OPTIMIZING INFORMATION TRANSFER IN A STREAM-GAGING NETWORK

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations 30-75

U.S. DEPARTMENT OF THE INTERIOR
MARCH 3, 1849

BIBLIOGRAPHIC DATA SHEET

1. Report No.

2.

3. Recipient's Accession No.

4. Title and Subtitle
OPTIMIZING INFORMATION TRANSFER IN A STREAM GAGING NETWORK

5. Report Date
September 1975

6.

7. Author(s)
P. H. Carrigan, Jr. and H. G. Golden

8. Performing Organization Rept. No. USGS/WRI-30-75

9. Performing Organization Name and Address
U.S. Geological Survey - Water Resources Division
National Center, Mail Stop 415
12201 Sunrise Valley Drive
Reston, Va. 22092

10. Project/Task/Work Unit No.

11. Contract/Grant No.

12. Sponsoring Organization Name and Address
U.S. Geological Survey - Water Resources Division
National Center, Mail Stop 415
12201 Sunrise Valley Drive
Reston, Va. 22092

13. Type of Report & Period Covered
Interim

14.

15. Supplementary Notes

16. Abstracts Networks of small stream (less than 50 sq. mi.) gages have been operated by the U.S. Geological Survey throughout the country for a number of years to supplement flood information already available for large streams. The goal in operation of these networks has been to obtain sufficient data for estimating flood frequency at ungaged sites with the equivalent accuracy expected from 10 years of observed flood records. Most networks have accumulated sufficient data to satisfy these accuracy requirements. A review of these data, looking toward possible reduction of the number of gages in these networks, is now timely.

Thomas Maddock in 1974 developed a rational method for selecting gages to be retained in a reduced hydrologic network. This method of network analysis will result in selecting the optimum set of gages retained for a given level of annual operating costs with the information content of the reduced network being the factor optimized. Application of Maddock's method demonstrated a considerable number of gages can be eliminated from a network without grossly decreasing its information content.

Maddock's method of analysis is described and its use described in detail for a hypothetical network of gages. The method is applied to actual networks in Illinois, Georgia, and Montana.

17. Key Words and Document Analysis. 17a. Descriptors

Network design, stream gaging stations, operations research, computer models, linear programing, optimum design

17b. Identifiers/Open-Ended Terms

Mixed integer programing

17c. COSATI Field/Group

18. Availability Statement
Release unlimited

19. Security Class (This Report)
UNCLASSIFIED

20. Security Class (This Page
UNCLASSIFIED

21. No. of Pages
25

22. Price

FORM NTIS-35 (REV. 10-73) ENDORSED BY ANSI AND UNESCO. THIS FORM MAY BE REPRODUCED USCOMM-DC 8265-P74

OPTIMIZING INFORMATION TRANSFER IN A STREAM-GAGING NETWORK

By P.H. Carrigan, Jr. and H.G. Golden

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations 30-75

September 1975

UNITED STATES DEPARTMENT OF THE INTERIOR

GEOLOGICAL SURVEY

Vincent E. McKelvey, Director

For additional information write to:

U.S. Geological Survey, WRD
National Center, Mail Stop #415
Reston, Virginia 22092

CONTENTS

ILLUSTRATIONS

TABLES

OPTIMIZING INFORMATION TRANSFER IN A STREAM-GAGING NETWORK

By P. H. Carrigan, Jr.
and H. G. Golden

ABSTRACT

Networks of small stream (drainage area less than 50 square miles or 130 square kilometres) flood gages have been operated throughout the country for a number of years to supplement flood information already available for large streams. The goal in operating these networks has been to obtain sufficient data for estimating flood frequency at ungaged sites with the equivalent accuracy expected from 10 years of observed flood records. In some areas the networks have accumulated sufficient data to satisfy these accuracy goals. A review of these networks, looking toward possible reduction of the number of gages, is now timely. Continued operation of a few selected gages may be desirable to provide a longer time-sample base for improving the flood-frequency estimating equation and(or) to expand the area over which the equations apply.

In 1974, Thomas Maddock III developed a rational method for selecting gages to be retained in a reduced hydrologic network. This method of network analysis seeks the optimum set of gages to be retained for a given level of annual operating costs with the information content of the reduced network being the factor optimized. Application of Maddock's method demonstrated that a considerable number of gages could be eliminated from a network without grossly decreasing its information content.

Maddock's method of analysis is described in detail for a hypothetical network of gages. The method also is applied to actual networks in Montana, Illinois, and Georgia.

The analysis of networks in Montana illustrates the basic approach to selecting an optimal subset from the existing set of gages. The Illinois analysis demonstrated that by retaining only 26 percent of the gages, nearly 63 percent of the original information is retained in the reduced network. Application of the procedure shows how the design of networks in Georgia may be modified because of hydrologic considerations to meet budgetary constraints.

INTRODUCTION

Networks of stream gages in small drainage basins have been operated throughout the country for a number of years. These networks were established to supplement flood-frequency information already being obtained for larger drainage basins (greater than 50 mi^2 or 130 km^2). The supplemental small stream flood gaging program has been operated by the Geological Survey in cooperation with State or Federal agencies. State highway departments and the Federal Highway Administration have supported the program in most States for the purpose of obtaining flood information needed for design of bridges, culverts, and other highway related drainage structures.

The Geological Survey (Benson and Carter, 1973) and an Interagency Advisory Committee (Federal Interagency Work Group on Hydrologic Data for Small Watersheds, 1974) have considered the question of how much small stream flood frequency data might be desirable. As an accuracy goal for small stream flood frequency programs they recommended collection of sufficient data to allow estimation of the flood characteristics at any ungaged point on any stream with an accuracy at least equal to the accuracy of a frequency relation defined from 10 years of records collected at that site. This goal implies the use of some technique for transferring information from gaged to ungaged sites. Mathematical relations between flood peak flows and drainage basin characteristics have been defined by multiple regression analysis for many regions, thus, providing a means of transferring information to ungaged sites. In the Interagency Advisory Committee report (1974) a density of one small stream gage in an area of 500 mi^2 (1,300 km^2) was suggested as necessary to define transfer relations which would meet this accuracy goal. The Committee recommended that 20 percent of these gages be operated for at least 20 years and that the remainder be operated for 10 to 15 years. More recently Moss and Karlinger (1974) proposed analytical techniques for assessing the adequacy of available data for defining multiple regression relations to meet the recommended minimum accuracy goals and for further assessing any additional data needs.

Once the minimum accuracy goals have been met, data collection might cease. But Fiering (1965) and Matalas (1967) show that continued operation of a reduced number of stations will improve the time-sampling accuracy for both the continued and discontinued sites. There are additional reasons why it might be desirable to continue operation of only a subset of gages in an existing network; for example, when budgetary constraints force reduction of the gaging effort. For whatever reason, when a data collection effort is to be reduced, it is necessary to select a subset of the existing network for continued operation.

Maddock (1974) developed a rational method for selecting stations to be retained in a reduced network. This technique considers all feasible subsets of stations to maximize the sum of information contributed directly from the retained stations and indirectly through correlation with the discontinued station. The feasible combinations are delimited by a budgetary constraint (sum of operating costs for subset of retained stations cannot exceed selected budget limit), by information transferability constraints (correlation coefficients which define information may be required to be greater than zero and must be absolutely greater than square root of reciprocal of years of record minus two), and information is transferred from only one station.

Maddock's technique for station selection has been applied to small stream flood gaging networks in several States. The technique involves selecting stations to be retained in each network, and indicates the pairing of stations for maximum information transfer. The results of applying this technique will be illustrated in this report. The technique, however, should only be considered as a tool from which decisions would be made about final selection of stations in the network. Approaches to analyzing results of Maddock's method for selecting stations will be illustrated by considering decisions which could be reached regarding networks in Montana, Illinois, and Georgia.

