Cost Effectiveness of the Stream-Gaging Program in Maine— A Prototype for Nationwide Implementation

United States
Geological
Survey
Water-Supply
Paper 2244

Cost Effectiveness of the Stream-Gaging Program in Maine—A Prototype for Nationwide Implementation

By R. A. Fontaine, M. E. Moss, J. A. Smath, and W. O. Thomas, Jr.

U. S. GEOLOGICAL SURVEY WATER-SUPPLY PAPER 2244

UNITED STATES DEPARTMENT OF THE INTERIOR
WILLIAM P. CLARK, Secretary

GEOLOGICAL SURVEY
Dallas L. Peck, Director

UNITED STATES GOVERNMENT PRINTING OFFICE: 1984

For sale by Distribution Branch
Text Products Section
U.S. Geological Survey
604 South Pickett Street
Alexandria, Virginia 22304

Library of Congress Cataloging in Publication Data
Main entry under title:
Cost-effectiveness of the stream-gaging program in Maine.
(Water-supply paper / United States Geological Survey ; 2244)
Bibliography: p.
Supt. of Docs. no.: I 19.13:2244
1. Stream measurements—Maine—Cost-effectiveness.
I. Fontaine, Richard A. II. Series: Geological Survey water-supply paper ; 2244.
TC801.U2 no. 2244 553.7'0973s 83-600267
[TC177] [551.48'3'0723]

PREFACE

The collection of surface-water data is a major activity of the U.S. Geological Survey's (USGS) Water Resources Division (WRD). Approximately $40 million was spent in 1982 by WRD in cooperation with State and local governments and other Federal agencies in the collection of these data. This major expenditure of funds for hydrologic data collection should be evaluated with respect to the needs of the data users and the utility of the data. It is essential, therefore, that a rigorous analysis be made of the stream-gaging program to assure maximum cost-effectiveness. USGS is undertaking a nationwide analysis of its stream-gaging program over the next 5 years. The results from such an analysis should satisfy both local and national water-data needs within budget constraints while maintaining quality control.

This report for the State of Maine represents the first in a series of statewide reports describing this analysis. The proposed techniques and methodology for completing this nationwide analysis are described and documented in this report by application to the Maine stream-gaging program.

Analysis of the stream-gaging program is designed to define and document the most cost-effective means of furnishing streamflow information. The stream-gaging activity is no longer considered a network of observation points, but rather an information system in which data are provided by both observation and synthesis. Alternative methods of providing streamflow information such as flow routing and statistical methods are investigated as to their cost-effectiveness, accuracy, and information content.

Recently, new techniques for evaluating the cost-effectiveness of data-collection programs have been developed. These techniques, Kalman filtering and mathematical programming, are utilized to define strategies for operating the stream-gaging program so that the uncertainty in the streamflow records is minimized. The USGS first applied these techniques to a stream-gaging program in the Lower Colorado River Basin. Subsequently, the techniques have been expanded and improved, and are being applied to the present nationwide study of the USGS stream-gaging program. No doubt these techniques will continue to be modified and improved over the duration of the study.

The analysis of the stream-gaging program is a part of the continuing effort of the USGS to evaluate the Nation's water resources. The national stream-gaging program that results from this analysis should be responsive to the needs of local, State, and Federal agencies and provide streamflow information in the most cost-effective manner.

CONTENTS

FIGURES

TABLES

Cost-Effectiveness of the Stream-Gaging Program in Maine—a Prototype for Nationwide Implementation

By R. A. Fontaine, M. E. Moss, J. A. Smath, *and* W. O. Thomas, Jr.

Abstract

This report documents the results of a cost-effectiveness study of the stream-gaging program in Maine. Data uses and funding sources were identified for the 51 continuous stream gages currently being operated in Maine with a budget of $211,000. Three stream gages were identified as producing data no longer sufficiently needed to warrant continuing their operation. Operation of these stations should be discontinued. Data collected at three other stations were identified as having uses specific only to short-term studies; it is recommended that these stations be discontinued at the end of the data-collection phases of the studies. The remaining 45 stations should be maintained in the program for the foreseeable future.

The current policy for operation of the 45-station program would require a budget of $180,300 per year. The average standard error of estimation of streamflow records is 17.7 percent. It was shown that this overall level of accuracy at the 45 sites could be maintained with a budget of approximately $170,000 if resources were redistributed among the gages.

A minimum budget of $155,000 is required to operate the 45-gage program; a smaller budget would not permit proper service and maintenance of the gages and recorders. At the minimum budget, the average standard error is 25.1 percent. The maximum budget analyzed was $350,000, which resulted in an average standard error of 8.7 percent.

Large parts of Maine's interior were identified as having sparse streamflow data. It was determined that this sparsity be remedied as funds become available.

INTRODUCTION

The U.S. Geological Survey (USGS) is the principal Federal agency collecting surface-water data in the Nation. The collection of these data is a major activity of the Water Resources Division of the USGS. The data are collected in cooperation with State and local governments and other Federal agencies. The USGS presently (1983) operates approximately 8,000 continuous-record gaging stations throughout the Nation. Some of these records extend back to the turn of the century. Any activity of long standing, such as the collection of surface-water data, should be reexamined at intervals, if not continuously, because of changes in objectives, technology, or external constraints. The last systematic nationwide evaluation of the streamflow information program was completed in 1970 and is documented by Benson and Carter (1973). The USGS is presently undertaking another nationwide analysis of the stream-gaging program that will be completed over a 5-year period with 20 percent of the program being analyzed each year. The objective of this analysis is to define and document the most cost-effective means of furnishing streamflow information.

For every continuous-record gaging station, the analysis identifies the principal uses of the data and relates these uses to funding sources. Gaged sites for which data are no longer needed are identified, as are deficient or unmet data demands. In addition, gaging stations are categorized as to whether the data are available to users in a real-time sense, on a provisional basis, or at the end of the water year.

The second aspect of the analysis is to identify less costly methods of furnishing the needed information; among these are flow-routing models and statistical methods. The stream-gaging activity no longer is considered a network of observation points, but rather an integrated information system in which data are provided both by observation and synthesis.

The final part of the analysis involves the use of Kalman-filtering and mathematical-programming techniques to define strategies for operation of the necessary stations that minimize the uncertainty in the streamflow records for given operating budgets. Kalman-filtering techniques are used to compute uncertainty functions (relating the standard errors of computation or estimation of streamflow records to the frequencies of visits to the stream gages) for all stations in the analysis. A steepest descent optimization program uses these uncertainty functions, information on practical stream-gaging routes, the various costs associated with stream gaging, and the total operating budget to identify the visit frequency for each station that minimizes the overall uncertainty in the streamflow. The stream-gaging program that results from this analysis will meet the expressed water-data needs in the most cost-effective manner.

This report is organized into five sections; the first is an introduction to the stream-gaging activities in Maine and to the study itself. The middle three sections each contain discussions of individual steps of the analysis. Because of the sequential nature of the steps and the dependence of subsequent steps on previous results, summaries of conclusions are given at the end of each middle section. The complete study is summarized in the final section.

History of the Stream-Gaging Program in Maine

The program of surface-water investigations by the USGS in Maine has grown steadily through the years as Federal and State interest in water resources has increased. The Maine office of the USGS began collecting surface-water data after the establishment of six gaging stations in 1901. Four of these gages were in Maine and two were in New Hampshire. These first stations were operated primarily to evaluate the power potential of major rivers in the State. From this modest beginning, the program gradually expanded to the point where, in 1928, the USGS operated 26 gaging stations in the State. During the 1930's, despite the Depression, the program operated by the Maine office increased to 40 stations. Two of these new stations were supported by the U.S. Army Corps of Engineers as part of special flood studies following the major flooding that occurred in 1936. The era witnessed the shift of data needs from primarily power evaluation to design and planning of many types of hydraulic and hydrologic structures and resources assessment and management. Although the war effort from 1941 to 1946 curtailed expansion of the program, by 1960 USGS was operating 55 surface-water gaging stations in Maine.

A study of characteristics of peak flows on streams with drainage areas of less than 15 mi^2 was started in 1963. Five continuous-record gaging stations and 23 crest-stage partial-record stations were operated as part of the program. A study by Hayes and Morrill (1970) described the development of Maine's surface-water program and proposed a program to meet the future needs of water-data users. At the time of the study, the Maine program had 59 continuous and 23 partial-record stations.

Subsequent to a publication by Morrill (1975), operation of the 23 crest-stage partial-record stations was terminated. Between 1970 and 1981, 8 continuous stream gages were added to and 16 continuous stream gages were eliminated from the Maine gaging program. The decision to drop these gages was based on a Network Analysis for Regional Information (NARI) study by Morrill (written commun., 1983). These reductions leave the Maine office with 51 stations. Included in this total are five gages currently being operated as part of special projects scheduled to be completed in the near future and three gages situated in New Hampshire.

The number of continuous stream gages historically operated within the State of Maine is given in figure 1; gages in New Hampshire that are operated by the Maine office are not included in figure 1.

Current Maine Stream-Gaging Program

Maine can be divided into four major physiographic regions as noted by Prescott (1963)—the Coastal Lowlands, the Central Uplands, the Moosehead Plateau, and the Aroostook Valley. The locations of these regions and the distribution of the 51 stream gages currently operated by the Maine office of the USGS are shown on figure 2. Twenty-four gages are in the Coastal Lowlands, 12 are in the Central Uplands, 9 are in the Aroostook Valley, and 6 are in the Moosehead Plateau. Figure 2 demonstrates that the majority of the gages are in the Coastal Lowlands, the Central Uplands, and the northern portions of the Moosehead Plateau and Aroostook Valley. Large areas almost totally devoid of gaging stations are evident throughout the Moosehead Plateau and parts of the Aroostook Valley.

The cost of operating these 51 stream gages in fiscal year 1982 was $211,000.

Selected hydrologic data, including drainage area, period of record, and mean annual flow, for the 51 stations are given in table 1. Station identification numbers used throughout this report are the last five digits of the USGS's eight-digit downstream-order station number; the first three digits of the standard USGS station number for all stations used in this report are 010. Table 1 also provides the official name of each stream gage, as well as an abbreviated version of each name; abbreviated names are used in the remainder of this report.

USES, FUNDING, AND AVAILABILITY OF CONTINUOUS STREAMFLOW DATA

The relevance of a stream gage is defined by the uses that are made of data produced from the gage. The uses of the data from each gage in the Maine program were identified by a survey of known data users. The survey documented the relative importance of each gage and identified gaging stations that may be considered for discontinuation.

Data uses identified by the survey were categorized into nine classes, defined below. The sources of funding for each gage and the frequency at which data are provided to the users were also compiled.

Data-Use Classes

The following definitions were used to categorize each known use of streamflow for each continuous stream gage.

Regional Hydrology

For data to be useful in defining regional hydrology, a stream gage must be largely unaffected by manmade storage or diversion. In this class of uses, man's effects on streamflow are not necessarily small, but the effects considered are limited to those of land-use and climate changes. Large amounts of manmade storage may exist in the basin provided the outflow is uncontrolled. These stations are useful in developing regionally transferable information about the relationship between basin characteristics and streamflow.

Twenty-eight stations in the Maine network are classified in the regional hydrology data-use category. Four of the stations are special cases in that they are designated bench-mark or index stations. One hydrologic bench-mark station in Maine serves as an indicator of hydrologic conditions in watersheds relatively free of manmade alteration. Three index stations in different regions of the State are used to indicate current hydrologic conditions. The locations of stream gages that provide information about regional hydrology are given in figure 3.

Hydrologic Systems

Stations that can be used for accounting—that is, to define current hydrologic conditions and the sources, sinks, and fluxes of water through hydrologic systems including regulated systems—are designated hydrologic systems stations. They include diversions and return

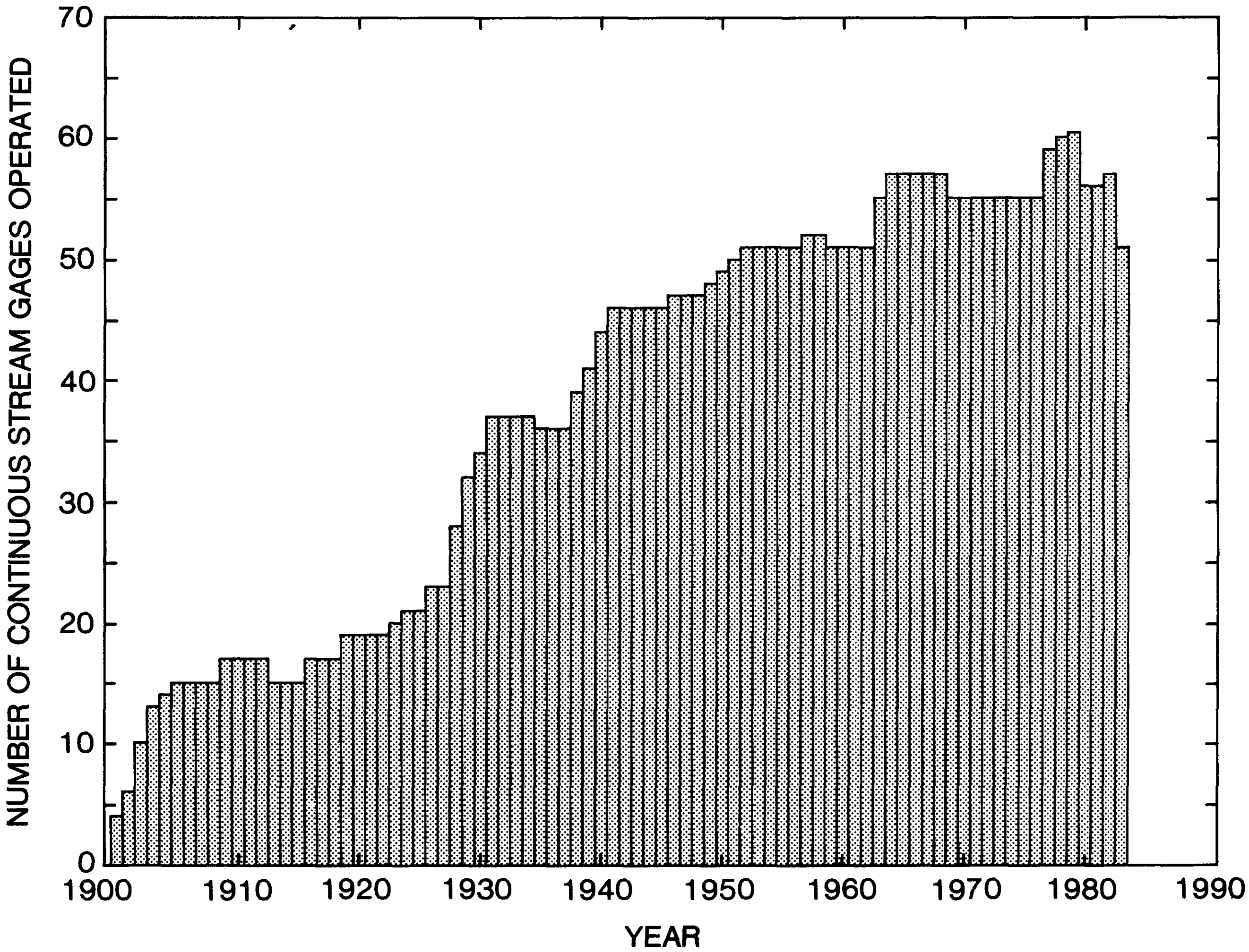

Figure 1. History of continuous stream gaging in Maine.

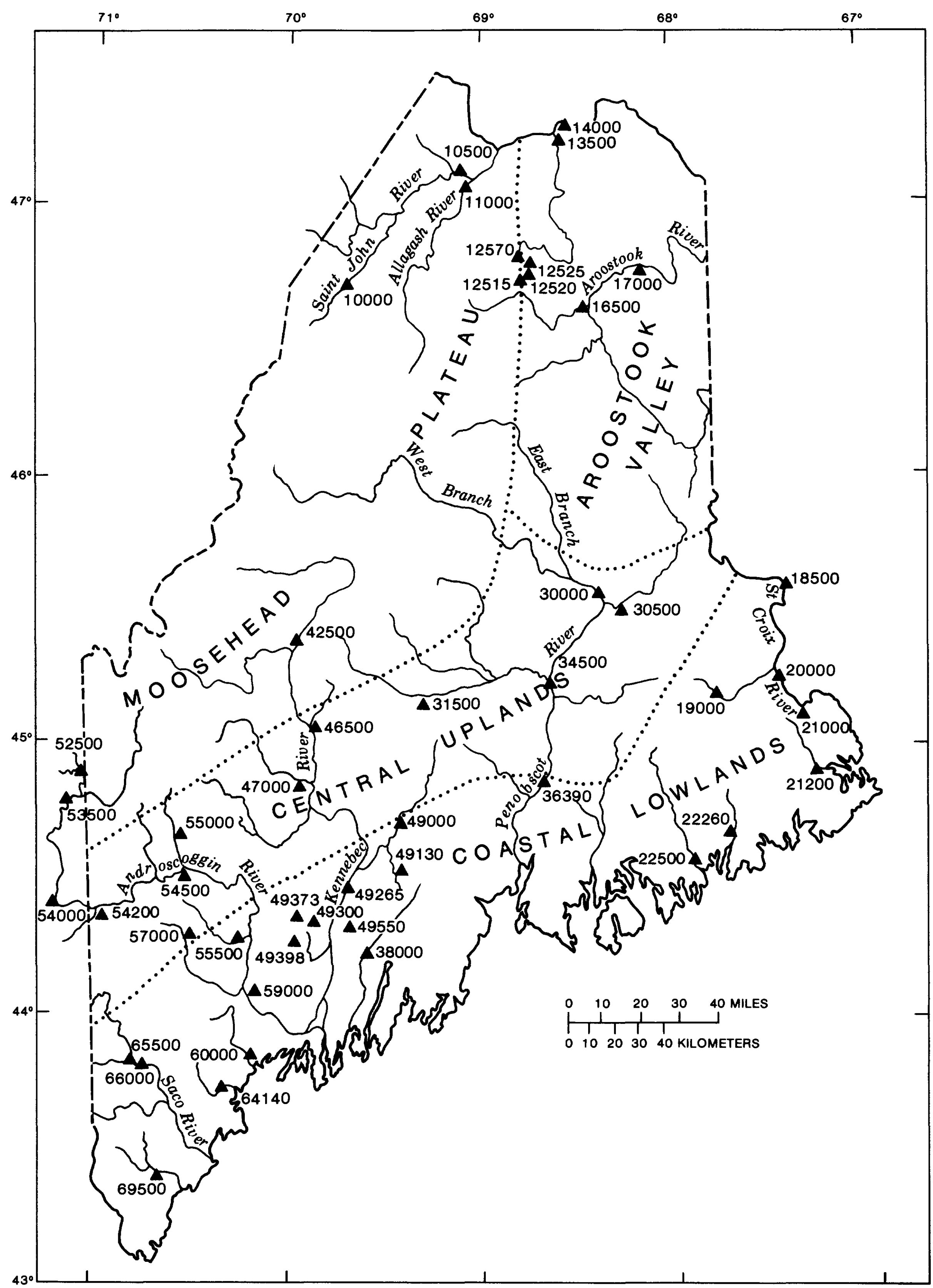

Figure 2. Locations of stream gages.

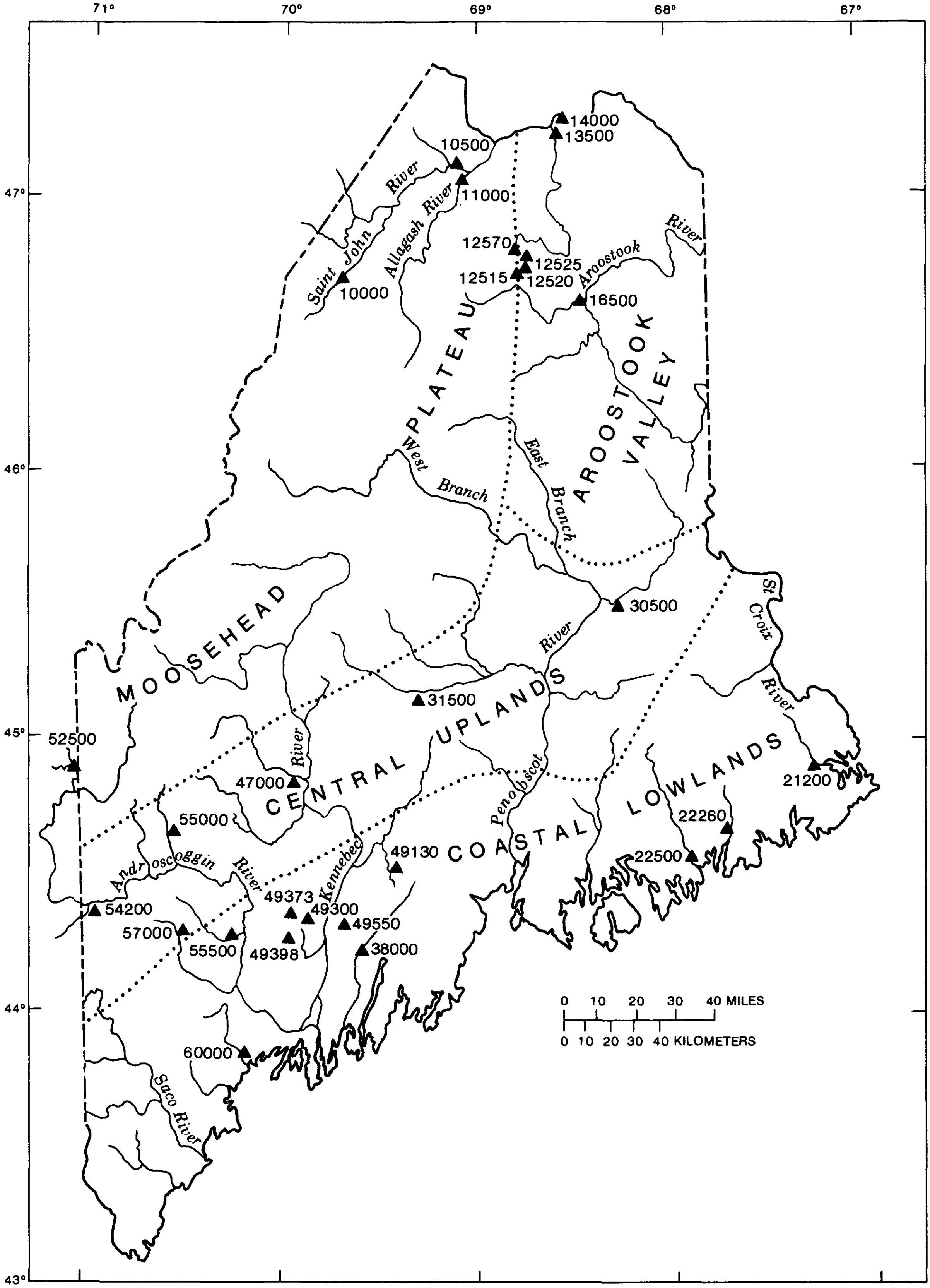

Figure 3. Locations of regional hydrology stream gages.

