Analysis of Potential Errors in Real-Time Streamflow Data and Methods of Data Verification by Digital Computer
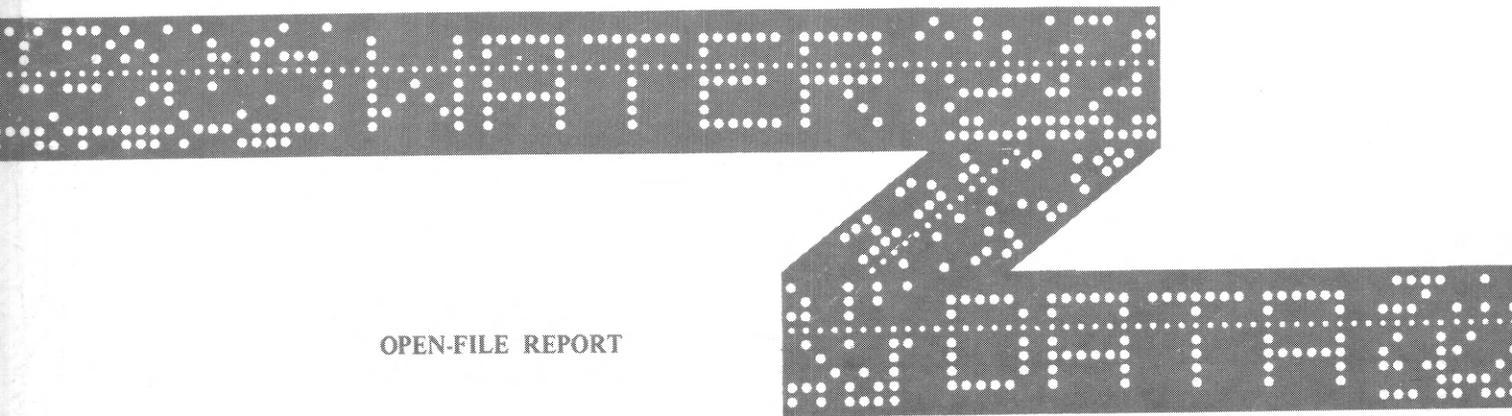
Analysis of Potential Errors in Real-Time Streamflow Data and Methods of Data Verification by Digital Computer

--

By David J. Lystrom

--

CONTENTS

ILLUSTRATIONS

iii

TABLE

ANALYSIS OF POTENTIAL ERRORS IN REAL-TIME STREAMFLOW DATA AND METHODS OF DATA VERIFICATION BY DIGITAL COMPUTER

--

By David J. Lystrom

--

ABSTRACT

The magnitude, frequency, and types of errors inherent in real-time streamflow data are presented in part I. It was found that real-time data are generally less accurate than are historical data, primarily because real-time data are often used before errors can be detected and corrections applied.

Various methods of verifying real-time streamflow data are outlined in part II. Relatively large errors (those greater than 20-30 percent) can be detected readily by use of well-designed verification programs for a digital computer, and smaller errors can be detected only by discharge measurements and field observations. The capability to substitute a simulated discharge value for missing or erroneous data is incorporated in some of the verification routines described. The routines represent concepts ranging from basic statistical comparisons to complex watershed modeling and provide a selection from which real-time data users can choose a suitable level of verification.

INTRODUCTION

Real-time data are any information reported at the time of or shortly after occurrence. An automated network for collecting, processing, and disseminating real-time hydrometeorological data is being developed in the Pacific Northwest by an interagency group. This network (referred to as the hydromet network) will utilize a dedicated computer system to convert field data from more than 400 data-collection sites into timely information needed for efficient water management and water-quality surveillance. The reporting interval for real-time data will range from hourly to daily.

As a participant in the development and operation of this network, the Geological Survey anticipates providing the technology for processing streamflow information for more than 150 gaging stations proposed

1

for automation, and assuming responsibility for quality control (verification).

At this point in time, not much is known about the potential accuracy of streamflow information acquired by way of a real-time system, nor about accuracy requirements of real-time data uses. These data are proposed to be used both in a large-scale digital model of the Columbia Basin and on an independent station-by-station basis. Because the usefulness of the data will vary with accuracy, it has been assumed that data are desired to be as accurate as possible. The obvious limit to accuracy is cost.

The accuracy of historical streamflow data has been maintained by applying after-the-fact corrections to data as much as a year or more after occurrence. These corrections are determined by analysis of such factors as discharge measurements, field observations, weather records, and comparison of hydrographs of adjacent streamflow stations. Because real-time data are often used within hours after acquisition, meaningful corrections can be applied only forward in time after an error has been detected. Consequently, traditional methods of verification are impractical, and real-time data will be less accurate than historical data. The accuracy of real-time data can be improved by increasing the frequency of field visits to gaging stations and/or by expedient verification techniques. Logically, verification must be done mostly by automated means as part of the network data-processing system. The costs of various computer verification techniques are approximated in this report in terms of computer programming and computer facilities required.

The degree of verification desired for optimum quality control depends on (1) the magnitude and frequency of potential errors intrinsic to a real-time data-acquisition system; (2) the cost of computer hardware, programming, and processing time required for verification; and (3) the need for accuracy. The data users participating in the development of the hydromet network must select an optimum level of verification on the basis of quality requirements compared to cost of verification.

This report, presented in two parts, is intended to provide data users with information on the potential quality of real-time streamflow data and several methods by which quality can be improved through computerized verification. The frequency, type, and magnitude of potential errors are estimated from a simulated version of real-time data in part I. Concepts of various verification routines applicable to computerized detection of errors, and which can be incorporated readily in a processing system, are outlined in part II.

The real-time data-verification concepts in part II were designed in consideration of the following:

2

1. Magnitude and frequency of errors inherent in real-time streamflow data.

2. Adaptability to digital computer processing.

3. Density of streamflow stations in the hydromet network.

Because the list of hydromet stations is presently (1971) in a proposed status, only general functions and applicability of the various verification routines are discussed in part II.

The scope of this project does not include computer programming for quality control nor for data processing. These programs would be prepared after data-user agencies have selected the verification routines necessary to meet the required level of quality control and have designated the computer hardware and software to be implemented in the hydromet system.

PART I

ANALYSIS OF POTENTIAL INACCURACIES IN STREAMFLOW DATA ACQUIRED
VIA A REAL-TIME SYSTEM

Scope of the Analysis

The purpose of part I is to determine the magnitude and frequency of potential data errors in a real-time streamflow-data-collection system.

Most streamflow data in the Pacific Northwest hydromet system will be acquired by transmitting stream stages from traditional USGS-type gaging stations by radio to a central computer where discharges will be computed from a stage-discharge relation (rating curve). Lindsay, Kohler, and Paulhus (1949, p. 182-189) describe USGS-type gaging stations. The transmission interval may range from hourly to daily. The assumption is made that the potential magnitude and frequency of errors in hourly discharges are virtually equivalent to those in mean daily discharges. Mean daily discharge is defined as the average rate of streamflow during a day, and is referred to as "daily discharge" in this report. Potential inaccuracies are defined in this study on the basis of daily discharges and are considered applicable to discharges for shorter periods.

Potential errors in real-time daily discharges analyzed for this report are categorized into three groups according to their source, as follows:

1. Changes in the stage-discharge relation caused by shifting channel.

2. Backwater from ice or other obstructions.

3. Sensing and recording equipment malfunctions.

Failure of transmitting and interfacing devices was not included in this study because adequate information on the reliability of these components was not available at the time of this analysis.

Procedures and Concepts

Water years 1964, 1965, and 1967 were selected for the study period because they represent recent conditions of average, high, and low flows, respectively. Historical records for 109 stations proposed for inclusion in the North Pacific hydromet network were analyzed for occurrences of conditions that could cause errors in the acquisition of real-time data. Table 1 shows a listing of the conditions and provides a comparative reference to the susceptibility of each of the 109 stations to the three types of errors itemized above.

Following analysis of the 109 station records, 14 stations, identified by asterisks in table 1, were selected for recomputation of daily discharges in a simulated real-time concept. In this simulation, daily discharges were computed in the conventional (USGS) manner by using the same gage heights and results of discharge measurements as were used for historical records, but application of adjustments was conducted within the following constraints:

1. Real-time processing does not allow for after-the-fact reconsiderations; therefore, all adjustments for ice, shifting channel, and equipment malfunction were applied only forward in time, relative to when the condition was detected by field observation.

2. A lag period of 5 days was assumed between a field visit and the time when field notes were available for application. Hopefully, in actual operation of a real-time system, this lag period can be reduced.

3. Only actual field measurements and observations were used as a basis for adjustments. The intent is to show data errors that would occur if real-time data were not verified and extra field visits were not made.

The recomputed daily records for the 14-station group were compared with corresponding historical records for the 3-year study period. Errors were computed as the absolute percentage deviation of the recomputed discharges from published discharges. Daily errors were categorized according to source and magnitude. When simultaneous effects of two types of errors occurred on the same day (such as rating change and

4

ice effect), the error was attributed only to the more significant of the two effects for that day.

The discrepancies that were found represent errors that would have occurred if records for the 14-station group had been processed in a real-time system within the constraints itemized above. Because 5 percent is generally presumed to be the average accuracy of historical daily discharge data, errors of less than 5 percent were not evaluated in this study.

Analysis of the 109-station group (data shown in table 1) includes only occurrences of conditions that could cause the three types of errors in real time, whereas the error analysis of the 14-station group provides a quantitative sampling of the number of days of error produced by each condition. The average number of days of error for the 109-station group was approximated by multiplying the average number of days of error in the 14-station group by ratio of the average number of error-causing conditions per station year in the 109-station group versus the average number of conditions sampled in the 14-station group.

The three types of errors were evaluated separately, and the results are given in the following sections.

Evaluation of Real-Time Data Errors

Errors Caused by Changes in the Stage-Discharge Relation

When the channel geometry at a gaging station changes through natural or manmade causes, the stage-discharge relation is altered. As a result, either a new rating or an adjustment to the existing rating must be applied. Unless rating changes are detected as soon as they occur, error is introduced into real-time data. The magnitude of error depends on the degree of channel change, whereas the duration of the error depends on promptness of detection and correction of the rating definition used by the processing computer. Table 1 indicates the occurrence of 535 rating changes greater than 5 percent during the study period for all 109 stations, or an average of 1.64 rating changes greater than 5 percent per station year. Only nine of the 109 stations had no rating change during the 3-year period.

A comparison of 3 years of simulated real-time data for 14 stations with historical records revealed potential errors due to rating changes. The magnitudes of these errors and total number of days of occurrence are shown below.

Percent error	≥ 5	≥ 10	≥ 20	≥ 30	≥ 50	≥ 100
Number of days	2,567	968	272	86	24	14

Table 1.--Occurrence of conditions that can cause errors in real-time data, water years 1964, 1965, and 1967

Station number	Station name	Number of rating changes		Number of ice-affected days	No. of days of equipment malfunctions	
		≧5%	≧10%		Record-er**	Sens-or***
12027500*	Chehalis R. nr Grand Mound, Wash.	10	6	0	9	0
12031000	Chehalis R. at Porter, Wash.	5	5	0	4	9
12089500	Nisqually R. at McKenna, Wash.	9	5	0	0	0
12092000	Puyallup R. nr Electron, Wash.	15	12	0	0	0
12093500	Puyallup R. nr Orting, Wash.	10	7	0	0	66
12094000	Carbon R. nr Fairfax, Wash.	7	6	0	34	174
12098500	White R. nr Buckley, Wash.	9	4	0	0	4
12101500	Puyallup R. at Puyallup, Wash.	7	6	0	19	0
12108500	Newaukum Cr. nr Black Diamond, Wash.	15	12	0	47	2
12113000	Green R. nr Auburn, Wash.	3	2	0	28	0
12134500	Skykomish R. nr Gold Bar, Wash.	4	0	0	18	4
12149000	Snoqualmie R. nr Carnation, Wash.	2	2	0	32	13
12189500	Sauk R. nr Sauk, Wash.	5	4	0	28	0
12194000	Skagit R. nr Concrete, Wash.	1	0	0	65	0
12205000*	N.F. Nooksack R. nr Glacier, Wash.	13	11	4	50	4
12210500	Nooksack R. at Deming, Wash.	17	15	2	0	6
12301300	Tobacco R. nr Eureka, Mont.	6	5	59	0	0
12303000*	Kootenai R. at Libby, Mont.	9	6	41	9	20
12304500	Yaak R. nr Troy, Mont.	3	0	75	54	0
12307500	Moyie R. at Eileen, Idaho	2	0	81	0	0
12340500*	Clark Fork ab Missoula, Mont.	8	5	41	0	0
12353000	Clark Fork bl Missoula, Mont.	1	1	3	35	0
12354000	St. Regis R. nr St. Regis, Mont.	9	7	1	0	0
12354500	Clark Fork at St. Regis, Mont.	3	1	9	0	2
12358500	M.F. Flathead R. nr West Glacier, Mont.	4	1	0	39	5

See footnotes at end of table.