ACKNOWLEDGMENTS

Computer programs needed to carry out computational tasks were developed by Thomas Maddock III, H. E. Robinson, and Gerald Knecht of the U.S. Geological Survey. Information about the networks analyzed was supplied by G. M. Pike, L. A. Martens, and J. R. George, District Chiefs, U.S. Geological Survey, in, respectively, Montana, Illinois, and Georgia.

MATHEMATICAL DESCRIPTION OF TECHNIQUE

The gages in the reduced network can be selected by considering, separately and jointly, the relative quantity of information in the original and reduced network and desirable budget levels.

One means of weighing the information directly contributed by each station in the network is to consider the error, $\sigma(\hat{u})$, in estimating its mean annual flood, $\hat{u}$ - the smaller the error, the greater the information. Fisher (1949) defined this measure of information as

$$I_R = \frac{N_1}{\sigma^2(u)} ,$$

where N_1 is the period of record and I_R is the information with respect to the mean annual flood contributed by a station retained in the network.

Even though a station is eliminated from the network, information will continue to be acquired for the site as a result of correlation with data collected at another station retained in the network. The information to be derived for the discontinued station through correlation between stations i and j, called I_C $(N_1 + N_2)$, is a function of the correlation coefficient for their common period of record, the period of common record (N_1), and the anticipated period over which the reduced network will be operated (N_2).

Information with respect to the mean annual flood may be transferrable (see eq. 5) but information with respect to the 100-year flood may not be. The information for the discontinued station i, derived through correlation can be compared to the information furnished by the retained station j in the ratio

$$I_{i,j} = \frac{I_C\ (N_1 + N_2)}{I_R\ (N_1)} \ .$$

If $I_{i,j} > 1$, additional information is furnished to the discontinued site. Multiplying this ratio by $(N_1 + N_2)/N_1$ normalizes $I_{i,j}$, that is $0 < I_{i,j} \leq 1$, provided the squared correlation coefficient between stations i and j exceeds $1/(N_1 - 2)$; if the squared correlation coefficient is less, then $I_{i,j} = 0$. Maddock (1974) called the ratio the information content.

Using the normalized ratio of information content implies that all stations in the network have about the same period of record (N_1). This operational assumption holds for the examples treated in this report. (Recently, computer programs have been modified to treat unequal records).

The objective of the network analysis is to retain as much information as possible in the reduced network, subject to certain constraints on gaging costs. This objective may be expressed for a network of M stations by

$$Z = \max \sum_{i=1}^{M} \sum_{j=1}^{M} I(i,j)\ \delta\ (i,j) \tag{1}$$

where Z is called the objective function and is the maximum over all possible combinations of $\delta(i,j)$. The decision variable $\delta(i,j)$, equal to 0 or 1 indicates, whether a station (i=j) is to be kept in the network (1), or discontinued (0), and whether information (i≠j) can be transferred from station j to station i (1) or not (0). Information from only one station j can be transferred to another station i, thus the constraint

$$\sum_{j=1}^{M} \delta(i,j) = 1,\ i=1,\ ...,M \tag{2}$$

must be imposed. Furthermore, information must be obtained at a station, i.e., $I(i,j) > 0$, if information is to be transferred to others, so that the constraint,

$$A\ \ \delta(j,j) - \sum_{\substack{i=1 \\ i\neq j}}^{M} \delta(i,j) \leq 0,\ j=1,\ ...,M \tag{3}$$

where A is a number larger than M, must also be included.

The information I(i,j) obtained be retaining a station (i=j, $\delta(i,i) = 1$) or by transferring information from station j to station i(i≠j, $\delta(i,j) = 1$)

was formulated by Maddock (1974) in the following simplified form:

$$I(i,j) = \frac{1}{1 + N_2\left[\left(\frac{1-r^2(i,j)}{N_1}\right)\left(\frac{2-N_1}{3-N_1}\right)\right]} \quad (4)$$

Equation 4 is a function of the correlation coefficient $r(i,j)$ between stations i and j for their concurrent period of record, N_1 years, and of the expected duration for the continued operation of the reduced network, N_2 years. Study of this equation will show $I(i,j)$ increases, as would be expected, with an increase in $r(i,j)$ and N_1 and with a decrease in N_2.

Maddock wished to emphasize the relative contribution to the information content of the reduced network which is furnished through correlation. Therefore, he first compared the gain in information about station j which is obtained through correlation with station i to the information obtained at station i. Second, he adjusted this comparison so that $I(i,j) = 1$ for $i=j$. He also recognized two limitations which apply or may be imposed on equation 4. If the population correlation coefficient $r(i,j)$ and the period of record N_1 satisfy the constraint

$$r^2 > \frac{1}{N_1-2},$$

information transfer is possible; on the otherhand,

$$I(i,j) = 0 \text{ if } r^2(i,j) \leq \frac{1}{N_1-2}, \quad (5)$$

and information is not transferrable. The derivation of the correlation constraint on information transfer is given by Fiering (1963).

Generally, no hydrologic significance is attached to a correlation if $r(i,j) < 0$; and the analyst can optionally specify

$$I(i,j) = 0 \text{ if } r(i,j) < 0. \quad (6)$$

The final constraint to be applied to the analysis is

$$\sum_{i=1}^{M} C_i \quad \delta(i,i) \leq B \quad (7)$$

where C_i is the annual operating cost for station i which is retained in the network, and B is the anticipated budget for the reduced network.

In the application of Maddock's technique, only sample values of $r(i,j)$ are available, and, hence, only sample values of $I_{i,j}$ are determined. The user must be award of this distinction; he should realize the actual information transferred may be different from the computed information (even zero) because population correlation coefficients may be different from sample coefficients. (See Maddock, 1974, p. 339).

EXAMPLE OF APPLYING TECHNIQUE

As an example, assume a network of 5 stations, numbered 1, ... 5, with annual operating costs of, respectively, $500, $1,200, $1,200, $1,000, and $800, for which the following matrix of information I(i,j) was computed:

STATION	Station 1	Station 2	Station 3	Station 4	Station 5
1	1	0.386	0.593	0.373	0.458
2	.386	1	.513	.878	.675
3	.593	.513	1	.520	.736
4	.373	.878	.520	1	.665
5	.458	.675	.736	.665	1

Note the matrix is symetric.

The present budget for the 5 stations is $4,700 per year. The budget goal for a reduced network is $2,300. All stations will be considered candidates for continuation or elimination. The total information available from the five stations is 5.0

To illustrate station selection for retention in the network, consider the case where 3 stations are to be retained. There are 10 possible combinations of stations taken 3 at a time: 1-2-3, 1-2-4, 1-2-5, 1-3-4, 1-3-5, 1-4-5, 2-3-4, 2-3-5, 2-4-5, and 3-4-5. For each combination of 3 stations there are 9 possible pairings with stations eliminated from the network. For instance, the combination of stations 1, 4, and 5, which meets the target budget ($2,300), can be paired with the discontinued stations 2 and 3 to produce the indicated total information (computed from equation 4) as shown in table 1.

Table 1.--Possible pairings for transfer of information from stations 1, 4, and 5 to stations 2 and 3.

Pairings	Total Information
1-2, 1-3	3.98
1-2, 4-3	3.91
1-2, 5-3	4.12
1-3, 4-2	4.47
1-3, 5-2	4.27
4-2, 5-3	4.61
4-2, 4-3	4.40
4-3, 5-2	4.20
5-2, 5-3	4.41

As listed in table 1, the maximum total information occurs for the pairings of stations 4-2 and 5-3 (best transfer pairs) where $I_{1,1} + I_{4,4} + I_{5,5} + I_{4,2} + I_{5,3} = 4.61$.

In a similar manner all possible combinations for retaining 1, 2, or 4 stations can be considered and the best transfer pairs found for each combination. The results of such an analysis is shown in table 2.