Table 1. Selected hydrological data for stations in the Maine surface-water program
[All stations are located in Maine except as noted]

Station no.	Station name (Abbreviated name)	Drainage area (mi^2)	Period of record	Mean annual flow (ft^3/s)
10000	St. John River at Ninemile Bridge (Ninemile)	1,341	October 1950-	2,310
10500	St. John River at Dickey (Dickey)	2,680	September 1946-	4,745
11000	Allagash River near Allagash (Allagash)	1,229	September 1931-	1,924
12515	Clayton Stream at Outlet Clayton Lake (Clayton Stream)	13.9	July 1982-	__ 1/
12520	Bald Mountain Brook near Bald Mountain (Bald Mountain)	1.69	October 1980-	__ 1/
12525	Bishop Mountain Brook near Bishop Mountain (Bishop Mountain)	1.04	November 1981	__ 1/
12570	Fish River at Inlet Fish River Lake (Fish River Lake)	70.3	July 1982-	__ 1/
13500	Fish River near Fort Kent (Fish)	873	July 1903-December 1908 and May 1911 November 1911- 2/ September 1929-	1,402
14000	St. John River below Fish River, at Fort Kent (Fort Kent)	5,665	October 1926- 3/	9,682
15800	Aroostook River near Masardis (Masardis)	892	September 1957-	1,505
16500	Machias River near Ashland (Machias)	329	June 1951-	570
17000	Aroostook River at Washburn (Washburn)	1,654	August 1930-	2,647
18500	St. Croix River at Vanceboro (Vanceboro)	413	October 1928-	720
19000	Grand Lake Stream at Grand Lake Stream (Grand Lake Stream)	227	October 1982-	396
20000	St. Croix River near Baileyville (Baileyville)	1,315	October 1919-	2,344
21000	St. Croix River at Baring (Baring)	1,374	October 1958-	2,732
21200	Dennys River at Dennysville (Dennysville)	92.9	October 1955-	195
22260	Pleasant River near Epping (Epping)	60.6	July 1980-	__ 1/

See footnotes at end of table.

Table 1. Selected hydrological data for stations in the Maine surface-water program—Continued

Station no.	Station name (Abbreviated name)	Drainage area (mi^2)	Period of record	Mean annual flow (ft^3/s)
22500	Narraguagus River at Cherryfield (Cherryfield)	227	February 1948-	500
30000	Penobscot River near Mattawamkeag (Mattaseunk)	3,356	June 1940-	5,756
30500	Mattawamkeag River near Mattawamkeag (Mattawamkeag)	1,418	October 1934-	2,491
31500	Piscataquis River near Dover-Foxcroft (Dover-Foxcroft)	298	August 1902-	602
34500	Penobscot River at West Enfield (West Enfield)	6,671	November 1901-	11,889
36390	Penobscot River at Eddington (Eddington)	7,764	Aprril 1979-	___ 1/
38000	Sheepscot River at North Whitefield (North Whitefield)	148	October 1938-	246
42500	Kennebec River at The Forks (The Forks)	1,589	September 1901-	2,596
46500	Kennebec River at Bingham (Bingham)	2,715	June 1907- June 1910/, October 1930-	4,415
47000	Carrabassett River near North Anson (North Anson)	353	June 1902-May 1907/, August 1925-	715
47730	Wilson Stream at East Wilton (Wilson Stream)	45.8	February 1977-	___ 1/
49000	Sebasticook River near Pittsfield (Pittsfield)	572	October 1928-	953
49130	Johnson Brook at South Albion (Johnson Brook)	2.92	May 1980-	___ 1/
49265	Kennebec River at North Sidney (Sidney)	5,403	October 1978	___ 1/
49300	North Branch Tanning Brook near Manchester (Tanning)	0.93	November 1963	2.01
49373	Mill Stream at Winthrop (Mill)	32.7	October 1977-	___ 1/
49396	Jock Stream at South Monmouth (Jock)	13.7	October 1977-	___ 1/
49500	Cobbosseecontee Stream at Gardiner (Cobbossee)	217	June 1890-September 1964/October 1976-	341

See footnotes at end of table.

Table 1. Selected hydrological data for stations in the Maine surface-water program—Continued

Station no.	Station name (Abbreviated name)	Drainage area (mi^2)	Period of record	Mean annual flow (ft^3/s)
49550	Togus Stream at Togus (Togus Stream)	23.7	October 1981-	— 1/
52500	Diamond River near Wentworth Location, N.H. (Diamond)	152	July 1941-	349
53500	Androscoggin River at Errol, N.H. (Errol)	1,046	January 1905- 4/	1,900
54000	Androscoggin River near Gorham, N.H. (Gorham)	1,361	October 1913- 5/	2,458
54200	Wild River at Gilead (Gilead)	69.6	July 1964-	179
54500	Androscoggin River at Rumford (Rumford)	2,069	May 1892-	3,707
55000	Swift River near Roxbury (Roxbury)	96.9	June 1929-	198
55500	Nezinscot River at Turner Center (Turner)	169	1941-	303
57000	Little Androscoggin River near South Paris (South Paris)	75.8	September 1913-April 1924/ October 1931-	138
59000	Androscoggin River near Auburn (Auburn)	3,263	October 1928-	6,119
60000	Royal River at Yarmouth (Royal)	141	October 1949-	273
64140	Presumpscot River near West Falmouth (West Falmouth)	598	October 1975-	974
65500	Ossipee River at Cornish (Ossipee)	452	July 1916-	874
66000	Saco River at Cornish (Cornish)	1,293	June 1916-	2,700
69500	Mousam River near West Kennebunk (Mousam)	99.0	October 1939-	179

See footnotes at end of table.

Table 1. Selected hydrological data for stations in the Maine surface-water program—Continued

1/ No mean annual flow published; less than 5 years of streamflow record.

2/ Published as "at Wallagrass."

3/ Prior to October 1931 published as "at Fort Kent."

4/ Prior to 1922 published as "at Errol Dam."

5/ October 1922 to February 1929, monthly discharge only, published in Water-Supply Paper 1301. Prior to October 1928, published as "at Berlin."

flows and stations that are useful for defining the interaction of water systems.

The bench-mark and index stations are included in the hydrologic systems category because they account for current and long-term conditions of the hydrologic systems that they gage. Federal Energy Regulatory Commission (FERC) stations and international gaging stations, on significant rivers that cross national boundaries, also are included. The four international stations in Maine are located on the St. John and St. Croix Rivers, which flow into Canada, and provide data for the proper management of potentially conflicting uses of the river's resources by both countries.

The data collected at the 14 FERC stations are used to monitor the compliance of control structures with downstream flow requirements determined by FERC.

Two other stations are included in this category and are operated to ensure the compliance of wastewater-treatment plants with State-issued permits.

Legal Obligations

Some stations provide records of flows for the verification or enforcement of treaties, compacts, and decrees. The legal obligation category contains only those stations that USGS must operate to satisfy a legal responsibility.

No stations in the Maine program exist to fulfill a legal responsibility of the USGS.

Planning and Design

Gaging stations in this category of data use are used for the planning and design of a specific project (for example, a dam, levee, floodwall, navigation system, water-supply diversion, hydropower plant, or waste-treatment facility) or group of structures. The planning and design category is limited to those stations that were instituted for such purposes and for which this purpose is still valid.

Currently, no stations in the Maine program are operated for planning or design purposes.

Project Operation

Gaging stations in this category are used, on an ongoing basis, to assist water managers in making operational decisions such as reservoir releases, hydropower operations, or diversions. The project operation use generally implies that data are routinely available to operators on a rapid-reporting basis. For projects on large streams, data may only be needed every few days.

There are 18 stations in the Maine program that are used in this manner. Sixteen of these are used to aid operators in the management of reservoirs and control structures that are part of hydropower production systems. The remaining two stations are used to assist wastewater-treatment plant operators.

Hydrologic Forecasts

Gaging stations in this category are regularly used to provide information for hydrologic forecasting. Such information might include flood forecasts for a specific river reach, or periodic (daily, weekly, monthly, or seasonal) flow-volume forecasts for a specific site or region. The hydrologic forecast use generally implies that the data are routinely available to the forecasters on a rapid-reporting basis. On large streams, data may only be needed every few days.

Stations in the Maine program that are included in the hydrologic forecast category are those used for flood forecasting and for forecasting inflows to reservoirs that are a part of hydropower generating systems. Data are used by the U.S. National Weather Service (NWS); the

Flood Forecast Center of Fredericton, New Brunswick, Canada; and the Water Survey of Canada to predict floodflows at downstream sites. Additionally, NWS uses the data at some stations as input to longer range prediction models of the probability of snowmelt floods.

Water-Quality Monitoring

Gaging stations where regular water-quality or sediment-transport monitoring is conducted and where the availability of streamflow data contributes to the utility or is essential to the interpretation of the water-quality or sediment data are designated as water-quality-monitoring sites.

One such station in the program is a designated bench-mark station and five are National Stream Quality Accounting Network (NASQAN) stations. Water-quality samples from bench-mark stations are used to indicate water-quality characteristics of streams that have been and probably will continue to be relatively free of man-made influence. NASQAN stations are part of a country-wide network designed to assess water-quality trends of significant streams.

Research

Gaging stations in this category are operated for specific research or water-investigations studies. Typically, these are operated for only a few years.

Eight stations in the Maine program are used in support of research activities, including a phosphorus-loading study and a project to assess the impacts of a proposed copper mine. The State of Maine, the Department of Environmental Protection, and the University of Maine use data from a number of sites for research activities that involve phosphorus loading, waste-load allocation, and lake restoration.

Other

In addition to the eight data-use classes described above, five stations are used to provide streamflow information for recreational planning, primarily for canoceists, rafters, and fishermen.

Funding

The four sources of funding for the streamflow-data program are as follows:

1. *Federal program.*—Funds that have been directly allocated to USGS.
2. *Other Federal Agency (OFA) program.*—Funds that have been transferred to the USGS by OFA's.
3. *Coop program.*—Funds that come jointly from USGS cooperative-designated funding and from a non-Federal cooperating agency. Cooperating agency funds may be in the form of direct services or cash.
4. *Other non-Federal.*—Funds that are provided entirely by a non-federal agency or a private concern under the auspices of a Federal agency. In this study, funding from private concerns was limited to licensing and permitting requirements for hydropower development by the Federal Energy Regulatory Commission. Funds in this category are not matched by USGS cooperative funds.

In all four categories, identified sources of funding pertain only to the collection of streamflow data; sources of funding for other activities, particularly collection of water-quality samples, that might be carried out at the site are not necessarily the same as those identified herein.

Seventeen entities currently contribute funds to the Maine stream-gaging program.

Frequency of Data Availability

Frequency of data availability refers to the times at which streamflow data may be furnished to the users. In this category, three distinct possibilities exist. Data can be furnished by direct-access telemetry equipment for immediate use, by periodic release of provisional data, or in publication format through annual data reports published by the USGS for Maine (U.S. Geological Survey, 1981). These three categories are designated T, P, and A, respectively, in table 2. In the current Maine program, data for all 51 stations are made available through annual reports, data from 21 stations are available on a real-time basis, and data are released on a provisional basis at 12 stations.

Data-Use Presentation

Data-use and ancillary information are presented for each continuous gaging station in table 2. The entry of an asterisk in the table indicates that the station is used by the Geological Survey for regional hydrology purposes, and (or) the station is operated from Federal funds appropriated directly to the Survey.

Conclusions Pertaining to Data Uses

Concurrently with this study, the NARI procedure is being applied in Maine. NARI is a procedure for

Table 2. Data-use table

STATION NUMBER	DATA USE: REGIONAL HYDROLOGY	HYDROLOGIC SYSTEMS	LEGAL OBLIGATIONS	PLANNING & DESIGN	PROJECT OPERATION	HYDROLOGIC FORECASTS	WATER-QUALITY MONITORING	RESEARCH	OTHER	FUNDING: FEDERAL PROGRAM	OFA PROGRAM	CO-OP PROGRAM	OTHER NON-FEDERAL	FREQUENCY OF DATA AVAILABILITY
10000	*				1	1 2			3	*				AT
10500	*				1	1 4 2			3			5		AT
11000	*				1	1 4			3			5		AT
12515	*							6				5		A
12520	*							6				5		A
12525	*							6				5		A
12570	*							6				5		A
13500	*				1	1 4 2			3			5		AT
14000	7	7 8			1	1 4 2			3		9			ATP
15800		10			11	11 4						5	11	AT
16500	*									*				A
17000		10			11	11 4						5	11	AT
18500		8			12	12					9			A P
19000		10			12								12	A
20000		8			12	12						5		A P
21000		8			1	4					9			ATP
21200	*					4						5		AT
22260	*											5		A
22500	*						13					5		A
30000		10											14	A
30500	*											5		A
31500	7	7				4						5		ATP
34500		10			15	15 4							15	A
36390						4	13					5		AT
38000	*					4						5		A P
42500		10											16	A
46500		10											17	A
47000	*											5		A P
47730		18			18							19		A P
49000								20 21				5		A
49130	*							22				5		A
49265						4	13					5		AT
49300	*											5		A
49373	*											5 23		A
49396	*							24				5 23		A P
49500		10											25 26	A P
49550	*	18			18						27	5		A
52500	*				28	28						5		AT
53500		10			28	28							28	A
54000		10			28	28							28	A
54200	29	29				4	29			*				AT
54500		10				4							30	AT
55000	*					4						5		AT
55500	*					4						5		ATP
57000	7	7				4						5		ATP
59000		10			28	4							28	AT
60000	*					4						5		AT
64140		10					13					5	31	AT
65500		32										5		A
66000		10					13						16	A
69500								20				5		A

See footnotes at end of table.

Table 2. Data-use table—continued

1. New Brunswick (Canada) Electric Power Commission hydropower system operation.
2. Flood forecasting - Flood Forecast Center, Fredericton, N.B., Canada and Water Survey of Canada.
3. Streamflow data requests for recreational planning.
4. Flood forecasting - U.S. National Weather Service.
5. State of Maine, Maine Geological Survey, program co-ordinator for most state agencies.
6. Copper mine hydrology project.
7. Long-term index gaging station.
8. International gaging station, Boundary Waters Treaty of 1909.
9. International Joint Commission (State Dept.).
10. Federal Energy Regulatory Commission hydropower licensing requirements.
11. Maine Public Service Co. hydropower system operation.
12. Georgia-Pacific Co. hydropower system operetion.
13. NASQAN station.
14. Great Northern Paper Company.
15. Bangor Hydroelectric Co. hydropower system operation.
16. Central Maine Power Company.
17. Kennebec Water Power Company.
18. Maine Department of Environmental Protection waste effluent discharge permic requirements.
19. Town of Wilton, Maine.
20. Water quality research activities - State of Maine Dept. of Environmental Protection.
21. Water quality research activities - University of Maine, Orono, Maine.
22. Johnson Brook phosphorus loading study.
23. Cobbossee Watershed District.
24. Phosphorus loading - Cobbosseecontee Lake.
25. Swift River Company.
26. Gardiner Water District.
27. United States Veterans Administration.
28. Union Water Power Co. hydropower system operation.
29. Hydrologic benchmark stations.
30. Boise-Cascade Corp.
31. S.D. Warren Company.
32. Determination of inflow between control structure and downstream gage.

identifying the contributions to error reduction in a regional regression analysis of statistical characteristics of streamflow that can be expected from future stream-gaging activities. These activities include extending data collection at existing stream gages, establishing new stream gages, or various combinations of these activities (Moss and others, 1982). Preliminary results of the NARI analysis in Maine indicate that accuracy goals for regional streamflow estimates established by Carter and Benson (1970) were met for four of the six characteristics investigated. The two for which goals were not achieved were annual minimum 7-day mean flows and peak flows with 50-year recurrence intervals (R.A. Morrill, written commun., 1983). Attempts to improve the relationships for these two characteristics failed in part because of the poor spatial distribution of existing gages on unregulated streams as demonstrated in figure 3. The USGS has operated only 36 stream gages, each with more than 15 years of record, on unregulated streams. This small sample restricted efforts to develop individual regressions for each physiographic province. Therefore, the authors of the NARI study suggest that additional stream gages be established whenever possible on unregulated streams throughout Maine, but especially in the western part of the Coastal Lowlands and central and western parts of the Moosehead Plateau, where streamflow information is particularly sparse (R.A. Morrill, written commun., 1983).

A review of the data-use and funding information presented in table 2 indicates that five stations are currently operated solely to support short-term hydrologic studies. Gaging stations at Clayton Stream (12515), Bald Mountain (12520), Bishop Mountain (12525), and Fish River Lake (12570) are run as part of a study of the hydrologic impacts of a proposed open-pit copper mine. Johnson Brook (49130) is operated as part of a phosphorus runoff study.

Three of the four copper-mine-study sites, Bald Mountain, Bishop Mountain, and Fish River Lake, will be affected by operation of the proposed mine. Resulting regulation of flow would prevent use of data from these sites in a regression analysis to determine streamflow characteristics. Therefore, operation of these stations should be terminated at the end of the current project.

The fourth station in the copper mine study, Clayton Stream, is located above the proposed mine impact area; data from this site therefore will not be affected by regulation. Based on the drainage area and location of this station, it should be continued in operation beyond the life of the research project. Similarly, the Johnson Brook basin is not subject to flow regulation above the gaging station, and this station also should be kept in operation after the research project is terminated.

Three stream gages, Machias (16500), North Whitefield (38000), and Tanning (49300), are operated primarily for regional hydrologic uses (table 2). No interest was expressed in funding these stations beyond the current year, and the NARI study indicates that additional data from these sites would not improve significantly the current regional regression equations (R.A. Morrill, written commun., 1983). Therefore, these stations should be terminated at the end of the current water year (1983).

Based on the above conclusions, stream gages on Clayton Stream and Johnson Brook will be included for analysis in the following sections of this report. The stations on Bald Mountain, Bishop Mountain, Fish River Lake, Machias, North Whitefield, and Tanning will not be considered further in this report.

ALTERNATIVE METHODS OF DEVELOPING STREAMFLOW INFORMATION

The second step of the analysis of the stream-gaging program is to investigate alternative methods of providing daily streamflow information in lieu of operating continuous-flow gaging stations. The objective of the analysis is to identify gaging stations where alternative technology, such as flow-routing or statistical methods, will provide information about daily mean streamflow in a more cost-effective manner than operating a continuous stream gage. No guidelines concerning suitable accuracies exist for particular uses of the data; therefore, judgment is required in deciding whether the accuracy of the estimated daily flows is suitable for the intended purpose. The data uses at a station will influence whether a site has potential for alternative methods. For example, stations for which real-time flood hydrographs are required, such as hydrologic forecasts and project operation, are not candidates for the alternative methods. Likewise, legal obligation to operate a gaging station would preclude utilizing alternative methods. The primary candidates for alternative methods are stations operated upstream or downstream of other stations on the same stream. The accuracy of the estimated streamflow at these sites may be suitable because of the high redundancy of flow information between sites. Similar watersheds, in the same physiographic and climatic area, also may have potential for alternative methods.

All stations in the Maine stream-gaging program were categorized as to their potential utilization of alternative methods and selected mehods were applied at four subsequent sections of this report. This section briefly describes the two alternative methods used in the Maine analysis and documents why these methods were chosen.

Because of the short timeframe of this analysis, only two methods were considered. Desirable attributes of a proposed alternative method are (1) the proposed method should be computer oriented and easy to apply, (2) the proposed method should have an available interface with the USGS WATSTORE Daily Values File (Hutchinson, 1975), (3) the proposed method should be technically sound and generally acceptable to the hydrologic community, and (4) the proposed method should permit easy evaluation of the accuracy of the simulated streamflow records. The desirability of the first attribute above is obvious. Second, the interface with the WATSTORE Daily Values File is needed to easily calibrate the proposed alternative method. Third, the alternative method selected for analysis must be technically sound or it will not be able to provide data of suitable accuracy. Fourth, the alternative method should provide an estimate of the accuracy of the streamflow to judge the adequacy of the simulated data. The above selection criteria were used to select two methods—a flow-routing model and multiple-regression analysis.

