U.S. Department of the Interior
U.S. Geological Survey

Estimating the Magnitude of Peak Flows for Streams in Kentucky for Selected Recurrence Intervals

By **Glenn A. Hodgkins** *and* **Gary R. Martin**

Water-Resources Investigations Report 03-4180

In cooperation with the Kentucky Transportation Cabinet, Department of Highways

Louisville, Kentucky
2003

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

U.S. GEOLOGICAL SURVEY
Charles G. Groat, Director

For additional information contact:

District Chief, Kentucky District
U.S. Geological Survey
9818 Bluegrass Parkway
Louisville, KY 40299–1906
http://ky.water.usgs.gov

Copies of this report can be purchased from:

U.S. Geological Survey
Information Services
Box 25286
Denver, CO 80225–0286

CONTENTS

PLATE

FIGURES

TABLES

CONVERSION FACTORS

Multiply	By	To obtain
inch (in.)	25.4	millimeter
inch per hour (in/h)	0.0254	meter per hour
foot (ft)	0.3048	meter
foot per mile (ft/mi)	0.1894	meter per kilometer
cubic foot (ft^3)	0.02832	cubic meter
cubic foot per second (ft^3)	0.02832	cubic meter per second
mile (mi)	1.606	kilometers
square mile (mi^2)	2.590	square kilometer

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

°C = (°F - 32) / 1.8

Estimating the Magnitude of Peak Flows for Streams in Kentucky for Selected Recurrence Intervals

By Glenn A. Hodgkins *and* Gary R. Martin

Abstract

This report gives estimates of, and presents techniques for estimating, the magnitude of peak flows for streams in Kentucky for recurrence intervals of 2, 5, 10, 25, 50, 100, 200, and 500 years. A flowchart in this report guides the user to the appropriate estimates and (or) estimating techniques for a site on a specific stream.

Estimates of peak flows are given for 222 U.S. Geological Survey streamflow-gaging stations in Kentucky. In the development of the peak-flow estimates at gaging stations, a new generalized skew coefficient was calculated for the State. This single statewide value of 0.011 (with a standard error of prediction of 0.520) is more appropriate for Kentucky than the national skew isoline map in Bulletin 17B of the Interagency Advisory Committee on Water Data.

Regression equations are presented for estimating the peak flows on ungaged, unregulated streams in rural drainage basins. The equations were developed by use of generalized-least-squares regression procedures at 187 U.S. Geological Survey gaging stations in Kentucky and 51 stations in surrounding States. Kentucky was divided into seven flood regions. Total drainage area is used in the final regression equations as the sole explanatory variable, except in Regions 1 and 4 where main-channel slope also was used. The smallest average standard errors of prediction were in Region 3 (from -13.1 to +15.0 percent) and the largest average standard errors of prediction were in Region 5 (from -37.6 to +60.3 percent).

One section of this report describes techniques for estimating peak flows for ungaged sites on gaged, unregulated streams in rural drainage basins. Another section references two previous U.S. Geological Survey reports for peak-flow estimates on ungaged, unregulated, urban streams. Estimating peak flows at ungaged sites on regulated streams is beyond the scope of this report, because peak flows on regulated streams are dependent upon variable human activities.

INTRODUCTION

Estimates of the magnitude of peak streamflows (such as the 50-year-recurrence-interval peak flow) are necessary to safely and economically design bridges, culverts, and other structures that are in or near streams. These estimates also are needed by Federal, State, regional, and local officials for effective flood-plain management. This report, prepared by the U.S. Geological Survey (USGS) in cooperation with the Kentucky Transportation Cabinet (KTC), will help KTC and many others make improved estimates of the magnitude of peak flows for Kentucky streams.

This report gives estimates of, and presents techniques for estimating, the magnitude of peak flows for streams in Kentucky for recurrence intervals of 2, 5, 10, 25, 50, 100, 200, and 500 years. Peak flows are listed for USGS streamflow-gaging stations with 10 years or more of recorded annual peak flows through water year[1] 2000.

[1]A water year is the 12-month period from October 1 through September 30, and it is designated by the calendar year in which it ends.

A technique is presented for estimating the peak flows for ungaged, unregulated streams in rural drainage basins. Techniques also are described for estimating peak flows at ungaged sites on gaged streams (for unregulated sites in rural drainage basins). Two reports are referenced for estimating peak flows on ungaged, unregulated streams in urbanized drainage basins. A technique for estimating peak flows for ungaged sites on regulated streams is beyond the scope of this report, although a possible approach is mentioned and cautions about inappropriate approaches are given.

Various peak-flow studies have been published applicable to all or parts of Kentucky since 1958 (McCabe, 1958, 1962; Speer and Gamble, 1964, 1965; Hannum, 1976; Wetzel and Bettandorff, 1986; Choquette, 1988). Each succeeding report generally used more years of hydrologic data and more rigorous statistical techniques. This report supersedes Choquette (1988) and the other reports in that the estimates and estimating techniques described in this report should provide improved estimates of rural peak flows for Kentucky. Advances in techniques for this report included the development of a generalized skew for Kentucky and the use of generalized-least-squares regression (explained later in the report).

The U.S. Army Corps of Engineers (USCOE) and the Tennessee Valley Authority generously provided peak-flow estimates for many regulated rivers. William J. Byron, Jr., USCOE, Louisville office, especially was helpful. This report would not be possible without nearly 100 years of peak-flow data collection, often under hazardous conditions, by USGS hydrologic technicians and hydrologists. This historical-data collection was done by the USGS in cooperation with the Commonwealth of Kentucky.

DESCRIPTION OF STUDY AREA

The Commonwealth of Kentucky encompasses an area of 40,395 mi^2 in the east-central United States. The major drainage basins in Kentucky—Big Sandy, Licking, Kentucky, Salt, Cumberland, Green, and Tennessee Rivers—are tributaries of the Ohio River (fig. 1). Variations in climate, physiography, and geology cause localized variations in streamflow characteristics in Kentucky.

Climate

Kentucky has a moist-continental climate with distinct seasonal variations and changeable weather patterns. Winter temperatures are moderate, rarely below 0°F; typical summer temperatures are warm and rarely above 100°F. Average annual snowfall is about 20 in., but the snow cover rarely remains longer than 3 days at a time. Weather patterns in Kentucky are affected variably by the meeting of cold, continental air masses arriving from the northwest and warm, moist air masses moving up the Mississippi and Ohio River Valleys from the southwest (Conner, 1982).

Annual precipitation in Kentucky averages about 47 in. The distribution of precipitation varies areally, annually, and seasonally. The mean annual precipitation in Kentucky ranges areally from about 41 to 53 in. Rainfall generally decreases to the north, reflecting the increase in distance from the source of moisture, which primarily is the subtropical Atlantic Ocean and Gulf of Mexico. Kentucky has considerable year-to-year variation in precipitation. During the period 1951-80, annual precipitation at reporting stations ranged from 14.5 to 78.6 in. Large amounts of precipitation in Kentucky have been associated with tropical cyclones moving north from the Gulf of Mexico (Conner, 1982).

Precipitation falls throughout the year but the sources and amounts of precipitation vary seasonally. Although March generally is the wettest month of the year, averaging from 4 to 6 in., the precipitation pattern is bimodal with a second peak, averaging from 3.3 to 5.5 in., occurring in July. October generally is the driest month when precipitation averages from 2 to 3 in. Mean seasonal precipitation in Kentucky is about 13.5 in. in spring (March through May), 12.4 in. in summer (June through August), 9.8 in. in fall (September through November), and 11.5 in. in winter (December through February) (Conner, 1982).

