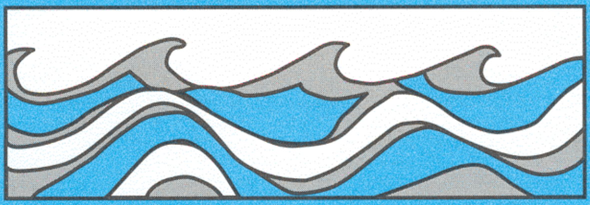
University of Washington
Department of Civil and Environmental Engineering

1861

EVALUATION OF THE EXPANSION OF WALSH LAKE

Robert W. Schanz
Richard N. Palmer



Water Resources Series
Technical Report No.96
June 1985

Seattle, Washington
98195

Department of Civil Engineering
University of Washington
Seattle, Washington 98195

EVALUATION OF THE EXPANSION OF WALSH LAKE

Robert W. Schanz
Richard N. Palmer

Water Resources Series
Technical Report No. 96

June 1984

ACKNOWLEDGEMENTS

Funding for this project was provided by the Seattle Water Department. This study was supervised by Richard N. Palmer, Assistant Professor of Civil Engineering and the work presented herein served as the basis for the first author's Masters thesis.

The authors wish to acknowledge the support and suggestions of David Parkinson of the Seattle Water Department. Lori Nagle and Stephanie Boyd produced this report and their patience with our countless revisions is appreciated. The authors also wish to thank Diana Leytham for her editorial support and helpful suggestions. Previous research by Dr. Robert Hirsch of the U.S. Geological Survey served as a point of departure for much of this study, and his comments are greatly appreciated.

ABSTRACT

This report addresses issues involved in the estimation of inflows for a proposed reservoir at Walsh Lake. Reliable streamflow data are scarce for the proposed reservoir site. Therefore, the first issue examined is the estimation of streamflow statistics from regional data. Drainage basin characteristics and flow data for nearby streams are used to derive estimates of flow statistics for the reservoir inflows. Reservoir operation is then simulated using these regional estimates of inflows to obtain preliminary evaluations of reservoir performance. Regionally estimated inflows are not sufficiently accurate for design purposes. A second issue is therefore the development of sampling strategies for the accurate estimation of reservoir performance indices. It is shown that after a number of years the value of additional data becomes insignificant relative to data collection costs. Methods for extending existing inflow data are examined as a means of decreasing data requirements. These record extension techniques make use of correlations between concurrent flows at different sites within a region. The application of record extension to the reservoir site is simulated for various record lengths. Results indicate that the use of streamflow record extension significantly increases the accuracy of reservoir reliability extimates and reduces the number of years of streamflow data needed at the reservoir site.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Number Page

EXECUTIVE SUMMARY

A proposed reservoir at Walsh Lake was examined as a means of increasing water supply yield from the Cedar River. Storage in this reservoir is to be used for meeting instream flow requirements for fish in the Cedar River, thus allowing for increased withdrawals at the Landsburg diversion. Proper evaluation of this alternative is limited by the scarcity of data that is available for the streams that contribute to inflows to Walsh Lake. To address these limitations three research tasks were accomplished: (1) Initial estimates of annual and monthly streamflows into Walsh Lake were made using currently available data; (2) Estimates of Walsh Lake performance were made based upon these initial estimates; and (3) A sampling strategy was developed to determine the length of sampling needed in the Walsh Lake basin to accurately estimate reservoir reliability.

Initial estimates of streamflows into Walsh Lake were developed using four techniques including: use of a limited set of historical data, water balance computations, correlation with adjacent basins, and regional regressions. Historical flow data for Walsh Lake indicate a mean annual inflow volume of 15,300 acre-feet (1.89×10^7 m^3). However, the accuracy of this data is extremely suspect due to unrecorded diversions at the stream gage site and changes in the Walsh Lake drainage basin. Water balance computations indicate a potential mean annual runoff volume of 25,600 acre-feet (3.16×10^7 m^3). Streamflow data for the Issaquah Creek basin were used to derive lower bounds on inflow statistics; these computations result in a mean annual inflow volume of 12,600 acre-feet (1.55×10^7 m^3). Regional regressions for Western Washington indicate a mean annual inflow volume of 24,800 acre-feet (3.06×10^7 m^3). Of the four techniques, regional regression is the most appropriate for the Walsh Lake watershed and the value of 24,800 acre-feet is considered the most accurate of the four estimates.

Regionally estimated inflow statistics were used in computer simulations to derive preliminary estimates of reservoir reliability performance. Simulations using regional regression inflow statistics indicate that an annual storage volume of 20,000 acre-feet (2.47×10^7 m^3) is necessary for reliable operation of the reservoir during the entire 3-month low flow period. Simulations using lower bound reservoir inflow statistics derived from Issaquah Creek data indicate that the reservoir cannot be operated reliably for the entire low flow period. If operated during July and August only, preliminary simulations indicate a storage volume of 14,000 acre-feet (1.73×10^7 m^3) is sufficient for reliable reservoir operation. The above estimates of storage requirements are not based on actual reservoir inflow data and are only preliminary estimates.

Regionally estimated inflow statistics are not sufficiently accurate for design purposes. It is therefore recommended that streamflow data be collected at the reservoir site. Once collected, this streamflow data can be used in combination with data for a nearby stream with a longer streamflow record to derive an extended inflow record. The extended inflow sequence can then be used to size the reservoir and evaluate its reliability under proposed operational conditions. In this study, record extension was found to significantly improve the accuracy of estimates of reservoir reliability. Six years of data were found to be sufficient when used in combination with record extension.

Chapter 1

INTRODUCTION

A common problem encountered in evaluating the feasibility of water resource projects is the scarcity of hydrologic data. Knowledge of anticipated inflows is vital to the efficient sizing and operation of any water development project. Continuous information on streamflow is available only for large streams and for streams involved in specific projects. For small undeveloped streams there is commonly no reliable information on historical discharges. In such situations, the designer is faced with either constructing the project based on minimal data or installing stream gages and waiting until sufficient information has been collected to design a reliable facility. In the first case, there is the possibility of either overdesigning or underdesigning the project. In the second case, costs of data collection or project delay may outweigh the value of the information collected. This conflict suggests the following questions:

1. What procedures exist to estimate streamflow in basins with no reliable historical streamflow data?
2. How should estimates of stream flow be used to evaluate reservoir reliability?
3. How many years of data should be collected to improve the accuracy of initial inflow estimates?
4. How should the collected data be used to improve initial reliability estimates?

This report addresses these questions for the proposed addition at Walsh Lake for to the City of Seattle's water supply system.

An option for meeting increasing demands on the Seattle water supply system is to withdraw of more water from the Cedar River at the Landsburg diversion. Currently, withdrawals at Landsburg are constrained by instream flow requirements for the Cedar River at Renton. To maintain the Cedar River as a salmon habitat and spawning ground, minimum flow requirements were adopted in 1969 by the State of Washington in the Minimum Water Flows and Levels Act.

The project studied in this report is a proposed reservoir which could be used to supplement Cedar River flows below Landsburg.

Located north of Landsburg is Walsh Lake, a small lake resting on glacial till and peat deposits. An artificial channel, Walsh Lake Ditch, drains the Lake and empties into the Cedar River below Landsburg. In addition, Rock Creek flows just to the east of the Lake and enters the Cedar River above Landsburg. The proposed reservoir is an enlargement of Walsh Lake, created by damming Walsh Lake Ditch and Rock Creek and storing their flows in the depression surrounding Walsh Lake. Due to the underlying peat deposits in the region, the quality of water in the reservoir is not adequate for direct use as municipal water supply. However, during low flow months this water could be piped into the Cedar River below Landsburg to meet the instream flow requirements at Renton. This would allow the City of Seattle to increase withdrawals at the Landsburg diversion during the low flow periods of July, August, and September.

To evaluate this alternative, streamflow data are needed to determine if the proposed Walsh Lake expansion can provide sufficient seasonal storage to meet instream flow requirements in the Cedar River. Knowledge of the magnitude and variability of monthly inflows can be used to compute the reliability of the reservoir for a given capacity and operating procedure. Monthly flow sequences for Walsh Lake Ditch and Rock Creek can be routed through the reservoir, subject to withdrawals for instream flow requirements, to determine how often the reservoir would fail. This information can be used to estimate the required volume necessary to maintain reliability requirements. Since no usable data are currently available at the site, inflows can be estimated from stream gaging records for nearby streams. Regional information can be used to determine parameters for stochastic streamflow models. This will enable the development of synthetic flow traces that can be used to estimate the reliability of the reservoir under a wide range of conditions.

Estimates of monthly flows obtained from regional information are uncertain. To decrease this uncertainty in flow estimates, and therefore in projected reliability, it is necessary to install stream gages on Walsh Lake Ditch and Rock Creek. As streamflow data are collected for the Walsh Lake inflows, it can be correlated with concurrent flow data on streams with longer historical gaging records. This information can be used to develop relationships between the reservoir inflows and recorded flows in the long-record stream. The short inflow record can then be extended to the length of the longer record, thus adding to the value of the collected inflow data. Although inflow estimates will always be expected to improve as more data are collected, the relative improvement in the accuracy of inflow estimates may be insignificant relative to data collection costs. The number of years of data needed is a function of:

1. the variability of monthly inflows
2. the amount of correlation between the inflows and the flows in the long-record stream from which the inflow record will be extended
3. the net benefit of improved inflow estimates.

Several methods exist for extending short flow records from longer records. Each of these are evaluated in this report using different inflow record lengths to determine the best combination of record extension technique and record length.

The contents of this report are organized as follows. Chapter 2 gives a detailed physical description of the Walsh Lake drainage basin. Included in this chapter are relevant hydrologic, meteorologic, and geologic data. In Chapter 3 the various methods for estimation of streamflow parameters in ungaged catchments are discussed and applied to the Walsh Lake basin. In Chapter 4 these initial estimates are used to derive preliminary reservoir reliability values for a range of reservoir capacities and operating rules. Chapter 5 describes several of the methods available for extending short samples from correlated longer samples. Chapter 6 provides a description of computer simulation experiments for the determination of ideal sampling strategies. Chapter 7 summarizes the findings of this report and indicates potential areas of further research.

Chapter 2

DESCRIPTION OF THE WALSH LAKE DRAINAGE BASIN

Walsh Lake is located some 22 mi southeast of Seattle. Figure 1 illustrates the drainage area for the proposed Walsh Lake Reservoir. The area is owned by the City of Seattle and is regulated as part of the Cedar River Watershed. Past land use has included coal mining and some limited logging. Major topographic features in the region include Taylor Mountain, with a peak elevation of 2600 ft (793 m) above mean sea level (MSL), and Brew Hill, at an elevation of 2530 ft (771 m). Inflows to the proposed reservoir drain the steep southern slopes of Taylor Mountain and Brew Hill. Major inflows to Walsh Lake include Webster Creek and Hotel Creek. Webster Creek extends up from its mouth at Walsh Lake to the top of Taylor Mountain. Hotel Creek drains low elevation lands at the foot of the mountain. Walsh Lake is a glacial kettle lake with a water surface elevation of approximately 725 ft (221 m) and is drained by Walsh Lake Ditch. The ditch is an artificial channel with a designed trapezoidal cross section. The bottom of the ditch consists of an armored layer of cobbles and large pebbles. Commonly, Walsh Lake Ditch has a bottom width of 12 to 15 ft and side slopes of 40 degrees.

Rock Creek flows in a southwesterly direction and is fed by tributaries flowing off the southeast slope of Brew Hill. In its upper reaches, Rock Creek is a relatively steep, rocky channel. However, as it enters the lowlands surrounding Walsh Lake, the grade decreases. Here, the channel is characterized by meandering bends and point bar deposits. The channel bottom consists of fine silty-sand and occasional deposits of cobbles and large pebbles. Inflows to Rock Creek in the lowlands come off of peat bogs and are much more brackish than the Rock Creek streamflow.

The soils in the upland parts of the basin are thin, with relatively low organic content. In the lowlands surrounding Walsh Lake, the soils consist of sandy loams with a high organic content. In the marshy areas, peat deposits are predominant. Most of the area above Walsh Lake is forested; the land immediately surrounding the lake is a swamp, making the lake virtually inaccessible. Water draining from the Taylor mine sites is of poor quality and has been diverted out of the Cedar River watershed since 1975. A

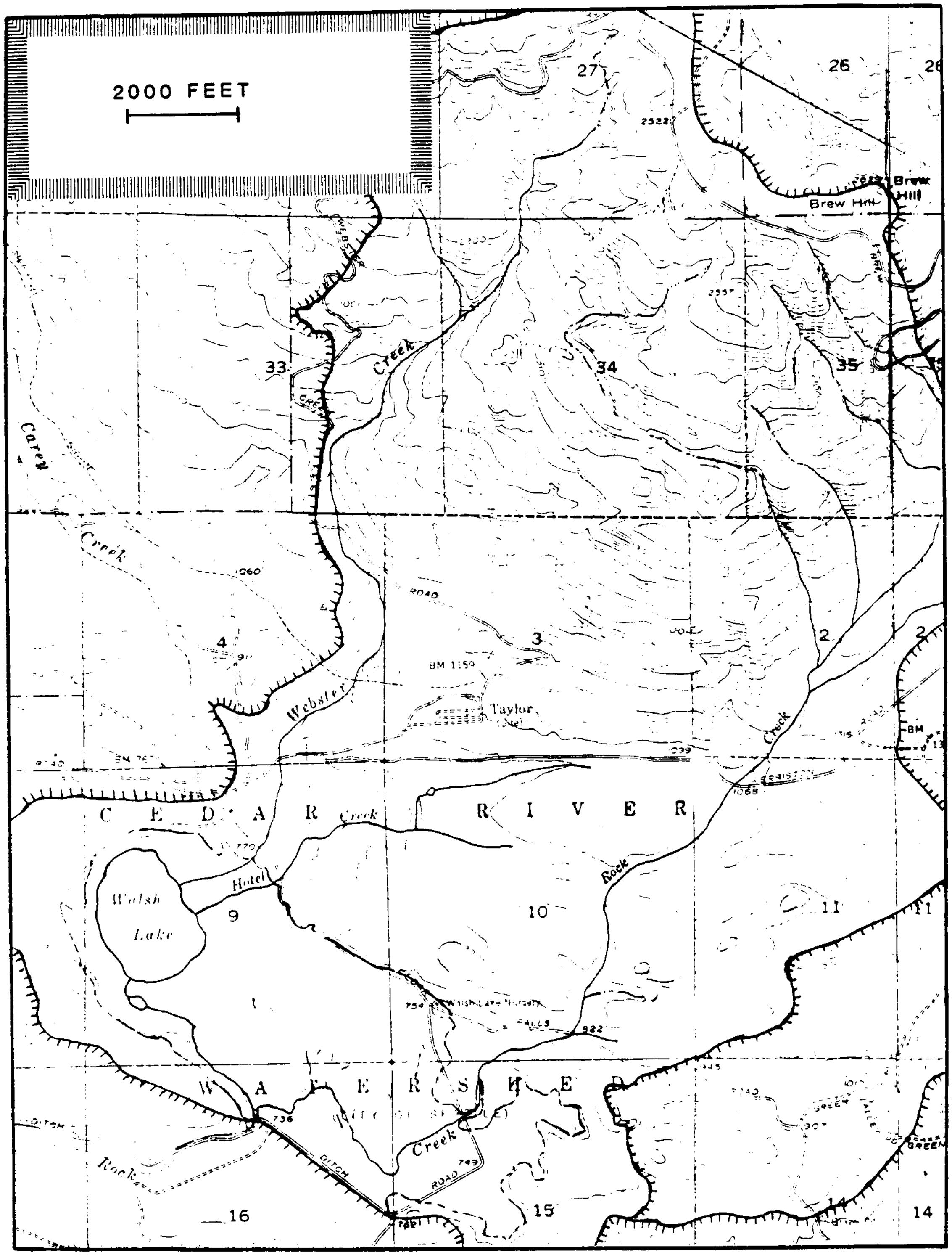


Figure 1. Map of Walsh Lake Basin

diversion ditch cuts across the site and carries the water draining from the mines across the basin divide and into the Issaquah Creek basin. The U.S. Geological Survey lists the diverted drainage area as 1.9 mi^2 or 4.9 km^2; drainage divides near the mines are difficult to delineate, and 1.9 mi^2 appears to be a high estimate of the actual diverted area. However, for all estimates of runoff in this report, it was assumed that all of the water from this area is diverted out of the Walsh Lake basin. This assumption should result in conservative estimates of reservoir inflows. In estimating inflows, a drainage area of 2.57 mi^2, or 6.66 km^2 was used for Walsh Lake Ditch at the dam site, while an area of 4.13 mi^2, or 10.7 km^2 was used for Rock Creek.

Geology of the Walsh Lake Region

The geology of the Walsh Lake Region was studied in detail in a U.S. Geological Survey Professional Paper by Vine (1969). This report describes the geology of the three U.S.G.S. quadrangles covering much of the Walsh Lake and Issaquah Creek basins. The paper was prepared as an evaluation of coal resources in the area and is summarized in the following paragraphs.

The bedrock geology of the region consists of a series of Eocene sedimentary formations. The lowest mapped formation in the area is the Raging River formation. This geologic unit is estimated to be up to 2900 ft in thickness and consists of massive outcrops of thick-bedded sandstone, siltstone, and conglomerate. These marine sedimentary rocks are generally fine grained and were formed from volcanic material. The top 160 ft of the formation consists of a silty claystone.

The Raging River formation is overlain by 200 ft of nonmarine sedimentary rocks. This sequence is referred to as the Tiger Mountain formation and is characterized by micaceous sandstone beds interstratified by siltstone and coal beds. The Tiger Mountain formation is exposed only in steep stream canyons. It is overlain by glacial deposits, except on Taylor Mountain where the Tukwila formation is exposed. The Tukwila formation is up to 6500 ft in thickness on Taylor Mountain. It consists of volcanic sandstones, siltstones, and shales which are highly resistant to erosion. The upper layers of this formation are subject to surface weathering, as indicated by the exposure of massive sapprolite blocks. On the southern part of Taylor

Mountain, the Tukwila unit is overlain by the Renton formation. These rocks consist of fine-to-medium grained sandstone, siltstone, shale, and coal. The Renton formation is approximately 2200 ft thick and differs from the older Eocene formations in that it is not made up of volcanic debris.

Quaternary deposits in the Walsh Lake region consist of various forms of Vashon drift. These glacial deposits were left by the last Pleistocene ice sheet to occupy the entire Puget Sound lowlands. This ice sheet was confined to the east by the Cascade range and to the west by the Olympic Mountains. Glacial deposits in the Walsh Lake area are classified as either till or terrace gravel and stratified drift. The Vashon till, in the form of ground moraine, forms a mantle 5.0 to 15.0 ft thick on the slopes of Taylor Mountain. This material is relatively compact and impermeable. The impermeability of these till deposits results in the occurrence of undrained depressions and small lakes.

Terrace gravel and stratified drift deposits occupy abandoned stream valleys. These valleys were formed by melt water channels during the retreat of the Vashon ice sheet. Stratified sand and gravel were deposited on the floors of these abandoned channels. The upper surfaces of these deposits form gravel terraces. This drift is well sorted, permeable, and relatively well drained. However, lakes and swamps occur in the melt water channels in areas where large ice blocks were buried in the ice sheet. These ice blocks melted during the retreat of the glaciers, leaving behind kettle lakes.

