NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM
REPORT **136**

ESTIMATING PEAK RUNOFF RATES FROM UNGAGED SMALL RURAL WATERSHEDS

REFER TO:	ACT	INF.	SIG.
MATERIALS ENGR.			
OFFICE MANAGER			
QUALITY CONTROL			
PROJ. RECDS. & SUMM.			
ASSOC. MATLS. ENGR. II			
PROJ. DEVELOPMENT			
SOILS & FOUND. ENGR.			
SOILS			
SOIL MECHANICS			
CHIEF GEOLOGIST			
DRILL CREWS			
ASSOC. MATLS. ENGR. I			
RESEARCH			
TESTING ENGR.			
CHEMISTRY			
ASPHALT			
STRUCTURES			
ASPHALT MIX			
AGGREGATES			
RECEIVING			
MAINTENANCE			
E. I. T.			

HIGHWAY RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM REPORT **136**

ESTIMATING PEAK RUNOFF RATES FROM UNGAGED SMALL RURAL WATERSHEDS

P. BOCK, I. ENGER, G. P. MALHOTRA,
AND D. A. CHISHOLM
THE TRAVELERS RESEARCH CORP.
HARTFORD, CONNECTICUT

RESEARCH SPONSORED BY THE AMERICAN ASSOCIATION
OF STATE HIGHWAY OFFICIALS IN COOPERATION
WITH THE FEDERAL HIGHWAY ADMINISTRATION

AREA OF INTEREST:
HIGHWAY DRAINAGE

HIGHWAY RESEARCH BOARD
DIVISION OF ENGINEERING NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING 1972

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Systematic, well-designed research provides the most effective approach to the solution of many problems facing highway administrators and engineers. Often, highway problems are of local interest and can best be studied by highway departments individually or in cooperation with their state universities and others. However, the accelerating growth of highway transportation develops increasingly complex problems of wide interest to highway authorities. These problems are best studied through a coordinated program of cooperative research.

In recognition of these needs, the highway administrators of the American Association of State Highway Officials initiated in 1962 an objective national highway research program employing modern scientific techniques. This program is supported on a continuing basis by funds from participating member states of the Association and it receives the full cooperation and support of the Federal Highway Administration, United States Department of Transportation.

The Highway Research Board of the National Academy of Sciences-National Research Council was requested by the Association to administer the research program because of the Board's recognized objectivity and understanding of modern research practices. The Board is uniquely suited for this purpose as: it maintains an extensive committee structure from which authorities on any highway transportation subject may be drawn; it possesses avenues of communications and cooperation with federal, state, and local governmental agencies, universities, and industry; its relationship to its parent organization, the National Academy of Sciences, a private, nonprofit institution, is an insurance of objectivity; it maintains a full-time research correlation staff of specialists in highway transportation matters to bring the findings of research directly to those who are in a position to use them.

The program is developed on the basis of research needs identified by chief administrators of the highway departments and by committees of AASHO. Each year, specific areas of research needs to be included in the program are proposed to the Academy and the Board by the American Association of State Highway Officials. Research projects to fulfill these needs are defined by the Board, and qualified research agencies are selected from those that have submitted proposals. Administration and surveillance of research contracts are responsibilities of the Academy and its Highway Research Board.

The needs for highway research are many, and the National Cooperative Highway Research Program can make significant contributions to the solution of highway transportation problems of mutual concern to many responsible groups. The program, however, is intended to complement rather than to substitute for or duplicate other highway research programs.

NCHRP Report 136

Project 15-4 FY '68
ISBN 0-309-02021-2
L. C. Catalog Card No. 72-9483

Price $4.60

This report is one of a series of reports issued from a continuing research program conducted under a three-way agreement entered into in June 1962 by and among the National Academy of Sciences-National Research Council, the American Association of State Highway Officials, and the Federal Highway Administration. Individual fiscal agreements are executed annually by the Academy-Research Council, the Federal Highway Administration, and participating state highway departments, members of the American Association of State Highway Officials.

The study reported herein was undertaken under the aegis of the National Academy of Sciences—National Research Council. The National Cooperative Highway Research Program, under which this study was made, is conducted by the Highway Research Board with the express approval of the Governing Board of the NRC. Such approval indicated that the Board considered that the problems studied in this program are of national significance; that solution of the problems requires scientific or technical competence, and that the resources of NRC are particularly suitable to the conduct of these studies. The institutional responsibilities of the NRC are discharged in the following manner: each specific problem, before it is accepted for study in the Program, is approved as appropriate for the NRC by the Program advisory committee and the Chairman of the Division of Engineering of the National Research Council.

The specific work to be performed in each problem area is defined by an advisory panel that then selects a research agency to do the work, monitors the work, and reviews the final reports. Members of the advisory panels are appointed by the Chairman of the Division of Engineering of the National Research Council. They are selected for their individual scholarly competence and judgment, with due consideration for the balance and breadth of disciplines.

Responsibility for the definition of this research project and for the publication of this report rests with the advisory panel. However, the opinions and conclusion expressed or implied are those of the research agency that performed the research, and are not necessarily those of the Highway Research Board, the National Research Council, the Federal Highway Administration, the American Association of State Highway Officials, nor the individual states participating in the Program.

Although reports in this category are not submitted for approval to the Academy membership nor to the Council, each report is reviewed and processed according to procedures established and monitored by the Academy's Report Review Committee. Such reviews are intended to determine, *inter alia,* whether the major questions and relevant points of view have been addressed, and whether the reported findings, conclusions and recommendations arose from the available data and information. Distribution of the report is permitted only after satisfactory completion of this review process.

Published reports of the

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

are available from:

Highway Research Board
National Academy of Sciences
2101 Constitution Avenue
Washington, D.C. 20418

(See last pages for list of published titles and prices)

FOREWORD

By Staff

Highway Research Board

A problem that continues to plague highway agencies is that of selecting optimum culvert sizes for waterways based on (1) estimates of the magnitude and frequency of peak flows from small rural watersheds (less than 25 square miles), (2) the relative cost of facilities necessary to accommodate the estimated flows, and (3) the possible effects of flows in excess of the estimates used for design purposes. Although this report does not resolve the problem, it does provide valuable insight into the difficulties associated with attempts to develop improved methods for estimating peak runoff rates of various return periods for small ungaged rural watersheds throughout the United States. It should be of considerable practical value to agencies that lack well-developed local or regional methods for predicting flood flows and frequencies. Hydrologists and researchers working in this problem area undoubtedly will find the report of interest and value.

A basic problem in designing highway bridges or culverts for stream crossings is the determination of the flow to be accommodated. This involves estimating the magnitude of peak flows or floods at various frequencies for the particular drainage area under consideration. For major stream crossings, such an estimate normally is made on the basis of hydrologic analysis of the drainage area and the stream, characteristics of the climate, and accumulated stream flow data. However, probably all small rural watersheds are ungaged. Thus, the engineer generally is nearly always required to estimate the design flow for small drainage areas on the basis of limited topographic and climatic data.

In the late 1940's, cooperative stream gaging programs were begun between state highway departments and the U.S. Geological Survey to collect runoff data from selected small rural watersheds. Other agencies also have been gathering information to obtain a better understanding of the phenomena involved in the generation of runoff from small drainage areas. NCHRP Project 15-4 was undertaken with the expectation that the data and experiences accumulated since about 1950 from the gaging programs would provide a basis for the development of improved practical methods for predicting flood flows for small ungaged rural watersheds. To accomplish project objectives, researchers of The Travelers Research Corporation (now The Center for the Environment and Man) conducted an analysis involving a stepwise multiple regression technique using predictor variables. The selection of effective predictors from a large set of possible choices was based on computed coefficients of correlation between the predictant (peak runoff) and each predictor.

When the study was initiated, it was anticipated that data would be available for about 1,000 small watersheds throughout the United States. However, about one-half of the original watersheds identified were eliminated from the study because of the short period for which hydrologic data were available and the lack of

adequate topographic information. After careful screening, the data sample finally selected consisted of 493 watersheds. All are 25 square miles or less in area, are rural in nature, are without significant pondage, have a minimum of 12 years of acceptable annual runoff records, and have adequate topographic, physiographic, and climatic information available. The accumulated data, consisting of 116 pieces of information for each of the 493 watersheds, have been compiled as the National Small Streams Data Inventory (NSSDI) and placed on magnetic tape (available from the Assistant Chief Hydrologist, Scientific Publications and Data Management, U.S. Geological Survey, Washington, D.C. 20242).

Practically all previous studies have suffered from a lack of adequate verification of the flood prediction method that was developed. It is well known that a prediction method that produces satisfactory results when tested on its own developmental data sample may fail when applied to other problems. For this reason, the analysis program was conducted by using data from 395 of the watersheds, and an independent sample of 98 watersheds was withheld for verification of the prediction equations. In addition, the independent sample was used to evaluate prediction methods currently being used by state highway departments.

As a result of the analysis, it was found that topographic characteristics of the basins have higher predictive capabilities for estimating peak runoffs than do hydrologic-climatic or physiographic variables. Three sets of prediction equations were finally selected as appearing to have a predictive capability for flood flows of various frequencies for the entire U. S. that was about equal to the predictive capability of the aggregate of the 31 state highway department methods when each state method was applied within its own state. It should be noted, however, that this study indicates that approximately two-thirds of highway department hydrologic predictions for small rural drainage basis in the contiguous U.S. may be in error by more than 25 percent, and that one in five probably is overestimated by a factor of 3.

The findings of this study indicate that presently used methods for estimating runoff from ungaged rural watersheds are unsatisfactory on a nationwide basis. Consequently, designers should make the best possible use of existing prediction methods, with full realization of the high probability of error, and give careful consideration to the increased cost of overdesign versus the possible consequences of an underestimation of peak flow. Short-range research efforts should be concentrated on the development of improved local and regional methods for estimating peak flows based on use of data collected during this study and obtained in future years as the data base is strengthened.

CONTENTS

ACKNOWLEDGMENTS

The research reported herein was performed under NCHRP Project 15-4 by The Travelers Research Corporation, with Paul Bock and Isadore Enger as co-principal investigators. They were assisted in the research and the writing of the report by G. P. Malhotra and D. A. Chisholm. Currently, Dr. Bock is Professor of Hydrology and Water Resources, University of Connecticut, Storrs, Conn.; Mr. Enger is Director, Mathematical Statistics Section, GeoMet, Inc., Rockville, Md.; Mr. Malhotra is Design Engineer, Beas Designs Organization, Nangal Township, Punjab, India; and Mr. Chisholm is Research Physicist, Airforce Cambridge Research Laboratory, Bedford, Mass.

Other Travelers personnel giving significant support included Patricia Bernier, Richard M. Davis, Richard H. Coté, Guy T. Merriman, and Charles J. Mulé.

The following Federal Government organizations contributed to various phases of the study: U.S. Geological Survey, ESSA Weather Bureau, Soil Conservation Service, Agricultural Research Service, Federal Highway Administration, Forest Service, and Office of Water Data Coordination. State-level agencies that contributed include some 31 state highway departments, together with the State Conservationists and the USGS District Engineers in nearly all of the contiguous states.

Many other research investigators and engineers, too numerous to name, contributed helpful suggestions at various phases of the study.

In particular, grateful acknowledgment is given for the advice of Professor Ven Te Chow, who acted as consultant throughout the study.

ESTIMATING PEAK RUNOFF RATES FROM UNGAGED SMALL RURAL WATERSHEDS

CHAPTER ONE

INTRODUCTION AND RESEARCH APPROACH

One of the classical hydrologic problems yet unsolved is that of estimating floods of various frequencies from ungaged small rural watersheds. This problem exists today because of lack of basic understanding of hydrologic phenomena and the lack of systematic observational data. The lack of understanding and data have inhibited both the development of concepts and the verification and improvement of existing methodologies.

Many design engineers and hydrologists consider present methods as inadequate for estimating peak flow rates from ungaged small rural drainage basins. As a result there is no generally accepted design method. The plethora of methods being used throughout the United States and within individual states has produced inconsistent estimates of magnitudes of floods of various frequencies.

The societal purpose of small drainage facilities such as highway culverts is to provide for the safety and convenience of the public in an economic manner. The sound hydrologic design of small drainage facilities affects commerce, industry, transportation, and practically every section of public and private engineering works.

In particular, the economic importance of highway drainage structures is becoming increasingly evident. Approximately $500 million is being spent annually for highway culverts and small bridges. This represents 15 percent of the total annual cost of interstate and state highways for construction and maintenance.

Runoff data from small watersheds have increased greatly in the last decade. In the late 1940's, cooperative stream gaging programs began between a few state highway departments and the U.S. Geological Survey. In fiscal 1968, the total funds for all cooperative USGS-state highway department programs amounted to some $1 million. This program includes some 42 states involving nearly 2,000 gaged watersheds. In addition, a few other states have sponsored similar studies through other funding. Thus, at the start of the study in 1967 the base of runoff data from small watersheds had grown to the point where meaningful studies on a nationwide basis could be expected.

In parallel with the growth of runoff data, considerable progress had been made in recent years in research methodologies of watershed modeling, in stochastic approaches to the generation of synthetic series of hydrological events, in statistical procedures for analyzing hydrological data, in the accumulation of hydrologically-related data (such as meteorological, physiographic, and geologic data), and in the development of computer capabilities. These developments also provided a basis for the initiation of comprehensive studies of peak flows from small watersheds.

OBJECTIVES

The over-all objective of this study was to develop a better method(s) for estimating the magnitude and frequency of runoff from small rural watersheds (approximately 20 square miles or less).

The estimation method(s) were required to adhere to the following constraints:

1. Require only data that can be readily obtained by the designers.
2. Use parameters and functional relationships that are logically justified.
3. Take cognizance of differences due to geographical characteristics.
4. Present the information desired in a readily usable form.

This over-all objective and the four constraints were formulated to adhere to the practical requirements of highway design engineers in the hydrologic design of small drainage structures, particularly culverts and small bridge openings.

It is emphasized that the objective of the study was to develop a practical prediction method for estimating peak runoff rates of various return periods for small ungaged rural watersheds throughout the United States. This objective suggested certain limitations to the range of research approaches. The pursuit of better physical understanding of hydrological phenomena *per se*, though a laudable objective, could only be a supporting objective. Approaches

aimed at development of strictly local relationships, or requiring considerable development of general research methodologies, or research aimed primarily at elucidating physical understanding of rainfall-runoff relationships that were also considered to offer only long-range promise for the development of a practical prediction method, were discarded as being not appropriate for the stated objective of this study. However, considerable effort was devoted to establishing a comprehensive rational basis for the development of a practical prediction method.

THE ENGINEERING DESIGN PROBLEM

The sizing of waterway openings for small highway drainage structures depends on a number of considerations, including hydrologic, structural, and economic factors; the amount of vehicular use of the road and the types of user traffic; the amount of expected future development of the area; past experience as translated into design factors; location in relation to nearby flood-sensitive areas (such as schools, railroad embankments, important road intersections or interchanges); damage potential; maintenance requirements; etc. Although all such factors are important to design, this study considers only the hydrologic factors that affect the estimation of magnitude and frequency of peak runoff from small rural watersheds. Urbanized watersheds also are economically important, but these fall outside the scope of the present study.

The hydrologic design problem most routinely faced by the highway design engineer is estimation of the peak rate of flow (cfs) of a given return period for ungaged drainage areas in which only easily obtainable standard information is available. To be widely useful, the design method for estimating peak flow must be relatively simple to apply. In the engineering office, the time required for the hydrologic design should be short; that is, less than one or two hours at the most for standard design cases. Once the design peak flow rate has been calculated, invert elevations, waterway openings, hydraulic grade lines, surcharge conditions, backwater curves, etc., can be determined as required.

For the design of most minor highway drainage structures, calculation of the entire hydrograph is not required; only the peak flow rate is computed. For special design situations where routing of various flows is required or where the effects of storage are to be assessed, the calculation of the whole hydrograph is required. Hydrograph synthesis, being a more specialized design problem, falls outside the scope of this study.

ORGANIZATION OF THE REPORT

The body of the report summarizes the main features of the study, including background information, research data and methodology, and findings and conclusions. Three new equations for estimating peak flows for ungaged small rural watersheds are presented. These equations are compared with methods presently used by state highway departments. An up-to-date bibliography is given.

Nine appendices are included. New information includes: 84 sets of flood-flow equations for small rural areas; a comprehensive analysis showing that the log-normal distribution is better than both the log-Pearson Type III and the Gumbel distribution for estimating the maximum annual peak runoffs from small rural basins; and computer-derived maps of 5-, 10-, 15-, and 30-min precipitation amounts for the 25-year frequency for the contiguous United States. (These maps are derived from the four appropriate sets of precipitation data, whereas *USWB Technical Paper No. 40* assumes a single constant relationship to the 60-min duration precipitation for each of the 5-, 10-, 15-, and 30-min durations.)

A separate volume containing computer listings of the "National Small Streams Data Inventory" (NSSDI) is available on a loan basis by contacting the Program Director, NCHRP. A description of the NSSDI is contained in Appendix E.

RESEARCH APPROACH

Simply stated, the over-all objective of this study was to develop a practical prediction method for estimating flood magnitudes of various frequencies from ungaged small rural basins (less than 25 sq mi) throughout the contiguous United States using readily available data. The rationale for selection and formulation of the research approach of this study was to use that approach believed to have the best likelihood of achieving the stated objective.

Hydrograph Synthesis

Several research approaches were considered. One important traditional approach involves synthesis of hydrographs for gaged areas. This requires development of rainfall-runoff relationships and subsequent translation of these relationships to ungaged areas using all available techniques, including statistical methods for generalizing many types of hydrologic variables, and use of judgment and empirical evidence. The hydrograph approach requires coordinated rainfall-runoff data (lacking for small rural basins on a national basis), estimation on a national basis of relevant precipitation and climatic characteristics that cause design-type floods, and translation of these relationships from gaged situations to ungaged situations. This approach seems most appropriate at the present state of knowledge for gaining understanding of hydrologic phenomena rather than for development of practical countrywide design procedures.

One of the best-known examples of hydrograph synthesis is the Stanford Watershed Model developed by Crawford and Linsley (*16, 17*) and adapted by Clarke (*15*) and Miller (*34*) to small watersheds. This mathematical simulation model of the land phase of the hydrologic cycle uses a moisture accounting procedure. The model provides a step-by-step accounting of precipitation, evaporation, interception and depression storage, soil moisture, ground water, subsurface flow, interflow, surface runoff, and stream flow. The computer program requires as input data hourly precipitation, average daily evaporation by 10-day periods, a time-area histogram used for channel routing, and values of the previously listed basin parameters, including four values describing initial moisture storage. The model syn-

thesizes a continuous hydrograph for gaged watersheds under specific input and initial conditions.

Clarke (*15*) applied the Stanford Watershed Model concepts to a single watershed in Kentucky to estimate flood peaks and to correlate the resulting runoff coefficient used in the Rational Equation ($Q = C\,I\,A$) to six arbitrarily selected basin characteristics. Miller (*34*) extended Clarke's study to 39 watersheds less than 20 sq mi in area with a minimum of 10 years stream flow data. Miller concluded: "Further studies are needed to verify this hypothesized relationship (between watershed characteristics and the Stanford Watershed Model parameters)."

The Stanford Watershed Model concept is an important research development, but its use is deemed premature at this time for developing practical design methods for ungaged rural areas on a national basis. The requirements for adequate data, particularly for rainfall and soil permeability by depth for all design situations throughout the country, pose serious difficulties in adapting its use for routine design.

Instantaneous Unit Hydrograph

The concept of the unit hydrograph proposed by Sherman (*48*) develops a hydrograph of direct surface runoff from a given basin due to a unit rainfall excess distributed uniformly over the entire basin for a duration less than the time of concentration. Many rainfall-runoff relationships based on the unit hydrograph concept have been developed. The method is most useful for areas gaged for both rainfall and runoff and depends on invariance and superposition principles and the evaluation of the "rainfall excess" parameter. Snyder (*51*) correlated basin variables with unit hydrograph variables of peak flow, basin lag, and the hydrograph time base.

In 1945 Clark (*14*) introduced the concept of the instantaneous unit hydrograph (IUH), a hypothetical unit hydrograph of unit rainfall excess and whose duration approaches zero as a limit. Other investigators, including Dooge (*21*), Nash (*37*), O'Donnell (*39*), Minshall (*35*), Gray (*25*), Singh (*49*), Blank and Delleur (*8*), and Schmer (*46*), provided theoretical, mathematical, and practical improvements and extensions to the IUH procedure.

Blank and Delleur (*8*) used a linear systems analysis technique to evaluate the kernel function within the convolution integral equation. Three theoretical examples, for which the kernels were known, were analyzed, resulting in an accurate reproduction of both the theoretical kernel and output functions (direct surface runoff hydrograph). The investigations are in the mathematical and conceptual development stage. Hydrologic, precipitation, and geomorphological data have been collected for 55 watersheds in Indiana ranging from about 2 to 300 sq mi. Future studies contemplate the collection of more data and the further development of design procedures.

The recent study by Schmer (*46*) using a linear convolution model for approximating the rainfall-runoff phenomena yielded good results for two small drainage basins in Texas. However, Schmer concludes, "The direct application of the proposed model is not feasible from an engineering design point of view until simple, inexpensive techniques have been developed for selecting the generalized transfer function." Other practical problems include lack of coordinated rainfall-runoff data on a national basis, correlation analysis of drainage basin and hydrologic-climatic characteristics, and sufficient verification of the IUH on independent data.

Hydrologic Engineering Design Approaches

Chow (*13*) has reported on a comprehensive summary of various methods for the hydrologic determination of waterway areas for the design of small drainage structures. More than 100 equations and methods are presented dating back to 1852. More than 50 variables of all kinds are included in the equations. Chow classifies the methods into nine categories—judgment, classification and diagnosis, empirical rules, formulas, tables and curves, direct observations, rational method, correlation analysis, and hydrograph synthesis. He also presents a summary of hydrologic design practices of 43 states as of 1962. Among the 43 states, the most widely used methods are: 58 percent use the Talbot method, 26 percent use U.S. Geological Survey methods, 23 percent use the rational method ($Q = C\,I\,A$), etc. Many states use several methods, depending on watershed size, degree of development, location, and other considerations. Some states apply several methods, choosing a design result based on subjective criteria. It is generally recognized that the Talbot method (1887) and the rational method (1889) lack an analytical basis and suffer from lack of developmental and verification data. As an example of the USGS method, the Bigwood-Thomas equation for Connecticut (*7*) is derived from statistical analysis of historical streamflow records, including the development of a mean annual flood value (MAF) for a given location which can be translated to flood flows of various frequencies. Regionalization techniques using basin and climatic variables are used to translate the results for use on ungaged areas. However, it is recognized that the central problem is that of translating from the gaged to the ungaged situation.

Some state highway departments have adopted the Bureau of Public Roads method (*42*), which requires computation of a topographic index (T), a rainfall index (P-index), and identification of the zone in which the watershed is located. Basically, the value of the 10-year flood is estimated and the magnitudes of other flood frequencies are found through use of curves related to the 10-year flood. States that use this method are Virginia, Pennsylvania, New York, Arkansas, Vermont, and Michigan. Other states, not containing any of the basins used in this study, also may use the BPR method.

Many state highway departments use methods that are strictly for local use. These methods include regression equations (curves developed from local data relating runoff with area, slope, vegetal cover, etc.), nomographs, precipitation indices, maps of runoff coefficients, etc.

The Soil Conservation Service (*53*) has developed a method for estimating peak flows for small farm-type basins (less than 2,000 acres and watershed slopes less than 30 percent). Peak discharge is related to drainage area, 24-hr rainfall amounts (from U.S. Weather Bureau

Tech. Paper No. 40), two types of rainfall time distributions, three categories of average watershed slopes, and watershed characteristics (land-use practices, hydrologic conditions, and hydrologic soil groups A, B, C, D). Charts are presented for easy application of the SCS method.

Statistical Approaches

In addition to the hydrograph synthesis methods, unit hydrograph methods, and locally derived methods, various statistical methods were considered for use in this study. These included stepwise multiple regression, principal component analysis, factor analysis, and synthetic hydrology. Diaz et al. (*20*) provides an evaluation of the various techniques:

> Opinions on the relative merits of normal multiple regression compared with multivariate analysis in studying hydrologic relationships differ with various investigators. Mustonen (*36*), in a study of the effects of climatic and basin characteristics on annual runoff, concluded that normal multiple regression is an appropriate method for studying hydrologic relationships and that, when studying general hydrologic laws, it is hardly worthwhile to use the more complicated multivariate methods. Matalas and Reiher (*32*) concluded that factor analysis is still largely undeveloped technically, and there are serious doubts as to its usefulness. However, the limitations of the normal multiple regression approach to water yield studies was pointed out in detail by Sharp et al. (*47*). Snyder (*51*) concluded that, for hydrologic studies, multivariate analysis offers the more satisfactory solution to the problem of estimating independent effects when the independent variables are correlated. He also suggested that component and factor analysis in hydrologic applications be investigated. Wallis (*58*), in a discussion of multivariate statistical methods in hydrology, recommends for multifactor hydrologic problems the use of principal component regression with varimax rotation of the factor weight matrix. The Tennessee Valley Authority (*55*) used factor analysis in the design of a hydrologic condition survey. Dawdy and Feth (*19*) used factor analysis in a study of ground water quality.

Wong (*62*) developed a multivariate statistical model for predicting mean annual floods in New England for basins of 10 to 2,000 sq mi using length of main stream and average land slope as predictor variables.

Lewis and Williams (*30*) applied multivariate statistical methods to five Montana watersheds for 50 runoff events using 29 independent variables. He concluded:

> The principal-component and rotated-factor regression equations for the peak discharge rate and runoff volume from the watersheds were not as consistent as expected. When variables were discarded for the rotated-factor regression equations, some relationships of dependent and independent variables became intuitively incorrect, especially when only six independent variables were used. In summary, it would seem that the equations for peak discharge rate and total runoff were theoretically accurate, but the sensibility of the coefficients was not as consistent as the literature had predicted.

Osborn and Lane (*40*), using simple linear regression models, found that peak rates of runoff for four small semi-arid watersheds (0.56 to 11.0 acres) were strongly correlated to 15-min depth of precipitation for short return period events. However, the authors state: "It is possible that these models will not accurately predict the low-frequency events."

In recent times, "synthetic or stochastic hydrology" has been used to develop a long hydrologic series from historical data based on the statistical parameters exhibited by the short sample. Benson and Matalas (*5*) recognize two major deficiencies—large errors due to sampling errors of the original sample, and the inability to generate a series for an ungaged basin. To adjust these deficiencies, in a study of the Potomac Basin, they used a multiple regression technique to derive relationships between monthly flows and physical and climatic variables of the basin. The six predictors used were: watershed area, mean annual precipitation (MAP), mean annual snowfall, stream slope, forested area (%), and area of lakes and ponds (%). The authors report: "The variables that are found to be related do not vary consistently from month to month, and the equations are therefore somewhat questionable."

The major problem faced by investigators attacking the classical hydrologic problem of flow estimation for ungaged basins is in the insufficient understanding of hydrologic phenomena. Hydrologists have not been able to develop general physical-mathematical equations that describe the causal relationships from input to a hydrologic system (such as a small rural drainage basin) to the output hydrograph. For example, Sittner et al. (*50*) wrote:

> In runoff analysis there is no rational technique for completely and accurately delineating the various flow components that together define the hydrograph. Further, the decision as to how many components to recognize is somewhat arbitrary.

The same views are held by Merva et al. (*33*):

> Hydrologists do not have at hand a general functional form relating watershed runoff to the climatic and physiographic parameters that must be used to describe the runoff process. Even more important, no analytical means of defining the proper parameters for a general functional form are available.

They also describe a probabalistic approach using the analogy between the path of water on a watershed surface and the phenomenon of Brownian motion. The work is in the theoretical stage of development.

The difficulty in relating runoff to watershed variables was stated by W. M. McMaster at a Technical Conference on Small Stream Flood Frequency attended by the U.S. Geological Survey personnel (*24*):

> In summary, it appears that a method for rigorous or even closely approximate evaluations of the effects of cover, soils, and geology on flood flow is not likely to be developed.

In addition to lack of basic hydrologic understanding of runoff relationships, practically all previous studies (including those reported here) have suffered from lack of verification of the flood-prediction equation using data that were not used in the development of the equation itself. It is well known that minimization techniques can produce optimum or near optimum results when the prediction equation is tested on its own developmental sample. However, verification using independent data is the only valid procedure for testing prediction equations for the general

case. Consequently, practically all the flood prediction equations, notably for small rural basins, lack adequate quantitative verification.

Stepwise Screening Regression Technique

Some writers in the hydrologic literature have implied that ordinary multiple regression analysis yields a valid prediction equation only if the predictors are truly independent. Independence in the statistical sense means that the simple correlation coefficients between predictors are zero. In hydrologic problems, it has been found that the predictor variables are usually correlated in varying degrees. Consequently, some hydrologic investigators have used multivariate techniques (such as principal component analysis) to overcome this difficulty even though this procedure has been demonstrated to reduce R^2 (the square of the multiple regression coefficient).

However, according to Brownlee (*64*, p. 422) there is no requirement in the regression analysis that predictors be independent. He states:

> We assume that η is a simple linear function of two "independent variables" x_1 and x_2:
>
> $$\eta = \alpha + \beta_1(x_1 - \overline{x_1}) + \beta_2(x_2 - \overline{x^2}).$$
>
> Although this is standard terminology for x_1 and x_2, it is misleading in that there is no requirement that x_1 and x_2 be independent in the statistical sense.

Stepwise multiple regression analysis using predictor variables regardless of independence is used in this study. The stepwise multiple regression technique (see Appendix G) is rapid and efficient and permits evaluation of the predictive capability of each predictor in a stepwise fashion until the point is reached where the addition of another predictor does not meet a selected significance level. The technique is flexible in that the matrix of predictors and the predictor forms can be easily changed. The stepwise multiple regression technique also serves as both the initial and final steps in the multiple regression analysis of nonlinear variables. It is recognized that physical interpretation cannot be imputed to the sign and magnitude of the regression coefficients. It is also recognized that although the regression equation serves as an efficient prediction equation of flood magnitudes (prediction is intended), it is not intended as a statement of hydrologic cause-effect relationships.

The selection of effective predictors from a large set of possible choices is conducted in an objective, stepwise manner. It is based on computed correlation coefficients between the predictand (peak runoff) and each of the predictors individually and in combination with the previously selected variables. The predictors can take on any number of numeric forms—arithmetic, logarithmic, binary, or nonlinear—either uniquely or in conjunction with other forms. Regardless of the form of the predictand and selected predictors, the predictand is expressed as a linear function of the selected predictors with the coefficients being determined by the method of least squares.

The accuracy of the estimation equations developed in this study and previously developed methods was evaluated by three statistical procedures: root-mean-square-error analysis, sign test comparison, and the frequency distribution of estimation errors. These and the regression analysis technique are described in detail in Appendix G.

Additionally, extreme-value statistical analysis techniques were utilized to compute return period values of peak runoff and short-duration precipitation from annual maximum observations. These techniques are discussed in Appendices B and C, respectively.

GENERAL OUTLINE OF SELECTED RESEARCH APPROACH

The research approach finally adopted in this study essentially followed along the lines of prediction schemes derived by meteorologists in attacking meteorological forecasting problems. For example, one meteorological problem is to forecast the ceiling heights at airports scattered throughout the northern hemisphere for periods of 1 hr, 2 hr, and 12 hr into the future, using routinely available meteorological data. The forecast technique must be readily usable by the weather forecaster in real time, be logically justifiable, and provide the forecaster with some measure of the reliability of the forecast. This meteorological forecast problem is analogous to the hydrologic problem of predicting flood flows of various frequencies.

The Data

The data used in this study are described in Appendices E (NSSDI) and F. Appendix F describes the criteria for selection of the study basins, defines the basin characteristics, the hydrologic-climatic and physiographic parameters, and describes the computer assembly of the data sample.

The NSSDI contains printed data concerning 116 pieces of information for each of the 493 watersheds used in this investigation. The specific variables are listed in four groupings: peak discharge, topographic, hydrologic-climatic, and physiographic.

Information about the magnetic tapes containing the complete data record for the 493 watersheds (and for another 179 stations with peak discharge records between 5 and 12 years) can be obtained by writing the Ass't. Chief Hydrologist, Scientific Publications and Data Management, U.S. Geological Survey, Washington, D.C. 20242.

In summary, the processed data sample consists of (a) peak discharge, (b) topographic parameters, (c) hydrological and climatic factors, and (d) physiographic (soil) parameters. The data samples were developed from 493 watersheds distributed across the United States (Fig. 1). These watersheds are of 25 sq mi or less, rural, and without any significant pondage, diversion, or regulation of the flow. Seventy-five percent of the runoff records comprised 15 years or more of acceptable annual values, no record was for less than 12 years, and the mean time of record was 18.3 years.

The analyses of peak runoff were developed from published USGS annual maximum peak discharge values. Topographic information was obtained from USGS quadrangle maps. Hydrologic and climatic data were drawn mainly from ESSA Weather Bureau sources. Physiographic information was obtained from SCS sources. Information

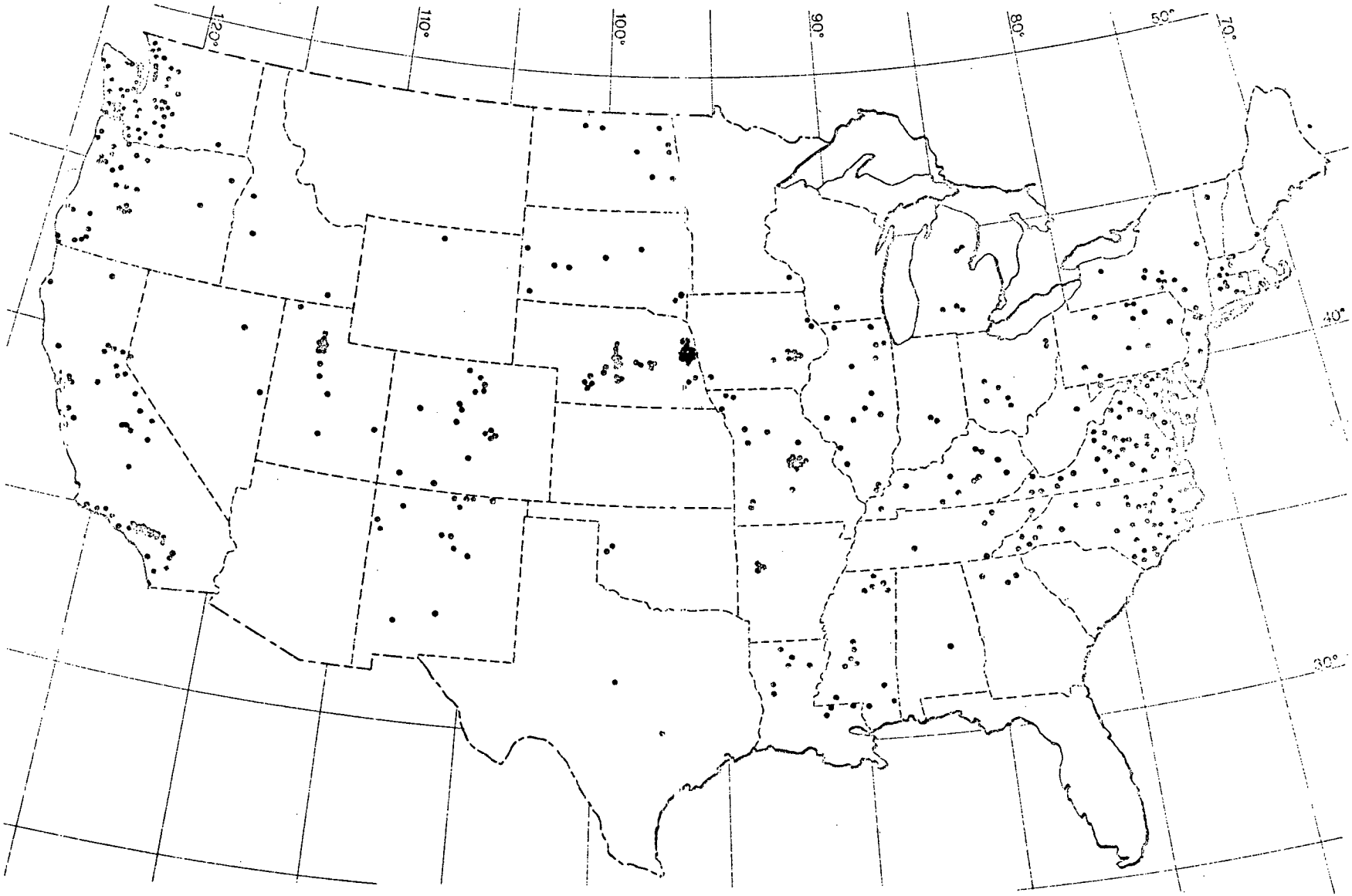


Figure 1. Geographic distribution of 493 watersheds used in the study.

obtained by questionnaires sent to USGS district engineers and SCS state conservationists supplemented and validated many of the published data.

Thus, the study is based on a data sample comprising more than 9,000 station-years of record from small rural watersheds throughout the United States. These data represent the largest number of basins, the largest number of hydrologic variables, the largest number of functional forms of the variables, and the broadest geographic coverage assembled to date concerning small rural drainage basins.

Framework of Experiments

A series of experiments (see Appendix H) was conducted to test various hypotheses leading toward the development of equations for estimating the magnitude and frequency of runoff from small rural watersheds. As used here, an experiment consists of applying a statistical technique to the data and interpreting the results based on hydrologic and statistical reasoning. The hypotheses tested fell into five general phases.

In Phase I, the predictor variables were divided into three basic types: topographic, hydrologic-climatic, and physiographic. The stepwise regression technique was applied to various sets of predictor variables to obtain equations expressing runoff as functions of only a restricted set of variables.

In Phase II, the form of the variables was investigated. The predictand variables, Q_5, Q_{10}, Q_{50}, were used in both their arithmetic form and in logarithmic form. The predictor variables were used in arithmetic, logarithmic, and nonlinear forms. Using the dependent sample of 395 watersheds, the stepwise regression technique was applied to develop estimation equations using various combinations of forms of the variables (e.g., both predictand and predictor in arithmetic form; log predictand and arithmetic predictors; log predictand and nonlinear predictors).

In Phase III, several methods of stratifying the data sample were formulated. The intent of stratification was to identify sets of hydrologically homogeneous drainage basins. Separate regression equations were developed for each subset. This approach differs from the first and second phases, wherein relationships were developed on a national basis using the full developmental sample of 395 watersheds to generate, for a given experiment, a single national equation.

In Phase IV, stratification factors were introduced into the development of national equations. Several of the stratification factors were considered in binary, or dummy variable, form as predictors, thus yielding a single national equation with terms denoting the stratification factors (e.g., MAF and region of the country).

Practicality was the dominant consideration in Phase V. All the possible predictors that had been considered in the earlier experimentation could be obtained, in practice, either through presently published sources or by means of information that could be developed from the data base, although in some cases considerable time would have to be expended to obtain them. From a design point of view, practicality and minimizing the time spent on determinations of design discharge must be taken into consideration. Therefore, a series of experiments was formulated and tested in which only predictors that are quickly and easily determined were considered.

A total of 24 hydrologic/statistical experiments (Table H-1) were performed to develop methods for estimating peak runoffs from small rural watersheds. These experiments were developed from as many as 223 total predictors (101 in arithmetic form, 47 in logarithmic form, and 75 in nonlinear form). These experiments produced 84 equations, of which 48 were single national equations and the remaining 36 were sets of stratified equations (see Appendix H).

The reduction of variance (R^2) for each prediction equation and for each selected predictor in each prediction equation was noted. Comparisons of errors of the various prediction equations with errors produced by design methods used by 31 state highway departments were made based on three verification schemes using an independent sample of 98 basins withheld for the purposes of verification of the prediction equations. The "error" is defined as the difference between the estimated peak flow and the peak flow based on a log-normal distribution of the historical record. The log-normal distribution was selected after exhaustive tests of goodness-of-fit in comparison with other normally used distributions (see Appendix B).

The three prediction equations finally recommended (see Chapter Two) consist of two sets of national equations with stratification factors and another set of equations for "hydrologically homogeneous" basins stratified according to indices of USGS mean annual flood. These three sets of new equations appear to possess about equal predictive capability among themselves for flood flows of various frequencies for the nation as a whole and to possess about the same predictive capability as the aggregate of the 31 state highway department design methods when each state method is applied within its own state.

TERMINOLOGY

For the convenience of the reader, the principal symbols used are listed in Appendix I; Table H-2 lists the 101 predictors (including 6 predictands) used in the stepwise regression; and Tables E-1, E-2, and E-3 list the symbols of the NSSDI.

CHAPTER TWO

FINDINGS

The results of the hydrologic-statistical experiments are contained in the 34 tables and the discussions in Appendix H. The main findings are summarized in the following.

The search for a set or sets of prediction equations for estimating peak runoff from small rural watersheds involved formulation and testing of a considerable number of hydrologically sound and statistically justifiable experiments. A total of 24 iterated detailed experiments were conducted that drew on the results of the preliminary stages to formulate the later stages. The experiments were organized into five phases to examine: (1) the predictive value of various sets of predictors; (2) the functional form of the variables; (3) the stratification of the watersheds into "homogeneous" (hydrologic, climatic, regional) subsets; (4) the utility of stratification factors in the framework of national equations; and (5) the degradation caused by utilizing easily obtained parameters in place of other parameters identified in the regression analysis.

In Phase I it was found that topographic variables, individually and as a set, have higher predictive value for peak runoff than have the climatic and physiographic (soil) variables. Among the topographic variables, the length of tributaries (TRIB), area (A), and stream slope (S_{10}) are the most important. Climatic variables as a set made significant independent increases over topographic variables in the reduction of variance of peak runoff (41.0 to 49.5 percent for the log Q_{25} relationship) whereas physiographic (soil-type) variables did not contribute.

Regarding the functional form of the variables tested in Phase II, an equation developed with a logarithmic predictand was retained by a process of elimination of other forms. A standard linear regression equation form (see Eq. G-1) was rejected because it can, on occasion, yield negative estimates of peak discharge due to the form of the predictand and the nature of regression. This problem is eliminated by using a logarithmic predictand. Application of a highly complex nonlinear regression technique to this problem did not yield significantly better results than the other simpler approaches; therefore, it was rejected due to its complexity not only in developing the relationships but also in their subsequent application to other data (ungaged watersheds).

Testing of nine alternative methods of stratifying into more "homogeneous" subsets and developing separate equations for each subset was conducted in Phase III. Of the nine alternatives considered, three were clearly better than the rest on the basis of root-mean-square error analysis (RMSE) and a frequency distribution of percent errors. They were stratifications based on five mean annual flood categories (MAF_5), three categories of mean annual temperature (MAT), and two categories of mean basin elevation ($\bar{E}$). Comparing these to the national equations retained from Phase II showed the stratified equations to be slightly better.

Several of the variables used to develop sets of stratification equations were then introduced as potential predictors of peak runoff in a national framework in Phase IV. It was found that (a) the stratification factors that were selected by the screening procedure increased the reduction of variance 5 to 8 percentage points over comparable equations not using them and (b) on a sign test comparison the national equation with stratification factors yields superior results to both the best methods of stratification and the best national equations not including stratification effects.

In Phase V two national equations limited to a few parameters that could be obtained quickly and easily by a design engineer were developed and compared to the best results achieved. Of the two, E-1 (which uses only A and $P_{10\text{-}360}$) compared poorly, whereas E-2 yielded just slightly poorer results than the best results achieved by the other equations. The comparatively high predictive capability of E-2 is probably due to the inclusion of geographic regional factors—REG_3 (Gulf), REG_2 (Midwest), and remainder of country—and the mean annual flood (MAF) categories.

THREE RECOMMENDED HYDROLOGIC DESIGN EQUATIONS

This study developed three sets of equations (D-3, E-2, C-1) for estimating peak flows for ungaged, small, rural basins of less than 25 sq mi that gave better results than the other 81 equations formulated and tested in this study.

These three recommended sets of equations are as follows:

Experiment D-3—(National Equation with Stratification Factors) Logarithmic Regression Equation

Q_{10} *Equation*

$$\widehat{(\log Q_{10})} = -2.54 + 0.77(\log \text{TRIB} + L) + (\log P_{10\text{-}360}) + 0.82(\log \text{MAF}_5) + 0.10(\text{REG}_2) - 0.66(\log \text{SHAPE}) + 2.01(\log \text{July } T) - 0.23(\text{REG}_3) \qquad \text{(AIV-19)}$$

$$\hat{Q}_{10} = 3.82 \times 10^{-3}(\text{TRIB} + L)^{0.77}(P_{10\text{-}360})^{1.14} (\text{MAF}_5)^{0.82}(\text{SHAPE})^{-0.66} (\text{July } T)^{2.01}\, 10^{0.10(\text{REG}_2)-0.23(\text{REG}_3)} \qquad \text{(AIV-20)}$$

Q_{25} *Equation*

$$\widehat{(\log Q_{25})} = -3.56 + 0.77(\log \text{TRIB} + L) + 1.18(\log P_{10-360}) + 0.79(\log \text{MAF}_5) + 0.13(\text{REG}_2) - 0.61(\log \text{SHAPE}) + 2.63(\log \text{July } T) - 0.28(\text{REG}_3) \quad \text{(AIV-21)}$$

$$\widehat{Q}_{25} = 3.69 \times 10^{-4}(\text{TRIB} + L)^{0.77}(\text{P}_{10-360})^{1.18} (\text{MAF}_5)^{0.79}(\text{SHAPE})^{-0.61} (\text{July } T)^{2.63}\, 10^{0.13(\text{REG}_2)-0.28(\text{REG}_3)} \quad \text{(AIV-22)}$$

Q_{50} *Equation*

$$\widehat{(\log Q_{50})} = -4.21 + 0.77(\log \text{TRIB} + L) + 1.20(\log P_{10-360}) + 0.78(\log \text{MAF}_5) + 0.15(\text{REG}_2) + 3.04(\log \text{July } T) - 0.31(\text{REG}_3) - 0.58(\log \text{SHAPE}) \quad \text{(AIV-23)}$$

$$\widehat{Q}_{50} = 8.60 \times 10^{-5}(\text{TRIB} + L)^{0.77}(P_{10-360})^{1.20} (\text{MAF}_5)^{0.78}(\text{July } T)^{3.04}(\text{SHAPE})^{-0.58}\, 10^{0.15(\text{REG}_2)-0.31(\text{REG}_3)} \quad \text{(AIV-24)}$$

Experiment E-2—(Simplified National Equation with Stratification Factors) Four or Six Predictor

Q_{10} *Equation*

$$\widehat{(\log Q_{10})} = -3.23 + 0.71(\log A) + 1.18(\log P_{10-360}) + 0.90(\log \text{MAF}_5) + 0.15(\text{REG}_2) + 2.41(\log \text{July } T) - 0.23(\text{REG}_3) \quad \text{(AV-11)}$$

$$\widehat{Q}_{10} = 7.95 \times 10^{-4}\, A^{0.71}(P_{10-360})^{1.18}(\text{MAF}_5)^{0.90} (\text{July } T)^{2.41} \times 10^{0.15(\text{REG}_2)-0.23(\text{REG}_3)} \quad \text{(AV-12)}$$

Q_{25} *Equation*

$$\widehat{(\log Q_{25})} = -4.26 + 0.70(\log A) + 1.21(\log P_{10-360}) + 0.87(\log \text{MAF}_5) + 0.19(\text{REG}_2) + 3.05(\log \text{July } T) - 0.28(\text{REG}_3) \quad \text{(AV-13)}$$

$$\widehat{Q}_{25} = 7.66 \times 10^{-5}\, A^{0.70}(P_{10-360})^{1.21}(\text{MAF}_5)^{0.87} (\text{July } T)^{3.05} \times 10^{0.19(\text{REG}_2)-0.28(\text{REG}_3)} \quad \text{(AV-14)}$$

Q_{50} *Equation*

$$\widehat{(\log Q_{50})} = -4.93 + 0.70(\log A) + 1.22(\log P_{10-360}) + 0.20(\text{REG}_2) + 0.86(\log \text{MAF}_5) + 3.46(\log \text{July } T) - 0.32(\text{REG}_3) \quad \text{(AV-15)}$$

$$\widehat{Q}_{50} = 1.74 \times 10^{-5}(A)^{0.70}(P_{10-360})^{1.22}(\text{MAF}_5)^{0.86} (\text{July } T)^{3.46} \times 10^{0.20(\text{REG}_2)-0.32(\text{REG}_3)} \quad \text{(AV-16)}$$

Experiment C-1—Stratification Based on Mean Annual Flood (MAF_5)

Q_{10} *Equation*

(1) $\text{MAF}_5 = 1 (0 \le \text{MAF} \le 10)$

$$\widehat{(\log Q_{10})} = -6.17 + 0.41(\log T) + 2.97(\log \text{MAT}) + 0.28(\log \text{TRIB}) + 1.72(\log \text{P days}) \quad \text{(AIII-1}a\text{)}$$

$$\widehat{Q}_{10} = 7.77 \times 10^{-7}(T)^{0.41}(\text{MAT})^{2.97}(\text{TRIB})^{0.28} (\text{P days})^{1.72} \quad \text{(AIII-2}a\text{)}$$

(2) $\text{MAF}_5 = 2 (10 < \text{MAF} \le 30$

$$\widehat{(\log Q_{10})} = -9.05 + 6.02(\log \text{July } T) + 0.36(\log \text{MAS}) + 0.20(\log \text{TRIB}) \quad \text{(AIII-3}a\text{)}$$

$$\widehat{Q}_{10} = 12.56 \times 10^{-10}(\text{July } T)^{6.02}(\text{MAS})^{0.36} (\text{TRIB})^{0.20} \quad \text{(AIII-4}a\text{)}$$

(3) $\text{MAF}_5 = 3 (30 < \text{MAF} \le 50)$

$$\widehat{(\log Q_{10})} = 2.12 + 0.82(\log A) \quad \text{(AIII-5}a\text{)}$$

$$\widehat{Q}_{10} = 168\, A^{0.82} \quad \text{(AIII-6}a\text{)}$$

(4) $\text{MAF}_5 = 4 (50 < \text{MAF} \le 90)$

$$\widehat{(\log Q_{10})} = -4.05 + 0.48(\log T) + 0.082(\log E) + 0.025(\log \text{TRIB}) + 3.52(\log \text{Pot ET}) + 0.48(\log A) + 0.26(\log \text{32 F days}) + 0.34(\log S_{10}) + 0.26(\log \text{T days}) \quad \text{(AIII-7}a\text{)}$$

$$\widehat{Q}_{10} = 0.000099(T)^{0.48}(E)^{0.082}(\text{TRIB})^{0.025} (\text{Pot ET})^{3.52}(A)^{0.48}(\text{32 F days})^{0.26} (S_{10})^{0.34}(\text{T days})^{0.26} \quad \text{(AIII-8}a\text{)}$$

(5) $\text{MAF}_5 = 5 (\text{MAF} > 90)$

$$\widehat{(\log Q_{10})} = 3.50 + 0.42(\log \text{TRIB}) - 0.42(\log S_{10}) \quad \text{(AIII-9}a\text{)}$$

$$\widehat{Q}_{10} = 3827.23(\text{TRIB})^{0.42}(S_{10})^{-0.42} \quad \text{(AIII-10}a\text{)}$$

Q_{25} *Equation*

(1) $\text{MAF}_5 = 1 (0 \le \text{MAF} \le 10)$

$$\widehat{(\log Q_{25})} = -0.61 + 0.82(\log T) + 2.13 (\log \text{MAT}) + 0.37(\log \text{TRIB}) - 0.85(\log L) \quad \text{(AIII-1}b\text{)}$$

$$\widehat{Q}_{25} = 0.30(T)^{0.82}(\text{MAT})^{2.13}(\text{TRIB})^{0.37} (L)^{-0.85} \quad \text{(AIII-2}b\text{)}$$

(2) $\text{MAF}_5 = 2 (10 < \text{MAF} \le 30)$

$$\widehat{(\log Q_{25})} = -10.16 + 6.68(\log \text{July } T) + 0.38(\log \text{MAS}) + 0.19(\log \text{TRIB}) \quad \text{(AIII-3}b\text{)}$$

$$\widehat{Q}_{25} = 1.06 \times 10^{-11}(\text{July } T)^{6.68}(\text{MAS})^{0.38} (\text{TRIB})^{0.19} \quad \text{(AIII-4}b\text{)}$$

(3) $\widehat{\text{MAF}_5} = 3(30 < \text{MAF} \leq 50)$

$$\widehat{(\log Q_{25})} = 1.75 + 0.87(\log A) + 0.75(\log \text{M24P}) \quad \text{(AIII-5}b\text{)}$$

$$\widehat{Q}_{25} = 77.39(A)^{0.87}(\text{M24P})^{0.75} \quad \text{(AIII-6}b\text{)}$$

(4) $\widehat{\text{MAF}_5} = 4(50 < \text{MAF} \leq 90)$

$$\widehat{(\log Q_{25})} = 5.34 + 0.024(\log T) - 0.016(\log \text{P days}) + 0.77(\log A) + 0.34(\log \text{T days}) \quad \text{(AIII-7}b\text{)}$$

$$\widehat{Q}_{25} = 250701.36(T)^{0.024}(\text{P days})^{-0.016}(A)^{0.77}(\text{T days})^{0.34} \quad \text{(AIII-8}b\text{)}$$

(5) $\widehat{\text{MAF}_5} = 5(\text{MAF} > 90)$

$$\widehat{(\log Q_{25})} = 4.16 + 0.77(\log \text{TRIB}) - 0.52(S_{10}) - 0.74(\log A) \quad \text{(AIII-9}b\text{)}$$

$$\widehat{Q}_{25} = 17939.01(\text{TRIB})^{0.77}(S_{10})^{-0.52}(A)^{-0.74} \quad \text{(AIII-10}b\text{)}$$

Q_{50} Equation

(1) $\widehat{\text{MAF}_5} = 1(0 \leq \text{MAF} \leq 10)$

$$\widehat{(\log Q_{50})} = -1.77 + 0.92(\log T) + 2.72(\log \text{MAT}) + 1.18(\log \text{DD}) - 1.35(\log \text{SHAPE}) \quad \text{(AIII-1}c\text{)}$$

$$\widehat{Q}_{50} = 2.04(T)^{0.92}(\text{MAT})^{2.72}(\text{DD})^{1.18}(\text{SHAPE})^{-1.35} \quad \text{(AIII-2}c\text{)}$$

(2) $\widehat{\text{MAF}_5} = 2(10 < \text{MAF} \leq 30)$

$$\widehat{(\log Q_{50})} = -11.55 + 7.48(\log \text{July } T) + 0.52(\log \text{MAS}) \quad \text{(AIII-3}c\text{)}$$

$$\widehat{Q}_{50} = 4.82 \times 10^{-10}(\text{July } T)^{7.48}(\text{MAS})^{0.52} \quad \text{(AIII-4}c\text{)}$$

(3) $\widehat{\text{MAF}_5} = 3(30 < \text{MAF} \leq 50)$

$$\widehat{(\log Q_{50})} = 1.81 + 0.89(\log A) + 0.83(\log \text{M24P}) \quad \text{(AIII-5}c\text{)}$$

$$\widehat{Q}_{50} = 91.73(A)^{0.89}(\text{M24P})^{0.83} \quad \text{(AIII-6}c\text{)}$$

(4) $\widehat{\text{MAF}_5} = 4(50 < \text{MAF} \leq 90)$

$$\widehat{(\log Q_{50})} = 5.84 + 0.0072(\log T) - 1.76(\log \text{P days}) + 0.78(\log A) + 0.35(\log \text{T days}) \quad \text{(AIII-7}c\text{)}$$

$$\widehat{Q}_{50} = 7.89 \times 10^{5}(T)^{0.0072}(\text{P days})^{-1.76}(A)^{0.78}(\text{T days})^{0.35} \quad \text{(AIII-8}c\text{)}$$

(5) $\widehat{\text{MAF}_5} = 5(\text{MAF} > 90)$

$$\widehat{(\log Q_{50})} = 4.35 + 0.82(\log \text{TRIB}) - 0.55(\log S_{10}) - 0.84(\log A) \quad \text{(AIII-9}c\text{)}$$

$$\widehat{Q}_{50} = 3.07 \times 10^{4}(\text{TRIB})^{0.82}(S_{10})^{-0.55}(A)^{-0.84} \quad \text{(AIII-10}c\text{)}$$

The previously listed equations give about equally accurate predictive capability. The advantages and disadvantages of each are given in Table 1. It should be noted that both D-3 and C-1 require the computation of lengths of tributaries, a time-consuming task. From a practical viewpoint E-2 is easiest to use because it contains only predictors that are easily computed. However, the choice among the three sets of equations ultimately lies with the designer.

