# REGIONALIZATION OF HARMONIC-MEAN STREAMFLOWS IN KENTUCKY

By Gary R. Martin and Kevin J. Ruhl

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 92-4173

Prepared in cooperation with the

KENTUCKY NATURAL RESOURCES AND ENVIRONMENTAL PROTECTION CABINET

Louisville, Kentucky

1993

The coefficient of multiple determination ($R^2$) for the regression model is 0.90. The standard error of estimate (of log $Q_h$)--a measure of the accuracy of the regression-model estimates compared to the observed values used in the regression--is 76 percent. The standard error of estimate was computed using the model root-mean-square error (Statistical Analysis System Institute, 1982) and information from Hardison (1971). The standard error of prediction (of log $Q_h$)--a measure of the accuracy of the regression estimates compared to observed data for stations excluded from the regression--is 78 percent, which is slightly higher than the standard error of estimate. Standard error of prediction was estimated as the square root of the PRESS divided by the error degrees of freedom (Statistical Analysis System Institute, 1982; Montgomery and Peck, 1982; 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 model predictions for ungaged sites similar to those used in the regression could be expected to compare favorably to the standard error of prediction. If all the assumptions for applying regression are met, two-thirds of the observations lie within one standard error of a regression line. For this regression, a 0.293 log units standard error, when untransformed, would place two-thirds of the observations within plus 96 percent and minus 49 percent of the regression line.

A scatter plot of the values of $Q_h$ computed from the streamflow-gaging station data and values computed using the regression model (fig. 9) shows reduced residuals and a slight tendency of the model to underpredict the values of $Q_h$ above about 50 $ft^3/s$. The underprediction tendency may be associated with increased error and bias in the values of mapped streamflow-variability index for large basins. The reduced residuals are probably related to generally reduced time-sampling error (long periods of record) for the stations having large values of $Q_h$. Also, less variability in the streamflow response would be expected for large basins as compared to small basins.

## ESTIMATING HARMONIC-MEAN STREAMFLOW AT STREAM SITES IN KENTUCKY

Procedures for obtaining $Q_h$ estimates differ depending on the location of the stream site in relation to stream gage locations where $Q_h$ 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 $Q_h$ is desired, the gage information is used where appropriate in making the estimate, as discussed below.

#### Sites at Gage Locations

Estimates of $Q_h$ values for 230 continuous-record streamflow-gaging stations are presented in tables 1 and 2. When an estimate of $Q_h$ is required at a stream site, refer to table 1 (if the site is unregulated), or to table 2 (if the site is regulated), to determine whether values have previously been estimated for the site.

#### Sites Near Gage Locations

If information is available for an unregulated stream where an estimate is desired, but not at the specific location, a weighting procedure can be employed (Carpenter, 1983). The first constraint to the use of this method is that the drainage area of the ungaged site differ by no more than 50 percent from that of the gaged