Other surficial deposits in the region include landslide debris and peat and swamp deposits. A large landslide covers a portion of the southeast side of Taylor Mountain. This landslide deposit is composed of glacial drift and rocks of the Tukwila formation. The slide is thought to have been an earthflow carrying large blocks of sandstone downslope. Peat and swamp deposits occur in the depression adjacent to Walsh Lake and are underlain by stratified drift. Vegetation fills the margin of the lake and peat bogs fill depressions that formerly contained other lakes. These swamp deposits are generally greater than a meter thick.

Coal in the Renton formation at the southern base of Taylor Mountain was mined at the Taylor mine sites. These mines were underground mines and had severe seepage problems. Often, continuous pumping of excess water was required. Tailings from these mines later led to water quality problems. As a result, runoff from the mine sites is diverted out of the Cedar River basin into the Issaquah Creek basin.

Figure 2 is a map of the surficial formations in the region, as mapped by the U.S. Geological Survey. The implications of these surficial formations on the hydrologic response of the basin are clear. Much of the area drained by Webster Creek is covered by glacial till deposits. These deposits tend to be relatively impermeable, indicating little groundwater movement in this area. In the upland areas on Taylor Mountain, there is potential for groundwater storage in the weathered sapprolite of the Tukwila formation. The land at the foot of Taylor Mountain, surrounding Walsh Lake and Rock Creek, is covered by permeable stratified drift. Here, one would suspect subsurface movement of water to be an important component of the hydrologic cycle. This is verified at road cuts below Walsh Lake where large amounts of seepage can be observed from the exposed drift deposits. Much of this seepage is water from Walsh Lake draining through peat deposits and moving downslope through the porous stratified drift. The peat deposits surrounding Walsh Lake also create water quality problems which preclude the direct use of the lake water for municipal water supply.

Hydrologic Regime

The climate of the Walsh Lake Region is characterized by frontal storms moving inland from the Pacific Ocean. These storms are heavily influenced by orographic lifting. Thus, annual rainfall increases with elevation. High precipitation in the upper elevation areas of the Cedar River watershed, coupled with the effects of snowmelt storage, make the Cedar River an ideal source of municipal water. The Walsh Lake basin is at a lower elevation than much of the Cedar River basin; snowmelt therefore does not play a critical role in the hydrology of the Walsh Lake inflows. Precipitation data for the region are available at Landsburg, elevation 640 ft and at Cedar Lake, elevation 1560 ft. Mean annual precipitation at Landsburg is 56 in, while 104 in/year is the average at Cedar Lake.

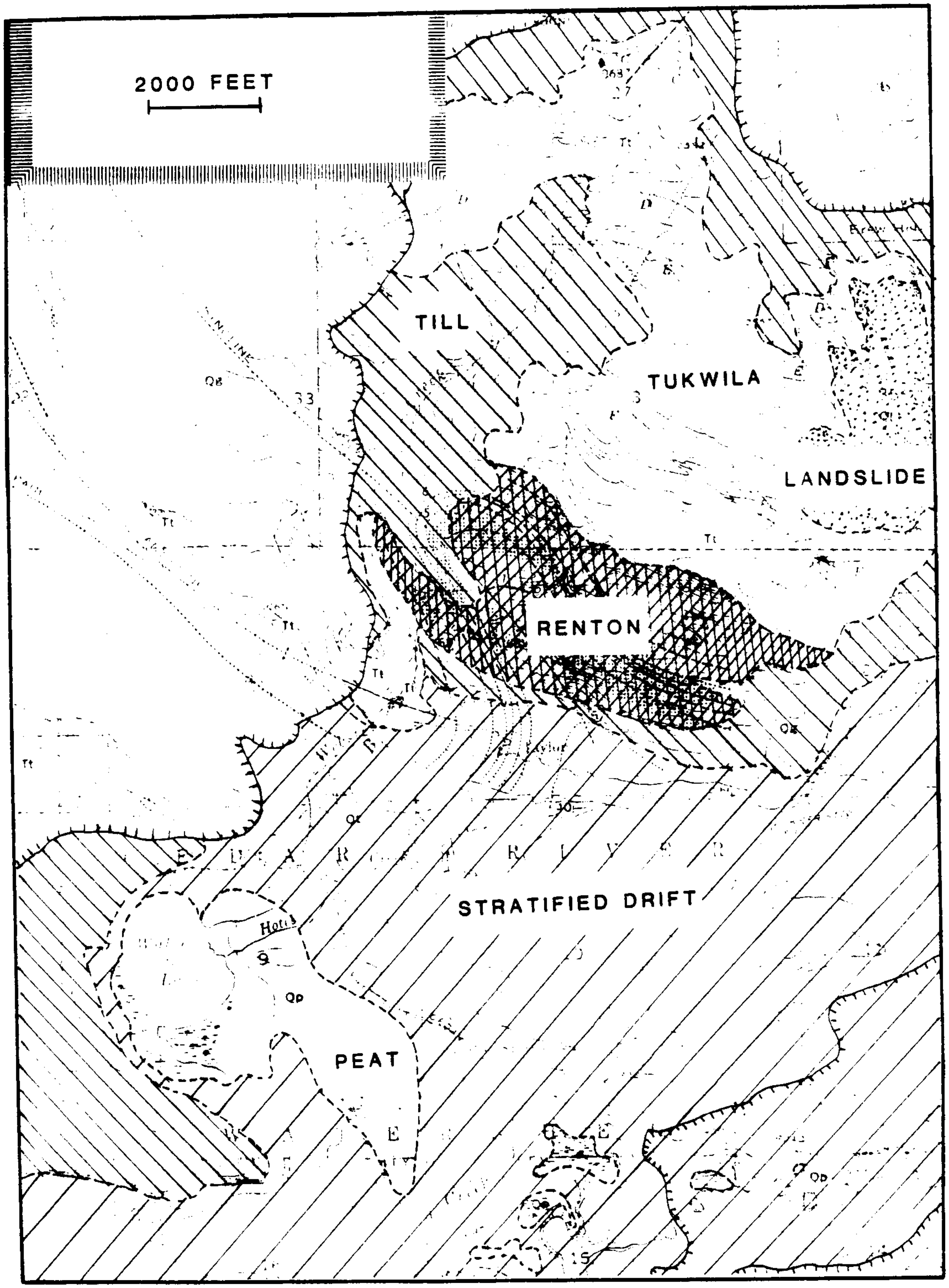


Figure 2. Surficial Geology of Walsh Lake Region

Assuming that annual precipitation increases linearly with elevation between these two stations, 89 in/yr of precipitation is estimated as the Walsh Lake mean. Evaporation data are scarce in western Washington; the Puyallup experimental station is the closest gage to Walsh Lake. Average monthly rainfall and pan evaporation rates are summarized in Table 1. The driest months of the year are July and August when monthly rainfall drops below 2 in. These are also the months with the highest evaporation rates, with pan evaporations of greater than 3.9 in per month. Wet months include November, December, and January when average precipitation rates exceed 7 inches per month.

Thus, the hydrologic regime of the Walsh Lake basin is characterized by high winter flows and low flow periods during the late summer when streams are fed principally by groundwater storage from the wet season.

Existing Flow Records for Walsh Lake Ditch and Rock Creek

From 1932 to 1948, a stream gage existed on Walsh Lake Ditch. During most of this period, Rock Creek flow was diverted from its natural channel into Walsh Lake Ditch. At times, however, the diversion dam was washed out and it is not known if all of the Rock Creek flow was diverted into the ditch continuously. From 1932 to 1944 streamflow in the ditch was measured only two to ten times per year; from July 1945 to October 1948 the gage was operated throughout the year and daily records are available. In 1945 the gage was moved 400 m downstream. From 1946 to 1948 there was regulation of flow at Walsh Lake by the City of Seattle. The gage drainage area was listed as 11 mi^2, or 28.5 km^2 and is presumed to include Rock Creek. Average discharge for the gaging period was 18,460 acre-feet/yr, or 0.722 m^3/s.

No record exists of the frequency of diversions between Rock Creek and Walsh Lake ditch during this period. Two diversion sites exist on these creeks. The downstream diversion consists of a wide rectangular channel through which Rock Creek flows were diverted into Walsh Lake Ditch. This diversion was operated at the discretion of the Seattle Water Department. Upstream of this point, a pipe 3.6 ft in diameter connects the two streams. During low flows, no flow occurs through the pipe. During higher flow periods, the direction of flow in the pipe is uncertain, although it appears that flow is generally from Walsh

Table 1

METEOROLOGICAL DATA, WALSH LAKE REGION

Month	Average Precip.*		Average Pan Evaporation**	
	cm	in.	cm	in.
Jan	18.4	7.24	0.00	0.00
Feb	14.4	5.66	0.00	0.00
Mar	13.2	5.20	1.14	0.45
Apr	11.3	4.46	5.84	2.30
May	8.41	3.31	9.09	3.58
Jun	7.98	3.14	11.8	4.64
Jul	3.38	1.33	14.1	5.56
Aug	4.62	1.82	11.8	4.63
Sep	8.43	3.32	6.91	2.72
Oct	14.0	5.53	3.23	1.27
Nov	18.9	7.45	1.01	0.40
Dec	20.1	7.91	0.18	0.07

*Precipitation data from continuous raingage at Landsburg

**Pan evaporation data from Puyallup experimental station

Lake Ditch to Rock Creek. No record exists indicating when this pipe was installed.

The 1932-1948 gaging record is suspect due to the unrecorded diversion of flow from Rock Creek. Assuming that conditions in the Walsh Lake basin have not changed, this record could be used to obtain lower bounds on reservoir inflows. Unrecorded diversions into Rock Creek could result only in underestimation of streamflow. However, significant changes have occurred in the drainage basin since 1948. In particular, the diversion of water from the Taylor mine sites results in a smaller drainage area and decreased streamflow. The change in drainage area can be accounted for by assuming that discharge per unit drainage area is constant thoughout the catchment. The recorded average annual flow for the 1932-1948 period is 25.5 cfs (2.32 cfs per mi^2). The flow lost from the 1.9 mi^2 of diverted area is then estimated as 4.38 cfs. Subtracting this figure from the recorded mean results in an estimated average annual inflow volume of 21.1 cfs (15,300 acre-feet). The assumptions behind this computation result in a conservative estimate of inflow volume; observed inflow volumes are expected to be greater.

Four instantaneous measurements of discharge were made on Rock Creek by the U.S. Geological Survey during 1976 and 1978. In addition, the City of Seattle Water Department made six measurements of discharge on both streams between 1982 and 1983. These instantaneous discharge measurements may be useful in characterizing discharge in low flow months (Table 2).

In summary, the streamflow data available for the Walsh Lake Inflows are not sufficiently reliable for use in the evaluation of reservoir performance. The 1932-1948 gaging record can be used to derive an estimate of a lower bound on the mean annual inflow volume. However, this flow record cannot be used to develop the monthly flow sequences necessary to model seasonal reservoir operation. To develop accurate flow sequences, it will be necessary to collect streamflow data for Walsh Lake Ditch and Rock Creek under current hydrologic conditions. This data can then be combined with regional data to develop reliable monthly flow sequences.

Table 2

SUMMARY OF RECENT DISCHARGE MEASUREMENTS ON WALSH LAKE INFLOWS

Location	Date	Discharge	
		m^3/s	cfs
Walsh Lake Ditch, at proposed reservoir site	8/5/82	.0714	2.52
	9/21/82	.0866	3.06
	2/16/83	.437	15.43
	4/26/83	.202	7.15
	5/31/83	.107	3.79
	6/27/83	.182	6.41
Rock Creek, at proposed reservoir site	8/5/82	.171	6.05
	9/21/82	.185	6.52
	2/16/83	.576	20.33
	4/26/83	.279	9.84
	5/31/83	.189	6.66
	6/8/83	.147	5.18
Rock Creek, near Mouth	7/26/76	.159	5.62
	8/17/76	.416	14.7
	10/3/77	.263	9.3
	11/28/77	1.36	47.9

Chapter 3

INITIAL ESTIMATION OF RESERVOIR INFLOWS

Approaches to estimation of streamflow parameters in data-poor situations fall into several categories. Regional approaches attempt to use streamflow data from nearby gaged streams to determine flow parameters for the basin of interest. This method assumes that, within a given region, relationships between runoff and basin characteristics are uniform. Rainfall-runoff models attempt to translate rainfall data into streamflow for the basin. This can be done by either using physically based conceptual models or by fitting response functions to rainfall-runoff data. Rainfall-runoff approaches generally require streamflow data for the site to use in calibration. Simple water balance computations use precipitation and evaporation data to estimate long term mean values of flow parameters. Such computations are usually limited to estimation of annual flows, but attempts have been made to compute monthly water balances (Thornthwaite and Mather, 1957).

Drainage Area Method

A direct method for estimating statistics for ungaged streams is to use streamflow and catchment data from a nearby gaged stream. The drainage basin for the two streams should have similar physical characteristics and should be subject to the same precipitation events. Flows for the gaged stream are adjusted by the ratio between the two drainage areas to obtain values for the ungaged stream. It is assumed that the gaged basin and the ungaged basin respond similarly to precipitation events; runoff produced per unit area will then be the same for the two catchments. Flows for the ungaged catchment are computed as (Hirsch, 1979):

$$Q_1 = Q_2(A_1/A_2) \tag{1}$$

where Q_1 is the observed flow volume in the gaged catchment; Q_1 is the computed flow volume for the ungaged catchment; A_1 is the drainage area for the ungaged catchment; and A_2 is the drainage area for the gaged basin.

Regional Regression Techniques

Regional techniques have been developed which attempt to use information from a large number of gaged catchments to estimate flows. The most common approach is to relate discharge parameters to measurable basin parameters for a set of representative gaged catchments. Generally, this is done using multiple regression techniques. Discharge parameters commonly used include mean discharge, discharge variance, lag-correlations, and flood discharges of specified frequency. Common basin parameters include drainage area, mean slope, land use coefficients, and elevation. Often, the logarithms of parameters are used in the regression analysis (Riggs, 1973), resulting in equations of the form:

$$Q = aA^{b}\,B^{c}\,C^{d}\,\ldots\ldots \tag{2}$$

a,b,c,d = constants
A,B,C = basin parameters
Q = streamflow parameter

To calculate parameters, a regression analysis is performed in which constants in the regression equation are chosen to minimize the sums of the squares of errors about the regression line (Haan, 1977).

Basin parameters used in multiple regression equations can be selected by one of two approaches. In the first approach, the modeler selects parameters which are thought to control runoff in the region. Multiple regression is then applied directly to the available data. If the parameters are incorrectly chosen, this approach can result in physically unreasonable regression equations (McCuen et al., 1979). This approach can also include basin parameters which do not add significantly to the predictive ability of the regression. Stepwise regression, the second approach, eliminates statistically insignificant basin parameters from the regression. Statistical significance is tested using an F-distributed statistic (Haan, 1977). The selection of parameters again depends upon the judgement of the modeler and it has been shown that this procedure can lead to biased coefficients (McCuen et al., 1979). With any ordinary least squares, multiple regression technique strong correlation between basin parameters results in instability in the estimated coefficients (Haan, 1977).

Attempts have been made to circumvent the problems associated with least squares multiple regression techniques. McCuen et al. (1979) compared multiple regression, stepwise regression, principal component analysis, polynomial regression, and constrained pattern search analysis. Pattern search methods and polynomial regression gave the best results. Haan and Allen (1972) compared principal component analysis and multiple regression for a set of thirteen catchments in Kentucky. Principal component analysis did not prove to be superior, although it was found to be computationally more difficult. Stedinger and Tasker (1984) compared ordinary, weighted, and generalized least squares regression techniques. Weighted and generalized least squares regressions proved to be more accurate than ordinary least squares in a series of Monte-Carlo tests. Generalized least squares is capable of accounting for cross correlation of concurrent flows between sites and was shown to be superior to weighted least squares regression when this cross correlation was 0.6 or greater.

Application of Regional Regressions to Western Washington

Regional techniques are less appropriate in areas where rainfall patterns are highly variable and cannot be well explained by the independent variables. In western Washington, orographic influences on frontal storms make regional approaches to streamflow estimation highly inaccurate. Upland catchments generally receive far more precipitation than do lowland catchments; however, elevation does not explain precipitation variation well. The distribution of glacial deposits, which govern the movement of subsurface water in the region, is highly variable and difficult to represent by parameters in regression equations. Regional approaches cannot supply more than a very cursory indication of runoff conditions in this area.

Regional analysis has been applied to the state of Washington in two studies by the U.S. Geological Survey. These studies evaluated the stream gaging program in the state to improve regional relationships. In the first report (Collings, 1971), regional relationships were judged adequate if the accuracy of regression equations were equivalent to the accuracy of estimates obtained from a 10-year flow record for minor streams, or the accuracy obtained from a 25-year record for principal streams. Collings divided the state into two regions: western Washington and eastern Washington. Regressions were performed using ordinary least-squares multiple regression. Flow parameters estimated included mean annual flow, mean monthly flows, standard deviations of flows, 7-day low flows, and flood peak flows. Basin parameters included drainage area, channel slope, percentage area of lakes, forest cover, mean elevation, mean annual precipitation, and minimum January temperature. Regression equations for monthly and annual flow means and standard deviations are listed in Table 3. Figure 3 compares the accuracy of regressions for various statistics to the goals described above. In all cases, the standard errors of the regressions were greater than the equivalent standard error of estimates obtained from a 10-year flow record.

These findings were verified in a second report by Moss and Hauschild (1978). In this report, a computer program called Network Analysis for Regional Information (NARI) was applied to the State of Washington stream gaging program. NARI uses Bayesian decision theory and Monte-Carlo experimentation to determine the number of stations needed to meet regional regression accuracy requirements (Moss and Karlinger, 1974). Moss and Hauschild divided the state into six regions, based upon hydrologic regime. The Walsh Lake basin lies in the Puget Sound lowland region, where the annual hydrograph is characterized by a single winter peak. Regressions were performed to determine annual flow means and standard deviations, mean annual peak flows, and flood flows of various frequencies. Only stations with a minimum of 15 years of data were included in the regressions and highly correlated variable pairs were removed from the equations when they did not add significantly to the quality of the regression. Moss and Hauschild concluded that the accuracies of the regional regressions were limited by the inappropriateness of the regression model for this region and not by the lack of regional information.