Comparison of the three sets of design equations developed in this study with methods presently used by state highway departments indicates that the three equations (D-3, E-2, C-1) are better in seven states, worse in eight states, and inconclusive in another 16 states. Comparisons were not made in the remaining 17 states of the contiguous United States either because none of the 98 independent basins used for verification was located in these states or because their presently used design methods were not available to this study.

The eight states that produced significantly better estimates of runoff than any equation developed in this study are Connecticut, Kentucky, Louisiana, Mississippi, New Mexico, South Dakota, Utah, and Washington.

The seven states that produced significantly poorer estimates of runoff than the three best equations developed in this study are California, Idaho, Iowa, North Carolina, Oregon, Pennsylvania, and Virginia.

On the independent set of 98 basins, the equations resulted in prediction errors of less than 25 percent in approximately 28 percent of the basins for peak runoffs

TABLE 1

GENERAL COMPARISON OF THREE BEST EQUATIONS DEVELOPED IN THIS STUDY

D-3 (National equation with stratification factors)

1. Gives slightly better results than E-2 or C-1; namely, smaller RMSE and fewer errors greater than 200%.
2. For $\widehat{Q}_{10}$, $\widehat{Q}_{25}$, $\widehat{Q}_{50}$, the same five predictors are used in a systematic manner.
3. Uses a regional variable.
4. Requires computation of lengths of tributaries.

E-2 (Simplified national equation with stratification factors)

1. Uses simple, readily computed predictors.
2. Gives results slightly poorer than D-3 and C-1.
3. For $\widehat{Q}_{10}$, $\widehat{Q}_{25}$, $\widehat{Q}_{50}$, the same four predictors are used in a systematic manner.
4. Uses a regional variable.

C-1 Stratification based on mean annual flood (MAF_5)

1. Stratified into five sets based on MAF.
2. Gives results slightly poorer than D-3 and slightly better than E-2.
3. Requires calculation of lengths of tributaries.
4. Uses a total of 15 equations for five MAF categories and three peak flow return periods. Uses differing predictors in the MAF categories and the three return periods.
5. Does not use a regional variable.

of the 10-year return period. For 74 of the 98 basins, 35 percent of the errors using state highway department methods were less than 25 percent. These two error statistics are not statistically different.

It thus appears that approximately two-thirds of highway hydrologic designs in the contiguous U.S. are in error by more than 25 percent for rural drainage basins of less than 25 sq mi. Further evaluations are given in Chapter Three.

CHAPTER THREE

INTERPRETATION, APPRAISAL, APPLICATION

COMPARISON WITH STATE HIGHWAY DEPARTMENT PROCEDURES

The main emphasis in this chapter is to evaluate in practical terms the three recommended design equations given in Chapter Two. At the same time, hydrologic design procedures used by many state highway departments are appraised against the available data.

A wide variety of methods for the estimation of floods from small rural areas are currently in use throughout the country. In general, the states differentiate between rural and urban watersheds and the use of the "rational" method is usually restricted to small watersheds of predominantly urban nature.

Many states have adopted the Bureau of Public Roads method for rural application. It requires computation of a topographic index (T), a rainfall index (P-index), and the zone in which the watershed is located, as read from maps. The value of Q_{10} is estimated and curves are generally available to relate Q_{10} to other flood frequencies. States that use this method are Virginia, Pennsylvania, New York, Arkansas, Vermont, and Michigan. There may be other states (not used for comparison in this study) that also use this method.

Another common method makes use of USGS curves developed from magnitude and frequency studies, from which the value of the mean annual flood can be determined from curves for a given watershed area in square miles. Another set of curves relates the mean annual flood (2.33-year return period) with the floods of other frequencies. Such curves have been developed by USGS for almost all of the United States.

Other methods include regression equations; curves developed from local data relating runoff with area, slope, vegetal cover, etc.; nomograms; precipitation indices; etc.

To assess the potential value of the equations developed in this study in relation to the existing methods, the equations developed in this study were compared with the state highway department procedures. Drainage manuals and other literature were obtained from the 31 states listed in Table 2. Each state's design practices were applied only to watersheds within the state and only when the procedure was clearly applicable. The total number of watersheds for which state highway department estimates could be made was 377, which represents about 75 percent of the total sample. Seventy-four of these watersheds were part of the independent portion of the data sample, which totaled 98 cases. Table 3 gives the pertinent data for the 74 basins of the independent sample.

A sign test comparison (see Appendix G) was made of the state's estimation errors and errors made by two of the equation sets (D-3 and C-1) developed in this study. The comparisons with D-3 and C-1 were made for the 377 drainage basins in the 31 states for Q_{10} estimates only.

On the basis of the sign test comparison, states that were better than the D-3 equation at the 95 percent confidence level based on the binomial test were Connecticut, Mississippi, Kentucky, South Dakota, New Mexico, Washington, and Utah; Connecticut, Louisiana, Kentucky, South Dakota, Washington, and Utah were similarly better than the C-1 equations.

On the same basis, the D-3 equation was better than the state methods in Pennsylvania, Virginia, Iowa, and

TABLE 2

STATES FROM WHICH HIGHWAY DEPARTMENT DRAINAGE MANUALS WERE OBTAINED

1. Alabama	16. Nebraska
2. Arkansas	17. New Mexico
3. California	18. New York
4. Colorado	19. North Carolina
5. Connecticut	20. North Dakota
6. Georgia	21. Oregon
7. Idaho	22. Pennsylvania
8. Illinois	23. South Dakota
9. Iowa	24. Tennessee
10. Kentucky	25. Texas
11. Louisiana	26. Utah
12. Massachusetts	27. Vermont
13. Michigan	28. Virginia
14. Minnesota	29. Washington
15. Mississippi	30. West Virginia
	31. Wisconsin

TABLE 3

COMPARISON OF Q_{10}, Q_{25}, AND Q_{50} VALUES FOR 74 INDEPENDENT WATERSHEDS BASED ON VARIOUS METHODS OF CALCULATION

WATERSHED			LOG-NORMAL ANAL.			STATE HWY. DEPT.			C-1 EQUATIONS [a]			D-3 EQUATIONS [b]		
STATE	USGS GAGE NO.	AREA (SQ MI)	Q_{10}	Q_{25}	Q_{50}	Q_{10}	Q_{25}	Q_{50}	Q_{10}	Q_{25}	Q_{50}	Q_{10}	Q_{25}	Q_{50}
Connecticut	1–1880.00	4.12	593	802	951	760	—	1400	645	1167	1429	1056	1556	2053
West Virginia	3–0525.00	14.5	1346	1714	2005	2650	3400	3950	2477	3054	5350	2531	3634	4714
Mississippi	2–4856.50	5.85	2931	3947	4784	3255	4092	4743	6368	10744	14752	3971	5776	7550
	7–2682.00	9.09	4907	7304	9448	3030	3838	4545	3956	6545	9018	1770	2529	3261
Idaho	13–2005.00	15.80	152	213	264	640	1074	1545	344	377	352	138	183	223
	10–1225.00	13.00	132	173	207	166	246	306	1252	1920	2635	678	1001	1322
Nebraska	6–7693.00	5.19	742	1455	2248	290	404	—	714	1163	1617	1313	2233	3190
	6–7778.00	2.04	365	926	1576	172	239	—	331	515	701	596	1042	1513
	6–7893.00	21.10	3138	7114	12076	805	1119	—	2146	3514	5036	2748	4739	6830
	6–8397.00	.72	234	368	476	343	477	—	62	152	267	111	195	282
	6–6078.00	4.08	1635	2106	2480	1729	2404	—	3145	10153	6954	2728	4675	6738
	6–8064.40	10.00	10062	22767	38602	3960	5505	—	4004	6980	9049	3836	6423	9096
	6–8104.00	.76	747	1014	1205	898	1248	—	258	429	507	488	830	1184
	6–6088.00	6.50	3992	6872	9763	2244	3120	—	3671	9846	8149	4233	7168	10266
Kentucky	3–2890.00	24.00	1550	1957	2274	2360	2840	3220	2515	4767	6822	3490	5785	8124
	3–3135.00	7.47	2011	2290	2491	1578	1817	2056	2732	2820	3994	918	1268	1598
North Carolina	3–4575.00	14.40	1251	1509	1703	2467	3538	4655	1652	1892	2588	2459	3236	3978
	2– 872.40	.25	132	175	209	49	71	93	458	766	1117	136	243	361
	2– 920.20	24.00	2153	3448	4675	1359	1949	2565	2178	3032	3025	1238	1950	2652
	2–1033.90	7.56	267	339	395	490	702	924	760	905	985	519	848	1188
	2–1335.90	4.66	166	222	267	164	236	310	678	807	1059	408	667	935
	2–1086.30	10.00	1189	1902	2577	529	758	998	1303	1714	2081	886	1434	2002
Virginia	2– 765.00	9.20	890	1190	1435	1850	2300	2750	2640	2866	3474	1644	2437	3224
	2– 156.00	11.30	1301	2059	2772	1364	1700	1900	714	1447	1958	991	1759	2286
	3–1686.00	.61	173	295	415	160	200	240	378	336	435	315	478	642
Oregon	14– 505.00	16.5	131	151	165	329	471	597	155	209	264	281	367	443
	14–1849.00	.89	108	132	150	379	544	689	158	229	303	125	169	210
	14–3121.00	2.42	251	290	318	600	860	1090	381	546	740	219	296	367
	14–3702.00	3.16	388	489	569	933	1338	1694	397	598	813	304	448	588
Louisiana	7–3663.80	.36	30	42	52	320	418	490	360	360	497	206	303	397
	7–3776.50	.73	250	429	582	823	1007	1193	419	808	861	420	625	828
Iowa	5–4537.00	1.95	534	821	1083	630	720	900	574	1028	1256	1828	1377	1938
Utah	0–1435.00	3.15	25	35	43	42	75	117	98	98	114	68	101	133
	0–1700.00	21.70	92	116	134	376	669	1043	459	478	580	299	426	548
Washington	2–1161.00	.19	102	128	149	21	26	30	29	47	43	75	96	115
	2–1570.00	15.40	240	276	302	218	272	313	560	1549	1445	939	1201	983
	4–2120.00	18.30	1413	1716	1945	1376	1548	1806	1211	1811	2342	1993	2651	3273
	2– 505.00	11.20	446	586	700	320	400	460	969	1194	1319	907	1195	1466
	2– 655.00	1.51	199	245	280	205	256	294	149	249	167	198	254	305
	2–1022.00	2.5	138	164	182	144	180	207	231	337	197	199	263	321
	2– 167.00	2.05	261	314	353	218	265	281	211	229	292	346	449	546
	2– 427.00	2.03	706	819	900	1243	1554	1787	88	196	179	331	430	523
	2–1072.00	2.17	133	175	209	96	120	138	351	332	433	230	309	382
	4–1252.00	4.10	352	431	492	384	432	504	556	571	773	235	312	382
	2–1355.00	8.31	1852	2324	2692	2048	2560	2944	652	939	862	655	855	1040
	2–2007.00	2.58	65	81	93	48	58	61	198	383	215	362	497	625
	4–2311.00	2.29	138	167	188	136	184	248	364	531	720	289	397	498
	4–2481.00	1.13	159	198	227	176	198	231	153	143	213	221	292	359
	2– 105.00	16.40	2380	2721	2967	1760	2200	2530	1303	1758	1965	882	1179	1452
Alabama	2–4100.00	5.10	1270	1645	1943	1170	1320	1450	1671	3023	3569	1632	2275	2893
Massachusetts	1–1740.00	3.39	268	344	403	520	—	940	996	1557	2165	457	669	885
Michigan	4–1410.00	1.20	237	412	588	202	260	310	736	1153	1606	181	290	399
Colorado	7–1005.00	3.41	22	31	39	680	—	—	112	137	155	121	184	246

WATERSHED			LOG-NORMAL ANAL.			STATE HWY. DEPT.			C-1 EQUATIONS [a]			D-3 EQUATIONS [b]		
STATE	USGS GAGE NO.	AREA (SQ MI)	Q_{10}	Q_{25}	Q_{50}	Q_{10}	Q_{25}	Q_{50}	Q_{10}	Q_{25}	Q_{50}	Q_{10}	Q_{25}	Q_{50}
New Mexico	7–2010.00	14.40	2202	3742	5272	1170	1665	2160	1172	1339	2562	923	1397	1870
	8–2535.00	2.50	17	21	25	68	91	106	33	73	74	75	103	130
	8–3177.00	1.50	1155	1899	2619	1053	1404	1872	519	781	1060	180	281	383
Tennessee	3–6005.00	17.50	2620	3266	3767	2360	3210	4040	1727	934	1532	1790	2693	3633
Pennsylvania	1–5525.00	23.80	3811	5086	6130	2750	3700	4100	3356	4353	5837	2412	3590	4754
South Dakota	6–4416.50	14.60	2402	5514	9434	829	1333	1705	198	244	212	437	731	1025
	6–4788.00	14.80	589	1219	1952	676	1046	1397	667	1274	1989	1072	1869	2008
New York	1–5080.00	3.12	361	465	548	610	790	900	527	629	647	535	764	986
	1–4155.00	14.10	1575	2061	2452	3066	4494	5782	1780	2285	2490	1620	2295	2946
	1–3280.00	14.70	1381	1714	1937	839	1251	1677	1696	3172	2869	1224	1741	2235
California	11– 565.00	3.23	75	154	230	1200	1418	1637	12	60	119	177	277	376
	11– 670.00	4.59	660	1212	1794	900	1125	1350	645	1371	1953	1246	1884	2526
	11– 860.00	7.24	445	1057	1850	1934	2285	2637	939	2040	2937	1639	2471	3306
	11–1000.00	9.71	1410	2671	4037	2934	3467	4000	1195	2636	3819	3332	4999	6694
	11–1825.00	5.89	1316	2324	3357	1108	1293	1477	1521	2643	3199	585	818	1040
	10–2580.00	16.70	901	2200	3918	3650	5070	6287	67	315	300	413	666	925
	10–2818.00	18.20	115	153	185	2100	2625	3150	2003	2419	3342	638	818	984
	11–3090.00	20.60	3025	4746	6349	2334	2771	3209	3403	4221	7444	3360	5093	6840
	11–4400.00	22.10	1711	2856	3976	3326	3717	4109	2350	4217	6076	1863	2771	3670
	10–3435.00	10.8	476	792	1101	2250	3000	3500	33	73	74	75	103	129
	11–1284.00	11.5	2890	5043	7228	2333	2800	3150	1373	3055	4443	1351	1863	2360

[a] Mean annual flood stratification (MAF_5). [b] National equation with stratification factors.

Idaho and the C-1 equations were better in Pennsylvania, Virginia, North Carolina, Iowa, Oregon, and California.

Of the eight states that produced better results than the D-3 or C-1 equations, seven used the USGS method of mean annual flood (or some variant) either as the sole method or as one of several methods employed in highway drainage design.

Of the seven states that produced poorer results than the D-3 or C-1 equations, the following methods were used: Bureau of Public Roads, Rational Method ($Q = C\,I\,A$), USGS MAF method, and various state-derived procedures.

Of the 31 states that could be tested, statistically significant comparisons could not be made between 16 states and D-3 or C-1 equations because of either an insufficient number of basins or lack of significant difference in the errors. These 16 states are Vermont, Massachusetts, New York, Georgia, Alabama, Arkansas, Tennessee, West Virginia, Michigan, Wisconsin, Minnesota, Illinois, North Dakota, Nebraska, Texas, and Colorado.

A test was made to evaluate the characteristics of the prediction errors in the state highway department methods for estimating Q_{10}. Only those states having 10 or more basins of the total sample of basins were used for this comparison to provide a fair basis for evaluation. The results are summarized in Table 4. The aggregate of the prediction errors in the 292 basins showed state methods overestimate 148 times and underestimate 144 times. This indicates that for these states as a whole the number of over- and underestimations tends to balance out.

However, individual states have biased methods. California and Oregon have a significant number of overestimations, whereas Nebraska, North Carolina, and Washington have a significant number of underestimations.

Considering the magnitude and sign of the mean prediction error (%), three states have sizable mean overestimations: California, +78 percent; Oregon, +58 percent; and Virginia, +144 percent. Only Nebraska, of the 11 states tested, appears to underestimate by a wide margin, an average of −42 percent. The other states underestimated by about 25 percent or less.

Miller (*34*) applied the Kentucky highway design procedures to 39 gaged watersheds in and near Kentucky and compared these with the results of frequency analysis of historical stream gage records. He reported that the methods consistently underestimated the flood peak. Although Table 4 shows that the mean prediction error for the 11 Kentucky basins compared in this study is −26.78 percent; the number of underestimations (6) and overestimations (5) is about the same.

Table 5 gives the frequency distribution of prediction errors for estimating Q_{10} using the state highway department methods on the researchers' independent sample. Seventy-four of the 98 basins could be compared. The remaining 24 basins could not be compared because either there were no independent basins in the 31 states whose drainage manuals were available to the study or drainage manuals were not available for those states that did contain basins in the independent sample. Although these basins are from this study's independent sample, there is no guarantee that some of the 74 basins were not part of the developmental sample used in formulating individual state highway department methods. This would introduce

TABLE 4

PREDICTION ERRORS IN ESTIMATING Q_{10} USING STATE HIGHWAY DEPARTMENT METHODS [a]

STATE	NO. OF BASINS	MEAN ERROR (%)	ESTIMATIONS (NO.) OVER	UNDER
California	54	+ 78.05	44	10
Kentucky	11	− 26.78	5	6
Mississippi	13	− 16.35	6	7
Nebraska	38	− 42.63	11	27
New Mexico	14	− 13.79	8	6
North Carolina	31	− 13.40	9	22
Oregon	29	+ 58.39	26	3
Pennsylvania	10	− 25.46	3	7
Utah	14	− 2.15	7	7
Virginia	10	+144.21	6	4
Washington	68	− 10.62	23	45
Total	292		148	144

[a] Includes only states containing 10 or more basins (either in the dependent or independent sample of the total 493 basins).

TABLE 5

FREQUENCY DISTRIBUTION OF PREDICTION ERRORS FOR ESTIMATING Q_{10} USING STATE METHODS ON INDEPENDENT [a] DATA (74 BASINS)

ERROR RANGE (%)	NO. OF ERRORS
−100 to − 50.1	8
− 50 to − 25.1	9
− 25 to − 10.1	6
− 10 to − 0	8
0 to + 10	5
+ 10.1 to + 25	7
+ 25.1 to + 50	4
+ 50.1 to +100	10
+100.1 to +200	1
>200	16
Total	74

[a] The basins selected were all from this study's independent sample of 98 basins withheld for the purpose of verification. There is no guarantee that some of the 74 basins are not part of the development sample used in developing individual state highway department methods.

an unknown favorable bias to the distribution of prediction errors for the state methods.

Table 5 indicates that about 35 percent of the errors fall within ±25 percent, about 38 percent fall within the range of +25 to +100 percent and −25 to −100 percent, and 21 percent are greater than +200 percent. This sample indicates that about one-third of the designs are of reasonable accuracy (±25 percent), whereas one in five is overdesigned by a factor of 3.

Similar comparisons of the D-3 and C-1 prediction errors for the 98 independent basins show that 28 percent of the errors fall within ±25 percent, about 45 percent fall within the range of +25 to +100 percent and −25 to −100 percent, and 16 percent are greater than +200 percent.

It seems, therefore, that the D-3 and C-1 prediction equations compare favorably with the individual state methods on the whole, being better in some states, worse in some, and essentially the same in the rest.

EXAMPLE OF APPLICATION OF DESIGN EQUATIONS

Of the 84 sets of equations developed in this study (Appendix A), 3 sets (D-3, E-2, C-1) were found to be superior and are regrouped in Chapter Two. A typical design problem is given in the following.

Illustrative Problem

Using the D-3, E-2, and C-1 equations, estimate the 10-year peak flood (Q_{10}) for the hydrologic design of a highway culvert draining an 11.60-sq-mi drainage basin located at Lat. 41.18.10 Long. 72.31.00 in Connecticut. The basin is predominantly rural, without significant man-made structures such as diversions, impoundments, or sewers, and only a small percentage of the total area is in lakes, ponds, and marshes. Other pertinent characteristics of the basin are given as required for the application of each equation.

National Equation with Stratification Factors (D-3)

This $\hat{Q}_{10}$ equation is Eq. AIV-20. Equations for $\hat{Q}_5$, $\hat{Q}_{25}$, and $\hat{Q}_{50}$ given as Eqs. AIV-18, AIV-22, and AIV-24.

$$\hat{Q}_{10} = 3.82 \times 10^{-3} (\text{TRIB} + L)^{0.77} (P_{10\text{-}360})^{1.14} (\text{MAF}_5)^{0.82} \times (\text{SHAPE})^{0.66} (\text{July } T)^{2.01} \times 10^{0.10(\text{Reg}_2) - 0.23(\text{Reg}_1)} \quad \text{(AIV-20)}$$

The values of the predictors in this equation are obtained as follows:

1. (TRIB) = 21.55 mi. This is the length of all tributaries measured to the nearest 0.1 mile using the blue lines on the topographic map (USGS quadrangle sheets 1:62,500 or 1:24,000). The tributary lines are not extended to the watershed boundary. The actual graphical measuring procedure for (TRIB) can be time consuming for basins with numerous tributaries.
2. (L) = 10.80 mi. This is the length of the main stem measured to the nearest 0.1 mile using the blue lines and dotted blue lines (on USGS quadrangle sheets 1:62,500 or 1:24,000) extended to the watershed boundary in accordance with the contour pattern and presence or absence of a headwater spring, lake, or marsh area.
3. (TRIB + L) = 32.35 mi = Item 1 + Item 2.
4. ($P_{10\text{-}360}$) = 3.50 in. The value of the 10-year 6-hr rainfall amount is directly read for the Connecticut location from Weather Bureau *Tech. Paper No. 40,* U.S. Department of Commerce, "Rainfall Frequency Atlas of the United States" (1961) (*61*). Use Chart 32, p. 71.

5. (MAF_5) $= 2$. The value of the Mean Annual Flood (MAF_5) is directly found for the Connecticut location from Fig. H-1.

6. (SHAPE) $= \frac{L}{2}\sqrt{\frac{\pi}{A}} = 280$. To compute the value of the shape factor, use (L) = 10.80 mi. The area (A) is planimetered from the map to the nearest 0.01 sq mi. (A) = 11.60 sq. mi.

7. (July T) = 70°F. The value of the mean July temperature is read to the nearest whole °F for the Connecticut location from *Climatic Maps of the United States* (*22*, p. 13).

8. (REG_1) $= 1$.

9. (REG_2) $= 0$.

From Figure H-8, determine the Region for the drainage basin under design. Assign the numerical value 1 to the REG predictor where the drainage basin is located and 0 to all other REG predictors. To solve for drainage basins in Connecticut, $REG_1 = 1$ and $REG_2 = 0$. To solve for drainage basins in Illinois, $REG_1 = 0$ and $REG_2 = 1$. For basins in other than Regions 1 and 2, $REG_1 = 0$ and $REG_2 = 0$.

The values of the predictors are substituted in Eq. AIV-20 and the equation is solved by slide rule:

$\widehat{Q}_{10} = 3.82 \times 10^{-3} (32.35)^{0.77} (3.50)^{1.14} (2)^{0.82} (2.80)^{-0.66} (70)^{2.01} (10)^{0} = 1{,}010$ cfs. This is the 10-year peak flood computed by the National Equation with Stratification Factors (D-3).

Simplified National Equation with Stratification Factors (E-2)

This $\widehat{Q}_{10}$ equation is Eq. AV-12.

$$\widehat{Q}_{10} = 7.95 \times 10^{-4} (A)^{0.71} (P_{10\text{-}360})^{1.18} (MAF_5)^{0.90} (\text{July } T)^{2.41}\, 10^{0.15\,(REG_2) - 0.23\,(REG_3)} \qquad \text{(AV-12)}$$

The values of the predictors in this equation are readily computed or directly read from maps and atlases. Unlike the D-3 equation, the E-2 equation does not contain (TRIB), which is usually time consuming to extract graphically from maps. Each of the predictors (A), ($P_{10\text{-}360}$), (MAF_5), (July T), (REG_2) and (REG_3) is explained in the previous section. Substitution of values in Eq. AV-12, and computation by slide rule gives

$$\widehat{Q}_{10} = 7.95 \times 10^{-4} (11.60)^{0.71} (3.50)^{1.18} (2)^{0.90} (70)^{2.41} \times 10^{0} = 1{,}040 \text{ cfs.}$$

Stratification Based on Mean Annual Flood (MAF_5) (C-1)

There are five equations for $\widehat{Q}_{10}$ based on (MAF_5): Equations AIII-2*a*, AIII-4*a*, AIII-6*a*, AIII-8*a*, and AIII-10*a*. The choice of the particular equation to use depends on the magnitude of the mean annual flood, grouped into five categories mapped as Figure H-1. For the Connecticut watershed, the value of (MAF_5) = 2 is directly read from Figure H-1. Therefore, the equation to use is AIII-4*a*, which corresponds to (MAF_5) = 2:

$$\widehat{Q}_{10} = 12.56 \times 10^{-10} (\text{July } T)^{6.02} (MAS)^{0.36} (TRIB)^{0.20} \qquad \text{(AIII-4}a\text{)}$$

In this equation (July T) = 70°F and (TRIB) = 21.55 mi, as in the preceding section on D-3. (MAS) = 50 in. The value of the mean annual snow to the nearest inch is found from *Climatic Maps of the United States* (*22*, p. 53). Substitution in Eq. AIII-4*a* and solution by slide rule gives $Q_{10} = 12.56 \times 10^{-10} (70)^{6.02} (50)^{0.36} (21.55)^{0.20} = 1{,}190$ cfs.

For comparison purposes, the method used by the Connecticut State Highway Department (*7*) was computed for the same 11.60-sq-mi drainage basin. The computed value of the 10-year flood peak (Q_{10}) is 1,050 cfs.

In a similar manner, the magnitudes of the 25-year flood and the 50-year flood can be computed. Notice however that in certain equations additional predictors are required. The main sources for obtaining values of the predictors are: The relevant USGS quadrangle maps at 1:62,500 or 1:24,000; *Climatic Atlas of the United States* (*22*); *Rainfall Frequency Atlas of the United States* (*61*); and *Water Atlas of the United States* (*63*) for (MAT) Plate 2, (T-days) Plate 4. In addition, the reader should consult Figures H-1 and H-8 herein.

CHAPTER FOUR

CONCLUSIONS AND SUGGESTED RESEARCH

The main research conclusions of the study are summarized in this chapter. The experimental findings are summarized and the three recommended hydrological design equations are given in Chapter Two. The latter are of more immediate interest to the design engineer.

1. Only 493 drainage basins in the United States were identified that met all selection criteria. To obtain even as many as 493 basins required that runoff records as short as 12 years be used.
2. The data base of 493 basins was insufficient to permit comprehensive testing of hydrologic relationships for "hydrologically homogeneous" stratifications of the basins.
3. The logarithmic form of the predictand (peak runoff for a given return period) should be used rather than the arithmetic form.
4. The topographic characteristics of the basins were shown to have higher predictive capabilities for estimating peak runoffs than hydrologic-climatic or physiographic variables.
5. Some widely used hydrologic parameters were found to contribute insignificantly to prediction of peak runoffs. These included precipitation parameters (obtained on a climatological basis), soil types, length of main stream, and stream slope.
6. The basins were stratified into groups in nine different "hydrologically homogeneous" ways. The resulting stratified equations did not improve the predictive capability over equations developed using the entire set of basins distributed over the United States.
7. Logarithmic transformations of the predictor variables improved the prediction equations. A nonlinear transformation (using a rational fraction form of first, second, and third degree) of the predictor variables did not significantly improve the equations.
8. Among the equations developed in this study, a single national equation employing stratification factors (D-3) yielded the best over-all result.
9. Easily determined variables may be substituted for some complicated variables with little reduction of predictive capability.
10. On a national basis, the over-all results of the study showed no major improvement in methods for estimating peak runoffs; however, for certain states, improvements over their methods could be demonstrated.
11. The study also produced the following new sources of research information:

    (a) An extensive computer compilation on magnetic tape of hydrologic/climatic/physiographic/topographic information on 493 basins plus limited information on another 179 stations.

    (b) 30 maps of precipitation-duration-frequency for the contiguous United States (for durations of 5, 10, 15, 30, and 60 min and for return periods of 2, 5, 10, 25, 50, and 100 years) developed by computer techniques using a large amount of previously unanalyzed ESSA Weather Bureau data.

RECOMMENDED ADDITIONAL RESEARCH

It is believed that considerably more basic understanding of hydrologic phenomena is required to systematically improve hydrologic design techniques and to guide selection of predictor variables for use in flood-flow equations for ungaged small rural watersheds. Research to gain such basic understanding should be encouraged. This causal knowledge provides the foundation for practical use by the design engineer.

Further efforts are suggested in the following areas:

1. Enlargement of the data sample of 493 basins throughout the United States is highly desirable. It is estimated that within a few years the data from another several hundred small rural basins will become available for study. Such a data base will provide greater flexibility and scope to studies using a variety of hydrologic/statistical analysis techniques, including the multivariate methods, principal component analysis, factor analysis, synthetic hydrology, etc. More kinds of stratification and more divisions within each stratification become possible as the data base increases. Findings of the present study are likely to be enlarged.
2. A considerable body of statistics has been generated concerning prediction errors of flood-flow equations used by some 31 state highway departments. Further study should be undertaken to determine the relationships, if any, between the characteristics of the various equations, the hydrologic/climatic/topographic/physiological features of the individual states—and the magnitude and sign of the prediction errors. It may be possible to identify common features of the "good" methods under particular hydrologic situations. Such information could provide immediate guidelines for improving (or discarding) present design methods.
3. Further in-depth studies of selected regions should be pursued, taking into account what are believed to be important "local effects." The stepwise multiple regression technique is a powerful predictive tool when the appropriate predictors are offered for screening. The region (stratification) should encompass a data base sufficiently large that the results can be tested statistically on an independent sample.
4. The results of the over-all study have been put into the form of equations for estimating flood magnitudes. In the interest of time saving in the engineering design office, these equations should be converted into design nomo-

graphs or curves. Such a task, although relatively simple, would ease the computational burden and tend to promote use of these results.

5. Predictors associated with short-duration precipitation amounts (5, 15, 30, 60 min) intuitively are closely related to peak runoffs from small drainage basins. Appendix C describes a computerized objective analysis procedure (CRAM) for developing maps from unevenly spaced points such as Weather Bureau stations. The study produced 30 maps of precipitation-intensity-duration frequency for durations of 5, 10, 15, 30 and 60 min for return periods of 2, 5, 10, 25, 50, and 100 years. These maps should be further analyzed, taking into consideration factors including: local climatic and topographic features, local anomalies that may cause irregular spacing or excessive curvatures in the isohyetals, and other additional data. Such rigorously derived maps for short-duration precipitation are likely to be useful to the design engineer concerned with areas having short hydrologic response times.

6. Fitting distribution functions to annual peak runoffs is essential in the development of flood-peak equations. Appendix B shows that the log-normal distribution fits better than the log-Pearson Type III and the Gumbel distribution for small rural watersheds grouped for the entire U.S. Considerable additional work could be done, particularly on the geographical variations of skewness and possible combinations of data from several basins to obtain more stable estimates of skewness coefficients. Skewness should be investigated for various climatological stratifications when sufficient data become available to warrant such an evaluation.

7. The general approach that studies individual storm events using coordinated rainfall-runoff measurements from individual basins should be further investigated. Such an approach promises eventual understanding of the rainfall-runoff processes; however, lack of coordinated measurements covering a sufficient range of conditions is likely to limit its practical value on a nationwide basis.

REFERENCES

1. Agriculture Research Service, U.S. Department of Agriculture, "Monthly Precipitation and Runoff for Small Agricultural Watersheds in the United States." (June 1957).
2. Agriculture Research Service, U.S. Department of Agriculture, "Hydrological Data for Experimental Watersheds in the United States." *Misc. Publ. No. 945* (1956-59), *No. 994* (1960-61), *No. 1070* (1962).
3. Benson, M. A., "Factors Influencing the Occurrence of Floods in a Humid Region of Diverse Terrain." *U.S. Geol. Survey Water Supply Paper 1580-B* (1962).
4. Benson, M. A., "Factors Affecting the Occurrence of Floods in the Southwest." *U.S. Geol. Survey Water Supply Paper 1580-D* (1964).
5. Benson, M. A., and Matalas, N. C., "Synthetic Hydrology Based on Regional Statistical Parameters." *Water Resources Res.*, Vol. 3, No. 4 (1967).
6. Benson, M. A., "Evolution of Methods for Evaluating the Occurrence of Floods." *U.S. Geol. Water Supply Paper 1580-A* (1962).
7. Bigwood, B. L., and Thomas, M. P., "A Flood-Flow Formulation for Connecticut." *U.S. Geol. Survey Circ. 365*, Washington, D.C. (1955).
8. Blank, D., and Delleur, J. W., "A Program for Estimating Runoff from Indiana Watersheds. Part 1. Linear System Analysis in Surface Hydrology and Its Application to Indiana Watersheds." *Tech. Rep. No. 4*, Water Resources Research Center, Purdue University (1968).
9. Bock, P., Enger, I., Malhotra, G. P., and Chisholm, D. A., "Development of Methods for Estimating Runoff Rates from Small Rural Watersheds: A Pilot Study." The Travelers Research Center (1969).
10. Bryan, J. G., "Fitting Regression and Probability Curves by Rational Fractions and Hyperbolic Tangents." Unpubl., The Travelers Research Center (Apr. 1969).
11. Chow, V. T. (chairman), "Report of the Committee on Runoff, 1955-56." *Trans. Amer. Geophysical Union*, Vol. 38, No. 3 (June 1967).
12. Chow, V. T. (ed.-in-chief), *Handbook of Applied Hydrology*. McGraw Hill (1964).
13. Chow, V. T., "Hydrologic Determination of Waterway Areas for the Design of Drainage Structures in Small Drainage Basins." *Eng. Exper. Station Bull. 462*, University of Illinois (1962).
14. Clark, C. O., "Storage and the Unit Hydrograph." *Trans. ASCE*, Vol. 110 (1945).
15. Clarke, K. D., "Applications of Stanford Watershed Model Concepts to Predict Flood Peaks for Small Drainage Areas." Div. of Research, Kentucky Dept. of Highways (1968).
16. Crawford, N. H., and Linsley, R. K., "The Synthesis of Continuous Streamflow Hydrographs on a Digital Computer." *Tech. Rep. No. 12*, Dept. of Civil Engineering, Stanford University (1962).
17. Crawford, N. H., and Linsley, R. K., "Digital Simulation in Hydrology: Stanford Watershed Model."

Tech. Rep. No. 39, Dept. of Civil Engineering, Stanford University (1966).

18. DALRYMPLE, T. (ed.), "Flood-Frequency Analysis." Manual of Hydrology, Pt. 3, Flood-Flow Techniques, *U.S. Geological Survey Water Supply Paper 1543-A* (1960).
19. DAWDY, D. R., and FETH, J. H., "Applications of Factor Analysis in Study of Chemistry of Ground-Water Quality, Mojave River Valley, California." *Water Resources Res.,* Vol. 3 (1967).
20. DIAZ, G., SEWELL, J. I., and SHELTON, C. H., "An Application of Principal Component Analysis and Factor Analysis on the Study of Water Yield." *Water Resources Res.,* Vol. 4, No. 2 (Apr. 1968).
21. DOOGE, J. C. I., "A General Theory of the Unit Hydrograph." *Jour. Geophys. Res.,* Vol. 64, No. 1 (1959).
22. Environmental Science Services Administration, "Climatic Maps of the United States." U.S. Dept. of Commerce (1966).
23. U.S. Geological Survey, *Proceedings of the Technical Conference on Small Stream Flood Frequency.* Southeastern Districts of the Water Resources Division, USGS (Sept. 1965).
24. U.S. Geological Survey, "Magnitude and Frequency of Floods in the Conterminous United States." *Water Supply Papers Nos. 1671-1689* (1964 and later).
25. GRAY, D. M., "Synthetic Unit Hydrographs for Small Watersheds." *Jour. Hydraulics Div.,* ASCE, Vol. 87 (1961).
26. GUMBEL, E. J., "Statistics of Extremes." Columbia Univ. Press, p. 228 (1958).
27. HARTER, H. L., "Percentage Points of the Pearson Type III Distribution." *Technomet.,* Vol. 11, No. 1, pp. 177-187 (1969).
28. IRZA, T. J., "Preliminary Flood-Frequency Relations for Small Streams in Kansas." Open file report, U.S. Geol. Survey (1966).
29. KENT, K. M., "A Method for Estimating Volume and Rate of Runoff in Small Watersheds." *SCS-TP-149,* Soil Conservation Service, U.S. Dept. of Agriculture (Jan. 1968).
30. LEWIS, G. L., and WILLIAMS, T. T., "Selected Multivariate Statistical Methods Applied to Runoff Data from Montana Watersheds." Dept. of Civil Engineering and Mechanics, Montana State Univ. (Aug. 1968).
31. LYON, T. L., BUCKMAN, H. O., and BRADY, N. C., "Nature and Properties of Soils." Macmillan (1952).
32. MATALAS, N. C., and REIHER, B. J., "Some Comments on the Use of Factor Analysis," *Water Resources Res.,* Vol. 3 (1967).
33. MERVA, G. E., BRAZEE, R. D., SCHWAB, G. O., and CURRY, R. B., "A Proposed Mechanism for Investigation of Surface Runoff from Small Watersheds. 1. Development." *Water Resources Res.,* Vol. 5, No. 1 (Feb. 1969).
34. MILLER, C. F., "Evaluation of Runoff Coefficients from Small Natural Drainage Areas." *Res. Rep. No. 14,* Univ. of Kentucky Water Resources Inst. (1968).
35. MINSHALL, N. E., "Predicting Storm Runoff on Small Experimental Watersheds." *Jour. of Hydraulics Div.,* ASCE, Vol. 86, No. 4 (HY 8), pp. 28-33 (1960).
36. MUSTONEN, S. E., "Effects of Climatological and Basin Characteristics on Annual Runoff." *Water Resources Res.,* Vol. 3 (1967).
37. NASH, J. E., "Systematic Determination of Unit Hydrograph Parameters." *Jour. Geophys. Res.,* Vol. 64, No. 1 (1959).
38. National Bureau of Standards, *Tables of Probability Functions.* Vol. 2, Federal Works Agency (1942).
39. O'DONNELL, T., "Instantaneous Unit Hydrograph Derivation by Harmonic Analysis." *Internat. Assn. of Scientific Hydrol.,* Vol. 51, pp. 546-557 (1960).
40. OSBORN, H. B., and LANE, L., "Precipitation-Runoff Relations for Very Small Semiarid Rangeland Watersheds." *Water Resources Res.,* Vol. 5, No. 2 (Apr. 1969).
41. OSTBY, F. P., PANDOLFO, J. P., VEIGAS, K. W., and SPIEGLER, D. B., "Ballistic Wind Study." *Tech. Rep. ECOM-01377-F,* Contract No. DA-28-043-AMC-01377(E), The Travelers Research Center (1966).
42. POTTER, W. D., "Peak Rates of Runoff from Small Watersheds." *Hydraulic Design Series, No. 2,* Div. of Hydraulic Res., Bur. of Pub. Roads (1961).
43. POTTER, W. D., STOVICEK, F. K., and WOO, D. C., "Flood Frequency and Channel Cross-Section of a Small Natural Stream." *Bull. Internat. Assn. of Scient. Hydrol.* (1968).
44. REICH, B. M., "Predicting High Flows from Small Watersheds." *Info. Rep. No. 55,* Inst. for Research on Land and Water Resources, The Pennsylvania State Univ. (June 1968).
45. REICH, B. M., "Design Hydrographs for Very Small Watersheds from Rainfall." *Rep. CER 62 BMR 41,* Dept. of Civil Engineering, Colorado State Univ. (July 1962).
46. SCHMER, F. A., "Investigation of a Linear Model to Describe Hydrologic Phenomenon of Drainage Basins." *Tech. Rep. 19,* Water Resources Inst., Texas A & M Univ. (Dec. 1969).
47. SHARP, A. L., GIBBS, A. E., OWEN, W. J., and HARRIS, B., "Application of Multiple Regression Approach in Evaluating Parameters Affecting Water Yield of River Basins." *Jour. Geophys. Res.,* Vol. 65 (1960).
48. SHERMAN, L. K., "Stream Flow from Rainfall by the Unit Graph Method." *Eng. News-Rec.,* Vol. 108, pp. 501-505 (1932).
49. SINGH, K. P., "A Non-Linear Approach to the Instantaneous Unit Hydrograph Theory." Ph.D. dissertation, Civil Engineering Dept., Univ. of Illinois (1962).
50. SITTNER, W. T., SCHAUSS, C. E., and MONRO, J. C., "Continuous Hydrograph Synthesis with an API-Type Hydrologic Model." *Water Resources Res.,* Vol. 5, No. 5 (Oct. 1969).
51. SNYDER, W. M., "Some Possibilities for Multivariate Analysis in Hydrologic Studies." *Jour. Geophys. Res.,* Vol. 67 (1962).
52. Soil Conservation Service, *National Engineering Handbook.* Sect. 4, Hydrology (1964).

53. Soil Conservation Service, U.S. Dept. of Agriculture, "Patterns of Soil Orders and Suborders of the United States" (1967).
54. Subcommittee on Hydrology, Inter-Agency Committee on Water Resources, "Methods of Flow Frequency Analysis." Supt. of Documents, Govt. Printing Office, Washington, D.C. (Apr. 1966).
55. Tennessee Valley Authority, "Design of a Hydrologic Condition Survey Using Factor Analysis." *Res. Paper No. 53,* Hydraulic Data Branch, Tennessee Valley Authority (Nov. 1970).
56. THOMASELL, A., HARRIS, R., and WELSH, J., "Studies of Techniques for the Analysis and Prediction of Temperature in the Ocean. Pt. III. Automated Analysis and Prediction." *Tech. Rep. 7421-213,* The Travelers Research Center (1966).
57. THORNTHWAITE, C. W., Associates Laboratory of Climatology, "Average Climatic Water Balance Data of the Continents. Pt. VII. United States." (1964).
58. WALLIS, J. R., "Multi-Variate Statistical Methods in Hydrology—A Comparison Using Data of Known Functional Relationship." *Water Resources Res.,* Vol. 1 (1965).
59. Water Resources Council: Hydrology Committee, "A Uniform Technique for Determining Flood Flow Frequencies." *Bull. No. 15,* Water Resources Council, Washington, D.C. (1967).
60. Weather Bureau, U.S. Department of Commerce, "Climatological Data." For various states (relevant years for each watershed).
61. Weather Bureau, U.S. Department of Commerce, "Rainfall Frequency Atlas of the United States." *Tech. Paper No. 40* (1961).
62. WONG, S. T., "A Multivariate Statistical Model for Predicting Mean Annual Flood in New England." *Ann. Assn. Amer. Geog.,* Vol. 53, No. 3 (1963).
63. Water Information Center, "Water Atlas of The United States, Basic Facts About the Nation's Water Resources." Port Washington, N.Y. (1962).
64. BROWNLEE, K. A., *Statistical Theory and Methodology in Science and Engineering.* 2nd Ed., Wiley (1967).

APPENDIX A

EIGHTY-FOUR SETS OF PREDICTION EQUATIONS DEVELOPED FROM 24 HYDROLOGIC/STATISTICAL EXPERIMENTS

The equations derived from the regression analysis for peak runoff from small rural watersheds have a general form of

$$\widehat{Q} = A_0 + A_1 X_1 + A_2 X_2 + \ldots + A_n X_n \qquad \text{(A-1)}$$

in which $\widehat{Q}$ is the predictand (in cfs), the A's are constant coefficients derived from the developmental sample, and the X's are the predictors selected by the screening procedure. In the experiments wherein the predictand was in logarithmic form ($\widehat{\log Q}$) the derived equations were of the form

$$(\widehat{\log Q}) = A_0 + A_1 (\log X_1) + A_2 (\log X_2) + \ldots + A_n (\log X_n) \qquad \text{(A-2)}$$

Eq. (A-2) can be transformed into exponential form with the predictand in cfs, or

$$\widehat{Q} = B_0 \, A'_0 \, X_1^{A_1} X_2^{A_2} \ldots X_n^{A_n} \qquad \text{(A-3)}$$

in which B_0 is the ratio $(\bar{Q}/\bar{Q}')$ and A'_0 is the antilogarithm (base 10) of A_0.

