COST-EFFECTIVENESS OF THE U.S. GEOLOGICAL SURVEY'S STREAM-GAGING PROGRAM IN CONNECTICUT

By T. B. Shepard and L. A. Weiss

U.S. Geological Survey

Water Resources Investigations Report 85-4333

Hartford, Connecticut

1988

UNITED STATES DEPARTMENT OF THE INTERIOR

DONALD PAUL HODEL, Secretary

GEOLOGICAL SURVEY

Dallas L. Peck, Director

For additional information
write to:

Chief, Connecticut Office
U.S. Geological Survey
Abraham A. Ribicoff Federal Building
450 Main Street, Room 525
Hartford, Connecticut 06103

Copies of this report can
be purchased from:

Open-File Services Section
Western Distribution Branch
U.S. Geological Survey
Box 25425, Federal Center
Denver, Colorado 80225
(Telephone [303] 236-7476)

CONTENTS

ILLUSTRATIONS

TABLES

FACTORS FOR CONVERTING INCH-POUND TO METRIC (SI) UNITS

Multiple 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)

COST-EFFECTIVENESS OF THE U.S. GEOLOGICAL SURVEY'S STREAM-GAGING PROGRAM IN CONNECTICUT

By Thomas B. Shepard and Lawrence A. Weiss

ABSTRACT

This report documents the results of a study of the cost-effectiveness of the stream-gaging program in Connecticut. Data uses and funding sources were identified for 59 stream-gaging stations currently operated in Connecticut. These 59 stations, and 3 stations in Massachusetts, are operated with a 1984 budget of $267,000; included in this budget figure is the operation of equipment at 17 ground water wells and 11 reservoirs.

The current policy for operation of the 90-site program would require a budget of $267,000 per year. The average standard error of estimation of streamflow records is 14.5 percent. It was shown that the overall level of accuracy could be reduced from 14.5 percent to 11.7 percent if current budget of $267,000 was reallocated among the gages.

A minimum budget of $255,000 is required to operate the program; a budget less than this does not permit proper service and maintenance of the gages and recorders. At the minimum budget, the average standard error is 16.3 percent. The maximum budget analyzed was $350,000, which resulted in an average standard error of 6.6 percent.

INTRODUCTION

The U.S. Geological Survey 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 Survey. The data are collected in cooperation with State and local governments and other Federal agencies. The Survey is presently (1984) operating approximately 8,000 continuous-record gaging stations throughout the Nation. Some of these records extend back to the turn of the century. Any long-term activity, 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 Survey presently is (1984) undertaking another nationwide analysis of the stream-gaging program that will be completed over a 5-year period (1983-87), with 20 percent of the program being analyzed each year.

Purpose and Scope

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. Gaging stations are categorized as to whether the data are available 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 alternative 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 estimate 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.

The standard errors of estimate given in the report are those that would occur if daily discharges were computed through the use of methods described in this study. No attempt has been made to estimate standard errors for discharges that are computed by other means. Such errors could differ from the errors computed in the report. The magnitude and direction of the differences would be a function of methods used to account for shifting controls and for estimating discharges during periods of missing record.

This report is organized into five sections; the first being a description of the stream-gaging activities in Connecticut and an introduction to the analysis. 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 the previous results, conclusions are given at the end of each of the middle three sections. The complete study is summarized in the final section.

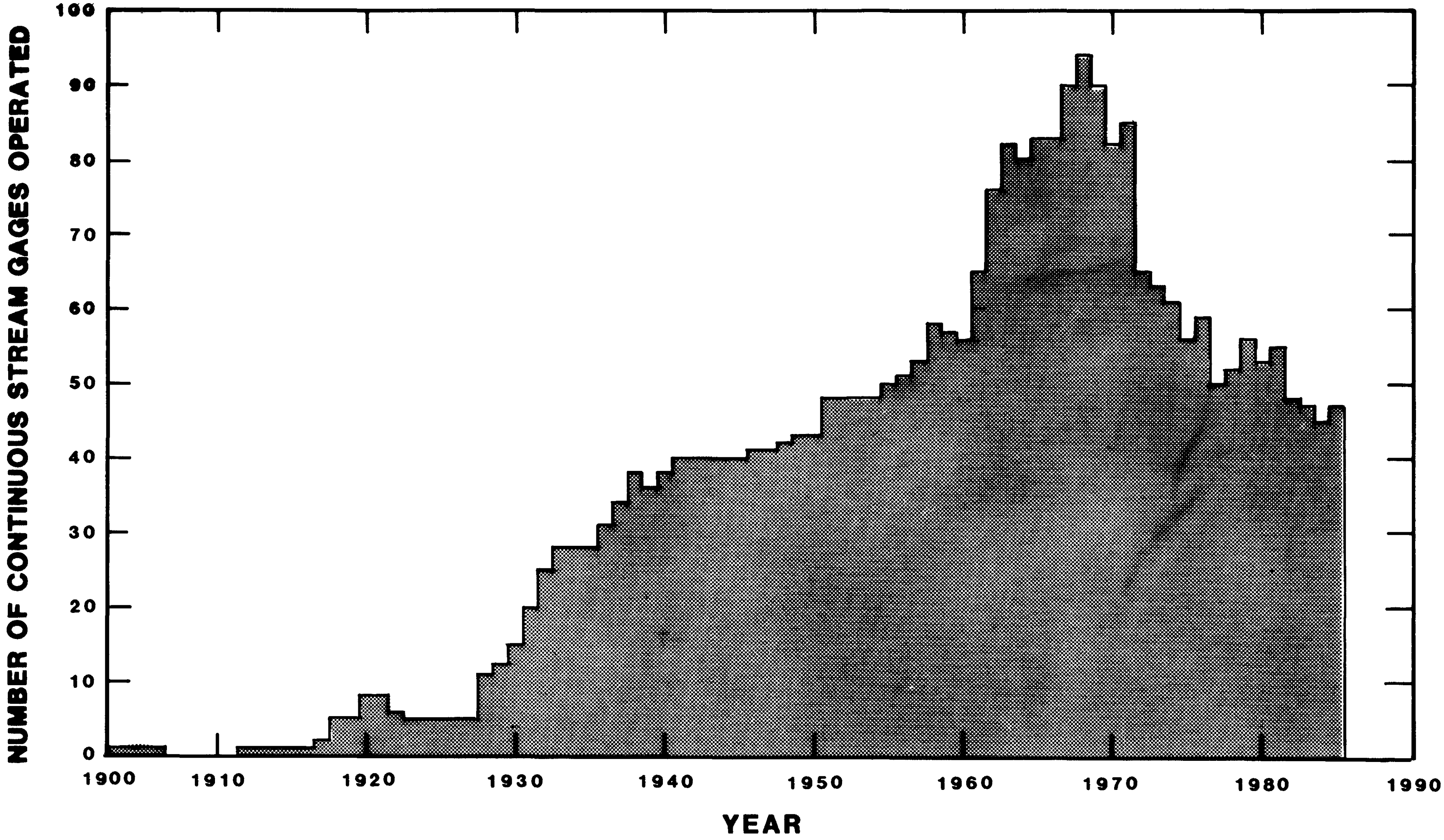


Figure 1.--History of continuous stream gaging in Connecticut.

History of the Stream-Gaging Program in Connecticut

The streamflow program of the U.S. Geological Survey in Connecticut has evolved through the years as Federal and State interests required information at specific sites for water management and for definition of surface-water hydrology throughout the State.

Prior to 1917 only one gaging station had been maintained in continuous operation in the State. Between 1917 and 1928 there was an average of five gaging stations. The impetus to expand the program was given in 1928 when in the course of litigation over the diversion of water from the Connecticut River basin to Metropolitan Boston there was brought to light an urgent need for more streamflow data. Four gaging stations were established in the lower Connecticut River basin, and one in both the Shetucket River basin and the Naugatuck River basin. The program gradually expanded to a maximum of 94 gaging stations in 1968.

In 1960, a network of 95 partial-record stations was established to define low-flow characteristics at sites other than those where daily records were collected. Forty-five of these partial-record stations also served as crest-stage stations to define peak flow characteristics for small (under 10 mi^2) drainage areas. This program continued through the 1978 water year when the low-flow characteristics portion was discontinued. The crest-stage program was discontinued at the end of the 1984 water year.

A report by Thomas and Cervione (1970) evaluated the surface-water program in Connecticut and provided guidelines for planning future programs. In 1972, using the results of the Thomas and Cervione report, 20 gaging stations were eliminated from the Connecticut gaging program.

Between 1972 and 1976, the number of gaging stations decreased to 59. In 1976, as a result of reduction in cooperative funding by the State of Connecticut a second analysis of the data collection program was undertaken using methodology of Thomas and Cervione (1970). Based on this analysis, eight additional gaging stations were eliminated from the Connecticut stream-gaging network.

Between 1977 and 1981, five gaging stations were added to the gaging program and, in 1982, seven gaging stations were eliminated from the program. The decision to drop these stations was based on a Network Analysis for Regional Information Study (NARI) by Weiss (1983). Two gaging stations were discontinued following completion of special projects in 1983. In 1984 one gaging station and two sites being measured for rating only were added to the program leaving the Connecticut office with 47 continuous-record gaging stations. In addition, Connecticut has three tidal sites, three sites on the Connecticut River used for tidal-volume interchange on an estuary, six sites measured for rating only, two sites located in Massachusetts where the Connecticut Office is conducting measurements and maintenance of the gages for the Boston Office of the New England District, and a new site, located in Holyoke, Massachusetts, that began in November, 1983 and is being operated for Northeast Utilities as mandated by the Federal Energy Regulatory Commission (FERC).

The history of gaging station operation in Connecticut is shown in figure 1; gaging stations in Massachusetts that are operated by the Connecticut office are not included in figure 1.

Current Connecticut Stream-Gaging Program

Connecticut can be divided into three major physiographic regions (Fenneman 1938): the New England Upland, the Connecticut Valley Lowlands, and the Seaboard Lowland. The locations of these regions and the distribution of the 59 gaging stations in Connecticut, are shown in figure 2. Thirty-three stations are located in the New England Upland, 11 are located in the Connecticut Valley Lowlands, and 15 are located in the Seaboard Lowland. Figure 2 shows gaging stations are fairly uniformly distributed across the State.

The cost of operating these 59 gaging stations, the three stations in Massachusetts, and the equipment at 17 ground water wells and 11 reservoirs in fiscal year 1984 was $267,000. Selected hydrologic data for the 59 stations and the FERC site including drainage area, period of record, and, for selected stations, mean annual flow, are given in table 1. Station identification numbers used throughout this report are the USGS's eight-digit downstream-order station number. Table 1 also provides the official name of each stream gage.

Table 1.--Selected hydrologic data for stations in the Connecticut surface-water program

[All stations are located in Connecticut except as noted]

Station no.	Station name	Drainage area (square miles)	Period of record	Mean annual flow (cubic feet per second)
01118300	Pendleton Hill Brook near Clarks Falls	4.02	July 1958-	8.50
01119040	Poquonock River near Groton	20.8	January 1973-	___ 1/
01119500	Willimantic River near Coventry	121	September 1931-	213 2/
01121000	Mount Hope River near Warrenville	28.6	July 1940-	50.9
01122000	Natchaug River at Willimantic	174	October 1930-	301 2/
01122500	Shetucket River near Willimantic	404	September 1928-	709 2/
01123000	Little River near Hanover	30.0	July 1951-	56.7
01124000	Quinebaug River at Quinebaug	155	September 1931-	270 2/
01124151	Quinebaug River at West Thompson	172	June 1966-	311 2/
01125500	Quinebaug River at Putnam	328	December 1929- September 1969	546 3/
01127000	Quinebaug River at Jewett City	713	July.1918-	1,273 2/
01127500	Yantic River at Yantic	89.3	October 1930-	163
01172003	Connecticut River at Holyoke, Mass.	8,309	November 1983-	___ 4/
01184000	Connecticut River at Thompsonville	9,660	July 1928-	16,380
01184100	Stony Brook near West Suffield	10.4	September 1960-April 1981 5/ May 1981-	___ 4/
01184490	Broad Brook at Broad Brook	15.5	August 1961-September 1976/ May 1982-	23.1
01186000	West Branch Farmington River at Riverton	131	August 1955-	249 2/
01186500	Still River at Robertsville	85.0	July 1948-September 1967/ July 1969-	173 2/
01187300	Hubbard River near West Hartland	19.9	January 1938-September 1955/ September 1956-	39.2
01188000	Burlington Brook near Burlington	4.10	September 1931-	8.20
01188090	Farmington River at Unionville	378	October 1977-	743 2/
01189000	Pequabuck River at Forestville	45.8	July 1941-	85.3 2/
01189995	Farmington River at Tariffville	577	October 1971-	1,270 2/
01190000	Farmington River at Rainbow	590	August 1928-	1,093 2/
01190070	Connecticut River at Hartford	10,493	January 1905-	___ 6/
01191000	North Branch Park River at Hartford	26.8	October 1936-	38.4
01192500	Hockanum River near East Hartford	73.4	September 1919-September 1921/ July 1928-September 1971/ 1972-76 7/ / October 1976-	114
01192883	Coginchaug River at Middlefield	29.8	October 1961- 8/	54.1
01193000	Connecticut River near Middletown	10,887	October 1965-	___ 6/
01193050	Connecticut River near Middle Haddam	10,897	October 1965-	___ 6/
01193500	Salmon River near East Hampton	100	July 1928-	183
01194825	Connecticut River at Old Saybrook	11,269	October 1979-	___ 1/
01195100	Indian River near Clinton	5.68	November 1981-	___ 4/
01195146	Pond Meadow Brook below Kroopa Pond at Killingworth	5.92	January 1983-	___ 4/
01196500	Quinnipiac River at Wallingford	115	October 1930-	211 2/
01196590	Mill River near Cheshire	5.54	June 1978-	___ 9/

See footnotes at end of table.

Table 1.--Selected hydrologic data for stations in the Connecticut surface-water program- Continued

[All stations are located in Connecticut except as noted]

Station no.	Station name	Drainage area (square miles)	Period of record	Mean annual flow (cubic feet per second)
01196600	Willow Brook near Cheshire	9.40	June 1978-	___ 9/
01196620	Mill River near Hamden	24.5	October 1978	56.8
01196651	West River near Westville	29.5	December 1983-	___ 9/
01199000	Housatonic River at Falls Village	634	July 1912-	1,086
01199050	Salmon Creek at Lime Rock	29.4	October 1961-	47.9
01199290	Housatonic River at Kent	756	June 1984-	___ 4/
01200000	Tenmile River near Gaylordsville	203	October 1929-	302
01200500	Housatonic River at Gaylordsville	996	July 1940-	1,684
01204000	Pomperaug River at Southbury	75.1	June 1932-	127
01205500	Housatonic River at Stevenson	1,544	August 1928-	2,605 2/
01205600	West Branch Naugatuck River at Torrington	33.8	August 1956-	58.0
01205700	East Branch Naugatuck River at Torrington	13.6	August 1956-	24.4 2/
01206900	Naugatuck River at Thomaston	99.8	October 1959-	201 2/
01208013	Branch Brook near Thomaston	20.8	June 1971-	35.9
01208171	Naugatuck River at Waterbury	174	November 1982-	___ 9/
01208325	Mad River at Waterbury	26.3	December 1983-	___ 9/
01208420	Hop Brook near Naugatuck	16.3	October 1969-	35.6
01208500	Naugatuck River at Beacon Falls	260	June 1918-September 1924/ September 1928- 10/	495 2/
01208873	Rooster River at Fairfield	10.9	June 1977-	17.9
01208925	Mill River near Fairfield	28.6	October 1972-	42.1
01208950	Sasco Brook near Southport	7.38	October 1964-	13.5
01208990	Saugatuck River near Redding	21.0	October 1964-	41.2
01209700	Norwalk River at South Wilton	30.0	September 1962-	55.8
01209788	Stamford Hurricane Barrier at Stamford	3.25	October 1972-	___ 1/

1/ No mean annual flow published, Tidal site. Maximum and minimum monthly tide published.

2/ Adjusted for storage.

3/ Adjusted for storage; currently being measured for rating only.

4/ No mean annual flow published, less than 5 years of streamflow record.

5/ Operated as a crest-stage indicator site.