Description of Flow-Routing Model

Hydrologic flow-routing mehtods use the law of conservation of mass and the relationship between storage in a reach and outflow from the reach. The hydraulics of the system are not considered. The method usually requires only a few parameters and treats the reach in a lumped sense without subdivision. The input is usually a discharge hydrograph at the upstream end of the reach and the output a discharge hydrograph at the downstream end. Several different types of hydrologic routing, such as Muskingum, Modified Puls, Kinematic Wave, and the unit-response flow-routing method, are available. The latter method was selected for this analysis. This method uses two techniques—storage continuity (Sauer, 1973) and diffusion analogy (Keefer, 1974; Keefer and McQuivey, 1974). These concepts are discussed below.

The unit-response method was selected because it fulfilled the criteria noted above. Computer programs for the unit-response method can be used to route streamflow from one or more upstream locations to a downstream location. Downstream hydrographs are produced by the convolution of upstream hydrographs with their appropriate unit-response functions. This method can only be applied at a downstream station where an upstream station exists on the same stream. An advantage of this model is that it can be used for regulated stream systems. Reservoir routing techniques are included in the model so flows can be routed through reservoirs if the operating rules are known. Calibration and verification of the flow-routing model is achieved using observed upstream and downstream hydrographs and estimates of tributary inflows. The convolution model treats a stream reach as a linear one-dimensional system in which the system output (downstream hydrograph) is computed by multiplying (convoluting) the ordinates of the upstream hydrograph by the unit-response function and lagging them appropriately. The model has the capability of combining hydrographs, multiplying a hydrograph by a ratio, and changing the timing of a hydrograph. In this analysis, the model is only used to route an upstream hydrograph to a downstream point. Routing can be accomplished using hourly data, but only daily data are used in this analysis.

Three options are available for determining the unit (system) response function. Selection of the appropriate option depends primarily upon the variability of wave celerity (traveltime) and dispersion (channel storage) throughout the range of discharges to be routed. Adequate routing of daily flows can usually be accomplished using a single unit-response function (linearization about a single discharge) to represent the system response. However, if the routing coefficients vary drastically with discharge, linearization about a low-range discharge results in overestimated high flows that arrive late at the downstream site; whereas linearization about a high-range discharge results in low-range flows that are underestimated and arrive too soon. A single unit-response function may not provide acceptable results in such cases. Therefore, the option of multiple linearization (Keefer and McQuivey, 1974), which uses a family of unit-response functions to represent the system response, is available.

Determination of the system's response to the input at the upstream end of the reach is not the total solution for most flow-routing problems. The convolution process makes no accounting of flow from the intervening area between the upstream and downstream locations. Such flows may be unknown or estimated by some combination of gaged and ungaged flows. An estimating technique that should prove satisfactory in many instances is the multiplication of known flows at an index gaging station by a factor (for example, a drainage-area ratio).

The objective in both the storage-continuity and the diffusion analogy flow-routing method is to calibrate two parameters that describe the storage-discharge relationship in a given reach and the travel time of flow passing through the reach. In the storage-continuity method, a response function is derived by modifying a translation hydrograph technique developed by Mitchell (1962) to apply to open channels. A triangular pulse (Sauer, 1973) is routed through reservoir-type storage and then transformed by a summation curve technique to a unit response of desired duration. The two parameters that describe the routing reach are K_s, a storage coefficient that is the slope of the storage-discharge relation, and W_s, the translation hydrograph time base. These two parameters determine the shape of the resulting unit-response function.

In the diffusion-analogy theory, the two parameters requiring calibration in this method are K_o, a wave dispersion or damping coefficient, and C_o, the floodwave celerity. K_o controls the spreading of the wave (analogous to K_s in the storage-continuity method) and C_o controls the travel time (analogous to W_s in the storage-continuity method). In the single-linearization method, only one K_o and one C_o value are used. In the multiple-linearization method, C_o and K_o are varied with discharge so a table of wave celerity (C_o) versus discharge (Q) and a table of dispersion coefficient (K_o) versus discharge (Q) are used.

In both the storage-continuity and diffusion-analogy methods, the two parameters are calibrated by trial and error. The analyst must decide whether suitable parameters have been derived by comparing the simulated discharge with the observed discharge.

Description of Regression Analysis

Simple- and multiple-regression techniques can also be used to estimate daily flow records. Regression equations can be computed that relate daily flows (or their logarithms) at a single station to daily flows at a combination of upstream, downstream, and (or) tributary stations. This statistical method is not limited, like the flow-routing method, to stations where an upstream station exists on the same stream. The explanatory variables in the regression analysis can be stations from different watersheds, or downstream and tributary watersheds. The regression method has many of the same attributes as the flow-routing method in that it is easy to apply, provides indices of accuracy, and is generally accepted as a good tool for estimation. The theory and assumptions of regression analysis are described in several textbooks such as those by Draper and Smith (1966) and Kleinbaum and Kupper (1978). The application of regression analysis to hydrologic problems is described and illustrated by Riggs (1973) and Thomas and Benson (1970). Only a brief description of regression analysis is provided in this report.

A linear regression model of the following form was developed for estimating daily mean discharges in Maine:

$$y_i = B_o + \sum_{j=1}^{p} B_j x_j + e_i$$

where

y_i is daily mean discharge at station i (dependent variable);

x_j is daily mean discharges at nearby stations (explanatory variables);

B_o and B_j is regression constant and coefficients;

and

e_i is the random error term.

The above equation is calibrated (B_o and B_j are estimated) using observed values of y_i and x_j. These observed daily mean discharges can be retrieved from the WATSTORE Daily Values File. The values of x_j may be discharges observed on the same day as discharges at station i or may be for previous or future days, depending on whether station j is upstream or downstream of station i. Once the equation is calibrated and verified, future values of y_i are estimated using observed values of x_j. The regression constant and coefficients (B_o and B_j) are tested to determine whether they are significantly different from zero. A given station j should only be retained in the regression equation if its regression coefficient (B_j) is significantly different from zero. The regression equation should be calibrated using one period of time and then verified or tested for a different period of time to obtain a measure of the true predictive accuracy. Both the calibration and verification period should be representative of the range of flows that could occur at station j. The equation should be verified by plotting the residuals e_i (difference between simulated and observed discharges) against the dependent and all explanatory variables in the equation, and by plotting the simulated and observed discharges versus time. These tests are intended to determine whether the linear model is appropriate or some transformation of the variables is needed and whether there is any bias in the equation such as overestimating low flows. These tests might indicate, for example, that a logarithmic transformation is desirable, that a nonlinear regression equation is appropriate, or that the regression equation is biased in some way. In this report these tests indicated that a linear model with y_i and x_j, in cubic feet per second, was appropriate. The application of linear-regression techniques to four watersheds in Maine is described in a subsequent section of this report.

It should be noted that the use of a regression relation to synthesize data at a discontinued gaging station entails a reduction in the variance of the streamflow record relative to that which would be computed from an actual record of streamflow at the site. The reduction in variance expressed as a fraction is approximately equal to one minus the square of the correlation coefficient that results from the regression analysis.

Categorization of Stream Gages by Their Potential for Alternative Methods

An analysis of the data uses presented in table 2 identified four stations at which alternative methods for providing the needed streamflow information could be applied. These four stations are Cherryfield (22500), Eddington (36390), Sidney (49265), and Ossipee (65500). Based on the capabilities and limitations of the methods and data availability, flow-routing techniques were used only at the Eddington and Ossipee gaging stations. Regression methods were applied to all four sites.

Eddington Flow-Routing Analysis

The purpose of this flow-routing analysis is to investigate the potential for use of the unit-response model for streamflow routing to simulate daily mean discharges at Eddington (36390). A schematic diagram of the Penobscot River study area is presented in figure 4. In this application, a best-fit model for the entire flow range is the desired product. Streamflow data available for this analysis are summarized in table 3.

Table 3. Gaging stations used in the Eddington flow-routing study

Station no.	Station name	Drainage area (mi^2)	Period of record
30500	Mattawamkeag	1,418	October 1934-present
31500	Dover-Foxcroft	298	August 1902-present
34500	West Enfield	6,671	November 1901-present
36390	Eddington	7,764	April 1979-present

The Eddington gage is located 32.7 mi downstream from the next upstream stream gage, West Enfield. In this reach, there are several small run-of-the-river impoundments which, under normal operating conditions, discharge approximately the inflow they receive. During low-flow periods when the dams are being operated, they can have a significant effect on streamflow at the Eddington gage. The intervening drainage area between West Enfield and Eddington is 1,093 mi², or 14 percent of the total drainage area contributing to the Eddington site. There are no stream gages within this 1,093 mi² intervening area. Another limitation on this analysis is the short period of available streamflow data at the Eddington gage.

To simulate the daily mean discharges, flows were routed from West Enfield to Eddington using the diffusion analogy method with a single linearization. The intervening drainage area was accounted for by using data from stations at Mattawamkeag and Dover-Foxcroft adjusted by drainage area ratios. The total discharge at Eddington was the summation of the routed discharge from West Enfield and adjusted discharge from Mattawamkeag and Dover-Foxcroft. The entire data set available for the Eddington site, water years 1980–82, was used to calibrate the model.

To route flow from West Enfield to Eddington, it was necessary to determine the model parameters C_o

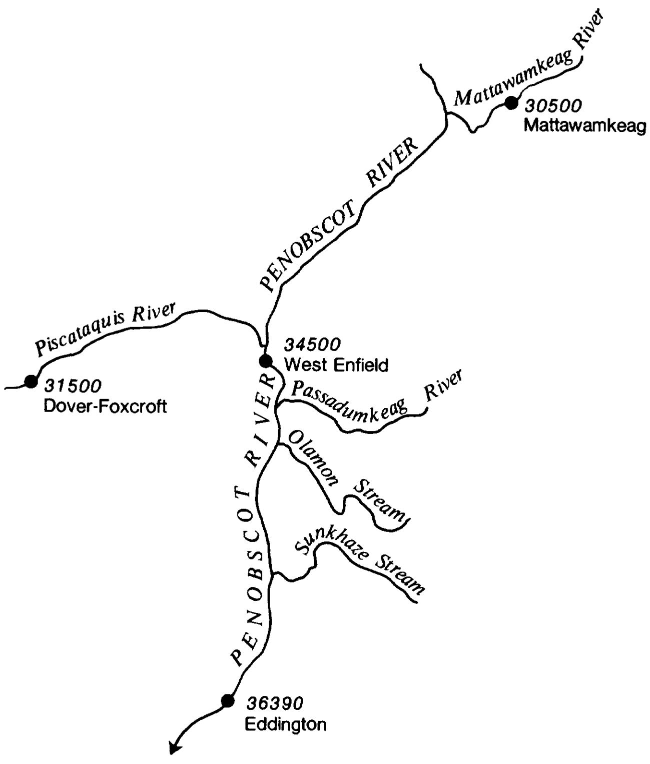

Figure 4. The Eddington study area.

(floodwave celerity) and K_o (wave-dispersion coefficient). The coefficients C_o and K_o are functions of channel width (W_o) in feet, channel slope (S_o) in feet per foot (ft/ft), the slope of the stage discharge relation (dQ_o/dY_o) in square feet per second (ft²/s), and discharge (Q_o) in cubic feet per second (ft³/s) representative of the reach in question and are determined as follows:

$$C_o = \frac{1}{W_o} \frac{dQ_o}{dY_o} \tag{1}$$

$$K_o = \frac{Q_o}{2\, S_o\, W_o} \,. \tag{2}$$

The discharge, Q_o, for which initial values of C_o and K_o were linearized, was the mean daily discharge for the West Enfield and Eddington gages as published for the 1981 water year (U.S. Geological Survey, 1981). The channel width, W_o, was calculated as the average for the 32.7-mile reach between the sites, and was measured from topographic maps. Channel slope, S_o, was determined by converting the corresponding gage heights of the initial discharges, Q_o, taken from the stage-discharge relationships at each gage to a common datum. The difference between these values was then divided by channel length to obtain a slope. The slope of the stage discharge relations, dQ_o/dY_o, was determined from the rating curves at each gage by using a 1-foot increment that bracketed the mean discharge, Q_o. The difference in the discharge through the 1-foot increment therefore represents the slope of the function at that point. The model parameters as determined above are listed in table 4.

Table 4. Selected reach characteristics used in the Eddington flow-routing study

Site	Q_o (ft³/s)	W_o (ft)	S_o (ft/ft)	$\frac{dQ_o}{dY_o}$ (ft²/s)	C_o (ft/s)	K_o (ft²/s)
West Enfield	14,110	1,240	6.896×10^{-4}	3,860	3.11	8,250
Eddington	14,420			3,100	2.50	8,430

For the first routing trial, average values for the model parameters, C_o=2.80 and K_o=8,340, were used. To simulate the intervening drainage area of 1,093 mi², an analysis was made of the general characteristics of the basins involved. These characteristics were then compared to those of the nearest stream gages, those at Dover-Foxcroft and Mattawamkeag. It was noted that the Passadumkeag, Sunkhaze, and Olamon River basins contained large percentages of swamps and lakes and were systems that responded relatively slowly during runoff events. These characteristics are traits of the Mattawamkeag River basin. A drainage area ratio calculated by dividing the combined area of the Passadumkeag, Sunkhaze, and Olamon River basins, 547 mi^2 (Fontaine, 1981), by the drainage area at the Mattawamkeag gage, 1,418 mi^2 (547/1,418=0.39), was applied to flows at the Mattawamkeag gage to simulate input from this portion of the ungaged intervening drainage. The remaining portion of the ungaged intervening drainage, 546 mi^3, is comprised of basins with less storage and larger slopes. These basins tend to respond faster during runoff events and are more accurately approximated by using the Dover-Foxcroft gage adjusted by the drainage area adjustment factor of 546 mi^2 divided by 298 mi^2, or 1.83.

Using the entire 3 water years of available data from Eddington as a calibration data set, several trials were made, adjusting both the values of C_o, K_o, and the drainage area adjustment factors. The best-fit single linearization model was determined to be that with a C_o=3.00, K_o=8,340 and the originally determined drainage area adjustment ratios. Attempts were made to improve the model using multiple linearization, splitting the year into open-water and backwater periods, and other stations to simulate intervening drainage. None of the combinations resulted in a better model for the calibration data set.

A summary of the simulation of mean daily discharge at Eddington for the 3 water years, 1980 through 1982, is given in table 5.

Table 5. Results of routing model for Eddington

Mean absolute error for 1,096 days	=	6.01 percent
Mean negative error (662 days)	=	-5.98 percent
Mean positive error (434 days)	=	6.06 percent
Total volume error	=	-1.29 percent
49 percent of the total observations had errors	≤	5 percent
85 percent of the total observations had errors	≤	10 percent
95 percent of the total observations had errors	≤	15 percent
98 percent of the total observations had errors	≤	20 percent
99 percent of the total observations had errors	≤	25 percent
1 percent of the total observations had errors	>	25 percent

This summary includes both periods of winter backwater and days of low flow when the run-of-the-river dams were exerting a strong influence on the discharge. By isolating the winter backwater portions of the 3 water years, it can be noted that these 284 days have a mean error of 7.32 percent and only 75 percent of the observations have prediction errors less than or equal to 10 percent. The remaining 812 days have a mean error of 5.54 percent, and 89 percent of the observations have prediction errors less than or equal to 10 percent. Of the 89 days (11 percent times 812 days) that have errors greater than 10 percent, 11 are days during which operational practices at the dams caused significant effects at Eddington.

Figure 5 is a comparison of the observed and simulated discharge for the Eddington gage during a spring high-water event. The fit for this period is very good with the exception of March 18 when the upstream dam exerted a substantial effect on the mean discharge for that day. Figure 6 is a comparison of the observed and simulated discharge for the Eddington gage during a late summer low-flow period. Again, this plot indicates the good fit of the routing model during open-water periods when there is no regulating effect by the upstream dams.

Ossipee Flow-Routing Analysis

A schematic diagram of the Ossipee study area is presented in figure 7. Gaging station data available for this analysis are summarized in table 6.

Table 6. Gaging stations used in the Ossipee flow-routing study

Station no.	Station name	Drainage area (mi^2)	Period of record
57000	South Paris	75.8	September 1913-April 1924/ October 1931-
60000	Royal	141	October 1949-
64500	Saco River near Conway, N.H.	385	February 1929-
65000	Ossipee River at Effingham Falls, N.H.	330	September 1942-
65500	Ossipee	452	July 1916-
66000	Cornish	1,293	June 1916-

The Ossipee gage (65500) is 17.6 mi downstream from the next upstream stream gage on the Ossipee River at Effingham Falls, and this reach is not subjected to any regulation. The intervening drainage area between Effingham Falls and Ossipee is 122 mi^2, or 27 percent of the total drainage area contributing to the Ossipee site. No stream gages are located within this 122-mi^2 area.

The approach used in this analysis was to route the flow downstream from Effingham Falls to Ossipee using the diffusion analogy method with single linearization. The intervening drainage area would be accounted for by using a station or stations from those listed in table 6 adjusted by proper drainage area ratios to account for the difference in size.

The routing parameters C_o and K_o were determined by using the techniques applied in the Eddington analysis and are summarized in table 7.

For the first routing trial, average values (see table 7) for the model parameters C_o=2.90 and K_o=1,580 were used. To simulate the intervening drainage, each of the stations listed in table 6 were used individually and

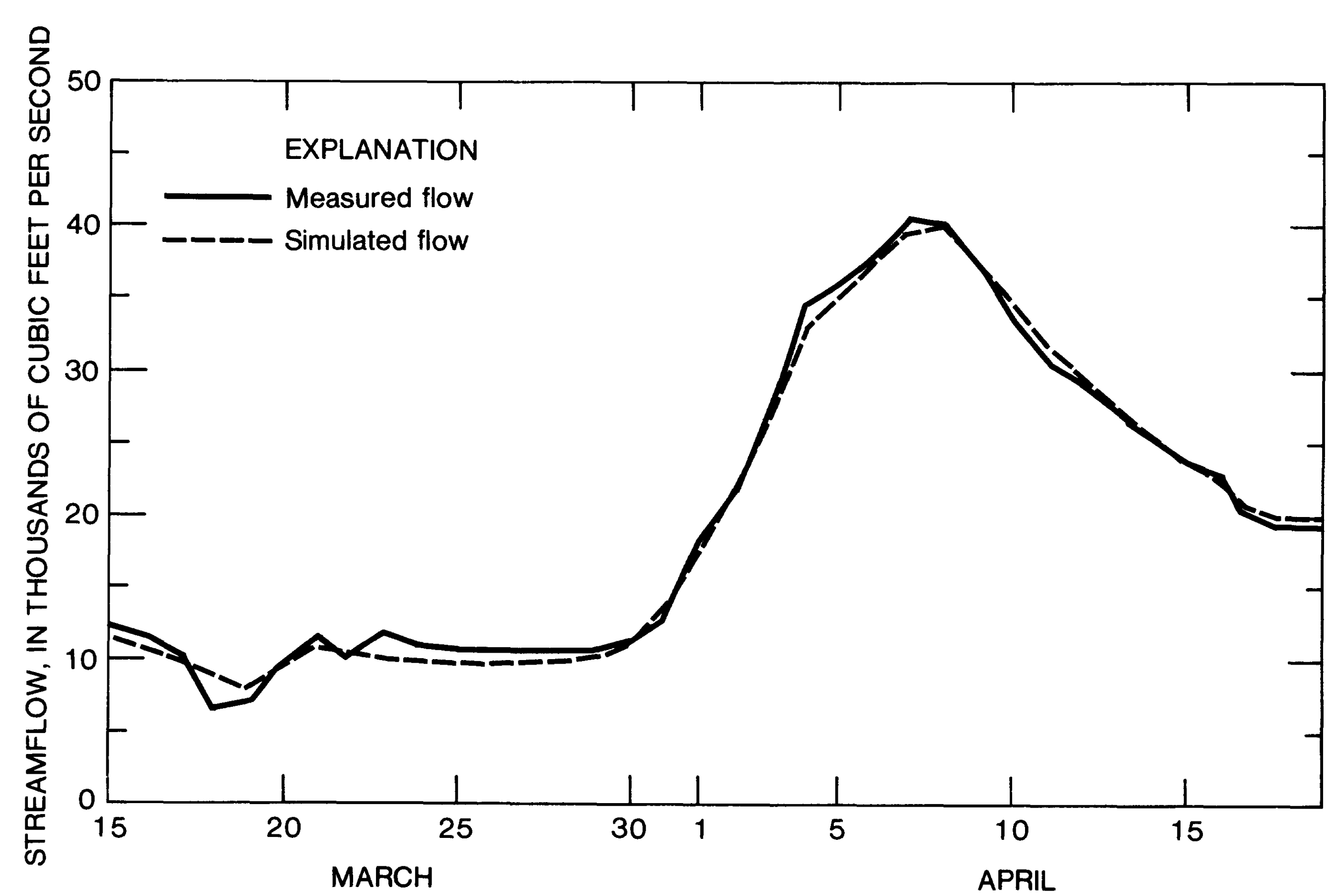

Figure 5. Daily hydrograph, Eddington, spring 1981.

