Design of Surface-Water Data Networks for Regional Information

By M. E. MOSS, E. J. GILROY, G. D. TASKER, and M. R. KARLINGER

U.S. GEOLOGICAL SURVEY WATER-SUPPLY PAPER 2178

UNITED STATES DEPARTMENT OF THE INTERIOR
JAMES G. WATT, Secretary

GEOLOGICAL SURVEY
Dallas L. Peck, Director

UNITED STATES GOVERNMENT PRINTING OFFICE 1982

For sale by Distribution Branch
Text Products Section
U.S. Geological Survey
604 South Pickett Street
Alexandria, Virginia 22304

Library of Congress Cataloging in Publication Data
Main entry under title:

Design of surface-water data networks for regional information

(Geological Survey water-supply paper ; 2178)
Bibliography: p.
Supt. of Docs. no.: I 19.13:2178
1. Information storage and retrieval systems—Hydrology.
I. Moss, Marshall E. II. Series: United States. Geological Survey. Water-supply paper ; 2178.
GB657.D47 553.7'8'028542 80-607878

CONTENTS

FIGURES

TABLE

LIST OF SYMBOLS

A_1	set of rainfall events not exceeding one precipitation unit
A_2	set of rainfall events exceeding one precipitation unit in a specified time period
a	intercept term in logarithmic regression
B_1	set of discharge events not exceeding 500 streamflow units
B_2	set of discharge events exceeding 500 streamflow units
b	slope term in logarithmic regression
$b_0, b_1, \ldots, b_k$	coefficients in general logarithmic regression function
C_v	true coefficient of variation
C_{vc}	coefficient of variation conditioned on flow being greater than zero
C_{vi}	sample value of coefficient of variation in ith gaging station
k	number of independent variables used in the regional regression
M_i	sample mean of annual flow values at ith gaging station
m_i	number of years of record at ith gaging station
$m_{i,j}$	number of concurrent years of record at ith and jth gaging stations
n	number of gaging stations used in the regional regression
NB	adjusted number of gaging stations used in the regional regression
NB^*	adjusted number of gaging stations in a possible future network configuration
NY	harmonic mean record length of the n records used in the regression
NY^*	harmonic mean record length associated with a possible future network configuration
p	probability of zero flow for a period of 1 year
$P(\bullet)$	probability of an event
$P(d_1, d_2)$	probability of joint occurrence of two events, d_1 and d_2
$P(d_1 \mid d_2)$	conditional probability of event d_1 given d_2 has occurred
Q_{ij}	recorded value of the annual hydrologic event in the ith year at the jth station
S_i	the sample standard deviation of annual flow values at station i
S_0	observed standard error of estimate in log e units of a regional regression
S_T	a random variable denoting the true standard error of estimate given a sample regression function
S_T^α	the αth percentile of the probability distribution of S_T; that value not exceeded by S_T with probability α
W_{ij}	natural logarithm of Q_{ij}
X	an estimate of a statistical parameter of streamflow
X_i	true value at station i of the streamflow parameter that is being regionalized
$\tilde{X}_i$	a sample estimate of X_i
$Y_1, Y_2, \ldots, Y_k$	physiographic parameters that describe the relevant portion of a drainage basin
y_i	surrogate physiographic parameter for the ith synthetic streamflow record
α	probability that the true standard error of a given network configuration does not exceed a specified value, S_T^α
γ	model error, the root mean square error of prediction of the regression relation if the regression parameters were defined without error
ϵ_i	a random component with zero mean and variance equal to γ^2
θ	true value of a streamflow parameter
ϱ_c	true value of cross correlation between two streamflow records
ϱ_{ij}	unadjusted sample cross correlation between two streamflow records
ϱ_X	true value of cross correlation between sample estimates of a streamflow parameter at two sites
σ_w^2	true value of variance of logarithms of a streamflow record
$\hat{\sigma}_w^2$	average value of sample variances of logarithms of streamflows at all sites used in the regionalization
τ	true value of cross correlation between logarithms of streamflows
$\hat{\tau}$	average value of sample interstation correlation between logarithms of streamflows for all stations usedin the regionalization

Design of Surface-Water Data Networks for Regional Information

By M. E. Moss, E. J. Gilroy, G. D. Tasker, *and* M. R. Karlinger

Abstract

This report describes a technique, Network Analysis of Regional Information (NARI), and the existing computer procedures that have been developed for the specification of the regional information-cost relation for several statistical parameters of streamflow. The measure of information used is the true standard error of estimate of a regional logarithmic regression. The cost is a function of the number of stations at which hydrologic data are collected and the number of years for which the data are collected. The technique can be used to obtain either (1) a minimum cost network that will attain a prespecified accuracy and reliability or (2) a network that maximizes information given a set of budgetary and time constraints.

INTRODUCTION

In the planning process for the development of the water resources of a particular region, certain hydrologic information is desirable. At the present time, however, the optimum level of hydrologic information demanded at any specific step of the planning process has not been defined. Generalized relations for the benefits derived from and the costs of obtaining hydrologic information have not yet been developed. As a result, hydrologic network design has not evolved beyond the stage in which the comparison of existing networks with arbitrarily defined criteria for information is used as the basis for the design procedure. To progress much beyond the current stage will require interaction between the hydrologist, who in essence will supply the information-cost relation, and the planner, who will provide the information-benefit relation. This report describes a technique, Network Analysis of Regional Information (NARI), and the existing computer procedures that have been developed for the specification of the regional information-cost relation for several statistical parameters for streamflow.

NARI is based on a regional regression approach (Benson and Matalas, 1967) for the definition of the streamflow parameters, and its output is an evaluation of the likelihood of various levels of improvement in the regression relations that may be obtained by the collection of additional streamflow data. NARI has evolved to fill a need that was highlighted by a national study of the U.S. Geological Survey's (USGS) streamflow data-collection program (Benson and Carter, 1973). This study compared the then existing level of regional information, as defined by the accuracy of a regional regression relation, with arbitrary criteria. In many cases the criteria were not met, and this elicited the question, "How much more streamflow data will be required to meet the criteria?" At the time of the national study this question could not be answered. Research toward meeting this need was first reported by Moss and Karlinger (1974). Subsequent testing and attempts at implementation have resulted in changes in some of the steps suggested by Moss and Karlinger; these changes have been incorporated into this paper.

To date (1981), the NARI technique has not been applied to a broad range of hydrologic and hydrometric settings. Therefore, it can not be considered a fully tested procedure. However, this paper describes the assumptions of the technique and provides sufficient insight into its underlying theories such that a careful user should be able to evaluate its use in a given implementation. To assist further in this evaluation, potential users of the NARI technique should also consult published applications such as those of Moss and Haushild (1978) and Tasker and Moss (1979). Moss (1979) also presents some valuable interpretive concepts.

OVERVIEW

Because time is required to collect hydrologic information and because site-specific demands for hydrologic data cannot be anticipated perfectly, the basic hydrologic-information network is comprised of two parts: (1) the set of stations at which hydrologic data are collected and (2) a mechanism for transferring information from the gaged sites to ungaged sites when the need arises. If saturation gaging (collecting lengthy records at every conceivable source of information demand) were economically feasible, the information transfer mechanism would be superfluous; however, this is seldom, if ever, the case in the real world.

One procedure that has been proposed as a transfer mechanism for streamflow information is logarithmic regression analysis (Benson and Matalas, 1967), which results in a relation of the form

$$X = b_0 Y_1^{b_1} Y_2^{b_2} \ldots Y_k^{b_k} \qquad (1)$$

in which X is an estimate of a statistical parameter of streamflow, such as the annual mean or standard deviation, $Y_1, Y_2, \ldots, Y_k$ are physiographic parameters that describe the drainage basin upstream from the site for which the estimate is being made, and $b_0, b_1, b_2, \ldots, b_k$ are coefficients that are defined by the regression analysis performed using data collected at existing gaging stations. As more streamflow information is collected in the region, the coefficients obtained in the regression analysis will approach their optimum values, and therefore equation 1 will tend to become a better predictor of the streamflow parameters at ungaged sites. A measure of the predictive accuracy of this regional information-transfer mechanism is given by S_0, the observed standard error of estimate associated with the regional regression analysis. S_0 is only an estimate of the true accuracy of the regression estimates obtained by using equation 1 and hence is statistical in character. If another set of gaging stations and(or) another period of years of record were used, it is highly probable that another value of S_0 would be obtained.

The true level of predictive accuracy may differ substantially from the value of the observed S_0, and hence, S_0 may be misleading. This devious behavior of S_0 can be attributed to the fact that the values of X used in the regression analysis are only estimates of the streamflow parameters, whose true values are denoted as θ. The discrepancies between the X's and the θ's, illustrated in figure 1, are caused by the time-sampling errors contained in the finite lengths of the streamflow records that were used to compute the X's. The statistical nature of the time-sampling errors may be such that it can cause S_0 to either underestimate or overestimate the true predictive error, S_T, of the regional regression relation. S_0 is simply a measure of how well the data points, X, fit the given model. The true standard error of estimate is the root-mean-square error of the difference between the values estimated from equation 1 and the true values, θ, averaged over all possible values of θ, X, Y_1, Y_2, $\ldots$, Y_k for a fixed set of values for b_0, b_1, b_2, $\ldots$, b_k.

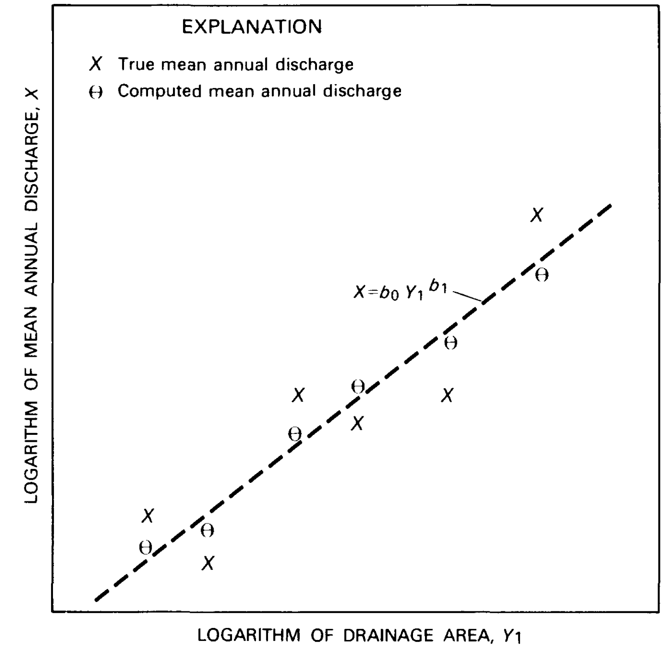


Figure 1. Example of streamflow regression with uncorrelated time-sampling errors.

In figure 1, the regression line is derived from the X's that represent the estimates of mean annual discharge derived from streamflow records. The value of S_0 pertaining to this regression is derived from the differences between the X values and their vertical projections onto the regression line. Associated with each X in figure 1 is the true value of mean annual discharge represented by the symbol θ. In the example, it can be seen that each θ is closer to the regression line than its associated X. If the set of θ's is representative of the total population of possible θ's, then S_0 could be approximated by taking deviations between the θ's and their vertical projections onto the regression line. These deviations, being much smaller than those of the X's, would result in an S_T that is smaller than S_0; thus the observable quantity, S_0, is an example of overestimation of the error contained in estimates of mean annual discharge derived from a regression equation.

If the streamflow records for the sites represented in figure 1 were lengthened, the X's would tend to converge to the value of θ. Because the θ's are above the regression line on its lower end and below it on its upper end, the added data would tend to cause the regression line to rotate clockwise, increasing b_0 and decreasing b_1. The values of S_0 and S_T would also tend to converge as more streamflow data are collected.

By defining the statistical character of both an observed standard error of estimate and true standard error of estimate as a function of the number of stations and the average length of record used in a regional regression analysis, the ability of a particular regression model to extract regional streamflow information from various regional streamflow data bases can be described. This ability provides a means to construct the information-cost curve for a streamflow parameter, where cost is measured both in terms of funds and in terms of time.

The information-cost curve derived in this manner has two noteworthy characteristics. First, it is statistical in nature because perfect information is not available for its definition. Thus future increments of information with increases in data must also be viewed in a statistical sense. On the average, increases in data result in increases in information, but in any individual real-world situation, the estimate of the information content of a data set may even decrease as data are added.

The second characteristic of the information-cost curve is that it is constrained by the regression model. Unless the regression model is truly a perfect model of the hydrologic region and its independent variables are measured with infinite accuracy and precision, the total amount of information that can be generated by collecting streamflow data is limited. This limit will prevent the design of networks with a least-cost criterion if the information requirement of the design exceeds the capabilities of the regression model. On the other hand, a design that maximizes the amount of information generated from a given level of funding is always possible.

THE STATISTICAL NATURE OF STANDARD ERRORS OF ESTIMATE

Not only are all the statistical assumptions underlying classical regression analysis almost always violated in the regional regression analysis as applied in hydrology but also the model assumptions embodied in equation 1 seldom, if ever, completely specify the relations existing between the hydrologic parameter X and the basin parameters, Y_1, Y_2, . . . Y_k. Hence the frequency distribution of S_0 is not known, and S_0 may not be the appropriate measure of predictive accuracy as discussed above. One measure of the goodness of the model is given by the model error, γ, which is the root-mean-square error of prediction of the regression relation if the relation is calibrated with an infinite number of years of streamflow records at each of an infinite number of stations. Model error measures the validity of the model in the absence of any error in the definition of its parameters. Obviously, the data base is not infinite in either the spatial or the temporal dimensions, and thus model error cannot be measured directly. Nevertheless, consideration of this error is essential to the probabilistic description of the predictive accuracy of the regional regression function and therefore is a controlling factor of the efficiency of the regional network. This can be illustrated by the fact that if model error is sufficiently large, even a massive data collection program may yield little or no improvement in the estimates at ungaged sites. In general, nothing is known about the magnitude of model error except for the information that can be distilled from some apparent measure of the accuracy of the regression such as S_0, the observed standard error of estimate. The distillation mechanism is Bayes' rule, one of the fundamental relations of probability theory that will be discussed in the following section.

The problem of describing the probabilistic nature of the measures of predictive accuracy of a regional regression is further complicated by the fact that model error is related to other parameters as well as to the observed standard error of estimate. A finite set of such parameters that appears to be the major influencing factor on the probabilistic nature of the standard errors of estimate and the model error has been set forth by Moss and Karlinger (1974). These additional parameters are (1) *NB*, the effective number of streamflow stations incorporated into the regression analysis; (2) *NY*, the harmonic mean record length of the streamflow records at the *NB* sites; (3) C_v, the average coefficient of variation of the streamflow records; and (4) ϱ_c, the effective cross correlation between pairs of streamflow records. Of these additional parameters, the latter two, C_v and ϱ_c, are known only with a limited degree of accuracy. Their levels of uncertainty are introduced into the analysis by assigning probabilities to their values and including these probabilities in the Bayesian analysis described in the next section.

The predictive accuracy of a particular sample regression equation, which is a measure of the effectiveness of the network, cannot be determined exactly but can be described probabilistically. If the values of γ, C_v, ϱ_c, *NB*, and *NY* are known, the probability distribution of the predictive accuracy that might result from a regression analysis can be determined (Moss and Karlinger, 1974).

BAYESIAN ANALYSIS

One means by which uncertainties can be handled in a decision-making context is Bayesian analysis—so named because of the frequent use of Bayes' rule (Raiffa, 1970 p. 17–18). Unknown or uncertain quantities are considered to be random variables, and the values that they may attain are described probabilistically. The analyst uses his best knowledge concerning the uncertain quantities to assign them initial probabilities. These probabilities, known as prior probabilities, may be based on purely subjective feelings or on a combination of feelings, hard facts, and data. The development of subjective prior probabilities is described by Raiffa (1970, p. 161–168).

In many instances, the analyst may have little or no prior information about the magnitude of an uncertain quantity. Prior probabilities that describe such a state are known as diffuse, or noninformative, priors. A truly noninformative prior is a function of the parameter or parameters that are uncertain and of the family of probability distributions that underlie the parameters. For example, the noninformative prior for the mean of a normal (Gaussian) distribution is that any value is as likely as any other value; that is, the probabilities are uniformly distributed over all possible values (Box and Tiao, 1973, p. 27). In the procedures that follow, the noninformative prior is used as the basis for all analyses. However, the analyst can override the noninformative prior if information is available.

After the prior probabilities are set, additional information concerning some or all of the uncertain quantities may become available. Bayes' rule provides an objective method by which new information is incorporated into the probabilistic description of the uncertain quantities. As an example of the use of Bayes' rule, consider a case of the joint frequency distribution of rainfall and runoff for a specific basin. Let A_1 denote the event in which the total rainfall during a given time period was less than or equal to one precipitation unit. Let A_2 denote the event in which the total rainfall during the same time period was greater than one unit. Let B_1 denote the event in which the peak discharge at a point in the basin affected by the rainfall was less than or equal to 500 streamflow units, while B_2 is the event in which the total discharge at the same point exceeded 500 units. Suppose that from past records or from personal judgments, estimates have been made of the probabilities of either discharge event occurring given that one of the rainfall events occurred in a specific preceding time period. Such a probability is denoted by $P(B_1|A_2)$, which is read as "the probability that event B_1 occurred given that event A_2 occurred" or "the probability that a discharge of less than 500 units occurred conditioned on the fact that precipitation in excess of one unit occurred." On the basis of past records, the unconditional probabilities, $P(A_1)$ and $P(A_2)$, of events A_1 and A_2 occurring during a specified time period have also been estimated. Two basic relationships of probability theory are fundamental to understanding Bayes' rule. First, the unconditional probability of an event, for example B_2, can be written in terms of a weighted sum of conditional probability as shown in equation 2:

$$P(B_2) = P(B_2|A_1)\,P(A_1) + P(B_2|A_2)\,P(A_2) \qquad (2)$$

where events A_1 and A_2 may not occur simultaneously, but one of them must occur. Second, the probability $P(B_2,A_2)$ of the joint occurrence of the two events, B_2 and A_2, can be written as

$$P(B_2,A_2) = P(B_2|A_2)\,P(A_2) \qquad (3)$$

or

$$P(B_2,A_2) = P(A_2|B_2)\,P(B_2). \qquad (4)$$

Equating the right-hand sides of equations 3 and 4 and solving for $P(A_2|B_2)$ yields Bayes' rule,

$$P(A_2|B_2) = \frac{P(B_2|A_2)\,P(A_2)}{P(B_2)}. \qquad (5)$$

The value of $P(B_2)$ is found by using equation 2.