Water Balance Approaches to Runoff Estimation

The simplified water balance for a region is given by the following equation:

$$DS = P + I - E - O - GW \tag{3}$$

DS = change in storage during a time period
P = precipitation

Table 3

REGIONAL REGRESSION RELATIONSHIPS, WESTERN WASHINGTON (COLLINGS, 1971)

$$Q = aA^{b}\ S^{c}\ L^{d}\ E^{e}\ F^{f}\ P^{g}\ T^{h}$$

Q = flow parameter, cfs
A = drainage area, mi^2
S = channel slope, ft/mi
L = percentage area of lakes
E = average basin elevation, feet divided by 1000
F = percentage area forested
P = mean annual precipitation, in.
T = average minimum January temperature

			Exponents for Basin Parameters					
Parameter Q	a	b	c	d	e	f	g	h
Mean Annual FLOW	3.24E-3	0.99	0.00	0.00	0.12	0.14	1.23	0.33
Monthly Means:								
Jan	6.02E-6	1.02	0.00	-0.02	0.00	0.26	1.28	2.14
Feb	7.76E-6	1.02	0.00	-0.02	0.00	0.29	1.20	2.10
Mar	2.14E-5	1.06	0.00	-0.02	0.00	0.24	1.15	1.80
Apr	1.03E-3	0.99	0.00	0.00	0.24	0.24	1.09	0.77
May	0.629	0.98	0.00	0.03	0.45	0.00	1.04	-0.86
Jun	35.5	0.97	0.00	0.06	0.58	-0.35	1.00	-1.61
Jul	55.0	1.07	0.14	0.07	0.37	-0.73	1.09	-1.80
Aug	6.31	1.20	0.29	0.05	0.00	-0.87	1.17	-1.55
Sep	11.2E-2	1.02	0.00	0.00	0.38	-0.82	1.34	0.00
Oct	2.95E-4	0.99	0.00	0.00	0.26	-0.33	1.73	0.85
Nov	2.51E-5	0.98	0.00	-0.3	0.18	0.00	1.51	1.68
Dec	1.26E-5	1.01	0.00	-0.02	0.09	0.24	1.38	1.80
Annual Standard Deviation	2.19E-3	1.02	0.00	0.00	0.82	0.29	1.03	0.00
Monthly Standard Deviations:								
Jan	4.90E-6	1.00	0.00	0.00	0.00	0.31	1.36	1.80
Feb	2.95E-5	0.98	0.00	0.00	0.00	0.34	1.25	1.36
Mar	4.57E-6	1.02	0.00	0.00	0.00	0.53	1.07	1.74
Apr	1.78E-3	0.99	0.00	0.00	0.07	0.35	1.16	0.00
May	4.57E-2	0.97	0.00	0.00	0.31	0.25	0.98	-0.63
Jun	2.04	0.95	0.00	0.06	0.65	0.00	0.95	-1.42
Jul	5.25	0.95	0.00	0.08	0.64	0.00	1.10	-2.06
Aug	0.251	0.96	0.00	0.07	0.43	-0.76	1.46	-0.95
Sep	5.37E-4	0.94	0.00	0.00	0.32	-0.44	2.02	0.00
Oct	9.17E-6	0.98	0.00	0.00	0.31	0.32	1.67	0.93
Nov	1.35E-5	0.98	0.00	0.00	0.21	0.52	1.32	1.24
Dec	6.03E-5	1.02	0.00	0.00	0.14	0.49	1.12	1.07

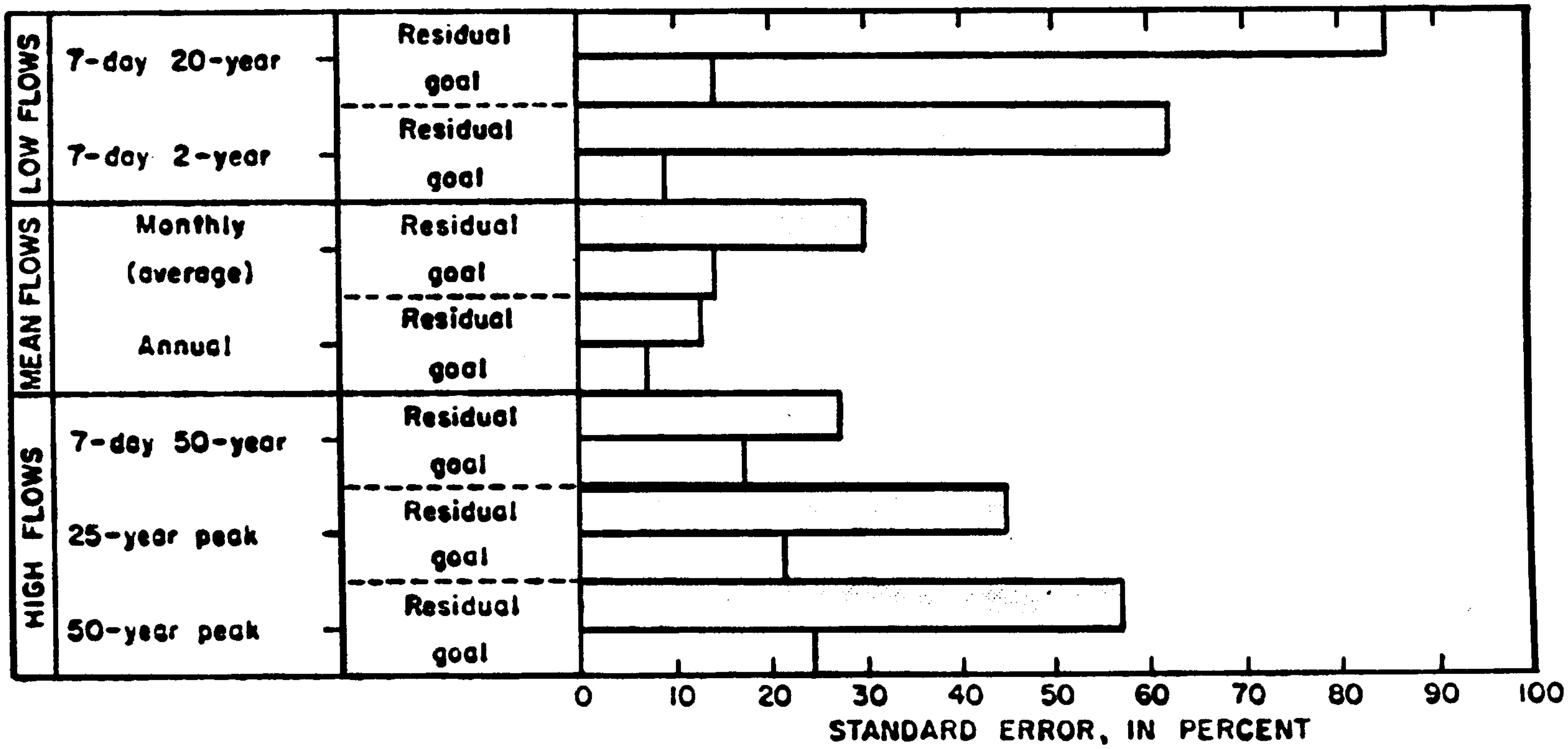


Figure 3. Accuracy of Western Washington Regional Regression (Collings, 1971)

E = evaporation
O = streamflow out of the basin
I = inflow to the basin
GW = subsurface seepage

Often this equation is used to determine the annual water balance for a region. For an annual balance, it is often assumed that subsurface seepage is negligible, and that there will be no change in storage from year to year. The equation can then be solved for streamflow, given precipitation and evaporation data. Because such techniques are applicable only when long term estimates of flow are sought, they cannot be used to accurately estimate flows on a time scale of interest in this research.

Inflow Estimation

The techniques described above were applied to the Walsh Lake basin to estimate flow parameters for Walsh Lake Ditch, Rock Creek, and the combined drainage area for the two streams. The simplified water balance is often used as an approximate computation of the amount of runoff that a catchment could potentially produce. Using an estimated mean annual precipitation depth of 89 in, and a mean evaporation rate of 25.6 in/yr, the annual water balance yields a mean annual inflow volume of 63.4 in, or 25,600 acre-feet.

The Drainage Area method was applied to the Issaquah Creek streamflow record for 1945 through 1964 to estimate annual and monthly flow parameters for Walsh Lake Ditch and Rock Creek. The Issaquah Creek basin has a drainage area of 27 mi^2, and lies at an average elevation of 940 ft. Estimated flow means, standard deviations, and drought statistics for Walsh Lake Ditch and Rock Creek are listed in Table 4. The estimated mean annual inflow volume for the proposed reservoir is 12,600 acre-feet (1.55×10^7 m^3) with a standard deviation of 2,600 acre-feet (3.21×10^6 m^3).

The regional regression relationships for Western Washington were applied to Walsh Lake Ditch and Rock Creek at the proposed reservoir site. These relationships were also applied to the Issaquah Creek basin. The Issaquah Creek basin is the nearest gaged basin in the region. It is used here to

Table 4

INFLOW STATISTICS ESTIMATED FROM ISSAQUAH CREEK RECORD

Parameter	Value of parameter (cfs)			
	Issaquah Creek	Walsh Lake Ditch	Rock Creek	Total Inflow
Monthly Means				
Jan	129	12.3	19.7	32.0
Feb	130	12.4	19.9	32.3
Mar	97.5	9.3	14.9	24.2
Apr	77.8	7.4	11.9	19.3
May	50.9	4.8	7.8	12.6
Jun	39.2	3.7	6.0	9.7
Jul	25.0	2.4	3.8	6.2
Aug	18.9	1.8	2.9	4.7
Sep	21.2	2.0	3.2	5.2
Oct	38.1	3.6	5.8	9.4
Nov	88.6	8.4	13.6	22.0
Dec	122	11.6	18.7	30.3
Monthly Standard Deviations				
Jan	46.3			11.5
Feb	48.0			12.0
Mar	38.8			9.7
Apr	22.7			5.6
May	19.5			4.8
Jun	16.6			4.1
Jul	7.6			1.9
Aug	4.2			1.0
Sep	7.6			1.9
Oct	21.4			5.3
Nov	43.8			10.8
Dec	47.0			11.5
Annual Mean	69.9	6.7	10.7	17.4
Annual Standard Deviations	14.8	1.4	2.3	3.7
20-year, 7-day low flow	11.0	1.1	1.7	2.8

evaluate the performance of the regressions in the Walsh Lake area. Measured basin parameters for each stream are listed in Table 5. Streamflow estimates derived from the regional regressions are listed in Tables 6 and 7. Also listed in Table 7 are the Issaquah Creek streamflow parameters computed from the observed record. For Issaquah Creek the monthly regressions developed by Collings (1971) overestimated mean monthly flows in the wet season by an average of 17 percent. Dry season flows were underestimated by 10.3 percent. This indicates that the regressions are not calculating sufficient basin storage of precipitation during wet months. The Collings regression did result in reasonably accurate estimates of mean annual flow. However, the 1978 regression by Moss and Hauschild overestimated the annual mean flow by 57 percent. It should be noted that the observed Issaquah means are based on 19 years of data and are themselves uncertain estimates of the population parameters. However, it can be concluded from comparisons with observed flows that regional estimates of monthly parameters can only be expected to be within 10 to 20 percent of the true value.

Comparison of Inflow Estimates

Although the Issaquah Creek basin is the most similar basin in the region, it receives on average 29 in/yr less precipitation than is estimated for Walsh Lake. The mean annual inflow volume computed from the Issaquah Creek record is 18 percent lower than the lower bound on Walsh Lake inflows derived from the 1946-1948 Walsh Lake Ditch streamflow record and 51 percent lower than the mean inflow volume computed from the annual water balance. Regional regression estimates of inflow statistics compare more favorably with observed inflows and the annual water balance. The inflow statistics computed from the Issaquah Creek record are therefore used only to derive conservative lower bounds on reservoir performance. Regional regression inflow statistics are thought to be more realistic; these are used in subsequent chapters to determine preliminary best estimates of reservoir performance parameters.

Table 5

MEASURED BASIN DATA FOR WALSH LAKE REGION

Basin:	Walsh Lake Ditch	Rock Creek	Total Drainage Basin	Issaquah Creek
Drainage Area				
(km^2):	6.66	10.7	17.4	69.9
(mi^2):	2.57	4.13	6.7	27.0
Percent Lakes:	4.59	0.01	1.77	0.08
Percent Forested:	88.1	98.5	94.5	91.0
Mean Annual Precip.				
(cm):	242.0	216.0	226.0	152.0
(in.):	95.4	85.0	89.0	60.0
Mean Elevation				
(m):	424.0	363.0	387.0	287.0
(ft/1000):	1.39	1.19	1.27	0.94
Min. January Temp.				
(^oC):	-1.1	-1.1	-1.1	-1.1
(^oF):	30.	30.	30.	30.
Channel Slope				
(%):	6.9	4.2	5.2	2.9
(ft/mi):	366	222	277	153

Table 6

ESTIMATED FLOW PARAMETERS, FROM REGIONAL REGRESSIONS (COLLINGS, 1971) WALSH LAKE DITCH AND ROCK CREEK

Parameter	Estimate for Walsh Lake Ditch (cfs)	Estimate for Rock Creek (cfs)	Total Inflow (cfs)
Mean Annual Flow	14.3	19.9	34.2
Standard Deviation of Annual Flows	3.1	4.0	7.1
Monthly Means:			
Jan	24.3	39.6	63.9
Feb	21.6	35.7	57.6
Mar	15.2	23.2	38.5
Apr	16.3	22.8	39.2
May	13.0	14.2	27.2
Jun	9.8	8.3	18.1
Jul	4.5	3.9	8.4
Aug.	2.6	2.3	4.9
Sep	3.8	4.5	8.3
Oct	8.9	10.8	19.7
Nov	19.0	27.1	46.3
Dec	23.5	37.1	60.6
Monthly Standard Deviations:			
Jan			27.5
Feb			24.9
Mar			16.1
Apr			10.7
May			9.3
Jun			8.5
Jul			4.9
Aug			1.6
Sep			4.1
Oct			11.6
Nov			24.8
Dec			23.4
10-yr, 7-day low flow	2.1	0.9	
20-yr, 7-day low flow	1.8	0.7	

Table 7

ESTIMATED AND OBSERVED FLOW PARAMETERS, ISSAQUAH CREEK

1. from 1971 regressions

Parameter	Estimate (cfs)	Observed Value (cfs)
Mean Annual Flow	79.8	69.9
Monthly Means:		
Jan	161.	129.
Feb	150.	130.
Mar	111.	97.5
Apr	93.2	77.7
May	59.7	50.9
Jun	37.4	39.2
Jul	21.9	25.0
Aug	15.2	18.9
Sep	18.9	21.2
Oct	36.7	38.1
Nov	99.2	88.6
Dec	141.	122.
20-year, 7-day low flow	6.82	11.0

2. from 1978 regression

Mean Annual Flow 110.cfs

Standard Deviation of Annual Flows 17.6 cfs

Chapter 4

PRELIMINARY EVALUATION OF RESERVOIR PERFORMANCE

Computer simulation experiments were conducted to derive preliminary estimates of the reliability of the proposed Walsh Lake Reservoir under a variety of conditions. These estimates of reliability are based upon inflow statistics determined in the preceding chapter. They are presented only to give an approximate indication of how Walsh Lake Reservoir may perform under actual operating conditions. More accurate estimates of reservoir reliability will be obtained from streamflow data collected at the reservoir site.

The Walsh Lake Reservoir was simulated using a mass balance model. Changes in reservoir storage were computed from estimated inflows and withdrawals needed to meet instream flow requirements. Operation of the reservoir was simulated using stochastic inflow sequences. Reliability was then computed as the percentage of years of simulation in which reservoir storage was insufficient to meet demand. The model was used to evaluate reservoir reliability for a range of reservoir capacities. Two operating rules were used in simulation. In addition, experiments were conducted using several estimates of inflow statistics. The mass balance model, development of stochastic flow sequences, and results of simulation experiments are described in detail in the remaining portions of this chapter.

Mass Balance Model of the Reservoir

To simulate the operation of Walsh Lake Reservoir, inflows were routed through the simulated reservoir, subject to withdrawals for instream flow requirements, to determine storage at the end of a given time period. Thus, the mass balance equation for the reservoir can be written as follows:

$$S_t = S_{t-1} + I_t - D_t \tag{4}$$

where S_t is the storage at the end of period t; I_t is the total inflow volume during time period t; and D_t is the draft during period t needed to meet instream flow requirements in the Cedar River. A monthly time step was chosen to simulate the reservoir because an annual time step was too large to model the seasonal fluctuations in flow volumes which govern the operation of the reservoir, and data were not available to model the reservoir on a time scale finer than a monthly scale.

The logic incorporated in the simulation model drafted Walsh Lake Reservoir only during low flow summer months. The natural Cedar River inflow between Landsburg and Renton is referred to as Inflow 3 in the operation of the Seattle Water Supply system. During the low flow periods, instream flow requirements in the Cedar River will be met by the sum of Inflow 3 and the draft from the Walsh Lake Reservoir. The draft from the reservoir in a given month, D_t, can therefore be determined as:

$$D_t = IFR_t - \text{Inflow } 3_t \tag{5}$$

where IFR_t is the normal instream flow requirements for the Cedar River during month t. Monthly instream flow requirements for the Cedar River at Renton are shown in Table 8.

Two operating policies are proposed for the Walsh Lake expansion. Ideally, IFR drafts would be made for the entire low flow period, July through September. However, the proposed reservoir may prove incapable of providing sufficient storage for the implementation of this policy. An alternative policy is to draft from the reservoir during July and August only. Drafting from the reservoir for this shorter period would reduce the stress on the reservoir. However, the benefits gained from increased withdrawals at Landsburg during September would be lost.

Development of Stochastic Inflow Sequences

To simulate operation of the reservoir, flow sequences for both the combined reservoir inflows and Inflow 3 are required. Since historical sequences are unavailable for the reservoir inflows, flow sequences were generated stochastically using estimated inflow statistics. The stochastic generation of streamflow sequences is a common procedure used in the simulation of hydrologic systems (Fiering and Jackson, 1971). Sequences of flows are computed which preserve the statistical properties of observed sequences. A probability distribution of flows is assumed. Flow sequences are generated by sampling from this distribution, taking care to preserve both the moments of

Table 8

INSTREAM FLOW REQUIREMENTS, CEDAR RIVER AT RENTON, WASHINGTON

Month	Normal Requirement (cfs)	Critical Requirement (cfs)
Jan	370	250
Feb	370	250
Mar	370	250
Apr	370	250
May	370	250
Jun	280	110
Jul	130	110
Aug	130	110
Sep	130	110
Oct	200	110
Nov	370	250
Dec	370	250

the distribution and correlations in time and space. Common probability distributions used in hydrology include the Normal distribution, the Three-Parameter Log Normal distribution (3-PLN), and the Gamma family of distributions. The 3-PLN and gammma distributions are used when data are thought to be significantly skewed.

In the simulation of the Walsh Lake Reservoir, several properties of the distribution of inflows were of particular interest. First, the means, standard deviations, and skew coefficients of monthly flows had to be preserved for both the reservoir inflows and Inflow 3 flows. Secondly, to account for the dependence of the flow in a given month on the flow in the previous month, it was important to maintain time-lag correlations. Lastly, it was necessary to preserve spatial correlations between reservoir inflows and Inflow 3 flows.

