	-.0018
10		.000	$1.90X10^{-18}$	.000	.0013

Partial Correlation Coefficients

Lag	Sample	ARMA(0,0)	AR(1)	MA(1)	ARMA(1,1)
1	.017	.000	.017	.013	-.0196
2	.140	.000	.000	-.00017	.0141
3	-.092	.000	.000	$2.20X10^{-6}$	-.0102
4	.496	.000	.000	$-2.86X10^{-8}$	.0074
5	-.050	.000	.000	$3.72X10^{-10}$	-.0053
6	-.150	.000	.000	$-4.83X10^{-12}$	.0038
7	.092	.000	.000	$6.28X10^{-14}$	-.0028
8	-.153	.000	.000	$-8.17X10^{-16}$	.0020
9	-.074	.000	.000	$1.06X10^{-17}$	-.0014
10	.005	.000	.000	$-1.38X10^{-19}$	.0010

Table 5.4

Comparision of Sample and Model Correlograms for Station 13186000 - Unextended

Serial Correlation Coefficients

Lag	Sample	ARMA(0,0)	AR(1)	MA(1)	ARMA(1,1)
1	-.048	.000	.0488	-.044	-.1370
2	.090	.000	.00238	.000	-.0989
3	-.294	.000	.000116	.000	-.0714
4	-.007	.000	$5.67X10^{-6}$	.000	-.0516
5	-.205	.000	$2.77X10^{-7}$	.000	-.0372
6	-.067	.000	$1.35X10^{-8}$	.000	-.0269
7	.122	.000	$6.59X10^{-10}$	.000	-.0194
8	-.253	.000	$3.22X10^{-11}$	.000	-.0140
9	.172	.000	$1.82X10^{-13}$	.000	-.0101
10	-.419	.000	$7.66X10^{-14}$	.000	-.0073

Partial Correlation Coefficients

Lag	Sample	ARMA(0,0)	AR(1)	MA(1)	ARMA(1,1)
1	-.048	.000	-.049	-.0444	-.137
2	.088	.000	.000	-.00198	-.120
3	-.289	.000	.000	$-8.79X10^{-5}$	-.106
4	-.037	.000	.000	$-3.91X10^{-6}$	-.0957
5	-.175	.000	.000	$-1.74X10^{-7}$	-.0867
6	-.182	.000	.000	$-7.74X10^{-9}$	-.0792
7	.137	.000	.000	$-3.44X10^{-10}$	-.0729
8	-.412	.000	.000	$-1.53X10^{-11}$	-.0674
9	.109	.000	.000	$-6.81X10^{-13}$	-.0625
10	-.552	.000	.000	$-3.03X10^{-14}$	-.0583

Table 5.5

Comparision of Sample and Model Correlograms for Station 13186000 - Extended

Serial Correlation Coefficients

Lag	Sample	ARMA(0,0)	AR(1)	MA(1)	ARMA(1,1)
1	-.037	.000	-.037	-.035	.0752
2	.046	.000	.00137	.000	-.0632
3	-.121	.000	$5.06X10^{-5}$	.000	.0531
4	.100	.000	$1.87X10^{-6}$	.000	-.0446
5	.087	.000	$6.93X10^{-8}$	.000	.0375
6	-.034	.000	$2.56X10^{-9}$	.000	-.0315
7	.163	.000	$9.49X10^{-11}$	.000	.0264
8	-.067	.000	$3.51X10^{-12}$	.000	-.0222
9	.129	.000	$1.30X10^{-13}$	.000	.0186
10		.000	$4.81X10^{-15}$	.000	-.0157

Partial Correlation Coefficients

Lag	Sample	ARMA(0,0)	AR(1)	MA(1)	ARMA(1,1)
1	-.037	.000	-.037	-.045	.0752
2	.044	.000	.000	-.00122	-.0692
3	-.118	.000	.000	$-4.28X10^{-5}$	.0640
4	.092	.000	.000	$-1.50X10^{-6}$	-.0595
5	.105	.000	.000	$-5.25X10^{-8}$	.0555
6	-.053	.000	.000	$-1.84X10^{-9}$	-.0519
7	.182	.000	.000	$-6.43X10^{-11}$	.0486
8	-.046	.000	.000	$-2.25X10^{-12}$	-.0456
9	.091	.000	.000	$-7.88X10^{-14}$	.0429
10	-.091	.000	.000	$-2.76X10^{-15}$	-.0406

Table 5.6

Summary of AIC(p,q) Values of Competing Models

Station	Record	Model	n yrs	s_e^2 (cfsd)^2	AIC(p,q)
12413000	Unextended	probablistic	44	3.869X10^{10}	1072.7
12413000	Unextended	AR(1)	44	3.947X10^{10}	1075.5
12413000	Unextended	MA(1)	44	3.861X10^{10}	1074.6
12413000	Unextended	ARMA(1,1)	44	3.559X10^{10}	1073.0
12413000	Extended	probablistic	63	4.148X101^{0}	1540.3
12413000	Extended	AR(1)	63	4.214X10^{10}	1543.2
12413000	Extended	MA(1)	63	4.148X10^{10}	1542.2
12413000	Extended	ARMA(1,1)	63	4.022X10^{12}	1542.3
13186000	Unextended	probablistic	38	8.611X10^{9}	869.3
13186000	Unextended	AR(1)	38	8.823X10^{9}	872.2
13186000	Unextended	MA(1)	38	8.593X10^{9}	871.2
13186000	Unextended	ARMA(1,1)	38	7.940X10^{9}	870.2
13186000	Extended	probablistic	72	9.042X10^{9}	1650.6
13186000	Extended	AR(1)	72	9.157X10^{9}	1653.5
13186000	Extended	MA(1)	72	9.031X10^{9}	1652.5
13186000	Extended	ARMA(1,1)	72	9.005X10^{9}	1654.3

ARMA(p,0): With q=0, the ARMA(p,q) model becomes identical to the AR(p) autoregressive model of order p. The AR(p) model accounts for the part of the total annual series variance (s_y^2) which can be explained by the linear dependence between successive annual flows separated by 1 to p years.

The general form of the AR(p) model and its correlograms are given by:

$$y_t = \overline{y} + \sum_{j=1}^{p} \phi_j (y_{t-j} - \overline{y}) + s_e \lambda_t \qquad (5.8)$$

$$s_e = s_y(1 - \sum_{j=1}^{p} \phi_j r_j) \qquad (5.9)$$

$$r_k = \sum_{j=1}^{p} \phi_j r_{k-j} \quad (\text{for } k > 0) \qquad (5.10)$$

$\phi_k(k)$ = peaks at lags 1 through k and then equals zero.

where:

ϕ_j = autoregression coefficient of order j

More specifically, the AR(1) model was fitted to each of the annual time series at stations 12413000 and 13186000. The correlograms of the AR(1) model reduce to:

$$r_k = \phi^k \qquad (5.11)$$

$$\phi_1(1) = r_1 \qquad (5.12)$$
$$\phi_k(k) = 0, \quad \text{for } k > 1$$

The autoregression coefficient ϕ_1 can be estimated by either the method of moments or the method of maximum likelihood. The method of maximum likelihood was used because generally it gives better parameter estimates (32) and s_e^2 in the Akaike Information Criteria is the maximum likelihood estimate of the residual variance. An approximate method of the maximum likelihood estimate was used where:

$$\phi_1 = d_{12}/d_{22} \qquad (5.13)$$

$$d_{12} = \frac{n}{n-1} \sum_{t=1}^{n-1} (y_t - \bar{y})(y_{t+1} - \bar{y}) \qquad (5.14)$$

$$d_{22} = \frac{n}{n-2} \sum_{t=2}^{n-1} (y_t - \bar{y})^2 \qquad (5.15)$$

The results of the calucations using equations 5.13 through 5.15 are listed in Table 5.7.

Table 5.7

Estimates of ϕ_1

Station	Record	d_{12}	d_{22}	Max Lik ϕ_1	Moment ϕ_1
12413000	Unextended	$-9.659X10^{10}$	$1.737X10^{12}$	-.0556	-.0562
12413000	Extended	$4.452X10^{10}$	$2.639X10^{12}$	.0169	.0169
13186000	Unextended	$-1.544X10^{10}$	$3.164X10^{11}$	-.0488	-.0482
13186000	Extended	$-2.330X10^{10}$	$6.298X10^{11}$	-.0370	-.0366

For comparision, the autoregression coefficients were also estimated by the method of moments. This method involves solving equation 5.11 using the sample moment estimate of r_1.

$$\phi_1 = r_1 \tag{5.16}$$

The method of moments gave almost the exact same estimate as the method of maximum likelihood, as shown in Table 5.7. In order for an AR(1) model to be stationary (statistics not changing with time), the absolute value of ϕ_1 must be less than one. All of the AR(1) models fitted to the historical records are stationary.

The low value of ϕ_1 can be interpreted physically to mean than the flows from the previous year have little effect on the following years' flows. As stated in section

5.1, this implies that these watersheds do not store large volumes of precipitation over a period of years.

Next, the AR(1) correlogram values were determined from equations 5.11 and 5.12 and are compared to the sample correlogram values in Tables 5.2 through 5.5.

The AIC(1,0) values were then determined by:

$$AIC(1,0) = (n)\ln(s_e^2) + 2$$

where the maxiumum likelihood estimate of s^2 is

$$s_e^2 = \frac{1}{n - 1}(d_{11} - \phi_1 d_{12}) \qquad (5.17)$$

$$d_{11} = \sum_{t=1}^{n}(y_t - y)^2 \qquad (5.18)$$

The values of d_{11} and s_e^2 for each annual series are shown in Table 5.8

Table 5.8

Maximum Likelihood Estimates of s_e^2

Station	Record	d_{11}	ϕ_1	s_e^2
12413000	Unextended	$1.7024X10^{12}$	-.0556	$3.9467X10^{10}$
12413000	Extended	$2.6136X10^{12}$	.0169	$4.2142X10^{10}$
13186000	Unextended	$3.2721X10^{11}$	-.0488	$8.8232X10^{9}$
13186000	Extended	$6.5103X10^{11}$	-.0870	$9.1573X10^{9}$

The resulting values of the AIC(1,0) listed in Table 5.6 are higher than the AIC(0,0) values. In addition, from Tables 5.2 through 5.5, it can be seen that the AR(1) model correlogram values are nearly equal to zero, as is the case of the pure probablistic model. Therefore, since no major imporovements were made by increasing the order of p to one, no higher orders of p were examined.

ARMA(0,q): With p=0, the ARMA(p,q) model becomes identical to the MA(q) moving average model of order q. The MA(q) model accounts for the part of the total annual series variance (s_y^2) which can be explained by the linear dependence between the residual values separated by 1 to q years.

The general form of the MA(q) model and its correlogram are given by:

$$y_t = \overline{y} - \sum_{J=0}^{q} \theta_J e_{t-J} \tag{5.19}$$

$$r_K = \frac{\sum_{J=0}^{q-K} \theta_J \theta_{J-1}}{\sum_{J=0}^{q} \theta_J} \qquad \text{for } k \leq q \tag{5.20}$$

$$r_K = 0 \qquad \text{for } k > q$$

$\phi_K(k)$ = infinite in extent and attenuates with a mixture of damped waves and/or damped expontentials.

where:

θ_J = moving average coefficient of j^{th} order ($\theta_0 = -1$)

The MA(1) model was fitted to each of the annual time series at stations 12413000 and 13186000. The correlogram of the MA(1) model reduces to:

$$r_1 = \frac{-\theta_1}{1 + \theta_1^2} \tag{5.21}$$

$$r_K = 0 \qquad \text{for } k > 1$$

$$\phi_K(k) = \text{infinite in extent and attenuates}$$

In order to construct the correlogram, the parameter θ_1 must be estimated. In 1976, McLeod (23) developed a modified sum of squares method which provides parameter estimates that are close approximations to the exact maximum likelihood estimates. However, an earlier method known as the unconditional sums of squares approach was used here, in which the maximum likelihood estimate of θ_1 can be approximated by finding the value of θ_1 which gives the minimum sum of the residuals (e_t) squared. The residuals of an MA(1) process can be found by:

$$e_t = y_t - \bar{y} + \theta_1 e_{t-1} \tag{5.22}$$

Therefore, the minimum sum of residuals squared can be written as:

$$S(\theta_1)_{min} = (y_t - \bar{y} + \theta_1 e_{t-1})^2 \tag{5.23}$$

The starting value of e_{t-1} is taken as 0, its expected value. Generally, the value of e_{t-1}only influences the first few residual values and does not significantly effect the estimate of θ_1(32).

The estimate of θ_1 for each of the annual series was found by computing the sum of the residuals squared for θ_1 in the range of -1 to 1 using increments of .1. In order for an MA(1) model to be valid, the absolute value of θ_1 must be less than 1. This range was then refined by taking smaller increments for the values of θ_1 until the minimum sum of residuals squared was found. These iterations are shown in Tables D.2 through D.5, included in Appendix D. The final values obtained for θ_1 are listed in Table 5.9.

Table 5.9

Maximum Likelihood Estimates of θ_1

Station	Record	$S(\theta_1)_{min}$	θ_1
12413000	Unextended	$1.6987X10^{12}$	.040
12413000	Extended	$2.6130X10^{12}$	-.013
13186000	Unextended	$3.2655X10^{11}$	.0445
13186000	Extended	$6.5024X10^{11}$	.035

With the estimates of θ_1 corresoponding to the minimum sum of residuals squared, the correlogram values for the MA(1) models could be computed using equation 5.20. The sample and MA(1) model correlogram values are compared in Tables 5.2 through 5.5.

Next, the AIC(0,1) value for each MA(1) model was determined using:

$$AIC(0,1) = (n)\ln(s_e^2) + 2$$

where:

$$s_e^2 = \frac{1}{n} S(\theta_1)_{min} \qquad (5.24)$$

and are listed in Table 5.6.

As can be seen by comparing the MA(1) model correlogram and AIC(0,1) values, with those for the ARMA(0,0) model, the MA(1) model does not appear to be significantly better than the pure probablistic model. As a result, no higher orders of q were considered.

Previously, it was mentioned that the MA component represents the portion of the annual precipitation from the previous q years that resulted in relatively rapid drainage. The MA(1) coefficient is practically equal to zero, meaning there is little linear dependence between each year's streamflow as explained by the previous year's rapid drainage. Also, the AR(1) models showed that there was no linear dependence between each year's streamflow as explained by groundwater discharge or long term storage outflow.

Both streams are perennial, and this would indicate that there is a significant groundwater component to sustain the flow during dry periods. Therefore, it was decided to try the ARMA(1,1) model which combines both the AR(1) and the MA(1) model.

ARMA(p,q): The autoregressive moving average model ARMA(p,q) combines the AR(p) model and the MA(q) model resulting in:

$$y_t = \bar{y} + \sum_{J=1}^{p} \phi_J (y_{t-J} - \bar{y}) - \sum_{J=0}^{q} \theta_J e_{t-J} \qquad (5.25)$$

The correlogram for the first q lags is a function of both ϕ_1 and θ_1, but for lags higher than q, the correlogram becomes only a function of ϕ_1. This can be seen by examining the correlograms of the AR(p) and MA(q) models. The MA(q) model goes to zero after q lags whereas the AR(p) correlogram is infinite in extent. The partial correlogram of the ARMA(p,q) model starts with the first p lags being irregular, followed by damped exponentials and/or damped waves.

More specifically the ARMA(1,1) model was now fitted to each of the annual series at station 12413000 and 13186000. The correlogram of the ARMA(1,1) model is expressed as:

$$r_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1}$$

$$r_K = \phi_1 r_{K-1} \qquad \text{for } k > 1 \qquad (5.26)$$

$\phi_K(k)$ = infinite in extent and attenuates

The maximum likelihood estimate of ϕ_1 and θ_1 are found in much the same way as θ_1 was found for the MA(1) model. Again, the maximum likelihood estimate of ϕ_1 and θ_1 can be approximated by the values of of ϕ_1 and θ_1 which together give the minimun sum of the residuals (e_t^2) squared. The residuals of an ARMA(1,1) process can be found by:

$$e_t = y_t - \bar{y} - \phi_1(y_{t-1} - \bar{y}) + \theta_1 e_{t-1} \qquad (5.27)$$

Consequently, the minimum sum of residuals squared can be written as:

$$S(\phi_1, \theta_1)_{min} = \left(y_t - \bar{y} - \phi_1(y_{t-1} - \bar{y}) + \theta\, e_{t-1}\right)^2 \quad (5.28)$$

The starting value of e_{t-1} again is taken as 0, its expected value. The minimum sum of residuals squared was found by calcuating the sum for a range of values of the parameters. Initially, ϕ_1 and θ_1 were both varied from -1 to 1 by increments of .1. Contour lines of equal sum-of-squares were drawn and the estimates of ϕ_1 and θ_1 were finally refined to increments of .001. Only the final iterations of these calculations are shown in Tables D.6 through D.9 in Appendix D, with the resulting values of ϕ_1 and θ_1 summarized in Table 5.10

Table 5.10

Maximum Likelihood Estimates of ϕ_1 and θ_1

Station	Record	$S(\phi_1, \theta_1)_{min}$	ϕ_1	θ_1
12413000	Unextended	1.530534×10^{12}	.713	.739
12413000	Extended	2.493270×10^{12}	-.741	-.722
13186000	Unextended	2.937970×10^{11}	.722	.965
13186000	Extended	6.393534×10^{11}	-.840	-.960

With the estimates of ϕ_1 and θ_1 corresponding to the minimum sum of residuals squared, the correlograms for the ARMA(1,1) models could be computed using equation 5.21. The sample and ARMA(1,1) correlogram values are compared in Tables 5.2 through 5.5.

Next, the AIC(1,1) value for each ARMA(1,1) model was determined using:

$$AIC(1,1) = (n)\ln(s_e^2) + 4$$

where:

$$s_e^2 = \frac{1}{n} S(\phi_1 \theta_1)_{min} \qquad (5.29)$$

and are listed in Table 5.6.

As can be seen by the comparision of the AIC(1,1) values, and the correlogram values, the ARMA(1,1) model does not significantly improve the description of the annual series. Therefore, no higher orders were considered of the combined autoregressive moving average model.

Based on the AIC(p,q,) criteria for each of the models, a pure probablistic model would be the choice. The correlograms for each series also substantiate this choice, since all of the model correlogram values are approximately equal to zero. Hence, the annual time series examined do not seem to possess a linear time-dependent structure. In addition, if the sample correlograms as illustrated in Figures B.13, B.26, (Appendix B) C.13, and C.26 (Appendix C) are examined, it can be seen that at the 95% significance level and up to twelve lags, nearly all of the correlation coefficients are statistically equal to zero. Consequently, the pure probablistic model was choosen for all of the annual series.

5.5 Residuals

Since the residuals of the pure probablistic model equal each value minus the mean annual flow, the residual lag-one serial correlation coefficients and skew coefficient would remain the same as for the actual annual series. In chapter 4, it was shown that each annual series could be assumed to follow a normal distribution, and as just mentioned, the correlograms show no significant correlation coefficients Therefore, the residuals of the pure probablistic models can be assumed to be normally distributed and temporally independent.

5.6 Model Summary

The models selected for annual streamflow generation are summarized below.

Station 12413000 - Unextended

$$y_t = 706854 + 198977\lambda_t \quad \text{(cfs days)} \quad (5.30)$$

Station 12413000 - Extended

$$y_t = 696295 + 205315\lambda_t \quad \text{(cfs days)} \quad (5.31)$$

Station 13186000 - Unextended

$$y_t = 294519 + 94040\lambda_t \quad \text{(cfs days)} \quad (5.32)$$

Station 13186000 - Extended

$$y_t = 271726 + 95757\lambda_t \quad \text{(cfs days)} \quad (5.33)$$

The fact that all of the annual models contained no serial correlation component, suggests that these streams do not have significant over-year storage capabilities, or that the carryover or groundwater level is about the same each fall. Furthermore, since no correlation was found between sucessive years' rapid drainage, there seems to exist a complex relationship between the storage and rapid drainage components of flow, with their relative contributions changing from year to year. Because there are so many variables affecting streamflow i.e., infiltration rate, ground cover, hydraulic conductivity, slope, aspect, temperature, etc., simple linear correlations just could not account for the complexity of the process.

Data extension did not change the form of the model at either station. However, as mentioned in Section 3.5, the extended records give more reliable estimates of the annual flow statistics and thus more reliable model parameter estimates. Consequently, the extended models would be a more accurate representation of the streamflow process at each station.

CHAPTER 6

DISAGGREGATION MODELS

A primary goal of this study was to examine synthetic monthly flow sequences generated from a disaggregation model, to determine their low flow characteristics. Consequently, an annual to monthly disaggregation model was developed for the four time series studied (12413000 - unextended, 12413000 - extended, 13186000 - unextended, 13186000 - extended).

The topics covered in this chapter are: the development of disaggregation models, model selection and assumptions, normality of monthly series, the lognormal transform, standardization of annual and monthly series, correlation coefficients between monthly and annual streamflow values, estimation of disaggregation model parameters, synthetic generation, hypothesis test for equality of historic and synthetic statistics, and conclusions.

6.1 Disaggregation Models

Traditionally, synthetic streamflow records have been generated by models designed to preserve the statistics at one time level. For instance, synthetic monthly streamflow records have often been generated by models developed from and designed to preserve the historical monthly record. Experience has shown that if each year's monthly flows are summed to form an annual series, the statistics of the generated annual series do not necessarily resemble the statistics of the historical annual record. This is because any modeling errors, whether due to unreasonable assumptions (i.e., linear correlation, normality, etc.) or poor parameter estimates are concentrated into the resulting annual series. Disaggregation models are designed to overcome the inconsistencies of series generated at different time levels.

Disaggregation modeling is a process by which a key series (such as an annual time series) is broken apart into subseries (smaller time increment series) which add to give the key series values. The key series could itself have been previously generated by an appropriate stochastic model designed to preserve its important statistics. Then generation of the subseries is accomplished by using a model designed to preserve the important statistical properties of not only the subseries itself, but also of the linear relationship between the key and subseries values. In this manner, statistical properties are preserved at both the key and subseries levels and the relationships between the two levels are maintained.

The first well-accepted disaggregation model was presented by Valencia and Schaake in 1973 (37). However, it possessed a number of disadvantages. First, the statistics being preserved for each subseries value were not consistent. For example, if a monthly series was being generated from an annual series, the last monthly value in the year would be generated preserving all the covariances between itself and the eleven preceeding months, while the first monthly value of the following year would be generated without preserving covariances between itself and any preceding months. Also, the number of estimated parameters was large. For an annual to monthly disaggregation, 156 parameters had to be estimated from the historical data. Consequently, the principle of parameter parsimony (section 3.1) was hard to satisfy.

Later in 1976, Mejia and Rousselle (26) modified the Valencia and Schaake model to preserve the covariances between the first subseries value in a year with the last subseries value in the preceding year. Though taking care of one disadvantage of the Valencia and Schaake model, it created other disadvantages: parameter estimation became more complicated and the number of estimated parameters increased. For the annual to monthly disaggregation, 168 parameters were needed.

Then in 1979, Lane (19) developed an approach which essentially sets to zero many parameters of the Mejia and

Rousselle model. His model was developed to generate each subseries value by considering only the correlation between the current subseries value and the preceding one, as well as the correlation between the key series and the subseries. The main advantage of this model is the fewer number of parameters which must be estimated, requiring only 36 for the annual to monthly disaggregation. Also, it is consistent in that Lane's model preserves the lag-one serial correlation between each month and the cross correlation between each month and the corresponding annual streamflow. On the other hand, by reducing the number of parameters, many of the moments preserved by the two earlier models are not directly preserved, although they may be indirectly preserved. Also, Lane's model does not assure that the generated subseries values will add to the key station value. However, this problem is also frequently encountered when using the two earlier models if any data transformation is used, and is easily overcome by the use of correction factors applied to the generated subseries.

6.2 Model Selection and Assumptions

Lane's model was the model chosen to be used in this study because of the ease of parameter estimation and the fewer number of parameters. Had either of the two earlier models been used, the principle of parameter parsimony would have been seriously violated, as the longest record available had 936 annual and monthly values, which results in an index of parameter parsimony (equation 3.4) of roughly 6 for both the Valencia-Schaake and the Mejia-Rousselle models. On the other hand, the index of parameter parsimony for the shortest record (499 monthly and annual values) using Lane's model is 14. Consequently, Lane's model is the only one of the three that satisfies the principle of parameter parsimony (section 3.1).

But, before the model parameters could be determined, a major assumption of Lane's model had to be examined - Lane's disaggregation model assumes that all the monthly and annual series are normally distributed with a mean of zero.

6.3 Normality of Monthly Records

In Chapter 4, the annual series (key) were previously found to approximate a normal distribution. Hence, only the monthly series (subseries) were further examined for normality. The normality of each monthly series was checked by determining if the coefficient of skew of each series was statistically equal to zero. Equation 6.1 (same as equation 3.2) gives the bounds for a coefficient of skew equal to zero at the 95% significance level. Therefore, if the coefficient of skew of a monthly series fell within the 95% limits, the series was considered normal. The results of this test for normality are summarized in Table 6.1.

$$g(95\%) = \pm\ 1.96(6/n)^{1/2} \qquad (6.1)$$

where:

$g(95\%)$ = limit for coefficient of skew equal to zero

n = number of monthly valuels

From Table 6.1, it can be seen that the assumption of normality was not valid for many of the months. This left two options: 1) model the skewed data and account for the skewness in the residual term, or 2) find an appropriate transformation that will convert the skewed sequences into normally distributed sequences. For the second option, transformed sequences must be used for model generation and the inverse transform applied to obtain the actual streamflow values. This option was selected because it was the proceedure recommended by Lane (19) when presenting his model.

6.4 Lognormal Transformations

A lognormal transform has frequently been used to reduce the skewness of hydrologic series. The general logrithmic transform can be expressed as

$$j = \log(x - c) \qquad (6.2)$$

where:

j = transformed monthly streamflow value

x = raw monthly streamflow value

c = constant

Table 6.1

Test for Normality of Monthly Series based on Coefficient of Skew

	12413000 - unextended			12413000 - extended		
Month	Skew Coef	95% Skew Coef	Normal	Skew Coef	95% Skew Coef	Normal
Oct	2.103	.724	No	2.249	.605	No
Nov	1.320	.724	No	2.562	.605	No
Dec	1.150	.724	No	2.696	.605	No
Jan	2.671	.724	No	2.385	.605	No
Feb	1.580	.724	No	1.490	.605	No
March	1.674	.724	No	1.405	.605	No
April	.365	.724	Yes	.339	.605	Yes
May	-.108	-.724	Yes	-.118	-.605	Yes
June	1.233	.724	No	1.180	.605	No
July	.532	.724	Yes	.723	.605	No
Aug	.394	.724	Yes	.567	.600	Yes
Sept	1.234	.724	No	1.602	.600	No

	13186000 - unextended			13186000 - extended		
Month	Skew Coef	95% Skew Coef	Normal	Skew Coef	95% Skew Coef	Normal
Oct	.381	.779	Yes	.359	.566	Yes
Nov	1.041	.779	No	2.104	.566	No
Dec	3.411	.779	No	3.169	.566	No
Jan	1.674	.779	No	1.746	.566	No
Feb	1.361	.779	No	1.354	.566	No
March	1.133	.779	No	.975	.566	No
April	.565	.779	Yes	1.366	.562	No
May	.221	.769	Yes	.427	.562	Yes
June	.351	.769	Yes	.077	.562	Yes
July	.990	.769	No	.822	.562	No
Aug	1.044	.769	No	.710	.562	No
Sept	.706	.769	Yes	.471	.562	Yes

Before applying this transformation, one problem associated with using transforms was considered. Frequently, when transformed series are modeled, the statistics of the transformed series (j) are preserved, but once the inverse transform is applied and the actual series examined, the historical (untransformed) statistics are not preserved.

This problem has prompted the development of formulas for the logarithmic transform which relate the moments of the historical record (x) to those of the transformed record (j) (22, 25). Use of these relationships help to preserve the actual historical statistics. These relationships for the logrithmic transform are listed below and were used to determine the statistics of the transformed series that, in turn, were used in estimating the parameters for the disaggregation models.

<u>For each monthly streamflow series</u>

if $j = \log(x - c)$ and $c = 0$ then

$$\bar{x} = \exp\left(\frac{s_J^2}{2} + \bar{j}\right) \tag{6.3}$$

$$s_x^2 = \exp\left[2(s_J^2 + \bar{j})\right] - \exp(s_J^2 + 2\bar{j}) \tag{6.4}$$

if $j = \log(x - c)$ and $c \neq 0$ then

$$\bar{x} = \exp\left(\frac{s_J^2}{2} + \bar{j}\right) + c \tag{6.5}$$

$$s_x^2 = \exp\left[2(s_J^2 + \bar{j})\right] - \exp(s_J^2 + 2\bar{j}) \tag{6.6}$$

$$g_x = \frac{\exp(3s_J^2) - 3\exp(s_J^2) + 2}{\left(\exp(s_J^2) - 1\right)^{3/2}} \tag{6.7}$$

For correlation between monthly series

if $j_{v-1} = \log(x_{v-1} - c_{v-1})$ and $j_v = \log(x_v - c_v)$

$$r_{xv} = \frac{\exp(s_{J,v-1} s_{J,v} r_{J,v}) - 1}{\left[\exp(s^2_{J,v-1}) - 1\right]^{1/2} \left[\exp(s^2_{J,v}) - 1\right]^{1/2}} \qquad (6.8)$$

if $j_{v-1} = \log(x_{v-1} - c_{v-1})$ and $j_v = x_v$

$$r_{J,v} = \frac{r_{X,v}(\exp(s^2_{J,v-1}) - 1)^{1/2}}{s_{J\,v-1}} \qquad (6.9)$$

For the correlation between monthly and annual values

$j_{v,t} = \log(x_{v,t} - c_v)$ and $j_t = y_t$

$$r_{Jy,v} = \frac{r_{JX,v}(\exp(s^2_{J,v}) - 1)^{1/2}}{s_{J,v}} \qquad (6.10)$$

where:

- j = transformed monthly streamflow value
- x = raw monthly streamflow value
- $\bar{x}$ = mean of raw monthly streamflow
- s_x = standard deviation of raw monthly streamflow
- s_J = standard deviation of transformed monthly streamflow
- g_x = skew coefficient of raw monthly streamflow
- r_x = lag one serial correlation coefficient between raw monthly streamflow values (equation 2.14)
- r_{xy} = cross correlation between raw monthly and annual streamflow values (equation 6.14)
- r_J = lag one serial correlation coefficient between transformed months
- r_{Jy} = cross correlation between transformed monthly and annual streamflow values.
- c = constant

6.5 Normalizing Monthly Streamflow Series

The lognormal transform (equation 6.2) was applied to each monthly streamflow series, both for the case where "c" equaled zero and where "c" equaled the value as computed from equations 6.5 through 6.7. Next, the skew coefficient of the transformed series was calculated in order to determine if it was reduced enough such that the assumption of normality could be satisfied. The computed skew coefficients for each transformed sequence are listed in Table 6.2.

The series corresponding to the skew coefficient in Table 6.2 marked with an "*" were the series used for modeling, although in several cases, the transformation did not produce a skew coefficient statistically equal to zero. However, it was decided to use these transforms since appropriate relationships exist relating the transformed and historical statistics, and such relationships for other transforms were not as readily available.

Based on the transformation selected in Table 6.2 for each monthly series, the appropriate relationships (equations 6.3 through 6.9) were solved to obtain the statistics of the transformed sequences. The resulting statistics are listed in Tables 6.3 and 6.4.

Table 6.2

Coefficients of Skew for Transformed and Untransformed Monthly Streamflow

	12413000 - unextended			12413000 - extended		
Month	j=x	j = log(x)	j = log(x-c)	j=x	j = log(x)	j = log(x-c)
Oct	2.103	+*1.3882	1.4290	2.249	+*1.1897	1.2416
Nov	1.320	* .1800	.6962	2.562	*.3880	.7074
Dec	1.150	* .0114	.7366	2.696	*.2002	.7144
Jan	2.671	-.2863	*-.1409	2.385	-.5414	*.2898
Feb	1.580	* .1410	.7059	1.490	*-.1656	.6386
Mar	1.674	-.3033	*-.0733	1.405	-.4072	*.0275
Apr	.365	-.4074	* .1397	.339	-.4245	*.1358
May	*-.108	-.8039	-.1561	*-.118	-.8925	-.1765
June	1.233	*-.0402	.3164	1.180	*-.2586	.3192
July	.532	*-.1142	.2565	.723	*-.0778	.2715
Aug	.394	-.1898	*.1033	.567	*-.0584	.1328
Sept	1.234	.6139	*.2077	1.602	.8662	*.2757

	13186000 - unextended			13186000 - extended		
Month	j=x	j = log(x)	j = log(x-c)	j=x	j = log(x)	j = log(x-c)
Oct	.381	*.0072	.1326	.3595	*.0312	.1614
Nov	1.041	.5190	*.1394	2.1036	.9889	*-.5261
Dec	3.411	1.8866	*.2525	3.1686	1.1634	-----
Jan	1.674	1.1560	*.6204	1.7464	1.0368	* .1260
Feb	1.361	.7975	*.2540	1.3538	.5295	*-.3742
Mar	1.133	.2244	*.1317	.9747	*.2229	.2693
Apr	.565	-.4014	*.2311	1.3655	-.1670	* .0621
May	.221	-1.4460	*.0716	.4269	-.8446	* .1234
June	.351	-.8818	*.1129	.0772	-2.0089	* .0286
July	.990	*-.1514	.4029	.8221	-.8289	* .2517
Aug	1.044	* .0435	.0800	.7104	-.3394	* .0656
Sept	.706	* .1869	.2237	.4714	-.1690	* .1173

* Transformation which produced a coefficient of skew closest to zero.

--- produced values for which the logrithum was undefined

\+ Skew coefficient is not statistically equal to zero.

Table 6.3

Statistics of Transformed Values
as Calculated from Moment Relationships

12413000 - unextended

Month	Mean	Standard Deviation	Lag-1 Ser Corr Coef	c
October	9.255	.5974	.5315	0
November	9.933	.6829	.7396	0
December	10.565	.7423	.5893	0
January	10.547	.6679	.4509	-2070
Feburary	10.717	.6904	.1940	0
March	11.228	.4833	.4095	-8422
April	13.118	.1206	.0032	-339862
May	173468	72660	.4614	0
June	10.915	.4971	.7119	0
July	9.892	.3553	.7779	0
August	10.000	.1300	.8657	-11263
September	8.447	.3775	.5616	3938

12413000 - extended

Month	Mean	Standard Deviation	Lag-1 Ser Corr Coef	c
October	9.254	.6206	.7102	0
November	9.888	.7906	.8180	0
December	10.473	.8515	.5072	0
January	10.771	.6217	.6663	-13259
Feburary	10.626	.7140	.2931	0
March	11.381	.4206	.4712	-19179
April	13.175	.1122	.0662	-362395
May	167862	70537	.4272	0
June	10.816	.5409	.7438	0
July	9.798	.3734	.8010	0
August	9.205	.2668	.8817	0
September	8.297	.4672	.4811	4283

Table 6.4

Statistics of Transformed Values
as Calculated from Moment Relationships

13186000 - unextended

Month	Mean	Standard Deviation	Lag-1 Ser Corr Coef	c
October	8.919	.1832	.7901	0
November	8.310	.3257	.5483	3345
December	7.726	.7681	.2558	4899
January	8.022	.4833	.8414	4332
Feburary	8.175	.4098	.6099	3541
March	9.204	.3508	.3772	1041
April	11.477	.1846	.3557	-59835
May	13.012	.0734	.6401	-363396
June	12.528	.1161	.6866	-201636
July	10.040	.5204	.9461	0
August	9.101	.3382	.9253	0
September	8.840	.2446	.9047	0

13186000 - extended

Month	Mean	Standard Deviation	Lag-1 Ser Corr Coef	c
October	8.865	.1972	.8741	0
November	7.741	.5714	.6509	4652
December	8.881	.3023	.3122	0
January	7.835	.4991	.8155	4503
Feburary	8.087	.4080	.6776	3324
March	9.240	.3287	.4866	0
April	10.562	.4110	.4429	-5990
May	12.336	.1407	.5507	-150363
June	14.159	.0257	.6816	-1339774
July	10.875	.2630	.8936	-31230
August	9.545	.2295	.9224	-5727
September	9.364	.1549	.9238	-5211

6.6 Standardizing Monthly Streamflow Values

Lane's model also assumes that the means of the normally distributed series equal zero. This assumption was satisfied by subtracting the means of the monthly transformed series (listed in Table 6.4). In addition, each transformed value was divided by its transformed standard deviation (Table 6.4) to create standardized series. Since the means and standard deviations of the transformed series as computed by equations 6.3 through 6.7 were used, the means and standard deviations of the standardized series may actually differ slightly from zero and one. But, as previously mentioned, the use of these relationships should help to preserve the statistics of the original historical sequences. Equation 6.11 illustrates the steps taken to arrive at the series actually used in modeling.

$$J = \frac{\log(x-c) - \bar{j}}{s_J} \quad (6.11)$$

where:

J = normally distributed monthly streamflow value with mean of zero and standard deviation of one.
x = historical monthly streamflow value
c = may or may not equal zero
log = transform used, except for several months where $\log(x-c)$ would be replaced simply by x
$\bar{j}$ = mean of transformed series
s_J = standard deviation of transformed series.

6.7 Lane's Disaggregation Model

Lane's model for an annual to monthly disaggregation may be written as

$$J_{v,t} = Q_v Y_t + G_{v,t} \lambda_{v,t} + H_v J_{v-1,t} \quad \text{(for v = 1 to 12)} \qquad (6.12)$$

where:

J_v = standardized, normalized monthly streamflow value (if v=1 then v-1 = 12 and t = t-1)
Y = prexisting normalized, standardized annual value corresponding to same year as monthly J value.
v = current month
t = year
Q,G,H = model parameters, change for each month

This model is designed to preserve the linear cross correlation between annual and monthly values along with the lag-one correlations, variances and means of the annual and monthly values. It accomplishes this by preserving the means and standard deviations through normalization while the correlation structure is preserved by the actual model. In order to use this model, the parameters Q, G, and H of equation 6.12 must first be estimated for each month of each time series. For the one-station temporal model using normalized and standardized sequences, the parameters can be estimated as follows:

$$Q_v = \frac{r_{yJ,v} - r_{J,v} \quad r_{yJ,v-1}}{1 - r^2_{yJ,v-1}} \qquad (6.13)$$

$$H_v = r_{J,v} - Q_v \, r_{yJ,v-1}$$

$$G_v = 1 - Q_v \, r_{yJ,v} - H_v \, r_{J,v}$$

where:

$r_{y_J,v}$ = correlation coefficient between each month and corresponding annual value

$r_{y_J,v}$ = correlation coefficient between previous monthly value and corresponding yearly value

$r_{J,v}$ = lag one serial correlation coefficient between monthly value

v = month

6.8 Monthly/Annual Correlation Coefficients

In order to estimate the parameters Q, G, and H the correlation coefficient between each month and corresponding annual streamflow value had to be calculated. Equation 6.14 was used to calculate these correlation coefficients between the untransformed monthly and annual values.

$$r_{y_J,v} = \frac{\sum_{i=1}^{n} (x_{v,t} - \bar{x}_v)(y_t - \bar{y})}{s_{x,v}\, s_y} \qquad (6.14)$$

where:

v = month

t = year

x = raw monthly streamflow value

y = raw annual streamflow value

$\bar{x}$ = mean of monthly streamflow

$\bar{y}$ = mean of annual streamflow

s_x = standard deviation of monthly streamflow

s_y = standard deviation of annual streamflow

However, the correlation coefficients as calculated from equation 6.14 were inappropriate for most of the monthly and corresponding annual series (except for the month of May at Station 12413000), since logarithmic transformations were used. Hence, once again, the relationships as developed for the logarithmic transform relating the transformed and historical statistics were utilized. In this particular case, the annual series was to be untransformed, while the monthly series was to be transformed. Thus, equation 6.10 was used to arrive at the correlation coefficients to be used in estimating the model parameters. The resulting correlation coefficients are listed in Table 6.5.

Table 6.5

Correlation Coefficients between Monthly and Annual Values

	12413000 Unextended		12413000 Extended		13186000 Unextended		13186000 Extended	
Mon	raw r_{xy}	trans r_{Jy}	raw r_{xy}	trans r_{Jy}	raw r_{xy}	trans r_{Jy}	raw r_{xy}	trans r_{Jy}
Oct	.3322	.3642	.3884	.4290	-.0560	-.0565	-.0079	-.0080
Nov	.5496	.6203	.5288	.6233	.4001	.4109	.3145	.3420
Dec	.4176	.4823	.4832	.5856	.6000	.7004	.5064	.5182
Jan	.5896	.6618	.6513	.7196	.7766	.8242	.6882	.7334
Feb	.2969	.3361	.3621	.4136	.6149	.6416	.5590	.5831
Mar	.3487	.3701	.4338	.4537	.5568	.5744	.5215	.5359
Apr	.6587	.6611	.5730	.5748	.6541	.6597	.6315	.6591
May	.6968	.6968	.6304	.6304	.8937	.8949	.8818	.8862
Jun	.6149	.6549	.5587	.6022	.8985	.9015	.9018	.9019
Jul	.5797	.5985	.5379	.5572	.8266	.8858	.8585	.8736
Aug	.5938	.5964	.5618	.5720	.8649	.8902	.8927	.9046
Sep	.2182	.2262	.2341	.2378	.7460	.7573	.8102	.8152

6.9 Disaggregation Model Parameters

The parameters Q, G, and H for Lane's disaggregation model were calculated using equations 6.13 and the statistics in Tables 6.3, 6.4, and 6.5. The resulting parameters are listed in Table 6.6.

Table 6.6

Parameters for Lane's Disaggregation Model

Month	12413000 - Unextended Q	G	H	12413000 - Extended Q	G	H
October	.2571	.8092	.4734	.2758	.6510	.6446
November	.4046	.5577	.5922	.3338	.4898	.6748
December	.1898	.7941	.4716	.4406	.7900	.2325
January	.5791	.7344	.1716	.5014	.6252	.3726
Feburary	.3696	.9411	-.0506	.4204	.9104	-.0094
March	.2621	.8783	.3214	.3122	.8349	.3421
April	.7647	.7038	-.2799	.6860	.7886	-.2450
May	.6960	.7172	.0013	.5747	.7722	.0969
June	.3087	.6665	.4968	.2212	.6459	.6044
July	.1560	.6172	.6758	.1174	.5912	.7303
August	.1219	.4909	.7928	.1170	.4617	.8165
September	-.1687	.8162	.6622	-.0556	.8755	.5129

Month	13186000 - Unextended Q	G	H	13186000 - Extended Q	G	H
October	-1.5352	-----	1.9527	-2.1481	-----	2.6252
November	.4433	.7096	.5733	.3473	.6751	.6537
December	.7162	.7129	-.0385	.4659	.8431	.1528
January	.4612	.4286	.5184	.4249	.4504	.5953
Feburary	.4333	.7536	.2528	.1865	.7244	.5408
March	.5651	.8185	.0146	.3821	.8166	.2638
April	.6797	.7510	-.0347	.5917	.7445	.1258
May	.8368	.4414	.0880	.9252	.4611	-.0591
June	1.4415	.3387	-.6034	1.3878	.3493	-.5483
July	.1758	.3148	.7876	.3626	.4207	.5666
August	.3275	.3474	.6352	.4172	.3286	.5580
September	-.2316	.4128	1.1108	-.1126	.3799	1.0256

At station 13186000, "G^2" for the month of October was undefined (negative). This seemed to be reasonable since the correlation coefficient between the month of October and the corresponding annual flows was very small (statistically equal to zero). Therefore, Lane's model, which accounts for this correlation, is inappropriate for this month. On the other hand, the lag-one serial correlation coefficient for the month of October was quite large. As a result, a simple AR(1) model was used for the month of October at Station 13186000, while Lane's model was used for the other eleven months.

6.10 Residuals

The residuals of the disaggregation models were not tested for normality and independence due to the transformations and monthly flow corrections which would distort the residual values. Instead, synthetic records were generated and their performance evaluated.

6.11 Generation of Synthetic Records

In order to check the performance of Lane's model, 500 years of monthly streamflow values were generated for each record: 12413000 - Unextended, 12413000 - Extended, 13186000 - Unextended, and 13186000 - Extended. The steps taken to generate these synthetic records are listed below:

1) Five hundred and fifty five annual streamflow values were generated using the following equations:

For Station 12413000 - Unextended

$$y_t = 706854 + 198977\lambda_{n,t} \qquad \text{(cfs days)}$$

For Station 12413000 - Extended

$$y_t = 933877 + 258015\lambda_{n,t} \qquad \text{(cfs days)}$$

For Station 13186000 - Unextended

$$y_t = 294519 + 94040\lambda_{n,t} \qquad \text{(cfs days)}$$

For Station 13186000 - Extended

$$y_t = 442814 + 130801\lambda_{n,t} \qquad \text{(cfs days)}$$

where:

λ_n = standard random normal deviate (same seed value was used for each 555 year sequence).

2) The first 50 values were discarded to avoid any startup bias.

3) The annual flow series generated in step 1 were then standardized.

$$Y_t = (y_t - \bar{y})/s_y$$

where:

y = generated annual streamflow value from step one (years 51 - 555)
$\bar{y}$ = mean annual streamflow
s_y = standard deviation of annual streamflow

Steps 4 through 7 were repeated for each annual value from step 3.

4) Lane's disaggregation model was applied:

$$J_{v,t} = Q_v Y_t + G_v \lambda_{v,t} + H_v J_{v-1,t} \quad \text{(for v = 1 to 12)} \qquad (6.12)$$

where:

Y = Standardized annual streamflow value from step 3
v = current month
Q, G, H = model parameters (Table 6.6)
λ = standard random normal deviate
J = previous month's flow as generated from Lane's model. Initial J taken as its expected value: 0. (if v=1 then t = t-1)

5) The inverse transforms were used to arrive at the actual monthly streamflow values. In most cases the inverse transform was: (except for the month of May at Station 12413000: x = J s + j)

$$x_{v,t} = \exp(J_{v,t}\ s_{Jv} + \bar{j}_v) + c_v$$

where:

s_J = standard deviation of transformed series (Tables 6.3 and 6.4)
$\bar{j}$ = mean of transformed series (Tables 6.3 and 6.4)
J = standardized, normalized monthly flow generated from step 4.
c = constant (Table 6.3 and 6.4)

If a negative monthly streamflow value was generated, it was retained for generating the next month's flow and then replaced by a positive generated value which was obtained by repeating step 4.

6) The monthly flows were adjusted such that they would sum to the annual streamflow value (y).

$$x_{v,t}* = \frac{x_{v,t} + (y_t - \sum_{v=1}^{2} x_{v,t})\, s_{x,v}}{\sum_{v=1}^{2} s_{x,v}}$$

Howver, in a few cases this correction produced negative values, in which case the following alternate adjustment was used

$$x_{v,t}* = \frac{x_{v,t}\ \ y_t}{\sum_{v=1}^{12} x_{v,t}}$$

where:

x^* = the adjusted monthly streamflow value
y = annual streamflow value (step 1)
x = monthly streamflow value from step 5
s_x = historical monthly standard deviation

7) The last month's value was normalized and standardized as it was used as the initial $J_{v-1,t}$ for the next year's disaggregation model.

8) The first 5 years of monthly streamflow values were discarded in order to avoid any startup bias.

6.12 Statistics of Synthetic Records

The statistics of the synthetic records were computed and compared to the statistics of the historical record. Hypothesis tests were performed on the means and standard deviations using the t- and F-statistic, respectively (Table 2.2). The resulting statistics are summarized in Tables 6.7 through 6.10 while the results of the hypothesis tests are presented in Tables 6.11 through 6.14

Table 6.7

Statistics for Synthetic Record
at Station 12413000 - Unextended
(Streamflow in cfs days)

Period	Mean	Variance	Standard Deviation	Skew	Coef Skew	n
October	12697	$5.960X10^7$	7720	$7.894X10^{11}$	1.716	500
November	26780	$3.600X10^8$	18973	$1.252X10^{13}$	1.834	500
December	52711	$1.599X10^9$	39984	$1.316X10^{14}$	2.059	500
January	44695	$8.233X10^8$	28693	$2.741X10^{13}$	1.160	500
February	55974	$1.600X10^9$	39994	$1.430X10^{14}$	2.236	500
March	76074	$1.465X10^9$	38270	$4.451X10^{13}$	0.794	500
April	160797	$3.442X10^9$	58667	$5.096X10^{13}$	0.252	500
May	178770	$4.360X10^9$	66032	$-.214X10^{14}$	-.074	500
June	64540	$1.079X10^9$	32846	$4.995X10^{13}$	1.410	500
July	21677	$6.195X10^7$	7871	$5.481X10^{11}$	1.124	500
August	11151	$8.309X10^6$	2883	$5.038X10^9$	0.210	500
September	8988	$3.913X10^6$	1978	$1.030X10^{10}$	1.330	500
Annual	714855	$3.762X10^{10}$	193966	$7.616X10^{14}$	0.104	500

Serial Correlation coefficients

Period	Lag (k) 1	2	3	4	5
October	.4434	.3547	.2691	.1879	.1722
November	.6999	.2922	.3569	.3065	.2657
December	.5181	.3014	.0961	.2473	.1957
January	.2110	.0814	.0732	.0398	.1337
February	.0299	.0072	-.0045	.0391	-.0067
March	.2358	.2476	.0351	.0121	.0675
April	-.0730	.0896	.2689	.0663	.1018
May	.3515	.1974	.0500	.3780	.1153
June	.6256	.3560	.1175	.0432	.2728
July	.7523	.5145	.3708	.1142	.0405
August	.8410	.6435	.5234	.3824	.1470
September	.5508	.4395	.2644	.2146	.1678
Annual	.0976	-.0758	-.0269	.0166	.0373

Table 6.8

Statistics for Synthetic Record
at Station 12413000 - Extended
(Streamflow in cfs days)

Period	Mean	Variance	Standard Deviation	Skew	Coef Skew	n
October	12477	$6.812X10^{7}$	8254	$1.358X10^{12}$	2.414	500
November	27000	$4.827X10^{8}$	21971	$2.179X10^{13}$	2.054	500
December	55018	$2.574X10^{9}$	50730	$4.083X10^{14}$	3.128	500
January	44936	$1.101X10^{9}$	33187	$5.204X10^{13}$	1.424	500
February	54112	$1.603X10^{9}$	40033	$2.266X10^{14}$	3.533	500
March	79999	$1.779X10^{9}$	42180	$8.607X10^{13}$	1.147	500
April	168018	$3.380X10^{9}$	58139	$4.498X10^{13}$	0.229	500
May	168121	$3.969X10^{9}$	62997	$2.233X10^{12}$	0.009	500
June	56862	$9.474X10^{8}$	30779	$4.197X10^{13}$	1.439	500
July	19054	$4.819X10^{7}$	6942	$3.844X10^{11}$	1.149	500
August	10206	$6.693X10^{6}$	2587	$1.003X10^{10}$	0.579	500
September	8748	$4.976X10^{6}$	2231	$1.690X10^{10}$	1.522	500
Annual	704551	$4.006X10^{10}$	200145	$8.367X10^{14}$	0.104	500

Serial Correlation coefficients

Period	Lag (k) 1	2	3	4	5
October	.6896	.3602	.2805	.1454	.1541
November	.7617	.5196	.4027	.3393	.2810
December	.4168	.2337	.0827	.1771	.1338
January	.3280	.1554	.0760	.0724	.1430
February	.0951	-.0945	-.0528	-.0667	-.0297
March	.3108	.2227	-.0032	-.0005	.0004
April	-.0218	.1127	.2453	-.0394	.0461
May	.3901	.0985	.0487	.2809	-.0191
June	.6443	.2999	.1289	.0639	.2594
July	.7373	.5374	.2691	.1429	.0765
August	.8465	.6437	.5263	.3494	.1518
September	.4035	.3116	.1739	.1404	.1372
Annual	.0976	-.0758	-.0269	.0166	.0373

Table 6.9

Statistics for Synthetic Record
at Station 13186000 - Unextended
(Streamflow in cfs days)

Period	Mean	Variance	Standard Deviation	Skew	Coef Skew	n
October	7681	$2.115X10^6$	1454	$9.267X10^{12}$	0.301	500
November	7695	$2.179X10^6$	1476	$3.610X10^{13}$	1.122	500
December	8045	$7.807X10^6$	2794	$5.809X10^{14}$	2.663	500
January	7827	$3.162X10^6$	1778	$6.143X10^{13}$	1.093	500
February	7415	$2.751X10^6$	1659	$4.382X10^{14}$	0.960	500
March	11485	$1.444X10^7$	3799	$4.817X10^{13}$	0.878	500
April	39480	$2.594X10^8$	16108	$1.363X10^{13}$	0.326	500
May	87187	$1.019X10^9$	31918	$-.209X10^{12}$	-.064	500
June	77756	$9.808X10^8$	31318	$2.188X10^{13}$	0.071	500
July	26912	$1.903X10^8$	13796	$3.659X10^{11}$	1.394	500
August	9622	$9.721X10^6$	3118	$1.630X10^{10}$	0.538	500
September	7194	$2.867X10^6$	1693	$1.520X10^{10}$	0.313	500
Annual	298300	$8.404X10^9$	91672	$8.040X10^{14}$	0.104	500

Serial Correlation coefficients

Period	Lag (k) 1	2	3	4	5
October	.7576	.6818	.5996	.5744	.4519
November	.5747	.4213	.3800	.3259	.3335
December	.2399	.0406	.0254	.0437	.0232
January	.7945	.3543	.0534	.0503	.0594
February	.6049	.4424	.2941	.0888	.0444
March	.3994	.3991	.2985	.3077	.0832
April	.2756	.3908	.4772	.3558	.2779
May	.5819	.5540	.5742	.6713	.5065
June	.6716	.5442	.4787	.5679	.7046
July	.8972	.6531	.5202	.4300	.5226
August	.9021	.8697	.7158	.5256	.4890
September	.8784	.7600	.7465	.5849	.3998
Annual	.0976	-.0758	-.0269	.0166	.0373

Table 6.10

Statistics for Synthetic Record
at Station 13186000 - Extended
(Streamflow in cfs days)

Period	Mean	Variance	Standard Deviation	Skew	Coef Skew	n
October	7214	$2.373X10^6$	1540	$-.363X10^8$	-.010	500
November	7366	$3.188X10^6$	1786	$6.737X10^9$	1.183	500
December	7592	$5.562X10^6$	2358	$9.615X10^9$	0.733	500
January	7399	$2.808X10^6$	1676	$4.670X10^9$	0.992	500
February	6803	$2.703X10^6$	1644	$4.964X10^9$	1.117	500
March	10952	$1.267X10^7$	3560	$1.372X10^{10}$	0.304	500
April	36505	$3.050X10^8$	17464	$8.887X10^{12}$	1.668	500
May	81403	$1.023X10^9$	31981	$8.870X10^{12}$	0.271	500
June	71307	$1.135X10^9$	33692	$4.547X10^{12}$	0.119	500
July	23760	$1.823X10^8$	13503	$1.652X10^{12}$	0.671	500
August	8685	$1.065X10^7$	3264	$1.519X10^{10}$	0.437	500
September	6590	$3.218X10^6$	1794	$1.118X10^9$	0.194	500
Annual	275576	$8.713X10^9$	93346	$8.488X10^{13}$	0.104	500

Serial Correlation coefficients

Period	Lag (k) 1	2	3	4	5
October	.8155	.7532	.6810	.6239	.5460
November	.6886	.5429	.4910	.4355	.4067
December	.3834	.2343	.1252	.1247	.1228
January	.7911	.4370	.2579	.1268	.1185
February	.7094	.5320	.3370	.2241	.0931
March	.4454	.4644	.3278	.2184	.0908
April	.4394	.3148	.3934	.2628	.1969
May	.4570	.5004	.4939	.6391	.4646
June	.6561	.5470	.4886	.5052	.6149
July	.8685	.6848	.5146	.4640	.4900
August	.9145	.8609	.7454	.5319	.5226
September	.9158	.8236	.7650	.6742	.4700
Annual	.0976	-.0758	-.0269	.0166	.0373

Table 6.11

Hypothesis Tests for Equality of Synthetic and Historical Variances

Station 12413000 - Unextended

Period	s_m (cfsd)	s_n (cfsd)	m (yrs)	n (yrs)	Sample F	95% F
October	8182	7720	44	500	1.12	1.50
November	20052	18973	44	500	1.12	1.50
December	43759	39984	44	500	1.20	1.50
January	35669	28693	44	500	*1.54	1.50
February	44747	39994	44	500	1.25	1.50
March	43353	38270	44	500	1.28	1.50
April	60740	58667	44	500	1.07	1.50
May	72660	66032	44	500	1.21	1.50
June	32945	32846	44	500	1.01	1.50
July	7871	7729	500	44	1.04	1.66
August	2901	2883	44	500	1.01	1.50
September	1978	1959	500	44	1.02	1.66
Annual	198977	193966	44	500	1.05	1.50

Station 12413500 - Extended

Period	s_m (cfsd)	s_n (cfsd)	m (yrs)	n (yrs)	Sample F	95% F
October	8681	8254	63	500	1.11	1.41
November	25090	21971	63	500	1.30	1.41
December	52424	50730	63	500	1.07	1.41
January	39677	33187	63	500	*1.43	1.41
February	43327	40033	63	500	1.17	1.41
March	42180	42120	500	63	1.00	1.51
April	59668	58139	63	500	1.05	1.41
May	70537	62997	63	500	1.25	1.41
June	33609	30779	63	500	1.19	1.41
July	7466	6942	63	500	1.16	1.41
August	2798	2587	64	500	1.17	1.41
September	2231	2210	500	64	1.02	1.51
Annual	205315	200145	63	500	1.05	1.41

* Standard deviations that were not statistically equal.

