University of Washington
Department of Civil and Environmental Engineering

FORECAST UNCERTAINTY IN RESERVOIR OPERATION: A CASE STUDY

Nabeel R. Mishalani
Richard N. Palmer

Water Resources Series
Technical Report No.107
August 1987

Seattle, Washington
98195

Department of Civil Engineering
University of Washington
Seattle, Washington 98195

FORECAST UNCERTAINTY IN RESERVOIR OPERATION: A CASE STUDY

Nabeel R. Mishalani
Richard N. Palmer

Water Resources Series
Technical Report No. 107

August 1987

University of Washington
Department of Civil Engineering
Environmental Engineering and Sciences Program
Seattle, WA 98195

FORECAST UNCERTAINTY IN RESERVOIR OPERATION: A CASE STUDY

By

Nabeel R. Mishalani

Richard N. Palmer

Water Resources Series

Technical Report No. 107

August 1987

Abstract

This research investigates the benefits of forecasting in water resources systems. Questions relating economic losses caused by water supply shortages to forecast period and accuracy are addressed. Some simple available forecasting techniques are assessed for their accuracy and applicability. These issues are addressed through the use of a simulation model of the Cedar and South Fork Tolt rivers. This system is represented as a single-purpose reservoir supplying municipal and industrial water to the Seattle metropolitan area. The following conclusions were made: 1) Economic losses caused by water deficits always increase with an increase in forecast uncertainty; 2) The optimal length of forecasting period is five months; 3) Up to 88% of the economic losses incurred by operating the reservoir without forecasting are due to the lack of forecasts; 4) The mean of the historic data is not recommended to predict future flows because Markov methods are always superior; and 5) Lag-one autoregressive Markov schemes exhibit about a 9% improvement over no forecasting.

ACKNOWLEDGMENTS

This study was supervised by Dr. Richard N. Palmer, Associate Professor of Civil Engineering. Mr. Nabeel R. Mishalani, Graduate Research Assistant in the Department of Civil Engineering, developed computer code and was responsible for the analysis of the computer results. The study served as the basis for Mr. Mishalani's MSCE thesis. The authors wish to acknowledge the valuable comments and suggestions made by Dr. Ronald E. Nece and Dr. Wen-Sen Chu.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

Forecasting plays an important role in our lives. Every individual makes innumerable predictions daily. These predictions, or forecasts, range widely in their importance, their accuracy, and the community they affect. In a five minute meeting between two business people, hundreds of predictions are made by each person as to the other person's reaction to a thought. Physicians, having experience with a drug, predict a patient's reaction to that drug. Financial brokers depend on forecasts to make decisions about buying and selling stocks and commodities. Other types of forecasts may affect more people and have graver consequences. Long-term earthquake forecasts may affect the migration into and out of a geographic area, while a short-term forecast may save thousands of lives. Highly uncertain weather forecasts affect thousands of people daily.

Increasing populations, higher water supply demands, and competing water uses have necessitated the incorporation of optimization techniques into water resources problems. The complexities of a multipurpose, multi-reservoir system have generally required release decisions to be determined by optimization or simulation models. The analysis of a complex water resources system involves models with thousands of decision variables and constraints. Once the objectives and constraints have been determined, most problems lend themselves to solution techniques developed in the fields of operations research and management sciences [Yeh, 1985].

The introduction of optimization methods into water resources problems has encouraged the investment of much research and effort into the subject of 'value of information'. The collection of data implies a future use of that data and therefore an inherent worth. Typically, the first data to be collected contain the most information and thus have the greatest value. As more data are collected, the information content, and hence the worth per additional unit of information decreases. On the other hand, the marginal cost of obtaining those data usually remains constant. Because of the decreasing marginal worth of the data, it should be possible to define a level of information at which the marginal worth and cost for data are equal. In principle, defining that optimal level of information is thus conceptually simple. In practice, however, assigning an economic worth for data has proven to be a very difficult task. Therefore, much effort has been devoted to the determination of surrogates of worth. This research employs the economic framework to evaluate the benefits of forecasting.

In order to devise an operating policy, optimization and simulation models require some form of forecast. Forecasts are made with the hope that they will enable a decision maker to take advantage of certain future events and, at the same time, reduce losses resulting from others. A forecaster employs a forecasting technique with the available data base to make a "best guess" at future events. The benefits of a forecast are measured by the degree to which losses resulting from the lack of such a forecast can be reduced. Economic losses resulting from water supply shortages are incurred when restrictions on water uses are initiated. Because forecasts are made to increase the net benefits of an activity, forecasts can be said to have a worth. This can be related to an increase in informational content of the historic record. Obviously, in some cases when forecasts are not accurate, or when a sufficiently long historic record does not exist, forecasts may lead the decision maker into making an inferior decision. Such a forecast can be thought of as supplying misinformation to the decision maker.

Water supply forecasts are essential to the operator of a water resources system. In order to satisfy competing water uses such as municipal and industrial water demand, irrigation, flood control, fish flow requirements, and recreation, an operator must make decisions concerning the apportionment of water releases among purposes, reservoirs, and time periods. Therefore, high quality forecasts become a necessity to the operator.

Numerous forecasting techniques with differing complexities and realisms are available to the operator. The operator is then faced with several questions. To what extent do the available data describe the system at hand? Which forecasting technique is best suited for that system? Would forecasts be of any value in the system operation? How accurate are these forecasting schemes? What length of forecasting period should be adopted? To what extent does this affect the operation and the resulting economic losses?

The purpose of this report is to study the relationship between economic losses caused by water supply shortages, forecast uncertainty, and length of operation period. Additionally, some simple forecasting techniques are employed to test their applicability and accuracy.

The above topics are addressed through the use of a computer model describing the water resources system supplying the Seattle metropolitan area. The model simulates the operation of a single reservoir with a single inflow and a single purpose, namely municipal and industrial (M&I) water supply for the City of Seattle. Forty-seven years of inflow data are used for the estimation of population statistics and for the reservoir operation.

The contents of this report are organized as follows: Chapter 2 is a review of worth of data literature and streamflow generating techniques. Chapter 3 contains a brief description of the Cedar and South Fork Tolt Rivers drainage basins. It also includes esimates of the water demand for the Seattle metropolitan area and the demand load factors used in this study. Chapter 4 describes the operating policy, the methodology and logic of the model, and the experimental design. Chapter 5 summarizes the findings of this report and indicates potential areas of further inquiry.

Chapter 2

LITERATURE REVIEW

With the development of optimization theory and methods, and their application to water resources problems, the subject of the value of information has become more and more important. How much data are really needed to make a decision? Is the available information enough, or should a decision be postponed until more data have been collected? In this case, what are the economic costs for waiting? Because of the relatively short historic records at most sites of interest, worth of data studies have generally employed synthetic streamflow generating techniques to provide a longer time series of flow data for use in simulation of reservoir operation. This chapter is devoted to the review of worth of data studies and the generating techniques that they use.

2.1 WORTH OF DATA LITERATURE :

As more information is available, the planner, or operator, is more likely to make a better decision. Hence, information can be said to have value. The worth of a specific set of data is dependent on the use or uses intended for those data. For a certain application, the worth of data may be defined in terms of the losses that would have been incurred if such data were lacking. Practically, such an economic evaluation of the worth of data has proven to be very difficult. Therefore, much attention has been devoted to the determination of surrogates to worth, usually some measure of accuracy and sample variability [Dawdy, 1979].

Studies of the worth of data have generally assumed three basic approaches. The first estimates the "true value" of a parameter from long streamflow records or stochastic streamflow traces. Worth of data is then measured in terms of the deviations in the value of those parameters from their "true value" when they are estimated using smaller sample sizes. The second approach attempts to evaluate the worth of data in an economic framework. The third uses Bayesian statistical techniques to update a subjective prior distribution of the parameter to be estimated.

2.1.1 Sample Variability as a Surrogate to Worth :

Dawdy [1979] traced the history of the search for surrogates of worth. The literature review classified network design studies according to four approaches : 1) the Information-Variance Approach, 2) the Transfer Function-Variance Approach, 3) the Economic Framework, and

4) Decision Theory and Bayesian Analysis. Papers had shown that the full Bayesian analysis could be circumvented by the use of simulation to overcome some of the computationally more difficult portions of the analysis. It was concluded that simplicity was required if there is to be any impact on decision makers, and that techniques that reduce the dimensions of the problem without reducing the physical realism of the solution must be developed.

Moss and Karlinger [1974] used the first approach in conjunction with the Bayesian theorem in the design of stream gaging networks. The study investigates the statistical characteristics of the accuracy of regression analyses as used in surface water regionalization by simulating logarithmic regressions of the streamflow parameters. According to the authors, generalized relationships for the economic evaluation of the benefits and costs of information had not yet been developed. Thus, a measure of regional information was estimated from the standard error of estimate of the regression analysis. This measure is expressed in terms of equivalent years of at-site record. Prior to this study, only a single observation of apparent equivalent years was available to define the level of regional information concerning a streamflow parameter. To improve upon this, the statistical nature of equivalent years of record must be considered. Hence, it is suggested that the network design criterion be expressed as a statement of probability, specifying the probability of attaining a desired level of information.

Tschannerl [1971] set forth two methods for computing the expected opportunity loss (EOL) for a single purpose reservoir. The first method is based on sampling from a synthetic flow sequence generated on the basis of estimates of the true values of the streamflow statistics. The other method designs a decision rule within the framework of subjective probabilities (or Bayesian statistics). EOL is defined as the loss resulting from overdesigning or underdesigning due to imperfect information. The study demonstrates the computation of EOL as a function of sample size. This represents a measure of the worth of data, and can be used by the planner to establish a cutoff point for data collection. It was concluded that the subjective approach allowed adjustments in design to account for uncertainty by incorporating the EOL into the design decision.

2.1.2 Economic Framework :

The major disadvantage of the first approach is the inadequacy of most hydrologic data; that is, streamflow records are often short or do not exist at points of interest. Jettmar and Young [1975] attempted to determine the consequences of using Markovian (simple) and self-similar (complex) synthetic data generators in the economic design of a multi-purpose reservoir. The lag-one Markovian model used the first-order serial correlation coefficient of the historical series as the explanatory parameter. The self-similar process used the Hurst coefficient as the driving parameter. Statistical analyses showed that, for the Markovian model, the short range fit was perfect, the intermediate poor, and the high lag range good. On the other hand, for the self-similar process, the short range fit was poor, the intermediate good, and the high lag range poor. Hence, the statistical analysis of the historical data was judged to be inconclusive with respect to discrimination between a Markovian or a self-similar generating process. To distinguish between the two models, a much greater time horizon was needed (T>10,000 [Chi et al. 1973; Mandelbrot, 1971; Wallis and Matalas, 1971]). Therefore, the economic framework approach was used. The economic factors considered were (1) reservoir size, construction, and cost; (2) water supply, target output, and penalty function; (3) storage and its relationship to recreation; and (4) avoidance of flooding and flood losses. It was concluded that the approach of viewing economic results to judge the significance of hydrologic assumptions was sound and capable of answering very basic issues. Given limited resources, a Markovian model can be used for both sizing and operating a multi-purpose reservoir. Furthermore, no discrimination between a simple Markovian model and a complex self-similar model could be found.

The second approach taken by worth of data studies evaluates the value of information in an economic setting. This approach uses a long streamflow record or stochastic streamflow traces to develop the "optimal" plan, and evaluates the economic impacts on this plan if shorter records, taken arbitrarily from the long record, are used in reservoir design. This approach permits the researcher to assign a value for information increments. Dawdy et al. [1970] assess the effect of sampling error on the value of data used for reservoir design. In the study, streamflow for Arroyo Seco near Soledad, California, are used. The value of data is assessed by measuring the net benefits foregone as a result of the lack of data. Optimum reservoir design is based on a 500-year stochastic sequence of flows generated from statistics of the historical flow. Records of various lengths are arbitrarily chosen from the 500-year base record and used to determine the apparently optimal design. A capacity/cost function is assumed and benefits are computed for flood control and water supply. The expected value of net benefits increase as more data are collected. Also, the results demonstrated a decrease in the added value per additional year as more years of data were collected.

Moss [1970] uses the data from Dawdy et al. [1970] with existing discharge and stage measurements to illustrate and test the applicability of a scheme for determining the optimum operating procedure at the gaging station on Arroyo Seco near Soledad, California. The optimum length of record was satisfactorily defined. For a situation where no prior streamflow data exist and the reservoir is to be built as soon as enough data are collected, it was found that the optimal length of record is nine years. On the other hand, the optimum frequency of discharge measurement was found to be indeterminate. Sensitivity analysis showed that the optimum length of record was more sensitive to the variability of the streamflow and the planned level of development at the site than to discount rate or cost of obtaining streamflow data.

It should be noted that extension techniques were not used to add to the value of the collected data.

Klemes [1977] also uses the economic approach to examine the value of information in optimization of storage reservoir operation. The range of aspects examined in relation to the value of information in reservoir operation includes the implications of using expected values as an optimality criterion, the relevance of the knowledge of population parameters, the value of real-time forecasting, the relation between the economic and hydrologic uncertainties, and the influence of reservoir size. The hypothetical problem is the annually updated operation of a multi-purpose single storage reservoir fed with a random input having a log normal probability density. An objective function of the form of a quadratic loss function is assumed. The policy is derived by minimizing the expected present value of loss using the standard backward explicit stochastic dynamic programming algorithm in conjunction with an assumed distribution fitted to a 'historic' input series. Numerical experiments demonstrate a negligible sensitivity of the results to the assumed form of input distribution, a near optimality of the policy - 'draft equals input mean' - for any reservoir size and any size of input sample, and a similarity between the effects of hydrologic and economic uncertainties. Finally, it was found that the degree of optimality of a given policy did not depend on the degree to which the underlying input model and its parameters agreed with the input population, but rather on the similarity of the features of the historic sample underlying a given policy and the future 10-year sample on which the policy performance was tested.

Yeh et al. [1982] assess the improved benefits that might be gained from the use of long-range streamflow forecasts in the operation of a multi-purpose reservoir. The Oroville-Thermalito reservoir system of the California State Water Project was selected for the study. The study relates incremental operational benefits to streamflow prediction period and accuracy. Applicable benefits are considered one at a time, independant of any others. No optimal combination of benefits was attempted since such optimality was highly subjective. In this study, the major potential benefits are increased hydropower generation, water conservation, and decreased seepage damage to crops during their critical growing periods. It is concluded that estimation of the future inflows as the historical averages provides significant benefits.

2.1.3 Bayesian Analysis :

The third common approach to the evaluation of the expected worth of additional data is through the use of Bayesian decision theory. In this approach, a subjective prior probability distribution of the quantity to be estimated is postulated. As more data are collected, the prior distribution is updated using conditional probabilities [Benjamin and Cornell, 1970] and becomes what is commonly referred to as the posterior density. For each additional sample, the resulting posterior distribution is used to compute the expected worth of the additional data. This approach is used by Lenton et al. [1974] to improve at-site estimates of the annual lag-one autocorrelation of streamflow series where the prior distribution was derived from regional information.

