t No. FHWA-RD-77-21

ASSESSMENT OF NATIONAL SMALL RURAL WATERSHEDS PROGRAM

Vol. 1. Technical Report

June 1977
Final Report

Document is available to the public through
the National Technical Information Service,
Springfield, Virginia 22161

Prepared for
FEDERAL HIGHWAY ADMINISTRATION
Offices of Research & Development
Washington, D. C. 20590

NOTICE

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-RD-77-21		

4. Title and Subtitle	5. Report Date
Assessment of National Small Rural Watersheds Program. Volume 1, Technical Report	June 1977
	6. Performing Organization Code
7. Author(s): M. B. Fiering, R. E. Sharpin, G. K. Young	8. Performing Organization Report No.
9. Performing Organization Name and Address: Meta Systems Inc. (in association with GKY & Associates) 10 Holworthy St. Cambridge, MA. 02138	10. Work Unit No.: FCP 35H3-092
	11. Contract or Grant No.: DOT-FH-11-8605
12. Sponsoring Agency Name and Address: Office of Research and Development Federal Highway Administration U.S. Department of Transportation Washington, D.C. 20590	13. Type of Report and Period Covered: Final Report
	14. Sponsoring Agency Code

15. Supplementary Notes

FHWA Contract Manager: Frank K. Stovicek

16. Abstract

A State-by-State assessment is made of the current status of the National Small Rural Watersheds Program with regard to adequacy of data collection and analysis. Methodology is recommended for flood frequency estimation to replace currently used biased approaches. Concepts of risk aversion are discussed, and decision criteria based on economic considerations are incorporated into the hydrologic evaluation. Stream gaging programs of various gaging densities for 48 States are evaluated and recommendations made for continuation or termination of the programs based on FHWA objectives of drainage culvert design.

Volumes 1 and 2 of the report are available upon request.

FHWA No.	Short Title
77-21	Technical Report (Volume 1)
77-22	Appendices (Volume 2)
77-23	Executive Summary

17. Key Words: Hydrology, Small Watershed Gaging Networks, Worth of Data, Statistical Analysis, Drainage Structures, Design Flow, Economic Analysis.

18. Distribution Statement: This document is available to the public through the National Technical Information Service, Springfield, Virginia 22151.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	226	

Form DOT F 1700.7 (8-69)

ACKNOWLEDGMENTS

Cooperation of numerous agencies and persons was necessary to successfully complete this study. Provision of economic data and highway plans by the State highway departments of Alabama, California, Georgia, Idaho, Maine, Missouri, Oregon, South Dakota, Texas and Virginia is gratefully acknowledged. Regional meetings related to small watershed problems were attended by Federal Highway Administration personnel from Connecticut, Massachusetts and New York as well as State highway and U.S. Geological Survey officials. Extensive input was received from the Systems Group of the U.S. Geological Survey, Reston, Virginia on hydrological concepts and methodologies. Many constructive comments and suggestions were received from reviewers of the draft report and are incorporated in the final report; their interest and contributions are appreciated. The participation of the Federal Highway Administration, especially the Environmental Design and Control Division, is acknowledged. Special thanks go to Mr. Frank Stovicek of the Federal Highway Administration, who served as the contract manager and provided extensive input to the report.

Table of Contents - Volume 1

		Page
Section 1	Introduction	1
Section 2	Restatement of the Problem	9
	General Literature Review	9
	Basic Documents	13
	Potter's Method (1961)	13
	Fletcher's Method (Utah State University, (1974)	15
	Travelers Research (1972)	18
	Interagency Committee (1974)	19
	Early Statistical Analyses	22
	U.S. Geological Survey Programs	25
	Basic Reports	25
	Equivalent Years of Record	28
	Recent Network Theory	30
	Sources of Error in Regression Models	30
	BIGBASIN	33
	Work Done by the States	37
	Position Held by the States	39
	A Revised Statement of the Problem	39
	Work Plan	48
	Theoretical Basis	48
	Distribution of Design Flow	54
	Alternative Algorithms	56
	Outliers and Skewness	60
	Arguments for BIGBASIN	64
	Extension of Gaging	68
Section 3	Economic and Hydrologic Studies	71
	Economic Analysis	71
	General	71
	Construction Cost Savings	74

Table of Contents (continued)

		Page
	Primary Data Sources	79
	Total Highway Costs per Mile	80
	Culvert Data Compiled from Plans	80
	Generalized Hydraulic Functions	87
	Generalized Cost Functions	89
	Sampling Culvert Costs	94
	Highway Needs and Potential Benefits	94
	Numerical Example	103
	Results of Hydrologic Analysis	106
	General	106
	Estimation of Q_{50}	106
	Regression Analysis	118
	Other Arguments for Use of BIGBASIN	125
Section 4	Decision Analysis	154
	Existing Hydrometric Network	154
	Development of a Decision Table	161
	Discussion of Results	183
	Implications of the Results	198
Section 5	Recommended Research for Improvement of Small Watershed Program	201
	Frequency Analysis Revisited	201
	Non-Parametric Technique	202
	Multinomial Logit Analysis	205
	A New Hydrologic Framework	207
	Summary Recommendations	207
	REFERENCES	208
	BIBLIOGRAPHY	212

Table of Contents - Volume 2

		Page
Appendix A	Statement of Work	1
Appendix B	Program Documentation	5
Appendix C	Gaging Stations by State	290

List of Figures

		Page
Figure 1	Review of Agency Involvement	10
Figure 2	Record of Annual Floods	42
Figure 3	Distribution Function for Pooled Annual Floods	42
Figure 4	Distribution Function for Segment of Record	42
Figure 5	Histogram of Estimates of Q_T	43
Figure 6	Theoretical Density Function for Annual Floods	43
Figure 7	Distribution of $\hat{Q}_{50}$ as a Function of Record Length	50
Figure 8	Reliability of Estimates of Q_{50}	51
Figure 9	Contours of Equal Design Flow	53
Figure 10	Flow-Duration Curve at 2229000	58
Figure 11	Flow-Duration Curve at 2229000	59
Figure 12	$\hat{Q}_{50}$-Duration Curve, Virginia Sites	62
Figure 13	$\hat{Q}_{50}$-Duration Curve, Virginia Sites, Skew $\leq$ 5	63
Figure 14	Typical Scaled Economic Data	76
Figure 15	Price Trends for Federal-Aid Highway Construction after U.S. Department of Transportation	81
Figure 16	Ten Regions for Highway Cost Designations	82
Figure 17	SCS Land Resource Regions	85
Figure 18	Estimating Design Flow from Culvert Area	90
Figure 19	Design Flow vs. Area	91
Figure 20	Pipe Culvert Costs after ARMCO [2]	92
Figure 21	Assumptions for Box Culvert Costs	93
Figure 22	Generalized Culvert Cost Functions	95
Figure 23	Economic Estimation Scheme	102
Figure 24	Flow Duration Curve for Georgia	107
Figure 25	Flow Duration Curve for Massachusetts	108
Figure 26	Flow Duration Curve for Missouri	109
Figure 27	Flow Duration Curve for Montana	110

List of Figures (continued)

		Page
Figure 28	Flow Duration Curve for New Mexico	111
Figure 29	Flow Duration Curve for Ohio	112
Figure 30	Flow Duration Curve for Oregon	113
Figure 31	Flow Duration Curve for Tenneseee	114
Figure 32	Flow Duration Curve for Utah	115
Figure 33	Flow Duration Curve for Virginia	116
Figure 34	Flow Duration Curve for Wyoming	117
Figure 35	ρ_{50} vs. ρ_a	150
Figure 36	Active and Inactive Precipitation Gages by State and SCS Land Resource Region	155
Figure 37	Active Runoff Gages, D.A. $\leq$ 50 mi^2 by State and SCS Land Resource Region	156
Figure 38	Inactive Runoff Gages, D.A. $\leq$ 50 mi^2 by State and SCS Land Resource Region	157
Figure 39	Eleven Representative Hydrologic Regions	175

List of Tables

		Page
Table 1	Minimum Criteria for Data Collection on Small Watersheds	21
Table 2	Statistics of Q_{50} Drawn from a Typical Site	47
Table 3	Highway Cost per Mile in 10 U.S. Regions	83
Table 4	SCS Classification Scheme	84
Table 5	Culvert Density and Area	86
Table 6	Culvert Dimensions	88
Table 7	Primary Economic Data	96
Table 8	Marginal Benefits	99
Table 9	Product-Moment and Spearman Correlation Coefficients Between WRC and WRC* Estimates of Q_{50}	119
Table 10	Regression Coefficients for $\hat{Q}_{50}$ on Basin Characteristics, 11 States, Using WRC and WRC* Estimates	121
Tables 11 through 21	ρ_a Correlations between Q_a-values and Overlapping Record Lengths, 15 sites:	
Table 11	Georgia	126
Table 12	Massachusetts	127
Table 13	Missouri	128
Table 14	Montana	129
Table 15	New Mexico	130
Table 16	Ohio	131
Table 17	Oregon	132
Table 18	Tennessee	133
Table 19	Utah	134
Table 20	Virginia	135
Table 21	Wyoming	136
Tables 22 through 32	ρ_{50}, Correlations Between $\hat{Q}_{50}$-values (10-year Estimates) and $\hat{Q}_{50}$-values (25-year Estimates) Based on Synthetic Annual Floods:	
Table 22	Georgia	138
Table 23	Massachusetts	139

List of Tables (continued)

		Page
Table 24	Missouri	140
Table 25	Montana	141
Table 26	New Mexico	142
Table 27	Ohio	143
Table 28	Oregon	144
Table 29	Tennessee	145
Table 30	Utah	146
Table 31	Virginia	147
Table 32	Wyoming	148
Table 33	Summary of State-wide Flood Correlations and Skew Coefficients	149
Table 34	The Hydrometric Network	158
Table 35	Correlations Between Population Density and Gage Density, for 50 States	161
Table 36	Decision Table	162
Table 37	Modified Decision Table for Reduced Network	186

METRIC CONVERSION FACTORS

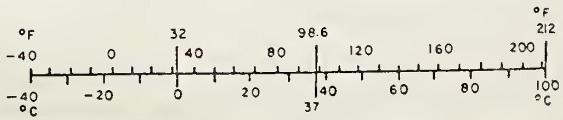
Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.

Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



Section 1

INTRODUCTION

This Technical Report, with Appendices A, B and C, is a technical appendix to the Managerial Summary, which is directed at the decision-makers in the Federal Highway Administration (FHWA) who bear the responsibility of managing programs for gaging small rural watersheds. These programs collect and analyse hydrologic data used to design drainage structures. As with virtually all problems of resource management in the public sector, the issues are not clean-cut and are subject to different interpretations and perceptions of the value of data collected under this or similar programs. Thus to make a binary decision on the basis of an objective function assumed to be appropriate to a particular agency or administrator would be a disservice to the scientific community and others who depend on routine collection of hydrographic information. Nonetheless, decisions must be made, programs must be retained or cancelled, and a whole range of sensitive political issues must be resolved on the basis of fragmentary economic and statistical information.

It therefore happens that this project was subjected to sophisticated statistical analysis because it is only through this sort of analysis that the limited data can be interrogated and extrapolated to render operational decisions. As a result of this sophistication, much of the technical manipulation is not essential for a political decision-maker or authority, whereupon the basic conclusions and documentary justifications are incorporated in the Managerial Summary which accompanies this Technical Report.

Section 2 of this Technical Report is a self-contained statement of the technical, statistical and institutional issues which together comprise the justification for this project. It contains a long summary of early and recent literature in the subject of statistical estimation of basin parameters and extreme flows, and then moves to a discussion of the current statistical techniques available for treating the instabilities inherent in estimating extrema from short hydrologic records. The results and citations are neither problem-specific nor related to any particular basin or site, but give background under which the statistical manipulations in subsequent sections are made. It emphasizes estimating procedures for parameters of skewed distributions, and indicates how statistical and economic issues can be interfaced to produce operational results.

Section 3 contains the heart of the economic and hydrologic analyses. It shows how scanty information on culvert frequency and cost is extrapolated to each of the States and how these data are used to impute benefit functions associated with reduction in design capacities of culverts.

The economic benefits of reduction in design flow are given, for each State, in dollars saved per percent reduction. Because the percentage is dimensionless, there is no question on the use of English or metric units; the tabulated values are not related to one system or another. However, some of the intermediate computations are not dimensionless and so are given in the most convenient units. For purposes of cost estimation, English units are used. Most of the hydrologic analysis is carried in logarithmic space, and these, too, are dimensionless units; the coefficient of variation is dimensionless and standard errors are expressed in log units. Thus once again, even though the basic data are in English units, the final results are expressed in percentages of reduction in design flow so that conversion of intermediate values to their metric equivalents is not indicated.

Highway bridges are not included in our economic analyses. It is assumed the drainage structures, pipe culverts, box culverts and bridges change in frequency of occurrences with increasing flows, that is, with increasing drainage areas. Bridges become more prevalent for large stream crossings. This would suggest relatively few bridges for small drainage areas; additionally, as the stream becomes larger, the probability of a gaging site on the stream and near the proposed bridge site increases, relieving the difficult task of transferring information from remote gaging sites.

To test the validity of our assumption, highway plans for projects in nine states were examined to determine the frequency of stream crossings. These covered approximately 250 highway miles. Stream crossings were divided into culverts and bridges, with culverts sub-divided into boxes and pipes. Bridges were sub-divided into classes according to length. Within the sample area there were a total of 865 stream crossings. Of these, 797 crossings were culverts and 68 crossings were bridges (92 percent and 8 percent, respectively). This percentage of bridge crossings (8 percent) is quite small.

Of the 797 culvert crossings, 121 were boxes and 676 were pipes. Of the 68 bridge crossings, 17 were greater than 200 feet in length and 51 were less than 200 feet.

Seventy five percent of bridges within the sample area are shorter than 200 feet, and below this length no generally reliable length-cost relationships could be developed. Thus 25 percent of 8 percent, or only 2 percent of all the structures in the study area, are bridges for which generalized cost relationships could be reliably constructed. This small frequency of occurrence suggests that bridge crossings can safely be ignored, given the coarseness of the economic analysis in this study. The remaining bridges occur at approximately one structure per 4.9 miles (7.84 km) and represent a cost of some $180,000 per structure or $37,000 per mile, ($23,125 per km). For example, this compares to average culvert costs (pipe and box) of $67,300 per mile for Alabama.

The results of our study indicate that due to the low cost of the stream gaging program, any region that shows an improvement in estimation of design flow (i.e., smaller variance) with longer record length is economically justified in continuing the gaging program. That is, the savings associated with smaller sized pipe and box culverts exceed the cost of gaging program continuation; bridges are undoubtedly more costly drainage devices with greater cost savings associated with them. Therefore, the results of our study are not invalidated by exclusion of bridge cost since no gaging program continuation or termination was predicated on inadequate marginal cost savings for the drainage structure but only on the capability of improving the design flow estimate by longer record length; the estimated cost savings may be conservative because of exclusion of bridges.

Most of the hydrologic analysis reported in Section 3 pertains to the calculation of regional basin parameters and regression coefficients used in the decision analysis described in Section 4. Thus the statistical material is descriptive rather than prescriptive; it does not provide decision rules but only the data manipulations and theoretical calculations required to implement the rules. The basic thrust of the argument is to show how unstable are the estimates of Q_{50} under normal conditions of record length, gage density and hydrologic variability. The necessary assumptions are defended in detail and enable us to evaluate existing gaging networks and justify the continuation or reduction of gaging programs and/or for the redirection of funds from gaging programs to analyses of hydrologic model error.

Section 4 contains the basic decision analyses under two important conditions. For each State, the existing gaging network, the regression analysis for estimating extrema from basin characteristics, and the extent to which additional sites and additional record lengths might be fruitfully used to improve estimates of Q_{50} at ungaged locations are studied. This work is contained in Table 36, which includes the analysis of 11 typical regions which together encompass all contiguous United States. It is shown that gaging extensions of five years do not generally produce increased information.

The gaging effort in each State is then limited to 25 locations in an effort to re-evalute the program with the money saved by reducing the gaging program. The results of this inquiry are contained in Table 37.

This critical limitation on the gaging program requires some justification. It was learned that the reliability of information transfer in a region does not improve significantly beyond the point at which there are 25 gages within that region. Thus a regression in a State with 50 gages would not be much more useful than a regression on 25 "independent" variables. It would generally be more effective to partition the 50 gages into hydrologically distinct sub-regions and to run regressions for each. In this manner, large States are still subject to representation by many more gages than small States but as sub-regions, each containing no more than 25 sites, the upper limit of efficient size.

Each State is represented by a single region in this report to demonstrate the methodology. This decision was based on data limitations. The analyses required annual floods and basin characteristics, and while many sites are tabulated in some form or location, not all these sites are listed on the U.S. Geological Survey (USGS) tape files which served as our basic data source. Therefore many States, including some large ones, are represented by surprisingly few complete gage records -- where completeness implies annual flood records and basin characteristics available on USGS tape files.

Subsequent applications of this methodology can utilize more complete data files, with the option of dividing States into hydrologically homogeneous sub-regions.

The information-economic evaluation problem was solved for each State rather than providing a general methodological solution. This was done because the actual solution required extensive interpolation in four dimensions from long and elaborate tables. Many look-ups were required; medians and modes were involved in interpolation routines. It required only a few hours to extend the results from 10 to 30 and then to all 48 States, and this effort was cost-effective from the Government's standpoint. Cost-effectiveness was the overriding factor in the decision to perform and summarize the analyses for all States as opposed to presenting a complicated algorithm (as suggested in the written response of the contractor to a questionnaire aimed at clarifying some items in the proposal submitted July 17, 1974).

Section 5 contains the technical information and prospectus for a set of design recommendations which should be further investigated, challenged, calibrated and tested and which might then become standard practice for the design of highway drainage structures. An early methodological investigation suggested by Harold Thomas, Jr. is updated to show its equivalence to the use of unbiased estimates of return interval of extreme events, and generalizations to network design are suggested but not elaborated.

Finally, three appendices to this Technical Report are attached; they give the statement of work, program documentation and gaging station identification.

In one other respect the original proposal has been changed; early in the study, after one meeting with regional officials, it was mutually agreed that further meetings would serve no purpose, so they were deleted. Thus the Managerial Summary does not reflect a consensus view gleaned from these regional meetings.

The objective of this study is to help field offices of FHWA define policy with regard to continuation or termination of funding for cooperative stream-gaging programs on small watersheds. The work tasks require statistical and economic measures to develop criteria for evaluation of program extension or termination; clearly an important issue is an attempt analytically to measure the effectiveness of the program and thereby to define whether or not it is worth continuing.

It has been held that an appropriate objective for the small-watershed gaging program is a gaging network sufficiently dense to guarantee that estimates made by regression of the T-year flow upon basin characteristics at ungaged sites would produce expected errors no larger than those anticipated if there were 10 years of actual record at the ungaged site. We study here the economic and hydrologic circumstances under which longer or shorter records would be appropriate to specify culvert design flows in small watersheds. In so doing, three new considerations are brought to this analysis.

First, skewness of annual floods is treated as an important statistic. It is generally agreed among hydrologists that annual floods are neither normally nor symmetrically distributed, so that a skewed distribution is appropriate. There are several candidates, including the commonly used two-parameter and three-parameter log-normal distributions, the log-Pearson distribution recommended by the Water Resources Council (WRC), the Weibull distribution, the Gumbel distribution, and what is known in this report as the modified WRC (or WRC*) distribution (a log-normal distribution whose moments are unbiased).

A consequence of attention to the skew coefficient is the acceptance of outliers, or extraordinary events which might be deleted from typical records. Unpublished USGS results indicate that in a very high proportion of short synthetic records (perhaps 30 or 40 percent of 10-year records) derived from log-normal populations with skew coefficient of the order of 5, at least one outlier was generated. This suggests the danger in suppressing or modifying such outliers so that they are brought more nearly into agreement with the other flows in the sample. The first of the unique features of this analysis is a consistent method for dealing with extreme events and their consequences. The second contribution offered here is the notion that designing for Q_T, the T-year event, is a statistical artifact. There is such an event, but we can never know it because there is no way of defining the entire population of events from which Q_T can be drawn. The distribution of events Q_T can be estimated, and it depends on the hydrologic variables and on the length of record or equivalent record at the site. The design flow must represent the economic and social issues which prevail at a site, so there must be some consideration of the risk of exposure and its economic consequence. These together enable a designer prudently to specify the design flow Q_d. To call this design flow the 50-year flow, the 100-year flow, or whatever, is immaterial; it is a label.

Third, by extending the analysis indicated in the two sections above, we use economic guidelines to define the adequacy of gaging networks and criteria for their extension or termination. Traditional techniques deal exclusively with the standard error and with "equivalent years of record." These fail properly to account for bias, skewness and other sampling problems, and do not explicitly treat economic and social considerations. Thus this last feature of the study introduces economics as an integral part of the decision-making process, not merely a component added to the analysis at its completion.

The skewness and sampling issues are addressed through the application of results recently and continuingly available from the USGS. Economic inputs to our decision-making mechanism are derived from data for a few States and Soil Conservation Service (SCS) regions, with results extrapolated to the entire nation. The effort was directed at obtaining culvert costs per square mile of drainage area for each State, from which (under some given design or decision rule) the culvert cost for the State is calculated. On the basis of additional years of equivalent record derived by regression, the confidence in the distribution of design flows would tend to increase, whereupon the design flow itself might be decreased, resulting (at a constant level of security) in a smaller culvert requirement. For each State we give the drainage cost reduction due to a unit or 1 percent reduction in design flow. This is a fundamental economic result of the work.

This reduced flow can be translated into a cost saving from generalized cost curves for the State. This benefit is contrasted to the cost of additional data collection to evaluate the gaging program.

Section 2

RESTATEMENT OF THE PROBLEM

GENERAL LITERATURE REVIEW

This review focuses on important research efforts of the Federal Highway Administration (FHWA) and the U.S. Geological Survey (USGS) relating to the estimation of flood peaks. As described in a subsequent section, the focus of this report is on the federal effort; little attention is given to research and methodology produced by the states.

Figure 1 reviews agency involvement in the estimation of flood peaks within small drainage areas. The FHWA is concerned with estimates of design flow for highway drainage, while the USGS has nationwide responsibility for hydrographic measurements so that its interest in flood peaks includes watersheds of all sizes. In the early 1960's the FHWA perceived a deficiency in the data base required for estimating rural flood peaks for the design of small drainage structures. This led to increased federal funding for securing and analysing runoff measurements on small watersheds. Program support comes from the FHWA through State Highway Departments to Geological Survey District Offices; in addition, there is direct State and matching support from the USGS. In any particular case the State and local offices are supported by a variety of funding sources, but gathering runoff data is coordinated by the District offices of the USGS.

The interactions of agency interests are shown in Figure 1. The FHWA has supported work relating flood peaks to basin characteristics of small watersheds. This work favors the insights of experienced hydrologists who apply their knowledge of watershed physics and their intuition to help define flow relationships.

The FHWA sponsored two major studies and NCHRP sponsored one to described statistically runoff and watershed data. These are:

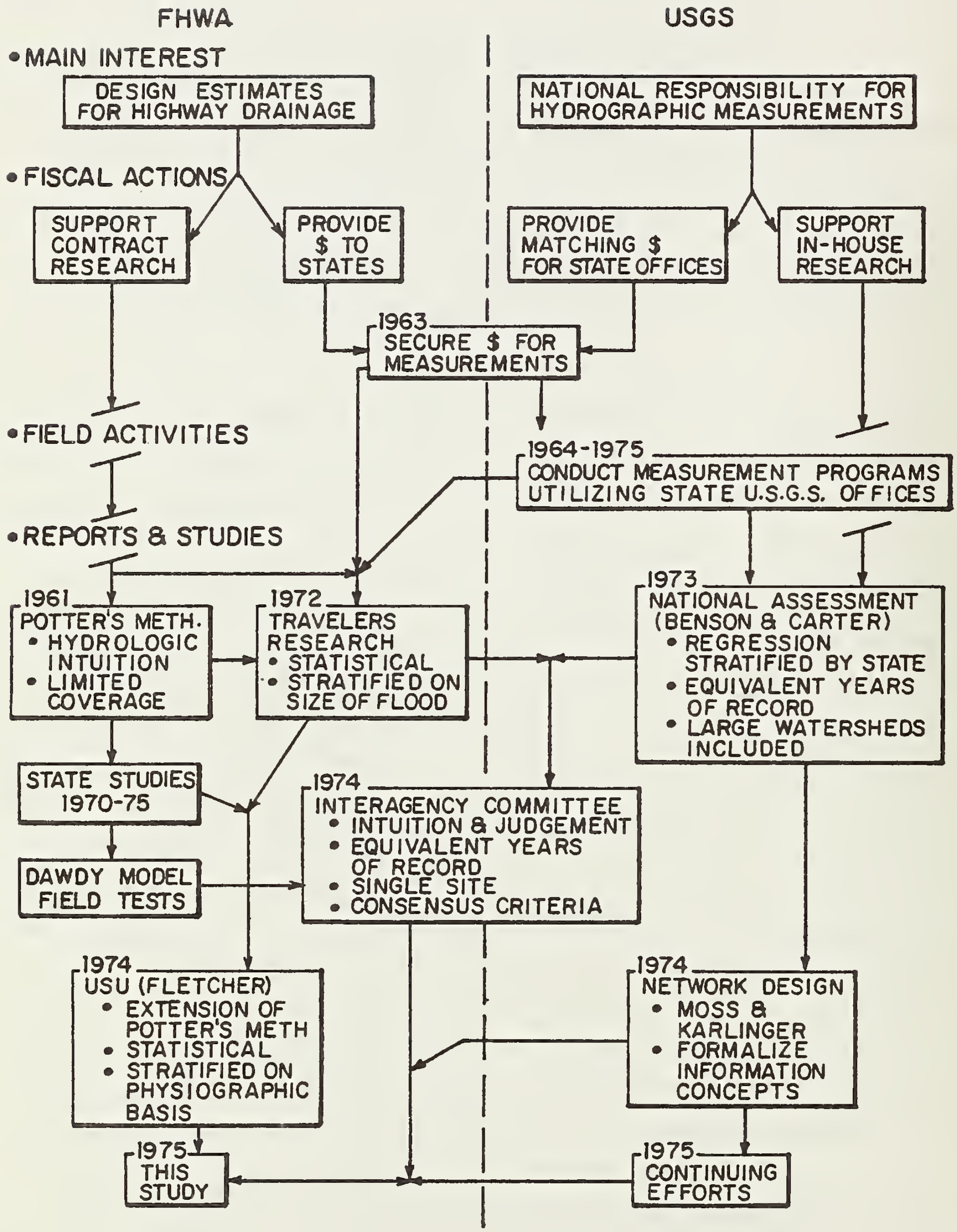


Figure 1. Review of Agency Involvement

1. Potter's Method.* A simple graphical approach is the basis of this work, performed in the FHWA by an experienced staff scientist.

2. Utah State University.** This is an extension of Potter's Method; it attempts to reduce estimating errors using an augmented data source and improved independent variables. Little structural change to Potter's basic approach is proposed.

3. Travelers Research.*** A data file for storing flood peaks and watershed characteristics is created, from which many linear and geometric regression relationships are developed and evaluated. For each of these, the correlation coefficients and standard errors are produced, giving some measure of the precision of the regression. Little attention is devoted to removing bias in the parameter estimates; the work has not been well received in the hydrologic community.

The USGS has a policy of encouraging its staff to prepare interpretive and scientific reports based on its basic mission of data gathering. The Geological Survey is also involved in estimating flood peaks on the basis of watershed characteristics, and more generally on the definition of the regional watershed parameters derived from geomorphologic and physiographic measurements. These studies are not limited to small watersheds, and have evolved quantitative approaches to the maximization of information, the transfer of information from gaged to ungaged sites, and the specification of optimal gaging networks.

---

* Potter, W. D., "Peak Rates of Runoff from Small Watersheds," Hydraulic Design Series No. 2, BPR, Washington, April 1961.

** Fletcher, J. E., et al., "Runoff Estimates for Small Rural Watersheds and Development of a Sound Design Method," Utah State University, 1974.

*** Bock, Paul, et al., "Estimating Peak Runoff Rates from Ungaged Small Rural Watersheds," National Cooperative Highway Research Program Report 136, Highway Research Board NRC, NAS/NAE, 1972.

Two recent studies by the USGS are representative of this approach. Their goals are similar to those of the FHWA in that estimates of hydrologic statistics are desired for ungaged locations. These studies are:

1. National Assessment.* Each USGS District Office developed regression relationships giving runoff as a function of watershed parameters. The USGS headquarters staff reviewed thse efforts and determined "equivalencies" among the regression relationships and actual records. Based on these, estimating errors are analyzed in an effort to formulate a rational sampling or gaging program.

2. Network Design.** Monte Carlo analysis (simulation) is used to evaluate the standard error of the network equivalencies in order to obtain unbiased quantification of criteria for the design of gaging networks.

In 1974 an Interagency Committee consisting of representatives of the Department of Transportation, Department of Interior, and other Federal organizations recommended gaging criteria based primarily on intuition and judgment expressed by group consensus. The participants were charged with considering small watersheds, for which there were essentially no design data prior to the early 1960's. A basic flaw with the report of this group*** is that it asks technologists how much data they need; the answers are predictable -- more, or much more. Little effort was devoted rationally to calculating how many more years

---

* Benson, M. A. and Carter, R. W., "A National Study of the Streamflow Data-Collection Program," USGS, WSP, No. 2028, Washington, 1973.

** Moss, Marshall E., and Karlinger, M. R., "Surface Water Network Design by Regression Analysis Simulation," WRR, 10: 3, June 1974.

*** Federal Interagency Work Group, "Hydrologic Data Requirements for Small Watersheds," U.S. Department of the Interior, December 1973.

would be justifiable; it was agreed, but not unanimously, that the equivalent of ten years would generally suffice. Some of the difficulties with implementing this criterion are described in subsequent sections; they are at the root of a partial disillusionment articulated by some State Highway Departments in measuring the value of the Federal cooperative stream-gaging program.

In specific response to the Scope of Work of this study, this section concludes with critical reviews of the Travelers Research Report and the Interagency Committee Report. Subsequent sections give critical reviews of the other citations listed above and of additional work and background for this study. There is no effort to present a chronological survey because it is not our intent here to document the development of the problem, merely to indicate its current status and to suggest some of the flavor for how we got to this point. Figure 1 shows something of the relation and history of the several major studies.

BASIC DOCUMENTS

Potter's Method (1961)

This work presents the results of research (within the Bureau of Public Roads) on runoff from small ($\leq$ 25 mi^2) watersheds each of the 105th meridian. This work led to the introduction of the hydrologic estimating procedure termed "Potter's Method."*

Potter's Method consists of the use of a series of graphs relating watershed area (A), watershed topographic index (T), and watershed precipitation (P) to an estimate of the 10-year peak flow. This 10-year peak ($\hat{Q}_{10}$) is the estimate of the peak runoff rate that may be expected to be equaled or exceeded on the average of once in 10 years. The method also presents a correction procedure for $\hat{Q}_{10}$ if the topographic index of the watershed under study differs significantly from the topographic index for the zone, or collection of watersheds on which the estimating graphs are based.

---

* Potter, W. D., op. cit.

The major problem underlying this research was development of a homogeneous data base for estimation procedure (or correlation analysis). The approach was first to divide the area under study into four zones based on the underlying lithology. Each of the four zones was then further sub-divided into physiographic areas based on Soil Conservation Service maps. This classification system formed the framework for organizing the data base.

Two hundred and forty-three (243) ungaged watersheds were classified by zone and physiographic area. Within each zone, physiographic areas are ranked according to the number of watersheds they contain; the area with the greatest number of watersheds was used for further analysis. For each zone a graphical correlation was developed for T as a function of A and P. This is written $\hat{T}_{AP}$.

A study of the error of estimate associated with the zone correlations and with the application of a drainage density variable led to the conclusion that a large error in $\hat{T}_{AP}$ indicated a watershed with different drainage characteristics, one which would need a correction to the peak discharge estimates.

The next step was to establish a sample set of gaged watersheds for each of the four zones. The candidate watersheds were initially screened to exclude those that:

1. had man-made controls;
2. had one percent (1 percent) or more of the area in lakes, swamps, or excessive floodplain storage;
3. had twenty percent (20 percent) or more of the area in urban development; or
4. had changing land-use.

Of the initial sample, 96 watersheds ranging in size from 1 to 16,000 acres and having typical natural cover were chosen for further study. The period of hydrologic record for these watersheds ranged from 6 to 38 years.

Frequency studies were performed on these watersheds to establish estimates of Q_{10} and Q_{50}.

In order to preserve homogeneity of drainage characteristics, the topographic index for each of the 96 watersheds was calculated and then estimated by T vs. A,P graphs. It was found that errors of ± 30 percent in $\hat{T}_{AP}$ had no significant effect on the magnitude of $\hat{Q}_{10}$.

Therefore, two groups of watersheds were established; Group 1, for which the error in $\hat{T}_{AP}$ was less than ± 30 percent, and Group 2, for which the error was greater than 30 percent. Of the 96 watersheds, 52 fell into Group 1 and 44 into Group 2.

The 52 watersheds in Group 1 were then placed in the proper zone, and frequency estimates of Q_{10} were graphically correlated with the watershed variables T, A and P. By employing these curves to estimate $\hat{Q}_{10(ATP)}$ for the remaining 44 watersheds and comparing the errors of estimate of $\hat{Q}_{10(ATP)}$ with $\hat{Q}_{10}$ derived from the frequency studies, a correction function for $\hat{Q}_{10(ATP)}$ was developed for watersheds with significantly different drainage characteristics. This function related $\hat{Q}_{10}/\hat{Q}_{10(ATP)}$ to $T/\hat{T}_{AP}$ for all zones.

The BPR Report presents a simplified methodology for estimating peak runoff rates from small watersheds. The simplifying assumptions are that the underlying lithology of a region is highly correlated with its physiographic characteristics and that the physiographic characteristics are highly correlated with the peak runoff characteristics. With these assumptions and limited data, graphical correlations are presented to define a design methodology.

Fletcher's Method (Utah State University, 1974)

This Report presents the results of work undertaken at Utah State University by Fletcher, Huber and Clyde.* Its objective was to revise

---

* Fletcher, J. E., et al., op. cit.

and improve the accuracy of Potter's Method. The work consisted of:

1. verifying Potter's curves by employing statistical curve-fitting techniques to increase data availability;
2. increasing the applicability of the methods by extending the geographical area on which the regressions are based;
3. evaluating the methodology for estimating Q_{10}; and
4. interviewing state agencies to determine the currently preferred design methodology.

The available (incomplete) draft form of this Report was reviewed. In this form the Report is unclear in stating the conclusions of the research effort. This review discusses what appear to be the results.

The major statistical techniques used in step 1 were the t-test for detecting significant differences and for evaluating the correlation coefficient between parameters derived by different analysts. In general, wherever comparisons are made, the t-score and the correlation coefficient are presented, but there is little or no qualitative interpretation of the statistical experiment.

The project also used Potter's watersheds and data to develop a series of least-square functions relating flows to Potter's parameters. Comparisons of the estimates made with the fitted functions and Potter's original curves showed no significant difference.

The Report suggests a substitution for Potter's topographic factor. The new factor is a slight simplification:

$$L^{1.5/\sqrt{\Delta E}}$$

where

L = length of stream channel (mi.).

ΔE = difference in elevation (ft.).

A further development was the presentation of a slightly improved correction function for flow estimates. It was suggested that the function replace Potter's "C" factor curve.

The effects of longer flow records than were initially available were examined by evaluating the t-statistics for the differences between estimates. The upper and lower frequency methods were employed to established theoretical Q_{10}. No change in estimating power of Potter's or the USU methodology was noted. It was also shown that the correlation between watershed parameters and 10-year peak flows improved with increasing record length. This is not a surprising result; increased record length will tend to reduce the noise associated with flow estimates, making the regressions more stable. Whether any of these improvements were significant was not shown.

The extension under step 2 of Potter's Method comprised two separate paths. The first was to apply the estimating procedures to states which had contained the original watersheds; the second was to extend the methodology to the entire United States.

The approach to estimating the accuracy of Potter's Method was randomly to select 25 watersheds and apply Potter's Method and three variations of the method to these watersheds. The range of accuracy was then calculated as a percent error for each of the methodological variations. The range of percent error was large in these states; no reasons are given. The only statement presented is that the error range is greater than in Potter's original work.

The task of extending the methodology to the entire United States was accomplished by adding parameters to Potter's equations and fitting a new set of curves. A total of 643 watersheds were used in this effort. No physiographic stratification was employed, and no discussion of the accuracy of the new estimating methodology is presented.

For step 3, new parameters are used in the USU estimating procedure. These are: drainage density, area of storage in watershed, 10-year 10-minute precipitation, stream length, and percent of normal annual April 1 snow-water equivalent. A series of curves are presented that estimate Q_{10} as functions of the above parameters along with Potter's original parameters (area, topographic factor and 10-year 60-minute precipitation).

The Report briefly describes the effects of the methods of flood frequency estimation on acceptance of regionalized flow estimates. The conclusion was that there was little effect on the results due to the destratification or methods chosen.

Travelers Research (1972)

With the goal of simplifying and nationalizing the estimating methodology for peak flows, a massive statistical analysis effort was undertaken by Travelers Research Corporation.* They attempted to develop better methods for estimating the magnitude and frequency of runoff from small rural watersheds (approximately 20 mi^2 or less).

Their approach was to use stepwise multiple regression analysis. Predictor variables are used regardless of independence. The data base for the regressions consisted of basin characteristics, hydrologic and climatologic factors, and physiographic parameters for 493 watersheds in the contiguous United States.

From this data base 24 statistical experiments were performed which produced 84 equations. Of these, 48 were single rational equations. The remaining 36 were sets of stratified equations. Three of these equations were suggested for use, of which two were rational equations with regional variables embedded. The third was stratified into hydrologically homogeneous regions by the magnitude of the mean annual flood.

The ability of these new equations to improve estimates of peak flow characteristics was tested by comparing them to 31 existing statewide estimates. In general there were no significant differences between the existing methods and the existing state methodologies.

---

* Bock, Paul, et al., op. cit.

Interagency Committee (1974)

This report is concerned with watersheds of less than 50 square miles; it

- reviews techniques for estimating flow characteristics on ungaged watersheds,
- summarizes the priorities for information on various streamflow characteristics,
- sets accuracy criteria for estimating flow characteristics on small watersheds,
- defines hydrologically homogeneous regions,
- summarizes existing data,
- determines minimal density for gaged watersheds by climatic region,
- tentatively identifies the number and type of additional gages needed by region, and
- estimates the total program cost.

The consensus of the committee was that regionalization provides the only acceptable process for making flow estimates at ungaged sites. No evaluation of regionalization methods is made, so that either runoff methods or rainfall-runoff modeling are deemed acceptable for estimating flow characteristics.

The report concludes that the available methods are satisfactory in transferring most flow characteristics, the major exception being low flow values. It was found that the priority ranking for data among Federal agencies is: flood-peak magnitudes, flood volumes, hydrograph characteristics and annual and mean flows.

An accuracy standard for all flow characteristics is set at 10 years of equivalent record. In setting this standard, the committee felt that estimates of flow characteristics having this accuracy would be satisfactory for use by planners and designers.

The identification of homogeneous hydrologic regions presented a complex problem to the committee. Therefore, with the time available, they adopted the Land Resource Regions identified by the Soil Conservation Service as the bases for planning zones. This classification provided 23 zones in the contiguous United States, Alaska, Hawaii, and Puerto Rico.

The report then established criteria for data collection systems. These criteria accounted for type of gage, length of record, and density of coverage. Five types of data requirements were established, and separate criteria reported for each one. Two of the data requirements were for urban hydrology and storage effects. The remaining three data requirements were for relationships between complete flow and precipitation, flood hydrograph and precipitation, and flood peak. Table 1 presents the criteria for the three major gage types.

The report inventories all existing gages by type and drainage area by region. This information, in conjunction with the minimal coverage criteria, allowed estimation of the number and type of required gages. The committee estimated that at natural flow sites there was a need for:

1. 295 complete record gages,
2. 111 flood hydrograph-precipitation gages,
3. 2,034 peak discharge gages, and
4. 1,016 precipitation gages.

The cost of installing and operating these gages to fulfill the established criteria was estimated at approximately 31.6 million dollars.

Both the USGS and the Interagency Report employ equivalent years of record as a measure of data accuracy. Research into this use of equivalent years is continuing within the USGS, and some of these results are reported and utilized here.

The 1974 Interagency Report considered basins up to 50 square miles (128 square km); this provides the basis for utilization of this basin size in this study. Another reason is that the paucity of data, severe for basins of 50 square miles or less, would be aggravated by further reducing the basin size.

Table 1. Minimum Criteria for Data Collection on Small Watersheds*

	Region of greater than 20" precipitation			Region of less than 20" precipitation		
*Basin size (mi^2)	Complete Flow & Precipitation	Flood Hydrograph & Precipitation	Flood Peak	Complete Flow & Precipitation	Flood Hydrograph & Precipitation	Flood Peak
Minimum	1	1	0.5	10	1	0.5
Maximum	50	50	50	50	50	50
*Area/gage (mi^2/gage)	≤ 2500	≤ 2500	≤ 2500	≤ 2500	≤ 2500	≤ 500
*Record length (years)	≥ 20	≥ 20	≥ 10	≥ 20	≥ 20	≥ 10

* These criteria are proposed for all regions except Alaska, where the gage density of about one-half that of other regions is proposed.

EARLY STATISTICAL ANALYSES

In the 1960's there was a flurry of activity in the application of mathematical statistics to information theory and the definition of optimal gaging networks. This was occasioned by the newborn interest in the planning and design of large-scale water resource systems. Powerful computers encouraged larger and larger models for planning. Thus it was necessary to accommodate the hydrologic information, mainly precipitation and streamflow, from a large number of gages and to analyse these so as to produce workable statistical representations of the hydrologic regime in the study area. These models were the forerunners of what has now become widely known as stochastic hydrology.

The fundamental problem, that of extending the record at a gaging station by means of correlation with a longer (but overlapping) record, has been addressed by several authors; a detailed bibliography and criticism of this early work appears in Fiering.* The basic difficulty with this early work is the assumption that population correlation and regression coefficients are known. This leads to the conclusion, obviously incorrect, that even in the absence of correlation between two overlapping records, regression does not result in dilution of useful information. The theory shows that at worst there is no improvement, overlooking the fact that in the absence of correlation the estimated missing values are pure noise so that while they increase the apparent record length, they do so at the cost of introducing very large variance. It becomes advisable to use only the actual observations without augmentation. Thus it follows that due to the correlation structure among many stations in a network, certain of these stations provide more information than others, and the intent of the analysis proposed originally by Fiering** was to identify that combination of

---

* Fiering, Myron B, "On the Use of Correlation to Augment Data," Journal of American Statistical Association, 67: 1962.

** Fiering, Myron B, "An Optimization Scheme for Gaging," WRR, 1: 4, 1965.

stations which provided the largest reduction in variance (or which minimized the residual or unexplained variance) in the network. This is the notion of relative information first introduced by Thomas* and subsequently used widely by Matalas and Langbein** and others.

The work by Matalas and Langbein began specifically to relate concepts of information transfer among a number of correlated gaging stations in a region in an effort to develop regional parameters for use at ungaged locations. The value of sample information was given as a function of serial correlation among the observations at given locations. Matalas subsequently generalized and extended the work of Fiering, and in a well-known paper*** gave an application of the theory to a stream gaging network.

More recently, Maddock[+] implemented the non-linear program for gaging station location which appeared in the original Fiering citation, and showed how, for a range of budgetary constraints and objectives (i.e., estimating first the mean and then the standard deviation) a range of different gaging programs could evolve.

The decision theory literature contains many references to the specification of optimal programs for data collection, but the applications do not readily fall into network-type problems. Much of the

---

* Thomas, Harold A., Jr., unpublished memorandum, Harvard Water Program, 1958.

** Matalas, Nicholas C., and Langbein, W. B., "The Relative Information of the Mean," JGR, 67: 9, 1962.

*** Matalas, Nicholas C., "Optimum Gaging Station Location," Proc. IBM Symposium on Water and Air Resource Management, IBM, Yorktown Heights, 1967.

\+ Maddock, Thomas, III, "An Optimum Reduction of Gages to Meet Data Program Constraints," Bull. Hydrological Sciences, XIX: 3.

original work has been summarized in Raiffa and Schlaiffer,* which develops a calculus for information collection. The approach is Bayesian, and has led to the introduction of acronyms such as EVPI (Expected Value of Perfect Information), EVSI (Expected Value of Sample Information), and similar expressions. The basis of the "value" computations inheres in the economic benefits associated with more appropriate actions or strategies derived from better information, recognizing that the incremental information can be based on direct observation or on Bayesian regression estimates. Raiffa and Schlaiffer provide tables which guide the specification of optimal sampling programs under a variety of prescribed conditions involving imperfect information on the population variances.

A report in the Hydrology Series of the Colorado State University** deals with rainfall-runoff relationships for very small drainage areas, many of which are less than one square mile. Five methods of flood prediction are appraised, concentrating on generally accepted formulae. It was found that results varied widely, which is not surprising. But no effort was spent on the statistical issues of bias, information, model error, sampling error, and related phenomena which are central to this work. A later report in the same series*** draws an important distinction between estimating specific floods from specific rain storms and defining design criteria from rainfall statistics. The study proposes a single parameter for expressing the time-distribution

---

* Raiffa, Howard, and Schlaiffer, Robert, Applied Statistical Decision Theory, (Harvard University Press, Cambridge), 1961.

** Hiemstra, Lourens, and Reich, Brian, "Engineering Judgment and Small Area Flood Peaks," Hydrology Paper No. 19, Colorado State University, April 1967.

*** Bell, Frederick C., "Estimating Design Floods from Extreme Rainfall," Hydrology Paper No. 29, Colorado State University, July 1968.

introduced by the watershed; this parameter is called the _representative lag_ and is related to the volume/peak ratio. Some generalizations are drawn concerning the equality of return periods for design floods and the corresponding extreme rainfalls. There is no mention, of course, of the (much later to be discovered) issues of bias in the process of attaching return intervals to extrema. In any case, estimation of 10-year events is shown to be poorly validated and subject to large sampling fluctuations.

U.S. GEOLOGICAL SURVEY PROGRAMS

Basic Reports

In 1970 the USGS initiated an evaluation of its program for streamflow data. The results of this survey are reported by Benson and Carter.* The report first sets forth a framework for categorization of uses of streamflow data. Four main categories are presented:

1. data for current use,
2. data for planning and design,
3. data for definition of long-term trends, and
4. data on the stream environment.

Category 2 was further divided into streams with natural flow and streams with regulated flow, and further divided into minor and principal streams. A minor stream is defined as a stream which has a drainage area of under 500 square miles. All other streams are principal streams.

With this framework, the report sets forth goals for the data program within each category. The objective is to establish the purpose and accuracy limits for the "information on the flow characteristics at any point on any stream" within each category.

---

* Benson, M. A. and Carter, R. W., _op. cit._

The program goal for category 1 -- current use data -- is to provide the particular information needed at specific sites for designated current use. Data within this category are generally used for operational decisions and thus may require a high degree of accuracy; therefore, due to changing demands, a collection network was not amenable to optimal design.

The program goal for collecting planning and design data is to define (within given accuracy) the statistical flow characteristics for all streams in the country. Because flow characteristics on ungaged streams must be estimated by a form of regionalized analysis, an accuracy goal on these estimates could be set. These were established in terms of equivalent years of record. This criteria specifies that "information provided for any ungaged point on a stream should be equivalent in accuracy to that which would have been attained by an actual record of some number of years at that point." Since it is possible to convert accuracy goals measured in terms of equivalent years of record to standard errors in percent of the mean, it is possible to establish accuracy goals for a given region from the coefficient of variation within a region. The accuracy goal for minor streams was set at ten equivalent years of record, and that for principal streams at 25.

The goal of the program for collecting data for analysing long-term trend is to operate indefinitely a representative sample of gaging stations on natural-flow streams in each region of the country, thereby to provide a continuing series of consistent observations. It was estimated that approximately 100 stations would provide the required information if two long-term gages were operated in each of the subregions of the United States, as defined by the Water Resources Council.

The goals of the data for stream environment, and the necessary accuracy, are set according to specific needs in the area. The report evaluates data currently available with the major portion allotted to data for planning and design. It was found that over half the ongoing streamflow data program is related to collecting data for current-use, and, in general, that requirements for these data were being fulfilled. The report presents no final evaluation of data for long-term trend or stream environment.

Evaluation of the data base for planning and design rested on the ability of the data to allow for regionalization of streamflow estimates by multiple regression. Flow data were employed to derive flow statistics which became the dependent variables in regressions on the basin characteristics. The flow-frequencies were defined by fitting a log-Pearson Type III distribution. The statistics developed for flood-frequency analysis at each site were limited to flows at recurrence intervals less than twice the record length of the site.

The general findings were:

1. Some or all accuracy goals were met, principally in the eastern half of the country.
2. Few or none of the goals were met in the western half of the country.
3. Regionalization was not applicable on principal streams. A network was established to allow for interpolation or modeling for flow estimates between gages on principal streams.
4. Accuracy goals for low-flow estimates were not met in any locality.
5. Deficiencies exist in information on small streams and on streams under urban conditions.

Evaluation by the USGS was a nation-wide examination of large drainage areas (> 50 mi^2). The need existed for the data network on smaller watersheds. This work was undertaken by the Interagency Advisory Committee on Water Data under the Office of Water Data Coordination of the USGS.

Equivalent Years of Record

The concept of equivalent years was introduced by Hardison* and represents a convenient way to measure the reliability of information at a site.

> The equivalent years of record at an ungaged site is the length of record which would be required at that site in order to produce parameter estimates which are equally reliable (that is, which have the same standard error of estimate) as those estimates which are made by transferring information through the use of a mathematical model from gaged sites elsewhere in the region. Because of the dependence on the standard error of a particular parameter or statistic, the equivalent years of record is a function of the parameter under estimate.

In the early work by Hardison, the equivalent years of record was shown to have properties which have been modified by the more recent work of Moss and Karlinger. They showed that the original Hardison definition contained certain biases, and they indicated how these biases might be compensated; massive Monte Carlo analyses were performed and the results were summarized in the engineering literature; tabular abstracts will be made available. In fact, most of the Moss-Karlinger work was published after the proposal and the Scope of Work for this contract were prepared, so the FHWA is in the position of being the first agency to apply this major theoretical advance in an important decision problem.

The basic concept in the early studies of network design, as expressed in some of the papers by Fiering,** Matalas,***

---

* Hardison, Clayton H., "Accuracy of Streamflow Characteristics," USGS Prof. Paper 650-D, 1969; Hardison, Clayton H., "Prediction Error of Regression Estimates of Streamflow Characteristics at Ungaged Sites," USGS Prof. Paper 750-C, 1971.

