COST-EFFECTIVENESS OF THE STREAMFLOW-GAGING PROGRAM IN WYOMING

By Stanley A. Druse and Kenneth L. Wahl

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 87-4264

Cheyenne, Wyoming

1988

DEPARTMENT OF THE INTERIOR

DONALD PAUL HODEL, Secretary

U.S. GEOLOGICAL SURVEY

Dallas L. Peck, Director

For additional information
write to:

District Chief
U.S. Geological Survey
1120 Capitol Avenue
P.O. Box 1125
Cheyenne, Wyoming 82003

Copies of this report can
be purchased from:

U.S. Geological Survey
Books and Open-File Reports Section
Federal Center, Bldg. 810
Box 25425
Denver, Colorado 80225

CONTENTS

FIGURES

TABLES

CONVERSION FACTORS

For use of readers who prefer to use metric (International System) units, conversion factors for inch-pound units used in this report are listed below:

Multiply inch-pound unit	By	To obtain metric unit
foot (ft)	0.3048	meter
mile (mi)	1.609	kilometer
square mile (mi^2)	2.590	square kilometer
cubic foot (ft^3)	0.02832	cubic meter
foot per second (ft/s)	0.3048	meter per second
square foot per second (ft^2/s)	0.0929	square meter per second
cubic foot per second (ft^3/s)	0.02832	cubic meter per second

COST-EFFECTIVENESS OF THE STREAMFLOW-GAGING PROGRAM IN WYOMING

By Stanley A. Druse and Kenneth L. Wahl

ABSTRACT

This report documents the results of a cost-effectiveness study of the streamflow-gaging program in Wyoming. One phase of evaluating the cost-effectiveness considers the use of alternative methods to simulate streamflow records. Regression analysis or hydrologic flow-routing techniques were considered for 24 combinations of stations from a 139-station network operated in 1984 to investigate suitability of techniques for simulating streamflow records. Only one station was determined to have sufficient accuracy in the regression analysis to consider discontinuance of the gage.

The evaluation of the gaging-station network, which included the use of associated uncertainty in streamflow records, is limited to the nonwinter operation of the 47 stations operated by the Riverton Field Office of the U.S. Geological Survey. The current (1987) travel routes and measurement frequencies require a budget of $264,000 and result in an average standard error in streamflow records of 13.2 percent. Changes in routes and station visits, using the same budget, could optimally reduce the standard error by 1.6 percent.

Budgets evaluated ranged from $235,000 to $400,000. A $235,000 budget would increase the optimal standard error for the current budget from 11.6 to 15.5 percent, and a $400,000 budget could decrease the optimal error to 6.6 percent. For all budgets considered, lost record accounts for about 40 percent of the average standard error.

INTRODUCTION

The U.S. Geological Survey is the principal Federal agency collecting surface-water data in the Nation. The collection of these data is a major activity of the Water Resources Division of the Geological Survey. The data are collected in cooperation with State and local governments and other Federal agencies (Schuetz, 1986). Currently (1987), the Geological Survey is operating approximately 7,000 continuous-record gaging stations throughout the Nation; some of these records date back to the turn of the century.

Any activity of long standing, such as the collection of surface-water data, as noted in Schuetz (1986), needs to be re-examined at intervals--if not continually--because of changes in objectives, technology, or external constraints. The last systematic nationwide evaluation of the streamflow information program was completed in 1970 and was documented by Benson and Carter (1973). The Wyoming contribution to that evaluation was done by Wahl (1970). In 1983, the Geological Survey undertook another nationwide analysis of the streamflow-gaging program. The analysis is to be completed over a 5-year period; 20 percent of the program is to be analyzed each year. The objectives of the nationwide analysis are to define and document the most cost-effective methods of furnishing streamflow information. Sections of this report that describe techniques or methodology are from earlier reports (Engel and others, 1984, and Fontaine and others, 1984).

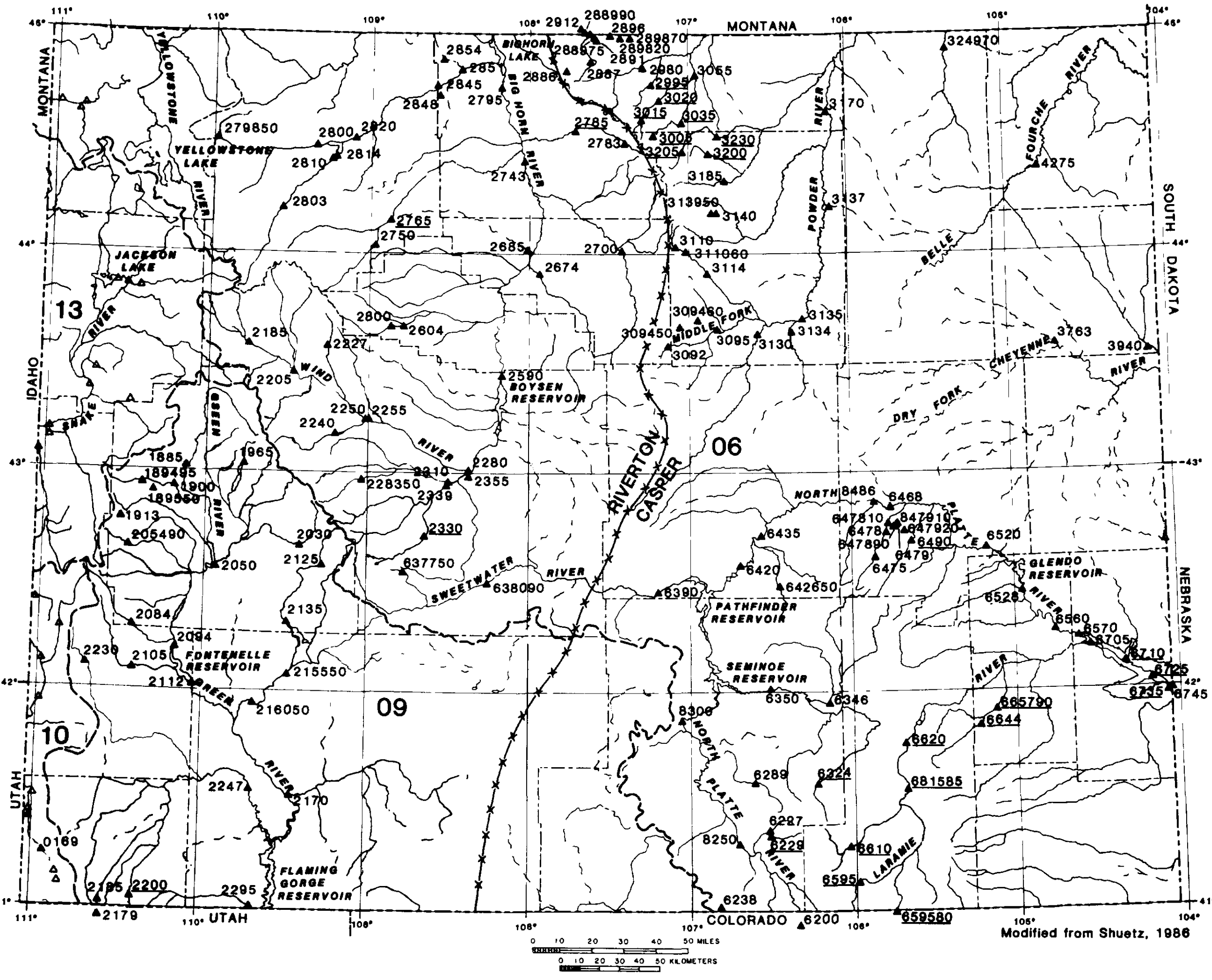
Purpose and Scope

The nationwide analysis of the streamflow-gaging program comprises three major phases of analysis. Data use and availability are analyzed in phase one; less-costly, alternate methods of furnishing streamflow information are identified in phase two; and statistical techniques are used in phase three for evaluating the operation of gaging-station networks using associated uncertainty in streamflow records for various operating budgets. The purpose of this report is to document phases two and three of the nationwide analysis, as applied to the Wyoming District of the Geological Survey.

Phases of the Analysis

Phase one of the three-phase analysis for Wyoming--to analyze data use and availability--was reported by Schuetz (1986). That report documented a survey that identified local, State, and Federal uses of data from 139 continuous-record, surface-water stations being operated in 1984 by the Wyoming District of the Geological Survey. Additionally, the report identified sources of funding relating to collection of streamflow data and presented frequency of data availability. The uses of data from the stations were categorized into seven classes: regional hydrology, hydrologic systems, legal obligations, planning and design, project operation, hydrologic forecasts, and water-quality monitoring. The report documented that sufficient use of surface-water data collected from the stations justified continued operation of all stations.

The second phase of the analysis--to identify less-costly, alternate methods of furnishing streamflow information--was applied to those stations in the statewide network that had sufficient correlation to warrant either regression analysis or flow-routing. The third phase of the analysis--to evaluate the operation of gaging-station networks using associated uncertainty in streamflow records for various operating budgets--was limited to the network of stations operated by the Riverton Field Office of the Wyoming District, U.S. Geological Survey (fig. 1). This network consists of stations in the Missouri River, Colorado River, and The Great basins in western Wyoming and is approximately half of the surface-water stations operated by the Wyoming District. The evaluation of that network was considered adequate to address the cost-effectiveness of the overall streamflow-gaging program in Wyoming and to provide a basis for considering changes in operating procedures.