The name, symbol, and units of the predictors included in the equations are as follows:

Predictor Name	*Symbol*	*Units*
Drainage area	A	sq mi
Main stream length	L	mi
Length of tributaries	TRIB	mi
Total length of streams	TRIB + L	mi
Elevation of stream gage	E	ft
Stream slope	S	ft-mi^{-1}
Drainage density	DD	mi^{-1}
Watershed shape	SHAPE	dimensionless
Travel time index	T	mi
Frequency duration precipitation	Pf-d	in.
Mean annual precipitation	MAP	in.
Mean wettest month	P_{wet}	in.
Mean driest month	P_{dry}	in.
Maximum 24-hr precipitation	M24P	in.
No. of 0.01-in. precipitation days/yr	P days	days
No. of thunderstorm days/yr	T days	days
Mean annual snowfall	MAS	in.
Maximum 24-hr snowfall	M4S	in.
No. of 1-in. snowcover days/yr	S days	days

Mean annual temperature	MAT	°F
Mean July temperature	July T	°F
No. of days minimum temperature <32°F/yr	32F days	days
Mean relative humidity	RH	%
Potential evapotranspiration	Pot ET	in.
Percent soil type A	SA	%
Mean annual flood—alternative 3	MAF_5	dimensionless
Geographical regions (REG_1, REG_2, REG_3)	REG	dimensionless
Latitude	LAT	degrees

PHASE I. EVALUATION OF PREDICTOR SETS (A-1 through A-5)

Experiment A-1. Hydrologic-Climatic Predictor Set

Q_{10} *Equation*

$$\widehat{(\log Q_{10})} = 2.04 + 1.03(P_{100-05}) - 0.00054\ (32\text{F days}) + 5.44\ (P_{5-120}) - 4.27\ (P_{5-60}) \qquad \text{(AI-1)}$$

$$\widehat{Q}_{10} = 108.5B_0\ 10^{[1.03\ (P_{100-05}) - 0.00054\ (32\text{F days}) + 5.44\ (P_{5-120}) - 4.27\ (P_{5-60})]} \qquad \text{(AI-2)}$$

Q_{25} *Equation*

$$\widehat{(\log Q_{25})} = 1.04 + 0.021\ (\text{July } T) + 0.028\ (\text{M24P}) - 0.0020\ (32\text{F days}) + 0.65\ (P_{5-120}) \qquad \text{(AI-3)}$$

$$\widehat{Q}_{25} = 10.94B_0\ 10^{[0.021\ (\text{July } T) + 0.028\ (\text{M24P}) - 0.0020\ (32\text{F days}) + 0.65\ (P_{5-120})]} \qquad \text{(AI-4)}$$

Experiment A-2. Topographic Predictor Set

Q_{10} *Equation*

$$\widehat{(\log Q_{10})} = 2.61 + 0.012\ (\text{TRIB}) - 0.000073(E) + 0.022(A) - 0.00058\ (S_{10}) \qquad \text{(AI-5)}$$

$$\widehat{Q}_{10} = 405.8B_0\ 10^{[0.012\ (\text{TRIB}) - 0.000073(E) + 0.22(A) - 0.00058\ (S_{10})]} \qquad \text{(AI-6)}$$

Q_{25} *Equation*

$$\widehat{(\log Q_{25})} = 2.77 + 0.014\ (\text{TRIB}) - 0.00060\ (S_{10}) - 0.000071(E) + 0.019(A) \qquad \text{(AI-7)}$$

$$\widehat{Q}_{25} = 592.9B_0\ 10^{[0.014\ (\text{TRIB}) - 0.00060\ (S_{10}) - 0.000071(E) + 0.019(A)]} \qquad \text{(AI-8)}$$

Experiment A-3. Hydrologic-Climatic and Topographic Predictor Set

Q_{10} *Equation*

$$\widehat{(\log Q_{10})} = 1.80 + 0.0095\ (\text{TRIB}) + 0.81\ (P_{100-05}) - 0.000056(E) + 0.035(A) - 0.10(T) + 4.10\ (P_{5-120}) - 3.16\ (P_{5-60}) \qquad \text{(AI-9)}$$

$$\widehat{Q}_{10} = 63.45B_0\ 10^{[0.0095\ (\text{TRIB}) + 0.81\ (P_{100-05}) - 0.000056(E) + 0.035(A) - 0.10(T) + 4.10\ (P_{5-120}) - 3.16\ (P_{5-60})]} \qquad \text{(AI-10)}$$

Q_{25} *Equation*

$$\widehat{(\log Q_{25})} = 1.90 + 0.0097\ (\text{TRIB}) + 0.87\ (P_{100-05}) - 0.000077(E) + 0.038(A) - 0.14\ (P_{\text{dry}}) - 0.13(T) + 3.49\ (P_{5-120}) - 2.51\ (P_{5-60}) \qquad \text{(AI-11)}$$

$$\widehat{Q}_{25} = 79.25B_0\ 10^{[0.0097\ (\text{TRIB}) + 0.87\ (P_{100-05}) - 0.000077(E) + 0.038(A) - 0.14\ (P_{\text{dry}}) - 0.13(T) + 3.49\ (P_{5-120}) - 2.51\ (P_{5-60})]} \qquad \text{(AI-12)}$$

Experiment A-4. Hydrologic-Climatic, Topographic, and Soil Predictor Set

Q_{10} *Equation*

$$\widehat{(\log Q_{10})} = 1.80 + 0.011\ (\text{TRIB}) - 0.000082(E) + 0.028(A) + 0.025\ (\text{Pot ET}) - 0.0048\ (\text{SA}) \qquad \text{(AI-13)}$$

$$\widehat{Q}_{10} = 63.01B_0\ 10^{[0.011\ (\text{TRIB}) - 0.000082(E) + 0.028(A) + 0.025\ (\text{Pot ET}) - 0.0048\ (\text{SA})]} \qquad \text{(AI-14)}$$

Q_{25} *Equation*

$$\widehat{(\log Q_{25})} = 1.50 + 0.011\ (\text{TRIB}) - 0.000080(E) + 0.017\ (\text{July } T) + 0.026(A) - 0.0041\ (\text{SA}) \qquad \text{(AI-15)}$$

$$\widehat{Q}_{25} = 31.59B_0\ 10^{[0.011\ (\text{TRIB}) - 0.000080(E) + 0.017\ (\text{July } T) + 0.026(A) - 0.0041\ (\text{SA})]} \qquad \text{(AI-16)}$$

Experiment A-5. Hydrologic-Climatic, Topographic, and Soil Predictor Set

Q_{10} *Equation*

$$\widehat{(\log Q_{10})} = 1.43 + 0.012\ (\text{TRIB}) - 0.000075(E) + 0.025(A) + 0.035\ (\text{Pot ET}) + 0.013\ (\text{S days}) - 0.0015\ (S_{10}) + 0.00093(S) \qquad \text{(AI-17)}$$

$$\widehat{Q}_{10} = 26.65B_0\ 10^{[0.012\ (\text{TRIB}) - 0.000075(E) + 0.025(A) + 0.035\ (\text{Pot ET}) + 0.013\ (\text{S days}) - 0.0015\ S_{10} + 0.00093(S)]} \qquad \text{(AI-18)}$$

Q_{25} *Equation*

$$\widehat{(\log Q_{25})} = 1.34 + 0.013\ (\text{TRIB}) - 0.000080(E) + 0.018\ (\text{July } T) + 0.024(A) \qquad \text{(AI-19)}$$

$$\widehat{Q}_{25} = 21.70B_0\ 10^{[0.13\ (\text{TRIB}) - 0.000080(E) + 0.018\ (\text{July } T) + 0.024(A)]} \qquad \text{(AI-20)}$$

PHASE II. FUNCTIONAL FORM OF VARIABLES (B-1 through B-5)

Experiment B-1. Standard Linear Regression

Q_5 *Equation*

$$\widehat{Q}_5 = -772.05 + 22.75\ (\text{TRIB}) + 822.31\ (P_{100-05}) + 10.34\ (\text{MAP}) + 47.47(A) - 270.73\ (P_{\text{dry}}) + 19.27\ (\text{T days}) - 0.10(E) + 1195.1\ (P_{50-10}) + 7532.3\ (P_{5-120}) - 5930.8\ (P_{10-60}) \qquad \text{(AII-1)}$$

Q_{10} *Equation*

$$\widehat{Q}_{10} = -2457.4 + 44.25\ (\text{TRIB}) + 4922.7\ (P_{100-05}) - 1137.1\ (P_{50-10}) + 1127.1\ (P_{5-120}) - 691.09\ (P_{\text{dry}}) + 8.43\ (\text{P days}) + 43.60(A) - 7444.4\ (P_{10-60}) \quad \text{(AII-2)}$$

Q_{25} *Equation*

$$\widehat{Q}_{25} = -3276.5 + 70.33\ (\text{TRIB}) + 42.71\ (\text{July } T) + 148.47\ (P_{\text{wet}}) - 1057.1\ (P_{\text{dry}}) + 494.26\ (P_{100-05}) - 0.38(E) + 45.6\ (\text{T days}) + 72.19\ A \quad \text{(AII-3)}$$

Q_{50} *Equation*

$$\widehat{Q}_{50} = -3538.9 + 131.41\ (\text{TRIB}) + 79.25\ (\text{July } T) - 1498.5\ (P_{\text{dry}}) - 0.62(E) + 65.24\ (\text{T days}) \quad \text{(AII-4)}$$

Experiment B-2. Logarithmic Regression

Q_5 *Equation*

$$\widehat{(\log Q_5)} = -5.23 + 0.45\ (\log \text{TRIB}) + 0.56\ (\log P_{\text{wet}}) + 3.46\ (\log \text{July } T) + 0.51\ (\log \text{MAP}) \quad \text{(AII-5)}$$

$$\widehat{Q}_5 = 0.0000059 B_0\ (\text{TRIB})^{0.45}\ (P_{\text{wet}})^{0.56}\ (\text{July } T)^{3.46}\ (\text{MAP})^{0.51} \quad \text{(AII-6)}$$

Q_{10} *Equation*

$$\widehat{(\log Q_{10})} = -4.35 + 0.44\ (\log \text{TRIB}) + 0.15\ (\log \text{M24P}) + 0.77\ (\log P_{\text{wet}}) + 3.37\ (\log \text{July } T) \quad \text{(AII-7)}$$

$$\widehat{Q}_{10} = 0.000065\ (\text{TRIB})^{0.44}\ (\text{M24P})^{0.15}\ (P_{\text{wet}})^{0.77}\ (\text{July } T)^{3.37} \quad \text{(AII-8)}$$

Q_{25} *Equation*

$$\widehat{(\log Q_{25})} = -4.96 + 0.43\ (\log \text{TRIB}) + 0.30\ (\log \text{M24P}) + 3.80\ (\log \text{July } T) + 0.60\ (\log P_{\text{wet}}) \quad \text{(AII-9)}$$

$$\widehat{Q}_{25} = 0.000017(\text{TRIB})^{0.43}\ (\text{M24P})^{0.30}\ (\text{July } T)^{3.80}\ (P_{\text{wet}})^{0.60} \quad \text{(AII-10)}$$

Q_{50} *Equation*

$$\widehat{(\log Q_{50})} = -4.60 + 0.42\ (\log \text{TRIB}) + 4.33\ (\log \text{July } T) + 1.02\ (\log P_{\text{wet}}) - 0.65\ (\log \text{P days}) \quad \text{(AII-11)}$$

$$\widehat{Q}_{50} = 0.000025\ B_0\ (\text{TRIB})^{0.42}\ (\text{July } T)^{4.33}\ (P_{\text{wet}})^{1.02}\ (\text{P days})^{-0.65} \quad \text{(AII-12)}$$

Experiment B-3. Logarithmic-Linear Regression

Q_5 *Equation*

$$\widehat{(\log Q_5)} = -1.74 + 0.41\ (\log \text{TRIB}) - 0.000086\ (E) - 0.00061\ (S_{10}) + 2.04\ (\log \text{Pot ET}) + 0.18\ (\log E) + 0.49\ (\log \text{MAP}) + 0.0024\ (\text{MAS}) \quad \text{(AII-13)}$$

$$\widehat{Q}_5 = 0.018\ \text{B}_0\ (\text{TRIB})^{0.41}\ (\text{Pot ET})^{2.04}\ (E)^{0.18}\ (\text{MAP})^{0.49} \times 10^{[0.0024(\text{MAS}) - 0.000086(E) - 0.00061(S_{10})]} \quad \text{(AII-14)}$$

Q_{10} *Equation*

$$\widehat{(\log Q_{10})} = -1.12 + 0.28\ (\log \text{TRIB}) - 0.00014\ (E) + 0.0067\ (\text{T days}) + 0.019\ (A) + 0.18\ (\log E) + 1.58\ (\log \text{MAT}) - 0.00069\ (S_{10}) + 0.22\ (\log S) \quad \text{(AII-15)}$$

$$\widehat{Q}_{10} = 0.10\ (\text{TRIB})^{0.28}\ (E)^{0.18}\ (\text{MAT})^{1.58}\ (S)^{0.22} \times 10^{[0.019(A) - 0.00014(E) - 0.00069(S_{10}) + 0.0067(\text{T days})]} \quad \text{(AII-16)}$$

Q_{25} *Equation*

$$\widehat{(\log Q_{25})} = -2.74 + 0.40\ (\log \text{TRIB}) - 0.00010\ (E) + 2.77\ (\log \text{July } T) + 0.24\ (\log E) - 0.00058\ (S_{10}) - 0.0015\ (32\ \text{F days}) \quad \text{(AII-17)}$$

$$\widehat{Q}_{25} = 0.0027 + (\text{TRIB})^{0.40}\ (\text{July } T)^{2.77}\ (E)^{0.24} \times 10^{[-0.00010(E) - 0.00058(S_{10}) - 0.0015(32)\ \text{F days})]} \quad \text{(AII-18)}$$

Q_{50} *Equation*

$$\widehat{(\log Q_{50})} = 0.92 + 0.39\ (\log \text{TRIB}) - 0.00011\ (E) - 0.021\ (\text{July } T) + 0.27\ (\log E) - 0.00056\ (S_{10}) - 0.0016\ (32\ \text{F days}) \quad \text{(AII-19)}$$

$$\widehat{Q}_{50} = 8.37\ B_0\ (\text{TRIB})^{0.39}\ (E)^{0.27} \times 10^{[-0.00011(E) - 0.021(\text{July } T) - 0.00056(S_{10}) - 0.0016(32\ \text{F days})]} \quad \text{(AII-20)}$$

Experiment B-4. Nonlinear Regression (8 Predictors)

$$\widehat{(\log Q_{25})} = -3.25 + 0.98\ [F_1\ (\text{July } T)] + 1.09\ [F_2\ (\text{TRIB})] + 1.01\ [F_3\ (\text{MAF}_5)] + 1.19\ [F_4(E)] + 0.91\ [F_5\ (A)] + 1.10\ [F_6\ (P_{10-360})] + 0.14\ (\text{REG}_3) - 0.46\ (\text{REG}_2) \quad \text{(AII-21)}$$

in which

$$F_1\ (\text{July } T) = 4.738 - 0.143\ (\text{July } T) + 21.45\ (\text{July } T)^2$$

$$F_2\ (\text{TRIB}) = -0.183 + 0.087\ (\text{TRIB}) - 0.054\ (\text{TRIB})^2$$

$$F_3\ (\text{MAF}_5) = 0.127 + 0.133\ (\text{MAF}_5)$$

$$F_4\ (E) = -0.228 + 0.00314\ \frac{E}{(1 + 0.0029\ E)^2} + 0.0000007\ \frac{E^2}{(1 + 0.0029\ E)^2}$$

$$F_5\ (A) = -0.117 + 0.081\ \frac{A}{(1 + 0.097\ A)}$$

$$F_6\ (P_{10-360}) = 0.359 - 0.128\ (P_{10-360}) + 2.475\ (P_{10-360})^2$$

Experiment B-5. Nonlinear Regression (6 Predictors)

$$\widehat{(\log Q_{25})} = 1.30 + 0.85\,[F_1\,(\text{TRIB})] + 1.09\,[F_2\,(\text{MAF}_5)] + 0.84\,[F_3\,(A)] + 1.01\,[F_4\,(P_{10\text{-}360})] + 0.33\,(\text{REG}_1) + 0.15\,(\text{REG}_3) \qquad \text{(AII-22)}$$

in which

$$F_1\,(\text{TRIB}) = -0.263 + 0.399\,\frac{\text{TRIB}}{(1 + 0.533\,\text{TRIB})^2} + 0.138\,\frac{\text{TRIB}^2}{(1 + 0.533\,\text{TRIB})^2}$$

$$F_2\,(\text{MAF}_5) = 0.130 + 0.159\,(\text{MAF}_5) - 0.0052\,(\text{MAF}_5)^2$$

$$F_3(A) = -0.306 + 0.262\,\frac{A}{(1 + 0.233\,A)}$$

$$F_4\,(P_{10\text{-}360}) = 0.309 + 0.194\,(P_{10\text{-}360})$$

PHASE III. STRATIFICATION INTO SUBSETS (C-1 through C-9)

Experiment C-1. Based on MAF_5 (Q_{10} Equation)

$\text{MAF}_5 = 1\ (0 \le \text{MAF} \le 10)$

$$\widehat{(\log Q_{10})} = -6.17 + 0.41\,(\log T) + 2.97\,(\log \text{MAT}) + 0.28\,(\log \text{TRIB}) + 1.72\,(\log \text{P days}) \qquad \text{(AIII-1}a\text{)}$$

$$\widehat{Q}_{10} = 7.77 \times 10^{-7}\,(T)^{0.41}\,(\text{MAT})^{2.97}\,(\text{TRIB})^{0.28}\,(\text{P days})^{1.72} \qquad \text{(AIII-2}a\text{)}$$

$\text{MAF}_5 = 2\ (10 < \text{MAF} \le 30)$

$$\widehat{(\log Q_{10})} = -9.05 + 6.02\,(\log \text{July } T) + 0.36\,(\log \text{MAS}) + 0.20\,(\log \text{TRIB}) \qquad \text{(AIII-3}a\text{)}$$

$$\widehat{Q}_{10} = 12.56 \times 10^{-10}\,(\text{July } T)^{6.02}\,(\text{MAS})^{0.36}\,(\text{TRIB})^{0.20} \qquad \text{(AIII-4}a\text{)}$$

$\text{MAF}_5 = 3\ (30 < \text{MAF} \le 50)$

$$\widehat{(\log Q_{10})} = 2.12 + 0.82\,(\log A) \qquad \text{(AIII-5}a\text{)}$$

$$\widehat{Q}_{10} = 168\,A^{0.82} \qquad \text{(AIII-6}a\text{)}$$

$\text{MAF}_5 = 4\ (50 < \text{MAF} \le 90)$

$$\widehat{(\log Q_{10})} = -4.05 + 0.48\,(\log T) + 0.082\,(\log E) + 0.025\,(\log \text{TRIB}) + 3.52\,(\log \text{Pot ET}) + 0.48\,(\log A) + 0.26\,(\log \text{32 F days}) + 0.34\,(\log S_{10}) + 0.26\,(\log \text{T days}) \qquad \text{(AIII-7}a\text{)}$$

$$\widehat{Q}_{10} = 0.000099\,(T)^{0.48}\,(E)^{0.082}\,(\text{TRIB})^{0.025}\,(\text{Pot ET})^{3.52}\,(A)^{0.48}\,(\text{32 F days})^{0.26}\,(S_{10})^{0.34}\,(\text{T days})^{0.26} \qquad \text{(AIII-8}a\text{)}$$

$\text{MAF}_5 = 5\ (\text{MAF} > 90)$

$$\widehat{(\log Q_{10})} = 3.50 + 0.42\,(\log \text{TRIB}) - 0.42\,(\log S_{10}) \qquad \text{(AIII-9}a\text{)}$$

$$\widehat{Q}_{10} = 3827.23\,(\text{TRIB})^{0.42}\,(S_{10})^{-0.42} \qquad \text{(AIII-10}a\text{)}$$

Experiment C-1. Based on MAF_5 (Q_{25} Equation)

$\text{MAF}_5 = 1\ (0 \le \text{MAF} \le 10)$

$$\widehat{(\log Q_{25})} = -0.61 + 0.82\,(\log T) + 2.13\,(\log \text{MAT}) + 0.37\,(\log \text{TRIB}) - 0.85\,(\log L) \qquad \text{(AIII-1}b\text{)}$$

$$\widehat{Q}_{25} = 0.30\,(T)^{0.82}\,(\text{MAT})^{2.13}\,(\text{TRIB})^{0.37}\,(L)^{-0.85} \qquad \text{(AIII-2}b\text{)}$$

$\text{MAF}_5 = 2\ (10 < \text{MAF} \le 30)$

$$\widehat{(\log Q_{25})} = -10.16 + 6.68\,(\log \text{July } T) + 0.38\,(\log \text{MAS}) + 0.19\,(\log \text{TRIB}) \qquad \text{(AIII-3}b\text{)}$$

$$\widehat{Q}_{25} = 1.06 \times 10^{-11}\,(\text{July } T)^{6.68}\,(\text{MAS})^{0.38}\,(\text{TRIB})^{0.19} \qquad \text{(AIII-4}b\text{)}$$

$\text{MAF}_5 = 3\ (30 < \text{MAF} \le 50)$

$$\widehat{(\log Q_{25})} = 1.75 + 0.87\,(\log A) + 0.75\,(\log \text{M24P}) \qquad \text{(AIII-5}b\text{)}$$

$$\widehat{Q}_{25} = 77.39\,(A)^{0.87}\,(\text{M24P})^{0.75} \qquad \text{(AIII-6}b\text{)}$$

$\text{MAF}_5 = 4\ (50 < \text{MAF} \le 90)$

$$\widehat{(\log Q_{25})} = 5.34 + 0.024\,(\log T) - 0.016\,(\log \text{P days}) + 0.77 \log A) + 0.34\,(\log \text{T days}) \qquad \text{(AIII-7}b\text{)}$$

$$\widehat{Q}_{25} = 250701.36\,(T)^{0.024}\,(\text{P days})^{-0.016}\,(A)^{0.77}\,(\text{T days})^{0.34} \qquad \text{(AIII-8}b\text{)}$$

$\text{MAF}_5 = 5\ (\text{MAF} > 90)$

$$\widehat{(\log Q_{25})} = 4.16 + 0.77\,(\log \text{TRIB}) - 0.52\,(S_{10}) - 0.74\,(\log A) \qquad \text{(AIII-9}b\text{)}$$

$$\widehat{Q}_{25} = 17939.01\,(\text{TRIB})^{0.77}\,(S_{10})^{-0.52}\,(A)^{-0.74} \qquad \text{(AIII-10}b\text{)}$$

Experiment C-1. Based on MAF_5 (Q_{50} Equation)

$\text{MAF}_5 = 1\ (0 \le \text{MAF} \le 10)$

$$\widehat{(\log Q_{50})} = -1.77 + 0.92\,(\log T) + 2.72\,(\log \text{MAT}) + 1.18\,(\log \text{DD}) - 1.35\,(\log \text{SHAPE}) \qquad \text{(AIII-1}c\text{)}$$

$$\widehat{Q}_{50} = 2.04\,(T)^{0.92}\,(\text{MAT})^{2.72}\,(\text{DD})^{1.18}\,(\text{SHAPE})^{-1.35} \qquad \text{(AIII-2}c\text{)}$$

$\text{MAF}_5 = 2\ (10 < \text{MAF} \le 30)$

$$\widehat{(\log Q_{50})} = -11.55 + 7.48\,(\log \text{July } T) + 0.52\,(\log \text{MAS}) \qquad \text{(AIII-3}c\text{)}$$

$$\widehat{Q}_{50} = 4.82 \times 10^{-10}\,(\text{July } T)^{7.48}\,(\text{MAS})^{0.52} \qquad \text{(AIII-4}c\text{)}$$

$\text{MAF}_5 = 3\ (30 < \text{MAF} \le 50)$

$$\widehat{(\log Q_{50})} = 1.81 + 0.89\,(\log A) + 0.83\,(\log \text{M24P}) \qquad \text{(AIII-5}c\text{)}$$

$$\widehat{Q}_{50} = 91.73\,(A)^{0.89}\,(\text{M24P})^{0.83} \qquad \text{(AIII-6}c\text{)}$$

$MAF_5 = 4\ (50 < MAF \leq 90)$

$$\widehat{(\log Q_{50})} = 5.84 + 0.0072\ (\log T) - 1.76\ (\log \text{P days}) + 0.78\ (\log A) + 0.35\ (\log \text{T days}) \quad \text{(AIII-7}c\text{)}$$

$$\widehat{Q}_{50} = 7.89 \times 10^5\ (T)^{0.0072}\ (\text{P days})^{-1.76}\ (A)^{0.78}\ (\text{T days})^{0.35} \quad \text{(AIII-8}c\text{)}$$

$MAF_5 = 5\ (MAF > 90)$

$$\widehat{(\log Q_{50})} = 4.35 + 0.82\ (\log \text{TRIB}) - 0.55\ (\log S_{10}) - 0.84\ (\log A) \quad \text{(AIII-9}c\text{)}$$

$$\widehat{Q}_{50} = 3.07 \times 10^4\ (\text{TRIB})^{0.82}\ (S_{10})^{-0.55}\ (A)^{-0.84} \quad \text{(AIII-10}c\text{)}$$

Experiment C-2. Based on P_{10-60}

$0 \leq P_{10-60} \leq 0.80$

$$\widehat{(\log Q_{25})} = 1.58 + 2.78\ (\log A) - 0.61\ (\log \text{T days}) + 2.25\ (\log \text{DD}) + 0.58\ (\log P_{wet}) - 0.017\ (\log L) - 0.49\ (\log \text{TRIB}) \quad \text{(AIII-11)}$$

$$\widehat{Q}_{25} = 78.37\ (A)^{2.78}\ (\text{T days})^{-0.61}\ (\text{DD})^{2.25}\ (P_{wet})^{0.58}\ (L)^{-0.017}\ (\text{TRIB})^{-0.49} \quad \text{(AIII-12)}$$

$0.80 < P_{10-60} \leq 1.50$

$$\widehat{(\log Q_{25})} = -8.67 + 0.38\ (\log \text{TRIB}) + 4.36\ (\log \text{July } T) - 0.24\ (\log S_{10}) + 2.55\ (\log \text{Pot ET}) \quad \text{(AIII-13)}$$

$$\widehat{Q}_{25} = 2.78 \times 10^{-9}\ (\text{TRIB})^{0.38}\ (\text{July } T)^{4.36}\ (S_{10})^{-0.24}\ (\text{Pot ET})^{2.55} \quad \text{(AIII-14)}$$

$P_{10-60} > 1.50$

$$\widehat{(\log Q_{25})} = 1.71 + 0.40\ (\log \text{TRIB}) + 7.12\ (\log P_{10-30}) \quad \text{(AIII-15)}$$

$$\widehat{Q}_{25} = 77.47\ (\text{TRIB})^{0.40}\ (P_{10-30})^{7.12} \quad \text{(AIII-16)}$$

Experiment C-3. Based on MAP

$0 \leq MAP \leq 34.0$

$$\widehat{(\log Q_{25})} = 3.35 - 0.43\ (\log E) + 0.40\ (\log \text{TRIB}) + 1.08\ (\log \text{M24P}) \quad \text{(AIII-17)}$$

$$\widehat{Q}_{25} = 3781.63\ (E)^{-0.43}\ (\text{TRIB})^{0.40}\ (\text{M24P})^{1.08} \quad \text{(AIII-18)}$$

$34.0 < MAP \leq 47.0$

$$\widehat{(\log Q_{25})} = -10.82 + 0.069\ (\log \text{TRIB}) + 6.56\ (\log \text{July } T) + 0.26\ (\log E) + 0.49\ (\log A) + 0.82\ (\log P_{wet}) \quad \text{(AIII-19)}$$

$$\widehat{Q}_{25} = 1.97 \times 10^{-11}\ (\text{TRIB})^{0.069}\ (\text{July T})^{6.56}\ (E)^{0.26}\ (A)^{0.49}\ (P_{wet})^{0.82} \quad \text{(AIII-20)}$$

$MAP > 47.0$

$$\widehat{(\log Q_{25})} = 1.80 + 0.75\ (\log A) + 0.44\ (\log \text{T days}) \quad \text{(AIII-21)}$$

$$\widehat{Q}_{25} = 90.44\ (A)^{0.75}\ (\text{T days})^{0.44} \quad \text{(AIII-22)}$$

Experiment C-4. Based on MAT

$MAT \leq 50.0$

$$\widehat{(\log Q_{25})} = 2.72 + 0.44\ (\log \text{TRIB}) - 0.000089\ (E) \quad \text{(AIII-23)}$$

$$\widehat{Q}_{25} = 817.8\ (\text{TRIB})^{0.44} \times 10^{-0.000089(E)} \quad \text{(AIII-24)}$$

$50.0 < MAT \leq 55.0$

$$\widehat{(\log Q_{25})} = 0.33 - 0.19\ (\log T) + 0.42\ (P_{10-60}) + 0.81\ (\log P_{wet}) + 0.61\ (\log A) + 0.013\ (\text{T days}) - 0.15\ (P_{dry}) - 0.00023(E) + 0.33\ (\log E) + 0.20\ (\log \text{TRIB}) + 0.013\ (\text{M24S}) \quad \text{(AIII-25)}$$

$$\widehat{Q}_{25} = 2.45\ (T)^{-0.19}\ (P_{wet})^{0.81}\ (A)^{0.61}\ (E)^{0.33}\ (\text{TRIB})^{0.20} \times 10^{[0.42(P_{10-60}) + 0.013(\text{T days}) - 0.15(P_{dry}) - 0.00023(E) + 0.013\ (\text{M24S})]} \quad \text{(AIII-26)}$$

$MAT > 55.0$

$$\widehat{(\log Q_{25})} = 3.25 + 0.38\ (\log \text{TRIB}) - 0.0014\ (S_{10}) - 0.0044\ (\text{32F days}) \quad \text{(AIII-27)}$$

$$\widehat{Q}_{25} = 2514.8\ (\text{TRIB})^{0.38} \times 10^{[-0.0014\ (S_{10}) - 0.0044\ (\text{32F days})]} \quad \text{(AIII-28)}$$

Experiment C-5. Based on Area

$0.0 < A \leq 5.0$

$$\widehat{(\log Q_{25})} = 2.79 + 1.12\ (\log P_{10-10}) + 0.29\ (\log \text{TRIB}) \quad \text{(AIII-29)}$$

$$\widehat{Q}_{25} = 1155.43\ (P_{10-10})^{1.12}\ (\text{TRIB})^{0.29} \quad \text{(AIII-30)}$$

$5.0 < A \leq 15.0$

$$\widehat{(\log Q_{25})} = 1.93 + 0.44\ (\log \text{M24P}) + 0.32\ (\log \text{TRIB}) + 0.77\ (\log P_{wet}) + 0.26\ (\log E) - 0.29\ (\log S_{10}) - 0.16\ (\log \text{MAS}) \quad \text{(AIII-31)}$$

$$\widehat{Q}_{25} = 120.79\ (\text{M24P})^{0.44}\ (\text{TRIB})^{0.32}\ (P_{wet})^{0.77}\ (E)^{0.26}\ (S_{10})^{-0.29}\ (\text{MAS})^{-0.16} \quad \text{(AIII-32)}$$

$A > 15.0$

$$\widehat{(\log Q_{25})} = -9.17 + 5.87\ (\log \text{July } T) + 0.84\ (\log \text{MAP}) + 1.01\ (\log \text{DD}) \quad \text{(AIII-33)}$$

$$\widehat{Q}_{25} = 9.01 \times 10^{-10}\ (\text{July } T)^{5.87}\ (\text{MAP})^{0.84}\ (\text{DD})^{1.01} \quad \text{(AIII-34)}$$

Experiment C-6. Based on Mean Basin Elevation

$\bar{E} \leq 1000.0$

$$\widehat{(\log Q_{25})} = -3.77 + 0.22\ (\log \text{TRIB}) + 3.20\ (\log \text{July } T) + 0.41\ (\log A) + 0.22\ (\log E) - 0.14\ (\log \text{MAS}) \quad \text{(AIII-35)}$$

$$\widehat{Q}_{25} = 0.00023\ (\text{TRIB})^{0.22}\ (\text{July } T)^{3.20}\ (A)^{0.41}\ (E)^{0.22}\ (\text{MAS})^{-0.14} \quad \text{(AIII-36)}$$

$\bar{E} > 1000.0$

$$\widehat{(\log Q_{25})} = -0.65 + 0.37\ (\log \text{TRIB}) - 0.50\ (\log E) - 0.38\ (\log S_{10}) + 3.32\ (\log \text{MAT}) \quad \text{(AIII-37)}$$

$$\widehat{Q}_{25} = 0.35\ (\text{TRIB})^{0.37}\ (E)^{-0.50}\ (S_{10})^{-0.38}\ (\text{MAT})^{3.32} \quad \text{(AIII 38)}$$

Experiment C-7. Based on Soil Erosion

SE = 1

$$\widehat{(\log Q_{25})} = 2.52 - 0.00022\ (E) + 0.050\ (A) - 2.23\ (P_{\text{dry}}) + 0.18\ (\text{M24P}) \quad \text{(AIII-39)}$$

$$\widehat{Q}_{25} = 384.1 \times 10^{[-0.00022(E) + 0.050(A) - 0.23(P_{\text{dry}}) + 0.18(\text{M24P})]} \quad \text{(AIII-40)}$$

$2 \leq \text{SE} \leq 4$

$$\widehat{(\log Q_{25})} = -8.91 - 0.082\ (\log \text{MAP}) + 0.043\ (\text{TRIB}) - 0.0024\ (S_{10}) + 6.97\ (\log \text{MAT}) \quad \text{(AIII-41)}$$

$$\widehat{Q}_{25} = 1.95 \times 10^{-9}\ (\text{MAP})^{-0.082}\ (\text{MAT})^{6.97} \times 10^{[0.043(\text{TRIB}) - 0.0024(S_{10})]} \quad \text{(AIII-42)}$$

SE = 5

$$\widehat{(\log Q_{25})} = 2.24 + 0.053\ (\log T) + 0.53\ (\log \text{TRIB}) - 0.000088\ (E) + 1.17\ (P_{10-10}) - 0.17\ (\text{DD}) - 0.22\ (P_{\text{dry}}) + 0.0053\ (\text{MAP}) \quad \text{(AIII-43)}$$

$$\widehat{Q}_{25} = 227.2\ (T)^{0.053}\ (\text{TRIB})^{0.53} \times 10^{[1.17(P_{10-10}) - 0.000088(E) - 0.17(\text{DD}) - 0.22(P_{\text{dry}}) + 0.0053\ (\text{MAP})]} \quad \text{(AIII-44)}$$

SE = 6

$$\widehat{(\log Q_{25})} = -7.47 + 0.0088\ (\text{TRIB}) + 5.02\ (\log \text{July } T) + 0.39\ (\log E) - 0.00016\ (E) + 0.28\ (\log A) \quad \text{(AIII-45)}$$

$$\widehat{Q}_{25} = 4.28 \times 10^{-8}\ (\text{July } T)^{5.02}\ (E)^{0.39}\ (A)^{0.28} \times 10^{[0.0088(\text{TRIB}) - 0.00016(E)]} \quad \text{(AIII-46)}$$

Experiment C-8. Based on Geological Zones

GZ = 1

$$\widehat{(\log Q_{25})} = 3.74 + 0.0022\ (\text{TRIB}) - 0.012\ (\text{32F days}) + 5.78\ (\log P_{10-30}) + 0.59\ (\log A) \quad \text{(AIII-49)}$$

$$\widehat{Q}_{25} = 6414.5\ (P_{10-30})^{5.78}\ (A)^{0.59} \times 10^{[0.0022\ (\text{TRIB}) - 0.012(\text{32F Days})]} \quad \text{(AIII-50)}$$

GZ = 2

$$\widehat{(\log Q_{25})} = 2.57 + 0.61\ (\log \text{TRIB}) \quad \text{(AIII-51)}$$

$$\widehat{Q}_{25} = 554.7\ (\text{TRIB})^{0.61} \quad \text{(AIII-52)}$$

GZ = 3

$$\widehat{(\log Q_{25})} = -4.44 + 0.81\ (\log T) + 4.23\ (\log \text{32F days}) - 0.047\ (\text{M24S}) \quad \text{(AIII-53)}$$

$$\widehat{Q}_{25} = 4.13 \times 10^{-5}\ (T)^{0.81}\ (\text{32F days})^{4.23} \times 10^{-0.047\ (\text{M24S})} \quad \text{(AIII-54)}$$

GZ = 4

$$\widehat{(\log Q_{25})} = 2.74 + 0.48\ (\log A) \quad \text{(AIII-55)}$$

$$\widehat{Q}_{25} = 676.0\ (A)^{0.48} \quad \text{(AIII-56)}$$

GZ = undefined

$$\widehat{(\log Q_{25})} = 0.82 + 0.23\ (\log \text{TRIB}) - 0.000070\ (E) + 0.048\ (\text{Pot ET}) + 0.36\ (\log A) + 0.012\ (\text{M24S}) \quad \text{(AIII-57)}$$

$$\widehat{Q}_{25} = 8.70\ (\text{TRIB})^{0.23}\ (A)^{0.36} \times 10^{[0.048(\text{Pot ET}) - 0.000070(E) + 0.012\ (\text{M24S})]} \quad \text{(AIII-58)}$$

Experiment C-9. Based on Geographic Regions

REG_1

$$\widehat{(\log Q_{25})} = 2.71 + 0.032\ (A) \quad \text{(AIII-59)}$$

$$\widehat{Q}_{25} = 600.47 \times 10^{0.032(A)} \quad \text{(AIII-60)}$$

REG_2

$$\widehat{(\log Q_{25})} = 0.86 + 0.016\ (\text{TRIB}) + 0.64\ (P_{10-360}) \quad \text{(AIII-61)}$$

$$\widehat{Q}_{25} = 8.46 \times 10^{[0.016(\text{TRIB}) + 0.64(P_{10-360})]} \quad \text{(AIII-62)}$$

REG_3

$$\widehat{(\log Q_{25})} = 3.16 + 0.057\ (A) - 0.010\ (\text{32F days}) \quad \text{(AIII-63)}$$

$$\widehat{Q}_{25} = 1385.11 \times 10^{[0.057(A) - 0.010(\text{32F days})]} \quad \text{(AIII-64)}$$

REG_4

$$\widehat{(\log Q_{25})} = 1.49 + 1.31\ (T) + 0.28\ (\mathrm{MAF}_5) \qquad \text{(AIII-65)}$$

$$\widehat{Q}_{25} = 33.62 \times 10^{[1.31(T)+0.28(\mathrm{MAF}_5)]} \qquad \text{(AIII-66)}$$

REG_5

$$\widehat{(\log Q_{25})} = -4.54 + 0.28(L) + 0.23\ (P_{10-360}) - 0.83\ (T) - 0.37\ (\mathrm{SHAPE}) + 0.045\ (\mathrm{RH}) + 0.045\ (\text{July } T) \qquad \text{(AIII-67)}$$

$$\widehat{Q}_{25} = 3.68 \times 10^{-5} \times 10^{[0.28(L)+0.23(P_{10-360}) -0.83(T)-0.37(\mathrm{SHAPE})+0.045(\mathrm{RH})+0.045(\text{July } T)]} \qquad \text{(AIII-68)}$$

REG_6

$$\widehat{(\log Q_{25})} = 0.88 + 0.034\ (\mathrm{RH}) + 0.019\ (\mathrm{TRIB}) \qquad \text{(AIII-69)}$$

$$\widehat{Q}_{25} = 8.75 \times 10^{[0.034(\mathrm{RH})+0.019(\mathrm{TRIB})]} \qquad \text{(AIII-70)}$$

PHASE IV. NATIONAL EQUATIONS WITH STRATIFICATION FACTORS (D-1, D-2, D-3)

Experiment D-1. Semi-Logarithmic Equation

Q_5 Equation

$$\widehat{(\log Q_5)} = 1.028 + 0.0057\ (\mathrm{TRIB}) + 0.18\ (P_{10-360}) + 0.15\ (\mathrm{MAF}_5) + 0.034\ (A) + 0.29\ (\mathrm{REG}_2) + 0.22\ (\mathrm{REG}_1) \qquad \text{(AIV-1)}$$

$$\widehat{Q}_5 = 10.66\ B_0\ 10^{[0.0057(\mathrm{TRIB})+0.18(P_{10-360}) +0.15(\mathrm{MAF}_5)+0.034(A)+0.29(\mathrm{REG}_2)+0.22\ (\mathrm{REG}_2)]} \qquad \text{(AIV-2)}$$

Q_{10} Equation

$$\widehat{(\log Q_{10})} = 1.15 + 0.0064\ (\mathrm{TRIB}) + 0.20\ (P_{10-360}) + 0.14\ (\mathrm{MAF}_5) + 0.034\ (A) + 0.35\ (\mathrm{REG}_2) + 0.20\ (\mathrm{REG}_1) \qquad \text{(AIV-3)}$$

$$\widehat{Q}_{10} = 16.60 \times 10^{[0.0064(\mathrm{TRIB})+0.20(P_{10-360})+0.14(\mathrm{MAF}_5) +0.034(A)+0.35(\mathrm{REG}_2)+0.20(\mathrm{REG}_1)} \qquad \text{(AIV-4)}$$

Q_{25} Equation

$$\widehat{(\log Q_{25})} = 1.28 + 0.0073\ (\mathrm{TRIB}) + 0.22\ (P_{10-360}) + 0.13\ (\mathrm{MAF}_5) + 0.033\ (A) + 0.41\ (\mathrm{REG}_2)\ 0.18\ (\mathrm{REG}_1) \qquad \text{(AIV-5)}$$

$$\widehat{Q}_{25} = 22.74 \times 10^{[0.0073(\mathrm{TRIB})+0.22(P_{10-360})+0.13(\mathrm{MAF}_5) +0.033(\mathrm{A})+0.41(\mathrm{REG}_2)+0.18(\mathrm{REG}_1)]} \qquad \text{(AIV-6)}$$

Q_{50} Equation

$$\widehat{(\log Q_{50})} = 2.74 + 0.0058\ (\mathrm{TRIB}) + 0.22\ (P_{10-360}) + 0.37\ (\mathrm{REG}_2) + 0.033\ (A) + 0.14\ (\mathrm{MAF}_5) - 0.033\ (\mathrm{LAT}) - 0.42\ (\mathrm{REG}_3) \qquad \text{(AIV-7)}$$

$$\widehat{Q}_{50} = 665.20 \times 10^{[0.0058(\mathrm{TRIB})+0.22(P_{10-360})+0.37(\mathrm{REG}_2) +0.033(A)+0.14(\mathrm{MAF}_5)-0.033(\mathrm{LAT})-0.42(\mathrm{REG}_3)]} \qquad \text{(AIV-8)}$$

Experiment D-2. Logarithmic-Linear Regression Equation

Q_5 Equation

$$\widehat{(\log Q_5)} = 1.12 + 0.22\ (\log \mathrm{TRIB}) + 1.26\ (\log P_{10-360}) + 0.17\ (\mathrm{MAF}_5) + 0.76\ (\log T) - 0.26\ (\mathrm{SHAPE}) - 0.30\ (\mathrm{REG}_3) + 0.32\ (\log S) + 0.18\ (\log \mathrm{T\ days}) \qquad \text{(AIV-9)}$$

$$\widehat{Q}_5 = 13.31\ B_0\ (\mathrm{TRIB})^{0.22}\ (P_{10-360})^{1.26}\ (T)^{0.76}\ (S)^{0.32}\ (\mathrm{T\ days})^{0.18} \times 10^{[0.17(\mathrm{MAF}_5)-0.26(\mathrm{SHAPE})-0.30(\mathrm{REG}_3)]} \qquad \text{(AIV-10)}$$

Q_{10} Equation

$$\widehat{(\log Q_{10})} = -2.72 + 0.24\ (\log \mathrm{TRIB}) + 1.15\ (\log P_{10-360}) + 0.16\ (\mathrm{MAF}_5) + 0.12\ (\log T) - 0.31\ (\mathrm{REG}_3) + 2.33\ (\log \text{July } T) + 0.27\ (\log A) \qquad \text{(AIV-11)}$$

$$\widehat{Q}_{10} = 2.46 \times 10^{-3}\ (\mathrm{TRIB})^{0.24}\ (P_{10-360})^{1.15}\ (T)^{0.12}\ (\text{July } T)^{2.23}\ (A)^{0.27}\ 10^{[0.16(\mathrm{MAF}_5)-0.31(\mathrm{REG}_3)]} \qquad \text{(AIV-12)}$$

Q_{25} Equation

$$\widehat{(\log Q_{25})} = -3.28 + 0.26\ (\log \mathrm{TRIB}) + 1.31\ (\log P_{10-360}) + 0.15\ (\mathrm{MAF}_5) + 0.16\ (\mathrm{REG}_2) + 0.36\ (\log A) + 2.52\ (\log \text{July } T) - 0.30\ (\mathrm{REG}_3) \qquad \text{(AIV-13)}$$

$$\widehat{Q}_{25} = 7.01 \times 10^{-4}\ (\mathrm{TRIB})^{0.26}\ (P_{10-360})^{1.31}\ (A)^{0.36}\ (\text{July } T)^{2.52} \times 10^{[0.15(\mathrm{MAF}_5)+0.16(\mathrm{REG}_2)-0.30(\mathrm{REG}_3)]} \qquad \text{(AIV-14)}$$

Q_{50} Equation

$$\widehat{(\log Q_{50})} = -0.75 + 0.23\ (\log \mathrm{TRIB}) + 1.62\ (\log P_{10-360}) + 0.14\ (\mathrm{MAF}_5) + 0.27\ (\mathrm{REG}_2) + 0.38\ (\log A) - 0.35\ (\log \mathrm{MAP}) - 0.35\ (\mathrm{REG}_3) + 1.79\ (\log \mathrm{Pot\ ET}) \qquad \text{(AIV-15)}$$

$$\widehat{Q}_{50} = 0.025\ (\mathrm{TRIB})^{0.23}\ (P_{10-360})^{1.62}\ (A)^{0.38}\ (\mathrm{MAP})^{-0.35}\ (\mathrm{Pot\ ET})^{1.79} \times 10^{[0.14(\mathrm{MAF}_5)+0.27(\mathrm{REG}_2)-0.35(\mathrm{REG}_3)]} \qquad \text{(AIV-16)}$$

Experiment D-3. Logarithmic Regression Equation

Q_5 Equation

$$\widehat{(\log Q_5)} = 0.97 + 0.79\ (\log \mathrm{TRIB} + L) + 1.24\ (\log P_{10-360}) + 0.76\ (\log \mathrm{MAF}_5) - 0.75\ (\log \mathrm{SHAPE}) + 0.14\ (\mathrm{REG}_2) \qquad \text{(AIV-17)}$$

$$\widehat{Q}_5 = 9.285\ B_0\ (\mathrm{TRIB} + L)^{0.79}\ (P_{10-360})^{1.24}\ (\mathrm{MAF}_5)^{0.76}\ (\mathrm{SHAPE})^{-0.75}\ 10^{0.14(\mathrm{REG}_2)} \qquad \text{(AIV-18)}$$

Q_{10} *Equation*

$$\widehat{(\log Q_{10})} = -2.54 + 0.77\ (\log \text{TRIB} + L) + 1.14\ (\log P_{10\text{-}360}) + 0.82\ (\log \text{MAF}_5) + 0.10\ (\text{REG}_2) - 0.66\ (\log \text{SHAPE}) + 2.01\ (\log \text{July } T) - 0.23\ (\text{REG}_3) \quad \text{(AIV-19)}$$

$$\widehat{Q}_{10} = 3.82 \times 10^{-3}\ (\text{TRIB} + L)^{0.77}\ (P_{10\text{-}360})^{1.14}\ (\text{MAF}_5)^{0.82}\ (\text{SHAPE})^{-0.66}\ (\text{July } T)^{2.01}\ 10^{0.10(\text{REG}_2) - 0.23(\text{REG}_3)} \quad \text{(AIV-20)}$$

Q_{25} *Equation*

$$\widehat{(\log Q_{25})} = -3.56 + 0.77\ (\log \text{TRIB} + L) + 1.18\ (\log P_{10\text{-}360}) + 0.79\ (\log \text{MAF}_5) + 0.13\ (\text{REG}_2) - 0.61\ (\log \text{SHAPE}) + 2.63\ (\log \text{July } T) - 0.28\ (\text{REG}_3) \quad \text{(AIV-21)}$$

$$\widehat{Q}_{25} = 3.69 \times 10^{-4}\ (\text{TRIB} + L)^{0.77}\ (P_{10\text{-}360})^{1.18}\ (\text{MAF}_5)^{0.79}\ (\text{SHAPE})^{-0.61}\ (\text{July } T)^{2.63}\ 10^{0.13(\text{REG}_2) - 0.28(\text{REG}_3)} \quad \text{(AIV-22)}$$

Q_{50} *Equation*

$$\widehat{(\log Q_{50})} = -4.21 + 0.77\ (\log \text{TRIB} + L) + 120\ (\log P_{10\text{-}360}) + 0.78\ (\log \text{MAF}_5) + 0.15\ (\text{REG}_2) + 3.04\ (\log \text{July } T) - 0.31\ (\text{REG}_3) - 0.58\ (\log \text{SHAPE}) \quad \text{(AIV-23)}$$

$$\widehat{Q}_{50} = 8.60 \times 10^{-5}\ (\text{TRIB} + L)^{0.77}\ (P_{10\text{-}360})^{1.20}\ (\text{MAF}_5)^{0.78}\ (\text{July } T)^{3.04}\ (\text{SHAPE})^{-0.58}\ 10^{0.15(\text{REG}_2) - 0.31(\text{REG}_3)} \quad \text{(AIV-24)}$$

PHASE V. SIMPLIFIED NATIONAL EQUATIONS (E-1, E-2)

Experiment E-1. Two-Predictor National Equation

Q_5 *Equation*

$$\widehat{(\log Q_5)} = 2.07 + 0.64\ (\log A) + 0.93\ (\log P_{10\text{-}360}) \quad \text{(AV-1)}$$

$$\widehat{Q}_5 = 116.90\ B_0\ (A)^{0.64}\ (P_{10\text{-}360})^{0.93} \quad \text{(AV-2)}$$

Q_{10} *Equation*

$$\widehat{(\log Q_{10})} = 2.21 + 0.64\ (\log A) + 1.05\ (\log P_{10\text{-}360}) \quad \text{(AV-3)}$$

$$\widehat{Q}_{10} = 163.75\ B_0\ (A)^{0.64}\ (P_{10\text{-}360})^{1.05} \quad \text{(AV-4)}$$

Q_{25} *Equation*

$$\widehat{(\log Q_{25})} = 2.37 + 0.64\ (\log A) + 1.18\ (\log P_{10\text{-}360}) \quad \text{(AV-5)}$$

$$\widehat{Q}_{25} = 366.54\ (A)^{0.64}\ (P_{10\text{-}360})^{1.18} \quad \text{(AV-6)}$$

Q_{50} *Equation*

$$\widehat{(\log Q_{50})} = 2.47 + 0.64\ (\log A) + 1.26\ (\log P_{10\text{-}360}) \quad \text{(AV-7)}$$

$$\widehat{Q}_{50} = 293.41\ B_0\ (A)^{0.64}\ (P_{10\text{-}360})^{1.26} \quad \text{(AV-8)}$$

Experiment E-2. Four- or Six-Predictor National Equation

Q_5 *Equation*

$$\widehat{(\log Q_5)} = 0.97 + 0.72\ (\log A) + 1.36\ (\log P_{10\text{-}360}) + 0.82\ (\log \text{MAF}_5) + 0.20\ (\text{REG}_2) \quad \text{(AV-9)}$$

$$\widehat{Q}_5 = 9.343\ B_0\ (A)^{0.72}\ (P_{10\text{-}360})^{1.36}\ (\text{MAF}_5)^{0.82}\ 10^{0.20(\text{REG}_2)} \quad \text{(AV-10)}$$

Q_{10} *Equation*

$$\widehat{(\log Q_{10})} = -3.23 + 0.71\ (\log A) + 1.18\ (\log P_{10\text{-}360}) + 0.90\ (\log \text{MAF}_5) + 0.15\ (\text{REG}_2) + 2.41\ (\log \text{July } T) - 0.23\ (\text{REG}_3) \quad \text{(AV-11)}$$

$$\widehat{Q}_{10} = 7.95 \times 10^{-4}\ (A)^{0.71}\ (P_{10\text{-}360})^{1.18}\ (\text{MAF}_5)^{0.90}\ (\text{July } T)^{2.41}\ 10^{0.15(\text{REG}_2) - 0.23(\text{REG}_3)} \quad \text{(AV-12)}$$

Q_{25} *Equation*

$$\widehat{(\log Q_{25})} = -4.26 + 0.70\ (\log A) + 1.21\ (\log P_{10\text{-}360}) + 0.87\ (\log \text{MAF}_5) + 0.19\ (\text{REG}_2) + 3.05\ (\log \text{July } T) - 0.28\ (\text{REG}_3) \quad \text{(AV-13)}$$

$$\widehat{Q}_{25} = 7.66 \times 10^{-5}\ (A)^{0.70}\ (P_{10\text{-}360})^{1.21}\ (\text{MAF}_5)^{0.87}\ (\text{July } T)^{3.05}\ 10^{0.19(\text{REG}_2) - 0.28(\text{REG}_3)} \quad \text{(AV-14)}$$

Q_{50} *Equation*

$$\widehat{(\log Q_{50})} = -4.93 + 0.70\ (\log A) + 1.22\ (\log P_{10\text{-}360}) + 0.20\ (\text{REG}_2) + 0.86\ (\log \text{MAF}_5) + 3.46\ (\log \text{July } T) - 0.32\ (\text{REG}_3) \quad \text{(AV-15)}$$

$$\widehat{Q}_{50} = 1.74 \times 10^{-5}\ (A)^{0.70}\ (P_{10\text{-}360})^{1.22}\ (\text{MAF}_5)^{0.86}\ (\text{July } T)^{3.46}\ 10^{0.20(\text{REG}_2) - 0.32(\text{REG}_3)} \quad \text{(AV-16)}$$

APPENDIX B

AN EVALUATION OF THREE CUMULATIVE DISTRIBUTION FUNCTIONS FITTED TO ANNUAL PEAK RUNOFF AMOUNTS

Flow frequency analysis methods are widely used in hydrologic problems. Many methods have been developed and modified over the years that fit one cumulative distribution function or another to a series of annual peak flow values for a gaging station to obtain return period (or recurrence interval) values. When sufficient data are available, the various methods yield results that are comparable within the range of the data. However, appreciable differences have occurred when the functions are extended beyond the range of the data (e.g., estimating a 100-year flood from 25 years of data). Therefore, it was not clear at the start of this study which cumulative distribution function should be used.

An *ad hoc* Work Group on Flow Frequency Analysis of the Subcommittee on Hydrology, Inter-Agency Committee on Water Resources, investigated the problem of recommending a cumulative distribution function to fit to annual peak runoff values. They recommended the log-Pearson Type III distribution (with the log-normal as a special case). They further recommended that if information exists that indicates some other distribution or technique is preferred, it may be used provided adequate justification is given for such use. In light of this latter recommendation, and because the Work Group based its analyses on data from just ten basins, it was decided to conduct an investigation of three well-known methods applied to the annual peak discharge values from small rural watersheds. The methods selected for the investigation were log-Pearson Type III, log-normal, and Gumbel.

The objective of the study was to identify the best method for this application by fitting the three functions to several hundred series of data and measuring the closeness of fit of each by statistical tests. The function exhibiting the best fit on the basis of the statistical tests would then be used to develop the peak discharge return period values to be used as the predictands in the regression experiments.

The three methods used in this evaluation are described in the following section.

METHODS FOR FITTING DISTRIBUTIONS

In hydrology, the return period $T(X)$ is defined as

$$T(X) = \frac{1}{1 - F(X)} \qquad \text{(B-1)}$$

in which $F(X)$ is the probability that the values of a variable, x, are equal to or smaller than a specified value, X. This can be represented by a cumulative distribution function that has been fitted to a series of annual values of the variable. In words, $T(X)$ is the number of years such that, on the average, there would be just one value of the variable equalling or exeeding the fitted value X.

The central problem is to compute X for assigned values of $T(X)$. In this study $T(X)$ has been assigned the values 2, 5, 10, 20, 25, 50, and 100 years. The solution is obtained by the following steps:

1. Assign values to $T(X)$; i.e., 2, 5, . . . , 100 years.
2. Solve Eq. B-1 for each $F(X)$; e.g., $F(X) = 0.95$ if $T(X) = 20$ years.
3. Fit the selected distribution function to a series of annual peak discharge values.
4. Set the fitted distribution function equal to the values obtained in step 2 and solve for X's.

The methods of fitting (step 3) and the method of solving for X (step 4) for each of the three distribution functions—log-normal, log-Pearson type III, and Gumbel—are given in subsequent sections.

Log-Normal

Let $x_1, x_2, \ldots, x_i, \ldots, x_N$ denote a series of annual values of a variable for N years. Let y_i be the logarithm of x_i (it is immaterial whether natural or other base logarithms are used).

Compute

$$\bar{y} = N^{-1} \sum_{i=1}^{N} y_i \qquad \text{(B-2)}$$

$$s = \left[(N-1)^{-1} \sum_{i=1}^{N} (y_i - \bar{y})^2 \right]^{1/2} \qquad \text{(B-3)}$$

For assigned values of $T(X)$ of 2, 5, 10, 20, 25, 50, and 100 years, the corresponding values of $F(X)$ are 0.50, 0.80, 0.90, 0.95, 0.96, 0.98, and 0.99. Given these values of $F(X)$ and published tables of the cumulative normal distribution (*38*), one obtains values of $k_{F(X)}$, a unit normal deviate for the probability $F(X)$.

