USGS
science for a changing world

In cooperation with the Kentucky Transportation Cabinet—Department of Highways

Estimating Mean Annual Streamflow of Rural Streams in Kentucky

Water-Resources Investigations Report 02-4206

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

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

Estimating Mean Annual Streamflow of Rural Streams in Kentucky

By **Gary R. Martin**

Water-Resources Investigations Report 02-4206

In cooperation with the Kentucky Transportation Cabinet—Department of Highways

Louisville, Kentucky
2002

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

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

For additional information write to:

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

Copies of this report can be purchased from:

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

CONTENTS

FIGURES

TABLES

CONVERSION FACTORS AND VERTICAL DATUM

CONVERSION FACTORS

Multiply	By	To obtain
inch (in.)	25.4	millimeter
foot (ft)	0.3048	meter
mile (mi)	1.609	kilometer
square mile (mi^2)	2.590	square kilometer
cubic foot per second (ft^3/s)	0.02832	cubic meter per second

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

$$°C = (°F - 32) / 1.8$$

VERTICAL DATUM

Sea level: In this report “sea level” refers 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 both the United States and Canada, formerly called Sea Level Datum of 1929.

Estimating Mean Annual Streamflow of Rural Streams in Kentucky

By Gary R. Martin

Abstract

Mean annual streamflow (Q_a), defined as the mean of the series of annual mean streamflow values, was determined for selected rural stream sites in Kentucky. Streamflow data for the available period of record through the 1999 water year (October 1, 1998–September 30, 1999) at 235 continuous-record streamflow-gaging stations with at least 5 years of record located in and adjacent to Kentucky were used in the analysis. Record-extension procedures were applied for selected gaging stations to reduce time-sampling error and, thus, improve estimates of the long-term Q_a.

Techniques to estimate the Q_a at ungaged stream sites in Kentucky were developed. One-, two-, and three-variable regression equations that included total drainage area, station latitude minus 36 degrees, and mean basin elevation as explanatory variables were developed by use of ordinary- and generalized-least-squares regression. The three-variable regression equation has an approximate average standard error of prediction of 13.7 percent. The one- and two-variable equations exhibit geographical biases, and the indicated standard errors of prediction may poorly estimate the true prediction errors, depending upon the location in the State. Therefore, the three-variable equation should be used for estimating mean annual streamflow of rural streams in Kentucky whenever possible.

INTRODUCTION

The U.S. Geological Survey (USGS) has collected continuous-record streamflow-gaging data in Kentucky since 1907 (Beaber, 1970); other agencies collected such data in Kentucky as early as 1890. Statistical characteristics of the streamflow data, such as the mean annual streamflow, are needed by water-resource managers and engineers for design of hydraulic structures constructed in the riverine environment.

Resource limitations make it unfeasible for the collection of data on every stream and at every stream site where streamflow characteristics may be needed; therefore, techniques for estimating the needed streamflow characteristics at ungaged stream sites have been developed. The USGS, in cooperation with the Kentucky Transportation Cabinet—Department of Highways, compiled the available continuous-record streamflow-gaging station data, computed the Q_a for these gaged stream sites, and developed regional equations for estimating Q_a at ungaged rural stream sites on the basis of selected drainage-basin characteristics.

Purpose and Scope

The purpose of this report is to provide (1) Q_a values at continuous-record streamflow-gaging stations having 5 or more years of record through the 1999 water year and (2) procedures for estimating the Q_a at rural ungaged stream sites where flows are not appreciably affected by local diversions. This report presents Q_a values for 235 continuous-record streamflow-gaging stations in the study area. Procedures for estimating the Q_a at ungaged stream sites are described and illustrated with example computations.

Previous Studies

Mean annual streamflows in Kentucky have been investigated in various previous studies. Beaber (1970) analyzed streamflow-data needs and applications in Kentucky. Statistical multiple-regression analyses were done to define relations between selected streamflow characteristics and drainage-basin characteristics. Various regression models to estimate mean annual streamflows at ungaged stream sites statewide were developed, including

$$Q_a = 1.19A^{1.02} \quad (1)$$

$$Q_a = 1.22A^{1.01}E^{0.10} \quad (2)$$

$$Q_a = 0.290A^{1.01}E^{0.25}I^{1.27} \quad (3)$$

$$Q_a = 0.270A^{1.01}E^{0.23}I^{1.36}St^{-0.14} \quad (4)$$

where,

Q_a is the mean annual streamflow, in ft^3/s;

A is the drainage area in mi^2;

E is the mean elevation of the basin, in thousands of feet above sea level;

I is the maximum 24-hour, 2-year rainfall intensity, in inches; and

St is the area of lakes and ponds in percent of drainage area (plus 1).

The regression analyses provided estimates of the accuracy of the relations, which improved as the number of explanatory (regressor) variables and model complexity increased. The average standard error of estimate ranged from 14.8 percent for equation 1 to 12.1 percent for equation 4.

Melcher and Ruhl (1984) computed mean annual streamflows for the available period of record through the 1982 water year. Results presented were for combined regulated and unregulated periods of record at those sites with regulated flows.

Wetzel and Bettandorff (1986) developed regression models for estimating streamflow characteristics, including Q_a, based on data for 629 streamflow-gaging stations in coal provinces covering parts of 11 States in the eastern United States, including the Eastern and Western Kentucky Coal Field physiographic regions. The basin characteristics used as explanatory variables in these regression models for estimation of Q_a included drainage area, mean annual precipitation, and mean basin elevation. The average standard error of estimate ranged from 35.7 percent for the drainage-area-only model to 17.1 percent for the model that included all three of these basin characteristics as explanatory variables. The model coefficients of determination (R^2) in log space ranged from 0.96 to 0.99.

Acknowledgments

The author would like to thank the Kentucky Transportation Cabinet for its support of this work. The author also wishes to thank the many local, State, and Federal agencies that have cooperated in the operation and maintenance of streamflow-gaging stations in Kentucky and surrounding States that were used as part of this study. The author wishes to express appreciation to the many other USGS employees who assisted with collection and analysis of streamflow data, measurement of basin characteristics, and preparation of this report.

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 and Mississippi Rivers (fig. 1). In a generalized water balance for Kentucky, approximately 60 percent of precipitation leaves drainage basins as evapotranspiration, and approximately 40 percent of precipitation leaves as direct runoff or shallow ground-water flow into the streams and rivers. Variations in climate, physiography, and geology cause localized variations in streamflow characteristics in Kentucky.

