21. W.L. Graf. The Arroyo Problem--Paleohydrology and Paleohydraulics in the Short Term, from Background to Paleohydrology. John Wiley and Sons, New York, 1983.
22. B.M. Reich. Recent Changes in a Flood Series. Proc., Hydrology Section--Arizona-Nevada Academy of Science, Vol. 14, 1984, pp. 231-238.
23. Peak Discharge at Sites for Flood of October 1983 (Status as of July 9, 1984). U.S. Geological Survey, Tucson, Ariz., 1984, p. 1.
24. W.C. Boughton and K.G. Renard. Flood Frequency Characteristics of Some Arizona Watersheds. Water Resources Bulletin, Vol. 20, No. 5, 1984, pp. 761-770.
25. U.R. Schneider and K.V. Wilson. Hydraulic Design of Bridges with Risk Analysis. Report FHWA-TS-226, Office of Development, FHWA, U.S. Department of Transportation, 1980, 145 pp.

# Cost-Effectiveness of the U.S. Geological Survey Stream-Gauging Program

MARSHALL E. MOSS, WILBERT O. THOMAS, Jr., and ARTHUR G. SCOTT

ABSTRACT

A summary of the results of a cost-effectiveness study of the U.S. Geological Survey stream-gauging program in 17 states is contained in this paper. The results are for the first year of a 5-year nationwide study undertaken by the U.S. Geological Survey. The objective of the study is to define and document the most cost-effective means of furnishing streamflow information. The first step of this study involved identification of data uses and funding sources for 1,939 continuous-record stations currently being operated with a budget of $11,425,650. Only 35 continuous-record stations were identified as not having sufficient justification to continue their operation. In addition, 31 more short-term special study stations were identified as not having justified data uses beyond completion of their respective studies. In the second step, evaluation of alternative methods of providing streamflow information, flow-routing and regression models were developed for estimating daily flows at 145 stations of the 1,939 stations analyzed. Only 6 of the 145 stations that were analyzed were considered to have acceptable accuracy of the simulated flows for the intended uses of the data. Based on the accuracy of the simulated flows, the operation of continuous-record gauging stations at these locations could be discontinued. In the third step of the analysis, relationships were developed between the accuracy of the streamflow records and the operating budget. For the current operating budget, the weighted average standard error was 21.0 percent for the programs analyzed in the 17 states. By redistribution of resources among the stations according to an optimization program, this weighted average standard error can be reduced to 19.0 percent. The current weighted average standard error of 21.0 percent can conversely be achieved with a reduced budget of $10,889,800, a total budget reduction of $535,850.

To provide basic information on the flow of the nation's streams and rivers is one of the major functions of the U.S. Geological Survey (USGS). The vast majority of this information is generated by the collection of streamflow data at some 15,000 locations throughout the United States. At approximately 8,000 of these sites, the flow of rivers, streams, or canals is continuously gauged. These gauged records are permanently stored in the Daily Values File of the USGS National Water Data Storage and Retrieval System (WATSTORE) (1). At the remaining sites, only partial records of the flow are collected. The partial-record station usually only provides data at the high (flood) or low (drought) ends of the streamflow spectrum. Many of these gauges provide the basic data required by state highway departments for the economical design of highway drainage structures. Other gauges provide data for research in rural and urban flood frequency estimation methods at gauged and ungauged sites, flow-backwater techniques, and risk analysis. In fiscal year 1983, more than $40 million was expended in the collection and processing of streamflow data by the USGS.

The first-line management of the USGS stream-gauging program is performed at the Water Resources Division (WRD) district level. WRD districts usually correspond geographically with the boundaries of one or more states. Exceptions are the Caribbean District, which includes Puerto Rico and the U.S. Virgin Islands, and the Hawaii District, which includes the Pacific Trust Territories and Hawaii.

Because of the large scale of this program and its hydrologic and managerial complexities, a considerable effort has been expended within the National Research Program of the WRD to develop technologies for the design and management of data collection programs. As new tools became operational, they were readily implemented in the district programs. The nationwide implementation of one such set of tools is described by Benson and Carter (2). Subsequent to this nationwide study in the early 1970s, no new technologies were developed that had sufficient impact on the program to warrant another nationwide study until 1980. In 1980, Moss and Gilroy (3) presented a new approach to measuring the cost-effectiveness of a stream-gauging program. By using this technique, called the Kalman-Filtering for Cost-Effective Resource Allocation (K-CERA), the manager of a stream-gauging program can evaluate allocations of gauging effort among the continuous stream gauges of the program such that the overall amount of information that is generated would be a maximum. The K-CERA is composed of a set of techniques and computer programs to estimate measures of the errors in streamflow estimates and to distribute fiscal resources in a network to minimize the sum of error variances of each site, which, in turn, maximizes information. However, the approach does not specify the set of gauges that should make up the program. To address this last point, other steps are required.

