COST EFFECTIVENESS OF THE STREAM-GAGING PROGRAM IN NORTH DAKOTA

By Gerald L. Ryan

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 89-4068

Bismarck, North Dakota

1989

DEPARTMENT OF THE INTERIOR

MANUAL LUJAN, JR., Secretary

U.S. GEOLOGICAL SURVEY

Dallas L. Peck, Director

For additional information
write to:

District Chief
U.S. Geological Survey
Water Resources Division
821 East Interstate Avenue
Bismarck, ND 58501

Copies of this report can
be purchased from:

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

CONTENTS

	Page
Abstract	1
Introduction	1
Phases of analysis	2
North Dakota stream-gaging program	2
Alternative methods of developing streamflow information	4
Discussion of methods	4
Description of statistical methods	4
Description of flow-routing methods	6
Potential for use of alternative methods	7
Results of statistical methods	10
Results of flow-routing methods	14
Summary of phase two of analysis	14
Cost-effective resource allocation	15
Discussion of model	15
Application of model in North Dakota	16
Definition of variance when gaging station is operating	16
Definition of variance when record is missing	18
Discussion of routes and costs	22
Results	24
Summary of phase three of analysis	29
Summary	29
References cited	30
Description of mathematical program	32

ILLUSTRATIONS

Figure 1. Map showing locations of stream-gaging stations in North Dakota and selected stream-gaging stations in Minnesota and South Dakota — 3

2. Graph showing typical uncertainty functions for instantaneous discharge — 21

3. Graph showing relation between average equivalent Gaussian spread and budget — 25

TABLES

Table 1. Gaging-station combinations screened in alternative-methods analysis and associated lag-time and correlation coefficient for April 1 through October 31 — 8

2. Summary of statistics used for estimating relations defined in alternative-methods analysis for North Dakota — 11

3. Summary of calibration and verification results for flow-routing model as applied to the reach of the Red River of the North between the Wahpeton (05051500) and Hickson (05051522) gaging stations — 15

4. Gaging-station list and summary of statistics used to define uncertainty functions — 19

5. Selected results of analysis — 26

SELECTED FACTORS FOR CONVERTING INCH-POUND UNITS TO METRIC UNITS

For those readers who may prefer to use metric (International System) units rather than inch-pound units, the conversion factors for the terms used in this report are given below.

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

COST EFFECTIVENESS OF THE STREAM-GAGING PROGRAM IN NORTH DAKOTA

By Gerald L. Ryan

ABSTRACT

This report documents results of a cost-effectiveness study of the stream-gaging program in North Dakota. It is part of a nationwide evaluation of the stream-gaging program of the U.S. Geological Survey.

One phase of evaluating cost effectiveness is to identify less costly alternative methods of simulating streamflow records. Statistical or hydrologic flow-routing methods were used as alternative methods to simulate streamflow records for 21 combinations of gaging stations from the 94-gaging-station network. Accuracy of the alternative methods was sufficient to consider discontinuing only one gaging station.

Operation of the gaging-station network was evaluated by using associated uncertainty in streamflow records. The evaluation was limited to the non-winter operation of 29 gaging stations in eastern North Dakota. The current (1987) travel routes and measurement frequencies require a budget of about $248,000 and result in an average equivalent Gaussian spread in streamflow records of 16.5 percent. Changes in routes and measurement frequencies optimally could reduce the average equivalent Gaussian spread to 14.7 percent.

Budgets evaluated ranged from $235,000 to $400,000. A $235,000 budget would increase the optimal average equivalent Gaussian spread from 14.7 to 20.4 percent, and a $400,000 budget could decrease it to 5.8 percent.

INTRODUCTION

The U.S. Geological Survey is the principal Federal agency collecting surface-water data in the Nation. Data are collected by the Water Resources Division in cooperation with State and local governments and other Federal agencies (Ryan, 1985). Currently (1987), the U.S. Geological Survey is operating about 7,000 continuous-record surface-water gaging stations throughout the Nation. Records for some of these gaging stations date back to the turn of the century.

Any activity of long standing, such as collection of surface-water data, needs to be reexamined at intervals, if not continually, because of changes in objectives, technology, or external constraints. The last systematic nation-wide evaluation of the stream-gaging program was completed in 1970 and is documented by Benson and Carter (1973). In 1983, the U.S. Geological Survey undertook another nationwide evaluation of the stream-gaging program. The evaluation is to be completed over a 5-year period, with 20 percent of the program being evaluated each year. The objective is to define and document the most cost-effective means of furnishing streamflow information. Sections of this report that describe techniques or methodology are from earlier reports (Fontaine and others, 1984, and Engel and others, 1984).

Phases of Analysis

Nationwide analysis of the stream-gaging program comprises three major phases. Data use and availability are analyzed in phase one, less costly alternative methods of furnishing streamflow information are investigated in phase two, and operation of the gaging-station network is analyzed in phase three. The purpose of this report is to document phases two and three of the nationwide analysis as applied to the North Dakota District of the U.S. Geological Survey.

Phase one, to analyze data use and availability, was completed by Ryan (1985). His report documents a survey that identified local, State, and Federal data uses for 94 continuous-record surface-water gaging stations operated by the North Dakota District in 1984. Additionally, the report identifies sources of funding related to collection of streamflow data and presents the frequency of data availability. Data uses were categorized into seven classes: Regional hydrology, hydrologic systems, legal obligations, planning and design, project operation, hydrologic forecasts, and water-quality monitoring. Ryan's (1985) report documents that use of surface-water data collected from the gaging stations justified continued operation of all gaging stations.

Phase two of the analysis, to identify less costly alternative methods of furnishing streamflow information, was applied to those gaging stations in the statewide network that appeared to have sufficient correlation to warrant either statistical or flow-routing methods. Phase three of the analysis, to evaluate the operation of gaging-station networks by using associated uncertainty in streamflow records for various operating budgets, was limited to the network of gaging stations operated by the Grand Forks Field Headquarters of the North Dakota District.

North Dakota Stream-Gaging Program

The North Dakota stream-gaging program has evolved through the years to meet local, State, and Federal needs for surface-water data. The stream-gaging program has remained stable since Ryan (1985) reported on the 94-station network that was in place in 1984. The network evaluated in this report is included in the gaging-station network described by Ryan (1985; fig. 1).

The U.S. Geological Survey operates its gaging-station network from field headquarters located in Dickinson and Grand Forks and from the District Office in Bismarck. The network operated by the Grand Forks Field Headquarters consists of 29 gaging stations in the Red River of the North basin in North Dakota and includes about one-third of the continuous-record surface-water gaging stations operated by the North Dakota District (fig. 1). Operations in the Grand Forks area are considered representative of the overall stream-gaging program in North Dakota and provide a basis for considering changes in operating procedures.

The alternative-methods section of this report will evaluate selected gaging stations from the 94-station network. The cost-effective resource-allocation phase of this report will analyze the nonwinter operation of the