6

Table 1.--Occurrence of conditions that can cause errors in real-time data, water years 1964, 1965, and 1967--Continued

Station number	Station name	Number of rating changes		Number of ice-affected days	No. of days of equipment malfunctions	
		≥5%	≥10%		Record-er**	Sens-or***
12363000	Flathead R. at Columbia Falls, Mont.	3	1	0	0	28
12370000	Swan R. nr Bigfork, Mont.	0	0	0	45	9
12389000	Clark Fork nr Plains, Mont.	1	0	9	76	0
12395000	Priest R. nr Priest River, Idaho	2	0	80	183	0
12396500	Pend Oreille R. nr Ione, Wash.	0	0	0	0	3
12398600	Pend Oreille R. at International Boundary	0	0	0	55	12
12404500*	Kettle R. nr Laurier, Wash.	6	0	55	36	0
12409000	Colville R. at Kettle Falls, Wash.	7	1	0	6	13
12413000	Coeur d'Alene R. at Enaville, Idaho	6	4	7	0	0
12414500	Placer Cr. at Wallace, Idaho	4	0	112	1	0
12435500	Feeder Canal at Grand Coulee, Wash.	0	0	0	32	12
12436500	Columbia R. at Grand Coulee, Wash.	0	0	0	57	0
12445000	Okanogan R. nr Tonasket, Wash.	4	2	14	61	0
12459000	Wenatchee R. at Peshastin, Wash.	0	0	21	17	0
12476000	Kachess R. nr Easton, Wash.	3	1	0	0	0
12484500	Yakima R. at Umtanum, Wash.	0	0	0	35	23
12494000*	Naches R. nr Naches, Wash.	11	2	8	11	11
12510500	Yakima R. at Kiona, Wash.	2	1	0	95	0
13037500*	Snake R. at Heise, Idaho	3	2	37	9	0
13058000*	Willow Cr. nr Ririe, Idaho	8	2	256	29	0
13060000	Snake R. nr Shelley, Idaho	2	0	163	53	0
13088000	Snake R. at Milner, Idaho	5	2	0	0	0
13205500*	Boise R. at Boise, Idaho	10	4	7	20	0
13249500	Payette R. nr Emmett, Idaho	2	0	0	69	0
13258500	Weiser R. nr Cambridge, Idaho	6	4	20	30	67

See footnotes at end of table.

Table 1.--Occurrence of conditions that can cause errors in real-time data, water years 1964, 1965, and 1967--Continued

Station number	Station name	Number of rating changes ≥5%	≥10%	Number of ice-affected days	No. of days of equipment malfunctions Record-er**	Sens-or***
13266000	Weiser R. nr Weiser, Idaho	7	2	103	0	0
13269000	Snake R. at Weiser, Idaho	4	1	0	0	93
13302500*	Salmon R. at Salmon, Idaho	6	1	78	0	0
13317000	Salmon R. at White Bird, Idaho	5	2	2	0	0
13323500*	Grande Ronde R. nr Elgin, Oreg.	2	1	60	77	2
13333000	Grande Ronde R. at Troy, Oreg.	3	1	4	13	0
13334300	Snake R. nr Anatone, Wash.	0	0	0	0	5
13336500	Selway R. nr Lowell, Idaho	6	1	7	92	5
13337000	Lochsa R. nr Lowell, Idaho	5	2	3	67	11
13342500	Clearwater R. at Spalding, Idaho	5	1	5	18	47
13351000	Palouse R. at Hooper, Wash.	2	0	6	8	0
14018500	Walla Walla R. nr Touchet, Wash.	7	7	11	18	20
14021000	Umatilla R. at Pendleton, Oreg.	5	4	0	5	41
14033500*	Umatilla R. at Umatilla, Oreg.	7	5	6	0	0
14046000	N.F. John Day R. at Monument, Oreg.	5	3	4	20	0
14046500	John Day R. at Service Creek, Oreg.	2	1	9	2	0
14048000	John Day R. at McDonald Ferry, Oreg.	3	3	7	16	22
14080500	Crooked R. nr Prineville, Oreg.	4	3	0	0	0
14101500	White R. bl Tygh Valley, Oreg.	6	4	5	56	5
14103000	Deschutes R. at Moody, Oreg.	1	1	0	0	0
14113000	Klickitat R. nr Pitt, Wash.	1	1	0	0	0
14123500	White Salmon R. nr Underwood, Wash.	0	0	0	75	0
14128500	Wind R. nr Carson, Wash.	3	2	0	17	0
14144800	M.F. Willamette R. nr Oakridge, Oreg.	8	6	0	25	47
14144900	Hills Cr. ab Reservoir, nr Oakridge, Oreg.	4	4	0	0	33

See footnotes at end of table.

8

Table 1.--Occurrence of conditions that can cause errors in real-time data, water years 1964, 1965, and 1967--Continued

Station number	Station name	Number of rating changes		Number of ice-affected days	No. of days of equipment malfunctions	
		≧5%	≧10%		Record-er**	Sens-or***
14148000	M.F. Willamette R. bl N.Fork nr Oakridge, Oreg.	1	1	0	37	0
14152000	M.F. Willamette R. at Jasper, Oreg.	3	3	0	50	26
14152500	C.F. Willamette R. at London, Oreg.	3	2	0	54	118
14154500*	Row R. ab Pitcher Creek nr Dorena, Oreg.	8	5	0	0	2
14157500	C.F. Willamette R. nr Goshen, Oreg.	3	3	0	41	0
14159000	McKenzie R. at McKenzie Bridge, Oreg.	3	1	0	0	0
14159200	S.F. McKenzie R. nr Rainbow, Oreg.	4	3	0	0	0
14161100	Blue R. bl Tidbits Cr. nr Blue River, Oreg.	4	3	0	9	22
14161500	Lookout Creek nr Blue River, Oreg.	5	4	0	68	0
14166000	Willamette R. at Harrisburg, Oreg.	6	6	0	60	0
14166500	Long Tom River nr Noti, Oreg.	9	8	0	40	0
14167000	Coyote Creek nr Crow, Oreg.	7	5	0	8	46
14170000	Long Tom River at Monroe, Oreg.	10	6	16	0	3
14172000	Calapooia R. at Holley, Oreg.	6	3	0	0	4
14174000*	Willamette R. at Albany, Oreg.	6	4	0	0	0
14178000	N. Santiam R. bl Boulder Cr. nr Detroit, Oreg.	6	1	0	0	0
14179000	Breitenbush R. ab Canyon Cr. nr Detroit, Oreg.	5	2	0	25	0
14183000	N. Santiam R. at Mehama, Oreg.	1	0	0	52	0
14185800	M. Santiam R. nr Cascadia, Oreg.	4	4	0	54	14
14185900	Quartzville Cr. nr Cascadia, Oreg.	3	1	0	47	71
14187500	S. Santiam R. at Waterloo, Oreg.	5	2	0	51	0
14189000	Santiam R. at Jefferson, Oreg.	4	1	0	56	0
14190500	Luckiamute R. nr Suver, Oreg.	8	3	0	18	0
14191000	Willamette R. at Salem, Oreg.	6	2	0	0	0
14194000	S. Yamhill R. nr Whiteson, Oreg.	9	2	7	29	80

See footnotes at end of table.

6

Table 1.--Occurrence of conditions that can cause errors in real-time data, water years 1964, 1965, and 1967--Continued

Station number	Station name	Number of rating changes ≥5%	Number of rating changes ≥10%	Number of ice-affected days	No. of days of equipment malfunctions Record-er**	No. of days of equipment malfunctions Sens-or***
14200000	Molalla R. nr Canby, Oreg.	4	2	0	8	0
14211000	Clackamas R. nr Clackamas, Oreg.	4	1	0	6	72
14211500	Johnson Cr. at Sycamore, Oreg.	8	8	0	41	1
14243000	Cowlitz R. at Castle Rock, Wash.	2	0	0	6	2
14319500	N. Umpqua R. at Winchester, Oreg.	5	1	0	0	40
14321000	Umpqua R. nr Elkton, Oreg.	5	1	0	55	21
14359000	Rogue R. nr Raygold, Oreg.	6	3	0	39	46
14361500	Rogue R. at Grants Pass, Oreg.	4	0	0	18	6
14366000	Applegate R. nr Applegate, Oreg.	5	3	0	48	58

* Station used in detailed study.

** Includes digital and analog recorders.

*** Includes stilling well with float-mechanism and gas-purge sensing systems.
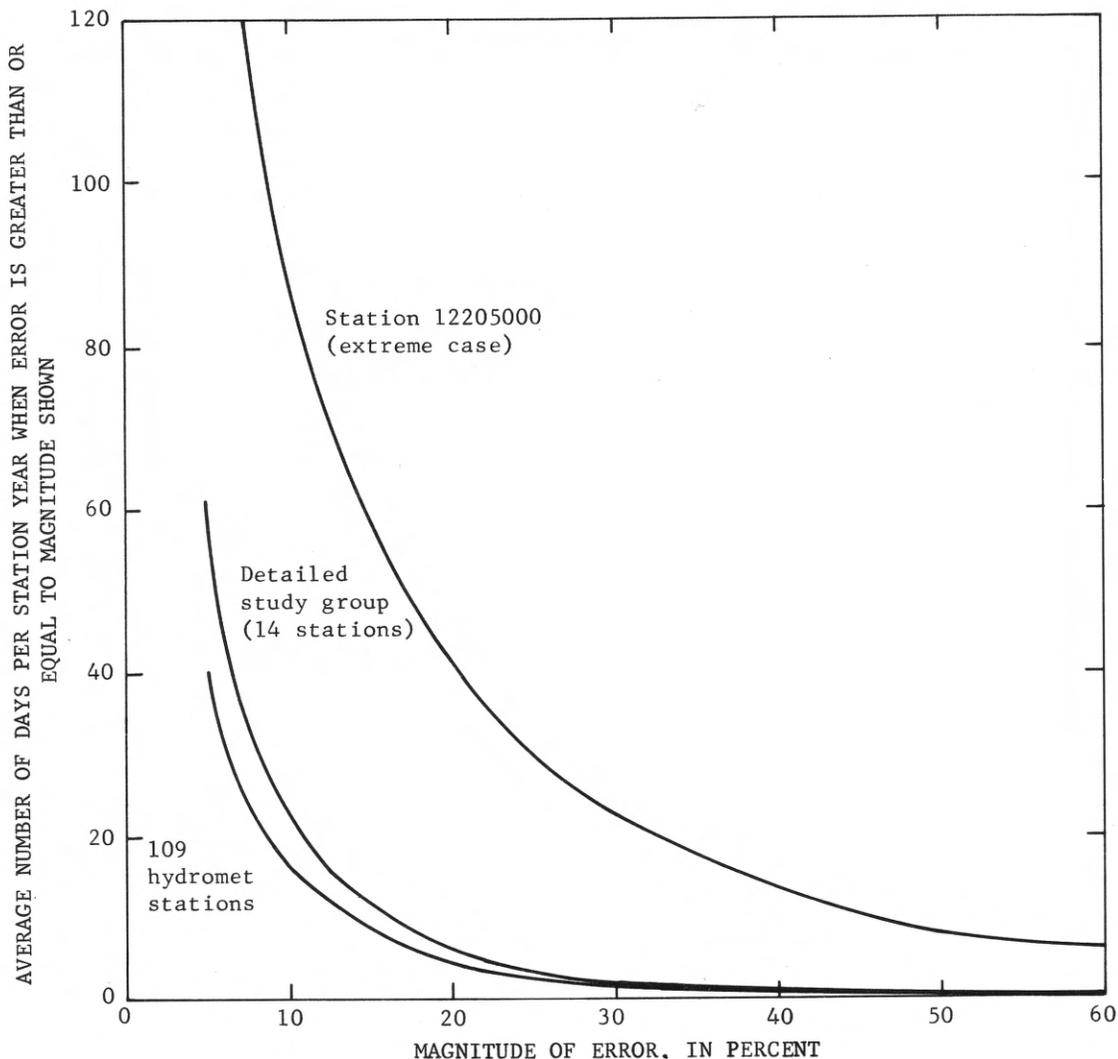
Figure 1.--Potential real-time data errors in mean daily discharge
caused by changes in the stage-discharge relation.

These data are also shown in figure 1, except that the number of days
plotted is the average number of days in error per station year.

The detailed-study group of 14 stations had a greater occurrence
of rating changes than did the entire group of 109 stations. The study
group had 1.28 rating changes with errors greater than 10 percent per
station year; the entire group had 0.92 rating changes of this magnitude
per station year. Thus, the curve in figure 1 for the entire group was
determined by adjusting the frequency (ordinate) of the study-group
curve by a factor of 0.72 (0.92/1.28=0.72).