** Fiering, Myron B, "On the Use of Correlation to Augment Data," op. cit.

*** Matalas, Nicholas C., "Optimum Gaging Station Location," op. cit.

Benson,* Carter,** Hardison,***, and others, and reinforced by the recent work of Moss and Karlinger,+ is the development of a mathematical model which relates measured parameters to some desired flow statistic. In all this work, as in the case of the small watershed program which is the focus of this study, there is postulated the existence of a regression model of the form

$$y = a x_1^{b_1} x_2^{b_2} \ldots x_p^{b_p} \tag{1}$$

in which y is the symbol for some output statistic (such as mean annual flow, T-year flood, T-year low-flow, etc.), the x_i are basin characteristics (such as drainage area, channel slope, precipitation intensity, soil index, etc.) and the a_i and b_i are coefficients of the estimating equation derived by least-squares or some other suitable technique. This functional form suggests an underlying linear relationship among the logarithms of the dependent and the several independent variables. This assumption represents a strong consensus reached by many investigators; it is not questioned here whether other functional forms are more appropriate because there seems to be little doubt that an exponential relationship is appropriate in the majority of basins.

The purpose of such a regression relationship is to enable designers to estimate design flows or other useful statistics at locations for which no flow measurements are available. Typically the independent variables, or x_i, can be measured at a site or estimated by examining maps, geological evidence, or other readily available sources of meteorologic data. It does not require many years of observation to produce

---

* Benson, Manuel A., "Factors Influencing the Occurrence of Floods in a Humid Region of Diverse Terrain," USGS, WSP 1580-B, 1962; Benson, Manuel A., "Factors Influencing the Occurrence of Floods in the Southwest," USGS, WSP 1580-B, 1962.

** Benson, M. A. and Carter, R. W., op. cit.

*** Hardison, Clayton H., USGS Prof. Papers 650-D and 750-C, op. cit.

\+ Moss, Marshall E., and Karlinger, M. R., op. cit.

these parameters, and therein lies the difference between measuring the parameters x_i and estimating the flow Q_T from sequences of actual data.

The coefficients of the regression relation represent regional characteristics, and clearly it is important that the gages be carefully chosen to assure that the coefficients are representative of those ungaged sites to which the regression might be applied. A State or region might appropriately be further divided into physiographic sub-regions according to geological factors, and a number of such relationships could be derived. This is suggested by the conclusion that more than 25 gages in a regression set can not contribute significantly to the collection and transfer of information. Thus many small sets or sub-regions are statistically more effective than a few large ones.

It is also important to recognize that large drainage basins behave differently than small ones, and that the difference is not always suitably accommodated by the inclusion of drainage area as one of the independent arguments x_i. In other words, there is reason to believe that the mechanism of watershed drainage changes appreciably with large and small drainage areas, so it is important to develop different regression relationships for each. In fact, there is strong evidence as reflected by significance testing on the coefficients themselves which suggests that different combinations of independent hydrologic variables are important in small and large drainage basins.

RECENT NETWORK THEORY

Sources of Error in Regression Models

The adoption of a regression model implies the acceptance of three sources of error, all of which are important to this analysis. First, there is time or sampling error. This would exist even if measurements were made directly at the site; it is that error due to finiteness of the record. Even if measurements were perfect, and contained no systematic or equipment error, it is clear they are derived from a small "window" on a long and continuing process. Therefore there remains the uncertainty associated with the fact that the measurements are made over a limited time horizon.

It is a common misconception that sampling error is serious only when applied to time intervals of 500-1,000 years; in fact, even if the 50-year event can be defined, the probability of x occurrences during (say) 100 years is a very flat function for small x, indicating substantial instability. For x = 0,1,2,3, the probabilities of precisely x floods (or 50-year events) in 100 years are: 0.13, 0.28, 0.28 and 0.18. For the 25-year flood, the equivalent probabilities are 0.02, 0.07, 0.14, and 0.20. The sampling errors typically associated with "p," the probability of a flood (or success) in any year, are so great as to render observed flood frequencies virtually meaningless as estimators of p, even if the "window" of observation represents a substantial fraction of the projected economic life. In other words, the two strings of probability values calculated above become indistinguishable, and the true or parent p is obscured.

Second, there is model error. This is perhaps the most important component of error in our study because it reflects the fact that the regression function may not be the best form for transferring information from the gaged sites to some ungaged location. The early works by Fiering and Matalas show the extent to which noise enters regression equations, and the consequence of that introduction in terms of standard error of the dependent variable. There may be other functional forms better than the exponential, and more importantly, there may be significant variables other than those actually retained by the estimating procedure. Of course, even if we had exactly the right causal model, and even if the correct set of independent variables were arguments in the model, measurements on each of those independent variables would themselves be subject to time error. It is generally impossible to discriminate between sources of error and to determine how much of the total error in the estimating equation can be traced to imprecisions in the model as opposed to unreliability in measurements of the independent variables themselves.

It is recognized that it is inadequate merely to demolish old techniques without suggesting replacements; this is handled in our final recommendations.

There is a third component of error; it is the spatial error associated with the fact that even if there is no time or sampling error, and even if the model includes the correct independent variables in the correct causal functional form, the array of independent locations may not be correct. That is, the model and the measurements on its parameters may be exact, but if the set of gages is not the unique set required to transfer regional information to the ungaged site, there will be an error in specification of the output variable y. This third source of error can not readily be distinguished from the other sources, so only gross estimates of the assignment of error to each of its three compartments can be made.

The point is to suggest that there might be so much noise attributed to the model error and its compounding by sampling and spatial errors that one might be better advised to use whatever data or estimating techniques are at hand and not further complicate the matter by introducing noise associated with transfers from other records. This procedure was first discussed by Fiering* in 1960, who showed that for purposes of estimating the mean and standard deviation of annual flow at a station it might sometimes be better to use an existing short record at that station than to augment the record by correlation with a neighboring gage; the criterion was the relative information of the parameter under estimate, or more precisely, the variance of that parameter using the regression as compared to that using the existing record alone. Correlation, if not strong, can add more noise than can be accounted for by the increased record length, and therefore it is not necessarily a useful technique. The criterion for including regression estimates becomes more severe for higher (i.e., more unstable)

---

* Fiering, Myron B, "Statistical Analysis of Streamflow Data," Ph.D. Dissertation, Harvard University, 1960.

moments; thus to estimate the variance requires better correlation than to estimate the mean. By induction, even stronger correlation is required to estimate Q_{50}.

The cost of collecting information and of transferring that to an ungaged site is measured along two axes, a monetary cost associated with the data collection and a statistical cost associated with the noise inherent in the three sources of error. If the value of the hydrologic information does not exceed the cost of obtaining it, the collection program should be abandoned.

Such abandonment would not imply that all hydrologic enterprise in the basin should be terminated, because significant improvements in estimating might be achieved through improvement of the model itself. This would redirect funding from data collection and manipulation to the development of a better understanding of the causal mechanisms and fundamental hydrologic relationships which govern the hydrology of extremes in those basins. The cure for inadequate data is not necessarily the collection of more data, but in some instances might be the development of better mechanisms for extracting information from the data already at hand. One of the conclusions to be drawn from this study is a procedure for making this distinction in a small watershed.

BIGBASIN

Moss and Karlinger* published an important paper whose analysis allows for the systematic evaluation of more gages and longer records. In other words, it offers formalisms for parsing the total error of estimate into its constituent parts. Their paper expands on the concept of equivalent years of record applied to gaging networks as a standard of accuracy for single stations. The basis of the analysis

---

* Moss, Marshall E., and Karlinger, M. R., op. cit.

is that a measure of regional information on a streamflow parameter can be approximated by the standard error of estimate of the regression analysis used to estimate the parameter. This estimate can be expressed in equivalent years of record.

The equivalent years of record is a random variable because it depends on sample statistics. Since the streamflow parameter estimates contain time (sampling) error and the associated equivalent year measure is a non-linear transformation of the standard error of estimate, it is unlikely that the equivalent year measure is a consistent unbiased estimator of regression accuracy. Monte Carlo simulation with regression analysis was employed to explore the statistical nature of equivalent years of record as a statistic, and thus to estimate its sampling properties.

A model is proposed that relates a streamflow parameter to drainage area (including a random component). The logarithms of the set of admissible areas in physiographic region are assumed to follow a rectangular distribution. A multivariate Markov streamflow generator is imposed to synthesize hydrology for hypothetical basins.

With this formalism it is possible to estimate two values of the equivalent years of record. The first is called the apparent equivalent years ($\hat{Y}$), whose estimate is developed by employing the standard error of estimate of the regression and simulation results. The second estimate is considered the true (or unbiased) best estimate of the equivalent years (Y). This is calculated by employing the standard deviation of the prediction errors in the cascade of equations which define equivalent years. The estimate of the expected value of true equivalent years is based on the entire population of drainage basins in the region whereas the estimate of the expected value of apparent equivalent years is based only on those sites used in the regression analysis. Therefore it is stated that the true equivalent years is a better (unbiased) estimate of the information content of the regression analysis.

The effects of varying the number of basins and length of record were studied; the anticipated result was that increasing either the number of basins or record length would increase the equivalent years of record. This was true for equivalent years but was reversed for apparent equivalent years.

Further analysis of the regression results led to the following conclusions:

1. Y is conditionally independent of $\hat{Y}$;
2. the marginal distribution of Y may be approximated by a β distribution; and
3. the marginal distribution of $\hat{Y}$ may be approximated by a γ distribution.

The presentation of apparent equivalent years of record as a random variable and its relationship to the true information content in the regression relationships leads to an approach for network design based on the confidence level desired among estimates of the true equivalent years. The USGS has adopted this method of network analysis, and has available a series of tables to assist network designers. These are tabulated as outputs from the programs identified as BIGBASIN and WORLDWAR I, from which the records are used to derive unbiased estimates of the moments of the distribution of Q_T. The program which performs this analysis is known as WORLDWAR I, which deals with observations at a single gage, not with networks. It is necessary therefore to extend the results of WORLDWAR I to apply to multiple sites in order to specify culvert design flows at locations where no gages exist (and to which information must be transferred from other locations). The program known as BIGBASIN can be used to evaluate the effects of networking. Moss has prepared a manual to help implement the technique.*

---

* Moss, Marshall E., "Design of Surface Water Data Networks for Regional Information (Technique Manual)," draft USGS Memorandum, 1975.

The effects of networks are built into the decision process through the equivalent years of record. The effect of a network, and of additional information obtained on that network, is generally (but not always) to increase the reliability of results at an ungaged site. This is done by sharpening the parameters of the distribution of Q_{50} so that estimates of the moments have the same reliability as those derived from a longer record. The relevant question is whether or not the incremental length of equivalent record is great enough to reduce the standard deviation s_{50} to the point at which enough culvert cost can be saved to justify continuation or extension of the network. Additional information, whether obtained directly at the site or transmitted through regressions, reduces sampling errors, but not necessarily to the level where additional collection costs can be justified.

A document* made available in draft form during the course of this study represents an effort undertaken by the U.S. Water Resources Council to identify a uniform technique for selecting the proper distribution to assign to flood events and thereby to determine flood frequencies. It was correctly noted that there was disagreement among many agencies, consultants and individual authors as to the best distribution to assume for flood data, and further, as to the correct way to estimate flood frequency parameters from the available data. A Uniform Technique for Determining Flood Flow Frequencies is an attempt to impose a single methodology so that all analysts confronted with the same data would develop the same flood-frequency curve. But the actual distribution may not matter very much from a decision viewpoint.

---

* U.S. Water Resources Council, A Uniform Technique for Determining Flood Flow Frequencies, draft report, December 1974.

In two pioneering studies, Slack, Wallis and Matalas* investigated the economic consequences of using different distribution functions for decision-making in hydrologic problems for which the underlying population is known. Depending on the population skew coefficient, the normal distribution appeared extremely robust and became less desirable only as the skew coefficient became better identified. Thus over a wide range of hydrologic uncertainty, given the difficulty associated with estimating the skew coefficient in the first place, and for a range of economic parameters, the normal distribution is extremely robust. When the skew coefficient is known within broad ranges, other distributions may become more appealing. It is well known that distributions of annual floods and fractiles such as Q_{50} have significant skewness, so the normal distribution might subject the analysis to valid criticism. We therefore select the log-normal density as appropriate for all the flow distributions, but note that trial calculations using the normal distribution do not produce significantly different results.

WORK DONE BY THE STATES

Within the past few years all the States have undertaken to prepare regression analyses in the spirit of this project, hoping to develop the coefficients whereby information could be transferred from gaged to ungaged locations. It is unnecessary to report on this work in great detail because there is nothing particularly significant about one set of regression coefficients as opposed to another; the important thing is the extent to which the States utilized the gaging information developed through the cooperative programs and the reliability placed by each of the States in the design flows which are deduced from their

---

* Slack, J., Wallis, James, and Matalas, Nicholas, "On the Value of Information to Flood Frequency Analysis," WRR, 11: 5, October 1975; Matalas, Nicholas, "A Mathematical Assessment of Synthetic Hydrology," WRR, 3: 4, 1967.

relationships. No important theoretical advances are offered by the States, most of which have routinely followed the form of analysis prescribed by Benson and Carter* and Hardison.** Only three States deviated from using routine USGS analysis: Alabama and Missouri made a significant breakthrough by issuing the first reports on the operational use of the small streams rainfall-runoff model. The investigators tested the Dawdy Model under field conditions. Wyoming used the model to establish a relation between peak discharges and volumes.

In addition, many of these states have prepared special reports dealing with hydrology, floods, rainfall-runoff relationships and other special features unique to their problems. Typical of reports that treat these issues are studies prepared by New Jersey,*** Texas,+ and Nevada.++ This list is merely representative, not exhaustive; many States issue special reports on environmental and hydrologic studies.

Field design practice has undergone a slow change over the years. For a long time the Rational Formula and its modifications were the basis of culvert design. More recently frequency curves developed by the USGS have been used, and the cooperative gaging programs have served to express designers' aspirations concerning the statistical validity of these curves.

---

* Benson, M. A. and Carter, R. W., op. cit.

** Hardison, Clayton H., USGS Prof. Papers 650-D and 750-C, op. cit.

*** State of New Jersey, "Magnitude and Frequency of Floods in New Jersey with Effects of Urbanization," Special Report 38, Department of Environmental Protection, with USGS, 1974.

\+ Texas Board of Water Engineers, "Texas Stream Gaging Program: Evaluation and Recommendations," with the USGS, October 1960.

++ Moore, Donald, "Estimating Mean Runoff in Ungaged Semi-Arid Areas," State of Nevada, Department of Conservation and Natural Resources, Water Resources Bulletin No. 36, 1968.

POSITION HELD BY THE STATES

State highway officials, appropriately enough, are concerned with criteria and guidelines for designing drainage systems; they are not principally concerned with statistical research, meteorological refinement, and the subtleties of regression errors. At the same time, the USGS is concerned with data collection and with the scientific underpinnings of network analysis. Thus it was reasonable that a cooperative program be initiated in the hope that the data themselves would provide the basis for theoretical analyses of interest to the Survey, and the actual design rules of interest to the States.

Our few interviews with State highway officials indicated their partial discontentment with the gaging program, even though they continue to participate. No simple rules or solutions have been advanced; of course, given the complexity of the problem, this is not surprising. The Geological Survey has made major theoretical advances but has not provided methodology readily to identify specific design flows. In the section on Recommendations, it is suggested that State Highway Departments utilize culvert performance data as a basis for a new design methodology.

A REVISED STATEMENT OF THE PROBLEM

An essential feature of drainage design is risk aversion. We assume that the design criterion for some hydraulic structure is the T-year event. It is impossible to define the T-year event; the best that can be done is to develop a reliable estimate. Hydrologic records are short samples of processes which have been continuing for millennia. There is no evidence to prove that these processes are stationary, and indeed there is some accumulation to support that at least many of them are not. Rivers meander, they deposit and scour their channels, they construct deltas, they flood and deposit sediment in lowlands, and generally change the landscape. The basic driving forces, which include precipitation and other meteorologic features, are non-stationary because they are subject to long-term climatological fluctuations and cycles and to shorter term perturbations. It is thus naive to suggest that any available hydrologic record is long enough to capture the richness of the hydrologic potential in a region.

Some parameters of the sample (or hydrologic record) are reasonably good estimators of their population counterparts. These include measures of central tendency, such as the mean and the median, and to a lesser extent, the second moment or variance. Higher moments are so unstable as to be virtually impossible to estimate reliably from customary hydrologic sample lengths. It follows that extreme events are particularly susceptible to sampling error; estimates derived from records, even from impressively long ones, are unstable. No amount of extrapolation no amount of massaging or manipulating data, can overcome the fact that extrema are elusive statistics, and that their proper estimation requires a healthy respect for their instability. This instability is typically measured by the standard deviation of estimates of the extreme event.

Suppose we have an estimate of the distribution of the T-year flow. We call this $\hat{Q}_T$, and in this study we take T = 50 although other values often are used in design. Every sample drawn from the population of annual floods will yield a different estimate of the T-year event; Q_T is subject to the vagaries of sampling error. If a large number of samples is available, and if a new $\hat{Q}_T$ is estimated from each sequence, it is possible to estimate a distribution of $\hat{Q}_T$ estimates. It is important to note that the characteristics and parameters of this distribution are strongly dependent on the length of record (N_Y years) from which $\hat{Q}_T$ is estimated, and longer records will have <u>better</u> estimates (in the sense that they are more stable) of the <u>true</u> or population value of Q_T. Longer records generally have smaller standard errors. This in no way implies that the true or population value of Q_T is necessarily closer to the expected value of the distribution based on a long record than to the expected value based on a short record. All we can say is that the reliabilities of the two results are typically different.

Suppose a very long record of annual floods is available -- say 200 years. It is divided into 20 sequences of 10 years, and each sequence is the data base for calculating an array of statistical parameters which together describe the sequence. If only one 10-year sequence were available, the estimates calculated from its elements would serve to estimate the parameters of the parent or population of annual floods. In this case, 20 arrays are calculated (Figure 2).

Q_T can be estimated in several ways. First, the distribution function of annual floods can be drawn from 200 pooled values, from which $\hat{Q}_T$ can be read as shown in Figure 3. Alternatively, each 10-year sequence can be used to estimate $\hat{Q}_T^i$, where the superscript identifies the sequence number i = 1,2,...,20. That is, each of 20 10-year traces yields a different $\hat{Q}_T^i$. Moreover, because the plotting postion of the largest flood does not extend to the 98 percent-event (T = 50 years), it follows that estimation of $\hat{Q}_T^i$ is an unstable process strongly dependent on assumptions concerning extrapolation of the distribution function beyond the range of observations. Two fitting techniques are contracted in this study. A typical distribution function, one of 20 possibilities, is shown in Figure 4.

Twenty estimates $\hat{Q}_T^i$ are drawn; which is "correct?" Which is "most likely?" All 20 are plotted in Figure 5.

Another alternative is to impose a specified probability density on annual floods, and to calculate $\hat{Q}_T$ from tables of that density, using moments of the annual events. This is shown in Figure 6.

All these techniques are (more or less) defensible, but <u>none</u> answers the following design question: What is Q_d, the design flow, to be? The fundamental contributions of this study are:

1. explicit recognition of the statistical uncertainties described above;
2. generalization of these to the case in which data points are not available except by transfer from remote sites (with consequent additional loss of reliability); and

Figure 2. Record of Annual Floods

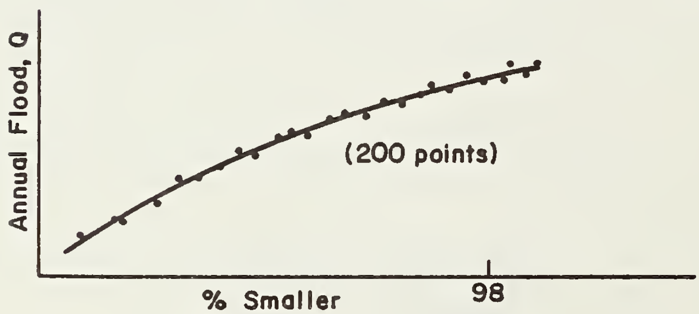


Figure 3. Distribution
Function for Pooled Annual Floods

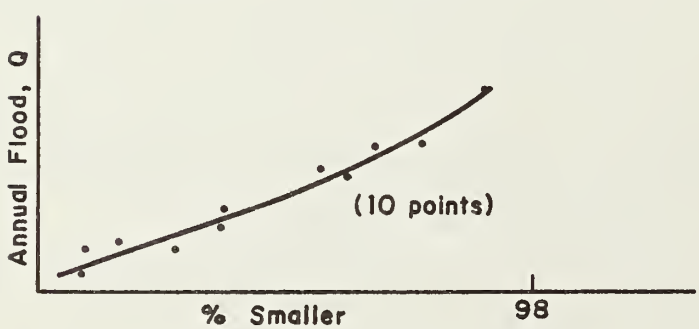


Figure 4. Distribution
Function for Segment of Record

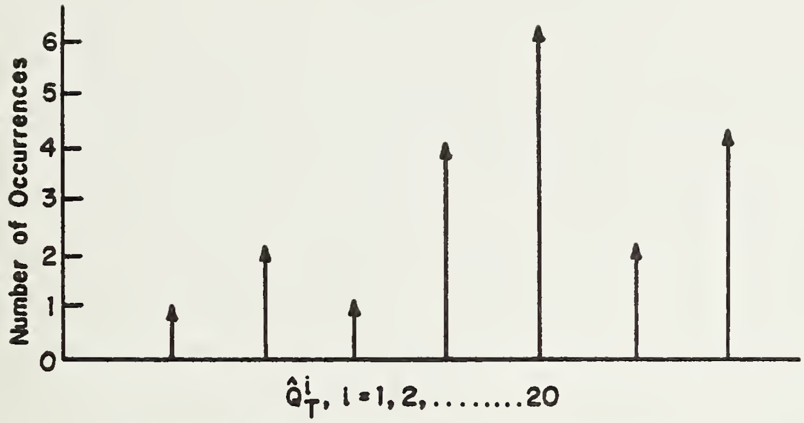


Figure 5. Histogram of Estimates of Q_T

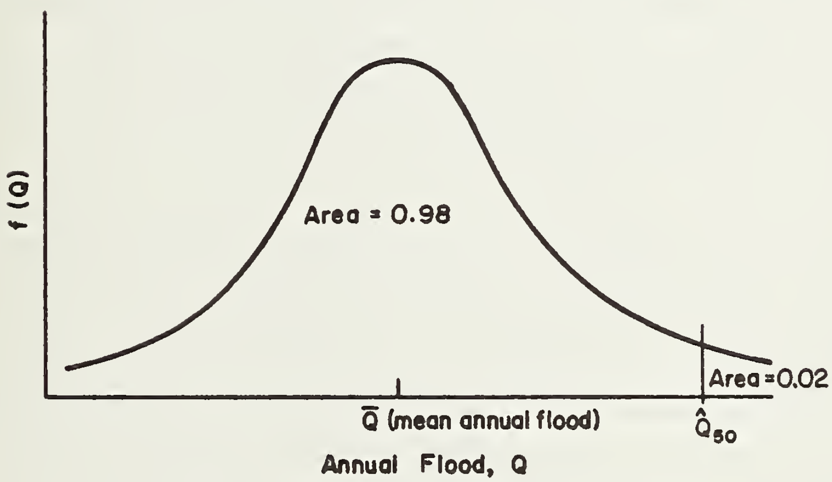


Figure 6. Theoretical Density Function for Annual Floods

3. introduction of economic criteria to help define the design flow and to make subsequent decisions on gaging strategy.

In other words, we deal with the distribution of $\hat{Q}_T$ rather than with some presumed scalar or representative quantity.

Let it be required to build a culvert whose capacity is to be the expected or average value of Q_{50}, and suppose that the distribution of Q_{50} is estimated so that it is possible (numerically or analytically) to calculate the expected value $\bar{Q}_{50}$. Assume that an alternative to building the culvert is to delay construction until additional gage measurements are available at the site, whereupon it might be possible to reduce the standard deviation of Q_{50} without necessarily changing its expectation. Assume that at the time of decision, it is not known to what extent the new data would change the expected value $\bar{Q}_{50}$. All one can say about the two alternative data bases is that the second (or delayed) estimate would have a smaller sampling variance or standard error of estimate.

If it is presumed that the expected values of both data bases or densities are identical, the only advantage in continued gaging comes from the greater confidence in the delayed estimate. If risk is not an issue in the design, <u>more confidence</u> can not be shown to be worth the cost of delay and more data collection.

On the other hand, if we add the criterion that the design flow should be $\bar{Q}_{50}$ plus some amount which encompasses a given percentage of the distribution of $\hat{Q}_{50}$, the peakedness or tightness of the distribution becomes important. For example, if the culvert is to be built in an extremely critical region for which flooding would be very expensive, the system should pass some high percentage (say 90 percent) of all <u>potential</u> events $\hat{Q}_{50}$. In other words, the design flow would be that flow larger than 90 percent of all potential events $\hat{Q}_{50}$ which define the distribution. The expected Q_{50}, written $\bar{Q}_{50}$, is not severe enough for design. If it appears that the distribution is symmetric (which it is not, as detailed in later assumptions), 50 percent of all events

Q_{50} are smaller than the mean $\bar{Q}_{50}$, so it is necessary to increment the design flow to include an additional 40 percent of all points in the distribution. The right-hand tail of the distribution then includes 10 percent of all flows, so there is a 10 percent chance that the design criterion will be exceeded.

This is not to say that the culvert system will be exceeded 10 percent of the time (or one year in ten). Recall that the event Q_{50}, if we were to know it exactly, would be exceeded on the average once every 50 years. Most designers will make allowance for this threat of economic disaster by adding a margin of safety to Q_d. Sometimes this is done implicitly rather than explicitly -- the choice of parameters can be subtly shaded. Selecting a design flow well out on the right-hand tail of the distribution of all events Q_{50} makes the design criterion even more conservative by selecting an event whose expected return interval is greater than 50 years. Thus there are really two levels of security in the specification of Q_T. The first is inherent in selection of T (or 50) years as the design criterion. This says something about the extent to which culvert failures can be tolerated. The second level of security lies in the confidence in specifying Q_{50}; it is this second level of security to which the gaging program is directed. Under certain assumptions concerning the distribution of Q_{50}, it is possible to estimate the return interval for which the design flow (specified as $\bar{Q}_{50} + \alpha s_{50}$) is chosen. In this equation, α is a parameter which represents a level of security or risk aversion; for $\alpha = 0$, we say that the decision-maker is indifferent to risk. For $\alpha > 0$, the decision-maker is risk averse; for $\alpha < 0$, the decision-maker is risk prone. The symbol s_{50} is the standard deviation of the distribution of estimates $\hat{Q}_{50}$.

Thus the efficacy of a gaging program, which lies in reducing s_{50}, must be measured in terms of a design criterion which in turn encompasses a parameter of risk aversion; this is identified as α. In engineering jargon, the additional carrying capacity imposed on the system (positive values of α) is a safety factor. For non-symmetric distribution of $\hat{Q}_{50}$, the mathematics becomes more troublesome but the basic explanation and motivation remain unchanged.

A few simple experiments illustrate these points. USGS Station 05014500, Montana, has a record length of 61 years. It is assumed throughout that annual floods are independent events, so their serial correlation is identically zero. Sampling with replacement, 100 random sequences of length 5, 10 and 25 years were drawn. From each of these 300 sequences, an unbiased estimate of Q_{50} was made using the latest USGS scheme for this calculation, after which the mean, standard deviation and extrema of each set (corresponding to each record length) were calculated and tabulated in Table 2.

These results are similar to, but less variable than, those of Moss and Karlinger. They do not contain model or spatial error because there is no transfer of information (regression) from one site to another; the only error is sampling error. And even this source of error is truncated because all random sequence of flows are drawn from actual observations, thus precluding extrema beyond the range of historical flow values. Nonetheless, despite these constraints on the variability of results, their instability is impressive. The single Q_{50} estimated from the entire long record is 3,283.6 cfs. Given that each random sub-set is equally likely as any other, we note that 10 years of record do not produce a stable estimate of Q_{50}.

Table 2 shows, on the assumption that the observations define the population of annual flood events, that $\hat{Q}_{50}$ as estimated from 10 years of actual record is not a stable statistic. Its standard deviation is 25,350, so if we assume that $\hat{Q}_{50}$ is normally distributed, about 95 percent of all estimates would lie between zero and 60,500. For 25 years of record the range is zero to 12,800. Given this high degree of instability, and given the extent to which model error makes it statistically inefficient to transmit information from gaged to ungaged sites, it is essential critically to assess the feasibility of the gaging program's objective.

Table 2. Statistics of Q_{50} Drawn from a Typical Site

Record Length	Min. Q_{50}	Max. Q_{50}	Mean	Standard Deviation	C_v
5	1,210	924,120	39,048	119,036	3.06
10	1,393	120,523	10,764	25,350	2.36
25	1,488	34,081	4,605	4,202	0.912

It would be premature to draw important conclusions from the sampling instability inherent in only one station. This contract does not encompass work items which would systematically generate tables similar to Table 2 for a large number of sites, and to do so for synthetic flows would essentially duplicate the work of Moss and Karlinger.

One alternative design criterion would be to base the design flow on a specified quantile of the distribution of Q_T, thus bypassing problems caused by those few outliers which make fitting by moments unattractive. Another criterion is tantamount to inverting the traditional design question (viz: what flow corresponds to the T-year return interval?) and to ask instead: What is the range of return intervals associated with a given flow? These questions and alternatives are treated below.

WORK PLAN

Theoretical Basis

Given the statistical instabilities associated with estimating Q_T, and given the inefficient transmission of information from gaged to ungaged locations, it is futile to maintain the objective of ten equivalent years of record from which Q_{50} should be estimated as the design flow. This analysis is predicated on explicit consideration of error introduced by statistical uncertainties and of economic consequences of these errors, which are then compared to the cost of collecting new information.

To evaluate the benefits associated with additional information, it is necessary to apply the same design criterion to two distribution functions; the first is derived from gaging information currently at hand and the second is based on information at hand plus that which could be added by continuation and extension of the gaging program. If the program is continued, the coefficients of the regression relationship between Q_{50} (the dependent variable) and the basin parameters (the independent variables) are more sharply defined (i.e., the regressions are "better"). Thus the model error associated with estimating Q_{50}, or any other Q_T, is reduced. However, the extent of this reduction is never so great that an additional year of gaging everywhere in the network will provide an additional equivalent year of record at ungaged sites because information is always lost in transferring unless the population correlation coefficient is known to be unity. Thus the length of equivalent record in the network increases slowly compared to the length of actual record at the gaged stations, whereupon the estimate of the design flow, because it is derived from the distribution of $\hat{Q}_{50}$, can not rapidly be reduced merely by increasing the length of gaging records at other network locations.

Figure 7 which shows a family of distributions of $\hat{Q}_{50}$ at a gaged location; the abscissa is the length of record of annual flood events at a particular site. The heavy line, not necessarily monotonic, represents the best estimate (the mean, median or some other statistic of central tendency) of $\hat{Q}_{50}$ which might be derived from a record of length N_Y years. There is no predetermined functional form for this locus, but in expectation it increases monotonically. The lines surrounding the locus represent boundaries within which a specified fraction of all estimates of Q_{50} will fall with a given probability. The figure is qualitatively suggestive, so no numerical values or theoretical significance should be attached to the representation. These boundaries are not necessarily symmetric with respect to the measure of central tendency or trend lines, but better estimates (i.e., more precise in the sense they have smaller sampling errors) generally are developed from longer records. Thus the loci which contain some given fraction of the potential estimates tend to funnel at the upper end of the function. Consider two sections passed vertically through Figure 7, which give distributions of $\hat{Q}_{50}$ for two alternative values of the record length; these are shown in Figure 8. The density f_0 has a smaller mean than does f_1, but this need not be the case. The second moment of f_1 is smaller than that of f_0 because its

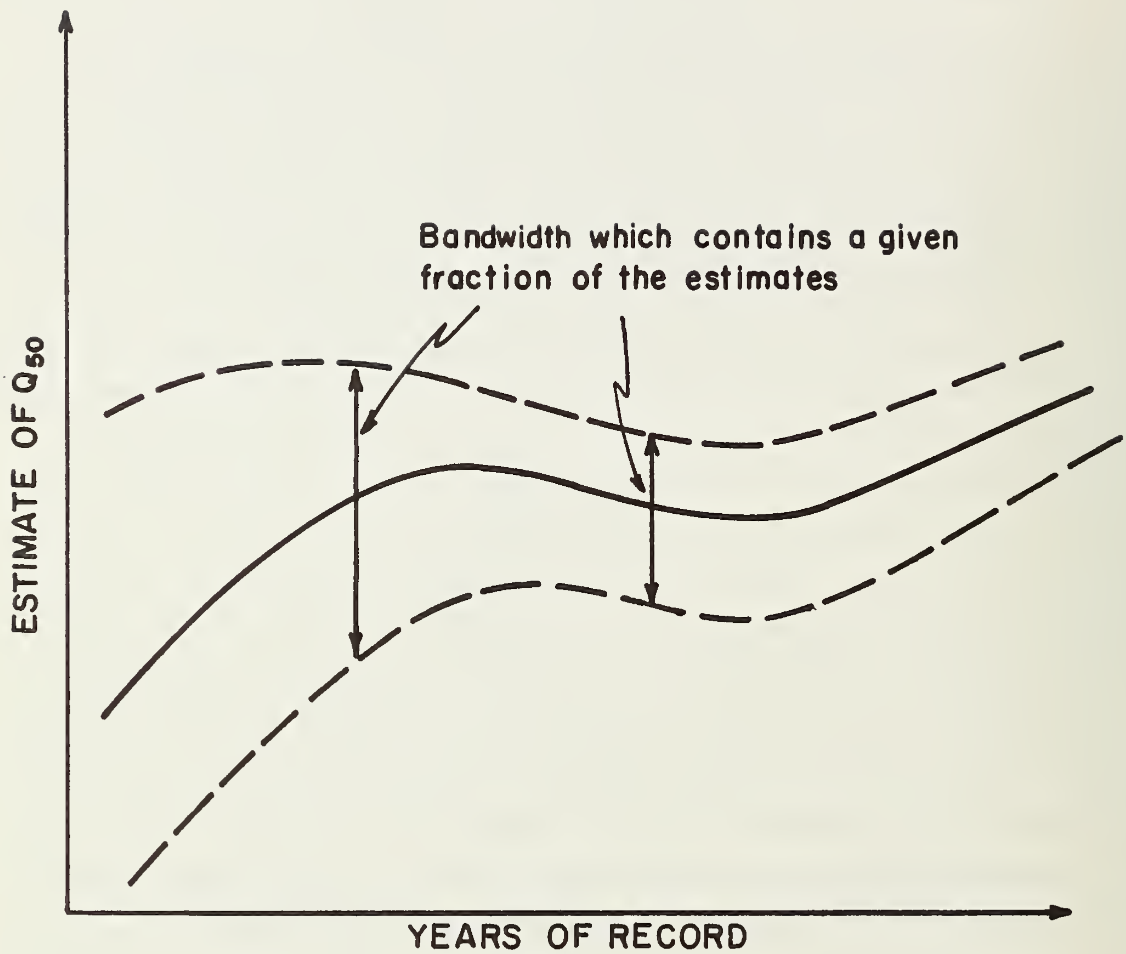


Figure 7. Distribution of $\hat{Q}_{50}$ as a Function of Record Length

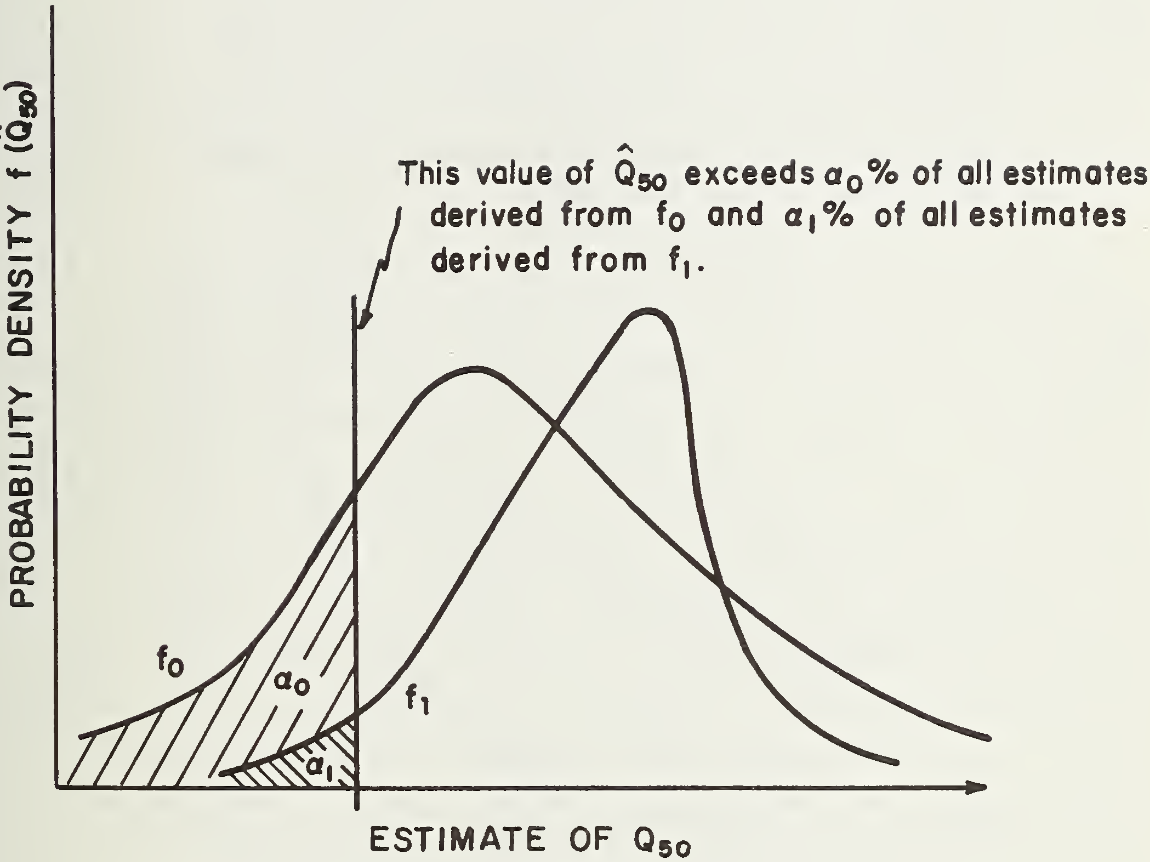


Figure 8. Reliability of Estimates of Q_{50}

record length is greater; the smaller second moment is important with respect to increasing the reliability of the design flow, and the second moment is reduced at a rate which makes it useful to have at least 10 years of (equivalent) record.

The design value, which need not be the expected value or median of the distributions in the figure, changes with the second moment. If the economic consequences of culvert failure are important, the design flow should be close to the right-hand tail of the densities f_0 or f_1. For undeveloped areas, where damages would be small, it might be appropriate to design closer to the left-tail, which is tantamount to a right-hand tail design for some other recurrence interval T. This is indicated in Figure 8.

The point is that flow can not be uniquely associated with a return interval but resides in the two-dimensional space of return interval and probability of exceedance. The contour map in Figure 9 represents this concept. It is seen from Figure 8 that every flow can be located on the density derived for a recurrence interval T. Upon locating this flow, a unique exceedance probability can be identified (analytically or numerically). Figure 9 does not represent actual contours, but shows the inverse relation between recurrence interval and exceedance probability. That is, the same flow might exceed (say) 90 percent of all estimates of Q_{25} but only 50 percent of all estimates of Q_{50}. Thus the hydrologic design problem, which is equivalent to selecting one of the contours q_i in Figure 9, requires specification of at least two parameters: the return period and the exceedance probability. These together uniquely define the design flow. For a given exceedance probability (which is mapped from a measure of risk aversion identified by the decision-making authority) design flow q_4 has return period T_4, flow q_3 has return period T_3, etc. Because $T_1 > T_2 > T_3 > T_4$, it follows that $q_1 > q_2 > q_3$ .... Similarly, for some specified recurrence period, the figure shows that a flow can be mapped into its associated exceedance probability.

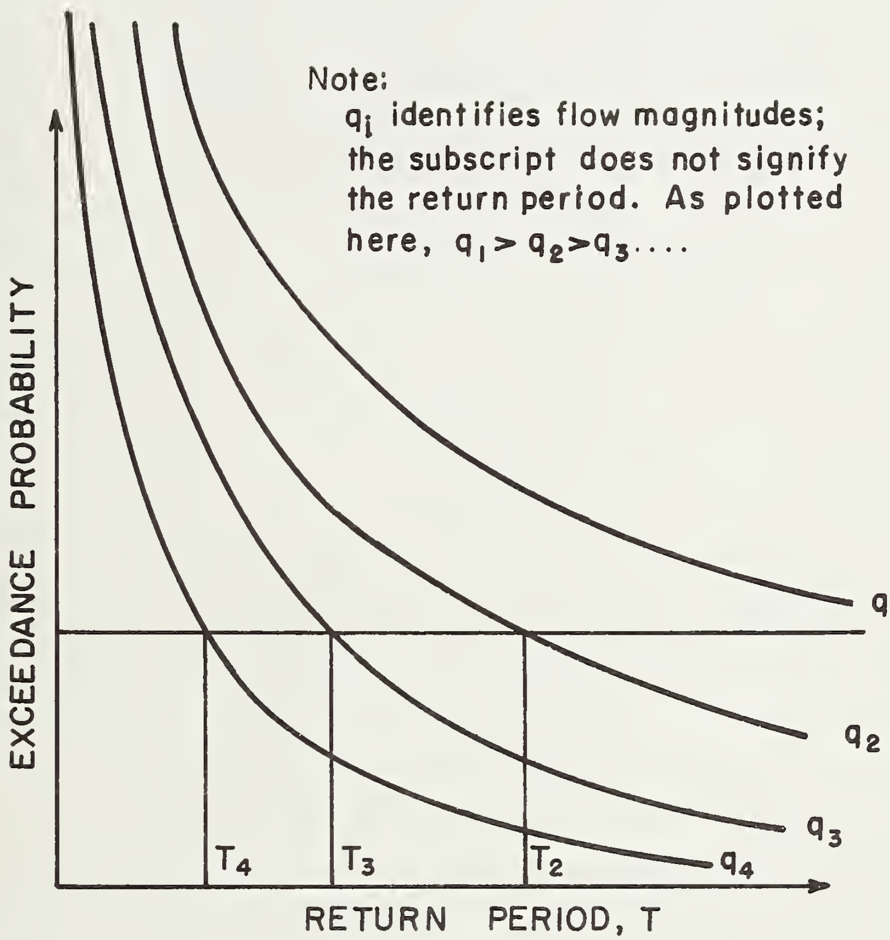


Figure 9. Contours of Equal Design Flow

Distribution of Design Flow

It is naive to speak of a design flow as if it were uniquely attached to some specific recurrence period. Instead, economic, social and institutional factors, are mapped into values along both axes of Figure 9 to specify a design flow. If we agree to deal only with the expected value of flow densities, and if these densities are symmetric so that their expectations are also their medians, then a unique mapping between design flow and return period can be deduced. This is equivalent to setting the exceedance probability at 0.5, but in the general case there is no economic justification for doing so. The T-year event is a random variable drawn from a distribution of potential events Q_T, and the design flow is to be selected at some percentile of this density. The bases for selection of the percentile are economic and institutional; the distribution becomes tighter as more information is available, and the effect of transferring information from gaged to ungaged sites is generally inefficient because of the dominance of model error.

To determine the distribution of Q_T and therefore to employ the economic analysis, the work plan is to develop the first two moments of the distribution and then to utilize BIGBASIN for each of the regions in question. This will enable maps such as Figure 8 to be drawn, from which the economic and institutional impacts of various levels of risk aversion can be deduced and incorporated into the final design. BIGBASIN tables have not in fact been developed for Q_{50}; the mean and standard deviation of flow are susceptible to BIGBASIN analysis. However, recall that there are (at least) two statistically interesting ways to calculate the design flow Q_d. First, assume a distribution for annual floods and make an unbiased estimate of Q_{50}. Note that the procedure recently developed by the USGS enables unbiased estimates to be made by changing (increasing, generally) the return period T and calculating Q_{T*}, $T^* > T$, as an estimate of the expected value of Q_T. The relationships between T* and T are given in the several WORLDWAR tables.

Second, the unbiased expectation (calculated as in the previous paragraph) can be incremented by an additive component whose sign and magnitude reflect the decision-maker's level of risk aversion. Knowing the unbiased estimates of the population moments of Q_T, it is then possible to assume a distribution for Q_T and apply BIGBASIN tables. Upon consultation with USGS staff members and close examination of the (limited) available theory, there was little reason *a priori* to discard the notion that Q_T is so distributed as to make BIGBASIN inapplicable. One of the most important published results to come from the USGS shows the relative insensitivity of optimal decisions to the *intentional mis-specification* of a distribution of floods; this study capitalizes on that work, and argues that the dangers of mis-specification are less important than exclusion of risk aversion. Moreover, if a decision-maker wishes to run no risk of mis-specification, the additive component is simply set to zero and the problem vanishes.

Arguments for entering the BIGBASIN tables are: N_B (the number of gages in the basin or region), N_Y (the length of record at each), ρ_{50} (the regional cross-correlation coefficient for events Q_{50}), η_{50} (the unbiased regional coefficient of variation, which is directly related to the skew coefficient of events Q_{50}), and the model error for the regression analysis applied in the region. It follows that the first task is to evaluate these five arguments for each of the regions* in question. The record length and extent of gage coverage (N_Y and N_B, respectively) are available trivially from the records. The regional coefficients and parameters require substantially more effort; their estimation is described in detail in a subsequent section.

---

* Again, note that a region, is a hydrologic, not a political, entity. In this study, States are designated as regions with a limited number of sites.

The use of BIGBASIN produces estimates of the true equivalent years of record for each region, and these, coupled with the use of WORLDWAR I (which gives unbiased estimates of the moments) determine the densities sketched in Figures 8 and 9.

Alternative Algorithms

Additional preliminary work is necessary to generate regressions (from which the regional model error can be estimated). These regressions give $\hat{Q}_{50}$ as the dependent variable, as a function of basin parameters, the independent variables. It is important to have a consistent method for estimating Q_{50} from the records at each site. At least three alternatives are available. These include: (i) the traditional method of fitting to the observations an empirical curve which gives the plotting position of each datum as $i/(N_Y + 1)$, where i is the rank of flow in question; (ii) the Water Resources Council (WRC) technique for estimating Q_{50} by fitting a log-Pearson function to the observations; (iii) the unbiased estimate of the expected value of Q_{50} using the latest USGS results in WORLDWAR I.

Each of these has advantages and disadvantages. The use of traditional plotting positions fails to account for statistical variations inherent in the observations so that extrapolation beyond the observed flows to the 50-year event is statistically indefensible and precarious. But merely because the technique is statistically indefensible one can not conclude that it is not useful; in fact, it has produced useful results in many cases and may be justified on the basis of empirical success alone. The second method, approved by the WRC, does not correct for bias in estimating parameters, whereupon some of the estimated values of Q_{50} might be significantly in error.

Finally, the USGS tabulations of the WORLDWAR I algorithm remove bias in the parameter estimates and is statistically most defensible. We call this modification the WRC* method. But as a result of this removal, it gives results which are sometimes difficult to accept. It is necessary to evaluate the true or population skew coefficient of the distribution of annual events in order properly to estimate the expected value of $\hat{Q}_{50}$, and for stations with short gaging records the skew is particularly vulnerable to enormous sampling fluctuations. The result is that one outlier among the data can introduce so much skewness that it might dominate the estimate $\hat{Q}_{50}$, leading to unusual results.

Under no circumstances should small-sample outliers be deleted from the record merely because they produce discomfiture; high sample skewness can not be overlooked, and although bounds are placed on its value, its estimation is central to the methodology of this study.

