MODELLING THE DYNAMIC RESPONSE OF FLOODPLAINS TO URBANIZATION IN EASTERN NEW ENGLAND

by

Donald O. Doehring
Mark E. Smith

January 1978

Completion Report No. 83

MODELLING THE DYNAMIC RESPONSE OF FLOODPLAINS TO URBANIZATION IN EASTERN NEW ENGLAND

Completion Report

OWRT Project No. B-147-COLO

by

Donald O. Doehring and Mark E. Smith

Department of Earth Resources

Colorado State University

submitted to

Office of Water Research and Technology
U.S. Department of Interior
Washington, D.C. 20240

January, 1978

The work upon which this report is based was supported in part by funds provided by the United States Department of the Interior, Office of Water Research and Technology, as authorized by the Water Resources Research Act of 1964, and pursuant to Grant Agreement No. 14-34-0001-6136.

ENVIRONMENTAL RESOURCES CENTER
Colorado State University
Fort Collins, Colorado

Norman A. Evans, Director

ABSTRACT

Sensible responses to natural hazards should encompass the complementary elements of likelihood and severity. Flood hazard evaluation accomplishes this by relating frequency of occurrence to magnitude with the methods of flood frequency analysis. The use of these methods in current engineering and planning practice is based on the implicit assumption that the 100-year floodplain is fixed in areal extent, and once delineated, it will not be affected by future urbanization of the watershed. To the extent that floodplains are dynamic, however, nonstructural management strategies such as the National Flood Insurance Program cannot be expected to fulfill their objectives. Research during the past three years has shown that this assumption of static floodplains is unjustified in southeastern New England where the hydrologic response of watersheds to extensive urban growth is conditioned by geologic, pedologic and morphometric parameters. Data from 18 watersheds located in eastern Massachusetts and Rhode Island are used to develop a methodology whereby the change in discharge corresponding to both one and two percent annual exceedance probabilities may be predicted. A group of secondary data sources, including topographic maps, surficial geologic quadrangles and land use maps, are employed to develop indices of urban land use change, surficial watershed properties and drainage network configuration. The dependent variable is derived from two separate estimates of flood expectancy which are found by standard analyses of nonoverlapping segments of a basin's hydrologic record. It is expressed as the ratio of change in the 50-year or 100-year flood expectancy to the mean annual discharge. Multiple regression techniques have yielded two equations corresponding to these two flood expectancies. The resultant models account for approximately 75 percent of the observed variation in the response, are statistically significant at the one percent level, and reproduce the observed values of the hydrologic indices with reasonable accuracy. Moreover, the urbanization index is by far the most important predictor, although the network parameter and the pervious index contribute substantially to the model. Incorporation of this methodology in the Metropolitan Landscape Planning Model, which is being developed at the University of Massachusetts, can be expected to refine its applicability to floodplain management in the southeastern New England region. Furthermore, the rationale by which this model has been constructed is recommended for the development of comparable techniques in other rapidly growing metropolitan areas.

ACKNOWLEDGMENTS

This study was funded by a grant from the U.S. Office of Water Resources and Technology (O.W.R.T.), Project Number B-147-COLO with matching funds provided by Colorado State University.

The investigators wish to extend their appreciation to a number of individuals and agencies that gave assistance. Individuals at the University of Massachusetts include, Professor Julius Fabos who originally suggested the problem and contributed important imputs on various aspects of land use planning, Professor Joseph H. Hartshorn who provided invaluable information and insights on the surficial geology of eastern Massachusetts, Professor Alan W. Niedoroda who assisted in the early stages of modeling and Professor William P. MacConnell who provided the land use maps vital to the study. Mr. Charles W. Kane, also of the University of Massachusetts, stimulated our thinking about somewhat more philosophical aspects of the work. The critical observations of Mr. Charles F. Kennedy, Massachusetts Water Resources Commission, regarding potential problems in development of variables contributed to the effectiveness of the resultant model and are gratefully acknowledged. The Water Resources Branch of the U.S. Geological Survey (Boston Office) provided hydrologic data.

Professors Perry J. Brown, Frank G. Ethridge, Stanley L. Ponce, and Stanley A. Schumm as well as Mr. Donald H. Simpson all of Colorado State University contributed to the study in numerous ways. To all of the above, we wish to extend our sincere gratitude, however, it should be understood that the views contained herein are solely the writers' responsibility.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

A	basin area (square miles)
D_d	drainage density (miles per square mile)
E	E-index (Jarvis, 1972); modified E-ratio (this paper)
G	skewness of the logarithms of discharge maxima
H	link distance (Jarvis, 1972)
I	area of impervious deposits (square miles)
$I_{h \cdot n}$	hydrologic index for the n-year flood expectancy
I_p	pervious index
I_w	wetland index
$I_{\Delta u}$	urbanization index
K	frequency factor in the Log Pearson Type 3 distribution
L	modified link distance (this paper)
M	stream magnitude at a junction within a drainage network; mean of the logarithms of discharge maxima; rank of discharge maxima in calculation of plotting positions
N	number of years in a hydrologic record; number of observations
n	weighting factor in the pervious index; designation of a recurrence interval in the hydrologic index
P	area of pervious deposits (square miles)
$\hat{Q}$	predicted or estimated value of discharge
Q_{ma}	mean annual flood (recurrence interval of ~2.33 years)
R	estimated recurrence interval
R_b	bifurcation ratio (Strahler, 1964)
R_c	circularity ratio (Miller, 1953)
R.V.	reduced variate transformation of plotting position
S	area of swamp deposits; standard deviation of the logarithms of discharge maxima

S_c	gradient of the main stream channel (feet per mile)
U_1 U_2	area of pre-urban and post-urban land use respectively (square miles); correspond to 1952 and 1972 land use maps in development of the model
$\overline{X}$	mean annual flood (see Q_{ma})
α	intercept of the Gumbel line; confidence level
β	slope of the Gumbel line
ΔQ_n	change in the n-year flood expectancy
ΣL	total length of stream channels in a watershed

INTRODUCTION

Man is periodically confronted by a variety of hazardous phenomena, some of which result from his own activities, and others that are related to the normal processes of nature. These natural hazards vary in both their frequency of occurrence and degree of severity. For example, Gulf Coast residents face the yearly threat of hurricanes, whereas destructive floods are a fairly remote danger in most parts of the country; areas with steep slopes and evidence of previous landslides are likely to be more dangerous development sites than steep terrain where soil creep is the dominant process. This suggests that sensible responses of individuals and society to such natural phenomena should take account of the complementary elements of severity and likelihood.

Unfortunately, the formalization of those rather simple concepts in the techniques and analytical methods of engineering, planning and other supporting disciplines has shrouded the decision-making process in an aura of scientific and technological elegance. In other words, the resultant strategies, which generally combine elements of control, accomodation or avoidance, are commonly perceived by laymen as deterministic, precise and without significant error. Actually, the methods are often probabilistic and may contain extensive approximations, safety factors and professional judgment. The public sometimes seems inclined to relinquish its participatory role in decision-making to the experts, perhaps for reasons of apathy or lack of understanding. This is potentially dangerous, however, because the ultimate decisions regarding natural hazards do not depend solely upon scientific and technical objectivity. Substantial economic and political pressures are exerted which may or may not be in the best interests of the general public. It is indeed proper to seek specific answers to land use problems, but it is equally important to maintain some perspective on the uncertainties associated with the development of effective strategies which deal with the destructive and unpredictable aspects of nature.

Statement of the Problem

The severity of flood hazards extends over a wide range; they can be mere nuisances or can pose an extreme danger for both life and property. Stream valleys have always been highly desirable sites of human activity for a variety of economic and aesthetic reasons. Whether by foresight or as a result of experience, older development has been situated largely on the safer floodplain margins and higher terraces. As a result, current residential and commercial activities are induced to occupy progressively more hazardous floodplain areas in rapidly growing metropolitan regions. Numerous flood control structures have been erected in response to the obvious dangers inherent in this pattern of development, but events such as Hurricane Agnes in 1972 are clear warnings that such measures can be only partial solutions in the long run.

As a direct response to steadily mounting flood losses despite massive investments in structural controls, Congress established the National Flood Insurance Program in 1968 and further amended it in 1973. The major thrust of this legislation is twofold. In the first place, it is designed to promote land use controls in flood hazard areas as a supplement to flood control structures. These nonstructural techniques help to resolve the dilemma of government investment of huge sums on flood control projects which cannot prevent economic and social disruption by catastrophic floods. Such an approach views avoidance of the hazard as being preferable to the illusory and all too often partial protection against it. Secondly, through the mechanism of an insurance program, the National Flood Insurance Program will ensure that some of the costs associated with floods will be borne by those who choose to occupy these hazardous areas, and less by the nation's taxpayers in the form of disaster relief.

Many of the problems and implications of this program have been discussed by Platt (1970). They concern (1) the constitutionality of restricting the use of privately owned land in floodplains; (2) the complex legal and administrative questions surrounding the application of subsidized and actuarial insurance rates to new construction; (3) the enormous technical obstacles to the "accurate" delineation of all flood hazard areas in the United States; and (4) the many difficulties of basing land use regulation on sound and defensible hydrologic criteria. His major conclusion is that floodplain delineation should proceed expeditiously with the best, readily available information in the interests of public safety and welfare, rather than be delayed in expectation of some costly and arbitrary degree of "accuracy".

Both floodplain delineation and longterm regulation are of primary interest in this study because most current approaches implicitly assume that floodplains are fixed in areal extent. Different recurrence intervals have been advocated as a rational planning and engineering standard, but any design frequency is a compromise between public safety and the analytical limitations of accurately estimating rare events. The 100-year flood is currently favored by federal agencies and is also the standard prescribed by the National Flood Insurance Act. Regardless of the particular level of risk judged to be acceptable, however, an important consideration remains: that floodplains with recurrence intervals relevant to hydraulic design and land use planning are not necessarily static where urbanization subjects the watershed to rapid and dramatic modification of its hydrologic properties.

Previous Research

Relationships between urbanization and the occurrence of floods have been studied extensively during the past two decades. In 1961, Savini and Kammerer reported on the lack of studies relating urban growth to changes in stream regimen, and made a number of recommendations for future research, stressing the importance to land use planning and engineering activities. During the next few years, a number of publications of the U.S. Geological Survey were devoted to flooding characteristics in urban settings (e.g., Waananen, 1961; Mitchell, 1961; Carter, 1961; Riggs, 1965; Crippen, 1965; Wilson, 1967). Although these studies addressed several diverse problems in different parts of the country, they may be characterized by their univariate approach, their emphasis on individual hydrographs rather than flood frequency, and their reliance on relatively short hydrologic records. Some of them also contain questionable analytical procedures such as the assumption of simple linearity between the magnitude of peak discharge and the percent of a watershed which is impervious, and the averaging of flood frequency curves to develop regional relationships.

More recently, Anderson (1970) analyzed information from 81 sites, principally in the vicinity of Washington, D.C. He concluded that improvements to urban drainages may reduce lag time by a factor of eight compared to natural channels, and may increase the observed peak flow by a factor ranging from two to eight because of greater runoff volumes. The study is basically a more sophisticated replication of Carter's earlier work. In another investigation of urbanizing watersheds near Philadelphia, Hammer (1972) related channel enlargements to a host of land use and other watershed parameters, the most important of which were type of impervious area, soil drainage characteristics and slope. He also discussed the relation of these findings to studies by Leopold on flood frequency and channel geometry. In the latter case, Leopold (1968) suggested that the frequency of occurrence of low and intermediate magnitude floods is greatly increased by urbanization, but that these effects become insignificant for rarer events.

One of the more interesting studies with respect to the research reported here was conducted by Espey and Winslow (1974). Working primarily with watersheds in Texas, they observed comparatively larger discharges for all return periods in urbanized basins than in nearby non-urbanized basins. In one case, the increase seemed to reach a maximum near the design frequency of the storm sewers. In a second case involving two watersheds, however, the impact appeared to extend over the entire range of expectancies, being greatest for the 100-year flood. Unfortunately, there is some question regarding the significance of their findings because of the limited hydrologic data and the small number of watersheds on which they are based.

In summary, previous work has focused on fairly small urban basins in geographic settings of gentle, fluvial topography, has tended to emphasize the effects of urbanization on hydrographs, and has dealt with flood expectancy in quite general terms. Most importantly, almost no research has been conducted on this topic in the New England region where very different hydrologic responses to urbanization might be anticipated because of its many unique features inherited from recent glaciation.

Research Objectives and Overview

As part of a Metropolitan Landscape Assessment Model (METLAND) under development at the University of Massachusetts, a study by Cole and others (1974) demonstrated marked changes in flood expectancy for a small watershed located in the METLAND study towns of Burlington, Wilmington and Tewksbury, Massachusetts. This provided the impetus for a pilot study to look for similar effects of urbanization elsewhere in the Boston metropolitan area, and to investigate potential ways in which this dynamic element of urban hydrology could be modelled for predictive purposes. Hydrologic analyses of 26 basins revealed that dramatic changes in flood expectancy were indeed common in this region where extensive urban growth had occurred. Furthermore, detailed study of a representative sample of five watersheds produced a tentative model for predicting this response (Doehring and others, 1975).

This paper is concerned with the culmination of that initial study. The overall objectives which have directed the course of this research are

(1) to identify the significant variables which influence runoff characteristics of watersheds in southeastern New England;

(2) to evaluate the importance of these variables, singly and in combination, on the dynamic response of these watersheds in order to develop a predictive model relating urbanization and flood expectancy; and

(3) ultimately to develop a methodology which can be used by

planners and other land managers to forecast the effects of proposed development on flood hazards by means of readily identifiable and measurable parameters.

The realization of these objectives will enable professionals involved in long-range planning to estimate the changes in discharge corresponding to relevant design frequencies accompanying projected changes in land use within a watershed. These estimates can then be applied in floodplain mapping to avoid jeopardizing new construction which otherwise might become situated in the expanded floodplain as urbanization proceeds. Failure to account for this dynamic behavior of floodplains in certain settings may lead to a serious malfunction of the National Flood Insurance Program by charging rates which are inappropriate to the true hazards. It must be recognized that changes in the areal extent of floodplains cannot be directly evaluated in a general manner because of the complex hydraulic relationships among stream discharge, water surface elevation and area of inundation. Since discharge is essentially conservative along a stream reach regardless of the morphology, the discussions which follow will address the impact of urbanization on this parameter. Nevertheless, the reader should be aware that changes in discharge correspond to definite, although unspecified changes in the extent of floodplains.

Initial considerations for this research were made in the context of a process-response model (Smith and others, 1974). This is a useful device for identifying relevant factors in a conceptual problem as well as describing their possible relations to one another. As shown in Figure 1, the process elements consist of both independent and semi-independent factors of which climate, bedrock geology and surficial geology affect morphometry and land use. Hydraulic flow parameters are dependent on

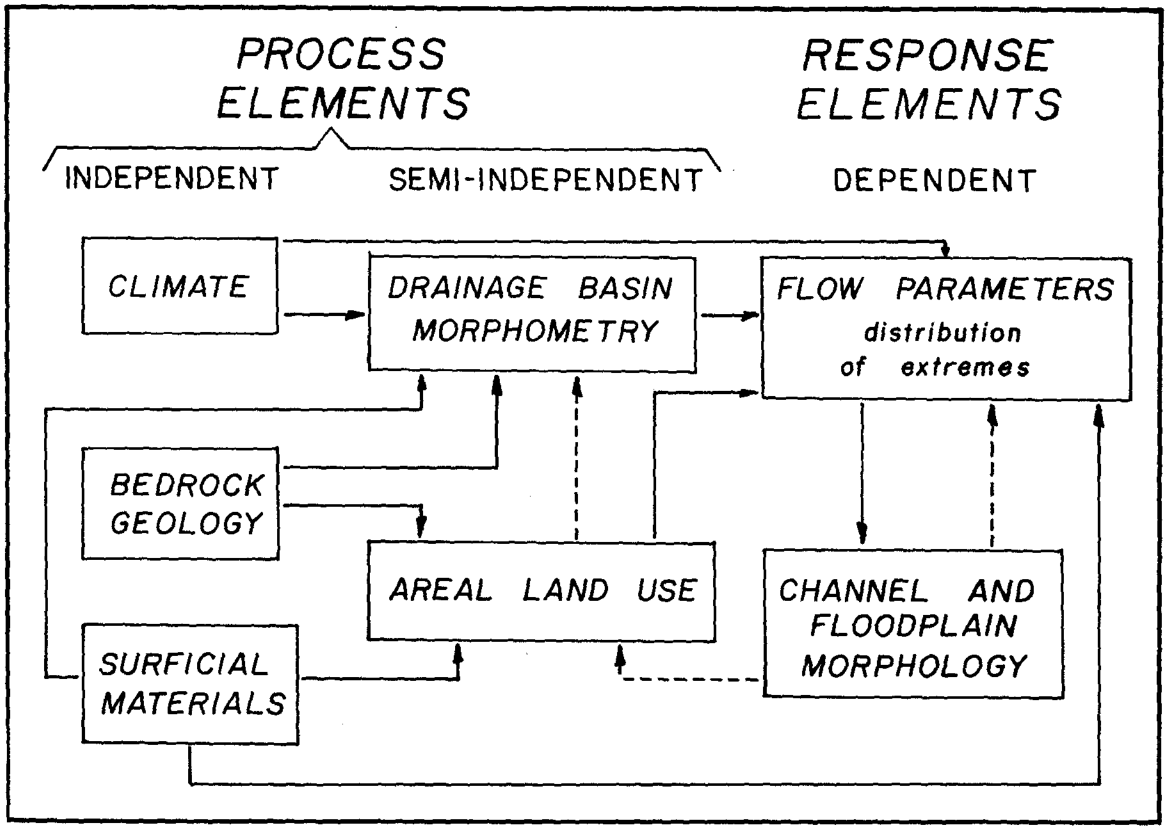
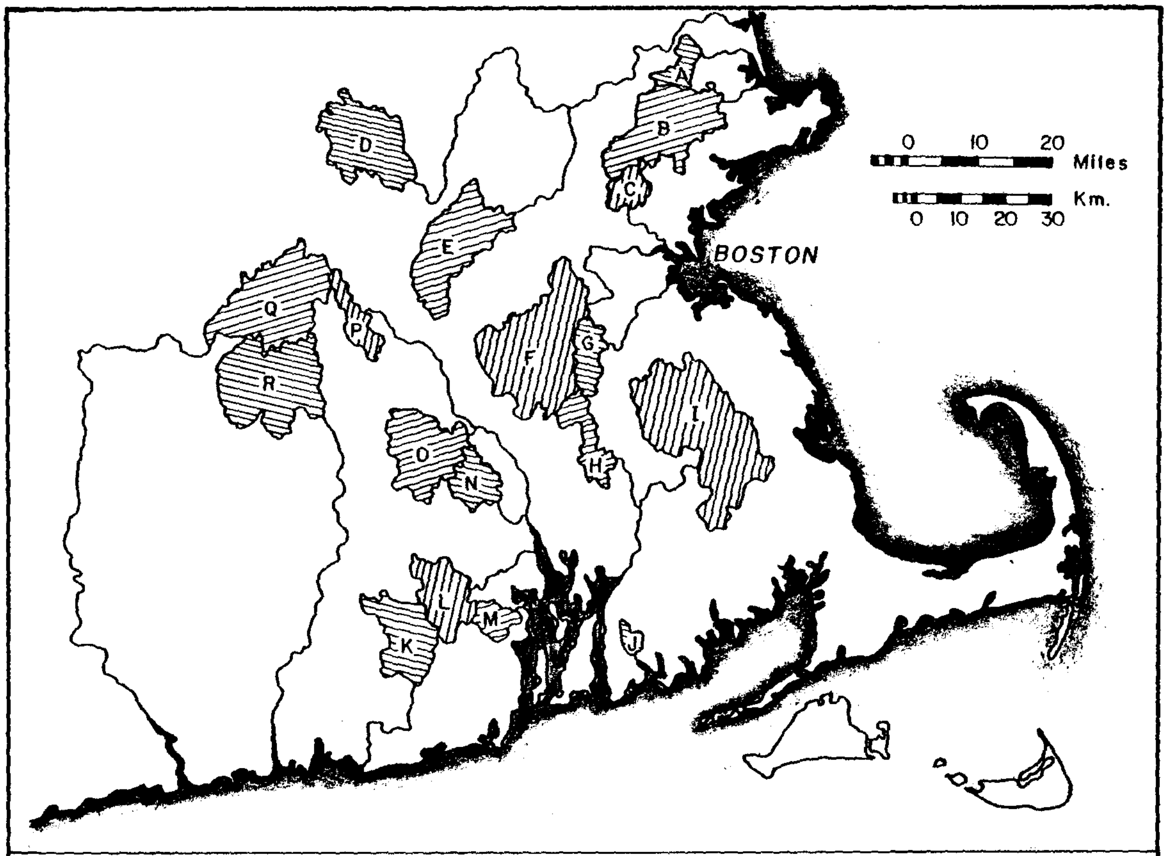


FIGURE 1. Process-response model of the research project.

surficial geology as well as on changes in climate, morphometry and land use. These parameters exercise controls on channel and floodplain morphology, which in turn influence land use, morphometry and flow parameters through feedback linkages shown as dashed arrows. Although this model does not contain all possible relationships, it has identified the most important variables and serves to define the system which will be quantitatively modelled.

It is reasonable to assume that increased urban land use within a basin will influence the frequency and magnitude of flood events in a variable manner. The exact response should depend on the combined and interacting effects of geology, topography and drainage network configuration. It is also likely that climatic fluctuations would influence the hydrologic regime, but detection of such changes would require detailed, closely-spaced meteorological data which do not exist. Therefore, climatic trends cannot be incorporated as a predictive factor in a model relating urbanization and changes in flood expectancy. Moreover, inclusion would have little usefulness since future climatic fluctuations cannot be estimated reliably.

The study area is located in southeastern New England as shown in Figure 2. The terrain of this region is distinctly glacial in origin. The western part is characterized by gently rolling, till-mantled hills with intervening ponds, swamps and stream valleys underlain by glacial sand and gravel. The eastern portions are generally flatter and consist of extensive outwash plains, wetland areas and scattered hills of till, bedrock or ice-contact glacial deposits. Owing to the humid climate, streams tend to be permanent and flow in deep, well-established channels. Twenty-six gaging sites were originally selected from all available sites in the region on the basis of the length of hydrologic records. Flood frequency relations were appraised for these watersheds and informally related to the historical pattern of urban growth. Eighteen of the original 26 watersheds were then identified as being suitable for model development on the basis of more detailed hydrologic information. The remaining eight were disqualified for several reasons including excessive regulation of streamflow, indications of major diversions during peak discharge, a drainage area larger than 500 square miles which was judged to be impractical and might produce a non-homogeneous sample, and finally redundancy of hydrologic information in the case of one watershed being a subdrainage of another. Since two of these criteria are a matter of degree, it is clear that the study watersheds are only relatively natural



A	PARKER RIVER at Byfield
B	IPSWICH RIVER near Ipswich
C	ABERJONA RIVER at Winchester
D	NORTH NASHUA RIVER near Leominster
E	ASSABET RIVER at Maynard
F	CHARLES RIVER at Charles River Village
G	NEPONSET RIVER at Norwood
H	WADING RIVER near Norton
I	TAUNTON RIVER near Bridgewater
J	ADAMSVILLE BROOK at Adamsville
K	WOOD RIVER at Hope Valley
L	SOUTH PAWTUXET RIVER at Washington
M	POTOWOMUT RIVER near East Greenwich
N	WOONASQUATUCKET RIVER at Centerdale
O	BRANCH RIVER at Forestdale
P	KETTLE BROOK at Worcester
Q	QUABOAG RIVER at West Brimfield
R	QUINEBAUG RIVER at Quinebaug

FIGURE 2. Index map and identification of gaging sites for 18 study watersheds.

hydrologic entities. However, the use of stricter selection criteria would have rendered the study infeasible.

The remainder of this paper will describe the development and application of a methodology by which the research goals have been achieved to a large extent. These efforts may be grouped into five distinct stages. First, a number of data sources were assembled to provide the requisite information on the geologic, morphometric and land use characteristics of the study watersheds. These data were subsequently converted into various parameters which would appropriately characterize each watershed. Second, hydrologic conditions of each basin were determined by standard analytical methods including flood frequency analysis. The response to urbanization was based on observed changes in frequency relations. Third, the resultant set of multivariate data was examined by statistical procedures which would identify the key variables for predictive purposes. Fourth, the relationships among geologic, morphometric and land use parameters and the hydrologic response were combined through multiple regression techniques and evaluated by means of several statistical criteria. Finally, the formal relationships derived in preceding stages were modified in several respects to facilitate their utilization by planners and others who are generally unfamiliar with the peculiarities of hydrologic models. The modifications also are designed to assure that the actual watershed response does not exceed the predicted response in the majority of cases.

DATA COLLECTION

The data requirements of this study were broad and could have been satisfied from a number of possible sources including maps and literature of various federal and state agencies, both published and unpublished information from universities and other research organizations, state and local records concerning land use, and original field work. As a result of this diversity, a fundamental decision was required at the outset in regard to (1) the adequacy of these sources to provide sufficient discrimination for the predictive model, (2) the feasibility of using them in terms of manpower, funding and time constraints, and (3) their eventual suitability for routine planning applications. It would have been beneficial from a strictly scientific viewpoint to rely solely on direct observations. However, the enormous area involved, as well as generally poor access due to land ownership and terrain, made such an approach impractical. Accordingly, secondary data sources were employed and were subject to limited field verification in situations where this was considered desirable and practical.

Watershed Morphometry

Standard U.S. Geological Survey 7½-minute quadrangles with a scale of 1:24000 were selected as the base map for this study, a choice which provided adequate detail for the purposes of the research. Furthermore, this particular scale is judged to be a desirable standard for such studies because of the general availability and familiarity of these maps. Smaller scale maps probably would not have been as effective, although they were not evaluated. Clearly, the use of widely different scales would introduce serious problems regarding data consistency when combined in a predictive model. Watershed boundaries were first drawn on the topographic quadrangles following generally accepted procedures which involve the isolation of all surface drainage above a stream gaging point and require the crossing of all contour lines at right angles. Although these techniques are something of an art, they are generally quite reproducible in fluvial terrain. In the glacial terrain of southeastern New England, however, greater uncertainty in the precise location of drainage divides arises from two factors: the presence of broad areas having low relief which is imperfectly portrayed by the fixed contour interval of ten feet; and the occurrence of wetland areas which drain both into and out of a watershed, and therefore lie on the divide. Such inaccuracies might be expected to cancel somewhat as they accumulate, but errors as large as a few percent may exist in various areal data.

Delineation of the drainage network was a matter of concern since several possible elements of the predictive model would be dependent on it. The glacial history of the study region has left a strong imprint on the drainage composition. Lakes, small ponds and wetland areas are abundant, and the drainage is deranged in many places, being neither closely related to the topography nor in a state of even approximate equilibrium. The majority of channels are permanent and some of the drainage is artificial. Because of these factors, any standard methods of quantitative measurement can only be approximate. One would expect all potential elements of the drainage network to be active during peak flow conditions. Therefore, it might be most appropriate to identify all such channels through detailed topographic analysis of contour crenulations and interpretation of stereo photographs. However, the great amount of labor involved in accomplishing this task for a total watershed area of 1500 square miles was prohibitive. It is also questionable whether the procedures could be performed reliably by non-specialists. Instead, a more easily replicated method has been selected whereby only those channels designated as permanent or intermittent on 1:24000 topographic quadrangles are included. This approach is judged to be the most uniform and reliable alternative, despite the likelihood of different interpretations by the various cartographers who produced the maps.