Krzysztofowicz [1983] notes that uncertainty is often ignored in the process of decision making and that conventional decision procedures utilize forecasts as if they were error free. He addresses the question of the value of an uncertain forecast when the decision procedure ignores the uncertainty, and investigates several fundamental aspects of this question by means of a simple decision model. In contrast with other studies which assumed optimal decision making, the economic value of a forecast is treated as being dependent on both the accuracy of the forecast and the degree of optimality of the decision. An analytic solution to a quadratic decision problem is extended in order to evaluate nonoptimal decision procedures. Two Bayesian information processors, one for categorical forecasts and the other for probabilistic forecasts, are formulated and applied to records of daily temperature forecasts. The probabilistic forecasts are of the type in which the forecaster quantifies his degree of uncertainty in terms of a fixed-probability central credible interval. It is demonstrated that probabilistic forecasts are likely to be more valuable than categorical forecasts, and that the relative gains from probabilistic forecasts are likely to be greater if suboptimal procedures are employed.

Krzysztofowicz and Watada [1986a] formulate a stochastic model to describe the uncertainties in the categorical forecasts of runoff volumes during the snowmelt season issued by the National Weather Service and the Soil Conservation Service. The model is a discrete-time, finite, continuous-space, nonstationary Markov process. The categorical forecasts consist of the "most probable" runoff x_n, expected to occur if precipitation subsequent to the date of forecast is median; the "reasonable minimum" runoff $\underline{x}_n$, expected to occur if precipitation subsequent to the date of forecast is equal to the 0.1th fractile; and the "reasonable maximum" runoff $\overline{x}_n$, expected to occur if precipitation subsequent to the date of forecast is equal to the 0.9th fractile. A categorical forecast is defined as a point predictor x_n, whose value depends upon the specific hydrometeorologic conditions in a given year. The above described forecast was termed as categorical because it specified a point predictor and a "naive" central credible interval $t_n = \overline{x}_n - \underline{x}_n$ (i.e., t_n was based only on a prior climatological distribution of precipitation and therefore did not reflect the uncertainty of the specific hydrometeorologic conditions in a given year). A probabilistic forecast is defined as one having both the point predictor x_n and the credible interval t_n depend upon the specific hydrometeorologic conditions in a given year. Thus, probabilistic forecasts were preferred to categorical ones in that the format of the forecast (that users have grown accustomed to) would not change, the credible interval would have attached to it an explicit probability (that can be set to any desired level), and that the information contained would completely specify the posterior density which would provide information for optimal decision making.

Krzysztofowicz [1986c] considers the decision maker's dilemma of when to make a commitment concerning water supply when such a commitment has to be made within a period prior to the forthcoming season. The more the commitment is delayed the more accurate is the forecast (on the average). Unfortunately, as the commitment is delayed, the potential benefits decrease. An optimal decision model for a class of commitment problems is presented. The model can serve purposes of real-time decision making, economic evaluation of forecasts, and analyses of improvements of the forecast system. The planning problem is formulated as a finite horizon, nonstationary Markovian stopping process with Bayesian updating of the distribution of the uncertain runoff, maximization of the expected utility of outcomes as the planning criterion, and dynamic programming as the solution algorithm. Results of numerical experiments show that even highly uncertain forecasts may be valuable provided they are employed in optimal decision models. Also, it is shown that substantial economic gains could be accrued from improvements in seasonal snowmelt runoff volume forecasts, particularly the early ones.

2.2 REVIEW OF FORECASTING TECHNIQUES :

The best way to answer questions about future flow patterns for a stream would be through the use of a complete theoretical model to determine the future course of the basin. Unfortunately, such models are simply not available. In the absence of such a complete model, hydrologists have attempted to devise models that provide possible flow traces of the stream under study. Such models are the core of synthetic or operational hydrology [Jackson, 1975]. Synthetic hydrology is a tool that may be used in evaluating the consequences in water supply management decisions [Hirsch, 1979]. According to Clarke [1977], synthetic hydrology may be defined as the application of Monte Carlo methods to estimation problems in hydrology. By using Monte Carlo techniques to generate long or multiple streamflow records and then analyzing the outputs of a water resource system with these records as inputs, one can test the system's response to periods that are longer and contain greater extremes than the historical record [Askew et al., 1971]. One application of synthetic hydrology is to the problem of forecasting. The following is a brief overview of some of the major families of models used in synthetic hydrology.

Thomas and Fiering devised a method for generating long traces of synthetic flows that are indistinguishable, within sampling variation, from the historical flow sequence. The Thomas-Fiering or lag-one Markov generation scheme can provide traces drawn from a population with a specified mean, variance, and first-order serial correlation coefficient [Jackson, 1975]. Such a synthetic streamflow generation scheme consists of two components : a deterministic part, d_i, and a random part, ε_i [Fiering & Jackson, 1971]. Because of the serial dependence of flows, it is necessary to include a non-zero deterministic component to reflect persistence in the generation process. The random component is assumed to be independently distributed with mean zero and constant variance. It is also assumed that the degree of persistence between successive flows does not depend on the level of those flows (obviously, this assumption is not completely realistic). Thus, the deterministic component will have the linear autoregressive form :

$$d_i = B_0 + B_1 q_{t-1} + B_2 q_{t-2} + \cdots + B_m q_{t-m}$$

where d_i is a linear combination of m previous flows.

The simplest model of this form is the Markovian, or lag-one, flow model which assumes that the entire influence of the past on the current flow is reflected in the previous flow value. The generating equation

$$q_i = \mu + \rho * (q_{i-1} - \mu) + t_i * \sigma * (1 - \rho^2)^{1/2}$$

where the t_i are independent normal sampling deviates with mean 0 and standard deviation 1, gives normally distributed synthetic flows that preserve the mean μ, standard deviation σ, and first order correlation coefficient, ρ, of the historical flow. This Markovian flow scheme assumes that a given flow depends on the preceding flow and a random component and on these factors alone.

Since the Markov models do not explain the Hurst phenomenon, Fiering continued on to consider the possibility that flow values depended not only on the immediately preceding flows but also on a longer portion of history of the flow sequence [Jackson, 1975]. Multilag generation schemes are used when the dependency on exactly one previous flow is not justifiable. Such schemes attempt to model the storage of water in underground aquifers from season to season, and the contribution of a fraction of it each season to form part of the total runoff. According to Fiering and Jackson [1971], this is "a model with a long memory." The deterministic part of multilag models would be of the form :

$$d_i = B_0 + B_1 q_{i-1} + \dots + B_m q_{i-m}$$

with m > 1. The solution to the problem of how many previous flows to include in the model is not trivial. Fiering and Jackson [1971] consider traditional statistical tests but reject them because of the non-circularity of historic flows. They adopt an operational procedure that includes lags as long as it is practical and profitable to do so. The procedure calls for measuring the goodness of fit (R^2) of a linear model for each level of serial dependence. A practical stopping rule is to stop including lags when the R^2 values reach a plateau.

Hirsch [1979] attempts to evaluate the performance of six different, single-site, monthly generators. He bases his method of evaluation on the synthetic streamflow generator's (SSG) ability to estimate the reliability of a water supply system. Other methods have been demonstrated by Askew et al. [1971] and Jettmar and Young [1975]. The six SSG's are applied to the case of the Potomac River at Point Rocks, Maryland, where an 81-year historic record was available. The models use three ways to describe the marginal probability distribution of the flows, and two ways to describe the mechanism of serial dependence (hence the six SSG's). The ability of one type of generator to produce water supply system reliabilities that are better supported by the relevant historical data than those produced by other SSG's is considered to be a demonstration of superiority of the former generator. Results show that preservation of statistical moments such as the mean, standard deviation, and the lag correlation coefficients, may be a misleading criterion for judging the ability of an

SSG to provide plausible estimates of water supply system performance. It is demonstrated that it is operationally superior to base the specification of marginal distributions of monthly streamflows on transformed values of historical streamflow data rather than on the data itself. Finally, a new type of autoregressive-moving average model is shown to be operationally superior to an autoregressive model.

Several families of models have been proposed to model long-term persistence, which is not properly accounted for by the Markov models [Burges and Lettenmaier, 1977]. The most extensively documented is the Fractional Gaussian Noise model of Mandelbrot and Wallis [1969]. In a series of papers, Mandelbrot and van Ness [1968] and Mandelbrot and Wallis [1968, 1969a,b,c,d,e] describe a family of models, the Fractional Gaussian Noise models, that do not belong to the Brownian domain of attraction [Jackson, 1975]. These models involve a complex correlation structure in which the current value of the process depends on the entire history of the process. Mandelbrot [1971] suggested a combination of fractional noise and Markov models, in fact, an approximation of fractional noise as a sum of Markov processes. The resulting process is referred to as a fast fractional noise process.

Another family of hydrologic models, the broken line model, has been suggested by Rodriguez-Iturbe et al. [1972], Mejia et al. [1972], and Garcia et al. [1972]. These authors suggest that a number of different simple broken line processes be added together to form a more complicated general broken line process. These models can be made to preserve a prescribed value of the second derivative of the correlation function at the origin. According to Jackson [1975], it is not clear what the second derivative at the origin means for a discrete parameter process.

2.3 Summary and Conclusions :

The purpose of this study is to investigate the relationship between economic costs and forecast uncertainty. It employs some of the simple available forecasting schemes to answer certain fundamental questions. How good are our forecasts? By how much can they be improved? What effect does the length of the forecasting period have on the quality of our forecasts?

In order to answer these questions, two sets of numerical experiments, having differing approaches and objectives, are conducted. The first produces a theoretical relationship between economic costs as a function of forecast period and accuracy. This is done by using, for our forecasts, the actual historical flow, perturbed by a noise term, and associating it with the resulting economic costs. The second set of experiments employs differing available generating techniques to make a best guess at future flows. It should be noted that the purpose of the generating techniques used in this study was not to augment existing flow records. Instead, they are used to test their potential use as forecasting schemes.

Chapter 3

WATERSHED DESCRIPTION

3.1 EXISTING WATER SUPPLY

The Seattle Water Department (SWD) supplies water to the metropolitan Seattle area from its existing Cedar and Tolt River sources. Both watersheds are located on the western slopes of the Cascade Mountains (Figure 1). The Cedar/Tolt Water Supply System represents a complex, multi-purpose reservoir system. In addition to providing municipal and industrial water supply to the Seattle metropolitan area, the Cedar reservoir assists in maintaining the salmon population by controlling floods in the winter and maintaining summer and fall flows greater than those that would naturally occur, provides high quality water to Lake Washington, and is a site for hydropower generation.

The Cedar/Tolt system includes reservoirs on the Cedar and South Fork Tolt Rivers, and diversion sites on the Cedar at Landsburg and on the South Fork of the Tolt River downstream of the dam. According to Erickson et al. [1982], active storage in the two reservoirs is sufficient for approximately two months of municipal and industrial (M&I) demand during the summer months at the current demand

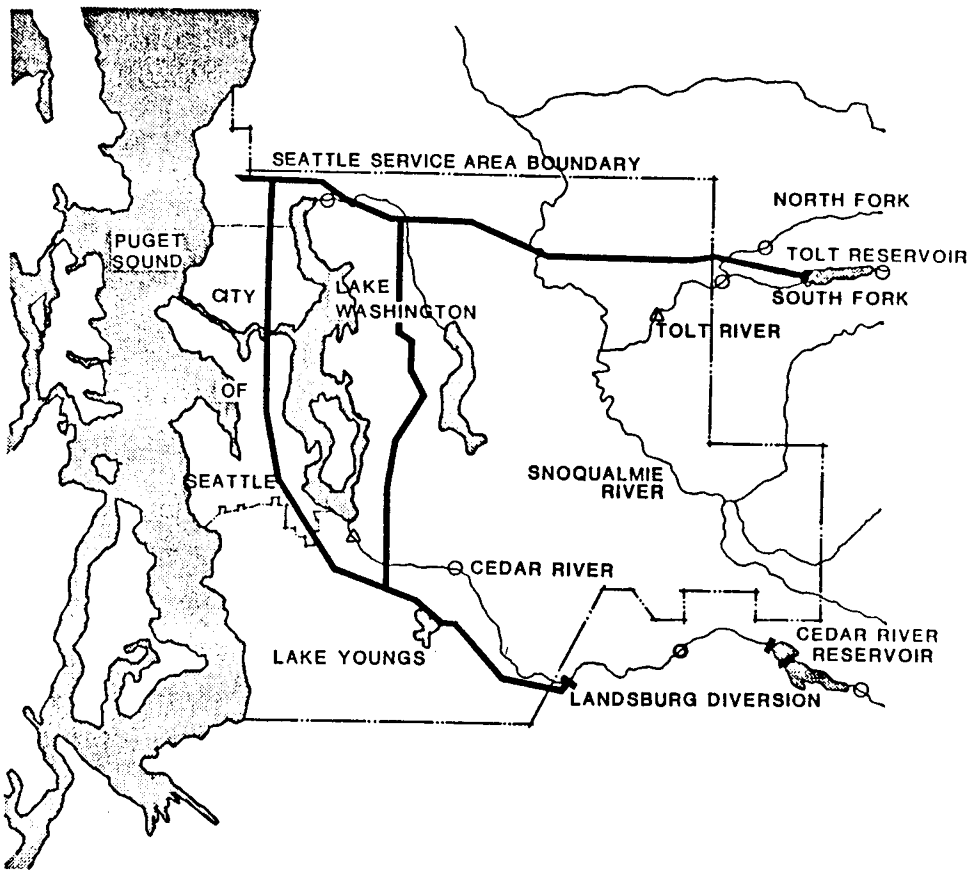


Figure 1

Water Supply System Location

level if all demands are satisfied. Using a reliability standard of 98 percent, SWD's current system yield is estimated at 169 mgd. This is obtained by adding the yields of the Cedar and Tolt river sources, where the yield for each source is calculated independently of the other. This yield, however, is affected by competing water uses and several water quality issues in the system's two watersheds.

Monthly varying flood storage requirements alter the available active storage capacity during the year. Table 1 summarizes the monthly variation of active capacities. This information includes SWD's latest update of June 1987.

3.1.1 Cedar System

The Cedar River Watershed is located in Southeast King County about 25 miles from downtown Seattle (Figure 2). The crest of the Cascade Mountains marks the eastern border of the watershed. The western edge is at the Landsburg water supply intake, approximately 24 miles downstream of the beginning of the Cedar River. The Cedar source supplies over 70 percent of the current SWD demand.