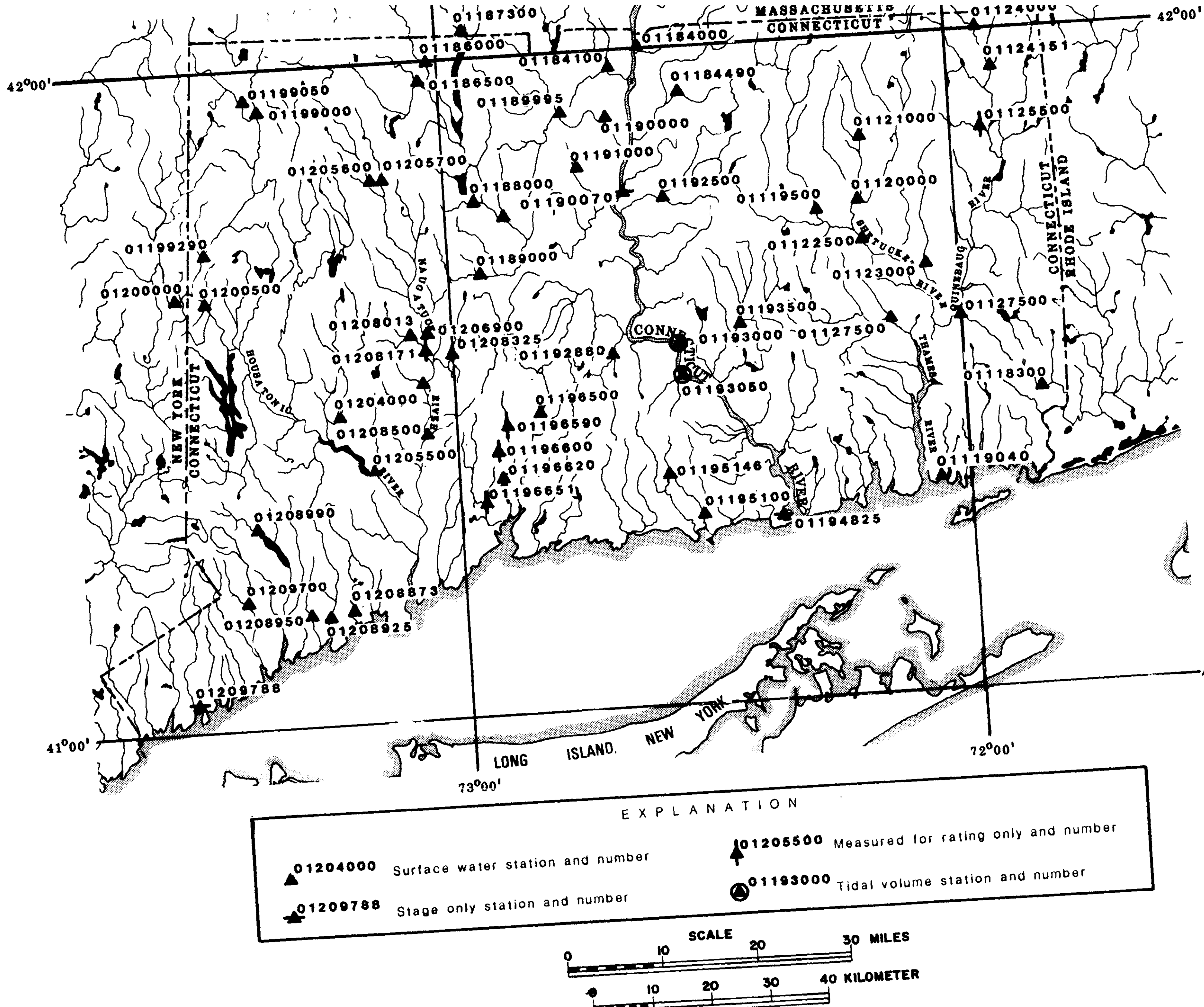
6/ Stage only; used in lower Connecticut River flow model.

7/ Operated as crest-stage indicator site water year 1972-1976.

8/ Prior to December 1980 published as "at Rockfall."

9/ Being measured for rating only.

10/ Prior to October 1955 published as "near Naugatuck."



stream gages in Connecticut.

USES, FUNDING, AND AVAILABILITY OF CONTINUOUS STREAMFLOW DATA

The relevance of a stream gage is defined by the uses that are made of the data that are produced from the gage. The uses of the data from each gage in the Connecticut program were identified by a survey of known data users. The survey documented the importance of each gage.

Data uses identified by the survey were divided into nine categories, 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 Categories

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

Regional Hydrology

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

Twenty-two gaging stations in the Connecticut network are classified in the regional hydrology data-use category. Five of the stations are designated index stations, and are used to indicate current hydrologic conditions. The locations of gaging stations that provide information about regional hydrology are given in figure 3.

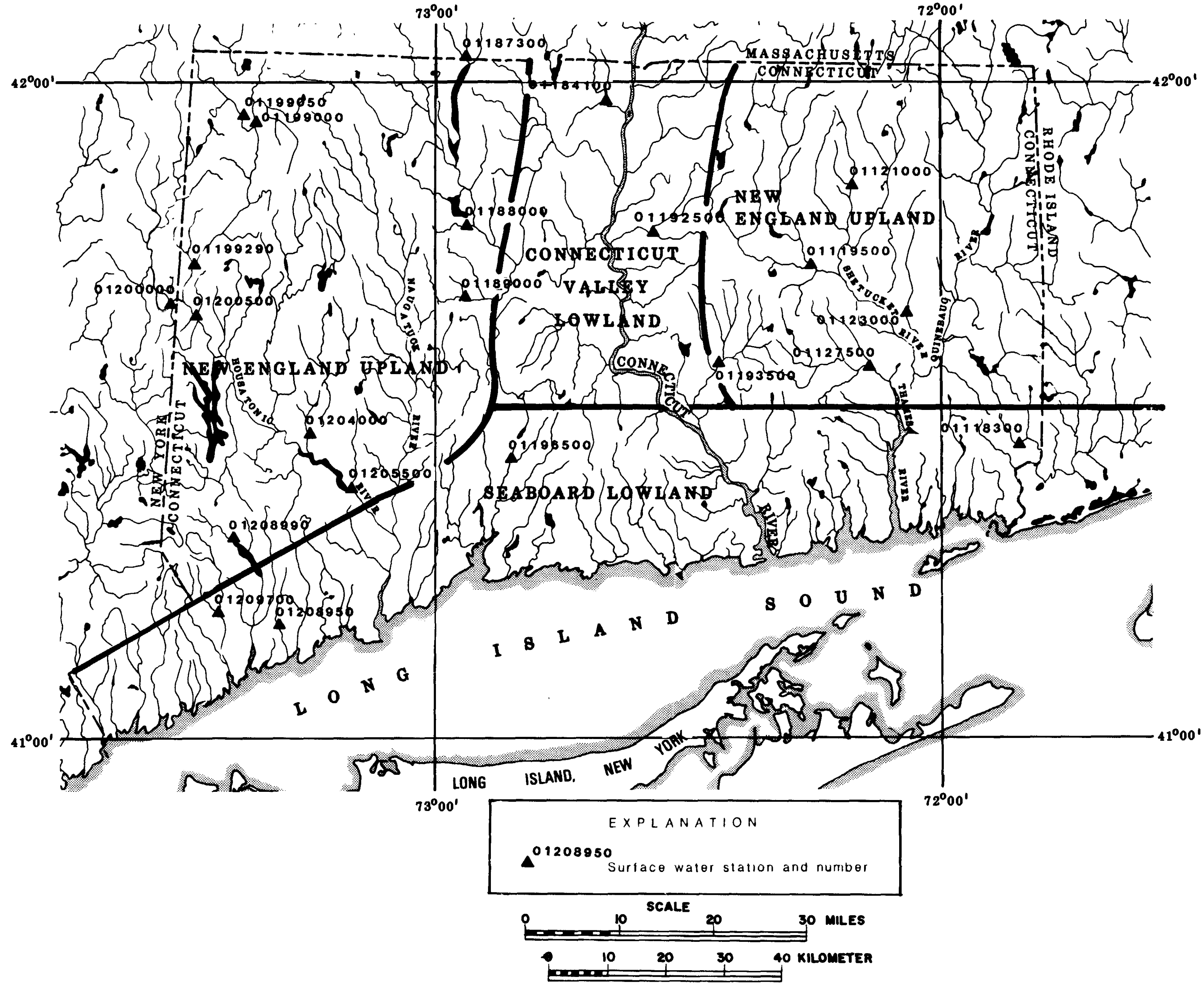


Figure 3.--Location of regional hydrology stream gages.

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 as hydrologic systems stations. They include diversions and return flows and stations that are useful for defining the interaction of water systems.

The index stations are included in the hydrologic systems category because they are accounting for current and long-term conditions of the hydrologic systems that they gage. The FERC station is also included. The data collected at the FERC station are used to monitor the compliance of control structures to downstream flow requirements determined by FERC.

In addition to the index stations and the FERC station, 27 stations are classified in the hydrologic systems data-use category in Connecticut.

Legal Obligations

Some stations provide records of flows for the verification or enforcement of existing treaties, compacts, and decrees. The legal obligation category contains only those stations that the U.S. Geological Survey is required to operate to satisfy a legal responsibility.

There are no stations in the Connecticut program used to fulfill a legal responsibility of the Survey.

Planning and Design

Gaging stations in this category 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 installed for such purposes and where these purposes are still valid.

Ten stations in the Connecticut program are used for planning and design: six of these are used by the Metropolitan District for public-water supply information, one by Connecticut Department of Environmental Protection, Water Compliance Unit, for a phosphorous loading study, two by the Town of Fairfield for land-use decisions and a river-trunk sewer project, and one for an Atlantic Salmon restoration study.

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 the data are routinely available to the operators on a rapid-reporting basis. For projects on large streams, data only may be needed every few days.

Twenty-five stations in the Connecticut program are used for project operations. Twenty-one of these are used to aid operators in management of reservoirs and control structures in a flood-control network; four stations are used to assist water-supply plant operators.

Hydrologic Forecasts

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

In Connecticut, seventeen gaging stations included in the hydrologic forecast category are used for flood forecasting by U.S. National Weather Service (NWS) and the NOAA River Forecast Center in Bloomfield. Additionally, NWS uses the data at some stations as input to longer range prediction models.

Water-Quality Sites

Gaging stations where regular water-quality or sediment-transport studies are being 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 sites.

Four National Stream-Quality Accounting Network (NASQAN) stations and one estuary-gaging water-quality station are maintained. NASQAN stations are part of a country-wide network designed to assess water-quality trends of significant streams. Twenty-one stations are part of a statewide water-quality network. In addition 13 water-quality sites use gaging stations to estimate flows by drainage-area ratios.

Research

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

One station in Connecticut is used in the support of research activity. The Connecticut Department of Environmental Protection, Natural Resources Center and the University of Connecticut are using this data in the development of a basin hydrologic model.

Other

In addition to the eight data-use categories described above, three stations are used to provide data on the tidal levels in Long Island Sound, three stations provide data on tidal volume, and 9 stations provide data for recreational use.

Funding

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

1. Federal program.--Funds that have been directly allocated to the 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. Funds in this category are not matched by USGS cooperative funds.

In all four funding categories the 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 may not necessarily be the same as those identified herein.

Eight Federal, State, and local agencies currently are contributing funds to the Connecticut stream-gaging program.

Frequency of Data Availability

Frequency of data availability refers to the time at which the streamflow data may be furnished by the users. 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 the annual data report published by the Survey for Connecticut (U.S. Geological Survey, 1983). These three categories are designated T, P, and A, respectively, in table 2. In the current Connecticut program, data for 54 of the 59 stations analyzed are made available through the annual report, data from 12 stations are available on a real-time basis, and data are released on a periodic basis at 19 stations.

Data-Use Presentation

Data-use and ancillary information are presented for each gaging station in table 2, which includes footnotes to expand the information conveyed. The entry of an asterisk in the table indicates that no footnote is required.

Conclusions Pertaining to Data Uses

A review of the data-use and funding information presented in table 2 indicates that all current gaging operations should be continued.

Table 2.--Data-use table.

[★, no footnote required]

STATION NUMBER	DATA USE									FUNDING				FREQUENCY OF DATA AVIALABILITY
	REGIONAL HYDROLOGY	HYDROLOGIC SYSTEMS	LEGAL OBLIGATIONS	PLANNING AND DESIGN	PROJECT OPERATION	HYDROLOGIC FORECASTS	WATER-QUALITY MONITORING	RESEARCH	OTHER	FEDERAL PROGRAM	OFA PROGRAM	CO-OP PROGRAM	OTHER NON-FEDERAL	
01118300	★	★										1		A
01119040									2			1		AP
01119500	★	★					3					1		AP
01121000	5	5				5						1		AP
01122000					★		3				6			ATP
01122500		★			★	7	3,8				6	1		ATP
01123000	5	5										1		A
01124000		★					3					1		AP
01124151					★		3				6			A
01125500					★		★				6			
01127000		★			★	7	3,8				6	1		AP
01127500	★	★				7						1		AP
01172003		9	★								10			
01184000		★				7	3,8		4			1		AYP
01184100	★	★					3					1		A
01184490								11				1		
01186000		★		12		★			4			1		A
01186500		★		12	★		3		4		6	1		A
01187300	★	★		12								1		AP
01188000	5	5				5	3			★		13		AP
01188090		★		12	★		3				6	1		A
01189000	★	★					3					1		A

1. Connecticut Department of Environmental Protection.
2. Interstate consideration, tidal action for Long Island Sound.
3. Connecticut Department of Environmental Protection - Water Compliance Unit.
4. Recreation.
5. Long-term index gaging station.
6. U.S. Army Corps of Engineers - New England Division.
7. Flood forecasting - U.S. National Weather Service.
8. NASQAN.
9. Federal Energy Regulatory Commission hydropower licensing requirements.
10. Northeast Utilities.
11. Connecticut Department of Environmental Protection - Natural Resources Center - University of Connecticut basin hydrologic model.
12. The Metropolitran District Commission diversion proposal.
13. City of New Britain - Board of Water Compliance.

Table 2.--Data-use table. --Continued

[★, no footnote required]

STATION NUMBER	DATA USE									FUNDING				FREQUENCY OF DATA AVIALABILITY
	REGIONAL HYDROLOGY	HYDROLOGIC SYSTEMS	LEGAL OBLIGATIONS	PLANNING AND DESIGN	PROJECT OPERATION	HYDROLOGIC FORECASTS	WATER-QUALITY MONITORING	RESEARCH	OTHER	FEDERAL PROGRAM	OFA PROGRAM	CO-OP PROGRAM	OTHER NON-FEDERAL	
01189995		★		12		7	3		4			1		AT
01190000				12	★	7	★		4		6	1		AT
01190070						7	3		14			1		ATP
01191000						7						1		AT
01192500	★	★				7	3			★				A
01192883							★					1		A
01193000					★	7			14	★	6			AT
01193050					★		15		14	★	6			A
01193500	5	5		★		5	3		4			1		AP
01194825									2			1		A
01195100												1		A
01195146					★								16	
01196500	★	★					3					1		AP
01196590					★								16	
01196600					★								16	
01196620					★								16	A
01196651											6			
01199000	★	★					3		4			1		A

1. Connecticut Department of Environmental Protection.
2. Interstate consideration, tidal action for Long Island Sound.
3. Connecticut Department of Environmental Protection - Water Compliance Unit.
4. Recreation.
5. Long-term index gaging station.
6. U.S. Army Corps of Engineers - New England Division
7. Flood forecasting - U.S. National Weather Service.
12. The Metropolitan district diversion proposal.
14. Tidal station - Tidal volume interchange on estuary.
15. CBR - Water-quality estuary monitoring station above a nuclear power station and below a conventional power station.
16. South Central Connecticut Regional Water Authority.

Table 2.--Data-use table. --Continued

[★, no footnote required]

STATION NUMBER	DATA USE									FUNDING				FREQUENCY OF DATA AVIALABILITY
	REGIONAL HYDROLOGY	HYDROLOGIC SYSTEMS	LEGAL OBLIGATIONS	PLANNING AND DESIGN	PROJECT OPERATION	HYDROLOGIC FORECASTS	WATER-QUALITY MONITORING	RESEARCH	OTHER	FEDERAL PROGRAM	OFA PROGRAM	CO-OP PROGRAM	OTHER NON-FEDERAL	
01199050	★	★										1		A
01199290	★	★					3					1		AP
01200000	★			3								1		A
01200500	★	★				7	3		4	★				A
01204000	5	5				5						1		AP
01205500	★	★			★	7	3,8		4	★	6			AP
01205600					★							1,17		A
01206700					★							1,17		A
01206900					★		3				6	1		AT
01208013					★		3				6			A
01208171					★						6			
01208325					★						6			
01208420					★						6			A
01208500					★		3				6	1		AT
01208873				★								18		AP
01208925				★								18		AP
01208950	★	★										1		A
01208990	★	★					3					1		AT
01209700	★	★					3					1		A
01209788					19				2		6			ATP

1. Connecticut Department of Environmental Protection.
2. Interstate consideration, tidal action for Long Island Sound.
3. Connecticut Department of Environmental Protection - Water Compliance Unit.
4. Recreation.
5. Long-term index gaging station.
6. U.S. Army Corps of Engineers - New England Division.
7. Flood forecasting - U.S. National Weather Service.
8. NASQAN.
17. Town of Torrington, Connecticut
18. Town of Fairfield, Connecticut
19. Hurricane Barrier.

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 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 gaging station. There are guidelines concerning suitable accuracies for particular uses of the data; therefore, judgement is required in deciding if the accuracy of the estimated daily flows is suitable for the intended purpose. The data uses at a station will influence its potential for alternative methods. For example, stations for which flood hydrographs are required in a real-time sense, such as hydrologic forecasts and project operation, are not candidates for the alternative methods. Likewise, a legal obligation to operate a gaging station would preclude utilizing alternative methods. The primary candidates for alternative methods are stations that are operated upstream or downstream of other stations on the same stream. The accuracy of the estimated streamflow at these stations may be suitable because of the high redundancy of flow information between stations. Similar watersheds, located in the same physiographic and climatic area, also may have potential for alternative methods.

All stations in the Connecticut program were categorized as to their potential utilization of alternative methods and selected methods were applied at three stations. The categorization of gaging stations and the application of the specific methods are described in subsequent sections of this report. This section briefly describes the two alternative methods that were used and documents why these specific methods were chosen.

