

Utah State University

DigitalCommons@USU

Reports

Utah Water Research Laboratory

January 1981

Model choice: An Operational Comparison of Stochastic Streamflow Models for Droughts

W. Robert James

David S. Bowles

Nath T. Kottegoda

Follow this and additional works at: https://digitalcommons.usu.edu/water_rep



Part of the [Civil and Environmental Engineering Commons](#), and the [Water Resource Management Commons](#)

Recommended Citation

James, W. Robert; Bowles, David S.; and Kottegoda, Nath T., "Model choice: An Operational Comparison of Stochastic Streamflow Models for Droughts" (1981). *Reports*. Paper 321.

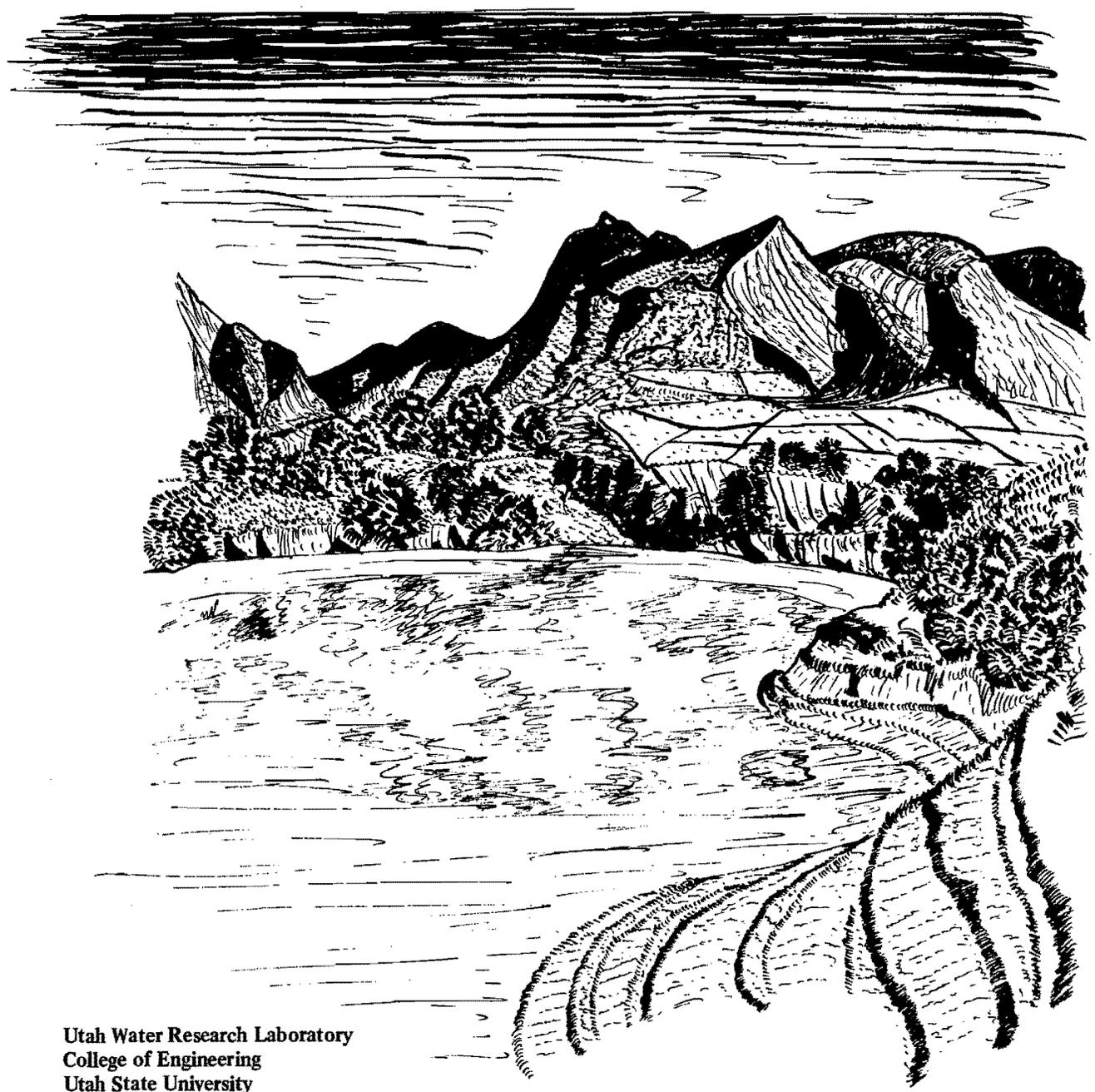
https://digitalcommons.usu.edu/water_rep/321

This Report is brought to you for free and open access by the Utah Water Research Laboratory at DigitalCommons@USU. It has been accepted for inclusion in Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



Model Choice: An Operational Comparison Of Stochastic Streamflow Models For Droughts

W. Robert James, David S. Bowles, Nath T. Kottegoda



Utah Water Research Laboratory
College of Engineering
Utah State University
Logan, Utah 84322

June 1981

WATER RESOURCES PLANNING SERIES
UWRL/P-81/03

MODEL CHOICE: AN OPERATIONAL COMPARISON
OF STOCHASTIC STREAMFLOW MODELS FOR DROUGHTS

by

W. Robert James
David S. Bowles
Nath T. Kottegoda

The work upon which this publication is based was supported in part by funds provided by the Office of Water Research and Technology (Agreement No. 14-34-001-9047), U.S. Department of the Interior, Washington, D.C., as authorized by the Water Research and Development Act of 1978.

WATER RESOURCES PLANNING SERIES
UWRL/P-81/03

Utah Water Research Laboratory
College of Engineering
Utah State University
Logan, Utah 84322

June 1981

Contents of this publication do not necessarily reflect the views and policies of the Office of Water Research and Technology, U. S. Department of the Interior, nor does mention of trade names or commercial products constitute their endorsement or recommendation for use by the U. S. Government.

ABSTRACT

The rapid development of stochastic or operational hydrology over the past 10 years has led to the need for some comparative analyses of the currently available long-term persistence models. Five annual stochastic streamflow generation models (autoregressive, autoregressive-moving-average (ARMA), ARMA-Markov, fast fractional Gaussian noise, and broken line) are compared on their ability to preserve drought-related time series properties and annual statistics. Using Monte Carlo generation procedures and comparing the average generated statistics and drought or water supply properties, a basis is established to evaluate model performance on four different Utah study streams.

A seasonal disaggregation model is applied to each of the generated annual models for each of the four study streams at a monthly disaggregation level. A model choice strategy is presented for the water resources engineer to select an annual stochastic streamflow model based on values of the historic time series' lag-one serial correlation and Hurst coefficient. Procedures are presented for annual and seasonal model parameter estimation, calibration, and generation. Techniques to ensure a consistent matrix for successful matrix decomposition are included such as normality, trend-analysis, and choice of model. User oriented model parameter estimation techniques that are easy and efficient to use are presented in a systematic manner.

The ARMA-Markov and ARMA models are judged to be the best overall models in terms of preserving the short and long term persistence statistics for the four historic time series studied. The broken line model is judged to be the best model in terms of minimizing the economic regret as determined by an agricultural crop production function.

Documentation and listings of the computer programs that were used for the stochastic models' parameter estimation, generation, and comparison techniques are presented in a supplementary appendix.

ACKNOWLEDGMENTS

This publication represents a report describing part of the work completed on a project which was supported with funds provided by the Office of Water Research and Technology (OWRT) through the Utah Center for Water Resources Research under agreement No. 14-34-0001-9047. Computer research funds were also provided in part by the College of Engineering, Utah State University. We also extend our thanks to Drs. Ronald V. Canfield, L. Douglas James, and J. Paul Riley for their careful and thoughtful review of the manuscript.

A summary completion report describing all parts of the work undertaken on this project was published by the Utah Water Research Laboratory. The report is "Vulnerability of Water Supply Systems to Droughts," Water Resources Planning Series, UWRL/P-80/08.

W. Robert James
David S. Bowles
Nath T. Kottegoda

TABLE OF CONTENTS

		Page
	ABSTRACT	iii
	ACKNOWLEDGMENTS	iv
	LIST OF FIGURES	viii
	LIST OF TABLES	x
Chapter		
1	INTRODUCTION	1
	Background	1
	Objective	2
	Summary of Contents	2
2	LITERATURE REVIEW	3
	Introduction	3
	Operational Hydrology	3
	Purpose	3
	Resemblance	3
	Systematic approach to hydrologic stochastic modeling	4
	Streamflow statistics	5
	Hydrologic nonhomogeneities	9
	State-of-the-Art for Annual Stochastic Streamflow Models	10
	Fractional Gaussian noise models	11
	ARMA models	11
	ARMA-Markov model	11
	Broken Line model	11
	Comparison of annual stochastic models and model choice considerations	12
	State-of-the-Art for Seasonal Stochastic Streamflow Models	14
	Model structure	15
	Parameter estimation solution procedures	16
	Modeling Drought Characteristics	16
	Models of Agricultural Economic Losses Due to Drought	19
3	SELECTION AND ANALYSIS OF STREAMFLOW TIME SERIES	21
	Introduction	21
	Selection of Study Streams	21
	Nonhomogeneity Analysis	21
	Annual Streamflow Statistics	22
	Monthly Streamflow Statistics	23
4	MODELING THE ANNUAL STREAMFLOW TIME SERIES	33
	Introduction	33

TABLE OF CONTENTS (Continued)

	Page
Second-order Autoregressive Model	33
Model structure	33
Generation procedure	34
ARMA (1,1) Model	34
Model structure	34
Calibration procedure	35
Generation procedure	35
ARMA-Markov Model	37
Model structure	37
Calibration to study streams	37
Generation procedure	38
Fast Fractional Gaussian Noise Model	40
Model structure	40
Calibration to study streams	41
Generation procedure	42
Broken Line Model	43
Model structure	43
Calibration to the study streams	45
Generation procedure	47
5 MODELING THE MONTHLY STREAMFLOW TIME SERIES	51
Model Structure	51
Calibration Procedure	51
Parameter estimation	51
Calibration to study streams	52
Comparison of historic and disaggregated monthly statistics	53
Generation Procedure	61
6 MODELING THE AGRICULTURAL ECONOMIC LOSSES FROM DROUGHT	63
Introduction	63
Crop Yield Model	63
Irrigation Diversion Rule	64
Irrigation Diversion Requirements	64
Calculation of Economic Regret	65
7 MODEL CHOICE FOR STUDY STREAMS	69
Introduction	69
Evaluation of Annual Performance	69
Beaver River	69
Blacksmith Fork	69
Logan River	71
Weber River	71
Summary	71
Evaluation of Seasonal Performance	73
Evaluation by Cost and Ease of Use	74
Comparison of Design Parameters	74
Evaluation by Economic Regret	80
A Model Choice Strategy	81

TABLE OF CONTENTS (Continued)

	Page
8 SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS	83
Summary	83
Conclusions	83
Recommendations	84
REFERENCES	85
APPENDIX A. LISTING OF MONTHLY STREAMFLOW DATA	89

LIST OF FIGURES

Figure	Page
1.1. Systematic procedure for stochastic modeling of hydrologic time series (adapted from Box and Jenkins (1970) and Salas and Smith (1980))	2
2.1. Systematic approach of hydrologic time series modeling (after Salas and Smith 1980)	4
2.2. Pox diagram of logarithm of the rescaled range for various series lengths (after O'Connell 1974 and James et al. 1979)	7
2.3. Crossing property definitions for droughts (after Salas et al. 1980)	18
3.1. Map of Utah showing location of study streamflow gaging stations	25
3.2. Recorded annual streamflow for Beaver River	26
3.3. Recorded annual streamflow for Blacksmith Fork River	26
3.4. Recorded annual streamflow for Logan River	27
3.5. Recorded annual streamflow for Weber River	27
4.1. Feasible range of $\rho(1)$ -K for ARMA(1,1) models (after Burges and Lettenmaier 1975)	36
4.2. Feasible range of $\rho(1)$ -K values for an ARMA-Markov (AMAK) model with $k_1 = 5$, $k_2 = 20$, $k_3 = 40$ (after Lettenmaier and Burges 1977)	40
4.3. Feasible range of $\rho(1)$ -h for FFGN model with $N = 50$ and $Q = 6$ (after Burges and Lettenmaier 1975)	44
4.4. A schematic representation of a simple Broken Line process (adapted from O'Connell 1974)	44
4.5. Solutions for broken line parameter a_1 using Equation 4.50 for $NL = 4$, $B = 5$	47
4.6. Solutions for broken line parameter a_1 using Equation 4.50 for $NL = 5$, $B = 3$	48
4.7. Solutions for broken line parameter a_1 using Equation 4.50 for $NL = 8$, $B = 3$	49
5.1. Procedure for examining the influence of several factors on parameter estimation for disaggregation models	53
5.2. Comparison of historic and disaggregated monthly statistics for Beaver	57
5.3. Comparison of historic and disaggregated monthly statistics for Blacksmith Fork	58
5.4. Comparison of historic and disaggregated monthly statistics for Logan	59

LIST OF FIGURES (Continued)

Figure	Page
5.5. Comparison of historic and disaggregated monthly statistics for Weber	60
5.6. Average absolute errors in elements of S_{YZ} which are separated by a J month lag	62
6.1. Irrigation diversion rule	65
7.1. Comparison of historic and generated $\rho(1)$ -K values for Beaver River	71
7.2. Comparison of historic and generated $\rho(1)$ -K values for Blacksmith Fork River	73
7.3. Comparison of historic and generated $\rho(1)$ -K values for Logan River	74
7.4. Comparison of historic and generated $\rho(1)$ -K values for Weber River	75
7.5. Comparison of historic and generated $\rho(1)$ values for the study streams	76
7.6. Comparison of the historic and generated K values for the study streams	76
7.7. Distribution of design parameters S^* and CD^* for Beaver River	77
7.8. Distribution of design parameters S^* and CD^* for Blacksmith Fork River	78
7.9. Distribution of design parameters S^* and CD^* for Logan River	79
7.10. Distribution of design parameters S^* and CD^* for Weber River	79
7.11. Recommended annual streamflow stochastic models based on $\rho(1)$ -K values	82

LIST OF TABLES

Table	Page
2.1. Explanations for Hurst Phenomenon (after Bowles 1979) . . .	8
2.2. Parameters preserved by annual stochastic models (after Bowles 1979)	10
2.3. Parameters preserved by seasonal and disaggregation models (Bowles 1979)	15
3.1. Streamflow gaging station location and watershed description	22
3.2. Annual streamflow data in acre-feet by water year . . .	23
3.3. Comparison of annual statistics of historic streamflow records	24
3.4. Monthly statistics for the Beaver River (1915-1978) . . .	28
3.5. Monthly statistics for the Blacksmith Fork (1914-1978)	29
3.6. Monthly statistics for the Logan River (1915-1978) . . .	30
3.7. Monthly statistics for the Weber River (1905-1978) . . .	31
4.1. AR2 model parameters and comparison of historic and generated statistics for study streams	34
4.2. ARMA(1,1) model parameters and comparison of historical and generated statistics for study streams	36
4.3. AMAK model parameters and comparison of historical and generated statistics for study streams	39
4.4. FFGN model parameters and comparison of historical and generated statistics for study streams	43
4.5. Broken line model parameters and comparison of historical and generated statistics for study streams	50
5.1. Summary of attempts to obtain real-valued parameters for disaggregation models	54
5.2. Parameter matrices for MR (m=3) disaggregation model of Beaver	55
5.3. Parameter matrices for VS disaggregation model of Blacksmith Fork	55
5.4. Parameter matrices for MR (m=2) disaggregation model of Logan	56
5.5. Parameter matrices for MR (m=2) disaggregation model of Weber	56
6.1. Calculation of monthly unit irrigation diversion requirements for corn at Logan, Utah	65

LIST OF TABLES (Continued)

Table	Page
6.2. Comparison of monthly irrigation requirements and range of irrigation diversion based on historic streamflows in ac. ft.	66
7.1. Overall summary of model results	70
7.2. Ranking of models by alternative model choice criteria	72
7.3. Comparison of cost and ease of use of annual streamflow models	77
7.4. Economic regret ^a in dollars per year for the study streams	80
A.1. Monthly streamflow data for Beaver River (ac. ft.)	89
A.2. Monthly streamflow data for Blacksmith Fork River (ac. ft.)	90
A.3. Monthly streamflow data for Logan River (ac. ft.)	91
A.4. Monthly streamflow data for Weber River (ac. ft.)	92

CHAPTER 1
INTRODUCTION

Background

One can approximate the flood flow for a return period equal to the length of gaged record as being equal to the largest flow recorded during the period of record, but a flood frequency analysis performed by fitting the data of record to a statistical distribution provides a much better estimate. Similarly, one can use the worst drought of record as a basis for water supply design, but assessment of the return period of that design drought requires a model that can generate flow sequences having the same magnitudes and the same patterns as to order as does the historical record.

Modeling to match historical distributions of magnitude in itself poses no real problem, but efforts to match order patterns is greatly complicated by the tendency for streams to have high flows followed by high flows, and low flows followed by low flows, as a result of moisture storage in the atmosphere, on the ground surface, or underground. Model problems in generating flow sequences that match this historical tendency have induced many researchers to investigate the nature of this behavior known as persistence in annual streamflows. This research has been directed towards determining the laws governing the stochastic processes determining streamflow. The search is complicated by the fact that persistence subjects the estimates of statistics used to characterize streamflow to large sampling errors. In fact, the errors may be so large that use of the estimates in the design of water storage systems may be very misleading.

Hurst (1951, 1956)¹ pioneered the study of long term persistence by empirical studies that led him to propose an empirical law that geophysical time series seem to follow and which can be used to express the degree of persistence as a coefficient. Since then, many advances in stochastic hydrological models have increased our ability to preserve the long term memory represented by the Hurst coefficient.

Prior to the recent emphasis on persistence in streamflow time series, annual flows were considered to be represented by independent random processes. Modeling advances that also preserve persistence mean that the probability of an extreme drought can be assigned with a greater degree of confidence

and the risk of water supply shortages can be better quantified.

Operational hydrology encompasses a variety of stochastic models for generating synthetic hydrologic time series that the water resource planner may then use to make realistic projections of future water supply conditions and associated estimates of the reliability of the supply. The modeling of hydrologic time series can be approached in a logical and systematic manner by using a form of the Box and Jenkins (1970) modeling approach. This iterative modeling approach, generalized to encompass a broader class of stochastic streamflow models, comprises the following five steps (see Figure 1.1.): 1) identification of the water resources system and model composition (e.g. univariate or multivariate, annual or seasonal); 2) choice of model type--short term or long term persistence model (e.g. autoregressive or fractional Gaussian noise); 3) identification of model form (e.g. order of the autoregressive model); 4) parameter estimation; and 5) model performance evaluation. Inadequate model performance may be judged by certain criteria such as preservation of historic statistics and reservoir storage requirements or runs. During model calibration¹ inadequate performance detected at step 5 may result in trial runs with alternative values assigned to model parameters or a change in the model form. If inadequate model performance persists over the entire range of model parameters and model forms it may be necessary to return to step 2 to select an alternative type of model, or to step 1 to simplify the system or model composition. Methods for implementing steps 1, 3, 4, and 5 are well covered in the literature, but a broadly applicable strategy for model choice in step 2 has yet to be developed. The

¹In this report the term model calibration is used to describe steps 2 and 3 of the systematic modeling procedure. It therefore includes the identification of model form, parameter estimation by rigorous mathematical techniques where applicable, and for some models, the trial-and-error determination of values for parameters for which rigorous parameter estimation techniques were not used (e.g. B and L in the fast fractional Gaussian noise model described in Chapter 4).

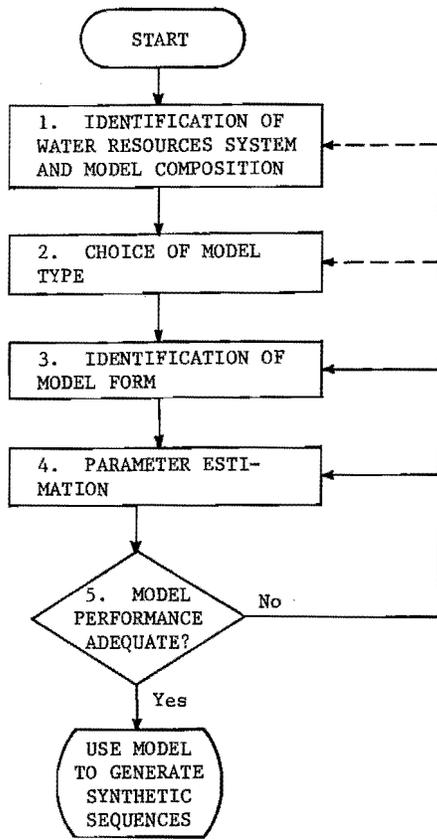


Figure 1.1. Systematic procedure for stochastic modeling of hydrologic time series (adapted from Box and Jenkins (1970) and Salas and Smith (1980)).

research reported herein is directed toward establishing a preliminary strategy for choosing between stochastic streamflow models.

Objective

The overall objective of this study is to compare the performance of five different stochastic streamflow models in order to develop a preliminary strategy for model choice in step 2 of the systematic modeling procedure (see Figure 1.1). The comparison was based on applying the five models to represent the streamflow records at four locations in Utah. Model performance was evaluated in terms of the following five

factors: the preservation of the short and long term persistence statistics for annual streamflows; the preservation of seasonal crossing properties; the cost and ease of model use; a comparison of reservoir capacity and critical drought design parameters; and minimizing economic regret associated with drought-related agricultural losses. Short term persistence was measured by the lag-one autocorrelation coefficient and long term persistence by the Hurst coefficient. Annual streamflow time series generated by the five stochastic models were disaggregated into monthly flow volumes. Reservoir capacity was calculated using the sequent peak algorithm and critical drought was calculated as the negative run-sum, defined with respect to monthly irrigation requirements, that has a 98 percent probability of nonexceedance. Economic regret was calculated based on a model for estimating agricultural economic losses from drought as a function of the shortfall of monthly irrigation diversions below crop water requirements.

Summary of Contents

A review of operational hydrology, univariate annual streamflow models, seasonal streamflow models, drought characteristics, and models of agricultural economic losses from drought is presented in Chapter 2. In Chapter 3, the selection of the study streams is described, followed by sections on non-homogeneity analysis of the streamflow records, the annual streamflow statistics, and the monthly streamflow statistics. Chapter 4 describes the structure, calibration, and generation procedures for each of the five annual stochastic streamflow models, and Chapter 5 describes the same items for the disaggregation models. The crop yield model, the calculation of irrigation diversions, the calculation of irrigation water requirements, and the procedure for calculating economic regret are explained in Chapter 6. In Chapter 7, model performance is evaluated with respect to the five factors listed under the statement of objective above and the proposed model choice strategy is presented. A summary of the research procedure is contained in Chapter 8, together with a list of study conclusions and recommendations for further work.

A user's manual for the computer programs for annual flow generation and disaggregation to monthly flows, the agricultural economic model, routines for calculating model performance statistics, parameter estimation for the disaggregation models, and calculation of the parameter a_1 in the broken line model has been printed separately and can be obtained by writing to the Utah Water Research Laboratory. A listing of the monthly streamflow time series for the four study streams can be found in Appendix A.

CHAPTER 2

LITERATURE REVIEW

Introduction

The review of literature begins with a description of the purpose of operational hydrology, a systematic approach to stochastic modeling in operational hydrology, and various streamflow statistics. The following sections contain a review of annual stochastic streamflow models, comparisons of these models, and a review of seasonal and disaggregation models. Other sections are brief reviews of models of drought characteristics and agricultural economic losses due to drought.

Operational Hydrology

Purpose

Operational hydrology, sometimes called synthetic or stochastic hydrology, denotes generation of hydrologic sequences for use by the water resource planner. They can be used to project future water supply availabilities and to assess drought risks. There are always uncertainties associated with planning a water resource system. The historic time series represents only one of many equally likely hydrologic time series that may occur at the same location. Operational hydrology does not provide new information, rather it is a means for extracting more information from the available historical record using a stochastic model of the underlying physical generating process. In practice, the underlying population distribution of the hydrologic time series is never known but must be inferred from the historical record. Also, the autocorrelation structure is not fully known, and a model must be chosen to approximate the apparent structure.

The statistical properties of a stationary time series can be obtained based on a single realization over a time interval or on several realizations at a particular time. The latter case does not occur in nature but is a necessary theoretical constraint in stochastic methods. The properties based on a single realization are known as time average properties. The properties based on several realizations at a given time are known as the ensemble properties, and while these cannot be estimated from historical records they can be represented by an "ensemble" of generated hydrologic sequences. If the time average properties and the ensemble properties are the same, the time

series is said to be ergodic. From a practical view point a time series can be said to be stationary if its time averaged properties estimated over several different time intervals do not change significantly from one time interval to another. Bendat and Piersol (1966) found that random data representing stationary physical phenomena are generally ergodic. Thus, it is usually assumed that the properties of a stationary random process can be estimated from a single historical record, since several simultaneous records are unavailable in hydrology, and used as a basis for parameter estimation for stochastic models. Stationary ergodic time series are important in hydrology for two reasons: 1) the mathematical techniques for modeling stationary series are well developed and 2) it is usually unclear how nonstationarities should be represented in future time periods unless they can be explained by such human activities as urbanization or the construction of reservoir storage.

Resemblance

O'Connell (1974) defines two types of resemblance between the statistics of historic time series and those of generated synthetic sequences. He refers to them as Type A and Type B resemblance. Under Type A resemblance the time-averaged values of the statistics of each generated sequence approach the values of the historical statistics asymptotically as the length of the generated sequence approaches infinity. Type B resemblance refers to the convergence of the ensemble-averaged statistics of the generated sequences, to the values of the historical statistics as the number of generated sequences is increased. For comparison purposes, the historic and synthetic sequences are made of equal length under Type B resemblance. Many model comparisons of the fractional Gaussian noise family of models are made by the Type A resemblance method in order to overcome small sampling biases for the long-term memory statistics such as the Hurst coefficient. However, Type B resemblance would appear to be the most appropriate in water resources planning because resemblance is achieved over a time period which is probably of length similar to that of the project life and because individual generated sequences may possess variability in their statistics consistent with the estimation error in the historic statistics.

Thus, the current state-of-the-art of stochastic modeling seeks to match statistics

computed from observed data sequences with corresponding statistics computed from an ensemble generated data sequence. Since available models cannot guarantee preservation of all statistics of interest over the entire range of values encountered in nature, one must often sacrifice in matching some statistics which are perhaps not of practical importance in order to do a better job of matching others. The implications of water resource design of not preserving certain parameters are not always considered.

Systematic approach to hydrologic stochastic modeling

A systematic procedure for stochastic modeling of hydrologic time series was briefly described in Chapter 1. A more

detailed description is presented by Salas and Smith (1980). Figure 2.1 is taken from their paper and illustrates the iterative procedure linking the six steps and four sets of factors that contribute to decisions that must be made at each step. The six steps are:

- 1) Identification of system - model composition
- 2) Selection of model type
- 3) Identification of model form
- 4) Estimation of model parameters
- 5) Testing of goodness-of-fit of the model
- 6) Evaluation of uncertainties

and the four sets of factors are grouped under the following headings:

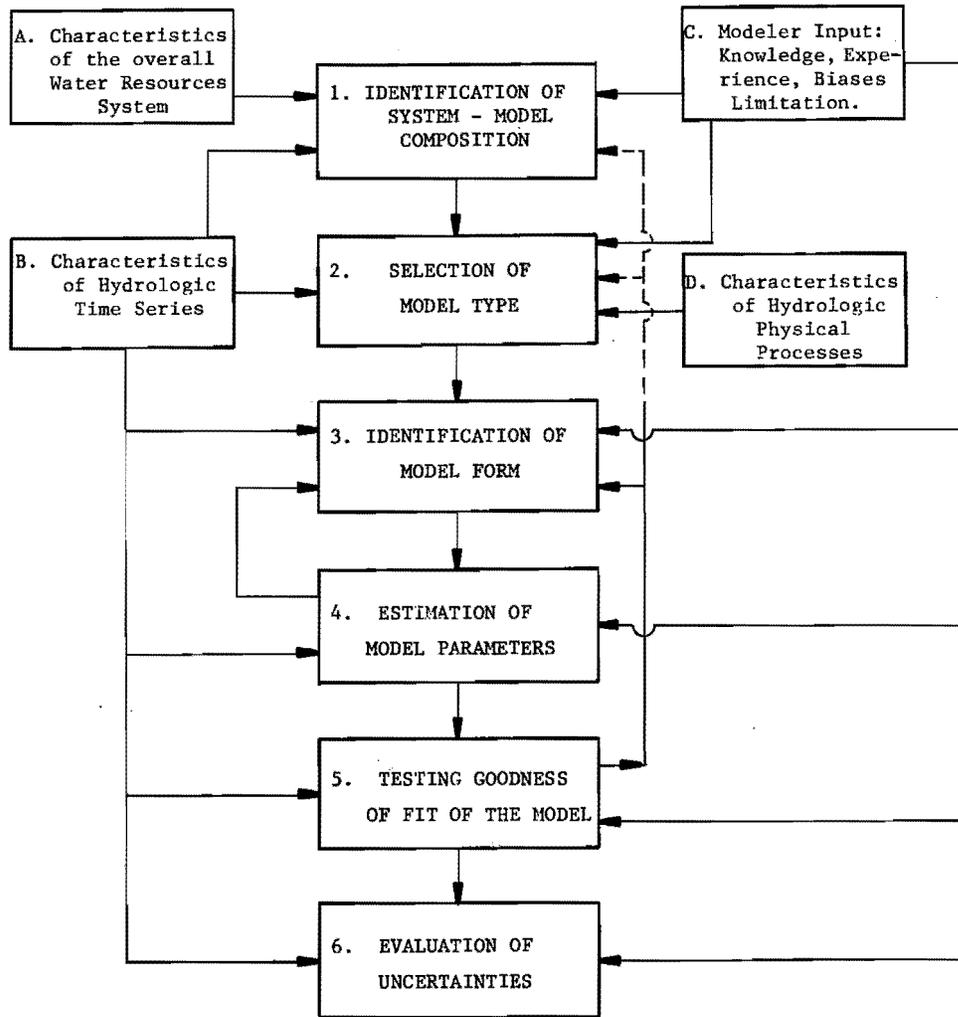


Figure 2.1. Systematic approach of hydrologic times series modeling (after Salas and Smith 1980).

- A) Characteristics of the overall water resources system
- B) Characteristics of the hydrologic time series
- C) Modeler input, knowledge, experience, biases, limitations
- D) Characteristics of hydrologic physical processes

The first step is concerned with whether the model should be univariate or multivariate, or a combination of an annual model and a seasonal disaggregation model. This decision will depend on 1) the characteristics of the overall water resource system, 2) cross-correlations between the hydrologic time series, 3) the modeler's biases and preferences, and 4) the temporal and spatial resolution of the generated data required to solve the design or planning problem.

Once a decision has been made on the model composition, the type of stochastic hydrology model must be selected in step two. The various types of models include autoregressive, ARMA, autoregressive-moving-average, ARMA-Markov, fractional Gaussian noise, and Broken Line. The choice of model type may be influenced by 1) the physical characteristics of hydrologic processes, (e.g. the influence of storage in the hydrologic system on the typical autocorrelation for the hydrologic variable being modeled), 2) the statistical characteristics of hydrologic time series relative to the feasible range of various statistics that can be preserved by each type of model, 3) the modeler's biases and preferences, and 4) prior unsuccessful attempts to use other types of models which might narrow the range of choice.

Once the model type has been selected, the form of the model needs to be identified. For example, the order of the autoregressive and moving average processes in an ARMA model must be determined. Also nonhomogeneities in the historic time series must be removed and a suitable transformation made to normality. Model form identification depends on 1) the characteristics of hydrologic time series and specifically the characteristics of the autocorrelation and partial autocorrelation functions in determining the order of an ARMA model, 2) the modeler's input, and 3) earlier unsuccessful attempts to use other model forms.

Step 4, parameter estimation is performed once the model form is identified. Parameter estimation is usually by rigorous procedures such as the method of moments or maximum likelihood, but in some cases parameters are estimated by trial-and-error procedures. Model performance or testing the goodness of fit of the model is evaluated at Step 5 by comparing the values of the statistics of the generated sequences to the historic values. These statistics should

include the mean, variance, skew, autocorrelogram, Hurst coefficient, and cross-correlation matrices for multivariate and disaggregation models. Statistical tests can be used to identify significant differences between generated and historic values of all of these statistics except skew. If these tests and comparisons indicate an unacceptable resemblance then the model form, the model type, or even the model composition may need to be changed and the steps following the change repeated until Step 5 is satisfactorily completed.

The sources of uncertainty in stochastic hydrologic modeling have been placed into three categories: 1) natural uncertainty, the uncertainty as to the size of a given event, 2) statistical uncertainty, the uncertainty in estimation of the parameters of the stochastic model due to limited data, and 3) model uncertainty, the uncertainty with respect to how well a particular stochastic model type and form approximate the true model. These sources of uncertainty must be taken into consideration for evaluating the model at Step 6.

Streamflow statistics

The historic time series of a hydrologic variable, such as annual streamflow, is considered to be a sample realization taken from the population of possible streamflow sequences. Quantities that are descriptive of a population are called parameters, and estimates of these quantities based on a sample realization are called sample estimates or statistics. Several important statistics in operational hydrology are defined and discussed below.

Mean. One important parameter or statistic is the mean, average, or expected value, which is a measure of the central tendency of the random variable. This parameter is estimated as follows for the random variable X:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad \dots \quad (2.1)$$

in which

\bar{X} = sample mean of X which is an unbiased estimator of the population mean, μ

x_i = observed values of X at time i

n = number of observed values of X in the historic record or sample

Standard deviation. The next sample statistic of interest is the variance or its square root, the standard deviation, which is a measure of dispersion of the random variable about its mean. From the estimated second moment about the mean, the following unbiased estimation, s^2 , of the variance, σ^2 , is obtained, as follows:

$$s^2 = \left(\frac{1}{n-1} \right) \sum_{i=1}^n (x_i - \bar{x})^2 \quad \dots \quad (2.2)$$

in which

s^2 = sample variance of X

Skew. A measure of asymmetry of the probability distribution of a hydrologic variable is the next statistic of interest. The coefficient of skewness is estimated from the third central sample moment and the standard deviation as follows:

$$g = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^{3/2}} \quad \dots \quad (2.3)$$

in which

g = the coefficient of skew and is a biased estimator of Y_1 , the population coefficient of skew

To unbiased the estimation for skewness, Bobee and Robitaille (1975) suggest the following bias correction of g :

$$g' = \frac{\sqrt{n(n-1)}}{(n-2)(1 + \frac{8.5}{n})} g \quad \dots \quad (2.4)$$

in which

g' = the unbiased coefficient of skew

Stochastic models are usually written in a form that assumes that the hydrologic variable being modeled is unskewed. If the hydrologic variable is significantly skewed it is necessary to apply a transformation to remove the skew. The Box and Cox (1964) transformation was used in this study. The general form of the Box-Cox transformation is as follows:

$$Y = \frac{X^\lambda - 1}{\lambda}, \lambda \neq 0$$

$$= \ln X, \lambda = 0 \quad \dots \quad (2.5)$$

in which

Y = transformed variable

X = untransformed variable

λ = coefficient

The coefficient, λ , is estimated such that the following goodness-of-fit statistic, T , is minimized (Hinckley 1977):

$$T = \frac{\bar{y} - y_{md}}{s_y} \quad \dots \quad (2.6)$$

in which

\bar{y} = sample mean of Y

y_{md} = sample median of Y

s_y = sample standard deviation of Y

Serial correlation. One measure of persistence in a time series is the degree of linear dependency or correlation between series values in consecutive time periods, called the serial correlation or autocorrelation. The sample lag-k correlation coefficient is given by:

$$r(k) = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x}_1)(x_{i+k} - \bar{x}_2)}{\left[\sum_{i=1}^{n-k} (x_i - \bar{x}_1)^2 \right]^{1/2} \left[\sum_{i=1}^{n-k} (x_{i+k} - \bar{x}_2)^2 \right]^{1/2}} \quad \dots \quad (2.7)$$

in which

$r(k)$ = sample estimate of population lag-k autocorrelation coefficient, $\rho(k)$, for X

k = number of time intervals involved in correlation

\bar{x}_1 = mean computed over interval $i=1$ to $n-k$

\bar{x}_2 = mean computed over interval $i = 1+k$ to n

Kendall (1954) showed that for a lag-one Markov model the small sample bias in the lag-one autocorrelation coefficient can be corrected as follows:

$$r'(1) = \hat{r}(1) - \frac{1}{n} (1 + 4\hat{r}(1)) \quad \dots \quad (2.8)$$

in which

$r'(1)$ = unbiased sample lag-one autocorrelation statistic

$\hat{r}(1)$ = sample lag-one autocorrelation coefficient

Wallis and O'Connell (1972) present small sample bias corrections for the lag-one Markov process. Small sample bias problems for approximations to discrete fractional Gaussian noise (FGN) models have been discussed by Wallis and Matalas (1971), Matalas and Wallis (1976), and Slack (1972). Estimates of sample variance from a lag-one Markov process are also biased and satisfy the following expression (Matalas 1966):

$$s'^2 = s^2 \left\{ 1 - \frac{2}{n(n-1)} \left[\frac{n r'(1)(1 - r'(1)) - r'(1)(1 - r'(1)^n)}{(1 - r'(1))^2} \right] \right\} \quad \dots \quad (2.9)$$

in which

- s'2 = unbiased estimate of variance
- s2 = biased population estimate of variance using Equation 2.2

However, the complexity of the available approximations precludes the analytical derivation of small sample properties for FGN (O'Connell 1974). For this study the biased sample estimates for variance and lag-one autocorrelation coefficients were used to avoid the complexity of bias corrections for the FGN model and because bias corrections are not available for some of the models which were applied.

Hurst phenomenon. Hurst (1951) calculated the range of cumulative departures from the sample mean, called the adjusted range, R^* , and normalized it by dividing by the estimated standard deviation, s , to develop a statistic called the Hurst coefficient that represents long-term persistence in hydrologic time series. He examined 690 annual time series of streamflow, river and lake levels, precipitation, temperature, pressure, tree ring growth, mud varves, sunspots, and wheat-price records for periods varying from 30 to 2000 years and found that the rescaled adjusted range can be represented as follows:

$$\frac{R^*}{s} = \left(\frac{n}{2}\right)^K \dots \dots \dots (2.10)$$

in which

K = estimate of the population Hurst coefficient, h

The mean value of K for the 690 series was found to be 0.729 with a standard deviation of 0.092. Hurst compared his empirical coefficient, K , with results from series of numbers taken from a normal distribution and found the latter K to equal 0.5. Feller (1951), using the theory of Brownian motion, found the same asymptotic results without assuming normality in the underlying process. The disagreement between the empirical Hurst coefficient of 0.729 and the theoretical value 0.5 has led to many efforts to explain this observed nonrandomness which is commonly referred to as the Hurst phenomenon.

Mandelbrot and Wallis (1969) suggested a more general form of Equation 2.10, as follows:

$$\frac{R^*}{s} = c n^H \dots \dots \dots (2.11)$$

in which

H = estimate of population Hurst coefficient, h

The estimate of the Hurst coefficient obtained by Hurst using Equation 2.10 is

normally designated K , while the estimate obtained from the 'pox plot' method is designated H . The two Hurst estimates can be quite different for the same time series. They found that values of H calculated from Equation 2.11 for data generated by a white noise process tended to be somewhat erratic for series lengths shorter than about 20 years. One would expect that, for natural series which have higher values of h , a longer record may be required to achieve a stable estimate. For this reason, Mandelbrot and Wallis proposed plotting a 'pox diagram' of R^*/s versus n_s , the subsequence length, on a log-log scale (See Figure 2.2) so that one could determine visually whether the length of record was sufficient to achieve a stable estimate.

Wallis and Matalas (1970) compared H estimates from the pox diagram method using Equation 2.11 with K estimates from Equation 2.10 for independent processes, lag-one Markov processes, and an approximation to discrete fractional Gaussian noise. They found that the pox diagram method showed less bias but greater variance.

Several explanations of the Hurst phenomenon have been proposed since Hurst's empirical findings in 1951 and these explanations are summarized in Table 2.1. These explanations are briefly discussed in the remaining paragraphs of this section.

1. Nonnormality of the marginal distribution. Marginal distribution as an explanation for the Hurst phenomenon has been stated

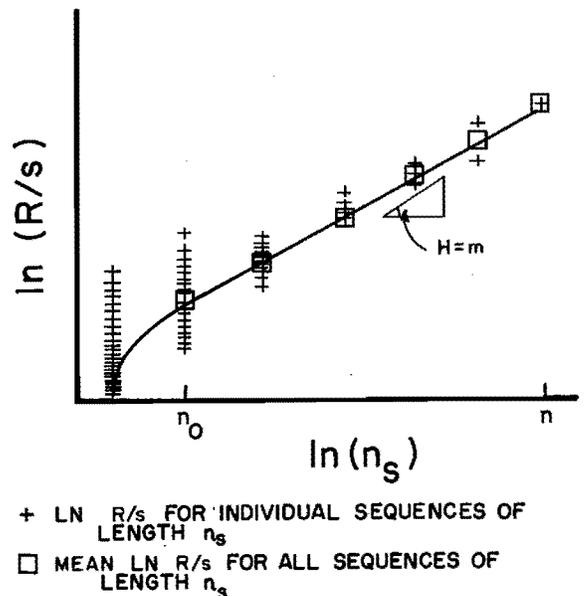


Figure 2.2. Pox diagram of logarithm of the rescaled range for various series lengths (after O'Connell 1974 and James et al. 1979).

Table 2.1. Explanations for Hurst Phenomenon (after Bowles 1979).

Explanation	Comments	Reference
Persistence	Tendency of high values of a time series to follow high values and low values to follow low values	Hurst (1951)
Autocorrelation	No simple correlations account for the phenomenon	Feller (1951)
Multiple-Lag Autocorrelation	Using 20 lags for N less than 60 K varied with N	Fiering (1967)
Infinite memory	Small but significant interdependence between values of a process at points in time long distant from each other	Mandelbrot and Wallis (1968, 1969)
Skewness (Nonnormality)	Many of Hurst's time series approximately normal. K might converge slowly to asymptotic value of 0.5	Lloyd (1967)
Nonstationarity	Fluctuating mean (climatic epochs)	Klemes (1974)
Semiinfinite Reservoir Storage	Hurst phenomenon result of many physical causes, one being storage systems	Klemes (1974)
Transience in small samples	Autocorrelation, nonstationarity and departure from normality all lead to transient behavior in small samples	Salas et al. (1977)

by Moran (1964) and Boes and Salas-LaCruz (1973) wherein they showed that the expected range, $E[R]$, and the expected adjusted range, $E[R^*]$, of the partial sums of independent, stable random variables behave asymptotically like n^h with $0.5 < h < 1.0$ thus paralleling the Hurst phenomenon. The effect of skewness on the marginal distribution and hence on the range statistics has been studied by Yevjevich (1965), Matalas and Huzzzen (1967), Moran (1968), Mandelbrot and Wallis (1969c), Anis and Lloyd (1975) and Salas et al. (1977). They have concluded that skewness has a small but detectable effect on the Hurst phenomenon in that it does affect the mean adjusted range, R^* in the transitional preasymptotic region but does not affect the mean rescaled adjusted range appreciably.

2. Autocorrelation structure. The autocorrelation structure of most geophysical time series describes the nature of its time dependent relationships. Feller (1951) was the first to suggest this as a possible cause for the Hurst phenomenon. Mandelbrot and Wallis (1969b) consider all Brownian domain models (which include: autoregressive models AR(p), autoregressive-moving-average models ARMA(p,q), and ARMA-Markov models) to follow the $R^*/s = cn^2$ law asymptotically. They stated that

...were the records in question generated by a random process such that observations far removed in time can be considered independent, $R(t,n)/s(t,n)$ should become asymptotically proportional to n^2 , which means that Hurst's law would have to 'break' for large enough

lags. But no such break has been observed. Thus, for practical purposes, geophysical records must be considered to have an 'infinite' span of statistical interdependence. (Mandelbrot and Wallis 1969b, p. 321.)

Most existing hydrologic time series are not sufficiently long to be asymptotic; but many existing models that belong to the Brownian domain of attraction and hence should not exhibit the Hurst phenomenon, do within the preasymptotic range. Gomide (1978) found that Markovian models can produce H estimates similar in magnitude to those found in geophysical time series by Hurst if the series are sufficiently long; that is $n > 1000$.

3. Nonstationarity in the mean. Nonstationarity of the mean as an explanation for the Hurst phenomenon was first suggested by Hurst (1956). Klemes (1974) conducted a series of Monte Carlo experiments with a fluctuating mean that varied normally but which remained constant for varying length intervals, called epochs, which were distributed exponentially, in order to show how various mechanisms produce different shapes of the plot R^*/s vs n_s . Similar experiments were also conducted by Potter (1975). Klemes concluded that the Hurst phenomenon cannot be attributed only to a specific physical cause; it may be caused by a type of infinite memory, nonstationarity in the process mean brought about by specific storage systems, and perhaps other causes as well. Salas et al. (1977) found that the "nonstationary" Hurst model has a correlation structure identical to the ARMA(1,1) model i.e. $\rho_k = \phi \rho_{k-1}$, $k=2,3,\dots$, where ϕ is the

parameter of the autoregressive component and the lag-one autocorrelation coefficient, ρ , is a function of both ϕ and θ , the parameter of the moving-average component. Their finding indicates the applicability of the ARMA (1,1) model given that the hydrologic time series is nonstationary. O'Connell (1971) implied that dependence-induced transience is a possible explanation of the Hurst phenomenon and Klemes (1974) and Potter (1975) working with nonstationary models showed the importance of transient effects on pox diagrams.

4. Transient behavior. Under this explanation the occurrence of values of $h > 0.5$ is attributed to a transient phenomenon for small sample size in the preasymptotic range, $100 < n < 1000$ (Anis and Lloyd 1975). For larger sample sizes $n > 1000$, h converges asymptotically to the Brownian domain of attraction value of $h = 0.5$.

5. Physical explanation. The final explanation for the Hurst phenomenon is long term persistence as a result of physical storage processes. A series of fractional Gaussian noise (FGN) models and approximations have been developed to preserve the Hurst phenomenon. Even though these FGN models can preserve values of h other than 0.5, they do so purely as operational tools developed apart from an understanding of the underlying physical processes. As a caution to those who would accept the mathematical basis of these models as an explanation of the physical basis for persistence in a geophysical system, Klemes (1974) noted:

It would be more realistic to say that: 1) fractional noises offer one possible explanation of the Hurst phenomenon; and 2) approximations to fractional noises provide a flexible operational tool for the simulation of series exhibiting the Hurst phenomenon. An ability to simulate, and even successfully predict, a specific phenomenon does not necessarily imply an ability to explain it correctly. A highly successful operational model may turn out to be totally unacceptable from the physical point of view (Klemes 1974, p. 675).