Table 2.--Total information and annual operating cost for all combinations of 1, 2, 3, and 4 station networks

Retained Station Combination	Best Transfer Pairs	Total Information	Annual Operating Cost
	1 Station discontinued		
1, 2, 3, 4 a/	3-5	4.74	$(3,900)
1, 2, 3, 5 a/	2-4	4.88*	(3,700)
1, 2, 4, 5 a/	5-3	4.74	(3,500)
1, 3, 4, 5	4-2	4.88	(3,500)
2, 3, 4, 5 a/	3-1	4.59	(4,200)

Table 2 continued.--Total information and annual operating cost for all combinations of 1, 2, 3, and 4 station networks

Retained Station Combination	Best Transfer Pairs	Total Information	Annual Operating Cost
	2 Stations discontinued		
1, 2, 3 a/	2-4, 3-5	4.61	(2,900)
1, 2, 4 a/	1-3, 2-5	4.27	(2,700)
1, 2, 5 a/	5-3, 2-4	4.61*	(2,500)
1, 3, 4	4-2, 3-5	4.61*	(2,700)
1, 3, 5	5-2, 5-4	4.34	(2,500)
1, 4, 5	4-2, 5-3	4.61*	2,300
2, 3, 4 a/	3-1, 3-5	4.33	(3,400)
2, 3, 5 a/	3-1, 2-4	4.47	(3,200)
2, 4, 5 a/	5-1, 3-5	4.19	(3,000)
3, 4, 5	3-1, 4-2	4.47	(3,000)
	3 Stations discontinued		
1, 2	1-3, 2-4, 2-5	4.15	$ 1,700
1, 3	3-2, 3-4, 3-5	3.77	1.700
1, 4	4-2, 1-3, 4-5	4.14	1,500
1, 5	5-2, 1-3, 5-4	3.93	1,300
2, 3 a/	3-1, 2-4, 3-5	4.21*	(2,400)
2, 4 a/	2-1, 4-3, 2-5	3.58	2,200
2, 5 a/	5-1, 5-3, 2-4	4.07	2,000
3, 4	3-1, 4-2, 3-5	4.21*	2,200
3, 5	5-1, 5-2, 5-4	3.80	2,000
4, 5	5-1, 4-2, 5-3	4.07	1,800
	4 Stations discontinued		
1	1-2, 1-3, 1-4, 1-5	2.81	500
2 a/	2-1, 2-3, 2-4, 2-5	3.45	1,200
3	3-1, 3-2, 3-4, 3-5	3.36	1,200
4	4-1, 4-2, 4-3, 4-5	3.44	1,000
5	5-1, 5-2, 5-3, 5-4	3.53*	800

a/ Network in which station 2 is retained
* Maximum information for number of stations
( ) Means target budget of $2,300 exceeded

For emphasis in scanning results shown in table 2, the maximum information for each combination of retained stations has been indicated and annual operating costs exceeding the target budget have been indicated.

Table 3 lists the variation of maximum information with number of stations retained, and figure 1 shows the variation graphically.

Table 3.--Maximum information and annual operating cost for optimum subsets of 1 to 5 stations

Number Stations Retained	Maximum Total Information	Number of Combinations With Some Information	Annual Cost
5	5.00	1	$4,700
4	4.88	2	3,500 - 3,700
3	4.61	4	2,300 - 2.900
2	4.21	2	2,200 - 2,400
1	3.53	1	800

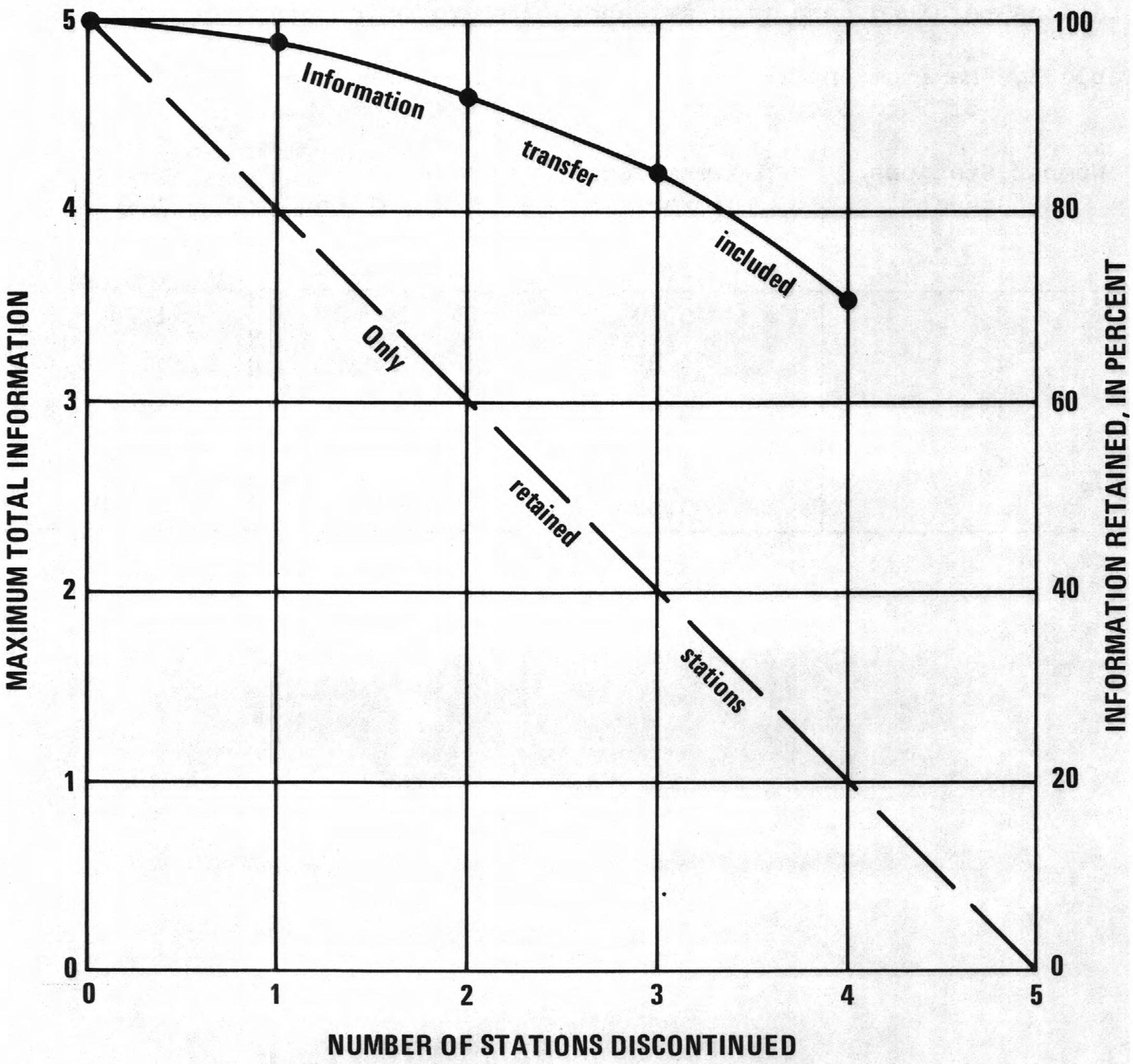


Figure 1.--Variation in maximum total information with number of stations discontinued from network.

In table 3 and figure 1 note that the maximum total information with retention of 1 station (3.53) is considerably greater than 1.0, the information obtained at the station itself. The information level (eq. 1) attained from the reduced network will exceed the number of stations retained provided information can be transferred to the discontinued stations. It is conceivable, however, that as a result of not exceeding the constraints in equations 5 and 6, the information retained would just equal the number of retained stations.

One other computational option is to declare the retention of one or more specific stations in the reduced network. Suppose that station 2 must be retained. Then, the combination of retained stations which does not exceed the target budget and provides the maximum total information is the station pair 2 and 5, with information of 4.07 and at a cost of $2,000 (table 2).