Table 6.12

Hypothesis Tests for Equality of Extended and Historical Means

Station 12413000 - Unextended

Period	$\bar{x}_m$ (cfsd)	$\bar{x}_n$ (cfsd)	m (yrs)	n (yrs)	Sp (cfsd)	Sample t	95% t
October	12697	12493	500	44	7758	.167	1.96
November	26780	26013	500	44	19061	.256	1.96
December	52711	51037	500	44	40296	.264	1.96
January	45501	44695	44	500	-	.146	2.01
February	57264	55974	44	500	40392	.203	1.96
March	76087	76074	44	500	38698	.002	1.96
April	161810	160797	44	500	58834	.109	1.96
May	178770	173468	500	44	66582	.506	1.96
June	64540	62219	500	44	32854	.449	1.96
July	21677	21068	500	44	7860	.493	1.96
August	11151	10951	500	44	2884	.441	1.96
September	8988	8944	500	44	1976	.142	1.96
Annual	714855	706854	500	44	194368	.262	1.96

Station 12413000 - Extended

Period	$\bar{x}_m$ (cfsd)	$\bar{x}_n$ (cfsd)	m (yrs)	n (yrs)	Sp (cfsd)	Sample t	95% t
October	12666	12477	63	500	8302	.170	1.96
November	27000	26926	500	63	22337	.025	1.96
December	55018	50805	500	63	50920	.619	1.96
January	44936	44500	504	63	-	.084	1.99
February	54112	53134	504	63	40410	.181	1.96
March	79999	76559	504	63	42173	.610	1.96
April	168018	167869	500	63	58310	.019	1.96
May	168121	167862	500	63	63874	.030	1.96
June	57651	56862	63	500	31104	.190	1.96
July	19300	19054	63	500	7002	.263	1.96
August	10303	10206	63	500	2612	.280	1.96
September	8758	8748	63	500	2229	.034	1.96
Annual	704551	696295	500	63	260723	.238	1.96

Table 6.13

Hypothesis Tests for Equality of Synthetic and Historical Variances

Station 13186000 - Unextended

Period	s_m (cfsd)	s_n (cfsd)	m (yrs)	n (yrs)	Sample F	95% F
October	1454	1403	500	38	1.07	1.74
November	1476	1433	500	38	1.06	1.74
December	2794	2731	500	38	1.05	1.74
January	1778	1756	500	38	1.02	1.74
February	1659	1651	500	38	1.01	1.74
March	3823	3799	38	500	1.01	1.54
April	18269	16108	38	500	1.29	1.54
May	33029	31918	39	500	1.07	1.53
June	32368	31318	39	500	1.07	1.53
July	14642	13796	39	500	1.13	1.53
August	3303	3118	39	500	1.12	1.53
September	1767	1693	39	500	1.09	1.53
Annual	94040	91672	38	500	1.05	1.54

Station 13186000 - Extended

Period	s_m (cfsd)	s_n (cfsd)	m (yrs)	n (yrs)	Sample F	95% F
October	1540	1437	500	72	1.15	1.49
November	1786	1683	500	72	1.13	1.49
December	2358	2330	500	72	1.02	1.49
January	1676	1523	500	72	1.21	1.49
February	1644	1504	500	72	1.19	1.49
March	3673	3560	72	500	1.06	1.39
April	18034	17464	73	500	1.07	1.39
May	32532	31981	73	500	1.03	1.39
June	36275	33692	73	500	1.16	1.39
July	14635	13503	73	500	1.17	1.39
August	3337	3264	73	500	1.04	1.39
September	1839	1794	73	500	1.05	1.39
Annual	95757	93346	72	500	1.05	1.39

Table 6.14

Hypothesis Tests for Equality of Extended and Historical Means

Station 13186000 - Unextended

Period	$\bar{x}_m$ (cfsd)	$\bar{x}_n$ (cfsd)	m (yrs)	n (yrs)	Sp (cfsd)	Sample t	95% t
October	7681	7596	500	38	1450	.348	1.96
November	7695	7629	500	38	1473	.266	1.96
December	8045	7945	500	38	2790	.213	1.96
January	7827	7755	500	38	1778	.241	1.96
February	7415	7402	500	38	1658	.046	1.96
March	11606	11485	38	38	3801	.189	1.96
April	39480	38289	500	38	16266	.435	1.96
May	87187	85770	500	39	31998	.266	1.96
June	77756	76265	500	39	31393	.286	1.96
July	26912	26255	500	39	13858	.285	1.96
August	9622	9488	500	39	3131	.257	1.96
September	7194	7115	500	39	1698	.280	1.96
Annual	298300	294519	500	38	91837	.245	1.96

Station 13186000 - Extended

Period	$\bar{x}_m$ (cfsd)	$\bar{x}_n$ (cfsd)	m (yrs)	n (yrs)	Sp (cfsd)	Sample t	95% t
October	7217	7214	72	500	1528	.015	1.96
November	7366	7361	500	72	1773	.022	1.96
December	7592	7532	500	72	2355	.202	1.96
January	7399	7367	504	72	1658	.153	1.96
February	6858	6803	72	500	1627	.268	1.96
March	10952	10873	500	72	3574	.175	1.96
April	36505	36046	500	73	17537	.209	1.96
May	81403	79775	500	73	32051	.405	1.96
June	71307	70187	500	73	34028	.263	1.96
July	23760	23451	500	73	13651	.181	1.96
August	8685	8619	500	73	32734	.161	1.96
September	6590	6587	500	73	1800	.013	1.96
Annual	275576	271726	500	72	93650	.326	1.96

Based on the hypothesis tests (Tables 6.11 to 6.14), only the standard deviations for January of the unextended and extended records at stations 12413000 were statistically unequal to the corresponding historical standard deviations. Sampling fluctuations seemed to be the reason for these differences because:

1) The January series at station 12413000 were sucessfully normalized, whereas the October series was not. Yet the statistics of October series were all statistically equal to their corresponding historical statistics. Therefore, the differences of the standard deviations was not caused by nonnormality.

2) There was nothing unusual about January's historical (Tables B.5, B.7) or transformed statistics (Table 6.4). They were neither the largest or smallest values, except the skew coefficient of the unextended record which was the largest. However, the skew coefficient for October of the same record was not much less than that of January. Hence, there was nothing unusual about the series statistics to suggest a modeling problem.

3) There was nothing unusual about the disaggreagation model parameters, as they were not extreme values (Table 6.6).

4) The standard deviations were just barely significantly different.

5) The fact that only January's standard deviations were statistically different at both the unextended and extended record was probably caused by using the same seed value for the random number generator. In other words, the same series of random numbers was used for both synthetic records, and therefore, they affected the same month in similar ways.

In conclusion, the synethic records were found to satisfactorally preserve the historical statistics. Therefore, Lane's disaggregation model was accepted for further use.

6.13 Conclusions

As can be seen by comparing the disaggregation model parameter estimates of the unextended and corresponding extended records (Table 6.6), the parameters of the extended series differed from those of the unextended series. This was to be expected since the historical statistics of these records changed after data extension (compare Table B.5 and B.7 with C.7 and C.8, respectively). As previously explained in section 3.5, the extended record statistics are a more reliable description of the streamflow process. Consequently, the disaggregation parameters estimated from these statistics would be more reliable. Therefore, data extension should be performed before model parameters are estimated if an appropriate record of longer length which is strongly correlated to the original record is available.

Figures 6.1 through 6.3 compare the disaggregation model parameters of the extended series at stations 12413000 and 13186000 to see how these parameters changed with location. The extended series were compared since they were felt to more accurately describe the respective streamflow processes. After examining these figures the following observations were made:

1) The parameter "Q" which relates the annual and monthly streamflow values followed the same general pattern except for the months of May and June. April and May are the months of maximum runoff at station 12413000, while both May and June have the heaviest runoff at station 13186000. The higher elevation and different climate at station 13186000 probably results in a later runoff series (extending between May and June) than at stations 12413000 (extending between April and May) possibly explaining some of the deviation in the "Q" parameters.

2) The parameter "G" which relates the residual and monthly streamflow values followed the same general pattern.

3) The parameter "H" which relates sucessive monthly streamflow values seemed to deviate more from any general pattern relative to the other two parameters.

Based on the preceding observations it was felt that the possibility of regionalizing disaggregation model

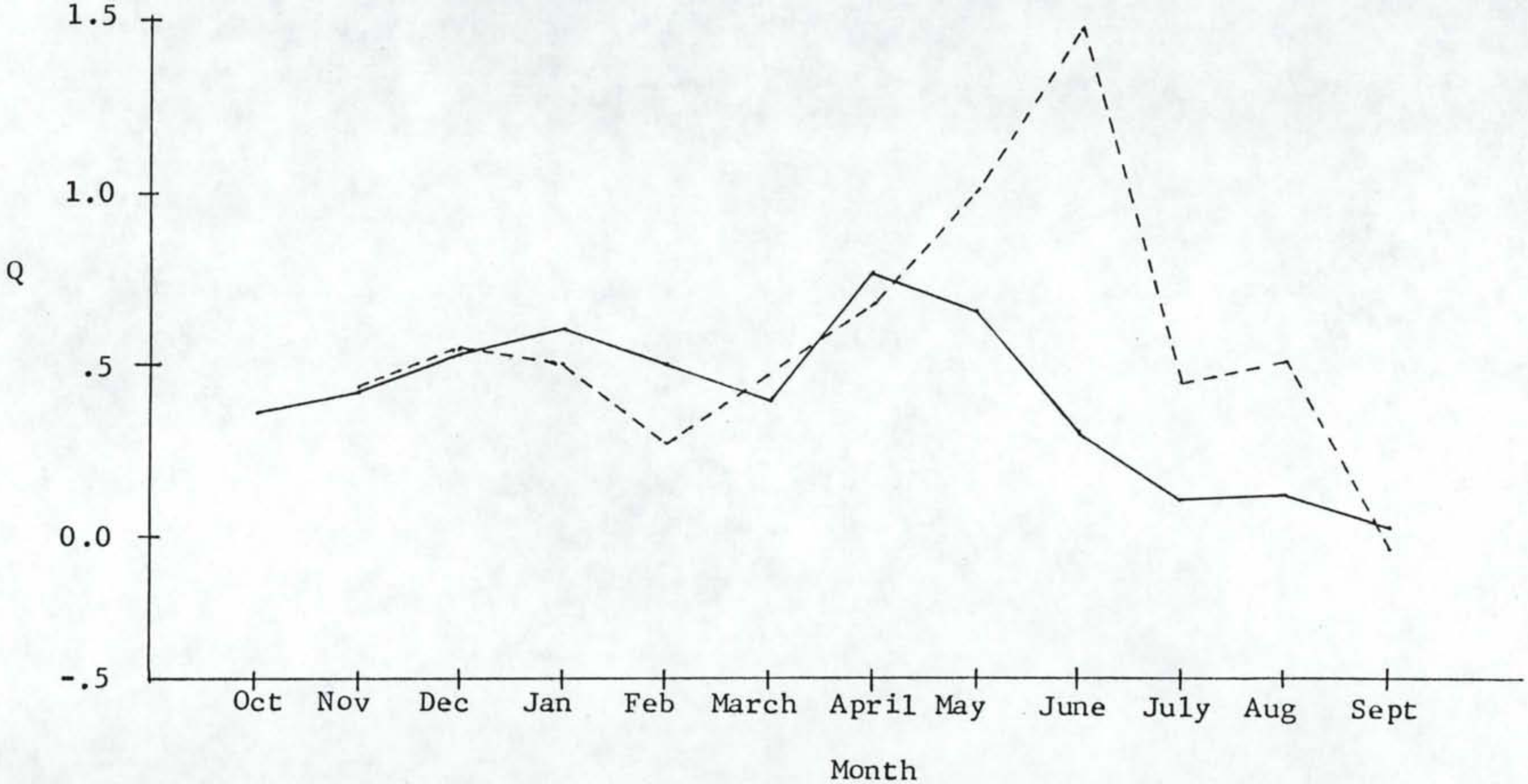


Figure 6.1 Comparison of Lane's Monthly Disaggregation Parameter "Q" from the Extended Records at Stations 12413000 and 13186000

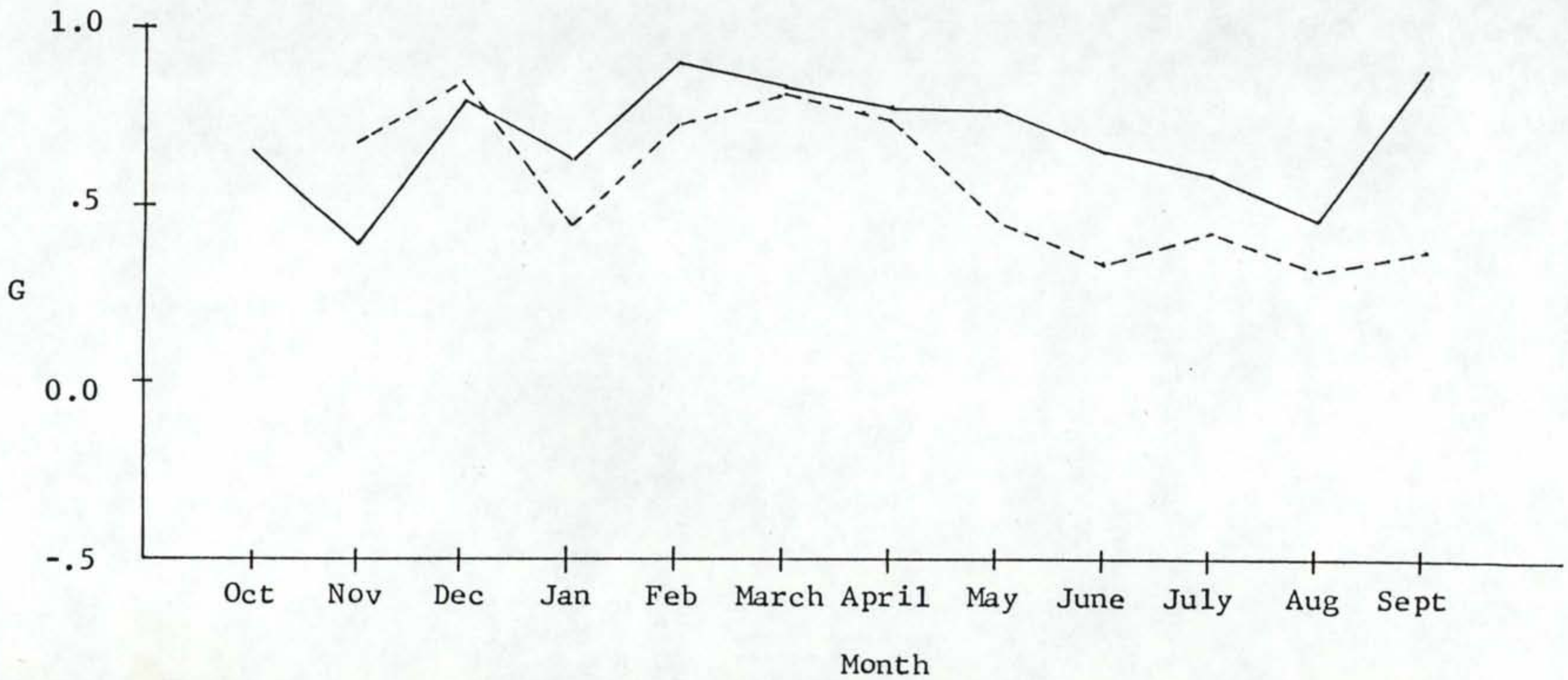


Figure 6.2 Comparison of Lane's Monthly Disaggregation Parameter "G" from the Extended Records at Stations 12413000 and 13186000

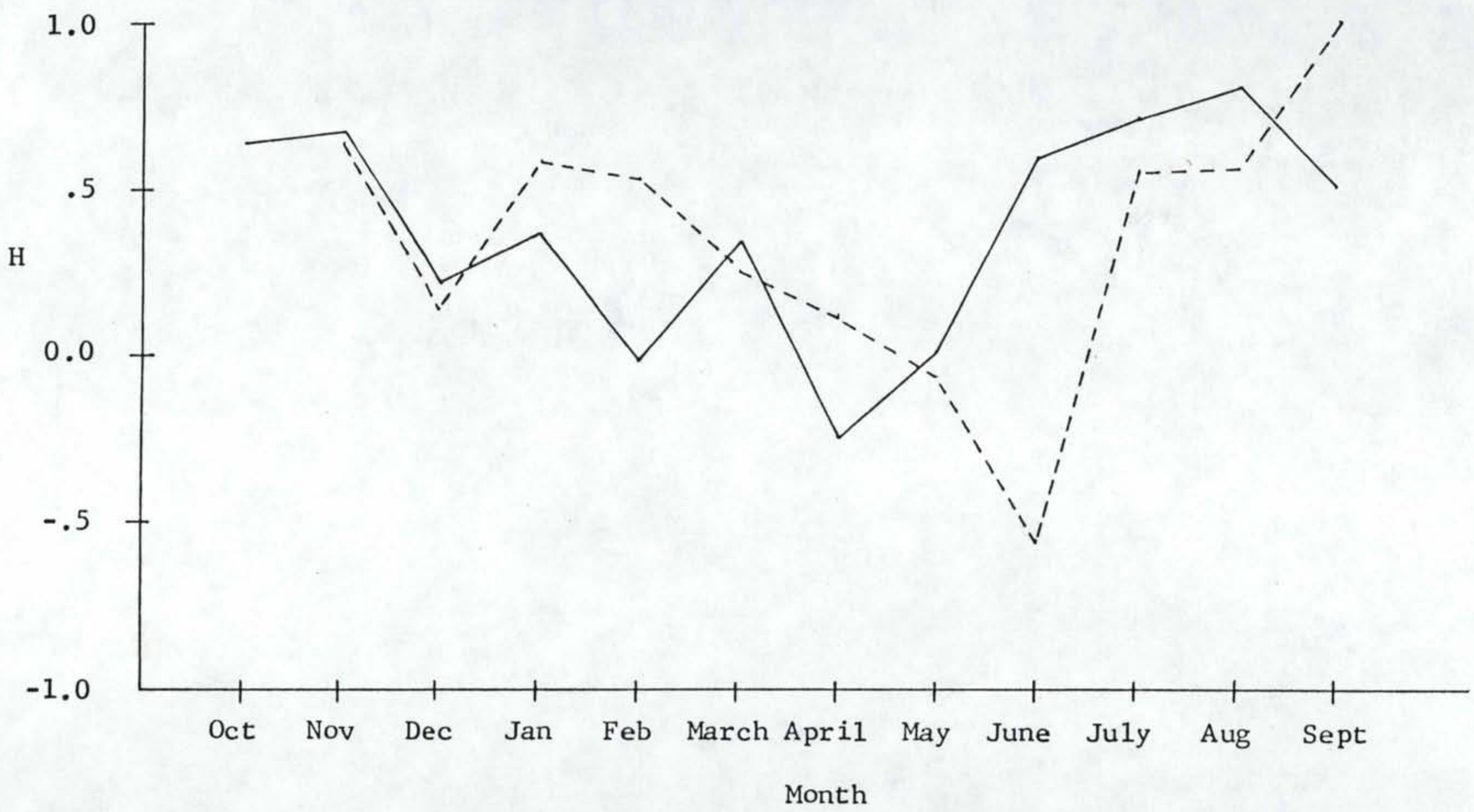


Figure 6.3 Comparison of Lane's Monthly Disaggregation Parameter "H" from the Extended Records at Stations 12413000 and 13186000

parameters was good. However, there does appear to be some differences between the two sets of parameters suggesting that a further refinement of hydrologic regimes may be necessary. In other words, the parameters seem to follow the same general trends, yet if further records were analyzed inbetween these two stations, a set of parameters averaging the two might be found, allowing further refinement of the region's disaggregation parameters. Also, disaggregation models seem to be very robust, as can be seen by the fact that the nonnormality of the October series at stations 12413000, the change in transformations for the month of May (untransformed) at station 12413000, the change in models for the month of October (AR(1)) at station 13186000, and the fact that the transformed series used in modeling did not have means of exactly zero and standard deviations of one, did not adversely affect the resulting synthetic series. Thus, if an appropriate estimate is obtained for each month's disaggregation model parameters, the resulting synthetic records would probably be reasonable.

From the performance of Lane's model described in this chapter, it seems that his model adequately preserves the important statistics of the annual and monthly time series. Thus the reduction in the number of parameters over the Valencia-Schaake and the Mejia-Rouselle models did not have severe adverse effects. In fact, several parameters of

Lane's model could probably be set to zero for several of the months, as they are very close to zero. In conclusion, not all of the parameters of the Valencia-Scaake and Mejia-Rouselle models are needed, since many of these parameters can be set to zero without severly affecting the performance of the disaggregation model.

CHAPTER 7

DROUGHT ANALYSES

This chapter presents the drought analyses which were performed on the historical and synthetic records at stations 12413000 and 13186000. The topics discussed include: the definition of droughts in a streamflow record; the need for accurate drought probability distributions; the generation of synthetic records and their use in determining the probabilities of maximum negative run-lengths and run-sums; effects of data extension on maximum run characteristics; the probabilities of historical droughts; the distributions of annual and monthly flows during drought years; and the use of the analysis results to suggest approaches for the design of storage reservoirs.

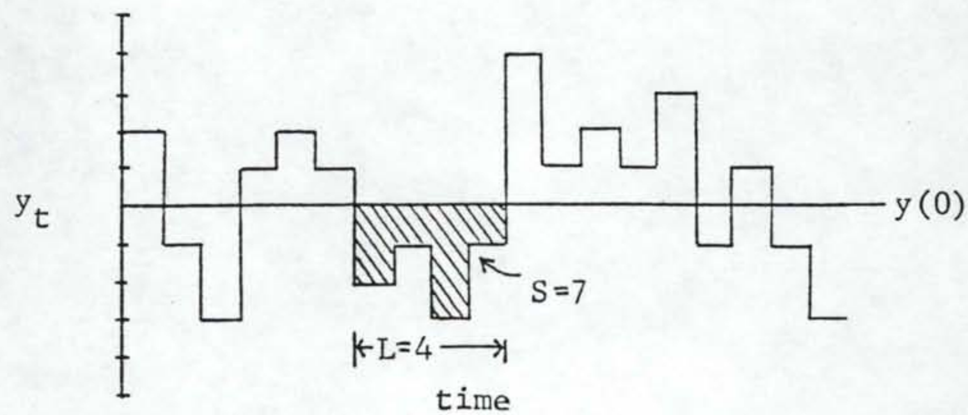
7.1 Definition of Runs

The statistics of a stationary time series are not a function of time. Figure 7.1 illustrates a discrete stationary time series $y_1, \ldots y_t$ which is divided into positive and negative deviations relative to a set value y(0) known as the truncation level. A run is defined as an uninterrupted sequence of either positive or negative deviations from the truncation level. Negative runs can objectively define the droughts of a stationary time series (42).

Two ways of describing negative runs were used in this study: length and sum. Run-length (L) is the duration of a run while run-sum (S) is the sum of the deviates within a run. Figure 7.1 illustrates a drought with a run-length of 4 and a corresponding run-sum of 7. When dealing with water resource problems, the truncation level, run-sum, and run-length are analogous to the demand upon a system, total water deficit, and duration of the drought, respectively.

Figure 7.1

Illustration of Runs



$y(0)$ = truncation level
y_t = stationary streamflow series
t = time units
L = run-length (time units)
S = run-sum (volume)

7.2 Need for Accurate Drought Probability Distributions

In the past, standard practice for designing reservoirs often relied on "critical" historical droughts. This method assumes that the most severe drought to be observed during any historical record would have an approximate return period equivalent to the record length. However, due to sampling fluctuation, the likelihood of this assumption being true is usually very small. A more accurate description of "critical" drought probability distributions is needed in order to better assess the probability of a particular "critical" dought occurrring during the lifetime of a project.

The maximum run-length, L(max) is the longest negative run found in a series of length n. Likewise, the maxiumum run-sum, S(max) is the largest deficit produced from a negative run present in a series of length n. The maximum run-sum of the historical record is usually the "critical" drought used in standard engineering practice. As a result, the expected maximum run-sum corresponding to a particular series length becomes a parameter of major concern. In this study, selected probability distributions of both the maximum run-length and maximum run-sum were investigated for the unextended and extended records at stations 12413000 and 13186000.

7.3 Monte Carlo Drought Analysis

In order to develop the "experimental" probability distributions of the maximum run-length and maximum run-sum, a large number of synthetic streamflow sequences were generated (Monte Carlo method) and then probabilities of the maximum run characteristics were assigned based on relative frequencies. "Experimental" refers to the probability distributions as defined by Monte Carlo methods using stochastic models. The synthetic streamflow sequences were generated using the annual flow models presented in section 5.6 and the disaggregation models formulated in chapter 6 to obtain monthly flow sequences. The stochastic models for both the unextended and extended records were used in order to observe the effects of data extension upon the drought characteristics.

Runs as an objective defintion of droughts can best be applied to stationary time series (10). The annual series at stations 12413000 and 13186000 were stationary while the monthly series were not. Therefore, the maximum run-lengths and maximum run-sums were identified in the synthetic annual records, and then the corresponding monthly sequences (from the disaggregation model) were investigated. The main advantage of this method is that it allows the use of the application of the theory of runs to a stationary annual series.

The distributions of run-lengths and run-sums are affected by the sample size and the selected truncation level. Consequently, the distributions of maximum run-length and maximum run-sum are also influenced by these variables, and their values must be specified if the distributions are to be compared.

The truncation level is usually expressed as a function of the quantile, $q(0)$, with $q(0) = P(y_t \leq y(0))$. As the truncation level changes, so does the frequency and severity of the associated run-lengths and run-sums. For instance, if water demand is high, say $y(0)=1.3\bar{y}$, then more values of y_t will result in negative deviations (deficits) than if water demand were only $.3\bar{y}$. Consequently, the run-lengths (durations) and run-sums (deficits) corresponding to the higher demand will be longer and more severe than for the same supply series (y_t) with a lower demand. Similarly, as the length of a series increases so does the probability of the presence of more extreme events, generally resulting in longer run-lengths and larger run-sums than are observed in shorter series.

For the purposes of this study, it was decided to examine the probability distributions of the maximum run-length and maximum run-sum for sample sizes of 25, 50, and 100 years using two different truncation levels which corresponded to $q(0)=.50$ and $q(0)=.35$. Therefore, synthetic records using the stochastic models for the unextended and extended records at stations 12413000 and 13186000 were needed.

MODELED RECORD LENGTH: The specification of a modeling length for the synthetic streamflow records was the first decision which had to be made. According to the central limit theorem, the sample mean run is normally distributed with a variance of

$$\sigma_R^2 / n \tag{7.1}$$

where:
σ_R^2 = variance of sample runs
n = number of samples

It was arbitrarily desired that the probability be 95% or greater that the computed mean run from the synthetic records be within $\pm .1\sigma_R$ of the population mean run. Using this criterion, the desired number of runs for any sample length was determined as follows:

$$P(\mu_R - .1\sigma_R) \le \bar{R} \le (\mu_R + .1\sigma_R) \ge .95 \tag{7.2}$$

Standardizing:

$$P(-\sqrt{n}/10 \le Z \le \sqrt{n}/10) \ge .95 \quad \text{or} \tag{7.3}$$

$$P(Z \le \sqrt{n}/10) - P(Z \le -\sqrt{n}/10) = .95 \tag{7.4}$$

or stated in terms of the two equal tail areas:

$$P(Z \le -\sqrt{n}/10) = .025 \text{ and } P(Z \le \sqrt{n}/10) = .025 \tag{7.5}$$

for Z corresponding to an area of .025 under a cumulative standardized normal curve:

$$-1.96 = -\sqrt{n}/10 \quad \text{or} \quad 1.96 = \sqrt{n}/10 \tag{7.6}$$

Therefore, by solving for n, the required number of runs is 384 or about 400.

Based on the above calculations, with an upper sample length of 100 years, 400 100-year synthetic streamflow sequences were generated, resulting in a total of 40,000 years of synthetic annual/monthly values. These sequences were then divided to obtain 50- and 25-year sequences. Table 7.1 summarizes the number of 100-, 50-, and 25-year sequences generated and analyzed to obtain the "experimental" drought probability distributions.

Table 7.1

Number of Generated Series

Series Length	100	50	25
Number of Series	400	800	1600

TRUNCATION LEVEL: Next, the two truncation levels for each model were determined. In Chapter 4, the unextended and extended annual series at stations 12413000 and 13186000 were shown to be normally distributed. Therefore, the truncation levels, y(0), were determined using the theoretical normal distribution. The quantile q(0) is equivalent to the cumulative area under a standard normal curve. Therefore, the standardized normal deviate, Z, corresponding to q(0) was used to determine the truncation levels as shown below:

$$Z = \frac{y(0) - \bar{y}}{s_y} \qquad (7.7)$$

The resulting truncation levels are presented in Table 7.2.

Table 7.2

Truncation levels

Station	Record	Mean (cfs)	Std Dev (cfs)	q(0)	Z	y(0) (cfs)
12413q00	Unextended	706854	198977	.50	.0000	706854
12413000	Unextended	706854	198977	.35	-.3854	630168
12413000	Extended	696295	205315	.50	.0000	696295
12413000	Extended	696295	205315	.35	-.3854	617167
13186000	Unextended	294519	94040	.50	.0000	294519
13186000	Unextended	294519	94040	.35	-.3854	258276
13186000	Extended	271726	95757	.50	.0000	271726
13186000	Extended	271726	95757	.35	-.3854	234821

COMPUTER MODEL: A computer program was developed to generate the synthetic sequences, identify the maximum run-lengths and maximum run-sums, determine their cumulative density functions (CDF's) and compute the statistics and flow distributions of the corresponding annual and monthly series. Following is a more detailed explanation of the steps the program actually went through to arrive at these results:

1) Four hundred 100-year annual/monthly synthetic streamflow sequences were generated with the disaggregation model developed from the unextended record at station 12413000, using the procedure described in section 6.11.

2) From each 100-year sequence the maximum run-length based on the truncation level corresponding to q(0)=.50 was identified.

2a. The CDF of maximum run-lengths from the 100-year series was determined:

$$CDF_i = \text{number of } L(max) \text{ with size} \leq i \ / \ 400 \qquad (7.8)$$

for i = 1 to longest run-length observed in all of the 100-year sequences

400 = total number of maximum run lengths examined (sample size)

2b. The statistics (equations 2.1 to 2.14) of the annual and monthly streamflow values which comprised the maximum run-length series were calculated.

2c. Using the same annual and monthly sequences as in step 2b, the number of streamflow values for each year and month which fell between predefined ranges were counted, to arrive at a histogram (flow distribution) for these annual and monthly streamflow values.

3) From each 100-year sequence the maximum run-sum based on the truncation level corresponding to q(0)=.50 was identified.

3a. The CDF of maximum run-sums from the 100-year series was determined:

$$CDF_i = \text{number of } S(max) \text{ with values} \geq i(s/2) \ / \ 400 \qquad (7.9)$$

for i = 0 to 60

400 = total number of maximum run-sums examined (sample size)

s = standard deviation of annual series

3b. The statistics (equations 2.1 to 2.14) of the annual and monthly streamflow values which comprised the maximum run-sum series were calculated.

3c. Using the same annual and monthly sequences as in step 3b, the number of streamflow values for each year and month which fell between predefined ranges were counted, to arrive at a histogram (flow distribution) for these annual and monthly streamflow values.

4) Each of the 400 100-year sequences were divided in half such that 800 50-year sequences resulted.

4a. Steps 2 and 3 were repeated, replacing "100-year" with "50-year" and "400" with "800".

5) Each of the 400 100-year sequences were divided in fourths such that 1600 25-year sequences resulted.

5a. Steps 2 and 3 were repeated, replacing "100-year" with "20-year" and "400" with "1600".

6) Steps 2 through 5 were repeated replacing "q(0) = .50" with "q(0) = .35" in steps 2 and 3.

7) Steps 1 through 6 were repeated replacing "unextended record at station 12413000" with "extended record at station 12413000" in step 1.

8) Steps 1 through 6 were repeated replacing "unextended record at station 12413000" with "unextended record at station 13186000" in step 1.

9) Steps 1 through 6 were repeated replacing "unextended record at station 12413000" with "extended record at station 13186000" in step 1.

Summarizing, the program produced the following output for the unextended and extended records at stations 12413000 and 13186000:

1) 40,000 years of monthly streamflow values

2) Cumulative density functions of the annual L(max) and S(max) for a 100-, 50-, and 25-year sequence (sample size) with y(0) corresponding to q(0)=.50 and q(0)=.35

3) Annual and monthly flow distributions and statistics corresponding to each annual series which composed the cumulative density functions listed above.

Since the actual truncation level y(0) changes for each model, hereafter this variable will frequently be referenced by its corresponding value of q(0).

7.4 Probabilties of Maximum Negative Run-Lengths

THEORETICAL DISTRIBUTION: The exact probability distribution of run-lengths for an independent normal series can be approximated by (8):

$$F_n(L+1) = \frac{1 - q(0)\Psi}{(L + 2 - (L + 1)\Psi\, p(0))} \; \frac{1}{\Psi^{n+1}} \qquad (7.10)$$

where:

$\Psi = 1 + p(0)q(0)^L + (L + 1)(p(0)q(0)^L)^2 + (L+1)^2(p(0)q(0)^L)^3 + \ldots$

$q(0) = P(y_t \leq y(0))$, truncation level

$p(0) = P(y_t > y(0))$; $p(0) = 1 - q(0)$

L = run-length size

$F_n(L+1)$ = probabilty of a run of size L+1 occurring for the first time in a series of length n_i, $i = 1, 2, 3, \ldots$

The CDF's of the maximum run-lengths can be determined by solving equation 7.10 with a constant truncation level q(0) and constant sample size n, for various values of run-length size (1, 2, . . . n/2). Note that for the independent normal process, the distribution is only a function of the truncation level and the series length.

EXPERIMENTAL CDF'S OF THE MAXIMUM RUN-LENGTHS: All of the annual series examined in this study were independent normal series, and according to equation 7.10 should all have identical CDF's of the maximum run-lengths for equal truncation levels and sample sizes. The experimental CDF's verified this conclusion. Tables E.1 through E.6 in

Appendix E list the exact and experimental CDF's of the maximum run-lengths (q(0) = .50 and .35; n = 100, 50 and 25 years) based on the stochastic models developed from the unextended and extended records at station 12413000 and 13186000.

Millan and Yevjevich (27) found that the lognormal probability distribution could be used as an approximation of the CDF of the maximum run-lengths. Figure 7.2 presents a log-probability plot of the experimental CDF's of the maximum run-length for each truncation level (q(0) = .50 and .35) and series length (100-, 50-, and 25-year) examined. For comparision, the theoretical CDF of the maximum run-length for each set of parameters was also plotted. A "best" straight line was drawn through the expermental points resulting in six CDF's of the maximum run-length.

RETURN PERIODS OF MAXIUMUM RUN-LENGTHS: The size of the maximum run-length corresponding to a probability of being exceeded or not exceeded 50% of the time is, by definition, the median of the probability distribution of the maximum run-length. In other words, the median size of the maximum run-length is the maximum run-length that would be exceeded or not exceeded 50% of the time if many samples of size n were generated using the same stochastic model. The median value, L_m from each CDF of the maximum run-length was defined as the representative maximum drought length for

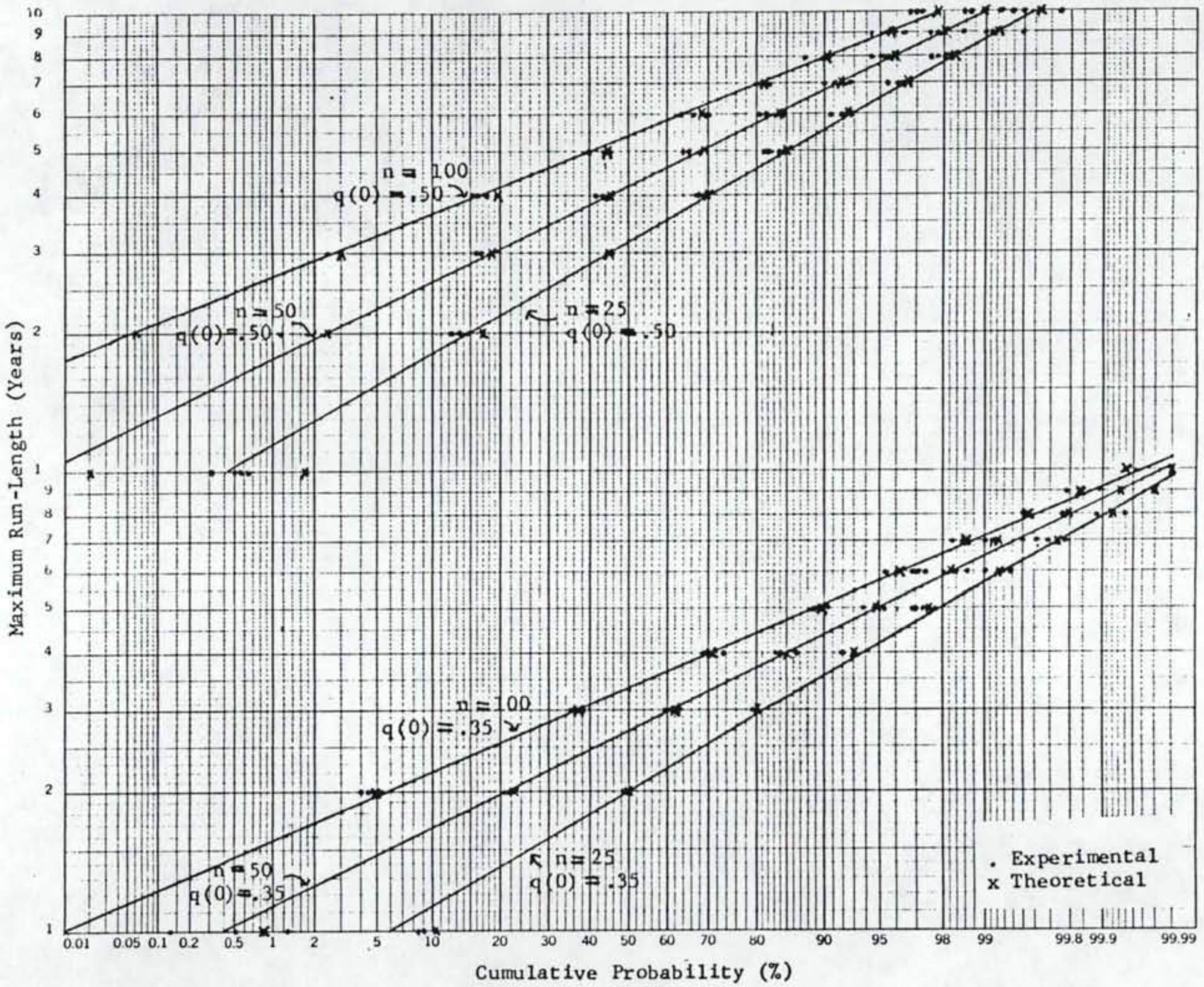


Figure 7.2 Experimental Cumulative Density Function of the Maximum Run-Length

a given sample size of n and truncation level. Moreover, for a normal series this median drought length equals the mean maximum run-length. The value of n associated with L_m becomes its annual return period, T.

The representative maximum drought lengths L_m, can be used to determine if the maximum run-length observed in an historical record is "representative" of that record. If the historical maximum run-length is close to the corresponding L_m value for the same sample size (record length) and truncation level, then the historical record accurately predicts the return period of the maximum run-length. If, however, the historical value of the maximum run-length corresponds to a very high or a very low probability of occurrence as determined from the CDF of the maximum run-lengths, then the historical maximum run-length would be considered "unrepresentative" of the historical record, and its return period, if based on the historical record, would be misleading.

Using the best straight line fits of the CDF's of the maximum run-lengths in Figure 7.2, the representative maximum drought lengths L_m, were determined for the six cases shown. These values of L_m are listed in Table 7.3.

Table 7.3

Experimental Representative Maximum Drought Lengths

T years	q(0)	L_m years	T years	q(0)	L_m years
25	.35	2.0	25	.50	3.2
50	.35	2.7	50	.50	4.2
100	.35	3.3	100	.50	5.4

The representative run-lengths with equal truncation levels were then plotted against the sample size on semi-log paper in Figure 7.3. A best fit curve was drawn through the points to estimate the relationship that exists between the maximum run-length and annual return period for an independent normal process with a truncation level of .50 and .35.

7.5 Probabilities of Maximum Negative Run-Sums

THEORETICAL DISTRIBUTIONS: The exact distriubtion of the run-sums for the independent normal process is more complex than for the run-lengths, and as a result will not be presented here. Instead, the results developed from Monte Carlo experiments by Millan and Yevjevich (27) will be reviewed and compared to the experimental CDF's of the maximum run-sums obtained in this study.

The probabiltiy distribution of run-sums was dependent upon the sample size, n, truncation level, q(0), serial

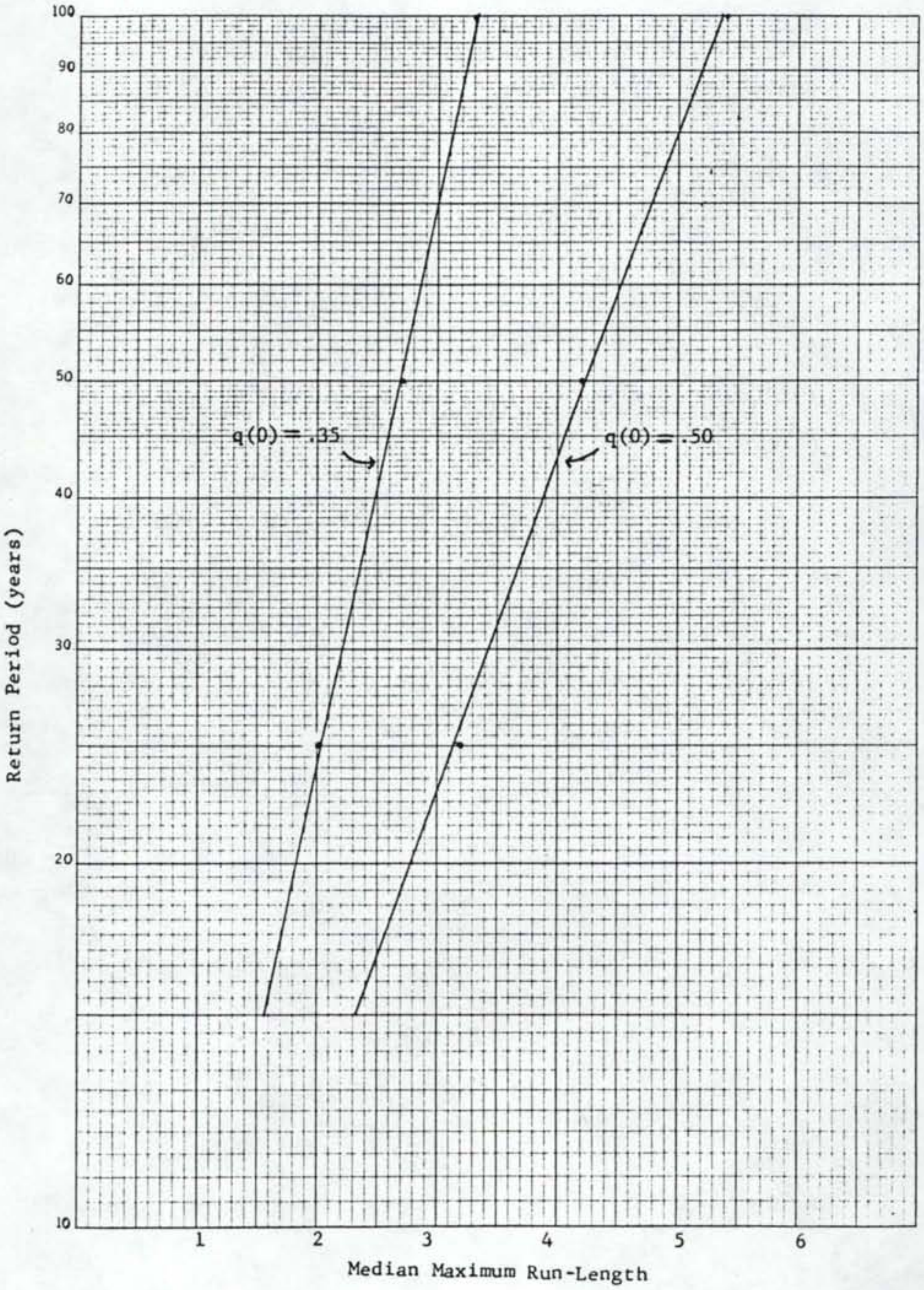


Figure 7.3 Relationship between the Median Maximum Run-Length and Return Period

correlation structure, ρ, and skewness, γ, of the process. In addition, the magnitude of the run-sums were directly proportional to the standard deviation of the process (34). Therefore, if the probability distribution of the maximum run-sums for a partiuclar set of variables (n, q(0), ρ, γ) is found for the standardized series (σ = 1), then the maximum run-sum distribution for any nonstandardized series ($\sigma \neq 1$) with the same characteristics (n, q(0), ρ, γ) could be found by multiplying the standardized distribution values by the standard deviation of the nonstandardized series.

As previously noted all of the annual series studied were independent ($\rho_K = 0$) and normally distributed ($\gamma = 0$). Hence, the standardized CDF's of the maximum run-sums for these series should be identical, given constant values of q(0) and n.

EXPERIMENTAL CDF'S OF THE MAXIMUM RUN-SUMS: Tables E.7 through E.12 in Appendix E list the experimental nonstandardized CDF's of the maximum run-sums (q(0) = .50 and .35; n = 100, 50 and 25 years) based on the stochastic models developed from the unextended and extended records at stations 12413000 and 13186000. These nonstandardized CDF's of the maximum run-sums were plotted on log-probabiltiy paper and the results are presented as Figures 7.4 through 7.7. A best fit straight line was drawn through the experimenatal points resulting in six CDF's of the maximum

Figure 7.4

Experimental Cumulative Density Function of the Maximum Run-Sum from the Unextended Record at Station 12413000

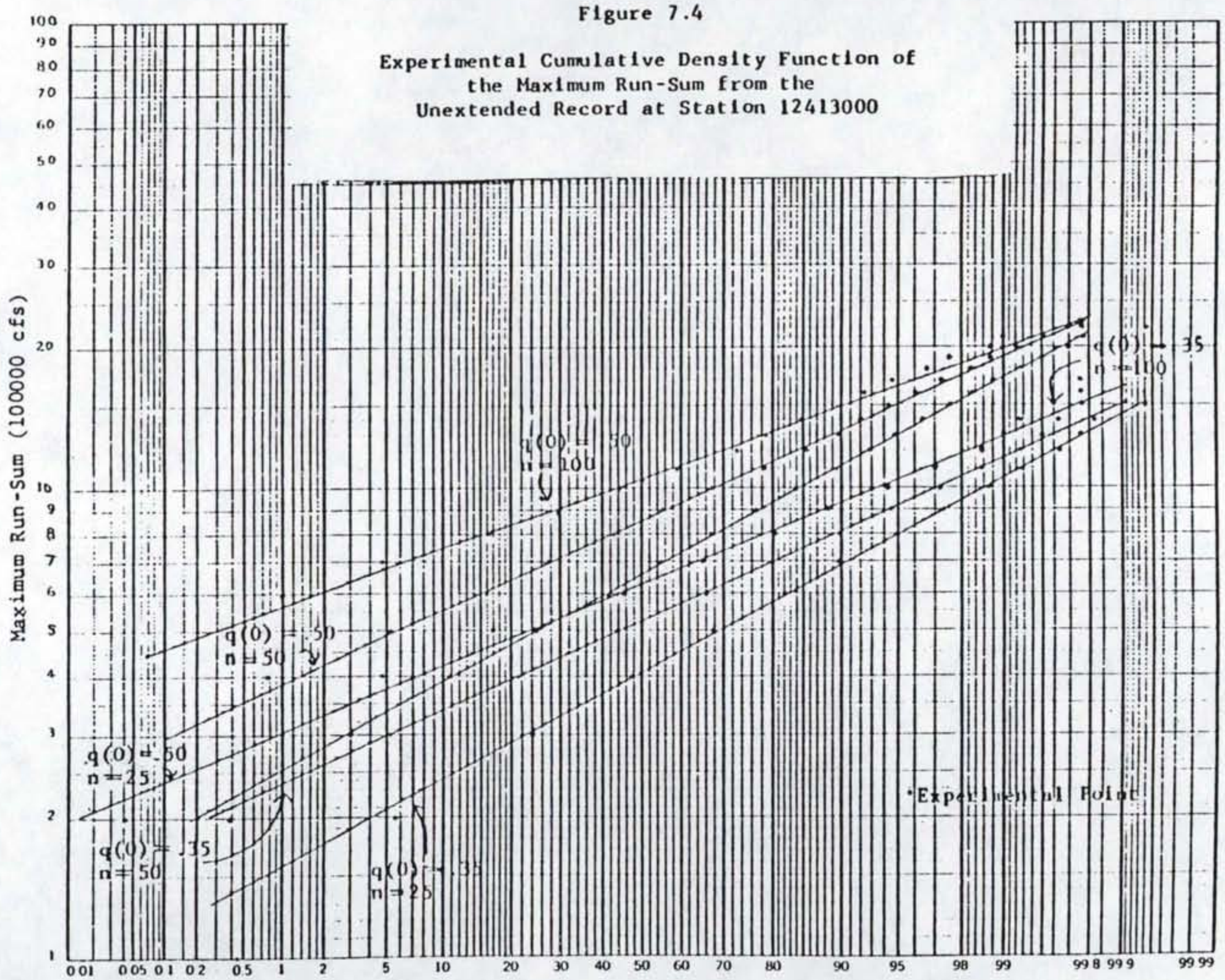


Figure 7.5

Experimental Cumulative Density Function of
the Maximum Run-Sum from the
Extended Record at Station 12413000

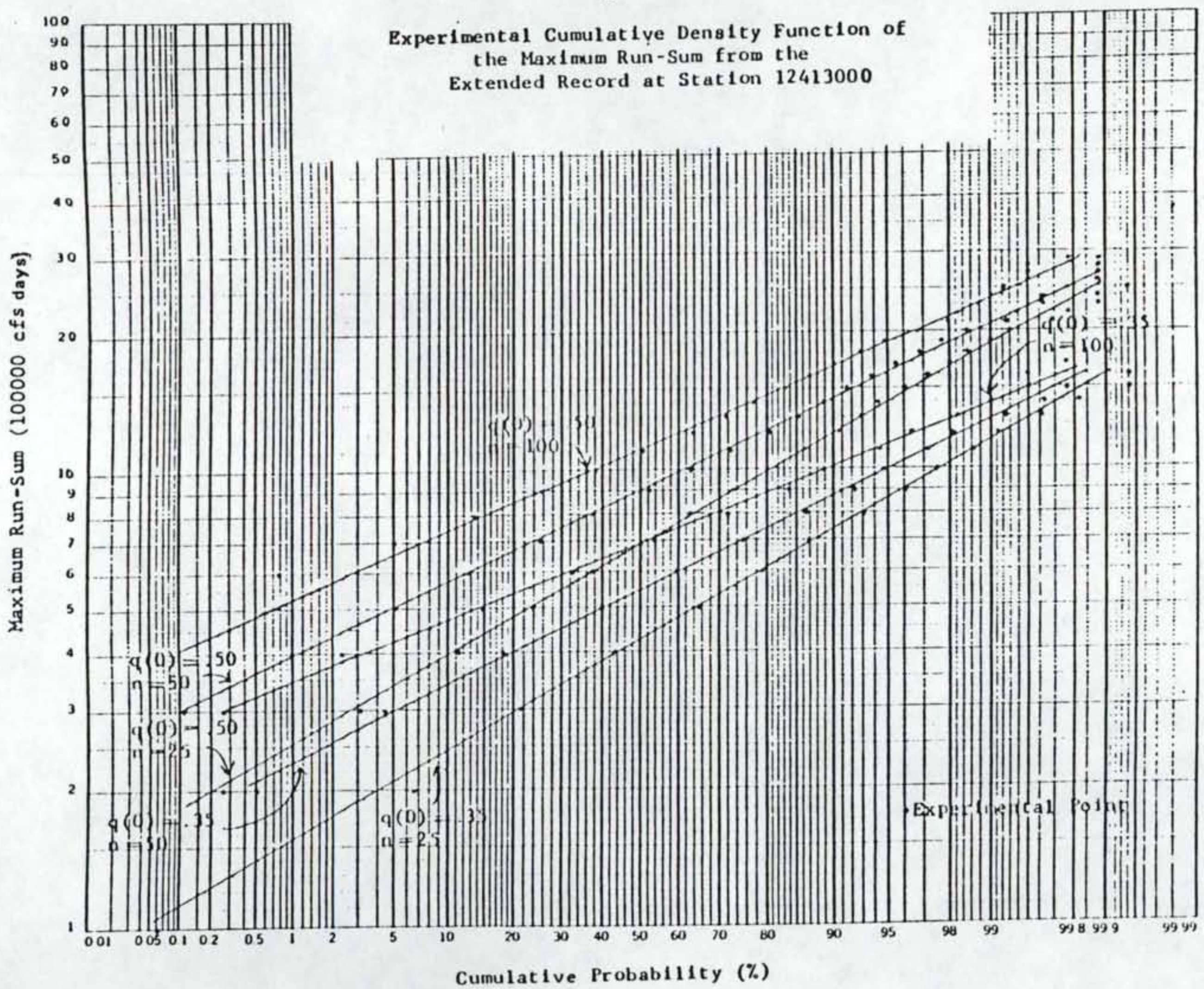


Figure 7.6

Experimental Cumulative Density Function of the Maximum Run-Sum from the Unextended Record at Station 13186000

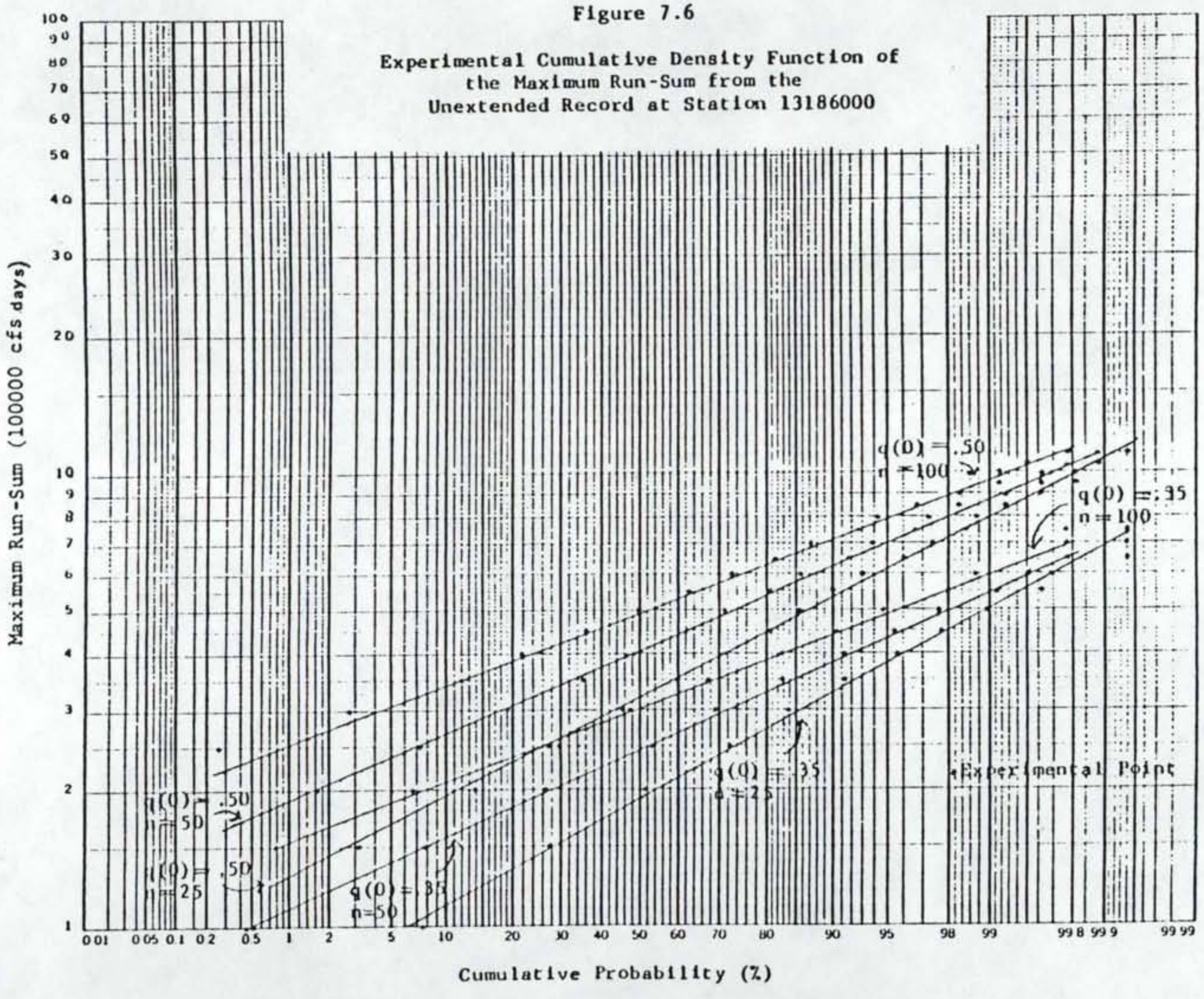
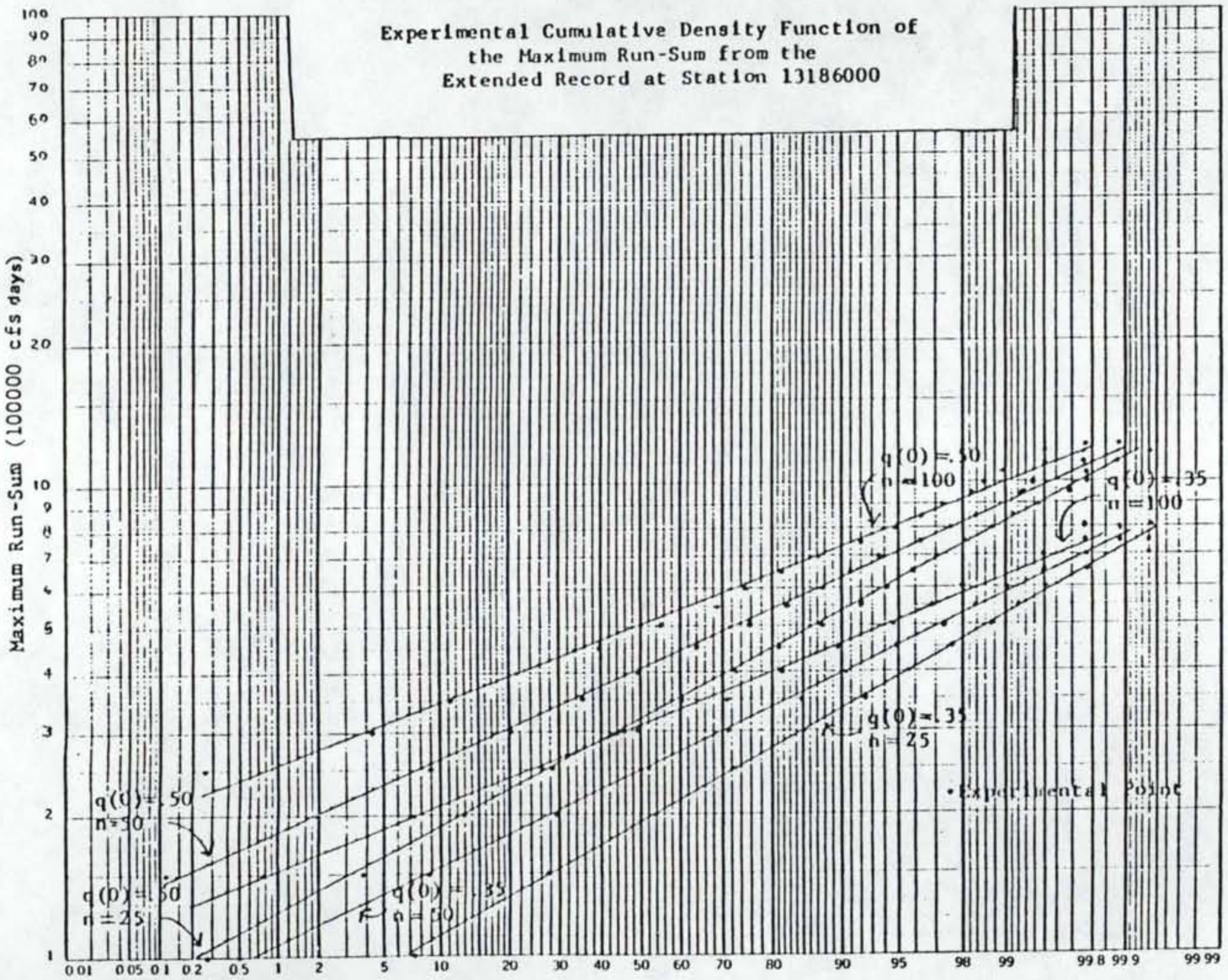


Figure 7.7

Experimental Cumulative Density Function of
the Maximum Run-Sum from the
Extended Record at Station 13186000



run-sums corresponding to each annual record at stations 12413000 and 13186000.

