

# A National Study of the Streamflow Data-Collection Program

By MANUEL A. BENSON *and* ROLLAND W. CARTER

---

GEOLOGICAL SURVEY WATER-SUPPLY PAPER 2028



**UNITED STATES DEPARTMENT OF THE INTERIOR**

**ROGERS C. B. MORTON, *Secretary***

**GEOLOGICAL SURVEY**

**V. E. McKelvey, *Director***

Library of Congress catalog-card No. 73-600157

---

For sale by the Superintendent of Documents, U.S. Government Printing Office  
Washington, D.C. 20402 — Price 45 cents domestic postpaid or 30 cents GPO Bookstore  
Stock Number 2401-02372

## CONTENTS

---

	Page
Abstract .....	1
Introduction .....	1
Acknowledgments .....	3
Framework for program design .....	3
Data for current use .....	4
Data for planning and design .....	6
Natural-flow streams .....	6
Minor streams .....	7
Principal streams .....	7
Regulated-flow streams .....	8
Data to define long-term trends .....	9
Data on the stream environment .....	9
Goals of the streamflow data program .....	9
Data for current use .....	10
Data for planning and design .....	10
Examples of accuracy standards .....	11
Data to define long-term trends .....	14
Data on the stream environment .....	14
Evaluation of available data .....	14
Data for current use .....	14
Data for planning and design .....	17
Natural-flow minor streams .....	17
Streamflow characteristics .....	18
Basin characteristics .....	19
Regression procedures .....	21
Results of regression analyses .....	21
Natural-flow principal streams .....	25
Regulated-flow streams .....	26
Data to define long-term trends .....	26
Data on the stream environment .....	27
Planning future streamflow data programs .....	27
Identification of gaging-station functions .....	29
Summary .....	30
References cited .....	33

## ILLUSTRATIONS

---

	Page
<b>FIGURE</b> 1. Graph showing relation of standard error to length of record in Minnesota .....	13
2. Graph showing standard error of Potomac River basin relations .....	23

	Page
FIGURES 3-11. Map showing—	
3. Ten-year accuracy goals for mean annual flow -----	36
4. Standard errors of estimate by regression for mean annual flow -----	37
5. Standard errors of estimate by regression for standard deviation of mean annual flow -----	38
6. Average 10-year accuracy goals for monthly means -----	39
7. Average standard errors of estimate by regression for monthly means -----	40
8. Ten-year accuracy goals for 50-year flood ---	41
9. Standard errors of estimate by regression for 50-year flood -----	42
10. Ten-year accuracy goals for 7-day 50-year flood volume -----	43
11. Standard errors of estimate by regression for 7-day 50-year flood volume -----	44

---

## TABLES

---

	Page
TABLE 1. Framework for design of data-collection program --	5
2. Accuracy standards for South Carolina and Arizona --	12
3. Complete-record stream-gaging stations, by district offices, that serve a current-use function as of fiscal year 1970 -----	16
4. Rank of variables in explaining streamflow characteristics -----	25
5. Present and proposed network of continuous-record stations, with primary purpose of stations in the proposed network -----	31

# **A NATIONAL STUDY OF THE STREAMFLOW DATA-COLLECTION PROGRAM**

---

By **MANUEL A. BENSON** and **ROLLAND W. CARTER**

---

## **ABSTRACT**

The streamflow data program of the U.S. Geological Survey was evaluated in a nationwide study during 1970. The principal elements of the study were (1) establishing the objectives and goals of the program, (2) analyzing all available data to determine which of the goals have already been met, (3) considering alternate means of meeting the remaining goals, and (4) identifying the elements which should be included in the future program.

The objective of the streamflow data program is to provide information on flow characteristics at any point on any stream. Program goals are related to four classifications of data: (1) Data for current use, (2) data for planning and design, (3) data for definition of long-term trends, and (4) data on the stream environment.

Analysis of the present data base shows that in the eastern part of the country, adequate data for planning and design can be provided for natural-flow streams by relating flow characteristics to the physical and climatic characteristics of drainage basins. Similar results were not obtained for the western part of the country.

More than half of the 8,300 streamflow stations now in operation are needed to provide current data for water management, flow forecasting, pollution control, or appraisal of current water conditions. These stations must be continued in operation as long as the need exists.

The present program is not providing homogeneous data for regulated streams. It is recommended that data for these streams be derived from mathematical models of stream systems that reflect the effect of storage, diversions, and management practices. Streamflow records can be used as input to the model.

According to the concepts used in the study, about 10 percent of the present streamflow program should be redirected to areas of higher priority.

## **INTRODUCTION**

During 1970 the U.S. Geological Survey initiated a nationwide study to evaluate all phases of its data-collection program. This program includes data on surface waters—streams, lakes, ponds, reservoirs, and estuaries; on ground-water availability; and on water quality. Because streamflow data are most numerous and the needs for streamflow data are well identified, the first phase

of the study was devoted exclusively to the evaluation of the streamflow data program.

Any program of long standing, whether it be a data-collection program, a manufacturing program, a management program, or other enterprise, should be reexamined at intervals, if not continuously. This is because objectives may change, technology surely changes, and external constraints and needs also change. No program adjusts itself automatically to such continually altered considerations. Programs must be periodically examined, evaluated, and revised to conform to current or projected conditions. Only in this way can a program be maintained which is most efficient and most effective in meeting the real needs of the moment.

It was in this light that the examination of the streamflow data program was made. The accelerated need for information, coupled with advances in water science and newly developed concepts of large-scale water management, argued for an improved approach to the planning of streamflow data programs. The use of high-speed computers made possible massive data handling, data storage, and data analysis that were a necessary part of this study. The objectives of the program were reviewed and stated as specific goals. All available data were examined to determine which goals had already been met. Where goals were found not to have been met, consideration was given to alternate means for meeting them, and finally, recommendations were made for a revised program in which all elements were identified as to their role in achieving the goals.

This report deals only with streamflow. Consideration is given, however, to the interrelation of streamflow to other aspects, such as the need for streamflow information in combination with ground-water information and for interpretation of observations of water quality.

Historically, streamflow data programs have developed in the Survey in response to local economic and hydrologic stimuli. Owing to their joint concern, other Federal agencies, the States, and many counties and municipalities have for 70 years contributed substantial funds to the Survey to obtain data directed to specific problems and also to intensify the general inventory of water resources. Although the program has largely evolved in response to specific and local area needs rather than by broadscale national planning, a wealth of information on streamflow has been accumulated during the past 70 years.

One of the chief advances in water science has been the development of methods of synthesizing data which greatly facilitates the definition of the effect of the environment on flow characteristics

and the statistical definition of flow distribution in time. These advances, coupled with the availability of digital computers, provide a more efficient means of utilizing observations of streamflow in hydrologic analysis which, in turn, governs to some degree the need for collection of data. Streamflow data programs should no longer be considered in terms of a network of observation points, but rather in terms of an information system in which data are provided both by observation and synthesis.

Ideally, the planning of data programs would be based on a complete knowledge of the needs for different kinds of data and the sensitivity of design decisions to the input of information on streamflow. This knowledge could be accumulated by examining models of river basin planning to test the response of optimum design to various levels of streamflow information. The value of data to project design could be determined by measuring the benefits foregone due to missing the optimum design through inadequate data or to the penalty cost of strategies utilized to meet uncertainties. Eventually, studies of this type will provide an improved scientific basis for design of data programs. In the meantime, planning can be based only in part on rigorous analysis, and must be supplemented by accumulated experience and judgment.

### ACKNOWLEDGMENTS

The general procedures for making this study, for setting goals, for evaluation of the available data, for codifying the purposes of stream gaging, and for determining the desirable program of the future were described by Carter and Benson (1969) in a report intended as a guideline to those carrying out the study. The actual work was done in 46 district offices of the Survey, which included all except the district office in Hawaii. The following acted as coordinators and advisers to the individual district offices: J. F. Bailey, Arlington, Va.; H. H. Barnes, M. A. Benson, R. W. Carter, C. H. Hardison, and D. M. Thomas, Washington, D.C.; H. H. Hudson, C. C. McDonald, and M. S. Peterson, Denver, Colo.; H. F. Matthai, Menlo Park, Calif.; and J. P. Monis, R. H. Tice, and I. D. Yost, St. Louis, Mo.

### FRAMEWORK FOR PROGRAM DESIGN

The general objective of the streamflow data program is to provide whatever information may be required by users of water data, at any point on any stream within the United States. This means that information is to be supplied for all sizes of streams, under natural or regulated conditions.

There are many different uses for water data in various fields, such as water supply, water power, irrigation, flood control, and industry. However, for all uses the types of information required may be classified into two distinct categories. The first category of information is that required for current use, and the second is that required for planning and design.

For purposes of data-collection program planning, two other classifications are used—specialized data collected to define long-term trends, which is actually a subclassification of data for planning and design, and data on stream environment, which consists of several subtypes, each of which is used for either current use purposes or for planning and design. This classification system is used in order to categorize all types of stream-related data other than discharge data.

Concepts of program design are presented with reference to the general framework shown in table 1. Streamflow data are classified into four types: (1) Data for current use, (2) data for planning and design, (3) data to define long-term trends, and (4) data on stream environment. For the second type of data, streams are classified as natural or regulated; each of these classifications is further subdivided into principal or minor, with the separation of the two at a drainage area of 500 square miles. The need for each type of data and the methods of obtaining it are described in the following sections.

#### DATA FOR CURRENT USE

Streamflow data are needed at many sites on a day-to-day basis for the management of water, for the assessment of current water availability, for the control of water quality, for the forecast of flow extremes, and for the surveillance necessary for legal requirements. This classification represents the need for information on the actual flow at any moment, or during any specified day, week, month, or year.

Streamflow data obtained for current use have a high payoff value, as a current knowledge of the rate of flow and storage at different points in the system provides a basis for water-management decisions that govern the economic efficiency of the operation.

Current-purpose data are obtained by operating gaging stations to obtain the data specifically required by water-management systems. Current-purpose data stations are placed in a separate category because (1) justification can be related to specific needs, (2) the data may have limited transfer value in a hydrologic



TABLE 1.—*Framework for design of data-collection program*

[Drainage area limits may be varied with terrain and hydrologic conditions]

	Current use	Planning and Design				Long-term trends	Stream environment
		Natural Flow		Regulated Flow			
		Minor streams	Principal streams	Minor streams	Principal streams		
Goals	To provide current data on streamflow needed for day-by-day decisions on water management as required.	To provide information on statistical characteristics of flow at any site on any stream to the specified accuracy.				To provide a long-term data base of homogeneous records on natural flow streams.	To describe the hydrologic environment of stream channels and drainage basins.
Drainage area limits.	None -----	Less than 500 sq mi.	Greater than 500 sq mi.	Less than 500 sq mi.	Greater than 500 sq mi.	None -----	None.
Accuracy	As required -----	Equivalent to 10 years of record.	Equivalent to 25 years of record.	Equivalent to 10 years of record.	Equivalent to 25 years of record.	Highest obtainable.	As required.
Approach	Operate gaging stations as required to provide specific information needed.	Relate flow characteristics to drainage basin characteristics using data for gaged basins.	Operate gaging stations to obtain 25 years of record (or the equivalent by correlation at a network of points on principal streams; interpolate between points.	Develop gaging stations that account for the effect of storage, diversion or regulation on natural flow characteristics.	Utilize analytical stream system with observed data as input to compute homogeneous records for both natural flow conditions and present development.	Operate a number of carefully selected gaging stations indefinitely.	Observe and publish information on stream environment.
Evaluate available data.	Identify stations where data is used currently and code the specific use of data.	Develop relationship for each flow characteristic and compare standard error with accuracy goal. Evaluate sample.	Lay out network of points on principal streams and compare data available at these points with goal.	Appraise type of regulation data available, and areas where relationships are needed.	Identify stream systems that should be studied using model approach and determine data requirements.	Select two stations in each WRC subregion to operate in definitely for this purpose.	Evaluate information available in relation to goals.
Design future program.	Identify goals that have not been attained. Consider alternate means of attaining goals. Identify elements of future program.						

sense, and (3) the location of the stations and the periods of operation are specified by the user of the data, who usually provides the financing.