Because of the short time frame of this analysis, only two methods were considered. Desirable attributes of a proposed alternate 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 Survey WATSTORE (Water Data Storage and Retrieval System) 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 rather 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 regression analysis.

Flow-Routing Model

Computer model CONROUT (Doyle and others, 1983) was selected to route streamflow from one or more upstream locations to a downstream location by the unit-response convolution method. Downstream hydrographs are produced by the convolution of upstream hydrographs with their appropriate unit-response functions. These functions were defined using the diffusion analogy method (Keefer, 1974; Keefer and McQuivey, 1974).

The convolution procedure 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. This model can only be applied at a downstream station when there is an upstream station 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 are achieved with observed upstream and downstream hydrographs and estimates of tributrary inflows. The model has the capability of combining hydrographs, multiplying hydrographs 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 location. Routing can be accomplished using any equal-interval streamflow data; only daily streamflow data are used in this analysis.

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 procedure 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 is satisfactory in many instances is the multiplication of known flows at an index gaging station by a factor--for example, a drainage-area ratio.

In the diffusion analogy method, the two parameters required to define the unit-response function are K_0, a wave dispersion or damping coefficient, and C_0, the flood wave celerity. K_0 controls the spreading of the wave and C_0 controls the traveltime. In the single linearization method, only one K_0 and C_0 value are used to define one unit-response function (linearization about a single discharge).

Adequate routing of daily flows can usually be accomplished using the single linearization method 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. In the multiple linearization method, C_0 and K_0 are varied with discharge so a table of wave celerity (C_0) versus discharge (Q) and a table of dispersion coefficient (K_0) versus discharge (Q) are used.

In the diffusion-analogy method, the two parameters are calibrated by trial and error. The analyst must decide if suitable parameters have been derived by comparing the simulated discharge to the observed discharge.

Flow-Routing Analysis--Shetucket River Results

A flow-routing model was developed to simulate daily mean discharges at Shetucket River near Willimantic, station 01122500, using upstream stations on Willimantic River and Natchaug River. A schematic diagram of the Shetucket River study area is presented in figure 4. Streamflow data available for this analysis are summarized in table 3.

Shetucket River gage is located 7.58 mi downstream from Willimantic River gage, 3.0 mi downstream from Natchaug River gage, however, and is 1.3 mi downstream from the confluence of the Willimantic and Natchaug Rivers. The sum of the drainage areas at Willimantic and Natchaug gages is 295 mi^2, leaving 109 mi^2 of ungaged area between them and the Shetucket gage. A coefficient of 0.89 times Willimantic River drainage area was used to account for this ungaged area. The sum of the adjusted Willimantic daily values and the Natchaug daily values were then added to the routed Willimantic values, resulting in a simulated daily discharge record for the Shetucket River. Diffusion analogy with a single linearization was used to simulate flows.

To route flow from the Willimantic to the Shetucket gaging station model parameters C_0 (floodwave celerity) and K_0 (wave dispersion coefficient) were determined. The coefficients C_0 and K_0 are functions of channel width (W_0) in feet (ft), channel slope (S_0) in feet per foot (ft/ft), the slope of the stage-discharge relation (dQ_0/dY_0) in square feet per second (ft^2/s), and the discharge (Q_0) in cubic feet per second (ft^3/s) which are representative of the reach being studied and are determined by the following equations:

$$C_0 = \frac{1}{W_0} \frac{dQ_0}{dY_0}, \quad \text{and} \qquad (1)$$

$$K_0 = \frac{Q_0}{2S_0 W_0} \qquad (2)$$

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

Station number	Station name and location	Drainage area (square miles)	Period of record used in study
01119500	Willimantic River near Coventry	121	October 1952 - September 1982
01122000	Natchaug River at Willimantic	174	October 1952 - September 1982
01122500	Shetucket River near Willimantic	404	October 1952 - September 1982

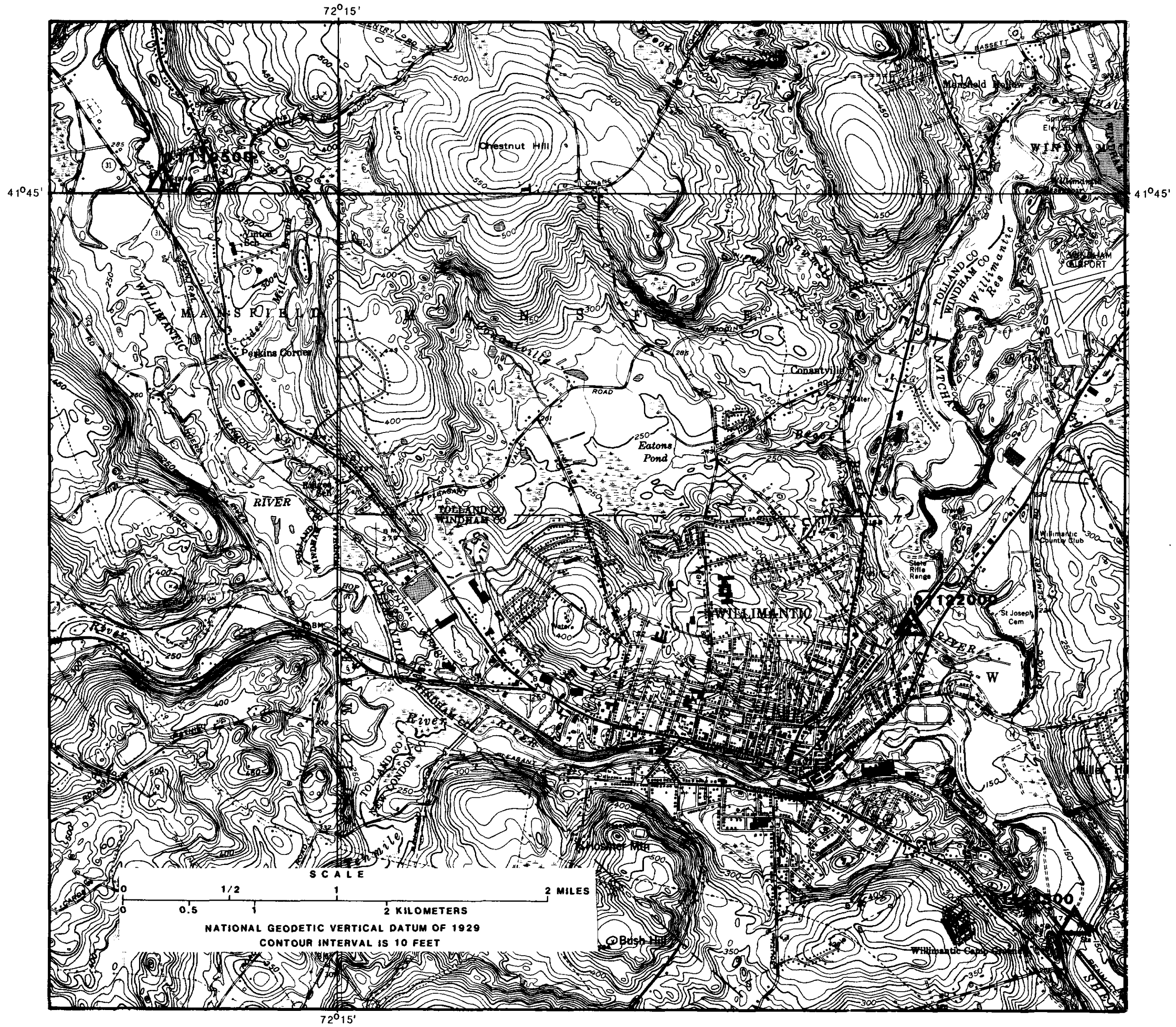


Figure 4.--The Shetucket River basin study area.

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

Station number	Type of flow	Discharge, Q_o (cubic feet per second)	Average width, W_o (feet)	Slope, S_o (feet/foot)	dQ_o/dY_o (feet squared per second)	C_o (feet per second)	K_o (feet squared per second)
01119500	Mean	213	70	2.708×10^{-3}	220	3.143	562
01119500	7-day, 10-year low flow	15.0	40	2.708×10^{-3}	110	2.750	69.2
01119500	Peak flow 10-year recurrence interval	5,240	500	2.708×10^{-3}	1,600	3.200	1,935
01122500	Mean	709	150	2.386×10^{-3}	500	3.333	991
01122500	7-day, 10-year low flow	44.7	100	2.386×10^{-3}	110	1.100	93.7
01122500	Peak flow 10-year recurrence interval	11,900	500	2.386×10^{-3}	1,800	3.600	4,987

The discharge, Q_0, for which C_0 and K_0 were initially linearized was the mean flow, 7-day, 10-year low flow, and for peak flow, the flow for a 10-year recurrence interval. Channel width (W_0) for low and mean flow was determined from discharge measurements made near the gages and for peak flows from flood studies of the Federal Emergency Management Agency. Channel slope, S_0, was determined by a technique used by Benson, 1962. The slope of the stage-discharge relations, dQ_0/dY_0, was determined from rating curves at each gage by using a 1-foot increment that bracketed the mean discharge, Q_0. The difference in discharge through the 1-foot increment then represents the slope of the function at that point. Rating model parameters as determined above are shown in table 4. Using the 1981 water year, daily mean flows of the Willimantic, Natchaug, and Shetucket Rivers for calibration, the single-linearization model was determined to be that with a C_0 = 1.7 ft/s, K_0 = 776 ft^2/s, and an ungaged drainage area ratio of 0.89.

A summary of the simulation of mean daily discharge for the Shetucket River for the 1981 water year is given in table 5. Although the root mean square error was less than 10 percent, and the total volume error was less than 1 percent, only 79 percent of all observations had errors less than 10 percent. A daily hydrograph for October and February (fig. 5) indicates winter flows are simulated as accurately as fall flows, and daily mean discharges in the low, mean, and high discharge ranges are simulated with equal accuracy. Use of the multiple-linearization model did not appreciably reduce the errors.

Table 5.--Results of routing model for Shetucket River for 1981 water year.

1981 WATER YEAR SUMMARY

Mean absolute error (%) for 365 days = 6.85
Mean - error (%) for 180 days = -6.85
Mean + error (%) for 185 days = 6.85
Q1 (observed) volume (SFD) = 162046
Q2 (simulated) volume (SFD) = 161449
Volume error (%) = 0.37
RMS error (%) = 8.95

45 percent of total observations had errors <= 5 percent
79 percent of total observations had errors <= 10 percent
89 percent of total observations had errors <= 15 percent
96 percent of total observations had errors <= 20 percent
98 percent of total observations had errors <= 25 percent
2 percent of total observations had errors > 25 percent

Final model parameters: C_o = 1.7 ft/s, K_o = 756 ft^2/s

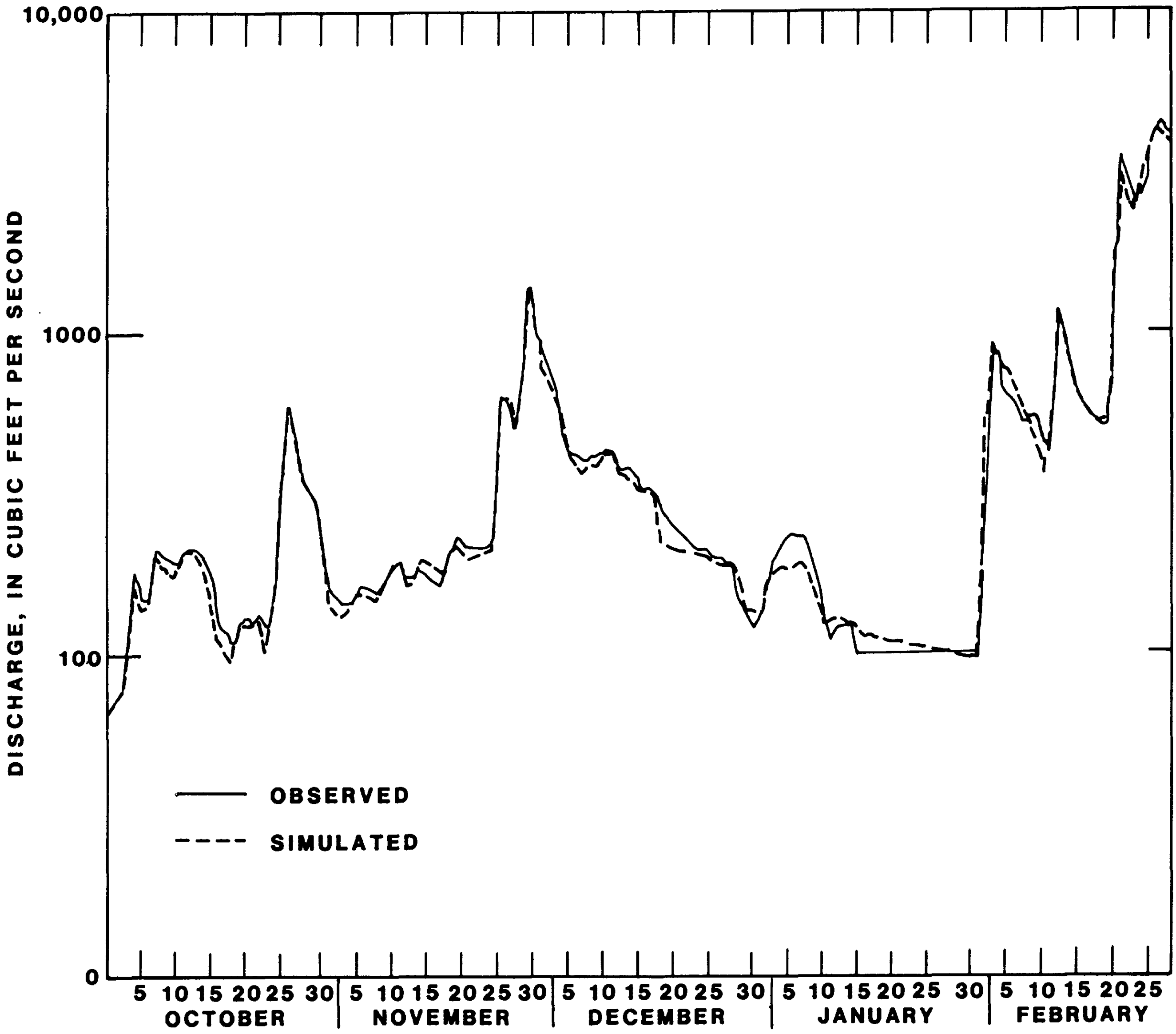


Figure 5.--Daily hydrograph of Shetucket River near Willimantic for fall and winter of 1981 water year.

Flow-Routing Analysis--Housatonic River Results

Housatonic River study area and gaging station locations are shown in figure 6. Gaging-station data available are summarized in table 6.

Gaging station 01199000 is just downstream from a Connecticut Light and Power Company hydroelectric plant (CL&P). The confluence of the Housatonic River and Salmon Creek is 1.7 mi downstream from station 01199000. The intervening drainage area between the Falls Village gage and the confluence is 41.0 mi^2, 29.4 mi^2 of which is gaged at station 01199050. The entire 41.0 mi^2 intervening area is simulated by adjusting the daily discharges at station 01199050 by a ratio of 1.39. This adjusted flow is then added to the daily discharges routed from station 01199000 to the confluence using a C_0 = 4.2 ft/s and K_0 = 2,800 ft^2/s. Model parameters were tested over the range of flows and are shown in table 7. Flows simulated at the confluence of Housatonic River and Salmon Creek were then routed 26.4 mi downstream to station 01200500 which is about 1,000 ft downstream from another CL&P hydroelectric plant. Tenmile River is tributary to Housatonic River 500 ft upstream from station 01200500 with a drainage area of 210 mi^2 . The intervening ungaged area of 115 mi^2 is adjusted by a factor of 1.565 times the daily discharges for station 01200000 and added to the routed flows at station 01200500 using values of C_0 = 4.2 ft/s and K_0 = 2,700 ft^2/s for the model parameters. Observed and simulated flows at station 01200500 match closely (fig. 7). Flow-routing model errors are summarized in table 8.