USGS Station 02229000, in Florida, is representative of this point; Figures 10 and 11 depict the difficulty. There are 12 annual values, of which 11 range from approximately 25 cfs to 2,000 cfs and of which the 12th is approximately 3,900 cfs. The mean of all 12 annual peaks is 1,250 cfs, with a standard deviation of 1,027 cfs; these statistics are shown on Figure 10. All flow values can reasonably be approximated by a normal distribution, which plots as a straight line on Figure 10. Extrapolation to the 98 percent exceedance level suggests the estimate of Q_{50} should be approximately 2,850 cfs, and that the recurrence interval associated with the 12th or outlier event (3,900 cfs) is of the order of 2,500 years. Figure 11 shows the same information plotted on log-probability paper, in which the solid portion above 2,000 cfs represents the actual observations while the dashed portion represents an extrapolation of the 11 smallest values. The estimate of Q_{50} on the basis of these 11 values alone is approximately 3,000 cfs, and the return interval associated with a flow of 3,900 cfs is approximately 2,000 years. These quantities agree closely with estimates from the arithmetic projections contained in Figure 10. But the consequences of a logarithmic representation are much more severe because if the solid flow-duration curve is projected beyond the largest observation it passes the 98 percent intercept at a flow near 300,000 cfs. In fact, using all 12 historical observations and the WRC* procedure, the estimate of the expected

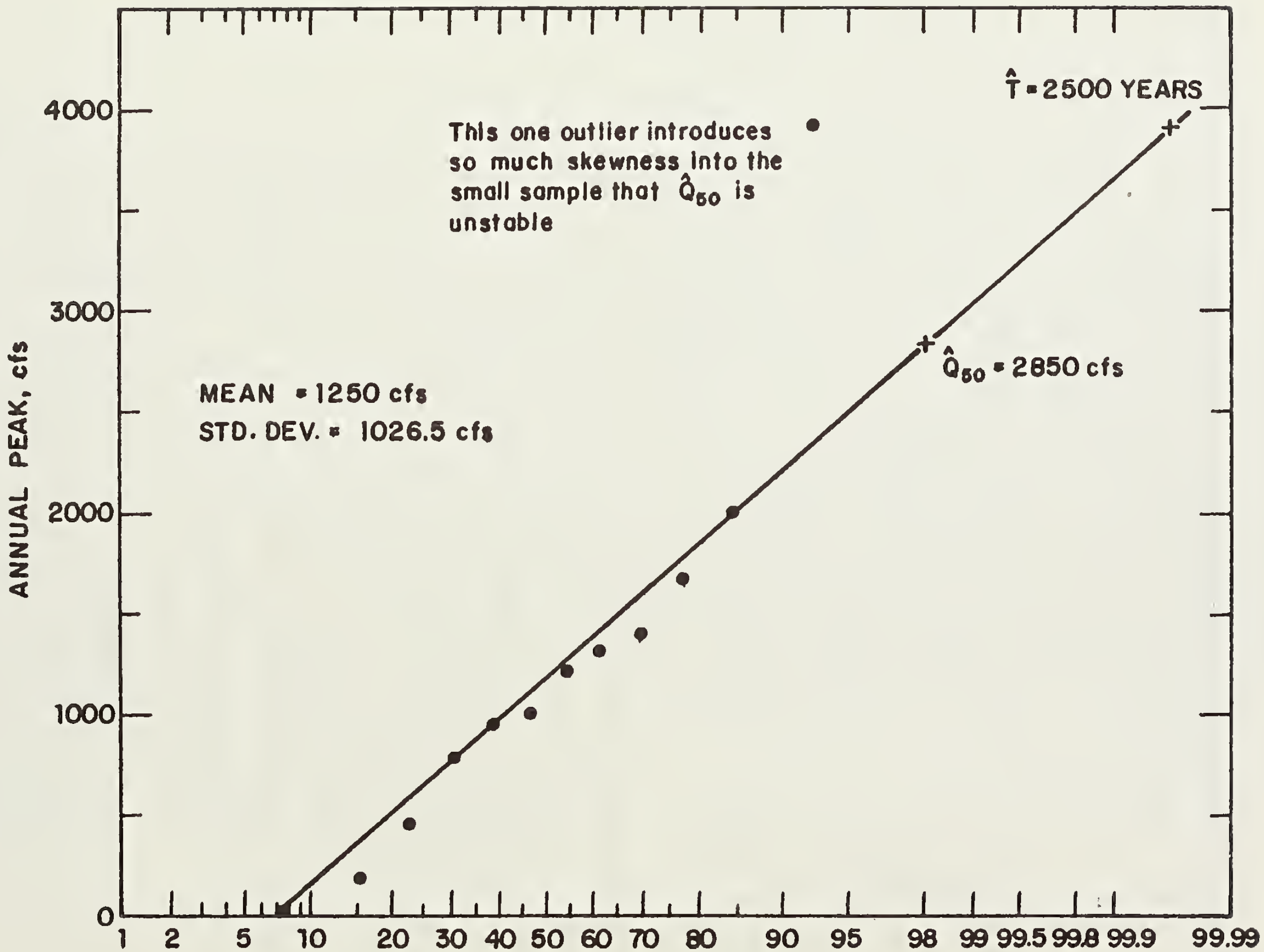


Figure 10. Flow-Duration Curve at 2229000

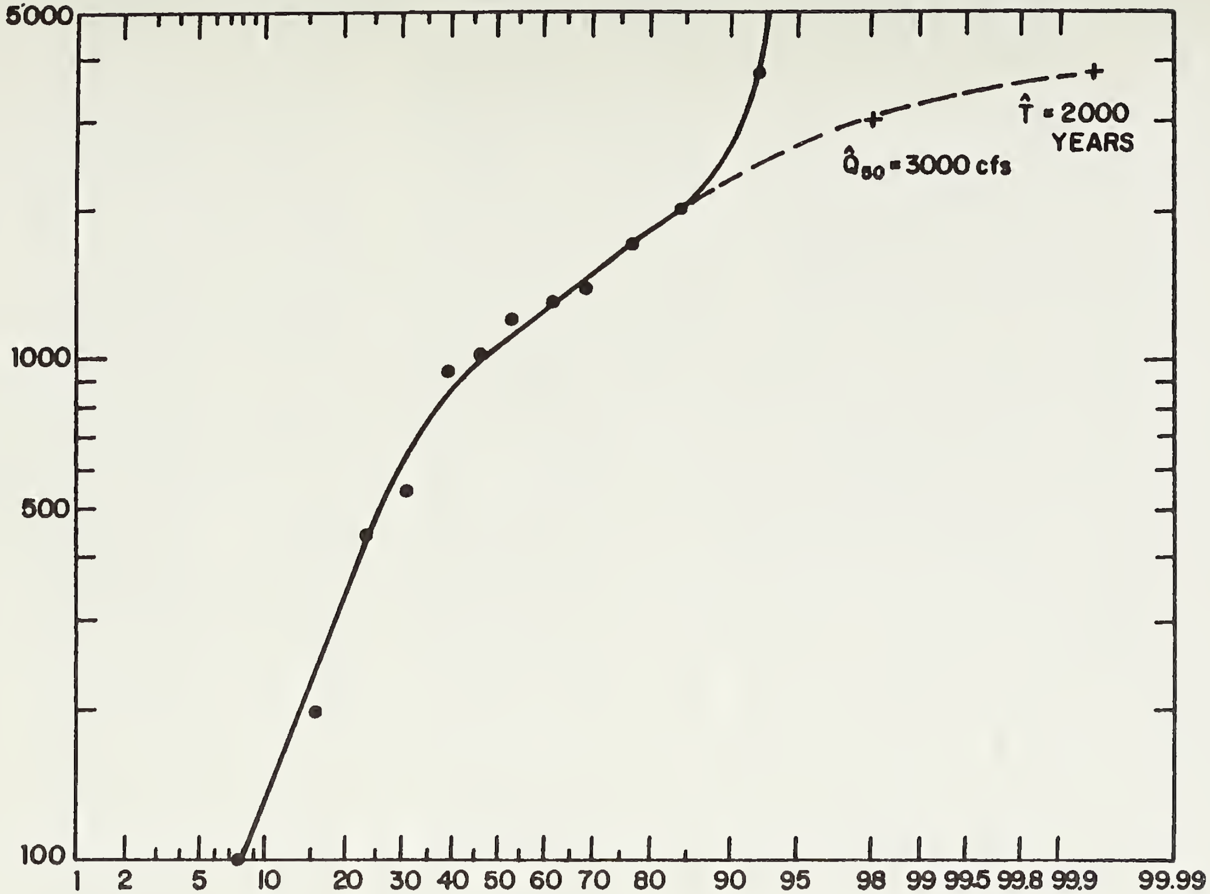


Figure 11. Flow-Duration Curve at 2229000

value of Q_{50} is 341,000 cfs, exceeding the largest observation by two orders of magnitude. This enormous flow can not be taken seriously as a design flow at or near this gage. The important question raised by alternative curve-fitting models is not the choice between one regression fit and another but that of inclusion, among the points which define the region, of data from sites with large sample skew coefficients.

Outliers and Skewness

Several options are available. First, the single offending event which introduces substantial skewness into the record could be deleted, leaving 11 rather than 12 data points from which to estimate Q_{50}. Second, that site could be deleted from the region (leaving 39 rather than 40 sites for Florida), thereby decreasing slightly the value of N_B. Third, a limit on the skewness of the annual events could be imposed so as to preclude such large estimates.

Under the first alternative, many of the observations would be discarded. The entire data array for all regions, for all sites, and all years was scanned, and it was determined that arbitrarily to discard extrema would be to strip the data of much of their richness and variability. Matalas has shown* that a substantial fraction of short streamflow traces drawn from skewed parent distributions display extreme values.

For the second alternative, the regional regressions for Florida were run using 39 stations, the notion being that discarding a station from a region would introduce less distortion into the results than discarding one or more data from a station. For the WRC estimates the multiple correlation coefficient decreased from 0.812 to 0.811, with the regression coefficients remaining essentially unchanged. Using the WRC* technique for estimating Q_{50} (the dependent variable), the

---

* Matalas, Nicholas C., private communication based on an unpublished study, USGS, 1976.

regression with 39 Florida stations had a multiple correlation coefficient of 0.705, a modest improvement over the original value of 0.647.

The third alternative, truncation of the skewness, is demonstrated by results in Figures 12 and 13. The State of Virginia has 145 sites, and the two estimating procedures produce plotting positions shown in Figure 12. The logarithms are plotted along the ordinate; log-probability paper is not used because the range of flows is so great that standard papers do not have enough cycles to accommodate the flows. The lower curve on Figure 12 represents the WRC technique, from which the mean of logarithms of events $\hat{Q}_{50}$ is 7.1565 with a standard deviation of 1.4912. The upper curve is the WRC* algorithm, from which the mean of logarithms is 8.6642 with a standard deviation of 2.1535. The bivariate correlation coefficient between logarithms of flows at the same site (as generated by the two algorithms) is 0.868, and the Spearman rank correlation ceofficient is 0.864 (this is a measure of how closely the two techniques reproduce the rank of events). Figure 13 contains the same information except that all skews are truncated at five.

Selection of five as the upper limit of population skew is not arbitrary. Kirby* showed that the upper limit of sample skew from n observations is (n-2)/(n-1). For samples of 10-15, typical values for hydrologic data on small watersheds, the sample skew is approximately 2.7-3.5. Matalas also showed that the expected population skew is about two to three times this sample value for samples of about 10-15, so that the maximal value of five is at the lower end of the range of products.

The moments of logarithms calculated from the WRC algorithm remain unchanged, indicating (for Virginia) that no sites generate skew coefficients in excess of five. For the WRC* algorithm, the

---

* Kirby, W., "Algebraic Boundedness of Sample Statistics," WRR, 10: 2, April 1974.

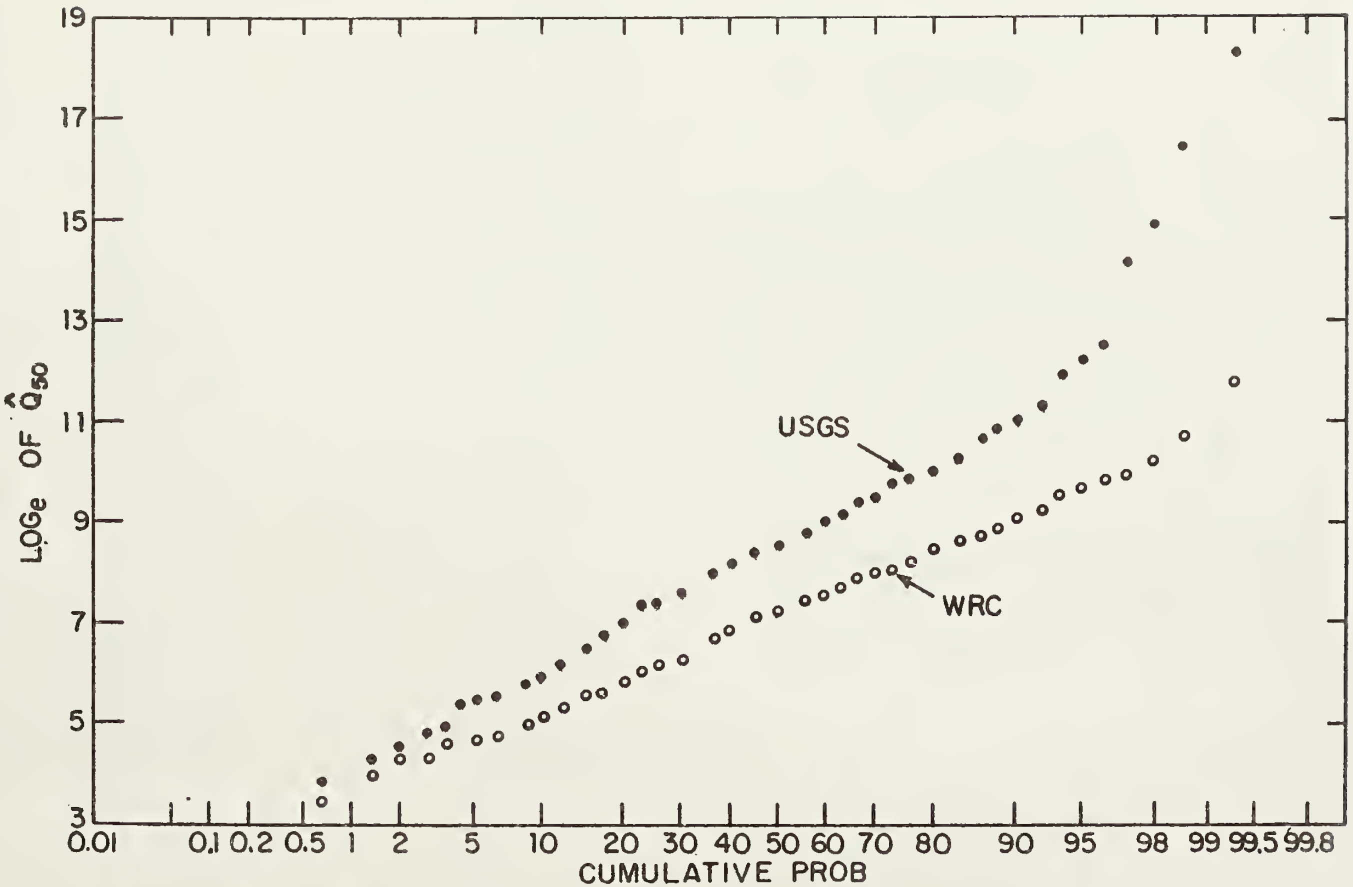


Figure 12. $\hat{Q}_{50}$-Duration Curve, Virginia Sites

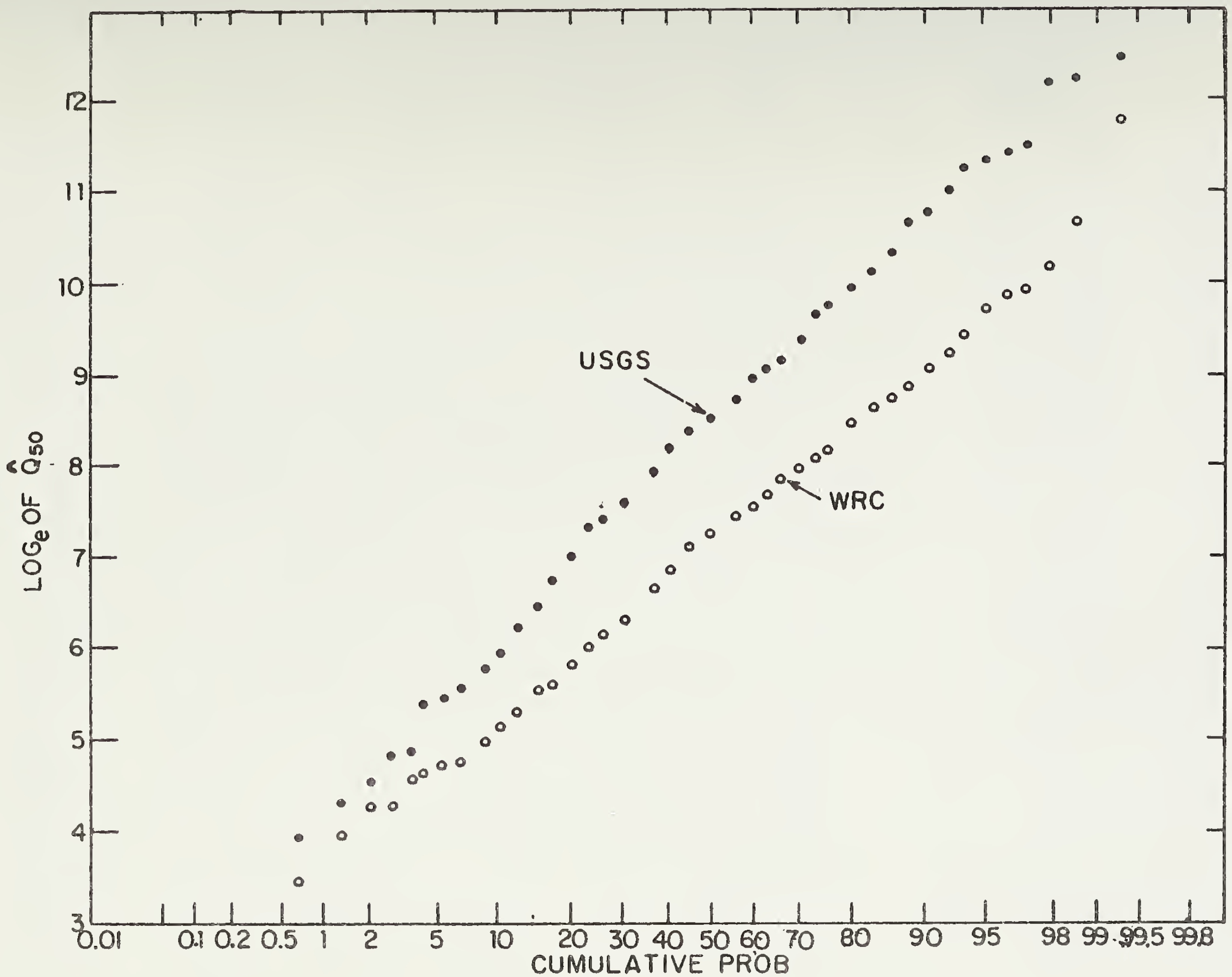


Figure 13. $\hat{Q}_{50}$-Duration Curve, Virginia Sites

Skew$\leq$5

mean logarithm is 8.4667 with a standard deviation of 1.7158. Thus, at least in log space, the coefficient of variation is substantially reduced by truncation of skewness. Both algorithms produce distribution functions which plot acceptably straight on Figure 13, and the correlation coefficient between magnitudes is 0.900. The Spearman rank correlation coefficient is 0.887. Thus both correlations are improved slightly by truncation.

The point of this discussion is to highlight instability inherent in estimating extrema by regression, and thereby to lay the groundwork for the poor success attributed to transfer of information from gaged to ungaged stations.

Arguments for BIGBASIN

Regression analysis is then performed in each region.* It is from these regressions that the model error in each region is estimated as the unexplained or residual variance in the regression of $\hat{Q}_{50}$ on basin characteristics. In this study all regression functions are exponential, so the unexplained variance is given in log units, as the standard error of the regression. Thus the standard error in absolute flow units is not independent of the arguments. In moving from small to large values of the arguments, the standard error tends to increase; if the regressions were not logarithmic, if the parameters of the regression were known with certainty, and if the normal distribution were the underlying parent, the standard error would be constant. It is therefore necessary to take an average standard error over the range of independent variables, which is accomplished by an approximation introduced by Slack, Wallis and Matalas.** The standard deviation in raw data space is given by the formulation

---

* Again, a State is designated a region, and conversely, in this study; in general, large States could have several regions or subregions $\leq$ 25 gaging stations.

** Matalas, Nicholas C., <u>et al.</u>, "Regional Skew In Search of a Parent," <u>WRR</u>, <u>11</u>: 6, December 1975.

$$\sigma = \left[\exp\,(2b + a^2)\ (\exp(a^2) - 1)\right]^{\frac{1}{2}} \tag{2}$$

where b is the mean in log space, and a is the standard deviation in log space of the dependent variable. The mean in log space is arbitrarily set to zero to give the conditional standard deviation rather than its absolute value (that is, the standard deviation about the regression line). The parameter a becomes the standard error of the regression in log units and σ becomes the standard deviation about the regression. This is analogous to a numerical estimate in raw data space of the average model error.

The computations for estimating ρ_{50}, the regional cross-correlation between estimates of Q_{50}, are tedious but conceptually simple. Each of the N_B gaging sites has available a record of annual flows Q_a, and these records are (approximately, with appropriate insertions and deletions made on an ad hoc basis) N_Y in length. We calculate the cross-correlation of annual floods between all pairs of gaging stations; there are $N_B(N_B - 1)/2$ different pairs derived from the array of N_B locations, whereupon an average value of the regional correlation between annual flows can be calculated.

The correlation between annual flows is not sufficient for purposes of entering the tables of BIGBASIN and WORLDWAR I because our method deals with the domain of 50-year events, for which one and only one estimate of Q_{50} can be made from the record of annual events at each site. Thus a correlation can not be calculated. Instead, we calculate the regional cross-correlation coefficient for annual events and then generate a long series of replicate synthetic traces at two stations to calculate the cross-correlation among 50-year events. Matalas* gives equations from which the parameters of the log-normal density can be calculated from the moments of the untransformed or raw data. It is then simply a matter of generating enough replicate

---

* Matalas, Nicholas C., "A Mathematical Assessment of Synthetic Hydrology," op. cit.

synthetic sequences of annual floods from which the extrema Q_{50} are estimated and correlated. In other words, assuming log-normal distributions of annual events, temporally independent annual flows, and a bivariate correlation coefficient equal to the regional cross-correlation coefficient, many replicate bivariate series of annual events are generated, each of length N_Y years, from which two 50-year events are estimated (WRC and WRC* algorithms). This gives pairs of values of $\hat{Q}_{50}$, with one element of the pair being an estimate of Q_{50} at one site and the other element being an estimate of Q_{50} at the correlated site. The correlation coefficients between series of 50-year events can readily be calculated, so that by Monte Carlo analysis a relation between the regional cross-correlation for annual events and the associated cross-correlation for extrema Q_{50} is deduced.

This numerical relation gives ρ_{50} as a function of ρ_a, N_Y and an estimate of the regional coefficient of variation (or skew coefficient). These parameters together define the log-normal densities from which the annual flood series are synthesized. The results can be presented in a set of contour maps showing, for a given record length N_Y, the relationship between ρ_{50} and ρ_a. A new map is required for each regional coefficient of variation (or skew coefficient). The coefficient of variation and the skew coefficient are used interchangably in specifying log-normal densities because there is a unique functional relationship between them (Aitchison and Brown*).

Estimation of the regional coefficient of variation of 50-year events is also conceptually simple. It requires first the estimation of the coefficient of variation of annual flows at each site, estimation of an unbiased skew coefficient, and ultimately the extraction from WORLDWAR I of unbiased estimates of the mean and variance of the 50-year events for all gaged sites. From these last two statistics

---

* Aitchison, J., and Brown, J. A. C., The Log-Normal Distribution, (Cambridge University Press, London), 1957.

the average unbiased coefficient of variation over the region is calculated by averaging the unbiased coefficients of variation at all sites. The statistic η_i, a biased estimate of the coefficient of variation of annual flows at the ith site, is calculated. The unbiased coefficient of skew for annual events is given by

$$G_i = \eta_i^3 + 3\eta_i \tag{3}$$

The tables of WORLDWAR I are entered with this unbiased skew coefficient to calculate unbiased estimates of the mean and standard deviation of 50-year events. This step generally requires linear interpolation because tabulated skew coefficients appear in large discrete steps. From the unbiased mean and standard deviation at each site, we calculate an unbiased estimate of the coefficient of variation for events $\hat{Q}_{50}$. These unbiased skew coefficients or unbiased coefficients of variation average to estimate an unbiased statistic for the region. Recently available unpublished tables (Slack, Wallis, Matalas, 1976) show that in expectation, for $N_Y = 10$ and regional skew of five, the coefficients of variation of log-normal events are essentially independent and that the average unbiased coefficient of variation for the region can closely be approximated by the average ratios of unbiased mean to unbiased standard deviation. Had the tables been available at the time the calculations were done, the exact values would have been used.

This completes the preliminary discussion of calculation of the arguments for the BIGBASIN tables, discussed below. There is a true return period τ associated with flows derived from a population characterized by a mean $\overline{Q}_{50}$ and a standard deviation s_{50}. The tables give standardized deviates for estimating τ. For example, if the annual events are log-normally distributed, if the regional skew coefficient of annual events is 0.5, and if N_Y is 10, then the true 50-year event has a standardized deviate of 2.313 derived from the 10-year entries in the tables. By interpolation in the log-normal row of the tables,

τ = 96 years at a deviate of 2.313. That is, the unbiased estimate of $\overline{Q}_{50}$ is that flow which, based on a sample of annual events, has a return interval of 96 years. When estimating the 50-year flow from a 10-year record of annual events, we seriously underestimate the true mean of all possible 50-year events which might be derived from the population which generated that 10-year sample sequence. The unbiased estimate of $\overline{Q}_{50}$ lies well beyond the biased single estimate suggested by the sample.

A similar procedure is used to extract unbiased estimates of the standard deviation, which in general is increased, so both the mean and standard deviation, when unbiased, exceed their single-sample counterparts. Thus the unbiased estimates of coefficients of variation of the 50-year flow, when measured at gaging locations throughout the basin, might not be greatly different from their biased values because both the numerator and denominator increase simultaneously. The ratio of unbiased moments is defined as the unbiased coefficient of variation at the ith gaging location, whereupon the average value over all locations gives the required average coefficient of variation. The unbiased coefficient of variation can be mapped directly into an unbiased skew coefficient for 50-year events using the above cubic equation for annual events, whereupon the average unbiased skew can be calculated for the region and used as the arguments for BIGBASIN.

Extension of Gaging

The economic criteria are now introduced. Decision variables are the extent of the gaging record, its length and areal coverage, as measured by N_Y and N_B.

Consider an incremental value to be added to the current value of N_Y; typically this will be five years or fewer. It would be ideal, of course, if the increment were one year and the gaging program re-evaluated on an annual basis. But the WORLDWAR I tables are developed for increments of no less than five years and often greater, so it was arbitrarily assumed to think in terms of extensions of five years.

The number of equivalent years is calculated twice. For lack of data, the regional regression analysis can not be redone because the extension for five years is conceptual rather than real so that no new data are in fact available. The parameters of the original record are used as if they were the parameters of the extended record. In expectation, the regression coefficients are not subject to change if they were estimated using unbiased and consistent techniques. Thus the design flow Q_d is reduced if it is chosen from the upper half of the distribution of $\hat{Q}_{50}$. The reduction in standard deviation is proportional to the square root of the ratio of equivalent record lengths.

For example, if the original number of equivalent years is six, and if five more years of gaging add two more equivalent years of record, the standard deviation is multiplied $(6/(6 + 2))^{\frac{1}{2}} = 0.866$, or reduced by 13.4 percent. The five additional years of gaging do not add five years to the length of equivalent record because the five years of data are diluted by model error when they are transferred to ungaged sites.

If the original number of equivalent years of record is small, and if the number of additional years developed by regression on the extended record is small, it suggests that extensive additions to the gaging program will not add significiantly to the information at the ungaged sites. Thus gaging should be terminated and effort devoted to improving the model so that future extensions of gaging can produce significantly more information at the ungaged sites. A most unlikely event, at least a priori, is that the original equivalent years of record is quite large whereupon additional gaging is presumably not indicated. This means that both the gaging program and any research into hydrologic modeling could profitably be discontinued.

The efficiency, or cost effectiveness, of continued gaging depends on the data at a site and on its transfer to other, ungaged, sites. Failure to reduce model error implies that gaging for transfer is inefficient, but it may still be useful to gage on the assumption that development might take place at, or very near, the gage.

Another area of consequence, namely the direct comparison between information gained and its cost, depends on the cost functions for the culverts involved and on the risk parameters α and T.

Section 3

ECONOMIC AND HYDROLOGIC STUDIES

ECONOMIC ANALYSIS

General

The major thrust of this study is the relationship between the cost of improved estimates and the economic benefits associated with them. This is in contrast to the usual criteria imposed on information systems: the collection of enough information to reduce to a given level the standard error of estimate of some parameter(s). Economic inputs do not appear explicitly in traditional analysis but are implicit in establishing the standard of precision below which the data base is inadequate. Tradition dominates the specification of system standards, whence the impact of economics is not explicit because performance criteria become habitual and therefore not the subject of explicit disciplined decision. To identify the way in which economic factors explicitly enter the decision-making process, this study presents its economic analysis as a coherent entity, showing details and assumptions, whereupon the decision-making framework might be better articulated and clarified. This section presents first an overview of our analysis and a discussion of why it takes the form reported herein. This is followed by tables which give the methodology and numerical results. The assumptions and approximations are cited throughout, as necessary.

It is generally true that more information improves parameter estimates. In the usual statistical sense, improved means that the standard deviation of the statistic under estimate, or its standard error of estimate, is reduced. The "best" estimate of the population mean derived from n observations, that value which has minimal standard error, is the sample mean. If the observations are independent it is well-known that the standard error of the mean is $\sigma/\sqrt{n}$, where σ is the population standard deviation. Unhappily, we rarely know σ so it must be estimated. But the sample mean remains the best, unbiased estimate of the population mean, so that the expected value from a sample of size n is μ, the population mean.

If the observations are not independent, it is clear that the value of an additional observation is not given by its full face value because we do not learn as much from a correlated as from an independent observation. For example, a positive serial correlation between consecutive flow values implies that high flow tends to cluster and low flows tend to cluster. Thus another datum replicates some of the information contained in the initial data; this redundancy is inherent in the persistence among the observations. We do not deal with this complication here because we have determined* that annual flood events can safely be taken to be independent; this assertion can not be made for mean annual flows, but is acceptable for extrema.

Consider records at nearby stations, say X and Y. They have a substantial period of overlap from which correlation can be estimated, but the record at X is longer than that at Y. It is desired to estimate the mean flow at Y, so correlation is utilized to estimate the missing values at Y from the longer record at X. Fiering** has shown that the use of regression estimates does not necessarily result in a better estimate of the mean at Y, but that criteria concerning record length and correlation must be met in order that augmentation be statistically useful. Consider the case in which records at X and Y are independent so that their correlation is zero. Under these circumstances, regression would add pure noise so that its effect would be to reduce the precision of the estimate of the population mean at Y even while its <u>apparent</u> effect is to increase the effective sample length at Y and thereby to improve the quality of the estimate of its mean. Similarly, if the correlation is perfect so that knowing x_i at X is equivalent to knowing y_i at Y, substitution between x_i and y_i

---

* Informal discussions with USGS personnel, 1975.

** Fiering, Myron B, "On the Use of Correlation to Augment Data," <u>op. cit.</u>

can be made with impunity so that the extended record at X can freely be used to augment the record at Y. It follows that somewhere between zero and unity there is a value of the correlation at which the standard error of estimate of the mean of the y_i is indifferent as to whether augmentation is undertaken or not. The trade-off occurs at that point where the standard error is unchanged by augmentation.

In the same paper, Fiering shows that if it is desired to estimate the variance of y_i there is a more rigorous restriction on the indifference level of the correlation because the sampling errors of the variance and other higher, moments increase quite rapidly. Thus the indifference level of the correlation must be significantly higher to control this potential source of increased error.

This early work is based on the sole statistical criterion that the standard error be unchanged. This section shows how economic criteria can be used to focus on a new definition of the indifference level, at which point the cost of including additional information through model-making and regression is compensated or balanced by the economic value of that information. Suppose it is desired to design a culvert* where no flow measurements are available. Current technique utilizes regional regression analysis, in which the dependent variable is the design flow Q_d at the ungaged location and the independent variables are basin characteristics. Data for these equations come from analysis of a large number of stations in the region, or vicinity of the ungaged location. The question is whether a longer record at the gaged site will produce a sufficiently more precise regression estimate of Q_d at the ungaged site to justify the cost of the data collection program. Arguing as in the previous paragraphs, it is not so important to determine whether the expected value of the design flow is subject to change as a function of increased record availability but to concentrate on its standard error, which might be reduced by increasing record length or the number of gaged sites.

---

* This analysis, for reasons given elsewhere, treats culvert only and ignores bridges.

We seek to determine the economic benefit associated with reducing the sampling error or standard deviation of the design flow, and to determine how much these reductions cost in terms of additional measurements (in time or space). These are brought together by comparing the value of additional information against the cost of its collection, and some conclusion reached concerning the adequacy of existing networks and the extent to which they should be continued. The economic considerations of how these savings are distributed among the State and Federal governments are not considered nor is the issue of whether these savings are derived from Interstate, primary or secondary road systems. It is recognized that various mixtures of road systems will result in different levels of cost sharing, and that each State actually pays a different portion of its total drainage need. Thus the States and Federal government perceive different levels of cost and benefit. This study considers total highway construction needs, without distinctions introduced by various Federal incentive and re-payment programs.

Construction Cost Savings

The basis for the argument that culvert construction cost savings are associated with improved flow estimates is:

1. hydrometric data networks increase the level of information with respect to the estimation of flood peaks and design flows;
2. additional data reduce the standard error of flood estimators;
3. a lower standard error, acquired through investment in the hydrometric network and in models for transferring information from gaged to ungaged sites, results in estimates of the design flow which decrease with decreasing standard error (assuming the same failure probability or return period is maintained); and
4. tightly distributed flood peaks, when utilized with nearly constant criteria of risk aversion, produce smaller design

flows which result in less costly drainage structures -- the savings being thereby a direct benefit of the hydrometric network and the models superimposed thereon.

The economic information derived from a methodologic study* and from a detailed case study of 20 culvert sites** is taken as typical for culvert costs. These are shown in Figure 14, for which the following items are relevant:

1. Risk is the expected value of losses associated with site damage, flood damage to the highway, and to appurtenant structures; flood damage is damage to the flooded area adjacent to the highway; and traffic delays are costs for extra mileage and time necessitated by routing traffic around damaged highways.

2. Risks are based on dynamic flood routing of families of inflow hydrographs deemed appropriate to each of the 20 case study sites.

3. The probability associated with each member of the hydrograph family is our best estimate, and is used to weight the economic losses which define the risk.

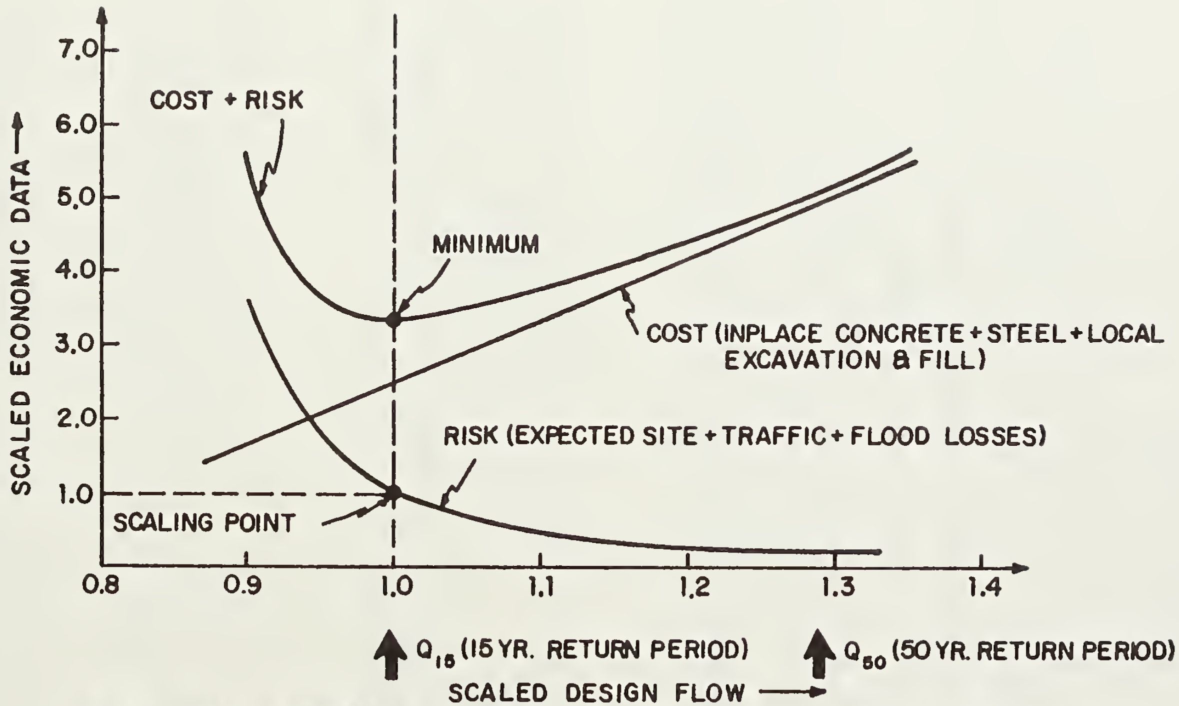
4. All 20 case study sites represent rural culverts on interstate roads located in Virginia, and are chosen to represent a range of physiographic conditions (mountains, piedmont and coastal plain) which is extrapolated to the nation. That is, the smooth curve is assumed applicable everywhere even though its parameters vary from State to State.

5. The optimization analyses conducted in the cited reports and displayed in Figure 14 indicate that for Virginia a return period of approximately 15 years is associated with the minimal cost-risk

---

* Young, G. K., et al., "Evaluation of the Flood Risk Factor in the Design of Box Culverts," Volume 1, Report, FHWA-RD-74-11, Federal Highway Administration, ORD, Washington, September 1970.

** Young, George K., et al., "Optional Design for Highway Drainage Culverts," J. ASCE, Hydraulic Div. HYT, July 1974.



NOTE: Based on analysis of 20 Virginia box culverts from hilly, rolling and flat land.

Figure 14. Typical Scaled Economic Data

combination (i.e., least-cost). Unfortunately, $\hat{Q}_{15}$ can not be estimated very precisely from most hydrologic samples. This conclusion is unique to Virginia and is not used elsewhere.

6. Risks associated with $\hat{Q}_{50}$ are very small. This suggests that $\hat{Q}_{50}$ might not be an appropriate statistic to serve as the design flow; indeed, there is evidence to suggest that the flow traditionally thought to be $\hat{Q}_{50}$ is generally associated with a much shorter return interval. The implication is that designers, over the course of decades and over a range of hydrologic and physiographic conditions, obtain acceptable results (where acceptability is measured in terms of economic losses) under the assumption that the design return period is 50 years when, in fact, based on unbiased estimates, it is much shorter.

7. The costs shown in Figure 14 are culvert barrel construction costs, which is that element of total cost that varies with culvert size. Other costs pertain to fill, pavement, entrance and discharge works, and local grading; these are fixed by considerations of highway and culvert alignment rather than hydraulics.

8. At the optimal or least-cost design there are considerable risks. Current design practice significantly increases the cost over the least-cost solution by accommodating a flood level with a return period of perhaps 50 years, and thereby reduces risk to very small levels.

9. There has heretofore been no rational assessment for specification of $\hat{Q}_{50}$ as the acceptable design criterion. It has nonetheless generated federal support by default; this support is apparent in design criteria specified for Interstate highways.

At design flows associated with 50-year return intervals there is virtually no statistical risk, so the cost consists almost entirely of construction costs rather than any cost attributed to culvert failure. This leads to the conclusion that economic benefits of improved estimates of the design flow are directly related to reduction in capital costs rather than to a trade-off between capital costs and increased risk. It is recognized that improved estimates of the design flow can have economic effects related to changes in the risk. However, the typical economic responses detailed in Figure 14 indicate that accounting for risk will produce results that remain dominated by the capital costs of culvert construction. This view of the economic data leads to the following working hypothesis: If current culvert design practice is maintained so that long return periods continue to dominate design, the economic factors remain dominated by capital costs. Risks in the vicinity of the design flow are relatively small and do not significantly contribute to the total cost of the culvert.

The capital costs are approximately 20 times greater than the risk in the neighborhood of design flows associated with 50-year return intervals. This is validated by case studies in the cited references. It is derived fron nation-wide data for rural areas; discrepancies among major cities could not be accommodated by this general statement, but this work pertains to small, rural watersheds and hence applies to these more nearly uniform areas.

Therefore, the remainder of this section is devoted to presentation of primary economic and culvert cost data to derive projected construction costs of culverts on a State-by-State basis. The objective is to estimate five-year expected culvert construction costs and to derive the marginal changes in these costs associated with small unit reductions in estimates of the design flow.* The methodology for generating these marginal costs data is directed at finding consistent and balanced estimates for each State. Had all emphasis been placed on one or two States, much sharper estimates would have been available; however, this study is concerned with nation-wide estimates. If future refinement should prove to be justifiable when this work is applied, the assumptions and methodology can be scaled to serve the appropriate model and its resolution.

---

* This is analogous to the use of a structural influence line or a unit hydrograph.

Primary Data Sources

The following groupings of primary data are used to estimate benefits associated with reduction in design flow:

1. Construction costs per mile of highway, for each of the three types of system (Interstate, primary and secondary), for each region of the United States.*

2. Highway plans for a range of road systems in nine States. These plans are analysed to obtain culvert densities (number of culverts per mile) and the culvert size distributions by State and by physiographic region within the States.**

3. Generalized relationship between design flow and culvert area. This is developed on the assumptions of full flow and velocity head recovery consistent with current improved inlet designs.

4. Generalized box culvert costs based on national average unit costs for locally available backfill, steel and concrete.***

5. A large, representative sample of typical pipe culvert costs.+

6. The five-year highway needs for reconstruction, isolated reconstruction and new locations, for each type of system (Interstate, primary and secondary), for each State.++ The published needs have

---

* U.S. Department of Transportation, FHWA, 1973 Highway Statistics, 1975.

** Again, note that bridges are excluded from the analysis because they do not matter at the margin.

*** Young, G. K., et al., "Evaluation of the Flood Risk Factor in the Design of Box Culverts," op. cit.

\+ ARMCO, Handbook of Drainage and Construction Products, (Middletown, Ohio), 1958.

++ U.S. Department of Transportation, "The 1974 National Highway Needs Report," Report of the Secretary of Transportation, House Document 94-95, 1975.

been reduced by 20 percent, as recommended by the FHWA to account for projected demand reductions in response to higher fuel costs. FHWA planners adjusted highway needs following the gasoline shortage in 1974.

The economic data contained in the primary sources are adjusted to 1974 prices using the price index curve in Figure 15.* The composite relation of that figure is used. The following sections discuss in detail the primary data sources listed above, including details of extrapolation to obtain national figures from State or regional data.

Total Highway Costs per Mile -- Figure 16 shows the ten regions of the United States for which there are published data on highway costs per mile.** The data are for 1964 and have been scaled to 1974 prices using the composite curve in Figure 15. Data for Interstate, primary and secondary systems are shown in Table 3. The highest costs are for Interstate roads in the Middle Atlantic region (approximately $2.4 million per mile), reflecting high land and labor costs. The lowest costs are for secondary roads in the mountain region (approximately $0.2 million per mile).

Culvert Data Compiled from Plans -- Data were collected on culvert density and cross-sectional area for pipe and box culverts in nine States: Alabama, California, Georgia, Idaho, Maine, Missouri, Oregon, South Dakota, and Texas. Several sets of highway drawings were obtained for each State. These drawings were examined to determine spacing (or culvert count per mile) and culvert size data, all of which were tallied and summarized. The data are then grouped according to Soil Conservation Service (SCS) land resource regions. The plans available to us describe highway designs in twelve of these SCS regions. The SCS classification scheme is given in Table 4; Figure 17 shows these land resource regions on a map of the contintental United States.

Table 5 shows the culvert density and average cross-sectional area by type of culvert (pipe or box) for the nine States and twelve SCS regions for which highway plans are available. To extrapolate

---

* U.S. Department of Transportation, FHWA, op. cit.

** Ibid.

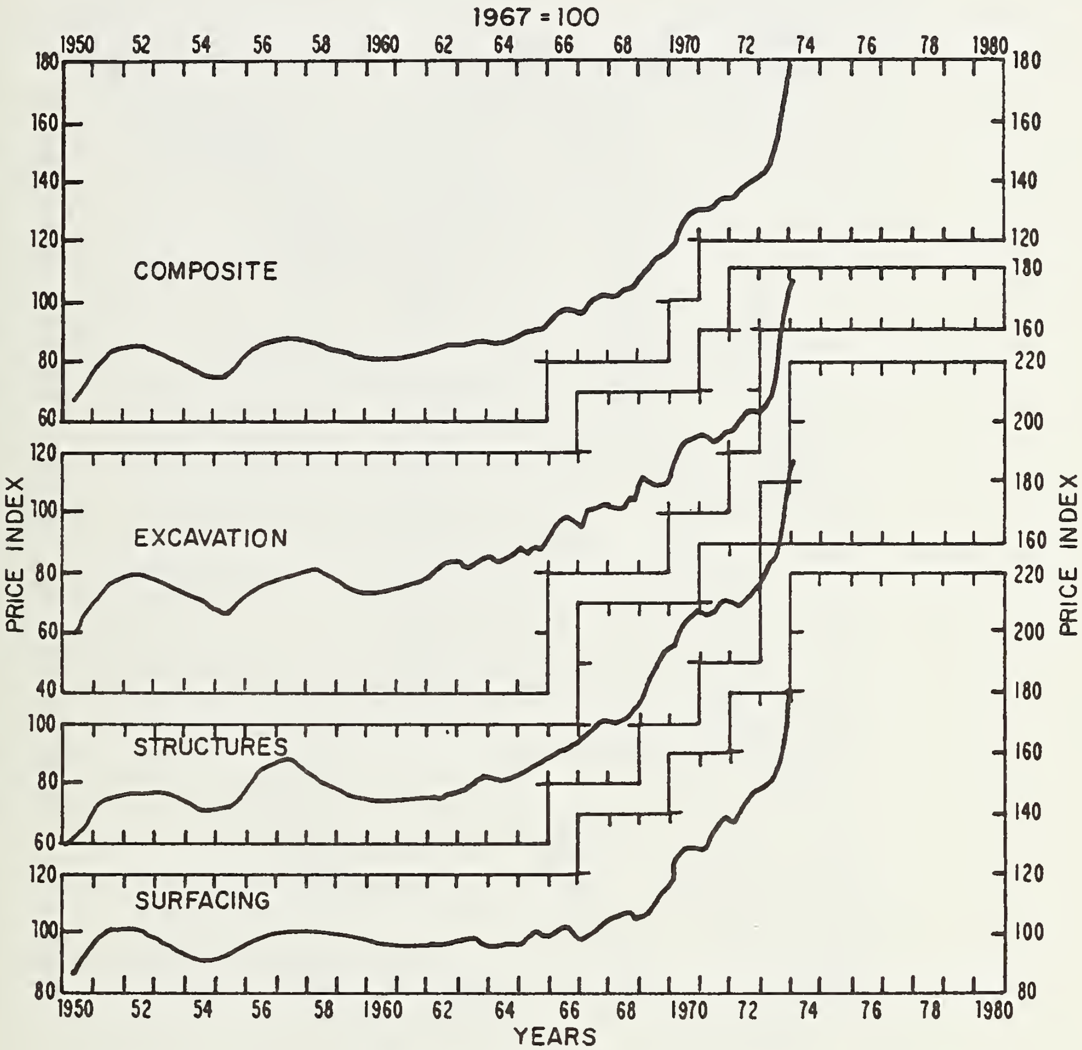


Figure 15. Price Trends for Federal-Aid Highway Construction After U.S. Department of Transportation [39]

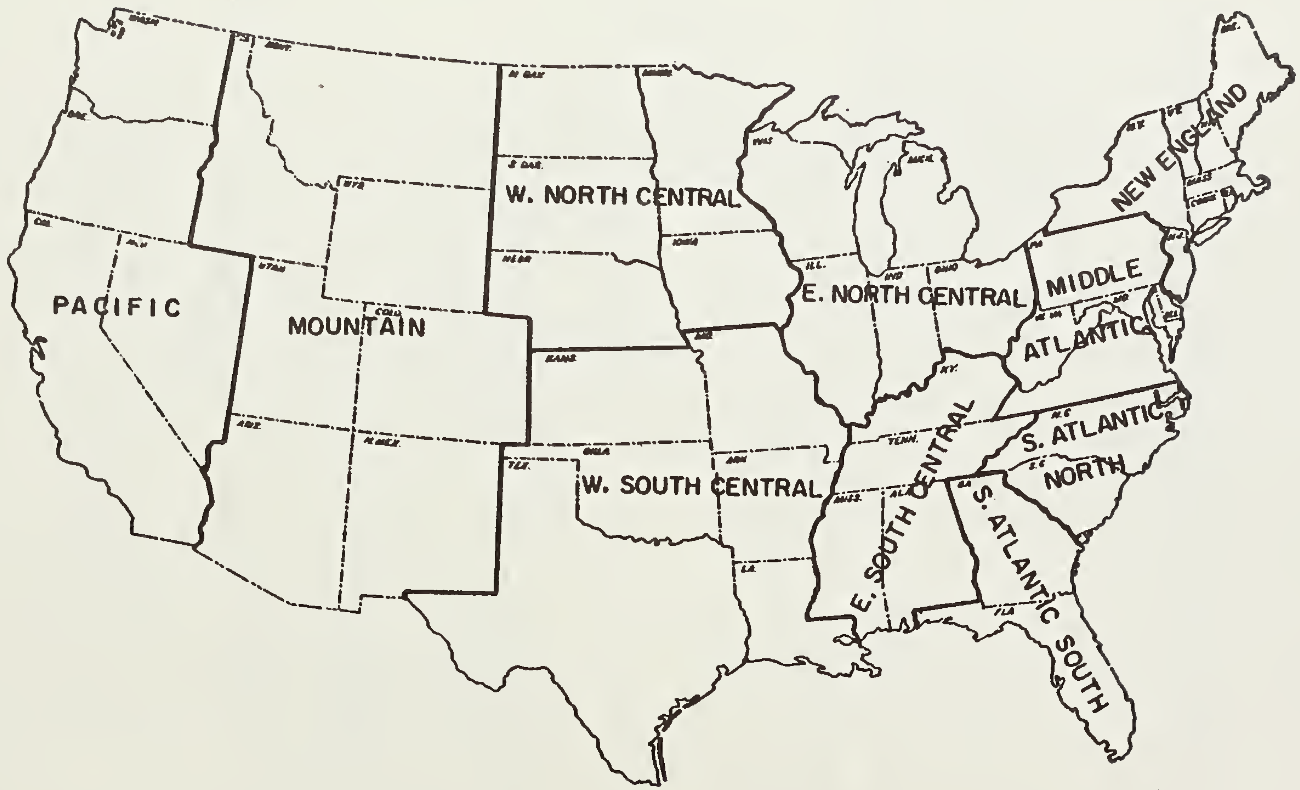


Figure 16. Ten Regions for Highway Cost Designations

Table 3. Highway Cost per Mile in 10 U.S. Regions

	Interstate		Primary		Secondary	
	1964	1974	1964	1974	1964	1974
1. New England	909,101	1,880,899	464,577	961,194	169,708	351,120
2. Middle Atlantic	1,170,299	2,421,308	539,210	1,115,607	192,555	398,390
3. So. Atlantic North	1,117,790	2,312,669	468,539	969,391	161,597	334,339
4. So. Atlantic South	507,607	1,050,221	287,751	595,347	101,938	210,906
5. East North Central	788,466	1,631,308	381,817	789,966	136,013	281,406
6. West North Central	515,535	1,066,624	172,094	356,057	114,971	237,871
7. East South Central	639,477	1,323,056	224,712	464,921	144,061	298,057
8. West South Central	659,328	1,364,127	194,340	402,082	123,266	255,033
9. Mountain	488,088	1,099,837	153,838	318,286	104,101	215,381
10. Pacific	718,471	1,486,492	313,452	648,521	125,705	260,079

Sources: Winfrey [44] and US DOT [39]

Scale factor = $\frac{180}{87} = 2.07$

1967 = 100

Table 4. SCS Classification Scheme

A. Northwestern Forest, Forage, and Specialty Crop Region
B. Northwestern Wheat and Range Region
C. California Subtropical Fruit, Truck and Specialty Crop Region
D. Western Range and Irrigated Region
E. Rocky Mountain Range and Forest Region
F. Northern Great Plains Spring Wheat Region
G. Western Great Plains and Irrigated Region
H. Central Great Plains Winter Wheat and Range Region
I. Southwestern Plateaus and Plains Range and Cotton Region
J. Southwestern Prairies Cotton and Forage Region
K. Northern Lake States Forest and Forage Region
L. Lake States Fruit, Truck, and Dairy Region
M. Central Feed Grains and Livestock Region
N. East and Central General Farming and Forest Region
O. Mississippi Delta Cotton and Feed Grains Region
P. South Atlantic and Gulf Slope Cash Crop, Forest, and Livestock Region
R. Northeastern Forage and Forest Region
S. Northern Atlantic Slope Truck, Fruit and Poultry Region
T. Atlantic and Gulf Coast Lowland Forest and Truck Crop Region
U. Florida Subtropical Fruit, Truck Crop and Range Region

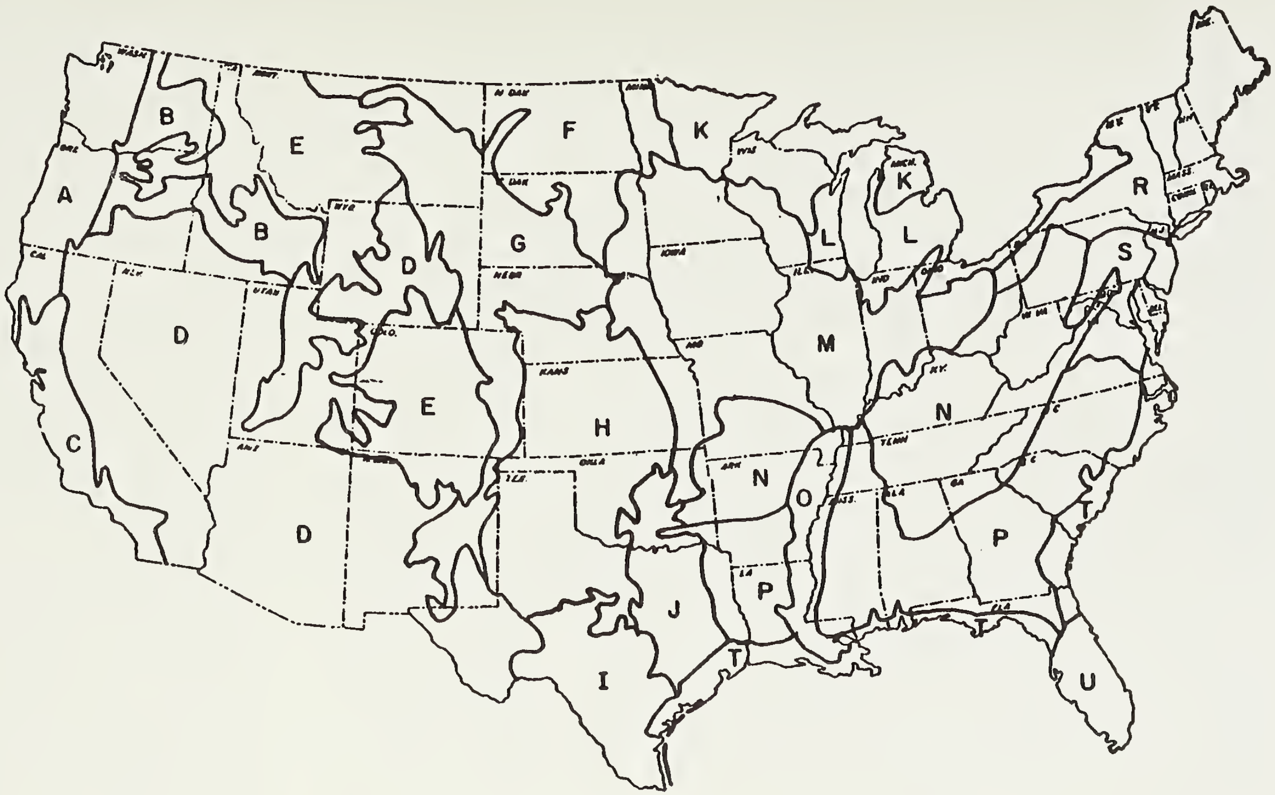


Figure 17. SCS Land Resource Regions

Table 5. Culvert Density and Area

	Culverts Per Mile	Percent Pipes	Avg. Area/Culvert		Equivalent SCS Regions
			Pipe	Box	
State					
1. Alabama	12.3	85.8	3.5	3.2	
2. California	5.5	95.2	6.2	84.3	
3. Georgia	9.0	92.1	3.1	54.0	
4. Idaho	4.8	93.3	4.7	84.3	
5. Maine	8.3	100.0	3.7	--	
6. Missouri	5.4	84.8	4.8	43.4	
7. Oregon	12.1	100.0	2.4	--	
8. So. Dakota	5.0	95.0	5.3	54.4	
9 Texas	1.1	62.8	4.8	32.5	
SCS Region					
1. A	11.7	100.0	2.3	--	
2. B	4.7	92.1	4.1	93.3	F
3. C	5.7	91.9	5.8	78.8	
4. D	3.5	95.2	7.0	52.2	I
5. E	6.0	96.6	6.0	30.0	
6. G	4.6	95.7	5.5	49.1	H
7. J	1.4	45.5	3.7	35.0	
8. M	5.1	86.6	4.4	62.0	L
9. N	9.2	90.3	3.4	38.6	
10. P	7.2	82.5	3.4	41.6	S
11. R	8.3	100.0	3.7	--	K
12. T	9.3	92.5	3.2	50.4	O, U, J

from the nine States with data to all forty-eight States in the continental United States, SCS regions within which culvert data are available are assumed hydrologically equivalent to those SCS regions for which no highway data are available. Table 5 tabulates the equivalencies. Extrapolation of culvert data is made on the assumption that equivalent regions have the same occurrence and area statistics as those regions for which data are available. This enables us to estimate culvert density and culvert area for each State, using weighting coefficients based on drainage areas, and yields for each state a set of defensible statistics to determine culvert costs.

