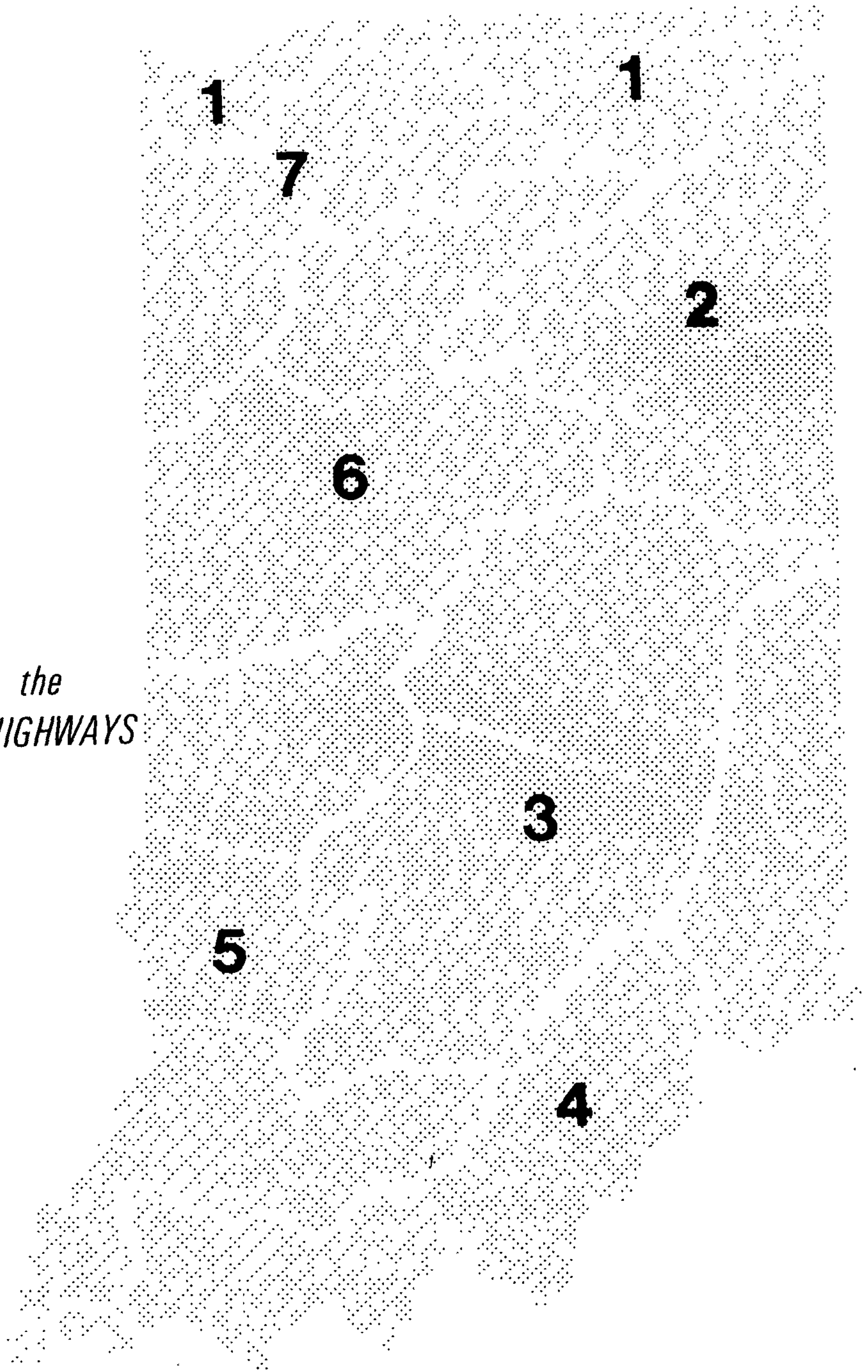
TECHNIQUES FOR ESTIMATING MAGNITUDE AND FREQUENCY OF FLOODS ON STREAMS IN INDIANA

U.S. GEOLOGICAL SURVEY
Water-Resources Investigations Report 84-4134



Prepared in cooperation with the
INDIANA DEPARTMENT OF HIGHWAYS
and the FEDERAL HIGHWAY
ADMINISTRATION

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO.	2.	3. Recipient's Accession No.

4. Title and Subtitle	5. Report Date
Techniques for Estimating Magnitude and Frequency of Floods on Streams in Indiana	1984
	6.
7. Author(s) Dale R. Glatfelter	8. Performing Organization Rept. No. USGS/WRI 84-4134
9. Performing Organization Name and Address U.S. Geological Survey, Water Resources Division 6023 Guion Road, Suite 201 Indianapolis, Indiana 46254	10. Project/Task/Work Unit No.
	11. Contract(C) or Grant(G) No. (C) (G)
12. Sponsoring Organization Name and Address U.S. Geological Survey, Water Resources Division 6023 Guion Road, Suite 201 Indianapolis, Indiana 46254	13. Type of Report & Period Covered Final
	14.

15. Supplementary Notes

Prepared in cooperation with the Indiana Department of Highways and the Federal Highway Administration

16. Abstract (Limit: 200 words)

Methods for estimating magnitude and frequency of floods on streams are necessary in the design of hydraulic structures and in flood-plain management. Equations for estimating the magnitude and frequency of floods on ungaged streams in Indiana were developed by multiple-regression analysis of basin characteristics and peak-flow statistical data from 242 gaged sites in Indiana, Ohio, and Illinois.

The State of Indiana was divided into seven areas on the basis of regression analysis. A set of equations for estimating peak discharges with recurrence intervals of 2, 10, 25, 50, and 100 years was developed for each area. The equations are valid for unregulated and nonurban streams. Significant basin characteristics in the equations are drainage area, channel length, channel slope, mean annual precipitation, intensity of precipitation, storage, and a runoff coefficient. Standard errors of estimate for the equations range from 24 to 45 percent.

A rainfall-runoff model was used to extend the length of record at 11 gaged sites on small streams. Flood-frequency data from the synthetic peaks and from the observed peaks were combined by weighting techniques.

17. Document Analysis a. Descriptors

*Flood frequency, *Flood peak, *Rainfall-runoff relationships, *Regional analysis, *Regression analysis, Flood recurrence interval, Frequency analysis, Indiana, Model studies, Regulated flow, Small watersheds, Statistical analysis, Stream discharge

b. Identifiers/Open-Ended Terms

Basin characteristics, Magnitude and frequency, Synthesis of discharge, Wabash River

c. COSATI Field/Group

18. Availability Statement No restriction on distribution. This report may be purchased from: Open-File Services Sect., Western Distribution Ctr., U.S.G.S., Box 25425, Federal Ctr., Denver, Colorado, 80225	19. Security Class (This Report) Unclassified	21. No. of Pages 110
	20. Security Class (This Page) Unclassified	22. Price

(See ANSI-Z39.18) See Instructions on Reverse

OPTIONAL FORM 272 (4-77)
(Formerly NTIS-35)
Department of Commerce

TECHNIQUES FOR ESTIMATING MAGNITUDE AND FREQUENCY OF FLOODS ON STREAMS IN INDIANA

By Dale R. Glatfelter

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 84-4134

Prepared in cooperation with the

INDIANA DEPARTMENT OF HIGHWAYS

and

FEDERAL HIGHWAY ADMINISTRATION

Indianapolis, Indiana

1984

UNITED STATES DEPARTMENT OF THE INTERIOR

WILLIAM P. CLARK, Secretary

GEOLOGICAL SURVEY

Dallas L. Peck, Director

For additional information write to:

District Chief
U.S. Geological Survey
6023 Guion Road, Suite 201
Indianapolis, Indiana 46254

Copies of this report
can be purchased from:

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

CONTENTS

ILLUSTRATIONS

TABLES

CONVERSION TABLE

The inch-pound system of units was used to develop the estimating equations within this report. Inch-pound units can be converted to the International System (SI) of units as follows:

Multiply inch-pound unit	By	To obtain SI unit
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
foot per mile (ft/mi)	0.1894	meter per kilometer (m/km)
square mile (mi^2)	2.590	square kilometer (km^2)
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

TECHNIQUES FOR ESTIMATING MAGNITUDE AND FREQUENCY OF FLOODS ON STREAMS IN INDIANA

by Dale R. Glatfelter

ABSTRACT

Equations are presented for estimating the magnitude and frequency of floods at ungaged sites on unregulated and nonurban streams in Indiana. The equations were developed by multiple-regression analysis of basin characteristics and peak-flow statistical data from 242 gaged locations in Indiana, Ohio, and Illinois. The State of Indiana was divided into seven areas on the basis of the regression analysis. A set of equations for estimating peak discharges with recurrence intervals of 2, 10, 25, 50, and 100 years was developed for each area. Significant basin characteristics in the equations are drainage area, channel length, channel slope, mean annual precipitation, storage, precipitation intensity, and a runoff coefficient. Standard errors of estimate for the equations range from 24 to 45 percent.

Methods are also presented for estimating flood magnitude and frequency at sites on gaged streams. Flood-frequency data based on observed peaks are given for 270 gaged locations. Twenty of these are on regulated streams, and six are on urban streams. Basin characteristics are also included for 245 of the gaged locations on unregulated and nonurban streams. No techniques are given for estimating flood magnitude and frequency at ungaged sites on regulated or urban streams.

A rainfall-runoff model was used to synthesize long-term peak data at 11 gaged locations on small streams. Flood-frequency curves developed from the long-term synthetic data were combined with curves based on short-term observed data to provide weighted estimates of flood magnitude and frequency at the rainfall-runoff stations.

INTRODUCTION

Purpose and Scope

The purpose of this report is to present techniques for estimating the magnitude and frequency of floods on streams in Indiana. This information is necessary in the design of culverts, bridges, and other hydraulic structures, and in flood-plain management. The contents of this report do not necessarily reflect the official views or policies of the Indiana Department of Highways or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

Background

A study designed primarily to define flood magnitude and frequency on small streams was begun in 1972 by the U.S. Geological Survey in cooperation with the Indiana Department of Highways and the Federal Highway Administration. Davis (1974) presented equations for estimating the magnitude and frequency of floods on any stream in Indiana that drained an area greater than 15 mi^2 and was not affected by regulation or urbanization. Gold (1980) presented equations for estimating the magnitude of floods having 2-year and 10-year recurrence intervals. The equations in Gold's interim report were valid for unregulated, nonurban streams of any size drainage area, but the standard errors of estimate for the equations were greater then those determined by Davis. An additional 10 years of peak data at most stations used in Davis' report, revised techniques for flood-frequency determination (U.S. Water Resources Council, 1981), and 10 years of peak data at gaged sites on small streams were used to update and improve the estimating procedures presented by Davis and Gold.

Methods of Study

Flood-frequency curves were developed from annual peak-discharge data for 242 gaging stations and crest-stage partial-record sites (3 in Ohio, 3 in Illinois, and 236 in Indiana) and guidelines given in U.S. Water Resources Council (1981). The flood-frequency curves from the observed data were then used along with basin characteristics in multiple-regression analysis to develop equations for estimating magnitude and frequency of floods. Basin characteristics and flood-frequency data are presented in tables.

A rainfall-runoff model developed by the U.S. Geological Survey (Dawdy and others, 1972) was used to synthesize peak data at 11 gaging stations on small streams. On the basis of the synthetic data, flood-frequency curves were developed by procedures discussed in Lichty and Liscum (1978). A weighting technique was used to combine the estimates of flood magnitude and frequency obtained from the observed and the synthetic peak data into one flood-frequency curve for the station for use in the regression analysis.

The estimating equations presented in the report were developed by multiple-regression techniques described in Helwig and Council (1979). Basin characteristics for 242 gaged locations were used as the independent variables, and corresponding peak-discharge statistics were used as the dependent variables. On the basis of regression analysis the State was divided into seven areas. A set of equations for estimating peak discharges with recurrence intervals of 2, 10, 25, 50, and 100 years was developed for each area. These equations are applicable only for locations on unregulated, nonurban streams. Examples showing use of the estimating equations are given in the section "Estimating Techniques."

Flood-frequency data are presented in the report for 20 sites on regulated streams and 6 sites on urban streams that were not used in regression analysis. The scope of the report does not include development of techniques for estimating magnitude and frequency of floods at ungaged sites on regulated and urban streams. Rather, the data obtained at sites on regulated and urban streams are presented for use in estimating flood magnitude and frequency at specific locations under current (1983) conditions. A change in regulatory practices or increased urbanization can greatly affect flow characteristics. Peak data should be thoroughly reviewed before a flood-frequency analysis is made.

Peak-discharge data from stations on the Wabash River were analyzed to show the effect of regulation on flood frequency. Results of separate analyses of unregulated peaks and of regulated peaks are presented in the report.

Acknowledgments

The report is the result of a cooperative agreement between the Indiana Department of Highways, the Federal Highway Administration, and the U.S. Geological Survey. Most of the small-stream data used in this report were collected under this cooperative program. The remainder of the streamflow data were collected for many years under various cooperative agreements with State and Federal agencies. Long-term daily precipitation and evaporation data, and rainfall data at 5-minute intervals from individual storms for use in rainfall-runoff modeling were obtained from the National Oceanic and Atmospheric Administration (NOAA) and have been stored in the Geological Survey computer files.