The potential impact of the K-CERA as a management tool led the USGS to initiate another nationwide analysis of its stream-gauging program in 1982 with this approach as its basis. Because of the relatively large initial investment of manpower in the implementation of the K-CERA, it was decided that the nationwide study would be performed over a 5 year period. In each of the 5 years, managerial units, usually districts, that contained 20 percent of the stream-gauging program would complete the analyses. Locations and areal extents of the studies performed during the first year are shown by the shaded areas in Figure 1. During 1983, the stream-gauging program was analyzed in Alaska, Arkansas, northern California, central Florida, Georgia, Hawaii (including the Pacific Trust Territories), Idaho, Illinois, Iowa, Kansas, Maine, Massachusetts, Nebraska, New Jersey, Pennsylvania, Rhode Island, and Washington. These analyses are summarized in this report for the first study year.


**FIGURE 1** Locations and areal extents of the studies performed during the 1983 fiscal year.

## APPROACH

An analysis of a stream-gauging program would, ideally, define the proper set of stream gauges to be operated and specify the most cost-effective way to operate those gauges. The K-CERA addresses the second aspect, but no robust technology for definition of the proper set of stream gauges currently exists. A pragmatic approach that consists of three sequential steps is therefore being used to analyze each of the continuous stream gauges in the USGS stream-gauging program. The first two steps involve screening each gauge in the program as to whether it should remain in use as a continuous stream gauge. In the first step, all known uses of the data that are generated at each continuous-record site are documented for comparison with other data collected as part of the USGS mission of generating streamflow information. Those stream gauges with uses that are not found to be sufficient and compatible with the USGS mission are suggested for discontinuance. Additionally, funding for the operation of each gauge and the frequency of the availability of its data are also documented.

In the second step, those gauges that passed the first screening are investigated as to whether a sufficient amount of the streamflow information contained in the streamflow data can be generated by means of either hydrologic models or statistical methods. These alternative methods of generating streamflow information are less costly than operating a continuous stream gauge. No guidelines concerning suitable accuracies exist for particular uses of the data. Therefore, judgment is required in deciding whether the accuracy of the estimated daily flows is suitable for the intended purpose. If the alternative method is successful for a particular stream gauge, then that stream gauge becomes a candidate for discontinuance.

Those gauges that pass the first two steps make up the continuous stream-gauging program that is to be subjected to the K-CERA analysis for the determination of its optimal operation in terms of cost-effectiveness.

A brief description of the content of each of these steps follows. However, if more details are desired, see the report by Fontaine et al. (4), which served as a prototype for all of the other areas analyzed in 1984.

## STEP ONE--CATEGORIZATION BY DATA USE, FUNDING, AND FREQUENCY OF AVAILABILITY

### Data Use Categories

The following definitions were used to categorize each known use of streamflow data for each continuous stream gauge. A given station may be included in more than one data use category.

### Regional Hydrology

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

### Hydrologic Systems

Stations that can be used for accounting (i.e., for defining current hydrologic conditions and the sources, sinks, and fluxes of water through hydrologic systems that include regulated systems) are designated as hydrologic systems stations. They include diversions and return flows and stations that are useful for defining the interaction of water systems.

The benchmark and index stations are included in the hydrologic systems category because they account for the current and long-term conditions of the hydrologic systems that they gauge. Federal Energy Regulatory Commission (FERC) stations and international gauging stations, located on significant rivers that cross national boundaries, are also included.

### Legal Obligations

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

### Planning and Design

Gauging stations in this category of data use are used for the planning and design of a specific project (for example, a dam, levee, floodwall, navigation system, water-supply diversion, hydropower plant, or waste-treatment facility) or group of structures. The planning and design category is limited to those stations where these purposes are currently valid.

### Project Operation

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

### Hydrologic Forecasts

Gauging stations in this category are regularly used to provide information for hydrologic forecasting. These might be flood forecasts for a specific river reach, or periodic (daily, weekly, monthly, or seasonal) flow-volume forecasts for a specific site or region. The hydrologic forecast use generally implies that the data are routinely available to the forecasters on a rapid reporting basis. On large streams, data may only be needed every few days. Data used for forecasting inflows or outflows solely for project operation are categorized as project operation and are not contained in the forecast category.

### Water-Quality Monitoring

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

### Research

Gauging stations in the research category are operated for a particular research or water-investigation study. Typically, these are only operated for a few years.

### Other

The eight categories described previously contain the majority of data uses. However, occasional data uses have been identified that do not fit into the scheme. Therefore, the "other" category is provided for such instances.