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

Station number	Station name and location	Drainage area (square miles)	Period of record used in study
01199000	Housatonic River at Falls Village	634	October 1961 - September 1982
01199050	Salmon Creek at Lime Rock	29.4	October 1961 - September 1982
01200000	Tenmile River near Gaylordsville	203	October 1961 - September 1982
01200500	Housatonic River at Gaylordsville	996	October 1961 - September 1982

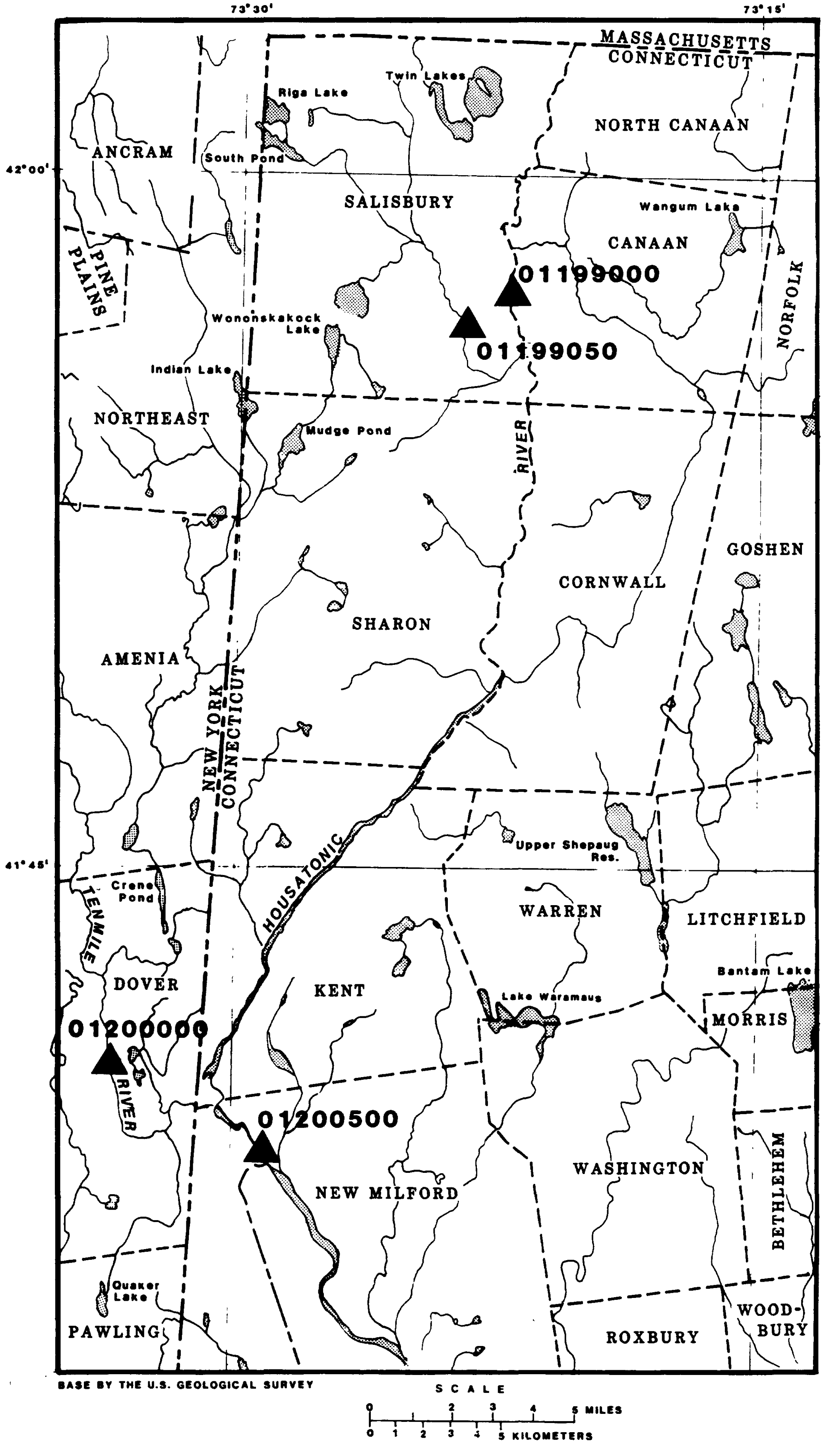


Figure 6.--The Housatonic River basin study area.

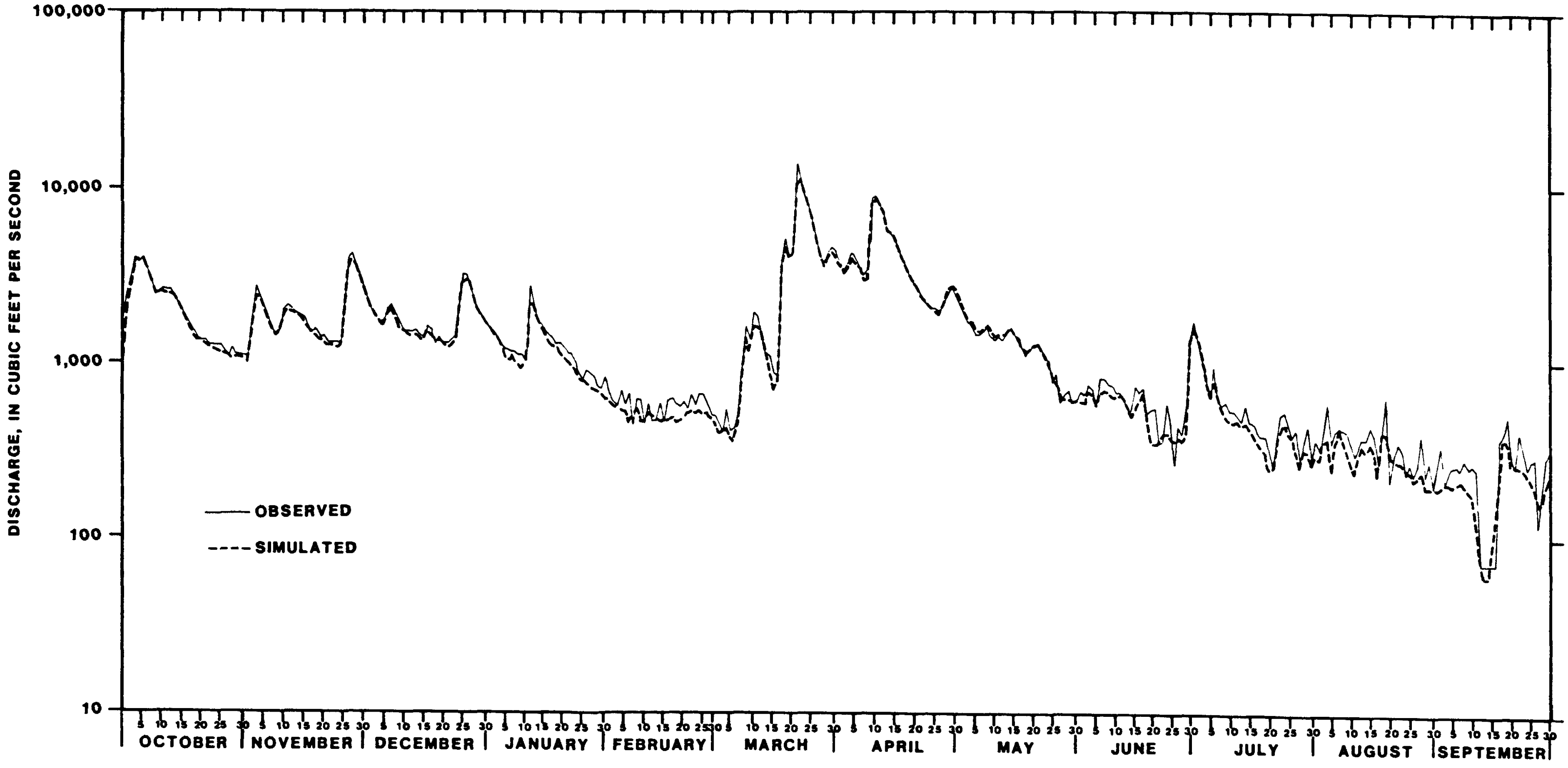


Figure 7.--Daily hydrograph of Housatonic River near Gaylordsville for 1980 water year

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

Station number	Type of flow	Discharge, Q_o (cubic feet per second)	Average width, W_o (feet)	Slope, S_o (feet/foot)	dQ_o/dY_o (feet squared per second)	C_o (feet per second)	K_o (feet squared per second)
01199000	Mean	1,086	160	1.21×10^{-3}	850	5.31	2,800
01199000	7-day, 10-year low flow	119	90	1.21×10^{-3}	270	3.00	546
01199000	Peak flow 10-year recurrence interval	12,000	300	1.21×10^{-3}	1,400	4.67	16,530
01199050	---------- Confluence of Salmon Creek and Housatonic River ----------						
01200500	Mean	1,684	225	1.42×10^{-3}	1,100	4.89	2,635
01200500	7-day, 10-year low flow	180	190	1.42×10^{-3}	250	1.32	334
01200500	Peak flow 10-year recurrence interval	22,200	500	1.42×10^{-3}	4,000	8.00	15,630

Table 8.--Results of routing model for Housatonic River for 1980 water year

1980 WATER YEAR SUMMARY

Mean absolute error (%) for 366 days = 10.41
Mean - error (%) for 300 days = -11.16
Mean + error (%) for 66 days = 7.00
Q1 (simulated) volume (SFD) = 532829
Q2 (observed) volume (SFD) = 566305
Volume error (%) = -5.91
RMS error (%) = 14.95

37 percent of total observations had errors <= 5 percent
62 percent of total observations had errors <= 10 percent
76 percent of total observations had errors <= 15 percent
84 percent of total observations had errors <= 20 percent
90 percent of total observations had errors <= 25 percent
10 percent of total observations had errors > 25 percent

Final model parameters: C_0 = 4.2 ft/s, K_0 = 2,800 ft^2/s

Although the volume and mean errors are less than -5.9 percent and 10.4 percent, respectively, the root mean square error is 15.0 percent and only 62 percent of all simulated values were within 10 percent of the observed flow.

The errors are weighted more to the negative than positive and much of this is a result of the regulation of flow at the CL&P dam just upstream from the Gaylordsville gage for hydroelectric power. Regulation occurs primarily during periods when flows are less than 1,000 ft^3/s. Multiple linearizations did not significantly reduce errors.

Description of Regression Analysis

Simple- and multiple-regression techniques also can 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, as is 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 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 Connecticut:

$$y_i = B_0 + \sum_{j=1}^{p} B_j x_j + e_i \qquad (3)$$

where

- y_i = daily mean discharge at station i (dependent variable),
- x_j = daily mean discharges at nearby stations (explanatory variables),
- B_0 and B_j = regression constant and coefficients, and
- e_i = the random error term.

The above equation is calibrated (B_0 and B_j are estimated) with 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_0 and B_j) are tested to determine if 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 on 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 i.

The equation should be verified by plotting the (1) residuals e_i (difference between simulated and observed discharges) against the dependent and all explanatory variables in the equation, and (2) simulated and observed discharges versus time. These tests are intended to identify if (1) the linear model is appropriate or whether some transformation of the variables is needed, and (2) 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 linear and log-linear models were appropriate. The application of both techniques to three watersheds in Connecticut 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.

Regression Analysis Results

Linear regression techniques were applied to three selected sites. The streamflow record for each station considered for simulation (the dependent variable) was regressed against streamflow records at stations upstream within the same basin (independent variables) during the period of concurrent record (calibration period). Linear regression models were developed to provide simulated daily streamflow records that were compared to observed streamflow records.

The percentage difference between simulated and observed daily discharge values are shown in table 9.

Regression equations were developed relating daily mean flow at Farmington River at Rainbow, station 01190000, 590 mi^2 to Farmington River at Tariffville, station 01189995, 577 mi^2, (table 9). Flow at Farmington River at Rainbow is affected by changes in reservoir storage brought about by powerplant operations 0.4 mi upstream of station 01190000. The reservoir has a total capacity of 244 million cubic feet, a usable capacity of 231 million cubic feet, and an average depth of 19 feet. The 13 mi^2 intervening drainage area drains into the reservoir.

During periods of high daily mean flow, such as April, simulated flows compare closely with observed flows (fig. 8) because inflow is not greatly affected by minor changes in reservoir storage, brought about by power plant operations. During periods of low daily mean flow, such as August (fig. 9), observed flows are generally greater than the simulated flows owing to increased flows associated with power generation.

Shetucket River is formed by the confluence of Willimantic River and Natchaug River in the New England Uplands (fig. 4). Gaging stations are located on Shetucket River near Willimantic, station 01122500 (404 mi^2), Willimantic River near South Coventry, station 01119500 (121 mi^2), and Natchaug River at Willimantic, station 01122000 (174 mi^2). Regression equations were developed relating daily mean flow in Shetucket River to flow at the two upstream stations on Willimantic River and Natchaug River (table 9). Flows simulated by the linear model were higher than observed flows, particularly during high flow events (figs. 10 and 11). The large model errors may be related to the high percentage of ungaged intervening drainage area and affects of regulation within the City of Willimantic.

Regression equations (table 9) were developed relating daily mean flow at Housatonic River near Gaylordsville, station 01200500 (993 mi^2) to the following stations: Housatonic River at Falls Village, station 01199000 (634 mi^2), Tenmile River near Gaylordsville, station 01200000 (203 mi^2), and Salmon Creek at Lime Rock, station 01199050 (29.4 mi^2), in the New England Uplands (fig. 6).

The 130 mi^2 ungaged area is similar in size and geology to that of the Shetucket River basin. A hydroelectric power station and reservoir is 0.4 mile upstream of station 01200500. Regression model errors (table 9) were similar to the Shetucket River model. Both the Shetucket and Housatonic models indicated that only 65 percent of the simulated values were within 10 percent of the observed using regression techniques. Flow routing is an alternate method that might reduce simulation errors. Typical calibration results for August and November of 1975 are presented in figures 12 and 13.

Conclusions Pertaining to Alternative Methods of Data Generation

The simulated discharges from the flow-routing method used for Housatonic and Shetucket stations and the regression method used for Farmington, Housatonic and Shetucket stations were not sufficiently accurate to substitute these methods for the operation of a continuous-flow stream gage. These stations should remain in operation and are included in the next step of this analysis.

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

Station	Model type	Model parameter	Percentage of simulated flow within 5 percent of observed	Percentage of simulated flow within 10 percent of observed	Calibration period (water years)
01122500 Shetucket River at Willimantic	log	$Q_{01122500} = 3.16\ (Q_{01119500})^{.436} \times (Q_{01122000})^{.549}$	39	65	1953-82
01190000 Farmington River at Rainbow	log	$Q_{01190000} = 1.59\ (Q_{01189995})^{.934}$	29	52	1971-82
01200500 Housatonic River near Gaylordsville	log	$Q_{01200500} = 3.72\ (Q_{01199000})^{.62} \times (Q_{01199050})^{.05} \times (Q_{01200000})^{.28}$	43	67	1962-82

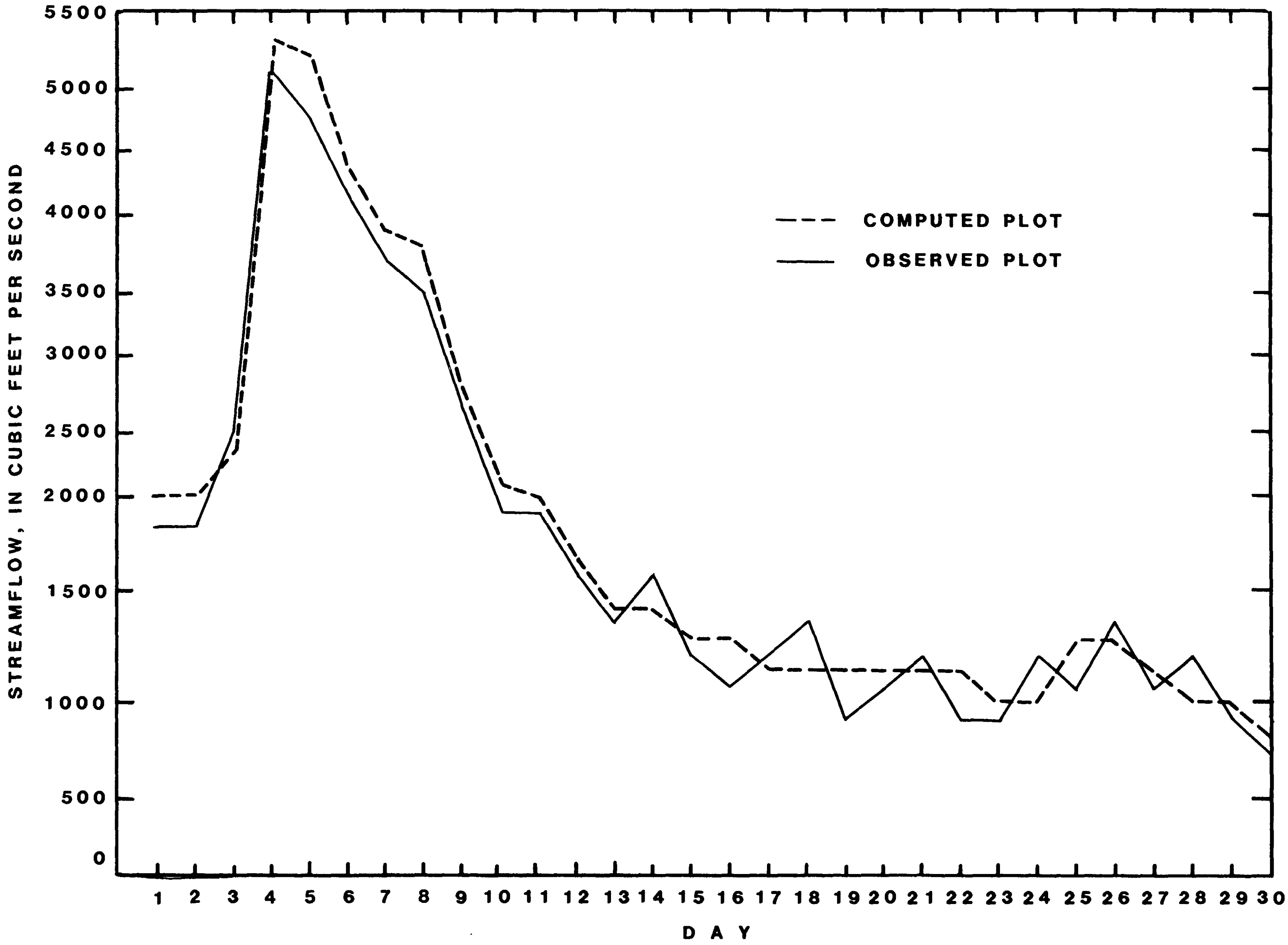


Figure 8.--Daily hydrograph of the Farmington River at Rainbow for April 1975.