Two common methods for the generation of stochastic flow sequences are the Lag-one Markov model and the Auto-regressive Moving Average model (ARMA(1,1)). The Lag-one Markov model regresses flow in a given time period with flows in the previous period. To preserve the mean and variance of the observed flows, the annual model takes the following form (Fiering and Jackson, 1971):

$$Q_t = \bar{Q} + \rho(Q_{t-1} - \bar{Q}) + e_t \sigma (1-\rho^2)^{1/2} \qquad (6)$$

where Q_t is the flow in period t; $\bar{Q}$ is the mean flow; σ is the standard deviation of flows; ρ is the correlation coefficent between Q_t and Q_{t-1}; and e_t is a normal random variable with zero mean and unit variance. To model monthly flows, one must account for seasonality. The seasonal Markov model is of the following form:

$$Q_{ij} = \bar{Q}_j + \rho_j \frac{\sigma j}{\sigma_{j-1}} (Q_{i,j-1} - \bar{Q}_{j-1}) + e_{ij}\sigma_j (1-\rho_j^2)^{1/2} \qquad (7)$$

where the subscript i denotes the year and j denotes the month, $\bar{Q}_j$ is the mean flow in month j, σ_j is the standard deviation of flows in month j, and e_j is the correlation between the flow in the current month and the flow in the previous month (Fiering and Jackson, 1971). It should be noted that although the seasonal model does preserve moments of the flow distributions for each season, properties of the annual flow distribution are not preserved. Thus, the model does not account for correlation with the previous annual flow volume.

An approach to seasonal modeling which preserves annual correlations is to generate flow sequences at an annual scale and disaggregate them into seasonal flows. Annual flow sequences can be generated using the annual Markov model, an ARMA(1,1) model, or a number of other stochastic models. The ARMA(1,1) model has the advantage of being able to maintain some level of long-term persistence in the data. Long-term persistence has been observed in a variety of hydrologic time series. It is often quantified by the Hurst coefficient. The Hurst coefficient is not an explicit parameter in the ARMA(1,1) model; however, other more complex models do exist which maintain a specified level of persistence (Lettenmaier and Burges, 1977).

For smaller basins, such as the Walsh Lake drainage basin, annual correlations are generally not significant. In addition, the operation of the proposed reservoir will be governed by the seasonal nature of Inflow 3 flows. Flows in the current water year are not considered to affect storage in the following water year. Therefore, reservoir inflows and Inflow 3 were modeled using the seasonal Markov model. To preserve spatial correlations, the multisite Markov model proposed by Matalas (1967) was used. If X is used to denote the vector of flows normalized by subtracting the mean at each site, the generating equation for a single season becomes:

$$X_{t+1} = AX_t + Be_{t+1} \tag{9}$$

For two sites, A and B will be 2x2 coefficient matrices, and e_{t+1} will be a 2x1 matrix of zero mean, unit variance random numbers. If M_0 is used to represent the lag-zero covariance matrix and M_1 to represent the lag-one covariance matrix, A is defined by:

$$A = M_1 M_0^{-1} \tag{10}$$

B is any matrix which satisfies the equation:

$$BB^T = M_0 - AM_1^T \tag{11}$$

where the superscript T denotes the transpose of a matrix. Since BB^T will be symmetric, B can be computed directly by assuming a lower triangular matrix. When B is computed in this manner, it is possible for the matrix to become singular; in such cases the elements of B are adjusted slightly until a nonsingular matrix results. When accounting for seasonality, A and B matrices must be computed for each month. An alternative method for preserving spatial correlations is to generate total flows for a region and disaggregate to single sites; details for this procedure are given in Lane (1978) and Salas et al. (1980).

Flows in the Walsh Lake Region were observed to be significantly skewed. Sequences were therefore generated using a three parameter Log-Normal distribution. Log-Normal parameters were estimated from sample statistics by the method of moments (Burges et al., 1975). In addition, a simplification of the multisite Markov model, as suggested by Fiering and Jackson (1971), was implemented. In this procedure, only lag-one autocorrelations and lag-zero cross correlations were explicitly preserved. Lag-one cross correlations were then determined by the Markov process, rather than by observed model parameters. This simplification reduced the number of parameters to be estimated from regional data and avoided some of the problems with singular B matrices.

Inputs for the Stochastic Streamflow Model

To estimate parameters for the multisite model, monthly means, standard deviations, skew coefficients, lag-one correlations, and spatial correlations were needed. Statistics for Inflow 3 were estimated from 48 years of flow data adjusted by the U. S. Army Corps of Engineers for the effects of reservoir storage. Means, standard deviations, skew coefficients, and lag-one correlations for Inflow 3 are listed in Table 9. Means and standard deviations for the Walsh Lake inflows were estimated using the regional techniques discussed in the previous chapter. Estimates derived from the

Table 9

HISTORICAL FLOW STATISTICS FOR CEDAR INFLOW 3

	Monthly Mean 10^7 m^3	(1000 acre-ft)	Standard Deviation 10^7 m^3	(1000 acre-ft)	Skew Coefficient	Lag-one Correlation
Oct	.847	6.87	.403	3.27	1.3	.499
Nov	1.22	9.93	.525	4.26	0.46	.520
Dec	1.87	15.2	.583	4.73	0.18	.305
Jan	1.85	15.0	.579	4.69	0.10	.240
Feb	1.53	12.4	.523	4.24	0.28	.225
Mar	1.52	12.3	.440	3.57	0.81	.264
Apr	1.27	10.3	.343	2.78	0.46	.262
May	.965	7.82	.232	1.88	1.07	.232
Jun	.746	6.05	.204	1.65	0.65	.587
Jul	.556	4.51	.144	1.17	0.46	.892
Aug	.411	3.33	.095	0.77	0.56	.726
Sep	.517	4.19	.239	1.94	1.91	.303

Issaquah Creek record are considered to be lower bounds on the magnitude of inflows. Reservoir reliabilities computed from these statistics are therefore expected to be extremely conservative. Reliabilities estimated using regional regression means and standard deviations given in Table 6 of Chapter 3 are expected to be more realistic.

Higher order statistics, including skews and correlations, cannot be estimated using standard regional techniques. These parameters were derived from streamflow records for two rainfall-dominated catchments in western Washington. These streams, the North River and the Chehalis River, do not have large snowmelt components of runoff in the spring and have been gaged for a sufficiently long period for reasonable estimation of higher statistical moments. Population statistics derived for the Walsh Lake inflows from these records are listed in Table 10. While it is recognized that these statistics may not reflect the actual values for the Walsh Lake basin, it is hoped that they give a reasonable structure for preliminary modeling of reservoir operation.

Results of Simulation

Streamflow sequences computed from the multisite model were used in conjunction with the mass-balance reservoir storage model to simulate operation of the reservoir; thus, the expected reliability of the reservoir could be estimated. Reliability was defined as the percentage of simulated years in which the reservoir did not fail. Operation was simulated for storage capacities ranging from 11,000 to 20,000 acre-feet (1.36×10^7 to $2.47 \times 10^7 \ m^3$).

Figure 4 shows the results of simulation using Issaquah Creek means and standard deviations. When operated during July, August, and September, the reservoir was found to be extremely unreliable for all reservoir sizes. A maximum reliability of 25.6 percent was computed for capacities greater than 19,000 acre feet ($2.34 \times 10^7 \ m^3$). Thus, reliability was limited more by inflow volume than the size of the reservoir. The reservoir did not consistently refill during the wet season. When operated in July and August

Table 10

POPULATION CORRELATIONS AND SKEWS FOR WALSH LAKE INFLOWS

Month	Lag-one Correlation	Skew Coefficient	Lag-zero Cross Correlation
Oct	.270	1.07	.840
Nov	.430	1.14	.919
Dec	.300	1.47	.890
Jan	.170	0.95	.791
Feb	.100	0.90	.754
Mar	.160	0.90	.880
Apr	.260	0.94	.860
May	.230	1.35	.794
Jun	.250	1.80	.791
Jul	.580	0.82	.916
Aug	.570	0.70	.857
Sep	.300	1.74	.913

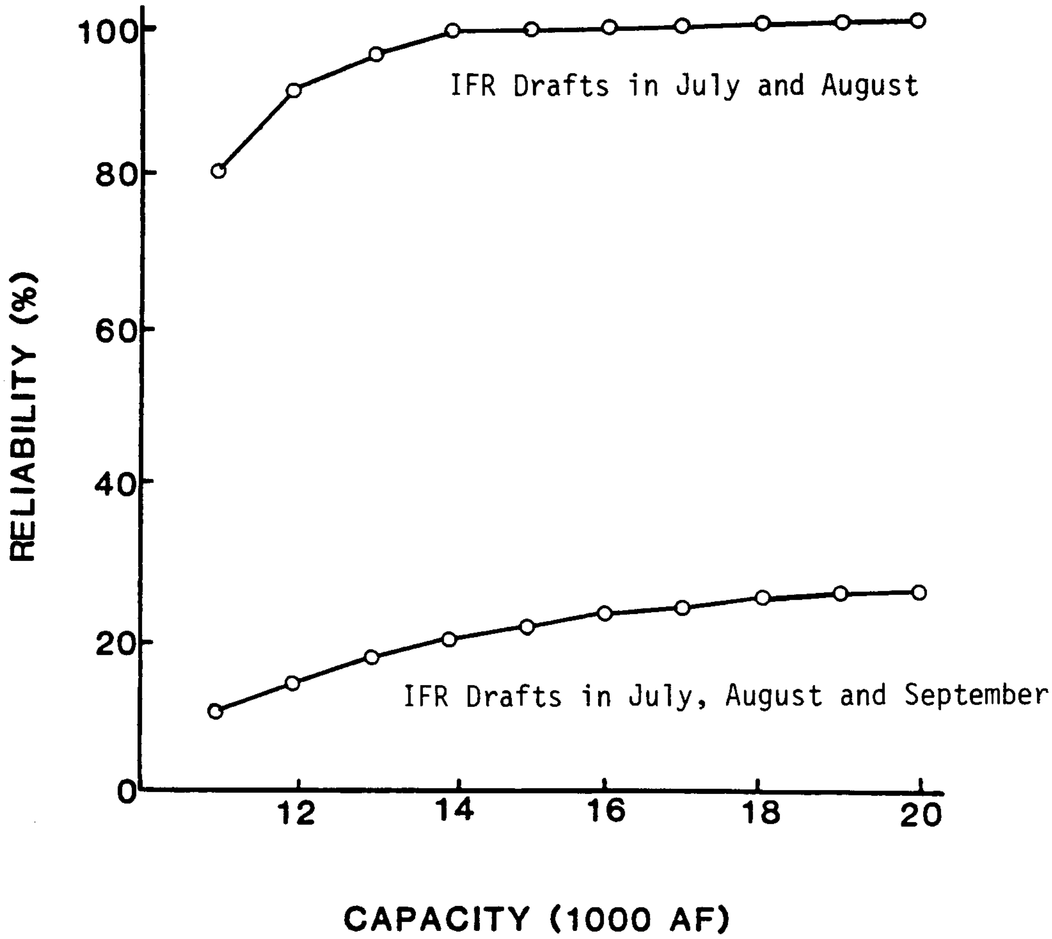


Figure 4. Reliability - Capacitance Curves, Issaquah Creek Inflow Estimates

only, the reservoir was found to be consistently reliable for reservoir sizes greater than 13,000 acre-feet (1.60×10^7 m^3).

Simulation using the higher regional regression inflow statistics gave similar results at low capacities (Figure 5). However, when operated in July, August, and September, reservoir reliability increased more rapidly with reservoir size. Reliabilities greater than 95 percent were obtained for reservoir sizes of 20,000 acre-feet or greater. Again, when operated in July and August only the reservoir proved to be much more reliable, with 99 percent reliability obtained at a capacity of 13,000 acre-feet.

It can be inferred from Figure 5 that when regional regression inflow estimates are used, operation in all three low flow months was limited by storage capacity and not by reservoir inflows. This was verified by modifying the model to require storage at the beginning of the dry season to always equal capacity. This modification eliminated the effects of variable wet season inflows. Figure 6 shows how reliability increased with capacity under these conditions, and verifies that reliability is indeed limited by capacity. To reach a reliability of 95 percent, a storage capacity of at least 20,000 acre-feet (2.47×10^7 m^3) will be needed to operate the reservoir during the entire dry season, regardless of inflow volume.

In summary, the following conclusions about the potential reliability of the Walsh Lake Reservoir can be reached from these simulation experiments:

1. Conservative estimates of inflow statistics, derived from Issaquah Creek data, indicate that the reservoir does not refill consistently and cannot be operated reliably for the entire low flow period.
2. Regional regression estimates of inflow statistics indicate that the reservoir does refill consistently between low flow periods.
3. If inflows are sufficient to fill the reservoir, a storage capacity of 20,000 acre-feet (2.47×10^7 m^3) is necessary for reliable operation of the Walsh Lake Reservoir during the entire 3-month low flow period.
4. If the Walsh Lake Reservoir is operated only during July and August, it is at least 99 percent reliable for storage capacities of 14,000 acre-feet (1.73×10^7 m^3) or greater.

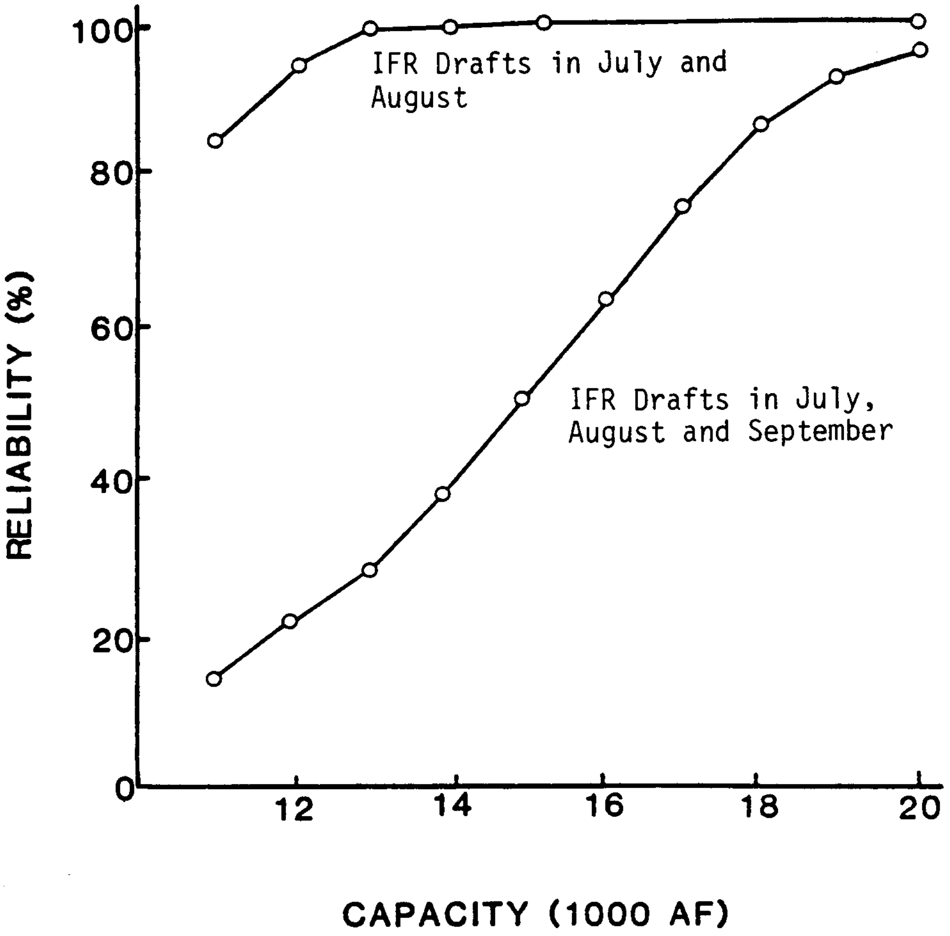


Figure 5. Reliability - Capacity Curves, Regional Regression Inflows

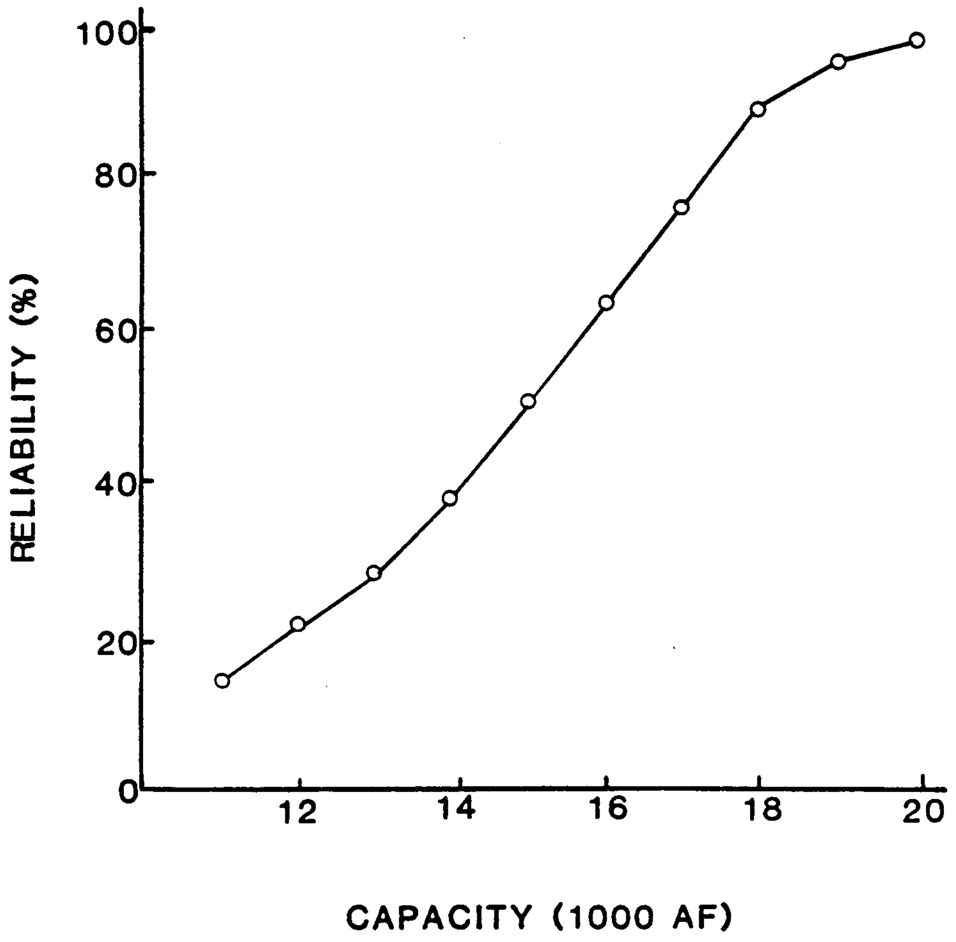


Figure 6. Reliability - Capacity Curve, Reservoir Full

Chapter 5

REVIEW OF RECORD EXTENSION TECHNIQUES

The initial estimates of flow statistics and reliability developed in the preceding two chapters were determined without the use of any at-site flow measurements. Thus, these estimates can only be expected to give an approximate idea of how the reservoir may perform under actual operating conditions. To gain improved understanding of reservoir performance, it will be necessary to monitor streamflow at the sites of interest. It remains to be determined how many years of data should be collected. Streamflow data collection involves costs which accumulate as the number of years of data collection increases. It is therefore desirable to gage the reservoir inflows for the minimum number of years necessary to accurately characterize reservoir performance. One way to augment the information gained from direct flow measurements is to correlate streamflow data with concurrent flow measurements at other sites. A relationship can then be developed between flows at the site of interest and flows at sites where more historical streamflow data are available. Thus, a very short streamflow record can be extended to the length of a record for a stream that has been gaged for a longer period of time. The accuracy of the extended data will depend upon the correlation between flows at the two sites and the length of the short record.

