

# **COST-EFFECTIVENESS OF THE STREAM-GAGING PROGRAM IN PUERTO RICO AND THE U.S. VIRGIN ISLANDS**

By Ken Reid

---

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 90-4088



San Juan, Puerto Rico

1991

**UNITED STATES DEPARTMENT OF THE INTERIOR**

**MANUEL LUJAN JR.,** *Secretary*

**U.S. GEOLOGICAL SURVEY**

**Dallas L. Peck,** *Director*

---

For additional information write to:

District Chief  
U.S. Geological Survey  
P.O. Box 364424  
San Juan, Puerto Rico 00936-4424

Copies of this report can be purchased from:

U.S. Geological Survey  
Books and Open-File Reports  
Federal Center  
Box 25425, Denver, Colorado 80225

# CONTENTS

	<b>Page</b>
Abstract .....	1
Introduction .....	1
History of the stream-gaging program in Puerto Rico and the U.S Virgin Islands .....	2
Present stream-gaging program in Puerto Rico and the U.S. Virgin Islands .....	3
Uses, funding, and availability of continuous-streamflow data .....	3
Data-use classes.....	3
Regional hydrology .....	3
Hydrologic systems .....	3
Legal obligations .....	9
Planning and design.....	9
Project operation.....	9
Hydrologic forecasts.....	9
Water-quality monitoring .....	9
Research.....	9
Other .....	9
Funding.....	9
Frequency of data availability .....	10
Conclusions pertaining to data uses .....	10
Alternative methods of developing streamflow information.....	10
Description of regression analysis .....	11
Regression results.....	12
Summary of second phase of analysis .....	16
Cost-effective resource allocation.....	16
Discussion of the model .....	16
Description of mathematical program.....	17
Application of the model in Puerto Rico and the U.S. Virgin Islands .....	18
Definition of variance when the station is operating .....	18
Definition of variance when record is missing .....	19
Discussion of routes and costs .....	21
Results .....	26
Summary of third phase of analysis .....	30
Summary .....	30
Selected References .....	32

## ILLUSTRATIONS

	Page
Figure	
1. Map showing geographic setting of Puerto Rico and the U.S. Virgin Islands.....	2
2. Graph showing history of continuous-record stream-gaging stations in operation in Puerto Rico and the U.S. Virgin Islands .....	3
3. Map showing location of continuous surface-water data-collection sites in Puerto Rico.....	4
4. Map showing location of continuous surface-water data-collection sites in the U. S. Virgin Islands .4	4
5. Hydrographs for observed and simulated streamflow for sites 50035000 - Río Grande de Manatí at Ciales, and 50038100 - Río Grande de Manatí at Highway 2 near Manatí, Puerto Rico.....	12
6. Graph showing rating curve plotted on logarithmic axes using straight-line segments.....	19
7. Graph showing typical uncertainty function for instantaneous discharge and number of visits for selected stations in Puerto Rico and the U.S. Virgin Islands.....	19
8. Graph showing average standard error per gaging station as a function of budget.....	26

## TABLES

	Page
Table	
1. Selected data for continuous-record gaging stations for Puerto Rico and the U.S. Virgin Islands, 1984.....	5
2. Data-use, source of funding, and frequency of data availability for continuous-record gaging stations in Puerto Rico and the U.S. Virgin Islands, 1984.....	7
3. Gaging stations used as dependent variables in the regression modeling of daily streamflow at selected sites in Puerto Rico and the U.S. Virgin Islands.....	12
4. Summary of calibration for regression modeling of daily streamflow at selected gage sites in Puerto Rico and the U.S. Virgin Islands.....	13
5. Summary of the autocovariance analysis.....	20
6. Statistics of record reconstruction.....	22
7. Summary of the routes that may be used to visit stations in Puerto Rico and the U.S. Virgin Islands.....	23
8. Selected results of K-CERA analysis.....	27

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

For the convenience of readers who may prefer to use metric (International System) units rather than the inch-pound units used in this report, values may be converted by using the following factors:

<b>Multiply inch-pound units</b>	<b>By</b>	<b>To obtain SI units</b>
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
square mile (mi <sup>2</sup> )	259.0	hectare (ha)
cubic foot per second (ft <sup>3</sup> /s)	0.02832	cubic meter per second (m <sup>3</sup> /s)

# **COST-EFFECTIVENESS OF THE STREAM-GAGING PROGRAM IN PUERTO RICO AND THE U.S. VIRGIN ISLANDS**

**By Ken Reid**

## **ABSTRACT**

*This report documents the results of a study of the cost-effectiveness of the stream-gaging program in Puerto Rico and in the U.S. Virgin Islands. Data uses and funding from 12 sources are identified for the 50 continuous-record surface-water gaging sites currently (1990) being operated in Puerto Rico and the U.S. Virgin Islands. With a budget of \$310,000, the average standard error of estimate of the present operation is 20.6 percent. However, the analysis indicates that with a budget of \$500,000, the average standard error of estimate could be reduced to 11.3 percent.*

*The present frequency of visits to continuous-record surface-water data-collection sites is monthly at all but one site in Puerto Rico. The four sites in the U.S. Virgin Islands are visited six times per year. Given a budget of \$350,000 (equivalent 1984 dollars), the average standard error could be reduced to 15.6 percent. However, this would require that 20 stations be visited at a frequency of two to three times per month, with the remainder visited as few as six times per year. The logistics required for assigning personnel and vehicles to the field at the computed frequencies would not be feasible. Therefore, the frequency of vehicle and personnel visits would probably be reduced to about twice per month at most sites. This alternate approach could result in a standard error of estimate of about 17 percent with a \$40,000 increase in the budget.*

*All stations were identified as necessary in the present network, and no stations could be replaced by data simulation using alternative methods, such as regression analysis and other simulation means, because these methods tend to be inaccurate. Also, most continuous-record stations are multiple-purpose ones where data are collected for more than one cooperator or Federal agency. There is a need for long-term Index and Benchmark continuous-record discharge stations in Puerto Rico and the U.S. Virgin Islands. These stations are needed on streams draining 1 to 15 square mile areas that are mainly unaffected by the activities of man. Possible sites for such stations are on streams draining areas from Federal and Commonwealth parkland. The data obtained from such areas could be used*

*to discriminate between changes due to man's activities and those due to natural trends. This information is particularly needed where tropical rain forests exist, such as in El Yunque, Puerto Rico.*

*There are no continuous-record stream-gaging stations on Puerto Rico's offshore islands of Vieques and Culebra, and only four such stations exist for the three major U.S. Virgin Islands. Sufficient stations are needed on these islands to obtain information on streamflow, flood peaks, low flows, and quality of water for hydrologic studies leading to an improved understanding of surface-water availability and quality.*

## **INTRODUCTION**

The U.S. Geological Survey is the principal Federal agency involved in the collection of surface-water data throughout the Nation, the Commonwealth of Puerto Rico, and the U.S. Trust Territories. Streamflow-data collection is a major activity of the Water Resources Division of the U.S. Geological Survey. The data are collected in cooperation with State and local governments and other Federal agencies. The U.S. Geological Survey operates about 7,000 continuous-streamflow record gaging stations throughout the Nation. The operation of some of these stations extends back to the turn of the century. In Puerto Rico, about 20 stations have been operated continuously since 1960.

Any activity of long standing, such as the collection of surface-water data, should be reexamined at intervals, if not continuously, because of the change in objectives, technology, or external constraints. The last systematic nationwide evaluation of the U.S. Geological Survey's streamflow information program was completed in 1970 and documented by Benson and Carter (1973). In 1983, the U.S. Geological Survey began a 5-year reevaluation of the national stream-gaging program, and 20 percent of the program is analyzed each year. The objective of this analysis is to define and document the most cost-effective means of furnishing streamflow information.

This report documents the analyses of the stream-flow data-collection program in Puerto Rico and the U.S. Virgin Islands (fig. 1). It is organized into five sections. The first section is an introduction to the stream-gaging activities and the current program in Puerto Rico and the U.S. Virgin Islands. The middle three sections each contain discussions of individual steps in the analyses. Because of the sequential nature of the steps and the dependence of subsequent steps on the previous results, conclusions and suggestions are made at the end of each of the middle three sections. The final section summarizes conclusions and recommendations. A similar report for the State of Maine (Fontaine and others, 1984) was used as a prototype for this report. Refer to the Maine report for more specific details of the methods used.

As the first phase of every continuous-record gaging station, an analysis identifies the principal uses of the data and relates these uses to funding sources. Gaged sites for which data are no longer needed, are deficient, or fail to meet user demands, are identified. In addition, gaging stations are categorized by whether the data are available to users in a near real-time sense, on a periodic basis, or at the end of the water year (October through September).

The second phase 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 is an integrated information system in which data are provided by measurement 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 error of estimate of instantaneous discharge to the frequency of visits to the gaging stations) for all stations in the analyses (Fontaine, and others, 1984). 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 data. The stream-gaging program that results from these analyses will meet the expressed water-data needs in the most cost-effective manner.

### History of the Stream-Gaging Program in Puerto Rico and the U.S. Virgin Islands

The Puerto Rico Water Resources Authority (now the Puerto Rico Electric Power Authority) began stream-flow gaging in Puerto Rico around 1943. The first gaging station was operated in 1943 at Río Caonillas near Utuado, Puerto Rico. The U.S. Geological Survey, under Public Law 29 (43 U.S.C. 47), began a stream-gaging program in Puerto Rico in 1957. A similar program was begun in 1961 in the U.S. Virgin Islands.

The number of continuous-record surface-water sites in the U.S. Geological Survey, Water Resources Division, Caribbean District varied from 1 in 1943 and

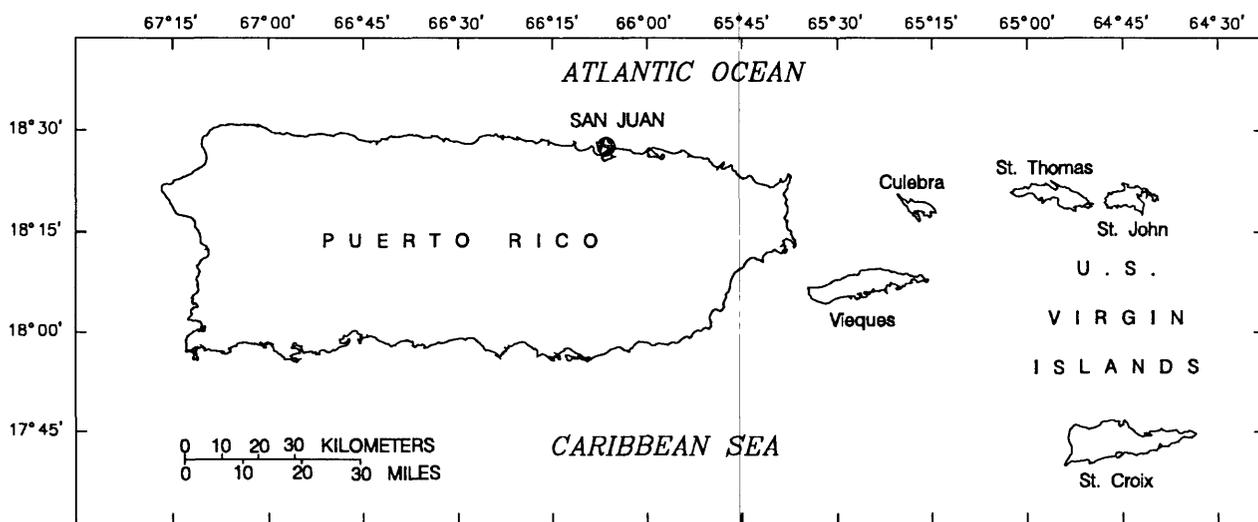


Figure 1.-- Geographic setting of Puerto Rico and the U.S. Virgin Islands.

1944 to 14 in 1950 and decreased to 6 by 1958. Some of the records collected by the Puerto Rico Water Resources Authority (PRWRA) are now in the files of the Caribbean District office of the U.S. Geological Survey, Water Resources Division in San Juan, Puerto Rico. The number of sites increased to a maximum of 60 in 1971, then decreased to 34 in 1978. During the 1984 water year, 58 sites were in operation, but by the end of the year, several sites had been discontinued. Fifty continuous surface-water data sites were continued and could be used for the last part of this study. The number of continuous surface-water stations for which records are collected for part or all of each water year since 1943 are summarized in figure 2.