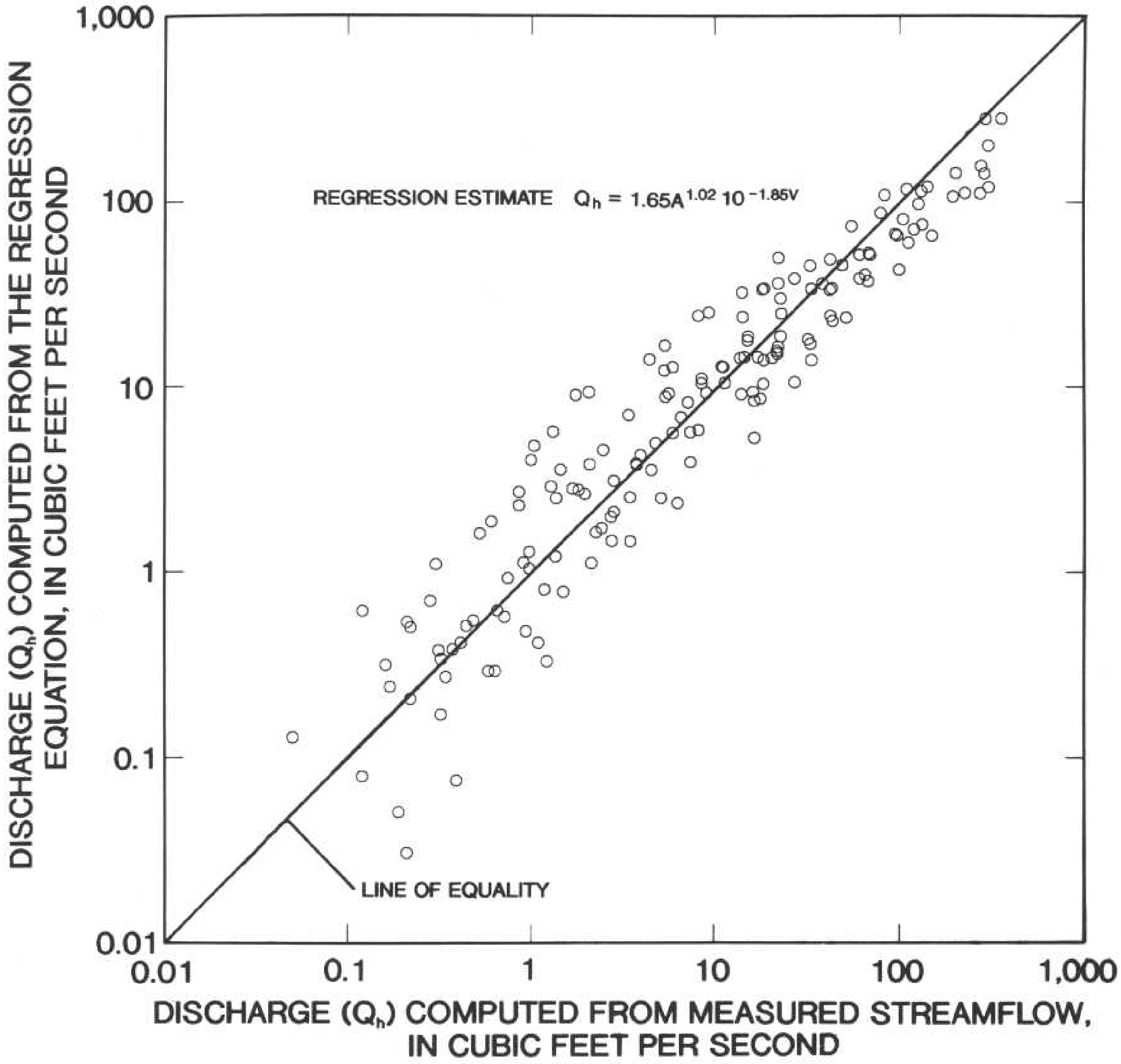


Figure 9.--Scatter plot of harmonic-mean streamflow computed from measured streamflow and from the regression equation for selected continuous-record streamflow-gaging stations in the study area.

site (to minimize the potential for hydrologic dissimilarity between the sites). The second constraint to the use of this method is that the entire drainage basin of the ungaged and gaged sites be within the same variability-index area (pl. 1), because the method assumes a linear relation between the flow values at the gaged and ungaged sites. This is not a valid assumption if the gaged and ungaged sites are affected by different basin geologic characteristics.

The first step in using the weighting procedure is to verify that the above two constraints are satisfied. If so, obtain the value of $Q_h$ computed using streamflow-gaging data at the gage site, $Q_{h\,g\,(d)}$, from table 1. Also, obtain the regression estimate at the gaged site, $Q_{h\,g\,(r)}$, using equation 4 or figure 7. Compute the correction factor at the gaged site ($C_g$) as the ratio of $Q_{h\,g\,(d)}$ divided by $Q_{h\,g\,(r)}$. A correction factor at the ungaged site, $C_u$, is computed based on $C_g$ and the difference in drainage area between the gaged and ungaged site as

$$C_u = C_g - \frac{2\Delta A}{A_g}(C_g - 1), \qquad (5)$$

where

$C_u$ is the correction factor for the ungaged site;
$C_g$ is the correction factor for the gaged site;
$\Delta A$ is the absolute value of the difference in drainage area between the gaged and ungaged site, in $mi^2$; and
$A_g$ is the drainage area of the gaged site, in $mi^2$.