#### DATA FOR PLANNING AND DESIGN

Designers and planners of water-control and water-related facilities increasingly utilize the statistical characteristics of streamflow rather than flow of specific periods in the past. The probability that the historical sequence of flow history observed at a given site will occur again is remote, and estimates of future flows needed in design and planning must consider all probable flows and sequences of flow. The need is to consider what may be expected to happen in the future, not in terms of specific events, but in terms of probability of occurrence over a span of years. For example, many highway bridges are designed on the basis of the flood that will be exceeded on the average of once in 50 years. Storage reservoirs can be designed on the basis of the probability of deficiency of storage for a given draft rate. The water available for irrigation, dilution of waste, or other purposes, may be stated in terms of the mean flow, or probability of flow magnitudes for periods of a year, season, month, week, or day. In addition, there is a marked trend toward simulation of streamflow data based on the statistical characteristics, such as the mean, standard deviation, and skew.

A record of streamflow of at least 25 years is the best basis for defining the statistical characteristics. It is not feasible to collect such records at every site where information may be needed, because there is an infinite number of such sites on all the streams in the country. The stream gaging that is done to provide information for planning and design must be considered as a representative sampling of these sites that is required to provide information for transfer to ungaged sites, or to sites where a small amount of streamflow data are available.

#### NATURAL-FLOW STREAMS

Transfer of runoff information for natural-flow streams may be accomplished in many ways, varying from extremely simple to extremely complex. Simple methods are interpolation between gaging points on the same stream, and extrapolation on the basis of drainage-area size. Information may be transferred from basins of like hydrologic characteristics. Mapping of gaged data may be used to define approximate lines of equal runoff values. Correlation of short records with long records can be used. A more com-

plex method that has been found most effective is to relate stream-flow characteristics to basin and climatic characteristics that affect streamflow by means of the statistical technique known as multiple-regression analysis.

These methods are not usually applicable to streams where the flow is affected by regulation and diversion. Because different techniques are required to provide information on natural and regulated streams, the two are considered separately in program design.

In setting the goals of a data-collection system, it is necessary to specify not only the scope and types of information that are to be furnished, but also the accuracy of such information. For the purpose of setting accuracy goals, streams are further classified by drainage area as minor streams (less than 500 square miles), and as principal streams (greater than 500 square miles). The intent is to use size of the drainage area as an index of worth of data. More costly water developments can be expected on larger streams, which justify a higher accuracy goal for principal streams than for minor streams. The given point of division between the two classes, 500 square miles, may be modified according to the hydrology of the region. For example, in arid regions the division point may be set at 1,000 square miles, because major developments are not likely on streams smaller than this.

#### MINOR STREAMS

Definition of flow characteristics of minor streams must be by some process of regionalization, because of the large number of such streams in the country. This can be accomplished for natural-flow streams by gaging at sample locations and relating the observed flow characteristics to basin parameters to provide definition for ungaged streams.

#### PRINCIPAL STREAMS

It has been found that techniques of regionalization cannot, in general, be used for principal streams because of higher accuracy requirements. Therefore, the proposed approach is to operate a network of gaging stations at selected locations on principal streams, and by interpolation or systems studies to estimate the flow characteristics at locations between stations. Experience gained heretofore in hydrologic analysis justifies a procedure for defining the network of principal stream stations as follows: (1) Select stations with drainage area of about 500 square miles on the most upstream segment of all streams and (2) select the additional stations on each stream from the upstream station to the

mouth at points where the drainage area has approximately doubled. The drainage area should be more than doubled if another principal stream enters between two principal stream stations.

#### REGULATED-FLOW STREAMS

The natural flow regime of many streams is altered by the construction of storage reservoirs, the return of pumped ground water into streams, and the diversion of water for consumptive use. This increases the scope of both the data collection and analysis that is required to provide information on the flow characteristics.

To be useful in statistical prediction, streamflow data must be homogeneous in time. Frequently, however, it is not possible to obtain a long record under one condition of development before additional changes occur.

Definition of the flow characteristics at any point on any stream is also much more difficult under conditions of regulation. The procedures used for natural streams—such as regression, correlation, and interpolation—cannot be applied.

For regulated streams, a systems approach appears to be the most efficient way of providing meaningful information on the statistical characteristics of flow. This approach requires some sort of analytical model of the stream system. Such models are simple in concept and usually consist of water-budget equations and flow-storage equations. However, in many cases the use of the digital computer is required to handle complex relations, or large volumes of data. A computer program tailored to the individual system can be prepared.

Development of such a model requires information on stage-capacity curves of reservoirs, stage-discharge curves at the outlets, operating-rule curves for the release of water, losses due to evaporation and seepage, the geometry of the stream channel, and records of diversions and return flow. Information on streamflow at some point or points is also needed as input to the model and to verify the output. In some cases aquifer characteristics and ground-water pumpage should be taken into account.

The model and the associated data can be used to derive homogeneous data for either the natural or the regulated condition. All historical streamflow records for both natural and regulated flows could be utilized as input to the model. Furthermore, data could also be derived for ungaged sites in the stream system.

### DATA TO DEFINE LONG-TERM TRENDS

A long continuing series of consistent observations on streamflow is needed (1) for analysis of the statistical structure of the hydrologic time series and (2) as a reference or comparative base for noting changes in the flow regime of streams that become increasingly regulated over a period of time.

Statistical statements on flow characteristics are based on the assumption that the data series is stationary in a statistical sense, and that the observed record is a representative sample of the population of flows. Long-term homogeneous streamflow data would provide a basis for checking these assumptions, and a basis for adjusting flow characteristics from short records to more nearly represent the characteristics of the flow population.

For these purposes the gages for defining long-term trends should be located on streams draining basins that have undergone no significant manmade changes and that are expected to remain in a comparable condition in the future. The gages should be well distributed areally and should be located on basins of different physical characteristics. The number of such gages can be small relative to the total number of gaging points.

### DATA ON THE STREAM ENVIRONMENT

Stream discharge and its use is intimately related to the environmental characteristics of the drainage basin. Environmental data include a wide variety of water-related information other than stream discharge. These data are necessary for hydrologic studies and for planning, designing, and operating systems for controlling water or pollution. For example, data on the geometry of a stream channel are needed to appraise the use of a stream for recreation or to determine its capacity to assimilate waste; profiles of flood elevations are required to determine areas subject to inundation by floods; and information on aquifer characteristics is essential to describe the variability of low flow or in planning the conjunctive use of surface water and ground water.

### GOALS OF THE STREAMFLOW DATA PROGRAM

The general objective of the streamflow data program is to provide information on flow characteristics at any point on any stream. Purposes for which streamflow data are used include the design of water-supply reservoirs, control of pollution, design of highway bridges and culverts, management of flood plains, development of recreation facilities, forecast and management of floods,

production of power, design and maintenance of navigation facilities, and allocation of water for irrigation and other uses. Data for one or more of these purposes will be needed on virtually every stream in the country, and the data system must be designed to produce the information in advance of the need.

The design of the streamflow data program should be based on specific goals that represent the type and accuracy of information that is needed. Acceptable accuracy levels need to be specified, because accuracy levels not only govern the cost and the techniques used in providing information, but also provide a measure of attainment of specific goals. Regional differences in the flow characteristics to be included in these goals may be expected because of variable hydrologic conditions and the need for different kinds of information; however, the general framework and accuracy levels used in this study are considered applicable to the entire country. Setting goals for each of the four types of data are described below.

#### DATA FOR CURRENT USE

The program goal for this type of data is to provide the particular information needed at specific sites for current use. This part of the program is tailored to fit the requirements for data on a current basis as specified by the user of the data. These specifications may include the data to be obtained, the time of reporting, and the accuracy requirements. This part of the program is not subject to advance design because its character changes frequently in response to changing need.

In general, a higher degree of stream-gaging accuracy is justified for current-purpose data used in the operation of water systems than for data to be used in the planning and design of water development projects. In contrast with the wide confidence intervals associated with the probability of future occurrences used in planning and design, the operation of water systems deals with known volumes of water that are subject to control for economic benefit. Accuracy of gaging at a given site depends on the requirements of the particular management system and can be met by intensified observations, or by more sophisticated instrumentation as needed.

#### DATA FOR PLANNING AND DESIGN

The program goal for this type of data is to define with a given accuracy the statistical flow characteristics for all streams in the country. This is indeed the major goal of the surface-water data

program. It includes not only all streams with natural flow, but also those streams that are affected by regulation and diversion. For the latter streams the goal includes definition of the flow characteristics for both natural and present conditions of development.

The statistical characteristics to be defined depend on the hydrology of the region. Typical characteristics are the mean and standard deviation of annual and monthly flow, the 50-year flood, and the 20-year 7-day low flow.

The flow characteristics of ungaged streams must be estimated by some process of regionalization. Accuracy goals proposed for these estimates are given in terms of equivalent years of record. This means it is specified that information provided for any ungaged point on a stream should be equivalent in accuracy to that which would have been attained by an actual record of  $x$  years at that point. For minor streams the accuracy goal is the equivalent of 10 years of record, and for principal streams the equivalent of 25 years of record. The accuracy goals proposed are based on judgment as to the worth of hydrologic data in planning and design and on experience with the approaches that are used to derive information on streamflow by observation or analysis.

Accuracy goals in terms of equivalent years of record can be converted to standard error in percent of mean, using the methods described by Hardison (1969). In summary, the standard error of a characteristic for a given number of years of record depends on the variability of the annual events as defined by their coefficient of variation or by the standard deviation of their logarithms. Using this methodology, the accuracy goals in terms of standard error that correspond to 10 and 25 years of record may be computed for a given State or region by consideration of the average coefficient of variation within the region.

#### EXAMPLES OF ACCURACY STANDARDS

Examples of how the goals were set are shown in table 2 for South Carolina and Arizona, which, respectively, represent humid and semiarid regions. The accuracy goals were computed as the equivalent of 10 years of record for minor streams and 25 years of record for principal streams. The accuracy figures shown represent a conversion from equivalent years of record to standard error in percent of mean.

