# GEOLOGICAL SURVEY RESEARCH 1969

## Chapter D

U.S. GEOLOGICAL SURVEY PROFESSIONAL PAPER 650–D

*Scientific notes and summaries of investigations in geology, hydrology, and related fields*

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON: 1969

# ACCURACY OF STREAMFLOW CHARACTERISTICS

By CLAYTON H. HARDISON, Washington, D.C.

*Abstract.*—Streamflow information, which is one of the end products of the Water Resources Division of the U.S. Geological Survey, may be obtained directly from a streamflow record or may be derived from base data by analytical methods. In either case, accuracy goals for the information obtained are needed to test the adequacy of the information and to design the data-collection system upon which the information is based. It is proposed that the accuracy goal for each streamflow characteristic be set equivalent to the accuracy with which that characteristic could be estimated from an observed record of some stipulated length. Curves and tables are presented in this paper to aid in setting such goals.

A streamflow characteristic as used in this paper is a characteristic of a population of flows, at a given site, of which the recorded flows are but a sample. Thus a streamflow information system, such as that operated by the U.S. Geological Survey, can only provide estimates of streamflow characteristics; that is, estimates of what flow may be expected in the future. These estimates are obtained either by analysis of the streamflow data collected at a specific site or by analytical methods, such as correlation and regression, that are used to extend or transfer streamflow information. Each estimate necessarily has an error associated with it that depends on the amount and kind of data and on the analytical methods used.

The accuracy of streamflow information required for specific uses, or a composite accuracy for all uses has never been determined. Studies are only now beginning that may eventually define the accuracy that will properly balance the cost of obtaining additional data and the project benefits accruing from streamflow information of increased accuracy. Until such studies have been made and evaluated, an interim means of setting accuracy goals is needed. It is proposed here that the accuracy goal for each streamflow characteristic be set equivalent to the accuracy that could be obtained from an observed record of some stipulated length.

This paper contains curves and tables from which the accuracy of streamflow characteristics based on records of stipulated lengths can be obtained.

A streamflow characteristic is anything that describes the flow to be expected at a given site. The 50-year peak flow, 20-year low flow, mean annual flow, mean monthly flow, and the standard deviation of the annual and monthly flows are examples of such characteristics. Streamflow characteristics can only be estimated; their true value can never be determined because there is a time-sampling error in every record of streamflow and a model error in every analytical method.

In this paper, the time-sampling errors of streamflow characteristics estimated from gaging-station records are computed using standard statistical methods and assuming no serial correlation between the annual occurrences. The measurement error in determining the discharge of the annual occurrence has been neglected as it is usually small in relation to the time-sampling error and, consequently, has only a minor effect on total error.

In the evaluation of the accuracy of streamflow characteristics given in this paper, the standard error of estimate of the characteristic is used as a single-valued index of accuracy. When the standard error of estimate is thus used, the reader should understand that only 68 percent of the estimates of that characteristic are within one standard error, plus or minus, of the true value. On the basis of theory of errors, he may also understand that about 95 percent of the estimates are within two standard errors of the true value, and that about 99.7 percent of the estimates are within three standard errors.

## ACCURACY OF THE MEAN

The standard error of the mean of any item of hydrologic data can be obtained from the standard deviation of the annual occurrences of that item by the formula $SE=SD/\sqrt{N}$ in which $SE$ is the standard error of the mean of the events, $SD$ is the standard deviation of the annual events, and $N$ is the number of events. The standard error of the mean in percent of the mean is given by

U.S. GEOL. SURVEY PROF. PAPER 650-D, PAGES D210–D214

$$\frac{100\,SE}{\overline{X}}=\frac{100\,SD}{\overline{X}\sqrt{N}}=\frac{100\,C_v}{\sqrt{N}},$$

in which $C_v$, equal to $SD/\overline{X}$, is the coefficient of variation and $\overline{X}$ is the mean of the annual events. This relation holds for all distributions. (The coefficient of variation is a dimensionless index of variability that allows the variability at several locations to be compared. It can also be averaged as a measure of regional variability.)