EXPLANATION

— — MAJOR DRAINAGE DIVIDE

06 MISSOURI RIVER BASIN

09 COLORADO RIVER BASIN

10 THE GREAT BASIN

13 SNAKE RIVER BASIN

x—x APPROXIMATE BOUNDARY BETWEEN AREAS SERVED BY RIVERTON AND CASPER FIELD OFFICES

2295 ▲ STREAMFLOW-GAGING STATION AND NUMBER--Operated by Wyoming District. (Underlined station number indicates station is operated by Wyoming State Engineers office.) Station numbers are abbreviated by not showing two-digit basin number and last two digits if zero.

△ STREAMFLOW-GAGING STATION--Operated in Wyoming by adjoining district, but not included in Wyoming report.

Figure 1.--Location of active continuous-record streamflow-gaging stations in Wyoming.

Wyoming Streamflow-Gaging Program

The Wyoming streamflow-gaging program has evolved since its beginning in the late 1800's to meet Federal, State, and local needs for surface-water data. The streamflow-gaging network of 139 stations as described by Schuetz (1986) includes the network evaluated in this report (fig. 1). The streamflow-gaging program has remained fairly stable since 1984.

The operation of the streamflow-gaging network is shared by personnel of the Geological Survey and the Wyoming State Engineer's Office as part of a cooperative program. The Geological Survey operates its network of streamflow-gaging stations from field offices located in Casper and Riverton. The Casper office operates stations located in the eastern half of the State and Riverton the western half (fig. 1). The State Engineer's Office currently (1987) operates 26 stations in the cooperative program, and these stations were included in the first and second phase of the analysis. Data from continuous-record streamflow-gaging stations operated by the State Engineer's Office as part of its own management program are not included in this analysis.

ALTERNATIVE METHODS OF DEVELOPING STREAMFLOW INFORMATION

The second phase of the analysis of the streamflow-gaging program was to investigate alternative methods of providing daily streamflow information instead of operating continuous-flow gaging stations. The objective of this phase of the analysis was to identify gaging stations where alternative technology, such as regression analysis or flow-routing methods, could efficiently provide accurate estimates of daily mean streamflow. No guidelines exist concerning suitable accuracies for particular uses of the data; therefore, judgment was required in deciding whether the accuracy of the estimated daily flows would be adequate for the intended purpose.

The data uses at a station affect whether or not information can potentially be provided by alternative methods. For example, those stations for which flood hydrographs are required on a current basis, such as hydrologic forecasts and project operation, are not candidates for the alternative methods. Likewise, there might be a legal obligation to operate an actual gaging station that would preclude using alternative methods. The primary candidates for alternative methods are stations that are operated upstream or downstream from other stations on the same stream. The accuracy of the estimated streamflow at these stations may be adequate if flows are highly correlated between stations. Gaging stations in similar watersheds, located in the same physiographic and climatic area, also may have potential for alternative methods.

Discussion of Methods

Desirable attributes of a proposed alternative method as described in Engel and others (1984) are: (1) The proposed method needs to be computer oriented and easy to apply, (2) the proposed method needs to have an available interface with the U.S. Geological Survey's WATSTORE Daily Values File (Hutchison, 1975), (3) the proposed method needs to be technically sound and generally acceptable to the hydrologic community, and (4) the proposed method needs to provide a measure of the accuracy of the simulated streamflow records. Because of the short duration of this analysis, only two methods were considered--regression analysis and hydrologic flow-routing.

Stations in the Wyoming streamflow-gaging program were screened to determine their potential for use of alternative methods, and selected methods were applied to the non-winter period at those stations where the potential was large. The applicability of alternative methods to specific streamflow-gaging stations is described in this section of this report.

Description of Regression Analysis

Simple- and multiple-regression techniques can be used to estimate daily flow records. Unlike hydrologic flow-routing, regression methods are not limited to locations where a station exists upstream from another station on the same stream. Regression equations can be computed that relate daily flows (or their logarithms) at a station (dependent variable) to daily flows at an index station or at a combination of upstream, downstream, or tributary index stations. The regression analysis can include independent variables computed for stations from different watersheds.

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

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

$$Y_i = B_o + \sum_{j=1}^{n} B_j X_j + e_i \qquad (1)$$

where

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

B_o = regression constant;

B_j = regression coefficient; and

X_j = daily mean discharge at index station (independent variable), with j indicating index stations i through n;

e_i = the random error term.

Equation 1 is calibrated (B_o and B_j are estimated) using observed values of Y_i and X_j. These observed daily mean discharges can be retrieved from the WATSTORE Daily Values File (Hutchison, 1975). The values of discharge for the index station may be observed on the same day as discharges at the independent station or may be for previous or future days, depending on whether the index station j is upstream or downstream from the independent station. During calibration, the regression constant (B_o) and coefficient (B_j) are tested to determine if they are significantly different from zero. A given independent variable is retained in the regression equation only if its regression coefficient is significantly different from zero.

The regression equation needs to be calibrated using one period of time and verified or tested using a different period of time to obtain a measure of the true predictive accuracy. Streamflow used for both the calibration and verification periods needs to be representative of the expected range of flows. The equation can be verified by: (1) Plotting the residuals (difference between simulated and observed discharges) versus both the dependent and the independent variables in the equation, and (2) plotting the simulated and observed discharges versus time. These tests are needed to confirm that the linear model is appropriate and that neither the date nor the variables in the equation change with time. The presence of either nonlinearity or bias requires that the data be transformed (for example, converted to logarithms) or that a different form of model be used.

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

Description of Flow-Routing Model

Hydrologic flow-routing methods use the law of conservation of mass and the relationship between the storage in a reach and the outflow from the reach. The hydraulics of the system are not considered. The methods usually require only a few parameters, and the reach is not subdivided. A discharge hydrograph is required at the upstream end of the reach, and the computations produce a discharge hydrograph at the downstream end. Hydrologic flow-routing methods include the Muskingum, Modified Puls, Kinematic Wave, and unit-response. The unit-response method uses one of two routing techniques--storage continuity (Sauer, 1973) and diffusion analogy (Keefer, 1974, and Keefer and McQuivey, 1974).

A computer program that uses the unit-response method to route streamflow from one or more upstream locations to a downstream location is available (Doyle and others, 1983). The model, referred to as CONROUT, treats a stream reach as a linear one-dimensional system in which the downstream hydrograph is computed by multiplying (convoluting) the ordinates of the upstream hydrograph by the unit-response function and lagging them appropriately. The model has the capabilities of combining hydrographs, multiplying a hydrograph by a ratio, and changing the timing of a hydrograph.

Daily flows usually can be routed using a single unit-response function (linearization about a single discharge) to represent the system response. However, if the routing coefficients vary significantly with discharge, linearization about a low-range discharge results in overestimated high flows that arrive late at the downstream site, and linearization about a high-range discharge results in low-range flows that are underestimated and arrive too soon. Multiple linearization (Keefer and McQuivey, 1974), in which separate unit-response functions are defined for different ranges of discharge, minimizes this problem.

Determination of the system's response to an upstream pulse is not the total solution for most flow-routing problems. The convolution process makes no accounting of flow from the intervening area between the upstream and downstream locations. Ungaged inflows usually are estimated by multiplying known flows at an index-gaging station by an adjustment factor (for example, the ratio of drainage area at the point of interest to that at the index gage).

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

Potential for Use of Alternative Methods

A two-level screening process was applied to gaging stations in Wyoming to evaluate the potential for use of alternative methods. The first-level screening was based only on hydrologic considerations, and the only concern at this level was whether it was hydrologically possible to simulate streamflow at a given station from streamflow at one or more other stations. The first-level screening was subjective; there was no attempt to apply mathematical procedures. For stations that passed the first-level screening, the second-level screening was used to determine whether simulated streamflow data would be acceptable according to data uses described by Schuetz (1986). Even if simulated streamflow data were not acceptable for the given data uses, the analysis continued. Mathematical procedures were applied to determine if it were technically possible to simulate streamflow data. This was done under the assumption that the data uses may change in the future. However, where data uses required continued streamflow gaging, the result was predetermined to require continued operation, even though alternative methods were technically possible.

Correlation methods were used as the second level of screening to determine the acceptability of gaging-station combinations for the regression analysis. Combinations of stations with concurrent, non-winter (April 1 to September 30) daily discharge record and that passed the first level of screening are listed in table 1. Table 1 includes the station whose record is being simulated from one or more index stations, the index stations, and the lag-time of streamflow between the two stations. The location of these stations are shown in figure 1. Correlation coefficients were determined for the combinations of stations shown in table 1. Combinations of stations with a correlation coefficient of more than 0.90 passed the second level of screening and were used in the regression analysis. Stations for which the initial results were unacceptable (less than 0.90 correlation coefficient) were eliminated from further consideration. Combinations of gages produced unacceptable results (less than 0.90 correlation coefficient) for several reasons. However, the most common reasons relate to the wide variability of runoff from streams originating in the mountains and flowing through the arid and semiarid parts of Wyoming and to variable effects of diversions, numerous small reservoirs, and irrigation return flows.

Regression Results

The regression equations and a measurement of accuracy of the similated discharge are presented in table 2. Two periods, representative of the range of flows, were used in the regression analysis, water years 1979-81 and 1982-84. The period 1982-84 was used for calibration and the period 1979-81 was used for verification. Results are presented for both the maintenance-of-variance model (MOVE.1) suggested by Hirsch (1982) and the ordinary least-squares model (OLS). All variables used in the regression equations in the table are statistically significant at the 0.01 level. The standard error of estimate for models, in the units of cubic feet per second, is not useful directly because the data are not homoscedastic (variance not constant throughout range of flow). Therefore, the individual errors were converted to percentage deviations, and the standard deviation of those percentage values was defined. The percentage deviations reported are plus or minus 5, 15, 25, and 40 percent of the observed discharge.