ESTIMATING TECHNIQUES

Sites on Ungaged Streams

Equations were developed to estimate flood magnitude at 2-, 10-, 25-, 50-, and 100-year recurrence intervals from basin characteristics at ungaged sites on unregulated, nonurban streams (table 1). These equations are not intended for use at sites on regulated or urban streams because changes in regulatory practices or increased urbanization can affect peak-flow characteristics. (See sections "Regulated Gaged Streams" and "Urban Gaged Streams.")

Annual peak-flow data from 236 gaging stations and crest-stage partial-record sites in Indiana (fig. 1) plus three stations in Ohio and three stations in Illinois (not on the map) were analyzed by techniques described in U.S. Water Resources Council (1981) to determine peak-flow statistics for each location (table 2, after References). On the basis of the analysis, flood magnitude and frequency were estimated for each of the gáged locations.

Table 1.--Equations for estimating magnitude and frequency of floods on streams in Indiana

Area 1 (16 stations)

Equations	Standard error of estimate		Equivalent years of record
	Log units	Percent	
$Q_2 = 6.72\ DA^{0.714}(STOR + 1)^{-0.289}(PREC - 30)^{0.965}$	0.114	27	3
$Q_{10} = 10.3\ DA^{0.701}(STOR + 1)^{-0.262}(PREC - 30)^{1.060}$	.149	35	3
$Q_{25} = 11.8\ DA^{0.697}(STOR + 1)^{-0.253}(PREC - 30)^{1.093}$	.165	39	3
$Q_{50} = 12.9\ DA^{0.696}(STOR + 1)^{-0.248}(PREC - 30)^{1.114}$	.176	42	4
$Q_{100} = 13.8\ DA^{0.695}(STOR + 1)^{-0.243}(PREC - 30)^{1.132}$	.186	45	5

Statistics of basin characteristics used in area 1 regression analysis

Basin characteristic	Maximum	Minimum	Mean	Median
DA	3,370 mi^2	0.17 mi^2	321 mi^2	79.1 mi^2
STOR	13.3 percent	0.0 percent	3.0 percent	1.3 percent
PREC	46.0 in.	34.0 in.	37.1 in.	35.3 in.

Table 1.--Equations for estimating magnitude and frequency of floods on streams in Indiana--Continued

Area 2 (31 stations)

Equations	Standard error of estimate		Equivalent years of record
	Log units	Percent	
$Q_2 = 26.4\ DA^{0.708}(STOR + 1)^{-0.207}RC^{0.479}(PREC - 30)^{0.653}$	0.104	24	4
$Q_{10} = 61.8\ DA^{0.655}(STOR + 1)^{-0.312}RC^{0.697}(PREC - 30)^{0.696}$	.120	28	4
$Q_{25} = 85.0\ DA^{0.635}(STOR + 1)^{-0.357}RC^{0.782}(PREC - 30)^{0.702}$	.134	31	5
$Q_{50} = 106\ DA^{0.619}(STOR + 1)^{-0.391}RC^{0.859}(PREC - 30)^{0.707}$	.147	35	6
$Q_{100} = 127\ DA^{0.608}(STOR + 1)^{-0.418}RC^{0.902}(PREC - 30)^{0.708}$	.156	37	7

Statistics of basin characteristics used in area 2 regression analysis

Basin characteristic	Maximum	Minimum	Mean	Median
DA	1,967 mi^2	0.17 mi^2	384 mi^2	270 mi^2
STOR	4.1 percent	0 percent	0.8 percent	0.3 percent
RC	0.8	0.5	0.7	0.8
PREC	39.0 in.	34.0 in.	36.7 in.	37.0 in.

Table 1.--Equations for estimating magnitude and frequency of floods on streams in Indiana--Continued

Area 3 (60 stations)

Equations	Standard error of estimate		Equivalent years of record
	Log units	Percent	
$Q_2 = 102\ DA^{0.758} SL^{0.273} (I_{24,2} - 2.5)^{0.948}$	0.150	36	3
$Q_{10} = 141\ DA^{0.772} SL^{0.384} (I_{24,2} - 2.5)^{0.894}$	.144	34	4
$Q_{25} = 158\ DA^{0.776} SL^{0.423} (I_{24,2} - 2.5)^{0.868}$	.150	36	5
$Q_{50} = 170\ DA^{0.777} SL^{0.445} (I_{24,2} - 2.5)^{0.847}$	.156	37	7
$Q_{100} = 181\ DA^{0.779} SL^{0.466} (I_{24,2} - 2.5)^{0.831}$	.163	39	9

Statistics of basin characteristics used in area 3 regression analysis

Basin characteristic	Maximum	Minimum	Mean	Median
DA	4,927 mi^2	0.31 mi^2	488 mi^2	85.1 mi^2
SL	149 ft/mi	2.0 ft/mi	17.2 ft/mi	9.0 ft/mi
$I_{24,2}$	3.15 in.	2.85 in.	2.97 in.	2.95 in.

Table 1.--Equations for estimating magnitude and frequency of floods on streams in Indiana--Continued

Area 4 (46 stations)

Equations	Standard error of estimate		Equivalent years of record
	Log units	Percent	
$Q_2 = 16.8\ DA^{0.435} SL^{0.528} L^{0.860} (I_{24,2} - 2.5)^{0.459}$	0.130	31	3
$Q_{10} = 24.1\ DA^{0.517} SL^{0.628} L^{0.769} (I_{24,2} - 2.5)^{0.445}$	.127	30	6
$Q_{25} = 27.4\ DA^{0.545} SL^{0.664} L^{0.741} (I_{24,2} - 2.5)^{0.448}$	.137	32	7
$Q_{50} = 29.6\ DA^{0.554} SL^{0.687} L^{0.738} (I_{24,2} - 2.5)^{0.458}$	.146	34	9
$Q_{100} = 32.0\ DA^{0.565} SL^{0.705} L^{0.730} (I_{24,2} - 2.5)^{0.464}$	.158	37	11

Statistics of basin characteristics used in area 4 regression analysis

Basin characteristic	Maximum	Minimum	Mean	Median
DA	1,224 mi^2	0.07 mi^2	110 mi^2	10.9 mi^2
SL	267 ft/mi	2.4 ft/mi	51.0 ft/mi	23.6 ft/mi
L	77.1 mi	0.3 mi	18.8 mi	8.6 mi
$I_{24,2}$	3.30 in.	2.80 in.	3.05 in.	3.05 in.

Table 1.--Equations for estimating magnitude and frequency of floods on streams in Indiana--Continued

Area 5 (35 stations)

Equations	Standard error of estimate		Equivalent years of record
	Log units	Percent	
$Q_2 = 45.5\ DA^{0.760}SL^{0.390}$	0.128	30	3
$Q_{10} = 67.7\ DA^{0.780}SL^{0.469}$	.138	33	5
$Q_{25} = 77.0\ DA^{0.790}SL^{0.499}$	.151	36	5
$Q_{50} = 83.8\ DA^{0.805}SL^{0.516}$	.163	39	7
$Q_{100} = 91.2\ DA^{0.811}SL^{0.529}$	.175	42	8

Statistics of basin characteristics used in area 5 regression analysis

Basin characteristic	Maximum	Minimum	Mean	Median
DA	11,125 mi^2	0.04 mi^2	583 mi^2	21.8 mi^2
SL	236 ft/mi	1.2 ft/mi	35.8 ft/mi	12.6 ft/mi

Table 1.--Equations for estimating magnitude and frequency of floods on streams in Indiana--Continued

Area 6 (32 stations)

Equations	Standard error of estimate		Equivalent years of record
	Log units	Percent	
$Q_2 = 681\ DA^{0.691} RC^{0.856} (I_{24,2} - 2.5)^{1.771}$	0.115	27	5
$Q_{10} = 2,177\ DA^{0.622} RC^{0.865} (I_{24,2} - 2.5)^{1.980}$	.125	29	7
$Q_{25} = 3,165\ DA^{0.598} RC^{0.852} (I_{24,2} - 2.5)^{2.035}$	.138	32	7
$Q_{50} = 3,908\ DA^{0.584} RC^{0.849} (I_{24,2} - 2.5)^{2.049}$	.146	34	10
$Q_{100} = 4,734\ DA^{0.570} RC^{0.834} (I_{24,2} - 2.5)^{2.068}$	.157	37	12

Statistics of basin characteristics used in area 6 regression analysis

Basin characteristic	Maximum	Minimum	Mean	Median
DA	856 mi^2	0.10 mi^2	164 mi^2	35.0 mi^2
RC	0.8	0.3	0.6	0.7
$I_{24,2}$	3.00 in.	2.70 in.	2.86 in.	2.85 in.

Table 1.--Equations for estimating magnitude and frequency of floods on streams in Indiana--Continued

Area 7 (22 stations)

Equations	Standard error of estimate		Equivalent years of record
	Log units	Percent	
$Q_2 = 22.6\ DA^{0.468} SL^{0.414} L^{0.624} RC^{0.846}$	0.109	26	3
$Q_{10} = 45.7\ DA^{0.350} SL^{0.439} L^{0.726} RC^{0.862}$	.122	29	4
$Q_{25} = 56.4\ DA^{0.318} SL^{0.458} L^{0.754} RC^{0.862}$	.137	32	4
$Q_{50} = 63.6\ DA^{0.300} SL^{0.473} L^{0.770} RC^{0.860}$	.149	35	5
$Q_{100} = 70.1\ DA^{0.285} SL^{0.488} L^{0.785} RC^{0.854}$	.160	38	6

Statistics of basin characteristics used in area 7 regression analysis

Basin characteristic	Maximum	Minimum	Mean	Median
DA	1,578 mi^2	0.39 mi^2	241 mi^2	99.8 mi^2
SL	39.7 ft/mi	0.9 ft/mi	7.4 ft/mi	2.7 ft/mi
L	78.6 mi	1.1 mi	23.7 mi	19.8 mi
RC	0.7	0.3	0.4	0.4

Basin characteristics were also determined for each of the 242 gaged locations (table 3, after References). The relation between peak-flow data and basin characteristics were analyzed by multiple-regression techniques. Detailed discussions of flood-frequency determination and multiple-regression analysis are presented later in the report.

On the basis of regression analysis, Indiana has been divided into seven areas (fig. 2). Equations for each area, to be used in estimating flood magnitude having recurrence intervals of 2, 10, 25, 50, and 100 years on unregulated, nonurban streams, are given in table 1. Statistics of the basin characteristics from the stations used in the regression analyses are also shown in the table. The estimating equations are valid at sites where the basin characteristics (particularly drainage area) are within the range listed for the seven areas in the table. Caution should be used when the basin characteristics of the ungaged site are outside the range of those used to develop the equations. The standard error of estimate (in log units and percent) and equivalent years of record for each equation included in table 1 are discussed in the section "Accuracy and Limitations."

Significant basin characteristics required to use the equations are defined as follows:

1. Contributing drainage area (DA), in square miles, is the area contributing directly to surface runoff. This area can be planimetered from topographic maps or can be obtained from the drainage-area report for Indiana (Hoggatt, 1975). Drainage area should be determined to the nearest 0.01 mi^2 in the range from 0.01 to 9.99 mi^2; to the nearest 0.1 mi^2, from 10.0 to 99.9 mi^2; and to the nearest 1 mi^2, for drainage areas greater than 99.9 mi^2.

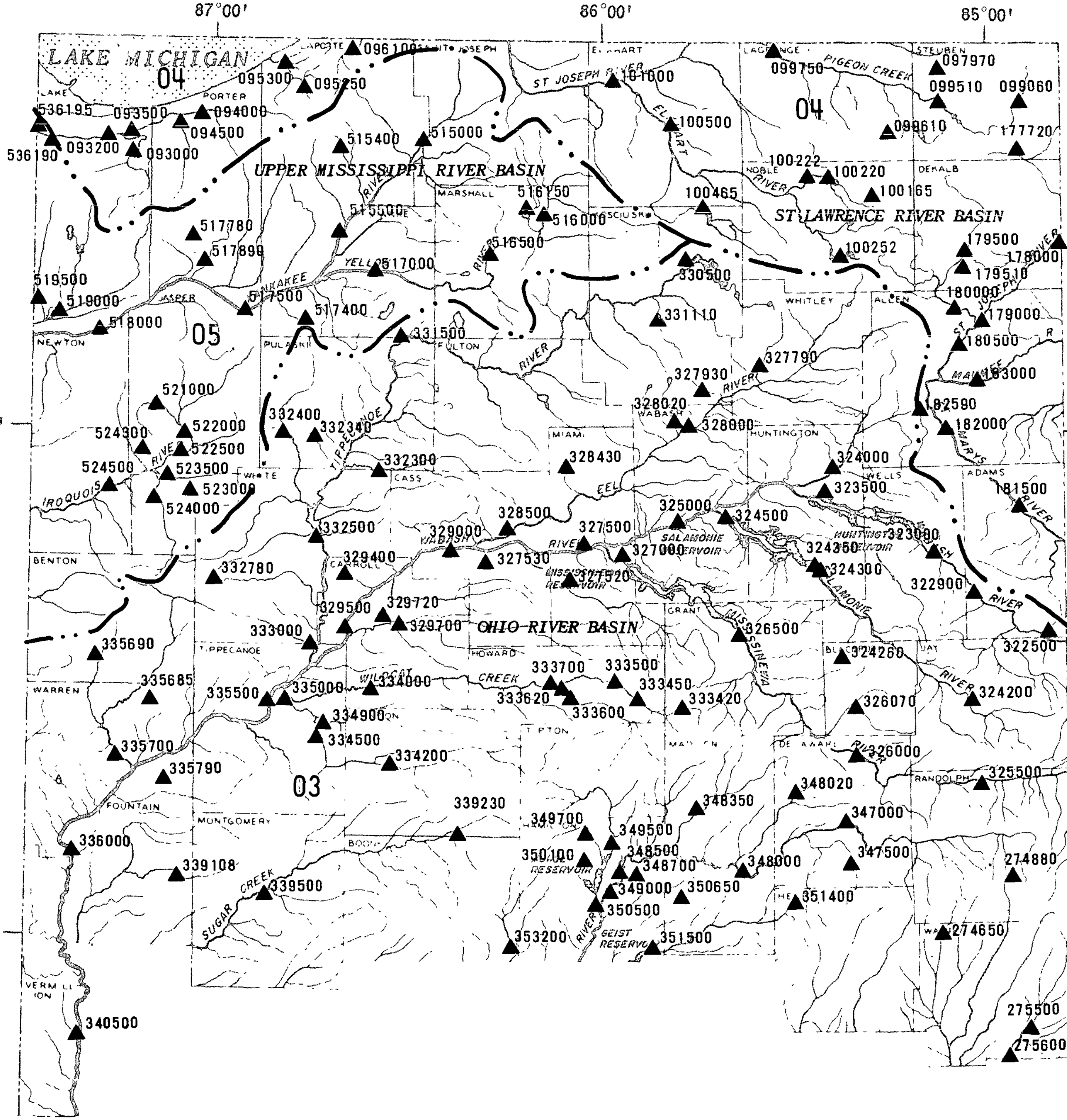
2. Main-channel slope (SL), in feet per mile, the slope of the streambed between points that are 10 and 85 percent of the distance from the location on the stream to the basin divide, is determined from topographic maps. Slope should be determined to the nearest 0.1 ft/mi.