RETURN PERIODS OF THE MAXIMUM RUN-SUMS: The median maximum run-sum S_m, was defined as the representative run-sum or drought deficit, from each CDF of the maximum run-sums. As was the case for L_m, the representative maximum run-sum is the maximum run-sum that would be exceeded or not exceeded 50% of the time if many samples of size n were generated using the same stochastic process. Furthermore, the corresponding sample size n, becomes the annual return period, T, associated with S_m.

Again, just as L_m could be used to determine if an historical maximum run-length was representative of its historical record, S_m can be used to deterimine if an historical maximum run-sum is representative of its historical record. If the historical standardized maximum run-sum is close to the corresponding standardized representative maximum run-sum value S_m for the same sample size and truncation level, then the historical maximum run-sum value is representatvie of that record and its return period based upon the historical record would also be considered representative.

Using the best straight line fits of the CDF's of the maximum run-sums in Figures 7.4 through 7.7, the representative maximum run-sum S was determined from each figure for the six cases illustrated. These values of S are listed in Table 7.4.

Table 7.4

Experimental Representative Maximum Run-Sums (cfs days)

Station	Record	q(0)	n yrs	S ($x10^3$)	Std Dev ($sx10^3$)	D = S/s
12413000	Extended	.35	25	428	205.30	2.08
12413000	Extended	.35	50	548	205.30	2.67
12413000	Extended	.35	100	690	205.30	3.36
12413000	Extended	.50	25	690	205.30	3.36
12413000	Extended	.50	50	900	205.30	4.38
12413000	Extended	.50	100	1100	205.30	5.36
12413000	Unextended	.35	25	405	198.98	2.04
12413000	Unextended	.35	50	522	198.98	2.62
12413000	Unextended	.35	100	630	198.98	3.17
12413000	Unextended	.50	25	660	198.98	3.32
12413000	Unextended	.50	50	850	198.98	4.27
12413000	Unextended	.50	100	1030	198.98	5.18
13186000	Extended	.35	25	192	95.76	2.00
13186000	Extended	.35	50	245	95.76	2.56
13186000	Extended	.35	100	300	95.76	3.13
13186000	Extended	.50	25	312	95.76	3.26
13186000	Extended	.50	50	403	95.76	4.21
13186000	Extended	.50	100	488	95.76	5.10
13186000	Unextended	.35	25	195	94.04	2.04
13186000	Unextended	.35	50	248	94.04	2.62
13186000	Unextended	.35	100	306	94.04	3.17
13186000	Unextended	.50	25	320	94.04	3.32
13186000	Unextended	.50	50	405	94.04	4.27
13186000	Unextended	.50	100	502	94.04	5.18

The representative maximum run-sum S_m values in Table 7.4 were standardized by dividing by the standard deviation of the corresponding annual series. These standardized values (D_m), with equal truncation levels were then plotted against sample size (return period) on semi-log paper and the results are presented as Figure 7.8. Also plotted in Figure 7.8 are the standardized values of D_m (Table 7.5) determined from the results of Millan and Yevjevich's experimental CDF's of the maximum run-sums for the case where $\rho = 0$ and $\gamma = 0$ (27). (Yevjevich's results were based on 95,000 years of synthetic record.)

Table 7.5

Standardized Representative Run-Sums after Yevjevich (Experimental)

T (yrs)	q(0)	D_m	T (yrs)	q(0)	D_m
25	.35	2.1	25	.50	3.4
50	.35	2.7	50	.50	4.3
100	.35	3.2	100	.50	5.2

Best-fit straight lines were drawn through the experimental points in Figure 7.8 for truncation levels. Although there was some scatter, only one line was drawn for each truncation level, because no justification could be found for assuming any differences between the unextended and extended records or between locations (all of the annual series were independent and normal). The experimental

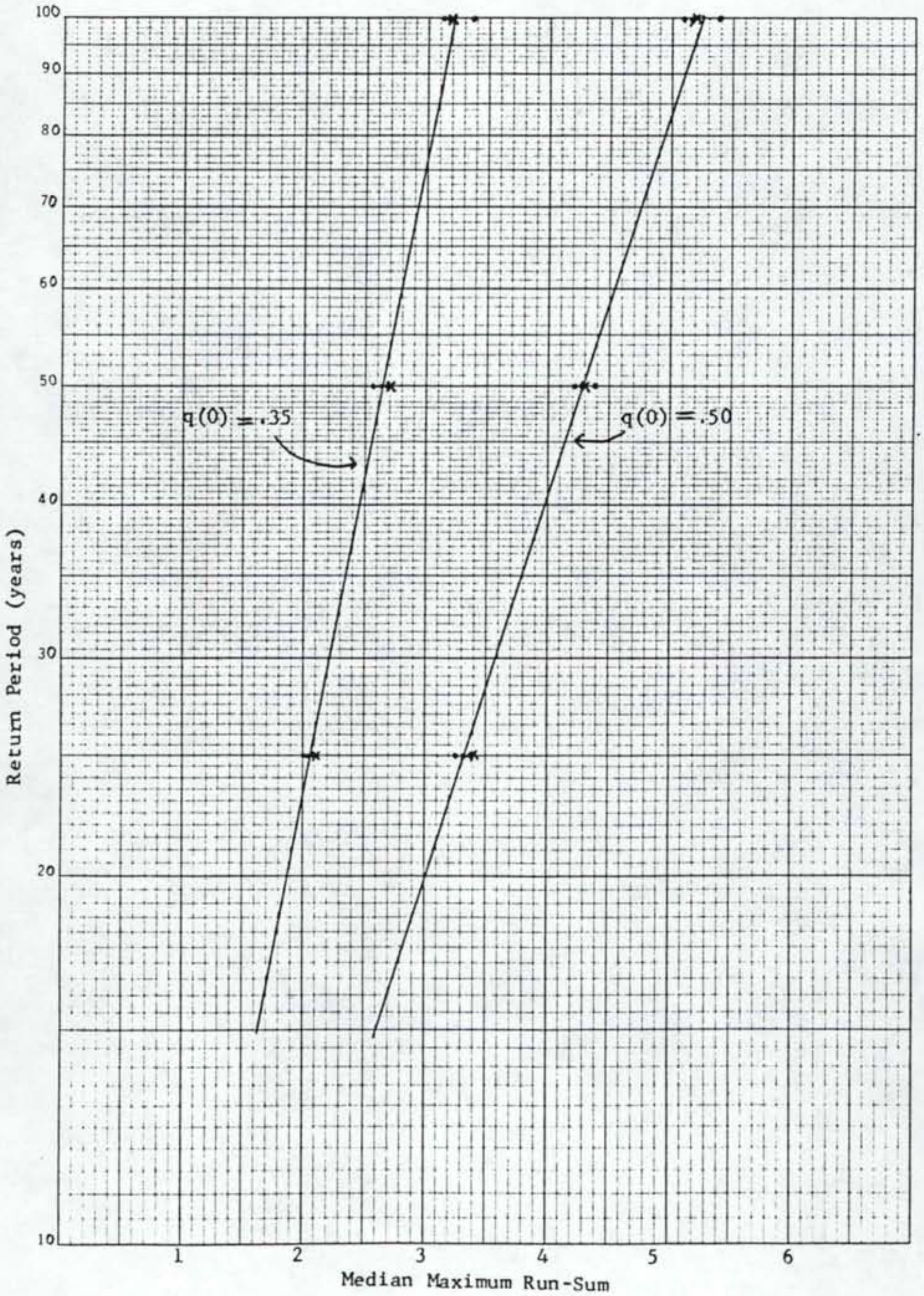


Figure 7.8 Relationship between the Median Maximum Run-Sum and Return Period

points corresponding to Yevjevich's results were weighted more heavily since they were based on many more samples.

7.6 Effects of Data Extension

While a preliminary assessment of the experimental CDF's of the maximum run-lengths and maximum run-sums indicate that data extension does not affect the relationships between the representative maximum run characteristics and their associated return periods (Figures 7.3 and 7.8), this result may be misleading for the following reasons:

1) Data extension would not affect these relationships except when the unextended and extended record estimates of ρ and γ are unequal. While in this study they were all equal to zero, this would not always be true for other records.

2) The actual median maximum run-sum S_m, equals the corresponding standardized maximum run-sum D_m, times the standard deviation σ, of the nonstandardized series ($S_m = D_m\sigma$). While D_m may be unaffected by data extension S is affected as σ changes. Table 7.6 summarizes the changes in σ resulting from data extension.

3) The truncation level y(0) greatly influences the CDF's of the maximum run-lengths and the maximum run-sums, which in turn influence the assignment of the return periods (34). Usually, as was the case in this study, the statistics of the unextended and extended records will vary somewhat, and these statistics are used to estimate the trunction levels based on the records' assumed probability distributions. Table 7.6 also summarizes the changes in the truncation level resulting from data extension.

Table 7.6

Changes in the Standarad Deviation and Truncation Level Resulting from Data Extension (cfs days)

Station	q(0)	Standard Deviation Unextend	Standard Deviation Extend	Truncation Level Unextend	Truncation Level Extend
12413000	.35	198977	205315	630167	617166
12413000	.50	198977	205315	706854	696295
13186000	.35	94040	95757	258276	234821
13186000	.50	94040	95757	294519	271726

As previously discussed in section 3.5, the population statistics estimated from the extended records at stations 12413000 and 13186000 are more reliable than those estimated from the corresponding unextended records, due to the larger sample sizes which produce smaller confidence intervals around the estimated statistics. Therefore, any parameters estimated from the extended sample statistics would be more reliable than those estimated from the unextended sample statistics. As a result, the maximum run characteristics developed from the extended records' estimates of ρ, γ, y(0) and σ would be more reliable than those developed from the unextended records. Consequently, subsequent analyses and presentations will concentrate on the results of the extended models, although historical droughts will be reviewed by examining both the unextended and extended records.

7.7 Probabilities of Historical Droughts

In this section, the maximum historical droughts from the unextended and extended records at stations 12413000 and 13186000 are assessed in terms of their representativeness of the sample period and long-term stochastic process.

13186000 - UNEXTENDED: Table 7.7 presents the maximum run-length and the maximum run-sum from the unextended record at station 13186000.

Table 7.7

Maximum Run Characteristics from the Unextended Record at Station 13186000

n years	q(0)	L(max) years	S(max) cfsd	Std Dev cfsd	D(max)
38	.50	6	399069	94040	4.24
38	.35	3	178635	94040	1.90

The return period of each maximum historical drought listed in Table 7.7 was assigned using the relationships developed in Figures 7.3 and 7.8, by letting L(max) = L_m and D(max) = D_m. The resulting return periods, T, listed in Table 7.8, represent the expected number of years that would elapse between the specified droughts based on the long-term stochastic model of the historical series.

Table 7.8

Return Period of Maximum Droughts from the Unextended Record at Station 13186000.

q(0)	L_m years	T years	q(0)	D_m	T years
.50	6	>100	.50	4.24	48
.35	3	68	.35	1.90	<25

As previously mentioned in section 7.2, standard engineering practice has often assumed that an historical drought has a return period similar to the record (sample) length. However, the analysis of these historical droughts at station 13186000 indicate how misleading this concept may be. For truncation levels of both .50 and .35, the maximum historical run-lengths are considerably longer than would be expected from the stochastic process as modeled.

The maximum historical run-sum corresponding to a truncation level of .50 appears to be representative of the record length (n = 38 years as compared to 48 years from Figure 7.8). However, at a truncation level of .35, the historical maximum run-sum is less severe than would reasonably be expected in a record of this length. Hence, a design based on this critical deficit may lead to a significant underdesign.

13186000 - EXTENDED: If the extended record at station 13186000 is analyzed as a new "quasi-historical" record of length n = 72 years, several significant changes occur in the definition of the critical drought periods:

1) Due to the changing (lowering) of the truncation levels, y(0), the historical maximum run-length of six years with q(0) = .50, is no longer the maximum run-length in the extended record. One of the annual flows comprising the six year drought is above the truncation level as determined from the extended record.

2) The maximum run-lengths at both truncation levels now occur in the extended portion of the "quasi-historical" record.

Table 7.9 presents the maximum run-length and the maximum run-sum from the extended record at station 13186000.

Table 7.9

Maximum Run Characteristics from the Extended Record at Station 13186000

n years	q(0)	L(max) years	S(max) cfsd	Std Dev cfsd	D(max)
72	.50	4	334110	95757	3.49
72	.35	4	208160	95757	2.17

The return period of each maximum drought listed in Table 7.9 was assigned using the relationships developed in Figures 7.3 and 7.8 by letting L(max) = L_m and D(max) = D_m, as was done with the historical droughts from the unextended record. The resulting return periods are listed in Table 7.10.

Table 7.10

Return Period of Maximum Droughts from the Extended Record at Station 13186000.

q(0)	L_m years	T years	q(0)	D_m	T years
.50	4	44	.50	3.49	27
.35	4	>100	.35	2.17	27

The return periods listed in Table 7.10 indicate that neither the extended record maximum run-lengths nor run-sums are really representative of the stochastic process, and reemphasize the importance of using the experimental results to determine the expected maximum drought length and severity for a given return period, rather than relying solely on the historical record.

12413000 - UNEXTENDED and EXTENDED: The unextended and extended records at station 12413000 were examined by using the same procedures previously described for station 13186000. The maximum droughts were identified and return periods assigned based on the relationships in Figures 7.3 and 7.8, and the results are summarized in Tables 7.11 and 7.12.

Table 7.11

Maximum Run Characteristics from the Unextended and Extended Records at Station 12413000.

Record	n years	q(0)	L(max) years	S(max) cfsd	Std Dev cfsd	D(max)
Unextended	44	.50	3	737265	198977	3.71
Unextended	44	.35	3	507204	198977	2.55
Extended	63	.50	4	933725	205315	4.55
Extended	63	.35	4	695778	205315	3.39

Table 7.12

Return Period of Maximum Droughts from the Unextended and Extended Records at Station 12413000.

Record	q(0)	L_m years	T years	q(0)	D_m	T years
Unextended	.50	3	<25	.50	3.71	32
Unextended	.35	3	68	.35	2.55	43
Extended	.50	4	44	.50	4.55	61
Extended	.35	4	>100	.35	3.39	>100

The historical maximum run-sums from the unextended record appear to be representative of the sample size (n = 44). Data extension yields a new record with the maximum run-sum, corresponding to a truncation level of .50, again representative of its 63 year record length. However, the maximum run-sum with a truncation level of .35 from the extended record is considerably more severe than could reasonably be expected in a 63 year record.

As previously stated, the design of water resource storage projects can best be based on the assignment of return periods using the probabilities determined by modeling. This method avoids the obvious inconsistencies that are apparent in these evaluations of the historical records when critical historical periods are arbitrarily assigned return periods equivalent to the record length. The previous analyses indicate that an assignment of return period based on the historical record length may yield results that are sometimes reasonable but at other times are unreasonably high or low.

Table 7.13 summarizes the 100-, 50-, and 25-year drought characteristics as determined from Figures 7.3 and 7.8 for stations 12413000 and 13186000. The maximum run-sums were calculated using the standard deviations from the extended records.

Table 7.13

Drought Characteristics of Stations
13186000 and 12413000.

T		13186000		12413000	
years	q(0)	L_m(yrs)	S_m(cfsd)	L_m(yrs)	S_m(cfsd)
25	.50	3.2	326000	3.2	698000
50	.50	4.2	412000	4.2	883000
100	.50	5.4	498000	5.4	1070000
25	.35	2.0	203000	2.0	435000
50	.35	2.7	256000	2.7	550000
100	.35	3.3	309000	3.3	663000

7.8 Distributions of Annual Flows During Drought Years

DROUGHT YEAR STATISTICS: The computer program used to analyze the characteristics of droughts from the modeled stochastic processes (section 7.3), developed a histogram and computed the statistics of the annual flow values which comprised the maximum run-length sequences. The same was also done for the annual flows which made up the maximum run-sum sequences. The resulting statistics for the extended records at stations 1241300 and 13186000 are listed in Table 7.14.

Table 7.14

Statistics for Annual Flows from Drought Periods

Station 12413000 - Extended.

		Based on Max Run-Sum			Based on Max Run-Length		
q(0)	n yrs	Mean cfsd	Std Dev cfsd	Coef Skew	Mean cfsd	Std Dev cfsd	Coef Skew
.50	25	493977	138218	-.69	533426	122637	-.95
.50	50	490639	139411	-.69	531997	123748	-.94
.50	100	487328	140802	-.64	530378	123787	-.92
.35	25	421759	128311	-.60	480248	108485	-1.04
.35	50	412886	134252	-.57	479890	108904	-1.05
.35	100	411691	135062	-.61	477306	109936	-1.04

Extended Record Statistics:

$\bar{y}$ = 696295 cfsd At q(0) = .50 y(0) = 696295 cfsd

s_y = 205315 cfsd q(0) = .35 y(0) = 617166 cfsd

g_y = -.187

Station 1318600 - Extended

		Based on Max Run-Sum			Based on Max Run-Length		
q(0)	n yrs	Mean cfsd	Std Dev cfsd	Coef Skew	Mean cfsd	Std Dev cfsd	Coef Skew
.50	25	178780	61886	-.58	197220	55557	-.86
.50	50	176772	62546	-.55	197519	55247	-.86
.50	100	174471	63381	-.55	197286	55360	-.87
.35	25	146713	56531	-.42	171738	48719	-.94
.35	50	143371	58331	-.40	172325	49138	-.95
.35	100	142698	58317	-.40	173362	48856	-.96

Extended Record Statistics:

$\bar{y}$ = 271726 cfsd At q(0) = .50 y(0) = 271726 cfsd

s_y = 95757 cfsd q(0) = .35 y(0) = 234821 cfsd

g_y = .271

After examining the statistics in Table 7.14, the following observations were made:

1) The sample size n did not seem to greatly affect the statistics for a given truncation level.

2) As expected, the truncation level q(0) had a large impact on the mean, since, as the truncation level increases, more "larger" flows are included in the negative run sequences.

3) The standard deviations also increased as the truncation level increased, which again can be explained by the "larger" flows which are considered as droughts as the trucation level increases.

4) Based on run-sums the skew coefficient increased (in a negative sense) as the truncation level increased. The opposite trend was observed for the skew coefficients from the maximum run-length series.

5) Droughts defined by the maximum run-sum are more severe than those defined by the maximum run-length, since the longest sequences of drought years may not contain extremely low flows. Also, the standard deviations of the maximum run-sum droughts are greater than the corresponding standard deviations of the maximum run-length droughts.

6) As expected, the mean, standard deviation and skew coefficients of the drought years are less than the values of the parent distribution because droughts represent a sample from the tail-area of the parent distribution. The skew coefficient becomes negative as a result of excluding all the large flow years (truncated distribution).

7) The skew coefficients from the run-lengths are larger (in a negative sense) than those based on run-sums.

DISTRIBUTION OF ANNUAL FLOWS IN DROUGHT YEARS: The actual distributions of the annual drought flows are probably very complex, as they are bounded by zero on the left and the truncation level on the right. However, one attempt was made to fit the probability density function of the annual drought flows for one set of statistics, i.e. the annual droughts based on the maximum run-sums at station 13186000 for n = 50 and q(0) = .35. Figure 7.9 presents the histogram of this set of annual drought flows.

The extreme value type III (Weibull) distribution was examined for a possible fit. This distribution has a lower bound (limit) and is usually skewed to the right, whereas Figure 7.9 indicates a skewed-left histogram. Therefore, the following transform was applied to the annual drought flows, which shifted and rotated the data in such a manner as to resemble the typical Weibull probability density function:

$$y' = y(0) - y_d \qquad (7.11)$$

where:

$y(0)$ = 234821 (truncation level corresponding to $q(0) = .35$.

y' = transformed annual drought flow

y_d = untransformed annual drought flow

The parameters (β, τ , ϵ) of the Weibull distribution were found by solving equations 7.12 through 7.14 simultaneously.

(7.12)

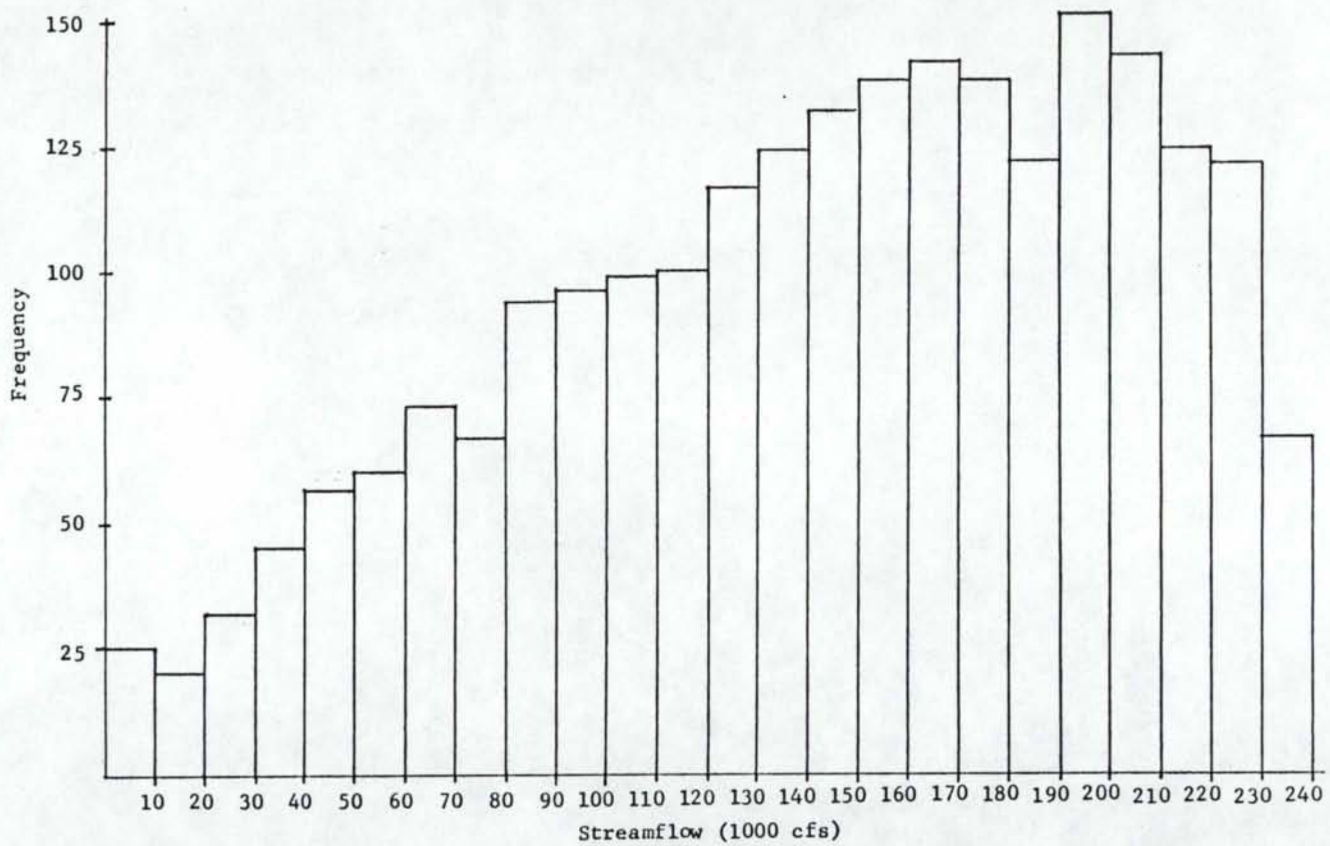


Figure 7.9 Histogram of Annual Drought Flows at Station 13186000 for q(0) = .35 and n = 50

$$g_{y'} = \frac{\Gamma(1 + 3/\tau) - 3\,\Gamma(1 + 2/\tau)\,\Gamma(1 + 1/\tau) + 2\Gamma^3(1 + 1/\tau)}{\left[\Gamma(1 + 2/\tau) - \Gamma^2(1 + 1/\tau)\right]^{3/2}}$$

$$\bar{y}' = \epsilon + (\beta - \epsilon)\,\Gamma(1 + 1/\tau) \qquad (7.13)$$

$$s^2_{y'} = (\beta - \epsilon)^2 \;\; \Gamma(1 + 2/\tau) - \Gamma^2(1 + 1/\tau) \qquad (7.14)$$

where:

$\bar{y}' = y(0) - \bar{y}_d$
$s_{y'} = s_y$
$g_{y'} = -g_y$
Γ = gamma function

for station 13186000, n=50, and q(0)=.35:

$\bar{y}$ = 143371
$y(0)$ = 234821
s_y = 58331
g_y = -.3954

The resulting parameters were:

τ = 2.419
β = 108378
ϵ = -41556

Table 7.15 presents the observed and expected frequencies of the annual drought flows. The expected frequencies were determined as follows:

$$P(y_d \le y_L) = 1 - P(y' \le y_L) \qquad (7.15)$$

Therefore,

$$CDF(y_d) = 1 - CDF(y') \qquad (7.16)$$

The CDF of the Weibull distribution for y' is

$$CDF(y') = 1 - \exp\left(-\left[(y' - \epsilon)/(\beta - \epsilon)\right]^{\tau}\right) \qquad (7.17)$$

Combining equations 7.16 and 7.17

$$CDF(y_d) = \exp\left(-\left[(y' - \epsilon)/(\beta - \epsilon)\right]^{\tau}\right) \qquad (7.18)$$

The expected frequency within a class interval can then be found by:

$$CDF(y_d \text{ upper limit}) - CDF(y_d \text{ lower limit}) \quad (7.19)$$

Table 7.15

Expected and Observed Frequencies of Annual Drought Flows based on Maximum Run-Sums at Station 13186000 (n=50 and q(0)=.35, Assumed Distribution: Weibull)

Class Range for y ($x10^3$) cfsd (1)	Observed Frequency (2)	Class Range for y' ($x10^3$) cfsd (3)	Expected Frequency (4)	Column $(2-4)^2/4$
0 - 20	45	234.8 - 214.8	30	7.50
20 - 40	77	214.8 - 194.8	54	9.80
40 - 60	117	194.8 - 174.8	89	8.81
60 - 80	140	174.8 - 154.8	134	0.27
80 - 100	190	154.8 - 134.8	185	0.14
100 - 120	199	134.8 - 114.8	236	5.80
120 - 140	241	114.8 - 94.8	277	4.68
140 - 160	270	94.8 - 74.8	298	2.63
160 - 180	280	74.8 - 54.8	292	0.49
180 - 200	273	54.8 - 34.8	258	0.87
200 - 220	267	34.8 - 14.8	202	20.91
220 - 234.8	189	14.8 - 0.0	105	59.92
	----		----	------
	2288		2160	121.81

$\chi^2(95\%) = 19.68$

Although the Weibull distribution does not pass the chi-squared test at the 95% significance level (Table 7.15), it appears to offer a reasonable representation of the annual drought flows except in the upper tail of the distribution. However, the lower tail area of the distribuition would be the most critical, as the lowest flows dictate most engineering designs.

MEAN ANNUAL DROUGHT: Millan and Yevjevich (27) arrived at the following general conclusions when considering the conditional probabilities of the maximum run-sums given run-length, and the maximum run-lengths given run-sums:

1) The run-length corresponding to the maximum run-sum is always smaller than the longest run-length for a given probability.

2) As the run-length increases, the distributions converge.

3) The run-sum corresponding to the maximum run-length is always smaller than the maximum run-sum for a given probability.

4) As the run-sum increases the two distributions converge.

The average annual flow during the maximum length and deficit periods do not necessarily behave in the same mannner as the above conditional probabilities. Table 7.16 presents a calculation of drought lengths, assuming that the mean annual drought (defined from maximum run-sum considerations) lasts long enough to produce a total deficit equivalent to S_m for that sample size. Then using Figure 7.3, the return period of the calculated drought length is determined and compared to the sample size. These calculations are explained in more detail below:

1) For each truncation level ($q(0)$ = .50 and .35) and sample size (n = 25, 50, and 100 years), the representative maximum run-sum S_m is determined using Figure 7.8 and the standard deviations from the extended records at stations 12413000 and 13186000 (same values as in Table 7.12).

2) The corresponding mean annual drought, y_d, from Table 7.14 is subtracted from the truncation level, thus defining an average annual deficit, $\bar{S}$.

3) The maximum median run-sum S_m, is divided by $\bar{S}$ which yields a drought length L. This length would produce S_m at a uniform annual flow rate of y_d.

4) Figure 7.3 is used to estimate the return period of L by assuming $L = L_m$.

Table 7.16

Return Periods of Drought Lengths Based on the Consideration of Mean Annual Drought Years

Station	q(0)	n yrs	Ave. Ann. Deficit cfsd $\bar{S}=y(0)-\bar{y}_d$	Figure 7.8 D_m	$S_m = D_m s_y$ cfsd	$L = S_m/\bar{S}$ yrs	Fig 7.3 T of L yrs
12413000	.50	25	202318	3.4	698000	3.45	30
12413000	.50	50	205656	4.3	883000	4.29	52
12413000	.50	100	208967	5.2	1070000	5.12	86
12413000	.35	25	195407	2.1	435000	2.22	31
12413000	.35	50	204280	2.67	550000	2.69	49
12413000	.35	100	205475	3.25	663000	3.23	92
13186000	.50	25	92946	3.4	326000	3.51	31
13186000	.50	50	94954	4.3	412000	4.34	54
13186000	.50	100	97255	5.2	498000	5.12	86
13186000	.35	25	88108	2.1	203000	2.30	33
13186000	.35	50	91450	2.67	256000	2.80	55
13186000	.35	100	92123	3.25	309000	3.35	93

From Table 7.16 it can be seen that the return periods of the run-length required for the mean annual deficits to equal the representative run-sum are consistent with the return periods of the representative run-sum. Consequently, the values of the mean annual droughts as based on the representative run-sum periods provide a reasonable description of an average flow during an n-year drought. If these flows continued at a uniform rate for an n-year drought-length, they would yield a total deficit that approximates the n-year representative deficit.

7.9 Distributions of Monthly Flows

The preceding analysis of annual flows indicate that an average drought year can be well defined by the distribution of flows during maximum run-sum periods. Therefore, this section will also concentrate on the monthly flow values associated with maximum run-sum periods.

STATISTICS OF MONTHLY FLOWS DURING DROUGHT YEARS: Tables 7.17 and 7.18 present the mean monthly flows based on the maximum run-sum periods for each n (25, 50, and 100 years) and q(0) (.50 and .35) combination at stations 12413000 and 13186000, respectively.

Table 7.17

Mean Monthly Flows Based on Maximum Run-Sum Periods
at Station 12413000 (cfs days)

	Ext Rec	q(0) = .50			q(0) = .35		
Mon	Mean	n=25	n=50	n=100	n=25	n=50	n=100
Oct	12666	9216	9115	9026	7850	7704	7844
Nov	26926	15132	15030	14679	11763	11511	11505
Dec	50805	28965	28739	28416	22603	22049	21648
Jan	44500	24343	23554	22648	18858	17619	16546
Feb	53134	37635	37158	36308	31846	30788	30165
Mar	76559	56666	56378	55768	49884	49784	50183
Apr	167869	130448	130381	129613	114864	112363	111444
May	167862	121162	120352	120897	102617	100715	101352
Jun	57651	38735	38309	38380	32364	31510	31901
Jul	19300	14979	14939	14943	13395	13213	13375
Aug	10303	8571	8566	8549	7927	7882	7940
Sep	8758	8125	8121	8101	7788	7749	7788

Table 7.18

Mean Monthly Flows Based on Maximum Run-Sum Periods
at Station 13186000 (cfs days)

	Ext Rec	q(0) = .50			q(0) = .35		
Mon	Mean	n=25	n=50	n=100	n=25	n=50	n=100
Oct	72178	6388	6172	6046	6541	6434	6328
Nov	73613	6103	5953	5889	6373	6324	6259
Dec	75324	5674	5569	5521	6158	6091	6018
Jan	73679	5896	5818	5813	6261	6222	6182
Feb	68583	5569	5488	5483	5933	5898	5846
Mar	108730	8000	7874	7845	8846	8766	8705
Apr	360469	20851	20426	20248	24774	24440	24142
May	797754	40871	39756	40035	51283	50837	50213
Jun	701874	29331	28604	28271	39737	39141	38473
Jul	234519	8814	8661	8539	12102	11938	11715
Aug	8619	4729	4637	4607	5707	5649	5591
Sep	6587	4486	4413	4401	5065	5031	4998

As with the annual means, it was observed from Tables 7.17 and 7.18 that the sample size n, has a much smaller effect on the monthly means than the truncation level q(0). A similar examination of the monthly standard deviations and skew coefficients (Tables E.14 through E.25 in Appendix E) based on the maximum run-sums leads to the same conclusion.

COMPARISION OF DROUGHT AND EXTENDED RECORD MONTHLY FLOWS: In order to reduce the number of subsequent analyses, the monthly statistics for only one sample size (n = 50) will be presented. Tables 7.19 throught 7.22 present a comparision of the monthly drought means and standard deviations for n = 50 at stations 12413000 and 13186000 with the extended data statistics.

Table 7.19

Reductions in Monthly Means at Station 12413000 for a sample size of 50 (Based on Maximum Run-Sum)

Mon	Ext Rec Mean cfsd	q(0)=.35 Mean cfsd	% of Column 2	q(0)=.50 Mean cfsd	% of Column 2
Oct	12666	7704	60.8	9115	72.0
Nov	29926	11511	42.7	15030	55.8
Dec	50805	22049	43.4	28739	56.6
Jan	44500	17619	39.6	23554	52.9
Feb	53134	30788	57.9	37158	70.0
March	76559	49784	65.0	56378	73.6
April	167869	112363	66.9	130381	77.7
May	167862	100715	60.0	120352	71.7
June	57651	31510	54.7	38309	66.4
July	19300	13213	68.5	14939	77.4
Aug	10303	7882	76.5	8566	83.1
Sept	8758	7749	88.5	8121	92.7
Annual	696295	412886	59.3	490639	70.5

Table 7.20

Reductions in Standard Deviations at Station 12413000 for a sample size of 50 (Based on Maximum Run-Sum)

Mon	Ext Rec Std Dev cfsd	q(0)=.35 Std Dev cfsd	% of Column 2	q(0)=.50 Std Dev cfsd	% of Column 2
Oct	8681	5525	63.6	6157	70.9
Nov	25090	10222	40.7	12251	48.8
Dec	52424	19803	37.8	24935	47.6
Jan	39677	15890	40.0	19498	49.1
Feb	43327	24749	57.1	28321	65.4
March	42120	31340	74.4	32632	77.5
April	59668	48893	81.9	50478	84.6
May	70537	51634	73.2	53525	75.9
June	33609	18137	54.0	19875	59.1
July	7466	5180	69.4	5546	74.3
Aug	2798	2202	78.7	2247	80.3
Sept	2210	1961	88.7	2047	92.6
Annual	205315	134254	65.4	139411	67.9

Table 7.21

Reductions in Monthly Means at Station 13186000 for a sample size of 50 (Based on Maximum Run-Sum)

Mon	Ext Rec Mean cfsd	q(0)=.35 Mean cfsd	% of Column 2	q(0)=.50 Mean cfsd	% of Column 2
Oct	7217	6172	85.5	6434	89.2
Nov	7361	5953	80.9	6324	85.9
Dec	7532	5569	73.9	6091	80.9
Jan	7367	5818	79.0	6222	84.5
Feb	6858	5488	80.0	5898	86.0
March	10873	7874	72.4	8766	80.6
April	36046	20426	56.7	24440	67.8
May	79775	39756	49.8	50837	63.7
June	70187	28604	40.8	39141	55.8
July	23451	8661	36.9	11938	50.9
Aug	8619	4637	53.8	5649	65.5
Sept	6587	4413	67.0	5031	76.4
Annual	271726	143371	52.8	176772	65.0

Table 7.22

Reductions in Standard Deviations at Station 13186000 for a sample size of 50 (Based on Maximum Run-Sum)

Mon	Ext Rec Std Dev cfsd	q(0)=.35 Std Dev cfsd	% of Column 2	q(0)=.50 Std Dev cfsd	% of Column 2
Oct	1437	1671	*	1510	*
Nov	1683	1468	87.2	1365	81.1
Dec	2330	2023	86.8	2014	86.4
Jan	1523	1259	82.7	1143	75.0
Feb	1504	1368	91.0	1286	85.5
March	3673	3040	82.8	3051	83.1
April	18034	11047	61.3	11703	64.9
May	32532	22228	68.3	23875	73.4
June	36275	20985	57.8	23468	64.5
July	14635	6715	45.9	7765	53.0
Aug	3337	2128	63.8	2210	66.2
Sept	1839	1532	83.3	1505	81.8
Annual	95757	58331	60.9	62546	65.3

* not calculated

After examining Tables 7.19 through 7.22 the following observations were made:

1) As expected, both the mean and standard deviations were reduced during the drought years as compared to the parent distribution values. The reduction was greatest at a truncation level of .35, due to the fact that the drought years corresponding to a truncation level of .50 consisted of some "larger" flows not defined as droughts when the truncation level equaled .35 (section 7.8).

2) In the record at station 13186000, October is an anomoly. The "reduced" standard deviation is actually larger than the original value. This month has a very small annual/monthly correlation

and was not modeled by Lane's disaggregation model as were the rest of the months (section 6.10).

3) The reduction is, in general, greater for those months with a large coefficient of variation, as determined from the extended data. The percentages from Tables 7.19 through 7.22 are plotted against the coefficient of variation in Figures 7.10 and 7.11 for a truncation level of .50 to illustrate this behavior.

4) As a another illustration, Figures 7.12 and 7.13 present plots for each station of the monthly mean percentages versus the monthly standard deviation percentages from Tables 7.19 through 7.22. The plot for station 12413000 (Figure 7.12) shows a fairly close relationship along a line of equal percentages, with the exception of the spring months (March through May). On the other hand, the plot for station 13186000 (Figure 7.13) exhibits considerable spread.

Based upon the preceding observations it was concluded that those months with the highest values of the coefficient of variation tend to have their statistics reduced the most during drought years. As the mean is reduced, the standard deviation tends to be reduced proportionally, preserving to some extent the historical coefficient of variation.

DISTRIBUTIONS OF MONTHLY FLOWS DURING DROUGHT PERIODS: The probability density functions of several monthly drought flows were examined by comparing their histograms to the extended data histograms (the parent distribution). Figures 7.14 through 7.25 present the monthly drought histograms based on the maximum run-sums for $n = 50$ and $q(0) = .50$ and

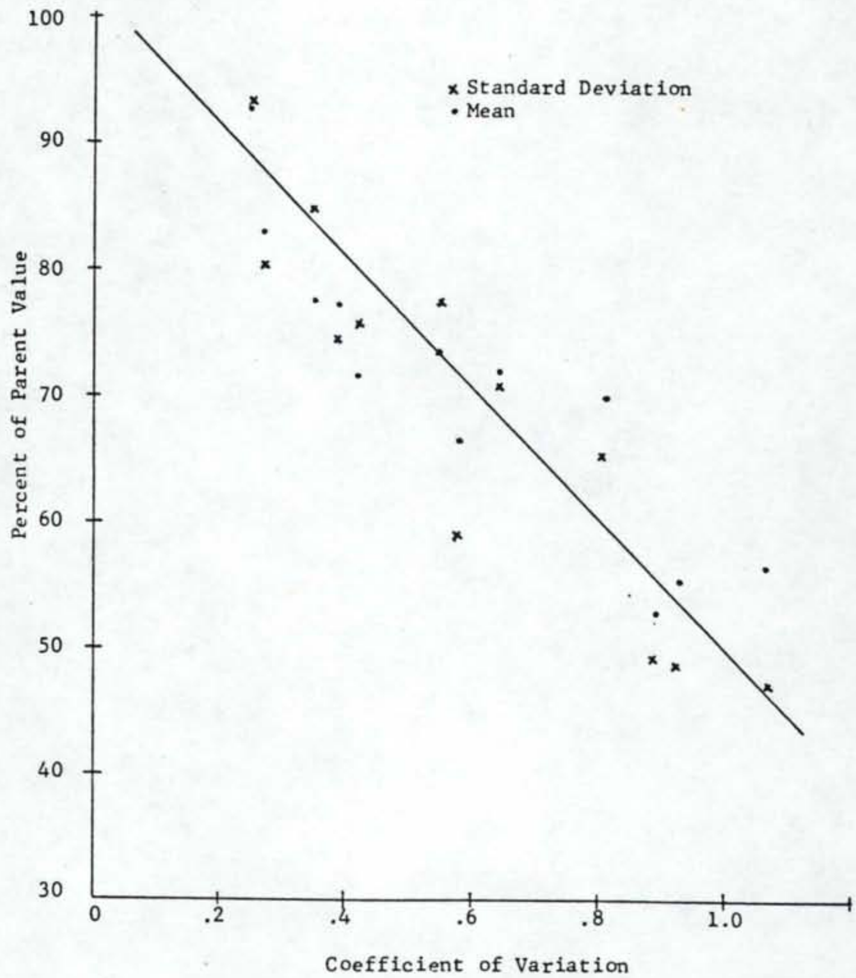


Figure 7.10 Relationship between Percent Reductions in Parent Monthly Statistics from a Drought with q(0) = .50 and n = 50, and Parent Coefficient of Variation for Station 12413000

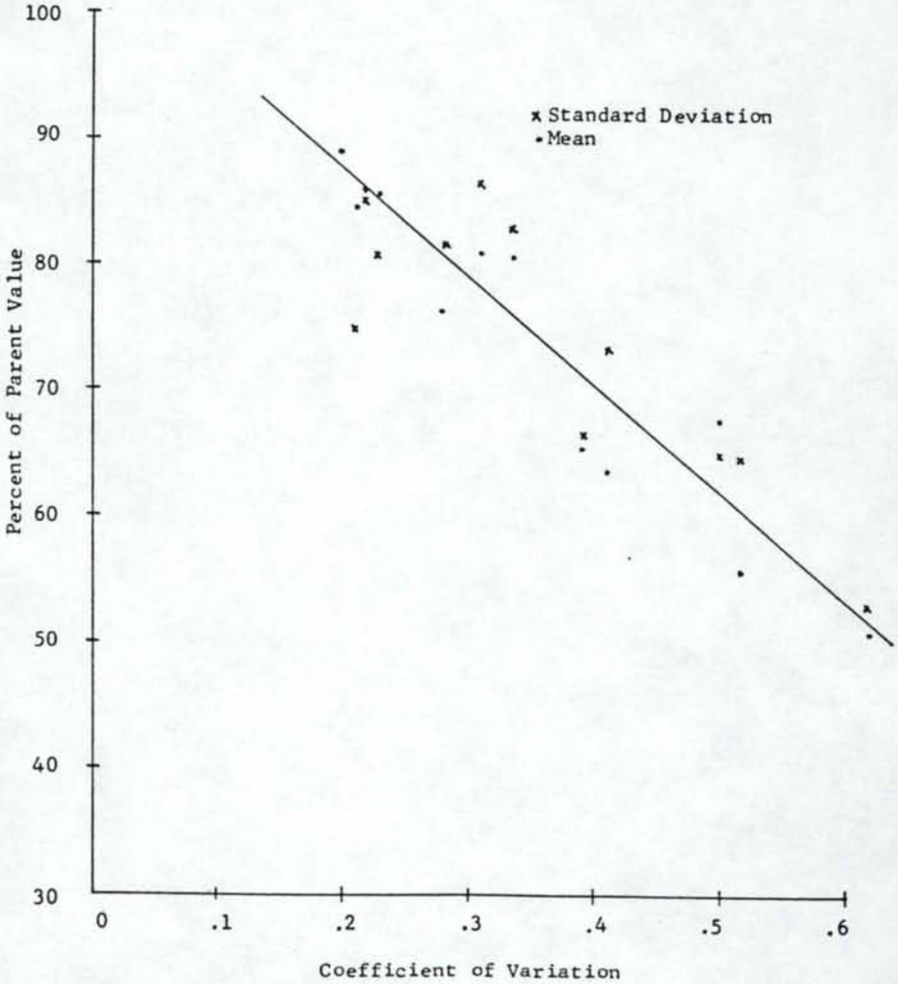


Figure 7.11 Relationship between Percent Reductions in Parent Monthly Statistics from a Drought with q(0) = .50 and n = 50, and Parent Coefficient of Variation for Station 13186000

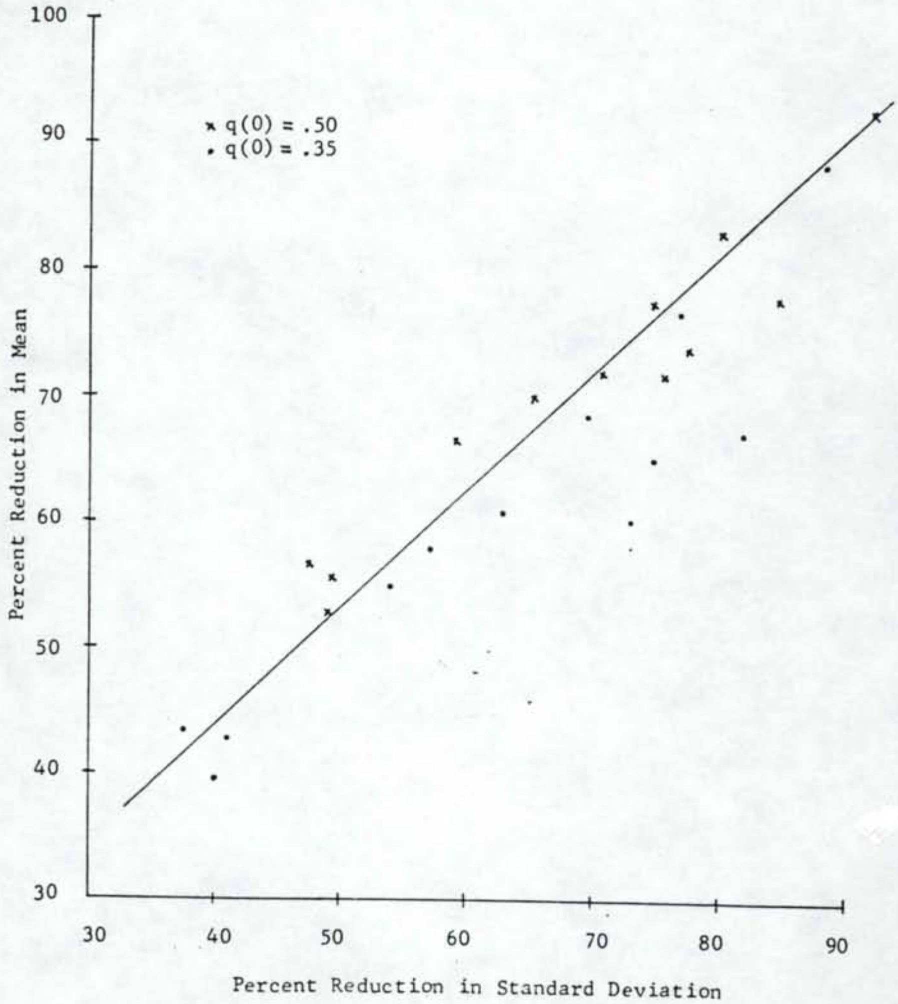


Figure 7.12 Relationship between Percent Reduction in Standard Deviation and Percent Recuction in Mean from a 50-year Drought at Station 12413000

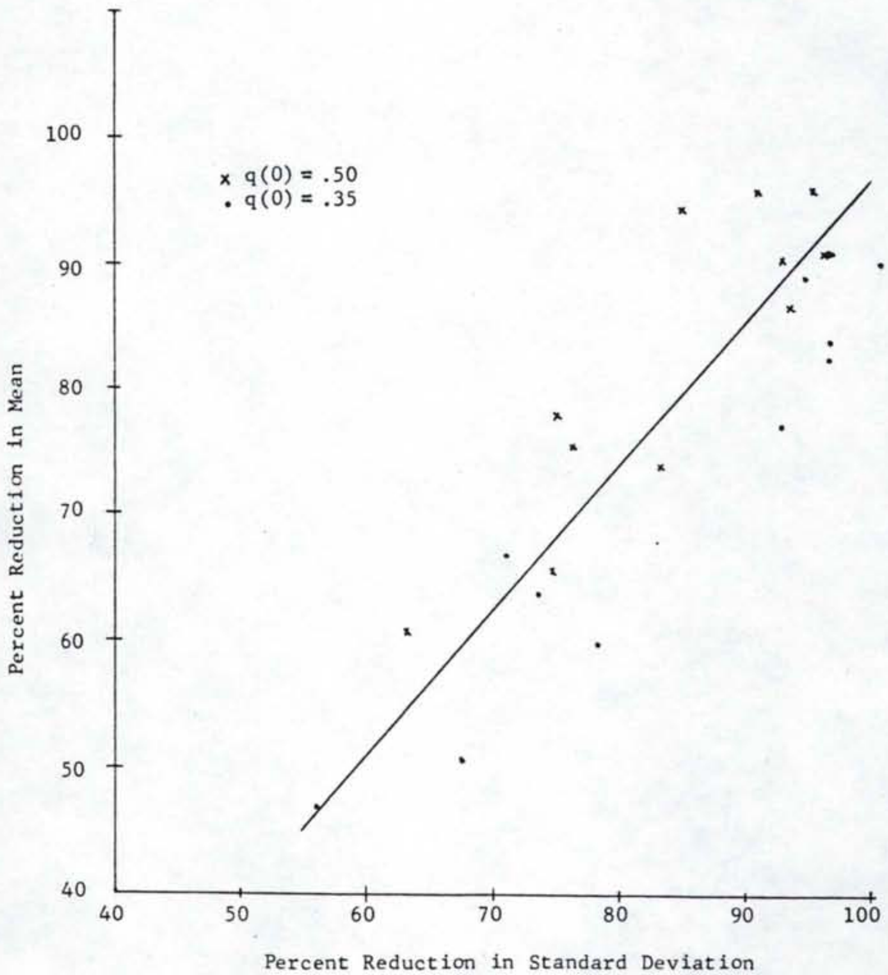


Figure 7.13 Relationship between Percent Reduction in Standard Deviation and Percent Recuction in Mean from a 50-year Drought at Station 1318600