The relation between $C_v$ and the standard error of the mean, in percent, is shown in figure 1 for five selected values of $N$. If, for example, the coefficient of variation of the annual discharge is 0.3, the standard error of the mean annual discharge based on 10 years of record is shown in figure 1 to be 9.5 percent.

If the logarithms of independent annual events, such as monthly or annual mean flows, are normally distributed, the coefficient of variation of the events can be estimated from $I_v$, the standard deviation of the common logarithms, by the relation

$$C_v^2=\exp[(2.3026\ I_v)^2]-1,$$

which is adapted from an equation given by Chow (1964, p. 17). Values of $C_v$ for selected values of $I_v$ based on this relation are given in table 1.

TABLE 1.—*Relation between standard deviation of the common logarithms,* $I_v$, *and the coefficient of variation,* $C_v$, *in a log-normal distribution*

| $I_v$ | $C_v$ | $I_v$ | $C_v$ | $I_v$ | $C_v$ |
|---|---|---|---|---|---|
| 0.06 | 0.139 | 0.32 | 0.849 | 0.58 | 2.22 |
| .08 | .186 | .34 | .920 | .60 | 2.40 |
| .10 | .233 | .36 | .994 | .62 | 2.58 |
| .12 | .282 | .38 | 1.072 | .64 | 2.78 |
| .14 | .330 | .40 | 1.155 | .66 | 3.01 |
| .16 | .381 | .42 | 1.24 | .68 | 3.26 |
| .18 | .432 | .44 | 1.34 | .70 | 3.53 |
| .20 | .486 | .46 | 1.44 | .72 | 3.82 |
| .22 | .540 | .48 | 1.55 | .74 | 4.15 |
| .24 | .598 | .50 | 1.66 | .76 | 4.51 |
| .26 | .656 | .52 | 1.78 | .78 | 4.91 |
| .28 | .718 | .54 | 1.92 | .80 | 5.36 |
| .30 | .782 | .56 | 2.06 | | |

## ACCURACY OF THE STANDARD DEVIATION

The accuracy of the standard deviation for normal distributions of annual events is discussed first, and then by analogy the resulting equations are applied to log-normal distributions. These expressions for the accuracy of the standard deviation are later used in the computation of the accuracy of an estimated T-year event.

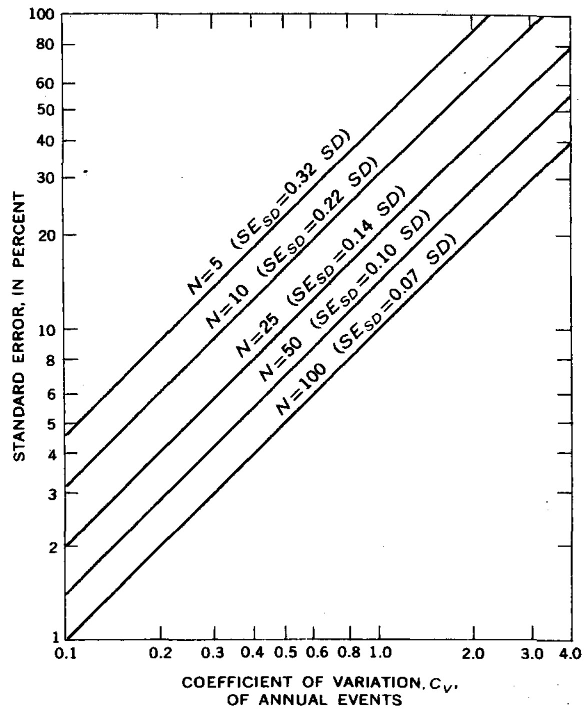


FIGURE 1.—Standard error of mean annual or mean monthly flow. Parameter, $N$, is number of years of record. Standard error of the standard deviation, $SE_{SD}$, for samples from a normal population is shown on parameter line.