There is no claim that the hydrologic and economic analyses are pursued with equal precision. Advanced statistical tools are applied to counteract bias in estimating Q_t, while coarse economic assessments are applied over large areas with little apparent regard for precision. But in fact only the paucity of economic data underlies this disparity; the methodological advance in hydrologic estimation is important enough to be displayed in detail in the hope that future applications will be based on better economic valuations. The concept of reducing design flow (and construction cost) as a consequence of improved information is an important step, and the apparent imbalance in technique should ultimately disappear.

It is also necessary to determine representative culvert lengths and fill heights. Disaggretation along these parameters at the State level is not available, so national estimates are used; these are based on the drawings from the nine representative States, and the data are presented in Table 6.

More precise culvert data can be utilized as it becomes available in the future; the new data is merely substituted and the methodology developed in this study used to evaluate the State program.

<u>Generalized Hydraulic Functions</u> -- State-by-State statistics on average culvert area are used to obtain representative values of State unit costs and State design flows. The assumption that the drainage system within each State can be represented by a single design flow

Table 6. Culvert Dimensions

A. Length (ft)

	Pipes	Boxes
Interstate	155	204
Primary	158	138
Secondary	103	89

B. Fill Height and Area

	Pipes	Boxes
Fill Height (ft)	8	23
*Cross-Sectional area (ft^2)	3.2	48

* This item also disaggregated by state (see Table III-3)

characteristic of that State is critical. Given the paucity of available highway information, and given that decisions on continuation of gaging will be made on a State-wide basis, it is reasonable to seek some single statistic which describes the drainage requirements (in physical structures as opposed to dollar costs) for each State. In subsequent analyses, based on regional data sets, this can be relaxed. Each State has a range of hydrologic and physiographic characteristics which govern its culvert design densities and areas. But by blending the different standards asssociated with the State's mixture of highway types (Interstate, primary and secondary) and by using the extrapolation algorithm described under Culvert Data Compiled from Plans above, there is little doubt that our statistics, while not ideal, are representative of the situation in any State.

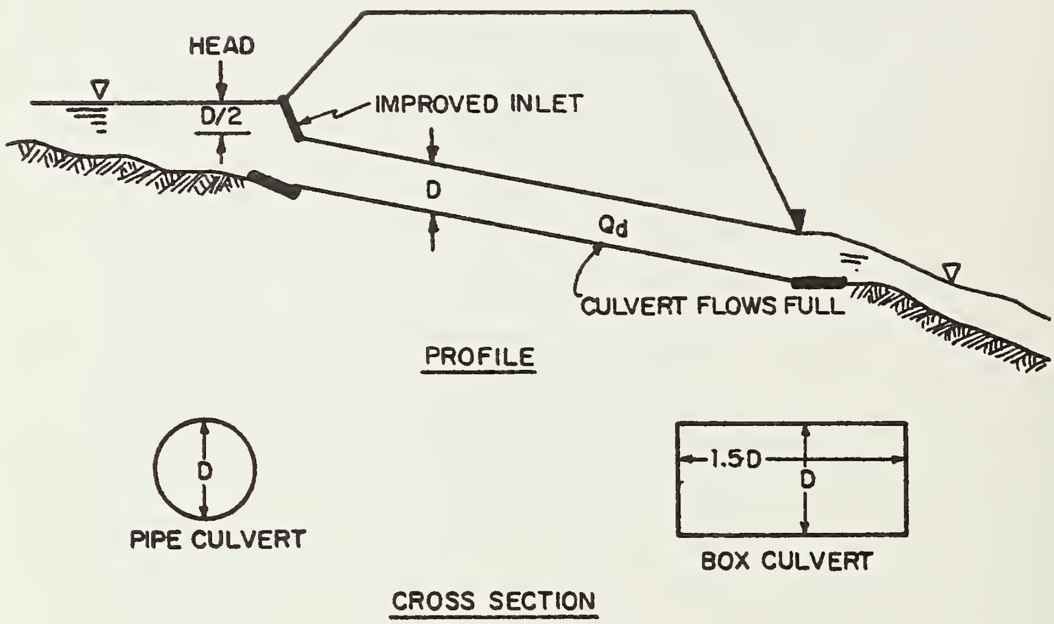
Design flows are therefore estimated using State-by-State cross-sectional areas. The hydraulic assumptions are shown in Figure 18. It is assumed that improved inlet design will be employed and that the conduits will flow full with velocity head equivalent to half the flow depth. A further assumption for box culverts is that the width is 1.5 times the depth. Design flow as a function of cross-sectional area is shown by the curve in Figure 19.

Generalized Cost Functions -- Culvert cost is related to culvert area. It is expressed in dollars per linear foot of culvert for pipe and box structures. Figure 20 shows the average unit bid price for metal and concrete pipe culverts, tabulated for several hundred jobs in 1953.* The prices are for the far western portion of the United States, which experiences construction costs in the middle of the cost range in Table 3. The 1953 prices for concrete pipe are scaled to 1974 using the price index data in Figure 15.

Prices for box culverts are not based on tabulated data but are determined from national average unit prices for 1974. Figure 21 gives

---

* ARMCO, op. cit.



PROBLEM: Given the area ($A = \pi D^2/4$ or $A = 1.5\,D^2$) in ft^2, estimate the design flow, Q_d, in cfs.

Assuming a modern improved inlet design with an associated velocity head of D/2, the design flow estimates are:

$$\boxed{\text{PIPE: } Q_d = 6.03A^{5/4}}$$

$$\boxed{\text{BOX: } Q_d = 5.12\,A^{5/4}}$$

Figure 18. Estimating Design Flow from Culvert Area

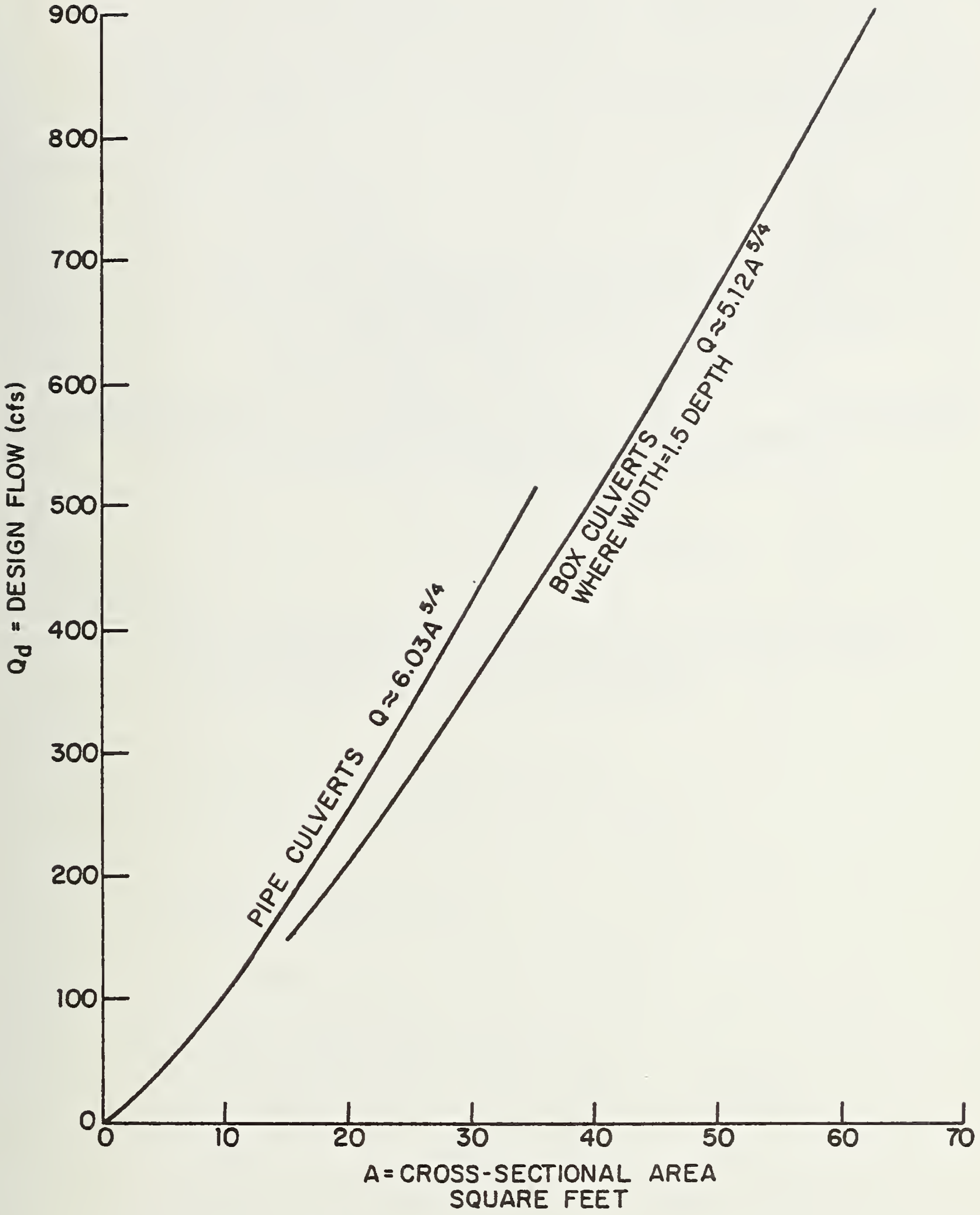


Figure 19. Design Flow vs Area

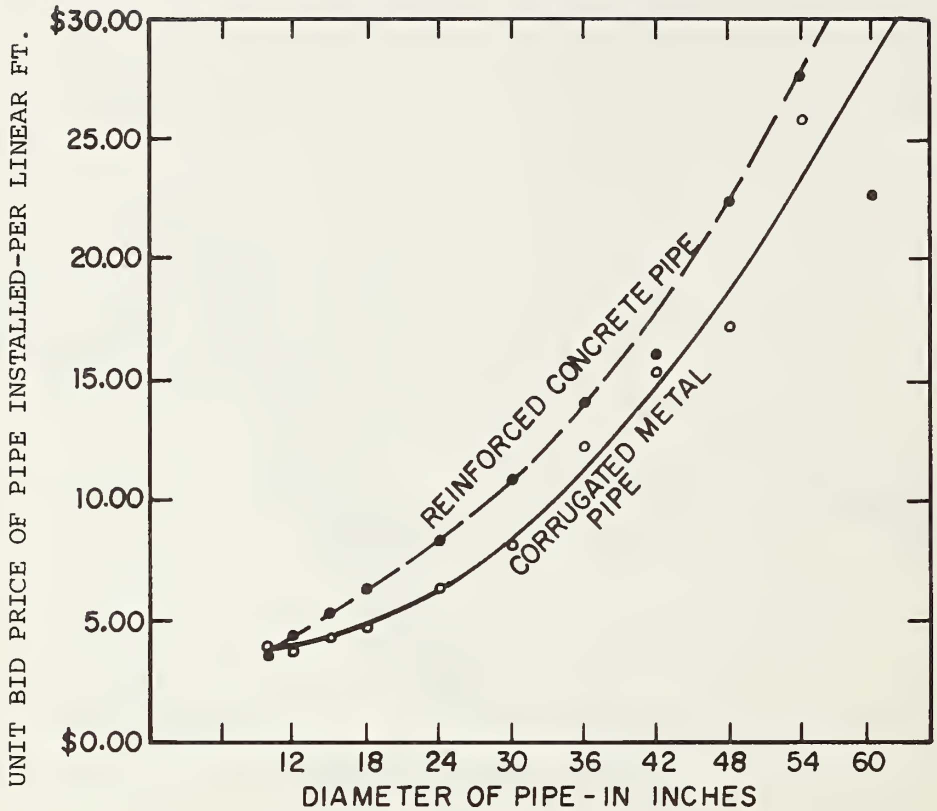
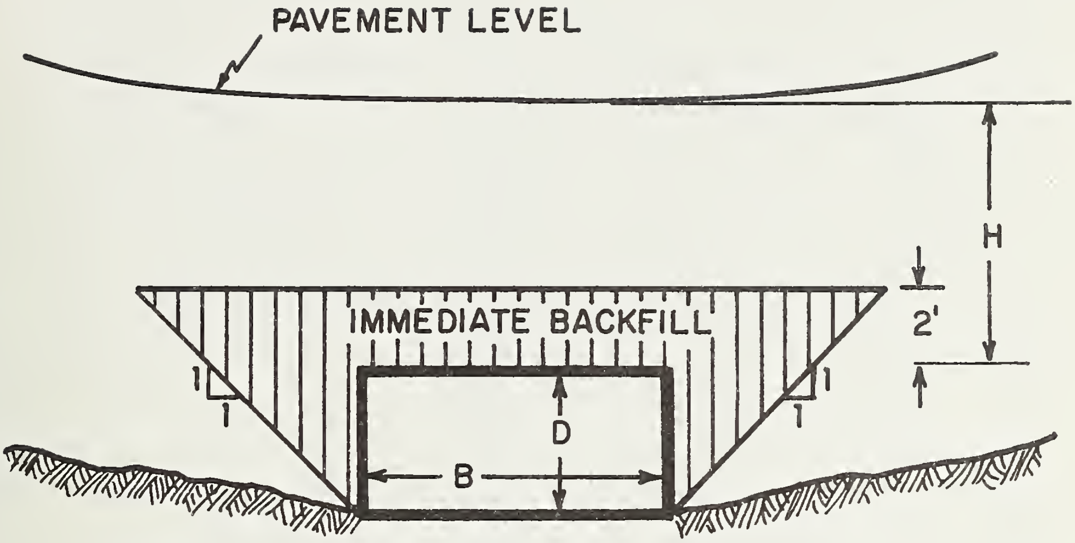


Figure 20. Pipe Culvert Costs after ARMCO [2]



NOTES:

1. Concrete and steel estimated from Virginia Highway Standards.

2. B/D = 1.5 for standard sizes.

3. Twin barrels used for larger sizes.

4. Backfill costs $3/cy.

5. Concrete in place costs $125/cy.

6. Steel in place costs 45¢/lb

Figure 21. Assumptions for Box Culvert Costs

the assumptions associated with calculation of box culvert costs. An average fill height of 23 feet is used, as indicated in Table 6. Standard designs* are consulted to obtain quantities of concrete and steel associated with a typical culvert installation. In addition, a width-to-depth ratio of 1.5 is used. Figure 22 gives generalized cost curves for box culverts as a function of cross-sectional area. In addition, the function for concrete pipe culvert as applicable to the smaller cross-sectional areas is shown in the figure. These unit costs in Figure 22 are the link between the actual highway plans examined for nine States and extrapolations of these density and area data to a State-by-State cost estimate for the nation. The culvert costs must be further identified with representative design flows to assess the benefit of improved estimation of the design flows. These improvements are obtained at the cost of improving and maintaining the hydrometric network.

Sampling Culvert Costs -- Figures 20, 21, and 22 present extrapolation from the published costs for pipe culverts and estimated costs of box culverts to each State, and serve therefore as the basis of a State-by-State evaluation of potential savings which might be effected through improvement of the estimates of the design flow Q_d.

Highway Needs and Potential Benefits -- Unit culvert costs and design flows are estimated as functions of culvert cross-sectional area. The functions represented in Figures 19 and 22, when used simultaneously, provide a mapping among the three quantities: culvert area, unit cost of culvert construction, and design flow. If any one of these is given, the functions can be used uniquely to estimate the other two.

Table 7 contains the economic data which forms the basis of the crucial trade-offs. In the first three columns the highway needs for

---

* Young, G. K., et al., "Evaluation of the Flood Risk Factor in the Design of Box Culverts," op. cit.

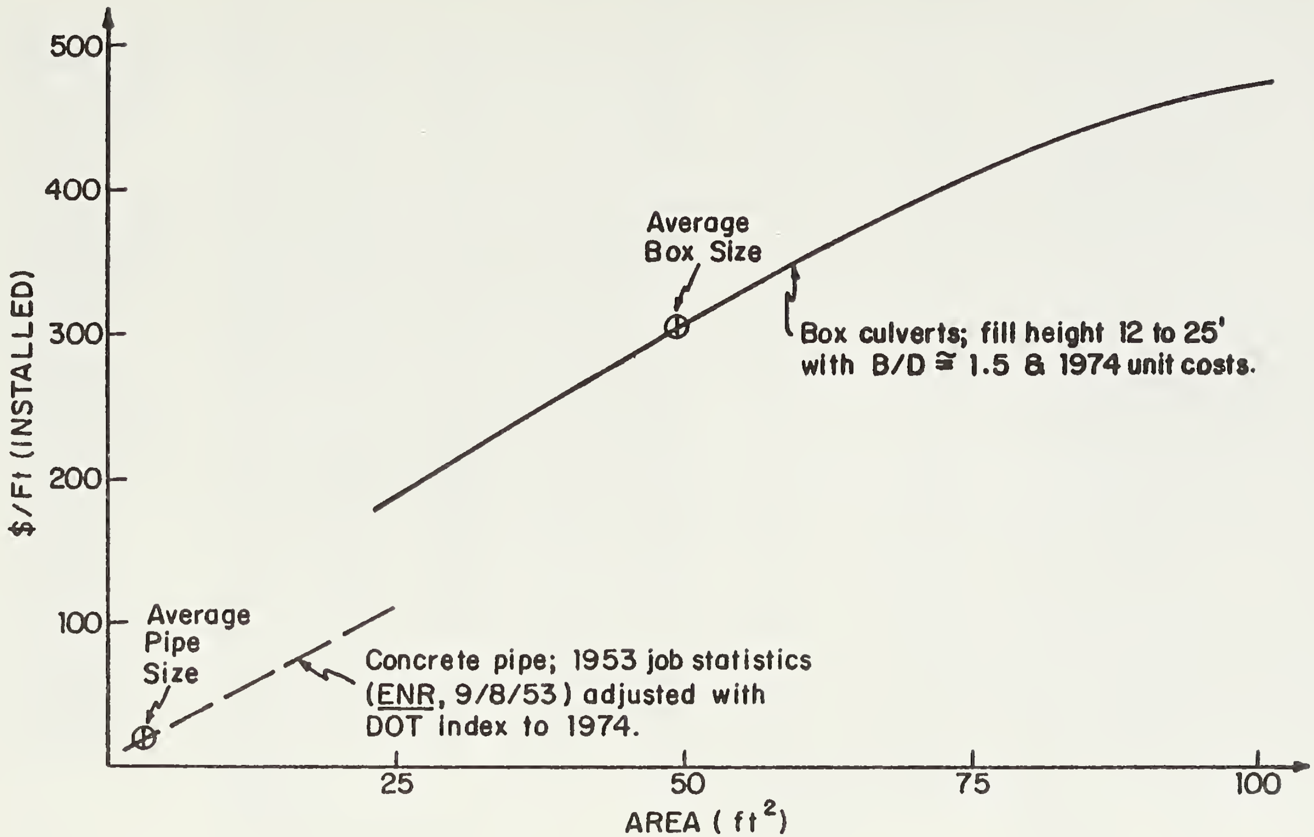


Figure 22. Generalized Culvert Cost Functions

Table 7. Primary Economic Data

	Hwy $/Mile(millions)			Culvert Data				Costs, $/lin.ft.		5-Yr Needs (millions)		
	Inter-State	Primary	Secondary	#/mi	Fract. Pipes	Avg Area Pipes, ft^2	Avg Area Boxes, ft^2	Pipe	Box	Inter-State	Primary	Secondary
Alabama	1.323	0.465	0.298	7.9	.849	3.4	41.1	21.80	273	0	593.8	239.9
Arizona	1.100	0.318	0.215	4.2	.96	7.1	52.8	37.50	330	24.4	254.9	90.5
Arkansas	1.364	0.402	0.255	8.6	.912	3.3	43.0	21.00	285	1.9	469.8	487.0
California	1.486	0.648	0.260	4.4	.956	6.2	62.3	33.50	365	59.2	1928.6	478.4
Colorado	1.100	0.318	0.215	5.6	.96	6.0	35.0	33.00	240	12.1	145.0	291.9
Connecticut	1.881	0.961	0.351	8.3	.96	3.7	48.3	23.50	310	3.9	146.9	132.2
Delaware	2.421	1.116	0.398	9.3	.92	3.2	50.4	20.50	310	0	101.2	16.3
Florida	1.050	0.595	0.211	9.0	.905	3.2	49.1	20.50	310	0	229.5	230.8
Georgia	1.050	0.595	0.211	7.8	.847	3.4	42.5	21.80	285	18.7	447.3	284.0
Idaho	1.100	0.318	0.215	5.5	.895	5.2	65.1	29.00	375	8.8	233.0	219.1
Illinois	1.631	0.790	0.281	5.1	.87	4.4	62.0	26.00	365	36.9	981.0	403.7
Indiana	1.631	0.790	0.281	5.7	.875	4.3	58.5	25.50	350	11.8	467.9	690.1
Iowa	1.067	0.356	0.238	5.1	.87	4.4	62.0	26.00	365	38.9	736.4	286.2
Kansas	1.364	0.402	0.255	4.7	.942	5.2	51.7	29.50	325	32.5	301.2	359.4
Kentucky	1.323	0.465	0.298	9.0	.892	3.4	38.9	21.80	260	24.6	1203.7	1238.2
Louisiana	1.364	0.402	0.255	8.4	.915	3.3	46.4	21.00	300	31.5	214.6	286.5
Maine	1.881	0.961	0.351	8.3	.96	3.7	48.3	23.50	310	19.1	265.8	543.3
Maryland	2.421	1.116	0.398	7.4	.918	3.4	41.3	21.80	275	31.0	426.6	143.7
Massachusetts	1.881	0.961	0.351	8.3	.96	3.7	48.3	23.50	310	4.2	325.2	185.6
Michigan	1.631	0.790	0.281	6.7	.935	4.1	50.2	25.00	320	55.4	867.5	572.8
Minnesota	1.067	0.356	0.238	5.1	.932	4.1	62.8	25.00	365	1.2	266.1	313.7
Mississippi	1.323	0.465	0.298	7.7	.845	3.4	43.8	21.80	290	8.3	658.5	720.0
Missouri	1.364	0.402	0.255	6.7	.883	4.0	53.2	25.00	330	20.7	697.6	663.1
Montana	1.100	0.318	0.215	5.3	.96	5.7	44.2	32.00	290	20.6	265.3	334.6
Nebraska	1.067	0.356	0.238	4.7	.936	5.2	52.3	29.00	330	0	336.2	163.2
Nevada	1.486	0.649	0.260	4.2	.96	7.1	52.8	37.50	330	3.9	64.0	53.1
N. Hampshire	1.881	0.961	0.351	8.3	.96	3.7	48.3	23.50	310	7.7	202.2	213.4
New Jersey	1.881	0.961	0.351	9.1	.936	3.3	50.4	21.00	320	27.3	502.6	228.2
New Mexico	1.100	0.318	0.215	4.6	.96	6.5	48.5	35.00	310	0	347.5	296.5
New York	1.881	0.961	0.351	7.7	.96	3.8	48.3	23.50	265	48.9	906.2	928.7

Table 7. (continued)

	Hwy $/Mile (millions)			Culvert Data				Costs, $/lin ft.		5-Yr. Needs (millions)		
	Inter-State	Primary	Secondary	#/mi	Fract. Pipes	Avg Area Pipes, ft^2	Avg Area Boxes, ft^2	Pipe	Box	Inter-State	Primary	Secondary
No. Carolina	2.313	0.969	0.334	7.9	.852	3.4	42.9	21.80	285	22.1	962.8	861.3
No. Dakota	1.067	0.356	0.238	4.7	.922	4.2	91.1	24.50	445	13.4	315.1	233.8
Ohio	1.631	0.790	0.281	6.7	.892	4.0	48.8	24.50	315	48.2	676.2	1126.4
Oklahoma	1.364	0.402	0.255	6.7	.931	4.4	49.4	26.00	315	9.7	369.6	270.2
Oregon	1.486	0.659	0.260	7.5	.96	4.3	58.8	25.50	350	74.4	1729.7	850.2
Pennsylvania	2.421	1.116	0.398	8.2	.928	3.5	41.7	22.00	285	301.2	1453.3	155.6
Rhode Island	1.881	0.961	0.351	8.3	.96	3.7	48.3	23.50	310	0.59	25.3	9.1
So. Carolina	2.313	0.969	0.334	8.0	.859	3.3	44.5	21.00	290	10.9	249.0	128.2
So. Dakota	1.067	0.356	0.238	4.7	.93	4.9	65.0	28.00	375	0	156.1	101.4
Tennessee	1.323	0.465	0.298	8.9	.89	3.4	40.2	21.80	270	1.9	504.5	457.4
Texas	1.364	0.402	0.255	5.5	.955	5.7	50.6	32.00	320	105.0	996.0	921.1
Utah	1.100	0.318	0.215	4.9	.96	6.7	43.7	36.00	290	0	133.1	180.0
Vermont	1.881	0.961	0.351	8.3	.96	3.7	48.3	23.50	310	0.40	179.9	233.7
Virginia	2.421	1.116	0.398	7.9	.833	3.4	41.1	21.80	270	22.9	436.3	883.4
Washington	1.486	0.649	0.260	8.3	.96	3.4	64.5	21.80	370	56.9	400.2	367.2
W. Virginia	2.421	1.116	0.398	9.2	.9	3.4	38.6	21.80	265	0	585.1	1016.3
Wisconsin	1.631	0.790	0.281	6.2	.916	4.2	57.2	25.00	345	26.3	736.3	333.1
Wyoming	1.100	0.318	0.215	4.8	.96	6.4	46.2	34.50	300	25.5	90.9	158.1

each State which appear in the 1974 Report of the Secretary of Transportation* are tabulated. The focus is on rural needs for new locations, reconstruction and isolated reconstruction for Interstate, primary and secondary systems. The Report indicates that 22.9 percent of the rural need is for new locations, 49.3 percent for reconstruction, and 4.7 percent for isolated reconstruction. On the basis of extrapolations described in the previous sections, the next four columns give the number of culverts per mile, the fraction of culverts which are pipe as opposed to box structure, the average area of pipe culverts and the average area of box culverts. The next two columns are costs in dollars per linear foot of pipe and box structures, corresponding to the cost functions associated with the design flows assigned to each State. They do not reflect culvert density or the fractions of culverts which are pipes or boxes, but are costs per linear foot of installed structure. Finally, the last three columns reflect the five-year needs in millions of dollars for Interstate, primary and secondary systems. These are total construction costs (or needs) and do not reflect classification into drainage and other costs. The Needs Report gives 18-year estimates, but five-year needs are judged to be convenient for analysis of hydrometric networks. Therefore we multiply the published 18-year needs by 5/18.

Table 8 contains the cost and benefit information derived from the primary economic data in Table 7. The first three columns represent the five-year drainage needs for all three highway systems, and are fractions of the total five-year needs in Table 7. These drainage needs are a different proportion of the total needs for each State, the difference lying in the fact that the culvert densities and design flows (and hence the cost of culvert construction) are different for each State. The second set of three columns represents the five-year drainage needs associated with a one percent reduction (i,e., unit reduction) in the design flow associated with each State. This flow

---

* U.S. Department of Transportation, "The 1974 National Highway Needs Report," op. cit. A required report on the Nation's highway needs is submitted to Congress every two years.

Table 8. Marginal Benefits

	5-Yr $ Drainage Needs (millions)			5-Yr $ Drainage Needs (millions) with 1% Reduction in Design flow			Marginal Benefits ($ millions)	% of 5-yr Needs
	Interstate	Primary	Secondary	Interstate	Primary	Secondary		
Alabama	0	87.202	35.707	0	86.616	35.467	.827	.67
Arizona	.775	25.280	8.671	.743	25.130	8.656	.208	.60
Arkansas	.092	65.348	69.535	.0916	64.914	69.074	.895	.66
California	1.419	95.307	38.503	1.375	91.913	37.361	.787	.58
Colorado	.426	16.872	31.442	.415	16.380	30.531	.293	.60
Connecticut	.099	6.695	10.781	.097	6.647	10.702	.129	.73
Delaware	0	5.419	1.590	0	5.387	1.581	.042	.59
Florida	0	24.350	44.952	0	24.198	44.671	.432	.62
Georgia	1.638	52.539	61.174	1.577	52.172	60.745	.858	.74
Idaho	.535	38.491	34.795	.532	38.283	34.608	.398	.54
Illinois	1.519	64.309	48.294	1.510	63.941	48.017	.652	.57
Indiana	.502	32.387	87.178	.499	32.195	86.662	.710	.59
Iowa	2.455	107.115	40.494	2.441	106.503	40.262	.858	.57
Kansas	.909	24.653	30.269	.904	24.504	30.083	.340	.61
Kentucky	1.461	162.320	169.480	1.451	161.193	168.303	2.314	.69
Louisiana	1.595	29.465	40.420	1.586	29.289	40.180	.425	.59
Maine	.503	12.122	44.306	.416	9.986	36.471	.402	.71
Maryland	.732	17.796	10.935	.727	17.676	10.862	.197	.67
Massachusetts	.115	40.598	15.133	.114	40.347	14.891	.493	.88
Michigan	1.794	48.385	58.469	1.748	48.056	58.072	.769	.71
Minnesota	.054	26.402	30.354	.053	26.260	30.189	.307	.54
Mississippi	.594	99.756	110.423	.590	99.066	109.660	1.457	.69
Missouri	1.145	102.788	100.168	1.138	102.134	99.531	1.298	.64
Montana	.704	28.546	34.771	.685	27.704	33.750	.440	.69
Nebraska	0	32.020	18.521	0	31.828	18.296	.417	.82
Nevada	.083	3.114	4.199	.081	2.974	4.008	.048	.65
N. Hampshire	.206	9.225	17.400	.205	9.160	17.278	.190	.71
New Jersey	.961	28.287	22.901	.954	28.104	22.753	.337	.65
New Mexico	0	35.268	29.086	0	33.957	28.000	.397	.62
New York	1.140	36.521	66.958	1.135	36.345	66.634	.504	.48

Table 8. (continued)

	5-Yr $ Drainage Needs (millions)			5-Yr $ Drainage Needs (millions) with 1% Reduction in Design Flow			Marginal Benefits $ Millions	% of 5-yr Needs
	Interstate	Primary	Secondary	Interstate	Primary	Secondary		
No. Carolina	.872	68.935	116.168	.866	68.452	115.355	1.302	.70
No. Dakota	.629	34.867	25.197	.627	34.698	25.074	.296	.49
Ohio	2.054	46.859	142.559	2.041	46.558	141.643	1.230	.64
Oklahoma	.394	42.104	31.644	.392	41.841	31.447	.463	.62
Oregon	2.527	116.136	92.962	2.512	115.430	92.396	1.288	.61
Pennsylvania	7.513	64.838	126.695	7.460	64.384	125.808	1.395	.70
Rhode Island	.015	1.155	.742	.015	1.148	.737	.012	.63
So. Carolina	.423	17.515	16.979	.421	17.399	16.867	.230	.66
So. Dakota	0	15.973	10.121	0	15.885	10.065	.145	.56
Tennessee	.115	69.376	63.841	.095	68.891	63.391	.956	.72
Texas	3.256	92.942	88.552	3.174	90.346	86.015	1.236	.67
Utah	0	14.472	18.937	0	13.781	18.015	.225	.67
Vermont	.0076	8.024	19.061	.0075	7.967	18.926	.191	.71
Virginia	.897	28.186	103.810	.891	28.004	103.139	.860	.65
Washington	1.974	27.424	40.968	1.963	27.254	40.715	.435	.62
W. Virginia	0	32.682	103.496	0	32.444	102.743	.992	.73
Wisconsin	.947	44.129	36.488	.940	43.861	36.265	.495	.61
Wyoming	.844	9.443	15.886	.815	9.084	15.277	.169	.65

reduction is assigned to the entire State and is carried by the blend of Interstate, primary and secondary systems for the State. The difference between the sum of costs in the first three columns and the sum in the second three is the marginal benefit, in millions of dollars for five years, which can be ascribed to a one percent or unit reduction in the design flow. Note that a unit reduction is not a target -- it is merely a scale of performance for the hydrometric network. Some of the largest entries in this column of marginal benefits are for States with large needs and with particularly difficult construction conditions which make the unit culvert costs relatively high.

It should be emphasized that no effort is made to justify or validate the highway and drainage needs in the national survey. There is little reason to believe that the published data represent least-cost solutions to highway needs, but we have no way of disaggregating the expressed values to closely scrutinize them and develop better estimates.

The highway needs reflect a 20 percent reduction from published values; this accounts for anticipated reduction in highway travel consequent to the energy crisis in the winter of 1973-1974. The selection of 20 percent is consistent with recommendations of the FHWA.

Finally, the last column in Table 8 gives the five-year marginal benefits as a percentage of the five-year drainage needs. These numbers have a mean of 0.647 percent with a standard deviation of 0.074 percent; they cluster very closely around their mean value. This is necessary but not sufficient to demonstrate that the method of calculating and presenting potential benefits is valid; it is encouraging to note the remarkable agreement across a wide range of hydrologic and physiographic variation. In subsequent studies, better resolution should be attained.

Figure 23 gives the scheme and order of computation for the economic estimation described in this section. The primary data sources are arrayed across the top, occupying seven columns. These are the

Independent Variables & Functions →

Dependent Variables & Functions ↓

Order of Computation ↓	Primary (see key)							Intermediate						
	Total hwy. cost per mile for 10 regions (W)	Culvert density per mile by state (P)	Average culvert area by state (P)	Cost vs. area (P,G)	Flow vs. area (G)	5-yr highway needs by state (N)	Network statistics (A)	Drainage cost/total cost for 10 regions	No. of culverts to be constructed, by state	Average design flow, Q_d, by state	Revised design flow with more info.	Revised 5-yr culvert cost flow reductn.	Cost savings due to flow reduction	
Drainage cost/total cost for 10 regions	X	X	X	X										INTERMEDIATE
5-yr culvert costs, by state						X		X						INTERMEDIATE
No. of culverts to be constructed by state			X	X					X					INTERMEDIATE
Avg. design flow, Q_d, by state			X		X									INTERMEDIATE
Revised design flow with more information							X				X			INTERMEDIATE
Revised 5-yr culvert cost with flow reduction				X	X					X		X		INTERMEDIATE
Cost savings due to flow reduction									X				X	FINAL

Key to Primary Sources

(P) Highway Plans (9 states)
(W) Winfrey [44]
(N) DOT [40]
(G) Generalized Relationships
(A) Network Analysis

Note: Culvert data are stratfied by pipes and boxes and needs data are stratified by the type of construction

Figure 23. Economic Estimation Scheme

independent variables and source relationships or functions for the six intermediate and one final set of computations arrayed in the first column. The ratio of drainage costs to total costs for each of the ten regions is computed first, using the first four primary data sources as arguments or independent variables. The five-year culvert costs are computed second, using the five-year highway needs and drainage/total costs ratios (just calculated) as the independent variables. Each row uses some primary data sources and some of the intermediate computations produced earlier. The last row uses two sets of earlier intermediate results to produce the final benefit and cost analysis associated with reduction in the design flow. This is identified as FINAL in Figure 23.

Numerical Example

Consider the calculation required for Alabama. The primary data are:

$x_1 = 7.9$	culvert density, number/mile
$x_2 = 0.849$	fraction in pipes, dimensionless
$1 - x_2 = 0.151$	fraction in boxes
$x_3 = 21.80$	unit cost of pipes, \$/ft.
$x_4 = 273$	unit cost of boxes, \$/ft.
$x_5 = 204.4$	Interstate box length, ft.
$x_6 = 155.3$	Interstate pipe length, ft.
$x_7 = 138.8$	primary box length, ft.
$x_8 = 157.8$	primary pipe length, ft.
$x_9 = 89.7$	secondary box length, ft.
$x_{10} = 103.6$	secondary pipe length, ft.
$x_{11} = 1,323,000$	total Interstate cost, \$/mi.
$x_{12} = 465,000$	total primary cost, \$/mi.
$x_{13} = 298,000$	total secondary cost, \$/mi.

$x_{14} = 0$ 5-year Interstate needs, \$-million

$x_{15} = 593.8$ 5-year primary needs, \$-million

$x_{16} = 239.9$ 5-year secondary needs, \$-million.

1. Calculate the total cost per culvert, the drainage cost per mile and the ratio of drainage cost to total cost for Interstate, primary and secondary systems. For each State, highway plans are used to determine the culvert denstiy x_1; the fraction of pipe and box culverts in the state x_2, and $1 - x_2$; the average culvert areas transformed (using Figure 22) to culvert unit costs per foot x_3 and x_4; and total lengths of culvert x_5 through x_{10}. Total costs are from Winfrey.*

Calculate y_1, y_2, y_3 = total cost per culvert for Interstate, primary and secondary systems:

$$y_1 = x_2\, x_3\, x_6 + (1 - x_2)\, x_4\, x_5 = 11{,}300$$

$$y_2 = x_2\, x_3\, x_8 + (1 - x_2)\, x_4\, x_7 = 8{,}642$$

$$y_3 = x_2\, x_3\, x_{10} + (1 - x_2)\, x_4\, x_9 = 5{,}615$$

Calculate y_4, y_5, y_6 = drainage costs per mile for Interstate, primary and secondary systems:

$$y_4 = y_1\, x_1 = 89{,}270$$

$$y_5 = y_2\, x_1 = 68{,}271$$

$$y_6 = y_3\, x_1 = 44{,}359$$

Calculate y_7, y_8, y_9 = ratios of drainage cost to construction cost:

$$y_7 = y_4/x_{11} = 0.067$$

$$y_8 = y_5/x_{12} = 0.147$$

$$y_9 = y_6/x_{13} = 0.149$$

---

* Winfrey, Robley, Economic Analysis for Highways (International Textbook Co., Scranton, Pennsylvania), 1969.

2. Calculate five-year drainage costs for the systems using data from DOT.*

Calculate y_{10}, y_{11}, y_{12} = five-year drainage needs by state:

$$y_{10} = y_7 x_{14} = 0$$

$$y_{11} = y_8 x_{15} = 87.180 \times 10^6$$

$$y_{12} = y_9 x_{16} = 35.71 \times 10^6$$

$$y_{10} + y_{11} + y_{12} = 122.89 \times 10^6$$

3. Divide the drainage needs by the cost per culvert to calculate the number of culverts to be constructed.

$$y_{10}/y_1 + y_{11}/y_2 + y_{12}/y_3 = 0 + 10{,}088 + 6{,}360 = 16{,}448$$

4. Utilizing the average culvert area for each state and Figure 19, calculate the design flows Q_d for pipe and box culverts. For Alabama these flows are 27.8 cfs and 533 cfs, respectively.

5. Determine the economic savings associated with a reduction in the design flow through continuation of the gaging program. Extension of the gaging program for Alabama produces no reduction in the design flow estimate as shown in Table 36, so the specific numeric example is not continued but the generalized procedure is as follows. Based on network statistics (i.e., more information from longer records), a revised design flow is determined from which a revised five-year culvert cost is calculated. Using Figures 19 and 22, determine a culvert cost per foot for a design flow reduction of 25 percent from the <u>initial</u> design flow. This large reduction more accurately extracts costs from Figure 22, that is, the consequences of small reductions (of one percent) are difficult to discern from the graph. Knowing the culvert cost per foot for the reduced design flow permits calculation of marginal cost savings per unit reduction (one percent) by

---

* U.S. Department of Transportation, "The 1974 National Highway Needs Report," <u>op. cit.</u>

dividing the cost differential for 25 percent by 25. Thus we assume linearity of the cost function in this range. The revised five-year culvert cost for the refined design flow is calculated; the cost saving due the flow reduction is easily determined.

RESULTS OF HYDROLOGIC ANALYSIS

General

The hydrologic analysis is designed to identify the extent to which additional information can reduce the design flow so that the cost of such reduction can be compared to the benefit associated with entries in the last column of Table 8. It is assumed the marginal benefits associated with a one percent reduction in design flow can be applied over the full range of potential flow reductions. In other words, it is assumed that the benefit function is linear in the vicinity of the actual decision. The study also deals in five-year benefits because it is posited that a decision to continue gaging implies a minimal institutional commitment of five years. Thus the hydrologic analysis should evaluate the distribution of design flows under the current hydrometric network and its distribution under a network configuration with five additional years of observation, which presumably will have a smaller standard deviation. The consequence of this reduction is a smaller design flow under the same level of risk aversion; the extent of this reduction determines the benefits (or construction savings).