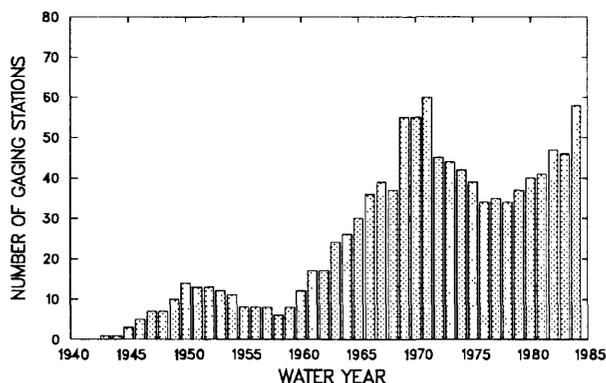


Figure 2.-- History of continuous-record stream-gaging stations in operation in Puerto Rico and the U.S. Virgin Islands.

A network of partial-record data-collection stations was begun in Puerto Rico in 1959 to 1960. The purpose of this network has been to define peak-flood characteristics used for highway design, bridge site analyses, delineation of flood-prone areas, and hydraulic analyses. The increasing costs of operation and constraints on funds and manpower has resulted in the discontinuation of most of the partial-record peak-flood stations (López and Fields, 1970). Only one partial-record peak-flood station (50115900 Río Portugués at Highway 14 at Ponce) is now being operated.

#### Present Stream-Gaging Program in Puerto Rico and the U.S. Virgin Islands

The Caribbean District of the U.S. Geological Survey is comprised of Puerto Rico and the U.S. Virgin Islands. The region includes an area of 3,605 mi<sup>2</sup> (square miles) (including the islands of Vieques and Culebra and other offshore islands). The Caribbean District is in the northeast corner of the Greater Antilles and about 1,000 mi (miles) east-southeast of Miami, Florida. The islands of the northeast part of the Greater Antilles

form a divide between the Atlantic Ocean to the north and east and the Caribbean Ocean to the south and southwest.

The Caribbean District currently (1990) operates 51 surface-water streamflow stations (50 continuous-record discharge sites and 1 partial record crest-stage station site). Four of the surface-water streamflow stations are operated in the U.S. Virgin Islands of which two are on St. Thomas, one on St. John, and one on St. Croix. No benchmark continuous-record streamflow sites have been established on any of the Caribbean District islands. Gage location is shown in figures 3 and 4; and station number, name, drainage area, and period of record of the continuous-record gages are shown in table 1.

#### USES, FUNDING, AND AVAILABILITY OF CONTINUOUS-STREAMFLOW DATA

The relevance of a stream-gaging station is defined by the uses that are made of the data that are produced from the station. The uses of the data from each stream-gaging station in the Puerto Rico and the U.S. Virgin Islands were identified from a survey of known data users (table 2). Also included as part of the survey were the sources of funding and the frequency of data availability for each station. The survey documented the importance of each station and identified gaging stations that may be considered for discontinuance.

#### Data-use Classes

Data uses identified by the survey are categorized into nine classes as defined below.

#### Regional Hydrology

To be useful in defining regional hydrology, the data from a gaging station must be largely unaffected by manmade storage or diversion. In this class of uses, the effects on streamflow are limited to those caused primarily by land-use and climate changes.

Large amounts of manmade storage may exist in a basin and stations in such basins are useful, provided the outflow is uncontrolled. These stations are useful in developing regionally transferable information about the relations between stations classified in the regional hydrology category. Forty-four stations in Puerto Rico and the U.S. Virgin Islands are in this category, as listed in table 2.

#### 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 sys-



Table 1.--Selected data for continuous-record gaging stations for Puerto Rico and the U.S. Virgin Islands, 1984  
 [mi.<sup>2</sup>, square mile; ft<sup>3</sup>/s, cubic feet per second; \*, not used in initial analysis; IND., indetermined;  
 --, not sufficient years of record to compute mean annual flow; PR, Puerto Rico; VI, U.S. Virgin Islands]

Station number	Station name	Drainage area (mi. <sup>2</sup> )	Period of record	Mean annual flow (ft <sup>3</sup> /s)
* 50010600	Río Guajataca above Lago Guajataca, PR	IND.	1984-	--
* 50011200	Río Guajataca below Lago Guajataca, PR	IND.	1984-	--
* 50011400	Río Guajataca above mouth near Quebradillas, PR	IND.	<sup>1</sup> 1969; 1969-1970, 1984-	--
* 50014800	Río Camuy near Bayaney, PR	IND.	1984-	--
* 50015700	Río Camuy near Hatillo, PR	IND.	1984-	--
50027750	Río Grande de Arecibo above Arecibo, PR	200	1982-	--
50028000	Río Tanamá near Utuado, PR	18.4	<sup>1</sup> 1944-58; 1959-	48.4
50028400	Río Tanamá at Charco Hondo, PR	57.6	1969-71, 1981-	--
50031200	Río Grande de Manatí near Morovis, PR	55.2	1965-	103
50035000	Río Grande de Manatí at Ciales, PR	128	<sup>1</sup> 946-53, <sup>3</sup> 1956-57; <sup>1</sup> 1959-60; 1960-	256
50038100	Río Grande de Manatí at Hwy 2 near Manatí, PR	197	<sup>1</sup> 1963-68; 1970-	362
50038320	Río Cibuco below Corozal, PR	15.1	1969-	28.3
50039500	Río Cibuco at Vega Baja, PR	99.1	1973-	119
50043000	Río de la Plata at Proyecto la Plata, PR	54.8	<sup>1</sup> 1958; <sup>1</sup> 1959-60; 1960-	111
50046000	Río de la Plata at Toa Alta, PR	200	<sup>1</sup> 1959; 1960-	267
50050900	Río Grande de Loiza at Quebrada Arenas, PR	6.0	1977-	30.6
* 50051150	Quebrada Blanca at Jagual, PR	3.25	1984-	--
* 50051180	Quebrada Salvatierra near San Lorenzo, PR	3.74	1984-	--
50051310	Río Cayaguas at Cerro Gordo, PR	10.2	1977-	47.9
* 50053050	Río Turabo at Borinquen, PR	7.89	<sup>1</sup> 1983-	--
50055000	Río Grande de Loiza at Caguas, PR	89.8	<sup>1</sup> 1959; <sup>1</sup> 1959; 1959-	218
* 50055650	Quebrada Caimito near Juncos, PR	0.82	1984-	--
50056400	Río Valenciano near Juncos, PR	16.4	1971-	49.4
* 50056900	Quebrada Mamey near Gurabo, PR	2.3	1983-	--
50057000	Río Gurabo at Gurabo, PR	60.2	<sup>1</sup> 1958; <sup>1</sup> 1959; 1959-	131
50061800	Río Canóvanas near Campo Rico, PR	9.84	1967-	27.5
* 50063440	Quebrada Sonadora near El Verde, PR	1.01	1983-	--
* 50063500	Quebrada Toronja at El Verde, PR	0.064	1983-	--
50063800	Río Espíritu Santo near Río Grande, PR	8.62	<sup>1</sup> 1959-63; 1966-	56.6
50065500	Río Mameyes near Sabana, PR	6.88	1967-73; 1983-	57.7
50067000	Río Sabana at Sabana, PR	3.96	1979-	17.5
50071000	Río Fajardo near Fajardo, PR	14.9	<sup>10</sup> 1960-61; 1961-	67.8
50075000	Río Icacos near Naguabo, PR	1.26	<sup>11</sup> 1945-53; <sup>4</sup> 1953-62; <sup>12</sup> 1962-66; 1979	15.3
50092000	Río Grande de Patillas near Patillas, PR	18.3	<sup>9</sup> 1959-65; 1966-	59.9
* 50106500	Río Coamo near Coamo, PR	46	1984-	--

Table 1.--Selected data for continuous-record gaging stations for Puerto Rico and the U. S. Virgin Islands, 1984--Continued

Station number	Station name	Drainage area <sup>2</sup> (mi <sup>2</sup> )	Period of record	Mean annual flow (ft <sup>3</sup> /s)
* 50108000	Río Descalabrado near Los Llanos, PR	12.9	1984-	--
* 50111500	Río Jacaguas at Juana Díaz, PR	49.8	1984-	--
50112500	Río Inabón at Real Abajo, PR	9.7	1962-63; 1964; 1964-70, 1971-	18.4
50114000	Río Cerrillos near Ponce, PR	17.8	1964; 1964-	35.1
50115000	Río Portugués near Ponce, PR	8.82	1964; 1964-	18.0
50124200	Río Guayanilla near Guayanilla, PR	18.9	1981-	--
* 50129900	Laguna Cartagena Outflow near Boquerón, PR	IND.	1984-	--
50136000	Río Rosario at Rosario, PR	16.4	1960-66; 1975	45.1
50138000	Río Guanajibo near Hormigueros, PR	120	1959; 1959-67; 1973-	219
50144000	Río Grande de Añasco near San Sebastián, PR	94.3	1963-	309
50147800	Río Culebrinas at Hwy 404 near Moca, PR	71.2	1967-	300
50252000	Bonne Resolution Gut at Bonne Resolution, St. Thomas, VI	0.49	1962-67; 1981; 1982-	0.21
50276000	Turpentine Run at Mariendal, St. Thomas, VI	2.97	1963-69; 1978-80; 1982-	1.04
50295000	Guinea Gut at Bethaney, St. John, VI	0.37	1963-67; 1982-	0.072
50345000	Jolly Hill Gut at Jolly Hill, St. Croix, VI	2.1	1963-68; 1962; 1969; 1982-	0.027

6

1 Monthly discharge measurements only.

2 Daily stage and two to four measurements per month by the Puerto Rico Water Resources Authority.

3 Unpublished, available in files of Caribbean District office and in the National Water Data Storage and Retrieval System, Reston, Virginia.

4 Annual maximum discharge only.

5 Occasional measurements only.

6 Measurements only.

7 Low-flow measurements only.

8 Occasional low-flow measurements only.

9 Annual low-flow and occasional measurements only.

10 Occasional low- and peak-flow measurements only.

11 Operated by the Puerto Rico Water Authority.

12 Annual low-flow measurements.

13 Gage-height records only; in files of the Puerto Rico Water Resources Authority.

Table 2. ---Data use, source of funding, and frequency of data availability for continuous-record gaging stations in Puerto Rico and the U.S. Virgin Islands, 1984

[1, North Coast Limestone project; 2, Puerto Rico Department of Natural Resources (PRDNR); 3, Puerto Rico Environmental Quality Board (PREQB); 4, Puerto Rico Department of Agriculture (PRDOA); 5, Long-term index gaging station, used for water resources review; 6, National Stream-Quality Accounting Network (NASQAN) Station; 7, Rio Grande de Loiza water-supply project; 8, Puerto Rico Aqueduct and Sewer Authority (PRASA); 9, U.S. Corps of Engineers (USCOE); 10, Sediment transport in Rio Grande de Loiza; 11, El Yunque watershed project; 12, Center of Energy and Environmental Research (CEER); 13, Pollution network; 14, G.W. Lejas Valley; 15, Puerto Rico Industrial Development Corporation (PRIDCO); 16, U.S. Virgin Islands Department of Public Works (VIDPW); A, annual; P, periodic; T, telemetric; \*, applies to this category; OFA, other federal agencies]

Station number	Data use				Source of funding			Frequency of data available	
	Regional hydrology systems	Hydrologic systems	Planning and design	Water quality monitoring	Research	Federal program	OFA program		Coop program
50010600	*				1			2	A
50011200	*				1			2	A
50011400	*			3	1			2	A
50014800	*				1			2	A
50015700	*				1			2	A
50027750	*		4	3				3,4	A
50028000	*	5	3	3				3	A
50028400	*		3,4					3,4	A
50031200	*		3	3				3	A
50035000	*		3					3	A
50038100	*	5	3	3,6		6		3	A,P
50038320	*			3				3	A
50039500	*	5		3				3	A
50043000	*	5		3				3	A
50046000	*			3,6		6		3	A
50050900	*		7	2,3,8			9	2,3,8	A
50051150	*			2,3,8			9	2,3,8	A
50051180	*			2,3,8			9	2,3,8	A
50051310	*		7	2,3,8			9	2,3,8	A
50053050	*			2,3,8			9	2,3,8	A
50055000	*	5		3,8				3,8	A,T
50055650	*			10			9	2,3,8	A
50056400	*		8,9	3			9	3,8	A
50056900	*			10				2,3,8	A
50057000	*	5	8,9	3				3,8	A
50061800	*							3	A
50063440	*		11		12			12	A
50063500	*		11		12			12	A
50063800	*			13				3	A
50065500	*		8					8	A

Table 2.--Data use, source of funding, and frequency of data availability for continuous-record gaging stations in Puerto Rico and the U.S. Virgin Islands, 1984--Continued

Station number	Regional hydrology	Hydrologic systems	Data use			Source of funding			Frequency of data available
			Planning and design	Water quality monitoring	Research	Federal program	OFA program	Coop program	
50067000	*		8					3	A
50071000	*	5	9					3	A,P
50075000	*			9,3	12		9	12	A
50092000	*	5		3,6			6		A
50106500	*		4	3				3,4	A
50108000	*		4	4				4	A
50111500	*		4	4				4	A
50112500	*	5	3					3	A,P
50114000	*		9	3			9	3	A,P
50115000	*		9	3,9			9	3	A
50124200	*	5	3					3	A
50129900	*				14			4,8,15	A
50136000	*		9				9	3	A
50138000	*	5	3	3				3,8	A
50144000	*	5		3,6			6		A
50147800	*	5	3	3				3	A
50252000	*		16					16	A
50276000	*		16					16	A
50295000	*		16					16	A
50345000	*		16					16	A

tems," including regulated systems, are designated as hydrologic system stations. They include stations used to gage diversions and return flows, and stations that are useful for defining the interaction of water systems. In Puerto Rico and the U.S. Virgin Islands, 13 continuous surface-water stations are included in this category.