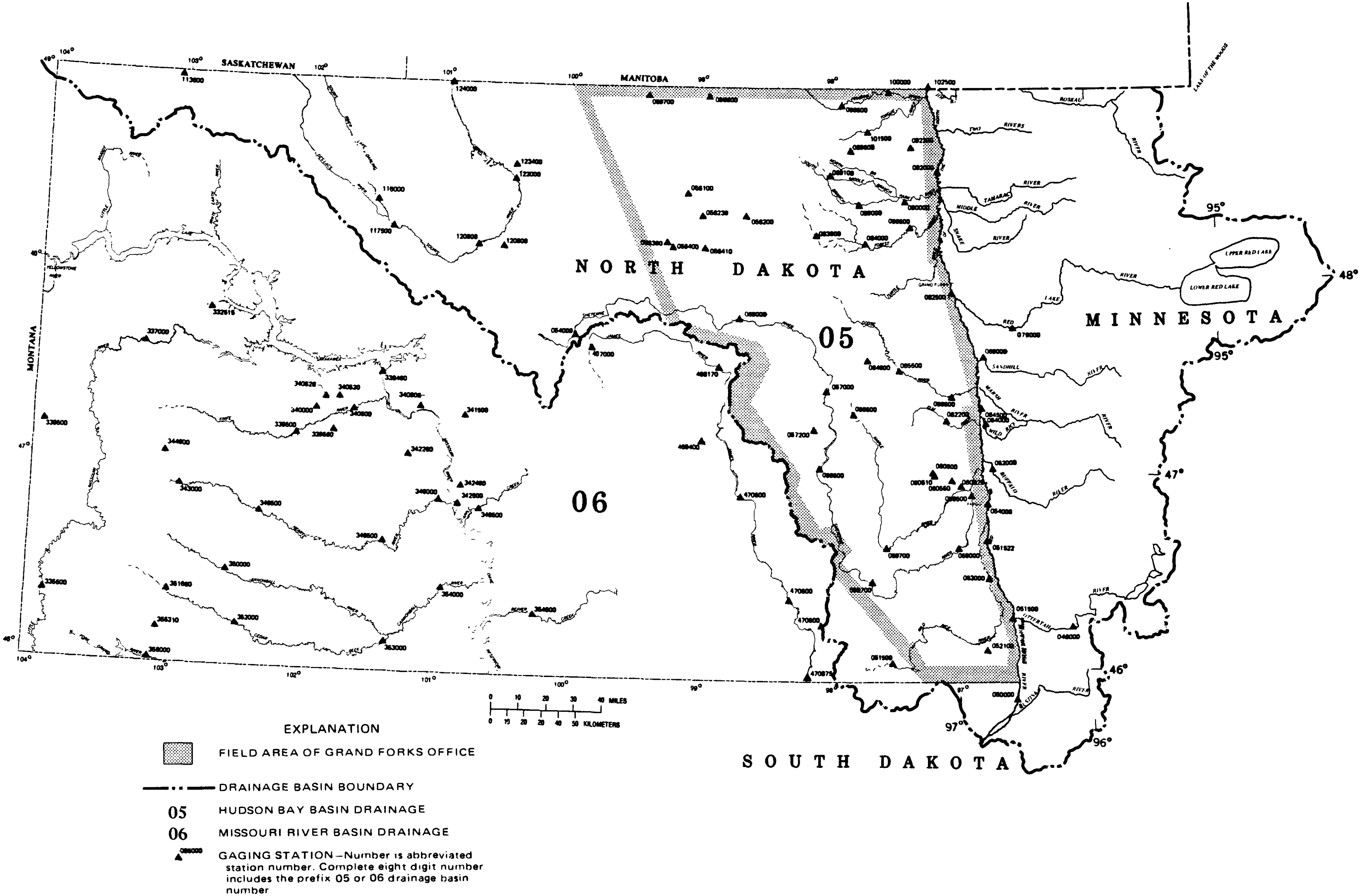

Figure 1.—Locations of stream-gaging stations in North Dakota and selected stream-gaging stations in Minnesota and South Dakota.

gaging-station network currently (1987) operated by the Grand Forks Field Headquarters.

ALTERNATIVE METHODS OF DEVELOPING STREAMFLOW INFORMATION

Phase two of the three-phase analysis of the stream-gaging program was to identify alternative methods of furnishing daily streamflow information in lieu of operating continuous-record surface-water gaging stations (Engel and others, 1984). The objective of this phase was to identify gaging stations where alternative methods, such as statistical analysis or hydrologic flow routing, could 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 estimated daily flows would be adequate for the intended purpose.

Use of data from a gaging station affects whether or not information potentially can be provided by alternative methods. For example, gaging stations for which flood hydrographs are required in a real-time sense, such as hydrologic forecasts and project operation, are not candidates for alternative methods. Also, there may be a legal obligation to operate an actual gaging station that would preclude using alternative methods. Primary candidates for alternative methods are gaging stations that are operated upstream or downstream from other gaging stations on the same stream. The accuracy of estimated streamflow may be adequate if flows are well correlated between gaging 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 by Fontaine 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 (Water Data Storage and Retrieval System) 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 simulated streamflow records. Because of the limited time available for the analysis, only two methods were considered--statistical analysis and hydrologic flow routing.

Gaging stations in the North Dakota stream-gaging program were screened to determine their potential for use of alternative methods. Selected methods then were applied to the nonwinter period for those gaging stations where potential was great.

Description of Statistical Methods

Simple- and multiple-regression methods can be used to estimate daily flow records. Unlike hydrologic flow routing, regression methods are not limited to locations where an upstream gaging station exists on the same stream. Regression equations can be computed that relate daily flows (or their logarithms) at one gaging station (dependent variable) to daily flows at another gaging station or at a combination of upstream, downstream, or tributary gaging stations. The independent variables in the regression equations can include gaging 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 Riggs (1973) and Thomas and Benson (1970). Only a brief description of a regression model is provided in this report.

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

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

where

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

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

B_0 and B_j = regression constant and coefficients, and

e_i = the random error term.

Equation 1 is calibrated (B_0 and B_j are estimated) using observed values of Y_i and X_j. The procedure for determining values of B_0 and B_j is called the method of ordinary least squares (OLS). The observed daily mean discharges can be retrieved from the WATSTORE Daily Values File (Hutchison, 1975). Values of discharge for the independent variables may be observed on the same day as discharges at the independent gaging station or may be for previous or future days, depending on whether gaging station j is upstream or downstream of gaging station i. During calibration, the regression constant and coefficients (B_0 and B_j) are tested to determine if they are significantly different from zero. A given independent variable is retained in the regression equation only if 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. Both the calibration and verification periods need 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) against 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 no time trends are reflected in either the data or the equation. The presence of either nonlinearity or bias requires that the data be transformed (for example, by converting 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 gaging station. 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 is important, OLS regression models are not appropriate.

Hirsch (1982) discussed this problem and described another method that preserves the variance of the original data. Maintenance of variance extension type 1 (MOVE.1) uses the same equation (eq. 1) as OLS, but the parameters B_0 and B_j are determined so that the mean and variance estimated using equation 1 would equal the sample mean and variance. Monte Carlo and empirical experiments with actual streamflow records by Hirsch (1982) showed that, even for a relatively small sample size, the MOVE.1 equation tends to produce a less biased estimate of the variance of an extended streamflow record than does OLS regression. The reason for record extension is to produce a time series that is relatively long and possesses the same statistical characteristics as those of the actual record. Hirsch (1982) demonstrated that the MOVE.1 equation procedure preserves the statistical characteristic of the actual record better than the OLS procedure.

Description of Flow-Routing Methods

Hydrologic flow-routing methods use the law of conservation of mass and the relationship between storage in a reach and outflow from the reach. 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 flow-routing methods. The unit-response method uses one of two routing techniques--storage continuity (Sauer, 1973) or 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 sites to a downstream site is available (Doyle and others, 1983). The model, referred to as CONROUT (A Digital Model For Streamflow Routing By Convolution Methods), 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 the problem.

Determination of the system's response to an upstream pulse is not the total solution for most flow-routing problems. The convolution process does not account for flow from the intervening area between the upstream and downstream sites. 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 gaging station).

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 simulated discharge to observed discharge.

Potential for Use of Alternative Methods

A two-level screening process was applied to gaging stations in North Dakota 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 gaging station from streamflow at other gaging stations. The first-level screening was subjective; no attempt was made to apply mathematical procedures. Gaging stations that passed first-level screening then were screened again to determine whether simulated streamflow data would be acceptable according to the data uses described by Ryan (1985). Even if simulated streamflow data were not acceptable for 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 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.