For samples drawn from a normal population of annual events, the standard error of the standard deviation is given by

$$SE_{SD}=\frac{\sigma}{\sqrt{2N}},$$

in which $\sigma$ is the standard deviation of the population and $N$ is the number of annual events in the sample. Using $SD$, the standard deviation of the sample, as being the best estimate of the population, standard deviation gives $SE_{SD}=SD\sqrt{2N}$. As the ratio $SE_{SD}/SD$ given by this equation is constant for a given $N$, the evaluation of $SE_{SD}$ in terms of $SD$ can be shown along the five lines in figure 1. As indicated in this figure, the $SE_{SD}$ for a given $N$ is obtained by multiplying the given constant by the standard deviation of the annual events; $SE_{SD}$ in percent is 100 times the given constant.

For log-normal distributions, the standard error of the logarithms, $I_v$, can be substituted for $SD$ in the equations for normal distributions, and the resulting standard error of the standard deviation, $SE_{Iv}$, will be in log units.

The coefficient of variation, $C_v$, is another commonly used measure of variability. If $C_v$ is defined as $SD/\overline{X}$, the standard error of $C_v$ for samples from normal populations is given by

$$SE_{C_v}=\frac{C_v}{\sqrt{2N}}\sqrt{1+2C_v^2}.$$

In this paper, $C_v$ is used as a measure of variability of untransformed data, and $I_v$ is used when a logarithmic transformation is applied to the data.

## ACCURACY OF AN ESTIMATED T-YEAR EVENT

The standard error of a T-year event, such as the 10-year flood, estimated from a record of annual occurrences depends on the type of distribution, the error of the mean event, and the error in the slope of the frequency curve (the standard deviation is a measure of the slope). It is evaluated here by adding the variance due to error in slope to the variance of the mean event. This evaluation purports to give the standard error of T-year events based on samples of size $N$ drawn from a normal population of known standard deviation. The magnitude of such T-year events would, of course, be computed from the mean and standard deviation of each sample, but the appraisal of their accuracy can be based on the standard deviation of the population, which is assumed to be known.

For samples from a normal population, the accuracy of estimated T-year events depends on the accuracy of the computed mean and of the computed standard deviation of the annual events in the sample. For a normal population of known standard deviation, the standard error of the sample mean equals $\sigma/\sqrt{N}$ and the standard error of the sample standard deviation equals $\sigma/\sqrt{2N}$ as discussed in the preceding sections. The variance due to error in slope at a point that is $k$ standard deviation units from the mean may be computed as $(k\sigma/\sqrt{2N})^2$; tables of $k$ values for a normal distribution are given in most statistical texts. By adding the variance of the mean event, $(\sigma/\sqrt{N})^2$, to the variance due to slope we obtain

$$SE_{x_T}=\sqrt{\left(\frac{\sigma}{\sqrt{N}}\right)^2+\left(\frac{k(\sigma)}{\sqrt{2N}}\right)^2}=\sigma\sqrt{\frac{1+k^2/2}{N}},$$

in which $SE_{x_T}$ is the standard error of the T-year event in the same units as $\sigma$.

For samples from a log-normal population, the standard deviation of the logarithms, $I_v$, can be substituted for $\sigma$ to obtain

$$SE_{x_T}=I_v\sqrt{\frac{1+k^2/2}{N}},$$

in which $k$ and $N$ are the same as for normal distributions and $SE_{x_T}$, the standard error of the T-year event, is in log units. This expression for the standard error defines the relations between variability of the annual events and the standard error of 2-, 10-, 20-, and 50-year events for selected lengths of record that are shown in figures 2 and 3.