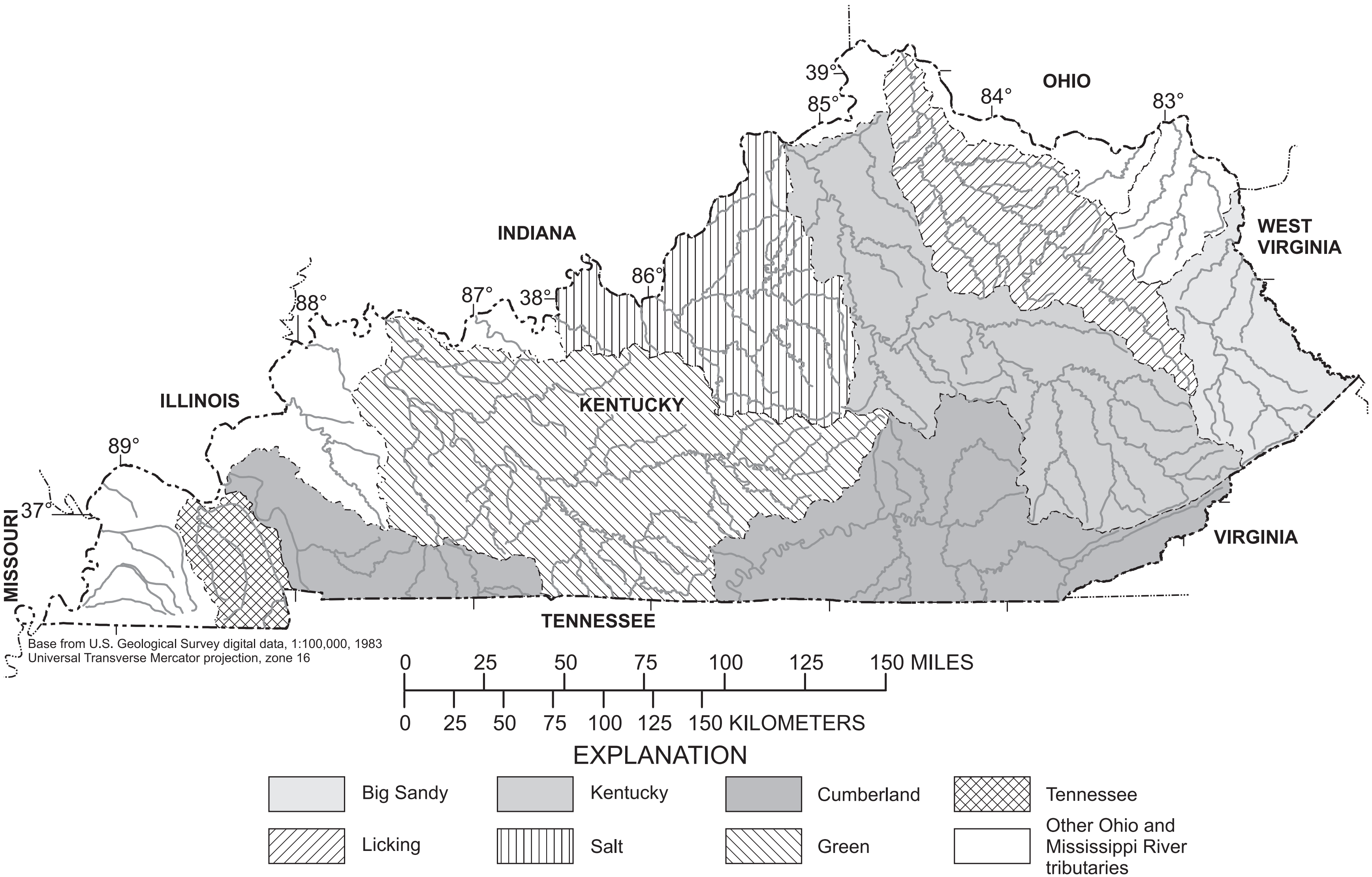


Figure 1. Major drainage basins in Kentucky.

Winter precipitation is characterized by frontal storm systems. Summer precipitation generally results from convective storm activity, commonly in the form of afternoon thunderstorms. Precipitation intensity generally is higher in summer than during other seasons, but the number of days having precipitation is similar in winter and summer. The Bermuda high-pressure system has a strong effect on seasonal precipitation patterns in Kentucky. In the fall, this high-pressure system generally moves inland from the southeastern coast of the United States and is centered over Kentucky and Tennessee, where it inhibits both convective activity and frontal storm movement and produces a dry season (Conner, 1982).

Physiography and Geology

Topographic relief in Kentucky (fig. 2) reflects the results of long-term stream-erosional processes in relation to the character of the rock formations. The upland areas—hills, ridges, mountains, and plateaus—generally consist of formations resistant to erosion. Western and central parts of Kentucky have rolling terrain, whereas the eastern part of Kentucky has rugged terrain with high relief. Land-surface elevations in Kentucky vary by more than 3,500 ft and range from 260 ft above sea level along the Mississippi River to 4,145 ft at the peak of Black Mountain in Harlan County near the Kentucky–Virginia border (McGrain and Currens, 1978).

The physiography of the State reflects the lithology of the surface rocks and largely is defined by the Cincinnati Arch (fig. 3). The axis of the Cincinnati Arch trends northward from south-central Kentucky to just south of the Outer Bluegrass boundary where it divides into two branches—Kankakee and Findlay Arches. The branches approximately are parallel but are separated by approximately 25 mi at the Ohio River (McFarland, 1950). Lithologic units dip away from the axis of the arch—a regional structural high—so that geologic features generally are symmetrical on each side of the arch.

Progressively younger rocks are exposed at the surface both east and west of the Cincinnati Arch. The oldest exposed rocks are part of the Jessamine Dome and adjacent areas; the location of this area corresponds approximately to the Inner Bluegrass region (fig. 3). These rocks consist of limestone, shale, and sandstone of Ordovician age. Narrow bands of shales and limestones of Silurian and Devonian age surround this area and correspond to The Knobs region. An expansive area of limestone of Mississippian age (Mississippian Plateaus Region) is exposed starting at the Ohio River in northeastern Kentucky, extending southwest to the State boundary, and extending northwest in a crescent-shaped area surrounding the Western Kentucky Coal Field. The eastern boundary of this area is the Cumberland Escarpment (fig. 3). Sandstones, shales, siltstones, and coals of Pennsylvanian age in eastern and northwestern Kentucky—the youngest rocks in Kentucky—compose the Eastern and Western Kentucky Coal Fields. Alluvial deposits of Cretaceous and Tertiary age are in extreme western Kentucky in the Mississippi Embayment.

Much of the Mississippian Plateau is characterized by carbonate rock and karst features such as sinkholes, caves, springs, and losing streams. Most well-developed karst features are located in a band originating in west-central Kentucky and extending to south-central Kentucky, southeast to the State boundary, east along the boundary, and then northeast and north (areas shown in black on fig. 4). Less well-developed karst features are in central and south-central Kentucky.

Seasonality of Peak Flows

Precipitation patterns strongly affect the magnitude and timing of peak flows. Seasonally changing conditions, such as evapotranspiration rates, antecedent soil moisture, and the extent, duration, and intensity of storm systems affect flood response in a given drainage basin.

The timing of peak flows varies with drainage-basin size. In basins with drainage areas from 50 to 1,000 mi^2, from 70 to 75 percent of the annual maximum peaks occur between January and April. About 45 percent of the peaks in basins from 0.1 to 10 mi^2 and 58 percent of the peaks in basins from 10 to 50 mi^2 occur between January and April.