The accuracy requirements illustrated in table 2 are typical of those for humid and semiarid conditions. The accuracy requirements vary with the variability of flow. Thus, the requirements are higher in the semiarid and arid regions, particularly for mean flows and floodflows. Values such as those shown here or those

TABLE 2.—*Accuracy standards for South Carolina and Arizona*

Streamflow characteristics	Accuracy (standard error in percent of mean)			
	South Carolina		Arizona	
	Minor streams	Principal streams	Minor streams	Principal streams
Mean annual discharge -----	15	10	33	21
Mean monthly discharge (average) --	29	19	65	41
Standard deviation of mean annual discharge -----	22	14	22	14
Standard deviation of monthly discharge (average) -----	22	14	22	14
Seasonal discharge (July–September) -----	--	--	40	25
50-year flood -----	35	20	60	37
50-year 7-day high flow -----	25	15	60	37
2-year 7-day low flow -----	18	12	18	12
20-year 7-day low flow -----	28	17	29	18

that would be calculated in other regions might appear high to those concerned with planning and design related to streamflow. Yet these accuracies are based on the nature of the phenomena being observed. These are the magnitudes of errors that occur in time sampling. Reduction of time-sampling errors by collection of longer records is accomplished at progressively higher cost, because the error varies inversely as the square root of the number of years of record. The improvement becomes progressively less as the length of record is increased. Figure 1 shows the variation of the standard error with years of record in Minnesota for several streamflow characteristics.

The value of accuracy of data used in planning and design must be considered in relation to the cost of attaining a given accuracy. For minor streams the regionalization procedure (such as by multiple-regression analysis) can eventually provide information equivalent to 10 years of record for any site at a relatively low cost. Limited studies have shown that such accuracy is on a par with, or better than, the accuracy associated with other elements than go into planning or into hydraulic design, such as the economic projections, political judgments, and the state of the art of other technical considerations. Experience has shown, moreover, that significantly greater accuracy cannot be achieved by methods of regionalization. A higher accuracy goal could be achieved for minor streams only by gaging for long periods at every site where information may conceivably be needed in the future. Because of the extremely large number of such sites, the cost of this alterna-



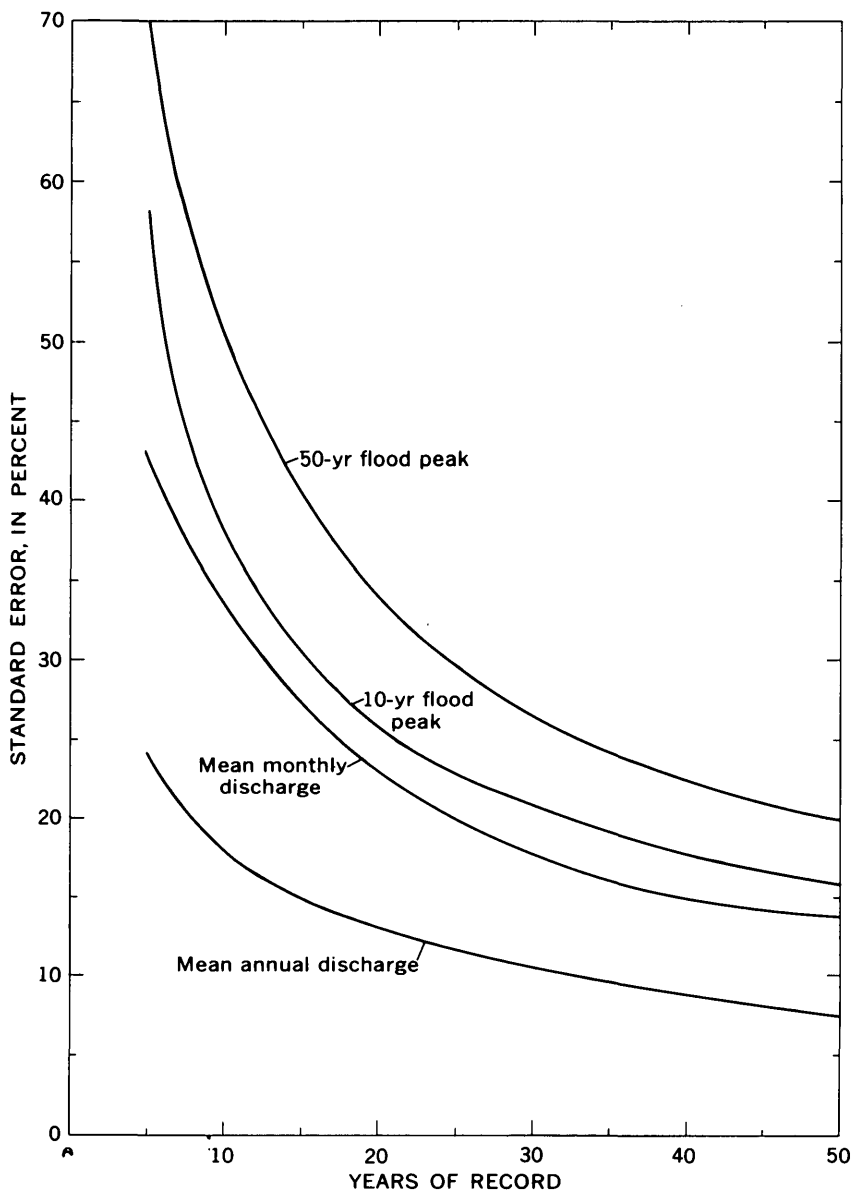


FIGURE 1.—Relation of standard error to length of record in Minnesota.

tive would be several orders of magnitude greater than for the regionalization scheme. In view of this, significantly greater accuracy goals for minor streams cannot be justified for this type of streamflow data.

### DATA TO DEFINE LONG-TERM TRENDS

The goal for this type of data is to operate indefinitely a representative sample of gaging stations on natural-flow streams in each region of the country to provide a continuing series of consistent observations. A total of about 400 such stations would provide an acceptable long-term data base. To achieve adequate geographical coverage, it was recommended that two such stations be operated primarily for this purpose in each of the subregions of the United States as identified by the Water Resources Council (1970). There are 220 such regions, 205 of which are in the conterminous United States.

### DATA ON THE STREAM ENVIRONMENT

The goals for this type of data should be set in response to the need for information. Typical data would include:

1. Stream-channel geometry, including widths, depths, slopes, hydraulic roughness, and description of bed and bank material.
2. Profiles of flood elevations and areas subject to inundation by floods.
3. Velocities and traveltimes of water and waste in channels.
4. Drainage basin characteristics, including geometry, land use such as urban areas, irrigated acres, water storage, or forested areas; and climatic characteristics. Also to be included are factors that express the recreational or nonutilitarian values of streams or valleys.
5. Aquifer characteristics, including location, extent, hydraulic connection to stream channel, and hydraulic characteristics.

### EVALUATION OF AVAILABLE DATA

Following the setting of the framework for program design, the general objective of the stream-gaging program was defined, and specific goals were set in terms of kinds of data to be furnished and accuracy standards for each. The next step was evaluating the present data base in relation to program goals. Deficiencies between present information and goals form the basis for the planned future program. A separate evaluation was made for each of the four types of data previously described.

### DATA FOR CURRENT USE

The demand for data for current use in a given region is a function of the degree of development and use of the water re-

source. Nationwide, about one-half of the total streamflow data program is presently related to this purpose. In general, it will be found that the requirements for this type of data are being met because the data are fundamental inputs to other specific and funded missions.

Identification of current-purpose stations is a prerequisite to the appraisal of the total data program. As the flow records for many of these stations have little or no transfer value in the hydrologic sense, their operation must be justified by the current need for data. Each current-purpose station was coded to identify specific uses being made of the data. This provides documentation of this segment of the program and information useful in selecting stations needed to amass the other types of data. Current uses of data to be considered in identifying stations were:

1. Assessment of current-water conditions: This is done, for example, by River Basin Commissions or Regional Planning Boards in assessing the water resources of a river basin or a region. Annual or seasonal account may be made of the inflow, outflow, rainfall, evaporation, water in storage, and water use.
2. Operation of single and multipurpose storage reservoirs: Day-by-day operations are carried on with the aid of current records of streamflow and amounts of water in storage.
3. Forecasting of flood peaks, low flows, or seasonal flows: Current flows are used together with known or assumed accretions of rainfall or snow, to make short-term forecasts of runoff and stage.
4. Disposal of waste and control of pollution: Current flow records are used to determine the amount of pollutants that can safely be discharged, or the amount of water to be released from storage pools in order to achieve desired dilution of pollutants.
5. Water-quality data programs for which discharge records are needed: Measures of concentration of chemical or sediment content need to be accompanied by discharge data in order that total loads may be computed.
6. Compact and legal requirements: Current records are required to be kept by permittees and licensees of the Federal Power Commission, and in accordance with interstate and international agreements.
7. Research or special studies: Records of only temporary interest may need to be collected in connection with research or special studies.

Many stations are used for more than one current-purpose function. The coding identifies from one to three of the functions served by each station. Table 3 shows by district offices, as of the time of the study, the total number of complete record stations

TABLE 3.—*Complete-record stream-gaging stations, by district offices, that serve a current-use function as of fiscal year 1970*

[Current-use functions are as listed on p. 15. The Maryland District Office includes Delaware and the District of Columbia, and the Massachusetts District Office includes Rhode Island, New Hampshire, and Vermont]

	Number of stations		Number of stations for indicated current-use function							Current-use functions (total)
	Total	Current-use functions	1	2	3	4	5	6	7	
Alabama -----	90	48	5	38	10	0	10	11	4	78
Alaska -----	123	47	14	15	10	0	2	0	26	67
Arizona -----	185	115	17	30	0	0	16	30	57	150
Arkansas -----	106	55	11	28	22	0	15	9	9	94
California -----	767	499	23	371	95	7	165	170	81	912
Colorado -----	371	299	15	173	6	1	33	60	91	379
Connecticut -----	81	40	16	12	22	2	16	1	9	78
Florida -----	214	117	51	72	2	4	8	5	23	165
Georgia -----	128	60	10	48	16	7	23	6	5	115
Idaho -----	254	212	13	161	90	4	52	39	29	388
Illinois -----	166	69	22	16	25	2	9	4	29	107
Indiana -----	204	114	21	53	28	6	38	18	17	181
Iowa -----	120	77	17	39	44	6	15	2	6	129
Kansas -----	156	125	18	56	45	1	81	5	23	229
Kentucky -----	139	100	14	62	35	0	6	1	28	146
Louisiana -----	83	53	9	17	9	1	27	4	17	84
Maine -----	62	46	14	24	3	0	4	16	11	72
Maryland -----	115	73	10	16	3	1	14	2	40	86
Massachusetts -----	213	119	13	87	28	1	10	23	14	176
Michigan -----	194	95	22	36	4	20	8	11	18	119
Minnesota -----	127	105	8	50	24	2	46	14	26	170
Mississippi -----	66	38	8	15	15	6	8	0	7	59
Missouri -----	161	125	8	52	56	4	37	1	37	195
Montana -----	174	160	22	86	60	0	31	49	24	272
Nebraska -----	147	124	7	87	28	0	21	81	7	231
Nevada -----	80	49	1	36	24	0	6	6	0	73
New Jersey -----	85	64	3	11	22	3	14	27	20	100
New Mexico -----	206	136	8	61	5	0	51	50	26	201
New York -----	192	117	26	62	14	2	2	7	27	140
North Carolina -----	174	61	10	27	13	2	1	11	21	85
North Dakota -----	107	80	8	71	55	5	15	12	3	169
Ohio -----	174	141	14	88	46	2	68	0	9	227
Oklahoma -----	141	127	7	82	19	0	59	15	36	218
Oregon -----	307	256	15	177	69	0	29	30	36	356
Pennsylvania -----	249	149	9	59	41	1	22	7	41	180
Puerto Rico -----	92	15	0	15	0	0	0	0	0	15
South Carolina -----	68	29	11	9	1	1	4	10	0	36
South Dakota -----	95	68	11	47	16	0	23	3	4	104
Tennessee -----	149	116	9	70	12	7	14	11	40	163
Texas -----	464	341	30	190	33	0	104	22	97	476
Utah -----	224	103	7	71	6	0	25	35	18	162
Virginia -----	176	157	27	63	34	10	52	6	98	290
Washington -----	311	186	27	119	29	1	6	33	25	240
West Virginia -----	107	54	7	43	10	0	1	0	3	64
Wisconsin -----	105	60	14	21	13	0	0	20	18	86
Wyoming -----	184	140	2	109	5	0	61	24	31	232
Total -----	8,136	5,364	634	3,075	1,147	109	1,252	891	1,191	8,299

in operation, the number that served a current-purpose function, and a breakdown of those functions. The table shows a total of 8,299 current-purpose uses for 5,359 stations, or an average of 1.55 uses per station.

