SURFACE WATER DATA NETWORK ANALYSIS FOR PUERTO RICO

By Patrick W. McKinley

U.S. GEOLOGICAL SURVEY
WATER-RESOURCES INVESTIGATIONS
REPORT 83-4055

Prepared in cooperation with the
PUERTO RICO DEPARTMENT OF NATURAL RESOURCES

San Juan, Puerto Rico

1985

UNITED STATES DEPARTMENT OF THE INTERIOR

WILLIAM P. CLARK, Secretary

GEOLOGICAL SURVEY

Dallas L. Peck, Director

For additional information
write to:

Chief, Caribbean District, WRD
U.S. Geological Survey
GPO Box 4424
San Juan, PR 00936
(Telephone: (809) 753-4414)

Copies of this report
can be purchased from:

Open-File Services Section
Western Distribution Branch
U.S. Geological Survey
Box 25425, Federal Center
Denver, Colorado 80225
(Telephone: (303) 234-5888)

CONTENTS

ILLUSTRATIONS

TABLES

ACKNOWLEDGMENTS

Drafting and Art......................José Alicea
Typing/Editing............Cármen A. García

SURFACE WATER DATA NETWORK ANALYSIS FOR PUERTO RICO

by
Patrick W. McKinley

ABSTRACT

The streamflow data collection network of Puerto Rico has been evaluated by a computer modeling technique known as NARI (Network Analysis for Regional Information). This technique utilizes the hydrologic information available from the present network to provide a basis for developing a future gaging network. The mean annual discharge, 2-year, 10-year, 50-year, and 100-year peak discharges were used as dependent or response variables in the regression analysis of the network. Other hydrologic characteristics such as low-flow were not analyzed.

Regression equations were calculated for the whole island of Puerto Rico using mean annual precipitation for the basin, and drainage area as independent or predictor variables. The standard error of estimate of regression varied from 21 percent for the mean annual discharge to 44 percent for the 100-year flood.

The results of the statistical analyses show a shortage of long-term records. The harmonic mean-record length for the flood analysis was only 8 years. A long-term network of 16 stations is proposed which if continued would decrease the statistical error in transferring data to ungaged sites, identification of trends, and future modeling efforts.

INTRODUCTION

Streamflow data-collection programs normally develop from the hydrologic-information needs of the data users. Special data-collection projects, which have a limited life and scope, often answer these needs. Hydrologists and planners also have recognized the need to anticipate data demands as insurance against decisions based on too little or inadequate hydrologic data (Moss, Lettenmaier, and Wood, 1978). In order to meet future data demands, the Commonwealth of Puerto Rico in cooperation with the U.S. Geological Survey developed and maintains a network of streamflow stations to serve as a base of hydrologic information that can be transferred to ungaged sites. This network must be periodically reevaluated and modified to reflect the influence of new data and a changing level of understanding of the hydrologic processes on the network design.

INTRODUCTION - Continued

In 1981, the U.S. Geological Survey, Water Resources Division, began a statistical evaluation of the current stream-gaging network in Puerto Rico. The project, in cooperation with the Puerto Rico Department of Natural Resources, presents a cost-effective gaging plan to provide basic hydrologic information for the mean-annual discharge, and peak discharges for the 2-, 10-, 50-, and 100-year frequency floods at ungaged sites. Low-flow characteristics were not investigated because of the uncertainty of statistical analyses by regression in low flow studies.

Regional regression equations were developed for Puerto Rico to predict each of these desired statistics at ungaged sites. Utilizing a statistical procedure known as Network Analysis for Regional Information, NARI, (Moss and others, 1982), a relation was developed between the standard error of regression, and the number of stations and record length. The network cost, represented as the number of stations operated, is then related to the standard error.