## Funding

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

1. Federal--Funds that have been directly allocated to the USGS.
2. Other federal agency (OFA)--Funds that have been transferred to the USGS by OFAs.
3. Cooperative (COOP)--Funds that come jointly from USGS cooperative-designated funding and from a nonfederal cooperating agency. Cooperating agency funds may be in the form of direct services or cash.
4. Other nonfederal--Funds that are provided entirely by a nonfederal agency or a private concern under the auspices of a federal agency. In this study, funding from private concerns was limited to that derived from the licensing and permitting requirements for hydropower development by the FERC. Funds in this category are not matched by USGS cooperative funds.

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

## Frequency of Data Availability

Frequency of data availability refers to the times at which the streamflow data may be furnished to the users. In this category, three distinct possibilities exist. Data can be furnished by (a) direct-access telemetry equipment for immediate use, (b) periodic release of provisional data, or (c) in the annual data reports published by the USGS.

## STEP TWO--CONSIDERATION OF ALTERNATIVE METHODS

Two methods were used to synthesize streamflow records at gauging stations where it was thought that the records were sufficiently correlated with the records of one or more other stations. These two methods are described briefly in the following. Usually no more than 10 percent of the gauges in any district program were candidates for the alternative-methods analysis.

### Description of Flow-Routing Model

Hydrologic flow-routing methods use the law of conservation of mass and the relationship between the storage in, and outflow from, a reach. The hydraulics of the system are not considered. The method usually requires only a few parameters and treats the reach in a lumped sense without subdivision. The input is usually a discharge hydrograph at the upstream end of the reach and the output, a discharge hydrograph at the downstream end. Several different types of hydrologic routing are available such as Muskingum, Modified Puls, Kinematic Wave, and the unit-response flow-routing method. The last method was selected for this analysis. Two techniques are used--storage continuity (5) and diffusion analogy (6,7). The computer program that utilizes these two techniques of flow routing is described by Doyle et al. (8).

### Description of Regression Analysis

Simple and multiple-regression techniques were also used to estimate daily flow records. Regression equations relate daily flows at a single gauge to daily flows at a combination of upstream, downstream, and (or) tributary gauges. This statistical method is not limited, as is the flow-routing method, to gauges where an upstream gauge exists on the same stream. The explanatory variables in the regression analysis can be data from gauges in different watersheds or in downstream and tributary watersheds. The regression method has many of the same attributes as the flow-routing method in that it is easy to apply, provides indexes of accuracy, and is generally accepted as a good tool for estimation. The theory and assumptions of regression analysis are described in several textbooks such as that by Draper and Smith (9) and that by Kleinbaum and Kupper (10). The application of regression analysis to hydrologic problems is described and illustrated by Riggs (11).

## STEP THREE--K-CERA

In a study of the cost-effectiveness of a network of stream gauges that was operated to determine the amount of water consumption in the Lower Colorado River Basin, a set of techniques called the K-CERA was developed (3). Because of the water-balance nature of that study, the measure of effectiveness of the network was chosen to be the minimization of the sum of variances of errors of estimation of annual mean discharges at each site in the network. This measure of effectiveness tends to concentrate stream-gauging resources on the larger, less stable streams where potential errors are greatest. Although such a tendency is appropriate for a water-balance network, in the broader context of the multitude of uses of the streamflow data collected in the USGS's Streamflow Information Program, this tendency causes undue concentration on larger streams. Therefore, the original version of the K-CERA was extended to include, as optional measures of effectiveness, the sums of the variances of the following: errors of annual mean discharge estimation in cubic feet per second and percent, and errors of average instantaneous discharge estimation in cubic feet per second and percent. The use of percentage errors does not, however, unduly weight activities at large streams to the detriment of records on small streams. In addition, the instantaneous discharge is the basic variable from which all other streamflow data are derived. For these reasons, this study used the K-CERA approach with the sums of the percentage error variances of the instantaneous discharges at all continuously gauged sites as the measure of the effectiveness of the data-collection activity.

Brief descriptions of the mathematical program that was used to optimize cost-effectiveness of the data-collection activity and of the application of Kalman filtering (12) to the determination of the accuracy of a stream-gauging record are presented in the following paragraphs. For more detail on the theory, the assumptions, or the applications of the K-CERA, see reports by Moss and Gilroy (3), Gilroy and Moss (13), and Fontaine et al (4).