Combinations of gaging stations that passed the first-level screening for concurrent, nonwinter (April 1 through October 31) daily discharge records are listed in table 1. After the first-level screening, the month of April was left out of all further analysis because it was found that during some years flows during parts or all of April were affected by ice. The gaging station whose record is being simulated from one or more index gaging stations, the index gaging stations, and the lag-time of daily discharge between the two gaging stations are included in table 1. Locations of these gaging stations are shown in figure 1. Correlation coefficients were determined for the combinations of gaging stations listed in table 1, and gaging stations that had

Table 1.--*Gaging-station combinations screened in alternative-methods analysis and associated lag-time and correlation coefficient for April 1 through October 31*

Gaging-station numbers and names (Index stations are indented)	Lag-time of daily discharge (in days)	Correlation coefficient
05051500 Red River of the North at Wahpeton, N.Dak.		0.764
05046000 Otter Tail River below Orwell Dam near Fergus Falls, Minn.	-1	
05050000 Bois de Sioux River near White Rock, S.Dak.	-2	
05051522 Red River of the North at Hickson, N.Dak.		.922
05051500 Red River of the North at Wahpeton, N.Dak.	-2	
05054000 Red River of the North at Fargo, N.Dak.		.915
05051522 Red River of the North at Hickson, N.Dak.	-1	
05053000 Wild Rice River near Abercrombie, N.Dak.	-3	
05056200 Edmore Coulee near Edmore, N.Dak.		.558
05056100 Mauvais Coulee near Cando, N.Dak.	0	
05059500 Sheyenne River at West Fargo, N.Dak.		.968
05059000 Sheyenne River near Kindred, N.Dak.	-1	
05064500 Red River of the North at Halstad, Minn.		.957
05054000 Red River of the North at Fargo, N.Dak.	-3	
05062000 Buffalo River near Dilworth, Minn.	-3	
05064000 Wild Rice River at Hendrum, Minn.	0	
05082500 Red River of the North at Grand Forks, N.Dak.		.965
05064500 Red River of the North at Halstad, Minn.	-3	
05069000 Sand Hill River at Climax, Minn.	-2	
05079000 Red Lake River at Crookston, Minn.	-2	
05092000 Red River of the North at Drayton, N.Dak.		.978
05082500 Red River of the North at Grand Forks, N.Dak.	-2	
05085000 Forest River at Minto, N.Dak.	-2	
05090000 Park River at Grafton, N.Dak.	-1	
05102500 Red River of the North at Emerson, Manitoba		.977
05092000 Red River of the North at Drayton, N.Dak.	-3	
05100000 Pembina River at Neche, N.Dak.	-1	
05101000 Tongue River at Akra, N.Dak.	-2	

Table 1.--*Gaging-station combinations screened in alternative-methods analysis and associated lag-time and correlation coefficient for April 1 through October 31*--Continued

Gaging-station numbers and names (Index stations are indented)	Lag-time of daily discharge (in days)	Correlation coefficient
05117500 Souris River above Minot, N.Dak.		0.971
05116000 Souris River near Foxholm, N.Dak.	-1	
05116500 Des Lacs River at Foxholm, N.Dak.	-1	
05120000 Souris River near Verendrye, N.Dak.		.945
05117500 Souris River above Minot, N.Dak.	-2	
05122000 Souris River near Bantry, N.Dak.		.934
05120000 Souris River near Verendrye, N.Dak.	-2	
05120500 Wintering River near Karlsruhe, N.Dak.	-2	
06337000 Little Missouri River near Watford City, N.Dak.		.879
06335500 Little Missouri River at Marmarth, N.Dak.	-3	
06336600 Beaver Creek near Trotters, N.Dak.	-2	
06340500 Knife River at Hazen, N.Dak.		.870
06339500 Knife River near Golden Valley, N.Dak.	-2	
06340000 Spring Creek at Zap, N.Dak.	-1	
06342500 Missouri River at Bismarck, N.Dak.		.909
06340500 Knife River at Hazen, N.Dak.	-1	
06338490 Missouri River at Garrison Dam, N.Dak.	-1	
06349000 Heart River near Mandan, N.Dak.		.929
06348000 Heart River near Lark, N.Dak.	-1	
06349500 Apple Creek near Menoken, N.Dak.		.491
06342450 Burnt Creek near Bismarck, N.Dak.	0	
06353000 Cedar Creek near Raleigh, N.Dak.		.776
06352000 Cedar Creek near Haynes, N. Dak.	-2	
06354000 Cannonball River at Breien, N.Dak.		.904
06353000 Cedar Creek near Raleigh, N. Dak.	-1	
06350000 Cannonball River at Regent, N.Dak.	-2	
06354500 Beaver Creek at Linton, N.Dak.		.581
06349500 Apple Creek near Menoken, N.Dak.	0	
06470500 James River at LaMoure, N.Dak.		.807
06470000 James River at Jamestown, N.Dak.	-4	

little correlation with corresponding index gaging stations were eliminated. Combinations of gaging stations that were well correlated were used in further statistical analyses, and the results are described in the next section of this report. Hydrologic flow routing of daily flows generally gives results that are comparable to results obtained using statistical methods. Therefore, routing methods were used on only one pair of gaging stations for demonstration purposes.

Results of Statistical Methods

Correlation methods used on the combinations of gaging stations listed in table 1 indicated that statistical methods would be unacceptable for some of the combinations. Those gaging stations that had correlation coefficients of less than 0.900 were considered unacceptable and were eliminated from further consideration. About 81 percent of the variance can be explained when the correlation coefficient is 0.900.

Combinations of gaging stations produced unacceptable results for several reasons. However, the most common reasons relate to the wide variability of runoff and channel storage that occurs in North Dakota and to variable effects of diversions, numerous small reservoirs, and irrigation return flows.

The results of statistical analyses for selected combinations of gaging stations are listed in table 2. Results are listed for both the OLS model and the MOVE.1 model suggested by Hirsch (1982). All variables shown in the table are statistically significant at the 0.01 level. The standard error of estimate for models, in units of cubic feet per second, is not directly useful because the data are not homoscedastic (the variance is not constant throughout the range of flow). Therefore, individual errors were converted to percentage deviations, and the standard deviation of those percentage values was defined. Percentage deviations reported are plus or minus 5, 15, 25, and 40 percent of observed discharge.

Two periods were used in the statistical analyses. Water years 1982-84 were used for calibration, and water years 1985-86 were used for verification.

The Missouri River at Bismarck (06342500) is the only gaging station that appears to be a true candidate for application of alternative methods. The gaging station at Bismarck cannot be discontinued, however, because a cooperator uses it to monitor flows out of Garrison Reservoir. The equations in table 2 for the gaging station at Bismarck could be used, if necessary, to compute missing record. None of the other relations are considered to be sufficiently accurate to consider discontinuing the gaging station. The equations listed in table 2 could be improved somewhat by changing the form of the model and by defining separate relations for various ranges of discharge. However, the conclusions would not change--gaging stations are still necessary.

Table 2.--*Summary of statistics used for estimating relations defined in alternative-methods analysis for North Dakota*

[Relations are based on the ordinary-least-squares (OLS) model and the maintenance-of-variance-extension type 1 (MOVE.1) model. The relations are defined only for May 1 through October 31]

	Percent of days in indicated percent range							
	Calibration (1982-84)				Verification (1985-86)			
Estimating relation	±5	15	25	40	±5	15	25	40
Red River of the North at Hickson, N.Dak.								
OLS								
05051522 = 39 + 0.971(050515L2)	45	83	89	92	43	73	82	87
MOVE.1								
05051522 = 34 + 0.979(050515L2)	47	84	89	91	48	75	83	91
Red River of the North at Fargo, N.Dak.								
OLS								
05054000 = 8 + 1.005(OHICNABR)	26	66	86	94	31	65	78	86
MOVE.1								
05054000 = -23 + 1.048(OHICNABR)	30	76	91	96	41	75	88	93
Sheyenne River at West Fargo, N.Dak.								
OLS								
05059500 = -14 + 1.194(050590L1)	22	60	81	92	26	66	90	97
MOVE.1								
05059500 = -20 + 1.234(050590L1)	20	56	75	86	26	58	81	95
Red River of the North at Halstad, Minn.								
OLS								
05064500 = -22 + 1.319(OFWFDILH)	30	70	87	94	16	52	74	85
MOVE.1								
05064500 = 64 + 1.317(OFWFDILH)	15	53	72	89	9	31	67	87
Red River of the North at Grand Forks, N.Dak.								
OLS								
05082500 = -136 + 1.120(OHALCRCL)	41	85	95	98	35	80	94	99
MOVE.1								
05082500 = -319 + 1.181(OHALCRCL)	35	71	88	95	22	60	78	89