The Cedar River Watershed administrative boundary encompasses 90,495 acres, of which the City of Seattle owns

TABLE 1

Active Reservoir Storage Capacities (acre-feet)

	CEDAR RIVER RESERVOIRS							S. FORK TOLT RESERVOIR			
	Chester Morse Lake			Masonry Pool			Total Cedar				COMBINED ACTIVE
MONTH	Upper !	Lower°°	Active	Upper	Lower	Active	Capacity	Upper	Lower	Active	CAPACITY
April	55506	36064	19442	16040	415	15625	35067	57900	26500	31400	66467
May	55506	36064	19442	23376	415	22961	42403	57900	26500	31400	73803
June	55506	36064	19442	29059	415	28644	48086	57900	26500	31400	79486
July	55506	36064	19442	29059	415	28644	48086	57900	26500	31400	79486
August	55506	36064	19442	29059	415	28644	48086	57900	26500	31400	79486
September	55506	36064	19442	29059	415	28644	48086	57900	26500	31400	79486
October	55506	36064	19442	16040	415	15625	35067	54900	26500	28400	63467
November	55506	36064	19442	3892	415	3477	22919	47100	26500	20600	43519
December	55506	36064	19442	3892	415	3477	22919	47100	26500	20600	43519
January	55506	36064	19442	3892	415	3477	22919	47100	26500	20600	43519
February	55506	36064	19442	3892	415	3477	22919	47100	26500	20600	43519
March	55506	36064	19442	3892	415	3477	22919	54900	26500	28400	51319

* All figures are in acre-feet.

! Upper = Upper bound on reservoir storage
°° Lower = Lower bound on reservoir storage

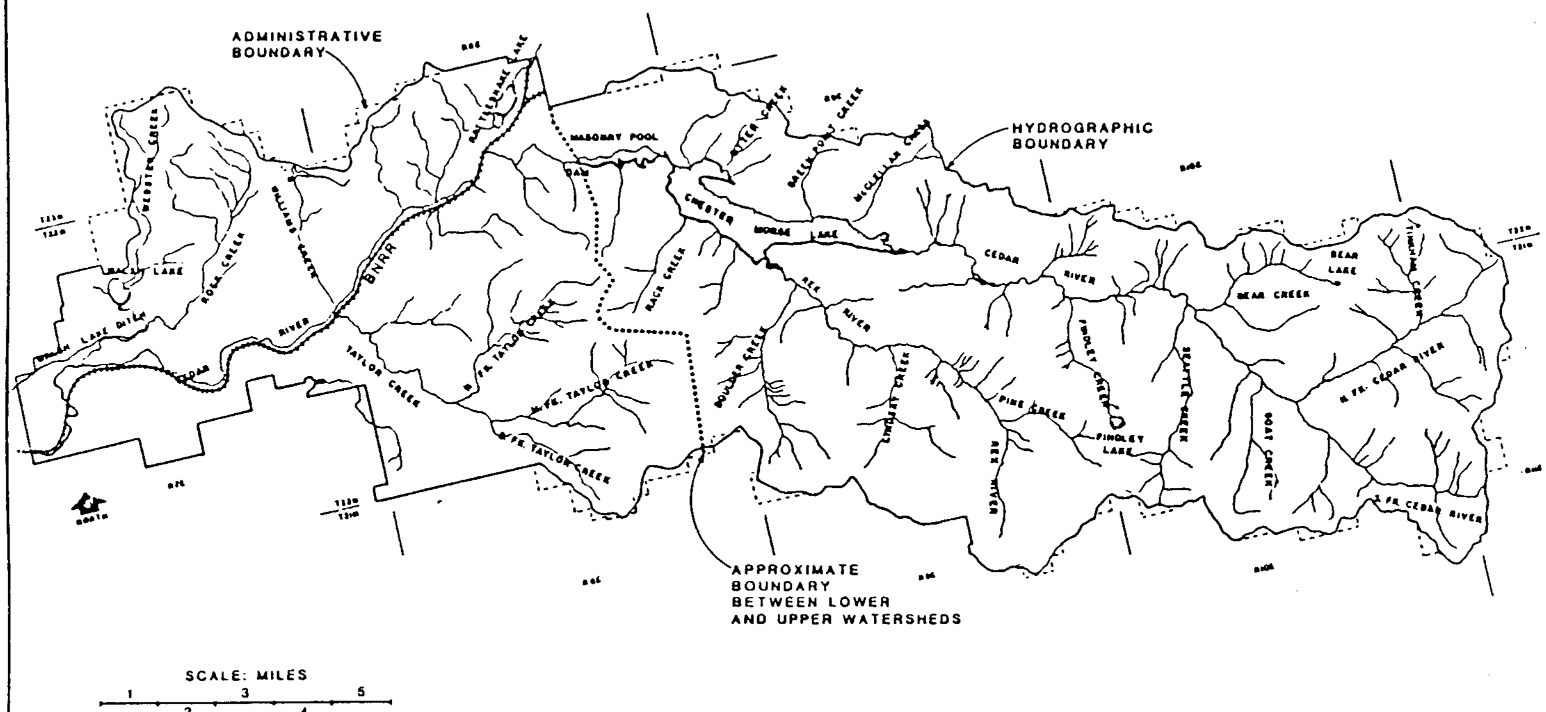


Figure 2
Cedar River Watershed

80.6 percent. Since the beginning of the century, the City has obtained total fee ownership of all lands within the administrative boundary of the Cedar Watershed.

The Cedar reservoir system is composed of two dams (Figure 3). Upstream is a low dam constructed of timber, known as the crib dam, that controls the elevation of Chester Morse Lake. The total storage capacity of Chester Morse Lake is 55,000 acre-feet (ac-ft), of which only about 19,000 ac-ft are useable because of the large volume of dead storage. Approximately 1.4 miles downstream is a higher masonry dam creating what is called the Masonry Pool. The total capacity of the reservoir system is 154,000 ac-ft. Severe seepage losses from the masonry pool, however, occur when the water depth exceeds 30 feet, limiting the total storage capacity to 74,000 ac-ft [Erickson et al., 1982]. Water lost to seepage from the reservoir flows into the Cedar River Morrain Aquifer. Losses from the Masonry Pool are estimated at 2300 to 23,000 ac-ft/month depending on the hydrostatic pressure in the reservoir as well as on the water storage in the aquifer.

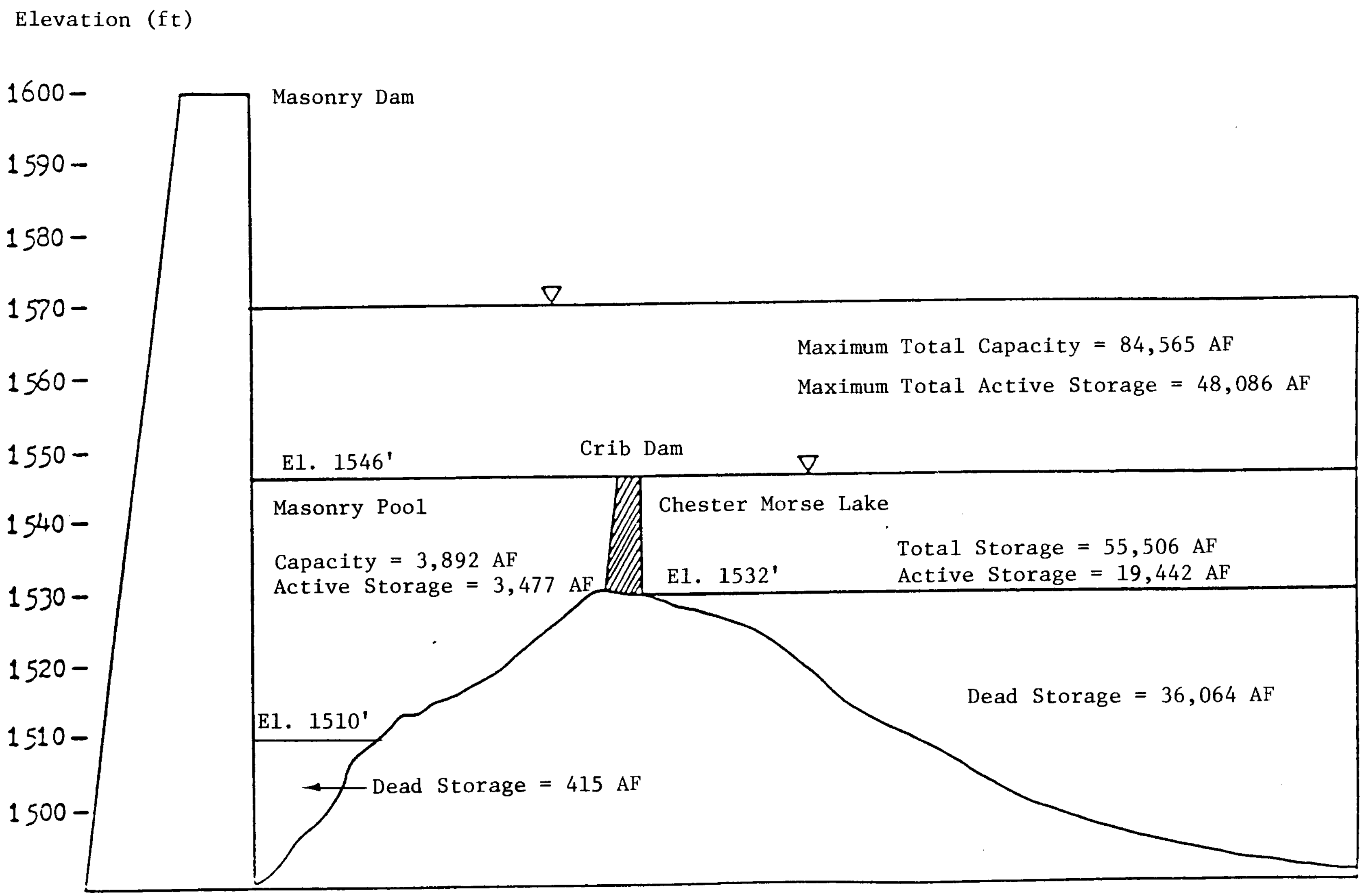


Figure 3

Cedar Reservoir: Schematic Diagram

3.1.2 Tolt System

Approximately 20 miles north of the Cedar River Watershed in northeastern King County are the two watersheds of the Tolt River Basin (Figure 4). Presently, only the South Fork Tolt Watershed is used for municipal water supply. The average current supply from the South Fork Tolt is 52 mgd. This is based on a limiting reservoir drawdown to 1730 feet to aid in minimizing turbidity problems.

In contrast to its actions on the Cedar system, the City of Seattle does not have a program for obtaining total fee ownership of either the North or South Fork Tolt Watersheds. The South Fork Tolt Watershed has a drainage area of about 13,390 acres of which the City owns about 28.6 percent. The City also owns 12.6 percent of the 25,500 acre North Fork Tolt Watershed.

The Tolt system drains into the 58,000 ac-ft earth fill South Fork Tolt Reservoir at an elevation of 1765 feet. The water is diverted into the Tolt Regulating Basin (elevation 760 feet) and then into the supply system via the Tolt River Pipeline No.1, which has a transmission capacity of about 110 mgd (SWD, COMPLAN 1985). Average annual supply from the South Fork Tolt is currently

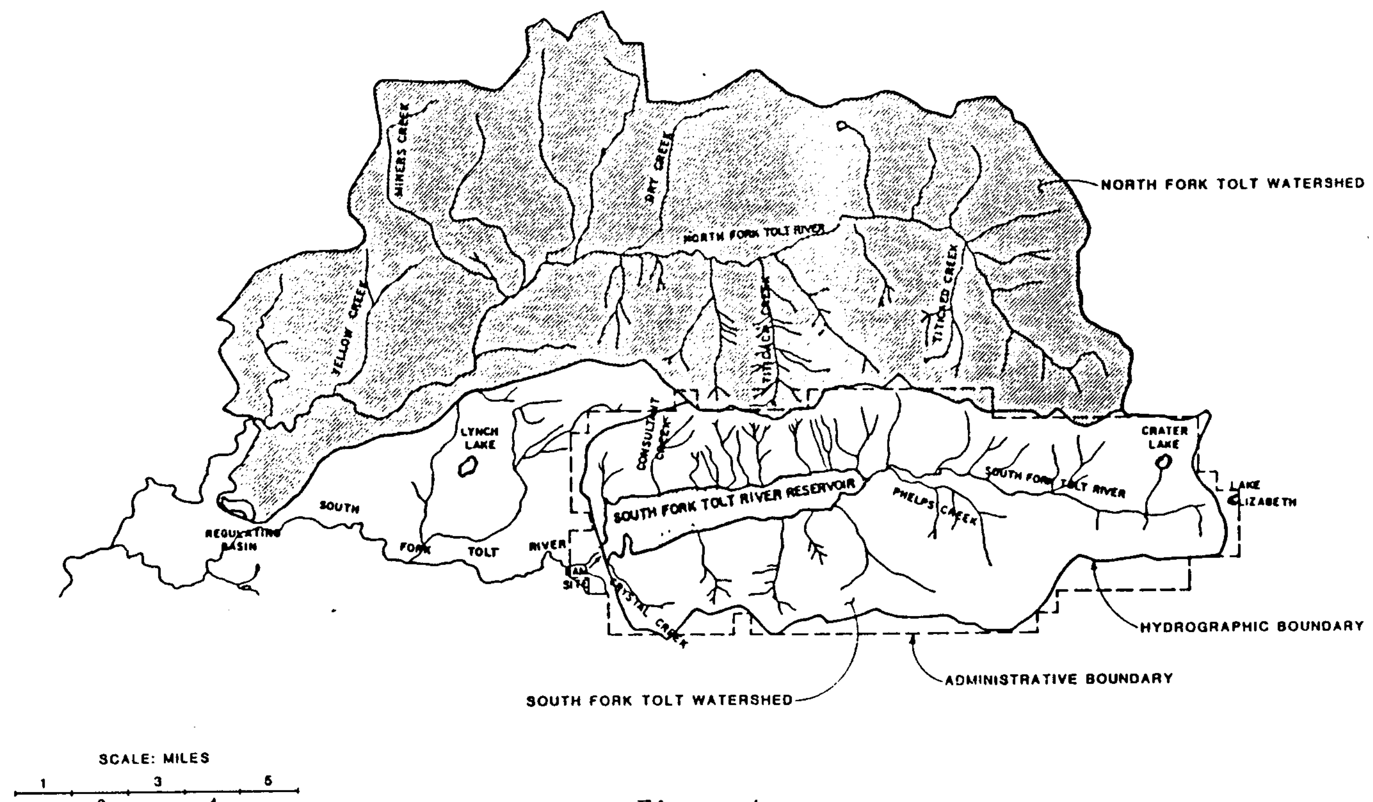


Figure 4

Tolt River Watershed

developed at 52 mgd. SWD has a water rights permit on the South Fork Tolt River for a total of 150 mgd.