### Description of Mathematical Program

One program in the K-CERA technique is called "the traveling hydrographer." This program attempts to allocate among stream gauges a predefined budget for the collection of streamflow data so that the field operation is the most cost-effective possible. The measure of effectiveness was discussed previously in this paper. The set of decisions available to the manager is the use frequency (number of times per year) of each of a number of routes that may be used to service the stream gauges and to make discharge measurements. The range of options within the program is from zero to daily usage for each route. (A route is defined as a set of one or more stream gauges and the least cost travel that takes the hydrographer from his base of operations to each of the gauges and back to base.) A route will be associated with an average cost of travel and an average cost of servicing each stream gauge visited along the way. The first step taken by the analyst is to define the set of practical routes. This set of routes will frequently contain the path to an individual stream gauge with that gauge as the lone stop and return to the home base so that the individual needs of a stream gauge can be considered in isolation from the other gauges.

The analyst then determines any special requirements for visits to each of the gauges for such things as necessary periodic maintenance, servicing of recording equipment, or required periodic sampling of water-quality data. Such special requirements are considered to be inviolable constraints in terms of the minimum number of visits to each gauge.

The final step is to use the traveling hydrographer program with all of the above to determine the number of times that the routes are used during a year such that (a) the budget for the network is not exceeded, (b) the minimum number of visits to each station is made, and (c) the total uncertainty in the network is minimized.

### Description of Uncertainty Functions

As noted earlier, uncertainty in streamflow records is measured in this study as the average relative variance of estimation of instantaneous discharges. The accuracy of a streamflow estimate depends on how that estimate was obtained. Three situations are considered in this study:

1. Streamflow is estimated from measured discharge and correlative data using a stage-discharge relation (rating curve),
2. The streamflow record is reconstructed using secondary data at nearby stations because primary correlative data are missing, and
3. Primary and secondary data are unavailable for estimating streamflow.

The variances of the errors of the estimates of flow that would be employed in each situation were weighted by the fraction of time each situation is expected to occur. The average relative variance would thus be

$$\bar{V} = \epsilon_f V_f + \epsilon_r V_r + \epsilon_e V_e \tag{1a}$$

with

$$1 = \epsilon_f + \epsilon_r + \epsilon_e \tag{1b}$$

where

- $\bar{V}$ = the average relative variance of the errors of streamflow estimates,
- $\varepsilon_f$ = the fraction of time that the primary recorders are functioning,
- $V_f$ = the relative variance of the errors of flow estimates from primary recorders,
- $\varepsilon_r$ = the fraction of time that secondary data are available to reconstruct streamflow records given that the primary data are missing,
- $V_r$ = the relative variance of the errors of estimation of flows reconstructed from secondary data,
- $\varepsilon_e$ = the fraction of time that primary and secondary data are not available to compute streamflow records, and
- $V_e$ = the relative error variance when both primary and secondary data are not available.

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

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

If the recorder at the primary site fails and there are no concurrent data at other sites that can be used to reconstruct the missing record at the primary site, there are at least two ways of estimating discharges at the primary site: a recession curve could be applied from the time of recorder stoppage until the gauge was again functioning, or the expected value of discharge for the period of missing data could be used as an estimate.

The expected-value approach is used in this study to estimate $V_e$, the relative error variance during periods of no concurrent data at nearby stations. If the expected value is used to estimate discharge, the value that is used should be the expected value of discharge at the time of year of the missing record because of the seasonality of the streamflow processes. The variance of streamflow, which also is a seasonally varying parameter, is an estimate of the error variance that results from using the expected value as an estimate.

The variance $V_r$ of the relative error during periods of reconstructed streamflow records is estimated on the basis of correlation between records at the primary site and records from other gauged nearby sites.

Because errors in streamflow estimates arise from three different sources with widely varying precision, the resultant distribution of those errors may differ significantly from a normal or log-normal distribution. This lack of normality causes difficulty in interpretation of the resulting average estimation variance. When primary and secondary data are unavailable, the relative error variance $V_e$ may be very large. This could yield correspondingly large values of $\bar{V}$ in Equation 1a even if the probability that primary and secondary information are not available ($\varepsilon_e$) is quite small.

A new parameter, the equivalent Gaussian spread (EGS), is introduced here to assist in interpreting the results of the analyses. If it is assumed that the various errors arising from the three situations represented in Equations 1a and 1b are log-normally distributed, the values of EGS was determined by the probability statement that

$$\text{Probability}\left\{ e^{-EGS} \leq [q_c(t)/q_T(t)] \leq e^{+EGS} \right\} = 0.683 \tag{2}$$

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

## SUMMARY OF INDIVIDUAL STUDIES

The 17 individual state studies are summarized with regard to the data use, alternative methods, and K-CERA analysis. For each step of the analysis, summary statistics are presented and an evaluation is made of what was learned. A total of 1,939 stations were analyzed in the 1983 fiscal year, which represents approximately 24 percent of the nationwide stream-gauging program.