Table 2.--*Summary of statistics used for estimating relations defined in alternative-methods analysis for North Dakota*--Continued

Estimating relation	Percent of days in indicated percent range							
	Calibration (1982-84)				Verification (1985-86)			
	±5	15	25	40	±5	15	25	40
Red River of the North at Drayton, N.Dak.								
OLS 05092000 = 155 + 1.023(OGFKMIGR)	28	75	93	98	25	67	83	95
MOVE.1 05092000 = 74 + 1.046(OGFKMIGR)	41	84	96	99	41	81	88	95
Red River of the North at Emerson, Manitoba								
OLS 05102500 = 228 + 0.885(ODRNECAK)	29	68	85	93	32	58	74	89
MOVE.1 05102500 = 158 + 0.904(ODRNECAK)	34	73	88	94	32	65	83	95
Souris River above Minot, N.Dak.								
OLS 05117500 = 2.36 + 0.967(OSORNDES)	20	46	64	76	22	45	59	68
MOVE.1 05117500 = 1.99 + 0.970(OSORNDES)	18	45	63	75	22	44	59	68
Souris River near Verendrye, N.Dak.								
OLS 05120000 = 47.49 + 1.036(051175L2)	8	18	28	34	2	18	26	32
MOVE.1 05120000 = 34.35 + 1.157(051175L2)	6	20	28	36	10	18	22	26
Souris River near Bantry, N.Dak.								
OLS 05122000 = 46.20 + 1.015(OVERNKAR)	6	20	28	42	8	16	24	36
MOVE.1 05122000 = 32.46 + 1.090(OVERNKAR)	6	19	32	51	10	21	34	48
Missouri River at Bismarck, N.Dak.								
OLS 06342500 = 2490 + 0.981(OHAZNDAM)	80	98	100	100	78	100	100	100
MOVE.1 06342500 = 1844 + 1.012(OHAZNDAM)	80	98	100	100	84	100	100	100

Table 2.--*Summary of statistics used for estimating relations defined in alternative-methods analysis for North Dakota*--Continued

Estimating relation	Calibration (1982-84) ±5	15	25	40	Verification (1985-86) ±5	15	25	40
Heart River near Mandan, N.Dak.								
OLS 06349000 = -3.23 + 1.296(063480L1)	8	26	46	67	7	23	41	61
MOVE.1 06349000 = -4.15 + 1.285(063480L1)	8	27	47	68	8	24	39	61
Cannonball River at Breien, N.Dak.								
OLS 06354000 = 78.55 + 1.262(ORALNREG)	4	12	22	36	2	10	18	24
MOVE.1 06354000 = 34.38 + 1.545(ORALNREG)	7	20	34	47	4	16	28	34

Percent of days in indicated percent range.

NOTE: Standard deviation is not shown because differences are not normally distributed. Numbers are percent of daily discharges that are within plus or minus 5, 15, 25, and 40 percent of observed discharge. Index stations used in the analysis were lagged the number of days shown in table 1--for example, 050515L2 is 05051500 lagged 2 days.

(OHICNABR) = 05051522 + 05053000

(OFWFDILH) = 05054000 + 05062000 + 05064000

(OHALCRCL) = 05064500 + 05069000 + 05079000

(OGFKMIGR) = 05082500 + 05085000 + 05090000

(ODRNECAK) = 05092000 + 05100000 + 05101000

(OSORNDES) = 05116000 + 05116500

(OVERNKAR) = 05120000 + 05120500

(OHAZNDAM) = 06340500 + 06338490

(ORALNREG) = 06353000 + 06350000

Results of Flow-Routing Methods

The CONROUT model (Doyle and others, 1983) requires three parameters. They are:

X = routing distance (miles),
C_0 = flood-wave celerity (controls travel time), and
K_0 = dispersion or damping coefficient (controls spreading of the wave).

C_0 and K_0 are approximated from the following equations:

$$C_0 = (1/W_0)(dQ_0/dy) \tag{2}$$

$$K_0 = Q_0/(2S_0W_0) \tag{3}$$

where

W_0 = average channel width (feet) in the reach,

dQ_0/dy = slope of stage-discharge curve,

Q_0 = stream discharge of interest (cubic feet per second), and

S_0 = average bed slope (feet/feet) in the reach.

These parameters were estimated for the reach of the Red River of the North between the Wahpeton (05051500) and Hickson (05051522) gaging stations and were refined based on application of the model to the calibration period, 1982-84. The calibrated model then was used to simulate mean daily discharges for the verification period, 1985-86. The final parameter values were:

X = 74.0 mi,
C_0 = 3.00 ft/s, and
K_0 = 13,000 ft^2/s.

The net contributing drainage areas are 4,010 mi^2 for Wahpeton and 4,300 mi^2 for Hickson. Because of the difference in contributing area between these gaging stations, the model was used to route the flow at Wahpeton plus the flow from the intervening drainage area to Hickson. The intervening flow was estimated using a ratio of the difference in drainage areas between Wahpeton and Hickson. Results of the calibration and verification (shown in table 3) are comparable to the results obtained by statistical analysis (table 2).

Summary of Phase Two of Analysis

None of the gaging stations presently in operation in North Dakota can be replaced by an alternative method of producing streamflow information. The relation defined for the Missouri River at Bismarck is accurate and is useful for estimating missing record. Data needs, however, cannot be met with a synthesized record; therefore, the gaging station will continue in operation.

Table 3.--*Summary of calibration and verification results for flow-routing model as applied to the reach of the Red River of the North between the Wahpeton (05051500) and Hickson (05051522) gaging stations*

Daily discharge errors	Percent of time daily discharges were within given discharge error: Calibration 1982-84	Percent of time daily discharges were within given discharge error: Verification 1985-86
Less than or equal to 5 percent	41	41
Less than or equal to 10 percent	73	72
Less than or equal to 15 percent	86	87
Less than or equal to 20 percent	90	92
Less than or equal to 25 percent	93	96
Greater than 25 percent	7	4
Total volume error (percent)	-0.25	0.39

COST-EFFECTIVE RESOURCE ALLOCATION

Discussion of 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) to study the cost effectiveness of gaging-station networks. The original application of the techniques was to analyze a network of gaging stations 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 stream-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 U.S. Geological Survey, concentration of effort on large rivers and streams is undesirable and inappropriate.

The original version of K-CERA, therefore, was 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 squared; annual mean discharge, in percent squared;

and average instantaneous discharge, in cubic feet per second squared, or average instantaneous discharge, in percent squared (Fontaine and others, 1984). The use of percentage errors effectively gives equal weight to both large and small streams, and 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 missing stage or other correlative records that are used to compute streamflow data. The probabilities of missing correlative records increase as the period between service visits to a gaging station increases. A procedure for dealing with the missing record was developed by Fontaine and others (1984) and was incorporated into this study.

Brief descriptions of the mathematical program used to minimize the total error variance of the data-collection activity for given budgets and of the application of Kalman filtering (Gelb, 1974) to the determination of the accuracy of a stream-gaging record are presented by Fontaine and others (1984). A modified version of the mathematical program description is provided in the Description of Mathematical Program section at the end of this report. 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 Model in North Dakota

Phases one and two of this analysis indicate that operation of the current gaging-station network in North Dakota needs to be continued. Phase three of the analysis was limited to the network of gaging stations operated by the Grand Forks Field Headquarters. Operations in the Grand Forks area are considered representative of the overall stream-gaging program in North Dakota and provide a basis for considering changes in operating procedures.