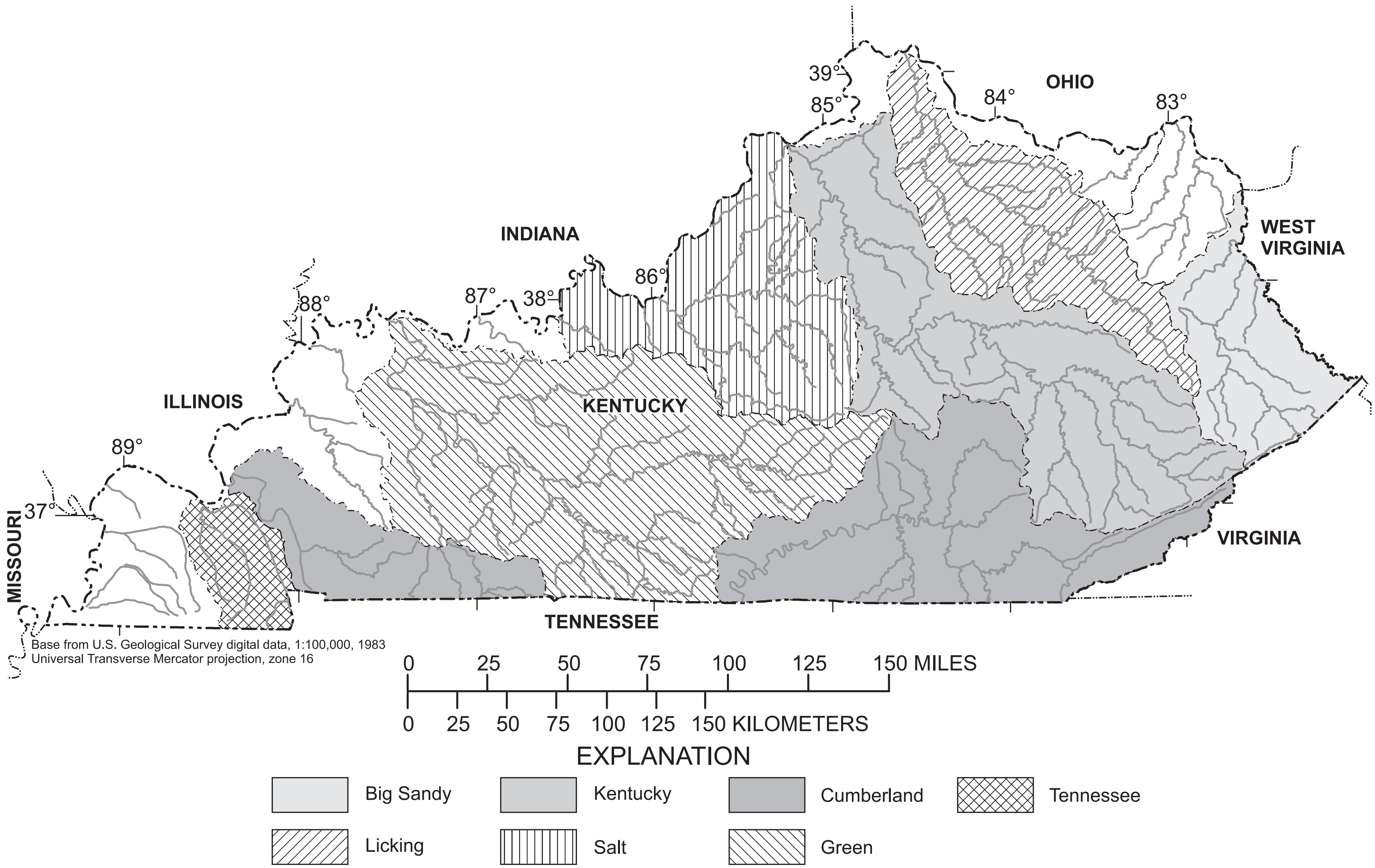


Figure 1. Major drainage basins in Kentucky.

Climate

Kentucky has a moist-continental climate with distinct seasonal variations and changeable weather patterns with generally short periods of extreme conditions. Winter temperatures are moderate, rarely below 0°F; typical summer temperatures are warm and rarely above 100°F. The weather patterns are affected variably by the meeting of cold-, arctic-, and continental-air masses arriving from the northwest and warm, moist-air masses moving up the Mississippi and Ohio River Valleys from the southwest.

Annual precipitation in Kentucky averages about 47 in. (Conner, 1982). The distribution of precipitation varies areally, year to year, 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 precipitation, which primarily is the subtropical Atlantic Ocean and Gulf of Mexico. Considerable year-to-year variation in precipitation results in Kentucky. During the period 1951-80, annual precipitation at reporting stations ranged from 14.5 to 78.6 in. (Conner, 1982). Large amounts of precipitation in Kentucky have been associated with tropical cyclones moving north from the Gulf of Mexico.

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 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 in fig. 4). Less well-developed karst features are in central and south-central Kentucky. River main channels in karst areas commonly have sustained base flow during dry-weather periods.

