

Prepared in cooperation with the  
Tennessee Department of Transportation

# Flood-Frequency Prediction Methods for Unregulated Streams of Tennessee, 2000

Water-Resources Investigations Report 03-4176





# Flood-Frequency Prediction Methods for Unregulated Streams of Tennessee, 2000

By George S. Law and Gary D. Tasker

---

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 03-4176

Prepared in cooperation with the  
Tennessee Department of Transportation

Nashville, Tennessee  
2003

U.S. DEPARTMENT OF THE INTERIOR  
GALE A. NORTON, Secretary

U.S. GEOLOGICAL SURVEY  
CHARLES G. GROAT, Director

Any use of trade, product, or firm names in this report is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

---

For additional information write to:

District Chief  
U.S. Geological Survey  
640 Grassmere Park, Suite 100  
Nashville, Tennessee 37211

Copies of this report may be purchased from:

U.S. Geological Survey  
Branch of Information Services  
Federal Center  
Box 25286  
Denver, Colorado 80225

# CONTENTS

- Abstract..... 1
- Introduction ..... 1
  - Purpose and scope ..... 2
  - Previous studies ..... 2
  - Description of study area ..... 3
  - Acknowledgments ..... 3
- Basin characteristics ..... 3
  - Hydrologic areas ..... 6
  - Physiographic-region factor ..... 8
- Flood-frequency prediction methods..... 8
  - Unregulated, gaged sites ..... 9
    - Recurrence intervals ..... 9
    - Bulletin 17B method..... 9
  - Unregulated, ungaged sites..... 10
    - Regional-regression method ..... 11
    - Region-of-influence method..... 16
    - Comparison of methods..... 18
    - Use of computer application..... 19
- Application of methods ..... 22
- Summary..... 23
- References ..... 24
- Appendix A. Calculation of the prediction error and prediction interval for flood-frequency predictions at unregulated sites in Tennessee..... 63
  - Table A-1. Model error variance ( $\gamma^2$ ) for the single-variable and multivariable regional-regression equations in tables 6 and 7..... 64
  - Table A-2. Matrix  $\{X^T \Lambda^{-1} X\}^{-1}$  for the single-variable regional-regression equations in table 6 ..... 64
  - Table A-3. Matrix  $\{X^T \Lambda^{-1} X\}^{-1}$  for the multivariable regional-regression equations in table 7..... 65
- Appendix B. Description of detailed output file produced by the region-of-influence method for Tennessee..... 67
  - Table B-1. Detailed output file produced by the region-of-influence method for Tennessee ..... 69
- Appendix C. Computing effective record length when historical information is available ..... 77
- Appendix D. Calculation of equivalent years of record for regression-predicted peak discharges ..... 79

## FIGURES

- 1. Map showing gaging stations, hydrologic areas, and physiographic provinces in the study area..... 4
- 2-4. Graphs showing:
  - 2. Regional-regression equations for the 25-year flood for Tennessee..... 15
  - 3. A segmented regional-regression equation used for hydrologic area 3 ..... 15
  - 4. Segmented regional-regression equation and region-of-influence regression equation for the 25-year flood at a 2,000-square-mile ungaged site in hydrologic area 3 ..... 18
- 5. Sample of summary output file produced by flood-frequency computer application ..... 21

## TABLES

1. Number of gaging stations by hydrologic area and state .....	6
2. Basin characteristics .....	7
3. Physiographic-region factor equations .....	8
4. Selected basin characteristics and flood-frequency estimates for 297 gaging stations located in Tennessee.....	27
5. Selected basin characteristics and flood-frequency estimates for 156 gaging stations located in adjacent states.....	49
6. Single-variable regional-regression equations and accuracy statistics .....	13
7. Multivariable regional-regression equations and accuracy statistics .....	14
8. Comparison of deleted-residual standard error for the region-of-influence method and regional-regression equations.....	19
9. Suggested ranges for contributing drainage area and main-channel slope for input to the computer application.....	20

## CONVERSION FACTORS, TEMPERATURE, DATUMS, AND ABBREVIATIONS

Multiply	By	To obtain
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
square mile (mi <sup>2</sup> )	2.59	square kilometer (km <sup>2</sup> )
cubic foot (ft <sup>3</sup> )	0.02832	cubic meter (m <sup>3</sup> )
cubic foot (ft <sup>3</sup> )	28.32	liter (L)
cubic foot (ft <sup>3</sup> )	28,320	cubic centimeter (cm <sup>3</sup> )
cubic foot per second (ft <sup>3</sup> /s)	0.02832	cubic meter per second (m <sup>3</sup> /s)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows: °F = (1.8 X °C) + 32

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows: °C = (°F – 32)/1.8

Vertical coordinate information is referenced to the National Geodetic Vertical Datum of 1929 (NGVD of 1929)—a geodetic datum derived from a general adjustment of the first-order level nets of the United States and Canada, formerly called Sea Level Datum of 1929.

Horizontal coordinate information is referenced to the North American Datum of 1927.

### SELECTED ABBREVIATIONS

CDA	Contributing drainage area
CF	Climate factor
CS	Main-channel slope
HA	Hydrologic area
LAT	Latitude
LNG	Longitude
MRE	Multivariable regional-regression equations
PF	Physiographic-region factor
ROI	Region-of-influence method
SRE	Single-variable regional-regression equations
TDOT	Tennessee Department of Transportation
USGS	U.S. Geological Survey

# Flood-Frequency Prediction Methods for Unregulated Streams of Tennessee, 2000

By George S. Law and Gary D. Tasker

## ABSTRACT

Up-to-date flood-frequency prediction methods for unregulated, ungaged rivers and streams of Tennessee have been developed. Prediction methods include the regional-regression method and the newer region-of-influence method. The prediction methods were developed using stream-gage records from unregulated streams draining basins having from 1 percent to about 30 percent total impervious area. These methods, however, should not be used in heavily developed or storm-sewered basins with impervious areas greater than 10 percent. The methods can be used to estimate 2-, 5-, 10-, 25-, 50-, 100-, and 500-year recurrence-interval floods of most unregulated rural streams in Tennessee. A computer application was developed that automates the calculation of flood frequency for unregulated, ungaged rivers and streams of Tennessee.

Regional-regression equations were derived by using both single-variable and multivariable regional-regression analysis. Contributing drainage area is the explanatory variable used in the single-variable equations. Contributing drainage area, main-channel slope, and a climate factor are the explanatory variables used in the multivariable equations. Deleted-residual standard error for the single-variable equations ranged from 32 to 65 percent. Deleted-residual standard error for the multivariable equations ranged from 31 to 63 percent. These equations are included in the computer application to allow easy comparison of results produced by the different methods.

The region-of-influence method calculates multivariable regression equations for each

ungaged site and recurrence interval using basin characteristics from 60 similar sites selected from the study area. Explanatory variables that may be used in regression equations computed by the region-of-influence method include contributing drainage area, main-channel slope, a climate factor, and a physiographic-region factor. Deleted-residual standard error for the region-of-influence method tended to be only slightly smaller than those for the regional-regression method and ranged from 27 to 62 percent.

## INTRODUCTION

Planners and engineers require reliable estimates of the magnitude and frequency of floods to design bridges, culverts, embankments, dams, levees, and buildings near unregulated streams and rivers. Flood-plain management needs up-to-date information and techniques for predicting floods to protect the public and minimize flood-related costs to government and private enterprise. Standardized techniques for the measurement and analysis of hydrologic data, especially through regionalization of streamflow and basin characteristics, are essential for understanding and predicting the magnitude and frequency of floods on unregulated streams of Tennessee.

The U.S. Geological Survey (USGS), in cooperation with the Tennessee Department of Transportation (TDOT), developed and tested a computer application that automates the complex calculations necessary to predict flood magnitude and frequency. The computer application allows planners and engineers to compare flood-frequency predictions for unregulated rivers and streams in Tennessee produced with regional-regression equations and the newer region-of-influence method.

This report describes the application of flood-frequency prediction methods in Tennessee based on statistical and hydrologic techniques and data developed by various Federal, State, and local government agencies that work cooperatively with the USGS. These agencies include the Federal Highway Administration, U.S. Army Corps of Engineers, National Weather Service, Tennessee Valley Authority, Tennessee Department of Environment and Conservation, TDOT, Metropolitan Government of Nashville and Davidson County, and other Federal, State, and local agencies.

## Purpose and Scope

The purpose of this report is to describe the development of linear-regression methods that can be used to predict flood frequency for unregulated streams in Tennessee. Regression methods used include the regional-regression method and the region-of-influence method. A computer application that automates these prediction methods is described in this report.

Flood-frequency prediction methods provided in this report are applicable in the State of Tennessee. The database of information used for this study is derived from 453 streamgaging stations located primarily in rural and lightly developed areas of Tennessee and the adjacent states of Georgia, North Carolina, Virginia, Alabama, Kentucky, and Mississippi (fig. 1). These stations measure flow in streams draining basins with 1 percent to about 30 percent total impervious area.

Gaging stations in the database were required to have at least 10 years of observed annual peaks and to be free of regulation from large dams and reservoirs. A number of urban sites in Nashville, Tennessee, having from 20 to 30 percent impervious ground cover, are included in the database because they have been shown to have streamflow characteristics similar to nearby undeveloped sites (Wibben, 1976). Flood-frequency prediction methods described in this report should not be applied to heavily developed basins or storm-sewered basins having greater than 10-percent impervious cover.

## Previous Studies

Previous reports by Jenkins (1960), Patterson (1964), Speer and Gamble (1964), Randolph and

Gamble (1976), and Weaver and Gamble (1993) provided methods to define flood frequency for rural streams in Tennessee. The first three of these reports used a graphical fit on Gumbel probability paper for gaging station flood-frequency analysis and the index-flood method (Dalrymple, 1960) to regionalize the results for application at ungaged sites. The first two reports were based on data collected mostly on the main channels of rivers.

Randolph and Gamble (1976) were the first to define flood frequency at gaging stations in Tennessee by using the log-Pearson Type III statistical distribution and methodology described in U.S. Water Resources Council Bulletin 17 (1976). Randolph and Gamble delineated four hydrologic areas that are based on physiographic provinces of Tennessee. Randolph and Gamble performed statistical analyses that showed each hydrologic-area set of stations was statistically different from a single set of all gaging stations in the study area. Flood-frequency analyses were performed for 281 gaging stations having 10 or more years of record through 1972. Ordinary least-squares regression was used to develop single-variable regional-regression equations for estimating flood frequency at rural unregulated streams in each of the hydrologic areas.

Weaver and Gamble (1993) defined flood frequency at gaging stations in Tennessee using the log-Pearson Type III statistical distribution and methodology described by the Hydrology Subcommittee of the Interagency Advisory Committee on Water Data Bulletin 17B (1982). Weaver and Gamble used flood-frequency analyses for 304 gaging stations having 10 or more years of record through 1986, and continued the use of the hydrologic areas for Tennessee that were previously established by Randolph and Gamble (1976). Weaver and Gamble were the first to use the operational generalized least-squares regression computer application (Tasker and Stedinger, 1989) to develop single-variable regional-regression equations to estimate flood frequency at rural unregulated streams in each of the hydrologic areas.

Recent flood-frequency studies in other states (Tasker and Slade, 1994; Tasker and others, 1996; Asquith and Slade, 1999; Pope and others, 2001; Feaster and Tasker, 2002) have introduced a new computer-based method to produce flood-frequency estimates at unregulated streams. The region-of-influence method has demonstrated advantages by building on the regional-regression method and



improving the accuracy of flood-frequency estimates at unregulated streams. The region-of-influence method is a computer application that can be revised by periodically updating the database, which contains gaging station flood-frequency values and basin characteristics used by the program. Tennessee's flood-frequency computer program is a result of these studies and incorporates both the regional-regression method and the region-of-influence method.

## Description of Study Area

Tennessee's diverse topography ranges from the lowlands of the Mississippi Valley and Coastal Plain Physiographic Provinces and the hills of the Western Valley Physiographic Province; to the gently rolling hills and glades of the Highland Rim and Central Basin Physiographic Provinces; across the elevated Cumberland Plateau section and the highly incised Sequatchie Valley Physiographic Province; to the steep hills of the Valley and Ridge and mountains of the Blue Ridge Physiographic Provinces (Fenneman, 1946; U.S. Geological Survey, 1970, p. 59; and Miller, 1974). Land-surface elevations range from about 250 ft above NGVD of 1929 along the Mississippi River in West Tennessee to over 6,600 ft in the mountains of East Tennessee.