The model assumes the uncertainty of discharge records at a given gaging station to be derived from three sources: (1) Errors that result because the stage-discharge relationship is not perfect (applies when the gaging station is operating), (2) errors in reconstructing records based on records from another gaging station when the primary gaging station is not operating, and (3) errors inherent in estimated discharge when the gaging station 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 squared in instantaneous discharge. 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 Gaging Station is Operating

The model used in this analysis assumes that the difference (residual) between instantaneous discharge (measured 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 three or more water years, and deviations (residuals) of the measured discharges from the rating discharge were determined. A time-series analysis of these residuals defined the 1-day lag (lag-one) autocorrelation coefficient and the process variance required by the K-CERA model. Finally, the error variance was defined within the model as a function of the lag-one autocorrelation coefficient, the process and measurement variances, and the frequency of discharge measurements.

In the North Dakota program analysis, definition of long-term rating functions is complicated by the fact that most gaging stations in the Grand Forks field area are affected by backwater from ice for about 6 months during the year. Rating curves based on open-water measurements are not applicable during the ice-affected periods.

In the pilot study for Maine, winter rating curves were replaced with regression relations relating the discharge at the ice-affected gaging station to the discharge at an ice-free station. 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 North Dakota because there are no ice-free stations in large areas of the State and because variability of winter flow, resulting from the temporary storage and subsequent release of ice, precludes the development of a winter rating. Reliable discharge records during the winter presently can 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 6 months in the Grand Forks field area, generally from November through April. The model was applied only to the 6 months (184 days) that are virtually free from ice effect. The study also assumed that, regardless of ice-free period visit requirements, four visits would continue to be made during the winter season.

Long-term rating curves applicable to ice-free periods were defined for each gaging station used in the evaluation. In some cases, existing ratings adequately defined the long-term condition and were used in the analysis. For a majority of gaging stations, however, 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 *B*1, *B*2, and *B*3 have the following physical interpretation: *B*1 is the logarithm of discharge for a flow depth of 1 ft, *B*2 is the gage height of zero flow, and *B*3 is the slope of the rating curve.

The residuals about the long-term rating for individual gaging stations defined the total variance. A review of discharge measurements made in North Dakota indicated that the average standard error of open-water measurements was about 3 percent. The measurement variance for all gaging stations, therefore, was defined as equal to the square of the 3-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 residuals 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.

The values of lag-one autocorrelation coefficient, process variance, and coefficient of variation are listed in table 4; length of period (184 days) 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 relationship 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 EGS (equivalent Gaussian spread) shown in figure 2 was introduced by Fontaine and others (1984, p. 26); the definition is included in the Description of Mathematical Program section at the end 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 uncertainty curves reflect a low process variance and high coefficient of variation for gaging station 05058000, a high process variance and high coefficient of variation for gaging station 05099600, and a low process variance and low coefficient of variation for gaging station 05051500. Lag-one autocorrelation coefficients are 0.93 or greater for all three gaging stations.

The residuals about rating curves for four gaging stations, 05057200, 05060500, 05089000, and 05098800, in the Grand Forks field area poorly approximate a continuous first-order Markov process. These gaging stations have moderate-to-significant changes in ratings resulting from channel changes, which usually are the result of beaver activity. These ratings 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. For these four gaging stations, process variance was assigned a value of 0.10 for the analysis. Additionally, gaging stations 05053000, 05056410, 05057000, and 05090000 were excluded from the analysis because the number of discharge measurements was insufficient.

Definition of Variance when Record is Missing

When stage record is missing at a gaging station, the model assumes the discharge record either is reconstructed using correlation with another gaging station or is estimated from historical discharge for that period. Fontaine and others (1984, p. 24) indicated that the fraction of time for which a

Table 4.--*Gaging-station list and summary of statistics used to define uncertainty functions*

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

Gaging-station number and name	Lag-one auto-correlation coefficient	Process variance	Coefficient of variation
05051500 Red River of the North at Wahpeton, N.Dak.	0.988	0.005527	0.793
05051522 Red River of the North at Hickson, N.Dak.	.978	.004335	.859
05054000 Red River of the North at Fargo, N.Dak.	.994	.003268	1.020
05056000 Sheyenne River near Warwick, N.Dak.	.979	.087160	1.940
05056400 Big Coulee near Churchs Ferry, N.Dak.	.989	.058000	1.430
05057200 Baldhill Creek near Dazey, N.Dak.	.992	.100000	1.700
05058000 Sheyenne River below Baldhill Dam, N.Dak.	.933	.003919	1.440
05058700 Sheyenne River at Lisbon, N.Dak.	.966	.008239	1.360
05059000 Sheyenne River near Kindred, N.Dak.	.965	.012078	1.150
05059500 Sheyenne River at West Fargo, N.Dak.	.984	.011648	.925
05059700 Maple River near Enderlin, N.Dak.	.991	.018010	2.060
05060500 Rush River at Amenia, N.Dak.	.992	.100000	1.890
05064500 Red River of the North at Halstad, Minn.	.827	.000369	.952
05064900 Beaver Creek near Finley, N.Dak.	.981	.050390	1.630

Table 4.--*Gaging-station list 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
05066500 Goose River at Hillsboro, N.Dak.	0.955	0.059625	1.740
05082500 Red River of the North at Grand Forks, N.Dak.	.914	.000557	.921
05083600 Middle Branch Forest River near Whitman, N.Dak.	.973	.042320	1.860
05084000 Forest River near Fordville N.Dak.	.992	.092720	1.910
05085000 Forest River at Minto, N.Dak.	.989	.029800	2.060
05089000 South Branch Park River below Homme Dam, N.Dak.	.998	.100000	2.040
05092000 Red River of the North at Drayton, N.Dak.	.922	.002193	.911
05098700 Hidden Island Coulee near Hansboro, N.Dak.	.955	.082214	2.010
05098800 Cypress Creek near Sarles, N.Dak.	.997	.100000	1.980
05099600 Pembina River at Walhalla, N.Dak.	.997	.077419	1.390
05100000 Pembina River at Neche, N.Dak.	.958	.033392	1.340

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 to failure since the last servicing visit. 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.