Suppose that for a certain basin the following probability estimates have been obtained

$P(B_1|A_1) = 0.8$	$P(B_1|A_2) = 0.3$
$P(B_2|A_1) = 0.2$	$P(B_2|A_2) = 0.7$
$P(A_1) = 0.6$	$P(A_2) = 0.4$

Substituting appropriate values into equation 2, we have

$$P(B_2) = (0.2)(0.6) + (0.7)(0.4) = 0.40,$$

and equation 5 gives

$$P(A_2|B_2) = \frac{(0.7)(0.4)}{(0.40)} = 0.7.$$

Thus, knowledge that the runoff event, B_2, has occurred has increased the probability that the rainfall event, A_2, had occurred from 0.4 to 0.7.

This rainfall-runoff example could be extended to a case of a finer partition of the possible values of rainfall and runoff. The equations remain the same except that equation 2 would include more terms in the summation.

Associated with Bayes' rule as formulated in equation 5 is the following terminology. $P(B_2|A_2)$ is the "likelihood" of the event B_2 having occurred if A_2 actually occurred. $P(A_2)$ is the "prior probability" of the event A_2. $P(A_2|B_2)$ is known as the "posterior probability" of the event A_2 after B_2 has occurred.

DISTRIBUTIONS OF MEASURES OF PREDICTIVE ACCURACY OF A REGRESSION

The probability distributions of the observed standard error of estimate and the true standard error of estimate are needed as functions of *NB, NY*, γ, C_v and ϱ_c to implement the NARI technique for hydrologic network design. Direct analytical derivation of these distributions from knowledge and reasonable assumptions about their casual factors has proved intractable. Monte Carlo simulation, which exploits the relative-frequency concept of probability that is familiar to hydrologists in the form of flood-frequency analysis, does provide a method for their definition. In a Monte Carlo simulation, a set of random numbers is used as input to a model of the mathematical or physical system being studied, and the computer model of the system processes the random numbers to derive an outcome or measure of the experiment. The experiment is then repeated a sufficient number of times with independent sets of random numbers so that the outcomes can be ordered (in the same way as flood peaks in a frequency analysis) and a probability distribution can be estimated.

The model used to derive the two probability distributions has two major parts: a simulator of simple regression analysis and a multisite synthetic streamflow generator. The regression simulator assumes an underlying regression of the form

$$\ln X_i = a + b \ln Y_i + \epsilon_i \tag{6}$$

where X_i is the value at station i of the streamflow characteristic that is being regionalized, Y_i is a surrogate for the basin physiographic and climatic characteristics at station i, ϵ_i is a random component with zero mean and variance equal to γ^2, and a and b are known coefficients. Equation 6 is a logarithmic transformation of the simple case of equation 1. The values of Y_i are assumed to fall randomly between a lower and upper limit. Selection of *NB* random values of Y_i and ϵ_i permits the evaluation of X_i at each of the *NB* hypothetical gaging stations. If X_i represents mean streamflow, the assumption of a constant coefficient of variation, C_v, within the area of interest specifies a standard deviation, σ_i, for each station,

$$\sigma_i = X_i C_v. \tag{7}$$

The further assumptions that streamflows are lognormally distributed with two parameters (mean and standard deviation), that interstation correlation, ϱ_c, is a constant between each pair of stations, and that serial correlation is insignificant provide the remaining information that is required for the synthetic streamflow generator.

The synthetic streamflow generator (Fiering and Jackson, 1971) is simply an algorithm that converts random numbers into a sequence of synthetic data that maintain a statistical similarity to a set of input statistics such as mean, standard deviation, and cross and serial correlations. These synthetic data can be used in the same manner as actual streamflow records to compute statistics or to design projects. In the NARI technique the synthetic streamflows are used to compute a set of estimates, $\tilde{X}_i$, of the streamflow characteristics, X_i, that then can be returned to the regression simulator for use as dependent variables in a regression analysis. The estimates, $\tilde{X}_i$, are based on a sequence of data of length *NY* at each station, that is, *NY* years of record.

The regression simulator performs the regression analysis by using the estimates, $\tilde{X}_i$, and derives estimates of and a and b and a value of the observed standard error of estimate S_0. Given the properties of the underlying regression and the estimates of a and b, it is possible to compute a value of the true standard error of estimate, S_T, a measure of the predictive ability of the regression relation. The pair of values S_0 and S_T summarizes the first experiment. Sufficient repetition of the experiment with fixed values of γ, C_v, ϱ_c, *NB* and *NY* provides pairs of S_0 and S_T to estimate their probability distributions within a given degree of accuracy. In the NARI technique a minimum of 2,500 repetitions was used for each evaluation of $P(S|\gamma,C_v,\varrho_c,NB,NY)$ and $P(S_T|\gamma,C_v,\varrho_c,NB,NY)$. By repeating the above exercise for representative values of γ, C_v, ϱ_c, *NB*, and *NY* over the relevant ranges, enough probability distributions have been generated so that interpolation schemes could be devised to estimate non-generated probability distributions for a large range of values of γ, C_v, ϱ_c, *NB*, and *NY*.

IMPLICATIONS OF MODEL ASSUMPTIONS

Several of the assumptions used to develop the probability distributions for NARI may not be fully consistent with the real world for any particular application. This section of the report discusses the implications of these assumptions and of their violation.

In the procedure used to develop the probability distributions, the simulated regressions were performed on data that were homoscedastic; that is, the variance of the residuals in the regression did not depend on the magnitudes of the associated independent variables.

Homoscedasticity was assured by three steps: (1) the model error did not vary with the magnitude of the independent variable, (2) the coefficient of variation of the streamflows was constant for all sites and thus did not vary with the independent variable, and (3) the length of record was assumed to be constant at each gaged site. Any step or combination of these three steps may be invalidated to some degree in the real world. For example, the coefficient of variation tends to decrease as one moves downstream in a basin because the flow is an integration of all upstream factors and these factors are not perfectly correlated. Thus, a logarithmic regression of a flow variable against drainage area will tend to show residuals of larger absolute value for the smaller sites. This facet also may be compounded by a gaging history in which the larger basins generally have a longer period of record than the smaller ones. The effect of this lack of homoscedasticity is that the level of information available will vary with the magnitude of the independent variable. In the above example, more is known about large basins than is known about smaller ones. To alleviate this problem, separate analyses may have to be performed on various partitions of the available data set.

A second assumption that is undoubtedly violated in the real world is that of a constant correlation coefficient between any pair of streamflow records. The correlation between two records is controlled by the similarity of the drainage basins and by the correspondence of their inputs of precipitation. Thus, records that come from sites that are closer together usually are more highly correlated than the records of those that are farther apart. Also sites on the same stream generally are more correlated than those on separate streams. The variability of individual correlation coefficients is treated in NARI by averaging the observed correlations. This averaging accounts for the redundancy in the existing streamflow records. However, if large increases in the number of gages are being analyzed, the average of the existing correlations will tend to underestimate the future level of data redundancy and thereby tend to overestimate the level of regional streamflow information. This overestimate could be modified by increasing the correlation with the areal density of the network, but an objective means to do so has not yet been developed.

Finally, the NARI synthetic streamflow generators and the frequency analyses were based on lognormal probability distributions. Thus, the questions of the lack of knowledge of the true underlying distributions (Wallis, Matalas, and Slack 1976) and the possible effects of unknown higher statistical moments (Wallis, Matalas, Slack 1977) were skirted by fiat. The immediate effect of this step would seem to be an increase in the estimated magnitude of the model error to account for this added variability. Overestimations of model error would cause underestimation of the regional information content of future stream-gaging activities.

The synthesis of the above discussions is that NARI is a first step in the quantitative analysis of regional information networks. It will provide answers that must be weighed in light of the degree of belief in conformance of the system being analyzed with the system that was simulated.

THE NARI PROCEDURE

The NARI strategy for network analysis based on the above considerations was outlined by Moss and Karlinger (1974). Subsequent use of this strategy has resulted in some changes that have been incorporated into this paper. An existing regression analysis for a mean, standard deviation, 2-, 10-, 50-, or 100-year event is necessary to implement the procedure. The design procedure can be partitioned into three parts: (1) definition of level of data availability, (2) definition of prior probability of unknown parameters, and (3) evaluation of current networks and design of a future network.

Definition of Level of Data Availability

The level of data availability is defined by the adjusted number of stations, *NB*, and the harmonic mean record length, *NY*, (Hardison, 1971) of the stations that were used in the regression analysis. The harmonic mean record length is used instead of the arithmetic mean record length because the former is directly related to the time sampling errors contained in the estimates of the streamflow parameters, which are the dependent variables in the regression analysis. Let m_i denote the length of record, in years, available at the ith station for $i = 1, 2, \ldots, n$. Then the harmonic mean record length, *NY*, is defined by

$$1/NY = \frac{1}{n} \sum_{i=1}^{n} (1/m_i) \qquad (8)$$

or

$$NY = (\sum_{i=1}^{n} (m_i n)^{-1})^{-1}. \qquad (9)$$

If more than one independent (explanatory) variable is used in the regression analysis, it is necessary to adjust the number of stations to obtain an effective number of stations. The adjusted number of stations is defined by

$$NB = n - k + 1 \tag{10}$$

in which k is the number of independent variables used in the regression analysis. Note that equation 10 differs from the formula for degrees of freedom because one independent variable is already accounted for in the NARI simulation.

The level of data availability, NY and NB, delimits feasible probability distributions of observed standard error of estimate that will be used as likelihoods in the Bayesian analysis of the uncertainty surrounding ϱ_c, C_v, and γ. The necessary distributions are available for the following ranges of values of NB and NY:

$$10 \leq NB \leq 50 \tag{11}$$

$$10 \leq NY \leq 50. \tag{12}$$

If either the adjusted number of stations or the harmonic mean of the years of record exceeds the above limit of 50, the analyst can ignore some of the available data or intelligently partition the data base of interest into smaller bases so that NB and NY fall in the above ranges.

Definition of Prior Probabilities of the Unknown Parameters

The values of cross correlation, ϱ_c, coefficient of variation, C_v, and model error, γ, are not known with a high degree of certainty, therefore, the effects of their accuracy of estimation must be included in the network analysis. This inclusion is achieved by employing basic Bayesian statistical techniques, which allow the designer to describe the uncertainties in a probabilistic manner.

Uncertainties surrounding the value of serial correlation have been excluded because realistic values of serial correlation, from 0 to 0.3, have been found to have negligible effects on the statistical nature of the standard error of estimate. The technique of network design formulated in this paper assumes a zero value of serial correlation.

Two of the input parameters, C_v and ϱ_c, describe the hydrology of the region under consideration. The population values of C_v and ϱ_c are never known. This uncertainty is expressed in terms of probabilities of possible values of C_v and ϱ_c. The task one is faced with is determining these probabilities. One straightforward and simple way of estimating the probabilities of C_v and ϱ_c is to calculate sample estimates of the C_v at each site and sample estimates of ϱ_c for each pair of sites. Let Q_{ij} for $j = 1, 2, \ldots, m_i$ denote the sequence of recorded values of an annual hydrologic event in the jth year at the ith station.

The necessary sample statistics, mean and covariance, are given by

$$M_i = (\sum_{k=1}^{m_i} Q_{ik})/m_i \tag{13}$$

and

$$S_{ij}^2 = \sum_{k=1}^{m_{ij}} (Q_{ik} - M_i)(Q_{jk} - M_j)/(m_{ij} - 1) \tag{14}$$

where m_{ij} is the length of concurrent record that stations i and j have in common. As estimate $\hat{C}_{v_i}$ at site i is evaluated by

$$\hat{C}_{v_i} = S_{ii}/M_i \tag{15}$$

for $i = 1, 2, \ldots, n$. The histogram of values of C_{v_i} can then be used to estimate prior probabilities of occurrence of possible values of C_v.

An estimate $\hat{\varrho}_{ij}$ of the cross correlation, ϱ_c, between flows at site i and site j is given by

$$\varrho_{ij} = (m_{ij} - 1)S_{ij}/((m_i - 1)(m_j - 1)S_{ii}S_{jj})^{1/2}. \tag{16}$$

Again the histogram of $n(n - 1)/2$ values of $\hat{\varrho}_{ij}$ thus obtained can be used to estimate prior probabilities of values of ϱ_c.

The probability of a joint occurrence of a specified value of ϱ_c and C_v is estimated by assuming independence of the C_{v_i}'s and the ϱ_{ij}'s and multiplying their probabilities.

This method, which is called the method of moments, for obtaining probabilities to associate with C_v and ϱ_c is good in that it makes no assumptions about the distribution of flows. However, this method has some undesirable properties. First, the estimate, C_{v_i}, of the C_v at site i cannot exceed the value $(m_i - 1)^{1/2}$ (Kirby, 1974). This upper bound results from the fact that streamflow magnitudes cannot be less than zero. This boundedness causes problems when one is investigating the likelihood of possible values of the true C_v. Research into the sampling properties of this moment estimator of C_v has been carried out (Slack, Wallis and Matalis, 1976).

Another problem arises from considering the estimates of C_v and ϱ_c to be independent.

One method of circumventing these troublesome sampling problems is to use an assumption of multivariate lognormality of the flows at the sites. The synthetic streamflow generating scheme underlying the NARI method rests on this assumption, so it seems reasonable to make use of it in estimating the probabilities associated with C_v and ϱ_c values. Under this assumption, certain relationships exist among the parameters of the probability distribution of the annual hydrologic event of interest and the parameters of the probability distribution of the logarithm of the event. Let

$$W_{ij} = \ln Q_{ij} \tag{17}$$

denote the natural logarithm of the streamflow Q_{ij} in the jth year at the ith station. The W_{ij} are then distributed as multivariate normal. The assumption of a constant C_v and ϱ_c for the Q_{ij} implies a constant variance σ_w^2 and cross correlation τ for the logarithms of the flows. The relations between C_v and ϱ_c and σ_w^2 and τ are given by Matalas (1967)

$$C_v^2 = e^{\sigma_w^2} - 1 \tag{18}$$

$$\varrho_c = (e^{\tau\sigma_w^2} - 1)/C_v^2. \tag{19}$$

From these two equations it is obvious that estimates of σ_w^2 and τ for the logarithms will lead to logical estimates of C_v and ϱ_c. Furthermore, probability statements about possible values of C_v and ϱ_c can be derived from probability statements about corresponding values of σ_w^2 and τ. Hence the task of estimating probabilities of certain C_v and ϱ_c values is reduced to estimating probabilities of values of σ_w^2 and τ. This latter problem has been solved in the context of Bayesian analysis by Geisser (1964), who gives the joint likelihood function, $g(\hat{\sigma}_w^2, \tau \mid \hat{\sigma}_w^2, \tau)$ where $\hat{\sigma}_w^2$ is an average of the at-site sample variances of the logarithms and $\hat{\tau}$ is an average of sample estimates of the cross correlation, τ, between the logarithm of flows at all pairs of sites. Using this likelihood function along with noninformative prior probability density functions on σ_w^2 and τ, gives a posterior probability density function of σ_w^2 and τ, denoted $f(\sigma_w^2, \tau \mid \hat{\sigma}_w^2, \hat{\tau})$. This posterior probability density function is conditioned on n and NY along with the statistics $\hat{\sigma}_w^2$ and $\hat{\tau}$. Joint probabilities of C_v and ϱ_c being in stated intervals can be evaluated using the identities, equation 18 and equation 19, along with the joint probability density function $f(\sigma_w^2, \tau \mid \hat{\sigma}_w^2, \hat{\tau})$.

In order that the NARI technique be applicable to regions in which there is a nonzero probability of no flow for a period, the values of C_v associated with probabilities estimated by the above Bayesian analysis must be adjusted for the effect of these zero flows. Let p denote the probability of zero flow. Let C_{vc} denote the coefficient of variation conditioned on the flow being greater than zero. Then the unconditional C_v, adjusted for zero flow, is given by

$$C_v = \left(\frac{C_{vc}^2 + p}{1 - p}\right)^{1/2}. \tag{20}$$

Bayesian description of the uncertainty in values of C_v and ϱ_c embodied in the joint posterior probability density function $f(\sigma_w^2, \tau \mid \hat{\sigma}_w^2, \hat{\tau})$ and the transformations given by equations 18 and 19 allow for continuous variation of C_v over all positive numbers and ϱ_c over the interval (0,1). However, because of financial limitations, the necessary probability distributions of S_0, the observed standard error, and S_T, the true standard error, have been evaluated only for C_v values between 0.1 and 5.0 and ϱ_c values in the interval (0,0.9). The NARI technique is constrained to a discrete description of the uncertainty so that it is necessary to divide the range of possible values of C_v and ϱ_c into a finite number of intervals and to use the mid-interval value as a surrogate for all values contained in the interval.

Because studies to date have not been able to discriminate model error, γ, from the other errors contained in regression analysis, very little knowledge is available concerning its value. Model error as used in this network design technique is defined to be a standard deviation and, therefore, must be a nonnegative number. Little else is known about it. Such a state is known as a diffuse prior (Benjamin and Cornell, 1970), and the unknown variable, γ, can be assumed to have a uniform (equally likely) probability distribution within a limited range in which its likelihood is significantly different from zero. In the NARI technique this locally uniform range of γ is assumed to have a lower bound of zero. This discretization of q is initiated by considering the interval 0.0–0.025. Subsequent intervals of width 0.05 are processed until a value of γ is reached for which no significant effect on the implementation of Bayes' rule is encountered. Sensitivity studies have shown that discretizations finer than intervals of 0.05 do not change the result. Because the width of the locally uniform prior is known at the beginning of the processing, the true probability associated with any interval is also unknown initially. It is fortunate, however, that Bayes' rule can be used with values for the priors that are relative weights as opposed to strict probability measures. This condition makes it possible to use nonzero priors for model error and thereby arrive at nonzero posterior probabilities and a solution to the design problem.

The relative weight attached to the first interval is 0.5 and for all other intervals is 1.0. The half weight is caused by the absence of probability on the negative side of zero. By using these weights the resulting solution will define the proper lack of prior knowledge concerning model error even though its priors are not in reality measures of probability.