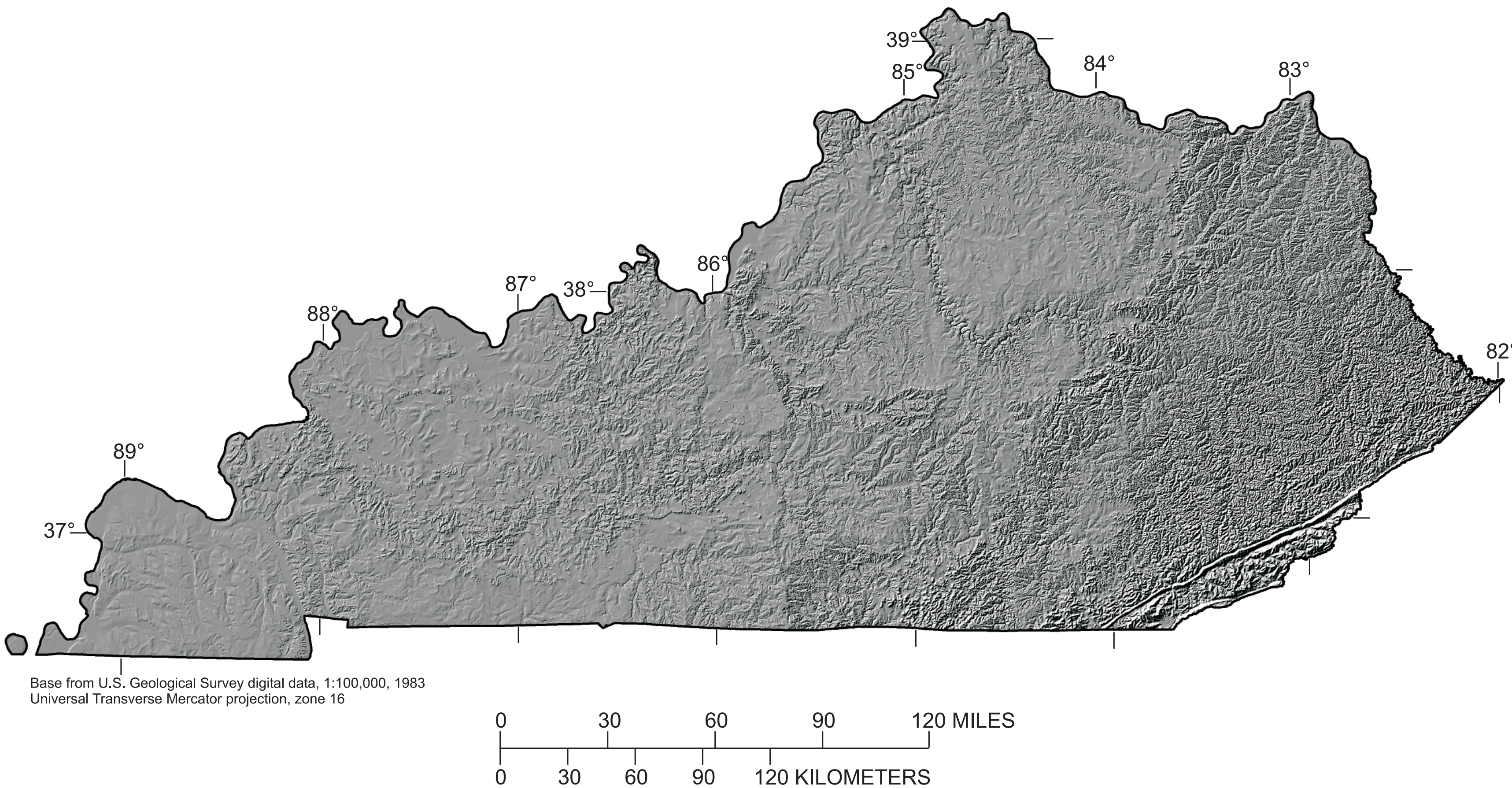
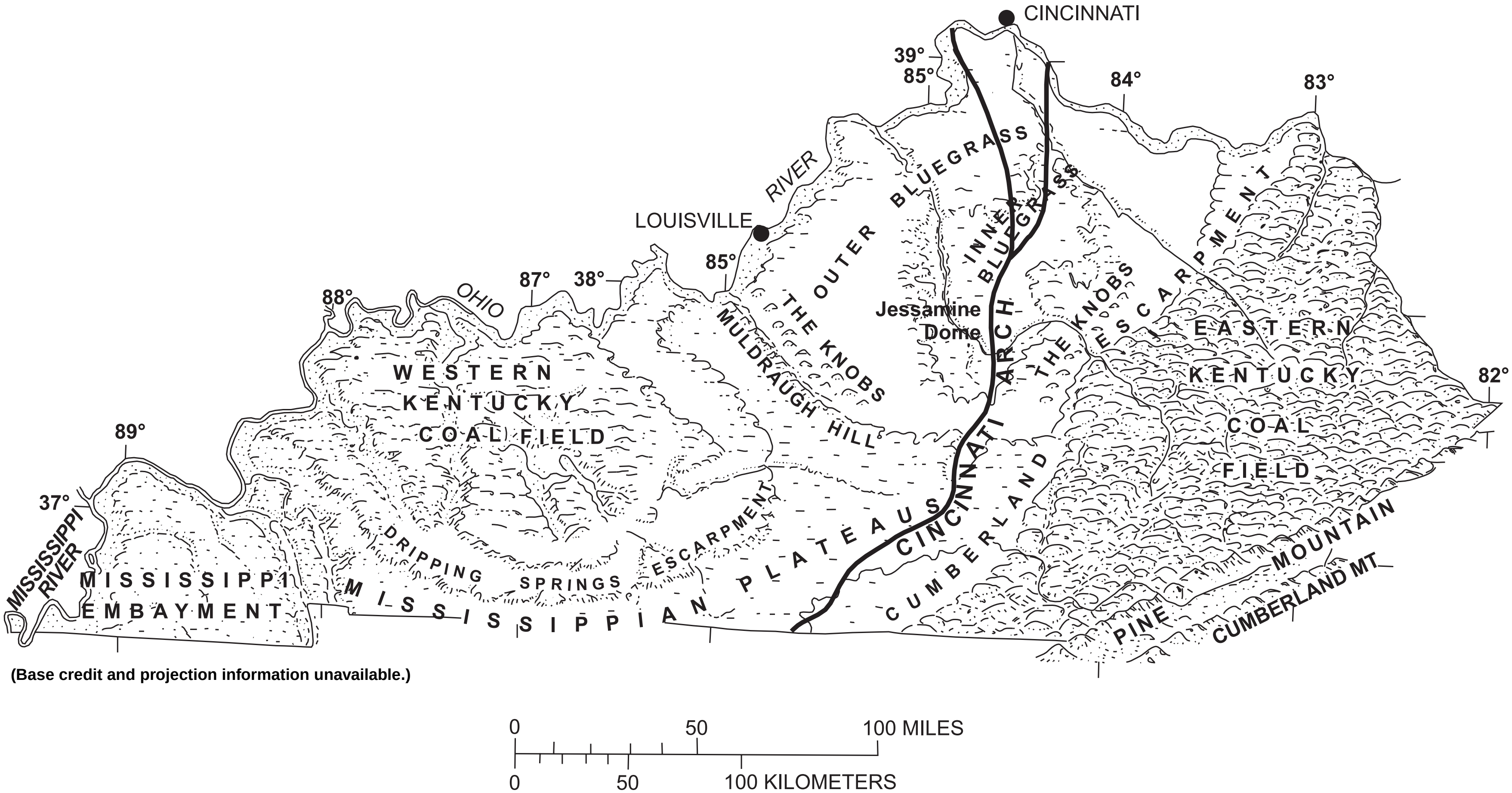
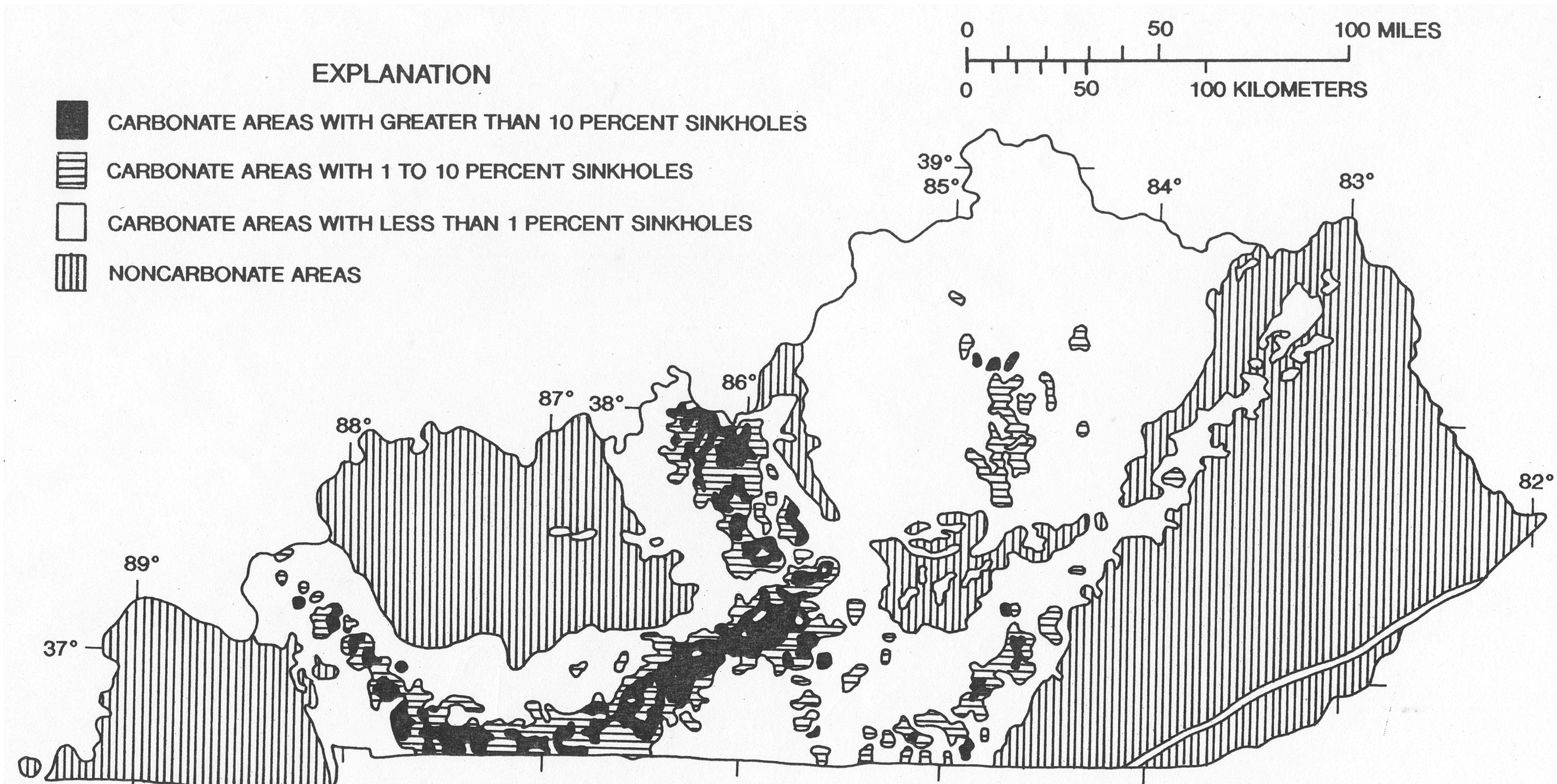