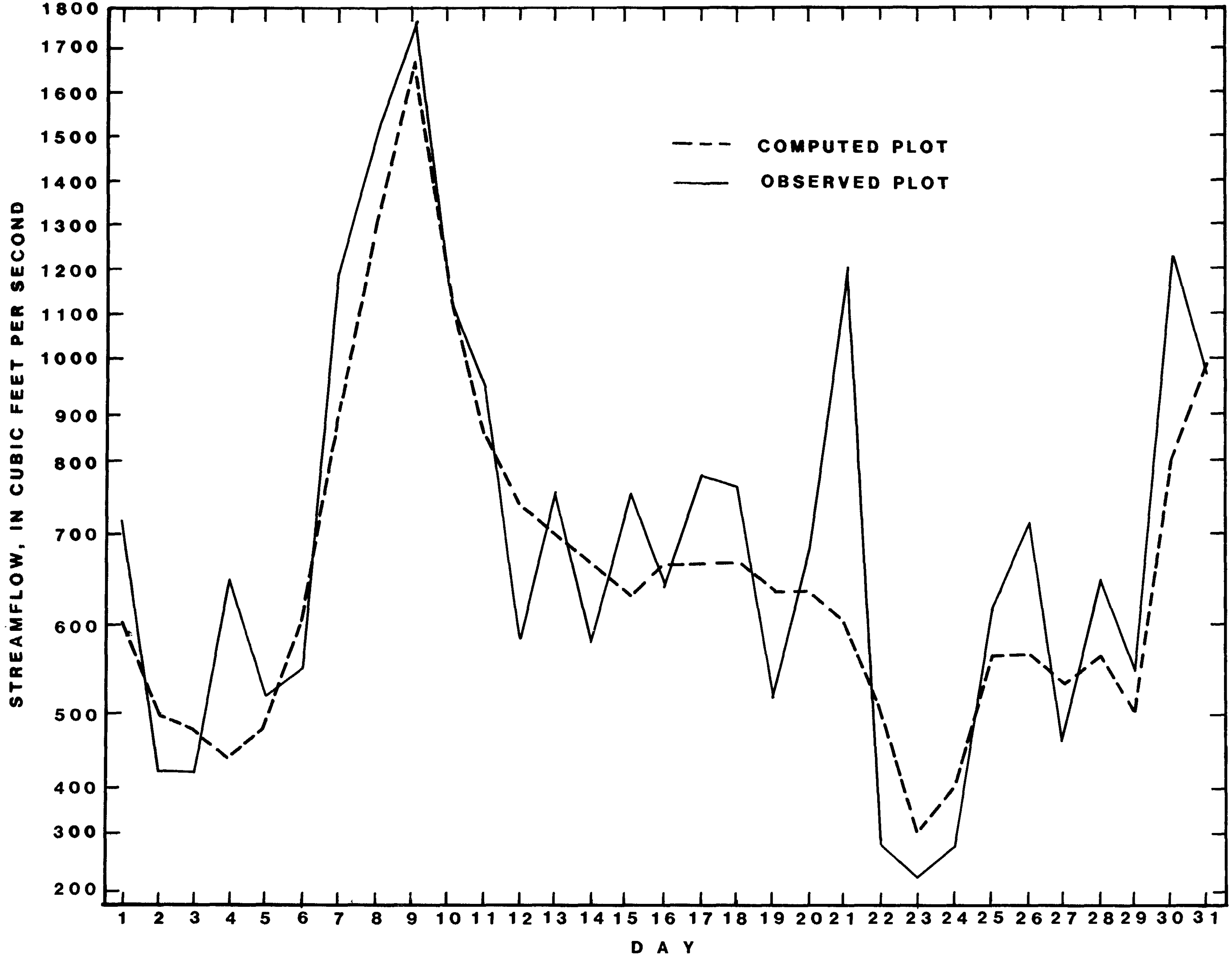


Figure 9.--Daily hydrograph of the Farmington River at Rainbow for August 1975.

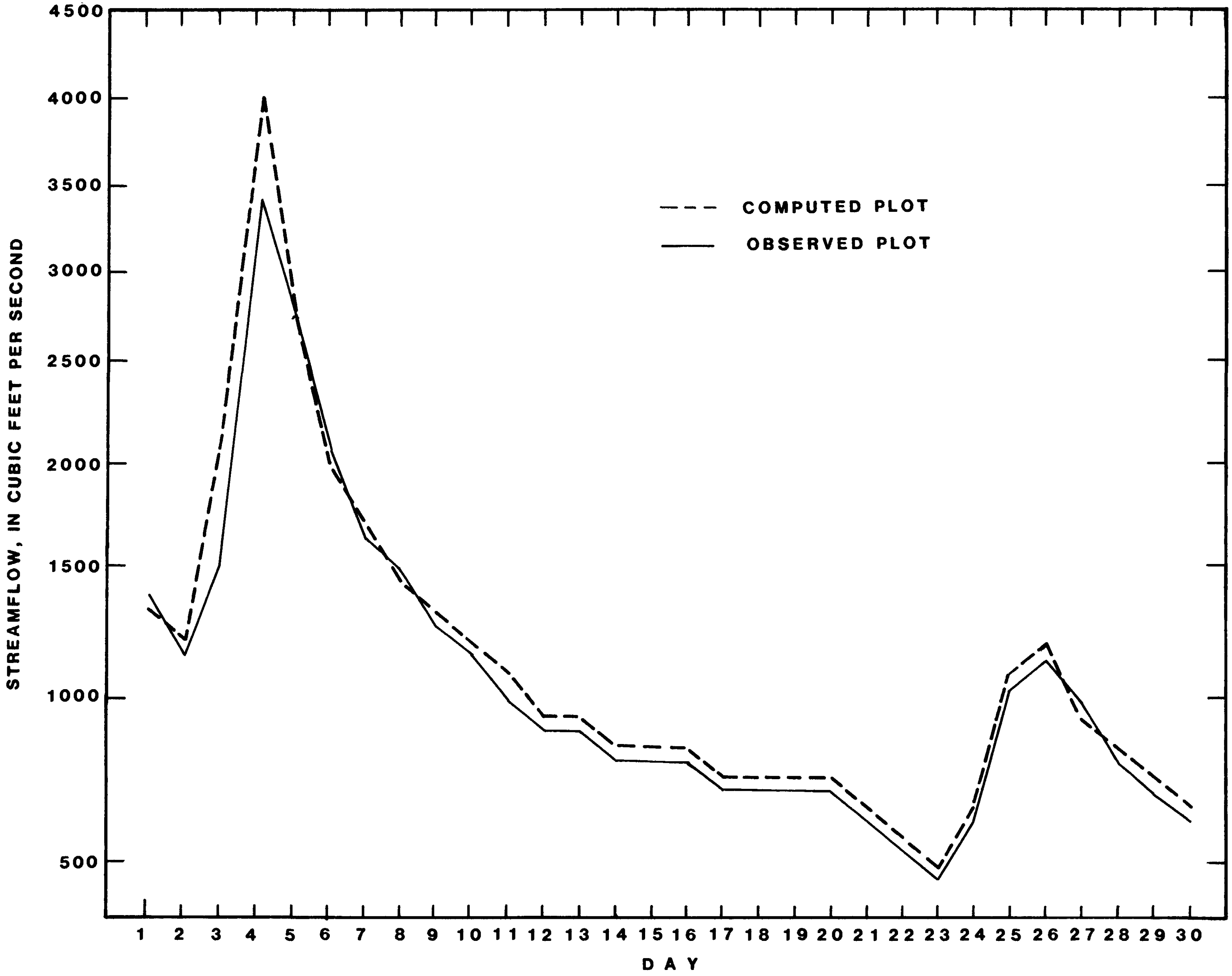


Figure 10.--Daily hydrograph of the Shetucket River near Willimantic for April 1975.

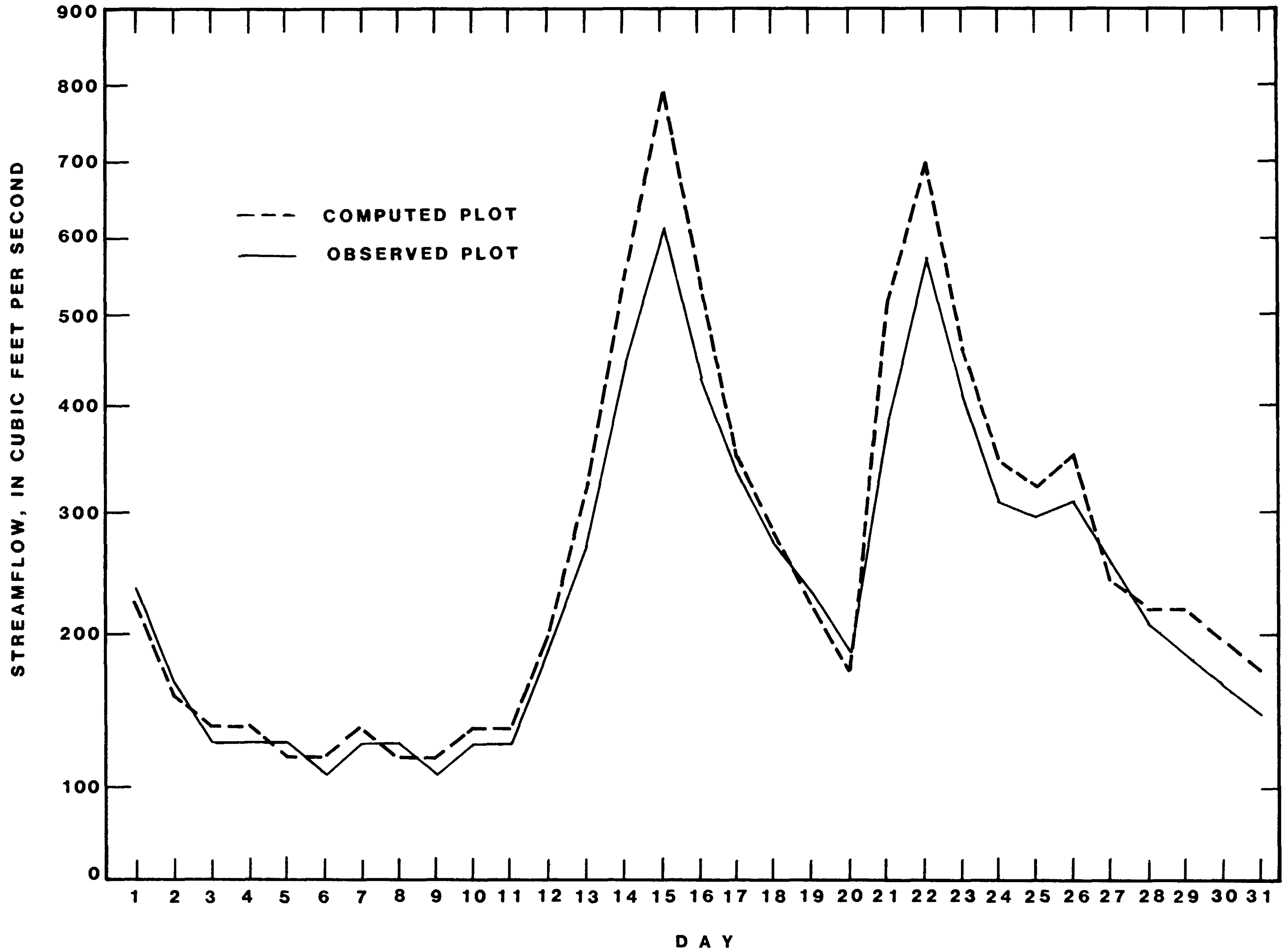


Figure 11.--Daily hydrograph of the Shetucket River near Willimantic for July 1975.

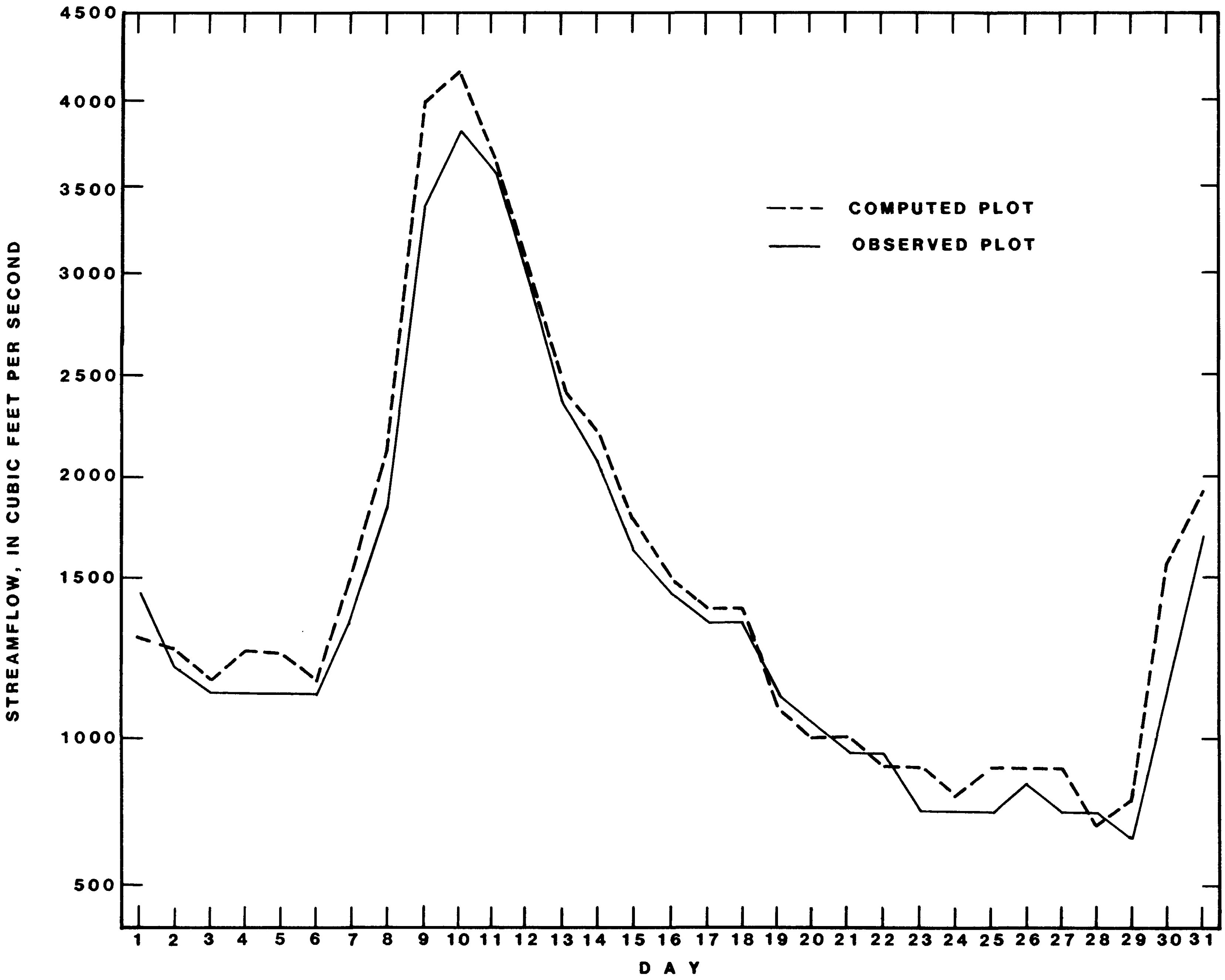


Figure 12.--Daily hydrograph of the Housatonic River near Gaylordsville for August 1975.