Streamflow can be simulated by regression analysis with sufficient accuracy for North Fork Crazy Woman Creek near Buffalo (06314000) but not for Laramie River near Lookout (06662000). The relation for North Fork Crazy Woman Creek tends to overpredict for discharges greater than 80 ft^3/s. That bias could be reduced and the accuracy improved by modifying the form of the relation or by defining separate relations for discharges greater than 80 ft^3/s. The two gages on North Fork Crazy Woman Creek (both discontinued in 1984) were used primarily for determining streamflow gain or loss through the reach they bracketed. Actual streamflow records were required for determining the gain or loss, and simulated records would have been unsatisfactory. The Laramie River relations are not valid for discharges of about 50 ft^3/s or less. Even if errors for discharges less than 50 ft^3/s were omitted, the regression analysis still does not reflect sufficient accuracy to warrant discontinuance of either gage.

Table 1.—**Gaging-station combinations considered in alternative-methods analysis, and associated lag-time and correlation coefficient for April 1 to September 30**

Gaging-station number and name (Index stations are indented)	Lag-time of daily discharge (days)	Correlation coefficient
06218500 Wind River near Dubois		0.6046
06280300 South Fork Shoshone River near Valley	0	
06220500 East Fork Wind River near Dubois		.6386
06218500 Wind River near Dubois	0	
06222700 Crow Creek near Tipperary		.4824
06260000 South Fork Owl Creek near Anchor	0	
06225500 Wind River near Crowheart		.8001
06218500 Wind River near Dubois	-3	
06220500 East Fork Wind River near Dubois	-2	
06225000 Bull Lake Creek near Lenore	0	
06233900 Popo Agie River near Arapahoe		.7405
06235500 Little Wind River near Riverton	0	
06276500 Greybull River at Meeteetse		.7357
06275000 Wood River at Sunshine	0	
06278500 Shell Creek near Shell		.2528
06278300 Shell Creek above Shell Reservoir	0	
06280000 North Fork Shoshone near Wapiti		.7148
06281000 South Fork Shoshone River above Buffalo Bill Reservoir	0	
06281000 South Fork Shoshone River above Buffalo Bill Reservoir		.7628
06280300 South Fork Shoshone River near Valley	0	
06299500 Wolf Creek at Wolf		.3910
06301500 West Fork Big Goose Creek near Big Horn	0	
06300500 East Fork Big Goose Creek near Big Horn		.5724
06301500 West Fork Big Goose Creek near Big Horn	0	
06309460 Beaver Creek above White Panther Ditch, near Barnum		.5847
06309200 Middle Fork Powder River near Barnum	0	

Table 1.—Gaging-station combinations considered in alternative-methods analysis, and associated lag-time and correlation coefficient for April 1 to September 30—Continued

Gaging-station number and name (Index stations are indented)	Lag-time of daily discharge (days)	Correlation coefficient
06309450 Beaver Creek below Bayer Creek, near Barnum		0.7785
06309460 Beaver Creek above White Panther Ditch, near Barnum	0	
06313500 Powder River at Sussex		.6906
06317000 Powder River at Arvada	0	
06314000 North Fork Crazy Woman Creek near Buffalo		.9674
06313950 North Fork Crazy Woman Creek below Pole Creek, near Buffalo	0	
06622900 South Brush Creek near Saratoga		.7249
06622700 North Brush Creek near Saragota	0	
06623800 Encampment River above Hog Park Creek, near Encampment		.5242
06625000 Encampment River at mouth, near Encampment	0	
06634600 Little Medicine Bow River near Medicine Bow		.6411
06635000 Medicine Bow River above Seminoe Reservoir, near Hanna	0	
06639000 Sweetwater River near Alcova		.7245
06638090 Sweetwater River near Sweetwater Station	0	
06661000 Laramie River near Filmore		.5666
06632400 Rock Creek above King Canyon Canal, near Arlington	0	
06662000 Laramie River near Lookout		.9454
06661585 Laramie River near Bosler	0	
09203000 East Fork River near Big Sandy		.7459
09196500 Pine Creek above Fremont Lake	0	
9209400 Green River near La Barge		.7974
09188500 Green River at Warren Bridge, near Daniel	-2	
09205000 New Fork River near Big Piney	-1	
9212500 Big Sandy River at Leckie Ranch, near Big Sandy		.7718
09203000 East Fork River near Big Sandy	0	

Table 2.--Accuracy of regression analyses for mean daily streamflow from April 1 to September 30

[Stations: 06314000, North Fork Crazy Woman Creek near Buffalo; 06313950, North Fork Crazy Woman Creek below Pole Creek near Buffalo; 06662000, Laramie River near Lookout; 06661585, Laramie River near Bosler]

Regression equation[1]		Percent daily discharge within indicated percentage deviation							
		Calibration period, water years 1982-84				Verification period, water years 1979-81			
Independent station	Index station	Percentage deviations (+ or -)							
		5	15	25	40	5	15	25	40
MOVE.1 model									
$Y_{06314000} = -0.42 + 1.002$	$X_{06313950}$	52	90	97	97	49	87	99	100
$Y_{06662000} = 23.6 + 0.982$	$X_{06661585}$	22	58	81	93	12	39	55	65
Ordinary least-squares model									
$Y_{06314000} = -9.23 + 0.998$	$X_{06313950}$	49	90	97	99	47	88	98	100
$Y_{06620000} = 29.7 + 0.972$	$X_{06661585}$	22	57	79	92	11	33	50	61

[1] Terms defined in equation 1.

Flow-Routing Results

Hydrologic flow-routing of daily flows generally gives results that are comparable to regression results. Therefore, because regression results were satisfactory for North Fork Crazy Woman Creek near Buffalo, only the flow-routing model, CONROUT, was used to simulate flow in a reach of the Laramie River between the Bosler (06661585) and Lookout (06662000) gages. The CONROUT model (Doyle and others, 1983) using the diffusion-analogy method of routing requires three parameters to describe flow; they are:

X = routing distance, in miles;

C_o = flood-wave celerity (controls travel time), in feet per second;

K_o = dispersion or damping coefficient (controls spreading of the wave), in square feet per second.

C_o and K_o are approximated from the following equations:

$$C_o = (1/W_o) \quad (dQ_o/d_y) \tag{2}$$

$$K_o = Q_o \;/\; (2\; S_o W_o) \tag{3}$$

where

W_o = average channel width, in feet, in the reach;

dQ_o/d_y = the slope of the stage-discharge curve;

Q_o = the stream discharge of interest, in cubic feet per second; and

S_o = average streambed slope, in feet per foot, in the reach.

These parameters were estimated for the reach of the Laramie River between the Bosler (06661585) and Lookout (06662000) gages and were refined--based on application of the model to the calibration period, 1982-84. No winter records have been collected at the Lookout gage (06662000) since 1972; therefore, the analysis used the April 1 to September 30 period for each year. The calibrated model was then used to simulate mean daily discharges for the verification period, 1979-81. The final values for the parameter describing the flow values were:

X = 25.6 miles;

C_o = 3.00 feet per second; and

K_o = 10,000 square feet per second.

The net contributing drainage areas are 1,507 square miles for Bosler and 1,571 square miles for Lookout (Druse and Rucker, 1985). Because there is little contributing area between these gages, the model was used to route the flow at Bosler to Lookout. Results of the calibration and verification are shown in table 3.

Table 3.—Accuracy of calibration and verification results for the flow-routing model as applied to the reach of the Laramie River between the Bosler (06661585) and Lookout (06662000) gages

Daily-discharge errors	Percent of daily discharges within given daily-discharge error Calibration 1982-84	 Verification 1979-81
Less than, or equal to, 5 percent	25	10
Less than, or equal to, 10 percent	52	23
Less than, or equal to, 15 percent	64	36
Less than, or equal to, 20 percent	76	43
Less than, or equal to, 25 percent	84	49
Greater than 25 percent	16	51
Total volume error	-0.9	-3.2

The results from the flow-routing model compare well with the results obtained by regression analysis (table 2). Discharges calculated by the flow-routing model are generally within 10 percent of the observed flow when discharge in the Laramie River near Lookout is greater than about 500 ft^3/s; however, percentage errors are significantly greater for daily flows less than 500 ft^3/s. The major difference in the degree of accuracy between the calibration period and verification period is a function of the average discharge during each period. The average discharge from April 1 to September 30 was 642 ft^3/s during 1982-84 and 203 ft^3/s during 1979-81.

Summary of Phase Two of Analysis

One of the stations investigated was found suitable for the application of alternative methods. Only at North Fork Crazy Woman Creek near Buffalo (06314000) is the accuracy of the regression relation sufficient to consider discontinuing the gage; however, the data uses in 1984 and in prior years required that both gages be operated.