It is assumed in the use of this network design technique that—although the estimates of the coefficient of variation and cross correlation are dependent on one another—the estimates of the model error are independent of the values of C_v and ϱ_c. This assumption permits the computation of the joint prior probability of a set of input parameter values by multiplication of the prior on model error by the joint prior on C_v and ϱ_c. For example, if the joint prior probability that $C_v = 0.5$ and $\varrho_c = 0.3$ is one fourth and the prior "probability" that $\gamma = 0.0$ is one half, then the joint prior "probability" of ($C_v = 0.5$, $\varrho_c = 0.3$, $\gamma = 0.0$) is the product $(0.25)(0.5) = 0.125$.

Evaluation of the Current Network and Design of a Future Network

Even if all uncertainty concerning the parameters C_v, ϱ_c, and γ were removed, the value of the true standard error, S_T, the measure of the predictive accuracy of a given network design, still could not be determined exactly. However, probability statements concerning possible ranges of values of S_T could be made by using the generated conditional frequency distributions, $P(S_T | C_v, \varrho_c, \gamma, NB, NY)$, discussed in a previous section. Some risk would still be involved in the design based on this conditional frequency distribution because the value of S_T is not directly determinable and is statistical in character. Although C_v, ϱ_c, and γ cannot be determined exactly, Bayes' rule can be used to make certain probability statements about possible values (Moss and Karlinger, 1974).

The basis of the design methodology employed in the NARI technique can be separated into two fundamental operations: (1) determining those values of the parameters C_v, ϱ_c, and γ that, with some degree of belief, could have resulted in the observed standard error of estimate, S_0, and determining probability weights to associate with values in this set, and (2) using these weights, determining an averaged probability distribution of S_T, the true standard error of estimate of the regression that would result from using various combinations of numbers of stations and lengths of record.

The operation of determining values of C_v, ϱ_c, and γ consists of first considering only those representative values of C_v and ϱ_c that have nonzero prior probabilities, $P(C_v, \varrho_c)$, as described above. The set of allowable γ values is then determined by accepting only those values of γ such that there is a nonzero probability of having observed the standard error of estimate. Hence, only those values of C_v, ϱ_c, and γ are considered feasible for which both the likelihood function, $P(S_0 | C_v, \gamma, NB, NY)$, and the joint prior probability $P(C_v, \varrho_c, \gamma)$ are nonzero. These probabilities are then used in Bayes' rule (see equation 5) to determine posterior probabilities of the particular combination, C_v, ϱ_c, and γ. These posterior probabilities are written as

$$P(C_v,\varrho_c,\gamma | S_0,NB,NY) = \frac{P(S_0 | C_v,\varrho_c,\gamma,NB,NB)\ P(C_v,\varrho_c,\gamma)}{P(S_0 | NB,NY)} \qquad (21)$$

with

$$P(S_0 | NB,NY) = \Sigma P(S_0 | C_v,\varrho_c,\gamma,NB,NY)\ P(C_v,\varrho_c,\gamma) \qquad (22)$$

where the summation in equation 22 is over all feasible values of C_v, ϱ_c, and γ. These posterior probabilities are the desired probability weights. In effect, this states that the probability is zero of the occurrence of any other combination of C_v,ϱ_c, and γ, given the observed standard error of estimate, S_0, along with NB stations and NY years of data.

The posterior probabilities developed above contain all of the available information concerning the unknown values of the parameters C_v, ϱ_c, and γ. Until further data are collected that can be used in Bayes' rule to update these probabilities, they will be used to estimate the probability distribution of the predictive accuracy of a specific network configuration. The feasible values of C_v and ϱ_c describe the hydrology of the region—for the purposes of this network design—while the feasible values of model error, γ, describe the "validity" of the regression model being employed.

To evaluate the present network the averaged probability distribution of the true standard error of the regression is given by

$$P(S_T | S_0,NB,NY) = \Sigma P(S_T | C_v,\varrho_c,\gamma,NB,NY)\ P(C_v,\varrho_c,\gamma | S_0,NB,NY) \qquad (23)$$

where the summation is over the same set of C_v, ϱ_c, γ used in equation 22. Given this averaged frequency distribution of S_T, the true standard error of the regression, statements about the reliability of the regression can be made in terms of the probability that the true accuracy of the regression is less than a stated value S_T^{α}. For example,

$$P(S_T \leq S_T^{\alpha} | S_0, NB, NY) = \alpha \qquad (24)$$

states that the probability of S_T being less than or equal to S_T^{α}, given S_0, NB, and NY, is α, where S_T^{α} is a reference value of standard error and α is the reliability associated with that value. If the analyst can accept the predictive accuracy S_T^{α} possessing reliability α, then the network is satisfactory. The acceptable level of predictive accuracy and the associated reliability must be determined externally to this design program. If the present network is deemed unsatisfactory, then the averaged probability distribution of S_T, given another combination of effective number of stations, NB^*, and number of years of record, NY^*, can be obtained using this same method. The required distribution is given by

$$\begin{aligned} &P(S_T | S_0, NB, NY, NB^*, NY^*) \\ &= \Sigma P(S_T | C_v, \varrho_c, \gamma, NB^*, NY^*) \qquad (25) \\ &P(C_v, \varrho_c, \gamma | S_0, NB, NY) \end{aligned}$$

where again the summation in equation 25 is over all feasible parameter values C_v, ϱ_c, and γ. The ultimate output of the design computer program is a listing of 12 percentage points $\{S_T^{\alpha_i}\}^{12}_{i=1}$ of such averaged distributions for as many design points, NB^* and NY^*, as the analyst wishes so long as

$$10 \leq NB^* \leq 50$$

and

$$10 \leq NY^* \leq 50.$$

The analyst specifies the values of NB^* and NY^* as input to the program, and one of these points should be the current stations-years combination, NB and NY. The values of α_i for which percentage points are determined are

$$\{0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 0.99\}$$

which should enable the analyst to interpolate S_T^{α} for other values of α. Two basic graphs employing these percentage points may be drawn. For a given reliability α, a plot of S_T^{α} for various values of NB^* and NY^* may be made as is done in figure 2. For figure 2, $\alpha = 0.5$, so $S_T^{0.5}$, the median true standard error, is plotted. Through this plot the analyst can determine which combination of number of gaging stations and years of record would achieve a given prediction accuracy level with the stated reliability. For a given predictive accuracy, a plot of the probabilities of achieving such accuracy could be made as is done in figure 3. From this plot the analyst could determine the probability of achieving the given accuracy level with NB^* stations and NY^* years of record. Plots such as these would define regions of achievable goals given budgetary constraints on values of NB^* and NY^*.

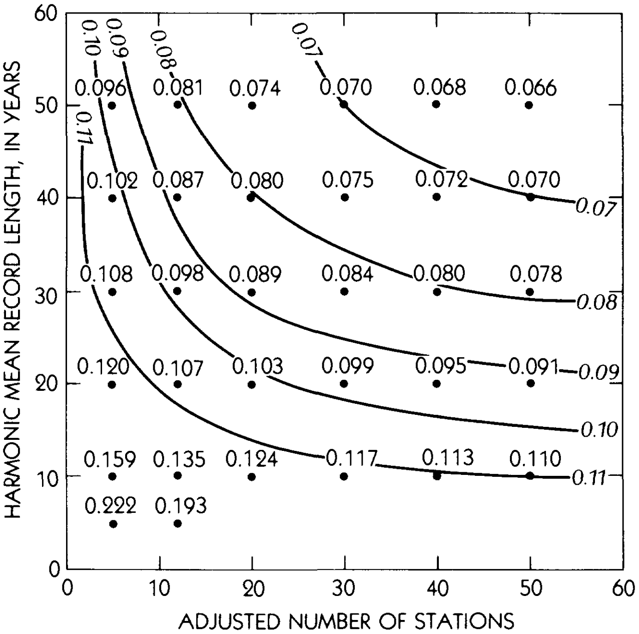


Figure 2. Median true standard error of 50-year event analysis (in natural logarithms).

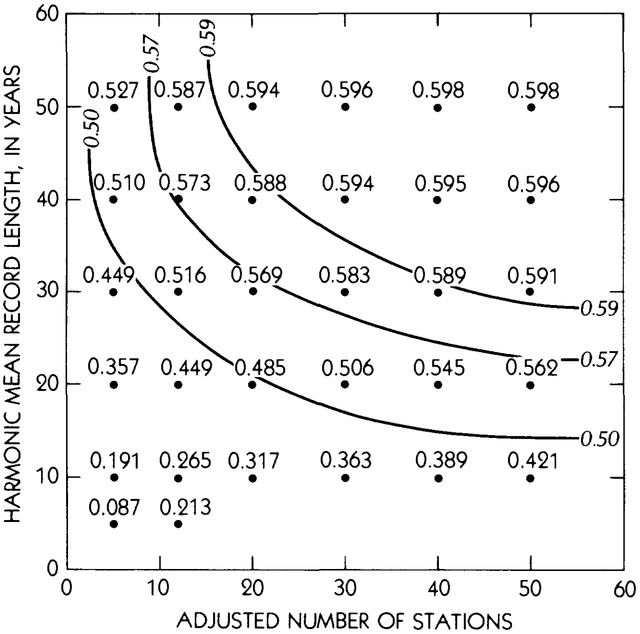


Figure 3. Probability of achieving a given accuracy.

It is possible that the desired goal is not achievable. In such a situation the analyst must reevaluate his criteria—levels of accuracy and(or) reliability—or seek a better information-transfer mechanism (model). Improvement of the model is also a valid means of reducing the required networks even when the information criterion can be met by additional gaging.

USES OF NARI IN NETWORK DESIGN

The end products of the NARI procedure are the plots illustrated by figures 2 and 3. Their use can be demonstrated in three different formulations of the design of streamflow-data networks: (1) a minimum cost network that will attain a prespecified accuracy and reliability, (2) a network that maximizes information (accuracy), given a set of budgetary and time constraints, and (3) a network that derives the maximum net benefit from the data-collection program. The first two types of designs are surrogates for the latter, which is not perfected to an operational state at this time. Were it operational, the maximum-net-benefits procedure would make others obsolete.

In most design situations, it is handy to have the plots of accuracy versus the number of stations and length of record expressed in a mathematical form. The relation

$$\log S_T^{\alpha} = a + b_1 NB^{-1} + b_2 NY^{-1} \qquad (26)$$

is a reasonable approximation of the NARI output for a fixed value of α. The coefficients a, b_1, and b_2 are defined by multiple regression using the NARI output as its dependent and independent variables.

An existing network that is used to perform a NARI analysis represents a single level of data availability, *NB* and *NY*. The goal of all three types of design mentioned above is to arrive at a level of improved accuracy. Thus, any proposed network modification represents a relocation on plots such as those in figures 2 and 3. There are numerous ways in which this relocation can take place. For example, suppose that an existing network consisted of 10 stations ($NB = 10$) each of which had 10 years of record ($NY = 10$) and that it was desired to increment the level of data availability to $NB = 20$ and $NY = 20$. Obviously, one way to do this would be to establish 10 new stations and to operate each of them for 20 years while operating the old stations for an additional 10 years. A second approach might call for adding 10 new gages to the network and operating the 20-station network for 17 additional years. Each strategy accomplishes the desired relocation. The first strategy requires 300 station years of data in comparison with 340 for the second. However, the first strategy attained the desired level after 20 years, while the second arrived at the goal after 17 years. Thus, there are different costs associated with the two strategies and different times at which the goal is attained. Unfortunately, the strategy with the lower cost does not have the shorter time of attainment, and therefore neither strategy clearly dominates the other. The choice of the stratey will depend on the type of objective and the constraints that exist in the formulation of the design of the network.

Minimum-Cost Design

For the minimum-cost network, the designer has an objective of minimizing the costs of collecting the data while attaining a prespecified level of regional accuracy. There may also be additional constraints such as the following: (1) the goal must be attained by a specific date, (2) the expenditure of funds must be as uniform from year to year as is practicable, or (3) the operation of each gage must be as continuous as is possible.

The first step in the design of minimum-cost networks is to define a set of values for *NB* and *NY* that will be candidates for the final design. To be a candidate the network must have the minimum value of *NY* that satisfies the accuracy criterion for a given value of *NB*. In other words, the combination of values of *NB* and *NY* must fall on the contour defined by the prescribed level of accuracy and reliability, S_T^{α}. This set may be defined graphically or by solution of equation 26 for *NY* given a set of values for *NB*.

Often, a network is established to provide information about more than one streamflow parameter. In such a situation, the set of candidate networks consists of those designs for which *NY* and *NB* satisfy or exceed the accuracy criterion for each parameter of interest. Figure 4 illustrates the candidates for a two-parameter design.

The second step in the design is to define the feasible strategies for each of the candidate designs. This step entails the specification of each strategy that does not violate any additional constraints that are pertinent to the problem. The constraints thus limit the number of strategies that must be considered for each candidate. In fact, some candidates may be eliminated completely at this step because all of their strategies violate at least one of the constraints.

The third step is comprised of ascertaining the minimum-cost, feasible strategy for each of the candidate designs. The final step is to choose the design that

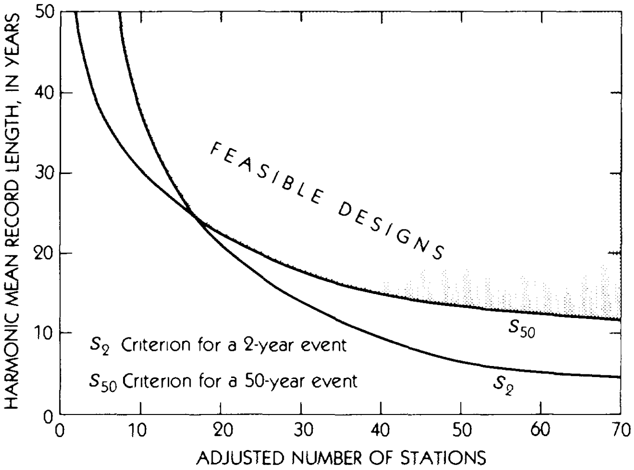


Figure 4. Feasible combinations of *NB* and *NY* for a two-parameter design.

has the minimum cost and to implement the strategy that is associated with this cost.

Maximum-Information Design

Because the arbitrary accuracy criterion required for minimum-cost designs is usually very difficult to defend, the maximum-information network is frequently the design procedure that is used for regional information networks. In this type of design the planning period is specified, and the budget is allocated for each year within the planning period. The objective is to maximize the regional information at the end of the planning period. Maximizing information is equivalent to minimizing the error of prediction associated with the streamflow-parameter estimates. Because each streamflow parameter will have a different maximum-information network, multiple-parameter design by this procedure becomes very difficult.

The procedure for maximum-information design is initiated by defining, for each of a set of values for *NB*, the maximum value of *NY* that may be attained while not exceeding the budget and any other constraints that may be in effect. For each value of *NB* a full array of strategies must be considered if the true maximum *NY* is to be located.

The second step in the procedure consists simply of searching the pairs of *NB* and *NY* values to locate the level of data availability that yields the lowest value of S_T^α (the maximum level of information). The strategy associated with this optimum pair should then be implemented.

Maximum-Net-Benefit Design

Although not currently perfected to the operational state, the maximum-net-benefit design procedure has been illustrated by Karlinger (1975), Moss (1976), and Attanasi and Karlinger (1977). This procedure is very involved, and the interested reader is directed to the above references. In summary, the actual monetary benefits that can be expected from improved decisions that are data dependent are used in conjunction with the cost of collecting the data to determine the net benefit that can be derived from various data networks. The costs include those incurred by delaying the decisions in order to collect the data. The objective of this procedure is to select that design that yields the maximum net benefit.

CATALOGED PROCEDURES

Procedures for the computations described above have been programed and cataloged for automatic computation on International Business Machines 370/155 computers operated by the USGS. The design procedure assumes that an existing regression analysis for a mean, standard deviation, 2-, 10-, 50-, or 100-year event is available and a standard error of estimate, in natural or common log units, has been computed.

Cataloged procedures BBPEAK and BBFLOW are used to retrieve data from the USGS peak-flow file or daily-values file, respectively, for input into cataloged procedure BBPOSPRI. Cataloged procedure BBVOLS formats low-flow data for entry into BBPOSPRI. BBPOSPRI computes joint probabilities for C_v and ϱ_c, computes an estimate of the probability of zero flows, adjusts the C_v's for zero flows, and computes the harmonic mean record length. The joint probability table produced by BBPOSPRI is intended to aid the user in the development of subjective prior probabilities of C_v and ϱ_c.

Prior probabilities of C_v and ϱ_c, harmonic record length, adjusted number of basins, and the standard error of estimate are input to cataloged procedure MODLVALU, which makes the computations for the evaluation of a network. In the following sections, input data for each of these cataloged procedures are described, an example illustrates their use.

BBPEAK

BBPEAK is a cataloged procedure that retrieves all available annual flood data from the USGS Annual Peak Flow File for a specified set of gaging stations within a specified time period. Input to the procedure consists of data cards that specify that the data are annual floods, a user-supplied title, the beginning and ending years of the time period, and the USGS station numbers. Station numbers, of which there must be from 3 to 70, should be in ascending order.

Five programs are used in this procedure. The printout from each program contains the following:

1. J663—Type of data retrieved, and numbers for the stations requested for retrieval.
2. SORT—Station numbers sorted in ascending order.
3. J664—No printout.
4. G745—Prints out the station numbers stating which ones have not been found in the Station Header File.
5. J980—Prints in the brief vector format the annual flood records retrieved from the Annual Peak Flow File.
6. J666—Periods of record requested and obtained, number of stations requested and obtained, station numbers for those successfully retrieved, warning messages (number of stations not retrieved, inadequate period of record, too few or too many stations requested), and listing of data actually retrieved beginning with water year INYR.

The program creates an output file on magnetic disk that contains the following: data type, title, first and last years of period of record, the number of stations, the station numbers, and the annual peaks for the period of record requested (may contain the null value −1) for each station. The disk file can be used by the cataloged procedure BBPOSPRI.