#### DATA FOR PLANNING AND DESIGN

An analysis of available data is necessary to determine which of the goals for this type of data have been met. Because the goal is to define flow characteristics at any point on any stream, techniques for generalizing the information obtained at gaging stations must be employed in the analysis. The discussion of procedures given below follows the general framework described on page 3. Additional information on the use of the procedures is given in the report for each district.

#### NATURAL-FLOW MINOR STREAMS

The first question to be answered is how accurately can the statistical characteristics that are listed as goals be defined by regionalization of the data now available.

The most effective way presently known for defining statistical streamflow characteristics on a broad scale is to relate them to basin characteristics in equations developed by use of multiple-regression techniques applied to past data.

Multiple-regression analysis is a statistical technique used for relating a dependent variable (here a streamflow characteristic) to a set of independent variables (here the basin and climatic characteristics). The regression may be linear or nonlinear. A linear relation is much simpler in concept and in application. Regressions between hydrologic variables have most frequently been found to be linear if the logarithms of all the variables are used. Such a regression equation would have the form:

$$\log Y = \log a + b \log A + c \log P + d \log S \dots$$

where  $Y$  is a streamflow characteristic, such as one of the nine items listed in table 2;  $A$ ,  $P$ ,  $S \dots$  are topographic or climatic characteristics; and  $a$ ,  $b$ ,  $c$ ,  $d \dots$  are coefficients evaluated as part of the multiple-regression procedure. This method was described by Benson (1962) and Thomas and Benson (1969) in applications defining streamflow relations. A simpler but exactly equivalent form of the equation is:

$$Y = aA^bP^cS^d \dots$$

## STREAMFLOW CHARACTERISTICS

As a minimum, the following flow characteristics were computed in virtually all the studies made for use in the regression analysis:

1. Mean annual flow,  $Q_a$ .
2. Mean flow for each calendar month,  $q_i$ .
3. Standard deviation of annual flows,  $SD_a$ .
4. Standard deviation of flows for each calendar month,  $SD_i$ .
5. Seven-day low flows of 2- and 20-year recurrence intervals,  $M_{7,2}$  and  $M_{7,20}$ .
6. Seven-day flood volumes of 2- and 50-year recurrence intervals,  $V_{7,2}$  and  $V_{7,50}$ .
7. Annual flood peaks of 2-, 5-, 10-, 25-, and 50-year recurrence intervals,  $Q_t$ .

In addition to these 35 streamflow characteristics, others were used in some studies. If the record lengths permitted, the 100-year floods were computed. Other durations and recurrence intervals were used for both low flows and flood volumes. Seasonal values were used, such as June through September mean flows. Percentage points on the daily-duration curve, such as the 10- or 90-percent point, were used. In each case, the streamflow characteristics computed were those that might be used for planning and design purposes within the particular region. The total number of streamflow characteristics varied from the minimum necessary to assess the adequacy of the streamflow program to all those that might conceivably be used.

Characteristics of annual and monthly flows (means and standard deviations) were computed by means of the Geological Survey's "Flow Variability" computer program (No. W4422) from the record of daily streamflows stored on magnetic tape.

Characteristics of daily flows, such as annual 7-day lows or 3-day highs for selected recurrence intervals, were obtained by means of the "Streamflow Statistics" computer program (No. A969) from the same tapes. Optional output contains the coordinates of a log-Pearson Type III distribution fitted to any selected lowest or highest flow, and plots of both the observed points and the computed frequency curve. For short records or where the log-Pearson distribution did not fit the data adequately, graphical frequency curves were used.

Annual flood peaks of 2-, 5-, 10-, 25-, 50-year, and other intervals were defined by a log-Pearson Type III fitting program (No. 4014), from a flood-peak data storage file. For some frequency curves, adjustment to the computed curves was made graphically

to give weight to historical data. In some cases high or low peaks were omitted as outliers by comparison with other data in the vicinity. Where the computed distribution did not fit satisfactorily within a region, graphical curves were used. In only 10 to 15 percent of the total frequency curves fitted were changes made from the originally fitted curves.

#### BASIN CHARACTERISTICS

The basin characteristics are those features of the basin or the climate prevailing over the basin that are correlated with the flow characteristics by means of multiple-regression techniques. Many basin characteristics may be evaluated, but the following list was suggested as a minimum set based on experience in previous analyses:

1. Drainage area,  $A$ , in square miles, as shown in the latest data publications. Contributing area was used where appropriate.
2. Main-channel slope,  $S$ , in feet per mile, as computed by the 85- to 10-percent method described by Benson (1962).
3. Stream length,  $L$ , in miles, available from the main-channel slope computation.
4. Area of lakes and ponds,  $S_l$ , in percent of drainage area, (plus one percent) which was measured from maps by use of transparent grids, and by planimetering in the case of large lakes only.
5. Mean basin elevation,  $E$ , in 1,000 feet above m.s.l., which was measured from topographic maps by a transparent grid-sampling method. The procedure was to select a grid of a size such that between 20 and 80 points were sampled (20 is minimum, used only in very flat terrain; 80 is a practicable maximum). The more rugged and varied the terrain, the more points are needed.
6. Forested area,  $F$ , in percent of drainage area, and increased by one percent. This was measured by the grid-sampling method from topographic maps having a green overprint which indicates forest cover.
7. Mean annual precipitation,  $P$ , in inches. The most recent and detailed of such maps appears in the "Climates of the States" series of the U.S. Weather Bureau publications (1959-61). Where a reliable isohyetal map was available, it was used with the grid-sampling method so that from 20 to 80 points were sampled in each basin. If a reliable isohyetal map was not available (this is true in most moun-

- tainous regions), an anomaly map was constructed as described by Benson (1964, p. 32-33).
8. Precipitation intensity,  $I$ , expected in 24 hours once each 2 years, in inches. This characteristic was estimated (for the basin, not the gage location) from U.S. Weather Bureau Technical Paper 40 (1961), and preferably from U.S. Weather Bureau Technical Paper 29 (1958), for the middle Atlantic region.
  9. Mean minimum January temperature,  $t_i$ , in degrees Fahrenheit, was also estimated from the "Climates of the States" series. This characteristic was used only in regions affected by snowmelt floods.
  10. For appropriate districts, an index of the snowfall,  $S_n$ . This index might be a measure of the total snowfall, in inches, as was used in the Potomac River basin study, or it might preferably be a measure of the April 1 (or May 1) water content of snow. Either index could be evaluated for all basins in a district by an anomaly map. A map of maximum water content of the March snow cover in the North Central States has been prepared by the U.S. Weather Bureau in Technical Paper 50 (1964).
  11. A soils index,  $S_s$ , was used to represent the effect of soils differences on streamflow characteristics. In most districts, a map showing the variation in the soils complex curve number was furnished by the local office of the Soil Conservation Service. The soils index is a value that represents the average infiltration into the soil during the maximum annual floods. In other districts, where studies were available, the value of average permeability of the surface soils was used as an index.

In addition to the suggested standard set of basin characteristics, district offices were encouraged to try other variables, if time permitted their exploration. Instead of mean basin elevation, mean elevation above some critical altitude, such as the foot of the mountains or the snowline, was used in some of the analyses. Latitude and longitude were used in some places. Such variables as orientation of the basin, distance from the nearby coastline, drainage density, and slope of the channel above the plains were used. Evaporation, moisture deficiency, and a climatic index were used, as was the average annual number of thunderstorm days. Geologic indices of various sorts were used, including depth of loess, percentage area of stratified drift, and percentage area of glaciers.



Transparent grids were used for measuring several basin characteristics. They provide a fast as well as accurate means of measurement. A set of four grids, to different map scales, were utilized. The grids were used in several ways. For determining areas, as of ponds, a grid can be laid over a pond and the number of squares and partial squares covering the pond can be totaled and converted to an area, or the number of grid intersections within the area can be multiplied by the area in each square. For determining such things as elevations and temperatures from a contour map, a reading is taken under each of several grid intersections and the readings are averaged. For determining percentages of forested area, or perhaps small ponds if very numerous, the number of grid intersections occurring over forest or over water within the basin are counted, multiplied by 100, and divided by the total number of grid intersections within the basin.

#### REGRESSION PROCEDURES

Relating basin characteristics to streamflow characteristics was done by use of multiple-regression techniques, applied by digital computer. Either the step-forward or the step-backward methods were used (Draper and Smith, 1967, p. 44).

In selecting streamflow records for inclusion in the regression analysis, only such records were chosen as were virtually in a natural condition. The extent of regulation was investigated for each, and a record was used to define only such characteristics as were not significantly altered. For example, regulation at low flows might affect daily low flows, but not mean monthly or annual flows. In other cases, peak flows might be affected excessively, yet annual flows might not be. The effect of regulation can be judged by correlation procedures before and after the start of regulation practices, or by mass-curve analysis.

The log-Pearson fitting procedure for floods, low floods, or flood volumes, furnished estimates of flood magnitudes for recurrence intervals up to any desired values. It was considered that estimates of magnitudes for recurrence intervals of more than twice the period of record are not dependable enough for developing regression relations; thus derived data for each site were limited to recurrence intervals of less than twice the record length.

#### RESULTS OF REGRESSION ANALYSES

The detailed formulas obtained by the multiple-regression procedure for the many streamflow characteristics that were investigated are shown in the open-file reports available at the 46 district

offices of the Survey except Hawaii. They are too numerous to attempt to assemble in this report. Summary maps and tables are included here to indicate the general nature of the results.

For each regression equation that is developed, a value is computed that is called the standard error of estimate, a general index of the accuracy of estimates obtained by use of the equation. (For estimates at individual sites, a more precise estimate is provided by the standard error of prediction, which is based on the basin characteristics at the individual site.)

Thomas and Benson (1969), in a pilot study that preceded the nationwide evaluation of the surface-water program, reported on the results of multiple-regression analyses in four areas of the country that represent a wide range in hydrologic conditions. Figure 2 illustrates the accuracy, in terms of the standard error in percent, of the regression equations defined for the Potomac River basin. The flows vary in magnitude from the left to the right side of the figure. It can be seen that the defined relations are most accurate for flows nearest the mean and become less accurate for extreme flows, being least accurate for extreme low flows. Similar trends with the range of flow have been found in all other regions where studies were made. In general, it was found that the method produces more accurate relations for the humid eastern and southern regions than for the arid central and western regions.