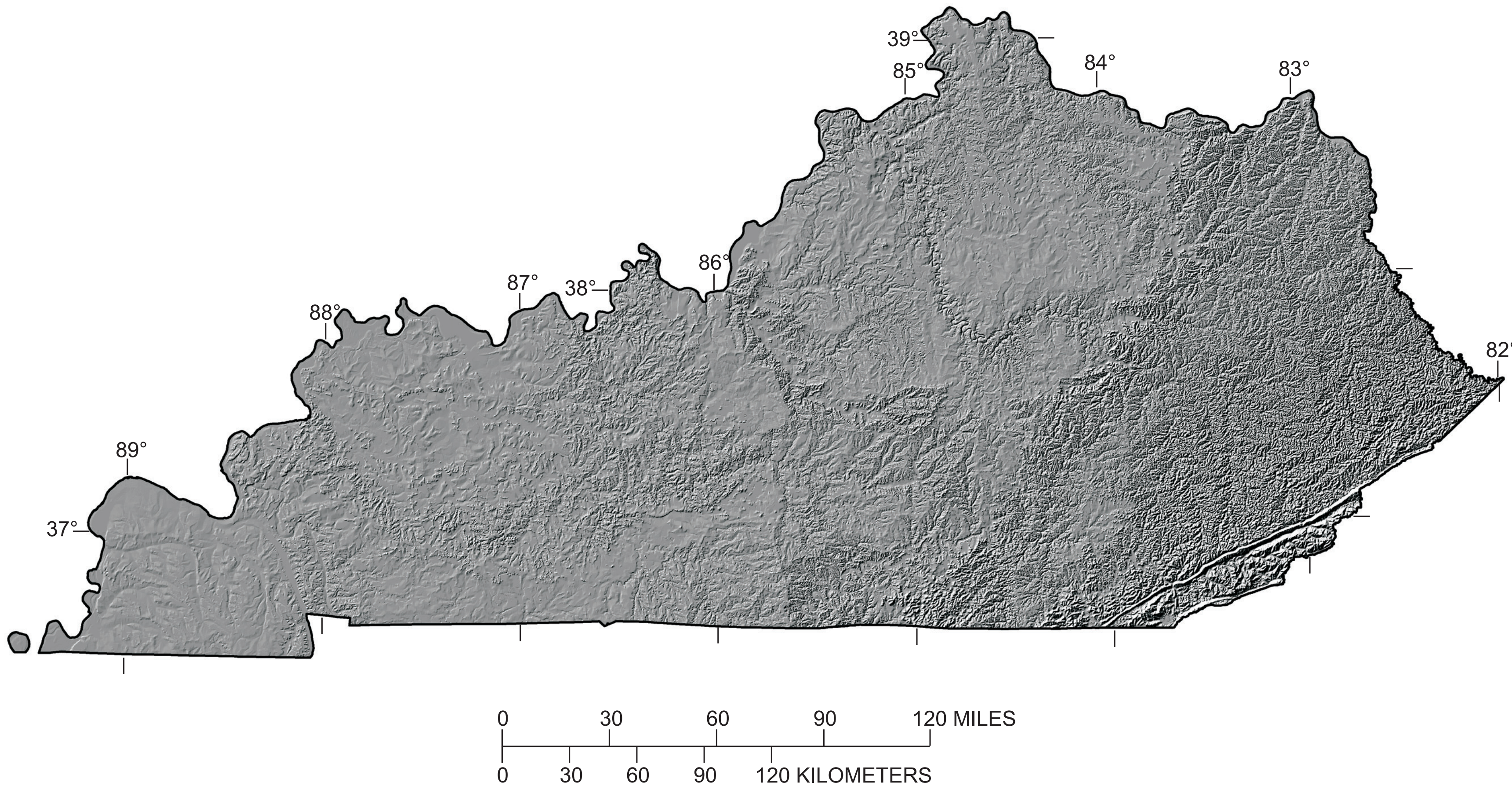


Figure 2. Shaded-relief image of landforms in Kentucky.

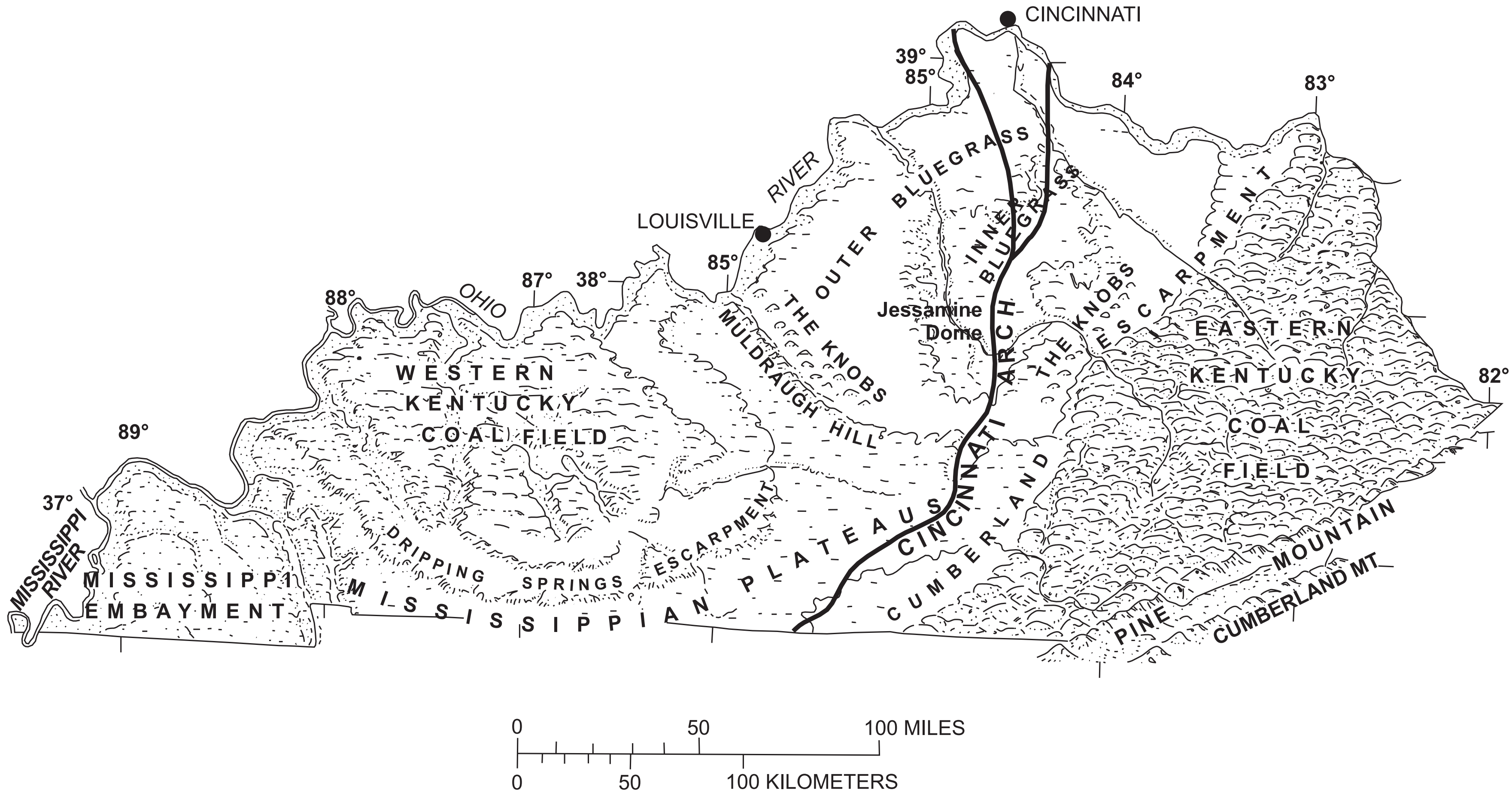


Figure 3. Physiographic regions in Kentucky [from Kentucky Geological Survey, 1980].