Geology in Tennessee is variable. West Tennessee is characterized by horizontal beds of unconsolidated sand, silt, clay, and gravel. Middle Tennessee is dominated by horizontal beds of karstic limestone. East Tennessee is characterized by folded beds of limestone and dolomite. The mountains of East Tennessee are underlain by folded beds of complex metamorphic and igneous rock.

Average precipitation in Tennessee varies from about 40 in. to nearly 80 in. per year, generally increasing from west to east (Dickson, 1960). Precipitation is lowest in the Mississippi Valley and Coastal Plain Physiographic Provinces of West Tennessee and the Central Basin Physiographic Province in Middle Tennessee where average annual precipitation totals about 45 in. Areas of the Highland Rim, Cumberland Plateau, and southern part of the Valley and Ridge Physiographic Provinces receive from 50 to 60 in. of precipitation annually. Maximums for the State occur along the foothills and peaks of the Great Smoky Mountains where average annual precipitation totals from 60 to 80 in.

Widespread flooding is uncommon in Tennessee, but typically occurs during the winter and early spring (December through March) when frequent frontal storms bring widespread rains of high intensity on already saturated ground. Localized flooding is common during the summer when thunderstorms often produce intense downpours. In the fall, while flood-producing rains are rare, the remnants of hurricanes sometimes cause serious flooding. The numerous dams constructed along the Tennessee and Cumberland Rivers and their tributaries are major features in the control of flood waters in the State (Dickson, 1960). Some of the more notable floods in Tennessee occurred in 1793, 1867, 1902, 1929, 1948, 1955, 1973, 1975, and 1984.

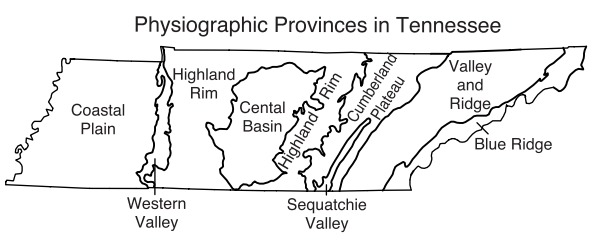
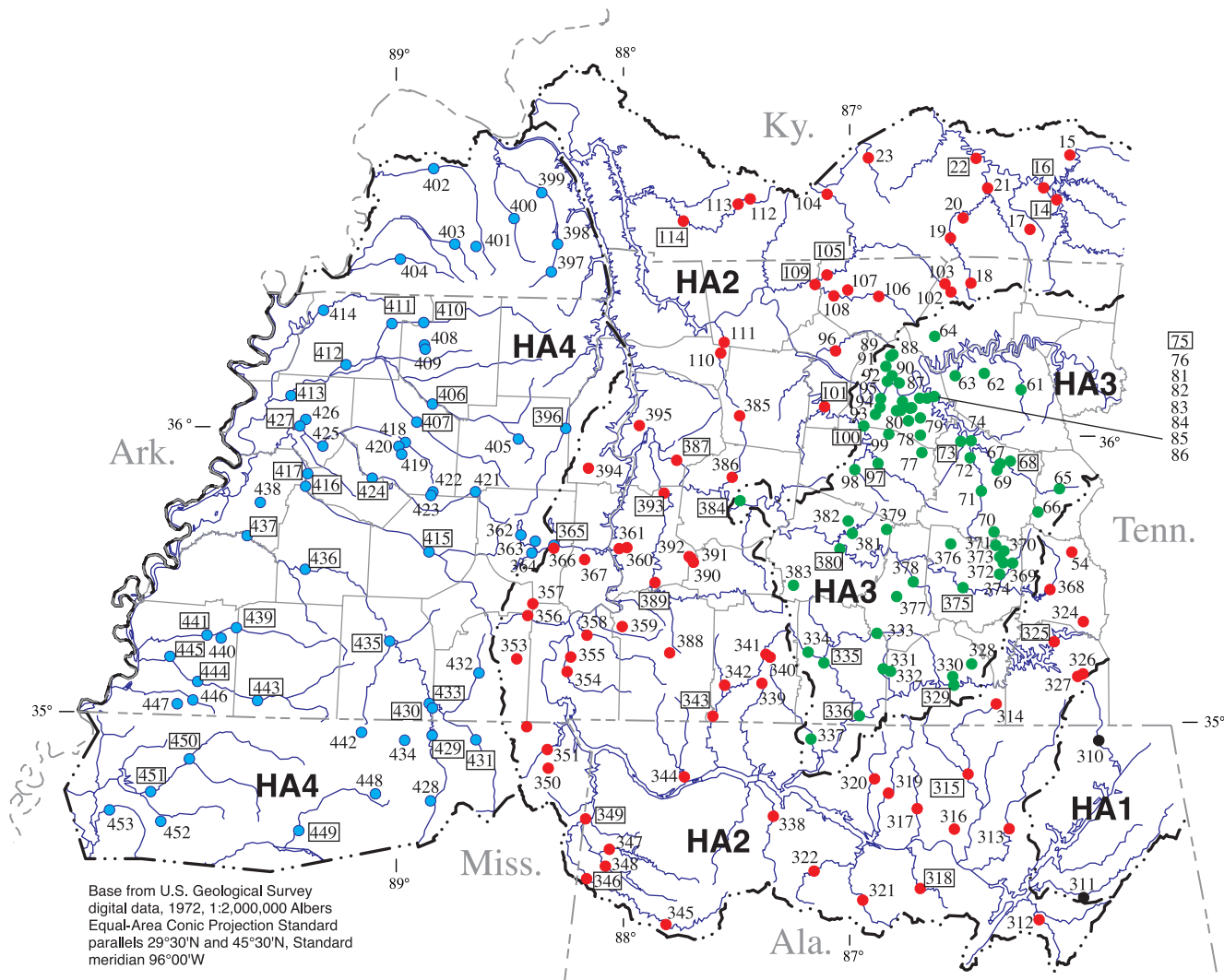
## Acknowledgments

We gratefully acknowledge the assistance and support of Mr. Jon Zirkle of the Tennessee Department of Transportation. Certain descriptions, definitions, methods, and processes described in this report were adapted from USGS Water-Resources Investigations Report 01-4207, "Estimating the Magnitude and Frequency of Floods in Rural Basins of North Carolina—Revised." We acknowledge the contribution of Tim Diehl of the USGS Tennessee District for proposing and developing the physiographic-region factor used in this study. We would like to recognize the valuable comments and suggestions made by Larry Bohman, Lamar Sanders, and Toby Feaster of the USGS. We also recognize the dedicated work of the USGS field office staff in collecting, processing, and storing most of the streamflow data used in this study.

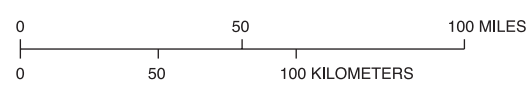
## BASIN CHARACTERISTICS

Basin characteristics are factors that describe the physical attributes of a drainage basin. Because differences in basin characteristics can be used to account for differences in flow magnitudes of Tennessee streams, these factors are often used as explanatory variables in regression equations and hydrologic models.

Selected factors that characterize size, shape, relief, geology, physiography, and climate were computed and compiled for the 453 gaging stations used in this study (fig. 1). Of the 453 stations, 297 are located in Tennessee, 21 in Georgia, 37 in North Carolina, 28 in Virginia, 20 in Alabama, 36 in Kentucky, and 14 in



Modified from Fenneman, 1946; Miller, 1974; and Weaver and Gamble, 1993



**EXPLANATION**

- HA1
- HA2
- HA3
- HA4
- Hydrologic area (HA) boundary
- 428 ● Gaging station number and location
- 318 Station used in region-of-influence example (figure 5 and table B-1)

**Figure 1.** Gaging stations, hydrologic areas, and physiographic provinces in the study area.

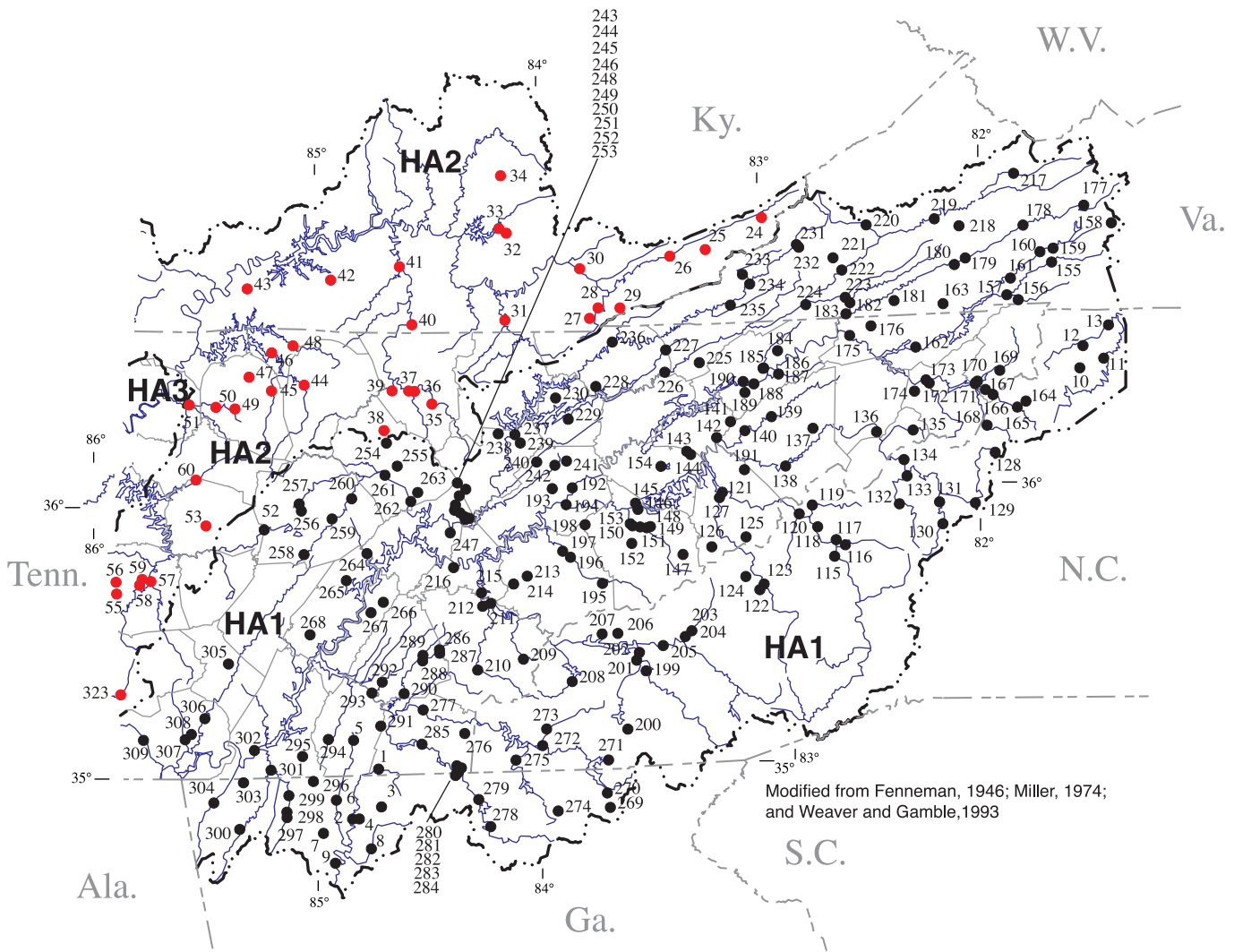


Figure 1. Gaging stations, hydrologic areas, and physiographic provinces in the study area—Continued.

Mississippi (table 1). The drainage basins measured by these stations represent a wide range of physical and climatic conditions within the study area. Basin characteristics that were analyzed for possible inclusion in the flood-frequency prediction methods include contributing drainage area (*CDA*), main-channel slope (*CS*), stream length (*L*), average basin elevation (*BE*), a basin shape factor (*SF*), selected recurrence-interval climate factors (*CF*), hydrologic region (*REG*), and a physiographic-region factor (*PF*) (table 2).

Stream length (*L*), main-channel slope (*CS*), and average basin elevation (*BE*) values were available for approximately 80 percent of the gaging stations used in this study. Values for the remaining 20 percent of the stations were computed using manual techniques. Values of *L* were determined by measuring along a stream from the gaging station proceeding upstream to the watershed divide. Values of *CS* were calculated as the change in land elevation divided by the distance between two points located 10 percent and 85 percent of the stream length upstream from the station. Values of *BE* were the average of 40 to 100 land elevations in the basin selected by using a grid-sampling method. These measurements can be calculated by using either manual or digital methods.