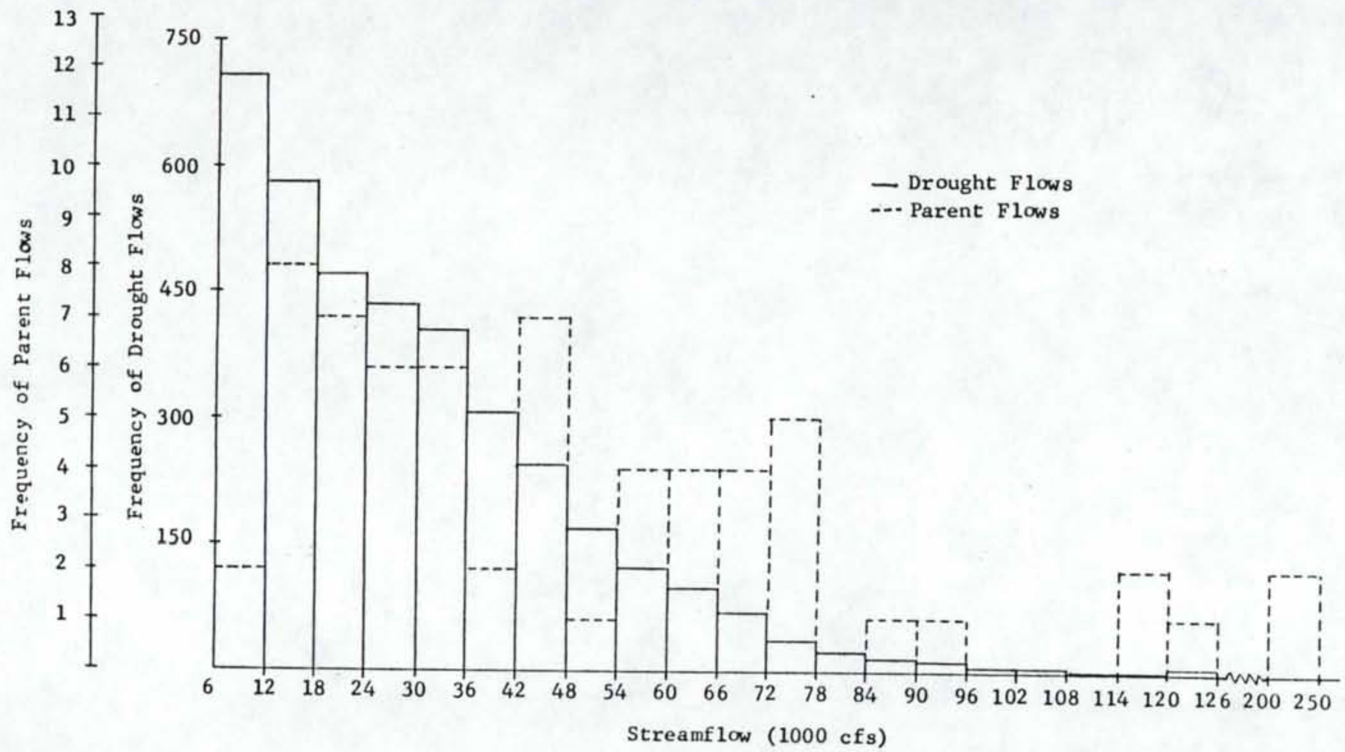


Figure 7.14 Histogram of January Parent and Drought Flows
(q(0) = .50, n = 50) at Station 12413000

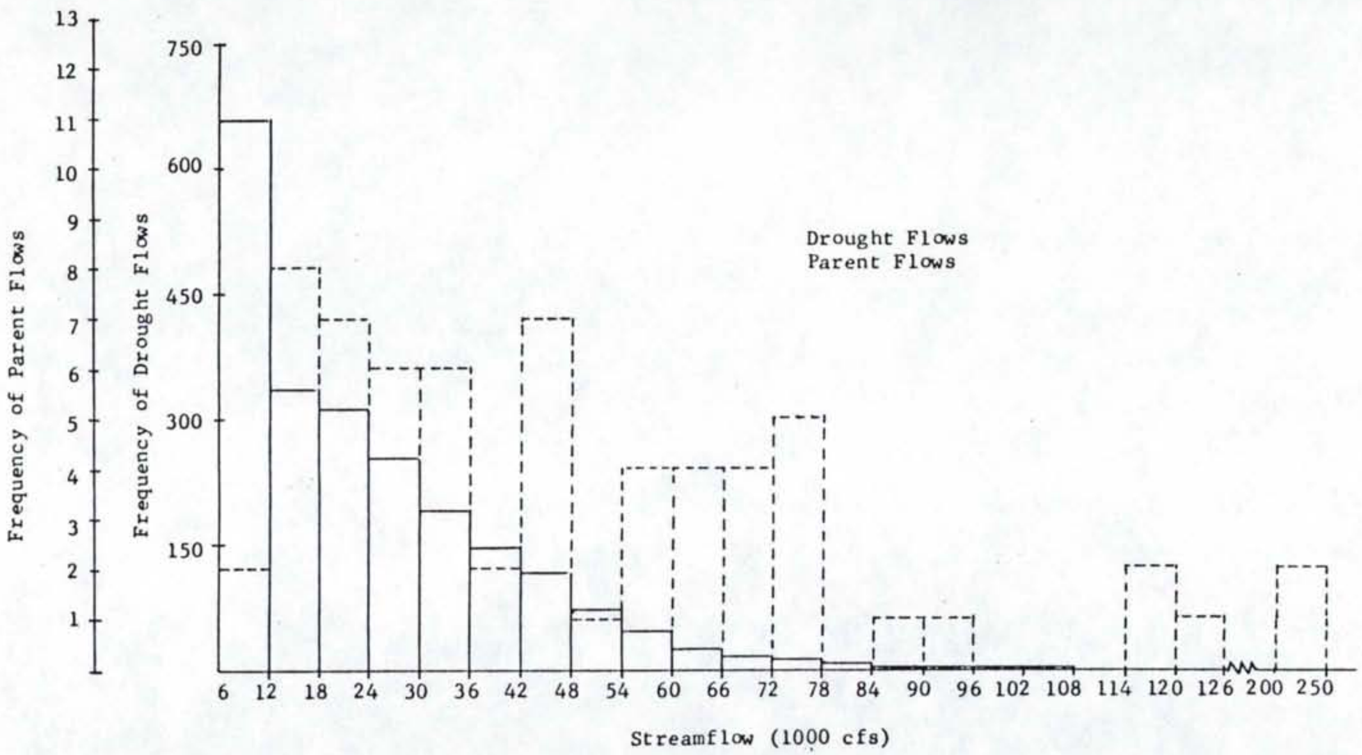


Figure 7.15 Histogram of January Parent and Drought Flows (q(0) = .35, n = 50) at Station 12413000

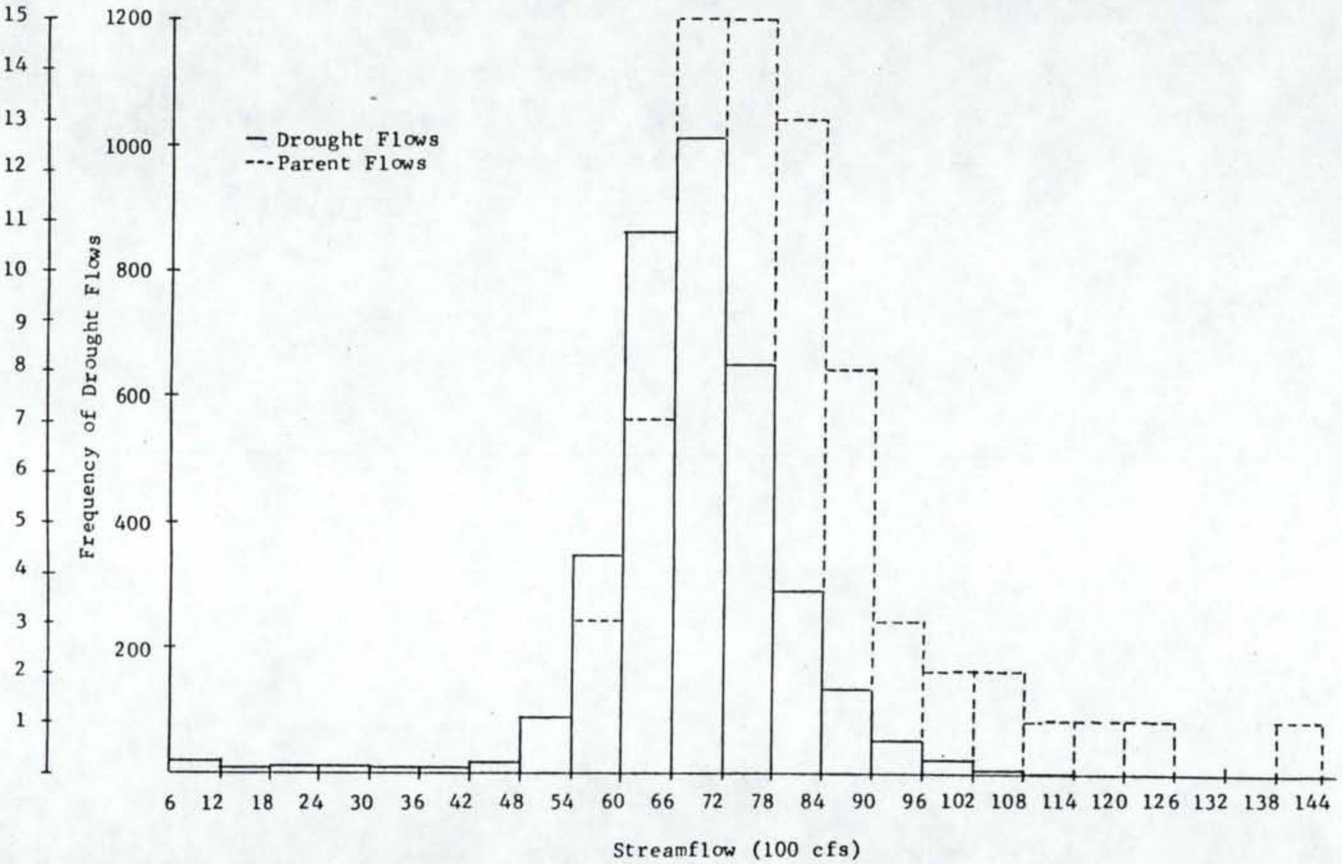


Figure 7.16 Histogram of January Parent and Drought Flows
($q(0)$ = .50, n = 50) at Station 13186000

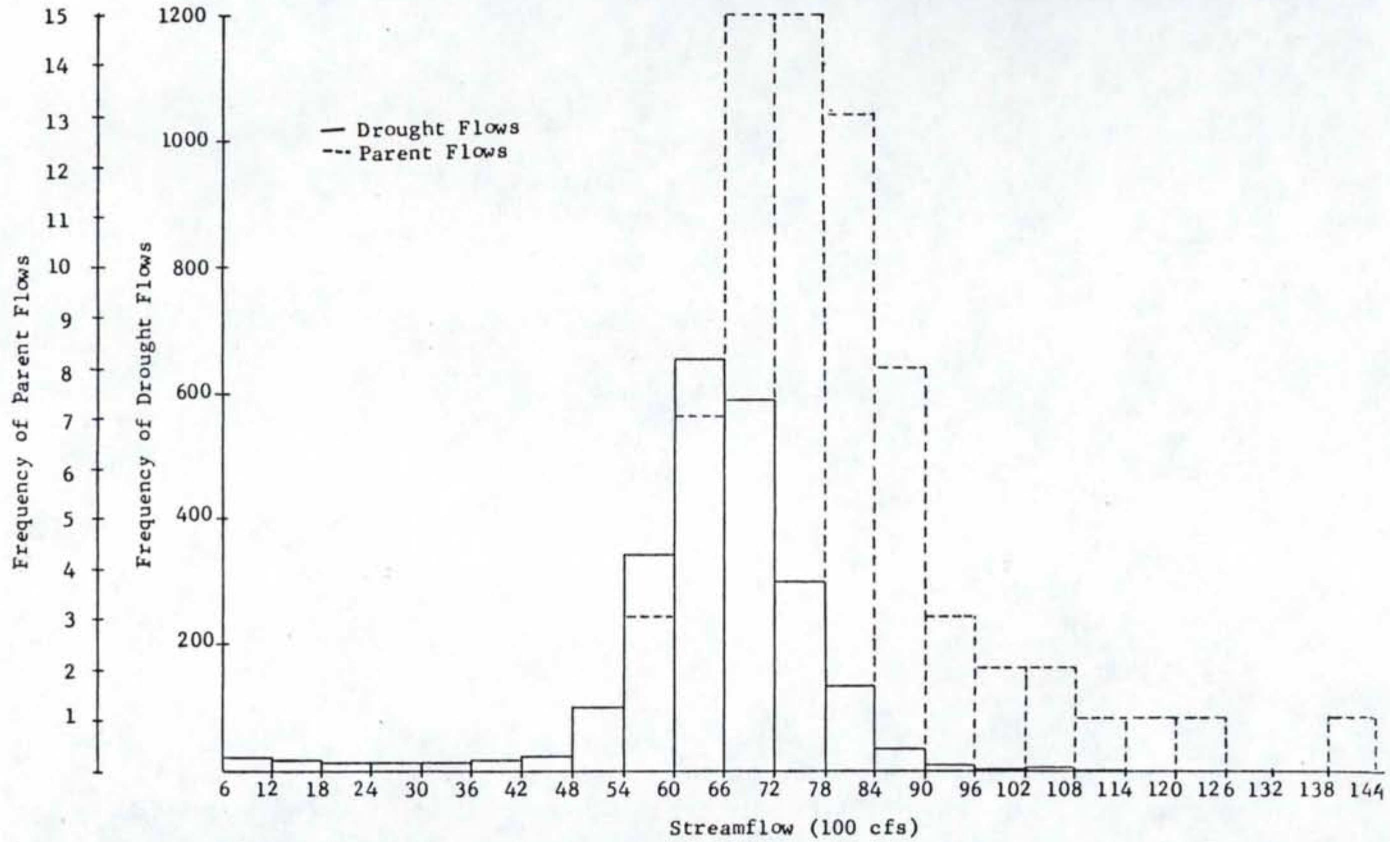


Figure 7.17 Histogram of January Parent and Drought Flows ($q(0) = .35$, $n = 50$) at Station 13186000

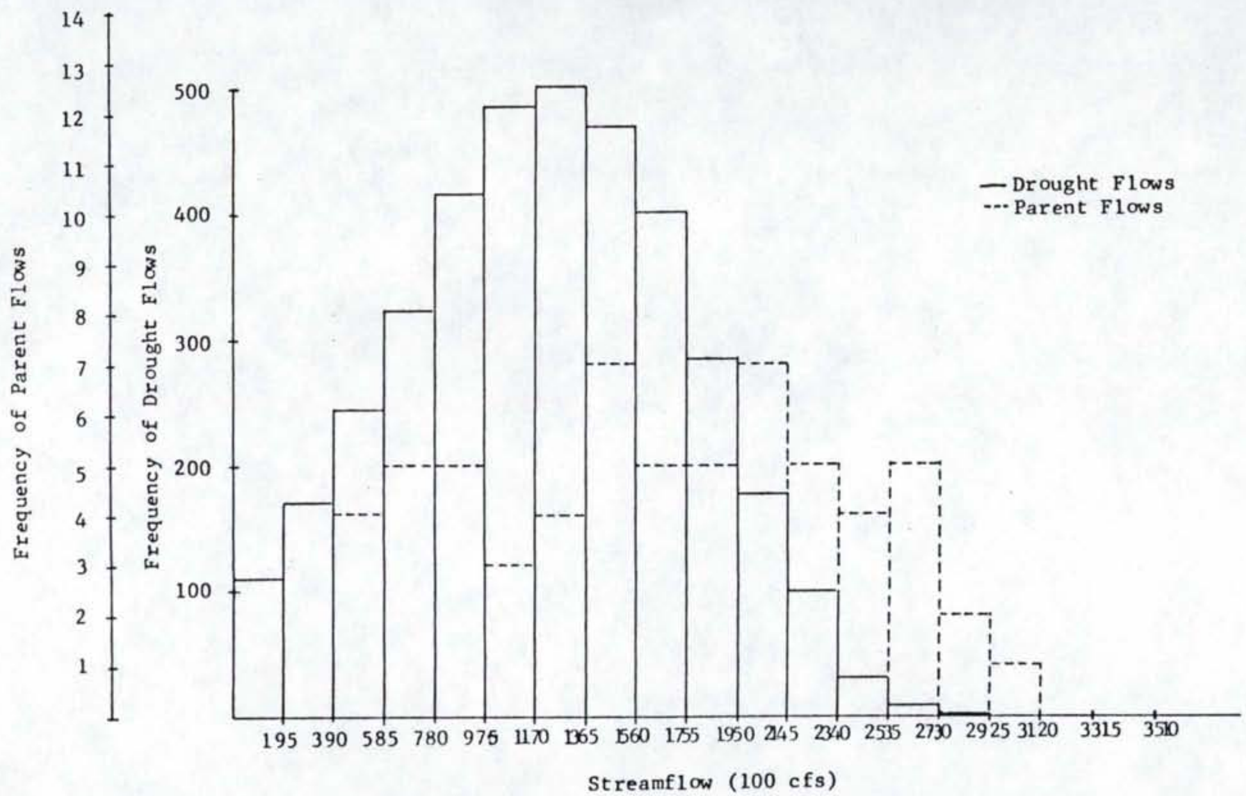


Figure 7.18 Histogram of May Parent and Drought Flows
($q(0) = .50$, $n = 50$) at Station 12413000

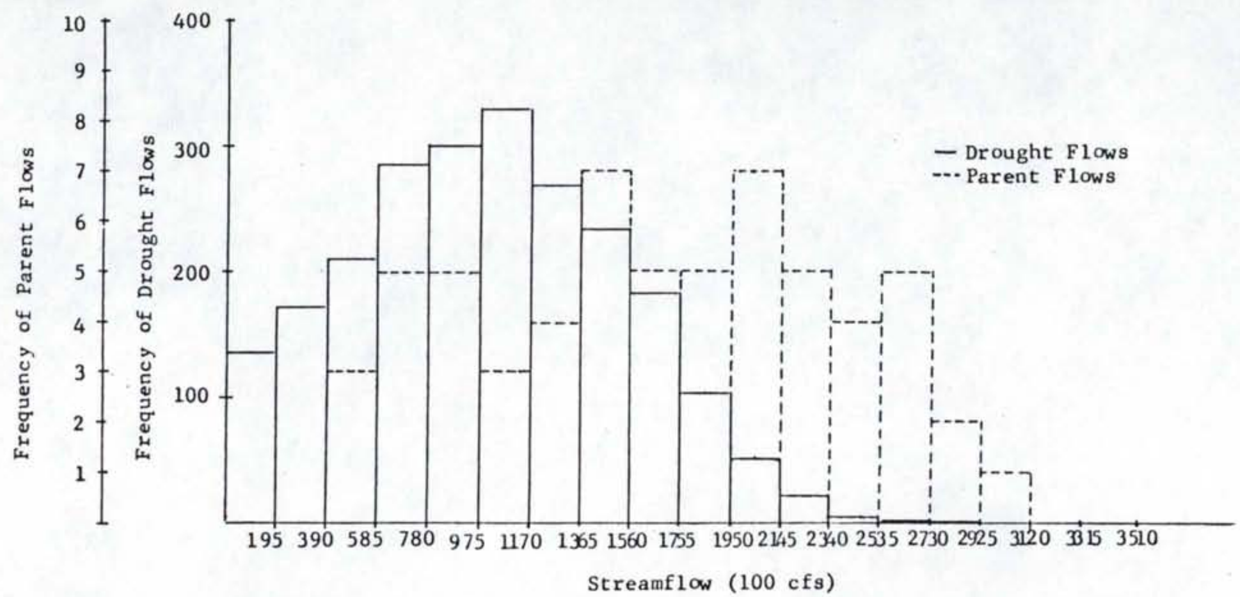


Figure 7.19 Histogram of May Parent and Drought Flows ($q(0) = .35$, $n = 50$) at Station 12413000

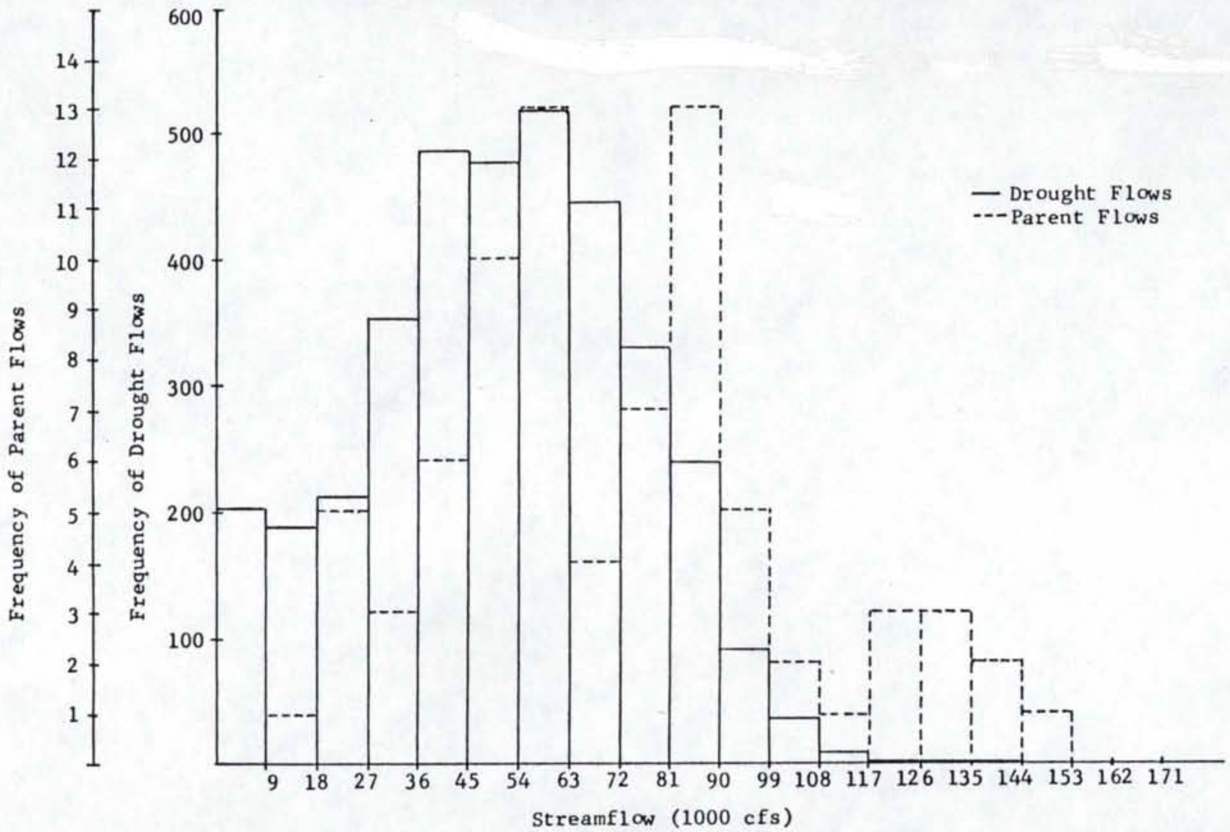


Figure 7.20 Histogram of May Parent and Drought Flows
($q(0) = .50$, $n = 50$) at Station 13186000

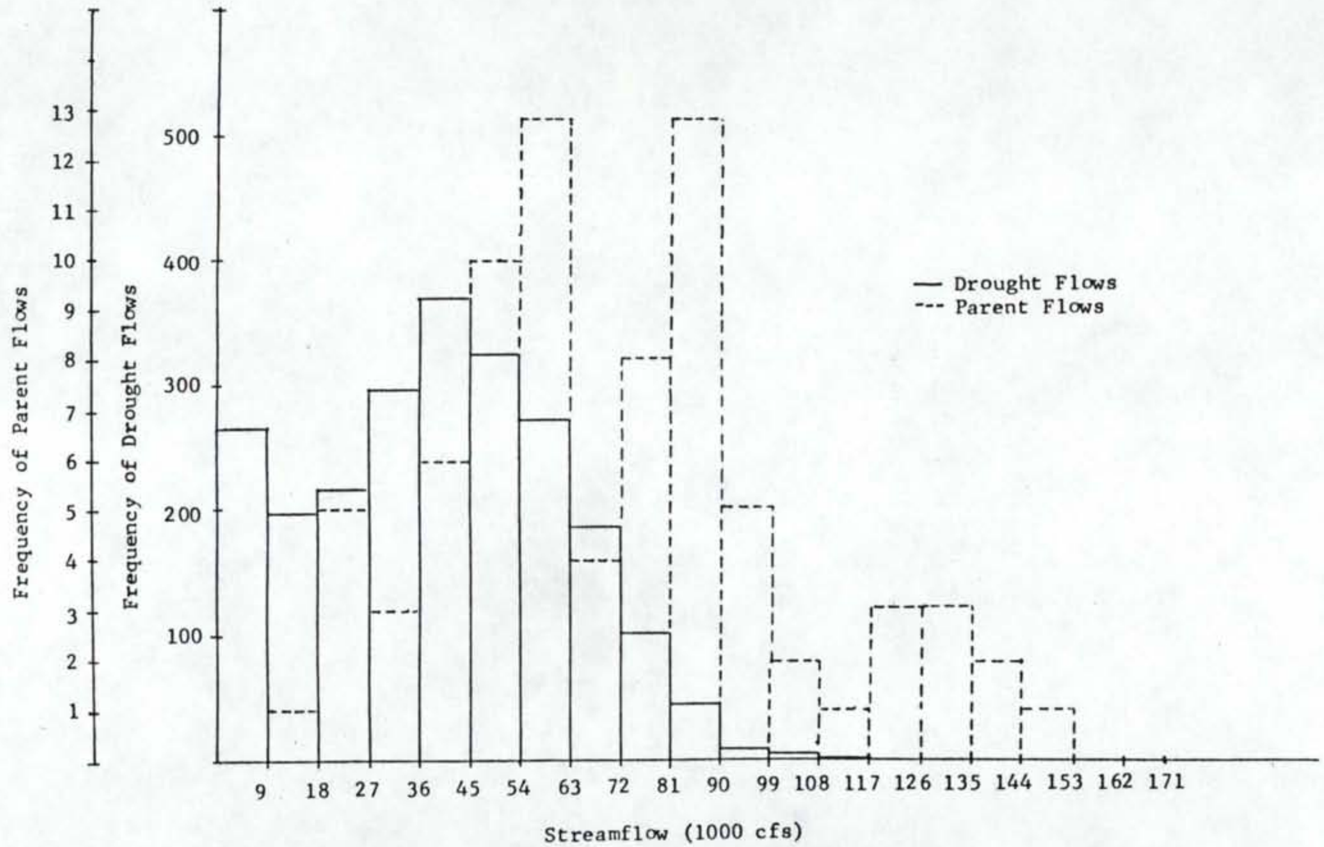


Figure 7.21 Histogram of May Parent and Drought Flows
($q(0) = .35$, $n = 50$) at Station 13186000

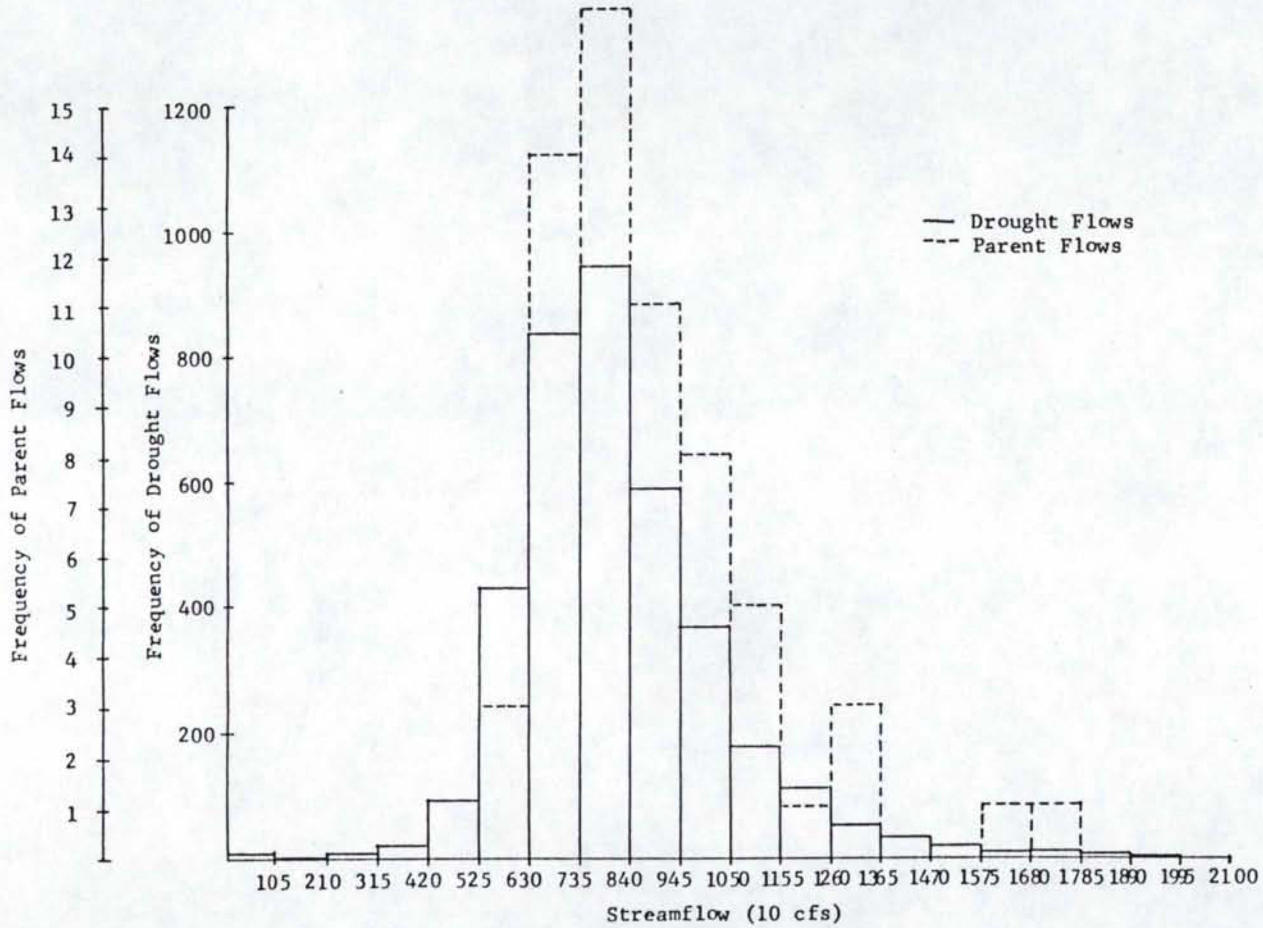


Figure 7.22 Histogram of September Parent and Drought Flows
($q(0)$ = .50, n = 50) at Station 12413000

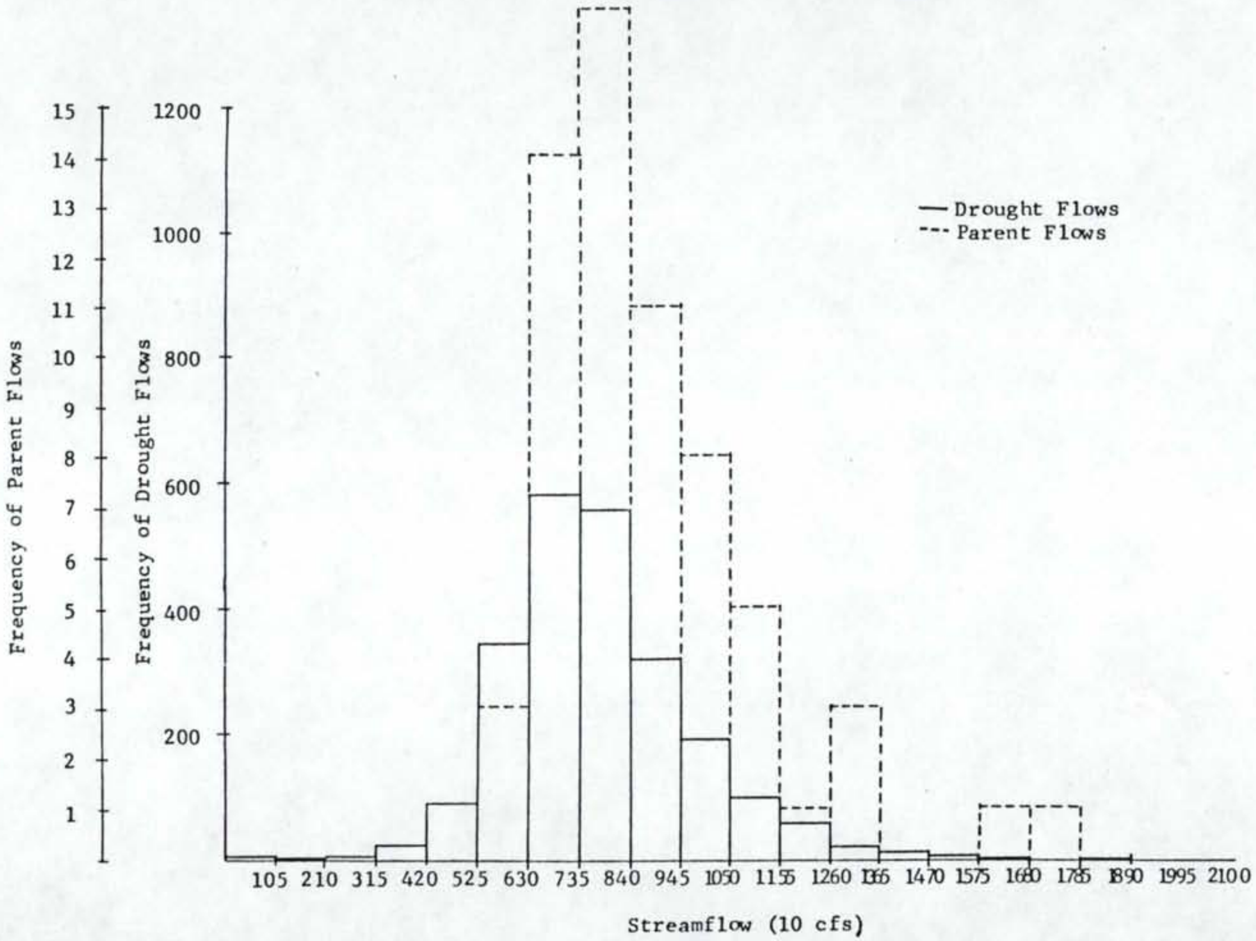


Figure 7.23 Histogram of September Parent and Drought Flows
($q(0)$ = .35, n = 50) at Station 12413000

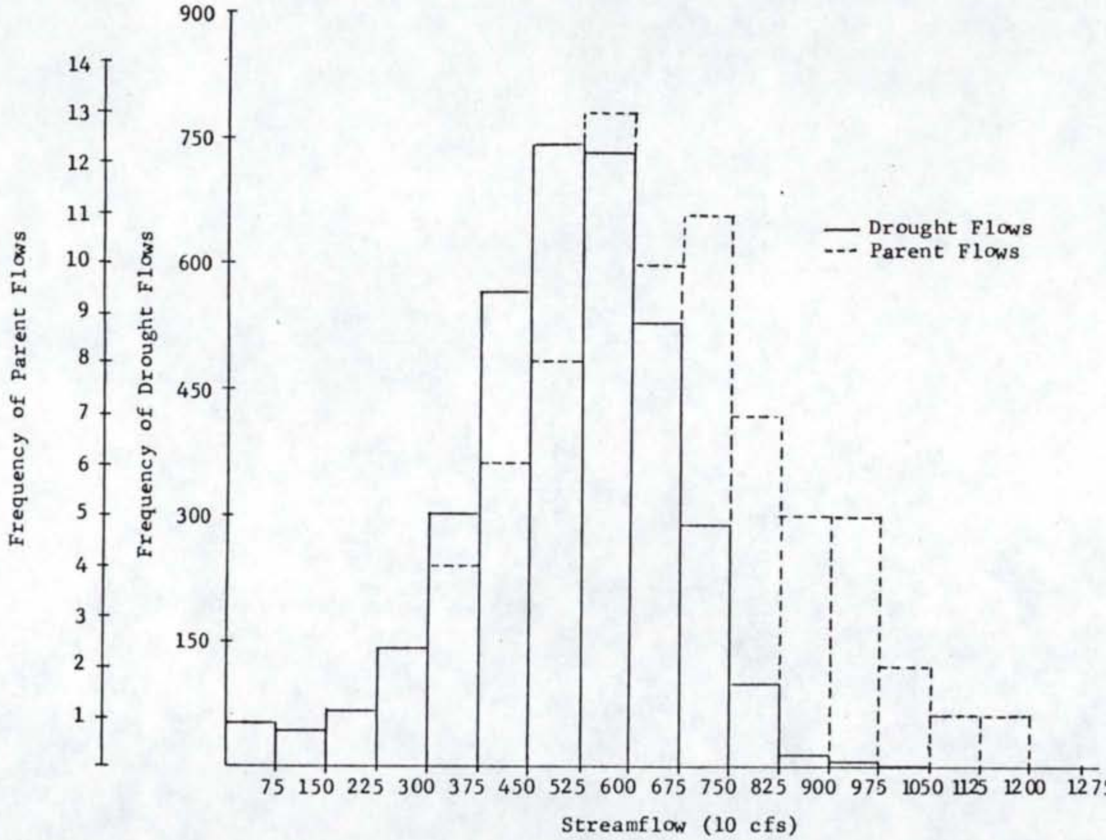


Figure 7.24 Histogram of September Parent and Drought Flows
($q(0) = .50$, $n = 50$) at Station 13186000

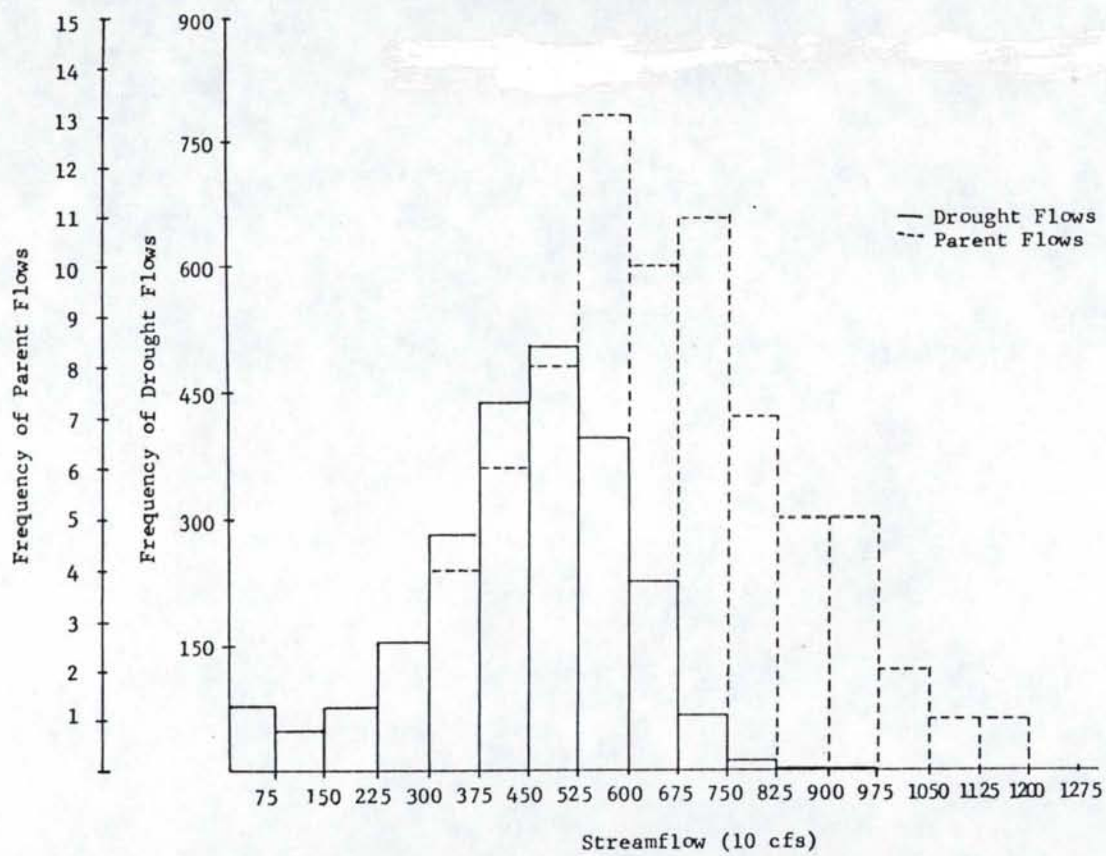


Figure 7.25 Histogram of September Parent and Drought Flows
($q(0) = .35$, $n = 50$) at Station 13186000

.35 for the months of January, May and September at both stations.

January represented a month of medium flow relative to the rest of the year. The parent distribution of January flows at both stations was previously assumed to be lognormal. The drought flows at station 12413000 seem to follow an exponential type distribution while the drought flows at station 13186000 appear to follow an extreme value type distribution. The coefficient of variation remains practically equal and the skew is less for the drought distributions as compared to the parent distributions. In all cases, the monthly distributions appear to represent a truncated version of the parent distributions distribution with the addition of more low flows.

May represented a month of high flow relative to the rest of the year. The parent distribution of May flows was lognormal at station 13186000 and normal at station 12413000. Yet, the drought distributions at both stations appear similar, and at the higher truncation level, are almost normal. As the truncation level decreases the distribtions became more uniform, with a straigth line slope to the right. Again, the drought distributions resemble a truncated version of the parent distributions with the addition of more low flows.

September represented a month of low flow relative to the rest of the year. The parent distributions of September flows were lognormal at both stations. Likewise, the drought distributions at both stations appear quite similar (possibly an extreme value distribution), and once again, resemble a truncated version of the parent distributions with the addition of more low flows.

In conclusion, it appears that the monthly drought distributions represent a truncated version of the parent distributions with the addition of more low flows. The higher the skew coefficient of the extended record, the fewer the number of low flows that appear in the drought distributions and the more closely the drought distribution resembles the parent distribution. However, no one distribution seems to fit all of the monthly drought flow distributions.

7.10 Use of Model Results for Storage Design

Section 7.7 provided an analysis of drought length and deficit as a function of return period for both stations. If a storage facility were to be designed for an economic life of 100 years, and the design return period selected to coincide with the economic life, then Table 7.13 would provide the median drought deficits for this return period (for two truncation levels). However, the concept of median or representative droughts is based on sampling theory, which says that if repeated samples of length 100 years were taken from the stochastic process, the 100-year representative drought would be exceeded in 50% of the 100-year samples.

The CDF's for the maximum run-sums (Figure 7.8) provide an estimate of the risk associated with a particular deficit. For example, at station 12413000 with n = 100, and q(0) = .35, Table 7.13 gives a median deficit of 663,000 cfs-days. Hence, in 50% of the samples with a 100-year length, the maximum deficit would be greater. Using the same concepts, from Figure 7.8, 100-year droughts corresponding to other exceedence probabilities can be determined as shown in Table 7.23.

Table 7.23

Exceedence Probabilities of 100-year Droughts at Station 12413000 (n=50, q(0) = .35)

Max Deficit (cfs days)	663000	870000	1000000	1100000
Risk of Being Exceeded in any 100-year period	50%	20%	10%	5%

A judicious design approach should include an evaluation of not only the median deficit but also the larger deficits associated with lower risk levels.

SUGGESTED APPROACH FOR DESIGN OF STORAGE RESERVOIR: Using the same example as represented in Table 7.23, the model output indicates that a large number of monthly streamflow traces were generated for maximum deficits corresponding to those shown. These are listed in Table 7.24.

Table 7.24

Number of Streamflow Sequences
within Stated Maximum Deficit Limits
Generated from Model at Station 12413000

Deficit Range (cfs days)	Average Risk Level (%)	No of Sequences
650000 - 700000	50	45
850000 - 900000	20	21
950000 - 1050000	10	17
1050000 - 1150000	5	15

As previously stated, the value in Table 7.24 corresponds to a truncation level of .35 which represents the average demand or desired streamflow yield. The design should consider the periodicity in the demand or yield. This can be done by constructing a simple reservoir optimiztion model, using monthly flows from the identified sequences above, and monthly design demands. Each sequence could be run through the model and evaluated. The model would include the flow sequences prior to and following the drought period to test for reservoir filling and refilling.

If a risk level of 10% were used in the example, 17 monthly flow records would be available from the disaggregation model to test in the optimization program. These sequences would probably yield a range of reservoir storage requirements, although the range should be fairly small. The average or median value of required storage would then be selected as the design value.

7.11 Conclusions

It was found that data extension improves the drought characteristic estimates at stations 12413000 and 13186000, since the population statistics estimated from the extended records are more reliable (section 3.5). The sample statistics are used to develop the annual and disaggregation models, which in turn, generate the output used for the drought analyses, and therefore, each step more accurately describes the respective streamflow processes if the best possible estimates of the population statistics are used.

From a comparision of the return periods of the historical droughts as defined by the record length and long-term stochastic processes, it was concluded that the return period of droughts should be assigned based on the long-term stochastic process. If the return period is assigned based on the historical record length, it may be low, reasonable, or high. The results were inconsistent and consequently, a design based on the historical record length as the return period may be under- or over-sized.

In general, the following trends were noticed in the annual and monthly drought statistics and distributions as determined from the maximum run-sum sequences:

1) The drought statistics are not greatly affected by the sample size, but both the mean and standard deviation are considerably affected by the truncation level (demand).

2) The drought distributions are a truncated version of the parent distribution (large flows deleted) with the addition of more low flow values.

3) The parent distributions with the largest coefficients of variation have the largest decreases in their corresponding drought statistics.

4) The coefficient of variation of the parent distribution tends to be preserved in the drought distributions.

5) The larger the skew coefficient of the parent distribution, the more the corresponding drought distribution resembles the parent distribution (in terms of over-all shape).

6) The mean annual and monthly drought flows provide a reasonable representation of the flow during drought periods.

7) The Weibull distribution provides a reasonable approximation of the annual drought streamflow distribution.

8) No one distribution appears to describe the distribtuions of the monthly drought flows.

9) The droughts as defined by the maximum run-sums are more severe than those defined by the maximum run-lengths for a given truncation level and sample size.

The method used in this chapter of identifying droughts in the annual series, followed by monthly disaggregation, allows the use of the theory of runs to define return periods for the stationary annual series. The relationships between the maximum run-length, maximum standardized run-sum and their associated return periods (Figures 7.3 and 7.8) can be used to assess the risk level of a design drought. Then the periodicity in the demand and supply can be further analyzed at the monthly level, with additional refinements accomplished by disaggregating the monthly flows into weekly, the weekly into daily, etc., until the desired time interval is reached.

The analyses in this and preceding chapters suggest several avenues which could be explored, depending upon the degree of sophistication desired, to develop regional low-flow characteristics of Idaho streams.

As stated in section 5.6, there seems to exist the possibility of regionalizing annual stochastic model parameters or similar annual flow characteristics, i.e.,ρ and γ. In the past, annual flow series have frequently been modeled by AR(1) models (19, 32). If the annual streamflow series in Idaho can be modeled with AR(1) models (or AR(0) when $p = 0$), then the CDF's of the maximum run-lengths and maximum standardized run-sums could be determined for each stream using the graphs developed by Millan and Yevjevich (27).

The CDF's of the maximum run-length and standardized run-sum could be used to estimate the relationships between the median maximum droughts and their return periods (same procedure as used to develop Figures 7.3 and 7.8). These relationships would then further provide an estimate of the maximum run-length and maximum standardized run-sum for a selected truncation level, return period and risk level. Next, an estimate of the annual standard deviation **σ**, would be needed, to calculate the actual deficit S, corresponding to the standardized run-sum D, ($S = D\sigma$).

Therefore, by regionalizing ρ, and γ, and estimating **σ**, the maximum deficit associated with a selected truncation level, return period, and risk level could be computed for any stream in Idaho.

If further detail is needed, several options may be possible:

OPTION 1: As mentioned in these conclusions, the sample size seems to have little affect, as compared to the truncation level, upon the statistics of the drought flows as defined by the maximum run-sums. Consequently, it may be possible to regionalize the percent of the mean annual historical flow represented by the mean annual drought flow for various truncation levels. Hence, by using an estimate of the mean annual historical flow at a stream, the mean annual drought flow which provides a reasonable estimate of flow during drought periods, could be determined.

Then, if the historical monthly flow statistics could be estimated, figures similar to Figure 7.10 could be used to estimate the percent reduction in each average monthly flow. These monthly drought flows would then be adjusted such that their sum equals the mean annual drought flow, and used to approximate the "typical" monthly flows during an average annual drought event.

OPTION 2: If sequences of annual drought flows were needed, and the percent reduction in the annual statistics (y, s and γ) corresponding to annual drought flows for a given truncation level could be regionalized, then the Weibull distribution along with ρ of the annual drought series could be used to construct an AR(1) model:

$$y_{d\,t} = r_d y_{d\,t-1} + s_d(1 - r_d^2)^{1/2}\lambda_w$$

where:

y_d = annual drought flow

r_d = lag-one serial correlation coefficient of annual drought series. Further research would be needed to develop this correlation, as it would probably be higher than for the entire annual series.

s_d = standard deviation of drought flows

λ_w = random standard deviate from Weibull distribution.

The parameters of the Weibull distribtion could be estimated using equtions 7.12 through 7.14. Therefore, if the percentages associated with the annual drought statistics could be regionalized, and the corresponding annual statistics estimated, then actual drought flow sequences could be generated. The lengths of the sequences would be determined by the design maximum deficit, when the sum of the annual deficits neared the design maximum deficit the sequence would be ended.

These annual sequences could then be disaggregated into monthly flows, if the parameters of the disaggregation models could be regionalized and estimates were avaiable for the statistics of the historical monthly streamflow values. Once again, further research would be needed to arrive at the disaggregation parameters corresponding to just the drought periods.

CHAPTER 8

SUMMARY AND CONCLUSIONS

This research study examined the application of stochastic disaggregation modeling techniques to two rivers in Idaho, both of which have the potential for future storage development or other regulation projects. These rivers were the Coeur d'Alene (station 12413000) and the South Fork of the Boise (station 13186000). This chapter summarizes the procedures used and the conclusions reached in this study.

8.1 Data Extension

Extended records at stations 12413000 and 13186000 were generated by using a multivarite model and the longer-term nearby records of stations 124135000 and 13185000, respectively. These two extended records along with the two original unextended records became the time series by which the effects of data extension upon stochastic model parameters and drought characteristics could be observed.

Four multivariate models were considered for use in extending the records at stations 12413000 and 13186000: simple linear regression, and models developed by Fiering (9), Lawrance (20), and Yevjevich (43). All of these models were developed to preserve the cross correlation between the shorter (subordinate: 12413000 and 13186000) and longer (key: 12413500 and 13185000) series, and except for the linear regression model, were also designed to preserve the lag-one serial correlation of both the shorter and longer series.

The residuals at the subordinate stations resulting from the Yevjevich model showed less time dependency and, for the most part, were less skewed than the residuals from the other multivariate models. The better performance of the Yevjevich model seemed to be due to the fact that the lag-one serial correlation coefficient of the subordinate record was used directly to relate the sucessive monthly flows, whereas the other multivariate models used a parameter which was only parially a function of the lag-one serial correlation coefficient, and the linear regression model did not even consider serial correlation. Consequently, the Yevjevich model was used to extend the records at both subordinate stations.

The extended portions of record were examined to make sure they were reasonable by comparing the statistics of the extended portions to the statistics of the actual historical records at each subordinate station. In addition, the statistics of the corresponding periods of time were computed for the key station records and compared to see if the same trends were observed between these statistics as in the corresponding subordinate station statistics. Several of the statistics seemed to be changing considerably. However, the same trends were seen between the corresponding statistics of the key and subordinate stations. These changes were found to be due to the drought condition which existed during the 1930's which was included in the extended portion of each record. Therefore, the extended portion of each record was considered reasonable and added to the historical record of each subordinate station.

As a result of this analysis, it was concluded that data extension should be performed prior to the estimation of any statistics, or model parameters, if an appropriate record of longer length which is strongly correlated to the shorter record, is available. The confidence intervals around the extended record statistics are smaller, and thus the statistics estimated from the extended record are more reliable than the statistics estimated from the shorter record. Consequently, any model parameters and record characteristics will be most accurately defined if developed from the most reliable statistic estimates.

8.2 Annual Models

Before estimating annual stochastic model parameters, the unextended and extended annual records at stations 12413000 and 13186000 were tested for normality by examining their histograms, coefficients of skew, and chi-squared values assuming a normal distribution. All of the annual series were found to approximate the normal distribution, and no change in the annual series distribution as caused by data extension or station location was noticed.

An ARMA(p,q) model was then fitted to each of the annual streamflow series. Annual ARMA(p,q) streamflow models utilize the serial correlation of streamflow values separated by 1 to p years, and the correlation that exists between sucessive residual values separated by 1 to q years.

There is a physical basis for the use of such models in describing annual flows (39). Annual streamflow for a given year is the result of effective precipitation occuring in that year plus a contribution from the previous years' precipitation in the form of groundwater discharge. Also, added to this is the effect of surface storage. The autoregressive component of the ARMA(p,q) model can be used to represent the contribution of streamflow from groundwater discharge (base flow) and long-term surface storage (such as a lake), while the moving average component can be related to the precipitation from the previous q years that resulted in relatively rapid drainage (overland flow and interflow).

As judged by the Akaike Information Criteria and a comparision of competing ARMA(p,q) models and historical correlograms, it was found that a pure probabilistic model (p = 0 and q = 0) most accurately represented the annual streamflow series at stations 12413000 and 13186000 based on the unextended and extended records. The fact that all of the annual models contained no serial correlation component (p = 0), suggests that these streams do not have significant over-year storage capabilities. In other words, the majority of the effective precipitation falling within a water year is discharged during that year. Furthermore, since no correlation was found between sucessive years' rapid drainage (q = 0), there seems to exist a complex relationship between the storage and rapid drainage components of flow, with their relative contributions from year to year being nonlinearly related.

Data extension did not change the form of the model at either station. However, the extended records give more reliable estimates of the annual flow statistics and thus more reliable model parameter estimates. Consequently, the extended models would be a more accurate representation of the streamflow process at each station. In addition, niether the form of the model nor the distribution of the annual flows changed between stations.

8.3 Disaggregation Models

Lane's condensed disaggregation model was chosen to be used in this study because of the fewer number of parameters. Had any of the earlier disaggregation models (Valencia-Schaake and Mejia-Rouselle models) been used, the principle of parameter parsimony would have been seriously violated. Lane's model is designed to preserve the linear cross correlation between annual and monthly values along with the lag-one serial correlations, variances and means of the annual and monthly values. It accomplishes this by preserving the means and standard deviations through normalization while the correlation structure is preserved by the actual model.