Table 7. Selected reach characteristics used in the Ossipee flow-routing study

Site	Q_o 1/ (ft³/s)	W_o (ft)	S_o (ft/ft)	$\frac{dQ_o}{dY_o}$ (ft²/s)	C_o (ft/s)	K_o (ft²/s)
Effingham Falls	687	200	1.233×10^{-3}	460	2.30	1,393
Ossipee	874			706	3.53	1,772

1/ Mean discharge calculated over the period of record.

adjusted by the ratio determined by dividing 122 mi² by the drainage area of the site being considered. Water years 1979 through 1981 were used as a calibration data set. The best-fit model from this analysis proved to be the one that used the Royal station adjusted by a ratio of 0.87 to simulate intervening drainage. Further refinement of this model found the best-fit values of C_o and K_o to be 1.50 and 1,580, respectively.

A summary of the simulation of mean daily discharge at Ossipee for the 3 water years 1979 through 1981 is given in table 8.

Table 8. Results of routing model for Ossipee

Mean absolute error for 1,096 days	=	11.47 percent
Mean negative error (783 days)	=	-11.16 percent
Mean positive error (313 days)	=	12.25 percent
Total volume error	=	-2.16 percent

36 percent of the total observations had errors ≤ 5 percent
60 percent of the total observations had errors ≤ 10 percent
73 percent of the total observations had errors ≤ 15 percent
83 percent of the total observations had errors ≤ 20 percent
89 percent of the total observations had errors ≤ 25 percent
11 percent of the total observations had errors > 25 percent

All attempts to refine this best-fit model failed to reduce the errors significantly. An analysis of the results indicated consistently large negative errors (simulated values too low) during the low-flow and baseflow periods of the simulation (fig. 8). Further study pointed out that

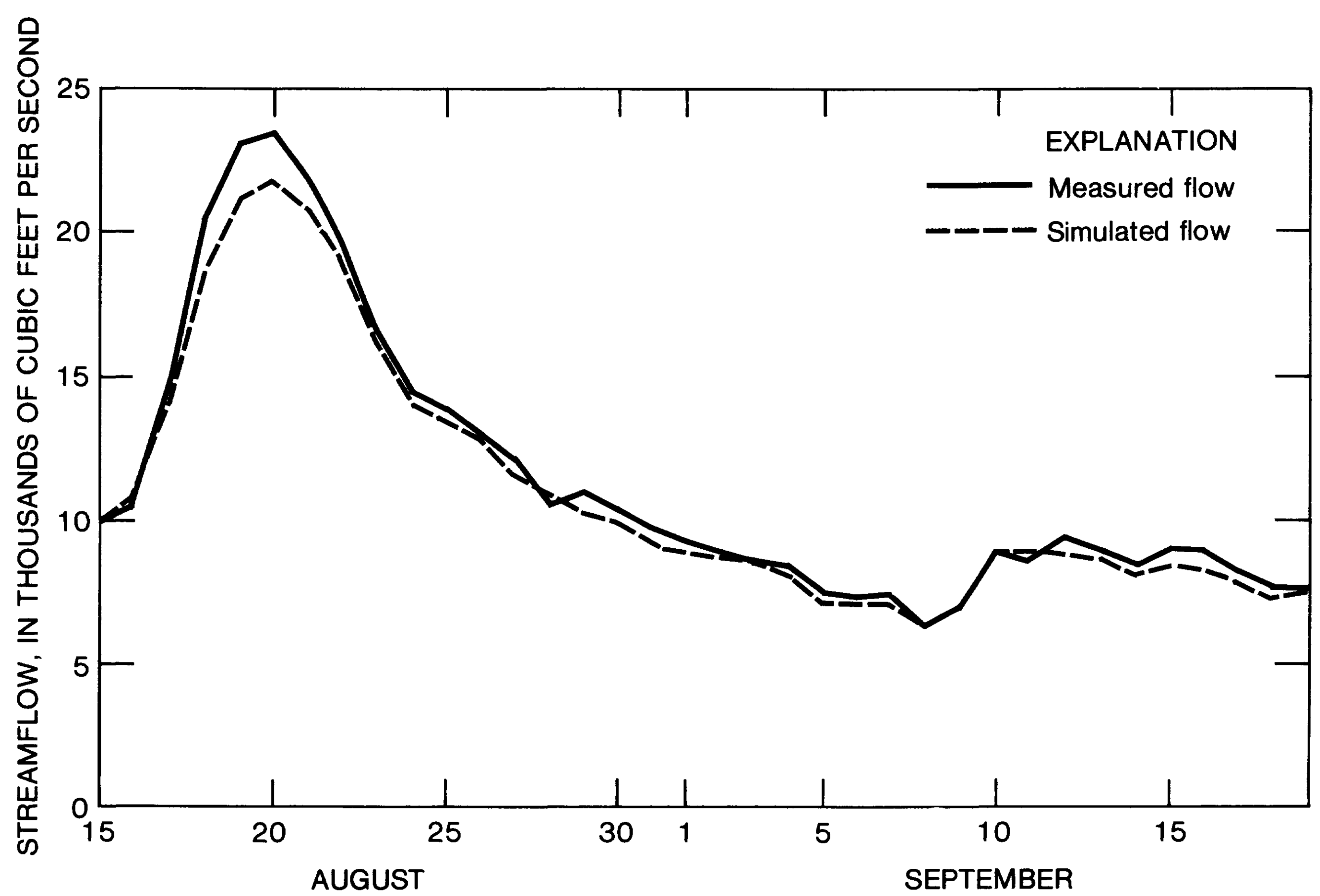

Figure 6. Daily hydrograph, Eddington, late summer 1981.

the intervening drainage area of 122 mi^2 between the Effingham Falls and Ossipee gages was shown by Prescott (1980) to be composed of significant deposits of sand and gravel and glacial till. These deposits would be a significant source of ground-water inflow to the Ossipee River during low-flow periods. None of the gages used in an attempt to simulate the intervening drainage had this characteristic and, as a result, the simulation has the significant errors noted.

Regression Analysis Results

Linear regression techniques were applied to all four of the selected sites. The streamflow record for each station considered for simulation (the dependent variable) was regressed against streamflow records at other stations (explanatory variables) during a given period of record (the calibration period). Best-fit linear regression models were developed and used to provide a daily streamflow record that was compared to the observed streamflow record. The percent difference between the simulated and actual record for each day was calculated. The results of the regression analysis for each site are summarized in table 9.

The streamflow record at Cherryfield (22500) was not reproduced with an acceptable degree of accuracy using regression techniques. The Cherryfield simulated data were within 10 percent of the actual record only 20 percent of the time during the calibration period. These results occurred when daily mean discharges at Dennysville (21200) were used as the explanatory variable.

The greatest hindrance to obtaining a satisfactory simulation in this case was that the station was regressed against stations in different drainage basins because no gage other than the Cherryfield station is currently operated in the Narraguagus River basin. Although the Cherryfield streamflow records were regressed against independent stations with similar basin and hydrographic characteristics, the differences in basin characteristics and, no doubt, precipitation patterns are sufficient to preclude adequate simulation.

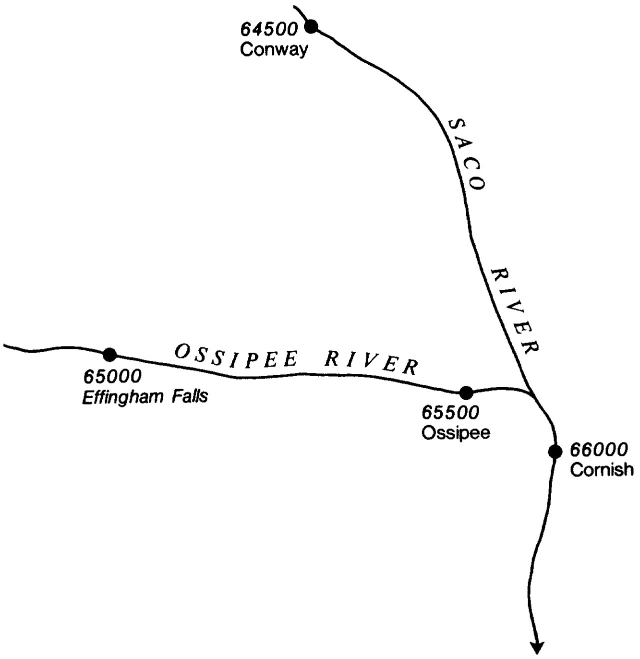

Figure 7. The Ossipee study area.

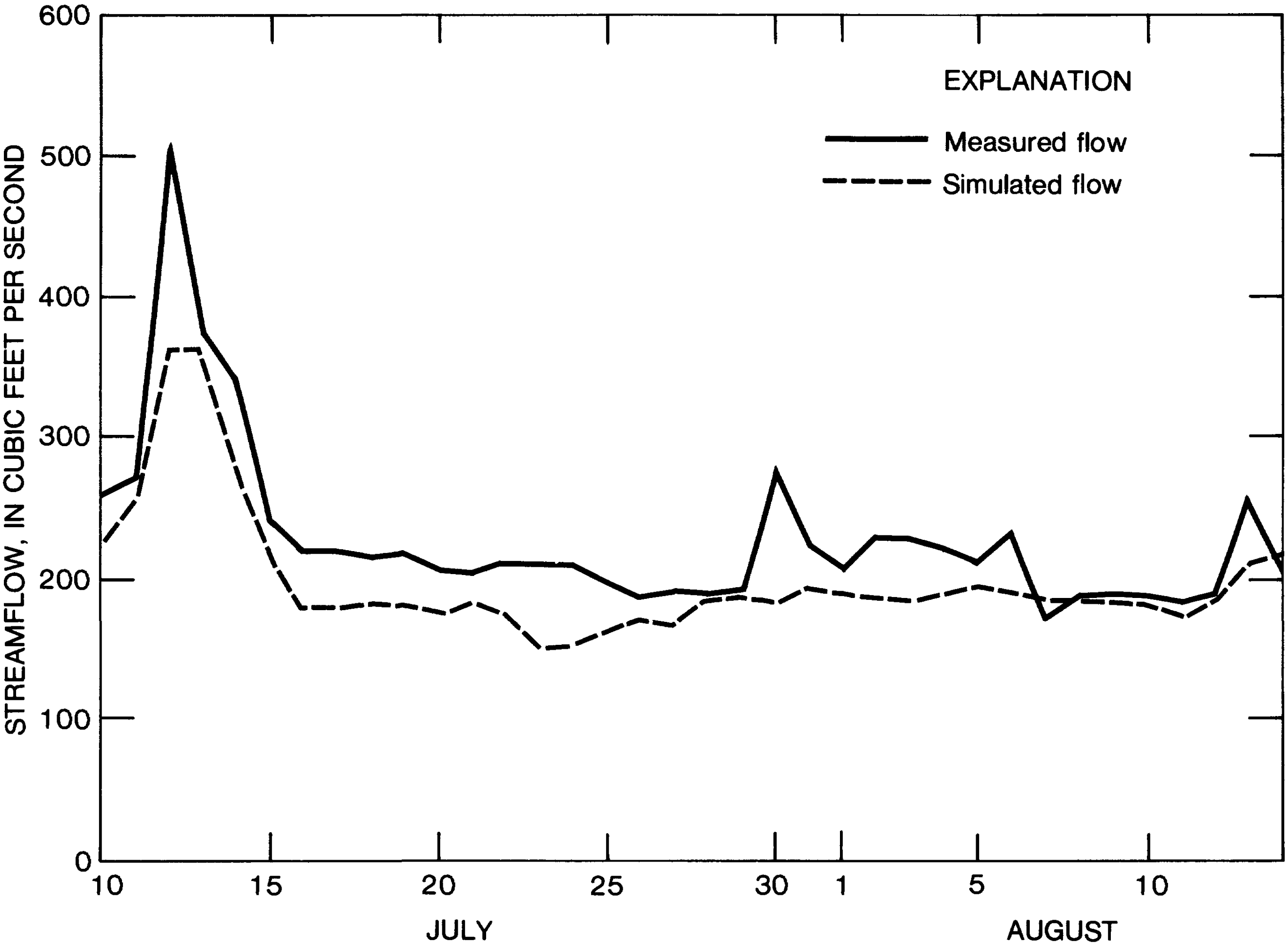

Figure 8. Daily hydrograph at Ossipee, summer 1980.

Table 9. Summary of calibration for regression modeling of mean daily streamflow at selected gage sites in Maine.

Station	Model	Percentage of simulated flow within 5% of actual	Percentage of simulated flow within 10% of actual	Calibration period (water years)
22500 Cherryfield	$Q_{22500} = 65.9 + 2.14\ (Q_{21200})$	10.3	19.6	1980-81
36390 Eddington	FOR: $Q_{34500} \leq 12000$ $Q_{36390} = 653 + 0.377\ (Q_{34500}) + 0.565\ (\text{LAG1 } Q_{34500}) + 0.570\ (Q_{30500})$ FOR: $Q_{34500} > 12000$ $Q_{36390} = 1018 + 0.554\ (Q_{34500}) + 0.530\ (\text{LAG1 } Q_{34500}) + 0.333\ (Q_{30500})$	61.2	86.2	1980-82
49265 Sidney	$Q_{49265} = 199 + 0.707\ (Q_{46500}) + 0.306\ (\text{LAG1 } Q_{46500}) + 1.06\ (\text{LAG1 } Q_{47000}) + 2.56\ (Q_{49000}) + 12.7\ (Q_{47730})$	29.3	60.7	1981-82
65500 Ossipee	$Q_{65500} = 27.4 + 0.782\ (Q_{65000}) + 0.376\ (\text{LAG1 } Q_{65000}) + 0.390\ (Q_{60000})$	33.2	57.0	1979-81

The more successful simulations of streamflow records at the Sidney, Ossipee, and Eddington stations were all produced from regressions with stations in the same basin. The streamflows at all of these stations experience varying degrees of regulation. The dependent streamflow records were regressed against upstream records on the main stem of the rivers as well as regulated and unregulated tributaries to the main stem. Special explanatory variables specified as LAG1 Q were created by lagging the discharges by 1 day. The interaction in a regression of the lagged and unlagged values for a given streamflow record acts to statistically route the flow from an upstream to a downstream site. The lagged discharge values account for the travel time between the two sites.

The regression model for Sidney (49265) includes five explanatory variables. The flow at Sidney was regressed against the lagged and unlaged flow at station 46500, the nearest upstream station on the main stem, and the flow at station 49000, a station on a major tributary. Both of these stations are below control structures that greatly regulate the flow. Two tributary sites, stations 47000 (lagged flow only) and 47730, served as indicators of unregulated inflow upstream from the Sidney station. Lagged values were used where appropriate.

The estimates from the regression model for Sidney simulated the actual record within 10 percent for 61 percent of the calibration period and within 5 percent for 29 percent of the period. The probable reason this simulation is not better is that there are numerous control structures on the main stem between station 46500 and Sidney. Any flow regulation by these structures cannot be accounted for with the regression model. This fact makes the Sidney station a poor candidate for statistical streamflow synthesis.

The streamflow record for the Ossipee station (65500) was simulated with a regression model that includes as explanatory variables the streamflow at station 65000, lagged streamflow at station 65000, and streamflow at station 60000. Station 65000 is located below a regulated site on the main stem, and there is very little regulation between it and the Ossipee site. Station 60000 is an out-of-basin unregulated site. It proved to be more significant to the model than an unregulated site in the Saco River basin, the major basin that includes the Ossipee drainage.

The simulated data for Ossipee were within 10 percent of the actual flows for 57 percent of the calibration period and within 5 percent for 33 percent of the period.

The most successful regression modeling for all four selected stations was that for the Eddington station. The model uses the lagged and unlagged streamflow record at station 34500 (West Enfield), and the record at station 30500, an unregulated tributary to the Penobscot. This model simulated the actual record within 10 percent for 83 percent and within 5 percent for 54 percent of the calibration period. The average error for the period is 5.7 percent, and the total flow volume error for the period is negligible.

Further improvement in the simulation was attempted by using two separate models, one for high flows ($Q>12,000$ ft³/s at West Enfield) and one for low flows ($Q\leq12,000$ ft³/s at West Enfield). At flows higher than approximately 12,000 ft³/s at West Enfield, the effects of regulation on the main stem streamflow are negligible. Using a high- and low-flow model, the models can accommodate change in travel time between West Enfield and Eddington at two different flow regimes.

The overall simulation for Eddington, using the two models, reproduced the actual Eddington record within 10 percent for 86 percent of the calibration period and within 5 percent for 61 percent of the period. The average error for the calibration period is 5.3 percent.

Conclusions Pertaining to Alternative Methods of Data Generation

The simulated data from both the flow-routing and regression methods for the Ossipee stream gage were not sufficiently accurate to substitute for the operation of a continuous stream gage. The same was true of regression results for Cherryfield and Sidney. Therefore all three stations should remain in operation as part of the Maine stream-gaging program. At the Eddington stream gage, both the flow-routing and the regression methods provided streamflow that may be accurate enough for the intended uses. The Eddington stream gage is operated primarily to supply discharges for water-quality data-collection activities. However, the modeling results are tentative because only 3 complete water years of data have been recorded at Eddington. These years were used for calibrating the models and no verification was attempted. Before the utility of the two models can be adequately assessed, they should be verified using a different data set. Operation of the stream gage should continue until sufficient data are available for verification.

In summary, all four stations considered in this section should remain in operation and will be included in the next step of this analysis.

COST-EFFECTIVE RESOURCE ALLOCATION

Introduction to Kalman-Filtering for Cost-Effective Resource Allocation (K-CERA)

In a study of the cost-effectiveness of a network of stream gages operated to determine water consumption in the Lower Colorado River Basin, a set of techniques called K-CERA were developed (Moss and Gilroy, 1980). Because that study concerned water balance, the network's effectiveness was measured in terms of the extent to which it minimized the sum of error variances in estimating annual mean discharges at each site in the network. This measure of effectiveness tends to concentrate stream-gaging resources on the larger, less stable streams where potential errors are greatest. While such a tendency is appropriate for a water-balance network, in the broader context of the multitude of uses of the streamflow data collected in USGS's Streamflow Information program, this tendency causes undue concentration on large streams. Therefore, the original version of K-CERA was extended to include, as optional measures of effectiveness, the sums of the variances of errors in estimating the following streamflow variables: annual mean discharge in cubic feet per second, annual mean discharge in percentage, average instantaneous discharge in cubic feet per second, and average instantaneous discharge in percentage. Using percentage errors does not unduly weight activities at large streams to the detriment of records on small streams. In addition, the instantaneous discharge is the basic variable from which all other streamflow data are derived. For these reasons, this study used the K-CERA techniques with the sums of the variances of the percentage errors of the instantaneous discharges at all continuously gaged sites as to measure the effectiveness of the data-collection activity.

The original version of K-CERA also failed to account for error contributed by missing stage or other correlative data that are used to compute streamflow data. The probabilities of missing correlative data increase as the period between service visits to a stream gage increases. A procedure for dealing with the missing record has been developed and was incorporated into this study.

Brief descriptions of the mathematical program used to optimize cost-effectiveness of collecting data and techniques of applying Kalman filtering (Gelb, 1974) to determine stream-gage record accuracy are presented below. For more detail on the theory or the applications of K-CERA, see Moss and Gilroy (1980) and Gilroy and Moss (1981).

Description of Mathematical Program

The program, called "The Traveling Hydrographer," attempts to allocate among stream gages a predefined budget for the collection of streamflow data in such a manner that the field operation is the most cost-effective possible. The measure of effectiveness is discussed above. The set of decisions available to the manager is the frequency of use (number of times per year) of each of a number of routes that may be used to service the stream gages and to make discharge measurements. The range of options within the program is from zero usage to daily usage for each route. A route is defined as a set of one or more stream gages and the least-cost travel that takes the hydrographer from his base of operations to each of the gages and back to base. A route will have associated with it an average cost of travel and average cost of servicing each stream gage visited along the way. The first step in this part of the analysis is to define the set of practical routes. This set of routes frequently will contain the path to an individual stream gage with that gage as the sole stop and return to the home base so that the individual needs of a stream gage can be considered in isolation from the other gages.

Another step in this part of the analysis is the determination of any special requirements for visits to each of the gages for such purposes as necessary periodic maintenance, rejuvenation of recording equipment, or required periodic sampling of water-quality data. Such special requirements are considered to be inviolable constraints in terms of the minimum number of visits to each gage.

The final step is to use all of the above to determine the number of times, N_i, that the i^{th} route for $i=1, 2, ..., NR$, where NR is the number of practical routes, is used during a year such that the budget for the network is not exceeded, the minimum number of visits to each station is made, and the total uncertainty in the network is minimized. Figure 9 represents this step in the form of a mathematical program. Figure 10 presents a tabular layout of the problem. Each of the NR routes is represented by a row of the table and each of the stations is represented by a column. The zero-one matrix, (ω_{ij}), defines the routes in terms of the stations that compose it. A value of one in row i and column j indicates that gaging station j will be visited on route i; a value of zero indicates that it will not. The unit-travel costs, β_i, are the per-trip costs of the hydrographer's travel time and any related per diem and operation, maintenance, and rental costs of vehicles. The sum of the products of β_i and N_i for $i=1, 2, ..., NR$ is the total travel cost associated with the set of decisions $N=(N_1, N_2, ..., N_{NR})$.