In support of this warning, Klemes showed that zero memory as well as infinite memory models can exhibit the Hurst phenomenon. Stochastic models that operate on stationarity of the mean can represent nonstationary time series for short time intervals, and stationarity or nonstationarity is a matter of time-series length. Experimentally, Klemes generated synthetic sequences from a distribution whose parameters were varied during the total simulation period but kept constant over shorter time periods called epochs. He showed that the infinite memory concept in FGN models can be a function of epoch length rather than of

total series length in accounting for the Hurst phenomenon. Specifically, Klemes attempted to show that a semiinfinite-storage reservoir model with various and diverse input processes might also explain the Hurst phenomenon. While Klemes was not able to prove or disprove specific physical explanations for the Hurst phenomenon, he did show that it can be generated using several models: long-term memory, nonstationarity in geophysical phenomena, or storage systems. Salas et al. (1977) have stated that the Hurst phenomenon might be explained by: "autocorrelation, nonstationarity, and departure from normality which either individually or combined accentuate a transient behavior, which is present in independent time series." In conclusion, one finds the Hurst phenomenon to be a property of hydrologic time series that can be quantified numerically but for which the physical causes are poorly understood and much debated.

Hydrologic nonhomogeneities

Nonhomogeneities may occur in hydrologic sequences due to natural or cultural underlying causes and may take on several different forms. These forms include periodicities, trends, and abrupt jumps in the mean or other statistics such as the variance, skew, serial and cross correlation coefficients. Natural causes include climate change and catastrophic events such as major earthquakes, landslides, and floods that change channel geometry or drainage systems. Cultural causes include the effects of reservoir construction, channelization, urbanization, and other land use changes. Nonhomogeneities must be identified and removed before parameter estimation can be performed. Methods for detecting nonhomogeneities include visual inspection of plots of raw data, cumulated data, and double mass plots; and statistical tests on split samples. Removal is achieved by describing the nonhomogeneity by a deterministic function of time, such as a step function, a sine wave, or a linear trend, and subtracting the values of this underlying deterministic component from the observed time series. An excellent discussion of hydrologic nonhomogeneities, their causes and removal, can be found in Kottegoda (1980).

The effects of nonhomogeneity on the persistence parameters of a time series can be significant. Yevjevich and Jeng (1969) demonstrate the effect of single and multiple jumps and linear and nonlinear trends on the probability distribution function, mean, variance, skewness, kurtosis and autocorrelation function of a time series. Systematic errors increase both variance and short-lag autocorrelations. Klemes (1974) and Potter (1975, 1976) show that stochastic models which exhibit random shifts in the mean level reproduce the Hurst phenomenon for very long time periods. These models can be considered as models of nonhomogeneity and therefore demonstrate that systematic errors can also inflate the Hurst coefficient of a time series.

State-of-the-Art for Annual
Stochastic Streamflow Models

Stochastic flow simulation began with the univariate models of the Harvard water program (Maass 1962). Markovian models were used in the Harvard program. However, Markov models do not replicate the Hurst phenomenon. Fractional Gaussian noise was introduced in the late 1960s in an attempt to preserve the Hurst phenomenon. The original form of FGN was too complex to model analytically and therefore approximations such as discrete fractional Gaussian noise and fast fractional Gaussian noise were developed. O'Connell (1971) proposed the autoregressive-moving-average (ARMA) process as an approximation to fractional Gaussian noise (O'Connell 1971). An alternative method for preserving the Hurst phenomenon is the Broken Line process (Rodriguez-Iturbe, Mejia, and Dawdy 1972 and Mejia 1972). A detailed historical account of these efforts has been recorded by O'Connell (1974). Each of these models is briefly described below. More detailed descriptions of the model structure, parameter estimation, and generation are given in Chapter 4. Table 2.2 contains a summary of the parameters preserved by several annual stochastic models. The short term models, which are not capable of preserving the Hurst coefficient, are listed first in Table 2.2, followed by the long term models, which can be used to preserve the Hurst coefficient.

Markov models

The Markov generating function for annual flows used by Thomas and Fiering (1962) is

$$X_t = \mu + A(X_{t-1} - \mu) + T_t \sigma \sqrt{1-\rho^2(1)} \quad (2.12)$$

in which

X_t = flow in time period t

μ = mean flow

A = regression coefficient of X_t on X_{t-1} (equal to the lag-one serial correlation coefficient, $\rho(1)$)

σ = standard deviation of flows

T_t = random variate from a standard normal distribution

Equation 2.12 is a first order autoregressive model in which the antecedent term, X_{t-1} , is sufficient to describe the current value, X_t . This can be extended to the n^{th} order model as follows:

$$X_t = \mu + A_1(X_{t-1} - \mu) + A_2(X_{t-2} - \mu) + \dots + A_n(X_{t-n} - \mu) + \sigma T_t \sqrt{1 - R^2} \quad (2.13)$$

in which

A_i = i^{th} multiple regression coefficient

R^2 = multiple regression coefficient of determination

Multiple autocorrelation analysis would then be used to determine the linear association within the historic time series through estimates of values for the A 's. Mathematically, one needs to estimate the partial autocorrelations to determine the order of the process. Physically, the maximum number of significant correlations is limited by the maximum duration of storage routing through aquifers making significant contribution to base flow. Fiering and Jackson (1971) describe tests for the significance of the parameters in autoregressive models and the order n required. Applications show the first-order model is adequate for most applications to annual streamflow and precipitation series.

Table 2.2. Parameters preserved by annual stochastic models (after Bowles 1979).

Model	Parameters Preserved							Reference
	μ	σ	γ_1	$\rho(1)$	$\rho(2)$	h		
1. AR1	X	X	X	X				Matalas (1967)
2. AR2	X	X	X	X	X			Matalas (1967)
3. Fast fraction Gaussian noise	X	X		X		X		Mandelbrot (1971)
4. ARMA (1,1)	X	X		X	X	X		O'Connell (1974)
5. ARMA-MARKOV	X	X		X		X		Burges and Lettenmaier (1977)
6. Broken Line	X	X	X	X		X		Mejia et al. (1972)

Fractional Gaussian noise models

Markov models do not replicate the Hurst phenomenon. The first models to succeed in preserving a Hurst coefficient exceeding 0.5 were the fractional Gaussian noise (FGN) models introduced to synthetic hydrology by Mandelbrot and Wallis (1968, 1969a,b,c). These models preserve long-term persistence by causing the autocorrelation function to decay increasingly slowly as h becomes larger than 0.5. The desired h is used as a model parameter and then preserved in the generated sequences. Fractional Gaussian noise sequences having h values exceeding 0.5 lie outside the Brownian domain, in that they do not satisfy the mixing property of Brownian motion. Specifically, past and future averages of the process become independent as the sample size approaches infinity. Mandelbrot and Van Ness (1968) and Mandelbrot and Wallis (1969c) define FGN mathematically, and O'Connell (1971) reviews the relationship of FGN to the Hurst phenomenon. The principal drawback of fractional Gaussian noise as a technique for generating flow sequences is its complexity and consequent high computer cost.

To simplify computation, Mandelbrot (1971) proposed an approximation of fractional Gaussian noise as a sum of Markov processes; he called the resulting model the fast fractional Gaussian noise model, FFGN. Chi, Neal, and Young (1973) have discussed Mandelbrot's FFGN model and outlined procedures for its implementation. An important aspect of FFGN is the use of autoregressive Markov-Gauss variables, instead of independent normal variables, which greatly reduces generation time (Lawrance and Kottegoda 1977). FFGN is composed of two additive components to represent the high-frequency and low-frequency effects. The choice of h and $\rho(1)$ completely determines the correlation structure of FFGN, and the other parameters B and N are determined based on empirical results by Chi, Neal, and Young (1973) and Kottegoda (1974).

ARMA models

The ARMA (p,q) acronym stands for autoregressive-moving-average, and the two numbers in parentheses p and q indicate the order of the autoregressive and the moving-average processes respectively. O'Connell (1974) found that for most annual streamflows an order of one for both the p and q would suffice. The equation for the ARMA (1,1) model is

$$X_t = \mu + \phi (X_{t-1} - \mu) - \theta \epsilon_{t-1} + \epsilon_t \quad (2.14)$$

in which ϵ_t is the error term. Comparison of Equations 2.14 and 2.12 shows the same form of relationship except that the memory of the preceding generated error term (ϵ_{t-1}) is added to maintain stability in the moving average. The parameters, ϕ and θ , in Equation 2.14 vary with the values for h and

$\rho(1)$ to be preserved. A disadvantage of the ARMA (1,1) model is that the ϕ and θ required to preserve given values of $\rho(1)$ and h cannot be determined explicitly. They have to be approximated empirically from tables given by O'Connell (1974) based on completed simulations or curves based on these tables given by Burges and Lettenmaier (1975, p. 17).

ARMA-Markov model

In order to have a generating model in which the parameters are an explicit function of the values of h and $\rho(1)$ to be preserved, Lettenmaier and Burges (1977) proposed combining Equations 2.12 and 2.14 into what they called an ARMA-Markov model having the form:

$$X_t = \mu + \rho^{(M)} (X_{t-1}^{(M)} - \mu) + \epsilon_t^{(M)} + \phi (X_{t-1}^{(AM)} - \mu) + \epsilon_t^{(AM)} - \theta \epsilon_{t-1}^{(AM)} \quad (2.15)$$

in which

M = label for Markov terms

AM = label for ARMA (1,1) terms

$\epsilon_t^{(M)}$ and $\epsilon_t^{(AM)}$ = independent noise processes

Variance for $\epsilon_t^{(M)}$ and $\epsilon_t^{(AM)}$ can be established from the other model parameters. The authors also provide a method for establishing values for ϕ and θ for the values of μ , σ , $\rho(1)$, and h estimated from the historical record. Details of this parameter estimation are given in Chapter 4. Lettenmaier and Burges (1977) found the ARMA (1,1) and ARMA-Markov models to provide reasonable approximation of the FGN process for values of $h \leq 0.80$ but for the results to become quite poor for higher values of h . The above mentioned ARMA (1,1) and similar models offer alternative approaches to univariate generation by which the Hurst phenomenon is maintained over finite time horizons.

Broken Line model

The simple Broken Line process results from the linear interpolation of equally-spaced independent random variables. The series is made stationary by imposing a random initial displacement. By summing N weighted simple Broken Line processes, a Broken Line process can be constructed that approximately reproduces the correlation function required to preserve the Hurst phenomenon. The weighting factor is derived from a relationship between the Hurst coefficient, h , and a quality parameter, B , and the time distance, a , between the random variables.

The Broken Line model has the advantage of preserving the second derivative of the

autocorrelation function at the origin, $\rho''(0)$, which is not possible in the FGN models, and provides better results with respect to crossing properties, extreme events, run lengths, and run sums for continuous time series. For the discrete time series used in hydrologic modeling, these advantages disappear. Curry and Bras (1978) showed that one could explicitly preserve the Hurst coefficient if one sacrificed the capability of preserving $\rho''(0)$. Use of the model is further handicapped by the fact that the parameters for the Broken Line model are difficult to compute as many investigations have found (Lawrance and Kottegoda 1977).

Comparison of annual stochastic models and model choice considerations

Water resource system simulation models are used to evaluate alternative system designs and operating policies. Either historical or generated streamflow data are input to these models, and computed time series of annual benefits and other performance information are output(s). Systems designs and operating policies are evaluated and ranked on the basis of summary statistics from these output time series (Fiering 1967). For example, if alternatives were to be ranked by mean annual net benefits, the better of two alternatives would tend to have the greater simulated mean annual net benefits. But the latter will not always be greater because of random variations in simulation outputs caused by random variations in the streamflow inputs. There exists some risk, therefore, that alternatives may not be correctly ranked depending on length of simulation runs, serial and cross-correlation of annual benefits, variance of annual benefits, and required level of resolution of differences between alternatives. (Vicens and Schaake 1980, p. 333.)

Stochastic hydrological models for generating streamflow or precipitation have been used by planners for evaluating various design alternatives. The role of operational hydrology in the design of water supply systems and the protection from floods and droughts has been well documented in the literature (e.g., Matalas 1975 and Lawrance and Kottegoda 1977). Obviously the severity of a drought will be greatest if several years of low flows follow each other and thus it is important to have a stochastic streamflow model that resembles this characteristic when applied to the system design. However, the interest of the engineer may not be in detailed representation of the streamflow but

in characterization of those properties of the time series that are pertinent to design in the planning situation being considered (Jackson 1975). These are not necessarily statistical moments of the flows; for example, maximum run-lengths and run-sums as measures of drought severity.

The operational hydrologic models applicable to the design of a water resource system can be evaluated with respect to several decision criteria. First, does the model adequately resemble the historic sequence in terms of its statistical moments and other statistical characteristics, such as crossing properties, important for system design.

The literature is replete with descriptions of research on the capability of the various models in generating synthetic flow sequences that preserve the statistical moments of the historic flow sequences. The coverage is much more meager on the importance of preserving these moments. For example, no one questions the need to preserve the mean and standard deviation; however, the skewness, lag-one serial correlation, and Hurst coefficient may also require close preservation. Sample size seriously affects estimates of parameters such as the coefficient of skew and the Hurst coefficient. This means that for the lengths of historic time series usually available in hydrology the historical estimates for these statistics are biased. On account of the uncertainty in these estimates it is questionable how closely they need to be preserved in the stochastic generation. Another aspect of how well it is necessary to preserve the coefficient of skew and Hurst coefficient is the sensitivity of the water resource system design to the accurate preservation of these statistics. In other words, does preserving the coefficient of skew or Hurst coefficient have a significant effect on the size of a reservoir, the severity of a water supply deficit or the magnitude of a flood?

Burges and Lettenmaier (1977, p. 1004) compare stochastic models on the basis of reservoir sizing and point out that:

Correct modeling of long term persistence (Hurst effect) will not in itself guarantee an accurate estimation of the storage distribution needed to satisfy a specific demand pattern. A careful analysis of the relative importance of parameters of the marginal distribution (coefficient of variation, CV, and skew, G) as opposed to persistence parameters ($\rho(1)$ and H) must be made, as well as between short and long-term persistence ($\rho(1)$ and H, respectively). The marginal distribution parameters, CV and G, have been shown to be much more important, even in cases where substantial long-term persistence is present, than had earlier

been thought. Specifically, for large H, $\rho(1)$ is relatively unimportant. The value of H is most important at high demand levels. For all other parameters held constant, increasing the magnitude of the marginal distribution skew reduces storage needs. Skew has most impact at moderate values of H and $\rho(1)$.

Some pertinent findings from the research by Burges and Lettenmaier (1977) are summarized below:

1) The importance of preserving the Hurst coefficient depends on the ratio of the annual demand to the average annual streamflow. If the demand ratio, D, is low, i.e. 0.5 to 0.7 then the Hurst coefficient has little effect on the distribution of reservoir size, but if the demand ratio is high, say $D > 0.7$, then the Hurst coefficient has greater importance.

2) FGN models generate fewer critical periods or droughts than Markov models.

3) Demand ratio (D), coefficient of variation (CV) and skew (G) are the most significant parameters affecting reservoir storage and critical droughts.

4) A low CV with high Hurst coefficient (i.e. ≥ 0.8) gives longer length critical droughts.

5) As CV increases so do reservoir storage requirements and the mean length of the critical drought. Based on the experimental work of Burges and Lettenmaier (1977), one would expect the seasonal operational models to follow the same patterns in model performance given the various combinations of CV, D, skew, $\rho(1)$, and Hurst coefficient.

Burges and Lettenmaier (1977) present a study of the sensitivity of reservoir capacity for 98 percent reliability to the lag-one serial correlation and Hurst coefficients. At an annual demand level of 0.9 and holding CV and G constant, a change in $\rho(1)$ from 0.2 to 0.45 caused an increase in reservoir capacity storage from 3.5 to 4.0 units; while as K was changed from 0.70 to 0.85 the reservoir capacity increased from 4.0 to 6.0 units. The ratio of the incremental increase in reservoir capacity to the incremental increase in $\rho(1)$ is

$$\frac{\Delta S}{\Delta \rho(1)} = \frac{0.50}{0.25} = 2.0 \quad \dots \quad (2.16)$$

in which

ΔS = change in reservoir capacity

$\Delta \rho(1)$ = change in lag-one serial correlation coefficient

For an incremental increase in the Hurst coefficient a similar ratio can be defined as follows:

$$\frac{\Delta S}{\Delta K} = \frac{2}{0.15} \quad \dots \quad (2.17)$$

in which

ΔK = change in Hurst coefficient

By dividing Equation 2.16 by Equation 2.17 we obtain:

$$\frac{\Delta K}{\Delta \rho(1)} = 0.15 \quad \dots \quad (2.18)$$

which indicates that the Hurst coefficient has more of an influence on reservoir capacity than does the lag-one serial correlation coefficient. This ratio is used in the analysis of the model performance in Chapter 7 to weigh the significance of errors in preserving $\rho(1)$ and K.

Burges and Linsley (1971) found that distributions of reservoir storage can be represented by the extreme value Type I (Gumbel) distribution. One would expect the critical drought to be distributed similarly to reservoir storage because of the similarity in their duration.

Another consideration for model selection is that the selected model should be the simplest of the models which are deemed adequate according to a statistical preservation criterion. Can a simple model replace a complex model if accuracy in certain types of resemblance can be sacrificed and what are the trade offs involved? Perhaps a simple stochastic streamflow model is adequate for the purposes involved in the water resource system design problem at hand, but what are the advantages, if any, to using a higher order or more complex model? By comparing different generating techniques, Askew, Yeh, and Hall (1971) found that for low-flow analysis the fractional Gaussian noise models were the most successful in resemblance of critical period characteristics, but that a simple Markovian model could be used, with some loss of representativeness, and with a saving in computational costs and time. By using an economic loss function, Jettmar and Young (1975) concluded that the problem of reservoir sizing is sensitive to the choice of stochastic model, while the problem of reservoir operation was insensitive to the choice of the stochastic model.

The need for research into model selection has been pointed out by several authors. In a discussion of a paper on advances in stochastic hydrology by Lawrance and Kottegoda (1977), P. E. O'Connell (p. 32) stated that,

...really formidable statistical difficulties arrive in applying some of these models about which the authors have said somewhat less, for example, questions of model choice and evaluation for a specific application, and parameter estimation in small samples...

it is suggested that future research should seek to establish the performance of such models in planning water resource systems, so that feedback can be obtained on where further model development should be concentrated.

Jackson (1975, p. 59) suggested that decision theory

...should be evaluated in terms of the benefits that they can be expected to provide, damages that they might cause, other costs, and similar measures; the models serve to allow the investigator to explore more fully the potential benefits and costs of the various decision choices.

Askew, Yeh, and Hall (1971) and Jettmar and Young (1975) compared alternative models and concluded that the simple Markovian models are sufficiently accurate and reasonable in cost for reservoir design purposes. The study on droughts by Askew, Yeh, and Hall (1971) did not use an economic loss function for evaluation of the different models. Research on model comparison using an economic loss function by Jettmar and Young (1975) was for a multipurpose reservoir; however, the results were heavily influenced by large flood damages. Most of the operational comparisons of annual stochastic streamflow models have not been in real world design situations. In general, researchers have been more statisticians than engineers and preferred to examine synthetic reservoir inflows and associated probability distributions of storage (Wallis and Matalas 1972, Slack, Wallis, and Matalas 1975, Burges and Lettenmaier 1975, 1977). Lawrance and Kottegoda (1977) do not consider that there is a best model available for all applications, especially with the unknown effects of future climatic changes. They consider that one should "let the historical data and physical understanding dictate the simplest model with which it is possible to get by." Also they see the need for some sort of reliability criterion for model choice as related to specific design purposes.

To evaluate the performance of the various models for a particular design purpose, several approaches have been used. Rogers (1978, p. 1004) applied the "type of errors" approach proposed by Thomas (1967).

Thomas considered two kinds of costs associated with errors in design parameters. He defined these costs in terms of the 'regret,' defined as the difference in benefits obtained with the best design possible and the benefits obtained when the design is based upon faulty estimates of the parameters implicit in making the design. The first kind of cost (or

regret), called 'type A error' is evaluated as the difference in benefits between the best possible design and the faulty design, the design benefits being evaluated, however, with the correct value of the design parameters. The second kind, called 'type B error' is similar but the faulty design is evaluated with the incorrect value of the design parameters.

Rogers adapted this approach using comparison of models instead of parameters and did pairwise comparisons of alternative models against some model which is believed to be superior or "best."

Jettmar and Young (1975) compared several hydrologic models using an economic loss function as the basis for regret (Thomas 1967). They found that the present value of the losses estimated with the economic simulation model was extremely sensitive to the Hurst coefficient while fairly insensitive to the lag-one serial correlation coefficient, thus indicating long-memory models more robust than short memory models for that particular design purpose. Askew, Yeh, and Hall (1971) found that fractional Gaussian noise models were superior to Markovian models in simulating the maximum permissible historic extraction rate per year from a reservoir and the historic deficiency accumulated relative to the mean flow.

The Akaike information criteria (a maximum-log-likelihood function) is one method for model selection recommended by McLeod and Hipel (1978) and by which they claim the ARMA model superior to the FGN models. However, McLeod and Hipel use a maximum likelihood estimator (MLE) of the Hurst coefficient which is highly correlated with the lag-one serial correlation thus biasing their work towards short term persistence (Todini and O'Connell 1979). Another disadvantage to the MLE method is the lengthy computer runs necessary to perform the estimation, McLeod and Hipel (1978) mention 11 minutes of computer cpu time for maximum likelihood estimation of the Hurst coefficient for a time series between 100 and 150 years in length.

State of the Art for Seasonal Stochastic Streamflow Models

Many critical water supply conditions exist only during the irrigation season when demands are high and streamflows are low. Therefore, operational hydrology includes the capability of generating flows over periods of less than a year in duration, for example a month, or a season. These models follow two approaches; first, the direct modeling of the seasonal periodicity by using a seasonal autoregressive model or a Box-Jenkins seasonal model; second, the disaggregation of generated annual flows to seasonal flows. The advantage of disaggregation is in the

preservation of the long-term annual statistics not necessarily preserved by the seasonal models. Table 2.3 contains a summary of the parameters preserved by six seasonal and disaggregation models. Model structure and parameter estimation solution procedures for some of these models are briefly presented below. More detailed descriptions of the two disaggregation models used in this study are contained in Chapter 5.

Model structure

The seasonal autoregressive Thomas-Fiering (1962) model and the seasonal Box-Jenkins (1970) models have been used to preserve periodicities in hydrologic and other phenomena. The disadvantages of these seasonal models have been their inability to preserve long-term properties, such as the Hurst coefficient, crossing properties and moments of annual flows; in fact, Box-Jenkins models do not preserve the periodicity in the variance. The annual stochastic models, such as fractional Gaussian noise, Broken Line and ARMA-Markov, that preserve the long-term properties cannot be adapted to preserve the seasonal short-term properties.

A disaggregation model was first developed by Valencia and Schaake (1973) and later modified by Mejia and Rousselle (1976), with another similar modification proposed by Hoshi and Burges (1979). Lane (1979) applied a disaggregation model to a multistation scheme with only selected cross-correlations being preserved.

The Valencia and Schaake (VS) model satisfactorily preserves within the year seasonal cross-correlations but does not preserve season-to-season correlation between water years. The Mejia and Rousselle (MR) model is an improved technique for preserving this season-to-season correlation between water years by incorporating an additional term containing some of the previous year's seasonal flows.

The VS model is represented by the following equations:

$$\underline{Y}_t = \underline{A}X_t + \underline{B}V_t \quad (2.19)$$

in which

\underline{Y} = n-vector of disaggregated standardized monthly flow volumes

X = standardized annual flow volume

\underline{V} = n-vector of random elements with zero mean and unit variance

\underline{A} = n-vector of model parameters

\underline{B} = nxn matrix of model parameters

n = number of months in a year (i.e. 12)

The MR model is a slightly expanded form of the VS model and is given by the following equation:

$$\underline{Y}_t = \underline{C}X_t + \underline{D}Z_t + \underline{E}V_t \quad (2.20)$$

Table 2.3. Parameters preserved by seasonal and disaggregation models (Bowles 1979).

Model type	Parameters Preserved ^a									Reference	
	μ	σ	$\rho(1)$	h	μ_i	σ_i	$\rho_i(1)$	$\rho_{ij}(0)$	$\rho_{ij}(1)$		
<u>SEASONAL</u>											
1. Lag-one Auto-regressive seasonal					X	X	X	X			Thomas-Fiering (1962)
2. Seasonal Auto-regressive-moving average					X	X	X	X	X		Box and Jenkins (1970)
<u>DISAGGREGATION</u>											
3. Disaggregation	X	X	X	X	X	X		X			Valencia and Schaake (1973)
4. Disaggregation	X	X	X	X	X	X		X	X		Mejia and Rousselle (1976)
5. Disaggregation	X	X	X	X	X	X		X	X ^b		Hoshi and Burges (1979)
6. Disaggregation	X	X	X	X	X	X		X ^c			Lane (1979)

^a i, j = season index, $\rho_{ij}(0)$ = cross correlation between seasons at lag zero
^b even years only (Bowles 1980)
^c only selected cross-correlations for multistation case

in which

- \underline{Z} = m-vector of standardized flow volumes for the last m months of the previous year
- \underline{C} = n-vector of model parameters
- \underline{D} = nxm matrix of model parameters
- \underline{E} = nxn matrix of model parameters
- m = number of months in (t-1)st year which are included in \underline{Z}_t

Parameter estimation solution procedures

The derivation of the coefficient matrix \underline{B} in the VS model and \underline{E} in the MR model requires the solution of a matrix equation with the unknown side in the form of a \underline{BB}^T or \underline{EE}^T . The solution for the \underline{B} (or \underline{E}) matrix can be accomplished by two general methods: 1) the principal components method and 2) Young's (1968) lower triangular method. Both methods are based on the property that \underline{BB}^T is an n dimensional real symmetric matrix.

Principal components method. The principal components method (Kendall 1961) utilizes the spectral theorem which applies to \underline{BB}^T because it is symmetric and leads to the following solution to \underline{B} :

$$\underline{B} = \underline{P}\underline{\lambda}^{1/2} \dots \dots \dots (2.21)$$

in which

- \underline{P} = matrix of eigen vectors
- $\underline{\lambda}$ = diagonal matrix of eigen values

In order to obtain a real-valued solution for \underline{B} , the matrix \underline{BB}^T must be positive semidefinite. This condition is met if all the eigen values of \underline{BB}^T are positive. Many computer applications of this solution procedure set small negative eigen values to zero which results in the corresponding columns of the \underline{B} matrix being zero.

Young's method. The other decomposition technique developed by Young (1968) also utilizes the principal of symmetry and is also known as the lower triangular solution. The matrix \underline{B} can be chosen as a lower triangular matrix such as:

$$\underline{B} = \begin{bmatrix} b_{11} & 0 & \dots & 0 \\ b_{21} & b_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix} \dots \dots \dots (2.22)$$

Therefore

$$\underline{BB}^T = \begin{bmatrix} b_{11}^2 & b_{11}b_{21} & \dots & b_{11}b_{n1} \\ b_{21}b_{11} & b_{21}^2 + b_{22}^2 & \dots & b_{21}b_{n1} + b_{22}b_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1}b_{11} & b_{n1}b_{21} + b_{n1}b_{22} & \dots & \sum_{i=1}^n b_{ni}^2 \end{bmatrix} \dots \dots \dots (2.23)$$

The solution for the b_{ij} 's is performed sequentially, starting with the first row in Equation 2.23 and moving to the right in that row. The solution then proceeds to the next row and goes along as the first, row to row until the general solution is obtained. Here, the problem of a nonpositive semidefinite matrix also arises when one tries to take the square root of a negative element. If it is only a small negative value the solution is to set the negative element to zero and then to use programming logic that avoids division by zero.

Valencia and Schaake (1973) and Slack (1973) discuss the importance of using proper estimators for the computation of serial and cross correlation matrices to ensure that the resulting coefficient matrices are positive semidefinite. Slack (1973) mentions as important conditions to help overcome the problem of inconsistent serial correlation observed in historical records, 2) the less complex the stochastic equation of a model, the more likely the availability is of parameter estimation techniques yielding an unconditional model. Other schemes for ensuring a positive semidefinite matrix are suggested by Fiering (1968) and Crosby and Maddock (1970) as corrections to make if the record lengths are not equal, but successful arrival to positive semidefinite coefficient matrices cannot be guaranteed (Slack 1973).

Modeling Drought Characteristics

Drought characteristics such as duration and severity are measured by the critical periods as defined by Askew, Yeh, and Hall (1971) or by the crossing properties (Kottegoda 1974) determined from the theory of runs (Yevjevich 1967). Millan and Yevjevich (1971, p. 1) discuss these properties in the following quotation:

In the past, standard practice for designing reservoirs relied heavily on the "critical period," defined as that period in time when the historic record would have been most critical with respect to water demands required from a system. It is claimed (Hall and Askew 1969)

[sic.] that design based on a critical period results in a reservoir storage capacity equal to the capacity obtained by using the total length of record. However, to determine this critical period accurately reliable knowledge of system performance, particularly demand patterns and operational rules and policies, is required. In the absence of this knowledge or because of complexity in obtaining this kind of information, the critical drought period is usually determined under simplified assumptions. Even though some current design practices take into account not only the critical drought period but also the total deficit of water supply by a reservoir under study, it still remains that this critical drought period represents, in most cases, the largest part of the deficit allowed by these design criteria.

W. Hall and A. I. Askew (1969) [sic.] found for 25 selected rivers across the continental United States that the dates of the critical periods agree with the dates of the major droughts in each region. Using this information, in general, a historic drought is considered that event for which most designs must perform satisfactorily. This is based on the assumptions that the most severe drought to be observed during the lifetime of a project will be about the same as a previously recorded maximum historic drought. The probability, however, that the critical drought period observed in the past will be the same as the critical drought period expected to be observed in the future is usually small. The probability is large that a very different drought will be observed.

Therefore replication of the historic maximum drought is not the purpose of operational hydrology. The purpose is instead the generation of different droughts which will make possible the estimation of a probability distribution of droughts and thus estimates of the probability of nonexceedance of any drought magnitude. Thus a stochastic model is not rejected because the generated maximum drought does not match the historic maximum. In fact, when long streamflow sequences are generated by a drought representative model one would expect to find more severe and some less severe droughts than in the historic record.

The crossing properties used to define drought duration and severity are illustrated in Figure 2.3. For a particular sequence of streamflows, $X(t)$, and a given demand level, X_0 , which could in general also be a

function of time and streamflow, $X_0(t,x)$, the periods when demands exceed the flows are readily indentified. Both the duration of each period, termed the run-length, and the volume or severity of the excess demands in each period, termed the run-sum, are readily calculated. The distribution of run-lengths can be used to assess the probability of different length droughts. The number of crossings of X_0 by $X(t)$ also can be used to characterize the time series. Thus crossing properties are random variables and their estimation is based on historical records in which large fluctuations in their observed values occur over the period of record. Millan and Yevjevich (1971) proposed a mathematical model for the distribution of the crossing properties. Their work was based on the following simplifying assumptions:

- 1) The observed time series is a stationary stochastic process
- 2) No trends or positive or negative jumps
- 3) No periodicities
- 4) Either independent process or simple dependent process such as a lag-one Markov process.

They further defined a representative drought by the expected value of run length and run-sum. Saldarriaga and Yevjevich (1970) have shown that run-length properties for flows following stationary processes are independent of the mean and the standard deviation of the underlying process, but they are dependent on the crossing level, X_0 , the lag-one serial correlation, $\rho(1)$, and the coefficient of skewness, g , of the underlying flow population distribution. The magnitude of the run-sum properties is also directly proportional to the standard deviation of the process (Millan and Yevjevich 1971).

The probability distributions of generated drought parameters of most interest are those associated with the largest drought in the sample. There are two probability distributions associated with the investigation of drought properties, which can be derived by Monte Carlo method: 1) the distribution of the longest run-length or drought duration, $P(L_{max})$ in the generated hydrologic time series; 2) the distribution of the largest run-sum or drought deficit, $P(D_{max})$ in the generated hydrologic time series. Millan and Yevjevich (1971) conducted a detailed analysis of regression coefficients and partial correlation coefficients of a mathematical relationship between crossing level, sample size, lag-one serial correlation coefficient, and coefficient of skewness as the independent variables and the crossing properties such as mean longest run-length, mean largest run-sum, as dependent variables. From their regression analysis they concluded that the most signif-

KEY

X = monthly streamflow in acre-feet

X_0 = level of demand for a certain water supply

L = run-length of negative deviations from the base level, X_0 ,
measure of the duration of the drought

D = run-sum of the negative deviations from the base level, X_0 ,
measure of the severity of a drought

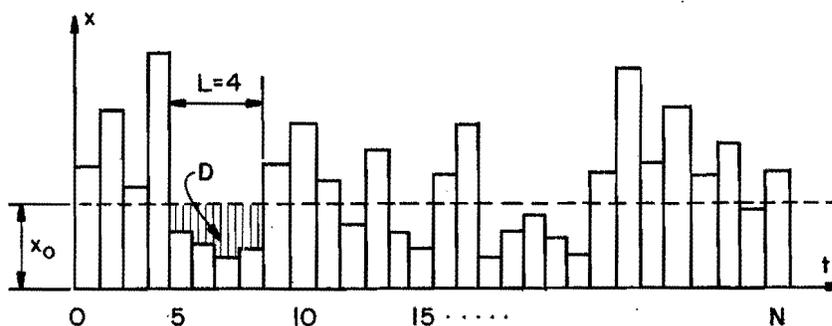


Figure 2.3. Crossing property definitions for droughts (after Salas et al. 1980).

icant (in terms of explained variance) independent variables are the crossing level and the sample size, next in importance is the lag-one serial correlation coefficient and last was the coefficient of skewness. As expected though, skewness was more important for the largest run-length than for the longest run-length. As the lag-one serial correlation coefficient increases, the run-length increases, and the number of crossings decrease. The importance of preserving the skewness of the historic series in regards to maintaining the crossing properties has been pointed out by Kottegeda (1974). Kottegeda's work also indicates that the simpler autoregressive model is sufficient for preserving the crossing properties when compared with the more sophisticated fractional Gaussian noise and Broken Line models. Kottegeda (1974) notes, however, that preservation of the crossing properties alone is not sufficient for water supply system design and that consideration of the effects of long-term persistence on storage requirements is also important.

An interesting result from the application of the probability distributions for the drought properties to annual streamflow series of 10 rivers, showed 3 rivers (Missouri, Rio Grande and MeKong) to have historical critical periods, more severe than would be predicted from generated sequences of flows, while the other 7 rivers, using first-order autoregressive linear models, gave on the average longer or larger representative droughts than were observed historically. These results do not support the conclusions by Hall, Askew, and Yeh (1969), that the historic droughts were significantly

more severe than the generated droughts. The models used by Hall, Askew, and Yeh (1969) were the autoregressive model and the fractional Gaussian noise model prior to the approximations of FGN developed since 1969. The selection of the crossing level is an important factor in defining the properties for both historic and generated droughts. The characteristics of the hydrologic time series plays an important part in determining the "best" type of model to apply.

In conclusion, the correct model should be fitted to the streamflow time series based on the importance and magnitude of the demand or crossing level for the water resource system and based on the statistics of the time series such as serial correlation and skewness. One would not be justified in aprior discarding operational hydrology models in favor of the historic statistics for a water supply design problem as recommended by Hall, Askew, and Yeh (1969). Neither would one be justified in always choosing a Markov model as recommended by Yevjevich (1964). Millan and Yevjevich (1971) showed that not all streamflows are well modeled by Markov models. Perhaps much of the confusion over which model to choose based on preservation of the drought crossing properties lies in the fact that the crossing properties have large sampling variations and are very sensitive to the correlation structure present in the historic streamflow time series. Therefore a strategy for model choice is needed in order to determine which type of model fits best the streamflow time series for the water resource design purposes at hand.

Models of Agricultural Economic
Losses Due to Drought

A nonlinear loss function has been used by Jettmar and Young (1975) for measuring model performance. For agricultural drought analysis, a crop production function that operates over an interseason time period provides the most information on the harm associated with various situations with respect to timing and severity of droughts. Such a crop yield function would respond to an inadequate water supply during a critical crop growth stage. Unfortunately, most crop yield functions that do not operate on a seasonal basis operate on a daily level and require daily meteorologic and irrigation data. Generating these data was beyond the scope of this study.

Crop yield is measured by the amount of vegetative matter produced and the weight of the vegetation is proportional to the amount of water transpired during the growing

season (Stewart et al. 1977). Hexem and Heady (1978) reviewed the literature concerning crop production functions basing a large part of their research on the pioneering work of Mitscherlich (Briggs 1925). Mitscherlich quantified the relationships between plant growth and environmental factors. Hexem and Heady (1978) using multiple regression analysis predicted crop yields on a seasonal basis from such variables as water supply, fertilizer, water holding capacity and other soil properties. Another study on crop yield models by Stewart et al. (1977) provided data comparing different models on a regional basis for corn.

Hanks (1974) developed a water-budget crop-yield model relating the yield as a function of evapotranspiration. The economic application to the Hanks model was performed by Gowon, Anderson, and Biswas (1978) with emphasis on critical growth stages for corn crops.

CHAPTER 3

SELECTION AND ANALYSIS OF STREAMFLOW TIME SERIES

Introduction

Both the selection of model type and form, and the acceptability of model performance depend heavily on the value of the streamflow statistics estimated from the historic record. Therefore, one of the first tasks in step one of the systematic modeling procedure (see Figure 1.1), is to obtain these estimates from the streamflow record. At the beginning of this chapter the selection of the study streams is described, followed by sections on nonhomogeneity analysis, the annual streamflow statistics, and the monthly streamflow statistics.

Selection of Study Streams

Synthetic traces from stochastic hydrologic models are only as good as the statistics used to generate them, hence, the importance of accurate measurements over a long enough period to provide a good data sample from which to estimate the statistics. Data collection requires searching out time series of recorded data (for this study, streamflow data), that are both reliable and long enough for good statistical estimation. Specific considerations in choosing the streamflow time series for comparison of stochastic models were: 1) a record of at least 60 years in length for both annual and monthly flows to provide good estimates of streamflow statistics; 2) to ensure a statistically homogeneous time series there should be no storage reservoirs, only small diversion and little change in watershed development above the gaging station over the period of record; 3) a wide range in important streamflow statistics (e.g. lag-one autocorrelation and Hurst coefficients) on which to test the models. Only the first of these three criteria would be necessary in evaluating a streamflow series for use in a design application. If the time series is statistically nonhomogeneous then various detrending techniques such as reverse reservoir routing or subtracting a known mean trend component can be used to approximately restore a streamflow series to a homogeneous series. The danger of detrending is that a trend caused by some physical change in the watershed or due to a natural climatic trend, perhaps with a long periodicity, may be

removed with relative ease: but the question is what assumptions should be made about its continuation in the future? Where possible these assumptions must be based on a cause-and-effect explanation of the trend; for example the growth of urbanization.

The four streamflows chosen for study are the Beaver River near Beaver, Utah, Blacksmith Fork above Utah Power and Light Company's Dam, near Hyrum, Utah, the Logan River above State Dam, near Logan, Utah, and Weber River near Oakley, Utah. The streamflow series appears to satisfy all three selection criteria as can be seen from the descriptions in Table 3.1 and the analysis of the series presented later in this chapter. The historic streamflow series was obtained from the Water Resources Data for Utah, Part 1, Surface Water Records (USGS 1979). Figure 3.1 shows the location of the streamflow gaging stations on a map of the State of Utah. Annual flows are listed in Table 3.2 and presented graphically in Figures 3.2-3.5. Monthly flows are listed in Appendix A.

Nonhomogeneity Analysis

Before estimation of either the annual or seasonal statistics of a streamflow time series is attempted it is necessary to verify that the series is statistically homogeneous. The following four approaches were used to investigate the annual streamflow records for trends or jumps:

- 1) Study of streamflow gage history and other information on upstream watershed development affecting streamflows.

- 2) Visual inspection of tabulated and plotted records (see Table 3.2 and Figures 3.2-3.5).

- 3) Visual inspection of double-mass plots and plots of cumulative streamflows.

- 4) Fisher-Behrens (Kendall and Stuart 1973) test on the difference in the means of split samples.

The only nonhomogeneities identified were a drop in the mean streamflows of the Logan

Table 3.1. Streamflow gaging station location and watershed description.

Stream	USGS Station No.	Period of Record (Water Years)	Drainage Area (Square Miles)	Remarks ¹
Beaver River	234500	1915-1978	91	No diversions upstream for irrigation. Water diverted for hydroelectric power, but returned to stream above station. Some regulation by power plants and several small reservoirs.
Blacksmith Fork	113500	1914-1978	268	A few small diversions for irrigation of about 200 acres above station. Flow is slightly regulated by power plant above station.
Logan River	109000	1901-1978	214	Flow effected by regulation and diversion above station for power, irrigation, and municipal culinary supply. Utah Power and Light Co. stopped diverting water from river November 1970 at which time the tailrace station was discontinued. Logan, Hyde Park and Smithfield Canal diversions are combined with the Logan River flows for purposes of this study.
Weber River	128500	1905-1978	162	Several small diversions for irrigation above station. Flow slightly regulated by several small lakes on head waters and a small reservoir on Smith and Morehouse Creek. Total capacity of lakes and reservoir is 3400 acre feet.

¹ Taken from USGS (1979).

and Weber Rivers which appears to have occurred between 1910 and 1920 (see Figures 3.4 and 3.5). It was also found that precipitation levels were higher before this period. The records for the Beaver and Blacksmith Fork Rivers did not begin early enough to clearly show the effects of this apparent climatic shift. This downshift in the mean streamflows was found to influence the outcome of parameter estimation for the disaggregation model of the Logan streamflows. Specifically, if statistics based on the entire period of record, including the earlier high flow years, were used it was not possible to obtain real-valued solutions to the disaggregation model parameters. However, if the starting year for estimating statistics was moved to 1913, to exclude the earlier high flow years, then real-valued model parameters were obtained. Therefore, the Logan record was used beginning with 1913. This parameter estimation problem did not occur for the Weber River and it was decided to use the entire period of record and not to omit the earlier high flow years on the basis that this apparent shift in the mean may provide a more realistic estimate in the long-term persistence as measured by the Hurst coefficient.

Annual Streamflow Statistics

Several annual streamflow statistics were calculated for the four study streams and are presented in Table 3.3. The statistics estimated are: mean, standard deviation, coefficient of variation, skew coefficient, lag-one autocorrelation coefficient, and Hurst coefficient. In addition, the expected run length and run sum with respect to a crossing level equal to the annual mean flow was also calculated. Both the K and H estimators of the Hurst coefficient were calculated.

The calculated statistics indicate average variabilities as measured by the coefficient of variation. Since the annual skews are low, no attempt was made to use a transformation to account for skew. There is greater variability in the value of H than K, as is characteristic of the H estimator (Wallis and Matalas 1970). The values of K are within the range normally found in streamflows (Hurst 1951). Values of the expected run length are quite similar for all four streams, perhaps suggesting that this statistic can be expected to be fairly stable in a given geographic region.

Table 3.2. Annual streamflow data in acre-feet by water year.

Year	Beaver River	Blacksmith Fork	Logan River	Weber River	Year	Beaver River	Blacksmith Fork	Logan River	Weber River
1901			235400.		1940	37354.	45450.	113110.	89290.
1902			213080.		1941	62110.	38350.	97250.	121260.
1903			228970.		1942	47290.	45100.	120000.	147680.
1904			270590.		1943	34890.	84810.	209410.	152980.
1905			180240.	118170.	1944	51410.	60650.	140570.	161200.
1906			222320.	186730.	1945	42840.	71920.	158980.	132700.
1907			373550.	300240.	1946	31310.	125020.	216570.	146260.
1908			204090.	158010.	1947	50430.	81500.	175490.	156790.
1909			326470.	277260.	1948	35760.	101460.	194300.	141850.
1910			281390.	183220.	1949	46200.	98470.	183570.	157960.
1911			284720.	184610.	1950	24464.	139480.	262780.	199490.
1912			297020.	202300.	1951	23894.	136870.	237410.	173820.
1913			185760.	168210.	1952	63644.	147820.	220330.	217300.
1914		117980.	231140.	241370.	1953	23188.	90960.	170400.	138630.
1915	45560.	67960.	136470.	140530.	1954	23466.	69830.	130200.	101360.
1916	49230.	119330.	224960.	200570.	1955	22066.	69630.	132430.	114350.
1917	38930.	148570.	233950.	235620.	1956	26432.	108540.	200050.	168330.
1918	31910.	105020.	194120.	149880.	1957	50347.	102120.	185590.	171360.
1919	35052.	73980.	154550.	121200.	1958	51520.	97080.	184330.	127490.
1920	49940.	109850.	220400.	197470.	1959	16791.	68910.	143640.	109180.
1921	53410.	154060.	280200.	270360.	1960	19504.	67010.	139100.	112630.
1922	58420.	140900.	244610.	231970.	1961	18603.	43940.	95380.	66950.
1923	51470.	137680.	239780.	202030.	1962	33266.	84670.	170890.	168330.
1924	29010.	93080.	171550.	110500.	1963	20445.	64150.	145920.	121580.
1925	36887.	86410.	171470.	139490.	1964	25620.	73750.	159200.	153430.
1926	44360.	67130.	135880.	132950.	1965	33807.	113590.	230190.	207720.
1927	36020.	88070.	187000.	166500.	1966	22059.	82320.	153480.	133850.
1928	40890.	90980.	200460.	168180.	1967	36205.	101030.	189560.	189970.
1929	45970.	88520.	182450.	180290.	1968	42344.	85370.	172000.	175940.
1930	36690.	66640.	145400.	133730.	1969	48218.	95350.	180240.	183350.
1931	18725.	43200.	92840.	78070.	1970	34790.	80590.	177910.	146930.
1932	37716.	102560.	215130.	182640.	1971	29300.	161470.	295180.	181810.
1933	33635.	74670.	170960.	137840.	1972	15155.	171510.	275540.	170580.
1934	16480.	42160.	91000.	56050.	1973	49759.	107580.	166570.	153950.
1935	36108.	52300.	141440.	142910.	1974	28120.	126780.	226270.	176160.
1936	48350.	110190.	234530.	179130.	1975	28584.	116650.	219540.	195750.
1937	56120.	86980.	165280.	128480.	1976	18023.	105010.	192160.	136150.
1938	43820.	88280.	190510.	155070.	1977	11686.	51190.	87270.	58520.
1939	25110.	63080.	137110.	109250.	1978	42970.	86150.	177200.	161970.