DESCRIPTION OF REGRESSION EQUATIONS

Drainage area and mean annual precipitation were used as independent or predictor variables in calculating the regression equations. Drainage area was determined from U.S. Geological Survey 1:20,000 topographic maps. Average annual precipitation was calculated from U.S. Weather Service data using the isohyetal method. For each equation an associated standard error of regression was computed (an estimate of how well the observed data fit the regression model). Residuals from the regression equations were plotted by latitude and longitude to identify areas of hydrologic similarity. The island was then divided into areas of similar hydrologic and geographic characteristics in an attempt to minimize the standard error. This method failed to decrease the standard error and therefore equations modeling the whole island were used instead of smaller regional equations.

A Statistical Analyses System (SAS) (Barr and others, 1976) was used to compute the regression equations and associated standard error. Discharge data were obtained from the U.S. Geological Survey WATSTORE files. The equations developed are shown in **table 1**.

DESCRIPTION OF REGRESSION EQUATIONS - Continued

Table 1.--Description of regression equations.

EQUATIONS	STANDARD ERROR, PERCENT	STANDARD ERROR NATURAL LOGARITHMIC UNITS
$Q_m = 0.0006\ A^{0.99}\ P^{1.86}$	21	0.211
$Q_2 = 1.35\ A^{0.77}\ P^{1.38}$	36	.353
$Q_{10} = 10.18\ A^{0.85}\ P^{1.06}$	35	.343
$Q_{50} = 38.86\ A^{0.90}\ P^{0.85}$	41	.393
$Q_{100} = 62.80\ A^{0.92}\ P^{0.77}$	44	.420

Where Q is the stream discharge in cubic feet per second, A is the drainage area in square miles and P is the average mean annual precipitation in inches per year. The subscripts of Q are m for the mean annual discharge and the 2-, 10-, 50-, and 100-year frequency flood peak.

A network of 31 stations used in the mean annual discharge regressions and 24 stations for the peak flow discharges are listed in table 2.

Statistical restrictions caused rejection of stations that had a short record-length or limited peak-flow record. The station locations (figure 1) represent the principal streams in Puerto Rico and/or sites established for specific areal studies. The report by López, Colón, and Cobb (1979) contains flood-frequency equations similar to the above regression equations, however, the equations in that report are not superceded by the equations in this report.

Table 2.--Stations used in developing the regression equations, years of record analysed, and selected basin characteristics.

MAP REFERENCE NUMBER	STATION NAME	CONTRIBUTING DRAINAGE AREA, IN SQ MI	ANNUAL PRECIPITATION, IN INCHES	YEARS OF RECORD ANALYZED
1	*50028000 Río Tanama nr Utuado.	18.4	79.7	18
2	50029000 Río Grande de Arecibo at Central Cambalache.	267.0	77.7	9
3	*50031200 Río Grande de Manatí nr Morovis.	55.2	79.7	12
4	*50034000 Río Bauta nr Orocovis.	16.7	81.1	8
5	*50035000 Río Grande de Manatí at Ciales.	128.0	75.9	16
6	50038100 Río Grande de Manatí at Hwy 2 nr Manatí.	159.0	73.7	8
7	*50038320 Río Cibuco at Corozal.	15.1	84.9	9
8	50039500 Río Cibuco at Vega Baja.	66.0	70.9	5
9	*50043000 Río de la Plata at Proyecto la Plata.	54.8	74.8	18
10	*50046000 Río de la Plata at Toa Alta.	200.0	70.1	17
11	50049000 Río Piedras at Río Piedras.	12.5	75.7	7
12	*50055000 Río Grande de Loíza at Caguas.	89.8	81.6	18
13	50056400 Río Valenciano nr Juncos.	16.4	84.9	7
14	*50057000 Río Gurabo at Gurabo.	60.2	80.0	18
15	*50061800 Río Canóvanas nr Campo Rico.	9.84	125.0	11