Estimation of Q_{50}

Figures 24 through 34 give cumulative probability densities (or exceedance probabilities) for estimates of Q_{50} for the following 11 States: Georgia, Massachusetts, Missouri, Montana, New Mexico, Ohio, Oregon, Tennessee, Utah, and Wyoming. The two functions plotted on each graph represent alternative methods of estimating Q_{50} from hydrologic records. The smaller values (dots) are calculated using the Water Resources Council (WRC) technique, which fits a log-Pearson

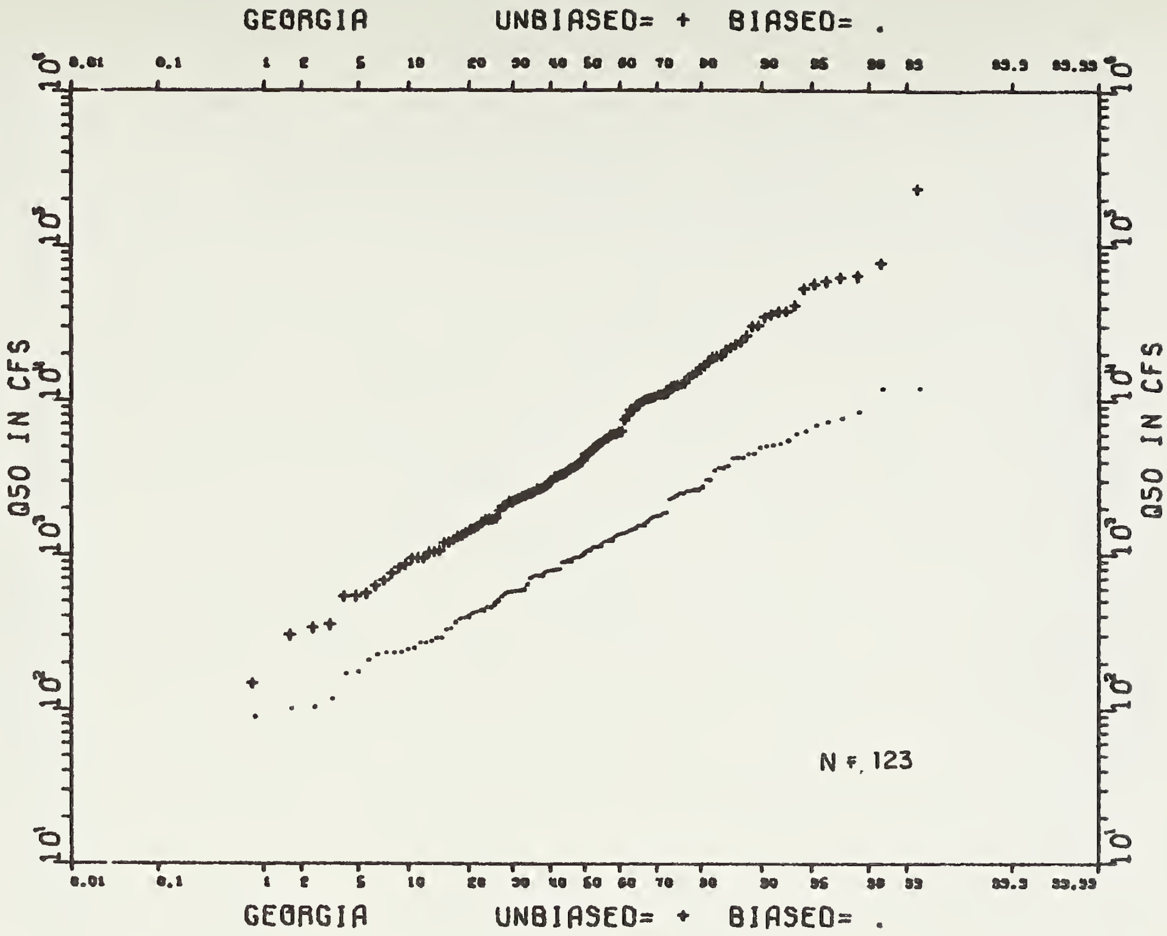


Figure 24. Flow Duration Curve for Georgia

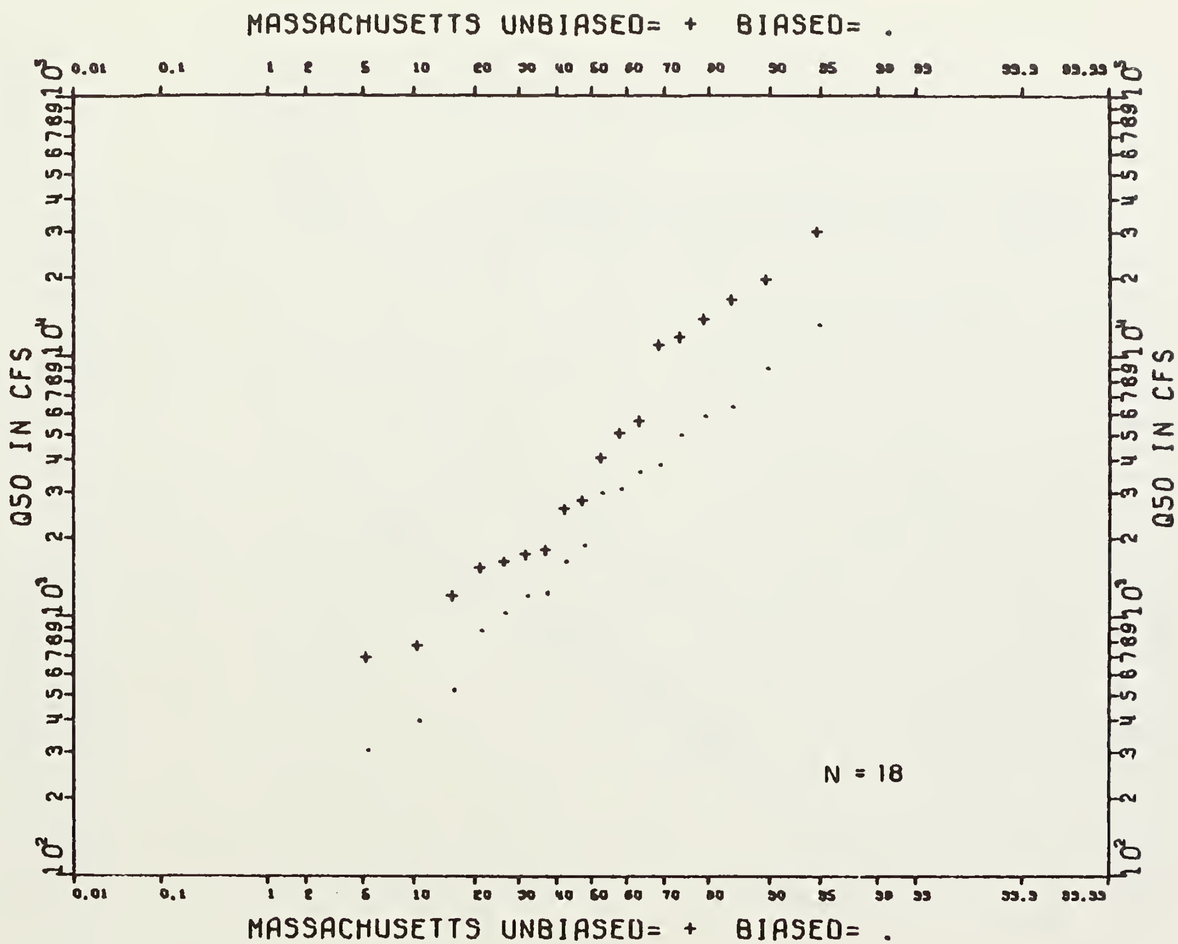


Figure 25. Flow Duration Curve for Massachusetts

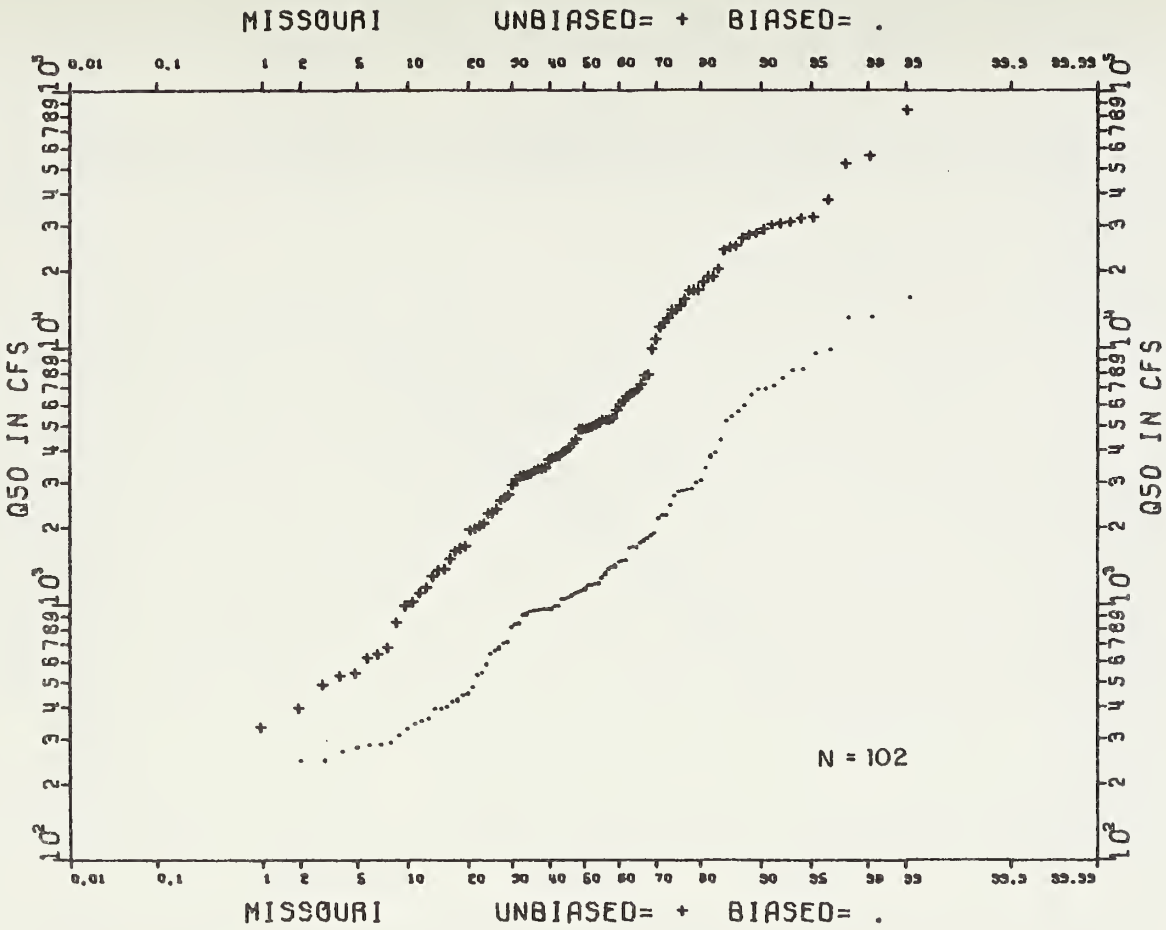


Figure 26. Flow Duration Curve for Missouri

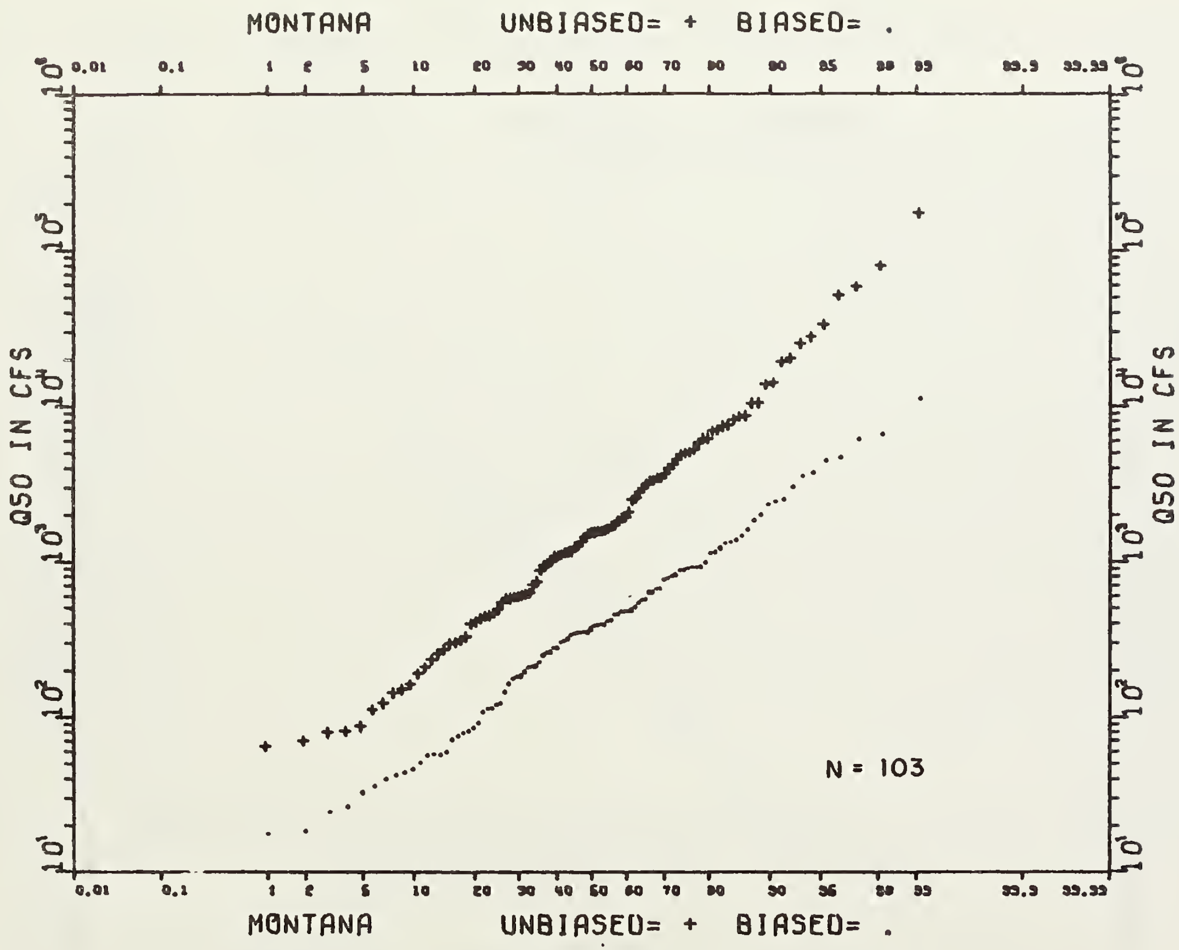


Figure 27. Flow Duration Curve for Montana

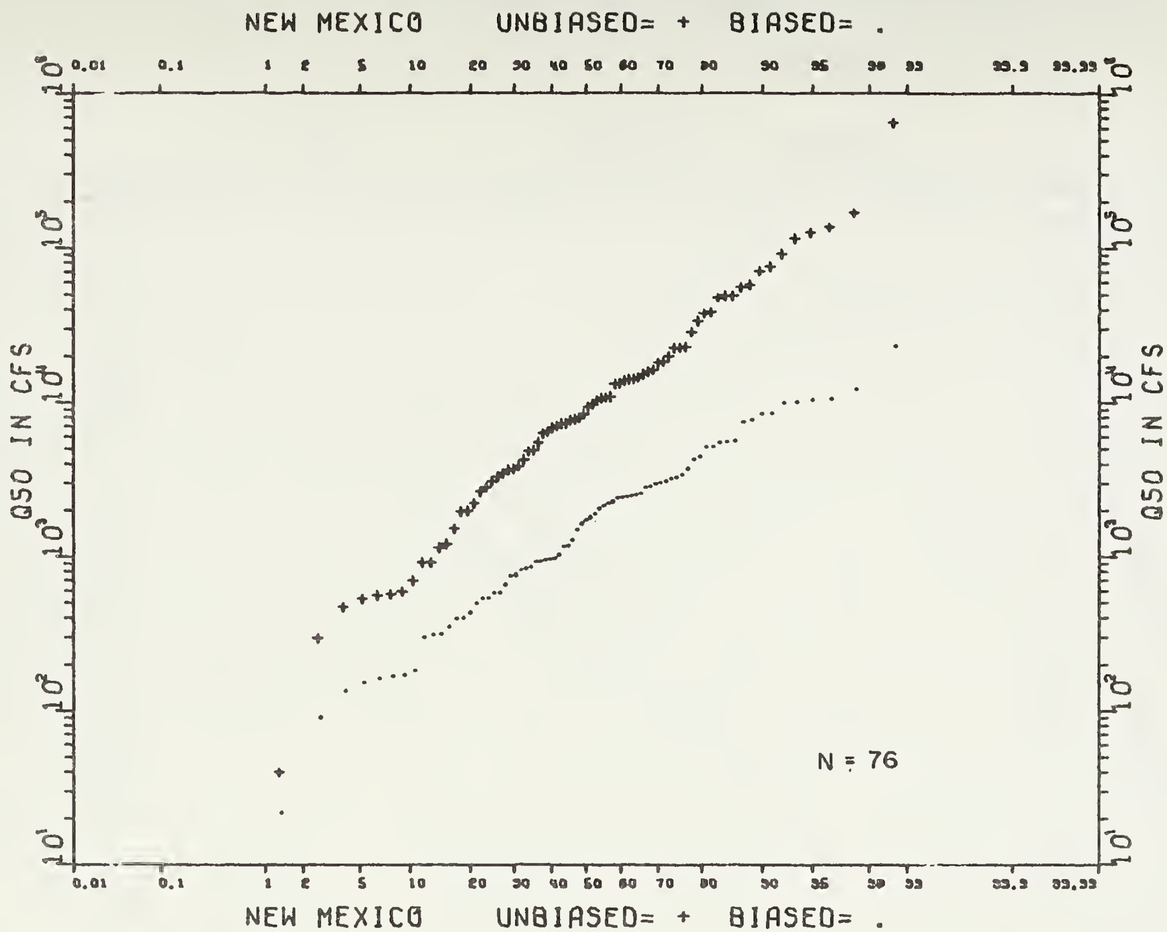


Figure 28. Flow Duration Curve for New Mexico

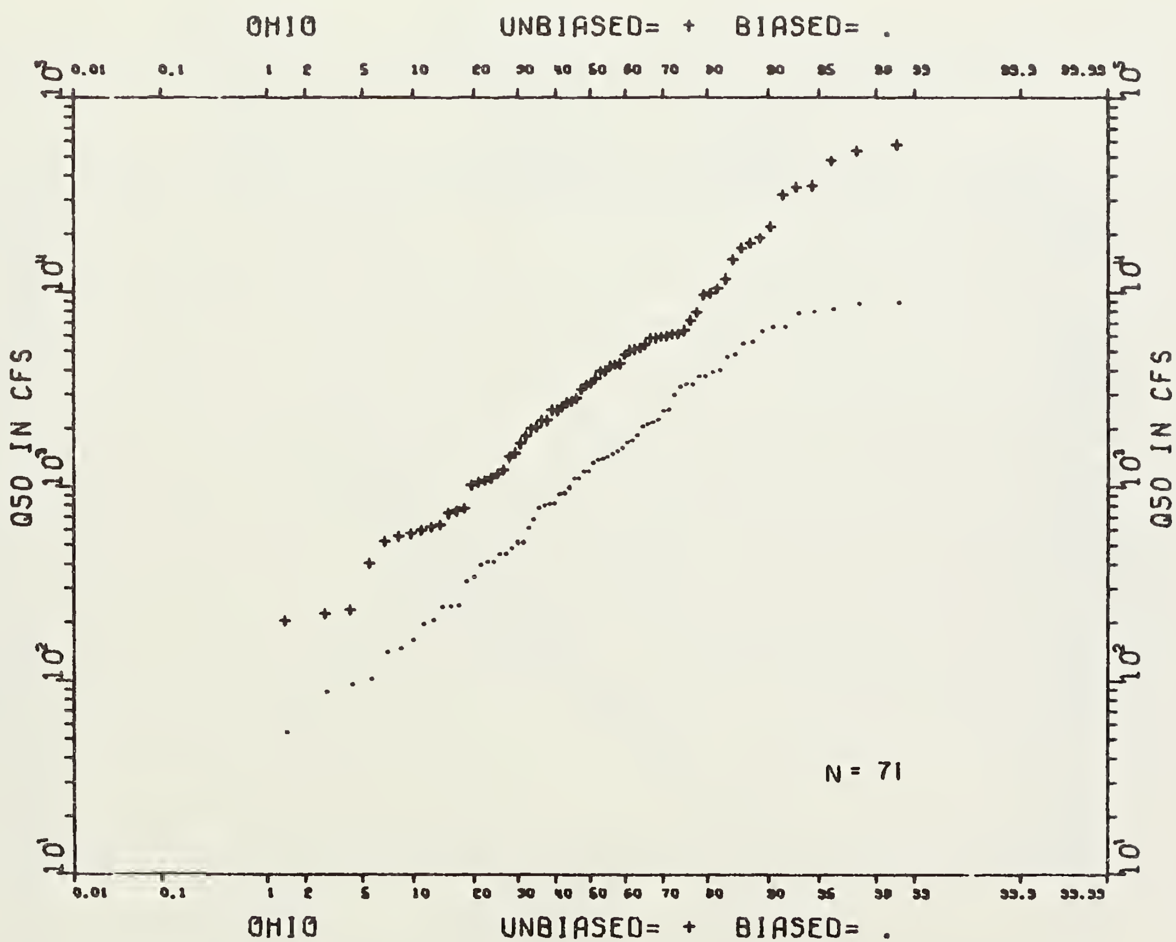


Figure 29. Flow Duration Curve for Ohio

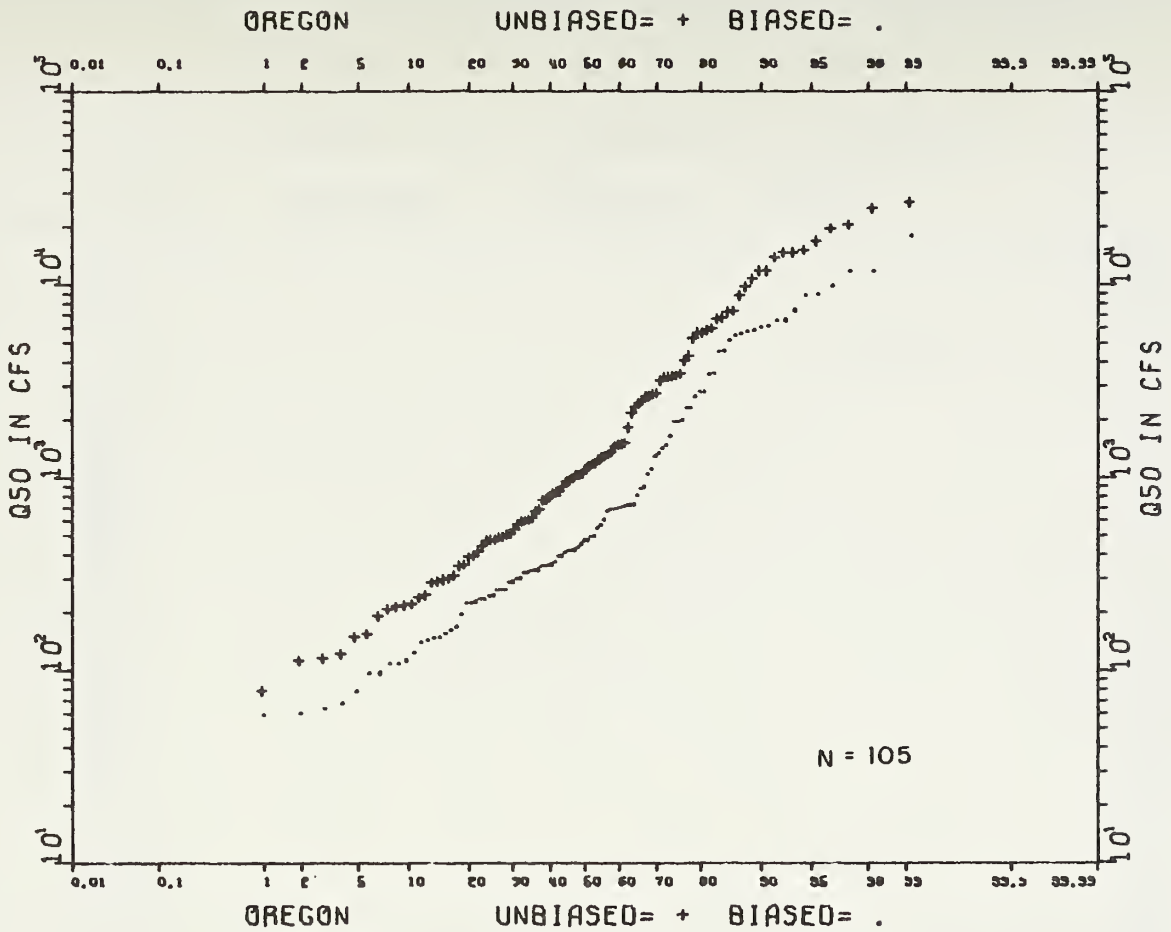


Figure 30. Flow Duration Curve for Oregon

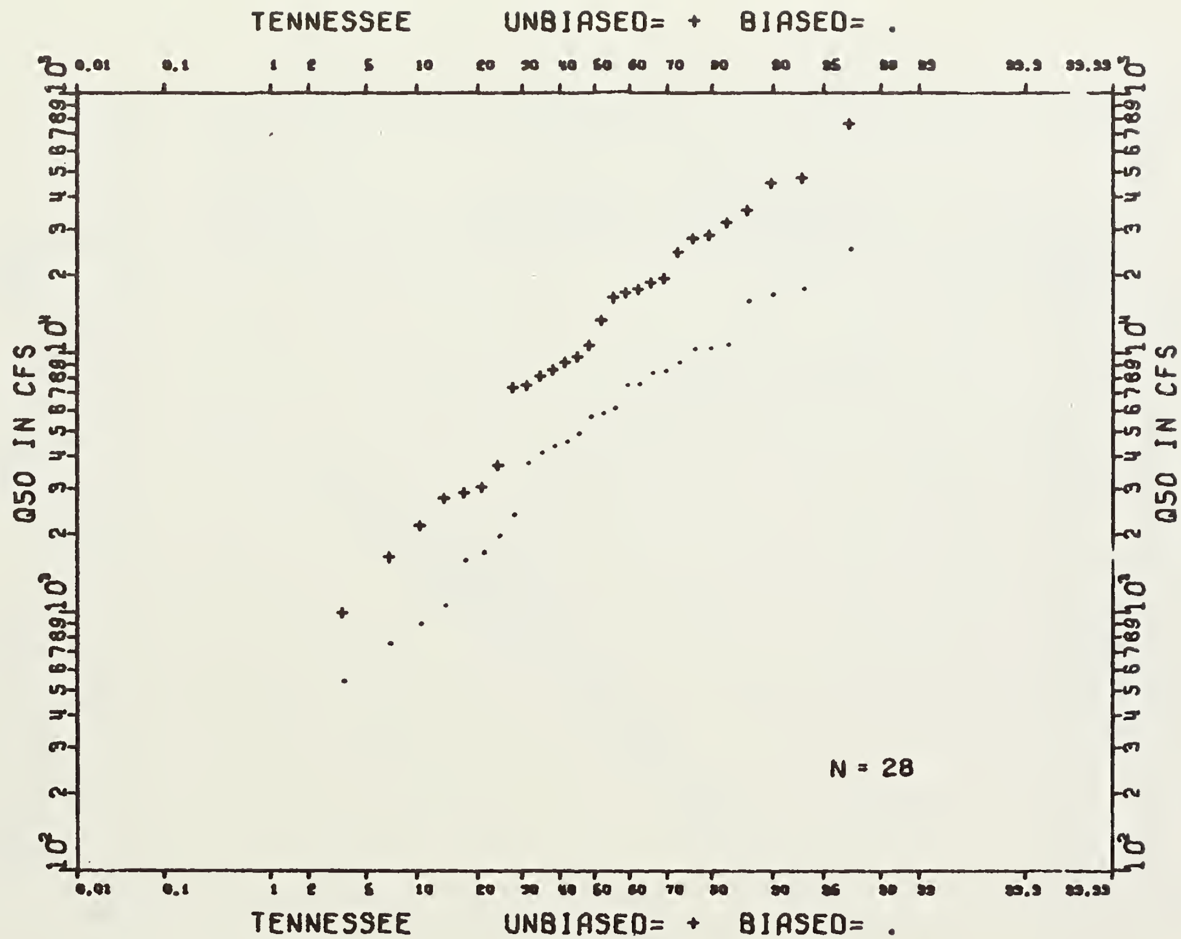


Figure 31. Flow Duration Curve for Tennessee

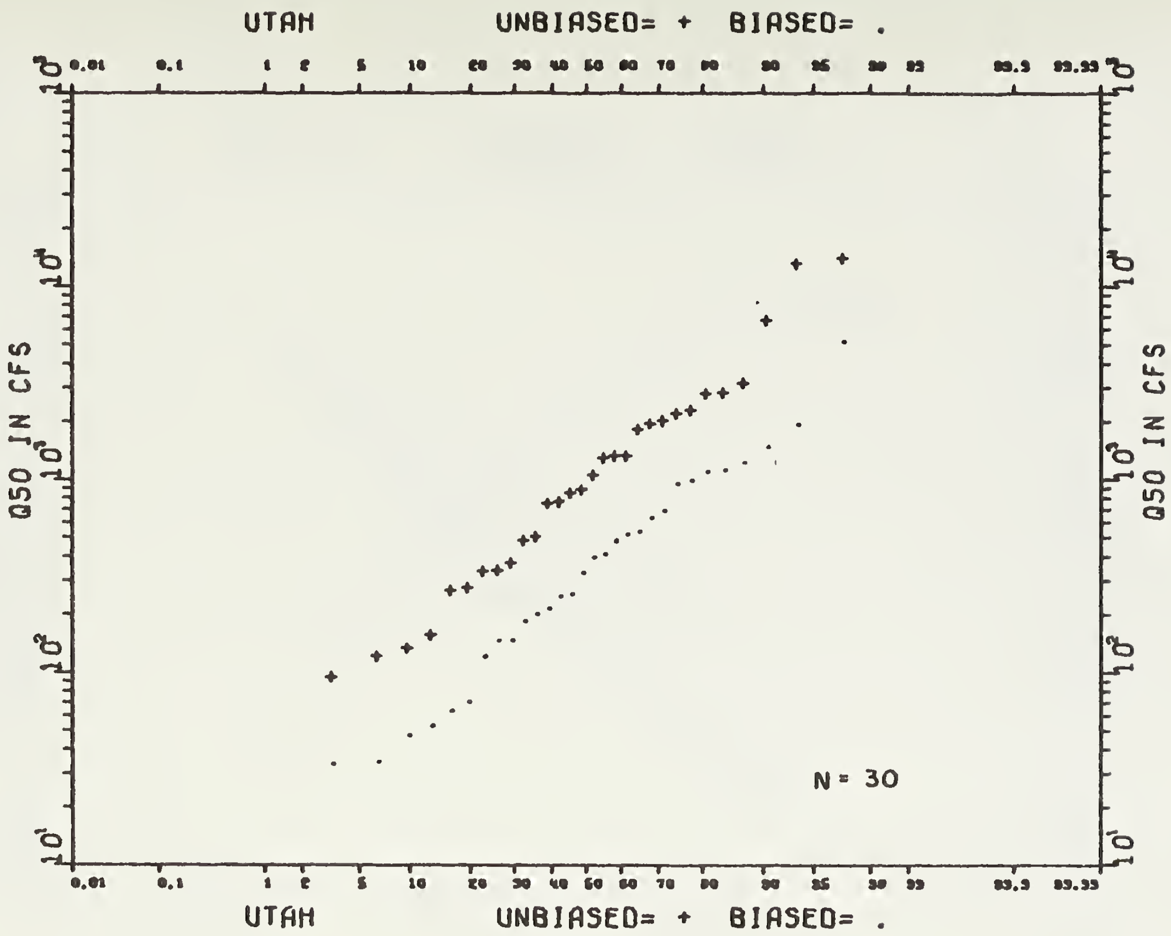


Figure 32. Flow Duration Curve for Utah

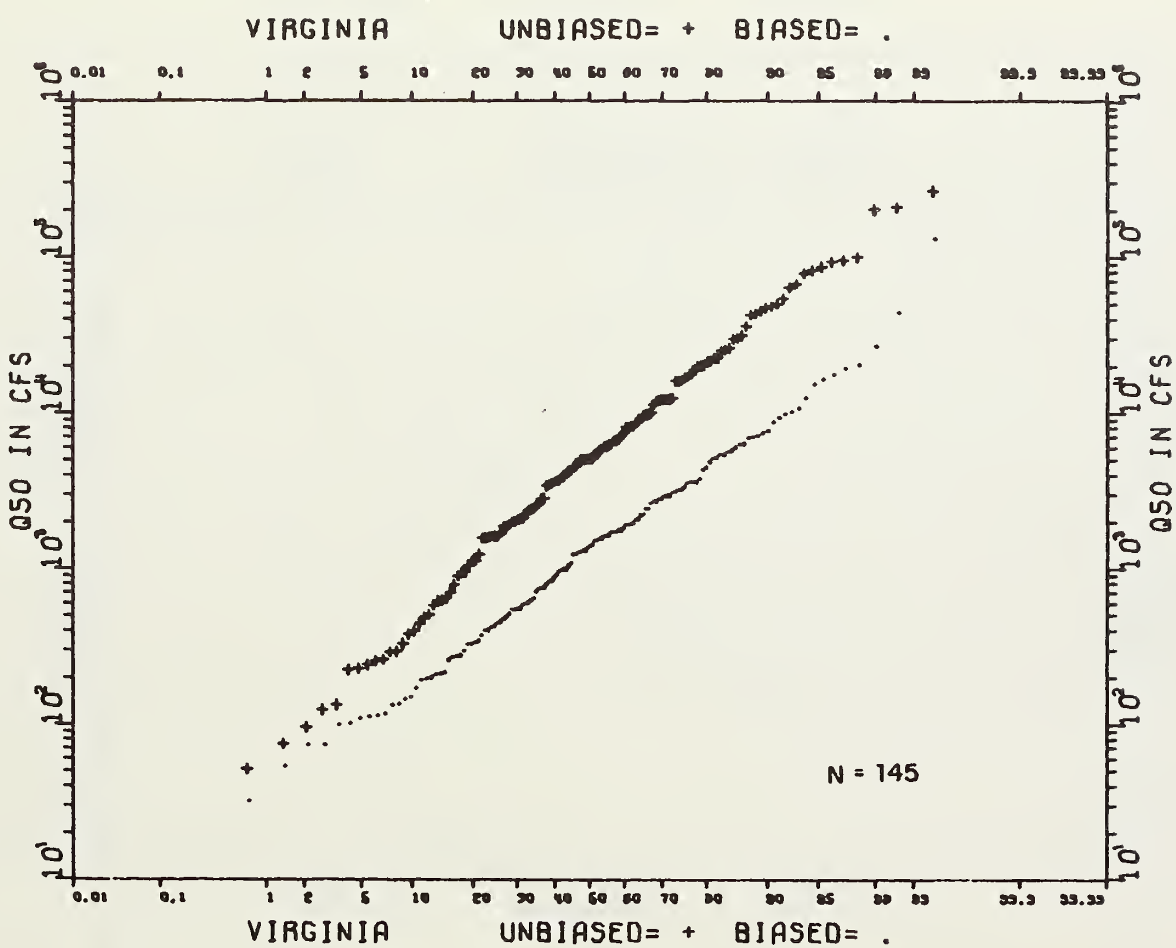


Figure 33. Flow Duration Curve for Virginia

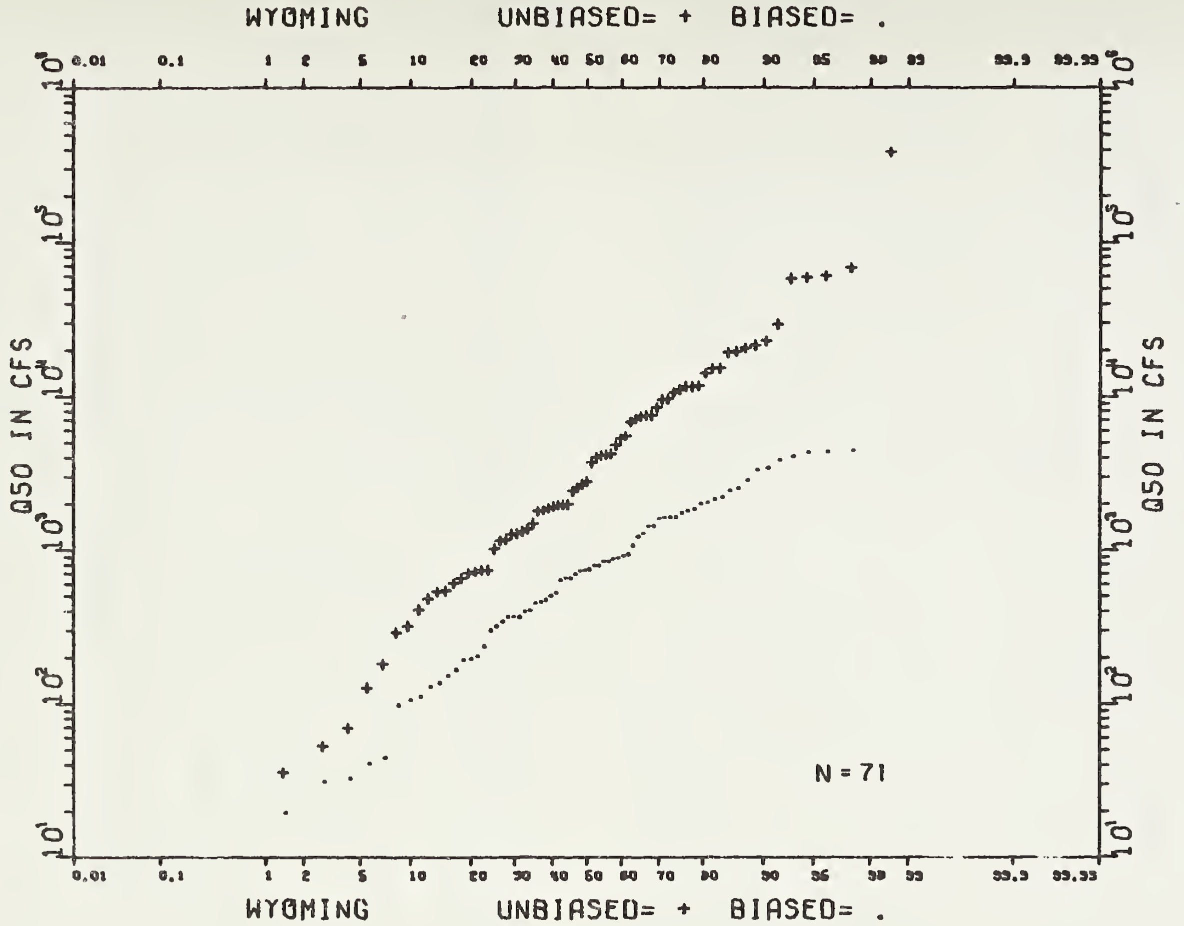


Figure 34. Flow Duration Curve for Wyoming

function; the larger values (crosses) represent the modification introduced by the USGS, which corrects for bias in estimating the moments. In this study the USGS technique is modified by imposing an upper bound on the skew coefficient. This upper bound (set at five) is consistent with the so-called "Kirby bound" for the sample skew coefficient and the bias correction introduced by Wallis et al.* The USGS estimates of Q_{50} become the dependent variables for the state-wide regression analyses which give estimates of the design flow (taken to be estimates Q_{50}) for all States in the analysis. These regressions are discussed below.

The Spearman rank correlation coefficients are tabulated in Table 9. These measure how well the rank order of one sequence is preserved by the rank order of another. In this study the sequences represent the estimates of Q_{50} from all the stations in a given State as calculated by the WRC and WRC* algorithms. The Spearman coefficient is a measure of how closely the two sequences agree in rank (but not in magnitude).

Regression Analysis

Regression analysis was performed using a standard statistical package** run for each State, with the dependent variables being estimates of Q_{50} and the independent variables being drainage area, channel slope, channel length, basin elevation, SCS soil index and precipitation. The program calculates regressions on all the independent variables, performing a stepwise regression on the most significant independent variable, and then on the two most significant, etc. For each combination, the multiple correlation coefficient and the standard error of estimate of the dependent variable are given. All the analyses

---

* Wallis, J. R., et al., "Just a Moment!," WRR, 10: 2, April, 1974.

** No documentation other than program identification is provided for standard programs.

Table 9. Product-Moment and Spearman Correlation Coefficients Between WRC and WRC* Estimates of Q_{50}

State	Sites	Correlation between $\ln \hat{Q}_{50}$	Spearman Rank Correlation	Test Statistic
Georgia	123	.756	.749	12.4
Massachusetts	18	.939	.930	10.1
Missouri	102	.859	.851	16.2
Montana	103	.762	.792	13.0
New Mexico	76	.780	.786	10.9
Ohio	71	.944	.942	23.2
Oregon	105	.950	.952	31.7
Tennessee	28	.961	.955	16.3
Utah	30	.929	.925	12.9
Virginia	145	.900	.887	22.9
Wyoming	71	.707	.659	7.3

in this study were performed using logarithmic transformations (to base *e*) of the raw data, so the coefficients define an exponential relationship among the dependent and independent variables. Table 10 gives the results of these analyses. The independent variables are tabulated in order of decreasing significance, so that where only one independent variable is given it is the most significant, followed by the most significant pair, the most significant triad, etc. In virtually all cases the most significant independent variable is the drainage area, followed by precipitation. However, in some of the analyses there are minor interchanges in the order of significance.

Two sequences provide the dependent variables for the regression. The first of these utilizes WRC estimates of Q_{50}, while the second utilizes the USGS modifications (or WRC*) which take account of bias. These two sets are plotted in Figures 24 through 34. In nine of the eleven States the multiple correlation coefficient for the WRC technique is a little higher than that for the USGS modification, while in two states the unbiased estimates of Q_{50} exhibit a marginally better correlation. These correlations, and the associated standard errors, are tabulated in the last two columns of Table 10. These indicators of the goodness-of-fit of the regressions form the basis of calculating the model error which is required for utilization of BIG-BASIN and, ultimately, estimation of the equivalent years of record.

Based on the equivalence identified by the SCS Land Resource Classification, Table 10 also indicates the other states (shown in parentheses) which are associated with the eleven representative States; there are judgmental issues involved in assignment of these equivalences, but for those few States which could be assigned to more than one representative state, or split among them, the cumulative effect on network decisions of error due to faulty assignment of regression coefficients is small enough to be ignored. It should be emphasized that the important issue here is not specification of the regression coefficients themselves -- the concern is not to define the best model, but rather how measures of a model may be used to effect policy analysis.

Table 10. Regression Coefficients for $\hat{Q}_{50}$ on Basin Characteristics, 11 States, Using WRC and WRC* Estimates

	Constant	Contrib. Area	Channel Slope	Length	Basin Elev.	SCS Index	Precip.	R	S.E.
Georgia (n=123) (Ala, Ark, Fla, La, Miss, N.C., S. .)									
WRC	6.098	.611						.778	.691
	-10.558	.597					4.300	.849	.587
	- 8.644	.609				-1.849	4.523	.867	.552
	- 8.463	.870		-.466		-1.688	4.471	.981	.547
	- 7.298	.864		-.460	.075	-1.690	4.050	.871	.549
	- 7.151	.871	.029	-.454	.050	-1.731	4.039	.871	.551
WRC*	7.689	.551						.557	1.149
	8.570	.559			-.142			.560	1.151
	- 3.912	.570			-.542		3.861	.582	1.135
	- 3.723	.786		-.383	-.549		3.870	.584	1.138
	- 3.227	.755		-.324	-.548	-.509	3.930	.585	1.141
	- 2.673	.780	.110	-.302	-.646	-.663	3.889	.586	1.145
Massachusetts (n=17) (Conn, Me, NH, NY, RI, Vt)									
WRC	5.027	.775						.730	.764
	1.908	.986	.700					.927	.434
	-15.247	.838	.447				4.863	.960	.337
	-12.500	.828	.339		.195		3.911	.965	.329
	-10.890	.833	.354		.208	-.258	3.547	.965	.340
WRC*	-31.600						10.452	.679	.869
	-29.974			1.076			9.325	.803	.722
	-17.967		.527	1.433			5.460	.874	.612
	-11.754	.563	.684	.454			3.832	.901	.566
	- 8.634	.604	.582	.332	.196		2.810	.906	.578
	- 3.863	.703	.619	.121	.249	-.617	1.723	.909	.598
Missouri (n=101) (Iowa, Minn)									
WRC	6.994	.690						.868	.529
	11.654	.692					-1.269	.873	.521
	12.672	.814	.343				-1.947	.882	.507
	12.970	.894	.333	-.149			-1.995	.882	.509
	12.787	.892	.336	-.146	.032		-2.009	.882	.512
WRC*	8.348	.637						.684	.910
	19.631	.642					-3.073	.714	.878
	21.168	.827	.518				-4.096	.730	.861
	20.612	.679	.538	.278			-4.005	.731	.864
	19.500	.670	.554	.298	.198		-4.090	.732	.867
	20.088	.675	.544	.284	.169	.132	-4.225	.732	.871

Table 10

(continued)

	Constant	Contrib. Area	Channel Slope	Length	Basin Elev.	SCS Index	Precip.	R	S.E.
Montana (n=103) (Id,N.D.)									
WRC	9.794		-.768					.573	1.156
	8.660	.393	-.700					.692	1.022
	6.451	.350	-.882				1.156	.743	.954
	9.304	.397	-.801		-.416		1.206	.746	.954
	9.347	.320	-.800	.161	-.438		1.214	.746	.958
WRC*	12.042		-.915					.577	1.362
	11.221	.284	-.866					.624	1.309
	9.875	.258	-.977				.705	.639	1.295
	14.770	.339	-.838		-.714		.791	.647	1.291
New Mexico (n=76) (Ariz, Okl, Tex)									
WRC	6.096			.725				.457	1.201
	35.299			1.022	-3.360**			.709	.958
	29.548			1.071	-2.370		-1.107	.728	.938
	28.808		-.082	1.056	-2.264		-1.020	.729	.943
WRC*	13.607		-.900					.427	1.546
	12.984		-.931	.487				.491	1.500
	36.801		-.436	.726	-3.026			.579	1.414
	35.699	-.300*	-.445	1.259	-2.933			.584	1.417
	34.881	-.268	-.418	1.218	-2.775		- .252	.585	1.426
Ohio (n=71) (Ill., Ind., Mich., Wis.)									
WRC	6.249	.611						.856	.672
	4.151	.857	.504					.899	.572
	-10.550	.800	.358				4.205	.911	.544
	-7.317	.794	.350		-.650**		4.555	.914	.540
	-8.743	.793	.339		-.694	.559	4.870	.916	.539
	-7.562	.674	.352	.220	-.738	.577	4.575	.916	.541
WRC*	6.940			.938				.730	.924
	-20.129			.964			7.435	.783	.847
	-23.530	.548		.037			8.492	.796	.830
	-16.348	.587	.322	.232			6.124	.809	.812
	-10.927	.523	.316	.330	-.998		6.515	.816	.805
	-13.521	.501	.297	.368	-1.093	1.100	7.082	.823	.798

Table 10

(continued)

	Constant	Contrib. Area	Channel Slope	Length	Basin Elev.	SCS Index	Precip.	R	S.E.
Oregon (n=105) (Cal, Wash)									
WRC	5.237	.669						.729	.975
	1.539	.736					.896	.801	.856
	1.124	.829				-.562	1.116	.858	.738
	.678	.857	.070			-.581	1.124	.859	.740
	1.852	.931	.187		-.218	-.498	1.023	.862	.736
	1.795	.786	.211	.300	-.241	-.490	.996	.863	.737
WRC*	6.001	.621						.677	1.047
	6.436	.688				-.498		.733	.972
	3.703	.755				-.614	.687	.774	.910
	2.654	.821	.166			-.657	.706	.779	.905
	3.569	.879	.256		-.169	-.593	.628	.781	.906
	3.512	.735	.280	.298	-.192	-.585	.601	.782	.908
Tennessee (n=28) (Ky, Pa, W Va)									
WRC	5.467	.977						.735	.701
	9.322	.929			-.552			.773	.669
	-8.619	.831			-.494		4.541	.800	.646
	-8.710	.609		.403	-.524		4.564	.807	.649
	-9.011	.558		.488	-.480	-1.031	4.905	.813	.655
	-8.849	.520	.147	.669	-.656	-.930	4.952	.814	.669
WRC*	6.349	.960						.662	.846
	12.346	.885			-.858			.745	.768
	-4.976	.791			-.802		4.384	.767	.754
	-4.815	.848	.135		-.947		4.449	.769	.767
	-4.584	.546	.480	.811	-1.378		4.661	.780	.767
	-4.492	.554	.495	.810	-1.408	.267	4.581	.781	.785
Utah (n=30) (Col, Nev)									
WRC	2.465	1.215						.871	.639
	3.301	1.392				-.726		.921	.519
	5.727	1.297				-.613	-.721	.929	.500
	5.127	1.381	.156			-.642	-.860	.932	.500
	5.057	1.471	.180	-.155		-.668	-.838	.933	.507
	1.707	1.489	.187	-.209	.403	-.716	-.893	.933	.515
WRC*	3.661	1.178						.831	.736
	4.532	1.363				-.757		.886	.623
	5.302	1.693		-.713		-.828		.907	.577
	8.492	1.520		-.617		-.667	-.978	.921	.546
	7.254	1.771	.327	-.768		-.746	-1.228	.932	.517
	-2.338	1.822	.347	-.921	1.155	-.883	-1.386	.938	.505

Table 10

(continued)

	Constant	Contrib. Area	Channel Slope	Length	Basin Elev.	SCS Index	Precip.	R	S.E.
Virginia (n=145) (Del, Md, NJ)									
WRC	6.190	.673*						.669	1.112
	11.236	.679					-1.348	.673	1.111
	11.458	.792		-.210			-1.380	.674	1.114
	10.584	.816		-.265	.052		-1.229	.675	1.116
	10.804	.785	-.140	-.333	.154		-1.227	.677	1.117
WRC*	7.678	.550						.465	1.556
	14.476	.558					-1.816	.472	1.555
	14.763	.704		-.272			-1.858	.473	1.559
	14.125	.721		-.312	.038		-1.747	.474	1.564
	14.388	.684	-.167	-.393	.159		-1.793	.478	1.566
Wyoming(n=70) (Kan, Neb, SD)									
WRC	5.451	.499						.488	1.172
	32.854	.829			-3.174			.728	.928
	27.759	.924			-2.406	-1.294		.762	.883
	28.095	1.394		-.823	-2.405	-1.207		.778	.864
	28.831	1.396	.118	-.783	-2.557	-1.239		.779	.868
	28.514	1.429	.111	-.846	-2.488	-1.141	-.140	.779	.874
WRC*	42.283				-3.872			.573	1.472
	53.318	.541			-5.244			.666	1.350
	46.106	.601			-3.985		-1.561	.727	1.252
	45.334	1.054		-.773	-3.802		-1.726	.735	1.247
	44.617	1.033		-.704	-3.710	-.315	-1.613	.735	1.255
	44.810	1.032	.029	-.691	-3.751	-.327	-1.608	.735	1.265

* The Contribution Area is not available for New Mexico and Virginia; the total basin area is used instead.

** The Basin Elevation is not available for New Mexico and Ohio; the channel elevation is used instead.

Legend

R: Correlation Coefficient
S.E.: Standard Error
n: Sample Size

Other Arguments for Use of BIGBASIN

Tables 11 through 21 are correlation matrices for the annual floods measured at pairs of stations within each State, one matrix per State. All available sites are not utilized in each matrix; the computational burden would be enormous and would afford little advantage. Calculations indicate that stable values of the State or regional correlations are obtained for approximately 15 sites. Moreover, for larger numbers of sites, numerous correlation inconsistencies are encountered. Because the annual flood data do not form a rectangular array of simultaneous observations at all sites, some combinations will undoubtedly develop for which the multiple correlation coefficients exhibit infeasible values.

For example, consider a simple three-stream array in which site 1 contains data from (say) 1940-1970, site 2 from 1940-1955, and site 3 from 1954-1970. Sites 2 and 3 have only two years in common: 1954 and 1955. Therefore, because these two data points uniquely determine a straight line, the sample correlation coefficient $\hat{\rho}_{23}$ is unity. Therefore the correlation $\hat{\rho}_{12}$ must equal the correlation $\hat{\rho}_{13}$ because $\hat{\rho}_{23}$ is unity so having values at site 2 is mathematically equivalent to having values at site 3 (they can be mapped linearly and unambiguously into each other). Equality between correlations $\hat{\rho}_{12}$ and $\hat{\rho}_{13}$ will virtually never occur; the correlation matrix derived from these sample estimates of the correlation coefficients is called <u>inconsistent</u>. This extreme example represents the difficulties that are encountered as the number of sites in a State or region grows large; the larger the number of sites, the more likely some of the anomolies related to non-overlapping or briefly overlapping records will exist. Fiering* proposed a correction for this condition; a surrogate for this correction, computationally less extensive, is used in this study.

---

* Fiering, Myron B, "Schemes for Handling Inconsistent Matrices," <u>WRR</u>, <u>4</u>: 2, April 1968.