Monthly Streamflow Statistics

Tables 3.4 through 3.7 contain the monthly streamflow statistics calculated for the four study streams. The mean monthly flows and the percentages of the mean annual flow represented by the mean monthly flow show that more than 50 percent of the annual flow occurs in a two or three month period in the late spring. Variability in the monthly flows is greatest during the late spring and early summer as indicated by the standard deviations and coefficients of variation of the monthly flows. The skew coefficient of monthly flows is generally small and positive. The largest values of the skew coefficient generally occur during the spring runoff period and especially in March and June. A Box-Cox transformation with $\lambda = 0.33$ was found to minimize the average monthly goodness-of-fit statistic, T (see Equation 2.6), for all study streams. The effect of this transformation on parameter estimation for the disaggregation models is

discussed in Chapter 5, and values of T are presented for various values of λ (see Table 5.1). The selected transformation is better at representing skew in the months with low skew than in the months with high skew (see Figures 5.2 through 5.5). However, it was decided to use one value of λ for all months rather than attempt to use different transformations for each month which might lead to inconsistencies in parameter estimation for the disaggregation model. The lag-one autocorrelation coefficient between monthly flows (e.g. between June and July) is consistently high in all months except during the spring. The high values occur because of the dominant influence of the groundwater recession in controlling flows in adjacent months. During the spring the influence of the groundwater recession is less than the influence of the snowmelt condition which runs from months to months with less serial correlation than the groundwater recession. Correlations between the monthly and annual flow volumes are least in the fall before the

Table 3.3. Comparison of annual statistics of historic streamflow records.

Symbol	Statistic Description	Units	Stream			
			Beaver	Blacksmith Fork	Logan	Weber
N	Length of record used ¹	yrs	64	65	66	74
\bar{X}	Mean	ac-ft	36,306	92,659	180,438	158,326
s	Standard deviation	ac-ft	12,706	31,402	47,005	46,570
CV	Coefficient of variation	-	0.35	0.34	0.26	0.29
g	Skew coefficient	-	0.08	0.50	0.15	0.52
r(1)	Lag-one autocorrelation coefficient	-	0.24	0.49	0.32	0.26
H	Hurst coefficient ²	-	0.61	0.74	0.73	0.84
K	Hurst coefficient ³	-	0.76	0.76	0.72	0.78
E(RL)	Expected run length ⁴	yrs	2.43	3.09	2.29	2.60
E(RS)	Expected run sum ⁴	ac-ft-yrs	23,940	69,189	79,809	83,632

¹Last year of record used was 1978.

²H estimator based on pox diagram.

³K estimator given in Equation 2.8.

⁴Expected run length and run sum are based on a crossing level of the annual mean flow.

winter snow influences runoff but increase in the spring due to the direct influence of snow runoff and are generally maintained at the higher levels in the summer due to the important indirect influence of the runoff from the snowpack in the preceding winter.

The two cross-correlation matrices presented in the lower part of Tables 3.4 through 3.7 are used to estimate the parameter matrices for the disaggregation models described in Chapter 5. S_{YY} , the matrix of cross correlations between monthly flows in the same water year is a symmetric matrix.

In general the correlations decrease with greater separation between the months, that is on diagonals more distant from the leading diagonal. The matrix of cross-correlations between monthly flows in adjacent years, S_{ZY} , contains the largest values in the lower left hand corner which is the cross-correlation between flows in the adjacent months of September in year t-1 and October in year t. These cross-correlations decrease on diagonals moving toward the top right hand corner although the year-to-year serial correlation induces an increase in the cross-correlations in the last few columns of S_{YZ} .

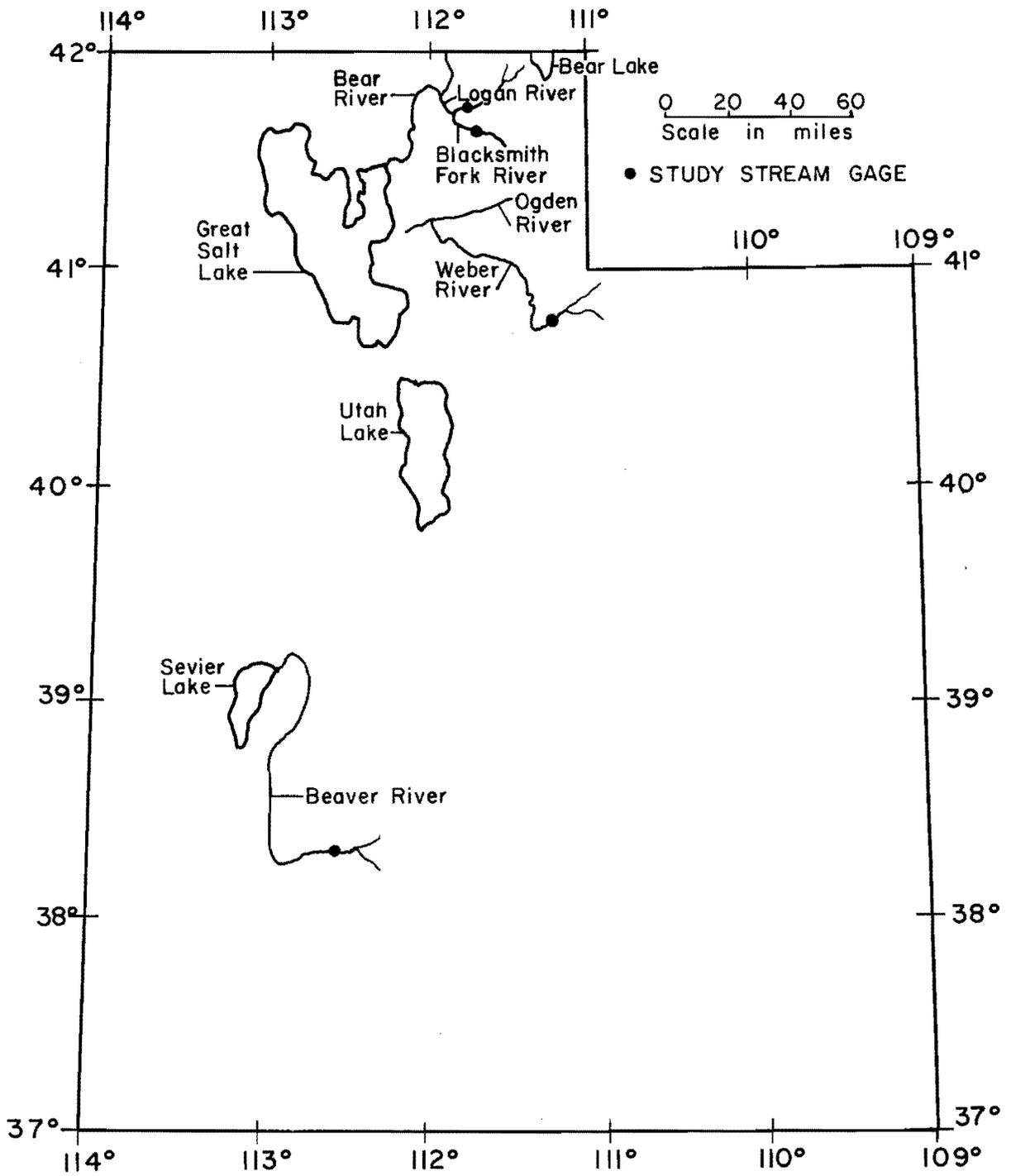


Figure 3.1. Map of Utah showing location of study streamflow gaging stations.

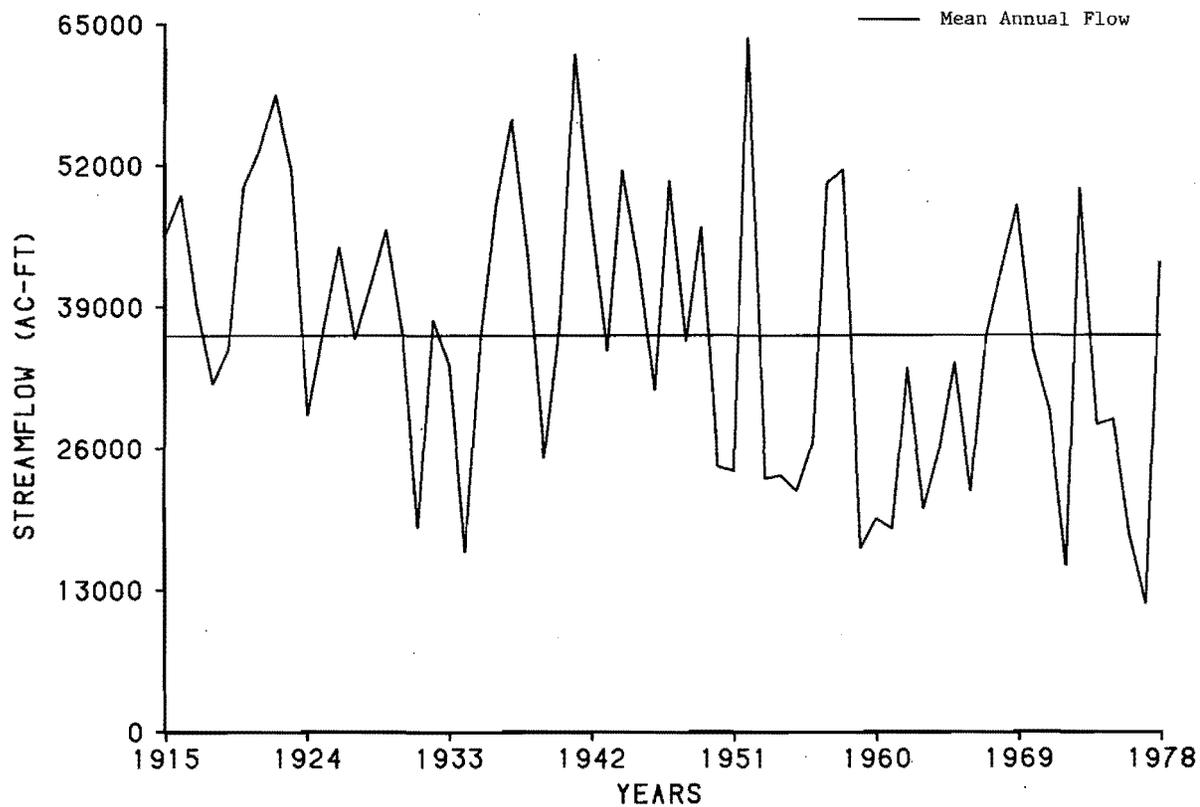


Figure 3.2. Recorded annual streamflows for Beaver River.

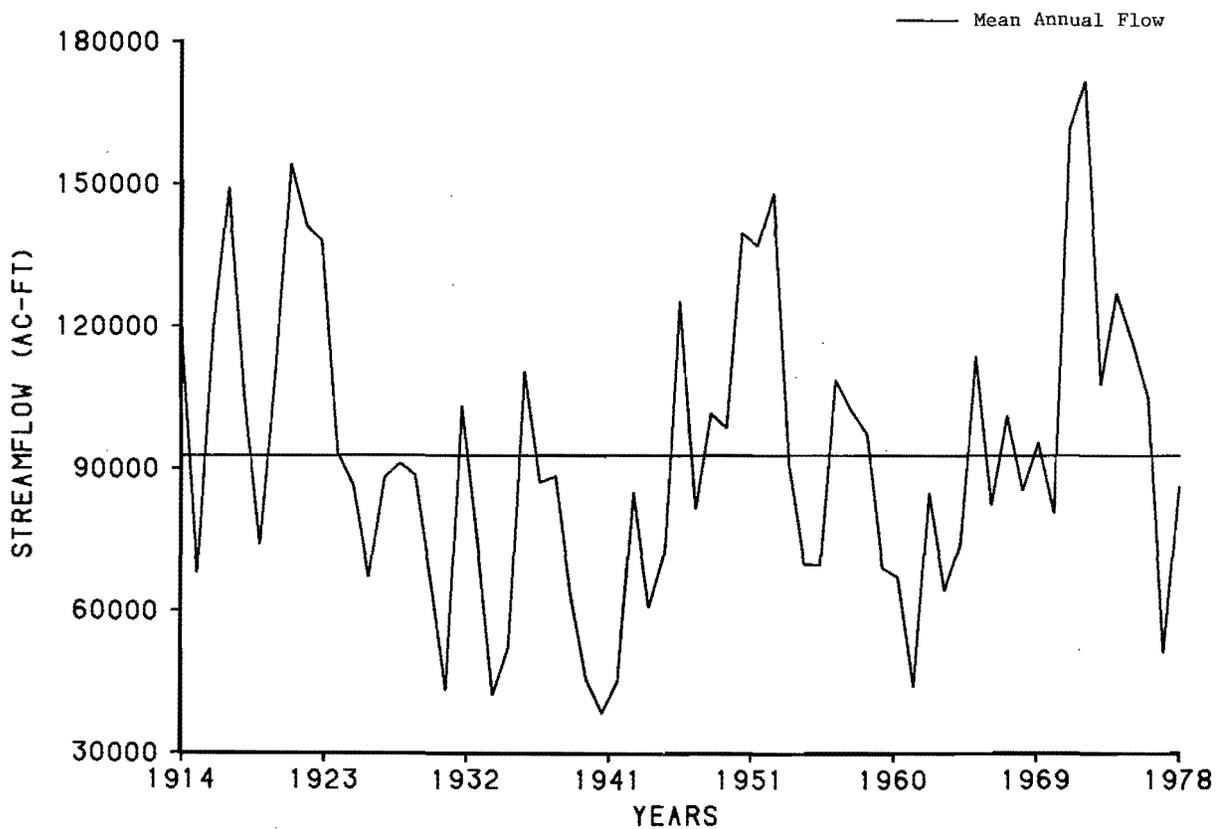


Figure 3.3. Recorded annual streamflow for Blacksmith Fork River.

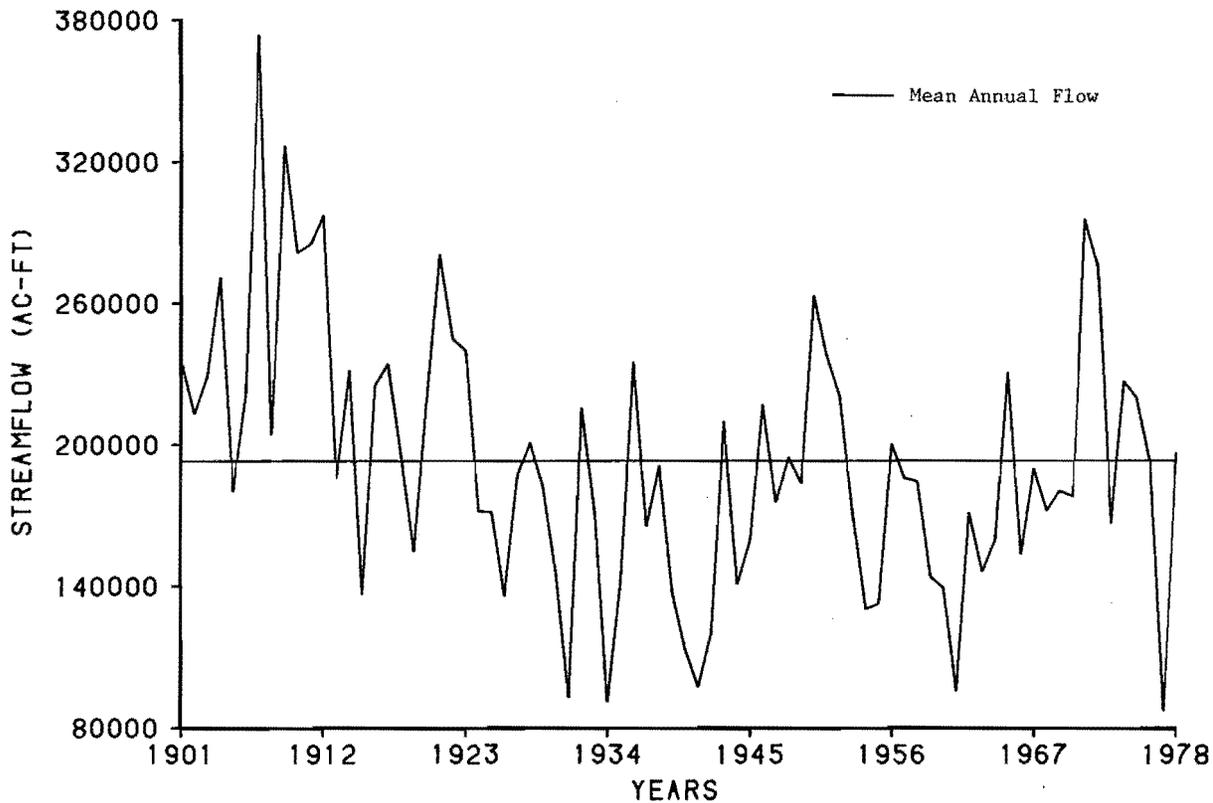


Figure 3.4. Recorded annual streamflow for Logan River.

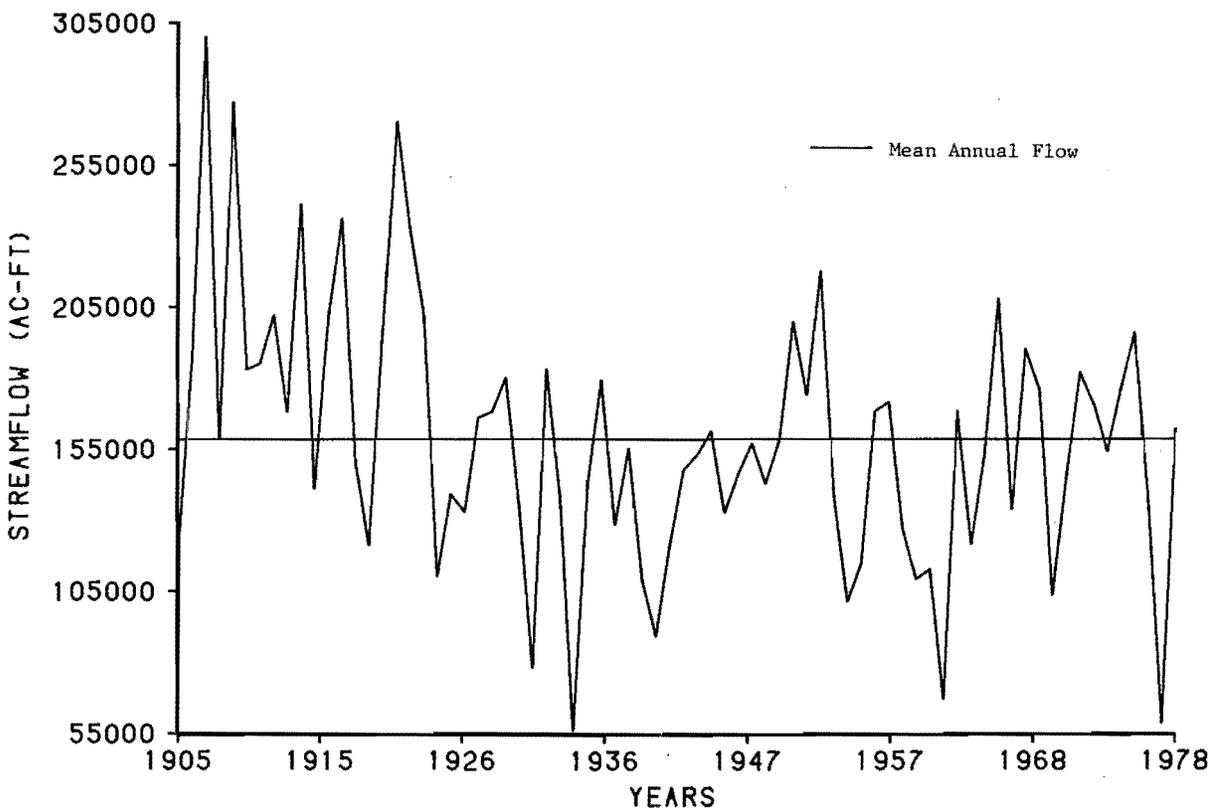


Figure 3.5. Recorded annual streamflow for Weber River.

Table 3.4. Monthly statistics for the Beaver River (1915-1978).

Month	Mean \bar{y} (ac-ft)	Percentage of mean annual flow in month	Standard Deviation s(ac-ft)	Coefficient of varia- tion CV	Skew coefficient g	Lag-one month autocorrelation r(1)	Correlation between monthly and annual flows s_{YX}
O	1,425	3.92	380	0.27	0.56	0.89	0.33
N	1,247	3.43	290	0.23	0.46	0.93	0.37
D	1,175	3.24	236	0.20	-0.10	0.86	0.38
J	1,101	3.03	227	0.21	0.30	0.86	0.34
F	1,049	2.89	202	0.19	0.35	0.80	0.35
M	1,350	3.72	335	0.25	1.65	0.81	0.32
A	3,173	8.74	1,402	0.44	0.76	0.53	0.35
M	10,332	28.46	5,466	0.53	0.69	0.39	0.89
J	8,387	23.10	5,086	0.61	0.92	0.57	0.84
J	3,560	9.81	1,791	0.50	0.31	0.85	0.82
A	2,059	5.67	820	0.40	0.44	0.86	0.80
S	1,450	3.99	419	0.29	-0.08	0.87	0.85

Matrix of cross-correlations between monthly flows in same water year¹ S_{YY} .

	O	N	D	J	F	M	A	M	J	J	A	S
O	1.00	0.93	0.81	0.76	0.81	0.62	0.39	0.34	0.06	0.01	0.04	0.27
N		1.00	0.86	0.81	0.85	0.69	0.42	0.39	0.06	0.03	0.09	0.33
D			1.00	0.86	0.77	0.60	0.43	0.40	0.05	0.09	0.13	0.37
J				1.00	0.80	0.54	0.33	0.31	0.11	0.05	0.10	0.30
F					1.00	0.81	0.47	0.32	0.07	0.02	0.13	0.36
M						1.00	0.53	0.31	0.03	-0.01	0.18	0.40
A							1.00	0.39	-0.03	0.03	0.18	0.32
M								1.00	0.57	0.59	0.57	0.70
J									1.00	0.85	0.77	0.69
J										1.00	0.86	0.76
A											1.00	0.87
S												1.00

Matrix of cross-correlations between monthly flows in adjacent years, t-1 and t, S_{ZY} .

	O	N	D	J	F	M	A	M	J	J	A	S
O	0.34	0.40	0.41	0.28	0.48	0.50	0.33	0.19	-0.10	-0.06	0.01	0.23
N	0.42	0.47	0.44	0.36	0.55	0.53	0.35	0.22	-0.12	-0.09	-0.02	0.24
D	0.48	0.53	0.49	0.35	0.58	0.56	0.40	0.18	-0.21	-0.13	-0.05	0.16
J	0.39	0.49	0.44	0.33	0.52	0.51	0.42	0.20	-0.25	-0.19	-0.10	0.12
F	0.49	0.55	0.46	0.43	0.60	0.48	0.34	0.21	-0.09	-0.10	-0.06	0.21
M	0.54	0.54	0.41	0.35	0.55	0.38	0.24	0.30	0.13	0.06	0.06	0.31
A	0.43	0.37	0.35	0.28	0.44	0.32	0.14	0.20	0.04	-0.04	0.03	0.20
M	0.65	0.66	0.70	0.72	0.63	0.43	0.21	0.21	0.06	-0.02	0.03	0.20
J	0.61	0.58	0.62	0.64	0.55	0.36	0.15	0.18	0.04	0.05	0.05	0.13
J	0.64	0.57	0.60	0.59	0.50	0.32	0.10	0.05	-0.07	-0.07	-0.05	0.00
A	0.78	0.70	0.66	0.63	0.60	0.40	0.22	0.25	0.00	0.00	0.00	0.13
S	0.89	0.85	0.78	0.72	0.72	0.52	0.30	0.31	0.00	-0.02	0.05	0.21

¹ S_{YY} is a symmetric matrix.

Table 3.5. Monthly statistics for the Blacksmith Fork (1914-1978).

Month	Mean \bar{y} (ac-ft)	Percentage of mean annual flow in month	Standard Deviation s (ac-ft)	Coefficient of varia- tion CV	Skew coefficient g	Lag-one month autocorrelation r(1)	Correlation between monthly and annual flows s_{YX}
O	5,678	6.13	1,657	0.29	0.37	0.99	0.56
N	5,177	5.59	1,337	0.26	0.28	0.99	0.58
D	5,043	5.44	1,264	0.25	0.24	0.94	0.62
J	4,810	5.19	1,200	0.25	0.43	0.96	0.69
F	4,416	4.77	1,031	0.23	0.40	0.90	0.70
M	6,081	6.56	2,271	0.37	2.28	0.68	0.72
A	13,040	14.08	6,845	0.52	0.99	0.62	0.77
M	18,538	20.01	10,310	0.56	0.72	0.63	0.92
J	10,117	10.92	4,716	0.47	1.08	0.91	0.91
J	7,464	8.06	2,740	0.37	0.52	0.96	0.98
A	6,527	7.05	2,260	0.35	0.48	0.99	0.98
S	5,750	6.21	1,855	0.32	0.45	0.99	0.98

Matrix of cross-correlations between flows in same water year¹ S_{YY} .

	O	N	D	J	F	M	A	M	J	J	A	S
O	1.00	0.99	0.93	0.87	0.80	0.57	0.18	0.34	0.41	0.51	0.54	0.55
N		1.00	0.94	0.88	0.81	0.60	0.21	0.36	0.42	0.53	0.56	0.57
D			1.00	0.97	0.88	0.62	0.28	0.37	0.44	0.56	0.59	0.60
J				1.00	0.90	0.68	0.37	0.45	0.53	0.64	0.67	0.68
F					1.00	0.68	0.44	0.45	0.53	0.66	0.69	0.71
M						1.00	0.62	0.47	0.52	0.65	0.67	0.67
A							1.00	0.63	0.57	0.70	0.70	0.70
M								1.00	0.91	0.92	0.91	0.89
J									1.00	0.96	0.94	0.93
J										1.00	0.99	0.98
A											1.00	0.99
S												1.00

Matrix of cross-correlations between monthly flows in adjacent years, t-1 and t, S_{ZY} .

	O	N	D	J	F	M	A	M	J	J	A	S
O	0.57	0.57	0.55	0.53	0.42	0.26	0.02	0.13	0.14	0.22	0.24	0.27
N	0.60	0.59	0.56	0.55	0.43	0.27	0.04	0.16	0.16	0.25	0.26	0.30
D	0.64	0.63	0.59	0.58	0.48	0.34	0.07	0.17	0.19	0.28	0.30	0.33
J	0.71	0.71	0.69	0.68	0.60	0.45	0.11	0.21	0.25	0.36	0.37	0.40
F	0.73	0.72	0.67	0.66	0.60	0.43	0.13	0.23	0.26	0.36	0.38	0.40
M	0.68	0.66	0.62	0.57	0.45	0.27	-0.05	0.24	0.32	0.34	0.34	0.37
A	0.69	0.67	0.60	0.51	0.46	0.22	-0.09	0.14	0.24	0.27	0.26	0.26
M	0.88	0.89	0.83	0.75	0.68	0.46	0.08	0.32	0.35	0.43	0.45	0.45
J	0.92	0.93	0.88	0.80	0.75	0.54	0.18	0.29	0.34	0.45	0.48	0.48
J	0.98	0.97	0.92	0.85	0.78	0.53	0.13	0.29	0.37	0.47	0.49	0.50
A	0.99	0.98	0.93	0.86	0.79	0.53	0.13	0.31	0.38	0.48	0.50	0.51
S	0.99	0.98	0.94	0.87	0.80	0.56	0.16	0.34	0.40	0.51	0.53	0.54

¹ S_{YY} is a symmetric matrix.

Table 3.6. Monthly statistics for the Logan River (1915-1978).

Month	Mean \bar{y} (ac-ft)	Percentage of mean annual flow in month	Standard Deviation s (ac-ft)	Coefficient of varia- tion CV	Skew coefficient g	Lag-one Month autocorrelation $r(1)$	Correlation between monthly and annual flows s_{YX}
O	9,241	5.12	2,142	0.23	-0.23	0.98	0.34
N	7,940	4.40	1,636	0.21	-0.19	0.98	0.38
D	7,330	4.06	1,393	0.19	-0.09	0.94	0.44
J	6,792	3.76	1,192	0.18	-0.07	0.96	0.55
F	6,029	3.34	959	0.16	0.26	0.90	0.59
M	7,452	4.13	1,861	0.25	2.29	0.68	0.60
A	15,378	8.52	6,083	0.40	0.86	0.55	0.58
M	37,820	20.96	12,747	0.34	0.18	0.62	0.85
J	38,526	21.35	16,323	0.42	0.31	0.67	0.91
J	20,659	11.45	8,501	0.41	0.95	0.94	0.88
A	13,055	7.24	3,726	0.29	0.22	0.96	0.96
S	10,218	5.66	2,587	0.25	0.02	0.99	0.97

Matrix of cross-correlations between flows in same water year¹ S_{YY} .

	O	N	D	J	F	M	A	M	J	J	A	S
O	1.00	0.98	0.93	0.88	0.80	0.55	0.11	0.20	0.11	0.13	0.27	0.31
N		1.00	0.94	0.90	0.83	0.58	0.17	0.25	0.13	0.15	0.30	0.34
D			1.00	0.96	0.87	0.61	0.23	0.32	0.20	0.19	0.35	0.39
J				1.00	0.90	0.67	0.30	0.38	0.32	0.32	0.46	0.49
F					1.00	0.68	0.29	0.44	0.36	0.35	0.49	0.54
M						1.00	0.55	0.46	0.40	0.37	0.48	0.57
A							1.00	0.62	0.32	0.31	0.40	0.43
M								1.00	0.67	0.58	0.71	0.75
J									1.00	0.94	0.95	0.93
J										1.00	0.96	0.93
A											1.00	0.99
S												1.00

Matrix of cross-correlations between monthly flows in adjacent years, $t-1$ and t , S_{ZY} .

	O	N	D	J	F	M	A	M	J	J	A	S
O	0.34	0.35	0.35	0.36	0.24	0.18	-0.05	0.04	-0.05	-0.02	0.02	0.06
N	0.38	0.37	0.37	0.38	0.26	0.23	-0.07	0.05	-0.02	0.01	0.05	0.10
D	0.43	0.42	0.41	0.42	0.29	0.23	-0.09	0.07	-0.01	-0.00	0.05	0.09
J	0.53	0.53	0.50	0.52	0.40	0.36	-0.03	0.09	0.03	0.04	0.10	0.14
F	0.58	0.57	0.52	0.52	0.43	0.37	-0.04	0.13	0.04	0.05	0.12	0.17
M	0.55	0.53	0.49	0.45	0.33	0.21	-0.18	0.05	0.05	0.11	0.14	0.17
A	0.50	0.45	0.37	0.30	0.28	0.15	-0.19	-0.04	-0.05	-0.01	0.02	0.03
M	0.77	0.75	0.68	0.63	0.53	0.23	-0.14	0.09	0.13	0.12	0.22	0.24
J	0.90	0.90	0.87	0.81	0.75	0.53	0.12	0.21	0.19	0.19	0.31	0.33
J	0.90	0.88	0.86	0.80	0.75	0.55	0.14	0.20	0.17	0.17	0.30	0.32
A	0.97	0.95	0.92	0.86	0.78	0.55	0.12	0.21	0.15	0.16	0.30	0.33
S	0.98	0.97	0.92	0.87	0.78	0.54	0.10	0.19	0.14	0.16	0.29	0.32

¹ S_{YY} is a symmetric matrix.

Table 3.7. Monthly statistics for the Weber River (1905-1978).

Month	Mean \bar{y} (ac-ft)	Percentage of mean annual flow in month	Standard deviation s (ac-ft)	Coefficient of varia- tion CV	Skew coefficient g	Lag-one month autocorrelation r(1)	Correlation between monthly and annual flows S_{YX}
O	4,817	3.02	1,385	0.29	0.69	0.68	0.38
N	4,163	2.61	997	0.24	0.83	0.84	0.46
D	3,743	2.35	808	0.22	0.45	0.84	0.49
J	3,468	2.18	677	0.20	0.67	0.79	0.48
F	3,174	1.99	604	0.19	0.59	0.88	0.47
M	4,060	2.55	1,243	0.31	3.11	0.64	0.37
A	10,530	6.61	5,060	0.48	1.33	0.64	0.28
M	42,379	26.59	13,867	0.33	0.36	0.39	0.50
J	55,171	34.61	26,012	0.47	0.52	0.13	0.87
J	16,354	10.26	12,967	0.79	3.71	0.66	0.75
A	6,785	4.26	2,430	0.36	0.88	0.79	0.84
S	4,757	2.98	1,458	0.31	0.98	0.86	0.70

Matrix of cross-correlations between flows in same water year¹ S_{YY} .

	O	N	D	J	F	M	A	M	J	J	A	S
O	1.00	0.84	0.68	0.59	0.58	0.49	0.25	0.33	0.19	0.13	0.19	0.25
N		1.00	0.84	0.74	0.68	0.68	0.37	0.43	0.24	0.12	0.22	0.31
D			1.00	0.79	0.73	0.67	0.41	0.37	0.27	0.22	0.28	0.31
J				1.00	0.88	0.69	0.42	0.41	0.28	0.15	0.25	0.27
F					1.00	0.64	0.34	0.36	0.31	0.18	0.22	0.26
M						1.00	0.64	0.39	0.12	0.08	0.04	0.06
A							1.00	0.39	0.03	0.05	0.04	0.02
M								1.00	0.13	0.07	0.26	0.25
J									1.00	0.66	0.76	0.62
J										1.00	0.79	0.59
A											1.00	0.86
S												1.00

Matrix of cross-correlations between monthly flows in adjacent years, t-1 and t, S_{ZY} .

	O	N	D	J	F	M	A	M	J	J	A	S
O	0.33	0.37	0.37	0.49	0.52	0.43	0.09	0.08	0.21	0.09	0.06	0.17
N	0.45	0.45	0.45	0.50	0.54	0.37	0.08	0.15	0.21	0.09	0.09	0.20
D	0.49	0.46	0.45	0.43	0.46	0.33	0.02	0.10	0.22	0.06	0.07	0.15
J	0.44	0.46	0.49	0.50	0.57	0.39	0.11	0.12	0.18	0.00	0.08	0.17
F	0.46	0.50	0.53	0.47	0.57	0.34	0.04	0.14	0.17	-0.07	0.02	0.14
M	0.29	0.25	0.30	0.31	0.45	0.12	-0.11	-0.01	0.18	0.07	0.01	0.05
A	0.14	0.04	0.10	0.09	0.20	-0.04	-0.13	-0.18	0.10	0.08	0.00	-0.02
M	0.19	0.23	0.23	0.30	0.35	0.15	-0.16	0.02	0.15	0.14	0.14	0.12
J	0.43	0.52	0.55	0.49	0.43	0.44	0.18	0.12	0.13	0.13	0.19	0.21
J	0.34	0.36	0.44	0.38	0.26	0.25	0.11	-0.02	0.07	0.12	0.20	0.14
A	0.54	0.56	0.56	0.54	0.38	0.43	0.29	0.19	0.04	0.16	0.24	0.20
S	0.68	0.71	0.64	0.62	0.51	0.60	0.44	0.36	0.05	0.14	0.20	0.13

¹ S_{YY} is a symmetric matrix.

CHAPTER 4

MODELING THE ANNUAL STREAMFLOW TIME SERIES

Introduction

This chapter describes the structure, calibration, and generation procedures for each of the five annual stochastic streamflow models used in this study. In the previous chapter the results of analyzing the streamflow time series were presented. According to the systematic modeling procedure given in Figure 1.1 the next step would be the choice of model type, but this step was omitted so that the performance of all five models could be evaluated and a model choice strategy recommended. The next steps in the systematic modeling procedure are step 3, model form identification, and step 4, parameter estimation. These steps can be described as the model calibration procedure.

The annual streamflow models described in this chapter are: the second order autoregressive model (AR2), the autoregressive moving-average model of the first order for each term (ARMA(1,1) denoted ARMA in this report), the ARMA-Markov model (AMAK), the fast fractional Gaussian noise model (FFGN) and the broken line model (BKL). The structure, calibration, and generation procedure for each model are presented starting with the simplest (AR2) and proceeding in order of increasing complexity.

Second-order Autoregressive Model

Model structure

If the streamflow time series exhibits an autocorrelation structure which decays approximately exponentially the time series can be modeled by an autoregressive model. If the autoregressive order, p , is one, then the autoregressive model of Thomas and Fiering (1962) is the simplest and most widely known model for simulating synthetic streamflows, the model which is defined by Equation 2.12. The autocorrelation term can be expanded to as many lags as needed.

The second order autoregressive (AR2) Markov model was used in this study because all the study streamflows were found to be at least first order autoregressive and one time series, Blacksmith Fork, appeared to be second order autoregressive. By modeling all

the streams with the same order autoregressive model, an equivalent comparison among streams could be made for the autoregressive model. The AR2 model is defined as follows:

$$X_t = \mu + A_1(X_{t-1} - \mu) + A_2(X_{t-2} - \mu) + \epsilon_t \quad (4.1)$$

in which

X_t = streamflow volume during time period t

μ = mean streamflow volume

A_1 = model parameter

A_2 = model parameter

ϵ_t = error term in time period t , distributed normally and independently with zero mean and variance, σ_ϵ^2

Kendall and Stuart (1968) and Jenkins and Watts (1968) have described the parameter estimation procedure for the AR2 model as follows:

$$A_1 = \rho(1) (1 - A_2) \quad (4.2)$$

$$A_2 = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} \quad (4.3)$$

$$\sigma_\epsilon^2 = \{1 - A_1 \rho(1) - A_2 \rho(2)\} \quad (4.4)$$

in which

$\rho(1)$ = lag-one autocorrelation of streamflow X

$\rho(2)$ = lag-two autocorrelation of streamflow X

Table 4.1 contains the AR2 model parameters and a comparison of the historical and generated statistics for the four study streams. The generated statistics are ensemble averages based on 50 synthetic streamflow sequences. Each sequence was

Table 4.1. AR2 model parameters and comparison of historical and generated statistics for study streams.

Stream	Sequence Type	Length of record (yrs)	Number of traces	Model Parameters		Statistics				
				A1	A2	CV ¹	$\bar{\rho}(1)^2$	$\bar{\rho}(1)^3$	\bar{K}^4	\bar{K}^5
Beaver	Historical	64	1	-	-	0.35	0.24	-	0.76	-
	Generated	64	50	0.27	-0.11	0.35	0.19	0.09	0.66	0.06
Blacksmith Fork	Historical	65	1	-	-	0.34	0.49	-	0.77	-
	Generated	65	50	-0.03	0.51	0.34	0.43	0.11	0.74	0.07
Logan	Historical	66	1	-	-	0.26	0.32	-	0.72	-
	Generated	66	50	-0.05	0.34	0.26	0.26	0.11	0.68	0.07
Weber	Historical	74	1	-	-	0.29	0.27	-	0.76	-
	Generated	74	50	0.03	0.26	0.29	0.22	0.10	0.69	0.07

- ¹Coefficient of variation
- ²Average lag-one autocorrelation coefficient
- ³Standard deviation of lag-one autocorrelation coefficient
- ⁴Average Hurst coefficient
- ⁵Standard deviation of Hurst coefficient

equal in length with the historic record used in this study. These results will be discussed in Chapter 7.

Generation procedure

The standardized form of the generation equation for the AR2 model is:

$$Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + T_t \sigma_e \quad (4.5)$$

in which

$$Z_t = (X_t - \mu) / \sigma$$

σ = standard deviation of streamflows

T_t = normally distributed random variate with zero mean and unit variance

The steps used to generate a synthetic streamflow sequence using the AR2 model are described below.

- 1) Input the historic statistics (μ , σ , $\rho(1)$, $\rho(2)$) and the length of the sequence to be generated, N.
- 2) Generate NT = (N + 10) standard normal random variates, T_t (t=1, NT).
- 3) Compute A_2 , A_1 and σ_e^2 using Equations 4.2, 4.3, and 4.4.
- 4) Initialize the first two generated values, Z_1 and Z_2 , to T_1 and T_2 .
- 5) Generate standardized streamflows, Z_t , t = 3, NT, using Equation 4.5 and discard the first 10 generated values in order to eliminate influence of Z_1 and Z_2 being assigned independently when they should be related through Equation 4.5.

6) Estimate mean, \bar{Z} , and standard deviation, \bar{Z} , of the synthetic sequence. Note these will be approximately 0 and 1 respectively.

7) Standardize the synthetic sequence to force it to be exactly zero mean and unit standard deviation as follows:

$$Z'_t = \frac{Z_t - \bar{Z}}{\bar{Z}} \quad (4.6)$$

8) Rescale the standardized synthetic time series to the mean and standard deviation of the historic time series as follows:

$$X_t = Z'_t \sigma + \mu \quad (4.7)$$

The purpose of steps 6 and 7 is to ensure that the synthetic sequence preserves exactly the historic mean and variance. At this stage in the generation procedure the synthetic sequence is in standardized form and therefore, it should have zero mean and unit standard deviation. The actual mean and standard deviation, \bar{Z} and \bar{Z} deviate from 0 and 1, respectively, due to sample error. Application of Equation 4.6 will ensure that \bar{Z} and \bar{Z} are exactly 0 and 1 and Equation 4.7 converts the standardized sequence to a sequence of synthetic streamflows with the historic mean and variance, μ and σ^2 , respectively.

ARMA (1,1) Model

Model structure

O'Connell (1974) evaluated the autoregressive moving average (ARMA) family of models proposed by Box and Jenkins (1970) for their suitability for approximating fractional Gaussian noise (FGN); and recommended the

use of the ARMA (1,1) model, which has first order autoregressive and moving average terms, to preserve long-term persistence as represented by the Hurst effect. To accomplish this, the AR parameter (ϕ) must have a value close to unity, so that the autocorrelation function (ACF) of the process will attenuate slowly and hence approximate the theoretical ACF (TACF) of FGN.

The autocorrelation function of the ARMA (1,1) process at lag-one is defined by the autoregressive and moving average parameters, ϕ and θ , respectively as follows:

$$\rho_{AM} = \frac{(\phi - \theta)(1 - \phi\theta)}{1 + \theta^2 - 2\phi\theta} \quad (4.8)$$

in which

ρ_{AM} = lag-one autocorrelation coefficient of ARMA (1,1) process which is used to preserve the value of $\rho(1)$ estimated from the historic record

ϕ = the first-order autoregressive parameter

θ = the first-order moving average parameter

At higher lags, l , the autocorrelation function depends only on autoregressive parameter ϕ as follows:

$$\rho(l) = \phi^l \quad l \geq 2 \quad (4.9)$$

The ARMA (1,1) model is defined as follows:

$$X_t = \mu + \phi(X_{t-1} - \mu) + \epsilon_t - \theta\epsilon_{t-1} \quad (4.10)$$

O'Connell (1974) found that the values of ϕ and θ that are used for modeling streamflow time series are in the range where both parameters are positive and ϕ is greater than θ . The relationship ($\phi - \theta$) determines the sign of the lag-one autocorrelation, ρ_{AM} , and for this range of values ρ_{AM} is always positive, a limiting hydrologic characteristic. When $\phi = \theta$ the process degenerates into a white noise or completely random process.

O'Connell (1974) also found the following interesting parameter characteristics:

1) If ϕ is held constant, as θ decreases ρ_{AM} increases.

2) If θ is constant, as ϕ increases the Hurst coefficient increases.

3) A high ϕ and θ gives a low ρ_{AM} .

4) As θ approaches zero, ρ_{AM} approaches ϕ and the ARMA (1,1) process approaches an ARMA (1,0 or AR(1) process.

5) As ϕ decreases, emphasis on low frequencies switches to high frequencies i.e., the Hurst coefficient and ρ_{AM} increases.

6) ϕ must be greater than ρ_{AM} .

7) The range of ρ_{AM} that can be modeled decreases as ϕ decreases.

8) The quality of the approximation decreases as ϕ exceeds 0.95 particularly for large values of ρ_{AM} .

9) For $\phi = 0.92$ and $\theta = 0.82$ the Hurst estimator K was unbiased at 0.7.

10) There is an envelope of feasible Hurst estimates, H or K , and $\rho(1)$ combinations that can be modeled by all fractional Gaussian noise models and their approximations, (shown in Figure 4.1) which was derived experimentally by O'Connell (1974). The $\rho(1)$ - K values for the four study streams are plotted on Figure 4.1 and it will be observed that all these points fall within the feasible region for the ARMA (1,1) model.

The parameters of the ARMA model ϕ and θ which preserve both $\rho(1)$ and the Hurst coefficient were derived by O'Connell on the basis of a large number of Monte Carlo experiments. Thus, the appropriate values for ϕ and θ can be obtained from tables given by O'Connell (1974) based on the estimated values of $\rho(1)$ and the Hurst coefficient.

Calibration procedure

The calibration procedure involves interpolating values of ϕ and θ from Tables 3.2 through 3.7 of O'Connell (1974). These tables do not apply to the same length of synthetic sequences as was used in this study. Therefore, it was necessary to refine the interpolated values of ϕ and θ by Monte Carlo generation based on a sequence length equal to the historic record and on the criterion of preserving $\rho(1)$ and the K estimate of the Hurst coefficient. Both $\rho(1)$ and K were calculated as ensemble averages. Table 4.2 contains a comparison of historical and generated statistics for several alternative sets of ARMA model parameters. The model parameter values selected for use in this study are labeled "generated 1". The $\rho(1)$ - K values preserved by this model and the historical values are plotted on Figure 4.1 for each stream.

Generation procedure

The standardized form of the generating equation for the ARMA (1,1) model is:

$$Z_t = \phi Z_{t-1} + (T_t - \theta T_{t-1})\sigma_\epsilon \quad (4.11)$$

in which

$$\sigma_\epsilon^2 = \left(\frac{1 - \phi^2}{1 - \theta^2 - 2\phi\theta} \right) \sigma^2 \quad (4.12)$$

Table 4.2. ARMA(1,1) model parameters and comparison of historical and generated statistics for study streams.