The unit-visit cost, α_j, is composed of the average service and maintenance costs incurred on a visit to the station plus the average cost of making a discharge measurement. The set of minimum visit constraints is denoted by the row $\lambda_j, j=1, 2, ..., MG$, where MG is the number of stream gages. The row of integers $M_j, j=1, 2, ... MG$ specifies the number of visits to each station. M_j is the sum of the products of ω_{ij} and N_i for all i and must equal or exceed λ_j for all j if N is to be a feasible solution to the problem.

$$\underset{\underline{N}}{\text{Minimize}}\ V = \sum_{j=1}^{MG} \phi_j\,(M_j)$$

$V \equiv$ total uncertainty in the network

$\underline{N} \equiv$ vector of annual number times each route was used

$MG \equiv$ number of gages in the network

$M_j \equiv$ annual number of visits to station j

$\phi_j \equiv$ function relating number of visits to uncertainty at station j

Such that

Budget $\geq T_c \equiv$ total cost of operating the network

$$T_c = F_c + \sum_{j=1}^{MG} \alpha_j M_j + \sum_{i=1}^{NR} \beta_i N_i$$

$F_c \equiv$ fixed cost

$\alpha_j \equiv$ unit cost of visit to station j

$NR \equiv$ number of practical routes chosen

$\beta_i \equiv$ travel cost for route i

$N_i \equiv$ annual number times route i is used (an element of $\underline{N}$)

and such that

$$M_j \geq \lambda_j$$

$\lambda_j \equiv$ minimum number of annual visits to station j

Figure 9. Mathematical-programming form of the optimization of the routing of hydrographers.

Route	Gage 1	2	3	4	.	j	.	MG	Unit Travel Cost	Uses
1	1	0	0	0	.	.	.	0	β_1	N_1
2	1	1	0	0	.	.	.	0	β_2	N_2
3	1	0	0	0	.	.	.	0	β_3	N_3
4	0	1	0	0	.	.	.	0	β_4	N_4
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
i	.	.	.	.	.	ω_{ij}	.	.	β_i	N_i
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
NR	0	0	0	0	.	.	.	1	β_{NR}	N_{NR}
Unit Visit Cost	α_1	α_2	α_3	α_4	.	α_j	.	α_{MG}		
Minimum Visits	λ_1	λ_2	λ_3	λ_4	.	λ_j	.	λ_{MG}		
Visits	M_1	M_2	M_3	M_4	.	M_j	.	M_{MG}		
Uncert. Function	ϕ_1	ϕ_2	ϕ_3	ϕ_4	.	ϕ_j	.	ϕ_{MG}		

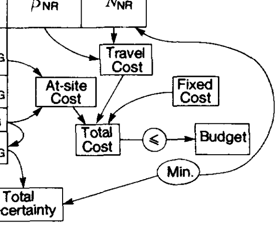

Figure 10. Tabular form of the optimization of the routing of hydrographers

The total cost expended at the stations is equal to the sum of the products of α_j and M_j for all j. The cost of record computation, documentation, and publication is assumed to be influenced negligibly by the number of visits to the station and is included along with overhead in the fixed cost of operating the network. The total cost of operating the network equals the sum of the travel costs, the at-site costs, and the fixed cost, and must be less than or equal to the available budget.

The total uncertainty in the estimates of discharges at the MG stations is determined by summing the uncertainty functions, ϕ_j, evaluated at the value of M_j from the row above it, for $j=1, 2, ..., MG$.

As pointed out in Moss and Gilroy (1980), the steepest descent search used to solve this mathematical program does not guarantee a true optimum solution. However, the locally optimum set of values for N obtained with this technique specify an efficient strategy for operating the network, which may be the true optimum strategy. The true optimum cannot be guaranteed without testing all undominated, feasible strategies.

Description of Uncertainty Functions

As noted earlier, uncertainty in streamflow records is measured in this study as the variance of the percentage errors of estimation of instantaneous discharges. This uncertainty is derived from three sources: (1) an error derived from uncertainties in the stage-discharge relationship (rating curve) or other functions that relate discharge to primary correlative data collected at the stream gage, (2) an error derived from reconstruction of streamflow records when the primary correlative data are missing, and (3) an error derived during periods when secondary data are not available to reconstruct streamflow records. The variances of the errors from these sources are weighted by the fractions of time during which each can be expected to occur and combined to estimate the expected error variance, which is the dependent variable of an uncertainty function. This relation can be expressed:

$$V_T=\epsilon_f V_f+\epsilon_r V_r+\epsilon_e V_e \qquad (3)$$

where

V_T is the expected total error variance,
ϵ_f is the fraction of time when the primary recorders are functioning,
V_f is the variance of the first error source described above,
ϵ_r is the fraction of time during which secondary data are available to reconstruct streamflow records given that the primary data are missing,
V_r is the variance of the second error source,
ϵ_e is the fraction of time during which no data are available to compute streamflow records,
V_e is the variance of the third error source.

The fractions of time for which each source of error is relevant are functions of the frequencies at which the recording equipment is serviced. It is assumed that the primary and secondary sites are serviced at the same frequency and at about the same times.

The time, τ, from the last service visit until failure of the recorder or recorders at the primary site is assumed to have a probability distribution defined by the truncated negative exponential family, that is

$$f_T=ke^{-kT}/(1-e^{-ks}) \qquad (4)$$

where

f_T is the probability density of failure times,
k is a coefficient,

and

e is the base of natural logarithms.

It is assumed that if a recorder fails it continues to malfunction until the next service visit. Thus,

$$\epsilon_f=1-E[d]/s \qquad (5)$$

where

d is downtime of the primary recorders,
$E[\cdot]$ is the expected value of the random variable contained within the brackets,

and

s is the interval between visits to the site.

$E[d]$ is derivable from equation 4, as is shown in the Appendix.

The fraction of time, ϵ_e, for which no records exist at either the primary or the secondary site also can be derived from a bivariate application of equation 4. (See appendix.) It is assumed that the times to failure at the primary and secondary sites are independent of each other and that they have identical probability density functions for failure times.

The fraction of time, ϵ_r, for which records are reconstructed based on data from a secondary site is determined by the equation

$$\epsilon_r=1-\epsilon_f-\epsilon_e. \qquad (6)$$

The variance, V_f, of the error derived from primary record computation is determined by analyzing a time series of residuals that are the differences between the measured discharge and the rating curve discharge. The rating curve discharge is determined from a relationship between discharge and some correlative data such as water-surface elevation for the gaging station. The measured discharge is the discharge determined by field observations of depths, widths, and velocities. The following variables are defined

$$x_2(t)=ln\Big(q_T(t)\Big)-ln\Big(q_R(t)\Big), \tag{7}$$

where $x_2(t)$ is the instantaneous difference between the logarithms of the true discharge, $q_T(t)$, and the rating-curve discharge $q_R(t)$. The variable $x_2(t)$ represents the true variability about the rating curve, but $x_2(t)$ is an unobservable random variable because $q_T(t)$ is unobservable. The residuals available to the analyst include measurement errors but also contain information about the structure of $x_2(t)$. These residuals, $z(t)$, are defined as

$$z(t)=x_2(t)+v(t)=ln\Big(q_m(t)\Big)-ln\Big(q_R(t)\Big), \tag{8}$$

where

$v(t)$ is the measurement error,

and

$q_m(t)$ is the measured discharge.

In the Kalman-filter analysis, the time series of $z(t)$ is analyzed to determine three site-specific parameters for each uncertainty function. The Kalman filter used in this study assumes that the difference $x_2(t)$ is a continuous first-order Markovian process that has an underlying Gaussian (normal) probability distribution with a zero mean and a variance (subsequently referred to as process variance) equal to $q/2\beta$. The variable q is the spectral density of the white noise that drives the Markovian process, and β is the reciprocal of the correlation time of the Markovian structure of $x_2(t)$. The 1-day autocorrelation coefficient, ρ, of $x_2(t)$ is a function of β. The variance of $z(t)$, α_z^2, is therefore defined as

$$\alpha_z^2=q/2\beta+r, \tag{9}$$

where r is the variance of the measurement error $v(t)$. The three parameters, q, β, and r, are computed by analyzing the statistical properties of the time series of residuals $z(t)$. These three site-specific parameters are needed to define this component of the uncertainty relationship. The Kalman filter utilizes these three parameters to determine the variance of the errors of estimation of discharge as a function of the number of discharge measurements per year (Moss and Gilroy, 1980).

If the recorder at the primary site fails and no concurrent data are available at other sites to reconstruct the missing record at the primary site, there would be at least two ways of estimating discharges at the primary site. A recession curve could be applied from the time of recorder stoppage until the gage was once again functioning, or the expected value of discharge for the period of missing data could be used as an estimate. The expected-value approach is used in this study to estimate V_e, the error variance during periods when concurrent data are unavailable at nearby sites. If the expected value is used to estimate discharge, the value used should be the expected value of discharge at the time of year for which the record is missing because of the seasonality of the streamflow processes. The variance of streamflow, which also is a seasonally varying parameter, is an estimate of the error variance that results from using the expected value as an estimate. Thus, the coefficient of variation, C_v, squared is an estimate of the required error variance V_e. Because C_v varies seasonally and the times of failures cannot be anticipated, a seasonally averaged C_v is used:

$$C_v=100\left(\frac{1}{365}\sum_{i=1}^{365}\left(\frac{\sigma_i}{\mu_i}\right)^2\right)^{1/2} \quad , \tag{10}$$

where

σ_i is the square root of the variance of daily discharges for the i^{th} day of the year,

and

μ_i is the expected value of discharge on the i^{th} day of the year.

The variance, V_r, of the error during periods of reconstructed streamflow records is estimated on the basis of correlation between records at the primary site and records from other gaged sites. The correlation coefficient, ρ_c, between the streamflows with seasonal trends removed (detrended) at the site of interest and detrended streamflows at the other sites is a measure of the soundness of their linear relationship. The fraction of the variance of streamflow at the primary site that is explained by data from the other sites is equal to ρ_c^2. Thus, the fraction of unexplained variance, that is, the error in reconstructed records at the primary site, is $(1-\rho_c^2)$. If the error variance is expressed in units of percentage squared, as is the case in this study, an estimate of the potential variance of streamflow for any day of the year is C_v^2 as defined in the paragraph above. Thus, V_r can be estimated as $(1-\rho_c^2)C_v^2$.

It is assumed in this study that the differences between the logarithms of the computed discharges and the true discharges at each instance are normally (Gaussian) distributed with a mean of zero and a variance of either V_f, V_r, or V_e depending on whether the at-site streamflow recorder was functioning (f), whether the record was reconstructed (r) from another primary source of data, or whether the record was estimated (*e*) without the aid of other concurrent data. Therefore, the resulting *a priori* distribution of errors is not normally distributed in terms of the logarithms of discharge data. This lack of normality causes difficulty in interpretation of the resulting errors of estimation, that is, the square root of the uncertainty contained in the streamflow record. If the logarithmic errors were normally distributed, approximately two-thirds of the time the true logarithmic error would be within the range defined by plus and minus one standard error from the mean. The lack of normality caused by the multiple sources of error increases the percentage of errors contained within this range above that of a Gaussian probability distribution of logarithmic errors with the same standard deviation.

To assist in interpreting the results of the analyses, a new parameter, equivalent Gaussian spread (EGS), is introduced. The parameter EGS specifies the range in terms of equal positive and negative logarithmic units from the mean that would encompass errors with the same *a priori* probability as would a Gaussian distribution with a standard deviation equal to EGS; in other words, the range from -1 EGS to +1 EGS contains about two-thirds of the errors. For Gaussian distributions of logarithmic errors, EGS and standard error are equivalent. EGS is reported herein in units of percentage and an approximate interpretation of EGS is "two-thirds of the errors in instantaneous streamflow data will be within plus or minus EGS percent of the reported value."

The Application of K-CERA in Maine

As a result of the first two parts of this analysis, it has been recommended that 45 of the currently existing stream gages in the State of Maine be continued in operation. These 45 stream gages were subjected to the K-CERA analysis with results that are described below.

Definition of Missing Record Probabilities

As described earlier, the statistical characteristics of missing stage or other correlative data for computation of streamflow records can be defined by a single parameter, the value of k in the truncated negative exponential probability distribution of times to failure of the equipment. In the representation of f_T as given in equation 4, the average time to failure is $1/k$. The value of $1/k$ will vary from site to site depending upon the type of equipment at the site and upon its exposure to natural elements and vandalism. The value of $1/k$ can be changed by advances in the technology of data collection and recording. To estimate $1/k$ in Maine, a period of actual data collection of 10 years duration in which little change in technology occurred and in which stream gages were visited on a consistent pattern of monthly frequency was used. During this 10-year period a gage could be expected to be malfunctioning an average of 5.6 percent of the time (G.R. Keezar, oral commun., 1983). There was no reason to distinguish between gages on the basis of their exposure or equipment, so the 5.6 percent lost record and a monthly visit frequency were used to determine a value for $1/k$ of 261 days, which was used to determine ϵ_f, ϵ_e, and ϵ_r for each of the 45 stream gages as a function of the individual frequencies of visit.

Definition of Cross-Correlation Coefficient and Coefficient of Variation

To compute the values of V_e and V_r of the needed uncertainty functions, daily streamflow records for each of the 45 stations for the last 30 years or the part of the last 30 years for which daily streamflow values are stored in WATSTORE (Hutchinson, 1975) were retrieved. For each of the stream gages that had 3 or more complete water years of data, the value of C_v was computed and various options, based on combinations of other stream gages, were explored to determine the maximum ρ_c. For the four stations that had less than 3 water years of data, values of C_v and ρ_c were estimated subjectively. In addition to other nearby stream gages, some of the stations had other means by which streamflow data could be reconstructed when the primary recorder was malfunctioning. Some stations are equipped with telemetry systems that operate independently from the primary recorder and are routinely queried either once or twice per day. At other locations, a local resident is hired to read and record stage at a station once or twice daily. At several sites nearby, hydropower plants have rated their turbines to determine the discharge that passes through them and keep flow records that can be used for streamflow reconstruction. At one site, an auxiliary recorder is operated at the station to provide backup stage record.

Analyses were performed to determine cross correlations, ρ_c, between daily discharges at sites with one or another of these types of auxiliary records. For the case of daily or twice-daily readings of stage (observer or telemetry), station 55000 (Swift), which had the highest observed value of C_v (142) yielded a ρ_c of 0.96 for daily readings and 0.99 for twice-daily readings. Because the high C_v indicates a relatively flashy stream, these values of ρ_c were assumed to be worst cases and were used for all other stations that were read either once or twice daily.

A worst-case situation, station 59000 (Auburn), for those stations with nearby hydropower records was analyzed. This site had the largest intervening flow between the gage and the power plant of all stations in this category. The ρ_c developed between the Auburn stream gage and the Gulf Island Power Plant was 0.99. This value was used for all other stations with nearby power records.

In the case of the auxiliary recorder at the gaging station (53500, Errol), an 11-year history of operation in this manner from 1970 through 1980 was inspected. During this period, one recorder or the other always produced valid stage record. However, there is no reason to believe that this will always be the case, so a ρ_c of 0.99 was assumed between the primary and auxiliary records at this site.

The set of parameters for each station and the auxiliary records that gave the highest cross correlation coefficient are listed in table 10.

Table 10. Statistics of record reconstruction

Station no.	C_v	ρ_c	Source of reconstructed records
10000	107	0.914	11500 14000
10500	103	.948	14000 11500
11000	89.2	.934	11500 14000
12515*	86	.70	
13500	82.6	.96	Telemetry; read daily.
14000	86.7	.946	10500 10000
15800	101	.99	Telemetry; read twice daily.
17000	106	.99	Telemetry; read twice daily.
18500	72.5	.96	Observer; read daily.
19000	68.7	.96	Observer; read daily.
20000	56.0	.99	Upstream hydropower plant.
21000	57.2	.977	20000
21200	83.4	.808	22500
22260*	91	.82	
22500	99.3	.836	21200 31500
30000	49.3	.99	Upstream hydropower plant.
30500	104	.901	34500
31500	132	.779	57000
34500	64.6	.99	Observer; read twice daily.
36390	51.3	.980	31500 30500 34500
42500	65.7	.99	Upstream hydropower plant.
46500	55.2	.99	Upstream hydropower plant.
47000	135	.914	57000 31500
47730	101	.702	31500 57000
49000	112	.702	57000 31500
49130*	100	.61	
49265	53.7	.669	49000 46500
49373	76.5	.608	49396
49396	104	.608	49373
49500	79.4	.505	55500
49550*	78	.61	
52500	114	.96	Telemetry; read daily.
53500	42.8	.99	Supplemental recorder at site.
54000	42.0	.99	Upstream hydropower plant.
54200	136	.836	57000 55000
54500	63.2	.99	Upstream hydropower plant.
55000	142	.96	Telemetry; read daily.
55500	123	.920	57000
57000	132	.920	55500
59000	73.0	.99	Upstream hydropower plant.
60000	125	.802	55500
64140	48.7	.573	60000
65500	76.4	.972	65000
66000	77.8	.912	64500 65000
69500	88.4	.659	60000

*Less than 3 water years of data are available. Estimates of C_v and ρ_c are subjective.

Kalman-Filter Definition of Variance

The determination of the variance V_f for each of the 45 stream gages required the execution of three distinct steps: (1) long-term rating analysis and computation of residuals of measured discharges from the long-term rating, (2) time-series analysis of the residuals to determine the input parameters of the Kalman-filter streamflow records, and (3) computation of the error variance, V_f, as a function of the time-series parameters, the discharge-measurement-error variance, and the frequency of discharge measurement.

In the Maine program analysis, definition of long-term rating functions was complicated by the fact that several stream gages in Maine have the dual seasonal characteristic of a summer or open-water period and a winter or backwater period. As a result of this characteristic, a single rating function to define the entire year is not feasible. Of the 45 stations included for analysis in this portion of the report, 22 have both a winter and summer period and required two rating curves to define discharge throughout the year. Fontaine (1982) has previously documented the fact that, for the open-water periods, existing rating curves, in most cases, defined the long-term rating function required in the analysis. In a majority of the cases where this is not true, the shifts in the curves have been extensions at the high end of the curves or slight adjustments in the extreme low ends of the curves. In these cases a mean curve was determined graphically. For Maine, the rating function for Jock (49396) was the only one that required development of a new rating. The rating function determined for Jock was of the form

$$LQM = B1 + B3 * ln(GHT - B2), \tag{11}$$

in which

- LQM is the logarithmic (base e) value of the measured discharge,
- GHT is the recorded gage height corresponding to the measured discharge,
- $B1$ is the logarithm of discharge for a flow depth of 1 foot,
- $B2$ is the gage height of zero flow,

and

- $B3$ is the slope of the rating curve.

The values of $B1$, $B2$, and $B3$ for this station were determined to be 3.06, 1.42, and 1.96, respectively.

Rating curves for the winter portions of the year have previously been determined and documented by Fontaine (1983). In summary, the methods utilized involved application of general linear models to solve for the dependent variable, measured discharge, as a function of groupings of independent variables. The independent variables included in the analysis for each winter discharge could be classified into three categories; data from the site for which a rating was desired, climatological data, and data from other stream gages. Data from the site in question include measured stage and the discharge corresponding to the measured stage determined from open-water rating. Climatological data taken from the National Weather Service sites closest to the stream gages in question include the maximum, minimum, and mean temperatures for the given day in question, the total precipitation that occurred as rain for the day and previous day in question, and finally the monthly mean maximum and minimum temperatures for the month being considered and the heating degree day units up to that time in the winter season. Data from other stream gages include the indicated mean daily discharge, based on the open-water rating curve, for sites that are both proximate and (or) physiographically similar to the site being considered.

Results of the winter rating analyses often yield ratings about which there was a large amount of variance, but some of the ratings had very tight fits about the available discharge measurements. Examples of both types of ratings are given below for the typical winter backwater periods in Maine. The rating curves were developed using discharge in cubic feet per second and the residuals were converted to logarithmic units (base e) before the auto covariance analysis.

The general linear rating function at West Enfield (34500) is given by the formula

$$Q=-1243+0.142(INDQ)+54.84(MonthMax) +0.621(SUNKHAZE), \tag{12}$$

where

- Q is the discharge at West Enfield in cubic feet per second,
- $INDQ$ is the indicated discharge at West Enfield in cubic feet per second,
- $MonthMax$ is the average of daily maximum temperatures for the month in °C,

and

- $SUNKHAZE$ is the furnished flow data from Sunkhaze power station in cubic feet per second.

The coefficient of determination (R^2) for this model is 0.970.

A tabular presentation of the residuals of the measured discharges about the ice period rating curve (measured discharge minus rated discharge) for West Enfleld is given in table 11.