3. Channel length (L), in miles, the distance measured along the main channel from the location on the stream to the basin divide, is determined from topographic maps. Length should be determined to the nearest 0.1 mi.

4. Storage (STOR), the percentage of the contributing drainage area covered by lakes, ponds, and wetlands, is determined from topographic maps. A constant of 1 percent is added to characteristic STOR for use in the estimating equations. Storage should be determined to the nearest 0.1 percent.

5. Mean annual precipitation (PREC), in inches, the 1941-70 average annual precipitation, is determined from figure 3 (Stewart, 1983). A constant of 30 inches is subtracted from the characteristic PREC for use in the estimating equations. The basin centroid should be plotted in figure 3, and mean annual precipitation for that point should be determined to the nearest 0.5 in. by interpolation between lines of equal precipitation.

6. Precipitation intensity ($I_{24,2}$), in inches, the maximum 24-hour precipitation having a recurrence interval of 2 years, is determined from figure 4 (Hershfield, 1961). A constant of 2.5 inches is subtracted from the characteristic $I_{24,2}$ for use in the estimating equations. The basin centroid


87°00'
86°00'
85°00'
41°00'
40°00'
LAKE MICHIGAN
04
05
03
UPPER MISSISSIPPI RIVER BASIN
ST LAWRENCE RIVER BASIN
OHIO RIVER BASIN
LAKE
PORTER
LAPORTE
ST JOSEPH
ELKHART
LAGRANGE
STEUBEN
NOBLE
DEKALB
MARSHALL
KOSCIUSKO
WHITLEY
ALLEN
JASPER
NEWTON
PULASKI
FULTON
WABASH
MIAMI
HUNTINGTON
WELLS
ADAMS
WHITE
CASS
BENTON
CARROLL
GRANT
HOWARD
TIPPECANOE
WARREN
TIPTON
FOUNTAIN
MONTGOMERY
BOONE
RANDOLPH
VERMILLION
ST JOSEPH RIVER
PIGEON CREEK
ELKHART
KANKAKEE
YELLOW
RIVER
IROQUOIS
TIPPECANOE
EEL
WABASH
SALAMONIE RESERVOIR
MISSISSINEWA RESERVOIR
MISSISSINEWA
SALAMONIE
MAUMEE
ST MARYS
WILDCAT CREEK
SUGAR CREEK
GEIST RESERVOIR
096100
095300
095250
536195
536190
093500
094000
094500
093200
093000
515400
515000
101000
100500
099750
097970
099510
099060
099610
177720
100222
100220
100165
100465
516150
516000
515500
517780
517890
516500
517000
179500
178000
179510
100252
330500
519500
519000
518000
517500
517400
331500
331110
180000
179000
180500
327790
327930
183000
521000
328020
328000
182590
182000
332400
332340
524300
522000
522500
524500
523500
523000
524000
332300
328430
324000
323500
181500
325000
324500
332500
329000
328500
327500
327530
327000
327520
324350
323000
324300
329400
332780
322900
329500
329720
329700
333000
326500
322500
335690
324260
333700
333500
333450
333620
333600
333420
335685
335500
335000
334000
334900
334500
326070
324200
326000
335700
335790
334200
325500
348020
339230
348350
347000
349700
349500
348500
350100
348700
348000
347500
274880
336000
339108
339500
349000
350650
350500
351400
353200
351500
274650
340500
275500
275600

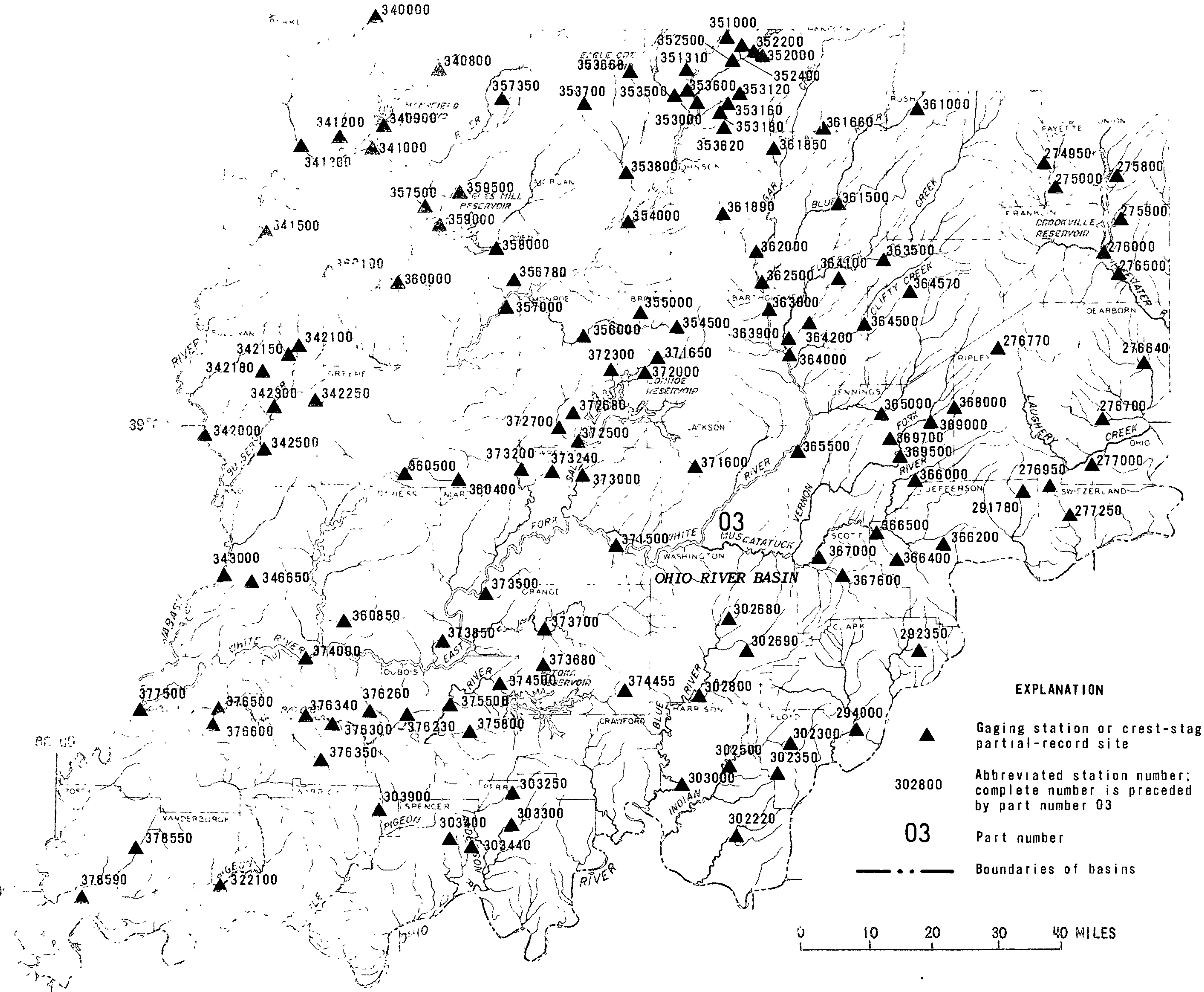


Figure 1.-- Locations of streamflow data-collection sites in Indiana.

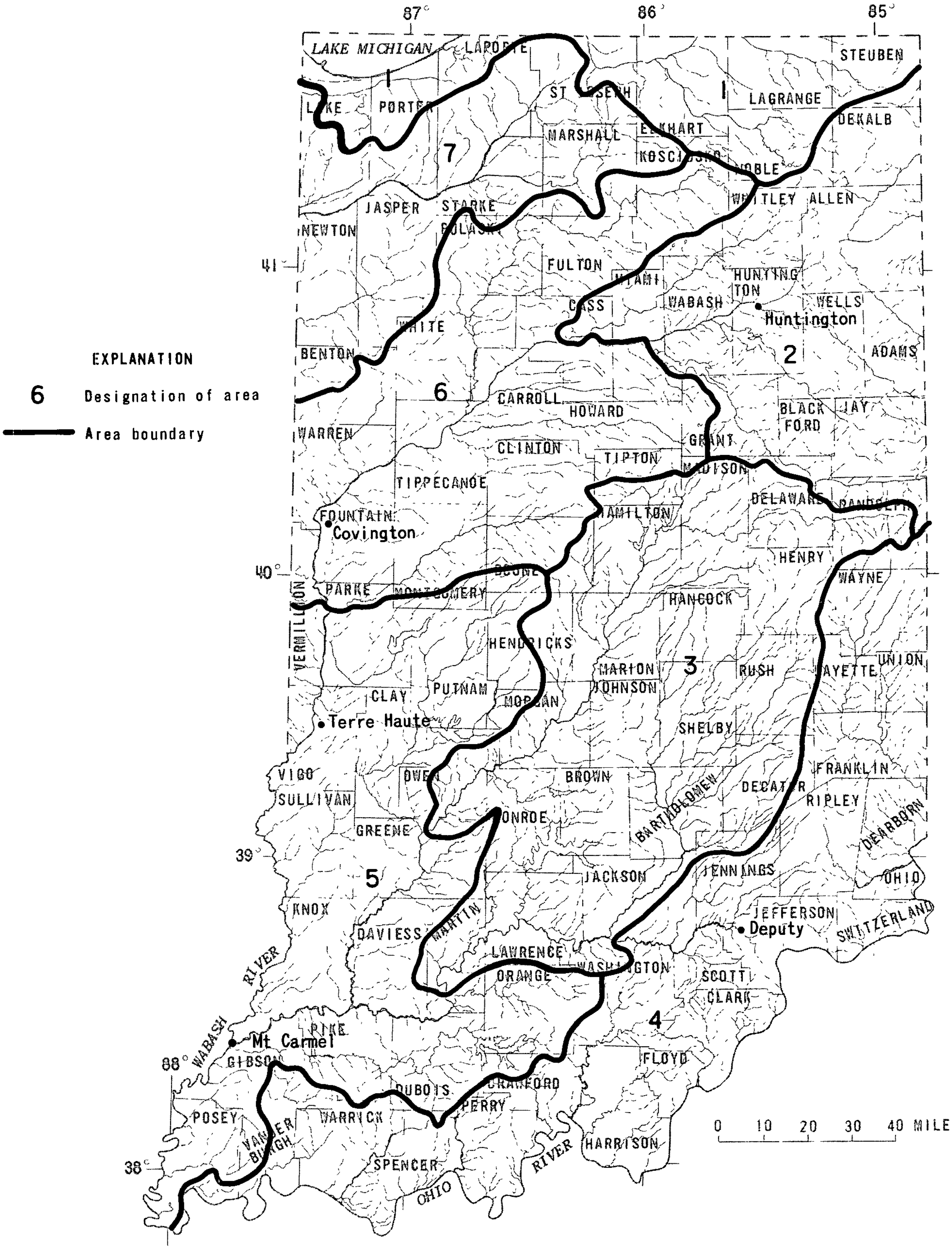
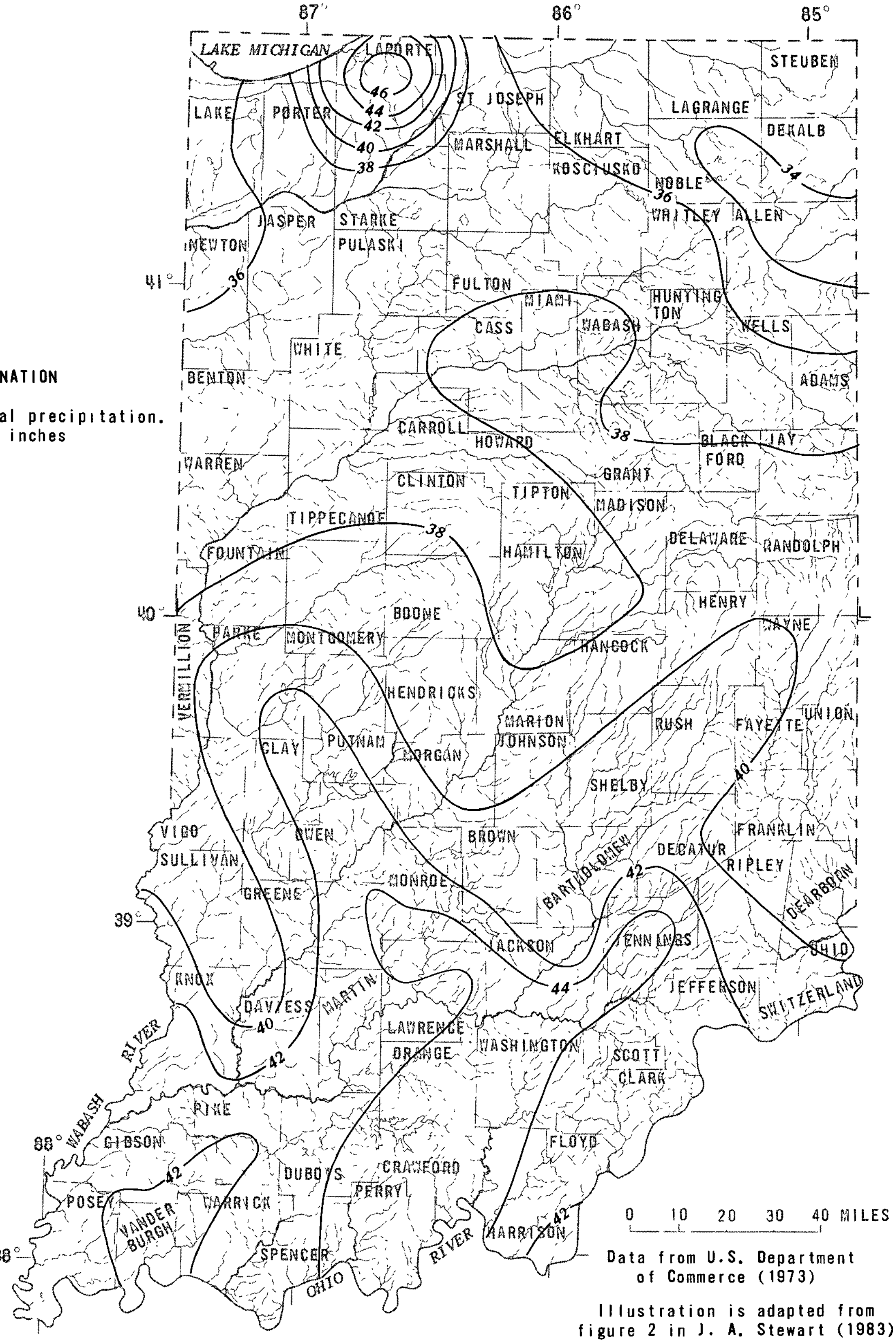