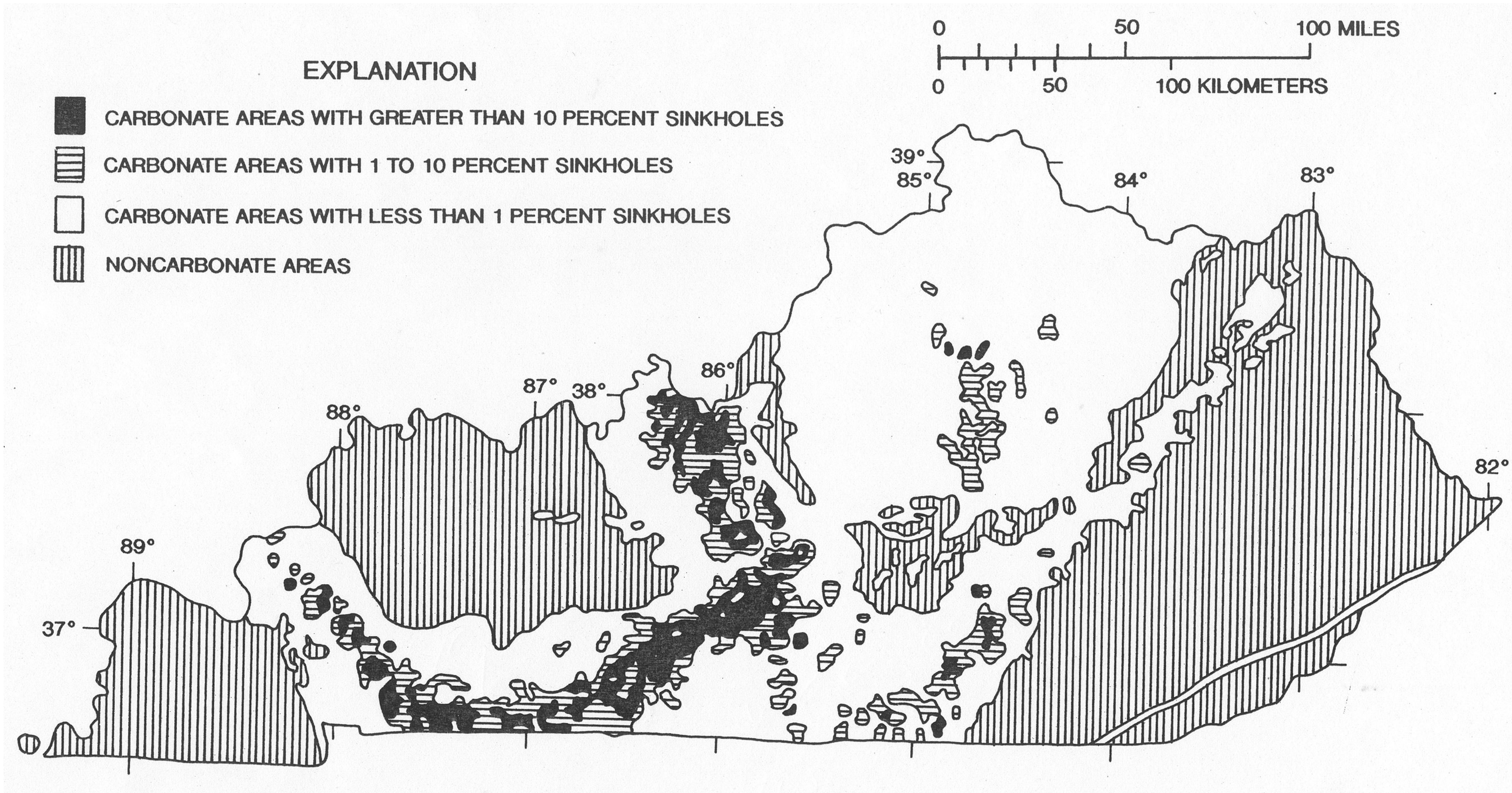


Figure 4. Generalized carbonate areas and surficial karst development in Kentucky [from Crawford and Webster, 1986].

COMPILATION AND REVIEW OF STREAMFLOW DATA

Daily mean streamflow data for 235 continuous-record streamflow-gaging stations located in Kentucky and adjacent states were retrieved by use of the USGS Automated Data Processing System (Bartholoma, 1997). The streamflow data in surrounding States were retrieved to provide additional information for use in the regionalization of Q_a values. The data were checked and verified by comparing computed yearly and monthly summary statistics of the daily mean streamflows to published values (U.S. Geological Survey, 1958a, 1958b, 1964a, 1964b, 1962-65, 1966-75, and 1976-2000).

Annual mean streamflows at 235 stations (198 in Kentucky and 37 in surrounding States) were tested for trends ($p <= 0.10$) by use of the Kendall's Tau test, which indicated that 27 stations had an increasing trend in streamflow and 3 had a decreasing trend in streamflow. These trends appeared climate-related and consistent with reported trends for mid-range (median) flows in this region (Lins and Slack,1999). Precipitation-adjusted annual mean streamflows were approximated roughly as the residuals from regressions of annual mean streamflow with annual precipitation data (by water year; see Glossary for definition) at Louisville, Ky., which is centrally located in relation to the gaging stations in the data set. Kendall's Tau tests of these residuals indicated that 28 had trend: 19 with increasing trend and 9 with decreasing trend.

Daily mean streamflows at many gaging stations in Kentucky are affected by regulation and (or) local diversions. Regulation by multipurpose or flood-control reservoirs reduce peak flows and generally augment low flows on a seasonal time scale; however, when streamflows are averaged over an annual time step, as in the case of the annual mean streamflow statistic, regulation generally has no effect on this statistic when the volume of water stored is released during the same water year. Two-sample statistical comparisons (Mathsoft, Inc., 1999a and 1999b) of pre- and post-regulation, precipitation-adjusted annual mean streamflow values (the residuals of the regression of annual mean streamflow with annual precipitation) downstream from some of the major reservoirs in Kentucky failed to indicate a significant difference in the sample means ($p = 0.05$).