The curve for station 12205000, North Fork Nooksack River near
Glacier, Wash., is shown in figure 1 to represent an extreme case of
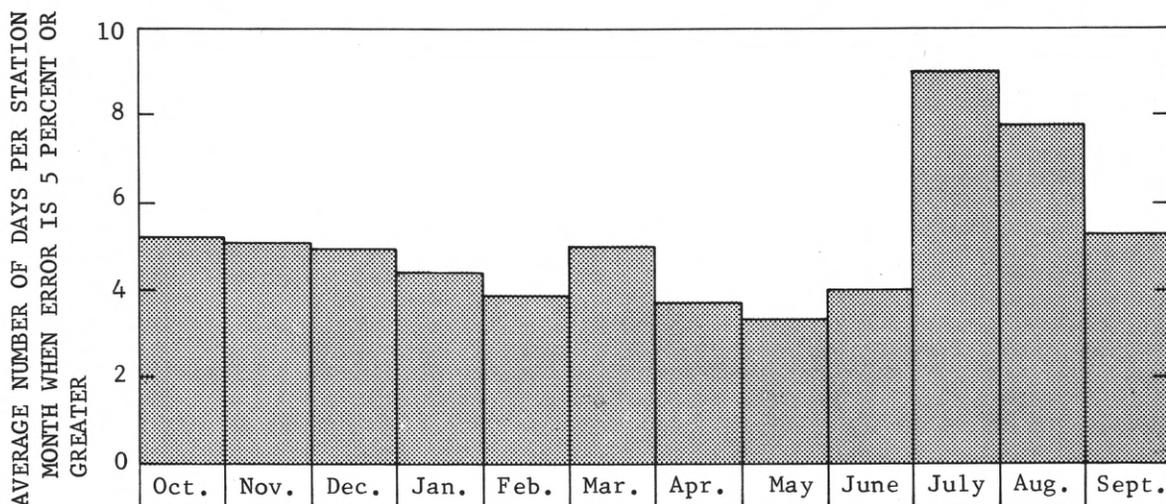real-time record error caused by rating changes.

11

Figure 2.--Average monthly distribution of errors in real-time mean daily
   discharges caused by changes in the stage-discharge relation for 14
   stations during water years 1964, 1965, and 1967.

The seasonal distribution of errors caused by rating changes is
shown in figure 2 for the 14-station group. For this group, rating
changes occur generally throughout the year, with somewhat more frequent
occurrence during late summer. Rating changes occur more frequently
during the low flows of late summer because a very slight channel change
represents a greater percentage of discharge than at high flows. At
some stations rating changes regularly occur during specific seasons.

Errors Caused by Backwater from Ice

The formation of ice in a stream channel may cause a significant
change in an established stage-discharge relation. Unless the condition
is detected and adjustments are made, a large error can be introduced
into real-time data. Historical records are generally computed for ice-
affected periods by a combination of methods utilizing hydrographic com-
parison with adjacent streams, weather records, and the appearance of
the stage-time or the discharge-time graph. This traditional method of
computing discharge for ice-affected periods is a manual procedure and
is not readily adaptable to a real-time data system unless it can be
converted and programmed for computer application. In the computation
of simulated real-time record for this study, errors caused by ice con-
ditions were based on the difference between published daily figures
and daily values computed by applying only flat corrections between
field measurements. Of the 14 stations selected for detailed study, 10
had ice effect during the 3-year study period.

The magnitude and duration of the errors presumably caused by ice, as determined for the 10-station study group, are as follows:

Percent error	≥5	≥10	≥20	≥30	≥50	≥100
Number of days	799	675	525	409	237	38

Because the magnitude of backwater from ice changes continually as the ice builds or melts, the correction determined by a discharge measurement may apply only for a short time. Therefore, the assumption is invalid that the correction remains nearly constant until the next measurement. This is apparent from comparison of the simulated real-time record with historical record. The simulated real-time record appears to show 799 days in error by more than 5 percent, whereas only 593 days are actually shown as ice affected in the historical record. This means that there were periods of no ice effect between measurements, when no corrections should have been applied. This indicates that it may not be advisable to apply flat adjustments for ice unless measurements are made more frequently than normally required for historical record.

The simulated real-time record for ice-affected periods was recomputed without applying any corrections for ice. Paradoxically, the duration of errors for this uncorrected record, as shown below for the 10-station group, is generally less than that for the corrected record. However, the number of days in error by more than 50 percent increased.

Percent error	≥5	≥10	≥20	≥30	≥50	≥100
Number of days	532	469	393	342	278	162

The above data are also shown in figure 3, but the number of days plotted is the average number of days in error per year. Here it is apparent that less error will occur if flat ice corrections are not applied during intervals between normally spaced measurements.

Figure 3 also shows curves representing errors caused by ice for a 40-station group, the entire 109-station group, and one station extremely affected by ice. The 40-station group includes all stations affected by ice during the 3-year study period. The 40-station curve was approximated by adjusting the frequency (ordinate) of the 10-station curve by a ratio of the average number of ice-affected days per station year in the 40-station group versus that for the 10-station group. The 109-station group includes 69 stations not affected by ice. The 109-station curve is included in figure 3 to show the overall effect of ice and to allow comparison with other types of errors shown in figures 1 and 5.
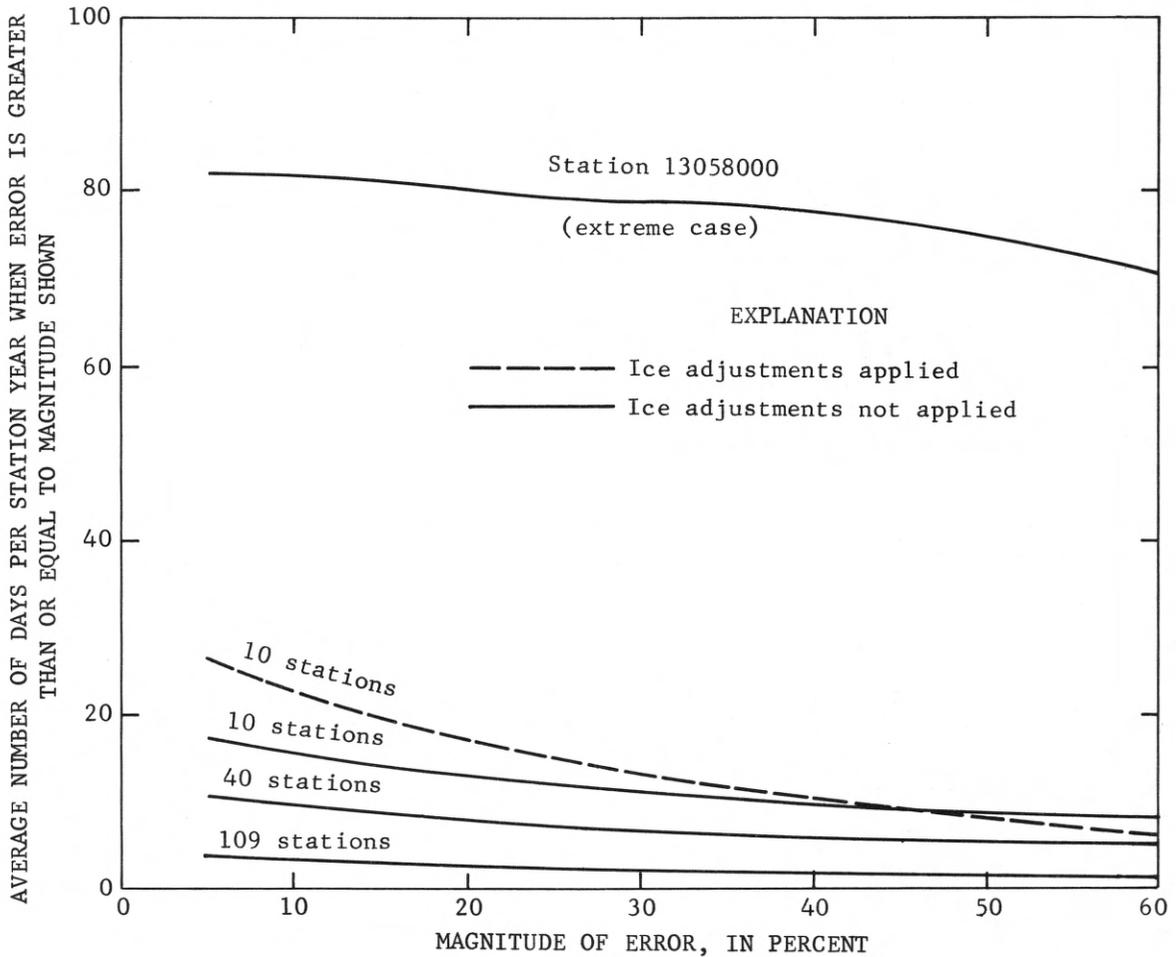
Figure 3.--Potential real-time data errors in mean daily discharges
caused by backwater from ice.

The seasonal distribution of errors caused by ice is shown in
figure 4 for the 10-station group. Ice effect, as expected, occurs
mostly during December through February.

Errors Caused by Malfunctions of Stage-Sensing Equipment

The hydromet system will be instrumented with one or both of two
general types of data-transmission devices, as follows:

1. The transmitter connected directly to a stage-sensing mechanism,
    through an encoder; that is, the transmitter is linked directly
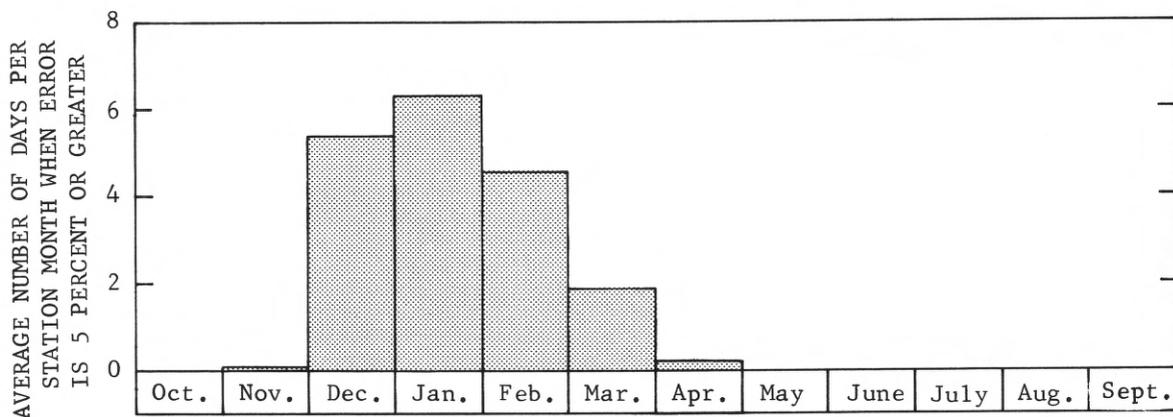   to the float wheel which activates the recorder.

14

Figure 4.--Average monthly distribution of errors in real-time mean
daily discharges caused by backwater from ice for 14 stations during
water years 1964, 1965, and 1967.

2. The transmitter connected to the on-site recorder, such as the
    "telekit" attachment for a Fischer & Porter digital-punch
    recorder.

     The first of these depends only on the reliability of the sensing
mechanism, whereas the second is affected by malfunctions of the re-
corder as well as by the sensing mechanism that activates the recorder.
Therefore, the errors caused by stage-sensing malfunctions were cate-
gorized separately, as follows:

1. Errors caused by malfunctions of the stage-sensing apparatus (that
    is, float assembly, stilling well intakes, or bubble manometer
    apparatus).