Stream	Type and Run No. ⁶	Length of record (yrs) NR	Number of traces NS	Model Parameters		Statistics				
				ϕ	θ	CV ¹	$\bar{\rho}(1)$ ²	$\bar{\sigma}(1)$ ³	\bar{K} ⁴	\bar{K}^5
Beaver	Historical	64	1	-	-	0.35	0.24	-	0.76	-
	Generated 1	64	50	0.96	0.76	0.35	0.20	0.15	0.78	0.07
	2	64	10	0.75	0.50	0.35	0.22	0.12	0.72	0.06
Blacksmith Fork	Historical	65	1	-	-	0.34	0.49	-	0.77	-
	Generated 1	65	50	0.92	0.52	0.34	0.49	0.16	0.84	0.06
	2	65	50	0.88	0.52	0.34	0.41	0.16	0.81	0.07
	3	65	10	0.88	0.52	0.34	0.38	0.13	0.79	0.05
	4	65	10	0.92	0.52	0.34	0.46	0.14	0.83	0.04
Logan	Historical	66	1	-	-	0.26	0.32	-	0.72	-
	Generated 1	66	50	0.75	0.45	0.26	0.28	0.13	0.74	0.07
	2	66	50	0.92	0.64	0.26	0.33	0.16	0.80	0.07
	3	66	50	0.92	0.64	0.26	0.29	0.14	0.79	0.04
	4	66	10	0.88	0.60	0.26	0.28	0.13	0.77	0.05
Weber	Historical	74	1	-	-	0.29	0.27	-	0.79	-
	Generated 1	74	50	0.96	0.72	0.29	0.30	0.16	0.80	0.07
	2	74	10	0.96	0.72	0.29	0.25	0.15	0.79	0.04

- ¹ Coefficient of variation
- ² Average lag-one correlation coefficient
- ³ Standard deviation of lag-one correlation coefficient
- ⁴ Average Hurst coefficient
- ⁵ Standard deviation of Hurst coefficient
- ⁶ Run 1 is the final calibrated model used for generation

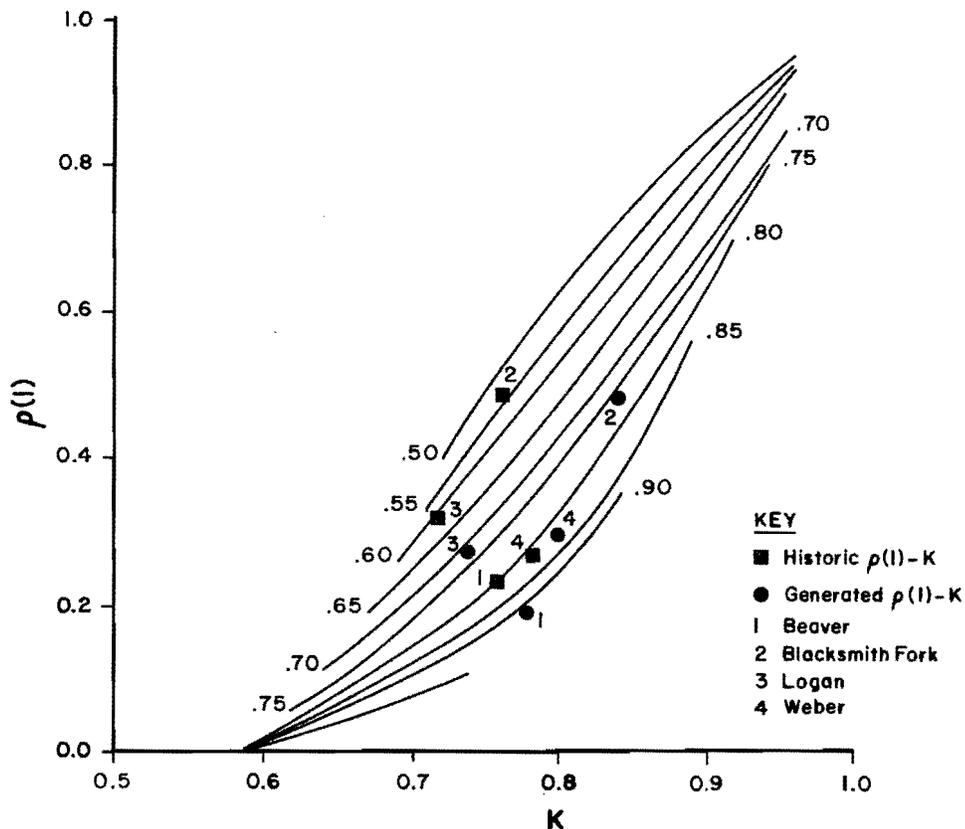


Figure 4.1. Feasible range of $\rho(1)$ -K for ARMA(1,1) models (after Burges and Lettenmaier 1975).

The procedure for generating the ARMA (1,1) process once the parameters μ , σ , ϕ , θ are estimated are summarized below:

- 1) Input the historic statistics and model parameters (μ , σ , ϕ , θ ,) and the length of sequence to be generated, N.
- 2) Generate NT (= N + 10) standard normal random numbers, T.
- 3) Compute σ_e^2 using Equation 4.12.
- 4) Initialize the first generated value, Z_1 , to T1
- 5) Generate standardized streamflows, Z_t , $t = 2, NT$, using Equation 4.11 and discard the first 10 generated values.
- 6) Estimate mean, \bar{Z} , and standard deviation, \bar{Z} , of the synthetic sequence. Note these will be approximately 0 and 1 respectively.
- 7) Standardize the synthetic sequence to force it to be exactly zero mean and unit standard deviation using Equation 4.6.
- 8) Rescale the standardized synthetic time series to the mean and standard deviation of the historic time series using Equation 4.7.

ARMA-Markov Model

Model structure

The ARMA-Markov Model (AMAK) was developed by Lettenmaier and Burges (1977) as an alternative approximation for fractional Gaussian noise. It is a combination of the ARMA (1,1) model used by O'Connell (1974) and the Markov or first order autoregressive model. The AMAK model attempts to satisfy the requirements for modeling both high and low frequency persistence as well as being economical to use in terms of computer time. An advantage of the AMAK model is that the Hurst coefficient, h, is an explicit parameter as it is for FGN models.

The ARMA (1,1) process has the advantage that the decay rate of the autocorrelation function is not dependent on the lag one autocorrelation coefficient. Lettenmaier and Burges (1977) used this feature in the AMAK model to maintain long term persistence at high lags by adjusting the parameters in their model to fit the theoretical autocorrelation function (ACF) of FGN at three arbitrarily selected lags. The theoretical ACF for FGN is:

$$\rho_f(l, h) = 0.5[(l+1)^{2h} - 2l^{2h} + (l-1)^{2h}] \quad (4.13)$$

in which

l = lag

h = Hurst coefficient

ρ_f = theoretical autocorrelation function for FGN

At large lags the theoretical ACF for FGN can be approximated by:

$$\rho_f(l, h) = h(2h-1)l^{2h-2} \quad (4.14)$$

The AMAK model is defined as follows:

$$X_t = \mu + \rho_M(X_{t-1}^{(M)} - \mu) + \epsilon_t^{(M)} + \phi(X_{t-1}^{(AM)} - \mu) +$$

$$\epsilon_t^{(AM)} - \phi \epsilon_{t-1}^{(AM)} \quad (4.15)$$

in which

M = label for Markov terms

AM = label for ARMA(1,1) terms

$\epsilon_t^{(M)}$ and $\epsilon_t^{(AM)}$ are independent processes having the following variances:

$$\sigma_{\epsilon}^2(M) = C_1(1 - \rho_M^2) \quad (4.16)$$

$$\sigma_{\epsilon}^2(AM) = C_2 \frac{(1 - \phi^2)}{(1 + \phi^2 - 2\phi\theta)} \quad (4.17)$$

in which

C_1 = fraction of variance explained by Markov component

C_2 = fraction of variance explained by ARMA (1,1) component

Calibration to study streams

Three alternative methods of parameter estimation for the AMAK model were used in the study. Each method is described below and is denoted by the name of its originator. The first two methods are designed to fit the autocorrelation function of the AMAK model to the theoretical autocorrelation function of FGN (see Equations 4.13 and 4.14). The third method, which is proposed herein, attempts to preserve only $\rho(1)$ and the Hurst coefficient. The parameters to be estimated are C_1 , C_2 , ρ_M , ρ_{AM} and ϕ and each is correlated to lie between 0 and 1.

Lettenmaier and Burges' method. The LB method was proposed by the originators of the AMAK model (Lettenmaier and Burges 1977) and is based on fitting the TACF of FGN at 3 arbitrary lags l_1 , l_2 and l_3 . Lettenmaier and Burges found it convenient to use lags of $N/8$, $N/2$, and N where N is the length of the sequence being generated. Parameter estimation by the LB method requires the solution of the five simultaneous equations

$$C_1 + C_2 = 1 \quad (4.18)$$

$$\rho(1) = C_1 \rho_M + C_2 \rho_{AM} \quad (4.19)$$

$$\rho_f(l_1, h) = C_1 \rho_M^{|l_1|} + C_2 \rho_{AM}^{\phi |l_1-1|} \quad (4.20)$$

$$\rho_f(l_2, h) = C_1 \rho_M^{|l_2|} + C_2 \rho_{AM} \phi^{|l_2-1|} \dots \quad (4.21)$$

$$\rho_f(l_3, h) = C_1 \rho_M^{|l_3|} + C_2 \rho_{AM} \phi^{|l_3-1|} \dots \quad (4.22)$$

θ , the moving average parameter in the ARMA (1,1) process, is obtained from Equation 4.8 using the estimated values of ϕ and ρ_{AM} . Equations 4.18 through 4.22 are solved simultaneously using Newton's method to give parameter estimates based on the values of $\rho(1)$ and h estimated from the historic record.

Kottegoda's method. The second parameter estimation procedure for the AMAK model was recommended by Kottegoda (1980). The method is based on a least squares fitting of the first nine lags of the TACF of FGN to the ACF of the AMAK model. The Kottegoda (K) method was implemented as follows:

- 1) Select ϕ and c_1 values.
- 2) Vary ρ_M and ρ_{AM} values over range of feasible values.
- 3) Calculate theoretical autocorrelation function of AMAK model at first nine lags, as follows:

$$\rho_{AMAK}^{(l)} = C_1 \rho_M^{|l|} + C_2 \rho_{AM} \phi^{|l|-1}, \quad l = 1, 9 \quad (4.23)$$

- 4) By least squares comparison of TACF of AMAK with TACF of FGN and select best values of ρ_M and ρ_{AM} for given ϕ and C_1 values.

- 5) Repeat steps 1 through 4 for other combinations of ϕ and C_1 values and select best parameter set for fitting TACF of FGN. Calculate C_2 using Equation 4.18. This trial and error method is not as costly as the first method in terms of computer time but requires more user time. The fitting of the AMAK TACF to the FGN TACF could be automated using a curve fitting procedure.

James' method. The third method of parameter estimation procedure is proposed herein. It involves the following steps:

- 1) Set ρ_M equal to $\rho(1)$.
- 2) Select ϕ and θ using O'Connell's (1974) parameter estimation procedure for the ARMA (1,1) model which is based on preserving $\rho(1)$ and K estimated from the historic record.
- 3) Set C_1 by trial and error fitting of $\rho(1)$ and K based on analysis of the values of $\rho(1)$ and K preserved in generated sequences. Note that C_2 is given by Equation 4.18.

A comparison of the results of the three parameter estimation procedures is given in Table 4.3. For each generated case the method of parameter estimation is indicated

in the last column. In several cases the James method was initialized using parameter estimates from one of the other methods (denoted by LB+J or K+J).

Comparison of the three parameter estimation methods indicates that the James' method generally provides parameter estimates which lead to a better preservation of $\rho(1)$ and K values. Experience demonstrated that the James' method was the easiest to apply. The LB method appeared to preserve lower values of $\rho(1)$ and in several cases it gave very little weight to the Markov component because of the small values of C_1 . The values of $\rho(1)$ and K preserved in the generated sequences are not very sensitive to the values assigned to C_1 and C_2 : for example compare runs 4 and 5 for Blacksmith Fork. Figure 4.2 gives the feasible region for the AMAK model in terms of $\rho(1)$ and K and also shows the historic $\rho(1)$ - K values for the study streams. From this figure it will be seen that the generated persistence statistics are generally very close to their historic values.

Generation procedure

The Markov and ARMA (1,1) components of the AMAK model must be generated separately because the lagged error terms for the ARMA (1,1) model, $\epsilon_t^{(AM)}$, must be kept separate from the Markov error term. Thus, the generation equations for the AMAK model, written in their standard form, are given as follows:

$$Z_t^{(M)} = \rho_M Z_{t-1}^{(M)} + T_t^{(M)} \sigma_\epsilon^{(M)} \dots \quad (4.24)$$

$$Z_t^{(AM)} = \phi Z_{t-1}^{(AM)} + T_t^{(AM)} \sigma_\epsilon^{(AM)} - \sigma T_t^{(AM)} \sigma_\epsilon^{(AM)} \quad (4.25)$$

and the standardized generated streamflows are obtained by summing these two components, as follows:

$$Z_t^{(AMAK)} = Z_t^{(M)} + Z_t^{(AM)} \dots \quad (4.26)$$

The generating process for the AMAK model is described in the following steps:

- 1) Input the parameters μ , σ , C_1 , C_2 , ρ_M , ϕ and θ and the length of sequence to be generated, N .
- 2) Generate standard normal random variates, $T_t^{(M)}$, $T_t^{(AM)}$ ($t=1, NT$) ($NT = N+10$).
- 3) Compute the error variances, $\sigma_\epsilon^{2(M)}$ and $\sigma_\epsilon^{2(AM)}$ using Equations 4.16 and 4.17.
- 4) Initialize the first generated values $Z_1^{(M)}$ and $Z_1^{(AM)}$ to $T_1^{(M)}$ and $T_1^{(AM)}$ respectively.
- 5) Generate NT standardized Markov variables using Equation 4.24 and discard the first 10 generated values in order to elimi-

Table 4.3. AMAK model parameters and comparison of historical and generated statistics for study streams.

Stream	Type and Run No. ⁷	Length of record (yrs) NR	Number of traces NS	Model Parameters						Statistics					Parameter Estimation Method ⁶
				C1	C2	ρ_M	ρ_{AM}	ϕ	θ^9	CV ¹	$\bar{\rho}(1)^2$	$\bar{\rho}(1)^3$	\bar{K}^4	\bar{K}^5	
Beaver	Historical	64	1	-	-	-	-	-	-	0.35	0.24	-	0.76	-	
	Generated 1	64	50	0.40	0.60	0.24	0.43	0.96	0.76	0.35	0.23	0.15	0.73	0.08	J
	2	64	10	0.38	0.62	0.24	0.17	0.67	0.47	0.35	0.23	0.11	0.69	0.08	LB + J
	3	65	10	0.02	0.98	0.98	0.23	0.67	0.47	0.35	0.19	0.11	0.71	0.09	LB
Blacksmith Fork	Historical	65	1	-	-	-	-	-	-	0.34	0.49	-	0.77	-	
	Generated 1	65	50	0.40	0.60	0.49	0.43	0.92	0.52	0.34	0.46	0.12	0.80	0.07	J
	2	65	10	0.38	0.62	0.49	0.40	0.96	0.56	0.34	0.52	0.12	0.80	0.08	K
	3	65	10	0.38	0.62	0.49	0.40	0.92	0.52	0.34	0.50	0.12	0.80	0.08	K + J
	4	65	10	0.38	0.62	0.49	0.40	0.88	0.52	0.34	0.46	0.13	0.77	0.08	K + J
	5	65	10	0.62	0.38	0.49	0.40	0.88	0.52	0.34	0.47	0.11	0.76	0.08	K + J
	6	65	10	0.80	0.20	0.49	0.40	0.88	0.52	0.34	0.47	0.09	0.74	0.09	K + J
	7	65	10	0.38	0.32	0.49	0.37	0.96	0.56	0.34	0.51	0.12	0.78	0.09	K + J
	8	65	10	0.08	0.92	0.99	0.44	0.73	0.36	0.34	0.38	0.14	0.75	0.09	LB
9	65	10	0.42	0.58	0.73	0.16	0.98	0.59	0.34	0.65	0.11	0.81	0.09	LB ⁸	
Logan	Historical	66	1	-	-	-	-	-	-	0.26	0.32	-	0.72	0	
	Generated 1	66	50	0.40	0.60	0.32	0.36	0.75	0.45	0.26	0.30	0.12	0.72	0.06	J
	2	65	10	0.38	0.62	0.31	0.32	0.91	0.72	0.26	0.27	0.13	0.73	0.09	K
	3	65	10	0.07	0.92	0.99	0.24	0.78	0.59	0.26	0.19	0.13	0.72	0.09	LB
Weber	Historical	74	1	-	-	-	-	-	-	0.29	0.27	-	0.79	-	
	Generated 1	74	50	0.40	0.60	0.26	0.47	0.96	0.72	0.29	0.27	0.12	0.77	0.06	J
	2	65	10	0.23	0.77	0.99	0.29	0.86	0.66	0.29	0.22	0.16	0.74	0.09	LB
	3	65	10	0.38	0.62	0.26	0.60	0.86	0.66	0.29	0.25	0.12	0.71	0.08	LB + J

¹Coefficient of variation

²Average lag one autocorrelation coefficient

³Standard deviation of lag one autocorrelation coefficient

⁴Average Hurst coefficient

⁵Standard deviation of Hurst coefficient

⁶Parameter estimation: LB = Lettenmaier and Burges' method, K = Kottegoda's method, J = James' method

⁷Run 1 is the final calibrated model used for generation

⁸Parameters from Table 1c in Lettenmaier and Burges (1977)

⁹Obtained from ρ_{AM} and ϕ using Equation 4.8

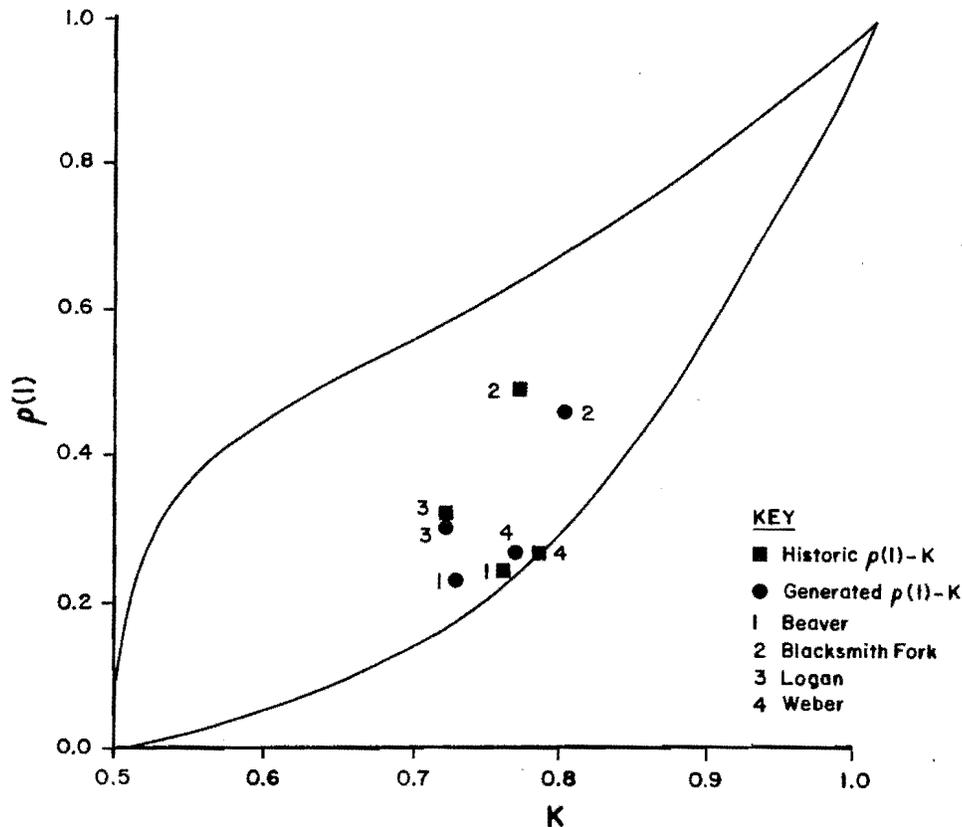


Figure 4.2. Feasible range of $\rho(1)-K$ values for an ARMA-Markov (AMAK) model with $\lambda_1 = 5$, $\lambda_2 = 20$, $\lambda_3 = 40$ (after Lettenmaier and Burges 1977).

nate influence of Z_1 and Z_2 being assigned independently when they should be related through Equation 4.24.

6) Generate NT standardized ARMA (1,1) variates using Equation 4.25 and discard the first 10 generated values in order to eliminate influence of Z_1 and Z_2 being assigned independently when they should be related through Equation 4.25

7) Combine the Markov and ARMA sequences using Equation 4.26.

8) Estimate mean, \bar{Z} , and standard deviation, \bar{Z} , of the synthetic sequence. Note these will be approximately 0 and 1 respectively.

9) Standardize the synthetic sequence to force it to be exactly zero mean and unit standard deviation using Equation 4.6.

10) Rescale the standardized synthetic time series to the mean and standard deviation of the historic time series using Equation 4.7.

Fast Fractional Gaussian Noise Model

Model structure

A concise history of the development of fractional Gaussian noise models can be found in O'Connell (1974). Fractional Gaussian noise is defined as the smoothed derivative of fraction Brownian motion. A Brownian motion process is a stochastic process, $B(t)$, defined in continuous time, such that its increments $B(t+u) - B(t)$ are Gaussian with zero mean and variance u and are independent for nonoverlapping time intervals. A fractional Brownian motion can be defined as the moving average of the incremental continuous time process $dB(t) = B(t+dt) - B(t)$ in which past increments of $B(t)$, $dB(s)$, are weighted by $(t-s)^{h-0.5}$; where h is the Hurst coefficient, (Lawrance and Kottegoda, 1977). Mandelbrot and VanNess (1968) define fractional Brownian motion (fBm), $B_h(t)$, as a function of $B(t)$ as follows:

$$B_h(t_2) - B_h(t_1) = \frac{1}{\sqrt{h+0.5}} \left[\int_{-\infty}^{t_2} (t_2-u)^{h-0.5} dB(u) \right]$$

$$- \int_{-\infty}^{t_1} (t_1 - u)^{h-0.5} dB(u) \dots \dots \dots (4.27)$$

in which dB(u) is an infinitesimal increment of ordinary Brownian motion. Each increment of fractional Brownian motion is a weighted average of all past increments of a Brownian motion process. The weighting function (t-u)^{h-0.5} is a function of the time increment and the Hurst coefficient which results in the present increment exerting a nonnegligible influence on all future increments, a property which is called infinite memory. Another important property of the fractional Brownian motion process is that the increments are self-similar, that is, the process over two intervals are generated by the same probabilistic mechanism. The self-similarity of the increment requires that the expected values of the Hurst coefficient be constant over all intervals. The increments are also Gaussian and independent (i.e. white noise) with zero mean and variance equal to the increment length. The discrete time fractional Gaussian noise (dfGN) models were developed by Mandelbrot and Wallis (1968) but were quite cumbersome and expensive to operate. Mandelbrot (1971) developed an approximation to the autocorrelation function of the dfGN (Equations 4.13 and 4.14), and called it fast fractional Gaussian noise, FFGN. The model is essentially a sum of high and low frequency terms, the high frequency represented by a lag-one Markov process and the low frequency represented by a weighted sum of several lag-one Markov processes specified by a choice of two parameters called the base B and the number of low frequency terms L.

The FFGN model is defined by the following equations:

$$X_t^{(HF)} = \mu + \rho(1)^{(HF)} (X_t^{(HF)} - \mu) + T_t^{(HF)} (1 - \rho(1)^{(HF)2})^{1/2} \dots \dots \dots (4.28)$$

$$X_t^{(LF)} = \mu + \sum_{i=1}^L W_i (C_i(X_{i,t-1}^{(LF)} - \mu) + T_{i,t}^{(LF)} (1 - C_i^2)^{1/2}) \dots \dots \dots (4.29)$$

$$X_t = \mu + \sigma^{(HF)} (X_t^{(HF)} - \mu) + (X_t^{(LF)} - \mu) \dots \dots \dots (4.30)$$

- in which
- HF = label for high frequency terms
 - LF = label for low frequency terms
 - L = number of low frequency terms
 - W_i = weight for ith low frequency term or contribution of vari-

ance of ith low frequency term to unit variance of FFGN process, X

$$= \left(\frac{h(2h-1)(B^{1-h} - B^{h-1})B^{2(h-1)i}}{\Gamma(3-2h)} \right)^{1/2} \dots \dots \dots (4.31)$$

- B = base
- Γ = gamma function
- ρ(1)^(HF) = lag-one autocorrelation coefficient of high frequency term

$$\rho_f(1,h) = \frac{\sum_{i=1}^L W_i^2 C_i}{1 - \sum_{i=1}^L W_i^2} \dots \dots \dots (4.32)$$

- C_i = lag-one autocorrelation coefficient of ith low frequency term
- = e^{-B⁻ⁱ} \dots \dots \dots (4.33)

$$\sigma^2(HF) = \text{residual variance of high frequency component}$$

$$= 1 - \frac{B^{(h-1)}h(2h-h)}{\Gamma(3-2h)} \dots \dots \dots (4.34)$$

The quality factor of the FFGN model, Q, is a function of L, B, and N as follows:

$$L = \left\| \frac{\log(Q*N)}{\log B} \right\| \dots \dots \dots (4.35)$$

in which

- ||X|| denotes the smallest integer above X
- N = length of generated sequence
- Q = quality factor

Calibration to study streams

Parameter estimation procedures for the FFGN model are fairly simple and begin with estimation of μ, σ, ρ(1), and the Hurst coefficient from the historical record. The parameters to be estimated are B and L. Q is not estimated since it is uniquely determined by L using Equation 4.35 when N and B are specified. Mandelbrot (1971) recommends using a value of B in the range 2, 3, or 4, while Chi, Neal, and Young (1973) suggest setting B in the range 2 to 3 and L equal to 20. In both cases the authors generated a sequence of length N equal to 10,000 to achieve a Type A resemblance. In our study the object is Type B (not to be confused with the base parameter, B) resemblance to preserve ensemble average persistence statistics and the errors in approximation addressed by Mandelbrot and Chi, Neal, and Young are not applicable here. Also Lettenmaier and Burges

(1977) showed that the maximum error in the ACF for FFGN varied over various L, B and h values and that increasing L from 10 to 20 was only advantageous for h greater than 0.85. The value of B used in this study was read from Figure 2 of Chi, Neal, and Young (1973) which indicates that, for L = 20, B = 2 to be best for H = 0.7 and B = 3 for H = 0.8. Chi, Neal, and Young (1973) did not obtain much improvement in accuracy in resemblance with L greater than 20. Lettenmaier and Burges (1977) found that for high values of H, a decrease in L from 20 to 10 can increase accuracy and reduce computer costs. For the study streams the parameters and a comparison of historic and generated statistics are given in Table 4.4. To be consistent with the other annual models the K, rather than the H, estimator was used for the application of the FFGN model. A value of L = 10 was used for the Weber streamflows with the following results:

1) As indicated by Lettenmaier and Burges (1977) the approximation for the lag-one autocorrelation was improved by decreasing L from 20 to 10 for the higher Hurst coefficients, noting Table 4.4, run 5 compared with runs 3 or 7, the Hurst coefficient generated was closer to the historic estimate, K = 0.79, for L = 10 than L = 20.

2) The lag-one autocorrelation and Hurst coefficient were found to be fairly insensitive to the value of B, compare runs 5 to 6 and 3 to 9.

The high frequency lag-one autocorrelation coefficient, $\rho(1)^{(HF)}$, is usually computed using Equation 4.32. If the calculated value of $\rho(1)^{(HF)}$ is outside the feasible region, $0 < \rho(1)^{(HF)} < 1.0$, then the value of B should be adjusted to bring $\rho(1)^{(HF)}$ into the feasible region (Chi, Neal, and Young 1973). In this study it was found that by setting $\rho(1)^{(HF)}$ equal to the historic estimate of $\rho(1)$ the resemblance of $\rho(1)$ and K was generally improved. Only in the last numbered run for each stream (see Table 4.4) was $\rho(1)^{(HF)}$ assigned using Equation 4.32; in all other runs, including the final calibration run in the historic estimate of $\rho(1)$ was used.

The calibration results for the study streams show that the following parameter assignments were made: B = 2.0, L = 10, and $\rho(1)^{(HF)}$ equal to the historic estimate of $\rho(1)$. The feasible region for the FFGN model defined in terms of $\rho(1)$ and K for given values of L and B is shown in Figure 4.3. The higher the value of K the higher the minimum feasible value of $\rho(1)$. Figure 4.3 shows that two of the study streams, Beaver and Weber, fall outside the feasible region. The generated $\rho(1)$ -K points tend to be grouped close to the theoretical autocorrelation function.

Generation procedure

The standardized form of the generation equations for the FFGN model are given as follows:

$$Z_t^{(HF)} = \rho^{(HF)}(1) Z_{t-1}^{(HF)} + T_t^{(HF)} \left(1 - \rho^{(HF)}(1)^2\right)^{\frac{1}{2}} \quad (4.36)$$

$$Z_t^{(LF)} = \sum_{i=1}^L W_i \left(C_i Z_{i,t-1}^{(LF)} + T_{i,t}^{(LF)} (1 - C_i)^{\frac{1}{2}} \right) \quad (4.37)$$

$$Z_t = \sigma^{(HF)} Z_t^{(HF)} + Z_t^{(LF)} \quad (4.38)$$

The generation process for the FFGN model comprises the following steps:

- 1) Input the parameters μ , σ , L, $\rho(1)$, K and the length of sequence to be generated, N.
- 2) Generate NT (= N + 10) standard normal random variates, $T_t^{(HF)}$, t = 1, NT.
- 3) Initialize the first generated value $Z_t^{(HF)}$ to $T_t^{(HF)}$.
- 4) Generate NT standardized high frequency FFGN variates using Equation 4.36 and discard the first 10 generated values in order to eliminate influence of $Z_t^{(HF)}$ and $Z_t^{(LF)}$ being assigned independently when they should be related through Equation 4.36.
- 5) Generate L x NT standard normal variates, $T_{i,t}^{(LF)}$, t = 1, NT, i=1, L.
- 6) Compute W_i and C_i using Equations 4.31 and 4.33, respectively.
- 7) Generate NT low frequency standardized FFGN variates using Equation 4.37 and discard the first 10 generated values in order to eliminate influence of $Z_t^{(LF)}$ and $Z_t^{(HF)}$ being assigned independently when they should be related through Equation 4.37.
- 8) Compute residual variance of high frequency component using Equation 4.34.
- 9) Combine the low and high frequency components into the standardized FFGN process using Equation 4.38.
- 10) Estimate mean, \bar{Z} , and standard deviation, \bar{Z} , of the synthetic sequence. Note these will be approximately 0 and 1 respectively.
- 11) Standardize the synthetic sequence to force it to be exactly zero mean and unit standard deviation using Equation 4.6.

12) Rescale the standardized synthetic time series to the mean and standard deviation of the historic time series using Equation 4.7.

Table 4.4. FFGN model parameters and comparison of historical and generated statistics for study streams.

Stream	Type and Run No. ¹	Length of record (yrs)	Number of traces	Model Parameters				Statistics				
				NR	NS	B	L	$\rho(1)$	h	CV ²	$\bar{\rho}(1)$ ³	$\bar{\rho}(1)$ ⁴
Beaver	Historical	64	1	-	-	-	-	0.35	0.24	-	0.76	-
	Generated 1	64	50	2.0	10	0.24	0.76	0.35	0.33	0.13	0.75	0.07
	2	64	50	2.0	20	0.24	0.76	0.35	0.34	0.11	0.74	0.06
	3	64	10	2.0	10	0.24	0.76	0.35	0.30	0.08	0.74	0.05
	4	65	10	2.0	20	0.24	0.72	0.35	0.32	- ⁷	0.70	- ⁷
	5	65	10	2.0	20	0.24	0.76	0.35	0.33	0.12	0.70	0.08
Blacksmith Fork	Historical	65	1	-	-	-	-	0.34	0.49	-	0.77	-
	Generated 1	65	50	2.0	10	0.49	0.76	0.34	0.44	0.10	0.78	0.06
	2	65	50	2.0	10	0.49	0.74	0.34	0.44	0.10	0.77	0.06
	3	65	50	2.0	10	0.49	0.72	0.34	0.43	0.10	0.76	0.06
	4	65	10	2.0	20	0.49	0.74	0.34	0.46	- ⁷	0.74	- ⁷
	5	65	10	2.0	20	0.49	0.76	0.34	0.46	0.11	0.73	0.08
Logan	Historical	66	1	-	-	-	-	0.26	0.32	-	0.72	-
	Generated 1	66	50	2.0	10	0.32	0.72	0.26	0.34	0.15	0.74	0.07
	2	65	10	2.0	20	0.31	0.73	0.26	0.36	- ⁷	0.71	- ⁷
	3	65	10	2.0	20	0.31	0.72	0.26	0.36	0.11	0.71	0.07
Weber	Historical	74	1	-	-	-	-	0.29	0.27	-	0.79	-
	Generated 1	74	50	2.0	10	0.27	0.78	0.29	0.33	0.12	0.74	0.09
	2	74	50	2.0	10	0.26	0.78	0.29	0.33	0.12	0.74	0.07
	3	65	10	3.3	20	0.26	0.84	0.29	0.35	- ⁷	0.71	- ⁷
	4	65	10	2.0	5	0.26	0.84	0.29	0.30	0.14	0.74	0.07
	5	65	10	2.0	10	0.26	0.84	0.29	0.36	0.14	0.77	0.06
	6	65	10	3.0	10	0.26	0.84	0.29	0.37	0.17	0.78	0.09
	7	65	10	3.0	20	0.30	0.84	0.29	0.37	0.13	0.71	0.08
	8	65	10	3.0	20	0.15	0.84	0.29	0.31	0.13	0.70	0.08
	9	65	10	4.0	20	0.26	0.84	0.29	0.35	0.13	0.71	0.07
	10	65	10	2.0	20	0.10	0.84	0.29	0.32	0.13	0.72	0.08
11	74	50	2.0	10	0.25	0.79	0.29	0.33	0.12	0.74	0.07	

¹Run 1 is the final calibrated model used for generation
²Coefficient of variation
³Average lag-one autocorrelation coefficient
⁴Standard deviation of lag-one autocorrelation coefficient
⁵Average Hurst coefficient
⁶Standard deviation of Hurst coefficient
⁷These statistics are not available

Broken Line Model

Model structure

The general broken line (BKL) flow generating process developed by Mejia, Rodriquez-Iturbe, and Dawdy (1972) is the sum of NL+1 simple broken line processes. The simple broken line process, which is illustrated in Figure 4.4, is derived from linear interpolation between uniformly spaced independent Gaussian variates and is defined as follows (Curry and Bras 1978):

$$YA(t-\lambda a) = \sum_{j=0}^{N'} \frac{(e_j + (e_{j+1} - e_j)(t - ja)) I_{(ja, (j+1)a)}(t)}{a} \quad (4.39)$$

in which

- YA = simple standardized broken line time series
- e_j = identically independent distributed random variables with zero mean and variance, s^2
- λ = random variable uniformly distributed over the interval (0,1) at the origin of time for first spacing interval to provide stationarity
- j = number of intervals in the time series
- a = time interval between the random variables e_j

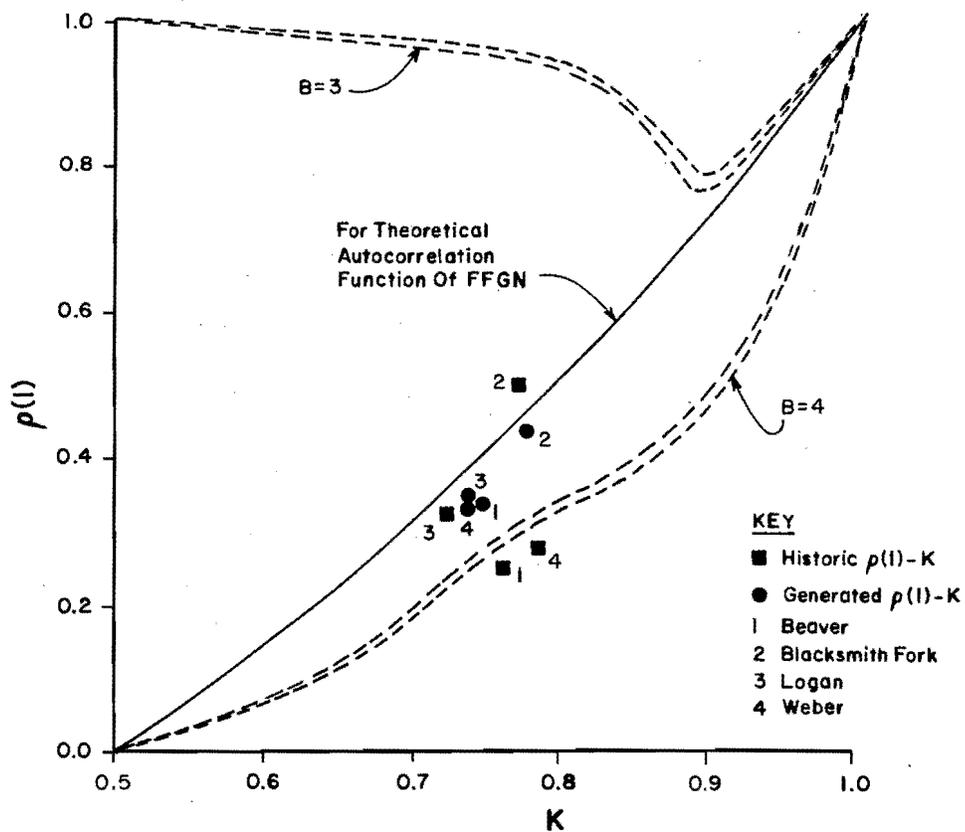


Figure 4.3. Feasible range of $\rho(1)-h$ for FFGN model with $N = 50$ and $Q = 6$ (after Burges and Lettenmaier 1975).

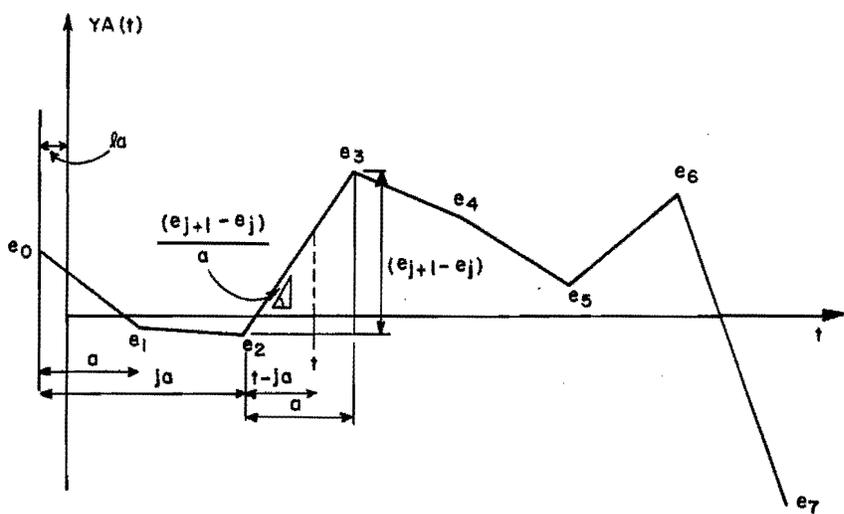


Figure 4.4. A schematic representation of a simple Broken Line process (adapted from O'Connell 1974).

N' = number of time intervals needed to cover required length of synthetic sequence
 $= \frac{N+l}{a}$ rounded up to the nearest integer

$I(ja, (j+1)a)(t)$ = indicator function which keeps the linear interpolation within the correct interval
 $= \begin{cases} 1 & \text{for } ja < t \leq (j+1)a \\ 0 & \text{otherwise} \end{cases}$

The sum of the simple broken line processes is performed vertically and algebraically to produce the general broken line process,

$$X_t = \mu + \sum_{i=1}^{NL} YA_i(t)W_i \quad (4.40)$$

in which

i = index for simple broken line process

NL = number of simple broken lines to be summed

W_i = process weighting function

$$W_i = \left(\frac{a_i^{2(h-1)} BB(B^{(h-1)} - B^{(1-h)}) B^{2(h-1)(i-1)}}{2(h-1)} \right)^{\frac{1}{2}} \quad (4.41)$$

$$W_o = \left(\frac{NL}{1 - \sum_{i=1}^{NL} W_i^2} \right)^{\frac{1}{2}} \quad (4.42)$$

a = the time interval between random variables e_j for the first line

B = base

$$BB = \frac{h(2h-1)(2h-2)(2h-3)(2h-4)(2h-5)}{6(2^{3-2h} - 1)} \quad (4.43)$$

There are two restrictions imposed upon the processes by Mejia, Rodriguez-Iturbe, and Dawdy (1972) in order to reduce the large number of parameters required in the summation:

$$a_i = a_1 B^{(i-1)} \quad (4.44)$$

$$s_i = s_1 \left(\frac{a_1}{a_i} \right)^{\frac{1}{2}} \quad (4.45)$$

in which

s_i = standard deviation of one of the simple broken line processes. Substituting Equation 4.44 into Equation 4.45, s_i is given by:

$$s_i = s_1 B^{\frac{(i-1)}{2}} \quad (4.46)$$

By varying NL , the number of simple broken lines, and W_i the weighting of the individual lines, it is possible to generate a process that has as many degrees of freedom as desired. Thus, it is possible to preserve a memory of the historic hydrologic time series that reproduces the historical correlation structure and Hurst effect (Curry and Bras 1978). The memory of the broken line process is equal to the time lag at which the autocorrelation function becomes zero and is equal to $2a_n$ which is controlled by l and NL . The high frequency properties of the BKL model are a function of the short simple broken lines and the low frequency properties are a function of the long simple broken lines. The parameter a_1 is used to fit the correlation function of the broken line process to the historical lag-one correlation coefficient, $\rho(1)$.

The simple broken line process YA has a mean of zero and a variance $= \frac{2}{3} s^2$ and a function defined as follows:

$$\rho(k, a_i) = \begin{cases} 1 - \frac{3}{4} \left(\frac{k}{a_i} \right)^2 \left(2 - \frac{k}{a_i} \right) & 0 \leq k \leq a_i \\ \frac{1}{4} \left(\frac{2-k}{a_i} \right)^3 & a_i < k \leq 2a_i \\ 0 & 2a_i < k < \infty \end{cases} \quad (4.47)$$

in which

$\rho(k, a_i)$ = correlation of the broken line process at lag k and time interval a_i

Calibration to the study streams

When fitting the BKL model to the lag-one autocorrelation coefficient several modifications to Equation 4.47 are in order according to Curry and Bras (1978). First, the lag-one autocorrelation function for a single broken line with parameter a_1 is given by

$$\rho_1(1, a_1) = \begin{cases} 1 - \frac{3}{4} \left(\frac{1}{a_1} \right) \left(2 - \frac{1}{a_1} \right) & a_1 \geq 1 \\ \frac{1}{4} \left(2 - \frac{1}{a_1} \right)^3 & 0.5 < a_1 < 1 \\ 0 & a_1 \leq 0.5 \end{cases} \quad (4.48)$$

A high frequency simple broken line with parameter a_0 is added to minimize the effect on the Hurst phenomenon caused by the low frequency terms, the high a_i 's. The $\rho(1)$ of the new simple broken line process is given by

$$\rho(1) = \frac{a_1^{2h-2}}{2^{(h-1)}} \sum_{i=0}^{NL-1} BB(B^{h-1} - B^{1-h}) \rho_{i+1}(1, a_1 B^i) B^{2(h-1)i} + \left(1 - \frac{BBa_1^{(2h-2)} B^{(1-h)}}{2-2h}\right) \rho_0(1, a_0) \dots \dots \dots (4.49)$$

The value of a_0 will determine what range of $\rho(1)$ can be satisfied by varying the parameter a_1 . The behavior of Equation 4.49 has been shown by Curry and Bras (1978) to have the following characteristics:

- 1) The lag-one autocorrelation of a single broken line (Equation 4.48) increases as parameter a_1 increases. However, the lag-one autocorrelation of the BKL model, which is derived from Equation 4.49 with $a_0 < 0.5$, decreases as a_1 increases due to the rapidly decreasing weighting of the simple broken lines with parameters $a_1, a_1 B, \dots, a_1 B^{NL-1}$. The value of $\rho(1)$ given by Equation 4.49 will approach $\rho(1, a_0)$ as a_1 increases for any value of a_0 .
- 2) To insure that an arbitrarily small $\rho(1)$ can be preserved, a_0 must be less than or equal to 0.5 so that $\rho(1, a_0) = 0$, Curry and Bras (1978) recommend setting $a_0 = 0.49$.

Parameter estimation for the BKL model requires fitting the historic lag-one autocorrelation and Hurst coefficients to the BKL lag-one autocorrelation coefficient. An alternative fitting procedure is to preserve the $\rho''(0)$, the second derivative of the lag-zero autocorrelation coefficient at the origin instead of the Hurst coefficient. However, since this derivative does not exist for discrete series, such as annual streamflow volumes, this procedure was not followed. The following equation for the BKL lag-one autocorrelation coefficient is solved for by trial-and-error to derive the parameter a_1 :

$$\rho_1 = \frac{\frac{a_1^{(2h-2)}}{(2h-1)} (B^{h-1} - B^{1-h}) \sum_{i=0}^{NL-1} \left\{ \left[1 - \frac{3}{4} \left(\frac{1}{a_1 B^i}\right)^2 \right. \right.}{\left. \left. \left(2 - \frac{1}{a_1 B^i}\right) \right] I_1(a_1 B^i) + \frac{1}{4} \left(2 - \frac{1}{a_1 B^i}\right)^3 I_2(a_1 B^i) \right\} B^{2(h-1)i}}{\dots \dots \dots}$$

$$+ \left[\left(1 - \frac{BBa_1^{(2h-2)} B^{(1-h)}}{2-2h}\right) \left\{ \left(1 - \frac{3}{4} \left(\frac{1}{a_1}\right) \left(2 - \frac{1}{a_1}\right) I_1(a_1) + \frac{1}{4} \left(2 - \frac{1}{a_1}\right)^3 I_2(a_1) \right\} \right. \right. \\ \left. \left. \dots \dots \dots \right] \text{Term 2} \dots \dots \dots (4.50)$$

in which

$$I_1(X) = \begin{cases} 1 & \text{if } X \geq 1 \\ 0 & \text{otherwise} \end{cases} \\ I_2(X) = \begin{cases} 1 & \text{if } 0.5 \leq X \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

An explanation of the derivation of Equation 4.50 can be found in Curry and Bras (p. 37, 1978) along with a discussion of the behavior of the parameters a_1 and a_2 . Essentially the first term in Equation 4.50 results from substituting Equation 4.48 into Equation 4.49 and the second term represents the effect of adding a high frequency broken line with $a_1 = a_0$. Figures 4.5 through 4.7 contain solutions for a_1 obtained from Equation 4.50 for specified values of $\rho(1)$, NL, B and h. For a small number of broken lines (e.g. NL = 4) the value for the Hurst coefficient, h, is less important in determining the parameter a_1 at a given value of the lag-one autocorrelation coefficient. The discontinuity that appears in the function moves to the right (i.e. higher $\rho(1)$ values), when the Hurst coefficient is increased. Increasing the value of NL also moves the discontinuity to the right. Equation 4.50 for estimating a_1 from a given NL, B, h and $\rho(1)$ is programmed and contained in Appendix C. Table 4.5 contains the BKL model parameters and a comparison of the historical and generated persistence statistics for each study stream. Only two parameters, NL and a_1 were found to have much effect on the persistence statistics, $\rho(1)$ and K, of the generated sequences. The procedure to calibrate the model parameters for replication of the historic lag-one autocorrelation and Hurst coefficients is as follows:

- 1) Referring to Figures 4.6 through 4.7 find the value of NL that gives a value of a_1 approximately equal to one and preferably less than two.
- 2) Using the NL value determined from step 1 and using the estimated $\rho(1)$ and K values compute the exact value for the a_1 parameter using Equation 4.50.
- 3) Evaluate term 1 of Equation 4.50 at the value of a_1 obtained in the previous step to obtain the theoretical maximum value for $\rho(1)$ that can be generated with this BKL model. If the historic estimate of $\rho(1)$ exceeds this theoretical maximum value then set a_0 equal to a_1 rather than 0.49.