COMPUTER PROGRAMS

Equations 1 to 4 and 7 are solved using mixed integer programing which is part of the mathematical programming package available from International Business Machines, Inc. (1971).

As input to the mixed integer program, the correlation coefficients for the period of record common to each pair of stations have to be computed, the common period of record of each possible pair has to be determined, and the correlation coefficients have to be screened according to the criterion set forth in equation 5. The user may elect to apply the screening criterion set forth by equation 6, and the program user has to select the number of years for continued operation of reduced network, set a budget limit, and ascertain each station annual operating cost.

The user may wish to retain selected stations because the continued collection of data is necessary for some other purpose, flood-profile studies for instance. The user may also decide the selection process would be improved if data from continuous record stations, located in the vicinity of the crest-stage-gage network, are added to the analyses; he would retain continuous-record stations because (1) they are part of another network and (2) the purpose of their record is more comprehensive.

The annual flood data for the analyses are retrieved from the U.S. Geological Survey Peak Flow File. This computer file contains for each station a listing of the date, magnitude, stage, and special qualifications (for instance, peak affected by regulation) for each annual flood.

A flow chart for the programs used to apply Maddock's technique is shown in figure 2.

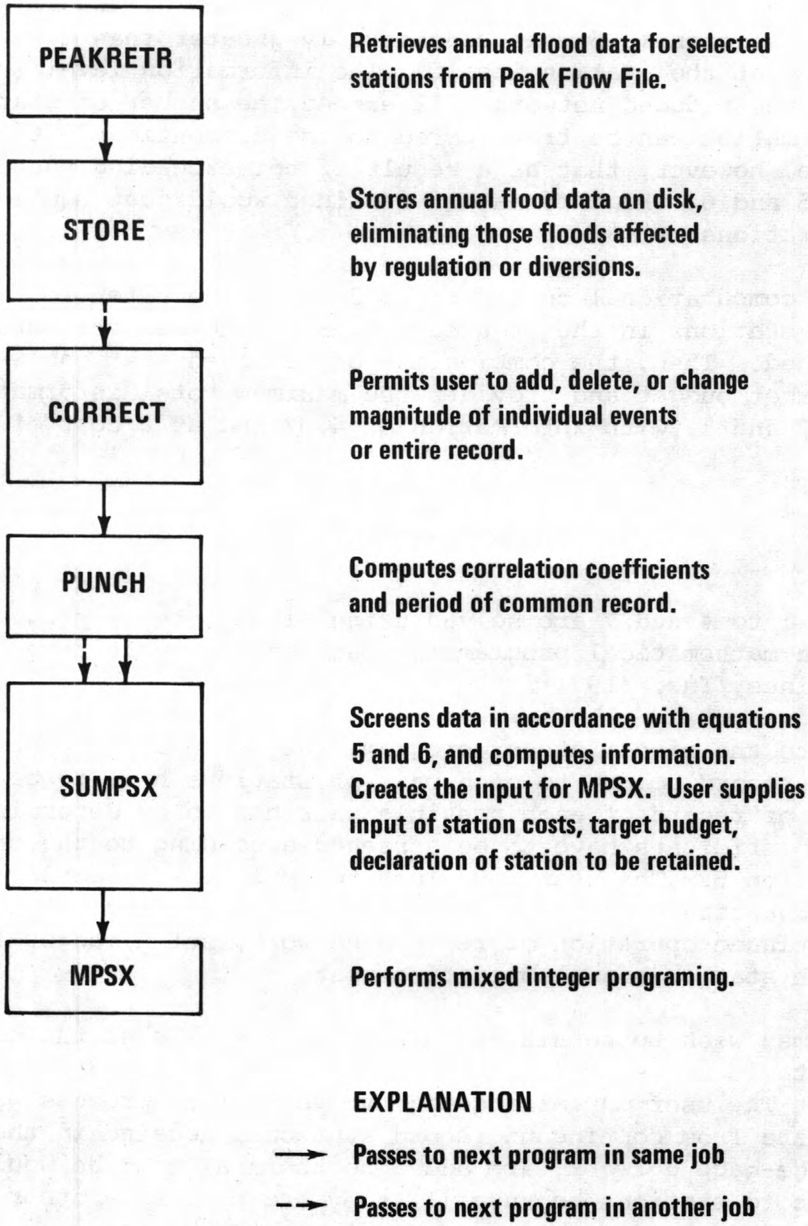


Figure 2.--Flow chart for programs which select stations to be retained in network.

APPLICATIONS OF TECHNIQUE

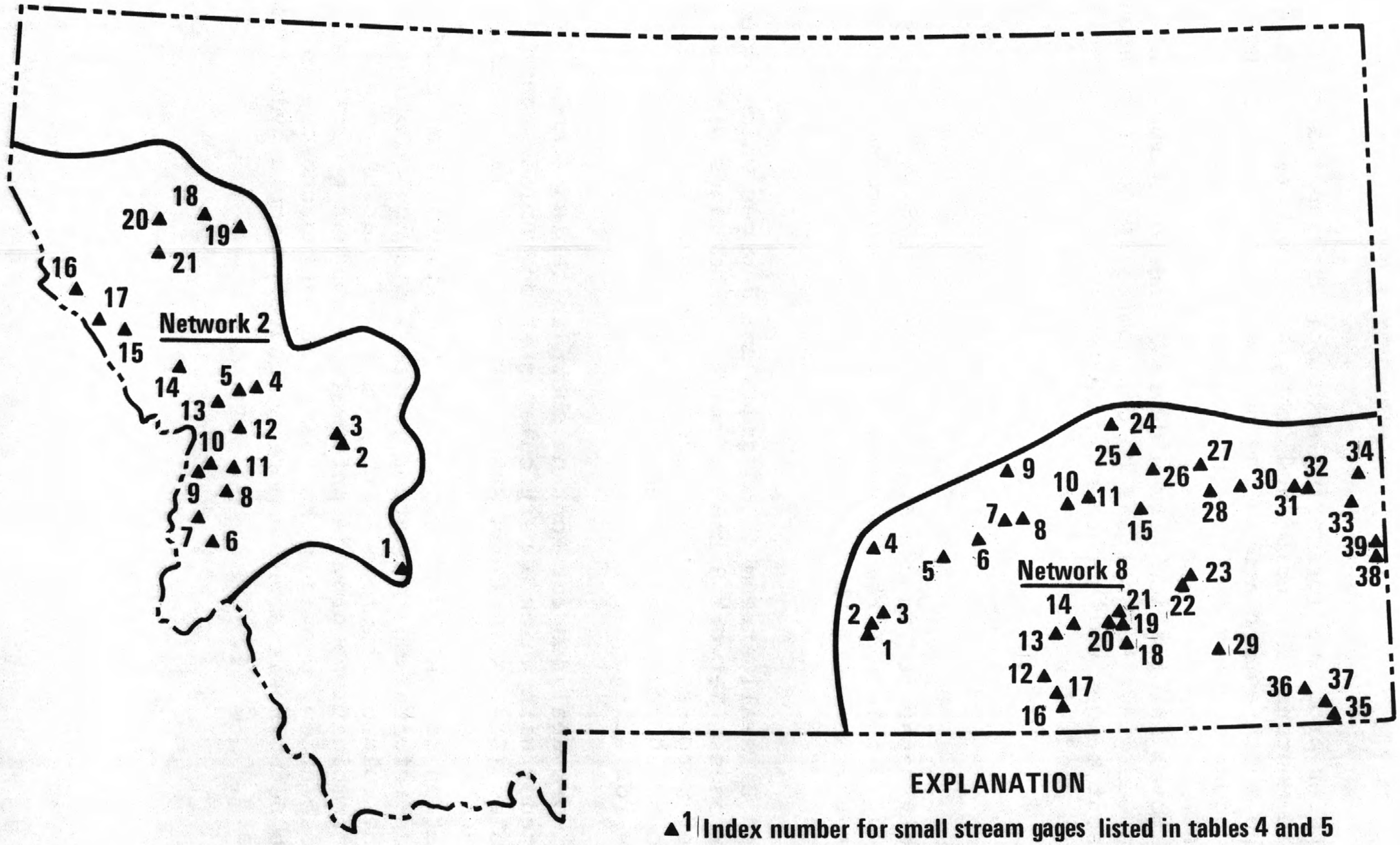
The system of computer programs (preceding section) was applied to analyses of networks in Montana, Illinois, and Georgia. Two networks in Montana, located in the southeastern and in the northwestern parts of the State, were used to suggest changes in the composition of the network as the number of stations was reduced. The analysis in Illinois was undertaken to study variation in total information with a decrease in the number of retained stations. In Georgia, the goal was to cut operating costs to about 20 percent of the present budget while retaining those stations that sample a balance of hydrologic conditions.