Table 2.--Continued

MAP REFERENCE NUMBER	STATION NAME	CONTRIBUTING DRAINAGE AREA, IN SQ MI	ANNUAL PRECIPITATION, IN INCHES	YEARS OF RECORD ANALYZED
16	*50063800 Río Espiritu Santo nr Río Grande.	8.62	148.0	11
17	*50064200 Río Grande nr El Verde.	7.31	159.0	9
18	*50065700 Río Mameyes at Hwy 191 at Mameyes.	11.8	140.0	12
19	*50071000 Río Fajardo nr Fajardo.	14.9	106.0	17
20	*50075000 Río Icacos nr Naguabo.	1.26	194.0	3
21	*50082800 Río Guayanes nr Colonia Laura.	4.69	120.0	9
22	50090500 Río Maunabo at Lizas.	5.38	113.0	7
23	*50092000 Río Grande de Patillas nr Patillas.	18.3	86.5	12
24	*50112500 Río Inabon at Real Abajo.	9.7	81.1	12
25	*50114000 Río Cerrillos nr Ponce.	17.8	87.0	14
26	*50115000 Río Portugues nr Ponce.	8.82	86.6	14
27	*50136000 Río Rosario nr Rosario.	17.6	76.1	8
28	50138000 Río Guanajibo nr Hormigueros.	120.0	72.9	5
29	*50141000 Río Yahuecas nr Adjuntas.	15.4	79.0	11
30	*50144000 Río Grande de Añasco nr San Sebastián.	130.0	89.9	15
31	*50147800 Río Culebrinas at Hwy 404 nr Moca.	71.2	93.4	11

* Used in peak-flow regression.