Once the watershed and drainage network were delineated, values of basin area, basin perimeter and the total length of stream channels were obtained by means of an electronic graphics calculator having a linear resolution of 0.01 inches. The drainage network was evaluated further in terms of its morphometric properties. Two general approaches have become prominent, the one based on stream orders and the other based on stream magnitude. With regard to the process-response model described earlier, the objective was to discriminate among different drainage patterns with respect to their hydraulic characteristics. Ideally, this would incorporate factors of channel length, slope and cross-sectional area, as well as the branching structure. However, the measurement of all these elements would have involved an unreasonable amount of labor, and therefore, attention has focused on the branching structure.

Networks have been coded by a method adapted from Smart (1970) and other workers. The coding is based on the individual stream link which is a channel segment bounded by junctions or sources. This enables the drainage network to be represented by a vector of integers in which exterior links or sources are distinguished from interior links. Instead of a simple binary coding, however, the technique used here accounts for the abundance of glacial ponds and their possible influence on the hydrologic response. This coding system is applied to a hypothetical network in Figure 3. Source links are coded as +2 or +1, depending on the presence or absence of a pond in the link. Interior links are coded as 0, -1 or -2, depending on the occurrence of ponds and their position with respect to the upper junction. The vector of integers is derived by proceeding headward from the outlet and first including the subdrainage with the least number of sources at each junction. Hence, the basin depicted below would be coded as

(0 -2 1 1 0 2 0 0 1 2 -1 1 0 1 1).

The main channel of each watershed is then specified as the channel whose source corresponds to the last element in this vector, and its length is measured from the topographic maps.

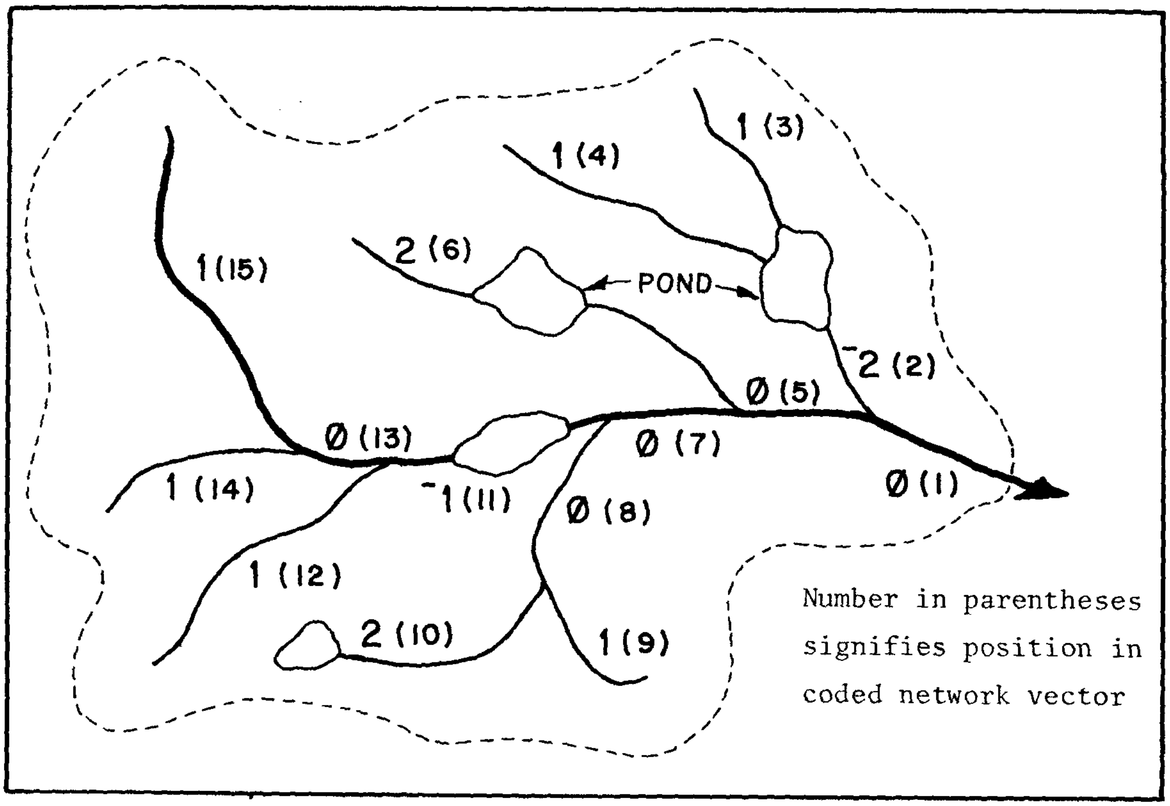


FIGURE 3. Method of coding the drainage network.

Thus, the complete set of basic morphometric data for each watershed includes area, perimeter, total length of stream channels, length of the main channel, and a coded vector of network composition. All of these properties are potentially important in evaluating the hydrologic response to urbanization because they represent the spatial pattern within which hydrologic processes occur and reflect the amount and time distribution of runoff. The drainage network of the Neponset basin is presented in Figure 4 on which both the Strahler ordering and the coded network vector are shown. This unusual and irregular channel pattern is fairly typical of the glacial terrain in the study area.

Surficial Materials

Information concerning the character of the land surface within each watershed was derived from a combination of topographic map interpretation and U.S. Geological Survey surficial geologic quadrangles. Areas in Rhode Island were evaluated by means of groundwater maps which closely correspond with the surficial geology. Soils maps of the U.S. Soil Conservation Service were not utilized because of their reliance on criteria other than texture. No uniform data source exists which would provide complete coverage of the study area at a suitable scale, hence some interpretation of landforms was clearly necessary. Previous experience of many workers (J.H. Hartshorn, personal communication) has demonstrated the excellent correspondence between landforms and surficial deposits in this region. That is, the pattern of individual deposits is a complex mosaic in which the differing topographic expression of particular kinds of material, such as sand and gravel, has been closely controlled by different processes associated with deposition. For example, kame deltas

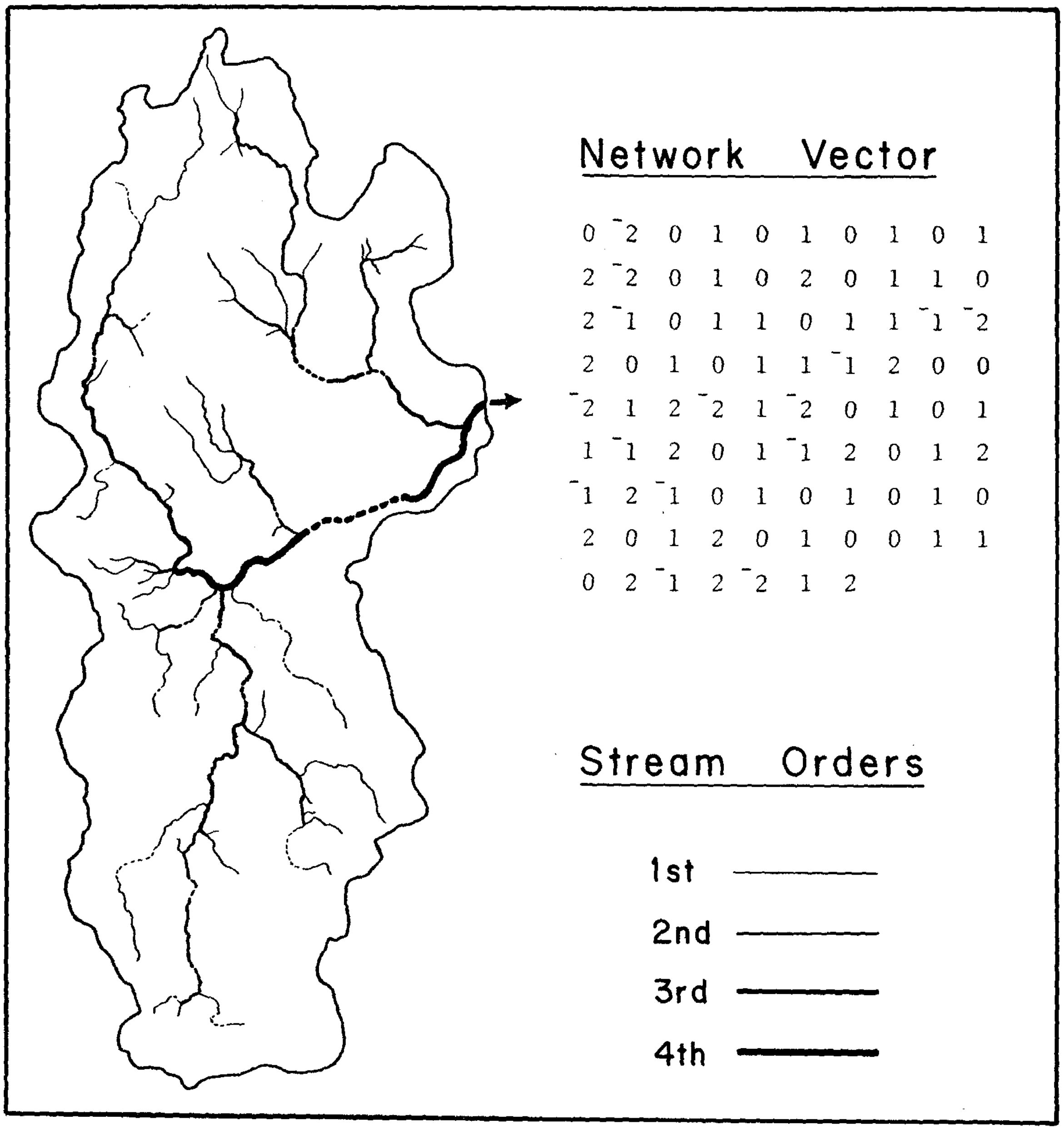


FIGURE 4. Drainage network of the Neponset basin.

and terraces typically display pronounced breaks in slope and occupy particular locations in the terrain, whereas till generally produces more subdued landforms and slope elements. In retrospect, total reliance on landform interpretation might have been preferable for the sake of uniformity. Nevertheless, it was used in a complementary fashion with available geologic maps in the study. Limited verification of this method was carried out in the early phases of the work and the degree of accuracy further justified its use.

While recognizing that the types of surficial units are neither homogeneous nor discrete with respect to texture, three principal types were focused on here.

(1) Ponds and swamp deposits generally occur as isolated units or along reaches of streams. Rainfall on these areas rapidly becomes part of streamflow or is otherwise detained at the surface owing to low permeability of the substrate.

(2) Impervious deposits include bedrock, glacial till and lake beds, and some fine-grained, postglacial marine deposits. Bedrock and till usually occur in upland areas, while lake beds and marine deposits are found in lower areas associated with swamp deposits. Runoff to streams is relatively rapid from these materials.

(3) Pervious deposits include glaciofluvial sand and gravel, and are found adjacent to upland areas as kame terraces, along streams as valley train and modern stream terrace deposits, and as scattered positive relief elements in lowland areas. Runoff to streams is relatively slow because of depression storage on irregular surfaces and high infiltration capability of the coarse-grained materials.

The abundance and distribution of these types of materials in a watershed can be expected to have a profound influence on the ultimate amount of runoff resulting from rainfall, snowmelt or a combination of the two.

The areal extent of these surficial units in the Neponset basin is illustrated in Figure 5. The amount of pervious deposits in this watershed is about average for the study region, although considerable variation exists among the 18 basins. The typical abundance of well-sorted,

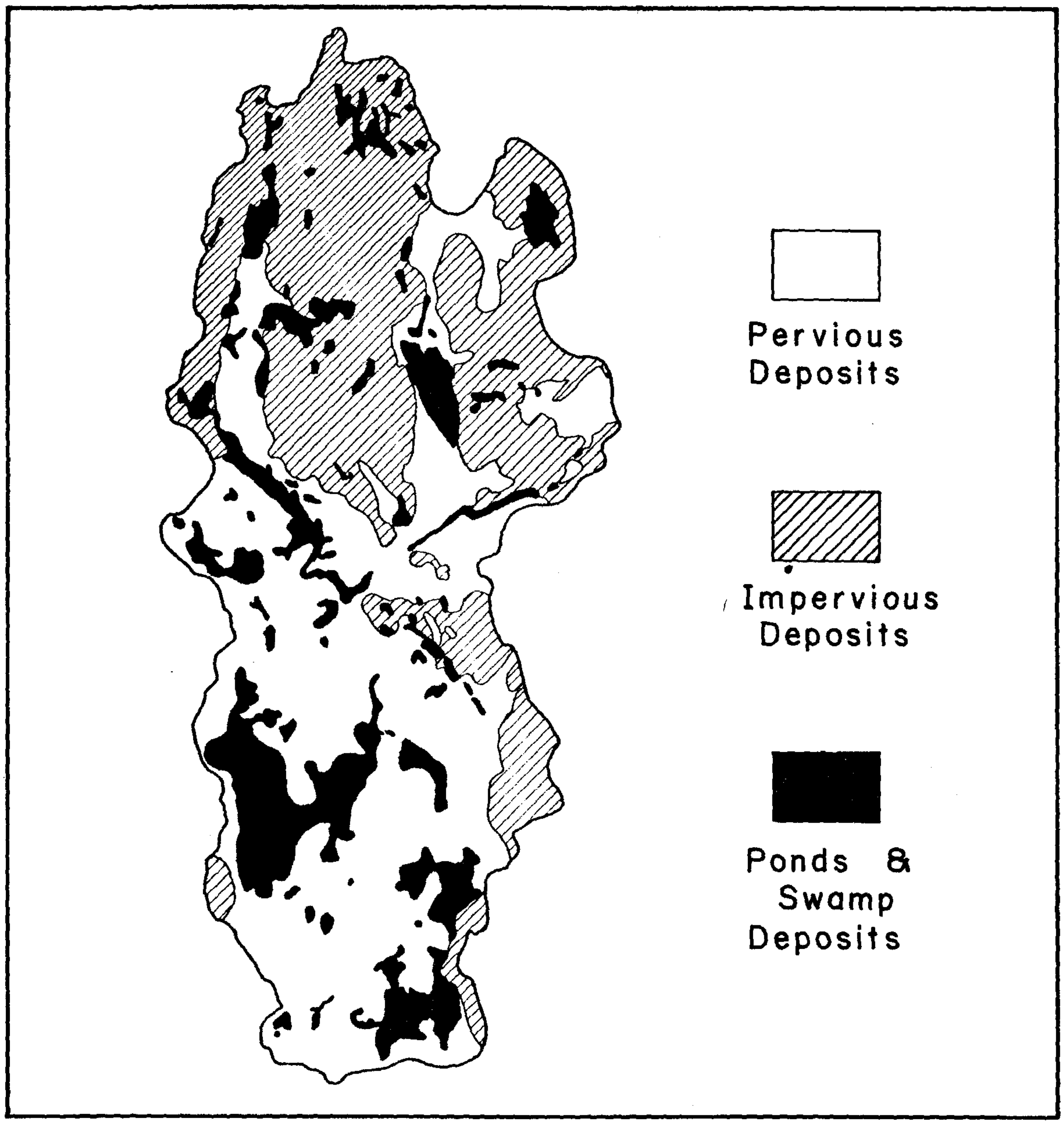


FIGURE 5. Areal distribution of surficial materials in the Neponset basin.

coarse-grained glacial outwash, as well as the desirability of these materials for development sites, is illustrated in Figure 6.

Land Use Data

This study was provided with an extremely valuable data resource in the form of detailed land use maps. They were compiled by William

FIGURE 6. Exposure of an extensive outwash deposit showing nearby suburban development.

MacConnell (1973) of the Forestry Department, University of Massachusetts on the basis of aerial photography taken in 1952 and 1972. Since Rhode Island coverage is limited to the second set of land use maps, topographic maps issued in the early 1950's provided the necessary land use information in five of the watersheds (K through O in Figure 2). The validity of this substitution was evaluated by carefully studying areas in Massachusetts where both the earlier topographic maps and the 1952 land use maps were available.

MacConnell's detailed land use classification scheme was simplified into three major types including wetland areas and water bodies, open or forested land, and urban land. Some difficulties were encountered owing to the imperfect correspondence between specific land use types on the two sets of maps. Assignment of specific types to a major group was accomplished by inspection of numerous areas on the maps and examination of classification descriptions provided by MacConnell. Areas of major land use types were obtained by dot-area grid since the intricate pattern was unsuitable for measurement by planimeter. Grids were drawn for the two map scales (1:31680 and 1:24000) with each dot representing 0.00585 square miles or 3.75 acres. Land use in the Neponset basin for 1952 and 1972 is illustrated in Figure 7. Urban land use in this watershed has increased from about 15 to 25 percent of basin area.

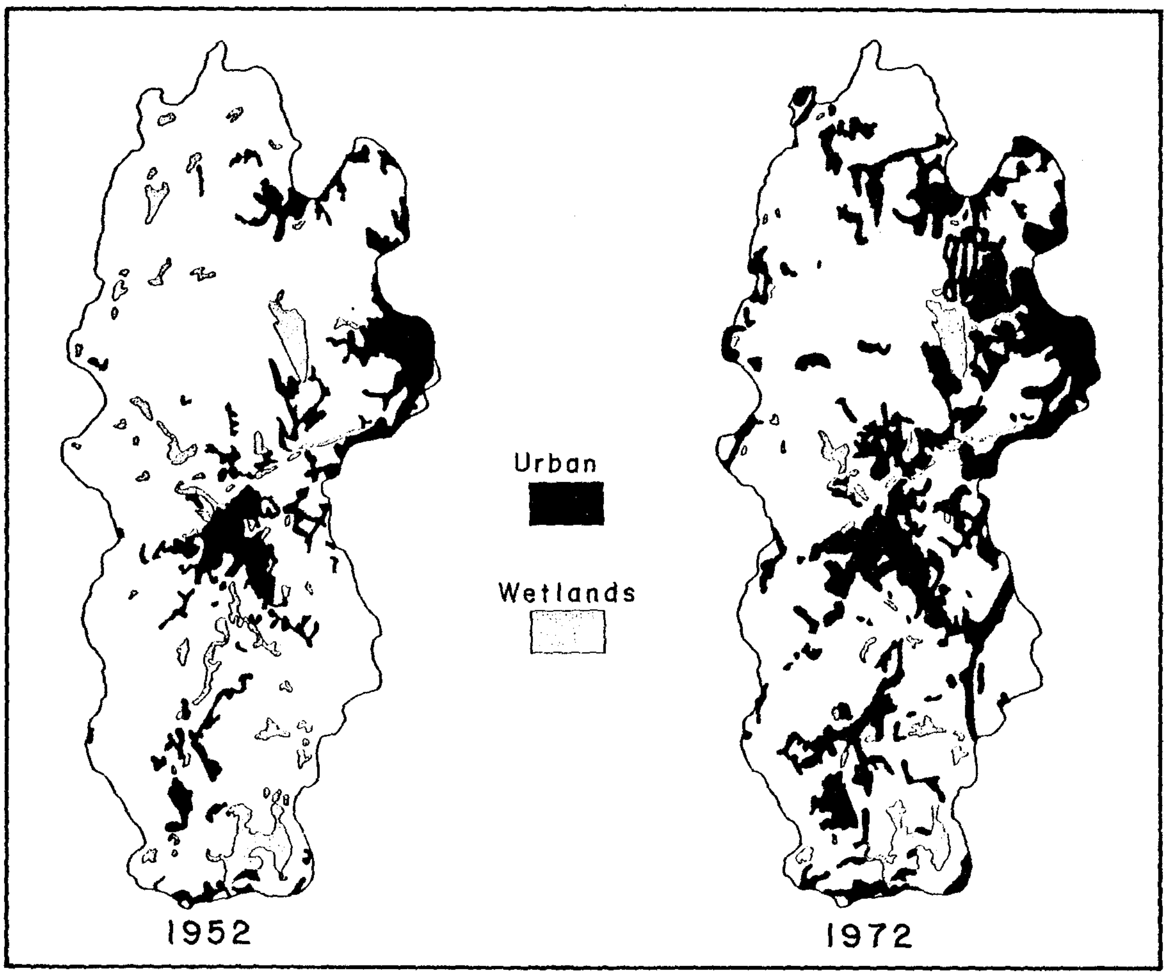


FIGURE 7. Land use in the Neponset basin for 1952 and 1972.

The patterns of land use may be contrasted between the rural, western parts of the study area as shown in Figure 8, and the more urbanized conditions of easterly parts of the region near Boston as depicted in Figure 9. In the second case, the relation between suburban development and the distribution of pervious surficial materials is again evident, as is the loss of dense forest which often accompanies modern development. The less attractive side of urbanization may be observed in Figure 10. Here, one of numerous wetland areas in eastern Massachusetts has been pre-empted by a parking lot with a concurrent loss of water supply, wildlife habitat and floodwater storage capacity. During the period of time considered in this study, the obliteration of wetlands has been one of the most insidious effects of expanding urban and suburban areas (Larson, 1973). Development in floodplains is the central concern of this study, and Figure 11 illustrates a new residential area along the Assabet River near Concord, Massachusetts. The continued approval and construction of such developments, despite their aesthetic appeal, is both shortsighted and irresponsible.

Hydrologic Data

All discharge data required in this study have been compiled by the U.S. Geological Survey and published in various Water Supply Papers. Primary hydrologic data included values of annual momentary peak discharge with one observation per year of record. The length of complete records in the 18 study watersheds ranges from 36 to 56 years with many of the records beginning around 1940. In addition, estimated values of peak discharge for the 1936 flood have been made at seven of these sites by the Geological Survey on the basis of indirect stage-discharge

FIGURE 8. Aerial photograph of western part of study area.

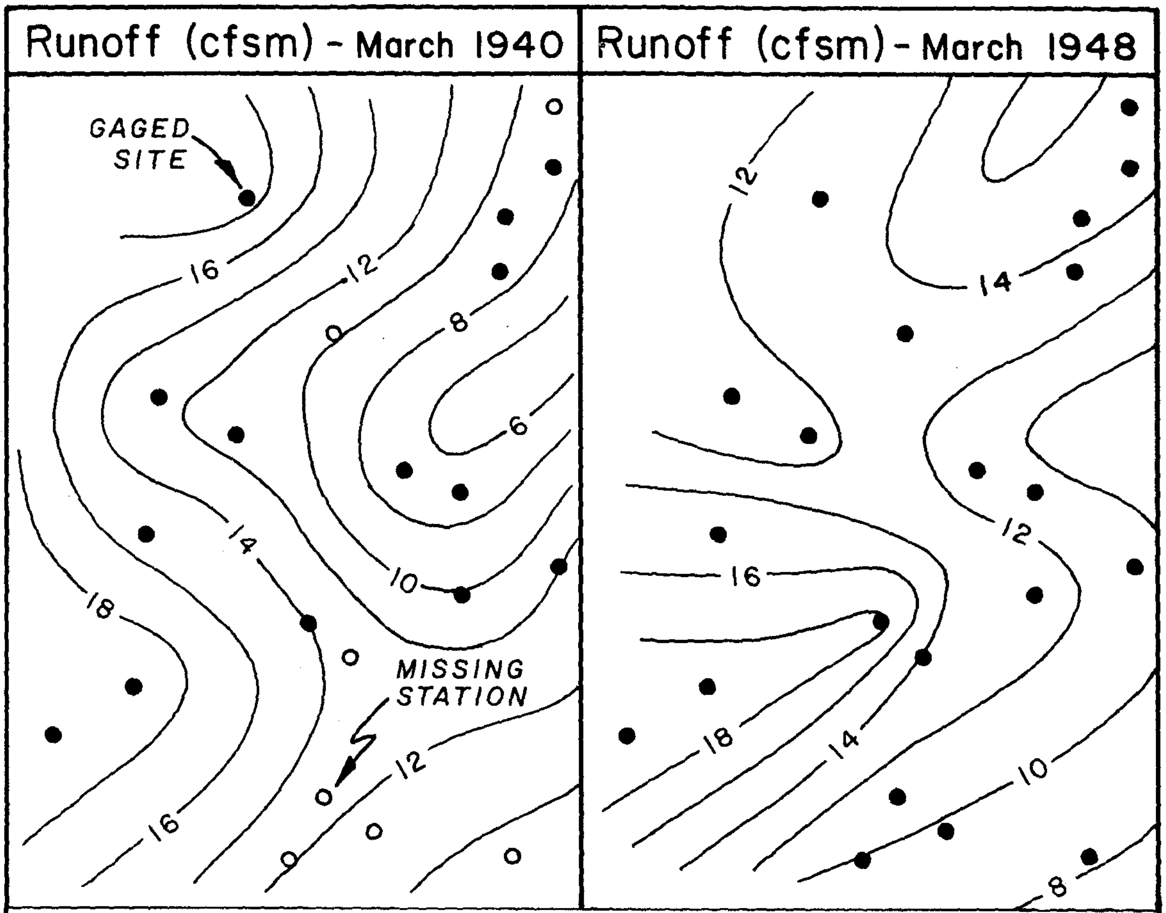
FIGURE 9. Aerial photograph of eastern part of study area.

FIGURE 10. Wetland filling in Burlington, Massachusetts.

FIGURE 11. Floodplain development near Concord, Massachusetts.

relations. Secondary data included monthly maximum values of average daily discharge which were employed in two supplementary roles. One of these is related to a regionalization technique which was developed to extend all records of annual maxima back to 1940. It involved the estimation of either one or two values for seven watersheds during years of about average or below average maximum discharge. Twenty-five distinct runoff events were selected from the hydrologic years 1940 to 1949, and the maximum average daily discharge was converted to discharge per square mile for each gaged site. These values were plotted at the visually estimated watershed centroids and a contour map was developed where all stations were represented. These maps were then utilized as possible models for other runoff events in which all of the gaging sites were not represented.

The largest runoff events of 1940 and 1941 were identified and compared to patterns of later events. The most similar pattern served as a guide for extending contours in the vicinity of the ungaged sites, and this extension of contour lines was guided to a greater extent by plotted values at stations near the ungaged site than by those farther away. It should also be noted that later events which served as models were only selected from the same part of the year as the runoff event which was being synthesized. The estimated values of unit areal discharge at the ungaged sites were converted to total discharge of the watershed. The final step was the calculation of momentary peak discharge from maximum daily runoff by means of linear relationships which had been developed for this purpose. An example of this regionalization procedure for the South Pawtuxet basin is presented in Figure 12.



Runoff (cfsm) - March 1940

Runoff (cfsm) - March 1948

Estimated areal runoff = 12.8 cfsm

Watershed area = 61.54 sq. mi.

Maximum average daily discharge ($\overline{Q}_d$) = 787.7 cfs

Peak discharge = 1.135 ($\overline{Q}_d$) - 24.5

= 870 cfs

FIGURE 12. Regionalization method of synthesizing hydrologic data on an areal basis.