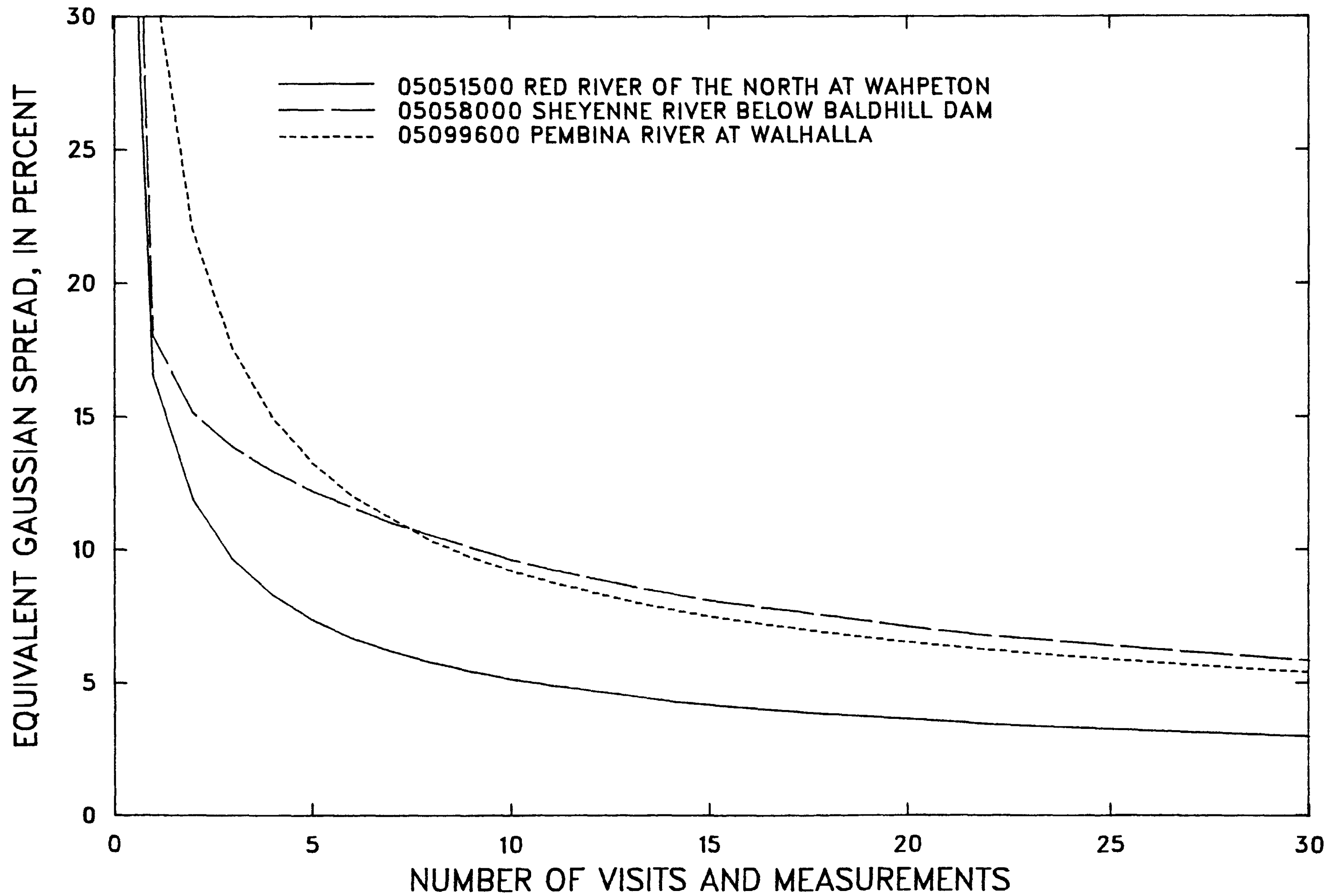

Figure 2.—Typical uncertainty functions for instantaneous discharge.

Data collected in North Dakota in recent years were reviewed to define the average time to failure for recording equipment and stage-sensing devices. Little change in technology occurred during the period reviewed, and gaging stations were visited in a consistent pattern of about nine visits per year. During this period, gages were found to be malfunctioning an average of about 4.8 percent of the time. Because the K-CERA model analysis in North Dakota was confined to a 6-month nonwinter period, there was no reason to distinguish differences between gaging stations on the basis of exposure of equipment. The 4.8-percent missing record and a visit frequency of five times in 6 months (184 days) were used to determine an average time to failure of 408 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 gaging stations.

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 gaging station(s) used in reconstruction and the coefficient of variation of daily discharges at the gaging station. Daily streamflows for the last 30 water years (or period of record if less) were used to define seasonally averaged coefficients of variation for each gaging station (table 4).

Many different sources of information are used in reconstructing periods of missing record. These sources include, but are not limited to, recorded ranges in stage (for graphic recorders with clock stoppage), known discharges on adjacent days, recession analysis, observer's staff-gage readings, weather records, high-water-mark elevations, and comparison with nearby gaging stations. However, most of these techniques are unique to a given gaging station or to a specific period of missing record. Using all information available, short periods (several days) of missing record usually can be reconstructed quite accurately. An even longer period (more than a month) of missing 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 missing record at individual gaging stations.

Discussion of Routes and Costs

Twenty-nine continuous-record surface-water gaging stations in the Grand Forks field area as well as seven seasonal gaging stations (record from March through September), six stage and lake stations, and several crest-stage gages and miscellaneous sites are serviced on field trips. All sites except the continuous-record surface-water gaging stations are considered to be null stations in that they do not contribute to the uncertainty of the continuous-record network. Operating budgets and associated costs for the null stations are included in the surface-water operating budget being analyzed in phase three.

As previously indicated, uncertainty functions cannot be defined for 4 of the 29 continuous-record surface-water gaging stations. These four gaging stations are treated as null stations.

Minimum visit constraints were defined for each of the 29 gaging stations prior to defining the practical service routes. A minimum of three visits was established for all gaging stations in the network (including all the null stations) for the 184-day study period in order to minimally maintain equipment. However, only three visits during the period probably would lead to increased incidence of equipment failure.

Practical routes to service the 29 gaging stations were determined after consultation with personnel responsible for maintaining the gaging stations, and after the uncertainty functions and minimum visit requirements were considered. Forty-two routes to service all the gaging stations in the Grand Forks field area were identified. These included routes that currently are used, alternative routes under consideration as future possibilities, routes used to service certain key gaging stations, and route combinations that grouped proximate gaging stations where levels of uncertainty indicated 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 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. Fixed costs were calculated for all gaging stations in the network.

Visit costs are those associated with paying the hydrographer for the time actually spent at a gaging station making a discharge measurement. 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 generally is made. An average visit time, in hours, was estimated for each gaging station, based on historical operations. The average number of hours then was multiplied by the average hourly salary of the hydrographers in Grand Forks to determine visit costs for each gaging station.

Route costs include the vehicle cost associated with the number of miles required to cover the route, the cost of the hydrographer's time while in transit, the cost of servicing the equipment at the gaging stations, and any per diem associated with the time needed to complete the trip.

The model was run using a 184-day period with the added requirement that four visits would be made during the remaining 181 days of the year. Fixed costs were calculated on an annual basis, but visit and route costs were applied only when a trip was made. In order for all costs to be applied on an annual basis, visit and route costs for the four winter visits to each gaging station were added to fixed costs for each gaging 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 stream-gaging program. The objective function in the optimization process is the sum of the variances of the errors of instantaneous discharge (in percent squared) for the entire gaging-station network.

Present practices to define the associated, total uncertainty were simulated by restricting the specific routes and the number of visits to each gaging station to those now being used. This was done only to compute the EGS of present practice; no optimization was done. The restrictions then were removed, and the model was allowed to define optimal visit schedules for the current budget. The optimization procedure was repeated for other possible budgets. Results for both the present operation and the optimal solutions are shown in figure 3 and in table 5. Both standard error and EGS are included in table 5 for comparison purposes.

The results in figure 3 and table 5 are based on the assumption that a discharge measurement is made each time a gaging station is visited. The percentage values also represent only the 6 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 represents the minimum level of uncertainty that can be obtained for a given budget, with existing technology. An additional assumption to consider when interpreting the results is the applicability of the Markov process to all gaging stations.

The current operating policy results in an average EGS of about 16.5 percent for nonwinter streamflow. This policy is based on a budget of about $248,000 for operating the 29-station stream-gaging network. Using the current budget, the average EGS could be reduced to about 14.7 percent by altering route schedules to achieve more frequent visits to gaging stations where uncertainty is large, accompanied by less frequent visits to gaging stations where uncertainty is small.

A budget of about $235,000 could be used to operate the program, but the magnitude of errors would increase. Gaging stations would have to be eliminated from the program if the budget were less than $235,000. At a budget level of $235,000, the optimal average EGS is increased to about 20.4 percent.

The maximum budget analyzed was $400,000, about 60 percent more than the current budget. This budget resulted in an optimal average EGS of about 5.8 percent. Thus, a 60-percent increase in the budget would reduce the optimal average EGS obtainable under the current budget from about 14.7 percent to about 5.8 percent.

For the current operation, missing record also impacts EGS. Thus, improvements in instrumentation, increased use of local observers, increased use of data telemetry equipment, and changes in routes could have a positive impact on uncertainties of instantaneous discharges.