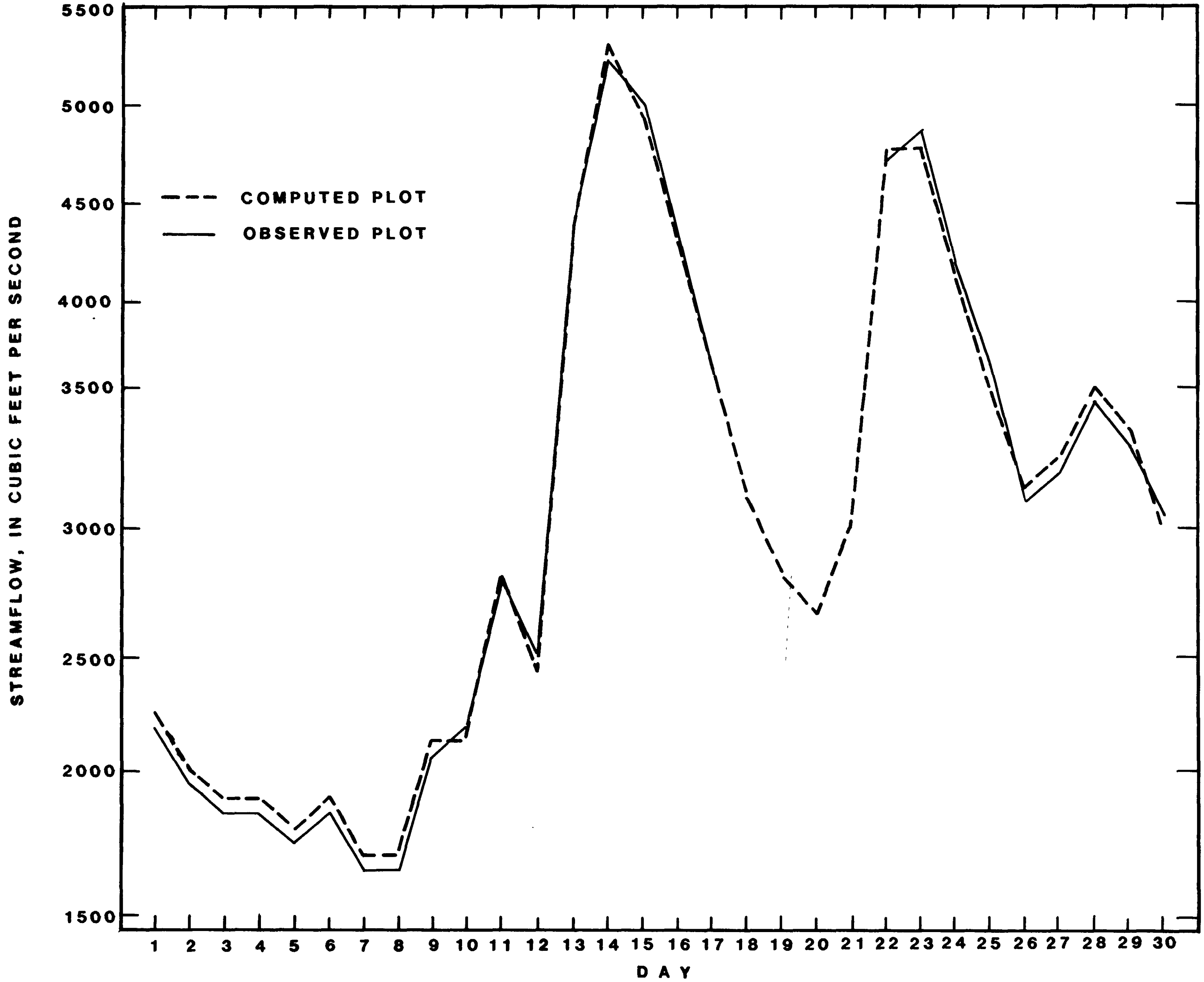


Figure 13.--Daily hydrograph of the Housatonic River near Gaylordsville for November 1975.

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 (Kalman Filtering for Cost-Effective Resource Allocation) 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. Although 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 the measure of 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 (Fontaine and others, 1984) 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

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 commonly will include the path to an individual stream gage with that gage as the lone 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 water-quality sampling. 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 βi= 1, 2, ..., NR, where NR is the number of practical routes, is used during a year such that (1) the budget for the network is not exceeded, (2) the minimum number of visits to each station is made, and (3) the total uncertainty in the network is minimized. Figure 14 represents this step in the form of a mathematical program. Figure 15 presents a tabular layout of the problem. Each of the NR routes is represented by a row in the table, and each of the stations is represented by a column. The zero-one matrix, (W_{ij}), defines the routes in terms of the stations that comprise 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 βi and Ni for i = 1, 2, ..., NR is the total travel cost associated with the set of decisions $\underline{N} = (N_1, N_2, \ldots, N_{NR})$.

The unit-visit cost, α_j, is comprised of the average cost of making a discharge measurement. The set of minimum visit constraints is denoted by the row λ_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 W_{ij} and N_i for all i and must equal or exceed λ_j for all j if $\underline{N}$ is to be a feasible solution to the decision problem.

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 $\underline{N}$ obtained with this technique specifies 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.

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

$V \equiv$ 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 14.--Mathematical-programing form of the optimization of the routing of hydrographers.

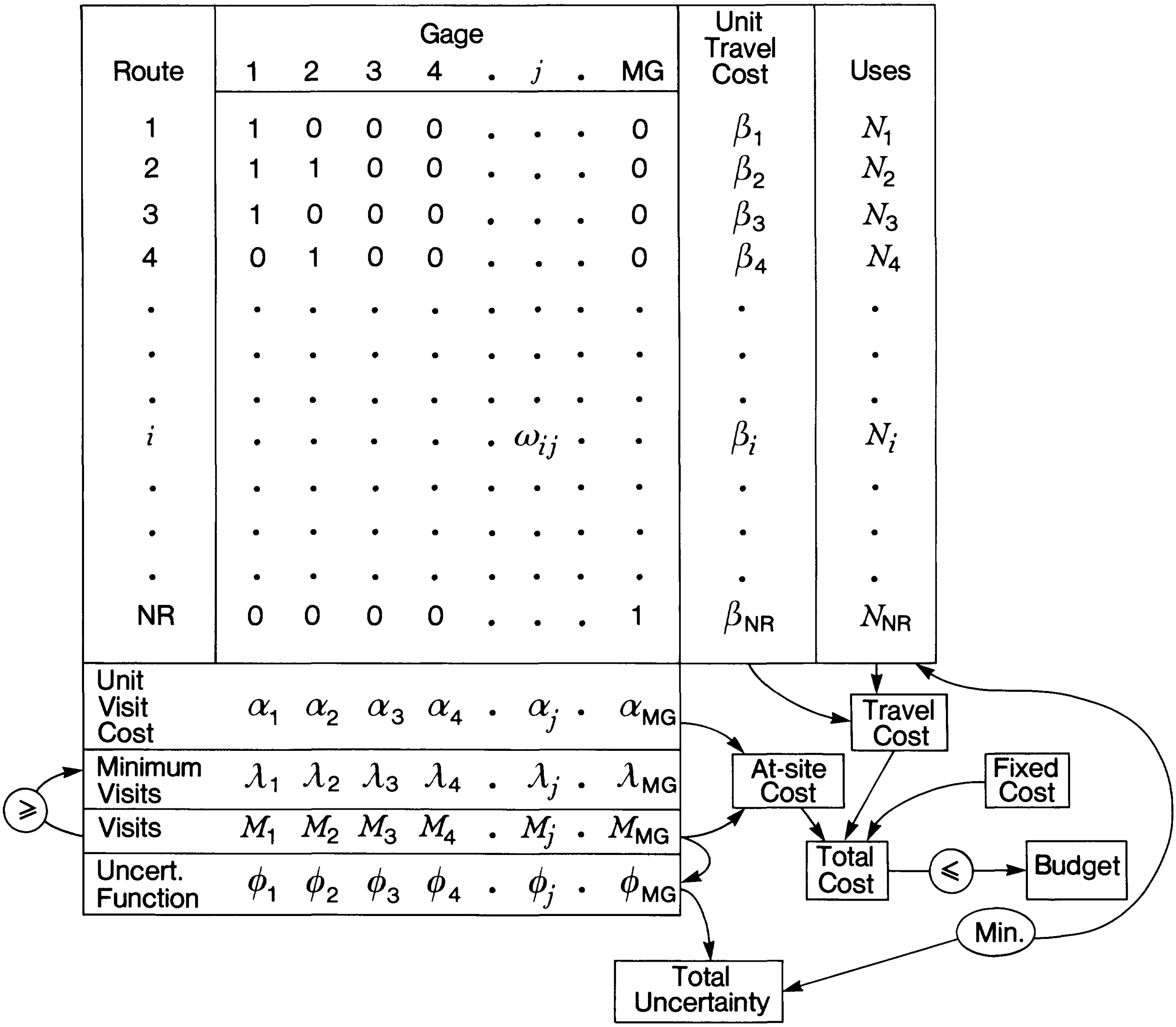


Figure 15.--Tabular form of the optimization of the routing of hydrographers.

Description of Uncertainty Functions

As noted earlier, uncertainty in streamflow records is measured in this study as the average relative variance of estimation of instantaneous discharges. The accuracy of a streamflow estimate depends on how that estimate was obtained. Three situations are considered in this study: (1) Streamflow is estimated from measured discharge and correlative data using a stage-discharge relation (rating curve), (2) the streamflow record is reconstructed using secondary data at nearby stations because primary correlative data are missing, and (3) primary and secondary data are unavailable for estimating streamflow. The variances of the errors of the estimates of flow that would be employed in each situation were weighted by the fraction of time each situation is expected to occur. Thus, the average relative variance would be

$$\bar{V} = \varepsilon_f V_f + \varepsilon_r V_r + \varepsilon_e V_e$$

with (4)

$$1 = \varepsilon_f + \varepsilon_r + \varepsilon_e$$

where

- $\bar{V}$ is the average relative variance of the errors of streamflow estimates,
- ε_f is the fraction of time that the primary recorders are functioning,
- V_f is the relative variance of the errors of flow estimates from primary recorders,
- ε_r is the fraction of time that secondary data are available to reconstruct streamflow records given that the primary data are missing.
- V_r is the relative variance of the errors of estimation of flows reconstructed from secondary data,
- ε_e is the fraction of time that primary and secondary data are not available to compute streamflow records, and
- V_e is the relative error variance of the third situation.

The fractions of time that each source of error is relevant are functions of the frequencies at which the recording equipment is serviced.

The time τ since the last service visit until failure of the recorder or recorders at the primary site is assumed to have a negative-exponential probability distribution truncated at the next service time; the distribution's probability density function is

$$f(\tau) = ke^{-kt}/(1-e^{-ks}) \quad (5)$$

where

- k is the failure rate in units of (day) $^{-1}$,
- e is the base of natural logarithms, and
- s is the interval between visits to the site in days.

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

$$\varepsilon_f = (1-e^{-ks})/(ks) \quad (6)$$

(Fontaine and others, 1984, eq. 21).

The fraction of time ε_e that no records exist at either the primary or secondary sites can also be derived assuming that the times between failures at both sites are indepenent and have negative exponential distributions with the same rate constant. It then follows that

$$\varepsilon_e = 1 - [2(1-e^{-ks}) + 0.5(1-e^{-2ks})]/(ks)$$

(Fontaine and others, 1984, eqs. 23 and 24).

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

$$\begin{aligned}\varepsilon_r &= 1 - \varepsilon_f - \varepsilon_e \\ &= [(1-e^{-ks}) + 0.5(1-e^{-2ks})]/(ks) \qquad (7)\end{aligned}$$

The relative 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 logarithms of 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 at the gaging station. The measured discharge is the discharge determined by field observations of depths, widths, and velocities. If $q_T(t)$ is the true instantaneous discharge at time t and $q_R(t)$ is the value that would be estimated using the rating curve, then:

$$x(t) = \ln q_T(t) - \ln q_R(t) = \ln [q_T(t)/q_R(t)] \qquad (8)$$

is the instantaneous difference between the logarithms of the true discharge and the rating-curve discharge.

In computing estimates of streamflow, the rating curve may be continually adjusted on the basis of periodic measurements of discharge. This adjustment process results in an estimate, $q_c(t)$, that is a better estimate of the stream's discharge at time t. The difference between the variable $\hat{x}(t)$, which is defined as:

$$\hat{x}(t) = \ln q_c(t) - \ln q_R(t) \qquad (9)$$

and x(t) is the error in the streamflow record at time t. The variance of this difference over time is the desired estimate of V_f.

Unfortunately, the true instantaneous discharge, $q_T(t)$, cannot be determined, thus, x(t) and the difference, $x(t) - \hat{x}(t)$, cannot be determined as well. However, the statistical properties of $x(t) - \hat{x}(t)$, particularly its variance, can be inferred from available discharge measurements. Let the observed residuals of measured discharge from the rating curve be z(t) so that

$$z(t) = x(t) + v(t) = \ln q_m(t) - \ln q_R(t) \qquad (10)$$

where

v(t) is the measurement error, and

$\ln q_m(t)$ is the logarithm of the measured discharge equal to $\ln q_R(t)$ plus v(t).

In the Kalman-filter analysis, the z(t) time series was analyzed to determine three site-specific parameters. The Kalman filter used in this study assumes that the time residuals x(t) arise from a continuous first-order Markovian process that has a Gaussian (normal) probability distribution with zero mean and variance (subsequently referred to as process variance) equal to p. A second important parameter is β, the reciprocal of the correlation time of the Markovian process giving rise to x(t); the correlation between $x(t_1)$ and $x(t_2)$ is $\exp[-\beta|t_1-t_2|]$. Fontaine and others (1984) also define q, the constant value of the spectral density function of the white noise that drives the Gauss-Markov x-process. The parameters, p, q, and β are related by

$$Var[x(t)] = p = q/(2\beta) \quad (11)$$

The variance of the observed residuals z(t) is

$$Var[z(t)] = p + r \quad (12)$$

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

If the recorder at the primary site fails and there are no concurrent data at other sites that can be used to reconstruct the missing record at the primary site, there are 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 relative error variance during periods of no concurrent data at nearby stations. If the expected value is used to estimate discharge, the value that is used should be the excepted value of discharge at the time of year of the missing record because of the seasonality of streamflow. The variance of streamflow, which is also a seasonally variable parameter, is an estimate of the error variance that results from using the expected value as an estimate. Thus the coefficient of variation squared $(C_v)^2$ is an estimate of the required relative error variance V_e. Because C_v varies seasonally and the times of failures cannot be anticipated, a seasonally averaged value of C_v is used:

$$\bar{C}_v = \left(\frac{1}{365}\sum_{i=1}^{365} \left(\frac{\sigma_i}{\mu_i}\right)^2\right)^{1/2} \quad (13)$$

where

σ_i is the standard deviation of daily discharges for the i^{th} day of the year,

μ_i is the expected value of discharge on the i^{th} day of the year, and $(\bar{C}_v)^2$ is used as an estimate of V_e.

The Application of K-CERA in Connecticut

As a result of the first two parts of this analysis, it has been determined that 59 of the currently existing gaging stations in the State of Connecticut and three in Massachusetts be continued in operation. Forty one of the gaging stations were subjected to the K-CERA analysis with results described below. The remaining 21 stations were not included in the analysis because these stations had less than five years of data; had less than 20 measurements; or they were tidal sites. The cost to operate these sites will be taken into account in subsequent sections of the report.

Definition of Missing Record Probabilities

As was 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 ε_f as given in equation 4, the average time to failure is 1/k. The value of 1/k will differ from site to site depending upon the type of equipment a 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. A period of actual data collection of 5 years duration in which little change in technology occurred and in which stream gages were visited on a fairly consistent pattern of a 5-week frequency, was used to estimate 1/k in Connecticut. During this 5-year period a gage could be expected to malfunction an average of 5.5 percent of the time. There was no reason to distinguish between gages on the basis of their exposure or equipment, so the 5.5 percent lost record and a six-week visit frequency were used to determine a value for 1/k, which was used to determine ε_f, ε_e, and ε_r for each of the 41 stream gages as a function of the individual frequencies of visit.

The variance V_r of the relative 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 nearby sites. The correlation coefficient ρ_c between the streamflows with seasonal trends removed at the site of interest and at the other sites is a measure 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 relative error variance of flow estimates at the primary site obtained from secondary information will be

$$V_r = (1-\rho_c^2)\ \bar{C}_v^2 \qquad (14)$$

Because errors in streamflow estimates arise from three different sources with widely varying precisions, the resultant distribution of those errors may differ significantly from a normal or log-normal distribution. This lack of normality causes difficulty in interpretation of the resulting average estimation variance. When primary and secondary data are unavailable, the relative error variance V_e may be very large. This could yield correspondingly large values of $\bar{V}$ in equation (4) even if the probability that primary and secondary information are not available, ε_e, is quite small.

A new parameter, the equivalent Gaussian spread (EGS), is introduced here to assist in interpreting the results of the analyses. If it is assumed that the various errors arising from the three situations represented in equation (4) are log-normally distributed, the value of EGS is determined by the probability statement that

$$\text{Probability } [e^{-EGS} \leq (q_c(t)\ /\ q_T(t) \leq e^{+EGS}] = 0.683 \qquad (15)$$

Thus, if the residuals $\ln q_c(t) - \ln q_T(t)$ were normally distributed, $(EGS)^2$ would be their variance. Here EGS is reported in percent because EGS is defined so that nearly two-thirds of the errors in instantaneous streamflow data will be within plus or minus EGS percent of the reported values.

Definition of Cross-Correlation Coefficient and Coefficient of Variation

The values of V_e and V_r of the needed uncertainty functions were computed using daily streamflow records for each of the 41 stations for the 30-year period (1953-82, water years) for which daily streamflow values are stored in WATSTORE (Hutchinson, 1975) were retrieved. For each of the stream gages that had 5 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. The set of parameters that gave the highest (Pc) and second highest (R2) cross correlation coefficients are listed in table 10.

Kalman-Filter Definition of Variance

The determination of the variance V_f for each of the 41 gaging stations 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 measurements.

Long-term ratings for the Connecticut gaging stations were determined by applying a non-linear statistical fitting routine (Helwig and Council, 1979) to discharge measurements and correlative data. The correlative data for a discharge rating function is the gage height. The rating function that was fit to this data was of the general form:

$$LQM = B_1 + B_3 * \ln (GHT-B_2) \qquad (16)$$

where

- LQM is the logarithmic (base e) value of the measured discharge,
- GHT is the recorded gage height corresponding to the measured discharge,
- B_1 is the logarithm (base e) of discharge for an effective flow depth of 1 foot,
- B_2 is the gage height of zero flow,
- B_3 is the slope of the rating curve, and
- ln is the natural logarithm function.

The fitting routine computed the values for B1, B2, and B3 that best fit the given data sets. The best-fit rating function was then used to compute the rated discharge for the given gage heights. Residuals were computed as the rated discharge minus the measured discharge. The residuals divided by the rated discharge gives the percent error.