COST-EFFECTIVE RESOURCE ALLOCATION

Discussion of the Model

A set of techniques known as the K-CERA (Kalman filtering for Cost-Effective Resource Allocation) model was developed by Moss and Gilroy (1980) for studying the cost-effectiveness of networks of streamflow gages. Engel and others (1984) noted the original application of the techniques was used to analyze a network of streamflow gages being operated to determine water consumption in the Lower Colorado River Basin (Moss and Gilroy, 1980). Because of the water-balance orientation of that study, the minimization of the total variance of errors of estimation of annual mean discharges was chosen as the measure of effectiveness of the network. This total variance is defined as the sum of the variances of errors of mean annual discharge at each site in the network. This measure of effectiveness tends to concentrate streamflow-gaging resources on the large rivers and streams where discharge and, consequently, potential errors (in cubic feet per second) are greatest. Although this measure may be acceptable for a water-balance network, considering the many uses of data collected by the Geological Survey, concentration of effort on large rivers and streams is undesirable and inappropriate.

The original version of K-CERA was therefore altered to include as optional measures of effectiveness, the sums of the variances of errors of estimation of the following streamflow variables: annual mean discharge, in cubic feet per second; annual mean discharge, in percent; average instantaneous discharge, in cubic feet per second; or average instantaneous discharge, in percent (Fontaine and others, 1984). The use of percentage errors effectively gives equal weight to both large and small streams. In addition, instantaneous discharge is the basic variable from which all other streamflow data are derived. For these reasons, this study used the K-CERA techniques with the sums of the variances of the percentage errors of the instantaneous discharges at continuously gaged sites as the measure of the effectiveness of the data-collection activity.

The original version of K-CERA also did not account for error contributed by lost stage or other correlative data that are used to compute streamflow data. The probabilities of missing correlative data increase as the period between service visits to a stream gage increases. A procedure for dealing with the lost record was developed (Fontaine and others, 1984) and was incorporated into this study.

Brief descriptions of uncertainty functions and of the mathematical program used to minimize the total error variance of the data-collection activity for given budgets and of the application of Kalman filtering (Gelb, 1974) to the determination of the accuracy of a streamflow-gaging record are presented by Fontaine and others (1984); a modified version of that description is provided in the Supplemental Information section at the end of this report. For this study, uncertainty at each site is defined as the variance of the percent error in the instantaneous discharge. More detail on either the theory or the applications of the K-CERA model is provided by Moss and Gilroy (1980) and Gilroy and Moss (1981).

Application of the Model in Wyoming

Phases one and two of this analysis showed that operation of the current network of stream gages in Wyoming needs to be continued. Phase three of the analysis was limited to the network stations operated by the Riverton Field Office. The evaluation of that network, using the K-CERA model, was considered adequate to address the cost-effectiveness of the overall streamflow- gaging program in Wyoming and to provide a basis for considering changes in operating procedures.

The model assumes the uncertainty of discharge records at a given gage to be derived from three sources: (1) Errors that result because the stage-discharge relationship is not perfect (applies when the gage is operating); (2) errors in reconstructing records based on records from another gage when the primary gage is not operating; and (3) errors inherent in estimated discharge when the gage is not operating and correlative data are not available to aid in record reconstruction. These uncertainties are measured as the variance of the percentage errors in instantaneous discharge and vary in magnitude relative to the number of visits and discharge measurements made per season at a gage. The proportion of time that each source of error applies is dependent on the frequency interval at which the equipment is serviced.

Definition of Variance When Station is Operating

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

Computation of the error variance about the stage-discharge relation was done in three steps. A long-term rating was defined, generally based on measurements made during 3 or more water years, and deviations (residuals) of the measured discharges from the rating discharge were determined. A time-series analysis of these residuals defined the 1-day lag (lag-one) autocorrelation coefficient and the process variance required by the K-CERA model. Finally, the error variance is defined within the model as a function of: the lag-one autocorrelation coefficient, the process and measurement variances, and the frequency of visits and discharge measurements made at a gage.

In the Wyoming program analysis, definition of long-term rating functions was complicated by the fact that most stream gages in Wyoming are affected by backwater from ice for about 5 months during the year. Rating curves based on open-water measurements are not applicable during the ice-affected periods.

In the Maine pilot study, winter rating curves were replaced with regression relations relating the discharge at the ice-affected station to the discharge at an ice-free station, stage, climatic variables, and reservoir releases. The model used this relationship in place of a standard stage-discharge relationship, and uncertainties of the ice-affected and ice-free periods were evaluated separately (Fontaine and others, 1984). This approach does not work well in Wyoming because of the distances between gages and the variability of flow resulting from the temporary storage and subsequent release of ice. Reliable discharge records during the winter can presently be produced only by making periodic visits and discharge measurements to document the degree of ice-effect.

Review of past discharge records indicates that the average period of significant ice-effect lasts about 5 months in Wyoming, generally from November through March. The model was applied only to the approximately 7 months (214 days) that are virtually free from ice-effect. The study also assumed that, regardless of ice-free period visit requirements, 3 visits will continue to be made during the winter season.

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

$$LQM = B1 + B3\ (LOG\ (GHT - B2)) \tag{4}$$

where

LQM = the logarithmic (base 10) value of the measured discharge, and

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

The constants B1, B2, and B3 were determined by a non-linear regression procedure (Helwig and Council, 1979) and have the following physical interpretation: B1 is the logarithm of discharge for a flow depth of 1 ft; B2 is the gage height of zero flow; and B3 is the slope of the rating curve.

The residuals about the long-term rating for individual gages defined the total variance. A review of discharge measurements made in Wyoming indicated that the average standard error of open-water measurements was about 3.5 percent. The measurement variance for all gages, therefore, was defined as equal to the square of the 3.5-percent standard error. The process variance required in the model is, thus, the variance of the residuals about the long-term rating minus the constant measurement variance.

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

Definition of Variance When Record is Lost

When stage record is lost at a gaging station, the model assumes that the discharge record is either reconstructed using correlation with another gage or is estimated from historical discharge for that period. Fontaine and others (1984, p. 24) indicated that the fraction of time for which a record must be either reconstructed or estimated can be defined by a single parameter in a probability distribution of times-to-failure of the equipment. The reciprocal of the parameter defines the average time, since the last servicing visit, to failure. The value of average time-to-failure varies from site to site, depending on the type of equipment at the site and on exposure to natural elements and vandalism. In addition, the average time-to-failure can be changed by advances in the technology of data collection and recording equipment.

Data collected in Wyoming during 1984 and 1985 were reviewed to define the average time-to-failure for recording equipment and stage-sensing devices. Little change in technology occurred during the period examined, and streamflow gages were visited in a consistent pattern of about 10 visits per year. Gages were malfunctioning an average of about 6 percent of the time.

Because the K-CERA model analysis in Wyoming was confined to a 7-month non-winter period, there was no reason to distinguish differences between gages on the basis of exposure or equipment. The 6-percent lost record and a visit frequency of 7 times in 7 months (214 days) were used to determine an average time-to-failure of 417 days after the last visit. This average time-to-failure was used to determine the fractions of time, as a function of the frequency of visits, that each of the three sources of uncertainty were applicable for individual streamflow gages.

The model defines the uncertainty as both the sum of the multiples of the fraction of time each error source (rating, reconstruction, or estimation) is applicable and the variance of the error source. The variance associated with reconstruction and estimation of a discharge record is a function of the coefficient of cross correlation with the station(s) used in reconstruction and the coefficient of variation of daily discharges at the station. Daily streamflows for the last 30 water years 1956 to 1985 (or period of record if less) were used to define seasonally-averaged coefficients of variation for each station (table 4). In addition, cross-correlation coefficients (with seasonal trends removed) were determined for various combinations with other stations.

Many different sources of information are used in reconstructing periods of lost record. These sources include, but are not limited to, recorded ranges in stage (for graphic recorders with clock stoppage), known discharges on adjacent days, recession analysis, observer's staff-gage readings, weather records, highwater-mark elevations, and comparison with nearby stations. However, most of these techniques are unique to a given station or to a specific period of lost record. Using all the information available, short periods (several days) of lost record usually can be reconstructed quite accurately. An even longer period (more than a month) of lost record can be reconstructed with reasonable accuracy if observer's readings are available. If none of these data are available, however, lengthy reconstructions can be subject to large errors. This study could not reasonably quantify the uncertainty associated with all the possible methods of reconstructing lost record at the individual sites.

Discussion of Variance at Wyoming Gages

The values of lag-one autocorrelation coefficient, process variance, and coefficient of variation are listed in table 4; length of period (214 days), crosscorrelation coefficients between the gage and an index gage(s) (selected correlation coefficients shown in table 1), and data from the definition of missing record probabilities are used jointly to define uncertainty functions for each gaging station. The uncertainty functions give the relation of error variance to the number of visits, assuming a measurement is made at each visit. Examples of typical uncertainty functions are given in figure 2. The uncertainty curve reflects a high-process variance and high coefficient of variation for station 06260400, a high-process variance and low coefficient of variation for station 06231000, and a low-process variance and high coefficient of variation for station 09224700. Both 06279500 and 09196500 are representative curves for stations with low-process variance and low coefficient of variation. Lag-one autocorrelation coefficient is 0.96 or greater for all five stations.