Job Control Language—The following cards are required to execute BBPEAK: (△ indicates a space and brackets indicate optional parameter fields. Do not punch the brackets.)

```
/ /xxxxxxxx△JOB△(-------)
/ /PROCLIB△DD△DSN = WRD.PROCLIB,DISP = SHR
/ /△EXEC△BBPEAK,MUNIT=unit,MVOL=volume,MVNAME='whatever' [,MDISP=disp]
/ /SYSIN△DD△*
      Data-type card
      User-supplied title
      Period-of-record card
      Station number cards
/*
Other job steps if needed
/*
/ /
$$$
```

The parameters of the catalog procedure are

MUNIT—The unit onto which the output file is to be written.

MUNIT = ONLINE for creation of permanent data set on the online disk. (If ONLINE is used, the MDISP parameter should be used and set = CATLG.)

MUNIT = SYSDK for creation of a temporary data set. If this option is used procedure BBPOSPRI must be executed in the same job.

MUNIT = xxxx is used for creation of a data set file on a private disk pack.

MVOL—The volume serial number of the disk containing the file. If MUNIT = ONLINE or MUNIT = SYSDK, MVOL is not used. If MUNIT = xxxx is used, MVOL = name of the particular private disk pack.

MVNAME—Data set name of the file. If MUNIT = SYSDK, then MVNAME = '&&name'. If MUNIT = ONLINE, follow data set naming conventions described in the USGS Computer User's Manual.

Input—The input consists of data-type, title, and period-of-record cards followed by 3-70 station-number cards.

Card 1

The data-type card indicates that the cataloged procedure is to retrieve data from the Annual Peak Flow File. The format is as follows:

Column	Format	Contents
1-4	4A1	Enter PEAK
5-8		Blank

Card 2

The user-supplied title card may contain up to 30 characters, which can be used to describe the region to be analyzed. Entries on this card must be left justified. The following format is used:

Column	Format	Contents
1-30	30A1	Any description of from 1 to 30 characters chosen by user. Left justified.
31-80		Blank

Card 3

The period-of-record card specifies the range of water years for which data are to be retrieved. It has the following form:

Column	Format	Variable	Contents
1-4	I4	INYR	The initial water year for which data are retrieved
5-8	I4	LAYR	The last water year for which data are retrieved
9-80			Blank

Card 4

The station-number cards give the station numbers. Each card contains one station number. The cards should be assembled in increasing order of station numbers. The format of each card is

Column	Format	Variable	Contents
1	1X		Blank
2-9	I8	CARD (J)	The *j*th station downstream order number.
10-80			Blank

Error Messages—The following messages can be generated by program J666 in BBPEAK.

1. 'NUMBER OF STATIONS = ( ) > 70.'—The number of stations for which data are requested exceed the maximum allowable, 70. The job is aborted with a condition code of 16.

2. "LENGTH OF RECORD EXCEEDS THE MAXIMUM OF 75 YEARS.'—Record length requested is in excess of 75 years; that is, LAYR-INYR + 1 > 75. The job is aborted with a condition code of 16.

3. '( ) IS A DUPLICATE STATION NUMBER. PROCESSING ENDED.'

4. 'NUMBER OF STATIONS = ( ) < 3. JOB ABORTED.'—Number of stations requested is less than 3 required. Job is aborted with a condition code of 16.

5. 'SPECIFIED RECORD LENGTH IS LESS THAN 3 YEARS. JOB ABORTED.'—Record length requested is less than the minimum allowable, 3. Job is aborted with a condition code of 16.

BBFLOW

BBFLOW is a cataloged procedure that retrieves all available streamflow data from the USGS Daily Values Files for a specified set of gaging stations within a specified time period and computes the annual and monthly flows in each water year. Input to the procedure consists of data cards that specify that the data are monthly or annual flows, a user-supplied title, the beginning and ending years of the time period, and the USGS station numbers. Station numbers, of which there must be between 3 and 70, should be in ascending order.

Six programs are in the procedure. The printout from each program contains the following:

1. J663—Type of data retrieved and numbers for the stations requested for retrieval.
2. SORT—Station numbers sorted in ascending order.
3. J664—No printout.
4. G475—Station name, drainage area, gage datum and other information retrieved from WRD Station Header File, and station number for records not found in file.
5. G490—Indicates number of stations and station years of record retrieved.
6. J667—Periods of record requested and obtained, number of stations requested and obtained, station numbers for those successfully retrieved, warning messages (number of stations not retrieved, inadequate period of record, too few or too many stations requested), and listing of data actually retrieved beginning with water year INYR (missing data indicated with −1.).

The program creates an output file on magnetic disk, which contains the following: data type, title, first and last years of period of record, the number of stations, the station numbers, and the annual and monthly flows for the period of record requested (may contain the null value −1.) for each station. The disk file may ultimately be used by the cataloged procedure BBPOSPRI.

Job Control Language—The following cards are required to execute BBFLOW; (△ indicates space and brackets indicate optional parameter fields. Do not punch in brackets.)

```
/ /xxxxxxxx△JOB△(-------)
/*SETUP△△△△tape #/H (ONE SETUP PER TAPE IS REQUIRED)
/ /PROCLIB△DD△DSN = WRD.PROCLIB,DISP = SHR
/ /△EXEC△BBFLOW,MUNIT=unit,MVOL=volume,MVNAME=‘whatever’ SP4=num,
[MDISP=disp]
/ /△AGENCY=USGS,VOL1=tape#1[,TIME1=,VOL2=tape#2,TAPE3=tape#3,
/ /△TAPE4=tape#4,TAPE5=tape#5,TAPE6=tape#6]
/ /SYSIN△DD△*
      Data-type card
      User-supplied title
      Period-of-record card
      Station number cards
/*
Other job steps if needed
/*
/ /
$$$
```

The parameters of the catalog procedure are

MUNIT—The unit onto which the output file is to be written.

MUNIT = ONLINE for creation of permanent data set on the online disk. (If ONLINE is used the MDISP parameter should be used and set = CATLG.)

MUNIT = SYSDK for creation of a temporary data set. If this option is used procedure BBPOSPRI must be executed in the same job.

MUNIT = xxxx is used for creation of a data set file on a private disk pack.

MVOL—The volume serial number of the disk containing the file. If MUNIT = ONLINE or MUNIT = SYSDK, MVOL is not used. If MUNIT = xxxx is used, MVOL = name of the particular private disk pack.

MVNAME—Data set name of the file. If MUNIT = SYSDK, then MVNAME = '&&name'. If MUNIT = ONLINE, follow data set naming conventions described in the USGS Computer User's Manual.

SP4—Space needed to store retrieved station years of record. Set SP4 to the next multiple of 10 larger than $NB \cdot NY/7$.

TIME1—Time needed for execution of program J667. TIME1 not needed if time for program J667 execution is less than 2 minutes.

VOL1—Tape number of historical data required for retrieval.

VOL2—Tape number of historical data if second tape required.

TAPE3—Tape number of historical data if third tape required.

TAPE4—Tape number of historical data if fourth tape is required.

TAPE5—Tape number of historical data if fifth tape required.

TAPE6—Tape number of historical data if sixth tape requird.

Note—At most, six Daily Value Historic Tapes may be used. If more than one tape is required, the tape specifications on the execute card must be in state-code order. The "/*SETUP" card is used only when accessing historical data from tape. The last two water years of data are on disk.

Input—The input consists of data type, title, and period-of-record cards followed by 3–70 station-number cards.

Card 1

The data-type card indicates that the cataloged procedure is to retrieve data from the Daily Values File. The format is as follows:

Column	Format	Contents
1–4	4A1	Enter MEAN
5–80		Blank

Card 2

The user-supplied title card may contain up to 30 characters, which can be used to describe the region to be analyzed. Entries on this card must be left justified. The following format is used:

Column	Format	Contents
1–30	30A1	Any description of 1–30 characters chosen by user. Left justified.
31–80		Blank (NOTE: To suppress printing of monthly and annual means, enter an * in column 1.)

Card 3

The period-of-record card specifies the range of water years for which data are to be retrieved. It has the following form:

Column	Format	Variable	Contents
1–4	I4	INYR	The initial water year for which data are retrieved
5–8	I4	LAYR	The last water year for which data are retrieved
9–80			Blank

Card 4

The station-number cards give the station numbers. Each card contains one station number. The cards should be assembled in increasing order of station numbers. The form of each card is

Column	Format	Variable	Contents
1	1X		Blank
2-9	I8	CARD (J)	The *j*th station downstream order number.
10-80			Blank

Error Messages—The following messages can be generated by program J667 in BBFLOW.

1. 'NUMBER OF STATION = ( ) > 70.'—The number of stations for which data are requested exceed the maximum allowable, 70. The job is aborted with a condition code of 16.

2. "LENGTH OF RECORD EXCEEDS THE MAXIMUM OF 75 YEARS.'—Record length requested is in excess of 75 years; that is, LAYR-INYR +1 > 75. The job is aborted with a condition code of 16.

3. 'NUMBER OF STATIONS = ( ) < 3. JOB ABORTED.'—Number of stations requested is less than 3 required. Job is aborted with a condition code of 16.

4. 'SPECIFIED RECORD LENGTH IS LESS THAN 3 YEARS. JOB ABORTED.'—Record length requested is less than the minimum allowable, 3. Job is aborted with a condition code of 16.

BBVOLS

BBVOLS is a cataloged procedure that rearranges the magnetic file of the annual high- and low-flow statistics created by program A969 (Meeks, 1975) to the format required for input to the cataloged procedure BBPOSPRI. The data set contains the lowest and highest average *N*-consecutive-day mean discharge values for each year of record for each station requested. The values of *N* are 1, 3, 7, 14, 30, 60, 90, 120, and 183 for low flows and the same for high flows except that 15 replaces 14 in the above list. BBVOLS identifies this data set with the following information: data type, title, the first and last years of record retrieved, the number of stations contained in the data set, and the identification number of each station. Periods of missing data are denoted by −1 in the data set.

Job Control Language—The following cards are required to execute BBVOLS; (△ indicates a space and brackets indicate optional parameter fields. Do not punch in brackets.)

```
/ /xxxxxxxx△JOB△(-------)
/ /PROCLIB△DD△DSN = WRD.PROCLIB,DISP = SHR
/ /△EXEC△BBVOLS,BKREC = 'name1.id1.note1',BKU = unit,BKVOL = volume1,
/ /△MVNAME = 'name2.id2.note2',MUNIT = unit2,MVOL = vol2[,MDISP = disp,]
/ /SYSIN△DD△*
    Data-type card
    User-supplied title card
    Period-of-record and stations card
/*
Other job steps if needed
/*
/ /
$$$
```

The parameters of the catalog procedure are

BKREC—The data set name for the data set previously created by program A969.

BKU—The unit on which the data set given by BKREC has been written.

BKVOL—The volume serial number of the disk containing the file BKREC.

MUNIT—The unit onto which the output file is to be written.

MUNIT = ONLINE for creation of permanent data set on the online disk. (If ONLINE is used, the MDISP parameter should be used and set = CATLG.)

MUNIT = SYSDK for creation of a temporary data set. If this option is used procedure BBPOSPRI must be executed in the same job.

MUNIT = xxxx is used for creation of a data set file on a private disk pack.

MVOL—The volume serial number of the disk containing the newly created file. If MUNIT = ONLINE or MUNIT = SYSDK, MVOL is not used. If MUNIT = xxxx is used, MVOL = name of the particular private disk pack.

MVNAME—The data set name of the file to be created by BBVOLS. If MUNIT = SYSDK, MVNAME = '&&name'. If MUNIT = ONLINE, follow data set naming conventions described in the USGS Computer User's Manual.

Input—The input consists of data type card, a title card, and an array-definition card.

Card 1

The data-type card indicates that the cataloged procedure is to operate on data retrieved from the Daily Values File by program A969. The format is as follows:

Column	Format	Contents
1-4	4A1	Enter VOLS
5-80		Blank

Card 2

The user-supplied title card may contain up to 30 characters, which can be used to describe the region to be analyzed. Entries on this card must be left justified. The following format is used:

Column	Format	Contents
1-30	30A1	Any descriptor of 1-30 characters chosen by user. Left justified.
31-80		Blank

Card 3

The period-of-record card and number of stations card specifies the period of record and the number of stations for which data have been retrieved. The following format is used:

Column	Format	Variable	Contents
1-5	I5	INYR	The initial water year for which data have been retrieved
6-10	I5	LAYR	The last water year for which data have been retrieved
11-15	I5	NUMSTA	The number of stations for which data have been retrieved

Error Messages—The following messages can be generated by program J667 in BBFLOW.

1. 'NUMBER OF STATIONS = ( ) > 70.'—The number of stations for which data are requested exceed the maximum allowable, 70. The job is aborted with a condition code of 16.

2. 'LENGTH OF RECORD EXCEEDS THE MAXIMUM OF 75 YEARS.'—Record length requested is in excess of 75 years; that is, LAYR - INYR +1 > 75. The job is aborted with a condition code of 16.

3. 'NUMBER OF STATIONS = ( ) < 3. JOB ABORTED.' Number of stations requested is less than the 3 required. Job is aborted with a condition code of 16.

4. 'SPECIFIED RECORD LENGTH IS LESS THAN 3 YEARS. JOB ABORTED.' Record length requested is less than the minimum allowable, 3. Job is aborted with a condition code of 16.

BBREVISE

The cataloged procedure BBREVISE permits the user to do the following operations to the file created by procedure BBPEAK or BBFLOW:

1. Delete all data for a station.
2. Add data for a new station.
3. Change individual flow values.

It should be noted that BBREVISE cannot expand arrays; that is, the addition of data prior to INYR or subsequent to LAYR in the procedure BBFLOW or BBPEAK cannot be made.

The printout from the program in the procedure will indicate in the heading (data type, title) that a revised data set has been created. A list of stations in the revised data set will be given, stations that have been added or deleted will be identified, and the operations that were requested will be documented.

Prior to revising the data set by program J669, the input cards will be sorted in ascending order according to station number, which is coded to indicate revision operation, card number, and water year.

Job Control Language—The following cards are required to execute BBREVISE:
(△ indicate space)

```
/ /xxxxxxxx△JOB△(----)
/ /PROCLIB△DD△DSN = WRD.PROCLIB,DISP = SHR
/ /△EXEC△BBREVISE,MUNIT=unit,MVOL=number,MVNAME=name
/ /SORT.SORTIN△DD△*
      —Input cards—
/ /
```

The parameters of the catalog procedure are

MUNIT—The unit from which file BBPEAK or BBFLOW is read. The default is 3330.

MVOL—The volume serial number of the unit described above. The number must be identical with the number in the procedure BBPEAK or BBFLOW.

MVNAME—The data set name from BBPEAK or BBFLOW.

Input—The input for BBREVISE consists of one type of card with the following format:

Column	Format	Variable	Contents
1	I1	ICODE=	1, Delete station 2, Add station and its data 3, Change flows
2			Blank
3-10	I8	IS	Downstream order number of station
11			Blank
12	I1	ICD	Card number, 1 or 2. Two cards are required for entering annual and monthly flows. Only card number 1 is required for deletion of a station or for entering annual floods
13-16	I4	IYR	Water year. Blank is a station deletion.
17-72	7F8.0	Q(J)	Flow data, subject to the following rules: a. Blank if station deletion b. Use only columns 17-24 for annual floods c. Enter mean annual flow in columns 17-24 and monthly flows for October through March on card 1 (columns 25-72) d. Enter monthly flows for April through September on card 2 (columns 17-64) e. Blank if no change. Q(J) = −1. to delete value. Q(J) ≥ 0 to change data.

	f. Two cards always required for any changes to annual and monthly flow for a given year
	g. Right justify, punch decimal if required
73-80	Blank

Note—If a new station is being added, ICODE must be 2 for all cards.

Error Message—The following error messages can be generated within the program J669:

1. 'MORE THAN ONE TYPE OF UPDATING CODE SPECIFIED FOR STATION iiiiiii'—The ICODE parameter is 2 on one or more cards *and* 3 on one or more other cards.
2. 'INSUFFICIENT DATA FOR STATION iiiiiii in year yyyy'—A revision operation with ICODE = 2 or 3 for annual and monthly flows does not have cards 1 and 2 for year yyyy.
3. 'STATION iiii, TO BE DELETED, COULD NOT BE FOUND IN DATA SETS'—Station was not in file created by BBPEAK or BBFLOW.
4. 'CANNOT ADD STATION iiiiiii TO DATA SET'—There are 70 stations (maximum) in the file.
5. 'STATION iiiiiii IS NOT IN DATA SET AND COULD NOT BE UPDATED'—An assumption is made that station is already in file (ICODE = 3), whereas, it is not.

BBJOIN

Through the use of cataloged procedure BBJOIN, two data sets that were created by either procedure BBPEAK or BBFLOW may be merged into another single data set. The input data sets are not deleted upon execution of BBJOIN.

The output consists of a printout of the combined list of stations and creation of a data set on a magnetic disk on which are listed a heading indicating data type, a title, first and last years of period of record, station numbers, and the flows which occurred during the period of record for each station.

The title of the second input data set will be assumed for the output data set. The station numbers for the second input data set will be added onto those for the first input data set. The flow records for the second input data set will follow those for the first input data set. No sorting capability exists in this procedure.

Any revisions to the data set by cataloged procedure BBREVISE must be executed before executing BBJOIN. Such revisions must include the elimination of any duplicate data from one of the original data sets.

Job Control Language—The following cards are required to execute BBJOIN. (△ indicates space)

```
/ /xxxxxxxx△JOB△(----)
/ /PROCLIB△DD△DSN = WRD.PROCLIB,DISP = SHR
/ /△EXEC△BBJOIN, VOL1 = volume1,UNIT1 = unit1,NAME1 = name1,
/ /△VOL2 = volume2,UNIT2 = unit2,NAME2 = name2,
/ /△JVOL = volume3,JUNIT = unit3,JNAME = name3
/ /
```

The parameters of the procedure are

VOL1—The volume serial number of the disk containing the first input data set.

VOL2—The volume serial number of the disk containing the second input data set.

JVOL—The volume serial number of the disk which will contain the merged output data set.

UNIT1—The unit from which the first input data set is read.

UNIT2—The unit from which the second input data set is read.

JUNIT—The unit on which the output data set is written.

NAME1—The name of first input data set.

NAME2—The name of second input data set.

JNAME—The name of merged output data set.

Input—No card input is required.

Error Messages—The following error messages can be generated when procedure BBJOIN is executed.

1. 'TOTAL NUMBER OF STATION > 70—JOB ABORTED'—The total number of stations in two input data sets exceeds 70.