The computed value (peak discharge in this case) in logarithmic form is determined from

$$Y_{F(X)} = \bar{y} + k_{F(X)}\, s \qquad \text{(B-4)}$$

which can be converted back to original form by

$$X_{F(X)} = a \exp Y_{F(X)} \qquad \text{(B-5)}$$

in which a is the logarithm base selected initially. The values $X_{F(X)}$ are the solution to the central problem; i.e., for $T(X)$ future observations of the variable, x, on the average there would be just one value of x equalling or exceeding $X_{F(X)}$.

Log-Pearson Type III

The log-normal distribution function is a special case of log-Pearson Type III, the only difference being that the latter requires computation and use of the skewness of the distribution. Again, the logarithm of each of a series of annual values, $x_1, x_2, \ldots, x_N$ is computed and Eqs. B-2 and B-3 are applied to compute $\bar{y}$ and s.

It is now necessary to compute the coefficient of skew

$$g = N\,[(N\text{-}1)\,(N\text{-}3)\,s^3]^{-1} \sum_{i=1}^{N} (y_i - \bar{y})^3 \qquad \text{(B-6)}$$

Given g and each value of $F(X)$, a series of values, $k_{F(X)}$, is obtained from published tables of the Pearson Type III distribution (*27*). Each $k_{F(X)}$ is entered in Eq. B-4 to obtain a series of $Y_{F(X)}$ values that in turn are entered in Eq. B-5 to obtain a series of $X_{F(X)}$ values yielding the solution.

Gumbel

The cumulative distribution function of the Gumbel distribution is

$$F(X) = \exp\,(-\,e^{-Z}) \qquad \text{(B-7)}$$

in which

$$Z = \alpha(X - u) \qquad \text{(B-8)}$$

The intermediate variable, Z, in Eq. B-8 is often referred to as the "reduced" variable, α (or sometimes $1/\alpha$) is the scale parameter, and u is the mode of the Gumbel distribution.

The parameters α and u are estimated from the data. As before, let $x_1, x_2, \ldots, x_i \ldots, x_N$ denote a series of N annual values. Compute the mean

$$\bar{x} = (N)^{-1} \sum_{i=1}^{N} x_i \qquad \text{(B-9)}$$

and standard deviation

$$s = [(N\text{-}1)^{-1} \sum_{i=1}^{N} (x_i - \bar{x})^2]^{1/2} \qquad \text{(B-10)}$$

Then,

$$\frac{1}{\alpha} = \frac{s}{\sigma\,(N)} \qquad \text{and} \qquad u = \bar{x} - y(N)/\alpha \qquad \text{(B-11)}$$

in which $\sigma(N)$ and $y(N)$ are obtained from published tables (*26*).

As before, a series of $F(X)$ values is given. The series of $Z_{F(X)}$ values is obtained from Eq. B-7. These in turn are entered in Eq. B-8 with the computed α and u to obtain a series of $X_{F(X)}$ values, which constitute the solution to the central problem.

TESTS OF FIT

The chi-square and binomial tests were used to evaluate the agreement between a fitted distribution and the distribution of observed data. Before discussing the two tests a term needs to be defined. In the previous section methods were described for computing a series of values of $X_{F(X)}$ for the three distribution functions where $F(X)$ is some value between 0 and 1. For ease of notation p is used in place of $F(X)$ and X_p is termed the pth percentile. If the fitted distribution fits the observations perfectly, $100 \cdot p$ percent of the observations will be equal to or less than X. Similarly, $100 \cdot (p_2 - p_1)$ percent of the observations will lie between X_{p2} and X_{p1}.

Chi-Square Test

To apply the chi-square test, a series of p-values is specified. Let these be denoted by $p_1, p_2, \ldots, p_m$, where m is the number of p-values. Let $p_0 = 0$ and $p_{m+1} = 1$. Using the methods of the previous section, m percentile values are computed, $X_{p1}, X_{p2}, \ldots, X_{pm}$. Let $X_{p0} = -\infty$ and $X_{pm+1} = +\infty$.

Then compute

$$\chi^2 = \sum_{j=1}^{m+1} \{[n_j - N(p_j - p_{j-1})]^2 / N(p_j - p_{j-1})\} \qquad \text{(B-12)}$$

in which n_j is the number of observations falling in the jth group (i.e., lying between X_{pj} and X_{pj-1}), N is the total number of observations, and $N(p_j - p_{j-1})$ is the expected number in the jth group.

If the distribution of observed data agrees perfectly with the fitted distribution, $\chi^2 = 0$ because the actual number of observations for any group is equal to the expected number for that group. If the observed and fitted distributions do not agree, χ^2 becomes large. The question arises as to what is meant by "large" because perfect fit is highly unlikely to occur with any sample of data and therefore χ^2 will exceed zero even if the distribution of observed values does, in fact, follow the theoretical distribution. The answer is that if observed data follow a theoretical distribution, χ^2, computed by Eq. B-12, is expected to equal the number of degrees of freedom.

At this point, it is necessary to discuss the concept of degrees of freedom because many statistics textbooks are misleading on the subject. The rule is that the degrees of freedom, f, is found from

$$f = M - H - 1 \qquad \text{(B-13)}$$

in which M is the number of groups and H is the number of parameters of the distribution function estimated from the sample.

Specification of five percentile values divides the data into six groups. For the Gumbel and log-normal distributions the number of estimated parameters is 2; for log-Pearson Type III, 3. Eq. B-13 is the one presented in statistics texts. What is often omitted is that the rule applies only when the parameters are estimated from the group frequencies. In this study, the estimating parameters were computed from the total samples, in which case f is not known exactly. However, it is known that f lies between $M - 1$ and $M - H - 1$. (See Chernoff, H., "Degrees of Freedom for Chi-Square." *Technometrics*, Vol. 9, No. 3, 1967.) As an example, if the number of groups is set at, 6, in Eq. B-13 $M = 6$ and $H = 2$ for the log-normal and Gumbel distributions so that the true value of f lies between 5 and 3, whereas for the log-Pearson Type III distribution it lies between 5 and 2 because $H = 3$. Similarly, if the number of groups is set at 2, the true value of f, which can never be negative, lies between 0 and 1.

Binomial Test

In those cases where only one percentile value is to be evaluated, the binomial test can be applied to test the goodness of fit. Let X_p be the percentile value. A "success" is defined as the event when an observation exceeds X_p. Then $1 - p$ is the probability of success. Let W be the number of trials and let w be the number of successes. With large W (which is true in this study), the normal approximation to the binomial is quite accurate. Then

$$B = \frac{w - (1 - p)W}{[p(1 - p)W]^{1/2}} \qquad \text{(B-14)}$$

is a normal deviate with a mean of 0 and a standard deviation of 1.

A perfect fit would result in $B = 0$. As noted previously, a perfect fit is highly unlikely even if the distribution of observed values does follow the theoretical distribution. If the latter is true, then, because B is a normal deviate (0,1), there is a probability of 0.05 that B lies between -1.96 and 1.96 and a probability of 0.01 that B lies in the range ± 2.58. Values of B outside these ranges are evidence that the theoretical distributions do not fit well. In comparing the fit of two distributions to the same data, the one resulting in a B nearest zero fits best.

MAXIMUM ANNUAL PEAK RUNOFF

Annual peak runoff data had been collected and processed into computer-usable form for 459 gaging stations distributed across the continental United States. (Although data for 493 watersheds were used throughout most of the study, data were available for only 459 gaging stations for this portion of the analysis.) The stations pertained to rural watersheds with area of 25 sq mi or less, and not subjected to any significant natural or man-made diversion or control. The number of annual values available for each station ranged upward from a minimum of 12 years, with a mean of 22.4 years of record. The total of data available for the evaluation comprised 10,287 station-years.

The three cumulative distribution function methods (log-Pearson Type III, log-normal, and Gumbel) were programmed for computer utilization and each of the 459 series of annual peak runoffs was fitted separately by each method.

Five Percentile Values

Five percentile values ($X_{0.50}$, $X_{0.60}$, $X_{0.70}$, $X_{0.80}$, and $X_{0.90}$) were computed for each distribution function for each of the 459 series of data. The evaluation was initiated at $X_{0.50}$, rather than some lower percentile, because 0.50 corresponds to a return period of two years, the lowest period of interest in the study. Using Eq. B-12, 459 values of χ^2 were computed for the three distribution functions. The averages of these 459 values of χ^2 for each distribution function are given in Table B-1.

Table B-1 indicates that the mean chi-squares for the log-normal and Gumbel methods do lie in the proper range, as discussed earlier, but the mean chi-square for log-Pearson Type III lies outside its acceptable range. Based on this test, this last distribution fits annual peak runoff data less well than the other two.

TABLE B-1

MEAN CHI-SQUARE FOR 459 RUNOFF STATIONS

DISTRIBUTION FUNCTION	MEAN CHI-SQUARE
Log-Pearson type III	5.53
Log-normal	3.91
Gumbel	4.15

One Percentile Value

From a hydrological viewpoint it is important that the distribution functions fit the observed values at percentiles higher than $X_{0.90}$ (10-year return period) the largest percentile used in the immediately preceding section. Consequently, four percentiles were chosen for investigation corresponding to return periods of 10, 20, 50, and 100 years. Each percentile was used separately to divide the data into two groups (i.e., individual values of peak runoff are either below or above the specified percentile). Chi-square was computed for each of the 459 series of data for each two-group separation for each distribution function. A sum of chi-square values was obtained for each function for each percentile; i.e., 459 chi-square values were summed. The results are given in Table B-2. From theoretical considerations, for a single station's data divided into two groups the expected value of χ^2 lies between 0 and 1. Therefore, the expected value of 459 stations lies between 0 and 459. Both the log-normal and Gumbel values meet this criterion, but three of the log-Pearson Type III values are too high.

The next test involved counting the number of observations that fell above the computed percentile (return period values) for a basin and then summing over all basins. There were a total of 10,287 observations. Table B-3 gives the number of observations falling above each of the four return period values. The last line of the table gives the number of observations "expected" to fall above the percentile. This row is computed by $(1 - 0) \times 10{,}287$. It should be noted that for log-Pearson Type III there are 418 values exceeding the computed 50-year return

TABLE B-2

SUMS OF CHI-SQUARE ABOVE SINGLE RETURN PERIOD VALUES

TYPE OF DISTRIBUTION	VALUE FOR RETURN PERIOD OF			
	10 YR	20 YR	50 YR	100 YR
Log-Pearson III	424	536	1086	2338
Log-normal	197	206	197	170
Gumbel	259	275	347	453

TABLE B-3

NUMBER OF OBSERVATIONS ABOVE RETURN PERIOD VALUES

TYPE OF DISTRIBUTION	NO. FOR RETURN PERIOD OF			
	10 YR	20 YR	50 YR	100 YR
Log-Pearson III	912	565	418	381
Log-normal	881	440	202	129
Gumbel	724	371	145	73
"Expected" no.	1029	514	206	103

values (0.98 percentile point) when only 206 were expected. This indicates that the log-Pearson Type III distribution return period values are too low. The third line indicates that the Gumbel values are too high, because only 145 of the observations fall above the computed value, whereas 206 would be expected. Of the three, the log-normal distribution gave the best results.

The binomial test, rather than the chi-square test, can be applied to Table B-3 as an additional test of the goodness of fit of the distributions. Eq. B-14 was applied to each of the four values for each function in Table B-3. It will be recalled that in an earlier section it was shown that B is approximately normally distributed. Under the hypothesis that the distributions do fit the observations, the probability of obtaining an absolute value of $B \geq 1.96$ is ≤ 0.05. Table B-4 gives the values obtained from Eq. B-14 for each function. This test shows the poorer fit by the log-Pearson Type III distribution function to peak runoff data from small rural watersheds, because all values lie outside the specified range. It verified that the log-Pearson values are too small for the longer return periods,

TABLE B-4

BINOMIAL TEST APPLIED TO DATA IN TABLE B-3

TYPE OF DISTRIBUTION	VALUE FOR RETURN PERIOD OF			
	10 YR	20 YR	50 YR	100 YR
Log-Pearson III	−3.9	2.3	14.9	21.4
Log-normal	−4.5	−3.3	−0.3	2.6
Gumbel	−10.0	−6.5	−4.3	−3.0

because the B-values get progressively larger. Smaller values of B were obtained for the log-normal technique for each return period, compared to the Gumbel method, substantiating that the log-normal distribution function fits these data better than either the Gumbel or the log-Pearson Type III methods. On the basis of these tests, log-normal was selected for use in the study to estimate peak discharge return period values for small rural watersheds.

Many more data were used in this investigation than in other studies. However, the issue is far from closed. Considerable additional work could be done, particularly on the geographical variation of skewness and possible combinations of data from several basins to obtain more stable estimates of skewness coefficients. Further, although on the average the log-normal distribution gave the best results, there were many basins where the fit was not as good as desired. It should be recalled that the objective here is not to study goodness of fit *per se,* but rather to choose a distribution function in a logical manner for use in this study. This rather extensive investigation has achieved that result. In addition, a contribution has been made to the difficult problem of fitting distribution functions to annual peak runoff amounts.

APPENDIX C

ANALYSIS OF SHORT-DURATION PRECIPITATION AMOUNTS FOR VARIOUS RETURN PERIODS

Inasmuch as the primary purpose of this study was to develop methods for estimating runoff rates from small rural watersheds, it was believed that information related to short-duration precipitation amounts might prove to be useful. However, almost all of the rural stream gaging stations used in this study do not have a rain gage nearby; therefore it would be necessary to transpose the precipitation data horizontally to stream gage stations. It was decided to do this in a systematic, objective manner that would take into account variations in precipitation amounts between measurement locations. This was accomplished in a three-step approach. First, the precipitation records were analyzed using the Gumbel frequency distribution method to determine precipitation amounts for various return periods. Second, objective mappings were prepared of precipitation amounts for various short

durations and various return periods. Finally, the short-duration precipitation data were determined for each of 493 watersheds by interpolating to each location on the objectively drawn maps. This appendix describes the methodology for preparing the analyses and a discussion of the maps. It is thought that the maps may prove to be useful in other hydrologic investigations.

The Gumbel distribution method was selected to determine return period precipitation values because it had been used previously by Hershfield (*61*) in his preparation of *U.S. Weather Bureau Technical Paper 40*. This reference, which includes maps for precipitation durations ranging from 30 min through 24 hr, is widely used in hydrological applications and so it was chosen as a guide and control over this study, which is essentially an extension of Hershfield's earlier efforts.* The Gumbel method is described in the next section.

The objective analysis procedure that was selected to generate the precipitation maps is called the Conditional Relaxation Analysis Method (CRAM). It is a computerized procedure that has been used on several analysis problems at the research agency conducting the present study. The CRAM technique is described in detail subsequently under "Development of Rainfall Maps."

GUMBEL DISTRIBUTION METHOD

The Method

In many hydrology or hydrometeorological problems there is concern for statistics related to the recurrence of events exceeding some threshold value(s). Such statistics are generally expressed in terms of return period (or recurrence interval), expressed as

$$T(X) = \frac{1}{1 - F(X)} \qquad \text{(C-1)}$$

in which $F(X)$ is the probability that the values of a variable, x, are equal to or smaller than a specified value, X. This can be represented by a cumulative distribution function (in this case the Gumbel function) that has been fitted to a series of annual values of the variable. In words, $T(X)$ is the number of years such that, on the average, there would be just one value of the variable equalling or exceeding the fitted value X. In general, values of X are computed for assigned values of $T(X)$.

The cumulative distribution function of the Gumbel method is

$$F(X) = \exp\left(-e^{-Z}\right) \qquad \text{(C-2)}$$

$$Z = \alpha\,(X - u) \qquad \text{(C-3)}$$

The intermediate variable Z is often referred to as the "reduced" variable, α is the scale parameter, and u is the mode of the Gumbel distribution.

Given a series of N annual values of maximum rainfall intensity for a given time interval, such as 15 min, defined by $x_1, x_2, x_3, \ldots x_N$; the parameters α and u are estimated from the data. First the mean and standard deviation are computed for the series from:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \text{(C-4)}$$

and

$$s_x = \left[\frac{1}{(N-1)}\sum_{i=1}^{N} (x_i - \bar{x})^2\right]^{1/2} \qquad \text{(C-5)}$$

Then α and u are computed from

$$\alpha = \frac{\sigma(N)}{s} \quad \text{and} \quad u = \bar{x} - y(N)/\alpha \qquad \text{(C-6)}$$

in which

$\sigma(N)$ and $y(N)$ are obtained from published tables (*26*).

In application, the Gumbel method is fitted to a series of data by the following steps:

1. Initially values of $T(X)$ are assigned; in the present case the values desired were 2, 5, 10, 25, 50, and 100 years.
2. Corresponding values of $F(X)$ are then determined from Eq. C-1 (e.g., for $T(X) = 50$ years, $F(X) = 0.98$).
3. Solve Eq. C-2 to obtain a series of $Z_{F(X)}$-values, which are then entered in Eq. C-3.
4. Solve for α and u from Eq. C-6 as functions of the series size (N).
5. Compute the return period precipitation values, $X_{F(X)}$, from Eq. C-3.

Figure C-1 shows the result of such an analysis of 15 years of maximum annual 15-min precipitation amounts for Bridgeport, Conn. The actual values have been plotted at their computed position, $(N + 1)/m$, in which N is the total number of amounts and m is their rank in descending magnitude and the Gumbel at their return period (years) position.

Gumbel Analysis

The data on which this study was based were the maximum annual precipitation records obtained from the Office of Hydrology (ESSA Weather Bureau). These data include the date of occurrence and amount of highest 5-, 10-, 15-, 30-, 60-, 120-, and 1440-minute precipitation for each year of operation of some 175 to 200 first-order Weather Bureau stations. For the current over-all purposes, the data for 60-min duration and less for some 167 stations were used.

Figure C-2 shows the spatial distribution of the 167 precipitation stations. The minimum number of annual values for any station is 15 and most of the stations have in excess of 30 years of data.

The Gumbel distribution method was programmed for the research agency's IBM 360-40 computer and applied to the 5-, 10-, 15-, 30-, and 60-min maximum annual precipitation values for each station to obtain return period (or recurrence interval) estimates for 2, 5, 10, 25, 50, and 100 years. The fitted return period values, $X_{F(X)}$, in Eq. C-3 thereby provided the input data to the objective analysis procedure.

Goodness of Fit

Prior to conducting the objective analysis phase of the study, the "goodness of fit" of the Gumbel distribution to the annual values was investigated. The chi-square and

* The researchers are grateful to Mr. Hershfield for encouraging this effort and for many fruitful discussions. Any errors are, of course, the researchers' own.

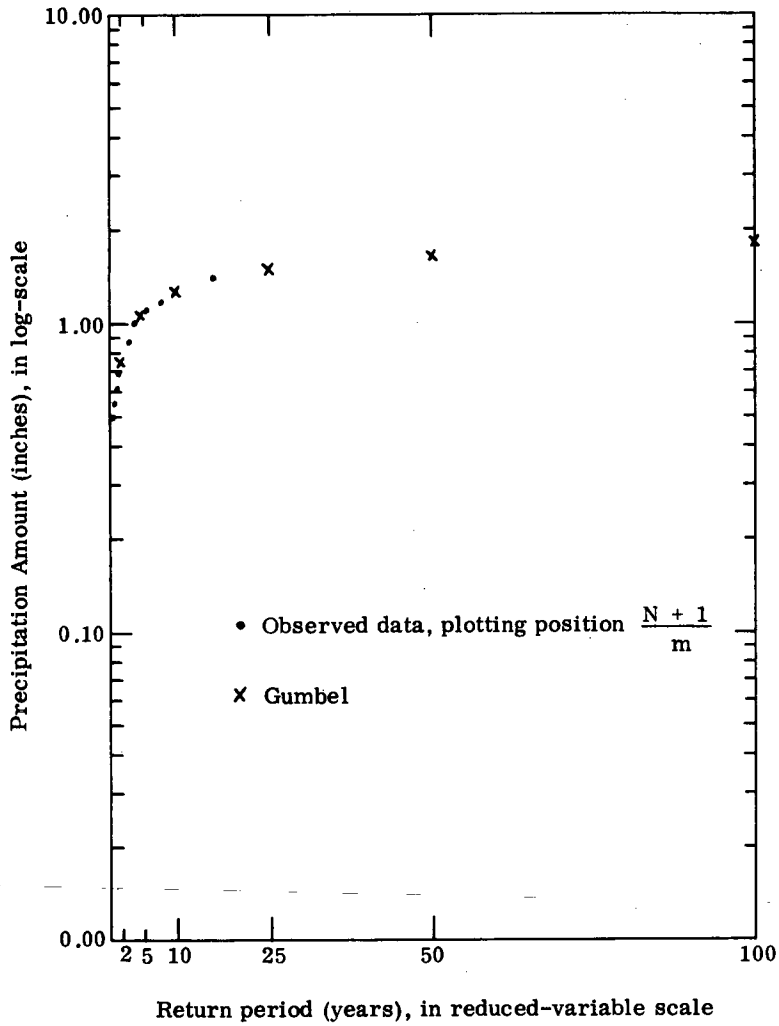


Figure C-1. Frequency plot of 15-min precipitation amounts at Bridgeport, Conn., estimated by Gumbel method and by (N+*1*)/m *plotting position formula.*

binomial tests were used for this purpose. For simplicity of notation p was used in place of $F(X)$ and X_p is the pth percentile. If the fitted distribution fits the observations perfectly, $100 \cdot p$ percent of the observations will be equal to or less than X_p. Similarly, $100 \cdot (p_2 - p_1)$ percent of the observations will lie between X_{p_2} and X_{p_1}.

To apply the chi-square test, a series of p-values is specified, denoted by $p_1, p_2, \ldots p_m$, where m is the number of p-values. Let $p_0 = 0$ and $p_{m+1} = 1$. The m percentile values are then computed from the data, $x_{p_1}, x_{p_2}, \ldots x_{p_m}$. Let $x_{p_0} = -\infty$ and $x_{p_{m+1}} = +\infty$.

Then compute

$$\chi^2 = \sum_{j=1}^{m+1} \left[[n_j - N(p_j - p_{j-1})]^2 / N(p_j - p_{j-1}) \right] \quad \text{(C-7)}$$

in which n_j is the number of observations falling in the jth group (i.e., lying between X_{p_j} and $X_{p_{j-1}}$), N is the total number of observations, and $N(p_j - p_{j-1})$ is the expected number in the jth group.

The binomial test can be applied in those cases where only one percentile value has been computed. Herein, let X_p be the percentile value. If a success is defined as an event when an observation exceeds X_p, the probability of success equals $1 - p$. Further, let W be the number of trials and w be the number of successes. With large W (the case in this study) the normal approximation to the binomial is quite accurate. Thus,

$$B = \frac{w - (1-p)W}{[p(1-p)W]^{1/2}} \quad \text{(C-8)}$$

is the normal deviate with mean of 0 and a standard deviation of 1.

From a hydrological viewpoint it is important that the Gumbel distribution fit the observed values at percentiles higher than $X_{0.90}$, which corresponds to a 10-year return period. Consequently, four percentiles were chosen for investigation corresponding to return periods of 20, 25, 50, and 100 years. Each percentile was used separately to divide the data into two groups (i.e., individual values of precipitation are either above or below the specified per-

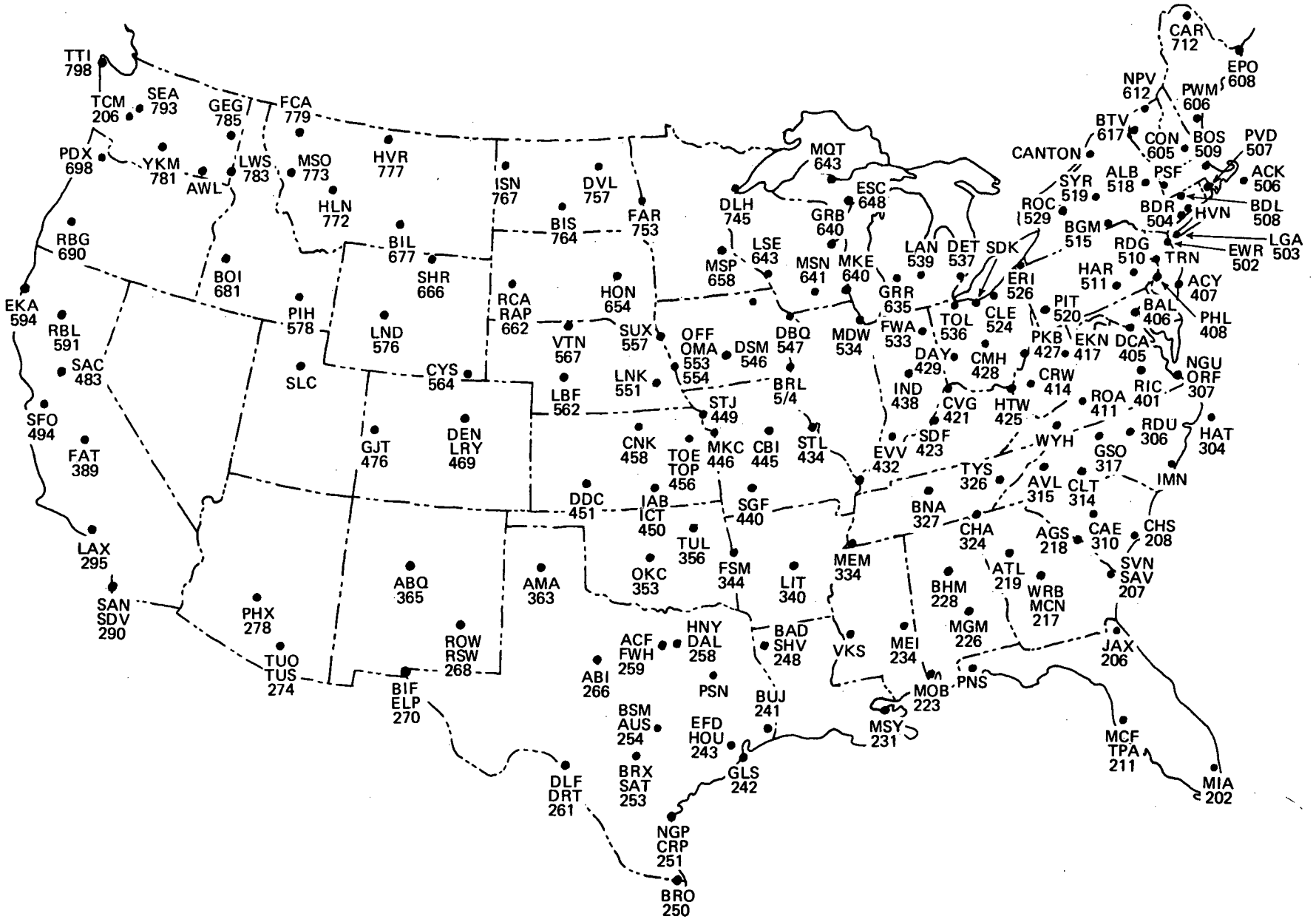


Figure C-2. Geographic distribution of 167 precipitation stations.

centile). Chi-square was computed for each of the series of data for each two-group separation. A sum of chi-square values was obtained for each duration for each percentile (i.e., 167 chi-square values were summed). The results are given in Table C-1.

From theoretical considerations, each sum in Table C-1 should be less than 167 if the Gumbel fits closely (see Appendix B for reasons). All of the sums meet this criterion, so it appears that the Gumbel distribution fits quite well to annual maximum precipitation amounts for durations of 1 hr and less.

The lack of exactitude in the "chi-square to be expected" on theoretical grounds led to another test. Using only the 136 stations with 50 years or more of record, each station's data were divided into two parts by selecting at random 20 percent of the observations. The Gumbel distribution was fitted to the remaining 80 percent and four percentile values corresponding to return periods of 20, 25, 50, and 100 years were computed.

Each such percentile value was used to compute a chi-square value on the independent data; i.e., the 20 percent selected at random. Whenever independent data were used, it is known that the degrees of freedom equal $M - 1$, or 1, inasmuch as M, the number of groups, is 2.

Table C-2 gives the sums of chi-square for the 136 stations having 50 years or more of record. The values in Table C-2 average out to 131, which is quite close to the expected value of 136.

Table C-3 gives the actual number of independent data observations at all stations falling above the stated percentile. For example, line one shows that there were 1,661 observations of annual maximum 5-min duration precipitation amounts in the independent samples of all 136 stations. Of these 1,661 observations, 81 equalled or exceeded their respective $X_{0.95}$ percentile value, $57 \geq X_{0.96}$, $22 \geq X_{0.98}$ and $13 \geq X_{0.99}$.

The binomial test, rather than the chi-square test, can be applied to these numbers to test the goodness of fit of the Gumbel distribution. Thus, Eq. C-8 was applied to each of the 20 values in Table C-3. Under the hypothesis that the Gumbel distribution does fit the observations, the probability of obtaining a value of the normal deviate $B \geq 1.96$ is ≤ 0.05. Only one B-value out of 20 exceeded 1.96, which is not an unusual result because with 20 trials one exceedance is to be expected. Again, this is strong evidence that annual maximum precipitation amounts for short durations are fitted well by the Gumbel distribution.

TABLE C-1

SUM OF CHI-SQUARE VALUES FOR 167 STATIONS

DURATION (MIN)	VALUE FOR RETURN PERIOD OF 20 YR	25 YR	50 YR	100 YR
5	99	104	102	84
10	84	92	101	87
15	73	82	96	93
30	67	77	87	91
60	65	70	78	91

TABLE C-2

SUM OF CHI-SQUARE VALUES FOR 136 STATIONS

DURATION (MIN)	VALUE FOR RETURN PERIOD OF 20 YR	25 YR	50 YR	100 YR
5	177	175	102	108
10	117	108	113	143
15	139	102	108	98
30	143	113	112	159
60	127	117	168	193

TABLE C-3

SUM OF NUMBER OF INDEPENDENT OBSERVATIONS FOR 136 STATIONS

DURATION (MIN)	NUMBER FOR RETURN PERIOD OF 20 YR	25 YR	50 YR	100 YR	TOTAL
5	81	57	22 [a]	13	1661
10	73	52	26	16	1663
15	71	53	24	12	1604
30	84	60	28	17	1610
60	76	62	36	21	1671

[a] Significant at 0.05 level.

DEVELOPMENT OF RAINFALL MAPS

The Objective Analysis Procedure

The problem of interpolating values at a set of equally spaced points from a set of observations at unequally spaced points is quite common and important in meteorology. Several mathematically sophisticated and very efficient programs are available for performing the necessary calculations by computer. The method used here transforms observations to points on an equally spaced grid and outputs two products—contoured maps and a list of values interpolated from the map at any specified set of locations (watersheds in the current case). Thus, the 167 values of one set of precipitation duration-frequency, Pf-d, are read into the computer, which produces maps of the United States with isopleths of Pf-d and a list of values of Pf-d, one value for each of the 493 watersheds. This program was used to produce 30 maps of Pf-d, covering six return periods and five durations. The objective here is to give a brief description of the method, the Conditional Relaxation Analysis Method (CRAM), which was designed and written by James Welsh of The Travelers Research Center based on the analysis method originally proposed by Carstensen.

The CRAM procedure (*41, 56*) requires that the analyzed gridpoint values satisfy Poisson's equation. The method used for translating the observations to gridpoints requires a representative initial guess of the short-duration precipitation frequency value at each gridpoint. The first

phase of the analysis consists of correcting the initial guess values by using the observations. Where more than one observation influences a gridpoint value, the mean of the differences between the initial guess and observations is used to correct the initial guess. The value at the corrected gridpoint is then treated as an internal boundary value.

When the initial guess, the perimeter boundary values, and the internal boundary values are defined, the analysis is computed by requiring all nonboundary gridpoint values to satisfy Poisson's equation

$$\nabla^2 R(i, j) = F(i, j) \qquad \text{(C-9)}$$

in which R is the analyzed precipitation at gridpoint (i, j), F is a forcing function defining the shape of the precipitation field (computed as the Laplacian of the smoothed initial guess), and ∇^2 is the finite-difference Laplacian operator.

Eq. C-9 is solved by a relaxation procedure. In such a procedure a number of "passes" are made over the data where at each pass the equation is solved at each nonboundary point. A number of passes are required because solving the equation at one point causes a slight change in the values at adjacent points, thus causing Eq. C-9 not to hold exactly at such previously solved points. At pass n, the value of $R(i, j)$ is equal to the value of $R(i, j)$ at pass n-1 plus a small increment, which can be either positive or negative.

$$R(i, j)^{(n)} = R(i, j)^{(n-1)} + \alpha \nabla^2 R(i, j)^{(n)} - F(i, j) \qquad \text{(C-10)}$$

The parameter α is termed the relaxation coefficient and is specified by the investigator for each analysis. The incremental change in value becomes smaller and smaller with each pass until it falls below some pre-set very small value, ϵ, which terminates the first phase of the analysis.

The field of gridpoint values computed by the foregoing procedure generally contains unwanted, unreal details introduced by the procedure. Such detail is reduced by smoothing the data using

$$\tilde{R}(i, j) = \frac{R(i, j) + b\,\bar{R}(i, j)}{1 + b} \qquad \text{(C-11)}$$

in which $\bar{R}$ is the mean R at the four gridpoints surrounding point $R(i, j)$ and b is a smoothing parameter that is chosen by the investigator.

A comparison is made between the smoothed field values and the original 167 station values. If the differences between the two sets is deemed to be too large, the entire procedure is repeated. The smoothed field values serve as input to the relaxation analysis procedure, which is then smoothed again. The resmoothed field is compared with the original 167 station values and again the differences between the two are examined. The process is repeated until a satisfactory agreement is obtained.

The parameters α and b are assigned at the discretion of the investigator and serve to control the amount of relaxation and smoothing in the analysis. In this study the 60-min data were analyzed first and α and b were varied until values were found that produced maps that were similar to those of Hershfield. These same values of α and b were retained for all the remaining analyses.

The Analyses

The analysis grid for this study contained 1,628 points (37 rows × 44 columns) with a grid interval such that the distance between gridpoints is approximately 50 nautical miles. Such a grid density is justified on the basis of the geographical distribution of observations that in most sections of the analysis area are 50 miles or so apart.

Development of the precipitation mappings was conducted in two phases. In the first phase the 60-min precipitation data (measured in inches) were used to arrive at analysis criteria (smoothing and relaxation parameters) that yielded analyses similar to those developed by Hershfield (*61*). In phase two, the objective analyses of the precipitation frequency data for durations of less than 60 min and maps of the ratio of precipitation frequency values for durations less than 60 min to the 60-min value were generated.

Figure C-3 is the 25-year, 60-min prccipitation map developed by Hershfield. This map was used as the control analysis for the experimental testing with the analysis method. The researchers' initial map (Fig. C-4) was generated with no smoothing applied to the analysis. It clearly shows that there is much more detail in this analysis than is present in the control analysis. This is particularly true over the region adjacent to and including the Appalachian Mountains. Trial analyses were then performed varying the smoothing parameter (b in Eq. C-11) until an analysis comparable in detail and features to the control analysis was achieved. This final 25-year 60-min rainfall map, produced by CRAM, is shown in Figure C-5. The similarity of this map to Figure C-3 is obvious.

The relaxation and smoothing parameters determined with the 25-year, 60-min analysis were used in the second phase of the analysis, in which objective mappings were generated for precipitation durations of 5, 10, 15, 30, and 60 minutes for return periods of 2, 5, 10, 25, 50, and 100 years, or 30 maps in all.* Each map was analyzed independently of the others, using the observations from the 167 stations, with the computer output being in two forms—contoured (banded) printout and cathode ray tube (CRT) microfilm and hardcopy. Figures C-6 through C-9 show, respectively, the analysis for 5-, 10-, 15-, and 30-min precipitation for a return period of 25 years developed from CRT microfilm. The computer-prepared base map for the analyses includes an outline of the United States (heavy-lined) and 5° latitude and longitude lines. The precipitation isopleths are in hundredths of an inch and the plotted values represent extrema points in the data field. Note the similarity in the geographical distribution from the shortest (5-min) to the longest (30-min) duration precipitation. The axis of maximum precipitation intensity lies up the Mississippi Valley, with a secondary axis over the Chesapeake Bay region in the lee of the Appalachians. The flat gradient across the Rockies to the West Coast can be attributed in part to the paucity of data in that region, as shown in Figure C-2. Unfortunately, the full value of short-duration precipitation as a predictor of stream run-

* The complete set of 30 precipitation maps reproduced from CRT microfilm is available to qualified researchers on a loan basis on request to the Program Director, National Cooperative Highway Research Program.

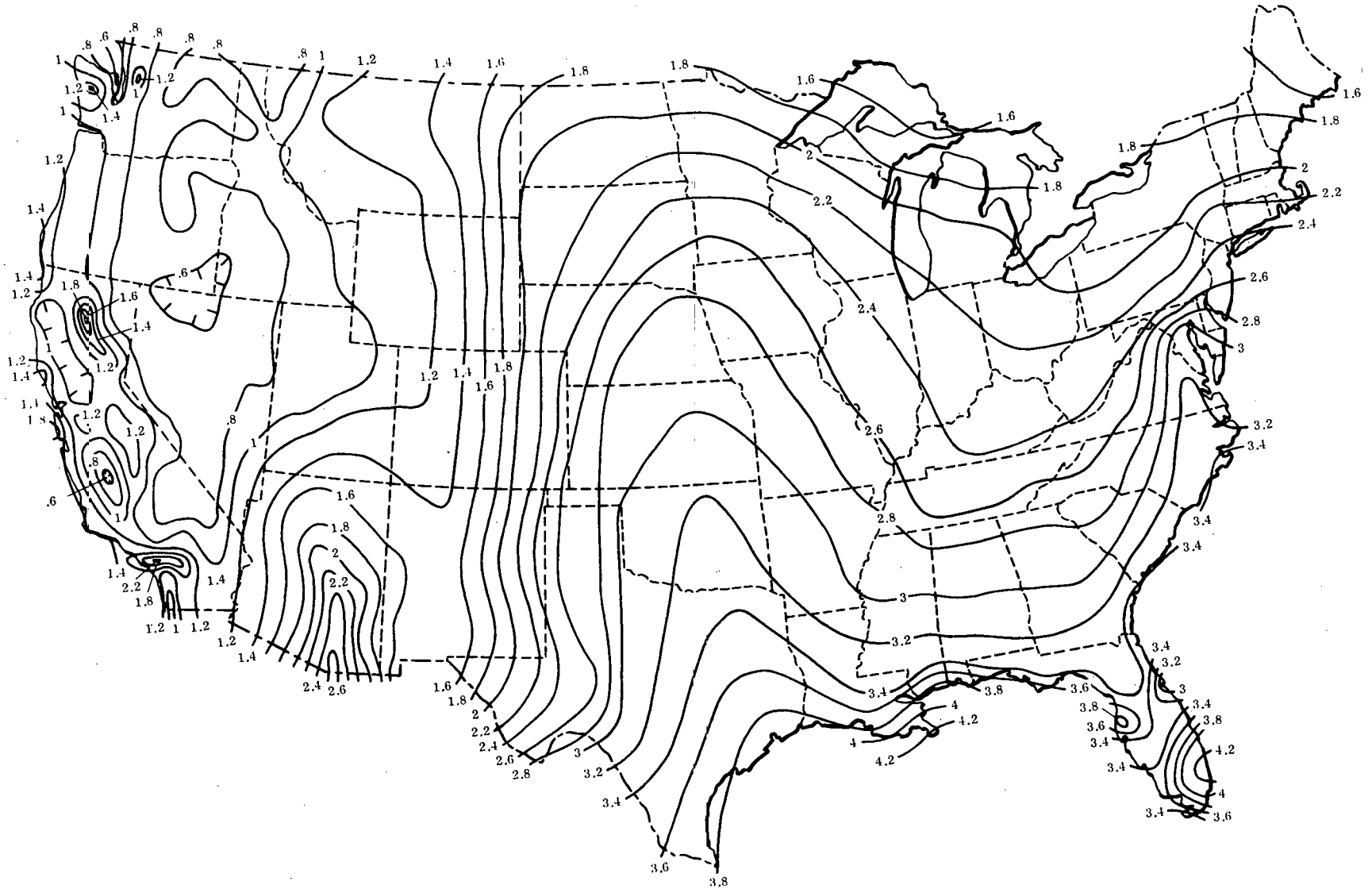


Figure C-3. Map of 60-min precipitation amount, 25-year frequency (P_{25-60}), by Hershfield. Used as the control in the study.

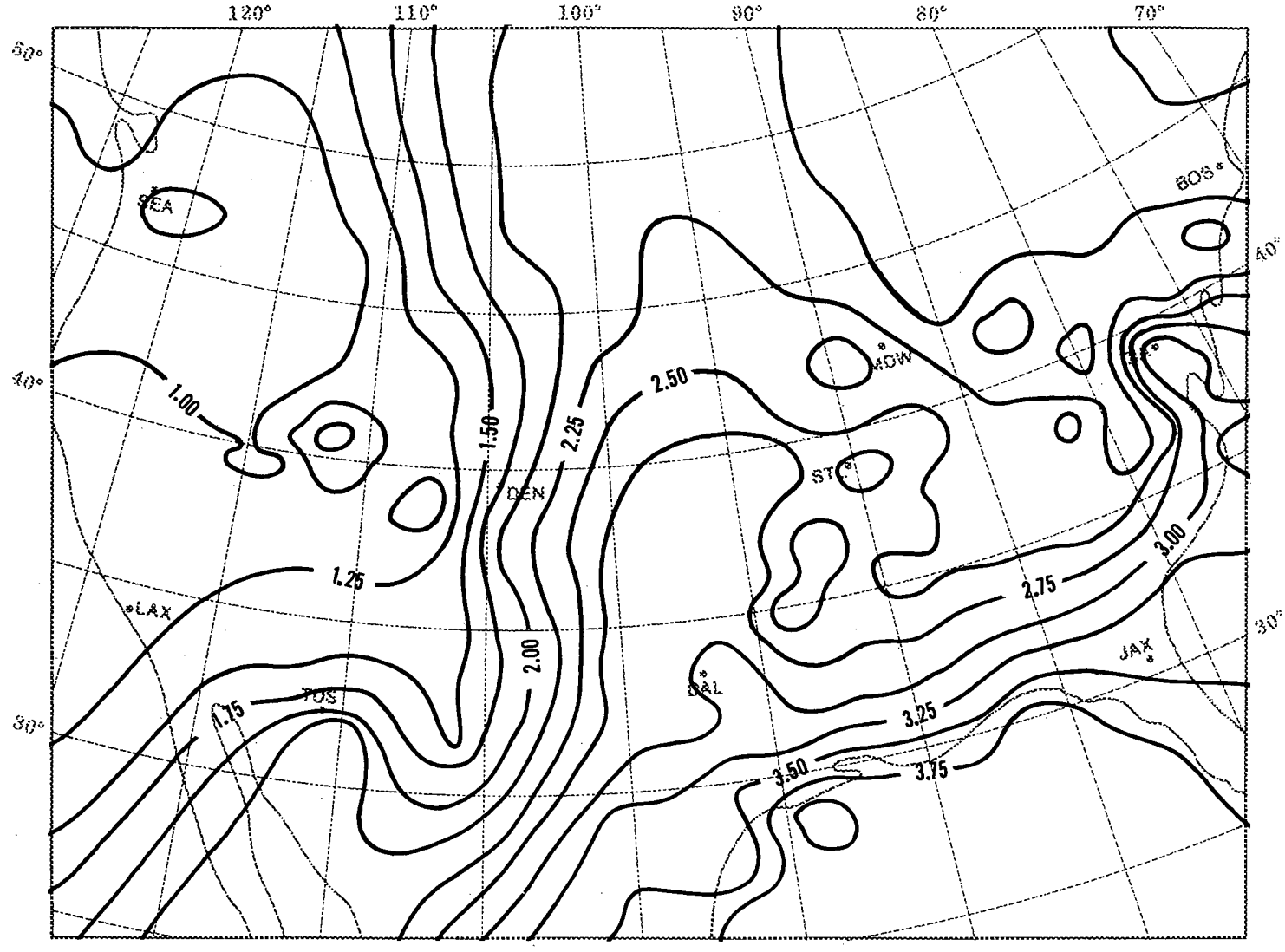


Figure C-4. Map of 60-min precipitation amount, 25-year frequency ($P_{25\text{-}60}$), *developed by using computerized CRAM, no smoothing (initial map).*

off may have been lessened in the regression experiments due to the paucity of data in the Far West, wherein a sizable portion of the stream runoff stations were located.

The primary justification for conducting this thorough analysis was that the procedure for obtaining short-duration precipitation return period values suggested in Ref. (*61*) uses a single multiplying factor for the U.S. for each duration period. These factors, which are multiplied by the 60-min precipitation amount to obtain the desired amount, are 0.29 for a 5-min duration, 0.45 for 10 min, 0.57 for 15 min, and 0.79 for 30 min. Because these factors are constants, the short-duration precipitation values would be statistically identical to the 60-min value as far as the regression technique used to develop the runoff relationships is concerned. The generation of ratio maps allowed determination of the extent to which the analyses that had been developed varied from an analysis based on a single ratio factor for the U.S. Some 24 ratio maps were prepared—four each for 2-, 5-, 10-, 25-, 50-, and 100-year return period.† The ratio maps for the 25-year return period are shown in Figures C-10 through C-13 corresponding to 5, 10, 15, and 30 min. The ratios are shown in whole percentage values (i.e., ratio times 100) with isopleths for each percentage unit. The base map is similar to that of Figures C-6 through C-9. Although the sharp gradients in the western third of the map must be discounted to a large extent due to the paucity of data, there is a fairly substantial variation over the eastern two-thirds wherein ample data were available. One can conclude from these results that the geographical variation is sufficient to justify taking it into account for certain applications, whereas for other,

† The complete set of 24 ratio maps reproduced from CRT microfilm is available to qualified researchers on a loan basis on request to the Program Director, National Cooperative Highway Research Program.

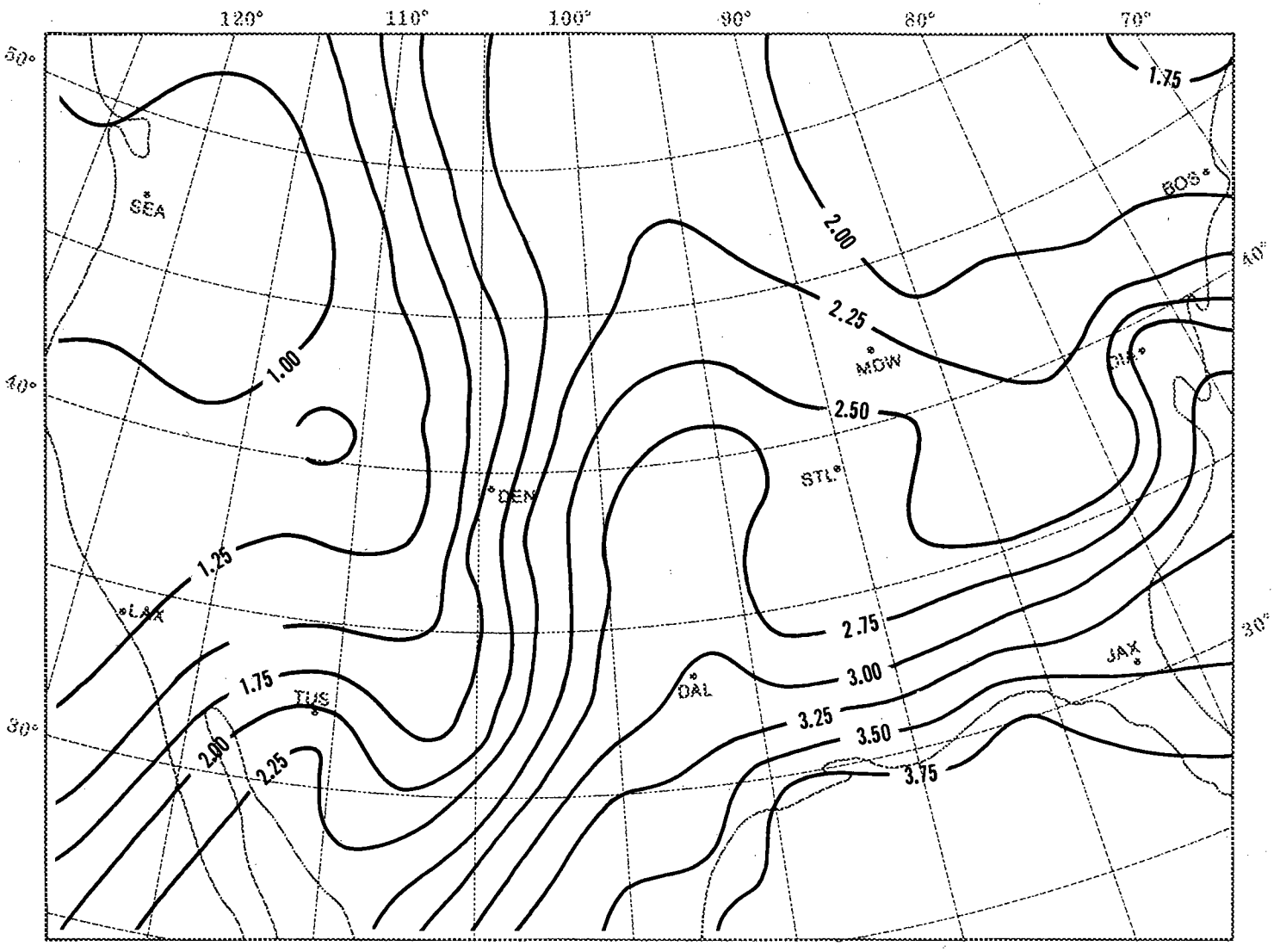


Figure C-5. Map of 60-min precipitation amount, 25-year frequency (P_{25-60}), *developed by using computerized CRAM, with smoothing (final map).*

less critical, needs a single value for the whole country may suffice.

The ratio of the 5-min amount to the 60-min amount (Fig. C-10), varies from about 0.20 over the southeastern states to about 0.32 over the Great Lakes. Although about one-half the ratio map has values greater than Hershfield's single value of 0.29, most of the region with ample data has a ratio less than this value. A similar pattern is noted in the 10-, 15-, and 30-min ratio maps: lowest values over the southeast; high over the Great Lakes; large, and apparently unreal, gradients over the Far West. Perhaps the relative ratio minimum of the southeast can be attributed to the incidence of tropical cyclones, which yield 60-min precipitation amounts proportionately larger than 60-min storms in other parts of the country.

The last phase of the computerized analysis procedure was to obtain rainfall values from the analyses by interpolating to each of the 493 stream runoff stations. These interpolated values, 30 for each watershed, were written on magnetic tape for use in the regression experiments.