Table 4.--Gaging stations and summary of statistics used to define uncertainty functions

[Daily streamflows for the last 30 water years (or period of record record if less) were used to define seasonally averaged statistics. Process variance units are base 10 logarithms]

Gaging-station number and name		Lag-one auto-correlation coefficient	Process variance	Coefficient of variation
06218500	Wind River near Dubois	0.985	0.00071	0.38
06220500	East Fork Wind River near Dubois	.985	.00502	.53
06222700	Crow Creek near Tipperary	.970	.01641	.59
06224000	Bull Lake Creek above Bull Lake	.958	.00074	.47
06225000	Bull Lake Creek near Lenore	.971	.00055	1.03
06225500	Wind River near Crowheart	.993	.00171	.38
06228000	Wind River at Riverton	.979	.00079	.65
06228350	South Fork Little Wind River above Fort Washakie	.963	.00132	.45
06231000	Little Wind River above Arapahoe	.967	.01696	.54
06233900	Popo Agie River near Arapahoe	.973	.00181	.48
06235500	Little Wind River near Riverton	.992	.00274	.58
06259000	Wind River below Boysen Reservoir	.960	.00009	.45
06260400	South Fork Owl Creek below Anchor Reservoir	.978	.01042	1.01
06270000	Nowood River near Tensleep	.990	.00308	.59
06274300	Bighorn River at Basin	.975	.00032	[1].50
06275000	Wood River at Sunshine	.947	.07667	.71
06278300	Shell Creek above Shell Reservoir	.984	.00689	.79
06279500	Bighorn River at Kane	.990	.00022	.45
06280000	North Fork Shoshone River near Wapiti	.999	.00354	.36
06280300	South Fork Shoshone River near Valley	.966	.00032	.43
06281000	South Fork Shoshone River above Buffalo Bill Reservoir	.989	.01726	.55
06281400	Diamond Creek near Cody	.970	.01503	.65
06282000	Shoshone River below Buffalo Bill Reservoir	.983	.00053	.48
06284500	Bitter Creek near Garland	.975	.00092	.38
06284800	Whistle Creek near Garland	.971	.00205	.94
06285100	Shoshone River near Lovell	.989	.00568	.61
06285400	Sage Creek at Sidon Canal, near Lovell	0.960	0.00057	0.45

Table 4.--Gaging stations and summary of statistics used to define uncertainty functions--Continued

Gaging-station number and name	Lag-one auto-correlation coefficient	Process variance	Coefficient of variation
06637750 Rock Creek above Rock Creek Reservoir	.960	.00034	.53
06638090 Sweetwater River near Sweetwater Station	.981	.00129	.60
09188500 Green River at Warren Bridge, near Daniel	.930	.00067	.36
09196500 Pine Creek above Fremont Lake	.987	.00281	.53
09203000 East Fork River near Big Sandy	.960	.00094	.63
09205000 East Fork River near Big Piney	.883	.00038	.47
09209400 Green River near La Barge	.976	.00124	.41
09210500 Fontenelle Creek near Herschler Ranch	.971	.00720	.52
09211200 Green River below Fontenelle Reservoir	.971	.00049	.53
09212500 Big Sandy River at Leckie Ranch, near Big Sandy	.981	.00050	.57
09213500 Big Sandy River near Farson	.947	.00154	.66
09215550 Big Sandy River below Farson	.985	.00349	.87
09216050 Big Sandy River at Gasson Bridge, near Eden	.984	.00424	.80
09217000 Green River near Green River	.996	.00119	.55
09218500 Blacks Fork near Millburne	.987	.00295	.71
09223000 Hams Fork below Pole Creek, near Frontier	.970	.00461	.59
09224700 Blacks Fork near Little America	.981	.00167	.89
09229500 Henrys Fork near Manila, Utah	.972	.00280	.79
10016900 Bear River at Evanston	.991	.00036	[1].60

[1] Estimated, data insufficient to calculate the value.

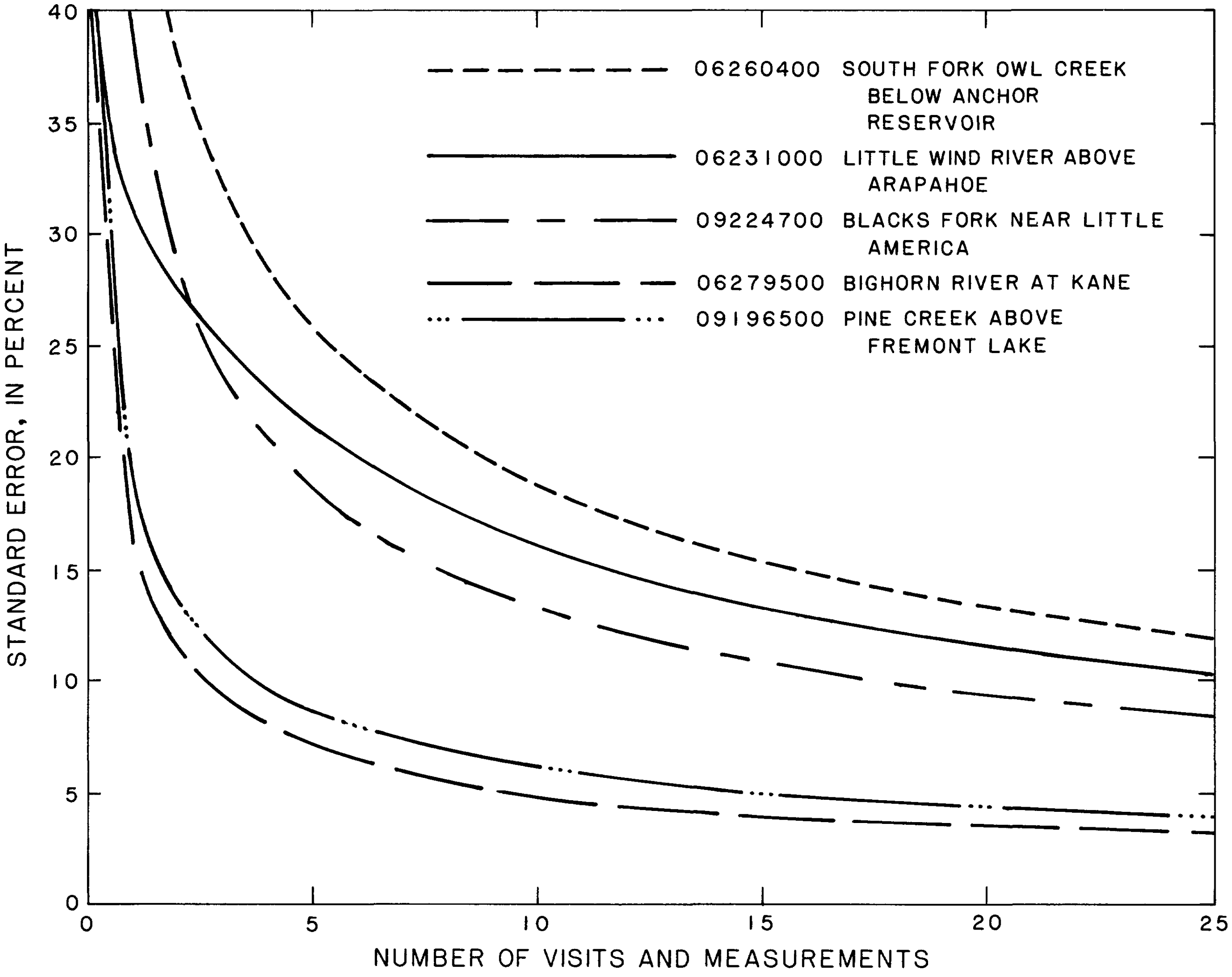


Figure 2.--Typical uncertainty functions for instantaneous discharge.

The residuals about rating curves for many stations in Wyoming only poorly approximate a continuous first-order Markov process. These stations have moderate-to-significant changes in ratings resulting from channel changes that usually result from periodic floods. These may shift with each flood but will not necessarily return to the original rating after a change. The process may be Markovian but is not continuous because no meaningful, long-term rating exists. Additionally, station 06267400 was excluded from the analysis because of insufficient discharge measurements for meaningful statistical computations.

Discussion of Routes and Costs

Forty-seven continuous-record surface-water stations in the Riverton field area, as well as several ground-water observation wells and water-quality stations, are serviced on field trips. The observation wells and water-quality stations are considered to be null stations because they do not add uncertainty to the network. Additionally, the operating budgets for these other types of stations, and associated costs, are not included in the surface-water operating budget being analyzed in phase three.

As previously indicated, uncertainty functions could not be defined for one (station 06267400) of the 47 continuous-record surface-water stations. This station was treated as a null station with the exception that all operating costs were included in the analysis.

Minimum visit constraints were defined for each of the 47 stations prior to defining the practical service routes. A minimum of two visits was established for all stations in the network (including the null station) where there was no uncertainty determined, for the 214-day study period. Two visits were regarded as the minimum necessary to minimally maintain equipment and to get a discharge measurement at a high stage during the snowmelt runoff period. However, only two visits during the period would probably lead to increased incidence of equipment failure. Current practice does require special trips to be made to make discharge measurements; however, no special trips are necessary for cooperator requirements.

Practical routes to service the 47 stations were determined after consultation with personnel responsible for maintaining the stations and after the uncertainty functions and minimum visit requirements were considered. Forty-four routes to service all the stream gages in the Riverton field area were identified. These included: routes that describe the current operating practice; alternate routes under consideration as future possibilities; routes used to service certain key stations; and route combinations grouping proximate gages having levels of uncertainty indicating that more frequent visits might be useful.

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

Fixed costs typically include charges for equipment rental, batteries, electricity, data processing and storage, maintenance, and miscellaneous supplies, in addition to supervisory charges and the costs of computing the record. The fixed costs were average values for the statewide network.