Local diversions—localized transfers of water such as water-supply withdrawals or wastewater discharges—artificially decrease or increase streamflows within a reach. Local diversions are common near municipalities and in urban areas. The extent of alterations in natural streamflows caused by local diversions was reviewed based on available water-withdrawal and permitted-wastewater-discharge data. (A.C. Downs, U.S. Geological Survey, written commun., 2002; and S. Bolssen, Kentucky Natural Resources and Environmental Protection Cabinet, Division of Water, written commun., 2002). Localized diversions were deemed minor in relation to Q_a where available data indicated the flows diverted annually probably would not exceed 10 percent of Q_a.

MEAN ANNUAL STREAMFLOW ESTIMATES FOR GAGED STREAM SITES

Annual mean streamflows were computed from daily mean streamflows by use of the National Water Information System program DVMAS (Daily Values Monthly and Annual Statistics) (Bartholoma, 1997). The mean annual streamflow (Q_a) is defined as

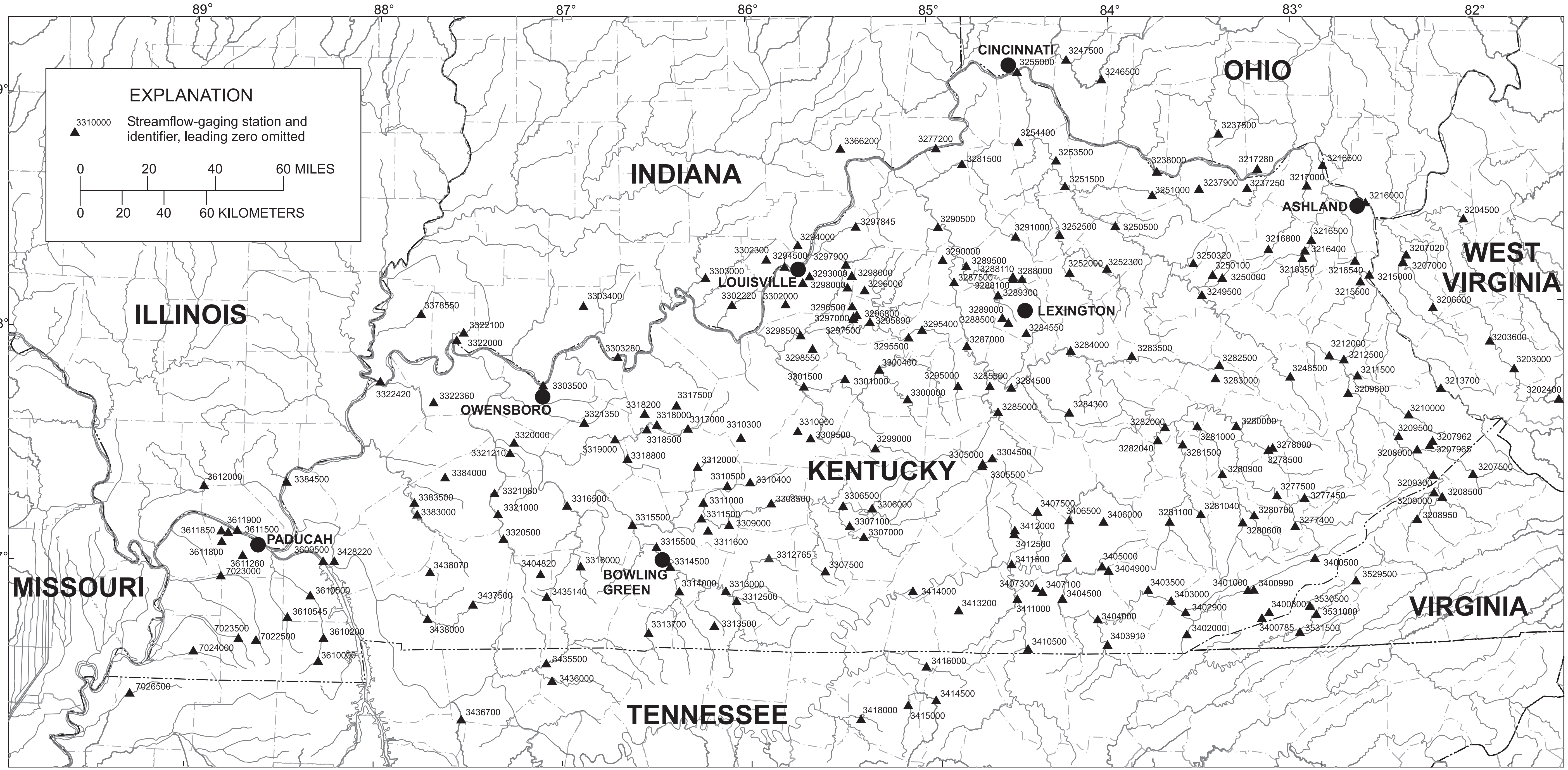
$$Q_a = \left(\sum_{i=1}^{N_a} Q_{ai} \right) / N_a , \qquad (5)$$

where

Q_{ai} is annual mean streamflow for the ith year and

N_a is the number of annual mean streamflows in the gaging station period of record.

The computed values of Q_a at streamflow-gaging stations having at least 5 years of record are shown in table 1 (back of report). Locations of these gaging stations are shown in figure 5.



Base from U.S. Geological Survey digital data, 1:100,000, 1983
Universal Transverse Mercator projection, Zone 16

Figure 5. Locations of continuous-record streamflow-gaging stations in Kentucky and surrounding States for which data are presented in this report. (See table 1.)

Streamflow statistics are subject to error associated with the particular time period sampled (time-sampling error). Gaged record may occur during either an abnormally wet or dry period, thus making it unrepresentative of long-term average climatic conditions. Time-sampling error decreases as record length increases. An example of the effects of time-sampling error is reflected in the 100 ft^3/s difference (reduction) of the computed Q_a value at Barren River at Lock 1 at Greencastle (station 0331500) compared to the computed Q_a value upstream at Barren River at Bowling Green (station 03314500) as shown in table 1.