The bench-mark and index stations (there are none in Puerto Rico or in the U.S. Virgin Islands) are included in the hydrologic systems category because they document current and long-term conditions of the hydrologic systems that they gage. Federal Energy Regulatory Commission (FERC) stations and international gaging stations, located on significant rivers that cross national boundaries, would also be included in this category. Bench-mark stations in areas such as El Yunque and Commonwealth parks in both Puerto Rico and its islands of Vieques and Culebra and in the U.S. Virgin Islands would provide useful hydrologic data for the assessment of man's activities on streamflow on these islands.

#### ***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 its legal responsibility. There are no stations in Puerto Rico and the U.S. Virgin Islands program used for that purpose.

#### ***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 instituted for such purposes and where this purpose is still valid. Currently, 29 stations in Puerto Rico and no stations in the U.S. Virgin Islands are being operated for planning and design purposes.

#### ***Project Operation***

Gaging stations in this category are regularly used, on an ongoing basis, to assist water managers in making operational decisions such as reservoir releases, hydro-power 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 may only be needed every few days. There are no stations in Puerto Rico and the U.S. Virgin Islands that are in current use for this purpose.

#### ***Hydrologic Forecasts***

Gaging stations in this category are regularly used to provide information for hydrologic forecasting. The latter includes flood forecasts for a specific river reach, or periodic (daily, weekly, monthly, or seasonal) flow-volume forecasts, which are routinely available to the forecasters on a rapid-reporting basis. On large streams, data may only be needed every few days. No stations in Puerto Rico and in the U.S. Virgin Islands are included in the hydrologic forecast category.

#### ***Water-Quality Monitoring***

Only gaging stations where water-quality or sediment-transport monitoring is being conducted, and the streamflow data contribute to the analysis, are designated as water-quality monitoring sites. There are 30 such continuous-record surface-water stations, of which 29 are in Puerto Rico and 1 is in the U.S. Virgin Islands. Four of these stations are National Stream Quality Accounting Network (NASQAN) sites, which are part of a U.S. Geological Survey national network designed to define water-quality trends in principal streams.

#### ***Research***

Gaging stations in this category are operated for a particular water-resource investigation. Typically, these are operated from 2 to 5 years. There are nine such stations used in support of research activities in Puerto Rico. However, hydrologic analyses are often required for estimating flow characteristics, such as low- and peak-flow frequency prediction for resource studies, and the stochastic statistics used usually require 30 or more years of continuous data to develop reliable models.

#### ***Other***

Stations in this category provide streamflow information for recreational planning, primarily for canoeists, rafters, and fishermen. No stations in Puerto Rico or the U.S. Virgin Islands are in this category.

#### ***Funding***

The sources of funding for the Puerto Rico and the U.S. Virgin Islands streamflow-data program are

1. Cooperative program.--Funded jointly by the U.S. Geological Survey and a non-Federal cooperating agency. Cooperating agency funds may be in the form of direct services or money. There are 48 continuous-record surface-water stations in this category with 44 in Puerto Rico and four in the U.S. Virgin Islands.

2. Other Federal Agencies (OFA) program.--Funds that have been transferred to the U.S. Geological Survey by OFA's. There are nine OFA funded sites in Puerto Rico and none in the U.S. Virgin Islands.

3. Federal programs.--Funds that have been directly allocated to the U.S. Geological Survey for the collection of streamflow or other water-related data. There are four federally funded sites in Puerto Rico and none in the U.S. Virgin Islands. In these categories, the identified sources of funding pertain only to the collection of continuous-streamflow data; sources of funding for other activities, particularly collection of water-quality samples that may be carried out at the site, may not necessarily be the same as those identified in table 2. Funds for the stream-gaging in Puerto Rico and the U.S. Virgin Islands are contributed from 16 sources (headnotes in table 2).

### Frequency of Data Availability

Frequency of data availability refers to the times at which the stream-flow data may be furnished to the users. 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 reports for Puerto Rico and the U.S. Virgin Islands (Curtis and others, 1985). These three categories are designated in table 2 as T, P, and A, respectively. Data for all 50 stations in the current Puerto Rico and U.S. Virgin Island program, are made available through the annual report. Data are available for one station on a real-time basis; at four stations, quarterly; and at two stations every 2 months.

### Conclusions Pertaining to Data Uses

A review of table 2 shows that all stations have at least one use, and most of the continuous-record surface-water stations in Puerto Rico, and the U.S. Virgin Islands have multiple uses. Most of the stations are used continuously for accounting and hydrologic operations. Although the stations may have been established for a single purpose, the data are available for other projects. As an example, gaging stations at Quebrada Sonadora near El Verde (50063440), and Quebrada Toronja at el Verde (50063500) in Puerto Rico are research stations in the Caribbean National Forest that will probably be discontinued at the end of the project. Both stations are in a tropical rainforest area minimally affected by cultural activities and are located at relatively high elevations (1,230 and 876 feet, respectively). Data for such stations in tropical areas are scarce but needed for comparison with streamflow affected by cultural activities. The information from such sites can be used to evaluate environmental changes and their effects on the

surface water, ground water, and quality of water of small basins. There is a need to continue these sites indefinitely and to establish additional similar sites in other areas of Puerto Rico and the U.S. Virgin Islands.

### ALTERNATIVE METHODS OF DEVELOPING STREAMFLOW INFORMATION

The second phase of the analyses of the stream-gaging program is to investigate alternative methods of providing daily streamflow information, instead of operating continuous-flow gaging stations. The objective of this part of the analyses is to identify gaging stations where alternative technology, such as flow routing or statistical methods, could provide accurate estimates of daily mean stream flow. There are no accuracy guidelines for the data; therefore, judgment is required in deciding whether the accuracy of the estimated daily flows would be adequate for the intended purpose.

The data uses at a station affect whether or not information can potentially be provided by alternative methods. For example, those 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, there might be a legal obligation to operate an actual gaging station that would preclude using alternative methods. The primary candidates for alternative methods are stations that are operated upstream or downstream from other stations on the same stream. The accuracy of the estimated streamflow at these sites may be adequate if flows are highly correlated between sites. Alternative methods could also be employed at gaging stations in similar watersheds, located in the same physiographic and climatic area.

All stations in the Puerto Rico and the U.S. Virgin Islands stream-gaging program were categorized as to their potential for utilization of alternative methods. Because of time limitations in this study, only the regression method (described below) was applied to five stations that best met the criteria as candidates for simulation. The categorization of gaging stations and the application of the specific methods are described in subsequent sections of this report. Hydrologic flow routing methods were not applied because there are very few pairs of stations on the same stream where they might work, and regression analysis provides similar results.

Desirable attributes of a proposed alternative method are (1) the method needs to be computer-oriented and easy to apply; (2) the method needs to have an available interface with the WATSTORE Daily Values File (Hutchison, 1975); (3) the method needs to be technically sound and generally acceptable to the hydro-

logic community; and (4) the proposed method needs to provide a measure of the accuracy of the simulated streamflow records. The regression method has these attributes.

### Description of Regression Analysis

Simple and multiple regression techniques can be used to estimate daily flow records. Unlike hydrologic routing, regression methods are not limited to locations where an upstream station exists on the same stream. Regression equations can be used to compute daily flows at a station (dependent variable) from measured daily flows at another station or combination of stations (independent variable). The independent variables in the regression analyses can include stations from different watersheds.

The regression method is easy to apply, provides indices of accuracy, and is widely used and accepted in hydrology. The theory and assumptions of the method are described in numerous textbooks such as those by Ezekiel and Fox (1930), Draper and Smith (1966), and Kleinbaum and Kupper (1978). The application of regression methods to hydrologic problems is described and illustrated by Riggs (1973) and Thomas and Benson (1970). Only a brief description of the regression analyses is provided in this report.

A linear regression model of the following form is commonly used for estimating daily mean discharges:

$$Y_i = B_0 + \sum_{j=1}^n B_j X_j + e_i, \quad (1)$$

where

$Y_i$  = daily mean discharge at station  $i$  (dependent variable);

$X_j$  = daily mean discharge(s) at  $n$  station(s)  $j$  (independent variables), these values may be lagged to approximate travel time between stations  $i$  and  $j$ ;

$B_0$  and  $B_j$  = regression constant and coefficients; and

$e_i$  = the random error term.

The above equation is calibrated using observed values of  $Y_i$  and  $X_j$  ( $B_0$  and  $B_j$  are estimated). The observed daily mean discharges can be retrieved from the WATSTORE Daily Values File (Hutchison, 1975). The values of discharge for the independent variables may be observed on the same day as discharges at the independent station or may be for previous or future days, depending on whether station  $j$  is upstream or downstream of station  $i$ . During calibration, the regression constant and coefficients ( $B_0$  and  $B_j$ ) are tested to determine if they are significantly different from zero. A

given independent variable is retained in the regression equation only if the variable's regression coefficient is significantly different from zero.

The regressions are calibrated using one period of time and verified or tested using a different period of time, to obtain a measure of the true predictive accuracy. The calibration and verification periods must be representative of the expected range of flows. The equation can be verified in two ways: (1) plotting the residuals (difference between simulated and observed discharges) against both the dependent and the independent variables in the equation, and (2) plotting the simulated and observed discharges over time. These tests are intended to determine whether the linear model is appropriate or some transformation of the variables is needed and whether there is any bias in the equation. These tests might indicate; for example, that a nonlinear regression equation is appropriate, or that the regression equation is biased in some way.

The use of regression to produce a simulated record at a discontinued gaging station causes the variance of this record to be less than the variance of an actual record of streamflow at the site. The reduction in variance is not a problem if the only concern is deriving the best estimate of a given daily mean discharge record. If, however, the simulated discharges are to be used in additional analyses where the variance of the data are important, least-squares regression models are not appropriate. Hirsch (1982) discusses this problem and describes several models that preserve the variance of the original data.

A two-level screening process was applied to gaging stations in Puerto Rico and the U.S. Virgin Islands, to evaluate the potential for use of alternative methods. The first level was based only on hydrologic considerations. The only concern at this level was whether it was hydrologically possible to simulate flows at a given station from information at other gages. The first-level screening was subjective; there was no attempt at that level to apply any mathematical procedures. Those stations that passed the first level of screening were then screened again to determine if the simulated data would be acceptable in view of the data uses in table 2. Even if simulated data were not acceptable for the given data uses, the analyses continued. This was done under the assumption that the data uses may change in the future. Where data uses required continuation of gaging, however, the result was predetermined. Although alternative methods were technically possible, they were unacceptable given the present uses of the data and the accuracy of simulated daily values obtained for sites with the best simulated figures of daily discharges.