Montana

Two networks, widely separated in the State (fig. 3), were examined by the technique described in this report: network 2 in the mountainous northwestern part of the State, and network 8 in the southeastern part of the State.

Network 2 contained 21 small stream flood gages and 7 gages which were operated for other purposes. Network 8 had 39 small stream gages with no stations used for other purposes. For both networks the planning horizon (N_2) was considered to be 20 years and the annual operating costs were considered to be the same for all gages.

The results of network analyses for Montana shown in tables 4 and 5 illustrate the consistency with which a particular station replaces another on the switching of roles (Sta. A replaces Sta. B or Sta. B replaces Sta. A) as the network decreases in size. In network 2 for example, station 7 provides transfer information to station 10 with 3 and 5 stations eliminated, but station 10 provides information to station 7 with 7 stations eliminated. This interchange of the role of stations as the number of stations eliminated grows, or as feasible combinations provide information content nearer and nearer to the optimal solution, provides the means for making choices in final selection of the network sites which more adequately sample hydrologic variations in the area.



EXPLANATION

▲1 Index number for small stream gages listed in tables 4 and 5

Figure 3.--Sketch map of network areas in Montana.

Table 4.--List of stations which provide transfer data for given reduction in size of network 2 in Montana

Index Number	Station Number	Small Stream	Other Purpose Record	Index Number of Equivalent Gage — Number of Stations Eliminated: 3	5	7
1	12-3233.00	X				
2	12-3247.00	X				
3	12-3248.00	X				
4	12-3399.00	X				
5	12-3402.00	X			12	
6	12-3443.00	X				
7	12-3458.00	X				10
8	12-3485.00	X		28*	28*	28*
9	12-3502.00	X				
10	12-3505.00	X		7	7	
11	12-3510.00	X		26*	26*	26*
12	12-3514.00	X				5
13	12-3522.00	X				
14	12-3534.00	X			21	21
15	12-3538.00	X				16
16	12-3585.50	X				
17	12-3541.00	X				24*
18	12-3705.00	X				
19	12-3709.00	X				
20	12-3743.00	X				
21	12-3757.00	X				
22	12-3907.00		X			
23	12-3895.00		X			
24	12-3540.00		X			
25	12-3465.00		X			
26	12-3320.00		X			
27	12-3241.00		X			
28	12-3434.00		X			

*Other purpose station

Table 5.--List of stations which provide transfer data for given reduction in size of network 8 in Montana

Index Number	Station Number	Index Number of Equivalent Gage					
		Number of Stations Eliminated					
		4	15	18	20	26	27
1	6-2162.00		3	3	2	14	2
2	6-2163.00	18	3	3		26	
3	6-2165.00	1			2	14	2
4	6-2177.00	28	9	9	39	9	39
5	6-2178.00				16	38	38
6	6-2944.00					27	27
7	6-2948.00		6		6	9	35
8	6-2948.50						
9	6-2949.00				39		39
10	6-2950.20			24			
11	6-2950.50		31	31	31	27	27
12	6-2951.00						
13	6-2951.30		39			39	39
14	6-2952.00						
15	6-2961.00					26	2
16	6-3069.00			5		35	35
17	6-3069.50					35	35
18	6-3076.40		26	21	2	26	2
19	6-3076.60					12	12
20	6-3077.60		19	19	19	12	12
21	6-3077.80					39	39
22	6-3082.00			17			
23	6-3083.00			17	22	22	22
24	6-3090.20					10	
25	6-3090.40			37	37	36	22
26	6-3090.60			15	2		2
27	6-3090.80						
28	6-3090.90		9	9	39	9	39
29	6-3247.00	11	31	31	31	27	24
30	6-3264.00		6	6	6	22	22
31	6-3266.00					39	24
32	6-3266.50			27	22		
33	6-3267.00					10	10
34	6-3268.00		17	17	17	22	22
35	6-3329.00		16		2		
36	6-3341.00		16	35	16		35
37	6-3342.00		25			35	35
38	6-3346.40				27		
39	6-3347.20						

Illinois

The gaging stations of Illinois were divided into two networks of stations for the analysis as the large number of available stations would have required an inordinate amount of time for processing as a single network in the MPSX computer program (fig. 4). Network 1 contained 47 stations of which 44 were part of the small streams supplementary flood data program gages and 3 were operated for other purposes. Network 2 contained 48 stations of which 44 were part of the program of small-stream gages and four were operated for other purposes.

For this study, the assumption was made that all other purpose stations must be retained in the network (only small stream gages were candidates for elimination). The cost of operating all stations were assumed to be equal. The planning horizon (N_2 in eq. 4) was chosen to be 20 years. Six computer runs were made with data from each network, eliminating in turn 5, 10, 15, 20, 30, and 35 gages from each network. As shown in figure 5, the information content decreases almost linearly with an increase in the number of stations eliminated from the network. With 35 stations eliminated from a network (26 percent remaining), nearly 63 percent of the information available from the original network is retained. This relatively high retention of information in the reduced network illustrates a high degree of information transfer and is the reason that the curve in figure 5 shows no abrupt change in slope toward zero information when a large number of stations has been eliminated.

Georgia

Small stream flood gages have been operated in clusters as shown in figure 6. The networks were originally selected to sample three geographical provinces, Ridge and Valley, Piedmont, and Coastal Plain and, in general, the analysis to select an optimum subset followed that sampling structure. Because the coastal zone had a large number of gages, that network was divided into two parts, 3A and 3B, to reduce computer costs.

The numbers of small stream flood gages and other stations in present and reduced networks with present, target, and achieved annual costs are listed in table 6.