Figure 2. Shaded-relief image of landforms in Kentucky.



(Base credit and projection information unavailable.)

Figure 3. Physiographic regions in Kentucky [from Kentucky Geological Survey, 1980].



(Base credit and projection information unavailable.)

Figure 4. Generalized carbonate areas and surficial karst development in Kentucky [from Crawford and Webster, 1986].

Basins less than 10 mi^2 show a more uniform distribution of peaks throughout the year, with a particularly high percentage of floods occurring from June through September (about 28 percent) in comparison to the basins larger than 50 mi^2, where only about 10 percent of the annual floods occurred during these months. A similar pattern of summer flooding also occurred in the 10- to 50-mi^2 basins where about 20 percent of the peaks occurred from June through September. The annual peaks in small drainages (generally less than 10 mi^2) are more frequently caused by convective summer storms, which generally are of more limited areal extent, shorter duration, and higher intensity than the frontal storms in winter and spring that frequently cause the annual peaks in large basins (generally greater than 50 mi^2) (Choquette, 1988).

DATA USED FOR PEAK-FLOW ESTIMATES AND ESTIMATING TECHNIQUES

The USGS has been collecting and publishing continuous-record streamflow data for gaging stations in Kentucky since 1907 (Beaber, 1970). The data currently (2003) are published by the USGS in the annual report series titled “Water Resources Data—Kentucky.” For the section of this report titled “Estimates of Peak Flows at USGS Streamflow-Gaging Stations” (page 19), peak flows are reported for 222 Kentucky stations with 10 or more years of annual peak-flow data that are considered representative of current peak-flow conditions (table 1, page 33). For the section of this report titled “Estimating Peak Flows for Ungaged, Unregulated Streams in Rural Drainage Basins” (page 25), the data for 238 streamflow-gaging stations—187 in Kentucky, 7 in West Virginia, 8 in Virginia, 6 in Ohio, 13 in Indiana, 6 in Illinois, and 11 in Tennessee—with at least 10 years of rural, unregulated, annual peak flows were used (table 2, page 61). These data include the pre-regulation period from various gaging stations where flows currently (2003) are regulated. For both sections of this report, any sites with flow diversions or sites likely to be urbanized were not used. Flow diversions are documented in the annual report series titled “Water Resources Data—Kentucky.” Urban drainage basins in Jefferson County were documented in Martin and others (1997). USGS field personnel identified which drainage basins in the rest of Kentucky were likely currently to be urbanized.

The peak flows from various gaging stations were not used for various reasons. The data at two stations were combined into one station if the drainage area for a station was less than 10-percent different from the drainage area of another station and if doing so appeared reasonable based on the data. A drainage-area correction was applied when combining the stations if the drainage areas differed from 3 to 10 percent. Drainage-area corrections were not applied to stations for which the drainage areas differed by less than 3 percent. Data from the following stations were combined: Cumberland River near Pineville, Ky. (USGS gaging-station number 03403000), was combined with Cumberland River at Pine Street Bridge at Pineville, Ky. (03402900).

The peak flows for selected recurrence intervals for Salt River at Glensboro (03295400, 172 mi^2, 11 years of record) are not reported because the annual peak flows at this station appear to have been collected during an unrepresentative short period as compared to Salt River near Van Buren (03295500, 196 mi^2, 44 years of record).

Regression equations are used to estimate a response variable (in this case, a peak flow for a given recurrence interval) for an ungaged drainage basin by measuring explanatory variables (such as drainage area). Explanatory variables should make hydrologic sense, explain a large amount of the variability of the response variable, and be reasonably easy to measure. A set of explanatory variables that were qualitatively judged to best meet these criteria was selected for testing.

For the section of this report titled “Estimating Peak Flows for Ungaged, Unregulated Streams in Rural Drainage Basins” (page 25), the values of 27 explanatory variables were determined for gaged, unregulated streams in rural drainage basins in Kentucky and surrounding States. These 27 explanatory variables were:

total drainage area (TDA), in mi^2, the area measured in a horizontal plane that is enclosed by a drainage divide, measured by planimeter, digitized, or measured by grid method from USGS 7.5-minute topographic quadrangle maps;

contributing drainage area, in mi^2, is the total drainage area excluding any parts characterized by internal drainage, such as by way of sinkholes in karstic terrain;

main-channel length, in mi, the length measured along the main stream channel from the gage to the basin divide (by extension of the mapped main channel up to the divide), following the longest tributary as determined from USGS 7.5-minute topographic quadrangle maps;

main-channel slope (S), in ft/mi, computed as the difference in elevation between points located at 10 and 85 percent of the main-channel length from the gage, divided by the stream length between these two points, as determined from USGS 7.5-minute topographic quadrangle maps;

basin length, in mi, the straight-line distance from the gage to the basin divide (defined by the main-channel length);

mean basin width, in mi, calculated by dividing the total drainage area by basin length;

basin-shape factor, the ratio of basin length, in mi, squared to total drainage area, in mi^2;

main-channel sinuosity, the ratio of main-channel length, in mi, to basin length, in mi;