Figures 3 to 11 (at end of report) show by district office the results of the regression analyses for selected streamflow characteristics. It may be noted that, in general, a district covers a complete State. Exceptions are the district office in Boston, Mass., which handles Massachusetts, Rhode Island, New Hampshire, and Vermont, and the district office in Towson, Md., which handles Maryland and Delaware. For these two districts, the mapped results apply to the complete area handled by each. Within many of the States, the regression analyses were best done by subdividing the entire area into two, three, or four parts and developing separate relations within each part. The dashed lines show these subdivisions. (For exact delineation of the subdivision boundaries, refer to individual open-file reports published by each district office.) Omission of a value within any State or subdivision means that no regression relation was developed for the streamflow characteristic in question.

Figures 3, 6, 8, and 10 show for several streamflow characteristics the required accuracies for ungaged areas that represent the equivalent accuracy that would be obtained by an actual 10-year

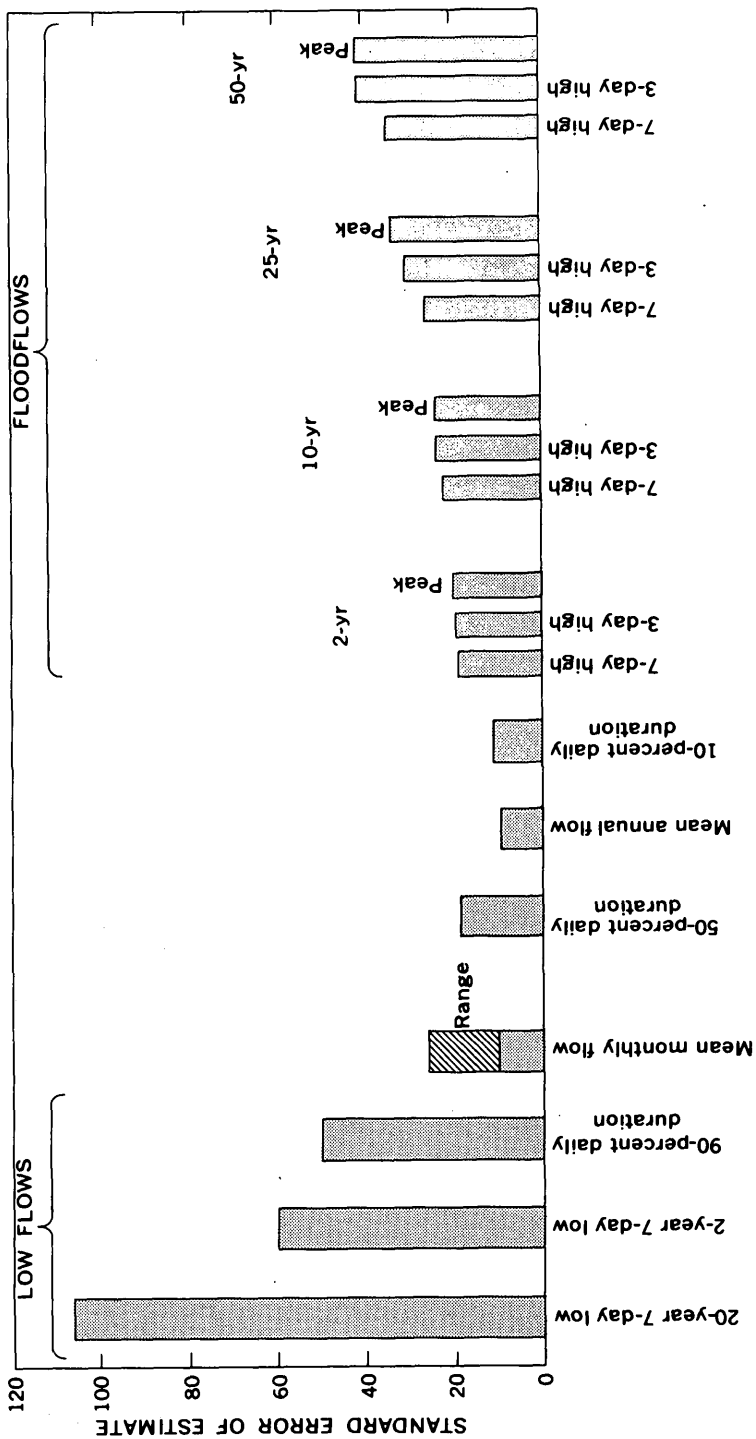


FIGURE 2.—Standard error of Potomac River basin relations.

record. The values shown represent the goals to be attained. Figures 4, 5, 7, 9, and 11 show the standard errors of estimate of the regression equations developed within each region of analysis (State, States, or subdivision of a State). Comparison for each streamflow characteristic of the accuracy goal with the standard error of estimate indicates whether or not the goal has been met by the multiple-regression method of transferring information. In figures 4, 5, 7, 9, and 11 the crosshatched areas are those within which the goals have been substantially met (standard error of estimate exceeds the accuracy goal by no more than 10 percent).

It will be noted that goals for mean and floodflow characteristics have been met over most of the eastern part of the conterminous United States, but have not been met in the West or in Alaska or Puerto Rico.

Results for low-flow stream characteristics have not been shown on these maps. Standard errors of estimate were invariably very high and did not meet the accuracy goals anywhere. The primary reason for this is that low flow is highly related to aquifer characteristics, and data on these are too scanty to permit inclusion of aquifer indices in the multiple-regression relations.

Table 4 is a summary that indicates, nationwide, the relative importance of the basin characteristics in explaining various categories of streamflow. The categories are represented by  $Q_a$ , the mean annual flow;  $SD_a$ , the standard deviation of mean annual flow;  $Q_f$ , all floodflows; and  $V_{d,f}$ , all flood volumes. The ranking is based on a count, nationwide, of the number of regression equations, within each category, in which the particular basin characteristic was found to be statistically significant at at least the 5-percent level. Drainage area,  $A$ , was the most important variable for all categories. Mean annual precipitation was the second most important for all but flood peaks, for which slope was second in importance.

The rankings of table 4 do not necessarily represent the relative importance for any particular region, for several reasons. For example:

1. Snowfall,  $S_n$ , is not important everywhere, so does not appear as a variable in most States; its importance in the northern States is larger than indicated by the nationwide ranking.
2. Evaporation,  $E_v$ , was not investigated as a variable in many places. Potentially, it is probably more important than indicated.
3. Slope,  $S$ , is, next to the drainage area, most important in defining floods. Its ranking for other categories may be misleading. This is because elevation,  $E$ , may usually be used

TABLE 4.—*Rank of variables in explaining streamflow characteristics*[Symbols are defined on p. 19 except for *Ev*, which represents evaporation index]

	<i>A</i>	<i>P</i>	<i>F</i>	<i>E</i>	<i>I</i>	<i>S</i>	<i>S<sub>t</sub></i>	<i>S<sub>i</sub></i>	<i>t<sub>i</sub></i>	<i>L</i>	<i>S<sub>n</sub></i>	<i>Ev</i>
<i>Q<sub>a</sub></i> -----	1	2	3	4	6	9	8	5	7	12	10	11
<i>SD<sub>a</sub></i> -----	1	2	5	4	6	7	3	10	8	9	12	11
<i>Q<sub>i</sub></i> -----	1	3	7	5	8	2	6	4	10	9	11	12
<i>V<sub>a, i</sub></i> -----	1	2	4	6	3	5	7	8	9	10	11	12
Total -----	4	9	19	19	23	23	24	27	34	40	44	46

interchangeably with slope; where one appears in an equation, the other usually does not.

#### NATURAL-FLOW PRINCIPAL STREAMS

In general, the accuracy goals for natural-flow principal streams cannot be met by the results of multiple-regression studies; more intense gaging is usually required.

The suggested method of attaining the accuracy goal was to set up a system of principal-stream gaging points as described on page 7. Such a system of points is established regardless of whether or not the stream in current condition represents natural or regulated flow, because these points represent sites at which it is considered desirable to define natural flow even though the streamflow may be currently regulated.

The accuracy goal could be met at each of the principal-stream gaging sites by an actual 25-year record of natural flow. In some cases, the goal could also be met by a record of less than 25 years at each site, extended by correlation with a nearby natural-flow station with a longer record to provide the equivalent of a 25-year record of monthly and annual discharge.

Streamflow information on principal streams between principal-stream gaging sites would be based on interpolation supplemented by information for tributaries, or would be obtained from relations based on several of the principal-stream gaging sites. It is assumed that with 25-year accuracy at the selected principal stream points, streamflow information can be developed for all sites throughout the principal-stream system with little or no decrease from the 25-year accuracy standard.

Evaluation of existing data for natural-flow principal streams in relation to the goals was accomplished as follows: (1) Identification of the system of principal-stream gaging points, (2) identification of such of those points where 25 years of natural-flow record were already available or where records could be

extended by regression to obtain the equivalent of 25 years of record, and (3) identification of points where a station should be continued or a new station installed to obtain the equivalent of 25 years of record of natural flow.

#### REGULATED-FLOW STREAMS

As discussed earlier in this report, a systems approach was considered necessary if meaningful information were to be provided on regulated streams. The goal is to define statistical flow characteristics for both the natural and regulated condition. Model studies will usually be required.

Existing data are difficult to assess in relation to the goals for regulated streams. It is evident, however, that existing streamflow programs are not providing the information required. In many cases it is not possible to develop even a simple water-budget equation for stream systems from the data available. The available records are not homogeneous in time, and are thus of limited use in statistical prediction. Little or no information is available for ungaged sites.

For the purpose of program design it was considered useful to identify the regulated stream systems that should be studied, consider the approach to be used, and assess the data requirements. Some cases involved a single stream with a simple pattern of regulation; other cases, an entire basin with a complex system of regulatory works, diversions, and management practices. In each case some type of analytical model is necessary by which the entire system is described. Consideration of the model approach also indicates deficiencies in the present data-collection scheme.

A total of about 300 regulated stream systems were identified as requiring modeling studies. In some districts a start has already been made in such studies.

#### DATA TO DEFINE LONG-TERM TRENDS

The goal for this type of data was to operate two natural-flow stations indefinitely in each Water Resources Council Subregion. The Federal network of recently established bench-mark stations provides this type of data. This network, however, includes only 57 stations because of the difficulty that was experienced in finding basins that are expected to remain in a natural condition in future years. Therefore, to augment the long-term data base, other stations that now represent essentially natural flow were selected to be operated until such time as the flow is markedly affected by developments in the basin. These stations were selected from the

existing network, giving consideration to current length of record, chance of the basin remaining in natural condition, conditions for accurate gaging, and use of the station records for other purposes.

#### DATA ON THE STREAM ENVIRONMENT

The assessment of this type of data required no special techniques. It was accomplished by comparing the information available with the goals that were established.

#### PLANNING FUTURE STREAMFLOW DATA PROGRAMS

The study as outlined in this report yielded the goals of the streamflow data program in each locality where it was studied, and led to an evaluation of currently available data in relation to those goals. The information developed in various segments of the study could be merged and plans for the future data program developed that would eventually attain as many of the goals as possible within the limits of available funds.

From the evaluation study it was found that some streamflow characteristics could and some could not be predicted within the standards of accuracy that have been specified. Where streamflow characteristics have been adequately defined, it is wasteful to continue to collect further data on the same scale. An elimination or reduction in collection of that type of data is justified unless it is needed for some other purpose. If, on the other hand, the accuracy of definition of streamflow is deficient, the various remedial steps that can be taken should be considered in the planning. Among the alternatives are (1) continued or augmented collection of data necessary to increase time or geographic coverage of the sampling, (2) improvement in analytical methods by research, and (3) change in the analytical methods—possibly a change from generalization to intensive measurement programs for special conditions where generalization may not furnish the desired accuracy. On regulated streams there seems to be no alternative to system studies if the effect of various operational patterns are to be predicted.