Table 11. ρ_a, CORRELATIONS BETWEEN Q_a-VALUES, AND OVERLAPPING RECORD LENGTHS, 15 SITES
GEORGIA

COR MATRIX R(I,J)= COR OF I AND J IF I<J AND=# OF YEARS IN COMMON FOR I>J

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000	0.389	0.265	0.000	0.441	0.156	0.036	0.196	-0.231	-0.030	0.078	0.001	0.713	0.264	0.555
2	20.000	1.000	0.375	-0.039	0.956	0.014	0.275	0.045	-0.025	0.137	-0.032	0.183	0.156	0.037	0.322
3	19.000	24.000	1.000	0.114	0.503	0.347	0.341	0.534	0.214	0.544	-0.052	0.086	0.267	0.504	0.152
4	18.000	20.000	20.000	1.000	0.100	-0.042	-0.126	-0.148	0.267	0.548	-0.207	-0.263	0.109	0.129	-0.037
5	19.000	23.000	23.000	20.000	1.000	0.019	0.339	0.176	0.196	0.391	-0.052	0.096	0.260	0.327	0.516
6	19.000	23.000	23.000	20.000	23.000	1.000	0.307	0.200	0.284	0.296	0.027	-0.164	0.029	0.264	-0.026
7	19.000	25.000	24.000	20.000	23.000	23.000	1.000	0.461	0.320	0.112	0.300	0.474	0.016	0.215	0.256
8	19.000	24.000	24.000	20.000	23.000	23.000	24.000	1.000	0.407	0.313	0.325	0.436	0.232	0.465	0.405
9	19.000	24.000	24.000	20.000	23.000	23.000	24.000	24.000	1.000	0.729	0.171	0.174	-0.119	0.426	0.320
10	19.000	20.000	20.000	19.000	20.000	20.000	20.000	20.000	20.000	1.000	-0.160	-0.123	-0.044	0.502	0.225
11	19.000	23.000	23.000	20.000	23.000	23.000	23.000	23.000	23.000	20.000	1.000	0.749	-0.124	0.209	-0.050
12	19.000	26.000	23.000	20.000	23.000	23.000	24.000	23.000	23.000	20.000	23.000	1.000	-0.074	0.273	0.094
13	19.000	23.000	23.000	20.000	23.000	23.000	23.000	23.000	23.000	20.000	23.000	23.000	1.000	0.299	0.409
14	19.000	24.000	24.000	20.000	23.000	23.000	24.000	24.000	24.000	20.000	23.000	23.000	23.000	1.000	0.347
15	20.000	32.000	24.000	20.000	23.000	23.000	25.000	24.000	24.000	20.000	23.000	26.000	23.000	24.000	1.000

105 CROSSCR, MEAN= 0.205

Table 12. ρ_a, CORRELATIONS BETWEEN Q_a-VALUES, AND OVERLAPPING RECORD LENGTHS, 15 SITES MASSACHUSETTS

COR MATRIX R(I,J)= COR OF I AND J IF I<J AND=# OF YEARS IN COMMON FOR I>J

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000	0.927	0.325	0.503	0.263	0.022	0.084	0.444	0.093	0.107	-0.115	-0.101	-0.144	-0.011	0.406
2	29.000	1.000	0.454	0.608	0.266	0.208	0.083	0.480	0.132	0.195	-0.080	0.112	0.072	0.137	0.459
3	29.000	35.000	1.000	0.905	-0.003	0.451	0.507	0.336	0.337	0.323	0.726	-0.029	-0.199	-0.073	0.250
4	28.000	36.000	34.000	1.000	0.249	0.371	0.394	0.350	0.229	0.235	0.477	0.211	0.128	0.050	0.390
5	28.000	36.000	34.000	48.000	1.000	0.803	0.449	0.580	0.481	0.947	0.022	0.700	0.847	0.698	0.587
6	29.000	37.000	35.000	48.000	56.000	1.000	0.698	0.461	0.747	0.772	0.636	0.633	0.703	0.617	0.397
7	28.000	35.000	34.000	35.000	35.000	35.000	1.000	0.571	0.623	0.465	0.751	0.330	0.353	0.397	0.372
8	27.000	27.000	27.000	27.000	27.000	27.000	27.000	1.000	0.374	0.608	0.319	-0.002	0.029	-0.110	0.109
9	27.000	27.000	27.000	26.000	26.000	27.000	26.000	26.000	1.000	0.579	0.482	0.054	0.193	0.151	0.493
10	29.000	37.000	35.000	37.000	37.000	38.000	35.000	27.000	27.000	1.000	0.724	0.569	0.755	0.668	0.370
11	28.000	28.000	28.000	28.000	28.000	28.000	28.000	27.000	26.000	28.000	1.000	0.216	0.146	0.096	0.041
12	28.000	36.000	34.000	38.000	38.000	38.000	35.000	27.000	26.000	37.000	28.000	1.000	0.917	0.874	0.569
13	28.000	36.000	34.000	42.000	42.000	42.000	35.000	27.000	26.000	37.000	28.000	38.000	1.000	0.872	0.548
14	28.000	36.000	34.000	43.000	43.000	43.000	35.000	27.000	26.000	37.000	28.000	38.000	42.000	1.000	0.737
15	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	1.000

105 CROSSCR, MEAN= 0.378

Table 13. ρ_a, CORRELATIONS BETWEEN Q_a-VALUES, AND OVERLAPPING RECORD LENGTHS, 15 SITES

MISSOURI

COR MATRIX R(I,J)= COR OF I AND J IF I<J AND=# OF YEARS IN COMMON FOR I>J

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000	0.230	0.247	-0.117	-0.090	[illegible]	0.456	0.364	0.099	0.406	0.059	0.370	0.084	0.076	[illegible]
2	24.000	1.000	0.834	0.288	0.424	0.234	0.011	0.440	0.157	0.076	-0.130	-0.010	-0.320	-0.030	0.374
3	24.000	24.000	1.000	0.323	0.377	0.183	-0.103	0.475	0.233	-0.113	-0.278	-0.086	-0.383	-0.101	0.22[illegible]
4	24.000	25.000	24.000	1.000	0.423	0.120	0.289	0.201	0.064	-0.130	-0.034	0.022	0.173	0.085	0.245
5	19.000	20.000	19.000	20.000	1.000	0.351	0.370	0.039	-0.147	0.166	0.291	-0.054	-0.202	-0.115	-0.010
6	19.000	20.000	19.000	20.000	20.000	1.000	0.340	0.449	0.053	0.083	0.292	0.362	0.012	0.148	0.376
7	19.000	20.000	19.000	20.000	20.000	20.000	1.000	0.224	-0.452	0.667	0.418	0.266	0.220	0.069	0.012
8	21.000	22.000	21.000	22.000	17.000	17.000	17.000	1.000	0.250	0.167	-0.076	0.025	0.064	0.242	0.511
9	24.000	24.000	24.000	24.000	19.000	19.000	19.000	21.000	1.000	-0.001	0.100	0.106	-0.079	0.060	0.367
10	25.000	25.000	25.000	25.000	20.000	20.000	20.000	22.000	25.000	1.000	0.395	0.300	0.005	-0.050	0.291
11	26.000	25.000	25.000	25.000	20.000	20.000	20.000	22.000	25.000	26.000	1.000	0.633	0.134	0.022	-0.265
12	26.000	24.000	24.000	24.000	19.000	19.000	19.000	21.000	24.000	25.000	26.000	1.000	0.263	-0.156	0.122
13	18.000	19.000	18.000	19.000	19.000	19.000	19.000	16.000	19.000	19.000	19.000	18.000	1.000	-0.213	-0.123
14	24.000	25.000	24.000	25.000	20.000	20.000	20.000	22.000	24.000	25.000	25.000	24.000	19.000	1.000	0.298
15	23.000	24.000	23.000	24.000	20.000	20.000	20.000	21.000	23.000	24.000	24.000	23.000	19.000	24.000	1.000

105 CROSSCR, MEAN= 0.142

Table 14. ρ_a, CORRELATIONS BETWEEN Q_a-VALUES, AND OVERLAPPING RECORD LENGTHS, 15 SITES

MONTANA

COR MATRIX R(I,J)= COR OF I AND J IF I<J AND=# OF YEARS IN COMMON FOR I>J

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000	0.875	0.349	-0.128	0.038	-0.186	-0.031	-0.233	0.712	-0.048	-0.237	-0.341	0.032	0.024	0.015
2	24.000	1.000	0.424	-0.181	-0.046	-0.117	-0.011	-0.150	0.825	0.057	-0.149	-0.308	-0.097	-0.003	-0.008
3	23.000	27.000	1.000	0.154	-0.169	0.159	0.412	0.157	0.321	0.191	0.234	0.258	0.084	0.504	0.107
4	17.000	17.000	17.000	1.000	0.501	0.004	0.021	-0.166	-0.178	0.115	0.404	0.124	0.058	-0.005	-0.024
5	24.000	33.000	27.000	17.000	1.000	0.314	0.205	-0.300	0.214	0.502	0.082	0.096	0.369	0.129	0.514
6	19.000	19.000	19.000	15.000	19.000	1.000	0.415	-0.072	-0.002	0.791	-0.124	0.402	-0.107	0.059	0.265
7	19.000	19.000	19.000	15.000	19.000	19.000	1.000	-0.137	0.193	0.469	-0.167	0.251	0.157	0.489	0.563
8	18.000	18.000	18.000	14.000	18.000	18.000	18.000	1.000	0.020	-0.103	0.140	0.081	-0.130	-0.175	-0.338
9	18.000	18.000	18.000	15.000	18.000	18.000	18.000	17.000	1.000	0.262	-0.224	-0.207	-0.015	-0.089	-0.108
10	19.000	19.000	19.000	15.000	19.000	19.000	19.000	18.000	18.000	1.000	-0.139	0.409	0.061	-0.132	0.042
11	19.000	19.000	19.000	15.000	19.000	19.000	19.000	18.000	18.000	19.000	1.000	0.187	0.215	-0.005	0.022
12	19.000	19.000	19.000	15.000	19.000	18.000	18.000	17.000	18.000	18.000	18.000	1.000	0.368	-0.051	-0.205
13	19.000	19.000	19.000	15.000	19.000	19.000	19.000	18.000	18.000	19.000	19.000	18.000	1.000	0.371	-0.065
14	16.000	16.000	16.000	15.000	16.000	16.000	16.000	15.000	16.000	16.000	16.000	16.000	16.000	1.000	0.675
15	20.000	22.000	22.000	17.000	22.000	16.000	16.000	15.000	16.000	16.000	16.000	17.000	16.000	16.000	1.000

105 CROSSCR, MEAN= 0.102

Table 15. ρ_q, CORRELATIONS BETWEEN Q_q-VALUES, AND OVERLAPPING RECORD LENGTHS, 15 SITES

NEW MEXICO

COR MATRIX R(I,J)= COR OF I AND J IF I<J AND=# OF YEARS IN COMMON FOR I>J

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000	0.400	0.211	0.542	0.557	0.196	-0.327	0.361	-0.164	-0.090	-0.485	0.450	0.040	-0.314	-0.393
2	34.000	1.000	0.674	0.493	0.570	0.290	0.030	-0.055	-0.130	0.250	-0.175	0.138	-0.323	0.023	0.137
3	33.000	36.000	1.000	0.661	0.684	-0.101	0.221	-0.157	-0.243	-0.232	-0.310	-0.076	-0.370	0.031	0.305
4	35.000	36.000	35.000	1.000	0.830	-0.064	0.032	0.042	-0.298	-0.240	-0.502	-0.321	-0.280	-0.014	0.395
5	37.000	37.000	36.000	38.000	1.000	0.022	-0.029	0.013	-0.313	-0.245	-0.522	-0.216	-0.304	-0.056	0.244
6	29.000	28.000	27.000	29.000	31.000	1.000	-0.048	0.033	0.685	0.112	-0.410	0.205	-0.078	-0.086	-0.252
7	28.000	30.000	29.000	30.000	31.000	27.000	1.000	-0.230	0.190	0.029	-0.041	-0.402	0.260	0.109	0.174
8	19.000	22.000	22.000	21.000	22.000	13.000	15.000	1.000	0.111	0.142	-0.097	-0.013	-0.227	-0.103	-0.296
9	17.000	20.000	20.000	19.000	20.000	12.000	14.000	20.000	1.000	-0.046	0.556	0.106	-0.067	-0.174	-0.011
10	18.000	21.000	21.000	20.000	21.000	13.000	14.000	21.000	19.000	1.000	0.258	0.135	0.309	0.324	-0.065
11	15.000	16.000	16.000	15.000	16.000	7.000	10.000	16.000	15.000	15.000	1.000	0.037	-0.053	-0.143	-0.049
12	17.000	20.000	20.000	19.000	20.000	11.000	14.000	20.000	19.000	19.000	16.000	1.000	-0.025	0.032	-0.335
13	18.000	21.000	21.000	20.000	21.000	12.000	14.000	21.000	19.000	20.000	15.000	19.000	1.000	-0.084	-0.129
14	22.000	25.000	24.000	24.000	25.000	16.000	19.000	21.000	20.000	20.000	16.000	20.000	20.000	1.000	0.368
15	18.000	21.000	21.000	20.000	21.000	14.000	15.000	20.000	19.000	19.000	14.000	18.000	19.000	19.000	1.000

105 CROSSCR, MEAN= 0.021

Table 16. ρ_a, CORRELATIONS BETWEEN Q_a-VALUES, AND OVERLAPPING RECORD LENGTHS, 15 SITES

OHIO

COR MATRIX R(I,J)= COR OF I AND J IF I<J AND=# OF YEARS IN COMMON FOR I>J

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000	0.437	0.243	-0.018	0.064	0.032	0.351	0.279	0.025	0.245	0.251	-0.106	0.373	0.280	0.287
2	27.000	1.000	0.163	-0.008	-0.013	0.217	0.346	0.060	-0.028	0.093	0.241	-0.067	0.135	0.363	0.232
3	32.000	27.000	1.000	0.260	0.574	0.171	0.409	0.127	-0.020	0.013	0.284	0.124	0.099	0.417	0.337
4	30.000	25.000	35.000	1.000	0.771	0.600	0.002	-0.002	-0.073	-0.207	-0.026	0.703	0.388	0.270	-0.030
5	30.000	25.000	35.000	35.000	1.000	0.696	0.328	0.217	-0.119	-0.031	0.100	0.480	0.415	0.424	0.176
6	30.000	25.000	33.000	33.000	33.000	1.000	0.075	-0.049	-0.139	-0.070	0.014	0.360	0.217	0.346	-0.044
7	27.000	27.000	27.000	25.000	25.000	25.000	1.000	0.713	-0.019	0.356	0.472	-0.110	0.562	0.472	0.853
8	27.000	27.000	27.000	25.000	25.000	25.000	27.000	1.000	0.409	0.624	0.457	-0.215	0.357	0.362	0.634
9	27.000	27.000	27.000	25.000	25.000	25.000	27.000	27.000	1.000	0.766	0.345	-0.067	-0.297	-0.063	-0.045
10	26.000	26.000	26.000	24.000	24.000	24.000	26.000	26.000	26.000	1.000	0.506	-0.074	-0.012	0.135	0.203
11	27.000	27.000	27.000	25.000	25.000	25.000	27.000	27.000	27.000	26.000	1.000	0.011	0.300	0.415	0.477
12	29.000	27.000	31.000	29.000	29.000	29.000	27.000	27.000	27.000	26.000	27.000	1.000	0.364	0.060	-0.175
13	27.000	27.000	27.000	25.000	25.000	25.000	27.000	27.000	27.000	26.000	27.000	27.000	1.000	0.516	0.428
14	27.000	27.000	27.000	25.000	25.000	25.000	27.000	27.000	27.000	26.000	27.000	27.000	27.000	1.000	0.247
15	27.000	27.000	27.000	25.000	25.000	25.000	27.000	27.000	27.000	26.000	27.000	27.000	27.000	27.000	1.000

105 CROSSCR, MEAN= 0.219

Table 17. ρ_a, CORRELATIONS BETWEEN Q_a-VALUES, AND OVERLAPPING RECORD LENGTHS, 15 SITES

OREGON

COR MATRIX R(I,J)= COR OF I AND J IF I<J AND=# OF YEARS IN COMMON FOR I>J

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000	0.150	0.070	0.329	0.038	-0.051	0.115	0.200	0.138	0.252	-0.034	0.182	0.287	0.415	0.316
2	49.000	1.000	0.029	0.460	0.366	0.020	0.327	0.057	0.004	0.142	-0.208	0.143	0.206	0.347	0.259
3	38.000	38.000	1.000	0.303	0.230	0.598	0.508	0.393	0.131	0.308	0.297	0.434	0.395	0.391	-0.134
4	38.000	38.000	32.000	1.000	0.605	0.500	0.569	0.501	0.539	0.509	0.205	0.785	0.691	0.666	0.141
5	55.000	50.000	38.000	39.000	1.000	0.214	0.053	0.022	0.133	0.218	0.042	0.458	0.438	0.262	0.067
6	49.000	48.000	38.000	37.000	48.000	1.000	0.543	0.367	0.305	0.509	0.397	0.620	0.605	0.565	-0.181
7	22.000	22.000	17.000	22.000	22.000	22.000	1.000	0.687	0.209	0.279	0.303	0.487	0.629	0.524	0.237
8	40.000	40.000	35.000	37.000	40.000	40.000	22.000	1.000	0.499	0.437	0.312	0.457	0.426	0.399	0.380
9	34.000	34.000	29.000	34.000	34.000	34.000	22.000	34.000	1.000	0.662	0.350	0.647	0.567	0.544	0.103
10	23.000	23.000	18.000	23.000	23.000	23.000	22.000	23.000	23.000	1.000	0.181	0.665	0.657	0.685	0.086
11	23.000	23.000	18.000	23.000	23.000	23.000	22.000	23.000	23.000	23.000	1.000	0.363	0.430	0.242	-0.203
12	47.000	47.000	38.000	37.000	47.000	47.000	22.000	40.000	34.000	23.000	23.000	1.000	0.845	0.824	-0.021
13	47.000	47.000	38.000	37.000	47.000	47.000	22.000	40.000	34.000	23.000	23.000	47.000	1.000	0.721	0.112
14	33.000	33.000	28.000	33.000	33.000	33.000	22.000	33.000	33.000	23.000	23.000	33.000	33.000	1.000	0.098
15	22.000	22.000	17.000	22.000	22.000	22.000	21.000	22.000	22.000	22.000	22.000	22.000	22.000	22.000	1.000

105 CROSSCR, MEAN= 0.333

Table 18. ρ_a, CORRELATIONS BETWEEN Q_a-VALUES, AND OVERLAPPING RECORD LENGTHS, 15 SITES

TENNESSEE

COR MATRIX R(I,J)= COR OF I AND J IF I<J AND=# OF YEARS IN COMMON FOR I>J

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000	0.458	0.012	0.353	0.321	0.108	0.234	0.062	0.324	0.093	0.232	-0.274	0.410	0.206	-0.190
2	20.000	1.000	0.292	0.529	0.255	0.376	0.264	0.409	0.354	0.342	0.428	0.002	0.606	0.508	0.177
3	19.000	19.000	1.000	0.443	0.443	0.573	0.099	0.429	0.224	0.020	0.094	0.039	0.025	-0.442	-0.033
4	20.000	20.000	24.000	1.000	0.560	0.708	0.789	0.471	0.789	0.505	0.514	-0.049	0.604	-0.131	0.031
5	20.000	20.000	20.000	20.000	1.000	0.806	0.290	0.211	0.402	0.219	-0.036	-0.199	0.110	-0.096	-0.210
6	20.000	20.000	20.000	20.000	21.000	1.000	0.402	0.576	0.588	0.460	0.240	0.106	0.337	0.002	0.055
7	16.000	16.000	25.000	24.000	17.000	17.000	1.000	0.091	0.575	0.615	0.148	0.218	0.485	0.304	0.216
8	17.000	17.000	16.000	17.000	17.000	17.000	13.000	1.000	0.563	0.462	0.554	0.047	0.789	0.224	0.266
9	20.000	20.000	22.000	20.000	21.000	21.000	19.000	17.000	1.000	0.496	0.484	-0.262	0.590	-0.111	-0.015
10	20.000	20.000	20.000	20.000	21.000	21.000	17.000	17.000	21.000	1.000	0.300	-0.066	0.608	0.505	0.200
11	20.000	20.000	19.000	20.000	20.000	20.000	16.000	17.000	20.000	20.000	1.000	-0.101	0.744	-0.118	-0.079
12	16.000	16.000	17.000	16.000	17.000	17.000	13.000	13.000	13.000	17.000	16.000	1.000	0.051	0.434	0.478
13	21.000	20.000	20.000	20.000	21.000	21.000	17.000	17.000	21.000	21.000	20.000	17.000	1.000	0.316	0.095
14	14.000	14.000	23.000	22.000	15.000	15.000	28.000	11.000	17.000	15.000	14.000	16.000	15.000	1.000	0.558
15	20.000	20.000	20.000	20.000	21.000	21.000	17.000	17.000	21.000	21.000	20.000	17.000	21.000	15.000	1.000

105 CROSSCR, MEAN= 0.273

Table 19. ρ_a, CORRELATIONS BETWEEN Q_a-VALUES, AND OVERLAPPING RECORD LENGTHS, 15 SITES

UTAH

COR MATRIX R(I,J)= COR OF I AND J IF I<J AND=# OF YEARS IN COMMON FOR I>J

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000	0.223	0.226	0.753	0.911	[illegible]	-0.060	0.103	0.132	-0.057	0.462	0.203	0.697	0.344	[illegible]
2	18.000	1.000	0.696	0.338	0.901	[illegible]	0.204	0.813	0.191	0.422	0.458	0.494	0.745	0.784	0.668
3	18.000	24.000	1.000	0.664	0.776	0.378	-0.068	0.834	0.107	0.675	0.417	0.706	0.543	0.564	0.481
4	16.000	18.000	18.000	1.000	0.938	0.597	0.121	0.663	0.186	0.287	0.739	0.557	0.814	0.835	0.828
5	16.000	18.000	18.000	18.000	1.000	0.761	0.208	0.324	0.165	0.220	0.797	0.554	0.760	0.793	0.654
6	18.000	24.000	24.000	18.000	18.000	1.000	0.456	0.467	-0.094	0.015	0.721	0.395	0.761	0.748	0.698
7	18.000	24.000	24.000	18.000	18.000	24.000	1.000	-0.102	0.125	-0.186	0.150	-0.311	0.138	0.161	0.158
8	17.000	19.000	19.000	17.000	17.000	19.000	19.000	1.000	0.023	0.497	0.608	0.629	0.518	0.615	0.478
9	19.000	17.000	17.000	17.000	17.000	17.000	17.000	16.000	1.000	-0.094	-0.049	-0.265	-0.101	-0.070	-0.065
10	20.000	16.000	16.000	15.000	15.000	16.000	16.000	15.000	18.000	1.000	0.232	0.404	0.265	0.306	0.351
11	19.000	26.000	22.000	18.000	18.000	22.000	22.000	17.000	20.000	18.000	1.000	0.522	0.799	0.922	0.719
12	20.000	18.000	18.000	18.000	18.000	18.000	18.000	17.000	20.000	19.000	21.000	1.000	0.626	0.613	0.562
13	17.000	18.000	18.000	18.000	18.000	18.000	18.000	17.000	18.000	16.000	19.000	19.000	1.000	0.954	0.879
14	21.000	19.000	19.000	17.000	17.000	19.000	19.000	18.000	19.000	19.000	20.000	21.000	18.000	1.000	0.932
15	17.000	18.000	18.000	18.000	18.000	19.000	18.000	17.000	18.000	16.000	19.000	19.000	19.000	18.000	1.000

105 CRUSSCR, MEAN= 0.449

Table 20. ρ_a, CORRELATIONS BETWEEN Q_a-VALUES, AND OVERLAPPING RECORD LENGTHS, 15 SITES

VIRGINIA

COR MATRIX R(I,J)= COR OF I AND J IF I<J AND=# OF YEARS IN COMMON FOR I>J

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000	0.674	-0.072	0.115	-0.103	-0.033	0.022	-0.121	0.953	0.314	0.050	0.057	0.213	0.147	-0.005
2	25.000	1.000	0.128	0.682	0.152	0.333	0.224	0.318	0.611	0.458	-0.006	0.151	0.185	0.179	-0.233
3	27.000	25.000	1.000	0.235	0.714	-0.155	0.764	0.197	-0.129	0.533	0.126	-0.013	0.382	-0.126	-0.139
4	24.000	24.000	24.000	1.000	0.448	0.207	0.393	0.713	0.105	0.268	0.109	-0.166	0.307	-0.084	-0.267
5	27.000	25.000	27.000	24.000	1.000	0.019	0.500	0.378	-0.146	0.334	0.113	0.085	0.131	-0.032	-0.338
6	24.000	24.000	24.000	24.000	24.000	1.000	-0.187	0.622	0.058	-0.087	0.124	0.688	0.018	-0.122	0.048
7	25.000	25.000	25.000	24.000	25.000	24.000	1.000	0.204	-0.019	0.552	0.232	0.031	0.134	0.100	-0.387
8	25.000	25.000	25.000	24.000	26.000	24.000	25.000	1.000	-0.045	0.067	0.300	0.079	0.240	-0.160	-0.146
9	25.000	25.000	25.000	24.000	25.000	24.000	25.000	25.000	1.000	0.283	0.224	-0.049	0.115	0.386	0.194
10	27.000	25.000	27.000	24.000	30.000	24.000	25.000	26.000	25.000	1.000	0.384	0.019	0.253	0.312	-0.169
11	27.000	25.000	27.000	24.000	29.000	24.000	25.000	26.000	25.000	29.000	1.000	0.019	0.240	0.444	-0.066
12	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	1.000	-0.154	-0.102	-0.020
13	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	24.000	1.000	-0.026	0.119
14	27.000	25.000	27.000	24.000	29.000	24.000	25.000	26.000	25.000	29.000	29.000	24.000	24.000	1.000	-0.139
15	24.000	23.000	24.000	23.000	26.000	23.000	23.000	24.000	23.000	26.000	26.000	23.000	23.000	26.000	1.000

105. CROSSCR, MEAN= 0.151

Table 21. ρ_a, CORRELATIONS BETWEEN Q_a-VALUES, AND OVERLAPPING RECORD LENGTHS, 15 SITES

WYOMING

COR MATRIX R(I,J)= COR OF I AND J IF I<J AND=# OF YEARS IN COMMON FOR I>J

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000	0.784	0.759	0.060	0.787	0.802	-0.019	0.805	0.018	-0.102	-0.110	0.715	-0.241	0.046	-0.101
2	17.000	1.000	0.441	0.074	0.589	0.232	0.028	0.846	-0.013	-0.201	0.218	0.387	-0.022	-0.213	-0.065
3	17.000	20.000	1.000	-0.063	0.787	0.533	0.156	0.804	0.139	-0.073	0.072	0.667	-0.135	0.435	0.039
4	16.000	26.000	19.000	1.000	0.127	0.036	-0.066	-0.055	0.295	-0.036	0.031	0.256	0.172	-0.145	0.247
5	14.000	24.000	17.000	22.000	1.000	0.517	-0.232	0.708	0.268	-0.140	0.212	0.845	0.173	-0.030	0.039
6	17.000	22.000	20.000	21.000	19.000	1.000	0.109	0.917	0.047	-0.118	0.301	0.364	-0.053	-0.055	-0.072
7	14.000	14.000	14.000	13.000	13.000	14.000	1.000	0.270	-0.223	-0.130	-0.239	-0.437	-0.128	0.669	-0.171
8	13.000	13.000	13.000	12.000	12.000	13.000	13.000	1.000	0.089	-0.148	0.288	0.494	-0.176	0.157	-0.060
9	15.000	28.000	18.000	25.000	23.000	20.000	12.000	11.000	1.000	0.419	0.638	0.467	0.504	-0.150	0.531
10	15.000	26.000	18.000	24.000	21.000	20.000	12.000	11.000	31.000	1.000	0.146	0.113	0.153	0.057	-0.051
11	15.000	21.000	18.000	20.000	18.000	20.000	12.000	11.000	21.000	21.000	1.000	0.205	0.440	-0.354	0.152
12	15.000	28.000	18.000	25.000	23.000	20.000	12.000	11.000	32.000	30.000	21.000	1.000	0.199	-0.189	0.053
13	17.000	30.000	20.000	26.000	24.000	22.000	14.000	13.000	32.000	30.000	23.000	32.000	1.000	-0.078	0.380
14	15.000	15.000	15.000	14.000	13.000	15.000	14.000	13.000	13.000	13.000	13.000	13.000	15.000	1.000	0.149
15	17.000	20.000	20.000	19.000	17.000	20.000	14.000	13.000	18.000	18.000	18.000	18.000	20.000	15.000	1.000

105 CROSSCR, MEAN= 0.184

The use of 15 sites per State (or region) to identify the regional correlation does not imply a further limitation on the number of sites per region (or regression). It suggests only that few sites have sufficiently great overlap among their records to calculate reasonably stable correlations for inclusion in the averaging process in the region. The resulting sample size, 15 x 14/2 = 105, is adequate for defining the mean, particularly given the coarse grid of ρ in the BIGBASIN tables.

Using a two-parameter log-normal density at each site, 3,500 years of synthetic floods are generated, consisting of 100 ten-year sequences and 100 twenty-five-year sequences for testing the sensitivity of the regional correlations ρ_{50} to record length. The symbol adopted for the regional correlation coefficient among the estimates $\hat{Q}_{50}$ is ρ_{50}, and there is one value for each State or region. Tables 22 through 32 are matrices of pair-wise correlations ρ_{50} based on 100 replications of ten years' duration and 100 replications of 25 years' duration. Because of the symmetry of the correlation matrices, the top half is devoted to ten-year values and the bottom half to 25-year values. Average values are taken for each State. Using the same regional equivalencies as in the economic analysis to extrapolate from the 11 analyses to the nation, we obtain the State-wide estimates of ρ_a and two estimates of ρ_{50} shown in Table 33. The two synthetic record lengths, ten and twenty-five years, give essentially identifical results (Figure 35).

Estimates of the regional skew coefficient are made by averaging technique which utilizes the WORLDWAR I tables. The skew coefficient is calculated from the coefficient of variation for the log-normal density; this in turn is calculated from unbiased estimates of the mean and standard deviation of the distribution of Q_{50}. The USGS recently produced tables of the unbiased estimates of the coefficient of variation; these were not available during our calculations, but it is recommended that further extension of this work be based on utilization of these important values. As indicated above, the error

Table 22. ρ_{50}, CORRELATIONS BETWEEN $\hat{Q}_{50}$-VALUES (10-YEAR ESTIMATES) AND $\hat{Q}_{50}$-VALUES (25-YEAR ESTIMATES) BASED ON SYNTHETIC ANNUAL FLOODS, GEORGIA

COR. MATRIX FOR Q50 I<J = 10 YEAR SETS, I>J = 25 YEAR SETS

Top Diagonal
10 YEAR ESTIMATES (MEAN=.130)

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.0	0.093	0.174	0.074	0.437	0.296	0.374	0.282	0.190	-0.095	-0.217	-0.152	0.347	-0.035	0.276
2	0.315	0.0	0.534	0.131	0.814	0.027	-0.052	-0.156	-0.055	-0.180	-0.007	-0.182	0.213	-0.019	0.001
3	0.255	0.355	0.0	0.289	0.352	0.154	0.021	0.121	-0.023	0.629	0.031	-0.048	-0.052	-0.048	0.217
4	-0.089	0.078	0.080	0.0	0.398	0.213	0.057	0.137	0.682	0.405	-0.107	0.082	-0.064	-0.187	0.233
5	0.406	0.834	0.292	0.006	0.0	-0.048	0.298	-0.080	-0.103	0.339	-0.203	0.059	0.209	0.063	0.536
6	0.233	0.042	0.181	-0.111	0.006	0.0	0.108	-0.040	0.024	-0.170	-0.079	-0.176	0.222	0.354	-0.074
7	-0.112	-0.061	0.155	-0.167	0.035	0.150	0.0	0.317	0.346	-0.022	-0.031	0.140	-0.109	0.048	0.038
8	0.375	0.005	0.525	0.383	-0.163	0.168	0.414	0.0	0.151	-0.106	0.404	0.271	0.294	0.335	0.203
9	0.306	-0.255	-0.081	-0.029	-0.037	0.146	-0.030	0.003	0.0	0.508	0.343	-0.046	0.117	0.181	0.219
10	0.068	-0.308	0.473	0.397	0.263	0.141	-0.044	0.023	0.521	0.0	0.021	-0.004	-0.032	0.225	0.165
11	0.032	0.002	-0.143	0.370	-0.248	-0.125	0.076	0.041	0.059	-0.241	0.0	0.610	0.071	-0.050	0.279
12	-0.188	0.031	-0.067	-0.076	-0.032	-0.072	0.334	0.187	0.047	-0.091	0.557	0.0	-0.016	0.138	0.305
13	0.456	0.128	0.039	-0.149	0.079	0.095	-0.012	0.259	-0.166	0.138	0.337	-0.155	0.0	0.030	0.353
14	0.232	0.115	0.521	0.129	0.214	0.201	0.312	0.150	0.392	0.172	-0.032	0.334	0.085	0.0	0.119
15	0.427	-0.096	0.123	0.328	0.146	0.065	0.122	0.133	0.220	0.042	-0.107	0.031	0.105	0.081	0.0

Lower Diagonal
25 YEAR ESTIMATES (MEAN = .115)

Table 23. ρ_{50}, CORRELATIONS BETWEEN $\hat{Q}_{50}$-VALUES (10-YEAR ESTIMATES) AND $\hat{Q}_{50}$-VALUES (25-YEAR ESTIMATES) BASED ON SYNTHETIC ANNUAL FLOODS, MASSACHUSETTS

Top Diagonal
10 YEAR ESTIMATES (MEAN = .235)

COR. MATRIX FOR Q50 I<J = 10 YEAR SETS, I>J = 25 YEAR SETS

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.0	0.937	0.211	0.308	0.103	0.245	0.271	0.462	0.175	0.007	-0.299	-0.140	0.055	0.034	0.223
2	0.887	0.0	0.656	0.668	0.339	0.160	-0.166	0.623	0.007	-0.203	-0.195	-0.054	0.188	-0.005	0.125
3	0.229	0.477	0.0	0.690	0.027	0.179	0.141	0.055	0.065	0.297	0.212	-0.066	-0.140	0.004	0.235
4	0.367	0.250	0.731	0.0	0.334	0.063	0.025	0.042	0.641	-0.024	0.126	-0.019	-0.020	-0.211	0.574
5	0.524	0.311	-0.026	0.205	0.0	0.614	0.415	0.393	0.189	0.851	-0.072	0.682	0.871	0.148	0.558
6	0.146	0.041	0.269	0.006	0.615	0.0	0.308	0.282	0.464	0.593	0.121	0.426	0.621	0.458	0.204
7	-0.103	-0.114	0.409	-0.091	0.115	0.620	0.0	0.418	0.637	0.093	0.148	-0.026	-0.023	0.093	0.137
8	0.453	0.172	0.326	0.493	0.091	0.266	0.500	0.0	0.151	0.256	0.108	0.112	0.089	0.129	-0.006
9	0.283	-0.214	-0.019	-0.050	0.233	0.526	0.343	-0.006	0.0	0.399	0.425	-0.121	0.091	-0.009	0.352
10	0.090	-0.288	0.323	0.345	0.834	0.524	0.161	0.367	0.443	0.0	0.110	0.188	0.227	0.115	0.239
11	-0.230	-0.065	0.323	0.142	-0.199	0.265	0.171	0.165	0.400	0.577	0.0	-0.049	0.164	-0.007	0.289
12	-0.188	0.120	-0.073	-0.076	0.491	0.480	0.219	-0.089	0.050	0.517	0.080	0.0	0.928	0.646	0.441
13	-0.251	0.078	0.072	-0.112	0.657	0.546	0.185	0.115	-0.040	0.474	0.125	0.347	0.0	0.920	0.354
14	0.195	0.150	-0.000	0.092	0.537	0.468	0.444	-0.014	0.164	0.390	-0.188	0.643	0.867	0.0	0.604
15	0.156	-0.007	0.151	0.436	0.322	0.423	0.188	-0.003	0.316	0.185	-0.044	0.395	0.153	0.508	0.0

Lower Diagonal
25 YEAR ESTIMATES (MEAN = .231)

Table 24. ρ_{50}, CORRELATIONS BETWEEN $\hat{Q}_{50}$-VALUES (10-YEAR ESTIMATES) AND $\hat{Q}_{50}$-VALUES (25-YEAR ESTIMATES) BASED ON SYNTHETIC ANNUAL FLOODS, MISSOURI

Top Diagonal
10 YEAR ESTIMATES (MEAN = .090)

COR. MATRIX FOR Q50   I<J = 10 YEAR SETS,   I>J = 25 YEAR SETS

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.0	-0.034	0.437	0.322	-0.057	0.301	0.135	0.274	0.141	0.395	-0.138	-0.043	-0.004	-0.040	0.058
2	0.179	0.0	0.273	0.373	0.288	0.250	-0.222	-0.045	-0.105	-0.216	0.070	-0.151	0.150	-0.022	0.055
3	0.170	0.384	0.0	0.467	0.176	0.089	0.069	0.099	-0.096	-0.128	0.110	0.042	-0.020	-0.035	0.210
4	-0.051	0.021	0.293	0.0	0.498	0.089	-0.002	0.020	0.191	-0.094	-0.007	-0.074	-0.063	-0.170	0.237
5	0.658	0.432	-0.090	0.167	0.0	0.363	0.461	-0.175	-0.137	0.278	-0.054	-0.013	0.164	0.145	0.457
6	0.217	0.091	0.092	-0.110	0.092	0.0	0.133	0.252	-0.017	-0.208	-0.172	0.065	0.202	0.160	0.179
7	0.193	-0.145	-0.023	-0.142	-0.036	0.179	0.0	0.468	0.105	0.372	0.143	-0.013	-0.056	0.157	-0.077
8	0.498	0.025	0.295	0.326	-0.165	0.199	0.363	0.0	0.025	-0.125	0.114	0.072	0.098	0.036	0.176
9	0.287	-0.196	-0.023	-0.088	0.021	0.048	0.140	-0.045	0.0	-0.048	0.259	-0.082	0.107	-0.045	0.249
10	0.238	-0.243	0.058	0.055	-0.022	-0.029	0.372	-0.048	0.023	0.0	0.330	-0.075	-0.047	-0.022	0.221
11	0.163	0.060	0.092	0.227	-0.080	-0.019	0.205	-0.011	0.017	0.054	0.0	0.509	0.082	-0.078	0.306
12	-0.057	-0.002	-0.030	-0.035	-0.035	0.117	0.190	-0.081	0.060	-0.030	0.531	0.0	0.205	0.109	0.226
13	-0.229	0.179	0.055	-0.075	-0.024	0.127	0.046	0.128	-0.189	0.124	0.389	-0.150	0.0	-0.110	0.073
14	0.139	0.152	-0.050	0.090	0.131	0.145	0.104	0.052	0.123	0.059	-0.100	-0.074	-0.018	0.0	0.081
15	0.001	0.032	0.038	0.388	0.001	0.402	-0.043	0.196	0.240	0.147	0.035	0.023	0.031	0.049	0.0

Lower Diagonal
25 YEAR ESTIMATES (MEAN = .083)

Table 25 ρ_{50}, CORRELATIONS BETWEEN $\hat{Q}_{50}$-VALUES (10 YEAR ESTIMATES) AND $\hat{Q}_{50}$-VALUES (25-YEAR ESTIMATES) BASED ON SYNTHETIC ANNUAL FLOODS, MONTANA

Top Diagonal
10 YEAR ESTIMATES (MEAN = .098)

COR. MATRIX FOR Q50   I<J = 10 YEAR SETS,   I>J = 25 YEAR SETS

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.0	0.251	0.306	0.292	0.029	0.266	0.009	0.290	0.608	0.096	-0.209	-0.033	0.008	0.033	0.092
2	0.825	0.0	0.459	0.130	0.393	-0.048	-0.279	0.023	0.365	-0.069	0.097	-0.055	0.065	-0.113	0.057
3	0.220	0.406	0.0	0.284	-0.024	0.104	0.107	0.223	0.082	0.072	0.098	0.202	-0.175	-0.025	0.198
4	-0.082	0.194	0.109	0.0	0.692	0.320	-0.158	0.208	0.402	0.081	0.166	-0.061	-0.101	-0.164	0.234
5	0.533	0.037	-0.206	0.350	0.0	-0.081	0.481	-0.063	-0.083	0.086	-0.139	0.041	0.391	0.021	0.510
6	0.123	0.078	0.143	-0.045	0.271	0.0	0.092	-0.148	-0.094	0.360	-0.019	0.314	0.031	0.171	-0.107
7	0.062	-0.366	0.220	-0.144	-0.106	0.404	0.0	0.128	0.011	0.293	-0.100	0.135	-0.115	0.336	0.182
8	0.118	-0.016	0.268	0.351	0.034	0.197	0.232	0.0	-0.107	-0.074	0.107	0.230	0.150	0.177	-0.173
9	0.506	0.493	0.025	0.005	-0.051	0.084	0.071	-0.066	0.0	0.242	0.092	-0.103	0.083	-0.074	0.125
10	-0.104	-0.135	0.043	0.134	0.270	0.553	0.152	-0.062	-0.015	0.0	-0.013	0.037	0.011	-0.131	-0.141
11	-0.012	0.177	-0.014	0.124	-0.137	0.173	-0.189	0.033	0.012	-0.384	0.0	0.199	0.097	-0.095	0.147
12	-0.027	0.111	0.003	-0.166	-0.057	0.176	0.150	-0.048	0.110	0.078	0.249	0.0	0.124	0.045	0.194
13	-0.257	-0.001	0.023	-0.126	0.139	0.050	0.032	0.001	-0.157	0.028	0.337	0.001	0.0	0.173	0.071
14	0.212	0.089	0.489	0.068	0.145	0.185	0.241	0.012	0.091	0.101	-0.132	-0.151	0.146	0.0	0.553
15	-0.151	-0.163	0.097	0.245	0.212	0.279	0.432	0.044	0.125	-0.219	-0.142	0.178	-0.037	0.471	0.0

Lower Diagonal
25 YEAR ESTIMATES (MEAN = .089)

Table 26. ρ_{50}, CORRELATIONS BETWEEN $\hat{Q}_{50}$-VALUES (10 YEAR ESTIMATES) AND $\hat{Q}_{50}$-VALUES (25-YEAR ESTIMATES) BASED ON SYNTHETIC ANNUAL FLOODS, NEW MEXICO

COR. MATRIX FOR Q50 I<J = 10 YEAR SETS, I>J = 25 YEAR SETS

Top Diagonal
10 YEAR ESTIMATES (MEAN=.109)

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.0	0.103	0.218	0.324	0.422	0.027	0.316	0.232	0.291	-0.136	0.268	0.106	-0.038	0.186	0.292
2	0.381	0.0	0.705	0.611	0.673	-0.180	-0.192	-0.127	-0.064	-0.201	0.005	-0.065	0.327	-0.000	-0.017
3	0.121	0.627	0.0	0.443	0.615	0.008	0.005	0.045	-0.093	-0.182	0.012	-0.082	0.003	-0.062	0.245
4	0.462	0.147	0.516	0.0	0.799	0.083	-0.021	0.092	0.236	-0.019	0.448	0.111	-0.017	-0.095	0.515
5	0.506	0.530	0.547	0.763	0.0	0.134	0.093	-0.190	-0.078	0.649	0.350	0.325	0.199	0.127	0.350
6	0.320	0.110	-0.006	0.283	-0.083	0.0	-0.132	-0.098	-0.063	-0.077	0.517	0.076	-0.018	0.244	-0.115
7	0.356	-0.188	0.072	-0.307	-0.086	-0.199	0.0	0.181	0.050	0.015	-0.156	0.239	-0.056	0.241	-0.193
8	0.595	0.019	0.088	0.238	-0.157	0.006	0.310	0.0	-0.096	-0.029	0.195	0.051	0.311	0.341	-0.114
9	0.275	-0.187	-0.108	0.202	0.053	0.016	-0.028	-0.098	0.0	-0.144	0.325	-0.067	0.241	-0.004	0.178
10	0.170	-0.266	0.204	0.151	0.186	0.352	-0.013	-0.155	0.071	0.0	-0.222	-0.124	-0.036	-0.056	0.026
11	0.436	0.059	0.025	0.689	0.318	0.256	-0.148	-0.099	0.455	-0.095	0.0	0.134	0.024	-0.090	0.251
12	-0.049	0.157	-0.040	0.001	0.019	0.238	0.201	-0.092	0.022	0.186	-0.018	0.0	0.083	0.167	0.238
13	-0.265	0.237	0.143	0.024	0.099	0.040	0.097	0.006	-0.227	0.248	0.344	-0.178	0.0	0.025	0.068
14	0.293	0.065	0.042	0.109	0.077	0.057	0.166	0.006	0.121	0.140	0.087	-0.004	-0.058	0.0	-0.005
15	0.097	-0.107	0.200	0.487	0.012	0.349	0.143	0.045	0.122	-0.123	-0.111	0.431	-0.013	0.104	0.0

Lower Diagonal
25 YEAR ESTIMATES (MEAN=.120)

Table 27. ρ_{50}, CORRELATIONS BETWEEN $\hat{Q}_{50}$-VALUES (10 YEAR ESTIMATES) AND $\hat{Q}_{50}$-VALUES (25 YEAR ESTIMATES) BASED ON SYNTHETIC ANNUAL FLOODS, OHIO

Top Diagonal
10 YEAR ESTIMATES (MEAN= .131)

COR. MATRIX FOR Q50   I<J = 10 YEAR SETS, I>J = 25 YEAR SETS

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.0	0.181	0.220	0.237	0.051	0.130	0.317	0.319	0.225	0.093	-0.100	-0.136	0.148	0.007	0.089
2	0.399	0.0	0.091	0.123	0.776	-0.161	-0.009	-0.152	0.017	-0.195	-0.019	-0.136	0.144	-0.051	-0.061
3	0.198	0.165	0.0	0.284	0.418	0.253	0.055	0.129	-0.145	-0.015	0.282	-0.011	-0.128	-0.101	0.272
4	-0.033	-0.014	0.238	0.0	0.665	0.093	-0.056	0.034	0.046	0.033	0.062	0.343	0.040	-0.205	0.108
5	0.514	0.120	0.429	0.546	0.0	0.452	0.198	0.013	-0.062	0.183	-0.158	0.170	0.660	0.074	0.410
6	0.181	-0.014	0.092	0.119	0.426	0.0	-0.112	-0.152	-0.106	-0.041	-0.041	0.306	0.062	-0.029	-0.215
7	0.072	-0.032	0.246	-0.068	0.020	-0.011	0.0	0.431	0.171	0.003	0.323	-0.062	0.112	0.119	0.761
8	0.325	0.012	0.264	0.306	-0.135	0.031	0.649	0.0	0.367	0.237	0.607	0.184	0.161	0.253	0.460
9	0.124	-0.183	-0.112	-0.001	0.034	-0.075	-0.054	0.083	0.0	0.459	0.202	-0.142	0.269	-0.058	0.082
10	0.110	-0.303	0.199	-0.030	0.071	0.126	0.082	0.405	0.441	0.0	0.311	-0.019	0.027	0.005	0.148
11	0.068	0.023	-0.055	0.068	-0.255	0.159	0.250	0.028	0.211	0.179	0.0	0.212	0.111	0.193	0.354
12	-0.122	0.127	-0.079	0.507	0.265	0.269	-0.064	-0.034	0.090	-0.105	-0.012	0.0	0.169	0.090	0.181
13	-0.048	0.062	0.021	0.130	0.124	0.063	0.423	0.365	-0.207	0.103	0.497	-0.100	0.0	0.456	0.297
14	0.251	0.218	0.425	0.266	0.297	0.302	0.504	0.083	0.070	0.039	0.059	-0.037	0.257	0.0	0.033
15	0.061	-0.128	0.206	0.152	0.017	0.121	0.769	0.312	0.021	-0.012	-0.009	0.135	0.154	0.060	0.0

Lower Diagonal
25 YEAR ESTIMATES (MEAN= .129)

Table 28. ρ_{50}, CORRELATION BETWEEN $\hat{Q}_{50}$-VALUES (10 YEAR ESTIMATES) AND $\hat{Q}_{50}$-VALUES (25 YEAR ESTIMATES) BASED ON SYNTHETIC ANNUAL FLOODS, OREGON

Top Diagonal
10 YEAR ESTIMATES (MEAN=.182)

COR. MATRIX FOR Q50 I<J = 10 YEAR SETS, I>J = 25 YEAR SETS

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.0	-0.079	0.187	0.127	-0.019	0.238	0.255	0.273	0.197	0.098	-0.211	-0.111	0.077	0.040	0.094
2	0.035	0.0	-0.035	0.533	0.603	0.062	0.037	-0.102	-0.068	-0.220	0.129	-0.126	0.259	-0.020	0.043
3	0.141	0.093	0.0	0.294	0.121	0.213	0.178	0.110	-0.085	0.267	0.232	0.320	0.120	-0.117	0.056
4	0.188	0.131	0.219	0.0	0.680	0.010	0.167	0.155	0.696	0.355	0.010	0.435	0.496	0.026	0.334
5	0.607	0.334	0.124	0.498	0.0	0.035	0.124	-0.133	-0.102	0.255	-0.229	0.126	0.500	0.030	0.497
6	0.122	-0.018	0.500	0.122	0.023	0.0	0.164	0.191	0.001	-0.048	-0.173	0.391	0.507	0.664	-0.050
7	0.045	-0.019	0.417	0.112	-0.043	0.420	0.0	0.575	0.252	-0.036	0.095	0.188	0.140	0.105	0.027
8	0.309	0.023	0.357	0.525	-0.152	0.227	0.622	0.0	0.280	-0.037	0.396	0.344	0.137	0.247	0.223
9	0.302	-0.233	-0.106	0.176	0.083	0.163	-0.059	0.084	0.0	0.376	0.299	0.368	0.082	0.329	0.116
10	0.130	-0.275	0.319	0.342	0.163	0.360	0.033	0.146	0.428	0.0	-0.063	0.346	0.294	0.672	0.126
11	-0.058	-0.015	-0.055	0.174	-0.262	0.076	0.060	0.041	0.232	-0.213	0.0	0.249	0.251	0.050	0.184
12	-0.151	0.192	0.129	0.628	0.226	0.424	0.390	0.143	0.367	0.297	0.230	0.0	0.871	0.631	0.186
13	-0.108	0.118	0.105	0.349	0.179	0.389	0.514	0.431	0.248	0.761	0.625	0.310	0.0	0.643	0.100
14	0.192	0.249	0.405	0.647	0.168	0.418	0.452	0.141	0.497	0.338	-0.078	0.594	0.580	0.0	-0.037
15	0.068	-0.134	0.053	0.314	-0.024	0.011	0.086	0.143	0.127	-0.038	-0.071	0.013	0.025	-0.008	0.0

Lower Diagonal
25 YEAR ESTIMATES (MEAN=.190)

Table 29. ρ_{50}, CORRELATION BETWEEN $\hat{Q}_{50}$-VALUES (10 YEAR ESTIMATES) AND $\hat{Q}_{50}$-VALUES (25 YEAR ESTIMATES) BASED ON SYNTHETIC ANNUAL FLOODS, TENNESSEE

Top Diagonal
10 YEAR ESTIMATES (MEAN=.165)

COR. MATRIX FOR Q50 I<J = 10 YEAR SETS, I>J = 25 YEAR SETS

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.0	0.209	0.208	0.178	0.177	0.258	0.294	0.193	0.295	-0.003	-0.112	-0.091	0.196	-0.007	0.112
2	0.378	0.0	0.276	0.557	0.304	0.372	0.045	0.257	0.177	-0.124	0.148	-0.180	0.594	0.020	0.029
3	0.144	0.254	0.0	0.357	0.280	0.217	-0.088	0.235	-0.022	0.012	0.090	-0.050	-0.174	0.432	0.134
4	0.185	0.172	0.329	0.0	0.222	0.295	0.557	0.246	0.700	0.373	0.339	-0.033	0.358	-0.130	0.301
5	0.598	0.234	0.165	0.556	0.0	0.490	0.114	0.072	0.343	0.063	-0.323	0.210	0.180	0.139	-0.001
6	0.169	0.135	0.284	0.269	0.621	0.0	0.084	0.407	0.227	-0.079	-0.019	-0.098	0.207	0.010	-0.085
7	-0.026	0.004	-0.050	0.537	0.044	0.239	0.0	0.179	0.434	0.217	-0.031	0.013	0.004	0.099	0.014
8	0.222	0.111	0.337	0.528	0.046	0.218	0.195	0.0	0.340	0.063	0.588	0.103	0.580	0.118	0.110
9	0.332	-0.063	-0.044	0.523	0.289	0.364	0.217	0.146	0.0	0.070	0.347	-0.106	0.110	-0.032	0.097
10	0.090	-0.214	0.106	0.334	0.273	0.313	0.330	0.207	0.238	0.0	-0.015	-0.118	0.261	0.315	0.152
11	0.062	0.190	-0.094	0.197	-0.174	-0.075	-0.078	0.254	0.374	-0.056	0.0	0.232	0.515	0.010	0.241
12	-0.134	0.157	-0.039	-0.153	0.072	-0.059	0.094	-0.019	0.046	-0.078	-0.073	0.0	0.192	0.303	0.344
13	0.001	0.429	-0.051	0.230	0.107	0.135	0.403	0.712	0.349	0.670	0.818	-0.208	0.0	0.081	0.107
14	0.171	0.297	-0.002	0.043	0.046	0.085	0.366	0.047	0.064	0.178	0.037	0.315	0.046	0.0	0.331
15	-0.137	-0.159	0.049	0.269	-0.026	0.131	0.075	0.054	0.196	0.021	-0.098	0.259	0.010	0.257	0.0

Lower Diagonal
25 YEAR ESTIMATES (MEAN = .160)

Table 30. ρ_{50}, CORRELATION BETWEEN $\hat{Q}_{50}$-VALUES (10 YEAR ESTIMATES) AND $\hat{Q}_{50}$-VALUES (25 YEAR ESTIMATES) BASED ON SYNTHETIC ANNUAL FLOODS, UTAH

COR. MATRIX FOR Q50 I<J = 10 YEAR SETS, I>J = 25 YEAR SETS

Top Diagonal 10 YEAR ESTIMATES (MEAN=.314)

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.0	-0.097	0.156	0.402	0.555	0.297	0.396	0.234	0.289	-0.071	-0.044	-0.131	0.324	0.020	0.536
2	0.104	0.0	0.807	0.847	0.858	0.497	-0.102	0.879	0.202	-0.170	0.163	0.160	0.709	0.206	0.151
3	0.238	0.627	0.0	0.458	0.727	0.196	-0.057	0.469	-0.095	0.761	0.389	0.577	0.381	0.007	0.341
4	0.601	0.601	0.516	0.0	0.728	0.255	-0.061	0.261	0.521	0.144	0.599	0.132	0.719	0.453	0.906
5	0.816	0.870	0.688	0.754	0.0	0.570	0.213	0.727	-0.061	0.246	0.627	0.162	0.679	0.623	0.595
6	0.281	0.491	0.224	0.334	0.545	0.0	0.194	0.194	0.052	-0.216	0.188	0.096	0.626	0.674	0.683
7	-0.129	-0.041	-0.168	-0.275	-0.009	0.362	0.0	0.112	0.257	-0.009	0.052	0.068	-0.116	0.134	-0.026
8	0.287	0.452	0.799	0.640	0.628	0.279	0.298	0.0	-0.160	0.045	0.625	0.436	0.278	0.411	0.265
9	0.131	-0.133	-0.110	-0.003	0.087	-0.002	-0.072	-0.042	0.0	-0.194	0.419	-0.032	0.256	-0.066	0.099
10	0.083	-0.144	0.593	0.147	0.153	-0.071	-0.031	0.223	0.099	0.0	-0.074	-0.024	-0.004	0.101	0.225
11	0.207	0.227	0.068	0.412	0.654	0.464	-0.045	0.329	0.007	-0.151	0.0	0.364	0.646	0.848	0.524
12	-0.152	0.411	0.494	0.259	0.359	0.154	-0.006	0.336	0.066	0.041	0.351	0.0	0.466	0.413	0.406
13	0.451	0.588	0.185	0.615	0.524	0.607	0.079	0.514	-0.240	0.282	0.800	0.065	0.0	0.900	0.704
14	0.328	0.575	0.580	0.821	0.646	0.625	0.160	0.277	0.065	0.070	0.795	0.371	0.866	0.0	0.889
15	0.563	0.286	0.340	0.807	0.322	0.663	-0.013	0.174	0.089	0.118	0.348	0.386	0.657	0.859	0.0

Lower Diagonal

25 YEAR ESTIMATES (MEAN=.310)

Table 31. ρ_{50}, CORRELATION BETWEEN $\hat{Q}_{50}$-VALUES (10 YEAR ESTIMATES) AND $\hat{Q}_{50}$-VALUES (25 YEAR ESTIMATES) BASED ON SYNTHETIC ANNUAL FLOODS, VIRGINIA .