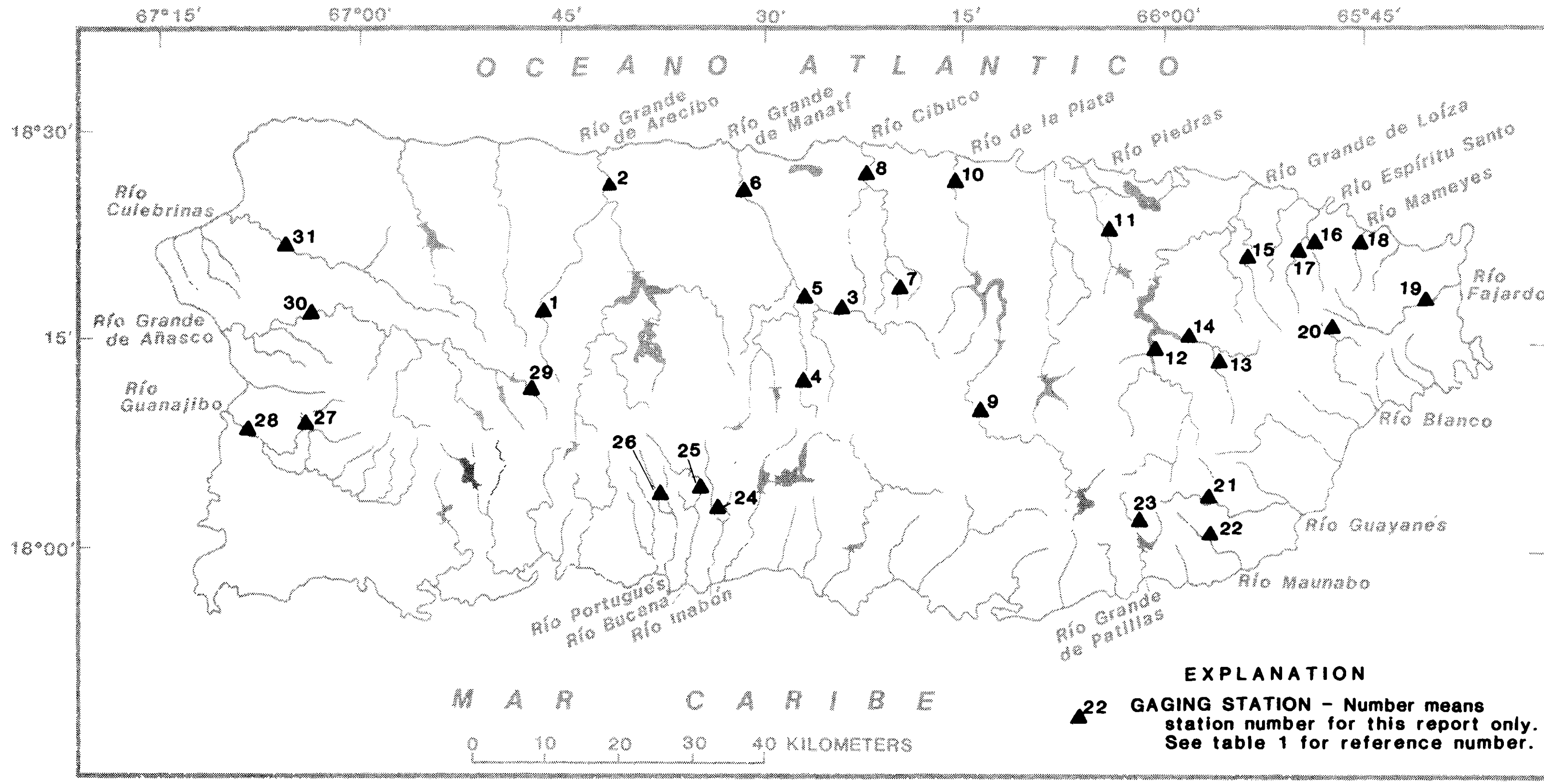

Figure 1.--Geographic location of gaging stations used in regressions.

METHOD OF EVALUATION

The method of evaluation used was a series of computer programs collectively known as NARI, Network Analysis for Regional Information, which is based on the technique of network design introduced by Moss and Karlinger, (1974).

The NARI technique takes the true standard error of estimate of a regional regression (Moss and others, 1982) as a measure of the regional information contained in a data network. Uncertainty in the true standard error of estimate, S_t, is handled by treating S_t as a random variable with a probability density function. Within a hydrologic region, this function is assumed to be dependent only on the following parameters:

(1) NB, the adjusted number of stations in the regional regression. NB is defined by:

$$NB=n-k+1,$$

Where k is the number of independent variables used in the regression (in this application k=2) and n is the number of stations used in the regression.

(2) NY, the harmonic mean record length of the stations in the regression. NY is defined by:

$$NY=(\sum_{i=1}^{n}(m_i n)^{-1})^{-1}$$

Where m is the record length, in years, at the ith station.

(3) S_o, the observed standard error of estimate. S_o is an estimate of the standard deviation of the regional regression residuals.

Output from a NARI analysis includes a family of probability distributions of the true standard error indexed by, or conditioned on, values of NB and NY. The 50-percentile point or median true standard error for each distribution is selected and the results plotted as a function of NB and NY in figure 2 for each flow characteristic.

These graphs represent changes in hydrologic information provided by the network. The standard error decreases as the number of stations and (or) their record lengths increase. The network design, and cost, are affected by the information needs. The available budget may limit the design and acceptable information error.