Before the model parameters could be determined the normality of the monthly series was tested by examining the coefficients of skew. This examination indicated that the assumption of normality for many of the months was not valid. There are two options for dealing with non-normality: 1) model the skewed data and account for the skewness in the residual term, or 2) find an appropriate transformation that would convert the skewed sequences into normally distributed sequences. For the second option, transformed sequences must be used for model generation and the inverse transform applied to obtain the actual streamflow values. This option was selected because it was the procedure recommended by Lane (19) when presenting his model.

Lognormal transforms were used to normalize the monthly series, since formulas exist which relate the statistics of the historical record to those of the transformed record (22, 25), and help to preserve the actual historical statistics during the modeling process. These relationships were therefore used to determine the statistics of the transformed series, which, in turn, were used in estimating the parameters for the disaggregation models. Although in several cases the monthly skew was not reduced enough to satisfy the assumption of normality, these few violations were accepted since no other transforms offered any significant advantages over the lognormal transforms.

Lane's model also assumes that the means of the normally distributed series equal zero. This assumption was satisfied by subtracting the means of the monthly transformed series. In addition, each transformed value was divided by its transformed standard deviation to create a standardized series, and this series was then used to estimate the parameters of Lane's model. Since the means and standard deviations as computed by the statistical relationships were used, the means and standard deviations of the standardized series differed slightly from zero and one.

At station 13186000, one parameter in Lane's model for the month of October was undefined. This seemed to be reasonable since the correlation coefficient between the month of October and the corresponding annual flows was very small (statistically equal to zero). Therefore, Lane's model, which accounted for this correlation, was inappropriate for this month. On the other hand, the lag-one serial correlation coefficient for the month of

October was quite large. As a result, a simple AR(1) model was used for the month of October at Station 13186000, while Lane's model was used for the other eleven months.

In order to check the performance of Lane's model, 500 years of monthly streamflow values were generated for each record: 12413000 - Unextended, 12413000 - Extended, 13186000 - Unextended, and 13186000 - Extended. The statistics of the synthetic records were computed and compared to the statistics of the historical records, and hypothesis tests were performed on the means and standard deviations.

Based on the hypothesis tests only the standard deviations for January of the unextended and extended records at station 12413000 were statistically unequal to the corresponding historical standard deviations. Sampling fluctuation seemed to be the reason for these differences. Also, a visual comparision was made of the synthetic and corresponding historical skew coefficients and correlation coefficients to see how well they were preserved. With only a few exceptions all of thse statistics appeared to be preserved quite well. As a result, it was concluded that the synthetic records satisfactorally preserved the historical statistics, and Lane's disaggregation model was accepted for further use.

A comparison of the disaggregation model parameter estimates of the unextended and corresponding extended records showed that the parameters of the extended series differed from those of the unextended series. This was to be expected since the historical statistics of these records changed after data extension. Since the extended record statistics were a more reliable description of the streamflow process, the disaggregation parameters estimated from these statistics should also be more reliable.

The disaggregation model parameters of the extended series at stations 12413000 and 13186000 were also compared and the following observations were made:

> 1) The parameter "Q" which relates the annual and monthly streamflow values followed the same general pattern except for the months of May and June. April and May are the months of maximum runoff at station 12413000, while both May and June have the heaviest runoff at station 13186000. The higher elevation at station 13186000 probably results in a later runoff series (extending between May and June) than at stations 12413000 (extending between April and May) possibly explaining some of the deviation in the "Q" parameters.
>
> 2) The parameter "G" which relates the residual and monthly streamflow values followed the same general pattern.
>
> 3) The parameter "H" which relates sucessive monthly streamflow values seemed to deviate more from any general pattern relative to the other two parameters.

Based on the preceding observations it was felt that the possibility of regionalizing disaggregation model parameters was good. However, there does appear to be some differences between the two sets of parameters suggesting that a further refinement of hydrologic regimes may be necessary. In other words, the parameters seem to follow the same general trends, yet if further records were analyzed between these two stations a set of parameters averaging the two might be found, allowing a further refinement of the region's disaggregation parameters.

Also, it appears that Lane's condensed disaggregation model is very robust, as can be seen by the fact that the nonnormality of the October series at stations 12413000, the change in transformations for the month of May (untransformed) at station 12413000, the change in models for the month of October (AR(1)) at station 13186000, and the fact that the transformed series used in modeling did not have means of exactly zero and standard deviations of one did not adversely affect the resulting synthetic series. Thus, if an appropriate estimate is obtained for each month's disaggregation model parameters, the resulting synthetic records would probably be reasonable.

Furthermore, it seems that since Lane's model adequately preserves the important statistics of the annual and monthly time series, the reduction in the number of parameters over the Valencia-Schaake and the Mejia-Rouselle models did not have severe adverse effects. In fact, several parameters of Lane's model could probably be set equal to zero for several of the months, as they are very close to zero. In conclusion, not all of the parameters of the Valencia-Scaake and Mejia-Rouselle models are needed, since many of these parameters can be set equal to zero without severly affecting the performance of the disaggregation model.

8.4 Droughts

The theory of runs was used to identify and assign probabilities to critical drought events. Runs as an objective defintion of droughts can best be applied to stationary time series (10). The annual series at stations 12413000 and 13186000 were stationary while the monthly series were not, and therefore, the maximum run-lengths and maximum run-sums were identified in the annual records, and then the corresponding monthly sequences were investigated.

In order to develop the "experimental" probability distributions of the maximum run-length (L(max)) and maximum run-sum (S(max)) based on the long-term stochastic processes at each station, a large number of synthetic streamflow sequences were generated (Monte Carlo method) and then probabilities of the maximum run characteristics were assigned based on relative frequencies. The synthetic streamflow sequences were generated using the annual and disaggreagion streamflow models previously developed for the unextended and extended records at stations 12413000 and 13186000.

For the purposes of this study, it was decided to examine the probability distributions of the maximum run-length and maximum run-sum for sample sizes of 25, 50, and 100 years using two different truncation levels (y(0)) which corresponded to q(0)=.50 and q(0)=.35 ($q = p(y_t \le y(0))$). Therefore, a computer program was developed to generate the synthetic sequences, identify the maximum run-lengths and maximum run-sums, determine their cumulative density functions (CDF's) and compute the statistics and flow distributions of the corresponding annual and monthly series. Following is a summary of the computer program's output for each record:

1) 40,000 years of monthly streamflow values

2) Cumulative density functions of the annual L(max) and S(max) for a 100-, 50-, and 25-year sequence (sample size) with y(0) corresponding to q(0)=.50 and q(0)=.35

3) Annual and monthly flow distributions and statistics corresponding to each annual series which composed the cumulative density functions listed above.

Since all of the annual series examined in this study were independent normal series, they had identical CDF's of the maximum run-length and standardized maximum run-sum ($D(max) = S(max)/\sigma$) for equal truncation levels and sample sizes.

While a preliminary assessment of the experimental CDF's of the maximum run-lengths and maximum run-sums indicated that data extension did not affect the relationships between the representative maximum run characteristics and their associated return periods, this result was considered misleading for the following reasons:

1) Data extension does not affect these relationships only when the unextended and extended record estimates of ρ and γ are equal. While in this study they were all equal to zero, this would not always be true for other records.

2) The actual maximum run-sum, S(max), equals the corresponding standardized maximum run-sum D(max), times the standard deviation σ, of the nonstandardized series ($S_m = D_m\sigma$). While D(max) may be unaffected by data extension, S(max) is affected as σ changes. Table 8.1 summarizes the changes in σ resulting from data extension.

3) The truncation level y(0) greatly influences the CDF's of the maximum run-lengths and the maximum run-sums, which in turn influence the assignment of the return periods (34) Usually, as was the case in this study, the statistics of the unextended and extended records will vary somewhat, and these statistics are used to estimate the truncation levels based on the records' assumed probability distributions. Table 8.1 also summarizes the changes in the truncation level resulting from data extension.

Table 8.1

Changes in the Standarad Deviation and Truncation Level Resulting from Data Extension (cfs days)

Station	q(0)	Standard Deviation Unextend	Standard Deviation Extend	Truncation Level Unextend	Truncation Level Extend
12413000	.35	198977	205315	630167	617166
12413000	.50	198977	205315	706854	696295
13186000	.35	94040	95757	258276	234821
13186000	.50	94040	95757	294519	271726

The size of the maximum run-length and maximum standardized run-sum corresponding to the median of their respective CDF's were defined as the representative maximum drought length and drought deficit, respectively, for a given sample size (return period) and truncation level. The representative maximum drought characteristics were used to determine if the maximum run-length and maximum run-sum observed in the historical records were "representative" of their respective sample sizes (record length). If the historical maximum run characterisitic was close to the corresponding maximum median drought characteristic for the same sample size (return period) and truncation level, then it was assumed that the historical record accurately predicted the return period of the maximum drought characteristic. Table 8.2 compares the return periods of the maximum droughts from the unextended and extended records at stations 12413000 and 13186000 based on the experimental CDF's and the historical record length.

Table 8.2

Comparison of Return Periods of Maximum Droughts as Based on Experimental CDF's and the Historical Record Length

q(0)	Record Length years	Record L(max) years	CDF T years	q(0)	Record D(max)	CDF T years
			Station 13186000 - Unextended			
.50	38	6	>100	.50	4.24	48
.35	38	3	68	.35	1.90	<25
			Station 13186000 - Extended			
.50	72	4	44	.50	3.49	27
.35	72	4	>100	.35	2.17	27
			Station 12413000 - Unextended			
.50	44	3	<25	.50	3.71	32
.35	44	3	68	.35	2.55	43
			Station 12413000 - Extended			
.50	63	4	44	.50	4.55	61
.35	63	4	>100	.35	3.39	>100

From this comparision of the return periods of the historical droughts as defined by the record length and long-term stochastic processes, it was concluded that the return period of droughts should be assigned based on the long-term stochastic process. If the return period is assigned based on the historical record length, it may be low, reasonable, or high. The results were inconsistent and consequently, a design based on the historical record length as the return period may be under- or over-sized.

As previously mentioned, the population statistics estimated from the extended records at stations 12413000 and 13186000 are more reliable than those estimated from the corresponding unextended records, due to the larger sample sizes which produce smaller confidence intervals around the estimated statistics. Therefore, any parameters estimated from the extended sample statistics would be more reliable than those estimated from the unextended sample statistics. As a result, the maximum run characteristics developed from the extended records' estimates of ρ, γ, $y(0)$ and σ would be more reliable than those developed from the unextended records. Consequently, subsequent analyses concentrated only on the results of the extended models.

The statistics of all the annual and monthly flows which comprised the maximum drought sequence in each sample period were then reviewed. Based on this review, the following trends and results were observed for the maximum run-sum sequences:

1) The drought statistics are not greatly affected by the sample size, but both the mean and standard deviation are considerably affected by the truncation level (demand).

2) The drought distributions are a truncated version of the parent distribution (large flows deleted) with the addition of more low flow values.

3) The parent distributions with the largest coefficients of variation have the largest decreases in their corresponding drought statistics.

4) The coefficient of variation of the annual parent distribution tends to be preserved in the annual drought distributions.

5) The larger the skew coefficient of the parent distribution, the more the corresponding drought distribution resembles the parent distribution (in terms of over-all shape).

6) The mean annual and monthly drought flows provide a reasonable representation of the flows during drought periods.

7) The Weibull distribution provides a reasonable approximation of the annual drought streamflow distribution.

8) No one distribution appears to describe the distribtuions of the monthly drought flows.

9) The droughts as defined by the maximum run-sums are more severe than those defined by the maximum run-lengths for a given truncation level and sample size.

8.5 Storage Reservoirs and Low-Flow Regionalization

The median value of the maximum run-sum for a given truncation level and return period would be exceeded 50% of the time if many sample sizes corresponding to its return period were analyzed. The CDF's for the maximum run-sums can provide an estimate of the risk associated with a particular drought deficit. For example, instead of the median value, one corresponding to a 10% chance of exceedence for a particular truncation level and return period could be used for design considerations.

A storage reservior design should consider the periodicity in demand and/ or yield. This could be done by constructing a simple reservoir optimiztion model, using monthly flows from drought sequences corresponding to the selected maximum run-sum value (given truncation level, return period, and risk level) and monthly design demands. Each sequence could be run through the model and evaluated. The model would include the flow sequences prior to and following the drought period to test for reservior filling and refilling. These sequences would probably yield a range of reservoir storage requirements, although the range should be fairly small. The average or median value of required storage would then be selected as the design value.

The analyses of this study suggested several avenues which could be explored, depending upon the degree of sophistication desired, for developing regional low-flow characteristics of Idaho streams.

There seems to exist the possibility of regionalizing annual stochastic model parameters or similar annual flow characteristics, i.e., ρ and γ. In the past, annual flow series have frequently been modeled by AR(1) models (19, 32). If the annual streamflow series in Idaho can be modeled with AR(1) models (or AR(0) when $\mathbf{P} = 0$), then the CDF's of the maximum run-lengths and maximum standardized run-sums could be determined for each stream using the graphs developed by Millan and Yevjevich (27).

The CDF's of the maximum run-length and standardized run-sum could be used to estimate the relationships between the median maximum droughts and their return periods. These relationships would then further provide an estimate of the maximum run-length and maximum standardized run-sum for a selected truncation level, return period and risk level. Next, a regionalized estimate of the annual standard deviation σ, would be needed, to calculate the actual deficit S, corresponding to the standardized run-sum D, ($S = D\sigma$).

Therefore, by regionalizing ρ, σ, and γ the maximum deficit associated with a selected truncation level, return period, and risk level could be computed for any stream in Idaho.

If further detail is needed, several options may be possible:

OPTION 1: As mentioned in these conclusions, the sample size seems to have little affect, as compared to the truncation level, upon the statistics of the drought flows as defined by the maximum run-sums. Consequently, it may be possible to regionalize the percent of the mean annual historical flow represented by the mean annual drought flow for various truncation levels. Hence, by using an estimate of the mean annual historical flow at a stream, the mean annual drought flow which provides a reasonable estimate of flow during drought periods, could be determined.

Then, if the historical monthly flow statistics could be estimated, figures relating the percent reduction of each monthly mean flow corresponding to a given drought could be used to estimate the average monthly drought flows. These monthly drought flows would then be adjusted such that their sum equals the mean annual drought flow, and used to approximate the "typical" monthly flows during an average annual drought event.

OPTION 2: If sequences of annual drought flows were needed, and the percent reduction in the annual statistics (y, s and γ) corresponding to annual drought flows for a given truncation level could be regionalized, then the Weibull distribution along with ρ_d of the annual drought series could be used to construct an AR(1) model:

$$y_{d,t} = r_d \, y_{d,t-1} + s_d (1 - r_d^2)^{1/2} \lambda_w$$

where:

- y_d = annual drought flow
- r_d = lag-one serial correlation coefficient of annual drought series. Further research would be needed to develop this correlation, as it would probably be higher than for the entire annual series.
- s_d = standard deviation of drought flows
- λ_w = random standard deviate from Weibull distribution.

The parameters of the Weibull distribtion could be computed using the estimates of the annual drought statistics. Therefore, if the percentages associated with the annual drought statistics could be regionalized, and the corresponding annual statistics estimated, then actual drought flow sequences could be generated. The lengths of the sequences would be determined by the design maximum deficit, and when the sum of the annual deficits neared the design maximum deficit the sequence would be ended.

These annual sequences could then be disaggregated into monthly flows, if the parameters of the disaggregation models could be regionalized and estimates were avaiable for the statistics of the historical monthly streamflow values. Once again, further research would be needed to arrive at the disaggregation parameters corresponding to just the drought periods.

REFERENCES

1. Akaike, H., "A New Look at the Statistical Model Identification", Institute of Electrical and Electronic Engineers Transactions on Automatic Control, Vol. AC-19, No. 6, Dec. 1974, pp. 716-723.

2. Anderson, R. L., "Distribution of the Serial Correlation Coefficients", Annals of Mathematical Statistics, Vol. 8, No. 1, March 1941, pp 1-13.

3. Bhuiya, R. K., Yevjevich, V., "Effects of Truncation on Dependence in Hydrologic Time Series", Hydrology Papers, No. 31, Colorado State University, Fort Collins, Col. Nov. 1968.

4. DeLucia, A., The Compact Atlas of Idaho, University of Idaho, Moscow, Idaho, 1981, pp. 8-43.

5. Devore, Jay, "The Analysis of Categorical Data", Probability and Statistics for Engineering and the Sciences, Brooks/Cole Publishing Company, Monterey, Calif., 1982, pp.522-526.

6. Durbin, J., "The Fitting of Time Series Models", Review of International Institute Statastics., Vol 28, No. 233, 1960.

7. Feller, W., "The Asymptotic Distribution of the Range of Sums of Independent Random Variables", Annals of Mathematical Statistics, Vol. 22, 1951, pp 427-432.

8. Feller, W., An Introduction to Probability Theory and Its Applications, Vol. 1., John Wiley and Sons, Inc., New York, N. Y., 1957.

9. Fiering, M. B., "Multivariate Technique for Synthetic Hydrology", Journal of Hydraulics Division, American Society of Civil Engineers, Vol. 90, No. HY5, Sept. 1964, PP. 43-59.

10. Guerrero-Salazar, P., "Analysis of Drought Charactertistics by the Theory of Runs", Hydrology Papers, No. 80, Colorado State University, Fort Collins, Col., Sept 1975.

11. Hipel, K. W., McLeod, A. I., "Advances in Box-Jenkins Modeling. 1. Model Construction", Water Resources Research, Vol 13, No. 3, June 1977, pp 567-574.

12. Hipel, K. W., McLeod, A. I., "Preservation of the Rescaled Adjusted Range, 2. Simulation Studies Using Box-Jenkins Models", Water Resources Research, Vol 14, No. 3, June 1978, pp 509-515.

13. Horn, Dennis, "Analysis and Generation of Low-Flow Sequences in Idaho Streams Using Disaggregation Modeling", University of Idaho, Moscow, Idaho, 1984.

14. Hurst, H. E., "Long-term Storage in Reservoirs", Proceeeding of the Institute of Civil Engineers, No. 116, 1951, pp. 776-808.

15. Hurst, H. E., "Methods of using Long-term Storage in Reservoirs", Proceeeding of the Institute of Civil Engineers, Vol. 5, No., 5, Sept. 1956, pp. 519-590.

16. Idaho Water Resource Board, "Atlas, Idaho Water Resources Inventory", Maps: M1-M38, University of Idaho, Moscow, Idaho, 1968.

17. Idaho Water Resource Board, State of Idaho: The State Water Plan - Part II, Bosie, Idaho, December 1976, p.15.

18. Kahan, J. P., "A Method for Maintaining Cross and Serial Correlations and the Coefficient of Skewness under Generation in a Linear Bivariate Regression Model", Water Resources Research, Vol. 10, No. 6, Dec. 1974, pp. 1245-1248.

19. Lane, W. L., Applied Stochastic Techniques, Bureau of Reclamation, Denver, Colorado, 1983.

20. Lawrance, A. J., "A Reconsideration of the Fiering Two-Station Model", Journal of Hydrology, Vol. 29, No. 1/2, March 1976, pp. 77-85.

21. Mandelbrot, B. B., Wallis, J. R., "Computer Experiments with Fractional l Gaussian Noise. Parts I and II., Water Resources Research, Vol 5, No. 1, Feb. 1969, pp. 228-241, 242-259.

22. Matalas, N. C., "Mathematical Assessment of Synthetic Hydrology", Water Resources Research, Vol. 3, No. 4, 1967, pp. 937-946.

23. McLeod, A. I., "Improved Box-Jenkins Estimators", Report for Department of Statistics, University of Waterloo, Ontario, Canada, 1976.

24. Mejia M. M., Rodriguez-Iturbe, I., Dawdy, D. R., "Streamflow Simulation, 2. The Broken Line Process as a Potential Model for Hydrologic Simulation", Water Resources Research, Vol. 8, No. 4, Aug. 1972, pp. 931-941.

25. Mejia, M. M., Rodriguez-Iturbe, I., Cordova, J. R., "Multivariate Generation of Mixtures of Normal and Lognormal Variables", Water Resources Research, Vol. 10, No. 4, Aug. 1974, pp. 691-693.

26. Mejia, M. M., Rousselle, J., "Disaggregation Models in Hydrology Revisited", Water Resources Research, Vol. 12, No. 2, April 1976, pp. 185-186.

27. Millan, J., Yevjevich, V., "Probabilities of Observed Droughts", Hydrology Papers, No. 50, Colorado State University, Fort Collins, Col., June 1971.

28. Molnau, M., "Climate and Hydrology Data for Idaho", Agriculture Experiment Station, Miscellaneous Series No., 32, University of Idaho, Moscow, Idaho, 1983.

29. O'Connell, P. E., "Simple Stochastic Modeling of Hurst's Law", Proceeding on Mathematical Models in Hydrology from the Warsaw Symposium, Vol. 1., No. 14, July 1971, pp. 169-187.

30. Phien, Sunchindah, and Patnaik, "Normality of Hydrological Data", Water Resources Bulletin, Vol. 18, No. 1, Feb 1982, pp. 37-43.

31. Salas, J. D., et al, "Hurst Phenomenon as a Pre-assymptotic Behavior", Journal of Hydrology, Vol 44., No. 1/4, Nov. 1979, pp. 1-15.

32. Salas, et al, Applied Modeling of Hydrologic Time Series, Water Resources Publications, Littleton, Colorado, 1980.

33. Salas, J. D., Boes, D. C., Smith, R. A., "ARMA Modeling of Seasonal Hydrologic Series", Paper submitted for publication to the Journal of Hydraulics Division, ASCE, 1980.

34. Sen, Z., "Statistical Analysis of Hydrologic Critical Droughts", Journal of Hydraulics Division, ASCE, Vol. 106, No. HY1, Jan. 1980, pp. 99-114.

35. United States Department of Interior: Water Pollution Control Administration, Twin Springs Project, Boise, Idaho, 1966.

36. United States Geological Survey, Water Resource Data, Boise, Idaho, 1965-1983.

37. Valencia, D. R., Schaake, J. C., "Disaggregation Processes in Stochastic Hydrology", Water Resources Research, Vol. 9, No. 3, June 1973, pp. 580-585.

38. Wilson, E. B., and Hilferty, M. M., "The Distribution of chi-squared", Proceedings of the National Academy of Science, USA, Vol. 17, 1931, pp. 684-688.

39. Yevjevich, V., "Fluctuations of Wet and Dry Years", Hydrology Papers, No. 1, Colorado State University, Fort Collins, Col. July 1963.

40. Yevjevich, V., "The Application of Surplus, Deficit and Range in Hydrology", Hydrology Papers, No. 10, Colorado State University, Fort Collins, Col. Sept. 1965.

41. Yevjevich, V., "An Objective Approach to Definitions and Investigations of Continental Hydrologic Drougths", Hydrology Papers, No. 23, Colorado State University, Fort Collins, Col. Aug. 1967.

42. Yevjevich, V., "Properties of Non-homogeneous Hydrologic Series", Hydrology Papers, No. 32, Colorado State University, Fort Collins, Col. April 1969.

43. Yevjevich, V., "Generation of Hydrologic Samples, Case Study of the Great Lakes", Hydrology Papers, No. 72, Colorado State University, Fort Collins, Col. May 1975.

APENDIX A

SELECTION OF STREAMFLOW RECORDS

Twenty-one stations were found to meet the first three selection criteria as stated in section 1.1, and these stations are listed in Table A.1.

Table A.1

Streamflow Stations With at Least 30 Years Record, Little Diverion and/or Regulation, and of at Least "fair" Quality.

Station Number	Station Name	Area sq mi	Record Length	* Remark
12306500	Moyie River, Eastport	570	1929-83	G-F
12307500	Moyie River at Eileen	755	1925-77	E-F
12411000	Coeur d'Alene R. above Shoshone Crk, Prichard	335	1950-83	G
12413000	Coeur d'Alene, Enaville	895	1939-83	G
12413500	Coeur d'Alene, Cataldo	1220	1920-72	G-F
12414500	St. Joe River. at Calder	1030	1920-83	G-F
13336500	Selway River at Lowell	1910	1929-83	G, SD
13337000	Lochsa R. near Lowell	1180	1929-83	G
13317000	Salmon R at White Bird	13550	1910-81	E
13235000	S. F. Payette at Lowman	456	1941-83	G, SD
13261000	Little Weiser below Mill Creek near Indian Valley	82	1938-71	G-F
13185000	Boise R. at Twin Springs	830	1911-83	G
13186000	S.F. Boise, Featherville	635	1945-83	G, SD
13200000	Mores Creek above Robie Creek near Arrowrock Dam	399	1950-83	G-F SD

Table A.1 (Continued)

Streamflow Stations With at Least 30 Years Record, Little Diverion and/or Regulation, and of at Least "fair" Quality.

Station Number	Station Name	Area sq mi	Record Length	* Remark
13092000	Rock Crk near Rock Crk	80	1943-74	G-F
10041000	Thomas Fork, ID-WY brd	113	1949-83	G-F
13011500	Pacific Crk, Moran, WY	169	1944-75 1978-83	F
13011900	Buffalo Fork above Lava Creek near Moran, WY	323	1944-60 1965-83	F
13120000	N. F. Big Lost River at Wildhorse near Chilly	114	1944-83	G,SD
13023000	Greys River abv Reserv. near Alpine, WY	448	1953-83	F,SD
10093000	Cub River near Preston	31.6	1940-52 1955-83	F

* E = Excellent; G = Good; F = Fair; P = Poor; SD = small diversion

Table A.2 identifies the best secondary station for the extension of each record in Table A.1. Also, the number of years that the original record could be extended and the number of years the two records temporally overlap are shown. When a record could not be extended by a period of at least 15 years or did not overlap the second station by at least 20 years, the station was no longer considered. The stations no longer considered are marked in Table A.2 with an "*" by the condition that prevented their use.

Table A.2

Streamflow Records that could be used for Data Extension

Unextended Station		Nearby Stations		Extended Rec.	
Station Number	Record Length	Station Number	Record Length	Years Extend	Years Overlap
12306500	1929-83	12307500	1925-77	* 4	49
12307500	1925-77	12306500	1929-83	* 6	49
12411000	1950-83	12413500	1920-72	30	23
12411000	1950-83	12414500	1920-83	30	34
12413000	1939-83	12413500	1920-72	19	34
12413000	1939-83	12414500	1920-83	19	45
12413500	1920-72	12413000	1939-83	* 11	34
12414500	1920-83	None		* 0	* 0
13336500	1929-83	13331700	1910-81	19	52
13337000	1929-83	13331700	1910-81	19	52
13317000	1910-81	None		* 0	* 0
13235000	1941-83	13185000	1911-83	30	43
13261000	1938-71	13235000	1941-83	* 12	33
13185000	1911-83	None		* 0	* 0
13186000	1945-83	13185000	1911-83	34	39
13200000	1950-83	13185000	1911-83	39	34
13092000	1943-74	None		* 0	* 0
10041000	1949-83	10093000	1940-52 1955-83	* 9	32
13011500	1944-75 1978-83	None		* 0	* 0
13011900	1944-60 1965-83	None		* 0	* 0

Table A.2 (Continued)

Streamflow Records that could be used for Data Extension

Unextended Station		Nearby Stations		Extended Rec.	
Station Number	Record Length	Station Number	Record Length	Years Extend	Years Overlap
13120000	1944-83	None		* 0	* 0
13023000	1953-83	13011500	1944-75 1978-83	* 9	38
10093000	1940-52 1955-83	None		* 0	* 0

* Station no longer considered because the record could not be extended by a least 15 years or the records did not temporally overlap by at least 20 years.

In Table A.3 the stations that could be extended (Table A.2) are compared in terms of drainage area and record length. The comparisions are made between stations that are as far apart geographically as possible. Station combinations which did not have drainage areas within 100% or record lengths within 50% of each other are marked with an "*". Station combinations marked with an "*" were no longer considered for use in this study.

Table A.3

Comparison of Drainage Area and Record Length of Stations which could be Extended

Station Number	Record Length	Drain Area	Station Number	Record Length	Drain Area	% Difference Length	% Difference Area
12341100	34	335	13235000	43	456	26	36
12341100	34	335	13186000	39	635	15	89
12341100	34	335	13200000	34	399	0	19
12413000	45	895	13235000	43	456	5	96
12413000	45	895	13186000	39	635	15	41
12413000	45	895	13200000	34	399	32	* 124
13337000	55	1180	13235000	43	456	28	* 159
13337000	55	1180	13186000	39	635	41	86
13337000	55	1180	13200000	34	399	* 62	* 196
13336500	55	1910	13235000	43	456	28	* 319
13336500	55	1910	13186000	39	635	41	* 201
13336500	55	1910	13200000	34	399	* 62	* 379

* Stations no longer considered

Because of geogrpahical distance, station 13337000 was no longer considered since other combinations existed which were further apart. Also, station 13235000 on the South Fork of the Payette River was eliminated based on the higher probability that seems to exist for flow regulation on the Boise River than on the South Fork of the Payette River. This higher probability was partly assumed based on the fact that a study has already been made to assess the possiblity of constructing a storage reservoir near Twin Springs on the Boise River (35), and the larger population density near the Boise River Basin.

A closer examination was then made of the remaining streamflow records. Table A.4 gives a more detailed description of each record still under consideration.

Table A.4

Detailed Look at Candidate Records

Station Number	Area sq mi	Record Length	Years Extend	Remarks
12411000	335	34	30	Records good. No regulation or diversion above station.
12413000	895	45	19	Records good. No appreciable regulation or diversion above station.
13186000	635	39	34	Records good. No regulation. Diversion above station for irrigation of about 450 acres
13200000	399	34	39	Records good except winter - fair. Small diversion above station for irrigation.

It was finally decided to select station 13186000 over station 1320000 because: 1) a better estimation of the amount of diverted flow could be made, and 2) the record at station 13186000 was of better quality. From Table A.3, it can also be seen that the drainage area of station 12413000 was closest in size to the drainage area at station 13186000. Therefore, stations 13186000 and 12413000 were chosen for use in this study.

Now, the records to be used for data extension had to be selected. By returning to Table A.2, the possible records for data extension could be found. As can be seen in Table A.2, two stations extisted which could be used to extend the record at station 12413000. Station 12414500 overlapped temporally by the largest number of years, yet it was felt Station 12413500 was hydrologically more similar to Station 12413000. Based on the advantages of both records, it was decided to use Station 12413500 for data extension, reasoning that the greater hydrologic similarity would compensate for the loss of eleven years of data. Meanwhile, only station 13185000 was available for data extension at station 13186000 and was therefore used.

APPENDIX B

HISTORICAL STREAMFLOW LISTINGS AND STATISTICS

Table B.1

Coeur d'Alene River near Enaville, Idaho
Station 12413000

Year	Oct	Nov	Dec	Jan	Feb	March	April	May	June	July	Aug	Sept	Annual
39-40	6862	6851	27153	15544	42271	146330	140330	71680	19072	9147	5928	6150	497318
40-41	7250	9809	23859	22771	22480	67182	57720	59190	33539	12835	8201	10576	335412
41-42	12671	26985	105789	27296	21693	41982	145630	73110	50230	26639	11019	7523	550567
42-43	7821	42217	29603	25892	19294	60174	296520	149500	71270	28450	12314	7880.	750935
43-44	9609	8492	23915	10041	12936	19608	94984	58910	22081	9706	6523	7624	284429
44-45	5828	8688	9896	45415	53211	56210	105980	195940	41066	14851	8142	10035	555262
45-46	10734	33361	59836	53611	22910	96580	243580	212670	55330	17746	8361	7522	822241
46-47	12893	58693	156680	63925	89550	109700	149760	105950	39627	16651	10087	9282	822798
47-48	34333	38640	40214	56917	37945	48590	197070	299160	91570	31119	18854	10960	905372
48-49	10298	16465	14797	9936	32705	79850	259200	258050	40889	15486	9433	8647	755756
49-50	11371	35797	34470	50659	56580	140610	198550	288830	136650	36354	14399	9940	1014210
50-51	24862	45475	115410	55010	140400	43775	173410	148340	38647	16106	9167	8187	818789
51-52	37498	47791	60349	17996	30699	32071	253170	160380	38640	18733	9936	7741	715004
52-53	6340	5920	6536	82672	104900	54158	144060	162680	67630	19699	10315	7322	672232
53-54	7202	11847	42189	38728	61680	81260	201960	256960	81590	28346	13436	11550	836748
54-55	12727	27468	20903	15830	28569	17764	101120	242320	103410	34583	13219	9981	627894
55-56	29051	68420	129260	63550	28316	61514	289540	245960	59930	24958	12904	8418	1021821
56-57	12399	11969	53632	17994	31555	85480	191740	275900	49648	17249	10012	7266	764844
57-58	9124	10462	17249	22457	85786	71480	176600	154890	30361	14220	7731	7551	607911
58-59	10328	70062	72320	113440	37851	48740	209230	203820	85610	20953	11370	12624	896348
59-60	27300	84785	62420	25838	47107	103417	204720	172250	67812	18023	11828	8657	834157
60-61	9935	31135	17261	35930	190800	105480	141160	206760	64746	16979	8671	8174	837031
61-62	9965	7539	15316	28412	48824	41928	225760	158610	51078	15529	9392	7837	620190
62-63	12268	34949	65560	41180	96760	69340	111490	87780	29757	13020	7925	6834	576863
63-64	7728	22658	14716	17909	16843	24038	144320	271210	147960	28963	15693	12779	724817
64-65	16230	35895	158750	53370	64480	76900	207620	154150	58080	19181	12104	9837	866597
65-66	8237	12327	15179	27322	12626	79581	169630	121320	36089	15045	8850	6748	512954
66-67	6862	19295	67079	69344	66220	66539	101020	200180	75276	18004	8825	6569	705213
67-68	13114	22811	23052	37821	138070	118400	75750	83610	34660	14574	10079	15786	587727
68-69	36649	71690	66990	70800	24697	55362	242990	190870	51810	18701	9830	8094	848483
69-70	9075	7997	13397	36619	47890	65210	106520	220560	72480	20135	10648	8577	619108
70-71	9546	17309	24531	74031	111370	56420	182010	248730	77750	38045	13633	10079	863454
71-72	10775	13241	14992	53710	80900	248760	147250	259130	90540	28929	13722	10059	972208
72-73	8880	9338	33228	57463	19174	51530	68950	66760	23545	11025	6895	6755	363543
73-74	7327	34332.	95300	214790	53650	87850	243640	247660	161080	30833	12831	8404	1197697
74-75	7389	14170	14276	21334	20948	48672	90710	264630	113900	25825	13116	8364	643334
75-76	10595	23490	112490	65330	41020	39690	173880	207820	55020	21039	14799	8693	773866
76-77	7710	8136	7774	6798	10198	20421	72236	42794	19600	9557	6242	7291	218757
77-78	7201	22025	117860	40015	34268	123450	143540	141430	52197	19591	14421	11317	727315
78-79	7941	8261	7224	6481	23601	95358	134930	193930	35043	14532	7882	6460	541643
79-80	7428	6168	26854	28064	35274	60590	137270	96540	72210	26534	12700	11268	520900
80-81	7953	19528	146603	67300	94730	58980	105720	90240	88900	32000	13571	9284	734809
81-82	9681	15503	40324	23833	196675	138570	144590	170150	59890	22366	10965	8566	841113
82-83	10692	16574	40398	88650	82180	148290	113770	111030	41441	34710	15856	10307	713898

Table B.2

Coeur d'Alene River near Cataldo, Idaho
Station 12413500

Year	Oct	Nov	Dec	Jan	Feb	March	April	May	June	July	Aug	Sept	Annual
19-20											13751	18757	
20-21	33290	46160	60620	155390	124400	203290	246030	271460	68890	25165	14574	11688	1260957
21-22	12698	14626	94870	19700	13875	25730	147560	286020	83610	22570	13454	10865	745578
22-23	10878	11590	21270	94560	17250	61875	273580	239440	122590	34889	18185	11946	918053
23-24	14252	16547	26180	17083	138360	73490	149780	172600	32352	16428	11105	9160	677337
24-25	11130	30524	55980	44223	189630	150170	338060	208860	64450	25888	15947	11350	1146212
25-26	10358	10888	28089	19565	53340	101850	151630	67510	23027	12528	9748	17026	505559
26-27	32686	56624	103740	52570	73420	83320	239200	296810	153020	37628	16797	25182	1170997
27-28	61512	195870	115360	88690	51360	152250	174010	258870	51350	22795	13072	10173	1195312
28-29	13907	16888	11464	9468	7722	54660	136630	172460	47220	19113	11087	10033	510652
29-30	9816	7138	15829	7462	42760	69320	143600	79610	45550	18619	10207	8355	458266
30-31	10361	12840	8550	15354	29242	106117	196950	129960	27582	12881	8549	9531	567917
31-32	9969	11831	10931	23632	57411	150360	362580	338010	102000	28087	14335	11412	1120558
32-33	13224	73780	72722	53228	21293	70429	264600	282090	203070	39575	17706	14542	1126259
33-34	27005	63918	410219	258010	119410	201190	190090	75760	29131	15981	9927	8944	1409585
34-35	19706	57214	57552	68630	64310	93130	237450	285770	81630	25053	14312	9901	1014658
35-36	10039	10137	9836	17045	11099	62800	321900	195190	52252	20341	10767	9632	731038
36-37	8827	7590	16165	8353	7747	43186	193110	250780	69800	25343	15008	10862	656771
37-38	8704	45944	78780	84350	35780	116530	322020	181220	56770	22789	12611	9701	975199
38-39	11548	12658	20661	25208	16326	86730	236480	153520	42580	20191	10247	8852	645001
39-40	9814	9697	33892	22905	57818	174340	179990	106200	28632	13230	8470	8626	653614
40-41	10772	16055	36528	33385	30714	80660	74680	83160	47140	18215	11344	15996	458649
41-42	20972	38274	131190	38760	35490	62168	186130	105390	73560	39159	15779	10958	757830
42-43	10175	51417	42590	36398	28265	83032	377030	202250	107310	42004	18060	11640	1010171
43-44	11899	13176	29325	13452	17495	25960	117150	88750	31011	14045	9258	10213	381734
44-45	8548	11619	12046	61879	70569	71028	130150	258140	60590	20604	11648	14367	731188
45-46	14770	44280	80289	79800	30012	125000	292510	275290	77090	26460	12323	10752	1068576
46-47	18678	87866	201260	78480	110510	135900	198700	161150	59170	22866	13154	12681	1100415
47-48	44697	50470	51770	79768	48624	61330	251910	395280	132360	46880	27830	15058	1205977
48-49	15218	23470	20960	14050	44190	110890	314820	325410	62650	22672	13915	12112	980357
49-50	15166	48485	52292	71680	89550	185620	251870	359530	200610	59100	22710	14538	1371151
50-51	32719	62600	145000	80030	186250	71490	218920	203650	65880	27481	14248	11465	1119733
51-52	44811	57480	75630	26753	49514	47121	310710	214060	60500	29503	14603	11628	942313
52-53	9918	9297	10572	104966	123050	67660	175370	217540	104800	31701	15616	10800	881290
53-54	10442	15536	57034	53360	88660	105470	247240	336890	116880	43906	22787	18085	1116290
54-55	20387	43242	29620	22296	38464	25113	125790	300820	141640	48430	20578	15634	832014
55-56	39628	90180	172320	84650	35626	85315	373020	327260	91030	34921	18782	12858	1365590
56-57	19820	19182	74601	25786	43262	113660	244260	370160	71990	28671	17116	12191	1040699
57-58	13502	17636	31465	39298	115670	85870	210770	214520	46350	20738	11752	10790	818361
58-59	13481	82760	89160	143600	53640	67450	252910	261720	126940	34952	17179	18271	1162063
59-60	40100	100860	73820	36496	63520	131870	253230	213220	101010	28970	16613	12108	1071817
60-61	13607	38910	24814	49936	247390	135990	184400	279080	96440	25036	12872	11379	1119854
61-62	12998	11176	20816	39860	65270	58687	301620	220400	75900	23970	14162	11498	856357
62-63	18101	46775	84150	51110	132730	92070	151010	126860	44230	20648	11736	10038	789458
63-64	10963	26524	19528	23612	23547	34739	182950	324650	192720	39913	20606	16973	916725
64-65	20931	43341	208940	73900	89070	97620	270140	205940	86680	30805	18407	14139	1159913
65-66	11621	16456	19099	37309	17138	107120	224240	169800	56780	21632	12146	9849	703190

Table B.2 (continued)

66-67	9967	23759	78960	93510	86740	90660	125730	264000	114600	28493	12450	9668	938537
67-68	18937	31306	30801	50337	173440	141920	94380	116890	57810	23953	15742	22988	778504
68-69	47939	92610	87010	89810	34310	80310	313580	255500	73560	27202	13964	11335	1127130
69-70	13110	12093	20230	50412	63790	80000	134670	283300	103620	28573	15426	12387	817611
70-71	14455	26273	36471	93755	138700	70460	219480	336950	117670	53100	20775	15357	1143446
71-72	16449	17883	20941	71472	104860	320510	183580	355390	139480	42651	17886	14030	1305132

Table B.3
South Fork of Boise River near Featherville, Idaho
Station 13186000

Year	Oct	Nov	Dec	Jan	Feb	March	April	May	June	July	Aug	Sept	Annual
44-45								60180	57220	21252	7253	5613	
45-46	6147	6637	7466	7270	5932	12339	66360	93530	60880	19320	8187	6349	300417
46-47	8617	7984	8619	6696	7169	16142	39275	82440	44352	16234	6798	5591	249917
47-48	7821	6900	6165	6900	6005	6655	25231	71600	73130	17009	7506	5990	240912
48-49	6889	6857	6534	6460	5595	9719	44839	87770	48081	13976	6341	5209	248270
49-50	6612	7471	6302	6996	6512	9181	41900	86520	95660	39065	11372	9030	326621
50-51	9610	12074	9577	8225	9390	11140	70528	116330	76140	32891	12515	7609	376029
51-52	9763	8408	9184	8640	7538	8834	63016	136620	88510	26639	10547	6845	384544
52-53	6877	6334	7112	8260	6793	12015	40689	55120	90710	39418	10123	6558	290009
53-54	6340	7037	6414	7155	7607	11664	43841	97390	53550	25612	8604	6248	281462
54-55	6565	6430	5712	5966	4961	5952	12275	51390	69320	18841	6583	5158	199153
55-56	6398	7853	13807	11005	7816	14373	75787	135720	98740	27556	10198	6873	416126
56-57	8143	7487	7706	6758	7311	11347	31838	107260	89720	23376	8320	6398	315664
57-58	7557	6594	7200	7097	7531	8946	26704	151120	84620	21783	9742	6969	345863
58-59	7112	8049	9386	7930	6514	8939	36283	49500	59869	14605	6997	8511	223695
59-60	10779	7865	6290	6832	5856	12226	38084	54198	51091	11175	6222	5691	216309
60-61	5987	6352	5538	5841	6578	8874	20844	45571	33173	7351	4844	5236	156189
61-62	6134	6180	5976	6062	6561	7213	49829	65620	81030	24514	9797	6602	275518
62-63	8367	7550	7358	6271	12206	10334	19655	75510	68800	22771	8686	7304	254812
63-64	7068	8173	6907	6647	6156	7337	27571	66740	68010	22869	7859	6185	241522
64-65	6231	6456	21153	13783	11706	14221	64722	125140	144040	60299	19937	11894	499582
65-66	9666	8713	7346	7476	6164	10275	34004	56974	28997	9665	5195	4481	188956
66-67	5280	5722	5469	6486	5748	8899	16301	82356	99770	30745	9611	6904	283291
67-68	8011	7200	6749	6525	8111	13029	20703	40959	47517	12820	10360	7454	189438
68-69	8526	9392	7968	10067	8052	12518	77830	134890	72410	20194	8419	6911	377177
69-70	7380	6185	6378	7794	7438	12273	20774	80190	91380	26063	9272	7385	282512
70-71	7388	10045	8682	10568	11067	12700	49021	138680	125660	49512	14383	9766	447472
71-72	9606	8729	8435	8812	7920	23308	35440	97800	114530	31047	11607	8416	365650
72-73	8982	8096	7752	7390	6380	9142	22054	54102	34515	12380	5806	5664	182263
73-74	6157	10793	8469	11796	8022	20563	65384	109660	124960	35464	11542	7344	420154
74-75	7771	7403	6775	7505	7625	13420	14672	80359	117880	60481	13981	8409	346281
75-76	9053	9161	9726	8471	7942	9304	38014	96660	55190	18562	11368	9816	283267
76-77	8510	6783	6191	5967	5671	6301	10362	13031	14776	6673	4293	4254	92812
77-78	4978	5368	7894	6310	6059	17748	42272	78560	86560	37725	11308	9522	314304
78-79	7459	6746	6608	6419	6342	9935	16358	56431	32902	9751	6684	5215	170850
79-80	6453	5793	5931	6992	6895	9460	45957	87720	69590	31045	10214	8741	294791
80-81	7526	7342	8412	7548	7820	12034	29344	57940	48648	12936	6873	5348	211771
81-82	6825	8708	9986	8202	9918	14541	36421	132720	132650	54783	13650	9490	437894
82-83	10075	9034	8717	9564	8383	18112	40781	130740	139750	57530	17046	10509	460241

Table B.4
Boise River near Twin Springs, Idaho
Station 13185000

Year	Oct	Nov	Dec	Jan	Feb	March	April	May	June	July	Aug	Sept	Annual
10-11							53900	111940	160910	53870	15428	11705	0
11-12	12335	11863	12339	15718	12223	16418	56300	136280	162910	46923	20465	14991	518765
12-13	13944	14307	10850	10926	11247	17326	72110	132410	104650	41356	20535	13161	462822
13-14	14885	18011	12824	13894	13898	37884	88430	127760	81760	30961	12823	12673	465803
14-15	16990	12886	8592	9495	10207	17635	43745	62050	47245	22756	11041	10064	272706
15-16	10121	9706	11478	9589	9220	31843	92350	119820	148040	82250	21423	13605	559445
16-17	13846	9820	9518	8680	8751	13037	41201	135060	156210	66615	15735	10501	488974
17-18	10094	9875	27117	20579	13992	32899	77570	102000	137790	28350	14020	11126	485412
18-19	12815	10985	9505	8230	8281	19048	73540	124120	59919	17455	9532	8244	361674
19-20	10273	12275	9010	13495	8197	13252	32503	120710	97460	31206	13151	12578	374110
20-21	15303	17497	16366	17968	17635	40776	70560	190120	176380	43382	16518	13419	635924
21-22	12456	16222	18146	11602	10545	18196	46469	141210	155470	35827	14827	10137	491107
22-23	10105	10491	11690	11584	10280	16881	47190	105670	89610	47487	14472	10300	385760
23-24	12014	10841	10739	10200	13324	12200	30367	64260	21684	9942	7330	7184	210085
24-25	9445	12552	9620	10082	20898	31610	92060	161230	90600	39025	14791	11459	503372
25-26	12393	10899	12411	11816	11258	23679	49833	60540	27132	11306	8119	7722	247108
26-27	8383	15257	18705	12750	16319	27246	64130	130640	170070	58763	17505	13683	553451
27-28	15007	32959	22633	19334	16314	42336	56330	202350	78150	28923	13234	10185	537755
28-29	11166	10428	10006	9300	8725	18296	31067	82360	75400	24312	10195	8555	299810
29-30	9218	8199	15290	9102	14904	19797	62210	68430	61400	18164	10599	8976	306289
30-31	12653	9470	8290	9290	9113	17677	40728	66230	31421	10162	6985	6993	229012
31-32	8265	8139	8343	9057	8382	21351	57087	133450	109700	34247	12765	9427	420213
32-33	9851	10673	8463	9920	8760	13229	41570	74580	137710	26358	10966	8647	362727
33-34	9669	9964	11602	14716	15701	36332	64910	56690	23979	10233	6931	6702	267429
34-35	9361	12090	11201	11486	12239	15962	58646	99810	91370	24247	10294	7752	364458
35-36	8461	8610	8224	9062	8600	16301	105423	142390	76200	19768	10656	8935	422630
36-37	8635	7878	9044	8400	8437	15394	34790	90120	46093	14814	8028	7066	258699
37-38	8566	10416	25821	14501	12974	27301	80049	141840	129780	46371	15768	10819	524206
38-39	12639	12782	13044	11089	10084	29258	63140	80680	33840	15269	8117	7744	297686
39-40	9513	8215	11352	11576	14720	39527	73270	107900	58835	16521	8625	9668	369722
40-41	11999	12975	11881	11180	12295	20522	37050	83950	66740	21123	13157	10904	313776
41-42	11036	16086	25457	13872	12478	16614	66820	73560	87810	35067	12219	9050	380069
42-43	9216	14219	15803	24709	19900	36125	169750	147900	141100	92210	25372	13752	710056
43-44	14641	15964	12242	11274	10637	13554	33388	68680	58320	25479	11067	8872	284118
44-45	9211	12133	9882	11700	13303	18091	39956	112180	108470	40856	14249	10547	400578
45-46	10687	12428	16467	16357	13292	35573	110240	138410	92090	32914	13714	11381	503553
46-47	15654	16690	25498	14530	17658	34990	66050	141970	82030	32897	13352	10808	472127
47-48	14241	12773	12695	16932	12139	14003	54510	126340	114640	29852	13233	10451	431809
48-49	11236	12047	11527	10594	8895	23341	80989	142130	79730	24436	11572	9200	425697
49-50	11182	12918	12276	14461	15328	25353	72290	121690	133510	65149	19362	13832	517351
50-51	15884	19713	18479	14386	20890	22115	100980	149680	106650	51249	18965	11991	550982
51-52	18874	15348	19129	14067	14611	17436	110660	174740	119090	41234	17198	11801	574188
52-53	10578	9406	12079	17664	15371	22115	64143	95110	151490	69536	17873	11143	496508
53-54	10436	12162	11611	12668	17428	27823	84532	148990	95610	51157	16221	11332	499970
54-55	11120	10899	10077	10148	8479	10681	23562	101440	117060	38127	13680	9676	364949
55-56	11530	16663	46805	30819	16550	35189	115860	190100	141610	48451	17711	12065	683353
56-57	13842	13240	15576	11575	17176	28427	67560	173390	125640	37201	14391	10908	528926

Table B.4 (continued)

57-58	12416	11000	12840	11956	19346	19605	49944	208850	122070	34329	15577	11296	529229
58-59	10914	15281	18897	16009	13060	17972	60071	81010	102070	28895	12960	15032	392171
59-60	21663	15079	12249	12238	11178	31800	67910	87750	90490	21471	11982	9841	393651
60-61	9969	10825	10174	9474	12430	18546	39840	83390	65316	14705	9814	9880	294363
61-62	11510	11439	11744	12540	15335	15256	86710	94740	105900	36235	15319	10457	427185
62-63	16922	15210	17514	13044	32769	20693	42368	110930	99740	37934	14528	12286	433938
63-64	11976	14304	11500	11937	11149	14014	48740	108780	105000	38462	14212	11519	401593
64-65	10493	11358	54203	27873	25360	28997	108410	153130	176270	82910	27664	17512	724180
65-66	14839	13059	10804	11918	9236	21900	51320	78390	45903	15148	8696	7628	288841
66-67	8566	9715	10070	12119	10948	16373	25889	112688	127690	44120	13567	10136	401881
67-68	12901	12078	8842	9478	22016	29436	40110	69070	74530	21882	17688	12503	330534
68-69	14885	20475	15135	26288	16012	27936	112730	152920	90770	31478	13877	11112	533618
69-70	11970	10410	11083	20679	18490	26315	34094	129441	141430	52806	16901	12807	486426
70-71	12436	22269	19990	28535	27506	31762	90060	190920	163100	74870	21945	14692	698085
71-72	14741	13642	13423	15396	16847	66920	67950	169630	172150	51374	19243	14085	635401
72-73	13920	12735	13574	14168	11611	18962	37954	89740	63330	21328	10880	10363	318565
73-74	10210	24835	19145	30571	16984	42050	107550	163210	204110	70060	22851	13112	724688
74-75	12736	12406	11458	12722	11707	22112	29715	115000	155790	87935	20317	12456	504354
75-76	14274	14914	20308	14622	15128	19934	66801	148540	87400	34679	18400	15469	470469
76-77	13074	10558	10064	9201	8542	10113	21500	24240	27683	10884	7755	7894	161508
77-78	8513	10862	26489	14350	14573	48084	75860	108890	119760	63758	17936	14533	523608
78-79	11608	10255	10392	9585	11028	21497	32072	91990	55910	16625	10758	8479	290199
79-80	10389	9079	10361	15350	17446	21832	80597	127120	95800	46172	15701	14003	463850
80-81	12094	13824	20937	16412	20298	24987	49425	85470	75550	22590	11347	9046	361980
81-82	10766	15638	21517	14070	28822	34046	62190	167680	156980	82240	25420	16169	635538
82-83	18215	16950	17637	22281	17301	50380	64930	158090	159380	74150	26454	15763	641581

Table B.5

Statistics for Station 12413000
(Streamflow in cfs days)

	Mean	Variance	Stand Dev	Coeff Var	Skew	Coeff Skew	n
Oct	12493	$6.695X10^{7}$	8182	0.655	$1.152X10^{12}$	2.103	44
Nov	26013	$4.021X10^{8}$	20052	0.771	$1.064X10^{13}$	1.320	44
Dec	51037	$1.915X10^{9}$	43759	0.857	$9.636X10^{13}$	1.150	44
Jan	45501	$1.272X10^{9}$	35669	0.784	$1.212X10^{14}$	2.671	44
Feb	57264	$2.002X10^{9}$	44747	0.781	$1.415X10^{14}$	1.580	44
Mar	76087	$1.880X10^{9}$	43353	0.570	$1.364X10^{14}$	1.674	44
Apr	161810	$3.690X10^{9}$	60740	0.375	$8.175X10^{13}$	0.365	44
May	173468	$5.280X10^{9}$	72660	0.419	$-.416X10^{14}$	-.108	44
Jun	62219	$1.085X10^{9}$	32945	0.529	$4.409X10^{13}$	1.233	44
Jul	21068	$5.974X10^{7}$	7729	0.367	$2.456X10^{11}$	0.532	44
Aug	10951	$8.414X10^{6}$	2901	0.265	$9.617X10^{9}$	0.394	44
Sep	8944	$3.837X10^{6}$	1959	0.219	$9.276X10^{9}$	1.234	44
Yr	706854	$3.959X10^{10}$	198977	0.281	$-.215X10^{16}$	-.273	44

Serial Correlation coefficients

	Lag (k)								
Period	1	2	3	4	5	6	7	8	9
Oct	.497	.094	.110	.120	.064	-.070	.036	.307	.094
Nov	.698	.388	-.083	-.008	.051	-.011	-.085	-.112	.047
Dec	.527	.302	.227	-.018	.089	.214	-.024	-.254	-.176
Jan	.390	.293	.017	-.096	-.298	-.179	-.159	-.169	-.289
Feb	.160	.078	-.006	-.053	-.102	.071	.172	.207	.080
March	.365	.101	-.025	-.047	-.145	.032	.031	.152	-.006
April	.003	-.145	.232	.260	.545	.459	.251	.065	.070
May	.460	.094	-.019	.197	-.112	.170	.170	.089	.103
June	.668	.204	.028	-.020	.448	.026	.157	-.070	-.100
July	.753	.522	.225	.109	.068	.357	.104	.126	.043
Aug	.852	.678	.537	.239	.115	.003	.301	.231	.213
Sept	.547	.385	.343	.167	-.100	.121	.150	.126	.023
Year	-.056	.204	.040	.160	-.119	-.256	-.024	-.215	.001

Table B.6

Statistics for Station 12413500
(Streamflow in cfs days)