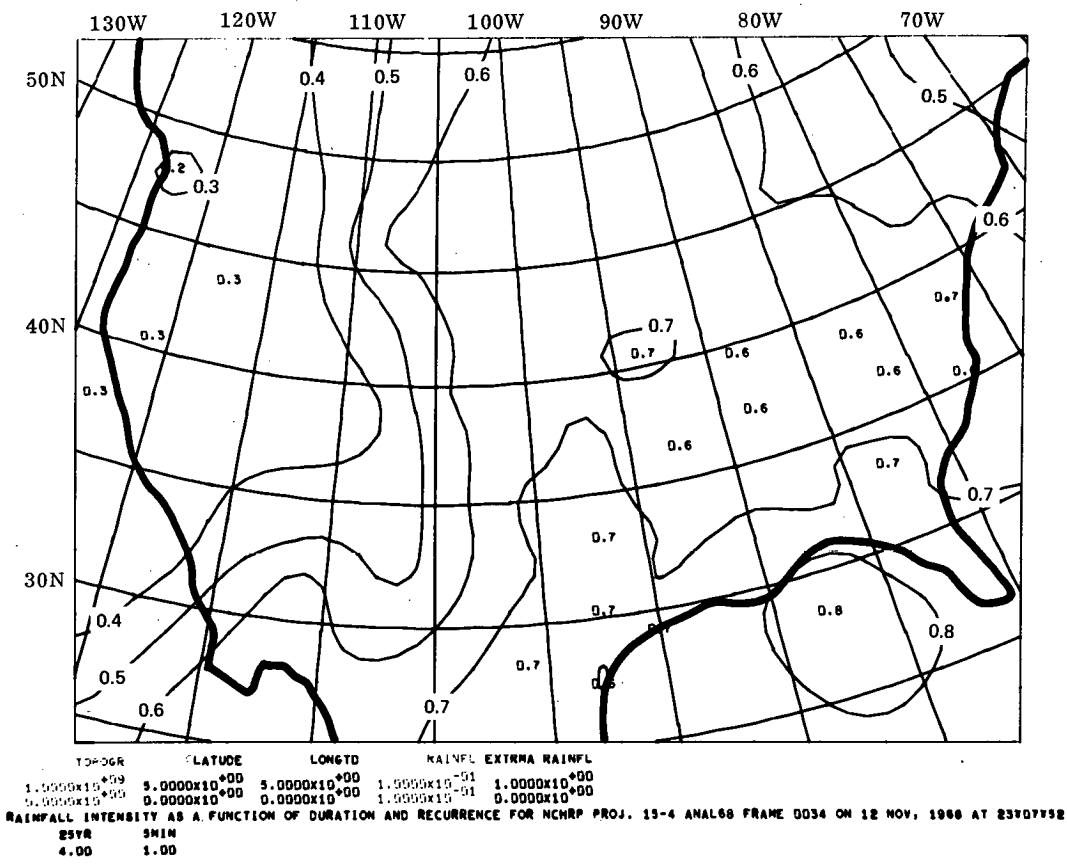


Figure C-6. Map of 5-min precipitation amount, 25-year frequency (P_{25-5}), *developed by using computerized CRAM.*

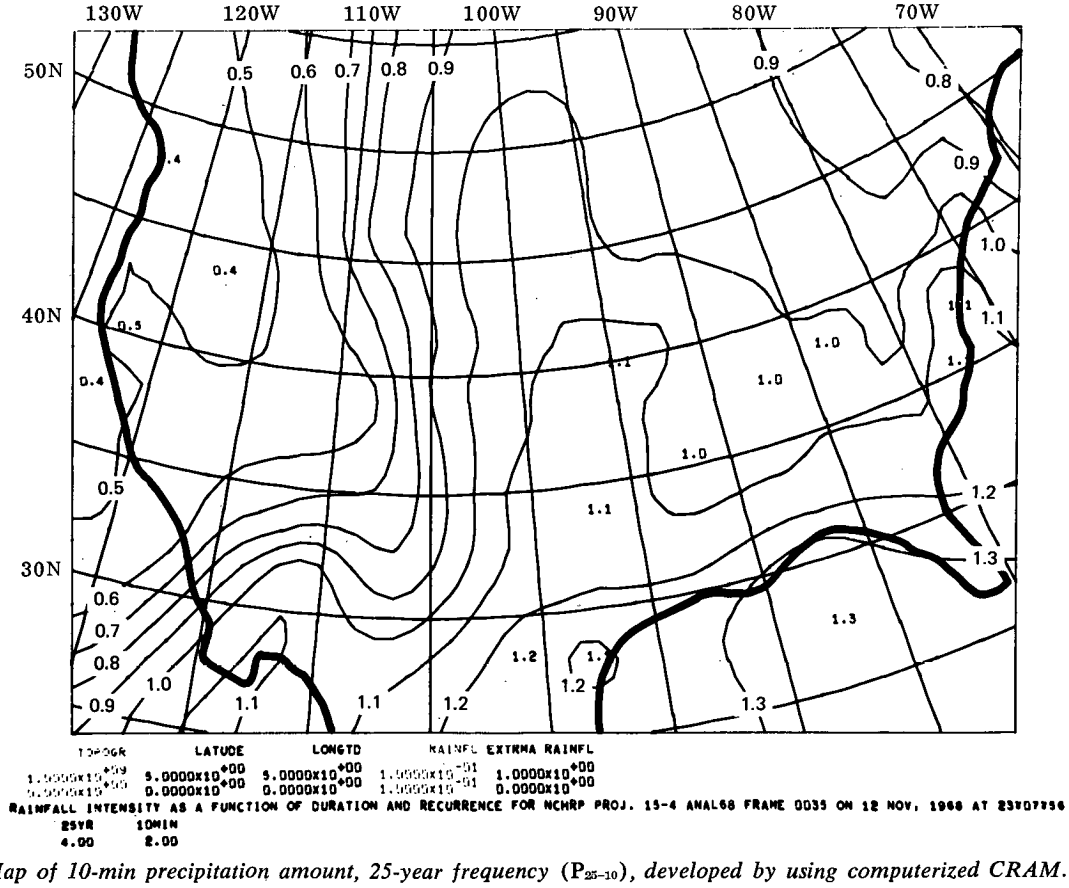


Figure C-7. Map of 10-min precipitation amount, 25-year frequency (P_{25-10}), *developed by using computerized CRAM.*

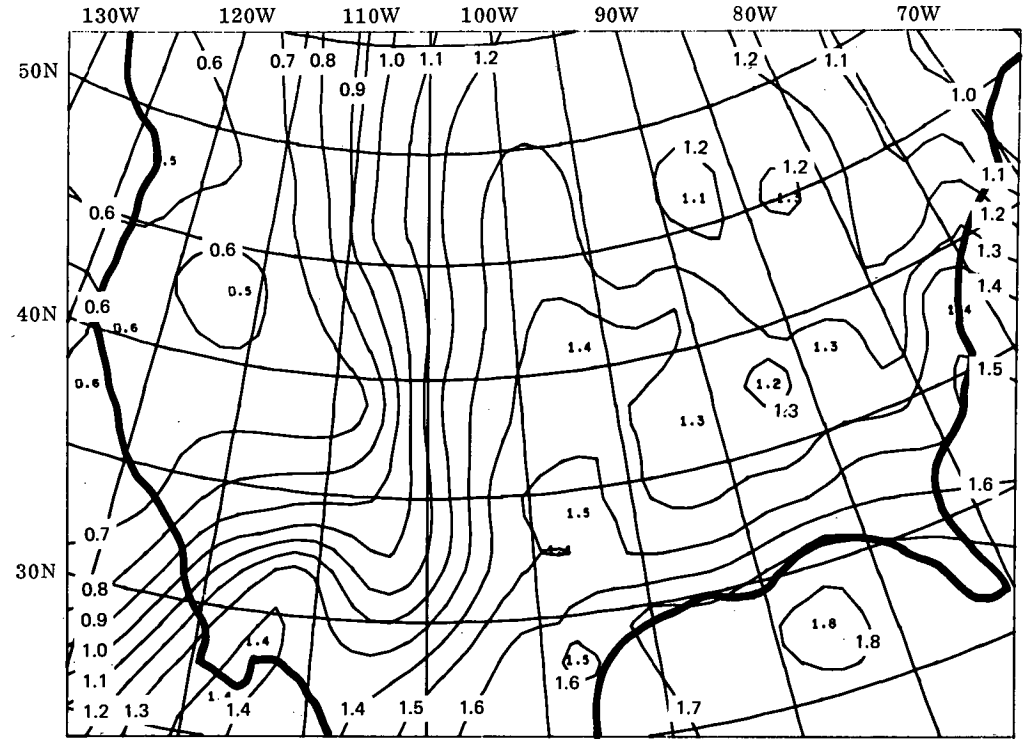


Figure C-8. Map of 15-min precipitation amount, 25-year frequency (P_{25-15}), *developed by using computerized CRAM.*

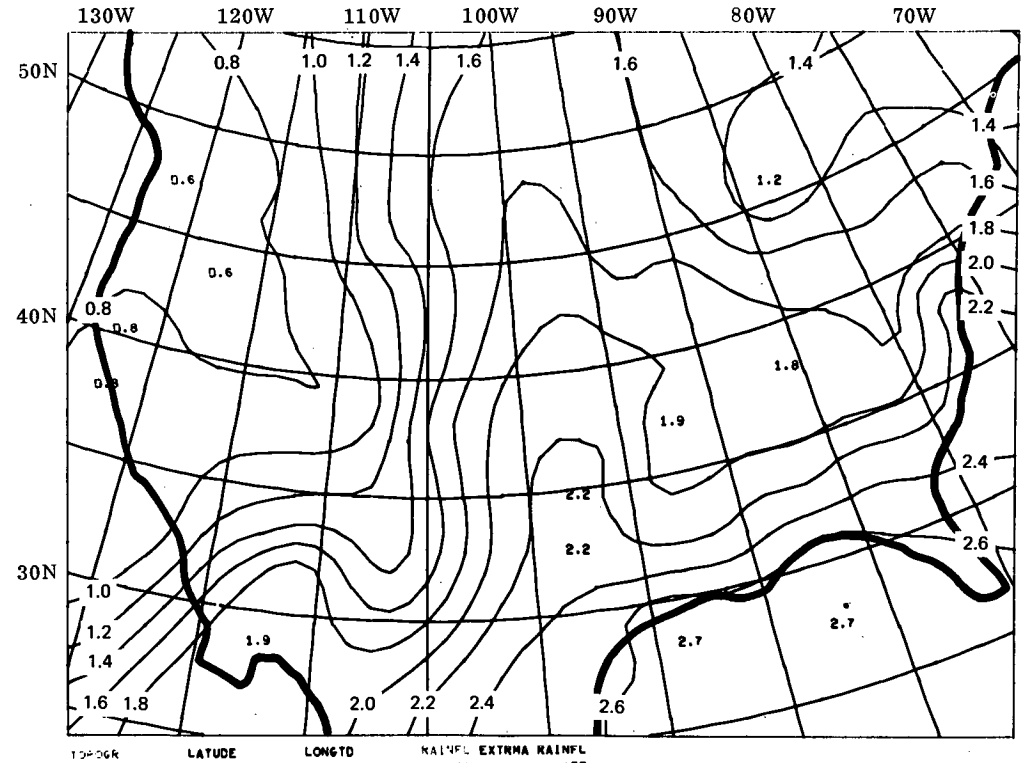


Figure C-9. Map of 30-min precipitation amount, 25-year frequency (P_{25-30}), *developed by using computerized CRAM.*

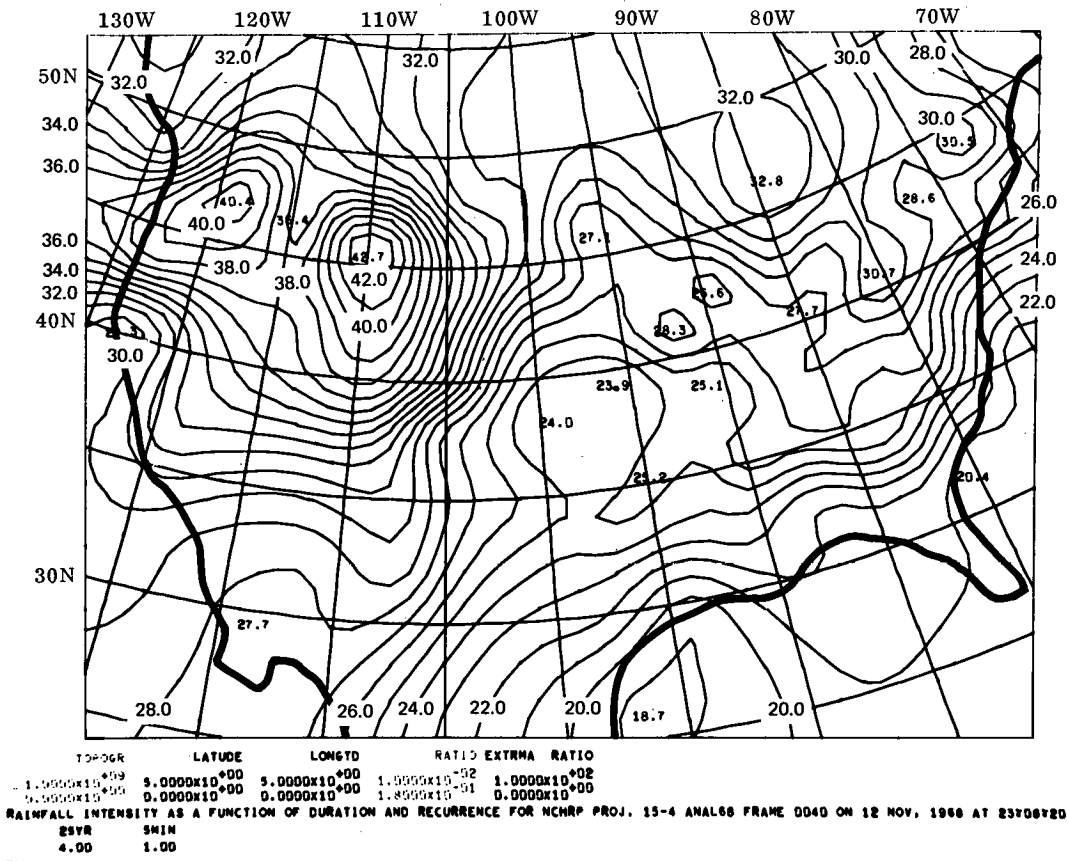


Figure C-10. Ratio map ($P_{25\text{-}5}/P_{25\text{-}60}$).

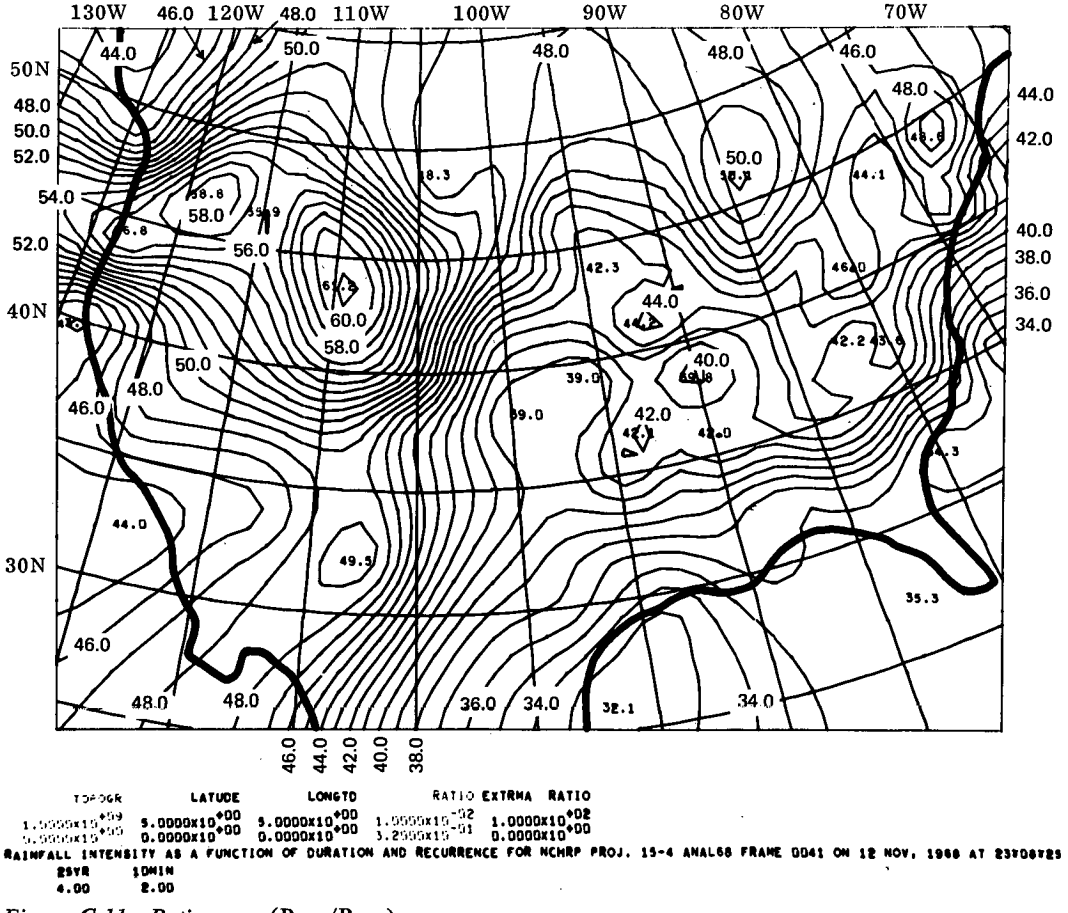


Figure C-11. Ratio map ($P_{25\text{-}10}/P_{25\text{-}60}$).

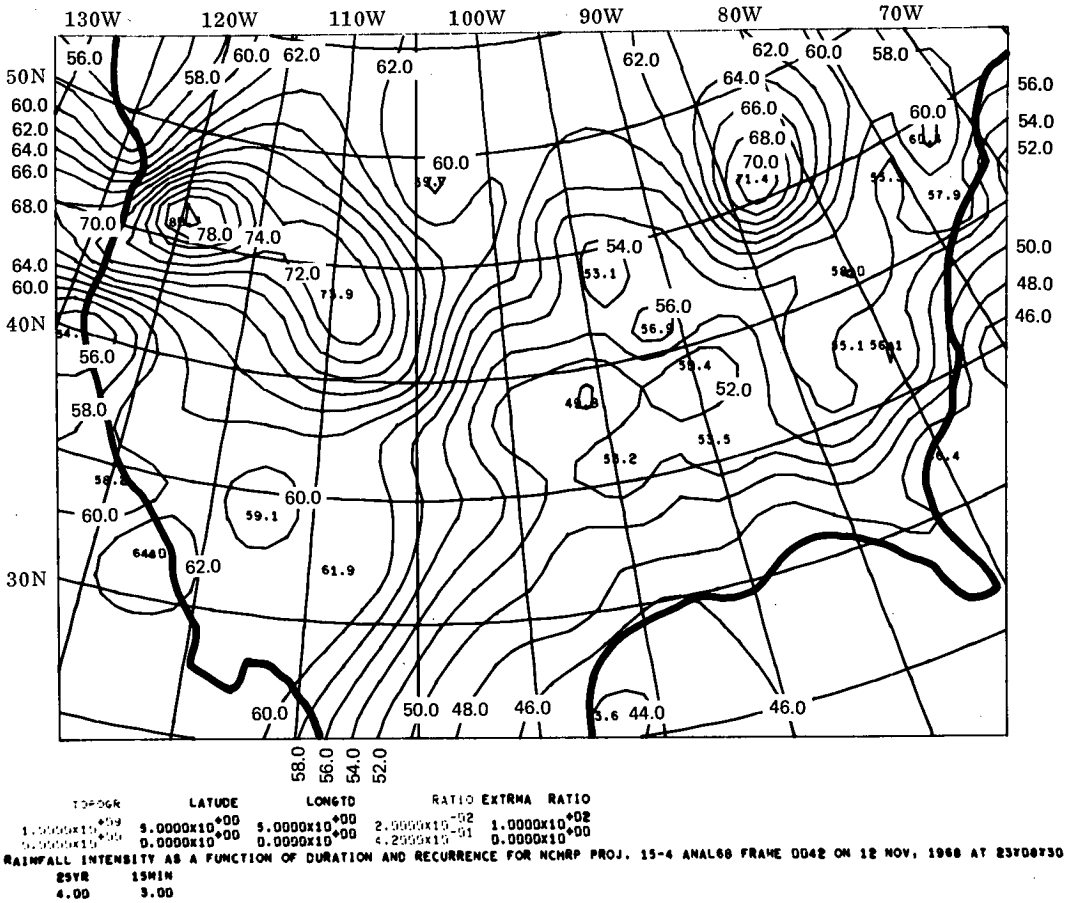


Figure C-12. Ratio map (P_{25-15}/P_{25-60}).

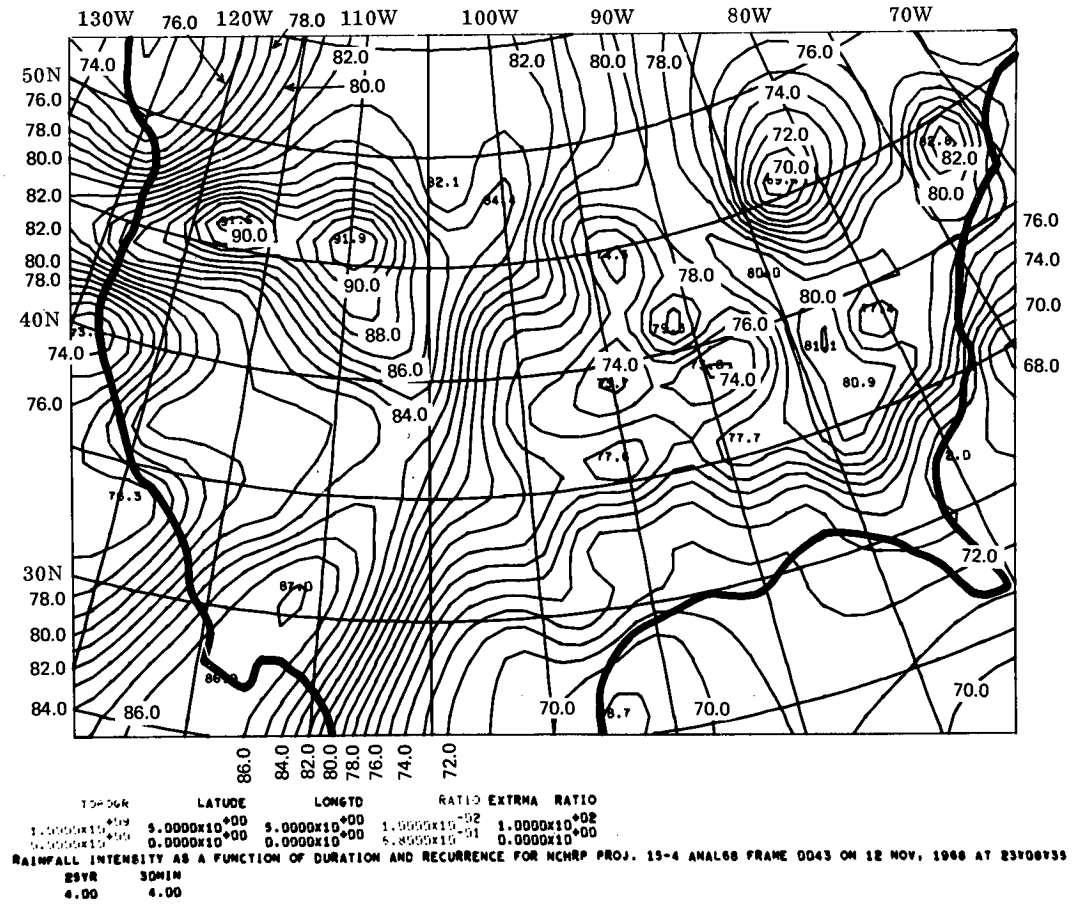


Figure C-13. Ratio map (P_{25-30}/P_{25-60}).

APPENDIX D

QUESTIONNAIRES

Two questionnaires were used in this study either: (1) to secure basic data that could not be obtained elsewhere or (2) to validate data obtained from other sources. The information contained in the returned questionnaires provided an essential check on the sources of national data. Many individuals either helped to formulate the questionnaires or completed and returned them.

One questionnaire (Fig. D-1) was sent to the U.S. Geological Survey District Offices for each of the original 1,000 or so watersheds considered in the study. The return rate on these exceeded 90 percent. In many cases comments appended to the questionnaire contained pertinent information regarding peculiar characteristics of a watershed that aided in the final selection of the study sample.

The other questionnaire (Fig. D-2) was sent to the Soil Conservation Service (SCS) state conservationist for each of the 493 watersheds that had been selected for study. The return rate on these was about 85 percent.

Information requested by The Travelers Research Center, Inc., Hydrology & Water Resources Division, 250 Constitution Plaza, Hartford, Connecticut, 06103; in connection with National Cooperative Highway Research Project No. 15-4.

Information concerning USGS basins less than 25 square miles

District Office_____

_____

1. Name of basin_____
2. USGS station number_____
3. Area_____
4. Would you characterize this basin as:
   (a) urban
   (b) semi-urban
   (c) rural

   If urban or semi-urban, please list the name of the city who should be contacted for further information concerning storm drains, sewers, etc. in the basin.

   If known, please indicate the approximate percentage of the total watershed area which is sewered.

5. Considering the existing man-made structures, e.g. highway embankments, dams, etc., would you recommend this basin for inclusion in a regional rainfall-runoff study for rural areas?

6. Is the watershed area predominantly
   (a) forest
   (b) grassland
   (c) cultivated

   If known, please indicate the approximate percentage of the watershed area falling in the above categories.

7. Any other remarks which in your opinion are pertinent to peak runoff characteristics of the watershed.

Figure D-1. Questionnaire sent to USGS District Offices covering some 1,000 watersheds considered in this study.

Please make reasonable judgment estimates in answering the three questions listed below. Attached is a topographic map with the watershed boundary indicated by a heavy black line. A small locator map is attached to the topographic map. The results of this effort will be directed toward improving runoff estimates for the design of culverts and also conservation measures.

Watershed Identification:

Longitude and Latitude of Basin Gage:

Area of Basin:

(1) What percentage of the watershed is covered by each hydrologic soil group?

Hydrologic soil groups	Percent of watershed
A	%
B	%
C	%
D	___%
	100%

(2) What is the predominant vegetation type in this watershed? (Check one.)

(a) Forest____ (b) Grassland____ (c) Cultivated and/or bare____ [1]

[1] Include "urban" and "semi-urban" with this vegetation type.

If more detailed information is available, what is the area covered by each type of vegetation within each hydrologic soil group?

Hydrologic soil groups	Percent area covered by		
A	Forest	____%	100%
	Grassland	____%	
	Cultivated/bare	____%	
B	Forest	____%	100%
	Grassland	____%	
	Cultivated/bare	____%	
C	Forest	____%	100%
	Grassland	____%	
	Cultivated/bare	____%	
D	Forest	____%	100%
	Grassland	____%	
	Cultivated/bare	____%	

(3) Is the watershed principally urban____, semi-urban____, or rural____? (Check one.)

(Urban is defined as areas drained by storm sewers; semi-urban areas partly drained by storm sewers; and rural, areas not sewered.)

If you can be more precise, indicate what percentage of the watershed is urban____, semi-urban____, and rural____.

Figure D-2. Questionnaire sent to Soil Conservation Service State Conservationists for each of the 493 watersheds selected for the study.

APPENDIX E

DESCRIPTION OF THE NATIONAL SMALL STREAMS DATA INVENTORY

The purpose of this appendix is to document and present the data that were collected and processed for the study reported herein. As such, these data constitute what the researchers have chosen to call the National Small Streams Data Inventory (NSSDI). The data of the NSSDI have been extracted from several sources, processed into a computer-usable form, and placed on punch cards and magnetic tape. The data included in the NSSDI can be grouped as follows: peak discharge, topographic, hydrologic-climatic, and physiographic. The specific variables included in each grouping are described in Appendix F.

Watersheds included in the sample were limited to:

1. Those gaged by the USGS.
2. Drainage area of 25 sq mi or less.
3. Rural characteristics.
4. Maximum annual peak discharge values available for 12 years or more.
5. Minimal diversion, regulation, or other man-made controls.
6. Minimal noncontributing drainage area (swamps, pondage, etc.)
7. Complete topographic map coverage of scale 1: 62,500 or smaller.
8. Those not disapproved by the local USGS district engineers.

Due to these limitations, the processed data sample comprised 493 watersheds selected from about 5,000 gaged watersheds that were potentially available. To the best of the researchers' knowledge, this represents the largest number of basins covering the broadest geographical region assembled to date in computer-usable form.

The basic data for each of the 493 watersheds have been tabulated and are available to qualified researchers on a loan basis on request to the Program Director, National Cooperative Highway Research Program. Magnetic tapes containing the complete data record for the 493 stations included in the listing and for another 179 stations meeting all requirements except that there were less than 12 annual maximum peak discharge values (but more than 5 values in each case), have been forwarded to the Office of Water Data Coordination, Department of the Interior, Washington, D.C., as part of the research agency's contractual commitment. Documentation of the tape format and contents have also been provided to OWDC. Inquiries regarding the availability and/or acquisition of copies of the data tapes should be directed to OWDC rather than NCHRP or the research agency. Besides the basic data presented in the listings, the data tape also includes, for each watershed, the magnitude and date of occurrence of each maximum annual peak discharge.

Assembling of the NSSDI was possible only because of the excellent cooperation received from several governmental agencies in providing the basic data used in the study. These agencies included:

1. U.S. Geological Survey (Department of the Interior):
   (a) Data Reports Unit (Washington, D.C.)
   (b) Surface Water Branch (Washington, D.C.)
   (c) Water Resources Division (State of Connecticut)
   (d) District Chiefs, Water Resources Division
2. ESSA Weather Bureau (Department of Commerce):
   (a) Office of Hydrology (Washington, D.C.)
   (b) Hydrometeorology Division (Washington, D.C.)
3. Soil Conservation Service (Department of Agriculture):
   (a) Soil Survey Operations (Washington, D.C.)
   (b) State Conservationist (Storrs, Conn. and Amherst, Mass.)
   (c) State Conservationists (various states)
4. Agricultural Research Service (Department of Agriculture):
5. State Highway Department
   (a) Connecticut (Hartford)
   (b) Missouri (Jefferson City)
   (c) Kansas (Topeka)
   (d) Texas (Austin)
   (e) Georgia (Atlanta)

DATA LISTINGS

The rationale used in presenting the basic data listings of the NSSDI was to exclude those variables on the data tape, such as logarithmic forms, that are obtained by one manipulation of a piece of basic data. Those "computed" variables obtained by two or more manipulations have been included. Similarly, the individual annual maximum peak discharge values have not been printed for two reasons: (1) they can be obtained conveniently either from published USGS surface water reports or from the magnetic data tape of this project on file at OWDC, and (2) they presented an awkward printing problem because of the variable number of values from one watershed to another. The printed data include 116 pieces of information for each of 493 watersheds.

The specific variables included in each of four defined groupings (peak discharge, topographic, hydrologic and climatic, and physiographic) are listed in the order they are presented in the data listing. Specific information regarding the geographic location of a watershed included latitude (3) and longitude (4) in hundredths of degrees, and gage elevation (5), in addition to its USGS station number (2) and name.

Peak discharge information included is summarized in Table E-1. Appendix F gives details regarding these variables.

The variables that were determined from USGS topographic quadrangle maps and included in this data listing are summarized in Table E-2.

Hydrologic and climatic variables in the data listings are summarized in Table E-3.

The physiographic (or soil) variables included in the data listings of the NSSDI are given in Table E-4.

TABLE E-1

PEAK DISCHARGE INFORMATION IN NSSDI [a]

DATA LISTING NO.	SYMBOL	VARIABLE NAME	UNITS	WHERE DISCUSSED
6	1st yr	First year of peak discharge	yr	App. F
7	No. yr	Total number of peaks used	yr	App. F
8	Gage type	Type of stream gage	None	App. F
		1—continuous recording		App. F
		2—crest-stage		App. F
		3—combination of 1 and 2		App. F
9	Q_2	Peak runoff—2 yr return	cfs	App. F
10	Q_5	Peak runoff—5 yr return	cfs	App. F
11	Q_{10}	Peak runoff—10 yr return	cfs	App. F
12	Q_{20}	Peak runoff—20 yr return	cfs	App. F
13	Q_{25}	Peak runoff—25 yr return	cfs	App. F
14	Q_{50}	Peak runoff—50 yr return	cfs	App. F
15	MAF_1	Mean annual flood (12 categories)	cfs	App. F
16	MAF_3	Mean annual flood (5 categories)	cfs	App. F

[a] National Small Streams Data Inventory.

TABLE E-2

TOPOGRAPHIC INFORMATION INCLUDED IN NSSDI [a]

DATA LISTING NO.	SYMBOL	VARIABLE NAME	UNITS	WHERE DISCUSSED
17	A	Drainage area	sq mi	App. F
18	L	Main stream length	mi	App. F
19	TRIB	Length of tributaries	mi	App. F
20	DD	Drainage density	mi^{-1}	App. F
21	SHAPE	Watershed shape	None	App. F
22	T	Travel time index	mi	App. F
23	EL_1	Stream elevation, headwater	ft	App. F
24	EL_2	Stream elevation—0.9(L)	ft	App. F
25	EL_3	Stream elevation—0.8(L)	ft	App. F
26	EL_4	Stream elevation—0.7(L)	ft	App. F
27	EL_5	Stream elevation—0.6(L)	ft	App. F
28	EL_6	Stream elevation—0.5(L)	ft	App. F
29	EL_7	Stream elevation—0.4(L)	ft	App. F
30	EL_8	Stream elevation—0.3(L)	ft	App. F
31	EL_9	Stream elevation—0.2(L)	ft	App. F
32	EL_{10}	Stream elevation—0.1(L)	ft	App. F
33	S_3	Stream slope	ft mi^{-1}	App. F
34	S_4	Stream slope	ft mi^{-1}	App. F
35	S_5	Stream slope	ft mi^{-1}	App. F
36	S_{10}	Stream slope	ft mi^{-1}	App. F
37	S	Stream slope	ft mi^{-1}	App. F

[a] National Small Streams Data Inventory.

TABLE E-3

HYDROLOGIC AND CLIMATIC VARIABLES INCLUDED IN NSSDI [a]

DATA LISTING NO.	SYMBOL	VARIABLE NAME	UNITS	WHERE DISCUSSED
38	MAT	Mean annual temperature	°F	App. F
39	P_{wet}	Mean wettest month	in.	App. F
40	P_{dry}	Mean driest month	in.	App. F
41	MAP-S	Mean annual precipitation	in.	App. F
42	M24P	Maximum 24-hr precipitation	in.	App. F
43	MAS	Mean annual snowfall	in.	App. F
44	M24S	Maximum 24-hr snowfall	in.	App. F
45	P-days	No. of .01″ precipitation days/yr	days	App. F
46	S-days	No. of 1″ snowcover days/yr	days	App. F
47	T-days	No. of thunderstorm days/yr	days	App. F
48	32F days	No. of days minimum temp. $< 32°F/yr$	days	App. F
49	MAP	Mean annual precipitation	in.	App. F
50	RH	Mean relative humidity	%	App. F
51	July T	Mean July temperature	°F	App. F
52	POT E-T	Potential evapotranspiration	in.	App. F
53–100	PF-D	Frequency-duration precipitation	in.	App. F

[a] National Small Streams Data Inventory.

TABLE E-4

PHYSIOGRAPHIC VARIABLES INCLUDED IN NSSDI [a]

DATA LISTING NO.	SYMBOL	VARIABLE NAME	UNITS	WHERE DISCUSSED
101	SA	Percent soil type A	%	App. F and Fig. D-2
102	SB	Percent soil type B	%	App. F and Fig. D-2
103	SC	Percent soil type C	%	App. F and Fig. D-2
104	SD	Percent soil type D	%	App. F and Fig. D-2
105	URB	Percent urban	%	Figs. D-1 and D-2
106	SUB	Percent suburban	%	Figs. D-1 and D-2
107	RUR	Percent rural	%	Figs. D-1 and D-2
108	FOR	Percent forest	%	Figs. D-1 and D-2
109	GRAS	Percent grassland	%	Figs. D-1 and D-2
110	CULT	Percent cultivated	%	Figs. D-1 and D-2
111	CS	Composite soil parameter	dimensionless	App. F
112	CC	Composite cover parameter	dimensionless	App. F
113	SO	Soil order category	dimensionless	App. F
114	SSO	Soil suborder category	dimensionless	App. F
115	SE	Soil erosion category	dimensionless	App. F
116	GZ	Geologic zone category	dimensionless	App. F

[a] National Small Streams Data Inventory.

APPENDIX F

THE DATA AND THE ASSEMBLY OF THE DATA SAMPLE

The research approach outlined in Chapter One required a large body of data in order to establish peak runoff prediction equations applicable to ungaged watersheds with widely differing hydrologic, physiographic, and climatic characteristics. These data essentially comprise the measured or estimated values of maximum peak annual runoff, the topographic characteristics of the watershed, climatic statistics, and soil parameters. Only a very small

percentage of small rural watersheds (less than 25 sq mi) in the United States are gaged. In fact, of hundreds of thousands of basins in this category, less than 5,000 are gaged by various federal, state, local, and private agencies. Because of the research approach used and the relevancy of the data available, the study has been based on rural stream locations gaged by the U.S. Geological Survey (USGS). Some of these represent locations having continuous recording stream gages; other locations use crest-stage gages.

An examination of the continuous recording gaging stations on watersheds of less than 25 sq mi maintained by the USGS revealed:

1. Of about 2,700 locations that have been in operation at least one year, approximately 1,250 are presently operating.
2. Of these stations, 700 or so have been operating for at least 15 years.
3. About 1,100 have data for 10 years or more.

These totals are complemented by another 2,000 or so crest-stage measurements varying in length from one year upwards, of which about 300 include data for 15 years or more. Thus, there were approximately 1,000 gaging stations on watersheds of less than 25 sq mi with a period of record that spanned at least 15 years. Of these 1,000 drainage basins, only a few have recording rain gages located within the basin. This lack of coordinated rainfall-runoff gaging on a national basis precluded the development of an adequate data base for precipitation amounts on basins gaged for streamflow. The process of selecting watersheds for the study from the 1,000 or so potentially available is discussed in Chapter One.

Although the main objective of the study was to derive equations to forecast peak runoff for different return periods for a given watershed, it was believed that the physiographic and other data collected for these watersheds for this purpose could provide a source of basic information for future research on this and other related subjects. These data have, therefore, been recorded separately. (See Appendix E regarding availability.)

To ensure that the watersheds selected and the data developed and processed for them would be of value for the specific purpose of the study, certain selection criteria were established. In certain instances, these criteria were revised according to the situations encountered. The following section describes the basic policy and criteria for data collection.

CRITERIA FOR THE SELECTION OF BASINS

The project as originally conceived defined a "small watershed" as "less than 20 square miles." However, during the actual selection of data it was felt that for the sake of a larger and more representative sample watersheds with areas between 20 and 25 sq mil could also be included in the study sample.

The data used in this study pertain to predominantly rural watersheds. A questionnaire (see Appendix D) issued to the district offices of the U.S. Geological Survey provided information regarding the estimated percentage of urban, semi-urban and rural areas within each of the watersheds. Those having an urban and semi-urban subarea of about 20 percent or more of the watershed were deleted from the study sample.

Because the primary objective of this study was to predict both the magnitude and the frequency (return period) of peak runoff, the minimum number of years of runoff record to use to estimate flood frequencies up to 50-year return period was established. The problem was addressed from a statistical and data availability point of view. Although it was of concern that the return period estimates be stable (based on sufficient data) it also was necessary to provide for enough basins to conduct the subsequent analyses. Based on these constraints a preliminary requirement was set of 15 years of annual peak discharge values for a basin to be considered. (According to a study by Benson, reported in Dalrymple (*18*), 18 years of records are required to define a 10-year flood within ±25 percent 19 times in 20. As is shown later, the average length of record for the 493 basins used in this study was 18.3 years.)

Examination of the initial geographical distribution of watersheds having a minimum of 15 years of data revealed a paucity of data in certain areas, particularly in the upper Midwest. It was found, however, that lowering the years-of-record requirement to 12 years in these regions would increase the number of gagings, particularly in those states wherein the cooperative program of installing and monitoring crest-stage gages between the local state highway departments and USGS had started back in 1954 or 1955.

Other constraints placed on watersheds for study selection included (1) complete coverage by USGS topographic quadrangle maps of a scale 1:24,000 or 1:62,500, and (2) minimum natural or man-made surface water control, such as regulation, diversion, pondage, or other non-contributing drainage area.

BASIN CHARACTERISTICS

Literally scores of basin or regional characteristics have an influence on the magnitude of a small stream flood. It was believed, therefore, that all reasonably relevant parameters should be considered in the study, allowing the hydrological statistical techniques to select the subset of parameters that contribute most to prediction of peak flows.

The parameters used in this study can be classified into three categories:

1. Topographic.
2. Hydrologic and climatic.
3. Physiographic (soils).

The various parameters included in each category and the processing executed to arrange them in acceptable form for computer analysis are discussed in subsequent sections.

Definition of Basin Characteristics

DRAINAGE AREA, A—Clearly, the drainage area of a watershed has an important effect on the magnitude of runoff from that watershed. The drainage areas published by

USGS were used. However, the watershed boundary was delineated on the USGS topographic maps and the areas were planimetered. This was done basically to check the delineated watershed boundary, which could be used for the development of other topographic variables. The drainage area is expressed in square miles.

LENGTHS OF MAIN STREAM, L, and TRIBUTARIES (TRIB); and DRAINAGE DENSITY (DD)—The blue line depicting the main stream of the watershed on the USGS topographic map was extended to the watershed boundary. Herein, the main stream is defined as the stream that drains the greatest area. The length of the main stream was determined by summing a series of short straight lines by use of a worksheet and a topographic map.

It has been widely known that the travel time of flow in a watershed (stream) is related to the length of the main stream and the lengths of tributary streams. The length of each tributary to the main stream was also measured using the same worksheet. The tributary blue lines on the topographic maps were not extended to the watershed boundary.

From the lengths of the main and tributary streams the following variables were developed:

$$\text{Total length of streams in the basin} = \text{TRIB} + L \qquad \text{(F-1)}$$

$$\text{Drainage density (DD)} = (\text{TRIB} + L)/A \qquad \text{(F-2)}$$

STREAM SLOPE—The development of stream slope parameters was limited to the main stream, not giving any specific consideration to tributary streams. In general, the slope of the main stream varies throughout its length, but it is desirable to develop a single measure of slope that represents the whole stream. Several such measures are possible.

The main stream length can be subdivided into ten equidistant segments. Then, straight lines extending from the outfall (stream gage) represent the slope between the upstream end of each segment and the outfall (i.e., $0.7L$ represents the stream slope based on seven-tenths of the main stream length).

The over-all slope is

$$S = H/L \qquad \text{(F-3)}$$

in which S is the over-all slope, in feet per mile; H is the total fall, in feet; and L is the total length of the main stream, in miles.

The over-all slope can, in some instances, be unrepresentative of the true slope because short lengths of stream with high slopes may have an effect on the average slope value out of proportion to their effect on travel time and/or peak discharge. In an effort to overcome this potential inconsistency, the following more detailed measures of slope were computed:

$$S_3 = (S + 0.3L + 0.7L)/3 \qquad \text{(F-4)}$$

$$S_4 = (S + 0.2L + 0.5L + 0.8L)/4 \qquad \text{(F-5)}$$

$$S_5 = (S + 0.2L + 0.4L + 0.6L + 0.8L)/5 \qquad \text{(F-6)}$$

$$S_{10} = (S + 0.1L + 0.2L + 0.3L + 0.4L + 0.5L + 0.6L + 0.7L + 0.8L + 0.9L)/10 \qquad \text{(F-7)}$$

Slopes S_3 through S_{10} are progressive refinements of the average slope of a stream. It was believed that the final results may show a high degree of correlation among these slopes, thus one of them may be as important (or unimportant) as the others as far as their influence on runoff is concerned.

Several other measures of stream slope proposed in the literature appear to be variations of the slopes considered herein.

WATERSHED SLOPE, R—Watershed slope has been shown previously to affect stream discharge. However, development of the watershed slope parameter presents a difficult problem, because the surface slope varies in multidimensions whereas stream slope is measured along the stream channel. Thus, most of the watershed slope methods suggested in the literature are some form of an areal averaging procedure that is tedious and time consuming. Because watershed slope does not represent information that can be easily computed by the designer or the field engineer, investigation was made of the feasibility of using another topographic variable that might be highly correlated with the watershed slope but simpler to compute. The most obvious one appears to be a stream slope just previously discussed. A randomly selected test sample of 47 basins was obtained for testing this hypothesis. These basins were all Agriculture Research Service (ARS) basins that had been used in a pilot study conducted earlier by Bock et al. (*9*). The watershed slope for these basins was computed from

$$R = \frac{C \Sigma L_{cm}}{A} \qquad \text{(F-8)}$$

in which R is the watershed slope, in feet per mile; C is the contour interval, in feet; L_{cm} is the length of individual contour line, in miles; and A is the area of the basin, in square miles.

Simple correlation coefficients, developed between R and the various measures of stream slope, were 0.604 with S, 0.537 with S_3, 0.531 with S_4, 0.520 with S_5, and 0.516 with S_{10}. These correlations are considered strong enough in view of the complexities of computing watershed slope to permit the deletion of R from the study.

EFFECTIVE SHAPE OF THE WATERSHED—This term indicates the combined effect of the shape of the basin and the configuration of the drainage network on the runoff. The shape factor was obtained by dividing the length of the main stream by the diameter of the circle having the same area as the watershed; that is,

$$\text{SHAPE} = L \Big/ \sqrt{\frac{4A}{\pi}} = \frac{L}{2}\sqrt{\frac{\pi}{A}} \qquad \text{(F-9)}$$

TRAVEL TIME INDEX—As an index of the time of travel of flow in the main stream, the following two indices were used:

$$T = L/\sqrt{S_{10}}; \qquad T_S = L/\sqrt{S} \qquad \text{(F-10)}$$

in which L is the main stream length, in miles; S_{10} is the most detailed of the stream slopes available, in feet per mile; S is the over-all slope, in feet per mile. In effect, T_S

would be the simpler travel time index to compute in actual design practice.

ELEVATION, LATITUDE and LONGITUDE of the stream gage—The stream gage elevation in feet above mean sea level (MSL) and the latitude and longitude (in degrees, minutes, and seconds) of the gage were used to characterize the basins.

Runoff Index RI

The U.S. Geological Survey (USGS) has made a number of regional and statewide studies of the magnitude and frequency of floods in the United States (*24*). These studies have included basins of all sizes, both rural and urban, and have been completed to the extent that each major region has been divided into smaller hydrologically similar subregions. For each of these subregions, the researchers developed curves based on the USGS studies for predicting magnitude and frequency of floods as a function of watershed area.

From these studies a runoff index was developed based on the value of mean annual flood for an area of 20 sq mi for each subregion. The subregions were classified according to the ranges of mean annual flood as given in Table F-1. In the study, however, only five classifications or runoff indices were used because the twelve classifications produced data gaps in several of the categories. After further study, the five categories given in Table F-2 provided a satisfactory distribution of data. Geographical distribution of these classifications is shown in Figure H-1.

TABLE F-1

CLASSIFICATION OF SUBREGIONS IN THE UNITED STATES ACCORDING TO THE MEAN ANNUAL FLOOD (MAF) FOR AN AREA OF 20 SQ MI AND ONE SQ MI

CLASSIFICATION OR RUNOFF INDEX (RI)	MEAN ANNUAL FLOOD (MAF) FOR	
	20 SQ MI (CFS)	1 SQ MI (CSM)
1	0–100	$0 \leq MAF \leq 5$
2	101–200	$5 < MAF \leq 10$
3	201–400	$10 < MAF \leq 20$
4	401–600	$20 < MAF \leq 30$
5	601–800	$30 < MAF \leq 40$
6	801–1000	$40 < MAF \leq 50$
7	1001–1400	$50 < MAF \leq 70$
8	1401–1800	$70 < MAF \leq 90$
9	1801–2500	$90 < MAF \leq 125$
10	2501–3000	$125 < MAF \leq 150$
11	3001–4000	$150 < MAF \leq 200$
12	> 4000	$200 < MAF$

TABLE F-2

MEAN ANNUAL FLOOD CATEGORIES (MAF_5)

MAF	RI	MAF (CSM)
MAF_1	1–2	$0 \leq MAF \leq 10$
MAF_2	3–4	$10 < MAF \leq 30$
MAF_3	5–6	$30 < MAF \leq 50$
MAF_4	7–8	$50 < MAF \leq 90$
MAF_5	9–12	$90 < MAF$

ASSEMBLY OF DATA SAMPLE

Development of a high quality, computer-usable data sample for the research study was an important but difficult process. It was particularly time consuming because none of the data used had been subjected to any computer-oriented processing prior to initiation of the study. Thus, considerable care had to be taken throughout the process of assembling the data sample to ensure that data errors were minimized. Quality control procedures were established for each phase of data identification, extraction, calculation, interpolation, encoding, and key punching. These control procedures are described as they pertained to the various data sets discussed in the following.

The procedures for assembling the data sample varied according to the type of data being processed and the sources from which they were derived. For the purposes of this presentation they have been divided into five classes—peak discharge, topographic, generalized hydrologic or climatic, physiographic, and short-duration precipitation. Detailed descriptions of the data extraction and processing are presented in the following sections.

Peak Discharge (Runoff)

Annual peak discharge values were obtained primarily from two types of USGS publications. For data prior to water year 1961, the Water Supply Paper Compilation Reports were utilized. Data for water years 1961 through 1966 were extracted from annual reports of Water Resources Data (Part 1: Surface Water Records) published by the individual USGS district offices.

Because the primary requirements for selection of a watershed in this study were a minimum of 12 years of runoff record and a watershed area of 25 sq mi or less, the WSP Compilation Reports were used to identify potentially acceptable locations. Many of these basins were subsequently eliminated from the sample for various reasons. Some were rejected due to hydrologic considerations, such as excessive urbanization, diversion, regulation, pondage, and noncontributing drainage area within the watershed. Others were dropped because USGS topographic mappings had not been prepared for the watershed either in total or in part. Finally, a certain percentage was rejected on the basis of recommendations of USGS district engineers in response to a questionnaire (Fig. D-1) formulated jointly with USGS headquarters personnel in Washington and distributed to the district offices under their auspices.

The extraction and preparation of annual peak discharge for computer processing consisted of the following steps for each of the 493 selected watersheds:

1. Peak discharge values and their dates of occurrence from the first year of operation of the gage through water year 1960 were extracted from WSP Compilation Reports and coded for punch-card generation.
2. Similar data for water years 1961 through 1966 were extracted from the yearly USGS district reports of surface

water records. In most instances, there was an incompatibility in USGS identification numbers due to the initiation of a "downstream order" station numbering system by USGS in the 1950's. In these cases, the latitude, longitude, and name of the gaged stream location had to be used for identification, which served to complicate the data extraction process.

3. After the extracted data had been hand-checked for correctness, they were key-punched and verified.

Within the framework of the study, surface runoff data were needed in the form of return period values rather than individual observations. A detailed evaluation of three methods of developing return period values was conducted, as described in Appendix B. On the basis of this study, the log-normal distribution was used and return period values (2, 5, 10, 20, 25, and 50 yr) were generated by computer for the 493 watersheds and placed on magnetic tape for further computer processing. Table F-3 gives a breakdown of the number of basins as a function of the number of years of annual peak discharge values available for the log-normal analysis. The mean length of record was 18.3 years.

TABLE F-3

AVAILABILITY OF ANNUAL PEAK DISCHARGE VALUES FROM SMALL RURAL WATERSHEDS

NO. OF YEARS OF DATA	NO. OF STATIONS
Less than 15	132
15 to 25	275
26 to 35	43
More than 35	43

Topographic Parameters

One of the requirements for selection of a watershed for this study was the availability of detailed USGS topographic quadrangles covering the entire extent of the watershed. Maps of 1:24,000 scale were used for those watersheds for which they were available (65 percent); for the remaining 35 percent, maps of 1:62,500 scale were used. Consideration was not given to watersheds whose mapped coverage was of a larger scale than 1:62,500 to ensure compatibility in identification and measurement of topographic features for the various watersheds.

Determination of topographic parameters from the quadrangle maps was the most tedious task of the project. As the first step, the gaging station had to be located precisely. In some cases the gaging station is identified (premarked) on the map. For most gages, however, one must use the latitude, longitude, elevation, and physical description of the gage location to identify it on the map. The second step preliminary to determining topographic features is to delineate the watershed area contributing to stream runoff past the gaging station. The defined area was measured (planimetered) and compared to the USGS published area. The delineated drainage area was accepted when the computed area fell within 5 percent of the published area.

The main stream length was measured after extending the marked stream (blue line) up to the watershed boundaries as dictated by the contours. Tributary streams were not extended to watershed boundaries, but were measured as defined by solid and/or dashed blue lines on the topographic maps. The main stream was divided into ten equidistant segments and eleven elevations were determined (gaging station, stream headwater, and nine internal points).

The second step in preparing the topographic information for computer use involved hand coding and keypunching the data onto computer punch cards. Because this step offered the potential for many data transposition errors, the procedures of coding and punching were executed twice (independent of one another). A computer program was prepared to check the two sets of data against each other, digit by digit. Errors were then identified and corrected; the cross-verification was repeated until all such differences were eliminated.

Topographic parameters were organized on punch cards in two data sets. Each punch card of the first set contained, for a single watershed, the watershed area (0.01 sq mi), main stream length (0.01 mi), length of tributaries (0.01 mi), and gage location information (latitude, longitude, USGS identification number). Each punch card of the second set contained, for a single watershed, the USGS identification number, the main stream length, and the eleven elevations (whole feet) from the main stream headwater to the outflow used to compute the stream slope terms described earlier under "Basin Characteristics."

Hydrologic-Climatic Parameters

The primary sources of hydrologic-climatic information are of two types—climatological summary statistics for individual weather stations (*60*) and climatological atlases (*22*) containing maps of the normal geographical distribution of certain parameters.

The statistics available at individual weather stations are compiled for a limited number of locations referred to by the U.S. Weather Bureau as first-order stations. The nearest of these first-order stations was identified for each of the 493 watersheds. The vast majority of the watersheds were within 50 miles of a weather station. Climatic data extracted from these summaries included: (1) mean annual temperature (° F); (2) mean wettest monthly precipitation (inches); (3) mean driest monthly precipitation (inches); (4) mean annual precipitation (inches); (5) maximum 24-hr precipitation (inches); (6) mean annual snowfall (inches); (7) maximum 24-hr snowfall (inches); (8) mean number of days per year of (a) 0.01 in. or more of precipitation, (b) 1.0 in. of snow cover, (c) thunderstorm occurrence, and (d) minimum temperature 32° F or less.

The procedure of data transposition from published tabulations to coding sheet and then to computer punch cards was similarly executed twice independently. A computer program edited the two sets of data, respectively, until all discrepancies were resolved. The variables described were organized such that the information for one watershed (specified by its USGS identification number) was included on a single data card. Also included on the

card was the World Meteorological Organization (WMO) call-letter designator of the weather station located closest to the gage location from which the information had been obtained.

Certain other climatic variables were extracted from climatic atlases (*22, 61*). They included mean relative humidity (percent), potential evapotranspiration (inches), mean July temperature (° F), and the mean annual precipitation (inches) and the 3- and 6-hr precipitation recurrence interval values (inches). These values were obtained, for a given watershed, by interpolating to its location (given by latitude and longitude expressed in degrees, minutes, and seconds). The process of coding, keypunching, and cross-verifying via a computer program was also used here. An acceptable tolerance for interpolation differences was assigned to each variable, rather than demanding an exact match. One set of data cards included the data from the climatic atlas and another the recurrence interval precipitation data.