Top Diagonal
10 YEAR ESTIMATE (MEAN = .093)

COR. MATRIX FOR Q50 I<J = 10 YEAR SETS, I>J = 25 YEAR SETS

VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.0	0.424	0.011	0.191	[illegible]	[illegible]	-0.105	0.248	0.954	-0.059	-0.172	-0.093	0.039	0.105	0.034
2	0.153	0.0	0.169	0.489	0.157	0.146	-0.059	0.016	0.354	-0.142	-0.108	-0.170	0.252	-0.025	0.150
3	0.293	0.166	0.0	0.369	0.706	0.719	0.053	0.301	-0.033	0.508	0.007	0.156	0.059	0.004	0.069
4	0.051	0.416	0.190	0.0	0.050	0.283	0.214	0.353	0.044	0.199	0.160	0.072	0.015	-0.183	0.066
5	-0.023	0.279	0.282	0.271	0.0	0.065	0.140	0.277	-0.155	-0.067	0.042	-0.006	0.071	0.537	-0.085
6	0.035	0.110	-0.043	-0.062	-0.054	0.0	0.090	0.213	-0.108	-0.102	-0.100	0.028	0.102	0.082	-0.155
7	-0.181	-0.197	0.352	-0.051	0.378	-0.073	0.0	0.188	-0.106	0.602	0.008	-0.029	-0.108	0.029	-0.171
8	-0.209	0.054	0.221	0.561	0.168	0.409	0.435	0.0	-0.086	-0.116	0.085	0.161	0.033	0.120	-0.159
9	0.844	0.312	0.122	-0.087	-0.179	-0.010	-0.011	-0.032	0.0	0.235	-0.076	-0.057	0.160	-0.007	0.144
10	-0.125	-0.130	0.428	0.178	0.278	-0.043	0.172	-0.097	0.138	0.0	0.274	-0.005	-0.052	0.166	0.217
11	0.093	-0.037	-0.004	0.338	-0.143	-0.085	-0.007	0.289	0.147	0.123	0.0	0.033	0.139	0.160	0.212
12	-0.066	0.076	-0.137	-0.245	-0.016	0.084	0.106	-0.026	0.028	-0.119	-0.048	0.0	-0.011	0.220	0.069
13	-0.149	0.105	-0.015	0.057	0.060	0.025	0.276	0.196	-0.101	0.326	0.071	-0.242	0.0	-0.067	0.123
14	0.236	0.148	-0.085	0.045	0.092	0.076	-0.125	0.003	0.121	-0.054	-0.029	0.240	-0.065	0.0	-0.088
15	-0.162	-0.034	0.053	0.122	0.017	0.165	0.141	-0.055	0.114	0.037	-0.069	-0.260	0.010	0.015	0.0

Lower Diagonal
25 YEAR ESTIMATE (MEAN = .069)

Table 32. ρ_{50}, CORRELATION BETWEEN $\hat{Q}_{50}$-VALUES (10 YEAR ESTIMATES) AND $\hat{Q}_{50}$-VALUES (25 YEAR ESTIMATES) BASED ON SYNTHETIC ANNUAL FLOODS, WYOMING

Top Diagonal

10 YEAR ESTIMATES (MEAN = .174)

COR. MATRIX FOR Q50   I<J = 10 YEAR SETS, I>J = 25 YEAR SETS

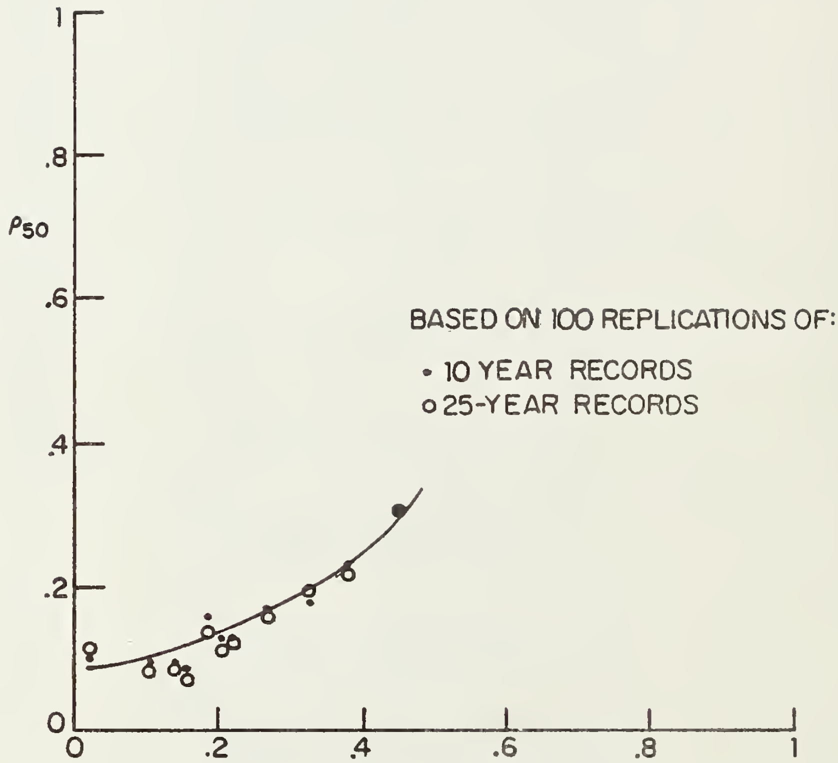
VAR#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.0	0.819	0.602	0.104	0.610	0.659	0.291	0.751	0.171	0.021	-0.249	0.467	0.131	-0.053	0.091
2	0.705	0.0	0.589	0.179	0.687	0.848	-0.187	0.524	-0.062	0.095	-0.068	0.027	0.143	0.051	0.066
3	0.612	0.408	0.0	0.343	0.735	0.188	-0.089	0.312	-0.077	-0.040	0.059	0.572	-0.214	0.005	0.148
4	-0.041	0.082	-0.008	0.0	0.478	0.174	-0.023	0.224	0.610	-0.030	-0.076	-0.032	-0.011	0.021	0.458
5	0.639	0.556	0.541	0.052	0.0	0.401	0.116	0.540	-0.065	0.293	-0.085	0.545	0.183	0.075	0.486
6	0.795	0.730	0.232	-0.121	0.204	0.0	-0.025	0.752	-0.048	-0.129	-0.247	-0.011	0.204	-0.008	-0.109
7	-0.110	-0.061	-0.069	-0.306	0.044	-0.022	0.0	0.207	0.009	-0.053	0.065	0.213	-0.104	0.273	-0.077
8	0.778	0.787	0.819	0.267	0.345	0.766	0.372	0.0	-0.035	-0.083	0.509	0.600	0.218	0.191	-0.139
9	0.282	-0.282	-0.054	-0.013	-0.035	0.044	-0.161	-0.011	0.0	-0.026	0.451	0.194	0.120	-0.014	0.342
10	0.112	-0.206	0.114	0.020	0.099	-0.150	-0.057	-0.157	0.183	0.0	-0.099	-0.138	-0.042	-0.065	0.074
11	-0.051	0.043	-0.121	0.205	-0.170	-0.027	-0.120	0.023	0.559	-0.257	0.0	0.196	0.261	0.181	0.287
12	0.338	0.353	0.447	-0.106	0.795	0.162	0.133	0.169	0.266	-0.102	0.104	0.0	0.176	0.124	0.197
13	-0.204	0.083	0.042	-0.079	0.018	0.119	0.013	0.048	0.243	0.268	0.631	-0.240	0.0	0.076	0.332
14	0.275	0.136	0.418	0.122	0.287	0.052	0.539	0.067	0.049	0.034	0.166	-0.042	-0.018	0.0	-0.020
15	-0.143	-0.127	0.063	0.353	-0.009	0.045	-0.086	-0.067	0.326	-0.058	-0.141	0.009	0.083	0.113	0.0

Lower Diagonal

25 YEAR ESTIMATES (MEAN = .141)

Table 33. Summary of State-Wide Flood Correlations and Skew Coefficients

Region	States	ρa	$\rho_{50}(10)$	$\rho_{50}(25)$	Regional Skew
Georgia	Ala., Ark., Fla., La., Miss., N.C., S.C.	.205	.130	.115	.728 ± .280
Massachusetts	Conn., Me., N.H., N.Y., R.I., Vt.	.378	.235	.231	.985 ± .266
Missouri	Ia., Minn.	.142	.094	.083	.588 ± .220
Montana	Id., N.D.	.102	.098	.089	.793 ± .584
New Mexico	Ariz., Okl., Tex.	.021	.109	.120	.501 ± .234
Ohio	Ill., Ind., Mich., Wis.	.219	.131	.129	.754 ± .236
Oregon	Calif., Wash.	.333	.182	.190	1.156 ± .472
Tennessee	Ky., Pa., W. Va.	.273	.165	.160	1.018 ± .365
Utah	Col., Nev.	.449	.314	.310	.859 ± .349
Virginia	Del., Md., N.J.	.151	.093	.069	.644 ± .186
Wyoming	Kan., Neb., S.D.	.184	.174	.141	1.185 ± .527



ρ_a, Regional Corr. of Annual Flood:

Figure 35. ρ_{50} vs. ρ_a

introduced by the approximation is managably small. Available USGS tables were used to derive close approximations to the unbiased coefficients of variation.

This study utilized two volumes of tables derived from the WORLD-WAR I analysis. These are based on Monte Carlo analyses and give the deviates for unbiased estimates of the mean and standard deviation, respectively, of sample statistics corresponding to various return periods. Use of the tables to derive an unbiased estimate of the mean is explained in an earlier section; their use to estimate the standard deviation is shown here. The ratio closely approximates the unbiased coefficient of variation of Q_{50}. The table for the mean is entered in the usual fashion, from which the unbiased estimate of the expected value of Q_{50} is derived by reading some other flow, say Q_T, from the distribution of annual floods Q_a. If that return period T is then applied to the tables for the standard deviation, a new deviate is read by interpolation. This deviate then defines a new return period (say V) at which the corresponding flow is an unbiased estimate of the standard deviation of Q_{50}.

Thus the sequence is to enter Volume 1 with the desired return period (T = 50), to read an appropriate deviate from the "true" distribution, to apply that deviate to the actual density and thereby to interpolate the return interval T which defines Q_{50} from the distribution of annual floods Q_a, to utilize T to develop a new deviate from Volume 2, and finally to apply that new deviate to determine the return interval V whose corresponding flow (from the density or distribution of Q_a) defines the standard deviation of Q_{50}. The ratio of the standard deviation to the mean defines the coefficient of variation of Q_{50}, from which the coefficient of skewness can be derived.

The regional skew is estimated at the same 15 gages used in the correlation analysis. If all the skew coefficients at the 15 gages covering each of the 11 representative States are averaged, a regional skew coefficient for that State and region can be developed. These regional skew coefficients are given in Table 33, along with the regional coefficients of correlation for annual and 50-year flood events.

The 50-year correlations, derived by 100 replications of sequences of annual floods, appear to be independent of the length of record or trace. There is virtually no difference between the 10- and 25-year results, as shown in Figure 34 and Table 33, so a single relationship between ρ_a and ρ_{50} is used. If one contrasts the 15 individual sites in each State and the average over the 105 or ((15 x 14)/2) combinations, the agreement is less pronounced. In particular, for those pairs of sites characterized by small correlations between annual floods, 50-year correlations are quite widely scattered. The average or regional values tend to be dominated by those few combinations which have large correlations. Weighting the individual correlations by some measure of their overlapping record lengths was considered, but the average values are used to represent regional correlations and regional skew coefficients because no valid procedure was developed. The gaging stations associated with each of the 15 sites in the representative States are reported in Appendix C.

The 15 gages used in calculating regional parameters are sub-sets of gages which are themselves sub-sets of the total array of gages in a given State. For example, from Table 10, Massachusetts offers 17 sites for regression analysis of $\hat{Q}_{50}$ on basin characteristics, yet there are many more gages available in Massachusetts. Only those gages were used which had full, or relatively full, sets of basin characteristics so the remaining gages, approximately 125 in Massachusetts could not be used in generating regression estimates. Many locations have reported data that are not routinely available on USGS data files, so that a small portion of sites were usable. However, if it turns out that transfer of information is useful in Massachusetts, such transfer might be effected through expansion of the gaging network to accommodate additional "independent" sites. Thus some of the 125 gages might become part of the information network by measuring basin characteristics or by placing currently existing basin data into usable form;

this would be cheaper and more efficient than starting a new gage elsewhere in the State. This problem is confronted more fully in connection with calculating the parameter N_B, the number of gages in the network, where the problem of augmenting information through the use of regression models in each of the States is discussed.

This completes the preliminary economic and hydrologic analyses. A State-by-State examination of the value of improving hydrologic information appears in the next section. The first part of the section describes the existing hydrometric network in each region and gives current criteria for gage location. The second part utilizes BIGBASIN tables to calculate the equivalent years and, ultimately, the economic value of continued and extended networks.

Section 4

DECISION ANALYSIS

EXISTING HYDROMETRIC NETWORK

The existing hydrometric network for precipitation and streamflow gages for small drainage areas is summarized in Figures 36 through 38 and in Table 34. In addition, the information on these figures and tables appears in greater detail on three large format maps which have been prepared for FHWA as part of this contract. These three maps with overlays were prepared for use as visual aids for discussion of the gaging program. These include plotted locations of USGS (1) active streamflow gages, (2) inactive streamflow gages and (3) precipitation gages (active and inactive) shown on a continental United States map of 1 inch to 50 mile scale. Both crest and continuous streamflow gages are included. The active/inactive status represents the operating condition as of the fall, 1975 according to USGS computer files. Only areas of 50 square miles or less (i.e., small watershed basins) are considered.

The drainage areas are grouped according to Soil Conservation Service land resource regions. These regions, selected by the Federal Interagency Work Group on Hydrologic Data for Small Watersheds* as the best homogeneous geographical unit for evaluating hydrologic data bases, are delineated as broad geographic areas having similar patterns of soil, slope, climate, water resources, land use and type of farming. Since some States fall into more than one SCS region, the designated regions on the maps do not always coincide with State boundaries.

Figure 36 shows the active and inactive precipitation gages in the continental United States, distributed by State and SCS resource regions. In this study the FHWA has made a first attempt to show the

---

* Federal Interagency Work Group, op. cit.

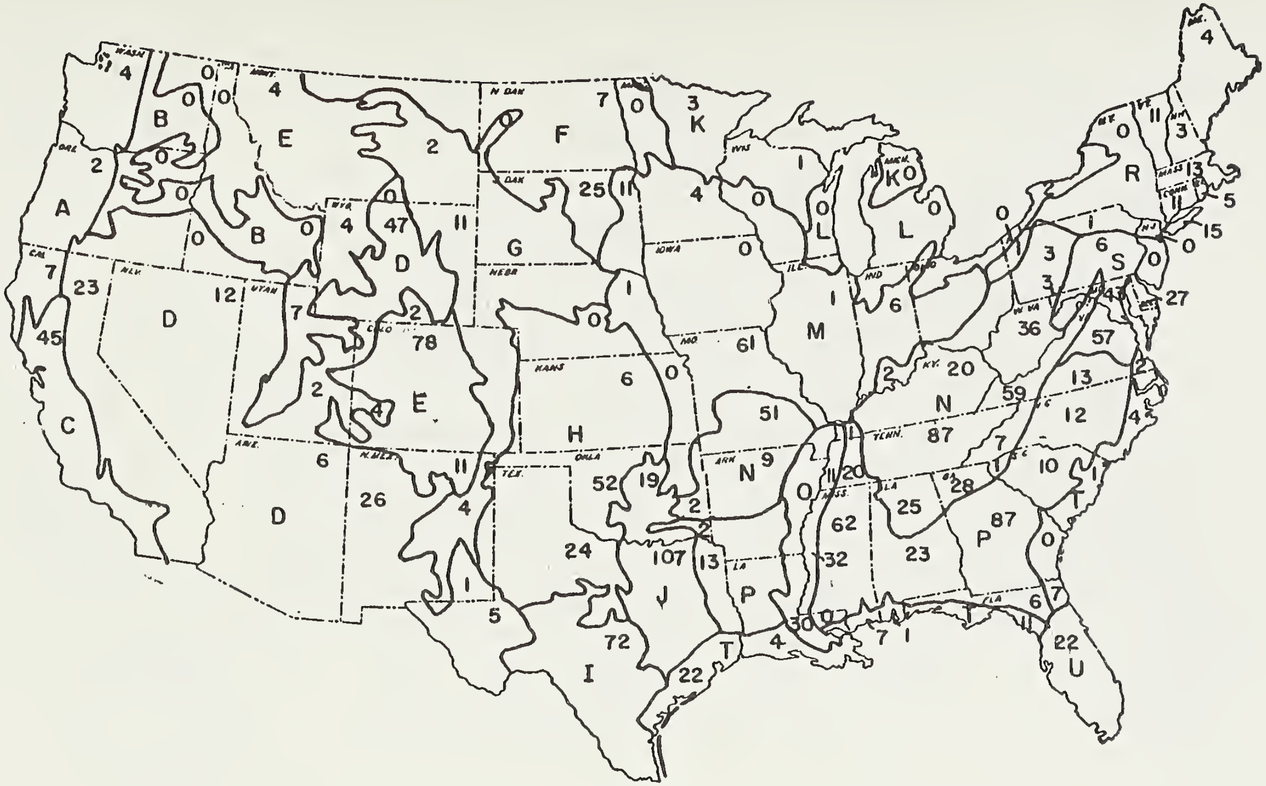


Figure 36. Active and Inactive Precipitation Gages by State and SCS Land Resource Regions

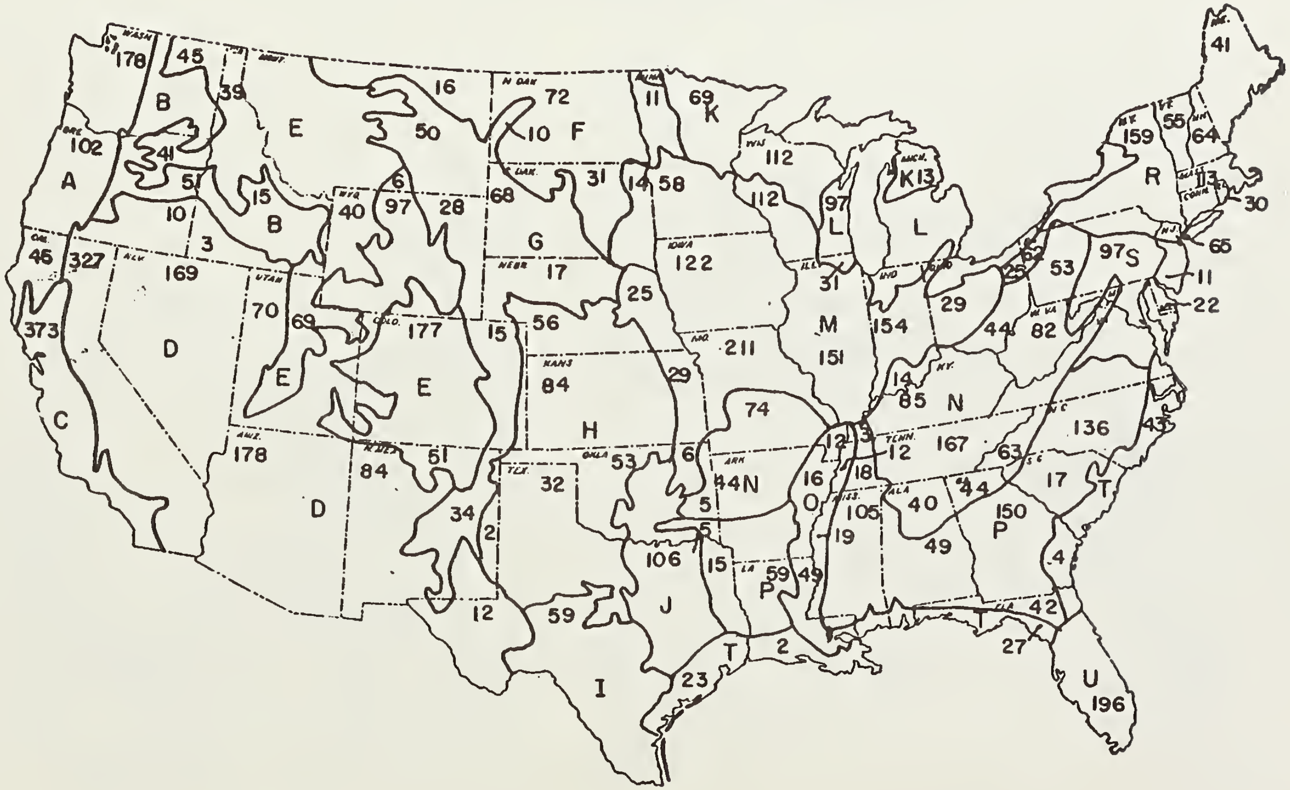


Figure 37. Active Runoff Gages, D.A.$\leq$50mi^2, by State and SCS Land Resource Regions

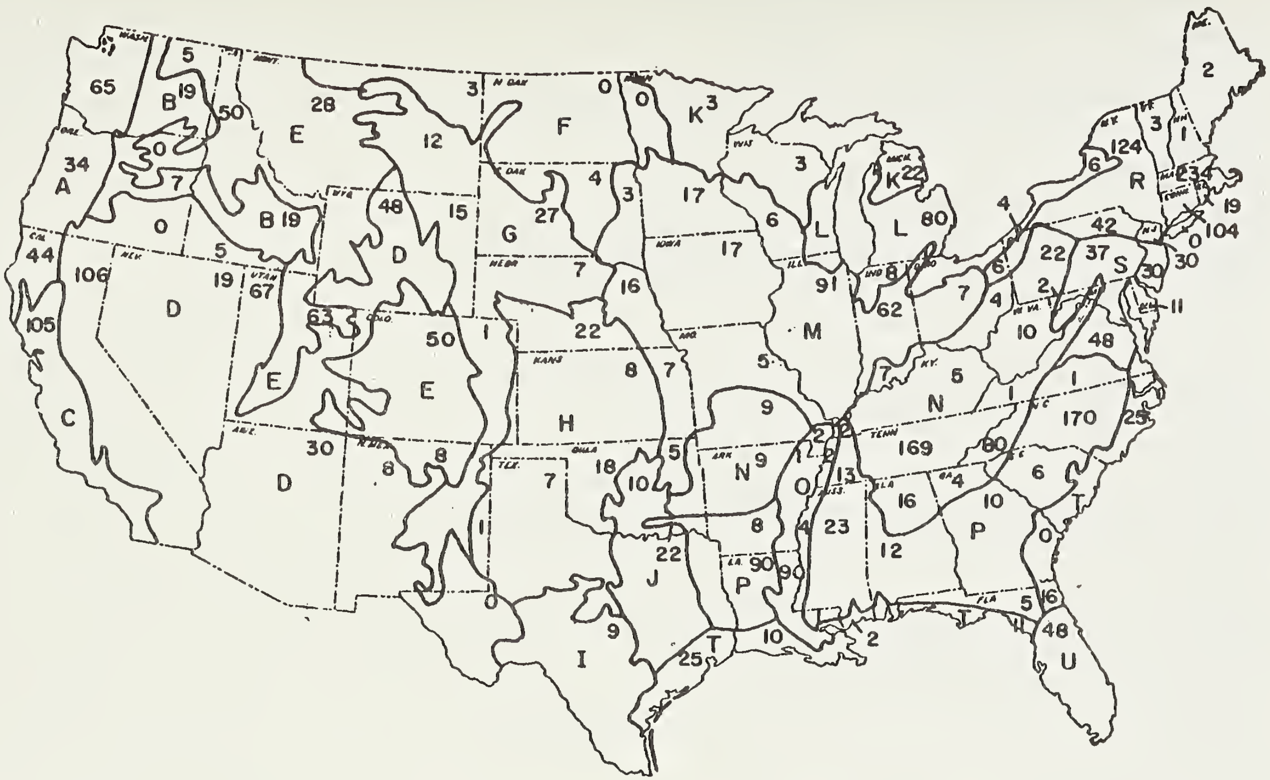


Figure 38. Inactive Runoff Gages, D.A.$\leq$50mi^2
by State and SCS Land Resource Regions

Table 34. The Hydrometric Network

	Precipitation Active	Precipitation Inactive	Streamflow (D.A. ≤ 50 mi.^{2}) Active	Streamflow (D.A. ≤ 50 mi.^{2}) Inactive	Land Area (mile^2)	Streamflow Gage Density (#/1000 mi.^{2})	Population per mi^2 (1960)
Alabama	37	12	93	28	51060	2.37	63
Alaska	2	0	137	75	571065	.37	.04
Arizona	6	0	178	30	113575	1.83	11
Arkansas	21	0	104	18	52499	2.32	34
California	54	21	745	255	156573	6.39	99
Colorado	95	4	211	66	103884	2.67	17
Connecticut	10	1	155	104	4899	52.87	506
Delaware	26	1	33	11	1978	22.24	217
Florida	41	5	265	80	54252	6.36	85
Georgia	113	2	198	14	58274	3.64	67
Hawaii	28	1	129	23	6415	23.69	99
Idaho	3	0	57	74	82708	1.58	8
Illinois	1	0	182	92	55930	4.90	179
Indiana	9	0	203	77	36185	7.74	128
Iowa	0	0	122	17	56032	2.48	49
Kansas	6	0	113	15	82048	1.56	26
Kentucky	20	1	88	7	39863	2.38	75
Louisiana	55	31	110	191	45106	6.67	67
Maine	4	0	41	2	31012	1.39	29
Maryland	44	5	116	17	9874	13.47	293
Massachusetts	13	0	143	234	7867	47.92	624
Michigan	0	0	125	103	57019	4.00	134
Minnesota	7	0	138	32	80009	2.12	41
Mississippi	97	4	127	29	47223	3.30	46
Missouri	109	6	297	16	69138	4.53	62
Montana	1	1	194	43	145736	1.63	4.6
Nebraska	0	0	98	45	76612	1.87	18
Nevada	12	0	169	19	109788	1.71	2.6
New Hampshire	3	0	64	1	9014	7.21	65
New Jersey	0	0	145	60	7521	27.26	774

Table 34. (continued)

	Precipitation Active	Precipitation Inactive	Streamflow (D.A. ≤ 50 mi.^{2}) Active	Streamflow (D.A. ≤ 50 mi.^{2}) Inactive	Land Area (Mi^2)	Streamflow Gage Density (#/1000 mi^2)	Population per mi^2 (1960)
New Mexico	41	1	171	17	121510	1.55	7.8
New York	16	0	221	130	47939	7.32	339
No. Carolina	15	8	242	279	49067	10.62	86
No. Dakota	7	0	82	0	69457	1.18	8.9
Ohio	10	0	113	11	40972	3.03	235
Oklahoma	64	22	87	41	68887	1.86	33
Oregon	2	0	204	41	96248	2.55	18
Pennsylvania	9	1	221	105	45007	7.24	250
Rhode Island	5	0	30	19	1058	46.31	708
So. Carolina	11	1	18	8	30272	.86	77
So. Dakota	86	9	113	27	76378	1.83	8.8
Tennessee	79	39	197	184	41762	9.12	84
Texas	193	51	247	87	262840	1.27	36
Utah	8	1	139	77	82339	2.62	10
Vermont	11	0	55	3	9276	6.25	41
Virginia	127	4	249	51	39838	7.53	97
Washington	3	1	318	89	66709	6.10	42
West Virginia	35	1	82	10	24079	3.82	77
Wisconsin	0	0	321	11	54705	6.07	70
Wyoming	34	30	165	88	97411	2.60	3.4

Other	18	12	266	81			
Total	1591	277	8321	3277			
	1,868		11,598				

hydrologic gaging network densities by physiographic region boundaries, rather than by State boundaries, as commonly reported by other agencies. There is no discrimination between active and inactive precipitation gages since they could not be separated in available data files. A total of 1,868 gages is distributed among the 50 States and the several territories and possessions. Table 34 includes information on the numbers of active and inactive precipitation gages, but these are listed by State only without reference to SCS resource regions. Similarly, Figure 37 shows the number of active streamflow gages for small drainage basins (with drainage areas not exceeding 50 square miles), and Figure 38 the number of inactive streamflow gages; these are organized by State and SCS region.

It was noted that gage counts available from current USGS data files do not always agree with those available from various State documents and other sources. Investigation of several of these discrepancies revealed that some State agencies impose additional criteria for publication of records. For example, in some States record lengths must exceed a threshold, while USGS files are more complete. Some State documents do not show gages on drainage canals, floodways and other hydraulic conveyances; again, these are listed in the USGS documents (with zero drainage areas). We were able adequately to explain the discrepancies in each State studied, and feel confident the USGS documentation and State reports could be made to agree if all the restrictions and constraints were carefully considered. The data in Table 34 are taken as the definitive USGS counts on small drainage area gages, both active and inactive.

In addition, Table 34 contains the land area of each State and the gage density, in number of active and inactive gages per 1,000 square miles. The last column of Table 34 gives the population per square mile based on 1960 Census data. The raw correlation between gage density and population density is 0.85; various other correlations can be calculated for combinations of logarithms and raw data, as shown in Table 35.

Table 35. Correlations Between Population Density and Gage Density, for 50 States

	Population	log Population
no. gages	.85	.58
log no. gages	.76	.78

The point is that a substantial portion of the gaging network is associated, but not necessarily causally, with population density. The gages are where the people are, and the people are where the economic action is. This suggests that gage locations have heretofore not been selected so much because they help reduce variance or because they provide equivalent or actual years of information but rather because they are located where there is a large potential for economic loss. This argument supports the position adopted by this study -- that the location of gages should be dictated by economic considerations in concert with statistical criteria, and that economic considerations have in fact, explicitly or otherwise, been part of the location decision for a long time.

DEVELOPMENT OF A DECISION TABLE

Table 36 contains the heart of the analysis. This section shows how the entries in each column of that table are prepared from the material in earlier sections. The format for this presentation is to number each column and to step through the headings and definitions. The initial column of the table defines the region of the study. In our work a region is defined by one of 11 representatives, for which the basic hydrologic parameters (but not necessarily economic benefits and costs) are assumed to be homogeneous. Separate computations have been developed for each of the 50 States, but these have been grouped

Table 36. Decision Table

1		2	3	4	5	6	7	8
Region	State	N_R	N_B	N_L	N_Y	ρ_{50}	G	η
1	Alabama		93		17.1			
	Arkansas		104		13.4			
	Florida		265		13.1			
	Georgia*	123	198	24.3	12.2	.123	.728	.237
	Louisiana		110		12.9			
	Mississippi		127		18.7			
	N. Carolina		242		15.6			
	S. Carolina		18		12.2			
2	Connecticut		155		12.2			
	Massachusetts*	17	143	37.7	15.9	.233	.985	.315
	Maine		41		10.4			
	New Hampshire		64		19.1			
	New York		221		21.8			
	Rhode Island		30		11.9			
	Vermont		55		11.2			
3	Iowa		122		15.4			
	Minnesota		138		12.3			
	Missouri*	101	297	24.0	16.1	.089	.588	.193
4	Idaho		57		11.1			
	Montana*	103	194	23.4	14.1	.094	.793	.258
	N. Dakota		82		15.6			

Table 36. (continued)

	9	10	11	12	13	14	15	16
State	SE(R)	$\mu(\ell n)$	$\sigma(\ell n)$	$\hat{Y}$	γ	Y	N_Y*	Y*
Alabama					.51	0.28	22.1	0.28
Arkansas					.51	0.28	18.4	0.28
Florida					(DR)			
Georgia*	1.135	9.572	0.723	0.406	.51	0.28	17.2	0.28
Louisiana					(DR)			
Mississippi					(DR)			
N. Carolina					(DR)			
S. Carolina					.86	0.10	17.2	0.10
Connecticut					.17	3.3	17.2	3.4
Massachusetts*	0.566	8.576	0.774	1.870	.19	2.7	20.9	2.7
Maine					.18	3.0	15.4	3.0
New Hampshire					(DR)			
New York					(DR)			
Rhode Island					.20	2.2	16.9	2.3
Vermont					.17	3.3	16.2	3.4
Iowa					(DR)			
Minnesota					.21	2.0	17.3	2.0
Missouri*	0.861	9.152	0.872	1.026	(DR)			
Idaho					.50	0.30	16.1	0.30
Montana*	1.291	8.078	0.814	0.398	(DR)			
N. Dakota					(DR)			

Table 36. (continued)

State	17 $\sqrt{Y/Y^*}$	18 SE*(R)	19 Q_d	20 Q^*_d	21 % Red'n	22 $/% x 10^6
Alabama	1.00	1.135	93,410	93,410	0	
Arkansas	1.00	1.135	93,410	93,410	0	
Florida					0	
Georgia*	1.00	1.135	93,410	93,410	0	
Louisiana					0	
Mississippi					0	
N. Carolina					0	
S. Carolina	1.00	1.135	93,410	93,410	0	
Connecticut	0.99	0.558	13,493	13,307	1.4	0.129
Massachusetts*	1.00	0.566	13,493	13,493	0	
Maine	1.00	0.566	13,493	13,493	0	
New Hampshire					0	
New York					0	
Rhode Island	0.98	0.554	13,493	13,218	2.0	0.737
Vermont	0.99	0.558	13,493	13,307	1.4	0.191
Iowa					0	
Minnesota	1.00	0.861	39,052	39,052	0	
Missouri*					0	
Idaho	1.00	1.291	27,123	27,123	0	
Montana*						
N. Dakota						

Table 36. (continued)

State	23 $ Saved	24 $ Cost	25 $ Net Benefits
Alabama			
Arkansas			
Florida			
Georgia*			
Louisiana			
Mississippi			
N. Carolina			
S. Carolina			
Connecticut	180,600	60,500	120,100
Massachusetts*			
Maine			
New Hampshire			
New York			
Rhode Island	1,474,000	36,300	1,437,700
Vermont	267,400	60,500	206,900
Iowa			
Minnesota			
Missouri*			
Idaho			
Montana*			
N. Dakota			

Table 36. (continued)

1		2	3	4	5	6	7	8
Region	State	N_R	N_B	N_L	N_Y	ρ_{50}	G	η
5	Arizona		178		10.3			
	New Mexico*	76	171	28.3	19.2	.115	.501	.166
	Oklahoma		87		11.4			
	Texas		247		10.8			
6	Illinois		182		16.8			
	Indiana		203		12.4			
	Michigan		125		14.7			
	Ohio*	71	113	29.7	20.4	.130	.754	.247
	Wisconsin		321		13.2			
7	California		745		13.8			
	Oregon*	105	204	39.2	14.4	.186	1.156	.368
	Washington		318		15.5			
8	Kentucky		88		20.5			
	Pennsylvania		221		13.6			
	Tennessee*	28	197	23.2	12.9	.163	1.018	.327
	W. Virginia		82		14.0			
9	Colorado		211		12.7			
	Nevada		169		10.8			
	Utah*	30	139	22.1	19.9	.312	.859	.279

Table 36. (continued)

	9	10	11	12	13	14	15	16
State	SE(R)	$\mu(\ell n)$	$\sigma(\ell n)$	$\hat{Y}$	γ	Y	N_Y*	Y*
Arizona					.50	.30	15.3	.30
New Mexico*	1.414	8.001	0.891	0.397	(DR)			
Oklahoma					(DR)			
Texas					(DR)			
Illinois					(DR)			
Indiana					.27	1.1	17.4	1.2
Michigan					(DR)			
Ohio*	.798	7.637	0.775	0.943	(DR)			
Wisconsin					.28	1.2	18.2	1.2
California					.58	.30	18.8	.30
Oregon*	.905	7.656	0.626	0.478	(DR)			
Washington					(DR)			
Kentucky					(DR)			
Pennsylvania					(DR)			
Tennessee*	.754	9.364	0.659	0.764	.30	1.1	17.9	1.1
W. Virignia					(DR)			
Colorado					.18	2.7	17.7	2.7
Nevada					.17	2.9	15.8	3.0
Utah*	.505	6.327	0.648	1.647	(DR)			

Table 36. (continued)

	17	18	19	20	21	22
State	$\sqrt{Y/Y^*}$	SE*(R)	Q_d	Q^*_d	% Red'n	$\$/\%\times10^6$
Arizona	1.00	1.414	30,764	30,764	0	
New Mexico*					0	
Oklahoma					0	
Texas					0	
Illinois					0	
Indiana	0.96	0.764	7,736	7,315	5.4	0.710
Michigan					0	
Ohio*	1.00	0.798	7,736	7,736	0	
Wisconsin					0	
California	1.00	0.905	9,407	9,407	0	
Oregon*					0	
Washington					0	
Kentucky					0	
Pennsylvania					0	
Tennessee*	1.00	0.754	40,461	40,461	0	
W. Virginia					0	
Colorado	1.00	0.505	1,287	1,287	0	
Nevada	0.98	0.497	1,287	1,269	1.4	0.048
Utah*					0	

Table 36. (continued)

	23	24	25
State	$ Saved	$ Cost	$ Net Benefit
Arizona			
New Mexico*			
Oklahoma			
Texas			
Illinois			
Indiana	3,834,000	60,500	3,773,500
Michigan			
Ohio*			
Wisconsin			
California			
Oregon*			
Washington			
Kentucky			
Pennsylvania			
Tennessee*			
West Virginia			
Colorado			
Nevada			
Utah*	67,200	60,500	6,700

Table 36. (continued)

Region	State	N_R	N_B	N_L	N_Y	ρ_{50}	G	η
	1	2	3	4	5	6	7	8
10	Delaware		33		12.7			
	Maryland		116		17.8			
	New Jersey		145		24.0			
	Virginia*	145	249	26.4	13.8	.081	.644	.212
11	Kansas		113		15.6			
	Nebraska		98		17.1			
	S. Dakota		113		11.9			
	Wyoming*	70	165	23.7	13.1	.158	1.185	.377

Table 36. (continued)

	9	10	11	12	13	14	15	16
State	SE(R)	$\mu(\ell n)$	$\sigma(\ell n)$	$\hat{Y}$	γ	Y	N_Y*	Y*
Delaware					.65	0.20	17.7	0.20
Maryland								
New Jersey								
Virginia*	1.555	9.798	0.944	0.369				
Kansas					.59	0.34	20.6	.35
Nebraska					.59	0.34	22.1	.35
S. Dakota					.59	0.33	16.9	.34
Wyoming*	1.247	7.555	0.658	0.278	.59	0.34	18.1	.35

Table 36. (continued)

	17	18	19	20	21	22
State	$\sqrt{Y/Y*}$	SE*(R)	Q_d	Q^*_d	%Red'n	$/%x10^6
Delaware	1.00	1.555	234,157	234,157	0	
Maryland					0	
New Jersey					0	
Virginia*					0	
Kansas	0.99	1.229	14,951	14,502	3.0	0.340
Nebraska	0.99	1.229	14,951	14,502	3.0	0.417
S. Dakota	0.99	1.229	14,951	14,502	3.0	0.145
Wyoming*	0.99	1.229	14,951	14,502	3.0	0.169

Table 36. (continued)

State	23 $ Saved	24 $ Cost	25 $ Net Benefit
Delaware			
Maryland			
New Jersey			
Virginia*			
Kansas	1,020,000	60,500	959,500
Nebraska	1,251,000	60,500	1,190,500
S. Dakota	435,000	60,500	374,500
Wyoming*	507,000	60,500	446,500

LEGEND OF TABLES 36 and 37

N_R:	Number of stations in regression analysis
N_B:	Number of active gages in the State
N_L:	Average length of record for N_R
N_Y:	Average length of record for N_B
ρ_{50}:	Regional correlation for $\hat{Q}_{50}$ events
G:	Regional skew coefficient for $\hat{Q}_{50}$ events
η:	Regional coefficient of variation for $\hat{Q}_{50}$ events
SE(R):	Standard error from regression analysis (in logarithm units)
μ(ln):	Average of the mean (in logarithm space) of the estimates of Q_{50} events
σ(ln):	Average of the standard deviation (in logarithm space) of the estimates of Q_{50} events
$\hat{Y}$:	Apparent equivalent record length (years)
γ:	Modal value of the model error
Y:	True equivalent record length (years)
N_Y*:	True equivalent augmented record length
(DR):	Dominated result

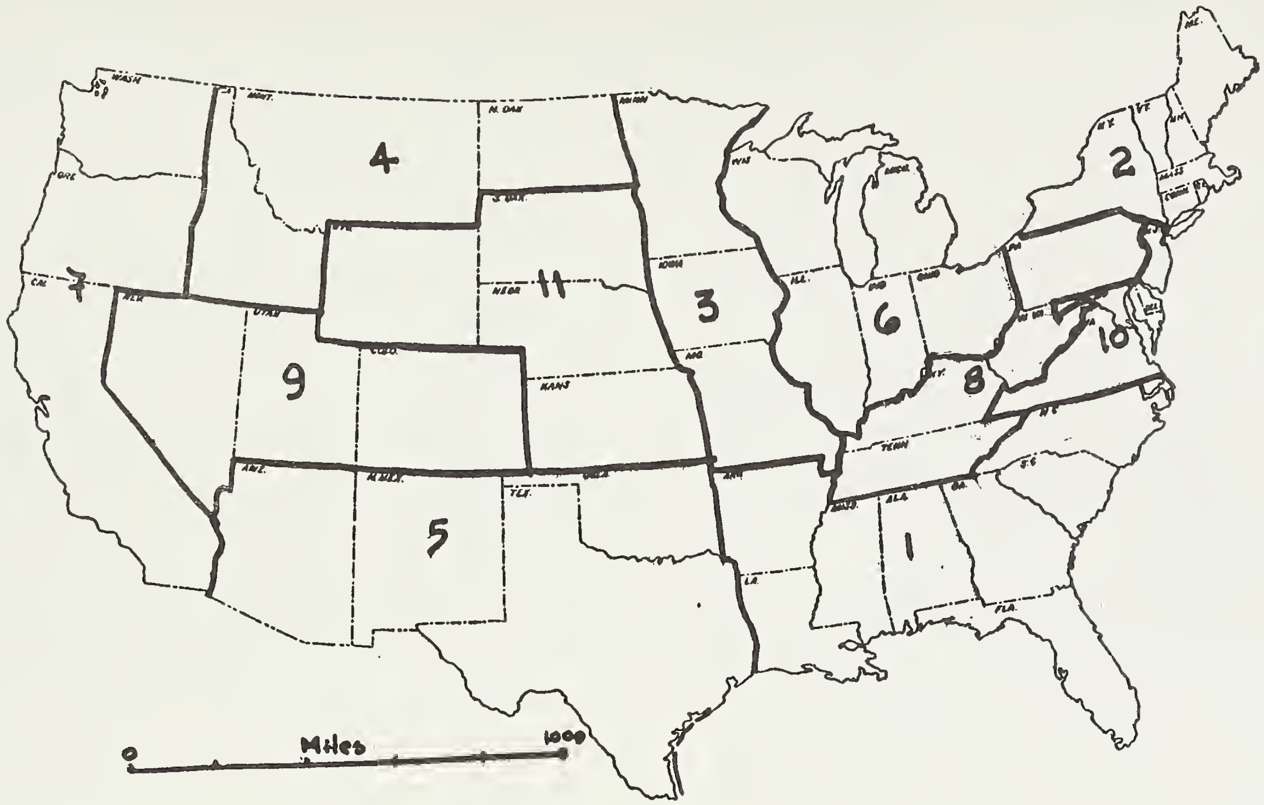


Figure 39. Eleven Representative Hydrologic Regions

into the 11 regions for which regional hydrologic parameters are homogeneous. The several States are listed alphabetically within each of the 11 regions, and a map appears as Figure 39. If a more definitive criterion for identifying hydrologic regions is available to a State, it may disaggregate all its gages accordingly and re-evaluate the program for each region.

1. All States within each region are listed. That State which is representative of the region (for example, for Region 1, the representative State is Georgia) is identified with an asterisk.

2. N_R is the number of stations used in the regression analysis for that region. For Georgia the value is 123.

3. N_B is the number of active gages in each of the States. These values also appear in Table 34, and are based on gage counts for drainage areas not exceeding 50 square miles. For example, Georgia has 198 active gages (but of these only 123 have complete basin characteristics and can be used in the regression analysis for that State and the region).

4. N_L is the average length of record for the N_R gages used in the regression. For example, for Georgia it is the average number of station-years of observation at the 123 locations. This statistic is reported only for that State which characterizes the region.

5. N_Y is the average length of record for the N_B active gages within the State. This represents the average length of record for the complete pool of active gages which could be used as part of the information network if complete basin characteristics were available. The amount of such data to be collected at each of these locations is a function of the number of steps (i.e., number of independent variables) utilized in the regression analysis and of the amount of data already available. With respect to the first criterion, it does not follow that "better" regressions are obtained with more independent variables because it is well known that the addition of such variables might introduce too much noise. This is tested by analysis of variance at successive steps

of the regression, using the t- and F-statistics. The reliability of each step is summarized by listing the standard error. These values, in Table 10, indicate when the regression gets better as more variables are added and also when it begins to get worse (i.e., the standard error increases), indicating that too many independent variables are included. The point at which this reversal occurs identifies the number of independent variables which should be included, and defines the data needs for that State or region.

6. This column gives the regional correlation for events Q_{50} based on 105 pairwise combinations deduced from 15 gages within each State. These values are abstracted from Table 33, taking an average of those values derived for 10 years of simulation and those for 25. The symbol ρ_{50} is used for this parameter.

7. The regional skew coefficient for events Q_{50}, also abstracted from Table 33, is given here. The symbol G represents this parameter.