Figure 2.-- Areas for selecting flood-frequency estimating equations.

EXPLANATION

—42— Line of equal precipitation. Interval 2 inches



Data from U.S. Department of Commerce (1973)

Illustration is adapted from figure 2 in J. A. Stewart (1983)

Figure 3.-- Mean annual precipitation, 1941-70.

should be plotted in figure 4, and precipitation intensity for that point should be determined to the nearest 0.05 in. by interpolation between lines of equal precipitation.

7. Runoff coefficient (RC), a coefficient that relates storm runoff to soil permeability by major hydrologic soil groups, is determined from figure 5 (Davis, 1975). Values of the coefficient (fig. 5) range from 0.30, for hydrologic soil-group A, to 1.00, for hydrologic soil-group E. If the drainage area covers more than one hydrologic soil group, the runoff coefficient should be an areally weighted average determined to the nearest 0.05.

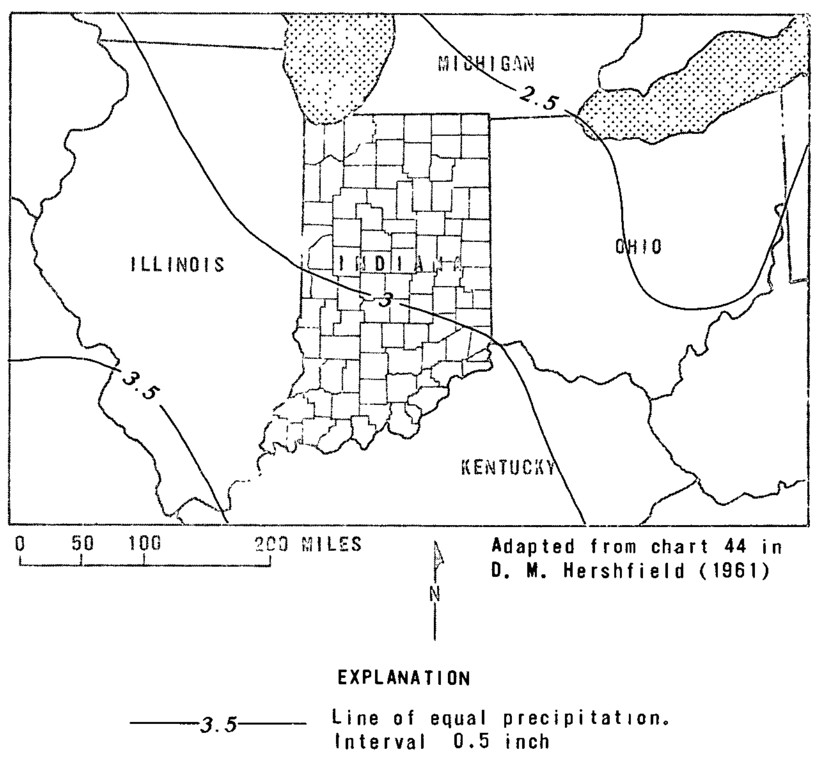


Figure 4.-- Two-year, 24-hour precipitation.

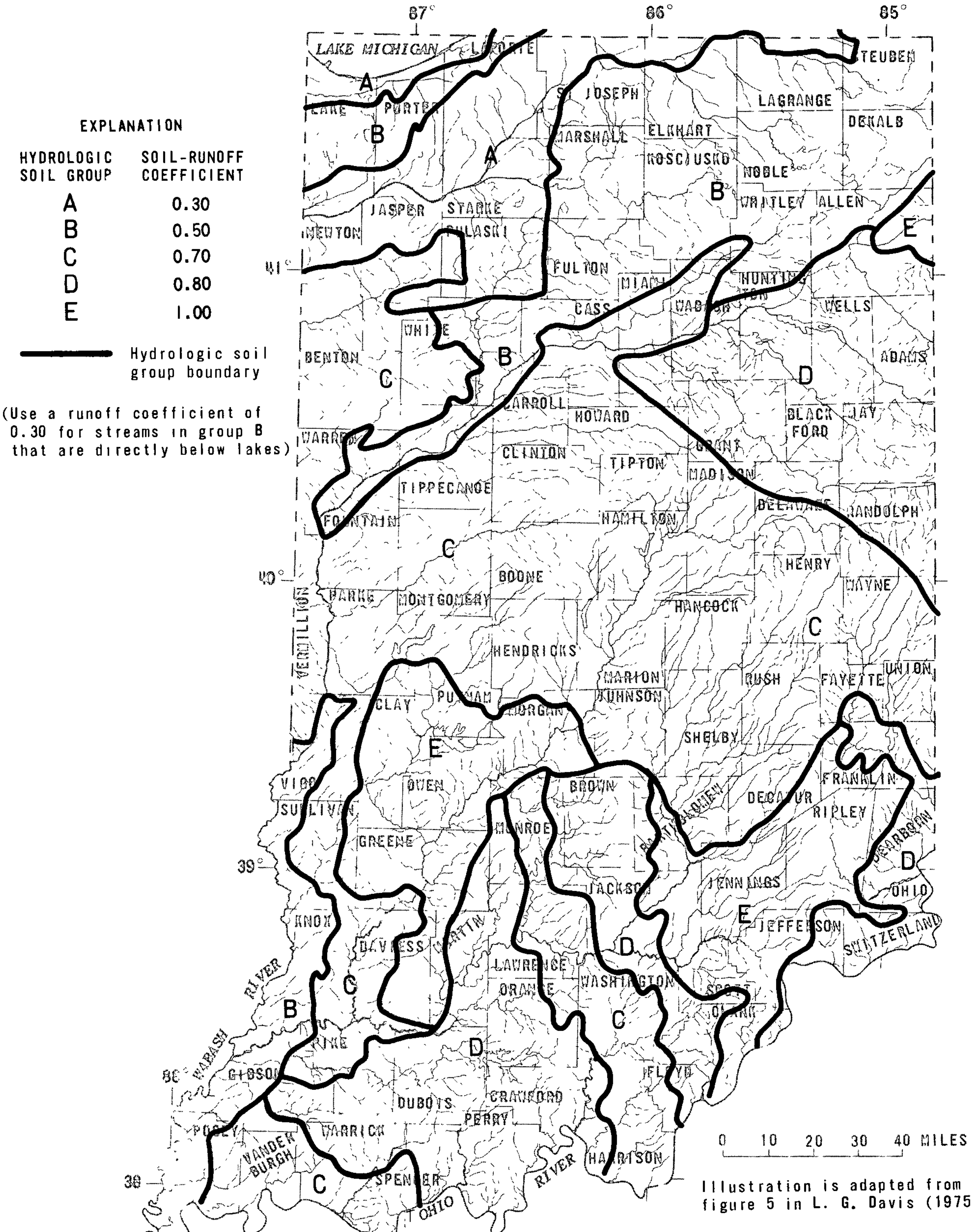


Figure 5.-- Major hydrologic soil groups.

Use of the estimating equations is shown in the following example: A highway engineer is given the task of designing a culvert to pass the 100-year flood on a small stream in Brown County, Ind. From figure 2, the location of the site is found to be in area 3. The equations for estimating flood peaks in area 3 (table 1) require contributing drainage area (DA), channel slope (SL), and 24-hour, 2-year rainfall ($I_{24,2}$) as independent variables. Physical characteristics of the basin determined from a topographic map are as follows:

contributing drainage area, 6.94 mi^2;

channel length, 4.40 mi;

elevation of the channel at a point 10 percent of the length (0.4 mi) upstream, 652 ft;

elevation of the channel at a point 85 percent of the length (3.7 mi) upstream, 824 ft;

distance between points 10 and 85 percent of the length upstream, 3.7 - 0.4 = 3.3 mi;

channel slope, $\frac{824 - 652}{3.3} = 52.1$ ft/mi.

From figure 3, the 24-hour, 2-year precipitation is determined to be 3.05 in.

The equation for estimating the 100-year peak discharge for a site on an unregulated, nonurban stream in area 3 (table 1) is:

$$Q_{100} = 181 \ DA^{0.779} \ SL^{0.466} \ (I_{24,2} - 2.5)^{0.831}.$$

Substituting the values of basin characteristics for the ungaged site in the equation yields:

$$Q_{100} = 181 \times 6.94^{0.779} \times 52.1^{0.466} \times (3.05 - 2.5)^{0.831} = 3,140 \ ft^3/s.$$

Sites on Gaged Streams

Unregulated and Nonurban Gaged Streams

Flood magnitude having a specific recurrence interval can be estimated for a site on an unregulated, nonurban stream by one of the following procedures:

1. If the site is at a gaged location, the weighted estimate of Q_T from table 4 (after References) should be used.

2. If the drainage area of an ungaged site on a gaged stream is less than 50 percent or greater than 150 percent of the drainage area of a gaged site on the same stream, the discharge should be estimated from the appropriate equation in table 1 as if the site were on an ungaged stream. An example showing how to use the estimating equations is shown in the section "Sites on Ungaged Streams."

3. If the drainage area of an ungaged site on a gaged stream is between 50 and 150 percent of the drainage area of a gaged site on the same stream, the discharge should be an estimate calculated from both gaged data (table 4) and estimating equations (table 1). An estimate of the T-year peak discharge at an ungaged site is determined by first computing the ratio:

$$R = \frac{Q_{T_W} \text{ (gaged site)}}{Q_{T_R} \text{ (gaged site)}},$$

where Q_{T_W} (gaged site) is the weighted estimate of the T-year flood at the gaged site and Q_{T_R} (gaged site) is the estimate of the T-year flood at the gaged site determined by a regional estimating equation (table 1). This ratio is the correction needed to adjust the regional value to the weighted value at the gaged site. Values of Q_{T_W} and Q_{T_R} for 245 gaged sites are listed in table 4. The weighting factor (R_W) to be applied to the estimate of Q_T at the ungaged site is computed as:

$$R_W = R - \frac{2\Delta A}{A_G}(R - 1),$$

where R is the ratio defined above, ΔA is the absolute value of the difference between the drainage areas of the gaged and ungaged sites, and A_G is the drainage area of the gaged site. The T-year peak discharge at the ungaged site is then determined by the equation:

$$Q_T = Q_{T_R} \text{ (ungaged site)} \times R_W,$$

where Q_{T_R} (ungaged site) is the estimate of the T-year flood at the ungaged site determined by a regional estimating equation (table 1) and R_W is the weighting factor defined above. The effect of R_W is phased out as ΔA increases to 50 percent of A_G.