TABLE 2

Summary of Existing Water Supply

	Cedar	South Fork Tolt
Watershed		
Area (acres)	90,495	13,390
Seattle Ownership (%)	80.6	28.6
Water Supply		
Year of First Delivery	1901	1964
Firm Supply (mgd)	117	52
Reservoir Inflow (cfs)		
Mean Annual Flow	6,001	2,202
Maximum Annual Flow	10,544	3,497
Minimum Annual Flow	3,417	1,385

3.2 WATER DEMAND

The water demand incorporated in reservoir operation experiments for this study is based on forecasts made by SWD and presented in the 1985 Seattle Comprehensive Regional Water Plan (SWD, COMPLAN 1985). SWD developed a recommended long-term demand forecast as a tool for preparing a sound management plan for meeting the future needs. The recommended demand curve (Figure 5) is a product of a forecast model and a series of underlying assumptions related to planning area, base year, time frame, population projections, estimated future rates, and estimated price elasticities. When all the variables were

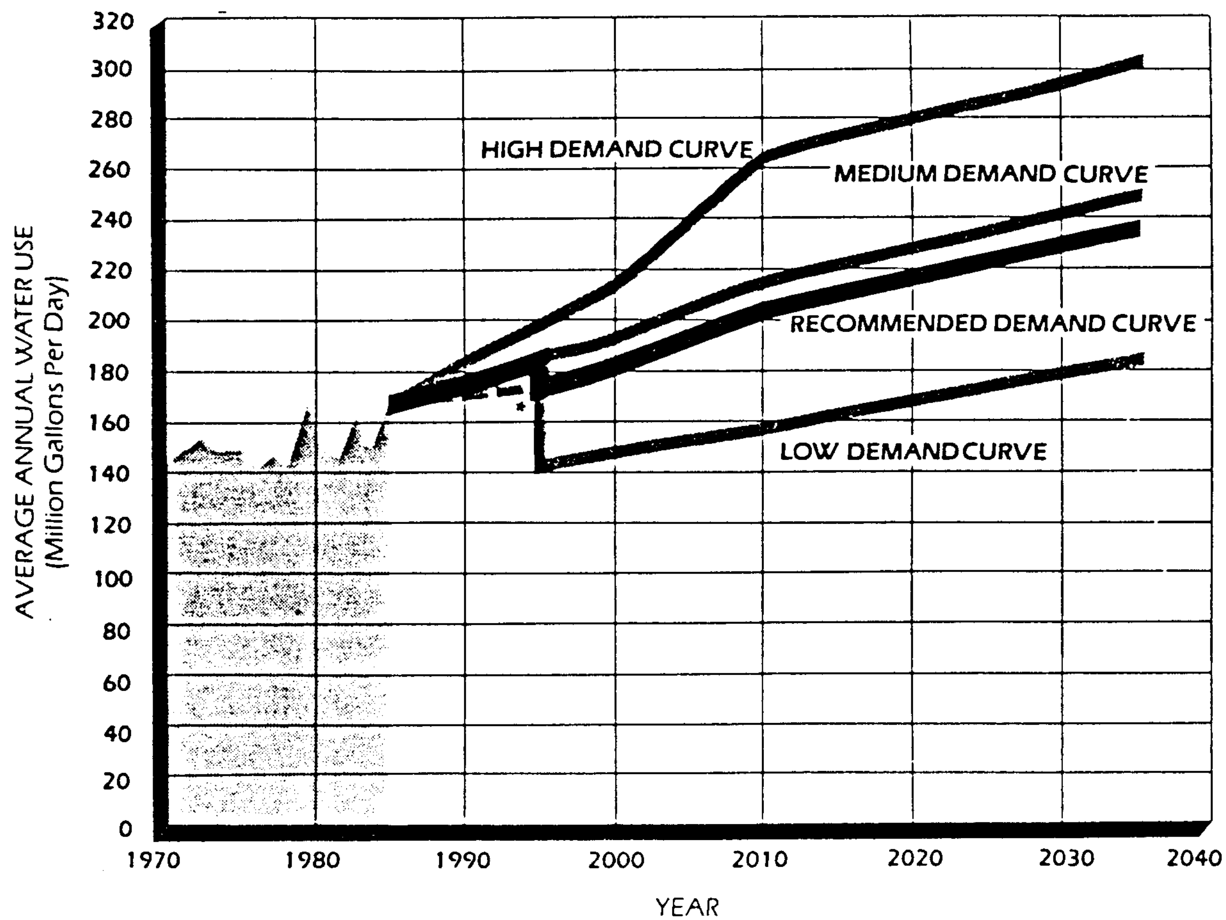


Figure 5
Water Demand Forecasts

included, SWD developed estimates for high, medium, and low demand scenarios through 2035 (Figure 5). The recommended demand curve depicts the most probable scenario during the planning period. It also assumes that the effects of the Long-Term Conservation Plan will be felt by 1995.

The choice of base demand to be used in the numerical experiments depended on several factors. The recommended demand forecasts set forth by the SWD represent M&I water demand of the Seattle metropolitan area. In view of the system's multi-purpose function, using the recommended curve to represent the total demand would yield unrealistically high system reliabilities. In order to account for the other purposes of the system and, at the same time, to keep the model as simple as possible, it was judged justifiable to increase the level of the recommended base demand. Therefore, a value of 235 MGD (22,000 ac-ft/month) for the base demand, corresponding to the high demand scenario for the year 2005, is used in the operation of the reservoir system. This choice was seen to increase the credibility of the study without adding any complexity.

Actual monthly demand levels used in the reservoir operation algorithm are obtained by multiplying the base demand by the corresponding demand load factors (Table 3).

Conclusion:

This chapter described the general layout of the reservoir system. The research employs a computer simulation model that reduces the system to a single-purpose reservoir. The following chapter is devoted to the description of the model and experiments.

TABLE 3

Monthly M&I Demand Factors

Month	Demand Factor	Demand MGD	Demand ac-ft/month
APR	0.85	200	18,700
MAY	1.00	235	22,000
JUN	1.20	282	26,400
JUL	1.45	341	31,900
AUG	1.40	329	30,800
SEP	1.00	235	22,000
OCT	0.85	200	18,700
NOV	0.85	200	18,700
DEC	0.85	200	18,700
JAN	0.85	200	18,700
FEB	0.85	200	18,700
MAR	0.85	200	18,700

Chapter 4

MODEL DESCRIPTION AND EXPERIMENTAL DESIGN

This chapter describes the computer simulation model employed and the experiments conducted in this study. The following questions are addressed: 1) How are economic costs from water supply shortages related to forecast accuracy? 2) Does increasing the length of the operation period reduce losses, and can this be related to an increase in forecast accuracy? 3) What are the potential benefits from increasing forecast accuracy? 4) How good are the available forecasting schemes, and to what extent can they be improved?

The simulation model is a simple conceptualization of the multi-purpose, multi-site Cedar/Tolt reservoir system, in which the system is reduced to a single-purpose reservoir. Such an idealization is judged appropriate for the purposes of the study. The operating policy is first described, followed by the methodology and the logic of the model, and finally by a description of the experiments.

4.1 THE OPERATING POLICY

Proper management of a water-resources system should mobilize the latent usefulness of natural water resources while allaying their potential destructiveness. According to Bower et al. [1962],

> " . . . an operating procedure is a set of rules for storing and releasing water from surface and ground-water reservoirs in a given water-resources system."

In multi-purpose, multi-site systems, these rules are the result of three major types of decisions concerning the release and storage of water : apportionment among reservoirs, among purposes, and among time periods. Although the system simulated in this study consists of two reservoirs, the author's time limits necessitate the simplicity of the model; hence, the system is modeled as a single reservoir with a single combined inflow. For the objectives of this study, water supply is assumed to be the sole purpose of this simplified water resources system.

Since the problem is that of a single-site, single-purpose reservoir, the only decision concerning the storage and release of water is that of apportionment over time. Ideally, it would be most desirable to release the exact demand at all times, and store whatever is in excess of the demand for later use during periods of low flow. In order to prevent deficits, however, it is required to have perfect forsight of all possible droughts at the design stage of the water resources system. Obviously, this is not possible. Reservoirs are designed with the knowledge and acceptance of a certain degree of risk of failure. Accepting the system's inability to prevent water shortages, it is most desirable to minimize economic losses due to such deficits. Foreknowledge of reservoir inflows over the operating period allows such an apportionment of releases that minimizes losses due to water supply shortages.

It is sometimes economical to accept a small current deficit so as to decrease the probability of a more severe water shortage later in the drawdown-refill cycle. The resulting hedging procedure is introduced early in the cycle, with small shortages accepted early and excessively large ones thereby avoided.

Economically, this can be justified only if the proposed uses of water have convex (non-linear) loss functions; that is, if severe droughts are penalized proportionately more than mild deficits. The Seattle Water Department has developed a series of curves to represent the economic costs caused by a deficit (Figure 6). The loss function used in this study has a parabolic shape and is given by :

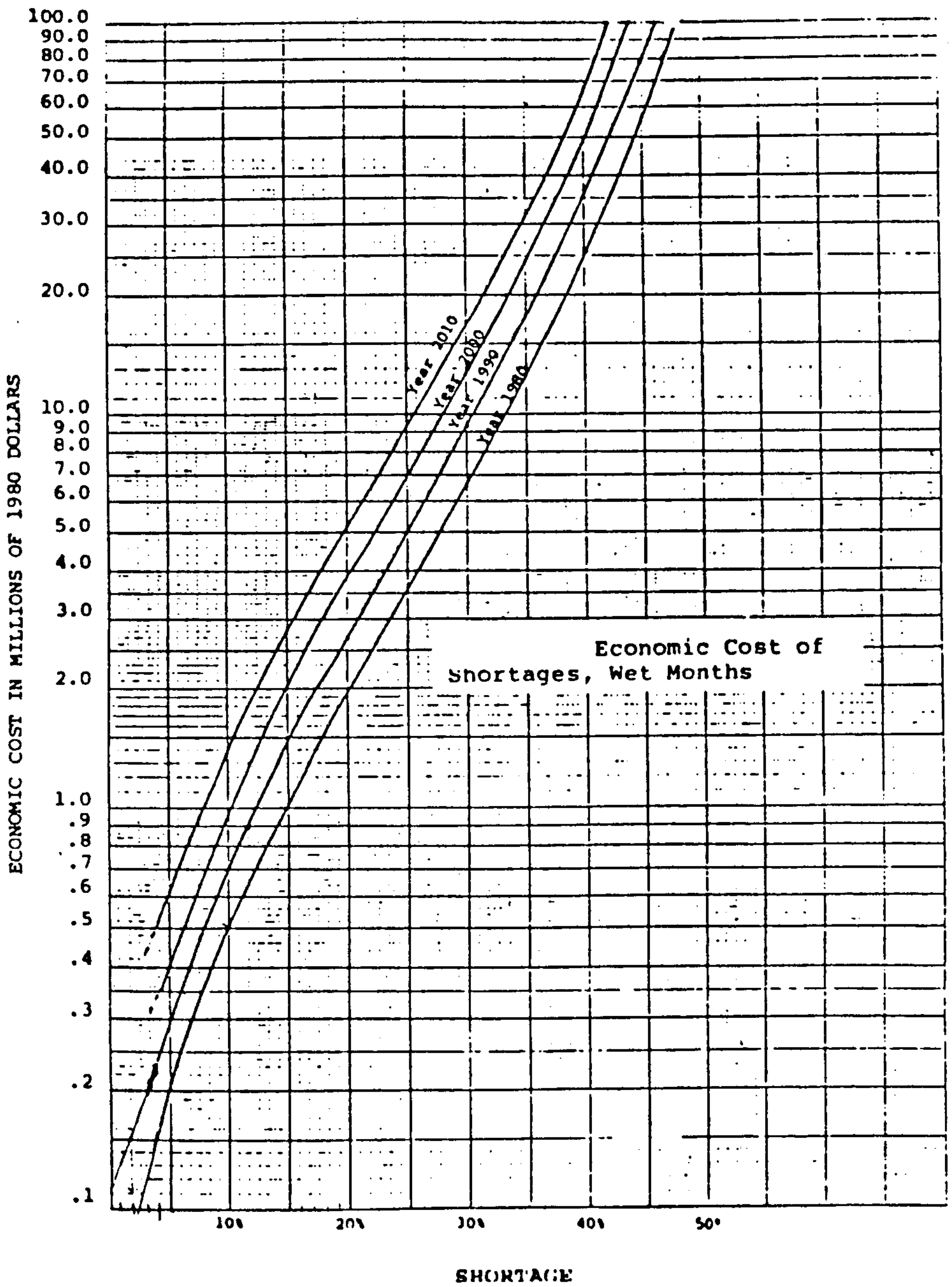


Figure 6. Economic Cost of Percent Shortage

Source: Applied Economic Associates Inc., 1983.

$$LOSS (\$) = 10,000 * (SHORTAGE/DEMAND*100)^2$$

It is not the purpose of this study to debate the appropriate loss function for the City of Seattle. The loss function used justifies the use of hedging by complying with the non-linearity requirement. Furthermore, in order to increase the credibility of the estimated economic losses, the coefficients are chosen so that the resulting loss function approximates that recommended by SWD for the year 2000 for shortages between 10% and 20%.

Due to the non-linearity of the loss-function, hedging suggests accepting small present economic losses in order to decrease future losses resulting from severe droughts. Hedging results in reducing shortage intensities; the cumulative deficit, however, remains unchanged.

Because hedging only decreases the intensities of the shortages and does not actually reduce the cumulative deficits, and because of the non-linearity of the loss function, it is most desirable to use an operating policy that results in equalizing the anticipated individual deficits for each month over the operating period. The procedure used in this study for obtaining the optimum operating policy is an iterative one. Based on streamflow forecasts and water supply requirements, the optimum operating policy apportions water releases in such a way to minimize economic losses resulting from water shortages. This iterative approach could have been avoided through the use of linear programming employing a simple operating rule such as the linear decision rule. In this study, however, the use of linear programming is made unattractive by the large number of constraints that need to be solved as compared to efficient and simple simulation.

4.2 METHODOLOGY :

The model simulates the operation of a single reservoir with a single inflow, a single outflow, and a single purpose, namely, water supply for the Seattle metropolitan area. Reservoir storage is assumed to be limited by the sum of the active storage capacities of the Cedar and the Tolt reservoirs. As has been discussed earlier, these are not constant over the year. Table 1 summarizes the monthly varying active storage capacities of the reservoirs. Estimates of monthly inflow data for the period October 1929 to September 1977 for the rivers were provided by the Seattle Water Department. In this study, it was more appropriate to define a water year as starting in April. Therefore, data for 47 years, from April 1930 to March 1977, are used in the experiments. Outflow, or release, from the reservoir is controlled by the level of the demand and the sum of the present storage and forecasted inflow. Demand is calculated by multiplying the constant base demand of 235 MGD (22,000 ac-ft/month) by its appropriate demand load factor (Table 3).