Record-extension (augmentation) techniques may be used to reduce time-sampling error in streamflow values and statistics. Record extension is achieved by relating concurrent streamflows (and streamflow statistics) at a short-term and a nearby long-term (index) station that is hydrologically similar. The Q_a at the index station and the relation between the concurrent daily mean streamflows at both stations may be used to provide an estimate of the long-term Q_a at the short-term station. A mathematical record-extension technique, Maintenance of Variance Extension Type 1 (MOVE.1) as described by Hirsch (1982), was used in this study. The estimate was computed by use of log-transformed values of the concurrent nonzero daily mean streamflows as

$$Log\ Q_{a(s)} = M_s + \left(\frac{S_s}{S_l}\right) x\ (Log\ Q_{a(l)} - M_l)\,, \quad (6)$$

where

$Q_{a(s)}$ is the estimated long-term Q_a for the short-term station;

$Q_{a(l)}$ is the Q_a for the long-term station;

M_s, M_l are the mean of the log-transformed daily mean streamflows for the concurrent period at the short- and long-term stations, respectively; and

S_s, S_l are the standard deviations of the log-transformed daily mean streamflows for the concurrent period at the short- and long-term stations, respectively.

MOVE.1 was applied to improve Q_a estimates for Ohio River main-stem stations only (table 1). The two long-term index stations on the Ohio River at Louisville, Ky. (03294500) and at Metropolis, Ill. (03611500) each had 71 full water years of record (1929-99). Stations designated “short-term” in equation 6 had fewer than 71 years of record. The adjustments reduced time-sampling errors and improved consistency in Q_a and drainage-area-standardized values of Q_a along the river.

DEVELOPMENT OF THE TECHNIQUE FOR ESTIMATING MEAN ANNUAL STREAMFLOW FOR UNGAGED, RURAL STREAM SITES

A regression was used to develop regional equations for estimating Q_a at ungaged, rural sites. Drainage-basin characteristics, including climate, affect streamflow patterns. Relations among selected basin characteristics and computed Q_a were investigated by methods of linear correlation and multiple-linear regression.

Basin Characteristics

Various drainage-basin characteristics were tested for applicability in the regionalization of Q_a. Selection of basin characteristics for inclusion in exploratory scatter plots, linear correlation analysis, and subsequent multiple-linear-regression analysis was based on (1) the possible hydrologic importance of the characteristic in relation to the Q_a statistic, (2) the availability of previously determined basin characteristics for the study basins, and (3) results of previous regionalization studies of other streamflow statistics (Beaber, 1970; Wetzel and Bettandorff, 1986; Choquette, 1988; Ruhl and Martin, 1991; and Martin and Ruhl, 1993).

Basin characteristics tested for significance in the regression analysis included the following:

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

contributing drainage area, in square miles, is the total drainage area excluding any parts characterized by internal drainage, such as by way of sinkholes in karst terrain;
main-channel length, in miles, the length measured along the main stream channel from the station to the basin divide, following the longest tributary as determined from USGS 7.5-minute topographic quadrangle maps;
main-channel slope, in feet per mile, 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 miles, the straight-line distance from the streamflow-gaging station to the basin divide (defined by the main-channel length);
mean basin width, in miles, calculated by dividing the total drainage area by basin length;
basin shape, the ratio of basin length, in miles, squared to total drainage area, in square miles;
main-channel sinuosity, the ratio of main-channel length, in miles, to basin length, in miles;
mean basin elevation (E), in thousands of feet above sea level, computed as the average elevation of the basin from a 1:250,000-scale digital elevation model converted to a grid coverage in ARC/INFO;
average basin elevation index, in thousands of feet above sea level, 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;
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 intensity, in inches, with recurrence intervals of 2 and 10 years (Hershfield, 1961);
maximum 24-hour precipitation intensity, 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), is a measure of potential infiltration based on basin vegetative cover, soil infiltration rate, and soil water storage;
soil infiltration index, in inches per hour, is based on minimum infiltration rates for the U.S. 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; and 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; and Dempster, 1990) of daily mean streamflow for the entire period of record; "mapped" values of variability index tested in the regression were computed as areally weighted average values from the regionalized variability index (Ruhl and Martin, 1991);

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

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

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

drainage-basin centroid latitude, in decimal degrees minus 36.0°, determined in geographical information system (GIS) by means of the "centrallabels" command as applied to the basin-boundary polygons in ARC/INFO; and

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

Regression Analysis

A multiple-linear-regression model was developed to relate Q_a (dependent variable) to selected basin characteristics ("independent" or explanatory variables). Included in the regression analysis were 170 streamflow-gaging stations with at least 10 years of record where Q_a was deemed not significantly affected by local diversions (identified as "minor" local diversions in table 1). The regression analysis included an exploratory phase using ordinary-least-squares (OLS) regression to select appropriate explanatory variables and a final phase using generalized-least-squares (GLS) regression. GLS regression compensates for differences in the variability and reliability of, and correlation among, the Q_a estimates at stations included in the analysis.

Inspection of scatter plots showing relations among dependent and explanatory variables and plots of residuals from initial linear regressions indicated that logarithmic (base 10) transformation of the dependent and most of the explanatory variables would be appropriate. This transformation generally helped make the relations more linear and the residuals more uniform in variance about the regression line than before transformation. The relations between dependent and explanatory variables after transformation were consistent with the assumed linear form of the model.

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

$$\log(Q_a) = b_o + b_1 \log X_1 + b_2 \log X_2 + \ldots + b_n \log X_n + \varepsilon, \quad (7)$$

where

Q_a is mean annual streamflow,

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 a random error component, and

n is the total number of explanatory variables.

The algebraically equivalent form when the log (base 10) transformation is used and when the equation is re-transformed to the original units is

$$Q_a = 10^{bo} X_1^{b1} X_2^{b2} \ldots X_n^{bn}. \quad (8)$$

The alternative OLS regression models were generated by all-possible-regression and stepwise-regression procedures (Statistical Analysis System Institute, Inc., 1985) using the prospective explanatory variables listed in "Basin Characteristics." Various factors were considered in evaluating alternative regression models, including (1) the coefficient of determination, 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 alternative 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 regression exponents.

The best one-, two-, and three-variable OLS regression models included total drainage area (A), latitude of the gaging station minus 36° (Lat_g-36), and mean basin elevation (E) in thousands of feet above sea level as explanatory variables. The drainage-area-only model accounts for a large part of the variability of Q_a; however, the drainage-area-only model exhibits geographical bias, which was reduced progressively by adding the second and third explanatory variables to the regression models. The locational variable, Lat_g, may serve to integrate and index statewide variations in precipitation and evapotranspiration. Mean basin elevation also may serve as an index to a combination of other factors that are difficult to evaluate, such as radiation, temperature, wind, vegetation, and basin relief, which can cause streamflow variations (Thomas and Benson, 1970).