mean basin elevation, in thousands of ft, computed in ARC/INFO as the average elevation of the basin from a 1:250,000-scale digital elevation model (where elevations are referenced to the National Geodetic Vertical Datum of 1929, NGVD of 1929);

average basin elevation index, in thousands of ft, determined by averaging main-channel elevations at points 10 and 85 percent of the distance from a specified location on the main channel to the topographic divide, as determined from USGS 7.5-minute topographic quadrangle maps (where elevations are referenced to the National Geodetic Vertical Datum of 1929, NGVD of 1929);

storage area, in percent, plus 1.00 percent, that part of the contributing drainage area occupied by lakes, ponds, and swamps, as shown on USGS 7.5-minute topographic quadrangle maps, not including temporary storage as a result of detention basins or ponding at roadway embankments;

mean annual precipitation, in inches, minus 30 in., estimated from Kentucky Department for Natural Resources and Environmental Protection (1979) and Conner (1982);

maximum 24-hour precipitation frequencies, in inches, with recurrence intervals of 25 and 50 years (Hershfield, 1961);

maximum 24-hour precipitation, in inches, occurring during the 30-year interval of 1951-80 (Glenn Conner, Kentucky Climate Center, written commun., 1986);

soils index, in inches ("S"; U.S. Department of Agriculture, 1969), a measure of potential infiltration based on basin vegetative cover, soil infiltration rate, and soil-water storage;

soil infiltration index, in in/h, based on minimum infiltration rates for the U.S. Natural Resources Conservation Service (formerly Soil Conservation Service) hydrologic soil groups (Musgrave, 1955) for soil series in Kentucky (U.S. Department of Agriculture, 1975 and 1984);

forested area, as a percentage of the contributing drainage area, plus 1.00 percent, measured from USGS 7.5-minute topographic quadrangle maps by use of the transparent-grid sampling method;

streamflow-recession index, defined as the number of days it takes base streamflow to decrease one log cycle, or one order of magnitude, as determined graphically from hydrograph plots of daily mean streamflow during representative periods of streamflow recession (Riggs, 1964; Bingham, 1982, Ruhl and Martin, 1991);

streamflow-variability index, (Lane and Lei, 1950) at a station ("station" value) is computed as the standard deviation of the logarithms of the 19 discharges at 5-percent class intervals from 5 to 95 percent on the flow-duration (cumulative-frequency) curve (Searcy, 1959; Dempster, 1990) of daily mean streamflow for the entire period of record;

azimuth, measured in degrees from north to the line defining basin length;

gaging-station latitude, in decimal degrees, minus 36.0°, determined from USGS 7.5-minute topographic quadrangle maps;

gaging-station longitude, in decimal degrees, minus 81.0°, determined from USGS 7.5-minute topographic quadrangle maps.

drainage-basin centroid latitude, in decimal degrees minus 36.0°, determined in a geographic information system (GIS) by means of the “centroidlabels” command as applied to the basin-boundary polygons in ARC/INFO.

drainage-basin centroid longitude, in decimal degrees, minus 81.0°, determined in a GIS as described for centroid latitude.

climate factor for the 2-, 25-, and 100-year recurrence intervals, an index integrating the effects of climate on flood frequency as interpolated from climate factor isolines presented by Lichty and Karlinger (1990).

regional indicator variables X1, X2, ... X7, which were set to a value of 1, if a site was in the selected region, or 0 if the site was not in the selected region.

region, a single regional indicator variable, which was set to integer values 1, 2, ...7 depending on the particular region in which the gaging station was located.

DEVELOPMENT OF PEAK-FLOW ESTIMATES AND ESTIMATING TECHNIQUES

Peak-flow estimates for selected recurrence intervals at gaging stations were developed based on the guidelines of the Interagency Advisory Committee on Water Data (1982) (Bulletin 17B). Peak-flow regression equations for ungaged locations were developed by use of ordinary-least-squares (OLS) and generalized-least-squares (GLS) regression techniques. The peak flows at the gaging stations then were weighted with regression-equation peak-flow estimates at the gaging stations.

Peak Flows at Gaging Stations

The 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year peak flows for individual streamflow-gaging stations discussed in this section were calculated by use of the guidelines of the Interagency Advisory Committee on Water Data (1982) (Bulletin 17B). The calculations involved fitting the Pearson Type III probability distribution to the logarithms (base 10) of the observed annual peak flows at a gaging station. This fitting required computation of the mean, standard deviation, and skew of the logarithms of the annual peak-flow data. The peak flow for any selected recurrence interval was determined from the fitted curve.

Detailed Bulletin 17B Analyses

Bulletin 17B analyses require that the peak-flow data used for statistical analysis at a gaging station be a reliable and representative sample of random, homogeneous events. The annual peak flows at gaging stations described in this report are assumed to be random, reliable, and independent of each other.

The peak flows in a drainage basin will not be homogeneous if the hydrologic conditions in the basin change appreciably over time because of urbanization or other human activities. A two-sided Mann-Kendall trend test (Helsel and Hirsch, 1992) was done on the annual peak flows at gaging stations to test for trends over time. To produce accurate results for the significance of a trend, this test requires that the data have no serial correlation. Serial correlation, in this case, is the dependence or correlation in time sequence between annual peak flows. Annual peak-flow data can exhibit some serial correlation. This correlation can cause the Mann-Kendall trend test to indicate a significant trend when there is none, especially at gaging stations with less than 30 years of peak-flow data (G.D. Tasker, U.S. Geological Survey, written commun., 1997). For this reason, some judgment is necessary to determine whether the results of the Mann-Kendall trend test are significant. The Mann-Kendall test was not done at stations with less than 25 years of peak-flow data because trends cannot be distinguished from serial correlation at stations with this data length. Ten (7 percent) of 142 gaging stations tested had significant trends (at a significance level of 5 percent) over time, with 6 positive trends and 4 negative trends. These trends all had significance levels ranging from 1 to 5 percent. One-hundred two of these sites were located in Kentucky and 40 sites were located in surrounding States. The number of stations with significant trends are close to the number expected simply by chance and are distributed rather uniformly between positive and negative trends. The significant trends are believed to be chance occurrences rather than true trends; no sites were removed from the analyses.

The annual peak flows at all stations were plotted to look for large changes in the distribution of peak flows over time, especially at gaging stations whose basins now are regulated. A station was considered significantly regulated if its drainage basin had more than 4.5 million ft^3 of usable reservoir storage per mi^2 (Benson, 1962) or if pre-regulation peaks were significantly different from post-regulation peaks. The pre- and post-regulation annual peaks from all gaging stations downstream of USCOE dams were tested for significant differences. Significance was established with the Mann-Whitney test (also known as the Wilcoxon rank-sum test) (Helsel and Hirsch, 1992) at a one-sided significance level of 0.05. The results of the regulation analyses are listed in table 1 (page 33). The Cumberland River at Pine Street Bridge at Pineville, Ky. (USGS gaging-station number 03402900), had a *p*-value of 0.075. Given this *p*-value and the fact that the data at both upstream and downstream gaging stations indicated significant regulation, this gaging station also was considered regulated. The section of this report titled “Estimating Peak Flows for Ungaged, Unregulated Streams in Rural Drainage Basins” (page 25), used annual peak flows from pre-regulation time periods only. When peak flows were computed by use of techniques described in Bulletin 17B, only post-regulation data from stations were used (see the section of this report titled “Estimates of Peak Flows at USGS Streamflow-Gaging Stations,” page 19), because the pre-regulation data are no longer representative of current (2003) flows.

Bulletin 17B guidelines were followed for the treatment of high and low outliers, for the conditional probability adjustment, for the adjustment for historical information, and for weighting the station skew coefficient with a generalized skew coefficient. The station skew was not weighted with the generalized skew if the annual peak flows at a gaging station were significantly affected by regulation. The annual peak flows from the gaging stations in this study did not show obvious evidence of being caused by multiple generating mechanisms; therefore, the procedures used to handle this situation were not used. Expected probability adjustments were not made; these adjustments are explained in Bulletin 17B.