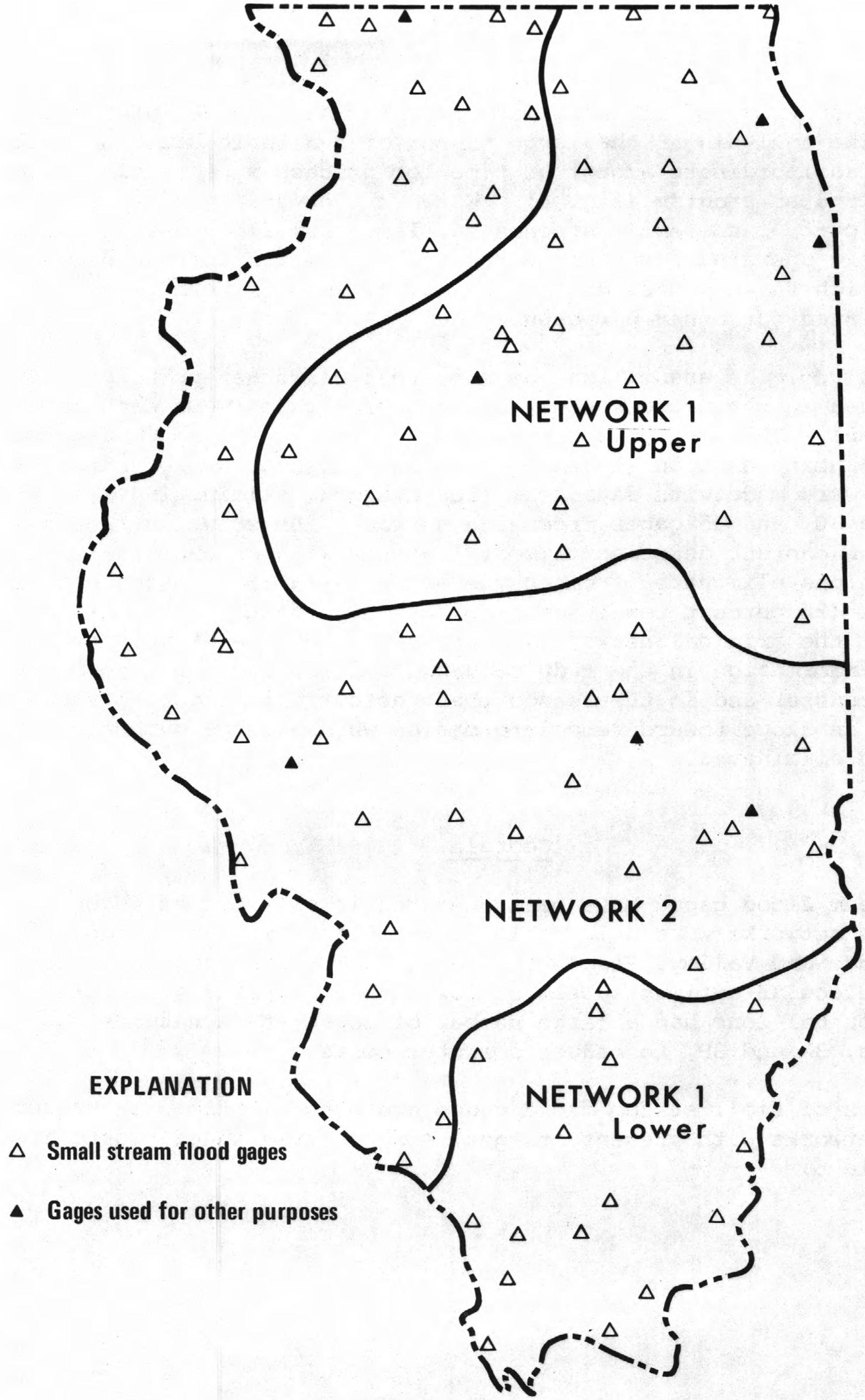


Figure 4.--Sketch map of network areas in Illinois.

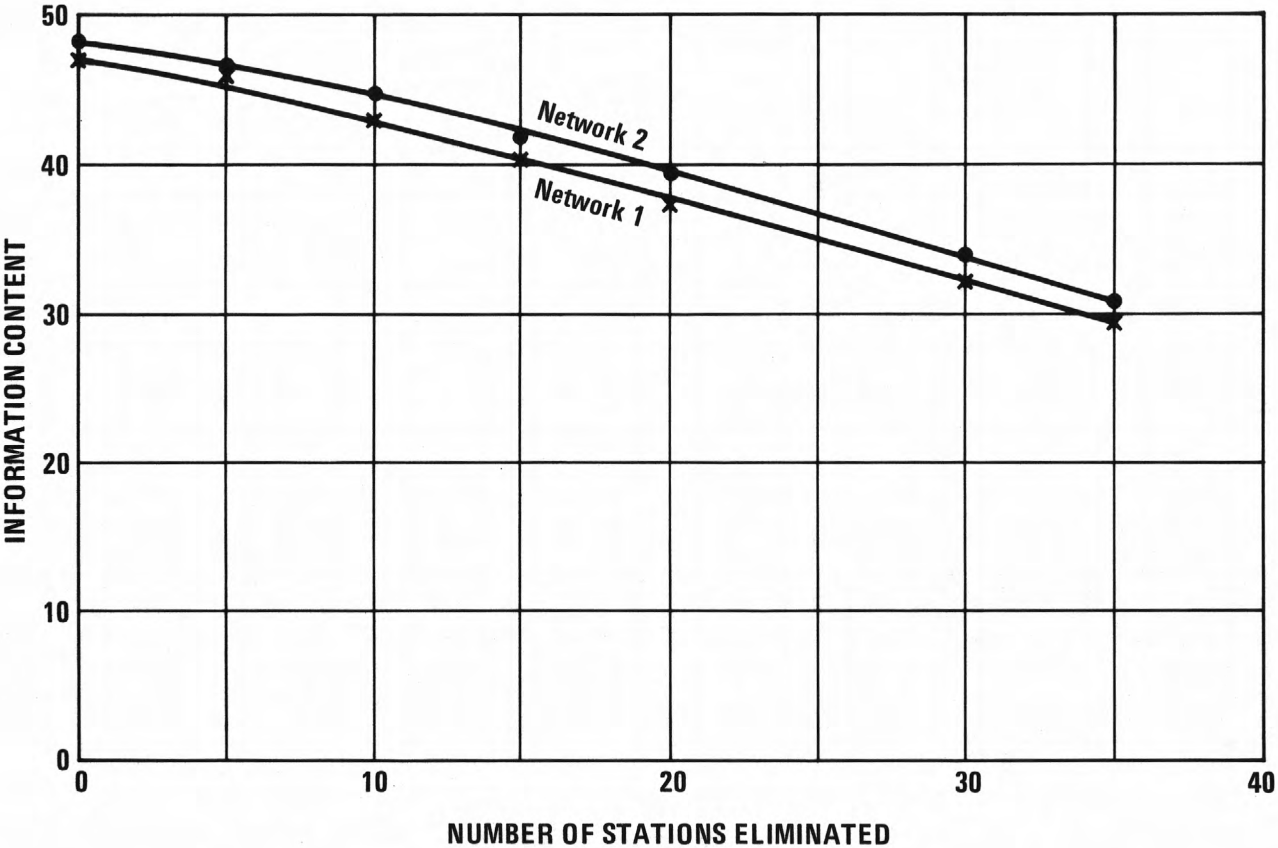


Figure 5.--Variation in information content with number of stations in Networks 1 and 2 of Illinois.