Table 11. Residual data for West Enfield

Observation no.	Measurement no.	Date	Measured discharge (ft^3/s)	Residual (ft^3/s)	Percent error
1	483	Jan. 12, 1971	6360	-591.5	-9.300
2	484	Feb. 10, 1971	8130	615.6	7.572
3	485	Mar. 16, 1971	10400	421.1	4.049
4	492	Jan. 7, 1972	5090	319.4	6.274
5	493	Feb. 11, 1972	4800	868.7	18.098
6	494	Mar. 9, 1972	5090	-1109.2	-21.791
7	501	Jan. 12, 1973	8790	-626.8	-7.130
8	502	Feb. 16, 1973	14400	315.3	2.190
9	503	Mar. 22, 1973	29600	-124.8	-0.422
10	513	Jan. 29, 1974	12400	2130.7	17.183
11	514	Mar. 14, 1974	8240	-850.5	-10.322
12	520	Feb. 19, 1975	6460	-265.7	-4.113
13	527	Jan. 13, 1976	6560	-696.2	-10.613
14	528	Mar. 10, 1976	11700	932.4	7.970
15	535	Dec. 15, 1976	8920	670.7	7.519
16	536	Jan. 25, 1977	8270	202.0	2.443
17	537	Feb. 23, 1977	6480	-1539.4	-23.756
18	544	Dec. 15, 1977	9560	446.4	4.669
19	546	Feb. 14, 1978	10800	910.1	8.427
20	550	Jan. 16, 1979	7460	499.5	6.696
21	551	Feb. 9, 1979	11100	-1704.8	-15.359
22	557	Mar. 4, 1980	4660	-756.7	-16.238
23	562	Jan. 14, 1981	5860	-66.4	-1.132

The general linear rating function at Roxbury (55000) is given by the formula

$$Q=-7.64+25.80(INDSTAGE)+0.29(SPARIS) +0.03(DIAMOND), \tag{13}$$

where

- Q is the discharge at Roxbury in cubic feet per second,
- $INDSTAGE$ is the stage at Roxbury in feet,
- $SPARIS$ is the indicated discharge at South Paris gaging station in cubic feet per second,

and

- $DIAMOND$ is the indicated discharge at Diamond gaging station in cubic feet per second.

The coefficient of determination (R^2) for the Roxbury model is 0.897. A tabular presentation of the residuals of the measured discharges about the ice period rating curve for Roxbury is given in table 12.

The general linear rating function for Ossipee (65500) is given by the formula

$$Q=79.61+1.12(EOSSIPEE)+0.06(ROYAL), \tag{14}$$

where

- Q is the discharge at Ossipee River in cubic feet per second,
- $EOSSIPEE$ is the indicated discharge at Effingham Falls gaging station in cubic feet per second,

Table 12. Residual data for Roxbury

Observation no.	Measurement no.	Date	Measured discharge (ft^3/s)	Residual ft^3/s	Percent error
1	368	Dec. 16, 1970	75.6	3.649	4.827
2	370	Feb. 15, 1971	166.0	-9.434	-5.683
3	371	Mar. 17, 1971	158.0	15.902	10.065
4	385	Dec. 15, 1971	58.6	-0.433	-0.739
5	387	Feb. 16, 1972	75.3	-8.836	-11.734
6	388	Mar. 21, 1972	147.0	-8.675	-5.901
7	394	Dec. 12, 1972	99.1	-8.075	-8.148
8	396	Feb. 16, 1973	146.0	32.466	22.237
9	406	Jan. 30, 1974	137.0	-6.223	-4.543
10	411	Dec. 26, 1974	66.6	-6.066	-9.108
11	419	Mar. 10, 1976	119.0	25.351	21.303
12	425	Dec. 15, 1976	93.0	10.057	10.814
13	427	Mar. 10, 1977	58.9	0.389	0.660
14	433	Dec. 13, 1977	113.0	19.109	16.911
15	434	Feb. 6, 1978	161.0	-14.618	-9.079
16	443	Nov. 30, 1978	23.3	-11.057	-47.455
17	444	Jan. 16, 1979	62.9	-16.214	-25.777
18	451	Jan. 4, 1980	66.7	-3.463	-5.191
19	453	Mar. 18, 1980	43.2	-13.831	-32.017

ROYAL is the indicated discharge at Royal gaging station in cubic feet per second.

The coefficient of determination (R^2) for the Ossipee model is 0.959. A tabular presentation of the residuals of the measured discharges about the ice period rating curve for Ossipee is given in table 13.

The general linear rating function for Diamond (52500) is given by the formula

$$Q=-1570+0.22(INDQ)+26.23(MonthMax) +0.63(DEGDAY), \tag{15}$$

where

Q is the discharge at Diamond River in cubic feet per second,

$INDQ$ is the indicated discharge at Diamond River in cubic feet per second,

$MonthMax$ is the average of the daily maximum temperature for the month in °C,

and

$DEGDAY$ is the total heating degree days in °C from the beginning of the winter season to the date of interest.

The coefficient of determination (R^2) for the Diamond model is 0.765. A tabular presentation of the residuals of the measured discharges about the ice period rating curve for Diamond is given in table 14.

The time series of residuals (in logarithmic units) is used to compute sample estimates of q and β, two of the three parameters required to compute V_f, by determining a best fit autocovariance function to the time series of residuals. Measurement variance, the third parameter, is determined from an assumed constant percentage standard error. For the Maine program, all open-water measurements were assumed to have a measurement error of 2 percent and all ice measurements an assumed measurement error of 10 percent.

Table 13. Residual data for Ossipee

Observation no.	Measurement no.	Date	Measured discharge (ft^3/s)	Residual (ft^3/s)	Percent error
1	408	Mar. 10, 1971	406	-15.18	-2.057
2	409	Dec. 9, 1971	444	-44.84	-21.769
3	410	Jan. 18, 1972	738	-29.82	-6.598
4	415	Feb. 15, 1972	206	-21.11	-3.825
5	416	Mar. 14, 1972	452	-28.15	-6.134
6	417	Jan. 15, 1973	552	80.47	-9.422
7	418	Jan. 15, 1974	459	184.59	18.571
8	424	Feb. 14, 1975	854	28.49	5.458
9	434	Feb. 9, 1976	994	-150.91	-12.472
10	441	Dec. 21, 1976	522	-17.20	-2.073
11	447	Mar. 19, 1977	210	63.03	10.593
12	456	Dec. 10, 1977	830	-48.95	-5.021
13	457	Feb. 1, 1978	322	89.63	4.871
14	458	Jan. 16, 1979	595	-96.05	-17.182
15	462	Feb. 14, 1979	975	-99.32	-21.876
16	463	Feb. 11, 1980	840	-6.51	-2.310
17	468	Mar. 12, 1980	559	48.32	-13.650
18	469	Dec. 31, 1980	454	87.59	20.370
19	473	Jan. 27, 1981	282	-24.06	-7.591

Table 14. Residual data for Diamond

Observation no.	Measurement no.	Date	Measured discharge (ft^3/s)	Residual (ft^3/s)	Percent error
1	246	Dec. 17, 1970	149.0	21.68	14.55
2	247	Jan. 14, 1971	83.6	4.89	5.84
3	248	Feb. 17, 1971	155.0	2.47	1.59
4	249	Mar. 18, 1971	330.0	-0.41	-0.12
5	256	Nov. 10, 1971	96.7	-35.25	-36.45
6	257	Dec. 16, 1971	159.0	-24.28	-15.27
7	258	Jan. 20, 1972	109.0	-57.87	-53.09
8	259	Feb. 16, 1972	76.5	22.47	29.37
9	266	Jan. 17, 1973	135.0	-12.06	-8.93
10	267	Feb. 12, 1973	224.0	18.74	8.37
11	274	Jan. 31, 1974	242.0	88.13	36.42
12	275	Mar. 30, 1974	371.0	-13.66	-3.68
13	281	Feb. 13, 1975	71.6	18.73	26.16
14	287	Mar. 24, 1976	684.0	159.88	23.37
15	291	Dec. 16, 1976	196.0	102.53	52.31
16	297	Dec. 16, 1977	173.0	7.57	4.38
17	298	Feb. 2, 1978	275.0	-79.54	-28.92
18	304	Dec. 11, 1978	73.5	-26.44	-35.98
19	305	Jan. 17, 1979	113.0	-168.90	-149.47
20	311	Jan. 3, 1980	142.0	53.25	37.50
21	312	Feb. 12, 1980	44.4	37.75	85.02
22	313	Mar. 19, 1980	57.1	-103.93	-182.02
23	319	Jan. 8, 1981	87.9	-15.75	-17.92

As discussed earlier, q and β can be expressed as the process variance of the shifts from the rating curve and the 1-day autocorrelation coefficient of these shifts. Table 15 presents a summary of the autocovariance analysis expressed in terms of process variance and 1-day autocorrelation. In table 15, a 1 was added to the last digit of the station number to denote the winter portion of the year at the respective site. Typical fits of the covariance functions for selected stations in Maine are given in figures 11–13.

The autocovariance parameters, summarized in table 15, and data from the definition of missing record probabilities, summarized in table 10, are used jointly to define uncertainty functions for each gaging station. The uncertainty functions give the relationship of total error variance to the number of visits and discharge measurements. The stations for which graphic fits of the autocovariance functions were previously given present typical examples of uncertainty functions and are given in figure 14. These functions are based on the assumption that a measurement was made during each visit to the station.

In Maine, feasible routes to service the 45 stream gages were determined after consultation with personnel in the Hydrologic Data Section of the Maine office and after review of the uncertainty functions. In summary, 89 routes were selected to service all the stream gages in Maine. These routes included all possible combinations that describe the current operating practice, alternatives that were under consideration as future possibilities, routes that visited certain key individual stations, and combinations that grouped proximate gages where the levels of uncertainty indicated more frequent visits might be useful. These routes and the stations visited on each are summarized in table 16.

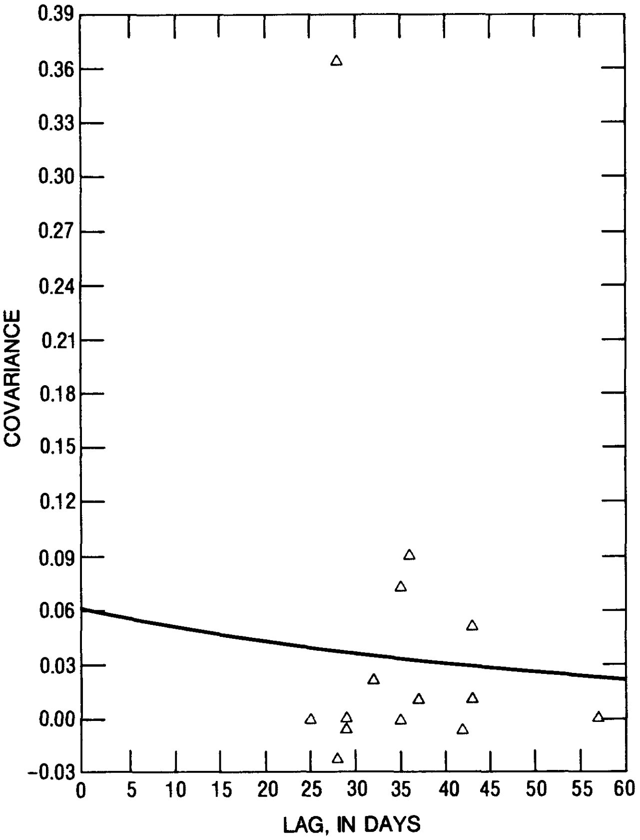

Figure 11. Autocovariance function for the winter period at Fort Kent.

The costs associated with the practical routes must be determined. Fixed costs to operate a gage typically include equipment rental, batteries, electricity, data processing and storage, computer charges, maintenance, miscellaneous supplies, and analysis and supervisory charges. For Maine, average values were applied to each station in the program for all the above categories except analysis and supervisory costs. Costs of analysis and supervision form a large percentage of the cost at each gaging station and can vary widely. These costs were determined on a station-by-station basis from past experience. Because 22 of the 45 stations have been split into two seasons, the associated costs for these sites had to be subdivided as well. The total fixed costs at each gage, not including analysis and supervisory expenditures, were prorated based on the lengths of the seasons into which the year had been divided. For example, at Ninemile (10000) the average length of the winter backwater period is 121 days and the average length of the summer open-water period is 244 days. Costs at this site would be allocated as follows:

$$\text{Winter fixed costs} = \text{total fixed cost} \times \frac{121}{365}$$

$$\text{Summer fixed costs} = \text{total fixed cost} \times \frac{244}{365}.$$

The analysis and supervisory costs were allocated in a different method. These costs are not merely a function of the length of a season but also of the difficulty associated with data interpretation during the period. Work on winter records requires a significantly larger portion of the funds than a simple pro rating based on time would indicate. These charges were allocated as follows. If the winter period was longer than 3 months, 60 percent of analysis and supervisory costs were charged to the winter station and 40 percent to the summer station. If the winter period was shorter than 3 months, 50 percent was charged to each of the seasons. These divisions were based on past experience.

Visit costs are those associated with paying the hydrographer for the time actually spent at a station servicing the equipment and making discharge measurements. These costs vary from station to station as a function of the difficulty and time required to make the discharge measurement. Average visit times were calculated for each station based on an analysis of discharge measurement data available. This time was then multi-

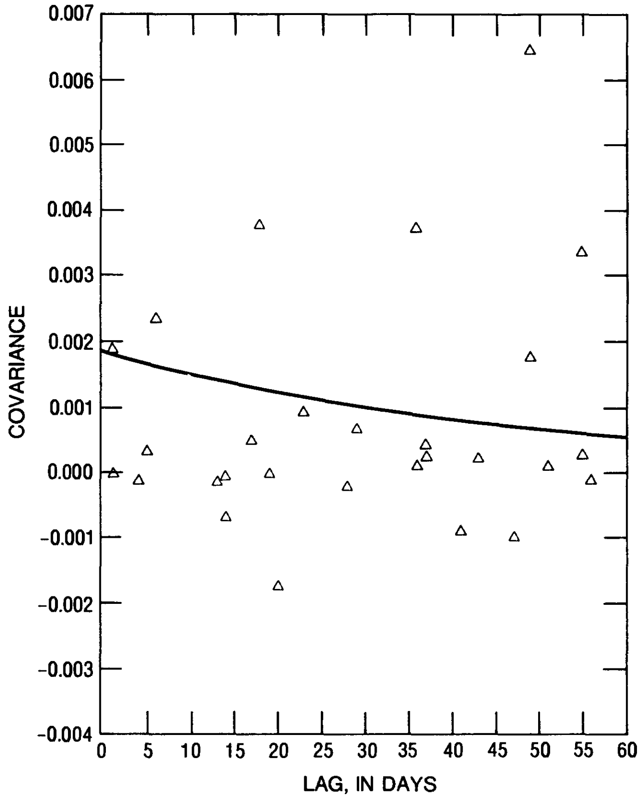

Figure 12. Autocovariance function for the summer period at Gilead.

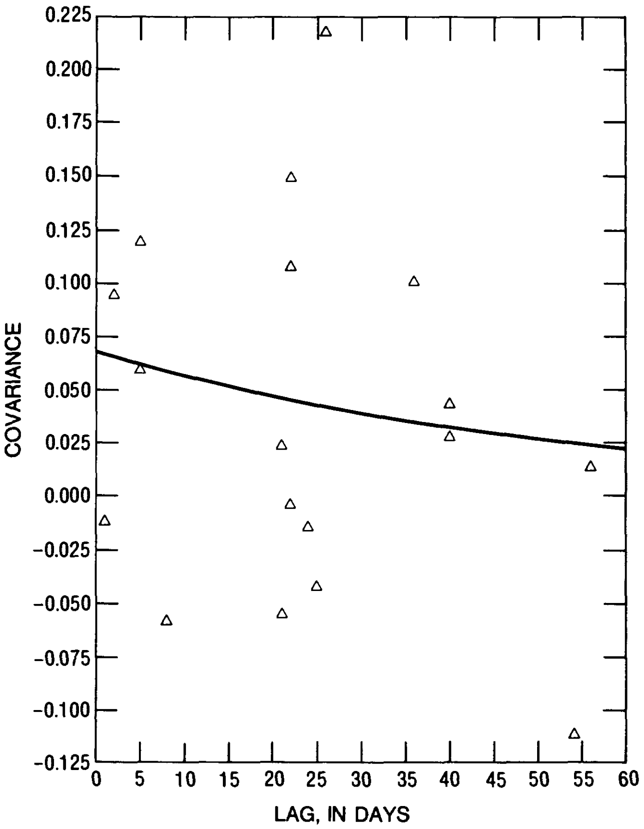

Figure 13. Autocovariance function for complete year at West Falmouth.

Table 15. Summary of the autocovariance analysis

Station no.	Station name	RHO*	Measurement variance (log base e)^2	Process variance (log base e)^2	Length of period (days)
10000	Ninemile summer	0.408	.0004	.0001	244
10001	Ninemile winter	0.452	.0100	.0790	121
10500	Dickey summer	0.959	.0004	.0029	244
10501	Dickey winter	0.927	.0100	.0833	121
11000	Allagash summer	0.992	.0004	.0004	244
11001	Allagash winter	0.917	.0100	.1091	121
12515	Clayton Stream all year	0.001	.0004	.0001	365
13500	Fish summer	0.999	.0004	.0023	244
13501	Fish winter	0.898	.0100	.0218	121
14000	Ft. Kent summer	0.850	.0004	.0009	244
14001	Ft. Kent winter	0.983	.0100	.0583	121
15800	Masardis summer	0.952	.0004	.0006	244
15801	Masardis winter	0.938	.0100	.1034	121
17000	Washburn summer	0.972	.0004	.0015	244
17001	Washburn winter	0.963	.0100	.0502	121
18500	Vanceboro all year	0.680	.0004	.0022	365
19000	Grand Lake Stream all year	0.971	.0004	.0007	365
20000	Baileyville all year	0.889	.0004	.0006	365
21000	Baring all year	0.889	.0004	.0004	365
21200	Dennysville all year	0.988	.0004	.0009	365
22260	Epping all year	0.973	.0004	.0004	365
22500	Cherryfield summer	0.779	.0004	.0380	275
22501	Cherryfield winter	0.908	.0100	.0031	90
30000	Mattaseunk summer	0.832	.0004	.0056	213
30001	Mattaseunk winter	0.993	.0100	.0082	152
30500	Mattawamkeag summer	0.989	.0004	.0013	261
30501	Mattawamkeag winter	0.827	.0100	.0966	104
31500	Dover-Foxcroft summer	0.809	.0004	.0017	261
31501	Dover-Foxcroft winter	0.943	.0100	.0890	104
34500	West Enfield summer	0.507	.0004	.0016	261
34501	West Enfield winter	0.626	.0100	.0021	104
36390	Eddington all year	0.833	.0004	.0002	365
42500	The Forks all year	0.964	.0004	.0006	365

See footnote at end of table.

Table 15. Summary of the autocovariance analysis—Continued

Station no.	Station name	RHO*	Measurement variance (log base e)^2	Process variance (log base e)^2	Length of period (days)
46500	Bingham all year	0.996	.0004	.0007	365
47000	North Anson summer	0.932	.0004	.0010	261
47001	North Anson winter	0.947	.0100	.1000	104
47730	Wilson Stream summer	0.977	.0004	.0022	275
47731	Wilson Stream winter	0.952	.0100	.2934	90
49000	Pittsfield summer	0.982	.0004	.0015	306
49001	Pittsfield winter	0.985	.0100	.0664	59
49130	Johnson Brook all year	0.847	.0004	.0083	365
49265	Sidney all year	0.986	.0004	.0029	365
49373	Mill all year	0.973	.0004	.0062	365
49396	Jock all year	0.981	.0004	.0677	365
49500	Cobbossee all year	0.982	.0004	.0219	365
49550	Togus Stream all year	0.263	.0004	.0005	365
52500	Diamond summer	0.964	.0004	.0006	261
52501	Diamond winter	0.556	.0100	.1571	104
53500	Errol all year	0.957	.0004	.0001	365
54000	Gorham all year	0.712	.0004	.0002	365
54200	Gilead summer	0.979	.0004	.0018	261
54201	Gilead winter	0.886	.0100	.0467	104
54500	Rumford all year	0.686	.0004	.0001	365
55000	Roxbury summer	0.999	.0004	.0095	261
55001	Roxbury winter	0.977	.0100	.0164	104
55500	Turner summer	0.980	.0004	.0016	275
55501	Turner winter	0.963	.0100	.1100	90
57000	South Paris all year	0.445	.0004	.0012	365
59000	Auburn all year	0.781	.0004	.0015	365
60000	Royal summer	0.913	.0004	.0019	320
60001	Royal winter	0.526	.0100	.0320	45
64140	West Falmouth all year	0.709	.0004	.0030	365
65500	Ossipee summer	0.952	.0004	.0010	275
65501	Ossipee winter	0.979	.0100	.0052	90
66000	Cornish summer	0.921	.0004	.0017	275
66001	Cornish winter	0.911	.0100	.0154	90
69500	Mousam all year	0.972	.0004	.0146	365

*One-day autocorrelation coefficient.