Procedures for use in estimating peak discharge at a specific recurrence interval at a gaged site and at an ungaged site on the same stream and near the gaged site are given in the examples that follow.

If an estimate of the 100-year peak discharge is needed for the gaging station on the Muscatatuck River near Deputy, Ind. (03366500), one can be obtained from table 4. The table contains three estimates of Q_{100} for this station: The upper number (40,900 ft^3/s) is from flood-frequency analysis of the observed data, the middle number (44,600 ft^3/s) is from the regression equation for area 4 (table 1), and the lower number (41,200 ft^3/s) is from weighting the two independent estimates. The weighting procedure and analysis of observed peak data are described in the section "Flood-Frequency Analysis." The best estimate of Q_{100} for the gaging station on the Muscatatuck River would be the weighted estimate, 41,200 ft^3/s.

An estimate of Q_{100} is also needed on the Muscatatuck River downstream from the gaging station near Deputy, Ind. (03366500). Q_{100} for the ungaged location is first estimated by the regression equation for area 4 (table 1) which is of the form:

$$Q_{100} = 32.0\ DA^{0.565}\ SL^{0.705}\ L^{0.730}\ (I_{24,2} - 2.5)^{0.464}.$$

From topographic maps and figure 4, basin characteristics for the ungaged site are determined to be: DA, 359 mi^2; SL, 6.2 ft/mi; L, 68.8 mi; and $I_{24,2}$, 3.00 inches. By substitution:

$$Q_{100} = 32.0 \times 359^{0.565} \times 6.2^{0.705} \times 68.8^{0.730} \times (3.00 - 2.5)^{0.464}$$
$$= 51,200\ ft^3/s.$$

Because the drainage area at the ungaged site is between 50 and 150 percent of the drainage area at the gaged location this number is then weighted by a factor that reflects how well the estimating equations match values from flood-frequency analysis of observed peaks at the gaged location. In the previous example, the equation for estimating Q_{100} in area 4 produced 44,600 ft^3/s at the gaging station near Deputy, Ind. Flood-frequency analysis of the station record gave 40,900 ft^3/s for Q_{100}. These two estimates were combined by a weighting technique previously mentioned to produce the weighted estimate 41,200 ft^3/s for Q_{100} at the gaging station. The weighting factor to be applied to the estimate of Q_{100} from the regression equation at the ungaged location is calculated as follows:

$$R_W = R - \frac{2\Delta A}{A_G}(R - 1).$$

By substitution:

$$R_W = \frac{41,200}{44,600} - \frac{(2)(359 - 293)}{293}\left(\frac{41,200}{44,600} - 1\right) = 0.958.$$

The best estimate of Q_{100} at the ungaged site on the Muscatatuck River then becomes:

$$Q_T = 51,200 \times 0.958 = 49,000\ ft^3/s.$$

Regulated Gaged Streams

Flood magnitude and frequency at gaged sites on regulated streams should be estimated on the basis of the best available streamflow data for that site, not on estimating equations. Peak-discharge data are available for many sites on regulated streams in Indiana. Gaging stations on streams affected by regulation are listed in table 5 (after References). The period of record for each station shown in table 5 was split at the time when regulation began. If more than 10 years of unregulated annual-peak data were available, an unregulated flood-frequency curve was determined (table 4). Stations marked with an asterisk (*) in table 4 are currently (1983) regulated, but flood-frequency data from the unregulated period of record at these sites were used in the regression analysis to develop estimating equations. Flood-frequency estimates for current (1983) conditions at each of these stations should be based on peak data from the regulated period and used with caution.

Annual peak discharges affected by regulation were not used in determining flood-frequency curves for use in developing estimating equations. However, if the period of record during regulated flow is of sufficient length, these data can be used to estimate flood magnitude and frequency at a specific site on a regulated stream under current (1983) conditions. Flood-frequency data for eight such sites are shown in table 4. Flow characteristics at sites on regulated streams could be greatly altered by a change in regulatory practices; peak data should be thoroughly reviewed before a flood-frequency analysis is made. Regulated and unregulated peak data should not be combined in determining the flood-frequency curve for a gaged site.

An example of the effect of regulation on flood frequency was obtained by analysis of peak-discharge data from stations on the Wabash River. Streamflow in the Wabash River in the reach downstream from Huntington, Ind., has been regulated since 1968 by flood-control reservoirs operated by the U.S. Army Corps of Engineers. Huntington Reservoir (717 mi^2), Salamonie Reservoir (553 mi^2), and Mississinewa Reservoir (807 mi^2) control more than 25 percent of the drainage area of the Wabash River from Huntington, Ind., to Covington, Ind. Reservoirs on tributaries control a small part of the drainage area of the Wabash River from Terre Haute, Ind., to Mt. Carmel, Ill.

The magnitude and frequency of floods based on analysis of unregulated annual peaks through 1967 at 12 gaging stations on the Wabash River from Huntington to Mt. Carmel are shown in table 6 (after References). Estimates of flood magnitude and frequency for the period of regulated flow (Indiana Department of Natural Resources, 1981) are also given in the table. Regulation has substantially reduced the estimate of flood magnitude for all recurrence intervals at each station.

Urban Gaged Streams

Flood-frequency data from six gaged sites on urban streams are listed in table 4 but were not used in the regression analyses to develop the estimating equations. The data are presented for use in estimating flood magnitude and frequency at specific locations under current (1983) conditions. Peak discharge on an urban stream is dependent on the degree of urbanization within the basin. The imperviousness of the land surface associated with an urban basin is generally greater than that of a nonurban basin, and peak discharge from an urban basin is generally larger than that from a nonurban basin of similar size. Thus, the estimating equations shown in table 1 could underestimate flood magnitude. Conversely, ponding behind a highway embankment, with available storage capacity and with a culvert to allow outflow, could reduce the peak discharge on an urban stream. In this case, flood magnitude in the channel downstream from the embankment could be overestimated by use of the equations shown in table 1. No methodolgy is given in this report for estimating flood magnitude and frequency at ungaged sites on urban streams.

Accuracy and Limitations

The accuracy of the estimating equations in table 1 is expressed as standard error of estimate (log units and percent) and equivalent years of record. The standard error of estimate is a measure of how well the discharges determined by the equations compare with the discharges from the individual station flood-frequency curves that were used to develop the equations. Because of the transformation of the variables to corresponding base 10 logarithmic values before regression analysis, the standard error of estimate was determined in log units and was converted to percent and equivalent years of record by techniques given in Hardison (1971). On the average, two-thirds of the observations of discharge from flood-frequency curves based on observed data lie within one standard error of estimate (expressed in log units) of corresponding values computed by the equations. For example, the standard error of estimate for the Q_{100} equation in area 1 is 0.186 log unit. This means that two-thirds of the time logarithms of the Q_{100} values from flood-frequency analysis of observed peaks will be within 0.186 log unit of the logarithms of the Q_{100} values computed from the equation for area 1. The standard error of 0.186 log unit was converted to 45 percent by the conversion table in Hardison (1971). The standard error of estimate in log units was also converted to equivalent years of record by use of Hardison's equation:

$$N_U = R^2[\overline{I}_V/SE]^2,$$

where N_U is equivalent years of record, R is a function of recurrence interval and mean logarithmic skew, $\overline{I}_V$ is mean logarithmic standard deviation, and SE is the standard error in log units. Using this equation and the statistical analyses of flood frequency for stations in area 1, the author converted the standard error of estimate (0.186 log unit) to an accuracy equivalent of 5 years. Thus, the estimate of a 100-year peak discharge at a site in area 1 computed from the estimating equation has an accuracy similar to that obtained by flood-frequncy analysis of 5 years of peak-discharge data collected at the site.

Split-sampling techniques were used in area 3 to verify the predictive accuracy of the estimating equations. The 60 stations in area 3 were divided into two sets, one set for developing equations and the other for measuring the accuracy of prediction by the equations. The stations were first arranged by size of drainage area and were then alternately assigned to the predicting and estimating sets, beginning with the smallest and ending with the largest. This procedure of data splitting resulted in an estimating set of 30 stations and a predicting set of 30 stations. A regression analysis using data from the 30 stations in the estimating set produced an equation for Q_{100} having a standard error of estimate of 39 percent (0.163 log unit). Independent variables in the equation were the same ones shown to be significant by analysis of data from all 60 stations in area 3. Using this equation, the author computed peak discharges having a 100-year recurrence interval for the 30 stations in the predicting set. The standard error of estimate of the observed values of Q_{100} for stations in the predicting set compared with Q_{100} values for these stations computed by the equation from analysis of data in the estimating set is 46 percent. This approximates the standard error of estimate (39 percent) for Q_{100} where data from all 60 stations in the area were used.

The equations in table 1 are for estimating magnitude and frequency of floods on unregulated, nonurban streams. Statistics of the basin characteristics used in developing the individual area equations are also given in the table. The equations are valid at sites where the basin characteristics fall within the range shown in the table. The equations should not be used for estimating discharge on an urban or a regulated stream; the flood-frequency curve reflecting current conditions at a site should be used in planning and design. No methodology is given for estimating flood magnitude and frequency at ungaged sites on urban or regulated streams.

DATA ANALYSIS

Annual peak-discharge data and basin characteristics from 242 continuous-record and crest-stage partial-record stations having at least 10 years of observed record were used in a multiple-regression analysis to develop equations for estimating magnitude and frequency of floods. Synthetic peak-discharge data generated by a rainfall-runoff model were used to extend the length of record at 11 stations. (See section "Extending Length of Record by a Rainfall-Runoff Model.") Locations of the 236 stations in Indiana used in developing the estimating equations are shown in figure 1. Locations of 26 stations on regulated streams and 6 stations on urban streams not used in developing the estimating equations are also shown in figure 1. Six stations used in the regression analysis (three in Ohio, and three in Illinois) are not shown in figure 1.

Long-term daily and unit-precipitation data for use in the rainfall-runoff model were obtained from the National Oceanic and Atmospheric Administration (NOAA) for Indianapolis, Ind.; Fort Wayne, Ind.; Chicago, Ill.; Peoria, Ill.; Springfield, Ill.; Cairo, Ill.; and Louisville, Ky. Long-term daily-evaporation data for use in the model were obtained for Oaklandon, Ind. (Geist Reservoir).

Peak-discharge frequency data and basin characteristics were determined for each gaged site on naturally flowing streams in Indiana. The State of Indiana was divided into seven areas on the basis of regression analysis. Flood-frequency equations for each of the seven areas were developed by multiple-regression techniques. These equations can be used to estimate the magnitude and frequency of floods on any unregulated, nonurban stream in Indiana.

Flood-Frequency Analysis

Flood-frequency analyses were done for 270 continuous-record stations and crest-stage partial-record sites having at least 10 years of peak-flow data to determine flood-frequency curves for each site. For these analyses, guidelines of the U.S. Water Resources Council (1981) were used to fit the logarithms of the annual peak discharges to a Pearson type-III distribution. Historical peaks and high outliers were given weight, low outliers were omitted, and station skew was weighted with skew values from a generalized skew map in the reference.

The technique for fitting a log-Pearson type-III distribution to observed annual peak discharges is to compute the base 10 logarithm of the discharge (Q) at a selected probability of occurence (P) by the equation:

$$\log Q = \bar{x} + KS,$$

where $\bar{x}$ is the mean of the logarithms of the annual peak discharges, S is the standard deviation of the logarithms of the annual peak discharges, and K is a function of the WRC-weighted skew coefficient (G) and the selected probability of occurrence (P). Values of K can be obtained from U.S. Water Resources Council (1981). A summary of the statistics of the logarithms of the annual peak discharges used in developing flood-frequency curves for the gaged sites is shown in table 2.

Flood-frequency analysis is done to define the relation of flood magnitude (instantaneous maximum discharge) to probability of occurrence or to recurrence interval. Probability of occurrence (P) is the percent chance of a given flood magnitude being exceeded in any 1 year. Recurrence interval (T), which is the reciprocal of the probability of occurrence multiplied by 100, is the average number of years between exceedances of a given flood magnitude. The recurrence interval is an average interval, and the occurrence of floods is random in time; no schedule of regularity is implied. The occurrence of a flood having a 50-year recurrence interval (2-percent probability of occurrence) is no guarantee, therefore, that a flood of equal or greater magnitude will not occur the following year, or even the following week.