Under the first alternative, partial-record gaging probably offers the most efficient means of collecting information for minor streams, particularly if the accuracy goals for annual and monthly flow have been met. For example, if the failure is in areal definition of floodflows, the collection of data on storm rainfall and runoff to define the parameters of a rainfall-runoff model is probably the best approach, anticipating that records could be extended

in time by use of the model together with long-term rainfall records. If the results of regionalization do not apply to small-sized basins or urbanized basins because these conditions were not adequately represented by the available data, then an increase in this activity is needed in complete-record gaging or partial-record gaging for these types of basins in order to attain the goals.

Under the second alternative, the reason for deficiencies in the multiple-regression method need to be examined. The method may be deficient because the model is inadequate or because indices of all the important basin characteristics have not been included or are not adequately defined. The effects of geology or soils, or urbanization, for example, are difficult to evaluate numerically. Research may indicate more suitable models and means of developing better indices of basin characteristics.

One example of action under the third alternative is the use of specialized miscellaneous measurements. In defining low-flow characteristics, a number of locations can be measured during periods of base flow, and by correlation with index stations, the desired characteristics may be determined. If this approach is to be used, the availability of index stations to be used for this purpose must be considered.

In general, it was found that the accuracy goals for principal streams could not be met by techniques of regionalization. It was found, however, that 25 or more years of record was already available for most of the sites in the principal-stream network with natural flow. Thus, it was found possible to discontinue a number of gaging stations in this category, and to continue to gage only those sites where the goal had not been attained.

The regulated portion of principal streams will offer the greatest challenge in the future program. The present surface-water data programs are not yielding all the information required on regulated streams, because the effort has been generally limited to operation of gages at particular points with little consideration as to what is being measured or what related data are necessary to provide meaningful answers. These streams should be considered as a flow system and plans made to obtain the required input data. Where, for example, is water diverted from or returned to the stream channel? How much water is diverted or returned at each point? Where are the storage reservoirs in the basin, and what are the operational characteristics? What is the net loss due to evaporation from reservoirs or evapotranspiration from irrigated lands in the basin? What is the effect of ground-water pumpage in the basin on streamflow? Is sufficient information



available to develop a water budget or flow-storage model of the stream system? A model approach is considered to be necessary if meaningful information is to be provided on regulated streams. Full advantage should be taken of studies made and of models developed by agencies primarily concerned with design and operation of water systems.

Planning the surface-water information system cannot be done by formula. It must be done by hydrologists who are familiar with the hydrology of the region, the needs for information, the information currently available, and methods of hydrologic analysis. The results of an evaluation such as described in this paper, however, provided a firm basis for planning the future program. Decisions on number, location, and length of operation of continuous stations and partial-record stations were based on these results. Less productive elements could be weeded out and the effort shifted to important areas of streamflow information that are now receiving little attention. The plan provides for a continuous interaction between data collection and analysis, not only to gain a better understanding of the hydrologic system, but to guide the future data-collection program.

## IDENTIFICATION OF GAGING-STATION FUNCTIONS

Within each district, the end result of the program evaluation study was a proposed plan of operation consisting of gaging stations to be operated and special and analytical studies to be made. Each station proposed to be operated in the future program was identified as having a primary function and possibly other secondary functions. The functions were considered, ranked in order of importance, as (1) long-term or bench-mark stations, designated "B," (2) principal-stream stations, designated "P," (3) hydrologic stations for general-purpose information, designated "H," (4) current-purpose stations, designated "C," and (5) stations needed for hydrologic modeling of regulated basin systems, designated "R." The primary function of a station is the highest ranked function of the one or more that might be assigned to it.

Table 5 shows, for each district, the total number of continuous-record stations in the 1970 system, the number proposed to be discontinued and added, and the net total proposed for the future program. It also shows the number of stations with primary purpose in each of the five functions. The percentage distribution of the primary purposes of the continuous-record stations in the proposed network is as follows:

<i>Station purpose</i>	<i>Percentage</i>
Bench-mark -----	7.7
Principal-stream -----	12.9
Hydrologic -----	33.4
Current-purpose -----	44.4
Regulated -----	1.6
Total -----	100.0

A total of 1,007 gaging stations were proposed for discontinuance. Such recommendations were made for several reasons. The stations may have served the purpose for which they had originally been established, or in the case of some long-record stations, the information to be gained from their continued operation was not worth the expenditure, particularly in view of high-priority needs such as in urban hydrology, in small-area gaging, or in hydrologic modeling of regulated basins.

The 625 gaging stations proposed to be added to the network were recommended for several reasons. Some were to fill gaps for bench-mark stations, principal-stream stations, or for hydrologic stations where deficiencies were found in the sampling of the range of basin characteristics. The stations proposed represent the first stage in a long-range plan to shift to a more efficient system of stream gaging, to cover the highest priority needs in the most effective manner, and to provide a maximum return of information per dollar of expenditure.

### SUMMARY

During 1970 the Geological Survey initiated a nationwide study to evaluate all phases of its data-collection program. Because streamflow data are most numerous and the needs for streamflow data are best known, the 1970 effort was devoted almost exclusively to the evaluation of the streamflow data-collection program.

The study involved both evaluation of the present system and examination of ways in which it might be improved. The principal elements in the study were:

1. Establishment of the objectives and goals of the program.
2. Examination and analysis of all available data to determine which goals had been met.
3. Consideration of alternate means of meeting the goals.
4. Identification of elements to be included in the future programs.

Streamflow information was classified as either for current-purpose use, as in management and operation, or for use in planning and design. The general objective was stated as furnishing information of either type wherever it might be wanted. Goals

TABLE 5.—*Present (1970) and proposed network of continuous-record stations, by district offices, with primary purpose of stations in the proposed network*

[Primary purpose as listed on p. 29. The Maryland District Office includes Delaware and the District of Columbia, and the Massachusetts District Office includes Rhode Island, New Hampshire, and Vermont]

	Present stations (1970)	Proposed changes		Proposed network					
		Dis-continue	Add	Primary purpose					Total
				B	P	H	C	R	
Alabama -----	90	29	15	17	33	19	7	0	76
Alaska -----	123	8	69	9	57	105	13	0	184
Arizona -----	185	12	17	8	82	90	10	0	190
Arkansas -----	106	27	3	11	47	24	0	0	82
California -----	767	39	14	34	19	233	442	14	742
Colorado -----	371	22	18	15	29	150	171	2	367
Connecticut -----	81	9	0	2	1	39	30	0	72
Florida -----	214	24	0	22	7	67	94	0	190
Georgia -----	128	31	15	15	7	31	59	0	112
Idaho -----	254	20	22	13	21	56	156	10	256
Illinois -----	166	54	11	26	21	51	25	0	123
Indiana -----	204	30	0	13	6	56	98	1	174
Iowa -----	120	33	8	11	49	23	12	0	95
Kansas -----	156	19	15	13	23	28	86	2	152
Kentucky -----	139	20	32	18	34	36	63	0	151
Louisiana -----	83	15	0	12	5	14	37	0	68
Maine -----	62	9	6	5	30	24	0	0	59
Maryland -----	115	22	2	13	10	71	1	0	95
Massachusetts -----	213	12	0	7	3	58	108	25	201
Michigan -----	194	30	6	17	35	60	58	0	170
Minnesota -----	127	22	53	15	97	38	8	0	158
Mississippi -----	66	26	21	11	17	4	29	0	61
Missouri -----	161	18	10	18	10	37	88	0	153
Montana -----	174	2	0	14	11	44	103	0	172
Nebraska -----	147	1	37	9	17	88	67	2	183
Nevada -----	80	5	19	6	0	43	45	0	94
New Jersey -----	85	4	11	4	0	21	64	3	92
New Mexico -----	206	24	24	12	0	61	102	31	206
New York -----	192	43	25	11	1	47	108	7	174
North Carolina -----	174	65	5	7	13	64	25	0	114
North Dakota -----	107	5	23	12	27	58	33	0	130
Ohio -----	174	26	0	13	0	7	125	3	148
Oklahoma -----	141	7	4	2	20	11	105	0	138
Oregon -----	307	23	16	16	20	91	173	0	300
Pennsylvania -----	249	51	11	8	17	61	123	0	209
Puerto Rico -----	92	3	12	7	0	82	12	0	101
South Carolina -----	68	12	1	7	24	25	1	0	57
South Dakota -----	95	9	20	11	27	27	39	2	106
Tennessee -----	149	13	12	6	8	40	94	0	148
Texas -----	464	35	9	35	46	184	163	10	438
Utah -----	224	18	0	12	55	106	33	0	206
Virginia -----	176	12	4	16	5	21	125	1	168
Washington -----	311	55	0	23	13	106	103	11	256
West Virginia -----	107	20	4	11	3	29	48	0	91
Wisconsin -----	105	18	6	15	10	19	49	0	93
Wyoming -----	184	25	40	14	33	46	106	0	199
Total -----	8136	1007	625	596	998	2595	3441	124	7754

were set in terms of accuracy of information for various types of streams (unregulated, regulated, principal, minor) and for various categories of flow (such as mean flows, high flows).

Results may be summarized as follows:

1. In many localities, notably in the eastern half of the country, all or some of the goals have been met. Where this is so, discontinuance of part of the gaging effort is indicated as desirable, in favor of other types of data collection for which needs exist.
2. In some localities, mostly in the western half of the country, few or none of the goals were met. Even where this was so, the study showed that some long-record stations could profitably be discontinued in favor of other gaging, because the marginal information furnished after many years of operation is small.
3. Classification of each gaging site as to its specific utility disclosed some for which no management function was known; if planning and design information goals had already been met, such stations were recommended for discontinuance.
4. A network of stations was designated for indefinite operation to maintain a record of long-term trends.
5. Among the management types of stations, a network of stations was designated to maintain a surveillance and assessment of flow into and out of sub-basins identified for planning purposes by the Water Resources Council.
6. Basins were identified that, because of extensive regulation, require the development of hydrologic modeling to answer future needs for information.
7. Goals in some places were not met for floodflow definition, and in all places goals for low-flow information were not met. Both of these types of information can be furnished more efficiently by partial-record gaging rather than complete-record gaging. Where appropriate, a shift was planned from complete-record to partial-record gaging.
8. There are present deficiencies in information on small streams and on streams under urban conditions. Therefore, where goals had been met in whole or part, a shifting of gaging effort was planned to meet these needs.
9. Of the 8,136 complete-record stations in operation, 1,007, or about 12 percent, have been recommended for discontinuance. Implementation of such action, with shifting to higher-priority efforts, are contingent on approval of State or other federal agencies that cooperate in the work of the Water Resources Division.
10. In recommendations for the network of the future, the specific purpose of each gaging station is identified.

The study, in general, has led to an extremely worthwhile examination of the streamflow program, and to plans for redirection

in many phases towards a more efficient system for satisfying present and future needs. A similar analysis was being initiated in fiscal year 1971 for the water quality data-collection program.