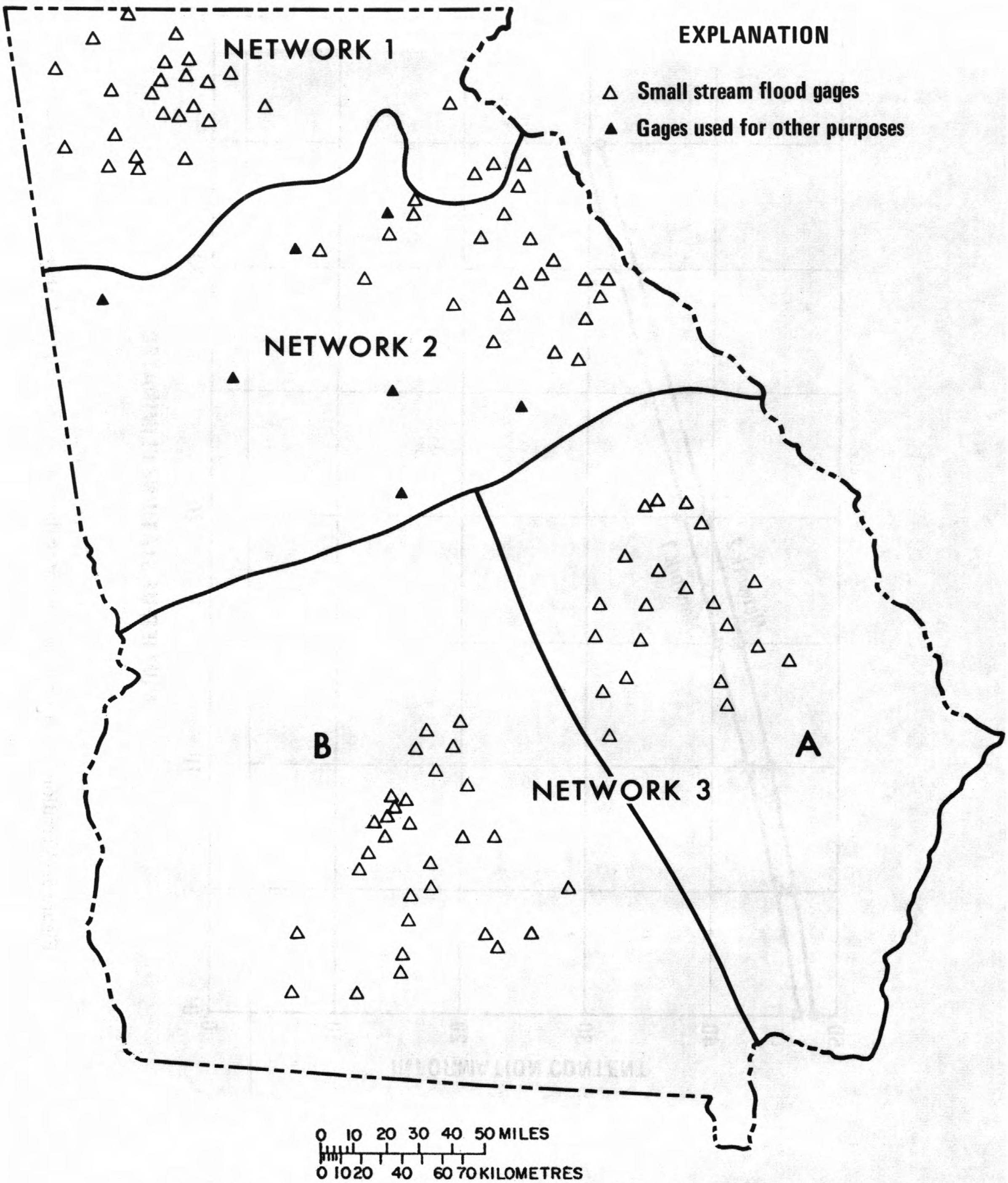


Figure 6.--Sketch map showing locations of existing gages and network areas in Georgia.

Table 6.--Number of stations in Georgia networks including annual costs for present network, target network, and achieved network

Network	Number of Stations				Actual Cost		
	Present		Reduced				
	Small Stream	Other Purpose	Small Stream	Other Purpose	Present	Target	Achieved
1	26	0	5	0	$18,500	$3,000	$2,400
2	23	7	6	7	19,500	3,900	4,300
3A	21	0	5	0	13,650	2,730	2,100
3B	30	0	5	0	18,900	3,780	2,600

The planning horizon for continued operation of the reduced network (N_2 in eq. 4) was 15 years. The annual operating costs for the small-stream flood gages ranged from $400 to $1,200 per year.

As indicated in figure 7, the reduced network finally selected for continued operation was different from the network designated using the technique described in this report. These modifications in station selection were desired because those stations optimally designated by the computer program tended to cluster in a small area of the existing network and failed to maintain a desired representative sample of either basin characteristics or areal dispersion. The final selection of stations for a reduced network more fully sampled variations in the basin characteristics. Each station to be included in the final network was indicated by substitution pairings of stations in intermediate results in the MPSX program.

The MPSX program seeks the optimal solution for the mixed-integer programing problem in steps for which the solutions are feasible, that is, properly satisfy constraints in equations 2, 3, 5, and 6. The information level found in each successive step in the program will be greater than the preceding step. In the final step, the optimal solution is reached. The results of each intermediate step will indicate feasible pairings of stations for which information transfer would occur. Often the intermediate feasible solutions will indicate information is transferred from station 4 to 7 in one step and from station 7 to 4 in the next step. This documentation in the program of alternate stations in the intermediate feasible solutions is often advantageous. Stations selected in an intermediate solution may be better suited for sampling hydrologic conditions than in the optimal solution. Thus, the final network chosen may not be optimal from the standpoint of the mathematical criteria of this technique, but is more desirable from an overall management viewpoint.

Results of the analysis, combining the results of the computer selection with the appraisal of hydrologic conditions is summarized in table 7.

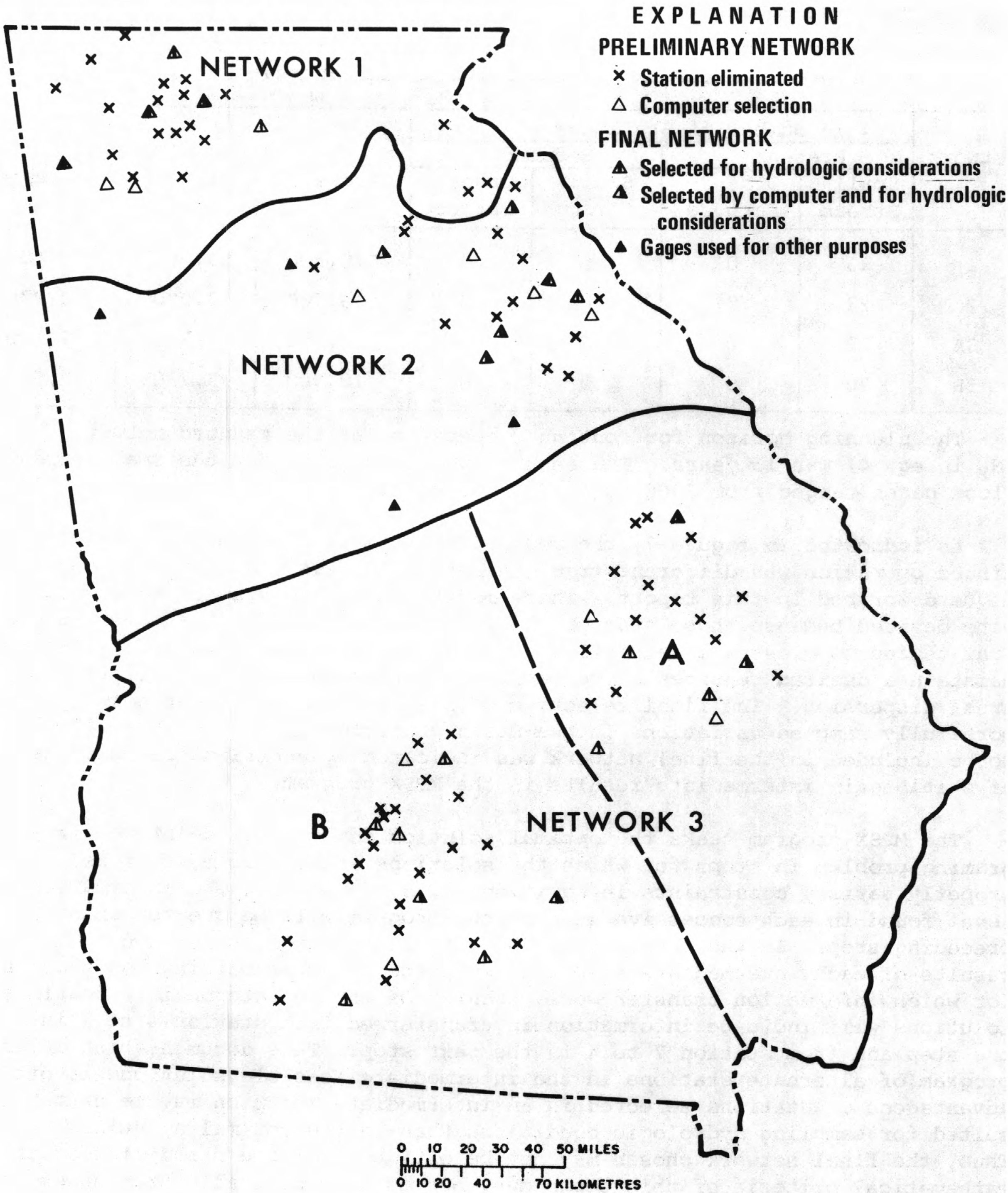


Figure 7.--Sketch map of location of gages in the reduced network in Georgia.