The simulation model employs two independent variables, forecast period length and forecast accuracy, to investigate their impact on economic losses resulting from water supply shortages. The model first reads 47 years of monthly inflow data. For a specific level of forecast uncertainty and a specific length of forecast period, monthly flows are predicted. The shortfalls of a normal operation of the reservoir are then evaluated. The term 'normal operation' refers to reservoir operation employing the policy 'release equals demand'. If such a policy results in no water shortages, then it is used in the actual operation of the reservoir. Otherwise, if it does result in deficits, another policy that minimizes economic losses resulting from such deficits is devised. The forecasted flow for the first month is then updated by the actual flow and the reservoir is operated for one month according to the adopted policy. This procedure is repeated over the entire record, each time advancing the period of operation by one month. Losses over the entire record are summed. The resulting loss represents the economic costs that are expected if the reservoir is operated over the 47 year historical record using that forecasting scheme and that length of operating period.

The procedure described above (Figure 7) is the core of all the simulations conducted in this research. Variations in forecast period, accuracy, and scheme are employed to investigate the basic questions addressed at the beginning of this chapter.

4.3 EXPERIMENTAL DESIGN

This study addresses these questions through two basic approaches. The first establishes a relationship of economic costs as a function of forecast period and accuracy. The second applies variations of the lag-one autoregressive Markov forecasting technique in which the system is operated using variable forecasting periods to relate economic losses of water shortages to the forecast period length associated with that predicting scheme. Results from both approaches are compared to assess the accuracy of the Markov scheme as a forecasting technique and the extent to which it can be practically improved. Section 4.3.1 describes the experimental design used in the first approach to develop a relationship between economic

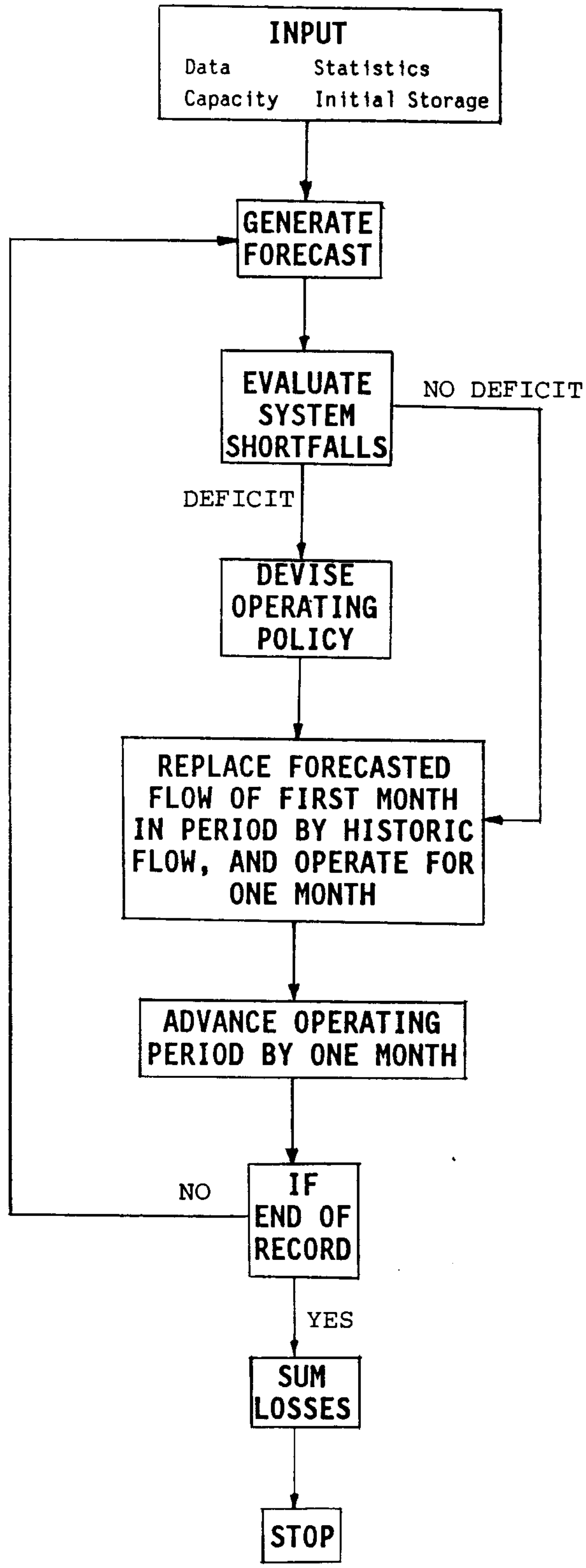


Figure 7
The Simulation Model: Flow Chart

losses as a function of forecast period and accuracy. Section 4.3.2 describes the second approach in which some simple available forecasting schemes are employed to assess their accuracy and applicability.

4.3.1 Parametrically Varied Forecast Period and Accuracy:

The first three questions are answered by relating incremental operational losses to parametrically varied streamflow prediction period and accuracy. This approach does not employ any streamflow generating schemes. Results are used to develop a relationship representing economic costs as a function of both the level of forecast uncertainty and the length of the forecasting period. Forecast uncertainty is represented as a noise term added to the forecast. A noise term of zero implies a perfect forecast while a large magnitude for that term implies a very poor forecast.

The loss analyses are based on monthly streamflows over 47 years (1929-1976). Losses are related to streamflow forecast period and accuracy. Forecast periods range from one month to one year and streamflows are predicted as actual historical flows perturbed by forecast error. Forecast error is generated by a normally distributed random number (ε) with zero mean and unit variance multiplied by a coefficient (σ) and by the actual historic flow. The coefficient, σ , is parametrically varied between zero and 0.9 to reflect forecast uncertainty. Hence, for a given month, j, the forecast, F_j, is represented by

$$F_j = AF_j + \varepsilon_j * \sigma_f * AF_j$$

where AF_j is the actual historic flow during month j and ε_j a random number, normally distributed with zero mean and unit variance. σ_f is a coefficient parametrically varied to reflect forecast uncertainty.

The level of forecast uncertainty is reflected by the degree to which the noise term affects the forecast. σ_f is parametrically increased from 0 to 0.9. A value of zero for σ_f implies a perfect forecast. In this case, flow forecasts as well as the predicted losses from these forecasts are identical to reality. For each value of σ_f, the reservoir is operated over the entire 47 year record 100 times. Analysis indicates this to be a sufficient number of simulations to limit sampling error. The losses obtained from these 100 simulations are averaged and their mean represents the expected losses associated with that level of uncertainty. ε_j is generated for each month and is reset to its original value for each new value of σ_f.

The experiment described above assumes a constant length of forecasting period. Therefore, the results define the relationship between forecast uncertainty level and the expected economic losses, associated with that length of forecasting period. The procedure is repeated using differing forecasting period lengths ranging from one month to one year. The results describe the relationship of economic costs of water supply shortages as a function of forecast period and accuracy.

4.3.2 Markovian Forecasting Schemes :

The fourth question is addressed through the application of the Markov lag-one autoregressive scheme for streamflow forecasting. To predict future inflows, the noise term is dropped from the generating equation. The forecasting equation thus consists of a deterministic part only. It is assumed that the entire influence of the past on the current flow is reflected in the previous flow value. Therefore, flows are predicted as the sum of the historic mean of that month and a fraction of the past month flow's deviation from its historical mean. The forecasting scheme is thus given by :

$$F_i = \mu_i + \rho_i * (\sigma_i / \sigma_{i-1}) * (AF_{i-1} - \mu_{i-1})$$

where μ is the historical mean, σ the standard deviation, F the predicted flow, AF the actual historic flow, and ρ the correlation coefficient.

Four variations on the scheme are made by assuming different characteristics of the inflow distribution. The first assumes no correlation between months. In this case, the historic mean of monthly flows is used to predict future inflows. The second assumes a normal distribution of inflows, while the third and fourth assume log-normal inflow distributions.

4.3.2.1 Historical Mean as Forecast

This experiment does not use any forecasting equation. Next month's forecast is the historic flow mean for that month. Of course, this method does not make use of much of the information provided by the historical record.

4.3.2.2 Normal Inflow Distribution

In order to make use of the historical flow record, a normal inflow distribution is assumed. The above stated equation is employed on the untransformed historical data. Negative generated flows are used to generate the next month's flow, and then equated to zero for use in the reservoir operation algorithm.

4.3.2.3 Log-Normal Inflow Distribution

This experiment assumes a normal distribution function for the natural logarithm transforms of the historical data. Streamflow, assumed to be a random variate X with mean μ_x and standard deviation σ_x, was transformed to a normally distributed random variate Y with mean μ_y and standard deviation σ_y by Y=lnX. The mean, standard deviation, and correlation coefficient are calculated for the natural logarithm transforms of the monthly flow data, and used in the generating equation. According to Hirsch [1979], such a generating scheme does not assure that the synthetic flows will resemble the historical flows in the long run with respect to the historical means, standard deviations, and correlation coefficients. What it does assure is that the logarithms of the generated flows will resemble the logarithms of the historical flows in the long run with respect to the first three moments.

4.3.2.4 Log-Normal Inflow Distribution

The assumption of a log-normal inflow distribution is also made in this experiment. However, to preserve the mean and standard deviation of the forecast flows in real space, the log normal transformation method described by Burges and Hoshi [1978] is used to determine the parameters to be used in the forecasting equation. The appropriate equations for μ_y and σ_y are :

$$\sigma_y = [\ln((\sigma_x/\mu_x)^2+1)]^{1/2}$$

$$\mu_y = \ln(\mu_x) - 0.5*\sigma_y^2$$

The correlation coefficient, ρ_y, is calculated in the usual manner using the natural logarithm transforms of the historic flows.

Summary:

This chapter described the model used to simulate the Cedar/Tolt water supply system. Some basic questions about the impact of forecast period and accuracy on economic losses resulting from water shortages were addressed. Two approaches were used to address these questions. The first produced a theoretical relationship between economic costs as a function of forecast period and accuracy. The second approach considered some practical aspects of forecasting schemes. The following chapter presents the results of the experiments described in this chapter. A detailed analysis is made and some conclusions drawn.

Chapter 5

RESULTS AND CONCLUSIONS

The preceding chapter described the simulation model used and the experiments conducted in this study. A set of questions were posed, and were addressed through two experimental approaches. This chapter presents the output from these experiments, compares the results from the different approaches, and attempts to interpret them in a manner useful to the operator.

5.1 IMPACT OF FORECASTS ON RESERVIOR OPERATION:

The theoretical approach addresses the first three questions : 1) How are economic costs resulting from water supply shortages related to the accuracy of the forecast? 2) Does increasing the length of the forecasting period reduce economic losses, and can this be related to an increase in forecast accuracy? 3) What are the potential benefits from increasing forecast accuracy?

5.1.1 Relationship between Losses and Forecast Accuracy :

These questions are addressed by simulating reservoir operation using forecasts whose length and uncertainty are parametrically varied, and noting the impacts of such variations on reservoir operation. In this approach, forecasts are calculated as the sum of two terms. The first term is the actual historic flow that occured during that month. The second term is a noise term normally distributed with a zero mean and unit variance, and multiplied by a coefficient that is parametrically varied between zero and nine tenths the value of the actual historic flow. A value of zero for the coefficient implies that the noise term does not contribute to the forecast. This represents a perfect forecast. It is expected that operating the reservoir with a forecast uncertainty level of zero produces the least water shortages and minimum economic losses. This is verified by the results of the experiments (Table 4 and Figure 8). As the contribution of the noise term is parametrically increased to its maximum value of 0.9 of the actual historic flow, it is observed that there is a consistent increase in water supply shortages, resulting in greater economic losses. Table 4 indicates that economic costs corresponding to the maximum uncertainty level may be up to $ 408 million (or 2000%) higher than those corresponding to perfect forecasts (for a twelve month forecasting period).

Therefore, the quality of the forecast has a major impact on economic losses from water supply shortages. For any length of forecasting period, there is a clear and

TABLE 4

Economic Losses (10^7 \$) vs. Forecast Period and Accuracy

Forecast Period Length	Forecast Uncertainty									
(months)	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
1	11.70	12.40	13.44	15.02	17.53	21.48	25.44	28.73	31.41	33.87
2	5.72	6.58	7.64	9.21	12.16	15.96	20.07	24.48	28.62	32.12
3	3.45	4.05	5.02	6.84	9.92	14.43	19.34	24.25	28.48	32.70
4	2.32	2.83	3.76	5.60	8.90	13.49	18.37	23.11	27.68	32.42
5	1.88	2.33	3.30	4.94	7.86	12.34	18.19	24.10	29.75	35.31
6	1.88	2.30	3.17	4.91	8.60	13.79	19.31	25.23	31.25	37.18
7	1.88	2.30	3.22	4.97	8.57	13.75	20.25	27.02	33.56	39.87
8	1.88	2.30	3.23	4.90	8.30	13.55	20.36	27.69	34.63	41.28
9	1.88	2.30	3.23	4.95	8.31	13.46	19.54	26.88	34.49	41.52
10	1.88	2.30	3.22	4.84	8.08	13.09	19.12	26.09	33.25	40.39
11	1.89	2.29	3.16	4.98	8.29	13.36	19.68	26.41	33.55	40.66
12	1.90	2.35	3.37	5.21	8.78	14.21	21.30	28.90	36.04	42.71

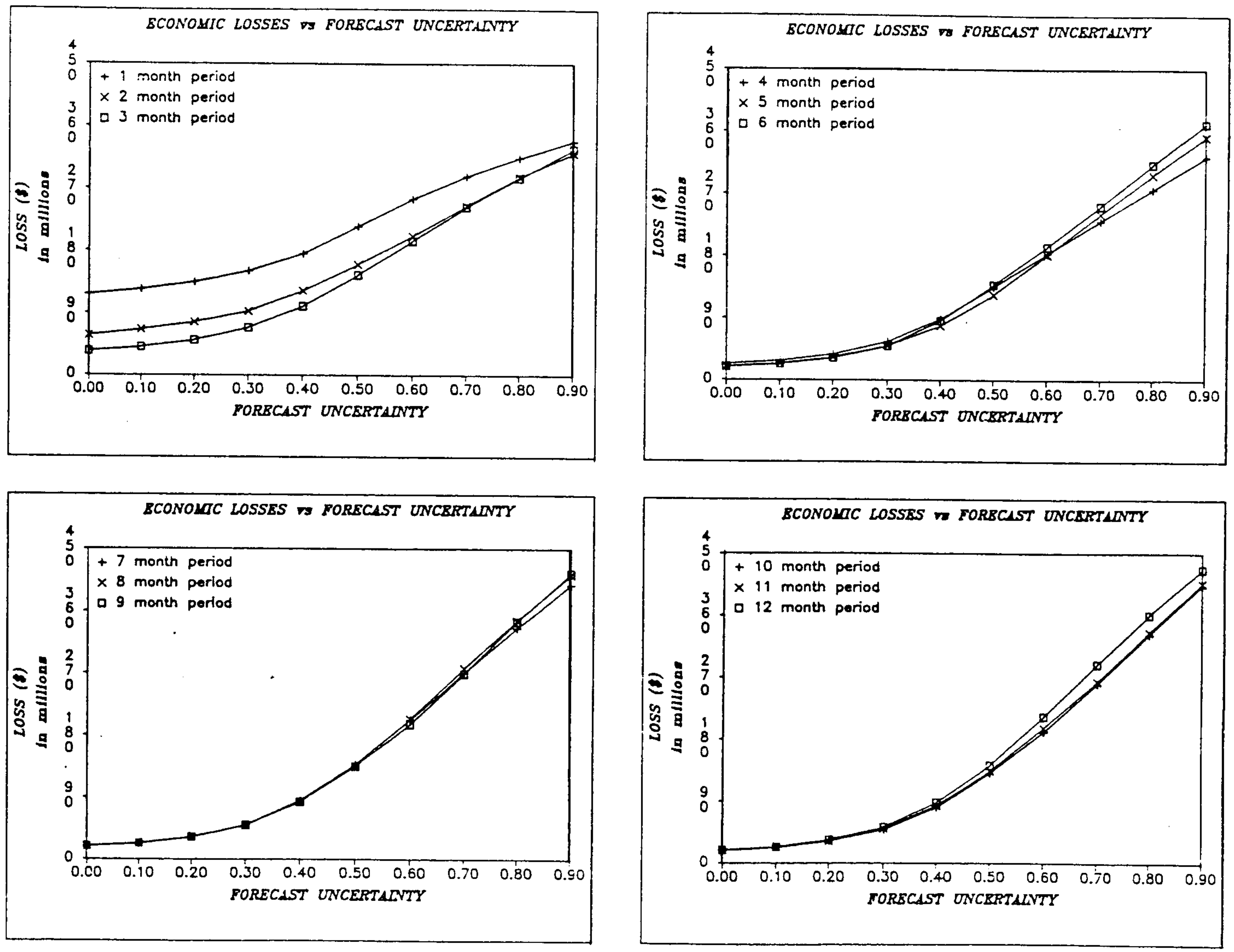


Figure 8

Economic Losses vs. Forecast Period and Accuracy consistent increase in economic costs with an increase in forecast uncertainty. Figure 8 also shows that, generally, for a high quality forecast a unit change in quality produces a smaller change in losses than it would for a low quality forecast. This is expected because the value of information is highest when there is little or no information, and decreases as more information is gathered.