Physiographic (Soils) Parameters

Accumulation of acceptable information regarding the characteristics of the land cover and soil structure in the watersheds of concern was a difficult task and one in which no particular success was attained. Information needed for a study such as this is not readily available and the task of collecting it for the 493 watersheds was far beyond the scope of this study. It was also considered that soils data that are not readily available would not be used by engineers in routine design of small hydrologic structures. Therefore, generally available soil classification schemes were used. Reliance also was placed on information received from USDA Soil Conservation Service (SCS State Conservationists) through a questionnaires (Fig. D-2) formulated jointly with SCS personnel familiar with the proposed study objectives and distributed under their auspices. Detailed information was obtained on the percentage of the various hydrologic soil groups within a watershed. These soil groups, well-known to hydrologists and soil conservationists, have been defined (*52*) on the basis of the effect of the soil structure on surface runoff (or, conversely, the soil's capacity to absorb moisture). Also included were detailed percentages of land cover.

Each item obtained by questionnaire was placed on punch cards and validated in a similar fashion to the other punch card data to minimize data transposition errors.

Two composite parameters were developed for each watershed documented by SCS conservationists. In each, the individual percentages defined for each category of soil type (or land cover) are translated into a single number depicting the total watershed. The composite soil parameter (CS) is defined by

$$CS = \frac{4(\%\ \text{Soil Type D}) + 3(\%C) + 2(\%B) + (\%A)}{100} \tag{F-11}$$

and the composite land cover parameter (CC) by

$$CC = \frac{1}{100}\left[\begin{array}{l} 4(\%\ \text{Impervious Surface}) + 3(\%\ \text{Cultivated}) \\ \quad + 2(\%\ \text{Pasture}) + (\%\ \text{Forest}) \end{array}\right] \tag{F-12}$$

In each case, then, the composite parameter has a linear range of 1.0 to 4.0, corresponding at the lower value to conditions retarding runoff volume (i.e., forested land and/or soil of high infiltration capacity), and at the higher value to conditions favoring runoff.

The generalized soil information taken from published sources is summarized as follows:

1. Soil classification from Soil Order Map of the Soil Conservation Service. There are 10 classifications based on the structure of the soil (*53*).
2. Geological zone in which each watershed is located. The classification of these zones is the same as recommended by Potter (*42*). The four geological zone classifications are:
   - Zone I—Glacial drift and loess.
   - Zone II—Sandstone and shale.
   - Zone III—Limestone.
   - Zone IV—Schist.
3. Different degrees of soil erosion taken from SCS Soil Erosion Map (*31*). The six classes of soil erosion are:
   - (a) Severe sheet and gully erosion.
   - (b) Moderate to severe erosion of mesas and mountains.
   - (c) Moderate to severe wind erosion with some gullying.
   - (d) Moderate sheet and gully erosion with some wind action.
   - (e) Moderate sheet and gully erosion.
   - (f) Erosion rather unimportant.

Short-Duration Precipitation Data

All of the basic data considered in previous sections could be obtained by an engineer designing hydraulic structures in a manner similar to that used by the researchers. In fact, one of the primary considerations in formulating this research was that the methods developed and the information used in their solution had to be easily accessible to the design engineer. There was, however, one fundamental source of arithmetic data, desirable for the study, that required a sizable effort in data processing and analysis to transpose to a form usable for the study and by the engineer in design practice. These unprocessed data, which the researchers undertook to process, were tabulation records of short-duration precipitation intensity, which, from a hydrologic-physical point of view, appeared to offer potential value in a study of peak discharge from small watersheds. Previously published mappings of precipitation recurrence (*61*) are limited to durations of 30 min or longer.

The study data consisted of hand tabulations of maximum annual precipitations at 167 U.S. Weather Bureau first-order stations. Included in the tabulations were the date of occurrence and the maximum amount of precipitation in a 5-, 10-, 15-, 30-, 60-, and 120-min period on a year-by-year basis. The vast majority of these stations have been operating for more than 50 years; thus, these data represent approximately 8,000 station-years of record equivalent to about 50,000 values. The data were extracted from the tabulations, coded, placed on punch cards,

and verified. The details of the subsequent analysis of the station data using the Gumbel method to generate recurrence interval values of short-duration precipitation and the development of objective mappings of the recurrence-interval values utilizing a computer analysis technique called CRAM (Conditional Relaxation Analysis Method) are presented in Appendix C. The output of this process was a magnetic tape containing values of short-duration precipitation for the 493 basins.

Merge of Data into Computer-Usable Form

Organization of the punch card and magnetic tape data into a form compatible with the computerized statistical techniques was performed by a specially prepared computer program package to merge the various sets of data onto a single magnetic tape such that all data for one watershed were placed in a single magnetic tape record. Initially, the several sets of data on punch cards were merged onto a magnetic tape by a computer program that also computed several additional parameters, described earlier, such as:

1. Stream slopes (S, S_3, S_4, S_5, S_{10}).
2. Drainage density (DD).
3. Watershed shape (SHAPE).
4. Travel time index (T).
5. Composite soil parameter (CS).
6. Composite land cover parameter (CC).

The program examined the various sets of data, using the USGS identification number to match all data common to a particular watershed, made the necessary computations, and generated a magnetic tape on which the watersheds were organized in a ranked order according to their identification number to facilitate subsequent processing. By use of another program, the peak discharge (runoff), short-duration precipitation, and predictor data tapes were merged in a similar manner, and logarithms (base 10) were computed of many quantitative variables.

Special procedures had to be formulated to handle the logarithm calculation of those variables that had one or more cases in which their value was zero. In each case (there were six such variables), a meaningfully small value was assigned to the variable in order to compute its logarithm. The values used were: (1) gage elevation, 1.0 ft; (2) mean driest monthly precipitation, 0.01 inch; (3) mean annual snowfall, 0.1 inch; (4) mean number of thunderstorm days per year, 1; (5) mean number of days per year with minimum temperature 32° F or lower, 1; and (6) length of tributaries, 0.1 mile.

Additional data manipulations performed by the computer programs included classification of each watershed into geographical zones (as described in Appendix H), and generation of dummy (binary or dichotomous) variables from several of the qualitative variables.

A dummy variable is one that can take on only two values: it is either equal to zero or equal to one. By way of illustration, consider the four types of geologic zones defined by Potter (*42*). A set of four dummy variables can be generated for each case of the sample by determining to which zone a specific basin belongs and assigning a one (1) to that dummy variable and a zero (0) to the three remaining dummy variables. In this manner, qualitative variables are put in a suitable quantified form for the statistical experiments. Four of the qualitative variables were processed in this way. Potter's geologic zones (GZ) were categorized into four dummy variables and an extra dummy for those cases that were in regions undefined geologically. Soil erosion (SE) was categorized in four dummy variables; class 1, classes 2-4, class 5, and class 6. The mean annual flood categories (MAF) were put into five dummy variable categories. Finally, the six geographical zones (REG) were made into six dummy variables.

In developing new methods for estimating some parameter (peak runoff in this case), it is essential to set aside a certain portion of the processed data sample for the purpose of testing the reliability of the new relationships on data not used in developing the relationships. Therefore, approximately 20 percent of the 493 basins were selected on a random sampling basis. The developmental sample from which the relationship was developed consisted of 395 basins; the other 98 basins were withheld for verification purposes. Table F-4 gives a breakdown, by area, of the watersheds in the dependent and independent samples. The distribution through the full range of areas is fairly uniform, although there are proportionally more small watersheds (less than 5 sq mi) than there are through the other ranges of watershed area.

TABLE F-4

DISTRIBUTION OF WATERSHEDS IN DEPENDENT AND INDEPENDENT SAMPLES ACCORDING TO AREA

AREA (SQ MI)	NUMBER OF STATIONS		
	DEPENDENT	INDEPENDENT	TOTAL
.01–1	33	9	42
1.01–2	37	7	44
2.01–3	31	11	42
3.01–4	13	8	21
4.01–5	16	5	21
5.01–6	19	5	24
6.01–7	16	2	18
7.01–8	21	3	24
8.01–9	23	2	25
9.01–10	14	5	19
10.01–11	13	2	15
11.01–12	13	4	17
12.01–13	7	2	9
13.01–14	9	1	10
14.01–15	9	8	17
15.01–16	17	3	20
16.01–17	11	3	14
17.01–18	7	2	9
18.01–19	12	3	15
19.01–20	18	0	18
20.01–21	10	2	12
21.01–22	16	3	19
22.01–23	14	3	17
23.01–24	6	4	10
24.01–25	10	1	11
Total	395	98	493

APPENDIX G

STATISTICAL TECHNIQUES USED IN THIS STUDY

Two basic statistical techniques were used to develop equations expressing runoff as a function of other variables: step-wise linear regression and nonlinear regression. To verify the accuracy of the estimates of runoff made by these two techniques, three additional statistical techniques were employed: root-mean-square-error, sign test, and frequency distribution of errors of estimation. It is the purpose in this appendix to describe the five statistical techniques.

Additional statistical techniques were used to compute runoff frequency values and to compute return period amounts for precipitation of various durations. These techniques are described in Appendix B and Appendix C.

STEPWISE REGRESSION

A stipulated variable (e.g., peak runoff) called the *predictand* is the object of estimation. The variables used to make the estimation of the predictand are termed predictors. The number of plausible predictors that could be used to estimate runoff is rather large. On the other hand, the engineer applying the procedure in practice has only a limited amount of time. Furthermore, it is well known from statistical theory that the larger the number of predictors, the greater is the "shrinkage" in accuracy of estimation when the procedure is applied to new data. This situation imposes the practical necessity of selecting a manageable number of predictors. The step-wise regression technique makes a preferential selection of effective predictors from a large set of possible choices. Experiments comparing performance on independent data of estimation functions using large numbers of predictors with those using selectively chosen subsets of such variables have shown, as a rule, that whatever estimation accuracy resides in the large set is almost wholly contained in the much smaller subset. The objective selection of such a small subset is termed a step-wise procedure. After the procedure has been applied, the redundant or noncontrolling predictors are eliminated from subsequent analyses, and a multiple regression equation is developed using only the selected predictors.

In multiple regression, the predictand, Q, is expressed as a linear function of a number (P) of predictor variables:

$$\widehat{Q} = A_0 + A_1X_1 + A_2X_2 + \ldots + A_PX_P \qquad \text{(G-1)}$$

in which the coefficients $A_p (p = 0, 1, \ldots, P)$ are determined by least squares and the X's are the predictors. To select the first predictor, the simple linear correlation is computed between the predictand and each predictor. The predictor having the highest simple correlation coefficient is selected first. Next, partial correlations between each of the remaining predictors and the predictand (holding the first selected predictor constant) are examined and the predictor associated with the best partial coefficient is then selected as a second predictor. Additional predictors are selected in a similar manner. Selection is halted on the basis of an F-test criterion.

NONLINEAR REGRESSION

The nonlinear regression technique develops an equation expressing the predictand as a function of nonlinear functions of each of the predictors selected by the step-wise regression technique. The equation is of the form

$$\widehat{Q} = B_0 + B_1f_1(X_1) + B_2f_2(X_2) + \ldots + B_Pf_P(X_P) \qquad \text{(G-2)}$$

in which the B's are constants determined by least squares and the f's are nonlinear functions of the predictor variables. The equation is developed in four steps, as follows:

1. The step-wise regression procedure of the previous section is applied to the raw predictor variables to develop a regression equation such as Eq. G-1.
2. The "residual" procedure, described later, obtains "net residuals" for each predictor selected by the step-wise regression technique.
3. The "rational fit" procedure, also described later, uses the net residuals to compute a nonlinear function of each predictor.
4. The nonlinear functions serve as input to the step-wise regression technique, which selects a subset of them and computes Eq. G-2.

The nonlinear prediction technique was developed by Joseph G. Bryan of The Travelers Research Corporation. This exposition is based on Bryan's unpublished reports. Many of the details and fine points, and most of the proofs, are omitted. It is strongly urged that persons contemplating applying the procedure consult Bryan's original work (*10*).

Residual Procedure

The step-wise regression technique has selected a set of raw predictor variables and has developed a regression equation (such as Eq. G-1) expressing runoff as a linear function of selected variables. The objective of the residual procedure is to compute smoothed values of "net residuals" for each of the selected variables. Such smoothed values serve as input to the "rational fraction fit" procedure.

The net residual for the kth predictor variable is defined by

$$E_{ki} = Q_{ki}' - Q_i \qquad \text{(G-3)}$$

in which Q is the observed runoff in the ith basin in the sample of N basins used to develop the step-wise regression equation. Q_{ki}' is computed by Eq. G-1 with the term involving X_k omitted. Eq. G-3 is applied to each of the

N basins to obtain a series of E_{ki} values. This is repeated for each predictor variable.

To save computer time in the rational fraction fit procedure, the E_{ki} values are smoothed. Experience has indicated that smoothing has little, if any, effect on the ultimate results. The smoothing proceeds as follows:

1. For each of the P predictor variables, a set of class limits, $L_1 < L_2 < L_3 \ldots$ are specified by the investigator. Up to 29 limits may be specified.
2. The predictor data are grouped by the class limits; i.e., all values $<L_1$ are placed in group 1, values between L_1 and L_2 are in group 2, etc. The predictor values within a group are averaged to obtain group means.
3. The net residual values corresponding to the predictor values within a group are also averaged to obtain group means of net residuals.

For each predictor variable there are now two small series of means, ≤ 30 values in each series. The two series serve as input to the rational fraction fit procedure.

Rational Fraction Fit Procedure

The objective of the rational fraction fit procedure is to develop a nonlinear function of each of the predictor variables. The function is of the form

$$f_k(X_k) = C_{k0} + C_{k1}\frac{X_k}{(1+RX_k)^j} + C_{k2}\frac{X_k^2}{(1+RX_k)^j} \qquad \text{(G-4)}$$

in which k denotes the kth predictor variable; j is the degree of the equation ($=1$, 2, or 3); R is a constant ≥ 0 determined by a series of successive approximations; and the C's are constants determined by least squares. If $R=0$, Eq. G-4 is a polynomial in X. If $j=1$, $A_{k0}=0$, and $A_{k1}=1$, then $f_k(X_k)=X$. If $j=1$ and $R \neq 0$, even the first-degree form of Eq. G-4 is a nonlinear function of X.

The first problem faced in developing Eq. G-3 is to decide what degree to use. This is solved by first developing three equations, the first with $j=1$, a second with $j=2$, and a third with $j=3$. A statistical test is then applied to choose the degree that fits the data best.

Computation of each of these three equations is somewhat complicated and is described fully in Bryan's papers. The major computational burden is to fix the constant R. This is accomplished by a technique of successive trials. A value for R is assumed, then values of C are computed by least squares. A statistical test is applied to see how close the resulting function fits the data. A new value of R is assumed and the process is repeated to obtain another measure of closeness of fit. A third and successive values of R are assumed, and a third and successive measures of closeness of fit are computed. By examining the measures of closeness of fit it is possible to close in on the value of R that gives a function that fits the data closely. The final R and its associated C-values give an equation such as Eq. G-2. As previously stated, there are three such equations and a statistical test is applied to choose the best of the three.

The final output of the rational fraction fit procedure is one nonlinear function of each predictor variable. The functions are applied to the data to obtain a series of N values for each predictor variable. These values serve as input to the step-wise regression technique and Eq. G-2 is computed.

LOGARITHMIC FORM OF PREDICTAND

The two techniques were applied, in some instances, to develop equations for estimating the logarithm of Q rather than Q itself. This presents a statistical problem. The natural tendency is to take as the estimate of Q the antilogarithm of the estimate of log Q, or

$$Q' = \text{antilogarithm } (\widehat{\log Q}) \qquad \text{(G-5)}$$

However, the mean of the Q' values will be lower than the mean of the observed Q values. This is the result of the well-known mathematical fact that the logarithm of the mean of a set of values is greater than the mean of the logarithms of the individual values in the set.

It was deemed desirable to have the mean of the estimates made equal to the mean of the observed peak runoffs to ensure an unbiased estimater. This was accomplished by multiplying each $\bar{Q}'$ by $(\bar{Q}/\bar{Q}')$, or

$$\tilde{Q} = Q'(\bar{Q}/\bar{Q}') \qquad \text{(G-6)}$$

in which $\bar{Q}$ is the mean of the observed peak runoff values and Q' is from Eq. G-5. (For convenience, the term $(\bar{Q}/\bar{Q}')$ is called B_0 in some of the equations listed in Appendix A).

VERIFICATION TECHNIQUES

In the experimentation described in subsequent sections, the entire set of basins was divided into two sets—a developmental set to be used to develop equations for estimating Q and a test set to be used to verify the estimation procedures. Several verification techniques are used. It is the purpose in this section to describe the verification techniques.

Root-Mean-Square-Error

The root-mean-square-error of a series of N estimates is defined as

$$\text{RMSE} = \left[(1/N)\sum_{i=1}^{N}(\widehat{Q}_i - Q_i)^2\right]^{1/2} \qquad \text{(G-7)}$$

in which Q_i is the observed peak runoff for basin i and $\widehat{Q}_i$ is an estimate of Q_i. The RMSE is an excellent verification statistic whenever $(\widehat{Q}_i - Q_i)$ is normally distributed. However, experience indicates that this is not true in general and that, in fact, RMSE can be dominated by one or two large values. Nevertheless, it is a standard verification measure and does convey some information, provided extreme care is taken in its interpretation.

Sign Test

The sign test is used to compare the accuracy of two estimation methods. Let Q_i, $i=1, 2, \ldots, N$ denote peak runoff values for N basins and let Q_{1i} denote estimates made by one procedure and Q_{2i} denote estimates made by another procedure. Let

$$E_{1i} = |Q_{1i} - Q_i| \qquad (G\text{-}8)$$
$$E_{2i} = |Q_{2i} - Q_i|$$

denote, respectively, the error of forecast method one and of forecast method two in estimating Q_i. Now, the smaller of the two E's tells which forecast method is best. The sign test counts how many times each forecast method was best. Let N_1 denote the number of times forecast system one was best and $N_2(= N - N_1)$ the number of times system two was best. Then, by applying the binomial test to N_1 and N_2, one can test the hypothesis that there is no difference in forecast systems.

The sign test does not assume any distribution for the errors, only that they be continuous variables. The null hypothesis tested is that

$$\text{Probability}(E_{1i} > E_{2i}) = \text{Probability}(E_{2i} > E_{1i}) = ½ \qquad (G\text{-}9)$$

Under this null hypothesis it is expected that $N_1 = N_2$. However, just by chance there could be deviations from equality in any specific case even when Eq. G-9 is true. Tables of the binomial distribution can be entered with N_1 and N to determine the probability that the observed value of N_1 could have arisen by chance if Eq. G-9 is true. If this probability is small, there is strong evidence that one estimation procedure is better than the other. If $N_1 > N_2$, the first procedure is better, and vice versa.

TABLE G-1

PERCENT ERROR (PE) DISTRIBUTION CATEGORIES

CATEGORY	RANGE (PERCENT)
1	$-10 < \text{PE} \leq 10$
2	$10 < \text{PE} \leq 25$
	$-25 < \text{PE} \leq -10$
3	$25 < \text{PE} \leq 50$
	$-50 < \text{PE} \leq -25$
4	$50 < \text{PE} \leq 100$
	$-100 < \text{PE} \leq -50$
5	$100 < \text{PE} \leq 150$
	$-150 < \text{PE} \leq -100$
6	$150 < \text{PE} \leq 200$
	$-200 < \text{PE} \leq -150$
7	$200 < \text{PE}$
	$\text{PE} \leq -200$

Frequency Distribution of Errors

This is a presentation of data for visual inspection rather than a formal verification procedure. The "percent error" is defined as

$$\text{PE}_i = E_i / Q_i \qquad (G\text{-}10)$$

in which E_i is the error of a forecast (Eq. G-8), and Q_i is the observed peak runoff. The PE_i values are ranked from lowest to highest. The frequency distribution of errors is a count of the number of errors within specified ranges. In this study, seven ranges were used (Table G-1). The error distribution for a number of estimation procedures provides a convenient format for subjective comparison of the accuracy attained by the alternative procedures.

APPENDIX H

HYDROLOGIC/STATISTICAL EXPERIMENTS

FRAMEWORK OF EXPERIMENTATION

Table H-1 summarizes the 24 experiments that were conducted in the order discussed here. Details regarding the particular predictors considered, those selected, the equations developed, and the significance of the results for each experiment are presented in subsequent sections.

In each of the experiments summarized in Table H-1 the basic statistical technique utilized was screening regression. However, the functional form of the relationship developed varies from experiment to experiment due to the treatment of the possible predictors and the predictand. In some cases the arithmetic form of the variables is retained (e.g., area in sq mi); in others, logarithmic form (all logarithms are base 10) is considered (e.g., log (A) in sq mi); in others generalized nonlinear form is considered (e.g., see Eq. G-4).

The process of developing statistical relationships using screening regression is conducted in two steps. Initially, individual correlation coefficients are developed between each of the predictands (peak runoff) and all of the possible predictors, and also among all of the predictors themselves. This information is retained in a covariance matrix used in the second step wherein a subset of predictors is selected as described in Appendix G. Table H-2 gives, for each experiment, the predictors that were made available to the screening selection procedure and the predictand (peak runoff) variables for which relationships were developed. Reference should be made to Table H-1 regarding

TABLE H-1

SYNOPSIS OF 24 HYDROLOGIC/STATISTICAL EXPERIMENTS TO DEVELOP PEAK RUNOFF ESTIMATION RELATIONSHIPS

EXPER. NO.	PREDICTAND FORM	PREDICTOR SETS	PREDICTOR FORM	NATIONAL OR STRATIFIED [a]
		A—EVALUATION OF PREDICTOR SETS (PHASE I)		
A-1	Logarithmic	Hydrologic-climatic	Arithmetic	N
A-2	Logarithmic	Topographic	Arithmetic	N
A-3	Logarithmic	Hydrologic-clim.; topo.	Arithmetic	N
A-4	Logarithmic	Hydrologic-clim.; topo.; soil	Arithmetic	N
A-5	Logarithmic	Hydrologic-clim.; topo.; soil	Arithmetic	N
		B—FUNCTIONAL FORM OF THE VARIABLES (PHASE II)		
B-1	Arithmetic	Hydrologic-clim.; topo.	Arithmetic	N
B-2	Logarithmic	Hydrologic-clim.; topo.	Logarithmic	N
B-3	Logarithmic	Hydrologic-clim.; topo.	Arith.; log.	N
B-4	Logarithmic	Hydrologic-clim.; topo.	Nonlinear	N
B-5	Logarithmic	Hydrologic-clim.; topo.	Nonlinear	N
		C—STRATIFICATION INTO SUBSETS (PHASE III)		
C-1	Logarithmic	Hydrologic-clim.; topo.	Logarithmic	S(MAF_5)
C-2	Logarithmic	Hydrologic-clim.; topo.	Logarithmic	S(P_{10-60})
C-3	Logarithmic	Hydrologic-clim.; topo.	Logarithmic	S(MAP)
C-4	Logarithmic	Hydrologic-clim.; topo.	Arith.; log.	S(MAT)
C-5	Logarithmic	Hydrologic-clim.; topo.	Logarithmic	S(A)
C-6	Logarithmic	Hydrologic-clim.; topo.	Logarithmic	S($\bar{E}$)
C-7	Logarithmic	Hydrologic-clim.; topo.	Arith.; log.	S(SE)
C-8	Logarithmic	Hydrologic-clim.; topo.	Arith.; log.	S(GZ)
C-9	Logarithmic	Hydrologic-clim.; topo.	Arithmetic	S(REG)
		D—MODIFIED NATIONAL EQUATIONS INCLUDING STRATIFICATION FACTORS (PHASE IV)		
D-1	Logarithmic	Hydrologic-clim.; topo.	Arith.; binary	N
D-2	Logarithmic	Hydrologic-clim.; topo.	Arith.; log.; bin.	N
D-3	Logarithmic	Hydrologic-clim.; topo.	Log.; binary	N
		E—SIMPLIFIED NATIONAL EQUATIONS (PHASE V)		
E-1	Logarithmic	Hydrologic-clim.; topo.	Log.; binary	N
E-2	Logarithmic	Hydrologic-clim.; topo.	Log.; binary	N

[a] N = National; S( ) = stratified (basis of stratification).

the form of the predictand and predictors in a given experiment. For example, in experiment A-1 predictor No. 15 is considered in its arithmetic form, whereas in experiment B-2 it is considered in logarithmic form. In most of the experiments that were conducted four predictands were considered—Q_5, Q_{10}, Q_{25}, and Q_{50}; or log Q_5, log Q_{10}, log Q_{25}, and log Q_{50}, depending on the predictand form for the particular experiment.

INDIVIDUAL CORRELATION COEFFICIENTS

As a necessary preliminary step to the experimentation, simple linear correlation coefficients were computed between the predictand variables and the predictor variables. Such coefficients are of interest in themselves because they measure how well a single characteristic of a basin can estimate runoff in that basin.

The correlations were based on 395 watersheds. Both arithmetic values and logarithms of the four predictands, Q_5, Q_{10}, Q_{25}, and Q_{50}, were used, as were arithmetic values and logarithms of some 101 predictor variables. It is important to note that correlations involving arithmetic values of the predictands cannot be compared with those involving logarithms of predictands because a different variable is being estimated. Correlations involving arithmetic and logarithmic values of *predictor* variables are comparable wherever the same predictand is being considered.

Table H-3 gives the individual correlation coefficients, the highlights of which are discussed in the following.

First, the correlations are low. The highest correlation involving an arithmetic predictand is only 0.580 (i.e., between tributary length, TRIB, and Q_{10}). Many of the coefficients are close to zero, indicating little or no association between predictor and predictand.

Second, the magnitude of the coefficients decreases with increasing length of return period. This is most likely due to the short periods of record for many watersheds, which causes larger errors in estimating longer return period runoffs. Such larger errors tend to decrease any correlation that may exist between a predictor and runoff.

TABLE H-2

LIST OF 101 PREDICTORS AVAILABLE FOR SELECTION BY STEPWISE REGRESSION

NO.	NAME	SYMBOL	A-1	A-2	A-3	A-4	A-5	B-1	B-2	B-3	B-4	B-5	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	D-1	D-2	D-3	E-1	E-2
1	Peak runoff, 2-yr return	Q_2																								
2	Peak runoff, 5-yr return	Q_5				Q	Q	Q	Q	Q			Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
3	Peak runoff, 10-yr return	Q_{10}	Q	Q	Q	Q	Q	Q	Q	Q			Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
4	Peak runoff, 20-yr return	Q_{20}																								
5	Peak runoff, 25-yr return	Q_{25}	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
6	Peak runoff, 50-yr return	Q_{50}				Q	Q	Q	Q	Q			Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
7	Area	A		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
8	Main stream length	L		X	X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		X
9	Gage elevation	E		X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X		
10	Stream slope-1	S_3																								
11	Stream slope-2	S_4																								
12	Stream slope-3	S_5																								
13	Stream slope-4	S_{10}		X	X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
14	Over-all stream slope	S		X	X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
15	Mean annual temperature	MAT	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		X
16	Mean wettest month	P_{wet}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
17	Mean driest month	P_{dry}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
18	Mean annual precip.-station	MAP-S	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
19	Max. 24 hr precipitation	M24P	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
20	Mean annual snowfall	MAS	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
21	Max. 24-hr snowfall	M24S	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
22	0.01-in. precipitation days	P days	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		X
23	1-in. snow cover days	S days	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
24	Thunderstorm days	T days	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
25	32°F or lower days	32F days	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
26	Mean annual precip., map	MAP	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		X
27	Mean relative humidity	RH	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
28	Mean July temperature	July T	X		X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X		X
29	Potential evapotranspiration	Pot E T	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
30	Length of tributaries	TRIB		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
31	Drainage density	DD		X	X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
32	Watershed shape factor	SHAPE		X	X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
33	Travel time index	T		X	X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
34	3-hr precip., 5-yr return	P_{5-180}																								
35	3-hr precip., 10-yr return	P_{10-180}																								
36	3-hr precip., 25-yr return	P_{25-180}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
37	3-hr precip., 50-yr return	P_{50-180}																								
38	3-hr precip., 100-yr return	$P_{100-180}$																								
39	6-hr precip., 5-yr return	P_{5-360}																								
40	6-hr precip., 10-yr return	P_{10-360}									X	X													X	X
41	6-hr precip., 25-yr return	P_{25-360}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
42	6-hr precip., 50-yr return	P_{50-360}																								
43	6-hr precip., 100-yr return	$P_{100-360}$																								
44	5-min precip., 5-yr return	P_{5-05}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
45	5-min precip., 10-yr return	P_{10-05}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
46	5-min precip., 25-yr return	P_{25-05}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
47	5-min precip., 50-yr return	P_{50-05}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
48	5-min precip., 100-yr return	P_{100-05}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
49	10-min precip., 5-yr return	P_{5-10}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
50	10-min precip., 10-yr return	P_{10-10}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
51	10-min precip., 25-yr return	P_{25-10}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
52	10-min precip., 50-yr return	P_{50-10}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
53	10-min precip., 100-yr return	P_{100-10}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		

NO.	NAME	SYMBOL	A-1	A-2	A-3	A-4	A-5	B-1	B-2	B-3	B-4	B-5	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	D-1	D-2	D-3	E-1	E-2
54	15-min precip., 5-yr return	P_{5-15}																								
55	15-min precip., 10-yr return	P_{10-15}																								
56	15-min precip., 25-yr return	P_{25-15}																								
57	15-min precip., 50-yr return	P_{50-15}																								
58	15-min precip., 100-yr return	P_{100-15}																								
59	30-min precip., 5-yr return	P_{5-30}																								
60	30-min precip., 10-yr return	P_{10-30}																								
61	30-min precip., 25-yr return	P_{25-30}																								
62	30-min precip., 50-yr return	P_{50-30}																								
63	30-min precip., 100-yr return	P_{100-30}																								
64	60-min precip., 5-yr return	P_{5-60}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
65	60-min precip., 10-yr return	P_{10-60}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
66	60-min precip., 25-yr return	P_{25-60}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
67	60-min precip., 50-yr return	P_{50-60}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
68	60-min precip., 100-yr return	P_{100-60}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
69	120-min precip., 5-yr return	P_{5-120}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
70	120-min precip., 10-yr return	P_{10-120}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
71	120-min precip., 25-yr return	P_{25-120}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
72	120-min precip., 50-yr return	P_{50-120}	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
73	120-min precip., 100-yr return	$P_{100-120}$	X		X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X		
74	Percentage of soil type A	SA				X																				
75	Percentage of soil type B	SB				X																				
76	Percentage of soil type C	SC				X																				
77	Percentage of soil type D	SD				X																				
78	Composite soil parameter	CS				X	X																			
79	Composite land cover parameter	CC				X	X																			
80	Mean annual flood, altern. 1	MAF_{12}																								
81	Mean annual flood, altern. 2	MAF_{6}																								
82	Mean annual flood, altern. 3	MAF_{5}									X	X										X	X	X		X
83	Soil order class	SO																								
84	Soil suborder class	SSO																								
85	Erosion class	SE																								
86	Geologic zone	GZ																								
87	Geographic region	REG																								
88	MAF_5, dummy variable 1	MAF_5-1																				X	X	X		
89	MAF_5, dummy variable 2	MAF_5-2																				X	X	X		
90	MAF_5, dummy variable 3	MAF_5-3																				X	X	X		
91	MAF_5, dummy variable 4	MAF_5-4																				X	X	X		
92	MAF_5, dummy variable 5	MAF_5-5																				X	X	X		
93	SE, dummy variable 1	SE1																				X	X	X		
94	SE, dummy variable 2	SE2																				X	X	X		
95	SE, dummy variable 3	SE3																				X	X	X		
96	SE, dummy variable 4	SE4																				X	X	X		
97	GZ, dummy variable 1	GZ1																				X	X	X		
98	GZ, dummy variable 2	GZ2																				X	X	X		
99	GZ, dummy variable 3	GZ3																				X	X	X		
100	GZ, dummy variable 4	GZ4																				X	X	X		
101	GZ, dummy variable 5	GZ5																				X	X	X		
102	REG, dummy variable 1	REG1									X	X										X	X	X		X
103	REG, dummy variable 2	REG2									X	X										X	X	X		X
104	REG, dummy variable 3	REG3									X	X										X	X	X		X
105	REG, dummy variable 4	REG4									X	X										X	X	X		X
106	REG, dummy variable 5	REG5									X	X										X	X	X		X
107	REG, dummy variable 6	REG6									X	X										X	X	X		X

TABLE H-3

SIMPLE CORRELATION COEFFICIENTS BETWEEN PREDICTORS AND FOUR PREDICTANDS IN ARITHMETIC AND LOGARITHMIC FORM

PREDICTOR SYMBOL	Q_5	LOG Q_5	Q_{10}	LOG Q_{10}	Q_{25}	LOG Q_{25}	Q_{50}	LOG Q_{50}
A	0.513	0.504	0.492	0.501	0.433	0.488	0.377	0.476
log A	0.458	0.560	0.440	0.557	0.388	0.544	0.338	0.530
L	0.448	0.459	0.450	0.465	0.422	0.463	0.387	0.458
log L	0.425	0.510	0.419	0.513	0.382	0.507	0.343	0.498
E	—0.183	—0.259	—0.169	—0.250	—0.141	—0.238	—0.117	—0.229
log E	—0.019	—0.057	0.002	—0.032	0.028	—0.009	0.042	0.005
S_{10}	—0.273	—0.306	—0.265	—0.311	—0.237	—0.309	—0.207	—0.304
log S_{10}	—0.288	—0.279	—0.283	—0.291	—0.254	—0.295	—0.224	—0.293
S	—0.253	—0.257	—0.245	—0.262	—0.218	—0.261	—0.189	—0.257
log S	—0.258	—0.232	—0.254	—0.245	—0.229	—0.250	—0.201	—0.250
MAT	0.190	0.249	0.190	0.259	0.172	0.268	0.150	0.274
log MAT	0.187	0.254	0.187	0.262	0.167	0.268	0.145	0.272
P_{wet}	0.058	0.047	0.007	—0.003	—0.046	—0.052	—0.073	—0.079
log P_{wet}	0.108	0.118	0.060	0.071	0.006	0.025	—0.024	—0.003
P_{dry}	0.203	0.299	0.142	0.259	0.059	0.216	0.005	0.190
log P_{dry}	0.154	0.142	0.091	0.091	0.013	0.038	—0.034	0.006
MAP-S	0.169	0.193	0.094	0.129	0.004	0.065	—0.047	0.027
log MAP-S	0.180	0.218	0.112	0.158	0.028	0.096	—0.022	0.060
M24P	0.126	0.234	0.146	0.253	0.152	0.268	0.146	0.276
log M24P	0.153	0.253	0.172	0.272	0.178	0.287	0.172	0.293
MAS	—0.028	0.022	—0.031	0.013	—0.034	0.003	—0.034	—0.003
log MAS	—0.048	—0.081	—0.067	—0.109	—0.080	—0.136	—0.081	—0.153
M24S	—0.033	—0.017	—0.041	—0.045	—0.043	—0.071	—0.040	—0.086
P days	—0.027	—0.041	—0.096	—0.113	—0.163	—0.183	—0.192	—0.222
log P days	0.015	—0.004	—0.054	—0.073	—0.126	—0.141	—0.159	—0.180
S days	—0.053	—0.017	—0.057	—0.026	—0.057	—0.036	—0.054	—0.043
T days	0.283	0.205	0.281	0.219	0.253	0.227	0.222	0.229
log T days	0.224	0.171	0.213	0.174	0.181	0.174	0.151	0.171
32F days	—0.015	—0.056	—0.007	—0.044	0.004	—0.036	0.010	—0.033
log 32F days	0.054	—0.009	0.025	—0.038	—0.009	—0.069	—0.027	—0.088
MAP	0.014	0.012	—0.055	—0.056	—0.124	—0.121	0.156	—0.157
log MAP	0.063	0.102	—0.009	0.030	—0.087	—0.039	—0.127	—0.078
July T	0.321	0.328	0.340	0.369	0.330	0.402	0.304	0.418
log July T	0.311	0.328	0.330	0.369	0.320	0.403	0.296	0.419
Pot ET	0.247	0.273	0.240	0.284	0.209	0.292	0.178	0.296
log Pot ET	0.250	0.289	0.243	0.299	0.214	0.306	0.183	0.310
TRIB	0.573	0.522	0.580	0.530	0.546	0.529	0.499	0.524
log TRIB	0.490	0.618	0.487	0.621	0.449	0.615	0.405	0.606
DD	0.181	0.158	0.221	0.178	0.250	0.197	0.255	0.207
log DD	0.213	0.222	0.250	0.242	0.274	0.259	0.274	0.268
SHAPE	0.190	0.197	0.217	0.214	0.234	0.226	0.234	0.231
log SHAPE	0.200	0.227	0.219	0.240	0.228	0.249	0.222	0.251
T	0.455	0.407	0.452	0.415	0.416	0.410	0.376	0.404
log T	0.478	0.531	0.470	0.541	0.427	0.539	0.380	0.532
P_{25-180}	0.293	0.359	0.296	0.376	0.272	0.387	0.242	0.389
P_{25-360}	0.240	0.334	0.240	0.344	0.220	0.349	0.195	0.349
P_{10-05}	0.303	0.322	0.296	0.337	0.262	0.344	0.226	0.344
log P_{10-05}	0.281	0.308	0.280	0.328	0.252	0.340	0.220	0.344
P_{10-10}	0.282	0.298	0.281	0.317	0.255	0.329	0.223	0.333
log P_{10-10}	0.267	0.290	0.269	0.312	0.246	0.328	0.218	0.334
P_{10-15}	0.270	0.285	0.273	0.306	0.250	0.321	0.222	0.326
log P_{10-15}	0.259	0.279	0.263	0.304	0.243	0.321	0.216	0.328
P_{10-30}	0.257	0.270	0.263	0.295	0.245	0.319	0.219	0.318
log P_{10-30}	0.250	0.269	0.256	0.295	0.239	0.314	0.214	0.321
P_{10-60}	0.249	0.261	0.257	0.287	0.241	0.306	0.217	0.313
log P_{10-60}	0.244	0.262	0.251	0.289	0.236	0.309	0.212	0.317
P_{10-120}	0.242	0.253	0.251	0.280	0.238	0.300	0.216	0.308
P_{25-05}	0.307	0.325	0.303	0.340	0.270	0.349	0.234	0.350
log P_{25-05}	0.285	0.313	0.286	0.335	0.260	0.349	0.229	0.354
P_{25-10}	0.286	0.303	0.286	0.322	0.260	0.334	0.228	0.338
log P_{25-10}	0.270	0.295	0.274	0.319	0.252	0.335	0.223	0.341
P_{25-15}	0.276	0.291	0.278	0.312	0.254	0.326	0.224	0.331
log P_{25-15}	0.263	0.286	0.268	0.311	0.248	0.329	0.220	0.335
P_{25-30}	0.265	0.280	0.269	0.302	0.248	0.318	0.221	0.324
log P_{25-30}	0.256	0.277	0.262	0.303	0.243	0.322	0.217	0.330

TABLE H-3 (*continued*)

PREDICTOR SYMBOL	Q_5	LOG Q_5	Q_{10}	LOG Q_{10}	Q_{25}	LOG Q_{25}	Q_{50}	LOG Q_{50}
P_{25-60}	0.259	0.273	0.264	0.296	0.245	0.313	0.219	0.319
log P_{25-60}	0.251	0.272	0.258	0.299	0.240	0.318	0.216	0.326
P_{25-120}	0.253	0.267	0.259	0.291	0.242	0.309	0.217	0.315
SA	—0.172	—0.212	—0.172	—0.223	—0.159	—0.023	—0.144	—0.232
SB	—0.123	—0.083	—0.102	—0.074	—0.064	—0.065	—0.036	—0.060
SC	0.212	0.194	0.176	0.171	0.117	0.146	0.074	0.131
SD	0.011	0.004	0.028	0.029	0.043	0.053	0.050	0.069
CS	0.175	0.171	0.168	0.182	0.145	0.189	0.122	0.194
log CS	0.193	0.199	0.183	0.208	0.155	0.214	0.129	0.217
CC	0.157	0.086	0.191	0.121	0.215	0.151	0.217	0.166
log CC	0.169	0.096	0.200	0.129	0.219	0.157	0.218	0.170

Third, the predictor variables with the best coefficients are length of tributary (TRIB), area of watershed (A), and length of main stream (L) with coefficients of 0.580, 0.492, and 0.450, respectively, with Q_{10}. These, and the other higher coefficients, pertain to variables defining the unique physical characteristics of the watersheds.

Fourth, the correlations between return period precipitation and return period runoff average about 0.3. This is rather low, but it should be recalled that the variation in size of the watershed will tend to drive down the correlation between runoff and predictor variables not associated with basin size (such as precipitation intensity).

Finally, many of the simple correlation coefficients have little or no physical meaning. For example, those involving the slope variables are all negative, which is contrary to hydrologic reasoning. Also, the soil type correlations are smaller than expected, whereas those involving some climatic-type variables (e.g., July temperature) are relatively large. However, estimation of the simple correlation coefficient serves as an initial step leading to multiple regression equations aimed at improving prediction capability.

EVALUATION OF PREDICTOR SETS

The predictor variables were separated into three sets: topographic, hydrologic-climatic, and physiographic (soil). The sets were evaluated individually and in combination with other sets as summarized in Table H-1, experiments A-1 through A-5. For each of these five experiments, the predictors available for selection by the stepwise regression technique are given in Table H-2. The variables selected and discussion of the results are presented in the following. The regression equations for each experiment are given in Appendix A.

Hydrologic-Climatic Predictor Set—Experiment A-1

The developmental sample of 395 watersheds spread across all of the states in the sample was used to develop relationships between peak runoff in logarithmic form and hydrologic-climatic information. Such information is easily obtainable from tabulations and maps and does not require any determination from topographic (or other) maps of the watersheds. Thus, hydrologic-climatic information represents the simplest approach to the problem, because no requirement is made for a topographic map of each watershed.

The hydrologic-climatic predictor variables (Col. A-1, Table H-2) include such factors as mean annual temperature, snowfall, precipitation, and relative humidity; average number of days in the year with thunderstorms, below freezing, 1 in. or more of snowfall, and 0.01 in. or more of precipitation; monthly values such as July temperature, precipitation during wettest and driest month; and, finally, return period precipitation values. This set of hydrologic-climatic variables was presented to the step-wise regression technique, which selected out a small set of them. As indicated in Appendix G, under "Stepwise Regression," the small set is believed to contain all, or nearly all, of the estimation information contained in the largest. The variables selected for estimating log Q_{10} and log Q_{25} are given in Table H-4. Also listed with each variable is the square of the multiple correlation coefficient (also known as the reduction in variance), which is a measure of how well the variables selected to that point can estimate runoff. The correlations are rather low, the highest being only 20 percent, and it is concluded that hydrologic-climatic information by itself is of limited value in estimating peak runoff.

TABLE H-4

VARIABLES SELECTED FOR HYDROLOGIC-CLIMATIC PREDICTOR SET (EXPERIMENT A-1)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) log Q_{10}	
1	P_{100-5}	0.137
2	32F days	0.152
3	P_{5-120}	0.177
	(*b*) log Q_{25}	
1	July T	0.162
2	M24P	0.181
3	32F days	0.187
4	P_{5-120}	0.205

Topographic Predictor Set—Experiment A-2

In the second experiment of predictor sets, only topographic variables were considered. Here again the utility of a single source of basic information is examined. The predictors made available for selection (Col. A-2, Table H-2) were obtained from the topographic maps and represent unique signatures of the watershed. The variables selected for estimating log Q_{10} and log Q_{25} are given in Table H-5. Comparison of these results with those obtained in experiment A-1 indicates that topographic variables are more highly related to peak runoff, as a group and/or individually, than are hydrologic-climatic variables. Here, the reduction of variance has increased to more than 40 percent, which is about twice that achieved with climatic data.

TABLE H-5

VARIABLES SELECTED FOR TOPOGRAPHIC PREDICTOR SET (EXPERIMENT A-2)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) log Q_{10}	
1	TRIB	0.267
2	$\bar{E}$	0.358
3	A	0.398
4	S_{10}	0.426
	(*b*) log Q_{25}	
1	TRIB	0.265
2	S_{10}	0.353
3	$\bar{E}$	0.388
4	A	0.410

TABLE H-6

VARIABLES SELECTED FOR HYDROLOGIC-CLIMATIC AND TOPOGRAPHIC PREDICTOR SET (EXPERIMENT A-3)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) log Q_{10}	
1	TRIB	0.267
2	P_{100-05}	0.364
3	$\bar{E}$	0.400
4	A	0.446
5	T	0.455
6	P_{5-120}	0.466
7	P_{5-60}	0.488
	(*b*) log Q_{25}	
1	TRIB	0.264
2	P_{100-05}	0.367
3	$\bar{E}$	0.399
4	A	0.435
5	P_{dry}	0.452
6	T	0.464
7	P_{5-120}	0.482
8	P_{5-60}	0.495

However, some 60 percent of the variance of peak runoff is still unexplained by this relationship. Note that the same predictors were selected for both the log Q_{10} and log Q_{25} estimations. Although the order of selection differed between the two, the length of tributaries (TRIB) was selected first in both because it had the highest individual correlation coefficient of all the predictors available.

Hydrologic-Climatic and Topographic Predictor Set—Experiment A-3

In experiment A-3 the combined effect of hydrologic-climatic and topographic variables on peak runoff (log Q_{10} and log Q_{25}) is examined. The variables listed in Col. A-3, Table H-2, were made available for selection. The results are summarized in Table H-6. Examination of the variables selected finds a mixture of topographic and climatic variables. Again, length of tributaries is the single most important factor, with return period precipitation values being the dominant hydrologic-climatic variable selected. Here the reduction of variance has been increased to close to 50 percent, a statistically significant 8 to 10 percentage points more than topographic data alone.

Hydrologic-Climatic, Topographic Soils Predictor Sets—Experiments A-4 and A-5

For the experiments designed to examine the independent contribution of soil variables, it was necessary to use a smaller data sample of 335 cases. This was caused by the absence of soil information for 60 of the 395 cases of the dependent sample, due either to unanswered or incomplete questionnaires sent to SCS State Conservationists. The quantitative soil information obtained by questionnaire consisted of (a) percentages of hydrologic soil types (SA, SB, SC, SD) and (b) two composite parameters (CS and CC) based on the percentages of soil types and the percentages of land cover (as described in Appendix F) that attempt to depict the over-all character of the watershed. The other types of soil information are basically qualitative and as such are not suitable for consideration as predictors in a regression framework. Experiment A-4 considered both the individual soil percentages and the composite parameters, whereas A-5 considered only the composite parameters. In both experiments the predictands log Q_5 and log Q_{50} were included with log Q_{10} and log Q_{25}.

Tables H-7 and H-8 summarize the pertinent information related to the variables selected in experiments A-4 and A-5, respectively. It is obvious that the soil variables considered here are of little additional value to peak runoff estimation. In particular, the composite terms, which were computed to parameterize the total soil and land cover characteristics of the watershed, contribute little to the reduction of variance. It was only in the log Q_{50} relationship that a composite parameter was selected and it (CS) makes a minimal increase in the percent reduction of variance.

On the basis of experiments A-1 through A-5, it is concluded that: (1) topographic variables contain the most useful information regarding the estimation of peak runoff; (2) climatic factors add less, but still useful, information; and (3) soil variables (as considered in this study)

TABLE H-7

VARIABLES SELECTED FOR HYDROLOGIC-CLIMATIC, TOPOGRAPHIC, AND SOIL PREDICTOR SET (EXPERIMENT A-4)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(a) $\log Q_5$	
1	TRIB	0.285
2	$\bar{E}$	0.378
3	A	0.423
4	SA	0.443
5	Pot ET	0.463
	(b) $\log Q_{10}$	
1	TRIB	0.295
2	$\bar{E}$	0.381
3	A	0.422
4	Pot ET	0.447
5	SA	0.469
	(c) $\log Q_{25}$	
1	TRIB	0.295
2	$\bar{E}$	0.372
3	July T	0.409
4	A	0.446
5	SA	0.461
	(d) $\log Q_{50}$	
1	TRIB	0.291
2	July T	0.364
3	$\bar{E}$	0.408
4	A	0.441
5	SA	0.455

TABLE H-8

VARIABLES SELECTED FOR HYDROLOGIC-CLIMATIC, TOPOGRAPHIC, AND SOIL PREDICTOR SET (EXPERIMENT A-5)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(a) $\log Q_5$	
1	TRIB	0.285
2	$\bar{E}$	0.378
3	A	0.423
4	T days	0.443
5	T	0.459
	(b) $\log Q_{10}$	
1	TRIB	0.295
2	$\bar{E}$	0.381
3	A	0.422
4	Pot ET	0.447
5	S days	0.460
6	S_{10}	0.467
7	S	0.480
	(c) $\log Q_{25}$	
1	TRIB	0.295
2	$\bar{E}$	0.372
3	July T	0.409
4	A	0.446
	(d) $\log Q_{50}$	
1	TRIB	0.291
2	July T	0.364
3	$\bar{E}$	0.408
4	A	0.441
5	CS	0.452

make a minimal contribution. Therefore, soil variables were removed from further consideration in the study as predictors of peak runoff.

FUNCTIONAL FORM OF THE VARIABLES

A series of experiments was conducted to determine the functional form of the equations that estimate runoff. The form depends on the form of the variables in the equation, both predictand and predictors. Specifically, the predictands were expressed as arithmetic variables (e.g., Q_{10} in cfs), or in logarithmic form (e.g., $\log Q_{10}$). The predictors were expressed in three forms: arithmetic, logarithmic, and nonlinear.

Examples of the various functional forms follow. In these, Q represents a return period runoff and $X_1, X_2, \ldots$ represent predictor variables.

(a) Predictand and predictors in arithmetic form

$$\hat{Q} = A_0 + A_1X_1 + A_2X_2 + \ldots \qquad \text{(H-1)}$$

(b) Predictand and predictors in logarithmic form

$$\widehat{\log Q} = B_0 + B_1 \log X_1 + B_2 \log X_2 + \ldots \qquad \text{(H-2)}$$

which can be written as:

$$\tilde{Q} = 10^{B_0} X_1^{B_1} X_2^{B_2} \ldots \qquad \text{(H-3)}$$

(c) Predictand logarithmic and predictors arithmetic and logarithmic

$$\widehat{\log Q} = C_0\, C_1X_1 + C_2 \log X_2 + C_3 \log X_3 \ldots \qquad \text{(H-4)}$$

which can be written as:

$$\tilde{Q} = 10^{[C_0 + C_1X_1]} X_2^{C_2} X_3^{C_3} \qquad \text{(H-5)}$$

(d) Predictand logarithmic and predictors nonlinear

$$\widehat{\log Q} = D_0 + D_1f(X_1) + D_2f(X_2) + \ldots \qquad \text{(H-6)}$$

in which the f's are functions of the predictors as described in Chapter Four.

(e) Predictand logarithmic and predictors arithmetic

$$\widehat{\log Q} = E_0 + E_1X_1 + E_2X_2 + \ldots \qquad \text{(H-7)}$$

which can be written as:

$$\tilde{Q} = 10^{(E_0 + E_1 + E_2X_2 + \ldots)} \qquad \text{(H-8)}$$

The methodology for determining the functional form to be used in the regression equations was to develop estimation equations (such as Eq. H-1 through H-8), using the developmental sample of 395 watersheds, and then to compare the accuracy of the equations by applying them and equations developed in experiment A-3 to the independent sample of 98 watersheds. The form giving the best esti-

mates is then adopted. The techniques for developing the equations are described in Appendix G, as also are methods for verifying the accuracy of estimates made by the equations.

The estimation equations are discussed in the following section, followed by a comparison of the accuracy of the equations.

Development of Estimation Equations

Five sets of experiments were conducted: predictand and predictors in arithmetic form (B-1); predictand and predictors in logarithmic form (B-2); predictand in logarithmic form and predictors in both arithmetic and logarithmic form (B-3); and predictand in logarithmic form and predictors in nonlinear form (B-4 and B-5). Predictors made available to the screening procedure were from the climatic and topographic sets, as noted in Table H-2, Cols. B-1 through B-5.