Table 7.--Summary of analyses for Georgia networks

Network	Number of stations in reduced network				Total information in network		
	Computer selection		Based on hydrologic considerations	Final network	Original	Computer solution	Final
	Original	Final					
1	6	4	1	5	26.0	19.8	18.6
2	6	2	4	6	30.0	22.2	21.2
3A	5	3	2	5	21.0	12.9	12.4
3B	7	5	2	7	30.9	19.8	18.4

The aforementioned use of intermediate solutions is illustrated by the results in the analysis summarized for network 3B in table 7. For this network only five of the original seven computer selected stations were retained; two other stations were selected based on hydrologic considerations. Examination of figure 8 indicates the logic for selecting alternate stations based on hydrologic considerations. One of the computer-selected stations that was eliminated has about the same size drainage area and channel slope as those of stations that were retained (area, 1.34 mi^2 or 3.47 km^2; slope, 25.2 ft/mi or 4.77 m/km) but does not provide the desired variability in basin characteristic; the substitute station has an area of 6.4 mi^2 (16.6 km^2). An extremely small basin (0.14 mi^2 or 0.36 km^2) was indicated as desirable in the computer analysis but was eliminated because of operation difficulties and land-use changes in the basin. The adjustments for hydrologic considerations reduced the total information content from 19.8 to 18.4, or about 7 percent. The management rationale expressed herein is to adjust the network of stations so as to obtain a suitable range of basin characteristics in the reduced network without significantly reducing the information content. This general procedure was followed in the management analysis of all four Georgia networks.

CONCLUSIONS

A mathematical programming technique developed by Maddock (1974) has been used to identify the optimum subsets of stations to be continued as an operating network of small-stream gages in Montana, Illinois, and Georgia. The technique identifies the combination of stations that would retain the maximum information content in the reduced network and, if desired, limits the cost so that it does not exceed some predetermined budget level.

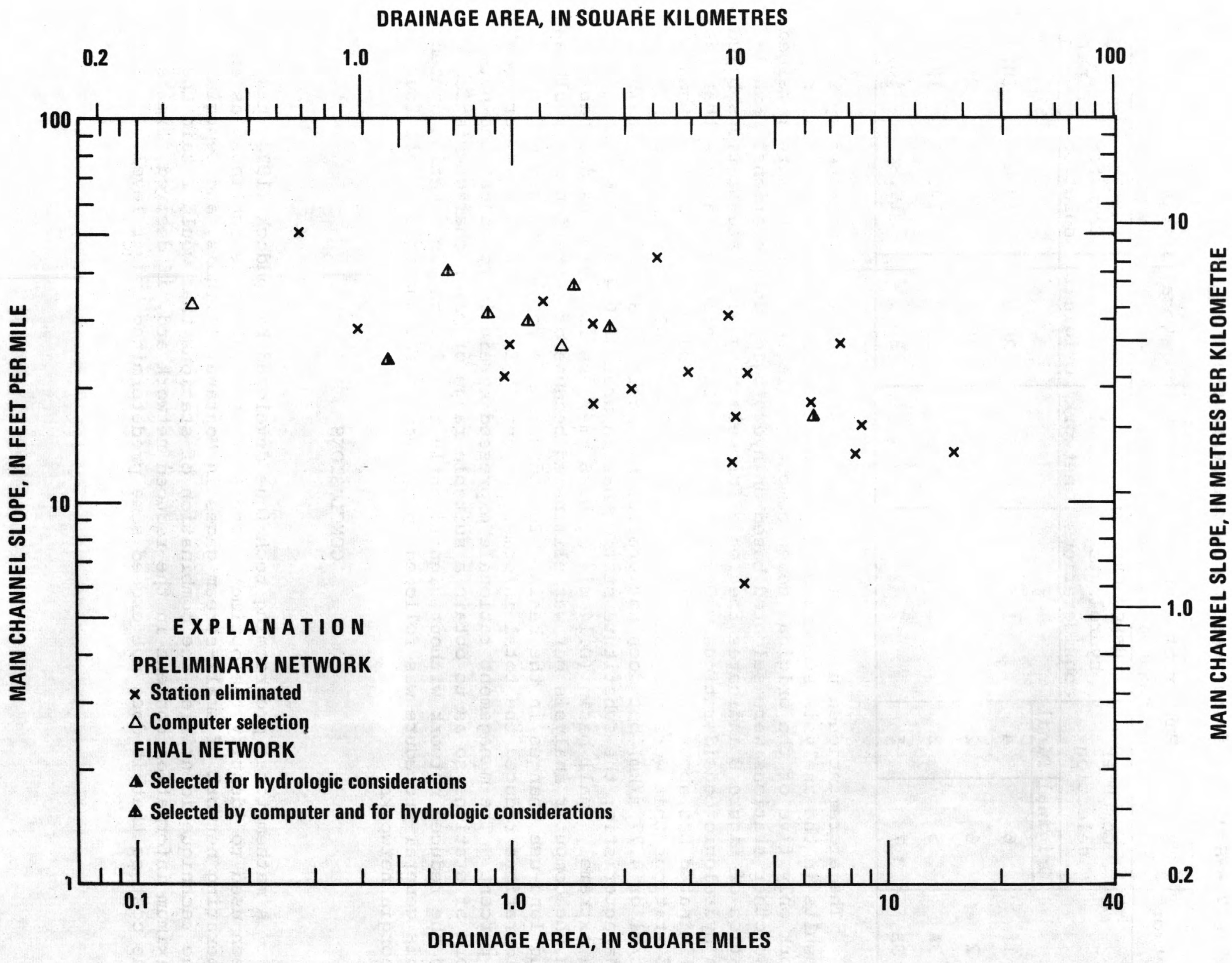


Figure 8.--Drainage area and channel slope characteristics of stations in Georgia Network 3B.

Applications of the technique demonstrate that the decrease in information level may be much less than the decrease in station numbers as increasingly greater numbers of stations are eliminated from a network. This high retention of information in a severely reduced network offers a great incentive to evaluate present network configurations.

As results of analysis of networks in Georgia indicate, it became apparent that the optimal solution may not provide an adequate selection of stations for sampling of hydrologic condtions. Intermediate results from computer processing or multiple computer processing may be used to provide alternate selections of stations which would provide a guide as to how stations in the reduced network can be selected to satisfy these hydrologic sampling needs.

REFERENCES

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

Federal Interagency Advisory Committee on Water-Data, 1974, Hydrologic data need for small watersheds, streamflow and related precipitation data: U.S. Geol. Survey, Reston, Va.

Fiering, M. B., 1963, Use of correlation to improve estimates of the mean and variance: U.S. Geol. Survey Prof. Paper 434-C.

______ 1965, An optimization scheme for gaging: Water Resources Research, v. 1, no. 4, p. 463-469.

Fisher, R. A., 1949, The design of experiments: London, UK, Oliver and Boyd, p. 182.

International Business Machines, Inc., 1971, Mathematical programming system extended. Mixed integer programming. Program description: IBM, p. 5734-XM4.

Maddock, Thomas, III, 1974, An optimum reduction of gauges to meet data program constraints: Bull., Intl. Assoc. Hydrologic Sciences, v. 19, no. 3, p. 337-345.

Matalas, N. C., 1967, Optimum gaging station location: Proc. IBM Scientific Computing Symposium on Water and Air Resource Management, October 23-25, 1967, Yorktown Heights, New York, p. 85-94.

Moss, M. E., and Karlinger, M. R., 1974, Surface water network design by regression analysis simulation: Water Resources Research, v. 10, no. 3, p. 427-433.