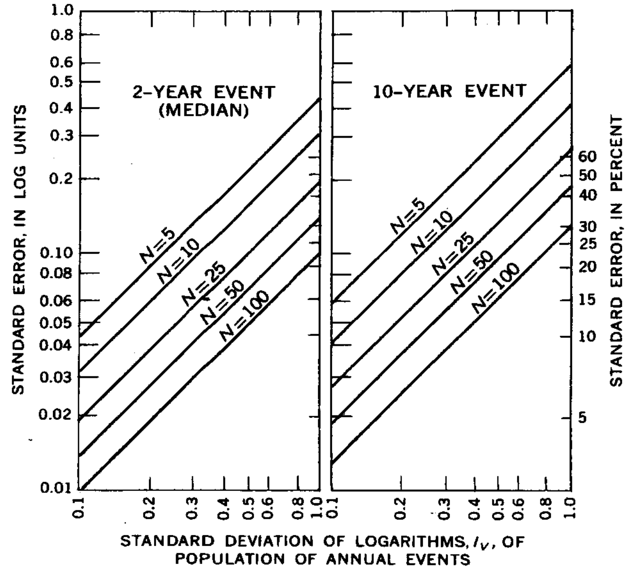


FIGURE 2.—Standard error of median and of 10-year event for log-normal distributions of annual events. Parameter, $N$, is number of years of record.

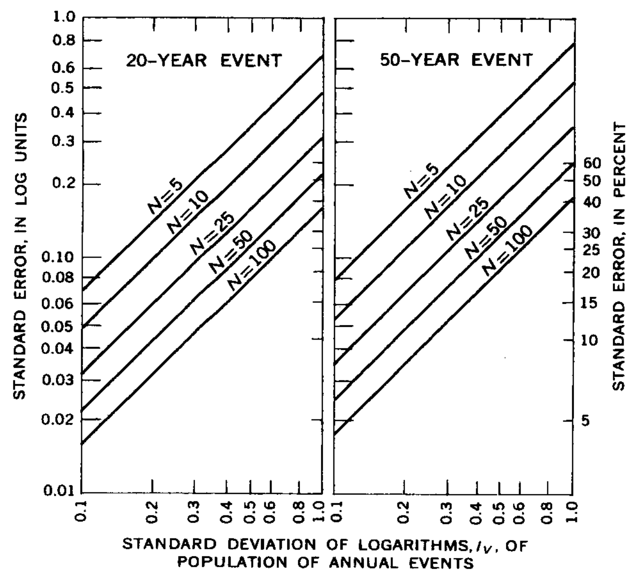


FIGURE 3.—Standard error of 20- and 50-year events for log-normal distributions of annual events. Parameter, $N$, is number of years of record.

The ordinate scales of figures 2 and 3 show the standard error in log units, and auxiliary ordinate scales show the standard error in percent. For a given number of log units, the standard error in percent is the average of the positive departure in percent and the negative departure in percent. The relation between log units and percentage is shown in table 2.

TABLE 2.—*Relation between log units and percentages*

| Log units | Percentage Plus | Percentage Minus | Percentage Average | Log units | Percentage Plus | Percentage Minus | Percentage Average |
|---|---|---|---|---|---|---|---|
| 0.01 | 2.3 | 2.3 | 2.3 | 0.16 | 44.5 | 30.8 | 37.6 |
| .02 | 4.7 | 4.5 | 4.6 | .17 | 47.9 | 32.4 | 40.1 |
| .03 | 7.2 | 6.6 | 6.9 | .18 | 51.4 | 33.9 | 42.6 |
| .04 | 9.5 | 8.8 | 9.2 | .19 | 54.9 | 35.4 | 45.1 |
| .05 | 12.2 | 10.9 | 11.6 | .20 | 58.5 | 36.9 | 47.6 |
| .06 | 14.8 | 12.9 | 13.9 | .21 | 62.2 | 38.3 | 50.2 |
| .07 | 17.4 | 14.9 | 16.2 | .22 | 66.1 | 39.7 | 52.8 |
| .08 | 20.2 | 16.8 | 18.5 | .23 | 69.8 | 41.1 | 55.4 |
| .09 | 23.1 | 18.7 | 20.9 | .24 | 73.8 | 42.5 | 58.1 |
| .10 | 25.9 | 20.6 | 23.2 | .25 | 77.8 | 43.8 | 60.8 |
| .11 | 28.8 | 22.4 | 25.6 | .26 | 81.9 | 45.0 | 63.5 |
| .12 | 31.8 | 24.1 | 28.0 | .27 | 86.2 | 46.3 | 66.2 |
| .13 | 34.9 | 25.9 | 30.4 | .28 | 90.5 | 47.5 | 69.0 |
| .14 | 38.0 | 27.6 | 32.8 | .29 | 95.0 | 48.7 | 71.8 |
| .15 | 41.3 | 29.2 | 35.2 | .30 | 99.5 | 49.9 | 74.7 |