Compute the regression estimate of discharge at the ungaged site, $Q_{h\,u\,(r)}$, using equation 4 (or fig. 7) and multiply this value by the correction factor, $C_u$ from equation 5, to obtain the stream-gage-weighted value of the $Q_h$ estimate at the ungaged site. The equation is

$$Q_{hw} = C_u Q_{hu\,(r)}, \ \textit{if}\ \Delta A < 0.5 A_g \qquad (6)$$

where

$Q_{hw}$ is the stream-gage-weighted $Q_h$ determined at the ungaged site, in $ft^3/s$;
$C_u$ is the correction factor for the ungaged site (from eq. 5);
$Q_{hu\,(r)}$ is the regression estimate of $Q_h$ (from eq. 4), in $ft^3/s$.
$\Delta A$ is the absolute value of the difference in drainage area between the gaged and ungaged site, in $mi^2$; and
$A_g$ is the drainage area of the gaged site, in $mi^2$.

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

#### Sites Between Gage Locations

If a $Q_h$ estimate is desired between two gage locations on the same stream, the value can be estimated by linear interpolation, using the $Q_h$ values and corresponding drainage areas at the two gaged sites. As with the previous method, the technique should not be used where the reach extends over, or is drained by more than one variability-index area. When this condition exists, the relation between the two gaged sites is not linear. The method described previously for unregulated sites near gage locations can, however, be used, if the basin of the stream site where $Q_h$ is desired is in the same variability-index area as one of the two gages.

## 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 equation 4 can be used directly to estimate $Q_h$ . This equation, or the nomograph shown in figure 7, can be used to estimate values of $Q_h$ at ungaged, unregulated stream sites in Kentucky.

Total drainage area of the site of interest should be determined from USGS 7.5-minute topographic maps. The drainage areas for many locations along streams in Kentucky are listed in Bower and Jackson (1981). A map value of streamflow-variability index is obtained from plate 1. The percentage of total drainage area within each streamflow-variability-index area will also need to be determined. Examples of numerical and graphical procedures for obtaining the estimated $Q_h$ values from basins lying entirely within one index area and those in two or more index areas are given in the following sections.

### Sites With Drainage Basins in One Index Area

Estimates of $Q_h$ at an ungaged site that is entirely within the same streamflow-variability index area is computed using the following method. Determine the total drainage area of the site from USGS 7.5-minute topographic maps and the streamflow-variability index from plate 1. Substitute the values into equation 4 as shown below. The example assumes the site has a total drainage area of 155 $mi^2$ and is entirely within the variability-index area of 0.70.

$$Q_h = 1.65A^{1.02}10^{-1.85V} \qquad \text{(eq. 4)}$$

$$Q_h = 1.65\,(155)^{1.02}10^{-1.85\,(0.70)}$$

$$Q_h = 14 \text{ ft}^3\text{/s} \text{ (rounded to the nearest tenth or two significant figures)}$$

A graphical solution can be obtained from the nomograph shown in figure 7. Enter the plot on the abscissa scale at 155 $mi^2$ and proceed upward to the 0.70 streamflow-variability-index curve. From there, proceed across to the ordinate scale to obtain the estimated $Q_h$ value.

### Sites With Drainage Basins in More Than One Index Area

If the drainage area for a desired site location includes more than one variability-index area, the following method is used to estimate values of $Q_h$. Determine the total drainage area of the site and the percentage of the drainage basin located within each of the streamflow-variability-index areas. For this example, assume that an estimate of $Q_h$ is desired for a 300 $mi^2$ basin having 65 percent of the drainage area within a variability-index area of 0.70. The remaining 35 percent is contained in an area having a variability index of 0.90. The numerical solution is as follows. First, obtain a value for $Q_h$ as if all of the basin were contained in the 0.70 variability-index area.

$$Q_h = 1.65A^{1.02}10^{-1.85V} \qquad \text{(eq. 4)}$$

$$Q_h = 1.65\,(300)^{1.02}10^{-1.85\,(0.70)}$$

$$Q_h = 28 \quad \text{ft}^3/\text{s}$$

Next, assume the entire area lies within the 0.90 variability-index area and compute the flow.

$$Q_h = 1.65A^{1.02}10^{-1.85V} \qquad \text{(eq. 4)}$$

$$Q_h = 1.65\,(300)^{1.02}10^{-1.85\,(0.90)}$$

$$Q_h = 12 \quad \text{ft}^3/\text{s}$$

Each of these $Q_h$ estimates can also be obtained graphically using figure 7.