Pope and others (2001) indicated that the primary climatic characteristics relevant to flood frequency in a basin are the intensity, duration, and amount of rainfall, as well as other meteorologic inputs that control evaporation and transpiration. Lichty and Liscum (1978) suggested the use of a regional climate factor,  $CF_t$ , where  $t = 2-, 25-,$  and 100-year recurrence intervals, which integrates long-

term rainfall and pan evaporation information and represents the effect of these climatic influences on flood frequency. In this study, a refined version of  $CF_t$ , as developed and described by Lichty and Karlinger (1990), is used to characterize climatic effects of flood frequency. Climate factors,  $CF_t$ , for each site are computed by using a computer program that includes the maps of climate-factor isolines presented in Lichty and Karlinger (1990), and the latitude and longitude of a site to interpolate values for the three climate factors,  $CF_2$ ,  $CF_{25}$ , and  $CF_{100}$ . This climate-factor computer program is part of the flood-frequency computer application for Tennessee that is described in this report.

## Hydrologic Areas

Parts of eight physiographic regions defined by Fenneman (1946), the U.S. Geological Survey (1970, p. 59), and Miller (1974) are represented by distinct hydrologic, geologic, and topographic characteristics in Tennessee (fig. 1). Four hydrologic areas (HA1-4), previously defined by Randolph and Gamble (1976) and Weaver and Gamble (1993), were slightly modified for use in this analysis of flood frequency and follow the general physiographic province boundaries.

HA1 contains 211 stations (table 1) and includes most of the Cumberland Plateau Physiographic Province and all of the Valley and Ridge and Blue Ridge Physiographic Provinces of East Tennessee. These areas are distinct physiographically, although their flood statistics are similar, therefore these three regions are treated as a single hydrologic area. HA2 contains 115 stations and includes almost all of the

**Table 1.** Number of gaging stations by hydrologic area and state  
[See figure 1 for station and hydrologic area locations]

State	Number of stations by hydrologic area				Total stations by State
	1	2	3	4	
Georgia	21	0	0	0	21
Tennessee	123	67	64	43	297
North Carolina	37	0	0	0	37
Kentucky	0	28	0	8	36
Virginia	28	0	0	0	28
Alabama	2	17	1	0	20
Mississippi	0	3	0	11	14
Total stations by hydrologic area	211	115	65	62	453

**Table 2.** Basin characteristics

[See figure 1 for hydrologic area locations; ----, dimensionless characteristic; NGVD, National Geodetic Vertical Datum]

Basin characteristic	Unit of measure	Definition
<b>Physical characteristics</b>		
<i>LAT</i>	dd mm ss	Latitude, in degrees, minutes, and seconds, at the site of interest.
<i>LNG</i>	dd mm ss	Longitude, in degrees, minutes, and seconds, at the site of interest.
<i>CDA</i>	mi <sup>2</sup>	Contributing drainage area is the watershed area, in square miles, that contributes directly to surface runoff.
<i>CS</i>	ft/mi	Main-channel slope, in feet per mile, measured between points 10 and 85 percent of the stream length upstream from the site of interest.
<i>L</i>	mi	Stream length, in miles, measured along stream channel from the site of interest to watershed divide.
<i>BE</i>	ft	Average basin elevation, in feet above NGVD of 1929, measured from topographic maps using a grid-sampling method (40 to 100 points in each basin were sampled).
<i>SF</i>	----	Shape factor is a dimensionless watershed descriptor defined as $CDA/L^2$ .
<b>Climatic characteristics</b>		
<i>CF<sub>2</sub></i>	----	2-year recurrence-interval climate factor
<i>CF<sub>25</sub></i>	----	25-year recurrence-interval climate factor
<i>CF<sub>100</sub></i>	----	100-year recurrence-interval climate factor
<b>Regional identifiers</b>		
<i>REG</i>	----	1, if site is in hydrologic area 1; 2, if site is in hydrologic area 2; 3, if site is in hydrologic area 3; or 4, if site is in hydrologic area 4.
<b>Physiographic characteristics</b>		
<i>PF</i>	----	Physiographic-region factor is used in the region-of-influence method to capture the uniqueness of flood-magnitude potential inherent in the hydrologic areas. It is the ratio of the 2-year peak discharge from a regression equation for a hydrologic area divided by the 2-year peak discharge from a regression equation for the entire study area.

Highland Rim Physiographic Province, which is a dissected limestone plateau with karst features. In addition, HA2 includes parts of the Cumberland Plateau and Western Valley Physiographic Provinces. HA3 contains 65 stations and closely conforms to the Central Basin Physiographic Province, which is a less karstic area underlain by limestone that has less relief than the Highland Rim. HA4 contains 62 stations and includes all of the Coastal Plain Physiographic Province and the western part of the Western Valley Physiographic Province (Weaver and Gamble, 1993).

Hydrologic areas presented by Weaver and Gamble (1993) were slightly modified in two places for use in this study. First, approximately 200 mi<sup>2</sup> of

land drained by the Elk River and its tributaries, in Coffee, Franklin, and Grundy Counties (see map on inside of front cover) formerly in HA1, was reassigned to HA2. This change allows the hydrologic area boundary to trace the regional drainage basin divide and conform more closely to the physiography and geology of the area. Second, about 75 mi<sup>2</sup> of the Duck River Basin, in Hickman County formerly in HA2, was reassigned to HA3. This change extends HA3 farther down the Duck River, which exhibits flood characteristics associated with this hydrologic area. HA4 was not modified for this study.

The hydrologic area for each site of interest can be determined by examining the study area map

(fig. 1) or, if necessary, by using more detailed maps. The integer value for the dominant hydrologic area that each gaging station measures was assigned to the region variable (*REG*). The dominant hydrologic area was assigned to the *REG* for each gaging station, even if the drainage basin for that station lies in two hydrologic areas, thus allowing for the database to be easily sorted by hydrologic area for regional flood-frequency analyses.

### Physiographic-Region Factor

Physiographic information can be used in flood-frequency analysis in several ways. Previous studies in Tennessee (Randolph and Gamble, 1976; Weaver and Gamble, 1993), as well as the current study (2002), analyzed flood frequency separately in the four hydrologic areas. An alternative approach, used in this study, is to compute a dimensionless basin characteristic that quantifies the effect of the physiographic provinces on flood statistics at gaged and ungaged sites. This factor, known as the physiographic-region factor (*PF*), is treated as an explanatory variable in further statistical analyses, such as those performed by the region-of-influence method, which combines data from all the physiographic regions. *PF* allows the region-of-influence method to capture some of the uniqueness in flood-magnitude potential inherent in the physiographic province-based hydrologic areas (fig. 1) (T.H. Diehl, U.S. Geological Survey, written commun., 2001). The physiographic-region factor is computed as

$$PF = Q_{2,REG} / Q_{2,ALL}, \quad (1)$$

where

$Q_{2,REG}$  is the 2-year recurrence-interval peak discharge computed by using a single-variable

ordinary least-squares regression equation developed for each of the hydrologic-area groupings of stations, and

$Q_{2,ALL}$  is the 2-year recurrence-interval peak discharge computed using a single-variable ordinary least-squares regression equation developed using all 453 gaging stations in the study area.

The 2-year recurrence-interval peak discharge ( $Q_2$ ) was used as an indicator of the response of floods within a physiographic region because the  $Q_2$  is a common event and an indicator of the amount of water that will run off during flood conditions. *PF* is computed at sites of interest in Tennessee using the hydrologic area (*HA*) and the contributing drainage area (*CDA*) of the site of interest. Contributing drainage area is the most important basin characteristic in flood-frequency prediction in Tennessee. *PF* equations (table 3) are incorporated into the flood-frequency computer application for Tennessee.

## FLOOD-FREQUENCY PREDICTION METHODS

Flood discharges for 453 gaging stations located in Tennessee and six adjacent States (fig. 1) with 10 or more years of record through water year 1999 were used to develop the regression methods presented in this report. Water year refers to the period of record beginning October 1<sup>st</sup> and ending September 30<sup>th</sup> of the designated year. For example, the 1999 water year is from October 1, 1998, through September 30, 1999. Flood discharges for these gaging stations were computed by fitting the peak streamflow data and supplemental historic information for each station to the log-Pearson Type III distribution as described in Bulletin 17B of the Interagency Advisory Committee on Water Data (1982).

**Table 3.** Physiographic-region factor equations

[OLS, ordinary least-squares regression;  $Q_2$ , 2-year recurrence-interval peak discharge in cubic feet per second; *CDA*, contributing drainage area in square miles]

Hydrologic area	Physiographic-region factor equations	OLS $Q_2$ equations	
		For each hydrologic area	For the entire study area
1	$0.6124CDA^{0.0626}$	$125.6CDA^{0.7482}$	$205.1CDA^{0.6855}$
2	$1.0394CDA^{0.0353}$	$213.2CDA^{0.7208}$	$205.1CDA^{0.6855}$
3	$1.7057CDA^{-0.0242}$	$349.9CDA^{0.6613}$	$205.1CDA^{0.6855}$
4	$2.0156CDA^{-0.1540}$	$413.4CDA^{0.5313}$	$205.1CDA^{0.6855}$

Gaging stations are grouped by hydrologic area and related to contributing drainage area (*CDA*), main-channel slope (*CS*), and a climate factor (*CF*) to produce the regional-regression equations. The regional-regression equations, in particular the single-variable regression equations, which are easy to solve manually, are an alternative that can be used to obtain estimates of flood frequency at unregulated sites in Tennessee if the computer application, and therefore the region-of-influence method, is not available.

The region-of-influence method by Tasker and others (1996), required the development of a computer application to derive prediction equations that relate recurrence-interval flood discharges for gaging stations, computed using Bulletin 17B of the Interagency Advisory Committee on Water Data (1982), to *CDA*, *CS*, *CF*, and a physiographic-region factor (*PF*). The physiographic-region factor allows the region-of-influence method to capture the uniqueness in flood-magnitude potential inherent in the four hydrologic areas in Tennessee, which are based on physiographic provinces. Similar to the regional-regression method, the region-of-influence method uses generalized least-squares regression to compute flood-frequency prediction equations. However, the region-of-influence regression analysis is applied to 60 of the most similar stations chosen from the database of 453 gaging stations, rather than the four hydrologic-area groupings of stations.

### Unregulated, Gaged Sites

Different methods are used to compute flood frequency at gaged sites than at ungaged sites. The methodology described in the following paragraphs of this section describes the prediction of flood frequency at gaged sites on unregulated streams in Tennessee.

### Recurrence Intervals

Flood-frequency estimates for given stream sites are typically presented as sets of exceedance probabilities or, alternatively, recurrence intervals along with the associated discharges. Exceedance probability is defined as the probability of exceeding a specified discharge in a 1-year period and is expressed as decimal fractions less than 1.0 or as percentages less than 100. A discharge with an exceedance probability of 0.10 has a 10-percent chance of being exceeded in any given year. Recurrence interval is defined as the number of years, on average, during

which the specified discharge is expected to be exceeded one time and is expressed as number of years. A discharge with a 10-year recurrence interval is one that, on average, will be exceeded once every 10 years.

Recurrence interval and exceedance probability are the mathematical inverses of each other; thus, a discharge with an exceedance probability of 0.10 has a recurrence interval of 1/0.10 or 10 years. Note: Recurrence intervals, regardless of length, always refer to an estimated average number of occurrences over a long period of time; for example, a 10-year flood discharge is one that might occur about 10 times in a 100-year period, rather than exactly once every 10 years. A 10-year flood discharge might occur 3 years consecutively. Thus, exceedance probability and recurrence interval do not indicate when a particular flood discharge will occur.

### Bulletin 17B Method

Flood-frequency estimates for gaged sites are computed by fitting the series of annual peak flows to a known statistical distribution. For the purposes of this study, estimates of flood-flow frequency are computed by fitting the logarithms (base 10) of the annual peak flows to a log-Pearson Type III distribution, following the guidelines and using the computational methods described in Bulletin 17B of the Hydrology Subcommittee (Interagency Advisory Committee on Water Data, 1982). The equation for fitting the log-Pearson Type III distribution to an observed series of annual peak flows is as follows:

$$\log_{10} Q_t = \bar{X} + KS \quad (2)$$

where

$Q_t$  is the  $t$ -year recurrence-interval peak discharge, in cubic feet per second,

$\bar{X}$  is the mean of the log (base 10)-transformed annual peak flows,

$K$  is a factor dependent on recurrence interval and the skew coefficient of the log (base 10)-transformed annual peak flows, and

$S$  is the standard deviation of the log (base 10)-transformed annual peak flows.

Values for  $K$  for a wide range of recurrence intervals and skew coefficients are published in appendix 3 of Bulletin 17B (Interagency Advisory Committee on Water Data, 1982).