DEVELOPMENT OF VARIABLES

On the basis of the conceptual model, there were two possible structures within which a predictive model could be developed. One alternative would be a regionalization method in which the discharge corresponding to a specific recurrence interval is predicted from a collection of factors including the degree of urbanization. The second alternative would take the more direct approach of predicting the dynamic response of basin hydrologic regime from changes in land use and other relatively fixed watershed parameters. The choice is not a simple one, however, because there are problems associated with either approach.

In the first case, an artificial doubling of the degrees of freedom would result from having two sets of observations on land use. That is, there would be associated with each basin two measurements of urban land use, two estimates of the N-year discharge, and a series of replicated watershed parameters related to area, drainage network properties and surficial characteristics. Although greater degrees of freedom are clearly an advantage, the statistical validity of such a model with non-independent observations of morphometry and surficial properties is questionable. Primarily for this reason, it was not an acceptable model for this research. In the second case, the problem of spurious correlation arises from the use of indices and ratios needed to quantify observed changes. The purely statistical arguments have been lucidly presented in a paper by Benson (1965). He demonstrates how apparent correlations may arise among unrelated variables which are reconstructed as complex ratios with common elements. Therefore, he correctly urges that great care be exercised when such variables are used to make statistical inferences. On the other hand, one should not lose sight of the objective which, in the present instance, is to predict a change in flood expectancy. Therefore, precautions have been taken to ensure that any ratios or indices developed truly represent the situation and are not artificial constructs. This will become clearer in the sequel.

Morphometric Variables

It was necessary to express quantitatively the hydraulic efficiency of the drainage system and also describe the overall shape of the watershed. Several mathematical expressions have been proposed for basin shape, but because of the deranged drainage and prominent linearity exhibited by some basins and their networks, and also because it is a nondirectional measure, the circularity ratio (Miller, 1953) was selected. It is defined as the ratio of basin area (A) to the area of a circle having a perimeter (P) equal to that of the basin, and is calculated as

$$R_c = (4\pi A) / P^2 . \qquad (1)$$

The value is sensitive to the detail with which the watershed boundary is drawn, but given a common map scale, the effect should be uniform within the study area.

The hydraulic efficiency is conceptually related to three aspects of the drainage. First, the degree of dissection is ordinarily expressed as the drainage density

$$D_d = (\Sigma L) / A \qquad (2)$$

where the total length of channelized flow (ΣL) is divided by basin area. Higher values are associated with more efficient and rapid runoff, depending on the intensity and duration of rainfall and the directness with which flow proceeds to the outlet. Second, the gradient of channels will influence the magnitude of discharge at the outlet by accelerating or retarding the flow. An expedient measure of average channel gradient is found by dividing the elevation difference of the main channel by its length, and is represented here as S_c in feet per mile. Third, the timing of runoff, and consequently the flood hydrograph, are dependent on the drainage composition or branching pattern since this governs the way in which increments of basin discharge are combined. The bifurcation ratio (R_b) is generally regarded as a standard measure of this property. It is derived by regressing the logarithms of the number of Strahler stream segments of each order against the corresponding order. Then

$$R_b = 10^{-m} \qquad (3)$$

where m is the slope of the regression. The value may be calculated from the network vector described earlier by means of the APL defined function STRAHLER listed in Appendix A. However, the bifurcation ratio is a fairly conservative property and does not discriminate very well between different network configurations.

An alternative to the bifurcation ratio has been proposed by Jarvis (1972). It is based on the concepts of stream links and magnitude which were introduced earlier. A link is a channel segment bounded by sources or junctions, and the magnitude of any junction refers to the number of sources or first-order streams located above the junction. As a more discriminating measure of network composition, Jarvis proposed an E-index defined as

$$E = (\Sigma M_i H_i) \;/\; (\Sigma M_e H_e) \qquad (4)$$

in which M is magnitude, H is link distance or number of links between the junction and the outlet, and subscripts i and e refer to interior and exterior links respectively. Since all of the M_e terms are equal to one, the E-index may be conceived as a comparison between the sequential accumulation of flow components within the drainage network and the external elongation of the drainage. Interior and exterior links are distinguished in Figure 13, where link distance and magnitude have been noted for several of the network junctions.

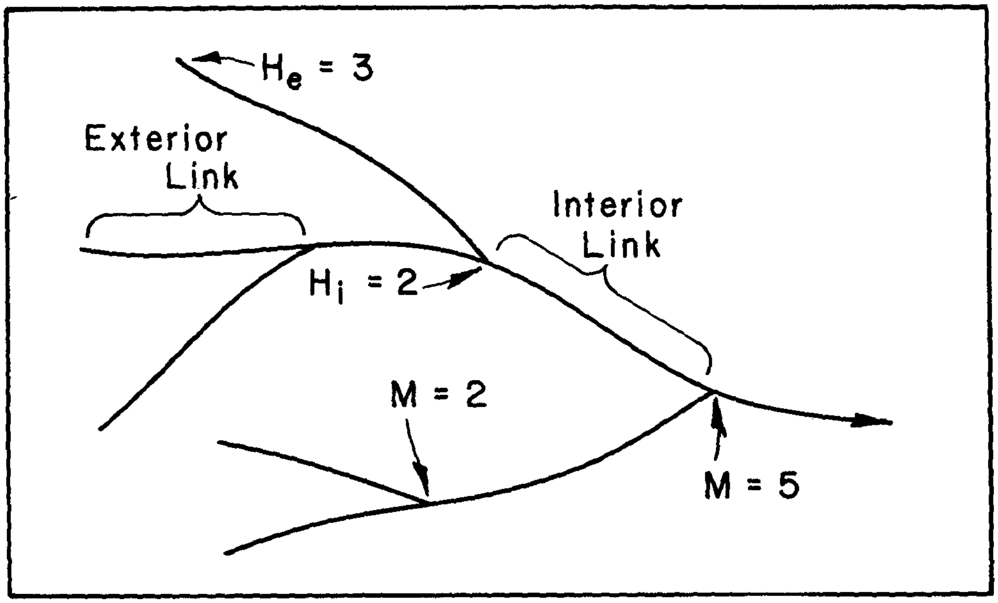


FIGURE 13. Illustration of terminology related to the E-index.

Two modifications of the E-index were made by this author. First, the link distance of any junction was increased by one for each intervening pond as a representation of the hydraulic attenuation. This increases the value where ponds are abundant on the mainstream or near the outlet, and decreases it where most ponds are located at the periphery of the basin. Second, it was observed that the E-index is dependent on overall network size, and it was found empirically that dividing by the logarithm of basin magnitude would standardize the relationship. In other words, this modification yields a network measurement which is independent of its overall size. The parameter used in this study, which will be referred to as the modified E-ratio, is given by

$$E = \frac{(\Sigma M_i L_i) \ / \ (\Sigma L_e)}{\log_{10}(\Sigma M_e)} \tag{5}$$

in which L is modified link distance, and the remaining terms are as previously defined. The modified E-ratio may be calculated by using the APL defined function ERATIO listed in Appendix A. The relationship between this parameter and several hypothetical network configurations is illustrated in Figure 14.

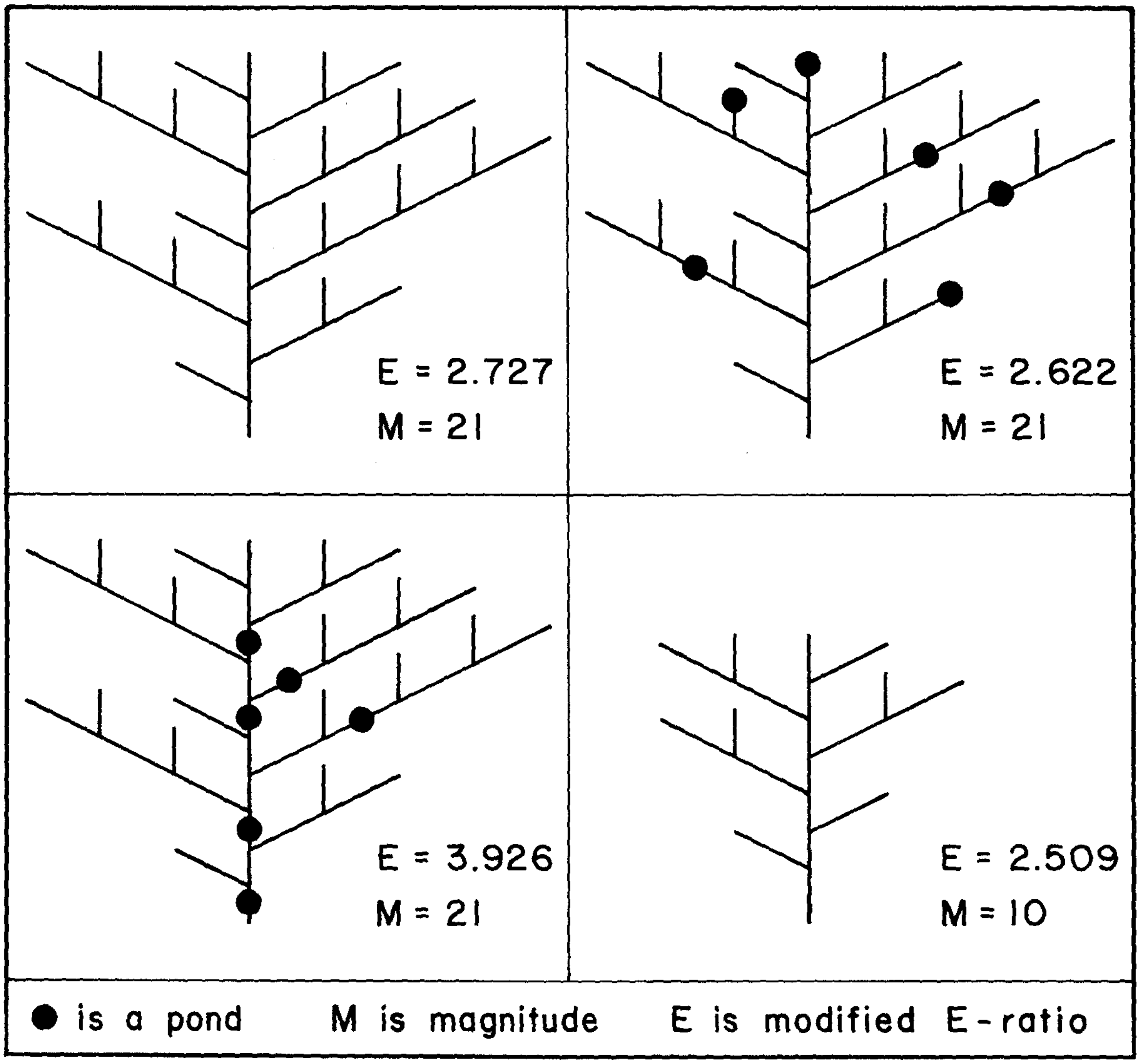


FIGURE 14. Network patterns and values of the modified E-ratio.

Land Use Changes and Urbanization

Data were collected on three land use types as described earlier. During the course of the study, however, it became clear that delineation of wetland areas and measurements of change between 1952 and 1972 were not reliable, largely because of the high seasonal variation in these zones and the imprecise criteria used in their identification on the maps. Therefore, land use considerations focus exclusively on the urban category.

In order to quantify the changes that have occurred in urban land area over the 20 year interval, an index was developed to conform with several criteria. First, the index ought to be normalized to basin area. Second, it should be sensitive not only to the absolute change, but also to the initial and final degree of urbanization. It would be unreasonable to expect the same hydrologic response to accompany a change from 10 to 20 percent urbanized as that which accompanies a change from 50 to 60 percent. Lastly, a simple percentage was considered to be undesirable for use in a parametric model since it is constrained by an upper bound. The formulation of the urbanization index which was finally selected is

$$I_{\Delta u} = \left[\frac{U_2 - U_1}{A - U_2}\right] \cdot \left[1 - \frac{U_1}{A}\right] \qquad (6)$$

in which U_1 and U_2 are the areal extent of urban land in 1952 and 1972 respectively, and A is watershed area. The values of this index for various combinations of U_1 and U_2 are presented as a series of curves in Figure 15. It may be noted that the values tend to converge as the final extent of urban land use becomes large, and also that no upper bound exists for this parameter. It appears to be most discriminating in its lower region where nearly all of the study watersheds would be plotted.

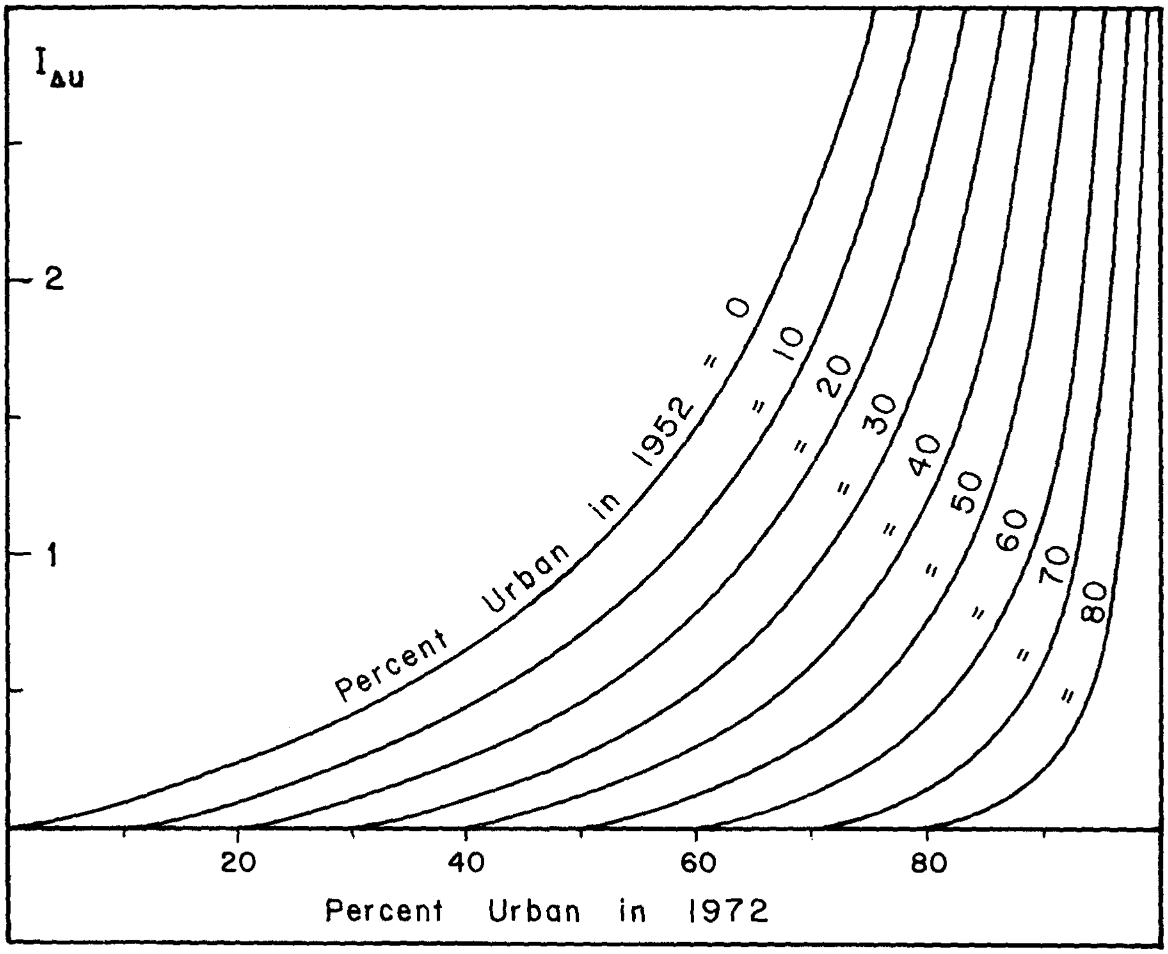


FIGURE 15. Distribution of values for the urbanization index.

Surficial Properties

In order to characterize the integrated hydrologic response of a watershed's surface, a pervious index was developed. The value of percent pervious was rejected as a predictive variable for reasons cited in the discussion of the urbanization index. The ratio of (Area of Pervious Deposits) to (Area of Impervious Deposits) was considered, but it does not incorporate all available information. An index of the form

$$I_p = P \;/\; (I + n \cdot S) \tag{7}$$

was preferred in which P is the area of pervious deposits, I is the area of impervious deposits, S is the area of swamp deposits and wetlands, and n is a weighting factor to account for the differing hydrologic response of the latter two surficial units. A value of two was found by means of iteration to yield the optimal correlation with the hydrologic response variable for the southeastern New England region. This is reasonable since runoff from wetland areas may be expected to contribute to streamflow much more rapidly than runoff from till and bedrock located in the interfluves.

A wetland index was also derived as

$$I_W = S / I \tag{8}$$

in which all terms have been defined previously. This was inspired by the findings of Larson (1973) who has associated changes in flood occurrence with recent filling and obliteration of wetland areas. It was also included as a possible surrogate for the wetland category of land use which was omitted from the model.

Hydrologic Response

The hydrologic response variable developed in this study is based on changes in flood expectancy, a concept which is generally approached by means of flood frequency analysis. The rationale, limitations and assumptions of flood frequency techniques will be discussed for those readers unfamiliar with them. Flood frequency analysis attempts to answer the question "What discharge should one expect to be equalled or exceeded at this particular site for a given probability?" on the basis of previous observations. Certain assumptions are required to formulate an answer. A hypothetical population of annual peak discharges is assumed to exist which is unlimited in time and magnitude; physical limitations of rainfall amount and intensity are ignored. The observations on which flood expectancy is based are assumed to be a random sample from that population, and consequently, they are assumed to be statistically independent events. Finally, it is assumed that the population may be described mathematically in terms of the frequency with which events of different magnitude occur. The expected frequency of occurrence in the population diminishes as the magnitude increases.

The rationale of flood frequency analysis is to examine the sample of observations (the hydrologic record) on the proposition that it is a good representation of the population about which we wish to make inferences. Since the frequency distribution of the population is unknown, we must select some mathematical distribution as potentially appropriate. Those used in such analyses belong to a family of curves which differ in shape according to the values of their parameters. In order to fit a distribution to a collection of observations, these parameters are estimated from statistical properties of the sample. The estimated parameters are then presumed to describe the population of floods.

Unfortunately, there are no clearly established guidelines for assessing the correctness of a particular distribution, and it is often simply a matter of judgment. Three parameter distributions have the advantage of flexibility over two parameter distributions, but there is less reliability in estimating a greater number of parameters from a given sample. The Log Pearson Type 3 is a three parameter distribution which is currently favored by federal agencies and others for most applications. It is fitted on the basis of the first three moments of the logarithms of discharge which are the mean (M), the standard deviation (S) and the skewness (G). Estimates of discharge ($\hat{Q}$) having some specified recurrence interval are given by

$$\hat{Q} = 10^{(M + K \cdot S)} \tag{9}$$

in which K is a frequency factor related to the skewness and the recurrence interval of interest.

Another distribution commonly used in flood frequency analysis is the Gumbel. Each annual maximum discharge is assigned a plotting position according to the expression (N+1)/M, where M is rank in descending order of magnitude and N is the number of observations. Therefore, the largest discharge in 49 years has a recurrence interval of 50 years, the second largest of 25 years (having been equalled or exceeded twice in 49 years), and the smallest of 1.02 years. Hence, no finite discharge can be assigned a recurrence interval of one year. A transformation of this plotting position into a reduced variate

$$\text{R.V.} = -\log_e\left[-\log_e\left[1 - \frac{M}{(N+1)}\right]\right] \tag{10}$$

enables a curve to be fitted by least squares of the form

$$\hat{Q} = \alpha + \beta\cdot\text{R.V.} \tag{11}$$

where α and β are the intercept and slope of the calculated regression. In most of the study watersheds, the Log Pearson distribution is clearly superior to the Gumbel in describing the observed frequencies. Nevertheless, the reduced variate transformation has been used in a graphic flood frequency algorithm which was developed for this project and is presented in Appendix A. The analytical methods which lead to the hydrologic response variable will now be discussed.

The overall objective is to quantify any change in flood expectancy which may have occurred between 1952 and 1972, the dates of the land use maps. Two separate flood frequency relations were required for each watershed corresponding to time frames defined in this study as pre-urban and post-urban. Subdivision of the entire record to obtain these separate estimates was based on several factors. Foremost was the correspondence of flood frequency relations with the land use maps. The set of hydrologic data represent an interval of time, whereas the maps do not. Although perfect correspondence was not possible, a break at 1952 or shortly thereafter was considered acceptable. Secondly, it is beneficial in flood frequency analysis to have as long a record as possible on which to base the estimates. However, it clearly would be invalid to use the entire record for the post-urban segment since this would result in the comparison of some maxima with themselves. Consequently, there could be no overlap of the two segments. Thirdly, it is desirable to have segments of nearly equal size so that the largest values of each segment are assigned similar plotting positions to avoid unnecessary bias. A uniform split could not achieve equal subsample size due to the unequal lengths of hydrologic records, but a break in the range 1952 to 1954 would optimize this criterion. Since no large floods occurred in this interval, complete uniformity was not considered essential.

A final consideration in selecting data for the two flood frequency analyses encompasses the outlier problem. An outlier is an observation which is so large that it is not representative of the population with respect to the size of sample in which it has been observed. In other words, the properties of the population should not be inferred from its observed frequency of occurrence. It is both important and desirable to include any observed extremes in appraising changes in the occurrence of rare events, but the mathematical properties of the distributions, especially values of skewness in the Log Pearson, are strongly dependent on the largest observations. Therefore, the inclusion of extremely rare events could distort the analyses and bias the separate estimates.

The problem of identifying outliers was dealt with in this study in the following manner. It was first recognized that all annual maxima are statistically independent, hence more than one outlier may occur in a single record regardless of its length. The observation of a 500-year flood and a 1000-year flood at a single station during 20 years would be considered unusual, but it would not be impossible (Gretener, 1967). Therefore, the entire record is evaluated sequentially, starting with the largest observed maximum discharge. At each step, the estimated recurrence interval of the suspected outlier is determined on the basis of the frequency curve for the remaining maxima as shown in Figure 16.

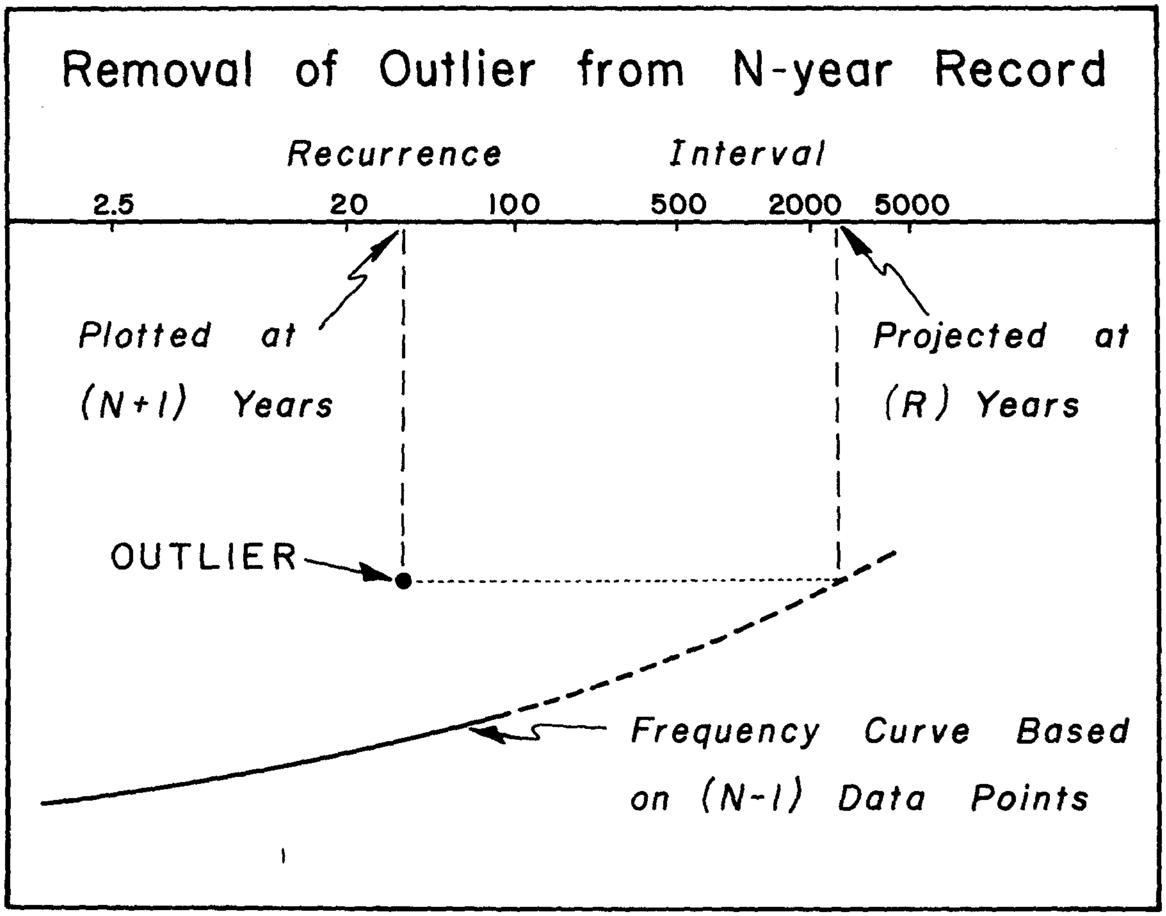


FIGURE 16. Method used in the identification of outliers.

The question is then asked, "What is the chance of so rare an event occurring in the period of time represented by this hydrologic record"? Since the occurrence or non-occurrence of a single extreme is a binomial phenomenon, the probability of the outlier may be described by a Poisson distribution for small samples. Consequently, the probability of occurrence (P) in N years of the suspected outlier with estimated recurrence interval R is given by the expression

$$P = 1 - e^{-(N/R)} \tag{12}$$

where e is the Naperian base. The decision of retaining or rejecting the suspected outlier is made by comparing a pre-determined significance level with this value. In this study, $\alpha=0.05$ was chosen as a reasonable compromise between Type I error (rejecting the outlier when it should be retained) and Type II error (retaining the outlier when it should be rejected. When an outlier is rejected from the record by this process, it is replaced by the next largest observation from the same hydrologic year. These substitute values were obtained by converting a value of monthly maximum average daily discharge to a peak discharge using the linear relationships described earlier. In all cases, the value was less than the mean annual flood. The complete procedure is then repeated for the next largest observation remaining in the hydrologic record until a suspected outlier is retained. Finally, the resultant record of maxima is divided as specified above and the analysis of these segments follows.