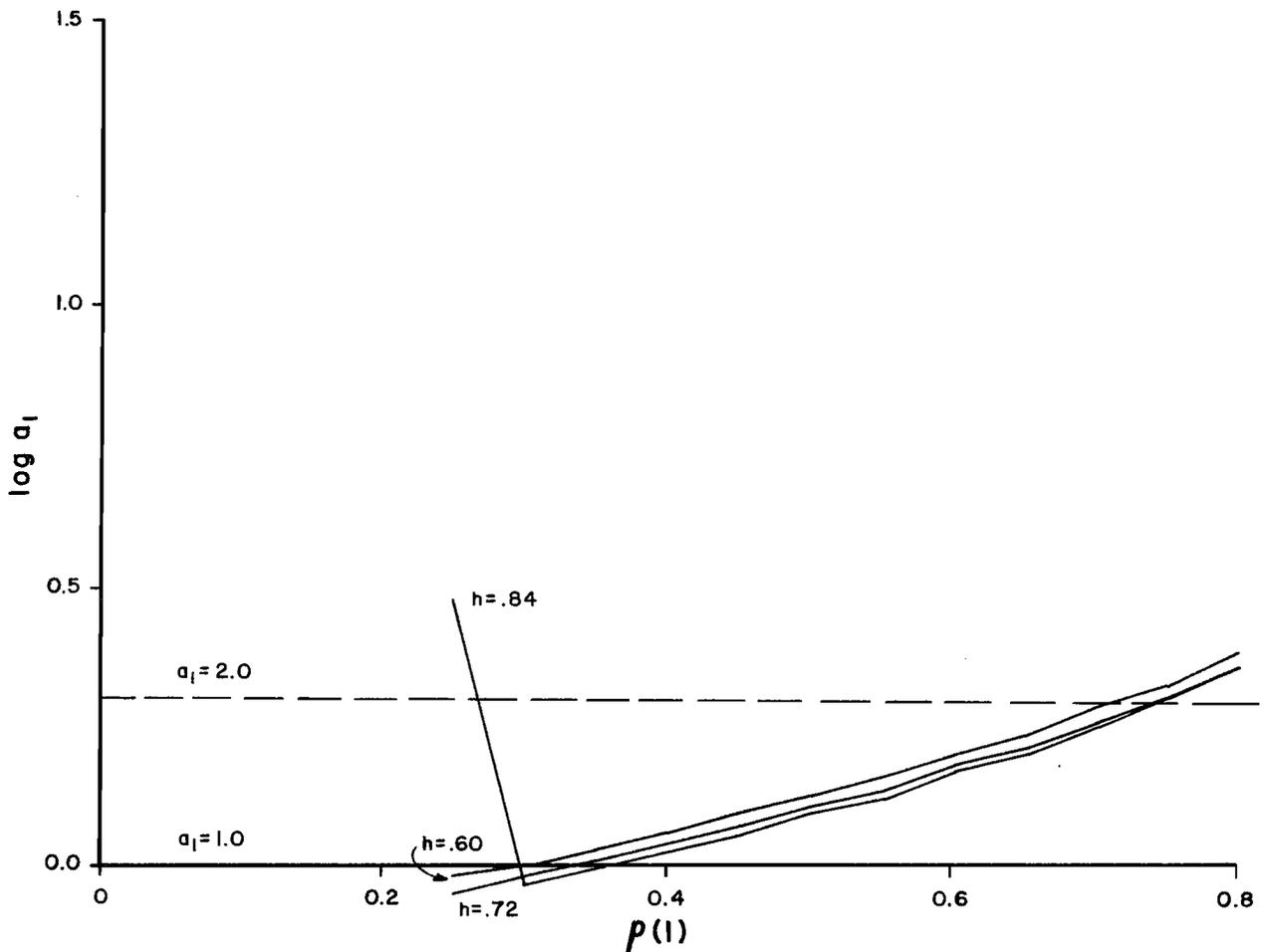


Figure 4.5. Solutions for broken line parameter a_1 using Equation 4.50 for $NL = 4$, $B = 5$.

4) Generate ensemble of sequences (at least 10 and preferably 50 or more) with the BKL model and compute the ensemble averages for $\rho(1)$ and K and compare with historic.

5) If the resemblance is poor, try another NL and a_1 combination.

An alternative calibration procedure is to start out with the lowest NL value of 4 and obtain a_1 from Equation 4.50 and then generate sequences for comparison of the persistence statistics. If the resemblance is poor then try the next higher NL value, until a good resemblance is obtained. However, increasing the value of NL does not necessarily mean an increase in accuracy of preserving $\rho(1)$ and K . Referring to Table 4.5 increasing NL , from 4 to 8 provided a better preservation of $\rho(1)$ for the Logan River (see runs 2 and 1 respectively, however, increasing NL from 4 to 6 did not improve the preservation of $\rho(1)$ and K for the Weber River (see runs 1 and 3 respectively). Experience gained in calibrating the study streams indicated that using a value of NL which maintained the parameter

a_1 between 1 and 2 usually lead to a more accurate preservation of the persistence statistics. The BKL was not capable of preserving the high lag-one autocorrelation coefficient of 0.49 for Blacksmith Fork, unless a_0 was set equal to a_1 (see step 3 of the calibration procedure); since the historic estimate of $\rho(1)$ for Blacksmith Fork exceeds the theoretical maximum value of $\rho(1)$ with the BKL model being used (see run 10, Table 4.5).

Generation procedure

The standardized forms of the generation equations for the BKL model are as follows:

$$YA_i(t - \lambda_i a_i) = \sum_{j=0}^{N'} S \left[\frac{T_{i,j} + (T_{i,j+1} - T_{i,j})(t - ja_i)}{a} \right] I_{(ja_i, (j+1)a_i)}(t) \quad (4.51)$$

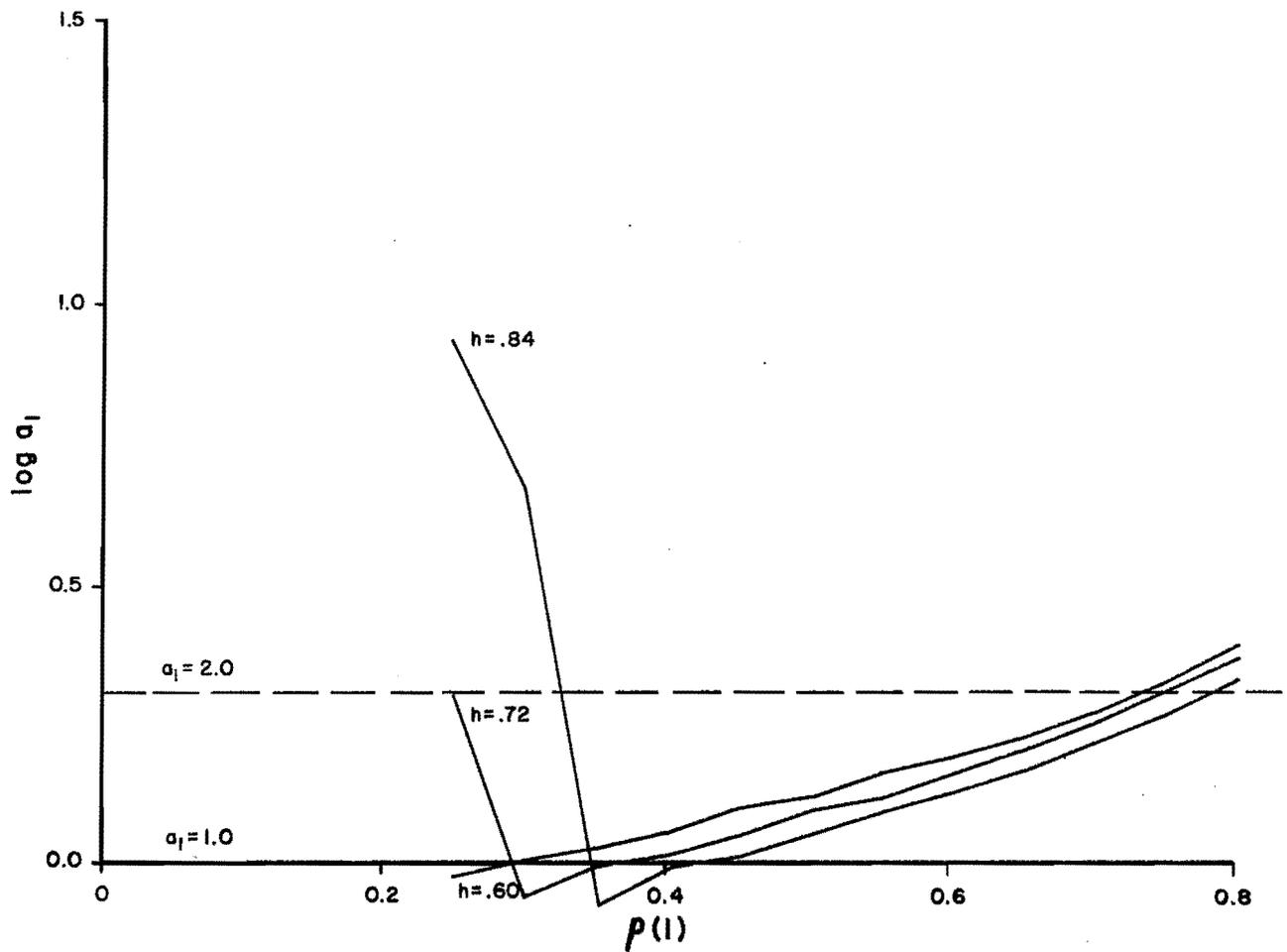


Figure 4.6. Solutions for broken line parameter a_1 using Equation 4.50 for $NL = 5$, $B = 3$.

$$Z_t = \sum_{i=0}^{NL} YA_i(t)W_i \quad \dots \quad (4.52)$$

The generation procedure for the BKL model is summarized in the following steps:

- 1) Input historic statistics and model parameters (μ , σ , K , B , NL , and a_1) and the length of sequence to be generated, N .
- 2) Compute BB using Equation 4.43.
- 3) Compute weighting function W_i using Equation 4.41 and W_0 using Equation 4.42.
- 4) Set the initial time interval $a_0 = 0.49$, the second time interval equal to the input value a_1 , and the succeeding intervals, a_i , using Equation 4.44.
- 5) Generate $NL + 1$ uniform random variates, U_i , $i = 0, NL$ and set origin of each simple broken line at $t = -a_i U_i$.

- 6) Generate N_i ($NL + 1$) standard normal random variates, $T_{i,j}$, $j = 1, N_i$, $i = 1, NL + 1$.

- 7) Generate $NL + 1$ simple broken line processes using Equation 4.51.

- 8) Sum the $NL + 1$ simple broken lines into a general BKL sequence using the weighting function, W_i following Equation 4.52.

- 9) Estimate mean, \bar{Z} , and standard deviation, $\hat{\sigma}$, of the synthetic sequence. Note these will be approximately 0 and 1 respectively.

- 10) Standardize the synthetic sequence to force it to be exactly zero mean and unit standard deviation using Equation 4.6.

- 11) Rescale the standardized synthetic time series to the mean and standard deviation of the historic time series using Equation 4.7.

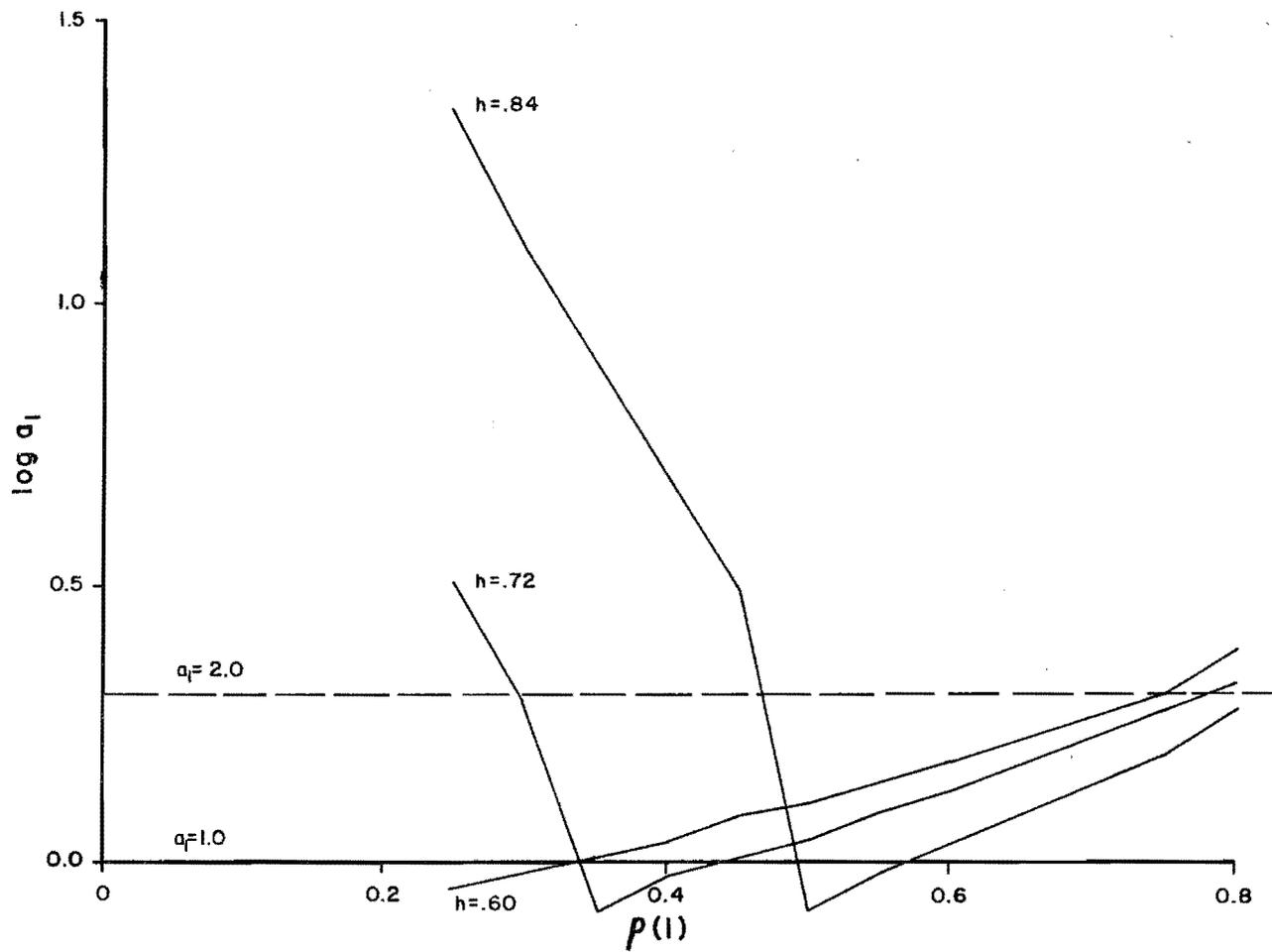


Figure 4.7. Solutions for broken line parameter a_1 using Equation 4.50 for $NL = 8$, $B = 3$.

Table 4.5. Broken line model parameters and comparison of historical and generated statistics for study streams.

Stream	Type and Run No. ¹	Length of record (yrs)	Number of traces	Model Parameters				Statistics				
				NR	NS	B	NL	a ₁	h	CV ²	$\bar{\rho}(1)$ ³	$\bar{\rho}(1)$ ⁴
Beaver	Historical	64	1	-	-	-	-	0.35	0.24	-	0.76	-
	Generated 1	64	50	3.0	4	0.86	0.76	0.35	0.24	0.11	0.72	0.08
	2	64	50	3.0	5	3.89	0.76	0.35	0.16	0.12	0.72	0.06
Blacksmith Fork	Historical	65	1	-	-	-	-	0.34	0.49	-	0.77	-
	Generated 1	65	50	3.0	5	1.16	0.77	0.34	0.48	0.12	0.76	0.07
	2	65	50	3.0	5	1.18	0.74	0.34	0.27	0.13	0.72	0.07
	3	65	50	3.0	5	1.20	0.72	0.34	0.28	0.12	0.74	0.07
	4	65	50	3.0	8	1.02	0.76	0.34	0.22	0.13	0.72	0.08
	5	65	50	3.0	5	2.00	0.76	0.34	0.28	0.14	0.74	0.07
	6	65	10	3.0	5	1.18	0.74	0.34	0.27	0.13	0.74	0.08
	7	65	10	3.0	4	1.26	0.74	0.34	0.24	0.12	0.72	0.06
	8	65	10	3.0	8	1.07	0.74	0.34	0.26	0.11	0.72	0.08
	9	65	50	3.0	5	0.89	0.77	0.34	0.27	0.13	0.73	0.07
10	65	50	3.0	5	1.16	0.76	0.34	0.29	0.14	0.73	0.08	
Logan	Historical	66	1	-	-	-	-	0.26	0.32	-	0.72	-
	Generated 1	66	50	3.0	8	1.77	0.72	0.26	0.24	0.13	0.72	0.07
	2	66	50	3.0	4	0.98	0.72	0.26	0.21	0.15	0.72	0.06
	3	65	10	3.0	8	1.94	0.73	0.26	0.24	0.11	0.73	0.07
4	65	10	3.0	8	0.98	0.73	0.26	0.28	0.10	0.75	0.08	
Weber	Historical	74	1	-	-	-	-	0.29	0.27	-	0.79	-
	Generated 1	74	50	3.0	4	0.88	0.78	0.29	0.28	0.13	0.75	0.07
	2	74	50	3.0	5	3.87	0.78	0.29	0.19	0.17	0.75	0.07
	3	74	50	3.0	6	5.26	0.78	0.29	0.17	0.16	0.74	0.08
	4	65	10	3.0	4	0.88	0.84	0.29	0.31	0.12	0.75	0.05
5	65	10	3.0	8	18.39	0.84	0.29	0.10	0.15	0.74	0.09	

¹Run 1 is the final calibrated model used for generation

²Coefficient of variation

³Average lag one autocorrelation coefficient

⁴Standard deviation of lag one autocorrelation coefficient

⁵Average of Hurst coefficient

⁶Standard deviation of Hurst coefficient

CHAPTER 5

MODELING THE MONTHLY STREAMFLOW TIME SERIES

Introduction

This chapter describes the structure, calibration, and generation procedures for the disaggregation models used in this study. The function of a disaggregation model is to divide a generated sequence of annual flows into seasonal flows while preserving some of the important correlation relationships between seasonal flow volumes and with the annual flow volume. In this study the seasonal level of disaggregation used was a month. Two types of disaggregation models were used and are described in the first section. The next section describes the parameter estimation procedure, calibration experience with the study streams and the calibration results. As with the annual models, calibration includes the model form identification and parameter estimation (steps 3 and 4) of the systematic modeling procedure (Figure 1.1). The final section summarizes the generation procedure for disaggregation.

Model Structure

Two alternative seasonal disaggregation models were used in this study: the Valencia-Schaake (VS) model (Valencia and Schaake 1973) and the Mejia-Rousselle (MR) model (Mejia and Rousselle 1976). Properties of each model were compared in the literature review presented in Chapter 2. The equations for each model are repeated in this section with the seasonal level being months. The VS model is written as follows:

$$\underline{Y}_t = \underline{A}X_t + \underline{B}V_t \quad \dots \quad (5.1)$$

in which

\underline{Y} = n-vector of disaggregated standardized monthly flow volumes

X = standardized annual flow volume

\underline{V} = n-vector of random elements with zero mean and unit variance

\underline{A} = n-vector of model parameters

\underline{B} = nxn matrix of model parameters

n = number of months in a year (i.e. 12)

The MR model is similar to the VS model except that it includes an additional term (\underline{DZ}_t) to preserve some of the serial correlations between adjacent months in successive water years. The equation for the MR model is as follows:

$$\underline{Y}_t = \underline{C}X_t + \underline{DZ}_t + \underline{E}V_t \quad \dots \quad (5.2)$$

in which

\underline{Z}_t = m-vector of standardized flow volumes for the last m months of the previous year

\underline{C} = n-vector of model parameters

\underline{D} = nxm matrix of model parameters

\underline{E} = nxn matrix of model parameters

m = number of months in (t-1)st year which are included in \underline{Z}_t

If the preservation of the serial correlations between adjacent months in successive water years is not important in a particular application then the VS model should be used. However, if these serial correlations are important in terms of the simulation and design for which the disaggregated flows are to be used then the MR model should be used and the extra model parameters in \underline{D} then must be estimated.

Calibration Procedure

In this section the equations for estimating the parameter matrices of the VS and MR models are first presented. In this study we first attempted to apply the MR model to all four study streams but encountered some difficulties which are described together with the procedure for resolving these difficulties. In the final subsection a comparison of the historic and disaggregated monthly flow statistics is presented and discussed.

Parameter estimation

The \underline{A} and \underline{B} parameters of the VS model are estimated as follows:

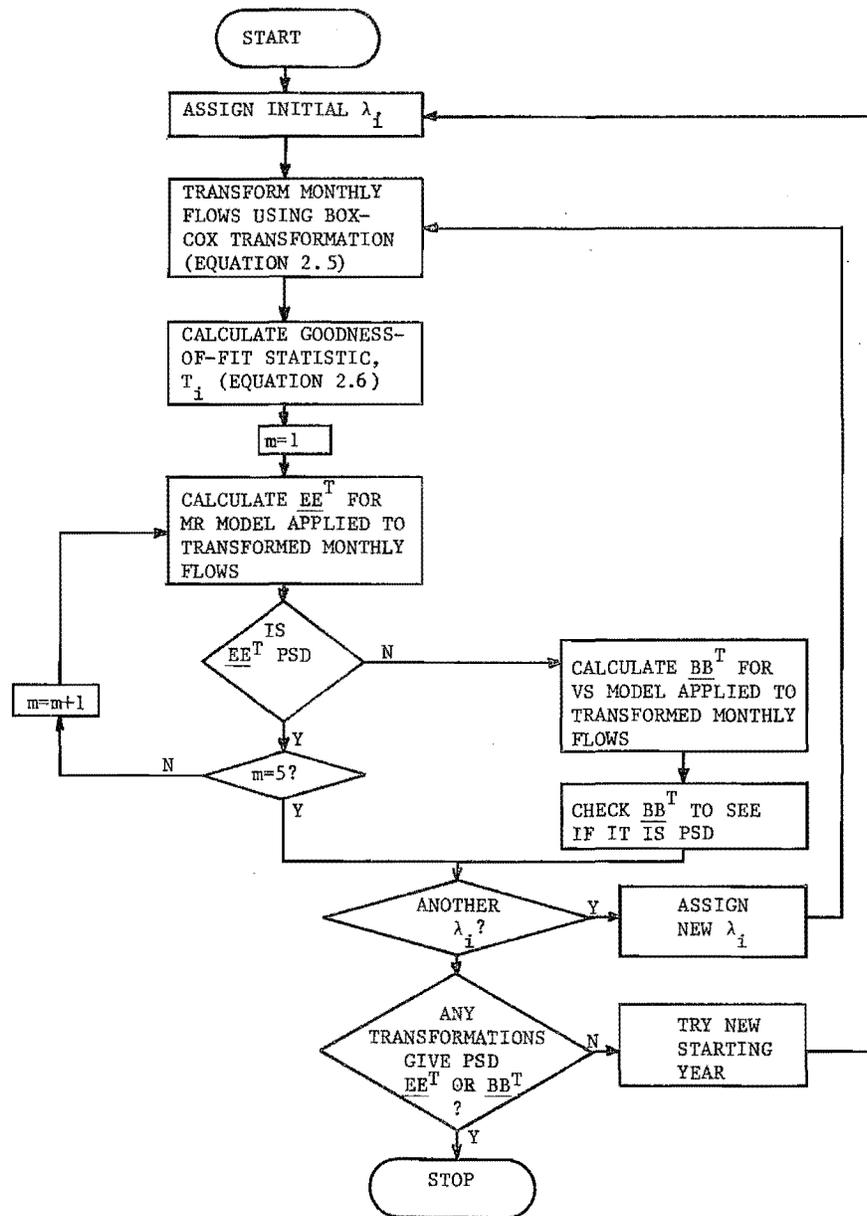


Figure 5.1. Procedure for examining the influence of several factors on parameter estimation for disaggregation models.

high flows (see Chapter 3). Parameter matrices for the selected disaggregation models for each study stream are contained in Tables 5.2 through 5.5.

Examination of Table 5.1 indicates several successful cases of parameter estimation for each stream. The cases selected for use are labeled by footnote d. These cases were selected to keep the same value of λ for all streams for comparative purposes, to keep m as high as possible to give the best preservation of over-the-year serial correlations, and to make use of the longest

length of homogeneous historical record to improve parameter estimates.

Comparison of historic and disaggregated monthly statistics

Figures 5.2 through 5.5 contain graphical comparisons of several historic and disaggregated monthly statistics for each of the four study streams. The disaggregated statistics are based on applying the disaggregation model selected in Table 5.1 to the historical annual flow volumes. The monthly statistics presented are the means,

Table 5.1. Summary of attempts to obtain real-valued parameters for disaggregation models.

Stream	Case No.	Starting year used	Disaggregation model ^a	Number of months in previous year (m)	Box-Cox Transformation Parameter (λ) ^b	Smallest Eigen Value	Goodness-of-fit (T)	
							Average Year	Average growing season
Beaver	1	1915	MR	1	0.333	0.285	-0.040	-0.006
	2	1915	MR	2	0.333	0.185	-0.040	-0.006
	3 ^d	1915	MR	3	0.333	0.054	-0.040	-0.006
	4	1915	MR	4	0.333	-0.181	-0.040	-0.006
Blacksmith Fork	1	1914	MR	1	0.200	-0.113	-0.014	-0.007
	2	1914	MR	1	0.150	-0.048	-0.022	-0.018
	3 ^d	1914	VS	-	- ^c	0.001	0.122	0.157
	4	1914	VS	-	0.333	0.102	0.010	0.022
	5	1924	MR	1	0.333	-0.507	0.010	0.022
	6	1924	MR	1	0.200	-0.044	-0.014	-0.007
	7	1924	MR	1	0.150	-0.017	-0.022	-0.018
	8	1924	MR	1	0.100	-0.007	-0.031	-0.028
Logan	1	1913	MR	1	0.333	0.103	-0.041	-0.078
	2 ^d	1913	MR	2	0.333	0.108	-0.041	-0.078
	3	1913	MR	3	0.333	-0.238	-0.041	-0.078
	4	1913	VS	-	0.333	0.163	-0.041	-0.078
	5	1923	MR	1	0.333	-0.141	-0.041	-0.078
	6	1923	MR	2	0.333	-0.469	-0.041	-0.078
	7	1923	MR	1	0.250	-0.033	-0.052	-0.092
	8	1923	MR	1	0.200	-0.014	-0.059	-0.101
	9	1924	MR	1	0.200	0.010	-0.059	-0.101
	10	1924	MR	2	0.200	0.009	-0.059	-0.101
	11	1924	MR	2	0.333	0.099	-0.041	-0.078
	12	1924	MR	3	0.333	-29.330	-0.041	-0.078
Weber	1	1905	MR	1	0.333	0.774	0.018	-0.033
	2 ^d	1905	MR	2	0.333	0.478	0.018	-0.033
	3	1905	MR	3	0.333	-0.663	0.018	-0.033

^aVS = Valencia-Schaake, MR = Mejia-Rousselle.

^bEigen value for EE^T for MR model and for BB^T VS model.

^cNo transformation used for this case.

^dThis is the case selected for use.

standard deviations, skew coefficients, lag-one autocorrelation coefficients between monthly flows (e.g. between June and July) $r(1)$ and the correlation coefficients between monthly and annual flow volumes, s_{yx} . These statistics are presented for each of the 12 calendar months.

For all four study streams the disaggregated means and standard deviations are very close approximations to the historical values in all months. Disaggregated values of the skew coefficient do not closely approximate the historical values for all months because the same value of λ used for all 12 calendar months was not the best value for every month, although it did minimize the average monthly goodness-of-fit statistics, over the year T (see Table 5.1). In general months with lower skew coefficient were modeled better than months with higher skew coefficients. Monthly values of s_{yx} are quite well preserved for all streams as would be expected since this parameter is

explicitly incorporated into the parameter estimation for both the MR and VS models.

The lag-one serial correlation between months, $r(1)$, is not explicitly incorporated into the parameter estimation for the MR model and consequently the disaggregated values closely resemble the historic values. However, as would be expected due to its lack of capability for preserving over-the-year serial correlations disaggregated flows from the VS model do not resemble the historic value of $r(1)$ at the beginning of the water year. This characteristic can be seen most clearly in Figure 5.2 for the Beaver in which the disaggregated values of $r(1)$ are shown for both the VS model (equivalent to MR model with $m=0$) and the MR model with $m=1, 2$, and 3. Results from these four models of the Beaver were analyzed to investigate the influence of different values of m on preserving S_{zy} , the matrix of serial correlations between months in successive water years. The absolute values of the differ-

Table 5.2. Parameter matrices for MR (m=3) disaggregation model of Beaver.

	$C = \begin{bmatrix} 0.03 \\ 0.02 \\ 0.03 \\ 0.04 \\ 0.03 \\ 0.04 \\ 0.15 \\ 0.83 \\ 1.01 \\ 0.64 \\ 0.40 \\ 0.25 \end{bmatrix}$	$D = \begin{bmatrix} -0.00 & -0.02 & 0.76 \\ -0.11 & 0.06 & 0.75 \\ -0.11 & 0.02 & 0.66 \\ 0.10 & -0.11 & 0.39 \\ 0.03 & -0.07 & 0.42 \\ 0.01 & -0.08 & 0.47 \\ -0.25 & 0.24 & 0.62 \\ -0.43 & 0.51 & 0.38 \\ 0.31 & -0.10 & -1.35 \\ 0.14 & -0.03 & -0.80 \\ 0.16 & -0.32 & -0.13 \\ 0.04 & -0.19 & 0.15 \end{bmatrix}$
$E = \begin{bmatrix} -0.06 & -0.16 & -0.17 & -0.34 & -0.59 & -0.49 & 0.10 & -0.55 & 0.67 & 0.54 & 0.23 & 0.38 \\ 0.13 & 0.36 & 0.13 & -0.22 & -0.20 & -0.16 & 0.09 & -0.31 & 0.73 & 0.35 & 0.26 & 0.25 \\ 0.12 & -0.20 & -0.31 & 0.13 & 0.18 & 0.31 & 0.49 & -0.43 & 0.69 & 0.08 & 0.22 & 0.19 \\ -0.13 & 0.16 & 0.02 & -0.11 & 0.23 & 0.49 & 0.67 & -0.43 & 0.68 & 0.43 & 0.22 & 0.20 \\ 0.03 & -0.26 & 0.47 & -0.21 & 0.04 & 0.30 & -0.19 & -0.17 & 0.74 & 0.38 & 0.44 & 0.39 \\ -0.03 & 0.07 & -0.23 & -0.01 & 0.13 & 0.27 & -0.96 & 0.31 & 1.21 & 0.45 & 0.71 & 0.93 \\ 0.01 & 0.01 & -0.03 & -0.09 & -0.16 & 0.24 & 0.03 & -0.07 & -0.98 & -0.02 & -3.16 & 4.52 \\ 0.01 & 0.01 & -0.06 & -0.28 & -0.26 & 0.47 & -0.08 & 0.09 & -0.35 & -0.41 & -3.25 & 3.37 \\ 0.02 & 0.00 & -0.08 & -0.22 & -0.30 & 0.50 & -0.06 & 0.28 & -0.80 & 1.49 & 0.46 & -4.52 \\ -0.01 & 0.03 & -0.01 & -0.06 & -0.14 & 0.27 & -0.33 & -1.12 & -0.13 & -2.20 & 0.55 & -2.54 \\ 0.01 & -0.03 & -0.07 & -0.51 & 1.14 & -0.05 & 0.28 & 1.20 & 0.37 & -1.53 & 0.84 & -1.03 \\ -0.01 & 0.01 & 0.08 & 0.49 & -0.66 & 0.27 & 0.23 & 0.70 & 0.74 & -0.66 & 0.37 & -0.13 \end{bmatrix}$		

Table 5.3. Parameter matrices for VS disaggregation model of Blacksmith Fork.

	$A = \begin{bmatrix} 0.19 \\ 0.17 \\ 0.17 \\ 0.18 \\ 0.17 \\ 0.19 \\ 0.64 \\ 0.91 \\ 0.60 \\ 0.46 \\ 0.42 \\ 0.37 \end{bmatrix}$	
$B = \begin{bmatrix} 0.17 & -0.12 & 0.02 & -0.18 & -0.20 & 0.03 & 0.28 & -0.81 & -0.17 & -0.88 & -0.44 & -4.15 \\ -0.21 & 0.16 & -0.06 & 0.04 & -0.22 & -0.05 & 0.25 & -0.69 & -0.21 & -0.70 & -0.25 & -3.49 \\ 0.07 & 0.01 & -0.08 & 0.50 & -0.20 & -0.33 & -0.37 & 0.36 & -0.29 & -0.47 & 0.09 & -3.27 \\ -0.04 & 0.02 & -0.00 & -0.41 & -0.20 & -0.40 & -0.60 & 0.70 & -0.26 & -0.16 & 0.17 & -2.85 \\ -0.01 & -0.00 & 0.04 & 0.00 & -0.28 & 0.49 & 0.68 & 1.14 & -0.39 & -0.04 & 0.40 & -2.23 \\ 0.02 & -0.00 & -0.01 & 0.01 & -0.19 & -0.09 & 0.08 & -0.02 & 2.27 & 1.49 & 2.21 & -2.11 \\ 0.01 & 0.01 & -0.04 & -0.02 & -0.39 & -0.17 & 0.10 & -0.19 & -0.65 & -0.43 & 6.07 & 3.25 \\ 0.01 & 0.01 & -0.04 & -0.02 & -0.44 & -0.23 & 0.17 & 0.24 & 1.18 & -1.55 & -3.12 & 4.02 \\ 0.03 & 0.06 & -0.11 & -0.01 & -0.58 & 0.19 & -0.19 & -0.33 & -0.81 & 1.87 & -2.55 & 1.27 \\ -0.10 & -0.23 & 0.31 & 0.09 & -0.18 & -0.44 & 0.22 & -0.07 & -0.42 & 0.73 & -0.83 & 0.34 \\ 0.10 & 0.25 & 0.18 & -0.04 & 0.14 & -0.54 & 0.50 & 0.08 & -0.34 & 0.56 & -0.61 & 0.09 \\ -0.02 & -0.11 & -0.44 & -0.05 & 0.20 & -0.44 & 0.51 & 0.16 & -0.32 & 0.55 & -0.45 & -0.11 \end{bmatrix}$		

Table 5.4. Parameter matrices for MR (m=2) disaggregation model of Logan.

$C = \begin{bmatrix} -0.00 \\ 0.01 \\ 0.03 \\ 0.06 \\ 0.07 \\ 0.14 \\ 0.39 \\ 0.72 \\ 1.00 \\ 0.72 \\ 0.45 \\ 0.37 \end{bmatrix}$	$D = \begin{bmatrix} -0.20 & 1.14 \\ 0.16 & 0.53 \\ 0.26 & 0.28 \\ 0.07 & 0.39 \\ -0.08 & 0.44 \\ -0.09 & 0.41 \\ 0.60 & -0.88 \\ 0.35 & -0.69 \\ -0.11 & -0.43 \\ -0.31 & 0.03 \\ -0.06 & 0.01 \\ -0.23 & 0.25 \end{bmatrix}$
$E = \begin{bmatrix} -0.08 & -0.32 & 0.17 & 0.08 & -0.05 & 0.10 & -0.40 & -0.20 & 0.16 & -0.01 & 0.08 & 0.21 \\ -0.09 & 0.05 & -0.24 & 0.48 & -0.09 & 0.03 & -0.25 & -0.26 & 0.24 & -0.15 & 0.05 & 0.31 \\ -0.14 & -0.02 & -0.19 & -0.30 & -0.01 & 0.42 & 0.15 & -0.82 & 0.22 & -0.51 & -0.01 & 0.43 \\ 0.16 & 0.04 & 0.17 & 0.18 & 0.01 & 0.41 & 0.21 & -0.96 & 0.38 & -0.52 & -0.20 & 0.34 \\ 0.00 & 0.05 & 0.05 & -0.11 & -0.35 & -0.62 & -0.27 & -0.93 & 0.36 & -0.45 & -0.20 & 0.52 \\ 0.01 & 0.01 & 0.01 & -0.04 & -0.05 & 0.07 & -0.19 & 0.80 & 1.27 & -1.98 & -1.14 & 1.47 \\ 0.02 & 0.02 & -0.00 & -0.02 & -0.19 & 0.15 & -0.19 & 0.08 & -0.84 & 0.53 & -2.99 & 6.95 \\ 0.03 & 0.03 & 0.00 & -0.04 & -0.32 & 0.20 & -0.22 & 0.27 & -0.01 & 0.12 & 4.31 & 3.60 \\ 0.02 & 0.01 & -0.01 & -0.01 & -0.29 & 0.17 & -0.28 & 0.12 & -1.69 & -1.09 & -0.38 & -4.83 \\ 0.03 & -0.02 & -0.04 & -0.03 & -0.50 & 0.19 & 0.13 & 0.28 & 1.07 & 1.24 & -1.30 & -3.65 \\ -0.12 & 0.20 & 0.21 & -0.04 & 0.13 & 0.21 & -0.66 & -0.04 & 0.34 & 0.58 & -0.27 & -1.53 \\ 0.17 & -0.05 & -0.21 & -0.11 & 0.23 & 0.04 & -0.75 & -0.17 & 0.27 & 0.39 & -0.09 & -0.92 \end{bmatrix}$	

Table 5.5. Parameter matrices for MR (m=2) disaggregation model of Weber.

$C = \begin{bmatrix} 0.04 \\ 0.06 \\ 0.07 \\ 0.06 \\ 0.06 \\ 0.06 \\ 0.10 \\ 0.36 \\ 1.16 \\ 0.85 \\ 0.37 \\ 0.24 \end{bmatrix}$	$D = \begin{bmatrix} -0.12 & 0.80 \\ 0.15 & 0.32 \\ 0.05 & 0.31 \\ 0.05 & 0.24 \\ -0.05 & 0.29 \\ -0.21 & 0.66 \\ -0.46 & 1.17 \\ -0.59 & 1.10 \\ 0.15 & -1.04 \\ 0.55 & -1.04 \\ 0.28 & -0.39 \\ 0.32 & -0.45 \end{bmatrix}$
$E = \begin{bmatrix} -0.23 & 0.13 & -0.34 & -0.28 & 0.02 & -0.00 & 1.95 & 0.62 & 2.25 & -0.49 & 0.40 & 0.11 \\ 0.50 & -0.20 & -0.15 & 0.63 & -0.41 & 0.06 & 0.41 & 0.46 & 1.70 & -0.76 & 0.10 & 0.66 \\ -0.27 & 0.03 & 0.52 & 0.93 & 0.30 & 0.76 & -0.48 & 0.46 & 1.58 & -0.41 & -0.23 & 0.53 \\ 0.20 & 0.61 & -0.14 & -0.47 & 0.07 & 0.84 & -0.88 & 0.17 & 1.06 & -0.66 & -0.21 & 0.59 \\ -0.07 & -0.58 & -0.21 & -0.71 & -0.01 & 0.81 & -0.83 & 0.12 & 1.30 & -0.54 & -0.16 & 0.45 \\ -0.18 & 0.08 & -0.49 & 0.23 & -0.39 & -1.11 & -1.32 & -0.40 & 1.64 & -0.85 & -1.16 & 1.46 \\ -0.03 & -0.00 & -0.09 & 0.05 & -0.24 & 0.33 & 0.38 & 0.46 & -0.94 & 0.07 & -6.66 & 5.81 \\ -0.11 & 0.04 & -0.23 & 0.19 & -0.66 & 0.49 & 0.03 & -0.13 & -0.96 & 0.73 & 4.20 & 8.25 \\ -0.12 & 0.03 & -0.22 & 0.20 & -0.72 & 0.50 & 0.07 & 0.21 & -1.24 & -4.63 & -0.23 & -6.62 \\ -0.06 & 0.04 & -0.14 & 0.13 & -0.63 & 0.44 & 0.02 & -0.75 & 0.41 & 5.88 & -1.01 & -5.35 \\ -0.01 & -0.03 & -0.81 & 0.40 & 0.66 & 0.16 & -0.32 & 1.87 & -1.05 & 1.54 & 0.49 & -1.54 \\ -0.02 & 0.03 & 0.48 & -0.34 & -0.58 & -0.43 & -0.36 & 2.84 & 0.89 & 1.14 & 0.48 & -0.54 \end{bmatrix}$	

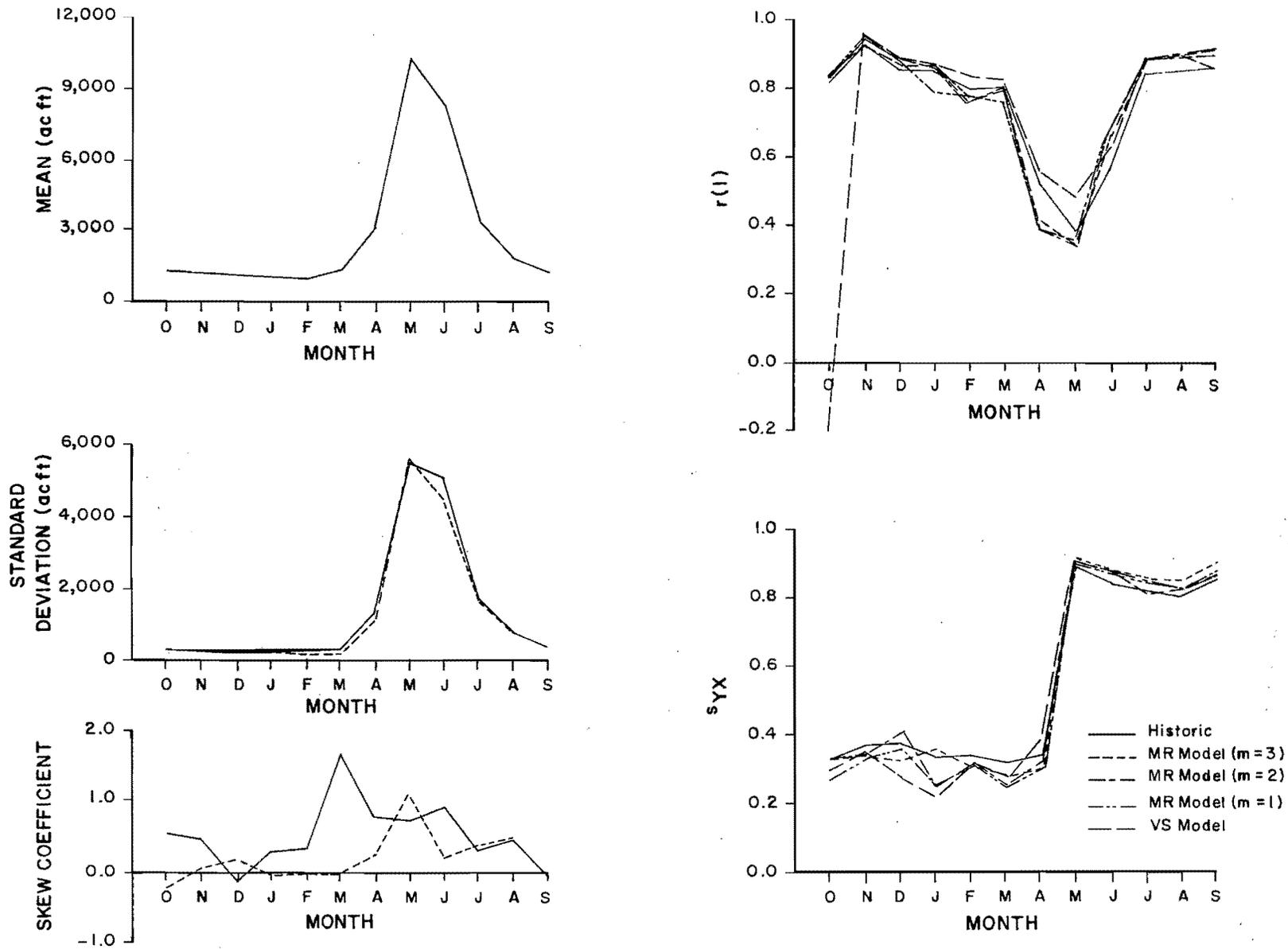


Figure 5.2. Comparison of historic and disaggregated monthly statistics for Beaver.

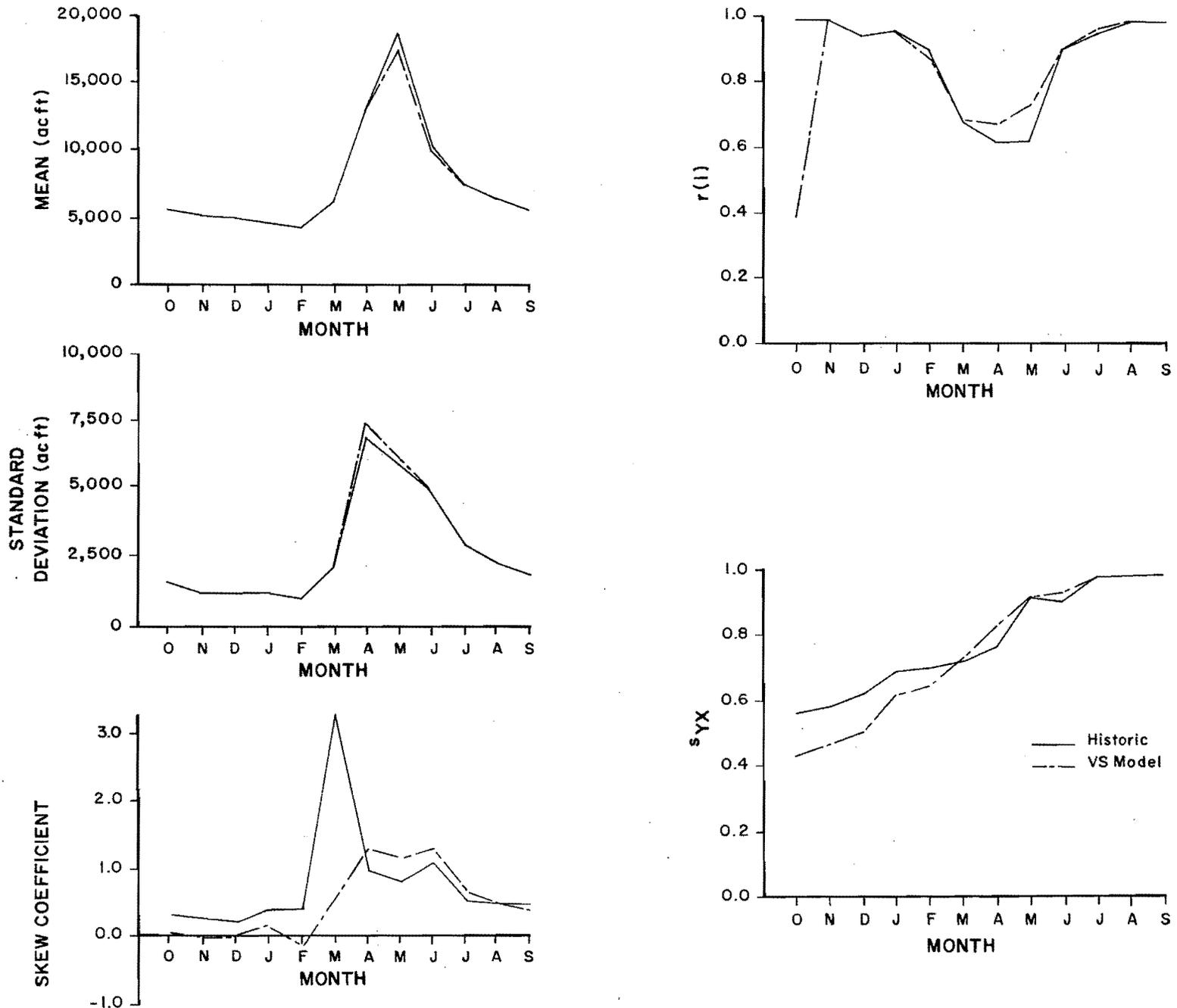


Figure 5.3. Comparison of historic and disaggregated monthly statistics for Blacksmith Fork.