Fitting the log-Pearson Type III distribution to a long-term, well-distributed series of annual peak flows generally is straightforward; however, a series of peak flows may include low or high peak flows that depart noticeably from the trend in the data. The station record also may include information about maximum peak flows that occurred outside of the period of regularly collected, or systematic, record. Such peak flows, known as historic peaks, are often the maximum peak flows known to have occurred during an extended period of time, longer than the period of data collection. Interpretation of outliers and historic peak information in the fitting process can affect the final flood-frequency estimate.

Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) provides guidelines for detecting and interpreting outliers and historic peaks and provides computational methods for making appropriate adjustments to the distribution to account for their presence. In some cases, high or low outliers are excluded from the record, so that the number of systematic peaks may not be equal to the number of years in the period of record.

Statistical measures of data, such as mean, standard deviation, or skew coefficient, can be described in terms of the sample or computed measure and the population or true measure. In terms of annual peak flows, the period of collected record can be thought of as a sample, or small part, of the entire record, or population. Statistical measures computed from the sample record are estimates of what the measure would be if the entire population were known and used to compute the given measure. The accuracy of these estimates depends on the nature of the specific measure and the given sample of the population.

Skew coefficient measures the symmetry of the distribution of a set of peak flows about the median of the distribution. A peak-flow distribution with the mean equal to the median is said to have zero skew. A positively skewed distribution has a mean that exceeds the median, typically as a result of one or more extremely high peak flows. A negatively skewed distribution has a mean that is less than the median, typically because of one or more extremely low peak flows.

The computed skew coefficient for the peak-flow record of a given station is very sensitive to extreme events; therefore, the sample skew coefficient for short records may not provide an accurate estimate of the population skew. This is problematic because

the  $K$ -factor in equation 2 for a given recurrence interval is dependent only on skew coefficient; therefore, an inaccurate skew coefficient will result in a flood-frequency estimate that is not representative of the true, or population, value.

An improved estimate of skew coefficient at a site can be obtained by using a weighted average of the sample skew coefficient estimate with a generalized, or regional, skew coefficient. A generalized skew coefficient is obtained by combining skew estimates from nearby, similar sites. A nationwide generalized skew study was conducted as documented in Bulletin 17B (Interagency Advisory Committee on Water Data, 1982). Skew coefficients for long-term gaged sites from across the Nation were computed and used to produce a map of isolines of generalized skew. The nationwide map of generalized skews was used in the computation of the weighted skew coefficient used to determine the  $K$ -factor in equation 2.

Peak discharges for recurrence intervals of 2, 5, 10, 25, 50, 100, and 500 years were determined for each of the 453 gaging stations by using data collected through the 1999 water year and the methodology described above (table 4 at back of report). For those streams where regulation now exists, the discharge values calculated are based on streamflow data collected prior to regulation. Flood-frequency estimates for 156 gaging stations located in adjacent states (table 5 at back of report) were used strictly to supplement the database used by the flood-frequency computer application for Tennessee; these estimates for sites in other states should not be used for design purposes.

Flood-frequency estimates for the 156 stream gages located outside of Tennessee should be obtained from the most recently published flood-frequency report for that state (Landers and Wilson, 1991; Stamey and Hess, 1993; Bisese, 1995; Atkins, 1996; Pope and others, 2001; and Hodgkins and Martin, in press). Any significant difference in flood-frequency estimates provided in these reports and the supplemental data used in this study likely is caused by differences in historical-record adjustment methodology, inclusion of additional systematic data, and the use of the nationwide skew map in this study.

## Unregulated, Ungaged Sites

Regional regression can be used to estimate flood frequency for all unregulated streams and rivers



and allows planners, hydrologists, and engineers to enhance the value of discharge records measured at gaging stations. Because streamflow is recorded at only a few of the many sites where information is needed, gaging-station information must be transferred to ungaged sites. Regional regression provides a tool for doing this. In addition, a regional regression may produce improved estimates of streamflow characteristics at the gaged sites (Riggs, 1973).

Two regression methods were developed that estimate flood discharges for unregulated sites in Tennessee. The first method, regional regression, uses generalized least-squares regression to define a set of predictive equations that relate peak discharges for various recurrence intervals to selected basin characteristics for unregulated streams and rivers in each of four hydrologic areas of Tennessee (fig. 1). The second method, the region-of-influence, required the development of a computer application to derive unique predictive relations that relate peak discharges to selected basin characteristics at unregulated sites in Tennessee. Just as in the regional-regression method, generalized least-squares regression is used to develop these predictive relations; however, in the region-of-influence method, regression analysis is applied to a subset of gaged sites chosen from the entire database of gaged sites, rather than the regional groupings of gaged sites.

### Regional-Regression Method

The four hydrologic area groups of streamgaging stations were analyzed to ensure that these regional groups (fig. 1) contribute to improved flood-frequency predictions in Tennessee. Regional-regression equations used to estimate flood frequency in Tennessee were developed by applying statistical techniques of ordinary and generalized least-squares regression to the hydrologic area groups of stations (table 1). Single-variable and multivariable regression equations that relate flood frequency to the best combination of explanatory basin characteristics (table 2) are presented in this section of the report.

The validity of the hydrologic areas were examined by performing a Wilcoxon signed-ranks test (Tasker, 1982) using a single-variable ordinary least-squares regression equation for the 50-year recurrence-interval peak discharge ( $Q_{50}$ ) developed from all 453 stations used in the study. Additionally, a test was conducted by introducing the regional identifiers (*REG*) (table 2) into the single-variable regres-

sion equations developed using all 453 stations in the study area. For each station, *REG* was set either at 1, if the site was in a particular region, or 0, if not. A multivariable ordinary least-squares regression equation developed using all 453 stations and (1) *CDA*, (2) *REG*, and (3) *REG* multiplied by *CDA*, was constructed for  $Q_{50}$  in each of the four hydrologic areas. For each equation, a significant coefficient for *REG* indicates a difference in the intercept between stations in that hydrologic area and stations in the rest of the study area; a significant coefficient for the product of *REG* and *CDA* indicates a difference in the coefficients of *CDA* between stations in that hydrologic area and stations in the rest of the study area. In this study, a 95-percent confidence interval was specified for significance testing. Each hydrologic-area group of gaging stations was shown to be significant by either the Wilcoxon signed-ranks test or the multivariable ordinary least-squares regression equations developed using the regional identifiers. Therefore, the hydrologic areas proposed for use in this study were accepted.

Ordinary least-squares regression is an appropriate and efficient method for use when flow estimates that are used as response variables are independent of one another (no correlation exists between pairs of sites) and when the reliability and variability of flow estimates that are used as response variables are approximately equal. Flood-frequency estimates for streams (Interagency Advisory Committee on Water Data, 1982) used in this study were calculated from peak-flow records measured at gaging stations throughout Tennessee and in parts of adjacent states. Systematic periods of record for the gaging stations used in this study range from 10 years to about 100 years. Records from gaging stations on the same stream within the same basin or even in adjacent basins may be highly correlated because the peak flows result from the same rainfall events, similar antecedent conditions, and similar basin characteristics. However, records from other gaging stations, in more remote basins, have varying degrees of correlation. In general, correlation between pairs of gaging stations can be described as a function of the distance between stations. Additionally, the reliability of the flood-frequency estimates computed using methods from Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) generally is a function of record length and, as such, cannot be considered equal for all gaging stations. Variability of the flow estimates,

characterized by the standard deviation of the peak-flow record that was used to compute the flow estimate, depends in large part on characteristics of the basin and cannot be considered equal for all gaging stations used in the study. For these reasons, ordinary least-squares regression was used only as an exploratory technique in this study to identify the basin characteristics most likely to be significant in the regression equations and to validate the hydrologic areas. The coefficients for the final regression equations were calculated by using generalized least-squares regression.

Generalized least-squares regression, as described by Stedinger and Tasker (1985), is a regression technique that takes into account the correlation between, as well as differences in, the variability and reliability of the flow estimates used as dependent, or response, variables. These factors are accounted for in generalized least-squares regression by assigning different weights to each observation of the response variable used in the regression, based on its contribution to the total variance of the sample-flow statistic used as the response variable. In contrast, ordinary least-squares regression assumes equal reliability and variability in flow estimates at all gaging stations that are assigned equal weight in the regression.

The use of generalized least-squares regression techniques to model the relations between peak discharges and basin characteristics of unregulated streams in Tennessee requires estimates of the cross-correlation coefficients and standard deviation of the peak-flow records that were used to compute peak discharges for the selected recurrence intervals. For each of the four hydrologic areas, a scatter plot of sample correlation coefficients versus distance between stations was constructed for gaging station pairs with at least 30 years of concurrent record. A graphical “best-fit” line to these points was used to define the relation between cross-correlation coefficient and distance between stations. This relation was then used to populate a cross-correlation matrix for the stations within each area. Variability of each peak-flow estimate is measured by the standard deviation of the peak-flow record used to compute that estimate. For each hydrologic area, a generalized least-squares regression of the sample standard deviations against *CDA* was used to obtain estimates of the standard deviations of the peak-flow records at each station. These regression estimates of the standard deviations were used to assign weights to flow estimates because they are

independent of the sample standard deviation estimates used to compute the flow estimate. Finally, length of record at each gaging station, which at many stations is adjusted for historical information, was used as a direct measure of the relative reliability of the flow estimates computed from those records.

Generalized least-squares regression was used to improve the single-variable and multivariable regional-regression equations determined by exploratory analysis using ordinary least-squares regression. Single-variable regional-regression equations are provided for all four hydrologic areas of Tennessee (table 6). In HA1, the inclusion of multiple variables in the regression equations marginally improves their predictive ability when compared to the single-variable regression equations. In HA2, 3, and 4, the inclusion of multiple variables in the regression equations provides little or no improvement in predictive ability when compared to the single-variable regression equations. However, for comparison purposes in the computer application, multivariable regression equations are provided for HA1, 2, and 3, but not for HA4 (table 7).

Regional-regression equations for HA1, 2, and 4 are single-segment linear equations. However, regression equations for HA3 are two-segment linear equations (fig. 2). Segmented equations were necessary in HA3 to account for curvature in the explanatory data. Determining the causes of the curvature in the explanatory data for HA3 (fig. 3) requires further study.

The final single-variable regression equations for each of the hydrologic areas relate peak discharge to *CDA* (table 6). The multivariable regression equations for HA1, 2, and 3 include *CDA* and *CS*, and in HA1, *CF*<sub>2</sub> (table 7). In each of the regression methods described in this report, *CF*<sub>2</sub> is renamed *CF* for simplicity.

Uncertainty in a flow estimate that was predicted for a site of interest, indexed by *i*, by using the regional-regression equations can be measured by the standard error of prediction,  $S_{p,i}$ , which is computed as the square root of the prediction error variance ( $MSE_p$ ). The  $MSE_p$ , as described by Stedinger and Tasker (1985), is the sum of two components—the model error variance described by Moss and Karlinger (1974) that results from the regression equation,  $\gamma^2$ , and the sampling error variance ( $MSE_{s,i}$ ) which results from estimating equation coefficients from samples of the population. The model error variance,  $\gamma^2$ , is a characteristic of the regression equation and is assumed

**Table 6.** Single-variable regional-regression equations and accuracy statistics

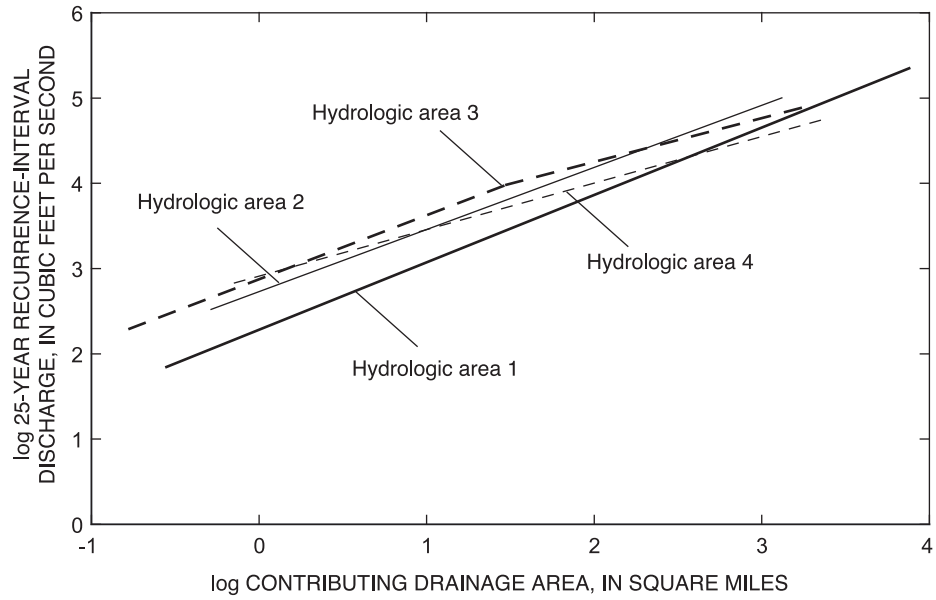
[ft<sup>3</sup>/s, cubic feet per second; *CDA*, contributing drainage area in square miles; see figure 1 for hydrologic area locations; mi<sup>2</sup>, square miles]