### Uses, Funding, and Availability of Continuous Streamflow Data

The analysis of data uses in the previously mentioned 17 states verified that data obtained in the national stream-gauging program are utilized for a variety of purposes by state and local governments, other federal agencies, and private industry. Of the 1,939 stations analyzed, nearly all had one or more data uses and only 35 continuous-record stations were identified as not having sufficient justification to continue their operation. In addition, 31 more short-term special study stations were identified as not having justified data uses beyond completion of their respective studies. The 66 stations that were suggested for discontinuance in the near future represent about 3 percent of the 1,939 stations analyzed.

A summary by state of the number of stations in each data-use category is given in Table 1. The data in Table 1 show that regional hydrology and hydrologic systems are the two primary data uses; 55 and 50 percent of the stations, respectively, are classified in these two categories. Streamflow data are utilized about equally in making decisions about the operation of water-resources projects, in making hydrologic forecasts of potential flooding, and in

TABLE 1 The Number of Stations in Each Data Use Category for the Stream-Gauging Program Analyzed in the 1983 Fiscal Year

| State | Total No. of Stations | Regional Hydrology | Hydrologic Systems | Legal Obligations | Planning and Design | Project Operation | Hydrologic Forecasts | Water-Quality Monitoring | Research | Other |
|---|---|---|---|---|---|---|---|---|---|---|
| Alaska | 110 | 91 | 40 | 0 | 36 | 7 | 21 | 38 | 17 | 4 |
| Arkansas | 49 | 34 | 29 | 1 | 4 | 34 | 11 | 23 | 0 | 1 |
| Northern California | 127 | 30 | 73 | 0 | 15 | 49 | 24 | 19 | 9 | 2 |
| Central Florida | 94 | 81 | 86 | 0 | 27 | 21 | 1 | 23 | 0 | 0 |
| Georgia | 98 | 64 | 51 | 0 | 11 | 34 | 31 | 36 | 1 | 0 |
| Hawaii | 124 | 56 | 65 | 0 | 1 | 0 | 0 | 7 | 0 | 5 |
| Idaho | 156 | 85 | 108 | 3 | 43 | 84 | 76 | 49 | 13 | 14 |
| Illinois | 138 | 88 | 87 | 0 | 37 | 17 | 52 | 96 | 90 | 0 |
| Iowa | 110 | 64 | 42 | 0 | 0 | 76 | 80 | 17 | 1 | 0 |
| Kansas | 140 | 73 | 88 | 5 | 2 | 92 | 114 | 115 | 0 | 0 |
| Maine | 51 | 28 | 16 | 0 | 0 | 18 | 26 | 6 | 8 | 5 |
| Massachusetts | 76 | 15 | 55 | 0 | 5 | 23 | 9 | 22 | 10 | 1 |
| Nebraska | 145 | 62 | 133 | 11 | 15 | 135 | 87 | 41 | 3 | 4 |
| New Jersey | 101 | 87 | 29 | 3 | 7 | 37 | 33 | 41 | 10 | 2 |
| Pennsylvania | 223 | 145 | 27 | 3 | 4 | 67 | 61 | 142 | 8 | 14 |
| Rhode Island | 15 | 5 | 5 | 0 | 7 | 1 | 2 | 10 | 5 | 0 |
| Washington | 182 | 66 | 29 | 4 | 32 | 37 | 51 | 57 | 33 | 62 |
| Total | 1,939 | 1,074 | 963 | 30 | 246 | 732 | 679 | 742 | 208 | 114 |

monitoring the water quality of the nation's streams. Of the stations classified, 38, 35, and 38 percent, respectively, were contained in the project operation, hydrologic forecasts, and water-quality-monitoring categories. The legal obligations and planning and design categories contained a relatively low percentage of stations (1.5 and 13 percent, respectively). The research and "other" categories were also relatively small (11 and 6 percent, respectively). The research category has decreased significantly in recent years with the completion of many small streams rainfall-runoff modeling projects and the curtailment of activity in the coal and oil-shale hydrology programs. The "other" category includes uses that do not fit into the other eight categories. Many districts included stations that were operated for recreational purposes in this category.

A funding summary for the stream-gauging program that is analyzed is given in Table 2. A shown, the primary funding source for the stream-gauging program is the COOP program. Approximately 61 percent of the stations that were analyzed in the 1983 fiscal year were financed by the COOP program. The next major category was the OFA program with approximately 32 percent of the stations. The federal and other nonfederal programs are about equal with 9 percent of the stations in each of these categories. With the exception of the other nonfederal program, the percentages reported above agree fairly well with the values given by Gilbert and Buchanan (14) for the entire 1981 USGS water-data program. They reported that 60.6 percent of the funding was provided through the COOP program, 27.3 percent through the OFA program, 11.8 percent through the federal program, and 0.3 percent through the other nonfederal program. At least from a funding standpoint, the stream-gauging program analyzed in the 1983 fiscal year appears to be fairly representative of the nationwide program.