## Regression Results

Regression methods were applied at streamflow stations shown in table 3 that exhibited high coefficients of correlation (at least 0.85) with other stations in the initial analyses. The initial results showed that regression methods would be unacceptable at many of the possible sites. Stations for which the initial correlation were not promising were eliminated from further consideration. Those stations with high correlations are listed in table 4. The data for the correlations were

converted to logarithmic form. Formulas from these correlations are in logarithms (Log) to the base ten format, as presented in table 4. The purpose of developing these formulas was to use them to compute synthetic daily discharge figures that could be compared to discharges in the Daily Values File (Hutchison, 1975). The percentage of time that the simulated streamflow is within 5, 10, and 15 percent of actual streamflow is listed in table 4. Hydrographs for simulated and observed discharges at two stations are plotted in figure 5.

Table 3.--*Gaging stations used as dependent variables in the regression modeling of daily streamflow at selected sites in Puerto Rico and the U.S. Virgin Islands*  
[mi<sup>2</sup>, square miles; Period of record, water years considered for the regression analysis]

Station number	Station name	Drainage	
		area (mi <sup>2</sup> )	Period of record
50031200	Río Grande de Manatí near Morovis, PR	55.2	1965-84
50035000	Río Grande de Manatí at Ciales, PR	128	1960-84
50038100	Río Grande de Manatí near Manatí, PR	197	1970-84
50071000	Río Fajardo near Fajardo, PR	14.9	1961-84
50114000	Río Cerrillos near Ponce, PR	17.8	1964-84

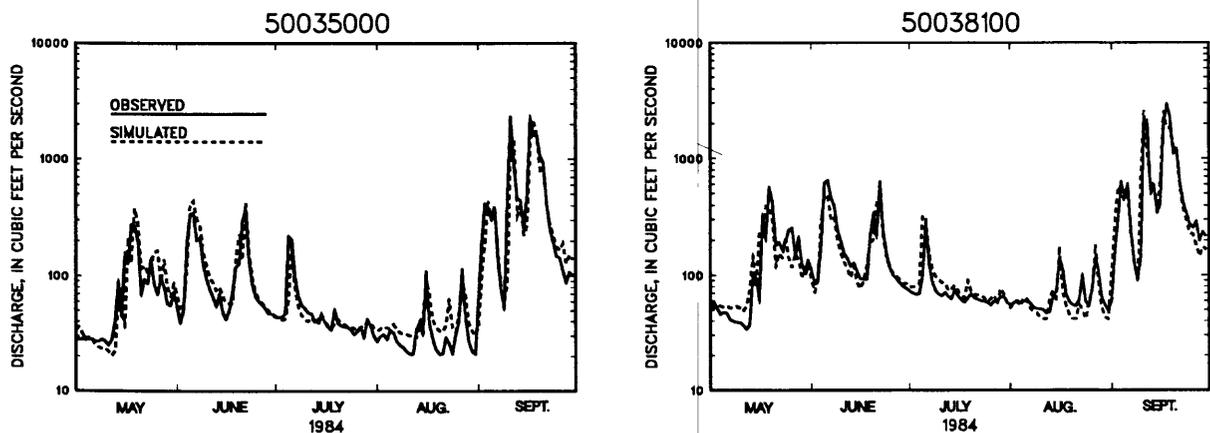


Figure 5.--Hydrographs for observed and simulated streamflow for sites 50035000-Río Grande de Manatí at Ciales, and 50038100-Río Grande de Manatí at Highway 2 near Manatí, Puerto Rico.

Table 4.--Summary of calibration for regression modeling of daily streamflow at selected gage sites in Puerto Rico and the U.S. Virgin Islands

Station number and name	Model	Percentage of simulated flow within 5 percent of observed	Percentage of simulated flow within 10 percent of observed	Percentage of simulated flow within 15 percent of observed	Calibration period (water years)
50031200-Río Grande de Manatí near Morovis, PR					
	November through May--				
	Log Q0312 = $-0.1438 + 0.07046 \text{ Log Q0350} + 0.0607 \text{ Log Q0381} + 0.2117 \text{ Log Q03832}$	18.5	38.5	61.3	1979-84
	Log Q0312 = $-0.2702 + 0.8331 \text{ Log Q0350} + 0.1107 \text{ Log Q0381}$	18.1	37.5	56.4	1979-84
	Log Q0312 = $-0.1153 + 0.7556 \text{ Log Q0350} + 0.2152 \text{ Log Q03820}$	18.7	39.0	61.3	1979-84
	Log Q0312 = $-0.2215 + 0.9312 \text{ Log Q0350}$	17.0	37.9	57.3	1979-84
	Log Q0312 = $0.6800 + 0.8747 \text{ Log Q03832}$	17.5	29.2	40.9	1979-84
	May through November--				
	Log Q0312 = $-0.1237 + 0.6203 \text{ Log Q0350} + 0.0820 \text{ Log Q0381} + 0.2787 \text{ Log Q3832}$	16.4	30.9	43.6	1979-84
	Log Q0312 = $-0.0703 + 0.6845 \text{ Log Q0350} + 0.2897 \text{ Log Q0381}$	17.6	33.3	51.2	1979-84
	Log Q0312 = $-0.0840 + 0.6886 \text{ Log Q0350} + 0.2836 \text{ Log Q03832}$	17.3	31.6	43.6	1979-84
	Log Q0312 = $-0.1749 + 0.8883 \text{ Log Q0350}$	10.7	20.6	36.9	1979-84
	Log Q0312 = $0.7169 + 0.8636 \text{ Log Q03832}$	10.2	18.2	25.8	1979-84
50035000-Río Grande de Manatí at Ciales, PR					
	November through May--				
	Log Q0350 = $-0.2893 + 1.0407 \text{ Log Q0381}$	25.8	48.5	65.0	1979-84
	Log Q0350 = $0.3525 + 0.0838 \text{ Log Q0395} + 0.9198 \text{ Log Q0312}$	18.4	37.5	53.3	1979-84
	May through November--				
	Log Q0350 = $-0.2472 + 1.02483 \text{ Log Q0381}$	22.3	43.1	59.1	1979-84

Table 4.--Summary of calibration for regression modeling of daily streamflow at selected gage sites in Puerto Rico and the U.S. Virgin Islands--Continued

Station number and name	Model	Percentage of simulated flow within 5 percent of observed	Percentage of simulated flow within 10 percent of observed	Percentage of simulated flow within 15 percent of observed	Calibration period (water years)
50038100-Rio Grande de Manati at Highway 2 near Manati, PR					
	November through May--				
	Log Q0381 = 0.3191 + 0.7404 Log Q0350 + 0.2276 Log Q0395	26.9	45.0	57.0	1982-84
	Log Q0381 = 0.4399 + 0.8859 Log Q0350	30.5	56.8	70.7	1979-84
	Log Q0381 = 0.8647 + 0.8294 Log Q0395	11.8	19.4	29.5	1982-84
	May through November--				
	Log Q0381 = 0.3248 + 0.7414 Log Q0350 + 0.2117 Log Q0395	35.6	58.7	72.5	1982-84
	Log Q0381 = 0.4657 + 0.8736 Log Q0350	59.6	74.4	80.3	1979-84
	Log Q0381 = 0.8928 + 0.7692 Log Q0395	12.0	23.6	35.5	1982-84
50071000-Rio Fajardo near Fajardo, PR					
	November through May--				
	Log Q0710 = -0.3459 + 0.8213 Log Q0638 + 0.2258 Log Q0655 + 0.3230 Log Q1150	5.8	13.6	19.4	1979-84
	Log Q0710 = 0.6642 + 0.2707 Log I125 + 0.6160 Log I150	5.4	10.5	14.9	1979-84
	May through November--				
	Log Q0710 = -0.2996 + 0.6936 Log Q0638 + 0.1332 Log Q0655 + 0.6142 Log Q1150	8.4	17.9	27.6	1979-84
	Log Q0710 = 0.4637 + 0.4709 Log Q1125 + 0.3936 Log Q1150 + 0.3936 Log Q1150	14.7	29.8	45.8	1979-84

Table 4.--Summary of calibration for regression modeling of daily streamflow at selected gage sites in Puerto Rico and the U.S. Virgin Islands--Continued

Station number and name	Model	Percentage of simulated flow within 5 percent of observed	Percentage of simulated flow within 10 percent of observed	Percentage of simulated flow within 15 percent of observed	Calibration period (water years)
50114000-Rio Cerrillos near Ponce, PR	November through May--				
	Log Q1140 = 0.3183 + 0.4412 Log Q1125 + 0.4805 Log Q1150 + 0.0389 Log Q0920	22.4	40.8	55.1	1984-84
	Log Q1140 = 0.3587 + 0.4416 Log Q1125 + 0.4976 Log Q1150	23.7	42.1	55.6	1979-84
	May through November--				
	Log Q1140 = 0.4202 + 0.4740 Log Q1125 + 0.3815 Log Q1150 + 0.0341 Log Q0920	13.6	31.3	45.5	1979-84
	Log Q1140 = 0.4657 + 0.4709 Log Q1125 + 0.3936 Log Q1150	14.8	29.8	45.8	1979-84

Most of the combinations of stations tested produced unacceptable results. The most common reason for poor correlation and regression results in the Puerto Rico and the U.S. Virgin Islands probably relates to orographic effects. The central mountain range of Puerto Rico induces higher precipitation along the northern coast, and creates a rain shadow along the southern coast. Also, along the northwest coast of Puerto Rico, streamflow relations are effected by a coastal karst limestone belt where there is a substantial exchange of water between the streams and the aquifer. Intense rainfall associated with tropical storms have no typical pattern and can cause extensive flooding in any part of the islands, even in areas with low average annual rainfall.

On the south coast of Puerto Rico, many streams lose water to the alluvial aquifer and correlation of flows between stream sites generally is poor. On the offshore islands of Puerto Rico (Vieques, Culebra, and Mona) and in the U.S. Virgin Islands, the amount of surface water data collected is not adequate for analyzing meaningful correlations. However, streamflow conditions similar to Puerto Rico probably exist in the smaller islands on a reduced scale, because similar characteristics in geography, geology, and weather occur in the islands of Culebra, Vieques, St. Thomas, St. John, and St. Croix. However, there are almost no data available for statistical evaluations of surface-water hydrology in these areas.

The results of regression analyses for selected combinations of streamflow sites are summarized in table 4. The differences between observed and simulated streamflows for discharges that are within 5, 10, and 15 percent of actual flow are given in percentage of simulated flows.

### Summary of Second Phase of Analysis

For the purposes of this study, acceptable accuracy is defined as 90 percent or more of the regression estimated streamflows being within 15 percent of the measured flow. None of the stations studied were within the range of acceptable accuracy for the application of alternative methods. At only two stations, Río Grande de Manatí at Ciales (50035000) and Río Grande de Manatí at Highway 2 near Manatí, Puerto Rico (50038100), were the regression models able to predict flows close to acceptable limits (fig. 5). From November through May, estimated flows at the two gages were within 15 percent of the measured flows 65 and 71 percent of the time, respectively. From May to November, estimates were within 15 percent of measured flows 59 percent of the time at site 50035000, and 80 percent of the time at site 50038100. For periods of missing

record, the regression procedures may be useful for estimating discharge record, but further verification of the calibrated model is needed.

Simulation of streamflow in Puerto Rico and the U.S. Virgin Islands may be a suitable alternative in those few instances where the inaccuracies of simulated information is acceptable, or where data for periods of missing record are needed. In these instances, the streamflow computed from a model should be compared with other methods, and the best results used on the basis of experience and judgment.

## COST-EFFECTIVE RESOURCE ALLOCATION

### Discussion of the Model

A set of techniques called K-CERA (Kalman filtering for Cost-Effective Resource Allocation) was developed by Moss and Gilroy (1980) to study the cost-effectiveness of networks of stream gages. The original application of the technique was in the analysis of a streamflow network operated to determine water consumption in the Lower Colorado River Basin (Moss and Gilroy, 1980). Because of the water-balance nature of that study, minimization of the total variance of errors of estimation of annual mean discharges was chosen as the measure of effectiveness of the network. This total variance is defined as the sum of the variances of errors of mean annual discharge at each site in the network. This measure of effectiveness is utilized on the large rivers and streams where discharge and, consequently, potential errors (in cubic feet per second) are greatest. Although this measure may be acceptable for a water-balance network, considering the many uses of data collected by the U.S. Geological Survey, concentration of effort on large rivers and streams is undesirable and inappropriate.