Recurrence interval, in years	Peak-discharge equation, in ft <sup>3</sup> /s	Average prediction error, in percent	Prediction-error departure	
			Under-estimation, in percent	Over-estimation, in percent
<b>Hydrologic area 1 (CDA=0.20 to 9,000 mi<sup>2</sup>)</b>				
2	119 <i>CDA</i> <sup>0.755</sup>	42.9	-33.7	+50.9
5	197 <i>CDA</i> <sup>0.740</sup>	42.2	-33.3	+49.9
10	258 <i>CDA</i> <sup>0.731</sup>	43.0	-33.8	+51.0
25	342 <i>CDA</i> <sup>0.722</sup>	44.9	-34.9	+53.6
50	411 <i>CDA</i> <sup>0.716</sup>	47.0	-36.1	+56.4
100	484 <i>CDA</i> <sup>0.710</sup>	49.5	-37.4	+59.7
500	672 <i>CDA</i> <sup>0.699</sup>	56.1	-40.7	+68.7
<b>Hydrologic area 2 (CDA=0.47 to 2,557 mi<sup>2</sup>)</b>				
2	204 <i>CDA</i> <sup>0.727</sup>	32.0	-26.8	+36.7
5	340 <i>CDA</i> <sup>0.716</sup>	30.2	-25.6	+34.4
10	439 <i>CDA</i> <sup>0.712</sup>	31.2	-26.3	+35.6
25	573 <i>CDA</i> <sup>0.709</sup>	33.4	-27.7	+38.4
50	677 <i>CDA</i> <sup>0.707</sup>	35.6	-29.2	+41.3
100	785 <i>CDA</i> <sup>0.705</sup>	37.9	-30.7	+44.2
500	1,050 <i>CDA</i> <sup>0.702</sup>	43.9	-34.3	+52.2
<b>Hydrologic area 3 (CDA=0.17 to 30.2 mi<sup>2</sup>)</b>				
2	280 <i>CDA</i> <sup>0.789</sup>	34.3	-28.4	+39.6
5	452 <i>CDA</i> <sup>0.769</sup>	34.1	-28.3	+39.4
10	574 <i>CDA</i> <sup>0.761</sup>	34.6	-28.5	+39.9
25	733 <i>CDA</i> <sup>0.753</sup>	35.5	-29.2	+41.1
50	853 <i>CDA</i> <sup>0.748</sup>	36.5	-29.8	+42.5
100	972 <i>CDA</i> <sup>0.745</sup>	37.7	-30.5	+43.9
500	1,250 <i>CDA</i> <sup>0.739</sup>	40.8	-32.5	+48.1
<b>Hydrologic area 3 (CDA=30.21 to 2,048 mi<sup>2</sup>)</b>				
2	679 <i>CDA</i> <sup>0.527</sup>	27.4	-23.6	+30.9
5	1,040 <i>CDA</i> <sup>0.523</sup>	28.0	-24.0	+31.6
10	1,280 <i>CDA</i> <sup>0.523</sup>	29.6	-25.1	+33.6
25	1,590 <i>CDA</i> <sup>0.525</sup>	32.5	-27.1	+37.2
50	1,800 <i>CDA</i> <sup>0.527</sup>	34.9	-28.8	+40.4
100	2,020 <i>CDA</i> <sup>0.529</sup>	37.7	-30.5	+43.9
500	2,490 <i>CDA</i> <sup>0.537</sup>	44.4	-34.6	+52.9
<b>Hydrologic area 4 (CDA=0.76 to 2,308 mi<sup>2</sup>)</b>				
2	436 <i>CDA</i> <sup>0.527</sup>	38.7	-31.2	+45.3
5	618 <i>CDA</i> <sup>0.545</sup>	37.2	-30.3	+43.4
10	735 <i>CDA</i> <sup>0.554</sup>	38.0	-30.7	+44.3
25	878 <i>CDA</i> <sup>0.564</sup>	40.1	-32.0	+47.1
50	981 <i>CDA</i> <sup>0.570</sup>	42.2	-33.3	+49.9
100	1,080 <i>CDA</i> <sup>0.575</sup>	44.7	-34.7	+53.2
500	1,310 <i>CDA</i> <sup>0.586</sup>	51.1	-38.2	+61.8

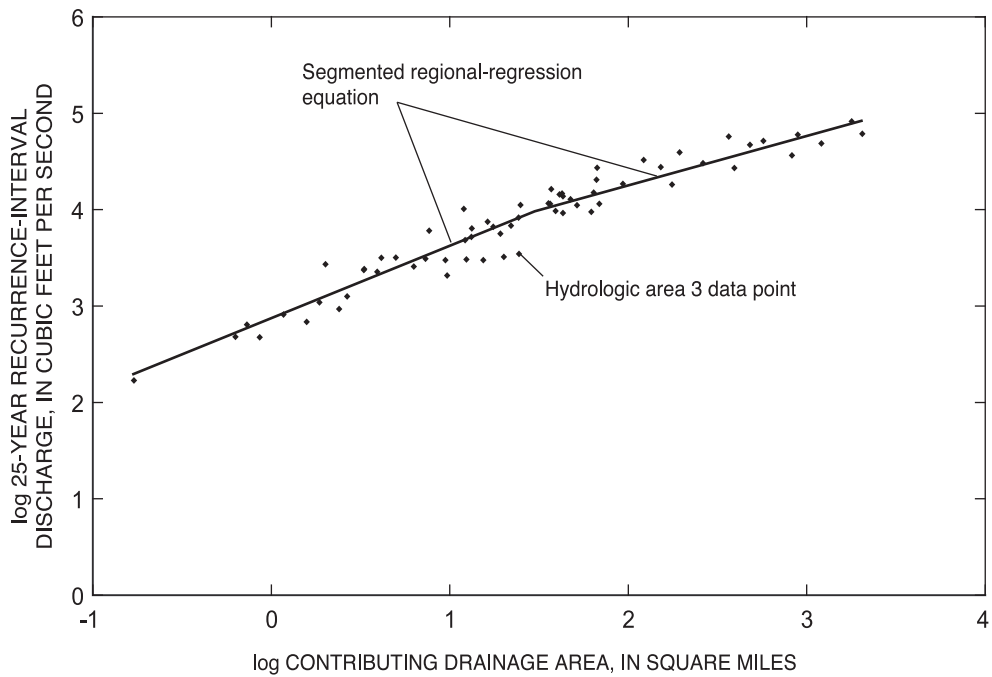
**Table 7.** Multivariable regional-regression equations and accuracy statistics

[ft<sup>3</sup>/s, cubic feet per second; *CDA*, contributing drainage area in square miles; *CS*, main-channel slope in feet per mile; *CF*, 2-year recurrence-interval climate factor; see figure 1 for hydrologic area locations; mi<sup>2</sup>, square miles]

Recurrence interval, in years	Peak-discharge equation, in ft <sup>3</sup> /s	Average prediction error, in percent	Prediction-error departure	
			Under-estimation, in percent	Over-estimation, in percent
<b>Hydrologic area 1 (CDA=0.20 to 9,000 mi<sup>2</sup>)</b>				
2	1.72 <i>CDA</i> <sup>0.798</sup> <i>CS</i> <sup>0.112</sup> <i>CF</i> <sup>4.581</sup>	39.2	-31.5	+45.9
5	3.41 <i>CDA</i> <sup>0.783</sup> <i>CS</i> <sup>0.114</sup> <i>CF</i> <sup>4.330</sup>	38.2	-31.3	+45.6
10	5.34 <i>CDA</i> <sup>0.775</sup> <i>CS</i> <sup>0.116</sup> <i>CF</i> <sup>4.087</sup>	40.1	-32.0	+47.1
25	9.00 <i>CDA</i> <sup>0.766</sup> <i>CS</i> <sup>0.117</sup> <i>CF</i> <sup>3.778</sup>	42.7	-33.6	+50.6
50	12.8 <i>CDA</i> <sup>0.760</sup> <i>CS</i> <sup>0.117</sup> <i>CF</i> <sup>3.560</sup>	45.2	-35.0	+53.8
100	17.9 <i>CDA</i> <sup>0.754</sup> <i>CS</i> <sup>0.117</sup> <i>CF</i> <sup>3.354</sup>	47.9	-36.5	+57.6
500	36.1 <i>CDA</i> <sup>0.742</sup> <i>CS</i> <sup>0.114</sup> <i>CF</i> <sup>2.904</sup>	55.2	-40.3	+67.5
<b>Hydrologic area 2 (CDA=0.47 to 2,557 mi<sup>2</sup>)</b>				
2	106 <i>CDA</i> <sup>0.787</sup> <i>CS</i> <sup>0.151</sup>	30.5	-25.8	+34.8
5	170 <i>CDA</i> <sup>0.779</sup> <i>CS</i> <sup>0.158</sup>	28.5	-24.4	+32.2
10	218 <i>CDA</i> <sup>0.776</sup> <i>CS</i> <sup>0.160</sup>	29.4	-25.0	+33.3
25	285 <i>CDA</i> <sup>0.772</sup> <i>CS</i> <sup>0.160</sup>	31.8	-26.7	+36.4
50	340 <i>CDA</i> <sup>0.769</sup> <i>CS</i> <sup>0.159</sup>	34.1	-28.3	+39.4
100	397 <i>CDA</i> <sup>0.766</sup> <i>CS</i> <sup>0.157</sup>	36.7	-29.9	+42.7
500	547 <i>CDA</i> <sup>0.761</sup> <i>CS</i> <sup>0.151</sup>	43.1	-33.8	+51.1
<b>Hydrologic area 3 (CDA=0.17 to 30.2 mi<sup>2</sup>)</b>				
2	211 <i>CDA</i> <sup>0.815</sup> <i>CS</i> <sup>0.063</sup>	35.2	-28.9	+40.7
5	329 <i>CDA</i> <sup>0.798</sup> <i>CS</i> <sup>0.071</sup>	34.9	-28.8	+40.4
10	405 <i>CDA</i> <sup>0.793</sup> <i>CS</i> <sup>0.078</sup>	35.4	-29.1	+41.0
25	497 <i>CDA</i> <sup>0.789</sup> <i>CS</i> <sup>0.086</sup>	36.4	-29.7	+42.3
50	565 <i>CDA</i> <sup>0.786</sup> <i>CS</i> <sup>0.092</sup>	37.4	-30.4	+43.6
100	632 <i>CDA</i> <sup>0.785</sup> <i>CS</i> <sup>0.096</sup>	38.6	-31.1	+45.2
500	789 <i>CDA</i> <sup>0.781</sup> <i>CS</i> <sup>0.102</sup>	40.5	-32.5	+47.7
<b>Hydrologic area 3 (CDA=30.21 to 2,048 mi<sup>2</sup>)</b>				
2	409 <i>CDA</i> <sup>0.584</sup> <i>CS</i> <sup>0.102</sup>	27.9	-23.9	+31.4
5	767 <i>CDA</i> <sup>0.558</sup> <i>CS</i> <sup>0.061</sup>	28.6	-24.4	+32.3
10	980 <i>CDA</i> <sup>0.554</sup> <i>CS</i> <sup>0.054</sup>	30.3	-25.7	+34.5
25	1,200 <i>CDA</i> <sup>0.557</sup> <i>CS</i> <sup>0.056</sup>	33.4	-27.7	+38.4
50	1,330 <i>CDA</i> <sup>0.562</sup> <i>CS</i> <sup>0.061</sup>	35.9	-29.4	+41.7
100	1,430 <i>CDA</i> <sup>0.568</sup> <i>CS</i> <sup>0.068</sup>	38.6	-31.1	+45.2
500	1,600 <i>CDA</i> <sup>0.587</sup> <i>CS</i> <sup>0.090</sup>	45.7	-35.3	+54.6
<b>Hydrologic area 4 (CDA=0.76 to 2,308 mi<sup>2</sup>)</b>				
No multivariable regression equations developed for this region (see table 6).				