Visit costs are those associated with paying the hydrographer for the time actually spent making a discharge measurement and servicing equipment. These costs vary from station to station depending on the difficulty incurred in making the measurement which can vary because of channel configuration, uniformity of flow, and whether a wading or cable-type measurement is generally made. Average visit time, in hours, was estimated for each station, based on historical operations. Average number of hours was then multiplied by the average hourly salary of the hydrographers in Wyoming to determine visit costs for each station.

Route costs include the vehicle cost associated with driving the number of miles required to cover the route, the cost of the hydrographer's time while in transit, and any per diem (1987 dollars) associated with the time needed to complete the trip.

The model was run on a 214-day period with the added requirement that 3 visits would be made during the remaining 151 days of the year. The fixed costs were computed on an annual basis, but the visit and route costs are only applied when a trip is made. In order for all costs to be applied on an annual basis, the visit and route costs for the three winter visits to each station were added to the fixed cost for each station.

Results

The "Traveling Hydrographer Program" (Moss and Gilroy, 1980) uses the uncertainty functions along with the appropriate cost data, route definitions, and minimum visit constraints to optimize the operation of the streamflow-gaging program. The objective function in the optimization process is the sum of the variances of the errors of instantaneous discharge (in percent) for the entire gaging-station network.

Present (1987) practices to define the associated, total uncertainty were simulated by restricting the specific routes and number of visits to only streamflow gages now being used. This was done only to compute the standard errors of present practice and calibrate the model; no optimization was done. The restrictions were then removed, and the model was allowed to determine optimal visit schedules for the current budget. The optimization procedure was repeated for other possible budgets. The results for both the present operation and the optimal solutions are shown in figure 3 and in table 5.

The Equivalent Gaussian Spread (EGS) shown in table 5 was introduced by Fontaine and others (1984, p. 26); the definition is included in the Supplemental Information section of this report. The approximate interpretation of EGS is, "Two-thirds of the errors in instantaneous streamflow data will be within plus or minus EGS percent of the reported value."

The analysis was repeated for each budget under the assumption that no stage record is lost. Analysis results indicate the average standard error of estimate for instantaneous discharge attainable when perfectly reliable systems are available to measure and record stage.

The results in figure 3 and table 5 are based on the assumption that a discharge measurement is made each time a station is visited. The percentage values also represent only the 7 months that are virtually free from ice effect. No estimate is made of the probable errors during ice-affected periods. The curve in figure 3 indicating "with lost record" represents the minimum level of uncertainty that can be obtained, with existing technology for a given budget. Additional assumptions to consider when interpreting the results is the applicability of the Markov process to all stations.

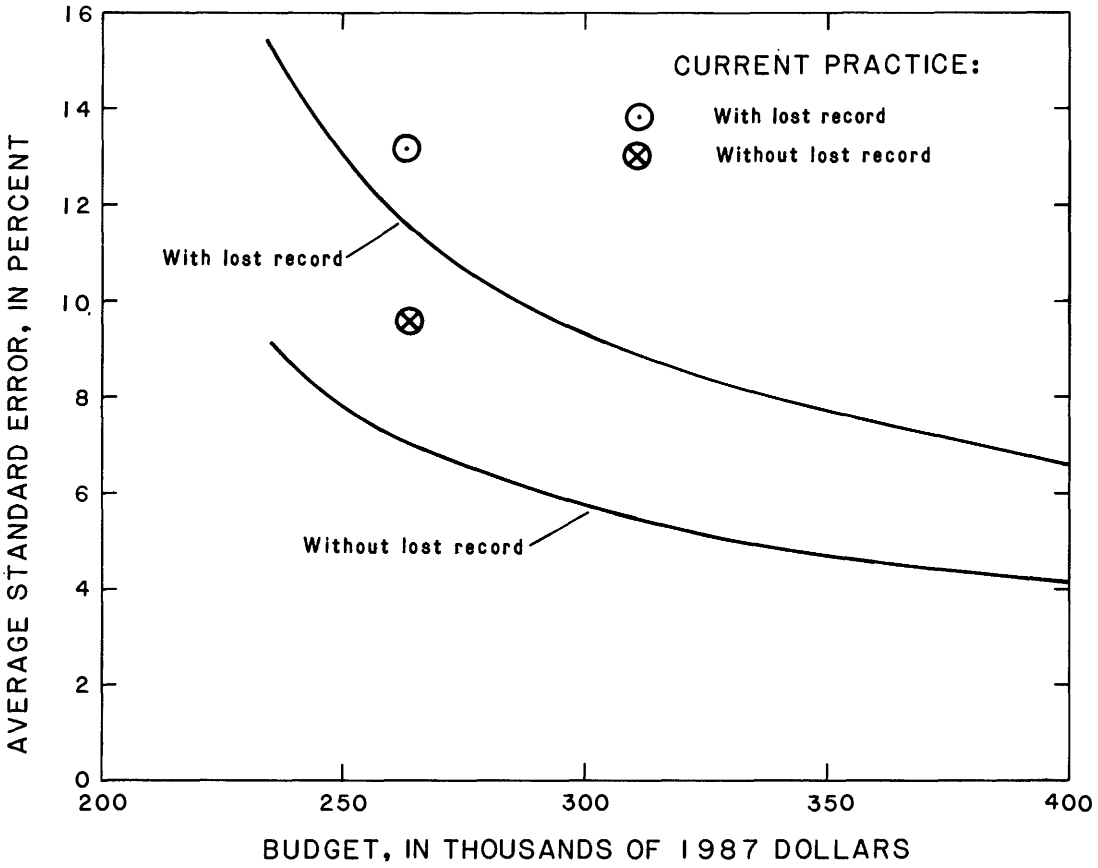


Figure 3.--Relation between average standard error per station and budget.

Table 5.--Selected results of the analysis

[SE, standard error of instantaneous discharge in percent; EGS, Equivalent Gaussian Spread; visits, optimized number of visits per study period to site]

Identification	Budget, in thousands of 1987 dollars						
	Current (1987)	Alternative values					
	264	235	250	280	310	350	400
	Average values for the network						
Average SE per station[1]	13.2	15.5	13.1	10.4	8.9	7.7	6.6
EGS for the program	5.1	7.2	6.0	4.8	4.1	3.5	3.1
	Average values of SE, [EGS], and (visits) for individual stations						
06218500	6.5	9.6	7.6	6.0	5.4	4.6	4.0
	[2.8]	[4.3]	[3.4]	[2.7]	[2.4]	[2.1]	[1.8]
	(7)	(3)	(5)	(8)	(10)	(14)	(19)
06220500	10.3	15.1	12.0	9.6	8.6	7.3	6.3
	[7.1]	[10.8]	[8.4]	[6.6]	[5.9]	[5.0]	[4.3]
	(7)	(3)	(5)	(8)	(10)	(14)	(19)
06222700	18.7	21.5	16.7	14.1	12.0	10.6	9.1
	[17.4]	[20.1]	[15.4]	[12.9]	[10.9]	[9.6]	[8.2]
	(7)	(5)	(9)	(13)	(18)	(23)	(32)
06224000	6.2	8.8	7.8	5.6	4.9	4.2	3.7
	[4.3]	[5.8]	[5.3]	[3.9]	[3.5]	[3.0]	[2.6]
	(7)	(3)	(4)	(9)	(12)	(17)	(23)
06225000	16.8	18.1	15.8	12.4	10.6	9.0	8.0
	[3.3]	[3.6]	[3.1]	[2.5]	[2.2]	[1.8]	[1.6]
	(7)	(6)	(8)	(13)	(18)	(25)	(32)
06225500	6.0	7.9	6.0	5.1	4.3	3.9	3.2
	[3.0]	[4.0]	[3.0]	[2.5]	[2.1]	[1.9]	[1.6]
	(7)	(4)	(7)	(10)	(14)	(17)	(26)
06228000	8.9	11.6	10.4	8.3	6.6	6.0	5.2
	[3.4]	[4.4]	[4.0]	[3.2]	[2.6]	[2.3]	[2.0]
	(7)	(4)	(5)	(8)	(13)	(16)	(21)
06228350	7.7	10.8	9.7	8.3	6.6	6.1	5.2
	[5.5]	[7.4]	[6.8]	[5.9]	[4.7]	[4.4]	[3.7]
	(7)	(3)	(4)	(6)	(10)	(12)	(17)