Flood frequency curves were calculated and plotted together for pre-urban and post-urban data segments as shown for the Neponset basin in Figure 17. The vertical scale is coded in units of the mean annual flood ($\bar{X}$) to accomodate basins of different area and runoff volumes. Both the linear reduced variate scale and a recurrence interval scale are shown

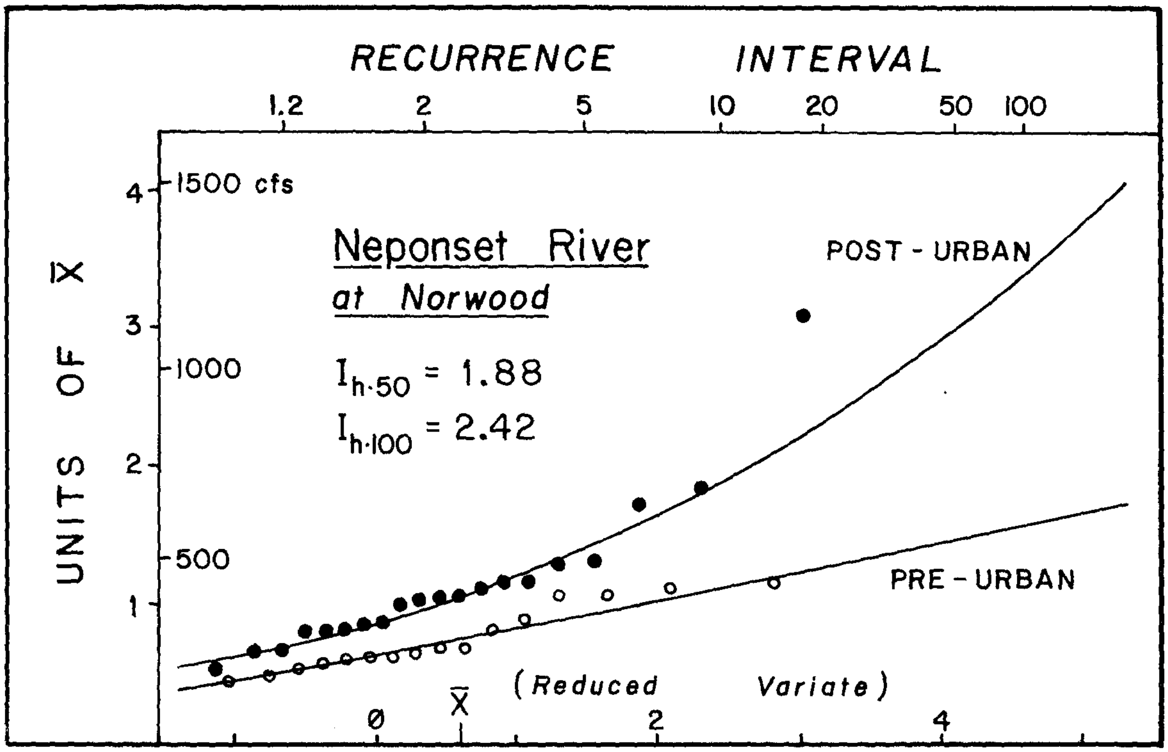


FIGURE 17. Flood frequency curves for the Neponset basin.

on the abscissa. Flood frequency curves, based on the Log Pearson Type 3 distribution, are presented in Appendix B for the 18 study watersheds, and should be examined carefully during the discussion which follows.

Results of these analyses were combined into an index of hydrologic change for each basin. Since the model was intended to describe a dynamic process and included basins of different area, the options for an appropriate response variable were restricted. There was no apparent way to combine the three parameters of the Log Pearson distribution into a single index to represent the change, and for planning applications, the index must be keyed to some design frequency such as the 100-year discharge required by the National Flood Insurance Program. The simple algebraic difference between pre-urban and post-urban expectancies was obviously unsatisfactory since it would not represent a homogeneous response for basins of different area. The remaining two alternatives were the ratio of the two expectancies, or the ratio of the algebraic difference (ΔQ_n) to the pre-urban mean annual flood (Q_{ma}). The latter was selected since it is conceptually more satisfactory and incorporates more information from the flood frequency curves. The mathematical form is

$$I_{h \cdot n} = \Delta Q_n \,/\, Q_{ma} \tag{13}$$

where $I_{h \cdot n}$ is the hydrologic index for recurrence interval n. Although the major focus of the predictive model is the 100-year expectancy, the index was calculated for the 50-year event as well, because of greater reliability associated with its estimation from short records. These values are recorded for the Neponset basin in Figure 17.

During development of this hydrologic response variable, the problem of mixed distributions in flood frequency analysis and its bearing on the study were considered. The use of annual maxima is widely accepted as the proper method of calculating annual flood expectancies. However, if two or more populations corresponding to different parts of the year are sampled in these annual data, the resulting estimates may be incorrect by an indeterminate amount. The distinct seasonality of different flood-generating events in the Northeast (spring runoff, late summer hurricanes) makes the existence of a mixed distribution quite likely. The effect on the hydrologic index is a matter of speculation, however, because the index is derived from three discharge estimates - the mean annual flood and two expectancies of large recurrence interval. The mixed distribution may have different or opposite effects on each of these estimated values. Clearly, the resolution of this problem is well beyond the scope of the present study, and consequently, the hydrologic index has been based on the standard method of analysis.

Before proceeding to an account of the statistical modelling, a brief recapitulation of the variables would be appropriate. The morphometric parameters of drainage density (D_d), circularity ratio (R_c), main channel gradient (S_c), bifurcation ratio (R_b), and modified E-ratio (E) encompass most hydraulic characteristics of the watershed such as the efficiency and timing of runoff. Changes in land use, specifically urbanization, are represented by an index ($I_{\Delta u}$) which is sensitive to the absolute change, the individual pre-urban and post-urban values, and basin area. The surficial character of each watershed, which governs the amount of runoff and interacts with the drainage network properties to affect timing, is represented by a pervious index (I_p) and a wetland index (I_w) which are derived from the areal extent of pervious, impervious and swamp deposits. Finally, a hydrologic index ($I_{h.n}$) has been developed from flood frequency analyses of two non-overlapping segments of the record of annual maxima, and it relates the change in a predicted design event to the mean annual flood. All of these values, in addition to other basic data for the study watersheds, are contained in Appendix C.

STATISTICAL ANALYSIS AND MODELLING

Models are developed with different objectives in mind. They may be useful in describing some physical relationships or tentatively explaining an observed phenomenon; they may serve as a theoretical framework for subsequent research; or they may be developed to predict future trends or the expected response of a complex system. Finally, they may be comprehensive, having more than one of these qualities, or highly specific. In general, however, models are organizational and synthesizing tools which simplify the complexities of the real world into a representation that can be used for some purpose. Thus, a model should not be confused with reality, nor should it be applied carelessly in ignorance of this imperfect correspondence.

The principal objective of model development in this research is to predict the impact of urban growth on flood expectancy amidst other relatively fixed watershed characteristics. This goal was pursued in two separate stages.

Relationships Among Pairs of Variables

Prior to generating any predictive equations, it was considered advisable to examine the relationships among potential independent variables, as well as between each of them and the dependent hydrologic variable. Therefore, correlation matrices were computed using both the Pearson product-moment and Spearman rank formulations. The purpose of this step was to identify useful predictors having a high degree of correlation with the hydrologic response, but low correlation with other predictors. The matrix of Pearson correlation coefficients is presented in Table 1. The majority of these coefficients are small and insignificant. With respect to the two hydrologic indices, however, the modified

S_c	R_c	E	I_p	I_w	$I_{\Delta u}$	I_{h50}	I_{h100}	
-.15	.02	.54	-.06	.03	-.23	-.23	-.23	D_d
	.53	-.33	-.41	-.46	-.21	-.02	-.01	S_c
		-.51	-.09	-.34	.03	.09	.06	R_c
			-.15	.07	-.22	-.53	-.51	E
				.55	.58	.54	.53	I_p
					.12	.14	.20	I_w
						.76	.76	$I_{\Delta u}$
							.96	I_{h50}

TABLE 1. Pearson correlation coefficients for potential independent variables and the hydrologic indices for N=18.

E-ratio, the pervious index and the urbanization index reveal fairly strong relationships. Some of the correlations among independent variables are interesting, but not particularly relevant to the predictive model. For example, basins with higher E-ratios, which indicate a more compact stream network, tend to have denser channels and more efficient drainage. It may also be noted that steeper channel gradients are associated with basins of more circular shape. On the other hand, relationships among the first three variables identified above (E, I_p and $I_{\Delta u}$) are highly relevant to the predictive model, especially the correlation between the pervious and urbanization indices. This result is not entirely unexpected because of the many regulatory inducements to the siting of development in areas underlain by pervious deposits.

Correlation analysis has the effect of consolidating the covariance of these data into a collection of single numbers which tends to mask the important details of individual observations. Consequently, graphic plots were prepared for many of these relationships in order to reveal the presence of any nonlinear trends or unusual responses of individual watersheds with respect to a particular parameter. The selection of variables is more likely to be reliable when based on this combined evidence (Draper & Smith, 1966, p.239). Four of these relationships are presented in Figure 18 where the 100-year hydrologic index is plotted against the modified E-ratio, the pervious index and the urbanization index, and the latter two are plotted against each other. All relations seem to be essentially linear, with the possible exception of the E-ratio. The appearance of a hyperbolic trend prompted the adoption of its reciprocal for the next stage of modelling. The relationship between urbanization and the hydrologic index is fairly well-defined for all but watersheds F and B, and is in the expected direction. The apparent modification of this primary response by the pervious index may be seen by comparing watersheds H and D, M and B, or N and G. For similar, moderate degrees of urbanization, the hydrologic response is greater in the more pervious watershed. In several other cases, however, the association between urbanization and the pervious index tends to mask this relation. The effect of the modified E-ratio is evident from comparisons between watersheds G and E, or B and M in which lower values of this parameter appear to enhance the hydrologic response to urban growth. It may prove helpful for the reader to examine the flood frequency curves in Appendix B in the context of the preceding discussion. The association between flood expectancy and urban growth is further illustrated for two basins in Figure 19, where frequency curves are accompanied by graphs of percent urban in 1952 and 1972. These diagrams strongly suggest that the effect being modelled is real and not an artifact of data manipulation.

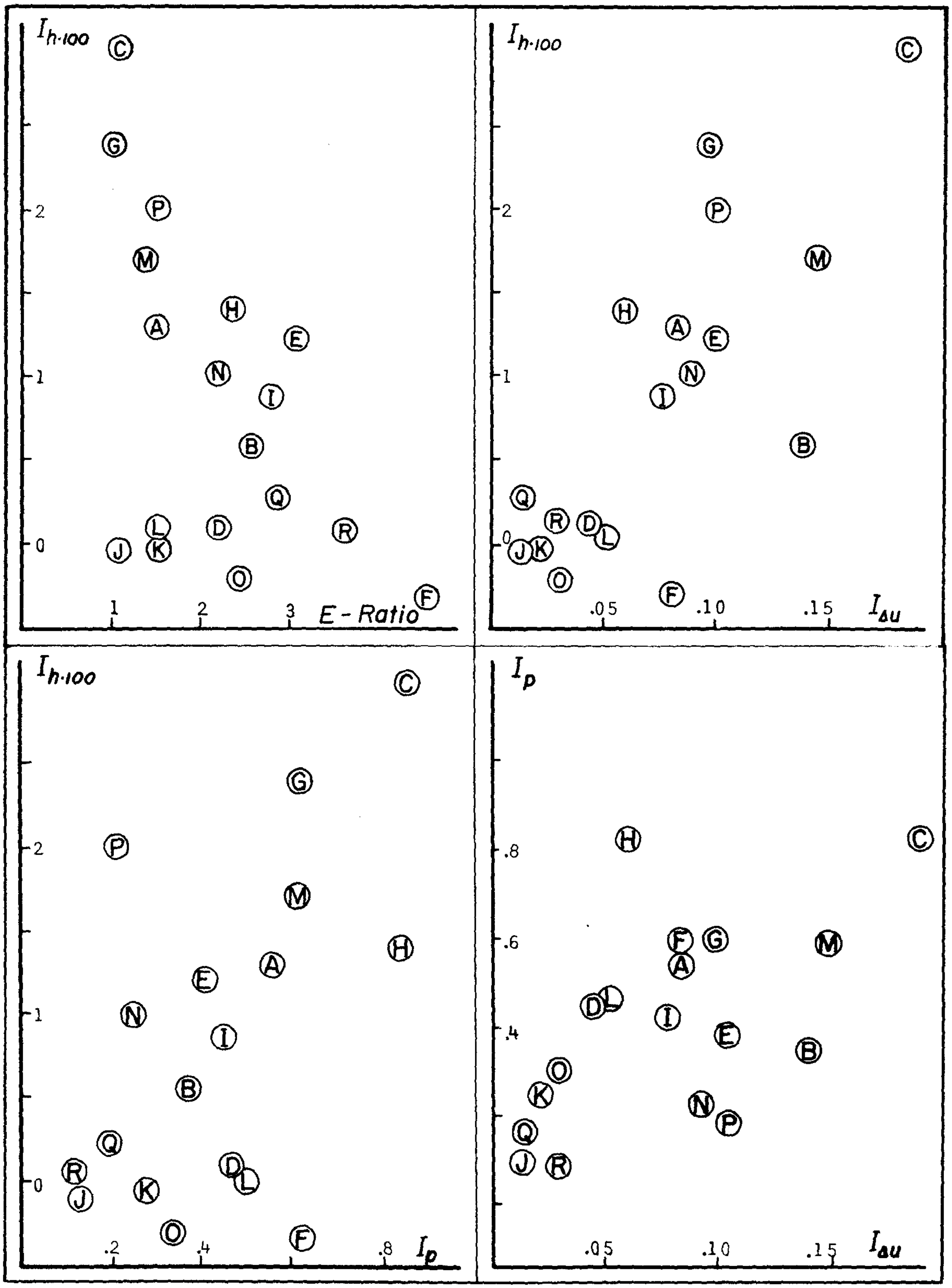


FIGURE 18. Graphic relationships among variables. The letters refer to the index map on page eight.

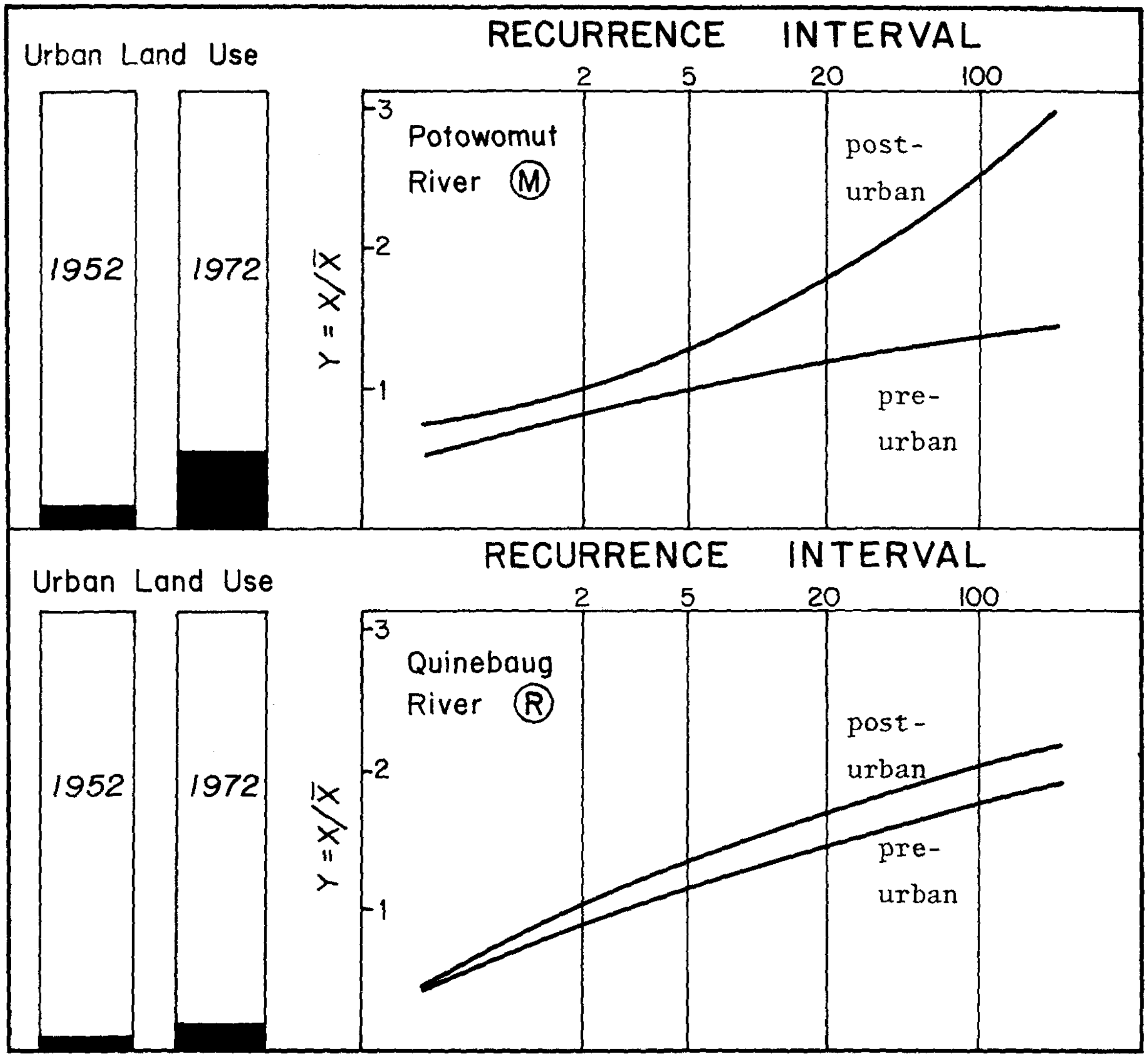


FIGURE 19. Examples of the association between changes in flood expectancy and urban growth.

In the case of some variables, one or more of the 18 basins display a response that appears anomalous with respect to the remaining observations. This is not surprising in view of the complex physical system which is being modelled, and it may indicate a source of variation that has not been accounted for, such as the spatial distribution of urban land use and details of the topography. Several compound variables, including $I_p \cdot I_{\Delta u}$, R_c/E, $I_{\Delta u}/E$ and $E \cdot D_d$, were also investigated as predictors, but they do not make any significant additional contribution beyond the single variables. They are also more difficult to interpret from a physical standpoint. Therefore, the final set of independent variables selected for the development of predictive equations includes the modified E-ratio, the pervious index and the urbanization index.

Derivation of Predictive Equations

In conceptual terms, a predictive equation is a mathematical relationship which estimates the most likely response of a physical system under some specified conditions. A rational basis for such estimates is prior observation of the system under a variety of conditions, which generally implies that one or more factors have been identified as being related in some way to its behavior. It is advantageous to gather the necessary data within a well-defined experimental design, but in hydrology, data must often suffice which are already in existence. One of the formal, statistical procedures whereby predictive relationships are derived is known as regression analysis which employs the least squares criterion. In simplified terms, the desired relation between the system response and one or more possible controlling factors satisfies the condition that the sum of squared deviations between the actual observations and the predicted responses is minimized.

The computational methods of regression are complex and need not be discussed here. A stepwise multiple regression routine (listed in Appendix A) was developed for this project, although the stepwise feature was not needed in the final modelling. The algorithm was adapted from one given by M.A. Efroymson (Draper & Smith, pp.178-195) and is described briefly with the program listing. Regression analyses of data obtained in this study produced a total of four predictive equations which are presented in Table 2. $I_{h \cdot n}$ is the hydrologic index of the N-year flood, $I_{\Delta u}$ is the urbanization index, I_p is the pervious index, E is the modified E-ratio, and N is the number of observations. Graphic plots of the residuals ($Y - \hat{Y}$) for these equations did not indicate the presence of any bias or violation of the statistical assumptions.

$$I_{h \cdot 50} = -0.70 + 8.57\ I_{\Delta u} + 0.35\ I_p + \frac{1.12}{E} \qquad N = 18 \qquad (14)$$

$$I_{h \cdot 100} = -1.08 + 12.18\ I_{\Delta u} + 0.45\ I_p + \frac{1.49}{E} \qquad N = 18 \qquad (15)$$

$$I_{h \cdot 50} = -0.57 + 8.39\ I_{\Delta u} + 0.66\ I_p + \frac{0.76}{E} \qquad N = 17 \qquad (16)$$

$$I_{h \cdot 100} = -0.91 + 11.95\ I_{\Delta u} + 0.86\ I_p + \frac{1.01}{E} \qquad N = 17 \qquad (17)$$

TABLE 2. Predictive equations for the two hydrologic indices.

In addition to the full complement of 18 watersheds, a subset of 17 observations was used as the basis for two equations. The tentative omission of the Charles basin (designated as F in Figure 18 and elsewhere) is considered to be warranted by its unusual response in regard to both urbanization and pervious indices. Although this watershed has undergone a moderate amount of urbanization comparable to others nearby, and also displays a high degree of perviousness, its hydrologic indices are the lowest of all 18 watersheds. A detailed examination of the hydrologic data and other physical characteristics of the Charles basin reveals nothing which satisfactorily accounts for this radical departure from the responses of the remaining study watersheds.

The evaluation of these alternative equations employed a number of statistical and non-statistical criteria because no single candidate was the obvious best choice. Three statistical criteria were used to judge the overall effectiveness: (1) the significance of the computed regression as measured by an F-ratio, (2) the standard error of estimate as a percentage of the mean response, and (3) the percent explanation of the total observed variance in the hydrologic response. While the given order corresponds to their priority in this study, a balanced combination of these criteria is considered preferable to an excessively large value of any single criterion. This rationale is based on their partial redundancy resulting from a common basis in the residual mean square.

It is also important to evaluate each of the independent variables in terms of its relative contribution. Sequential contributions are shown by each stepwise increase in percent explanation as predictors are added. A more realistic inference is gained from the partial F-ratios for variables in the regression. These values indicate the relative importance of a particular variable as if it were the last to enter the regression (Draper & Smith, p.119). Values of the overall F-ratio, standard error of estimate (S.E.), percent explanation (R^2) and partial F-ratios for each independent variable are summarized in Table 3. Note that the equations in which N=17 are slightly superior on the basis of all but one criterion, although the differences are not pronounced.

In addition to these formal statistical criteria, there are certain non-statistical considerations. For conceptual reasons embodied in Figure 1, it is important that the final predictive equation incorporate representatives from each of the major groups of variables - land use, morphometry and surficial properties. Furthermore, only one relationship for each hydrologic index should be selected for clarity in planning applications, and the choice should fulfill the study objectives.

Equation		Dependent Variable	R^2	F	S.E.	Partial F		
						$I_{\Delta u}$	I_p	1/E
N = 18	(14)	$I_{h\cdot 50}$	73%	12.3	56%	11.5	0.4	6.6
	(15)	$I_{h\cdot 100}$	71%	11.5	67%	11.3	0.3	5.6
N = 17	(16)	$I_{h\cdot 50}$	77%	14.3	47%	13.4	1.6	3.1
	(17)	$I_{h\cdot 100}$	75%	12.8	57%	12.6	1.2	2.5
			$F_{(3,13,0.01)} = 5.7$			$F_{(1,13,0.05)} = 4.7$		

TABLE 3. Statistical properties of the predictive equations.

The simultaneous evaluation of these diverse considerations resulted in the choice of equations (16) and (17) as being optimal. The overall F-ratio and the partial F-ratios for $I_{\Delta u}$ are significant at $\alpha=0.01$ for all four equations, and therefore, neither is a factor in the decision. Greatest importance is attached to the relatively low values of standard error of estimate and the balanced contribution of all independent variables. Furthermore, the coefficients are similar to the corresponding values in equations (14) and (15), especially for $I_{\Delta u}$. Considering the focus of this project, the chosen relationships appear to be the best predictive equations and the most appropriate ones for planning studies.

Omission of the Charles basin, which has an unusually low hydrologic response for the urban growth observed, tends to make the resultant model more conservative. This may be seen in Table 4, where observed values of the 50-year and 100-year hydrologic indices are listed with the responses predicted by each of the equations. The most accurate estimate has been underlined for each basin. A comparison of these values reveals superior estimates in the majority of cases for the favored equations (with N=17) which generally predict a somewhat larger response than the model based on all 18 watersheds. This is also true for values of the index above 0.5 which perhaps has greater practical significance. In the face of uncertainty, error in a conservative direction is judged to be preferable to an underestimate of the impact of future urbanization.

$I_{h\cdot 50}$ Observed	N = 18 Predicted	N = 17 Predicted	$I_{h\cdot 100}$ Observed	N = 18 Predicted	N = 17 Predicted
0.965	0.966	1.006	1.291	1.190	1.243
0.435	1.054	1.129	0.595	1.353	1.451
2.371	2.219	2.234	2.968	2.931	2.951
0.255	0.372	0.471	0.105	0.373	0.502
1.061	0.668	0.781	1.225	0.809	0.958
-0.167	0.461	-----	-0.309	0.518	-----
1.830	1.441	1.386	2.383	1.831	1.758
0.979	0.582	0.797	1.391	0.657	0.940
0.732	0.517	0.636	0.885	0.590	0.747
0.093	0.475	0.315	-0.051	0.488	0.277
0.137	0.327	0.297	-0.015	0.296	0.257
0.140	0.634	0.662	0.073	0.724	0.760
1.317	1.569	1.591	1.703	2.038	2.067
0.945	0.675	0.697	1.010	0.813	0.842
0.070	0.156	0.233	-0.254	0.077	0.179
1.320	0.979	0.914	2.001	1.225	1.140
0.291	-0.117	-0.061	0.274	-0.299	-0.225
0.118	-0.087	-0.030	0.094	-0.244	-0.170

TABLE 4. Observed and predicted values of the hydrologic indices.

CONCLUSIONS

This investigation has been quite successful in appraising the dynamic aspects of flood expectancy in the southeastern New England region and developing a methodology by which such changes may be estimated in the future. A large number of variables, which encompass the morphometric, surficial and land use characteristics of a watershed, have been examined for this study. In addition to examining the general relation of these factors to basin hydrology, the work has focused principally on the hydrologic regime of peak discharge. The prevailing assumption that regulatory floodplains are static in areal extent has thus been challenged and quantitatively evaluated in this particular regional setting. Application of standard analytical methods used in flood frequency analysis leaves little doubt that dramatic increases in flood expectancy have accompanied urban growth in the study area. Continuation of this trend, regardless of the precise causal linkages, will have potentially grave consequences for all local and regional floodplain management strategies.

Appraisal of Independent Variables

The most important factors which appear to govern the observed hydrologic response include the degree of urbanization, the branching structure of the drainage network and the extent of pervious glacial deposits. Changes in urban land use have been characterized by an index which is sensitive to the actual areal extent of urban land use rather than the change alone, and its correlation of 0.76 with the hydrologic indices is highly significant at the one percent level. In addition, a modified E-ratio has been developed to discriminate between slight differences in branching patterns. This parameter has a correlation of -0.53 with the dependent variables which is significant at the five percent level. This indicates that elongation of the network tends to diminish the impact of urbanization on flood expectancy, a finding which is consistent with the observations of Strahler (1964) regarding an apparent relation between bifurcation ratio and hydrograph shape. The present work suggests that the "flashiness" of compact stream networks (having low values of R_b or E) is relevant to both common and extreme runoff events. Finally, a pervious index has been developed to describe the particular combination of surficial materials in each basin. This variable displays a correlation of 0.54 with the hydrologic indices, and is also significant at the five percent level. This indicates that basins with more extensive deposits of pervious sand and gravel may experience a greater hydrologic response to urbanization than less pervious basins. Furthermore, this correlation suggests that the dramatic increases in flood expectancy which have been documented result partly from abundant surficial materials with the capability to absorb and detain significant quantities of water, even during extreme rainfall and snowmelt. Unfortunately, this relationship between flood expectancy and perviousness is complicated by an association of the latter with urbanization.