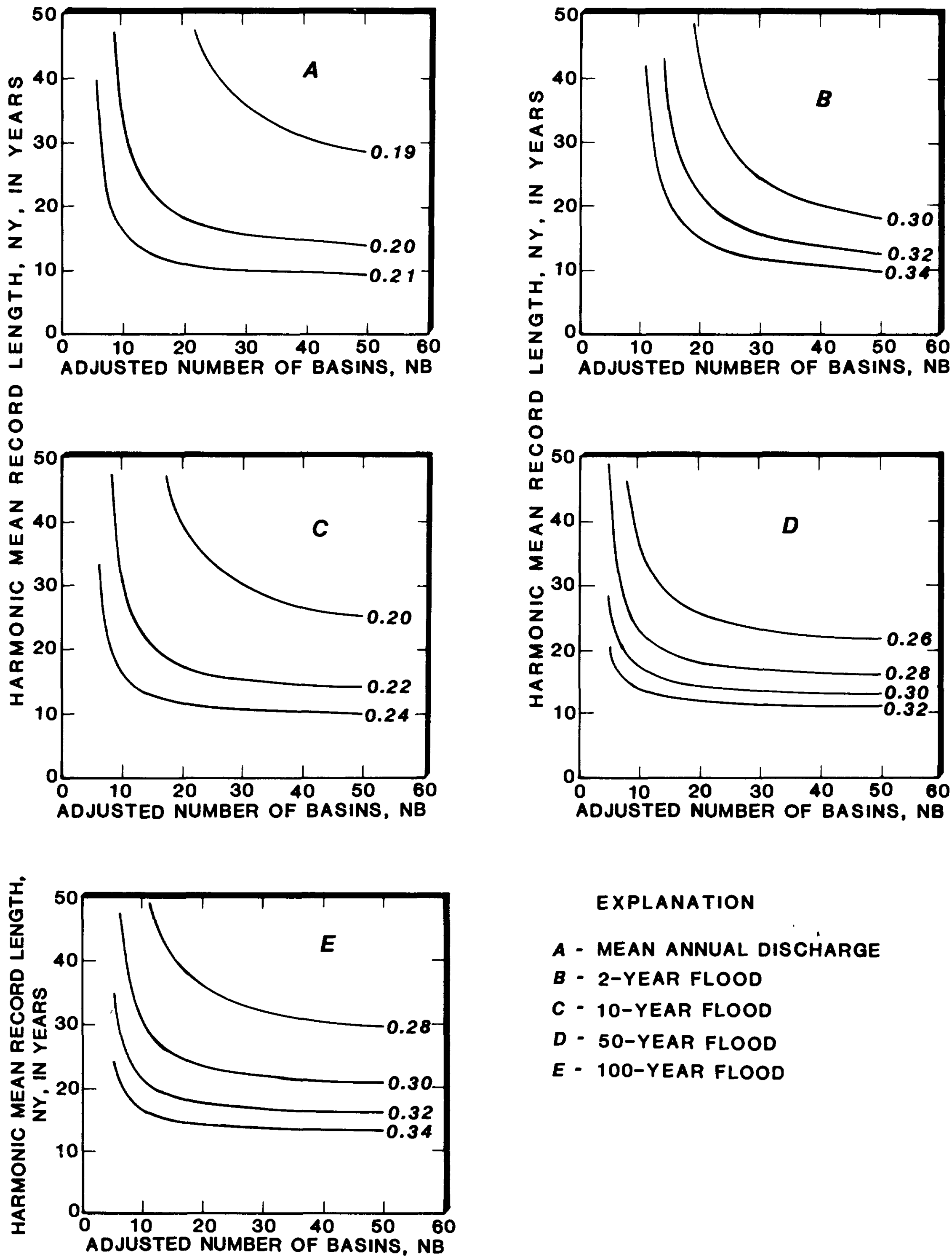

Figure 2.--True standard error for the mean annual discharge, 2-year, 10-year, 50-year, and 100-year flood.

GAGING-NETWORK DESIGN

The goal of network design is to identify a future gaging plan that is efficient in collecting regional information and also meets the data requirements of special projects.

The NARI technique allows the network designer to preselect the number of gaging sites that must be operated because of project requirements or separate cooperator funding. Stations that are expected to be operated in Puerto Rico during the next 20 years are listed in **table 3.** Three gaging networks described as planning horizons (5-, 10-, or 20- years) are distinguished in the list of stations. The 5-year horizon differs from the 10- and 20-year horizon in that it contains short-term stations which are usually associated with specific projects. The 10 and 20-year plans are representative of long-term stations.

Table 3.--Stations included in the 5-, 10-, and 20-year planning horizons.