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 Connecticut program, all open water measurments were assumed to have a measurement error of 2 percent.

Table 10.--Statistics of record reconstruction

Station no.	C_v in percent	e_c	R2	Station or source of reconstructed records
01118300	116	0.825	--	01127500
	--	--	0.718	01119500
01119500	109	.927	--	01121000, 01127500
	--	--	.916	01121000
01121000	136	.916	--	01119500, 01122000
	--	--	.729	01122000
01122000	115	.729	--	01121000, 01122500
	--	--	.927	01122500
01122500	110	.966	--	01121000, 01122000
	--	--	.927	01122000
01123000	112	.891	--	01122000, 01122500
	--	--	.874	01122500
01124000	115	.908	--	01121000, 01124151
	--	--	.846	01124151
01124151	89.2	.853	--	01124000, 01127000
	--	--	.846	01124000
01127000	97.4	.920	--	01124000
	--	--	.821	01124151
01127500	136	.927	--	01122500, 01123000
	--	--	.920	01123000
01184000	76.6	.890*	--	01190070**, 01193000**
01186000	91.9	.569	--	01186500, 01188090
	--	--	.561	01188090
01186500	133	.824	--	01186000, 01188000
	--	--	.813	01188000
01187300	169	.912	--	01186500, 01188000
	--	--	.899	01186500
01188000	137	.834	--	01186500, 01187300
	--	--	.819	01187300
01188090	70.3	.852	--	01186000, 01189995
	--	--	.838	01189995
01189000	116	.883	--	01188000
	--	--	.860	01188000, 01188090
01189995	72.5	.917	--	01188090, 01190000
	--	--	.878	01190000
01190000	100	.878	--	01189995
	--	--	.867	01188090, 01189995
01191000	174	.742	--	01119500, 01192500
	--	--	.737	01119500
01192500	79.8	.862	--	01119500, 01191000
	--	--	.852	01119500
01193500	124	.917	--	01121000, 01204000
	--	--	.894	01121000
01196500	94.3	.891	--	01193500, 01196620
	--	--	.877	01193500

Table 10.--Statistics of record reconstruction-continued

Station no.	C_v in percent	ρ_c	R2	Station or Source of reconstructed records
01196620	87.8	0.762	--	01196500, 01209700
	--	--	0.750	01196500
01199000	94.3	.977	--	01200500
	--	--	.864	01205500
01199050	102	.885	--	01200000, 01205600
	--	--	.861	01200000
01200000	125	.928	--	01199050, 01200500
	--	--	.921	01200500
01200500	97.0	.982	--	01199000, 01205500
	--	--	.977	01199000
01204000	136	.903	--	01193500, 01200000
	--	--	.842	01193500
01205500	113	.896	--	01200500
	--	--	.864	01199000
01205600	139	.884	--	01205700, 01206900
	--	--	.830	01205700
01205700	126	.830	--	01205600
	--	--	.721	01206900
01206900	129	.870	--	01205700, 01208500
	--	--	.866	01208500
01208013	104	.753	--	01206900, 01208420
	--	--	.724	01206900
01208420	120	.822	--	01208013, 01208500
	--	--	.794	01208500
01208500	115	.886	--	01205700, 01206900
	--	--	.866	01206900
01208873	101	.723	--	01208925, 01208950
	--	--	.702	01208950
01208925	110	.813	--	01208950
	--	--	.666	01208873
01208950	139	.901	--	01208925, 01209700
	--	--	.877	01209700
01208990	130	.946	--	01209700
	--	--	.854	01208950
01209700	132	.955	--	01208950, 01208990
	--	--	.946	01208990

* Estimated.
** Estimated discharge based on stages at slope gages.

Table 11.--Summary of the autocovariance analysis

Station no.	Station name	RHO*	Measurement variance (log base e)^2	Progress variance (log base e)^2
01118300	Pendleton Hill Brook near Clarks Falls	0.972	0.0004	0.0049
01119500	Willimantic River near Coventry	.997	.0004	.0091
01121000	Mount Hope River near Warrenville	.974	.0004	.0037
01122000	Natchaug River at Willimantic	.889	.0004	.0016
01122500	Shetucket River near Willimantic	.947	.0004	.0018
01123000	Little River near Hanover	.990	.0004	.1057
01124000	Quinebaug River at Quinebaug	.985	.0004	.0067
01124151	Quinebaug River at West Thompson	.981	.0004	.0056
01127000	Quinebaug River at Jewett City	.983	.0004	.0032
01127500	Yantic River at Yantic	.993	.0004	.0085
01184000	Connecticut River at Thompsonville	.991	.0004	.0795
01186000	West Branch Farmington River at Riverton	.947	.0004	.0009
01186500	Still River at Robertsville	.981	.0004	.0049
01187300	Hubbard River near West Hartland	.827	.0004	.0077
01188000	Burlington Brook near Burlington	.981	.0004	.0036
01188090	Farmington River at Unionville	.991	.0004	.0075
01189000	Pequabuck River at Forestville	.997	.0004	.0488
01189995	Farmington River at Tariffville	.989	.0004	.0019
01190000	Farmington River at Rainbow	.969	.0004	.0528
01191000	North Branch Park River at Hartford	.994	.0004	.0223
01192500	Hockanum River near East Hartford	.990	.0004	.0017
01193500	Salmon River near East Hampton	.992	.0004	.0685

* One-day autocorrelation coefficient.

Table 11.--Summary of the autocovariance analysis-continued

Station no.	Station name	RHO*	Measurement variance (log base e)^2	Progress variance (log base e)^2
01196500	Quinnipiac River at Wallingford	0.989	0.0004	0.0060
01196620	Mill River near Hamden	.980	.0004	.0087
01199000	Housatonic River at Falls Village	.675	.0004	.0012
01199050	Salmon Creek at Lime Rock	.972	.0004	.0053
01200000	Tenmile River near Gaylordsville	.875	.0004	.0018
01200500	Housatonic River at Gaylordsville	.974	.0004	.0047
01204000	Pomperaug River at Southbury	.963	.0004	.0019
01205500	Housatonic River at Stevenson	.944	.0004	.0057
01205600	West Branch Naugatuck River at Torrington	.975	.0004	.0101
01205700	East Branch Naugatuck River at Torrington	.961	.0004	.0163
01206900	Naugatuck River at Thomaston	.958	.0004	.0032
01208013	Branch Brook near Thomaston	.987	.0004	.0303
01208420	Hop Brook near Naugatuck	.953	.0004	.0121
01208500	Naugatuck River at Beacon Falls	.997	.0004	.0529
01208873	Rooster River at Fairfield	.995	.0004	.1717
01208925	Mill River near Fairfield	.990	.0004	.0384
01208950	Sasco Brook near Southport	.956	.0004	.0137
01208990	Saugatuck River near Redding	.973	.0004	.0142
01209700	Norwalk River at South Wilton	.991	.0004	.0625

* One-day autocorrelation coefficient.

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 11 presents a summary of the autocovariance analysis expressed in terms of process variance and 1-day autocorrelation. A typical fit of the covariance function for a selected station in Connecticut is given in figure 16.

The autocovariance parameters, summarized in table 11, 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. Three stations that present typical examples of uncertainty functions are given in figure 17. These functions are based on the assumption that a measurment was made during each visit to the station.

Feasible routes servicing 62 gaging stations, 11 reservoirs, and 17 wells, were determined after consultation with personnel in the Hydrologic Data and Analysis Section and review of the 41 uncertainty functions. Selected 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 50 routes and the stations visited on each are summarized in table 12.

Costs associated with practical routes were determined. Fixed costs to operate a gage typically include equipment rental, batteries, electricity, data processing and storage, computer charges, maintenance and miscellaneous supplies, and analysis and supervisory charges. For Connecticut, 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 differ widely. These costs were determined on a station-by-station basis from past experience.

Measurement costs are those associated with paying the hydrographer for the time actually spent at a station making a discharge measurement. These costs vary from station to station and are a function of the difficulty and time required to make the discharge measurement. Average measurement times were calculated for each station based on an analysis of discharge measurement data avaialable. This time was then multiplied by the average hourly salary of hydrographers in the Connecticut Office to determine total measurement 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 the time actually spent servicing the equipment.

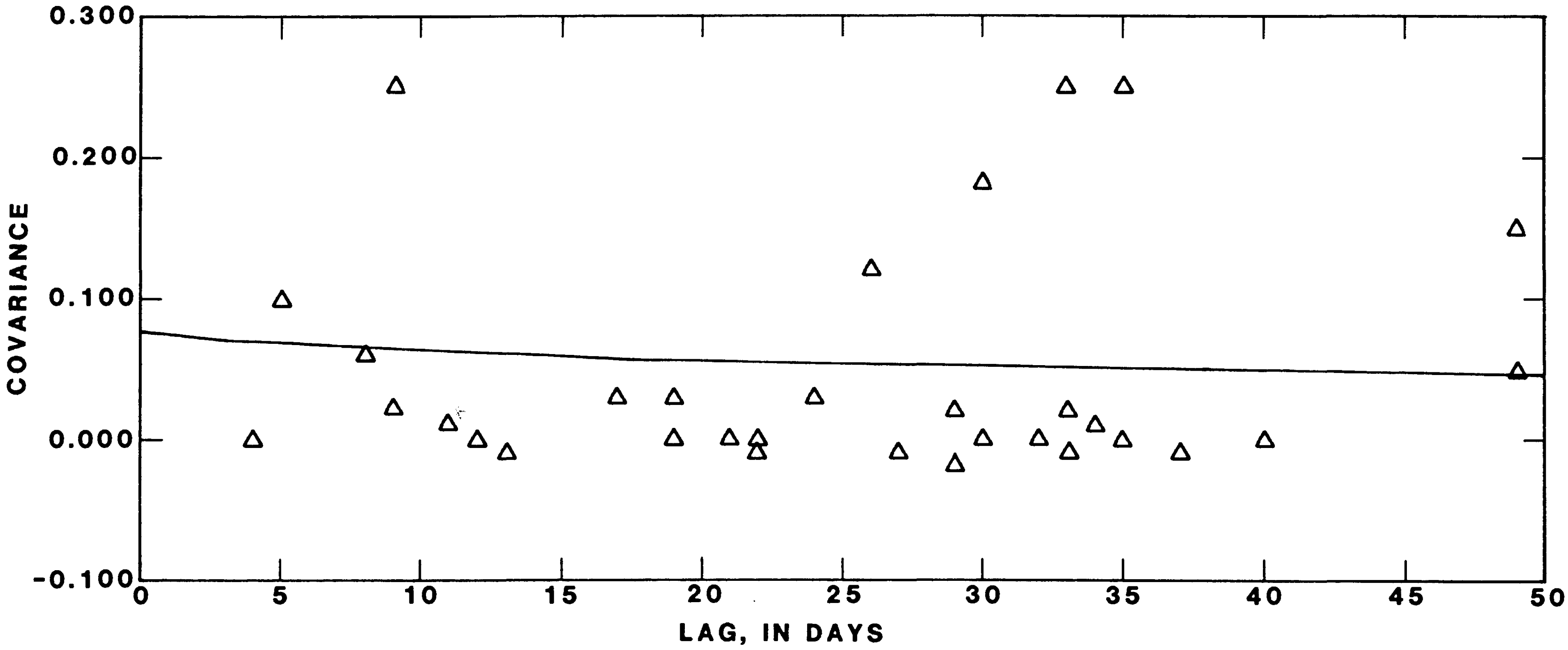


Figure 16.--Autocovariance function for complete year for Salmon River near East Hampton.

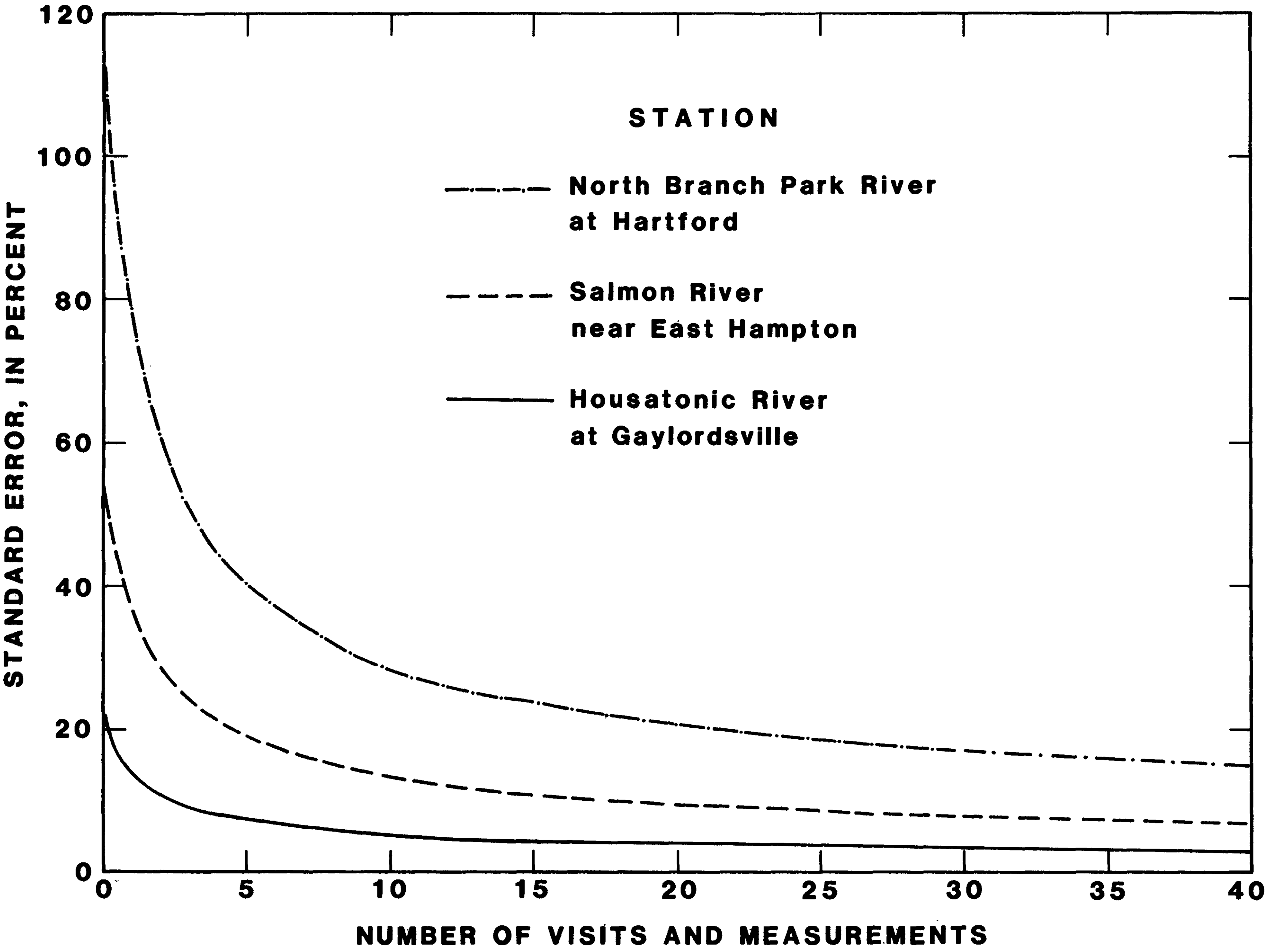


Figure 17.--Typical uncertainty function for instantaneous discharge.

Table 12.--Summary of the routes that may be used to visit stations in Connecticut

Route Number	Stations serviced on the route*									
1	-1	-2	01193500	-3	-4	-5				
2	01127000	-1	-2	01193500	-3	-4	-5			
3	01127000	01123000	01122500	01122000						
4	01118300	-01119040	01127500							
5	-01194825	-01195100	-01195146							
6	-01196651	-01192883								
7	01121000	-6	-7	-8						
8	01121000	-6	-7	01184000	-8					
9	01124000	01124151	-01125500							
10	01189000	01188090	01191000							
11	01189995	-01184100	01190000							
12	01192500	-01184490	01119500							
13	-01190070	-01193000	-01193050							
14	01199050	01199000	-01197500							
15	-01199290	01200500	01200000							
16	-9	-10	-11	-12	-13	01188000				
17	-01185500	01186500								
18	01186000	01187300								
19	01205600	01205700	01206900							
20	01208013	-01208171								
21	01204000	-14	-15	-16	-17	-18				
22	01204000	-14	-15	01205500	-16	-17	-18			
23	-01196600	-01196590	01196620	01196500						
24	-01208325	01208420	01208500							
25	01208873	01208925	01208950							
26	01208990	01209700	-01209788							
27	01205500									
28	-28	-27	-26	-25	-23	-24	-21	-22	-20	-19
29	01127000									
30	01184000									
31	-01172003									
32	01199000									
33	01200500									
34	01118300	01127500								
35	01123000									
36	-01125500									
37	01189995	01190000								
38	01192500	01119500								
39	01189000	01191000	01190000							
40	01189000									
41	01190000									
42	01191000									
43	01199050	01199000								
44	01200500	01200000								
45	01186500									
46	01205600	01205700								
47	01208013									
48	-01196600	-01196590								
49	01208420	01208500								
50	01208990	01209700								

*Negative numbers indicate ground-water and surface-water stations that are visited on a route but were not included in the uncertainty analysis.