Table 5.--Selected results of the analysis--Continued

	Budget, in thousands of 1987 dollars						
	Current (1987)	Alternative values					
Identification	264	235	250	280	310	350	400
	Average values of SE, [EGS], and (visits) for individual stations--Continued						
06231000	18.6	17.6	14.7	11.8	10.1	9.0	8.3
	[18.2]	[17.2]	[14.2]	[11.3]	[9.6]	[8.3]	[7.9]
	(7)	(8)	(12)	(19)	(26)	(35)	(38)
06233900	7.5	10.6	9.5	8.0	6.7	5.9	5.0
	[5.6]	[7.9]	[7.1]	[6.0]	[5.0]	[4.4]	[3.7]
	(7)	(3)	(4)	(6)	(9)	(12)	(17)
06235500	7.4	9.7	9.7	7.4	6.2	5.5	4.6
	[3.9]	[5.2]	[5.2]	[3.9]	[3.3]	[2.9]	[2.4]
	(7)	(4)	(4)	(7)	(10)	(13)	(19)
06259000	5.8	10.5	8.7	6.8	5.5	4.7	4.2
	[1.5]	[2.3]	[2.0]	[1.7]	[1.4]	[1.3]	[1.2]
	(7)	(2)	(3)	(5)	(8)	(11)	(14)
06260400	22.2	22.2	19.6	15.8	13.3	11.7	9.9
	[12.3]	[12.3]	[10.8]	[8.6]	[7.2]	[6.3]	[5.3]
	(7)	(7)	(9)	(14)	(20)	(26)	(36)
06270000	11.7	17.5	15.3	11.7	10.4	8.7	7.6
	[4.6]	[7.3]	[6.2]	[4.6]	[4.1]	[3.4]	[3.0]
	(7)	(3)	(4)	(7)	(9)	(13)	(17)
06274300	7.1	10.7	9.3	7.7	6.7	5.7	4.9
	[2.4]	[3.4]	[3.0]	[2.5]	[2.2]	[1.9]	[1.7]
	(7)	(3)	(4)	(6)	(8)	(11)	(15)
06275000	48.1	36.2	30.5	24.4	20.9	18.2	15.7
	[48.0]	[35.8]	[30.1]	[24.0]	[20.5]	[17.8]	[15.4]
	(7)	(14)	(20)	(31)	(42)	(55)	(73)
06278300	16.7	18.0	15.6	12.8	11.1	9.5	8.2
	[8.6]	[9.3]	[8.0]	[6.5]	[5.6]	[4.8]	[4.1]
	(7)	(6)	(8)	(12)	(16)	(22)	(30)
06279500	5.8	8.7	7.6	6.2	5.4	4.6	4.0
	[1.4]	[2.1]	[1.8]	[1.5]	[1.3]	[1.1]	[0.9]
	(7)	(3)	(4)	(6)	(8)	(11)	(15)

Table 5.—Selected results of the analysis—Continued

	Budget, in thousands of 1987 dollars						
	Current (1987)	Alternative values					
Identification	264	235	250	280	310	350	400
	Average values of SE, [EGS], and (visits) for individual stations--Continued						
06280000	5.7	10.4	8.6	6.2	5.3	4.6	4.1
	[1.8]	[3.5]	[2.7]	[1.9]	[1.6]	[1.5]	[1.3]
	(7)	(2)	(3)	(6)	(8)	(11)	(14)
06280300	5.9	10.4	8.7	6.4	5.6	4.8	4.3
	[2.7]	[4.2]	[3.6]	[2.8]	[2.5]	[2.2]	[2.0]
	(7)	(2)	(3)	(6)	(8)	(11)	(14)
06281000	12.4	14.6	12.4	9.6	8.3	7.1	6.2
	[10.9]	[13.0]	[10.9]	[8.2]	[7.1]	[6.1]	[5.3]
	(7)	(5)	(7)	(12)	(16)	(22)	(29)
06281400	20.0	23.1	20.0	15.6	13.6	11.6	10.1
	[16.7]	[19.4]	[16.7]	[12.9]	[11.1]	[9.4]	[8.2]
	(7)	(5)	(7)	(12)	(16)	(22)	(29)
06282000	5.4	9.6	8.0	5.8	5.1	4.3	3.9
	[2.6]	[4.5]	[3.8]	[2.8]	[2.4]	[2.1]	[1.9]
	(7)	(2)	(3)	(6)	(8)	(11)	(14)
06284500	7.7	13.4	11.3	9.0	7.7	6.9	5.8
	[4.0]	[6.6]	[5.7]	[4.6]	[4.0]	[3.5]	[3.0]
	(7)	(2)	(3)	(5)	(7)	(9)	(13)
06284800	18.2	21.3	18.2	14.6	12.6	10.9	9.2
	[6.3]	[7.3]	[6.3]	[5.1]	[4.4]	[3.8]	[3.2]
	(7)	(5)	(7)	(11)	(15)	(20)	(28)
06285100	8.6	12.8	11.3	9.3	8.1	7.0	6.0
	[6.4]	[9.9]	[8.5]	[6.9]	[5.9]	[5.1]	[4.3]
	(7)	(3)	(4)	(6)	(8)	(11)	(15)
06285400	8.8	15.3	12.9	10.2	8.8	7.8	6.6
	[3.8]	[5.7]	[5.1]	[4.3]	[3.8]	[3.4]	[2.9]
	(7)	(2)	(3)	(5)	(7)	(9)	(13)
06637750	8.1	12.0	9.5	8.1	6.8	5.8	5.3
	[2.9]	[3.9]	[3.3]	[2.9]	[2.5]	[2.2]	[2.0]
	(7)	(3)	(5)	(7)	(10)	(14)	(17)

Table 5.--Selected results of the analysis--Continued

Identification	Budget, in thousands of 1987 dollars						
	Current (1987)	Alternative values					
	264	235	250	280	310	350	400
	Average values of SE, [EGS], and (visits) for individual stations--Continued						
06638090	9.5	12.3	10.2	8.4	7.0	6.0	5.3
	[4.1]	[5.4]	[4.4]	[3.7]	[3.1]	[2.6]	[2.3]
	(7)	(4)	(6)	(9)	(13)	(18)	(23)
09188500	7.3	11.6	10.0	7.7	6.9	5.8	5.2
	[4.8]	[6.4]	[5.9]	[5.0]	[4.6]	[4.0]	[3.6]
	(7)	(2)	(3)	(6)	(8)	(12)	(16)
09196500	7.1	12.5	10.5	7.7	6.7	5.5	4.8
	[5.0]	[9.3]	[7.7]	[5.5]	[4.7]	[3.8]	[3.4]
	(7)	(2)	(3)	(6)	(8)	(12)	(16)
09203000	9.4	16.2	13.7	10.1	8.9	7.3	6.4
	[4.8]	[7.3]	[6.5]	[5.1]	[4.6]	[3.8]	[3.4]
	(7)	(2)	(3)	(6)	(8)	(12)	(16)
09205000	6.7	11.0	9.3	7.1	6.4	5.4	4.9
	[4.1]	[5.1]	[4.7]	[4.2]	[4.0]	[3.6]	[3.3]
	(7)	(2)	(3)	(6)	(8)	(12)	(16)
06209400	6.4	10.8	9.2	6.8	6.0	5.2	4.5
	[4.4]	[7.4]	[6.4]	[4.8]	[4.2]	[3.6]	[3.1]
	(7)	(2)	(3)	(6)	(8)	(11)	(15)
09210500	13.5	21.2	18.7	14.4	12.7	10.5	9.2
	[11.5]	[18.1]	[16.1]	[12.3]	[10.8]	[8.9]	[7.7]
	(7)	(2)	(3)	(6)	(8)	(12)	(16)
09211200	4.5	7.7	6.5	4.9	4.3	3.6	3.1
	[3.1]	[4.9]	[4.3]	[3.3]	[2.9]	[2.5]	[2.2]
	(7)	(2)	(3)	(6)	(8)	(12)	(16)
09212500	7.1	12.9	10.7	7.7	6.7	5.5	4.8
	[2.7]	[4.6]	[3.9]	[2.8]	[2.5]	[2.1]	[1.8]
	(7)	(2)	(3)	(6)	(8)	(12)	(16)
09213500	13.1	16.8	14.1	11.7	9.9	8.5	7.4
	[6.7]	[8.1]	[7.1]	[6.1]	[5.3]	[4.6]	[4.0]
	(7)	(4)	(6)	(9)	(13)	(18)	(24)

Table 5.—Selected results of the analysis—Continued

Identification	Budget, in thousands of 1987 dollars						
	Current (1987)	Alternative values					
	264	235	250	280	310	350	400
	Average values of SE, [EGS], and (visits) for individual stations--Continued						
09215550	11.8	15.5	12.8	10.5	8.7	7.5	6.5
	[6.1]	[8.1]	[6.6]	[5.3]	[4.4]	[3.8]	[3.3]
	(7)	(4)	(6)	(9)	(13)	(18)	(24)
09216050	11.5	15.0	12.5	10.2	8.5	7.3	6.3
	[6.8]	[9.1]	[7.4]	[6.0]	[5.0]	[4.3]	[3.7]
	(7)	(4)	(6)	(9)	(13)	(18)	(24)
09217000	4.1	7.5	6.2	4.4	3.8	3.1	2.7
	[2.0]	[4.0]	[3.1]	[2.2]	[1.9]	[1.6]	[1.3]
	(7)	(2)	(3)	(6)	(8)	(12)	(16)
09218500	13.0	16.9	14.0	11.5	9.6	8.2	7.1
	[5.2]	[6.9]	[5.6]	[4.6]	[3.8]	[3.2]	[2.8]
	(7)	(4)	(6)	(9)	(13)	(18)	(24)
09223000	12.7	16.0	13.5	11.3	9.5	8.1	7.1
	[9.4]	[11.9]	[10.0]	[8.4]	[7.0]	[6.0]	[5.2]
	(7)	(4)	(6)	(9)	(13)	(18)	(24)
09224700	15.7	20.5	16.9	13.9	11.6	9.6	8.4
	[4.7]	[6.2]	[5.1]	[4.2]	[3.5]	[2.9]	[2.6]
	(7)	(4)	(6)	(9)	(13)	(19)	(25)
09229500	15.5	20.1	16.7	13.8	11.5	9.8	8.5
	[7.1]	[9.1]	[7.6]	[6.3]	[5.3]	[4.5]	[3.9]
	(7)	(4)	(6)	(9)	(13)	(18)	(24)
10016900	8.3	15.2	12.5	9.0	7.8	6.4	5.2
	[1.7]	[3.1]	[2.5]	[1.8]	[1.5]	[1.3]	[1.1]
	(7)	(2)	(3)	(6)	(8)	(12)	(16)

---

[1] The average standard error per station is the square root of the average station variance.