Two families of techniques are described here: linear regression and maintenance of variance techniques. Linear regression techniques are based on least-squares regression. These techniques are designed to minimize the error between observed and predicted data. Maintenance of Variance Extension (MOVE) techniques are designed to preserve specific characteristics of the observed sequence. Each model type can be used for either seasonal or nonseasonal flows; in the latter case, one assumes that the relationship between flows at the various sites does not vary seasonally.

In describing the available record extension techniques, the notation of Matalas and Jacobs (1964) will be used. N_1 refers to the number of years of data available at the short-record site of interest. N_2 refers to the additional years of data available at a long-record site; the total length of record at this site will be $N_1 + N_2$. R is the sample correlation coefficient between flows at the two sites. The observed flows at the short-record site will be represented by Y. Y_t will be used to denote the flow estimated from record extension during time period t. The overall series at the long-record site is denoted by X; X_1 are the observed flows for N_1 years and X_2 are flows for N_2 years. The sample mean of a series will be represented by M( ). The sample standard deviation will be denoted by S( ).

Linear Regression Techniques

The most common method for extending streamflow records is ordinary least-squares regression. The relationship between short-record flows and long-record flows is then determined by minimizing the sum of the squares of the errors about the regression relationship for the observed flow data (Fiering, 1963). Thus, the relationship between Y_t and X_t for two sites becomes:

$$Y_t = M(Y) + R\frac{S(Y)}{S(X_1)}(X_t - M(X_1)) \tag{12}$$

If the relationship between Y and X appears to be geometric rather than linear, the above equation can be applied to the log-transformed data, whereby means, variances, and correlations are computed in log-space (Hirsch, 1979).

Multiple least-squares regression techniques can be used to extend a short record from longer records at several sites. Alley and Burns (1983) developed a technique whereby a different base station is used during different periods to extend short record flows. The station used at a particular time is chosen based on the standard error of regression for the various base stations. However, extension from multiple sites is generally difficult to implement in regions where the spatial variability of runoff conditions is important.

Linear regression gives an unbiased estimate of the mean of the extended series. However, Matalas and Jacobs (1964) show that the estimate of the variance is biased downwards. For some applications, this underestimation of variance in the extended flow series may not be a serious drawback. However, in the study of reservoir performance extreme events are of particular interest. Reduction of variance in extended sequences can lead to overestimation of reservoir reliability (Hirsch, 1982).

Matalas and Jacobs (1964) developed a regression technique which eliminates problems associated with variance reduction. In this procedure, an independent random noise term is added to the regression equation such that the expected variance of the extended sequence is unbiased. The extension equation then becomes:

$$Y_t = M(Y) + R\frac{S(Y)}{S(X_1)}(X_t - M(X_1)) + \alpha(1-R^2)^{1/2}S(Y)e_t \tag{13}$$

where

$$\alpha^2 = \frac{N_2(N_1-4)(N_1-1)}{(N_2-1)(N_1-3)(N_1-2)} \tag{14}$$

with e_t defined as a normally distributed random variable with zero mean and unit variance. While this equation does give unbiased flows, use in practice is limited by the fact that it does not produce a unique series of extended flows. Because of the random noise term, each use of this equation results in a different flow sequence. The equation can be used to derive improved estimates of the mean and variance at the short record site (Matalas and Jacobs, 1964). The improved estimate of the mean is given by:

$$\hat{M}(Y) = M(Y) + \frac{N_2}{N_1+N_2} R\frac{S(Y)}{S(X_2)} (M(X_2) - M(X_1)) \tag{15}$$

The variance is estimated by:

$$\hat{S}^2(Y) = \frac{1}{N_1+N_2-1} (N_1-1)S^2(Y) + (N_2-1)\, R^2 \frac{S^2(Y)}{s^2(X_1)} S^2(X_2) \tag{16}$$

$$+ (N_2-1)\alpha^2(1-R^2)S^2(Y) + \frac{N_1N_2}{(N_1+N_2)} R^2 \frac{S^2(Y)}{S^2(X_1)} (M(X_2) - M(X_1))^2$$

These are commonly referred to as the Matalas-Jacobs estimates of the mean and variance. In a series of Monte-Carlo experiments, Hirsch (1982) showed that the Matalas-Jacobs estimates are indeed unbiased.

The Matalas-Jacobs estimates of the mean and variance, while unbiased, are not always the best estimates of the population parameters. For low values of N_1 and ρ, where ρ is the population correlation coefficient between sites, the sample estimate of means and variances are superior. $\hat{M}(Y)$ will have a smaller variance than does $M(Y)$, and therefore be a better estimate, if:

$$\rho^2 > (N_1-1)^{-1} \tag{17}$$

Maintenance of Variance Extension Techniques

MOVE techniques are similar to linear regression extension in that they involve linear relationships between flows at the site of interest and flows at other sites. However, MOVE techniques are generally designed to preserve specific characteristics of the observed flow distribution. They do not minimize the errors between observed and predicted flows. MOVE techniques used in hydrologic engineering include MOVE.1 and MOVE.2 (Hirsch, 1982), and MOVE.3 and MOVE.4 (Vogel and Stedinger, 1984). Each of these techniques preserves the mean and variance in the extended sequence; they differ in the estimates of means and variances to be preserved.

MOVE.1 has been described by Hirsch (1982). MOVE.1 maintains the observed sample mean and variance for the short-record station. The extension equation is:

$$Y_t = M(Y) + \frac{S(Y)}{S(X_1)} (X_t - M(X_1)) \tag{19}$$

When using this equation, the means of extended sequences will be unbiased. The expected variance of extended flows was shown by Hirsch (1982) to be:

$$E[S^2(Y)] = \frac{\sigma_y^2}{N_1+N_2-1} \left[N_1-1 + N_2\left(\frac{N_1-1-2\rho^2}{N_1-3}\right) \right] \tag{20}$$

where σ_y^2 is the population variance of flows, and ρ is the population correlation coefficient between the two sites. Numerically, this equation indicates that the variance for MOVE.1 will be biased slightly upward. The magnitude of this bias, however, is much lower than that of linear regression. Upward bias in the variance will result in slightly conservative estimates of reservoir reliability and is therefore not as undesirable as is downward bias.

A second technique introduced by Hirsch (1982) is MOVE.2. This technique preserves the Matalas-Jacobs estimates of the mean and variance, $\hat{M}(Y)$ and $\hat{S}^2(Y)$. The extension equation then becomes:

$$Y_t = \hat{M}(Y) + \frac{\hat{S}(Y)}{S(X)} (X_t - M(X)) \tag{21}$$

Note that M(X) and S(X) are the mean and standard deviation of the entire long-record sequence. Analytical expressions for bias of the mean and variance for MOVE.2 have not as yet been derived. However, Hirsch found that in Monte-Carlo experiments consisting of 2000 trials with various combinations of N_1, N_2, and ρ that the mean of extended sequences was unbiased. Variances were found to be only slightly biased. The advantages of this technique over MOVE.1 lie in the statistics it preserves. The Matalas-Jacobs estimates of the mean and variance take advantage of historical information at other sites; they are therefore better estimates of the mean and variance than are the estimates obtained from the observed sequence alone, given that correlations between sites are sufficiently strong.

Hirsch (1982) conducted Monte-Carlo experiments designed to evaluate the effects of bias in linear regression, MOVE.1, and MOVE.2 extensions. Means, variances, and order statistics of randomly generated sequences were estimated and compared to known population values. Order statistics included the first, second, and fifth lowest flow in a given sequence. Linear regression was found to be biased for all statistics but the mean. For MOVE.1, the variance was biased upwards, and the order statistics were biased downwards. In general, MOVE.2 proved to be unbiased. In these experiments, the accuracy of extension techniques was measured by the root mean square error (rmse). The rmse is a measure of the deviation between estimates of a parameter and the true value of the parameter, and is computed by the following equation:

$$\text{rmse} = \left[\frac{1}{N} \sum_{i=1}^{N} (X_i - T)^2\right]^{1/2} \tag{22}$$

where N is the number of samples, X_i is the estimate of the parameter derived from sample i, and T is the true value of the parameter. A high rmse indicates a high level of inaccuracy in estimates of the parameter of interest.

MOVE.3 and MOVE.4 were introduced by Vogel and Stedinger (1984). MOVE.3 is similar to MOVE.2 in that it is designed to preserve the Matalas-Jacobs estimates of the mean and variance. However, while MOVE.2 preserves these statistics for the extended flows alone, MOVE.3 preserves $\hat{M}(Y)$ and $\hat{S}^2(Y)$ for the entire flow sequence, including both the observed flows and the extended flows. The extension equation becomes:

$$Y_t = a + b(X_t - M(X_2)) \tag{23}$$

where:

$$a = \frac{(N_1+N_2)\hat{M}(Y) - N_1M(Y)}{N_2} \tag{24}$$

and

$$b^2 = \frac{1}{(N_2-1)S^2(X_2)} (N_1+N_2-1)\hat{S}^2(Y) - (N_2-1)S^2(Y) - N_1(\hat{M}(Y)-M(Y))^2$$
$$- N_2(a - \hat{M}(Y))^2 \tag{25}$$

Monte-Carlo experiments conducted by Vogel and Stedinger indicated that MOVE.2 and MOVE.3 were virtually indistinguishable in terms of the rmse between sample estimates and population statistics.

As noted previously, the Matalas-Jacobs estimates of extended means and variances are not necessarily the estimates with the lowest variances. Vogel and Stedinger (1984) developed minimum variance estimators which result in improved estimates of the mean and variance regardless of the strength of cross-correlation. The minimum variance estimator of the mean is given by:

$$M^*(Y) = (1-\theta_1)M(Y) + \theta_1\hat{M}(Y) \tag{26}$$

where:

$$\theta_1 = \frac{(N_1-3)R^2}{(N_1-4)R^2+1} \tag{27}$$

The minimum variance estimator of the variance is:

$$S^{*2}(Y) = (1-\theta_2)S^2(Y) + \theta_2\hat{S}^2(Y) \tag{28}$$

The expresion for θ_2 is difficult to solve analytically. However, Vogel and Stedinger derived the following approximate expression for θ_2:

$$\theta_2 = \frac{(N_1-4)R^2}{(N_1-8.5)R^2+4.5} \tag{29}$$

This approximation was found to be adequate in a series of Monte-Carlo tests. Monte-Carlo tests also showed that the minimum variance estimators had rmse's between predicted and population values equal to or less than those of the Matalas-Jacobs estimators (Vogel and Stedinger, 1984).

This technique, denoted as MOVE.4, is consequently designed to preserve the minimum variance estimators of the mean and variance. The extension equation is then derived by replacing the Matalas-Jacobs estimators in the expression for MOVE.3 with the minimum variance estimators. It should be noted that θ_2 becomes undefined for N_1 less than nine. Thus, MOVE.4 is only appropriate when the length of record is nine or greater. It is not limited, however, by the strength of correlation between flows at the two sites.

Seasonal Versus Nonseasonal Extension

Each of the extension methods described above can be applied to seasonal or nonseasonal flows. The choice between a seasonal or nonseasonal model is not clear; each alternative involves significant trade-offs. If a seasonal model is used, the variation in correlation between sites over the year can be modeled. Seasonal effects such as snowmelt runoff and convective storms can then be reflected in the relationship between flows at the various sites. When a nonseasonal extension equation is used, the flow in each season is treated as part of a single stationary time series. The advantage of this approach is that it increases the number of data points from which a relationship can be developed. For instance, if a site were gaged for N years, a seasonal model could only be developed from N data points. A different extension equation would be used for each season. If monthly flows were being extended, a nonseasonal model would be estimated from 12 x N data points. Only one extension equation would be used for the entire water year. Alley and Burns (1983) investigated the applicability of the two types of models in several Virginia catchments. Models were chosen based on the standard errors of the various regressions. Neither seasonal or nonseasonal models were found to be superior in all instances and no reliable guidelines were developed for the choice of a model.

Summary of Extension Techniques

Record extension techniques available to the hydrologic engineer include Linear Regression, Linear Regression plus noise, MOVE.1, MOVE.2, MOVE.3, and MOVE.4. Ordinary least squares regression has been found to be significantly biased, whereas regression plus noise is unbiased but does not produce a unique sequence of flows. Regression can also result in inferior estimates of means and variances if between-site correlations are not sufficiently strong. Each of the maintenance of variance (MOVE) techniques preserve different estimates of the mean and variance of flows at a site. In general, the variance of MOVE.1 flows will be biased slightly upwards. MOVE.2, MOVE.3, and MOVE.4 are all unbiased. In terms of rsme's, MOVE.2 has been found to be superior to MOVE.1; virtually no difference has been found between the performances of MOVE.2 and MOVE.3. Each of the above techniques can be used seasonally or nonseasonally, depending on the variation in cross correlations over the water year. None of the extension models has yet been found in the literature to be superior in all situations. As a result, computer simulation experiments as described in the following chapter were conducted to determine the best technique for evaluating Walsh Lake inflows.

Chapter 6

EVALUATION OF SAMPLING STRATEGIES

Given that sampled streamflows can be used to extend a short record to the length of other records in a region, the length of time data should be collected remains to be determined. The length will be a function of the level of accuracy desired, the probability distributions of flows at the various sites, and the record extension technique used. Attempts have been made to evaluate the worth of additional streamflow data; these are summarized in the first section of this chapter. The worth of streamflow data varies from site to site, therefore, computer simulation experiments were conducted to evaluate changes in information gained over time at Walsh Lake. These experiments were executed for several record extension techniques and are discussed in the latter section of this chapter.

Review of Worth of Data Literature

Information, in the form of streamflow data, can be said to have worth in that as more is collected, the likelihood of making a correct decision increases. It should be noted that this is true only in expectation; it is quite possible that more data could lead to an inferior decision. Ideally, the worth of data can be measured in terms of economic benefits; however, the economic evaluation of the worth of data has proven to be a very difficult problem. Although evaluations have been attempted, the worth of data has generally been measured in terms of measures of accuracy and sample variability. Dawdy (1979) reviews past research on this topic and discusses the philosophy behind such evaluations.

Two basic approaches have been used to evaluate the worth of additional hydrologic data. In the first approach, long streamflow records or stochastic streamflow traces are used to determine the true value of a parameter to be estimated. The long record is then broken up into smaller samples, from which estimates of the parameter are made. The worth of data can then be measured in terms of deviations from the true value of the parameter. In the second approach, Bayesian statistical techniques are used to update a subjective prior distribution of the parameter to be estimated.

Dawdy et al. (1970) evaluated the economic worth of data for a hypothetical reservoir on Arroyo Seco near Soledad, California. A 500-year stochastic sequence of flows, generated from statistics of historical flows, was used to determine the optimal design of the reservoir. A capacity-cost function was assumed and benefits were computed for flood control and water supply. Multiple samples of various lengths were then taken from the 500-year record. Thus, for a record length of 10 years, fifty separate samples were taken from the sequence. Each sample was used to design the reservoir and an expected value of net benefits was computed as the mean of all sample net benefits for a particular record length. The expected value of net benefits increased as more data were collected. However, the added value per additional year of record decreased as more years of data were collected.

Moss (1970) used the data from Dawdy et al. (1970) with a cost of data collection function to determine where data collection costs were greater than the value of increased information. Discounting of net benefits over time was considered and several different cases were examined for the timing of data collection and reservoir construction. For a situation where no prior streamflow data existed and the reservoir was to be built as soon as enough data was collected, Moss found that the optimal strategy was to collect 9 years of flow data. It should be emphasized that extension techniques were not used to add to the value of the collected data and the conclusions are only applicable to the Arroyo Seco site.

Tschannerl (1971) conducted a similar study in which Markov-generated synthetic streamflows were used to determine a true optimal reservoir design. Again, no extension of data was used. An expected opportunity loss, defined as the difference between benefits from the true optimal design and the expected benefits from short record designs, was computed for each sample size. The expected opportunity loss decreased over time although the rate of decrease was small after 30 years of data collection. Costs of data collection were not considered and optimal data collection strategies were not suggested. McMahon and Codner (1972) used historical data for four Australian Rivers as the population for a sampling strategy study. They concluded that 36 years of data were sufficient for the design of reservoirs at the sites considered.

The second common approach to evaluate the expected worth of additional data is 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). The updated prior distribution is commonly referred to as the posterior distribution. For each additional sample, the posterior distribution is used to compute the expected worth of the additional data. Lenton et al. (1974) used this approach to improve at-site estimates of the annual lag-one autocorrelation of streamflow series. Regional information was used to derive the prior distribution. Kuczera (1982) used a similar approach to combine at-site and regional estimates of flood discharges. Tschannerl (1971) derived prior and posterior distributions of mean streamflows for each length of record. Expected opportunity losses were then computed based on the distributions of streamflow means. Davis and Dvoranchik (1978) computed the expected value of new data for the design of a bridge pier. Monte-Carlo experiments were used to derive probability distributions. Bayesian analysis, in conjunction with simulation, has also been used in the design of stream gaging networks (Moss and Karlinger, 1974).

In summary, worth of data studies have produced results which vary widely from site to site. Simulation of sampling strategies can produce realistic results for specific situations. However, economic evaluation of the results of simulation studies requires more economic data than commonly available for proposed reservoir projects. Economic evaluation of the worth of data has been successfully applied only to hypothetical projects where arbitrary cost and benefit functions could be assumed. Bayesian decision theory has appeared to have potential. However, prior and posterior distributions for output parameters such as reservoir reliability and optimal storage capacities are difficult to specify. Thus, simulation of sampling strategies with accuracy criteria as surrogates for the value of data is the most tractable solution to the worth of data problem.

Computer Simulation of Sampling Strategies

To determine the best combination of record extension technique and number of years of streamflow sampling, Monte-Carlo computer simulation experiments were conducted. In Monte-Carlo simulations, sufficient replicates of experiments are conducted to estimate the probability distributions of system parameters. These types of experiments are often performed when analytical solutions for the moments of probability distributions are not available. For the Walsh Lake reservoir, Monte-Carlo experiments were conducted to estimate measures of the expected sampling error for each record length and extension technique. Samples of various lengths were taken from a population of streamflows. Each sample was then extended and used to evaluate reservoir performance. The results for each sample size were then compared to the true measures of reservoir performance derived from the population of streamflows.

The stochastic streamflow sequences developed in Chapter 4 were used as the population of streamflows. Streamflow sequences were generated for the reservoir inflows and Inflow 3 using the seasonal multisite Markov model. Regional regression estimates of inflow statistics were used in the Markov model. Since 48 years of actual data are available for Inflow 3, stochastic streamflow sequences of length 48 years were generated to impart realism on the experiment. Inflow 3 flows were found to be significantly correlated with observed flows in Issaquah Creek and nearby Taylor Creek. Inflow 3 was therefore used both to simulate reservoir operation and to extend sampled inflow records. Monthly correlations between Walsh Lake reservoir inflows and Inflow 3 used as population statistics in the model are listed in Table 11.

The inflow sample size in the experiment is represented by N. For each replicate of the experiment, concurrent sequences of N years of monthly inflow data and 48 years of Inflow 3 data were generated. Correlations between the inflow sequence and the last N years of Inflow 3 data were computed. Forty-eight years of inflow data were then estimated from the N years of sampled inflow data and the 48 years of Inflow 3 data using one of several record extension techniques. This extended inflow sequence was subsequently used in combination with the Inflow 3 sequence to simulate 48 years of operation of the Walsh Lake reservoir. Simulation of reservoir operation was performed using the mass-balance model of the reservoir described in Chapter 4. The reservoir was drafted during the entire low flow season, July, August and September, and simulation was conducted using a reservoir capacity of 14,000 acre-feet (1.73×10^7 m^3).