The original version of K-CERA was, therefore, altered 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 percent; average instantaneous discharge, in cubic feet per second; or average instantaneous discharge, in percent (Fontaine and others, 1984). The use of percentage errors effectively assigns equal weight to large and small streams. In addition, 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 continuously gaged sites as the measure of effectiveness of the data-collection activity.

The original version of K-CERA did not account for errors contributed by missing stage or other correlative data that are used to compute streamflow. Missing correlative data are more likely to increase with fewer service visits to a stream gage. 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 minimize the total error variance of the data-collection activity for given budgets and of the application of Kalman filtering (Gelb, 1974) used to determine the accuracy of a stream-gaging record are presented by Fontaine and others (1984). For more detail on either the theory or the applications of the K-CERA model, see Moss and Gilroy (1980) and Gilroy and Moss (1981).

### Description of Mathematical Program

The K-CERA methodology considers the cost effectiveness of a network of stream gages. This is determined by the total variance uncertainty, in either the annual mean discharge or the instantaneous discharge at all sites involved in the stream-gaging program, and by the cost of achieving that uncertainty. For the present study, the measure of uncertainty at each site was taken to be the variance of the percent of error in the instantaneous discharge. (See Fontaine and others, 1984).

The first step in estimating a site-specific uncertainty function (a relation between variance and number of visits to the site) is to determine a logarithmic discharge rating curve relating instantaneous discharge to gage height for each station involved in the stream-gaging program. The sequence of discharge residuals (in logarithmic units) from this rating (the discharge measurement minus the rating value) is analyzed as a time series.

The second step is to fit a lag-one-day autoregressive model to this temporal sequence of discharge residuals. The three parameters obtained from this analysis are: (1) the measurement variance - a measure of the variability of a current-meter measurement at the site, (2) the process variance - a measure of the variability about the rating in the absence of measurement error, and (3) the lag-one-day autocorrelation coefficient (RHD) - a measure of the memory in the sequence of discharge residuals. These three parameters determine the variance,  $V_f$ , of the percentage error in the estimation of instantaneous discharge whenever the gage height data at the site is available for use in the rating equation. Kalman filter theory, along with the assumption of a first-order Markovian process, is used to

determine this variance,  $V_f$ , as a function of the number of discharge measurements per year (Moss and Gilroy, 1980).

If the gage height data at the site is not available, the discharge may be estimated by correlation with nearby sites. The correlation coefficient shows the linear relations ( $\rho_c$ ) between streamflows with seasonal trends removed (detrended) at the site of interest and detrended streamflows at the other sites. The fraction of the variance of the streamflow at the primary site that is explained by data from other sites is  $\rho_c^2$ . The coefficient of variation of daily streamflows at the primary site, in percent is taken to be

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

where  $\sigma_i$  is the square root of the variance of daily discharges for the  $i$ th day of the year and  $\mu_i$  is the expected value of discharge on the  $i$ th day of the year. Thus the variance,  $V_r$ , of the percentage error during periods of reconstructed streamflow records is

$$V_r = (1 - \rho_c^2) C_v^2, \quad (3)$$

and the variance,  $V_e$ , of the percentage error during periods when neither primary correlative data nor reconstructed streamflow from nearby sites is available, is

$$V_e = C_v^2. \quad (4)$$

If the fraction of time when primary correlative data are available is denoted by  $\epsilon_f$  and the fraction of time when secondary streamflow data are available for reconstruction is  $\epsilon_r$  and  $\epsilon_e = 1 - \epsilon_f - \epsilon_r$ , the total percentage error variance,  $VT$ , is given by

$$VT = \epsilon_f V_f + \epsilon_r V_r + \epsilon_e V_e. \quad (5)$$

The fraction uptime,  $\epsilon_f$ , of the primary recorders at the site of interest is modeled by a truncated negative exponential probability distribution, which depends on  $\tau$ , the average time between service visits, and the reciprocal of the average time to failure when no visits are made to the site. The fraction concurrent downtime of the primary and secondary site is found by assuming independence of downtime between sites (Fontaine and others 1984).

The variance,  $VT$ , given by equation (5), which is a function of the number of visits to each site, is determined in the stream-gaging network. For a given site visitation strategy, the sum of the variance,  $VT$ , over all sites is taken as the measure of the uncertainty of the network. The variance,  $VT$ , given by equation (5) is one measure of the spread of a probability density function,

$gT$ . The function,  $gT$ , is a mixture of three probability density functions -  $gf$ ,  $gr$ , and  $ge$  - each of which is assumed to be a normal, or Gaussian, probability density with a mean equal to zero and the variance,  $V_f$ ,  $V_r$ , and  $V_e$ , respectively. Such a mixture is denoted by

$$gT = \epsilon_f gf + \epsilon_r gr + \epsilon_e ge. \quad (6)$$

In general, the density  $gT$  will not be a Gaussian probability density and the interval from the negative square root of  $VT$  to the positive square root of  $VT$  may include much more than 68.3 percent of the errors. This will occur because, while  $\epsilon_e$  may be very small,  $V_e$  may be extremely large. Actually, this standard error interval may include up to 99 percent of the errors.

To assist in interpreting the results of the analyses, a new parameter, Equivalent Gaussian Spread (EGS), was introduced by Fontaine and others (1984). The parameter EGS specifies the range in terms of equal positive and negative logarithmic units from the mean that would encompass errors with the same a priori probability as would a Gaussian distribution with a standard deviation equal to EGS; in other words, the range from -1 EGS to +1 EGS contains about two-thirds of the errors. For Gaussian distributions of logarithmic errors, EGS and standard error are equivalent. EGS is reported herein in units of percentage and an approximate interpretation of EGS is "two-thirds of the errors in instantaneous streamflow data will be within plus or minus EGS percent of the reported value." Note that the value of EGS is always less than or equal to the square root of  $VT$  and ordinarily is closer to  $V_f$ , which is the measure of uncertainty applicable during periods of no missing record the greatest portion of the time.

The cost portion of the input to the K-CERA methodology consists of determining practical routes to visit the stations in the network, the costs of each route, the cost of a visit to each station, the fixed cost of each station, and the overhead associated with the stream-gaging program.

The next step is to determine the frequency of visits to each of the gages for periodic maintenance, rejuvenation of recording equipment, or required periodic sampling of water-quality data.

All these costs, routes, constraints, and uncertainty functions are then used in an iterative search program to determine the number of times that each route is used during a year. The objectives of the program are the following: (1) the budget for the network should not be exceeded, (2) at least the minimum number of visits to each station should be made, and (3) the total uncertainty in the network should be minimized. This allocation of the predefined budget among the stream

gages is taken to be the optimal solution to the problem of cost-effective resource allocation. Due to the high dimensionality and non-linearity of the problem, the optimal solution may really be "near optimal." (See Moss and Gilroy, 1980, or Fontaine and others, 1984, if greater detail is desired.)

### Application of the Model in Puerto Rico and the U.S. Virgin Islands

The operation of the existing network was analyzed by the K-CERA techniques to consider alternative operative funding and collection of field data. The results of this phase of the analysis are described in the remainder of this section. The model assumes the uncertainty of discharge records at a given gage to be derived from three sources: (1) errors that result because the stage-discharge relationship is not perfect (applies when the gage is operating), (2) errors in reconstructing records based on data from another gage when the primary gage is not operating, and (3) errors inherent in estimated discharge when the gage is not operating and no correlative data are available to aid in record reconstruction. These uncertainties are measured as the variance of the percentage errors in instantaneous discharge. The proportion of time that each source of error applies depends on the frequency at which the equipment is serviced.

#### *Definition of Variance When Station is Operating*

The model used in this analysis assumes that the difference (residual) between instantaneous discharge (measurement discharge) and rating curve discharge is a continuous first-order Markovian process. The underlying probability distribution is assumed to be Gaussian (normal) with a zero mean and the variance of this distribution is referred to as process variance. Because the total variance of the residuals includes error in the measurements, the process variance is defined as the total variance of the residuals minus the measurement error variance.

Computation of the error variance about the stage-discharge relation was performed in three steps. A long-term rating was defined, generally based on measurements made during three or more water years, and deviations (residuals) of the measured discharges from the rating discharge were determined. A time-series analysis of these residuals defined the 1-day lag (lag-one) autocorrelation coefficient and the process variance required by the K-CERA model. Finally, the error variance is defined within the model as a function of the lag-one autocorrelation coefficient, the process and measurement variances, and the frequency of discharge measurements.

Long-term applicable rating curves were defined for each station used in the evaluation. In some cases, existing ratings adequately defined the long-term condition and were used in the analysis. At a majority of gages, however, this was not the case, and new rating were developed. The rating function used was of the following form:

$$LQM = B1 + B3 ( \text{LOG}(GHT - B2) ), \quad (7)$$

where

LQM = the logarithmic (base e) value of the measured discharge, and

GHT = the recorded gage height corresponding to the measured discharge.

The constants B1, B2, and B3 were determined by a graphical fit of straight line segments as illustrated in figure 6. The residuals about the long-term rating for individual gages defined the total variance. A review of discharge measurements made in Puerto Rico and the U.S. Virgin Islands indicated that the average standard error of open-water measurements was about 4 percent. The measurement variance for all gages, therefore, was defined as equal to the square of the 4 percent standard error. The process variance required in the model is, thus, the variance of the residuals about the long-term rating, minus the constant measurement variance.

Time-series analyses of the process variance were used to compute sample estimates of the lag-one autocorrelation coefficient; this coefficient is required to compute the variance during the time when the recorders are functioning.

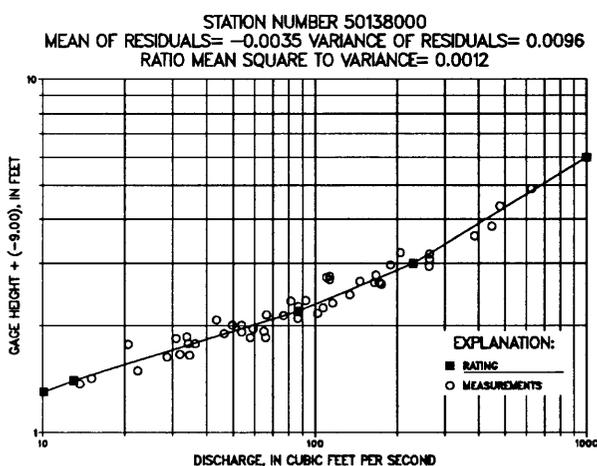


Figure 6.--Rating curve plotted on logarithmic axes using straight-line segments.

The values of lag-one autocorrelation coefficient, measurement and process variances, length of season (365 days), and data from the definition of missing record probabilities (6 percent - an average of missing record for all sites used in the computations) are used jointly to define uncertainty functions for each gaging station. The uncertainty functions give the relation between error variance and the number of visits (12 and 6 per year), assuming a measurement is made at each visit. Statistics of the uncertainty curves are given in table 5, and examples of typical uncertainty functions are shown in figure 7. The uncertainty curve for station 50345000 is representative of stations with a large process variance and that for station 50031200 represents stations with relatively small process variance. The uncertainty, in percent standard error of estimate for the number of visits per year (12), is 48 and 15 percent for stations 50345000 and 50031200, respectively, as shown in figure 7.

A total of 16 of the 50 stations in Puerto Rico were excluded from the analysis because the records were too short and the number of discharge measurements was insufficient to meet the assumptions of the model. These stations are marked by an asterisk in table 1 and are used as null sites in the final analysis so they will be accounted for in the cost and route computation analysis.

#### Definition of Variance When Record is Missing

When stage record is missing at a gaging station, the model assumes that the discharge record is either reconstructed using correlation with another gage or estimated from historical discharge for that period.