A summary of data availability is given in Table 3. As given, data for nearly all stations are published in the annual data report of the USGS. Only 5 of 1,939 stations do not have data published in the annual report. These stations are primarily short-term stations operated for special studies. Of the stations, approximately 27 percent have data available on a real-time basis from either a satellite data-collection platform or some type of landline

TABLE 2 The Number of Stations in Each Funding Category for the Stream-Gauging Program Analyzed in the 1983 Fiscal Year

| State | Total No. of Stations | Federal Program | OFA Program | COOP Program | Other Nonfederal Programs |
|---|---|---|---|---|---|
| Alaska | 110 | 25 | 26 | 52 | 16 |
| Arkansas | 49 | 6 | 33 | 25 | 1 |
| Northern California | 127 | 3 | 23 | 67 | 40 |
| Central Florida | 94 | 0 | 10 | 84 | 0 |
| Georgia | 98 | 10 | 44 | 32 | 19 |
| Hawaii | 124 | 7 | 10 | 108 | 0 |
| Idaho | 156 | 12 | 58 | 94 | 0 |
| Illinois | 138 | 4 | 46 | 91 | 0 |
| Iowa | 110 | 18 | 76 | 52 | 0 |
| Kansas | 140 | 10 | 62 | 72 | 1 |
| Maine | 51 | 3 | 15 | 23 | 14 |
| Massachusetts | 76 | 2 | 16 | 60 | 0 |
| Nebraska | 145 | 32 | 26 | 99 | 0 |
| New Jersey | 101 | 5 | 12 | 67 | 22 |
| Pennsylvania | 223 | 20 | 109 | 169 | 16 |
| Rhode Island | 15 | 3 | 1 | 12 | 0 |
| Washington | 182 | 9 | 48 | 84 | 47 |
| Total | 1,939 | 169 | 615 | 1,191 | 176 |

Note: A single station may have multiple sources of funding. Therefore, the total number of stations may be less than the sum of the stations listed under specific programs (e.g., 110 versus 119 for Alaska).

TABLE 3 The Number of Stations in Each Data Availability Category for the Stream-Gauging Program Analyzed in the 1983 Fiscal Year

| State | Total No. of Stations | Annual Report | Real Time | Provisional |
|---|---|---|---|---|
| Alaska | 110 | 110 | 18 | 4 |
| Arkansas | 49 | 49 | 17 | 32 |
| Northern California | 127 | 127 | 17 | 58 |
| Central Florida | 94 | 94 | 2 | 6 |
| Georgia | 98 | 98 | 39 | 49 |
| Hawaii | 124 | 124 | 0 | 0 |
| Idaho | 156 | 156 | 67 | 44 |
| Illinois | 138 | 138 | 37 | 25 |
| Iowa | 110 | 110 | 76 | 9 |
| Kansas | 140 | 140 | 54 | 8 |
| Maine | 51 | 51 | 21 | 12 |
| Massachusetts | 76 | 74 | 18 | 4 |
| Nebraska | 145 | 145 | 15 | 75 |
| New Jersey | 101 | 98 | 32 | 9 |
| Pennsylvania | 223 | 223 | 39 | 20 |
| Rhode Island | 15 | 15 | 2 | 0 |
| Washington | 182 | 182 | 79 | 18 |
| Total | 1,939 | 1,934 | 533 | 373 |

telemetry. As can be noted in Table 3, Iowa has telemetry at almost 70 percent of their stations. It is anticipated that the percentage of stations available on a real-time basis will increase nationwide in the future.

## Alternative Methods of Developing Streamflow Information

Flow-routing and regression models were developed for 145 different stations as part of the 1983 analysis of the stream-gauging program. This represents 7.5 percent of the 1,939 stations analyzed. Flow-routing methods (8) were applied to 52 stations and regression methods were applied to 129 stations for a total of 181 applications of an alternative method. There were 35 stations for which both the flow-routing and regression models were applied. Of the 145 stations analyzed, only 6 were considered to have acceptable accuracy of the simulated flows. Two of these stations were being utilized in a real-time data collection program, so they were not suggested for discontinuance. The other four stations were suggested for discontinuance conditioned on agreement from the cooperators. There were 14 additional stations where the analysts reported promising results and suggested further study to refine the models and to pursue discussions with the data users to define acceptable accuracy.