**Figure 2.** Regional-regression equations for the 25-year flood for Tennessee.



**Figure 3.** A segmented regional-regression equation used for hydrologic area 3.

constant for all sites.  $MSE_{s,i}$  for a given site, however, depends on the values of the explanatory variables used to develop the flow estimate at that site. The standard error of prediction for a site,  $i$ , is computed as:

$$S_{p,i} = (\gamma^2 + MSE_{s,i})^{1/2}, \quad (3)$$

and, therefore, varies from site to site. If the values of the explanatory variables for the gaging stations used in the regression are assumed to be representative of all sites in the region, then a general measure of the prediction accuracy of the regression equation can be determined by computing the average prediction error:

$$S_p^- = \left[ \gamma^2 + (1/n) \sum_{i=1}^n MSE_{s,i} \right]^{1/2}. \quad (4)$$

The average prediction error for a regression equation can be transformed from log (base 10) units to percent error by equation 5. Negative and positive prediction-error departures, in percent of the predicted value in cubic feet per second, may be calculated by equations 6 and 7 as follows:

$$\%SE_p = 100\{[e^{5.302 (S_p^-)^2} - 1]^{1/2}\}, \text{ and} \quad (5)$$

$$\%SE_{p(-\text{departure})} = 100[10^{-(S_p^-)} - 1], \text{ and} \quad (6)$$

$$\%SE_{p(+\text{departure})} = 100[10^{(S_p^-)} - 1]. \quad (7)$$

Average prediction errors provide a measure of potential underestimation or overestimation of a regression method. Computation of  $S_{p,i}$  for a given ungaged site,  $i$ , involves complex matrix algebra (appendix A). The average prediction error and the negative and positive prediction-error departures computed by using  $S_p^-$  provide an overall measure of the predictive ability of a regression equation. Average prediction errors for the regional-regression equations range from about 27 to 56 percent (tables 6 and 7). The negative and positive prediction-error departures for the single-variable and multivariable regional-regression equations range from about -24 to -41 percent and +31 to +69 percent, respectively (tables 6 and 7).

Another useful measure of the quality of a discharge estimate is the prediction interval for the esti-

mate. A prediction interval consists of an upper limit and a lower limit for a discharge estimate for a given level of confidence. A reduced prediction interval for a given level of confidence indicates a better discharge estimate. In this study, a 90-percent level of confidence is used to compute prediction intervals, which means there is a 95-percent chance that the true discharge value lies between the upper and lower limits. Computational procedures and the matrices needed to compute prediction intervals are provided in appendix A.

## Region-of-Influence Method

Another technique for estimating flood frequency at unregulated sites is the region-of-influence method (Tasker and Slade, 1994; Hodge and Tasker, 1995; Tasker and others, 1996; Asquith and Slade, 1999; Pope and others, 2001). In this method, multi-variable regression equations for each recurrence-interval peak flow are developed by using explanatory data from a unique group of similar gaging stations selected from all the stations in the study area. This unique group of stations that are most similar to the site of interest is called the "region-of-influence" by Burn (1990a, b) and suggested by Acreman and Wiltshire (1987). In this method, the similarity of a gaging station to the site of interest is measured not by the physical distance between the sites, but by the similarity in terms of the basin characteristics. The mathematical formula for the similarity between sites  $i$  and  $j$  is defined by the Euclidean distance metric:

$$d_{ij} = \left\{ \sum_{k=1}^p [(x_{ik} - x_{jk})/sd(X_k)]^2 \right\}^{1/2}, \quad (8)$$

where

$d_{ij}$  is the distance between sites  $i$  and  $j$  in terms of basin characteristics,

$p$  is the number of basin characteristics used to calculate  $d_{ij}$ ,

$X_k$  is the  $k$ th basin characteristic,

$sd(X_k)$  is the sample standard deviation for  $X_k$ , and

$x_{ik}$  is the value of  $X_k$  at the  $i$ th site.

This distance metric is directly analogous to the more familiar equation for distance ( $D$ ) between two points,  $(x_1, y_1)$  and  $(x_2, y_2)$  in a two-dimensional rectangular coordinate system:

$$D = [(x_2 - x_1)^2 + (y_2 - y_1)^2]^{1/2}, \quad (9)$$

where the only difference is the use of sample standard deviation to standardize the different basin characteristics and the slight notational difference of using an additional subscript  $k$  rather than changing variable symbols ( $x, y$ ).

Using  $CDA$ ,  $CS$ , and  $CF$ , the distances or similarities ( $d_{ij}$ 's) between a given site of interest and all the gaged sites are computed and ranked; the number of gaging stations ( $N$ ) with the smallest  $d_{ij}$  compose the region-of-influence for the site of interest. Once the region-of-influence is determined, generalized least-squares regression techniques are used to develop the unique predictive relations between flood discharge and the basin characteristics  $CDA$ ,  $CS$ ,  $CF$ , and  $PF$ , and estimates of the recurrence-interval flood discharges at the site of interest are computed.

The number ( $p$ ) and identity of basin characteristics that are used to compute  $d_{ij}$  and the  $N$  gaging stations that compose the region-of-influence are specific to a given set of flood-discharge estimates and basin characteristics. In order to adapt the region-of-influence method to that data set, these parameters must be determined. In addition to these parameters, the set of basin characteristics also must be chosen for use as explanatory variables in the generalized least-squares regression equations developed for each recurrence-interval peak discharge at the site of interest.

A subtle but important distinction exists between the two sets of basin characteristics—the first is used to define the region-of-influence for the site of interest; the second serves as explanatory variables that may or may not be used in the unique predictive equations that are developed for the site. These two sets of basin characteristics need not be identical but are in some cases. In other cases, such as in this study, the set of basin characteristics used to define the region-of-influence is a fixed subset ( $CDA$ ,  $CS$ , and  $CF$ ) of the set of characteristics that potentially can be included in the predictive equations for the site of interest ( $CDA$ ,  $CS$ ,  $CF$ , and  $PF$ ).

The number of gaging stations ( $N$ ) and the basin characteristics that are used to define the region-of-influence for unregulated sites in Tennessee were selected by using a computer program that computes prediction error for various combinations of  $N$  and basin characteristics. One of the best measures of the quality of a regression equation is the Prediction Error Sum of Squares ( $PRESS$ ) statistic (Helsel and Hirsch, 1992).  $PRESS$  is a validation-type estimator of error. Instead of splitting a data set in half, one half to

develop the equation, and the second to validate the equation, the  $PRESS$  statistic uses  $N-1$  observations to develop the equation, then estimates the value of the observation left out. The  $PRESS$  statistic then changes the omitted observation, and repeats the process for each observation. The prediction errors are squared and summed. In multiple regression, the  $PRESS$  statistic is a useful estimate of the quality of competing regression equations. An interactive computer program that computes the  $PRESS$  statistic was used to determine the characteristics of the region-of-influence method in Tennessee. Various combinations of  $N$  (20, 30, 40, 50, 60, and 70) and basin characteristics ( $CDA$ ,  $CS$ ,  $CF$ , and  $PF$ ) were compared by using the  $PRESS$  computer program and a trial and error process to select the characteristics of the region-of-influence computer application in Tennessee.

As implemented in Tennessee, the region-of-influence method compares basin characteristics for all 453 gaging stations and selects 60 sites having basin characteristics most similar to the site of interest.  $CDA$ ,  $CS$ , and  $CF$  are the basin characteristics used in the distance or similarity metric that defines the region-of-influence for unregulated sites in Tennessee.

To estimate recurrence-interval discharges at an unregulated site of interest, the region-of-influence method performs generalized least-squares regression using  $CDA$ ,  $CS$ ,  $PF$ , and  $CF$  from the 60 most similar sites. Because generalized least-squares regression was used to develop the predictive equations, the prediction-error departures and the 90-percent prediction interval are computed for each recurrence-interval peak discharge as described in appendix A.

The region-of-influence computer application for Tennessee will add or drop basin characteristics to or from a given recurrence-interval regression equation by performing a significance test ( $\alpha = 0.10$ , two-tailed  $t$  test) for each basin characteristic. Therefore, a site of interest can have recurrence-interval regression equations with different combinations of basin characteristics. This freedom was built into the region-of-influence method to maximize flexibility, but occasionally minor inconsistencies are produced in the recurrence-interval discharge estimates for a given site of interest.

For sites of interest having combinations of basin characteristics near the outer limits of the basin-characteristic data space, a subsequent recurrence-interval discharge estimate may be less than the previous lower recurrence-interval discharge estimate, for

example,  $Q_{100}$  less than  $Q_{50}$ . When inconsistent discharge estimates occur, the region-of-influence method uses a smoothing procedure to adjust the inconsistent values.

If the inconsistent point is an interior point ( $Q_5$  to  $Q_{100}$ ), then the point is estimated based on a linear interpolation on a log-probability scale defined by the preceding and following points. For example, if the region-of-influence method estimates a  $Q_{100}$  less than the  $Q_{50}$ , then a new  $Q_{100}$  is estimated based on a straight line on a log-probability scale between the  $Q_{50}$  and  $Q_{500}$  estimates.

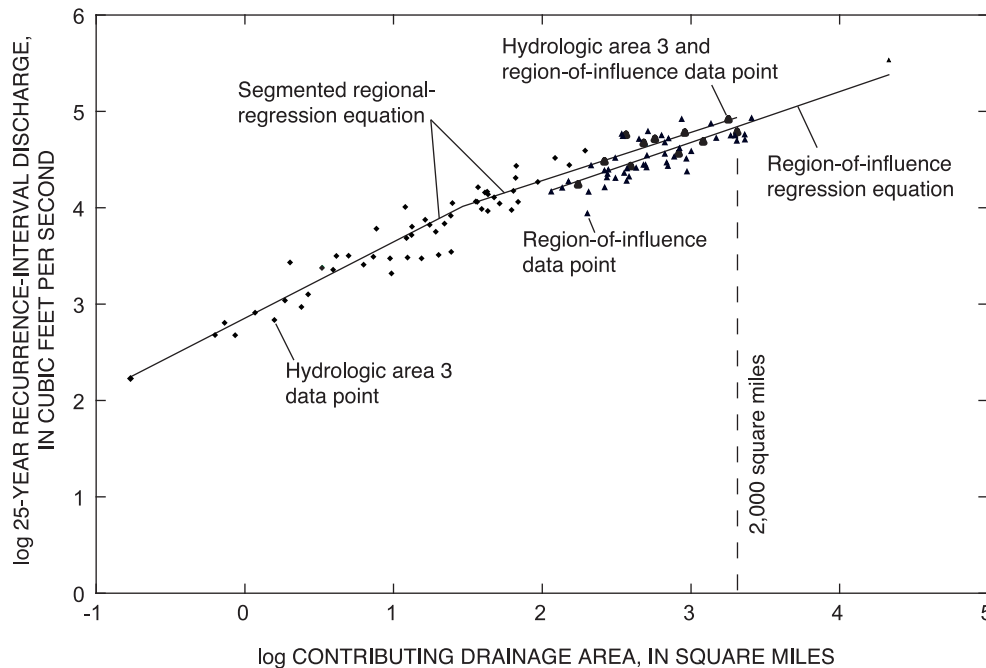
If the inconsistent point is an end point, for example  $Q_{500}$  less than  $Q_{100}$ , then the next-to-end point and the end point are adjusted based on a straight line on a log-probability scale from the second-to-the-end point with a slope defined as the average between a slope from the second-to-the-end point through the next-to-end point and a slope from the second-to-end point and the end point, which was estimated by regression. When the smoothing procedure is used, the region-of-influence method provides both the unadjusted and adjusted discharge estimates for easy identification and comparison by the user.

### Comparison of Methods

When comparing accuracy estimates for the regional-regression method and the region-of-

influence method at a particular site of interest, the following points should be considered. Occasionally, the scatter of data about a regional-regression equation has a subtle downward curving appearance. This slight curvature can be overcome by segmenting the data into two drainage-area ranges and fitting a regression equation to each range (fig. 3). This is essentially what the region-of-influence method does by placing the site of interest as near the center of a regression equation as possible. The negative and positive prediction-error departures are calculated assuming that the scatter about the fitted regression equation is uniform throughout the range of the data for every recurrence interval, which may not always be the case. In such cases, the regional-regression method, which uses the average scatter for the entire range of the data in the calculation, may produce a relatively poor estimate of the prediction-error departures for a particular site.