STATION NUMBER	STATION NAME	PLANNING HORIZON
50028000	Río Tanamá nr Utuado	5,10,20
50029000	Río Grande de Arecibo at Cambalache	5,10,20
50030460	Río Orocovis nr Orocovis	5
50035000	Río Grande de Manatí Hwy 2 nr Manatí	5,10,20
50039500	Río Cibuco at Vega Baja	5,10,20
50046000	Río de la Plata at Toa Alta	5,10,20
50055000	Río Grande de Loíza at Caguas	5,10,20
50056400	Río Valenciano nr Juncos	5,10,20
50057000	Río Gurabo at Gurabo	5,10,20
50067000	Río Sabana nr Sabana	5
50071000	Río Fajardo nr Fajardo	5,10,20
50075000	Río Icacos nr Naguabo	5
50075500	Río Blanco at Florida	5,10,20
50081000	Río Humacao at Las Piedras	5
50092000	Río Grande de Patillas nr Patillas	5,10,20
50112500	Río Inabón at Real Abajo	5,10,20
50114000	Río Cerrillos nr Ponce	5
50115000	Río Portugués nr Ponce	5
50124200	Río Guayanilla nr Guayanilla	5,10,20
50126150	Río Yauco above Diversion Monserrate nr Yauco	5
50128000	Río Yauco nr Yauco	5
50138000	Río Guanajibo nr Hormigueros	5,10,20
50141000	Río Yahuecas nr Adjuntas	5,10,20
50144000	Río Grande de Añasco nr San Sebastián	5,10,20
50147800	Río Culebrinas a Hwy 404 nr Moca	5,10,20

GAGING-NETWORK DESIGN - Continued

NARIPLOT, a computer procedure of NARI, (G.D. Tasker, written commun., 1980), was used to calculate the minimum standard error for each planning horizon as a function of the number of stations operated (fig. 3). The small square on each dashed line denotes the minimum standard error, optimum number of stations in the regression equation, and optimum harmonic mean-record length for the indicated number of stations operated.

Minimum regional information is gained by operating more than 25 stations. This is indicated by the small decrease in standard error of estimate as the number of stations increases from 25 to 45. Selecting the twenty year planning horizon, the most economical plan graphed is for 25 stations. The maximum information that can be attained from this network when NY equals 29 years and NB equals 31 stations is shown in figure 3C. In this scenario, 25 stations are operated during any one year but a total of 31 stations are used in the regression. The standard error for this example is 0.299 (natural logarithmic units).

The advantage of maintaining stations over a long period of time is illustrated in figures 2 and 3. The present gaging network in Puerto Rico is represented by the small triangle near the bottom of the graph, (figure 3). The adjusted number of basins is 23. This is less than the actual number of gaging stations presently operating on the island, but represents those stations that have sufficient years of record to be analysed statistically.