K-CERA Results

The Traveling Hydrographer Program uses the uncertainty functions, along with the appropriate cost data and route definitions, to compute the most cost-effective way of operating the stream-gaging program. In this application, the first step is to simulate the current practice and determine the associated total uncertainty. To accomplish this, the number of visits made to each stream gage and the specific routes that are used to make these visits were fixed. In Connecticut, current practice indicates that discharge measurements are made 90 percent of the time that a station is visited. The resulting average error of estimation for the current practice in Connecticut is 14.5 percent (figure 18).

The solid line in figure 18 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 operations, other than budget, were defined as described below.

The minimum number of times each station must be visited was determined by giving consideration 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 Connecticut a minimum requirement of four visits per year was calculated and applied to all stations with 30-minute recorders. At stations with 15-minute recorders the minimum was five.

Minimum visit requirements should also reflect the need to visit stations for special reasons such as water-quality sampling. In Connecticut, 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 18 and table 13 summarize the K-CERA analysis and are predicated on a discharge measurement being made each time that a station is visited.

It should be emphasized that figure 18 and table 13 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.

The current policy of servicing stations at 6-week intervals results in an average standard error of estimate of streamflow of 14.5 percent. This policy requires a budget of $267,000 to operate the 62 stream-gaging stations (59 in Connecticut and 3 in Massachusetts), and the operation of equipment at 11 reservoirs, and 17 ground-water wells in Connecticut. The range in standard errors is from a low of 5.2 percent for station 01199000, Housatonic River at Falls Village, to a high of 32.1 percent at station 01191000, North Branch Park River at Hartford. A 12.7 percent average standard error could be obtained with a $264,000 budget by changing the schedule of field data-collection activities. This policy would result in no increase in standard error at station 01199000, while the standard error at station 01191000 would decrease from 32.1 to 18.9 percent.

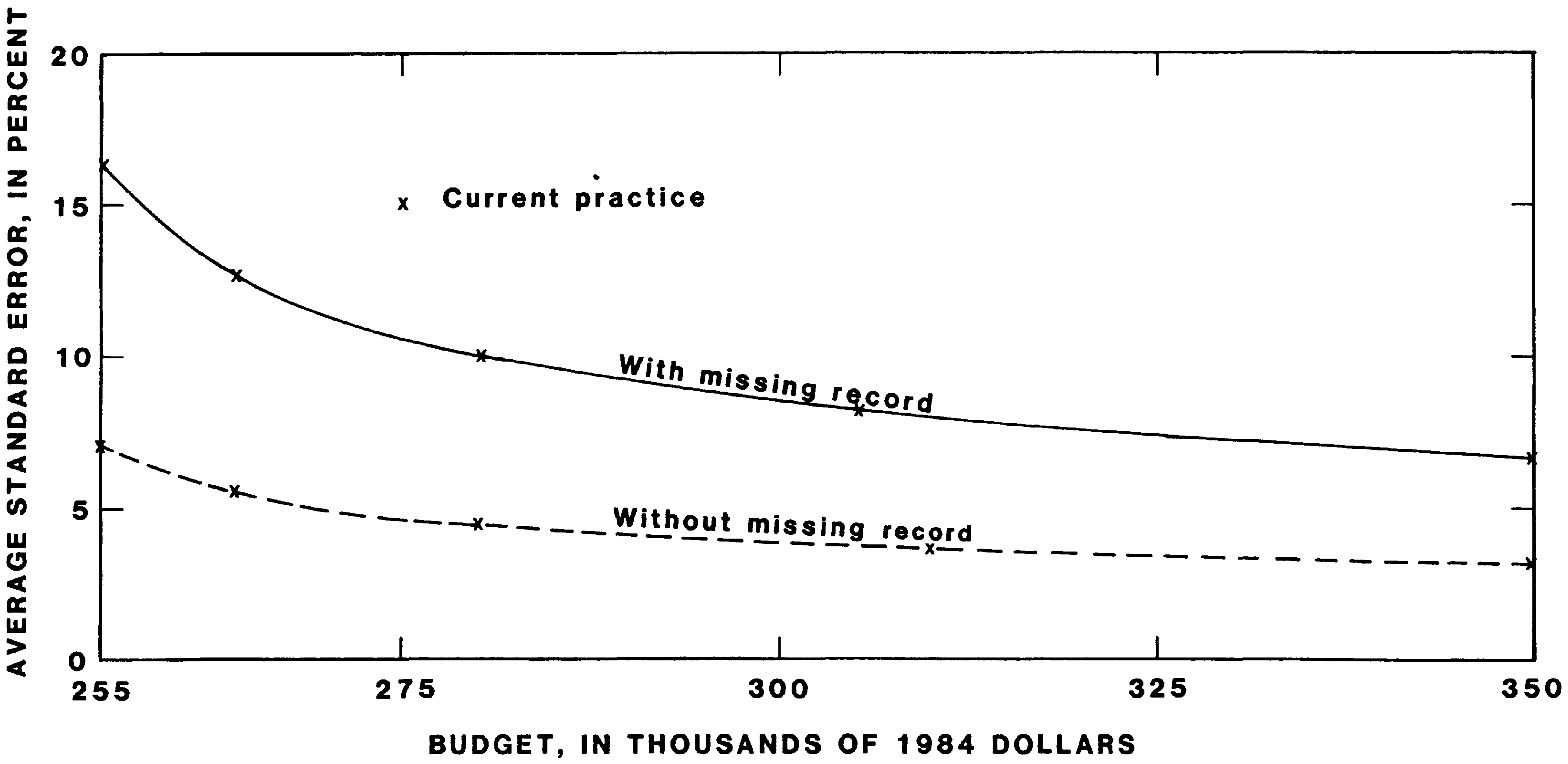


Figure 18.--Temporal average standard error per stream gage.

Table 13.--Selected results of K-CERA analysis

Station	Current, 1984, operation (267)	255	264	280	305	310
	Standard error of instantaneous discharge, in percent [Equivalent Gaussian spread, in percent] (Number of visits)	U.S. Geological Survey stream-gaging budget in Connecticut, in thousands of in 1984 dollars				
01127500	12.3 [3.5] (8)	12.3 [3.5] (8)	10.1 [2.8] (12)	7.9 [2.2] (20)	6.5 [1.8] (29)	6.2 [1.7] (32)
01184000	10.4 [9.1] (16)	18.2 [16.7] (5)	14.7 [13.2] (8)	11.6 [10.2] (13)	9.4 [8.1] (20)	8.9 [7.7] (22)
01186000	17.8 [2.6] (8)	20.4 [2.8] (6)	14.1 [2.2] (13)	10.9 [1.8] (22)	8.7 [1.5] (35)	8.4 [1.4] (37)
01186500	19.4 [4.3] (8)	22.2 [4.9] (6)	16.6 [3.6] (11)	12.7 [2.8] (19)	10.3 [2.2] (29)	10.2 [2.2] (30)
01187300	17.7 [8.8] (8)	19.8 [9.1] (6)	14.6 [8.2] (13)	12.0 [7.5] (22)	10.0 [6.7] (35)	9.8 [6.6] (37)
01188000	17.3 [3.0] (12)	17.3 [3.0] (12)	17.3 [3.0] (12)	14.6 [2.5] (17)	11.5 [2.0] (28)	11.3 [1.9] (29)
01188090	8.1 [3.6] (8)	11.3 [5.2] (4)	8.7 [3.9] (7)	6.2 [2.7] (14)	4.8 [2.1] (24)	4.6 [2.0] (26)
01189000	14.6 [5.4] (8)	18.3 [7.1] (5)	12.5 [4.5] (11)	9.4 [3.3] (20)	7.7 [2.7] (30)	7.2 [2.5] (34)
01189995	6.7 [2.1] (8)	7.1 [2.2] (7)	7.1 [2.2] (7)	6.0 [1.9] (10)	4.6 [1.4] (17)	4.8 [1.5] (16)
01190000	18.5 [16.3] (8)	18.5 [16.3] (8)	15.7 [13.7] (12)	12.4 [10.7] (20)	10.0 [8.6] (31)	9.4 [8.1] (35)

Table 13.--Selected results of K-CERA analysis-Continued

Station	Current, 1984, operation (267)	U.S. Geological Survey stream-gaging budget in Connecticut, in thousands of in 1984 dollars 255	264	280	305	310
	Standard error of instantaneous discharge, in percent [Equivalent Gaussian spread, in percent] (Number of visits)					
01191000	32.1 [5.3] (8)	26.4 [4.2] (12)	18.9 [2.9] (24)	14.2 [2.2] (43)	11.3 [1.8] (68)	11.0 [1.7] (72)
01192500	9.3 [1.9] (8)	9.9 [2.0] (7)	9.3 [1.9] (8)	7.6 [1.6] (12)	6.1 [1.2] (19)	6.1 [1.2] (19)
01193500	12.4 [8.2] (12)	12.4 [8.2] (12)	12.4 [8.2] (12)	12.4 [8.2] (12)	10.8 [7.0] (16)	10.5 [6.8] (17)
01196500	8.7 [3.6] (8)	10.0 [4.1] (6)	9.3 [3.8] (7)	6.6 [2.7] (14)	5.4 [2.2] (21)	5.2 [2.1] (23)
01196620	13.8 [5.7] (8)	15.8 [6.5] (6)	14.7 [6.0] (7)	10.6 [4.3] (14)	8.7 [3.5] (21)	8.3 [3.3] (23)
01199000	5.2 [3.6] (9)	6.2 [3.8] (5)	5.2 [3.6] (9)	4.6 [3.5] (14)	4.2 [3.4] (23)	4.2 [3.4] (23)
01199050	13.5 [5.2] (8)	16.8 [6.4] (5)	12.8 [4.9] (9)	10.4 [4.0] (14)	8.1 [3.1] (23)	8.1 [3.1] (23)
01200000	11.0 [4.1] (8)	11.6 [4.2] (7)	11.6 [4.2] (7)	9.2 [3.9] (12)	7.7 [3.6] (18)	7.2 [3.4] (21)
01200500	7.0 [6.0] (9)	7.5 [6.3] (7)	7.5 [6.3] (7)	6.4 [5.6] (12)	5.6 [5.0] (18)	5.3 [4.7] (21)
01204000	13.8 [2.9] (12)	13.8 [2.9] (12)	13.8 [2.9] (12)	12.7 [2.8] (14)	10.2 [2.2] (22)	9.8 [2.1] (24)
01205500	11.6 [7.3] (9)	14.2 [7.8] (5)	14.2 [7.8] (5)	10.5 [7.0] (12)	8.9 [6.5] (20)	8.6 [6.3] (22)

Table 13.--Selected results of K-CERA analysis-Continued

Station	Current, 1984, operation (267)	U.S. Geological Survey stream-gaging budget in Connecticut, in thousands of in 1984 dollars				
		255	264	280	305	310
	Standard error of instantaneous discharge, in percent [Equivalent Gaussian spread, in percent] (Number of visits)					
01205600	15.4 [6.7] (8)	17.6 [7.5] (6)	11.8 [5.1] (14)	9.0 [3.9] (24)	7.5 [3.2] (35)	7.1 [3.1] (39)
01205700	19.7 [10.1] (8)	22.3 [11.2] (6)	15.3 [8.0] (14)	11.8 [6.2] (24)	9.8 [5.1] (35)	9.3 [4.8] (39)
01206900	14.8 [4.5] (8)	16.9 [5.0] (6)	12.2 [3.9] (12)	9.6 [3.1] (20)	7.9 [2.6] (30)	7.9 [2.6] (30)
01208013	18.1 [8.7] (8)	20.8 [10.2] (6)	16.3 [7.7] (10)	12.6 [5.8] (17)	10.0 [4.6] (27)	9.7 [4.4] (29)
01208420	19.8 [9.2] (8)	22.4 [10.2] (6)	16.5 [8.0] (12)	12.7 [6.2] (21)	10.4 [5.1] (32)	9.9 [4.8] (35)
01208500	11.7 [5.4] (8)	13.4 [6.4] (6)	9.6 [4.4] (12)	7.3 [3.3] (21)	6.0 [2.7] (32)	5.7 [2.6] (35)
01208873	18.2 [12.4] (8)	22.9 [16.2] (5)	14.4 [9.6] (13)	11.3 [7.4] (21)	9.2 [6.0] (32)	9.1 [5.9] (33)
01208925	17.2 [8.6] (8)	21.4 [11.2] (5)	13.6 [6.6] (13)	10.7 [5.1] (21)	8.7 [4.1] (32)	8.6 [4.0] (33)
01208950	17.5 [9.6] (8)	21.3 [11.1] (5)	14.1 [7.9] (13)	11.3 [6.4] (21)	9.2 [5.2] (32)	9.1 [5.1] (33)
01208990	12.2 [8.2] (8)	12.2 [8.2] (8)	12.2 [8.2] (8)	9.8 [6.5] (13)	8.2 [5.4] (19)	7.8 [5.1] (21)
01209700	12.9 [10.3] (8)	12.9 [10.3] (8)	12.9 [10.3] (18)	10.2 [7.9] (13)	8.5 [6.5] (19)	8.1 [6.1] (21)

1/Square root of averaged station variance.

A minimum budget of $255,000 is required to operate the 90-site 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 16.3 percent. The minimum standard error of 6.2 percent would occur at station 01199000 while the maximum of 26.4 percent would occur at 01191000.

The maximum budget analyzed was $350,000, which resulted in an average standard error of estimate of 6.6 percent. Thus, increasing the budget by a third in conjunction with policy change would halve the average standard error that results from the current policy and current budget. 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 that was added to the stream-gaging records because of less than perfect instrumentation. The curve, labeled "Without missing record" on figure 18, 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 $255,000, the impacts of less than perfect equipment are greatest; average standard errors increase from 7.2 to 16.3 percent. At the other budgetary extreme of $350,000, under which stations are visited more frequently and equipment should be more reliable, average standard errors increased from 3.1 percent for ideal equipment to 6.6 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 possibly could be anticipated for the stream-gaging program in Connecticut.

Using the current operating policy budget and altering visits to seven locations as a result of a comparison of standard error of estimate of current policy and minimum visit constraints, an analysis was made with new visit constraints at these seven locations and 51 routes. The analysis resulted in a budget and average standard error for Connecticut that was approximately equal to budget and standard error of current policy. The analysis, however, yielded a lower standard error of estimate at one site, with four additional visits, while maintaining the same standard error of estimate at the other six locations with one less visit at each site.

Conclusions from the K-CERA Analysis

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

1. Present scheduling of field activities in the streamgaging program could be altered to reduce the current average standard error of 14.5 percent to 11.7 percent with the current budget of $267,000. This shift would result in some increases and decreases in accuracy of records at individual sites. The funding for stations with unacceptable accuracies for the data uses could be renegotiated with the data users.

2. It would be useful to repeat the K-CERA analysis with new stations included whenever sufficient information about the characteristics of the new stations has been obtained.

3. Schemes for reducing the probabilities of missing record, such as increased use of local gage observers and satellite relay of data, may be cost-effective alternatives to current methods for providing streamflow information.

SUMMARY

Currently 62 stream-gaging stations (59 in Connecticut and 3 in Massachusetts) and equipment at 11 reservoirs and 17 ground-water wells are operated by the Connecticut Office at a cost of $267,000. Seven separate sources of funding contribute to this program and eight separate uses were identified for data from a single gage.

In an analysis of data uses, no stations were identified for discontinuation based on a lack of need for the streamflow data.

The current (1984) policy for operation of the 90-site program would require a budget of $267,000 per year. The overall level of accuracy of the records at the 90 sites could be reduced from 14.5 to 11.7 percent (Figure 18) if the allocation of resources among the gages was altered.

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 manfunctions 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.

Analyses of the cost-effectiveness of the stream-gaging program may yield useful information. Future analyses may include investigation of the optimum ratio of discharge measurements to total site visits for each station and cost-effective ways to reduce the probabilities of lost correlative data. Future studies also may 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., 1962, Factors influencing the occurrence of floods in a humid region of diverse terrain: U.S. Geological Survey Water-Supply Paper 1580-B, p. B23-B25.

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.

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

Fenneman, N. M., 1938, Physiography of Eastern United States: New York, McGraw-Hill Book Co., 714 p.

Fontaine, R. A., Moss, M. E., Smath, J. A. and Thomas, W. O., Jr., 1984, Cost effectiveness of the stream-gaging program in Maine-a prototype for nationwide implementation: U.S. Geological Survey Water Supply Paper 2244, 39 p.

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

Helwig, J. T., and Council, K. A., editors, 1979, SAS User's Guide, SAS Institute Inc., Raleigh, North Carolina, p. 391-396.

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.

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.

Riggs, H. C., 1973, Regional analysis of streamflow characteristics: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. 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 characteristics: U.S. Geological Survey Water-Supply Paper 1975, 55 p.

Thomas, M. P., and Cervione, M. A., Jr., 1970, A proposed streamflow data program for Connecticut: Connecticut Water Resources Bulletin no. 23, 18 p., 1 pl.

U.S. Geological Survey, 1983, Water Resources Data Connecticut, Water Year 1982: U.S. Geological Survey Water-Data Report CT-82-1, 327 p.

Weiss, L. A., 1983, Evaluation and design of a streamflow-data network for Connecticut: Connecticut Water Resources Bulletin no. 36, 30 p.

* GOVERNMENT PRINTING OFFICE: 1988 - 501093 / 80004