TABLE H-9

VARIABLES SELECTED FOR STANDARD LINEAR REGRESSION FORM (EXPERIMENT B-1)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) Q_5	
1	TRIB	0.304
2	P_{100-05}	0.377
3	MAP	0.415
4	A	0.438
5	P_{dry}	0.460
6	T days	0.472
7	$\bar{E}$	0.508
8	P_{50-10}	0.517
9	P_{5-120}	0.530
10	P_{10-60}	0.540
	(*b*) Q_{10}	
1	TRIB	0.315
2	P_{100-05}	0.385
3	P_{50-10}	0.405
4	P_{5-120}	0.441
5	P_{dry}	0.473
6	P days	0.488
7	A	0.505
8	P_{10-60}	0.320
	(*c*) Q_{25}	
1	TRIB	0.283
2	July T	0.346
3	P_{wet}	0.363
4	P_{dry}	0.387
5	P_{100-05}	0.411
6	$\bar{E}$	0.421
7	T days	0.435
8	A	0.449
	(*d*) Q_{50}	
1	TRIB	0.241
2	July T	0.297
3	P_{dry}	0.317
4	$\bar{E}$	0.348
5	T days	0.380

Predictand and Predictors in Arithmetic Form—Experiment B-1

Both the return period runoff values and the predictor variables were used in their arithmetic form. In line with the results of the previous section on "Evaluation of Predictor Sets," in which it was found that topographic and climatic predictors are most useful, these types of variables were made available for selection. This experiment is quite similar to experiment A-3, with the exception that the predictand (peak runoff) in that experiment was in logarithmic form.

The variables selected by the screening procedure and the reduction of variance are given in Table H-9 for the predictands Q_5, Q_{10}, Q_{25}, and Q_{50}. Here again the selected predictors are about evenly mixed among topographic and climatic, with the length of tributaries (TRIB) being selected first for each predictand. The accumulative reduction of variance gets smaller as the return period increases from 5 year (Q_5) to 50 year (Q_{50}). This can be attributed to the fact that the 25- and 50-year log-normal estimates of Q represent, in many cases, extrapolations beyond the period of record of annual peaks, whereas the 5- and 10-year values were obtained by interpolation within the period of record.

The relationship developed for Q_{10} illustrates the specific form obtained in this phase; that is,

$$\begin{aligned}\widehat{Q}_{10} = &-2457.41 + 44.25(\text{TRIB}) + 4922.72(P_{100-5}) \\ &- 1137.13(P_{50-10}) + 11271.00(P_{5-120}) \\ &- 691.09(P_{dry}) + 8.43(\text{P-days}) \\ &+ 43.60(A) - 7444.42(P_{10-60}) \qquad \text{(H-9)}\end{aligned}$$

Eq. H-9 is a prediction equation; it is not an equation intended to describe the physical relationship between peak runoff and the other variables. It is invalid to draw physical meaning from the sign and magnitude of coefficients associated with the individual independent variables in this equation because in many cases they are compensating factors to achieve a least-square fit to the regression equation. Other statistical procedures, such as multivariate analysis and principal components analysis, are intended to identify independent sets of variables that may (or may not) be attached with physical or hydrological meaning.

Logarithmic Regression—Experiment B-2

In experiment B-2 the possible predictors were restricted to the logarithmic form of the topographic and climatic variables. The predictands were also considered in logarithmic form. Table H-10 gives pertinent information regarding the selected predictors for these relationships. It is acceptable to compare these reduction of variance statistics with those achieved in experiment A-3, because the predictand in both cases is log Q. In that regard the results are practically identical. In experiment A-3, for (log Q_{10}), $R^2 = 0.488$, whereas in this experiment $R^2 = 0.493$. One interesting difference in the variables selected is that, with the exception of log (TRIB), climatic variables were preferred in this experiment, whereas A-3 yielded a mix of topographic and climatic terms. The equation initially

TABLE H-10

VARIABLES SELECTED FOR LOGARITHMIC REGRESSION FORM (EXPERIMENT B-2)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) $\log Q_5$	
1	log TRIB	0.345
2	log P_{wet}	0.450
3	log July T	0.496
4	log MAP	0.509
	(*b*) $\log Q_{10}$	
1	log TRIB	0.344
2	log M24P	0.426
3	log P_{wet}	0.451
4	log July T	0.493
	(*c*) $\log Q_{25}$	
1	log TRIB	0.333
2	log M24P	0.424
3	log July T	0.458
4	log P_{wet}	0.483
	(*d*) $\log Q_{50}$	
1	log TRIB	0.322
2	log July T	0.423
3	log P_{wet}	0.467
4	log P days	0.482

developed by the regression technique for this form of relationship is

$$\widehat{(\log Q_{10})} = -4.35 + 0.44(\log \text{TRIB}) + 0.16(\log \text{M24P}) + 0.77(\log P_{wet}) + 3.38(\log \text{July } T) \quad \text{(H-10)}$$

which can be rewritten in nonlogarithmic form:

$$\hat{Q}_{10} = 0.000065(\text{TRIB})^{0.44}(\text{M24P})^{0.16}(P_{wet})^{0.77}(\text{July } T)^{3.38} \quad \text{(H-11)}$$

Logarithmic-Linear Regression—Experiment B-3

In this experiment, the topographic and hydrologic-climatic predictors in both logarithmic and arithmetic form were considered as possible predictors and the predictand was in logarithmic form. Table H-11 summarizes the variables selected. Log TRIB, having the highest individual correlation coefficient with the log-form of Q_5, Q_{10}, Q_{25}, and Q_{50}, was selected first for each relationship. There was, for each relationship, a mixture of logarithmic and arithmetic forms of both topographic and climatic variables. The form of the Q_{10} equation after transformation to eliminate logarithms is

$$\hat{Q}_{10} = 0.10(\text{TRIB})^{0.28}(E)^{0.18}(\text{MAT})^{1.58}(S)^{0.22} 10^{[0.019(A) - 0.00014(E) - 0.00069(S_{10}) + 0.0067(T \text{ days})]} \quad \text{(H-12)}$$

TABLE H-11

VARIABLES SELECTED FOR LOGARITHMIC-LINEAR REGRESSION FORM (EXPERIMENT B-3)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) $\log Q_5$	
1	log TRIB	0.345
2	$\bar{E}$	0.472
3	S_{10}	0.500
4	log Pot ET	0.518
5	log $\bar{E}$	0.533
6	log MAP	0.553
7	MAS	0.562
	(*b*) $\log Q_{10}$	
1	log TRIB	0.344
2	$\bar{E}$	0.468
3	T days	0.499
4	A	0.514
5	log $\bar{E}$	0.532
6	log MAT	0.546
7	S_{10}	0.557
8	log S	0.568
	(*c*) $\log Q_{25}$	
1	log TRIB	0.333
2	$\bar{E}$	0.450
3	log July T	0.496
4	log $\bar{E}$	0.515
5	S_{10}	0.532
6	32F days	0.543
	(*d*) $\log Q_{50}$	
1	log TRIB	0.322
2	$\bar{E}$	0.432
3	July T	0.489
4	log $\bar{E}$	0.513
5	S_{10}	0.526
6	32F days	0.537

Nonlinear Regression—Experiments B-4 and B-5

The nonlinear regression technique (described in Appendix G) was applied to the developmental sample of 395 watersheds to develop two estimation equations of the form

$$\widehat{(\log Q)} = B_0 + B_1 f_1(X_1) + B_2 f_2(X_2) + \ldots + B_p f_p(X_p) \quad \text{(G-2)}$$

in which the B's are constants determined by least squares and the f's are nonlinear functions of the predictor variables. The two equations differ in the respect that two of the eight predictor variables appearing in Eq. H-13 are omitted in Eq. H-14. The two variables (July T and E) were omitted to obtain an estimation equation with nonlinear variables that could be compared to one of the equations with linear variables only. The equation with all eight variables was retained for comparison with another one of the linear-variable equations. The objective of this section is to present and discuss the two equations with nonlinear variables. Selection of these variables for this phase of the study was based on a fairly exhaustive examination of some

20 arithmetic variables that were analyzed individually using the residual and rational fraction fit procedures described in Appendix G. Those retained for these experiments were the only ones exhibiting significant nonlinear characteristics identifiable by the procedure.

It will be recalled from Appendix G that the first step in the nonlinear technique is to apply the step-wise regression procedure to the arithmetic form predictor variables to develop an estimation equation. The two equations developed in experiments B-4 and B-5 are

$$\widehat{(\log Q_{25})} = -3.30 + 2.64(\text{July } T)0.23(\text{TRIB}) + 0.13(\text{MAF}_5) + 0.73(E) + 0.64(A) + 0.10(\text{REG}_3) - 0.31(\text{REG}_2) + 1.23(P_{10\text{-}360}) \qquad \text{(H-13)}$$

$$\widehat{(\log Q_{25})} = 1.30 + 0.88(\text{TRIB}) + 1.01(\text{MAF}_5) + 0.84(A) + 0.13(\text{REG}_1) + 0.39(\text{REG}_3) + 1.12(P_{10\text{-}360}) \qquad \text{(H-14)}$$

Because the development of nonlinear relationships represents a time-consuming computer effort, only Q_{25} was used in the initial investigation of the nonlinear regression technique.

The next step is to compute nonlinear functions of each of the predictor variables in Eqs. H-13 and H-14. First, however, a discussion of the regional variables is needed. These variables are 0-1 variables, with a variable taking a value of 1 for basins located inside the region indicated and a value of 0 for basins not inside that region. It is not possible to develop nonlinear functions of 0-1-type variables; thus, they were, by necessity, excluded from this phase of the evaluation.

A second feature of nonlinear representations of a variable is to note that although some variables are common to both Eq. H-13 and Eq. H-14 (e.g., area), their nonlinear functions are not the same. This can be shown by considering the method for computing nonlinear variables. It is recalled, from Appendix G, that net residuals are computed, defined as

$$E_{ki} = \log Q_{ki} - \log Q_i \qquad \text{(G-3)}$$

in which Q_i is the observed return period runoff in the ith basin and Q_{ki} is computed by Eq. H-13 or Eq. H-14. Because Q_{ki} is different between the two equations, E_{ki} is different; and because E_{ki} is used in compilation of the nonlinear functions, these too will be different.

The nonlinear functions are of the form

$$f(X) = C + C_1 j \frac{X}{(1+RX)^j} + C_2 j \frac{X^2}{(1+RX)^j} + \ldots \qquad \text{(G-4)}$$

in which j is the degree of the equation and can be either 1, 2, or 3, and the C's and R are constants determined by the procedure.

The nonlinear functions for the six variables of Eq. H-13 are

$$f(\text{July } T) = 4.74 - 0.14(\text{July } T) + 21.45(\text{July } T)^2$$
$$f(\text{TRIB}) = -0.18 + 0.087(\text{TRIB}) - 0.054(\text{TRIB})^2$$
$$f(\text{MAF}_5) = 0.13 + 0.13(\text{MAF}_5)$$
$$f(E) = -0.23 + 0.0031\frac{E}{(1+0.0029E)^2} + 0.0000007\frac{E^2}{(1+0.0029E)^2}$$
$$f(A) = -0.12 + 0.081\frac{A}{(1+0.097A)}$$
$$f(P_{10\text{-}360}) = 0.36 - 0.13(P_{10\text{-}360}) + 2.48(P_{10\text{-}360})^2$$

Similarly, the nonlinear functions for the four variables in Eq. H-14 are

$$f(\text{TRIB}) = -0.26 + 0.40\frac{\text{TRIB}}{(1+0.53(\text{TRIB})^2} + 0.14\frac{\text{TRIB}^2}{(1+0.53(\text{TRIB})^2}$$
$$f(\text{MAF}_5) = 0.13 + 0.16(\text{MAF}_5) - 0.0052(\text{MAF}_5)^2$$
$$f(A) = -0.31 + 0.26\frac{A}{(1+0.23A)}$$
$$f(P_{10\text{-}360}) = 0.31 + 0.19(P_{10\text{-}360})$$

For Eq. H-13, the mean annual flood (MAF_5) result yielded a best-fit relationship that was a linear function rather than a nonlinear one. Thus, for this particular variable for this particular equation, a nonlinear form was not found that was better correlated with the runoff residual than was the original arithmetic variable. On the other hand, the gage elevation result is quite nonlinear, the equation being of second degree with a denominator term. The result for gage elevation was expected because a preliminary plot of gage elevation vs log Q_{25} exhibited nonlinear tendencies.

The six nonlinear variables of Eq. H-13 listed and the two 0-1 variables of Eq. H-13 processed by the step-wise regression technique to produce

$$\widehat{\log Q_{25}} = -3.25 + 0.98[f(\text{July } T)] + 1.09[f(\text{TRIB})] + 1.01[f(\text{MAF}_5)] + 1.19[f(E)] + 0.91[f(A)] + 0.14(\text{REG}_3) - 0.46(\text{REG}_2) + 1.10[f(P_{10\text{-}360})] \qquad \text{(H-15)}$$

The corresponding result for Eq. H-14 is

$$\widehat{\log Q_{25}} = 0.85[f(\text{TRIB})] + 1.09[f(\text{MAF}_5)] + 0.84[f(A)] + 0.33(\text{REG}_1) + 0.15(\text{REG}_3) \qquad \text{(H-16)}$$

The four variables common to Eqs. H-15 and H-16 (TRIB, MAF_5, A, REG_3) have quite similar coefficients. This indicates that the importance of these variables in estimating Q_{25} is nearly the same, no matter which equation (H-13 or H-14) was the starting point.

Comparison of Alternative Forms

The relationships developed to estimate the 25-year return period peak runoff (Q_{25}) were evaluated on the independent data set as the basis for comparison of the alternative functional forms that were considered. The relationships were developed from six experiments that represent the various functional forms:

A-3 (predictand logarithmic and predictors arithmetic)

B-1 (predictand and predictors both arithmetic)

B-2 (predictand and predictors both logarithmic)

B-3 (predictand logarithmic and predictand both arithmetic and logarithmic)

B-4 (predictand logarithmic, six nonlinear predictors plus two regional predictors in 0-1 form)

B-5 (same as B-4 with two nonlinear predictors omitted)

Evaluation of the peak runoff estimation equations on watersheds not included in the developmental data sample provides a valid measure of the value of the equations to design engineers. Methods that work well on the data used to develop the equations and then fail badly on other data are useless to an engineer attempting to estimate runoff from an ungaged watershed. Therefore, the evaluation of alternative functional forms, and subsequent experimentation discussed later, is based on the performance of the equations when applied to independent data. The watersheds randomly selected and withheld from the developmental sample in this study consisted of 98 locations.

The six equations were applied by computer to the data for the 98 basins. In the cases where the predictand was in logarithmic form (each equation except Eq. B-1), the estimates were converted to arithmetic units, as described in Appendix G. The ratio $(\bar{Q}/\bar{Q}')$ was computed on the developmental sample cases and applied to each case of the independent sample.

The Q_{25} estimates were verified against the corresponding observed values of Q_{25} (i.e., the 25-year return period values determined by fitting the log-normal distribution to observed values of annual peak runoff). Two types of verification were used (see Appendix G):

1. The root-mean-square (RMS) of the differences (D) between estimated and observed values,

$$\text{RMS} = \left\{ (1/98) \sum_{i=1}^{98} D^2 \right\}^{1/2} \qquad \text{(G-7)}$$

2. Frequency distribution of "percent errors," defined as

$$\text{PE}_i = (D_i/Q_i) \times 100 \text{ percent} \qquad \text{(G-10)}$$

in which Q_i is the observed 25-year return period runoff for the ith watershed. The frequency distribution is a count of the number of errors within each of seven specified ranges of values from small errors of −10 to +10 percent to very large errors of more than 200 percent.

The results of this evaluation are summarized in Table H-12, which includes the RMSE and percent error distribution (PE) statistics. The RMSE is in units of cubic feet per second (cfs), while the number of cases falling in each error category is denoted. There is little difference between the various statistics presented. The nonlinear experiment (B-5) yielded the lowest RMSE (2,108), which is less than 10 percent lower than the rest. Although there are some small differences between the error distributions of the six relationships, they are generally the same. Experiment B-3 has the greatest number of errors less than 10 percent (12), but it also yielded a large number of errors in excess of 200 percent (20).

A characteristic of the relationship in arithmetic form (B-1) that would limit its acceptability by design engineers is that there were several cases (about 15 percent of the independent sample) in which the peak discharge estimated by the equation was a negative quantity. In just about every case the watershed for which this occurred was either very small or was in a region of small runoff such that the error of estimate was not particularly large but the estimate itself was negative. On this basis, and the fact that logarithmic relationships yielded comparable statistics, subsequent experimentation did not consider relationship in which the predictand is in arithmetic form.

TABLE H-12

EVALUATION OF ALTERNATIVE FUNCTIONAL FORMS ON INDEPENDENT DATA FOR Q_{25} ESTIMATES

	A-3	B-1	B-2	B-3	B-4	B-5
RMSE (CFS)	2397	2385	2342	2315	2315	2108
$-10 < \text{PE} \leq 10$	7	9	6	12	7	7
$10 < \text{PE} \leq 25$ $-25 < \text{PE} \leq -10$	19	8	10	10	12	10
$25 < \text{PE} \leq 50$ $-50 < \text{PE} \leq -25$	17	17	22	15	21	24
$50 < \text{PE} \leq 100$ $-100 < \text{PE} \leq -50$	28	15	25	23	29	24
$100 < \text{PE} \leq 150$ $-150 < \text{PE} \leq -100$	6	9	11	11	7	10
$150 < \text{PE} \leq 200$ $-200 < \text{PE} \leq -150$	1	10	6	7	8	9
$200 < \text{PE}$ $\text{PE} \leq -200$	20	30	18	20	14	14

Development of the nonlinear relationships (experiments B-4 and B-5) is a time consuming computer procedure and application of the resulting relationship in practice would also be appreciably more complicated. Although the results obtained here suggest a slight improvement using nonlinear techniques, it was concluded that the improvement was not sufficient to warrant its continued use through the rest of the experimentation.

Thus, by elimination, a functional form consisting of a logarithmic predictand is preferred. The choice regarding the form of the predictors can not be resolved on the basis of these results and was, therefore, considered further in subsequent analyses.

STRATIFICATION INTO SUBSETS

All of the prior experimentation was limited to the development of single national equations; i.e., one equation to apply to all watersheds. An alternative to this approach is to attempt to separate the basins into "hydrologically homogeneous" sets using some criterion thought to be important from a hydrologic viewpoint. Separate regression equations are then developed within each set. The assumption is that the relationship between runoff and the predictor variables varies significantly from set to set.

Nine different criteria were used to categorize the watersheds into homogeneous sets. These are identified in Table

H-1 as experiments C-1 through C-9. The criteria were: mean annual flood (MAF_5), 10-year 60-min precipitation (P_{10-06}), mean annual precipitation (MAP), mean annual temperature (MAT), area of watershed (A), mean basin elevation (E), soil erosion classification (SE), geological zone classification (GZ), and geographic location of the watershed (REG). The number of sets of watersheds varied from two for basin elevation (E) to six for geographic location (REG). The criteria were applied first to the dependent sample of 395 watersheds, then to the independent sample of 98 watersheds.

The stratified sets of watersheds of the dependent sample were used to develop equations for estimating runoff. The procedure was similar for all sets for all nine criteria. The predictand variables were in logarithmic form; i.e., log Q_5, log Q_{10}, log Q_{25}, log Q_{50}. The predictors were topographic and climatic variables. The step-wise regression technique was applied to select out a small set of predictor variables and develop a regression equation expressing log of peak runoff as a linear function of the predictor variables.

The relative predictive capability of each of the nine methods of stratifications was determined by comparing the accuracy of estimates made by the stratified equations on independent watersheds with the accuracy of estimates made by national equations. The methodology of comparison is similar to that described earlier under "Functional Form of the Variables, Comparison of Alternative Forms." Accuracy is measured in two ways—by the root-mean-square of the differences between estimated and observed runoff values, and by the frequency distribution of the percent errors.

The nine stratification criteria fall into five groups. The following presents the method of stratification, the number of watersheds within each set, and one regression equation (log Q_{25}) for each set. The remaining equations (for log Q_5, log Q_{10}, log Q_{50}) are omitted to save space. The verification results for Q_{25} also are presented in a subsequent section. Verifications for the other three predictands were quite similar and discussions are omitted.

Mean Annual Flood (MAF_5) Stratification—Experiment C-1

Each watershed of the sample had been classified into one of five categories of mean annual flood frequency as described in Appendix F. For this stratification experiment the cases of the developmental sample for each of the five categories were utilized separately to develop a set of regression equations. Figure H-1 shows the categories of mean annual flood (MAF_5).

Table H-13 summarizes the predictors selected by the screening procedure for each category for the Q_{25} relationships. The regression equations for Q_5, Q_{10}, Q_{25}, and Q_{50} are given in Appendix A. The selected predictors are headed by a topographic variable with the exception of category b (10 to 30 csm). There are several climatic variables (MAT, July T, MAS) selected for the various regions, although topographic terms (T, TRIB, L, A) are preferred more often. Similar variables were selected for the other predictands.

TABLE H-13

VARIABLES SELECTED FOR STRATIFICATION BASED ON MAF_5 FOR LOG Q_{25} EQUATIONS (EXPERIMENT C-1)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(a) $MAF_5=1$ ($0 < MAF \leq 10$)	
1	log T	0.30
2	log MAT	0.42
3	log TRIB	0.47
4	log L	0.55
	(b) $MAF_5=2$ ($10 < MAF \leq 30$)	
1	log July T	0.26
2	log MAS	0.43
3	log TRIB	0.49
	(c) $MAF_5=3$ ($30 < MAF \leq 50$)	
1	log A	0.47
2	log M24P	0.52
	(d) $MAF_5=5$ ($50 < MAF \leq 90$)	
1	log T	0.58
2	log P days	0.64
3	log A	0.72
4	log T days	0.75
	(e) $MAF_5=5$ ($90 < MAF$)	
1	log TRIB	0.41
2	log S_{10}	0.50
3	log A	0.57

It is meaningless to compare the various statistics among the several categories. Comparing the residual standard deviation (σ_r) achieved by a comparable unstratified experiment (B-2) to one computed from the weighted average procedure provides a relative measure of the relationships developed with regard to the dependent sample. The procedure for computing the weighted residual standard deviation is

$$\sigma_r' = \frac{1}{N}\sum_{i=1}^{g} \sigma_{ri}\, n_i \qquad \text{(H-17)}$$

in which σ_{ri} is the residual standard deviation for category i, n is the category i size, g is the number of categories, and N is the sample size. Such a comparison for this method of stratifying yields a residual standard deviation of 0.47 for the national experiment (B-2), whereas stratifying yields a weighted value of 0.37. On this basis, the MAF_5 stratification yields a better relationship than the single national equation.

Climatic Stratifications—Experiments C-2, C-3, and C-4

Three alternative methods for stratifying the sample based on hydrologic-climatic considerations were formulated. The first used the variable P_{10-60}, the 10-year return period value of maximum 60-min precipitation; the second used mean annual precipitation; the third was based on mean annual temperature. Each is discussed separately in the following.

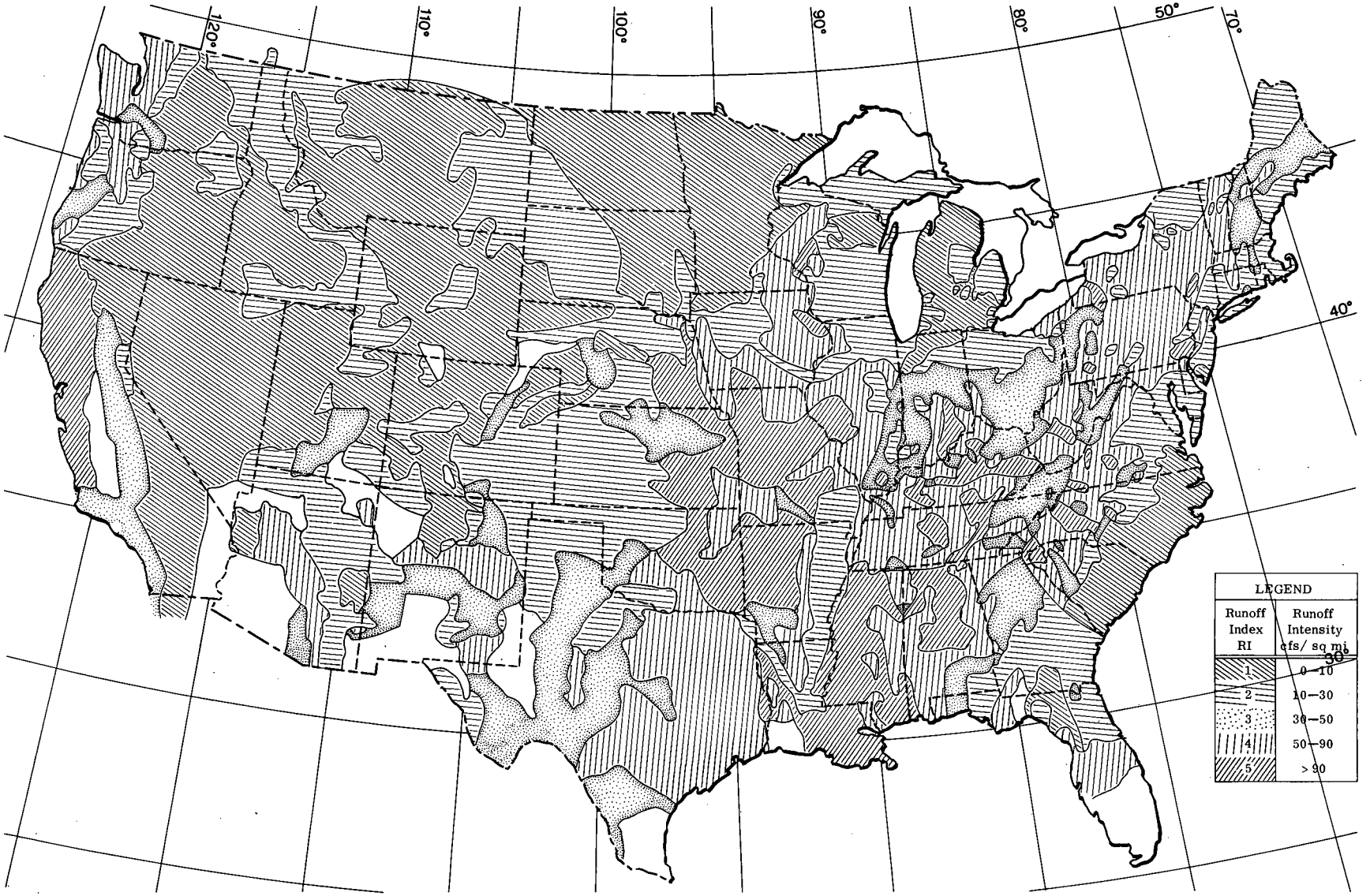


Figure H-1. Map of mean annual flood (MAF_5) regions in the contiguous United States.

The dependent and independent data samples were stratified in three categories of P_{10-60} wherein the categorical limits were selected such that the samples were broken into subsets with approximately the same number of cases in each. The categories so defined were; category a, $P_{10-60} \geq 0.80$ in.; category b, $0.8 < P_{10-60} \geq 1.5$; and category c, $P_{10-60} > 1.5$. These regions are shown on Figure H-2, which also gives the number of cases in each category for the dependent and independent samples.

The variables selected for the relationships developed for Q_{25} are given in Table H-14 and the regression equations are given in Appendix A. The selection of log TRIB first for categories b and c (higher intensity precipitation) suggests the relative importance here of the drainage network (i.e., the ability of the watershed to discharge its runoff quickly) as compared to regions where rainfall intensities are lighter. The weighted residual standard deviation for this stratification (computed from Eq. H-17) is 0.425 compared to 0.467 for the national experiment (B-2). Again, the stratified equations seem to give somewhat better results.

TABLE H-14

VARIABLES SELECTED FOR STRATIFICATION BASED ON P_{10-60} FOR LOG Q_{25} EQUATION (EXPERIMENT C-2)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) $0.0 < P_{10-60} \leq 0.8$	
1	log A	0.36
2	log T days	0.53
3	log DD	0.57
4	log P_{wet}	0.60
5	log L	0.62
6	log TRIB	0.65
	(*b*) $0.8 < P_{10-60} \leq 1.5$	
1	log TRIB	0.33
2	log July T	0.48
3	log S_{10}	0.50
4	log Pot ET	0.54
	(*c*) $1.5 < P_{10-60}$	
1	log TRIB	0.28
2	log P_{10-30}	0.37

TABLE H-15

VARIABLES SELECTED FOR STRATIFICATION BASED ON MAP FOR LOG Q_{25} EQUATION (EXPERIMENT C-3)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) MAP ≤ 34.0	
1	log $\bar{E}$	0.24
2	log TRIB	0.42
3	log M24P	0.46
	(*b*) $34.0 <$ MAP ≤ 47.0	
1	log TRIB	0.25
2	log July T	0.39
3	log $\bar{E}$	0.45
4	log A	0.48
5	log P_{wet}	0.52
	(*c*) $47.0 <$ MAP	
1	log A	0.53
2	log T days	0.62

With regard to the stratification based on mean annual precipitation, the procedure of dividing the sample into nearly equal subsets was used to resolve categorical limits. Thus, category a included all cases of MAP ≤ 34 in.; category b, $34 <$ MAP ≤ 47; and category c, MAP > 47. Figure H-3 shows these regions for the U.S., as well as the breakdown of the dependent and independent cases.

Table H-15 gives the variables selected for three categories of MAP for the predictand Q_{25}. Here again, topographic features are more frequently selected than are climatic variables. The weighted residual standard deviation (σ_r') is computed to be 0.436 for the regionalization, which is slightly better than the national equation, whose σ_r is 0.467.

The last method of stratification based on hydrologic-climatic factors considered mean annual temperature (MAT). Figure H-4 shows the isotherms delineating the categorical limits determined for this variable. These limits were also determined by dividing the dependent sample into three subsets such that the subset sample sizes were approximately equal. The number of cases in the dependent and independent samples for each category of MAT also are shown in Figure H-4.

In two of the three categories of MAT, the variable log TRIB was selected first, as shown in Table H-16, which summarizes the selected variables. The residual standard deviation for this method of stratification, as computed by Eq. H-17, is 0.395, with the relationship developed for category b contributing most to this lower value.

Topographic Stratification—Experiments C-5 and C-6

Two of the many topographic factors available were selected for the purpose of evaluating stratified relationships. They were the size of the watershed area (A) and the mean basin elevation ($\bar{E}$), noted in Table H-1 as experiments C-5 and C-6.

With regard to the area stratification experiment, the dependent sample was separated as follows: category a, $A \leq 5$ sq mi; category b, $5 < A \leq 15$; category c, $A > 15$. The cases falling in these categories for the dependent and independent samples are given in Table H-17.

The selected variables are given in Table H-18 for this stratification. An interesting difference is that the first variable selected in each case was a climatic factor. Inasmuch as the cases within a given subset of areas (e.g., 0 to 5 sq mi) are distributed all over the U.S., climatic variables, to a certain extent, act as geographical indices. This may explain why they were preferred within this stratification although not in earlier experimentation. The residual standard deviation for the area stratification was

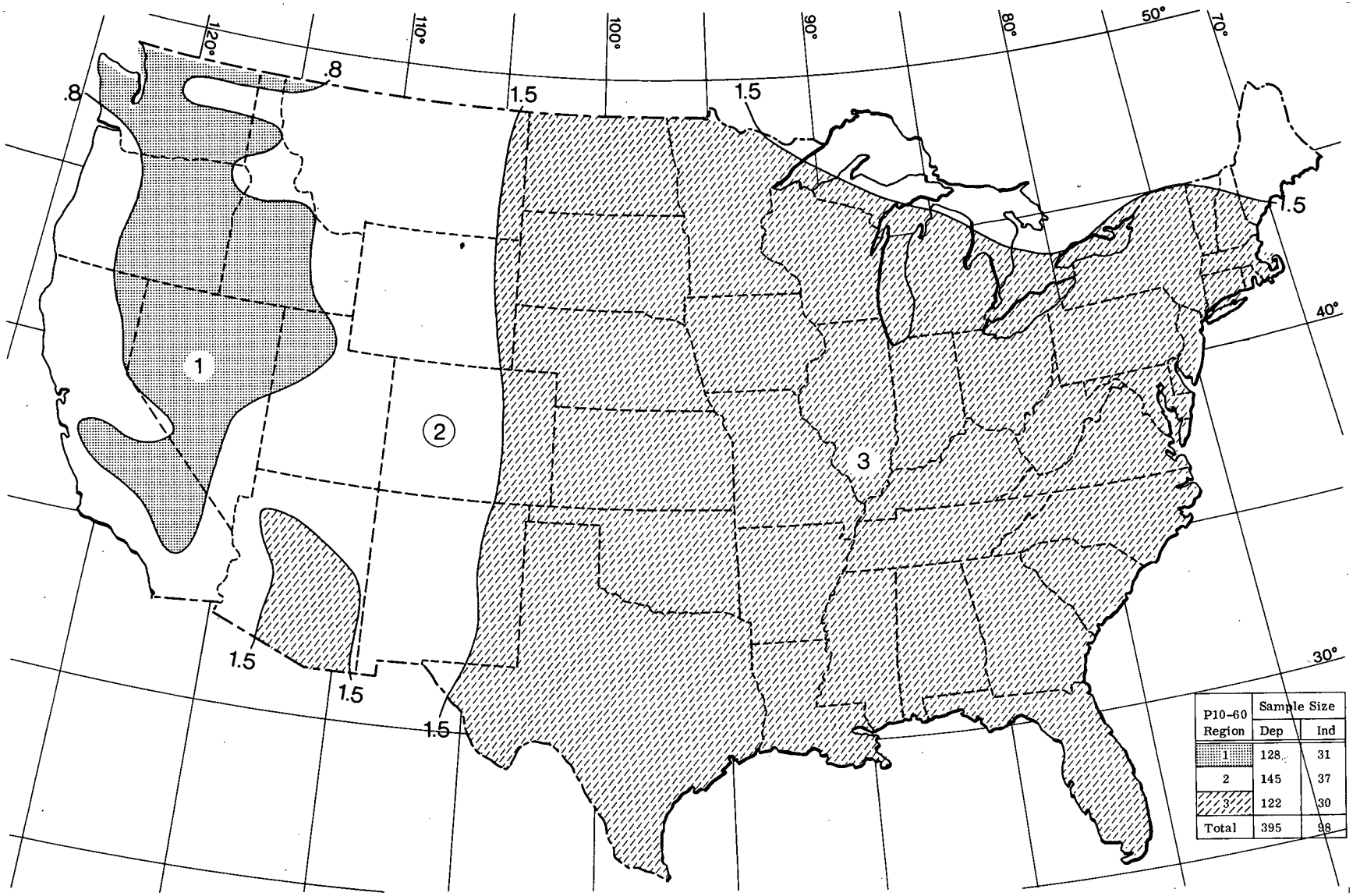


Figure H-2. Map of 10-year frequency, 60-min rainfall amount (P_{10-60}) regions in the contiguous United States.

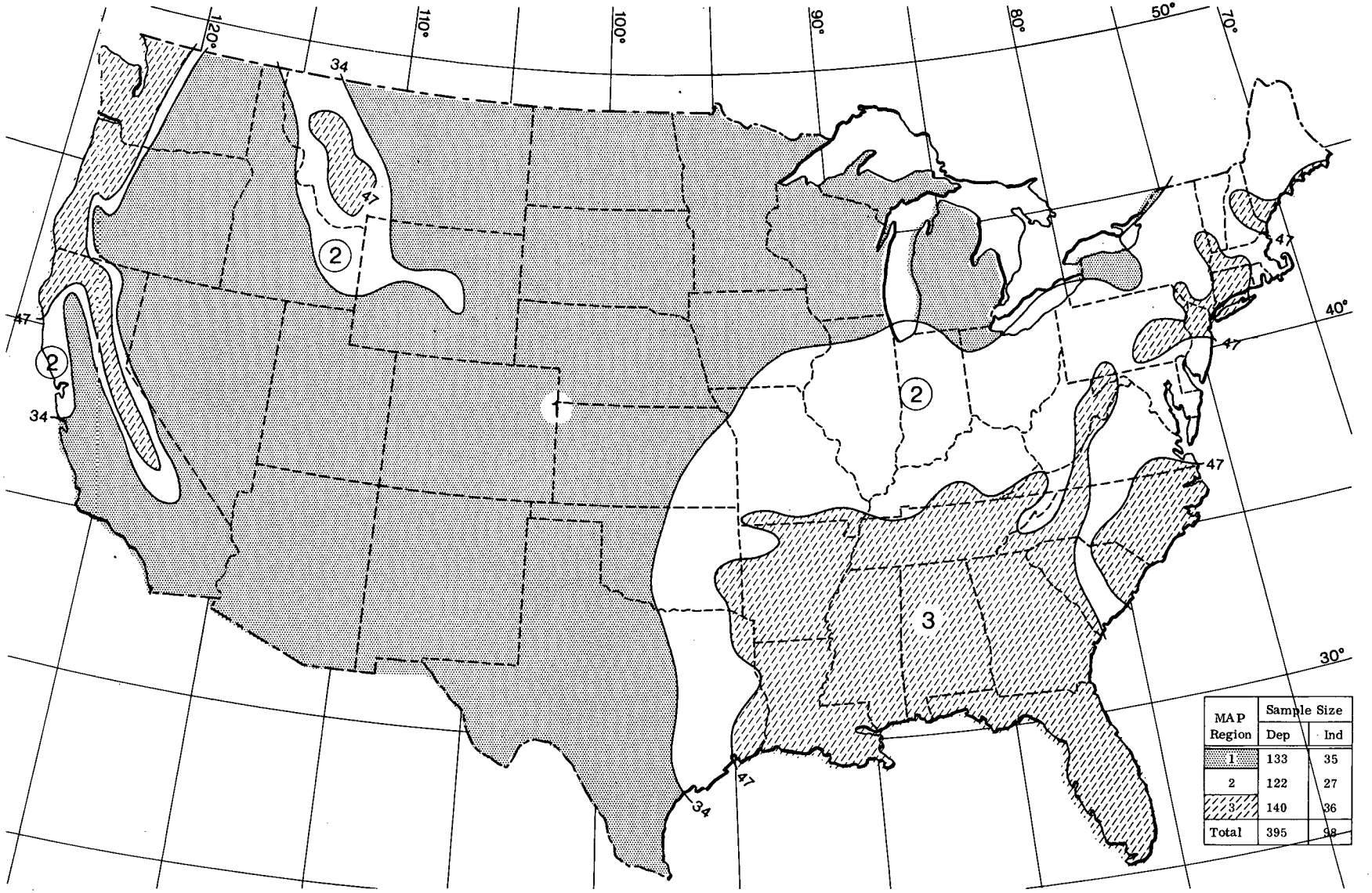
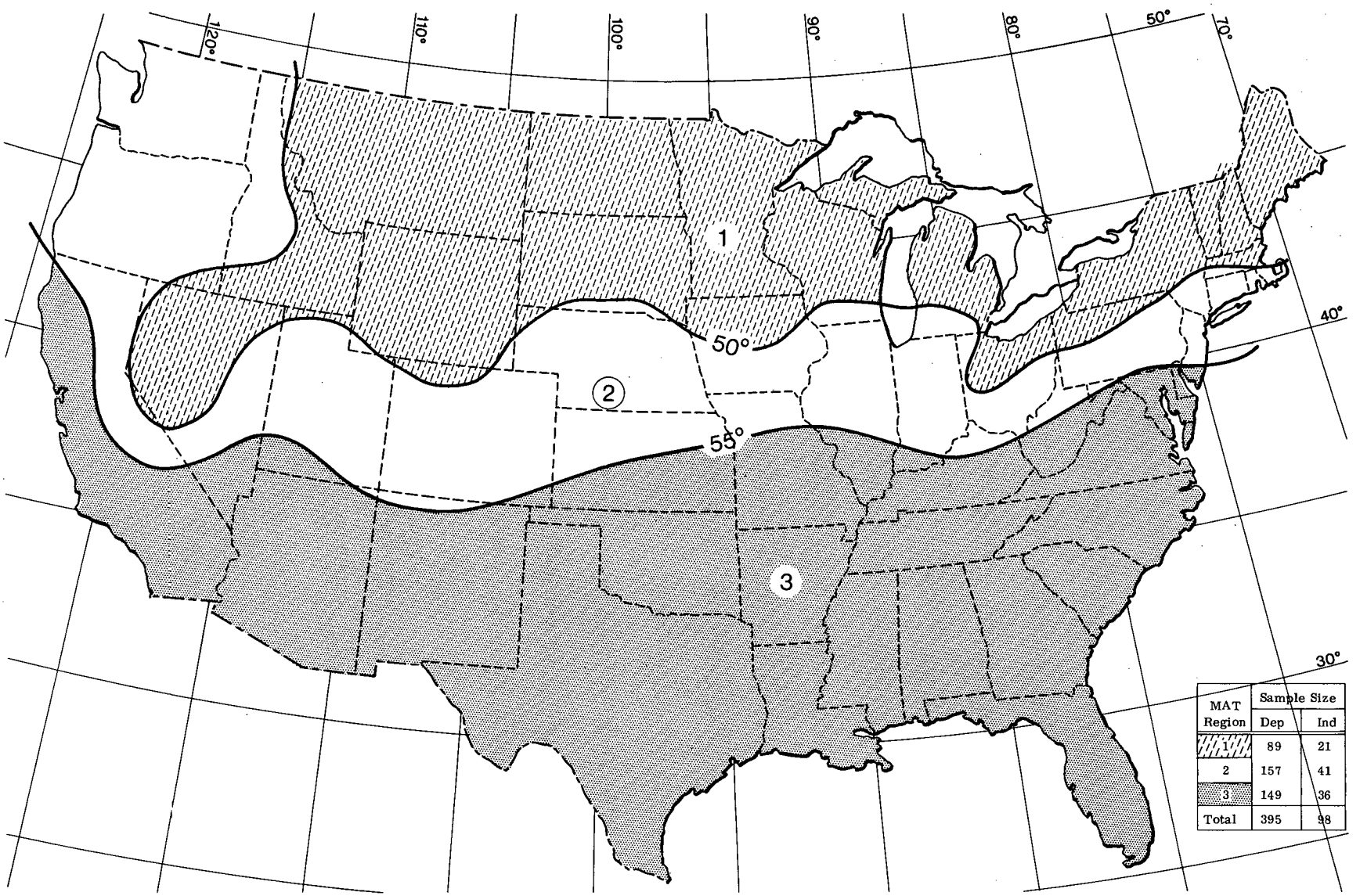


Figure H-3. Map of mean annual precipitation (MAP) regions in the contiguous United States.



MAT Region	Sample Size Dep	Sample Size Ind
1	89	21
2	157	41
3	149	36
Total	395	98

Figure H-4. Map of mean annual temperature (MAT) regions in the contiguous United States.

TABLE H-16

VARIABLES SELECTED FOR STRATIFICATION BASED ON MAT FOR LOG Q_{25} EQUATION (EXPERIMENT C-4)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) MAT $\leq$ 50.0	
1	log TRIB	0.32
2	$\bar{E}$	0.50
	(*b*) 50.0 $<$ MAT $\leq$ 55.0	
1	log T	0.45
2	P_{10-60}	0.56
3	log P_{wet}	0.59
4	log A	0.65
5	T days	0.68
6	P_{dry}	0.69
7	$\bar{E}$	0.72
8	log $\bar{E}$	0.76
9	log TRIB	0.78
	(*c*) 55.0 $<$ MAT	
1	log TRIB	0.27
2	S_{10}	0.34
3	32F days	0.43

TABLE H-17

DISTRIBUTION OF CASES FOR AREA STRATIFICATION

CATEGORY	SAMPLE SIZE DEP.	SAMPLE SIZE IND.
(a) $A \leq 5.0$	130	40
(b) $5.0 < A \leq 15.0$	144	34
(c) $15.0 < A$	121	24
Total	395	98

TABLE H-18

VARIABLES SELECTED FOR STRATIFICATION BASED ON A FOR LOG Q_{25} EQUATIONS (EXPERIMENT C-5)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(a) $A \leq 5.0$	
1	log P_{10-10}	0.22
2	log TRIB	0.32
	(*b*) 5.0 $< A \leq$ 15.0	
1	log M24P	0.14
2	log TRIB	0.21
3	log P_{wet}	0.25
4	log $\bar{E}$	0.28
5	log S_{10}	0.33
6	log MAS	0.37
	(*c*) 15.0 $< A$	
1	log July T	0.26
2	log MAP	0.36
3	log DD	0.46

0.451, just slightly better than the national equation value of 0.467.

The stratification based on basin elevation (MSL) involved dividing the watersheds into two groups: (1) 1,000 ft and lower, and (2) over 1,000 ft. The main stream elevation at the point 0.5L was used as the mean basin elevation for this experiment. Figure H-5 shows the regions of the country in excess of 1,000 ft elevation (hatched area), as well as the number of cases in the dependent and independent sample for each of the two categories.

Variables selected for these two categories of mean basin elevation for the predictand Q_{25} are given in Table H-19. Even though basin elevation has been used to stratify the sample, the variable log E has been selected in both categories. The value of σ_r' for this experiment is 0.433.

Physiographic Stratifications—Experiments C-7 and C-8

The stratification experiments based on knowledge of the physiographic (soil) features were limited to the soil erosion variable and the geological zones defined by Potter (*42*). The other soil classifications were not considered, because, being defined into upward of ten categories each, there was not sufficient sample size in most categories to warrant statistical analysis. Also, it did not appear desirable to combine two or more categories into a single one of mixed or unknown hydrologic characteristics.

Concerning the soil erosion stratification scheme, classes 2, 3, and 4 (refer to Appendix F for definition) have been combined into a single category for stratification because there were not enough cases in each individually to warrant statistical analysis. The geographical distribution of the four stratification categories is shown in Figure H-6. It should be noted in Figure H-6 that most cases of the dependent sample fall in category c (201 cases) and category d (114 cases).

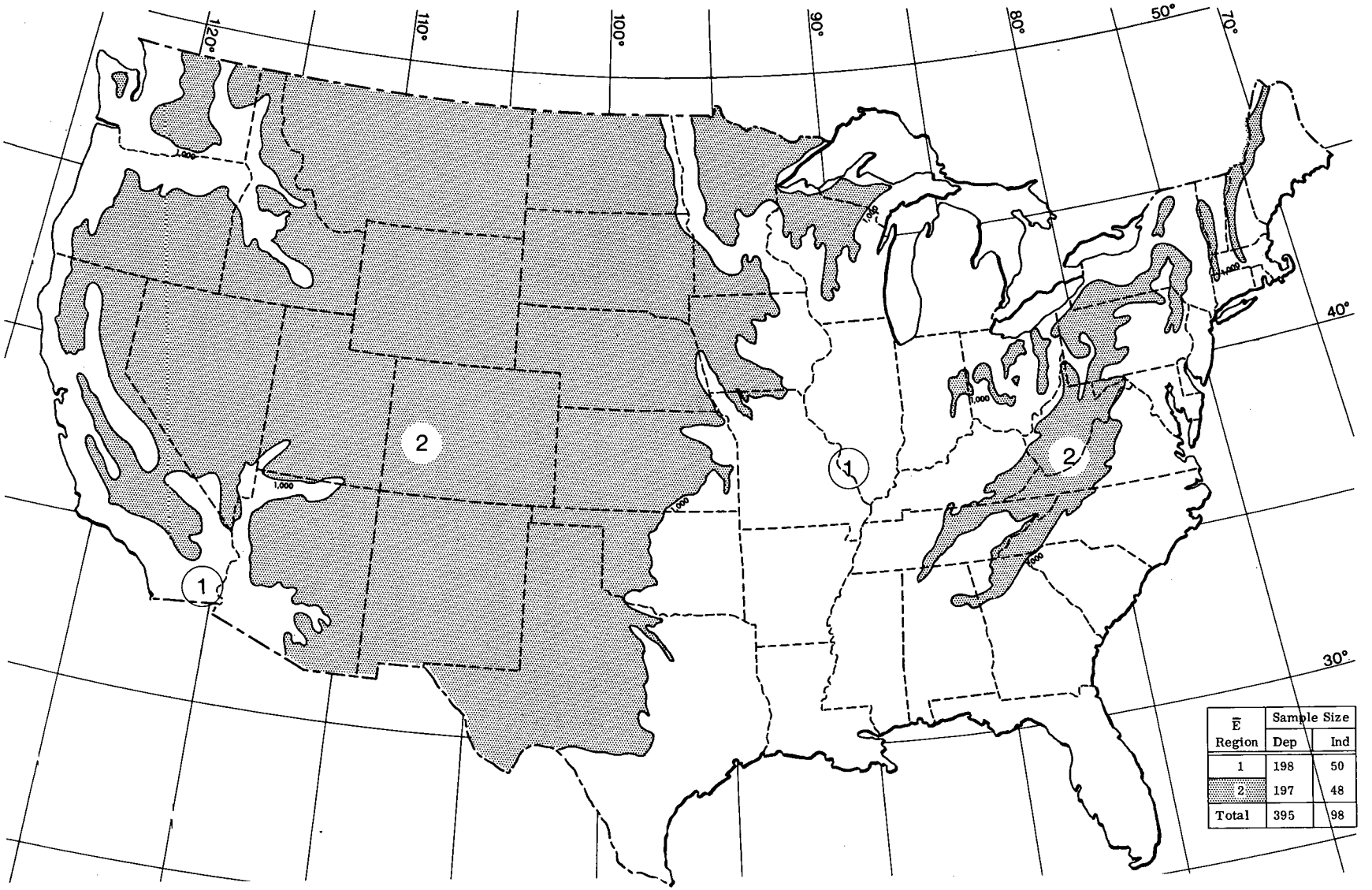
Table H-20 gives, for the Q_{25} relationships, the variables selected for each region of soil erosion. The equations are given in Appendix A. The residual standard deviation here is equal to 0.375.

Potter did not delineate the geologic zones (Fig. H-7) for the entire nation. Thus, many watersheds in the data sample were not classified geologically. In fact, 236 of the 395 basins of the developmental sample were undefined, as shown in Figure H-7. For the purpose of developing stratification equations, these cases were evaluated as a separate group identified as GZ_5 (category e in Table H-21).

The results of the selection procedure are given in Table H-21 for this method of stratifying; the weighted σ_r' is 0.388 compared to the national equation value of 0.467 for the dependent sample.

Geographical Stratification—Experiment C-9

The last method of stratification that was investigated was geographical. The country was divided into six regions



$\overline{E}$ Region	Sample Size Dep	Ind
1	198	50
2	197	48
Total	395	98

Figure H-5. Map of mean basin elevation ($\overline{E}$) regions in the contiguous United States.

TABLE H-19

VARIABLES SELECTED FOR STRATIFICATION BASED ON $\bar{E}$ FOR LOG Q_{25} EQUATIONS (EXPERIMENT C-6)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) $\bar{E} \leq 1000.0$	
1	log TRIB	0.49
2	log July T	0.57
3	log A	0.59
4	log $\bar{E}$	0.62
5	log MAS	0.64
	(*b*) $\bar{E} > 1000.0$	
1	log TRIB	0.24
2	log $\bar{E}$	0.38
3	log S_{10}	0.43
4	log MAT	0.48

(Fig. H-8) dictated to a certain extent by the availability of cases in the data sample. Longitude 81°W defines the western boundary of region 1 (REG_1), which in effect isolates most watersheds draining ultimately into the Atlantic Ocean from the rest. Regions 2 and 3 (REG_2 and REG_3) are separated at latitude 36.5°N such that to the north appreciable snow cover is generally the rule and spring peak floods, due in part to snowmelt, are more common than farther south where peak floods are generally associated with summer convective activity. The Rocky Mountain basins are grouped into the region bounded by longitude 104°W and 116°W (REG_4); the Far West is stratified into two regions by latitude 42°N, placing the states of Washington and Oregon in one (REG_5) and California and Nevada in the other (REG_6). For the purposes of this study, this last group is, practically speaking, restricted to California due to the availability of just one watershed in Nevada. Cases falling into each of these regions for the data sample also are given in Figure H-8.

Variables selected by the screening procedure for each region are given in Table H-22. Using Eq. H-17, the weighted residual standard deviation for the dependent sample for this method is 0.389.