The region-of-influence method takes advantage of the non-uniform distribution of the data (scatter), limiting the data used to develop regression equations and associated error estimates to a small range around *CDA* for the particular site (fig. 4). Thus, in some hydrologic areas, the region-of-influence method can be expected to provide a better “local” estimate of the peak at the site of interest. Further, the region-of-influence method also may provide a better estimate of the “local” accuracy of that peak than the regional-regression method, even in those instances where the



**Figure 4.** Segmented regional-regression equation and region-of-influence regression equation for the 25-year flood at a 2,000-square-mile ungaged site in hydrologic area 3.



estimates of the prediction error from the computer application are smaller for the regional-regression method.

The deleted-residual standard error,  $S_{(-)}$ , for a regression method is the square root of the average prediction error sum of squares,  $(PRESS/N)^{1/2}$ . The  $S_{(-)}$  is used to compare the predictive ability of regression methods with differing degrees of freedom. The deleted-residual standard error for a regression method in percent,  $\%S_{(-)}$ , is computed as:

$$\%S_{(-)} = 100\{[e^{5.3026(PRESS/N)} - 1]^{1/2}\}. \quad (10)$$

*PRESS* is the sum of the squared residuals obtained by subtracting the flood-frequency estimate determined by using Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) from the flood-frequency computed using a regression method. The *PRESS* statistic was previously described in the Region-of-Influence Method section of this report; and *N* is the number of residuals summed to produce the *PRESS* statistic. Comparison of the  $\%S_{(-)}$  values indicates that, in general, the region-of-influence method is slightly more accurate than the regional-regression method (table 8). In most cases, deleted-residual standard errors are slightly less for the region-of-influence method than for the regional-regression equations, and about 5 percent less than some of the single-variable equations.

Using the computer application, little difference exists in the ease of application between the region-of-influence method and the regional-regression equations. A comparison of the region-of-influence method and the regional-regression equations based on the overall predictive ability of the methods indicates that the region-of-influence method is, on average, the better of the two methods tested for predicting flood frequency for unregulated streams and rivers in Tennessee.

### Use of Computer Application

Application of the single-variable regional-regression equations requires much less effort than the multivariable regional-regression equations or the region-of-influence method. The single-variable regional-regression equations require input of *CDA* only, and the computation of the estimate is simple. Therefore, the single-variable equations should be used in the absence of the flood-frequency computer application. The need to provide *CS*, and possibly *CF*,

**Table 8.** Comparison of deleted-residual standard error for the region-of-influence method and regional-regression equations

[See figure 1 for hydrologic area locations. ----, not applicable]

Recurrence interval, in years	Deleted-residual standard error, in percent		
	Region-of-influence method	Regional-regression equations	
		Multivariable	Single-variable
<b>Hydrologic area 1</b>			
2	40.8	40.8	45.7
5	40.3	41.1	45.7
10	41.9	43.0	47.2
25	45.1	46.7	50.1
50	48.3	50.0	52.9
100	52.2	53.7	56.1
500	61.6	63.4	64.8
<b>Hydrologic area 2</b>			
2	29.4	32.2	33.5
5	27.5	31.1	32.3
10	29.3	33.1	34.0
25	33.3	36.8	37.4
50	37.0	40.1	40.3
100	40.3	43.5	43.5
500	50.7	52.0	51.5
<b>Hydrologic area 3</b>			
2	32.2	34.7	33.2
5	31.8	34.3	33.0
10	32.4	35.4	34.2
25	35.0	37.6	36.5
50	37.7	39.7	38.6
100	40.2	42.0	41.0
500	45.5	48.0	47.0
<b>Hydrologic area 4</b>			
2	38.0	----	41.7
5	37.2	----	40.6
10	40.0	----	41.7
25	43.0	----	44.3
50	46.3	----	46.7
100	50.2	----	49.4
500	57.1	----	56.4

make the multivariable regional-regression equations difficult to apply manually. The region-of-influence method is computationally intensive and is not suitable for manual application. However, each of the methods can be easily applied using a personal computer.

The flood-frequency computer application for Tennessee estimates flood frequency at unregulated sites by using all three methods for easy comparison by the user. Therefore, in addition to *CDA* and *CS*, the latitude (*LAT*), longitude (*LNG*), and hydrologic area(s) (*HA*) of the site of interest must be specified. The explanatory variables *CF* and *PF* are automatically computed using the *LAT*, *LNG*, and *HA*(s) of the site of interest. Tennessee's flood-frequency computer application automatically adjusts flood discharges for watersheds draining two hydrologic areas.

The flood-frequency computer application for Tennessee includes the following six files (approximate size shown in parentheses): (1) an executable main-program file named *TDOTv203.exe* (437 kilobytes); (2) an external subroutine used by the main executable program named *tnff.cmn* (1 kilobyte); and four supporting data files: (3) *cgrid.krg* (41 kilobytes), (4) *v203inp.txt* (91 kilobytes), (5) *v203M1* (413 kilobytes), and (6) *v203M2* (413 kilobytes). These files should be located in a common directory on the computer hard drive for the flood-frequency application to function properly. The flood-frequency computer application can be downloaded from the Tennessee District homepage at <http://tn.water.usgs.gov>.

Each time the flood-frequency computer application is executed, flood-frequency estimates are produced by using the single-variable and multivariable regional-regression equations, and the region-of-influence method. The computer application produces on-screen summary of results and generates two user-named output files containing the results of flood-frequency estimates at unregulated sites in Tennessee. The first user-named output file (fig. 5), which is identical to the on-screen output, contains discharge pre-

dictions, negative and positive prediction-error departures, and 90-percent prediction intervals for each recurrence interval. The second output file (table B-1 in appendix B) contains detailed diagnostic information for the region-of-influence method including a listing of the gaging stations in the region-of-influence and their respective basin characteristics; and the significant regression coefficients for each recurrence-interval discharge, the observed and regression-predicted discharges, residual and influence statistics for the stations in the region-of-influence including standardized residual, leverage, and *Cook's D*; and overall quality measures for the regression.

Suggested procedures for estimating flood frequency at unregulated streams and rivers in Tennessee are as follows:

- Determine the latitude (*LAT*) and longitude (*LNG*), in degrees, minutes, and seconds, of the site of interest.
- Determine the hydrologic area(s) (*HA*) of the drainage basin upstream from the site of interest.
- Determine the contributing drainage area (*CDA*), in square miles, and the main-channel slope (*CS*), in feet per mile, of the site of interest using the best available information. If there are two *HAs*, determine the proportion of *CDA* that lies within each *HA*.

To assist the user of the flood-frequency computer application for Tennessee, the following suggested ranges for *CDA* and *CS* (table 9) are provided on screen while the computer application is in use. Supplying input to the computer program that is within these ranges will decrease the chance of generating an extrapolated estimate beyond the range of the basin-characteristic data. However, values of *CDA* and *CS* that are within the ranges shown in table 9, when taken in combination, could be outside the basin-characteristic data space, thus producing an extrapolated result at the site of interest.

**Table 9.** Suggested ranges for contributing drainage area and main-channel slope for input to the computer application

Hydrologic area	Contributing drainage area, in square miles		Main-channel slope, in feet per mile	
	Lower	Upper	Lower	Upper
1	0.20	9,000	3.29	950
2	.47	2,557	1.90	343
3	.17	2,048	2.12	132
4	.76	2,308	.89	63

TDOT Version 2.0.3

SINGLE-VARIABLE REGIONAL-REGRESSION EQUATION (SRE) METHOD FOR TENNESSEE

Flood frequency estimates for:

Big River at Centerville, TN

Hydrologic Areas (percent): HA 3 ( 80.0) HA 2 ( 20.0)

LAT: 35 50 10 LNG: 87 25 30

Explanatory variable:

Contributing drainage area: 2000.00 square miles

RI	DISCHARGE	- SE (%)	+ SE (%)	90% PRED. INTERVAL	
	(cfs)				
2	39700.0	-24.4	32.3	25000.0	63200.0
5	59600.0	-24.5	32.4	37400.0	94900.0
10	73700.0	-25.5	34.3	45200.0	120000.0
25	92400.0	-27.4	37.7	54300.0	157000.0
50	107000.0	-29.0	40.9	60600.0	189000.0
100	122000.0	-30.7	44.3	66400.0	224000.0
500	160000.0	-34.7	53.2	78500.0	324000.0

MULTIVARIABLE REGIONAL-REGRESSION EQUATION (MRE) METHOD FOR TENNESSEE

Flood frequency estimates for:

Big River at Centerville, TN

Hydrologic Areas (percent): HA 3 ( 80.0) HA 2 ( 20.0)

LAT: 35 50 10 LNG: 87 25 30

Explanatory variables:

Contributing drainage area: 2000.00 square miles

Channel slope: 2.50 ft/mi

RI	DISCHARGE	- SE (%)	+ SE (%)	90% PRED. INTERVAL	
	(cfs)				
2	39900.0	-24.9	33.1	24800.0	64200.0
5	59400.0	-25.0	33.4	36800.0	95900.0
10	73400.0	-26.2	35.5	44400.0	122000.0
25	92100.0	-28.2	39.3	53100.0	160000.0
50	107000.0	-29.9	42.7	59200.0	193000.0
100	122000.0	-31.7	46.4	64800.0	230000.0
500	160000.0	-35.9	55.9	76500.0	334000.0

REGION-OF-INFLUENCE (ROI) METHOD FOR TENNESSEE

Flood frequency estimates for:

Big River at Centerville, TN

Hydrologic Areas (percent): HA 3 ( 80.0) HA 2 ( 20.0)

LAT: 35 50 10 LNG: 87 25 30

Explanatory variables:

Contributing drainage area: 2000.00 square miles

Channel slope: 2.50 ft/mi

Climate factor: 2.38

Log(Physiographic Factor): 0.152( HA 3 ) 0.133( HA 2 )

RI	DISCHARGE	- SE (%)	+ SE (%)	90% PRED. INTERVAL	
	(cfs)				
2	38800.0	-19.7	24.6	26900.0	55900.0
5	56300.0	-20.2	25.3	38700.0	81900.0
10	68400.0	-20.9	26.5	46300.0	101000.0
25	88900.0	-22.4	28.8	58400.0	135000.0
50	101000.0	-23.4	30.5	65200.0	158000.0
100	114000.0	-24.4	32.3	71600.0	182000.0
500	145000.0	-27.2	37.4	85500.0	246000.0

Figure 5. Sample of summary output file produced by flood-frequency computer application.

## APPLICATION OF METHODS

Methods of estimating flood discharges for unregulated streams in Tennessee vary depending on the amount of data available at a site of interest. These methods are designed for use at streams with unregulated flows, including sites on streams and rivers that flow into Tennessee from adjacent states.

Several points to consider when estimating flood-frequency of streams and rivers in Tennessee are as follows:

- Determine that the stream or river is not appreciably regulated; if regulated, regression methods presented in this report should not be used.
- Search for streamgage data at the site of interest; if available, this information should be weighted with the regression estimate using the methods presented in this section.
- Search for streamgage data for nearby stations on the same stream; if available, this information should be combined with the regression estimate using the methods presented in this section.

Flood-peak estimates suitable for design purposes at gaged sites can best be determined by a combined use of the log-Pearson Type III station estimates (Interagency Advisory Committee on Water Data, 1982) and the regression-method estimates. In this study, region-of-influence method estimates are used in the computation of weighted discharge estimates at gaging stations in Tennessee.

Weighted discharge estimates computed from station estimates and regression estimates are given for 297 gaging stations located in Tennessee (table 4 at back of report). The weighted value is based on the effective record length, in years, at the gaging station (table 4 and appendix C at back of report) and the equivalent years of record for the region-of-influence method estimate (example in table B-1; and appendix D at back of report). Weighted discharge values in table 4 were computed using the dominant hydrologic area for each station. The equation below is used to compute the weighted value at gaging stations:

$$\log_{10}(Q_t(w)) = \{[\log_{10}(Q_t(g))N_e] + [\log_{10}(Q_t(r))EY]\} / (N_e + EY), \quad (11)$$

where

$\log_{10}(Q_t(w))$  is the logarithm of the weighted discharge at the gaging station for recurrence interval  $t$ ;

$\log_{10}(Q_t(g))$  is the logarithm of the discharge for recurrence interval  $t$  determined using systematic and historical peak-flow record from the gaged site;

$\log_{10}(Q_t(r))$  is the logarithm of the discharge for recurrence interval  $t$  determined using the region-of-influence method;

$N_e$  is the number of systematic peaks in the gaging-station record, or the effective record length, in years, (table 4) computed using the method described in appendix C if adjusted for historical information;

$EY$  is the equivalent years of record for the region-of-influence method estimate (example in table B-1).