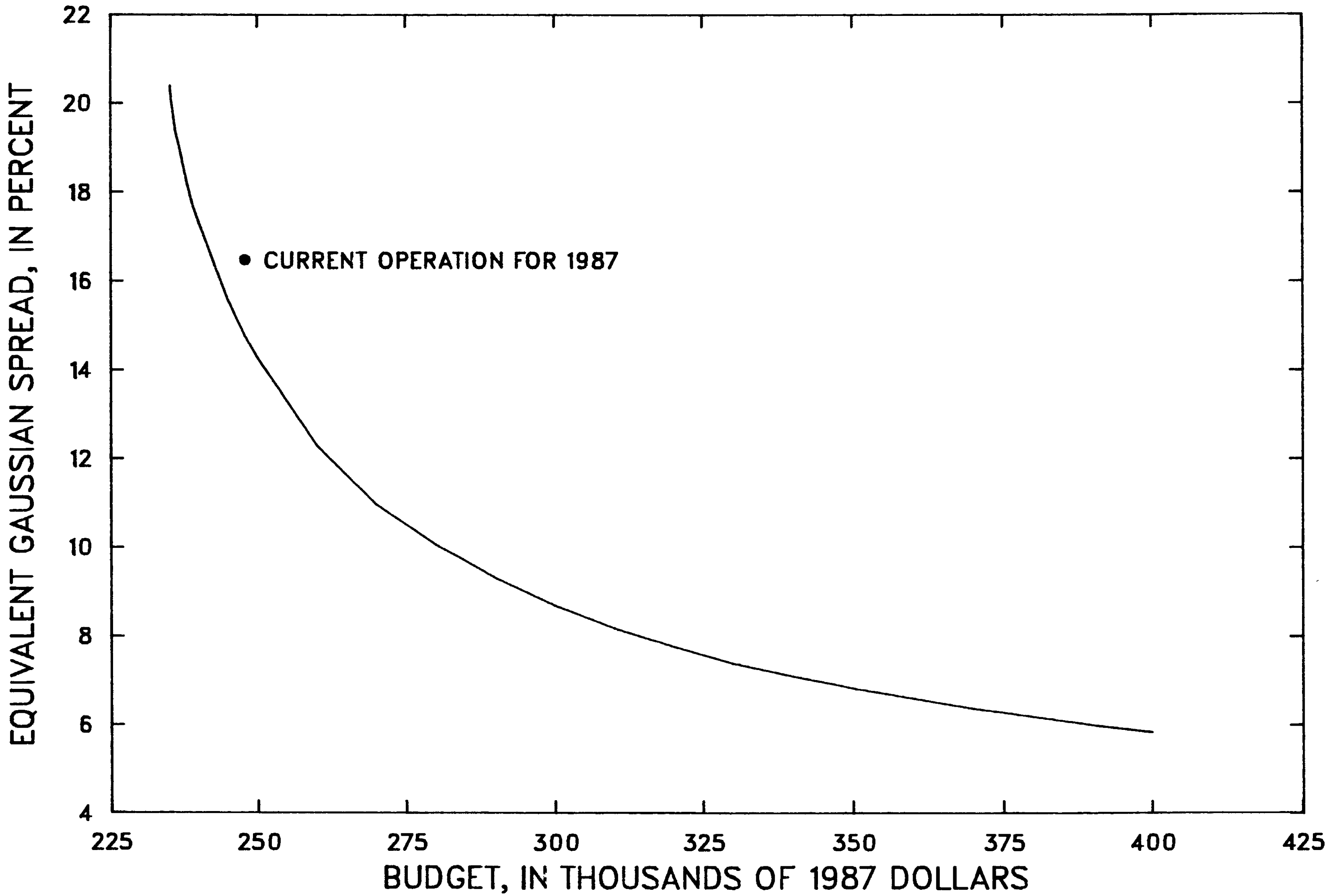

Figure 3.—Relation between average equivalent Gaussian spread and budget.

Table 5.--*Selected results of analysis*

Gaging station number	Standard error of instantaneous discharge, in percent [Equivalent Gaussian spread (EGS), in percent] (Number of visits per season)						
	Budget, in thousands of 1987 dollars						
	Current operation	Optimized values					
	248	235	248	280	310	350	400
Average standard error for the network	25.0	30.1	21.6	14.9	12.2	10.2	8.7
Average EGS for the network	16.5	20.4	14.7	10.1	8.2	6.8	5.8
05051500	12.7 [7.4] (5)	16.1 [9.6] (3)	16.1 [9.6] (3)	11.6 [6.7] (6)	9.6 [5.4] (9)	8.0 [4.5] (13)	6.3 [3.5] (21)
05051522	10.6 [8.6] (5)	13.2 [10.8] (3)	13.2 [10.8] (3)	9.8 [7.9] (6)	8.1 [6.4] (9)	6.8 [5.3] (13)	5.4 [4.2] (21)
05054000	9.4 [4.1] (5)	12.0 [5.4] (3)	12.0 [5.4] (3)	8.6 [3.7] (6)	7.1 [3.0] (9)	5.9 [2.5] (13)	4.7 [2.0] (21)
05056000	40.4 [34.9] (6)	48.8 [42.9] (4)	31.4 [26.6] (10)	21.0 [17.4] (22)	17.7 [14.5] (31)	14.5 [11.9] (46)	12.5 [10.2] (62)
05056400	24.7 [22.3] (5)	31.5 [29.2] (3)	24.7 [22.3] (5)	17.5 [15.4] (10)	13.8 [12.0] (16)	11.8 [10.2] (22)	10.1 [8.7] (30)
05057200	37.2 [25.3] (5)	47.6 [33.9] (3)	31.5 [20.9] (7)	21.5 [13.9] (15)	17.4 [11.1] (23)	14.6 [9.3] (33)	12.4 [7.9] (46)
05058000	19.8 [12.2] (5)	24.2 [13.8] (3)	19.8 [12.2] (5)	14.2 [9.2] (11)	11.3 [7.4] (18)	9.5 [6.3] (26)	8.3 [5.5] (34)

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

Gaging station number	Standard error of instantaneous discharge, in percent [Equivalent Gaussian spread (EGS), in percent] (Number of visits per season)						
	Budget, in thousands of 1987 dollars						
	Current operation	Optimized values					
	248	235	248	280	310	350	400
05058700	16.4 [14.1] (5)	19.7 [17.0] (3)	16.4 [14.1] (5)	11.5 [9.8] (11)	9.1 [7.7] (18)	7.6 [6.4] (26)	6.6 [5.6] (34)
05059000	17.2 [17.2] (5)	20.2 [20.0] (3)	17.2 [17.0] (5)	12.3 [12.0] (11)	9.7 [9.4] (18)	8.1 [7.8] (26)	7.1 [6.8] (34)
05059500	12.3 [11.9] (5)	15.4 [15.0] (3)	15.4 [15.0] (3)	11.3 [10.9] (6)	9.3 [8.8] (9)	7.5 [7.0] (14)	6.1 [5.7] (21)
05059700	34.2 [11.5] (5)	43.7 [15.4] (3)	29.0 [9.6] (7)	20.0 [6.4] (15)	15.8 [5.0] (24)	13.1 [4.2] (35)	11.5 [3.7] (46)
05060500	38.0 [25.3] (5)	42.4 [28.8] (4)	28.4 [18.2] (9)	19.6 [12.3] (19)	15.9 [9.9] (29)	13.1 [8.1] (43)	10.9 [6.8] (62)
05064500	7.1 [4.3] (5)	8.5 [4.6] (3)	8.5 [4.6] (3)	7.6 [4.5] (4)	6.6 [4.2] (6)	6.0 [4.1] (8)	5.2 [3.7] (12)
05064900	36.2 [27.6] (5)	40.2 [30.9] (4)	24.7 [18.2] (11)	15.7 [11.3] (27)	13.0 [9.3] (40)	10.9 [7.8] (57)	9.6 [6.9] (73)
05066500	38.5 [35.5] (8)	40.8 [37.6] (7)	29.7 [27.1] (14)	19.3 [17.3] (33)	15.6 [14.0] (50)	13.2 [11.8] (70)	11.4 [10.1] (94)
05082500	6.3 [4.6] (6)	6.0 [4.4] (7)	6.3 [4.6] (6)	6.3 [4.6] (6)	5.3 [4.0] (10)	4.5 [3.4] (15)	4.0 [3.1] (19)