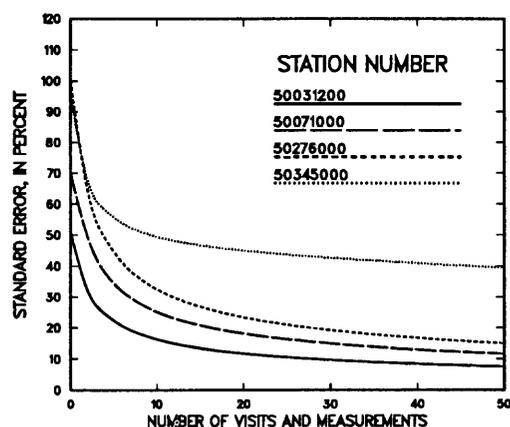


Figure 7.--Typical uncertainty function for instantaneous discharge and number of visits for selected stations in Puerto Rico and the U.S. Virgin Islands.

Table 5.--*Summary of the autocovariance analysis*

Station number	Number of measurements analyzed	RHO (1-day autocorrelation coefficient)	Measurement variance [(log base e)**2]	Process variance [(log base e)**2]
50027750	28	0.989	0.000302	0.00879
50028000	208	0.995	0.000302	0.00879
50028400	34	0.569	0.000302	0.00052
50031200	157	0.988	0.000302	0.00164
50035000	256	0.991	0.000302	0.01300
50038100	179	0.978	0.000302	0.00554
50038320	67	0.992	0.000302	0.02147
50039500	67	0.982	0.000302	0.01417
50043000	26	0.988	0.000302	0.02176
50046000	68	0.972	0.000302	0.01852
50050900	64	0.981	0.000302	0.00788
50051310	25	0.994	0.001599	0.04398
50055000	59	0.976	0.000302	0.00385
50056400	116	0.991	0.000302	0.00962
50057000	336	0.997	0.000302	0.22240
50061800	27	0.784	0.000302	0.00275
50063800	181	0.986	0.000302	0.00820
50065500	31	0.987	0.000302	0.02725
50067000	65	0.972	0.000302	0.01198
50071000	94	0.995	0.000302	0.05174
50075000	63	0.937	0.000302	0.00134
50092000	78	0.975	0.000302	0.04425
50112500	50	0.975	0.000302	0.09350
50114000	38	0.972	0.000302	0.02026
50115000	66	0.994	0.000302	0.09684
50124200	25	0.986	0.000302	0.07405
50136000	57	0.979	0.000302	0.00724
50138000	47	0.986	0.000302	0.00912
50144000	51	0.978	0.000302	0.00259
50147800	117	0.989	0.000302	0.00859
50252000	21	0.999	0.000302	0.15640
50276000	35	0.994	0.000302	0.25700
50295000	23	0.540	0.000302	0.03175
50345000	18	0.540	0.000302	0.03175

Fontaine and others (1984, p. 24) indicate that the fraction of time a record must be either reconstructed or estimated can be defined by a single parameter in a probability distribution of equipment failure times. The reciprocal of the parameter represents the average time to the average of all failure times since the last service visit. The average time to failure varies from site to site depending on the type of equipment at the site, the exposure to natural elements, such as floods and vandalism.

Data collected in Puerto Rico and in the U.S. Virgin Islands in recent years were reviewed to define the average time to failure for recording equipment and stage-sensing devices. In Puerto Rico, stream-gaging stations were examined 12 times per year, except for one station. Visits to the latter station and the four stations in the U.S. Virgin Islands were made six times per year. The average amount of missing record for stations was 6 percent. This average was computed over an 8-month (243-day) period for stations visited on a monthly basis and over a 16-month (486-day) period for stations visited on a bi-monthly basis.

The model defines the uncertainty as the sum of the multiples of the fraction of time each error source (rating, reconstruction, or estimation) is applicable and the variance of the error source. The variance associated with reconstruction and estimation of a discharge record is a function of the coefficient of cross correlation with the stations used in reconstruction and the coefficient of variation ( $C_v$ ) of daily discharges at the station. Daily streamflows for the last 26 water years were used to define seasonally averaged coefficients of variation for each station. In addition, cross-correlation coefficients (with seasonal trends removed) were defined for various combinations with other stations.

In current practice, many different sources of information are used to reconstruct periods of missing record. These sources include, but are not limited to, recorded ranges in stage (for graphic recorders with clock stoppage), known discharges on adjacent days, recession analyses, observer's staff-gage readings, weather records, highwater-mark elevations, and comparisons with nearby stations. However, most of these techniques are unique to a given station or to a specific period of missing record. Using all the information available, short periods (several days) of missing record usually can be reconstructed quite accurately. Even longer periods (more than a month) of missing record can be reconstructed with reasonable accuracy if observer's readings are available. If, however, none of these data are available, reconstruction of long periods of record can be subject to large errors. The present study could not reasonably quantify the uncertainty as-

sociated with all the possible methods of reconstructing missing record at the individual sites.

Historically, operating procedures have caused most periods of missing record to be measured in days rather than months. Given the low cross correlations and the relatively high variability of flow that usually occurs in Puerto Rico and the U.S. Virgin Islands, the model may overstate the uncertainty associated with short periods of missing record. For two stations a cross-correlation coefficient was not computed due to short or poor records, and a cross-correlation coefficient of 0.50 was arbitrarily used. For two stations cross-correlation coefficients less than 0.50 were computed and used. In reconstructing records, the cross correlation coefficient was, therefore, used as a surrogate for the knowledge of basin response that remains unquantified in the present model. This assumption is believed to be reasonable for short periods of missing record; it may cause the uncertainty to be overstated for long periods of missing record.

Uncertainty functions were defined for 34 of the 50 stations operated in the streamflow programs in Puerto Rico and the U.S. Virgin Islands. Statistics used to define those uncertainty functions are given in table 6.

#### *Discussion of Routes and Costs*

Although there are only 50 continuous surface-water stations in the network, a crest-stage gage (operated to record peak stages), ground-water observation wells, and quality of water monitors are serviced on the same field trips. The operating budgets for these other types of stations are not included in the surface-water budget being analyzed; however, the investigation could not ignore the additional mileage required to include these stations on field trips. These stations were, therefore, added to the 50 continuous surface-water stations to define the mileage associated with practical operating routes. These added stations acted as null stations in the analyses because there were no uncertainty functions or annual operating costs defined for them. Routes were defined for a total of 76 stations, including the null stations as listed in table 7. Uncertainty functions could not be defined for 16 of the 50 continuous surface-water stations. These 16 stations were treated like null stations except that all operating costs were included in the analyses.

Minimum visit constraints were defined for each of the 76 stations prior to defining the practical service routes. Minimum visits are dependent on the types of equipment and uses of the data. For example, water-quality samples generally are required on a monthly basis, so those stations where samples are collected must be visited at least once a month (or 12 times during

Table 6.--*Statistics of record reconstruction*

[Cv, coefficient of variation; CROSS, Cross Correlation Coefficients]

Station number	Cv	CROSS	Stations used in record reconstruction		
50027750	0.780	0.360	50038100	50028400	50035000
50028000	0.910	0.690	50144000	50031200	50035000
50028400	0.580	0.680	50028000	50035000	50038100
50031200	1.220	0.910	50035000	50038320	50038100
50035000	1.390	0.890	50031200	50038100	50039500
50038100	1.150	0.890	50035000	50039500	50027750
50038320	1.340	0.830	50039500	50031200	50046000
50039500	1.150	0.780	50038100	50046000	50035000
50043000	1.960	0.790	50055000	50046000	50057000
50046000	1.620	0.770	50038100	50043000	50039500
50050900	0.940	0.740	50055000	50056400	50043000
50051310	0.880	0.800	50050900	50056400	50055000
50055000	1.420	0.830	50057000	50050900	50056400
50056400	1.300	0.790	50057000	50061800	50075000
50057000	1.580	0.820	50055000	50056400	50061800
50061800	1.470	0.730	50063800	50067000	50071000
50063800	1.400	0.800	50061800	50075000	50057000
50065500	0.990	0.790	50061800	50063800	50071000
50067000	0.940	0.690	50071000	50063800	50065500
50071000	1.330	0.850	50063800	50065500	50075000
50075000	0.740	0.640	50065500	50056400	50071000
50092000	1.330	0.770	50050900	50051310	50071000
50112500	1.050	0.780	50114000		
50114000	1.010	0.850	50112500	50115000	50092000
50115000	1.190	0.790	50114000	50092000	50124200
50124200	0.710	0.520	50115000	50114000	50112500
50136000	0.930	0.680	50138000	50114000	50112500
50138000	1.070	0.700	50136000	50144000	50147000
50144000	0.830	0.760	50147800	50136000	50028000
50147800	1.140	0.540	50144000	50138000	50028000
50252000	1.100	0.500	50276000	50345000	
50276000	1.100	0.410	50252000	50345000	
50295000	1.160	0.610	50252000		
50345000	1.100	0.500	50276000		

Table 7.--*Summary of the routes that may be used to visit stations in Puerto Rico and the*

*U.S. Virgin Islands*

[C, crest-gage stage-discharge station; Q, water-quality monitor;  
 U, undefined uncertainty curve for this station, null station;  
 W, ground-water observation well]

Route	Stations serviced on the route--(third to the eighth digit of station number used)							
1S	027750	028400						
2S	035000	W70	W135					
3A	031200	038320						
4S	039500	046000	W70					
5A	061800	U063440						
6A	U063500	065500						
7A	067000	075000						
8S	039500	W70						
9S	031200	035000						
10S	038320	039500	W70					
11S	046000	W70						
12S	035000	039500	W70					
13S	031200	039500	W70					
14S	061800	U063440	U063500					
15S	U063440	U063500						
16S	065500	067000						
17S	U063440	U063500	075000					
18S	065500	075000						
19S	061800	075000						
20S	067000	071000						
21A	U055650	071000						
22S	056400	U055650	W96					
23A	056400	U056900						
24A	U056900	057000						
25S	055000	U056900						
26A	055000	057000						
27S	U053050	055000						
28A	U051150	U051180	U053050					
29S	U053050	U051180						
30S	056400	071000						
31S	043000	U053050						
32S	U051150	U053050						
33A	050900	051310						
34S	U051150	051310						
35S	050900	U051150						
36S	051310	055000						
37S	050900	055000						
38S	U051150	055000						
39S	U051180	055000						
40A	092000	W6	W96					
41S	043000	W6	W87					
42S	043000	U106500	W87					
43S	U106500	U108000	W87					
44S	U106500	U108000	U11500	114000	115000	124200	U129900	136000
	138000	W87	W132	W141	W143	C114400	C115900	
45S	U010600	U011200	U011400	U014800	U015700	027750	028000	028400
	144000	147800	W135					

Table 7.--Summary of the routes that may be used to visit stations in Puerto Rico and the U.S. Virgin Islands--Continued

Route	Stations serviced on the route--(third to the eighth digit of station number used)							
46S	U051150	U051180	051310					
47S	U051150	U051180	055000					
48S	U108000	W6	W87	W96				
49A	U010600	U011200	U014800	U015700	027750	028400	147800	W135
	W70							
50A	U010600	U011200	U014800	U015700	027750	028400	144000	147800
	W135	W70						
51A	U108000	U111500	124200	U129900	136000	W87	W132	W141
	W143	C114400	C115900					
52A	252000	276000	295000	345000				
53S	027750	028400						
54A	039500	046000						
55S	038320	039500						
56S	035000	039500						
57S	031200	039500						
58A	U055650	056400						
59S	043000	U106500						
60S	U106500	U108000						
61A	U106500	U108000	U111500	114000	115000	124200	U129900	136000
	138000							
62A	U010600	U011200	U011400	U014800	U015700	027750	028000	028400
	144000	147800						
63B	U011400	Q011400						
64B	028000	Q028000						
65B	031200	Q031200						
66B	038320	Q038320						
67B	039500	Q039500						
68B	043000	Q043000						
69B	046000	Q046000						
70B	055000	Q055000						
71B	063800	Q063800						
72B	071000	Q071000						
73B	092000	Q092000						
74B	U106500	Q106500						
75B	114000	Q114000						
76B	115000	Q115000						
77B	138000	Q138000						
78B	144000	Q144000						
79S	U010600							
80S	U011200							
81S	U011400							
82S	U014800							
83S	U015700							
84S	027750							
85S	028000							
86S	028400							
87S	031200							
88A	035000							

Table 7.--*Summary of the routes that may be used to visit stations in Puerto Rico and the U.S. Virgin Islands--Continued*

Route	Stations serviced on the route--(third to the eighth digit of station number used)
89A	038100
90S	038320
91S	039500
92A	043000
93S	046000
94S	050900
95S	U051150
96S	U051180
97S	051310
98S	U053050
99S	055000
100S	U055650
101S	056400
102S	U056900
103S	057000
104S	061800
105S	U063440
106S	U063500
107S	063800
108S	065500
109S	067000
110S	071000
111S	075000
112A	092000
113S	U106500
114S	U108000
115S	U111500
116A	112500
117S	114000
118S	115000
119S	124200
120S	U129900
121S	136000
122S	138000
123S	144000
124S	147800
125S	252000
126S	276000
127S	295000
128S	345000

the water year). It is estimated that visits to each gage are required about every other month, just to maintain the equipment. Therefore, unless a more stringent requirement exists, a minimum of six visits during the 12-month period are specified for all gages, except for the water-quality sampling sites.