5.1.2 Impact of Forecast Period Length on Losses :

The second question addressed is that of the impact of forecast period length on economic costs. When is it beneficial to use a longer forecasting period? To what extent does a longer forecasting period reduce economic losses? Can reduced losses from increased forecast period lengths be related to increased accuracies?

These issues are addressed through the same approach described above. The procedure is repeated for variable lengths of forecasting periods. It is expected that a longer forecast period should provide the operator with more information and result in improved reservoir operation. Although this is observed for high quality forecasts (Table 4), the relationship does not hold for poor forecasts. As the quality of the forecast decreases (worse than 0.5), there is no direct relationship between economic costs and length of forecasting period. Results show that long-term forecasts using a poor forecasting scheme give results that are, on the average, inferior to those given by short-term forecasts using the same scheme. This is best portrayed for forecasts with uncertainty levels of 0.9 (Table 4). For such forecasting schemes, a two month forecasting period produces $ 321 million in economic losses which is about 25% lower than the $ 427 million incurred if the reservoir is operated using a twelve month forecasting period.

Table 4 shows that a two month forecasting period produces results that are always superior to those produced by a one month forecast. This holds for any forecast uncertainty level. This reflects the fact that a one month forecast provides very little information, and a two month forecast always provides more information, even if the forecast quality is poor. This implies that a forecasting period of one month should never be employed. A superior operation is always achieved by using forecast period lengths longer than one month.

Table 4 also shows that as the accuracy of the forecast increases, it becomes more advantageous to use forecasting periods longer than two months. However, for any accuracy level, there is an optimal forecast period length. For a perfect forecast, the optimal length of forecast period is five months. Using longer periods produces little variation in losses. For forecast accuracies ranging between 0.2 and 0.4, using a forecasting period of five to eleven months produces optimal reservoir operation. However, there is a clear deterioration in the operation if a twelve month forecasting period is used. Finally, for poorer forecasting schemes, the optimal length ranges between two and five months. In contrast with higher quality forecasts where there is a range of optimal lengths of forecasting periods, in poor forecasts the use of forecast periods longer than the optimal produces inferior results .

Therefore, longer forecast periods generally provide more information than short forecasts do. However, depending on the quality of the forecast, there is an optimal length of forecast period, beyond which a longer period will provide more bad information than useful information, and thus produces inferior results.

5.1.3 Potential Benefits of Forecasting :

The potential advantages of forecasting are assessed through an analysis of benefits. In order to make such an assessment, economic losses incurred by operating the reservoir with a certain forecasting scheme must be compared to those incurred if no forecasting is used. Therefore, the reservoir system is operated over the entire historical record without forecasting, employing the policy 'release equals demand'. This operating policy resulted in an estimated economic losses of $ 155,488,700. The benefits associated with a forecasting scheme are then measured by the degree to which economic losses could be reduced if that technique is used. The percent improvement of a technique is calculated by normalizing the resulting benefits from that technique with respect to the economic losses resulting from operating the system without forecasting. 'Benefits' and 'Improvement' are defined by:

BENEFITS(I) = 155,488,700 - LOSS(I)

% IMPROVEMENT(I) = BENEFITS(I)/155,488,700 * 100

where I refers to an experiment.

According to these definitions, if the use of a forecasting technique results in no economic losses, this technique is said to improve the operation of the system by 100% over operation with no forecasting. As expected, this never occurs in the study. Even if it were possible to operate the reservoir with perfect foresight, the best that can be achieved is an 88% improvement over operation with no forecasting. This demonstrates the limitations of the reservoir in handling drought conditions. Therefore, no matter how good the available forecasting techniques are, shortages cannot be prevented due to the inherent limitations in the reservoir design. Inspite of this fact, it has been clearly shown that through the use of appropriate forecasting techniques, economic losses may be reduced by up to 88% over those resulting from operating the system without forecasting.

On the other hand, a forecasting technique resulting in a negative improvement implies a negative value for the information obtained by that forecast, or misinformation. Table 5 presents the percent improvement associated with a forecast uncertainty level and a forecast period length. It shows that any forecast having a level of accuracy corresponding to 0.6, or worse, should never be employed because, on the average, information gained through its use will lead to inferior decisions. Similarily, forecast schemes having an uncertainty level of 0.4 should not be employed with a one month forecasting period, and those with a 0.5 level of uncertainty should not be employed with one or two month forecasting periods.

TABLE 5

Percent Improvement vs. Forecast Period and Accuracy

Forecast Period Length	Forecast Uncertainty									
(months)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	24.73	20.23	13.54	3.38	-12.77	-38.18	-63.66	-84.82	-102.06	-117.89
2	63.20	57.68	50.88	40.78	21.77	-2.67	-29.11	-57.48	-84.11	-106.63
3	77.81	73.93	67.70	55.99	36.18	7.17	-24.41	-56.00	-83.21	-110.36
4	85.08	81.77	75.83	63.97	42.78	13.22	-18.17	-48.67	-78.07	-108.56
5	87.91	85.02	78.75	68.20	49.46	20.62	-17.02	-55.04	-91.38	-127.15
6	87.91	85.24	79.59	68.39	44.66	11.29	-24.22	-62.30	-101.03	-139.18
7	87.91	85.20	79.29	68.05	44.90	11.55	-30.27	-73.82	-115.89	-156.48
8	87.91	85.18	79.25	68.46	46.59	12.83	-30.98	-78.13	-122.77	-165.55
9	87.91	85.18	79.23	68.19	46.52	13.41	-25.70	-72.92	-121.87	-167.10
10	87.91	85.22	79.32	68.90	48.04	15.79	-23.00	-67.84	-113.90	-159.83
11	87.84	85.27	79.65	67.94	46.68	14.06	-26.60	-69.90	-115.83	-161.57
12	87.78	84.89	78.31	66.50	43.53	8.59	-37.02	-85.91	-131.84	-174.75

Table 5 also confirms the fact that a two month forecast is always better than a one month forecast. For schemes with forecast uncertainties ranging from zero to 0.3, a two month forecast produces results that are up to 40% better than those of a one month forecast. This clear improvement in reservoir operation is continued to be observed up to a forecast period of five months. Figure 9 demonstrates that forecast periods longer than five months show little or no improvement. On the other hand, the optimal forecast period to be used with schemes having accuracy levels of 0.4 and 0.5 is five months, and a longer period produces inferior results.

The benefits analysis has shown that by the use of proper forecasting techniques, reservoir operation may be improved by up to 88%. It has also shown that forecast schemes with 0.6 levels of uncertainty (or worse) should never be used because the operator is more likely to operate the system with less shortages even without the use of any forecasting scheme. Furthermore, for forecasting schemes with an uncertainty level of 0.5 or better, the optimal length of forecasting period is five months.

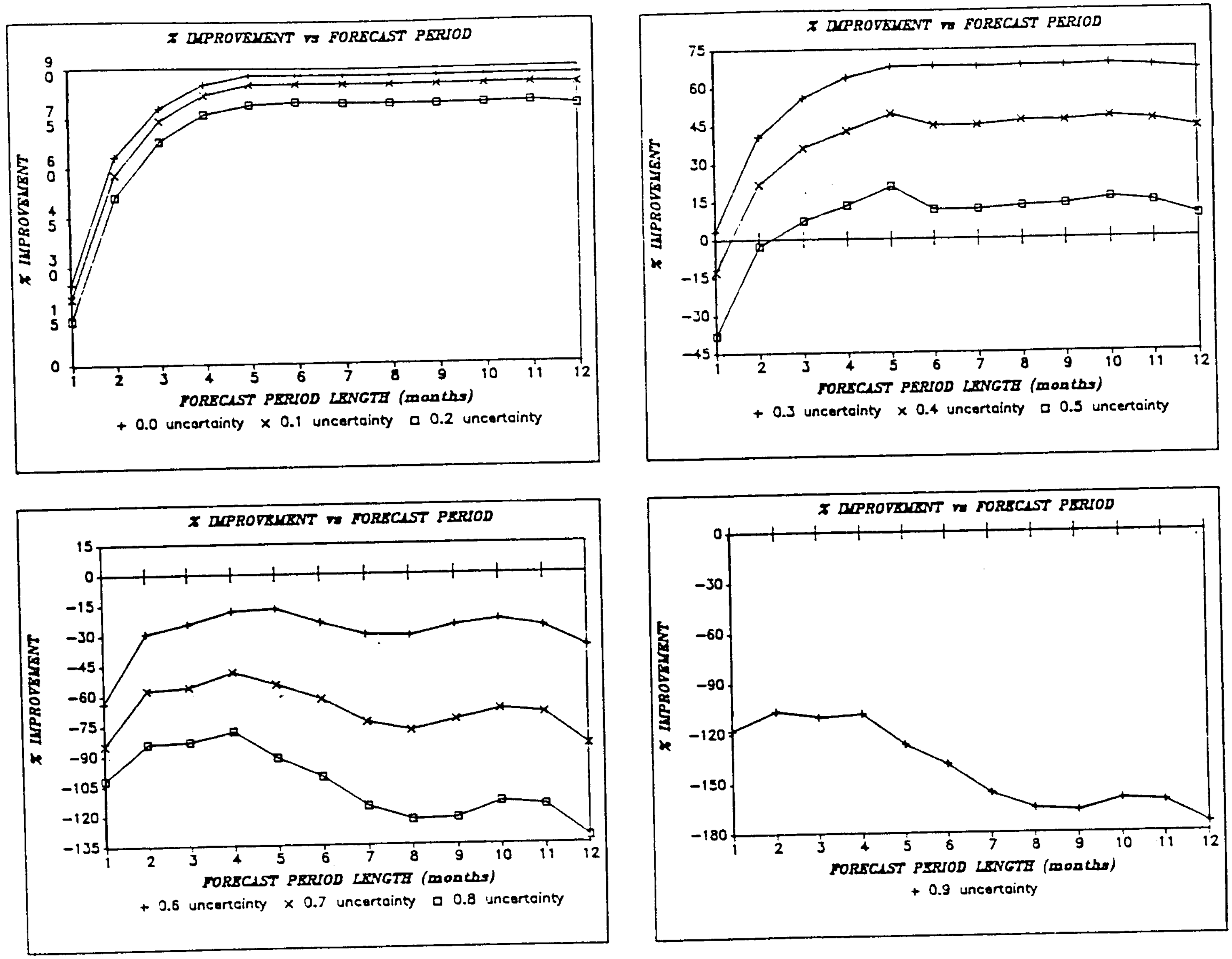


Figure 9
Percent Improvement vs. Forecast Period & Accuracy

5.2 RESULTS FROM EMPLOYING SIMPLE FORECAST SCHEMES :

In section 5.1, the accuracy and the length of period of a forecast were parametrically varied to observe their impact on economic losses resulting from water supply shortages. In contrast, this section evaluates some simple available forecasting techniques. The economic losses resulting from the use of such techniques are compared to the results from section 5.1 to assess the technique's accuracy. Forecast period length is parametrically varied between one month and one year to observe its impact on reservoir operation.

The simplest prediction method, and the least accurate, is using the historical mean of the monthly inflow data to predict future inflows (Table 6-method 1). This method only makes use of the information provided by the mean. The standard deviation as well as the correlation coefficient are completely ignored. In order to make better use of the information furnished by the historic data, a normal distribution of reservoir inflows is assumed. A simple lag-one autoregressive Markov scheme makes use of the first three moments of the untransformed inflow distribution to generate the forecasts (Table 6-method 2). In an attempt to improve the quality of the forecasts, a log-normal inflow distribution is assumed

(Table 6-method 3). The data are transformed to their natural logarithms, the first three moments calculated, and the same Markov scheme applied to the transformed data and their moments. However, this scheme preserves the three statistics of the data in log space only. Finally, in order to preserve the three moments in real space, the log normal transformation method described by Burges and Hoshi [1978] is employed (Table 6-method 4).

The previous paragraph described the four simple forecasting schemes employed in this study. Table 6 and Figure 10 summarize the results of the experiments and the economical losses expected from using these schemes.

Figure 10 indicates that if a forecasting period of one month is to be used (very unlikely), the best available forecasting scheme is using the mean of the historic flow data as the forecasts. However, this scheme is superior only if a one month forecast period is used. The use of Markovian forecasting techniques with the assumption of a normal inflow distribution function shows a clear improvement over using the mean of the historic data to predict future inflows. The assumption of a log normal inflow distribution, without the use of the log normal transformation method described by Burges and Hoshi [1978], produces results that are inferior to those realized by the assumption of a normal distribution of reservoir inflows. Finally, the use of the log normal transformation method to preserve the first three moments of the historic data in real space shows the best results.

The equations in section 5.1.3 are used to make an analysis of benefits accrued from using the above described forecasting schemes. Table 7 and Figure 10 summarize the results of the benefits analysis and demonstrates that the mean of the historic data should never be used as a forecasting method. The use of this scheme with forecasting periods of one or two months actually produces results that are inferior to those if no forecasting is used. Even with longer forecasting periods, its use improves reservoir operation only marginally (2.4% improvement over no forecasting). Larger benefits are realized by the use of any variation on the Markov forecasting schemes. Although a log normal inflow distribution function employing the log normal transformation method described by Burges and Hoshi [1978] shows greatest improvements, a clear conclusion about which method is the best cannot be made. All Markov methods seem to improve reservoir operation by 9%, plus or minus 1%.