To obtain a solution, multiply each flow value computed above by the corresponding percentage of basin drainage area and sum the resulting values to determine the weighted average $Q_h$ estimate.

$$28\ \text{ft}^3/\text{s}\,(0.65) = 18\ \text{ft}^3/\text{s}$$
$$12\ \text{ft}^3/\text{s}\,(0.35) = 4\ \text{ft}^3/\text{s}$$

$$\text{weighted average } Q_h = 22\ \text{ft}^3/\text{s}$$

## SUMMARY

The values of harmonic-mean streamflow, $Q_h$, were determined at selected streamflow-gaging stations in Kentucky. Daily mean streamflows for the available period of record through the 1989 water year at 230 continuous-record streamflow-gaging stations in Kentucky and just outside Kentucky in bordering States were used in the analysis. Periods of streamflow record affected by regulation were analyzed separately from periods unaffected by regulation. Record extension at short-term stations was accomplished using the MOVE.1 technique to reduce time-sampling error and, thus, improve estimates of long-term $Q_h$ values.

Techniques to estimate $Q_h$ streamflow at ungaged stream sites in Kentucky were developed. A multiple-linear-regression analysis was used to relate $Q_h$ values to drainage-basin characteristics. A regression model that included total drainage area and streamflow-variability index as explanatory variables was defined. Example applications of the model are presented. The regression model has a standard error of estimate of 76 percent and a standard error of prediction of 78 percent.

## REFERENCES CITED


Allen, D.M., 1971, Mean square error of prediction as a criterion for selecting prediction variables: Technometrics, v. 13, p. 469-475.

_____1974, The relationship between variable selection and data augmentation and a method for prediction: Technometrics, v. 16, no. 1, p. 125-127.

Beaber, H.C., 1970, A proposed streamflow data program for Kentucky: U.S. Geological Survey Open-File Report (unnumbered), 48 p.

Bingham, R.H., 1982, Low-flow characteristics of Alabama streams: U.S. Geological Survey Water-Supply Paper 2083, 27 p.

Bower, D.E., and Jackson, W.H., 1981, Drainage areas of streams at selected locations in Kentucky: U.S. Geological Survey Open-File Report 81-61, 118 p.

Carpenter, D.H., 1983, Characteristics of streamflow in Maryland: Maryland Geological Survey Report of Investigations No. 35, 237 p.

Choquette, A.F., 1988, Regionalization of peak discharges for streams in Kentucky: U.S. Geological Survey Water-Resources Investigations Report 87-4209, 105 p.

Conner, Glen, 1982, Monthly, seasonal, and annual precipitation in Kentucky 1951-1980: Kentucky Climate Center Publication Number 25, Bowling Green, Ky., Western Kentucky University, 30 p.

Crawford, N., and Webster, J., 1986, Karst hazard assessment of Kentucky: Sinkhole flooding and collapse: Bowling Green, Ky., Western Kentucky University, Center for Cave and Karst Studies, prepared for the U.S. Environmental Protection Agency, Region IV, Atlanta, Ga., scale 1:1,000,000.

Davis, R.W., Lambert, T.W., and Hansen, A.J., 1983, Surface geology and ground-water resources of the Jackson Purchase region, Kentucky: U.S. Geological Survey Water-Supply Paper 1987, 66 p., 11 pls.

Dempster, G.R., Jr., 1990, National water information user's manual, v. 2, chap. 3; Automated data processing system: U.S. Geological Survey Open-File Report 90-116, 321 p.

Friel, E.A., Embree, W.N., Jack, A.R., and Atkins, J.T., Jr., 1988, Low-flow characteristics of streams in West Virginia: U.S. Geological Survey Water-Resources Investigations Report 88-4072, 34 p.

Grubb, H.F., and Arthur, J.K., 1991, Gulf coast regional aquifer-system analysis--A Kentucky perspective: U.S. Geological Survey Water-Resources Investigations Report 90-4138, 28 p.

Hardison, C.H., 1971, Prediction error of regression estimates of streamflow characteristics at ungaged sites: U.S. Geological Survey Professional Paper 750-C, 9 p.

Hayes, D.C., 1991, Low-flow characteristics of streams in Virginia: U.S. Geological Survey Water-Supply Paper 2374, 69 p.

Hirsh, R.M., 1982, A comparison of four streamflow record extension techniques: Water Resources Research, v. 18, no. 4, p. 1082-1088.