Numerical values of the ratio $SE_{xT}/I_v$ are given in table 3 for use in plotting working copies of the graphs in figures 2 and 3. If the distribution of annual events is normal instead of log-normal, the standard error of the T-year event, in cubic feet per second can be computed from the ratio $SE_{xT}/\sigma$ shown in table 3.

TABLE 3.—*Ratio of standard error of T-year events to standard deviation of annual events for normal and log-normal distributions*

[For normal distributions, figures in table represent $SE_{xT}/\sigma$, in which $\sigma$ is the standard deviation of the annual events; for log-normal distributions they represent $SE_{xT}/I_v$ in which $I_v$ is the standard deviation of the logarithms of the annual events]

| Recurrence interval (years) | Length of record, in years: 5 | 10 | 25 | 50 | 100 |
|---|---|---|---|---|---|
| 2 | 0.447 | 0.316 | 0.200 | 0.141 | 0.100 |
| 10 | .604 | .427 | .270 | .191 | .135 |
| 20 | .686 | .485 | .307 | .217 | .153 |
| 50 | .788 | .558 | .353 | .249 | .176 |

To obtain the standard error in percent for T-year events based on normal distributions of annual events, the standard error in cubic feet per second obtained by use of table 3 must be divided by the average size of the T-year event, which equals $\mu+k\sigma$, in which $\mu$ is the population mean and $\sigma$ is the population standard deviation. As shown by Nash and Amorocho (1966, p. 193), the standard error in percent could thus be related to $C_v$, the coefficient of variation of the annual events, in which case the entries in table 3 multiplied by $100C_v/(1+kC_v)$ give the standard error in percent. Values of $k$ for normal distributions are given in most statistical texts.

For log-normal populations, the standard errors in log units obtained by multiplying the values in table 3 by $I_v$ can be used to define curves such as those shown in figures 2 and 3 or they can be converted into percentage by use of table 2.

If the population from which the annual events are drawn is neither normal nor log-normal and if it is assumed to be a Pearson type-III distribution of known coefficient of skew, the standard error of T-year events can be computed by the following equation:

$$SE_{x_T}=\frac{\sigma}{\sqrt{2N}}\sqrt{b^2k^2+2.828\ rbk+2},$$

in which $b^2=(0.75C_s^2+1)$ varies with $C_s$, the coefficient of skew, $k$ is from Harter's (1969) tables, and $r$ is the correlation coefficient of the sample means and sample standard deviations. This equation is based on the equation for variance of the standard deviation given by Kendall (1952, p. 224) and the relation between $\beta_1$ and $\beta_2$ given by Elderton (1953, p. 57) for a Pearson type-III distribution. Values of $r$ for use in this equation have been determined by sampling to be about 0.3 for $C_s$ of 0.5, 0.5 for $C_s$ of 1.0, and 0.65 for $C_s$ of 1.5. For negative skew coefficients, values of $r$ are opposite in sign to those for the corresponding positive skew coefficient. For log-Pearson type-III populations, $I_v$, the standard deviation of the logarithms of the annual events, can be substituted for $\sigma$ and the coefficient of skew of the logarithms for $C_s$; the resulting $SE_{x_T}$ will be in log units.

The error equation for Pearson type-III distributions is not evaluated here because in most regions and for most streamflow characteristics the standard error of T-year events obtained by assuming a normal or log-normal distribution will be sufficiently accurate for the proposed uses described in the next section.