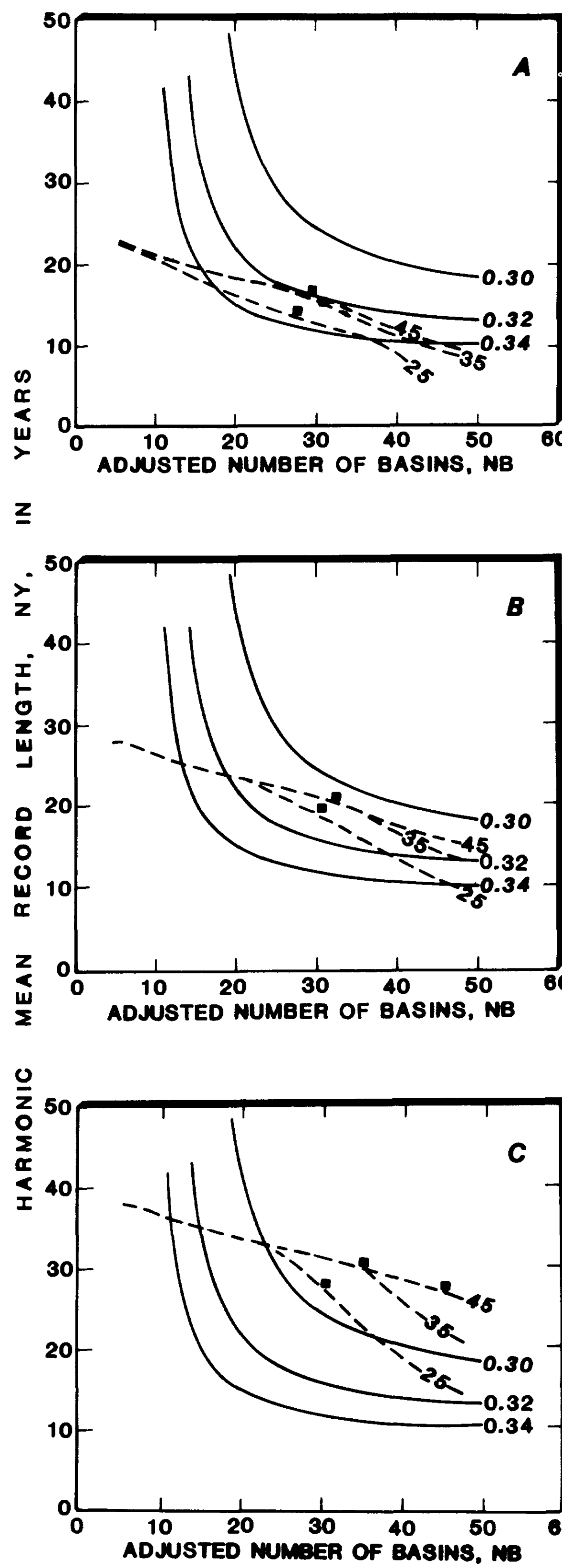

Figure 3.--NB-NY frontier overlaid on lines of standard error of estimate for the two year peak flood analysis.

GAGING-NETWORK DESIGN - Continued

To achieve the 20-yr planning horizon point of minimum error for 25 stations, the 16 designated stations in table 2 would need to be operated for 20 years. Nine additional stations, of which up to 5 could be new stations, would also need to be maintained. This gaging plan clearly will "move" the network along a path of relatively rapid decrease in standard error of estimate. For future modeling work and prediction of significant statistics of Puerto Rico's water resources, it is essential that long-term stations be maintained. Presently, there are only six operating stations that have 20 or more years of record.

In developing this plan, consideration was given only to estimating the mean-annual flow and the flood-frequency statistics previously described. Low-flow statistics, perhaps more important than mean and peak flows, cannot be estimated from these procedures. If the only purpose for collecting data at a particular site is to collect regional information about the mean and peak flows, stations not included in table 3, could be discontinued if the regression model and standard error described are acceptable to the data user. However, many of those stations are special purpose at the request of other Federal and Commonwealth agencies and are designed to provide other flow statistics.

A word of caution must be added to avert a tendency to over simplify. Actual deviation from predicted discharges at individual stations will differ from the standard error noted for the regression equations. This difference may be quite large in either the positive or negative direction. Therefore, when the regression equation is used to define discharges at a site with little or no data, the manager must remember that he is working with statistical averages and not absolutes.

CONCLUSIONS

The evaluation of Puerto Rico's surface-water stream-gaging network by the NARI technique suggests a more efficient network. It is clear from the analysis that some stations can be discontinued while decreasing the sampling error associated with estimation of mean-annual flow and flood frequency. The recommended network is listed in table 4. Sites listed as long-term is considered the foundation of the dicharge data collection program. A station for a special project that is presently unforeseen can be added to this list, but, should not replace a long-term station unless the overall system is reanalyzed. Old project stations and network stations that were originally designed as short term stations were selected to be discontinued.

There are other hydrologic occurrences, such as low flows, that are equally as important in network planning as those analyzed in this report. This should be kept in mind when reviewing the network set forth here. Additional stations may be required to satisfy other needs, but are not considered in this study.

Table 4.--Recommended network and cooperators.