2. 'DATA TYPES ON FIRST & SECOND SETS NOT THE SAME—JOB ABORTED'—The data type on one data set is 'PEAK' and on the other is 'MEAN'.

BBPOSPRI

The cataloged procedure BBPOSPRI uses the disk file created by the cataloged procedure BBPEAK, BBFLOW, BBVOLS, BBREVISED, or BBJOIN to compute joint probabilities of coefficients of variation, C_v, and cross-correlation coefficients, ϱ_c, for those stations that have sufficient records available. The statistics are for annual floods if BBPEAK is used, for annual and monthly flows if BBFLOW is used, and for *N*-day high or low flows if BBVOLS is used. The minimum length of record that is sufficient is an input parameter to the procedure BBPOSPRI. The cross-correlation coefficients are adjusted for non-concurrence of records as described in the main text; an estimate of the probability of zero flow during a one-year period is given, and the C_v's are adjusted for zero flow. The average variance and average cross correlation of the logarithms of flows are given along with estimates of the coefficient of variation and the cross correlation of the flows derived from the estimates of the variance and the cross correlation of the logarithms. BBPOSPRI also provides the harmonic mean record length of the input streamflow data.

BBPOSPRI can be used to compute the joint probabilities of C_v and ϱ_c for any desired combination of annual or monthly flows or high or low flows by coding an indicator IDO(k) with a 0 if undesired and with a 1 if desired. The index k designates annual flow (k = 1) or a monthly flow (k = 2, October; k = 3, November; . . . , k = 13, September). If the input data are annual peaks, set IDO(1) = 1. If the input data were created by BBVOLS, the index k would run from 1 to 18. The index k designates 1-day low flow (if k = 1), a 3-day low flow (if k = 2), . . . , a 183-day low flow (if k = 9), a 1-day high flow (if k = 10), . . . , a 183-day high flow (if k = 18).

Job Control Language—The following cards are required to execute BBPOSPRI: (△ indicates space and brackets indicate optional parameter fields. Do not punch in brackets.)

```
/ /xxxxxxxx△JOB△(----)
/ /PROCLIB△DD△DSN = WRD.PROCLIB,DISP = SHR
/ /△EXEC△BBPOSPRI,MUNIT=unit,MVOL=number,NVNAME=name,MDISP=KEEPorDELETE
/ /SYSIN△DD△*
          — INPUT cards —
/ /
```

(To obtain more time for execution, after the MDISP parameter punch TIME = time in minutes.)

The parameters of the procedure[1] are

MUNIT—The unit from which the file from BBPEAK or BBFLOW is read. Unit ordinarily must be identical with unit in the BBPEAK or BBFLOW procedure.

MVOL—The volume serial number on the unit described above. The number must be identical with the number in the procedure BBPEAK, BBFLOW, BBVOLS, or BBJOIN.

MVNAME—The data set name of the file from BBPEAK, BBFLOW, BBVOLS, or BBJOIN. (Note that if the name has any special characters (that is, periods, commas, ampersands) it must be enclosed in single quotes.)

MDISP—If no further use is to be made of the file, code DELETE. If the file is to be kept for further use, code KEEP.

Input—The input for BBPOSPRI consists of one card. This card delcares the minimum concurrent record and sets the indicator IDO.

[1] If BBJOIN was used to concatenate two data sets, the parameters must be consistent with the concatenated data set.

Card 1

Column	Format	Variable	Contents
1-3			Blank
4-5	I2	MINYR	The minimum number of years of concurrent record for the computation of cross-correlation coefficients
6	I1	IDO(1)	IDO(k) = 1 if analysis of data type corresponding
7	I1	IDO(2)	to index k is to be performed. IDO(k) = 0 other-
	⋮		wise. Correspondence of index k to data type is
23	I1	IDO(18)	shown in table.

Output—The output from BBPOSPRI consists of a title, which is carried through from the cataloged procedure BBPEAK, BBFLOW, BBVOLS, or BBJOIN and a listing of the station numbers. Following the station list are the joint probability tables for C_v and ϱ_c as described above. A sample output is show in figure 5.

Error Messages—The following error messages can be generated within the BBPOSPRI program:

1. 'INSUFFICIENT CONCURRENCE TO COMPUTE RHO.'—No pair of stations in the input list had MINYR or more years of concurrent data. The program proceeds to the next step, computation of harmonic mean record length.

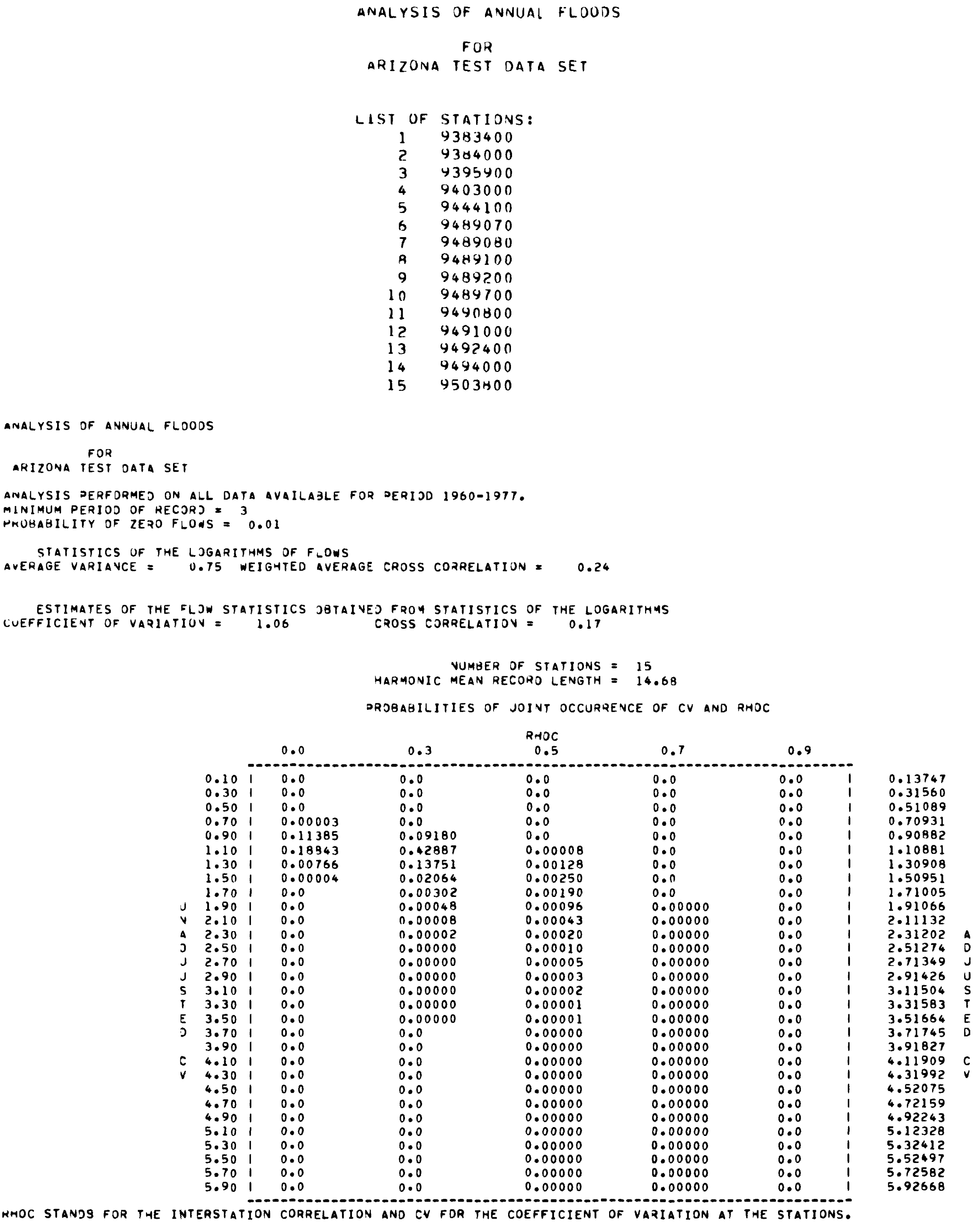
2. 'NUMBER OF STATIONS = < 3. JOB ABORTED'—The number of stations that had streamflow data was less than 3, which is required. Job is stopped.

Table 1. Relation of the index k to data type.

k	If BBFLOWS was used mean streamflow or	If BBVOLS was used
1	year	1 day low flow
2	October	3
3	November	7
4	December	14
5	January	30
6	February	60
7	March	90
8	April	120
9	May	183 day low flow
10	June	1 day high flow
11	July	3
12	August	7
13	September	15
14		30
15		60
16		90
17		120
18		183 day high flow

MODLVALU

MODLVALU is a cataloged procedure that performs a network evaluation and computes points for the definition of network design graphs based on the standard error of estimate of a regional regression analysis. The dependent variable of the regression must be associated with one of the following statistics: mean, standard deviation, or 2-, 10-, 50-, or 100-year event.



```
                                   ANALYSIS OF ANNUAL FLOODS

                                              FOR
                                    ARIZONA TEST DATA SET


                                   LIST OF STATIONS:
                                         1     9383400
                                         2     9384000
                                         3     9395900
                                         4     9403000
                                         5     9444100
                                         6     9489070
                                         7     9489080
                                         8     9489100
                                         9     9489200
                                        10     9489700
                                        11     9490800
                                        12     9491000
                                        13     9492400
                                        14     9494000
                                        15     9503800

ANALYSIS OF ANNUAL FLOODS

         FOR
 ARIZONA TEST DATA SET

ANALYSIS PERFORMED ON ALL DATA AVAILABLE FOR PERIOD 1960-1977.
MINIMUM PERIOD OF RECORD =  3
PROBABILITY OF ZERO FLOWS =  0.01

    STATISTICS OF THE LOGARITHMS OF FLOWS
AVERAGE VARIANCE =    0.75  WEIGHTED AVERAGE CROSS CORRELATION =    0.24


    ESTIMATES OF THE FLOW STATISTICS OBTAINED FROM STATISTICS OF THE LOGARITHMS
COEFFICIENT OF VARIATION =    1.06           CROSS CORRELATION =    0.17


                                        NUMBER OF STATIONS =  15
                                 HARMONIC MEAN RECORD LENGTH =  14.68

                                 PROBABILITIES OF JOINT OCCURRENCE OF CV AND RHOC

                                                     RHOC
                         0.0            0.3            0.5            0.7            0.9
                      ----------------------------------------------------------------------
                0.10 I  0.0            0.0            0.0            0.0            0.0       I   0.13747
                0.30 I  0.0            0.0            0.0            0.0            0.0       I   0.31560
                0.50 I  0.0            0.0            0.0            0.0            0.0       I   0.51089
                0.70 I  0.00003        0.0            0.0            0.0            0.0       I   0.70931
                0.90 I  0.11385        0.09180        0.0            0.0            0.0       I   0.90882
                1.10 I  0.18843        0.42887        0.00008        0.0            0.0       I   1.10881
                1.30 I  0.00766        0.13751        0.00128        0.0            0.0       I   1.30908
                1.50 I  0.00004        0.02064        0.00250        0.0            0.0       I   1.50951
                1.70 I  0.0            0.00302        0.00190        0.0            0.0       I   1.71005
              U 1.90 I  0.0            0.00048        0.00096        0.00000        0.0       I   1.91066
              N 2.10 I  0.0            0.00008        0.00043        0.00000        0.0       I   2.11132
              A 2.30 I  0.0            0.00002        0.00020        0.00000        0.0       I   2.31202  A
              D 2.50 I  0.0            0.00000        0.00010        0.00000        0.0       I   2.51274  D
              J 2.70 I  0.0            0.00000        0.00005        0.00000        0.0       I   2.71349  J
              U 2.90 I  0.0            0.00000        0.00003        0.00000        0.0       I   2.91426  U
              S 3.10 I  0.0            0.00000        0.00002        0.00000        0.0       I   3.11504  S
              T 3.30 I  0.0            0.00000        0.00001        0.00000        0.0       I   3.31583  T
              E 3.50 I  0.0            0.00000        0.00001        0.00000        0.0       I   3.51664  E
              D 3.70 I  0.0            0.0            0.00000        0.00000        0.0       I   3.71745  D
                3.90 I  0.0            0.0            0.00000        0.00000        0.0       I   3.91827
              C 4.10 I  0.0            0.0            0.00000        0.00000        0.0       I   4.11909  C
              V 4.30 I  0.0            0.0            0.00000        0.00000        0.0       I   4.31992  V
                4.50 I  0.0            0.0            0.00000        0.00000        0.0       I   4.52075
                4.70 I  0.0            0.0            0.00000        0.00000        0.0       I   4.72159
                4.90 I  0.0            0.0            0.00000        0.00000        0.0       I   4.92243
                5.10 I  0.0            0.0            0.00000        0.00000        0.0       I   5.12328
                5.30 I  0.0            0.0            0.00000        0.00000        0.0       I   5.32412
                5.50 I  0.0            0.0            0.00000        0.00000        0.0       I   5.52497
                5.70 I  0.0            0.0            0.00000        0.00000        0.0       I   5.72582
                5.90 I  0.0            0.0            0.00000        0.00000        0.0       I   5.92668
                      ----------------------------------------------------------------------
RHOC STANDS FOR THE INTERSTATION CORRELATION AND CV FOR THE COEFFICIENT OF VARIATION AT THE STATIONS.
```

Figure 5. Output of BBPOSPRI for sample problem.

Job Control Language—The following cards are required to execute MODLVALU: (△ indicates space):

```
/ /△xxxxxxxx△JOB△(-------)
/ /PROCLIB△DD△DSN = WRD.PROCLIB,DISP = SHR
/ /△EXEC△MODLVALU△TIME.G=x
/ /G.SYSIN△DD△*
          (input cards for MODLVALU)
/*
/ /
```

where x is the time in minutes to execute the GO step. The value of x can be estimated from figure 6 if the number of feasible parameter combinations and the requested number of design points are known. For example, if the name of feasible parameter combinations was 160 and the number of requested design points was 20, the estimated time for the GO step would be 8 minutes and x could be set equal to 10 to allow for error in the estimate. A value of $x = 2$ should be sufficient to obtain at least one design point and the number of feasible parameter combinations.

Input—Each network evaluation to be carried out requires a set of at least five cards. Input to the procedure consists of one or more such sets. The first two cards of each set are header cards and should contain any information considered necessary by the user in identifying the network evaluation being carried out. Entries in the next four types of data cards should be right justified.

Cards 1 and 2

Column	Format	Variable	Contents
1-80	20A4	ID	Any identifier or descriptor for job

Card 3

Column	Format	Variable	Contents
1-4	F4.0	NB	Adjusted number of stations used in regression analysis
5-8	F4.0	NY	Harmonic mean record length of stations used in regression
9-12	I4	NCVRHO	Number of values of CV and RHO that have nonzero priors

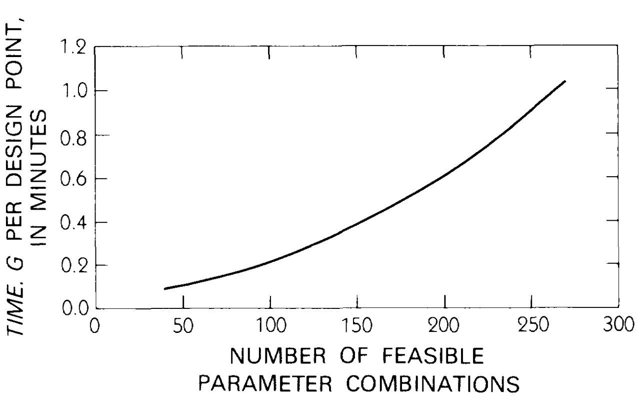


Figure 6. Estimating *TIME.G* = x.

Column	Format	Variable	Contents
13-16	I4	NUMDSN	Number of points for which output is to be printed enabling definition of network design graphs
17-20	I4	IT	Statistical code associated with dependent variables in the regression analysis *Code* *Statistic* 0 mean 1 standard deviation 2 2-year recurrence interval 10 10-year recurrence interval 50 50-year recurrence interval 100 100-year recurrence interval
21-25	F5.3	SEN	Observed value of standard error of estimate of the regression in natural (base e) log units
26-30	F5.3	SEC	Observed value of standard error of estimate of the regression, in common (base 10) log units—code only if SEN is not coded.
36	I1	ICARD	Code '1' for punch card output of design points
41-42	I2	NGAM	Number of model errors for which prior probabilities are nonzero. If blank or zero program assigns prior probabilities.

Card 4
[of which there are NCVRHO]

Column	Format	Variable	Contents
1-5	F5.3	RHO(I)	*I*th value of RHOC for which a nonzero prior exists (*I*=1 to NCVRHO)
6-10	F5.3	CV(I)	*I*th value of CV for which a nonzero prior exists
11-15	F5.3	P[RHO(I),CV(I)]	Joint prior probability associated with the *I*th value of CV and RHOC

Card 5

Column	Format	Variable	Contents
1-5, 11-15, 21-25, . . . 71-75	F5.0	NBDSN(J)	Number of stations associated with the *J*th point for which distribution of true standard error is to be evaluated (*J*= to NUMDSN)
6-10, 16-20, 26-30, . . . 76-80	F5.0	NYDSN(J)	Harmonic mean record length, in years, associated with the *J*th point for which a distribution of true standard error is to be evaluated.

Code this card only if NGAM is greater than zero.

Card 6

Column	Format	Variable	Contents
1–4, 9–12, . . . 65–68	F4.2	GAMMA(J)	Value of model error for which a nonzero prior probability is entered ($J=1$, NGAM)
5–8, 13–16, . . . 69–72	F4.2	PRIGAM(J)	Nonzero probability associated with GAMMA(J), $J=1$, NGAM

Values of the input parameters are subject to the following restrictions:

10. ≤ NY ≤50.

10. ≤ NB ≤ 50.

1 ≤ NCVRHO ≤ 100

1 ≤ NUMDSN ≤ 99

IT ϵ {0, 1, 2, 10, 50, 100}

0.0 < SEN

0.0 < SEC

0.1 ≤ CV ≤ 5.0

0.0 ≤ RHOC ≤ 0.9

10 ≤ NBDSN ≤ 50.

10. ≤ NYDSN ≤ 50.