2. Errors caused by malfunctions of the on-site recorder plus the
    stage-sensing apparatus.

     Of the 14 detailed-study stations, eight stations had equipment
malfunctions during the 3-year period. The magnitude and duration of
errors caused by equipment malfunctions were determined for the 14-
station study group, as follows:

Percent error	≥ 5	≥ 10	≥ 20	≥ 30	≥ 50	≥ 100
Number of days of sensor errors	30	27	21	12	0	0
Number of days of recorder and sensor errors	162	144	103	86	45	2

15

Figure 5 shows these data, but the number of days plotted is the average number of days in error per year.
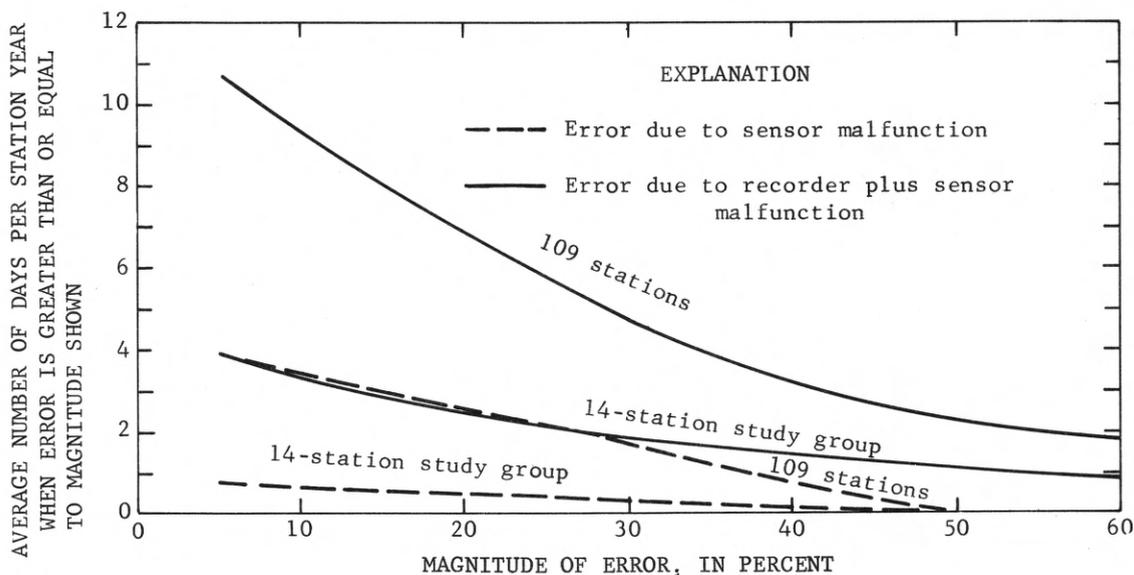

Figure 5.--Potential real-time data errors in mean daily discharges caused by sensor and recorder malfunctions.

Curves for the entire 109-station group were approximated by adjusting the frequency (ordinate of the 14-station curve) by a ratio of the average number of days of malfunction per station year in the 109-station group versus the 14-station group.

Although the frequency of error for both categories of malfunction is small, it is apparent that a sensor-dependent transmitting device is subject to fewer errors than is a transmitting device dependent on both a sensor and a recorder.

The distribution of errors caused by equipment malfunctions was found to be generally random except for December and January, which showed increased frequency of recorder malfunctions. Other studies have shown that malfunctions occur most frequently during cold, moist weather.

Summary of the Analysis

Probable errors in real-time data were approximated for 109 hydromet stations for a period of 3 years. The concepts used in computing simulated real-time data are considered to be logical and represent

16

the likely mode of processing in the operation of a real-time data-collection system.

The magnitude and duration of all errors that were accounted for in the entire 109-station group for the 3-year period are shown in figure 6. The curve is based on the total number of occurrences of the three types of errors analyzed in this study, including errors caused by rating changes, sensor and recorder malfunctions, and ice (without applying flat corrections).
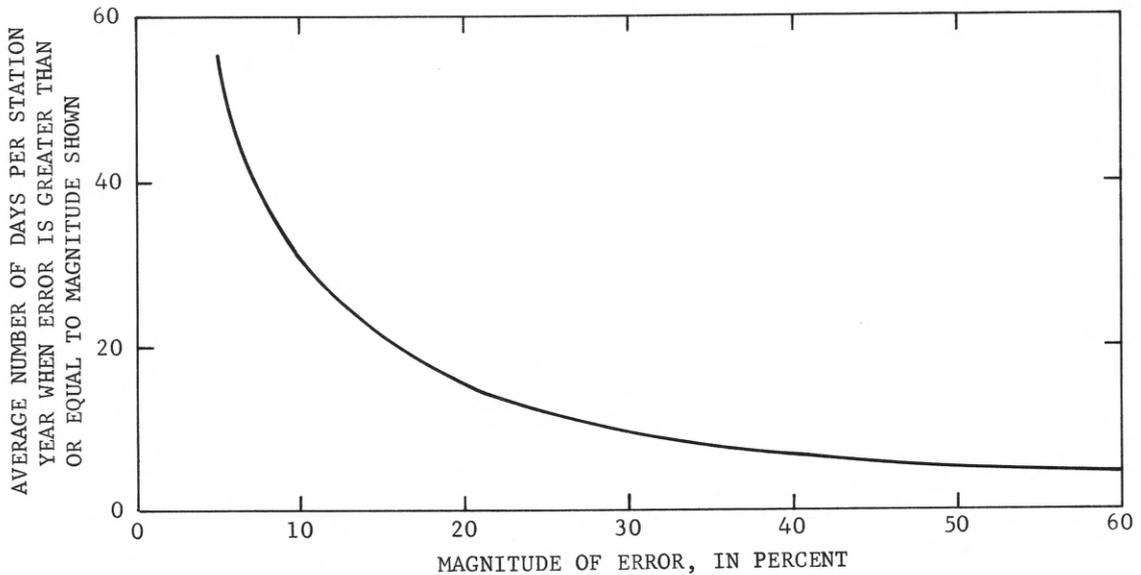
Figure 6.--Magnitude and frequency of all errors in real-time mean daily discharges for 109 hydromet stations.

Although this evaluation is based on data errors resulting from a simulated version of real-time data processing, it provides an estimate of the inaccuracies to be expected in a real-time data-collection system. It should be noted that this analysis does not include potential errors resulting from malfunctions of data-transmission devices.

The duration of inaccuracies can be reduced by timely detection, diagnosis, and correction. To accomplish this, verification must be conducted to some extent as part of the real-time processing. Concepts of verification by digital computer are discussed in part II.

PART II

TECHNIQUES FOR VERIFYING STREAMFLOW DATA IN A REAL-TIME SYSTEM

Purpose and Scope

Part II of this study defines various streamflow-data-verification concepts and outlines the computer routines required for implementation in a real-time data-collection system.

Because stations presently selected for the hydromet system are in a proposed status, subject to change, only the general scope of application of the verification concepts is presented.

The following sections present data-verification concepts in approximate order of complexity, starting with the more basic techniques. The discussions of computer applications of each of these techniques describe the overall function and the data-storage requirements.

Statistical Verification Based on Historical Records

Historical records of streamflow data reveal seasonal trends that characterize the hydrologic and climatic properties of each drainage basin. These trends, defined by the historical range in discharge for each season, month, or day provide a statistical basis for gross verification of future real-time streamflow data. Historical data can also be used to define correlation models between two or more streamflow stations that have similar flow characteristics and seasonal trends. These correlations provide a means to verify the flow at one station based on the flow at adjacent stations.

Most streamflow stations in the hydromet system have historical records of daily discharges for 10 to 50 years or more available on a magnetic-tape backfile. Because the historical backfile contains daily records, statistical verification methods are essentially limited in application to daily flows. Verification of hourly flows may be feasible for larger streams where hourly flows differ only slightly from daily flows.

To accomplish the daily or hourly verification of a large volume of input data to a real-time system, a computerized process must be used almost exclusive of manual intervention.

The following sections describe the various statistical techniques that can be formulated using historical data and suitable computer routines that apply these techniques to daily verification of real-time data.

Flagging Extreme Daily Discharges

Large errors in real-time data can be detected by a computer routine that flags daily discharges falling outside the historical range of flows for a given station. Such a computer routine would be simple in design, and its basic function would be to compare each real-time daily discharge to a predetermined historical maximum and minimum flow for the given day of the year. To determine if a flagged discharge is actually in error, it would be necessary for a hydrologist (or possibly a computer routine) to determine if current climatic occurrences could account for such an extreme streamflow event. When a discharge is found in error, a correct discharge can be estimated by one of the more complex routines described in subsequent sections of this report.

Historical daily maximums and minimums can be efficiently extracted, by computer, from the magnetic-tape backfile of daily discharges. Figure 7 shows two sample plots of the historical extremes extracted for each day during 1930-68 and 1929-68. Each plot reflects seasonal trends that are characteristic of that stream basin.

Three basic methods for storing extremes on a magnetic tape or diskfile in the real-time computer system are:

1. Store extremes for each day of the year, requiring 730 words (2 x 365) of data storage per station.

2. Use of only the highest and lowest points in successive 5-, 10-, or 20-day periods. This may be more representative of a long period of record because the daily extremes fluctuate considerably from day to day, as shown in figure 7.

3. Fit the extracted daily maximums and minimums to two polynomial regression equations. These equations would be of the form:

$$\log_{10} Q = c + c_1 D + c_2 D^2 + \ldots c_n D^n$$

where Q is an extreme daily discharge,

   D is the day of water year (1-365),

   n is the order number of the equation, and

   c's are regression constants.

Only the constants would be stored, and the number of storage words required would depend on the number of orders of n needed to fit the equation to the data. Possibly, n would range from 3 to 12, requiring 6 to 24 words of storage. Examples of polynomial curve fitting are shown in figure 7. The approximated maximum and minimum limits would be computed for any day of the year, but the limits would tend to be the average of the day-to-day fluctuations.
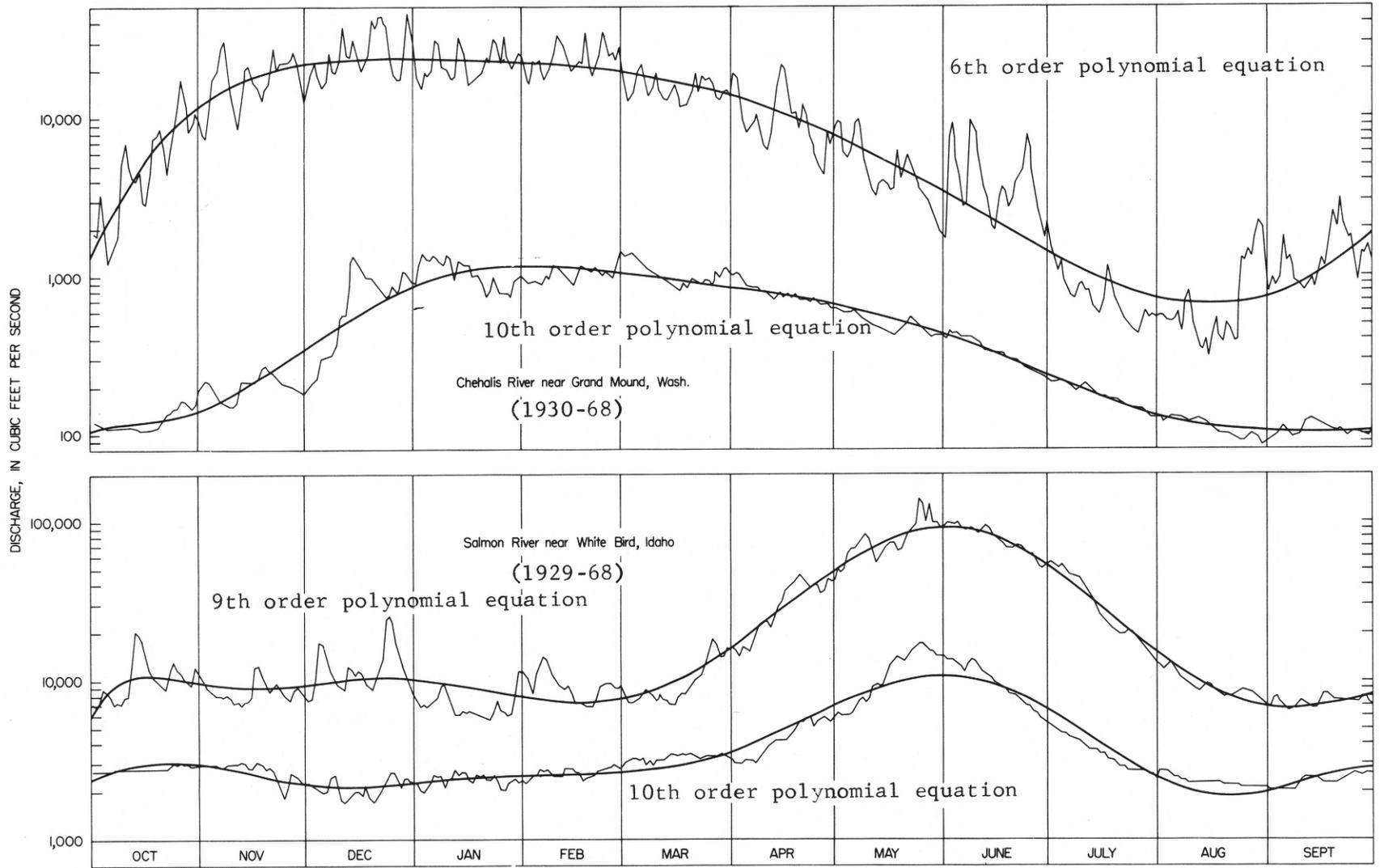
Figure 7.--Historical maximum and minimum daily discharges.

DISCHARGE, IN CUBIC FEET PER SECOND

6th order polynomial equation

10th order polynomial equation

Chehalis River near Grand Mound, Wash.
(1930-68)

Salmon River near White Bird, Idaho
(1929-68)

9th order polynomial equation

10th order polynomial equation

Axis values (top panel): 10,000 — 1,000 — 100

Axis values (bottom panel): 100,000 — 10,000 — 1,000

Months: OCT NOV DEC JAN FEB MAR APR MAY JUNE JULY AUG SEPT

Of the three methods of storing extremes, the method using the highest and lowest historical discharge in successive periods of about 10 days would provide adequate representation with a storage requirement of about 74 words per station. It is estimated that approximately 1 to 2 man-months would be required to program or modify an available program and implement this routine for the hydromet computer system.