Results of flood-frequency analysis of observed annual peaks at 259 individual stations are given in table 4. (Flood-frequency data for 12 stations on the Wabash River downstream from Huntington Reservoir are shown in table 6.) Peak discharges having recurrence intervals of 2, 10, 25, 50, and 100 years estimated by analysis of the observed data are shown in table 4 as the upper number for each station. Because the T-year flood estimated from the log-Pearson type-III distribution of the logarithms of the annual peak discharges and the corresponding estimate from the regression equations (table 1) are considered to be independent, a technique for weighting the two estimates is recommended (Curtis, 1977a, p. 4). The best estimate of flood magnitude at a selected recurrence interval for a gaged location is obtained by the equation:

$$\log Q_T = \frac{(\text{sta yrs rec})(\log \text{sta } Q_T) + (\text{eq yrs rec})(\log \text{reg } Q_T)}{(\text{sta yrs rec}) + (\text{eq yrs rec})}.$$

In the preceding equation, log sta Q_T (log station Q_T) is the upper number for each site in table 4 converted to a logarithm; sta yrs rec (station years of record) is determined from table 2; log reg Q_T (log regression Q_T) is computed as the logarithm of the discharge computed by the estimating equations in table 1 or obtained from table 4 (middle number); and eq yrs rec (equivalent years of record, which is the accuracy of the regression equation) is determined from table 1. The antilog of the calculated log Q_T is the best estimate of flood magnitude at a selected frequency. Weighted estimates of flood magnitude and frequency at each of the stations used in the regression analysis are shown as the lower number in table 4.

Extending Length of Record by a Rainfall-Runoff Model

A long-term record (60-70 years) of synthetic flood peaks was generated for each of 11 stations on small streams by a rainfall-runoff model developed by the U.S. Geological Survey (Dawdy and others, 1971; and Carrigan and others, 1977). The purpose of generating the synthetic data was to increase the effective length of record at the small-stream gaging stations, where short-term concurrent rainfall and discharge data had been collected. Flood hydrographs for each station were generated from daily-rainfall, daily-evaporation, and unit-rainfall data. The model deals with three components of the hydrologic cycle--antecedent soil moisture, storm infiltration, and surface-runoff routing. The two phases involved in using the model are calibration and synthesis.

During calibration of the model, daily rainfall, daily pan evaporation, and concurrent values of unit streamflow and unit rainfall were used to optimize the 10 parameters defined in table 7. Seven of the parameters define the volume of surface runoff, and three control the shape of the flood hydrograph. Several parameters are considered to vary only slightly (Lichty and Liscum, 1978). By holding these parameters constant, the fitting process improves the values of the remaining parameters. The values of DRN and TP/TC were held constant at 1.000 and 0.500 throughout the calibration. Optimum values of the 10 parameters obtained in calibrating the model are shown in table 7 for each of the 11 rainfall-runoff stations.

The optimum values of parameters from calibration of the rainfall-runoff model were used with long-term precipitation and evaporation data provided by the National Oceanic and Atmospheric Administration to generate a long-term series of flood peaks. Long-term evaporation data from Oaklandon, Ind. (Geist Reservoir), was used for each of the 11 gaging stations. However, data from seven long-term-precipitation stations were available for use in synthesis of long-term peak discharge. The choice of which long-term-precipitation record to use was based on techniques in Lichty and Liscum (1978) and Curtis (1977b).

At each of the seven long-term-precipitation stations, synthetic data were generated, and rainfall-runoff model estimates of T-year floods were related to the parameters of the model. Replicate synthesis using the optimum model parameters from each of the 11 gaging stations resulted in 77 synthetic

Table 7.--Results from calibration of the rainfall-runoff model

PSP Product of moisture deficit and suction at the wetted front for soil moisture at field capacity.
KSAT The minimum (saturated) hydraulic conductivity used to determine infiltration rates.
DRN A constant drainage rate for redistribution of soil moisture.
RGF Ratio of the product of moisture deficit and suction at the wetted front for soil moisture at the wilting point to that at field capacity.
BMSM Soil moisture storage volume at field capacity.
EVC Coefficient to convert pan evaporation to potential evapotranspiration.
RR Proportion of daily rainfall that infiltrates the soil.
KSW Time characteristic for linear reservoir routing.
TC Length of the base of the triangular translation hydrograph.
TC/TP Ratio of time to peak to base length of the triangular translation hydrograph.

Station number	PSP	KSAT	DRN	RGF	BMSM	EVC	RR	KSW	TC	TP/TC
03275800	3.998	0.287	1.000	16.664	3.904	0.864	0.855	0.316	10.4	0.500
03276640	1.017	.062	1.000	28.958	2.653	.808	.771	.344	21.7	.500
03324260	3.182	.335	1.000	22.820	7.018	.704	.740	.298	44.6	.500
03329720	2.759	.061	1.000	20.180	2.606	.873	.890	5.721	113.0	.500
03334200	.807	.114	1.000	10.716	1.486	.810	.960	6.989	51.3	.500
03335790	4.003	.145	1.000	11.860	2.625	.967	.730	.397	61.4	.500
03352400	3.374	.042	1.000	23.270	8.337	.778	.911	2.265	97.6	.500
03364100	1.514	.085	1.000	11.880	2.793	.952	.616	2.850	49.1	.500
03366400	1.712	.053	1.000	39.206	3.359	.968	.618	.316	43.2	.500
03373680	4.954	.164	1.000	15.390	2.507	.655	.755	.176	15.4	.500
03378590	1.894	.049	1.000	11.847	1.227	.584	.913	1.063	25.7	.500

annual-flood series (11 gaging stations times 7 precipitation records). A log-Pearson type-III distribution was used to quantify synthetic T-year flood estimates for each of the 77 synthetic peak-data sets.

Regression analyses were used to relate the synthetic estimate of peak discharge at a specified recurrence interval (Q_{T_S}) to a combination of optimum parameters from the rainfall-runoff model (table 7) that define the volume and shape of the hydrograph. The equation for estimating flood magnitude and frequency at a rainfall-runoff station from precipitation data collected at a long-term-precipitation station is as follows:

$$Q_{T_S} = a\ VAR^{b}\ FR^{c}\ DA,$$

where

Q_{T_S} is the synthetic T-year flood estimate, in cubic feet per second, based on precipitation data collected at the respective precipitation station,

a the regression constant,

VAR[1] an index of the dispersion about the mean arrival time (lag), in hours, that describes the hydrograph shape,

FR[2] the infiltration rate, in inches per hour, that describes the hydrograph volume,

b and c the regression coefficients,

and

DA the contributing drainage area, in square miles.

Regression analysis showed the regression coefficient "b" to be constant for all stations and recurrence intervals, and the regression coefficient "c" to be a function of "a", the regression constant. The equation for estimating the synthetic discharge for a specific recurrence interval was transformed to:

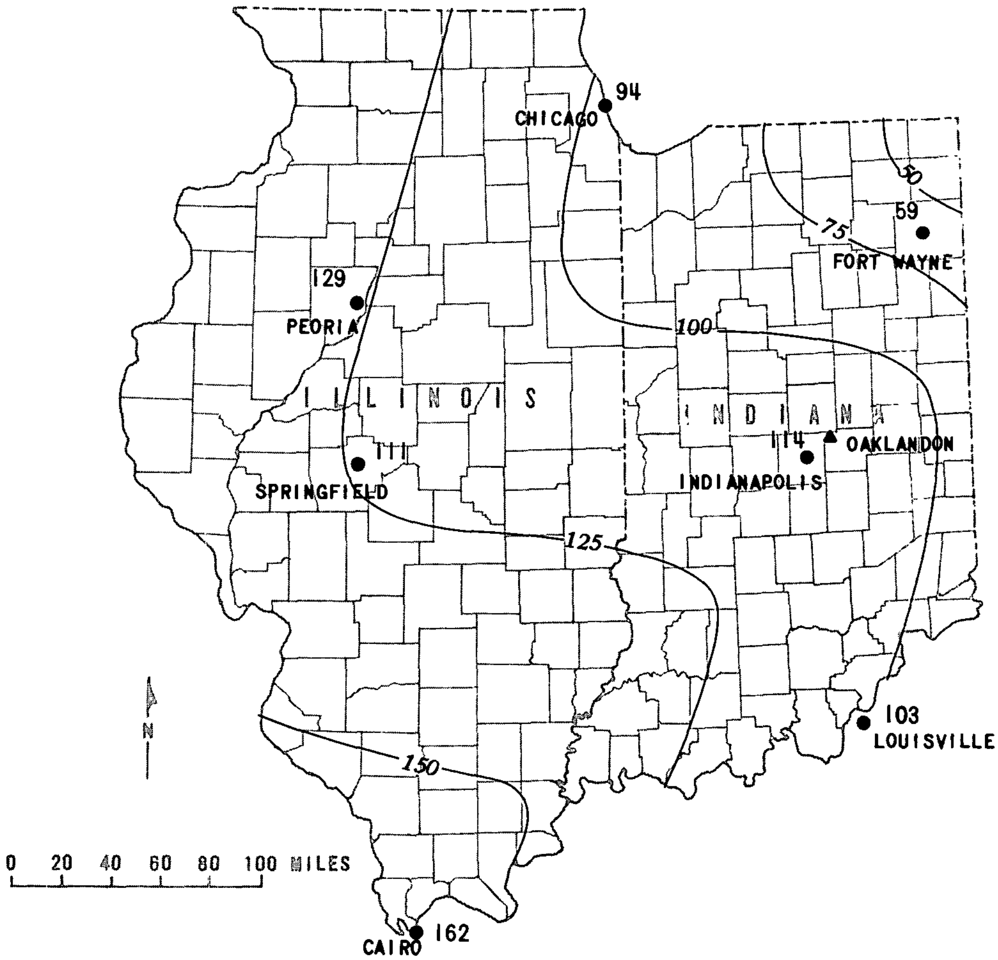
$$Q_{T_S} = a_T\ VAR^{-0.310}\ FR^{0.790 \log a_T - 2.266}\ DA.$$

The average standard error of estimate of Q_{T_S} was less than 20 percent. The only variable in this equation is "a_T" because VAR, FR, and DA are known for a given set of model parameters. Site-to-site variability in the magnitude of the regression coeffecient "a_T" is interpreted as reflecting the spatially varying influence of local climatic factors. Values of "a_T" for each of the seven long-term precipitation stations were plotted on a map for recurrence intervals of 2, 10, 25, 50, and 100 years (figs. 6-10). Lines of equal climatic factor drawn on each of the five maps can be used to estimate "a_T" for any location in Indiana. Values of "a_T", DA, VAR, and FR for the 11 rainfall-runoff stations are listed in table 8. Values of Q_{T_S} were calculated from these data.

---

[1]VAR is defined by the equation $VAR = KSW^2 + (TC/60)^2/24$

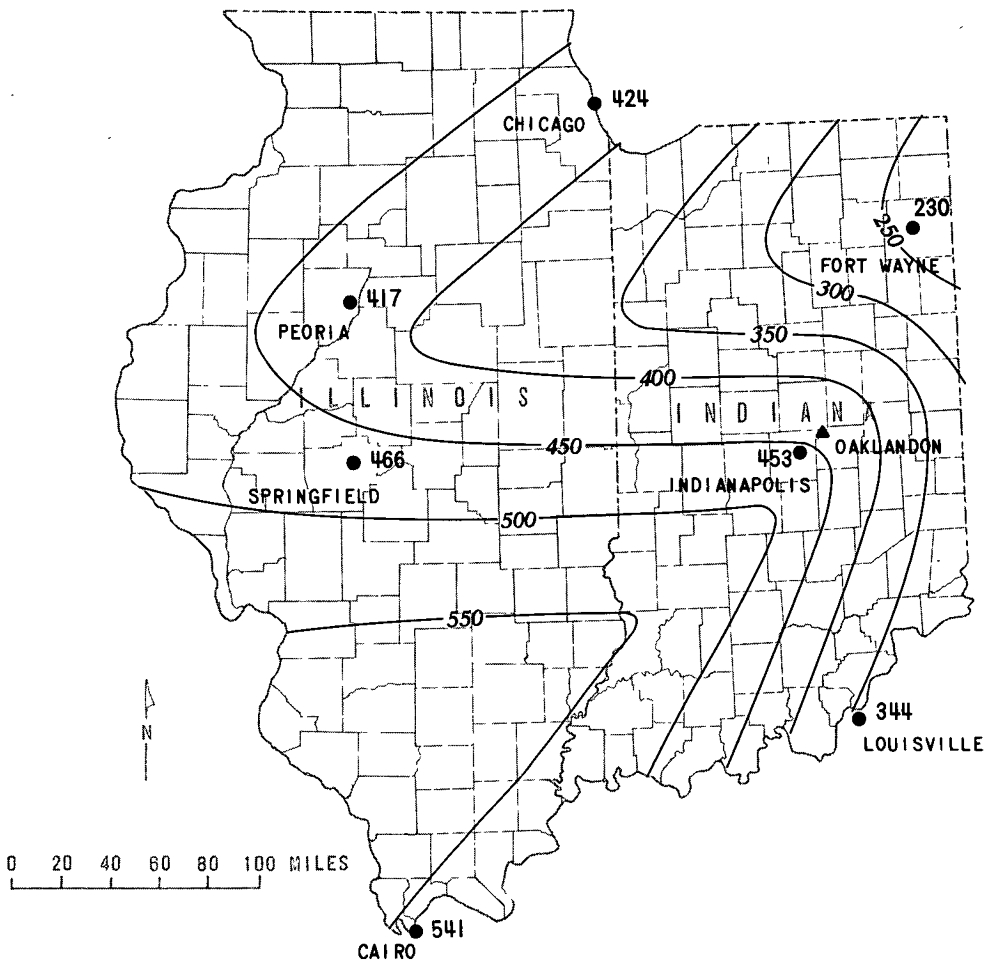
[2]FR is defined by the equation $FR = KSAT\ [1.0 + 0.50\ PSP(0.15\ RGF + 0.85)]$



EXPLANATION

—— 125 ——	Line of equal climatic factor a_2
114	Point value of climatic factor a_2
●	Long-term precipitation station
▲	Long-term evaporation station

Figure 6.-- Climatic factor a_2 for estimating synthetic Q_2 at a rainfall-runoff station.