The OLS regression coefficients all are statistically different from zero (p-value less than 0.01). Regression residuals were analyzed to (1) identify outliers and high-leverage stations for examination, (2) confirm normality and homogeneity of variance, and (3) identify and remedy geographic bias.

The regression models (table 2) were finalized by use of GLS regression techniques (Stedinger and Tasker, 1985; and Tasker and Stedinger, 1989), which were implemented in 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 streamflow statistics: (1) the errors of the streamflow statistic are homogeneous among the observations and (2) the observations statistically are independent. Error in streamflow statistics vary with the length of record, which differs among the gaging stations, and streamflows at the set of gaging stations are correlated because the same climatic conditions and weather events generally affect most of the streams within a hydrologic region.

Stedinger and Tasker (1985, 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 coefficients, and almost unbiased estimates of the model error when compared to OLS regression. GLS regression gives more weight to long-term gaging stations (with less time-sampling error) than short-term gaging stations and more weight to stations where flows are least correlated to flows at other gaging stations. GLS regression procedures use weighting matrices to proportionately account for the cross-correlation of streamflows and for the variations in time-sampling error of the streamflow statistic among the gaging stations.

Table 2. Equations (derived by generalized-least-squares regression) for estimating mean annual streamflow in Kentucky

[Q_a, mean annual streamflow in cubic feet per second; A, total drainage area, in square miles; Lat_g, latitude of the gage, or basin outfall, in decimal degrees; E, mean basin elevation, in thousands of feet above sea level; --, not applicable]

Equations	Range of explanatory variable			Approximate average standard error of prediction (percent)	Average equivalent years of record
	A	Lat_g	E		
$Q_a = 1.38\,A^{1.01}$	0.67–2,762	--	--	15.8	10.9
$Q_a = 1.42\,A^{1.01}(Lat_g\text{-}36)^{-0.18}$	.67–2,762	36.341–39.140	--	14.2	14.9
$Q_a = 1.39\,A^{1.00}(Lat_g\text{-}36)^{-0.15}E^{\,0.12}$	.67–2,762	36.341–39.140	0.391–2.414	13.7	15.4

The cross-correlations were estimated as an empirical, best-graphical-fit function of the distance between pairs of long-term gaging stations having at least 50 years of concurrent record. GLS regression required matrices of the mean, standard deviation, and skew of the annual mean streamflows associated with the matrix of the Q_a. A regional estimate of the standard deviations of annual mean streamflows independent of the Q_a estimating equation was developed within GLSNET by use of regression of the standard deviation against drainage area and mean basin elevation. A skew matrix was estimated by use of the observed skew of the series of annual mean streamflows at the gaging stations. The length of record at each gaging station was used as a measure of the reliability of the Q_a estimates. GLSNET enables partitioning of total regression-model error into model error and sampling error, which consists of both time- and space-sampling error. Model error arises from limitations of the model formulation, and it cannot be reduced by additional data collection. Time- and space-sampling error, however, are reduced through additional data collection by extending the period of data collection and by expanding the variety of basin characteristics of the sites where data are collected, respectively.

Limitations and Accuracy

The one-, two-, and three-variable regional regression models for estimating Q_a at ungaged stream sites have varying limitations and accuracies (table 2). As indicated previously, the one- and two-variable equations exhibit geographical biases, and the indicated standard errors of prediction may poorly estimate the true prediction errors, depending on the location in the State. The one- and two-variable models are suitable for initial approximate Q_a estimates; however, the three-variable equation should be used whenever possible.

The regional regression models are applicable to rural streams in Kentucky that are not appreciably affected by local diversions, which commonly are associated with urban development. Caution is warranted when applying the regression models in areas where streamflows are affected by hydrologic discontinuities such as large springs and sinks common to karst terrain in areas underlain by limestone (see fig. 4). Streamflows in these areas may vary unpredictably in karst drainageways. It may be difficult (if not impossible) to determine an accurate basin drainage area in karst terrain solely on the basis of topographic divides.

The regression model was developed by use of basin characteristics within a certain range of values. Drainage areas (A) of stations used in the regression analysis ranged from 0.67 to 2,762 mi^2, gage latitudes (Lat_g) ranged from 36.341 to 39.140°, and mean basin elevations (E) ranged from 0.391 to 2.414 in thousands of feet above sea level. Application of the regression models for Q_a estimates in basins outside these ranges is an extrapolation; therefore, these models probably should not be applied for this situation.

The standard error of prediction (of log Q_a) of the three-variable model—a measure of the accuracy of the regression estimates compared to observed data for stations excluded from the regression—is 13.7 percent. Standard error of prediction was estimated as the square root of the PRESS divided by the error degrees of freedom (Statistical Analysis System Institute, Inc., 1985; Montgomery and Peck, 1982; and Choquette, 1988). The procedure used for computing PRESS is considered a form of data splitting and can be applied as a model-validation tool. The accuracy of the three-variable model predictions for ungaged sites similar to those used in the regression could be expected to compare favorably to the standard error of prediction. A scatter plot of the values of Q_a computed from the streamflow-gaging station data and values computed using the regression model are shown in figure 6.