Flagging Improbable Changes or No-Changes in Sequential Daily Discharges

In addition to flagging extreme magnitudes of discharge, a routine that flags abnormally abrupt changes in sequential daily discharges could point out the occurrences of typical errors caused by backwater from ice or obstructions, and errors caused by equipment malfunctions.

The magnitudes of changes between consecutive daily discharges are dependent on basin characteristics and climatic variation. Consequently, the maximum changes in the historical record will define a seasonal pattern that is typical for each basin. Figure 8 shows maximum and minimum daily changes (in percent of discharge) that have occurred at Chehalis River near Grand Mound, Wash., and Salmon River near White Bird, Idaho, during 1930-68 and 1929-68, respectively. Although the graphs for the maximum change are jagged day to day, seasonal trends can be recognized. It is assumed that a much longer period (say 500 years or more) of record would produce a smoother graph that would follow closely the highest points on the sample graphs. Thus, a realistic value of the maximum change between daily discharges can be selected for each month. See dashed lines in figure 8. Any change exceeding this value would indicate that one of the daily discharges may be in error.

A computer-verification routine that flags improbable daily changes would be simple in design. The percentage change in real-time discharge (current day versus the previous day) would be compared to the maximum daily change previously recorded for the given month. Only discharges showing a change greater than the historical change for the given month would be flagged. The data user might conclude that a flagged figure was in error unless an extreme climatic condition had occurred. This verification routine would require only 12 words of data storage per streamflow station.

Because floods characteristically have a rise that is more rapid than the recession, consideration should be given to differentiation between plus and minus maximum daily changes. This would require 12 additional words of data storage per station.

Errors caused by sensor malfunctions may show a no-change condition more often than an abrupt change. The plots of minimum daily changes shown in figure 8 indicate that no change occurred on most days during the 40 years of historical record; therefore, a probable error can be
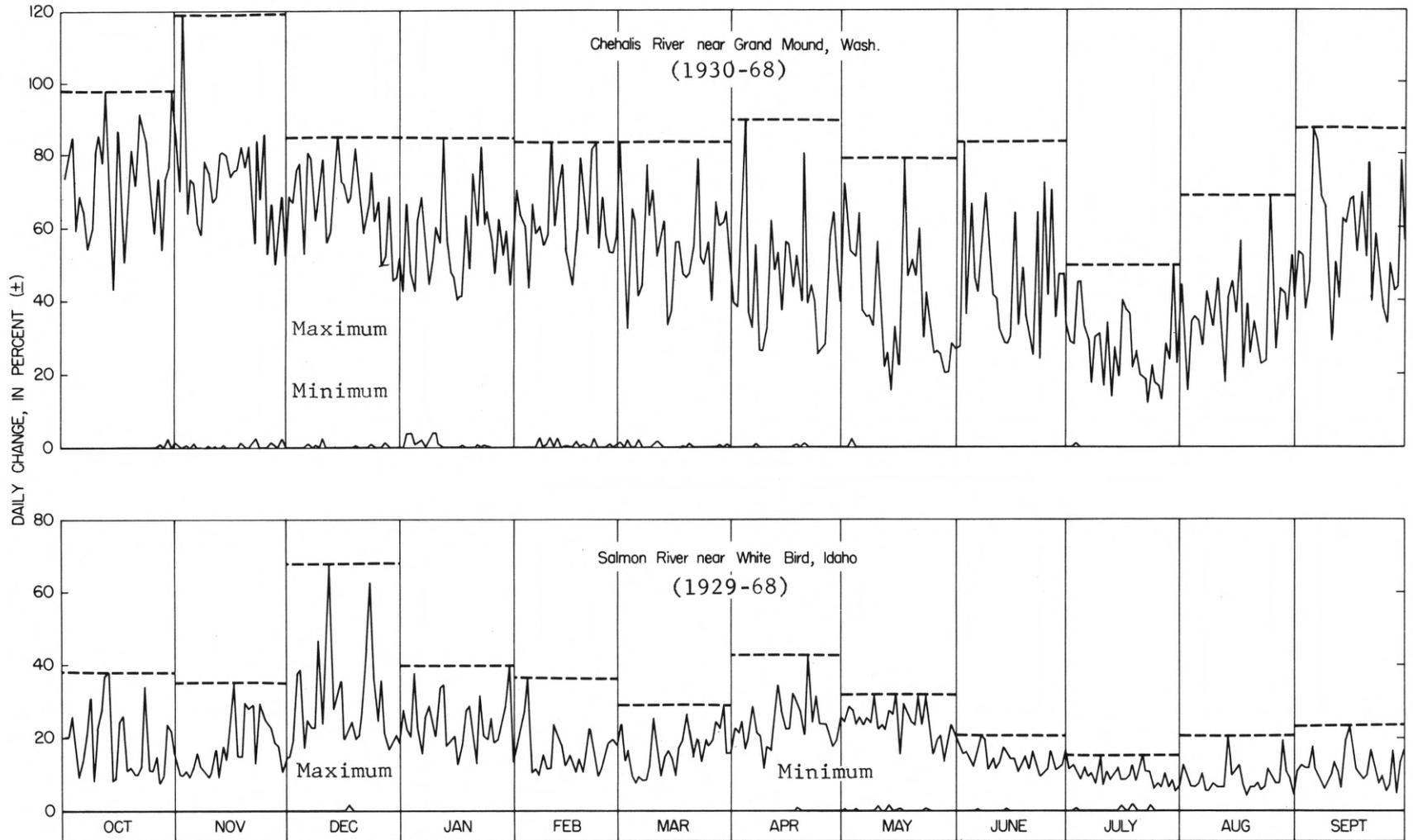
Figure 8.--Historical maximum and minimum changes in sequential daily discharges.

meaningfully detected only when a no-change condition persists for several successive days. To be most useful, the routine should be designed to footnote a no-change condition and also to indicate the number of successive days of occurrence. A reference table of the longest period of no change could be compiled from historical records for each station. This information could be used as a flagging limit in the computer routine or as a visual reference for the data user.

It is estimated that approximately 1 to 2 man-months would be required to program and implement this routine in the real-time computer system.

Correlation of Independent Observations

Correlation of concurrent daily discharges.--Experience has shown that natural-flow streams with similar hydrologic and climatic characteristics yield daily flows that can be expected to correlate well. It has become standard practice to verify historical records by superimposing and comparing daily discharge hydrographs of two or more stations that are known to correlate. This same method can be used to verify real-time data; but because graphic procedures are not readily adapted to computers, it is necessary to design a digital technique that can be adapted to a real-time computer-processing system.

Logarithms of daily flows for hydrologically similar streams generally correlate as a straight-line function. Thus, a relation can be developed using a simple linear regression model to define an equation of the form:

$$\log Q_1 = b' + m \log Q_2$$

By taking the antilog, the equation becomes

$$Q_1 = bQ_2^m$$

where b equals antilog b',

Q_1 and Q_2 are the dependent and independent daily flows, and

b and m are regression constants.

Figures 9 and 10 show examples of logarithmic correlations of daily flows for Clearwater River near Spalding, Idaho, versus Lochsa River near Lowell, Idaho. These correlations represent daily flows that occurred on October 15 and December 15 during 1930-68. Regression equations for the two correlations were computed by the method of least squares, and the curves are shown in the figures. The standard errors of estimate were computed (14.1 percent and 21.9 percent) and are shown in figures 9 and 10 as dashed lines. The deviations in a logarithmic model generally fall in a band paralleling the curve of best fit. Thus,
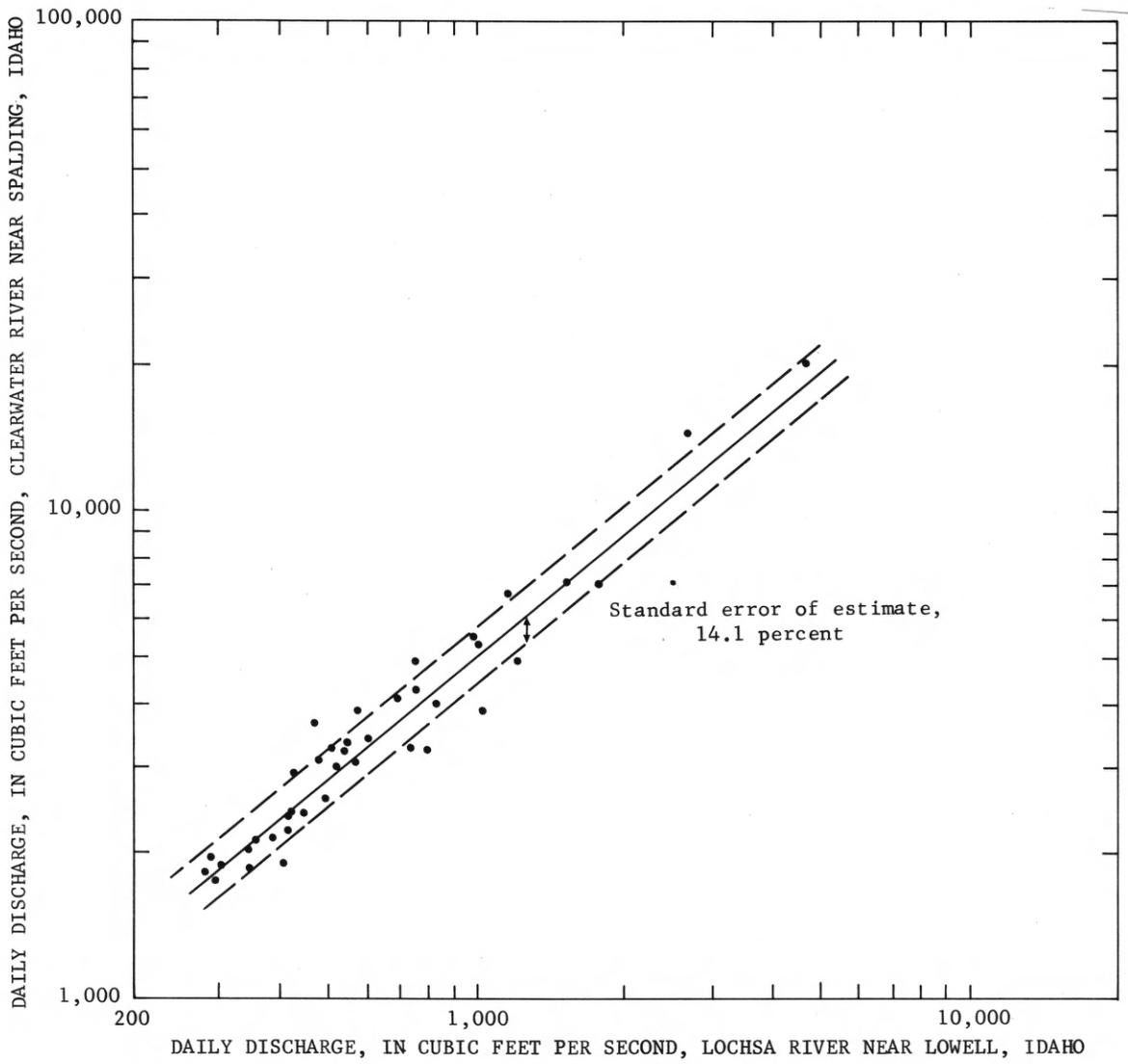
Figure 9.--Correlation of October 15 daily discharges, 1930-68.
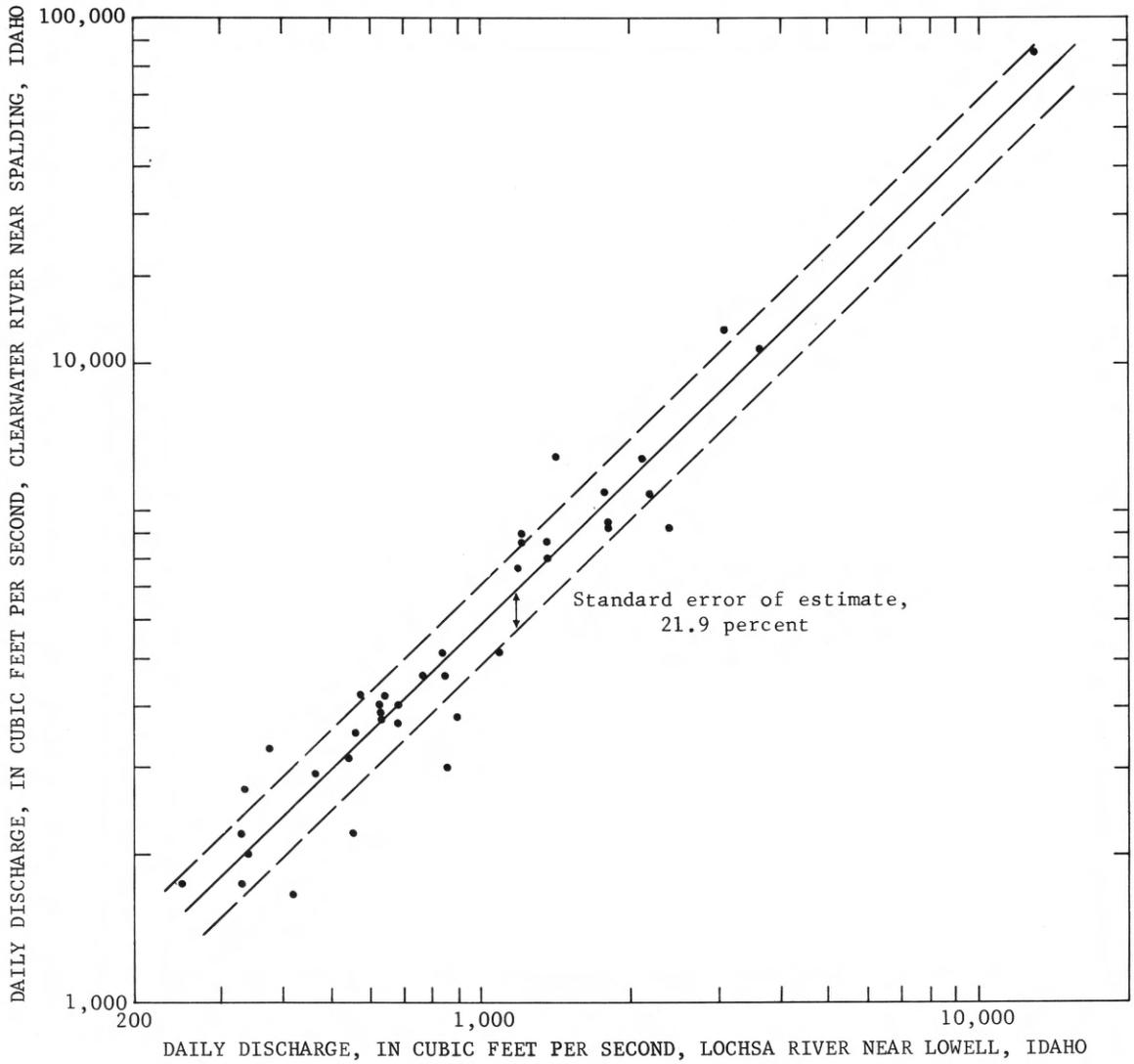
Figure 10.--Correlation of December 15 daily discharges, 1930-68.