Generalized Skew for Kentucky

Four methods were analyzed to find the most accurate generalized skew for Kentucky to use in the Bulletin 17B analyses. The first method was to compute an arithmetic mean of the station skews. To compute this skew coefficient, the station skews from 102 gaging stations in Kentucky were computed by use of the procedures in Bulletin 17B. The stations used all had at least 25 years of unregulated annual peak-flow data. None of these stations were significantly affected by diversions or urbanization. The computed station skews for this method, and the following methods, were adjusted for bias (Tasker and Stedinger, 1986). The 102 stations had an average of 40.4 years of annual peak-flow data.

In the second method, mean skews were calculated for stations that drain karst basins and stations that drain other basins. For this method, the karst stations were defined as stations with total drainage areas that were different from their contributing drainage areas. By this criteria, there were 82 non-karst drainage basins and 20 karst drainage basins. The two samples were different from each other at a significance level of 0.094 when the Mann-Whitney test was applied. The weak *p*-value, combined with mean skews that are similar (0.16 for the karst sites and -0.02 for the non-karst sites), lead to the decision not to separate these two populations of gaging stations in Kentucky.

In the third method of computing a generalized skew, an attempt was made to create a State skew-isoline map by plotting the station skews on a map at the centroid of their drainage basins.

The stations, however, showed no obvious geographic pattern. Positive and negative skew values both were scattered throughout the State.

In the fourth method, an attempt was made to develop a multiple-regression equation with station skew as the response variable and drainage basin characteristics (such as drainage area and stream slope) as the explanatory variables. There were 95 of the 102 gaging stations with the following basin characteristics available for testing as explanatory variables: contributing drainage area, main-channel slope, mean basin elevation, average basin elevation index, forested area, mean annual precipitation, maximum 24-hour precipitation frequencies with recurrence intervals of 25 and 50 years, basin length, and azimuth. All-possible-subsets multiple OLS regression and the Mallow's Cp statistic were used to find the best combinations of variables (Helsel and Hirsch, 1992). Some combinations were eliminated from consideration if individual explanatory variables were not significant. Combinations also were eliminated if they contained both measures of precipitation intensity and each measure had opposite signs. The best remaining regression equation contained the single variable—percent forested area. This regression explained too little of the variation in skew values ($r^2 = 0.076$) to use, given the risk that the form of the regression equation likely is to be imperfect (Helsel and Hirsch, 1992).

It is obvious that the Bulletin 17B generalized skew (the national skew-isoline map) is not representative of station skews in Kentucky. The Bulletin 17B map isolines indicate negative skews for all of Kentucky except the far eastern part. As discussed earlier, both positive and negative computed station skews for 102 sites in Kentucky were scattered throughout the State.

The mean skew for Kentucky was used as the generalized skew in the Bulletin 17B flood-frequency analyses. This new skew coefficient is 0.011, with a standard error of prediction of 0.520.

Peak Flows at Ungaged Locations

Regression equations are used to compute peak flows at ungaged locations. OLS regression techniques (Helsel and Hirsch, 1992) were used in an exploratory data analysis to develop linear regression equations to relate peak flows, Q_T ("response" or "dependent" variable, where T = 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence intervals), to selected basin characteristics ("independent" or "explanatory" variables). The most appropriate explanatory variables for estimating peak flows were selected in this OLS exploratory phase. The final regression coefficients and regression errors then were computed by use of GLS regression. GLS regression compensates for differences in the reliability of, and correlation among, the Q_T estimates at stations included in the analysis. The regression analysis included the 238 streamflow-gaging stations in Kentucky and nearby in neighboring States with at least 10 years of peak-flow record that was not affected appreciably by regulation (table 2–page 61, and plate 1 at back of report).

Inspection of scatter plots in the OLS exploratory phase, which showed the relations between response and explanatory variables, indicated that logarithmic (base 10) transformations of the response variable and most of the explanatory variables were appropriate. This transformation generally helped make the relations more linear and the residuals (errors) more uniform in variance about the regression line than before transformation. The relations between response and explanatory variables after transformation were consistent with the assumed linear form of the model.

The general form of the regression equations developed in this study is

$$\log(Q_T) = b_o + b_1\log X_1 + b_2\log X_2 + \ldots + b_n\log X_n + \varepsilon \,, \qquad (1)$$

where

Q_T the response variable, is the peak flow of estimated long-term average recurrence interval T,

b_o is a constant,

b_i (i=1 to n) is the regression coefficient for the ith explanatory variable,

X_i (i=1 to n) is the ith explanatory variable,
ε is the random-error component, and
n is the total number of explanatory variables.

The algebraically equivalent form when the equation is retransformed to the original units is

$$Q_T = 10^{bo} X_1^{\ b1} X_2^{\ b2} \ldots X_n^{\ bn} . \qquad (2)$$

Defining Flood Regions

During the exploratory data analysis, the seven peak-flow hydrologic regions defined by Choquette (1988) were assessed based on the updated peak-flow values. The peak-flow characteristics of the seven regions were evaluated graphically and statistically. A statewide OLS regression of log-transformed 50-year peak-flow values with the log-transformed total drainage areas ($log\ Q_{50} = b_0 + b_1\ log\ TDA$) was completed using all 238 gaging stations. Also, this same OLS regression was completed for each of three aggregated regions formed by combining Regions 1 and 2; Regions 3, 4, and 5; and Regions 6 and 7 (plate 1). The geographical distribution of the residuals (errors) from the statewide regression and the residuals from the three regressions for the aggregated regions generally conformed to the regional boundaries defined by Choquette (1988). Residuals from the statewide regression also were grouped by river basin for comparison; however, no distinctive pattern in residuals relative to river basins was apparent. This lack of pattern may be related to the large geologic and physiographic variability and climatic variability spanned by Kentucky drainage basins. Plots of the residuals in downstream order by river basin showed a tendency for the residual signs and magnitudes to vary locally with changes in geologic and physiographic characteristics.

Density plots and box plots of the residuals from the statewide regression (not shown) displayed patterns similar to those described by Choquette (1988) where residuals for Regions 1, 3, 5, and 7 appear similar in sign and magnitude and the residuals for Regions 2, 4, and 6 appear similar in sign and magnitude. A Kruskal-Wallis test indicated that for at least one region, the central tendency of the residuals differ among the seven regions (at an attained significance level (p-value) of less than 0.001). Residuals from the statewide regression were compared to residuals from each region by use of the nonparametric Mann-Whitney U test. The statewide residuals were significantly different ($p<0.05$) from the residuals of all regions except for Region 6, where the test-statistic p-value was 0.076.

Regional location-indicator variables were used in the OLS regressions to test for statistically significant differences among the seven regions. In these statistical tests, a 95-percent confidence level was defined as significant. The location-indicator variables were set either at 1, if the station was in a particular region, or 0, if not. Methods described by Montgomery and Peck (1982) were used to test for significant variations of the slopes (coefficients) and intercepts among regional regressions of the 50-year peak flow with total drainage area. Intercepts for the OLS-regression equations for Regions 2, 4, and 6 differ from those of Regions 1, 3, 5, and 7, and there is variation among slopes of the seven regional regression equations.

Regional location-indicator variables also were used to directly compare regressions for each region to regressions for the group of all the other stations combined (Pope and others, 2001). This comparison was done by adding the location-indicator variable and the product of the location-indicator variable and the total-drainage-area term to the regression. Thus, a three-variable OLS-regression equation to estimate Q_{50}, with all available stations and utilizing (1) total drainage area, (2) location-indicator variable, and (3) the product of the location-indicator variable and total drainage area as explanatory variables, was developed for each of the seven regions. In each regional model, a significant location variable indicates a difference in the regression intercept between the stations in that region and the stations in the rest of the State; a difference in the product of the location variable and total drainage area term indicates a difference in the coefficients for total drainage area between the stations in that region and the stations in the rest of the State. Based on the levels of significance of the indicator-variable terms in these regressions, only the Region 3 regression equation failed to indicate a significant difference (0.05) from the regression for the group of all other stations combined. However, further location-indicator regression comparisons showed that

Region 3 differs from Region 4, Region 3 differs from Region 2, and Region 3 also differs from the group of stations in the neighboring regions combined (Regions 2, 4, and 5).