The current (1987) operating policy results in an average standard error of estimate of nonwinter streamflow of about 13.2 percent. This policy is based on a budget of $264,000 for operating the 47-station streamflow-gaging network. Without lost record, the current standard error would be 9.7 percent. The current practice is within about 1.6 percent of the optimal value of standard error (11.6 percent) for the present budget when lost record is considered and is about 2.7 percent when lost record is not considered (fig. 3). The average standard error could be reduced as much as 1.6 percent by altering the route schedules to achieve more frequent visits to sites where uncertainty is large, and less frequent visits to sites where uncertainty is small. EGS values, in percent, are slightly less than half the corresponding standard errors.

A budget of about $235,000 could be used to operate the program. Stations would have to be eliminated from the program if the budget were less than this minimum. At that budget level, the optimal average standard error per station is about 15.5 percent, an increase of 34 percent compared to the optimal accuracy possible under the present budget (11.6 percent).

The maximum budget analyzed was $400,000, about 50 percent more than the present budget. This resulted in an optimal average standard error of estimate of about 6.6 percent. Thus, a 50-percent increase in the budget would decrease the optimal average standard error obtainable under the current budget by approximately 50 percent.

For all budgets considered, the effects of lost record account for about 40 percent of the standard error. Thus, improvements in instrumentation as well as increased use of local observers and data telemetry equipment could have a positive effect on uncertainties of instantaneous discharges.

Summary of Phase Three of Analysis

As a result of phase three of the analysis, conclusions are as follows:

1. The travel routes and measurement frequencies now (1987) in use are near the optimal level for the current budget. Changes in routes and station visits could optimally result in a 1.6 percent decrease in the standard error.

2. Any decrease in the current (1987) budget of $264,000 would result in an increase in the average standard error. A decrease to a budget of $235,000 would increase the optimal standard error from 11.6 percent to 15.5 percent. Any increase in budget could produce a reduction in the average standard error. A 50-percent increase in the current budget would decrease the optimal average standard error by approximately 50 percent.

3. Methods for decreasing the probabilities of lost record need to be explored. These methods may include improved instrumentation as well as increased use of local observers and data telemetry equipment.

SUMMARY

The first phase of the analysis of the Wyoming streamflow-gaging program categorized the uses of data at the 139 stations that were operated in 1984. Schuetz (1986) determined that current uses of the surface-water data justified continued operation of all stations.

Potential for use of alternative methods to simulate streamflow records was investigated by using regression analysis and hydrologic flow-routing techniques in the second phase of the analysis. Accuracy of the alternative methods was sufficient to consider discontinuing only one gage--at North Fork Crazy Woman Creek near Buffalo (06314000). However, the uses of data from both stations on North Fork Crazy Woman Creek in 1984 and in prior years required that both gages be operated.

The third phase of the analysis, to evaluate the operation of the gaging-station network by using associated uncertainty in streamflow records for various operating budgets, was limited to nonwinter operation of the network of 47 stations operated by the Riverton Field Office of the Geological Survey. The evaluation of that network is considered sufficient to address the effectiveness of the stream-gaging program in Wyoming and to provide a basis for considering changes in operating procedures.

The travel routes and measurement frequencies currently operated with a budget of $264,000 in the Riverton field area are near the optimal level. Changes in routes and station visits could optimally reduce the standard error by 1.6 percent, from 13.2 to 11.6 percent.

Any decrease in the current (1987) budget of $264,000 would result in an increase in the average standard error. A decrease to a budget of $235,000 would increase the optimal standard error from 11.6 to 15.5 percent. Conversely, a 50-percent increase in the current budget would decrease the optimum average standard error to 6.6 percent.

For all budgets considered, the effects of lost record account for about 40 percent of the average standard error. Thus, improvements in instrumentation as well as increased use of local observers and data telemetry equipment could have a positive effect on uncertainties of instantaneous discharges.

REFERENCES CITED

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

Doyle, W.H., Jr., Shearman, J.O., Stiltner, G.J., and Krug, W.R., 1983, Digital model for streamflow routing by convolution methods: U.S. Geological Survey Water-Resources Investigations Report 83-4160, 130 p.

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

Druse, S.A., and Rucker, S.J., IV, 1985, Water resources data for Wyoming, water year 1984: U.S. Geological Survey Water Data Report WY-84-1, 470 p.

Engel, G.B., Wahl, K.L., and Boohar, J.A., 1984, Cost-effectiveness of the stream-gaging program in Nebraska: U.S. Geological Survey Water-Resources Investigations Report 84-4098, 76 p.

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

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

Gilroy, E.J., and Moss, M.E., 1981, Cost-effective stream-gaging strategies for the Lower Colorado River Basin: U.S. Geological Survey Open-File Report 81-1019, 38 p.

Helwig, J.T., and Council, K.A., eds., 1979, SAS user's guide, 1979 edition: Raleigh, N.C., SAS Institute, Inc., 494 p.

Hirsch, R.M., 1982, A comparison of four streamflow record extension techniques: Water Resources Research, v. 18, no. 4, p. 1081-1088.

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

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

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

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

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

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

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

Schuetz, J.R., 1986, Use and availability of continuous streamflow records in Wyoming: U.S. Geological Survey Open-File Report 85-685, 33 p.

Thomas, D.M., and Benson, M.A., 1970, Generalization of streamflow characteristics from drainage-basin characteristics: U.S. Geological Survey Water-Supply Paper 1975, 55 p.

Wahl, K.L., 1970, A proposed streamflow data program for Wyoming: U.S. Geological Survey open-file report, 44 p.

SUPPLEMENTAL INFORMATION

The following description of the computations and mathematical relations, together with illustrations, is modified from Fontaine and others (1984, p. 22-24).

Description of Mathematical Program and the Uncertainty Functions

In a study of the cost-effectiveness of a network of stream gages operated in the lower Colorado River basin, a methodology called K-CERA was developed (Moss and Gilroy, 1980). The K-CERA methodology considers the cost effectiveness of a network of stream gages to be determined by the total variance--uncertainty--in either the annual mean discharge or the instantaneous discharge at all sites involved in the streamflow-gaging program and the cost of achieving that uncertainty. For the present study the measure of uncertainty at each site was taken to be the variance of the percent error in the instantaneous discharge. (See Fontaine and others, 1984, for the argument for this measure of uncertainty.)

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

The second step is to fit a lag-one-day autoregressive model to this temporal sequence of discharge residuals. The three parameters obtained from this analysis are (1) the measurement variance--actually estimated--a priori, (2) the process variance--a measure of the variability about the rating in the absence of measurement error, and (3) RHO, the lag-one autocorrelation coefficient--a measure of the memory in the sequence of discharge residuals. These three parameters determine the variance, V_f, of the percentage error in the estimation of instantaneous discharge whenever the primary correlative data at the site is available for use in the rating equation. Kalman filter theory, along with the assumption of a first-order Markovian process, is used to determine this variance V_f as a function of the number of discharge measurements per year (Moss and Gilroy, 1980).

If the primary correlative data at the site are not available, the discharge may be estimated by correlation with nearby sites. The correlation coefficient, r_c between the streamflows with seasonal trends removed (detrended) at the site of interest and detrended streamflows at the other sites is a measure of the soundness of their linear relationship. The fraction of the variance of the streamflow at the primary site that is explained by data from other sites is r_c^2. The variance of the percent error in streamflows at the primary site in the absence of primary data at both the principal site and nearby sites is taken to be

$$C_v = 100\ (1/365 \sum_{i=j}^{365} (s_i/u_i)^2)^{1/2} \qquad (1)$$

where s_i is the square root of the variance of daily discharges for the ith day of the year and u_i is the expected value of discharge on the ith day of the year. Thus the variance, V_r, of the percentage error during periods of reconstructed streamflow records is

$$V_r = (1-r_c^2)C_v^2 \qquad (2)$$

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

$$V_e = C_v^2\ . \qquad (3)$$

If the fraction of time when primary correlative data are available is denoted by e_f and the fraction of time when secondary streamflow data is available for reconstruction is e_r and $e_e=1-e_f-e_r$, the total percentage error variance, V_T, is given by

$$V_T = e_f\ V_f + e_rV_r + e_eV_e\ . \qquad (4)$$

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

The variance V_T given by equation 4, and which is a function of the number of visits to the site, is determined for each site in the stream-gaging network. For a given site visitation strategy, the sum of the variance, V_T, over all sites is taken as the measure of the uncertainty of the network. The variance V_T given by equation 4 is one measure of the spread of a probability density function, g_T. The function g_T is a mixture of three probability density functions: g_f, g_r, and g_e; each of which is assumed to be a normal, or Gaussian, probability density with mean zero and variance V_f, V_r, and V_e respectively. Such a mixture is denoted by

$$g_T = e_f g_f + e_r g_r + e_e g_e \quad . \tag{5}$$

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

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

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

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

All these costs, routes, constraints, and uncertainty functions, are then used in an iterative search program to determine the number of times that each route is used during a year such that (1) the budget for the network is not exceeded, (2) at least the minimum number of visits to each station are made, and (3) the total uncertainty in the network is minimized. This allocation of the predefined budget among the stream gages is taken to be the optimal solution to the problem of cost-effective resource allocation. Due to the high dimensionality and non-linearity of the problem, the optimal solution may really be "near optimal." (See Moss and Gilroy, 1980, or Fontaine and others, 1984, for greater detail.)