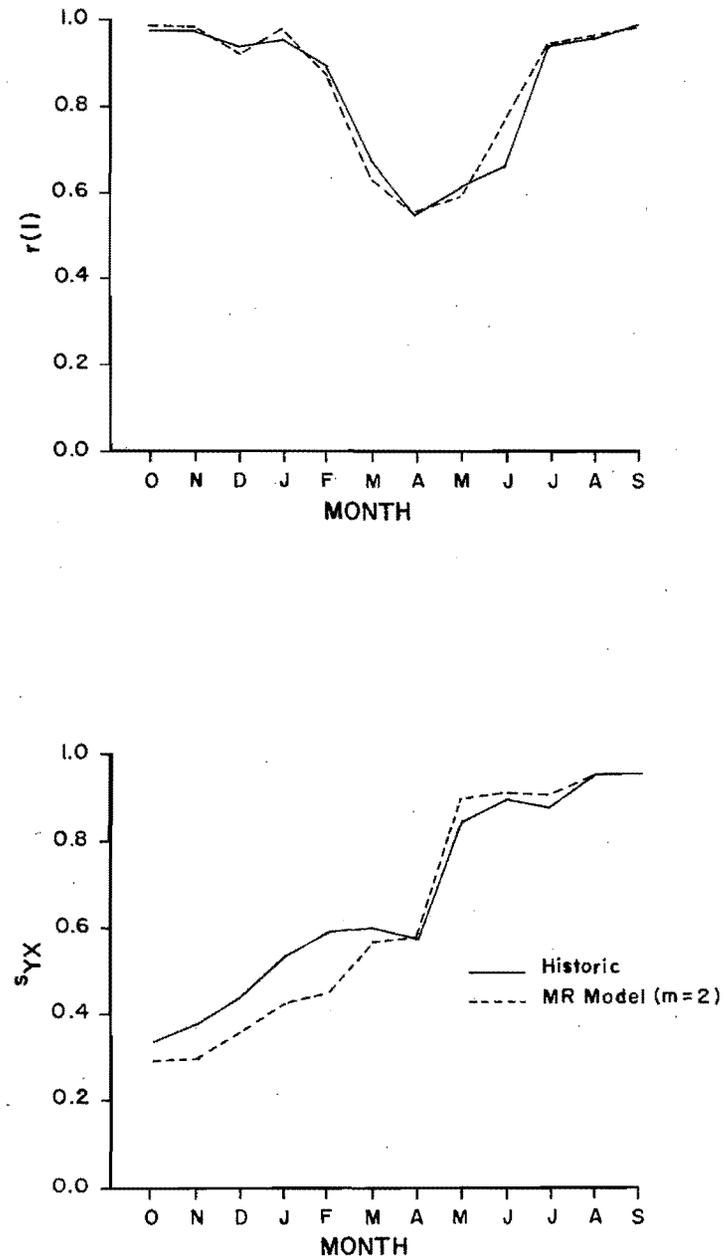
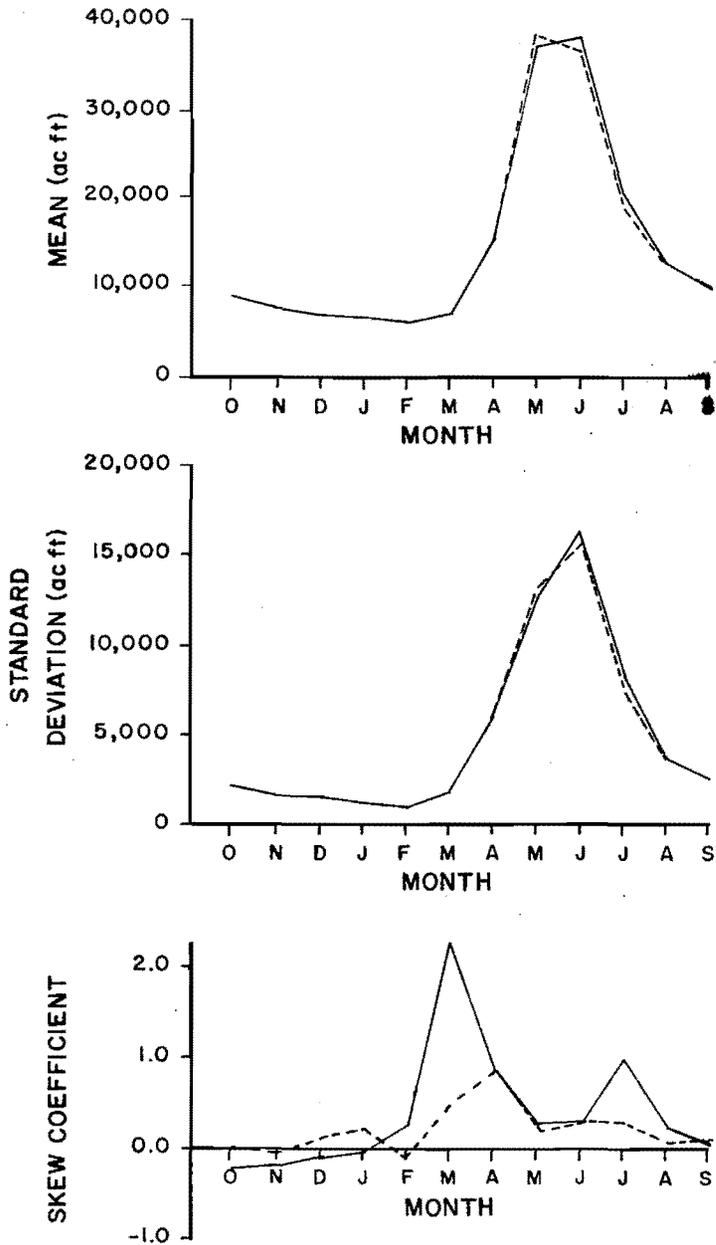


Figure 5.4. Comparison of historic and disaggregated monthly statistics for Logan.

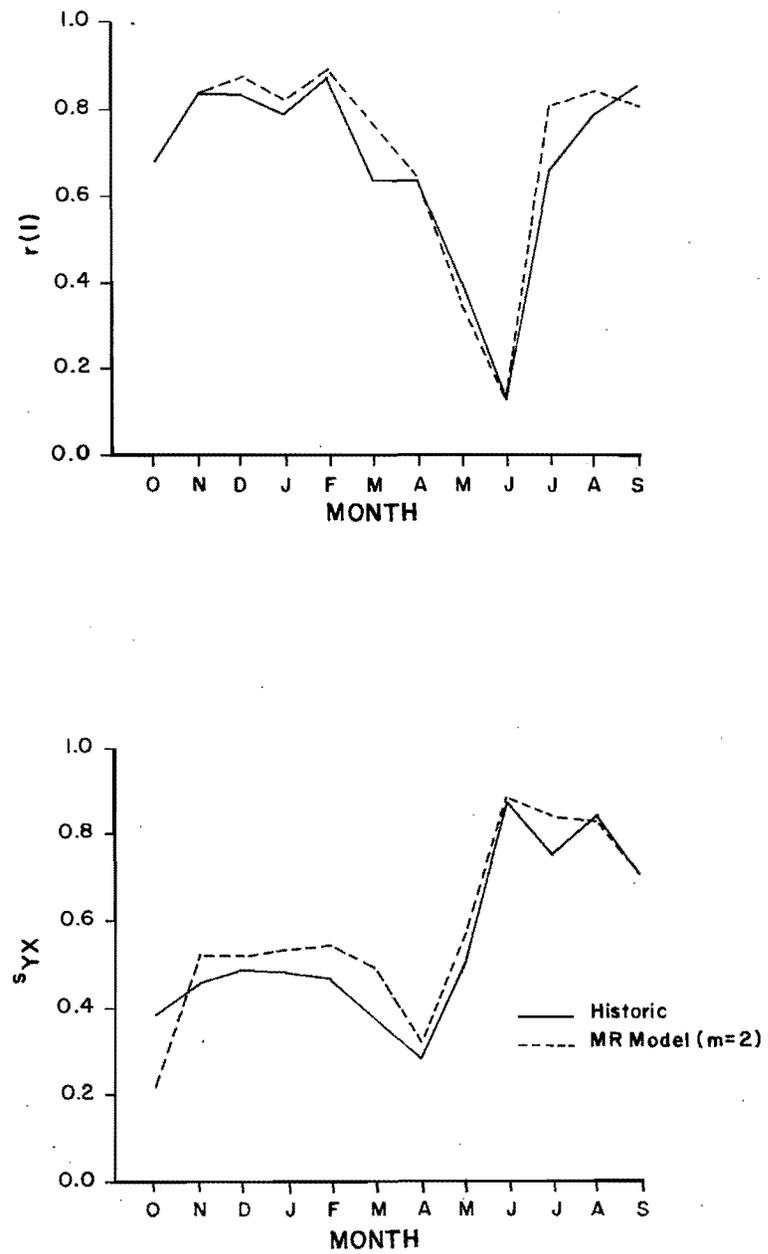
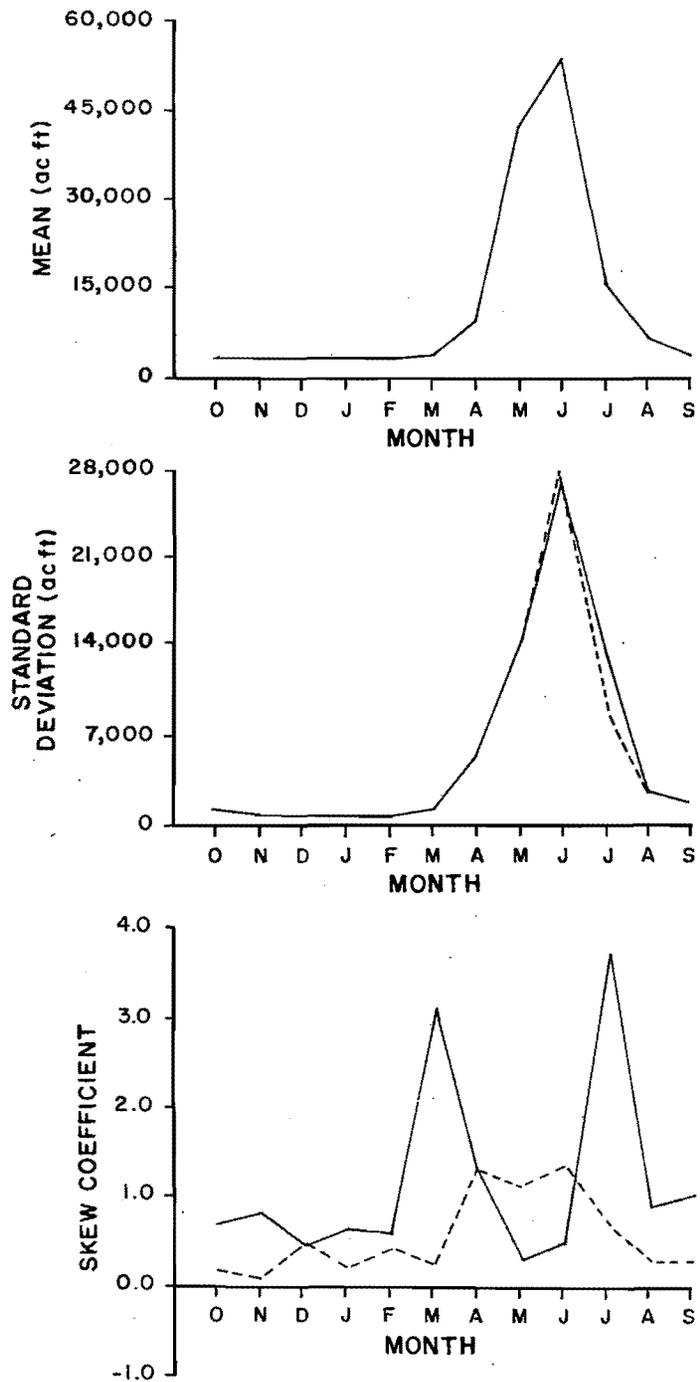


Figure 5.5. Comparison of historic and disaggregated monthly statistics for Weber.

ences between the disaggregated and historic values of each element of S_{ZY} were calculated and averaged for all pairs of months in successive water years which are separated by a J month lag as follows:

$$e_J = \frac{1}{J} \sum_{j=1}^J |S_{YZ}^{DISAG}(12 - J + j, j) - S_{YZ}^{HIST}(12 - J + j, j)| \dots (5.8)$$

in which

e_J = average absolute error between the disaggregated and historical values of elements of S_{ZY} which are separated by a J month lag

J = lag between pairs of months in successive water years (e.g. for J=3: July and October, August and November, and September and December, where the first month is in year t-1 and the second in year t)

j = index

S_{YZ} = element of S_{YZ} for disaggregation (DISAG) or historical (HIST) cases

Figure 5.6 contains plots of e_J vs J for the four disaggregation models applied to the Beaver and for lags up to J = 12. This figure shows that, as would be expected, the errors in preserving the across-the-year serial correlations are much smaller for the MR model than for the VS model. Values of e_J increase with increasing J for the MR model indicating that serial correlations with lower lags are better preserved. A decrease in the values of e_J with increasing J for the VS model can probably be attributed to the indirect preservation of these serial correlations via the autocorrelation structure at the annual level and the cross-correlation between annual and monthly flows S_{YX} , which are preserved by the VS model. The very high values of e_J for the VS model at low lags indicate that there is no preservation of S_{ZY} close to the boundary between water years. As m increases the magnitude of e_J tends to decrease demonstrating that a higher order Z matrix does improve the preservation of S_{ZY} , as would be expected. However, the magnitude of changes in e_J is small. This can be clearly seen by summing the 12 values of e_J for each

model: 7.97, 1.19, 1.09, 0.99 for m = 0 (VS model), 1, 2, and 3 respectively. Figure 5.6 shows no detectable improvement in the preservation of elements of S_{ZY} when $J \leq m$ as might be expected.

When the comparison of historic and disaggregated statistics is restricted to the growing season, which is the last five months of the year (May through September), the resemblance is a little better than for the entire year. As will be explained in the following chapter the only monthly flows used in this study were those in the growing season.

Generation Procedure

The procedure for generating a monthly streamflow sequence from an annual sequence using a seasonal disaggregation model for which the parameter matrices have been estimated is summarized below:

- 1) Input the transformation used in the parameter estimation procedure, λ ; the generated annual streamflow sequence; X_t , m monthly streamflows from the year before the starting year, Z_t , (for MR model only); the number of years to be disaggregated, N; and parameter matrices (A and B for VS model or C, D and E for MR model).
- 2) Transform X and Z_t using the Box-Cox transformation with the parameter λ input.
- 3) Subtract the corresponding transformed mean from the annual and monthly transformed inputs.
- 4) Generate 12 standard normal random numbers, V_t , for each of the N years.
- 5) Using either Equation 5.1 for the VS model or Equation 5.2 for the MR model disaggregate the annual streamflow, X_t , to the monthly streamflow Y_t .
- 6) For the MR model save the last m disaggregated monthly streamflows in Z_t , for the next year's disaggregation.
- 7) Rescale the standardized generated monthly streamflows by adding the transformed monthly means.
- 8) Inverse transform the rescaled generated monthly streamflows, Y_t .

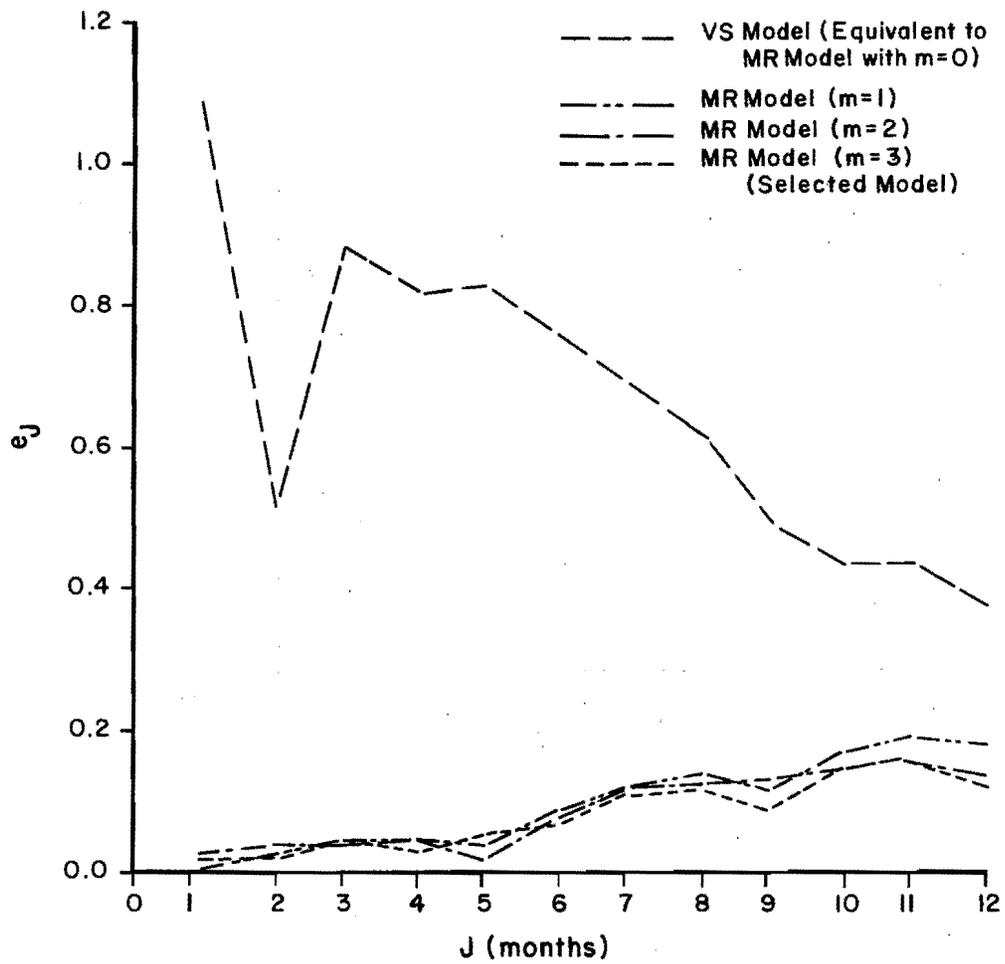


Figure 5.6. Average absolute errors in elements of \underline{S}_{YZ} which are separated by a J month lag.

CHAPTER 6
 MODELING THE AGRICULTURAL ECONOMIC
 LOSSES FROM DROUGHT

Introduction

Different patterns of water supply shortfall cause different amounts of economic loss. In agriculture the most critical period of water supply for grain corn is the pollination growth stage since soil moisture deficit at this stage will result in the most significant impact on crop yield, and therefore agricultural benefits. Most of the previous evaluations of stochastic streamflow models have emphasized the preservation of statistics of the streamflow time series and comparison of reservoir capacities based on the use of different models. In this study model performance was also compared based on the economic regret, measured in terms of the losses in the value of agricultural production, associated with choosing a particular model. A diversion rule was applied to the generated monthly streamflows to calculate monthly diversions available for irrigation. The quantity of water available for diversion and the irrigation water requirement were used as inputs to a model for estimating crop yield and hence the decrease in the value of agricultural production during water short years was estimated. The remainder of this chapter is divided into four sections describing the crop yield model, the calculation of irrigation diversions, the calculation of irrigation water requirements, and the procedure for calculating economic regret.

Crop Yield Model

The work on production functions reported by Stewart et al. (1977) and specifically the corn production function model developed by Hanks (1974) and statistically analyzed by Gowon, Anderson, and Biswas (1978), was selected as a basis for the agricultural loss function used in this study. The model was chosen because of its simplicity. Many such models require daily input of several meteorologic variables in addition to soil moisture or water supply and this level of detail was not practical in this study. The model relates the yield of grain corn to the ratio of actual to potential evapotranspiration in three growth stages, as follows:

$$\frac{Y}{Y_P} = C \left(\frac{ETA}{ETP} \right)_v^{\lambda_1} \left(\frac{ETA}{ETP} \right)_p^{\lambda_2} \left(\frac{ETA}{ETP} \right)_m^{\lambda_3} \dots \quad (6.1)$$

in which

- Y = yield of harvested grain corn (bushels/acre)
- Y_P = potential yield based on the highest measured yield for corn at the study location (bushels/acre)
- ETA = actual evapotranspiration
- ETP = potential evapotranspiration
- C = constant coefficient
- λ_i = exponential coefficients for three growth stages
- v = subscript denoting vegetative growth stage of corn
- p = subscript denoting pollination growth stage of corn
- m = subscript denoting maturation growth stage of corn

Gowon, Anderson, and Biswas (1978) applied a logarithmic transformation to Equation 6.1 and estimated the coefficients C and λ_i by linear regression based on data collected at Logan, Utah, during a two year period, 1974-1975. The resulting equation is:

$$Y = 0.97 R_v^{0.347} R_p^{0.574} R_m^{0.330} Y_P \dots \quad (6.2)$$

in which

R = ratio of actual to potential evapotranspiration

ETA is calculated based on an irrigation diversion rule applied to the monthly generated streamflows. Values of ETP for each growth stage are calculated using the modified Blaney-Criddle method for estimating consumptive use. Calculations for both ETP and ETA are reported in the following sections of the chapter.

The three growth stages were defined as follows:

1. Vegetative - time from planting to appearance of first tassel.
2. Pollination - time from appearance of first tassel to blister kernal.
3. Maturation - time from blister kernal to physiological maturity.

Average durations of these three periods at Logan were found to be 63, 26, and 43.5 days respectively. The pollination growth stage for grain corn is the most critical in terms of water deficit as can be seen from the high value of λ_2 in Equation 6.2. At Logan, Utah, the frost free growing season generally lasts from May to September. Irrigation diversions were generated on a monthly basis using the disaggregation models described in Chapter 5. Therefore, the three growth stages were approximated as follows:

1. vegetative - May and June
2. pollination - July
3. maturation - August and September

The potential yield, YP, for grain corn was taken to be 86.63 bushels/acre based on data contained in the Utah Agricultural Statistics (1978) for the years 1970-1977.

Grain corn was chosen because it is a highly drought sensitive crop. In contrast silage corn is less sensitive because its most critical growth stage is the vegetative stage when water supply is more plentiful and Stewart et al. (1977) found that stress at this early stage can actually condition the crop to water shortages later and result in increases in yield. No attempt was made in this study to represent these conditioning effects and thus the effect on crop yield of the evapotranspiration ratio in each season was considered independently of the other seasons.

Irrigation Diversion Rule

To calculate the quantity of irrigation water diverted, an irrigation diversion rule was applied to the generated monthly stream-flow volumes. The rule was based on the Kimball Decree, established in 1923, which defines water rights on the Logan River including the diversion of water to the Logan, Hyde Park, and Smithfield Canal. The decree defines the diversion in cfs as a function of the Logan River streamflow in cfs. For study purposes this diversion rule was converted to acre-feet/month and is represented graphically in Figure 6.1.

Irrigation Diversion Requirements

Monthly irrigation requirements for corn were estimated using the Blaney-Criddle method as modified by the Soil Conservation Service (1970). Consumptive use is calculated as follows:

$$U = k_t k_c t_p \dots \dots \dots (6.3)$$

in which

- U = consumptive use (in.)
- k_t = temperature coefficient
= $0.0173t - 0.314$
- k_c = monthly crop coefficient for corn
- t = average daily temperature for month ($^{\circ}$ F)
- p = fraction of total daylight hours in year which occur in month

Table 6.1 contains the calculation of consumptive use and monthly irrigation diversion requirements expressed in feet of water at the point of diversion. The following assumptions were made in applying the Blaney-Criddle method:

1. All inputs are obtained for Logan, Utah (latitude: $45^{\circ}45'$ N), and no attempt was made to adjust inputs to the location of the other study streams.
2. Water availability is assumed to be not limiting for the calculation of U.
3. Values of k_c are taken from Soil Conservation Service (1970).

The monthly irrigation requirement was calculated by subtracting the average monthly precipitation from the consumptive use. Based on these calculations no irrigation is needed in May because there is adequate precipitation. An irrigation conveyance efficiency of 50 percent was assumed in order to calculate the irrigation diversion requirement.

Table 6.2 contains the calculation to convert the irrigation diversion requirement in feet to acre-feet based on an irrigated area which varies from stream to stream. In order to provide an adequate test of the ability of the model to preserve drought properties under extreme conditions it was considered necessary to base the irrigation diversion requirements on larger irrigated areas than actually exist currently. The diversion at the mean monthly flow minus one standard deviation was calculated for each month of the irrigation season (June-September) and for each stream using Figure 6.1. The irrigated area was then fixed for each stream such that the monthly irrigation diversion requirements exceeded the diversion at the mean monthly flow minus one standard deviation in July through September (see Table 6.2). The irrigated areas associated with the study streams are 500, 1500, 5000, and 2000 acres for the Beaver, Blacksmith Fork, Logan, and Weber respectively.

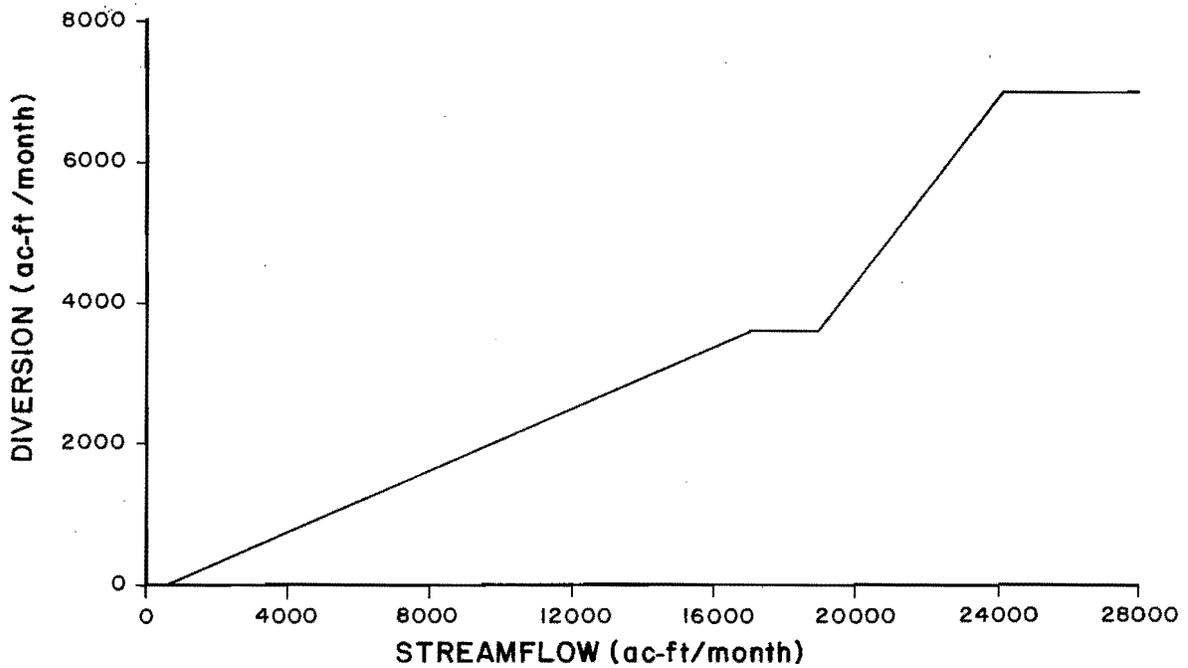


Figure 6.1. Irrigation diversion rule.

Table 6.1. Calculation of monthly, unit irrigation diversion requirements for corn at Logan, Utah (Latitude: 41°45' N).

Variable	Symbol	Units	Month					Irrigation Season Total
			May	June	July	August	September	
Crop coefficient	K_c		0.53	0.82	1.05	1.00	0.88	-
Average temperature	t	°F	56.3	63.1	72.9	71.4	62.0	-
Daylight hours	P	%	10.10	10.20	10.32	9.60	8.40	-
Consumptive use	U	in	1.81	4.10	7.48	6.31	3.48	23.18
Average precipitation	R	in	1.86	1.78	0.34	0.87	0.94	5.79
Unit irrigation requirement	U-R	in	0	2.32	7.14	5.44	2.54	17.44
Unit irrigation diversion requirement	ETR ^a	ft	0	0.39	1.19	0.91	0.42	2.91

^a Monthly unit irrigation diversion requirement is given in ft and is based on an irrigation efficiency of 50 percent. Thus $ETR = (U-R)/(12 \times 0.50)$ in which 12 converts (U-R) in inches to feet.

Table 6.2. Comparison of monthly irrigation requirements and range of irrigation diversion based on historic streamflows in ac. ft.

Stream	Month Phenological stage	June	July	August	September	Irrigation Season Total
		Vegetative	Pollination	Maturation		
Beaver	Diversion at mean monthly flow ¹	1,746	679	347	213	2,985
	Diversion at mean monthly flow minus one standard deviation ¹	622	283	166	120	1,191
	Monthly irrigation diversion requirements based on 500 ac ²	195	595	455	210	1,455
Blacksmith Fork	Diversion at mean monthly flow ¹	2,128	1,542	1,335	1,163	6,168
	Diversion at mean monthly flow minus one standard deviation ¹	1,056	936	835	753	3,580
	Monthly irrigation diversion requirements based on 1500 ac	585	1,785	1,365	630	4,365
Logan	Diversion at mean monthly flow ¹	7,010	5,399	2,784	2,154	17,347
	Diversion at mean monthly flow minus one standard deviation ¹	7,017	2,555	1,956	1,579	12,107
	Monthly irrigation diversion requirements based on 5000 ac	1,950	5,950	4,550	2,100	14,550
Weber	Diversion at mean monthly flow ¹	7,010	3,506	1,392	943	12,851
	Diversion at mean monthly flow minus one standard deviation ¹	7,010	641	855	621	9,127
	Monthly irrigation diversion requirements based on 2000 ac	780	2,380	1,820	840	5,820
Column No.		(1)	(2)	(3)	(4)	(5)

¹ Diversions are calculated using the rule in Figure 6.1 and the monthly streamflow statistics presented in Table 3. . For example: Beaver in July, mean flow minus one standard deviation = 3560 - 1791 = 1769 ac ft, diversion = $-107.836 + 0.221 \times 1769 = 283$ ac ft.

² Monthly irrigation requirements are obtained in Table 6.1 by the irrigation area. For example: Beaver in June, monthly irrigation requirement = $0.39 \times 500 = 195$ ac ft.

The monthly irrigation diversion requirements are a measure of the agricultural demand. Thus if the actual diversions based on the rule in Figure 6.1 fall below these requirements ETA will fall below ETP and some reduction in yield can be expected based on Equation 6.2. The monthly irrigation diversion requirement can be considered to define a crossing level below which drought conditions are defined for irrigated agriculture.

Calculation of Economic Regret

Generated monthly streamflow volumes are used to calculate irrigation diversions (see Figure 6.1) and hence the ratio $R = \text{ETA}/\text{ETP}$ which makes possible an estimate of annual crop yield using Equation 6.2. The average annual economic benefit or value of the crop is estimated as follows:

$$B_j = \frac{1}{N} \sum_{t=1}^N Y_t P A \quad (6.4)$$

in which

B_j = average annual economic benefit based on streamflows generated by model j

N = length of generated streamflow sequence

t = year index

Y_t = crop yield for t^{th} year

P = price of grain corn

A = irrigated area

Thus, different values of B_j were calculated using sequences generated by each of the five annual streamflow models and in addition using the historic streamflow sequence. The price of grain corn was \$2.45 per bushel which is the 1977 price reported by Utah Agricultural Statistics (1978).

Economic regret is the difference in the value of B_j based on streamflows generated with model j if model i is the true model. Thus economic regret is defined as follows:

$$\alpha_{ij} = B_i - B_j \quad (6.5)$$

in which α_{ij} = economic regret in dollars for model i given that model j is true. Since in practice we do not know which model is true, or perhaps more accurately, the best representation of the annual streamflows, it is useful to calculate a total economic regret by assuming that each alternative model and the historic sequence is true, as follows:

$$R_i = \sum_{j=1}^m \alpha_{ij} = \sum_{j=1}^m (B_i - B_j) \quad (6.6)$$

in which

R_i = total economic regret for model i in dollars

m = number of alternative models including the historic sequence

The α_{ij} are summed algebraically and thus R_i can be positive or negative. It follows that the most desirable model, based on this criterion, is the one that minimizes the value of R_i .

CHAPTER 7

MODEL CHOICE FOR STUDY STREAMS

Introduction

In following the systematic modeling approach outlined in Figure 1.1 the second step, choice of model type, was omitted so that the performance of each model could be evaluated as a basis for developing a model choice strategy. Therefore, all five annual models (AR2, ARMA, AMAK, FFGN, and BKL) were calibrated to the four study streams (Beaver, Blacksmith Fork, Logan, and Weber). This chapter contains an evaluation of the performance of the annual models and a proposed model choice strategy. For each study stream, the performance of the annual models is evaluated in terms of the following: the preservation of the persistence statistics ($\rho(1)$ and K) for annual streamflows; the preservation of seasonal crossing properties; the cost and ease of model use; a comparison of reservoir capacity and critical drought design parameters; and minimizing economic regret associated with drought-related agricultural losses. A proposed model choice strategy is given for the annual models based on the $\rho(1)$ and K values estimated from the historic streamflow record.

Evaluation of Annual Performance

The calibration of each annual model to the four streams was described in Chapter 4 along with a comparison of the historic and generated values of the persistence statistics $\rho(1)$ and K . In Chapter 4 the presentation was ordered by model type. In this section the ability of the five models to preserve the annual streamflow statistics will be ordered by study stream. A summary comparison of the historic and generated annual statistics is included in Table 7.1. The match between historical and generated series was compared by Type B resemblance, wherein the models were used to generate 50 series of N years length which is equal to that of the historical record. The lag-one serial correlation and Hurst coefficients were calculated for each series. Then the mean and standard deviations of both coefficients over the 50 series were calculated. The means are shown in column 2 and 4, and the standard deviations are shown in columns 3 and 5 of Table 7.1. From column 1 of Table 7.1 it will be observed that the coefficient of variation of the annual streamflows was preserved exactly in every

case as would be expected with the generation procedures described in Chapter 4.

Beaver River

A comparison of the $\rho(1)$ - K persistence statistics preserved by the annual models of the Beaver streamflows is shown in Figure 7.1 along with the boundaries of the feasible regions for the ARMA, AMAK and FFGN models. The unbiased $\rho(1)$ is included for a comparison with the generated $\rho(1)$ statistics which in some cases provided a closer fit to the unbiased estimate. For the Beaver River the unbiased $\rho(1)$ - K point lies just outside the feasible region for the AMAK model and both the biased and unbiased $\rho(1)$ - K points are outside the FFGN feasible region (see Figure 7.1). As would be expected the $\rho(1)$ - K values preserved by each model lie within the feasible region for that model. The ARMA, AMAK, and BKL models grouped fairly closely around the historic estimates. The FFGN model was the only one to overestimate $\rho(1)$ while the ARMA model was the only one to overestimate the Hurst coefficient although not by much. It also underestimated $\rho(1)$ because fit of $\rho(1)$ was sacrificed to improve fit of K (see Figure 7.1). As would be expected the AR2 model underestimated K . Therefore, the best fitting models of the Beaver based on the minimum total deviation of the values of $\rho(1)$ and K preserved from their historic values are first the ARMA model with the AMAK ranked second. The ranking of the models in terms of preserving the persistence statistics using the criterion given in Equation 2.18 is: FFGN and then ARMA, see Table 7.2.

Blacksmith Fork

Figure 7.2 shows that the historic $\rho(1)$ - K value is within the feasible region for all models. All models fit the persistence statistics quite well. The closest model to the historic persistence statistics is the AMAK model with equal weighting and the FFGN model based on the unequal weighting based on Equation 2.18; the second closest model is the BKL for both the equal weighting and unequal weighting cases. The historic Hurst coefficient was well preserved by the AR2 model probably because of the high value of $\rho(1)$. This relationship was described by O'Connell (1974) for the AR1 model and can be expected to be similar for AR2.

Table 7.1. Overall summary of model results.

Stream	Model	Annual Statistics					Demand	Irrigation Season Statistics			Total Economic Regret ¹⁰	Design Parameters			
		\bar{C}_V^1	$\bar{\rho}(1)^2$	$\bar{\sigma}(1)^3$	\bar{K}^4	\bar{K}^5		D^{*6}	$E(ND)^7$	$E(RL)^8$		$E(RS)^9$	Reservoir design capacity ¹¹ S_{98}^*	Reliability of Historic storage estimate S^{*12}	Critical drought ¹³ CD_{98}^*
Beaver	Historical	0.35	0.24	-	0.76	-	0.49	48	2.15	0.14	-	(0.58)	-	(0.36)	-
	AR2	0.35	0.19	0.09	0.66	0.06	0.49	46.68	2.33	0.14	5,100	1.08	45	0.97	7
	ARMA	0.35	0.20	0.15	0.78	0.07	0.49	46.48	2.35	0.14	-4,500	1.24	41	0.98	12
	AMAK	0.35	0.23	0.15	0.73	0.08	0.49	46.64	2.35	0.14	-900	1.17	50	0.96	16
	FFGN	0.35	0.33	0.13	0.75	0.07	0.49	47.06	2.32	0.13	6,900	1.15	48	0.86	6
	BKL	0.35	0.24	0.11	0.72	0.08	0.49	46.22	2.32	0.13	-8,100	1.05	49	0.89	9
Blacksmith Fork	Historical	0.34	0.49	-	0.77	-	0.71	46	2.02	0.14	-	(0.59)	-	(0.38)	-
	AR2	0.34	0.43	0.11	0.74	0.07	0.71	41.86	2.02	0.15	-17,700	2.16	14	1.19	15
	ARMA	0.34	0.49	0.16	0.84	0.06	0.71	41.52	2.03	0.15	-15,900	2.79	13	1.57	20
	AMAK	0.34	0.46	0.12	0.80	0.07	0.71	42.20	2.02	0.15	13,500	2.36	13	1.31	23
	FFGN	0.34	0.44	0.10	0.78	0.06	0.71	41.82	2.01	0.15	-1,500	2.04	19	1.15	17
	BKL	0.34	0.48	0.12	0.76	0.07	0.71	42.30	2.01	0.15	-900	1.93	14	1.08	14
Logan	Historical	0.26	0.32	-	0.72	-	0.84	59	2.17	0.22	-	(0.60)	-	(0.55)	-
	AR2	0.26	0.26	0.11	0.68	0.07	0.84	61.10	2.13	0.21	-35,000	1.29	12	0.85	24
	ARMA	0.26	0.28	0.13	0.74	0.07	0.84	61.28	2.13	0.21	-41,000	1.46	13	0.88	14
	AMAK	0.26	0.30	0.12	0.72	0.06	0.84	61.26	2.17	0.21	31,000	1.60	16	1.02	27
	FFGN	0.26	0.34	0.15	0.74	0.07	0.84	60.84	2.16	0.21	-11,000	1.31	20	0.81	29
	BKL	0.26	0.24	0.13	0.72	0.07	0.84	60.78	2.18	0.21	7,000	1.46	15	0.83	22
Weber	Historical	0.29	0.27	-	0.78	-	0.45	60	2.08	0.08	-	(0.30)	-	(0.30)	-
	AR2	0.29	0.22	0.10	0.69	0.07	0.45	58.48	1.94	0.09	-52,000	0.53	17	0.48	9
	ARMA	0.29	0.30	0.16	0.80	0.07	0.45	57.78	1.96	0.09	-64,000	0.53	14	0.49	10
	AMAK	0.29	0.27	0.12	0.77	0.06	0.45	57.94	1.94	0.09	-70,000	0.51	14	0.49	13
	FFGN	0.29	0.33	0.12	0.74	0.07	0.45	57.50	1.95	0.09	-100,000	0.49	14	0.49	19
	BKL	0.29	0.28	0.13	0.75	0.06	0.45	58.38	1.95	0.09	-46,000	0.49	16	0.49	17
Column No.		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)

¹ Expected value of coefficient of variation of annual streamflow volumes
² Expected value of annual lag-one autocorrelation coefficient
³ Standard deviation of lag-one autocorrelation coefficient
⁴ Expected value of Hurst coefficient
⁵ Standard deviation of Hurst coefficient
⁶ Seasonal demand divided by mean seasonal diversions (see Table 6.2) (James, Bowles, and Kottegoda 1980)
⁷ Expected number of down crossings or droughts in the synthetic sequences with length equal to the historic record at seasonal demand level in column 6
⁸ Expected seasonal negative run lengths with respect to demand level in column 6
⁹ Expected seasonal negative run sum divided by mean seasonal diversion with respect to demand level in column 6
¹⁰ Economic regret calculation in Table 7.3 (James, Bowles, and Kottegoda 1980)
¹¹ Reservoir design capacity divided by average seasonal diversion at 98 percent reliability for each model. Note that for historic case reservoir storage is based on streamflow record and with estimates of its reliability given in column 12
¹² Percent reliability of reservoir estimated from historic record
¹³ Critical drought (maximum negative run sum) based on 98% probability of nonexceedance divided by average seasonal diversion for each model.
¹⁴ Note that for historic case critical drought is based on streamflow record and with estimates of its reliability given in column 13
 Probability of nonexceedance of historic critical drought (maximum negative run sum)

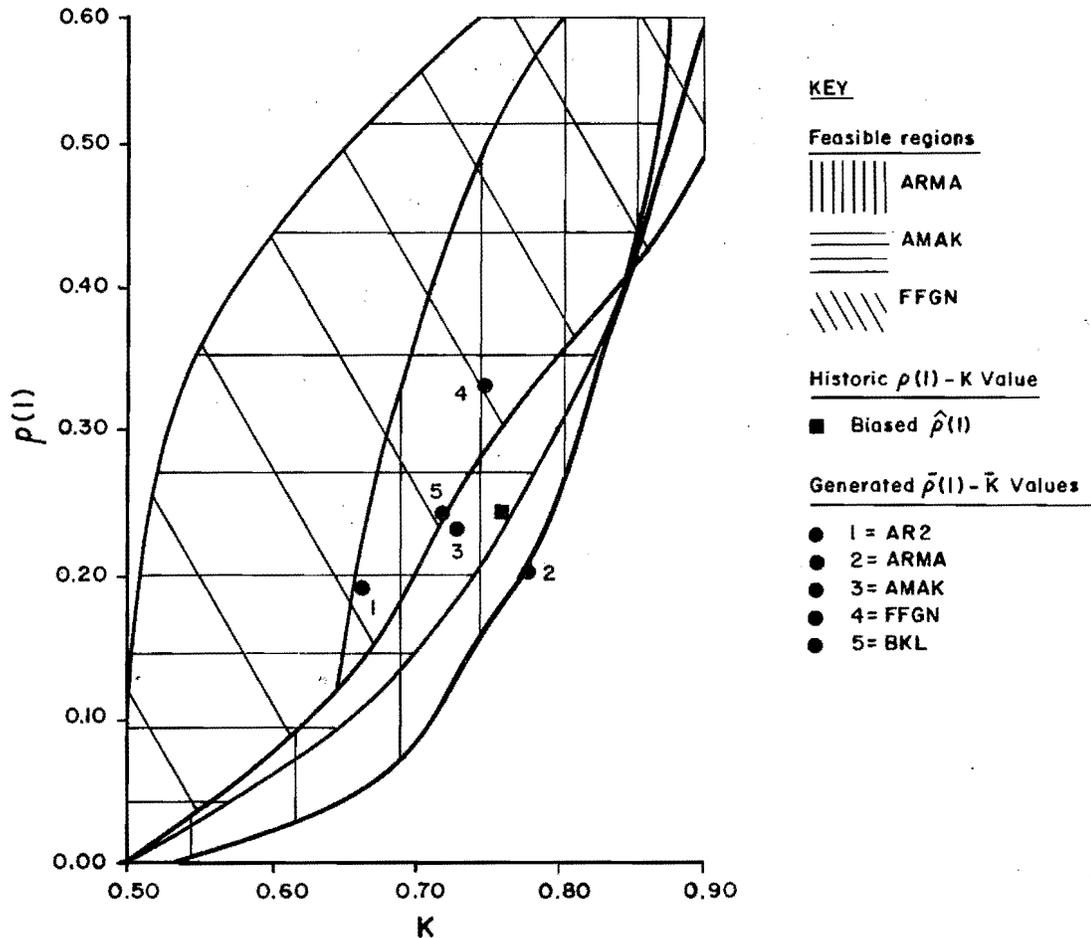


Figure 7.1. Comparison of historic and generated $\rho(1)$ -K values for Beaver River.

Logan River

Referring to Figure 7.3, the historic lag-one autocorrelation and Hurst coefficients are within the feasible region for all the models. The Hurst coefficient for the Logan streamflows is 0.72, which is the average value found by Hurst. At this value both the H and K estimation of h are approximately unbiased (Wallis and Matalas 1970). Results for all the models grouped closely around the historic values. Based on equal weighting the ARMA model is closest, because the overestimate in K offset the underestimate in $\rho(1)$, and for unequal weighting, Equation 2.18, the AMAK model is closest. The second closest models are AMAK for equal weighting and the BKL for unequal weighting, although it should be noted that the BKL preserved the worst lag-one autocorrelation coefficient fit, but this was discounted by the unequal weighting.

Weber River

Referring to Figure 7.4, the Weber historic lag-one autocorrelation and Hurst

coefficients are outside the feasible region for the FFGN model in the biased case and outside the feasible region for the AMAK model in the unbiased case. The closest fitting model in terms of $\rho(1)$ -K preservation is the AMAK regardless of the weighting factor, while the second closest model is the FFGN for equal weighting and the ARMA for unequal weighting.

Summary

Figures 7.5 and 7.6, which provide a graphical comparison of the preservation of the lag-one autocorrelation and Hurst coefficients for each stream, show that the preservation was generally good for the FFGN model or the FGN approximations (i.e. ARMA, AMAK, BKL) but poor for the AR2 model. The models tended to underestimate the biased lag-one autocorrelation (see Figure 7.5), possibly due to the small number of traces (O'Connell 1974). For the Hurst coefficient (see Figure 7.6) all models except ARMA underestimated K for the Beaver and Weber streamflows, because the $\rho(1)$ -K points for these two streams were outside of the feasible region. For the

Table 7.2. Ranking of models by alternative model choice criteria.