Table 11

POPULATION CROSS CORRELATIONS BETWEEN INFLOW 3 AND WALSH LAKE INFLOW

Month	Cross Correlation Between Sites
Oct	.840
Nov	.919
Dec	.890
Jan	.791
Feb	.754
Mar	.880
Apr	.860
May	.794
Jun	.791
Jul	.916
Aug	.857
Sep	.913

For each extended sequence, four measures of reliability were used to evaluate reservoir performance:

1. R_A = the percentage of time the reservoir did not fail in August.
2. R_S = the percentage of time the reservoir did not fail in September.
3. R_F = the percentage of time the reservoir filled to capacity between low flow periods.
4. DEF = the average deficit volume incurred in each year.

R_A and R_S indicate the length of the low flow season for which the reservoir can be reliably drafted, and are computed by:

$$R_A = 100\ (1-N_A/48) \tag{30}$$
$$R_S = 100\ (1-N_S/48) \tag{31}$$

where N_A is the number of Augusts in which the reservoir fails during the 48-year period, and N_S is the number of Septembers in which failure occurs. R_F is a measure of the reliability at which the reservoir refills during the wet season and is computed as the percentage of simulated years in which the reservoir fills to capacity by the end of the wet season. The average volume of deficits, DEF, is computed by:

$$\text{DEF} = \text{SDEF}/48 \tag{32}$$

where SDEF is the cumulative sum of deficits during the 48 years of simulated operation.

Prior to simulation of sampling strategies, the population sequence of stochastic streamflows was used to determine true values of R_A, R_S, R_F, and DEF for the simulated reservoir. These are listed in Table 12. Sample values of these parameters, computed from extended inflow samples, were then compared to the true values to measure the accuracy of a combination of record length and extension technique. In the simulation of each combination of record extension technique and sample size, 500 replicates of the experiment were performed. Thus, for each sample size and extension technique 500 samples of inflows were extended, and 500 values of R_A, R_S, R_F, and DEF were computed. The accuracy of a given sampling strategy was then measured by the

Table 12

POPULATION VALUES OF RESERVOIR PERFORMANCE INDICES BASED ON REGIONAL REGRESSION INFLOWS AND RESERVOIR CAPACITY = 14,000 ACRE-FEET

Performance Index	Population Value
August Reliability R_A	99.56 %
September Reliability R_S	37.89 %
Refill Reliability R_F	98.44 %
Average Deficit DEF	1620.0 Acre-Feet

rmse sampling error and the sampling bias. The rmse indicates the expected magnitude of errors, and is equivalent to the standard deviation of estimates when estimates are unbiased. When sampling bias is significant, the rmse reflects this bias, and is larger in magnitude than the standard deviation of estimates. Sampling bias indicates whether a technique consistently overestimates or underestimates the parameter of interest. A simple measure of bias is computed by:

$$\frac{\sum_{i=1}^{500} X_i}{500} - \text{TRUTH} \tag{33}$$

where X_i is the sample value of the parameter for sample i, and TRUTH is the true value of the parameter.

Inflow record lengths of 3, 4, 5, 6, 7, 10, 15, and 20 years were simulated. Experiments were conducted using no record extension, linear regression extension, MOVE.2 extension, and MOVE.4 extension. Each extension technique was applied both seasonally and nonseasonally. MOVE.2 has been show to be superior to MOVE.1 (Hirsch, 1982). MOVE.3 and MOVE.2 have been found to be indistinguishable in estimation of flow statistics (Vogel and Stedinger, 1984). Therefore, MOVE.1 and MOVE.3 were not tested for use at Walsh Lake.

Assumptions Behind Sampling Experiments

Before presenting the results of the computer simulation experiments described above, it is important to state the critical assumptions behind these experiments. The stochastic streamflow sequences used in the simulation represent only possible Walsh Lake inflow sequences. The estimated statistics used in computing these sequences are best estimates based on no at-site streamflow data and may not reflect the actual probability distribution of Walsh Lake inflows. However, it is assumed that the relationships between sampling accuracy and record length will be similar between the estimated distribution of reservoir inflows and the actual population of inflows. The stochastic streamflow model also assumes long-term persistence effects are negligible for this small catchment. If persistence effects are actually important, the length of record needed may increase significantly.

Results of Computer Simulation Experiments

As discussed previously, each sampling strategy was evaluated in terms of rmse and bias observed in estimates of the four reservoir performance indices. Figures 7 through 13 represent plots of rmse against inflow record length for each of the record extension techniques simulated. Errors for the four reservoir performance indices are plotted on each of these figures. Figure 7 shows the relationship between rmse and record length for no extension of sampled inflows. As expected, the rmse of estimates decreases as the record length increases. In this case, the accuracy of estimates of September reliability R_S is highly sensitive to inflow record length. The rmse drops from 28.7 percent to 18.6 percent when the inflow record length increases from 3 to 7 years. At the other extreme, the accuracy of estimates of reliability in August R_A is relatively insensitive to increased record length. The rmse for this parameter drops from 2.8 percent to 1.6 percent when the record length is increased from 3 to 7 years. This indicates that the reservoir is expected to be extremely reliable in August for a wide range of inflow volumes.

Root mean square errors for seasonal linear regression extension are plotted in Figure 8. For record lengths greater than 5 years, linear regresion extension results in improved accuracy of estimates of performance statistics over no extension. For record lengths of 5 years or less, extension using seasonal linear regression results in higher rmse's of estimates of refill reliability R_F and August reliability R_A. This indicates that correlations between the two sites are not sufficiently high for improved estimation of inflow statistics using seasonal linear regression extension of inflow records of length 5 years or less.

Figure 9 indicates the relationship between rmse and record length for nonseasonal linear regression extension. In all cases, accuracy is improved over no extension when nonseasonal linear regression extension is applied to inflow data. Comparison of Figure 8 and Figure 9 shows that nonseasonal linear regression is more accurate than seasonal linear regression in the estimation of R_A and R_S for record lengths of less than 6 years. Nonseasonal linear regression is more accurate in the estimation of R_F and average deficit