0 ≤ NGAM ≤ 41

Output—A sample of the output from MODLVALU is given in figure 7. The information from the top of the printout through the table of input parameters and their joint priors is a summary of the input to the analysis. The next table in figure 7 summarizes the posterior probabilities of the parameters CV, RHOC, and model error which have been obtained through the Bayesian analysis. NETWORK DESIGN POINTS is a listing of twelve confidence levels and the corresponding average true regression error for each design point (NBDSN, NYDSN) requested on the input data cards. The average true regression error is stated in log units and percentage. The confidence level, $\alpha(I)$, is the probability that the actual average regression error is less than or equal to the average regression error listed directly under $\alpha(I)$.

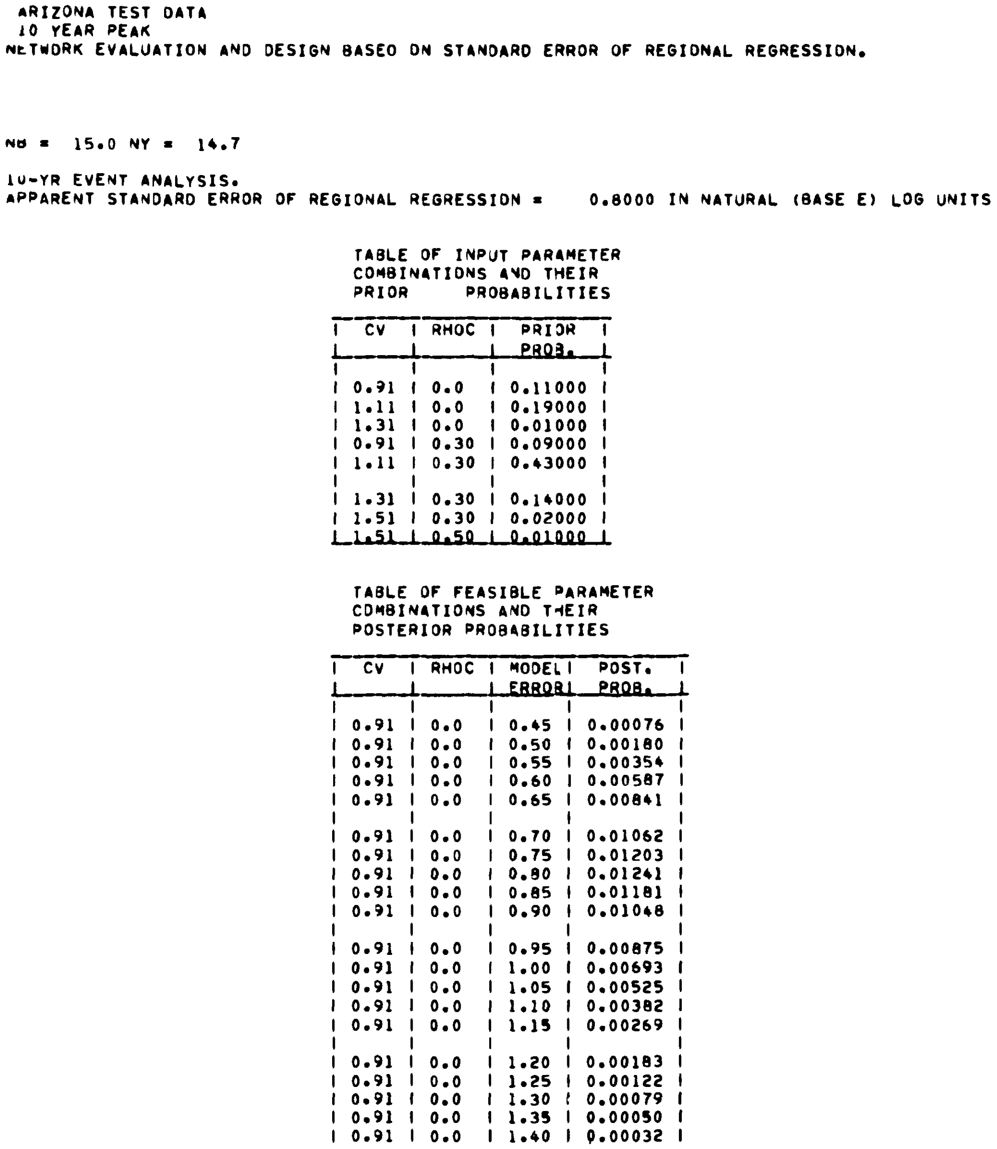
Error Messages—The following error messages can be generated while using the MODLVALU program:

1. 'NUMBER OF FEASIBLE PARAMETER COMBINATIONS EXCEEDS NUMBER ALLOTTED'—The number of combinations of RHOC, CV, and model error having an associated nonzero probability was greater than 100.

2. '**OBSERVED STANDARD ERROR APPEARS TO BE INCONSISTENT WITH THE OTHER PARAMETERS. THE CONDITIONAL PROBABILITY OF SUCH A COMBINATION OF PARAMETERS IS NEARLY ZERO'—No feasible values of CV, RHOC, and model error have been found. The values of the observed standard error, SEN or SEC, and the priors cannot yield a solution. The input data should be checked for correctness of entry and for validity of the priors. The program will attempt to read the next input data set.

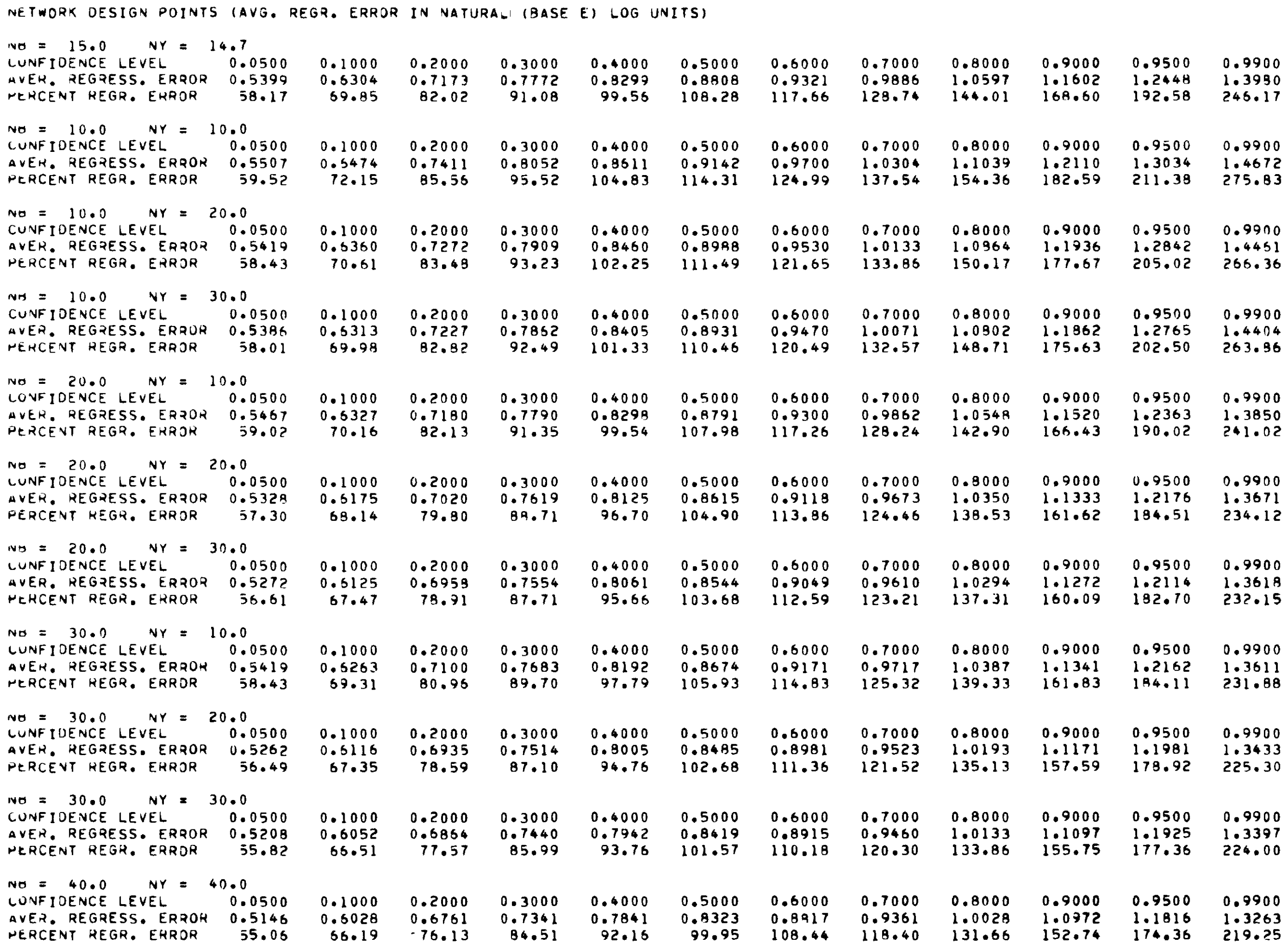
Warnings—The following warnings will be printed if input restrictions are violated. Program errors may result. In interpreting output, the user should be aware of the consequences of violating the following input restrictions:

1. '** NON-ALLOWABLE VALUE OF CV ENTERED.'
2. '** NON-ALLOWABLE VALUE OF RHO ENTERED.'
3. '** NON-ALLOWABLE NUMBER OF DESIGN POINTS REQUESTED.'
4. '** NON-ALLOWABLE NUMBER OF BASINS ENTERED.'
5. '** NON-ALLOWABLE RECORD LENGTH ENTERED.'
6. '** NON-ALLOWABLE NUMBER OF RHOC, AND CV COMBINATIONS ENTERED.'
7. '** PRIOR PROBABILITIES OF RHOC AND CV DO NOT ADD TO 1.'



```
ARIZONA TEST DATA
10 YEAR PEAK
NETWORK EVALUATION AND DESIGN BASED ON STANDARD ERROR OF REGIONAL REGRESSION.


NB =  15.0 NY =  14.7

10-YR EVENT ANALYSIS.
APPARENT STANDARD ERROR OF REGIONAL REGRESSION =    0.8000 IN NATURAL (BASE E) LOG UNITS

                    TABLE OF INPUT PARAMETER
                    COMBINATIONS AND THEIR
                    PRIOR     PROBABILITIES
                  ---------------------------
                  I  CV  I RHOC I  PRIOR  I
                  I______I______I__PROB.__I
                  I      I      I         I
                  I 0.91 I 0.0  I 0.11000 I
                  I 1.11 I 0.0  I 0.19000 I
                  I 1.31 I 0.0  I 0.01000 I
                  I 0.91 I 0.30 I 0.09000 I
                  I 1.11 I 0.30 I 0.43000 I
                  I      I      I         I
                  I 1.31 I 0.30 I 0.14000 I
                  I 1.51 I 0.30 I 0.02000 I
                  I_1.51_I_0.50_I_0.01000_I

                    TABLE OF FEASIBLE PARAMETER
                    COMBINATIONS AND THEIR
                    POSTERIOR PROBABILITIES
                  -----------------------------------
                  I  CV  I RHOC I MODELI   POST.  I
                  I______I______I_ERRORI__PROB.___I
                  I      I      I      I          I
                  I 0.91 I 0.0  I 0.45 I 0.00076  I
                  I 0.91 I 0.0  I 0.50 I 0.00180  I
                  I 0.91 I 0.0  I 0.55 I 0.00354  I
                  I 0.91 I 0.0  I 0.60 I 0.00587  I
                  I 0.91 I 0.0  I 0.65 I 0.00841  I
                  I      I      I      I          I
                  I 0.91 I 0.0  I 0.70 I 0.01062  I
                  I 0.91 I 0.0  I 0.75 I 0.01203  I
                  I 0.91 I 0.0  I 0.80 I 0.01241  I
                  I 0.91 I 0.0  I 0.85 I 0.01181  I
                  I 0.91 I 0.0  I 0.90 I 0.01048  I
                  I      I      I      I          I
                  I 0.91 I 0.0  I 0.95 I 0.00875  I
                  I 0.91 I 0.0  I 1.00 I 0.00693  I
                  I 0.91 I 0.0  I 1.05 I 0.00525  I
                  I 0.91 I 0.0  I 1.10 I 0.00382  I
                  I 0.91 I 0.0  I 1.15 I 0.00269  I
                  I      I      I      I          I
                  I 0.91 I 0.0  I 1.20 I 0.00183  I
                  I 0.91 I 0.0  I 1.25 I 0.00122  I
                  I 0.91 I 0.0  I 1.30 I 0.00079  I
                  I 0.91 I 0.0  I 1.35 I 0.00050  I
                  I 0.91 I 0.0  I 1.40 I 0.00032  I
```

Figure 7. Partial listing of output of MODLVALU.



```
NETWORK DESIGN POINTS (AVG. REGR. ERROR IN NATURAL (BASE E) LOG UNITS)

NB =  15.0     NY =  14.7
CONFIDENCE LEVEL     0.0500    0.1000    0.2000    0.3000    0.4000    0.5000    0.6000    0.7000    0.8000    0.9000    0.9500    0.9900
AVER. REGRESS. ERROR 0.5399    0.6304    0.7173    0.7772    0.8299    0.8808    0.9321    0.9886    1.0597    1.1602    1.2448    1.3980
PERCENT REGR. ERROR   58.17     69.85     82.02     91.08     99.56    108.28    117.66    128.74    144.01    168.60    192.58    246.17

NB =  10.0     NY =  10.0
CONFIDENCE LEVEL     0.0500    0.1000    0.2000    0.3000    0.4000    0.5000    0.6000    0.7000    0.8000    0.9000    0.9500    0.9900
AVER. REGRESS. ERROR 0.5507    0.6474    0.7411    0.8052    0.8611    0.9142    0.9700    1.0304    1.1039    1.2110    1.3034    1.4672
PERCENT REGR. ERROR   59.52     72.15     85.56     95.52    104.83    114.31    124.99    137.54    154.36    182.59    211.38    275.83

NB =  10.0     NY =  20.0
CONFIDENCE LEVEL     0.0500    0.1000    0.2000    0.3000    0.4000    0.5000    0.6000    0.7000    0.8000    0.9000    0.9500    0.9900
AVER. REGRESS. ERROR 0.5419    0.6360    0.7272    0.7909    0.8460    0.8988    0.9530    1.0133    1.0964    1.1936    1.2842    1.4461
PERCENT REGR. ERROR   58.43     70.61     83.48     93.23    102.25    111.49    121.65    133.86    150.17    177.67    205.02    266.36

NB =  10.0     NY =  30.0
CONFIDENCE LEVEL     0.0500    0.1000    0.2000    0.3000    0.4000    0.5000    0.6000    0.7000    0.8000    0.9000    0.9500    0.9900
AVER. REGRESS. ERROR 0.5386    0.6313    0.7227    0.7862    0.8405    0.8931    0.9470    1.0071    1.0802    1.1862    1.2765    1.4404
PERCENT REGR. ERROR   58.01     69.98     82.82     92.49    101.33    110.46    120.49    132.57    148.71    175.63    202.50    263.86

NB =  20.0     NY =  10.0
CONFIDENCE LEVEL     0.0500    0.1000    0.2000    0.3000    0.4000    0.5000    0.6000    0.7000    0.8000    0.9000    0.9500    0.9900
AVER. REGRESS. ERROR 0.5467    0.6327    0.7180    0.7790    0.8298    0.8791    0.9300    0.9862    1.0548    1.1520    1.2363    1.3850
PERCENT REGR. ERROR   59.02     70.16     82.13     91.35     99.54    107.98    117.26    128.24    142.90    166.43    190.02    241.02

NB =  20.0     NY =  20.0
CONFIDENCE LEVEL     0.0500    0.1000    0.2000    0.3000    0.4000    0.5000    0.6000    0.7000    0.8000    0.9000    0.9500    0.9900
AVER. REGRESS. ERROR 0.5328    0.6175    0.7020    0.7619    0.8125    0.8615    0.9118    0.9673    1.0350    1.1333    1.2176    1.3671
PERCENT REGR. ERROR   57.30     68.14     79.80     88.71     96.70    104.90    113.86    124.46    138.53    161.62    184.51    234.12

NB =  20.0     NY =  30.0
CONFIDENCE LEVEL     0.0500    0.1000    0.2000    0.3000    0.4000    0.5000    0.6000    0.7000    0.8000    0.9000    0.9500    0.9900
AVER. REGRESS. ERROR 0.5272    0.6125    0.6958    0.7554    0.8061    0.8544    0.9049    0.9610    1.0294    1.1272    1.2114    1.3618
PERCENT REGR. ERROR   56.61     67.47     78.91     87.71     95.66    103.68    112.59    123.21    137.31    160.09    182.70    232.15

NB =  30.0     NY =  10.0
CONFIDENCE LEVEL     0.0500    0.1000    0.2000    0.3000    0.4000    0.5000    0.6000    0.7000    0.8000    0.9000    0.9500    0.9900
AVER. REGRESS. ERROR 0.5419    0.6263    0.7100    0.7683    0.8192    0.8674    0.9171    0.9717    1.0387    1.1341    1.2162    1.3611
PERCENT REGR. ERROR   58.43     69.31     80.96     89.70     97.79    105.93    114.83    125.32    139.33    161.83    184.11    231.88

NB =  30.0     NY =  20.0
CONFIDENCE LEVEL     0.0500    0.1000    0.2000    0.3000    0.4000    0.5000    0.6000    0.7000    0.8000    0.9000    0.9500    0.9900
AVER. REGRESS. ERROR 0.5262    0.6116    0.6935    0.7514    0.8005    0.8485    0.8981    0.9523    1.0193    1.1171    1.1981    1.3433
PERCENT REGR. ERROR   56.49     67.35     78.59     87.10     94.76    102.68    111.36    121.52    135.13    157.59    178.92    225.30

NB =  30.0     NY =  30.0
CONFIDENCE LEVEL     0.0500    0.1000    0.2000    0.3000    0.4000    0.5000    0.6000    0.7000    0.8000    0.9000    0.9500    0.9900
AVER. REGRESS. ERROR 0.5208    0.6052    0.6864    0.7440    0.7942    0.8419    0.8915    0.9460    1.0133    1.1097    1.1925    1.3397
PERCENT REGR. ERROR   55.82     66.51     77.57     85.99     93.76    101.57    110.18    120.30    133.86    155.75    177.36    224.00

NB =  40.0     NY =  40.0
CONFIDENCE LEVEL     0.0500    0.1000    0.2000    0.3000    0.4000    0.5000    0.6000    0.7000    0.8000    0.9000    0.9500    0.9900
AVER. REGRESS. ERROR 0.5146    0.6028    0.6761    0.7341    0.7841    0.8323    0.8917    0.9361    1.0028    1.0972    1.1816    1.3263
PERCENT REGR. ERROR   55.06     66.19     76.13     84.51     92.16     99.95    108.44    118.40    131.66    152.74    174.36    219.25
```