## REFERENCES CITED

- Benson, M. A., 1962, Factors influencing the occurrence of floods in a humid region of diverse terrain: U.S. Geol. Survey Water-Supply Paper 1580-B, 62 p.
- 1964, Factors affecting the occurrence of floods in the Southwest: U.S. Geol. Survey Water-Supply Paper 1580-D, 72 p.
- Carter, R. W., and Benson, M. A., 1969, Concepts for the design of streamflow data programs: U.S. Geol. Survey open-file rept., 20 p.
- Draper, N. R., and Smith, H., 1967, Applied regression analysis: New York, John Wiley & Sons, 407 p.
- Hardison, C. H., 1969, Accuracy of streamflow characteristics, in Geol. Survey research, 1969, Chap. D: U.S. Geol. Survey Prof. Paper 650-D, p. 210-214.
- Thomas, D. M., and Benson, M. A., 1969, Generalization of streamflow characteristics from drainage-basin characteristics: U.S. Geol. Survey Water-Supply Paper 1975, 55 p.
- U.S. Weather Bureau, 1958, Rainfall intensity-frequency regime, part 3, the middle Atlantic region: Tech. Paper 29, 38 p.
- 1959-61, Climates of the States: U.S. Weather Bureau, Climatology of the United States, no. 60 [section for each State].
- 1961, Rainfall frequency atlas of the United States: Tech. Paper 40, 115 p.
- 1964, Frequency of maximum water equivalent of March snow cover in north central United States: Tech. Paper 50, 24 p.
- Water Resources Council, 1970. Water resources regions and subregions for the national assessment of water and related land resources: Washington, D.C., 188 p.



---

---

**FIGURES 3-11**

---

---

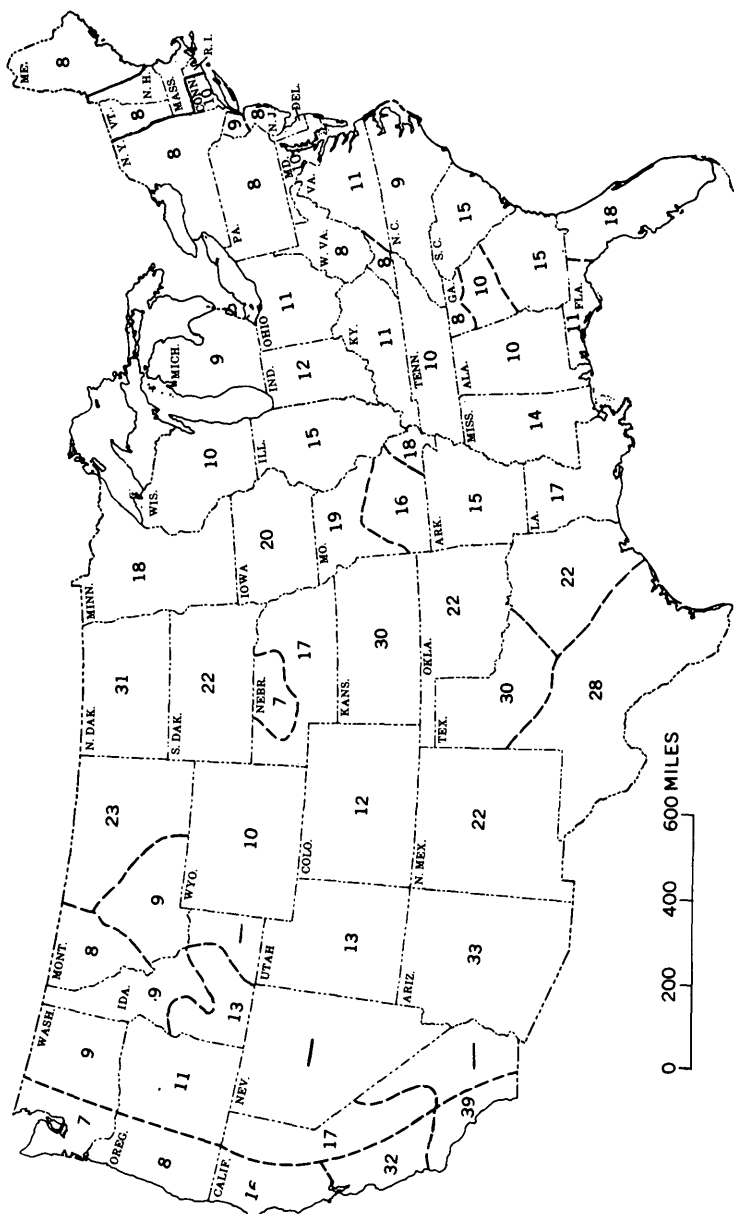


FIGURE 3.—Ten-year accuracy goals in percent for mean annual flow. The corresponding accuracy goal is 5 percent for Alaska and 9 percent for Puerto Rico. —, no accuracy goals developed.



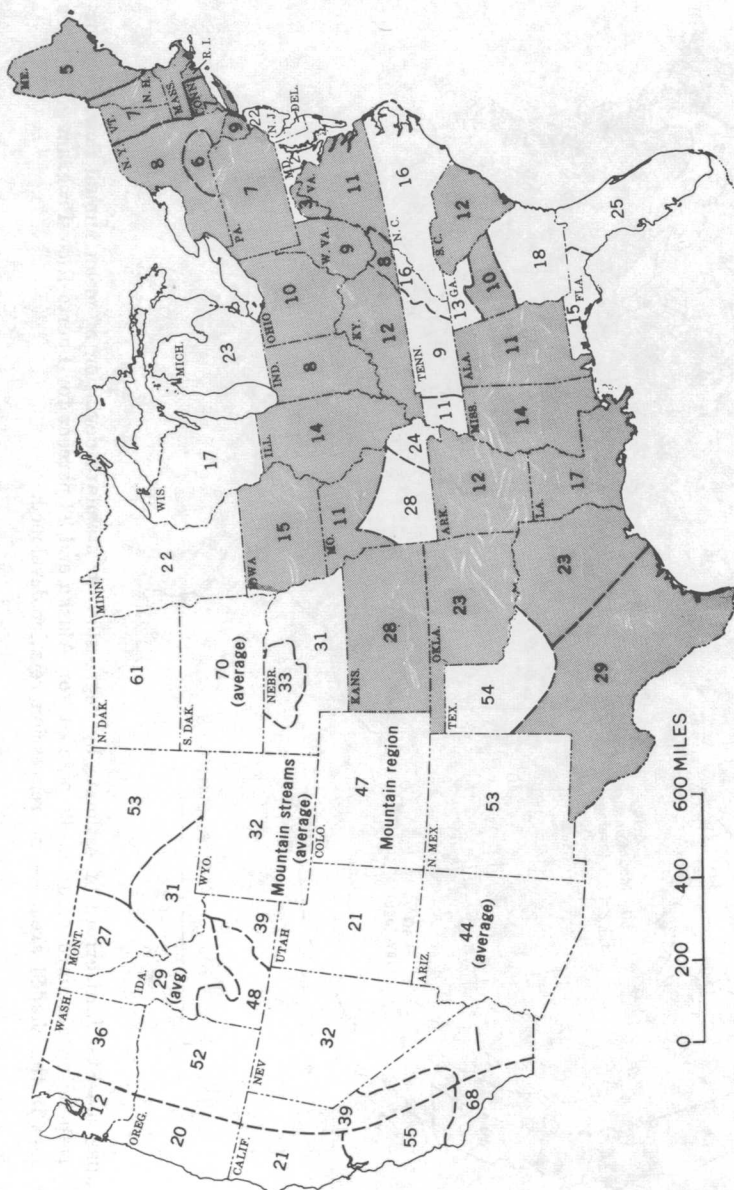


FIGURE 4.—Standard errors of estimate by regression for mean annual flow. The corresponding standard error is 39 percent for Alaska and 18 percent for Puerto Rico. Program goals were met in the shaded area. —, no regression relation developed.



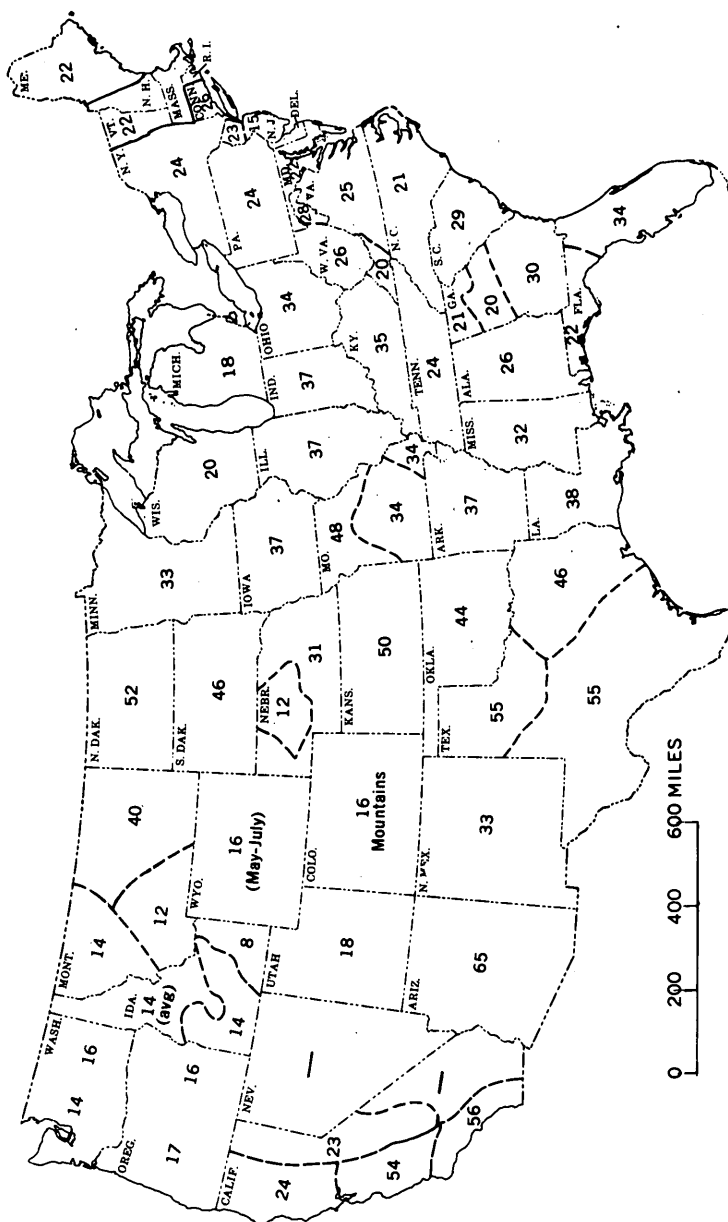


FIGURE 6.—Average 10-year accuracy goals in percent for monthly means. The corresponding accuracy goal is 12 percent for Alaska and 18 percent for Puerto Rico. —, no accuracy goals developed.