The remaining variables identified in the early stages of the study do not appear to be useful for assessing the effects of urbanization. Drainage density and circularity ratio show very little correlation with the hydrologic indices, but this may result largely from their low variance in the sample. One would expect highly dissected landscapes to respond quite differently from terrain of low drainage density. Perhaps the very low values recorded in this study area are a relevant factor.

The lack of correlation between channel gradient, the extent of wetland area, and the dependent variables is rather puzzling. One might anticipate more pronounced changes in flood hydrographs for basins with lower gradients as the volume and rate of runoff from new urban areas increase. Other aspects of channel hydraulics, such as the abundance of ponds and the non-equilibrium profiles, may overshadow the large differences in gradient ranging from four to 50 feet per mile. In the case of wetlands, the apparent lack of a relationship could be explained in several ways. Changes in wetland area, which have not been incorporated in the model, may have greater hydrologic effect than differences in total areal extent among basins. Furthermore, the effect might be limited primarily to lower magnitude floods due to the close proximity of stream channels to wetlands. The floodwater storage capacity of these wetlands may be exceeded rapidly during extreme floods but not during more modest runoff events. This response contrasts sharply with the apparent situation in areas underlain by highly pervious glacial deposits.

Discussion of the Predictive Model

The selection of three principal variables has enabled a mathematical relationship to be established between urbanization and changes in flood expectancy. Four multiple regression equations have been derived, and two of them have been chosen for predicting the 50-year and 100-year hydrologic indices. These equations are highly significant, explain approximately three-quarters of the observed variation in the dependent variable, and reproduce the observations with reasonable accuracy. The development of this predictive model represents a departure, both in methodology and results, from numerous investigations of these phenomena conducted during the past 20 years. In particular, this study has focused on flood expectancy changes instead of the changes in flood hydrographs; it has relied on measured areal change in urban land use rather than indicators of basin "imperviousness"; and it has incorporated the modifying effects of geologic and morphometric properties which have been ignored or only alluded to by earlier studies. In addition, relatively natural watersheds have been used which span a wide range in terms of area, degree of urbanization and other relevant physical characteristics. Most of the earlier studies have dealt with small watersheds in fairly restricted urban settings.

The implications of these comparisons are summarized in Figure 20, where the impact of urbanization on floodplain area is indicated by the elevation of a low recurrence interval (10-year) and a high recurrence interval (100-year) event. In (A), the effect is shown only for the low magnitude flood, and this corresponds to the static floodplain assumption cited earlier. Floods of both recurrence intervals are affected in (B), but the increase diminishes for the higher magnitude flood. The apparent situation in the southeastern New England study area is presented in (C) where the impact extends over all recurrence intervals and seems to be greater for rarer events. These illustrations suggest that an assumption of general applicability regarding any particular model is unjustified and may be imprudent.

The predictive model is presently an unfinished product since it is likely to evolve as new data become available and as it is implemented in the study region. At the same time, its validity is not indisputable, although that is not readily tested because of its probabilistic nature. On the other hand, it is conceptually reasonable in most details,

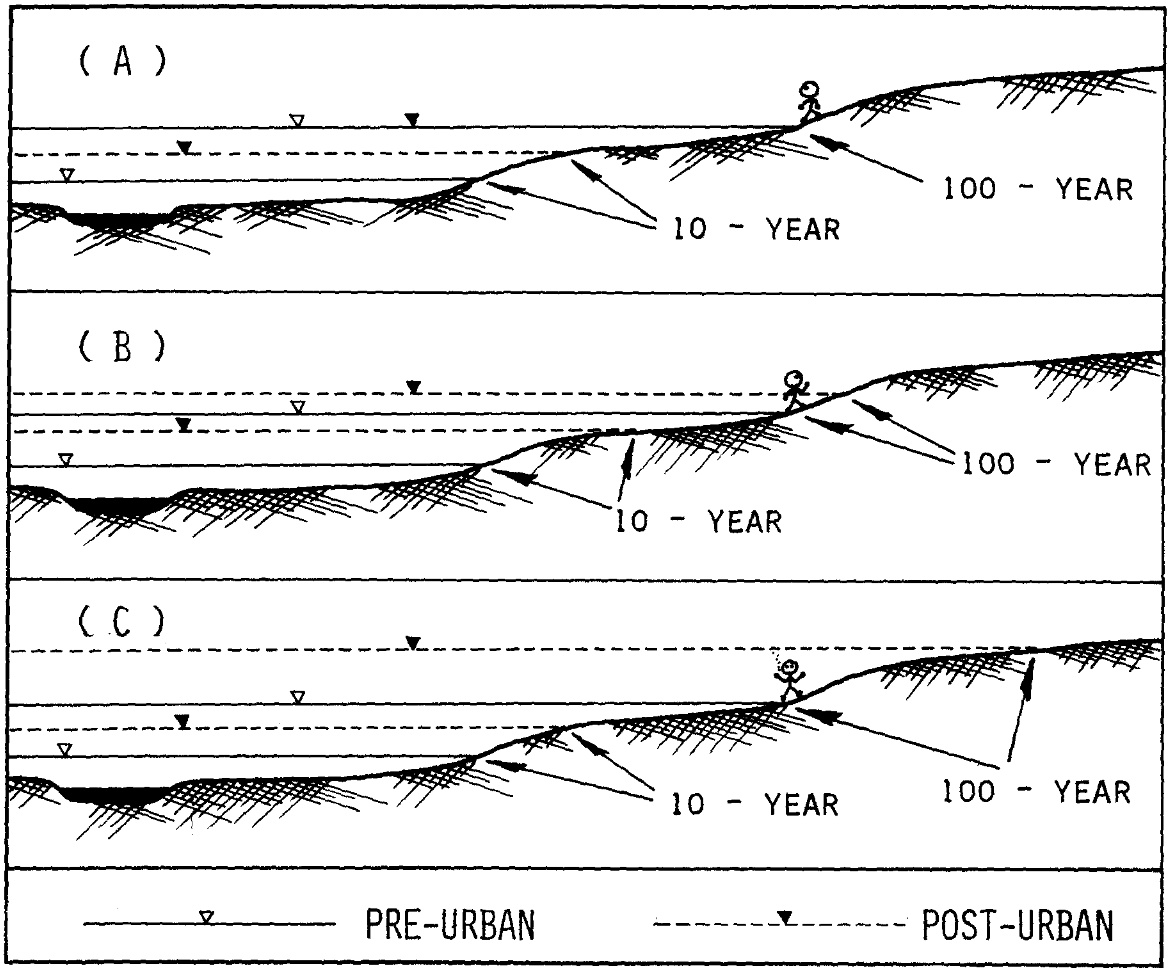


FIGURE 20. Alternative models of the impact of urbanization on flood expectancy.

including the modest but respectable degree of statistical explanation and precision of the estimates. In other words, a parsimonious model which perfectly described these complex relationships would be highly suspect. Although the method's applicability has not been thoroughly explored, earlier versions of the predictive equations (Doehring and others, 1975) have been, and are currently being applied by the METLAND research team in their environmental assessment program for the urbanizing fringe of the Boston Metropolitan Area. The more robust model described in this paper is expected to be a superior planning tool for future applications.

Implementation of the Methodology

A considerable proportion of the total effort expended during this project has involved a screening process to identify an efficient and reliable way for non-specialists to estimate the hydrologic impact of anticipated development. Despite the fairly simple appearance of the resulting equations, however, the implementation of this methodology is complicated by data collection requirements and the interpretation of results. The foregoing section related to data collection techniques is intended to be explanatory and should not be construed as a handbook. Many of the procedures which are briefly described and explained require considerable experience, and therefore should be performed only by or under the supervision of a qualified individual.

The valid application of this model is also restricted to a particular range of the independent variables. Specifically, regression analysis prohibits estimation of the dependent variable outside the range of any predictive variables used in the derivation of the equation. Instead of this strict approach, however, some flexibility is incorporated in the following suggested guidelines :

(1) total basin area between 10 and 200 square miles;
(2) total area of urban land less than 50 percent of basin area;
(3) modified E-ratio between 1.0 and 4.0
(4) area of pervious deposits less than 50 percent of basin area;
(5) area of wetlands less than 25 percent of basin area.

In addition, applications should be limited to the southeastern New England region, which is defined as those parts of Massachusetts, Rhode Island and Connecticut lying to the east of the Connecticut River. It should be noted that MacConnell's land use maps are not available at the present time for Connecticut, but other means of urban land use delineation could be substituted. Finally, it is recommended that the method be used only for watersheds above gaged stream sites with 10 or more years of record, although high quality, regionalized estimates may be employed with caution by experienced hydrologists.

Application of the proposed methodology can best be illustrated by means of a hypothetical example. Suppose we wish to estimate the increase in 100-year flood expectancy for a watershed in which 2500 acres are to be developed during the next three years. In addition, assume that a suitable gaging site exists with 18 years of available data, and that the following measurements have been made of the watershed above the stream gage.

Watershed area = 46.0 square miles
Current urban land = 7.4 square miles
Projected urban land = 11.3 square miles
Area of pervious deposits = 18.6 square miles
Area of wetlands = 3.6 square miles
Modified E-ratio = 2.12

Finally, assume that Log Pearson Type 3 analysis of the 18 year record yields the following estimates.

Mean annual flood = 225 c.f.s.
100-year flood = 583 c.f.s.

The calculations shown below result in an estimate of the hydrologic index for the 100-year recurrence interval.

$$I_{\Delta u} = \left[\frac{U_2 - U_1}{A - U_2}\right] \cdot \left[1 - \frac{U_1}{A}\right] \tag{6}$$

$$= (\,3.9\ /\ 34.7\,) \cdot (\,0.84\,) = \underline{\underline{0.094}}$$

$$I_p = P / (I + n \cdot S) = (18.6) / (23.8 + 7.2) \tag{7}$$

$$= \underline{\underline{0.6}}$$

$$I_{h.100} = -0.91 + 11.95\, I_{\Delta u} + 0.86\, I_p + \frac{1.01}{E} \tag{17}$$

$$= -0.91 + 1.12 + 0.52 + 0.48$$

$$= \underline{\underline{1.21}}$$

Let Q_{100} signify the present estimate of the 100-year flood, Q'_{100} signify the future 100-year flood, and Q_{ma} signify the mean annual flood. Then

$$I_{h \cdot 100} = \frac{Q'_{100} - Q_{100}}{Q_{ma}} \tag{13}$$

$$Q'_{100} = Q_{ma} \cdot I_{h \cdot 100} + Q_{100} = 225 \cdot (1.21) + 583$$

$$= \underline{\underline{855 \text{ c.f.s.}}}$$

On the basis of the 17 watersheds from which equation (17) was derived, this value is the best point estimate of the future discharge having an annual exceedance probability of one percent. However, such an estimate is subject to a certain amount of error which is related to the residual or unexplained variance in the regression model. Therefore, the calculation of the upper 95-percent confidence limit on this predicted value of the hydrologic index would be a prudent course of action. The actual computation of this confidence limit requires complex matrix manipulations related to the derivation of the regression equations, and it will not be described here. (See Draper & Smith, p.121 for details.) A secondary estimate of the future 100-year flood expectancy may be obtained by inserting this larger value of the hydrologic index in equation (13). There would then be only about a five percent chance of the actual 100-year discharge exceeding the resultant prediction as the proposed development proceeds. Finally, spatial assessment of this estimated discharge may be accomplished by using a stage-discharge rating curve or the slope-area method in conjunction with topography of the floodplains in the area under study. This entire process could, of course, be updated as further development is proposed or anticipated, implying that the management of flood hazard areas should be an ongoing activity.

Discussion

The preceding methodology is directed towards the goal of basing floodplain delineation on estimates of future, rather than present flood expectancy. It would be difficult to dispute the worthiness of such an objective in theory, but in practice, certain matters should be considered which involve the defensibility of the predictive algorithm on both hydrologic and legal grounds. Although no absolute answers can be given, the ensuing discussion will attempt to partially resolve these questions.

Most analytical methods in hydrology are inherently inaccurate to some extent. This is due in part to the imprecision of measurements and in part to the random error associated with stochastic processes; both of these elements are found in the practice of floodplain delineation. On the one hand, discharge measurements, stage-discharge rating curves, and the projection of flood elevations along a stream and laterally to a point on the ground are all approximate methods, even with state-of-the-art equipment. On the other hand, questions related to the selection of the proper mathematical distribution, treatment of outliers and mixed distributions, and the possible influence of ice jams or dam failure are not easily resolved. Consequently, "accurate" floodplain delineation is an unattainable abstraction that is only approached to varying degrees by modern techniques (Dingman & Platt, 1977). Nevertheless, these techniques are considered by most hydrologists to be the optimal basis for the management of flood hazard areas.

This study has identified an additional element of uncertainty - what is the dynamic nature of floodplains as urbanization occurs? The effect which has been detected and documented in these New England basins is certainly a real one, regardless of the exact cause. Documentation of alternative causes, such as a change in major atmospheric circulation, would be exceedingly difficult in the time span considered here. The model resulting from this investigation of flood frequency relations in southeastern New England is not, and cannot be any more exact than many other hydrologic tools. However, it currently represents the best available means of estimating flood expectancy changes which may result from various factors including urban growth. For the short term, in which action is imperative to forestall spiralling flood losses, the implementation of this methodology is certainly preferable to an educated guess (which is indefensible under any criteria) or the continued presumption of static floodplains in the face of convincing evidence to the contrary.

On the other hand, implementation will clearly penalize (or appear to penalize) some landowners, but any exercise of the police power does that to some extent. The method is neither arbitrary, capricious nor a "taking" to any greater extent than other accepted procedures of flood hazard evaluation. In other words, it simply extends the rationale of flood frequency analysis as the basis of sound land use decisions with respect to floodplains. Therefore, virtually all of the same philosophical, legal and technical arguments may be invoked. Perhaps the more hydraulically based methods, such as backwater curves, seem to have greater technical justification, but one must distinguish between precision (reproducibility of results) and accuracy (truth of results). The most elegant methods of engineering rely on approximations, estimates (including flood expectancy) and judgment, all of which render accuracy a rather elusive quality. This fact has not been an obstacle to the utilization of such methods, because if they are employed conscientiously and in the interests of public safety and welfare, they are generally acceptable on legal grounds. Equivalent reasoning surely may be applied to the predictive methodology developed in this study.

As noted earlier, application of the model is definitely restricted in geographic scope. There is frequently a strong inclination to discover and formalize universal relationships, which may be a worthwhile goal if pursued critically and honestly. However, it might be a serious mistake in efforts to relate urbanization and flood expectancy because of the danger of mixing very different populations. For example, the salient relationships which apply in the New England study area might be totally irrelevant to predicting the response in the semi-arid Southwest. Therefore, as this study is concluded, it is recommended that the geographic scope of the rationale employed here be increased by replication in other rapidly growing areas such as the Colorado Front Range and various parts of the West Coast. These studies would not only serve as valuable planning tools comparable to the work described here, but a comparative analysis might also suggest important conceptual and technical improvements.

SELECTED BIBLIOGRAPHY

Anderson, D.G., 1970, Effects of urban development on floods in northern Virginia. U.S. Geol. Survey, Water Supply Paper 2001-C, 22p.

Benson, M.A., 1962, Factors influencing the occurrence of floods in a humid region of diverse terrain. U.S. Geol. Survey, Water Supply Paper 1580-B, 62p.

Benson, M.A., 1965, Spurious correlation in hydraulics and hydrology. A.S.C.E. Proceedings, Hydraulics Division, HY4, pp.35-42.

Carter, R.W., 1961, Magnitude and frequency of floods in suburban areas. U.S. Geol. Survey, Professional Paper 424-B, pp.B9-11.

Cole, S.G., D.O. Doehring & J.G. Fabos, 1974, Assessment of flood hazard increases due to urbanization in the Greater Boston area. Geol. Soc. America, Abstracts with Programs, v.5, no.1, p.14.

Crippen, J.R., 1965, Changes in character of unit hydrographs, Sharon Creek, California after suburban development. U.S. Geol. Survey, Professional Paper 525-D, pp.D196-198.

Dingman, S.L. & R.H. Platt, 1977, Floodplain zoning: implications of hydrologic and legal uncertainty. Water Resources Research, v.13, no.3, pp.519-523.

Doehring, D.O., J.G. Fabos & M.E. Smith, 1975, Modelling the dynamic response of floodplains to urbanization in southeastern New England. Pub. No.53, Water Resources Research Center, University of Massachusetts, Amherst, 29p.

Doehring, D.O. & M.E. Smith, 1975, An algorithm for modelling changes in flood hazards associated with selected geologic and hydrologic characteristics. EOS, Transactions American Geophysical Union, v.56, no.12, p.982.

Draper, N.R. & H. Smith, 1966, Applied Regression Analysis. John Wiley & Sons, Inc., New York, 407p.

Espey, W.H., Jr. & D.E. Winslow, 1974, Urban flood frequency characteristics. A.S.C.E. Proceedings, Hydraulics Division, HY2, pp.279-294.

Fabos, J.G., 1973, Model for landscape resource assessment: Part I of the "Metropolitan Landscape Planning Model" (METLAND). Research Bull. No.602, College of Food and Natural Resources, Agricultural Experiment Station, University of Massachusetts, Amherst, 141p.

Gretener, P.E., 1967, The significance of the rare event in geology. A.A.P.G. Bulletin, v.51, no.11, pp.2197-2206.

Hammer, T.S., 1972, Stream channel enlargement due to urbanization. Water Resources Research, v.8, no.6, pp.1530-1540.

Jarvis, R.S., 1972, New measure of the topologic structure of dendritic drainage networks. Water Resources Research, v.8, no.5, pp.1265-1271.

Larson, J.S., 1973, A guide to important characteristics and values of freshwater wetlands in the Northeast. Pub. No.31, Water Resources Research Center, University of Massachusetts, Amherst, 35p.

Leopold, L.B., 1968, Hydrology for urban land planning - a guidebook on the hydrologic effects of urban land use. U.S. Geol. Survey, Circular No.554, 18p.

MacConnell, W.P., 1973, Massachusetts map down. Pub. No.97, Planning and Resource Series No.25, Cooperative Extension Service, University of Massachusetts, Amherst, 19p.

Miller, V.C., 1953, A quantitative geomorphic study of drainage basin characteristics in the Clinch Mountain area, Virginia and Tennessee. Tech. Rept. No.3, Office of Naval Research, Columbia University.

Mitchell, W.D., 1961, Effect of artificial storage on peak flow. U.S. Geol. Survey, Professional Paper 524-B, pp.B12-14.

Platt, R.H., 1976, The National Flood Insurance Program: some mid-stream perspectives. Jour. American Inst. of Planners, v.42, no.3, pp.303-313.

Riggs, H.C., 1965, Effect of land use on the low flow of streams in Rappahannock County, Virginia. U.S. Geol. Survey, Professional Paper 525-C, pp.C196-198.

Savini, J. & J.C. Kammerer, 1961, Urban growth and the water regimen. U.S. Geol. Survey, Water Supply Paper 1591-A, 41p.

Smart, J.S., 1970, Use of topologic information in processing data for channel networks. Water Resources Research, v.6, no.3, pp.932-936.

Smith, M.E., D.O. Doehring & J.G. Fabos, 1974, Process-response modelling of flood hazards in 28 New England basins. Geol. Soc. America, Abstracts with Programs, v.7, no.1, p.120.

Smith, M.E. & D.O. Doehring, 1977, A planning algorithm to assess the impact of urban growth on flood expectancy. Proceedings Volume, Symposium on the Effects of Urbanization on the Hydrological Regime and on Water Quality, Amsterdam, UNESCO, 12p.

Strahler, A.N., 1964, Quantitative geomorphology of drainage basins and channel networks. in, Handbook of Applied Hydrology, V.T. Chow, McGraw-Hill Book Co., New York.

Waananen, A.O., 1961, Hydrologic effects of urban growth - some characteristics of urban runoff. U.S. Geol. Survey, Professional Paper 424-C, pp.C353-356.

Water Resources Council, 1967, A uniform technique for determining flood flow frequencies. Bulletin No.15, Washington, D.C., 15p.

Wilson, K.V., 1967, A preliminary report on the effect of urbanization on floods in Jackson, Mississippi. U.S. Geol. Survey, Professional Paper 575-D, pp.D259-261.

GLOSSARY

Coefficient of determination - the proportion (usually a percentage) of the total variance in a dependent variable which is accounted for by a linear regression equation; it is calculated as

$$\frac{\text{(Sum of Squares Due to Regression)}}{\text{(Total Corrected Sum of Squares)}}$$

which is independent of degrees of freedom, and may thus be inflated by increasing the number of independent variables.

Degrees of freedom - number of independent variables needed to completely specify a system; the variance of N observations has N-1 degrees of freedom, while a least squares estimate in regression has only 1.

Design frequency - a specified recurrence interval which corresponds to a magnitude of discharge that is relevant to some engineering or planning activity; a specified norm of flood expectancy.

Drainage composition - the spatial pattern of a stream network which describes the manner in which individual stream segments are joined.

Exceedance probability - the chance that discharge, equal to or greater than some particular value, will occur in a specified time span, usually one year; it is important to recognize that the 100-year flood is commonly misconstrued as a specific event which may occur once every 100 years; in fact, it may be *exceeded* (slightly or greatly) every year with a one percent chance.

Flood expectancy - a stream discharge having a specified exceedance probability or recurrence interval (e.g., the 100-year flood expectancy).

Flood frequency analysis - statistical and graphic techniques which estimate the relation between magnitude of discharge and exceedance probability.

Glacial till - sediment deposited at the base of a glacier or during ablation of the ice; composed of a wide range of particle sizes from clay and silt to gravel and boulders; generally much coarser in New England deposits than in areas of the Midwest.

Glaciofluvial - pertaining to streams emanating from glacial ice or to the deposits laid down by such streams; generally, such deposits are considerably coarser than adjacent modern stream deposits.

Kame terrace - a relatively flat-topped deposit of glacial sand & gravel along valley margins; remnant of a depositional valley surface which was constructed in contact with glacial ice.

Lag time - amount of time between peak discharge and the center of mass of excess rainfall.

Least squares criterion - a statistical property in which $\Sigma(X - \bar{X})^2$ is minimized; estimates based on this criterion in regression are unbiased and have minimum variance of any possible linear estimate.

Monthly maximum average daily discharge - the largest value of average discharge for each 24-hour period occurring in a month, generally less than the true maximum peak discharge.

Morphometry - the quantitative measurement and description of landscapes in terms of slope elements, the drainage system and other factors.

Partial F-ratio - a ratio of two mean squares (variances) in which the numerator corresponds to one specific effect with one or more other effects held fixed (or accounted for), and the denominator is a residual or unexplained component; it evaluates the effect in terms of the contribution it makes beyond the total contribution of all other effects.

Peak discharge - momentary maximum value of discharge which is measured independently of the continuous water stage recorder.

Recurrence interval - the average length of time in which a specified discharge will be equalled or exceeded once over a much longer time span; the reciprocal of exceedance probability.

Regionalization - a technique whereby streamflow at an ungaged site is estimated by establishing a relationship with other nearby gaged sites; one equation might estimate the mean annual discharge from watershed area, elevation and mean annual precipitation.

Standard error of estimate - the standard deviation of estimated values about the calculated regression line, which is assumed equal to the square root of the residual mean square; may be expressed as a percentage of the mean observed value of the dependent variable; indicates the expected precision of estimated values of the dependent variable.

Stream order (Strahler) - designation of relative position of stream segments in a drainage network by an integer series; first order refers to a source, second order is fed by at least two first order segments, third order is fed by at least two second order segments, and so forth.

Unit areal discharge - streamflow in cubic feet per second per square mile of watershed above the gaging site (cfsm); also referred to as unit discharge (csm).

Valley train - a narrow body of glaciofluvial outwash confined within a stream valley.

APPENDIX A: Computer Program Listings

1. APL defined function *STRAHLER*

```
∇ STRAHLER A;M;N;P;U

[1] U←⍳0
[2] U←U,+/A←A>0
[3] A←(((N≠M←(A=0)/A+1⌽A)∨N=0)/N←(A=0)/A+(1⌽A)+2⌽A)>1
[4] →(1<⍴A)/2
[5] 'ORDER:   ';5 0⍕ ⍳N←1+⍴U
[6] 'NUMBER: ';5 0⍕ *U←⍟U,1
[7] 'RB =      ';8 2⍕ *-((+/U×⍳N)-(+/U)×A÷N)÷((+/(⍳N)*2)-((A←+/⍳N)*2)÷N)

∇
```

A is a coded network vector which is recast into a binary string for manipulation in the algorithm. The result takes the form

```
ORDER:    1     2     3     4
NUMBER: 87     21     5     1
RB =    4.41
```

2. APL defined function *ERATIO*

```
∇ ERATIO A;I;M;Z;Q;MI;ME;NP;L;N;D;E;M1;H

 [1] N←(+/A≤0)+H←+/A>I←MI←ME←NP←L←M←0
 [2] I←I+1
 [3] →(A[I]>0)/6
 [4] M←M,M1←+/A[I+⍳((+\A[I+⍳N-I]>0)-+\A[I+⍳N-I]≤0)⍳2]>0
 [5] →(I=1+Z←0)/10
 [6] →(0=+/Z←(D←+\Q≤0)>E←+\0<Q←⌽(I-1)↑A)/8
 [7] →(0≠+/Z←D>E←E+((Z-1)⍴0),(I-Z←(D>E)⍳1)⍴1)/7
 [8] Z←+/(Z×Q=0)+2×(Z←D=E)×Q<0
 [9] →(A[I]>0)/11
[10] →((MI←MI+M1×((Z+1)×I≠1)+I=1)≠0)/12
[11] M←M,(1+ME-ME←ME+(Z+1)×1+A[I+1]>0),(A[I+1]>0)/1
[12] NP←NP+1<|A[I]-0.5
[13] L←L+(Z+1-(A[I]=¯1)÷2)×1<|A[I]-0.5
[14] →(N>I←I+(A[I+1]>0)∧A[I]>0)/2
[15] 'ERATIO:     ';10 4⍕ MI÷ME;  'MODIFIED:  ';10 4⍕ MI÷ME×10⍟+/A>0
[16] 'NO. PONDS: ';10 0⍕ NP
[17] 'DISTANCE:  ';10 4⍕ L÷NP
[18] 'MEAN PATH: ';10 4⍕ ME÷H
[19] 'CENTROID:  ';10 0⍕ Z⍳⌊/Z←|1↓M-H÷2;' LINK'

∇
```

The result takes the form

```
ERATIO:          12.4426  MODIFIED:         5.9143
NO. PONDS:       40
DISTANCE:        20.3750    (average link distance of ponds)
MEAN PATH:       24.5275    (average link distance from sources)
CENTROID:       127  LINK   (network point which divides total sources)
```

3. Graphic Flood Frequency Analysis

This group of four program files is designed for the HP 9830A with peripheral printer and plotter. The cassette tape also contains a series of data files for Log Pearson Type III coefficients and for hydrologic records. The package is self-contained as well as self-explanatory.