Table 16. Summary of the routes that may be used to visit stations in Maine

Route number	Stations serviced on the route								
1	10001	10501	11001	12515	13501	14001	15801	17001	
2	10001	10501	11001	13501	14001	15801	17001		
3	10501	11001	13501	14001					
4	10501	11001	14001						
5	10001	12515	15801	17001					
6	11001	15801							
7	10501	11001	15801						
8	10501	11001	15801	17001					
9	22501	22260	17001	15801	10501	11001	13501	14001	
10	10000	10500	11000	12515	13500	14000	15800	17000	
11	10000	10500	11000	13500	14000	15800	17000		
12	17000	13500							
13	17000	13500	14000						
14	12515	10000	15800	17000					
15	10500	11000	13500	14000					
16	22500	22260	21200	21000	20000	19000	18500	17000	15800
	10500	11000	12515	13500	14000				
17	22501	21000	36390						
18	22501	21000	36390	30501					
19	30000	30500	18500	19000	20000	21000	21200	22260	22500
20	30001	30501	18500	19000	20000	21000	21200	22260	22500
21	20000	21000	21200	22500					
22	18500	20000	21000	21200	22500				
23	18500	20000	21000						
24	19000	21200	22500						
25	22500	21000	36390						
26	30001	30500							
27	30000	30500							
28	31501	47001	34501	49001					
29	31501	34501	47001						
30	31501	34501							
31	31501	47001							
32	31501	47001	49001						
33	47001	49001							
34	36390	42500	46500	47000	31500	34500	49000		
35	42500	46500	47000	31500					
36	42500	46500	47000						
37	31500	34500							
38	36390	49000							
39	47731	55001	54201	52501					

Table 16. Summary of the routes that may be used to visit stations in Maine—*Continued*

Route number	Stations serviced on the route						
40	47731	52501					
41	47731	52501	54201				
42	52501	54201					
43	47730	54500	54200	55000	52500	53500	54000
44	47730	54200	55000				
45	47730	55000					
46	52500	53500	54000				
47	47730	55000	54500				
48	54200	64140	66000				
49	54200	66000					
50	60001	65501	66001				
51	65501	66001					
52	65500	66000	69500	60000	64140		
53	66000	69500	60000	64140			
54	66000	64140					
55	69500	65500	66000				
56	69500	65500					
57	57000	59000					
58	55500	49373	49396				
59	49373	49396					
60	49500	49550					
61	30000	30501	34501	31501	10001		
62	30001	30500	34500	31500	10000	42500	46500
63	10000						
64	22260						
65	22501						
66	30001						
67	30500						
68	31501						
69	47001						
70	47730						
71	47731						
72	49000						
73	49001						
74	49265						
75	49396						
76	49500						
77	52501						
78	54200						

Table 16. Summary of the routes that may be used to visit stations in Maine—*Continued*

Route number	Stations serviced on the route	
79	54201	
80	55000	
81	55001	
82	55500	
83	55501	
84	60001	
85	69500	
86	54500	55000
87	60000	64140
88	49130	
89	36390	

plied by the average hourly salary of hydrographers in the Maine office to determine total visit costs.

Route costs include the vehicle cost associated with driving the number of miles it takes to cover the route, the cost of the hydrographer's time while in transit, and any per diem associated with the time it takes to complete the trip.

K-CERA Results

The Traveling Hydrographer Program utilizes the uncertainty functions along with the appropriate cost

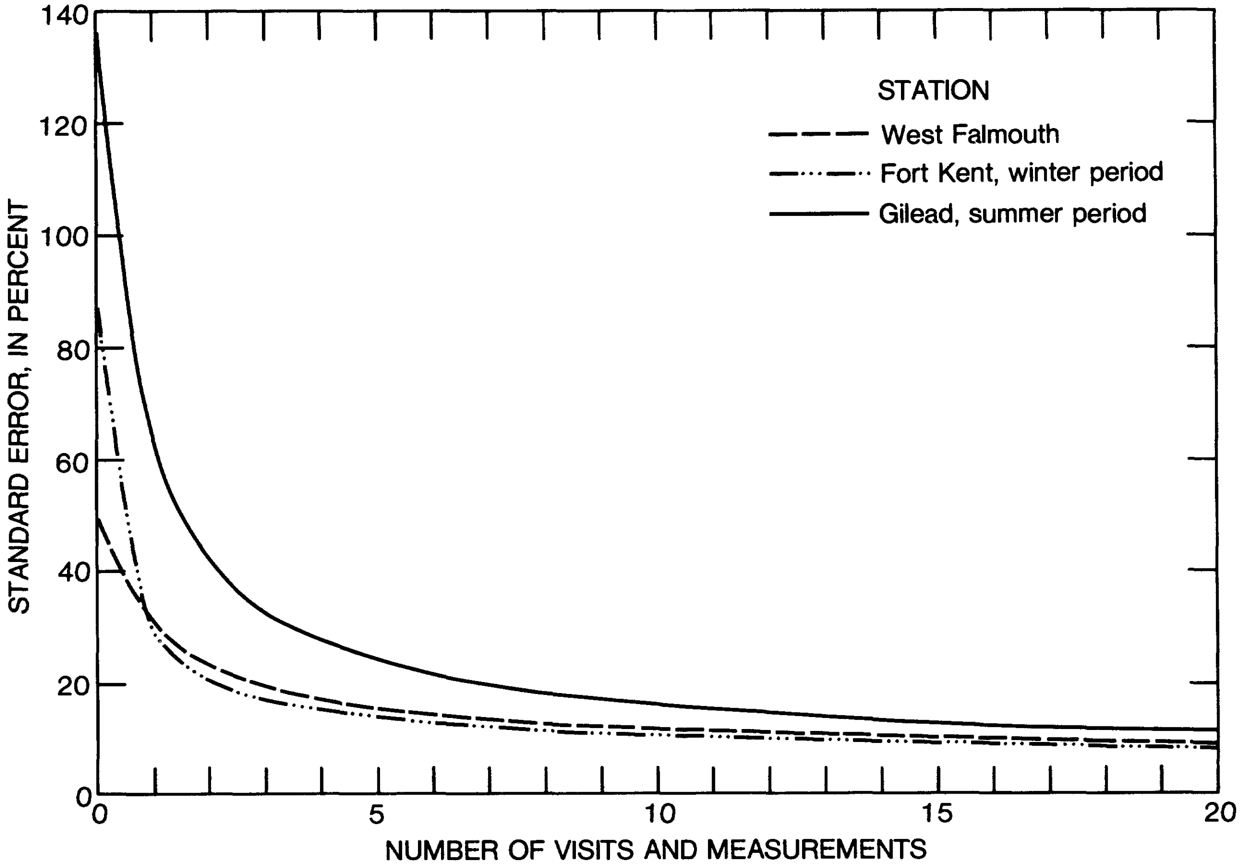

Figure 14. Typical uncertainty function for instantaneous discharge.

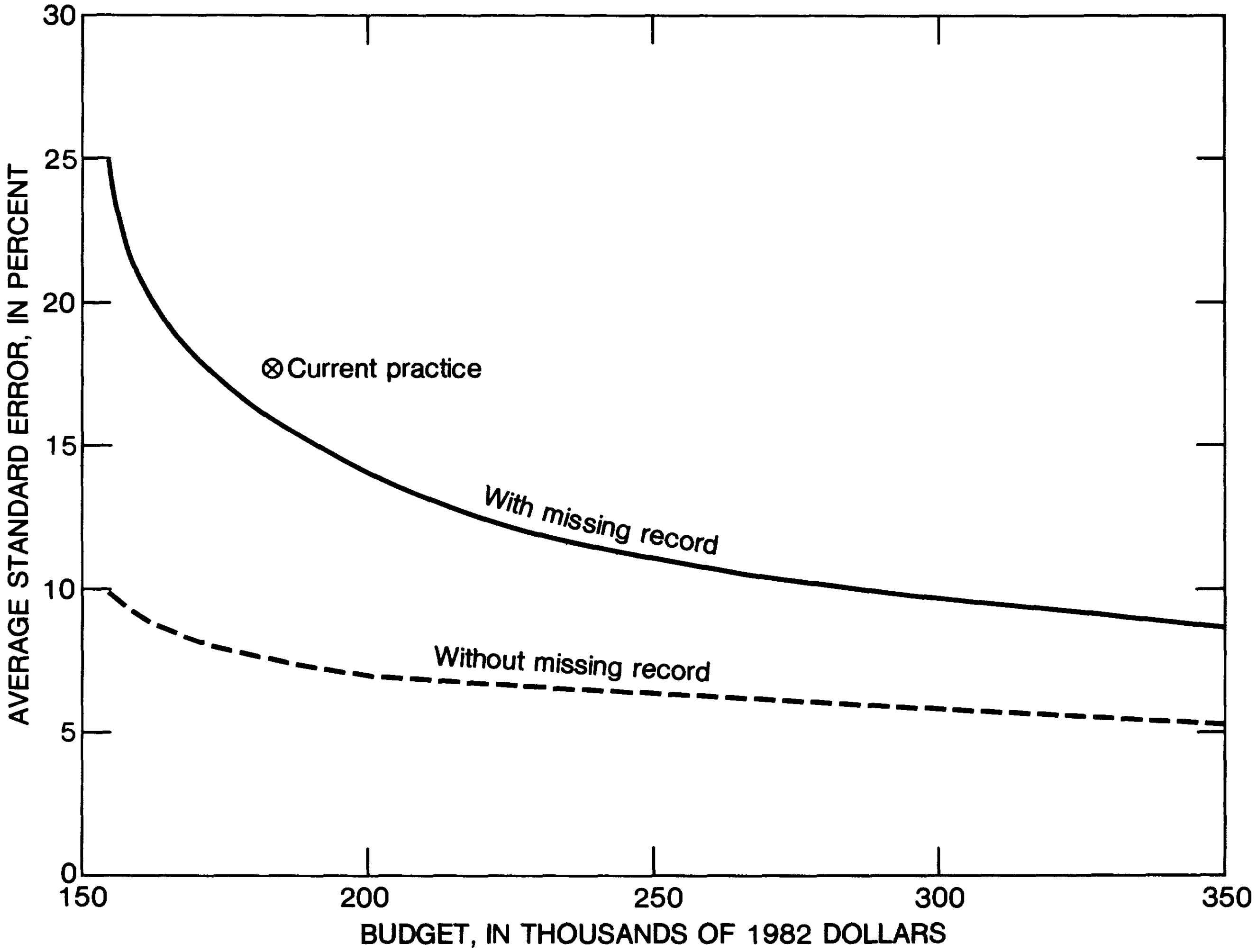

Figure 15. Temporal average standard error per stream gage.

data and route definitions to compute the most cost-effective way of operating the stream-gaging program. In this application, the first step was to simulate the current practice and determine the total uncertainty associated with it. To accomplish this, the number of visits made to each stream gage and the specific routes used to make these visits were fixed. In Maine, current practice indicates that discharge measurements are made 75 percent of the time that a station is visited. This value was determined as an average and applied to both the winter and summer periods. For gaging stations with seasonal ratings, the seasonal uncertainties must be weighted by the percentage of time that each applies to obtain a weighted average for the station's uncertainty function. The resulting average error of estimation for the current practice in Maine is plotted as a point in figure 15 and is 17.7 percent.

The solid line on figure 15 represents the minimum average standard error that can be obtained for a given budget with the existing instrumentation and technology. The line was defined by several runs of the Traveling Hydrographer Program with different budgets. Constraints on the operations other than budget were defined as described below.

To determine the minimum number of times each station must be visited, consideration was given only to the physical limitations of the method used to record data. The effect of visitation frequency on the accuracy of the data and amount of lost record is taken into account in the uncertainty analysis. In Maine, a minimum requirement of four visits per year was calculated and applied to all stations. At stations where the year was split into winter and summer seasons, the minimum was two visits for each season. This value was based on limitations of the batteries used to drive recording equipment, capacities of the uptake spools on the digital recorders, and the need to protect gages from freezing winter conditions in Maine (W.B. Higgins, written commun., 1983).

Minimum visit requirements should also reflect the need to visit stations for special reasons such as water-quality sampling. In Maine, all water-quality work is done on separate trips not integrated with the surface-water fieldwork and, therefore, did not influence minimum visit requirements.

The results in figure 15 and table 17 summarize the K-CERA analysis and are predicated on a discharge measurement made each time a station is visited. This is a change from the current policy, under which about three measurements are made for each four visits. It was felt that the new policy would improve the cost-effectiveness of the operation. Ideally, the ratio of measurements to visits would be optimized for each site individually. This step will be accomplished in a future evaluation of the Maine program.

Table 17. Selected results of K-CERA analysis

Identification	Standard error of instantaneous discharge (SE), in percent [Equivalent Gaussian spread (EGS)] (Number of visits per year to site)					
	Current operation	Budget, in thousands of 1982 dollars: 155	170	180.3	240	350
Average SE per station 1/	17.7	25.1	17.9	16.1	11.5	8.7
EGS for the program	4.2	5.4	4.1	3.8	2.9	2.3
10000 Ninemile-s 2/	16.2 [1.1] (5)	29.5 [1.5] (2)	22.6 [1.3] (3)	18.7 [1.2] (4)	13.1 [1.1] (7)	9.5 [1.0] (12)
10001 Ninemile-w 3/	30.8 [29.2] (3)	32.6 [29.9] (2)	32.6 [29.9] (2)	32.6 [29.9] (2)	29.8 [28.8] (4)	27.5 [27.2] (11)
10500 Dickey-s	14.4 [4.5] (5)	26.9 [6.5] (2)	16.6 [4.9] (4)	14.3 [4.5] (5)	9.8 [3.6] (9)	7.2 [2.8] (15)
10501 Dickey-w	27.5 [25.0] (3)	29.4 [27.0] (2)	29.4 [27.0] (2)	29.4 [27.0] (2)	22.6 [21.8] (5)	15.5 [15.0] (13)
11000 Allagash-s	12.6 [1.0] (5)	23.8 [1.7] (2)	14.7 [1.1] (4)	12.6 [1.0] (5)	8.6 [0.7] (9)	6.3 [0.5] (15)
11001 Allagash-w	31.0 [29.1] (3)	32.3 [31.0] (2)	32.3 [31.0] (2)	32.3 [31.0] (2)	26.2 [25.8] (5)	18.4 [18.2] (13)
12515 Clayton Stream	18.5 [0.4] (8)	26.5 [0.4] (4)	21.5 [0.4] (6)	19.8 [0.4] (7)	13.9 [0.3] (14)	9.9 [0.3] (27)
13500 Fish-s	10.6 [0.9] (5)	20.9 [1.7] (2)	12.4 [1.0] (4)	10.6 [0.9] (5)	6.9 [0.6] (9)	4.9 [0.5] (15)
13501 Fish-w	15.9 [13.7] (3)	17.8 [14.6] (2)	17.8 [14.6] (2)	17.8 [14.6] (2)	13.3 [12.3] (5)	9.8 [9.3] (12)
14000 Fort Kent-s	12.0 [3.1] (5)	22.7 [4.0] (2)	14.0 [3.3] (4)	12.0 [3.1] (5)	8.3 [2.8] (9)	6.2 [2.5] (15)
14001 Fort Kent-w	17.1 [13.3] (3)	19.5 [15.9] (2)	19.5 [15.9] (2)	19.5 [15.9] (2)	12.2 [10.4] (5)	7.8 [6.8] (12)
15800 Masardis-s	11.1 [2.1] (5)	23.7 [3.0] (2)	13.3 [2.3] (4)	11.0 [2.1] (5)	6.2 [1.7] (10)	4.3 [1.4] (16)
15801 Masardis-w	28.4 [25.5] (3)	29.9 [27.3] (2)	29.9 [27.3] (2)	29.9 [27.3] (2)	22.9 [22.2] (5)	15.4 [15.3] (13)
17000 Washburn-s	11.8 [2.9] (5)	25.0 [4.4] (2)	14.1 [3.2] (4)	11.7 [2.9] (5)	6.6 [2.1] (10)	4.6 [1.7] (16)
17001 Washburn-w	19.1 [15.7] (3)	21.6 [17.6] (2)	21.6 [17.6] (2)	21.6 [17.6] (2)	14.1 [13.1] (5)	9.3 [8.9] (12)
18500 Vanceboro	9.8 [5.0] (8)	15.4 [5.6] (4)	13.2 [5.4] (5)	11.7 [5.3] (6)	7.6 [4.8] (13)	6.1 [4.5] (22)
19000 Grand Lake Stream	8.5 [2.0] (8)	14.1 [2.7] (4)	12.0 [2.5] (5)	10.5 [2.3] (6)	6.1 [1.6] (13)	4.3 [1.3] (22)
20000 Baileyville	6.1 [2.4] (8)	10.6 [2.8] (4)	8.8 [2.7] (5)	7.6 [2.6] (6)	4.3 [2.2] (13)	3.0 [1.9] (22)
21000 Baring	4.9 [1.7] (12)	8.2 [2.0] (6)	7.2 [1.9] (7)	6.6 [1.9] (8)	4.0 [1.6] (16)	2.6 [1.3] (31)
21200 Dennysville	14.7 [1.6] (8)	21.9 [2.5] (4)	19.2 [2.1] (5)	17.3 [1.9] (6)	11.2 [1.3] (13)	8.4 [1.0] (22)
22260 Epping	17.3 [5.5] (8)	25.0 [6.6] (4)	22.2 [6.2] (5)	20.1 [5.9] (6)	13.0 [4.8] (14)	10.1 [4.0] (23)
22500 Cherryfield-s	15.0 [1.3] (8)	22.2 [1.8] (4)	19.5 [1.6] (5)	17.6 [1.5] (6)	11.5 [1.1] (13)	8.7 [0.8] (22)
22501 Cherryfield-w	21.5 [18.5] (4)	25.0 [20.2] (2)	25.0 [20.2] (2)	25.0 [20.2] (2)	21.1 [18.5] (4)	16.8 [15.5] (10)
30000 Mattaseunk-s	8.9 [7.2] (4)	12.3 [7.6] (2)	12.3 [7.6] (2)	10.0 [7.4] (3)	7.8 [7.0] (6)	6.9 [6.5] (10)
30001 Mattaseunk-w	5.5 [3.3] (4)	8.7 [4.5] (2)	8.7 [4.5] (2)	8.7 [4.5] (2)	5.2 [3.2] (4)	3.2 [2.3] (8)
30500 Mattawamkeag-s	19.8 [2.2] (4)	30.9 [3.5] (2)	23.8 [2.6] (3)	19.8 [2.2] (4)	12.9 [1.5] (8)	9.7 [1.2] (13)
30501 Mattawamkeag-w	30.6 [29.0] (4)	33.4 [31.5] (2)	33.4 [31.5] (2)	33.4 [31.5] (2)	28.7 [28.1] (5)	21.2 [20.8] (15)
31500 Dover-Foxcroft-s	28.0 [4.5] (5)	31.5 [4.7] (4)	25.4 [4.3] (6)	21.8 [4.1] (8)	15.9 [3.7] (15)	11.8 [3.2] (27)
31501 Dover-Foxcroft-w	32.6 [24.3] (3)	27.9 [21.9] (4)	27.9 [21.9] (4)	27.9 [21.9] (4)	18.5 [14.7] (10)	13.7 [10.9] (19)
34500 West Enfield-s	8.3 [4.3] (5)	16.4 [4.9] (2)	12.0 [4.6] (3)	9.7 [4.4] (4)	6.7 [4.2] (7)	5.4 [4.0] (11)
34501 West Enfield-w	6.8 [4.7] (3)	8.5 [4.8] (2)	8.5 [4.8] (2)	8.5 [4.8] (2)	8.5 [4.8] (2)	6.8 [4.7] (3)
36390 Eddington	4.3 [1.5] (12)	9.9 [1.9] (4)	9.9 [1.9] (4)	9.9 [1.9] (4)	6.4 [1.7] (7)	4.0 [1.5] (13)
42500 The Forks	10.2 [2.4] (5)	12.3 [2.6] (4)	10.2 [2.4] (5)	8.8 [2.2] (6)	5.0 [1.7] (12)	3.3 [1.4] (20)
46500 Bingham	8.5 [1.2] (5)	10.2 [1.3] (4)	8.5 [1.2] (5)	7.3 [1.0] (6)	4.0 [0.7] (12)	2.6 [0.5] (20)

See footnotes at end of table.