Different criteria relative to acceptable accuracy of the simulated flows were used in the individual studies. Therefore, the results reported above concerning the number of successful applications of alternative methods are not consistent across all states. A more consistent way of evaluating the alternative methods analysis would be to summarize these stations that meet certain accuracy requirements. A review of the individual state analyses indicated that 23 stations had 75 percent or more of the simulated flows within 10 percent of the observed flows for either the flow-routing model or regression model or both. Likewise, there were 13 stations that had 85 percent or more of the simulated flows within 10 percent of the observed flows and one station exceeded 95 percent. The best results for the flow-routing model were on the Rock River in Illinois (94 percent within 10 percent), the Ohio River in Pennsylvania (93 percent within 10 percent), and the Skagit River near Marblemount in Washington (93 percent within 10 percent). The best results for the regression model were also on the Skagit River (97 percent within 10 percent). These are all large rivers with relatively low variability of flow and a small percentage of intervening drainage area between stations. The application of the flow-routing or regression models on streams with a large percentage of intervening drainage area or regression modeling on nearby watersheds did not result in acceptable accuracy.

The flow-routing and regression models were both applied to 35 stations. A comparison was made of the accuracy of the two models by utilizing the results for only those stations that had at least 50 percent of the simulated flows within 10 percent of observed flows and for which the results were reported in the individual state analyses. An analysis of the 21 stations meeting these criteria revealed that the flow-routing results were most accurate for 10 stations, and the regression results were most accurate for 11 stations. It is fairly obvious that the two models are comparable in accuracy based on the assumption that each model was "adequately" calibrated for each station analyzed.

## Cost-Effective Resource Allocation

Suggestions were made after the data-use and alternative-methods analyses regarding the discontinuation of stations. In some studies, the stations suggested for discontinuance were omitted from the K-CERA analysis, whereas in others they were included pending discussions with the cooperators. As a result, 1,894 stations were included in the K-CERA analysis. These stations were included on various routes and the cost of operating these stations were included in the budget. However, uncertainty functions were developed for 1,714 stations. This implies that 180 stations were not used in computing the average standard errors for the various state programs. The primary reasons that uncertainty functions could not be developed for these 180 stations were lack of discharge measurements to develop a rating curve and the inappropriateness of the data at the site to fit the basic assumptions of the current K-CERA techniques. For the current operating procedures, the average standard errors ranged from 10 to 36 percent with a weighed average standard error of 21.0 percent for all 1,714 stations analyzed. The total budget for the current operating procedure is $11,425,650. By altering the field activities as determined in the individual K-CERA analyses and maintaining this current budget, this weighted average standard error can be reduced to 19.0 percent. The current weighted average standard error of 21.0 percent can conversely be achieved with a reduced budget of $10,889,800, a total reduction of $535,850. An example of the relationship between average standard error and budget is shown in Figure 2 for Maine (4).

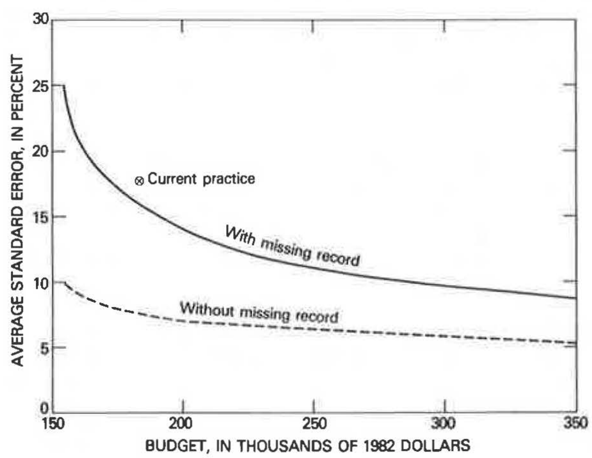


FIGURE 2 An example of the relationship between average standard error per station and budget.

Even though the EGS value was not computed for the entire stream-gauging program that was analyzed in 1983, a comparison of EGS and standard error values for an individual state analysis will illustrate the relative differences. Fontaine et al. (4) reported an average standard error of 17.7 percent and an EGS of 4.2 percent for the current operating practice. In using this study as a guideline, the weighted average standard error of 21.0 percent for the 17 studies is approximately equivalent to an EGS value of 5 percent. This comparison is predicated on the fact that the percentage of lost stage record did not vary significantly among most states. This implies that for two-thirds of the time, the error in estimating the instantaneous discharge is approximately 5 percent.

Some analysts developed an uncertainty-cost relation under the assumption that the instrumentation gave a complete stage record throughout the year. For the current budget and optimal operating practice, an analysis of nine different studies indicated a reduction in standard error of 7 percent if no missing stage record is assumed. If this reduction in standard error is indicative of all the states, this implies that the average standard error of 19.0 percent can be reduced to approximately 12 percent. It is obvious that the standard error of the missing record is a major portion of the total standard error (see Figure 2 for Maine). Nearly all analysts recognized this fact and suggested that satellite delay relay, landline telemetry, and observers be utilized to reduce the occurrence of the missing stage record.