STATION NUMBER	STATION NAME	COOPERATOR	STATUS
50028000	Río Tanamá nr Utuado	EQB	*/LT
50029000	Río Gde de Arecibo at Cambalache	EQB-DOA	*/LT
50030460	Río Orocovis nr Orocovis	DNR	*/D
50031200	Río Grande de Manatí nr Morovis	EQB	D
50034000	Río Bauta nr Orocovis	EQB	D
50035000	Río Gde de Manatí at Ciales	EQB	ST
50038100	Río Gde de Manatí Hwy 2 nr Manatí	EQB	LT
50038320	Río Cibuco below Corozal	EQB	ST
50039500	Río Cibuco at Vega Baja	EQB	ST
50043000	Río de la Plata at Proj. La Plata	EQB	ST
50046000	Río de la Plata at Toa Alta	EQB	LT
50049000	Río Piedras at Río Piedras	EQB	D
50050900	Río Gde de Loíza at Quebrada Arenas	COE-PRASA	*/ST
50051310	Río Cayaguas at Cerro Gordo	COE-PRASA	*/ST
50055000	Río Gde de Loíza at Caguas	EQB-PRASA	*/LT
50056400	Río Valenciano nr Juncos	EQB-PRASA	LT
50057000	Río Gurabo at Gurabo	EQB-PRASA	LT
50061800	Río Canóvanas nr Campo Rico	EQB	ST
50063800	Río Espíritu Santo nr Río Grande	EQB	D
50064200	Río Grande nr El Verde	EQB	D
50065700	Río Mameyes at Hwy 191 at Mameyes	EQB	D
50067000	Río Sabana nr Sabana	COE	*/ST
50071000	Río Fajardo nr Fajardo	COE	*/LT
50075000	Río Icacos nr Naguabo	COE	*/D
50076000	Río Blanco nr Florida	PRASA	N/LT
50082000	Río Humacao at Hwy 3 at Humacao	PRASA	N/ST
50082800	Río Guayanés nr Colonia Laura	EQB	D
50090500	Río Maunabo at Liza	EQB	ST
50092000	Río Gde de Patillas nr Patillas	NASQAN	LT
50112500	Río Inabón at Real Abajo	EQB	LT

Table 4.--Continued

STATION NUMBER	STATION NAME	COOPERATOR	STATUS
50114000	Río Cerrillos nr Ponce	COE	*/ST
50115000	Río Portugués nr Ponce	COE	*/ST
50124200	Río Guayanilla nr Guayanilla	EQB	LT
50126150	Río Yauco nr Monserrate Div. Yauco	DNR	*/ST
50128000	Río Yauco nr Yauco	DNR	*/ST
50128950	Lajas Irrigation	DOA	*/ST
50129300	Lajas Drainage Canal nr Ensenada	DOA	*/ST
50136000	Río Rosario at Rosario	EQB	ST
50138000	Río Guanajibo nr Hormigueros	EQB-PRASA	LT
50141000	Río Yahuecas nr Adjuntas	EQB	*/ST
50141100	Yahuecas Reservoir nr Castañer	DOA	*/LT
50141500	Guayo Reservoir nr Castañer	DOA	*/ST
50142500	Prieto Reservoir nr Castañer	DOA	*/ST
50144000	Río Gde de Añasco nr Sebastián	NASQAN	LT
50147800	Río Culebrinas at HWY 404 nr Moca	EQB	LT

EXPLANATION FOR TABLE 4

COOPERATORS:

EQB, Environmental Quality Board;
DOA, Department of Agriculture;
DNR, Department of Natural Resources;
PRASA, Puerto Rico Aqueduct and Sewer Authority;
COE, U.S. Army Corps of Engineers;
NASQAN, National Stream Quality Accounting Network.

STATUS:

*, presently a project site;
N, New project station;
LT, Long term-20 years or more;
ST, Short term-5 years or more;
D, Discontinuation recommended.

LIST OF REFERENCES

Barr, A. J., Goodnight, J. H., Sall, J. P., and Helwig, J. T., 1976, A user's guide to SAS 76: Raleigh, N.C., SAS Institute, Inc. 329 p.

López, M. A., Colón-Dieppa, Eloy, and Cobb, E. D., 1979, Floods in Puerto Rico, magnitude and frequency: U.S. Geological Survey Water Resources Investigations 78-141, 66 p.

Moss, M. E., Gilroy, E. J., Tasker, G. D., and Karlinger, M. R. 1982, Design of surface water data networks for regional information: U.S. Geological Survey Water Supply Paper 2178, 33 p.

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.

Moss, M. E., Lettenmaier, D. P., and Wood, E. F., 1978, On the design of hydrologic data networks: EOS, Transactions, American Geophysical Union, v. 59, no. 8, p 772-775.