## USE OF THE RELATIONSHIP CURVES

It is obvious from figures 1, 2, and 3 that any appraisal of the accuracy of streamflow characteristics obtained from a gaging-station record requires that an index of variability of the population of annual events be known or assumed. Thus the index of variability used to enter the curves should preferably represent an average for a region.

Average indices of variability ($C_v$ or $I_v$) for annual streamflow events in two separate regions have been used to obtain the standard errors for 25 years of record shown in table 4. Where appropriate, the type of distribution assumed for the population of annual events is shown by a letter symbol after the variability index. The standard errors for the first two characteristics in each region were obtained by entering the 25-year curve in figure 1 with the indicated $C_v$ (the type of distribution is immaterial). The standard errors for the third characteristic were determined directly from the note along the 25-year curve. The standard errors for the remaining characteristics were determined by entering the appropriate 25-year curves in figures 2 and 3.

TABLE 4.—*Standard error of selected items of streamflow information obtained from 25 years of record in two regions*

[$C_v$ is coefficient of variation of annual events, $I_v$ is standard deviation of the logarithms of the population of annual events, *LN* indicates a log-normal distribution]

| Flow characteristic | Potomac River basin | | | Part 8 in Texas | | |
|---|---|---|---|---|---|---|
| | Variability index | Standard error | | Variability index | Standard error | |
| | | Log units | Percent | | Log units | Percent |
| Mean annual | $C_v$=0.3 | -------- | 6 | $C_v$=0.7 | -------- | 14 |
| Mean monthly | $C_v$= .8 | -------- | 16 | $C_v$=1.4 | -------- | 28 |
| Standard deviation of annual and monthly means | -------- | -------- | 14 | -------- | -------- | 14 |
| 50-year flood | $I_v$= .22 *LN* | 0.078 | 18 | $I_v$= .31 *LN* | 0.11 | 26 |
| Median annual 7-day low | $I_v$= .20 *LN* | .040 | 9 | $I_v$= .47 *LN* | .094 | 22 |
| 20-year 7-day low | $I_v$= .20 *LN* | .061 | 14 | $I_v$= .47 *LN* | .144 | 34 |

The variability indices for the 50-year floods shown in table 4 were obtained by plotting regional flood-frequency curves (peak in ratio to mean annual peak) for each region on log-probability paper and estimating an average $I_v$ from the slope of the curves. The indices for the other items are the averages of those computed for the annual events observed during the period of record at several stations in each region. (The distribution of the logarithms of the annual 7-day low flows at several gaging stations in Texas appears to have a skew coefficient of about −1.0, but for the purpose of this appraisal were used as being log-normal). The variability of the monthly means in each region represent the average monthly $C_v$ at several stations. Tables similar to table 4 could be prepared for other lengths of record by using the appropriate curves in figures 1, 2, and 3 or the appropriate ratios in table 3.

Results such as those shown in table 4 can be used to appraise the results of analytical methods, such as regional regression analysis of streamflow characteristics and hydrologic parameters, in terms of equivalent length of record. In addition, the standard errors provide realistic guides for use in planning surface-water information programs in that they show the accuracy that could be obtained with a feasible length of record. The question of what length of record to use in setting accuracy goals for various classifications and size of stream is outside the scope of this paper.

## REFERENCES

Chow, V. T., 1964, Frequency analysis, pt. 1 *in* Statistical and probability analysis of hydrologic data, sec. 8 *of* Chow, V. T., ed. in chief, Handbook of applied hydrology: New York, McGraw-Hill Book Co., Inc., p. 8-1–8-42.

Elderton, W. P., 1953, Frequency curves and correlation, 4th ed.: Washington, D.C., Harren Press, 272 p.

Harter, H. L., 1969, A new table of percentage points of the Pearson Type III distribution: Technometrics, v. 11, no. 1, p. 177–187.

Kendall, M. G., 1952, The advanced theory of statistics, v. 1, 5th ed.: New York, Hafner Publishing Co., 457 p.

Nash, J. E., and Amorocho, Jamie, 1966, The accuracy of the prediction of floods of high return period: Water Resources Research, v. 2, no. 2, p. 191–198.