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

Gaging station number	Standard error of instantaneous discharge, in percent [Equivalent Gaussian spread (EGS), in percent] (Number of visits per season)						
	Budget, in thousands of 1987 dollars						
	Current operation	Optimized values					
	248	235	248	280	310	350	400
05083600	37.2 [27.2] (6)	44.6 [32.9] (4)	27.8 [19.9] (11)	18.4 [13.0] (25)	15.4 [10.7] (36)	12.8 [8.9] (52)	11.0 [7.7] (70)
05084000	25.0 [21.8] (6)	30.6 [27.2] (4)	18.5 [15.7] (11)	12.3 [10.2] (25)	10.3 [8.5] (36)	8.6 [7.1] (52)	7.5 [6.2] (70)
05085000	25.5 [14.7] (6)	35.5 [21.5] (3)	20.9 [11.8] (9)	14.1 [7.7] (20)	11.4 [6.3] (31)	9.7 [5.3] (43)	8.1 [4.5] (62)
05089000	28.0 [12.4] (5)	31.2 [14.1] (4)	21.0 [9.1] (9)	14.2 [6.1] (20)	11.7 [5.1] (30)	9.9 [4.4] (42)	8.3 [3.8] (61)
05092000	9.6 [9.2] (5)	10.6 [10.1] (3)	10.6 [10.1] (3)	10.6 [10.1] (3)	8.8 [8.5] (7)	6.9 [6.6] (14)	5.8 [5.6] (21)
05098700	55.7 [48.0] (6)	65.1 [56.2] (4)	40.7 [34.6] (12)	27.6 [23.1] (26)	23.1 [19.2] (37)	19.2 [15.9] (53)	16.1 [13.3] (75)
05098800	37.3 [15.3] (5)	47.8 [20.6] (3)	28.0 [11.1] (9)	18.0 [7.0] (22)	14.6 [5.7] (34)	12.1 [4.8] (50)	10.3 [4.2] (69)
05099600	15.7 [13.3] (5)	20.1 [17.6] (3)	20.1 [17.6] (3)	12.5 [10.3] (8)	10.7 [8.8] (11)	8.9 [7.3] (16)	7.7 [6.2] (22)
05100000	30.6 [30.4] (5)	35.5 [35.4] (3)	28.7 [28.4] (6)	18.0 [17.6] (17)	14.8 [14.4] (25)	12.2 [11.8] (37)	10.7 [10.3] (48)

Summary of Phase Three of Analysis

As a result of this phase of the analysis, conclusions are as follow:

1. Travel routes and measurement frequencies now in use are near the optimal level. Changes in routes and measurement frequencies optimally could result in a 1.8-percent decrease in the average EGS.
2. Decreasing the budget to about $235,000 would increase the optimal average EGS from 14.7 to 20.4 percent. Increasing the budget to about $400,000 would reduce the optimal average EGS from 14.7 percent to 5.8 percent.
3. Methods for decreasing the probabilities of missing record need to be explored. These methods may include improved instrumentation as well as increased use of local observers and data telemetry equipment.

SUMMARY

Phase one of the analysis of the North Dakota stream-gaging program was completed in 1985. Phase one identified data uses and funding sources for 94 gaging stations operated in 1984 and indicated that current uses of surface-water data justified continued operation of all gaging stations.

Phase two of the analysis was to investigate the potential for using methods to simulate streamflow records. Statistical and hydrologic flow-routing methods were chosen as the alternative methods. Accuracy of the alternative methods was sufficient to consider discontinuing only one gaging station--the Missouri River at Bismarck (06342500). Data needs, however, cannot be met with a synthesized record; therefore, the gaging station will continue in operation.

Phase three of the analysis was to evaluate the operation of the gaging-station networks by using associated uncertainty in streamflow records for various operating budgets. Phase three analysis was limited to the nonwinter operation of the network of gaging stations operated by the U.S. Geological Survey's Grand Forks Field Headquarters. Operations in the Grand Forks area are considered representative of the stream-gaging program in North Dakota and provide a basis for considering changes in operating procedures.

Travel routes and measurement frequencies now in use in the Grand Forks field area are near the optimal level. Changes in routes and measurement frequencies optimally could decrease the average equivalent Gaussian spread by 1.8 percent, from 16.5 to 14.7 percent.

Decreasing the budget to about $235,000 would increase the optimal average equivalent Gaussian spread from 14.7 to 20.4 percent. Conversely, increasing the budget to $400,000 would reduce the optimal average equivalent Gaussian spread from 14.7 percent to 5.8 percent.

For the current operation, missing record also affects the equivalent Gaussion spread. Thus, improvements in instrumentation, increased use of local observers, increased use of data telemetry equipment, and changes in routes could have a positive effect on uncertainities 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.

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 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., editor, 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.

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.

Ryan, G.L., 1985, Data uses and funding of the streamflow-gaging program in North Dakota: U.S. Geological Survey Open-File Report 85-349, 29 p.

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

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

DESCRIPTION OF MATHEMATICAL PROGRAM

The following description of the computations and mathematical relations is modified from Fontaine and others (1984, p. 22-26). In a study of the cost effectiveness of a network of gaging stations 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 gaging stations to be determined by the total variance (uncertainty) in either the annual mean discharge or the instantaneous discharge at all sites involved in the stream-gaging program and 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 (e.g., gage height) for each gaging station involved in the stream-gaging program. The sequence of discharge residuals (in logarithmic units) from this rating (the discharge measurement minus the rating value) is analyzed as a time series.

The second step is to fit a lag-1-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-1-day autocorrelation, 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 are available for use in the rating equation. The K-CERA methodology, along with the assumption of a first-order Markov process, is used to determine the variance, V_f, as a function of the number of discharge measurements per year (Moss and Gilroy, 1980).

If primary correlative data at the site are not available, the discharge may be estimated by correlation with nearby sites. The correlation coefficient, ρ_c, between the streamflows with seasonal trends removed (detrended) at the site of interest and streamflows detrended 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, which is explained by data from other sites, is ρ_c^2. The variance of the percent error in streamflows at the primary site, in the absence of data at both the primary site and nearby sites, is

$$C_V = 100 \frac{1}{365} \sum_{i=1}^{365} \frac{\sigma_i}{\mu_i}^{2\ \frac{1}{2}} \qquad (5)$$

where

σ_i = the square root of the variance of daily discharges for the ith day of the year, and

μ_i = 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

$$V_r = (1-\rho_c^2)C_v^2 \tag{6}$$

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

$$V_e = C_v^2. \tag{7}$$

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

$$V_T = \epsilon_f V_f + \epsilon_r V_r + \epsilon_e V_e. \tag{8}$$

The fraction uptime, ϵ_f, of the primary recorders at the site of interest is modeled by a truncated negative exponential probability distribution that depends on τ, 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 concurrent downtime of the primary and secondary sites is found by assuming independence of downtimes between sites (Fontaine and others, 1984).

The variance V_T given by equation 8, 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 variances, V_T, over all sites is taken as the measure of the uncertainty of the network. The variance V_T given by equation 8 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

$$g_T = \epsilon_f g_f + \epsilon_r g_r + \epsilon_e g_e. \tag{9}$$

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 ϵ_e may be very small, V_e may be extremely large. Actually, this standard error interval may include up to 99 percent of the errors.

To assist in interpreting the results of the analyses, a new parameter, equivalent Gaussian spread (EGS), 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 missing record, the greater part of the time.

The cost part of the input to the K-CERA methodology consists of determining practical routes to visit the sites in the stream-gaging network, the cost of each route, the cost of a visit to each site, the fixed cost of each site, 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 gaging stations for such purposes as necessary for periodic maintenance, rejuvenation of recording equipment, or required periodic sampling to obtain water-quality data. Such special requirements are considered to be inviolable constraints in terms of the minimum number of vists to each gaging station.

All costs, routes, constraints, and uncertainty functions then are 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 gaging station is 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 nonlinearity of the problem, the optimal solution may really be "near optimal." (See Moss and Gilroy, 1980, or Fontaine and others, 1984.)