the standard error of estimate provides an indication of reliability that is constant throughout the range of the model.

To correlate daily discharges throughout the year, it is necessary to compute regression equations for each day of the year for each pair of streamflow stations that correlate. The resulting regression constants, b and m, and the standard errors, SE, would be stored on magnetic tape or disk for access by the real-time computer-processing system. Figure 11 shows the daily regression constants and the standard errors for daily correlations of Clearwater River near Spalding versus Lochsa River near Lowell.

The basic function of a computer routine to verify real-time data by correlation would be to retrieve the appropriate constants, compute a simulated discharge, and compare the simulated discharge to the real-time discharge. If the observed discharge was not within one or two standard errors, SE, of the simulated discharge, it could be assumed that one or the other of the observed discharges may be in error. A correlation with a third station would be required to determine which of the two observations is in error. It is advantageous, then, to correlate any streamflow station with at least two others.

A specific advantage of this verification technique is that a simulated discharge can be substituted where an error is ascertained or where a data value is missing. The standard error of estimate can also be provided to indicate the general accuracy of the simulated discharge.

Provided correlations can be established with existing index stations within hydromet, the regression-model technique can be used to compute simulated discharges for streamflow stations that are not reported in the hydromet system. Regional correlation equations can also be developed to estimate flow for ungaged streams. The ungaged flow would be determined as a function of flow at an index station and of drainage area of the ungaged basin. This concept is mentioned here only for completeness of scope and was not studied under this report.

The data-storage requirements for this technique are 1,095 (3 x 365) words per pair of stations. This includes the two regression constants and the standard errors for each day of the year. For each group of three stations that cross-correlate, 3,285 words of storage are required, for each group of four stations, 6,570 words, and so on. This storage is best facilitated by an off-line storage medium such as magnetic tape. Only the constants for the day being processed need to be retrieved and placed in central memory, thereby requiring only three words of central-memory storage per pair of stations.
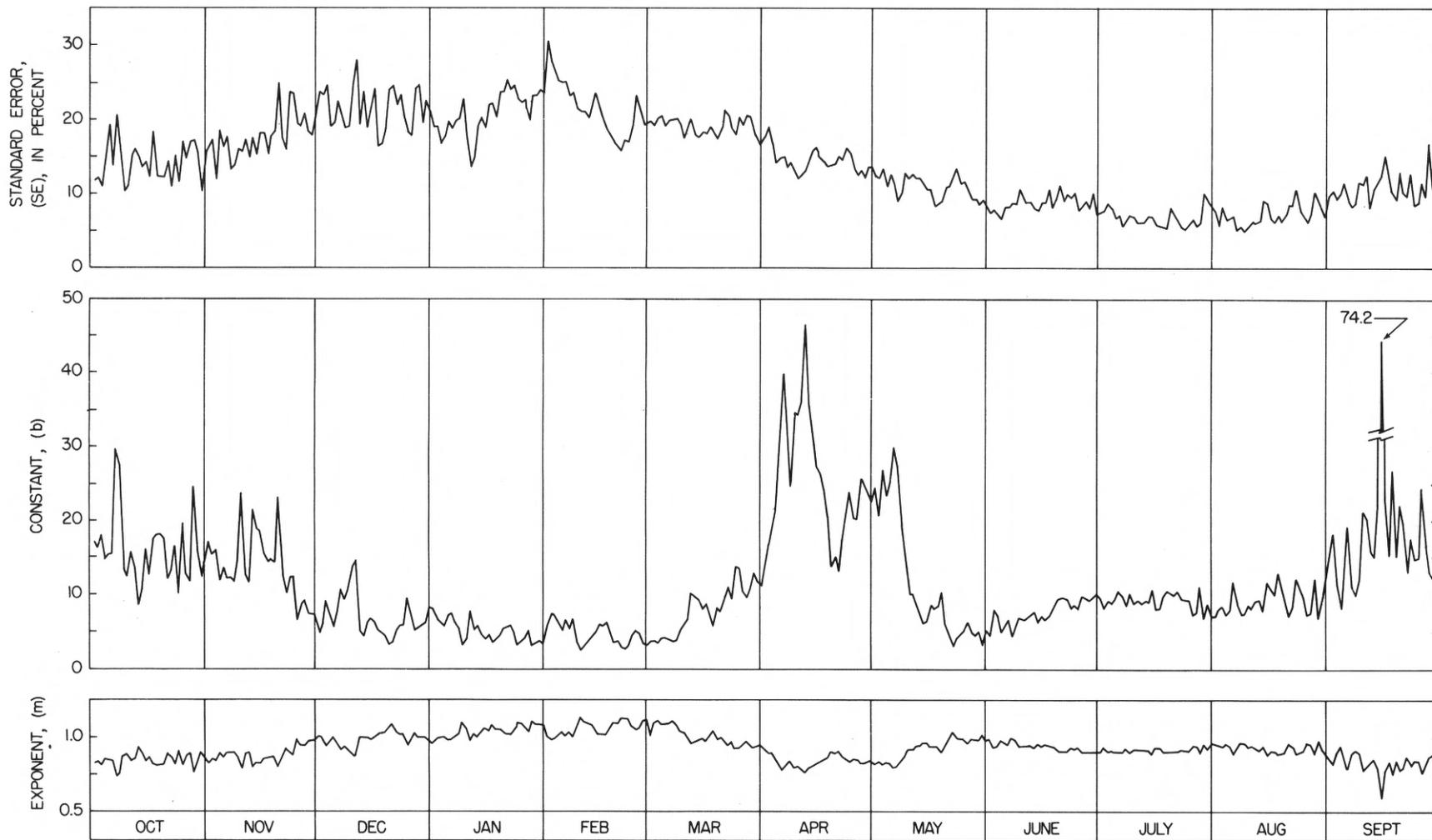
Figure 11.--Standard errors and constants for daily regression models of the form $Q_1 = bQ_2^m$, Clearwater River near Spalding, Idaho Q_1), versus Lochsa River near Lowell, Idaho (Q_2), 1930-68.

An attempt was made to reduce the required amount of off-line data storage by fitting each of the regression constants to a polynomial regression equation of the form:

$$C = a + a_1D + a_2D^2 + \ldots a_nD^n$$

where C is the constant (b or m),

D is the day of the water year (1-365),

a's are regression constants, and

n is the order of the polynomial used.

Only the a's and n would be stored, and the constant, C, could be approximated for any given day of the year. Polynomial equations provide an average fit through the daily fluctuations of b and m. By using these polynomial equations to recompute historical streamflow, it was found that the resulting standard errors were two to three times greater than the standard errors of the actual daily correlation models. This is a consequence of the "averaging" effect of polynomial curve fitting. Perhaps a Fourier series equation would be more successful. It was concluded that simulation of the regression constants by polynomial equation is not precise enough and that other methods of reducing off-line storage requirements may be explored at a later time, if needed.

Regression models can also be developed for daily discharges on a weekly or monthly basis rather than on a daily basis. Limited experience has shown that this approach may be adequate. Weekly or monthly regressions would be advantageous for stations with short historical records and would greatly reduce the requirement for off-line storage of regression constants.

In situations where one of two streams responds more quickly to climatic changes than does the other, consideration should be given to correlating the two streams on a nonconcurrent basis. Because a significant time lag may be caused by differences in the size, length, or geology of the drainage basins, it may be necessary to develop regression models incorporating a time lag of a day or more. On a real-time basis, this technique is limited to verifying only the lagging stream based on preceding flow at the leading stream.

An experimental program that retrieves historical discharge data from the existing tape backfile and computes daily or monthly regression models for pairs of stations has been developed for this study. This program is available for use in the hydromet system but requires some modification to provide the required output.

It is estimated that approximately 3 to 4 man-months would be required to program and implement a verification routine for correlation of daily discharges in the real-time computer system.

Correlation of exceedence probabilities.--An alternate method of correlating concurrent-flow events at two stations is to compare the exceedence probabilities of daily discharges. The exceedence probability of a real-time daily discharge is computed from the frequency distribution of historical flows for that day. It has been found that the exceedence probabilities of concurrent daily flows for two hydrologically similar stations correlate on a one-to-one basis; that is, the probabilities are approximately equal.

Figures 12 and 13 are typical examples of correlations of exceedence probabilities of historical daily flows for Clearwater River near Spalding versus Lochsa River near Lowell that occurred on October 15 and December 15 during 1930-68. The exceedence probabilities shown in these figures were computed by the plotting formula:

$$P = \frac{M}{N+1} \times 100 \text{ percent}$$

where P is the exceedence probability,

N is the number of years of record, and

M is the order of magnitude.

The dashed lines in the figures represent the standard error of estimate of the plotted positions about the line of equal probabilities.

To efficiently accomplish the correlation of exceedence probabilities on a daily basis in a real-time system, it is necessary to develop frequency-distribution equations, such as Log Pearson Type III, for each day of the year for all stations. The three constants for each equation would be stored on magnetic tape, requiring 1,095 words (3 x 365) of storage per station. On a daily basis, the real-time verification routine would retrieve the Log Pearson constants for that day and compute the exceedence probability of the reported discharge for each station. The routine would then compare probabilities for all pairs of stations that have a known correlation. Any station that does not compare within one or two standard errors with two or more stations would be flagged to alert the user of a probable error. The standard errors, used as a flagging reference, would have to be precomputed for each pair of stations (as shown in figs. 12 and 13) and stored on magnetic tape. This would require an additional 365 words of data storage per pair of stations.