File 1

```
10  COM TI[100],M,H5
11  DISP "EXPLANATION=1 ";
12  INPUT A
13  IF A≠1 THEN 25
14  PRINT "GRAPHIC FLOOD FREQUENCY ANALYSIS: PROGRAM & DATA FILES"
15  PRINT "FILE 1       DATA ACCESS AND STORAGE ON THIS TAPE."
16  PRINT "FILES 2&3    PERFORM LOG PEARSON TYPE III AND/OR GUMBEL
17  PRINT "             ANALYSES.  PLOTTING PAPER IS NEEDED."
18  PRINT "FILE 4       OPTIONAL OUTLIER ROUTINE:  FREQUENCY CURVES"
19  PRINT "             DRAWN FOR ALL BUT LARGEST OBSERVATION; OUT-"
20  PRINT "             LIER PLOTTED & PROJECTED TO CURVES.  PROBA-"
21  PRINT "             BILITY OF OCCURRENCE IN N YEARS CALCULATED"
22  PRINT "             FROM THE POISSON DISTRIBUTION."
23  PRINT "FILES 50-99 ARE DATA FILES; MISSING DATA CODED BY (-)."
24  PRINT
25  DISP "LOAD DATA FILE = 1 ";
26  INPUT A
27  IF A≠1 THEN 43
28  DISP "FILE NUMBER ";
29  INPUT Q
30  LOAD DATA Q,T
31  DISP "ADD DATA TO FILE = 1 ";
32  INPUT A
33  IF A≠1 THEN 63
34  DISP "NUMBER ADDED ";
35  INPUT N
36  A=T[1]-INT(T[1]/100)*100
37  FOR I=A+2 TO A+N+1
38  DISP 1900+INT(T[1]/100)+I-A-1;
39  INPUT T[I]
40  NEXT I
41  T[1]=T[1]+N*101
42  GOTO 59
43  DISP "YEAR OF MOST RECENT DATUM ";
44  INPUT Y
45  DISP "NUMBER OF ANNUAL MAXIMA ";
46  INPUT T[1]
47  DISP "ENTER MAXIMA - EARLY TO RECENT ";
48  FOR I=2 TO T[1]+1
49  IF I=2 THEN 51
50  DISP Y-T[1]+I-1;
51  INPUT T[I]
```

```
52  NEXT I
53  T[1]=T[1]+100*(Y-1900)
54  DISP "STORE NEW FILE = 1 ";
55  INPUT A
56  IF A≠1 THEN 63
57  DISP "FILE NUMBER ";
58  INPUT Q
59  STORE DATA Q,T
60  DISP "ANALYZE FILE IN MEMORY = 1 ";
61  INPUT A
62  IF A≠1 THEN 25
63  N1=N=T[1]-INT(T[1]/100)*100
64  FOR I=2 TO N1+1
65  IF T[I]>0 THEN 67
66  N=N-1
67  NEXT I
68  FORMAT "FILE",F4.0,":  RECORD FROM ",F5.0," TO",F5.0,"; ",
    F4.0," MAXIMA",/
69  WRITE (15,68)Q,INT(T[1]/100)+1901-N1,INT(T[1]/100)+1900,N
70  M=Z=R=Q=S1=S2=S3=S4=S5=M1=M2=H5=0
71  FOR I=2 TO N1+1
72  IF T[I]<0 THEN 74
73  M=M+T[I]/N
74  NEXT I
75  DISP "EVALUATE OUTLIERS = 1 ";
76  INPUT A
77  IF A=1 THEN 79
78  LOAD 2
79  DIM SI[100],DI[15],BI[15]
80  LINK 4
81  END
```

File 2

```
10  COM TI[100],M,H5
11  DIM SI[100],DI[15],BI[15]
12  SCALE -2,6.5,0,6.5
13  IF H5≠0 THEN 28
14  IF M>150 THEN 17
15  Z=100
16  GOTO 22
17  IF M>350 THEN 20
18  Z=200
19  GOTO 22
20  IF M>750 THEN 22
21  Z=500
22  PRINT "VERTICAL SCALE INTERVAL =";Z
23  PRINT
24  FOR I=Z/M TO 6.5 STEP Z/M
25  PLOT -2,I,-2
26  IPLOT 0.1,0,-1
27  NEXT I
```

```
28  M1=M2=S1=S2=S3=S4=S5=A=R=Z=Q=P=0
29  DISP "SELECT PART OF RECORD BY YEARS ";
30  INPUT B,C
31  PRINT "   ANALYSIS   OF   ";B;"TO   ";C
32  PRINT
33  B=B+T[1]-INT(T[1]/100)*101-1900
34  C=C+T[1]-INT(T[1]/100)*101-1900
35  FOR I=B+1 TO C+1
36  IF T[I]>0 THEN 38
37  P=P+1
38  NEXT I
39  N=C-B-P+1
40  FOR I=B+1 TO C+1
41  IF T[I]<=0 THEN 49
42  Z=Z+1
43  S[Z]=T[I]
44  M1=M1+T[I]/N
45  M2=M2+LGT(T[I])/N
46  R=R-(LOG(-LOG(Z/(N+1))))/N
47  IF A>T[I] THEN 49
48  A=T[I]
49  NEXT I
50  FOR I=1 TO Z
51  B=S[1]
52  FOR J=2 TO Z
53  IF B<S[J] THEN 55
54  B=S[J]
55  NEXT J
56  FOR K=1 TO Z
57  IF B=S[K] THEN 59
58  NEXT K
59  S[K]=A+1
60  S1=S1+(B-M1)↑2
61  S2=S2+(-LOG(-LOG(I/(N+1)))-R)↑2
62  S3=S3+(LGT(B)-M2)↑2
63  S4=S4+(LGT(B)-M2)↑2
64  S5=S5+(B-M1)*(-LOG(-LOG(I/(N+1)))-R)
65  PLOT -LOG(-LOG(I/(N+1)))+H5/33,B/M+(4-H5)/150,-2
66  IPLOT -H5/17,(H5-4)/75,-1
67  IPLOT (H5-1)/25,(H5+2)/75,-2
68  IPLOT (4-H5)/50,-H5/25,-1
69  NEXT I
70  DISP "LPT3 - 1;  GMBL - 2;  BOTH - 3 ";
71  INPUT I
72  LINK 3
73  END
```

File 3

```
10  DATA 40,200,400,600,700,800,900,950,960,975,980,990,995
11  IF I=2 THEN 28
12  S3=SQR(S3/(Z-1))
```

```
13  S4=(Z*S4)/((Z-1)*(Z-2)*S3↑3)
14  LOAD DATA INT(S4*10)+27,D
15  LOAD DATA 27-INT(-S4*10),B
16  FOR J=1 TO 13
17  Y=10↑(M2+S3*(D[J]/1000+(S4-INT(S4*10)/10)*(B[J]-D[J])/100))/M
18  IF J≠11 THEN 20
19  Q3=Y*M
20  IF J≠12 THEN 22
21  Q2=Y*M
22  READ G
23  PLOT -LOG(-LOG(G/1000)),Y,-2
24  NEXT J
25  PEN
26  RESTORE
27  IF I=1 THEN 52
28  S=S5/S2
29  X1=R-M1/S
30  X2=R+(6.5*M-M1)/S
31  IF X1>=-1.17 THEN 33
32  X1=-1.17
33  IF X2<=5.296 THEN 35
34  X2=5.296
35  PLOT X1,(M1+S*(X1-R))/M,-2
36  PLOT X2,(M1+S*(X2-R))/M,-1
37  DISP "95% CONF. CURVES ON GMBL = 1 ";
38  INPUT A
39  IF A≠1 THEN 51
40  T=10↑(0.2924+0.5192/(Z-2)+0.3598/(Z-2) 2-0.06744/(Z-2) 3)
41  FOR J=-10 TO 56 STEP 3
42  Y1=M1+S*(J/10-R)
43  Y2=T*SQR((S1-S*S5)/(Z-2)*(1/Z+(J/10-R)↑2/S2))
44  PLOT J/10,(Y1+Y2)/M,-2
45  PEN
46  PLOT J/10,(Y1-Y2)/M,-2
47  PEN
48  NEXT J
49  FORMAT F10.4
50  FORMAT F10.0
51  IF I=2 THEN 55
52  WRITE (15,49)"LOG MEAN ........... ",M2
53  WRITE (15,49)"LOG STND.DEV. ...... ",S3
54  WRITE (15,49)"LOG SKEWNESS ....... ",S4
55  WRITE (15,50)"MEAN ANNUAL FLOOD .. ",M1
56  IF I=2 THEN 60
57  WRITE (15,50)"Q-50, LPT3 ......... ",Q3
58  WRITE (15,50)"Q-100, LPT3 ........ ",Q2
59  IF I=1 THEN 63
60  WRITE (15,50)"Q-50, GMBL ......... ",M1+S*(3.902-R)
61  WRITE (15,50)"Q-100, GMBL ........ ",M1+S*(4.6-R)
62  PRINT "GUMBEL LINE:  Q =";INT(M1-S*R+0.5);"+ ";INT(S+0.5);"RV"
63  PRINT
64  H5=1
65  DISP "CONTINUE = 1 ";
```

```
66  INPUT Q
67  IF Q≠1 THEN 69
68  LOAD 2
69  LOAD 1
70  END
```

File 4

```
10  SCALE -2,10,0,10
11  DATA 40,200,400,600,700,800,900,950,960,975,980,990,995,
    999,999.9
12  DISP "INSERT OUTLIER PLOTTING PAPER ";
13  INPUT Q
14  FOR I=2 TO N1+1
15  IF A>T[I] THEN 17
16  A=T[I]
17  NEXT I
18  FOR I=2 TO N1+1
19  IF T[I]<=0 OR T[I]=A THEN 25
20  Z=Z+1
21  S[Z]=T[I]
22  M1=M1+T[I]/(N-1)
23  M2=M2+LGT(T[I])/(N-1)
24  R=R-(LOG(-LOG(Z/N)))/(N-1)
25  NEXT I
26  FOR I=1 TO Z
27  B=S[1]
28  FOR J=2 TO Z
29  IF B<S[J] THEN 31
30  B=S[J]
31  NEXT J
32  FOR K=1 TO Z
33  IF B=S[K] THEN 35
34  NEXT K
35  S[K]=A+1
36  S1=S1+(B-M1)↑2
37  S2=S2+(-LOG(-LOG(I/(Z+1)))-R)↑2
38  S3=S3+(LGT(B)-M2)↑2
39  S4=S4+(LGT(B)-M2)↑3
40  S5=S5+(B-M1)*(-LOG(-LOG(I/(Z+1)))-R)
41  PLOT -LOG(-LOG(I/(Z+1))),B/M,-2
42  PEN
43  NEXT I
44  S3=SQR(S3/(Z-1))
45  S4=(Z*S4)/((Z-1)*(Z-2)*S3↑3)
46  LOAD DATA INT(S4*10)+27,D
47  LOAD DATA 27-INT(-S4*10),B
48  FOR J=1 TO 15
49  Y=10↑(M2+S3*(D[J]/1000+(S4-INT(S4*10)/10)*(B[J]-D[J])/100))/M
50  READ G
51  PLOT -LOG(-LOG(G/1000)),Y,-2
52  NEXT J
```

```
53  PEN
54  RESTORE
55  S=S5/S2
56  PLOT -1,(M1-S*(1+R))/M,-2
57  PLOT 10,(M1+S*(10-R))/M,-1
58  PLOT -LOG(-LOG(N/(N+1))),A/M+0.04,-2
59  IPLOT 0,-0.08,-1
60  IPLOT -0.04,0.04,-2
61  PLOT 10,A/M,-1
62  PLOT 10,10,1
63  FORMAT 34X,"LPT3",10X,"GUMBEL"
64  FORMAT "Q-MAX",F7.0,4X,"RETURN PERIOD:  ",F6.0,10X,F6.0
65  FORMAT 11X,"PROB. IN",F4.0," YEARS:  ",F6.4,10X,F6.4,/
66  DISP "REDUCED VARIATES: LPT3, GMBL ";
67  INPUT R1,R2
68  R1=1/(1-EXP(-EXP(-R1)))
69  R2=1/(1-EXP(-EXP(-R2)))
70  WRITE (15,63)
71  WRITE (15,64)A,R1,R2
72  WRITE (15,65)N,1-EXP(-N/R1),1-EXP(-N/R2)
73  DISP "REMOVE OUTLIER FROM RECORD = 1 ";
74  INPUT Q
75  IF Q=1 THEN 79
76  DISP "CHANGE PLOTTING PAPER ";
77  INPUT Q
78  LOAD 2
79  FOR I=2 TO N1+1
80  IF T[I]=A THEN 82
81  NEXT I
82  PRINT I-1;" -";T[I];"  REMOVED"
83  PRINT
84  DISP "SUBSTITUTE ";
85  INPUT T[I]
86  M=M+(T[I]-A)/N
87  N=N-(T[I]≠0)
88  Z=R=Q=S1=S2=S3=S4=S5=M1=M2=A=0
89  DISP "EVALUATE NEXT OUTLIER = 1 ";
90  INPUT Q
91  IF Q=1 THEN 14
92  LOAD 2
93  END
```

4. Stepwise Multiple Regression Analysis

This group of three program files is designed for the HP 9830A with peripheral printer and plotter. The cassette tape also contains a data input program and a series of data vector files.

Data are first input from the keyboard or from stored files. Provision is made to select all or only some observations. Next, an augmented correlation matrix is formed as illustrated below.

Pearson Correlation Matrix in which last row and column correspond to the dependent variable.		Positive Identity Matrix Zero Row Vector
Negative Identity Matrix	Zero Column Vector	Zero Matrix

This matrix is modified by simple row and column operations as variables are entered in a stepwise manner. Order of entry is determined by the highest partial correlation of variables not yet entered, and the decision is based on an F-ratio. If two or more variables are in the regression, partial F-ratios are also given to decide on the omission of any redundant variables. All of these values, as well as various sums of squares and regression coefficients, are readily extracted from simple arithmetic combinations of matrix entries. (Draper & Smith, 1966, pp.178-195)

Program output at each stage includes (1) an analysis of variance table, (2) least squares estimates, (3) percent explanation, and (4) standard error of estimate. Final output tabulates the observed and predicted responses, the residual ($Y-\hat{Y}$), and the standard normal deviate of this residual; these values are also plotted.

File 1

```
10  DIM TS[6],VI[6],SS[6],CS[11,11],ES[11,11],LS[6],DS[6,30],
    US[31],OI[30]
11  FOR I=1 TO 6
12  T[I]=L[I]=0
13  V[I]=I
14  NEXT I
15  DISP "DATA FILE = 1 ";
16  INPUT Q
17  DISP "NO. INDEP. VARS. (5 MAX) ";
```

```
18  INPUT P
19  IF Q≠1 THEN 48
20  FOR I=1 TO P
21  DISP "VAR";I;"FILE NO. ";
22  INPUT X
23  LOAD DATA X,U
24  IF I>1 THEN 35
25  DISP "SELECT ( ALL =";U[1];") ";
26  INPUT N
27  N1=N-1
28  FOR J=1 TO N
29  IF N=U[1] THEN 33
30  DISP J;
31  INPUT O[J]
32  GOTO 34
33  O[J]=J
34  NEXT J
35  FOR J=1 TO N
36  D[I,J]=U[O[J]+1]
37  T[I]=T[I]+U[O[J]+1]/N
38  NEXT J
39  NEXT I
40  DISP "RESPONSE: FILE NO. ";
41  INPUT X
42  LOAD DATA X,U
43  FOR J=1 TO N
44  D[P+1,J]=U[O[J]+1]
45  T[P+1]=T[P+1]+U[O[J]+1]/N
46  NEXT J
47  GOTO 63
48  DISP "NO. OF OBSERV. (30 MAX) ";
49  INPUT N
50  N1=N-1
51  FOR I=1 TO P
52  FOR J=1 TO N
53  DISP "VAR";I;J;
54  INPUT D[I,J]
55  T[I]=T[I]+D[I,J]/N
56  NEXT J
57  NEXT I
58  FOR J=1 TO N
59  DISP "RESPONSE";J;
60  INPUT D[P+1,J]
61  T[P+1]=T[P+1]+D[P+1,J]/N
62  NEXT J
63  M=1+2*P
64  FOR I=1 TO M
65  FOR J=I TO M
66  C[I,J]=C[J,I]=0
67  NEXT J
68  NEXT I
69  FOR I=1 TO P+1
70  FOR J=I TO P+1
```

```
71  FOR K=1 TO N
72  C[I,J]=C[I,J]+D[I,K]*D[J,K]
73  NEXT K
74  C[I,J]=C[I,J]-N*T[I]*T[J]
75  NEXT J
76  S[I]=C[I,I]
77  NEXT I
78  FOR I=1 TO P
79  C[I,I]=1
80  FOR J=I+1 TO P+1
81  C[I,J]=C[J,I]=C[I,J]/SQR(S[I]*S[J])
82  NEXT J
83  NEXT I
84  C[P+1,P+1]=1
85  FOR I=P+2 TO M
86  C[I-P-1,I]=1
87  C[I,I-P-1]=-1
88  NEXT I
89  LINK 2
90  END
```

File 2

```
10  FOR I=1 TO P
11  IF V[I]≠0 THEN 13
12  NEXT I
13  A=C[I,P+1]*C[P+1,I]/C[I,I]
14  Q=I
15  IF I=P THEN 24
16  IF I>P THEN 29
17  FOR J=I+1 TO P
18  IF V[J]=0 THEN 23
19  B=C[J,P+1]*C[P+1,J]/C[J,J]
20  IF A>=B THEN 23
21  A=B
22  Q=J
23  NEXT J
24  V[Q]=0
25  F=(N1-1)*A/(C[P+1,P+1]-A)
26  DISP Q;INT(F*1000)/1000;N1-1;"ENTER=1 ";
27  INPUT B
28  IF B=1 THEN 30
29  LINK 3
30  FOR I=1 TO M
31  FOR J=1 TO M
32  IF I≠Q THEN 35
33  E[I,J]=C[I,J]/C[Q,Q]
34  GOTO 36
35  E[I,J]=C[I,J]-C[I,Q]*C[Q,J]/C[Q,Q]
36  NEXT J
37  NEXT I
38  FOR I=1 TO M
```

```
39  FOR J=1 TO M
40  C[I,J]=E[I,J]
41  NEXT J
42  NEXT I
43  N1=N1-1
44  IF N1=N-2 THEN 71
45  FORMAT "VAR.",F4.0,";  PARTIAL F =",F8.3
46  FOR I=1 TO P
47  IF V[I]≠0 THEN 49
48  WRITE (15,45)I,N1*C[I,P+1]↑2/C[P+1,P+1]/C[I+P+1,I+P+1]
49  NEXT I
50  PRINT
51  DISP "DELETE VARIABLE (0=NONE) ";
52  INPUT B
53  IF B=0 THEN 71
54  N1=N1+1
55  V[B]=B
56  L[B]=0
57  FOR I=1 TO M
58  FOR J=1 TO M
59  IF I≠B THEN 62
60  E[I,J]=C[I,J]/C[B+P+1,B+P+1]
61  GOTO 63
62  E[I,J]=C[I,J]-C[I,B+P+1]*C[B+P+1,J]/C[B+P+1,B+P+1]
63  NEXT J
64  NEXT I
65  FOR I=1 TO M
66  FOR J=1 TO M
67  C[I,J]=E[I,J]
68  NEXT J
69  NEXT I
70  IF N1≠N-2 THEN 46
71  FORMAT F4.0,2F12.4,F8.3
72  FORMAT "   R-SQR:",F6.3,"   STND.ERROR EST.:",2F8.3
73  FORMAT F2.0,F13.4,F14.4
74  PRINT "SOURCE     DF       SS          MS         F"
75  WRITE (15,71)"TOTAL     ",N-1,S[P+1]
76  G=C[P+1,P+1]*S[P+1]
77  B=S[P+1]-G
78  WRITE (15,71)"REGR.     ",N-1-N1,B,B/(N-1-N1),B*N1/G/(N-1-N1)
79  WRITE (15,71)"RESID.    ",N1,G,G/N1
80  PRINT "    LEAST SQUARES ESTIMATES     STND.ERROR"
81  Z=0
82  FOR I=1 TO P
83  IF V[I]≠0 THEN 88
84  B=C[I,P+1]*SQR(S[P+1]/S[I])
85  Z=Z+B*T[I]
86  L[I]=B
87  WRITE (15,73)"      B",I,"   =",B,SQR(G/N1*C[I+P+1,I+P+1]/S[I])
88  NEXT I
89  WRITE (15,73)"      B 0   =",T[P+1]-Z
90  WRITE (15,72)1-C[P+1,P+1],SQR(G/N1),SQR(G/N1)/T[P+1]
91  PRINT
```

```
92  L[6]=T[P+1]-Z
93  GOTO 1
94  END
```

File 3

```
10  FORMAT 4F11.3        (modify according to specific needs)
11  SCALE 0,2.2,0,1
12  PRINT "   OBSERVED  PREDICTED   RESIDUAL    Z-VALUE"
13  Y1=Y2=R1=0
14  Y3=Y4=R2=100000
15  FOR I=1 TO N
16  D[1,I]=D[1,I]*L[1]+L[6]
17  FOR J=2 TO P
18  D[1,I]=D[1,I]+D[J,I]*L[J]
19  NEXT J
20  A=D[P+1,I]
21  B=D[1,I]
22  WRITE (15,1)A,B,B-A,(B-A)/SQR(G/N1)
23  IF Y1>B THEN 25
24  Y1=B
25  IF Y2>A THEN 27
26  Y2=A
27  IF Y3<B THEN 29
28  Y3=B
29  IF Y4<A THEN 31
30  Y4=A
31  IF R1>B-A THEN 33
32  R1=B-A
33  IF R2<B-A THEN 35
34  R2=B-A
35  NEXT I
36  PRINT
37  Y1=Y1*(Y1 Y2)+Y2*(Y1 Y2)
38  Y3=Y3*(Y3 Y4)+Y4*(Y3 Y4)
39  Y1=Y1+(Y1-Y3)/10
40  Y3=Y3-(Y1-Y3)/11
41  R1=R1+(R1-R2)/10
42  R2=R2-(R1-R2)/11
43  XAXIS 0,1,0,1
44  YAXIS 1.2
45  XAXIS ABS(R2/(R1-R2)),1,2.2,1.2
46  YAXIS 0
47  FOR I=1 TO N
48  B=D[1,I]-D[P+1,I]
49  Y2=(D[1,I]-Y3)/(Y1-Y3)
50  Y4=(D[P+1,I]-Y3)/(Y1-Y3)
51  PLOT Y4,Y2,-2
52  PEN
53  CPLOT 0.3,-0.3
54  LABEL (*,2.2,1.8,0,10/22)O[I]
55  IPLOT 0,0,1
```

```
56  IPLOT 1.2,(B-R2)/(R1-R2)-Y2+0.01,-2
57  IPLOT 0,-0.02,-1
58  IPLOT -0.01,0.01,-2
59  IPLOT 0.02,0,-1
60  IPLOT B/(Y1-Y3),0.01,-2
61  IPLOT -0.02,-0.02,-1
62  IPLOT 0,0.02,-2
63  IPLOT 0.02,-0.02,-1
64  NEXT I
65  PLOT 2.2,1,0
66  END
```

File 4

```
10  DIM US[31]
11  DISP "NEW FILE = 1 ";
12  INPUT Q
13  IF Q≠1 THEN 25
14  DISP "NO. OF OBSERV. (30 MAX) ";
15  INPUT U[1]
16  IF U[1]=0 THEN 38
17  FOR I=2 TO U[1]+1
18  DISP I-1;
19  INPUT U[I]
20  NEXT I
21  DISP "STORE FILE NO. ";
22  INPUT Q
23  STORE DATA Q,U
24  GOTO 2
25  DISP "LOAD FILE NO. ";
26  INPUT Q
27  LOAD DATA Q,U
28  A=U[1]-1
29  DISP "INDEX & VALUE (STOP=0,0) ";
30  INPUT I,V
31  IF I=0 OR I>30 THEN 36
32  IF A>I THEN 34
33  A=I
34  U[I+1]=V
35  GOTO 29
36  U[1]=A+1
37  GOTO 23
38  END
```

APPENDIX B: Flood Frequency Analyses of Study Watersheds

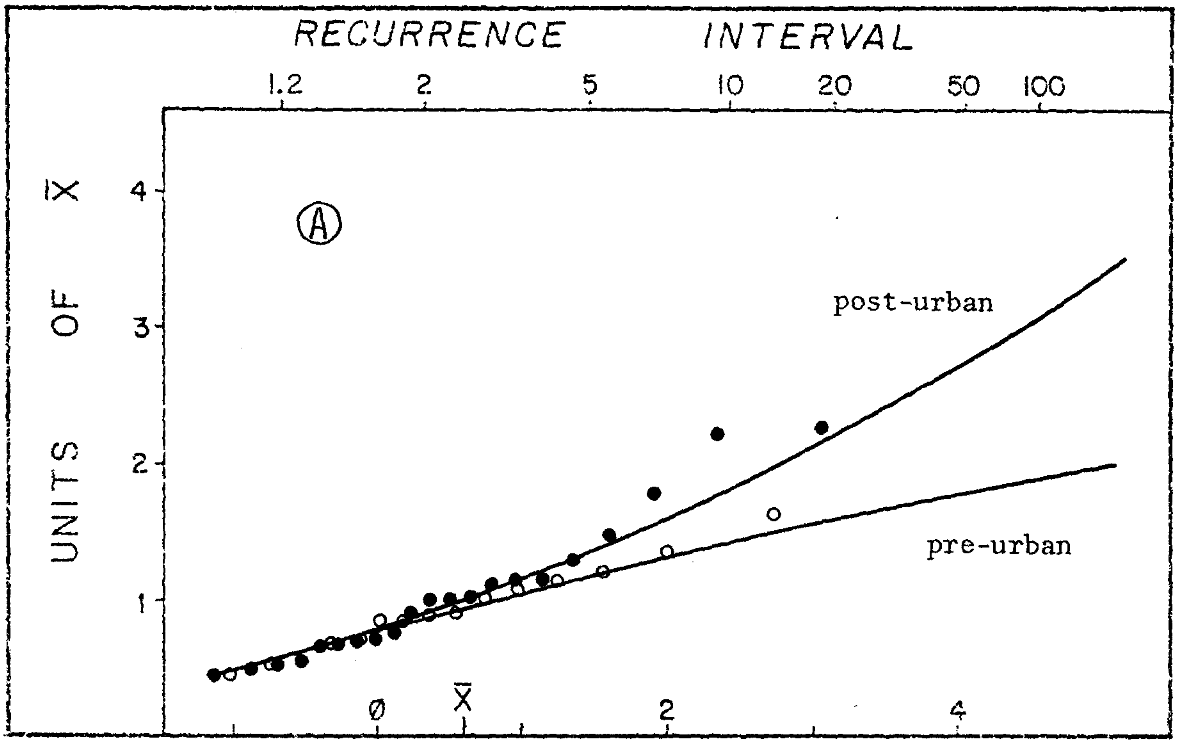


FIGURE 21. Flood frequency curves, Parker River.