8. The regional coefficient of variation of 50-year events is tabulated. This is given the symbol $(CV)_{50}$ or η.

9. The standard error based on the regression analysis is given here. This value is used in calculating the apparent equivalent record length; its symbol is SE(R). As described above, the minimal standard error is utilized from the alternative regression analyses. Table 10 gives two sets of analyses for each State; one is based on the WRC estimate and one on the WRC* estimate of events Q_{50}. Even though the correlation coefficients are generally higher for the WRC estimates, the WRC* standard errors are used because they are unbiased. Thus the standard errors are smaller for the WRC calculations, but this is a spurious advantage. The fit is better, but the data to which the fit is made are less reliable than those (which can be fit less well) from the WRC* technique. The standard errors are given in logarithmic units because the regression analyses themselves use exponential fitting procedures.

10. The mean μ (of the logarithms) of the 50-year estimates are averaged and tabulated in this column. The computation is based on an elaborate averaging scheme applied to each State. First the relationship between site 1 (in any State) and all the remaining 14 sites is considered, and that site identified with which site 1 has the longest overlapping record. This gives the best estimate of the mean and standard deviation, in raw data space, for site 1. The extent of overlap is noted in Tables 11-21. The material in columns 10 and 11 is based on the time or sampling error delineated by Moss and Karlinger; the average at a given site is taken over time. From the mean and standard deviation of estimates of Q_{50} it is a simple matter to compute the coefficient of variation and then uniquely to define the mean and standard deviation in log-space on the assumption of log-normality of the annual flood events, but these supplementary tabulations are not included in this Report. The analysis then considers the second site within this State, examines all the remaining 14 sites to determine that combination with the longest overlap, and proceeds to calculate the mean and standard deviation in raw data space and then ultimately in log-space for that pair. The computation proceeds through all 15 sites which comprise the State, whereupon the mean of the logs is averaged and reported in column 10. The standard deviation of the logs is preserved for subsequent calculation. This sequential procedure provides the "best estimate," or the longest record of overlap, at each step.

11. As described in the explanation for column 10, the standard deviation of logarithms is available at each site, and is averaged across all sites for that State; it is represented by the symbol $\sigma(\ln)$.

12. The apparent equivalent record length, in years, is given by the square of the ratio of column 11 to column 9. This is based on the results of Moss and Karlinger. It takes the symbol $\hat{Y}$.

Hardison* has proposed a simple correction for calculating the variance of Q_T from the variance σ^2 of the mean annual flood which

---

* Hardison, Clayton H., <u>USGS Prof. Paper 650-D</u>, <u>op. cit.</u>

would influence calculation of $\hat{Y}$. The standard error of Q_T is

$$SE(Q_T) = \sigma[(1 + k^2/2)/N]^{\frac{1}{2}} \tag{4}$$

where N is the record length (in years) and k is the standardized normal deviate corresponding to a recurrence interval of T years. For the 50-year flood, k = 1.64 so that the standard error of the event Q_{50} is 1.53 times the standard error of the mean annual flood. Unfortunately, use of this correction factor requires that the population standard deviation, σ, be known. As in most problems in applied statistics, this virtually is never the case in hydrologic practice. Unbiased estimates of the population standard deviation can be obtained, but these require knowledge of the population mean because it is necessary to know the coefficient of variation in order to unbias the results. In other words, the mean and standard deviation in the real situation are not independent, whereupon the assumptions which underlie Hardison's correction are violated because he assumes that the variance of Q_T is the sum of variances of the mean and of a multiple of the standard deviation, with independence between these two additive terms. This independence does not seem to be defensible, so the correction is not made.

One consequence of ignoring the correction is a slight shift in results owing to a change in the basic time scale. The BIGBASIN tables are derived for annual events, or events for which there is precisely one occurrence each year. While they might equally well be used for other events with different return intervals, the relationship between years of record and sample size must somehow be preserved. In other words, when used for events Q_T, there is no longer one event per year but rather one event for every T years and the relationship between T and N, the record length, becomes important. Thus a (say) 10-year record defines only one estimate of the floods Q_5, Q_{10}, Q_{25}, Q_{50},... there is not one event for each of N years but rather a vector of potential events to be estimated by extrapolation of the N-year record.

The strategy in this study is to utilize such characteristics as are available from the record to estimate parameters of the distributions of these several statistics Q_T and then to assume that these distributions are the <u>correct</u> distributions in the sense that they are already subjected to whatever adjustments and modifications are appropriate (such as the Hardison correction discussed above). Thus it is a moot point as to whether the Hardison (or any other) correction should be used at all, and in this study it was decided for consistency to use no correction rather than to introduce a correction of unknown properties. Among the candidates for consideration is a correction which shifts the time scale from years to decades, hoping to capture some of the flavor of the analysis by suggesting that each decade gives rise to a flood estimate Q_T (where T is typically 50 years), but this was rejected because the parameters of the distribution of Q_T do not seem to change significantly as the record length ranges from 10 to 25. This is the record length customarily available in hydrologic analysis of this sort, and if the parameters of various distributions of extrema (as evaluated by careful Monte Carlo analysis) do not change significantly, we are hard pressed to justify a general correction for the method.

In any event, the USGS indicates that tables are currently being prepared for more precise numerical evaluation of the sampling characteristics of extrema Q_T, so it is likely that the question of correcting the BIGBASIN tables to accommodate extrema will be resolved by the existence of tables based on the Monte Carlo analysis of the extrema themselves.

13. The modal value of the model error is calculated from Table 1 of BIGBASIN. A four-way linear interpolation rule is used. The arguments for using the table are N_B, N_Y, ρ_{50} and $(CV)_{50}$; these appear in columns 3, 5, 6 and 8, respectively of Table 36. The step sizes are large, so interpolation is always required. A linear interpolation routine is utilized to develop estimates of the modal value of the model error based on BIGBASIN tables. The order of interpolation is:

ρ_{50}, $(CV)_{50}$, N_B, N_Y, which minimizes the amount of manipulation required. The resulting model error is given the symbol γ.

14. The true equivalent years of record is read from Table 3 of BIGBASIN by entering with the same arguments as for column 13, augmented by the model error. Interpolation allows estimating the median value of the true equivalent years of record, or that value corresponding to the 0.5 probability of exceedance, which is tabulated.

15. N_Y* is the augmented record length, taken here to be $N_Y + 5$. We presume that program extensions of less than five years are infeasible because the minimum step in the BIGBASIN tables is five years and the corresponding increase in equivalent years of record is small. That is, extension of the program for only one year would not significantly improve the equivalent years of record for the gaging data.

16. The true value of equivalent years is estimated on the basis of the extended record length; it bears the symbol Y*. It is important here to note that the model error is assumed constant through the extended period of gaging. That is, the interpolation required to generate the values in column 13 is not repeated because the modal value of the model error is assumed not to change. Thus it is necessary only to change one of the arguments for utilizing Table 3 in BIGBASIN, and thereby directly to tabulate the true equivalent years of record under the extended gaging program.

17. The reduction in standard error of $\hat{Q}_{50}$ is identified by the square root of the ratio of true equivalent years, or $(Y/Y^*)^{\frac{1}{2}}$.

18. The modified standard error, SE*(R), is given by the product of columns 9 and 17.

19. The design flow, Q_d, under the original gaging program is tabulated here. On the assumption that the logarithms of all potential design events are normally distributed, and following upon the scheme portrayed in Figures 7 through 9, the design flow is that value corresponding to the logarithm which will be exceeded with probability

0.05. Given a distribution of logarithms of events Q_{50}, move toward the right-hand tail just far enough so that the area to the right of the cut-off is 0.05 or 5 percent of the (unit) total area under the distribution. This corresponds to a security level of $\alpha = 1.65$, so that calculation of the flow Q_d is a simple matter of adding the mean (contained in column 10) to 1.65 times the standard error (contained in column 9) and then taking the antilog.

20. The computation in column 19 is repeated except that the standard error is derived from column 18, that value based on the extended gaging program instead of the original gaging program. The security level of 1.65 is maintained. (It would be interesting in subsequent studies to evaluate the sensitivity of conclusions reached here to the security level α, but this is beyond the scope of this study.)

21. The percentage reduction in design flow, derived from columns 19 and 20, is tabulated. This becomes the basis of evaluating economic benefits associated with improving estimates of the design flow.

22. The reduction in cost associated with a unit (i.e., 1 percent) reduction in design flow, extracted from Tables 7 and 8, is repeated here.

23. The actual dollar savings, the product of columns 21 and 22, is given.

24. The cost of continuation of the gaging network for five years in each State is calculated on the basis of an O&M cost of $242 per gage per year (personal communication, USGS), or $1,210 over a five-year decision horizon. The result is tabulated in this column. This assumes the States pay only for crest stage type gages and the cost is divided equally between the States and the USGS. Amortization, a sunk cost, is already paid and is not a factor in this decision at the margin. It would make a significant difference.

25. The net benefits, derived by subtracting column 24 from column 23, are given here. Positive values indicate States in which the

gaging program should be continued for the next five years, while no entries indicate that the gaging program should be discontinued in its present form if serving the FHWA's needs for flood estimation is the program's sole objective. This does not imply that all gaging should be terminated because there are other purposes served by gaging.

DISCUSSION OF RESULTS

The computations in Table 36 are grouped according to the 11 representative States. Each State is associated with its own gaging intensity (column 3) and its average length of record (column 5). The hydrologic parameters for the region are given only for the representative State within each group. Because the BIGBASIN tables do not extend beyond 50 sites per gaging region (excepting a few incomplete results for 60), the results in Table 36 are based on reducing all values of N_B to 50 if there are more than 50 sites in any State. The amount of potential information gain beyond this point is negligible, so there is no significant error introduced by this truncation.

The most efficient and advantageous gaging program in a region is generally found in that State with the shortest record length, N_Y. It is assumed that all States in that region have the same apparent number of years of equivalent record, $\hat{Y}$ (column 12), and the "best" regression is that which produces that estimate of equivalent years from the smallest sample size. In other words, if column 12 is a surrogate for the precision of the regression in that it identifies the apparent number of equivalent years, it is better to do so with a shorter record length because that implies less noise in the regression. However, for very large values of N_Y, the sample correlation becomes a better estimate of the population value (even if ρ is small), so the regression becomes "better" again. Calculations are performed first for the most advantageous state or regression in a region because if the most advantageous regression analysis can not improve the results, then no inferior regressions can improve the estimates of Q_{50}, so these programs can be evaluated without recourse to calculations. Thus many entries in Table 36 are identified by the symbol (DR), which means the results for that State are dominated by at least one regression analysis in that region. The value "0" is inserted in column 21 to represent the percent reduction in design flow associated with dominated States.

The computation is completed for all undominated States using the basic hydrologic information contained in columns 6 through 12 for that region's representative State. The BIGBASIN tables contain discrete class entries for $\hat{Y}$ and model error (columns 11 and 12, respectively) using step sizes of 0.5 units. This is a coarse resolution, from which the true values of the expected equivalent years (with and without gaging extensions), given in columns 14 and 16, respectively, can reliably be interpolated to no more than one decimal place. Thus the reduction in standard error which can be attributed to gaging extension does not have very high precision. For example, Table 36 shows that all four States in Region 11 have a 1 percent reduction in the standard error of the design flow, or in the standard deviation of $\hat{Q}_{50}$, leading to a flow reduction of 3 percent, shown in column 21 for all four States in Region 11. This flow reduction is an extremely unstable estimate because of the imprecisions associated with interpolation.

The USGS is currently fitting analytical functions to the BIGBASIN tabular data, at least for selected combinations of arguments in BIGBASIN, and when these are available, it will be possible to make more precise estimates of the model error and equivalent years. In earlier sections it was noted that the interface between statistical and economic sections of this work represents potential inconsistency as between mathematical precision and economic interpolation. Here we note these potential errors in interpolation and the degree of resolution in the tables. Thus when applying the percentages in column 21 to the economic benefits associated with reducing design flows, the instabilities of both sources (statistical and economic) should be borne in mind.

It is clear from Table 36 that often there is little to gain from extending the gaging programs in their present form, particularly if the objective of such programs is limited to the design of drainage structures in small watersheds. Only nine of 48 States (Connecticut, Rhode Island, Vermont, Indiana, Nevada, Kansas, Nebraska, South Dakota and Wyoming) show any reduction in the design flow consequent upon five-year extensions of existing gaging programs. Eight of these nine (all but Rhode Island) have gaging programs with more than 50 gaged sites, so the statistical analysis presented in the table would be unchanged if programs in the eight States were limited to 50 sites and that in Rhode Island maintained for all 30. But even in some of these nine States the advantages of the gaging program are slender, and it is realistic to ask if similar results could be obtained under a reduction to 25 gages in each State. This analysis is reported in Table 37, which is similar to Table 36 except that N_B for each State is set at 25. Only South Carolina, with 18 gages, would be unable to meet this requirement; for purposes of consistency, this inability is ignored in the table.

For $N_B = 25$ the apparent inconsistence in estimating model error γ is more pronounced. That is, there is a stronger tendency for model error to increase with N_Y and then to reverse and decrease before N_Y becomes very large. Thus arguments of dominance can not readily be made in Table 37, and more computation was required.

The modified decision table also assumes that the modal model error, column 13 of Table 36, is unchanged under the new gaging assumption. This is reasonable because the error is based on regression results which, in turn, use the existing gaging network. Thus the reduced network would not increase the error because the old value, based on larger amounts of information, is still available. It could be argued that if the same regressions are deduced from $N_B = 25$ sites, the model error must change. But we assume that minor changes in the regression occur around a pivotal or fixed value of the model error.

Table 37 does not contain all the repetitive hydrologic information which appears in Table 36. The column numbers are preserved so that entries can readily be compared.

Table 37. Modified Decision Table for Reduced Network

1		2	3	4	5	6	7	8
Region	State	N_R	N_B	N_L	N_Y	ρ_{50}	G	η
1	Alabama		25		17.1			
	Arkansas		25		13.4			
	Florida		25		13.1			
	Georgia*	123	25	24.3	12.2	.123	.728	.237
	Louisiana		25		12.9			
	Mississippi		25		18.7			
	N. Carolina		25		15.6			
	S. Carolina		18		12.2			
2	Connecticut		25		12.2			
	Massachusetts*	17	25	37.7	15.9	.233	.985	.315
	Maine		25		10.4			
	New Hampshire		25		19.1			
	New York		25		21.8			
	Rhode Island		25		11.9			
	Vermont		25		11.2			
3	Iowa		25		15.4			
	Minnesota		25		12.3			
	Missouri*	101	25	24.0	16.1	.089	.588	.193
4	Idaho		25		11.1			
	Montana*	103	25	23.4	14.1	.094	.793	.258
	N. Dakota		25		15.6			

Table 37. (continued)

	9	10	11	12	13	14	15	16
State	SE(R)	$\mu(\ell n)$	$\sigma(\ell n)$	$\hat{Y}$	γ	Y	N_Y*	Y*
Alabama					(DR)			
Arkansas					(DR)			
Florida					(DR)			
Georgia*	1.135	9.572	0.723	0.406	.73	0.10	17.2	0.10
Louisiana					(DR)			
Mississippi					.67	0.20	23.7	0.20
N. Carolina					(DR)			
S. Carolina					.85	0.10	17.2	0.10
Connecticut					(DR)			
Massachusetts*	0.566	8.576	0.774	1.870	(DR)			
Maine					.21	2.1	15.4	2.1
New Hampshire					(DR)			
New York					.21	2.1	26.8	2.1
Rhode Island					(DR)			
Vermont					(DR)			
Iowa					(DR)			
Minnesota					.22	1.8	17.3	1.8
Missouri*	0.861	9.152	0.872	1.026	.23	1.6	21.1	1.6
Idaho					.74	0.10	16.1	0.10
Montana*	1.291	8.078	0.814	0.398	(DR)			
N. Dakota					.70	0.10	20.6	0.10

Table 37. (continued)

State	17 $\sqrt{Y/Y^*}$	18 SE*(R)	19 Q_d	20 Q^*_d	21 % Red'n	22 $/%
Alabama		1.135	93,410	93,410	0	
Arkansas		1.135	93,410	93,410	0	
Florida					0	
Georgia*	1.00	1.135	93,410	93,410	0	
Louisiana					0	
Mississippi	1.00	1.135	93,410	93,410	0	
N. Carolina					0	
S. Carolina	1.00	1.135	93,410	93,410	0	
Connecticut					0	
Massachusetts*					0	
Maine	1.00	0.566	13,493	13,493	0	
New Hampshire					0	
New York	1.00	0.566	13,493	13,493	0	
Rhode Island					0	
Vermont					0	
Iowa					0	
Minnesota	1.00	0.861	39,052	39,052	0	
Missouri*	1.00	0.861	39,052	39,052	0	
Idaho	1.00	1.291	27,123	27,123	0	
Montana*					0	
N. Dakota	1.00	1.291	27,123	27,123	0	

Table 37. (continued)

State	23 $ Saved	24 $ Cost	25 $ Net Benefits
Alabama			
Arkansas			
Florida			
Georgia*			
Louisiana			
Mississippi			
N. Carolina			
S. Carolina			
Connecticut			
Massachusetts*			
Maine			
New Hampshire			
New York			
Rhode Island			
Vermont			
Iowa			
Minnesota			
Missouri*			
Idaho			
Montana*			
N. Dakota			

Table 37. (continued)

1 Region	State	2 N_R	3 N_B	4 N_L	5 N_Y	6 ρ_{50}	7 G	8 η
5	Arizona		25		10.3			
	New Mexico*	76	25	28.3	19.2	.115	.501	.166
	Oklahoma		25		11.4			
	Texas		25		10.8			
6	Illinois		25		16.8			
	Indiana		25		12.4			
	Michigan		25		14.7			
	Ohio*	71	25	29.7	20.4	.130	.754	.247
	Wisconsin		25		13.2			
7	California		25		13.8			
	Oregon*	105	25	39.2	14.4	.186	1.156	.368
	Washington		25		15.5			
8	Kentucky		25		20.5			
	Pennsylvania		25		13.6			
	Tennessee*	28	25	23.2	12.9	.163	1.018	.327
	W. Virginia		25		14.0			
9	Colorado		25		12.7			
	Nevada		25		10.8			
	Utah*	30	25	22.1	19.9	.312	.859	.279

Table 37. (continued)

	9	10	11	12	13	14	15	16
State	SE(R)	$\mu(\ell n)$	$\sigma(\ell n)$	$\hat{Y}$	γ	Y	N_Y*	Y*
Arizona					.75	0.10	15.3	0.10
New Mexico*	1.414	8.001	0.891	0.397	.67	0.20	24.2	0.20
Oklahoma					(DR)			
Texas					(DR)			
Illinois					.32	0.80	21.8	0.80
Indiana					.31	0.82	17.4	0.85
Michigan					.32	0.80	19.7	0.80
Ohio*	.798	7.637	0.775	0.943	.33	0.77	25.4	0.80
Wisconsin					.31	0.82	18.2	0.85
California					.81	0.20	18.8	0.20
Oregon*	.905	7.656	0.626	0.478	(DR)			
Washington					.80	0.20	20.5	0.20
Kentucky					.36	0.70	25.5	0.70
Pennsylvania					.33	0.92	18.6	0.96
Tennessee*	.754	9.364	0.659	0.764	.33	0.92	17.9	0.95
W. Virginia					.33	0.93	19.0	0.96
Colorado					.20	2.0	17.7	2.1
Nevada					.20	2.0	15.8	2.0
Utah*	.505	6.327	0.648	1.647	.20	2.1	24.9	2.1

Table 37. (continued)

	17	18	19	20	21	22
State	$\sqrt{Y/Y^*}$	SE*(R)	Q_d	Q^*_d	% Red'n	$\$/\% \times 10^6$
Arizona	1.00	1.414			0	
New Mexico*	1.00	1.414			0	
Oklahoma					0	
Texas					0	
Illinois	1.00	0.798	7,736	7,736	0	
Indiana	0.98	0.784	7,736	7,555	2.3	0.710
Michigan	1.00	0.798	7,736	7,736	0	
Ohio*	0.98	0.784	7,736	7,555	2.3	1.230
Wisconsin	0.98	0.784	7,736	7,555	2.3	0.495
California	1.00	0.905	9,407	9,407	0	
Oregon*					0	
Washington	1.00	0.905	9,407	9,407	0	
Kentucky	1.00	0.754	40,461	40,461	0	
Pennsylvania	0.98	0.738	40,461	39,415	2.6	1.395
Tennessee*	0.98	0.742	40,461	39,668	2.0	0.956
W. Virginia	0.98	0.739	40,461	39,457	2.5	0.992
Colorado	0.98	0.493	1,287	1,262	1.9	0.293
Nevada	1.00	0.505	1,287	1,287	0	
Utah*	1.00	0.505	1,287	1,287	0	

Table 37. (continued)

	23	24	25
State	$ Saved	$ Cost	$ Net Benefit
Arizona			
New Mexico*			
Oklahoma			
Texas			
Illinois			
Indiana	1,633,000	30,250	1,602,750
Michigan			
Ohio*	2,829,000	30,250	2,798,750
Wisconsin	1,138,500	30,250	1,108,250
California			
Oregon*			
Washington			
Kentucky			
Pennsylvania	3,627,000	30,250	3,596,750
Tennessee*	1,912,000	30,250	1,881,750
West Virginia	2,480,000	30,250	2,449,750
Colorado	556,700	30,250	526,450
Nevada			
Utah*			

Table 37. (continued)

1		2	3	4	5	6	7	8
Region	State	N_R	N_B	N_L	N_Y	ρ_{50}	G	η
10	Delaware		25		12.7			
	Maryland		25		17.8			
	New Jersey		25		24.0			
	Virginia*	145	25	26.4	13.8	.081	.644	.212
11	Kansas		25		15.6			
	Nebraska		25		17.1			
	S. Dakota		25		11.9			
	Wyoming*	70	25	23.7	13.1	.158	1.185	.377

Table 37. (continued)

	9	10	11	12	13	14	15	16
State	SE(R)	$\mu(\ell n)$	$\sigma(\ell n)$	$\hat{Y}$	γ	Y	N_y*	Y*
Delaware					0.72	0.10	17.7	0.10
Maryland					0.68	0.10	22.8	0.10
New Jersey					0.62	0.20	29.0	0.20
Virginia*	1.555	9.798	0.944	0.369	0.71	0.10	18.8	0.10
Kansas					0.81	0.20	20.6	0.20
Nebraska					0.80	0.20	22.1	0.20
S. Dakota					0.84	0.10	16.9	0.10
Wyoming*	1.247	7.555	0.658	0.278	0.83	0.10	18.1	0.10

Table 37. (continued)

State	17 $\sqrt{Y/Y^*}$	18 SE*(R)	19 Q_d	20 Q^*_d	21 % Red'n	22 $/%
Delaware	1.00	1.555	234,157	234,157	0	
Maryland	1.00	1.555	234,157	234,157	0	
New Jersey	1.00	1.555	234.157	234,157	0	
Virginia*	1.00	1.555	234,157	234,157	0	
Kansas	1.00	1.247	14,951	14,951	0	
Nebraska	1.00	1.247	14,951	14,951	0	
S. Dakota	1.00	1.247	14,951	14,951	0	
Wyoming*	1.00	1.247	14,951	14,951	0	

Table 37. (continued)

State	23 $ Saved	24 $ Cost	25 $ Net Benefit
Delaware			
Maryland			
New Jersey			
Virginia*			
Kansas			
Nebraska			
S. Dakota			
Wyoming*			

The nine States wherein continuation of a 50-site (30 for Rhode Island) gaging program would improve design flow estimates exhibit benefits from $6,700 for Utah to $3,773,500 for Indiana; total savings of $8,515,900 are realized in the nine States of Connecticut, Rhode Island, Vermont, Indiana, Utah, Kansas, Nebraska, South Dakota and Wyoming. Reduction of all gaging networks to 25, except South Carolina with 18, produces the results in Table 37 (column 25). Seven States have programs that justify continuation on the basis of improved design flow estimates, namely Indiana, Ohio, Wisconsin, Pennsylvania, Tennessee, Colorado and West Virginia. Net benefits range from $526,450 for Colorado to $3,596,750 for Pennsylvania; total benefits are $13,964,450. No discounting is considered in these economic evaluations.

No further limit on the size of the gaging program is imposed and tested in this analysis. This is because reductions below $N_B = 25$ would probably impact objective functions other than efficiency in estimation of a design flow for drainage works. The potential uses and importance of gaging information are indicated elsewhere; all or some of these are served by information which would be derived from the network of 25 gages in each State. If reductions below 25 gages per State are to be made, they should be made on the basis of policy decisions which lie beyond the scope of this investigation.

IMPLICATIONS OF THE RESULTS

A net loss associated with the gaging program does not mean that the program should be completely abandoned. Several options are available. First, the program might be contracted so that by saving $242 per station per year its cost might be brought more nearly into line with benefits. If the program is reduced, the calculations in Table 36 would have to be re-evaluated for new values of N_B, which would result in new estimates of the percentage reduction (in design flow) and consequently in new values of net benefits. This is done in Table 37.

Analysis by Moss and Karlinger suggests that new information is not accumulated very rapidly when the number of gaged sites exceeds 25. This is another way of saying that if the correlation structure is strong, the significant variables will explain the bulk of the variation long before 25 sites are utilized, while if the underlying correlation structure is weak, the addition of more sites might provide more noise than information. Thus it does not follow that "more is better," and a reasonable way to effect a streamlined data network is first to reduce the number of gaging locations to approximately 25 and then to re-do the necessary calculations to determine if this network configuration could pay for itself in terms of reduced design flow.

Second, the gages in our study are assumed to be crest stage recorders rather than continuous monitors. There are institutional constraints under which the USGS might reasonably feel that if it is going to the trouble of installing and maintaining a gage, it might as well be the type that provides maximal information, thereby precluding crest gages. If a State's contribution to operating gages is not 50-50 with the USGS and the rate is not based on crest stage recorder, new benefit analysis must be done to evaluate the program.

Third, because model error is the dominant source of noise in the estimation procedure, it is appropriate for the USGS to continue some part of its gaging program to generate data to develop better models and thereby to reduce standard errors of estimate and increase the number of years of true equivalent record. We can not now specify how large such a gaging enterprise should be, but simply because a gaging network does not provide cost effective results for one user, there is no reason to terminate the complete program.

Further gaging is recommended in those States or regions where the net benefits are positive, or can be made positive by reducing the size of the network to approximately 25 or 30 active gages. If further reductions in network size are required to drive the losses to zero, the information derived from such reduced networks should be recalculated. The calculations outlined above and summarized in column 25 of

Tables 36 and 37 highlight the States where continuation of a gaging program, whether 25 or 50 sites, results in net benefits.

Section 5

RECOMMENDED RESEARCH FOR IMPROVEMENT OF SMALL WATERSHED PROGRAM

FREQUENCY ANALYSIS REVISITED

The material in this section is intended for the Headquarters staff of FHWA rather than for those responsible for immediate field implementation of various programs. Having suggested that current design techniques are statistically weak and that efforts to improve them are inefficient, we offer here some programs for further research and possible implementation. It is not argued here that the use of $\hat{Q}_{50}$ is wrong -- only that its estimate should be unbiased and that transfer of information by regression is ineffective.

Three schemes are proposed, increasing in cost and complexity. The first restates the argument given by Harold A. Thomas, Jr., in which standard plotting positions are bounded by confidence bands to show how unstable the return interval is. The second describes the use of multinomial logit analysis in drainage design, implementation of which would require an important commitment to data collection and manipulation. The third scheme would involve a major research venture whose results could fundamentally alter hydrologic science. These descriptions utilize more statistical and mathematical notation than has been used elsewhere in this Report; the nature of the subject requires this level of presentation.

The use of probability theory to specify confidence limits for flows of various magnitudes was suggested almost 30 years ago by Harold A. Thomas, Jr.,* who studied the range of recurrence intervals (or probabilities of occurrence) associated with the mth largest value in a series of annual flood events. His results are based on integrations

---

* Thomas, Harold A., Jr., "Frequency of Minor Floods," op. cit.

utilizing the incomplete-Beta function, which expresses the confidence limits surrounding estimates of the parameter p, or the probability of success, in a series of independent Bernoulli trials. These results are non-parametric; that is, they do not depend on the assumption or specification of a particular density function.

Non-Parametric Technique

Consider n annual events ranked so that Q_1 is the largest, Q_2 the next,... Q_n the smallest. Let p be the probability that any flood event (although floods are used here, the analysis is symmetric with respect to low flows) is smaller than some flood Q. Each year of record has precisely one flood event, so that if in the year with Q_1 there were flood events larger than Q_2, Q_3,..., these would not be available to the analysis.

Consider the probability that precisely (n-m) floods will be smaller than some given value Q, that (m-1) floods will be larger than Q, and that one flood will fall in the range dQ surrounding the magnitude Q. This is given as

$$ {}_nC_m \, mp^{n-m} \, (1-p)^{m-1} $$

where p is the probability that a flood of size Q will not be exceeded in any year. This probability is in a form similar to that of the binomial density, except for the presence of an extra parameter m and for the fact that the exponents on the probabilities do not sum to n but rather to (n-1). These are due to the fact that only (m-1) are "failures" in that they are larger than Q and that any one of m floods can be the one centered at Q. Thus the total number of ways that the conditions can be met is not given by the combinatorial term (or binomial coefficient) but by the binomial coefficient multiplied by m.

It is not possible precisely to determine the probability or p-value associated with a particular flood. But it is possible to ascertain confidence limits associated with the statement that the true recurrence probability lies within certain fixed limits, or within a fixed tolerance interval. Thomas integrates the probability given above to determine the probability θ that the actual p-value of the mth ranked of n floods is less than some value p_0:

$$\theta = \binom{n}{m} m \int_0^{p_0} p^{n-m} (1-p)^{m-1} dp \tag{5}$$

or, for the largest flood with m = 1:

$$\theta = n \int_0^{p_0} p^{n-1} dp = p_0^{\,n} . \tag{6}$$

He gives some interesting numerical examples. For instance, he calculates the chance that the largest flood of a 25-year record has a true average return period between 20 and 100 years. These return periods correspond to probabilities of 0.95 and 0.99, respectively. From the interval for m = 1 (given above) the requisite probability is $0.99^{25} - 0.95^{25} = 0.5004$. That is, there is approximately a 50 percent chance that the actual probability of recurrence of the flood of record lies between 0.95 and 0.99, and about a 50 percent chance that the actual p-value lies outside these limits (which imply return intervals of 20 and 100 years). The return interval is <u>not</u> notably stable!

Calculations for ranks other than m = 1 can be made by using tables of the incomplete-Beta function, or integral of the probability equation given above. Thomas calculated a few points, giving the 50 percent confidence limits associated with the average return periods of the five largest floods taken from a 25-year record. For each of the ranks m = 1,2,...,5, the limits (in years) for the average return period are (18, 87), (10, 26), (6.6, 14), (5.1, 9.8), and (4.2, 7.3). Lesser floods, associated with larger values of the rank m, have tighter confidence intervals and therefore can be better estimated with regard to their frequency of occurrence. The larger floods of record, even for a record of 25 years' duration, have broad confidence intervals so that their average return periods can be estimated, but not with much security.

The Thomas paper also presents an integral which gives the probability that in t future years the mth of n past floods will be exceeded precisely k times, or

$$\phi_k = \frac{m \binom{t}{k} \binom{n}{m}}{(m+k) \binom{t+n}{m+k}} \tag{7}$$

When k = 0, corresponding to the probability that in t future years the mth of n past floods will not be exceeded, the probability becomes

$$\phi_k = \frac{\binom{n}{m}}{\binom{t+n}{m}} \tag{8}$$

The important point here is that no prior probability density is assumed for the distribution of annual flood events, so that all of the probability statements are non-parametric. Apart from the theory which has been written about the recurrence interval and exceedance probability, the Thomas results show how unreasonable it is to attempt to deduce for purposes of design the 50-year flood on the basis of 10 years of real or equivalent record. In fact, because in any record less than 50 years, there is no rank m which can be used to approximate the 50-year flow, all that can be done is to consider a range of flow values without concern to their probability of recurrence and to ask for the confidence intervals in the manner suggested by Thomas.

This can form the basis of a design methodology; the exceedance probabilities for various ranks could be attached to economic losses, leading to new possibilities for combining data at gaged and ungaged sites. The method could not transfer <u>extreme</u> information, for which the sampling errors are large. The studied methods are directed at estimating Q_T, where T is large compared to the record length. But the Thomas method was developed for small floods, with no intent to analyse flows sufficiently large to be candidates for Q_{50}. The technique is more applicable to the design of small or temporary structures (e.g., cofferdams), which are to operate for a short time period and for which the consequences of a small overtopping are not significantly different from those of a large one.

The reason for discussing at length a potential design technique which seems to be disqualified because it deals only with minor floods rather than 50-year events is that this study shows that what hydrologists have typically regarded to be good estimates of the 50-year event are, in fact, estimates of much more common (or "minor") floods. Events usually taken to be $\hat{Q}_{50}$, for which satisfactory design decisions have historically been made, are much less extreme than anticipated. Thus a non-parametric technique such as the Thomas scheme might be utilized because existing techniques, currently employed with confidence and empirical success, are advertised to estimate Q_{50} but in fact do not do so by a wide margin. It might be appropriate, under a new research contract, seriously to consider whether specification of Q_{50} or any other Q_T is appropriate to define a design flow. This issue has been raised in an earlier context, where we deal with the specification both of confidence and tolerance limits for design flows.

Multinomial Logit Analysis

This is a form of multivariate analysis in which the dependent variable is divided into discrete classes rather than represented on a continuous scale, and from which the analysis gives the <u>probability</u>, p_i, that each of the discrete classes will be realized for a given set of independent variables. For example, a set of medical symptoms might represent disease i with probability p_i, where i ranges over a set of diseases for which the differential diagnosis is questionable. Another example is to let i range over a small set of possible meteorological episodes -- heavy rain, showers, no rain -- and to let the independent variables be a set of observations on the weather so that the probability p_i is the probability of rain, no rain, etc. Application to drainage design suggests that the independent variables might be all the relevant hydrologic information (for example, the moments of the annual floods), the basin characteristics, and some measure of economic assessment and risk aversion. The index i ranges over a small set of culvert design capacities, suggesting that a relatively small number of different designs might accommodate all the important cases. The p_i is then the probability that design capacity i is chosen, conditioned on the given combination of independent variables.

Consider a vector of variable values X_i which describes the state of a system. For example, X_1 is the mean annual flood, X_2 is the standard deviation, X_3 the skew coefficient, X_4 the regional correlation, X_5 the serial correlation, X_6 a measure of economic consequence, X_7 a measure of risk aversion, X_8 through X_{12} a group of basin characteristics, etc. The vector X defines all the inputs to a culvert design problem.

As the result of tabulating the X_i for many thousands of existing culverts, a large number seem to have the same X_i-values; but different designs have been selected. Suppose all designs could be lumped into three classes or groups: small, medium, and large. Of course, if three groups are too few, more could be added, but three are chosen for simplicity. The proportion of small culverts is p_1, of medium p_2, and of large p_3, where $p_1 + p_2 + p_3 = 1$.

Multinomial logit analysis enables the calculation of all the p_i from any combination of vector X_i.

If we examine all the <u>small</u> culverts, we note a failure rate (or probability) of π_1, and similarly for π_2 and π_3. If the sample is large enough, then $\pi_1 \geq \pi_2 \geq \pi_3$. The <u>design</u> problem is solved by calculating all p_i for any combination X_i and selecting that design (small, medium, or large) which meets failure criteria expressed by π.

Thus a new design technique, based on massive amounts of empirical data taken across a representative group of regions, could evolve. Collection of the requisite data base is recommended.

A New Hydrologic Framework

The so-called Rational Formula assumes no relationship between drainage area and runoff per unit area; the Meyer Formula, the Talbot Formula, and others set an arbitrary relationship. It is time to apply modern statistical theory to develop envelopes of discharge/area versus area for various C-values (100, 50, ...) and exceedance probabilities; there is now abundant background for making unbiased estimates of return intervals.

There is a wealth of hydrologic information contained in a hyetograph from which can be plotted on the abscissa the fraction of area (≤ 1) in the basin, and as the ordinate the fraction of maximal runoff at the outlet (≤ 1). Moments of this fraction or distribution (mean and standard deviation might suffice) connote a lot of information, and become arguments of a general runoff intensity function, $I = \phi(A, C(u), E(u), SD(u), m, n)$, where A is the drainage area, u the runoff ratio, C a runoff function, m the rank of the flood, and n the length of record. The properties of this function ϕ determine a design rule for drainage needs.

SUMMARY RECOMMENDATIONS

The decision as to which research program should be pursued is dependent upon budgetary and time constraints and therefore is one of public policy; we do not attempt to choose that policy. This study has recommended the WRC technique be modified by the USGS procedure to remove bias. Additionally, that the basin characteristics file of the USGS be updated to permit more extensive regional regression analysis. Pursuit of the research programs suggested in this section could progress in phases beginning with identification of data needs (hydrologic and economic) and time and cost requirements to collect and file the data. Such a program would be dynamic with continual evaluation and undoubtedly alterations. The existing gaging programs and design methodology (with recommended improvements) would continue until replaced by newly developed techniques.

REFERENCES

1. Aitchison, J., and Brown, J. A. C., The Log-Normal Distribution (Cambridge University Press, London), 1957.

2. ARMCO, Handbook of Drainage and Construction Products (Middletown, Ohio), 1958.

3. Bell, Frederick C., "Estimating Design Floods from Extreme Rainfall," Hydrology Paper No. 29, Colorado State University, July 1968.

4. Benson, Manuel A., "Factors Influencing the Occurrence of Floods in a Humid Region of Diverse Terrain," USGS, WSP, 1580-B, 1962.

5. Benson, Manuel A., "Factors Influencing the Occurrence of Floods in the Southwest," Ibid.

6. Benson, M. A. and Carter, R. W., "A National Study of the Streamflow Data-Collection Program," USGS, WSP, No. 2028, Washington, 1973.

7. Bock, Paul, et al., "Estimating Peak Runoff Rates from Ungaged Small Rural Watersheds," National Cooperative Highway Research Program Report 136, Highway Research Board NRC, NAS/NAE, 1972.

8. Bobée, Bernard, "The Log Pearson Type 3 Distribution and Its Application in Hydrology," WRR, 11: 5, October 1975.

9. Bobée, B., and Robitaille, R., "Correction of Bias in the Estimation of the Coefficient of Skewness," WRR, 11: 6, December 1975.

10. Federal Interagency Work Group, "Hydrologic Data Requirements for Small Watersheds," U.S. Department of the Interior, December 1973.

11. Fiering, Myron B, "Statistical Analysis of Streamflow Data," Ph.D. Dissertation, Harvard University, 1960.

12. Fiering, Myron B, "On the Use of Correlation to Augment Data," J. Amer. Stat. Assoc., 67: 1962.

13. Fiering, Myron B, "An Optimization Scheme for Gaging," WRR, 1: 4, 1965.

14. Fiering, Myron B, "Schemes for Handling Inconsistent Matrices," WRR, 4: 2, April 1968.

15. Fletcher, J. E., et al., "Runoff Estimates for Small Rural Watersheds and Development of a Sound Design Method," Utah State University, 1974.

16. Hardison, Clayton H., "Accuracy of Streamflow Characteristics," USGS Prof. Paper 650-D, 1969

17. Hardison, Clayton H., "Prediction Error of Regression Estimates of Streamflow Characteristics at Ungaged Sites," USGS Prof. Paper 750-C, 1971.

18. Hiemstra, Lourens, and Reich, Brian, "Engineering Judgment and Small Area Flood Peaks," Hydrology Paper No. 19, Colorado State University, April 1967.

19. Kirby, W., "Algebraic Boundedness of Sample Statistics," WRR, 10: 2, April 1974.

20. Maddock, Thomas III, "An Optimum Reduction of Gages to Meet Data Program Constraints," Bull. Hydrological Sciences, XIX: 3.

21. Matalas, Nicholas C., "Optimum Gaging Station Location," Proc. IBM Symposium on Water and Air Resource Management, IBM, Yorktown Heights, 1967.

22. Matalas, Nicholas, "A Mathematical Assessment of Synthetic Hydrology," Water Resource Research, 3: 4, 1967.

23. Matalas, Nicholas C., et al., "Regional Skew in Search of a Parent," WRR, 11: 6, December 1975.

24. Matalas, Nicholas C., private communication based on an unpublished study, USGS, 1976.

25. Matalas, Nicholas C., and Langbein, W. B., "The Relative Information of the Mean," JGR, 67: 9, 1962.

26. Moore, Donald, "Estimating Mean Runoff in Ungaged Semi-Arid Areas," State of Nevada, Department of Conservation and Natural Resources, Water Resources Bulletin No. 36, 1968.

27. Moss, Marshall E., and Karlinger, M. R., "Surface Water Network Design by Regression Analysis Simulation," WRR, 10: 3, June 1974.

28. Moss, Marshall E., "Design of Surface Water Data Networks for Regional Information (Technique Manual)," draft U.S. Geological Survey Memorandum, 1975.

29. State of New Jersey, "Magnitude and Frequency of Floods in New Jersey with Effects of Urbanization," Special Report 38, Department of Environmental Protection, with U.S. Geological Survey, 1974.

30. Potter, W. D., "Peak Rates of Runoff from Small Watersheds," Hydraulic Design Series No. 2, BPR, Washington, April 1961.

31. Raiffa, Howard, and Schlaiffer, Robert, Applied Statistical Decision Theory (Harvard University Press, Cambridge), 1961.

32. Shahane, Ashok, et al., "Role of Synthetic Time Series in Hydrometeorological Data Analysis," Water Resources Bulletin, paper 75039.

33. Slack, J., Wallis, James, and Matalas, Nicholas, "On the Value of Information to Flood Frequency Analysis," WRR, 11: 5, October 1975.

34. Texas Board of Water Engineers, "Texas Stream Gaging Program: Evaluation and Recommendations," with the U.S. Geological Survey, October 1960.

35. Thomas, D. M., "Planning Small-Streams Flood-Frequency Information Systems," USGS Memorandum, December 1969.

36. Thomas, Harold A., Jr., "Frequency of Minor Floods," JBSCE, 35, 1948.

37. Thomas, Harold A., Jr., Unpublished memorandum, Harvard Water Program, 1958.

38. U.S. Department of the Interior, Geological Survey, OWPC, "Hydrologic Data Needs for Small Watersheds," Reston, Virginia, December 1974.

39. U.S. Department of Transportation, FHWA, 1973 Highway Statistics, 1975.

40. U.S. Department of Transportation, "The 1974 National Highway Needs Report," Report of the Secretary of Transportation, House Document 94-95, 1975.

41. U.S. Water Resources Council, A Uniform Technique for Determining Flood Flow Frequencies, draft report, December 1974.

42. State of Virginia Highway Department, Design Standards.

43. Wallis, J. R., et al., "Just a Moment!," WRR, 10: 2, April 1974.

44. Winfrey, Robley, Economic Analysis for Highways (International Textbook Co., Scranton, Pennsylvania), 1969.

45. Winfrey, Robley and Fellner, Carl, "Summary and Evaluation of Economic Consequences of Highway Improvements," HRB, NRC, NAS/NAE, National Cooperative Highway Research Program Report 122, 1971.

46. Young, G. K., et al., "Evaluation of the Flood Risk Factor in the Design of Box Culverts," Vol. 1, Report, FHWA-RD-74-11, Federal Highway Administration, ORD, Washington, September 1970.

47. Young, George K., et al., "Optional Design for Highway Drainage Culverts," J. ASCE, Hydraulic Div. HYT, July 1974.

BIBLIOGRAPHY

Regional Flood Frequency Reports 1969 - August 1974

Alabama:

Hains, Charles F., 1973, Floods in Alabama, magnitude and frequency based on data through September 30, 1971: U.S. Geological Survey open-file report, Montgomery, Alabama.

McCain, J. F., 1973, Progress report on flood frequency synthesis for small streams in Alabama: Alabama Highway Department of Publications (in press).

Alaska:

Childers, Joseph M., 1970, Flood frequency in Alaska: U.S. Geological Survey open-file report, Anchorage, Alaska, 30 pp., 3 figures.

Arizona:

Aldridge, B. N. and Condes de la Torre, Al, 1970, Flood hydrology of small drainage areas in Arizona -- A progress report through June 1969: U.S. Geological Survey Administration Report, May 1970, Tucson, Arizona, 43 pp., 17 figures.

Arkansas:

Patterson, James L., 1971, Floods in Arkansas, magnitude and frequency characteristics through 1968: Arkansas Geological Commission, Water Resources Circular No. 11, Little Rock, Arkansas.

Colorado:

Gonzalez, D. D., and Ducret, G. L., Jr., 1971, Rainfall-runoff investigation in the Denver Metropolitan area, Colorado: U.S. Geological Survey open-file report, Denver, Colorado, September 1971.

Hedman, E. R., Moore, D. O., and Livingston, R. K., 1972, Selected streamflow characteristics as related to channel geometry of perennial streams in Colorado: U.S. Geological Survey open-file report, Denver, Colorado.

Colorado-Wyoming-Montana:

Hedman, E. R., and Kastner, W. M., 1974, Progress report on streamflow characteristics as related to channel geometry of streams in the Missouri River basin: U.S. Geological Survey open-file report, Lawrence, Kansas.

Georgia:

Golden, Harold G., 1973, Preliminary flood-frequency relations for small streams in Georgia: U.S. Geological Survey open-file report, Atlanta, Georgia, April 1973.

Idaho:

Thomas, C. A., Harenberg, W. A., Anderson, J. M., 1973, Magnitude and frequency of floods in small drainage basins in Idaho: U.S. Geological Survey Water Resources Inv. 7-73, National Technical Information Service, Springfield, Virginia 22151.

Illinois:

Carns, Jack M., 1973, Magnitude and frequency of floods in Illinois: State of Illinois Department of Transportation, Division of Water Resources Management, 1973.

Iowa:

Lara, Oscar, 1973, Floods in Iowa: Techniques manual for estimating their magnitude and frequency: State of Iowa Natural Resources Council, Bulletin No. 11, March 1973.

Maine:

Hayes, G. S., and Morrill, R. A., 1971, A preliminary report on small streams flood frequency in Maine: U.S. Geological Survey Administration Report, March 1971, Augusta, Maine, 10 pp., 1 figure.

Maryland:

Walker, Patrick N., 1971, Flow characteristics of Maryland streams: Maryland Geological Survey, Report of Investment, No. 16.

Massachusetts:

Johnson, Carl G., and Laraway, G. Alan, 1971, Flood magnitude and frequency of small Massachusetts streams, preliminary estimating relations: U.S. Geological Survey Administration Report, September 1971, Boston, Massachusetts, 26 pp., 3 figures.

Johnson, C. G., and Tasker, G. D., 1974, Progress report on flood magnitude and frequency of Massachusetts streams: U.S. Geological survey open-file report No. 74-131, March 1974, Boston, Massachusetts, 41 pp., 11 figures.

Minnesota-North Dakota:

Federal Interagency, 1971, Red River of the North regional flood analysis (Breckinridge to International Boundary): Federal Interagency Report, August 1971.

Mississippi:

Hudson, James W., 1970, Preliminary flood-frequency analysis on small streams in Mississippi: U.S. Geological Survey Administration Report, April 1970, Jackson, Mississippi, 14 pp., 7 figures.

Missouri:

Gann, E. E., 1971, Generalized flood-frequency estimates for urban areas in Missouri: U.S. Geological Survey open-file report, Rolla, Missouri, 18 pp., 3 figures.

Hauth, L. D., 1974, A technique for estimating the magnitude and frequency of Missouri floods: U.S. Geological Survey open-file report, Rolla, Missouri.

Hauth, L. D., 1974, Model synthesis in frequency analysis of Missouri floods: U.S. Geological Survey Circular (in press).

Nevada:

Moore, D. O., 1974, Estimating flood discharges in Nevada using channel-geometry measurements: Nevada Highway Department Hydrologic Report No. 1, Carson City, Nevada, 43 pp., 46 figures.

New Mexico:

Scott, Arthur G., 1971, Preliminary flood-frequency relations and summary of maximum discharges in New Mexico, A progress report: U.S. Geological Survey open-file report, June 1971, Albuquerque, New Mexico, 76 pp., 23 figures, 9 tables.

North Carolina:

Putnam, Arthur L., 1972, Effect of urban development on floods in the Piedmont Province of North Carolina: U.S. Geological Survey open-file report, Raleigh, North Carolina.

Ohio:

Webber, E. E., and Mayo, R. I., 1973, Flood magnitude and frequency on small streams in Ohio: U.S. Geological Survey Administration Report, Columbus, Ohio, April 1973.

Oklahoma:

Thomas, W. O., and Corley, R. K., 1974, Floodflows from small drainage areas in Oklahoma: Progress report and data compilation: U.S. Geological Survey open-file report, 1974.

Sauer, V. B., 1973, Flood characteristics of Oklahoma streams: U.S. Geological Survey Water-Resources Inv. Series 52-73. (NTIS).

Sauer, V. B., 1974, An approach to estimating flood frequency for urban areas in Oklahoma: Oklahoma City, Oklahoma, U.S. Geological Survey Water-Resources Inv. 23-74, 10 pp., 3 figures.

South Dakota:

Becker, L. D., 1974, A method for estimating the magnitude and frequency of floods in South Dakota: Huron, South Dakota, U.S. Geological Survey Water-Resources Inv. 36-74 (in press).

Texas:

Schroeder, E. E., 1973, Estimating the magnitude of peak discharges for selected flood frequencies on small streams in east Texas: U.S. Geological Survey Administration Report.

Utah:

Butler, Elmer, and Cruff, H. W., 1971, Floods of Utah, magnitude and frequency characteristics through 1969: U.S. Geological Survey open-file report, 1971.

Butler, Elmer, and Marsell, Ray E., 1972, Developing a State water plan, cloudburst floods in Utah, 1939-69: Utah Department Natural Resources, Division of Water Resources, Cooperative Inv. Report No. 11.

Vermont:

Johnson, Carl G., and Laraway, G. Alan, 1971, Flood magnitude and frequency of small Vermont streams, preliminary estimating relations: U.S. Geological Survey Administration Report, September 1971, Boston, Massachusetts, 24 pp., 3 figures.

Johnson, Carl G., and Tasker, Gary D., 1974, Flood magnitude and frequency of Vermont streams: Boston, Massachusetts, U.S. Geological Survey open-file report No. 74-130.

Virginia:

Miller, E. M., 1971 Virginia small streams program preliminary flood-frequency relations: U.S. Geological Survey open-file report, 1971, Richmond, Virginia, 28 pp., 13 figures.

West Virginia:

Frye, P. M., and Runner, G. S., 1969, Procedure for estimating magnitude and frequency of floods in West Virginia: West Virginia State Roads Commission design directive, 10 pp.

____________________, 1971, A preliminary report on small streams flood frequency in West Virginia: U.S. Geological Survey Administration Report: Charleston, West Virginia, 9 pp., 1 figure.

☆U.S. GOVERNMENT PRINTING OFFICE: 1977 728-329/1204 1-3

FHWA
R&D