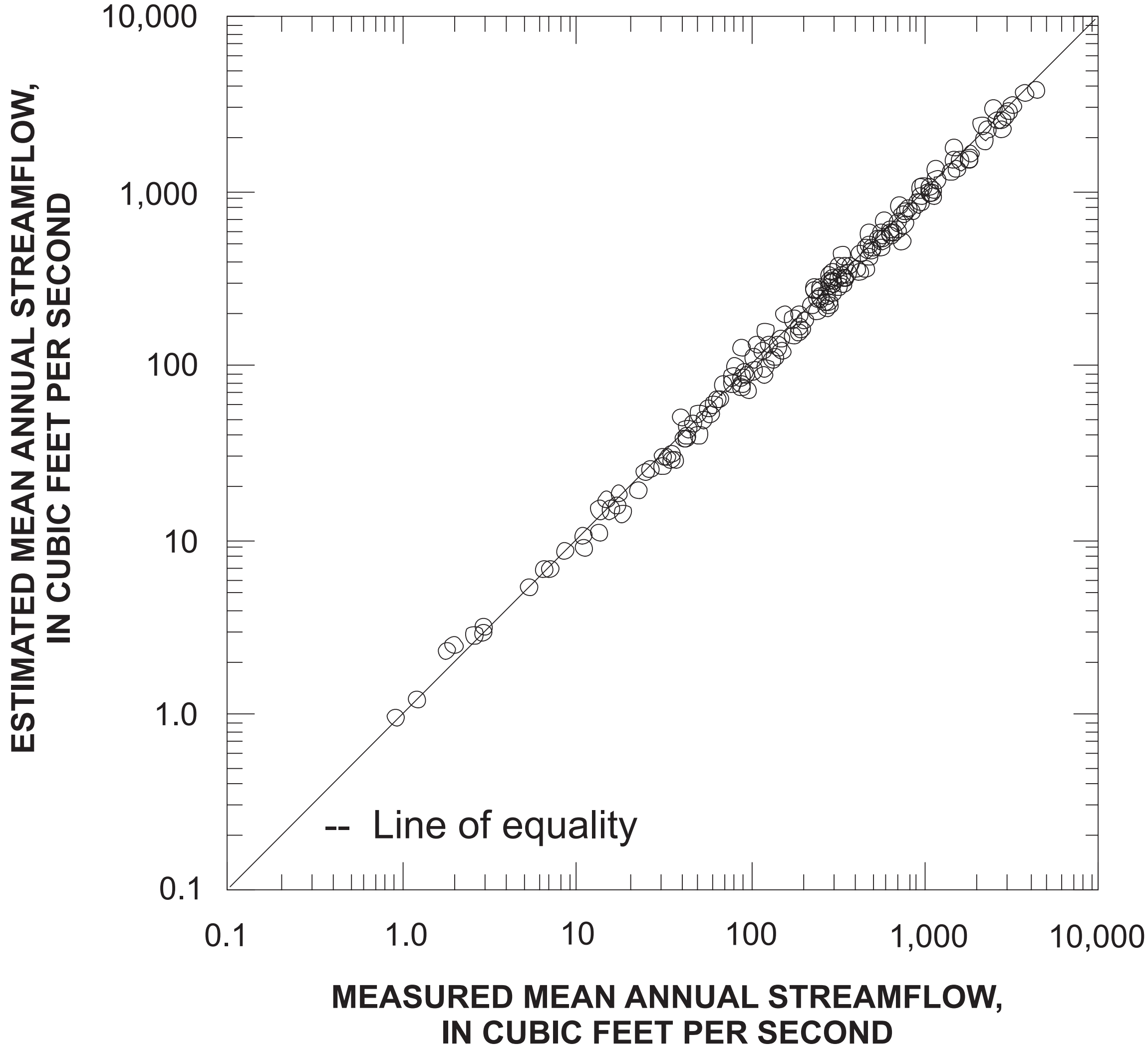


Figure 6. Comparison of measured mean annual streamflow and mean annual streamflow estimated by use of the three-variable regression equation for the 170 continuous-record streamflow-gaging stations in Kentucky and surrounding States used in the regression.

PROCEDURES FOR ESTIMATING MEAN ANNUAL STREAMFLOW AT STREAM SITES IN KENTUCKY

Procedures for obtaining Q_a estimates differ depending on the location of the stream site in relation to streamflow-gaging (gage) locations where Q_a has been determined. The appropriate procedures and examples are presented in the following sections.

Stream Sites With Gage Information

When streamflow-gaging information is available on the reach where an estimate of Q_a is desired, the gage information is used where appropriate in making the estimate, as discussed below.

Sites at Gage Locations

Estimates of Q_a values for 235 continuous-record streamflow-gaging stations are presented in table 1. When an estimate of Q_a is required at a stream site, refer to table 1 to determine whether values previously have been estimated for the site. At gage locations where the period of record is less than the equivalent years of record reported for the regression models (table 2), except on the Ohio River and where local diversions are significant, an improved estimate of Q_a may be obtained from a weighted average of the gaging-station estimate and regression-model estimate. The equivalent years of record can be used to weight the regression-model estimate, and the years of gaged record can be used to weight the gaging-station estimate.

Sites Near Gage Locations

If information is available for a stream reach where an estimate is desired, but not at the specific location, a weighting procedure can be used (Carpenter, 1983; and G.F. Koltun, U.S. Geological Survey, written commun., 2001). The drainage area of the ungaged site should differ by no more than 50 percent from that of the gaged site (ranging from 50 to 150 percent of the drainage area of the gaged site) to minimize the potential for hydrologic dissimilarity between the sites.

A weighted estimate of Q_a can be computed as

$$Q_{a_{uw}} = Q_{a_{ur}}\left[R - \left(\frac{2(|\Delta A|)(R-1)}{A_g}\right)\right], \qquad (9)$$

where R = $Q_{a_{gm}} / Q_{a_{gr}}$

and $Q_{a_{uw}}$ is the weighted mean annual flow, Q_a, for the ungaged site;

$Q_{a_{ur}}$ is the regression estimate of the Q_a for the ungaged site;

$Q_{a_{gm}}$ is the Q_a determined for the gaged site from measured streamflow data;

$Q_{a_{gr}}$ is the regression estimate of the Q_a for the gaged site;

$|\Delta A|$ is the absolute value of the difference between the drainage areas of the gaged site and the ungaged site; and

A_g is the drainage area of the gaged site.

As the difference in drainage area between the gaged and ungaged site approaches 50 percent, the value of the weighting factor in brackets in equation 9 approaches 1 and no longer has an effect on the regression estimate at the ungaged site.

Sites Between Gage Locations

If a Q_a estimate is desired between two gage locations on the same stream, the value can be estimated by linear interpolation by use of the Q_a values and corresponding drainage areas at the two gaged sites.

Stream Sites With No Gage Information

If no streamflow information is available at a stream site, or at a nearby stream site on the same stream reach so that the estimating methods in the previous section cannot be used, then the regional regression models (table 2) can be used directly to estimate Q_a.

Total drainage area of the site of interest should be determined from USGS 7.5-minute topographic maps or from other maps or GIS coverages of equivalent or improved accuracy. The drainage areas for many locations along streams in Kentucky are listed in Bower and Jackson (1981).