Table 17. Selected results of K-CERA analysis—Continued

Identification	Current operation	Budget, in thousands of 1982 dollars: 155	170	180.3	240	350
	Standard error of instantaneous discharge, in percent [Equivalent Gaussian spread] (Number of visits per year to site)					
47000 North Anson-s	21.4 [3.1] (5)	24.8 [3.4] (4)	21.4 [3.1] (5)	19.1 [2.9] (6)	12.6 [2.3] (12)	9.4 [1.9] (20)
47001 North Anson-w	29.8 [24.9] (3)	32.9 [28.2] (2)	32.9 [28.2] (2)	28.2 [24.9] (3)	19.8 [17.8] (7)	14.3 [12.9] (14)
47730 Wilson Stream-s	24.1 [3.5] (5)	38.5 [5.8] (2)	18.9 [2.8] (8)	16.9 [2.5] (10)	11.9 [1.8] (20)	8.8 [1.3] (36)
47731 Wilson Stream-w	44.8 [39.7] (3)	47.1 [45.5] (2)	36.7 [35.2] (4)	30.8 [29.3] (6)	20.5 [19.2] (14)	15.4 [14.4] (25)
49000 Pittsfield-s	28.0 [2.9] (5)	44.8 [5.1] (2)	20.7 [2.1] (9)	18.7 [1.9] (11)	12.8 [1.3] (23)	9.4 [1.0] (43)
49001 Pittsfield-w	18.6 [9.4] (3)	21.9 [11.6] (2)	21.9 [11.6] (2)	21.9 [11.6] (2)	13.9 [7.3] (5)	11.0 [5.8] (8)
49130 Johnson Brook	25.8 [9.5] (8)	35.4 [11.1] (4)	21.5 [8.8] (12)	18.9 [8.3] (16)	13.4 [6.8] (35)	9.8 [5.4] (68)
49265 Sidney	10.0 [2.4] (12)	17.3 [4.4] (4)	15.5 [3.8] (5)	14.1 [3.4] (6)	9.6 [2.3] (13)	7.0 [1.7] (24)
49373 Mill	18.8 [5.7] (8)	23.4 [7.1] (5)	13.7 [4.1] (15)	11.8 [3.6] (20)	8.2 [2.5] (42)	6.2 [1.9] (73)
49396 Jock	24.0 [12.9] (12)	32.2 [18.6] (6)	19.0 [10.4] (18)	16.4 [8.9] (24)	11.3 [6.0] (51)	8.5 [4.5] (91)
49500 Cobbossee	21.8 [8.9] (8)	29.6 [12.9] (4)	17.6 [7.1] (12)	14.9 [5.9] (17)	10.4 [4.0] (35)	7.6 [2.9] (66)
49550 Togus Stream	18.4 [2.6] (8)	25.9 [2.9] (4)	15.1 [2.5] (12)	12.7 [2.4] (17)	9.1 [2.3] (34)	7.0 [2.2] (61)
52500 Diamond-s	15.4 [2.2] (5)	18.1 [2.4] (4)	18.1 [2.4] (4)	15.4 [2.2] (5)	10.1 [1.7] (9)	7.2 [1.3] (15)
52501 Diamond-w	40.7 [39.8] (3)	41.5 [40.1] (2)	41.5 [40.1] (2)	41.5 [40.1] (2)	39.9 [39.5] (4)	32.1 [32.1] (25)
53500 Errol	6.5 [0.6] (5)	7.9 [0.6] (4)	7.9 [0.6] (4)	6.5 [0.6] (5)	4.0 [0.4] (9)	2.6 [0.4] (15)
54000 Gorham	6.6 [1.7] (5)	7.9 [1.8] (4)	7.9 [1.8] (4)	6.6 [1.7] (5)	4.1 [1.5] (9)	2.8 [1.4] (15)
54200 Gilead-s	18.8 [2.3] (9)	44.2 [5.1] (2)	21.6 [2.6] (7)	18.8 [2.3] (9)	13.3 [1.6] (17)	9.9 [1.2] (30)
54201 Gilead-w	28.1 [20.5] (3)	32.1 [22.1] (2)	27.7 [20.5] (3)	27.7 [20.5] (3)	18.4 [15.0] (9)	10.8 [9.0] (30)
54500 Rumford	9.7 [1.4] (5)	11.7 [1.5] (4)	11.7 [1.5] (4)	8.3 [1.4] (6)	5.4 [1.3] (10)	3.7 [1.2] (16)
55000 Roxbury-s	19.0 [1.8] (5)	22.4 [2.0] (4)	14.9 [1.5] (7)	11.6 [1.2] (10)	7.9 [0.9] (18)	5.6 [0.8] (31)
55001 Roxbury-w	16.1 [7.7] (3)	20.5 [9.2] (2)	20.5 [9.2] (2)	15.7 [7.7] (3)	10.2 [5.5] (6)	7.6 [4.3] (10)
55500 Turner-s	20.0 [2.9] (5)	36.6 [4.8] (2)	14.8 [2.2] (8)	12.2 [1.9] (11)	8.6 [1.4] (20)	6.3 [1.1] (35)
55501 Turner-w	26.3 [21.7] (3)	28.9 [25.4] (2)	24.3 [21.7] (3)	21.3 [19.2] (4)	14.5 [13.0] (9)	11.0 [9.8] (16)
57000 South Paris	19.1 [3.8] (8)	30.0 [4.3] (4)	19.1 [3.8] (8)	16.6 [3.7] (10)	11.9 [3.5] (18)	8.9 [3.3] (31)
59000 Auburn	8.2 [3.9] (8)	13.9 [4.4] (4)	8.2 [3.9] (8)	7.1 [3.8] (10)	5.0 [3.6] (18)	3.9 [3.2] (31)
60000 Royal-s	28.5 [4.7] (5)	47.2 [6.8] (2)	20.7 [4.0] (9)	17.8 [3.7] (12)	12.2 [2.9] (25)	8.9 [2.2] (47)
60001 Royal-w	21.4 [17.6] (3)	23.4 [18.2] (2)	23.4 [18.2] (2)	23.4 [18.2] (2)	20.0 [17.1] (4)	16.8 [15.3] (9)
64140 West Falmouth	12.2 [5.8] (9)	17.3 [6.7] (4)	11.3 [5.7] (11)	10.2 [5.5] (14)	8.0 [5.0] (27)	6.5 [4.5] (49)
65500 Ossipee-s	10.4 [2.9] (5)	20.7 [4.1] (2)	10.4 [2.9] (5)	8.1 [2.6] (7)	5.0 [2.0] (14)	3.7 [1.6] (23)
65501 Ossipee-w	7.5 [4.0] (3)	9.6 [4.7] (2)	9.6 [4.7] (2)	9.6 [4.7] (2)	7.3 [4.0] (3)	4.7 [2.9] (6)
66000 Cornish-s	9.5 [3.6] (9)	23.6 [5.6] (2)	15.2 [4.6] (4)	11.9 [4.0] (6)	8.0 [3.3] (12)	6.0 [2.6] (21)
66001 Cornish-w	13.7 [10.7] (3)	15.4 [11.9] (2)	15.4 [11.9] (2)	15.4 [11.9] (2)	13.2 [10.7] (3)	10.2 [8.7] (6)
69500 Mousam	21.6 [8.8] (8)	29.6 [12.0] (4)	22.7 [9.4] (7)	20.1 [8.3] (9)	13.9 [5.7] (19)	10.2 [4.2] (35)

See footnotes at end of table.

1/ Square root of seasonally averaged station variance.
2/ Summer season.
3/ Winter season.

It should be emphasized that figure 15 and table 17 are based on various assumptions (stated previously) concerning both the time series of shifts to the stage-discharge relationship and the methods of record reconstruction. Where a choice of assumptions was available, the assumption that would not underestimate the magnitude of the error variances was chosen.

It can be seen that the current policy results in an average standard error of estimate of streamflow of 17.7 percent. This policy requires a budget of $180,300 to operate the 45-station stream-gaging program. The range in standard errors is from a low of 4.3 percent for station 36390, at Eddington, to a high of 44.8 percent during ice-covered periods at station 47730, Wilson Stream (shown as station 47731 in table 17). It is possible to obtain this same average standard error with a reduced budget of about $170,000 with a change of policy in the field activities of the stream-gaging program. This policy and budget change would result in an increase in standard error from 4.3 to 9.9 percent at station 36390, while the standard error at station 47731 would decrease from 44.8 to 36.7 percent. However, these two stations would no longer have the lowest extremes of standard error. Station 21000, Baring, would have the standard error at 7.2 percent, while the ice-covered period at station 52500, Diamond (station 52501 in table 17), would have the highest at 41.5 percent.

It also would be possible to reduce the average standard error by a policy change while maintaining the $180,300 budget. In this case, the average standard error would decrease from 17.7 to 16.1 percent. Extremes of standard errors for individual sites would be 6.5 and 41.5 percent for stations 53500 and 52501, respectively.

A minimum budget of $155,000 is required to operate the 45-station program; a smaller budget would not permit proper service and maintenance of the gages and recorders. Stations would have to be eliminated from the program if the budget fell below this minimum. At the minimum budget, the average standard error is 25.1 percent. The minimum standard error of 7.9 percent would occur at two stations 53500 (Errol) and 54000 (Gorham), while the maximum of 47.1 percent would occur at 47731 (Wilson Stream-w).

The maximum budget analyzed was $350,000, which resulted in an average standard error of estimate of 8.7 percent. Thus, almost doubling the budget in conjunction with policy change would almost halve the average standard error that results from the current policy and current budget. For the $350,000 budget, the extremes of standard error are 2.6 for stations 21000 (Baring), 46500 (Bingham), 53500 (Errol), and 32.1 percent at station 52501. Thus, it is apparent that significant improvements in accuracy of streamflow records can be obtained if larger budgets become available.

The analysis also was performed under the assumption that no correlative data at a stream gage were lost to estimate the uncertainty added to the stream-gaging records because of less than perfect instrumentation. The curve, labeled "Without missing record" on figure 15, shows the average standard errors of estimation of streamflow that could be obtained if perfectly reliable systems were available to measure and record the correlative data. For the minimal operational budget of $155,000, the effects of less than perfect equipment are greatest; average standard errors increase from 9.8 to 25.1 percent.

At the other budgetary extreme of $350,000, under which stations are visited more frequently and the equipment should be more reliable, average standard errors increased from 5.4 percent for ideal equipment to 8.7 percent for the current systems of sensing and recording of hydrologic data. Thus, improved equipment can have a very positive impact on streamflow uncertainties throughout the range of operational budgets that might be anticipated for the stream-gaging program in Maine.

Conclusions From the K-CERA Analysis

As a result of the K-CERA analysis, the following conclusions are offered:

1. The policy for definition of field activities in the stream-gaging program should be altered to maintain the current average standard error of estimate of streamflow records of 17.7 percent with a budget of approximately $170,000. This shift would result in some increases and some decreases in accuracy of records at individual sites.
2. The funding for stations with unacceptable accuracies for the data uses should be renegotiated with the data users.
3. The funding made available by implementation of the first two conclusions should be used to establish two or more new stream gages in the Moosehead Plateau region of Maine, where data are particularly sparse.
4. The K-CERA analysis should be repeated with new stations included whenever sufficient information about the characteristics of the new stations has been obtained.
5. Schemes for reducing the probabilities of missing record, for example increased use of local gage observers and satellite relay of data, should be explored and evaluated as to their cost-effectiveness in providing streamflow information.

SUMMARY

Currently, 51 continuous stream gages are operated in Maine at a cost of $211,000. Seventeen separate sources of funding contribute to this program and eight separate uses were identified for data from a single gage. In spite of the size of the program, streamflow data for a large part of Maine's interior are too sparse to provide valid estimates of streamflow characteristics. This paucity should be remedied as funds can be made available.

In an analysis of the uses made of the data, three stations were identified as producing data that are no longer sufficiently needed to warrant continuing their operation. Operation of these stations should be discontinued. Three other stations were identified as having uses specific to short-term studies. These stations should also be deactivated at the end of the data-collection phases of the studies. The remaining 45 stations should be maintained in the program for the foreseeable future.

The current (1984) policy for operation of the 45-station program would require a budget of $180,300 per year. It was shown that the overall level of accuracy of the records at these 45 sites could be maintained with approximately a $170,000 budget, if the allocation of gaging resources among gages was altered. This alteration should take place and the remainder of the currently available money for stream gaging in Maine should be applied to redressing the paucity of data in the interior of the State.

A major component of the error in streamflow records is caused by loss of primary record (stage or other correlative data) at the stream gages because of malfunctions of sensing and recording equipment. Upgrading equipment and developing strategies to minimize lost record appear to be key actions required to improve the reliability and accuracy of the streamflow data generated in the State.

Studies of the cost-effectiveness of the stream-gaging program should be continued and should include investigation of the optimum ratio of discharge measurements to total site visits for each station, as well as investigation of cost-effective ways to reduce the probabilities of lost correlative data. Future studies also will be required because of changes in demands for streamflow information with subsequent addition and deletion of stream gages. Such changes will affect the operation of other stations in the program both because of the dependence between stations of the information that is generated (data redundancy) and because of the dependence of the costs of collecting the data from which the information is derived.

REFERENCES CITED

Benson, M. A., and Carter, R. W., 1973, A national study of the streamflow data-collection program: U.S. Geological Survey Water-Supply Paper 2028, 44 p.

Carter, R. W., and Benson, M. A., 1970, Concepts for the design of streamflow data programs: U.S. Geological Survey open-file report, 33 p.

Draper, N. R., and Smith, H., 1966, Applied regression analysis: New York, N.Y., John Wiley and Sons, 2d ed., 709 p.

Fontaine, R. A., 1981, Drainage areas of surface water bodies of the Penobscot River basin in central Maine: U.S. Geological Survey Open-File Report 78-556-F, 92 p.

———, 1982, Cost-effective stream-gaging strategies for Maine: U.S. Geological Survey Open-File Report 82-507, 43 p.

———, 1983, Uncertainties in records of annual mean discharge for Maine: U.S. Geological Survey Water-Resources Investigations Report 83-4025, 111 p.

Gelb, A., ed., 1974, Applied optimal estimation: The Massachusetts Institute of Technology Press, Cambridge, Mass., 374 p.

Gilroy, E. J., and Moss, M. E., 1981, Cost-effective stream-gaging strategies for the Lower Colorado River Basin: U.S. Geological Survey Open-File Report 81-1019.

Hayes, G. S., and Morrill, R. A., 1970, A proposed streamflow data program for Maine: U.S. Geological Survey open-file report, 38 p.

Hutchinson, N. E., 1975, WATSTORE User's guide, volume 1: U.S. Geological Survey Open-File Report 75-426.

Keefer, T. N., 1974, Desktop computer flow routing: American Society of Civil Engineers Proceedings, Journal of the Hydraulics Division, v. 100, no. HY7, p. 1047-1058.

Keefer, T. N., and McQuivey, R. S., 1974, Multiple linearization flow routing model: American Society of Civil Engineers Proceedings, Journal of the Hydraulics Division, v. 100, no. HY7, p. 1031-1046.

Kleinbaum, D. G., and Kupper, L. L., 1978, Applied regression analysis and other multivariable methods: North Scituate, Mass., Duxbury Press, 556 p.

Mitchell, W. D., 1962, Effect of reservoir storage on peak flow: U.S. Geological Survey Water-Supply Paper 1580, p. C1-C25.

Morrill, R. A., 1975, A technique for estimating the magnitude and frequency of flood in Maine: U.S. Geological Survey Open-File Report 75-292, 44 p.

Moss, M. E., and Gilroy, E. J., 1980, Cost-effective stream-gaging strategies for the Lower Colorado River Basin: U.S. Geological Survey Open-File Report 80-1048, 111 p.

Moss, M. E., Gilroy, E. J., Tasker, G. D., and Karlinger, M.R., 1982, Design of surface-water data networks for regional information: U.S. Geological Survey Water-Supply Paper 2178, 33 p.

Prescott, G. C., Jr., 1963, Reconnaissance of ground-water conditions in Maine: U.S. Geological Survey Water-Supply Paper 1669-T, 52 p.

———, 1980, Ground-water availability and surficial geology of the Royal, upper Presumpscot, and upper Saco River basins, Maine: U.S. Geological Survey Water-Resources Investigations 79-1287, 3 pl.

Riggs, H. C., 1973, Regional analysis of streamflow characteristics: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chapter B3, 15 p.

Sauer, V. B., 1973, Unit response method of open-channel flow routing: American Society of Civil Engineers Proceedings: Journal of the Hydraulics Division, v. 99, no. HY1, p. 179-193.

Thomas, D. M., and Benson, M. A., 1970, Generalization of streamflow characteristics from drainage-basin character-

istics: U.S. Geological Survey Water-Supply Paper 1975, 55 p.
U.S. Geological Survey, 1981, Water Resources Data for Maine, water year 1981: U.S. Geological Survey Water Data Report ME-81-1.

APPENDIX—DERIVATION OF ϵ_f, ϵ_e, AND ϵ_r by M. E. Moss

It is assumed that, if the sensing or recording equipment at a stream gage fails between service visits to the gage, the time, τ, from the last service visit until the failure has a conditional probability distribution that is defined by the truncated negative exponential family

$$f_\tau = ke^{-k\tau}/(1-e^{-ks}) \tag{16}$$

where s is the interval between visits and k is a parameter of the family of probability distributions ($1/k$ is the average time to failure). It also is assumed that the recorder continues to malfunction from the instant of failure until the next service visit. Thus, the fraction of time, ϵ_f, during which the gage can be expected to function properly is

$$\epsilon_f = 1 - E[d]/s \tag{17}$$

where $E[\cdot]$ is the expected value of the random variable contained in the brackets and d is the downtime of the recorder between visits. Down time is defined

$$d=\begin{cases} s-\tau \text{ if a failure occurs,} \\ 0 \text{ if no failure occurs} \end{cases} \tag{18}$$

as is shown in figure 16.

The expected value of down time is

$$E[d]=\int_0^s (s-\tau)f_\tau d\tau \tag{19}$$

which, when evaluated, results in

$$E[d]=(ks+e^{-ks}-1)/k. \tag{20}$$

Substituting equation 20 into equation 17 and simplifying result in

$$\epsilon_f=(1-e^{-ks})/ks. \tag{21}$$

The fraction of time, ϵ_e, for which no record is available at the station of interest and no record is available from an auxiliary site to reconstruct at the station of interest (both caused by equipment failures) is obtainable from a bivariate application of equation 16. If it is assumed that the probability distributions of failure times are identical and independent at the primary and auxiliary sites and that the primary and auxiliary sites are serviced at about the same times, ϵ_e can be evaluated as follows.

The concurrent downtime, d_2, of both stations is defined

$$d_2=\begin{cases} \min(s-\tau_a, s-\tau) \text{ if both stations fail,} \\ 0 \text{ otherwise,} \end{cases} \tag{22}$$

where τ_a is the time to failure at the auxiliary site. The case in which $s-\tau_a$ is the minimum and equals d_2 is shown in figure 17. The value of ϵ_e can be defined in terms of d_2 as

$$\epsilon_e=E[d_2]/s. \tag{23}$$

The expected value of concurrent downtime is

$$E[d_2]=\int_0^s (s-\tau)P[\tau_a\leqslant\tau]f_\tau d\tau+\int_0^s (s-\tau_a)P[\tau\leqslant\tau_a]f_{\tau_a}d\tau_a, \tag{24}$$

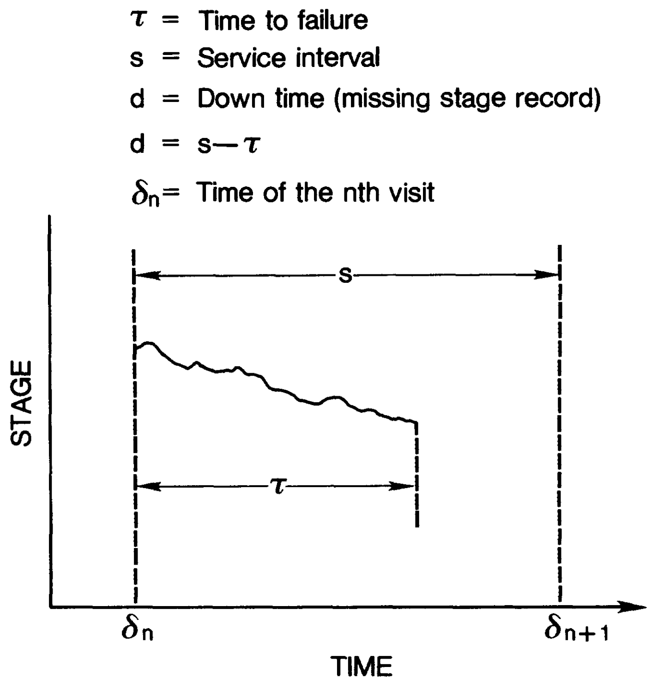

Figure 16. Definition of down time for a single station.

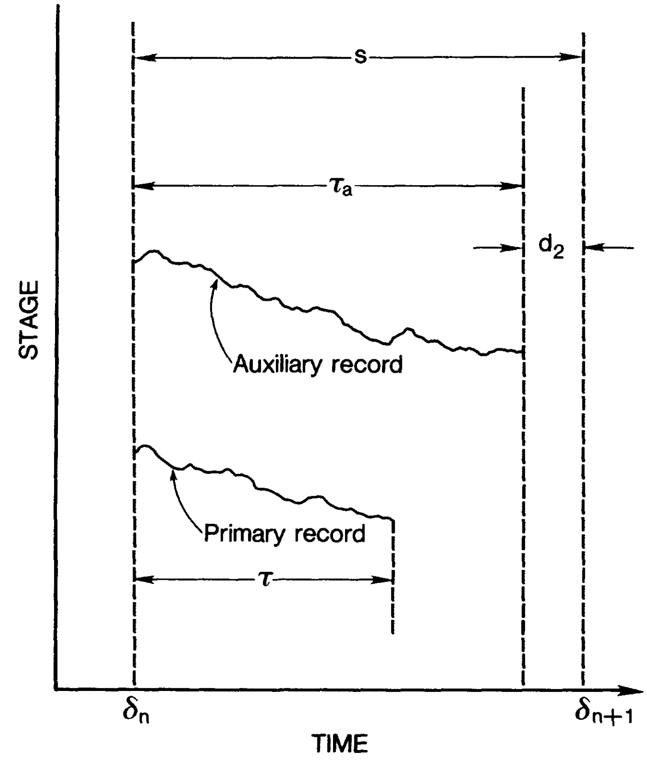

Figure 17. Definition of joint down time for a pair of stations.

where $P[\cdot]$ is the probability of the event contained within the brackets occurring. Evaluation of equation 24 under the given assumptions results in

$$E[d_2]=s-\frac{2}{k}(1-e^{-ks})-\frac{1}{2k}(1-e^{-2ks}), \qquad (24)$$

which can be substituted into equation 23 to obtain ϵ_e.

Because ϵ_f, ϵ_e, and ϵ_r are mutually exclusive and all encompassing

$$\epsilon_f+\epsilon_e+\epsilon_r=1. \qquad (26)$$

From equation 26, ϵ_r can be defined

$$\epsilon_r=1-\epsilon_f-\epsilon_e. \qquad (27)$$

Factors for Converting Inch-Pound Units to International System (SI) Units

Multiply inch-pound units	by	To obtain SI units
	Length	
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
	Area	
square mile (mi^2)	2.590	square kilometer (km^2)
	Volume	
cubic foot (ft^3)	0.02832	cubic meter (m^3)
	Flow	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)