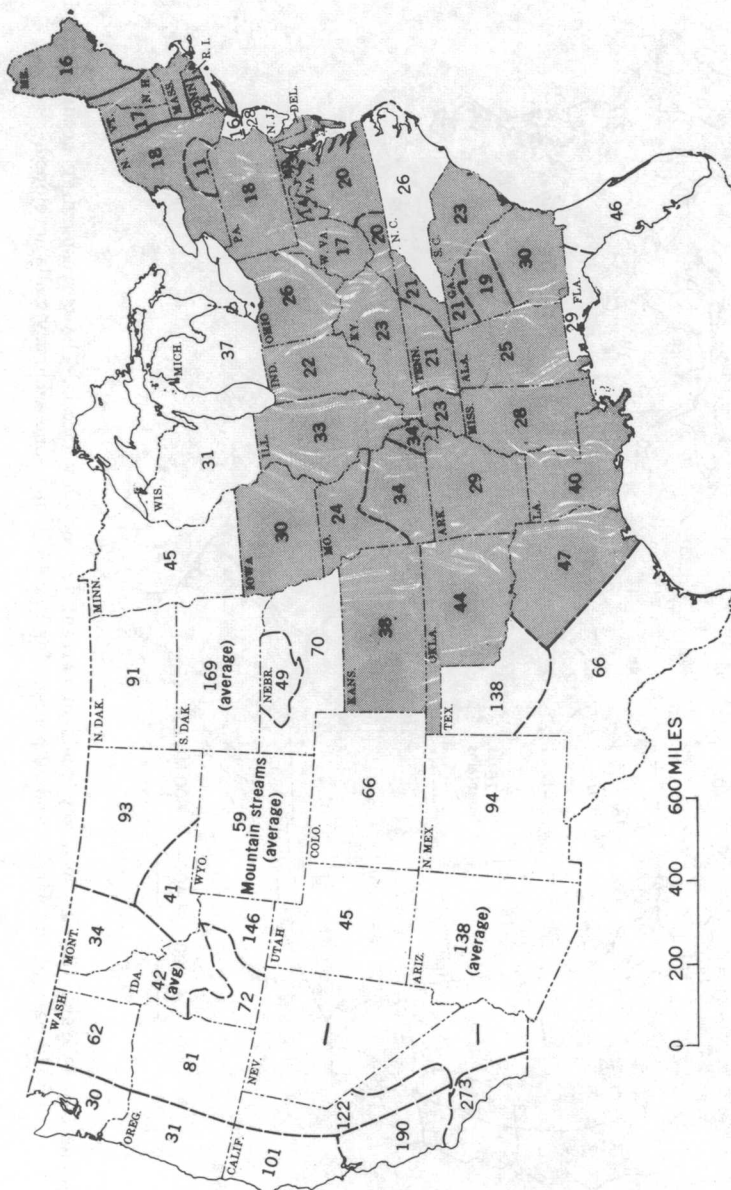


FIGURE 7.—Average standard errors of estimate by regression for monthly means. The corresponding standard error is 51 percent for Alaska and 41 percent for Puerto Rico. Program goals were met in the shaded area. —, no regression relation developed.

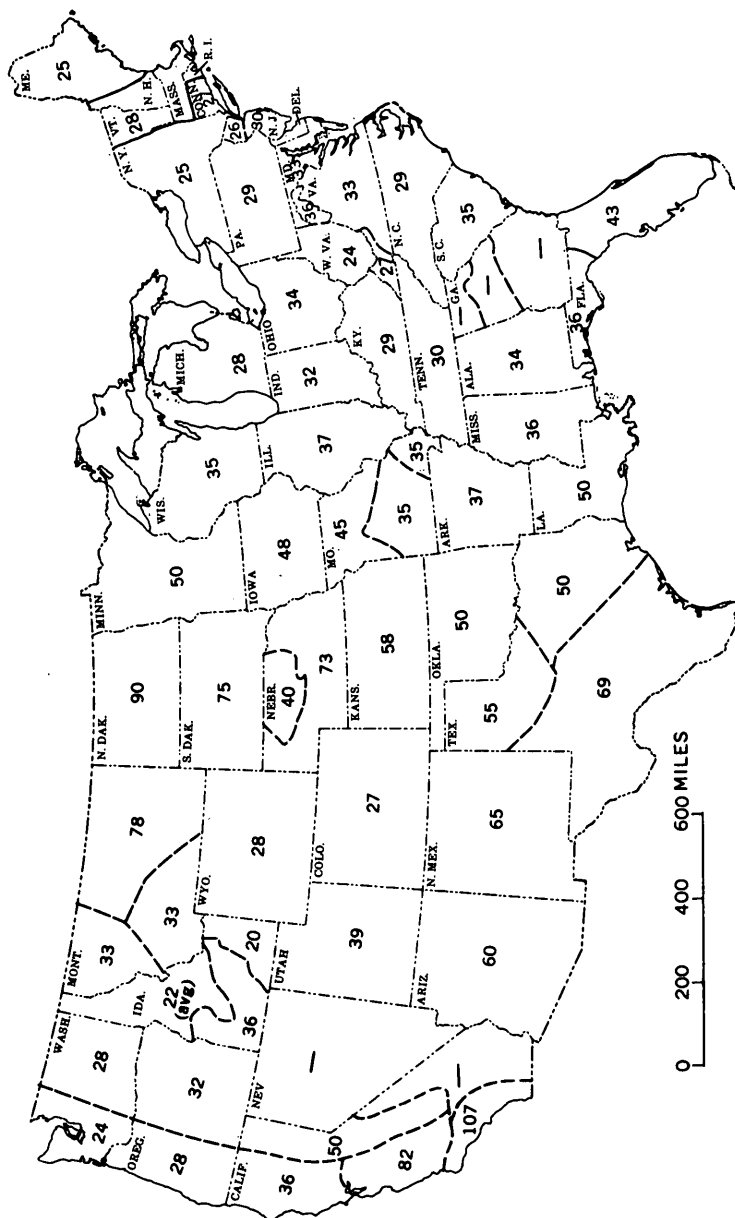


FIGURE 8.—Ten-year accuracy goals in percent for 50-year flood. The corresponding accuracy goal is 22 percent for Alaska and 50 percent for Puerto Rico. —, no accuracy goals developed.



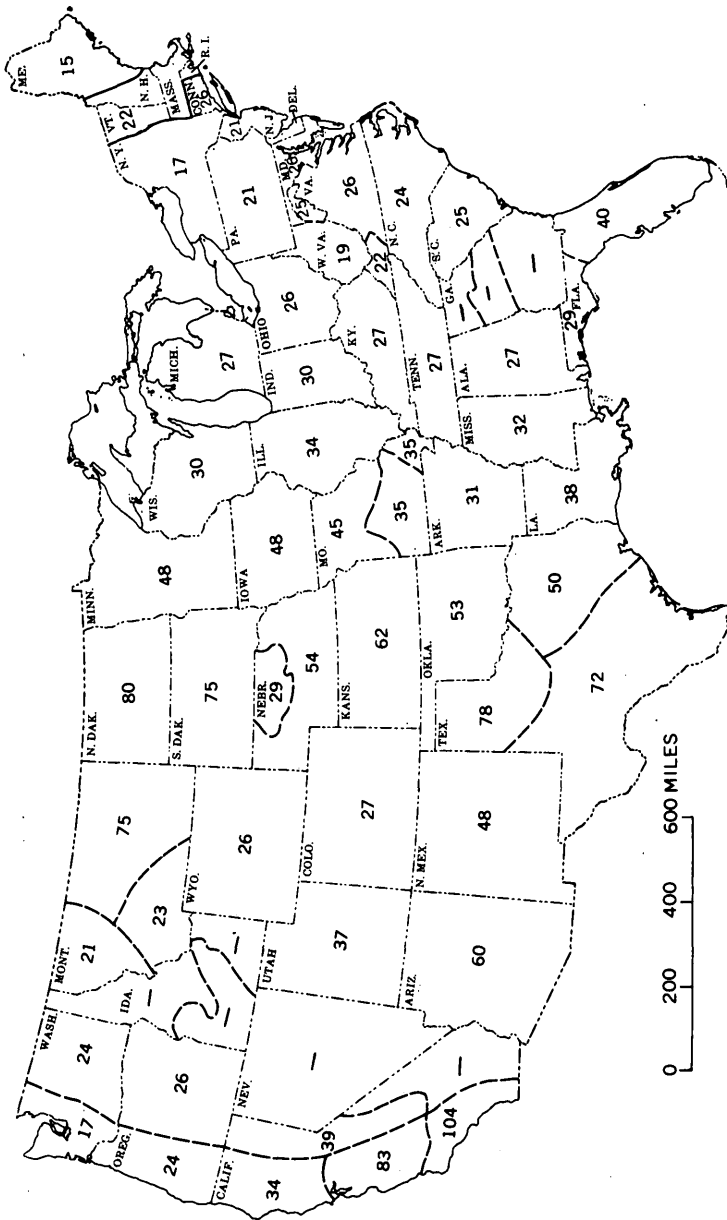


FIGURE 10.—Ten-year accuracy goals in percent for 7-day 50-year flood volume. The corresponding accuracy goal is 13 percent for Alaska. —, no accuracy goals developed.



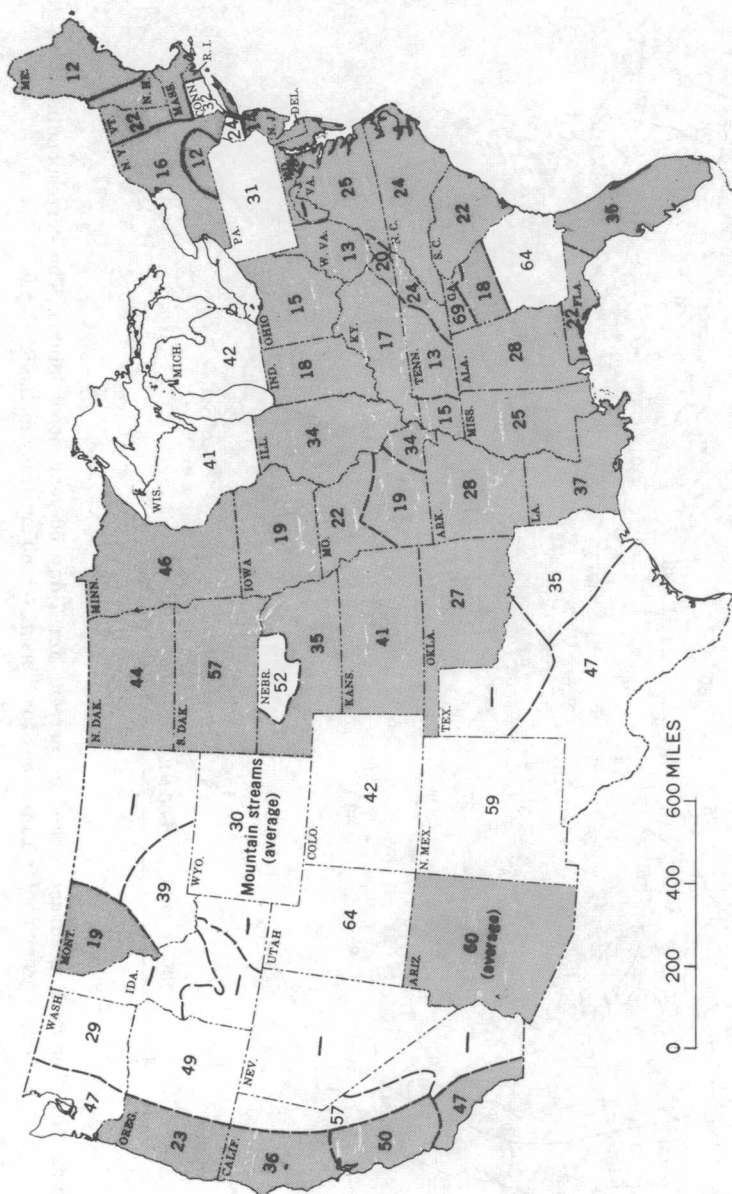


FIGURE 11.—Standard error of estimate by regression for 7-day 50-year flood volume. The corresponding standard error is 50 percent for Alaska. Program goals were met in the shaded area. —, no regression relation developed.