	Mean	Variance	Stand Dev	Coeff Var	Skew	Coeff Skew	n
Oct	18548	$1.446X10^8$	12027	0.648	$3.190X10^{12}$	1.834	52
Nov	38143	$1.151X10^9$	33927	0.889	$8.976X10^{13}$	2.299	52
Dec	63884	$4.695X10^9$	68521	1.073	$9.666X10^{14}$	3.005	52
Jan	56449	$1.953X10^9$	44193	0.783	$1.877X10^{14}$	2.174	52
Feb	69473	$2.836X10^9$	53258	0.767	$1.941X10^{14}$	1.285	52
Mar	98720	$2.801X10^9$	52926	0.536	$2.566X10^{14}$	1.731	52
Apr	221119	$5.530X10^9$	74364	0.336	$1.025X10^{14}$	0.249	52
May	228944	$7.459X10^9$	86366	0.377	$-.115X10^{15}$	-.179	52
Jun	83144	$1.904X10^9$	43639	0.525	$8.812X10^{13}$	1.060	52
Jul	28199	$1.108X10^8$	10526	0.373	$1.063X10^{12}$	0.911	52
Aug	14742	$1.565X10^7$	3957	0.268	$5.517X10^{10}$	0.891	53
Sep	12610	$1.218X10^7$	3490	0.277	$6.875X10^{10}$	1.617	53
Yr	933877	$6.657X10^{10}$	258015	0.276	$-.449X10^{16}$	-.262	52

Serial Correlation coefficients

Period	Lag (k) 1	2	3	4	5	6	7	8	9
Oct	.762	.247	.295	.358	.089	-.137	.064	.242	.183
Nov	.786	.671	.173	.246	.348	.140	-.056	.041	.090
Dec	.500	.408	.352	.207	.288	.539	.190	-.015	-.039
Jan	.675	.404	.310	.312	.061	.160	.369	.147	-.098
Feb	.314	.175	.058	.050	.017	.141	.199	.308	.178
March	.390	.381	.215	.164	.105	.200	.096	.201	.208
April	.094	-.135	.104	.106	.216	.191	.134	.125	.074
May	.370	.080	.005	.080	-.150	.142	.202	.223	.183
June	.702	.192	-.012	-.037	.114	-.090	.133	.015	.087
July	.858	.682	.267	.009	.032	.155	-.018	.165	.143
Aug	.909	.785	.719	.293	-.037	.047	.144	.006	.180
Sept	.495	.459	.493	.279	-.098	-.023	.129	.037	-.031
Year	.178	.080	-.052	.062	-.004	-.037	.012	-.065	-.094

Table B.7

Statistics for Station 13186000
(Streamflow in cfs days)

	Mean	Variance	Stand Dev	Coeff Var	Skew	Coeff Skew	n
Oct	7596	$1.969X10^6$	1403	0.185	$1.053X10^9$	0.381	38
Nov	7629	$2.052X10^6$	1433	0.188	$3.061X10^9$	1.041	38
Dec	7945	$7.460X10^6$	2731	0.344	$6.950X10^{10}$	3.411	38
Jan	7755	$3.084X10^6$	1756	0.226	$9.069X10^9$	1.674	38
Feb	7402	$2.724X10^6$	1651	0.223	$6.121X10^9$	1.361	38
Mar	11606	$1.461X10^7$	3823	0.329	$6.327X10^{10}$	1.133	38
Apr	38289	$3.338X10^8$	18269	0.477	$3.445X10^{12}$	0.565	38
May	85770	$1.091X10^9$	33029	0.385	$7.962X10^{12}$	0.221	39
Jun	76265	$1.048X10^9$	32368	0.424	$1.190X10^{13}$	0.351	39
Jul	26255	$2.144X10^8$	14642	0.558	$3.108X10^{12}$	0.990	39
Aug	9488	$1.091X10^7$	3303	0.348	$3.762X10^{10}$	1.044	39
Sep	7115	$3.122X10^6$	1767	0.248	$3.895X10^9$	0.706	39
Yr	294519	$8.844X10^9$	94040	0.319	$1.903X10^{14}$	0.229	38

Serial Correlation coefficients

	Lag (k)								
Period	1	2	3	4	5	6	7	8	9
Oct	.786	.711	.609	.632	.466	.361	.345	.484	.424
Nov	.538	.441	.412	.407	.383	.140	-.047	.123	.293
Dec	.220	-.014	-.013	.045	.122	.113	-.036	-.182	-.234
Jan	.797	.473	.034	.041	.093	.139	.135	.007	-.197
Feb	.585	.590	.417	.126	.029	.105	.164	.221	-.048
March	.360	.516	.340	.326	.088	.049	.057	.160	.159
April	.346	.294	.655	.519	.438	-.029	-.169	-.120	-.095
May	.637	.383	.537	.630	.475	.411	.036	-.188	-.039
June	.685	.342	.527	.520	.645	.457	.209	-.172	-.302
July	.905	.573	.279	.464	.537	.557	.442	.165	-.177
Aug	.911	.872	.647	.387	.494	.657	.674	.608	.245
Sept	.899	.793	.757	.549	.316	.437	.602	.571	.550
Year	-.048	.090	-.294	-.007	=.205	-.067	.122	-.253	.172

Table B.8

Statistics for Station 13185000
(Streamflow in cfs days)

	Mean	Variance	Stand Dev	Coeff Var	Skew	Coeff Skew	n
Oct	12108	$7.368X10^{6}$	2714	0.224	$1.984X10^{10}$	0.992	72
Nov	13155	$1.657X10^{7}$	4071	0.309	$1.459X10^{11}$	2.163	72
Dec	15001	$6.101X10^{7}$	7811	0.521	$1.397X10^{12}$	2.931	72
Jan	14072	$2.789X10^{7}$	5281	0.375	$2.466X10^{11}$	1.675	72
Feb	14261	$2.573X10^{7}$	5072	0.356	$1.743X10^{11}$	1.335	72
Mar	24808	$1.108X10^{8}$	10528	0.424	$1.557X10^{12}$	1.334	72
Apr	63788	$7.411X10^{8}$	27224	0.427	$2.096X10^{13}$	1.039	73
May	118829	$1.522X10^{9}$	39007	0.328	$9.532X10^{12}$	0.161	73
Jun	103389	$1.876X10^{9}$	43310	0.419	$5.340X10^{12}$	0.066	73
Jul	38505	$4.327X10^{8}$	20801	0.540	$7.344X10^{12}$	0.816	73
Aug	14603	$2.252X10^{7}$	4745	0.325	$7.246X10^{10}$	0.678	73
Sep	11093	$5.995X10^{6}$	2448	0.221	$5.021X10^{9}$	0.342	73
Yr	442814	$1.711X10^{10}$	130801	0.295	$5.080X10^{14}$	0.227	72

Serial Correlation coefficients

Period	Lag (k) 1	2	3	4	5	6	7	8	9
Oct	.740	.616	.495	.453	.363	.365	.294	.454	.346
Nov	.446	.388	.307	.238	.254	.090	-.033	.067	.295
Dec	.327	.008	.026	.022	.066	.062	-.058	-.237	-.222
Jan	.657	.542	.073	.120	.117	.101	.112	-.007	-.219
Feb	.506	.452	.369	.130	-.021	-.001	.032	.013	-.122
March	.423	.482	.348	.370	.120	.085	.052	.082	.060
April	.454	.340	.565	.432	.278	-.041	-.134	-.095	-.100
May	.525	.451	.440	.505	.375	.486	.120	.021	.077
June	.629	.315	.339	.354	.529	.371	.250	-.057	-.004
July	.826	.532	.408	.350	.434	.505	.355	.211	-.091
Aug	.907	.814	.616	.469	.408	.547	.576	.442	.315
Sept	.885	.756	.717	.567	.389	.423	.558	.530	.438
Year	-.026	.133	-.071	.120	.064	-.014	.144	-.064	.155

Figure B.1

Correlogram for October Streamflow at Station 12413000 (Unextended)

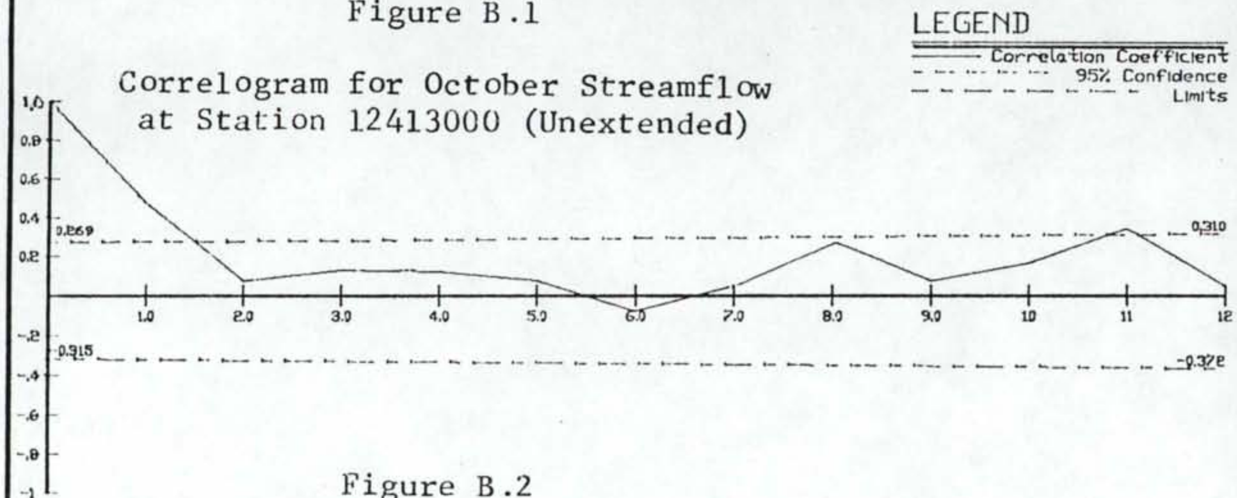


Figure B.2

Correlogram for November Streamflow at Station 12413000 (Unextended)

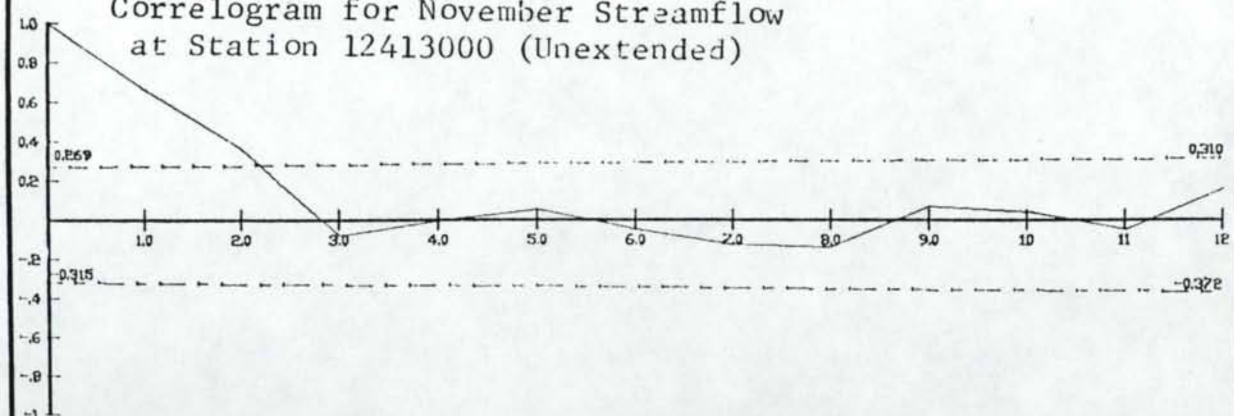


Figure B.3

Correlogram for December Streamflow at Station 12413000 (Unextended)

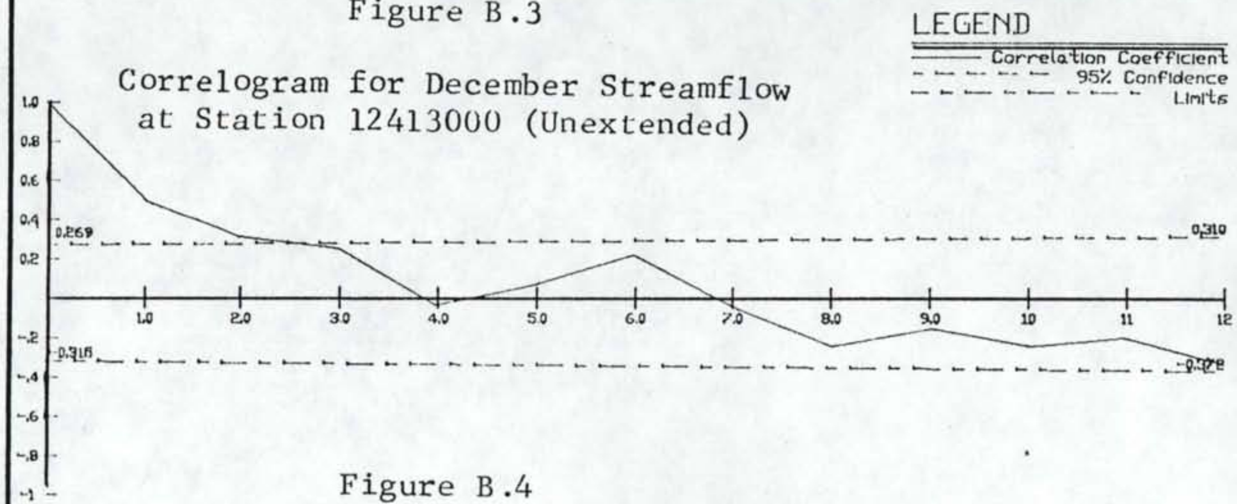


Figure B.4

Correlogram for January Streamflow at Station 12413000 (Unextended)

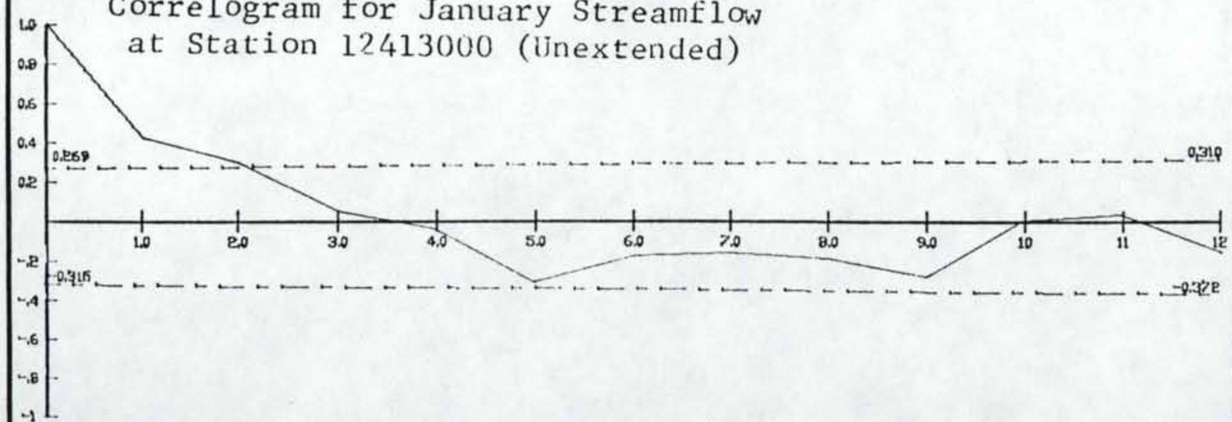


Figure B.5

Correlogram for February Streamflow at Station 12413000 (Unextended)

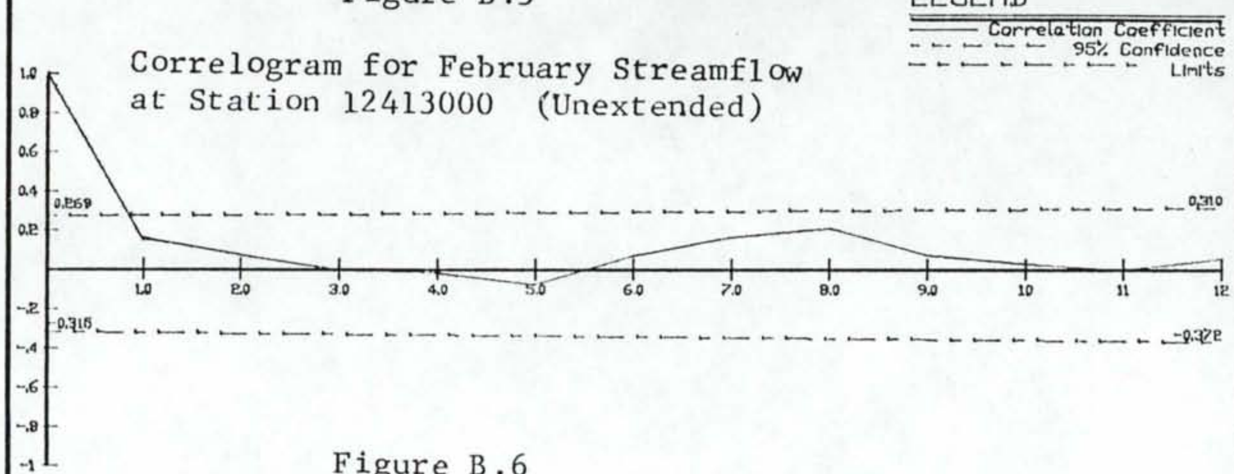


Figure B.6

Correlogram for March Streamflow at Station 12413000 (Unextended)

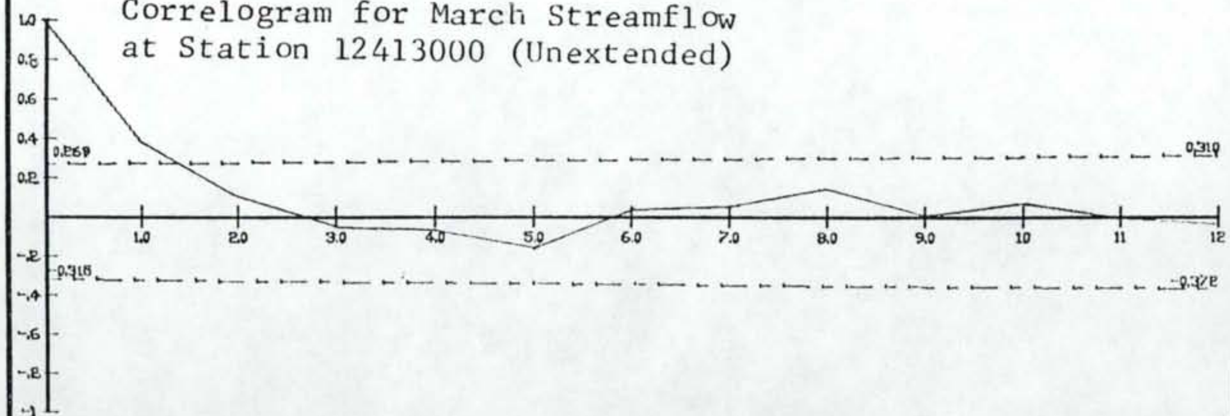


Figure B.7

Correlogram for April Streamflow at Station 12413000 (Unextended)

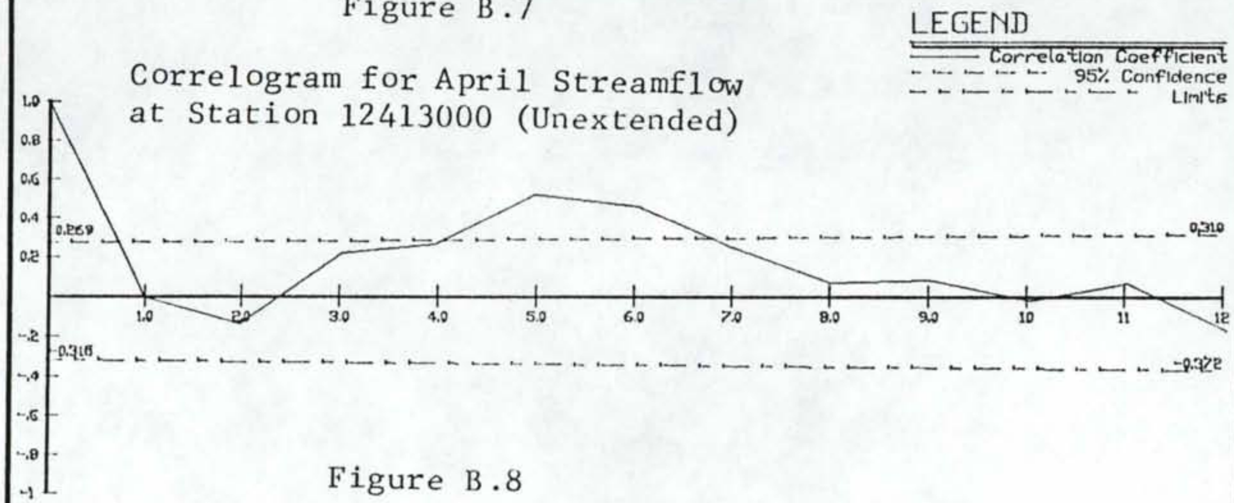
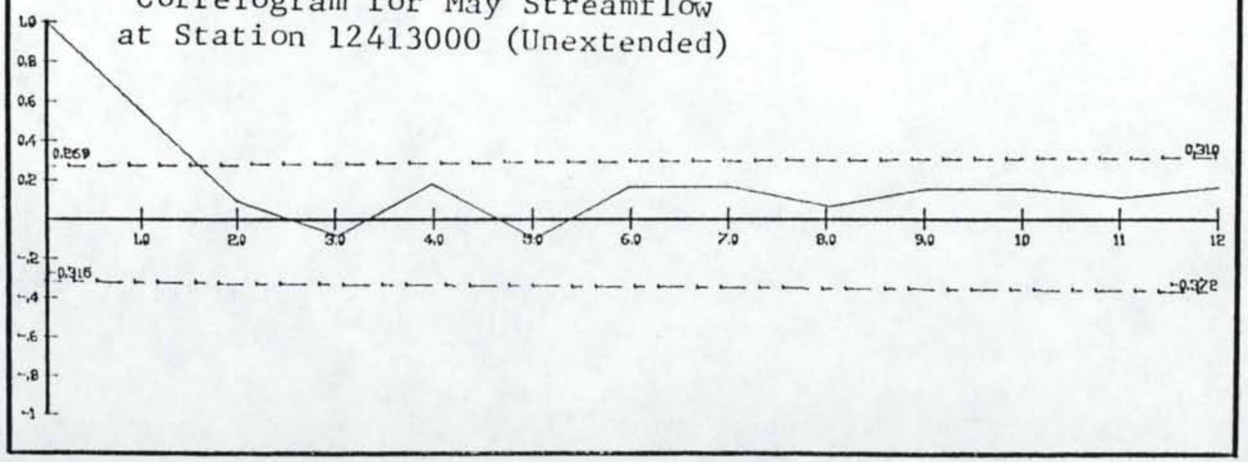


Figure B.8

Correlogram for May Streamflow at Station 12413000 (Unextended)



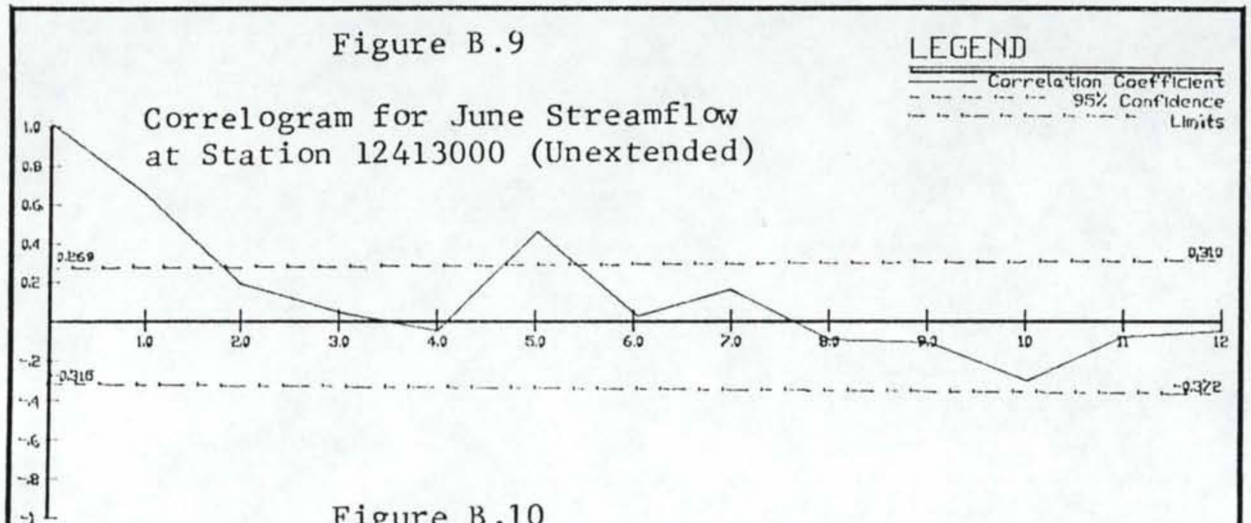


Figure B.9

Correlogram for June Streamflow at Station 12413000 (Unextended)

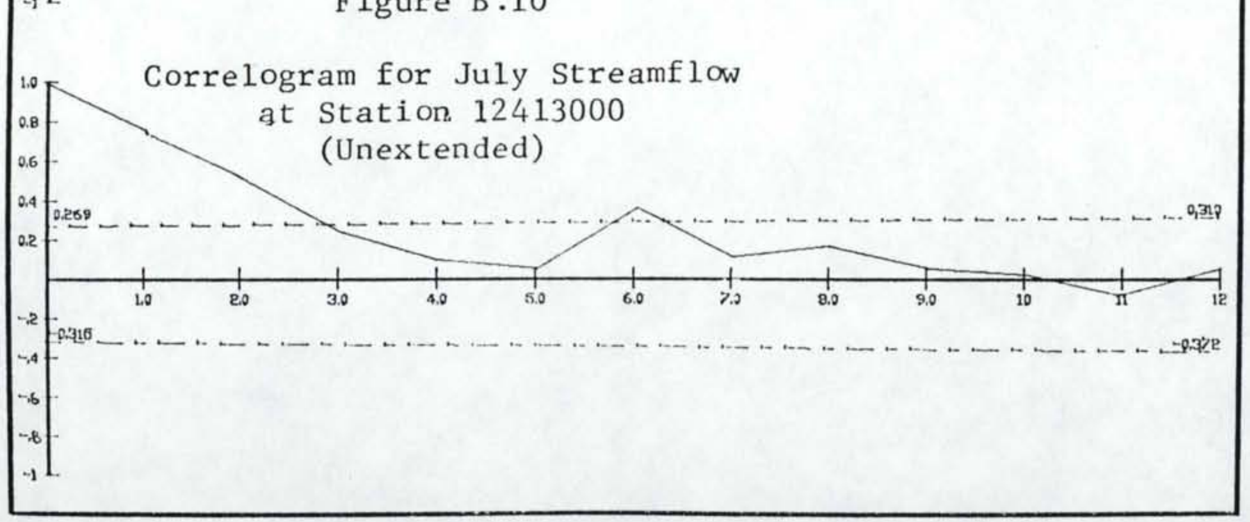


Figure B.10

Correlogram for July Streamflow at Station 12413000 (Unextended)

Figure B.11

Correlogram for August Streamflow at Station 12413000 (Unextended)

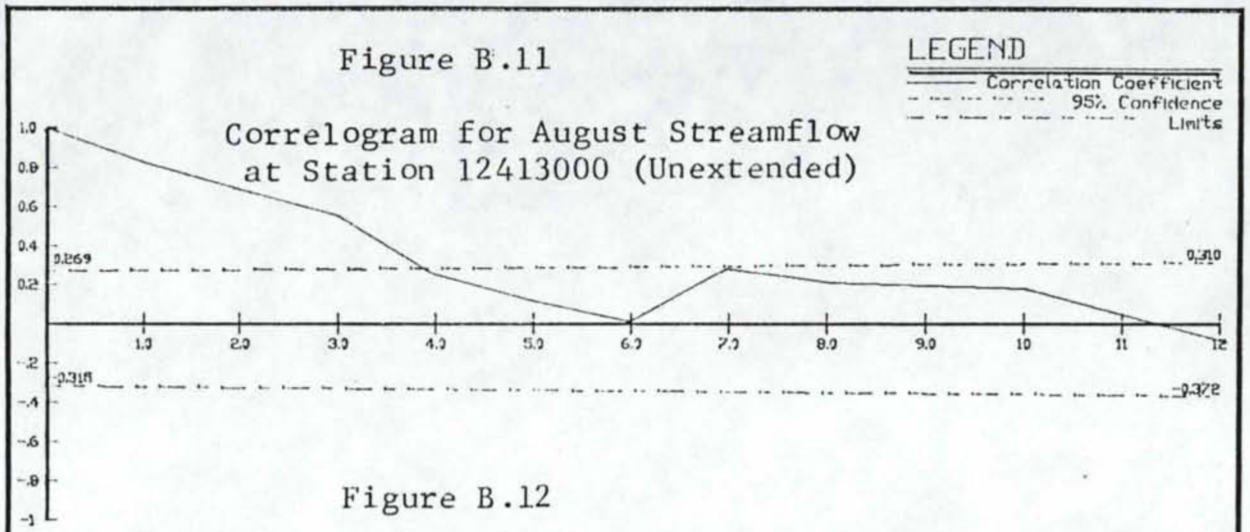
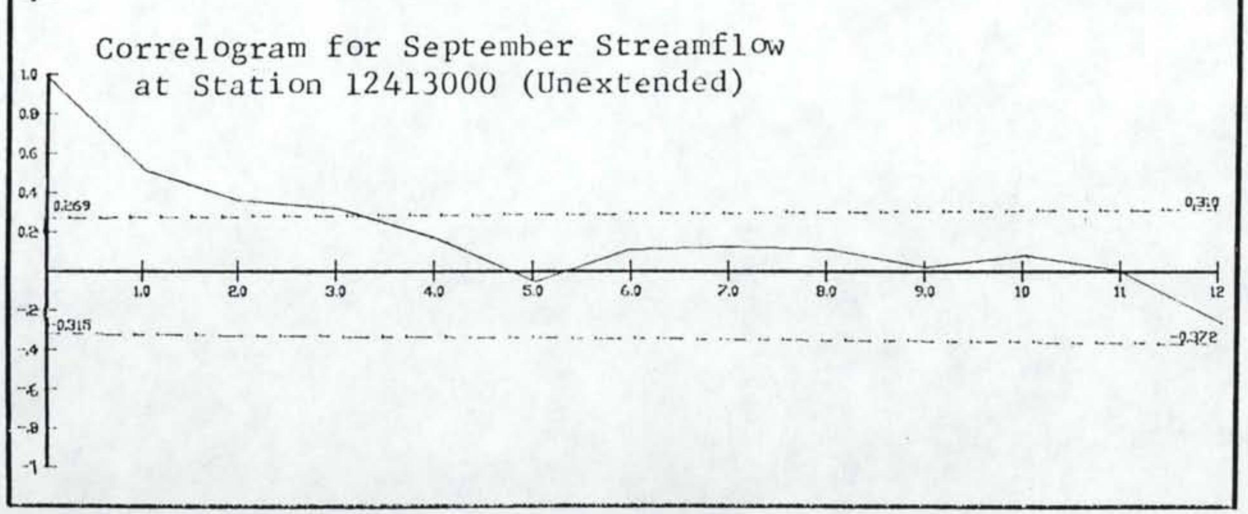


Figure B.12

Correlogram for September Streamflow at Station 12413000 (Unextended)



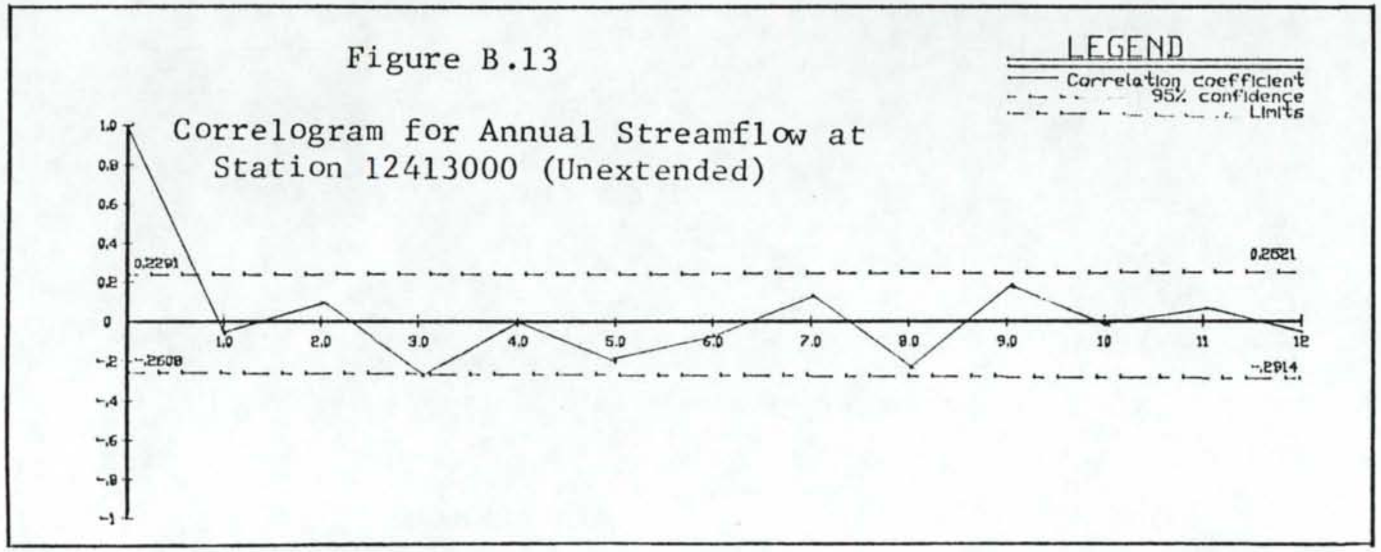


Figure B.13

Correlogram for Annual Streamflow at Station 12413000 (Unextended)

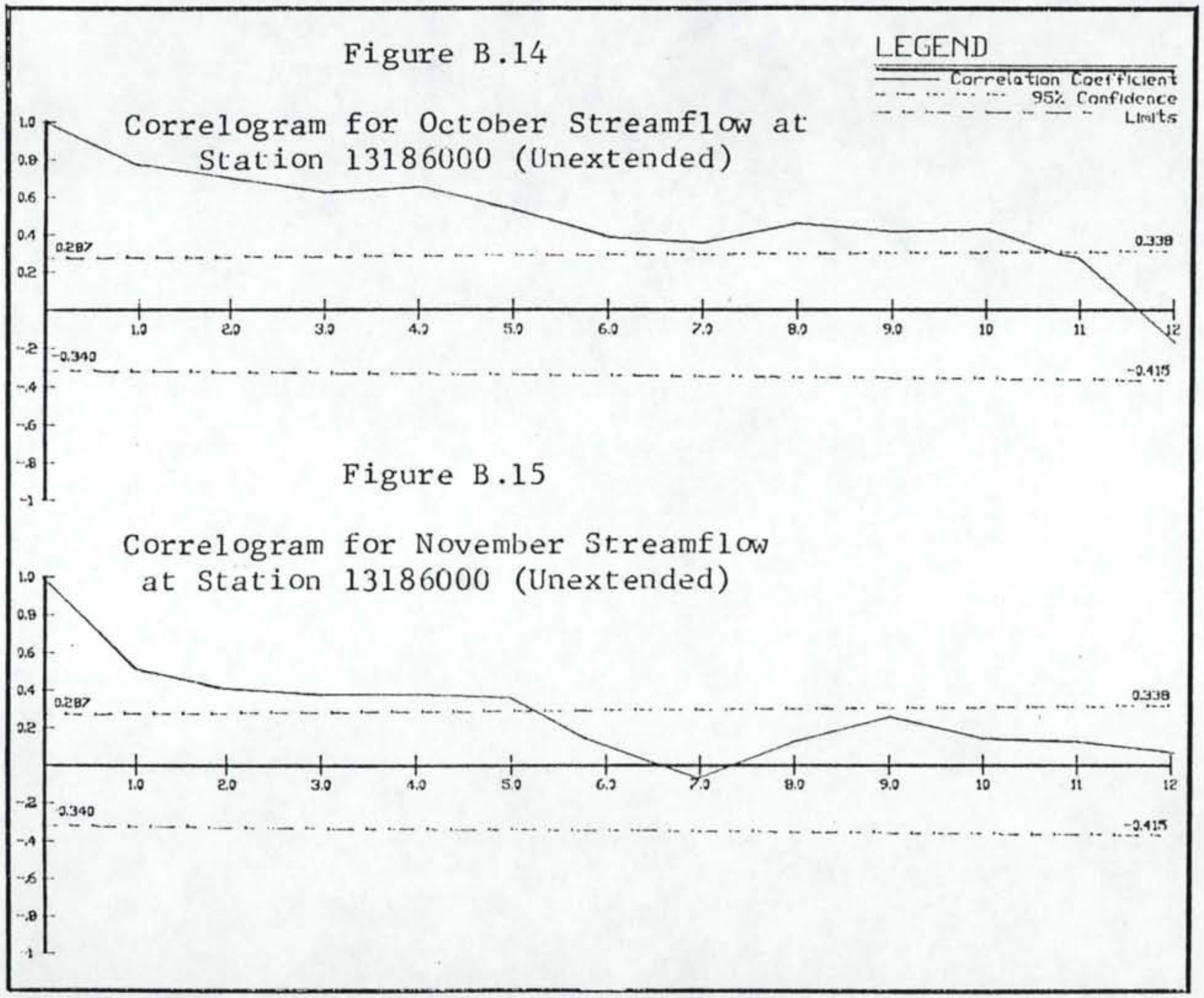


Figure B.14

Correlogram for October Streamflow at Station 13186000 (Unextended)

Figure B.15

Correlogram for November Streamflow at Station 13186000 (Unextended)

Figure B.16

Correlogram for December Streamflow at Station 13186000 (Unextended)

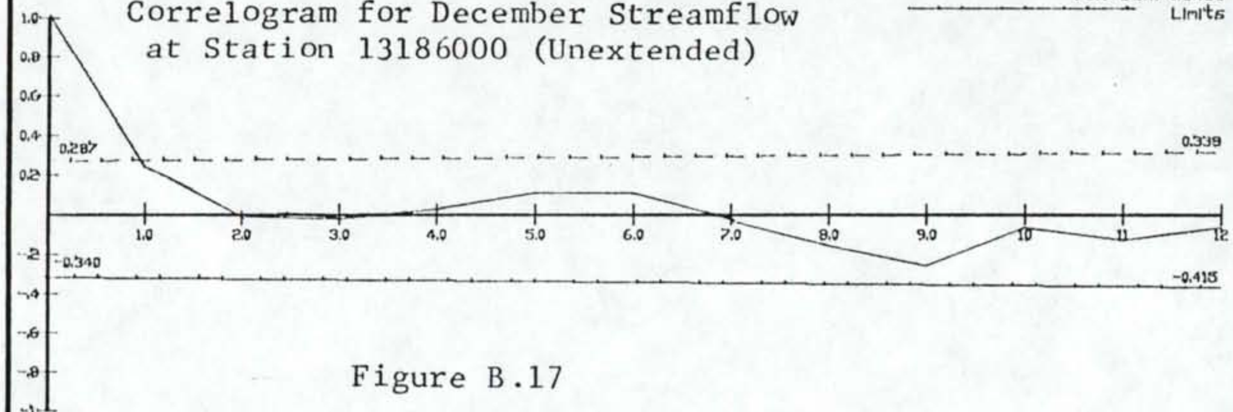
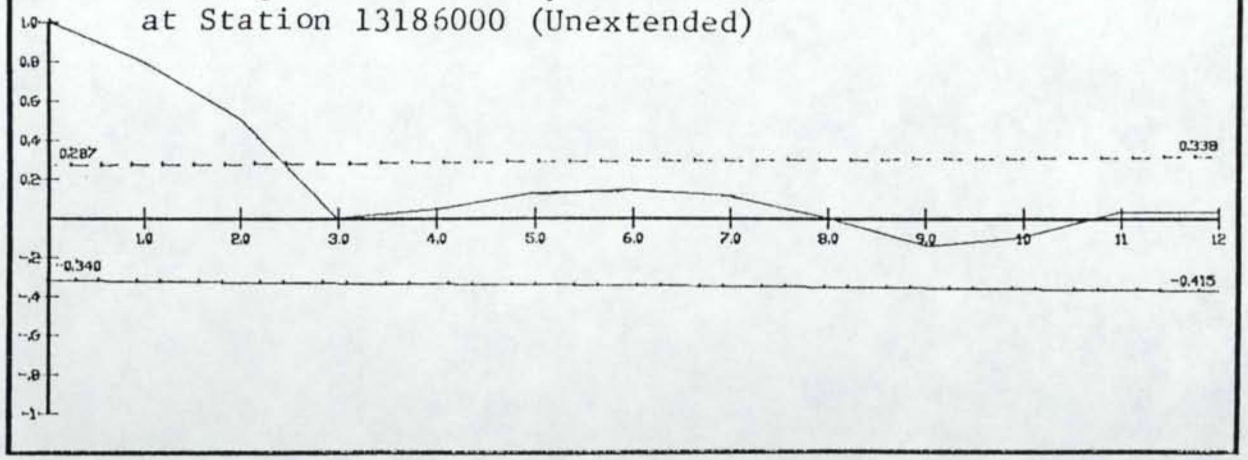


Figure B.17

Correlogram for January Streamflow at Station 13186000 (Unextended)



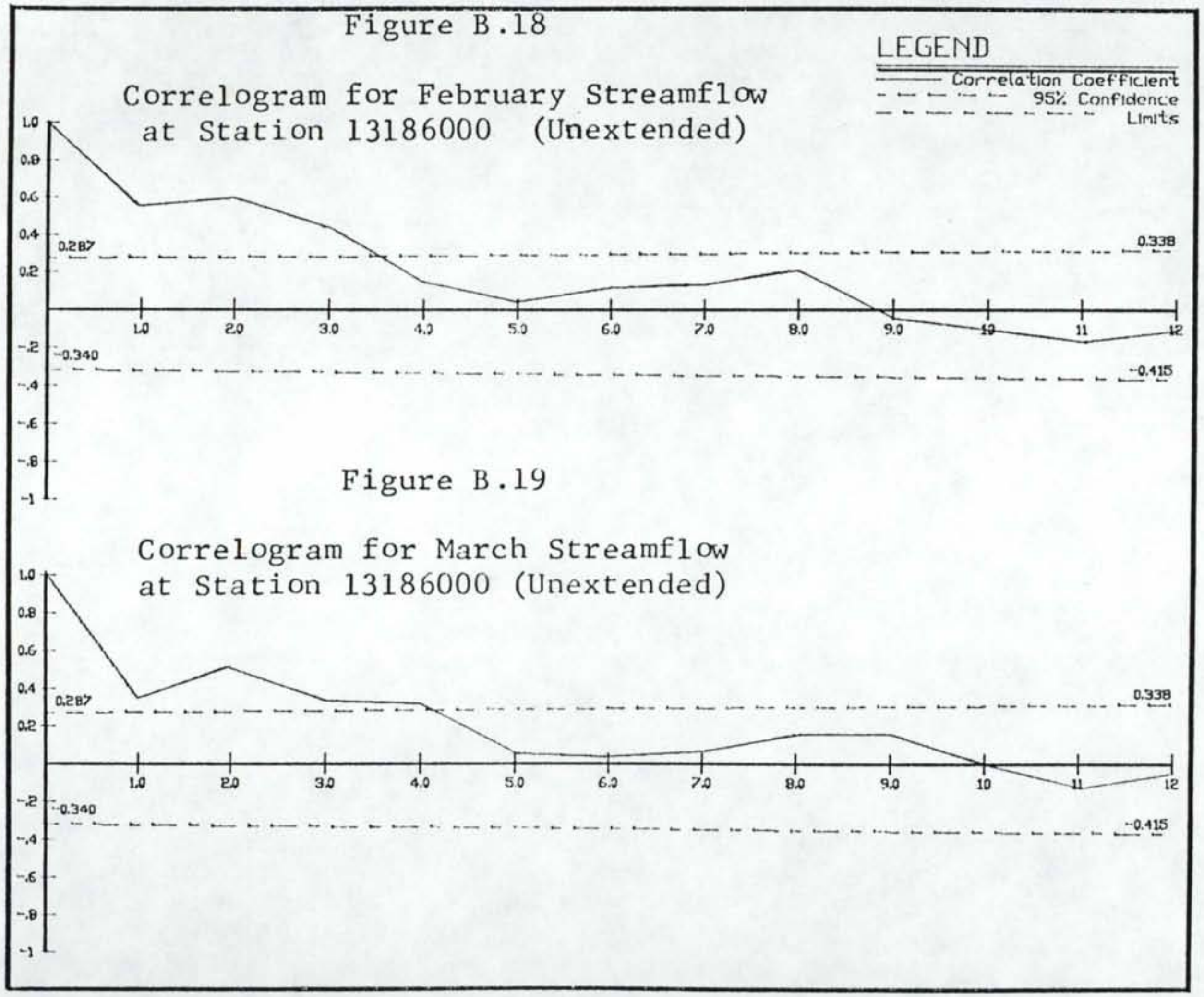


Figure B.18

Correlogram for February Streamflow at Station 13186000 (Unextended)

Figure B.19

Correlogram for March Streamflow at Station 13186000 (Unextended)

Figure B.20

Correlogram for April Streamflow at Station 13186000 (Unextended)

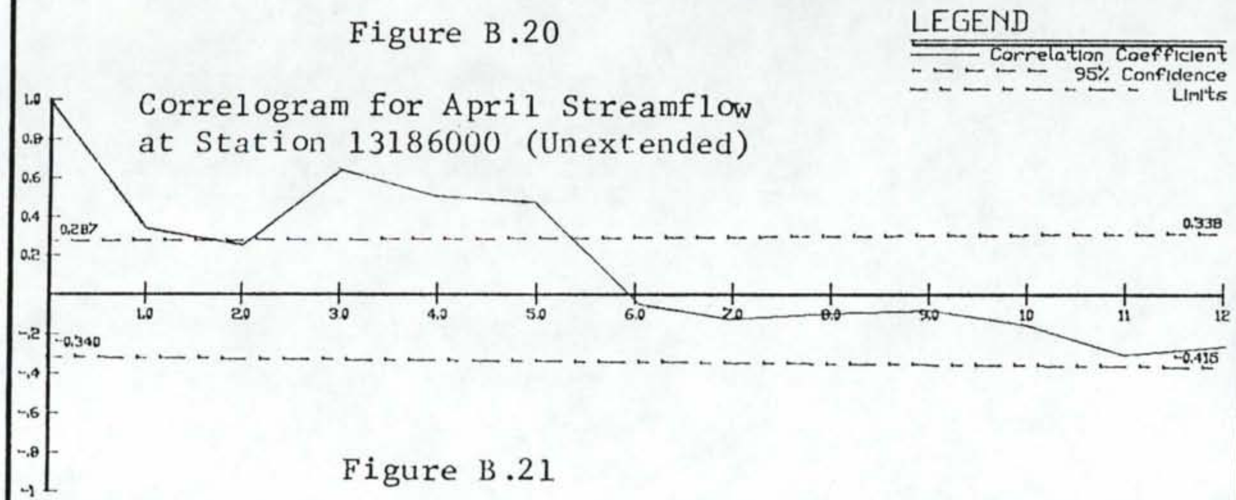
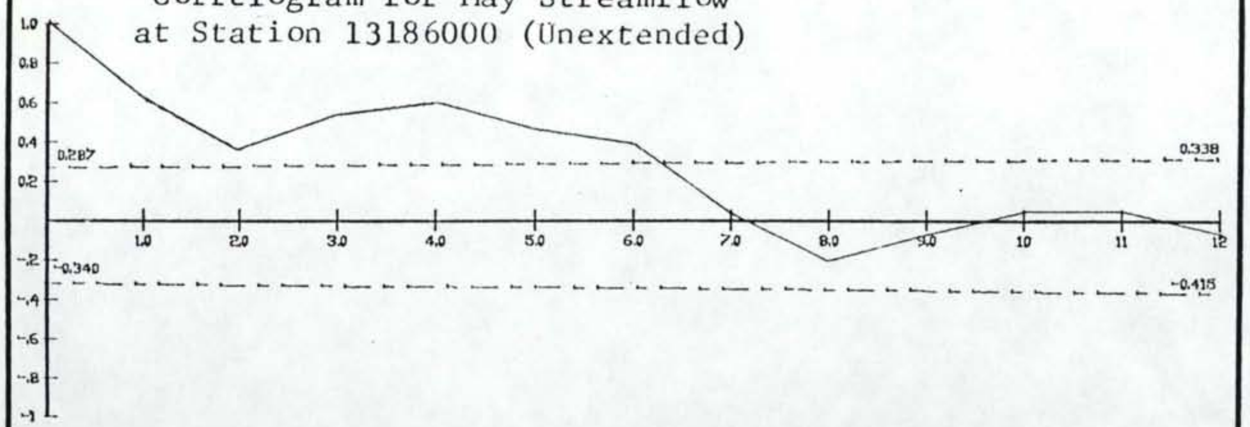


Figure B.21

Correlogram for May Streamflow at Station 13186000 (Unextended)



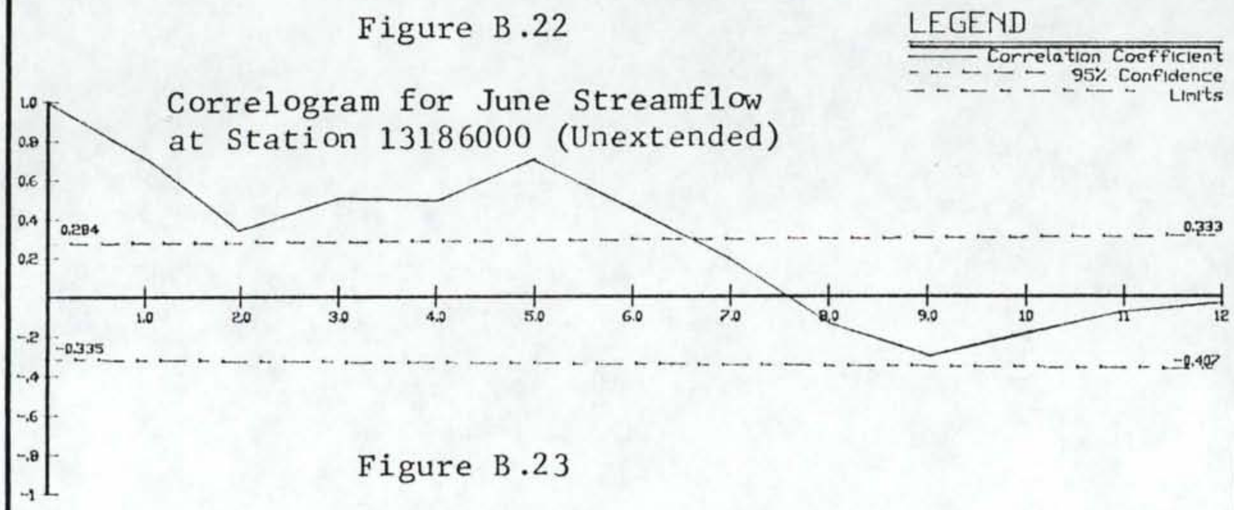


Figure B.22

Correlogram for June Streamflow at Station 13186000 (Unextended)

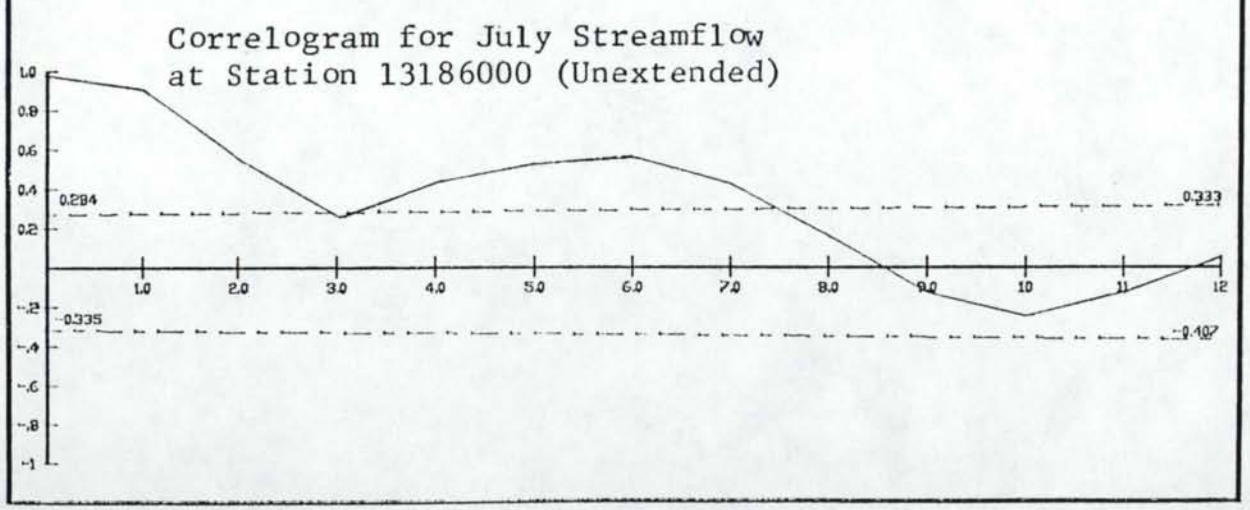


Figure B.23

Correlogram for July Streamflow at Station 13186000 (Unextended)

Figure B.24

Correlogram for August Streamflow at Station 13186000 (Unextended)

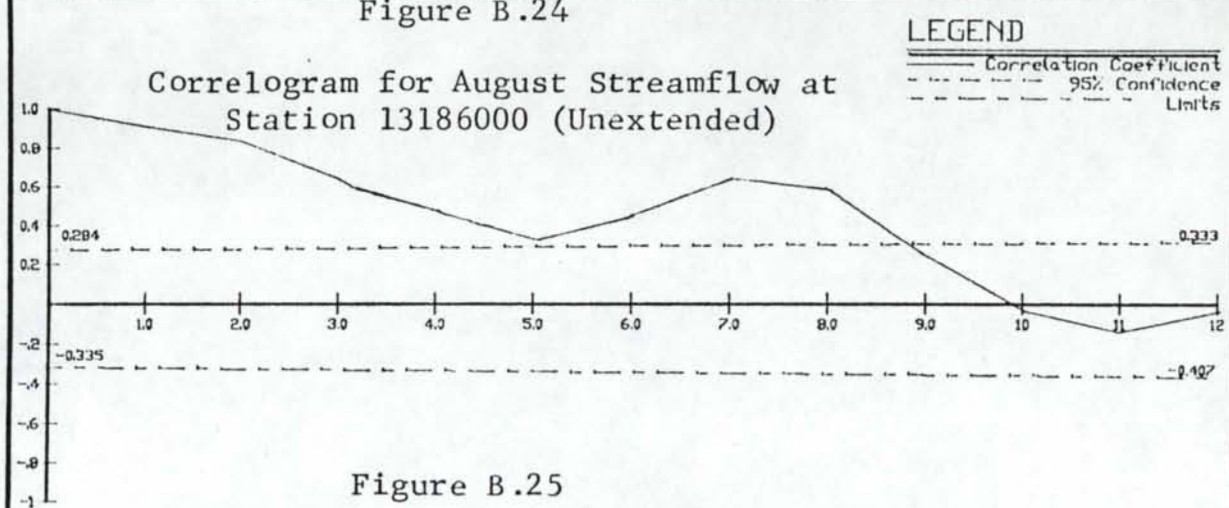
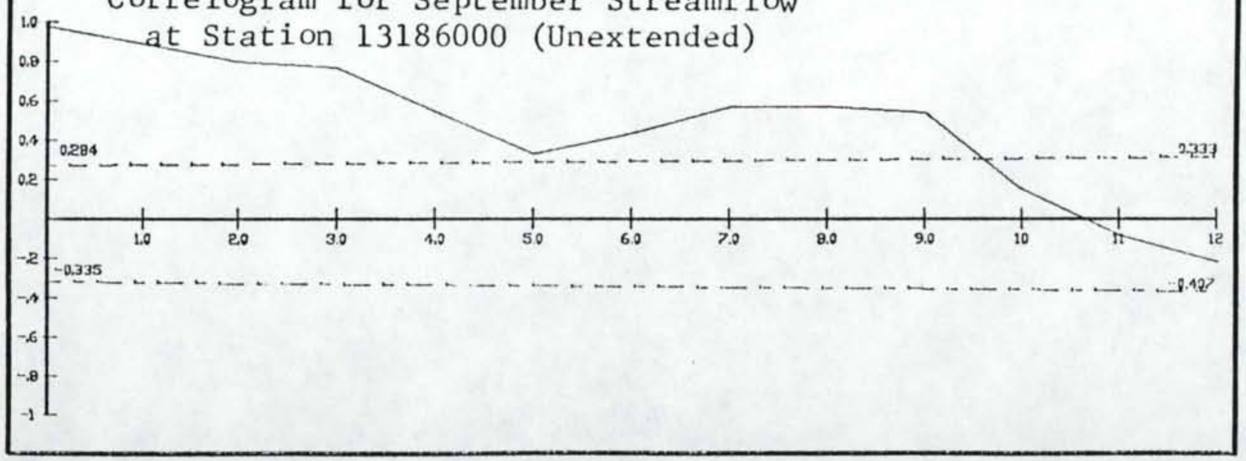