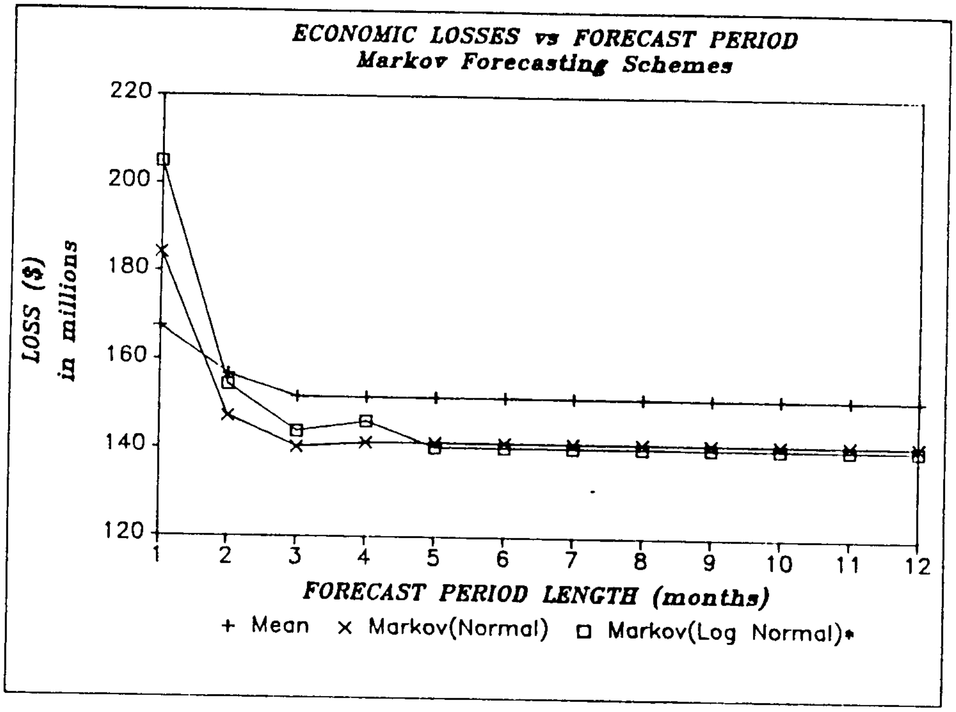
TABLE 6
Economic Losses vs. Forecast Period
for Various Schemes

Forecast Period Length (months)	Forecasting Method			
	Method 1	Method 2	Method 3	Method 4*
1	1.676E+08	1.841E+08	2.050E+08	2.049E+08
2	1.567E+08	1.471E+08	1.566E+08	1.544E+08
3	1.517E+08	1.402E+08	1.471E+08	1.438E+08
4	1.517E+08	1.413E+08	1.493E+08	1.461E+08
5	1.517E+08	1.413E+08	1.429E+08	1.403E+08
6	1.517E+08	1.413E+08	1.429E+08	1.403E+08
7	1.517E+08	1.413E+08	1.429E+08	1.403E+08
8	1.517E+08	1.413E+08	1.429E+08	1.403E+08
9	1.517E+08	1.413E+08	1.429E+08	1.403E+08
10	1.517E+08	1.413E+08	1.429E+08	1.403E+08
11	1.517E+08	1.413E+08	1.429E+08	1.403E+08
12	1.517E+08	1.413E+08	1.429E+08	1.403E+08

TABLE 7
Percent Improvement vs. Forecast Period
for Various Schemes

Forecast Period Length (months)	Forecasting Method			
	Method 1	Method 2	Method 3	Mehtod 4*
1	-7.84	-18.46	-31.86	-31.79
2	-0.78	5.34	-0.72	0.70
3	2.43	9.79	5.37	7.52
4	2.43	9.09	3.94	5.98
5	2.43	9.09	8.05	9.77
6	2.43	9.09	8.05	9.77
7	2.43	9.09	8.05	9.77
8	2.43	9.09	8.05	9.77
9	2.43	9.09	8.05	9.77
10	2.43	9.09	8.05	9.77
11	2.43	9.09	8.05	9.77
12	2.43	9.09	8.05	9.77

<u>Key to Forecast Methods</u>
Forecast Method 1 : Mean of Historic Data
Forecast Method 2 : Markov - Normal Inflow Distribution
Forecast Method 3 : Markov - Log Normal Inflow Distribution
Forecast Method 4*: Markov - Log Normal Inflow Distribution
(* = Log Normal Transformation Method suggested
by Burges and Hoshi [1978] is used)



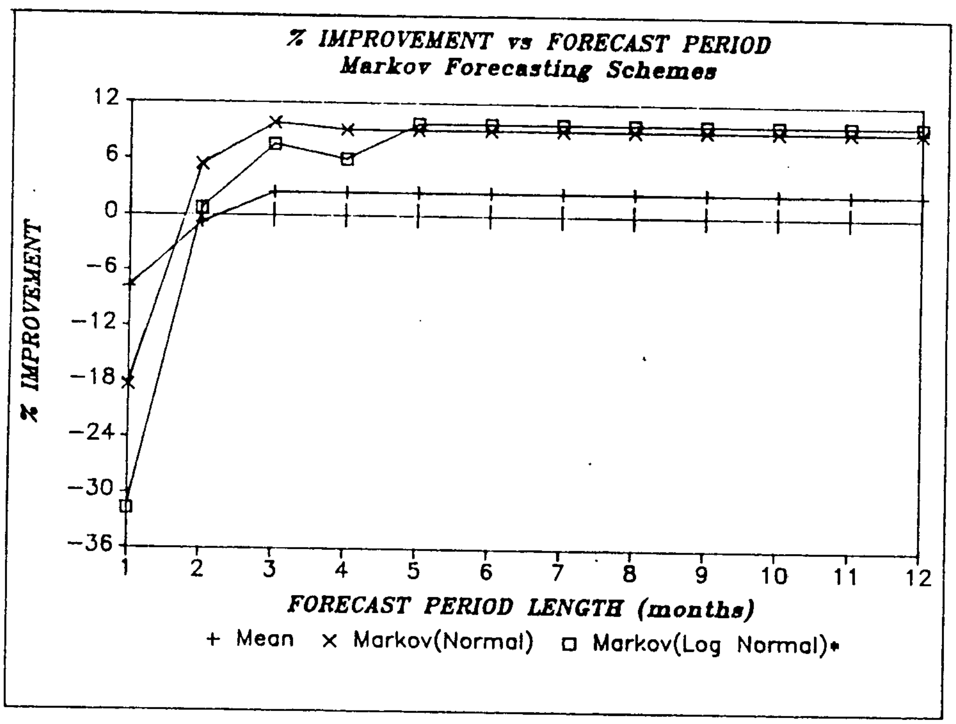


Figure 10

Economic Costs & Improvement of Markov Schemes

5.3 Performance of the Markov Models :

Figure 11 compares the improvement achieved by the two simple forecasting schemes to that achieved by a prediction technique having an uncertainty level of 0.5. For short forecast periods (one to two months), both forecasting techniques exhibit uncertainties better than 0.5. However, no advantage is gained by using such short-term forecasts (negative improvement). For longer forecast periods (three months or more), the forecasting techniques have an uncertainty level worse than 0.5. This demonstrates that the accuracy of a prediction method varies with the length of forecasting period.

5.4 Operating Period vs. Forecasting Period :

Throughout this report, the terms 'operating period' and 'forecasting period' are used interchangeably. Operating period refers to the period over which a policy that apportions water releases is devised. Forecast period refers to one over which reservoir inflows are predicted. In order to devise an operating policy, some form of forecast is required for each month of the operating period. Therefore, the lengths of the two periods should be equal; hence the interchangeable use of the terms.

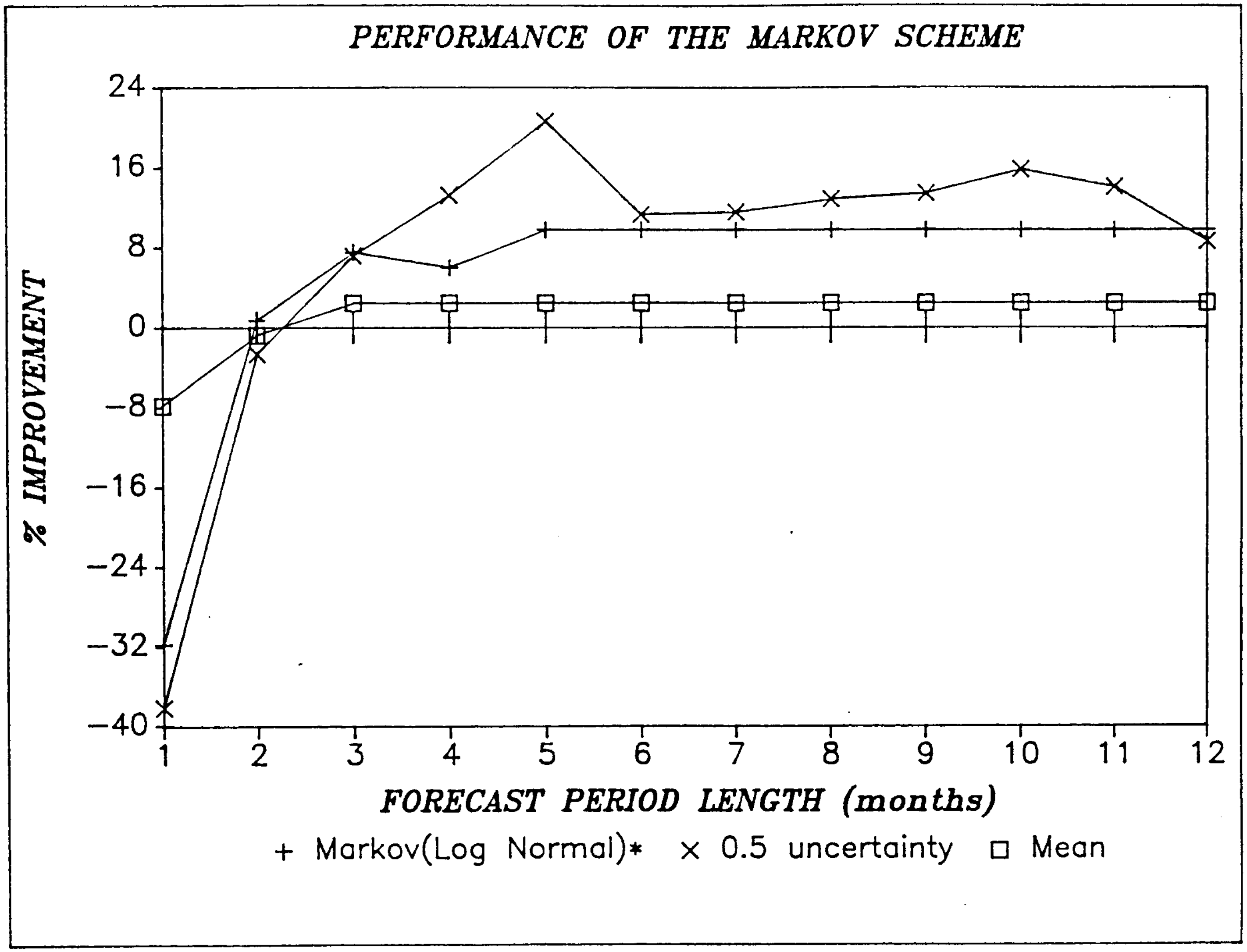


Figure 11

Performance of the Markov Scheme

In some cases, it may be beneficial to use a forecast period that is shorter than the operating period. For example, assume that a five month operating period is to be employed. On the other hand, the length of forecast period is chosen such that the operator feels confident of his predictions; assume a two month forecast period. Therefore, the operator relies on the available storage and predicted flows for the two months to devise an apportionment policy over the five month operating period. This assumes that there will be no inflow to the reservoir during the final three months of the operating period.

In order to assess the benefits of such a mode of operation, various combinations of operating/forecast periods are used. Forecasts are made using the lag-one auto-regressive Markov scheme with the assumption of a log normal inflow distribution function (Table 6-method 4). Operating periods of three months and five months are employed with shorter or same length forecast periods. The results of these simulations are presented in Table 8. For a three month operating period, it is best to predict inflows for every month in the period. On the average, for a five month operating period, it is best to predict inflows for the first four months, and assume that the inflow for the fifth month is zero.

TABLE 8

Impact of Unequal Forecast and Operating Periods on Economic Losses

Operating Period (months)	Forecast Period (months)	Economic Losses
3	0	54.2E+08
3	1	5.5E+08
3	2	1.5E+08
3	3	1.4E+08
5	0	198.4E+08
5	1	15.9E+08
5	2	5.5E+08
5	3	1.7E+08
5	4	1.4E+08
5	5	1.4E+08

These results show that the use of a forecast period that is much shorter than the operating period generally results in inferior results. This is due to the fact that zero inflows are predicted for some of the months, thus causing the initiation of water use restrictions when they are actually not needed. Such inferior results may be expected because the reservoir is operated over the entire historic record, the inflow mean of which is much higher than drought-time inflows. Therefore, such a policy where the forecast period is shorter than the operating period may be recommended only if a drought is anticipated.

5.5 Conclusions :

This chapter has presented and analyzed the results of the experiments conducted in this study. The following is a summary of the conclusions drawn from this analysis. It should be emphasized that what follows applies only to the simplified system under consideration. The results of this report should not be directly applied in other situations.

o A relationship between the economic losses that can be expected from water supply shortages as a function of forecast accuracy was established. It was concluded that the quality of the forecast has a major impact on the operation of the system, and, hence, on economic costs from water deficits. The expected economic losses of water supply shortages resulting from the operation of the reservoir system with a certain forecast accuracy level increased with an increase in forecast uncertainty. This relationship exhibited a decreasing slope for high quality forecasts, implying that a unit change in forecast quality has a bigger impact on economic losses for poor forecasts than for good ones. This confirms the fact that the value of information is highest when little or no information is available, and decreases as more information is gathered.

o A relationship of economic costs from deficits as a function of the forecast period was developed. Although it would be expected that longer forecast periods produce results that are always at least as good as shorter periods, this expectation is not always true. This relationship does hold for high quality forecasts. For perfect forecasts, there is a clear improvement in the reservoir operation with increasing forecast period lengths up to a length of five months, beyond which the use of longer forecast periods does not alter reservoir operation significantly. As the quality of the forecast decreases (0.2-0.4), the same plateau is observed. However, for such forecasts, the use of forecast periods longer than ten months produces inferior results. With a further decrease in forecast accuracy (worse than 0.4), the optimal length of forecast period ranges between two and five months. The use of longer forecast periods produces results that are clearly inferior to those realized by the use of the optimal length.

o An assessment of the potential benefits from forecasting is made. Benefits of a forecasting technique are defined by the degree to which economic losses can be reduced by the use of such a technique. It was shown that economic losses resulting from water supply shortages may be reduced by up to 88% just by the use of proper forecasting techniques. It was also shown that forecasting methods with 0.6 levels of uncertainty (or worse) are not recommended because the operator is more likely to operate the system with less shortages even without the use of forecasts. This benefits analysis also demonstrated that the optimal length of forecasting period is five months.

o The application of some simple forecasting schemes resulted in two basic conclusions : 1) The mean of the historic data should never be used to predict future inflows in reservoir operation; and 2) The use of the lag-one autoregressive Markov model has clear benefits; improvements of about 9% over no forecasting were realized.