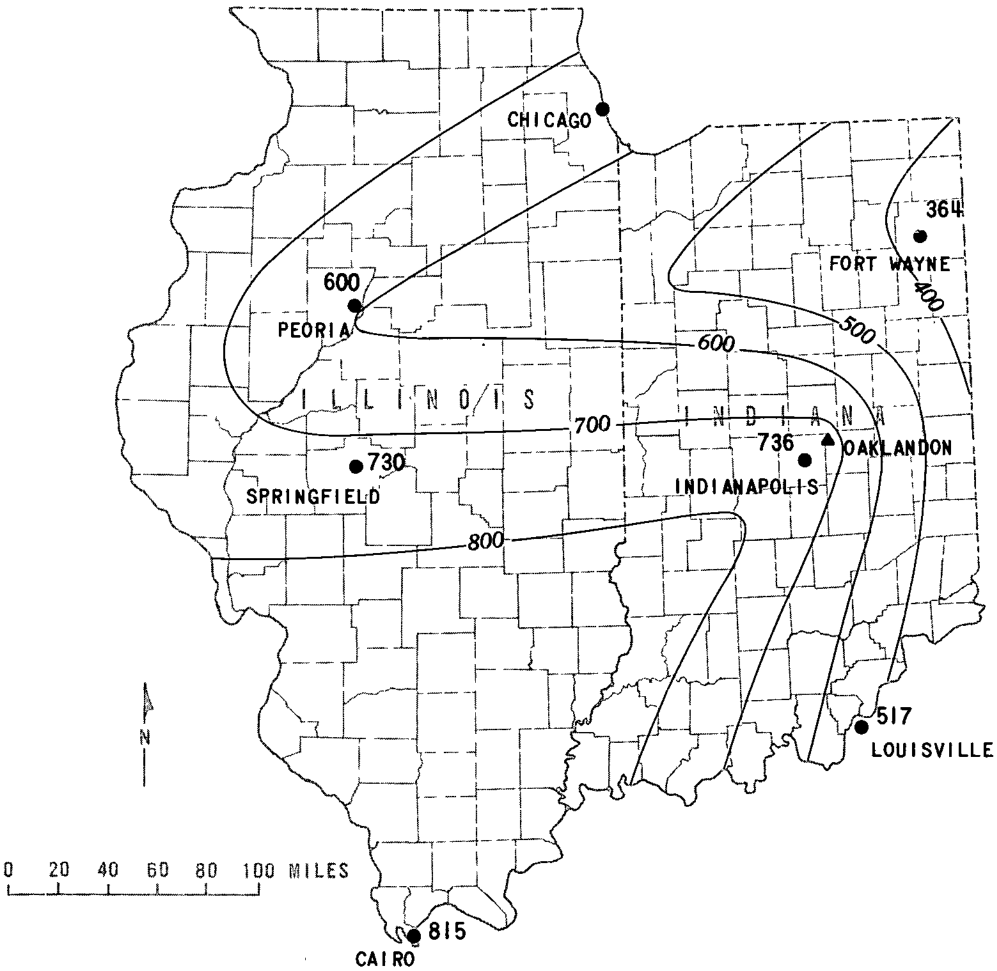


EXPLANATION

—350—	Line of equal climatic factor a_{10}
344	Point value of climatic factor a_{10}
●	Long-term precipitation station
▲	Long-term evaporation station

Figure 7.-- Climatic factor a_{10} for estimating synthetic Q_{10} at a rainfall-runoff station.

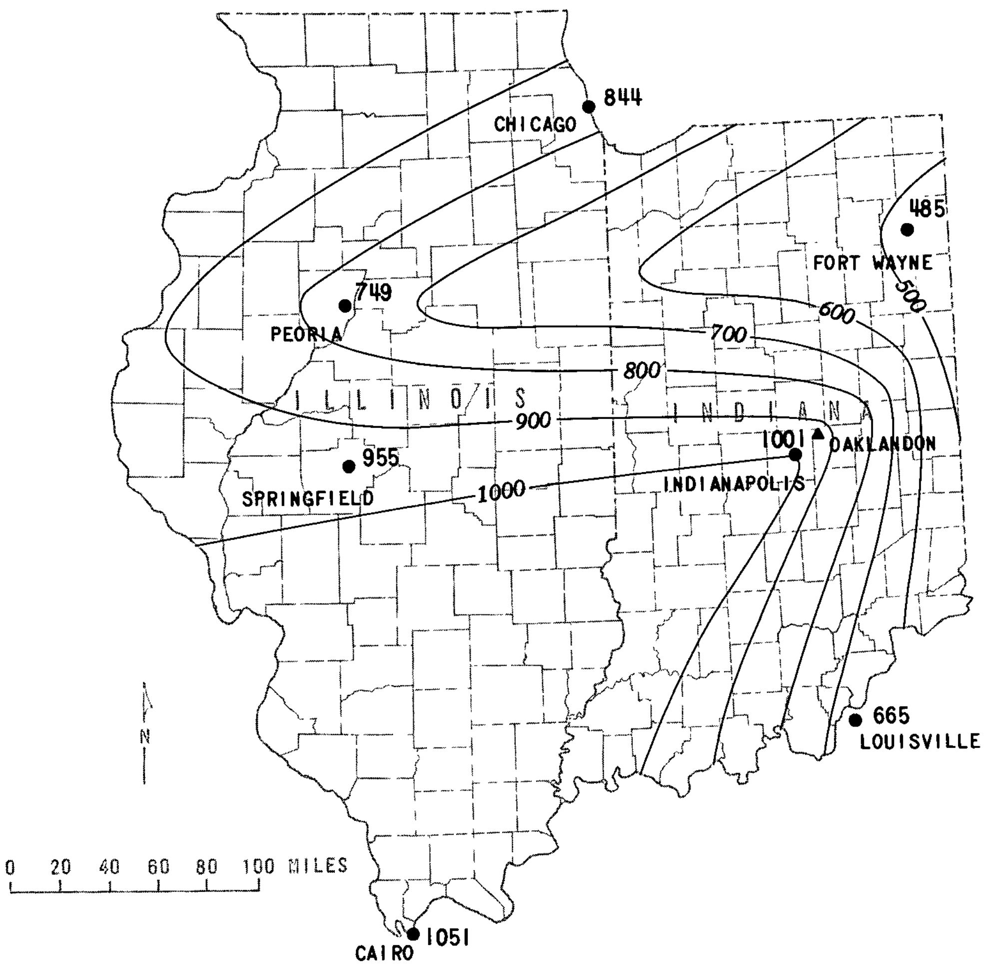
655



EXPLANATION

—— *600* ——	Line of equal climatic factor a_{25}
517	Point value of climatic factor a_{25}
●	Long-term precipitation station
▲	Long-term evaporation station

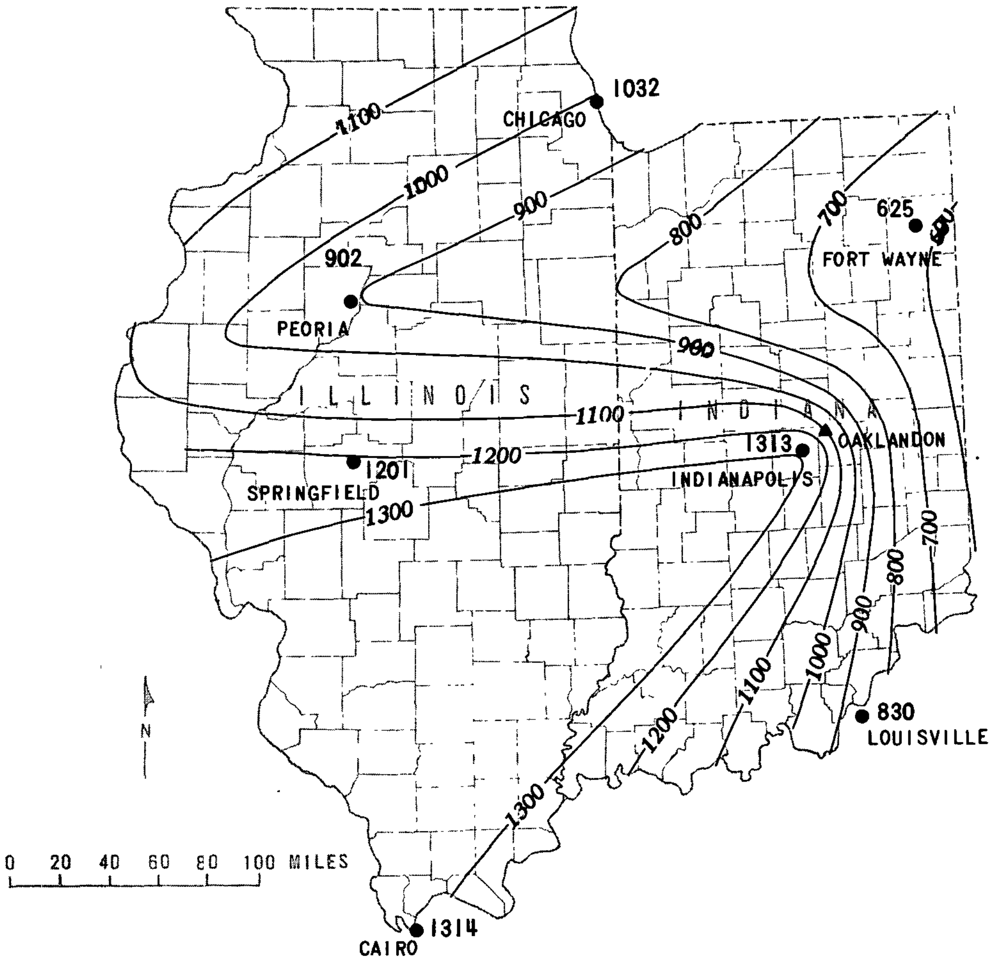
Figure 8.— Climatic factor a_{25} for estimating synthetic Q_{25} at a rainfall-runoff station.



EXPLANATION

—700—	Line of equal climatic factor a_{50}
665	Point value of climatic factor a_{50}
●	Long-term precipitation station
▲	Long-term evaporation station

Figure 9.-- Climatic factor a_{50} for estimating synthetic Q_{50} at a rainfall-runoff station.



EXPLANATION

—900—	Line of equal climatic factor a_{100}
1201	Point value of climatic factor a_{100}
●	Long-term precipitation station
▲	Long-term evaporation station

Figure 10.-- Climatic factor a_{100} for estimating synthetic Q_{100} at a rainfall-runoff station.

The preceding method was used to eliminate the need to select data from a single long-term-precipitation station to estimate synthetic Q_T at each of the 11 stations used in the modeling procedure. Furthermore, synthetic Q_T can also be estimated at any additional rainfall-runoff station whose record is adequate to define VAR and FR; synthesis of annual peak discharges at the various long-term-precipitation stations is not required.

Table 8.--Data used to estimate synthetic Q_T at rainfall-runoff stations

[DA is the drainage area, in square miles. VAR is an index of the dispersion about the mean arrival time (lag), in hours, that describes the hydrograph shape. FR is the infiltration rate, in inches per hour, that describes the hydrograph volume. a_T is the T-year climatic factor (from figs. 6-10)]

Station number	DA	VAR	FR	a_2	a_{10}	a_{25}	a_{50}	a_{100}
03275800	0.26	0.101	2.209	95	340	450	550	650
03276640	.19	.124	.226	90	300	400	500	600
03324260	.86	.112	2.612	95	320	500	600	700
03329720	5.62	32.878	.387	100	340	550	650	800
03334200	2.61	48.877	.227	110	400	650	800	950
03335790	1.22	.201	.908	110	400	650	800	1,000
03352400	.77	5.241	.350	115	450	700	1000	1,250
03364100	1.46	8.150	.254	110	420	650	800	1,000
03366400	.16	.152	.358	100	350	500	650	800
03373680	.29	.034	1.447	120	450	700	900	1,100
03378590	.32	1.138	.171	140	540	800	1,050	1,300

The synthetic flood-frequency curve was combined with the flood-frequency curve based on the 10 years of observed data, and the resultant flood-frequency data at each site was used in the regression analysis to develop estimating equations. A weighting procedure based on an analysis of variance (Lichty and Liscum, 1978, p. 21) and on equivalent years of record (W. O. Thomas, oral commun., 1983) was used to develop the final flood-frequency curve for each of the rainfall-runoff stations. In this procedure, the flood-frequency curves developed from the synthetic and the observed data are assumed to be unbiased and independent.

A value of equivalent years of record for the synthetic estimates of Q_T was determined from statistics of observed data by the equation:

$$N = R^2 \left(\frac{I_V}{SE_p}\right)^2,$$

where

N is equivalent years of record;

R a factor based on skew and recurrence interval relating standard error of a T-year flood to I_V and N (from Hardison, 1971, p. C230);

I_V the index of variability, equal to the standard deviation of the logarithms of the annual peaks (from Lichty and Liscum, 1978, p. 21);

and

SE_p the standard error of prediction, equal to the square root of the average variance of the synthetic estimate (from Lichty and Liscum, 1978, p. 29).