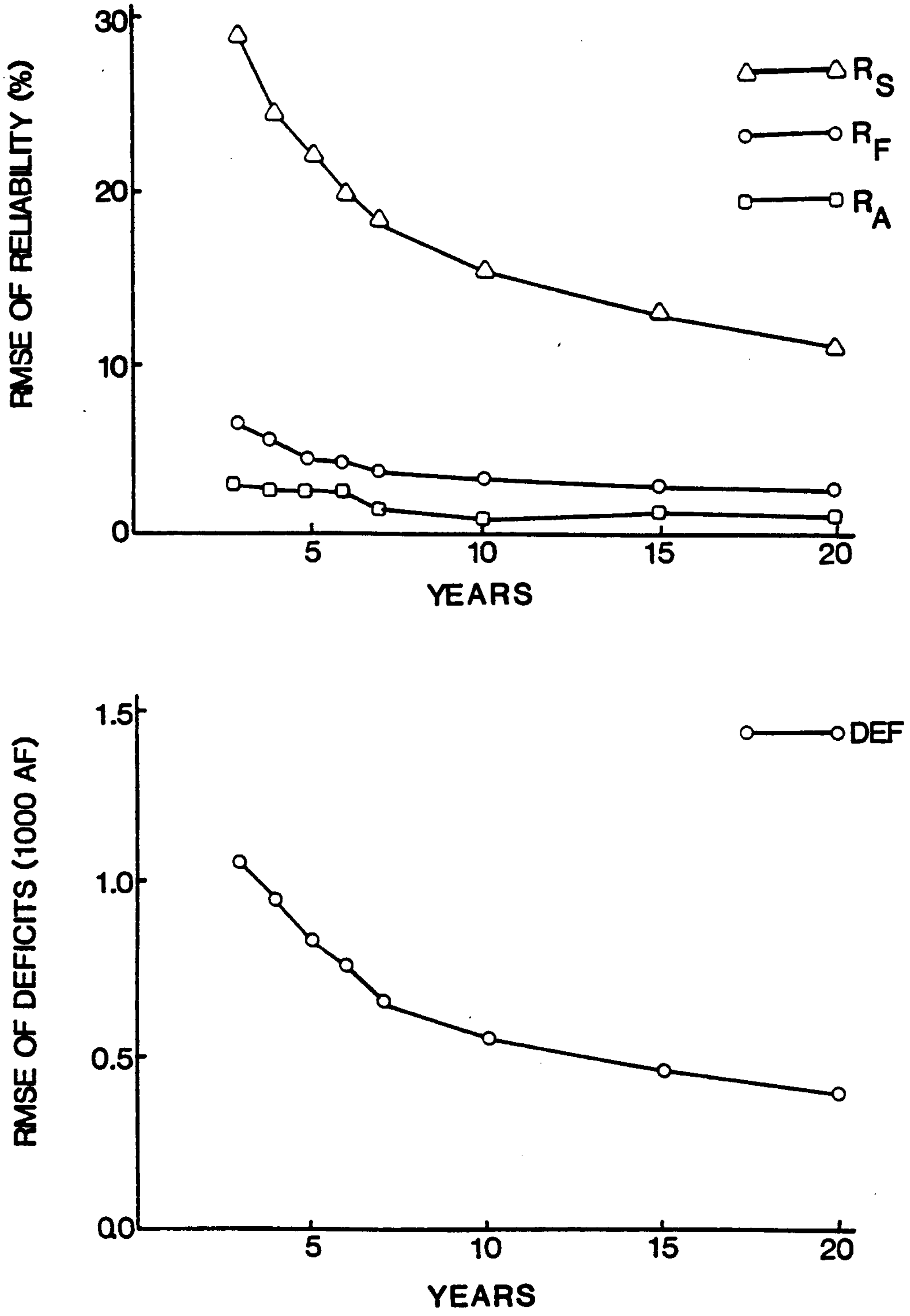


Figure 7. RMSE Curves, No Extension of Inflow Data

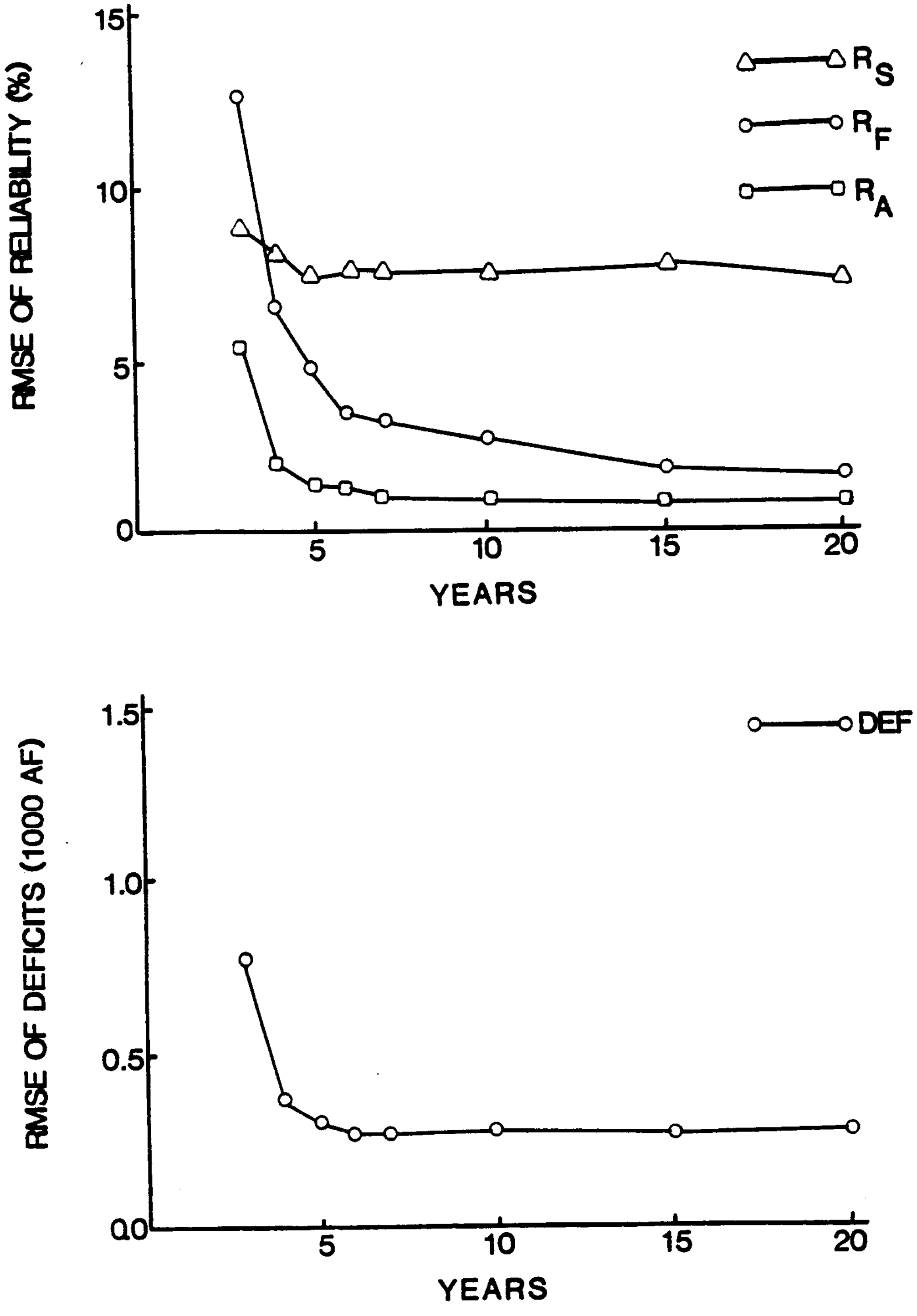


Figure 8. RMSE Curves, Seasonal Linear Regression Extension

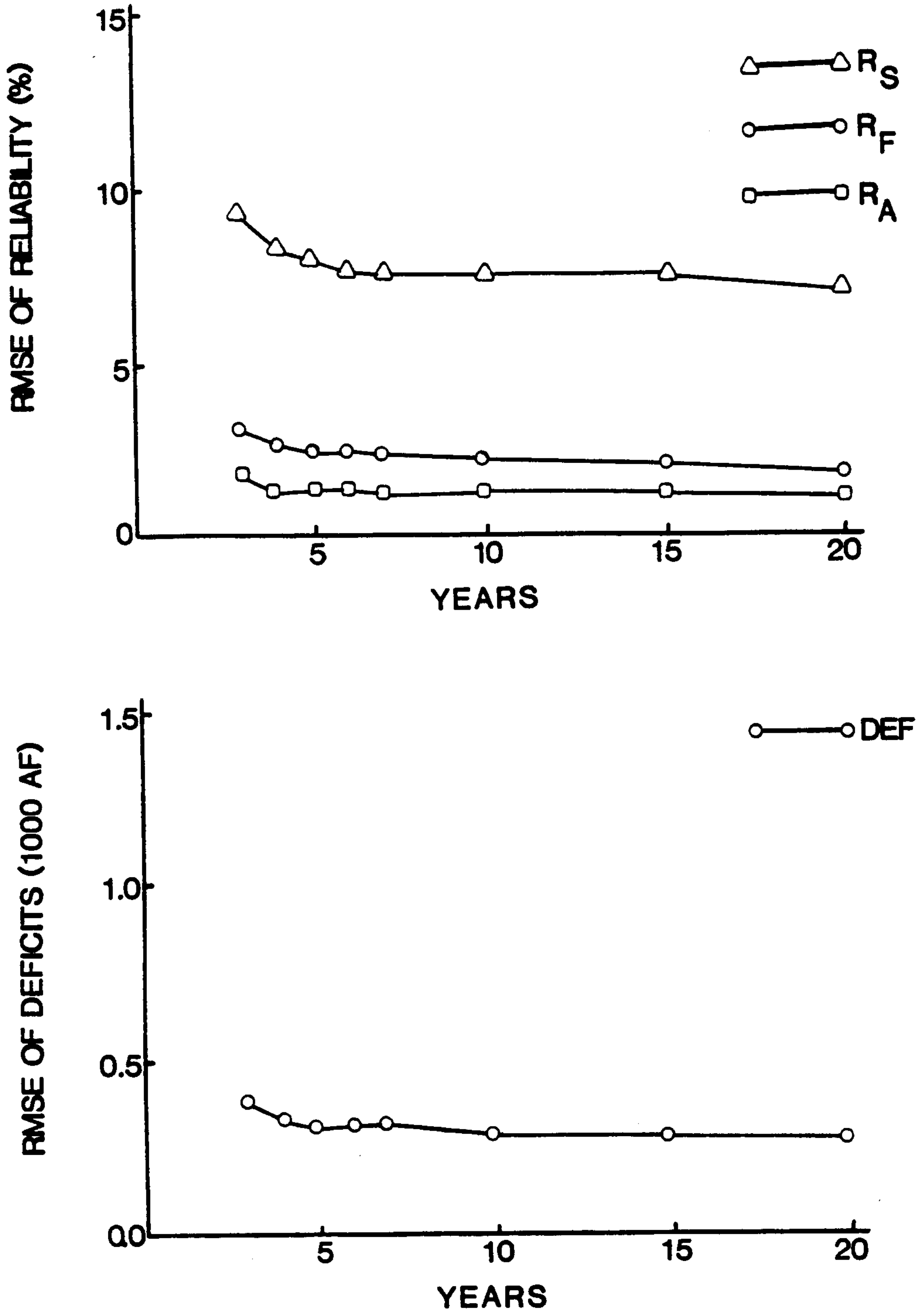


Figure 9. RMSE Curves, Non-Seasonal Linear Regression Extension

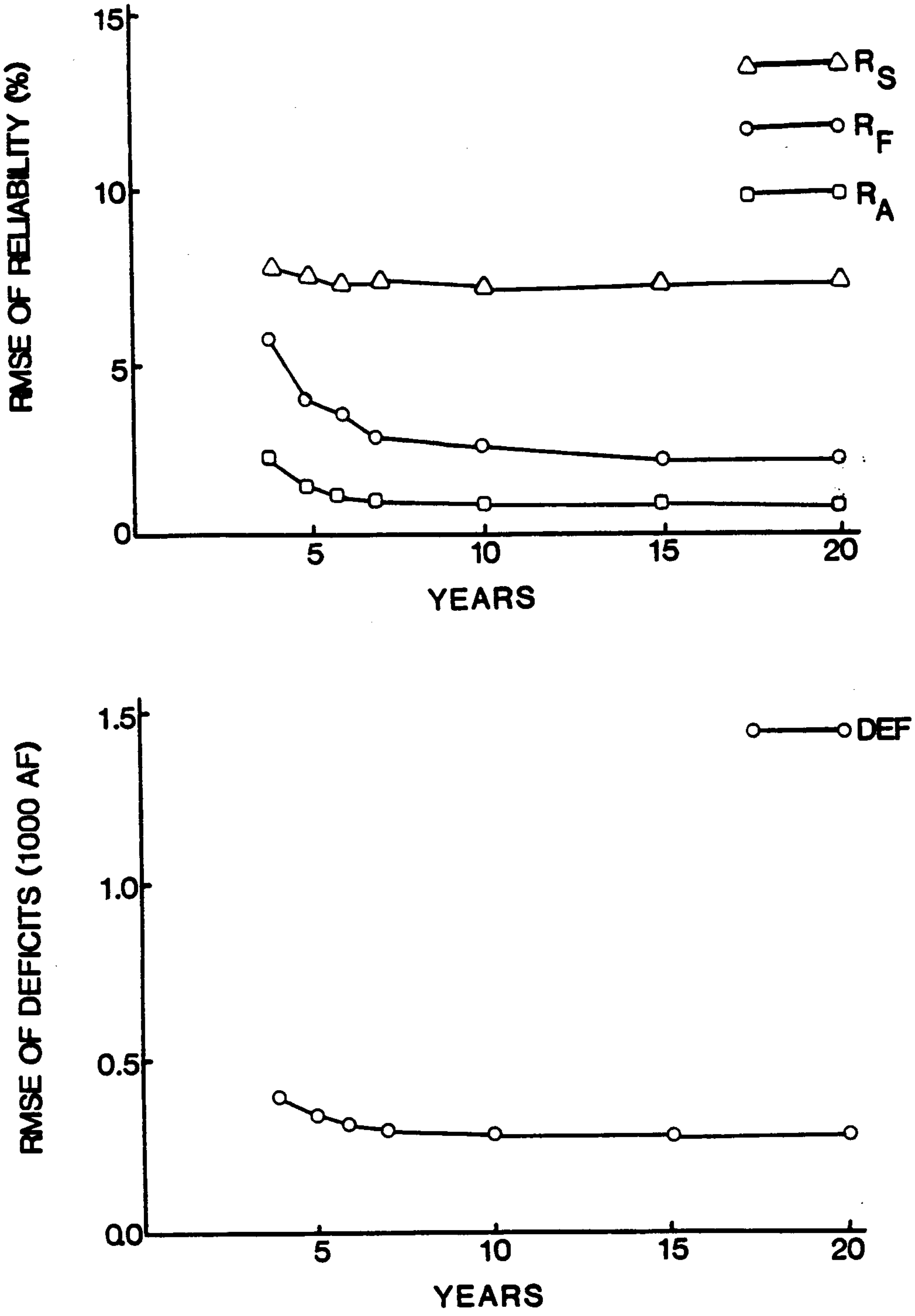


Figure 10. RMSE Curves, Seasonal MOVE.2

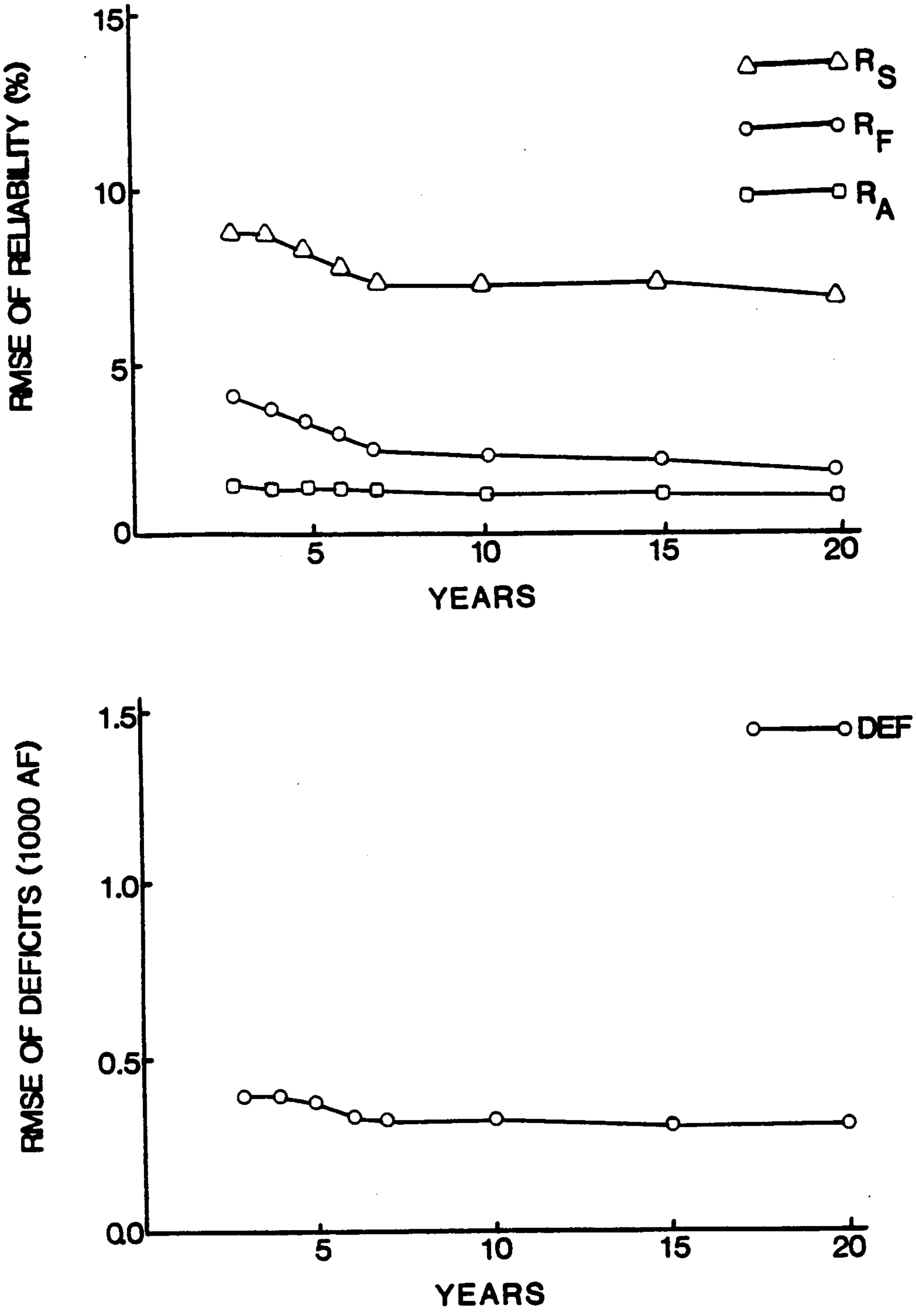


Figure 11. RMSE Curves, Non-Seasonal MOVE.2

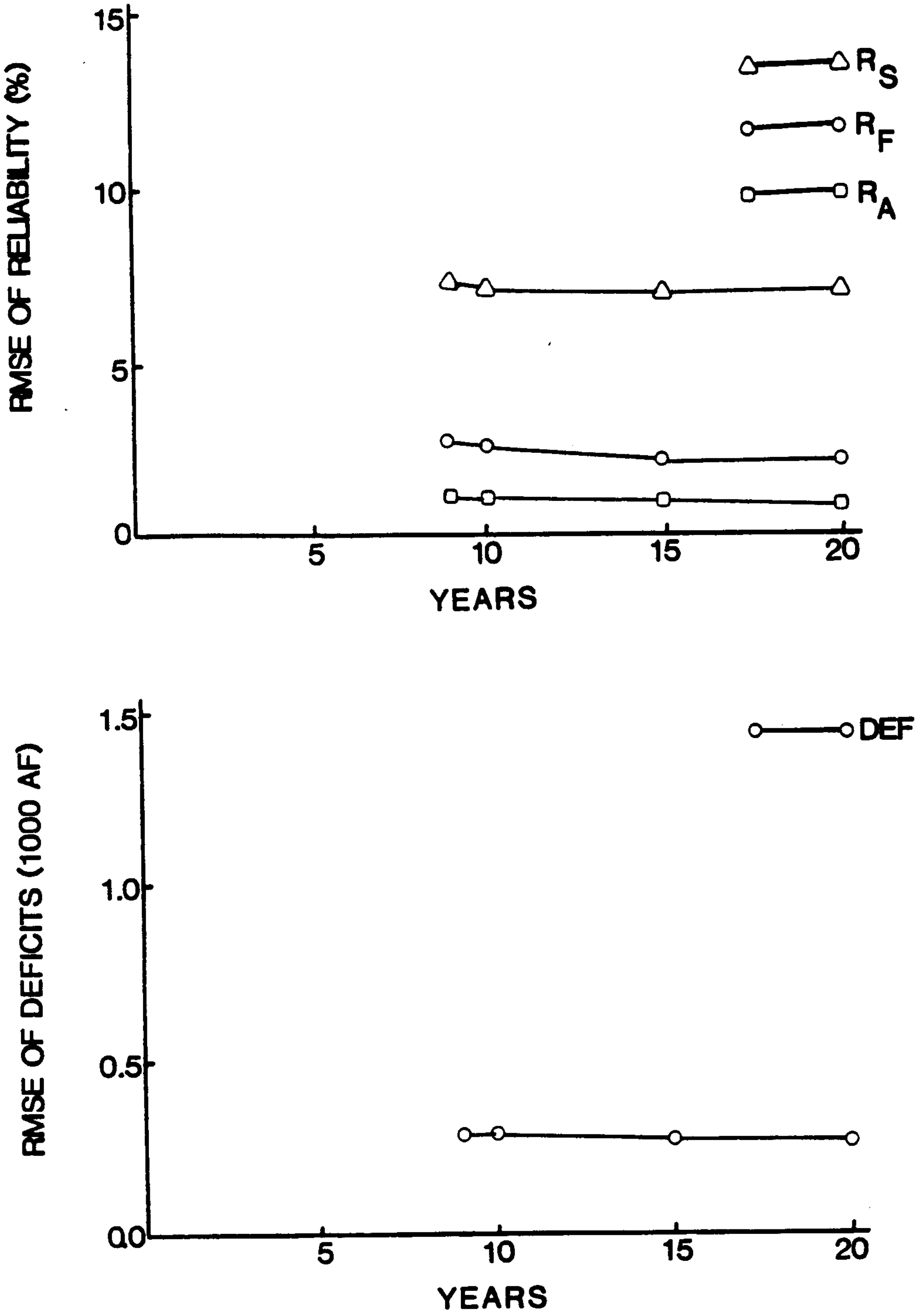


Figure 12. RMSE Curves, Seasonal MOVE.4

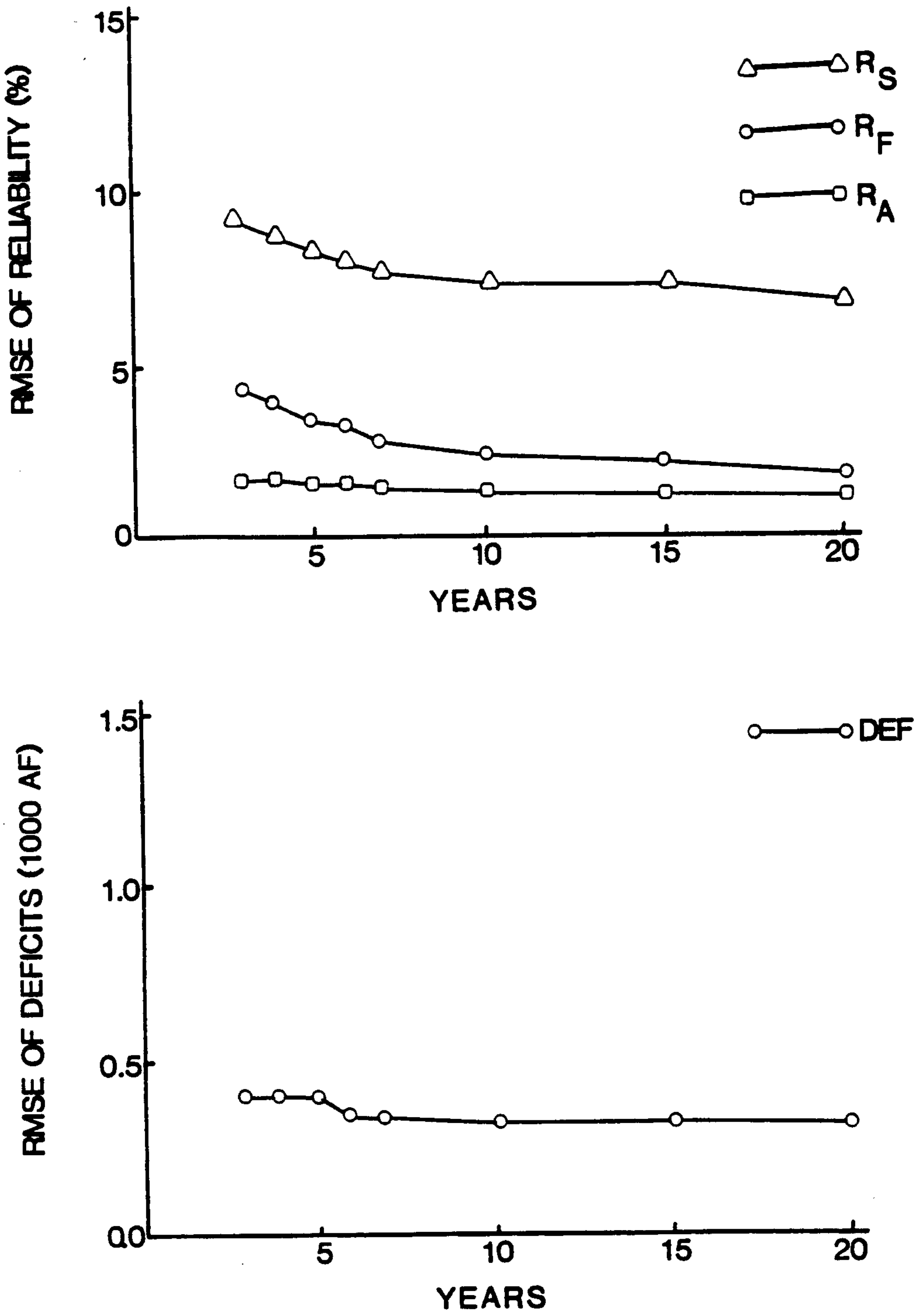


Figure 13. RMSE Curves, Non-Seasonal MOVE.4

volume DEF for record lengths of 10 years or less. When the record length is increased beyond these limits, seasonal linear regression extension is more accurate. This indicates that at short record lengths, the effects of the increase in sample size gained from nonseasonal extension outweighs the loss of seasonality in the model. As record length increases, the ability of seasonal linear regression to model seasonal differences in correlations has a greater effect on sample accuracy than does the increased sample size for nonseasonal extension.

Figures 10 and 11 show plots of rmse against sample size for seasonal and nonseasonal MOVE.2 extension, respectively. Seasonal MOVE.2 extension is defined only for record lengths greater than 3 years. MOVE.2 extension performs better than no extension in all cases. Seasonal MOVE.2 extension is slightly more accurate than seasonal linear regression for record lengths of 4 and 5 years. In all other cases, the accuracy of seasonal MOVE.2 extension is indistinguishable from that of seasonal linear regression. The rmse for seasonal MOVE.2 is less than that of nonseasonal MOVE.2 for estimation of September reliability, average deficit volume, August reliability at record lengths greater than 4 years, and reservoir refill reliability at record lengths greater than 5 years. Thus, in the majority of cases MOVE.2 extension is more accurate when applied seasonally. Exceptions occur in the estimation of August reliability and refill reliability at sample sizes less than 6 years. For these cases, nonseasonal extension is marginally more accurate.

Root mean square error plots for MOVE.4 extension are shown in Figures 12 and 13. Minimum variance estimators of the variance become undefined for sample sizes smaller than 9; seasonal MOVE.4 extension was therefore performed for record lengths of 9 years or greater. In all cases, MOVE.4 extension is more accurate than no extension of inflow data. In general, the performance of MOVE.4 extension is indistinguishable from that of MOVE.2 extension. rmses for MOVE.2 and MOVE.4 are nearly identical for all record lengths and for both seasonal and nonseasonal extension.

The rmse of estimation is indicative of the expected magnitude of sampling variability. This statistic does not indicate whether a method tends to overestimate or underestimate the value of the parameter of interest.

Therefore, it is also important to examine biases in estimates obtained from record extension techniques. A positive bias indicates that on average a parameter is being overestimated; a negative bias indicates a tendency to underestimate. Computed biases in estimates of reservoir performance indices for each record extension technique are listed in Table 13 for record lengths of 6 and 20 years. For short record lengths, seasonal linear regression is the least biased extension technique. Seasonal MOVE.2 is marginally more biased. Nonseasonal-linear regression is significantly positively biased in the estimation of September reliability. In general, nonseasonal techniques are more biased than seasonal techniques at short record lengths. As record length increases, the bias in seasonal linear regression remains significant. At a record length of 20 years, MOVE.2 is the least biased technique for the estimation of all statistics except September reliability. Nonseasonal MOVE.2 and MOVE.4 are the least biased techniques in the estimation of September reliability for a record length of 20 years.

Conclusions Based on Results of Simulation Experiments

Several conclusions can be drawn from the results of the simulation of sampling strategies. In nearly all cases, record extension results in significant decreases in sampling error. In terms of sampling accuracy, as measured by the rmse of estimation, seasonal MOVE.2 extension is the superior extension technique for record lengths of 6 years or longer. Nonseasonal MOVE.2 and nonseasonal MOVE.4 are marginally more accurate at record lengths less than 6 years. In terms of bias, seasonal linear regression is slightly less biased than seasonal MOVE.2 for record lengths of 6 years or less. For longer inflow records, MOVE.2 is generally less biased. All nonseasonal techniques exhibit significant biases when applied to short record lengths. Overall, these experiments indicate that MOVE.2 is the superior record extension technique for the estimation of the Walsh Lake reservoir performance indices.

The true optimal record length cannot be determined without considering the economic costs and benefits of streamflow sampling. However, conclusions regarding the relative improvement in sampling accuracy with increased record length can be made from the results of these experiments. Using seasonal MOVE.2 extension, the decrease in the rmse is negligible after 6 years of

Table 13

BIASES IN RECORD EXTENSION TECHNIQUES

Extension Technique	Record Length (years)	Bias R_A (%)	R_S (%)	R_F (%)	DEF (1000 acre-ft)
Seasonal Linear	6	0.03	-1.00	-0.56	0.07
Regression	20	0.16	-1.10	0.33	0.05
Nonseasonal	6	-0.19	2.56	-0.83	0.06
Linear Regression	20	-0.09	1.07	-0.24	0.05
Seasonal MOVE.2	6	-0.09	-1.05	-0.82	0.10
	20	0.00	-1.03	-0.14	0.33
Nonseas. MOVE.2	6	-0.25	1.91	-1.11	0.10
	20	-0.18	-0.08	-0.33	0.13
Nonseas. MOVE.4	6	-0.25	1.84	-1.13	0.10
	20	-0.19	-0.08	-0.33	0.13

record for the estimation of August reliability, September reliability, and average deficit volume. For the estimation of reservoir refill reliability, the rmse decreases from 3.3 percent to 2.1 percent when the record length is increased from 6 to 20 years. A record length of 6 years is therefore sufficient for the accurate estimation of Walsh Lake reservoir performance indices. The recommended sampling strategy is to sample reservoir inflows for 6 years and extend the sampled inflow record using seasonal MOVE.2 extension.