In an attempt to improve the operation of the reservoir, differing assumptions about the inflow distribution function were made. The use of such assumptions showed that reservoir operation is not sensitive to the form of the input distribution function.

o Finally, the benefits of operating the reservoir employing a forecast period that is shorter than the operating period were assessed. It was shown that such operation is not recommended except if droughts are anticipated. If this is the case, some benefits might be realized through such a mode of operation.

Recommendations for Further Study :

The benefits analysis showed that, theoretically, the use of a proper inflow forecasting scheme may reduce economic losses of water supply shortages by up to 88%. Unfortunatley, none of the forecasting techniques employed in this research approached this value. The highest degree of improvement realized is through the use of the lag-one autoregressive Markov scheme with the assumption of log-normal inflow distribution function and employing the log normal transformation method described by Burges and Hoshi [1978] to preserve the first three moments of the data in real space. This method resulted in a 9.8%

improvement over no forecasting.

Although the above described method resulted in a substantial improvement over no forecasting, the potential benefits of forecasting are far from being achieved. This research assessed the quality of only some very simple techniques. More complex forecasting techniques having higher degrees of realism are available. It is important that such techniques be tested for their accuracy and the extent to which they can reduce losses.

It is the author's belief that as long as forecasts are made based on mathematical models not describing the actual physical system, even complex schemes will not result in significant improvements in reservoir operation and reduction of losses. Therefore, more effort should be devoted to the development of higher quality forecasting techniques.

Much effort has gone into the 'equivalent years of data' issue. A related, but different, concept is that of relating forecast accuracy and length of historic record. This may prove to be a valuable tool in the design of network systems and in determining the cut-off point of data collection.

Finally, this research simplifies the actual water resources system supplying the Seattle metropolitan area by the use of a model that simulates the operation of a single-purpose reservoir. In order to increase the applicability of the results, a model that better simulates the actual system should be developed and incorporated into the framework of the experimental design. Furthermore, a one week time-step should be used instead of one month.

BIBLIOGRAPHY

Askew, A.J., Yeh, W. W-G. , and Hall, W.A. , "A Comparative Study of Critical Drought Simulation," Water Resources Research, Vol. 7, No. 1, 1971, pp. 52-62.

Berschauer, D., "Water Supply Planning Utilizing Synthetic Streamflows and System Simulation Model : The Seattle Water Supply System as a Case Study," paper presented to the University of Washington, at Seattle, WA, in 1984, in partial fulfillment of the requirements for the degree of Master of Science in Civil Engineering.

Burges, S.J., and Hoshi, K., "Approximation of a Normal Distribution by a Three-Parameter Log Normal Distribution," Water Resources Research, Vol. 14, No. 4, August 1978, pp. 620-622.

Burges, S.J., and Lettenmaier, D.P., "Comparison of Annual Streamflow Models," Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers, Vol. 103, No. HY9, September 1977, pp. 991-1006.

Chi, M., Neal, E., and Young, G.K. , "Practical Application of Fractional Brownian Motion and Noise to Synthetic Hydrology," Water Resources Research, Vol. 9, No. 6, 1973, p. 1523.

Clarke, R.T., "Problems and Methods of Univariate Synthetic Hydrology, in Mathemetical Models for Surface Water Hydrology," edited by T.A. Ciriani, U. Maione, and J.R. Wallis, John Wiley, New York, 1977, pp. 1-26.

Dawdy, D.R., "The Worth of Hydrologic Data," Water Resources Research, Vol. 15, No. 6, December 1979, pp. 1726-1731.

Dawdy, D.R., Kubik, H.E., and Close, E.R., "Value of Streamflow Data for Project Design - A Pilot Study," Water Resources Research, Vol. 6, No. 4, August 1970, pp. 1045-1050.

Erickson, K.A., Palmer, R.N., and Lettenmaier, D.P., "An Interactive Simulation Model for the Cedar/Tolt Water Supply System," Technical Report No. 73, Charles W. Harris Hydraulics Laboratory, Seattle, Washington, January 1982, 104 pp.

Fiering, M.B., and Jackson, B.B., "Synthetic Stremflows, Water Resources Monograph Series," Vol. 1, AGU, Washington, D.C., 1971, 98 pp.

Garcia, L.E., Dawdy, D.R., and Mejia, J.M., "Long Memory Monthly Streamflow Simulation by a Broken Line Model," Water Resources Research, Vol. 8, No. 4, 1972, pp. 1100-1105.

Hirsch, R.M., "Synthetic Hydrology and Water Supply Reliability," Water Resources Research, Vol. 15, No. 6, December 1979, pp. 1603-1615.

Jackson, B.B., "The Use of Streamflow Models in Planning," Water Resources Research, Vol. 11, No. 1, February 1975, pp. 54-63.

Jettmar, R.V., and Young, G.K., "Hydrologic Estimation and Economic Regret," Water Resources Research, Vol. 11, No. 5, 1975, pp. 648-656.

Klemes, V., "Value of Information in Reservoir Optimization," Water Resources Research, Vol. 13, No. 5, October 1977, pp. 837-850.

Krzysztofowicz, R., "Why Should a Forecaster and a Decision Maker Use Bayes Theorem," Water Resources Research, Vol. 19, No. 2, April 1983, pp. 327-336.

Krzysztofowicz, R., and Watada, L.M., "Stochastic Model of Seasonal Runoff Forecasts," Water Resources Research, Vol. 22, No. 3, March 1986a, pp. 296-302.

Krzysztofowicz, R., "Optimal Water Supply Planning Based on Seasonal Runoff Forecasts," Water Resources Research, Vol. 22, No. 3, March 1986c, pp. 313-321.

Lenton, R.L., Rodriguez-Iturbe, I., and Shaake, Jr. J.C., "The Estimation of ρ in the First-Order Autoregressive Model : A Bayesian Approach," Water Resources Research, Vol. 10, No. 2, 1974, pp. 227-241.

Mandelbrot, B.B., "A Fast Fractional Gaussian Noise Generator," Water Resources Research, Vol. 8, No. 3, 547, 1971.

Mandelbrot, B.B., and van Ness, J.W., "Fractional Brownian Motions, Fractional Noises, and Applications, SIAM Rev., 10(4), 1968, pp. 422-437.

Mandelbrot, B.B., and Wallis, J.R., "Noah, Joseph, and Operational Hydrology," Water Resources Research", Vol. 4, No. 5, 1968, pp. 909-918.

Mandelbrot, B.B., and Wallis, J.R., "Computer Experiments with Fractional Gaussian Noises, 1, Averages and Variances," Water Resources Research, Vol. 5, No. 1, 1969a, pp. 228-241.

Mandelbrot, B.B., and Wallis, J.R., "Computer Experiments with Fractional Gaussian Noises, 2, Rescaled Ranges and Spectra," Water Resources Research, Vol. 5, No. 1, 1969b, pp. 242-259.

Mandelbrot, B.B., and Wallis, J.R., "Computer Experiments with Fractional Noises, 3, Mathematical Appendix," Water Resources Research, Vol. 5, No. 1, 1969c, pp. 260-267.

Mandelbrot, B.B., and Wallis, J.R., "Some Long-Run Properties of Geophysical Records," Water Resources Research, Vol. 5, No. 2, 1969d, pp. 321-340.

Mandelbrot, B.B., and Wallis, J.R., "Robustness of the Rescaled Range R/S in the Measurement of Noncyclic Long-Run Statistical Dependence," Water Resources Research, Vol. 5, No. 5, 1969e, pp. 967-988.

Moss, M.E.,"Optimum Operating Procedure for a River Gaging Station Established to Provide Data for a Water Supply Project," Water Resources Research, Vol. 6, No. 4, 1970, pp. 1051-1061.

Moss, M.E., and Karlinger, M.R., "Surface Water Network Design By Regression Analysis Simulation," Water Resources Research, Vol. 10, No. 3, 1974, pp. 427-433.

Rodriguez-Iturbe, I., Mejia, J.M., and Dawdy, D.R., "Streamflow Simulation, 1, A New Look at Markovian Models, Fractional Gaussian Noise, and Crossing Theory," Water Resources Research, Vol. 8, No. 4, 1972, pp. 921-930.

Tschannerl, G., "Designing Reservoirs with Short Streamflow Records," Water Resources Research, Vol. 7, No. 4, 1971, pp. 827-833.

Wallis, J.R., and Matalas, N.C., "Correlogram Analysis Revisited," Water Resources Research, Vol. 7, No. 6, 1448, 1971.

Yeh, W. W-G., Becker, L., and Zettlemoyer, R., "Worth of Inflow Forecast for Resrvoir Operation," Journal of the Water Resources Planning and Management Division, Proceedings of the ASCE, Vol. 108, No. WR3, October 1982, pp. 257-269.

Yeh, W. W-G, "Reservoir Management and Operations Models : A State-of-the-Art Review," Water Resources Research, Vol. 21, No. 12, December 1985, pp. 1797-1818.

APPENDIX A

STATISTICS OF THE INFLOW DATA

This appendix presents the statistics of the monthly flow data used in the Markov forecasting schemes. The U.S. Corps of Engineers supplied the inflow records for the Cedar and the South Fork Tolt rivers. These flows were not measured directly, but were computed using gaged streamflows in surrounding basins. Computed data were available for the years 1929-1975. The Seattle Water Department extended this record up to 1977. For this study it was more appropriate to define a water year as starting in April. Therefore, data for the years April 1930 to March 1977 are used. Tables A-9, A-10, and A-11 represent the statistics of the combined flow data for the two rivers. The skew coefficient is not used in any of the experiments, but is presented here for the purposes of completeness.

Table A-9 presents estimates of the statistics of the data in normal space for use in the Markov forecasting model.

TABLE A-9

Historical Flow Statistics - Real Space

Month	Mean	Standard Deviation	Correlation Coefficient	Skew Coefficient
APR	51,362	13,461	0.38132	-0.15
MAY	63,188	19,248	0.24198	0.12
JUN	50,127	25,143	0.59938	0.69
JUL	20,710	13,828	0.87411	1.02
AUG	7,973	4,761	0.71300	1.16
SEP	12,528	10,901	0.31200	2.74
OCT	29,279	16,919	0.41228	0.58
NOV	52,403	27,477	0.53985	0.79
DEC	64,320	32,387	0.33690	1.62
JAN	55,924	29,044	0.26662	0.67
FEB	45,473	23,222	0.20552	0.95
MAR	41,384	19,251	0.25933	2.00

Table A-10 estimates the statistics of the natural logarithm transforms of the inlow data. Streamflow, assumed to be a random variate X with mean μ_x and variance σ_x^2, is transformed to a normally distributed random variate Y with mean μ_y and variance σ_y^2 by Y = lnX.

TABLE A-10

Historical Flow Statistics - Log Space

Month	Mean	Standard Deviation	Correlation Coefficient	Skew Coefficient
APR	10.808	0.296	0.41188	-1.01
MAY	11.004	0.328	0.26870	-0.51
JUN	10.687	0.556	0.70665	-0.63
JUL	9.723	0.673	0.85491	0.03
AUG	8.819	0.586	0.73923	-0.02
SEP	9.181	0.697	0.44633	0.30
OCT	10.082	0.704	0.48615	-0.78
NOV	10.695	0.677	0.70160	-1.46
DEC	10.959	0.483	0.42160	-0.12
JAN	10.787	0.569	0.10449	-0.40
FEB	10.600	0.512	0.22190	-0.08
MAR	10.549	0.393	0.23188	0.59

The use of the Markov model with the above estimates of the statistics preserves the first three moments of the data in log space only. In order to preserve the statistics of the historic data in real space, the log normal tranformation method described by Burges and Hoshi [1978] is used. The mean and standard deviation used in the forecasting scheme are given by :

$$\sigma_y = [\ln((\sigma_x/\mu_x)^2 + 1)]^{1/2}$$

$$\mu_y = \ln \mu_x - 0.5 * \sigma_y^2$$

Table A-11 presents the estimates of these statistics. The correlation coefficient is calculated in the normal manner.

TABLE A-11

Historical Flow Statistics - Log Space

Month	Mean*	Standard Deviation*	Correlation Coefficient
APR	10.813	0.258	0.41188
MAY	11.010	0.298	0.26870
JUN	10.710	0.474	0.70665
JUL	9.754	0.607	0.85491
AUG	8.831	0.552	0.73923
SEP	9.154	0.751	0.44633
OCT	10.141	0.537	0.48615
NOV	10.745	0.493	0.70160
DEC	10.959	0.475	0.42160
JAN	10.812	0.489	0.10449
FEB	10.609	0.481	0.22190
MAR	10.533	0.443	0.23188

* Using the log normal transformation method suggested by Burges and Hoshi [1978] to preserve the statistics of the data in real space.

The statistics presented in Tables A-12 and A-13 are not employed in reservoir operation experiments. They are presented here in order to give the reader an idea about reservoir inflows for the two rivers. All figures are in cubic feet per second (cfs).

TABLE A-12

Historical Flow Statistics - Cedar

Month	Mean	Standard Deviation	Skew Coefficient	Correlation Coefficient
APR	644	167	-0.30	0.25130
MAY	802	258	0.15	0.22700
JUN	644	334	0.69	0.57592
JUL	241	179	1.16	0.87336
AUG	83	53	1.04	0.74240
SEP	125	122	3.01	0.32538
OCT	323	204	0.70	0.49036
NOV	623	363	0.96	0.52268
DEC	781	438	1.92	0.30894
JAN	668	343	0.54	0.24542
FEB	567	296	0.92	0.23131
MAR	502	237	2.08	0.26521

TABLE A-13

Historical Flow Statistics - S.F. Tolt

Month	Mean	Standard Deviation	Skew Coefficient	Correlation Coefficient
APR	208	76	0.53	0.56103
MAY	243	78	-0.52	0.21359
JUN	185	93	0.46	0.66154
JUL	101	65	0.89	0.66581
AUG	49	34	1.30	0.56432
SEP	81	63	1.99	0.26858
OCT	161	86	0.16	0.24747
NOV	242	106	-0.02	0.57287
DEC	282	112	0.25	0.48983
JAN	251	146	1.03	0.36150
FEB	181	96	0.94	0.16340
MAR	180	90	1.42	0.31170