Stream	Criterion for ranking		
	Persistence Statistics: $\rho(1) - K$		Total Economic Regret
	Equal weighting Min ($\Delta\rho(1) + \Delta K$)	Unequal weighting Min ($0.15 \Delta\rho(1) + \Delta K$)	Minimum
<u>Beaver</u>			
1st	ARMA	FFGN	AMAK
2nd	AMAK	ARMA	ARMA
<u>Blacksmith Fork</u>			
1st	AMAK	FFGN	BKL
2nd	BKL	BKL	FFGN
<u>Logan</u>			
1st	ARMA	AMAK	BKL
2nd	AMAK	BKL	FFGN
<u>Weber</u>			
1st	AMAK	AMAK	BKL
2nd	FFGN	ARMA	AR2

Blacksmith and Logan the generated values of the Hurst coefficient were distributed above and below the historic values. In all cases the ARMA model gives the highest average estimate for the Hurst coefficient, providing the closest fit for Beaver and Weber. The lower lag-one autocorrelation coefficients for Beaver and Blacksmith were generally preserved better than the higher values. For the Hurst coefficient the best fit was for Logan which has an historic Hurst coefficient of 0.72 which is in the known unbiased range. As would be expected the AR2 model did poorly in preserving the Hurst coefficient. The BKL model consistently underestimated the Hurst coefficient for all the streams except Logan, but the historic value of the Hurst coefficient was always within one standard deviation of the average generated value. With the exception of the BKL model for the Beaver and Weber the AR2 and BKL models underestimated the lag-one autocorrelation coefficient. In most cases the models preserved the lag-one autocorrelation and Hurst coefficients within one standard deviation (see Table 7.1 for values of $\rho(1)$ and K). The exceptions were: the AR2 Hurst coefficient for the Beaver, the ARMA Hurst coefficient for the Blacksmith Fork, and the AR2 Hurst coefficient for the Weber. However, for all of these exceptions the generated values were within two standard deviations of the historic values.

The foregoing discussion has related to model performance with respect to preserving either $\rho(1)$ or K separately. In Table 7.2 the ranking of models by the persistence statistics criterion provides a means of evaluating the model performance with respect to preserving $\rho(1)$ and K simultaneously. In neither the equal or unequal weighting case is any model consistently ranked first or second for all four study streams. However, the AMAK model is ranked either first or second for all but one of the two weighting cases for the three streams, and therefore the AMAK models is judged the overall best based on its ability to preserve $\rho(1)$ and K . The ARMA model is judged the second best overall model. Several limitations should be borne in mind with regard to the generality of this assessment of the performance of the AMAK and ARMA models. It cannot be concluded that they are the best models for any stream as can be seen from the fact that AMAK is ranked first in only 50 percent of the study cases and ARMA in only 25 percent of the cases. The general applicability of these models to a wide range of streamflow sequences has not been demonstrated in this study or elsewhere at this time. It should also be noted that there is a subjective element in the calibration of the AMAK and ARMA models, and this study has not addressed the influence of this subjective element on the high ranking of the

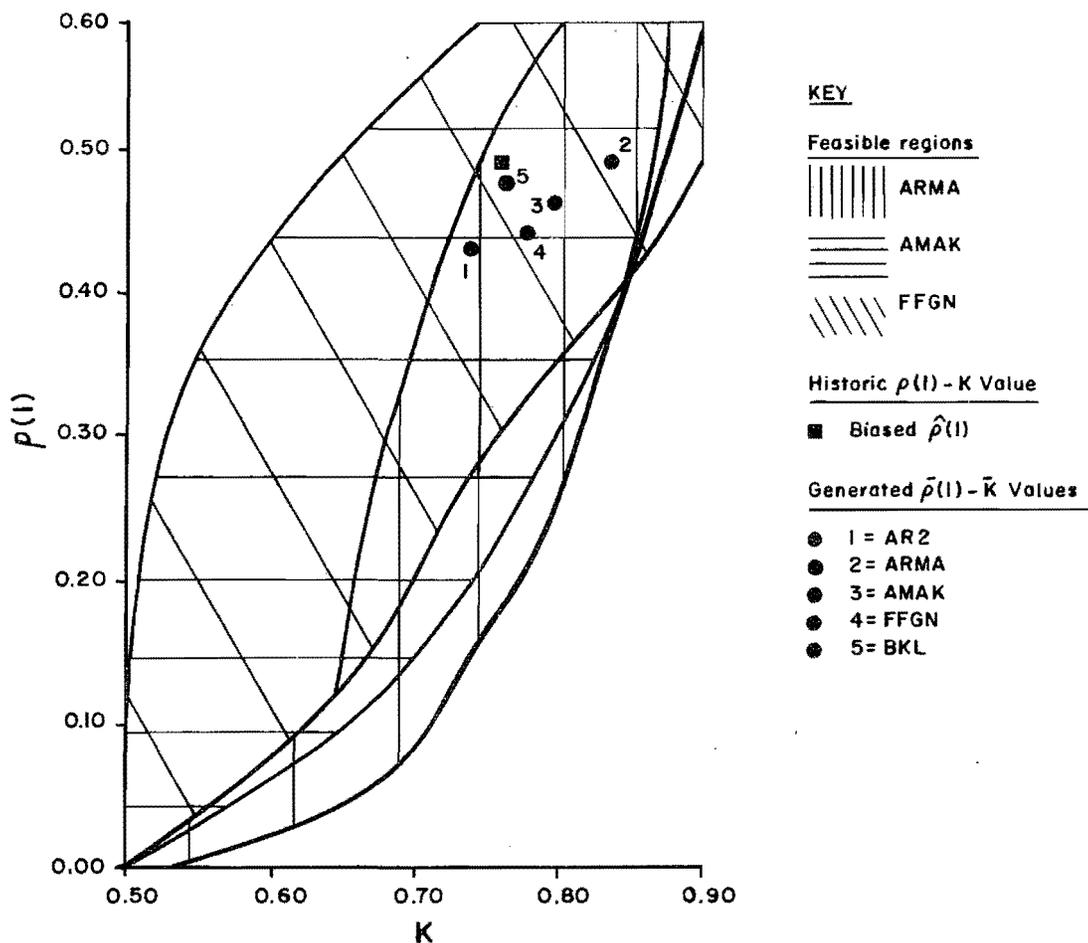


Figure 7.2. Comparison of historic and generated $\rho(1)$ -K values for Blacksmith Fork River.

AMAK and ARMA models. The five annual models were not evaluated based on their preservation of serial correlations greater than lag-one and therefore no conclusions can be made with regard to their ability to preserve the general autocorrelation structure of the historic streamflow time series.

Evaluation of Seasonal Performance

A summary of the historic and generated irrigation season statistics is included in Table 7.1. The irrigation season statistics are: E(ND), the expected number of down crossings or droughts in the N-year synthetic sequences, E(RL), the expected negative run length or drought duration in months and E(RS), the expected negative run sum or drought severity and they are contained in columns 7, 8, and 9 respectively. These statistics are drought crossing properties with respect to a crossing level defined by the agricultural demand (see column 6, Table

7.1) or irrigation requirements (see Table 6.2) which is assumed to be during the irrigation season. There is very little variation in the generated statistics between models for the same study stream. For all streams except Logan the E(ND) are slightly less than the historic number of droughts. The E(RL) for the Beaver are a little greater than historic RL, for the Blacksmith and Logan the E(ND) are approximately equal to their respective historic RL values, and for Weber E(ND) are a little less than the historic RL. The E(RS) for the Beaver are approximately equal to historic, for the Blacksmith and Weber the E(RS) are slightly greater than historic, and for the Logan the E(RS) are slightly less than historic. Thus the drought crossing properties do not appear to be very sensitive to either the choice of the annual model or the values of the lag-one autocorrelation and Hurst coefficients which are preserved.

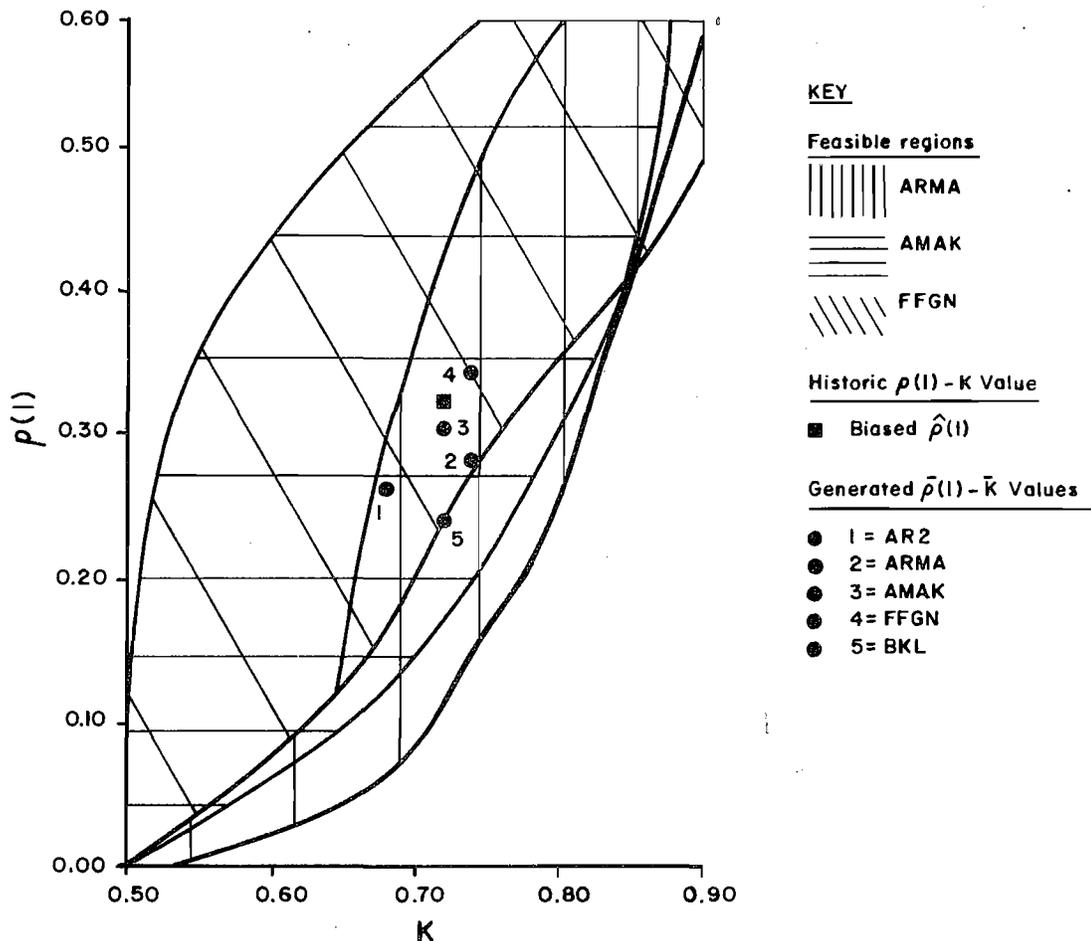


Figure 7.3. Comparison of historic and generated $\rho(1)$ -K values for Logan River.

Evaluation by Cost and Ease of Use

In addition to the adequacy of the performance of the annual models the cost and ease of their use must be considered when selecting a model. Table 7.3 contains a comparison of several measures of the cost and ease of use of the five annual models. The AR2, ARMA, and AMAK models are the least expensive to run and the FFGN is the most expensive, costing almost six times as much as the AR2. The level of effort required for parameter estimation varies from model to model and was described in Chapter 4. Most time consuming in this regard are the BKL model and AMAK models which require the use of separate programs for the estimation of model-specific parameters beyond the usual statistical moments and Hurst coefficient. Parameter estimation for the ARMA model can also require a moderate level of effort if values for ϕ and θ interpolated from O'Connell's (1974) tables are refined through Monte Carlo simulation for the sequence length and number of traces to be used in a particular application.

Comparison of Design Parameters

Two design parameters have been calculated based on the monthly flow volumes obtained by disaggregating the annual synthetic streamflow sequences generated by the five annual models. These parameters are a reservoir design capacity and a critical drought volume.

The reservoir design capacity associated with a particular reliability of supply is obtained from a probability distribution of reservoir storage volumes which are required to completely satisfy the irrigation water requirements for a hypothetical agricultural system described in Chapter 6. The reservoir storage volumes are obtained from applying the sequent peak algorithm to each of the 50 synthetic sequences. Burges and Linsley (1971) showed that reservoir storage volumes obtained from the sequent peak algorithm applied to a large number of streamflow sequences generated by a first-order autoregressive model approximately follow an extreme value type I (or Gumbel)

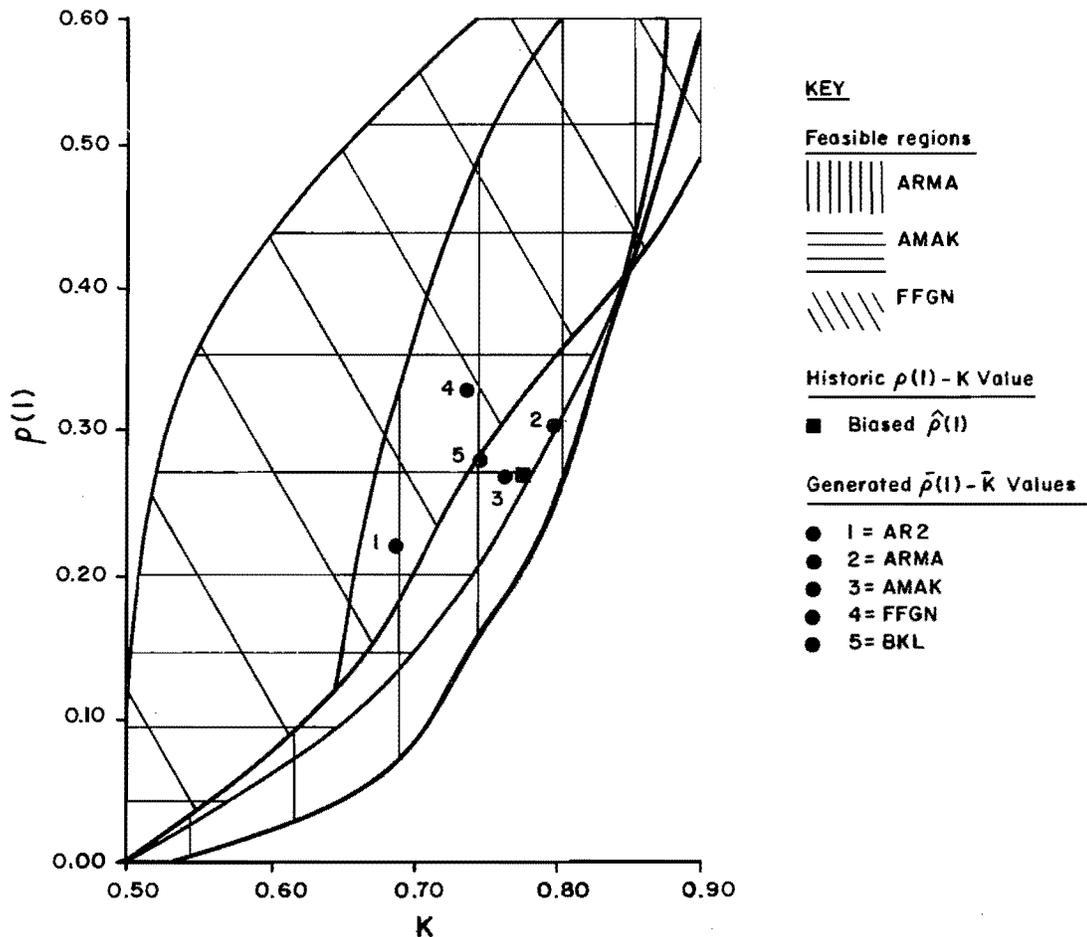


Figure 7.4. Comparison of historic and generated $\rho(1)$ - K values for Weber River.

probability distribution. Therefore, the reservoir storage volumes obtained in this study have been fitted to the Gumbel distribution. Burges and Lettenmaier (1977) have suggested that a safe practical design level of reservoir reliability is 98 percent; or equivalently a 2 percent or 1 in 50 year chance that the reservoir will fail to supply the required water. The 98 percent level of reliability has also been adopted in this study although it is recognized that different levels of reliability might be justified for different types of water use. For example, the 98 percent reliability may be too low for a cooling water supply for a nuclear reactor, but it might be unjustifiably high for water used exclusively for recreational purposes.

Distributions of reservoir storage volumes for the study streams are presented in Figures 7.7-7.10. The plotted values of storage are denoted S^* and are dimensionless storage ratios obtained by dividing the storage volumes by the mean irrigation

season diversion for each stream given in Table 6.2 (column 5).

To obtain the critical drought volume a probability distribution of the largest negative run-sums from each of the 50 synthetic traces is plotted. The negative run-sums or drought deficits are calculated with respect to the monthly irrigation requirements given in Table 6.2 for each study stream. The critical drought volume, CD_{98} is read from the distribution at the 98 percent probability of nonexceedance. Adoption of a 98 percent probability of nonexceedance was arbitrary in this case. It was found that the run-sums approximately followed the extreme value type I distribution which is not surprising because of the close relation of these negative run-sums and the reservoir storage volumes; both are range statistics. The distributions of the largest negative run-sums or largest drought deficits are presented in Figures 7.7-7.10. Plotted values of the drought deficits are denoted CD^* and are dimensionless drought

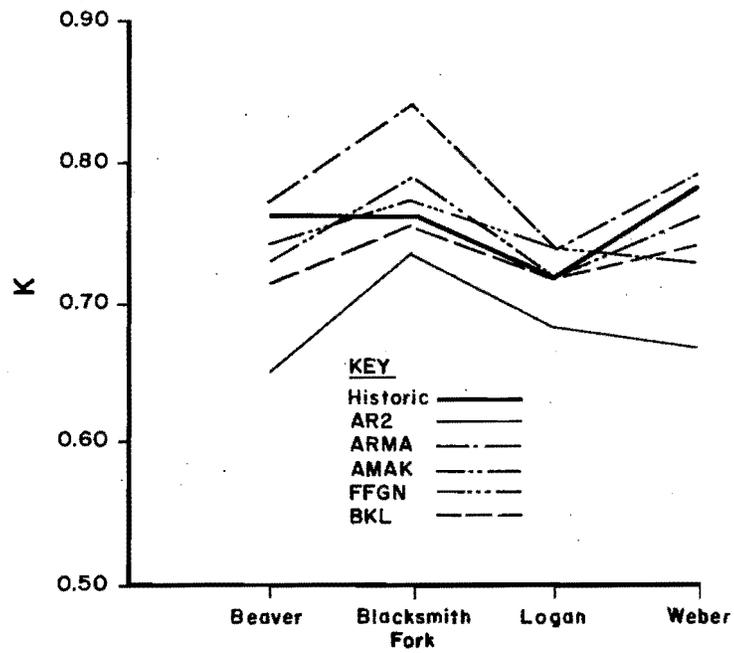


Figure 7.5. Comparison of the historic and generated $\rho(1)$ values for the study streams.

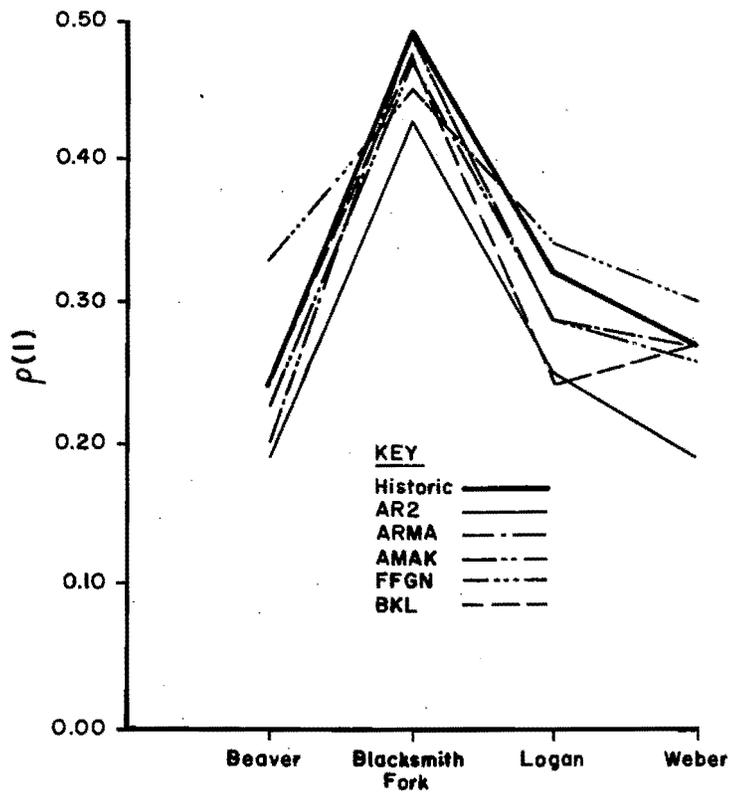


Figure 7.6. Comparison of the historic and generated K values for the study streams.

Table 7.3. Comparison of cost and ease of use of annual streamflow models.

Characteristic ^b	Model				
	AR2	ARMA	AMAK	FFGN	BKL
Lines of programming	38	40	42	114	115
Core storage (words)	8,568	8,554	8,724	14,106	11,028
Typical run time (secs)	18	19	27	105	45
Typical run cost ^a	\$0.90	\$0.95	\$1.35	\$5.25	\$2.25
Level of effort in parameter estimation	Minimal	Minimal	Minimal to Moderate	Minimal	Moderate

^a Typical run cost for generating 50 synthetic sequences approximately 65 years long.
^b For Burroughs 6800 Computer.

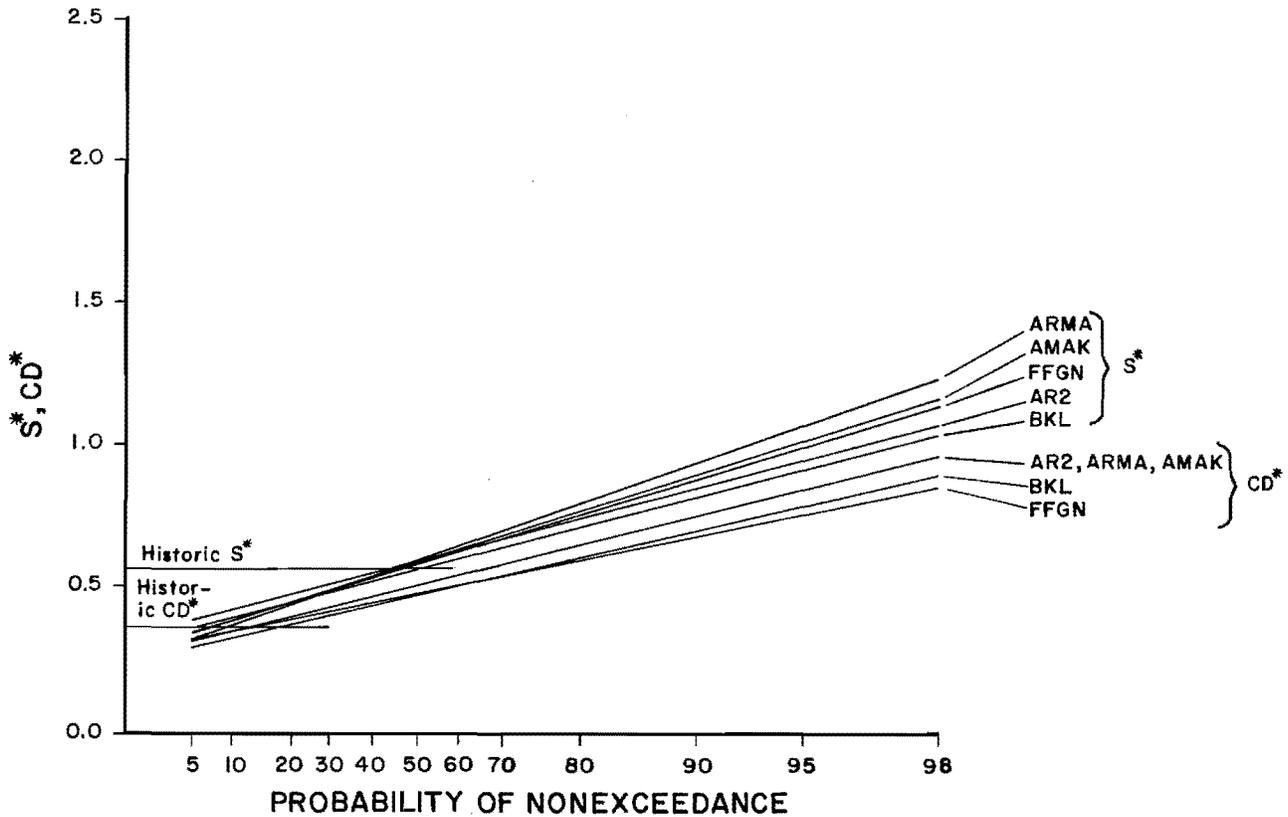


Figure 7.7. Distribution of design parameters S* and CD* for Beaver River.

ratios obtained by dividing the drought deficits by the mean irrigation season diversion for each stream given in Table 6.2 (column 5).

The plotted distributions of S* and CD* were obtained by the frequency factor approach of Chow (1951). This approach to

fitting the extreme value distribution to S* and CD* is considered preferable to the probability plotting approach because it eliminates the need for individual judgment and the possibility of personal bias. But this advantage can also be described as a disadvantage because the frequency factor procedure incorporates all data whereas the

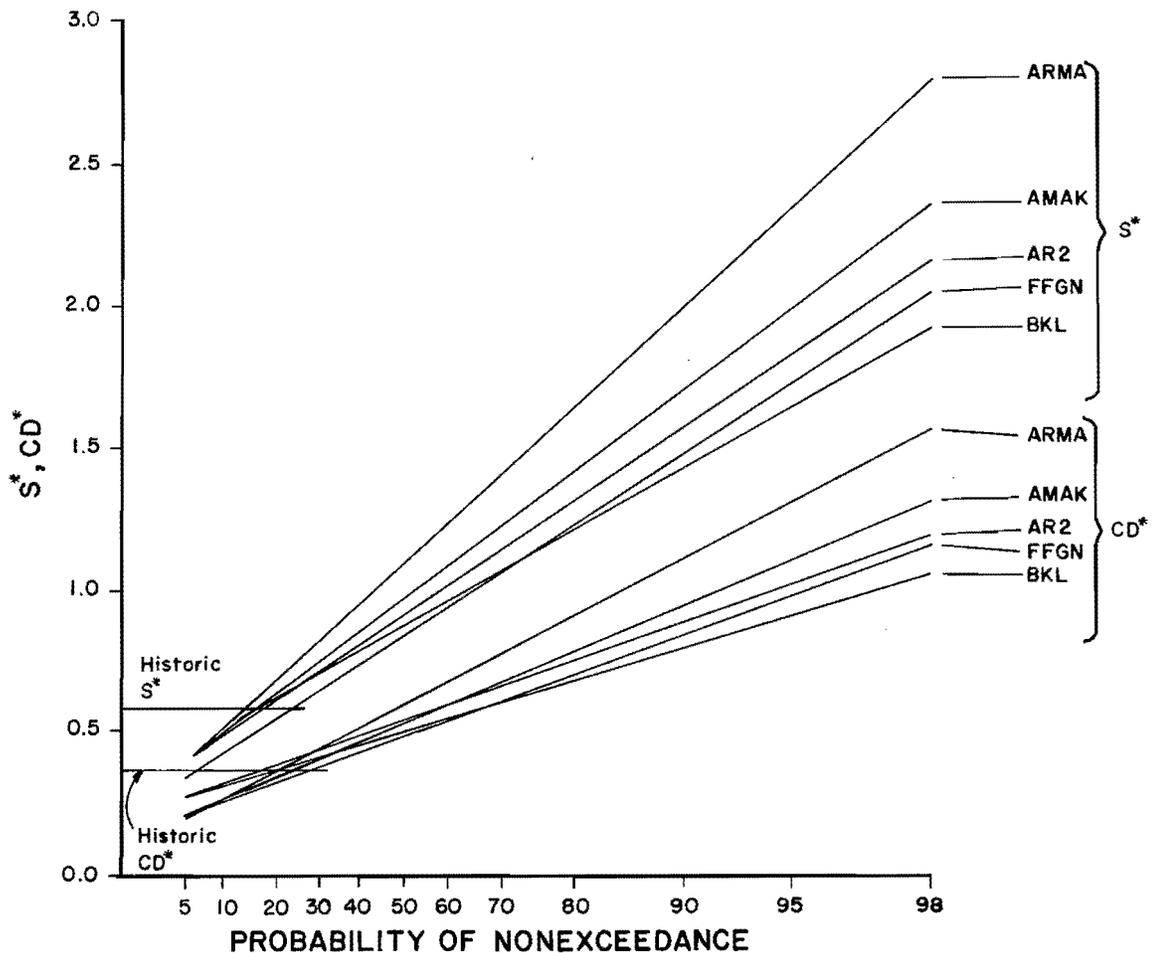


Figure 7.8. Distribution of design parameters S^* and CD^* for Blacksmith Fork River.

plotting approach allows the analyst to ignore obvious outliers in order to get a better fitting curve.

Values of S_{98}^* and CD_{98}^* obtained from Figures 7.7-7.10 are given in Table 7.1, columns 11 and 13, respectively. An examination of the ranking of these values reveals a fairly consistent trend which is similar for both design parameters in all study streams. The ARMA and AMAK models give the largest or most conservative values, the FFGN and BKL models the smallest or least conservative values, and the AR2 model generally gives values which lie in between those from the other models. It is interesting to note that since the ARMA and AMAK models were judged overall best in preserving the persistence statistics, the conservative estimates of the design parameters might be the ones that should be considered the most reliable.

Also presented in Table 7.1 in the rows labeled "Historical" are the reservoir design capacity and largest negative run-sum obtained from the historical streamflow records (columns 11 and 13, respectively).

Comparison of these values with the values obtained from the synthetic sequences indicates that the historical values are much smaller in all cases. This result refutes the results of some earlier workers in operational hydrology who were unable to generate more severe deficits than those obtained historically (e.g. Askew, Yeh, and Hall 1971). Another way of illustrating this same point is by obtaining the probability of nonexceedance (reliability for S^*) for S^* and CD^* from Figures 7.7-7.10. These probabilities are given for each stream in columns 12 and 14 of Table 7.1 based on each model. In all cases the probabilities are much less than the 98 percent values of the design parameters. In all cases the historic values of the design parameter are less than the mean of the distributions of these parameters obtained from the stochastic generation.

The values of reservoir storage, S^* , must be greater than or equal the values of drought deficit, CD^* , so that the drought deficit can be met from storage. The degree to which S^* exceeded CD^* varies from stream

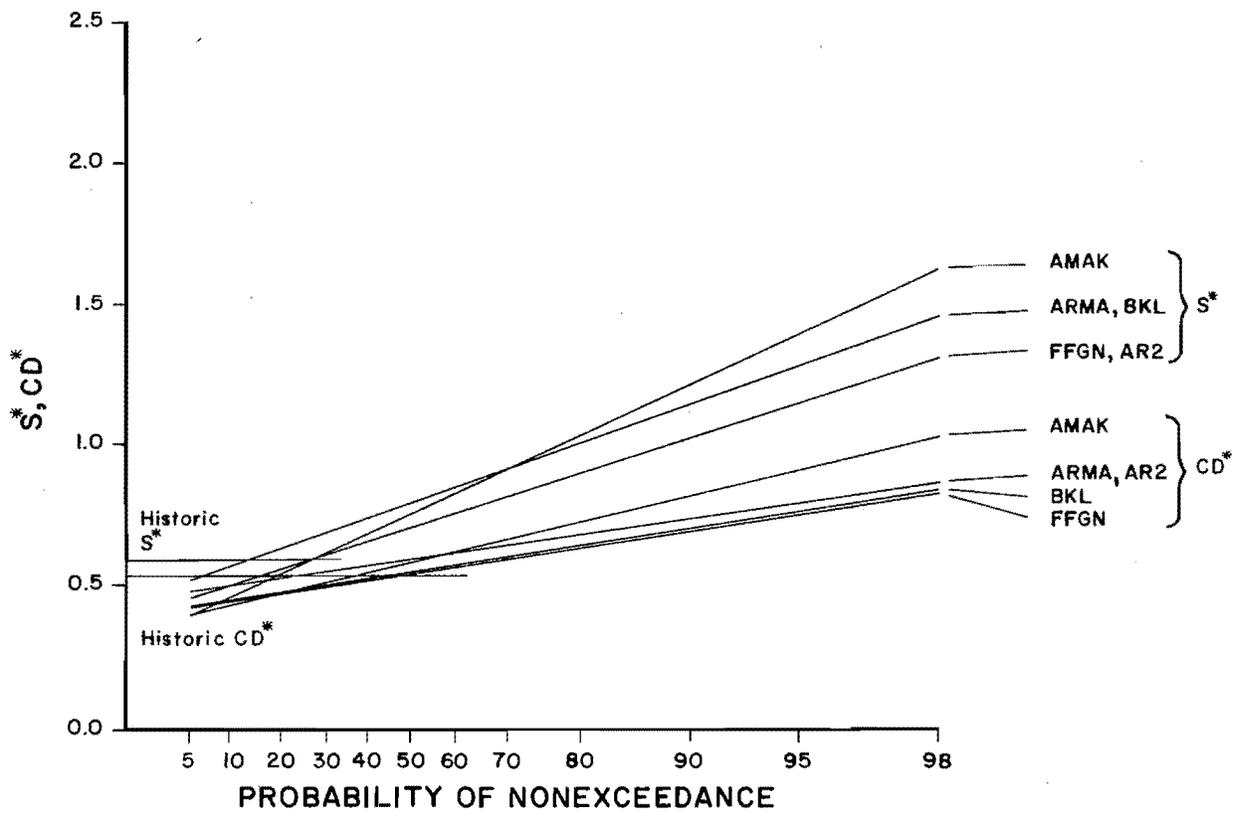


Figure 7.9. Distribution of design parameters S^* and CD^* for Logan River.

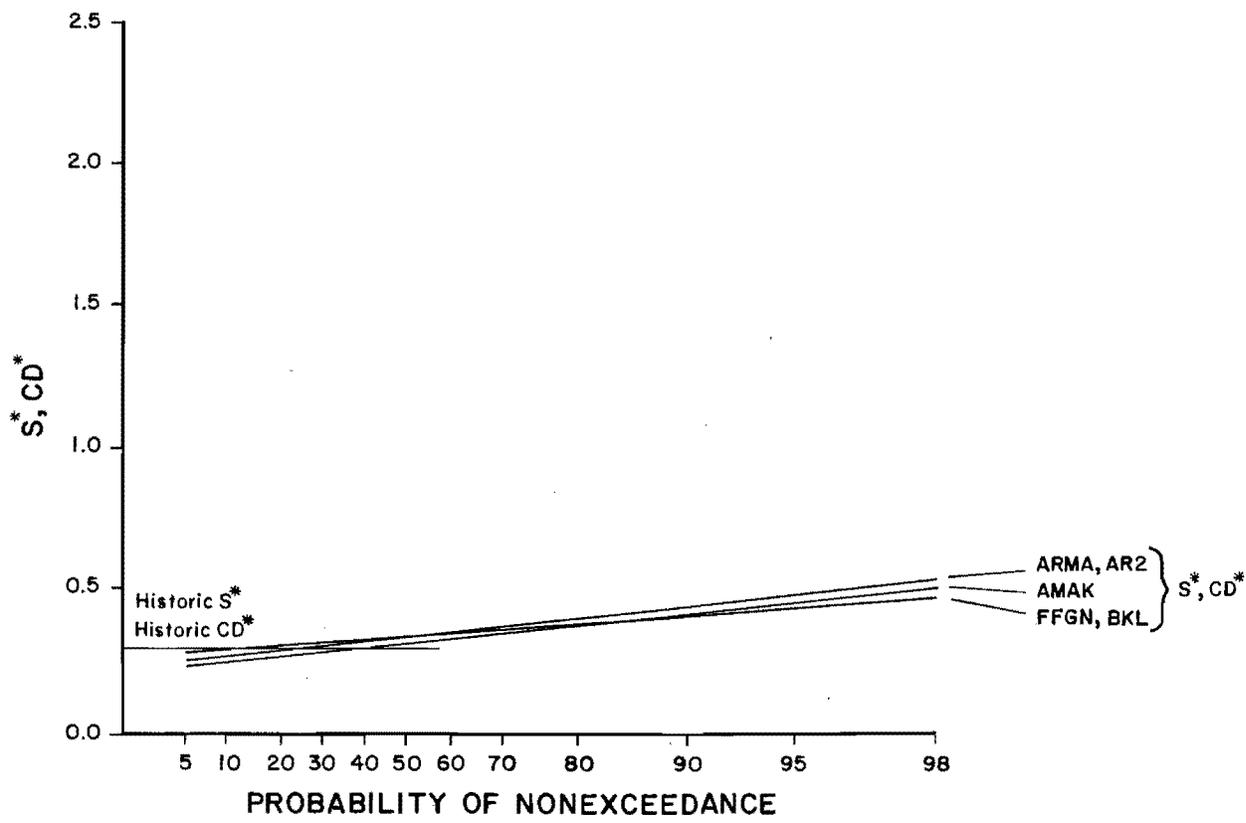


Figure 7.10. Distribution of design parameters S^* and CD^* for Weber River.

to stream as can be seen by comparing the difference in location of the distributions of the two distributions in Figures 7.7-7.10. For the Blacksmith Fork and Logan Rivers which have the largest values of the dimensionless demand, D^* (mean seasonal demand divided by mean seasonal diversion) the difference is greatest since there is a greater tendency to need carry-over storage after a drought to satisfy a shortage in the following year. The difference in the distributions of S^* and CD^* is also affected by the level of persistence in the streamflows. Hence, Logan, which has lower values of $\rho(1)$ and K has smaller differences between the distributions of S^* and CD^* than does Blacksmith Fork which has higher values of $\rho(1)$ and K . The effect of the lower persistence statistics for the Logan appears to offset the affect of its slightly higher D^* in this case. Beaver and Weber Rivers have lower values of D^* and hence the distributions of S^* and CD^* are closer together and are almost identical for the Weber which has the lowest value of D^* and a low $\rho(1)$.

Evaluation by Economic Regret

The average annual economic benefit from the annual series of crop yields was computed with each model for each stream using the crop production function described in Chapter 6. The crop production function was most sensitive to streamflow shortages during the critical pollination growth stage in that any water deficiency during the critical period resulted in a lower yield. For grain corn the critical month is July. Thus, prolonged low flows generated by the stochastic models would result in a lower crop yield as well as longer run lengths and larger run sums.

The estimate of the average annual economic benefit from crop yield was the basis for the calculation of economic regret. The total economic regret associated with the selection of each model was calculated using Equation 6.5. Table 7.4 contains the calculation of total economic regret for each model applied to the four study streams.

Table 7.4. Economic regret^a in dollars per year for the study streams.

Stream	Model assumed to be true	Model used to calculate benefits				
		AR2	ARMA	AMAK	FFGN	BKL
Beaver	Historic	-2,800	-2,700	-2,100	-800	-3,300
	AR2	0	100	700	2,000	500
	ARMA	-100	0	600	1,900	-600
	AMAK	-700	-600	0	1,300	-1,200
	FFGN	-2,000	-1,900	-1,300	0	-2,500
	BKL	500	600	1,200	2,500	0
	Total regret	5,100	-4,500	-900	6,900	-8,100
Blacksmith Fork	Historic	-6,400	-6,100	-1,200	-3,700	-3,300
	AR2	0	300	5,200	2,700	3,100
	ARMA	-300	0	4,900	2,400	2,800
	AMAK	-5,200	-4,900	0	-2,500	-2,100
	FFGN	-2,700	-2,400	2,500	0	400
	BKL	-3,100	-2,800	2,100	-400	0
	Total regret	-17,700	-15,900	13,500	-1,500	-900
Logan	Historic	-14,000	-15,000	-3,000	-10,000	-7,000
	AR2	0	-1,000	11,000	4,000	7,000
	ARMA	1,000	0	12,000	5,000	8,000
	AMAK	11,000	-12,000	0	-7,000	-4,000
	FFGN	4,000	-5,000	7,000	0	3,000
	BKL	-7,000	-8,000	4,000	-3,000	0
	Total regret	-35,000	-41,000	31,000	-11,000	7,000
Weber	Historic	-64,000	-66,000	-67,000	-72,000	-63,000
	AR2	0	-2,000	-3,000	-8,000	1,000
	ARMA	2,000	0	-1,000	-6,000	3,000
	AMAK	3,000	1,000	0	-5,000	4,000
	FFGN	8,000	6,000	5,000	0	9,000
	BKL	-1,000	-3,000	-4,000	-9,000	0
	Total regret	-52,000	-64,000	-70,000	-100,000	-46,000

^a Economic regret calculated using Equation 6.5.

For each stream the two models with the lowest total economic regret are listed in Table 7.2 together with the two models which result in the smallest deviations between the historic and generated values of $\rho(1)$ and K based on equal and unequal weighting of the deviations. The BKL model minimizes economic regret for all streams except the Beaver and is clearly the overall best model with respect to the regret criterion. The FFGN model appears to be the overall second best model. It is observed that for the Beaver the ARMA model was ranked first or second by all three criteria and for Blacksmith Fork the FFGN model was similarly placed. Also for three out of four of the study streams the same model is ranked first or second by the economic regret and by at least one of the persistence statistics criteria. Although the BKL model is ranked first for three out of four of the study streams based on economic regret it appears only once in second place based on the persistence statistics criterion. In fact the economic regret and persistence statistics criteria did not select any of the same overall best models. This implies that the objective of preserving the persistence statistics is not compatible with the objective of minimizing economic regret for the study streams. It should be noted that this conclusion is subject to the same limitations with respect to its generality as were discussed in the summary of the section on "Evaluation of Annual Performance." The low estimates of regret obtained from the BKL and FFGN models result from the tendency of these models to generate droughts with severities of magnitudes inbetween those generated by the other models. It should be noted that this property does not conflict with the fact that the BKL and FFGN models gave the smallest values of S98 and CD98 since these are extreme value statistics and economic regret is not.

A Model Choice Strategy

A decision regarding which stochastic streamflow model to choose for data generation should consider the following factors:

- 1) Ability to preserve relevant statistical characteristics of the historic streamflow time series.
- 2) Cost of using the technique measured in terms of computer costs and labor costs for calibration.
- 3) Economic regret resulting from the use of inaccurate design parameters obtained.

The work completed in this study has considered each of the above factors but only

for a very limited sample of four Utah streams. Thus, it is not possible to formulate a very generalized model choice strategy based only on this study. In addition, economic regret will vary so much for different uses of synthetic sequences that it is not possible to include it in a generalized model choice strategy. Therefore, the proposed model choice strategy will consider only the first two factors and is also based on the work of other researchers in order to broaden its applicability. To the extent that the proposed model choice strategy is based on work reported herein it assumes that preservation of $\rho(1)$ and K, and not the entire autocorrelation structure, is the goal of the analyst.

The proposed model choice strategy is limited to the selection of a univariate annual stochastic streamflow model. For annual models the model choice is based on the $\rho(1)$ -K values estimated from the historic record and the feasible regions for each of the five models considered in this study. Figure 7.11 contains the recommended initial model choice for each $\rho(1)$ -K combination and covers the usual range of values for these persistence statistics. Where feasible regions for different models overlap, selection of the recommended model was based on the ranking with respect to the preservation of persistence statistics and the cost and ease of use.

Hoshi, Burges, and Yamaoka (1978) showed that there was little advantage to using a long-term persistence model (i.e. ARMA, AMAK, FFGN, or BKL) if the value of K is less than 0.7. This result applies throughout the usual range of $\rho(1)$ values found in streamflows, that is $\rho(1)$ less than 0.6, and therefore the AR2 model is recommended for K less than 0.7 (see Figure 7.11). The feasible range of the FFGN model for K greater than 0.7 is completely covered by either the AMAK or ARMA models. Since these models are less expensive to run and were shown in this study to be more effective at preserving $\rho(1)$ and K than FFGN, the FFGN model is not included as a recommended model in Figure 7.11. The AMAK model is recommended over the ARMA model because of its superior performance in preserving the persistence statistics. Since there are no other choices below the lower boundary of that ARMA feasible region the BKL model is recommended in the region.

Based on Figure 7.11 the AMAK model would be selected for all the study streams. Since the AMAK model is ranked first or second in Table 7.2 in all but one of the cases using the persistence statistics criteria, this would be an acceptable choice.

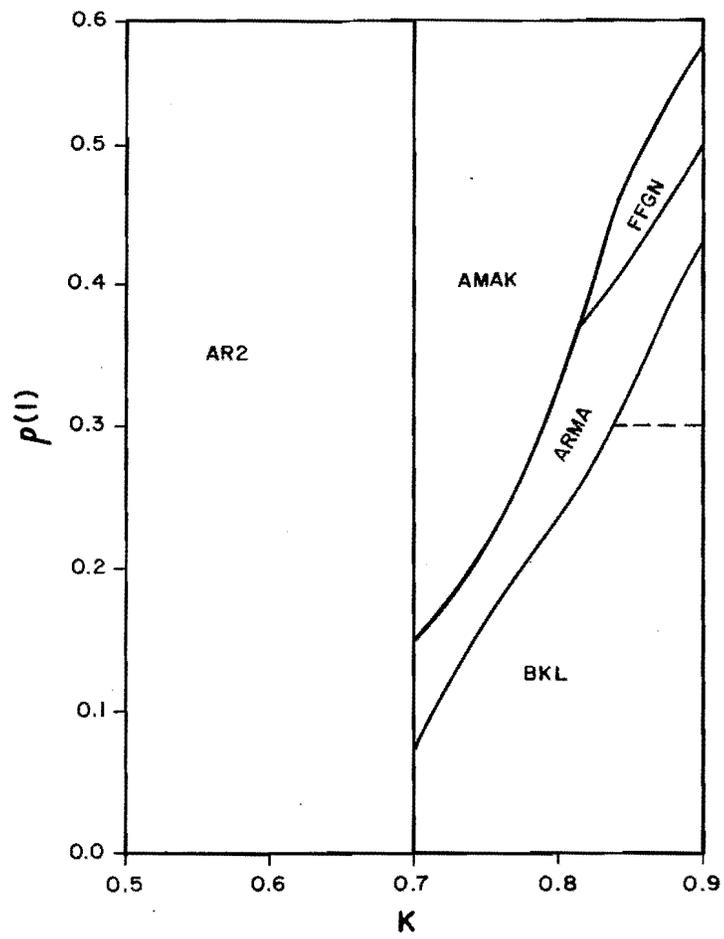


Figure 7.11. Recommended annual streamflow stochastic models based on $\rho(1)$ - K values.