Results of these statistical tests, which cannot be used to statistically verify the seven regions, nonetheless are supportive of the regionalization scheme. The tests, which compare the regional regression characteristics, indicate that each region represents a grouping of stations that is distinguishable from either all the other stations combined as a group or all the stations in the neighboring regions combined. Indeed, the central tendencies (means and medians) of the residuals of the statewide regression in each region are not statistically different from the central tendencies of each and every one of the other regions. As noted previously, the residuals in Regions 1, 3, 5, and 7 appear similar in sign and magnitude, and the residuals in Regions 2, 4, and 6 appear similar in sign and magnitude. However, the odd-numbered regions are geographically separated as are the even-numbered regions. Also, the regression equations for the individual regions (table 3) differ in their coefficients and in their accuracy. Based on the results of the graphical and statistical tests described in this section, the seven-region scheme developed by Choquette (1988) was accepted for the updated peak-flow values presented in this report.

Some minor adjustments in the region boundaries were made to improve alignment with the drainage-basin boundaries and the current residuals: An area of approximately 150 mi^2, which included the basin for the Goose Creek at Manchester station (03281100), was shifted from Region 4 into Region 3. An area of approximately 220 mi^2, which included the basin for the Barren River Tributary near Bowling Green station (03314750), was shifted from Region 6 into Region 5. An area of approximately 50 mi^2 in the lower Green River Basin, just downstream from the confluence with Pond River, was shifted from Region 7 into Region 6.

Choosing Explanatory Variables

OLS-regression equations were developed by all-possible-subsets regression procedures (Statistical Analysis System Institute, Inc., 1985) with the 2-, 50-, and 500-year peak flows for the 238 unregulated stations as response variables, initially utilizing all of the 27 prospective explanatory variables, or transformations thereof, discussed in the section of this report titled “Data Used for Peak-Flow Estimates and Estimating Techniques” (page 8). A subset of 54 of the stations had all of these explanatory variables available, because most of the explanatory variables at these stations were computed in a previous study (Choquette, 1988). The variables out of the 27 prospective variables that least improved the regression model, using data from the 54 stations, were dropped from the analyses. Also, variables were removed if they were highly correlated with a better-performing variable. As the number of explanatory variables was reduced, an increased number of stations had all of the variables. Regression equations were rerun with additional stations and fewer variables than before, and again the variables that least improved the equations were dropped. As more gaging stations were added, regression analyses also were done for individual regions. This iterative process continued as the best explanatory variables were retained, based on both the statewide and regional equations. Total drainage area, instead of contributing drainage area, was selected as the primary explanatory variable. Contributing drainage area was eliminated as an explanatory variable because of (1) the difficulty in determining this basin characteristic accurately from maps that generally are available, (2) the minimal overall improvement in accuracy in most regions, and (3) a reduced accuracy in Region 5. The top four explanatory variables were total drainage area, main-channel slope, main-channel sinuosity, and basin-shape factor. Based on the results from these regressions, it was not considered useful to compute values of the other explanatory variables that had not been determined previously.

Table 3. Regression equations and their accuracy for estimating peak flows for ungaged, unregulated streams in rural drainage basins in Kentucky

[*Q* is peak flow, in cubic feet per second; *TDA* is total drainage area, in square miles; *S* is main-channel slope, in feet per mile]

Peak-flow regression equation for given recurrence interval (recurrence intervals from 2 to 500 years)	Average standard error of prediction (percent)	$(PRESS/n)^{1/2}$ (percent)	Average equivalent years of record	Estimated model-error variance (base-10 logs)	Average sampling-error variance (base-10 logs)
Region 1 – 28 gaging stations					
$Q_2 = 312\ TDA^{0.673}$	49.7 to -33.2	49.9 to -33.3	1.3	0.0277	0.0030
$Q_5 = 493\ TDA^{0.651}$	48.0 to -32.4	48.6 to -32.7	1.9	.0259	.0031
$Q_{10} = 91.5\ TDA^{0.843}\ S^{0.451}$	46.2 to -31.6	47.2 to -32.1	2.8	.0230	.0042
$Q_{25} = 81.2\ TDA^{0.872}\ S^{0.535}$	49.3 to -33.0	51.1 to -33.8	3.6	.0253	.0050
$Q_{50} = 75.8\ TDA^{0.890}\ S^{0.587}$	52.9 to -34.6	55.5 to -35.7	3.9	.0283	.0057
$Q_{100} = 71.4\ TDA^{0.907}\ S^{0.632}$	57.3 to -36.4	61.0 to -37.9	4.1	.0321	.0066
$Q_{200} = 67.8\ TDA^{0.922}\ S^{0.673}$	62.4 to -38.4	67.3 to -40.2	4.2	.0367	.0076
$Q_{500} = 63.6\ TDA^{0.941}\ S^{0.722}$	69.7 to -41.1	76.6 to -43.4	4.3	.0438	.0090
Region 2 – 68 gaging stations					
$Q_2 = 152\ TDA^{0.728}$	44.0 to -30.6	45.1 to -31.1	1.9	.0238	.0013
$Q_5 = 239\ TDA^{0.721}$	39.5 to -28.3	40.8 to -29.0	3.0	.0197	.0012
$Q_{10} = 304\ TDA^{0.715}$	38.6 to -27.9	40.5 to -28.8	4.2	.0188	.0013
$Q_{25} = 393\ TDA^{0.709}$	38.7 to -27.9	41.5 to -29.3	5.8	.0187	.0015
$Q_{50} = 464\ TDA^{0.704}$	39.3 to -28.2	43.0 to -30.1	6.9	.0191	.0016
$Q_{100} = 538\ TDA^{0.699}$	40.4 to -28.8	44.8 to -30.9	7.9	.0199	.0018
$Q_{200} = 615\ TDA^{0.695}$	41.7 to -29.4	47.0 to -32.0	8.7	.0209	.0020
$Q_{500} = 721\ TDA^{0.690}$	43.6 to -30.4	50.2 to -33.4	9.6	.0225	.0022
Region 3 – 24 gaging stations					
$Q_2 = 187\ TDA^{0.748}$	25.9 to -20.6	29.0 to -22.5	5.9	.0081	.0019
$Q_5 = 355\ TDA^{0.712}$	23.7 to -19.1	27.5 to -21.5	9.8	.0064	.0021
$Q_{10} = 498\ TDA^{0.692}$	22.2 to -18.2	27.2 to -21.4	15.1	.0052	.0024
$Q_{25} = 714\ TDA^{0.670}$	20.7 to -17.2	27.5 to -21.5	24.6	.0040	.0027
$Q_{50} = 897\ TDA^{0.656}$	20.4 to -16.9	28.1 to -22.0	32.4	.0034	.0031
$Q_{100} = 1100\ TDA^{0.643}$	20.4 to -16.9	29.1 to -22.5	39.4	.0031	.0034
$Q_{200} = 1320\ TDA^{0.632}$	21.1 to -17.4	30.4 to -23.3	44.5	.0030	.0039
$Q_{500} = 1640\ TDA^{0.620}$	22.6 to -18.4	32.7 to -24.6	47.9	.0033	.0045
Region 4 – 17 gaging stations					
$Q_2 = 39.0\ TDA^{0.923}\ S^{0.204}$	31.4 to -23.9	32.3 to -24.4	3.5	.0112	.0029
$Q_5 = 69.8\ TDA^{0.894}\ S^{0.186}$	25.2 to -20.1	26.0 to -20.7	6.5	.0072	.0023
$Q_{10} = 92.7\ TDA^{0.882}\ S^{0.178}$	24.3 to -19.5	25.6 to -20.4	9.2	.0065	.0024
$Q_{25} = 121\ TDA^{0.873}\ S^{0.173}$	25.5 to -20.3	27.6 to -21.6	11.6	.0068	.0029