The weighting factor applied to the observed estimates of peak discharge (Q_T) was determined as the ratio of years of observed data to total years of record (observed and synthetic). The weighting factor applied to the synthetic estimate of peak discharge (Q_{T_S}) was equal to one minus the observed weighting factor. All information needed to determine the combined (weighted) flood-frequency curve at a rainfall-runoff station is given in table 9. Constant values of I_V (equal to S) and G were taken from Lichty and Liscum (1978, p. 21); R values were taken from Hardison (1971, p. C230); and SE_p values were determined by the equation $SE_p = \sqrt{VMM}$ and data in Lichty and Liscum (1978, p. 29). Data from the combined flood-frequency curves at the 11 rainfall-runoff stations were included in regression analysis to develop the estimating equations and are shown as the upper number in table 4.

Sample calculations to determine factors for weighting synthetic and observed estimates of a 25-year flood follow: Given: $I_V = S = 0.298$, $G = -0.109$, R (for $T = 25$ and $G = -0.109$) $= 1.512$, and $SE_p = \sqrt{VMM}$ (for $T = 25$) $= \sqrt{0.0110} = 0.105$. Substituting these values into the equation to calculate equivalent years of record gives:

$$N = R^2 \left(\frac{I_v}{SE_p}\right)^2 = (1.512)^2 \left(\frac{0.298}{0.105}\right)^2 = 18.4 \text{ years.}$$

The synthetic estimate of Q_{25} has an equivalency equal to 18.4 years compared to the 10 years of observed data. The weighting factor applied to the observed estimate of Q_{25} is:

$$\frac{N_{obs}}{N_{obs} + N_{syn}} = \frac{10}{10 + 18.4} = 0.35.$$

The weighting factor applied to the synthetic estimate of Q_{25} is:

$$1.00 - 0.35 = 0.65.$$

Therefore, the weighted estimate of Q_{25} from combining the observed and the synthetic frequency curves is calculated by the equation:

$$Q_{25_{weighted}} = (0.35)(Q_{25_{observed}}) + (0.65)(Q_{25_{synthetic}}).$$

Table 9.--Equations used to combine observed and synthetic estimates of Q_T at rainfall-runoff stations

[I_v and G are constant for all recurrence intervals. I_v, which is S in Lichty and Liscum (1978, p. 21), is 0.298. G, from Lichty and Liscum (1978, p. 21), is -0.109. T is the recurrence interval. R, from Hardison (1971, p. C230), is a function of T and G. SE_p is the square root of VMM, from Lichty and Liscum (1978, p. 29). N_{obs} is the number of observed peaks. N_{syn} is the equivalent years of record for the synthetic estimate of Q_T. $Q_{T_{obs}}$ is the estimate of Q_T from observed data. $Q_{T_{syn}}$ is the estimate of Q_T from synthetic data. $Q_{T_{wt}}$ is the weighted estimate of Q_T from combining $Q_{T_{obs}}$ and $Q_{T_{syn}}$.]

For T = 2 years, R = 0.999, SE_p = 0.140, N_{obs} = 10, and N_{syn} = 4.5

$$Q_{2wt} = (0.70)(Q_{2obs}) + (0.30)(Q_{2syn})$$

For T = 10 years, R = 1.298, SE_p = 0.102, N_{obs} = 10, and N_{syn} = 14.4

$$Q_{10wt} = (0.40)(Q_{10obs}) + (0.60)(Q_{10syn})$$

For T = 25 years, R = 1.512, SE_p = 0.105, N_{obs} = 10, and N_{syn} = 18.4

$$Q_{25wt} = (0.35)(Q_{25obs}) + (0.65)(Q_{25syn})$$

For T = 50 years, R = 1.841, SE_p = 0.112, N_{obs} = 10, and N_{syn} = 24.0

$$Q_{50wt} = (0.30)(Q_{50obs}) + (0.70)(Q_{50syn})$$

For T = 100 years, R = 2.192, SE_p = 0.118, N_{obs} = 10, and N_{syn} = 30.7

$$Q_{100wt} = (0.25)(Q_{100obs}) + (0.75)(Q_{100syn})$$

Regression Analysis

Multiple-regression analysis was used to develop the relation between flood magnitudes having 2-, 10-, 25-, 50-, and 100-year recurrence intervals (table 4, upper number) and basin characteristics (table 3) for 242 gaged locations in Indiana, Ohio, and Illinois. Independent variables (basin characteristics) and dependent variables (peak-flow statistics) were transformed to base 10 logarithms before analysis by multiple-regression techniques, and the equations were developed in log-linear form. Equations for estimating flood frequency are presented so that information from sites where peak data are available can be transferred to ungaged locations. These equations, which relate the most significant basin characteristics to peak discharge at specific recurrence intervals, are of the form:

$$\log Q_T = \log a + b \log A + c \log B + d \log C + \ldots n \log N$$

or

$$Q_T = a\, A^b\, B^c\, C^d \ldots N^n,$$

where Q_T is the flood magnitude, in cubic feet per second, having a recurrence interval of T years;

a the regression constant;

A, B, C...N the basin characteristics;

and b, c, d...n the regression coefficients.

Forward selection, backward elimination, and maximum R^2 improvement regression analyses described in Helwig and Council (1979) were done on data from throughout the State and on data from the seven areas used to define flood-frequency relations. For each area, the equations with the lowest standard error of estimate, independent variables significant at the 90-percent confidence level, and logical regression coefficients were chosen to estimate flood magnitude at recurrence intervals of 2, 10, 25, 50, and 100 years.

The State of Indiana was divided into seven areas (fig. 2) on the basis of regression analysis. Initially, basin characteristics and flood-frequency data from all 242 gaged sites were analyzed as a single area. Standard errors of estimate ranged from 38 percent for the 2-year flood to 50 percent for the 100-year flood. Grouping the stations by physiographic region and rerunning the regression analysis did not produce standard errors of estimate lower than those determined from analysis of a single area. Residuals (observed value minus the value computed from the estimating equation) from the single-area analysis were plotted on a State map. Stations were grouped by major river basins into areas having similar residuals and regression analysis was done on data from the stations in each area. The residuals from these analyses were plotted on a map. Stations were reassigned from one area to another on the basis of the residuals and regression analyses rerun. If the standard errors of estimate for both areas decreased, and regression coefficients were logical, the stations were kept in the new area; if not, the stations were kept in the old area. Some large areas were split into smaller subareas, and analyses were run on data from each of them. Conversely, several areas were combined into one large area for analysis. The regrouping of stations continued until no further improvement resulted from the reassignment of stations from one area to another. The groupings then consisted of seven geographical areas. Except for the White River and East Fork White River, basins drained by unregulated streams were not divided; all stations within a basin were reassigned from one area to another. Residuals from the final estimating equations were plotted against independent and dependent variables, and no trends were detected in the plots.

Equations for each of the seven areas, their accuracy, and statistical information about the independent variables used in the regression are shown in table 1. Standard errors of the equations are shown in log units and in percent. Accuracy of the equations is also shown as the number of years of record needed at an ungaged location in the area to produce an estimate as good as that produced by the equation (equivalent years of record).

Basin characteristics used as independent variables in the regression analyses included contributing drainage area, channel length, channel slope, average elevation, storage, forested area, mean annual precipitation (Stewart, 1983), precipitation intensity of a 24-hour, 2-year storm (Hershfield, 1961), precipitation index (Davis, 1974); and a soil runoff coefficient (Davis, 1975). Of these, average elevation, forested area, and precipitation index were insignificant in the estimating equations. Various combinations of the remaining variables were used in the final estimating equations. The equations for the individual areas are valid for all unregulated, nonurban streams in the area.

Additional analyses were done to determine whether equations for estimating flood magnitude and frequency could be developed on the basis of drainage-area size as was done by Davis (1974). Davis presented one equation for all locations draining 15 to 100 mi^2 and another for all locations draining more than 200 mi^2 (except the Wabash and White Rivers). He used a weighting procedure on streams draining 100 to 200 mi^2. Separating the 242 stations into groups based on drainage-area size did not produce equations with lower standard errors of estimate than those based on location (dividing the State into seven areas). The results of the analyses are shown in table 10. Comparison of the standard errors of estimate for the two sets of equations shows that equations based on location are better for estimating peak discharge on all sizes of streams and that the standard error of estimate for streams draining less than 100 mi^2 is virtually constant.

Split-sampling techniques were used to analyze the data from area 3 and to verify that the standard errors of estimate shown in table 1 are representative of the predictive accuracy of the estimating equations. Split sampling is discussed in the section "Accuracy and Limitations."

Table 10.--Standard errors of estimate of Q_{100} for area equations and equations developed by grouping stations according to size of drainage basin

Drainage area (square miles)	Number of stations	Standard error of estimate of Q100	
		area equations	basin size equations
<10	77	41 percent	56 percent
<15	84	41 percent	56 percent
<20	90	42 percent	54 percent
<50	117	43 percent	61 percent
<100	133	42 percent	58 percent
>100	109	32 percent	33 percent
>0.1	242	37 percent	50 percent

SUMMARY

Methods for estimating the magnitude and frequency of floods on any unregulated, nonurban stream in Indiana are given in this report. The State was divided into seven areas, and a set of equations for estimating peak discharges with recurrence intervals of 2, 10, 25, 50, and 100 years was developed for each area. Peak-discharge and basin-characteristics data from 242 gaging stations and crest-stage partial-record stations in Indiana and nearby Ohio and Illinois were used in multiple-regression analysis to develop the equations. A log-Pearson type-III frequency distribution based on guidelines of the U.S. Water Resources Council (1981) was used to develop flood-frequency curves for the individual stations. Basin characteristics shown to be significant in estimating flood magnitude included drainage area, channel length, channel slope, mean annual precipitation, precipitation intensity, storage, and a runoff coefficient. Standard errors of estimate ranged from 24 to 45 percent.

Peak-flow data synthesized by a rainfall-runoff model was used to extend the length of record at 11 small-stream gaging stations. The synthetic data were used to develop a flood-frequency curve for each station. These curves and flood-frequency curves developed from observed data were then combined into one curve for each station for use in regression analysis.

Flood-frequency data from stations on regulated and urban streams are presented for use in estimating flood magnitude and frequency at specific locations under current (1983) conditions. No methodology is given in the report for estimating magnitude and frequency of floods at ungaged sites on regulated or urban streams.

REFERENCES

Carrigan, P. H., Jr., Dempster, G. G., Jr., and Bower, D. E., 1977, User's guide for U.S. Geological Survey rainfall-runoff models--revision of Open-File Report 74-33: U.S. Geological Survey Open-File Report 77-884, 274 p.

Curtis, G. W., 1977a, Technique for estimating magnitude and frequency of floods in Illinois: U.S. Geological Survey Water-Resources Investigations Report 77-117, 70 p.

_____1977b, Frequency analysis of Illinois floods using observed and synthetic streamflow records: U.S. Geological Survey Water-Resources Investigations Report 77-104, 32 p.

Davis, L. G., 1974, Floods in Indiana: Technical manual for estimating their magnitude and frequency: U.S. Geological Survey Circular 710, 40 p.

_____1975, A manual for estimating the magnitude and frequency of floods on ungaged streams in Indiana: Indiana Department of Natural Resources, Division of Water Technical Memorandum 75-1.

Dawdy, D. R., Lichty, R. W., and Bergman, J. M., 1972, A rainfall-runoff simulation model for estimation of flood peaks for small drainage basins: U.S. Geological Survey Professional Paper 506-B, 28 p.

Gold, R. L., 1980, Flood magnitude and frequency of streams in Indiana: preliminary estimating equations: U.S. Geological Survey Open-File Report 80-759, 44 p.

Hardison, C. H., 1971, Prediction error of regression estimates of streamflow characteristics at ungaged streams, in Geological Survey Research 1971: U.S. Geological Survey Professional Paper 750-C, p. C228-C236.

Helwig, J. T., and Council, K. A., 1979, SAS user's guide 1979 edition: Raleigh, N.C., SAS Institute Inc., 494 p.

Hershfield, D. M., 1961, Rainfall frequency atlas of the United States: U.S. Department of Commerce Technical Paper 40, 115 p.

Hoggatt, R. E., 1975, Drainage areas of Indiana Streams: Indiana Department of Natural Resources, Division of Water, 231 p.

Indiana Department of Natural Resources, 1981, Coordinated discharges of selected streams in Indiana: Indiana Department of Natural Resources, Division of Water.

Lichty, R. W., and Liscum, F., 1978, A rainfall-runoff modeling procedure for improving estimates of T-year (annual) floods for small drainage basins: U.S. Geological Survey Water-Resources Investigations Report 78-7, 44 p

Stewart, J. A., 1983, Low-flow characteristics of Indiana streams: U.S. Geological Survey, Open-File Report 82-1007, 277 p.

U.S. Department of Commerce, 1973, Monthly normals of temperature, precipitation, and heating and cooling days, 1941-70: Ashville, N.C., U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Environmental Data Services, Climatography of the United States No. 81 (Indiana), 10 p.

U.S. Water Resources Council, 1981, Guidelines for determining flood flow frequency: Bulletin 17B of the Hydrology Committee, 183 p.