Hutchison, N.E., Shope, W.G., Jr., Morgan, C.O., Williams, O.O., Wilson, T.A., Stuthmann, N.G., Booker, R.E., Price, W.E., Jr., and Meeks, W.C., 1975, WATSTORE--National Water Data Storage and Retrieval System of the U.S. Geological Survey--User's Guide: U.S. Geological Survey Open-File Report 75-426, 7 v., 791 p.

Kentucky Geological Survey, 1980, Physiographic diagram of Kentucky: Lexington, Ky., University of Kentucky, 1 map, scale not specified.

Kentucky Natural Resources and Environmental Protection Cabinet, 1990, Kentucky water quality standards: Kentucky Administrative Regulations title 401, chap. 5.

Lane, E.W., and Lei, Kai, 1950, Streamflow variability: Proceedings of the American Society of Civil Engineers, Transactions, v. 115, p. 1084-1134.

Marquardt, D.W., 1970, Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation: Technometrics, v. 12, p. 591-612.

McDowell, R.C., Grabowski, G.J., Jr., and Moore, S.L., 1981, Geologic map of Kentucky: U.S. Geological Survey, 4 sheets, scale 1:250,000.

McFarland, A.C., 1950, Geology of Kentucky: Lexington, Ky., University of Kentucky, 531 p.

Melcher, N.B., and Ruhl, K.J., 1984, Streamflow and basin characteristics at selected sites in Kentucky: U.S. Geological Survey Open-File Report 84-704, 80 p.

Montgomery, D.C., and Peck, E.A., 1982, Introduction to linear regression analysis: New York, N.Y., Wiley and Sons, 504 p.

Musgrave, G.W., 1955, How much of the rain enters the soil?: U.S. Department of Agriculture, Yearbook of Agriculture: Water, p. 151-159.

Riggs, H.C., 1964, The base-flow recession curve as an indicator of groundwater: International Association of Scientific Hydrology Publication 63, Berkeley, Ca., p. 352-363.

Rossman, L.A., 1990a, DFLOW user's manual: U.S. Environmental Protection Agency, 26 p.

_____1990b, Design stream flows based on harmonic means: American Society of Civil Engineers Journal of Hydraulic Engineering, v. 116, no. 7, p. 946-950.

Ruhl, K.J., and Martin, G.R., 1991, Low-flow characteristics of Kentucky streams: U.S. Geological Survey Water-Resources Investigations Report 91-4097, 50 p.

Searcy, J.K., 1959, Flow-duration curves: U.S. Geological Survey Water-Supply Paper 1542-A, 33 p.

Statistical Analysis System Institute, Inc., 1985, SAS User's Guide: Statistics: Cary, N.C., Statistical Analysis System Institute, Inc., 956 p.

U.S. Army Corps of Engineers, 1981, Water resources development in Kentucky 1981: Louisville District Corps of Engineers, 119 p.

U.S. Department of Agriculture, 1969, National engineering handbook: Soil Conservation Service, Section 4, Hydrology, chap. 9-10.

_____1975, General soil map of Kentucky: Soil Conservation Service Map No. 4-R-34874, scale 1:750,000.

_____1984, Predicting soil loss in Kentucky: Soil Conservation Service Tech. Paper No. 4, Lexington, Kentucky, 152 p.

U.S. Environmental Protection Agency, 1986a, Quality criteria for water - 1986: Office of Water Regulations and Standards, EPA 440/5-86-001, variously paged.

_____1986b, Technical guidance manual for performing wasteload allocations: Book VI, Design conditions: Chapter 1 - Stream design flow for steady-state modeling, Office of Water Regulations and Standards, variously paged.

U.S. Geological Survey, 1958a, Compilation of records of surface waters of the United States through 1950, Part 3-A. Ohio River basin except Cumberland and Tennessee River basins: U.S. Geological Survey Water-Supply Paper 1305, 652 p.

_____1958b, Compilation of records of surface waters of the United States through 1950, Part 3-B. Cumberland and Tennessee River basins: U.S. Geological Survey Water-Supply Paper 1306, 353 p.

_____1964a, Compilation of records of surface waters of the United States, October 1950 to September 1960, Part 3-A. Ohio River basin except Cumberland and Tennessee River basins: U.S. Geological Survey Water-Supply Paper 1725, 560 p.