Table 3. Regression equations and their accuracy for estimating peak flows for ungaged, unregulated streams in rural drainage basins in Kentucky—*Continued*

[*Q* is peak flow, in cubic feet per second; *TDA* is total drainage area, in square miles; *S* is main-channel slope, in feet per mile]

Peak-flow regression equation for given recurrence interval (recurrence intervals from 2 to 500 years)	Average standard error of prediction (percent)	$(PRESS/n)^{1/2}$ (percent)	Average equivalent years of record	Estimated model-error variance (base-10 logs)	Average sampling-error variance (base-10 logs)
Region 4 – 17 gaging stations—*continued*					
$Q_{50} = 140\ TDA^{0.870}\ S^{0.173}$	27.3 to -21.5	30.2 to -23.2	12.5	0.0077	0.0033
$Q_{100} = 392\ TDA^{0.780}$	31.6 to -24.0	38.4 to -27.7	11.4	.0111	.0031
$Q_{200} = 441\ TDA^{0.778}$	33.8 to -25.3	41.3 to -29.2	11.7	.0125	.0035
$Q_{500} = 510\ TDA^{0.776}$	37.1 to -27.1	45.5 to -31.3	11.9	.0147	.0041
Region 5 – 40 gaging stations					
$Q_{2} = 260\ TDA^{0.704}$	34.5 to -25.7	36.3 to -26.6	3.3	.0151	.0015
$Q_{5} = 437\ TDA^{0.692}$	38.9 to -28.0	41.8 to -29.5	3.8	.0186	.0018
$Q_{10} = 571\ TDA^{0.686}$	43.9 to -30.5	48.0 to -32.4	4.2	.0228	.0022
$Q_{25} = 754\ TDA^{0.682}$	51.1 to -33.8	57.0 to -36.3	4.6	.0293	.0028
$Q_{50} = 901\ TDA^{0.679}$	56.7 to -36.2	64.0 to -39.0	4.7	.0347	.0033
$Q_{100} = 1060\ TDA^{0.677}$	62.4 to -38.4	71.3 to -41.6	4.8	.0405	.0039
$Q_{200} = 1220\ TDA^{0.676}$	68.4 to -40.6	78.7 to -44.0	4.9	.0467	.0045
$Q_{500} = 1450\ TDA^{0.674}$	76.4 to -43.3	89.0 to -47.1	4.9	.0555	.0053
Region 6 – 27 gaging stations					
$Q_{2} = 256\ TDA^{0.600}$	49.9 to -33.3	49.9 to -33.3	1.3	.0280	.0029
$Q_{5} = 397\ TDA^{0.586}$	45.1 to -31.1	46.4 to -31.7	2.0	.0234	.0027
$Q_{10} = 499\ TDA^{0.578}$	43.9 to -30.5	46.5 to -31.7	2.8	.0222	.0028
$Q_{25} = 636\ TDA^{0.569}$	44.1 to -30.6	48.4 to -32.6	3.8	.0221	.0031
$Q_{50} = 740\ TDA^{0.564}$	45.4 to -31.2	50.9 to -33.7	4.4	.0229	.0035
$Q_{100} = 846\ TDA^{0.559}$	47.1 to -32.0	53.9 to -35.0	4.9	.0243	.0038
$Q_{200} = 953\ TDA^{0.555}$	49.6 to -33.2	57.6 to -36.5	5.2	.0263	.0043
$Q_{500} = 1100\ TDA^{0.551}$	53.7 to -34.9	63.1 to -38.7	5.5	.0299	.0049
Region 7 – 34 gaging stations					
$Q_{2} = 293\ TDA^{0.623}$	56.6 to -36.1	58.2 to -36.8	1.4	.0350	.0029
$Q_{5} = 476\ TDA^{0.616}$	55.0 to -35.5	57.5 to -36.5	2.1	.0332	.0030
$Q_{10} = 614\ TDA^{0.613}$	54.5 to -35.3	58.1 to -36.8	3.0	.0325	.0032
$Q_{25} = 804\ TDA^{0.610}$	54.7 to -35.4	59.9 to -37.4	4.1	.0323	.0036
$Q_{50} = 956\ TDA^{0.610}$	55.3 to -35.6	61.6 to -38.1	5.0	.0326	.0039
$Q_{100} = 1110\ TDA^{0.609}$	56.1 to -35.9	63.6 to -38.9	5.8	.0332	.0042
$Q_{200} = 1280\ TDA^{0.610}$	57.3 to -36.4	66.0 to -39.7	6.6	.0342	.0045
$Q_{500} = 1510\ TDA^{0.610}$	59.2 to -37.2	69.5 to -41.0	7.5	.0358	.0050

Various factors were considered in evaluating alternative regression equations with the top four explanatory variables, including (1) the coefficient of determination, a measure of the proportion of the variation in the response variable explained by the regression equation; (2) the standard error of the estimate, a measure of model-fitting error; (3) the prediction sum of squares (PRESS) statistic, a measure of model-prediction error; (4) the statistical significance of each individual explanatory variable; (5) potential multicollinearity as indicated by the correlation of explanatory variables and the value of the variance inflation factor (Montgomery and Peck, 1982); (6) the effort and modeling benefit of determining the values of each additional explanatory variable; and (7) the hydrologic validity of the signs and magnitudes of the explanatory variables.

The best one-, two-, and three-variable regression equations in the final OLS all-possible-subsets regression run were determined for all available gaging stations in each region with the top four explanatory variables. Total drainage area was selected as an explanatory variable in all seven regions. The attained significance levels for the T-statistics on total drainage area were less than 0.0001 in all regions. Main-channel slope was chosen as a second explanatory variable in Regions 1 and 4. Addition of any of the three explanatory variables other than drainage area did not appreciably (more than 3 percentage points) reduce the regression standard errors of estimate in the other five regions. The values of drainage area and main-channel slope for all regions are listed in table 2 (page 61).

Regression diagnostic tools were used to test the adequacy of the final OLS regressions. The OLS-regression coefficients all are statistically different from zero (*p*-values less than 0.05). The influence of individual stations on the regressions was measured by Cook's D statistic (Helsel and Hirsch, 1992). Multicollinearity in the explanatory variables was tested with the variance inflation factor. There were no problems with high-influence points or multicollinearity between variables. Different types of residual plots were analyzed. The regression residuals were plotted against predicted values to look for linearity, homoscedasicity, normality, and the presence of outliers. Normal probability plots of the residuals also were analyzed. Residuals were plotted against the explanatory variables to look for biases in the explanatory variables over their range. All regression diagnostics indicated that the final explanatory variables in each region resulted in satisfactory regression equations.

Determining Final Regression Coefficients

The regression equations (table 3) were finalized by use of GLS-regression techniques (Stedinger and Tasker, 1985; Tasker and Stedinger, 1989), utilizing the computer program GLSNET (G.D. Tasker, K.M. Flynn, A.M. Lumb, and W.O. Thomas, U.S. Geological Survey, written commun., 1995). Two major assumptions of OLS regression commonly are violated in regression of peak flows and explanatory variables: (1) the errors in the computed peak flows of selected recurrence intervals are the same at all gaging stations and (2) the annual flows for overlapping years at different gaging stations are independent of each other (not cross-correlated). Error in the peak flows varies with the length of record, which differs among the gaging stations, and streamflows at gaging stations in a region usually are cross-correlated because the same climatic conditions and weather events can affect most of the streams within a hydrologic region.

Stedinger and Tasker (1985) and Tasker and Stedinger (1986) have shown that where streamflow records are of widely varying length and concurrent flows at different sites are highly correlated, GLS regression provides more accurate estimates of the regression coefficients, better estimates of the accuracy of the regression equations, and better estimates of the model error when compared to OLS regression. GLS regression gives more weight to long-term than short-term gaging stations and less weight to stations where flows are more highly correlated to flows at other gaging stations.