Practical routes to service the 76 stations were determined after consultation with personnel responsible for maintaining the stations and with consideration of the uncertainty functions and minimum visit requirements. A total of 128 routes were selected to service all the stream gages in Puerto Rico and in the U.S. Virgin Islands. These routes included all possible combinations that describe the current operating practice, alternatives that were under consideration as future possibilities, routes that visited certain key stations, and combinations that grouped proximate gages where the levels of uncertainty indicated more frequent visits might be useful.

The costs associated with the practical routes are divided into three categories. Those categories are fixed costs, visit costs, and route costs and are defined in the following paragraphs. Overhead is, of course, added to the total of these costs.

Fixed costs typically include charges for equipment rental, batteries, electricity, data processing and storage, maintenance, and miscellaneous supplies, in addition to supervisory charges and the costs of computing the record. Average values for Puerto Rico and the U.S. Virgin Islands generally were applied to individual stations. However, costs of record computation and supervision form a large percentage of the cost at each gaging station and can vary widely. These, as well as unusual equipment costs, were determined on a station-by-station basis from past experience.

Visit 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, depending on the difficulty of the measurement and the size of the channel. Average visit times were estimated for each station based on historical operations. This time was then multiplied by the average hourly salary of the hydrographers in Puerto Rico and the U.S. Virgin Islands to determine total visit costs.

Route costs include vehicle use, time spent servicing equipment, cost of the hydrographer's time while in transit, and any travel expenses. The fixed costs were computed on an annual basis, but the visit and route costs are only applied when a trip is made.

## Results

The "Traveling Hydrographer Program" uses the uncertainty functions along with the appropriate cost data, route definitions, and minimum visit constraints to optimize the operation of the stream-gaging program. The objective function in the optimization process is the sum of the variances of the errors of instantaneous discharge (in percent) for the entire gaging station network.

The current practices were simulated to define the total uncertainty associated with present practice. This was done by restricting the specific routes and number of visits to each stream gage to those now being used. This was done only to compute the standard errors of present practice; no optimization was done. The restrictions were then released and the model was allowed to define optimal visit schedules for the current budget. The optimization procedure was repeated for other possible budgets. The results for both the present operation and the optimal solutions are shown in figure 8 and presented in table 8.

The analysis was repeated for each budget under the assumption that no stage record was lost. Those results, labeled "Without missing record" in figure 8, show the average standard errors of estimate for instantaneous discharge attainable if perfectly reliable systems were available to measure and record stage for the present and other budgets. It also shows the error (or accuracy) that occurs when the gages are operating properly, which is about 94 percent of the time.

Assumptions made in the model need to be kept in mind when interpreting these results. Residuals about the ratings for 34 of the 50 stations in the surface-water network were judged to follow the first-order Markovian process assumed in the model. The remaining 16

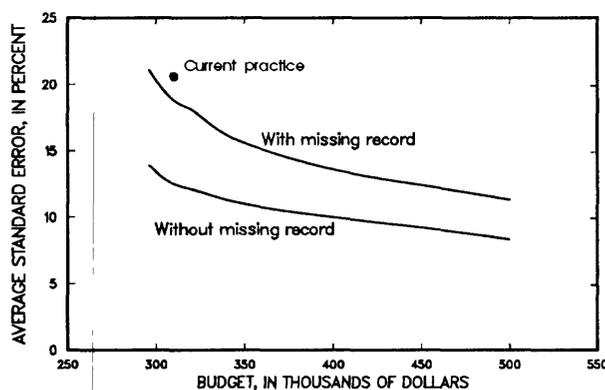


Figure 8.--Average standard error per gaging station as a function of budget.

Table 8.--Selected results of K-CERA analysis given in standard error of instantaneous discharge, in percent; equivalent Gaussian Spread, in brackets; and number of visits per year to site, in parentheses

Budget, in thousands of 1984 dollars										
Station number	Current operations									
	310	296	300	310	320	340	350	400	450	500
Average per station	20.6	21.1	20.3	18.8	18.1	16.2	15.6	13.6	12.4	11.3
50027750	12.8 [5.3] (12)	17.9 [7.6] (6)	17.9 [7.6] (6)	16.6 [7.0] (7)	15.6 [6.5] (8)	14.0 [5.8] (10)	12.8 [5.3] (12)	10.8 [4.4] (17)	9.5 [3.9] (22)	8.6 [3.5] (27)
50028000	16.9 [5.7] (12)	16.9 [5.7] (12)	16.9 [5.7] (12)	16.9 [5.7] (12)	18.4 [6.4] (10)	15.7 [5.3] (14)	15.2 [5.0] (15)	12.4 [4.0] (23)	11.0 [3.6] (29)	10.1 [3.3] (35)
50028400	11.5 [5.5] (12)	15.2 [6.0] (6)	15.2 [6.0] (6)	14.3 [5.9] (7)	13.6 [5.8] (8)	12.4 [5.6] (10)	11.5 [5.5] (12)	10.1 [5.3] (17)	9.2 [9.2] (22)	8.5 [5.1] (27)
50031200	12.9 [3.9] (12)	15.6 [4.9] (8)	14.8 [4.6] (9)	13.4 [4.1] (11)	12.4 [3.8] (13)	10.6 [3.2] (18)	10.3 [3.1] (19)	8.4 [2.5] (29)	7.5 [2.2] (37)	7.1 [2.1] (41)
50035000	17.5 [9.1] (12)	21.2 [11.5] (8)	20.1 [10.7] (9)	18.3 [9.5] (11)	17.5 [9.1] (12)	14.8 [7.5] (17)	14.1 [7.1] (19)	11.7 [5.8] (28)	10.8 [5.4] (33)	9.9 [4.9] (39)
50038100	15.3 [9.2] (12)	15.3 [9.2] (12)	15.3 [9.2] (12)	15.3 [9.2] (12)	15.3 [9.2] (12)	13.4 [7.9] (16)	13.0 [7.7] (17)	10.4 [6.0] (27)	9.6 [5.6] (32)	8.5 [4.9] (41)
50038320	20.8 [11.0] (12)	22.7 [12.2] (10)	21.7 [11.5] (11)	18.7 [9.7] (15)	18.1 [9.4] (16)	15.2 [7.7] (23)	14.6 [7.3] (25)	12.4 [6.2] (35)	10.8 [5.4] (46)	10.3 [5.1] (51)
50039500	21.3 [13.3] (12)	19.8 [12.2] (14)	18.6 [11.4] (16)	16.7 [10.1] (20)	15.6 [9.3] (23)	13.5 [8.0] (31)	12.9 [7.7] (34)	11.0 [6.4] (47)	10.1 [5.9] (56)	9.1 [5.3] (70)
50043000	31.8 [13.6] (12)	27.7 [11.5] (16)	26.9 [11.2] (17)	23.2 [9.4] (23)	22.3 [9.0] (25)	19.2 [7.6] (34)	18.7 [7.5] (36)	15.4 [6.2] (53)	13.8 [5.5] (67)	12.6 [5.1] (80)
50046000	30.6 [19.0] (12)	26.7 [16.4] (16)	25.3 [15.4] (18)	22.0 [13.3] (24)	21.2 [12.7] (26)	17.6 [10.4] (38)	16.7 [9.9] (42)	14.1 [8.3] (59)	12.7 [7.4] (74)	11.7 [6.8] (87)

Table 8.--Selected results of K-CERA analysis given in standard error of instantaneous discharge, in percent equivalent Gaussian Spread, in brackets; and number of visits per year to site, in parentheses--Continued

Station number	Budget, in thousands of 1984 dollars									
	Current operations									
	310	296	300	310	320	340	350	400	450	500
50050900	18.0 [10.2] (12)	20.6 [11.9] (9)	18.8 [10.7] (11)	16.8 [9.4] (14)	15.7 [8.8] (16)	13.5 [7.4] (22)	12.7 [6.9] (25)	10.8 [5.8] (35)	9.7 [5.3] (43)	8.9 [4.8] (52)
50051310	17.8 [13.4] (12)	19.4 [14.8] (10)	18.5 [14.0] (11)	16.5 [12.2] (14)	16.5 [12.2] (14)	13.3 [9.6] (22)	13.0 [9.4] (23)	11.4 [8.2] (30)	10.2 [7.3] (38)	9.3 [6.6] (46)
50055000	20.7 [8.1] (12)	21.6 [8.4] (11)	20.7 [8.1] (12)	18.1 [7.0] (16)	17.6 [6.8] (17)	14.6 [5.5] (25)	14.0 [5.3] (27)	11.6 [4.4] (40)	10.5 [4.0] (49)	9.6 [3.6] (59)
50056400	20.8 [7.9] (12)	21.6 [8.3] (11)	20.8 [7.9] (12)	18.1 [6.7] (16)	17.6 [6.5] (17)	14.9 [5.5] (24)	14.0 [5.1] (27)	12.0 [4.4] (37)	10.6 [3.8] (48)	9.7 [3.5] (57)
50057000	29.2 [20.9] (12)	25.4 [17.8] (16)	24.0 [16.7] (18)	21.8 [15.0] (22)	20.4 [14.0] (25)	17.4 [11.8] (35)	16.5 [11.1] (39)	14.0 [9.4] (55)	12.2 [8.2] (73)	11.4 [7.7] (84)
50061800	26.9 [12.2] (12)	26.1 [12.0] (13)	25.3 [11.9] (14)	22.8 [11.4] (18)	22.4 [11.3] (19)	18.9 [10.4] (29)	18.1 [10.2] (32)	15.3 [9.1] (48)	13.6 [8.3] (63)	12.6 [7.9] (74)
50063800	23.4 [12.4] (12)	20.5 [10.7] (8)	19.4 [10.1] (9)	16.8 [8.7] (12)	16.8 [8.7] (12)	13.8 [7.1] (18)	13.2 [6.7] (20)	11.0 [5.6] (29)	10.0 [5.0] (35)	9.4 [4.7] (40)
50065500	20.5 [15.6] (12)	21.4 [16.3] (11)	20.5 [15.6] (12)	17.4 [12.8] (17)	17.4 [12.8] (17)	14.1 [10.2] (26)	13.4 [9.7] (29)	11.6 [8.3] (39)	10.8 [7.7] (45)	9.9 [7.0] (54)
50067000	21.9 [14.9] (12)	22.8 [15.6] (11)	21.1 [14.4] (13)	18.6 [12.6] (17)	18.6 [12.6] (17)	15.5 [10.3] (25)	14.7 [9.7] (28)	12.0 [7.9] (42)	11.2 [7.2] (49)	10.2 [6.6] (59)
50071000	20.9 [13.2] (12)	21.8 [13.9] (11)	20.9 [13.2] (12)	18.2 [11.3] (16)	18.2 [11.3] (16)	15.3 [9.3] (23)	14.7 [8.9] (25)	12.0 [7.2] (38)	10.9 [6.6] (47)	9.9 [5.9] (56)
50075000	15.2 [6.8] (12)	18.2 [7.9] (8)	17.3 [7.6] (9)	15.2 [6.8] (12)	15.2 [6.8] (12)	12.7 [5.9] (18)	12.1 [5.6] (20)	10.2 [4.8] (29)	8.9 [4.3] (38)	8.3 [4.0] (44)
50092000	32.4 [27.2] (12)	32.4 [27.2] (12)	30.2 [25.1] (14)	26.1 [21.4] (19)	25.5 [20.8] (20)	21.3 [17.1] (29)	20.2 [16.2] (32)	17.1 [13.5] (45)	15.2 [12.0] (57)	14.7 [11.6] (61)