Figure 7. (continued)

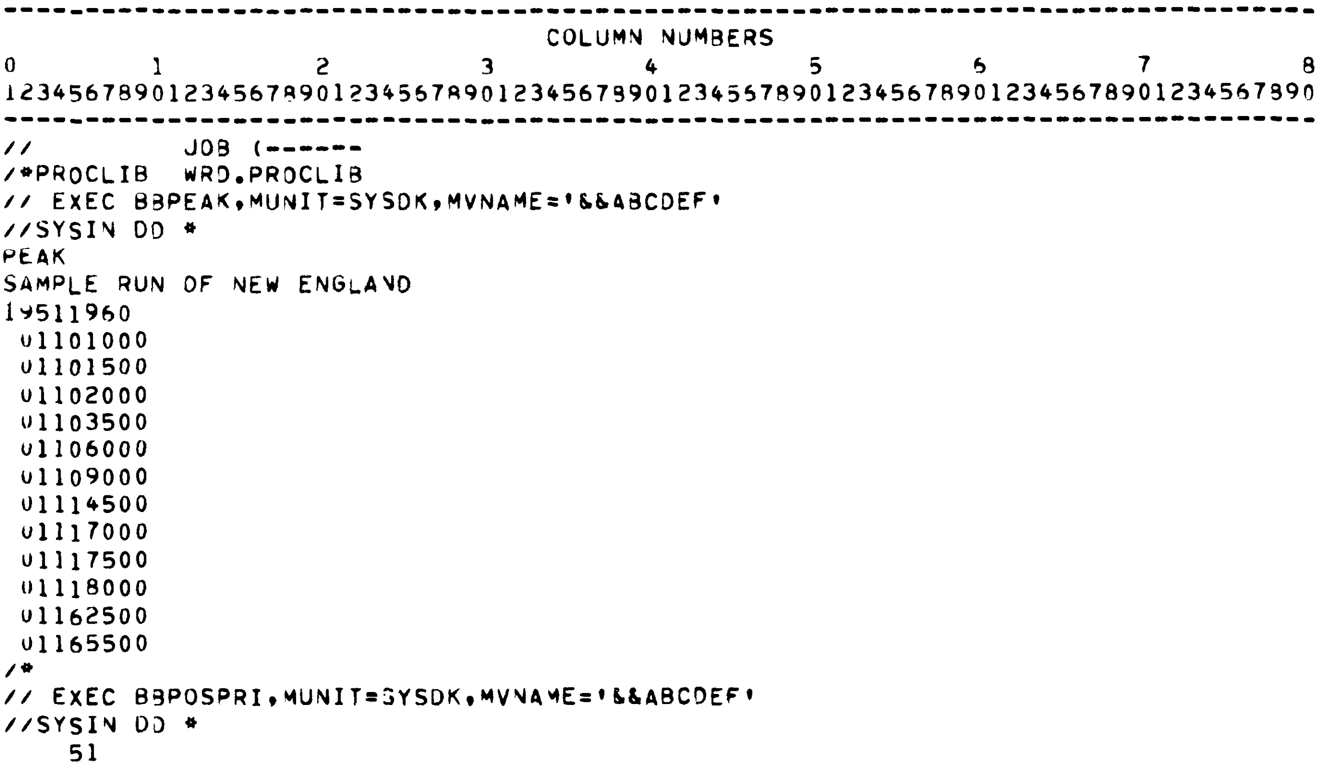
EXAMPLE

The following example illustrates the use of the procedures for network design. A regional regression analysis relating 50-year annual peaks for 12 stations in New England to one basic characteristic was available for study. The period of record for the stations ranged from 1951 to 1960. Standard error of estimate was given as 0.1000 natural log units.

The first step was to run BBPEAK and BBPOSPRI. Data cards were coded as shown in figure 8.

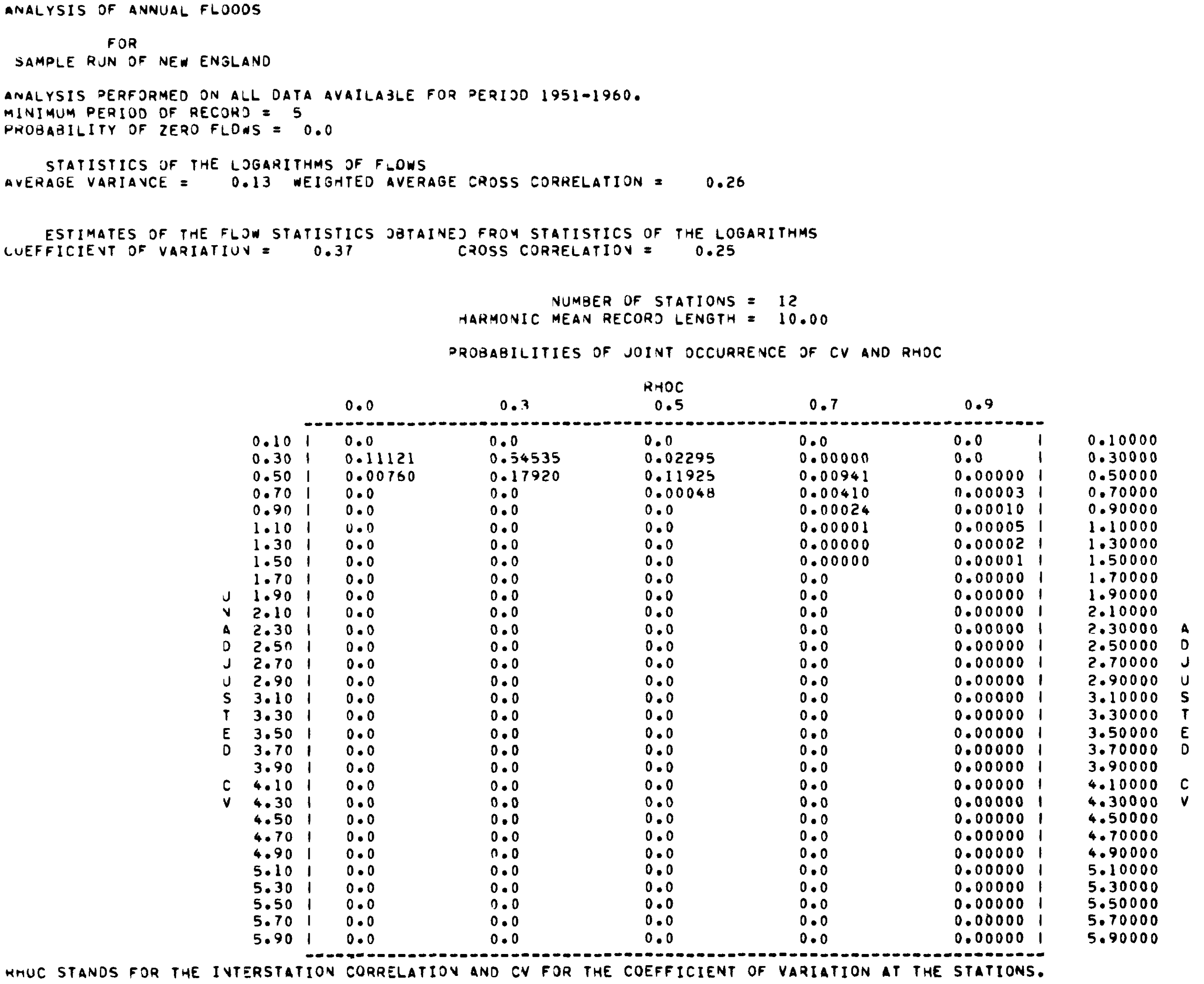
Output for BBPOSPRI is shown in figure 9. Prior probabilities for C_v and ϱ_c for input to MODLVALU were taken from the relative frequency tables of figure 9.

Input cards for MODLVALU are shown in figure 10. Output for MODLVALU is shown in figure 11. Regression errors for 50 percent confidence level are plotted as a function of *NB* and *NY* in figure 12.



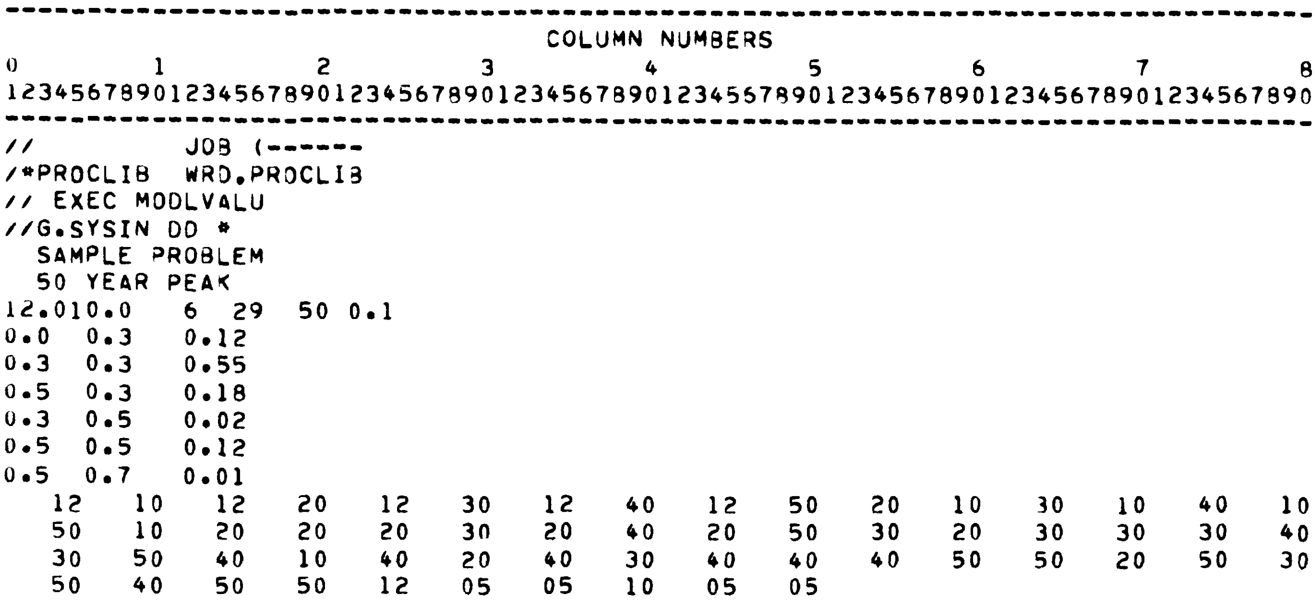
```
--------------------------------------------------------------------------------
                                 COLUMN NUMBERS
0        1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890
--------------------------------------------------------------------------------
//         JOB (------
/*PROCLIB  WRD.PROCLIB
// EXEC BBPEAK,MUNIT=SYSDK,MVNAME='&&ABCDEF'
//SYSIN DD *
PEAK
SAMPLE RUN OF NEW ENGLAND
19511960
 01101000
 01101500
 01102000
 01103500
 01106000
 01109000
 01114500
 01117000
 01117500
 01118000
 01162500
 01165500
/*
// EXEC BBPOSPRI,MUNIT=SYSDK,MVNAME='&&ABCDEF'
//SYSIN DD *
    51
```

Figure 8. Cards used to execute BBPEAK and BBPOSPRI for sample problem.



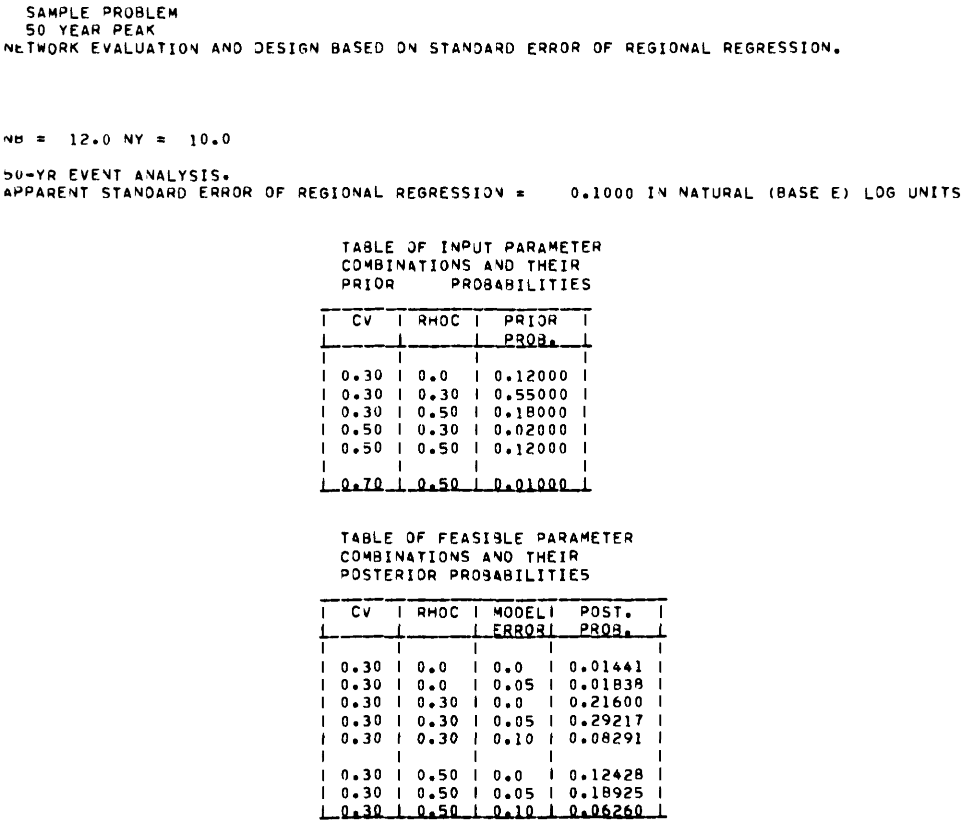
```
ANALYSIS OF ANNUAL FLOODS

         FOR
 SAMPLE RUN OF NEW ENGLAND

ANALYSIS PERFORMED ON ALL DATA AVAILABLE FOR PERIOD 1951-1960.
MINIMUM PERIOD OF RECORD =  5
PROBABILITY OF ZERO FLOWS =  0.0

    STATISTICS OF THE LOGARITHMS OF FLOWS
AVERAGE VARIANCE =     0.13  WEIGHTED AVERAGE CROSS CORRELATION =     0.26


    ESTIMATES OF THE FLOW STATISTICS OBTAINED FROM STATISTICS OF THE LOGARITHMS
COEFFICIENT OF VARIATION =     0.37            CROSS CORRELATION =     0.25


                                                 NUMBER OF STATIONS =  12
                                         HARMONIC MEAN RECORD LENGTH =  10.00

                                         PROBABILITIES OF JOINT OCCURRENCE OF CV AND RHOC

                                                               RHOC
                      0.0             0.3             0.5             0.7             0.9
                  ----------------------------------------------------------------------------
           0.10 I  0.0             0.0             0.0             0.0             0.0       I    0.10000
           0.30 I  0.11121         0.54535         0.02295         0.00000         0.0       I    0.30000
           0.50 I  0.00760         0.17920         0.11925         0.00941         0.00000   I    0.50000
           0.70 I  0.0             0.0             0.00048         0.00410         0.00003   I    0.70000
           0.90 I  0.0             0.0             0.0             0.00024         0.00010   I    0.90000
           1.10 I  0.0             0.0             0.0             0.00001         0.00005   I    1.10000
           1.30 I  0.0             0.0             0.0             0.00000         0.00002   I    1.30000
           1.50 I  0.0             0.0             0.0             0.00000         0.00001   I    1.50000
           1.70 I  0.0             0.0             0.0             0.0             0.00000   I    1.70000
        U  1.90 I  0.0             0.0             0.0             0.0             0.00000   I    1.90000
        N  2.10 I  0.0             0.0             0.0             0.0             0.00000   I    2.10000
        A  2.30 I  0.0             0.0             0.0             0.0             0.00000   I    2.30000   A
        D  2.50 I  0.0             0.0             0.0             0.0             0.00000   I    2.50000   D
        J  2.70 I  0.0             0.0             0.0             0.0             0.00000   I    2.70000   J
        U  2.90 I  0.0             0.0             0.0             0.0             0.00000   I    2.90000   U
        S  3.10 I  0.0             0.0             0.0             0.0             0.00000   I    3.10000   S
        T  3.30 I  0.0             0.0             0.0             0.0             0.00000   I    3.30000   T
        E  3.50 I  0.0             0.0             0.0             0.0             0.00000   I    3.50000   E
        D  3.70 I  0.0             0.0             0.0             0.0             0.00000   I    3.70000   D
           3.90 I  0.0             0.0             0.0             0.0             0.00000   I    3.90000
        C  4.10 I  0.0             0.0             0.0             0.0             0.00000   I    4.10000   C
        V  4.30 I  0.0             0.0             0.0             0.0             0.00000   I    4.30000   V
           4.50 I  0.0             0.0             0.0             0.0             0.00000   I    4.50000
           4.70 I  0.0             0.0             0.0             0.0             0.00000   I    4.70000
           4.90 I  0.0             0.0             0.0             0.0             0.00000   I    4.90000
           5.10 I  0.0             0.0             0.0             0.0             0.00000   I    5.10000
           5.30 I  0.0             0.0             0.0             0.0             0.00000   I    5.30000
           5.50 I  0.0             0.0             0.0             0.0             0.00000   I    5.50000
           5.70 I  0.0             0.0             0.0             0.0             0.00000   I    5.70000
           5.90 I  0.0             0.0             0.0             0.0             0.00000   I    5.90000
                  ----------------------------------------------------------------------------
RHOC STANDS FOR THE INTERSTATION CORRELATION AND CV FOR THE COEFFICIENT OF VARIATION AT THE STATIONS.
```

Figure 9. Output of BBPOSPRI for sample problem.