The real-time computer routine could be designed to print the exceedence probability with the associated discharge. This would provide a means for additional manual verification and would alert the user
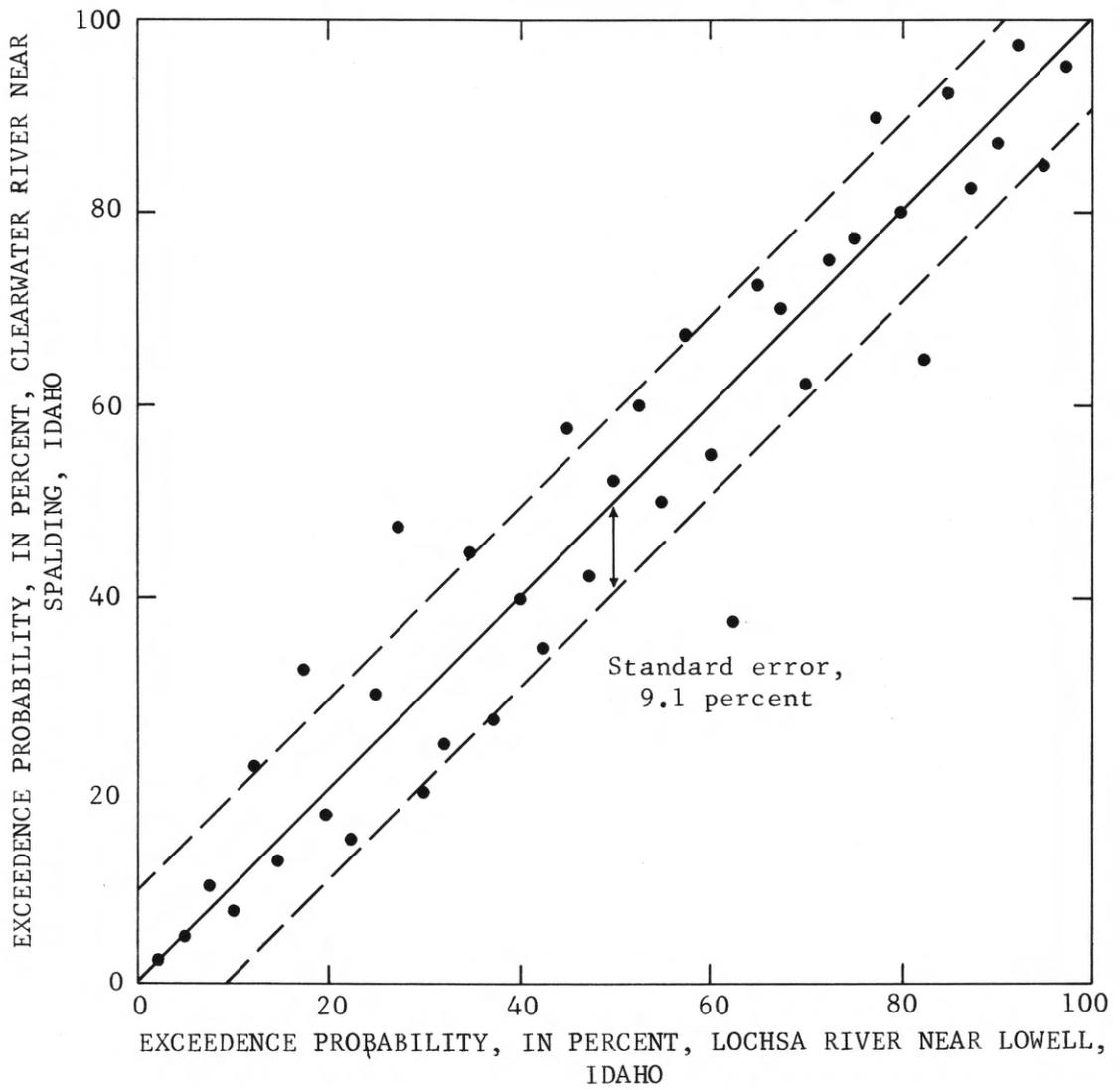
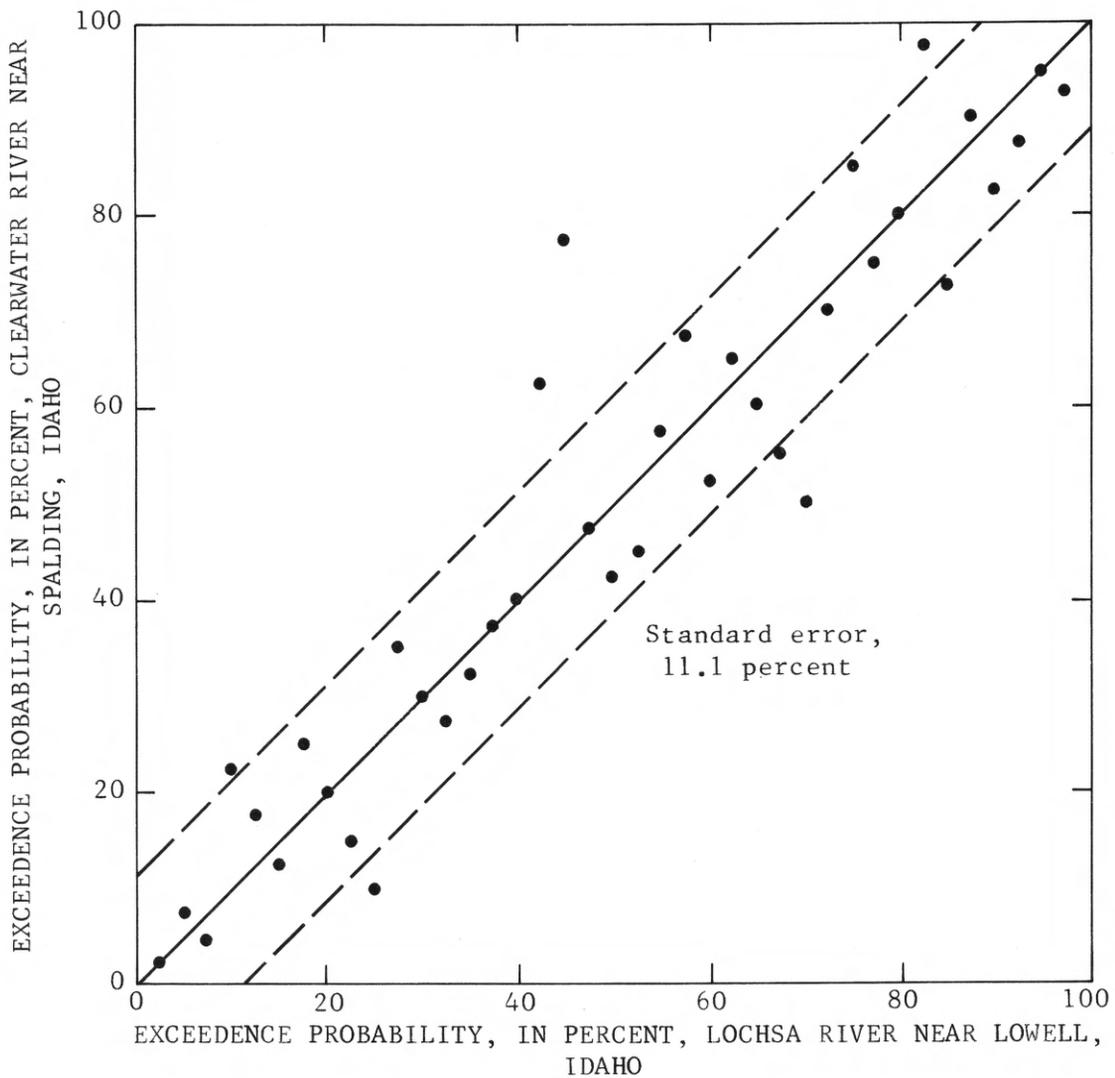Figure 12.--Correlation of exceedence probabilities of historical October 15 daily flows, 1930-68.

Figure 13.--Correlation of exceedence probabilities of historical December 15 daily flows, 1930-68.

when extreme events have occurred. The disadvantage of using exceedence probabilities rather than discharges for correlation is that a simulated discharge would not be readily available when data are missing or when obvious errors occur. A discharge could be computed from the Log Pearson equation using the exceedence probability from a station that is known to correlate; however, the accuracy of a simulated discharge determined in this manner could not be evaluated easily.

It is estimated that approximately 3 to 4 man-months would be required to program this concept and implement it in the real-time computer system.

Verification by Deterministic Techniques

Deterministic concepts, based on physical laws of hydrology, can be used to verify natural- and regulated-flow data. Three such techniques are outlined in the following sections.

Basic Continuity Test

Testing the continuity of flow at successive points in a river system is a traditional method of verifying historical as well as current streamflow data used for water management. The basic application can be represented by a general equation of continuity for a given river reach, as:

$$Q_o = Q_i + Q_t \overset{\Delta}{\pm} s - Q_d$$

where Q_o = outflow from reach,

Q_i = main-stem inflow (next upstream station),

Q_t = tributary inflow,

Δs = change in reservoir storage, and

Q_d = diverted flow.

Because this technique does not include all flow-routing parameters, such as channel storage and flow lag time, the technique is essentially limited to short river reaches under generally steady flow conditions.

A computer routine could be designed to retrieve the inflow and storage data required to solve the above equation. The outflow could then be computed and compared to the reported outflow. If the comparison were not successful within a predetermined limit, the outflow could be flagged to indicate a discontinuity and the computed figure printed to show the comparison. The predetermined comparison limit (expressed in percentage of flow) is based on the estimated accuracy of the inflow and storage data.

32

If the continuity test indicates the presence of a data error (by flagging the outflow figure), it does not necessarily mean that the outflow figure is incorrect, because any one of the inflow or storage values could be the actual source of error. However, if an inflow figure is in error, it may have been flagged by one of the other verification routines, such as one of the statistical verification routines. Thus, determining the source of error is in part a manual procedure. The basic purpose of this routine is to point out situations of obvious discontinuity.

When tributary inflow data are not available in the hydromet system, it may be necessary to develop a correlation model of the tributary inflow versus a reported hydromet index station. This procedure is described in the section entitled "Correlation of concurrent daily discharges."

A computer routine to test continuity would be simple in design and would require little data storage, because most of the data is retrieved directly from the real-time discharge data file. The configuration of inflow, outflow, and storage would be identified for each continuity test and placed on a magnetic tape or disk file. The data-storage requirement depends on the number of stations involved in each continuity test and on the number of correlation models required to simulate the tributary inflow.

It is estimated that approximately 1 to 2 man-months would be required to program this routine and implement it in the real-time computer system.

Flow Routing

Regulated flow data on the main stem and large tributaries of the Columbia River system can be verified by using streamflow-routing techniques. The development of a streamflow-routing model with specific application for verification of real-time data would be a major task. Therefore, the scope of this report is limited to (1) discussion pointing out the application of the technique for real-time data verification and (2) outlining the effort required to establish a digital flow-routing model.

The basic concept of flow routing is to compute simulated hydrographs (discharge versus time) at successive downstream points starting at a point with a known hydrograph (that is, a gaging station). Application of flow routing as a method of verification is accomplished by routing a series of real-time discharges between successive stations reported in the hydromet system in downstream sequence. Errors are detected by comparing the routed discharges to the reported values and flagging differences that exceed the estimated accuracy of the routing technique. It is evident that a flagged difference might indicate an error at either the upstream or downstream station, in the tributary inflow, or in reservoir regulation data. Because routing is reinitiated

for each independent reach based on reported discharges at the upstream station, each intermediate main-stem station is subject to a subsequent test when its reported flows are routed to the next downstream station. Thus, if an error is detected that was not also evident in the upstream or downstream reaches, the error could then be attributed to the tributary inflow or reservoir data used in the reach showing the error. In this case, the data user must subjectively determine the error source. In the event of missing or erroneous data for a main-stem station, this routing procedure can provide a simulated discharge value.

The theoretical mathematics of streamflow routing are quite complex, and simplifying assumptions are required to approximate a numerical approach. Several concepts of flow routing that may be practicably adapted to digital computers are described by Chow (1964) and by others. The computations involved are generally based on the physical characteristics of the stream channel and intervening reservoirs, the regulation imposed on the reservoirs, and tributary and local inflow. Most routing procedures can be adapted to time intervals ranging from a few minutes to 24 hours. It is necessary to determine local and tributary inflow based on estimates or correlations with adjacent stream stations.

The development of a computerized flow-routing model for the Columbia River system involves two major tasks: (1) the development of a computer program capable of performing the necessary computations and data handling, and (2) the determination of the physical characteristics of the river system and the incorporation of them into the computer program in digital form.

Because of the effort required to develop a flow-routing model, it is recommended that previously established routing models be utilized to the fullest extent possible for adaptation to verification of real-time data. One such comprehensive computer program capable of channel and reservoir routing, entitled "Hydro power system regulation analysis," has been developed and adapted to the Columbia River system by the U.S. Army, North Pacific Division, Corps of Engineers. The unpublished documentation of this program and the concepts by C. E. Abraham of the Corps of Engineers is assumed to be available and can be adapted for verification of data in the hydromet system. Personal communications with Mr. Abraham indicate that the adaptations required would involve primarily data retrieval and additional logic to accomplish the comparison of routed data versus reported data. The computer requirement is approximately 100,000 words of central memory. Indications are that computer processing time will be approximately 2 minutes to route hourly flows for a 1-week period for the main-stem Columbia River system, based on estimates on an IBM 360-50 computer.