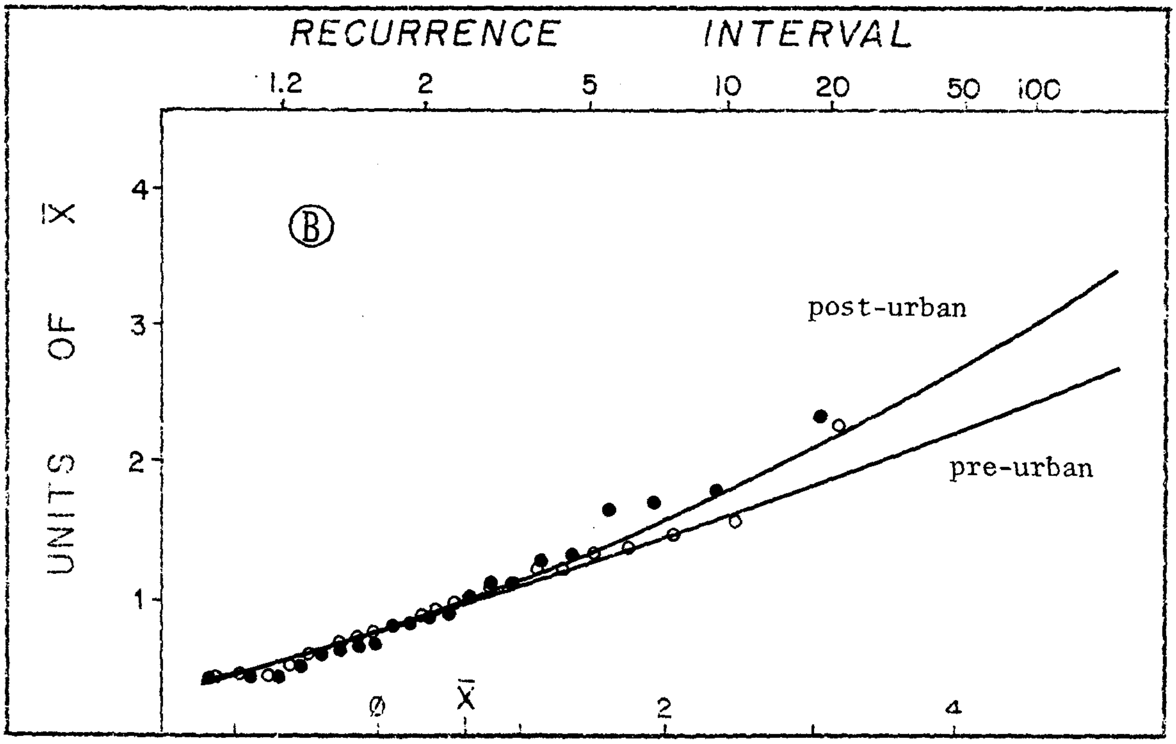


FIGURE 22. Flood frequency curves, Ipswich River.

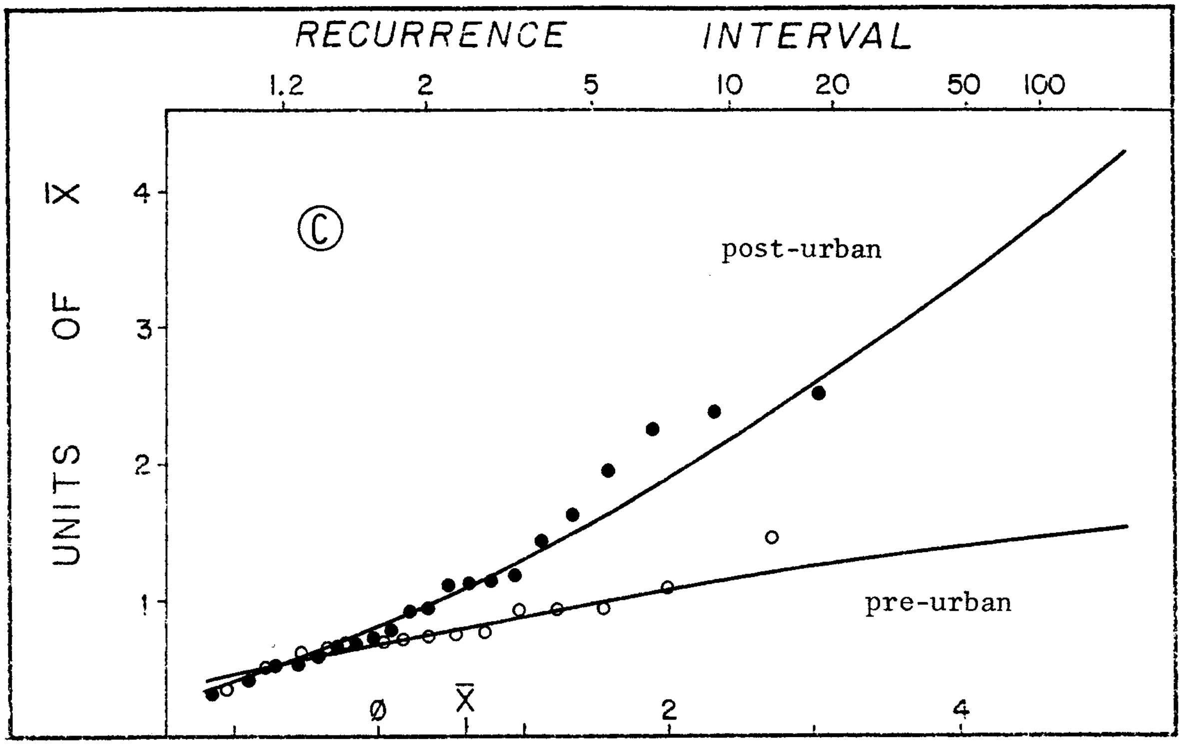


FIGURE 23. Flood frequency curves, Aberjona River.

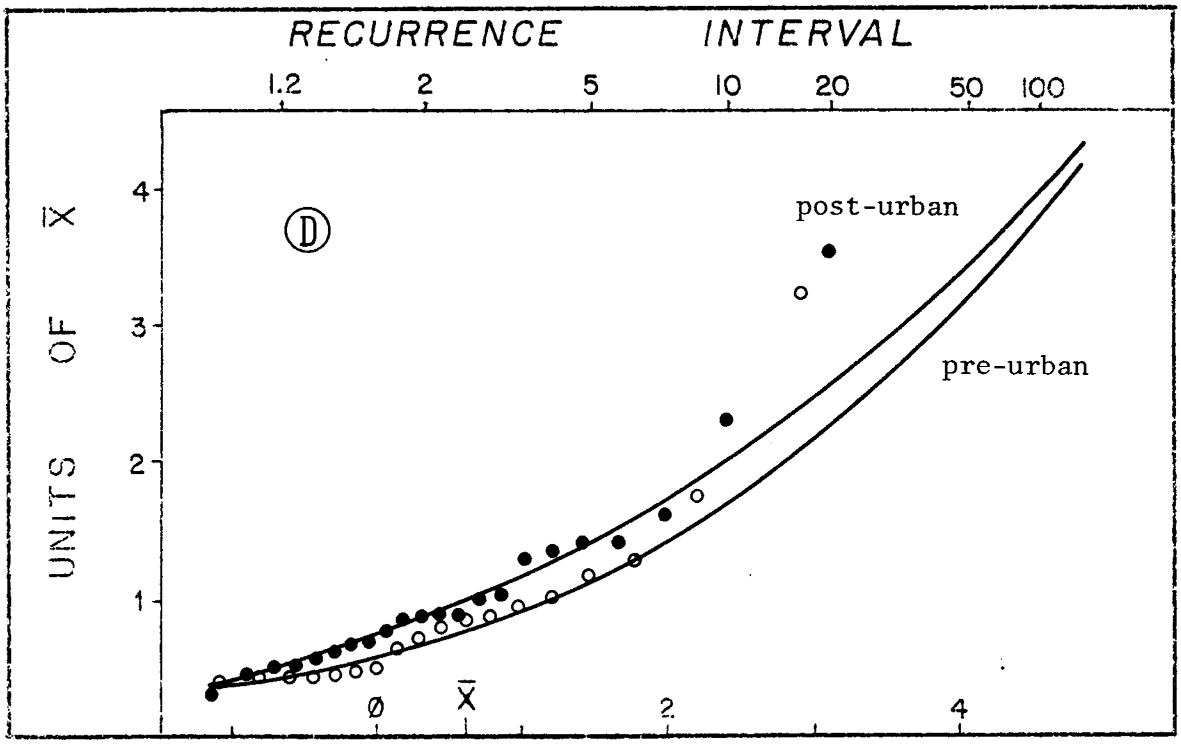


FIGURE 24. Flood frequency curves, North Nashua River.

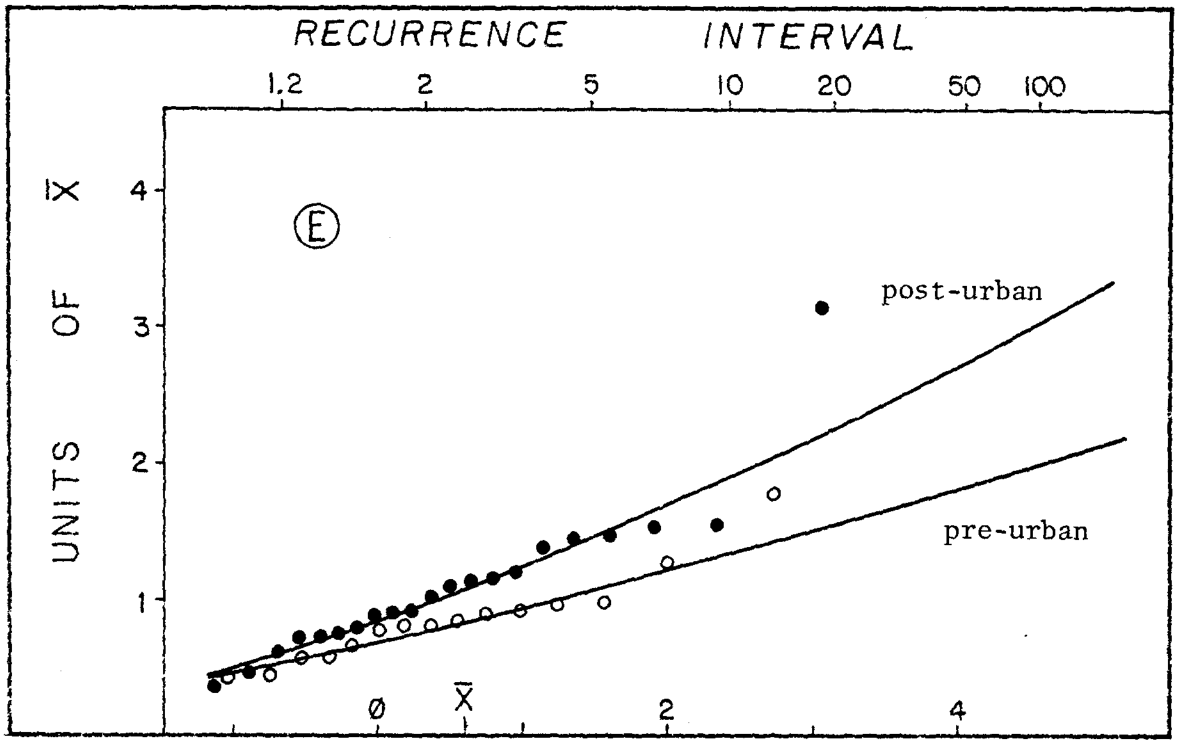


FIGURE 25. Flood frequency curves, Assabet River.

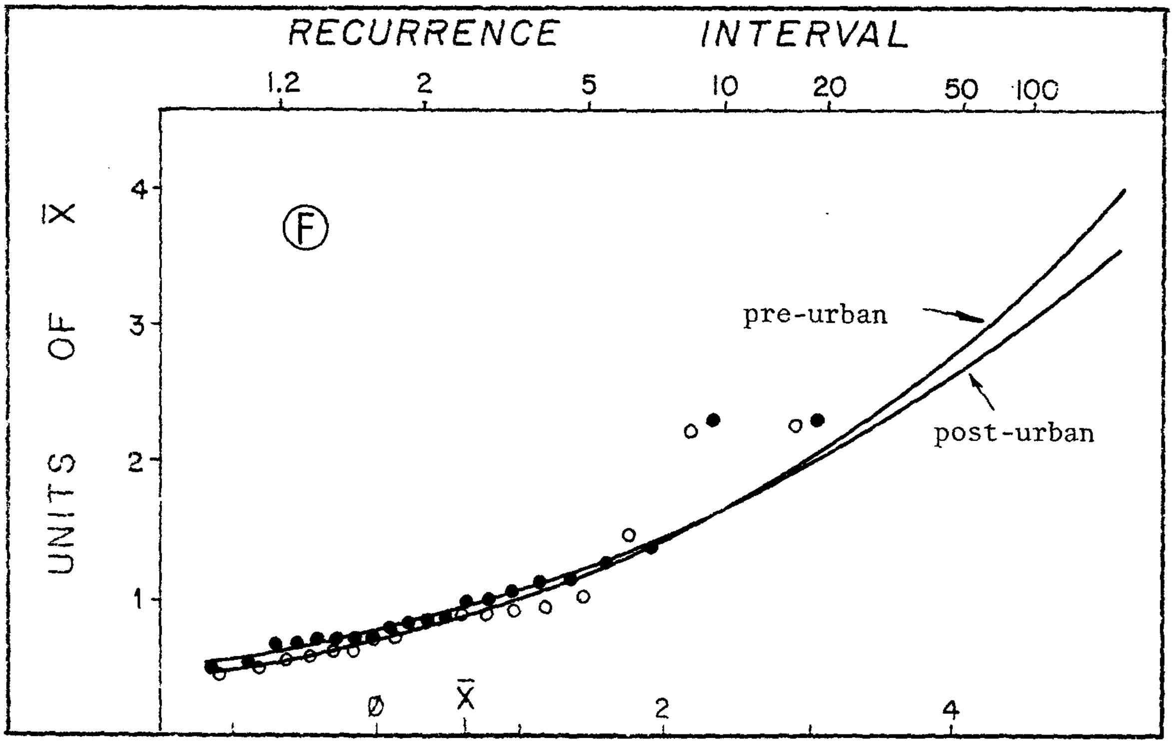


FIGURE 26. Flood frequency curves, Charles River.

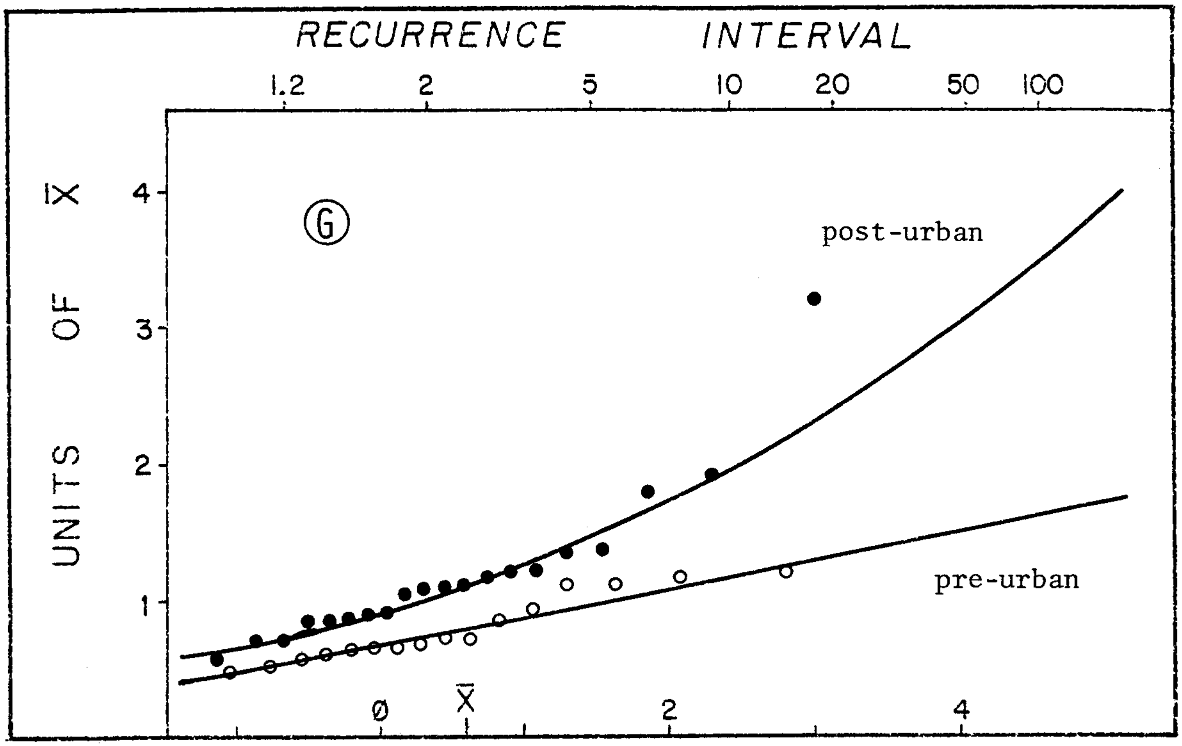


FIGURE 27. Flood frequency curves, Neponset River.

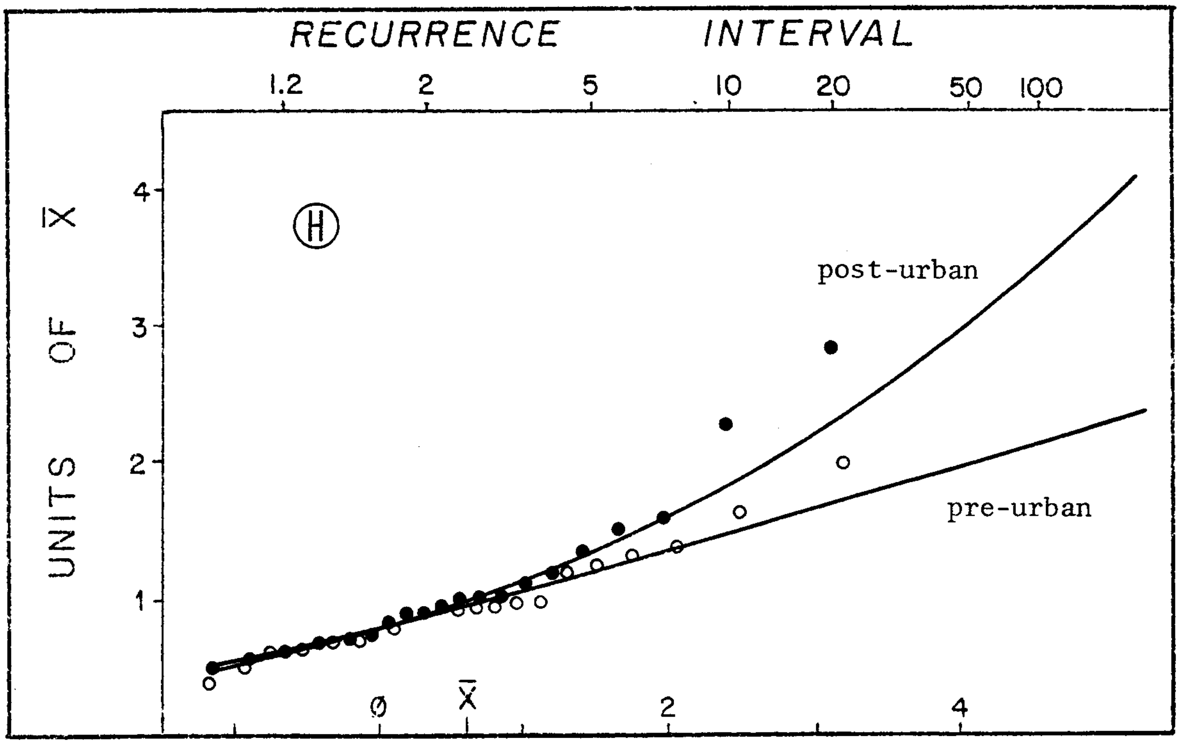


FIGURE 28. Flood frequency curves, Wading River.

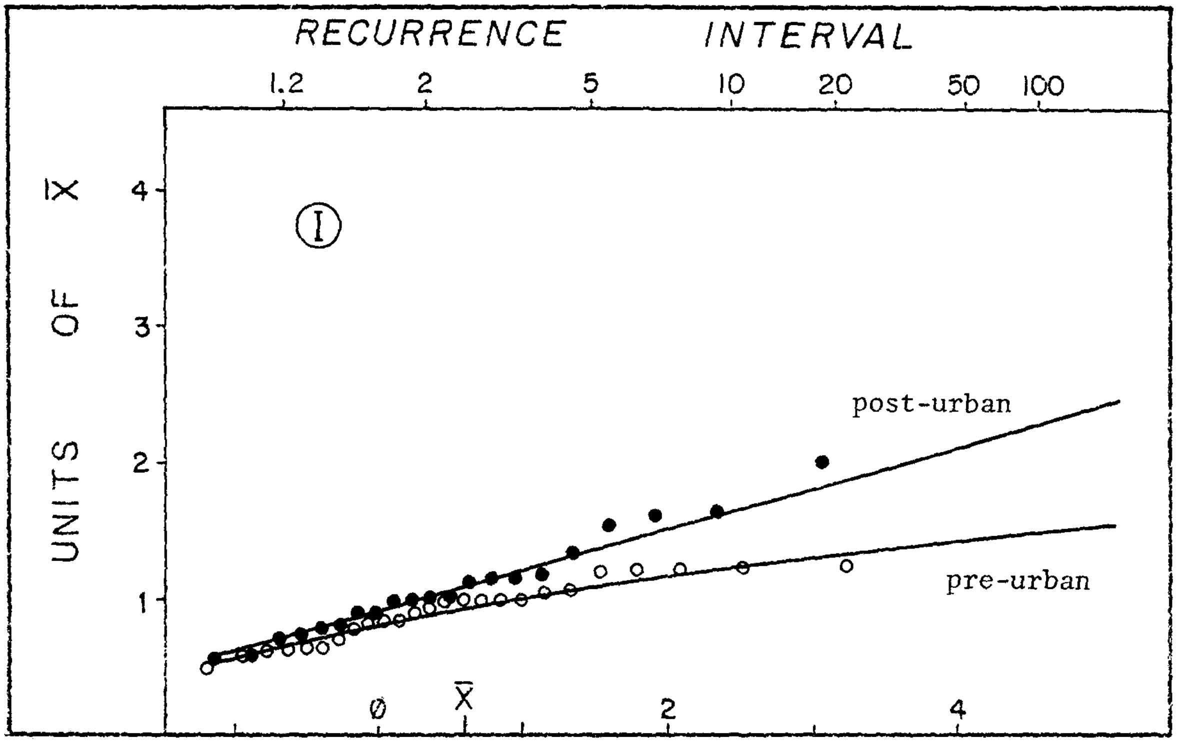


FIGURE 29. Flood frequency curves, Taunton River.

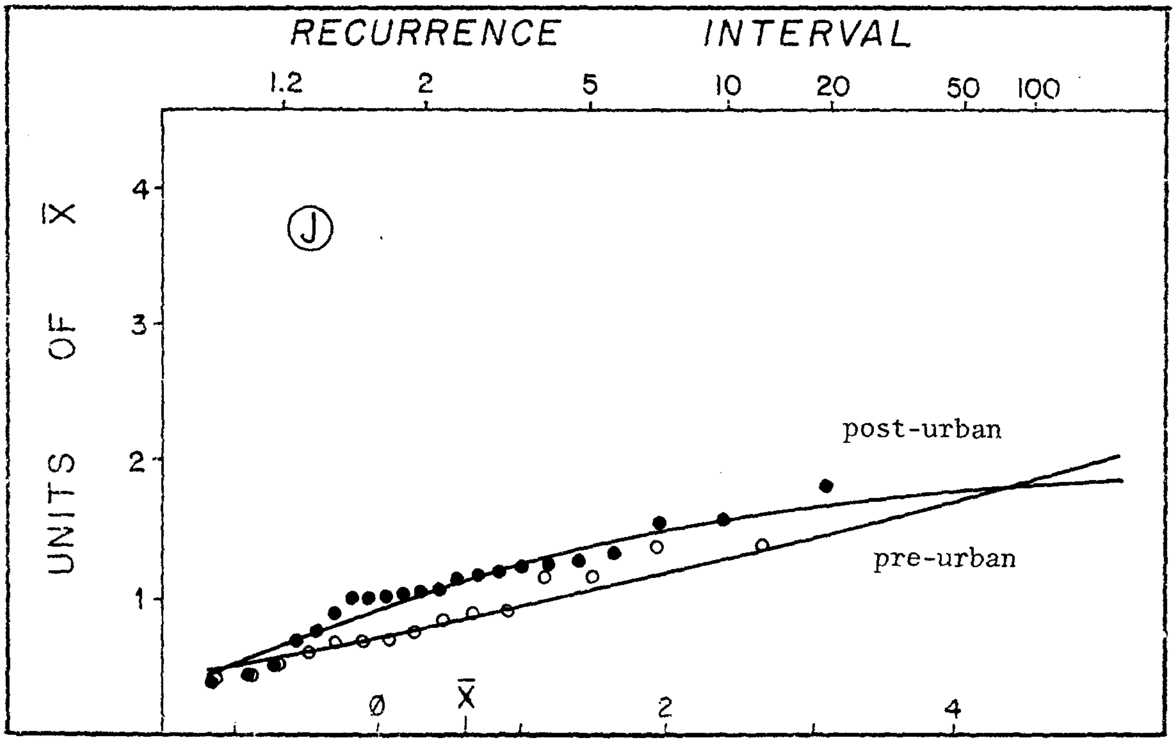


FIGURE 30. Flood frequency curves, Adamsville Brook.