Figure B.25

Correlogram for September Streamflow at Station 13186000 (Unextended)



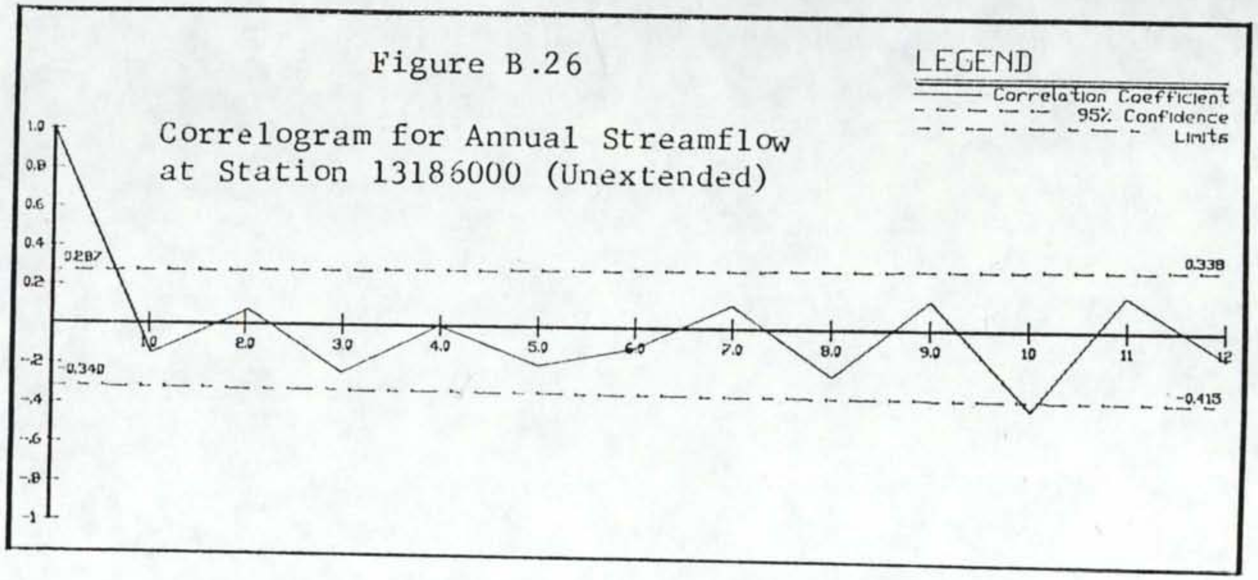


Figure B.26

Correlogram for Annual Streamflow at Station 13186000 (Unextended)

APPENDIX C

STATISTICS AND PARAMETERS OF DATA EXTENSION

Table C.1

Lag One Serial Correlation Coefficients and
Lag Zero Cross Correlation Coefficients

Period	Oct 1939 to Sept 1972			May 1945 to Setp 1983		
	*124130	124135		131860	131850	
	r_x	r_w	r_{xw}	r_x	r_w	r_{xw}
October	0.5502	0.5883	0.9909	0.7856	0.6068	0.9239
November	0.6870	0.7039	0.9909	0.5381	0.3471	0.8801
December	0.6096	0.6367	0.9975	0.2203	0.2606	0.9298
January	0.3472	0.3517	0.9941	0.7972	0.6339	0.8917
Feburary	0.2787	0.2687	0.9959	0.5853	0.3671	0.9162
March	0.2827	0.2816	0.9951	0.3600	0.2856	0.9452
April	-.0513	-.0013	0.9952	0.3458	0.3742	0.9786
May	0.3809	0.3715	0.9943	0.6365	0.5963	0.9459
June	0.7239	0.6944	0.9930	0.6853	0.6270	0.9499
July	0.8127	0.8495	0.9830	0.9055	0.8705	0.9699
August	0.8616	0.8881	0.9658	0.9110	0.8958	0.9547
September	0.5106	0.5388	0.9769	0.8991	0.8534	0.9605

* truncated station numbers

Table C.2

Statistics for Period of Overlapping Record
Streamflow in cfs days

Period	12413000 (10/39-9/72)			12413500 (10/39-9/72)		
	Mean	Std Dev	Coef Sk	Mean	Std Dev	Coef Sk
October	13845	9049	1.684	19230	11474	1.508
November	29304	21760	1.011	38809	27046	.938
December	48585	42032	1.397	63731	53439	1.475
January	41878	23937	.900	56752	30164	.726
Feburary	57816	42286	1.412	76905	53591	1.402
March	74983	44228	2.052	97789	55090	2.191
April	172436	60972	.174	217362	76735	.203
May	181806	71385	-.207	241186	88188	-.196
June	61055	29964	.177	89898	41269	.950
July	20574	7557	.814	30924	11112	.743
August	10683	2752	.850	15756	4234	.762
September	8994	2072	1.354	13043	3029	1.330

Period	13186000 (5/45 - 9/83)			13185000 (5/45 - 9/83)		
	Mean	Std Dev	Coef Sk	Mean	Std Dev	Coef Sk
October	7596	1403	.381	12823	2838	1.155
November	7629	1433	1.041	13750	3553	1.335
December	7945	2731	3.411	16647	9303	2.792
January	7755	1756	1.674	15711	5923	1.440
Feburary	7402	1650	1.361	16130	5613	1.112
March	11606	3823	1.133	26278	11477	1.530
April	38288	18269	.565	65950	27565	.267
May	85770	33029	.221	126353	39316	-.104
June	76265	32368	.351	111583	38434	.187
July	26255	14642	.990	43079	20705	.574
August	9488	3303	1.044	15982	4628	.701
September	7115	1767	.706	11877	2343	.427

Table C.3

Parameters for Simple Linear Regression

Period	12413000 & 13135000 B_R	A_R	13186000 & 13185000 B_R	A_R
October	.7815	-1183	.4569	1737
November	.7972	-1635	.3549	2749
December	.7846	-1416	.2730	3400
January	.7889	-2893	.2644	3601
Feburary	.7858	-2616	.2694	3057
March	.7989	-3143	.3148	3332
April	.7908	554	.6486	-4486
May	.8048	-12306	.7946	-14632
June	.7210	-3757	.8000	-13001
July	.6685	-97	.6859	-3294
August	.6277	794	.6812	-1399
September	.6683	278	.7242	-1487

Table C.4

Parameters for Yevjevich Model

Period	12413000 & 13135000 A_y	B_y	13186000 & 13185000 A_y	B_y
October	.8288	.1022	.6083	.1134
November	.7204	.0949	.7632	.3578
December	.7915	.0437	.9078	.3569
January	.9323	.1016	.5703	.1980
Feburary	.9565	.0863	.7733	.2437
March	.9545	.0948	.8849	.2956
April	.9951	.0841	.9187	.1907
May	.9194	.0981	.7311	.2457
June	.6863	.0706	.6954	.2162
July	.5769	.0821	.4173	.0769
August	.4934	.1192	.3951	.1183
September	.8406	.1808	.4288	.0880

Table C.5

Predicted Values from Yevjevich Model at
Station 12413000

Year	Oct	Nov	Dec	Jan	Feb	March	April	May	June	July	Aug	Sept	Annual
19-20											8775	12636	
20-21	22404	30406	38858	118274	94974	160427	190103	200518	45071	17433	10321	7880	936669
21-22	8747	10373	75526	16090	16875	15230	124214	218266	56801	14913	9517	7826	574380
22-23	7992	7004	14276	70759	15388	45648	227308	187815	87903	24422	13085	8842	710443
23-24	11514	14489	23303	9968	104099	52220	122715	135390	19893	10528	7942	6699	518760
24-25	7868	23015	44650	35605	143257	119684	262760	152599	39789	16412	10839	7597	864075
25-26	6291	5758	19065	13372	45134	82784	120771	44675	13732	8285	6915	11476	378258
26-27	24436	41363	75498	39460	63062	69133	189842	221083	101063	23462	10724	17210	876335
27-28	47244	153401	77938	66673	41307	115303	132334	192009	37904	15817	10879	7690	898499
28-29	10491	13826	11990	2085	5587	43981	117296	136417	37127	14359	8153	7578	408891
29-30	9263	9456	16994	3534	25552	48328	122348	51955	21498	10924	6343	5700	331896
30-31	8640	11584	8170	8829	21022	94630	147817	83834	11566	7247	5453	6144	414935
31-32	7225	8352	8563	19659	39255	111965	284745	267491	72576	18277	9556	8674	856338
32-33	11101	63400	61306	40350	14146	52627	212583	215808	141266	23249	10504	9493	855834
33-34	17630	45004	321012	200124	93552	155531	150299	45437	14199	9651	6661	6821	1065921
34-35	18733	51617	52197	56546	51828	70218	188874	218891	57292	17062	9951	6628	799837
35-36	8368	11482	12574	11225	6222	49065	263057	148322	34656	13792	7354	7299	573414
36-37	7820	11269	22623	7511	12657	32298	152926	176952	44147	16148	9764	7355	501470
37-38	3957	32108	54037	61961	26081	91582	259950	135772	36211	14377	8115	6735	730885
38-39	8526	7850	16516	19446	7795	64698	186165	109518	21654	12582	6718	6688	468158

Table C.6

Predicted Values from Yevjevich Model at
Station 13186000

Year	Oct	Nov	Dec	Jan	Feb	March	April	May	June	July	Aug	Sept	Annual
10-11							27043	67473	98154	28125	7928	6399	0
11-12	6997	7002	7017	7481	6237	9839	32908	85799	113465	29868	12486	9483	328583
12-13	8913	7642	6786	6534	5775	9015	40107	90968	70089	23632	11755	7726	288942
13-14	8584	9131	8313	7992	7520	17417	51267	93359	58915	20811	7877	7588	298773
14-15	9221	7341	6170	6575	5331	7158	18310	31082	13917	6555	5167	5440	122268
15-16	5817	6206	6297	5915	4833	11710	50116	71342	96432	47889	12117	8065	326739
16-17	8334	7021	6624	6301	5695	6917	21762	81745	110115	40640	8980	6201	310336
17-18	6667	6704	11128	9249	6695	14233	45377	70839	107377	21431	9349	7227	316275
18-19	8020	6357	5380	5389	5659	8383	32339	68664	18745	2533	4110	3787	169364
19-20	5352	6825	6585	7693	6000	9966	22500	84557	74587	23647	8911	7886	264509
20-21	8778	8332	7066	8256	7668	12944	35328	147538	139222	35751	11858	8952	431693
21-22	8109	9017	10657	7633	6042	8136	22476	99641	105404	23193	9043	6121	315471
22-23	6666	5687	5321	5946	5770	7813	22490	62693	69712	32322	9626	6678	240724
23-24	7466	6686	6579	6456	6441	7361	10883	45952	11626	3973	3631	3681	120737
24-25	5460	6578	5235	6093	7947	12541	50780	107778	71120	30467	10193	7600	321792
25-26	8088	7237	6785	6698	6012	10910	29314	32368	1259	2635	4119	4009	119434
26-27	5094	7473	8819	7793	7650	11804	37381	93039	134365	43392	12551	9467	378830
27-28	9490	15621	9114	8396	6794	16347	29045	153263	61549	19472	8051	5878	343020
28-29	6639	6438	6783	6744	5514	8152	10509	64437	58399	18223	6201	5143	203182
29-30	6062	5012	7576	6098	6662	6866	31914	37295	44433	12955	6386	5263	176523
30-31	7304	7355	6525	6679	5203	6078	20297	41379	7322	2215	2934	3307	116598
31-32	4981	5411	4798	5705	5055	8299	28825	92818	87965	24830	8429	5984	283099
32-33	6359	6167	4031	5358	4085	5995	29634	56633	97122	19105	7321	5367	247176
33-34	6180	6256	6861	7275	7266	13705	36496	29047	13977	5674	3664	3900	140302
34-35	5480	6159	5542	6041	5372	6504	28069	59926	49250	10675	5271	4205	192496
35-36	5157	5088	5075	5399	4742	5885	61042	100139	59219	14643	6404	5280	278073
36-37	5817	5842	5356	5752	4822	7649	20742	62571	21953	6210	4121	3743	154577
37-38	5238	5564	10560	7797	6759	11003	41872	98488	94355	29810	9675	6693	327814
38-39	7436	7631	7606	6646	6058	14417	38547	48725	13300	6611	3904	3893	164774
39-40	5374	5279	7296	6804	6346	16581	41153	66316	32857	7637	4013	5001	204656
40-41	6495	6815	7237	7162	6812	10094	25645	53882	33537	6894	6405	5676	176655
41-42	6301	7716	10023	7869	6544	6724	36667	33226	58211	21443	6937	5048	206709
42-43	5293	6837	7648	9730	8501	15019	109864	110289	105980	57844	14896	8155	460057
43-44	8397	10022	6591	6525	6780	8222	18939	35243	15362	6875	4827	4509	132291
44-45	5419	5623	7008	7726	7881	8123	16774	0	0	0	0	0	0

Table C.7

Statistics for Extended Record at Station 12413000
Streamflow in cfs days

	Mean	Variance	Stand Dev	Coeff Var	Skew	Coeff Skew	n
Oct	12666	7.536×10^{7}	8681	0.685	1.471×10^{12}	2.248	63
Nov	26926	6.295×10^{8}	25090	0.932	4.047×10^{13}	2.562	63
Dec	50805	2.748×10^{9}	52424	1.032	3.885×10^{14}	2.696	63
Jan	44500	1.574×10^{9}	39677	0.892	1.490×10^{14}	2.385	63
Feb	53134	1.877×10^{9}	43327	0.815	1.212×10^{14}	1.490	63
Mar	76559	1.774×10^{9}	42120	0.550	1.050×10^{14}	1.405	63
Apr	167869	3.560×10^{9}	59668	0.355	$-.412 \times 10^{13}$	0.339	63
May	167862	4.975×10^{9}	70537	0.420	4.481×10^{14}	-.118	63
Jun	576519	1.130×10^{9}	33609	0.583	3.007×10^{13}	1.180	63
Jul	19300	5.573×10^{7}	7466	0.387	1.242×10^{11}	0.723	64
Aug	10303	7.829×10^{6}	2798	0.272	1.728×10^{10}	0.567	64
Sep	8758	4.883×10^{6}	2210	0.252	$-.162 \times 10^{90}$	1.602	63
Yr	696295	4.215×10^{10}	205315	0.295	5.080×10^{16}	-.187	63

Serial Correlation coefficients

Period	Lag (k) 1	2	3	4	5	6	7	8	9
Oct	.669	.094	.153	.214	.065	-.064	.049	.250	.122
Nov	.773	.594	-.000	.094	.211	.084	-.013	-.034	.074
Dec	.423	.299	.026	.061	.172	.394	.079	-.102	-.114
Jan	.597	.323	.165	.148	-.126	.008	.157	-.011	-.196
Feb	.248	.174	.032	.026	.033	.122	.206	.232	.047
March	.424	.303	.156	.103	.025	.119	-.050	.098	.042
April	.063	-.109	.169	.100	.277	.229	.053	-.071	-.074
May	.426	.038	-.008	.105	-.148	.182	.188	.157	.084
June	.690	.216	-.048	-.028	.281	-.012	.152	-.025	.023
July	.779	.570	.195	.048	.091	.264	.035	.110	.059
Aug	.874	.692	.584	.189	.074	.081	.246	.116	.199
Sept	.473	.425	.429	.225	-.067	.061	.146	.063	.019
Year	.017	.141	-.085	.064	-.070	-.125	.053	-.161	-.045

Table C.8

Statistics for Extended Record at Station 13186000
Streamflow in cfs days

	Mean	Variance	Stand Dev	Coeff Var	Skew	Coeff Skew	n
Oct	72178	$2.065X10^6$	1437	0.199	$1.067X10^9$	0.360	72
Nov	73613	$2.833X10^6$	1683	0.229	$1.003X10^{10}$	2.104	72
Dec	75324	$5.430X10^6$	2330	0.309	$4.009X10^{10}$	3.169	72
Jan	73679	$2.319X10^6$	1523	0.207	$6.167X10^9$	1.746	72
Feb	68583	$2.261X10^6$	1504	0.219	$4.603X10^9$	1.354	72
Mar	108730	$1.349X10^7$	3673	0.338	$4.831X10^{10}$	0.975	72
Apr	360469	$3.252X10^8$	18034	0.500	$8.009X10^{12}$	1.366	73
May	797754	$1.058X10^9$	32532	0.408	$1.470X10^{13}$	0.427	73
Jun	701874	$1.316X10^9$	36275	0.517	$3.684X10^{12}$	0.077	73
Jul	234519	$2.142X10^8$	14635	0.624	$2.577X10^{12}$	0.822	73
Aug	8619	$1.114X10^7$	3337	0.387	$2.641X10^{10}$	0.710	73
Sep	6587	$3.381X10^6$	1839	0.279	$2.931X10^9$	0.471	73
Yr	271726	$9.169X10^9$	95757	0.352	$2.382X10^{14}$	0.271	72

Serial Correlation coefficients

Period	Lag (k) 1	2	3	4	5	6	7	8	9
Oct	.872	.794	.685	.684	.535	.408	.424	.508	.463
Nov	.615	.577	.537	.497	.466	.266	.185	.231	.390
Dec	.288	.088	.116	.134	.173	.158	.057	-.124	-.092
Jan	.796	.437	.121	.143	.158	.178	.150	.043	-.118
Feb	.654	.580	.355	.172	.122	.161	.169	.147	-.024
March	.469	.559	.383	.364	.137	.120	.118	.154	.136
April	.426	.336	.555	.391	.177	-.109	-.136	-.100	-.095
May	.533	.424	.481	.556	.392	.420	.105	.017	.050
June	.679	.342	.367	.413	.539	.389	.131	-.108	-.101
July	.881	.596	.404	.393	.489	.517	.371	.112	-.129
Aug	.920	.867	.669	.460	.429	.588	.616	.496	.194
Sept	.921	.815	.798	.608	.378	.440	.573	.560	.461
Year	-.037	.046	-.121	.100	.087	-.034	.163	-.067	.129

Figure C.1

LEGEND

——— Correlation coefficient
- -- - - 95% confidence Limits

Correlogram for October Streamflow at Station 12413000 (Extended)

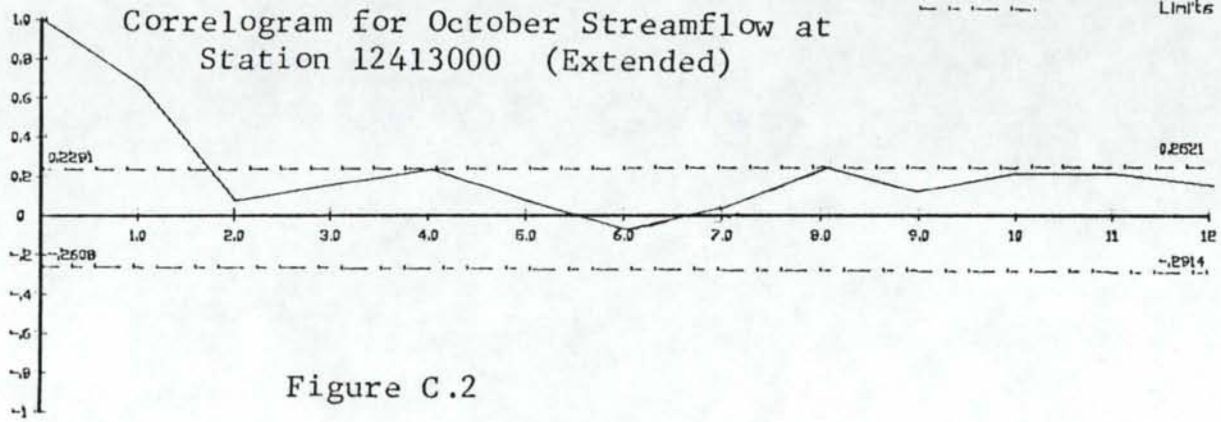


Figure C.2

Correlogram for November Streamflow at Station 12413000 (Extended)

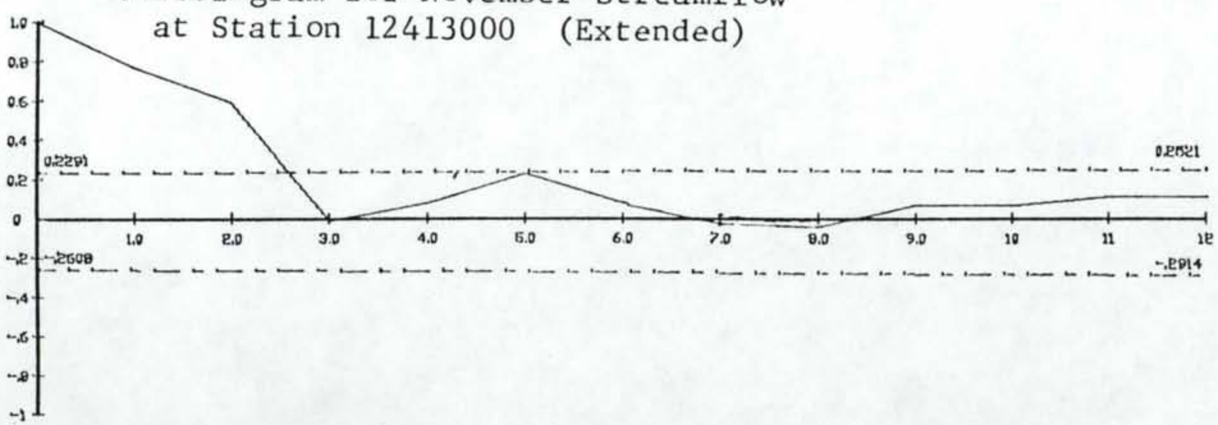


Figure C.3

Correlogram for December Streamflow at Station 12413000 (Extended)

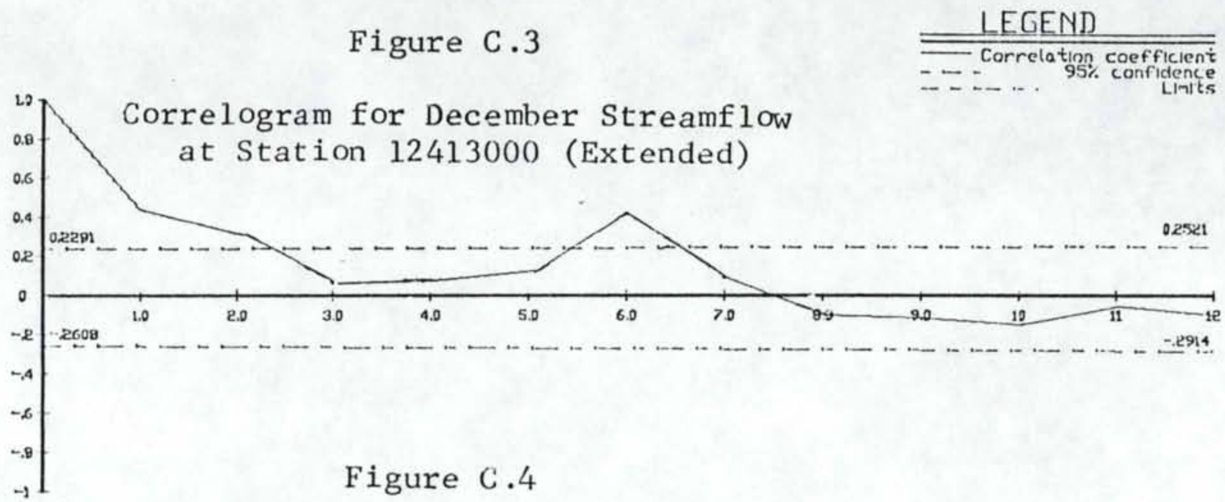


Figure C.4

Correlogram for January Streamflow at Station 12413000 (Extended)

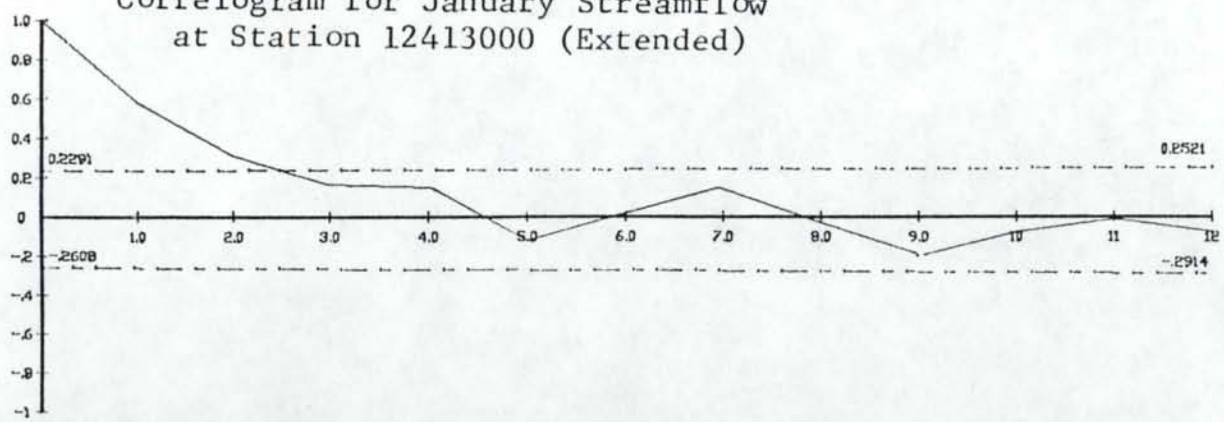


Figure C.5

Correlogram for February Streamflow at Station 12413000 (Extended)

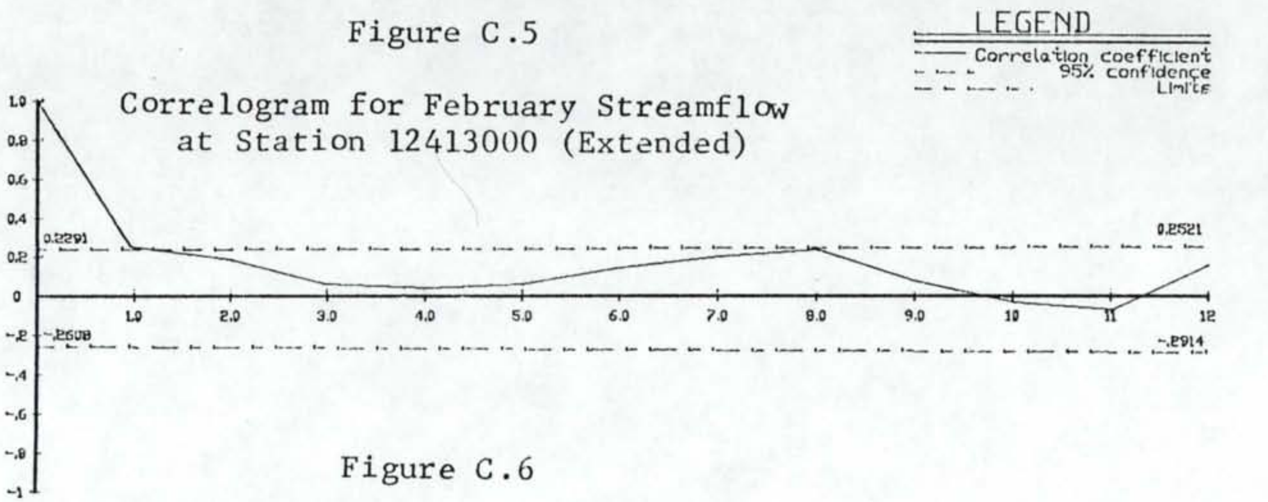


Figure C.6

Correlogram for March Streamflow at Station 12413000 (Extended)

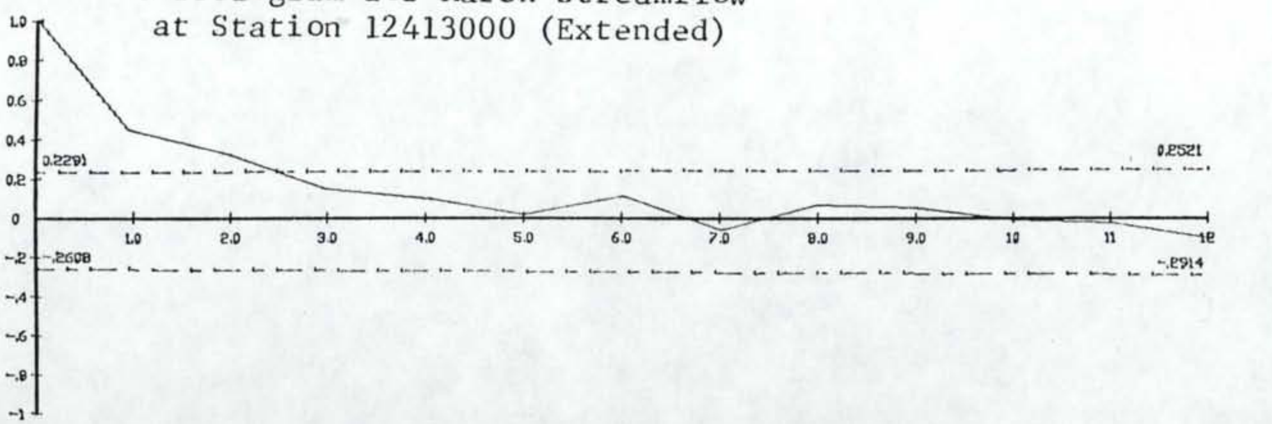


Figure C.7

Correlogram for April Streamflow
at Station 12413000 (Extended)

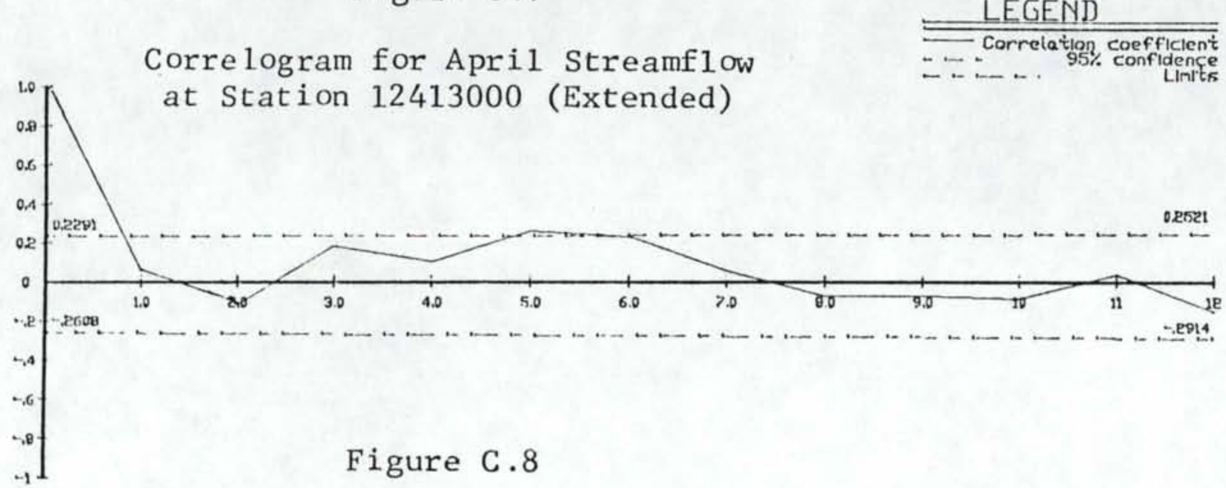


Figure C.8

Correlogram for May Streamflow
at Station 12413000 (Extended)

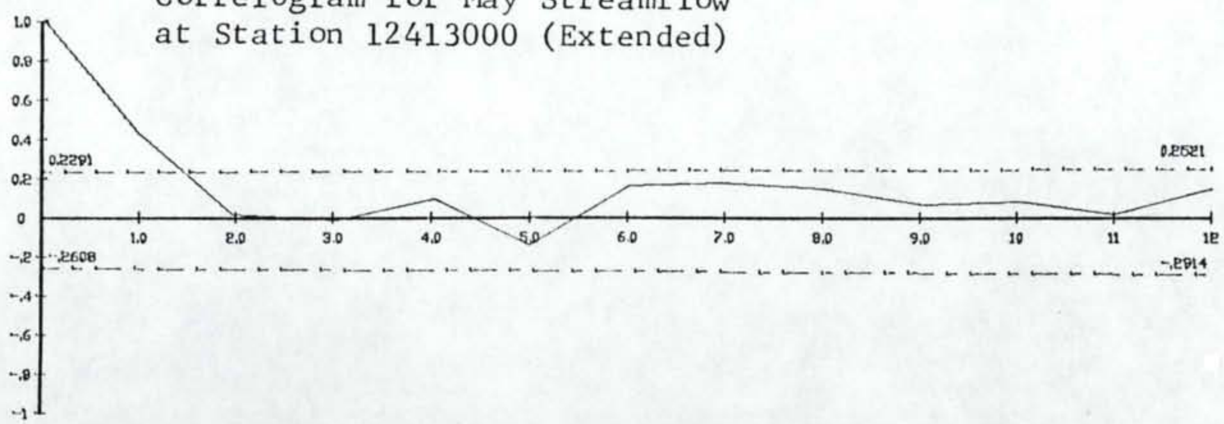


Figure C.9

Correlogram for June Streamflow at Station 12413000 (Extended)

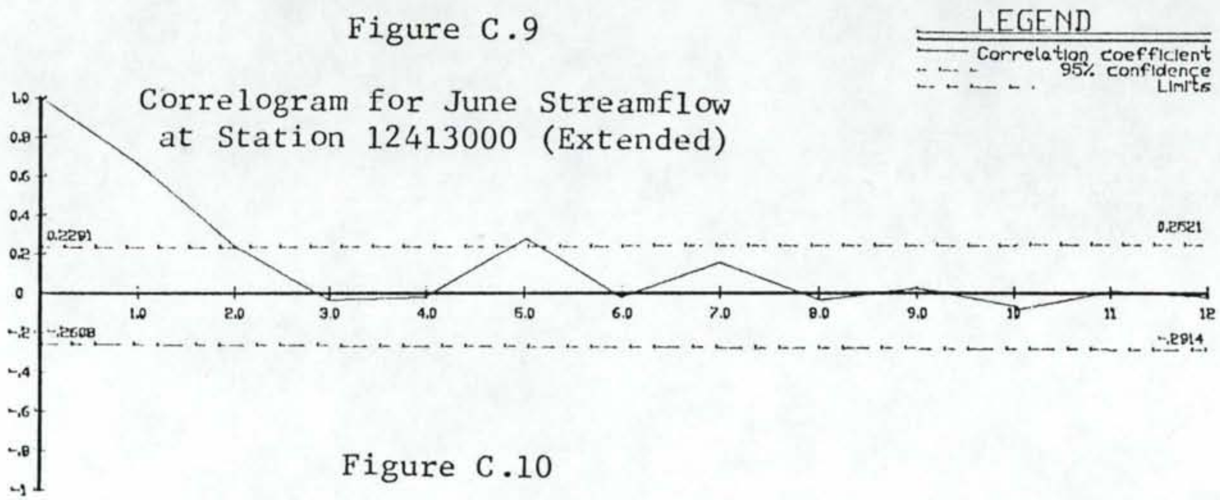


Figure C.10

Correlogram for July Streamflow at Station 12413000 (Extended)

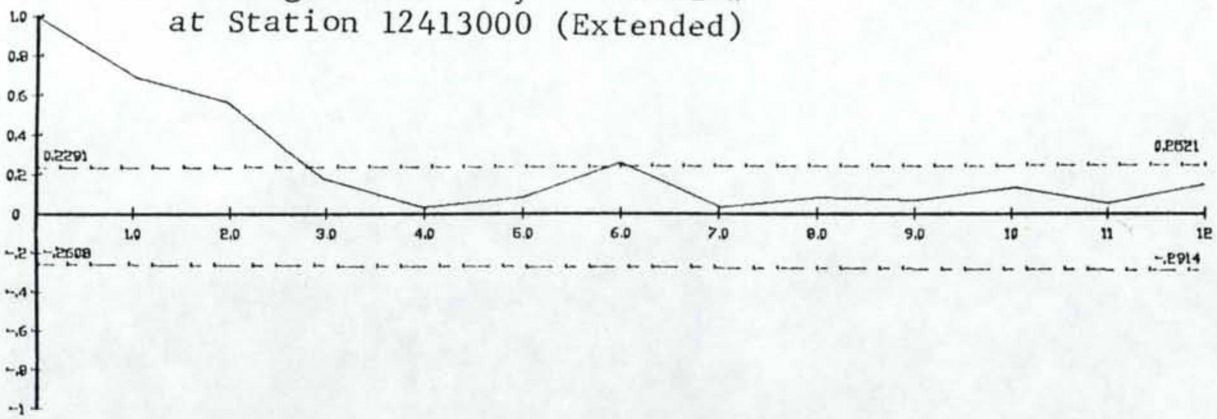


Figure C.11

Correlogram for August Streamflow at Station 12413000 (Extended)

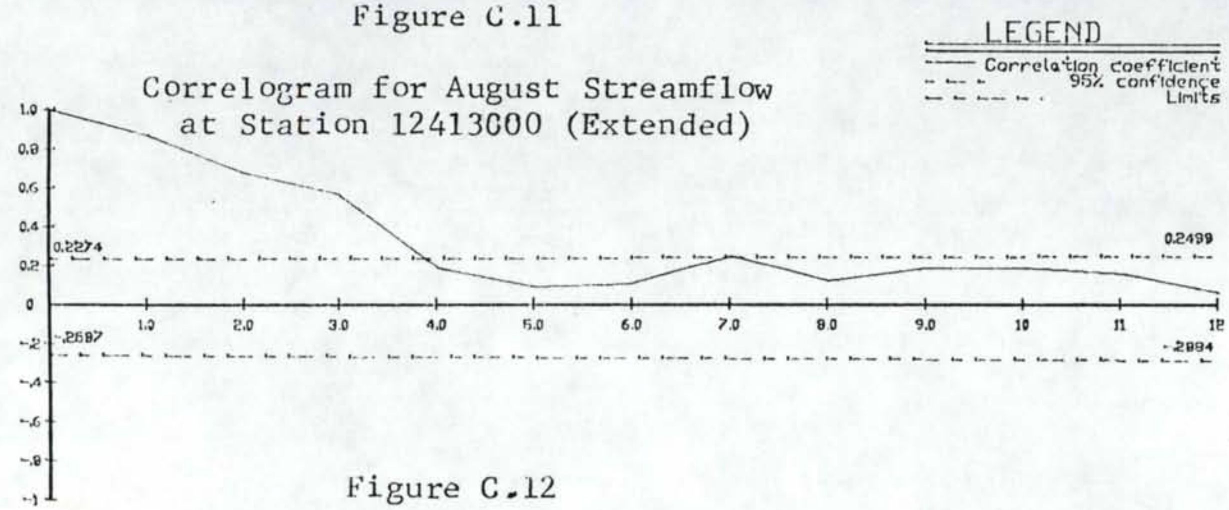
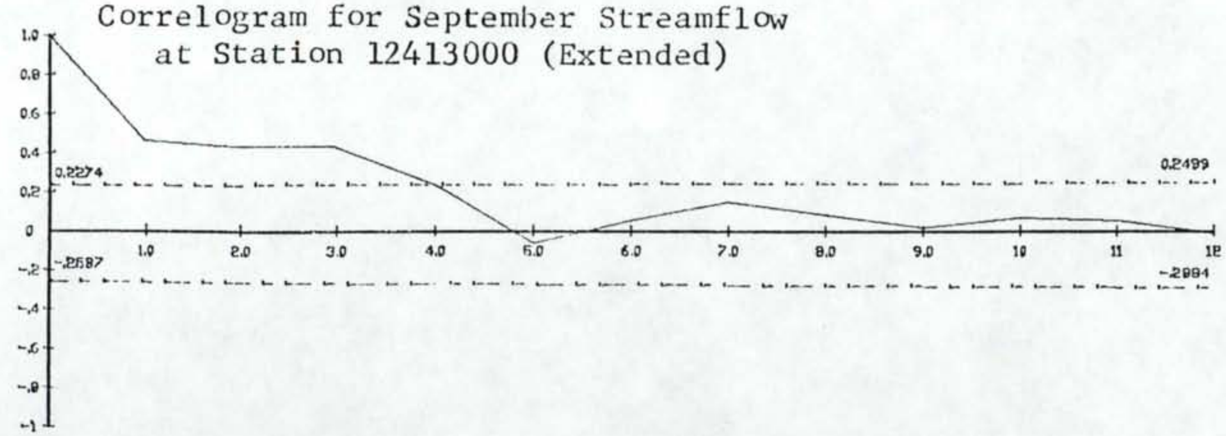


Figure C.12

Correlogram for September Streamflow at Station 12413000 (Extended)



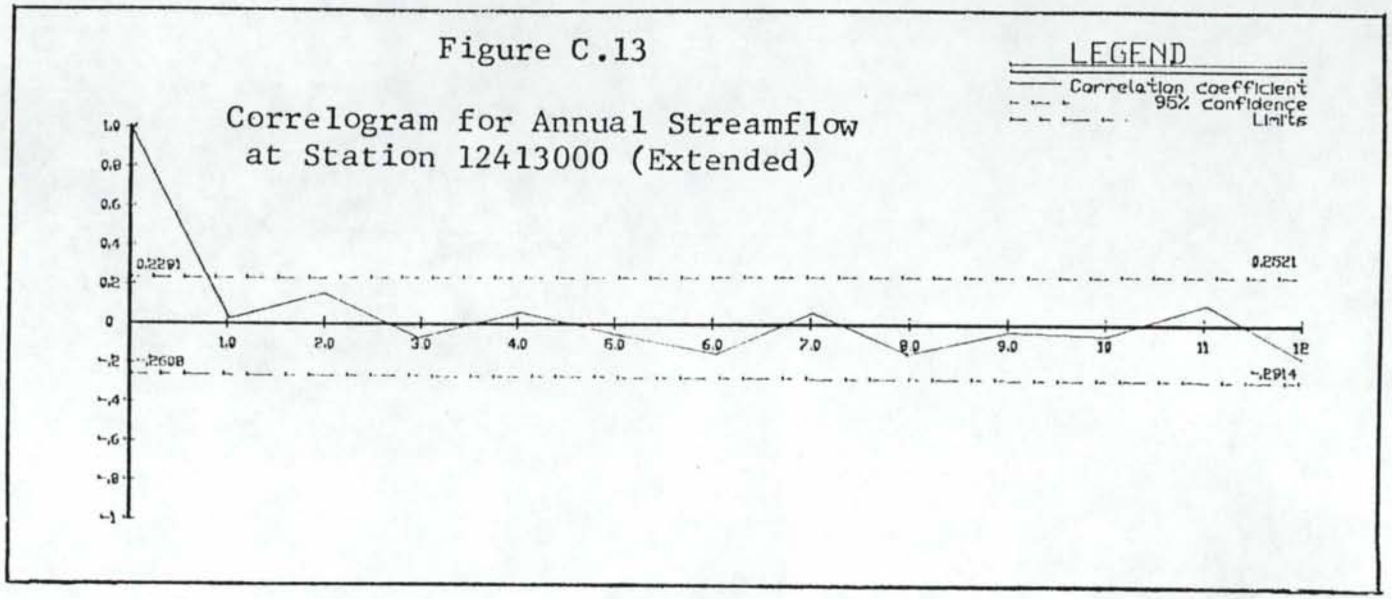


Figure C.13

Correlogram for Annual Streamflow at Station 12413000 (Extended)

Figure C.14

Correlogram for October Streamflow at Station 13186000 (Extended)

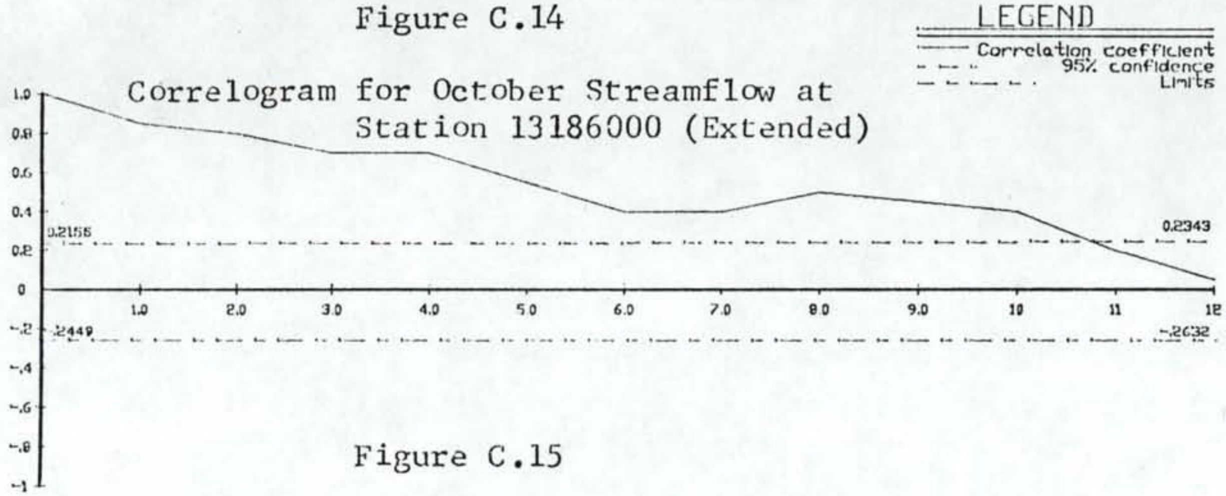


Figure C.15

Correlogram for November Streamflow at Station 13186000 (Extended)

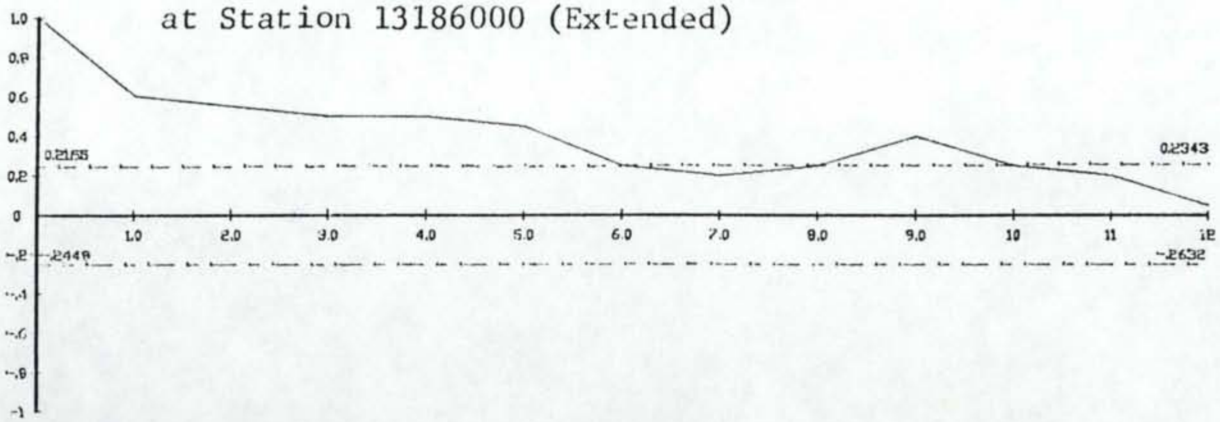


Figure C.16

Correlogram for December Streamflow at Station 13186000 (Extended)

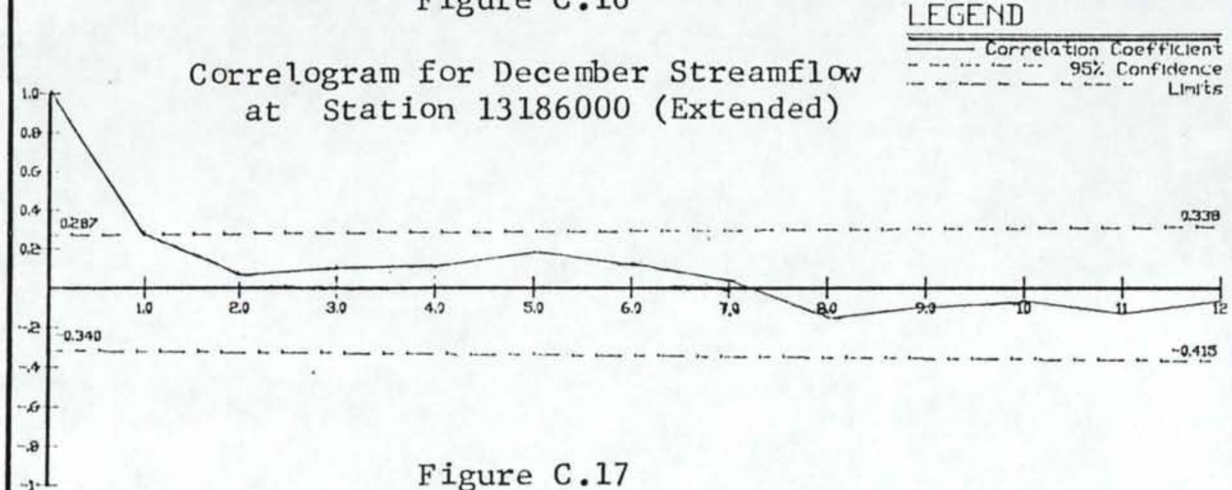


Figure C.17

Correlogram for January Streamflow at Station 13186000 (Extended)

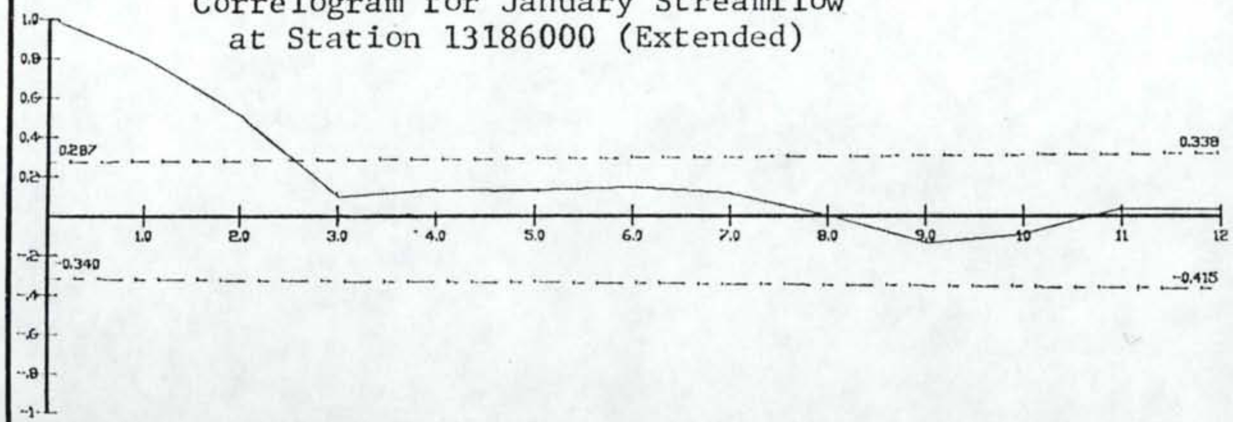


Figure C.18

Correlogram for February Streamflow at Station 13186000 (Extended)

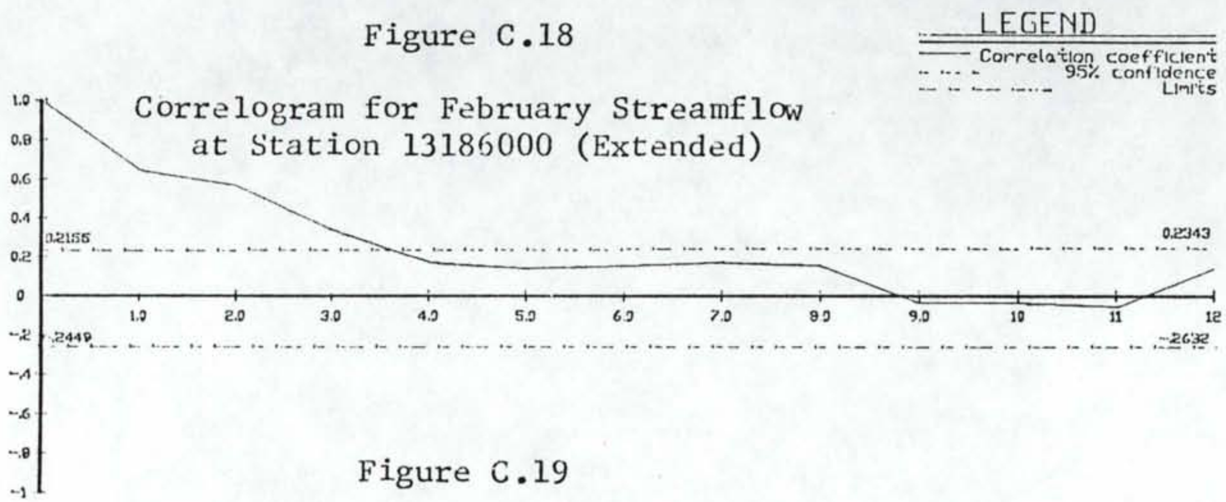


Figure C.19

Correlogram for March Streamflow at Station 13186000 (Extended)

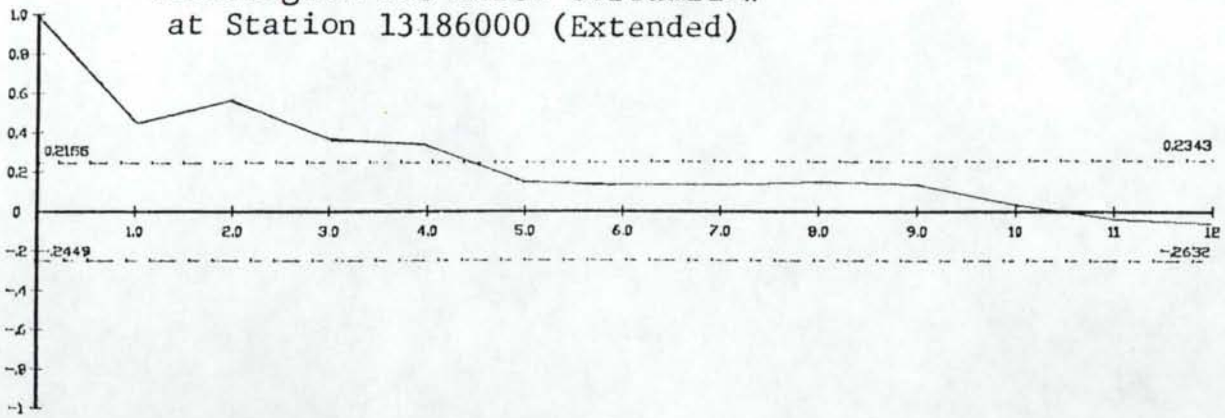


Figure C.20

Correlogram for April Streamflow at Station 13186000 (Extended)