Table 8.--Selected results of K-CERA analysis given in standard error of instantaneous discharge, in percent equivalent Gaussian Spread, in brackets; and number of visits per year to site, in parentheses--Continued

Station number	Budget, in thousands of 1984 dollars									
	Current operations									
	310	296	300	310	320	340	350	400	450	500
50112500	40.1 [39.1] (12)	36.2 [35.0] (15)	34.0 [32.8] (17)	30.0 [28.6] (22)	29.3 [28.0] (23)	24.4 [23.1] (33)	23.4 [22.0] (36)	19.2 [18.0] (53)	17.2 [16.0] (66)	15.8 [14.7] (78)
50114000	21.9 [19.2] (12)	25.9 [23.1] (8)	25.9 [23.1] (8)	22.8 [20.0] (11)	19.8 [17.2] (15)	18.2 [15.7] (18)	17.4 [14.8] (20)	14.5 [12.2] (29)	13.0 [10.9] (36)	11.7 [9.7] (45)
50115000	25.3 [19.7] (12)	29.0 [23.2] (9)	26.4 [20.6] (11)	24.3 [18.8] (13)	22.7 [17.4] (15)	19.8 [14.8] (20)	18.8 [14.1] (22)	15.5 [11.4] (33)	13.9 [10.2] (41)	12.8 [9.3] (49)
50124200	28.5 [26.2] (12)	31.1 [28.9] (10)	29.7 [27.5] (11)	26.4 [24.2] (14)	25.6 [23.3] (15)	21.6 [19.4] (21)	20.7 [18.5] (23)	16.8 [14.8] (35)	15.2 [13.3] (43)	14.4 [12.5] (48)
50136000	19.1 [10.3] (12)	26.1 [14.7] (6)	24.4 [13.6] (7)	23.0 [12.7] (8)	17.2 [9.1] (15)	18.4 [9.9] (13)	17.8 [9.5] (14)	14.6 [7.7] (21)	13.0 [6.7] (27)	12.1 [6.2] (31)
50138000	20.6 [9.5] (12)	24.9 [11.9] (8)	24.9 [11.9] (8)	23.6 [11.2] (9)	18.5 [8.4] (15)	19.1 [8.8] (14)	18.5 [8.4] (15)	15.0 [6.7] (23)	13.7 [6.1] (28)	12.4 [5.5] (34)
50144000	14.5 [6.4] (12)	15.8 [7.0] (10)	15.8 [7.0] (10)	15.8 [7.0] (10)	17.6 [7.8] (8)	15.1 [6.7] (11)	14.0 [6.1] (13)	11.6 [5.1] (19)	10.6 [4.6] (23)	9.8 [4.2] (27)
50147800	24.6 [8.2] (12)	28.2 [9.8] (9)	26.9 [9.2] (10)	24.6 [8.2] (12)	23.7 [7.9] (13)	19.7 [6.4] (19)	18.8 [6.1] (21)	15.8 [5.0] (30)	14.1 [4.5] (38)	13.0 [4.1] (45)
50252000	26.7 [14.4] (6)	26.7 [14.4] (6)	24.8 [13.2] (7)	24.8 [13.2] (7)	23.3 [12.3] (8)	20.0 [10.4] (11)	19.1 [9.9] (12)	15.7 [8.1] (18)	14.0 [7.2] (23)	13.0 [6.7] (27)
50276000	27.9 [14.7] (6)	25.9 [13.4] (7)	25.9 [13.4] (7)	24.3 [12.4] (8)	23.0 [11.8] (9)	20.8 [10.4] (11)	19.2 [9.6] (13)	15.9 [7.9] (19)	14.5 [7.1] (23)	13.6 [6.7] (26)
50295000	46.7 [44.0] (6)	46.7 [44.0] (6)	46.7 [44.0] (6)	46.7 [44.0] (6)	46.0 [43.6] (7)	45.5 [43.3] (8)	45.0 [43.1] (9)	43.0 [41.8] (16)	41.1 [40.4] (27)	37.1 [36.7] (59)
50345000	47.1 [44.1] (6)	47.1 [44.1] (6)	47.1 [44.1] (6)	47.1 [44.1] (6)	46.4 [43.7] (7)	45.3 [43.1] (9)	44.5 [42.7] (11)	42.0 [41.0] (22)	37.9 [37.4] (34)	33.4 [33.0] (95)

stations were retained in the analysis. This was done in the belief that, while the absolute values of standard error may be incorrect, the values had relative significance.

The current operating policy has resulted in an average standard error of estimate of streamflow of about 20.6 percent. This policy is based on a budget of \$310,000 to operate the 50-station stream-gaging program. This program provides for 12 measurements per year for most stations, except for five stations which are measured six times per year. Four of the latter are in the U.S. Virgin Islands.

For periods with missing record, the optimum standard error of estimate of streamflow is about 18.8 percent. Under optimum conditions, stream-gaging sites are visited and measured 6 to 25 times per year. These include 28 sites which are measured less than 11 times, and 10 sites measured more than 15 times per year.

For periods without missing record, the standard error of estimate of streamflow is about 12.5 percent. For optimum operation, from 6 to 35 stations are visited per year. These include 27 sites measured six times and four sites measured 25, 26, 27, and 35 times per year. Of the sites measured six times per year, 16 sites are ones for which uncertainty functions could not be defined. The logistics of either of these later operations, with or without missing record, is impractical in terms of equipment and manpower.

For periods of missing record, an increase in the operating budget of 12.9 percent (from \$310,000 to \$350,000 of 1984 dollars) results in a standard error of estimate of about 15.6 percent. The measurements per site ranges from 9 to 42 times per year with 24 sites measured less than 15 times per year, and 17 sites measured 21 or more times per year. The 17 sites are measured 21, 23, 25, 26, 27, 28, 32, 34, 36, 39, and 42 times per year. The 16 null sites are measured nine times per year. A more practical visit schedule would range from about 18 to 20 times per year or slightly less than a measurement every 3 weeks. This would result in a standard error of estimate of about 17 percent.

The maximum budget analyzed was \$500,000. The analyses using this budget resulted in an optimum average standard error of estimate of about 11.3 percent.

For the present operational budget of \$310,000, the effects of missing records adds about 6 percent to the average standard error. With a budget of \$350,000, stations would be visited more frequently, and the reduced number of missing records would decrease the average standard errors by about 3 percentage points. Also,

improvements in equipment can have an additional positive effect on uncertainties of instantaneous discharges.

### Summary of Third Phase of Analysis

The following are conclusions from this phase of the analyses:

1. The travel routes and measurement frequencies now in use could be modified in order to decrease the current 20.6 percent per station average standard error by 1.9 percent, given the present budget of \$310,000 1984 dollars. But with present manpower and equipment, only a 1 percent decrease in standard estimate of error is practical based on a compromise modification of present and computed measurement visit frequencies.

2. If the present operating budget were increased by about 12.9 percent (to \$350,000), the average standard error of data would decrease to about 17 percent from the present figure of about 20.6 percent. These figures are obtained using the computed optimum K-CERA station visits as a guide to achieve a practical routing of about 18 to 20 measurements per year for most stations in Puerto Rico and 12 measurements per year at most stations in the U.S. Virgin Islands.

3. Methods for decreasing the probabilities of missing record need to be explored. Missing record presently increases the average standard error by about 6 percentage points or about 30 percent of the present standard error of estimate. The methods of decreasing missing record might include increased measurements per year at each site, improved instrumentation and increased use of local observers and satellite relay of data.

### SUMMARY

Currently, there are 50 continuous-record streamflow sites being operated in Puerto Rico and the U.S. Virgin Islands at a cost of \$310,000 per year. Data from most stations have multiple uses and all of the stations are recommended for continuation.

Two stations (50063440 and 50063500) are used primarily for research and short-term investigations. However, these stations are located in critical tropical hydrologic areas where more data is needed and where additional data may prove useful well beyond the duration of the research projects. Thus, it would be desirable to continue these stations as index or bench mark stations and to establish other similar stations on the main island of Puerto Rico, as well as on Vieques, Culebra, and the U.S. Virgin Islands. The greatest need is for stations located in 1- to 15- square mile drainage basins located far from populated areas. Hydrologic information such as streamflow, quality of water, sediment

discharge, and precipitation are needed at stations such as these.

Flow routing and correlation and regression analyses were found to be unacceptable for estimating discharge. There were no sites at which flow routing could be attempted. Only four sites were identified at which correlation and regression analyses might have possibilities, and only at two sites were fair measurements of regression accuracy obtained. At these two sites, the best results were obtained by grouping the data into dry (November through May) and rainy (May through November) seasons where the simulated flow is within 15 percent of the measured discharge a significant percent of the time.

The current policy for operating the 50-station program requires a budget of \$310,000 per year. The travel routes and measurement frequencies now in use can be modified to decrease the present 20.6 percent standard error of estimate by 1.9 percent, while maintaining the

present budget. However, the number of discharge measurement visits required to obtain the 1.9 percent increase in accuracy is not practical with existing personnel and equipment. Discharge measurement site visits might be modified to decrease the standard error of estimate by about one percent, using present personnel and equipment.

A budget increase of 12.9 percent (to \$350,000 1984 dollars) and a modification of the measurements schedule at continuous stream-gaging sites (to 18 to 20 times per year) could result in a reduction of about 17.5 percent in the standard estimate of error (from 20.6 to about 17 percent).

Future studies of the stream-gaging program need to examine ways to decrease the probabilities of missing record and reduce the standard error of estimate with moderate increases in personnel and funding. Analyses similar to this one would be beneficial if repeated approximately every 10 to 15 years.

## SELECTED REFERENCES

- Benson, M.A., and Carter, R.W., 1973, A national study of the streamflow data collection program: U.S. Geological Survey Water-Supply Paper 2028, 44 p.
- Carter, R.W., and Benson, M.A., 1970, Concepts for the design of streamflow data programs: U.S. Geological Survey Open-File Report, 33 p.
- Curtis, R.E., Jr., Guzmán-Ríos, Senén, Díaz, P.L., 1985, Water Resources Data for Puerto Rico and the U.S. Virgin Islands, Water Year 1984: U.S. Geological Survey Water-Data Report PR 84-1, 374 p.
- Draper, N.R., and Smith, H., 1966, Applied regression analysis (second ed.): New York, John Wiley and Sons, Inc., 709 p.
- Ezekiel, E., Fox, K.A. 1930, Methods of correlation and regression analysis (third ed.): New York, John Wiley and Sons, Inc., 548 p.
- Fontaine, R.A., Moss, M.E., Smith, J.S., and Thomas, W.O., Jr., 1984, Cost-effectiveness of the stream-gaging program in Maine: U.S. Geological Survey Water-Supply Paper 2244, 39 p.
- Gelb, A., ed., 1974, Applied optimal estimation: Cambridge, Massachusetts, 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, 38 p.
- Hale, T.W., Stokes, W.R., III, Price, M., and Pearman, J.L., 1985, Cost-effectiveness of the stream-gaging program in Georgia: U.S. Geological Survey Water-Resources Investigations Report 84-4109, 144 p.
- Hirsch, R.M., 1982, A comparison of four streamflow record extension techniques: Water Resources Research, v. 18, no. 4, p. 1081-1088.
- Hutchison, N.E., 1975, WATSTORE User's guide, volume 1: U.S. Geological Survey Open-File Report 75-426, 315 p.
- Kleinbaum, D.G., and Kupper, L.L., 1978, Applied regression analysis and other multivariable methods: North Scituate, Massachusetts, Duxbury Press, 556 p.
- López, M.A., and Fields, F.K., 1970, A proposed streamflow-data program for Puerto Rico: U.S. Geological Survey Open-File Report, 39 p.
- López, M.A., Colón-Dieppa, Eloy, and Cobb, E.D., 1979, Floods in Puerto Rico, magnitude and frequency: U.S. Geological Survey Water-Resources Investigations Report 78-141, 66 p.
- McKinley, P.W., 1985, Surface-water data network for Puerto Rico: U.S. Geological Survey Water-Resources Investigations Report 83-4055, 14 p.
- Moss, M.E., and Gilroy, E.J., 1980, Cost-effective stream-gaging strategies for the Lower Colorado River Basin: U.S. Geological Survey Open-File Report 80-1048, 111 p.
- Riggs, H.C., 1973, Regional analysis of streamflow characteristics: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chapter B3, 15 p.
- 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.