Flood-frequency estimates at a site of interest that is on the same stream as a gaging station can be determined by using a combination of the regression estimate for the site of interest and the station estimate for the nearby gaged site. In order to make the appropriate adjustment, first compute the ratio,

$$R = Q_t(w) / Q_t(r), \quad (12)$$

for the gaged site by using  $(Q_t(w))$  and  $(Q_t(r))$  as defined in the preceding paragraph. Next, a correction factor,  $R'$ , is computed as follows:

$$R' = R - (\Delta CDA(R-1)/0.5CDA_g), \quad (13)$$

where

$\Delta CDA$  is the absolute value of the difference between the contributing drainage areas of the gaged site and site of interest, and

$CDA_g$  is the contributing drainage area of the gaged site.

If  $\Delta CDA/CDA_g$  is less than 0.5, then the corrected discharge for the site of interest,  $(Q_t(\text{corr}))$ , can be computed by multiplying the correction factor,  $R'$ , by the regression estimate for the site of interest  $(Q_t(r))$ . If  $\Delta CDA/CDA_g$  is greater than 0.5, or no station data are available, then select the regression method having the better prediction error and use the results without correction.

At times, flood-frequency estimates may be needed for a site of interest that is between two gaged sites on the same stream. In this case, select the gaged site for which  $\Delta CDA/CDA_g$  is less than 0.5, compute  $R'$ , and apply as described above. If  $\Delta CDA/CDA_g$  is

less than 0.5 for both gaged sites, compute  $R'$  for each. If both correction factors are greater than 1.0, then use the larger  $R'$ ; if both correction factors are less than 1.0, then use the smaller  $R'$ . If one correction factor is greater than 1.0 and the other smaller than 1.0, then an average of the two correction factors should be used.

If the drainage basin for a site of interest lies within two hydrologic areas ( $HA_i$  and  $HA_j$ ), then the computed discharge should be adjusted according to the proportion of the total contributing drainage area that lies within each hydrologic area. The adjusted discharge can be determined by the equation:

$$(Q_t)(\text{adjusted}) = Q_t(HA_i)(CDA_i/CDA_{\text{total}}) + Q_t(HA_j)(CDA_j/CDA_{\text{total}}), \quad (14)$$

where

$(Q_t)(\text{adjusted})$  is the adjusted discharge for the  $t$ -year recurrence interval,

$(Q_t)(HA_i)$  and  $(Q_t)(HA_j)$  are the discharges computed as if the entire contributing drainage area were within the hydrologic areas,  $HA_i$  and  $HA_j$ , respectively,

$CDA_i$  and  $CDA_j$  are the total contributing drainage areas within each of the respective hydrologic areas, and

$CDA_{\text{total}}$  is the sum of the total contributing drainage areas within each of the respective hydrologic areas.

## SUMMARY

Reliable and accurate estimates of the magnitude and frequency of floods are needed for the design of bridges and culverts, the delineation and management of flood zones, and the management of water-control structures. The U.S. Geological Survey, in cooperation with the Tennessee Department of Transportation, applied the region-of-influence method to improve estimates of flood frequency for unregulated streams and rivers in Tennessee. For comparison with the region-of-influence method, the regional-regression method for estimating flood frequency at unregulated sites was updated and expanded to include single-variable and multivariable regression equations. The prediction methods are part of an interactive com-

puter application used to estimate flood frequency at unregulated streams and rivers in Tennessee. The computer application allows for easy comparison of results from both of the regression methods.

Annual-peak streamflow records, historical flood information, and selected basin characteristics for streamgages in the study area with 10 or more years of record through water year 1999 were combined to form a database that was used to develop the prediction methods for use at unregulated sites in Tennessee. These stations measure the flow in streams draining basins with 1 percent to about 30 percent total impervious area; these methods should not be used on regulated streams, or in heavily developed or storm-sewered basins with impervious areas greater than 10 percent. Flood frequency at each of the gaging stations used in this study was computed by fitting the peak streamflow data and supplemental historic information for each station to the log-Pearson Type III distribution as described in Bulletin 17B of the Interagency Advisory Committee on Water Data (1982).

Basin characteristics and flood-frequency estimates for 453 gaging stations located in Tennessee and six adjacent States were merged to form the database that was used to develop the regional-regression equations described in this report. Of the 453 stations, 297 are located in Tennessee, 21 in Georgia, 37 in North Carolina, 28 in Virginia, 20 in Alabama, 36 in Kentucky, and 14 in Mississippi. For the regional-regression method, generalized least-squares regression was used to develop single-variable and multivariable regression equations for the hydrologic areas of Tennessee. The regional-regression equations can be used to compute the 2-, 5-, 10-, 25-, 50-, 100-, and 500-year recurrence-interval flood discharges at unregulated streams and rivers using contributing drainage area, main-channel slope, and a climatic factor.

The region-of-influence method was applied in Tennessee using the same 453 gaging stations that were used to develop the regional-regression equations. For an unregulated site of interest, the region-of-influence is defined as the 60 most similar stations selected from the database. The region-of-influence for a site of interest is determined by comparing the contributing drainage area, main-channel slope, and climate factor of the gaged sites to the site of interest. The region-of-influence method uses generalized least-squares regression to estimate the 2-, 5-, 10-, 25-,

50-, 100-, and 500-year recurrence-interval flood discharges at unregulated sites using contributing drainage area, main-channel slope, a climatic factor, and a physiographic-region factor as explanatory variables. The physiographic-region factor allows the region-of-influence method to capture the uniqueness in flood-magnitude potential inherent in the four hydrologic areas in Tennessee, which are based on physiographic provinces.

The regional-regression equations, in particular the single-variable regression equations, are easy to solve manually and are an alternative that can be used to obtain estimates of flood frequency at unregulated sites in Tennessee if the computer application, and therefore the region-of-influence method, is not available. A comparison of the regional-regression method to the region-of-influence method, based on average predictive ability of the methods, indicates that the region-of-influence method is the better method of the two methods tested for predicting flood frequency in Tennessee. The flood-frequency computer application for Tennessee can be downloaded from the website <http://tn.water.usgs.gov>.

## REFERENCES

- Acreman, M.C., and Wiltshire, S.E., 1987, Identification of regions for regional flood frequency analysis [abs.]: EOS, v. 68, no. 44, p. 1262.
- Asquith, W.H., and Slade, R.M., Jr., 1999, Site-specific estimation of peak-streamflow frequency using generalized least-squares regression for natural basins in Texas: U.S. Geological Survey Water-Resources Investigations Report 99-4172, 19 p.
- Atkins, J.B., 1996, Magnitude and frequency of floods in Alabama: U.S. Geological Survey Water-Resources Investigations Report 95-4199, 234 p.
- Bisese, J.A., 1995, Methods for estimating the magnitude and frequency of peak discharges of rural, unregulated streams in Virginia: U.S. Geological Survey Water-Resources Investigations Report 94-4148, 70 p.
- Bobee, Bernard, 1973, Sample error of T-year events computed by fitting a Pearson Type 3 Distribution: Water Resources Research, v. 9, no. 5, p. 1264-1270.
- Burn, D.H., 1990a, An appraisal of the "region of influence" approach to flood frequency analysis: Hydrological Sciences Journal, v. 35, no. 2, p. 149-165.
- Burn, D.H., 1990b, Evaluation of regional flood frequency analysis with a region of influence approach: Water-Resources Research, v. 26, no. 10, p. 2257-2265.
- Dalrymple, Tate, 1960, Flood frequency analyses: U.S. Geological Survey Water-Supply Paper 1543-A, 80 p.
- Dickson, R.R., 1960, Climates of the states—Tennessee: U.S. Department of Commerce, Weather Bureau, Climatology of the United States no. 60-40, 13 p.
- Feaster, T.D., and Tasker, G.D., 2002, Techniques for estimating the magnitude and frequency of floods in rural basins of South Carolina, 1999: U.S. Geological Survey Water-Resources Investigations Report 02-4140, 34 p.
- Fenneman, N.M., 1946, Physical divisions of the United States: U.S. Geological Survey special map, scale 1:7,000,000.
- Hardison, C.H., 1971, Prediction error of regression estimates of streamflow characteristics at ungaged sites, in Geological Survey Research 1971: U.S. Geological Survey Professional Paper 750-C, p. C228-C236.
- Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: Amsterdam, Elsevier, 522 p.
- Hodge, S.A., and Tasker, G.D., 1995, Magnitude and frequency of floods in Arkansas: U.S. Geological Survey Water-Resources Investigations Report 95-4224, 52 p.
- Hodgkins, G.A., and Martin, G.R., in press, Estimating the magnitude of peak flows for streams in Kentucky for selected recurrence intervals: U.S. Geological Survey Water-Resources Investigations Report 03-4180.
- Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency: U.S. Geological Survey, Office of Water Data Coordination Bulletin 17B, 186 p.
- Jenkins, C.T., 1960, Floods in Tennessee, magnitude and frequency: Tennessee Department of Highways, 68 p.
- Landers, M.N., and Wilson, K.V., Jr., 1991, Flood characteristics of Mississippi streams: U.S. Geological Survey Water-Resources Investigations Report 91-4037, 82 p.
- Lichty, R.W., and Karlinger, M.R., 1990, Climate factor for small-basin flood frequency: Water Resources Bulletin, v. 26, no. 4, p. 577-586.
- Lichty, R.W., and Liscum, Fred, 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.
- Miller, R.A., 1974, The geologic history of Tennessee: Tennessee Division of Geology Bulletin 74, 63 p.
- Moss, M.E., and Karlinger, M.R., 1974, Surface water network design by regression analysis simulation: Water Resources Research, v. 10, no. 3, p. 427-433.
- Patterson, J.L., 1964, Magnitude and frequency of floods in the United States, Part 7, Lower Mississippi River basin: U.S. Geological Survey Water-Supply Paper 1681, 636 p.
- Pope, B.F., Tasker, G.D., and Robbins, J.C., 2001, Estimating the magnitude and frequency of floods in rural basins of North Carolina—Revised: U.S. Geological Survey Water-Resources Investigations Report 01-4207, 44 p.

- Randolph, W.J., and Gamble, C.R., 1976, Technique for estimating magnitude and frequency of floods in Tennessee: Tennessee Department of Transportation, 52 p.
- Riggs, H.C., 1973, Regional analyses of streamflow characteristics: Techniques of Water-Resources Investigations of the U.S. Geological Survey, Book 4, Chapter B3, 15 p.
- Speer, P.R., and Gamble, C.R., 1964, Magnitude and frequency of floods in the United States, pt. 3B, Cumberland and Tennessee River basins: U.S. Geological Survey Water-Supply Paper 1676, 340 p.
- Stamey, T.C., and Hess, G.W., 1993, Techniques for estimating magnitude and frequency of floods in rural basins of Georgia: U.S. Geological Survey Water-Resources Investigations Report 93-4016, 75 p.
- Stedinger, J.R., and Cohn, T.A., 1985, Flood frequency analysis with historical and Paleoflood information: *Water Resources Research*, v. 22, p. 785-793.
- Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic regression: 1. Ordinary, weighted and generalized least squares compared: *Water Resources Research*, v. 21, no. 9, p. 1421-1432.
- Tasker, G.D., 1982, Simplified testing of hydrologic regression regions: *Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers*, v. 108, no. HY10, p. 1218-1222.
- Tasker, G.D., Hodge, S.A., and Barks, C.S., 1996, Region of influence regression for estimating the 50-year flood at ungaged sites: *Water Resources Bulletin*, v. 32, no. 1, p. 163-170.
- Tasker, G.D., and Slade, R.M., 1994, An interactive regional regression approach to estimating flood quantiles, *in* Fontane, D.G., and Tuel, H.N., eds., *Water policy and management, solving the problems, Proceedings of the Twenty-First Annual Conference*, May 23-26, 1994: Denver, Colo., Denver Society of American Society of Civil Engineers, p. 782-785.
- Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: *Journal of Hydrology*, v. 111, p. 361-375.
- Tasker, G.D., and Thomas, W.O., 1978, Flood frequency analysis with pre-record information: *Journal of Hydraulic Division of American Society of Civil Engineers*, v. 104, no. 2, p. 249-259.
- U.S. Geological Survey, 1970, *The national atlas of the United States of America*: Washington, D.C., U.S. Geological Survey, 417 p.
- U.S. Water Resources Council, 1976, *Guidelines for determining flood flow frequency*: U.S. Water Resources Council Bulletin 17, 221 p.
- Weaver, J.D., and Gamble, C.R., 1993, Flood frequency of streams in rural basins of Tennessee: U.S. Geological Survey Water-Resources Investigations Report 92-4165, 38 p.
- Wibben, H.C., 1976, Effects of urbanization on flood characteristics in Nashville-Davidson County, Tennessee: U.S. Geological Survey Water-Resources Investigations Report 76-121, 33 p.