## FUTURE DEVELOPMENTS

Research and development will continue on improving the methodology for cost-effective analysis of the stream-gauging program. For several studies completed in fiscal year 1983, a sample of stations with the highest standard errors will be analyzed to determine whether the Markovian model assumed for the Kalman filter is appropriate. This analysis should provide some guidelines on when the Markovian model is appropriate.

Improved estimates of the variance of the missing record are needed because of the importance of this factor on the total standard error at a station. The present model relies primarily on correlation with nearby stations to estimate the variance of the missing record. The flows for previous days at the station of interest are not considered. New techniques for estimating the variance of the missing record should include the length of the missing period and the correlation with flows just prior to the missing period. In this way, the hydrograph-recession characteristics can be utilized to estimate the variance of the missing record. This new technique will be incorporated into the appropriate computer program in the near future.

A study is planned to investigate the percentages of the missing record used in the K-CERA analysis to determine whether they are realistic. Because of the time limitations of the individual studies, most analysts were unable to do a detailed analysis of the missing record issue. A more detailed analysis is important because of the sensitivity of the total standard error to this factor.

Research is also continuing on more sophisticated models for the Kalman filter. As of 1984, work was underway to develop models for describing Kalman filters that are appropriate for sand channel streams and streams where artificial controls are regularly cleaned. These models will be used in studies as soon as they are operational. Every attempt is being made to utilize all information normally available to the analyst who computes the published record.

New studies initiated in fiscal year 1984 are utilizing the improvements noted in this paper as they become available. The primary objective of the present research is to make the K-CERA package of programs an effective tool for managing and determining the accuracy of data that is generated through the nationwide stream-gauging program.

## REFERENCES

1. N.E. Hutchison. WATSTORE: National Water Data Storage and Retrieval System of the U.S. Geological Survey. User's Guide, Vol. 1. Open-File Rept. 75-0426. U.S. Geological Survey, Denver, Colo., 1975, 791 pp.
2. M.A. Benson and R.W. Carter. A National Study of the Streamflow Data-Collection Program. Water-Supply Paper 2028. U.S. Geological Survey, Washington, D.C., 1973, 44 pp.
3. M.E. Moss and E.J. Gilroy. Cost Effective Stream-Gauging Strategies for the Lower Colorado River Basin. Open-File Rept. 80-1048. U.S. Geological Survey, Reston, Va., 1980, 128 pp.
4. R.A. Fontaine, M.E. Moss, J.A. Smath, and W.O. Thomas, Jr. Cost-Effectiveness of the Stream-Gauging Program in Maine. Water-Supply Paper 2244. U.S. Geological Survey, Washington, D.C., 1984, 39 pp.
5. V.B. Sauer. Unit Response Method of Open-Channel Flow Routing. Journal of the Hydraulics Division, ASCE, Vol. 99, No. HY1, 1973, pp. 179-193.
6. T.N. Keefer. Desktop Computer Flow Routing. Journal of the Hydraulics Division, ASCE, Vol. 100, No. HY7, 1974, pp. 1047-1058.
7. T.N. Keefer and R.S. McQuivey. Multiple Linearization Flow Routing Model. Proc., ASCE, Vol. 100, No. HY7, 1974, pp. 1031-1046.
8. W.H. Doyle, H.O. Shearman, G.J. Stiltner, and W.O. Krug. A Digital Model for Streamflow Routing by Convolution Methods. Water-Resources Investigations Rept. 83-4160. U.S. Geological Survey, Denver, Colo., 1983, 136 pp.
9. N.R. Draper and H. Smith. Applied Regression Analysis, 2nd ed. John Wiley and Sons, New York, N.Y., 1966.
10. D.G. Kleinbaum and L.C. Kupper. Applied Regression Analysis and Other Multivariable Methods. Duxbury Press, North Scituate, Mass., 1978.
11. H.C. Riggs. Regional Analysis of Streamflow Characteristics. Techniques of Water-Resources Investigations Book 4. U.S. Geological Survey, Alexandria, Va., 1973, 15 pp.
12. A. Gelb. ed. Applied Optimal Estimation. Massachusetts Institute of Technology Press, Cambridge, Mass., 1974.
13. E.J. Gilroy and M.E. Moss. Cost-Effective Stream-Gauging Strategies for the Lower Colorado River Basin. Open-File Rept. 81-1019. U.S. Geological Survey, 1981, 128 pp.
14. B.K. Gilbert and T.J. Buchanan. Water-Data Program of the U.S. Geological Survey. Circular 0863. U.S. Geological Survey, 1982, 16 pp.