Use of MOVE.2 at Walsh Lake

The first step in extending the Walsh Lake inflow record is to choose a long record, or base station, from which to extend. In this study, the deregulated Inflow 3 record was used for convenience of simulation. Other possibilities include Cedar River Inflows 1 and 2, and the Skykomish River near Gold Bar. Currently, 48 years of deregulated flows are available for each of the Cedar River Inflows. The Skykomish River gage has been in operation since 1928, and is not significantly affected by diversion or regulation. Walsh Lake monthly inflows are expected to be more significantly correlated with the deregulated Cedar River Inflows than with the Skykomish flows. However, use of the Cedar River Inflows for extension requires that deregulated Inflows be developed for the Walsh Lake gaging period.

The choice of a base station is dependent on the degree of monthly correlation between the various base stations and the sampled Walsh Lake inflows. For each possible base station, correlations between the Walsh Lake inflow volume and the base station flow volume are computed for each month. Thus, twelve monthly correlations based on 6 years of concurrent flow data are computed. The station for which monthly correlations are highest is then chosen as the base station for record extension.

Once a base station is chosen, the following statistics are computed for each month j:

N_1 = years of Walsh Lake inflow record.
N_2 = years of additional record available at the base station.
R = the correlation between Walsh Lake inflows and base station flows during month j.

$M(Y)$ = the mean Walsh Lake inflow volume during month j.

$S(Y)$ = the standard deviation of Walsh Lake inflow volumes during month j.

$M(X_1)$, $S(X_1)$ = the mean and standard deviation of base station flows in month j for N_1 years of record.

$M(X_2)$, $S(X_2)$ = the mean and standard deviation of base station flows in month j for N_2 years of record.

$M(X)$, $S(X)$ = the mean and standard deviation of base station flows in month j for $N_1 + N_2$ years of record.

$\hat{M}(Y)$, $\hat{S}(Y)$ = the Matalas-Jacobs improved estimates of the mean and standard deviation, computed from Equations 14, 15, and 16.

Once the above statistics are computed, the MOVE.2 extension equation (Equation 21) can be applied to derive a series of 48 years of Walsh Lake inflow data. For each month there is a different extension equation based on estimates of means and variances for that month.

The extended Walsh Lake monthly inflow record can be used in a variety of ways to evaluate the performance of the proposed reservoir. Reservoir simulation experiments such as those described in Chapter 4 can be conducted using the extended sequence in conjuction with Inflow 3 flow data and instream flow requirements. Forty-eight years of record may not adequately represent the range of possible inflow sequences. To model operation of the reservoir under a wide range of inflow conditions, the extended record can also be used to develop stochastic streamflow sequences. Thus, means, variances, and correlations can be computed from the extended inflow record and used to estimate parameters for a first-order Markov model, ARMA(1,1) model, or other stochastic streamflow models as described in Chapter 4. Multiple simulation of reservoir operation using these stochastic sequences can then be used to incorporate uncertainty into the design of the reservoir.

Chapter 7

SUMMARY AND CONCLUSIONS

A proposed expansion of Walsh Lake was examined as an option to increase yield for the Seattle water supply system. In particular, estimates of the Walsh Lake reservoir inflows were developed and used to test the performance of the Walsh Lake reservoir under a variety of conditions. No reliable streamflow data are available for the reservoir inflows. Therefore, computer simulations of streamflow data collection strategies were performed to determine efficient ways of collecting and using sampled flow data. Methods for extending sampled inflow records were tested for applicability to the Walsh Lake Reservoir. In addition, the number of years of data needed for accurate estimation of reservoir performance parameters was determined.

Summary of Results

Historical flow data for Walsh Lake indicate a mean annual inflow volume of 15,300 acre-feet (.598 cms). However, the accuracy of this data is suspect due to unrecorded flow diversions upstream of the gage and changes in the Walsh Lake drainage basin. Water balance computations result in a mean annual inflow volume of 25,600 acre-feet (161 cm). This figure indicates the potential runoff volume that could be produced by the Walsh Lake drainage basin. Actual inflow volumes may be substantially lower. Using data from the nearby Issaquah Creek basin, a mean annual inflow volume of 12,600 acre-feet ($1.55x10^7$ m^3) is estimated. The Issaquah Creek basin is believed to receive considerably less precipitation than does the Walsh Lake basin; inflow estimates developed from these data are lower bounds on possible Walsh Lake reservoir inflows. Regional regressions developed from flow data for numerous gaged basins in Western Washington result in a mean annual inflow volume of 24,800 acre-feet ($3.06x10^7$ m^3). Regional regressions have been found to be highly inaccurate in this region (Collings, 1971). Regional regression estimates of monthly flow statistics are nonetheless considered to be the best preliminary estimates of the actual flow statistics.

These initial estimates of reservoir inflow statistics were used to derive preliminary estimates of reservoir performance indices. Operation of the proposed reservoir was simulated using a simplified mass-balance computer model. Monthly stochastic flow sequences were developed from the Issaquah Creek estimates of inflow statistics and from regional regression inflow statistics. These sequences were then used in combination with Cedar River flow sequences and monthly instream flow requirements to model changes in reservoir storage over time. Simulations using the lower bound Issaquah Creek inflow estimates indicate that the reservoir cannot be operated reliably for the entire low flow season. If operated during July and August only, these simulations indicate that the reservoir can be operated reliably for a reservoir capacity of 14,000 acre-feet (1.73×10^7 m^3). Simulations using the Regional Regression inflow statistics indicate that the reservoir will consistently fill to capacity during the wet season. Under these conditions, the reservoir can be operated reliably for the entire low flow season with a capacity of 20,000 acre-feet (2.47×10^7 m^3).

The evaluations of reservoir performance described above are based upon regionally estimated inflow statistics. They are therefore only preliminary estimates of reservoir reliability, and are not sufficiently accurate for design purposes. To accurately evaluate how the reservoir will perform under actual operating conditions it is necessary to collect streamflow data at the Walsh Lake reservoir site. This information can then be used to develop improved estimates of inflow statistics and reservoir reliability. The amount of information gained from collected data can be increased using correlations with concurrent flows at a site with a longer streamflow data record. Techniques exist which can be used to extend the streamflow record at the site of interest to the length of other records in the region. Methods of record extension examined in this study include linear regression, MOVE.2, and MOVE.4 extension. Stochastic inflow sequences develped from regional regression inflow statistics were used in combination with Cedar River flows to simulate application of each record extension technique to inflow records of various lengths. Numerous inflow samples of a given size were extended and used to estimate reservoir performance indices. The estimates based on the extended samples were then compared with population values derived from long sequences to deterimine the expected accuracy of the various sampling strategies. The results of these experiments indicate that seasonal MOVE.2 extension is the superior method for this case, based on accuracy and bias criteria. In nearly all cases, record extension significantly improves sampling accuracy.

Nonseasonal techniques are significantly biased, and seasonal linear regression extension is less accurate than seasonal MOVE.2.

As more data are collected, the improvement in accuracy of inflow estimates becomes small relative to data collection costs. Previous studies indicate that the worth of streamflow data varies greatly from site to site, depending upon the variability of inflows and the use for which the data are intended. An economic evaluation of the worth of streamflow data was not conducted in this study. However, when using seasonal MOVE.2 extension the relative increase in sampling accuracy becomes small after 6 years of data collection. Therefore, it is recommended that 6 years of Walsh Lake reservoir inflow data be collected and extended using seasonal MOVE.2.

Recommendations for Further Study

The determination of the number of years of data needed for accurate estimation of Walsh Lake inflows is based on accuracy criteria alone. Thus, a potential area of further research is the economic evaluation of the costs and benefits of additional streamflow data. In general, the first year of data collection will be the most costly. During this period, costs include the costs of stream gage installation and of site visits to determine the rating curves for the gages. After the first year, the only costs involved are maintenance costs and the costs of periodic site visits to verify the rating curves. Thus, the determination of data collection costs is expected to be straightforward.

The benefits of streamflow data are more difficult to determine. The benefits of data are measured in terms of the value of making better decisions. To compute these benefits, it is necessary to know the projected benefits derived from operation of the proposed Walsh Lake reservoir. With this information, the expected benefits of a given reservoir design can be determined. Simulation of sampling strategies as described previously could then be conducted, where inflow records of varying lengths are extended and used to design the reservoir. For each record length, the expected benefits will be the average of the benefits derived from each reservoir design over all replicates of the simulation. An expected opportunity loss is defined as the difference between the expected benefits for a given record length and the benefits from the optimal design based on perfect knowledge. The worth of additional data is then measured by the decrease in the expected opportunity loss as record length increases. Data should be collected only as long as the costs of an additional year of data are less than the decrease in the expected opportunity loss.

A second subject warranting further study is the effect of varying degrees of long-term persistence in streamflow data on the length of record needed. Long-term persistence has been observed in many hydrologic sequences, and is indicative of the dependence of current hydrologic events on conditions occurring on the order of years prior to the present. In this study, flows were assumed to be derived from a first-order Markov process in which monthly flow volumes depend only on the flow in the previous month. Thus, to model the effects of long-term persisistence an ARMA(1,1) or other model capable of exhibiting this feature would be used to generate stochastic flow sequences. Sampling experiments as described in this study could then be conducted using these flow sequences. Persistence in flow data is often exhibited by long series of low flows followed by a series of high flows. Thus, it is expected that long-term persistence in inflow sequences will increase the length of record needed.

REFERENCES

Alley, W. M., and A. W. Burns, "Mixed Station Extension of Monthly Streamflow Records," Journal of Hydraulic Engineering, American Society of Civil Engineers, Vol. 109, No. 10, Oct. 1983, pp. 1272-1284.

Benjamin, J. R., and C. A. Cornell, Probability, Statistics, and Decision Theory for Civil Engineers, Mcgraw-Hill, New York, 1970.

Burges, S. J., D. P. Lettenmaier, and C. L. Bates, "Properties of the Three-Parameter Log Normal Probability Distribution, Water Resources Research, Vol. 11, No. 2, 1975, pp. 229-235.

Collings, M. R., "A Proposed Streamflow Data Program for Washington State," U. S. Geological Survey Open File Report, Tacoma, Washington, 1971.

Davis, D. R., and W. M. Dvoranchik, "Evaluation of the Worth of Additional Data," Water Resources Bulletin, Vol. 7, No. 4, 1978, pp. 700-707.

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

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

Fiering, M. B., "Use of Correlation to Improve Estimates of the Mean and Variance," U. S. Geological Survey Professional Paper 434-C, 1963.

Fiering, M. B., and B. B. Jackson, Synthetic Streamflows, Water Resources Monograph No. 1, American Geophysical Union, Washington, D. C., 1971.

Haan, C. T. and D. M. Allen, "Comparison of Multiple Regression and Principal Component Regression for Predicting Water Yields in Kentucky," Water Resources Research, Vol. 8, No. 6, 1972, pp. 1593-1596.

Haan, C. T., Statistical Methods in Hydrology, The Iowa State University Press, Ames, Iowa, 1977.

Hirsch, R. M., "An Evaluation of Some Record Reconstruction Techniques," Water Resources Research, Vol. 15, No. 6, 1979, pp. 1781-1790.

Hirsch, R. M., "A Comparison of Four Streamflow Record Extension Techniques," Water Resources Research, Vol. 18, No. 4, 1982, pp. 1081-1088.

Lane, W. L., Applied Stochastic Techniques (LAST Computer Package), User Manual, Division of Planning and Technical Services, Bureau of Reclamation, Denver, Colorado, 1979.

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

Lettenmaier, D. P., and S. J. Burges, "Operational Assessment of Hydrologic Models of Long-Term Persistence," Water Resources Research, Vol. 13, No. 1, 1977, pp. 113-124.

Matalas, N. C., "Mathematical Assessment of Synthetic Hydrology," Water Resources Research, Vol. 3, No. 4, 1967, pp. 937-946.

Matalas, N. C., and B. Jacobs, "A Correlation Procedure for Augmenting Hydrologic Data," U. S. Geological Survey Professional Paper 434-E, 1964.

McCuen, R. H., W. J. Rawls, and B. L. Whaley, "Comparitive Evaluation of Statistical Methods for Water Supply Forecasting," Water Resources Bulletin, Vol. 15, No.4, 1979, pp. 935-947.

McMahon, R. A., and G. P. Codner, "Inadequate Hydrologic Data and Reservoir Capacity," Decisions With Inadequate Hydrologic Data, Proceedings of the Second International Symposium in Hydrology, Fort Collins, Colorado, 1972, pp. 210-220.

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

Moss, M. E., and W. L. Hauschild, "Evaluation and Design of a Streamflow Data Network in Washington, U. S. Geological Survey Report Open File Report 78-167, Tacoma, Washington, 1978.

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

Riggs, H. C., "The Transfer Value of Information Collected on Representative Basins," International Association of Scientific Hydrology Publication 96, 1970, pp. 614-631.

Riggs, H. C., Regional Analysis of Streamflow Characteristics, U. S. Geological Survey Techniques of Water Resources Investigations, Book 4, Chapter B3, 1973.

Salas, J. D., J. W. Delleur, Y. Yevjevich, and W. L. Lane, Applied Modeling of Hydrologic Time Series, Water Resources Publications, Littleton, Colorado, 1980.

Stedinger, J. R., and G. D. Tasker, Regional Hydrologic Regression I, Ordinary, Weighted, and Generalized Least Squares Compared, submitted to Water Resources Research, 1984.

Thornthwaite, C. W., and J. R. Mather, Instructions and Tables for Computing Potential Evapotranspiration and the Water Balance, Laboratory for Climatology, Publication No. 10, Centerton, New Jersey, 1957.

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

Vine, James D., "Geology and Coal Resources of the Cumberland, Hobart, and Maple Valley Quadrangles, King County, Washington," U. S. Geological Survey Professional Paper 624, Washington, D. C., 1969.

Vogel, R. M., and J. R. Stedinger, Minimum Variance Streamflow Record Augmentation Procedures, submitted to Water Resources Research, May 1984.

APPENDIX A

ASSESSMENT OF POTENTIAL STREAM GAGING SITES

In this appendix, potential stream gaging sites on Rock Creek and Walsh Lake are assessed. Guidelines for the selection of gaging sites have been set forth in a series of U. S. Geological Survey publications, including the most recent Water Supply Paper (Vols. 1 and 2) by Rantz (1982). Much of the discussion given below is derived from Volume I. Stream gages measure stage (elevation) of the water surface relative to a datum. Discharge is then computed from the stage-discharge relationship for a given site. This relationship is established by measuring stage and discharge concurrently at a site until a curve can be fit to the measured data. Stage can be measured by a simple staff gage or a continuously recording bubble sensor gage.

For a natural gaging site to be ideal, it should be located on a relatively straight reach of the stream. Rantz recommends a straight reach of at least 300 feet in length. Flow in the cross section should be uniform and the cross section should be stable over time. The channel bed and banks should be resistant to scour and free from obstructions. No significant seepage should be occurring through the stream bed or banks. To develop an adequate stage-discharge relationship, the water level at the site must be sensitive to changes in discharge. This requires that a control, either natural or artificial, exist downstream of the gaging site. Natural controls often exist in streams with well developed pool and riffle sequences (Dunne and Leopold, 1978). In such cases, pools which are sensitive to changes in discharge develop behind sediment deposits and vegetative growth in the stream. Natural controls in a pool and riffle sequence are generally low flow controls; in higher flows, the controls are commonly flooded out. Constrictions in the channel often result in more stable controls. Large bedrock outcroppings into the channel are especially resistant to erosion.

If no controls exist on the stream, an artificial control must be constructed. These generally consist of either weirs or flumes. Weirs create an obstruction to flow, forming a pool upstream. Discharge can then be calibrated to the height of the water surface above the weir crest. During high flows, these weirs behave as broad-crested weirs. However, a V-notch is often placed in the center of the weir to increase its sensitivity to low flows. The weir should be installed in a straight section of the stream. Channel banks should be sufficiently stable to anchor the weir, and high enough to handle the backwater pool created by the weir. In streams with high sediment loads, sediment and debris can deposit behind the weir. In such cases, a flume is more desirable. Flumes are defined as control sections in which critical depth occurs (Kilpatrick and Schneider, 1983). In critical flumes, head is measured upstream of the critical section, whereas in supercritical flumes head is measured downstream of the critical section.

Survey of Sites on Rock Creek

Gaging sites on Rock Creek were primarily investigated in the vicinity of the Ditch Road crossing of the creek. This is the approximate location of the proposed dam on Rock Creek and Walsh Lake Ditch. Due to the uncertainty associated with diverted flows in the vicinity of the road crossing, it is recommended that gages on both Rock Creek and Walsh Lake Ditch be located upstream of the diversions. The City of Seattle Water Department currently has a staff gage between the two diversion sites. Although this site is accessible and has a stable cross section, it is inadequate due to the effects of the diversions and the lack of any natural control sections.

Immediately upstream of the pipe diversion there is a 200-foot straight reach of channel. At the downstream end of this reach, there is a low flow control consisting of a rock sandbar and clumps of vegetation in the channel. The pool upstream of this control is relatively wide. The disadvantages of this site are that the banks are low (2 feet above the channel bottom), and that the section is controlled only for normal flows. Due to the low banks, this could also be an inadequate site for a weir. Backwater effects from the weir will cause higher flows to go overbank and around the weir structure. Upstream of this reach, the channel meanders and becomes divided. Several small inflow channels also begin entering Rock Creek above this point. Therefore, to accurately characterize Rock Creek flows at the dam site, the stream gage must be located on the 200-foot straight reach immediately upstream of the diversions. A weir will be necessary to provide an adequate stage-discharge relationship for a wide range of streamflows.

Sedimentation is not seen as a major problem; it will not be necessary to install a flume on Rock Creek.

Survey of Sites on Walsh Lake Ditch

As in the case of Rock Creek, the City of Seattle has located a staff gage on the ditch between the two diversion sites. The site does have the advantages of accessibility and a downstream constriction control on the channel. However, the site is inadequate for the following reasons:

1. uncertainty associated with flow diversions
2. non-uniform flow in the cross-section
3. inadequately long straight reach.

Upstream of the flow diversions on the ditch, there are two relatively long, straight reaches of the channel. The first is approximately 180 feet long. The channel bottom is silty, and the banks are generally 3.5 feet above the channel bottom. The second reach is approximately 250 feet long. The channel bottom here is well armored with a layer of pebbles and cobbles. The banks are 10 feet high or greater. No natural controls exist on either reach, indicating that a weir is required. In this case, the second upstream reach is more appropriate due to its stable cross section and high banks. Another option for measuring flows on Walsh Lake Ditch is measure the stage at the Walsh Lake outlet. The lack of accessibility to the outlet due to dense vegetative growth in the marsh land surrounding the Lake may be a drawback to this option, however.

REFERENCES

Dunne, T. and L.B. Leopold, Water in Environmental Planning, W.H. Freeman and Co., San Francisco, 1978.

Kilpatrick, F.A. and V.R. Schneider, "Use of Flumes in Measuring Discharge," Techniques of Water Resources Investigations of the U.S.G.S., Book 3, Chapter A14, 1983.

Rantz, S.E., "Measurement and Computation of Streamflow: Volume 1. Measurement of Stage and Discharge," U.S. Geological Survey Water Supply Paper 2175, 1982.

Rantz, S.E., "Measurement and Computation of Streamflow: Volume 2. Computation of Discharge," U.S. Geological Survey Water Supply Paper 2175, 1982.