_____1964b, Compilation of records of surface waters of the United States, October 1950 to September 1960, Part 3-B. Cumberland and Tennessee River basins: U.S. Geological Survey Water-Supply Paper 1726, 269 p.

_____1962-65, Surface water records of Kentucky, 1961-64: U.S. Geological Survey (published annually).

_____1966-75, Water resources data for Kentucky, 1965-74--part 1. Surface-water records: U.S. Geological Survey Water-Data Reports (published annually).

_____1976-90, Water resources data for Kentucky, water years 1975-89: U.S. Geological Survey Water-Data Reports KY-75-1 to KY-89-1 (published annually).

Wetzel, K.L., and Bettandorff, J.M., 1986, Techniques for estimating streamflow characteristics in the Eastern and Interior Coal Provinces of the United States: U.S. Geological Survey Water-Supply Paper 2276, 80 p.

## GLOSSARY

COEFFICIENT OF MULTIPLE DETERMINATION.--The proportion of the variation in the dependent variable explained by the variables in a fitted regression model. Reported values are adjusted for error degrees of freedom.

HARMONIC-MEAN STREAMFLOW.--The reciprocal of the arithmetic mean of the reciprocals of a series of streamflows, or $Q_h = N/1/(Q_1 + 1/Q_2 + ... + 1/Q_N)$, where $N$ is the number of observations of daily mean streamflow.

LEVEL OF SIGNIFICANCE.-- The selected maximum probability of making a Type I error, or rejecting a true null hypothesis (0.05 for this report). Hypothesis tests were used to determine if statistically significant relations existed between dependent and explanatory variables of regression models.

LOCAL DIVERSION.--A localized transfer of water, such as a water-supply withdrawal or wastewater releases, that artificially increase or decrease streamflow in a reach.

MULTICOLLINEARITY.--The presence of a high correlation (near linear dependencies) between two or more explanatory variables of a regression. Multicollinearity causes instability in the estimates of the least squares regression coefficients.

MULTIPLE-LINEAR REGRESSION.--A method of regression wherein a linear relation between a dependent variable and more than one explanatory variable is defined.

ORDINARY-LEAST-SQUARES REGRESSION.--A method of fitting a regression model in which the sum of squared residuals (see residual) is minimized.

PREDICTION SUM OF SQUARES (PRESS) STATISTIC.--A measure of model-prediction error useful in regression-model selection. It is computed by summing the square of the prediction residuals resulting from the series of predictions of each observation by regressions defined using all other observations. Thus, each observation is in turn excluded from the regression data set and is not used in prediction of itself. This process simulates prediction using new data and is a form of data splitting useful for model validation (Allen, 1971, 1974; Montgomery and Peck, 1982).

REGULATED STREAMFLOW.--Streamflow controlled by upstream hydraulic structures such as dams.

RESIDUAL.--The difference between values of harmonic-mean streamflow computed using streamflow-gaging data and values estimated using a regression model.

STREAMFLOW-GAGING STATION.--An installation which provides systematic observations of stage from which streamflow is computed.

STREAMFLOW.--Discharge, measured as the volume of water that passes a given point within a given period of time ($ft^3/s$), that occurs in a natural channel whether or not it is affected by diversion or regulation.

STANDARD ERROR OF ESTIMATE.--A measure of model-fitting error, it is the standard deviation of the residuals of a regression adjusted for error degrees of freedom. Percentage values in this report were estimated using model root-mean-square error, or the square root of the sum of the squares of the residuals divided by the error degrees of freedom, n-k-1, where n is the number of observations and k is the number of explanatory variables in the regression (Statistical Analysis System Institute, 1982) and information from Hardison (1971).

STANDARD ERROR OF PREDICTION.--A measure of model-prediction error, it was estimated as the square root of the PRESS divided by the degrees of freedom for error (Statistical Analysis System Institute, 1982; Montgomery and Peck, 1982; Choquette, 1988). See Prediction Sum of Squares (PRESS) Statistic.

VARIANCE INFLATION FACTOR (VIF).--An indicator of multicollinearity, it is a measure of the combined effect of the dependencies among explanatory variables on the variance of each term in a regression model (Marquardt, 1970; Montgomery and Peck, 1982).

WATER YEAR.--The 12-month period beginning October 1 and ending September 30. It is designated by the calendar year in which it ends.