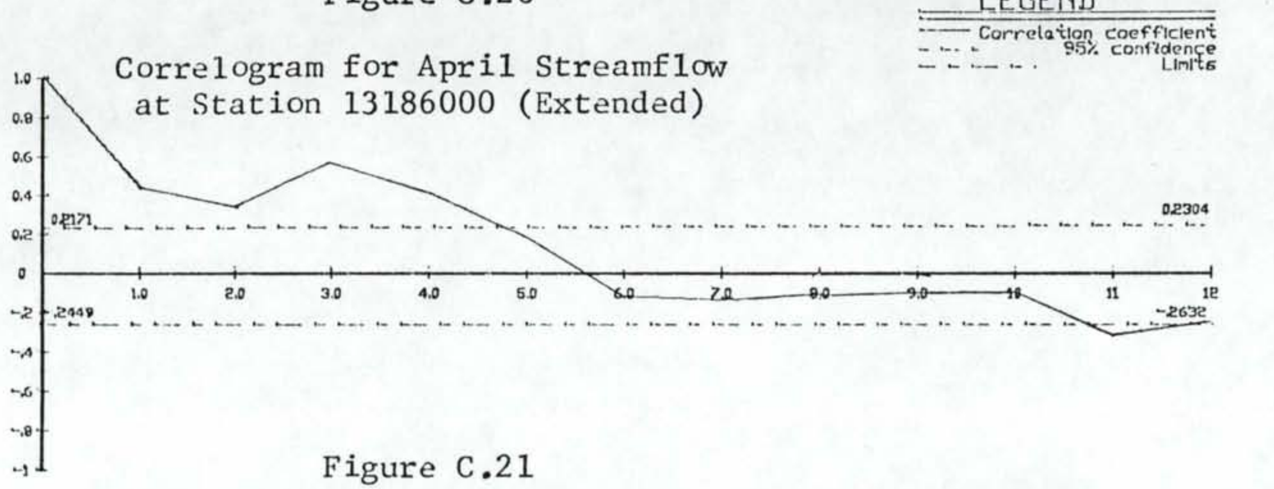


Figure C.21

Correlogram for May Streamflow at Station 13186000 (Extended)

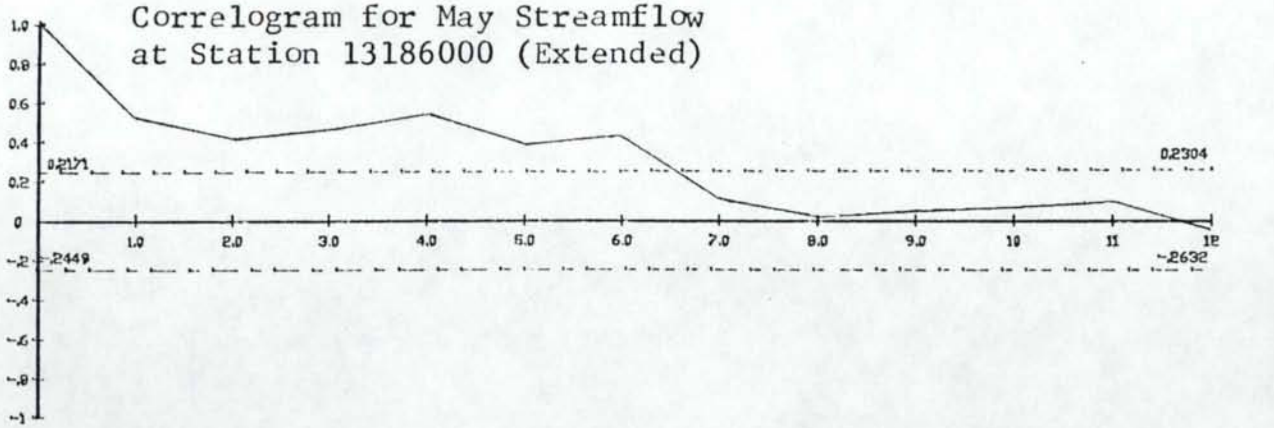


Figure C.22

Correlogram for June Streamflow at Station 13186000 (Extended)

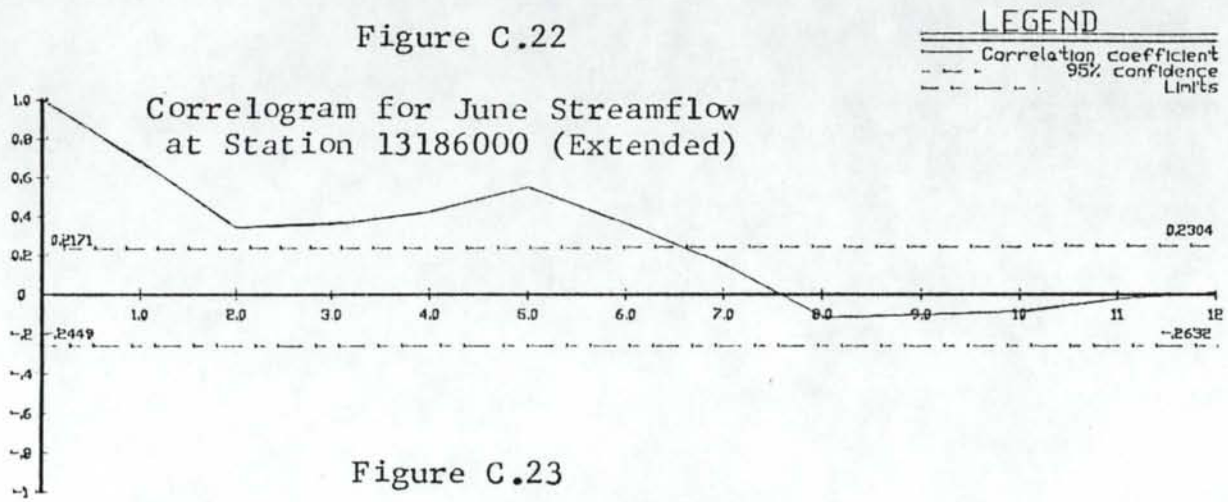


Figure C.23

Correlogram for July Streamflow at Station 13186000 (Extended)

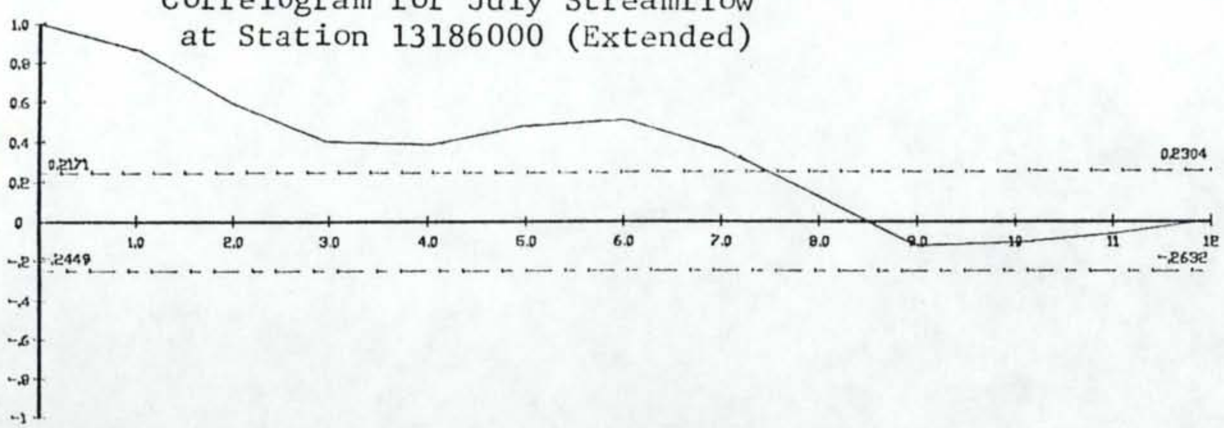


Figure C.24

Correlogram for August Streamflow at Station 13186000 (Extended)

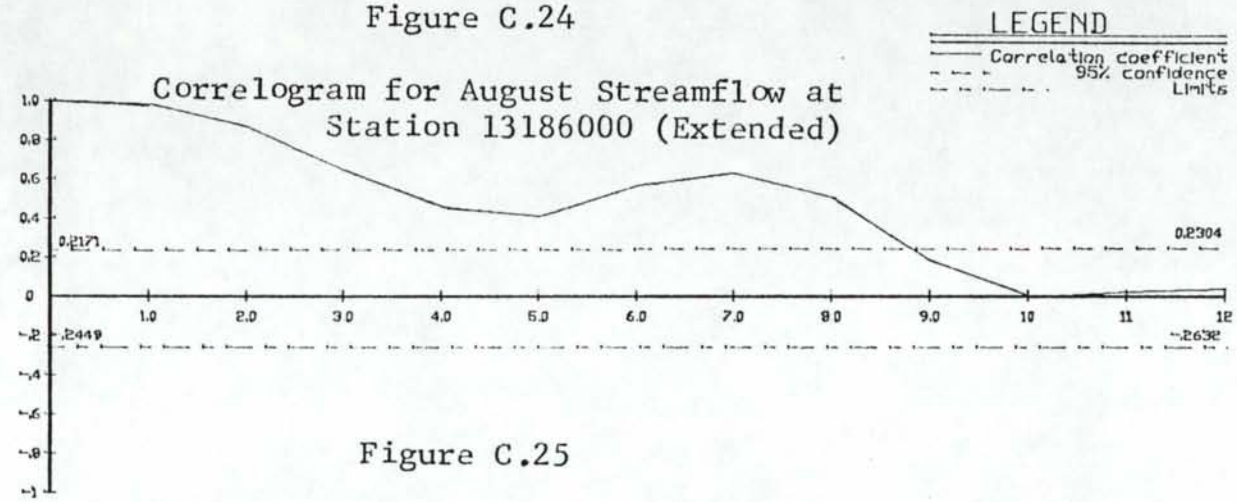


Figure C.25

Correlogram for September Streamflow at Station 13186000 (Extended)

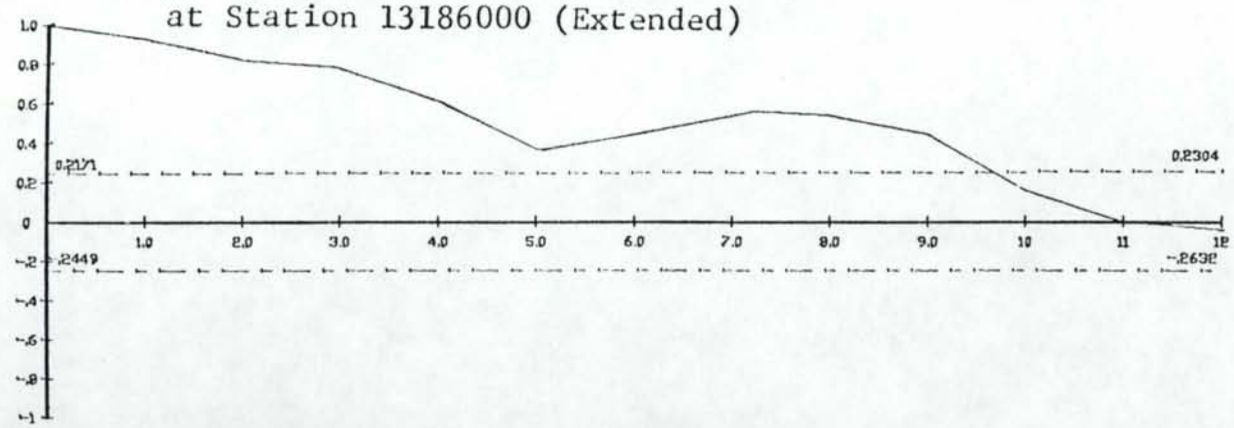
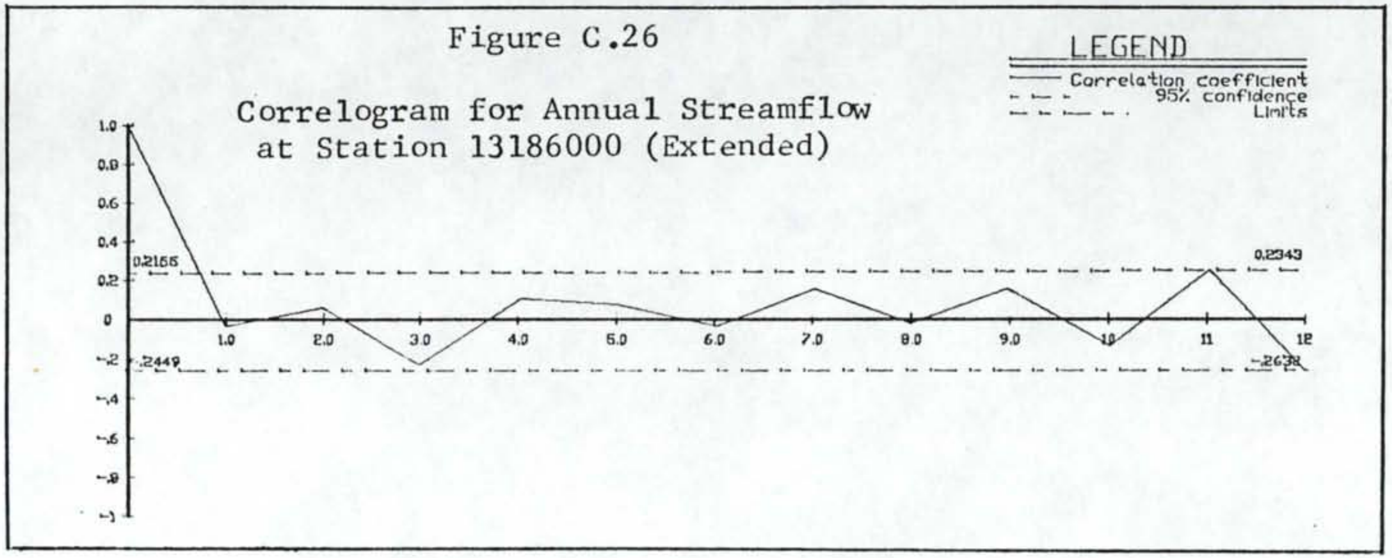


Figure C.26

Correlogram for Annual Streamflow at Station 13186000 (Extended)



APPENDIX D

DATA FOR ANNUAL MODELS

Table D.1

Partial Autocorrelation Coefficients
for Annual Series

k	12413000 Unextended	12413000 Extended	13186000 Unextended	13186000 Extended
1	-.0562	.0169	-.0482	-.0366
2	.2019	.1404	.0880	.0443
3	.0628	-.0917	-.2892	-.1179
4	.1298	.0496	-.0366	.0915
5	-.1314	-.0497	-.1747	.1047
6	-.3596	-.1496	-.1826	-.0526
7	-.0488	.0923	.1371	.1820
8	-.1215	-.1525	-.4115	-.0459
9	.0958	-.0743	.1095	.0906
10	.1641	.0050	-.5519	-.0906
11	.1648	.0626	-.0141	.2304
12	-.4103	-.1802	-.2427	-.3157

Table D.2

Sum of residuals squared for maximum likelihood estimate of θ_1 for unextended record at Station 12413000

θ_1	Sum e^2 X 10^{12}	θ_1	Sum e^2 X 10^{12}	θ_1	Sum e^2 X 10^{12}
-1.0	14.5012	-.10	1.7452	.030	1.698900
-0.9	8.4007	-.09	1.7387	.031	1.698854
-0.8	5.7908	-.08	1.7328	.032	1.698814
-0.7	4.1960	-.07	1.7273	.033	1.698778
-0.6	3.2280	-.06	1.7223	.034	1.698747
-0.5	2.6285	-.05	1.7178	.035	1.698721
-0.4	2.2462	-.04	1.7138	.036	1.698699
-0.3	1.9983	-.03	1.7102	.037	1.698682
-0.2	1.8396	-.02	1.7072	.038	1.698670
-0.1	1.7452	-.01	1.7046	.039	1.698663
0.0	* 1.7024	-.00	1.7024	.040	* 1.698660
0.1	1.7073	.01	1.7008	.041	1.698662
0.2	1.7627	.02	1.6996	.042	1.698669
0.3	1.8790	.03	1.6989	.043	1.698680
0.4	2.0762	.04	* 1.6987	.044	1.698696
0.5	2.3891	.05	1.6989	.045	1.698717
0.6	2.8804	.06	1.6996	.046	1.698743
0.7	3.6706	.07	1.7008	.047	1.698774
0.8	5.0189	.08	1.7025	.048	1.698809
0.9	7.5276	.09	1.7046	.049	1.698849
1.0	10.0878	.10	1.7073	.050	1.698923

* Minimum

Table D.3

Sum of residuals squared for maximum likelihood estimate of θ_1 for extended record at Station 12413000

θ_1	Sum e^2 X 10^{12}	θ_1	Sum e^2 X 10^{12}	θ_1	Sum e^2 X 10^{12}
-1.0	21.3243	-.10	2.6389	-.020	2.613158
-0.9	12.5554	-.09	2.6332	-.019	2.613115
-0.8	8.2445	-.08	2.6282	-.018	2.613078
-0.7	5.9302	-.07	2.6240	-.017	2.613049
-0.6	4.5807	-.06	2.6204	-.016	2.613026
-0.5	3.7616	-.05	2.6176	-.015	2.613010
-0.4	3.2506	-.04	2.6154	-.014	2.613001
-0.3	2.9308	-.03	2.6140	-.013	* 2.612998
-0.2	2.7386	-.02	2.6132	-.012	2.613003
-0.1	2.6389	-.01	* 2.6130	-.011	2.613013
0.0	* 2.6136	.00	2.6136	-.010	2.613031
0.1	2.6557	.01	2.6148	-.009	2.613055
0.2	2.7671	.02	2.6167	-.008	2.613086
0.3	2.9606	.03	2.6192	-.007	2.613124
0.4	3.2592	.04	2.6224	-.006	2.613168
0.5	3.7077	.05	2.6263	-.005	2.613219
0.6	4.3897	.06	2.6308	-.004	2.613277
0.7	5.4803	.07	2.6360	-.003	2.613341
0.8	7.4327	.08	2.6419	-.002	2.613412
0.9	11.7552	.09	2.6485	-.001	2.613490
1.0	22.6305	.10	2.6557	.000	2.613546

* Minimum

Table D.4

Sum of residuals squared for maximum likelihood estimate of θ_1 for unextended record at Station 13186000

θ_1	Sum e^2 X 10^{11}	θ_1	Sum e^2 X 10^{11}	θ_1	Sum e^2 X 10^{11}
-1.0	26.2485	0.00	3.2721	.030	3.266197
-0.9	14.9384	0.01	3.2694	.031	3.266105
-0.8	11.2336	0.02	3.2675	.032	3.266019
-0.7	8.2654	0.03	3.2662	.033	3.265941
-0.6	6.2813	0.04	* 3.2656	.034	3.265869
-0.5	5.0374	0.05	* 3.2656	.035	3.265803
-0.4	4.2658	0.06	3.2663	.036	3.265744
-0.3	3.7888	0.07	3.2676	.037	3.265692
-0.2	3.5008	0.08	3.2695	.038	3.265646
-0.1	3.3404	0.09	3.2720	.039	3.265607
0.0	* 3.2721	0.10	3.2751	.040	3.265574
0.1	3.2751	0.11	3.2789	.041	3.265548
0.2	3.3377	0.12	3.2832	.042	3.265528
0.3	3.4538	0.13	3.2880	.043	3.265515
0.4	3.6216	0.14	3.2935	.044	* 3.265508
0.5	3.8435	0.15	3.2995	.045	* 3.265508
0.6	4.1253	0.16	3.3060	.046	3.265514
0.7	4.4682	0.17	3.3132	.047	3.265527
0.8	4.8437	0.18	3.3208	.048	3.265546
0.9	5.2264	0.19	3.3290	.049	3.265571
1.0	7.2222	0.20	3.3378	.050	3.265603

* Minimum

Table D.5

Sum of residuals squared for maximum likelihood estimate of θ_1 for extended record at Station 13186000

θ_1	Sum e^2 X 10^{11}	θ_1	Sum e^2 X 10^{11}	θ_1	Sum e^2 X 10^{11}
-1.0	39.0256	-.10	6.6269	.020	6.503799
-0.9	22.8440	-.09	6.6085	.021	6.503612
-0.8	18.3584	-.08	6.5917	.022	6.503439
-0.7	14.2487	-.07	6.5765	.023	6.503278
-0.6	11.3307	-.06	6.5627	.024	6.503131
-0.5	9.4181	-.05	6.5505	.025	6.502997
-0.4	8.1848	-.04	6.5396	.026	6.502876
-0.3	7.3956	-.03	6.5302	.027	6.502768
-0.2	6.9052	-.02	6.5222	.028	6.502673
-0.1	6.6269	-.01	6.5156	.029	6.502592
0.0	* 6.5103	0.00	6.5103	.030	6.502523
0.1	6.5298	0.01	6.5064	.031	6.502468
0.2	6.6796	0.02	6.5038	.032	6.502425
0.3	6.9753	0.03	* 6.5025	.033	6.502396
0.4	7.4650	0.04	6.5026	.034	6.502379
0.5	8.2596	0.05	6.5039	.035	* 6.502376
0.6	9.6150	0.06	6.5065	.036	6.502386
0.7	12.1804	0.07	6.5104	.037	6.502409
0.8	17.9377	0.08	6.5156	.038	6.502444
0.9	35.0408	0.09	6.5221	.039	6.502493
1.0	169.2602	0.10	6.5298	.040	6.502555

* Minimum

Table D.6

Sum of Residuals squared (X 10^{12}) for Maximum Likelihood Estimates of ϕ_1 and θ_1 at Station 12413000 (Unextended)

$\phi_1 = .713 \quad \theta_1 = .739$

ϕ_1						
0.708	1.530600	1.530618	1.530643	1.530675	1.530715	1.530761
0.709	1.530573	1.530586	1.530605	1.530631	1.530665	1.530706
0.710	1.530555	1.530561	1.530575	1.530595	1.530623	1.530658
0.711	1.530544	1.530545	1.530553	1.530568	1.530590	1.530619
0.712	1.530542	1.530537	1.530539	1.530548	1.530565	1.530588
0.713	1.530548	1.530537	1.530534	1.530537	1.530548	1.530566
0.714	1.530562	1.530545	1.530536	1.530534	1.530539	1.530551
0.715	1.530584	1.530562	1.530547	1.530540	1.530539	1.530545
0.716	1.530614	1.530587	1.530567	1.530553	1.530547	1.530548
0.717	1.530652	1.530620	1.530594	1.530575	1.530563	1.530558
0.718	1.530699	1.530661	1.530629	1.530605	1.530588	1.530577
	0.737	0.738	0.739	0.740	0.741	0.742
				θ_1		

Table D.7

Sum of Residuals squared (X 10^{12}) for Maximum Likelihood Estimates of ϕ_1 and θ_1 at Station 12413000 (Extended)

$\phi_1 = -.741 \quad \theta_1 = -.722$

ϕ_1						
-.745	2.493314	2.493335	2.493366	2.493406	2.493455	2.493513
-.744	2.493290	2.493302	2.493323	2.493354	2.493393	2.493442
-.743	2.493278	2.493281	2.493293	2.493314	2.493344	2.493384
-.742	2.493279	2.493272	2.493275	2.493287	2.493308	2.493338
-.741	2.493293	2.493277	2.493270	2.493272	2.493284	2.493305
-.740	2.493320	2.493294	2.493277	2.493270	2.493272	2.493284
-.739	2.493359	2.493323	2.493297	2.493281	2.493273	2.493275
-.738	2.493411	2.493366	2.493330	2.493304	2.493287	2.493280
-.737	2.493475	2.493421	2.493375	2.493339	2.493313	2.493296
-.736	2.493553	2.493488	2.493433	2.493388	2.493352	2.493325
-.735	2.493643	2.493568	2.493504	2.493449	2.493403	2.493367
	-.724	-.723	-.722	-.721	-.720	-.719
				θ_1		

Table D.8

Sum of Residuals squared (X 10^{11}) for Maximum Likelihood Estimates of ϕ_1 and θ_1 at Station 13186000 (Unextended)

$\phi_1 = .722 \quad \theta_1 = .965$

ϕ_1						
0.715	2.938322	2.938228	2.938208	2.938265	2.938402	2.938624
0.716	2.938294	2.938184	2.938148	2.938188	2.938308	2.938512
0.717	2.938277	2.938152	2.938099	2.938122	2.938225	2.938411
0.718	2.938272	2.938131	2.938062	2.938069	2.938154	2.938323
0.719	2.938278	2.938121	2.938036	2.938027	2.938095	2.938246
0.720	2.938296	2.938124	2.938023	2.937996	2.938048	2.938181
0.721	2.938326	2.938138	2.938020	2.937977	2.938012	2.938127
0.722	2.938367	2.938163	2.938030	2.937970	2.937988	2.938086
0.723	2.938419	2.938200	2.938051	2.937975	2.937975	2.938056
0.724	2.938483	2.938248	2.938083	2.937991	2.937975	2.938038
0.725	2.938559	2.938309	2.938127	2.938019	2.937986	2.938032
	0.962	0.963	0.964	0.965	0.966	0.967

θ_1

Table D.9

Sum of Residuals squared (X 10^{11}) for Maximum Likelihood Estimates of ϕ_1 and θ_1 at Station 13186000 (Extended)

$\phi_1 = -.840 \quad \theta_1 = -.960$

ϕ_1						
-.845	6.394842	6.394544	6.394326	6.394181	6.394103	6.394087
-.844	6.394558	6.394280	6.394080	6.393952	6.393890	6.393887
-.843	6.394325	6.394066	6.393884	6.393773	6.393726	6.393737
-.842	6.394143	6.393903	6.393739	6.393644	6.393612	6.393637
-.841	6.394011	6.393790	6.393644	6.393565	6.393548	6.393586
-.840	6.393929	6.393728	6.393599	6.393536	6.393534	6.393586
-.839	6.393898	6.393716	6.393605	6.393558	6.393569	6.393635
-.838	6.393918	6.393754	6.393660	6.393629	6.393655	6.393734
-.837	6.393988	6.393843	6.393766	6.393750	6.393791	6.393882
-.836	6.394109	6.393982	6.393922	6.393922	6.393977	6.394081
-.835	6.394280	6.394172	6.394128	6.394143	6.394212	6.394329
	-.964	-.963	-.962	-.961	-.960	-.959

θ_1

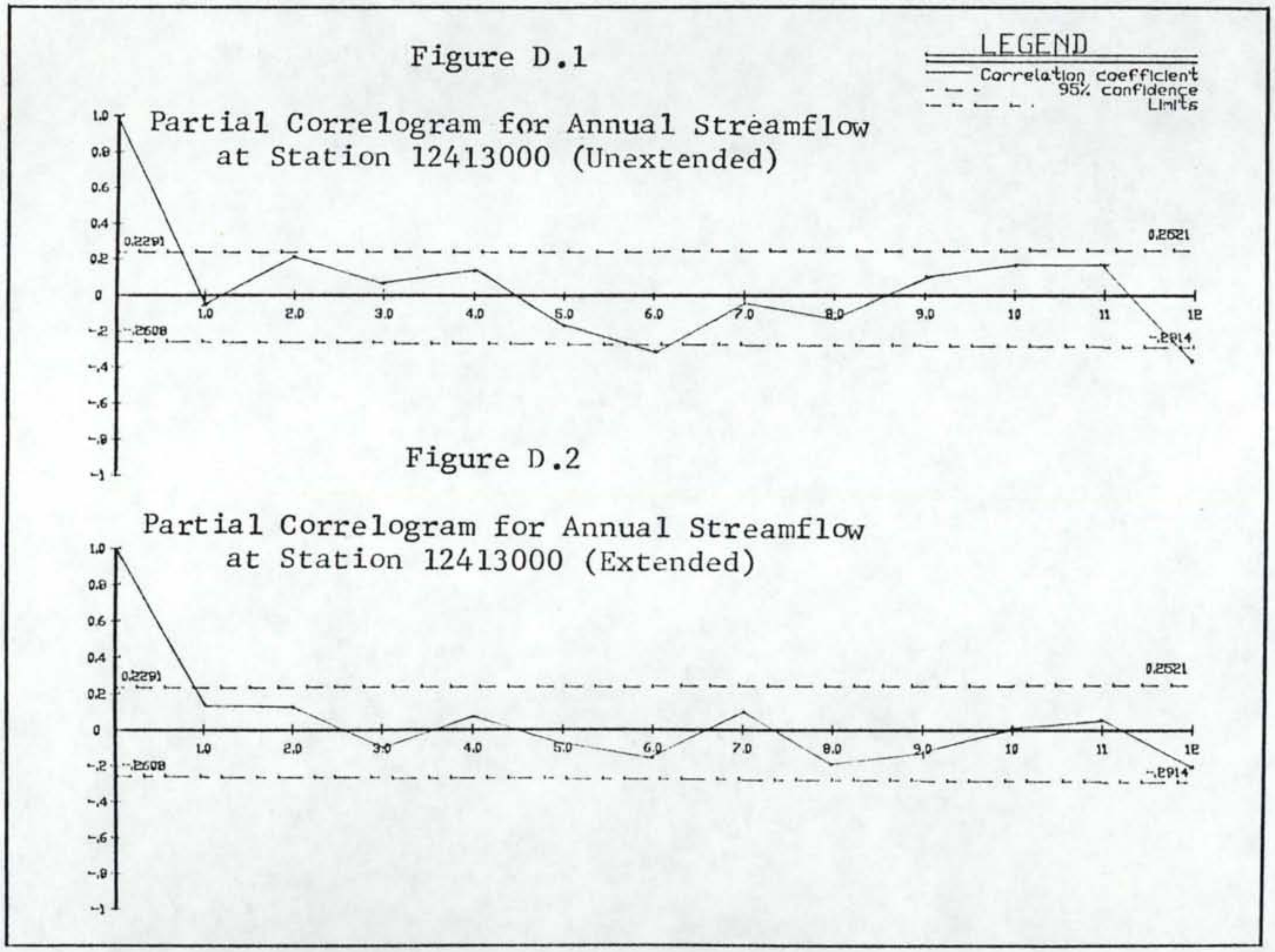


Figure D.1

Partial Correlogram for Annual Streamflow at Station 12413000 (Unextended)

Figure D.2

Partial Correlogram for Annual Streamflow at Station 12413000 (Extended)

Figure D.3

Partial Correlogram for Annual Streamflow at Station 13186000 (Unextended)

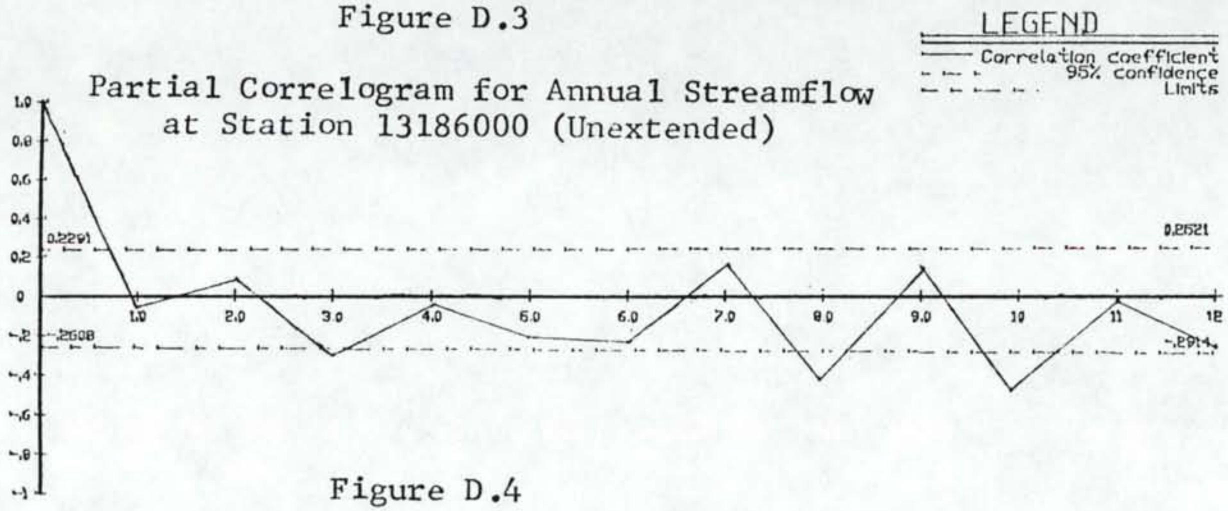
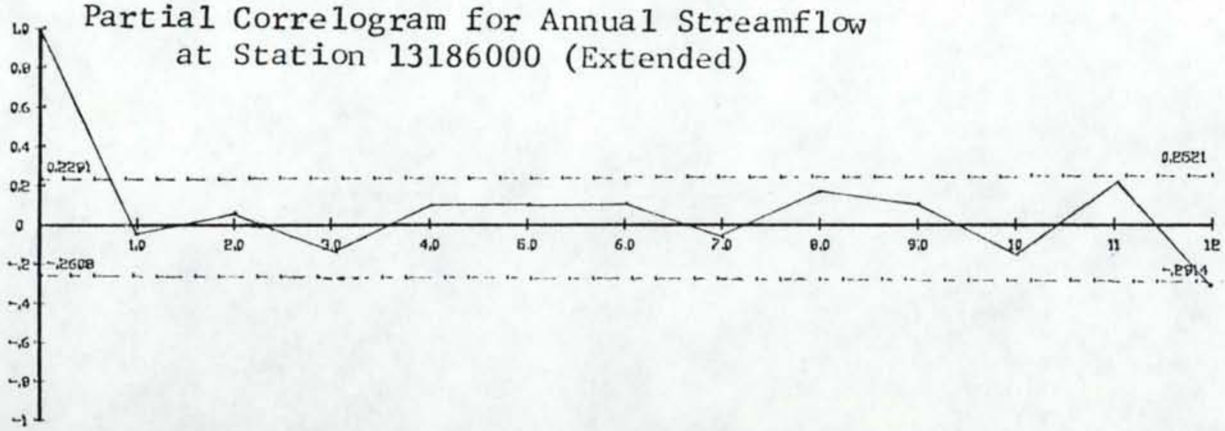


Figure D.4

Partial Correlogram for Annual Streamflow at Station 13186000 (Extended)



APPENDIX E

OUTPUT FROM DROUGHT ANALYSIS PROGRAM
DATA FOR ANNUAL MODELS

Table E.1

Experimental and Theoretical Cumulative Density Functions of the Maximum Run-Length for a Truncaton Level of .35 and a Sample Size of 100 years

Length years	12413000 Unextended	12413000 Extended	13186000 Unextended	13186000 Extended	Theor-etical
1	.0000	.0000	.0000	.0000	.0001
2	.0525	.0400	.0425	.0450	.0497
3	.3725	.3675	.3800	.3500	.3741
4	.7350	.6850	.7000	.6875	.7174
5	.9025	.8700	.8850	.8725	.8924
6	.9725	.9525	.9700	.9675	.9615
7	.9900	.9825	.9850	.9925	.9865
8	.9950	.9950	.9975	.9950	.9953
9	1.0000	1.0000	1.0000	.9975	.9984
10	1.0000	1.0000	1.0000	1.0000	.9994

Table E.2

Experimental and Theoretical Cumulative Density Functions of the Maximum Run-Length for a Truncaton Level of .35 and a Sample Size of 50 years

Length years	12413000 Unextended	12413000 Extended	13186000 Unextended	13186000 Extended	Theor- etical
1	.0050	.0125	.0012	.0125	.0086
2	.2175	.2075	.2188	.2175	.2327
3	.6262	.5950	.6100	.6075	.6230
4	.8625	.8350	.8338	.8375	.8537
5	.9512	.9375	.9450	.9362	.9478
6	.9862	.9800	.9862	.9838	.9819
7	.9950	.9912	.9925	.9962	.9938
8	.9975	.9975	.9988	.9975	.9979
9	1.0000	1.0000	1.0000	.9988	.9993
10	1.0000	1.0000	1.0000	1.0000	.9998

Table E.3

Experimental and Theoretical Cumulative Density Functions of the Maximum Run-Length for a Truncaton Level of .35 and a Sample Size of 25 years

Length years	12413000 Unextended	12413000 Extended	13186000 Unextended	13186000 Extended	Theor- etical
1	.0856	.0900	.0806	.0862	.1035
2	.4831	.4856	.4875	.4881	.5036
3	.7988	.7881	.7956	.7919	.8040
4	.9306	.9175	.9200	.9194	.9313
5	.9775	.9688	.9738	.9694	.9768
6	.9938	.9900	.9938	.9931	.9923
7	.9975	.9962	.9969	.9988	.9974
8	.9988	.9988	.9994	1.0000	.9992
9	1.0000	1.0000	1.0000	1.0000	.9997

Table E.4

Experimental and Theoretical Cumulative Density Functions of the Maximum Run-Length for a Truncaton Level of .50 and a Sample Size of 100 years

Length years	12413000 Unextended	12413000 Extended	13186000 Unextended	13186000 Extended	Theor-etical
1	.0000	.0000	.0000	.0000	.0000
2	.0000	.0000	.0000	.0000	.0004
3	.0225	.0225	.0300	.0300	.0300
4	.1550	.1525	.1575	.1775	.1950
5	.4525	.3975	.4375	.3975	.4584
6	.6950	.6250	.7025	.6625	.6854
7	.8200	.8000	.8450	.8350	.8315
8	.9000	.8750	.9050	.9025	.9134
9	.9550	.9425	.9625	.9550	.9563
10	.9700	.9675	.9775	.9725	.9782
11	.9875	.9850	.9875	.9900	.9892
12	.9900	.9975	.9975	.9950	.9946
13	.9950	.9975	.9975	1.0000	.9974
14	1.0000	.9975	.9975	1.0000	.9987
15	1.0000	1.0000	.9975	1.0000	.9994
16	1.0000	1.0000	.9975	1.0000	.9974
17	1.0000	1.0000	1.0000	1.0000	.9987

Table E.5

Experimental and Theoretical Cumulative Density Functions of the Maximum Run-Length for a Truncaton Level of .50 and a Sample Size of 50 years

Length years	12413000 Unextended	12413000 Extended	13186000 Unextended	13186000 Extended	Theor-etical
1	.0000	.0000	.0000	.0000	.0002
2	.0112	.0125	.0150	.0138	.0232
3	.1638	.1575	.1662	.1600	.1850
4	.4300	.4125	.4188	.4288	.4584
5	.6825	.6412	.6812	.6500	.6918
6	.8388	.8050	.8462	.8262	.8382
7	.9112	.9000	.9262	.9188	.9183
8	.9500	.9438	.9575	.9525	.9595
9	.9788	.9750	.9850	.9775	.9801
10	.9850	.9850	.9925	.9875	.9903
11	.9938	.9938	.9950	.9950	.9952
12	.9962	1.0000	.9988	.9975	.9977
13	.9988	1.0000	.9988	1.0000	.9989
14	1.0000	1.0000	.9988	1.0000	.9994
15	1.0000	1.0000	.9988	1.0000	.9997
16	1.0000	1.0000	.9988	1.0000	.9999
17	1.0000	1.0000	1.0000	1.0000	.9999

Table E.6

Experimental and Theoretical Cumulative Density Functions of the Maximum Run-Length for a Truncaton Level of .50 and a Sample Size of 25 years

Length years	12413000 Unextended	12413000 Extended	13186000 Unextended	13186000 Extended	Theor-etical
1	.0031	.0056	.0050	.0062	.0175
2	.1475	.1494	.1350	.1238	.1716
3	.4469	.4281	.4225	.4262	.4591
4	.6888	.6688	.6831	.6788	.7028
5	.8431	.8175	.8450	.8206	.8498
6	.9250	.9056	.9275	.9175	.9269
7	.9619	.9544	.9662	.9638	.9650
8	.9812	.9756	.9819	.9788	.9834
9	.9919	.9912	.9950	.9900	.9922
10	.9938	.9950	.9975	.9944	.9963
11	.9975	.9975	.9981	.9981	.9983
12	.9988	1.0000	.9994	.9988	.9992
13	1.0000	1.0000	.9994	1.0000	.9996
14	1.0000	1.0000	.9994	1.0000	.9998
15	1.0000	1.0000	.9994	1.0000	.9999
16	1.0000	1.0000	.9994	1.0000	1.0000
17	1.0000	1.0000	1.0000	1.0000	1.0000

Table E.7

Experimental Cumulative Density Functions
of the Maximum Run-Sum for the
Unextended Record at Station 12413000

Deficit	q(0) = .35			q(0) = .50		
cfsdX10^3	n=100	n=50	n=25	n=100	n=50	n=25
100	.0000	.0000	.0044	.0000	.0000	.0000
200	.0000	.0038	.0556	.0000	.0000	.0006
300	.0000	.0512	.2394	.0000	.0000	.0294
400	.0450	.2150	.4756	.0000	.0075	.1094
500	.1675	.4325	.6781	.0000	.0512	.2444
600	.4450	.6600	.8212	.0100	.1588	.4125
700	.6475	.7988	.8969	.0475	.2750	.5488
800	.8050	.8988	.9506	.1600	.4250	.6750
900	.8825	.9425	.9731	.2950	.5500	.7656
1000	.9425	.9712	.9862	.4425	.6788	.8400
1100	.9700	.9850	.9925	.5900	.7812	.8944
1200	.9850	.9925	.9962	.7250	.8525	.9294
1300	.9900	.9950	.9975	.7850	.8875	.9475
1400	.9925	.9962	.9981	.8475	.9200	.9638
1500	.9975	.9988	.9994	.8900	.9425	.9750
1600	.9975	.9988	.9994	.9225	.9600	.9831
1700	.9975	.9988	.9994	.9450	.9725	.9881
1800	1.0000	1.0000	1.0000	.9650	.9825	.9925
1900	1.0000	1.0000	1.0000	.9750	.9875	.9950
2000	1.0000	1.0000	1.0000	.9875	.9938	.9969
2100	1.0000	1.0000	1.0000	.9900	.9950	.9975
2200	1.0000	1.0000	1.0000	.9975	.9988	.9994
2300	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table E.8

Experimental Cumulative Density Functions
of the Maximum Run-Sum for the
Extended Record at Station 12413000

Deficit	q(0) = .35			q(0) = .50		
cfsdX10³	n=100	n=50	n=25	n=100	n=50	n=25
100	.0000	.0000	.0062	.0000	.0000	.0000
200	.0000	.0050	.0650	.0000	.0000	.0025
300	.0025	.0438	.2131	.0000	.0012	.0306
400	.0225	.1800	.4312	.0000	.0100	.1131
500	.1475	.4000	.6475	.0000	.0512	.2381
600	.3250	.5950	.7844	.0075	.1262	.3706
700	.5375	.7475	.8681	.0500	.2562	.5169
800	.7200	.8625	.9294	.1325	.3775	.6269
900	.8350	.9212	.9606	.2550	.5234	.7306
1000	.8900	.9475	.9738	.3800	.6325	.8044
1100	.9425	.9712	.9862	.5025	.7262	.8619
1200	.9625	.9812	.9912	.6325	.8075	.9056
1300	.9825	.9925	.9962	.7125	.8538	.9294
1400	.9900	.9962	.9981	.7675	.8812	.9419
1500	.9950	.9988	.9994	.8300	.9138	.9600
1600	.9950	.9988	.9994	.8700	.9375	.9719
1700	.9975	1.0000	1.0000	.9050	.9538	.9800
1800	1.0000	1.0000	1.0000	.9275	.9675	.9856
1900	1.0000	1.0000	1.0000	.9475	.9762	.9900
2000	1.0000	1.0000	1.0000	.9650	.9850	.9931
2100	1.0000	1.0000	1.0000	.9800	.9925	.9969
2200	1.0000	1.0000	1.0000	.9825	.9938	.9975
2300	1.0000	1.0000	1.0000	.9875	.9962	.9988
2400	1.0000	1.0000	1.0000	.9900	.9962	.9988
2500	1.0000	1.0000	1.0000	.9925	.9975	.9994
2600	1.0000	1.0000	1.0000	.9950	.9988	1.0000
2700	1.0000	1.0000	1.0000	.9950	.9988	1.0000
2800	1.0000	1.0000	1.0000	.9950	.9988	1.0000
2900	1.0000	1.0000	1.0000	.9975	.9988	1.0000
3000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table E.9

Experimental Cumulative Density Functions of the Maximum Run-Sum for the Unextended Record at Station 13186000

Deficit	q(0) = .35			q(0) = .50		
cfsdX10^3	n=100	n=50	n=25	n=100	n=50	n=25
50	.0000	.0000	.0056	.0000	.0000	.0000
100	.0000	.0050	.0756	.0000	.0000	.0044
150	.0075	.0788	.2750	.0000	.0000	.0306
200	.0675	.2612	.5131	.0000	.0162	.1381
250	.2725	.5188	.7219	.0025	.0725	.2894
300	.4750	.6988	.8362	.0275	.1900	.4494
350	.6750	.8275	.9131	.1200	.3550	.5969
400	.8275	.9112	.9550	.2175	.4825	.7094
450	.9025	.9525	.9775	.3625	.6200	.8038
500	.9450	.9750	.9888	.4900	.7162	.8538
550	.9800	.9912	.9962	.6250	.8025	.8994
600	.9875	.9950	.9969	.7225	.8575	.9300
650	.9975	.9988	.9994	.8250	.9150	.9606
700	.9975	.9988	.9994	.8700	.9388	.9731
750	.9975	.9988	.9994	.9250	.9650	.9844
800	1.0000	1.0000	1.0000	.9425	.9725	.9875
850	1.0000	1.0000	1.0000	.9650	.9825	.9925
900	1.0000	1.0000	1.0000	.9825	.9925	.9962
950	1.0000	1.0000	1.0000	.9925	.9962	.9981
1000	1.0000	1.0000	1.0000	.9925	.9962	.9981
1050	1.0000	1.0000	1.0000	.9950	.9975	.9988
1100	1.0000	1.0000	1.0000	.9975	.9988	.9994
1150	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table E.10

Experimental Cumulative Density Functions
of the Maximum Run-Sum for the
Extended Record at Station 13186000

Deficit	q(0) = .35			q(0) = .50		
cfsdX10^3	n=100	n=50	n=25	n=100	n=50	n=25
50	.0000	.0000	.0056	.0000	.0000	.0000
100	.0000	.0050	.0712	.0000	.0000	.0025
150	.0075	.0850	.2794	.0000	.0012	.0375
200	.0725	.2950	.5350	.0000	.0188	.1275
250	.2625	.5188	.7231	.0025	.0900	.2819
300	.4900	.7100	.8500	.0425	.2038	.4581
350	.7050	.8450	.9238	.1125	.3450	.6006
400	.8175	.9038	.9556	.2150	.4975	.7244
450	.8950	.9475	.9762	.3975	.6400	.8106
500	.9450	.9725	.9869	.5450	.7500	.8788
550	.9675	.9838	.9919	.6875	.8225	.9194
600	.9800	.9900	.9950	.7550	.8775	.9419
650	.9950	.9975	.9994	.8225	.9100	.9575
700	.9950	.9975	.9994	.8725	.9362	.9700
750	.9975	.9988	.9994	.9200	.9612	.9812
800	.9975	.9988	.9994	.9475	.9750	.9875
850	1.0000	1.0000	1.0000	.9625	.9838	.9919
900	1.0000	1.0000	1.0000	.9725	.9888	.9944
950	1.0000	1.0000	1.0000	.9825	.9925	.9969
1000	1.0000	1.0000	1.0000	.9850	.9938	.9975
1050	1.0000	1.0000	1.0000	.9875	.9938	.9975
1100	1.0000	1.0000	1.0000	.9950	.9975	.9988
1150	1.0000	1.0000	1.0000	.9950	.9975	.9988
1200	1.0000	1.0000	1.0000	.9975	.9988	.9994
1500	1.0000	1.0000	1.0000	.9975	.9988	.9994

Table E.11

Standard Deviations from Maximum Run-Sum Periods at Station 12413000 (Extended)

Mon	Ext Rec Std Dev	q(0) = .50 n=25	n=50	n=100	q(0) = .35 n=25	n=50	n=100
Oct	8681	6184	6157	6202	5587	5525	5687
Nov	25090	12201	12251	12114	9914	10222	10364
Dec	52424	24728	24935	24915	20349	19803	19321
Jan	39677	19739	19498	18725	16301	15890	15372
Feb	43327	29007	28321	28451	25756	24759	24630
Mar	42120	32251	32632	32616	30341	31340	32058
Apr	59668	50432	50478	49966	48718	48893	48794
May	70537	53552	53525	53928	51195	51634	52306
Jun	33609	19783	19875	20166	17655	18137	18255
Jul	7466	5485	5546	5628	5119	5180	5263
Aug	2798	2241	2247	2282	2146	2202	2236
Sep	2210	2020	2047	2045	1941	1961	1947

Table E.12

Standard Deviations from Maximum Run-Sum Periods at Station 13186000 (Extended)

Mon	Ext Rec Std Dev	q(0) = .50 n=25	n=50	n=100	q(0) = .35 n=25	n=50	n=100
Oct	1437	1496	1510	1535	1615	1671	1694
Nov	1683	1347	1365	1390	1416	1468	1448
Dec	2330	1999	2014	2024	1999	2023	2025
Jan	1523	1133	1143	1158	1192	1259	1244
Feb	1504	1275	1286	1276	1311	1368	1347
Mar	3673	3067	3051	3078	2975	3040	3052
Apr	18034	11784	11703	11575	10910	11047	10815
May	32532	23567	23875	23999	21827	22228	22340
Jun	36275	23433	23468	23631	20728	20985	20767
Jul	14635	7762	7765	7839	6653	6715	6670
Aug	3337	2205	2210	2276	2070	2128	2130
Sep	1839	1492	1505	1554	1475	1532	1554

Table E.13

Skew Coefficients from Maximum Run-Sum Periods at Station 12413000 (Extended)

	Ext Rec	q(0) = .50			q(0) = .35		
Mon	Coef Skew	n=25	n=50	n=100	n=25	n=50	n=100
Oct	2.248	1.610	1.586	1.629	1.710	1.681	1.869
Nov	2.562	1.900	1.815	1.779	1.941	2.192	2.267
Dec	2.696	2.228	2.569	2.501	2.010	1.864	1.784
Jan	2.385	1.338	1.411	1.219	1.273	1.443	1.504
Feb	1.490	2.104	1.964	2.057	2.165	1.930	2.000
Mar	1.405	0.881	0.872	0.900	0.945	0.949	1.054
Apr	0.339	0.076	0.071	0.020	0.137	0.069	0.031
May	-.118	-.007	-.012	-.013	0.169	0.152	0.136
Jun	1.180	0.991	0.989	0.971	0.918	1.046	0.825
Jul	0.723	0.650	0.633	0.728	0.405	0.361	0.340
Aug	0.567	0.243	0.230	0.306	0.047	-.006	-.055
Sep	1.602	1.153	1.169	1.127	0.938	0.825	0.604

Table E.14

Skew Coefficients from Maximum Run-Sum Periods at Station 13186000 (Extended)

	Ext Rec	q(0) = .50			q(0) = .35		
Mon	Coef Skew	n=25	n=50	n=100	n=25	n=50	n=100
Oct	0.360	-.100	-.130	-.155	-.179	-.198	-.106
Nov	2.104	0.107	0.080	0.096	-.388	-.534	-.463
Dec	3.169	0.201	0.184	0.117	0.071	-.008	0.030
Jan	1.746	-1.1400	-1.240	-1.3364	-1.662	-1.760	-1.714
Feb	1.354	-.393	-.496	-.708	-.852	-1.008	-.996
Mar	0.975	0.333	0.295	0.316	0.168	0.130	0.190
Apr	1.366	0.531	0.520	0.430	0.519	0.531	0.424
May	0.427	-.122	-.087	-.109	0.059	0.085	0.052
Jun	0.077	0.023	0.041	0.091	0.317	0.351	0.327
Jul	0.822	0.335	0.379	0.438	0.598	0.628	0.622
Aug	0.710	-.352	-.335	-.304	-.246	-.225	-.201
Sep	0.471	-.541	-.564	-.515	-.584	-.636	-.592

APPENDIX F: NOTATION

The following symbols are used in this paper:

A_L = parameter in Lawrance multivariate model
A_R = parameter in linear regression model
A_y = parameter in Yevjevich multivariate model
a = sample skew
a_L = lower limit parameter of confidence interal for standard deviation
a_U = upper limit parameter of confidence interal for standard deviation
B_L = parameter in Lawerance multivariate model
B_R = parameter in linear regression model
B_y = parameter in Yevjevich multivariate model
c = constant of 3-parameter lognormal distribution (cfs days)
CV = coefficient of variation
D = standardized Run-Sum (Deficit)
D_m = standardized Median Run-Sum (Deficit)
d_{11} = statistics used in maximum likelihood estimates of AR(1) parameters
d_{12} = statistics used in maximum likelihood estimates of AR(1) parameters
d_{22} = statistics used in maximum likelihood estimates of AR(1) parameters
E = standardized residual series
e = residual series
F = F-statistic
f_e = expected frequency
f_o = observed frequency
G = parameter in Lane's disaggregation model
g = sample coefficient of skew
H = parameter in Lane's disaggregation model
h = Hurst coefficient
J = standardized, normalized monthly streamflow value
j = normalized monthly streamflow value
$\bar{j}$ = mean of normalized series
K = number of class intervals used in performing ^{2} test
k = time lag
L = run-length (years)
L_m = median maximum run-length (years)
m = subseries length
n = sample size
n_o = number of temporally overlapping streamflow values between two records
n_p = number of estimated parameters
p = value of AR(p) component of an ARMA(p,q) model
Q = parameter in Lane's disaggregation model
q = value of MA(q)component of an ARMA(p,q) model
q(0) = quantile corresponding to $P(y_t \leq y(0))$

$p(0)$ = quantile corresponding to $P(y_t > y(0))$
R = range
R_n = rescaled range for sample size n
$r(k)$, r_k = sample correlation coefficient of lag k
S = run-sum (cfs days)
S_m = maximum median run-sum (cfs days)
$S(\theta_1)$ = sum of residuals squared from an MA(1) model
$S(\phi_1, \theta_1)$ = sum of residuals squared from an ARMA(p,q) model
$\bar{S}$ = average annual deficit (cfs days)
Sp = spooled variance (cfs days)
s = sample standard deviation (cfs days)
s^2 = sample variance (cfs days)
T = return period (years)
t = t-statistic
W = standardized monthly streamflow value from key station
w = monthly streamflow value at key station (cfs days)
w = mean monthly streamflow value at key station (cfs days)
X = standardized monthly streamflow value from subordinate station
x = raw monthly streamflow value (cfs days)
$\bar{x}$ = mean monthly flow (cfs days)
Y = standardized annual streamflow value
y = annual streamflow value (cfs days)
$\bar{y}$ = mean of annual values (cfs days)
$y(0)$ = truncation level
Z = standardized variable

α = population skew
$\alpha_{.05}$ = significance level
β = parameter of Weibull distribution
δ = index of parameter parsimony
ϵ = parameter of Weibull distribution
θ_j = moving average coefficient of order j
λ = standard random deviate
λ_g = standard gamma deviate
λ_n = standard normal deviate
μ = population mean
$\rho(k)$, ρ_k = population correlation coefficient of lag k
σ = population standard deviation
σ^2 = population variance
τ = parameter of Weibull distribution
ϕ_j = autoregressive coefficient of order j
ϕ_k = k partial autocorrelation coefficient for an AR(p) model
χ^2 = chi-squared
Ψ = parameter in theoretical cumulative density function of maximum run-lengths for an independent and normal series

Subscripts

e = of residual series
j = of transformed monthly series
k = time lag
m = of subseries with length m
n = of subseries with length n
t = year
v = month
w = of key station monthly series
x = of monthly and/or subordinate montly series
y = of annual series

Abbreviations

AIC(p,q) = Akaike Information Criteria for ARMA(p,q) model
Ann = annual
AR(p) = autoregressive model of order p
ARMA(p,q) = autoregressive moving average model of order p and q
Avg = average
CDF = cumulative denstiy function
cfsd = cubic feet per second - days
cfs days = cubic feet per second - days
CI = confidence interval
coef = coefficient
ext = extended
ft = feet
in = inches
MA(q) = moving average model of order q
max = maximum
min = minimum
Rec = record
sq mi = square miles
Std Dev = standard deviation
unext = unextended
yrs = years