It should be noted that the effort required to adapt an established flow-routing computer program for verification of real-time data is small compared to developing an entirely new flow-routing program. Depending on the degree of sophistication, 2-6 man-months would be

required for additional programming to adapt an existing flow-routing program.

Watershed Modeling

Real-time streamflow data can be verified on a broad spectrum by a system-model approach. Development of a mathematical watershed model of the entire Columbia River system would be a monumental task. It could possibly, however, be justified as a data-verification tool if an established model were adapted for this purpose. This section discusses the applicability of watershed modeling as a method of verification and outlines adaptations required to utilize an existing digital model of the Columbia River system.

The overall purpose of a mathematical watershed model is to determine discharge hydrographs (discharge versus time) at various points within a river system based on knowledge of the physical characteristics of the watershed and the stream system, a known set of initial hydrologic conditions, and reported meteorological data. This procedure evaluates the entire hydrologic process of snowmelt and (or) rainfall runoff from the watershed. The program then routes the outflow from the watershed to any number of downstream points in the stream system.

A digital watershed model can be utilized for verification of real-time streamflow data by comparing a series of simulated discharge values to the corresponding real-time values for each station in the hydromet system. Differences between these values do not necessarily indicate an error in the reported discharge. The discrepancy could also be the result of in orrect initial conditions, erroneous meteorological data, or overall inaccuracy of the model. It is evident that an experienced hydrologist must subjectively determine the actual source of error. It is conceivable, however, that a sophisticated computer program could be developed to indicate the most probable source of the error.

A digital watershed model is presently being used to forecast flows in the Columbia River system by the Portland River Forecast Center, operated by the National Weather Service, National Oceanic & Atmospheric Administration; the U.S. Army, North Pacific Division, Corps of Engineers; and Bonneville Power Administration. The basic model, together with its associated program, was originally designed by Rockwood (1958) and is now in its third generation of refinements, as described by Anderson (1971). The present computer program, entitled "Streamflow simulation and reservoir regulation" (SSARR), is written in Fortran IV and designed basically for an IBM 360-50 computer; however, a CDC 6400 version is also available. The SSARR program evaluates basin runoff, routes open-channel flow, and accounts for reservoir regulation. The incremental time periods for flow routing and discharge simulation may be varied between 0.1 hour and 24.0 hours as specified by the user. The operational data for the SSARR are obtained via an existing data-reporting network, a precursor to the proposed hydromet data network.

As a standard procedure, the River Forecast Center initiates each forecast by starting with data 1 or more days previous to the current day, whereby flows are simulated for the current day as well as for future days. This procedure provides the essence of real-time data verification in that simulated data may be used to check the validity of reported data.

It appears that the SSARR watershed model could be used effectively to verify real-time streamflow data. Adaptation to data-verification function would require the following modifications: (1) inclusion of applicable hydromet stations not presently used in the model, (2) provision for suitable printout of discharge comparisons and for flagging discrepancies that exceed the estimated accuracy limitations of the model, and (3) incorporation of logic that is functional in determining the most probable source of error.

The SSARR computer program, in its present configuration, requires about 200,000 words of central memory on an IBM 360-50 computer, with at least one random access disk drive. With this computer capability, the Columbia River system is modeled in four separate subbasins: (1) Upper Columbia River, (2) Snake River, (3) Lower Columbia River, and (4) Willamette River. The four computer runs require a total of about half a minute of central processing time per routing interval; that is, using a 6-hour time increment, about 2 minutes of computer-processing time is required to simulate a day's discharges throughout the Columbia River system. The estimates apply only to the above-described computer configuration and are included here for general planning purposes only.

Depending on the degree of sophistication required, 2 to 8 man-months would be required to adapt the SSARR program to be used specifically for streamflow-data verification.

In addition to verification of streamflow data, the watershed model can provide simulated stream discharges in lieu of missing or erroneous data. The SSARR watershed model can also provide verification of meteorological data reported by the hydromet network.

Recommended Field and Office Procedures

Frequency of field visits and the techniques for analyzing stage-discharge relations have been standardized for the collection and processing of historical streamflow records. Although the same basic methods will apply in a real-time system, it may be necessary to increase the frequency of field visits and expedite the rating analyses for efficient processing of real-time streamflow data.

As shown in part I of this study, there are three potential sources of error (not including data-transmission failure) inherent in real-time streamflow data. For the purpose of part I of this study, only the results of routine field measurements and observations were used to

detect and correct data errors. This procedure disclosed errors that can be expected if no analytical verification nor extra field visits to gaging stations are made. Part II of this study outlines various analytical verification techniques that can be employed to detect real-time data errors. Recommended field and office procedures for detecting and correcting errors are discussed in the following sections.

Errors Caused by Changes in the Stage-Discharge Relation

Figure 1 shows that errors caused by rating changes are generally less than 15 to 20 percent in magnitude. Because it is difficult to detect these relatively small errors by the methods outlined herein, rating changes must generally be detected and defined by field observations and discharge measurements.

It has been found that a measurement frequency of 4 to 8 weeks is sufficient to establish accurate rating definition for the computation of historical records. However, for real-time operation, where corrections can be applied only forward in time, discharge measurements must be made at more frequent intervals if a reduction of the duration of real-time data errors caused by rating changes is desired.

It would seem uneconomical to increase the measurement frequency at all hydromet stations. It would be beneficial, however, to schedule extra measurements during seasons when rating changes are expected to occur and only at stations having a history of frequent rating changes. About one-third of the stations analyzed in table 1 had an average of two or more rating changes per year. Many of these stations are subject to seasonal rating changes, whereby it is possible to predict the time of year when changes are likely to occur. It may also be beneficial to make additional measurements at a station after a rating change is first detected, so that the new rating definition or adjustment for shifting conditions can be more accurately established in a short time. Because of the cost of making additional measurements, it may also be desirable to schedule measurements in accordance with seasonal criticality of data for water management.

Errors Caused by Backwater from Ice

Figure 3 indicates that the magnitude of errors caused by ice are frequently large. The larger errors may often be detected by one of the various verification routines described in this study. Smaller, yet significant, errors and the precise times that ice effect begins and ends may be difficult to detect by the foregoing verification techniques. A more sensitive verification routine may be developed utilizing air and water temperatures. Accumulated thermal energy, based on an accumulation of air temperatures for several consecutive days, would serve as an index to the potential for increase or decrease in ice effect. Water temperatures would provide an indication of the possibility of ice starting to form in early winter.

Because of the time that would be required for calibration and testing, a routine utilizing air and water temperatures was not detailed as part of this study. However, the concept is believed to warrant consideration if backwater from ice cannot be adequately detected by the other methods. Consideration should, therefore, be given to sensing air and water temperatures at streamflow stations where ice is a potential problem.

To validate the accuracy of simulated or estimated data during ice-affected periods, additional field observations and discharge measurements are necessary. Because the error caused by ice changes continually as ice builds and melts, it would be necessary to make a discharge measurement every day to provide a high degree of accuracy. The cost of making daily measurements at all ice-affected stations would, of course, be prohibitive. Here again, the frequency of measurements needs to be determined station by station according to the criticality of data needs. About one-fifth of the 109 stations analyzed in table 1 are seriously affected by ice.

Errors Caused by Equipment Malfunctions

Data errors caused by equipment malfunctions are generally large in magnitude and random in occurrence. Figure 5 indicates that equipment errors are less frequent than are other sources of error; however, it is expected that data-transmission failure will add significantly to errors attributable to the equipment malfunctions analyzed in part I of this study.

Many of the larger errors caused by malfunctions can be detected by verification routines. Although a simulated discharge can be substituted, it is important that equipment malfunctions be corrected soon after detection.

Coordination of Data Verification and Field Visits to Gaging Stations

Although the bulk of the data processing will be accomplished by computer, considerable manual intervention and analysis will be necessary to maintain a high level of quality control.

The stage-discharge relation (rating curve) must be analyzed soon after each discharge measurement becomes available. As rating changes occur, new rating tables or corrections to existing tables must be prepared for use by the streamflow data-processing system.

Constant surveillance of the results of data-verification routines must be maintained to properly coordinate field visits when malfunctions or errors occur. The exact source of error or malfunction will, of course, be difficult to ascertain at times. 'Field visits will be required for (1) measuring the discharge when rating changes, ice effect,

or backwater conditions are suspected, (2) servicing and repairing sensing apparatus, and (3) servicing and repairing transmission equipment.

For planning purposes, it is estimated that a field visit or a discharge measurement will cost $50 to $150 depending on the travel distance and the number of technicians required.

Application of Verification Techniques to Streamflow Stations in the Hydromet System

Seven computer-oriented techniques of streamflow verification have been outlined in preceding sections. The first four are based on laws of statistical probability and pertain primarily to natural-flow (un-regulated) streams. The other three are based on deterministic principles and can be applied to both regulated- and natural-flow data. Approximately half the proposed hydromet stations are significantly affected by regulation. The applicability of each of the seven techniques is summarized as follows:

1. Flagging extreme daily discharges.--Although this routine applies most conveniently to natural flow, it is possible to develop an estimated probable range for regulated flow based on the current operating-rule criteria for upstream reservoirs. Conceivably, then, this routine can be applied to most streamflow stations in the hydromet system.

2. Flagging improbable changes (or no changes) in sequential daily discharges.--This routine applies readily to natural-flow data. For regulated flow, the maximum probable daily changes can be estimated based on the physical capability of reservoir outlets and current operational constraints. The maximum regulated daily changes can also be extracted from historical records provided a long record of regulated flow is available and the operational constraints have remained constant. Therefore, this routine applies to most stations in the hydromet system.

   Verification of streamflow data by testing for a no-change condition is applicable to both regulated- and natural-flow data.

3. Correlation of concurrent daily discharges, and

4. Correlation of exceedence probabilities.--The correlation techniques described in this report apply primarily to relatively unregulated streams. It is possible to correlate two regulated stations on the same stream provided there is little regulation in the intervening reach and the tributary inflow is a minor part of the total flow. The accuracy of correlations of various pairs of stations has not yet been established except for the few examples analyzed in this study. Additional study is required to determine which

stations correlate and the associated accuracy of the corre-
lations. It will also be necessary to determine a minimum cor-
relation-accuracy level that will provide meaningful verification.

5. Basic continuity test.--The primary application of the continuity
   test is to verify outflow from reservoirs based on the inflow and
   change in reservoir storage. This test will apply to most reser-
   voirs in the proposed hydromet system. In addition, the conti-
   nuity test can also be applied to several main-stem river
   reaches. The continuity test is limited to short river reaches
   under relatively steady-flow conditions. Each application must
   be tested for accuracy before being implemented in the real-time
   computer system.

6. Flow routing, and

7. Watershed modeling.--These techniques are unlimited in scope and
   may be used to verify regulated- and natural-flow conditions.

   In addition to seven computer-verification routines, specific
field and office procedures are recommended to verify real-time stream-
flow data and to determine corrections when inaccuracies occur.

Conclusions

   It is concluded that the duration of real-time data errors can be
reduced by (1) employing computerized techniques to detect errors, (2)
increasing the frequency of discharge measurements at selected gaging
stations, (3) timely analysis of stage-discharge relations, and (4)
scheduling field visits after streamflow-data errors have been detected.
The above techniques are essential to provide quality control in a real-
time data-collection system. The various data-verification techniques
described in this report can be used singly or in combination. It is
recommended that the simple and less costly verification techniques be
implemented first and their effectiveness evaluated before the more
complex and costly techniques are developed. The optimum level of
quality control must, however, be selected on the basis of the cost of
implementing these procedures and the need for accuracy.

40

REFERENCES

Anderson, J. A., 1971, Runoff evaluation and streamflow simulation by computer: U.S. Army, North Pacific Division, Corps of Engineers.

Carter, R. W., and Anderson, I. E., 1963, Accuracy of current-meter measurements: Am. Soc. Civil Engineers Jour., v. 89, no. Hy 4, p. 105-115.

Chow, V. T., 1964, Handbook of applied hydrology: New York, McGraw-Hill Book Co., sec. 25-II.

Lindsay, R. K., Kohler, M. A., and Paulhus, J. L. H., 1949, Applied hydrology: New York, McGraw-Hill Book Co., 689 p.

Riggs, H. C., 1966, Some statistical tools in hydrology: U.S. Geol. Survey Surface Water Techniques, book 4, chap. A1.

Rockwood, D. M., 1958, Columbia Basin streamflow routing by computer: Jour. Waterways and Harbors Div., Am. Soc. Civil Engineers, v. 84, pt. 1, paper 1874. Also published in ASCE Transactions, 1961, paper 3119.

Searcy, J. K., 1960, Graphical correlation of gaging-station records: U.S. Geol. Survey Water-Supply Paper 1541-C, p. C67-C100.

U.S. Water Resources Council, 1967, A uniform technique for determining flood flow frequencies: U.S. Water Resources Council Bull. 15, 15 p.

UNITED STATES DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY
WATER RESOURCES DIVISION
PORTLAND, OREGON