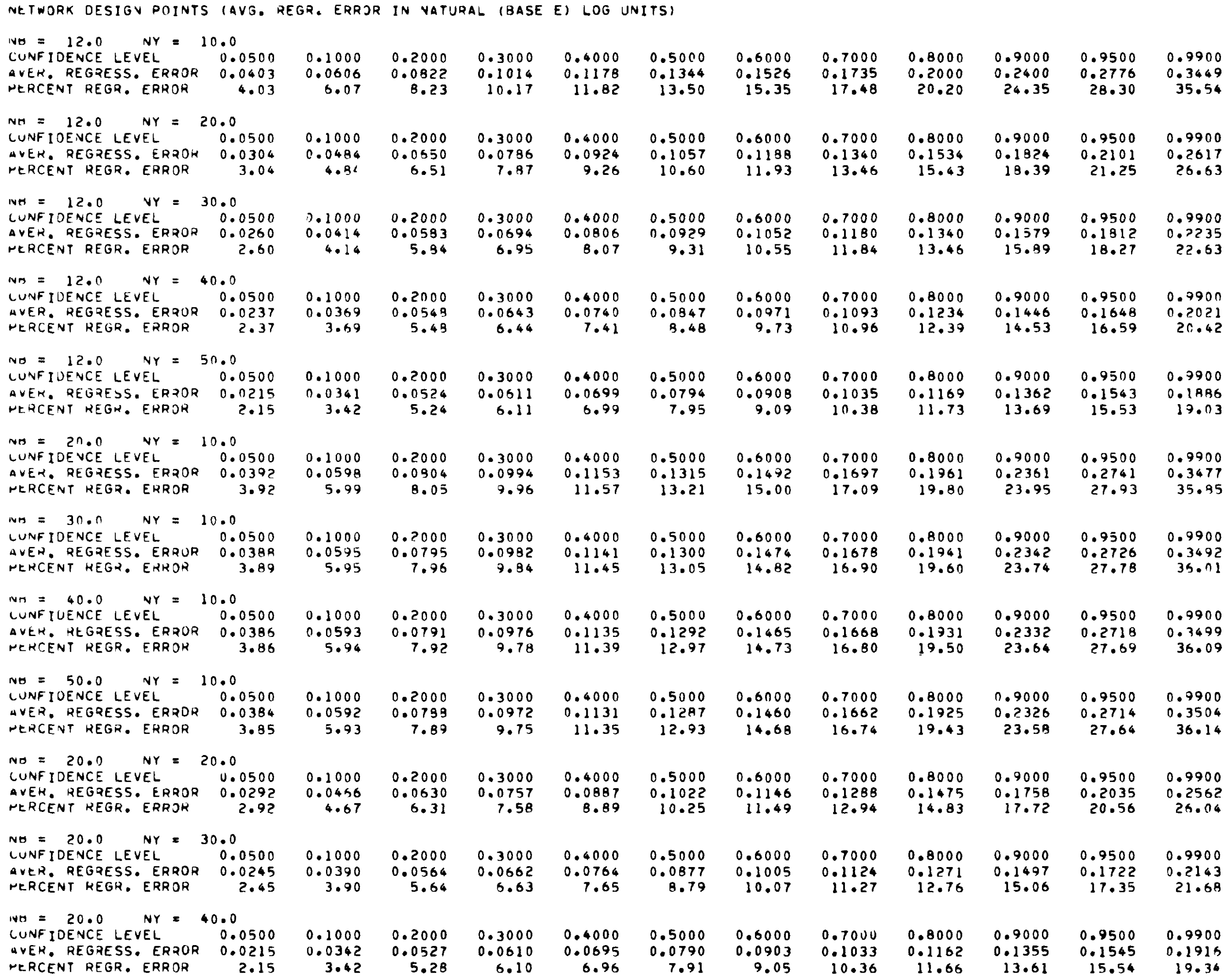
```
--------------------------------------------------------------------------------
                                COLUMN NUMBERS
0        1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890
--------------------------------------------------------------------------------
//         JOB (------
/*PROCLIB  WRD.PROCLIB
// EXEC MODLVALU
//G.SYSIN DD *
  SAMPLE PROBLEM
  50 YEAR PEAK
12.010.0   6  29  50 0.1
0.0  0.3   0.12
0.3  0.3   0.55
0.5  0.3   0.18
0.3  0.5   0.02
0.5  0.5   0.12
0.5  0.7   0.01
   12   10   12   20   12   30   12   40   12   50   20   10   30   10   40   10
   50   10   20   20   20   30   20   40   20   50   30   20   30   30   30   40
   30   50   40   10   40   20   40   30   40   40   40   50   50   20   50   30
   50   40   50   50   12   05   05   10   05   05
```

Figure 10. Cards used to execute MODLVALU for sample problem.



```
   SAMPLE PROBLEM
   50 YEAR PEAK
NETWORK EVALUATION AND DESIGN BASED ON STANDARD ERROR OF REGIONAL REGRESSION.



NB =  12.0 NY =  10.0

50-YR EVENT ANALYSIS.
APPARENT STANDARD ERROR OF REGIONAL REGRESSION =    0.1000 IN NATURAL (BASE E) LOG UNITS

                         TABLE OF INPUT PARAMETER
                         COMBINATIONS AND THEIR
                         PRIOR     PROBABILITIES

                        -------------------------
                        I  CV   I RHOC I  PRIOR  I
                        I_______I______I__PROB.__I
                        I       I      I         I
                        I 0.30  I 0.0  I 0.12000 I
                        I 0.30  I 0.30 I 0.55000 I
                        I 0.30  I 0.50 I 0.18000 I
                        I 0.50  I 0.30 I 0.02000 I
                        I 0.50  I 0.50 I 0.12000 I
                        I       I      I         I
                        I_0.70__I_0.50_I_0.01000_I

                         TABLE OF FEASIBLE PARAMETER
                         COMBINATIONS AND THEIR
                         POSTERIOR PROBABILITIES

                        ----------------------------------
                        I  CV   I RHOC I MODELI  POST.   I
                        I_______I______I_ERRORI__PROB.___I
                        I       I      I      I          I
                        I 0.30  I 0.0  I 0.0  I 0.01441  I
                        I 0.30  I 0.0  I 0.05 I 0.01838  I
                        I 0.30  I 0.30 I 0.0  I 0.21600  I
                        I 0.30  I 0.30 I 0.05 I 0.29217  I
                        I 0.30  I 0.30 I 0.10 I 0.08291  I
                        I       I      I      I          I
                        I 0.30  I 0.50 I 0.0  I 0.12428  I
                        I 0.30  I 0.50 I 0.05 I 0.18925  I
                        I_0.30__I_0.50_I_0.10_I_0.06260__I
```

Figure 11. Partial output of MODLVALU for sample problem.



```
NETWORK DESIGN POINTS (AVG. REGR. ERROR IN NATURAL (BASE E) LOG UNITS)

NB =  12.0     NY =  10.0
CONFIDENCE LEVEL        0.0500   0.1000   0.2000   0.3000   0.4000   0.5000   0.6000   0.7000   0.8000   0.9000   0.9500   0.9900
AVER. REGRESS. ERROR    0.0403   0.0606   0.0822   0.1014   0.1178   0.1344   0.1526   0.1735   0.2000   0.2400   0.2776   0.3449
PERCENT REGR. ERROR       4.03     6.07     8.23    10.17    11.82    13.50    15.35    17.48    20.20    24.35    28.30    35.54

NB =  12.0     NY =  20.0
CONFIDENCE LEVEL        0.0500   0.1000   0.2000   0.3000   0.4000   0.5000   0.6000   0.7000   0.8000   0.9000   0.9500   0.9900
AVER. REGRESS. ERROR    0.0304   0.0484   0.0650   0.0786   0.0924   0.1057   0.1188   0.1340   0.1534   0.1824   0.2101   0.2617
PERCENT REGR. ERROR       3.04     4.84     6.51     7.87     9.26    10.60    11.93    13.46    15.43    18.39    21.25    26.63

NB =  12.0     NY =  30.0
CONFIDENCE LEVEL        0.0500   0.1000   0.2000   0.3000   0.4000   0.5000   0.6000   0.7000   0.8000   0.9000   0.9500   0.9900
AVER. REGRESS. ERROR    0.0260   0.0414   0.0583   0.0694   0.0806   0.0929   0.1052   0.1180   0.1340   0.1579   0.1812   0.2235
PERCENT REGR. ERROR       2.60     4.14     5.84     6.95     8.07     9.31    10.55    11.84    13.46    15.89    18.27    22.63

NB =  12.0     NY =  40.0
CONFIDENCE LEVEL        0.0500   0.1000   0.2000   0.3000   0.4000   0.5000   0.6000   0.7000   0.8000   0.9000   0.9500   0.9900
AVER. REGRESS. ERROR    0.0237   0.0369   0.0548   0.0643   0.0740   0.0847   0.0971   0.1093   0.1234   0.1446   0.1648   0.2021
PERCENT REGR. ERROR       2.37     3.69     5.48     6.44     7.41     8.48     9.73    10.96    12.39    14.53    16.59    20.42

NB =  12.0     NY =  50.0
CONFIDENCE LEVEL        0.0500   0.1000   0.2000   0.3000   0.4000   0.5000   0.6000   0.7000   0.8000   0.9000   0.9500   0.9900
AVER. REGRESS. ERROR    0.0215   0.0341   0.0524   0.0611   0.0699   0.0794   0.0908   0.1035   0.1169   0.1362   0.1543   0.1886
PERCENT REGR. ERROR       2.15     3.42     5.24     6.11     6.99     7.95     9.09    10.38    11.73    13.69    15.53    19.03

NB =  20.0     NY =  10.0
CONFIDENCE LEVEL        0.0500   0.1000   0.2000   0.3000   0.4000   0.5000   0.6000   0.7000   0.8000   0.9000   0.9500   0.9900
AVER. REGRESS. ERROR    0.0392   0.0598   0.0804   0.0994   0.1153   0.1315   0.1492   0.1697   0.1961   0.2361   0.2741   0.3477
PERCENT REGR. ERROR       3.92     5.99     8.05     9.96    11.57    13.21    15.00    17.09    19.80    23.95    27.93    35.85

NB =  30.0     NY =  10.0
CONFIDENCE LEVEL        0.0500   0.1000   0.2000   0.3000   0.4000   0.5000   0.6000   0.7000   0.8000   0.9000   0.9500   0.9900
AVER. REGRESS. ERROR    0.0388   0.0595   0.0795   0.0982   0.1141   0.1300   0.1474   0.1678   0.1941   0.2342   0.2726   0.3492
PERCENT REGR. ERROR       3.89     5.95     7.96     9.84    11.45    13.05    14.82    16.90    19.60    23.74    27.78    36.01

NB =  40.0     NY =  10.0
CONFIDENCE LEVEL        0.0500   0.1000   0.2000   0.3000   0.4000   0.5000   0.6000   0.7000   0.8000   0.9000   0.9500   0.9900
AVER. REGRESS. ERROR    0.0386   0.0593   0.0791   0.0976   0.1135   0.1292   0.1465   0.1668   0.1931   0.2332   0.2718   0.3499
PERCENT REGR. ERROR       3.86     5.94     7.92     9.78    11.39    12.97    14.73    16.80    19.50    23.64    27.69    36.09

NB =  50.0     NY =  10.0
CONFIDENCE LEVEL        0.0500   0.1000   0.2000   0.3000   0.4000   0.5000   0.6000   0.7000   0.8000   0.9000   0.9500   0.9900
AVER. REGRESS. ERROR    0.0384   0.0592   0.0788   0.0972   0.1131   0.1287   0.1460   0.1662   0.1925   0.2326   0.2714   0.3504
PERCENT REGR. ERROR       3.85     5.93     7.89     9.75    11.35    12.93    14.68    16.74    19.43    23.58    27.64    36.14

NB =  20.0     NY =  20.0
CONFIDENCE LEVEL        0.0500   0.1000   0.2000   0.3000   0.4000   0.5000   0.6000   0.7000   0.8000   0.9000   0.9500   0.9900
AVER. REGRESS. ERROR    0.0292   0.0466   0.0630   0.0757   0.0887   0.1022   0.1146   0.1288   0.1475   0.1758   0.2035   0.2562
PERCENT REGR. ERROR       2.92     4.67     6.31     7.58     8.89    10.25    11.49    12.94    14.83    17.72    20.56    26.04

NB =  20.0     NY =  30.0
CONFIDENCE LEVEL        0.0500   0.1000   0.2000   0.3000   0.4000   0.5000   0.6000   0.7000   0.8000   0.9000   0.9500   0.9900
AVER. REGRESS. ERROR    0.0245   0.0390   0.0564   0.0662   0.0764   0.0877   0.1005   0.1124   0.1271   0.1497   0.1722   0.2143
PERCENT REGR. ERROR       2.45     3.90     5.64     6.63     7.65     8.79    10.07    11.27    12.76    15.06    17.35    21.68

NB =  20.0     NY =  40.0
CONFIDENCE LEVEL        0.0500   0.1000   0.2000   0.3000   0.4000   0.5000   0.6000   0.7000   0.8000   0.9000   0.9500   0.9900
AVER. REGRESS. ERROR    0.0215   0.0342   0.0527   0.0610   0.0695   0.0790   0.0903   0.1033   0.1162   0.1355   0.1545   0.1916
PERCENT REGR. ERROR       2.15     3.42     5.28     6.10     6.96     7.91     9.05    10.36    11.66    13.61    15.54    19.34
```

Figure 11. (continued)

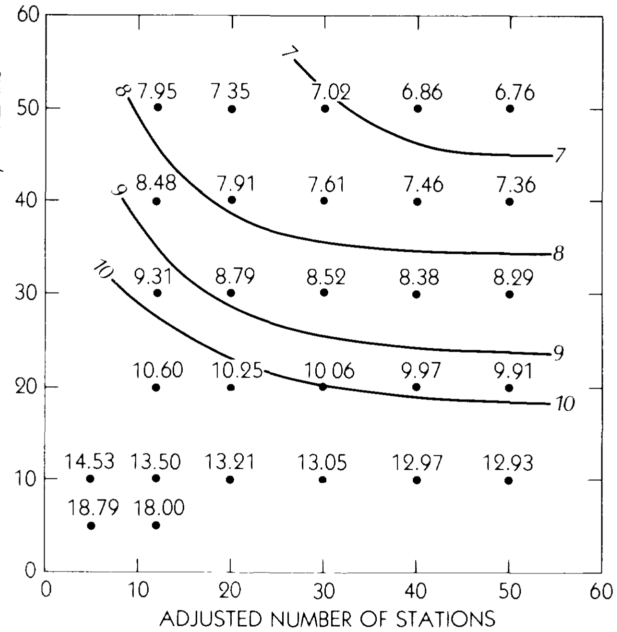


Figure 12. Median estimate of true standard error for sample problem (in percent).

EPILOG

The network analysis and design concepts described in this report have been developed only in the very recent past. Their evolution is expected to continue both with respect to the refinement of the analysis and to the extensions of the uses to which the technique may be applied. As an example to the latter, regional analysis of water-quality parameters such as was performed by Steele and Jennings (1972) seems to be a very fruitful field for the use of NARI.

REFERENCES CITED

Attanasi, E. D., and Karlinger, M. R., 1977, The economic basis of resource information systems. The case of streamflow data network design: Water Resources Research, v. 13, no. 2, p. 273–280.

Benjamin, J. R., and Cornell, C. A., 1970, Probability, statistics, and decision for civil engineers: New York, McGraw-Hill, 684 p.

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

Benson, M. A., and Matalas, N. C., 1967, Synthetic hydrology based on regional statistical parameters: Water Resources Research, 3(4), p. 931–935.

Box, G. E. P., and Tiao, G. C., 1973, Bayesian inference in statistical analysis: Reading, Massachusetts, Addison-Wesley, 588 p.

Fiering, M. B., and Jackson, B. B., 1971, Synthetic streamflows: American Geophysical Union, Water Resources Monograph, no. 1, 98 p.

Geisser, S., 1964, Estimation in the uniform covariance case: Journal of the Royal Statistical Society, Series B., v. 26, p. 477–483.

Hardison, C. H., 1971, Prediction error of regression estimates of streamflow characteristics at ungauged sites: U.S. Geological Survey Professional Paper 750-C, p. C288–C236.

Karlinger, M. R., 1975, Economic worth of hydrologic data in project design. An application to regional energy development: U.S. Geological Survey Open-File Report, 76-316, 63 p.

Kirby, W., 1974, Algebraic boundedness of sample statistics: Water Resources Research, v. 10, no. 2, p. 220–223.

Matalas, N. C., 1967, Mathematical assessment of synthetic hydrology: Water Resources Research, v. 3, no. 4, p. 937–945.

Meeks, W. C., 1975, Daily values statistics (Program A969), in WATSTORE User's Guide, v. 1: U.S. Geological Survey Open-File Report 75-426, Chapter IV, Section G, 32 p.

Moss, M. E., 1976, Information transfer for stream discharge. Hydrological Network Design and Information Transfer: World Meterological Organization Operational Hydrology Report No. 8, Geneva, Switzerland, p. 119–128.

Moss, M. E., 1979, Time, space, and the third dimension (model error): Water Resources Research, v. 15, no. 6, p. 1797–1800.

Moss, M. E., and Haushild, W. L., 1978, Evaluation and design of a streamflow-data network in Washington: U.S. Geological Survey Open-File Report 78-167, 43 p. and 1 plate.

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.

Raiffa, Howard, 1970, Decision analysis—introductory lectures on choices under uncertainty: Reading, Massachusetts, Addison-Wesley, 309 p.

Slack, J. R., Wallis, J. R., and Matalas, N. C., 1976, Distribution functions for statistics derived from bivariate normal and bivariate 2-parameter log-normal population, U.S. Geological Survey Open-File Report 76-214. (Also available as IBM Research Rept. RC 5794 (#259111).)

Steele, T. D., and Jennings, M. E., 1972, Regional analysis of streamflow chemical quality in Texas: Water Resources Research, v. 8, no. 20, p. 460–477.

Tasker, G. D., and Moss, M. E., 1979, Analysis of an Arizona flood-data network for regional information: Water Resources Research, v. 15, no. 6, p. 1791–1796.

Wallis, J. R., Matalas, N. C., and Slack, J. R., 1976, Effect of sequence length *n* on the choice of assumed distribution of floods: Water Resources Research, v. 16, no. 3, p. 457–471.

———1977, Apparent regional skew: Water Resources Research, v. 13, no. 1, p. 159–182.