TABLE H-20

VARIABLES SELECTED FOR STRATIFICATION BASED ON SE FOR LOG Q_{25} EQUATIONS (EXPERIMENT C-7)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) SE=1	
1	$\bar{E}$	0.48
2	A	0.64
3	P_{dry}	0.67
4	M24P	0.71
	(*b*) $2 \leq SE \leq 4$	
1	log MAP	0.15
2	TRIB	0.27
3	S_{10}	0.36
4	log MAT	0.60
	(*c*) SE=5	
1	log T	0.50
2	log TRIB	0.54
3	P_{10-10}	0.63
4	DD	0.66
5	P_{dry}	0.67
6	MAP	0.69
	(*d*) SE=6	
1	TRIB	0.25
2	log July T	0.33
3	log $\bar{E}$	0.37
4	$\bar{E}$	0.45
5	log A	0.49

TABLE H-21

VARIABLES SELECTED FOR STRATIFICATION BASED ON GZ FOR LOG Q_{25} EQUATIONS (EXPERIMENT C-8)

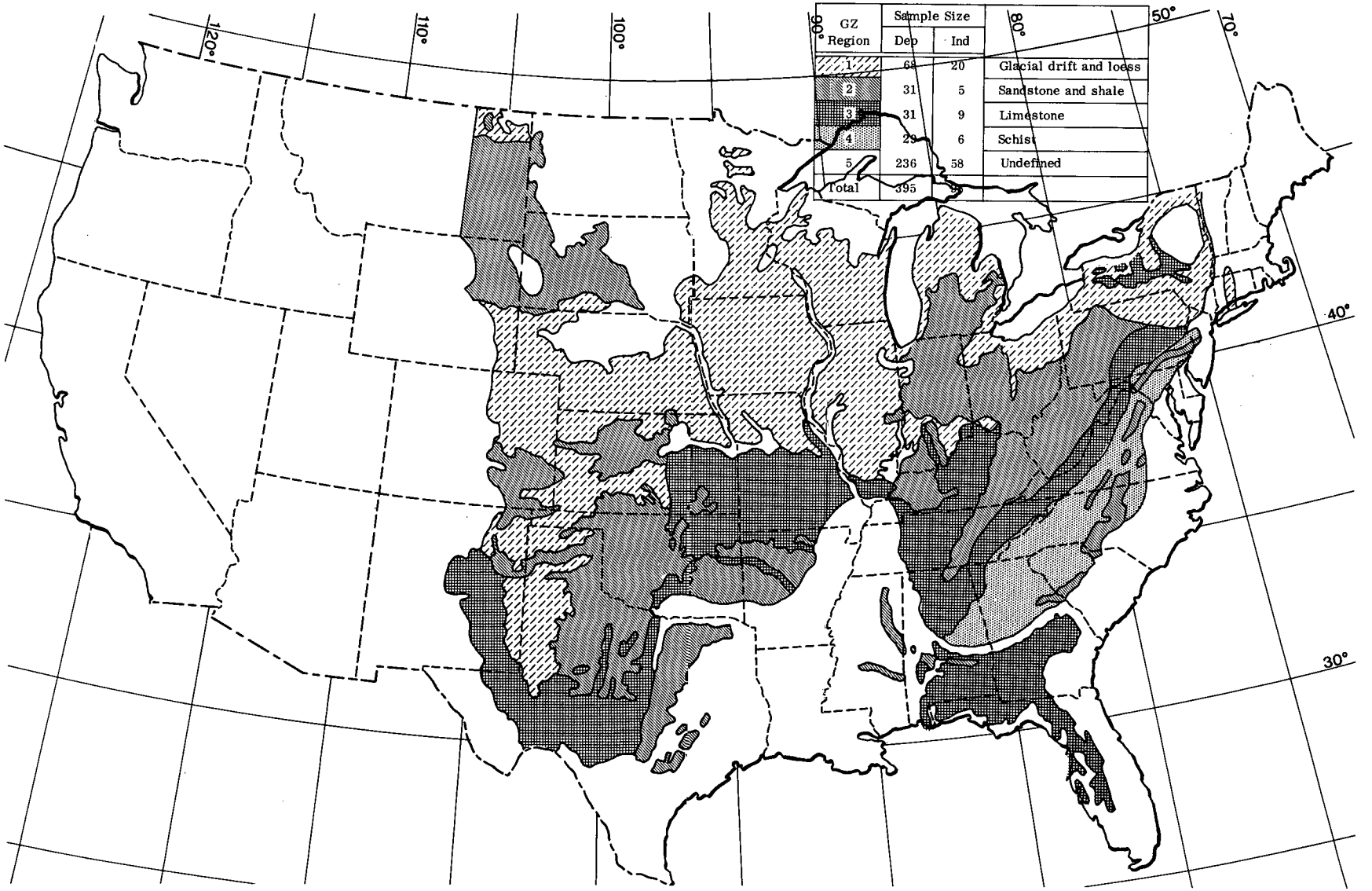
ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) GZ=1	
1	TRIB	0.24
2	32 F days	0.38
3	log P_{10-30}	0.46
4	log A	0.56
	(*b*) GZ=2	
1	log TRIB	0.49
	(*c*) GZ=3	
1	log T	0.41
2	log 32 F days	0.67
3	M24S	0.76
	(*d*) GZ=4	
1	log A	0.16
	(*e*) GZ undefined	
1	log TRIB	0.32
2	$\bar{E}$	0.54
3	Pot ET	0.54
4	log A	0.57
5	log $\bar{E}$	0.58
6	S_{10}	0.60
7	M24S	0.62
8	log MAS	0.63

Comparison of Alternative Methods of Stratification

The quantitative comparison of various methods of stratification was conducted with the independent data sample of 98 cases. Preliminary to this evaluation, each equation of each stratification method was applied to the cases of the dependent sample from which it was developed in order to compute the ratio $(\bar{Q}/\bar{Q}')$ needed to transform the logarithmic solutions of the independent sample to a usable form.

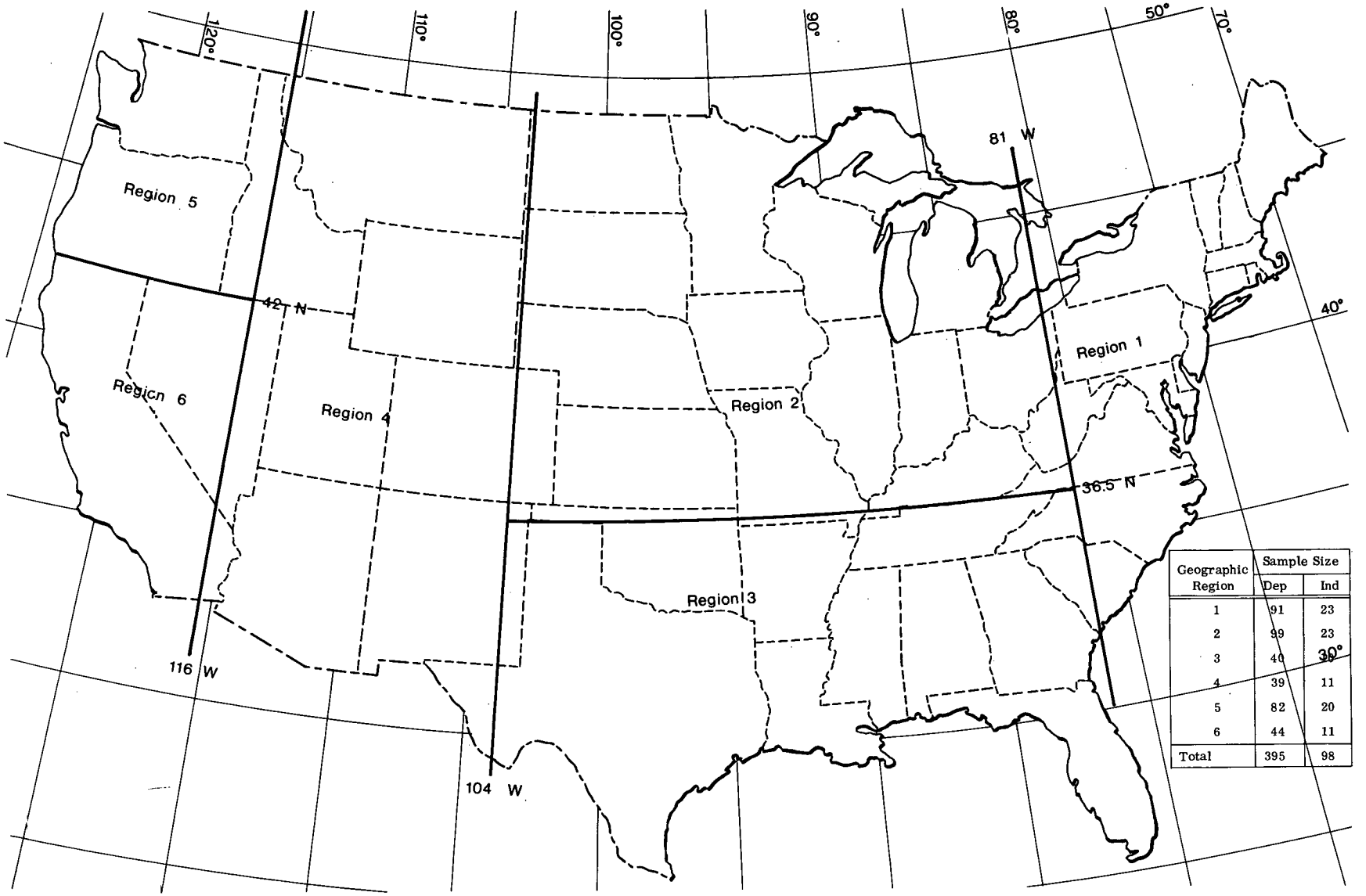
SE Region	Sample Size Dep	Sample Size Ind
1	57	16
2	23	3
3	201	54
4	114	25
Total	395	98

Figure H-6. Map of soil erosion class (SE) regions in the contiguous United States.



GZ Region	Sample Size Dep	Sample Size Ind	
1	68	20	Glacial drift and loess
2	31	5	Sandstone and shale
3	31	9	Limestone
4	29	6	Schist
5	236	58	Undefined
Total	395	98	

Figure H-7. Map of geological zone (GZ) regions in the contiguous United States.



Geographic Region	Sample Size	
	Dep	Ind
1	91	23
2	99	23
3	40	30
4	39	11
5	82	20
6	44	11
Total	395	98

Map of Geographic (REG)* regions in the contiguous United States.

*In using the equations, values for REG are the assigned binary values of 1 for the region of the subject drainage basin, and 0 for all other regions.

Example: Value for a drainage basin in Iowa (REG2) = 1
Value for all others (REG1, -3, -4, -5, -6) = 0

Figure H-8. Map of geographic (REG) regions in the contiguous United States.

TABLE H-22

VARIABLES SELECTED FOR STRATIFICATION BASED ON REG FOR LOG Q_{25} EQUATIONS (EXPERIMENT C-9)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) REG_1 (east of 81W)	
1	A	0.37
	(*b*) REG_2 (81 to 104W, north of 36.5N)	
1	TRIB	0.29
2	P_{10-360}	0.46
	(*c*) REG_3 (81 to 104W, south of 36.5N)	
1	A	0.42
2	32F days	0.57
	(*d*) REG_4 (104 to 116W)	
1	T	0.40
2	MAF_5	0.56
	(*e*) REG_5 (west of 116W, north of 42N)	
1	L	0.35
2	P_{10-360}	0.46
3	T	0.53
4	SHAPE	0.58
5	RH	0.62
6	July T	0.66
	(*f*) REG_6 (west of 116W, south of 42N)	
1	RH	0.35
2	TRIB	0.58

The root-mean-square-error (RMSE), expressed in cfs, and frequency distribution of errors (described in Appendix G) were used for this comparison. These statistics are given for Q_{25} in Table H-23. The negative percent error (PE) row headings are limited to errors of 100 percent or less (in a negative sense) because the logarithmic form, which yields only positive estimates, is being dealt with.

Based on the distribution of errors there is little difference between the methods of stratifying, with the exception that the mean basin elevation method has several more large errors than do the others. Comparing this against the RMSE statistics gives some insight into the magnitude of some of the larger errors. Stratifications P_{10-60}, SE, and GZ yielded some excessively large errors, which resulted in RMSE's considerably larger than those of the other methods.

Based on these results, three methods of stratifying (MAF_5, MAT, and $\bar{E}$) were retained for further evaluation against the national equations that evolved from the earlier evaluation. This comparison is summarized in Table H-24. From these statistics it can be seen that the stratifications MAF_5 (C-1) and MAT (C-4) have the least number of excessive errors (greater than 200 percent) and similarly the lowest RMSE. Stratification C-4 (based on MAT) also has the most estimates within an accuracy of 25 percent (30 cases out of 98).

MODIFIED NATIONAL EQUATIONS INCLUDING STRATIFICATION FACTORS—EXPERIMENTS D-1, D-2, D-3

Based on the results attained to this point, it appeared desirable to attempt to combine the useful information in some of the stratification factors into a modified national equation framework, with the intent of improving the accuracy of the estimation equations. Stratification factors that were introduced included the mean annual flood (MAF_5) in flood category index form and logarithmic form, and the geographic variable (REG). The (MAF_5) consisted of numerical representations of five flood categories and was considered as a possible predictor without modification. On the other hand, the geographical factor (REG) consists of an arbitrarily selected numerical representation of six sections of the country (e.g., REG_1 is East, REG_2 Midwest, etc.), thus it is descriptive of geographic location. It was necessary, therefore, to transform REG to

TABLE H-23

EVALUATION OF ALTERNATIVE METHODS OF STRATIFICATION ON INDEPENDENT DATA (Q_{25})

	C-1 (MAF_5)	C-2 (P_{10-60})	C-3 (MAP)	C-4 (MAT)	C-5 (A)	C-6 ($\bar{E}$)	C-7 (SE)	C-8 (GZ)	C-9 (REG)
RMSE (CFS)	2224	2843	2528	2116	2452	2275	3038	2962	2645
$-10 < PE \leq 10$	11	7	12	11	9	10	12	9	10
$10 < PE \leq 25$ $-25 < PE \leq -10$	14	7	11	19	13	12	12	18	12
$25 < PE \leq 50$ $-50 < PE \leq -25$	18	29	22	13	22	21	22	21	21
$50 < PE \leq 100$ $-100 < PE \leq -50$	30	23	18	23	24	24	27	23	32
$100 < PE \leq 150$	6	11	9	12	6	4	6	2	3
$150 < PE \leq 200$	3	5	7	5	4	4	2	6	5
$200 < PE$	16	16	19	15	20	23	17	19	15

TABLE H-24

COMPARISON OF NATIONAL AND STRATIFICATION EQUATIONS ON INDEPENDENT DATA (Q_{25})

	NATIONAL			STRATIFICATION		
	A-3	B-2	B-3	C-1 (MAF_5)	C-4 (MAT)	C-6 ($\bar{E}$)
RMSE (CFS)	2397	2342	2315	2224	2116	2275
$-10 < PE \leq 10$	7	6	12	11	11	10
$10 < PE \leq 25$ $-25 < PE \leq -10$	19	10	10	14	19	12
$25 < PE \leq 50$ $-50 < PE \leq -25$	17	22	15	18	13	21
$50 < PE \leq 100$ $-100 < PE \leq -50$	28	25	23	30	23	24
$100 < PE \leq 150$	6	11	11	6	12	4
$150 < PE \leq 200$	1	6	7	3	5	4
$200 < PE$	20	18	20	16	15	23

TABLE H-25

VARIABLES SELECTED FOR SEMI-LOGARITHMIC REGRESSION (EXPERIMENT D-1)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) log Q_5	
1	TRIB	0.26
2	P_{10-360}	0.38
3	MAF_5	0.45
4	A	0.51
5	REG_2	0.53
6	REG_1	0.55
	(*b*) log Q_{10}	
1	TRIB	0.27
2	P_{10-360}	0.40
3	MAF_5	0.45
4	A	0.50
5	REG_2	0.54
6	REG_1	0.55
	(*c*) log Q_{25}	
1	TRIB	0.26
2	P_{10-360}	0.40
3	MAF_5	0.44
4	A	0.48
5	REG_2	0.53
6	REG_1	0.55
	(*d*) log Q_{50}	
1	TRIB	0.26
2	P_{10-360}	0.40
3	REG_2	0.44
4	A	0.49
5	MAF_5	0.52
6	LAT	0.54
7	REG_3	0.56

dummy variable form—that is 0 or 1 (as described in Appendix F)—which in effect creates six new variables to replace the single geographical variable denoting six regions. Also, the length of tributary factor was modified by adding the main stream length to it in order to consolidate these two variables.

Three experiments (D-1, D-2, and D-3) were formulated for this phase of the study; D-1 in the functional form of experiment A-3, D-2 like B-3, and D-3 like B-2. As such, the predictand in each case was logarithmic. In D-1, the predictors are only in arithmetic and binary form (0-1); D-2 considers logarithmic, arithmetic, and binary; and D-3 only logarithmic and binary. Also, in experiment D-3 the modified variable log (TRIB + L) was used instead of log (TRIB). The dependent sample of 395 cases was used and new equations were generated for Q_5, Q_{10}, Q_{25}, and Q_{50}.

The variables selected are given in Tables H-25, H-26, and H-27. There are three points worth noting here. First, two stratification factors were selected by the screening procedure. Of the dummy variables depicting geographical regions, the Midwest (REG_2) and/or the Gulf (REG_3) appear in each relationship developed. Second, comparing the reduction of variance for these equations to comparable equations, experiments A-3, B-3, and B-2, respectively, one finds an improvement of 5 to 8 percentage points when stratifying factors are utilized. Third, the length of tributaries (TRIB, log TRIB or log (TRIB + L)) remains the single best estimator of peak runoff for all predictands on a national basis.

The national equations (modified with stratification factors) developed for Q_{10}, Q_{25}, and Q_{50} (see Appendix A) were applied to the developmental sample of 395 cases to determine the log-transform ratio ($\bar{Q}/\bar{Q}'$) for each. The equations were then applied to the independent sample of 98 cases. The estimates of peak runoff were evaluated by the RMSE, percent error distribution, and sign test against

TABLE H-26

VARIABLES SELECTED FOR LOGARITHMIC-LINEAR REGRESSION (EXPERIMENT D-2)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) log Q_5	
1	log TRIB	0.34
2	log P_{10-360}	0.50
3	MAF_5	0.57
4	log T	0.59
5	SHAPE	0.60
6	REG_3	0.61
7	log S	0.62
8	log T days	0.63
	(*b*) log Q_{10}	
1	log TRIB	0.34
2	log P_{10-360}	0.51
3	MAF_5	0.56
4	log T	0.59
5	REG_3	0.60
6	log July T	0.61
7	log A	0.62
	(*c*) log Q_{25}	
1	log TRIB	0.33
2	log P_{10-360}	0.50
3	MAF_5	0.54
4	REG_2	0.57
5	log A	0.59
6	log July T	0.60
7	REG_3	0.61
	(*d*) log Q_{50}	
1	log TRIB	0.32
2	log P_{10-360}	0.49
3	MAF_5	0.53
4	REG_2	0.56
5	log A	0.58
6	log MAP	0.59
7	REG_3	0.60
8	log Pot ET	0.61

TABLE H-27

VARIABLES SELECTED FOR LOGARITHMIC REGRESSION (EXPERIMENT D-3)

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(*a*) log Q_5	
1	log (TRIB+L)	0.33
2	log P_{10-360}	0.50
3	log MAF_5	0.57
4	log SHAPE	0.58
5	REG_2	0.59
	(*b*) log Q_{10}	
1	log (TRIB+L)	0.34
2	log P_{10-360}	0.51
3	log MAF_5	0.56
4	REG_2	0.58
5	log SHAPE	0.60
6	log July T	0.60
7	REG_3	0.61
	(*c*) log Q_{25}	
1	log (TRIB+L)	0.33
2	log P_{10-360}	0.50
3	log MAF_5	0.55
4	REG_2	0.57
5	log SHAPE	0.59
6	log July T	0.60
7	REG_3	0.61
	(*d*) log Q_{50}	
1	log (TRIB+L)	0.31
2	log P_{10-360}	0.49
3	log MAF_5	0.53
4	REG_2	0.56
5	log July T	0.57
6	REG_3	0.59
7	log SHAPE	0.60

the equations based on stratifications MAF_5 (C-1) and MAT (C-4). The error statistics (RMSE and distribution of errors) for the five methods are given in Table H-28.

The discussion in Appendix G indicates that the sign test represents a count of the number of times one forecast system yields errors smaller than another, and vice versa. The information given in Table H-29 summarizes these results. For example, for the 98 cases of the independent sample, the Q_{10} relationship for experiment D-1 had 60 smaller errors when compared to experiment C-4; whereas, conversely, there were just 38 cases wherein C-4 yielded smaller errors.

Summarizing both of these tables, the mean annual flood (MAF_5) method of stratifying yields the lowest RMSE for Q_{10}, but the national equation with regional factors (D-2) yields the lowest for Q_{25} and Q_{50}. There is little difference between the foregoing alternative methods on the basis of the error distribution. However, based on the sign test, which tallies the number of times method A is better than method B, the national equation with regional factors (D-3) is clearly the best relationship of the five. Between the two methods of regionalizing, C-1 (MAF_5) is slightly better than C-4 (MAT) on a sign test comparison.

SIMPLIFIED NATIONAL EQUATIONS—EXPERIMENT E-1, E-2

Among the variables included in the D-2 and D-3 relationships, two would require a certain amount of effort to compute or measure. These are the TRIB and SHAPE factors. In the last set of experiments conducted (described later), these variables were excluded and relationships were developed that were limited to easily obtained terms. The attempt in these experiments was to determine if the equations could be simplified without too much loss of prediction accuracy.

Two experiments were formulated; in each case the predictand and predictors were limited to logarithmic form, except for regional factors in binary form. The first experiment considered only area (A) and rainfall intensity (P_{10-360}), whereas the second considered A, L, and hydrologic-climatic and regional factors. National equations

TABLE H-28

COMPARISON OF MODIFIED NATIONAL AND SELECTED STRATIFICATION METHODS ON INDEPENDENT DATA[a] (Q_{10}, Q_{25}, Q_{50})

	Q_{10}					Q_{25}					Q_{50}				
	D-1	D-2	D-3	C-1	C-4	D-1	D-2	D-3	C-1	C-4	D-1	D-2	D-3	C-1	C-4
RMSE (CFS)	1233	969	1010	944	1195	2522	2093	2158	2224	2116	4169	3572	3668	3610	3660
$-10 < PE \leq 10$	16	13	16	15	4	7	9	10	11	11	13	7	9	13	7
$10 < PE \leq 25$ $-25 < PE \leq -10$	11	12	8	15	17	20	14	12	14	19	6	13	12	11	18
$25 < PE \leq 50$ $-50 < PE \leq -25$	17	14	17	16	22	16	16	21	18	13	22	21	20	15	11
$50 < PE \leq 100$ $-100 < PE \leq -50$	30	35	36	28	24	30	30	29	30	23	28	21	28	31	27
$100 < PE \leq 150$	6	8	8	6	9	5	12	10	6	12	9	7	8	8	14
$150 < PE \leq 200$	8	1	1	4	5	6	2	4	3	5	7	14	6	7	4
$200 < PE$	10	15	12	14	17	14	15	12	16	15	13	15	15	13	17

[a] D-1 = predictors in arithmetic and binary form; D-2 = predictors in arithmetic, logarithmic, and binary form; D-3 = predictors in logarithmic and binary form; C-1 = stratification based on MAF_5; C-4 = stratification based on MAT.

were generated for four predictands ($\hat{Q}_5$, $\hat{Q}_{10}$, $\hat{Q}_{25}$, and $\hat{Q}_{50}$) from the dependent sample of 395 cases.

The variables selected are given in Tables H-30 and H-31. The difference in the percent reduction of variance between the two predictor equations and the second set of equations that include regional factors is of the order of 10 to 12 percentage points. Also, in experiment E-2 the regional and other hydrologic-climatic factors are selected after the variables A and P_{10-360} in each case. Finally, note that L was not selected for these equations.

The Q_{25} equations for experiments E-1 and E-2 were applied to the dependent sample of 395 cases to obtain the log-transform constants ($\bar{Q}/\bar{Q}'$), and then reapplied to the 98 independent cases and evaluated against the best national (D-3) and best set of stratified equations (C-1). These results are summarized in Table H-32, which includes the RMSE, error distribution, and in Table H-33, which summarizes the sign test comparison.

Comparison of the results for experiment E-1 (the two-predictor equation) with the others, shows appreciably poorer scores. The RMSE is 400 to 500 cfs higher, there are several more large errors (in excess of 200 percent), and it performs poorly in a sign test comparison. There is only a slight deterioration in the results for the equation of the other simplified experiment (E-2), in which watershed area (A) was substituted for more complex topographic variables (TRIB, SHAPE, etc.) and regional factors (MAF_5, REG_2, etc.) were included. The positive contribution of regional factors is strongly suggested in this comparison because the only other climatic or topographic term in the equation was log July T.

TABLE H-29

SIGN TEST COMPARISON OF MODIFIED NATIONAL AND SELECTED STRATIFICATION METHODS ON INDEPENDENT DATA[a] (Q_{10}, Q_{25}, Q_{50})

METHOD	D-1	D-2	D-3	C-1	C-4	TOTAL
			(a) Q_{10}			
D-1	—	48[b]	47	47	60	202
D-2	50	—	41	54	56	201
D-3	51	57	—	55	62	225
C-1	51	44	43	—	54	192
C-4	38	42	36	44	—	160
			(b) Q_{25}			
D-1	—	56	45	51	50	100
D-2	48	—	45	48	52	192
D-3	53	53	—	51	59	216
C-1	47	51	47	—	51	196
C-4	48	46	39	47	—	180
			(c) Q_{50}			
D-1	—	48	42	43	50	183
D-2	50	—	50	49	56	205
D-3	56	48	—	54	58	216
C-1	55	49	44	—	51	199
C-4	48	42	40	47	—	177

[a] See Table H-28 for experiment descriptions.
[b] D-1 had 48 prediction errors smaller than D-2 out of a total of 98 cases.

SENSITIVITY OF REGRESSION EQUATION

One problem of concern to the design engineer is the sensitivity of a prediction technique (equation) to the accuracy of the independent variables (predictors) included in it. An evaluation of the Q_{25} prediction equation (from experiment D-3) was conducted to investigate this potential problem. Initially,

$$\hat{Q}_{25} = 3.69 \times 10^{-4}(\text{TRIB} + L)^{0.77}(P_{10-360})^{1.18}(\text{MAF}_5)^{0.79}(\text{SHAPE})^{-0.61}(\text{July } T)^{2.63} \quad \text{(H-18)}$$

was solved using the mean value of each variable obtained from the dependent sample and a value of 3 assigned to the mean annual flood term (MAF_5). The means for the four variables were: $\overline{TRIB + L} = 19.3$ mi, $\overline{P_{10-360}} = 3.3$ in., $\overline{SHAPE} = 1.6$, and $\overline{July\ T} = 71.6$ F. This yields a solution of 1,978.1 cfs for Q_{25}. Eq. H-18 was then solved by varying the value of each predictor, one at a time, by a defined percentage that represented a reasonable high error of determination on the part of the design engineer. A defined error of 20.0 percent in TRIB + L from its mean (or a value of 15.4 mi) and assuming no change in the other variables yields a solution of 1,661.7 cfs for Q_{25}. Similarly, an error of 5 percent in P_{10-360} (or a value of 3.1 in.) yields a solution of 1,837.5 cfs, a 10 percent error in SHAPE (a value of 1.45) yields a solution of 2,100.5 cfs, and a 5 percent error in July T (a value of 68.0 F) yields a solution of 1,727.4 cfs. Table H-34 gives the percent change in Q_{25} resulting from the defined percent change in the independent variables. The percent change (PC) in the Q_{25} prediction was determined from

$$PC = \frac{Q_{25} - Q_{25}'}{Q_{25}} \times 100. \qquad \text{(H-19)}$$

in which Q_{25} is the "correct" forecast of 1,978.1 cfs and Q_{25}' is the solution after altering the independent variable. Based on this illustration, the most sensitive term in the

TABLE H-30

VARIABLES SELECTED FOR EXPERIMENT E-1 [a]

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(a) log Q_5	
1	log A	0.28
2	log P_{10-360}	0.46
	(b) log Q_{10}	
1	log A	0.28
2	log P_{10-360}	0.47
	(c) log Q_{25}	
1	log A	0.26
2	log P_{10-360}	0.45
	(d) log Q_{50}	
1	log A	0.24
2	log P_{10-360}	0.44

[a] Only the predictors log A and log P_{10-360} were considered.

TABLE H-31

VARIABLES SELECTED FOR EXPERIMENT E-2

ORDER OF SELECTION	VARIABLE SYMBOL	REDUCTION OF VARIANCE, R^2
	(a) log Q_5	
1	log A	0.28
2	log P_{10-360}	0.46
3	log MAF_5	0.55
4	REG_2	0.57
	(b) log Q_{10}	
1	log A	0.28
2	log P_{10-360}	0.47
3	log MAF_5	0.53
4	REG_2	0.57
5	log July T	0.58
6	REG_3	0.59
	(c) log Q_{25}	
1	log A	0.26
2	log P_{10-360}	0.45
3	log MAF_5	0.51
4	REG_2	0.56
5	log July T	0.57
6	REG_3	0.58
	(d) log Q_{50}	
1	log A	0.24
2	log P_{10-360}	0.44
3	REG_2	0.50
4	log MAF_5	0.54
5	log July T	0.56
6	REG_3	0.57

TABLE H-32

ERROR COMPARISON OF SIMPLIFIED EQUATIONS WITH THE BEST NATIONAL AND STRATIFICATION EQUATIONS ON INDEPENDENT DATA (Q_{25})

	E-1	E-2	D-3	C-1
RMSE (CFS)	2638	2231	2158	2224
$-10 < PE \leq 10$	8	7	10	11
$10 < PE \leq 25$ $-25 < PE \leq -10$	12	8	12	14
$25 < PE \leq 50$ $-50 < PE \leq -25$	18	26	21	18
$50 < PE \leq 100$ $-100 < PE \leq -50$	22	26	29	30
$100 < PE \leq 150$	11	11	10	6
$150 < PE \leq 200$	6	3	4	3
$200 < PE$	21	17	12	16

TABLE H-33

SIGN TEST COMPARISON OF SIMPLIFIED EQUATIONS WITH THE BEST NATIONAL AND STRATIFICATION EQUATIONS ON INDEPENDENT DATA (Q_{25}) [a]

EXPER.	E-1	E-2	D-3	C-1	TOTAL
E-1	—	44	38	40	122
E-2	54	—	50	47	151
D-3	60	48	—	51	159
C-1	58	51	47	—	156

[a] E-1 = only used log A and log P_{10-360} as predictors; E-2 = simplified equation with stratification factors; D-3 = modified equation with stratification factors; C-1 = stratification based on MAF_5.

equation is obviously the July T, because the error in it is compounded more than 150 percent in the Q_{25} prediction. However, the likelihood of a measurable error in this term is dependent on the care and accuracy with which one locates a geographical location on a map and interpolates a value of July T from the map.

TABLE H-34

PERCENT CHANGE (PC) IN Q_{25} DUE TO CHANGES DEFINED IN INDEPENDENT VARIABLES

INDEPENDENT VARIABLE		Q_{25}	
SYMBOL	PC	SOLUTION	PC
TRIB+L	20.0	1661.7	16.0
P_{10-360}	5.0	1837.5	7.1
SHAPE	10.0	2100.5	6.2
July T	5.0	1727.4	12.7

APPENDIX I

LIST OF PRINCIPAL SYMBOLS

SYMBOL	DEFINITION	UNITS
A	Watershed area	sq mi
C	Contour interval	ft
CC	Composite land cover parameter	—[a]
cfs	Cubic feet per second	$ft^3\ sec^{-1}$
csm	cfs per square mile	cfs mi^{-2}
CS	Composite soil parameter	—[a]
DD	Drainage density	mi^{-1}
E	Gage elevation	ft
$\bar{E}$	Mean basin elevation	ft
GZ	Geologic zone	—[a]
H	Total fall along main stream	ft
July T	Mean July temperature	°F
L	Main stream length	mi
L_{cm}	Length of contour lines	mi
log	Logarithm to base 10	—[a]
MAF	Mean annual flood category	cfs mi^{-2}
MAP	Mean annual precipitation (map)	in.
MAP-S	Mean annual precipitation (station)	in.
MAS	Mean annual snowfall	in.
MAT	Mean annual temperature	°F
M24P	Maximum 24-hr precipitation	in.
M24S	Maximum 24-hr snowfall	in.
P days	Mean number of 0.01-in. or more precipitation days per year	days
P_{dry}	Mean driest monthly precipitation	in.
P_{f-d}	Frequency (f) and duration (d) precipitation	in.
Pot ET	Potential evapotranspiration	in.
P_{wet}	Mean wettest monthly precipitation	in.
Q_r	Peak runoff for return period p as estimated by log-normal	cfs
$\hat{Q}_P$	Peak runoff for return period p as estimated by regression equation	cfs
R	Watershed slope	ft mi^{-1}
REG	Defined geographical region	—[a]
RH	Mean relative humidity	%
RI	Mean annual flood runoff index	cfs
RMSE	Root-mean-square-error	—[a]
S	Stream slope	ft mi^{-1}
SA	Percentage of soil type A	—[a]
SB	Percentage of soil type B	—[a]
SC	Percentage of soil type C	—[a]
SD	Percentage of soil type D	—[a]
S days	Mean number of 1-in. snow cover days per year	days
SE	Soil erosion class	—[a]
SHAPE	Watershed shape	—[a]
SO	Soil order class	—[a]
SSO	Soil suborder class	—[a]
T	Travel time index	mi
T days	Mean number of thunderstorm days per year	days
TRIB	Length of tributaries	mi
TRIB + L	Total length of streams	mi
σ_r	Residual standard deviation	—[a]

[a] Dimensionless.

Published reports of the

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

are available from:

Highway Research Board
National Academy of Sciences
2101 Constitution Avenue
Washington, D.C. 20418

Rep. No. *Title*

—* A Critical Review of Literature Treating Methods of Identifying Aggregates Subject to Destructive Volume Change When Frozen in Concrete and a Proposed Program of Research—Intermediate Report (Proj. 4-3(2)), 81 p., $1.80

1 Evaluation of Methods of Replacement of Deteriorated Concrete in Structures (Proj. 6-8), 56 p., $2.80

2 An Introduction to Guidelines for Satellite Studies of Pavement Performance (Proj. 1-1), 19 p., $1.80

2A Guidelines for Satellite Studies of Pavement Performance, 85 p.+9 figs., 26 tables, 4 app., $3.00

3 Improved Criteria for Traffic Signals at Individual Intersections—Interim Report (Proj. 3-5), 36 p., $1.60

4 Non-Chemical Methods of Snow and Ice Control on Highway Structures (Proj. 6-2), 74 p., $3.20

5 Effects of Different Methods of Stockpiling Aggregates—Interim Report (Proj. 10-3), 48 p., $2.00

6 Means of Locating and Communicating with Disabled Vehicles—Interim Report (Proj. 3-4), 56 p. $3.20

7 Comparison of Different Methods of Measuring Pavement Condition—Interim Report (Proj. 1-2), 29 p., $1.80

8 Synthetic Aggregates for Highway Construction (Proj. 4-4), 13 p., $1.00

9 Traffic Surveillance and Means of Communicating with Drivers—Interim Report (Proj. 3-2), 28 p., $1.60

10 Theoretical Analysis of Structural Behavior of Road Test Flexible Pavements (Proj. 1-4), 31 p., $2.80

11 Effect of Control Devices on Traffic Operations—Interim Report (Proj. 3-6), 107 p., $5.80

12 Identification of Aggregates Causing Poor Concrete Performance When Frozen—Interim Report (Proj. 4-3(1)), 47 p., $3.00

13 Running Cost of Motor Vehicles as Affected by Highway Design—Interim Report (Proj. 2-5), 43 p., $2.80

14 Density and Moisture Content Measurements by Nuclear Methods—Interim Report (Proj. 10-5), 32 p., $3.00

15 Identification of Concrete Aggregates Exhibiting Frost Susceptibility—Interim Report (Proj. 4-3(2)), 66 p., $4.00

16 Protective Coatings to Prevent Deterioration of Concrete by Deicing Chemicals (Proj. 6-3), 21 p., $1.60

17 Development of Guidelines for Practical and Realistic Construction Specifications (Proj. 10-1), 109 p., $6.00

18 Community Consequences of Highway Improvement (Proj. 2-2), 37 p., $2.80

19 Economical and Effective Deicing Agents for Use on Highway Structures (Proj. 6-1), 19 p., $1.20

* Highway Research Board Special Report 80.

Rep. No. *Title*

20 Economic Study of Roadway Lighting (Proj. 5-4), 77 p., $3.20

21 Detecting Variations in Load-Carrying Capacity of Flexible Pavements (Proj. 1-5), 30 p., $1.40

22 Factors Influencing Flexible Pavement Performance (Proj. 1-3(2)), 69 p., $2.60

23 Methods for Reducing Corrosion of Reinforcing Steel (Proj. 6-4), 22 p., $1.40

24 Urban Travel Patterns for Airports, Shopping Centers, and Industrial Plants (Proj. 7-1), 116 p., $5.20

25 Potential Uses of Sonic and Ultrasonic Devices in Highway Construction (Proj. 10-7), 48 p., $2.00

26 Development of Uniform Procedures for Establishing Construction Equipment Rental Rates (Proj. 13-1), 33 p., $1.60

27 Physical Factors Influencing Resistance of Concrete to Deicing Agents (Proj. 6-5), 41 p., $2.00

28 Surveillance Methods and Ways and Means of Communicating with Drivers (Proj. 3-2), 66 p., $2.60

29 Digital-Computer-Controlled Traffic Signal System for a Small City (Proj. 3-2), 82 p., $4.00

30 Extension of AASHO Road Test Performance Concepts (Proj. 1-4(2)), 33 p., $1.60

31 A Review of Transportation Aspects of Land-Use Control (Proj. 8-5), 41 p., $2.00

32 Improved Criteria for Traffic Signals at Individual Intersections (Proj. 3-5), 134 p., $5.00

33 Values of Time Savings of Commercial Vehicles (Proj. 2-4), 74 p., $3.60

34 Evaluation of Construction Control Procedures—Interim Report (Proj. 10-2), 117 p., $5.00

35 Prediction of Flexible Pavement Deflections from Laboratory Repeated-Load Tests (Proj. 1-3(3)), 117 p., $5.00

36 Highway Guardrails—A Review of Current Practice (Proj. 15-1), 33 p., $1.60

37 Tentative Skid-Resistance Requirements for Main Rural Highways (Proj. 1-7), 80 p., $3.60

38 Evaluation of Pavement Joint and Crack Sealing Materials and Practices (Proj. 9-3), 40 p., $2.00

39 Factors Involved in the Design of Asphaltic Pavement Surfaces (Proj. 1-8), 112 p., $5.00

40 Means of Locating Disabled or Stopped Vehicles (Proj. 3-4(1)), 40 p., $2.00

41 Effect of Control Devices on Traffic Operations (Proj. 3-6), 83 p., $3.60

42 Interstate Highway Maintenance Requirements and Unit Maintenance Expenditure Index (Proj. 14-1), 144 p., $5.60

43 Density and Moisture Content Measurements by Nuclear Methods (Proj. 10-5), 38 p., $2.00

44 Traffic Attraction of Rural Outdoor Recreational Areas (Proj. 7-2), 28 p., $1.40

45 Development of Improved Pavement Marking Materials—Laboratory Phase (Proj. 5-5), 24 p., $1.40

46 Effects of Different Methods of Stockpiling and Handling Aggregates (Proj. 10-3), 102 p., $4.60

47 Accident Rates as Related to Design Elements of Rural Highways (Proj. 2-3), 173 p., $6.40

48 Factors and Trends in Trip Lengths (Proj. 7-4), 70 p., $3.20

49 National Survey of Transportation Attitudes and Behavior—Phase I Summary Report (Proj. 20-4), 71 p., $3.20

Rep. No.	Title
50	Factors Influencing Safety at Highway-Rail Grade Crossings (Proj. 3-8), 113 p., $5.20
51	Sensing and Communication Between Vehicles (Proj. 3-3), 105 p., $5.00
52	Measurement of Pavement Thickness by Rapid and Nondestructive Methods (Proj. 10-6), 82 p., $3.80
53	Multiple Use of Lands Within Highway Rights-of-Way (Proj. 7-6), 68 p., $3.20
54	Location, Selection, and Maintenance of Highway Guardrails and Median Barriers (Proj. 15-1(2)), 63 p., $2.60
55	Research Needs in Highway Transportation (Proj. 20-2), 66 p., $2.80
56	Scenic Easements—Legal, Administrative, and Valuation Problems and Procedures (Proj. 11-3), 174 p., $6.40
57	Factors Influencing Modal Trip Assignment (Proj. 8-2), 78 p., $3.20
58	Comparative Analysis of Traffic Assignment Techniques with Actual Highway Use (Proj. 7-5), 85 p., $3.60
59	Standard Measurements for Satellite Road Test Program (Proj. 1-6), 78 p., $3.20
60	Effects of Illumination on Operating Characteristics of Freeways (Proj. 5-2) 148 p., $6.00
61	Evaluation of Studded Tires—Performance Data and Pavement Wear Measurement (Proj. 1-9), 66 p., $3.00
62	Urban Travel Patterns for Hospitals, Universities, Office Buildings, and Capitols (Proj. 7-1), 144 p., $5.60
63	Economics of Design Standards for Low-Volume Rural Roads (Proj. 2-6), 93 p., $4.00
64	Motorists' Needs and Services on Interstate Highways (Proj. 7-7), 88 p., $3.60
65	One-Cycle Slow-Freeze Test for Evaluating Aggregate Performance in Frozen Concrete (Proj. 4-3(1)), 21 p., $1.40
66	Identification of Frost-Susceptible Particles in Concrete Aggregates (Proj. 4-3(2)), 62 p., $2.80
67	Relation of Asphalt Rheological Properties to Pavement Durability (Proj. 9-1), 45 p., $2.20
68	Application of Vehicle Operating Characteristics to Geometric Design and Traffic Operations (Proj. 3-10), 38 p., $2.00
69	Evaluation of Construction Control Procedures—Aggregate Gradation Variations and Effects (Proj. 10-2A), 58 p., $2.80
70	Social and Economic Factors Affecting Intercity Travel (Proj. 8-1), 68 p., $3.00
71	Analytical Study of Weighing Methods for Highway Vehicles in Motion (Proj. 7-3), 63 p., $2.80
72	Theory and Practice in Inverse Condemnation for Five Representative States (Proj. 11-2), 44 p., $2.20
73	Improved Criteria for Traffic Signal Systems on Urban Arterials (Proj. 3-5/1), 55 p., $2.80
74	Protective Coatings for Highway Structural Steel (Proj. 4-6), 64 p., $2.80
74A	Protective Coatings for Highway Structural Steel—Literature Survey (Proj. 4-6), 275 p., $8.00
74B	Protective Coatings for Highway Structural Steel—Current Highway Practices (Proj. 4-6), 102 p., $4.00
75	Effect of Highway Landscape Development on Nearby Property (Proj. 2-9), 82 p., $3.60
76	Detecting Seasonal Changes in Load-Carrying Capabilities of Flexible Pavements (Proj. 1-5(2)), 37 p., $2.00
77	Development of Design Criteria for Safer Luminaire Supports (Proj. 15-6), 82 p., $3.80
78	Highway Noise—Measurement, Simulation, and Mixed Reactions (Proj. 3-7), 78 p., $3.20
79	Development of Improved Methods for Reduction of Traffic Accidents (Proj. 17-1), 163 p., $6.40
80	Oversize-Overweight Permit Operation on State Highways (Proj. 2-10), 120 p., $5.20
81	Moving Behavior and Residential Choice—A National Survey (Proj. 8-6), 129 p., $5.60
82	National Survey of Transportation Attitudes and Behavior—Phase II Analysis Report (Proj. 20-4), 89 p., $4.00
83	Distribution of Wheel Loads on Highway Bridges (Proj. 12-2), 56 p., $2.80
84	Analysis and Projection of Research on Traffic Surveillance, Communication, and Control (Proj. 3-9), 48 p., $2.40
85	Development of Formed-in-Place Wet Reflective Markers (Proj. 5-5), 28 p., $1.80
86	Tentative Service Requirements for Bridge Rail Systems (Proj. 12-8), 62 p., $3.20
87	Rules of Discovery and Disclosure in Highway Condemnation Proceedings (Proj. 11-1(5)), 28 p., $2.00
88	Recognition of Benefits to Remainder Property in Highway Valuation Cases (Proj. 11-1(2)), 24 p., $2.00
89	Factors, Trends, and Guidelines Related to Trip Length (Proj. 7-4), 59 p., $3.20
90	Protection of Steel in Prestressed Concrete Bridges (Proj. 12-5), 86 p., $4.00
91	Effects of Deicing Salts on Water Quality and Biota—Literature Review and Recommended Research (Proj. 16-1), 70 p., $3.20
92	Valuation and Condemnation of Special Purpose Properties (Proj. 11-1(6)), 47 p., $2.60
93	Guidelines for Medial and Marginal Access Control on Major Roadways (Proj. 3-13), 147 p., $6.20
94	Valuation and Condemnation Problems Involving Trade Fixtures (Proj. 11-1(9)), 22 p., $1.80
95	Highway Fog (Proj. 5-6), 48 p., $2.40
96	Strategies for the Evaluation of Alternative Transportation Plans (Proj. 8-4), 111 p., $5.40
97	Analysis of Structural Behavior of AASHO Road Test Rigid Pavements (Proj. 1-4(1)A), 35 p., $2.60
98	Tests for Evaluating Degradation of Base Course Aggregates (Proj. 4-2), 98 p. $5.00
99	Visual Requirements in Night Driving (Proj. 5-3), 38 p., $2.60
100	Research Needs Relating to Performance of Aggregates in Highway Construction (Proj. 4-8), 68 p., $3.40
101	Effect of Stress on Freeze-Thaw Durability of Concrete Bridge Decks (Proj. 6-9), 70 p., $3.60
102	Effect of Weldments on the Fatigue Strength of Steel Beams (Proj. 12-7), 114 p., $5.40
103	Rapid Test Methods for Field Control of Highway Construction (Proj. 10-4), 89 p., $5.00
104	Rules of Compensability and Valuation Evidence for Highway Land Acquisition (Proj. 11-1), 77 p., $4.40

Rep. No. Title

105 Dynamic Pavement Loads of Heavy Highway Vehicles (Proj. 15-5), 94 p., $5.00
106 Revibration of Retarded Concrete for Continuous Bridge Decks (Proj. 18-1), 67 p., $3.40
107 New Approaches to Compensation for Residential Takings (Proj. 11-1(10)), 27 p., $2.40
108 Tentative Design Procedure for Riprap-Lined Channels (Proj. 15-2), 75 p., $4.00
109 Elastomeric Bearing Research (Proj. 12-9), 53 p., $3.00
110 Optimizing Street Operations Through Traffic Regulations and Control (Proj. 3-11), 100 p., $4.40
111 Running Costs of Motor Vehicles as Affected by Road Design and Traffic (Proj. 2-5a and 2-7), 97 p., $5.20
112 Junkyard Valuation—Salvage Industry Appraisal Principles Applicable to Highway Beautification (Proj. 11-3(2)), 41 p., $2.60
113 Optimizing Flow on Existing Street Networks (Proj. 3-14), 414 p., $15.60
114 Effects of Proposed Highway Improvements on Property Values (Proj. 11-1(1)), 42 p., $2.60
115 Guardrail Performance and Design (Proj. 15-1(2)), 70 p., $3.60
116 Structural Analysis and Design of Pipe Culverts (Proj. 15-3), 155 p., $6.40
117 Highway Noise—A Design Guide for Highway Engineers (Proj. 3-7), 79 p., $4.60
118 Location, Selection, and Maintenance of Highway Traffic Barriers (Proj. 15-1(2)), 96 p., $5.20
119 Control of Highway Advertising Signs—Some Legal Problems (Proj. 11-3(1)), 72 p., $3.60
120 Data Requirements for Metropolitan Transportation Planning (Proj. 8-7), 90 p., $4.80
121 Protection of Highway Utility (Proj. 8-5), 115 p., $5.60
122 Summary and Evaluation of Economic Consequences of Highway Improvements (Proj. 2-11), 324 p., $13.60
123 Development of Information Requirements and Transmission Techniques for Highway Users (Proj. 3-12) 239 p., $9.60
124 Improved Criteria for Traffic Signal Systems in Urban Networks (Proj. 3-5) 86 p., $4.80
125 Optimization of Density and Moisture Content Measurements by Nuclear Methods (Proj. 10-5A), 86 p., $4.40
126 Divergencies in Right-of-Way Valuation (Proj. 11-4), 57 p., $3.00
127 Snow Removal and Ice Control Techniques at Interchanges (Proj. 6-10), 90 p., $5.20
128 Evaluation of AASHO Interim Guides for Design of Pavement Structures (Proj. 1-11), 111 p., $5.60
129 Guardrail Crash Test Evaluation—New Concepts and End Designs (Proj. 15-1(2)), 89 p., $4.80
130 Roadway Delineation Systems (Proj. 5-7), 349 p., $14.00
131 Performance Budgeting System for Highway Maintenance Management (Proj. 19-2(4)), 213 p., $8.40
132 Relationships Between Physiographic Units and Highway Design Factors (Proj. 1-3(1)), 161 p., $7.20

Rep. No. Title

133 Procedures for Estimating Highway User Costs, Air Pollution, and Noise Effects (Proj. 7-8), 127 p., $5.60
134 Damages Due to Drainage, Runoff, Blasting, and Slides (Proj. 11-1(8)), 24 p., $2.80
135 Promising Replacements for Conventional Aggregates for Highway Use (Proj. 4-10), 53 p., $3.60
136 Estimating Peak Runoff Rates from Ungaged Small Rural Watersheds (Proj. 15-4), 85 p., $4.60

Synthesis of Highway Practice

No. Title

1 Traffic Control for Freeway Maintenance (Proj. 20-5, Topic 1), 47 p., $2.20
2 Bridge Approach Design and Construction Practices (Proj. 20-5, Topic 2), 30 p., $2.00
3 Traffic-Safe and Hydraulically Efficient Drainage Practice (Proj. 20-5, Topic 4), 38 p., $2.20
4 Concrete Bridge Deck Durability (Proj. 20-5, Topic 3), 28 p., $2.20
5 Scour at Bridge Waterways (Proj. 20-5, Topic 5), 37 p., $2.40
6 Principles of Project Scheduling and Monitoring (Proj. 20-5, Topic 6), 43 p., $2.40
7 Motorist Aid Systems (Proj. 20-5, Topic 3-01), 28 p., $2.40
8 Construction of Embankments (Proj. 20-5, Topic 9), 38 p., $2.40
9 Pavement Rehabilitation—Materials and Techniques (Proj. 20-5, Topic 8), 41 p., $2.80
10 Recruiting, Training, and Retaining Maintenance and Equipment Personnel (Proj. 20-5, Topic 10), 35 p., $2.80
11 Development of Management Capability (Proj. 20-5, Topic 12), 50 p., $3.20
12 Telecommunications Systems for Highway Administration and Operations (Proj. 20-5, Topic 3-03), 29 p., $2.80
13 Radio Spectrum Frequency Management (Proj. 20-5, Topic 3-03), 32 p., $2.80
14 Skid Resistance (Proj. 20-5, Topic 7), 66 p., $4.00
15 Statewide Transportation Planning—Needs and Requirements (Proj. 20-5, Topic 3-02), 41 p., $3.60

THE NATIONAL ACADEMY OF SCIENCES is a private, honorary organization of more than 700 scientists and engineers elected on the basis of outstanding contributions to knowledge. Established by a Congressional Act of Incorporation signed by President Abraham Lincoln on March 3, 1863, and supported by private and public funds, the Academy works to further science and its use for the general welfare by bringing together the most qualified individuals to deal with scientific and technological problems of broad significance.

Under the terms of its Congressional charter, the Academy is also called upon to act as an official—yet independent—adviser to the Federal Government in any matter of science and technology. This provision accounts for the close ties that have always existed between the Academy and the Government, although the Academy is not a governmental agency and its activities are not limited to those on behalf of the Government.

THE NATIONAL ACADEMY OF ENGINEERING was established on December 5, 1964. On that date the Council of the National Academy of Sciences, under the authority of its Act of Incorporation, adopted Articles of Organization bringing the National Academy of Engineering into being, independent and autonomous in its organization and the election of its members, and closely coordinated with the National Academy of Sciences in its advisory activities. The two Academies join in the furtherance of science and engineering and share the responsibility of advising the Federal Government, upon request, on any subject of science or technology.

THE NATIONAL RESEARCH COUNCIL was organized as an agency of the National Academy of Sciences in 1916, at the request of President Wilson, to enable the broad community of U. S. scientists and engineers to associate their efforts with the limited membership of the Academy in service to science and the nation. Its members, who receive their appointments from the President of the National Academy of Sciences, are drawn from academic, industrial and government organizations throughout the country. The National Research Council serves both Academies in the discharge of their responsibilities.

Supported by private and public contributions, grants, and contracts, and voluntary contributions of time and effort by several thousand of the nation's leading scientists and engineers, the Academies and their Research Council thus work to serve the national interest, to foster the sound development of science and engineering, and to promote their effective application for the benefit of society.

THE DIVISION OF ENGINEERING is one of the eight major Divisions into which the National Research Council is organized for the conduct of its work. Its membership includes representatives of the nation's leading technical societies as well as a number of members-at-large. Its Chairman is appointed by the Council of the Academy of Sciences upon nomination by the Council of the Academy of Engineering.

THE HIGHWAY RESEARCH BOARD, organized November 11, 1920, as an agency of the Division of Engineering, is a cooperative organization of the highway technologists of America operating under the auspices of the National Research Council and with the support of the several highway departments, the Federal Highway Administration, and many other organizations interested in the development of transportation. The purpose of the Board is to advance knowledge concerning the nature and performance of transportation systems, through the stimulation of research and dissemination of information derived therefrom.

HIGHWAY RESEARCH BOARD
NATIONAL ACADEMY OF SCIENCES—NATIONAL RESEARCH COUNCIL
2101 Constitution Avenue Washington, D. C. 20418

ADDRESS CORRECTION REQUESTED

NON-PROFIT ORG.
U.S. POSTAGE
PAID
WASHINGTON, D.C.
PERMIT NO. 42970

000015
MATERIALS ENGR

IDAHO DEPT OF HIGHWAYS
P O BOX 7129
BOISE ID 83707