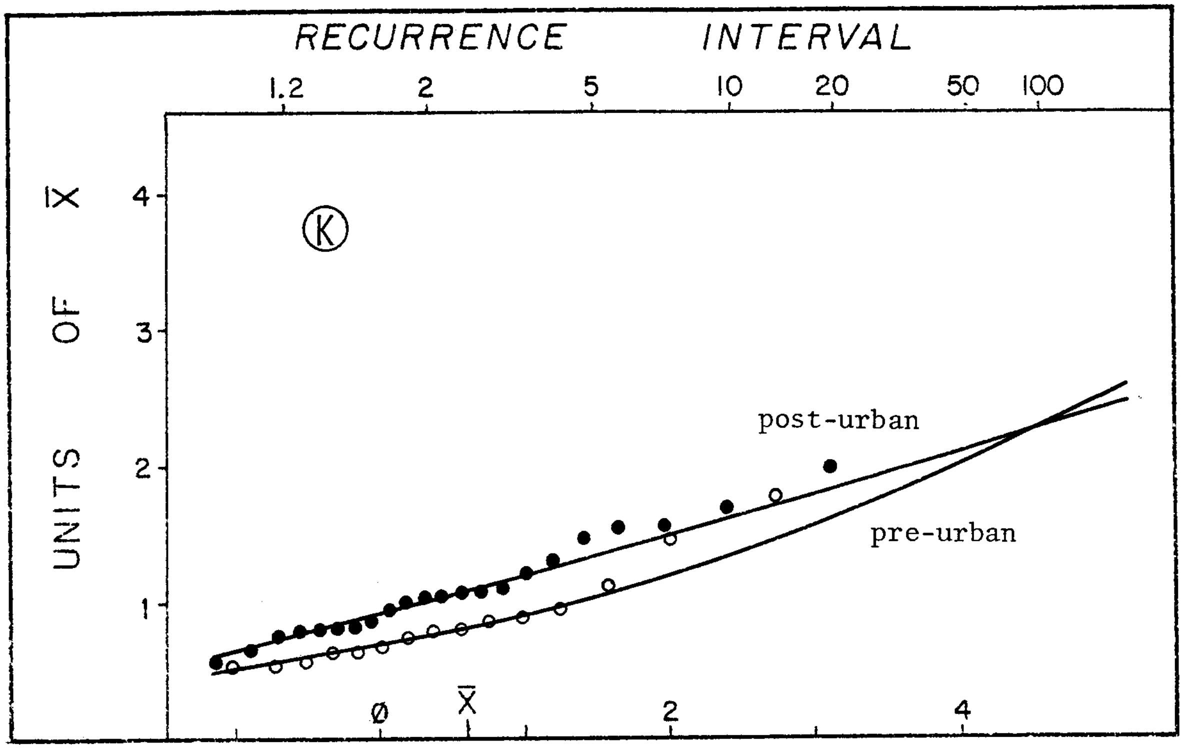


FIGURE 31. Flood frequency curves, Wood River.

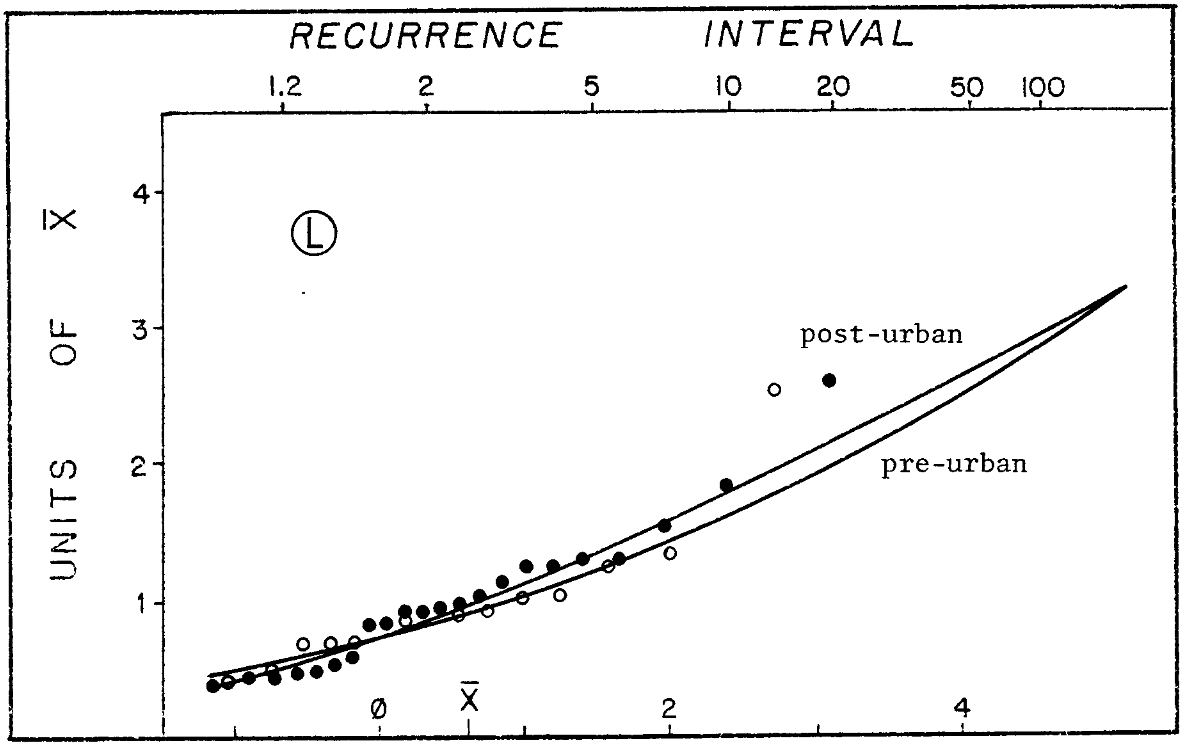


FIGURE 32. Flood frequency curves, South Pawtuxet River.

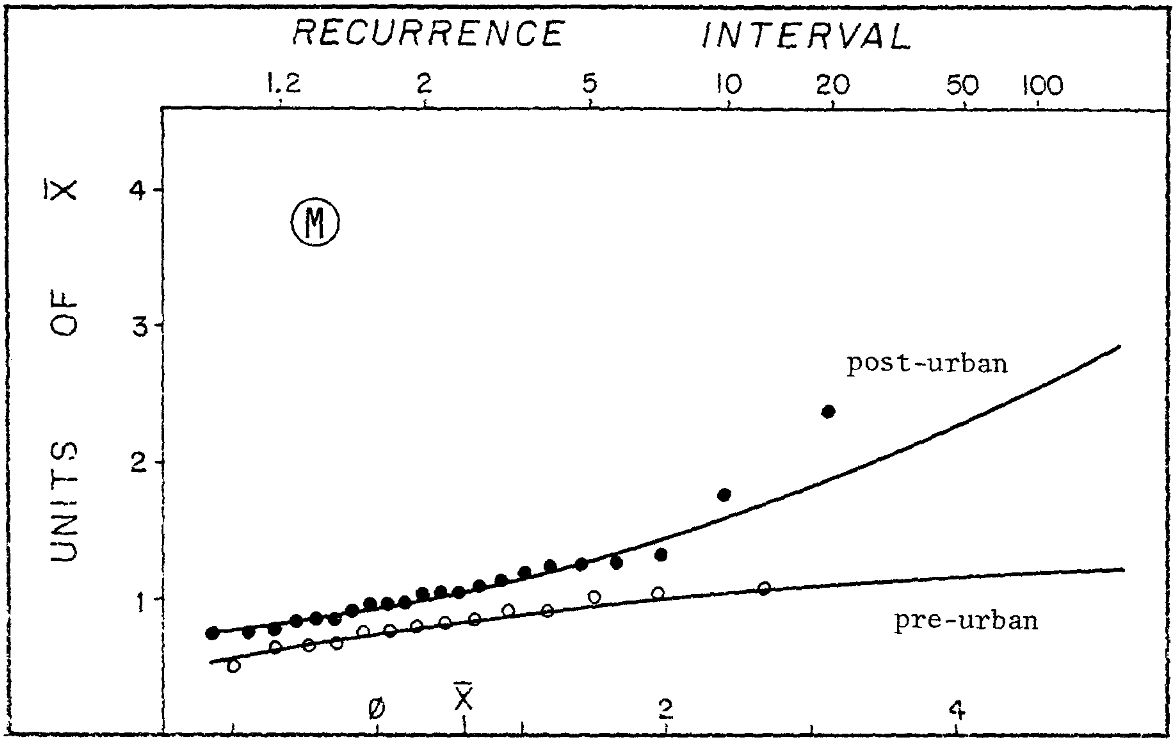


FIGURE 33. Flood frequency curves, Potowomut River.

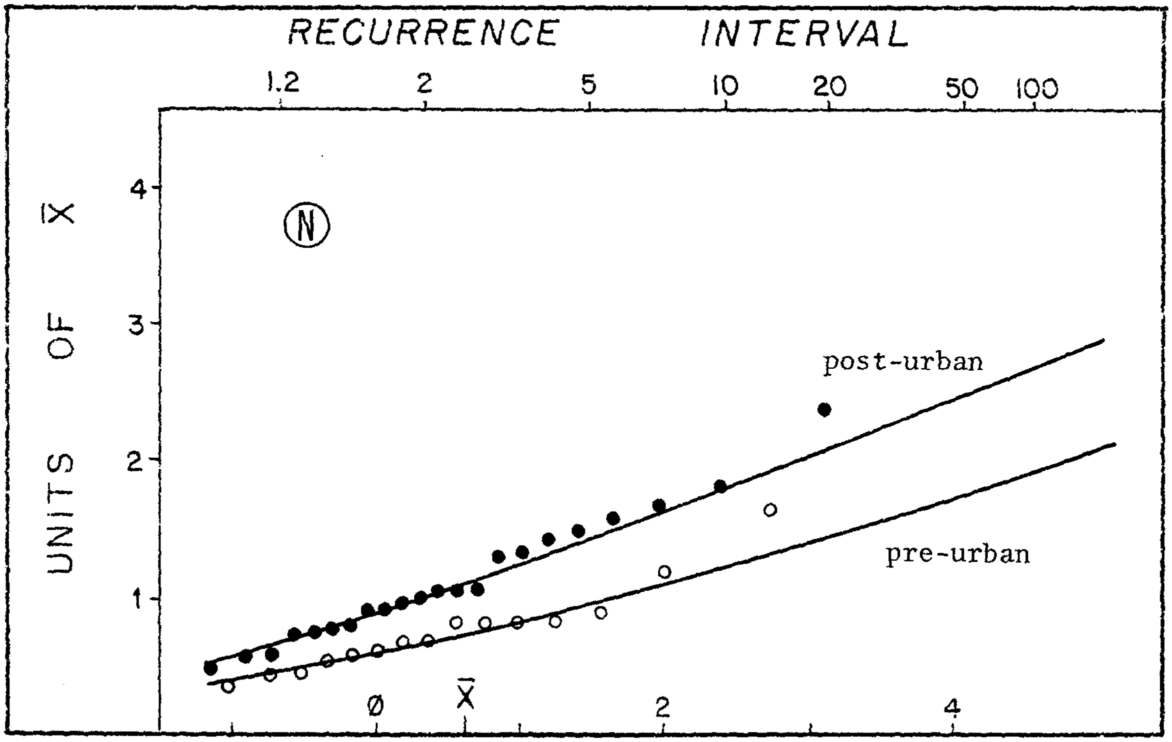


FIGURE 34. Flood frequency curves, Woonasquatucket River.

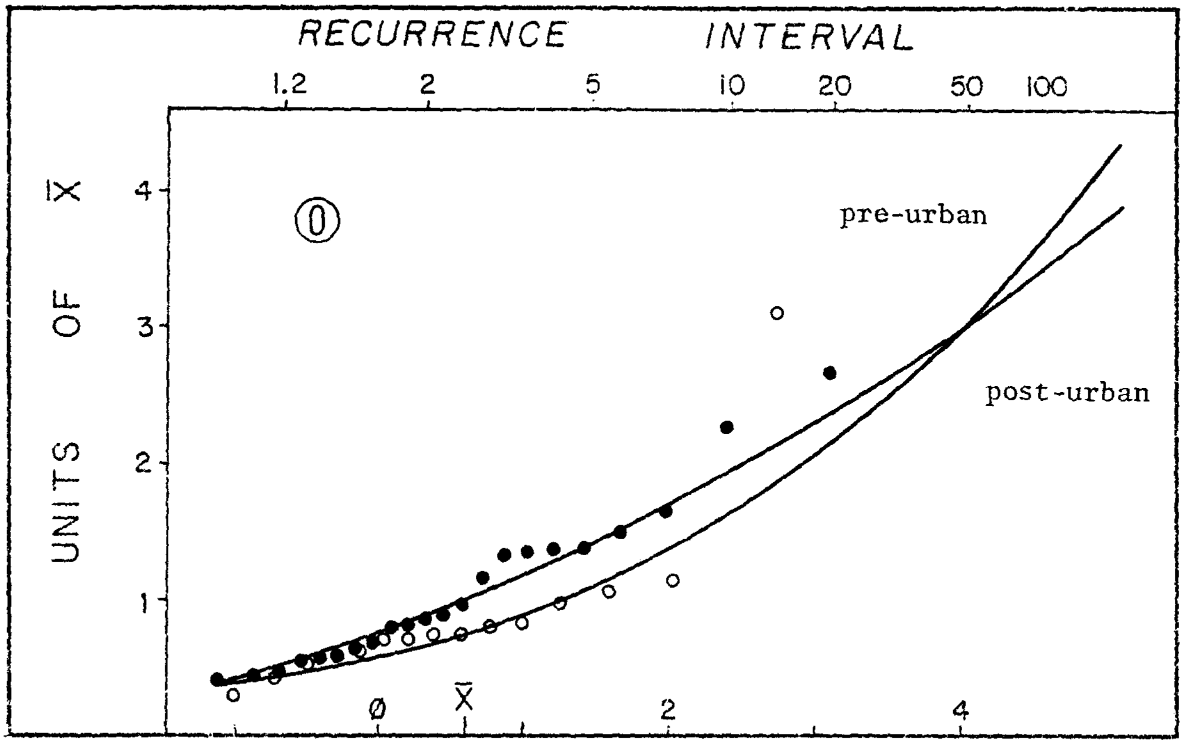


FIGURE 35. Flood frequency curves, Branch River.

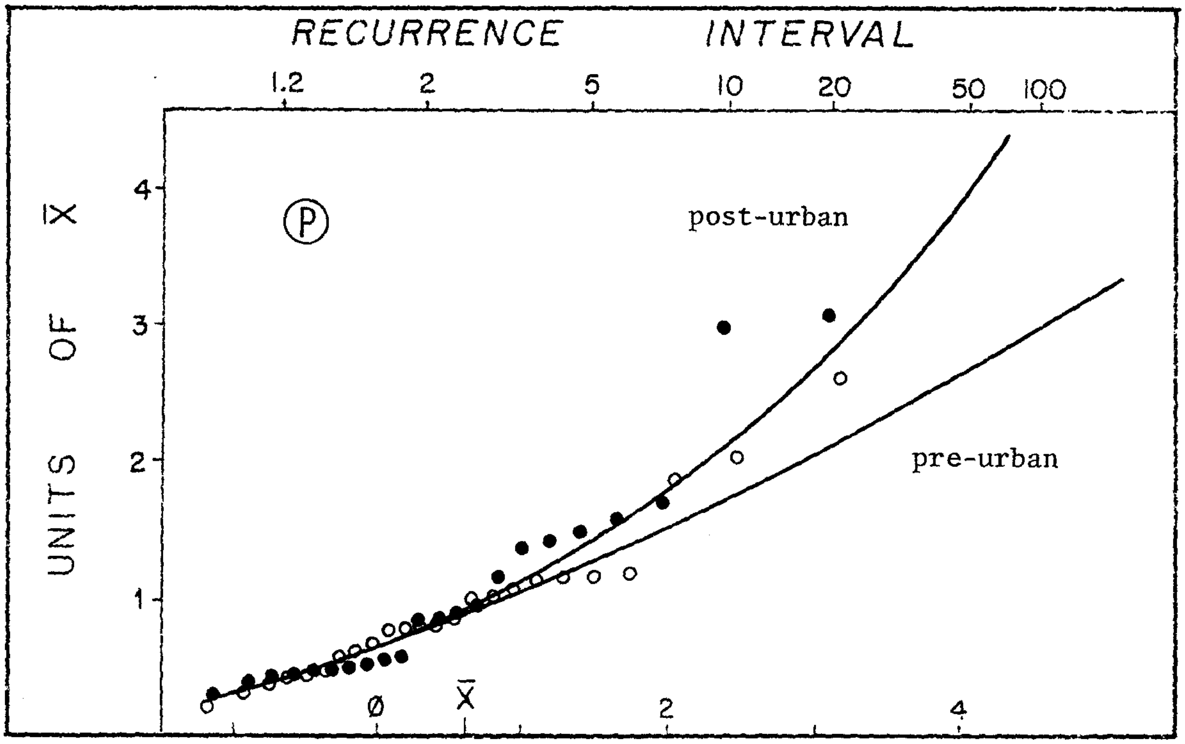


FIGURE 36. Flood frequency curves, Kettle Brook.

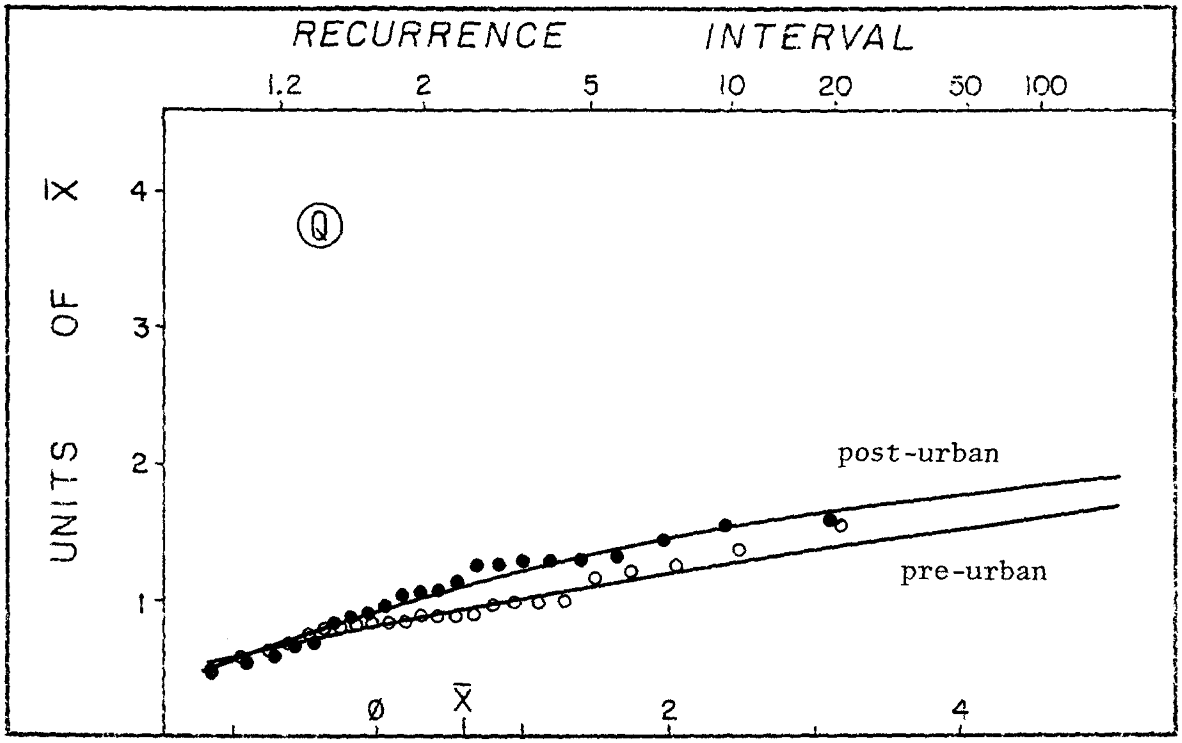


FIGURE 37. Flood frequency curves, Quaboag River.

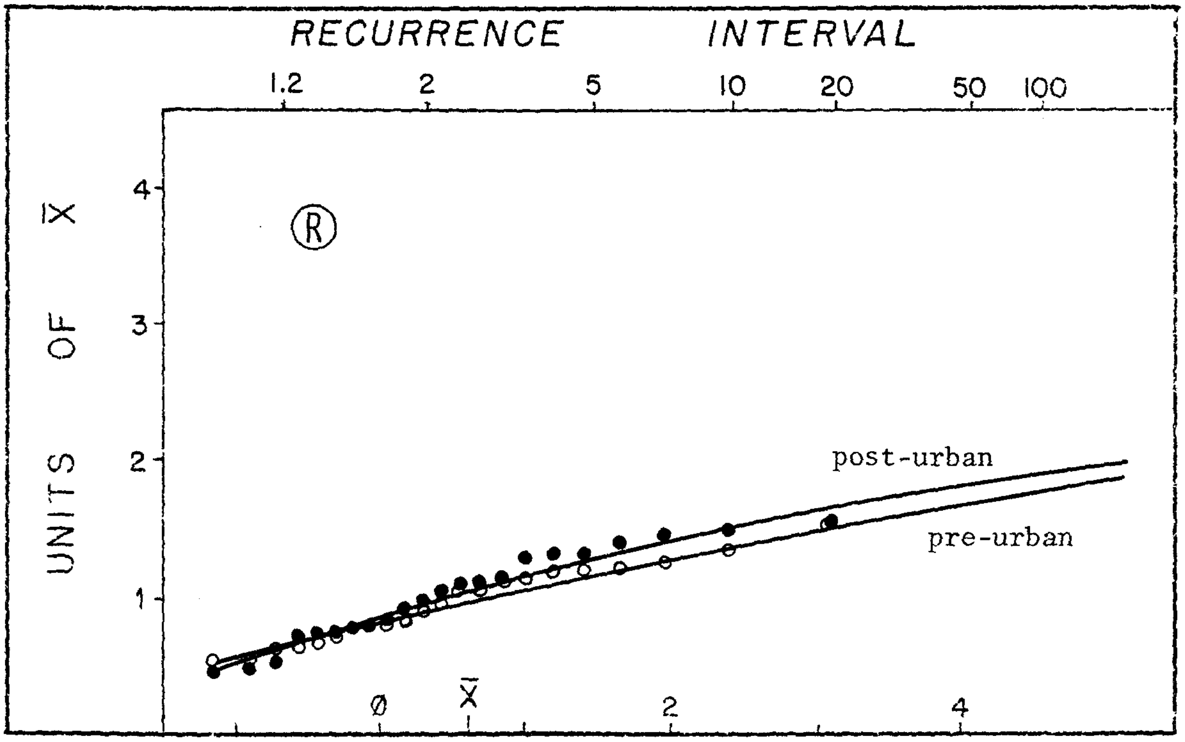


FIGURE 38. Flood frequency curves, Quinebaug River.

APPENDIX C: Data Tables

In the tabulations on this and following pages, all linear measurements are given in miles and all areal measurements are given in square miles, with the exception of channel gradient in feet per mile.

Basin	Area	Perimeter	Circularity Ratio (R_c)	Length of Main Channel	Main Channel Gradient (S_c)
A	22.09	34.43	0.234	12.56	10.35
B	121.80	84.07	0.217	32.80	3.54
C	25.04	34.59	0.263	8.93	8.96
D	107.90	70.47	0.273	23.66	36.35
E	110.86	71.74	0.271	26.82	11.82
F	181.65	93.54	0.261	47.51	8.80
G	33.88	34.00	0.368	11.14	18.85
H	42.00	51.51	0.199	16.59	13.26
I	219.13	110.75	0.225	31.07	7.63
J	8.85	16.05	0.432	5.81	33.22
K	72.92	54.89	0.304	17.82	18.13
L	61.54	47.82	0.338	16.05	17.76
M	23.82	25.13	0.474	9.52	40.23
N	38.09	39.48	0.307	14.35	26.41
O	91.32	64.04	0.280	19.52	27.46
P	31.40	38.54	0.266	13.25	48.08
Q	149.80	89.20	0.237	30.96	22.29
R	151.16	87.03	0.251	28.31	23.07

Data Tables (cont)

Basin	Drainage Density (D_d)	Bifurcation Ratio (R_b)	Basin Order	Modified E-Ratio (E)	Basin Magnitude (M)
A	1.569	3.26	4	1.500	30
B	1.439	3.49	5	2.569	127
C	1.651	3.36	4	1.082	35
D	1.733	4.98	4	2.211	138
E	1.936	3.87	5	3.062	180
F	2.199	4.49	5	4.546	352
G	1.684	3.54	4	1.032	44
H	1.900	4.34	4	2.335	72
I	1.866	4.29	5	2.795	339
J	1.953	3.46	3	1.118	12
K	1.379	3.68	4	1.526	52
L	1.595	3.96	4	1.534	63
M	1.779	3.26	4	1.367	31
N	2.345	3.19	5	2.186	90
O	1.711	3.54	5	2.477	155
P	1.148	3.98	4	1.522	68
Q	2.135	4.01	5	2.871	249
R	2.013	4.18	5	3.628	243

Data Tables (cont)

Basin	Area of Pervious	Percent	Area of Swamp Deposits	Percent	Pervious Index (I_p)	Wetland Index (I_w)
A	9.13	41	3.95	18	0.540	0.438
B	39.93	33	28.71	24	0.361	0.540
C	12.15	49	1.94	8	0.819	0.177
D	36.16	34	7.29	7	0.458	0.113
E	34.20	31	10.91	10	0.391	0.166
F	79.30	44	30.32	17	0.598	0.421
G	14.98	44	6.28	19	0.595	0.498
H	23.17	55	9.83	23	0.808	1.092
I	80.25	37	46.90	21	0.432	0.510
J	1.12	13	1.54	17	0.121	0.249
K	16.94	23	7.44	10	0.267	0.153
L	21.67	35	6.65	11	0.466	0.200
M	9.74	41	2.42	10	0.590	0.208
N	8.36	22	4.74	12	0.243	0.190
O	24.83	27	9.60	11	0.326	0.169
P	6.01	19	3.79	12	0.206	0.175
Q	26.21	17	18.30	12	0.185	0.174
R	16.26	11	11.55	8	0.111	0.094

Data Tables (cont)

Basin	Area of Urban Land - 1952	Percent	Area of Urban Land - 1972	Percent	Urbanization Index ($I_{\Delta u}$)
A	0.83	4	2.55	12	0.085
B	7.69	6	22.40	18	0.139
C	11.01	44	14.50	58	0.186
D	11.42	11	16.22	15	0.047
E	7.20	6	17.26	16	0.101
F	19.79	11	33.43	18	0.082
G	4.86	14	7.84	23	0.098
H	2.66	6	5.04	12	0.060
I	26.21	12	41.77	19	0.077
J	0.29	3	0.42	5	0.015
K	0.77	1	2.41	3	0.023
L	1.27	2	4.24	7	0.051
M	1.26	5	4.25	18	0.145
N	2.50	7	5.64	15	0.090
O	3.58	4	6.53	7	0.033
P	4.09	13	6.93	22	0.101
Q	3.88	3	6.02	4	0.015
R	4.21	3	8.70	6	0.031

Data Tables (cont)

	PRE-URBAN		POST-URBAN			
Basin	Q_{50} (cfs)	Q_{100} (cfs)	Q_{50} (cfs)	Q_{100} (cfs)	$I_{h \cdot 50}$	$I_{h \cdot 100}$
A	379	413	572	670	0.965	1.291
B	2513	2831	3000	3496	0.435	0.595
C	460	492	1074	1261	2.371	2.968
D	5285	6362	5776	6564	0.255	0.105
E	2056	2319	3082	3504	1.061	1.225
F	3772	4692	3552	4285	-0.167	-0.309
G	532	587	1050	1260	1.830	2.383
H	1000	1116	1477	1793	0.979	1.391
I	3519	3700	5162	5686	0.732	0.885
J	294	327	308	319	0.093	-0.051
K	1703	2004	1805	1993	0.137	-0.015
L	1745	2073	1842	2124	0.140	0.073
M	424	441	809	939	1.317	1.703
N	1032	1179	1464	1641	0.945	1.010
O	5380	6940	5494	6524	0.070	-0.254
P	1334	1558	1956	2501	1.320	2.001
Q	1841	1979	2157	2276	0.291	0.274
R	3210	3466	3422	3634	0.118	0.094