CHAPTER 8

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The purpose of this study was to perform an operational comparison of five annual stochastic streamflow models and to develop a strategy for model selection. Each stochastic model was applied to four Utah streams which were selected at locations above which little development has taken place. The annual models used were: second-order autoregressive (AR2), autoregressive moving-average (AKMA), ARMA-Markov (AMAK), fast fractional Gaussian noise (FFGN), and broken line (BKL). In applying these models to the study streams a systematic modeling procedure was utilized, comprising the following five steps: 1) identification of water resource system and model composition, 2) choice of model type, 3) identification of model form, 4) parameter estimation, and 5) model performance evaluation (see Figure 1.1). Step 1 typically involves decisions about the structure of a water resources simulation model, its inputs, state variables, outputs, temporal and spatial resolution, etc., needed to provide the required information for problem resolution. In this study step 1 involved a statement of the research objective in Chapter 1, a brief description of the four streamflow gaging sites in Chapter 3, and the recognition of the need for univariate model structure with disaggregation to monthly flows to be generated during the irrigation season. The second step, model choice, was omitted so that all five annual models could be applied to all streams and operational comparisons made. Based on these comparisons a strategy for model choice was recommended at the end of Chapter 7. Identification of model form (step 3) and parameter estimation (step 4) are described for the annual models in Chapter 4 and for the disaggregation models in Chapter 5. One of two types of disaggregation models was used to divide the generated annual flows into monthly flows. These models are the Valencia-Shaake (VS) and Mejia-Rousselle (MR) models. Step 5, model performance evaluation is described in Chapter 7. This step comprised an evaluation of the preservation of annual persistence statistics and seasonal crossing properties, the cost and ease of model use, and the magnitude of the economic regret associated with drought related agricultural losses, and a comparison of reservoir capacity and critical drought design parameters. Agricultural losses

were estimated using a crop yield model which is described in Chapter 6 together with the procedure for calculating economic regret. The model choice strategy recommended at the end of Chapter 7 is based on the $\rho(1)$ and K values estimated from the historic record.

Conclusions

The following conclusions have been developed from the results and experience gained during this study:

1. The AMAK and ARMA models were judged best in terms of preserving the lag-one autocorrelation and Hurst coefficients, which are measures of the short and long term persistence of the streamflow sequences.

2. The BKL model is judged the overall best model in terms of minimizing the economic regret. However, the BKL model performed poorly with respect to preserving the persistence statistics and appears to have underestimated the design parameters.

3. As would be expected the AR2 model does not adequately preserve the Hurst coefficient based on O'Connell (1974) except when the lag-one autocorrelation coefficient is quite large (e.g. 0.49, see Blacksmith Fork).

4. All five stochastic models generate average design parameters which are greater than the values based on historic record. The AMAK and ARMA models consistently gave the largest values of the design parameters, reservoir storage capacity and critical drought based on 98 percent probability of nonexceedance.

5. The seasonal crossing properties do not appear to be very sensitive to either model choice or the magnitude of $\rho(1)$ and K .

6. Comparisons of computer operating costs for the five models show that the FFGN model is the most expensive to run costing approximately six times as much as the AR2 model. Parameter estimation for the BKL and AMAK models requires the most effort by the analyst.

7. The Valencia and Schaake (VS) disaggregation model preserves the monthly

cross-correlations and lag-one autocorrelations as well as the Mejia and Roussele (MR) model. The only justification for using the more complicated MR model is to preserve the over-the-year monthly correlations. In many studies the timing of the end of the water year can be arranged to fall at a time when these serial correlations between months in adjacent years are not important operationally and therefore the VS model is suitable for many applications.

8. The positive semidefinite property of the BB^T (or EE^T) matrix for seasonal disaggregation model parameters was found to be sensitive to the transformation selected to remove skew from the historic streamflows, nonhomogeneities in the streamflow record, and the order m , of Z_t for the MR model. It was found that different choice of transformation or a slightly changed starting year of historic record could change BB^T (or EE^T) from negative to positive semidefinite.

9. A model choice strategy for selecting an annual stochastic streamflow model based on the values of $\rho(1)$ and K estimated from the historic streamflow record. This procedure does not necessarily select the best model for a particular stream but it does select one of the better models and will avoid the use of an unnecessarily complex model.

10. An alternative parameter estimation procedure for the AMAK model led to parameter values which preserved the persistence statistics better than the Burges and Lettenmaier (1977) method which is based on fitting the theoretical autocorrelation function for fractional Gaussian noise at three arbitrary lags. The alternative procedure was to use parameters for the ARMA part of the AMAK model which were estimated from O'Connell (1974), to use $\rho(1)$ for the Markov parameter,

and to estimate C_1 and C_2 by trial and error subject to the constraint that $C_1 + C_2 = 1$.

11. An alternative procedure for assigning the value of the lag-one autocorrelation coefficient of the high frequency component, $\rho(1)^{(HF)}$, of the FFGN model was found to give better preservation of $\rho(1)$. The alternative procedure was to set $\rho(1)^{(HF)}$ equal to the historic estimate of $\rho(1)$ instead of the conventional procedure of calculating it with Equation 4.32.

Recommendations

The following recommendations for further research are based on the experience gained during this study:

1. The alternative AMAK parameter estimation procedure used in the study should be compared with the procedure proposed by Burges and Lettenmaier (1977) to compare the effect of using each procedure on the design parameters.

2. The values of design parameters did not appear to be very sensitive to the model choice or the magnitudes of the persistence statistics for the four study streams. It is recommended that the sensitivity of these design parameters to a wide range $\rho(1)$ and K values be explored for each model. A sensitivity study of the effects of different values of the persistence statistics on the model regret should also be conducted. These sensitivity studies might provide information for improved model choice decisions near the boundaries of the feasible regions where the choice is between a complicated and a simple model because it might be possible to predict the effects of preserving slightly changed values of the persistence statistics by using the simpler model.

REFERENCES

- Anis, A. A., and E. H. Lloyd. 1975. Skew inputs and the Hurst effect. *Journal of Hydrology* 26:39-53.
- Askew, A. J., W. W-G. Yeh, and W. A. Hall. 1971. A comparative study of critical drought simulation. *Water Resources Research* 7(1):52-62.
- Bendat, J. S., and A. G. Piersol. 1966. *Random data: Analysis and measurement procedures*. Wiley-Interscience. New York. 390 p.
- Bobee, B. and R. Robitaille. 1975. Correction of bias in the estimation of the coefficient of skewness. *Water Resources Research* 11(6):851-854.
- Boes, D. C., and J. D. Salas-LaCruz. 1973. On the expected range and the expected adjusted range of partial sums of exchangeable random variables. *Journal of Applied Prob.* 10:671-677.
- Bowles, D. S. 1979. Unpublished class notes, CEE 644, Operational Hydrology. Department of Civil and Environmental Engineering, Utah State University, Logan, Utah.
- Bowles, D. S. 1980. Discussion of 'Disaggregation of streamflow volumes' by Hoshi and Burges. *Journal of the Hydraulics Division, Am. Soc. Civ. Engr.* 106(HY1):217-218.
- Box, G. E. P., and D. R. Cox. 1964. An analysis of transformations. *Journal Royal Stat. Soc.* B26:211.
- Box, G. E., and G. M. Jenkins. 1970. *Time series analysis: Forecasting and control*. Holden-Day, San Francisco, California. 553 p.
- Briggs, G. E. 1925. Plant yield and intensity of external factors. *Mitscherlich's Wirkungsgesetz*. *Ann. Bot.* 39:475-502.
- Burges, S. J., and D. P. Lettenmaier. 1975. Operational comparison of stochastic streamflow generation procedures. Charles W. Harris Hydraulics Laboratory Technical Report No. 45. University of Washington, Seattle, Washington. 112 p.
- Burges, S. J., and D. P. Lettenmaier. 1977. Comparison of annual streamflow models. *Journal of the Hydraulics Division, Am. Soc. Civ. Engr.* 103(HY9):991-1005.
- Burges, S. J., and R. K. Linsley. 1971. Some factors influencing required reservoir storage. *Journal of the Hydraulics Division, Proc. Am. Soc. Civ. Engr.* 103(HY7):977-991.
- Chi, M., E. Neal, and G. K. Young. 1973. Practical application of fractional Brownian motion and noise to synthetic hydrology. *Water Resources Research* 9(6):1523-1533.
- Chow, V. T. 1951. A generalized formula for hydrologic frequency analysis. *Trans. Am. Geophy. Union* 32(2):231-237.
- Crosby, D. S., and T. Maddock, III. 1970. Estimating coefficients of a flow generator for monotone samples of data. *Water Resources Research* 6(4):1079-1086.
- Curry, K., and R. L. Bras. 1978. Theory and applications of the multivariate broken line, disaggregation and monthly autoregressive streamflow generators to the Nile River. *Technology Adaptation Program Report 78-5*. Massachusetts Institute of Technology, Cambridge, Massachusetts. 416 p.
- Feller, W. 1951. The asymptotic distribution of the range of sums of independent random variables. *Ann. Math. Stat.* 22:427-432.
- Fiering, M. B. 1967. *Streamflow synthesis*. The McMillan Press Ltd., London. 139 p.
- Fiering, M. B. 1968. Schemes for handling inconsistent matrices. *Water Resources Research* 4(2):291-297.
- Fiering, M. B., and B. B. Jackson. 1971. *Synthetic streamflows*. *Water Resources Monograph 1*. American Geophysical Union. 98 p.
- Gomide, F. L. S. 1978. Markovian inputs and the Hurst phenomenon. *Journal of Hydrology* 37:23-45.
- Gowon, D. T., J. C. Anderson, and B. Biswas. 1978. An economic interpretation of impact of phenologically timed irrigation on corn yield. *Western Journal of Agricultural Economics*. 145-156.

- Hall, W. A., A. J. Askew, and W. W-G Yeh. 1969. Use of the critical period in reservoir analysis. *Water Resources Research* 5(6): 1205-1215.
- Hanks, R. J. 1974. Model for predicting plant yield as influenced by water use. *Agronomy Journal* 65:660-665.
- Hexem, R. W., and E. O. Heady. 1978. Water production functions for irrigated agriculture. Iowa State University Press, Ames, Iowa. 215 p.
- Hinkley, D. 1977. On quick choice of power transformation. *Appl. Statist.* 26:67-69.
- Hoshi, K., S. J. Burges, and I. Yamaoka. 1978. Reservoir design capacities for various seasonal operational hydrology models. *Proc. of Japan. Soc. Civ. Engr.* 273:121-134.
- Hoshi, K., and S. J. Burges. 1979. Disaggregation of streamflow volumes. *Journal of the Hydraulics Division, Am. Soc. Civ. Engr.* 105(HY1):27-41.
- Hurst, H. E. 1951. Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civ. Engrs.* 116:770-808.
- Hurst, H. E. 1956. Methods of using long-term storage in reservoirs. *Proc. Instn, Civ. Engrs.* 1:519-543.
- Jackson, B. B. 1975. The use of streamflow models in planning. *Water Resources Research* 11(1):54-63.
- James, L. Douglas, David S. Bowles, W. Robert James, and Ronald V. Canfield. 1979. Estimation of water surface elevation probabilities and associated damages for the Great Salt Lake. Report UWRL/P-79/03. Utah Water Research Laboratory, Logan, Utah. 182 p.
- Jenkins, G. M., and D. C. Watts. 1968. Spectral analysis and its applications. Holden-Day Inc., San Francisco, California. 525 p.
- Jettmar, R. V., and G. K. Young. 1975. Hydrologic estimation and economic regret. *Water Resources Research* 11(5):648-656.
- Kendall, M. G. 1954. Note on the bias in the estimation of autocorrelation. *Biometrika* 42:403-404.
- Kendall, M. G. 1961. A course in multivariate analysis. Charles Griffin and Co. Ltd., London. 185 p.
- Kendall, M. G., and A. Stuart. 1968. The advanced theory of statistics. Vol. 3. Charles Griffin and Co. Ltd., London. 522 p.
- Kendall, M. G., and A. Stuart. 1973. The advanced theory of statistics, Volume 2, Third Edition, Griffin, London. pp. 153-157.
- Klemes, V. 1974. The Hurst phenomenon: A puzzle? *Water Resources Research* 10(4):675-688.
- Kottegoda, N. T. 1974. Effect of skewness in three stochastic pentad river flow models on crossing properties of synthesized data. *Water Resources Research* 10(3):446-456.
- Kottegoda, N. T. 1980. Personal communication.
- Kottegoda, N. T. 1980. Stochastic water resources technology. John Wiley & Sons, New York. 384 p.
- Lane, W. L. 1979. Applied stochastic techniques, user manual. U. S. Bureau of Reclamation, Div. of Planning Tech. Services, Eng. and Research Center, Denver, Colo. 75 p.
- Lawrance, A. J., and N. T. Kottegoda. 1977. Stochastic modelling of riverflow time series. *Journal Royal Statistics Soc. Series A.* 140, Part 1:1-47.
- Lettenmaier, D. P., and S. J. Burges. 1977. Operational assessment of hydrologic models of long-term persistence. *Water Resources Research* 13(1):113-124.
- Lloyd, E. H. 1967. Stochastic reservoir theory. *Advances in Hydroscience.* Ed. V. T. Chow, 4:281-339.
- Maass, A. 1962. Design of water resources systems. Harvard University Press, Cambridge, Massachusetts. 660 p.
- Mandelbrot, B. B. 1971. A fast fractional Gaussian noise generator. *Water Resources Research* 7(3):543-553.
- Mandelbrot, B. B., and J. W. Van Ness. 1968. Fractional Brownian motions, fractional noises and applications. *Soc. for Industrial Appl. Math. Review* 10(4):422-437.
- Mandelbrot, B. B., and J. R. Wallis. 1968. Noah, Joseph, and operational hydrology. *Water Resources Research* 4(5):909-918.
- Mandelbrot, B. B., and J. R. Wallis. 1969a. Part 1, Averages and variances; computer experiments with fractional Gaussian noises. *Water Resources Research* 5(1):228-241.
- Mandelbrot, B. B., and J. R. Wallis. 1969b. Part 2, Rescaled ranges and spectra; computer experiments with fractional Gaussian noises. *Water Resources Research* 5(1):242-259.

- Mandelbrot, B. B., and J. R. Wallis. 1969c. Part 3, Mathematical appendix; computer experiments with fractional Gaussian noises. *Water Resources Research* 5(1):260-267.
- Matalas, N. C. 1966. Time series analysis. *Water Resources Research* 3(3):817-829.
- Matalas, N. C. 1967. Mathematical assessment of synthetic hydrology. *Water Resources Research* 3(4):931-935.
- Matalas, N. C. 1975. Developments in stochastic hydrology. *Reviews of Geophysics and Space Physics* 13(3):67-73.
- Matalas, N. C., and C. S. Huzzen. 1967. A property of the range of partial sums. *Proc. of the Intern. Hydrol. Symp. Fort Collins, Colorado.* 1:252-257.
- Matalas, N. C., and J. R. Wallis. 1976. Generation of synthetic flow sequences. In: A. K. Biswas (Ed.) *Systems Approach to Water Management.* McGraw-Hill, New York. 429 p.
- McLeod, A. I., and K. W. Hipel. 1978. Preservation of the rescaled range 1. A reassessment of the Hurst phenomenon. *Water Resources Research* 14(3):491-508.
- Mejia, J. M., I. Rodriguez-Iturbe, and D. R. Dawdy. 1972. 2-The broken line process as a potential model for hydrologic simulation: Streamflow simulation. *Water Resources Research* 8(4): 931-941.
- Mejia, J. M., and J. Rousselle. 1976. Disaggregation models in hydrology revisited. *Water Resources Research* 12(2):185-186.
- Millan, J. and V. Yevjevich. 1971. Probabilities of observed droughts. *Hydrology Paper No. 50, Colorado State University, Fort Collins, Colorado.* 50 p.
- Moran, P. A. P. 1964. On the range of cumulative sums. *Ann. Inst. Stat. Math.* 16:109-112.
- Moran, P. A. P. 1968. *An Introduction to probability theory.* Clarendon Press, Oxford. 542 p.
- O'Connell, P. E. 1971. A simple stochastic modeling of Hurst's law. *Proc. Inter. Symp. Math. Models Hydr. International Assoc. on Scientific Hydrology. Warsaw.*
- O'Connell, P. E. 1974. Stochastic modeling of long-term persistence in streamflow sequences. *Doctoral thesis, Civil Engineering Department, Imperial College, London, England.* 309 p.
- Potter, K. W. 1975. Comment on 'The Hurst Phenomenon: A puzzle?' by Klemes. *Water Resources Research* 11(2):373-374.
- Potter, K. W. 1976. Evidence for non-stationarity as a physical explanation of the Hurst phenomenon. *Water Resources Research* 12(5):1047-1052.
- Rodriguez-Iturbe, I., J. M. Mejia, and D. R. Dawdy. 1972. 1-A new look at Markovian models, fractional Gaussian noise and crossing theory: Streamflow simulation. *Water Resources Research* 8(4): 921-930.
- Rogers, P. 1978. On the choice of the 'appropriate model' for water resources planning and management. *Water Resources Research* 14(6):1003-1010.
- Salas, J. D., D. C. Boes, V. Yevjevich, and G. G. S. Pegram. 1977. On the Hurst phenomenon. Paper presented at the Third International Hydrology Symposium. Fort Collins, Colorado. 33 p.
- Salas, J. D., J. W. Delleur, V. Yevjevich, and W. L. Lane. 1980. Applied modeling of hydrologic time series. *Water Resources Publ. Littleton, Colorado.* 484 p.
- Salas, J. D., and R. A. Smith. 1980. Uncertainties in hydrologic time series analysis. *Am. Soc. Civ. Engr. Convention. Preprint 80-158, Portland, Oregon.* 22 p.
- Saldarriaga, J., and V. Yevjevich. 1970. Applications of run-lengths to hydrologic series. *Hydrology Paper No. 40, Colorado State University, Fort Collins, Colorado.* 56 p.
- Slack, J. R. 1972. Bias, illusion and denial as data uncertainties. *Intern. Symp. on Uncertainties in Hydro. and Water Resources Systems. University of Arizona. Tucson, Arizona, 1:122-132.*
- Slack, J. R. 1973. I would if I could (Self-denial by conditional models). *Water Resources Research* 9(1):247-249.
- Slack, J. R., J. R. Wallis, and N. C. Matalas. 1975. On the value of information to flood frequency analysis. *Water Resources Research* 11(5):629-647.
- Stewart, J. I., R. E. Danielson, R. J. Hanks, E. B. Jackson, R. M. Hoggan, W. O. Pruitt, W. T. Franklin, and J. P. Riley. 1977. Optimizing crop production through control of water and salinity levels in the soil. *Utah Water Research Laboratory. Report No. PRWG151-1, Utah State University, Logan, Utah.* 191 p.
- Thomas, H. A., Jr. 1967. Capacity expansion of public works report, Harvard University, Cambridge, Massachusetts. 14 p.
- Thomas, H. A., and M. B. Fiering. 1962. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In: *Design of Water Resources System,* edited by

- A. Maass et al., Harvard University Press, Cambridge, Massachusetts pp. 459-493.
- Todini, E., and P. E. O'Connell. 1979. Hydrological simulation of Lake Nasser. 1, Analysis and Results. Unpublished Report. 200 p.
- U. S. Geological Survey. 1979. Water resources data of Utah. Part 1. Surface water records. United States Department of the Interior. U. S. Government Printing Office, Washington, D. C. 299 p.
- U. S. Soil Conservation Service. 1970. Irrigation water requirements. U. S. Department of Agriculture. Technical Release No. 21. 88 p.
- Utah Agricultural Statistics. 1978. Department of Agriculture, Salt Lake City, Utah. 104 p.
- Valencia, R. D., and J. C. Schaake. 1973. Disaggregation processes in stochastic hydrology. Water Resources Research 9(3):580-585.
- Vicens, G. J., and J. C. Schaake. 1980. Design of length of water resource simulation experiments. Journal of the Water Resources Planning and Management Division. Am. Soc. Div. Engr. 106(WR1): 333-350.
- Wallis, J. R., and N. C. Matalas. 1970. Small sample properties of H and K estimators of the Hurst coefficient h. Water Resources Research 6(6):1583-1594.
- Wallis, J. R., and N. C. Matalas. 1971. Correlogram analysis revisited. Water Resources Research 7(6):1448-1459.
- Wallis, J. R., and N. C. Matalas. 1972. Sensitivity of reservoir design to the generating mechanism of inflows. Water Resources Research 8(3):634-641.
- Wallis, J. R., and P. E. O'Connell. 1972. Small sample estimation of λ . Water Resources Research 8(3):707-712.
- Yen, B. C., and A. H. A. Ang. 1971. Risk analysis in design of hydraulic projects. Proc. Symp. on Stochastic Hydraulics. University of Pittsburgh. pp. 694-709.
- Yevjevich, V. 1964. Fluctuations of wet and dry years, Part II. Analysis by serial correlation. Hydrology Paper No. 4, Colorado State University, Fort Collins, Colorado. 50 p.
- Yevjevich, V. M. 1965. The application of surplus, deficit and range in hydrology. Hydrology Paper No. 10. Colorado State University, Fort Collins, Colorado. 69p.
- Yevjevich, V. M. 1967. An objective approach to definitions and investigations of continental hydrologic droughts. Hydrology Paper No. 23, Colorado State University, Fort Collins, Colorado. 18 p.
- Yevjevich, V., and R. I. Jeng. 1969. Properties of non-homogeneous hydrologic series. Hydrology Paper No. 32. Colorado State University, Fort Collins, Colorado. 33 p.
- Young, G. K. 1968. Discussion of 'Mathematical assessment of synthetic hydrology' by N. C. Matalas. Water Resources Research 4(3):681-683.

APPENDIX A

LISTING OF MONTHLY STREAMFLOW DATA

Table A.1. Monthly streamflow data for Beaver River (ac. ft.).

Year	Month											
	O	N	D	J	F	M	A	M	J	J	A	S
1915	2550	1730	1630	1430	1260	1670	4740	10800	11400	4310	2260	1780
16	1700	1580	1390	1010	1510	2760	6250	13600	9100	4500	3390	2240
17	2480	1930	1190	1190	1470	1610	3140	6760	11800	3970	1830	1560
18	1520	1240	1270	1260	1080	1440	3180	8420	5880	3230	1920	1470
19	1730	1500	1400	1190	972	1380	5540	12400	3810	2150	1530	1450
20	1410	1410	1190	984	886	1010	1690	20000	12700	4390	2560	1710
21	1780	1510	1090	1250	1240	2150	2940	12900	18600	4890	2880	2180
22	1790	1560	1470	1280	1150	1440	2180	20800	17300	4630	2880	1940
23	1630	1520	1520	1470	1270	1510	3180	19200	10500	5200	2660	1810
24	1730	1510	1390	1360	1370	1450	4460	8300	3460	1680	1220	1080
25	1070	1050	1110	965	972	1440	3810	12700	5840	3890	2300	1740
26	1940	1690	1520	1290	1220	1730	5270	17200	6430	2670	1910	1490
27	1440	1280	1320	1320	1340	1500	3370	10900	5950	3310	2380	1910
28	1880	2000	1610	1650	1540	2180	3550	12900	6430	3280	2200	1670
29	1730	1550	1230	1230	1290	1750	2960	17800	8150	3810	2340	2130
30	1620	1490	1420	1230	1090	1430	5870	10000	6310	2560	2020	1650
31	1600	1210	1170	1110	1160	1320	2310	3820	2080	1260	953	732
32	1020	809	799	738	920	1150	3240	12900	8690	4460	1750	1240
33	1360	1110	1230	1540	1000	935	1300	5510	13000	3370	1890	1390
34	1380	1250	980	861	980	1410	2520	3000	1430	1140	867	662
35	825	857	922	984	1000	1190	2760	7260	13410	3660	1870	1370
36	1230	1180	1100	1210	1100	1300	5000	12230	12360	5660	3810	2170
37	1790	1460	1350	1230	1110	1460	4920	23640	10340	4780	2360	1680
38	1450	1380	1450	1120	1010	1420	4590	13010	10740	4200	1990	1460
39	1650	1320	1360	1320	1140	1730	3720	5680	2790	1730	1330	1340
40	1430	1100	1010	964	1060	1480	5050	14780	5360	2020	1550	1550
41	1470	1200	1170	1200	1120	1290	1560	20740	21220	6040	3090	2010
42	2110	1920	1710	1660	1380	1580	4800	15030	9820	3570	2190	1520
43	1500	1350	1260	1210	1300	1680	6970	8890	4820	2280	2200	1430
44	1550	1310	1220	1170	1110	1320	2310	14600	17010	5640	2380	1790
45	1580	1480	1380	1310	1200	1440	2470	12930	9610	5050	2610	1780
46	1690	1470	1450	1380	1240	1480	5220	7420	5010	1970	1710	1270
47	1950	1390	1320	1120	1110	1810	4290	18820	9210	4590	3240	1980
48	1830	1380	1400	1180	1090	1250	3030	12230	6540	2770	1760	1300
49	1250	1190	1200	1110	1030	1340	4480	12390	13450	4990	2250	1520
50	1820	1390	1270	1200	1210	1410	3030	5450	3650	1910	1170	954
51	978	1040	1010	891	825	1130	2080	5740	5280	2070	1650	1200
52	1160	1070	1030	922	952	1220	4970	22410	17970	6680	3220	2040
53	1610	1350	1330	1430	1210	1500	2010	3340	4940	1920	1560	988
54	1100	1040	956	950	940	1110	3400	6300	3760	1570	1270	1070
55	1020	948	922	889	807	1040	1850	5500	4570	2000	1490	1030
56	944	902	956	996	944	1320	2300	7000	5550	3010	1510	1000
57	1010	950	938	928	891	1080	2200	7080	22900	7380	3340	1650
58	1710	1410	1450	1240	1170	1310	3440	17150	12130	6450	2200	1860
59	1520	1380	1190	1160	1040	1270	2030	2590	1780	1040	990	801
60	895	799	821	893	789	1100	2430	5820	3000	1210	847	900
61	962	932	873	740	716	1020	2200	4870	2810	1080	1160	1240
62	1060	924	940	853	833	966	5070	7880	7030	4480	1970	1260
63	1250	1050	960	807	978	1030	1400	5710	3450	1350	1390	1070
64	948	954	867	803	713	835	1700	7330	5410	3450	1560	1050
65	952	944	940	906	799	916	1850	5480	9190	6000	3900	1930
66	1520	1230	1140	982	890	1220	2810	6530	2640	1260	944	893
67	936	865	1300	833	801	1090	1380	6200	10920	6640	3160	2080
68	1550	1210	1160	1010	984	1250	2250	10410	12430	5240	3260	1590
69	1480	1190	1130	1140	948	1170	4120	16440	10020	5730	3030	1820
70	1670	1340	1230	1190	1030	1200	1560	7750	8800	5030	2320	1670
71	1400	1280	1230	1070	1120	1270	2720	6220	6360	3440	1670	1320
72	1270	1140	1110	958	895	1470	1490	2400	1830	972	773	847
73	1190	1040	946	904	789	950	2070	15480	14460	7680	2890	1360
74	1270	1100	1140	1140	1060	1600	2470	8600	5950	1770	1080	940
75	934	1000	924	845	827	954	1110	5210	9430	4360	1760	1230
76	1140	962	990	918	873	958	1490	5320	2490	1210	908	764
77	823	738	612	613	635	791	1960	1580	1660	916	723	635
78	817	694	623	710	736	1190	2990	9670	15800	4360	3760	1620

Table A.2. Monthly streamflow data for Blacksmith Fork River (ac. ft.).

Year	Month											
	O	N	D	J	F	M	A	M	J	J	A	S
1914	6150	5470	4890	5270	4530	7620	21100	23700	12600	9590	8850	8210
15	6700	6070	6520	5740	5200	5650	6130	6210	5800	5040	4350	4550
16	4970	4760	4370	4210	4260	10800	27000	22800	11500	9350	7930	7380
17	6950	5820	5480	5080	4450	5010	13600	41900	25900	14000	11100	9280
18	8480	7500	7010	6270	5780	9840	12800	16300	8930	8300	7560	6250
19	6090	5480	5310	4880	4520	6030	9940	10200	6660	5230	4890	4750
20	4850	4350	4130	4270	3930	4730	9580	35200	13900	9410	8180	7320
21	7010	6600	6150	5710	5510	11500	17400	41900	20300	12200	10500	9280
22	8670	7620	7320	6890	5830	7440	14000	37900	16400	10600	9780	8450
23	7870	6960	6950	6640	5180	5850	14900	38800	14900	11400	9900	8330
24	8480	7260	6580	6270	5580	5820	12700	14700	7680	6640	5870	5500
25	5580	5040	5070	4440	4060	5960	13400	15700	8930	7070	5510	5650
26	5490	4880	5070	4840	3880	5010	10200	7870	5380	5140	4960	4410
27	4460	4390	4250	4210	3870	4940	12500	20200	10100	6950	6460	5740
28	5710	5390	5080	4630	4490	7320	12000	19700	7740	7010	6400	5510
29	5600	5330	4610	4300	3920	5490	10400	20100	9820	7010	6210	5730
30	5550	5210	5010	3900	4220	5020	8570	7870	5860	5320	5440	4670
31	4540	3900	3650	3580	3050	3680	3800	4480	3500	3070	3090	2860
32	2970	3000	3070	3010	2990	4670	17600	31900	12000	8360	7260	5730
33	5590	4960	4350	4050	3530	4270	7260	13800	10300	6460	5380	4720
34	4590	4250	4030	3730	3260	3820	3840	3610	2970	2780	2710	2570
35	2670	2740	2770	2660	2590	3230	7600	10500	6250	4280	3720	3290
36	3360	3310	3140	3100	2960	4330	24170	34810	10940	7860	6590	5620
37	5470	5220	4680	4250	3720	5060	9590	21910	9470	6920	5730	4960
38	5060	4770	4670	4310	3800	6260	15670	16270	8910	7020	6140	5400
39	5340	5070	4900	4280	3600	6140	9050	7530	5230	4370	3920	3650
40	3720	3510	3500	3490	3220	4140	5520	5440	3770	3280	3000	2860
41	3020	2880	2910	2840	2740	3160	3860	5130	3430	2970	2820	2590
42	2870	2760	2830	2730	2420	2910	6750	6170	4990	4090	3420	3160
43	3200	3270	3370	3530	3380	6030	20690	14480	8730	6900	6100	5130
44	5060	4540	4330	4080	3530	3860	5530	9190	6530	5340	4630	4030
45	4240	4110	3710	3580	3700	4330	5570	11810	12310	7060	6180	5320
46	5240	5060	5060	5000	4090	8130	33680	22600	12270	9240	7960	6690
47	6890	6160	5740	5160	4700	6480	9360	12540	7280	6330	5740	5120
48	5270	4790	4600	4280	4100	4850	12680	26870	12050	8500	7300	6170
49	6090	5720	5250	4890	4690	7090	16850	16500	9710	8140	7220	6320
50	6430	5540	5330	5970	5410	7710	22760	30760	18770	11770	10100	8930
51	8320	7690	7460	6460	6710	7050	24010	26060	13990	10920	9810	8390
52	8020	6950	6630	6050	5330	6060	24350	37160	16410	12110	10230	8520
53	8150	7150	6670	6510	5510	6430	8740	11260	11040	7470	6400	5630
54	5490	5210	5170	5050	4400	5630	10280	8560	5870	5160	4720	4290
55	4240	4010	3910	4070	3610	4030	7720	15140	7440	5710	5160	4590
56	4660	4330	6830	7080	5380	7600	20380	19290	10770	8580	7340	6300
57	6100	5650	5460	5040	5380	6690	10490	20580	13140	9100	7730	6760
58	6810	6070	5980	5480	5320	6200	11120	20760	9240	7500	6720	5880
59	6050	5610	5260	5020	4390	5650	8650	7890	5850	5070	4900	4570
60	4740	4290	4060	4390	3800	6180	10160	9680	6260	5110	4450	3890
61	4000	3900	4050	3740	3350	3630	4360	4610	3500	3130	2870	2800
62	3240	3050	3040	3000	4500	4130	21930	15820	8350	6520	5880	5210
63	4890	4460	4160	3810	5190	4720	7090	10900	6060	4770	4260	3840
64	3970	3910	3540	3580	3090	3110	8000	17930	9520	6650	5620	4830
65	4430	4150	5530	4880	5640	5670	19960	24400	13620	9590	8220	7500
66	6930	6210	5590	4980	4430	6690	12810	11200	7120	6170	5330	4860
67	4920	4510	4550	4430	3870	5750	9850	24630	15030	9240	7520	6730
68	6810	5780	5660	5230	5060	6110	7960	13880	9390	7040	6520	5930
69	6130	5550	5340	5160	4720	5640	17160	16260	9160	7760	6680	5790
70	5990	5320	5250	5160	4600	5200	6210	15170	9030	6920	6200	5540
71	5480	5310	5740	7100	6850	9590	26550	37730	21060	13800	11950	10310
72	9820	8300	7970	7980	7020	16730	27450	33430	18020	13580	11610	9600
73	9150	7840	7490	6870	5650	6910	10740	19640	10410	8500	7500	6880
74	6200	5800	5790	5760	5050	10530	18660	27350	14370	10810	8950	7510
75	7310	6600	6350	5900	4680	6660	8590	25140	18140	10870	9050	7360
76	7830	6640	6080	5830	5520	6830	15240	20800	9730	7960	6820	5730
77	5800	4990	4910	4570	3990	4220	4390	4690	4060	3460	3130	2980
78	3380	3510	3650	3460	3310	7460	16690	17680	9220	6640	5780	5370

Table A.3. Monthly streamflow data for Logan River (ac. ft.).

Year	Month											
	O	N	D	J	F	M	A	M	J	J	A	S
1901	13530	11908	10950	10520	9280	10330	15230	57860	41090	22190	17830	14690
2	13340	11480	11060	7690	7770	8610	11900	56360	57910	19450	15130	12380
3	12300	11070	11060	10660	9590	10760	15030	31820	54680	28060	18940	15000
4	11070	10890	8600	5200	9200	18750	22020	56140	57180	34010	21210	16320
5	13890	11430	10760	9650	8110	9770	11420	24170	33970	19860	15430	11780
6	10300	8870	8370	7470	6260	6850	16400	47200	54000	29300	16500	10800
7	8610	7450	8670	7720	9400	15900	39600	72400	96500	58100	29400	19800
8	16700	13900	12000	10900	9030	9040	14900	24500	44100	25000	14500	9520
9	9530	7320	6280	8880	5970	7390	24700	63600	98300	50500	28900	15100
10	10900	12100	9220	7870	5400	17900	41300	73800	47200	24100	18100	13500
11	12800	10100	8980	8300	9340	10700	22800	63500	77100	32100	16900	12100
12	11300	9100	7070	6950	7310	8790	16000	54600	94900	46500	19600	14900
13	11800	9520	8710	8210	7420	9800	21400	37800	28500	18100	13400	11100
14	10200	8840	8040	7730	6680	9250	23000	57300	48400	23800	15300	12600
15	11600	9620	7930	6970	6200	7160	15100	20200	21100	12500	9650	8440
16	8130	7150	6470	6320	6290	12000	25800	46100	48400	29600	16500	12200
17	11300	9230	8520	7290	6160	6850	11900	37200	59900	40500	20200	14900
18	12400	9740	9760	8100	7020	10900	17300	34900	41100	18700	13600	10600
19	10200	8640	8030	7020	6340	7450	12700	39000	24100	12700	9830	8540
20	8030	6990	6670	6440	6180	6690	10900	53200	60400	26300	16000	12600
21	11000	9920	8510	7810	6960	10800	19600	58800	74900	35900	20700	15300
22	13200	10400	10500	9060	7710	9440	13600	51100	60300	27000	17800	14500
23	11800	9760	9230	8900	6830	7360	16200	59900	50600	27800	17600	13800
24	12600	10900	9360	8280	7490	7380	19400	41500	21400	13800	10600	8840
25	8740	7200	6670	6190	5660	7210	15800	42100	30500	18600	12600	10200
26	8700	7690	7180	6590	5880	7960	17700	28700	17100	11600	9060	7720
27	7300	6620	6220	6060	5380	6720	13900	35800	47600	25000	14900	11500
28	10600	10200	8750	8070	6860	9680	13900	54100	35200	19100	13400	10600
29	9870	8400	7530	7230	6060	7460	11200	37400	40300	22200	13700	11100
30	9990	8410	7670	7020	6150	6780	16200	25200	25100	13900	10500	8480
31	7920	6570	6420	6030	5100	5740	7640	16800	11700	7400	6260	5260
32	4920	4570	4720	4670	4250	5700	15200	53900	57500	30600	16600	12500
33	10800	8960	7700	7280	5860	6520	10100	24400	48500	18600	12500	9740
34	8390	7130	6660	5980	5260	6870	12790	13040	8370	6340	5420	4750
35	4560	4460	4470	4360	4240	5000	11200	29470	37510	16910	10980	8280
36	7060	6150	5480	5130	4940	5820	24360	72950	51560	23720	15590	11770
37	10430	8640	7720	6740	5870	6520	10010	38410	31260	17900	12290	9490
38	8780	7700	7200	6730	5700	7330	19650	43450	40540	19620	13570	10240
39	9030	7930	7400	6650	5760	7840	16890	28660	18260	11990	9240	7460
40	7030	5970	5860	5630	5240	6420	9940	26550	15980	10150	7800	6540
41	6060	5300	4810	4590	4350	5280	7070	21050	15740	9800	7350	5850
42	5780	5210	5100	4870	4290	4940	13390	20230	25430	13790	9550	7420
43	6520	5910	5510	5420	5150	7030	29360	45600	46440	26090	15050	11330
44	10160	8280	7320	6560	5630	5720	8760	26660	27360	15280	10670	8170
45	7450	6500	5720	5580	5140	5580	7280	27920	38920	23580	14290	11020
46	9580	8510	7790	7290	6040	8910	33360	48000	39200	21850	14730	11310
47	10430	8870	8130	7050	6400	8450	12770	43590	30270	17320	12450	9760
48	8890	7630	7070	6510	5860	6000	13430	46880	46150	21040	13980	10860
49	9860	8480	7700	6990	6020	7670	18610	42900	33130	18350	13020	10840
50	10480	8060	7450	7660	6670	8610	21820	49110	68780	40020	19790	14330
51	12560	11040	9880	8570	8090	8800	27260	51750	46150	24790	16140	12380
52	11560	9720	8710	7900	6740	7080	20040	53190	45960	22500	15130	11800
53	10580	8980	8280	7830	6450	7450	11660	21640	41880	21980	13500	10170
54	9160	8140	7260	6830	5760	6800	12530	28260	17430	12140	8800	7090
55	6640	6000	5460	5320	4610	5090	8040	29680	28500	15090	10220	7780
56	7370	6490	8860	8010	6300	8200	22630	49230	41860	18920	12560	9620
57	8870	7640	7130	6530	5990	7230	11020	33790	49690	23570	13720	10410
58	9740	8260	7630	6730	6210	6990	12800	47220	38940	17630	12370	9810
59	8580	7690	7060	6270	5560	6510	12600	25700	29940	14910	10420	8400
60	8450	7030	6360	5910	5290	7510	16150	30120	22580	12600	9330	7770
61	7140	6580	5930	5580	4880	5430	7360	18040	14510	8160	6330	5440
62	5350	4900	4780	4580	5700	5650	25910	41400	32520	18020	11920	10160
63	9080	7640	6620	5820	6890	6460	9530	31960	29060	14340	10030	8490
64	7600	6710	6000	5510	4970	5200	9800	30110	39160	22080	12930	9930
65	8600	7480	8100	7550	7540	7510	17600	40560	61800	32370	17520	13560
66	11660	9650	8510	7610	6380	8590	18030	33250	19860	12500	9540	7900
67	7420	6740	6370	6320	5220	6520	9570	33280	52620	28350	15420	11730
68	10300	8750	7690	6960	6690	7720	10460	28060	39810	20830	13990	10740
69	9720	8600	7960	7670	6430	7120	19840	44360	28320	18450	12110	9660
70	8850	7330	6820	6680	5710	6230	8370	33530	48600	22450	13360	9980
71	8800	8600	8170	9150	8520	10240	25280	53480	80650	43750	22350	16190
72	13930	11380	9870	9390	7990	16350	24500	55320	63270	30390	18850	14300
73	12560	10630	9340	8220	6720	7220	9400	33670	30390	16980	11940	9500
74	8240	7610	7050	6710	5740	8710	16970	50690	58070	27030	16600	12850
75	10350	8490	7490	6930	5860	6820	7780	28490	59290	44750	19650	13640
76	11450	9480	8730	7730	7030	7960	15260	43710	35700	20410	13580	11120
77	9600	7980	7000	6520	5420	5780	7780	10140	9670	6770	5740	4870
78	4150	4460	4730	4450	4170	7810	18370	35600	48960	24260	11110	9130

Table A.4. Monthly streamflow data for Weber River (ac. ft.).

Year	Month											
	O	N	D	J	F	M	A	M	J	J	A	S
1905	4190	4220	3690	3380	2940	3950	7380	22900	48100	9590	4330	3500
6	4490	3560	2790	2640	2220	3020	10800	46700	68400	25800	9410	6900
7	4600	3770	3970	4060	3720	5610	20300	48900	93700	91400	13900	6310
8	5260	4260	4560	3570	3000	3210	9520	25500	59300	24500	9840	5490
9	7870	5220	4270	4060	3220	3570	8870	45800	130000	40500	14300	9580
10	7010	6660	5840	5230	4160	11100	30600	58500	36200	8670	4970	4280
11	4940	4060	3580	4610	4720	4320	11200	42100	78000	17600	5590	3890
12	5240	4550	4610	4000	4310	4220	5930	34800	95000	22400	10100	7140
13	8120	7260	4790	4300	3890	5170	13600	51100	40500	14300	7500	7680
14	8670	6430	4980	5300	4490	5810	17200	78700	76200	19600	8550	5440
15	7380	5230	4610	4920	4780	5050	13400	27000	46500	11900	4610	5150
16	5060	5430	5570	4000	3980	6270	17900	44800	78000	17200	7620	4740
17	7560	5690	6080	3010	3330	4360	8350	33300	103000	47100	8360	5500
18	4770	4460	4060	4060	3440	4850	7970	32300	66600	9040	4610	3720
19	6270	3810	4000	3070	3130	3710	9760	48600	24800	6010	3980	4060
20	5620	4520	3430	3460	3570	4000	5410	59100	79100	16800	7380	5080
21	6460	6430	4610	4300	4160	6700	13200	62100	127000	28600	9840	6960
22	6100	5350	4770	4300	3690	4240	7440	54600	106000	17300	8850	5330
23	4580	4550	4430	4430	4230	5140	8390	55800	69000	27700	8180	5600
24	6700	4620	3820	3690	3820	4590	7680	43400	19300	5900	3840	3140
25	3420	3420	3440	3140	3040	4510	12400	49100	34000	11500	5940	5580
26	6520	4550	4190	3070	3430	4930	15800	47800	27600	7320	4930	3310
27	3590	3310	3390	3380	3080	3600	8510	44500	66600	14600	6820	5120
28	5770	5900	4160	3690	3160	5120	8690	71300	39900	11000	5700	3790
29	3970	3920	3690	3380	3330	3900	6660	46400	71400	18600	8550	6490
30	4600	3760	3690	3070	3050	4300	14900	32100	42200	9530	7620	4910
31	6520	3390	3070	3380	3050	3060	5590	25200	15900	4250	2670	1990
32	2430	2830	3070	3070	3160	3740	9580	54400	75000	15600	5980	3780
33	3460	3140	2870	2770	2500	2820	4960	19400	78600	9720	4800	2800
34	2810	2620	2580	2640	2380	3480	10550	17540	4880	2560	2110	1960
35	2230	2560	2770	2770	2500	3070	7550	25090	74060	11280	5680	3350
36	3110	3080	3070	3070	2880	3380	16530	75600	40550	12660	9320	5880
37	4250	3870	3380	3070	2780	3380	8850	56840	24430	8750	5170	3690
38	4390	4460	3690	3380	3050	3720	13710	43490	53280	11170	6100	4630
39	4740	4830	4390	3690	3330	4590	13500	37200	18710	7010	4150	3110
40	3440	2920	2440	2340	2300	3230	8090	39080	13750	5380	3270	3050
41	4610	3400	3200	3070	2610	3420	5450	36530	38830	9630	6300	4210
42	4850	4610	3900	3480	3110	3680	16050	32060	55050	11080	5780	4030
43	3480	3250	3150	2880	2580	3550	20800	42700	46080	13780	6000	4730
44	4520	4100	3350	2980	2780	3100	4920	44960	62900	16110	6840	4640
45	4200	4040	3670	3530	3200	3720	5060	33840	41730	15300	8420	5990
46	4490	4840	3990	3650	3090	4600	22670	42660	36960	9420	5610	4280
47	4460	4180	4040	3380	3190	4280	9390	53580	44000	14030	6920	5340
48	4630	4220	4090	3940	3570	4290	8100	53590	37180	9420	5640	3180
49	3240	3390	3440	3440	3000	3830	14590	44010	54800	13320	6640	4260
50	5080	4410	3908	3570	3250	3660	13620	45970	77580	23640	8330	6480
51	5100	5070	4580	4170	3960	4080	12700	43880	58020	18120	8920	5220
52	6600	4600	3930	3490	2880	3440	15470	69160	73080	19100	9330	6220
53	5350	3790	3070	3180	2850	3740	7420	18710	65280	13400	7720	3920
54	3490	3290	3070	3130	2980	3070	10750	37930	17060	8320	4530	3740
55	3150	3000	2770	2770	2500	2460	5020	40690	33750	8690	5650	3900
56	3550	3410	4430	4160	2990	4560	14870	55640	54270	10270	5990	4170
57	3760	3270	2460	2570	2710	3220	5290	29860	78690	25430	8630	5470
58	4500	3720	3380	3100	3240	3520	6240	48250	35480	7660	4880	3520
59	3070	3220	3200	2950	2480	3270	7530	22880	42500	9860	4660	3560
60	6120	4480	3090	2820	2460	4010	11260	32430	32090	7560	3640	2670
61	2910	2980	2600	2880	2300	2890	4650	21710	12360	4260	2950	4960
62	5040	3930	3070	2770	3240	3570	17620	39490	61110	18230	6600	3660
63	3700	3000	2900	2770	3070	2920	5490	35650	42230	10320	5220	4310
64	3710	3460	3010	2610	2010	2270	4600	44120	56450	19190	7310	4690
65	3620	3440	3590	3200	3030	2770	8660	40260	78850	38040	13040	9220
66	6560	5150	4630	4110	3260	4720	14920	48180	24880	8350	5000	4090
67	5070	4190	3730	3500	2930	3910	5740	35390	75860	34510	9840	5300
68	4880	4220	3870	3400	3110	3810	6030	32200	77820	19880	9920	6800
69	5490	4580	4220	3800	3290	3790	15510	73500	42200	15600	7170	4400
70	4250	3830	3380	3540	3120	3330	4440	37690	57100	14940	6670	4640
71	4120	4220	4440	4900	4000	4440	14260	37690	72120	18690	7560	5370
72	4910	4650	4170	4110	3670	6670	13450	47160	60150	10990	6140	4510
73	5860	5050	4620	3940	3300	3560	5410	44670	48730	14140	8010	6660
74	5550	4500	3310	3070	2820	5470	9610	39720	58660	12820	6520	4110
75	4080	3650	3440	3080	2780	3320	3820	25390	77450	52790	9990	5960
76	5600	5110	4550	4180	3480	4140	10350	44980	34610	9840	5620	3690
77	3560	2960	2890	2300	2060	2210	4970	10940	16650	4890	3470	2800
78	3170	2240	1770	2380	2150	3410	12110	33910	74610	15770	6070	4380